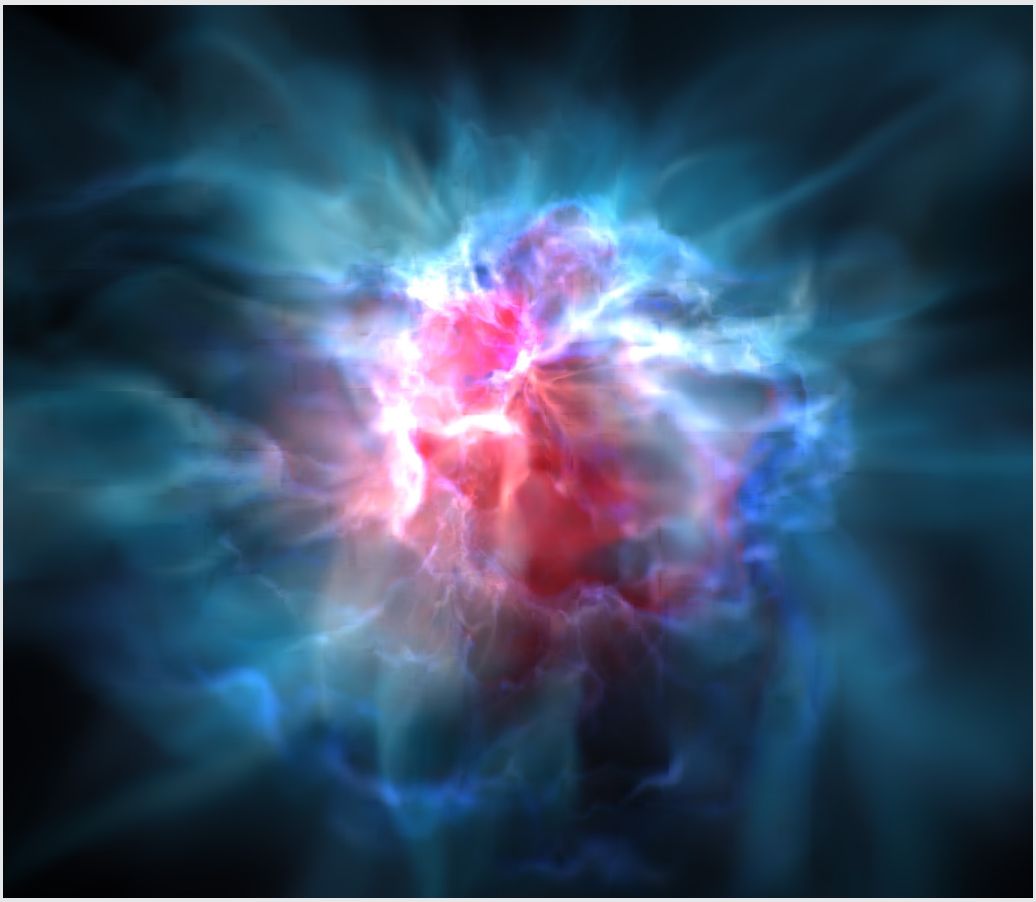
John von Neumann Institute for Computing

NIC Symposium 2025

6 – 7 March 2025 I Jülich, Germany

Ch. Peter, M. Müller, A. Trautmann (Editors)

Proceedings



NIC Series

Publication Series of the John von Neumann Institute for Computing (NIC)
NIC Series Volume 52

Forschungszentrum Jülich GmbH
John von Neumann Institute for Computing (NIC)

NIC Symposium 2025

6 – 7 March 2025 I Jülich, Germany

Ch. Peter, M. Müller, A. Trautmann (Editors)

Proceedings

Publication Series of the John von Neumann Institute for Computing (NIC)
NIC Series Volume 52

ISBN 978-3-95806-793-6

Bibliografische Information der Deutschen Nationalbibliothek.
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der
Deutschen Nationalbibliografie; detaillierte Bibliografische Daten
sind im Internet über http://dnb.d-nb.de abrufbar.

Herausgeber und Vertrieb:	Forschungszentrum Jülich GmbH Zentralbibliothek, Verlag 52425 Jülich Tel.: +49 2461 61-5368 Fax: +49 2461 61-6103 zb-publikation@fz-juelich.de www.fz-juelich.de/zb
Umschlaggestaltung:	Jülich Supercomputing Centre, Forschungszentrum Jülich GmbH
Druck:	Grafische Medien, Forschungszentrum Jülich GmbH
Copyright:	Forschungszentrum Jülich 2025

We thank S. Walch and B. Zimmermann (Universität zu Köln) for the image used in the cover design.

Publication Series of the John von Neumann Institute for Computing (NIC)
NIC Series Volume 52

ISBN 978-3-95806-793-6

Vollständig frei verfügbar über das Publikationsportal des Forschungszentrums Jülich (JuSER)
unter www.fz-juelich.de/zb/openaccess.

Preface

Christine Peter
Department of Chemistry, University of Konstanz, 78457 Konstanz, Germany
christine.peter@uni-konstanz.de

Marcus Müller
Institute for Theoretical Physics, Georg-August University, 37077 Göttingen, Germany
mmueller@theorie.physik.uni-goettingen.de

Alexander Trautmann
John von Neumann Institute for Computing, Jülich Supercomputing Centre,
Forschungszentrum Jülich, 52425 Jülich, Germany
a.trautmann@fz-juelich.de

In a longstanding tradition, the John von Neumann Institute for Computing (NIC) holds biennial symposia, accompanied by proceedings volumes – illustrating the broad range of modern computational science and the advances in high performance and data-intensive computing. Symposium and proceedings thus provide a glimpse into supercomputing-based research at its best and make it accessible both to the general public and to computational scientists across disciplinary boundaries. As such they foster exchange between different fields of natural science and engineering with respect to modern algorithms and computational strategies. To this end, on March 6th and 7th, 2025, computational scientists will again convene in Jülich for the 12th NIC symposium. We are very pleased that this time it is again possible to showcase the breadth of high-performance computing research supported by the NIC with contributions from astrophysics, elementary particle physics, and statistical physics of hard and soft condensed matter, computational chemistry and materials science, as well as computer science, fluid mechanics, and earth system modelling – covering both fundamental research and projects with a strong application orientation. We are also delighted to extend a very warm welcome to our colleagues from the Goethe University Frankfurt who have joined the NIC in 2024 as a new member institution. Together, we will further strengthen research in the field of computational science in Germany and Europe.

The NIC continuously provides the scientific community with essential high-performance computing resources and training. Within the framework of the Gauss Centre for Supercomputing (GCS), the Jülich Supercomputing Centre (JSC) has been operating the modular supercomputer JUWELS (Jülich Wizard for European Leadership Science) since 2020, which is composed of a CPU-based cluster and a GPU-based booster module. Thanks to the excellent training and user support by the technical experts from the JSC, the JUWELS architecture has been widely adopted across disciplines and communities. In particular, porting codes to the booster module and adapting algorithms to the GPU

architecture has been fundamental in getting the disciplines ready for the next generation of GPU-based exascale computing. After the decision by the European High Performance Computing Joint Undertaking (EuroHPC JU) that the Forschungszentrum Jülich is to operate the first exascale supercomputer in Europe, the JSC and the GCS have been preparing for JUPITER (Joint Undertaking Pioneer for Innovative and Transformative Exascale Research). The new system will become available in 2025. To optimally prepare applications and users for JUPITER and to facilitate the transition from current petascale and pre-exascale supercomputers to actual exascale computing, the JSC has launched JUREAP, the JUPITER Research and Early Access Program. In the first phase of 2024, users have participated in the Scalability and Performance Evaluation Phase (SPEP), an open call to test and demonstrate the performance and the scaling of the applications on test architectures. In September 2024, the GCS Exascale Pioneer Call has been initiated with two objectives: the successful projects are given early access to JUPITER during build-up, approximately from January 2025 onwards, and the call distributes JUPITER resources for the time period after the machine is officially operational until the end of October 2025 – thus enabling groundbreaking computational research for the German scientific community. A more detailed overview on JUPITER, the new opportunities that exascale computing opens up to all scientific communities, and in particular the shifts driven by the wave of developments in AI technologies and large foundation models are provided in the introductory article of the proceedings by Thomas Lippert and coauthors "Paradigm Change or Riding the Wave? Exascale-Computers to Train Foundation Models".

As one key element of its mission to promote innovative computing methodologies the NIC also supports several research groups at its member institutions.[a] The NIC research groups cover a broad spectrum of disciplines ranging from high-energy physics to biology, reflecting and reinforcing the research focus of its respective member institutions. Recent results are highlighted in a dedicated section in the proceedings with contributions from the Lattice QCD group (Owe Philipsen, GSI Darmstadt), the Elementary Particle Physics group (Stefan Schaefer, DESY-Zeuthen), and the Computational Structural Biology group (Alexander Schug, Forschungszentrum Jülich).

The subsequent section of the proceedings volume is dedicated to one of the other hallmarks of the NIC, the NIC excellence projects. Generally, NIC computing time is granted by a stringent peer-review process that focuses on the scientific quality of the proposed research. The international pool of expert reviewers and the NIC peer-review board, headed by Dietrich Wolf, play a vital role in sustaining the very high quality of the projects and in fairly and effectively allocating the valuable computational resources. At this point, we want to sincerely thank these reviewers for their engagement and time that they invest in fulfilling this essential task. One important element of this process is awarding the title "NIC excellence project" [b]. It is always a great pleasure to highlight these outstanding projects in a special section. The present proceedings volume features the following NIC excellence projects:

[a] Further information can be found at https://www.john-von-neumann-institut.de/en/research/research-groups
[b] https://www.john-von-neumann-institut.de/en/research/nic-excellence-projects

- Michael Rohlfing, Universität Münster, Spectra of 2D layered materials
- Holger Gohlke, Universität Düsseldorf, Structural dynamics of *apo*, agonist-, and antagonist-bound full-length ETR1
- Gerhard Gompper, Forschungszentrum Jülich, Collective Dynamics of Intelligent Microswimmers
- Johannes Knolle, Technische Universität München, Neural Wave Functions for Materials Physics

In order to further showcase the outstanding research enabled by the NIC that was made possible by the excellent computing environment and user services provided by the JSC, contributions to the symposium and the proceedings have been selected. These contributions provide an account of the projects and give a comprehensive review of the progress that has been made both to the general public and to the funding bodies. The contributions are arranged by scientific topics, thus nicely reflecting the vibrant activity across a broad variety of disciplines. It is our pleasure to thank all the authors of the contributions as well as the experts who wrote the section introductions. Neither the proceedings book nor the symposium would have been possible without the indispensable support from many people within the NIC and the JSC. We are very grateful to Martina Kamps who compiled all the texts and produced this high quality book. Finally, we want to thank Florian Janetzko, Johannes Simonis, and Janina Liebmann for their valuable help in organising the 12$^{\text{th}}$ NIC Symposium in Jülich.

Jülich, March 2025

Christine Peter Marcus Müller Alexander Trautmann

Contents

Astrophysics

Theoretical Chemistry

Elementary Particle Physics

Materials Science

Condensed Matter Theory

Computational Soft Matter Science

Earth and Environment

Computer Science and Numerical Mathematics

Fluid Mechanics

Plasma Physics and Charged Particle Dynamics

Paradigm Change or Riding the Wave? Exascale-Computers to Train Foundation Models

Thomas Lippert[1,2], Mathis Bode[1], Thomas Eickermann[1], Wolfgang Frings[1], Andreas Herten[1], Stefan Kesselheim[1], Benedikt von St. Vieth[1], and Kristel Michielsen[1,3]

[1] Jülich Supercomputing Centre, Forschungszentrum Jülich, 52425 Jülich, Germany
E-mail: th.lippert@fz-juelich.de

[2] Goethe University Frankfurt, Institut für Informatik, 60629 Frankfurt am Main, Germany

[3] RWTH Aachen, Physikzentrum, 52074 Aachen, Germany

A paradigm shift is underway in the field of high-performance computing (HPC). In addition to high-end simulation, the training of foundation models of artificial intelligence (AI) is becoming increasingly important. AI training and supercomputing have more or less become synonymous. Their union will change the way science and industry research complex phenomena and develop new technologies, and there is general consensus that it would be unwise to dismiss this development as mere hype. It is of crucial importance for Germany and Europe to take a leading role in this area and to secure their scientific and technological leadership and sovereignty, especially when it comes to industry and SMEs. JUPITER will take research in computer science and AI to a new level and, until German industry has built its own systems, JUPITER can play the role of an auxiliary bridge for industrial users; plans in this direction are in the works.

1 Foundation Models

The general public's access to ChatGPT at the end of 2022 has made society and science aware of the concept of so-called foundation models (aka base models), which are pre-trained deep learning models built on super-massive datasets. Since then, there has been an exponential increase in pertinent activities in this area, a selective overview of which can be found in Refs. 1,2. The experts agree that the transformative potential of this digital methodology is beyond measure.

Naturally, the focus of society and the public is on large language models (LLMs) and large multimodal models (LMMs), which combine different categories such as text and images or data from other modalities such as audio or other domain-specific data and text. This development leads to what is sometimes, perhaps somewhat euphemistically, referred to as the "democratisation of AI"[3].

It is reassuring that politics in democratic countries is observing the influence of these developments on societal transformations and is fulfilling its duty to implement regulatory measures[4], however, at the same time, the greatest attention must be paid to the potential transformative aspects of the new technology on the worldwide economy. This implies not only the availability of unimpeded access to instruments and infrastructures that enable secure data management of the highest data volumes and the creation of specific models of industrial stakeholders, but above all the sovereignty of all democratic countries in the provision and use of the most powerful computing systems, which cannot be compromised by political or trade policy[4].

Certainly, it is no mere coincidence that the new methodology, which is based on foundational models, was quickly adopted in science. After all, the use of machine learning methods of various origins has become part of the standard methodological repertoire in science, research and technology over many years. The 2024 Nobel Prizes in Physics and Chemistry, which are dedicated to the "invention" and application of the methodology, respectively, provide compelling testimony to its significance[5,6]. However, one should be aware that the use of generative AI and basic models as a tool in information systems and complexity research is still in its infancy.

On the one hand, researchers are convinced that basic models can be created in a wide range of areas where sufficient and sufficiently curated domain-specific data is available. These models can be used to carry out more realistic experiments, make certain types of quantitative studies feasible for the first time and make simulations more accurate or safer. For example, in the area of numerical weather prediction, we can assume that, based on previous weather data, trained forecasting models will exceed or significantly improve the length/accuracy of theoretically modelled forecasts, as has already been demonstrated in pilot studies by Schultz et. al[7] or elsewhere[8].

On the other hand, it is an extremely fascinating prospect that basic models can learn from a general, large and diverse database and develop the ability to deal with the widest range of different tasks from different domains and under different conditions. In this sense, LLMs might be enhanced with domain-specific scientific data and specific data from industry, commerce, finance, or logistics, among other fields, to capitalise on their transformative potential and one-shot learning capabilities in domains where data may be insufficient or too narrow to construct their own base model. This approach might even provide insights or a completely new understanding in areas where, due to limited experimentation or observational possibilities, progress appears to be very slow or could so far not be expected in academic time frames or even periods of lifetimes. A striking example is the understanding of the capabilities of the human brain[9].

All foundation models have an important aspect in common: the AI training of such large models requires unprecedented amounts of compute power of largest supercomputers aka exascale machines.

2 AI Meets Exascale

A central characteristic of foundation models is the existence of scaling laws: Increasing the scales in training leads to predictable improvements in model skills[10]. In the time before 2022, this insight has led to ever larger models being trained, with the surprising result that the limits of performance are apparently only given by the limited computing power [a]. Eventually, the development went exponential and led to revolutionary times: almost every day, the news report on new records set by a few leading US companies that install more than hundreds of thousands of GPUs. These systems are meanwhile capable of training LLMs with more than 400 billion parameters. In terms of AI, the machines used by science and research in the USA tend to come in second, although the two exascale systems Frontier at Oak Ridge National Laboratory and Aurora at Argonne National Laboratory are ranked first and second on the TOP500 list from June 2024.

[a] However, this correlation is of course closely linked to the availability of suitably large, independent data sets.

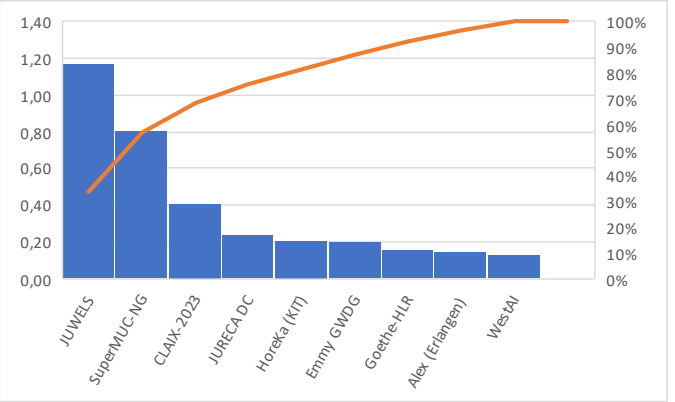


Figure 1. Supercomputer systems in Germany from academia and research capable to train large foundation models as of October 2024. The ordinate counts exaFLOP/s at 16 bit precision.

If we compare this situation with that in Germany, it is striking that there is not a single German company that can provide AI capacities on a similar scale as the US hyperscalers. The situation is not better throughout Europe as to industrial provision. The only systems worth mentioning for AI training in Germany – and in Europe – are provided in the field of science and research, financed by federal and state ministries and the European Commission [b].

Fig. 1 gives an impression of the current (October 2024) machine inventory in Germany with regard to systems from academia and research that can be used for AI training of foundation models in the range of 1 to 10 billion parameters. The most capable system, which has been in user operation since October 2020, is JUWELS at Forschungszentrum Jülich. The machine comprises almost 4000 NVIDIA A100 GPUs. Its suitability for training large AI systems was demonstrated with the 7-billion-parameter entry-level model of OpenGPT-X[11]. The next largest systems are suitable at best for development and test calculations. The figure makes it very clear that for the training of state-of-the-art models in the range of over 100 billion parameters, as is the case with the US giants, a performance increase by two orders of magnitude compared to the capabilities of JUWELS is necessary, and, more than that, several systems of such a size will be needed in Germany.

It is very gratifying to see how far the German scientific community has already adopted the methodology of AI, from long-established machine learning methods to large foundation models. JUWELS has played a major role in this process. The following Fig. 2 demonstrates this development.

For the spring call 2024 applications for resources on JUWELS Booster, more than 40 % of the projects were approved under the AI tag, and in autumn 2024 the figure is already over 50 %. The large foundation models are allocated under “Computer Science” (CS), i.e. more than 25 % of the resources are used for these activities.

In view of the encouraging acceptance of the now somewhat dated JUWELS system, a similar interest in the upcoming JUPITER exascale computer is to be expected. Indeed,

[b] Germany is even more underexposed as to systems suitable for providing AI inference computations.

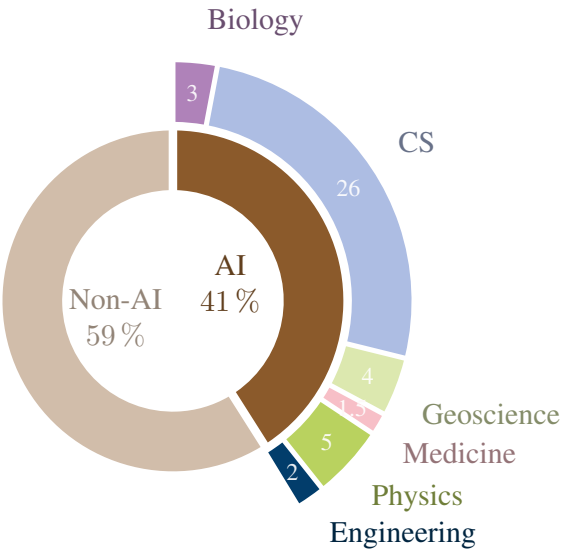


Figure 2. Distribution of allocated compute time per domain on JUWELS Booster in spring 2024.

this is reflected in the Jülich JUREAP initiative, see Sec. 4.3. JUPITER was designed from the ground up to be a system that can handle both large-scale simulation tasks and leading AI simulations, in anticipation of the current development. In fact, a specific balance between network performance and computing power was sought, see Sec. 4, which allows the maximum I/O performance of a GH200 GPU to be utilised. With this setting, JUPITER with its InfiniBand connectivity actually achieves more than 70 % of the point-to-point communication performance of an NVLINK network at comparable latency, and that across the entire machine, while NVLINK networks have so far been limited to 256 GH200 NVIDIA superchips [c].

If the expected AI performance of JUPITER is compared with the total performance available in Germany in October 2024, the system would increase it by a factor of about 20. This becomes frappantly obvious when the data from JUPITER is added to Fig. 1, cf. Fig. 3.

The considerations show that there is currently a lack of sufficient computing power for AI in such a highly industrialised country as Germany. The availability of sufficiently large and numerous training systems in science and industry is sobering and an emergency situation is becoming increasingly apparent; German providers currently play no significant role at all in society and business. JUPITER, primarily a system designated for science and research, will be a step in the right direction. If JUPITER is also partially opened up to the industry for commercial purposes, as being planned by the German Ministry for Education and Research (BMBF), it may be possible to bridge the gap of around three years until there are systems that can be set up and operated by industrial stakeholders in Germany.

[c] The DGX-Helios supercomputer is equipped with four DGX-GH200 systems connected by an Nvidia-Quantum-2 InfiniBand network (Mellanox).

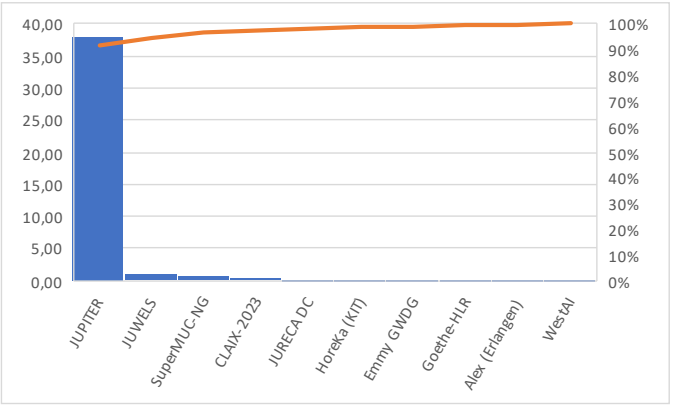


Figure 3. Supercomputer systems in Germany from academia and research capable to train large foundation models expected for May 2025. The ordinate counts exaFLOP/s at 16 bit precision.

In the following, we attempt to provide a quantitative understanding of the system dimensions required for state-of-the-art AI training. To do this, we will first discuss the definition of performance classes and the benchmarks explained in more detail in the following section. We will then take a closer look at the expected specifications of JUPITER and explain how we want to enable users to use the system as quickly as possible with maximum effectiveness.

3 How to Quantify the (AI) Performance of Supercomputers

In order to correctly classify the performance of the upcoming exascale supercomputers for the AI area, it is important to define the performance specification in relation to the area of application. In numerical and stochastic simulation processes, usually, the highest available machine precision, 64 bit, is used. For AI applications, in most cases it is advantageous to use lower precision such as 32 bit, 16 bit or 8 bit. This is explained in more detail in the following.

3.1 What is the Meaning of *Exascale*?

In the field of supercomputing, the term "exascale supercomputer" is defined internationally as a system that achieves a performance of at least 1 exaFLOP/s, or the capacity to perform at least 1 trillion[d], *i.e.* (10^{18}) IEEE 64 bit floating-point operations per second. More precisely, an exascale supercomputer is one that exceeds the threshold of 1 trillion FLOP/s with 64 bit precision when evaluated using a suitable benchmark, particularly the Linpack benchmark for the TOP500 list. For a proper entry in the ranking, it is required that the Linpack code runs with 64 bit precision[e].

[d] Here we refer to the European numbering system. In the US numbering system, 10^{18} is named 1 quintillion.
[e] The first machine worldwide passing the 1 exaFLOP/s threshold was Frontier at Oak Ridge National Laboratory end of 2022.

Following this classification, the European High Performance Computing Joint Undertaking (EuroHPC JU) distinguishes between exascale, pre-exascale and petascale systems. There are currently, as of 2024, five petascale supercomputers and three pre-exascale supercomputers co-financed by EuroHPC JU. Two exascale supercomputers are planned by 2025 and 2026; the first is JUPITER with a performance of at least 1 exaFLOP/s.

The three pre-exascale supercomputers, co-financed and owned by EuroHPC-JU, achieve a maximum peak performance of up to 0.4 exaFLOP/s:

LUMI	CSC-Finland	379 petaFLOP/s	0.38 exaFLOP/s)
LEONARDO	CINECA-Italy	241 petaFLOP/s	0.24 exaFLOP/s)
MareNostrum 5	BSC-Spain	175 petaFLOP/s	0.18 exaFLOP/s)

In this hierarchical scheme, the five petascale systems in EuroHPC are one order of magnitude less powerful. The fastest petascale supercomputer co-funded by EuroHPC achieves a maximum performance of 10.5 petaFLOP/s, i.e. just over 0.01 exaFLOP/s.

JUPITER, with its 1 exaFLOP/s Linpack performance, will instantaneously double the available computing time and thus performance of all other EuroHPC systems combined.

3.2 What Does *AI ExaFLOP/s* Mean?

For most applications in computational science and scientific computing, 64 bit precision as defined by the IEEE 754 standard is the gold standard, resulting in best results for numerical simulations. Data is stored in 64 bit size and computations are performed at the same precision or even higher internal precision.

The IEEE standard specifies in which form the 64 bits are used to represent a number: 52 bits are reserved for the *significand*, representing the fractional value independent of the magnitude; 11 bit are taken for the exponent, to move the significand into proper magnitude; and 1 bit represents the sign of the number (+ or -). Computational researchers have aligned to 64 bit-based computations, and many benchmarks evaluate hardware in this regard, for example the Linpack benchmark.

But not all computations require the full 64 bit precision for valid results. A prominent example are AI-based methods, which interlink layers of neural networks of potentially great depth to create results based on likelihood distributions. This enables the usage of precision lower than 64 bit: 32 bit, 16 bit, or even 8 bit. With corresponding hardware support each reduction of precision allows for more computations in the same time. Bandwidth is no limiter in this regard, as lower precision data words are as efficiently stored and transferred as higher-precision words – see Fig. 4.

Modern GPUs include hardware-acceleration for matrix-based computations. These *Tensor Cores* (NVIDIA), *Matrix Cores* (AMD), or the *Matrix Engine* (Intel) can perform FMA operations (fused multiply-add; a combined instruction for addition and multiplication, $a \times b + c$) with significant higher rate, compared to typical *vector* GPU operations. The Hopper GPU of JUPITER, for example, can execute 64 FMA instructions on each of the four Tensor Cores of each multiprocessor per clock cycle. In total, 66.9 TFLOP/s of FP64 performance can be reached per GPU by utilising Tensor Cores. Because of the nature of these matrix-compute-optimised execution units, a reduction in precision will disproportionately improve the effective throughput significantly. For example, for FP8 precision, the same Hopper GPU will perform with 1978.9 TFLOP/s – about $30\times$ more with an $8\times$ reduction in precision.

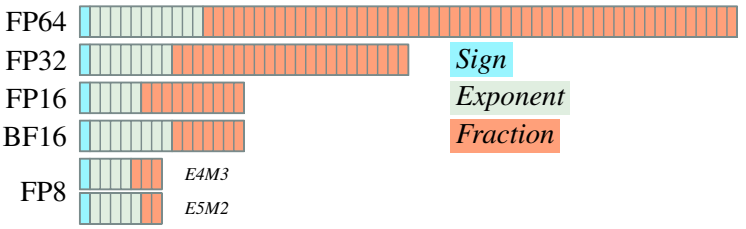


Figure 4. Comparison of precision formats. Each box indicates one bit.

Intermediate steps within the matrix multiplication operation can be performed with higher precision as for example an FP32 accumulate for an FP8 matrices, allowing for more fine-grained optimisation opportunities to retain stability.

Through the lower-precision Tensor Cores, GPUs are a great match for deep neural networks, which are utilising almost exclusively matrix multiplications and are robust in the employed precision (to a point). GPUs are core to the AI revolution, delivering high-throughput compute at excellent energy efficiency. Reduced precision is so deeply linked to AI, that parts of the community started calling the performance reached with lower precision *AI FLOP/s*.

Dedicated benchmarks have been created to test the reduced precision and AI capabilities of compute devices. HPL-MxP[f], for example, extends the Linpack benchmark with an iterative version utilising lower precisions (*mixed precision*, MxP) in intermediate steps; MLPerf[g] is a whole suite of tests for typical AI-patterns with different tasks. Both benchmarks report the results in lists, like the Top500 list for the HPL.

JUPITER Booster features about 24 000 Hopper GPUs as part of the GH200 superchips (see Sec. 4). The following figure estimates the available theoretical performance in dependency of the various precisions. Note: The reference data of NVIDIA is given for a dedicated H100 GPU operated at 1000 W TDP. The TDP of JUPITER Booster Hopper GPUs will be lower. As a first estimate, a reduction to 90 % is applied.

JUPITER Booster will deliver unprecedented performance; for classical numerical simulations, but especially for lower precision and AI-based applications. No HPL-MxP or MLPerf benchmarks have been run for JUPITER Booster, but it is expected that HPL-MxP may reach up to 10 exaFLOP/s performance.

3.3 Guesstimating the Requirements of a 100 Billion Parameter Model

The computing time required to train a state-of-the-art LLM or LMM with say 100 billion parameters (100B), a model that cannot yet be calculated in Germany, can be determined using analogous models that have already been trained. The computing time of LLM OpenGPT-X, which is known with great precision from JUWELS, serves as a reference point for the training time.

In the OpenGPT-X project, the training of several LLMs of size 7B, *i.e.*7 billion parameters, has been carried out on 256 A100 GPUs of the JUWELS Booster. A total of 0.8 mil-

[f] https://hpl-mxp.org/
[g] https://mlcommons.org/benchmarks/

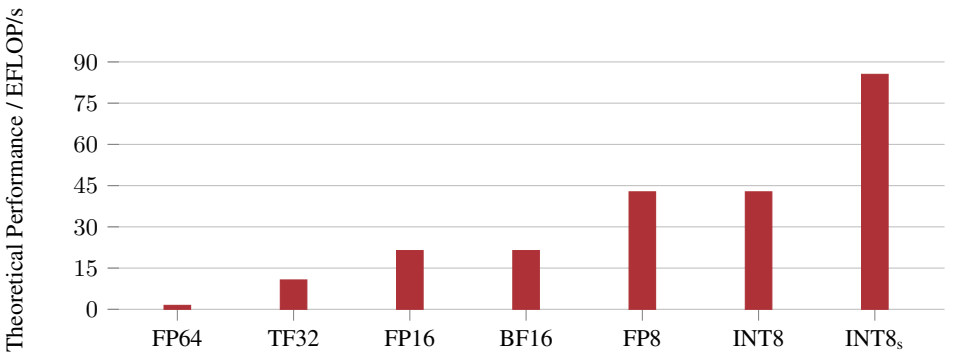


Figure 5. JUPITER Booster GPUs: Theoretical performance depending on utilised precision. $INT8_s$, the last bar, refers to computation with *sparsity*, a feature to exploit sparsity in neural networks if they are well-structured.

lion GPU-h were used. Based on early experiences on the JEDI system, the JUPITER GPUs are more than three times more powerful in practice. For a model with 100B parameters, trained under identical conditions, 3.9 million GPU-h are required on JUPITER. Assuming an average utilisation of half of JUPITER, *i.e.* 12,000 GPUs, training the 100B OpenGPT-X model takes less than 14 days. Training a 100B parameter model in less than two weeks underscores the new possibilities that JUPITER is opening up in the field of cutting-edge AI models.

3.4 Exascale beyond ExaFLOP/s – Interconnect, Storage, and External Connectivity

The performance of an exascale system is dominated by its compute capabilities. Nonetheless, a surrounding hardware ecosystem has to ensure that the compute units can communicate over a network optimised for high bandwidth and low latency without the risk of congestion, *e.g.* when applications scale beyond the GPU memory limits of a single node or even reach tens or hundreds of terabytes of memory and therefore need direct inter-node communication. In addition, the compute units have to be fed with the required data, thus, loading of data for AI training or writing checkpoints at a massive scale becomes the challenge.

Exascale is supposed to increase the requirements of both bandwidth as well as capacity of storage systems. Given the multi-decade experience of large HPC centres with parallel POSIX file systems and often thousands of nodes, scalability is not the main concern. With the utilisation of flash media, accessible by the high performance NVMe protocol, bandwidth aspects are covered by an increasing performance of the underlying storage media. For growing capacity demands, HDD-based storage systems touching hundreds of petabytes have to be provided, especially when it comes to storing multiple versions of datasets, *e.g.* from Earth System Modeling (ESM) or AI communities, which require hundreds of terabytes per domain.

In the past, HPC systems were installed as isolated units, with a focus on the cluster-internal network and its performance on the one hand and a strict eye on software maturity

on the other. However, this sometimes led to a lack of sufficiently frequent implementation of software and, in particular, security updates. Today, the trend is towards openness, both in terms of open software systems and the heterogeneous embedding of systems in a modular hardware environment. It is evident that this requires much stronger focus on security and certification than in the past.

Furthermore, to ensure that cluster-external services, *e.g.* external web services or databases can be used, communication patterns from internal compute nodes to the outside world are to be established more and more. With the raise of AI as the perfect use-case for strong HPC hardware, demand for high-bandwidth access to the internet got even higher. While downloading the hundreds of terabytes of training data can be bandwidth-dominated, limitations on the Domain Name System (DNS) and firewalls as well as rate-limiting on external web servers can also be dominant factors, depending on the local environment.

Optimised retrieval of datasets as well as relaxing limitations of network flows from and to the supercomputers is an active field of research to ensure that usability, but also security aspects are reflected reasonably.

4 JUPITER

With the selection of JSC as the Hosting Entity for the first European exascale supercomputer, challenging decisions had to be made during the design and procurement phase of the system. While the raw performance characteristics had to be defined in December 2021, one year before the generative AI wave got a significant peak with release of the first ChatGPT model by OpenAI, the JUPITER system was tailored to classical and future HPC activities but at the same time anticipating large-scale AI workloads and in particular training of deep neural networks like LLMs. The procurement was geared towards a large mix of applications, synthetic benchmarks, and a set of high-scaling applications that can utilise the full final exascale system. The technologies selected for JUPITER, which are briefly described in the following section, are the result of the best proposed solution for executing the JUPITER Benchmark Suite[12] as well as additional technical requirements, including energy consumption.

4.1 Technical Description

The JUPITER system is provided by the ParTec/Eviden supercomputer consortium, using the Eviden-Bull Sequana XH3000 hardware architecture for the compute intensive components of the system. The XH3000 is a direct-liquid-cooled rack solution, allowing for highest density as well as energy efficiency. The final system will implement the dynamic Modular Supercomputing Architecture (dMSA) and is powered by the highly-flexible JUPITER Management Stack integrating ParaStation Modulo, Eviden SMC xScale and JSC's xOPS software environment.

At the core of the JUPITER hardware is the accelerator-based JUPITER Booster module. It is utilising GPUs to achieve the best possible performance while keeping the energy consumption at lowest possible level. As a result of the aforementioned benchmark suite, the NVIDIA Grace Hopper (GH200) superchip was chosen as a combined CPU-GPU solution, integrated into the XH3000. With roughly 6000 compute nodes, the system is one of

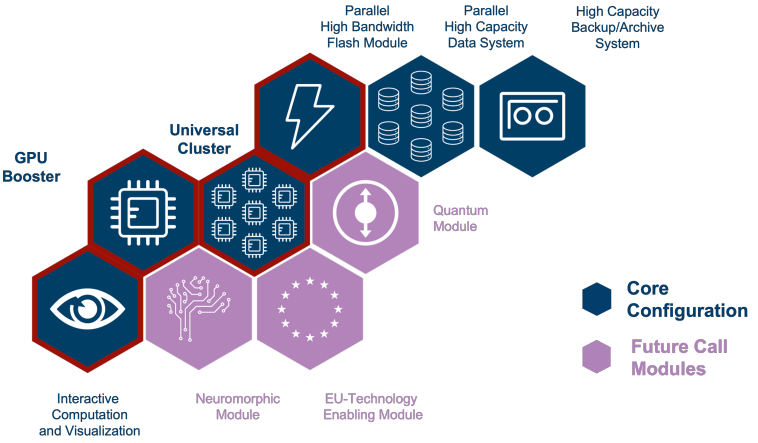


Figure 6. The JUPITER Modules.

the largest supercomputers to date. Each node incorporates four GH200 chips, four Grace CPUs paired with a Hopper-based GPU, as well as four high-speed NVIDIA InfiniBand NDR200 links as shown in Fig. 7, integrated into 125 XH3000 racks and interconnected by a vast InfiniBand DragonFly+ topology. JUPITER Booster is one of the largest, coherent AI training machines available in the world.

For JUPITER, an efficiency-optimised version of the GH200 chip with a CPU-GPU power draw of not more than 680 W was chosen. Depending on the applied performance optimisations, each chip can achieve more than 47 TeraFLOP/s HPL performance (FP64) or 700 TeraFLOP/s FP16. With 48 nodes per XH3000 rack, this renders to a power consumption of roughly 140 kW per rack and 17 MW of the final JUPITER Booster.

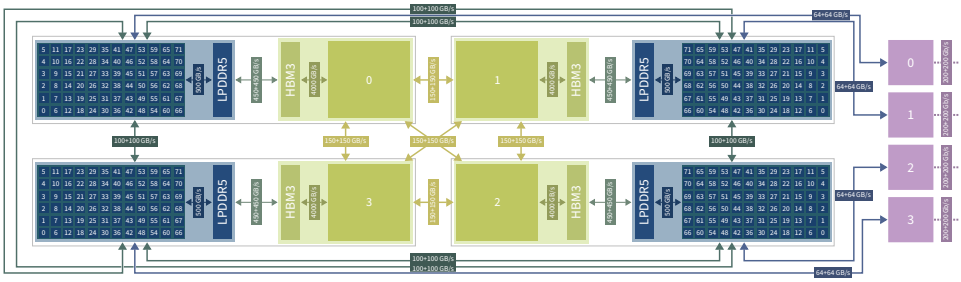


Figure 7. The JUPITER Booster Compute Node Design.

The JUPITER Cluster module is complementing the Booster by providing a general-purpose CPU-based architecture. It focuses on a high Byte-per-FLOP ratio to ensure that CPU-based applications can achieve the highest possible performance. The Cluster is utilising the Rhea1 processor, with roots in the European Processor Initiative and commer-

cialised by SiPearl. Rhea1, like Grace, implements the ARM instruction set architecture; Rhea1 utilises HBM memory to achieve highest-possible memory bandwidth.

Cluster and Booster are supported by multiple storage systems. The 29 PB ExaFLASH module, based on the latest generation IBM Storage Scale System (SSS) 6000 utilising NVMe media for excellent performance characteristics, is used for semi-temporary storage (SCRATCH). The 310 PB ExaSTORE module is provided for large datasets (DATA and HOME), also based on IBM SSS6000 and utilising HDD media. The directly accessible storage systems are supported by the 370 PB ExaTAPE tape infrastructure for archive and backup purposes (ARCHIVE). The HDD and tape systems will be upgraded during the JUPITER lifetime, depending on actual demand.

Thanks to its dMSA and large InfiniBand fabric, JUPITER is able to readily integrate future technology modules such as quantum computing and neuromorphic modules. The dMSA is supported by the novel Modular HPC Datacenter (MDC), which is the home of the JUPITER system on the Forschungszentrum Jülich campus.

4.2 A Remark on Records in and Demand for Energy Efficiency

In addition to the TOP500 list, which focuses on the pure computing performance in solving mathematical problems in modelling and simulation, since 2013 there has been a ranking for systems which can achieve the highest amount of computations per energy consumed. This list is using the Linpack benchmark and defines a power measuring methodology to generate a competitive GigaFLOP/s/W ranking. This list illustrates the energy efficiency of supercomputers of the TOP500 list and generates an incentive for optimising hardware as well as software for energy efficiency.

The JUPITER Exascale Development Instrument (JEDI), the first JUPITER component, was deployed in spring 2024 to support software preparation for both users and operators of the final JUPITER system. To evaluate the decisions made during the procurement for the final system, putting a focus on energy efficiency, the performance of JEDI was measured and is in the current Green500 list (June 2024) at 1st rank, making this module currently the most energy-efficient supercomputer in the world when running the Linpack benchmark.

Given the ever-increasing demand for computing resources by the research community, accelerated by public and private demand for AI computing/training time and thus access to power-hungry computing resources, energy efficiency of the utilised hardware is key to ensure that the impact on the environment, but also operational expenses, are kept at a reasonable level. This not only applies to the energy used for computations, but also for the surrounding infrastructure and in particular cooling of IT systems. JUPITER is designed to dissipate the heat it generates into a warm-water cooling system which uses free cooling, so no additional energy for decreasing cooling-loop temperatures over long periods of a year is needed. At the same time, the generated heat can be extracted for heat re-use.

4.3 JUREAP

JUREAP, the JUPITER Research and Early Access Programme, combines many activities that are designed to help JUPITER get off to an immediate and successful start: Key applications are optimised for exascale at an early stage, the JUPITER software environment

Figure 8. The JUPITER Exascale Development Instrument – JEDI.

is thoroughly tested and continuously expanded, and the system's structure is permanently monitored using application-related benchmarks. All this is intended to comprehensively prepare users and the system for production operation and ensure that the system can be used effectively right from the outset. Due to its high complexity, JUREAP requires HPC experts and domain scientists to work hand in hand.

The first phase of JUREAP, the Scalability and Performance Evaluation Phase (SPEP), started in January 2024 with an open call. Suitable applications were then integrated into a CI/CD environment *exaCB* and extensively tested with regard to their node performance on JEDI as well as large-scale capabilities on JUWELS boosters. In total, there were more than 100 interested applicants that were invited to integrate with exaCB.

More than 20 % of the SPEP applications are AI-related. On the one hand, there were applications from the "core AI" area, such as foundation models, large language models, and generative applications, which will substantially benefit from the high AI performance of the GH200 superchips. JUPITER will allow significantly more data to be considered for training in significantly less time, thus overcoming the physical limitations of current petascale and pre-exascale supercomputers. First evaluations show almost linear scaling and up to 3 to 4 times faster training speed compared to NVIDIA A100 GPUs.[h]

On the other hand, AI is also becoming an essential tool in more and more "classical" simulation workflows. Exascale simulations have the potential to generate extremely large amounts of data that are increasingly difficult to handle. Often, this is only possible using AI-supported data-driven approaches. In this use case, AI thus acts as an enabler to exploit the full potential of exascale simulations.

[h] In the most favourable case, one can expect a performance increase of a factor of 25 to 30 compared to JUWELS.

5 A Sea Change

The quest to use computers and data to understand the world's most complex phenomena is in turmoil. To use an astrophysics metaphor, large-scale AI methods are spreading at an almost inflationary rate, driving the construction and use of ever larger supercomputers that are suitable for them.

Although the initial driving force in public perception seems to come almost exclusively from the business world – the hyperscalers send their regards – , the Nobel Prize 2024 in Physics makes it very clear where it originates from: It is not science that is riding a wave created by industry, but science has helped to make AI a useable and most likely very profitable technology.

In fact, large-scale AI basic models are synonymous with HPC and have become the best-known HPC applications, even more so than weather forecasting. Without HPC computing, HPC storage and HPC networking, there would be no large-scale AI.

The extremely high and fast changing level of development of the fastest supercomputers is no coincidence either. Rather, it is the result of most consistent development efforts of the simulating and data processing sciences in association with leading manufacturers of processors, communication technologies, storage systems, and the ingenuity of integrators over the last 30 years. In this sense, the hyperscalers' AI gold rush benefits greatly from science without contributing much itself, and they are certainly not thinking of paying royalties.

The challenges faced by computational research in science and industry are obvious: The hyperscalers' supply of AI training and inference capacity – viewed on a global scale – is developing towards a monopoly, caused in particular through a GPU market being stirred up by their rigorous acquisitions, which the publicly funded organisations can hardly follow. As a consequence, users run the risk of losing their independence and countries their sovereignty, only being able to purchase their needs from commercial cloud services, with the result that they would have to endure a restrictive software service dictate and having restricted control over their prompt management while being cut off from their so fruitful interplay with machine development.

Given this situation, there is no doubt that we are on the threshold of a paradigm shift as to the future of our HPC-based methodologies. For science and research, it is important that this shift goes in a direction in which we do not lose control over our data, the prompt management, and own AI training capabilities. In particular, we need to be able to create LLMs, LMMs and other foundation models that are trained on open and publicly controlled data sets. Through our involvement in LAION[i], we were able to show how important the aspect of public accessibility of data sources is, especially in the field of LLMs and LMMs.

JUPITER is Germany's and Europe's ticket into the exclusive club of sovereign AI users. JUPITER will boost computational science and AI research to unprecedented heights, and until Germany's industry has built its own systems, JUPITER can provide initial support for industrial users.

With systems like JUPITER science and research in Germany and Europe can take up the challenges with confidence.

[i] `https://laion.ai/`

References

1. OECD, *Artificial Intelligence in Science*, 2023, https://www.oecd-ilibrary.org/content/publication/a8d820bd-en.
2. A. B. Rashid, MD A. K. Kausik, *AI revolutionizing industries worldwide: A comprehensive overview of its diverse applications*, Hybrid Advances **7**, 100277, 2024, https://www.sciencedirect.com/science/article/pii/S2773207X24001386.
3. C. J. Costa, M. Aparicio, S. Aparicio, J. T. Aparicio, *The Democratization of Artificial Intelligence: Theoretical Framework*, Applied Sciences **14**, 18, 2024, doi:10.3390/app14188236.
4. The Future of Life Institute, *The EU Artificial Intelligence Act: Up-to-date developments and analyses of the EU AI Act*, https://artificialintelligenceact.eu/.
5. The Nobel Prize organisation, *The Nobel Prize in Physics 2024*, https://www.nobelprize.org/prizes/physics/2024/summary/.
6. The Nobel Prize organisation, *The Nobel Prize in Chemistry 2024*, https://www.nobelprize.org/prizes/chemistry/2024/summary/.
7. L. H. Leufen, F. Kleinert, M. G. Schultz, *O3ResNet: A deep learning – based forecast system to predict local ground-level daily maximum 8-Hour average ozone in rural and suburban environments*, Artificial Intelligence for the Earth Systems **2**, 3, 202, e220085, American Meteorological Society, https://journals.ametsoc.org/view/journals/aies/2/3/AIES-D-22-0085.1.xml.
8. P. T. Vonich, G. J. Hakim, *Predictability limit of the 2021 Pacific Northwest heatwave from deep-learning sensitivity analysis.*, Geophysical Research Letters **51**, e2024GL110651, 2024, doi:10.1029/2024GL110651.
9. K. Amunts, Th. Lippert, *Brain research challenges supercomputing*, Science **374**, 1054-1055, 2021.
10. C. Schuhmann, R. Beaumont, R. Vencu, C. Gordon, R. Wightman, M. Cherti, T. Coombes, A. Katta, C. Mullis, M. Wortsman, J. Jitsev et al., *Laion-5b: An open large-scale dataset for training next generation image-text models*, Advances in Neural Information Processing Systems **35**, 25278-25294, 2022.
11. https://opengpt-x.de/en/about/
12. A. Herten, S. Achilles, D. Alvarez, J. Badwaik, E. Behle, M. Bode, T. Breuer, D. Caviedes-Voullième, M. Cherti, A. Dabah, S. El Sayed, W. Frings, A. Gonzalez-Nicolas, E. Gregory, K. Mood, T. Hater, J. Jitsev, C. John, J. Meinke, C. Meyer, P. Mezentsev, J. Mirus, S. Nassyr, C. Penke, M. Römmer, U. Sinha, B. von St. Vieth, O. Stein, E. Suarez, D. Willsch, and I. Zhukov, *Application-Driven Exascale: The JUPITER Benchmark Suite* in 2024 SC24: International Conference For High Performance Computing, Networking, Storage And Analysis SC, 468-512, 2024, https://doi.ieeecomputersociety.org/10.1109/SC41406.2024.00038

The NIC Research Groups

Constraining the QCD Phase Diagram

Francesca Cuteri[1], Alfredo D'Ambrosio[1,2], Michael Fromm[1], Reinhold Kaiser[1,2], Owe Philipsen[1,2], and Alessandro Sciarra[1]

[1] Institut für Theoretische Physik, Goethe Universität Frankfurt,
Max-von-Laue-Str. 1, 60438 Frankfurt, Germany
E-mail: philipsen@itp.uni-frankfurt.de

[2] John von Neumann Institute for Computing (NIC), GSI, Planckstr. 1, 64291 Darmstadt, Germany

Quantum Chromodynamics is the fundamental theory to describe the physics of strongly interacting particles, the hadrons. Its phase diagram plays an important role for the interpretation of experimental results in nuclear physics, heavy-ion collisions and nuclear astrophysics. Due to a severe fermionic sign problem, QCD at finite matter density cannot be simulated directly, and little reliable information on the phase diagram is available. Here we report on a long-term project to constrain the QCD phase diagram by simulating the theory away from its physical parameter values, in order to understand how the physical situation as a special case is embedded in the general parameter space of the theory. After producing data over many years, the first phenomenologically relevant bounds on the location of a possible chiral phase transition at finite baryon density are beginning to emerge.

1 Introduction

The fundamental theory describing the strong interactions in the framework of the Standard Model of Particle Physics is Quantum Chromodynamics (QCD). This quantum field theory is formulated in terms of quark and gluon fields, which are the elementary constituents of pions, kaons, nucleons etc., i.e., the strongly interacting particles (hadrons) that are responsible for the nuclear physics in atoms as well as within massive stars. For the purpose of such applications it is sufficient to restrict attention to the three light quark flavours, the u- d- and s-quarks. In the limit of massless u, d-quarks, $m_{u,d} = 0$, the theory displays a so-called chiral symmetry: it looks the same when the u- and d-degrees of freedom are exchanged or mixed. However, the vacuum state of this theory does not show this symmetry, which is then said to be spontaneously broken. As a consequence the Goldstone theorem predicts the existence of three massless bosons, the pions. In nature the u, d-quarks are very light but not exactly massless, representing a small distortion from the chiral (massless) limit. Correspondingly, the pions are not exactly massless, but carry a mass (~ 135 MeV) much smaller than that of ρ-mesons (~ 770 MeV) or nucleons (~ 930 MeV), thus identifying them as "would-be" Goldstone bosons.

When hadronic matter is either heated, such as in the early universe or heavy-ion collisions, or densely packed, such as in neutron stars, its properties are expected to change as a function of temperature and density. When temperature and/or matter density become large, the spontaneously broken chiral symmetry gets dynamically restored, and the properties of hadronic matter are believed to change fundamentally. In particular, the coupling strength of the quark-gluon-interaction reduces and one expects the hadrons to eventually melt into a quark gluon plasma with different properties, as sketched in the putative QCD phase diagram Fig. 1 left. Several heavy-ion experiments as well as astronomical observations of neutron stars and their mergers are under way to explore different regions of the

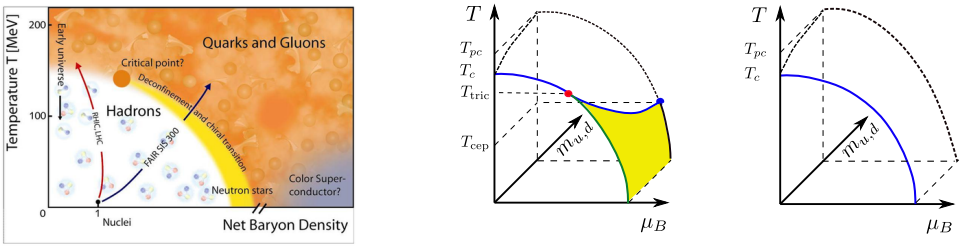


Figure 1. Left: Sketch of the expected phase diagram of QCD matter as a function of temperature and density. Middle + Right: Connection of the putative QCD phase diagram for physical light quark masses to the chiral limit with $m_{u,d} = 0$, in the front plane.

phase diagram. In the chiral limit of massless u, d-quarks, the chirally symmetric and broken phases must be separated by a non-analytic phase transition. For cold and dense matter several simplified models predict this transition to be of first order and proceed by bubble nucleation, similar to liquid gas transitions. On the other hand, for a hot and dilute gas one expects the transition to be of second order, i.e. to happen everywhere at the same time, such as the spontaneous magnetisation in a ferromagnet. This is sketched qualitatively in Fig. 1 middle, where the change from a first-order to a second-order phase transition is marked by a tricritical point. When the light quark masses are non-zero, as they are in nature, the chiral symmetry is broken explicitly. In this case there is never a fully chirally symmetric regime, and the regions with more or less chiral symmetry must be analytically connected. The second-order transition in this case is replaced by a smooth and analytic crossover, while there still can be remnants of a first-order transition, which then terminates in a critical endpoint. This is the scenario expected by a large part of the theoretical community based on various model studies[1–3].

However, the true phase diagram if QCD is still unknown today. The vacuum properties of QCD can be numerically simulated on a discretised space-time (lattice QCD) to give accurate predictions for the hadron mass spectrum, hadronic decay constants and many other properties observed in nature. Thermal lattice QCD can also be simulated, and we know that indeed the chiral transition at physical quark masses corresponds to an analytic crossover[4]. However, at finite density a fermionic sign problem prohibits Monte Carlo simulations, and not much is known for the dense situation with baryon chemical potential $\mu_B \neq 0$. This motivates the approach pursued in our NIC research group, namely to study how the nature of the QCD chiral transition changes away from the physical point as a function of the number of quark flavours, their masses and imaginary baryon chemical potential, which is unphysical but can be simulated straightforwardly. The parameter dependence of the QCD transition constrains the nature of the transition at the physical point, which one may hope to infer once sufficient information is available. In particular, one is interested in the nature of the transition in the chiral limit, in order to check the assumptions going into the scenarios in Fig. 1. As we shall see, all current lattice results are so far also consistent with the scenario shown in Fig. 1 right, where the transition is second order all the way in the chiral limit, and completely disappears into a crossover.

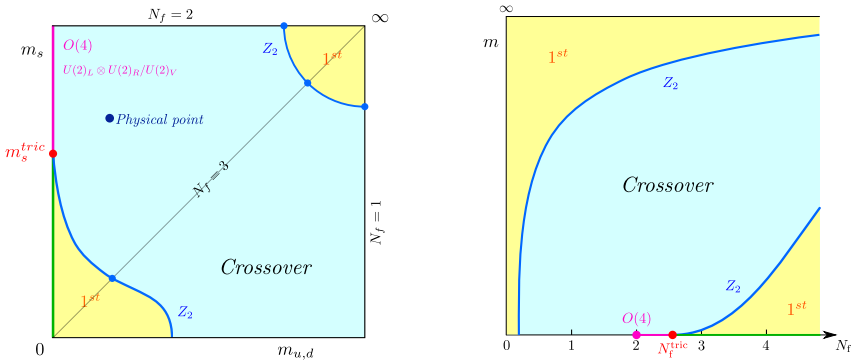


Figure 2. Left: The order of the QCD thermal transition as a function of the quark masses; possible scenario proposed in model studies[5] and observed on coarse lattices. Right: The analogous scenario for strictly degenerate quarks, interpolated to non-integer N_f [12].

2 The Columbia Plot at $\mu_B = 0$

Even at zero baryon density, QCD with massless quarks cannot be simulated because of diverging matrix inversions in that case. One can only approach this limit by simulations with gradually decreasing quark masses, at the expense of drastically increasing computational effort. The nature of the QCD transition as a function of quark masses at zero density is qualitatively shown in a so-called Columbia plot, Fig. 2. One possibility, which was first predicted nearly 40 years ago by the epsilon expansion applied to linear sigma models with symmetry breaking patterns as QCD[5], is schematically shown in Fig. 2 left. The theories with $N_f = 1, 2, 3$ mass degenerate quark flavours correspond to the right and top boundary lines, and the diagonal, respectively (quarks with infinite mass do not contribute to the dynamics and decouple from the theory). In the limits of $m_{u,d} = 0$ (left boundary), there are non-analytic phase transitions due to the spontaneous breaking of the chiral symmetry for massless quarks. Simulations at finite quark masses on coarse lattices appeared to confirm the predicted first-order region for $N_f = 3$[6,7]. On the other hand, one observes an analytic crossover at intermediate quark masses, with a second order boundary line separating these regions, which has been shown to belong to the $Z(2)$ universality class of the 3d Ising model[8,9]. Another possibility, which has also been seen in the staggered discretisation on coarse lattices[10], is for the first-order chiral transition region to extend all the way to the $N_f = 2$ theory in the upper left corner. Over the years the chiral critical boundary line was found to strongly recede towards smaller quark masses with decreasing lattice spacing[11]. However, it has remained open whether the chiral phase transition for two quark flavours is of first or second order in the continuum limit. Our original motivation for this long term project was to distinguish between these two scenarios.

3 Computational Strategy and Numerical Results

Rather than trying to approach the continuum chiral limit for a fixed number of quark flavours and masses, our strategy is to search for the tricritical point separating parameter

regions with first-order and second-order chiral transitions. To this end it is advantageous to change variables and consider strictly mass-degenerate quarks. The QCD partition function $\mathcal{Z}$ is expressed as a path integral over the gluon fields U, including a determinant of the Dirac operator for each species of quarks,

$$\mathcal{Z} = \int \mathcal{D}U \ (\det D[U; am])^{N_f} \exp\{-S_g[U]\} \ , \tag{1}$$

where $S_g[U]$ is the discretised gauge action and $D[U]$ the fermion Dirac operator, which we employ in the standard unimproved staggered discretisation. The bare microscopic parameters of the partition function are the lattice gauge coupling $\beta = 6/g^2$ and the bare quark mass am, given in units of the lattice spacing a. The temperature T is specified by the inverse temporal lattice extent

$$T = (a(\beta)N_\tau)^{-1} \ . \tag{2}$$

On a lattice with given N_τ, temperature is tuned by changing the lattice spacing a indirectly via the coupling $\beta(a)$. A continuum limit at fixed temperature implies $a \to 0, N_\tau \to \infty$, and larger values of N_τ imply smaller lattice spacings.

In this degenerate quark formulation one can continuously vary between $N_f = 2, 3$ by tuning N_f to non-integer values, rather than tuning $\infty \geq m_s \geq 0$. The Columbia plot from Fig. 1 left then gets replaced by the version on the right. A tricritical strange quark mass m_s^{tric} in the former version translates into a tricritical value $2 \leq N_f^{\mathrm{tric}} \leq 3$ in the latter. The chiral critical line is known to enter the tricritical point as[13]

$$am_c\big(N_f(N_\tau), N_\tau\big) = D(N_\tau)\big(N_f - N_f^{\mathrm{tric}}(N_\tau)\big)^{5/2} + \dots \ . \tag{3}$$

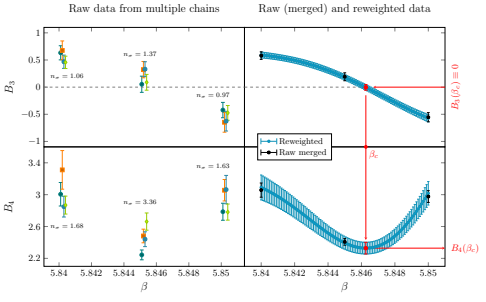
The benefit of this changed formulation is its generality, i.e. the tricritical point can be located at *any* value of N_f, in contrast to the model expectations on which the scenario in Fig. 2 is based. The task then is to follow the chiral critical line, which is known on coarse lattices, through parameter space as the lattice is made finer.

All numerical simulations have been performed using the publicly available OpenCL-based code `CL²QCD`, which is optimised to run efficiently on AMD GPUs and contains an implementation of the RHMC algorithm for unimproved rooted staggered fermions. Version `v1.0`[14] has been employed for simulations on smaller N_τ on the L-CSC supercomputer at GSI, while version `v1.1`[15] has been run on the HLR supercomputer at Goethe University to run the most costly simulations.

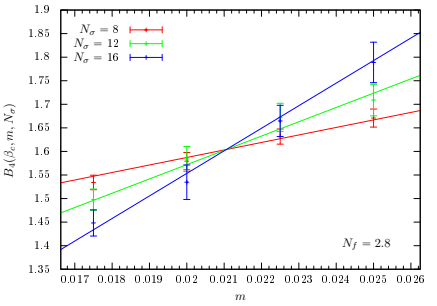
We locate and identify the nature of phase transitions by finite size scaling analyses of standardised moments of the distribution of an appropriate order parameter,

$$B_n(\beta, m, N_\sigma) = \frac{\langle(\mathcal{O} - \langle\mathcal{O}\rangle)^n\rangle}{\left\langle(\mathcal{O} - \langle\mathcal{O}\rangle)^2\right\rangle^{n/2}}. \tag{4}$$

For the chiral phase transition this is the chiral condensate, $\mathcal{O} = \overline{\psi}\psi$. We first locate the phase boundary β_c by the condition of vanishing skewness for the distribution of the observable $B_3(\beta_c, am, N_f, N_\tau, N_s) = 0$. The order of the transition can be determined by the kurtosis $B_4(\beta, m)$[16]. In the thermodynamic limit $N_\sigma \to \infty$, the kurtosis $B_4(\beta_c, m, N_s)$ takes the values of 1 for a first order transition and 3 for an analytic crossover, respectively, with a discontinuity when passing from a first order region to a

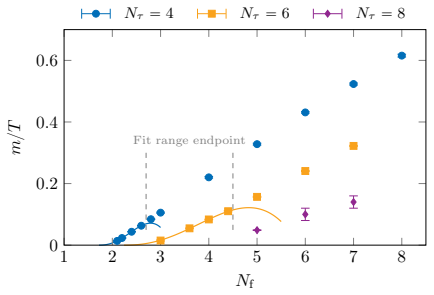


(a) Left: examples of B_3 for heavy quarks on a 6×36^3 lattice, obtained from different Markov chains, n_σ denotes the maximal difference between them in standard deviations. Merged raw and reweighted data for B_3 (top) and B_4 (bottom) are also shown, with the determination of β_c and $B_4(\beta_c)$ in red.



(b) The kurtosis on the phase boundary, $B_4(\beta_c)$, evaluated for different quark masses, on $N_\tau = 8$ lattices. Lines represent a common fit of all displayed volumes to the scaling formula Eq. 5.

Figure 3. Our general procedure of identifying phase transitions by finite size scaling analysis.



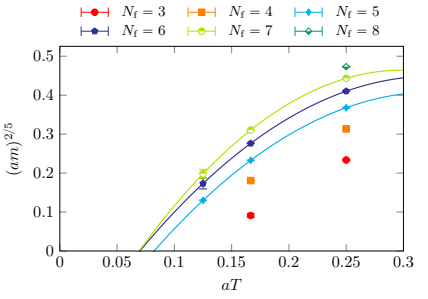


Figure 4. Left: The $Z(2)$-critical line separating first-order transitions (below) from crossover (above), for unimproved staggered fermions[12]. Right: The same data plotted in a different parameter pairing[12]. The continuum limit is in the lower left corner.

crossover region via a second order point, where it takes the value[17] 1.604 for the 3D Ising universality class. On finite but sufficiently large volumes, it can be expanded about the critical point,

$$B_4(\beta_c, am, N_f, N_\tau, N_s) = 1.604 + B\ (\beta_c, N_f, N_\tau)\ (am - am_c) N_s^{1/\nu} + \dots\ , \qquad (5)$$

through which it passes smoothly. As the volume is increased, the rate of the approach to the thermodynamic limit is governed by a 3D Ising critical exponent, $\nu = 0.6301$. Dots indicate additional terms that vanish in the infinite volume limit.

For each parameter combination, we generated statistics by simulating four independent Monte Carlo chains until their $B_{3,4}$-values agreed to within less than three standard deviations, upon which they were merged. The multi-histogram method was used to interpolate between simulated β-values[18], in order to locate the pseudo-critical coupling pre-

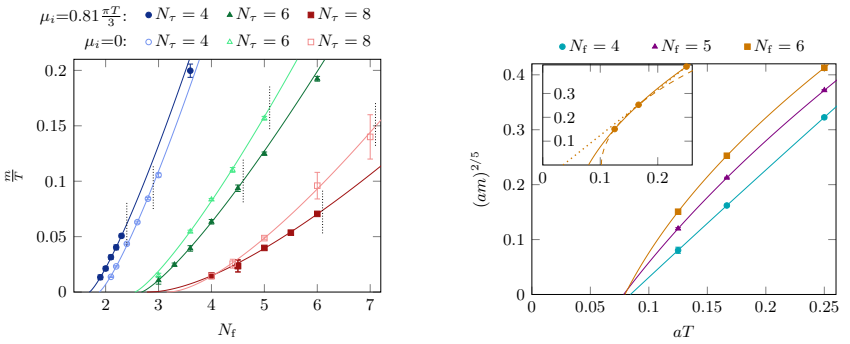


Figure 5. The same as in Fig. 4, but now for fixed imaginary quark chemical potential $\mu = i\mu_i$. On the left, the $\mu = 0$ curves are also displayed for comparison.

cisely. This is shown in Fig. 3(a). An example of a fit of our data to Eq. 5 is shown in Fig. 3(b). Altogether, the results presented in the following sections are based on ~ 200 million Monte Carlo trajectories spread over several hundred different parameter combinations, obtained over a span of several years.

Results of our investigation for $\mu_B = 0$ are shown in Fig. 4. The left plot is the numerical realisation of the chiral critical line, sketched schematically in Fig. 2 right, on lattices with $N_\tau = 4, 6, 8$ respectively. One observes the tricritical point, corresponding to the intercept of a fit to Eq. 3, to move to larger values as N_τ grows (i.e the lattice gets finer). While no continuum limit for the value of $N_f^{\rm tric}$ is available yet, it is obvious that $N_f^{\rm tric} > 3$, so that the chiral transition in the massless limit of the $N_f = 3$ theory is of second order. The same conclusion is reached by looking at the same critical line in a different parameter pairing, Fig. 4 right. The small curvature of the critical lines shows near-perfect tricritical scaling. All theories with $N_f \leq 7$ are consistent with a tricritical point at a finite $aT = N_\tau^{-1}$, which means that the first-order transition region under the critical line is not connected to the continuum limit, but a lattice artefact. Unless a new first-order region is found at even smaller masses, one concludes that for $N_f = 2 - 7$ the chiral phase transition in the continuum must be of second order. The Columbia plot in the continuum then differs from the proposal in Fig. 2 and instead looks as in Fig. 6 left.

4 The Columbia Plot with Imaginary Chemical Potential

The next step in our program is to determine how the nature of the chiral transition depends on chemical potential. Since a physical, real baryon chemical potential cannot be simulated, we chose an imaginary chemical potential, for which there is no sign problem. Earlier work on coarse lattices displays an analogous situation to $\mu_B = 0$, namely a first-order chiral transition region which terminates in a $Z(2)$-critical line. We have then considered a fixed quark chemical potential $\mu = i0.81\pi T/3$, for which there is an analogous Columbia plot with first-order regions, which on coarse $N_\tau = 4$ lattices are significantly wider than at zero density. However, upon increasing N_τ, i.e. making the lattices finer, this first-order region also shrinks and disappears at a tricritical point, as evinced by sim-

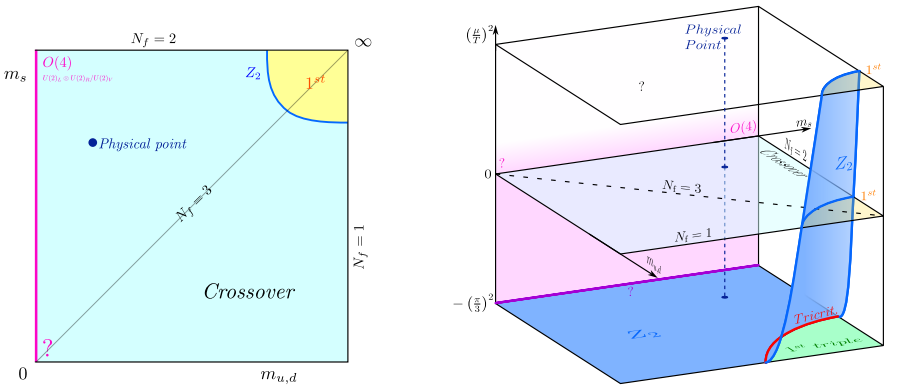


Figure 6. Left: The Columbia plot for $\mu_B = 0$ in the continuum according to our results[12]. Right: Extension to finite quark chemical potential. The lower half-space with imaginary chemical potential is accessible by Monte Carlo simulations, with the same result as at $\mu_B = 0$.

ilar scaling behaviour as at $\mu = 0$, which is shown in Fig. 5. Our results are consistent with other approaches with two different versions of improved staggered fermion actions at $\mu = i\pi T/3$: starting at the physical point and approaching the chiral limit, no sign of a first-order transition is found down to pion masses $\sim 50 - 60$ MeV[19,20]. Unless a completely different first-order transition is found at yet smaller quark masses in the future, one has to conclude that there is a second-order chiral transition in the limit of strictly massless quarks, and an analytic crossover for any non-vanishing quark mass. The Columbia plot including imaginary chemical potential then looks as in Fig. 6 right. Preliminary results have been published in conference proceedings[21] and a doctoral thesis[22], a journal article is in preparation.

5 Conclusions

Modern high performance computing allows to evaluate ever larger portions of the QCD parameter space, and to approach the theoretically interesting chiral limit of vanishing quark masses. With our ongoing long-term project reported here, we are presenting increasingly tight evidence that the QCD chiral phase transition in the limit of massless quarks at zero and imaginary baryon chemical potential is of second order for all $N_f = 2 - 7$. While this does not rule out a change to a first-order transition at some real chemical potential, our data are beginning to impose relevant bounds on the location of a possible critical endpoint. Fig. 1 middle and right show the remaining possibilities for the physical phase diagram and its connection to the chiral limit: while a first-order line closing to the T-axis is increasingly ruled out, a second-order line connecting the T- and μ-axes is still a possibility. In this case the transition for physical quark masses would correspond to crossover everywhere. In the more expected scenario with a first-order chiral phase transition at finite density, knowledge of T_c in the chiral limit and the curvature of the crossoverline at the physical point allows to bound the location of a possible critical endpoint in physical QCD to $\mu_B > 485$ MeV[11].

Acknowledgements

The authors acknowledge support by the Deutsche Forschungsgemeinschaft (DFG) through the grant CRC-TR 211 "Strong-interaction matter under extreme conditions". F.C. and O.P. in addition acknowledge support by the State of Hesse within the Research Cluster ELEMENTS (Project ID 500/10.006).

References

1. A. M. Halasz, A. D. Jackson, R. E. Shrock, M. A. Stephanov, and J. J. M. Verbaarschot, *On the phase diagram of QCD*, Phys. Rev. D, **58**, 096007, 1998.
2. M. A. Stephanov, K. Rajagopal, and E. V. Shuryak, *Signatures of the tricritical point in QCD*, Phys. Rev. Lett., **81**, 4816-4819, 1998.
3. Y. Hatta and T. Ikeda, *Universality, the QCD critical/tricritical point and the quark number susceptibility*, Phys. Rev. D, **67**, 014028, 2003.
4. Y. Aoki, G. Endrodi, Z. Fodor, S. D. Katz, and K. K. Szabo, *The Order of the quantum chromodynamics transition predicted by the standard model of particle physics*, Nature, **443**, 675-678, 2006.
5. R. D. Pisarski and F. Wilczek, *Remarks on the Chiral Phase Transition in Chromodynamics*, Phys. Rev., **D29**, 338-341, 1984.
6. F. R. Brown, F. P. Butler, H. Chen, N. H. Christ, Z. Dong, W. Schaffer, L. I. Unger, and A. Vaccarino, *On the existence of a phase transition for QCD with three light quarks*, Phys. Rev. Lett., **65**, 2491-2494, 1990.
7. Y. Iwasaki, K. Kanaya, S. Kaya, S. Sakai, and T. Yoshie, *Finite temperature transitions in lattice QCD with Wilson quarks: Chiral transitions and the influence of the strange quark*, Phys. Rev. D, **54**, 7010-7031, 1996.
8. F. Karsch, E. Laermann, and C. Schmidt, *The Chiral critical point in three-flavor QCD*, Phys. Lett., **B520**, 41-49, 2001.
9. P. de Forcrand and O. Philipsen, *The QCD phase diagram for three degenerate flavors and small baryon density*, Nucl. Phys., **B673**, 170-186, 2003.
10. C. Bonati, P. de Forcrand, M. D'Elia, O. Philipsen, and F. Sanfilippo, *Chiral phase transition in two-flavor QCD from an imaginary chemical potential*, Phys. Rev., **D90**, no. 7, 074030, 2014.
11. O. Philipsen, *Lattice Constraints on the QCD Chiral Phase Transition at Finite Temperature and Baryon Density*, Symmetry, **13**, no. 11, 2079, 2021.
12. F. Cuteri, O. Philipsen, and A. Sciarra, *On the order of the QCD chiral phase transition for different numbers of quark flavours*, JHEP, **11**, 141, 2021.
13. I. D. Lawrie and S. Sarbach, "Theory of tricritical points", in: Phase transitions and critical phenomena, C. Domb and J. L. Lebowitz (Eds.), **9**, 1, 1984.
14. C. Pinke, M. Bach, A. Sciarra, F. Cuteri, L. Zeidlewicz, C. Schäfer, T. Breitenfelder, C. Czaban, S. Lottini, and P. F. Depta, "`CL`^{2}`QCD - v1.0`", Sept. 2018.
15. A. Sciarra, C. Pinke, M. Bach, F. Cuteri, L. Zeidlewicz, C. Schäfer, T. Breitenfelder, C. Czaban, S. Lottini, and P. F. Depta, "`CL`^{2}`QCD - v1.1`", Feb. 2021.
16. K. Binder, *Finite size scaling analysis of Ising model block distribution functions*, Z. Phys., **B43**, 119-140, 1981.

17. A. Pelissetto and E. Vicari, *Critical phenomena and renormalization group theory*, Phys. Rept., **368**, 549-727, 2002.
18. A. M. Ferrenberg and R. H. Swendsen, *Optimized Monte Carlo analysis*, Phys. Rev. Lett., **63**, 1195-1198, 1989.
19. C. Bonati, E. Calore, M. D'Elia, M. Mesiti, F. Negro, F. Sanfilippo, S. F. Schifano, G. Silvi, and R. Tripiccione, *Roberge-Weiss endpoint and chiral symmetry restoration in $N_f = 2 + 1$ QCD*, Phys. Rev. D, **99**, no. 1, 014502, 2019.
20. F. Cuteri, J. Goswami, F. Karsch, A. Lahiri, M. Neumann, O. Philipsen, C. Schmidt, and A. Sciarra, *Toward the chiral phase transition in the Roberge-Weiss plane*, Phys. Rev. D, **106**, no. 1, 014510, 2022.
21. A. D'Ambrosio, O. Philipsen, and R. Kaiser, *The chiral phase transition at non-zero imaginary baryon chemical potential for different numbers of quark flavours*, PoS, **LATTICE2022**, 172, 2023.
22. A. D'Ambrosio, *The chiral phase transition in the bare parameter space of staggered lattice QCD*, PhD thesis, Frankfurt University, 2024.

Algorithms for Confined Gluons

Stefan Schaefer

John von Neumann Institute for Computing, DESY, Platanenallee 6, 15738 Zeuthen, Germany
E-mail: stefan.schaefer@desy.de

Gluons are the carriers of the strong force. Together with the quarks they form the protons and neutrons which make up the atomic nuclei. While gluons make up a significant fraction of those, a particular kind of particles are the so-called glueballs, which are thought to be made predominantly of gluons. Despite having been hypothesised half a century ago and significant effort to find them in experiments, glueballs have not been detected beyond doubt. Numerical computations using lattice quantum chromodynamics could shed light on the existence of these particles, but new algorithms and strategies are needed for this task. We discuss two promising strategies which could help achieve this goal: multi-level sampling and trivialising flows.

1 Introduction

The fundamental theory describing nuclear matter continues to pose a large array of challenges to theoretical physicists. Even fifty years after the formulation of quantum chromodynamics (QCD), it is still hard to calculate many quantities which physicists need to make progress in the field. QCD describes the observable particles, like the proton or the neutron, as being composed of quarks held together by gluons, the carriers of the strong force. While many features of strong interaction physics can be immediately understood from this theory, it is not easy at all to accurately predict elementary particle properties like their mass, their interactions and their structure.

The origin of these difficulties is in the nature of the system itself: the quarks and gluons are interacting strongly. This makes analytic methods, which are so successful in weakly interacting theories, less useful. In the last four decades, a powerful method to solve the theory numerically has been developed: lattice quantum chromodynamics. Today, lattice QCD computations are the prime source for quantities like the QCD running coupling or the quark masses.

Using lattice QCD, one can not only reproduce properties of known particles, but also study hypothetical particles. One particular class of particles, which have been predicted since the advent of QCD, are the so-called glueballs[1]. In a world without quarks, glueballs would be the bound state of gluons and the only strongly interacting "visible" particles, since single gluons cannot escape the confinement of the bound state. Glueballs as bound states are a direct manifestation of the special properties of quantum chromodynamics. Photons, the carriers of the electromagnetic force, do not exhibit a similar phenomenon. In numerical lattice computations, it has long been demonstrated that such glueballs exist in the theory without quarks[2,3] .

In QCD, quarks and gluons, however, always come together and cannot be separated. Therefore, these purely gluonic states will be different in the world we live in. They might still consist predominantly of gluons, with only a small contribution from the quarks, or their nature might change completely such that they have no resemblance with what has been established in the purely gluonic theory. For sure, they are no longer stable, to the contrary, they decay quickly with many possible decay channels.

In experiment, glueballs have been elusive. There have been experimental observations of candidates[4] and recent findings by BESIII confirming pseudoscalar quantum numbers for the X(2370) are reviving the interest in this field[5]. Reliable theoretical input from QCD would certainly help the interpretation of these findings.

The method requires significant computer resources and sustainable progress can only be made by exploring new numerical strategies and further developing proven algorithms. Simply using larger amounts of computer time is not an option. In the following, we will describe two such avenues to speed up lattice QCD computations.

2 Algorithms for Lattice Field Theory

As already indicated by the "lattice" in the name, space-time is discretised on a four-dimensional lattice. Once the theory is formulated on this lattice, what remains is integrating over the degrees of freedom attached to each of these lattice sites: the task to compute the expectation value of an observable $\langle O \rangle$ is to evaluate an integral with many millions to billions of integration variables

$$\langle O \rangle = \frac{1}{Z} \int [dU] e^{-S[U]} O[U] \,, \tag{1}$$

where $Z = \int [dU] e^{-S[U]}$ and U denoting the aggregate of the gluon fields.

Such high-dimensional integrals are typically tackled by Monte-Carlo integration. It means interpreting the Boltzmann weight $P[U] \propto e^{-S[U]}$ as a probability density and drawing all those millions of variables at once from this distribution.

Since it is virtually impossible to directly draw form this distribution, the field configurations are generated in the framework of a Markov Chain Monte Carlo: the field space is explored by deforming in a randomised way gauge field configurations such that one stays in the important region, where the importance is given by the probability $P[U]$. This gives a series of field configurations, where a field configuration U_i is a certain random value of each of the many integration variables

$$U_1 \to U_2 \to U_3 \to \cdots \to U_N \,. \tag{2}$$

From these variables, the quantities of interest, called $O[U]$ in the above formula, are then constructed. This is the measurement in lattice field theory language and the results for all the drawn realisations of the gauge fields are then averaged over

$$\bar{O} = \frac{1}{N} \sum_{i=1}^{N} O[U_i] \,. \tag{3}$$

The uncertainty of the estimate of $\bar{O}$ of the true value $\langle O \rangle$ decreases with $1/\sqrt{N}$, the inverse square root of the number N of samples drawn.

For certain types of observables, this procedure can lead to very satisfactory results. For many quantities of interest, this strategy, however, does not render competitive answers and we therefore have to work on new methods to go beyond the current state-of-the-art.

In this contribution, we discuss ideas to modify the above procedure in two ways. First we will challenge the idea that it is a good strategy to generate all field variables at once. Splitting up the update, one can devise improved estimators with a better scaling in the

number of measurements. The second avenue is a different way of generating the gauge fields. Instead of deforming probabilistically one global field into the next, we will present an effort to learn a map from a trivial distribution to the target distribution. If successful such an approach could avoid the problems arising from the correlation of subsequent field configurations.

2.1 Multilevel Algorithms

Many quantities of interest are extracted from correlations between operators O and O' at different points in space-time

$$\langle O(x)O'(y)\rangle\,, \tag{4}$$

e.g., the particle mass from the exponential decay rate with the distance. For such operator products, the statistical noise is frequently independent of the distance $|x-y|$ between the points, whereas the signal falls off exponentially with this distance. With growing distance, the computational cost to compute the signal to a certain level of accuracy grows exponentially with this separation. This problem has been known since the early days of lattice quantum field theory[6] and many strategies have been devised to fight it, but without finding a complete solution for general quantities. One successful strategy is the choice of good operators O and O' for the analysis. This will reduce the coefficient of the exponent and therefore mitigate the problem. A good choice of an operator basis is an essential ingredient in any modern computation and we will use a state-of-the-art setup in the following.

So-called multi-level algorithms are a possibility to improve the scaling of the Monte Carlo estimate of such products of operators. How is this done? The idea is to use the locality of the underlying theory, the property that local fields and their probabilities depend only on the other fields in their neighbourhood. We employ this in the concrete setup by decomposing the lattice into the two regions around the two points of the correlation function and a boundary where they meet, see Fig. 1. We can then independently use the Monte Carlo method to compute estimates for these two factors. For this to work, the fields in the boundary need to be kept fixed and we need to average over a certain number of

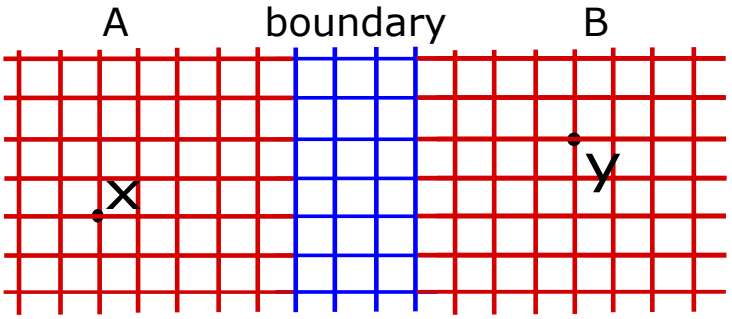


Figure 1. Simple decomposition of the lattice for a multilevel algorithm. We use a number N_0 of realisations for the boundary field in the centre. For each of these boundary fields, the lattice in region A and region B can now be sampled independently from the corresponding marginal distribution.

boundary field realisations. The strategy now translates to the following nested averaging scheme:

$$\langle O(x)O'(y)\rangle \to \langle\langle O(x)\rangle_{\rm A}\langle O'(y)\rangle_{\rm B}\rangle_{\rm bndry}\ . \tag{5}$$

where the averages $\langle O(x)\rangle_{\rm A}$ and $\langle O'(y)\rangle_{\rm B}$ depend on the boundary field, but not on the fields in the other region B and A, respectively.

The important result is that the uncertainty of the product potentially scales with the inverse square root of the product of the number of these two sub-measurements. If we make N_1 sub-measurements in each of the regions, the uncertainty drops with $1/N_1$ for the multilevel estimator, instead of $1/\sqrt{N_1}$ for the standard strategy. Note, however, that this is the ideal scaling which will be limited by the effect of the fixed boundary. Such ideas are not new and have been pursued over the years[7–9]. In a theory without fermions this has also been successful for certain quantities. Also for fermions, a possible strategy has been formulated[10,11]. That such an algorithm is efficient is not *a priori* clear. In particular, it might happen that the effect of the fixed region is so large that the signal has degraded beyond repair before the benefits of the new algorithm kicks in.

Since we have set out to study glueballs on the lattice, we tested this algorithm in an actual glueball computation, albeit without the presence of fermions[12]. This reduces the cost of the simulation significantly and we can therefore evaluate the idea in detail, also trying many variations. In Fig. 2, the noise-to-signal ratio of the correlation functions in different glueball channels is shown as a function of the distance between the operators. By its nature, the algorithm will be efficient once the two points are sufficiently far away from the respective side of the boundary. It is therefore no surprise that at shorter distances we observe an exponential degradation of the signal. Here, we can only use the standard estimator.

Depending on the channel, this changes at around $0.75r_0$ to r_0, with $r_0 \approx 0.5$ fm, where the degradation can be almost stopped. The plot also shows that the behaviour is physical in the sense that it does not depend on the lattice spacing, where the different values of $\beta = 5.8$, 6.08, and 6.2 correspond to a lattice spacing of 0.136 fm, 0.08 fm, and 0.068 fm, respectively. The use of the multilevel algorithm gives an essential window of opportunity for the measurements of the glueball's masses at the larger distances. While this is not necessary for the purely gluonic theory we used here, it will be essential for the next step where we want to compute in the full theory with the effect of the quark fields included.

2.2 Machine Learning

As already mentioned above, lattice gluon fields are typically drawn from the target distribution by using a Hamiltonian Monte Carlo[13], i.e. a method which continuously deforms the field while staying inside of the region of likely fields as defined by the probability distribution. This algorithm is widely used, not only in lattice quantum chromodynamics. It has the great advantage of being universally applicable as long as we have continuous variables.

The disadvantage is that it suffers from so-called critical slowing down as the lattice is made finer. There are significant correlations between subsequent configurations produced in this update procedure. They have their source in the general strategy for the algorithm:

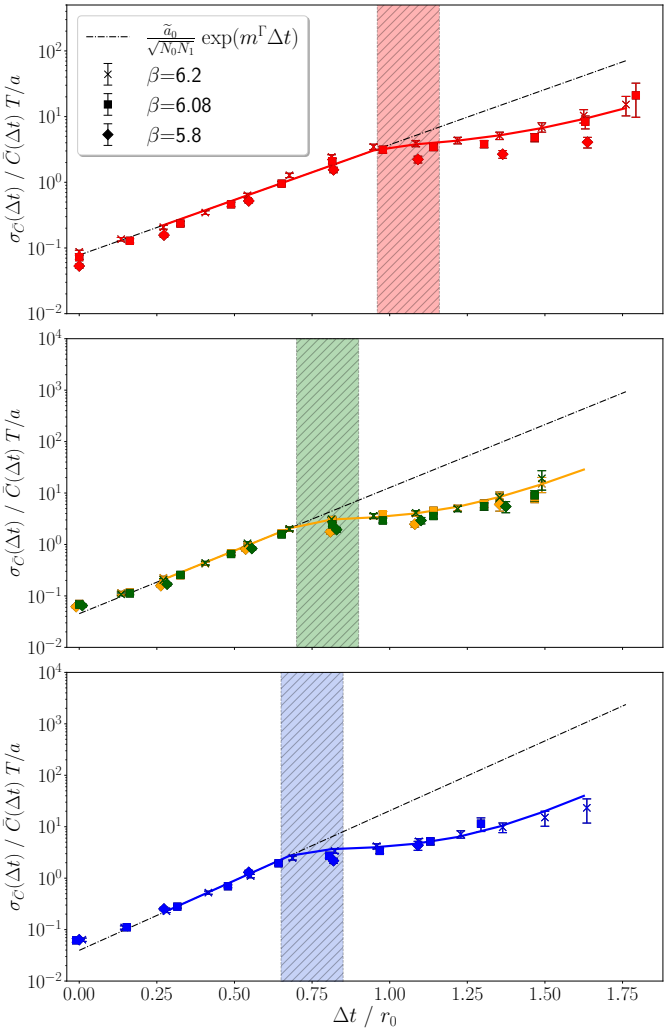


Figure 2. The noise-to-signal ratio as a function of the distance between the two operators for various types of glueball, A_1^{++} (red), E^{++} (green), T_2^{++} (orange), and A_1^{-+} (blue)[12]. At short distances, we observe the exponential deterioration of the signal compared to the uncertainty. From a certain distance on, the multilevel algorithm becomes efficient and this deterioration is slowed down. Plot from Barca et. al.[12]. The distance is given in units of $r_0 \approx 0.5$ fm.

it continuously deforms the fields, so there always is some "memory" left of the previous field configurations. These correlations now get worse on finer lattices. Detecting the correlation is to a certain extent also an art of its own. We have billions of variables and only certain will exhibit the worst of these correlations.

Particularly bad are observables linked to the topology of the underlying system. The topological sectors form quickly as the lattices are made finer and the simulation gets stuck in one of them. We therefore no longer draw from the full probability distribution

but only from a subset given by a certain topology of the fields. In a sense, this is even a good situation in which we have identified this particularly problematic observable and can therefore monitor the problem and know where the limits of the current setup are. There could very well be other quantities exhibiting much slower behaviour which we are unaware of.

These considerations trigger the wish to change the simulation strategy altogether. One such approach are so-called normalising flows. These are maps in fields space $U \to \mathcal{F}(U)$ such that in the integral Eq. 1 the Boltzmann weight $\exp(-S[U])$ gets replaced by a probability distribution from which we can sample trivially, without having to use methods like Hamiltonian Molecular Dynamics. For many applications this is the Gaussian distribution, hence the name. In our case it is a uniform distribution and we refer to them as trivialising flows.

If such a map can be found, the simulation setup changes drastically. Instead of producing a chain of correlated fields, we can draw uncorrelated fields from the uniform distribution to which we then apply the map. These maps are typically given in terms of a partial differential equation with parameters depending on the target distribution. These parameters have to be determined, learned in the language of machine learning, by minimising the distance of the distribution generated by the map and the target distribution given by the full theory $P[U] \sim \exp(-S[U])$.

This optimisation will never be perfect, and also the model will only have a limited number of parameters and terms. The question to answer is whether it is good enough to be used in actual applications.

This approach is very successful in many machine learning applications, however, these are frequently of very different type compared to the case of lattice quantum chromodynamics. On the lattice, we face a huge number of degrees of freedom on the one side. On the other side, we have a lot of symmetries in our system: translations for arbitrary shifts in the lattice, rotational symmetries and many more, also internal symmetries.

In recent years, there has been significant interest in this approach to the sampling problem in QCD[14]. With respect to earlier maps, we have proposed and implemented one with many orders of magnitude fewer parameters, implementing a large amount of symmetry[15]. It is based on an analytic approach proposed earlier by Lüscher[16], who also showed that in the given class of models such a map actually does exist, albeit with growing number of parameters. The low number of parameters of our model is a significant advantage if it comes to training. In this step, more parameters increase the cost and it will be more difficult to find optimal parameters with a given amount of resources. In case of an unsatisfactory match, it is also difficult to determine whether the problem is a lack of training or a lack of expressivity of the model.

In Fig. 3 we show the history of the training of the model in a two dimensional theory. The effective sampling size (ESS) is shown as it increases due to the training. The ESS gives an effective number of field configurations after taking into account the mismatch between the distribution produced by the model and the target. In our study, we could show that our reduced, physics driven approach significantly outperforms the previous efforts[15]. Using these methods is still in its infancy. In particular, there is so far one big hurdle: they scale very badly with the volume of the system. Some of this scaling might be overcome with a better model and better training, but this will only lead so far in the face of the billions of degrees of freedom of modern day lattice simulations.

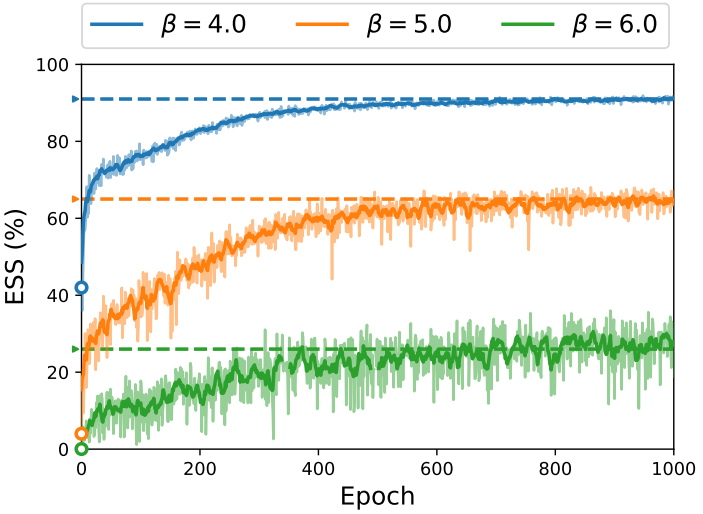


Figure 3. Training history of the effective sampling size (ESS), the effective fraction of field configurations generated after taking into account the mismatch between target distribution and the distribution generated by the flow[15].

3 Summary

Lattice quantum chromodynamics has come a long way since the first simulations at the beginning of the 1980s. Part of it is due to the increased availability of computer resources. A roughly even share is in the ability of the community to develop computational strategies to exploit the changing computer architectures and to invent new algorithms to address specific issues of the physical systems. In particular, it is time to develop algorithms with the specifics of the target observables in mind.

Here we discussed two such approaches taking specific advantage of key properties of the underlying theory. In the case of the multilevel sampling, we used the locality of the underlying theory to formulate an algorithm which could improve the scaling of the uncertainty of the result with the effort. For the trivialising flows, we used a model which implemented many of the symmetries of the theory, like the invariance under translations and rotations as well as internal symmetries directly. We demonstrated that this model is outperforming previously discussed ones by many orders of magnitude. This approach is, however, still far from being competitive with the Hamiltonian Monte Carlo.

These studies are not yet at the end, but an important step towards reach the goal of studying glueballs in the full theory of quarks and gluons.

Acknowledgements

We gratefully acknowledge the scientific support and HPC resources provided by the Forschungszentrum Jülich and Erlangen National High Performance Computing Center (NHR@FAU) of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU). This

work has in part been funded by the German Research Foundation (DFG) research unit FOR5269 “Future methods for studying confined gluons in QCD.”

References

1. H. Fritzsch and M. Gell-Mann, *Current algebra: Quarks and what else?*, eConf, **C720906V2**, 135-165, 1972.
2. G. S. Bali, K. Schilling, A. Hulsebos, A. C. Irving, C. Michael, and P. W. Stephenson, *A Comprehensive lattice study of SU(3) glueballs*, Phys. Lett. B, **309**, 378-384, 1993.
3. C. J. Morningstar and M. J. Peardon, *The Glueball spectrum from an anisotropic lattice study*, Phys. Rev. D, **60**, 034509, 1999.
4. V. Crede and C. A. Meyer, *The Experimental Status of Glueballs*, Prog. Part. Nucl. Phys., **63**, 74-116, 2009.
5. M. Ablikim et al., *Determination of Spin-Parity Quantum Numbers of $X(2370)$ as 0^{-+} from $J/\psi \to \gamma K_S^0 K_S^0 \eta'$*, Phys. Rev. Lett., **132**, 181901, May 2024.
6. G. Parisi, *The Strategy for Computing the Hadronic Mass Spectrum*, Phys. Rept., **103**, 203-211, 1984.
7. G. Parisi, R. Petronzio, and F. Rapuano, *A measurement of the string tension near the continuum limit*, Physics Letters B, **128(6)**, 418-420, 1983.
8. H. B. Meyer, *Locality and statistical error reduction on correlation functions*, JHEP, **01**, 048, 2003.
9. M. Lüscher and P. Weisz, *Locality and exponential error reduction in numerical lattice gauge theory*, JHEP, **09**, 010, 2001.
10. M. Cè, L. Giusti, and S. Schaefer, *Domain decomposition, multi-level integration and exponential noise reduction in lattice QCD*, Phys. Rev. D, **93(9)**, 094507, 2016.
11. M. Cè, L. Giusti, and S. Schaefer, *A local factorization of the fermion determinant in lattice QCD*, Phys. Rev. D, **95(3)**, 034503, 2017.
12. L. Barca, S. Schaefer, F. Knechtli, J. A. Urrea-Niño, S. Martins, and M. Peardon, *Exponential error reduction for glueball calculations using a two-level algorithm in pure gauge theory*, Phys. Rev. D, **110(5)**, 054515, 2024.
13. S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, *Hybrid Monte Carlo*, Phys. Lett. B, **195**, 216-222, 1987.
14. D. Boyda, G. Kanwar, S. Racanière, D. J. Rezende, M. S. Albergo, K. Cranmer, D. C. Hackett, and P. E. Shanahan, *Sampling using $SU(N)$ gauge equivariant flows*, Phys. Rev. D, **103(7)**, 074504, 2021.
15. S. Bacchio, P. Kessel, S. Schaefer, and L. Vaitl, *Learning trivializing gradient flows for lattice gauge theories*, Phys. Rev. D, **107(5)**, L051504, 2023.
16. M. Lüscher, *Trivializing maps, the Wilson flow and the HMC algorithm*, Commun. Math. Phys., **293**, 899-919, 2010.

Protein Structure Prediction in the Last 15 Years: From Inferring Sequence Coevolution to AlphaFold

Utkarsh Upadhyay[1], **Christian Faber**[1], **Oskar Taubert**[2], **and Alexander Schug**[1,3]

[1] John von Neumann Institute for Computing, Jülich Supercomputing Centre, Forschungszentrum Jülich, 52428 Jülich, Germany
E-mail: al.schug@fz-juelich.de

[2] Scientific Computing Center, Karlsruhe Institute of Technology, 76344 Eggenstein-Leopoldshafen, Germany

[3] Faculty of Biology, University of Duisburg-Essen, 45117 Essen, Germany

Biomolecules like Proteins and RNAs play a critical role in life processes at the molecular level, with their structures intricately linked to their function. However, experimentally determining their structures remains challenging. *In silico* techniques, such as computational structure prediction, offer a valuable complement to experimental approaches. Fifteen years ago, direct coupling analysis (DCA) was developed to infer co-evolutionary patterns and predict spatial relationships between residue pairs. Such spatial information could considerably improve 3D structure prediction. This paved the way for deep learning methods, which initially improved accuracy in predicting these relationships and eventually succeeded in generating full 3D structures. One of the most prominent achievements is AlphaFold, which has revolutionised the field. Its groundbreaking success earned Demis Hassabis and John Jumper one half of the 2024 Nobel Prize in Chemistry, highlighting its transformative impact on structural biology. This short review guides through the past 15 years of protein structure prediction for an interested public.

1 Introduction

Proteins and RNAs are fundamental building blocks of life at the molecular level and carry out key functions that sustain biological processes. Build from linear sequences of amino acids, proteins act as, e.g., enzymes, catalysts that speed up biochemical reactions, and serve as structural components of cells. Hemoglobin, for example, transports oxygen in the blood, while insulin regulates blood sugar levels. RNA is likewise a versatile molecule, being involved in protein synthesis and gene regulation, as seen with messenger RNA (mRNA) and ribosomal RNA (rRNA) but has also been shown to be able to react to external stimuli as riboswitches regulating genetic expression. Common to these biomolecules is the strong dependence on their three-dimensional structures, which dictate their interactions and functions with dysfunctions often related to disease[1,2].

However, determining these structures experimentally is a major scientific challenge. Techniques like X-ray crystallography, NMR and cryo-electron microscopy (cryo-EM) are powerful, yet time-consuming and technically demanding. For many proteins and RNA molecules, especially those that are large, flexible, or difficult to crystallise, obtaining high-resolution structural data remains (at a minimum) time consuming and expensive or even elusive.

Computational approaches, such as structure prediction techniques, have become valuable tools to complement experimental efforts. By using algorithms to predict 3D structures based on sequence data, these methods can provide insights where experiments may

be limited. The remarkable advancements in sequencing technologies have led to an exponential increase in the availability of biological sequence data[3]. These advances have opened up new opportunities for understanding the molecular basis of life, particularly in relation to proteins and ribonucleic acids (RNA). As sequencing techniques become more efficient and affordable, the sheer volume of data generated allows scientists to explore the diversity of proteins and RNA across different species, gaining insights into their evolution, structure, and function. In this mini-review, we will quickly summarise the basic of organising sequence information in multiple sequence alignments and inferring spatial or structural information from them by statistical physics and machine learning approaches.

2 Availability of Raw Data: Organising Sequence Data as Multiple Sequence Alignments

One of the key tools for analysing this vast amount of sequence data are multiple sequence alignments (MSA), which enable researchers to identify similarities and differences between sequences from different organisms. In short, MSAs are essential for inferring structural and functional relationships within and between phylogenetic trees. By aligning sequences, scientists can identify conserved regions that are crucial for the function of a protein or RNA, or regions that have (co-)evolved to confer new functions or maintain structural properties. Improved statistical methods and sophisticated alignment software have been instrumental in enhancing the accuracy of these analyses[4–10].

Freely accessible databases, such as UniProt, Pfam (the protein family database), and Rfam (the RNA family database), have emerged as invaluable resources for researchers seeking to analyse protein and RNA sequences. These databases curate and organise sequence data, providing researchers with comprehensive information on the structure, function, and evolutionary relationships of proteins and RNA. UniProt, for example, is a widely used protein database that contains detailed annotations on protein sequences, including information about protein function, domains, and interactions[11]. Similarly, Pfam offers curated data on protein families, allowing researchers to investigate conserved regions that are common to proteins with similar functions[12]. Rfam, on the other hand, focuses on RNA families, providing insights into the structure and function of non-coding RNAs, which play critical roles in gene regulation and other cellular processes[13].

The availability of these vast and well-organised data sets is essential for the broader scientific community and they serve as foundational resources for a wide range of applications by enabling development of sophisticated bioinformatics tools to expand our ability to study the molecular underpinnings of life. Importantly, these databases are freely accessible to researchers worldwide, fostering collaboration and accelerating discoveries.

3 2009 Structure Prediction by Tracing Co-Evolution: Direct Coupling Analysis (DCA)

Fifteen years ago, direct coupling analysis (DCA) and related methods revolutionised our ability to predict the spatial proximity of amino acid residues in proteins by detecting co-evolutionary patterns in sequence data. These techniques identify “contacts” or residues

that are likely to be close in three-dimensional space, by analysing patterns of linked mutations across evolutionary time[14,15]. Previous methods, such as those based on mutual information, struggled to distinguish between direct and indirect interactions, often misidentifying residues as interacting when they were not directly adjacent. DCA addressed this limitation by specifically inferring direct couplings between residues, which arise from their direct spatial proximity within the protein structure.

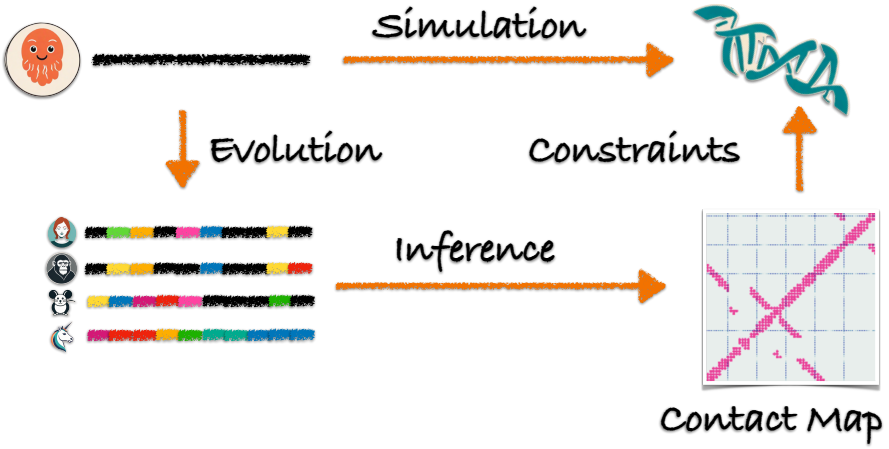


Figure 1. **General structure prediction workflow** The wild type sequence of biomolecules encodes a specific biomolecular structure. While maintaining this overall shape, evolution generates correlated mutational patterns in multiple sequence alignments, often due to spatial proximity of residues. These patterns can be used to statistically infer contact maps, which provide valuable constraints for predicting biomolecular structures. By incorporating these contact maps into prediction models, the accuracy of the predictions is significantly enhanced compared to models that rely solely on direct structure prediction without such constraints. This approach leverages evolutionary information to better capture the physical interactions between residues, ultimately improving the overall quality of the predicted structures.

DCA achieves this by applying an inverse Potts model to sequence data, allowing researchers to disentangle the complex web of evolutionary interactions. With $P(S)$ being the probability that a given sequence $S = a_1 a_2 ... a_L$ of length L, in which each state a_i is either a residue or a gap, is sampled over the course of evolution. $P(S)$ can be written using the Boltzmann law as:

$$P(S) = \frac{1}{Z} \exp(-\beta\, \phi), \tag{1}$$

where β is the inverse of the temperature, Z the partition function of the model and ϕ the Hamiltonian of the system, which in turn is expressed via a generalised Potts model as:

$$\phi = -\sum_{i<j}^{L} J_{ij}(a_i, a_j) - \sum_{i=1}^{L} h_i(a_i). \quad (2)$$

The parameters $h_i(a_i)$ measure the local field strength at site i occupied by state a_i and the coupling parameters $J_{ij}(a_i, a_j)$ quantify the coupling strength between pairs of sites i and j occupied by states a_i and a_j, respectively. The local field and the coupling parameters of the model are inferred from the MSA of homologous sequences to S using inverse statistical algorithms. Initially, DCA employed a message-passing algorithm to solve this inverse problem[15]. Subsequent advancements aimed at increasing efficiency and accuracy, including the development of the mean-field approach, which significantly sped up the inference process[16,17], using pseudolikelihood maximisation further improving prediction accuracy[18] or Boltzmann learning[19]. Furthermore, the principles behind DCA have been successfully extended to RNA molecules, enabling the prediction of structural contacts in RNA sequences as well[20].

In a similar idea to using experimental data as restraints in structural modelling[21–23], the contact information derived from DCA can serve as powerful structural constraints. This approach has been applied to a range of biomolecular modelling challenges[24,25], such as predicting the structure of protein complexes[14], globular proteins[26], mapping conformational transitions[27], serving as constraints in MD simulations[28] and even studying large-scale homodimer prediction[29]. DCA has also proven useful in applications like protein design, where it has been used to re-engineer protein signalling pathways[30–32] or predict fitness landscapes[33,34].

In addition to its success with proteins, DCA has been applied to RNA contact prediction, yielding great success in RNA structure prediction[20,35–37]. More recently, DCA has been combined with machine learning techniques to further improve its predictive power. Shallow learning approaches, which require fewer parameters, have been shown to enhance DCA's performance[38]. These advancements continue to expand the potential applications of DCA in biomolecular research, making it a valuable tool for understanding and modelling the complex structures of both proteins and RNA.

4 2017+ Shift to Deep Learning (DL) in Biomolecular Structure Prediction

In recent years, numerous scientific disciplines have been profoundly impacted by advancements in machine learning, driven by both theoretical innovations and vast technological improvements [a]. These breakthroughs have been supported by the development of specialised hardware such as novel graphical processing units (GPUs) optimised for machine learning architectures. In addition the availability of large, high-quality datasets (ideally fully annotated) supports deep-learning approaches. One of the fields most significantly transformed by these advancements is biological physics and in particular the focus of this mini-review: structure prediction. Here, the intersection of vast biological

[a] Interestingly, these approaches are conceptually linked to earlier developments in statistical physics, such as the Hopfield Network (e.g. the review[39]).

data, both on the sequential and structural level, combined with powerful machine learning models led astonishing progress in 3D structure prediction[40–42] including one of best known successes in deep learning AlphaFold[41]. These ML approaches, rely on deep neural networks with an accordingly massive number of free parameters. Naturally, training these deep networks require equally very large datasets of structurally known 3D structures and accompanying MSAs.

Early neural networks for protein contact prediction were based on convolutional neural networks (CNNs), which treated the problem similarly to image processing. These CNN-based models used output maps from direct coupling analysis (DCA) as input and refined them to predict contact likelihoods for each possible pair of residues, producing entire contact maps[43,38]. To better incorporate sequence-specific features, models introduced concatenated or summed embeddings of sequence tokens and sequence profiles, which contain alignment statistics for each sequence position[43]. Over time, these models such as AlphaFold advanced to predict distance maps, or distograms, instead of simple binary contacts, providing more granular spatial information about residue pairs[41].

The development of transformers and large language models marked a shift towards language processing approaches. The MSA Transformer[44] is one such protein language model, capable of extracting co-evolutionary patterns from multiple sequence alignments (MSAs) through self-supervised learning. It uses attention maps to predict residue contacts. More recent models, such as single-sequence transformers, embed evolutionary context directly into their model parameters, bypassing the need for MSAs in the input[45]. A third approach involves geometric or graph-based models, which analyse sampled or generated structures to improve structural predictions, although they do not directly generate structural candidates themselves[46].

In an ideal scenario, a model would directly predict the atomic coordinates of a protein as a point cloud, eliminating the need for complex post-processing and additional computational modelling. One of the first models to approach this was the recurrent geometric network (RGN)[47], which relied on long short-term memory (LSTM) networks[48]. However, more recent models like AlphaFold2[49] and RosettaFold[50] use attention-based architectures.

These models employ different sub-modules, including a token-level attention network and a geometric structure module. The latter incorporates inductive biases such as geometric transformations (e.g., SE(3) symmetry, used in the SE(3)-Transformer[51]), which helps the model better predict protein structures. Additionally, the token-level sub-module is trained using a self-supervised masked language task as an auxiliary loss, improving its performance. Both AlphaFold2 and RosettaFold are trained on large protein datasets, requiring hundreds of thousands of samples to achieve their high level of accuracy. Recently, AlphaFold3[52] employed a diffusion-based approach which further increased accuracy and range of predictable systems (proteins, nucleic acids, small molecules, ions and modified residues).

For RNA, there is a discrepancy in the availability of data. While there is massive sequence data (>30 Mio sequences), there exist only 8000 RNA structures in the PDB, many of them from related RNA families. Our recent method BARNACLE[53] first learns an optimal internal representation on the sequence data (upstream training) before fine-tuning to specific prediction tasks (downstream training) such as contact prediction, which relies on the less abundant structural data. Similarly, the recent development of RNAformer[54] is a transformer model designed for predicting RNA secondary structures and highlights

the growing application of machine learning techniques to RNA alongside proteins. AlphaFold3[52] also expanded its capabilities towards RNA structure prediction (cf. Fig. 2). Here, we observe a clear trend: structures with known structural similarity to the training data tend to be predicted more accurately, i.e. low in RMSD. However, the accuracy of predicted structures for sequences dissimilar to the training dataset still present challenges. The scoring of the models, i.e. the estimate of the predicted structure's quality, are quite reliable for high scoring models. Low scores indicate varying prediction quality, with some predictions being of high quality while others are structurally very dissimilar from the target structure.

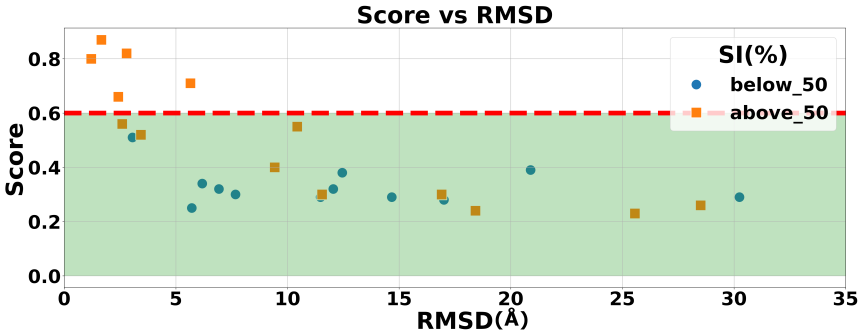


Figure 2. **AlphaFold3 RNA structure prediction** The figure shows blind RNA structure predictions made by AlphaFold3 (AF3) for RNAs that were not part of its training dataset but were experimentally resolved after its release. The score represents AF3's own confidence in the prediction, where values below 0.6 indicate low reliability. Sequences without similarity to known structures tend to be predicted less accurately, as reflected by their high root mean square deviation (RMSD) from experimentally measured structures. RMSD values below 5Å denote high-quality predictions, while higher values suggest poorer accuracy in the predicted structures.

5 Discussion and Summary

This article provided a comprehensive overview of advancements in protein and RNA structure prediction, focusing on the evolution from statistical methods to, more recently, deep learning approaches. It began with DCA, a statistical method that leverages co-evolutionary patterns in multiple sequence alignments to infer spatial contacts between residues. DCA marked a significant step forward, providing structural insights by revealing evolutionary constraints that indicate residue proximity. The field has transitioned to deep learning models, which have vastly improved predictive accuracy by directly learning complex patterns within large datasets with no direct assumption about the underlying evolutionary patterns[b]. In particular, the development of convolutional neural networks (CNNs) introduced the ability to treat contact prediction as an image-like problem, enhancing accuracy through feature extraction.

The article then discussed AlphaFold2, a deep learning model that employs attention-based architectures to predict full 3D structures with unprecedented accuracy. By integrating information from evolutionary data and protein sequence profiles, AlphaFold2 set a

[b]DCA only considers local field and two-body terms.

new standard in structure prediction, achieving results comparable to experimental techniques and was awarded part of the Nobel Prize 2024 in Chemistry, highlighting the successes in and impact of this field and the promises it holds for the future.

The covered advances have not only improved protein and RNA structure prediction but also open new possibilities in molecular biology and applied fields like pharmacology and biotechnology. One particularly promising direction is the reverse task of structure prediction: biomolecular design. In this area, deep learning approaches, especially those leveraging foundation models, could enable the design of biomolecules beyond the sequence space explored by evolution. This could lead to the creation of entirely new molecular structures with tailored functions, such as novel folds capable of catalysing specific reactions. Such innovations hold great potential for developing new therapies, materials, and biotechnological tools.

6 Acknowledgements

The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (`www.gauss-centre.eu`) for funding this project by providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS Supercomputer JUWELS at Jülich Supercomputing Centre (JSC).

We acknowledge the use of hotpot.ai for generating the logs in Figure 1 and GPT-4-turbo (Open AI, `https://chat.openai.com/`) for assistance in refining and clarifying text and to proofread the final draft.

References

1. J. Nasica-Labouze, P. H. Nguyen, F. Sterpone, O. Berthoumieu, N.-V. Buchete, S. Cote, A. De Simone, A. J. Doig, P. Faller, A. Garcia et al., *Amyloid β protein and Alzheimer's disease: When computer simulations complement experimental studies*, Chemical reviews, **115**, no. 9, 3518-3563, 2015.
2. A. Hautke, A. Voronin, F. Idiris, A. Riel, F. Lindner, A. Lelievre-Buttner, J. Zhu, B. Appel, E. Fatti, K. Weis et al., *CAG-repeat RNA hairpin folding and recruitment to nuclear speckles with a pivotal role of ATP as a cosolute*, Journal of the American Chemical Society, **145**, no. 17, 9571-9583, 2023.
3. B. Hwang, J-H. Lee, and D. Bang, *Single-cell RNA sequencing technologies and bioinformatics pipelines*, Experimental & Molecular Medicine, **50**, 1-14, 2018.
4. S. R. Eddy, *Hidden Markov models*, Curr. Opin. Struct. Biol., **6**, 361-365, 1996.
5. E. P. Nawrocki and S. R. Eddy, *Infernal 1.1: 100-fold faster RNA homology searches*, Bioinformatics, **29**, 2933-2935, 2013.
6. S. R. Eddy and R. Durbin, *RNA sequence analysis using covariance models*, Nucleic acids research, **22**, no. 11, 2079-2088, 1994.
7. J. D. Thompson, T. J. Gibson, and D. G. Higgins, *Multiple sequence alignment using ClustalW and ClustalX*, Curr. Protoc. Bioinf., **00**, 2.3.1-2.3.22, 2002.
8. H. McWilliam, W. Li, M. Uludag, S. Squizzato, Y. M. Park, N. Buso, A. P. Cowley, and R. Lopez, *Analysis Tool Web Services from the EMBL-EBI*, Nucl. Acids Res., **41**, W597-W600, 2013.

9. D. J. Lipman and W. R. Pearson, *Rapid and sensitive protein similarity searches*, Science, **227**, 1435-1441, 1985.
10. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, *Basic local alignment search tool*, J. Mol. Biol., **215**, 403-410, 1990.
11. The UniProt Consortium, *UniProt: a worldwide hub of protein knowledge*, Nucl. Acids Res., **47**, D506-D515, 2019.
12. R. D. Finn, A. Bateman, J. Clements, P. Coggill, R. Y. Everhardt, S. R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, E. L. L. Sonnhammer, J. Tate, and M. Punta, *Pfam: the protein families database*, Nucl. Acids Res., **42**, D222-D230, 2014.
13. S. Griffiths-Jones, A. Bateman, M. Marshall, A. Khanna, and S. R. Eddy, *Rfam: an RNA family database*, Nucl. Acids Res., **31**, 439-441, 2003.
14. A. Schug, M. Weigt, J. N. Onuchic, T. Hwa, and H. Szurmant, *High-resolution protein complexes from integrating genomic information with molecular simulation*, Proceedings of the National Academy of Sciences, **106**, no. 52, 22124-22129, 2009.
15. M. Weigt, R. A. White, H. Szurmant, J. A. Hoch, and T. Hwa, *Identification of direct residue contacts in protein–protein interaction by message passing*, Proc. Natl. Acad. Sci. U.S.A., **106**, 67-72, 2009.
16. F. Morcos, A. Pagnani, B. Lunt, A. Bertolino, D. S. Marks, C. Sander, R. Zecchina, J. N. Onuchic, T. Hwa, and M. Weigt, *Direct-coupling analysis of residue coevolution captures native contacts across many protein families*, Proc. Natl. Acad. Sci. U.S.A., **108**, E1293-E1301, 2011.
17. D. S. Marks, T. A. Hopf, and C. Sander, *Protein structure prediction from sequence variation*, Nature Biotechnol., **30**, 1072-1080, 2012.
18. M. Ekeberg, C. Lövkvist, Y. Lan, M. Weigt, and E. Aurell, *Improved contact prediction in proteins: Using pseudolikelihoods to infer Potts models*, Phys. Rev. E, **87**, 012707, Jan. 2013.
19. F. Cuturello, G. Tiana, and G. Bussi, *Assessing the accuracy of direct-coupling analysis for RNA contact prediction*, RNA, **26**, 637-647, 2020.
20. E. De Leonardis, B. Lutz, S. Ratz, S. Cocco, R. Monasson, A. Schug, and M. Weigt, *Direct-Coupling Analysis of nucleotide coevolution facilitates RNA secondary and tertiary structure prediction*, Nucleic acids research, **43**, no. 21, 10444-10455, 2015.
21. M. Weiel, I. Reinartz, and A. Schug, *Rapid interpretation of small-angle X-ray scattering data*, PLoS Computational Biology, **15**, no. 3, e1006900, 2019.
22. I. Reinartz, C. Sinner, D. Nettels, B. Stucki-Buchli, F. Stockmar, P. T. Panek, C. R. Jacob, G. U. Nienhaus, B. Schuler, and A. Schug, *Simulation of FRET dyes allows quantitative comparison against experimental data*, The Journal of Chemical Physics, **148**, no. 12, 123321, 2018.
23. M. Weiel, M. Götz, A. Klein, D. Coquelin, R. Floca, and A. Schug, *Dynamic particle swarm optimization of biomolecular simulation parameters with flexible objective functions*, Nature Machine Intelligence, **3**, no. 8, 727-734, 2021.
24. S. Cocco, C. Feinauer, M. Figliuzzi, R. Monasson, and M. Weigt, *Inverse statistical physics of protein sequences: a key issues review*, Rep. Prog. Phys., **81**, 032601, 2018.
25. T. A. Hopf, A. G. Green, B. Schubert, S. Mersmann, C. P. I. Schärfe, J. B. Ingraham, A. Toth-Petroczy, K. Brock, A. J. Riesselman, P. Palmedo, C. Kang, R. Sheridan,

E. J. Draizen, C. Dallago, C. Sander, and D. S. Marks, *The EVcouplings Python framework for coevolutionary sequence analysis*, Bioinformatics, **35**, no. 9, 1582-1584, 10 2018.
26. J. I. Sułkowska, F. Morcos, M. Weigt, T. Hwa, and J. N. Onuchic, *Genomics-aided structure prediction*, Proceedings of the National Academy of Sciences, **109**, no. 26, 10340-10345, 2012.
27. A. E. Dago, A. Schug, A. Procaccini, J. A. Hoch, M. Weigt, and H. Szurmant, *Structural basis of histidine kinase autophosphorylation deduced by integrating genomics, molecular dynamics, and mutagenesis*, Proceedings of the National Academy of Sciences, **109**, no. 26, E1733-E1742, 2012.
28. A. Voronin, M. Weiel, and A. Schug, *Including residual contact information into replica-exchange MD simulations significantly enriches native-like conformations*, PloS one, **15**, no. 11, e0242072, 2020.
29. G. Uguzzoni, S. John Lovis, F. Oteri, A. Schug, H. Szurmant, and M. Weigt, *Large-scale identification of coevolution signals across homo-oligomeric protein interfaces by direct coupling analysis*, Proceedings of the National Academy of Sciences, **114**, no. 13, E2662-E2671, 2017.
30. R. R. Cheng, F. Morcos, H. Levine, and J. N. Onuchic, *Toward rationally redesigning bacterial two-component signaling systems using coevolutionary information*, Proceedings of the National Academy of Sciences USA, **111**, no. 5, E563-E571, 2014.
31. M. Shibata, X. Lin, J. N. Onuchic, K. Yura, and R. R. Cheng, *Residue coevolution and mutational landscape for OmpR and NarL response regulator subfamilies*, Biophysical Journal, **123**, no. 6, 681-692, 2024.
32. A. Schug, *Residue coevolution and mutational landscape for OmpR and NarL: You can teach old dogs new tricks*, Biophysical Journal, **123**, no. 6, 653-654, 2024.
33. M. Figliuzzi, H. Jacquier, A. Schug, O. Tenaillon, and M. Weigt, *Coevolutionary landscape inference and the context-dependence of mutations in beta-lactamase TEM-1*, Molecular biology and evolution, **33**, no. 1, 268-280, 2016.
34. F. Pucci, M. B. Zerihun, M. Rooman, and A. Schug, *pycofitness – Evaluating the fitness landscape of RNA and protein sequences*, Bioinformatics, **40**, no. 2, btae074, 2024.
35. C. Weinreb, A. J. Riesselman, J. B. Ingraham, T. Gross, C. Sander, and D. S. Marks, *3D RNA and Functional Interactions from Evolutionary Couplings*, Cell, **165**, no. 4, 963-975, 2016.
36. F. Pucci and A. Schug, *Shedding light on the dark matter of the biomolecular structural universe: Progress in RNA 3D structure prediction*, Methods, **162-163**, 68-73, 2019.
37. F. Pucci, M. B. Zerihun, E. K. Peter, and A. Schug, *Evaluating DCA-based method performances for RNA contact prediction by a well-curated dataset*, RNA, rna-073809, 2020.
38. M. B. Zerihun, F. Pucci, and A. Schug, *CoCoNet – boosting RNA contact prediction by convolutional neural networks*, Nucleic Acids Research, **49**, no. 22, 12661-12672, 2021.
39. J. Martin, M. Lequerica Mateos, J. N. Onuchic, I. Coluzza, and F. Morcos, *Machine learning in biological physics: From biomolecular prediction to design*, Proceedings of the National Academy of Sciences, **121**, no. 27, e2311807121, 2024.

40. A. Kryshtafovych, T. Schwede, M. Topf, K. Fidelis, and J. Moult, *Critical assessment of methods of protein structure prediction (CASP) – Round XIII*, Proteins: Structure, Function, and Bioinformatics, **87**, no. 12, 1011-1020, 2019.
41. A. W. Senior, R. Evans, J. Jumper, J. Kirkpatrick, L. Sifre, T. Green, C. Qin, A. Žídek, A. W. R. Nelson, A. Bridgland, H. Penedones, S. Petersen, K. Simonyan, S. Crossan, P. Kohli, D. T. Jones, D. Silver, K. Kavukcuoglu, and D. Hassabis, *Improved protein structure prediction using potentials from deep learning*, Nature, **577**, no. 7792, 706-710, Jan 2020.
42. M. AlQuraishi, *End-to-End Differentiable Learning of Protein Structure*, Cell Systems, **8**, no. 4, 292-301.e3, 2019.
43. S. Wang, S. Sun, Z. Li, R. Zhang, and J. Xu, *Accurate De Novo Prediction of Protein Contact Map by Ultra-Deep Learning Model*, PLOS Computational Biology, **13**, no. 1, e1005324, 2017.
44. R M. Rao, J. Liu, R. Verkuil, J. Meier, J. Canny, P. Abbeel, T. Sercu, and A. Rives, *MSA Transformer*, in: Proceedings of the 38th International Conference on Machine Learning, **vol. 139** of *Proceedings of Machine Learning Research*, 8844-8856, 2021.
45. Z. Lin, H. Akin, R. Rao, B. Hie, Z. Zhu, W. Lu, N. Smetanin, A. dos Santos Costa, M. Fazel-Zarandi, T. Sercu, S. Candido et al., *Language models of protein sequences at the scale of evolution enable accurate structure prediction*, bioRxiv:2022.07.20.500902, 2022.
46. R. J. L. Townshend, S. Eismann, A. M. Watkins, R. Rangan, M. Karelina, R. Das, and R. O. Dror, *Geometric deep learning of RNA structure*, Science, **373**, no. 6558, 1047-1051, 2021.
47. M. AlQuraishi, *End-to-end differentiable learning of protein structure*, Cell systems, **8**, no. 4, 292-301, 2019.
48. S. Hochreiter and J. Schmidhuber, *Long Short-Term Memory*, Neural Computation, **9**, no. 8, 1735-1780, 11 1997.
49. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Zidek, A. Potapenko et al., *Highly accurate protein structure prediction with AlphaFold*, Nature, **596**, no. 7873, 583-589, 2021.
50. M. Baek, F. DiMaio, I. Anishchenko, J. Dauparas, S. Ovchinnikov, G. R. Lee, J. Wang, Q. Cong, L. N. Kinch, R. D. Schaeffer et al., *Accurate prediction of protein structures and interactions using a three-track neural network*, Science, **373**, no. 6557, 871-876, 2021.
51. F. B. Fuchs, D. E. Worrall, V. Fischer, and M. Welling, *Se(3)-transformers: 3d roto-translation equivariant attention networks* in: Proceedings of the 34th International Conference on Neural Information Processing Systems, **166**, 1970-1981, 2020.
52. J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A. J. Ballard, J. Bambrick et al., *Accurate structure prediction of biomolecular interactions with AlphaFold 3*, Nature, **630**, 493-500, 2024.
53. O. Taubert, F. von der Lehr, A. Bazarova, C. Faber, M. Weiel, P. Knechtges, C. Debus, D. Coquelin, A. Basermann, A. Streit, S. Kesselheim, M. Götz, and A. Schug, *RNA Contact Prediction by Data Efficient Deep Learning*, Communications Biology, **6**, 913, 2023.
54. J. K. H. Franke, F. Runge, and F. Hutter, *Scalable Deep Learning for RNA Secondary Structure Prediction*, 2023, arXiv:2307.10073.

The NIC Excellence Projects

Optically Excited States of Two-Dimensional Semiconductors

Michael Rohlfing

University of Münster, Institute of Solid-State Theory, 48149 Münster, Germany
E-mail: Michael.Rohlfing@uni-muenster.de

Two-dimensional or layered semiconducting systems are subject to various external stimuli. Their optical spectra can be modified by structural modifications, fields applied from outside, and more. In this paper we summarise our findings concerning (i) layered materials under pressure, (ii) modifications due to point defects, and (iii) magnetic proximity effects between a semiconducting monolayer and a two-dimensional ferromagnet. The computational framework of all these investigations is given by *ab initio* many-body perturbation theory (GW theory and the Bethe-Salpeter equation), based on conventional density-functional theory.

1 Introduction

Layered semiconducting materials are highly interesting physical systems in reduced dimension, and serve as platforms for potential applications in spintronics, optoelectronics, and quantum information technology. They are proposed as potential single-photon emitters, quantum qubits, and more, and they exhibit quantum-physical phenomena like topology, (quantum) (spin) Hall effect, Mott-Hubbard transitions, and exciton Bose-Einstein condensation, to name just a few. One important ingredient in all of this is the possibility to modify the materials' (optoelectronic) properties beyond the perfectly ordered materials, and to apply external stimuli for tuning. Here we discuss three such possibilities.

For the numerical determination of all such properties we employ the standard *ab initio* procedure to address optoelectronic excitations[1,2]. It starts with conventional density-functional theory (DFT). Total-energy optimisation within the DFT ground state yields the mechanical equilibrium geometry of a given system (plus, if necessary, nuclear vibrations and thermodynamic properties like phase transitions). Thereafter, for a given geometry we employ the GW theory (as part of *ab initio* many-body perturbation theory, MBPT) to determine electronic excitations. This includes single electrons and single holes, their excitation energy (i.e., the energy levels or band structure) and their single-particle wave functions. MBPT includes all relevant exchange and correlation issues in terms of the electronic self-energy operator. Finally, on the basis of the GW band structure we solve the Bethe-Salpeter equation (BSE) for correlated electron-hole pair states (i.e., excitons). These determine the optical properties of a material, i.e. optical absorption, emission, luminescence, reflectivity, etc.

(i) A simple external modification of transition metal dichalcogenides (TMDCs) is the application of pressure[3–11]. The materials have high mechanical flexibility combined with the ability to withstand high strain levels without breaking. Mechanical strain strongly modifies the electronic band structure and the fundamental optical transitions, leading, for example, to an energetic shift of the exciton resonances. This renders external strain – besides electric fields – an important means of controlling the optical properties of 2D semiconductors. In Sec. 3 we demonstrate that the optical spectra react sensitively to

pressure applied in terms of a diamond anvil cell, and a careful comparison of theoretical results with measured data obtained by close colleagues allows to address the question to which extent the mechanical stress of the anvil cell is transferred to the TMDC sample inside[12].

(ii) Complementary to the properties of a perfectly ordered periodic system (may it be a two-dimensional layer or a three-dimensional crystal), point defects embedded in it provide additional functionalities. They combine physics known from single atoms or molecules with the structural stability of a crystalline network, facilitating their investigation and usage. One prominent and timely example are point defects in hexagonal boron nitride (hBN)[13–20], which are being discussed especially as potential single-photon emitters for quantum information technology. They seem to be ubiquitous in both natural and artificially grown hBN and yield optical emission in the visible spectrum (whereas hBN by itself is transparent in the visible and soft ultraviolet range). However, in spite of the great significance, it is amazing that experimental data on their chemical and structural nature is still sparse and partially contradictory. In Sec. 4 we discuss optoelectronic details of various point defects in hBN that have been suggested as being responsible for the characteristic emission[21,22].

(iii) In addition to monolayers, homobilayers, and naturally grown bulk materials, there is another class of two-dimensional systems that is getting more and more attention: heterobilayers. In here, a monolayer of one 2D system is deposited on a monolayer of another 2D system, thus constituting a junction. Here we investigate the specific case of WSe_2 in contact with CrI_3[23–31], with particular focus on the excitons of the former. CrI_3 is a ferromagnetic semiconductor down to the monolayer, i.e. it exhibits magnetisation from spin polarisation and electronic states which are different in the majority spin channel (which constitutes the spin polarisation) and the minority spin channel. In contact to WSe_2, the spin polarisation partially transfers to the WSe_2 monolayer, which then also shows different physics in the majority and minority spin channels, finally affecting the WSe_2 excitons (see Sec. 5)[32].

Before discussing the above-mentioned three cases, a short summary of the theoretical formalism is presented in Sec. 2.

2 Theoretical Framework

The outline of *ab initio* many-body perturbation theory (MBPT) has been discussed in numerous publications (see, e.g., Ref. 1 and references therein). It usually starts from density-functional theory (DFT), which is typically carried out first anyway to provide the geometric structure of the material in question from total-energy minimisation. Thereafter, MBPT describes single-particle and two-particle excitations as effective individual particles on the background of the interacting many-electron system of the material, in terms of Green functions and their equation of motion[33–35]. Single-particle excitations refer to the addition of removal of a single electron, the energetics of which is reflected in the band structure or energy-level diagram of a system (occupied valence states refer to the removal of an electron, empty conduction states to the addition of an electron), and are described by the single-particle Green function. The net effects of electronic interaction and of exchange effects (from Pauli's principle) are expressed in form of the electronic self-energy operator, which is commonly evaluated on the level of the so-called GW approximation[1,33]. Two-

particle excitations occur in form of correlated electron-hole pair states, commonly known as excitons, and are described by the two-particle Green function and its equation of motion (the Bethe-Salpeter equation, BSE)[1,2]. Here we only mention some aspects relevant tor the application of the established theory to the case of two-dimensional semiconductor systems[36].

It has been well established that excitons in a semiconducting material are described by the Bethe-Salpeter equation (BSE), applied to a representation of the exciton in terms of a linear combination of independent interband transitions. Within the so-called Tamm-Dancoff approximation, an exciton can be expanded as

$$|S\rangle = \int d^3k \sum_{v,c} A_{v,c}^{(S)}(\mathbf{k}) \, |(v\mathbf{k}) \rightarrow (c, \mathbf{k}+\mathbf{Q})\rangle. \tag{1}$$

In here, v/c denotes occupied and empty bands, $\mathbf{Q}$ is the total momentum of the exciton (often close to zero) and the $\mathbf{k}$-space integration covers the first Brillouin zone. For numerical feasibility we have to replace the continuous integration by a finite summation oder $\mathbf{k}$ points:

$$|S\rangle = \sum_{\mathbf{k}_i} \sum_{v,c} A_{v,c}^{(S)}(\mathbf{k}_i) \, |(v\mathbf{k}_i) \rightarrow (c, \mathbf{k}_i+\mathbf{Q})\rangle. \tag{2}$$

Each $\mathbf{k}_i$ represents a volume V_i in reciprocal space (usually all V_i are of equal size and shape). The expansion coefficient $A(\mathbf{k}_i)$ is supposed to represent the average of the original ones, $A(\mathbf{k}_i) = 1/V_i \int_{V_i} A(\mathbf{k}) d^3k$. Finite sampling makes only sense if $A(\mathbf{k})$ varies only weakly within V_i. This set of $\mathbf{k}$-points defines all further requirements of the algorithm.

After defining the excitation in Eq. 1, its equation of motion is given by the BSE in the following way (omitting band indices for brevity sake):

$$\Delta E(\mathbf{k}) A(\mathbf{k}) - \int W(\mathbf{k}-\mathbf{k}') A(\mathbf{k}') d^3\mathbf{k}' = \Omega A(\mathbf{k}) \tag{3}$$

with $\Delta E(\mathbf{k})$ denoting band-energy differences, $W(\mathbf{q})$ denoting the screened Coulomb interaction and Ω the excitation energy. When using Eq. 2, the BSE turns into

$$\Delta E(\mathbf{k}_i) A(\mathbf{k}_i) - \sum_j \tilde{W}(\mathbf{k}_i - \mathbf{k}_j) A(\mathbf{k}_j) = \Omega A(\mathbf{k}_i). \tag{4}$$

In here,

$$\tilde{W}(\mathbf{k}_i - \mathbf{k}_j) := 1/V_j \int_{V_j} W(\mathbf{k}_i - \mathbf{k}') d^3\mathbf{k}' \tag{5}$$

is the integral of $W(\mathbf{q})$ over a (little) volume V_j around ($\mathbf{k}_i$-$\mathbf{k}_j$). For small reciprocal-space distance $\mathbf{k}_i$-$\mathbf{k}_j$, we employ an analytically known model for $W(\mathbf{k}_i - \mathbf{k}')$ and carry out the integration of Eq. 5 numerically[36]. For large reciprocal-space distance, $\tilde{W}(\mathbf{k}_i - \mathbf{k}_j) \approx W(\mathbf{k}_i - \mathbf{k}_j)$. Note that the treatment of W in the electron-hole interaction for the BSE must be equivalent to the treatment of W in the underlying GW (or LDA+GdW) band-structure calculation, especially concerning anisotropic behaviour of $W(\mathbf{q})$ at small momentum and the finite set of $\mathbf{k}$ points used in Eq. 2. These two issues, i.e. details of the band-structure calculation and of the BSE, must exactly correspond to each other to reach the numerical stability we need for the discussion in the next sections[36].

The underlying reason is the one-to-one correspondence (within MBPT) between the GW self-energy operator Σ^{GW} and the direct part of the corresponding electron-hole interaction kernel derived from Σ^{GW}:

$$\begin{aligned} \Sigma(1,2) &= iG(1,2)W(1^+,2) \\ \Longrightarrow K^d(13,24) &= \partial\Sigma(1,2)/\partial G(4,3) = iW(1^+,2)\delta(1,4)\delta(2,3) \end{aligned}$$

where we have made the usual approximation that $\partial W(1^+,2)/\partial G(4,3) \approx 0$. Apparently, for consistency between GW and BSE, the GW part of the MBPT should employ the identical screened Coulomb interaction as the BSE. This implies using exactly the same $\mathbf{q}$-point grid for the internal summation leading to the self energy, as well as, employing exactly the same modified interaction $\tilde{W}(\mathbf{q})$.

3 TMDC Materials under Pressure

When pressure is applied to a layered material, e.g. MoS_2, the effects on the structure are highly anisotropic. On the one hand, lateral compression will occur, as in any crystalline material, and with similar magnitude (up to 1 percent for 1 GPa of pressure). On the other hand, the weak interaction between the layers will result in much stronger vertical compression (several percent for 1 GPa of pressure) since the material is very soft in this direction. For small enough pressure, the compression (or, more generally, strain) is proportional to pressure (or, more generally, to the stress), as expressed by the elastic constants of the material. We have derived the elastic constants from a vast number of density-functional theory (DFT) calculation for various structural deformations, all carried out with the gradient-corrected PBE (Perdew-Burke-Ernzerhof) exchange-correlation functional[37] with semi-empirical van der Waals corrections as proposed by Grimme et al.[38]. Our results are: C_{11} = 218 GPa, C_{12} = 50 GPa, C_{13} = 5 GPa, and C_{33} = 21 GPa, in close agreement with available experimental data[12].

Concerning the response of optoelectronic excitations, it turns out that lateral compression shifts the excitons towards higher energy (blue-shift), while perpendicular compression shifts the excitons towards lower energy (red-shift). The latter effect is weaker than that of lateral compression, but would dominate if lateral compression were excluded. This might happen if a TMDC monolayer is solidly glued to an (incompressible) substrate, such that applied pressure leads exclusively to perpendicular compression. The spectra in Fig. 1 show the corresponding shifts of excitons, i.e. blue-shift to higher energies for hydrostatic, isotropic pressure including lateral compression (middle panel c), and red-shift to lower energies for anisotropic, exclusively perpendicular compression without change of the lateral lattice constant (right panel d).

The data shown in Fig. 1 can easily be interpreted in terms of geometric modifications. The vertical compression (see panels b and d) reduces only the layer-to-layer distance d and the thickness of the vacuum layer at a rate of about –0.10 Å/GPa at low pressure, gradually reducing to –0.03 Å/GPa at a higher pressure of 10 GPa. As a consequence of this layer-to-layer compression, the direct gap of the band structure shrinks from 2.53 eV at zero pressure to 2.43 eV at 10 GPa, i.e., by (on average) –10 meV/GPa. Consequently, all optical transitions are red-shifted upon increasing pressure by about this amount (see Fig. 1 d).

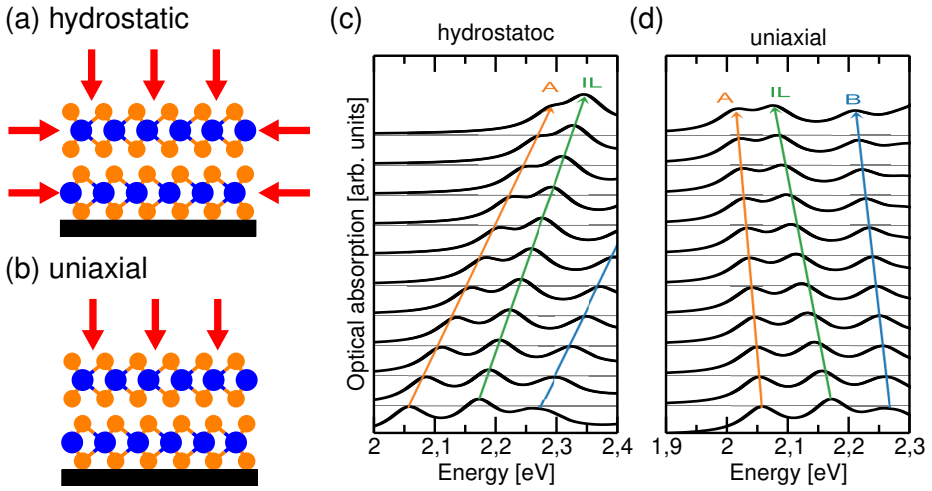


Figure 1. Left panels: schematic application of (a) isotropic (i.e., hydrostatic) pressure and (b) vertical uniaxial stress on a bilayer of MoS_2. Middle panel (c): response of BSE absorption spectra to pressure. The data show the optical absorption of a MoS_2 bilayer, for pressure from 0 GPa (bottom) to 10 GPa (top) in steps of 1 GPa. When hydrostatic/isotropic pressure is applied, the characteristic excitons (A, IL=interlayer, and B) excitons shift *up* in energy. Right panel (d): When stress is only applied perpendicular (e.g., because the bilayer is solidly fixed to a substrate and does not move/relay laterally, cf. schematic panel (b)), the characteristic excitons (A, IL, and B) excitons shift *down* in energy. All spectral data are calculated within many-body perturbation theory in the LDA+GdW approximation. For details, see Ref. 12.

In contrast, in the calculations for hydrostatic pressure, we let the lateral lattice constant a relax as well, at a rate of –7 mÅ/GPa. The effect of lateral (in-plane) deformation on the direct band gap is well-known. While biaxial tensile strain leads to a red-shift, in-plane compressive strain blue-shifts all optical excitations, also for MoS_2 monolayers. In our case, the band gap increases by about 30 meV/GPa, which overcompensates the gap reduction (–10 meV/GPa) of the simultaneous interlayer compression, such that in total the band gap grows from 2.53 to 2.75 eV under hydrostatic compression to 10 GPa, i.e., by +22 meV/GPa on average. This causes the strong blue-shift of all optical excitations upon hydrostatic pressure (Fig. 1 c). Note that the changes in the lateral lattice constant a are much weaker than those in the plane-to-plane distance d by a factor of $\delta d/\delta a \approx 3$, but the band structure is more sensitive to lateral in-plane compression than to out-of-plane compression by a factor of 10. Lateral stress will thus dominate, if permitted.

Corresponding experiments have been carried out in a diamond anvil cell with pressure of up to 10 GPa. The data show a red-shift slope of –3 meV/GPa for the A and IL (=inter-layer) excitons, which is in between our theoretical extrema of –10 meV/Gpa for uniaxial compression and +22 meV/GPa for the hydrostatic case. We conclude that the experimental reality lies in between our two extrema. The most plausible explanation is that the pressure of the diamond anvil cell is not transferred completely to the MoS_2 sample; instead, the lowest monolayer tends to remain stuck to the substrate, and in total the situation is closer to the one in Fig. 1 b. If we assume that the full truth is a partial lateral slipping, the interpolation between our two theoretical extrema would allow to assume

that the lateral pressure transfer succeeds to about 20 percent, i.e. the true geometry under pressure is a mixture of 20 % of the hydrostastic case and 80 % of the uniaxial case[12].

4 Hexagonal Boron Nitride and Point Defects therein

Hexagonal boron nitride (hBN) is a van der Waals material with an optical gap of about 6 eV[13]. The attention to point defects in this material has increased in the past decade due to their potential use as room-temperature stable two-dimensional (2D) single-photon emitters for quantum computing[14–16]. Recently, progress has been made in the fabrication of hBN quantum emitters with reproducible and controllable properties and their integration into quantum circuits[17–20]. The tunability of properties of quantum emitters, e.g., by electric fields or by strain, is a desirable feature for quantum technological applications. Conversely, this tunability contributes to the understanding of the atomic structure of the defect, which still poses an unsolved problem due to the variety of properties of hBN emitters. The observation of Stark shift with an electric field perpendicular to the layers of hBN means that some defects may break the planar symmetry of the 2D material. Phonon side bands or the influence of the emitter's distance to flake boundaries are specific to the atomic structure.

In experimental reality, defects in boron nitride occur both in naturally grown crystals and in artificially synthesised samples. So far, a unique identification of the elemental composition and chemical nature of the defects is difficult, since their existence is often only proved indirectly by the occurrence of corresponding optoelectronic features (in particular, visible-spectrum luminescence at energies deep within the boron nitride band gap of 6 eV, which is in the far ultraviolet). This indirect evidence and unknown character of the defect poses, of course, an unsatisfactory situation, and detailed theoretical understanding might significantly improve the situation.

In this work we investigate two aspects of point defects in hexagonal boron nitride[21,22]. On the one hand, we evaluate their optoelectronic excitation energies within *ab initio* MBPT. On the other hand, we investigate the change of the defect's local geometry when it is optically excited, and evaluate geometrical deformation, the related reorganisation energy, and resulting Stokes shift. This may then be used to judge whether a hypothetically assumed point defect is realistic, i.e. whether it can be made responsible for optoelectronic properties of defects in experimental reality. In measurements, some defects (in particular, the candidates for single-photon generation) show optical excitation near 2 eV excitation energy, and exhibit weak reorganisation energy and Stokes shift.

Tab. 1 summarises our findings for a number of various defects that we have investigated[21,22]. C_B and C_N refer to a substitutional carbon atom, which replaces a boron or nitrogen atom, respectively. V_N refers to a nitrogen vacancy, i.e. a nitrogen atom which is simply missing. V_{NB} refers to a double vacancy, in which two neighbouring atoms are missing (one nitrogen and one boron). C_BC_N refers to two substitutional carbon atoms at neighbouring positions (one instead of boron, and the other one instead of nitrogen). In case of C_BO_N the carbon atom substitutes boron, and the oxygen atom substitutes a neighbouring nitrogen. C_BV_N refers to a substitutional carbon atom instead of boron, while simultaneously a neighbouring nitrogen atom is missing, constituting a vacancy. Among these, the first three defects exhibit luminescence near 2 eV, which might indicate that they could be the candidates found (but not clearly identified) in experiment. Among the more

Defect	Absorption Energy [eV]	Reorganisation Energy [eV]	Stokes Shift [eV]
C_B	2.0	~ 0	~ 0
C_N	2.0	~ 0	~ 0
V_N	2.2	~ 0	~ 0
V_{NB}	4.9		
C_BC_N	4.3	0.1	0.2
C_BO_N	1.9	0.4	0.8
C_BV_N	1.8	0.9	1.8

Table 1. Excited-state data for various point defects in hexagonal boron nitride. Column 1: Chemical composition. Column 2: Vertical excitation energy (in eV) of the lowest dipole-allowed transition, with the geometry being given by the ground-state equilibrium. Column 3: Mechanical reorganisation energy (in eV) from ground-state equilibrium into the excited-state equilibrium, from geometry optimisation while being in the excited state. Column 4: Stokes shift (in eV), as the energy difference between the vertical transition in the ground-state geometry and the vertical transition in the excited-state geometry. For details, see Refs. 21, 22.

complex double defects (V_{NB}, C_BC_N, C_BO_N, and C_BV_N) the first two shows excitation at much too high energy, while the other two show very strong geometry reorganisation in the excited state and concomitantly large Stokes shift between absorption and emission, much higher than observed in experiment, which excludes them as being responsible for the observed properties.

5 The Two-Dimensional Heterostructure WSe_2-CrI_3

Heterostructures of two-dimensional transition-metal dichalcogenides and ferromagnetic substrates are important candidates for the development of viable new spin- or valleytronic devices. A particular example is the interface between the TMDC tungsten sulfied (WSe_2) and the two-dimensional ferromagnet chromium iodide (CrI_3)[32]. Our main interest is the A exciton of the WSe_2 monolayer, which is fully analogous to the A exciton of MoS_2 as shown in Fig. 1 (at 2.05 eV at zero pressure; in case of a WSe_2 monolayer we find it at 1.6 eV). For simplicity we have assumed in the current study that the two materials can be stacked on top of each other in a 1×1 unit cell, neglecting lattice mismatch and rotational misalignment. We have restricted the study to just one configuration in which the selenium and iodine anions form a hollow-site registry, such that no two atoms are on top of each other. For this specific situation we find a significant influence of the ferromagnetic spin polarisation of CrI_3 on the (opto-)electronic structure of WSe_2. CrI_3 has occupied majority spin electrons ($\uparrow$) that are several eV below their minority-spin counterpart orbitals. In this majority-spin channel this leads to the occurrence of weakly dispersing CrI_3 empty states (indicated as thick green bars in Fig. 2) slightly below the WSe_2 lowest conduction bands. The interesting physics of WSe_2 occurs at the K^- and K^+ point of its Brillouin zone, at which the relevant bands are completely spin polarised. At K^-, where we observe in the highest valence band the same spin in WSe_2 as the majority spin of the CrI_3 ferromagnet, the spins show exchange effects across the interface due to orbital overlap between the anions (S and I) and corresponding hybridisation. At K^+, on the other hand, the spin of the highest valence band of WSe_2 is the opposite of the CrI_3 majority spin, and does

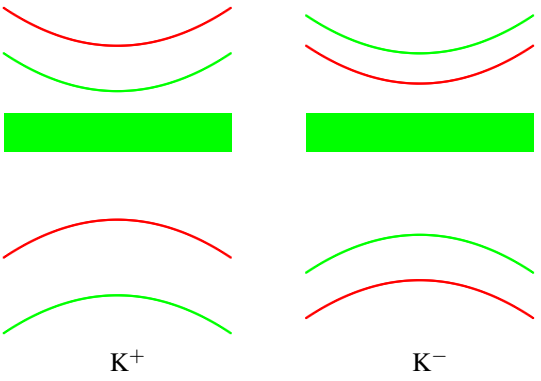


Figure 2. Schematic illustration of the dispersive valence and conduction bands of the WSe_2 monolayer and the nearly flat majority-spin conduction states of the crI_3 monolayer in direct contact of the two monolayers. All majority-spin (minority-spin) states are illustrated as green (red). Near the two independent K points of the Brioullin zone (K^+ and K^-) the WSe_2 bands are fully spin polarised, with alternating sequence. The sequence at K^+ is the opposite of the sequence at K^-. Due to magnetic proximity and the orbital overlap between S and I atoms, the spin-majority states of WSe_2 (green) are shifted in energy, while the spin-minority states (red) remain unaffected. For details, see Ref. 32.

not interact, such that no exchange is observed. In short, all WSe_2 majority-spin bands (shown in green in Fig. 2) are shifted in energy due to the magnetic proximity of CrI_3, while the WSe_2 minority-spin bands (shown in red in Fig. 2) remain unaffected. This leads to energetic splitting between the WSe_2 valence bands of 1.6 meV between K^- and K^+. This shift translates to an energetic splitting of 3.6 meV between the intravalley excitons A^- and A^+ at K^- and K^+. Since these two excitons can be measured by either positively or negatively oriented circularly polarised light, such energy splitting is directly observable in experiment.

Furthermore, we have found that the spin polarisation significantly changes the quantum-mechanical composition of the excitons, as well, in addition to the shifts in energy. At the K^+ points, the TMDC exciton (A^+) is in the spin channel corresponding to the magnet's minority spin ($\downarrow$, indicated as red in Fig. 2), preventing hybridisation. This exciton remains an *intralayer* exciton as known from a freestanding monolayer. At the K^- point, the TMDC exciton (A^-) is in the spin channel corresponding to the magnet's majority spin ($\uparrow$, indicated as green in Fig. 2). In this spin channel the magnet provides empty bands in the same energy range as the semiconductor conduction band. This provides charge-transfer configurations between semiconductor and magnet. They hybridise with the semiconductor intralayer excitations, which therefore acquires partial interlayer character. In the spectra this leads to significantly reduced dipole strength of one of the Zeeman-split peaks, as a clear signature of quantum-mechanical hybridisation across the van der Waals gap between the two systems. Such energetic splitting and difference in intensity has been observed experimentally, supporting our concept that the behaviour of the excitons at K^+ is basically different from those at K^-.

In addition, the different composition of the excitons at K^- and K^+ (with and without contribution of charge-transfer configurations across the interface) also leads to different behaviour in a magnetic fields. In such a field, the excitons observe Zeeman shifts, i.e. a

shift in excitation energy proportional to the field strength. The proportionality is expressed in terms of a so-called g factor (gyromagnetic ratio). In our calculations we find that the g factors of the two excitons (at K^- and K^+) differ by about 10 %, which should also be detectable in experiment.

6 Concluding Remarks

In this paper we have discussed excitonic states in two-dimensional semiconductors and their change when the system in question is more complex than just a simple monolayer or bulk material with perfectly ordered periodic crystal structure. Defects, mechanical deformation, and magnetic proximity effects significantly modify the (opto)electronic response, which can in turn be used to get detailed insight into the material's microscopic internal structure from optical experiment. In the three examples discussed here, (i) the energetic shifts of excitons of MoS_2 under pressure depend on the question if MoS_2 keep sticking to its substrate or not, (ii) the energy and line shape of defect states in hBN can be used to identify or rule out microscopic models of the defect, and (iii) proximity effects of a ferromagnet in direct contact splits the A^+ and A^- exciton of a WSe_2 monolayer, which are degenerate in the bare monolayer by itself. All these examples illustrate the crucial role of the structure and geometry of low-dimensional semiconductors for their optoelectronic properties.

Acknowledgements

The authors thanks T. Deilmann, J.-H. Graalmann, M.-C. Heissenbüttel, A. Kirchhoff, P. Krüger, and P. Marauhn for many fruitful discussions, and gratefully acknowledges the Gauss Centre for Supercomputing e.V. (`www.gauss-centre.eu`) for funding this project by providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS Supercomputer JUWELS[39] at Jülich Supercomputing Centre (JSC).

References

1. G. Onida, L. Reining, and A. Rubio, Rev. Mod. Phys. **74**, 601, 2002.
2. M. Rohlfing and S. G. Louie, Phys. Rev. B **62**, 4927, 2000.
3. M.-Y. Li et al., Nature **567**, 169-170, 2019.
4. J. R. Schaibley et al., Nat. Rev. Mater. **1**, 16055, 2016.
5. B. Urbaszek et al., Nature **567**, 39-40, 2019.
6. S. Bertolazzi et al., ACS Nano **5**, 9703-9709, 2011.
7. H. J. Conley et al., Nano Lett. **13**, 3626-3630, 2013.
8. A. Castellanos-Gomez et al., Nano Lett. **13**, 5361-5366, 2013.
9. J. Kern et al., Adv. Mater. **28**, 7101-7105, 2016.
10. R. Schmidt et al., 2D Mater. **3**, 021011, 2016.
11. X. He et al., Appl. Phys. Lett. **109**, 173105, 2016.
12. P. Steeger et al., Nano Lett. **23**, 8947-8952, 2023.
13. Y. Kubota, K. Watanabe, O. Tsuda, and T. Taniguchi, Science **317**, 932, 2007.
14. T. T. Tran et al., Nat. Nanotechnol. **11**, 37, 2016.

15. A. Kubanek, Adv. Quantum Technol. **5**, 2200009, 2022.
16. M. E. Turiansky, A. Alkauskas, and C. G. Van de Walle, Nat. Mater. **19**, 487, 2020.
17. S. Choi et al., Mater. Interfaces **8**, 29642, 2016.
18. N. Chejanovsky et al., Nano Lett. **16**, 7037, 2016.
19. Y. Chen et al., Small **17**, 2008062, 2021.
20. C. Zhang, Z. Shi, T. Wu, and X. Xie, Adv. Opt. Mater. **10**, 2200207, 2022.
21. A. Kirchhoff, T. Deilmann, P. Krüger, and M. Rohlfing, Phys. Rev. B **106**, 045118, 2022.
22. A. Kirchhoff, T. Deilmann, and M. Rohlfing, Phys. Rev. B **109**, 085127, 2024.
23. F. Subhan and J. Hong, J. Phys. Chem. C **124**, 7156, 2020.
24. Q. Zhang et al., Adv. Mater. **28**, 959, 2016.
25. B. Amin et al., Phys. Rev. B **92**, 075439, 2015.
26. L. Liang and V. Meunier, Nanoscale **6**, 5394, 2014.
27. Y. Gong et al., Nat. Mater. **13**, 1135, 2014.
28. Y. Yu et al., Nano Lett. **15**, 486, 2015.
29. A. Srivastava et al., Nat. Phys. **11**, 141, 2015.
30. C. Zhao et al., Nat. Nanotechnol. **12**, 757, 2017.
31. T. Li et al., Nat. Commun. **9**, 3344, 2018.
32. M.-C. Heißenbüttel, T. Deilmann, P. Krüger, and M. Rohlfing Nano Lett. **21**, 5173-5178, 2021.
33. L. Hedin, Phys. Rev. **139**, A796, 1965.
34. M. S. Hybertsen and S. G. Louie, Phys. Rev. B **34**, 5390, 1986.
35. G. Strinati, Rivista del Nuovo Cimento **11**, 1, 1988.
36. M. Drüppel, T. Deilmann, J. Noky, P. Marauhn, P. Krüger, and M. Rohlfing, Phys. Rev. B **98**, 155433, 2018.
37. J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865, 1996.
38. S. Grimme, J. Comput. Chem. **27**, 1787, 2006.
39. Jülich Supercomputing Centre, *JUWELS: Modular Tier-0/1 Supercomputer at the Jülich Supercomputing Centre*, Journal of large-scale research facilities **5**, A135, 2019, doi:10.17815/jlsrf-5-171.

Copper Transfer Mechanism to and Structural Dynamics of Plant Receptor ETR1

Lisa Sophie Kersten[1], Michele Bonus[1], Stephan Schott-Verdugo[2], and Holger Gohlke[1,2]

[1] Institute for Pharmaceutical and Medicinal Chemistry, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany
E-mail: {lisa.kersten, michele.bonus, gohlke}@hhu.de

[2] Institute of Bio- and Geosciences (IBG-4: Bioinformatics), Forschungszentrum Jülich, 52425 Jülich, Germany
E-mail: {s.schott-verdugo, h.gohlke}@fz-juelich.de

The structure, copper transport, and mechanism of action of the plant receptor ETR1 have long remained elusive, hampering the understanding of how ethylene perception is transformed into a downstream signal. We used *ab initio* modelling to generate the first structural model of the transmembrane sensor domain of ETR1. Through protein-protein docking and all-atom molecular dynamics simulations, we investigate the interaction between copper chaperones and ETR1. Additionally, all-atom molecular dynamics simulations were employed to explore the multimerisation of ETR1 and the interactions with the downstream target CTR1. This research project in combination with experimental validation of our data represents the most comprehensive analysis of ETR1 to date, providing new insights into its functional mechanisms.

1 Introduction

The small molecule ethylene is a gaseous plant hormone that affects various developmental processes in plants, such as seed germination, senescence, and fruit ripening. Ethylene is perceived by ethylene receptors located at the endoplasmatic reticulum membrane. In *Arabidopsis thaliana*, the model organism predominantly used to study ethylene-signalling, five receptor isoforms have been identified and associated with ethylene response, with ETR1 (Ethylene Response 1) being the best studied. ETR1 has three transmembrane α-helices at the N-terminus, forming the transmembrane sensor domain (TMD). The TMD also contains an essential cofactor, a Cu(I) ion, ensuring high affinity and specificity for binding of the chemically simple ethylene molecule. The cytoplasmic part of the receptor contains a GAF domain (named after its occurrence in cGMP-specific phosphodiesterases, adenylyl cyclases, and FhlA), followed by a dimerisation histidine-phosphotransfer domain (DHp), a catalytic ATP-binding domain (CD), and a receiver domain (RD). Functional ethylene receptors are homodimers, and higher-order oligomers have been described. Crystal structures of domains and structural homologs of the cytosolic domains of ETR1 are available, but there is no experimentally determined structural model of the TMD[1].

The molecular components involved in the transport of Cu(I) from the cellular plasma membrane to the ER membrane-bound ETR1 have been identified. They include the soluble copper chaperones ATX1 and CCH and the copper transporter RAN1. Their metal-binding domains share structural similarity[2]. However, a comprehensive understanding of the interactions between the copper chaperones and ETR1 at the atomistic level is lacking.

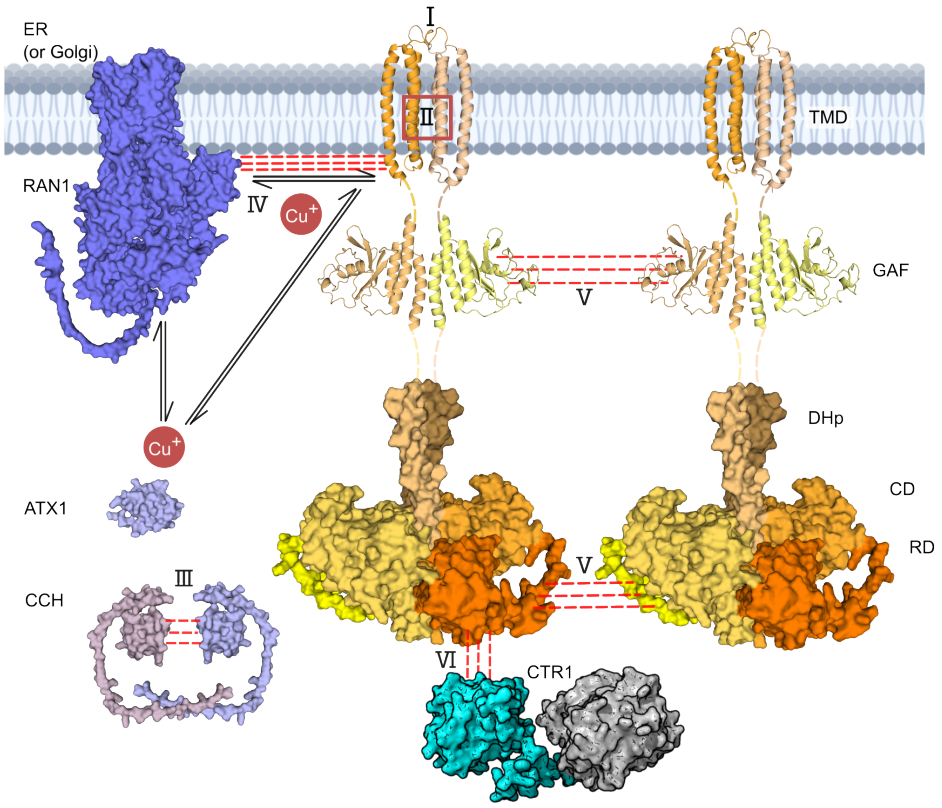


Figure 1. Overview of the project etr1. The first model of the TMD was created using *ab initio* modelling and experimentally validated (I). Further, the copper-binding site was characterised and possible binding sites of ethylene or antagonists were identified (II). Due to its special structural properties, the dimerisation behaviour of the copper chaperone CCH was investigated (III). Investigations are currently carried out to determine the atomic interactions of the chaperones and RAN1 with ETR1 (IV), as well as the multimerisation behaviour (V) and the interactions with the downstream target CTR1(VI). TMD: Transmembrane sensor domain, GAF: GAF domain, DHp: Dimerisation and histidine phosphotransfer domain, CD: Catalytic domain, RD: Receiver domain.

We thus aimed to elucidate the interactions established between the different components of the Cu(I) transport system and how Cu(I) is delivered to ETR1.

In the absence of ethylene, ETR1 functions as a negative regulator of the ethylene response. Once the plant biosynthesises ethylene, it diffuses throughout the plant or to neighbouring plants and coordinates with the Cu(I) cofactor. The Cu(I) ion is coordinated in ETR1 by Cys65 and His69 located in helix 2 of the TMD. Receptors bound to ethylene are thought to undergo conformational changes and, therefore, fail to activate downstream targets such as the serine/threonine-protein kinase CTR1[3], which finally triggers the ethylene response of the plant.

In our project, we generated the first structural model of the TMD of ETR1 and an

ETR1_TMD/Cu(I) dimer model by integrating *ab initio* structure prediction and coevolutionary information (Fig. 1 I). The obtained model was refined and independently validated by tryptophan scanning mutagenesis[4] as well as EPR spectroscopy[5]. Based on this model, we characterised the copper binding site in ETR1[6] and identified potential binding sites of ethylene and antagonists (Fig. 1 II). To shed light on how the Cu(I) cofactor is delivered to ETR1, we are investigating the dimerisation behaviour of CCH (Fig. 1 III), as well as the mechanism of Cu(I) transport from the chaperones to ETR1 (Fig. 1 IV). Finally, we are studying the multimerisation of ETR1 (Fig. 1 V) and its interaction with the downstream target CTR1 (Fig. 1 VI). This study is, to our knowledge, the most comprehensive analysis of ETR1, and it is expected to offer an in-depth understanding of its cellular functions.

2 Results

2.1 *Ab Initio* Modelling and Experimental Validation of the First Structural Model of the Transmembrane Sensor Domain in ETR1

The TMD of ETR1 from *Arabidopsis thaliana* was modelled *ab initio* before AlphaFold2 was available due to the lack of suitable homologous templates. Residues 1-117 were selected based on transmembrane topology, and secondary structure predictions. Using the RosettaMembrane membrane_abinitio2 protocol, 100,000 models were generated and filtered based on contact predictions and z-scores, yielding 5,217 structures. After clustering, the centroid structure of the largest cluster was selected and further refined, with side-chain configurations optimised through solvent accessibility analysis. In parallel, the TMD's Cu(I) stoichiometry was determined in Prof. Groth's laboratory at Heinrich Heine University Düsseldorf. Using the generated ETR1_TMD model and the determined Cu(I) stoichiometry, a dimeric model was generated using HADDOCK, coevolutionary signals, knowledge about lipophilicity regions, and characteristics of the copper binding site. The final dimer model was refined through molecular dynamics simulations and validated by alanine and tryptophan mutagenesis experiments resulting in the first structural model of ETR1_TMD[4] (Fig. 2A).

Soon after, AlphaFold2[7] predicted an alternative structural model of the ETR1 TMD, proposing a different helix arrangement, and hence, a different dimer interface and copper-binding site (UniProt: P49333) (Fig. 2B). To scrutinise which model better represents experimental findings, we combined site-directed spin labelling with electron paramagnetic resonance spectroscopy performed in collaboration with Prof. Drescher's laboratory at the University of Konstanz and obtained distance restraints for liposome-reconstituted ETR1 TMD on the orientation and arrangement of the transmembrane helices[5].

The experimental distance distributions were compared with distance distributions obtained by MMM, a programme for visualisation, inspection, generation, and improvement of models of proteins and protein assemblies based on restraints from multiple experimental techniques, using either TMD model. The experimental distance restraints are altogether in better agreement with the *ab initio* structural model[4] than with the AlphaFold2[7] prediction (Fig. 2C-E)[5]. However, since neither model is fully consistent with the EPR distances, work has always been continued with both models.

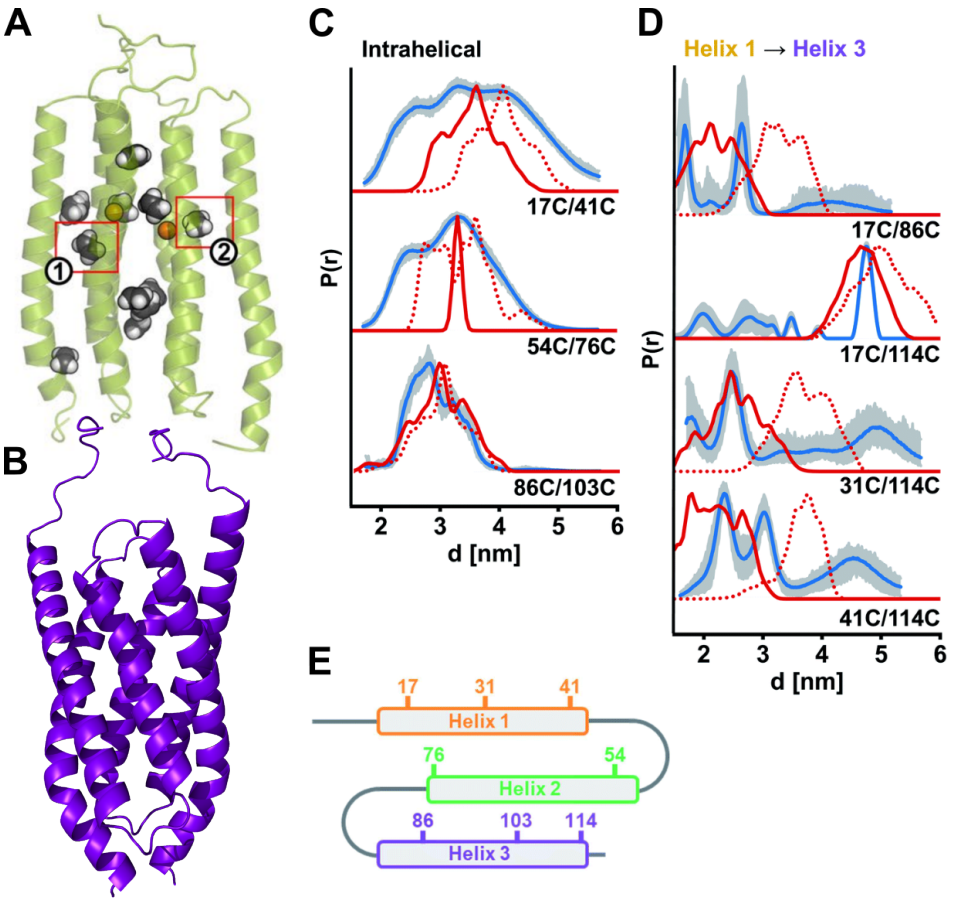


Figure 2. **A**: Representation of the *ab initio* structural model of the ETR1 TMD. **B**: Representation of the AlphaFold2 structural model of the ETR1 TMD. **C, D**: Experimental distance distributions obtained by DEER measurements (blue) with validation (grey area). Simulated distance distributions based on the *ab initio* model (2019, red) and the one from AlphaFold2 (2021, red dotted) are indicated. **C**: Intrahelical distances. **D**: Interhelical distances between helix 1 and helix 3. **E**: Schematic representation of the ETR1 TMD and spin-labelled sites used for DEER distance determinations. Panel A is taken from Ref. 4 and licensed under a Creative Commons Attribution 4.0 International License. Panels C, D and E were taken from Ref. 5 with permission from the Royal Society of Chemistry. Licensed under a Creative Commons Attribution 3.0 Unported Licence.

2.2 EXAFS and QM/MM Umbrella Sampling Simulation of the Copper Binding Site in ETR1

Extended X-ray Absorption Fine Structure (EXAFS) spectroscopy, performed in collaboration with Dr. Cutsail III and Prof. DeBeer at the MPI Mülheim, and with Professor Groth at Heinrich Heine University Düsseldorf, along with quantum mechanics/molecular mechanics umbrella sampling (QM/MM US) simulations, were used to further characterise the Cu(I) binding site in ETR1. The EXAFS results provided detailed insights into the local coordination environment of the copper ion. QM/MM US simulations completed these

findings by modelling the copper complex at the atomic level. The QM/MM US results agree with the EXAFS fit distance changes upon ethylene binding, particularly in the increase of the distance between His69 and Cu(I), and yield binding energetics comparable with experimental dissociation constants. Ethylene binding also results in changes to the C=C bond distance and dihedral angle of ethylene, consistent with hybridisation changes predicted by the Dewar-Chatt-Duncanson model. The observed changes in the copper coordination environment might be the triggering signal for the transmission of the ethylene response[6].

2.3 Structural Modelling of and Molecular Mechanics Generalised Born Surface Area (MM-GBSA) Calculations on the Copper Chaperone CCH Dimer

The copper chaperone CCH is one of the three chaperones characterised to be involved in Cu(I) transport to ETR1. While all share a common characteristic copper binding fold, CCH additionally features a C-terminal end, whose structure remains unresolved. Both TopProperty and DISOPRED suggest that this C-terminal region is intrinsically disordered. This hypothesis is further supported by NMR data from collaborators[8]. However, this C-terminal end may play a role in the dimerisation of CCH. To investigate the structure and dimerisation properties, ColabFold 1.5.2 was used to predict both the monomeric and dimeric structures of CCH. Based on the obtained CCH dimer, we conducted molecular dynamics simulations, computed on the JUWELS booster.

To pinpoint the amino acids critical for the stability of CCH-dimers, we conducted molecular mechanics/generalised Born surface area (MM-GBSA) calculations, including a per-residue decomposition of the effective energy. Our results indicate that residues of the dimer interface, as well as residues of the C-terminal end (T116, K117, and V121), significantly contribute to the binding energy (Fig. 3). Additionally, the Groth lab performed melting temperature measurements suggesting tighter monomer interactions in CCH compared to CCH lacking the C-terminal end. These results suggest that the C-terminal extension indirectly influences dimerisation and may play a role in copper transport and protection, highlighting distinct functional roles for CCH compared to its homolog ATX1[8].

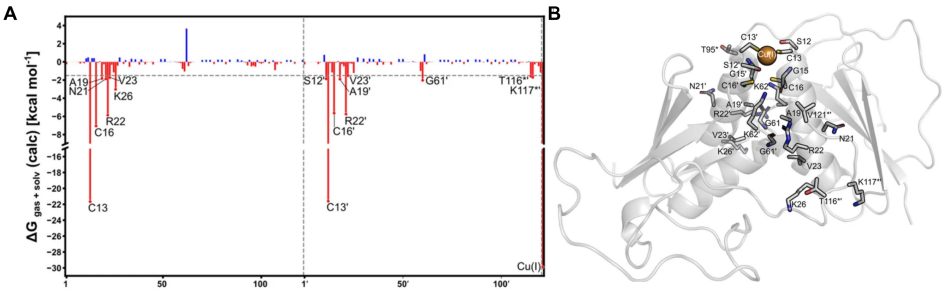


Figure 3. Identification of amino acids in the CCH-dimer that are crucial for dimer stability. **A**: Per residue decomposition of the binding effective energy of the Cu(I)-loaded CCH-dimer. **B**: Hot spot residues localised on the CCH-dimer. This figure was taken, in parts, from Ref. 8 and licensed under a Creative Commons Attribution 4.0 International License.

2.4 Mechanism of Cu(I) Transport from the Cell Membrane to ETR1, Multimerisation of ETR1, and its Interaction with the Downstream Target CTR1

To analyse how Cu(I) is transferred from the chaperones to the final target ETR1, we performed coevolution-informed protein-protein docking of ATX1, CCH, or RAN1 to ETR1. Possible interaction sites between ATX1, CCH, and RAN1 with ETR1 were predicted using GLINTER, a deep-learning method for predicting interaction sites in protein-protein complexes. The predicted protein-protein interactions were further used as ambiguous restraints to guide protein-protein docking with HADDOCK2.4. Currently, MD simulations of these complexes are performed on the JUWELS booster module. The obtained trajectories will be used to identify so-called 'hot-spot' residues, that significantly contribute to complex stability (Fig. 1 IV). Subsequently, these hot spots will be verified experimentally by mutagenesis studies to better understand the nature of complex formation and copper transfer dynamics. This research has the potential to provide new insights into the targeted regulation of ETR1 signalling.

Furthermore, the multimerisation behaviour of ETR1 (Fig. 1 V) and interaction with the downstream target CTR1 (Fig. 1 VI) are investigated on an atomistic level. We are currently performing unbiased MD simulations of the corresponding protein-protein complexes predicted with ColabFold 1.5.2. Therefore, fifty independent MD simulation replicas of 500 ns are performed on the JUWELS booster module. To identify 'hot-spot' residues, the obtained MD trajectories will be used to perform MM-GBSA calculations in combination with a decomposition of the effective energy of dimerisation at the single-residue level. Insights into the interaction of ETR1 with downstream targets and other receptors will enhance our understanding of ETR1's function in the broader cellular context, potentially offering new starting points for targeted regulation of ETR1 signalling.

3 Conclusion

Our research presented the first structural model of the transmembrane sensor domain (TMD) in ETR1, validated through mutagenesis studies. Comparison with AlphaFold2 predictions revealed that the *ab initio* model aligns more accurately with experimental data. We also confirmed that ethylene binds to the copper cofactor within the TMD, a finding further supported by spectroscopic approaches. Ongoing work is focused on investigating copper delivery to ETR1 and its related molecular components, the receptor's multimerisation behaviour, and its interactions with downstream targets such as CTR1. This study offers the most detailed structural analysis of ETR1 to date, advancing our understanding of the signalling mechanism and providing experimentally testable hypotheses on the biological functions. In the long term, these insights should contribute to ensure food security, as ETR1 plays a key role in post-harvest spoilage.

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Projektnummer 267205415 – SFB 1208 (project A03 to H. G.). The authors gratefully acknowledge fruitful discussions with Georg Groth, Dieter Willbold,

and Nils Lakomek, the wet-lab contributions of their colleagues as detailed in the referenced publications, and the computing time granted by the John von Neumann Institute for Computing (NIC) and provided on the supercomputer JUWELS at the Jülich Supercomputing Centre (JSC) (project IDs: HKF7, VSK33, etr1). We are grateful to the "Zentrum für Informations- und Medientechnologie" (ZIM) at the Heinrich Heine University Düsseldorf for computational support.

References

1. B. M. Binder, *Ethylene signaling in plants*, J Biol Chem, **295**, 7710-7725, 2020.
2. C. Hoppen, L. Muller, S. Hansch, B. Uzun, D. Milic, A. J. Meyer, S. Weidtkamp-Peters, and G. Groth, *Soluble and membrane-bound protein carrier mediate direct copper transport to the ethylene receptor family*, Sci Rep, **9**, 10715, 2019.
3. K. L. Clark, P. B. Larsen, X. Wang, and C. Chang, *Association of the Arabidopsis CTR1 Raf-like kinase with the ETR1 and ERS ethylene receptors*, Proc Natl Acad Sci U S A, **95**, 5401-5406, 1998.
4. S. Schott-Verdugo, L. Muller, E. Classen, H. Gohlke, and G. Groth, *Structural Model of the ETR1 Ethylene Receptor Transmembrane Sensor Domain*, Sci Rep, **9**, 8869, 2019.
5. A. Kugele, B. Uzun, L. Müller, S. Schott-Verdugo, H. Gohlke, G. Groth, and M. Drescher, *Mapping the helix arrangement of the reconstituted ETR1 ethylene receptor transmembrane domain by EPR spectroscopy*, RSC Advances, **12**, 7352-7356, 2022.
6. G. 3rd. Cutsail, S. Schott-Verdugo, L. Muller, S. DeBeer, G. Groth, and H. Gohlke, *Spectroscopic and QM/MM studies of the Cu(I) binding site of the plant ethylene receptor ETR1*, Biophys J, **121**, 3862-3873, 2022.
7. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Zidek, and A. Potapenko, *Highly accurate protein structure prediction with AlphaFold*, Nature, **596**, 583-589, 2021.
8. D. Dluhosch, L. S. Kersten, S. Schott-Verdugo, C. Hoppen, M. Schwarten, D. Willbold, H. Gohlke, and G. Groth, *Structure and dimerization properties of the plant-specific copper chaperone CCH*, Sci Rep, **14**, 19099, 2024.

Swimming and Swarming of Self-Propelled Cognitive Particles with Hydrodynamic Interactions

Segun Goh[1], Elmar Westphal[2], Roland G. Winkler[1], and Gerhard Gompper[1]

[1] Theoretical Physics of Living Matter, Institute for Advanced Simulation, Forschungszentrum Jülich, 52425 Jülich, Germany
E-mail: {s.goh, r.winkler, g.gompper}@fz-juelich.de

[2] Peter Grünberg Institute and Jülich Centre for Neutron Science, Forschungszentrum Jülich, 52425 Jülich, Germany
E-mail: e.westphal@fz-juelich.de

Sensing of the environment and information processing, combined with motility, are fundamental characteristics of life, from the largest animals to the smallest single-cell organisms. Adaptive self-steering gives rise to fascinating phenomena, ranging from large-scale collective behaviours denoted as swarming, as observed in mammalian herds, flocks of birds, schools of fish, or even cell layers and tissues, to the formation of bacterial biofilms. We study the collective behaviour of cognitive self-steering microswimmers using large-scale hydrodynamics simulations, applying a particle-based mesoscale hydrodynamics approach in combination with the squirmer model (prescribed surface flows on a spherical body) for microswimmers. The self-steering is governed by a combination of local alignment of the propulsion directions, and the joining of other swimmers due to (non-reciprocal) directional sensing. Our results show several types of self-organisation, like active turbulence, the formation of swirls and jets, and the emergence of elongated swarms, depending on the maneuverability of the microswimmer and the propulsion type (puller or pusher).

1 Introduction

The capability of motile organisms to sense the environment, to process information, and adapt their behaviour is a fundamental aspect of life. An important result of this ability is the collective dynamics of many identical individuals[1]. Examples range from macroscopic to microscopic length scales, from flocks of birds, schools of fish, mammalian herds, and groups of people, to swarms of insects, bacterial biofilms, and cellular aggregates. The purposes of these collective movements include the search for food, protection against predators, and enhanced motility. The mechanisms of collective motion deduced from the behaviour of natural systems can be employed in the design and construction of artificial systems, such as microscopic robots ("microbots")[2].

Many biological motile organisms live in aqueous environments, which implies that the hydrodynamics of the medium strongly affects or even dominates the collective dynamics[3]. It is of course also fundamental for the self-propulsion and navigation by swimming, as well as the hydrodynamic interactions between swimmers. The elucidation of the adaptive behaviour of microorganisms and microbots requires a suitable model for hydrodynamically self-steering cognitive microswimmers that can adjust their movement according to gathered information.

From a simulation point of view, studies of wet systems are much more computing-time intensive compared to dry systems[4], as many more degrees of freedom have to be taken into account, and also because hydrodynamic interactions are long range and decaying with a

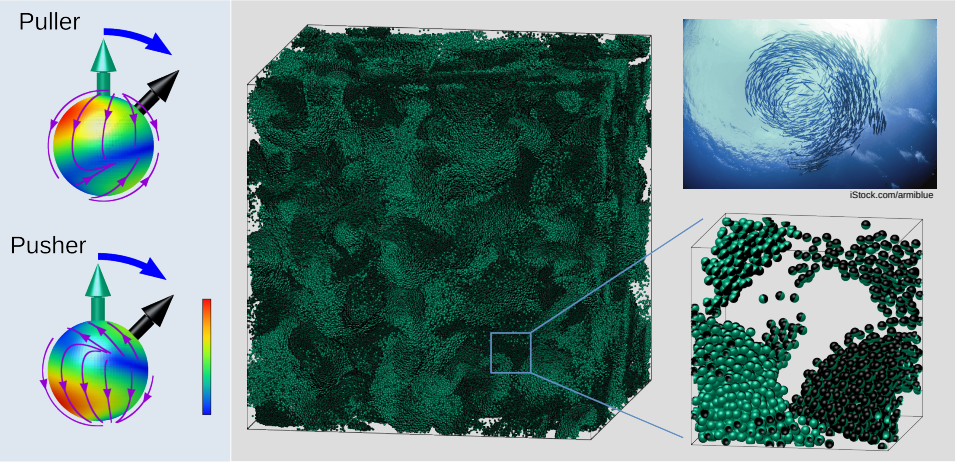


Figure 1. Left: Non-axisymmetric surface flows for pullers (upper panel) and pushers (lower panel), which allow microswimmers to self-steer, i.e., turn to a new direction of motion as indicated by blue arrows. Right: Emergent collective dynamics of self-steering microswimmers (pullers), from a simulation with 884,736 squirmers (with zoom-in – lower right panel), reminiscent to a fish school (upper right panel).

power law as a function of distance. In recent decades, efficient hydrodynamic simulation techniques have been developed for systems with characteristic mesoscopic length scales (tens of nanometres to hundreds of micrometers), such as the Lattice Boltzmann method (LBM), the Dissipative-Particle Dynamics (DPD), and the Multiparticle Collision Dynamics (MPC) approach[5]. In our simulations, we employ a newly developed code for MPC[6], which can run on an arbitrary number of GPUs in parallel. This facilitates simulations of very large wet systems of up to one million self-propelled particles.

By using the high-performance computing resources on JUWELS[7], we investigate the collective behaviour of intelligent active particles in a fluid environment. An illustration of the essential features of this study is depicted in Fig. 1, which indicates hydrodynamic self-steering for pullers and pushers, the self-organisation in large ensembles of pullers, and – as a real-world example – the structure formation in schools of fish.

2 Model

2.1 Self-Propulsion and Self-Steering

Microswimmers are modelled by the squirmer model, where non-zero surface slip velocity renders a squirmer self-propelled and self-steering. For a spherical body shape with the radius R_{sq} and the azimuthal and polar angles ϕ and θ, we consider the surface flow field[8]

$$u_\theta = \frac{3}{2} v_0 \sin\theta(1 + \beta\cos\theta) - \frac{1}{R_{\mathrm{sq}}^2}(\tilde{C}_{11}\cos\phi - C_{11}\sin\phi), \tag{1}$$

$$u_\phi = \frac{\cos\theta}{R_{\mathrm{sq}}^2}(C_{11}\cos\phi + \tilde{C}_{11}\sin\phi), \tag{2}$$

where the coefficients of the axisymmetric components v_0 and β (with $\beta < 0$ for pushers, $\beta > 0$ for pullers) denote the swim speed and the active stress, respectively. The coefficients of the non-axisymmetric components[9],

$$C_{11} = C_0 R_{\mathrm{sq}}^3 (\boldsymbol{e} \times \boldsymbol{e}_{\mathrm{aim}}) \cdot \boldsymbol{e}_x, \quad \tilde{C}_{11} = C_0 R_{\mathrm{sq}}^3 (\boldsymbol{e} \times \boldsymbol{e}_{\mathrm{aim}}) \cdot \boldsymbol{e}_y, \tag{3}$$

allow for self-steering[2,10], where reorientation towards the aiming direction $\boldsymbol{e}_{\mathrm{aim}}$ is given by

$$\dot{\boldsymbol{e}} = C_0 \, \boldsymbol{e} \times (\boldsymbol{e}_{\mathrm{aim}} \times \boldsymbol{e}). \tag{4}$$

Here, C_0 sets the maximum angular frequency of self-steering, corresponding to a limited maneuverability. The surface flow fields and the resultant self-steering are illustrated in Fig. 1 (left panel). Two dimensionless parameters, the Péclet number Pe and the maneuverability Ω, where

$$\mathrm{Pe} = \frac{v_0}{2 R_{\mathrm{sq}} D_R}, \quad \Omega = \frac{C_0}{D_R}, \tag{5}$$

are introduced to characterise the system. Here, D_R is the (passive) rotational diffusion coefficient of a spherical particle.

2.2 Alignment and Directional Sensing

Following our previous work on dry systems[11], we consider two types of sensing for self-steering, see Fig. 2(a) for illustration of their respective sensing ranges. First, our intelligent squirmers can autonomously align with the average self-propulsion direction of neighbouring squirmers (Fig. 2(c)), in the spirit of the (dry) Vicsek model[12,13], where alignment

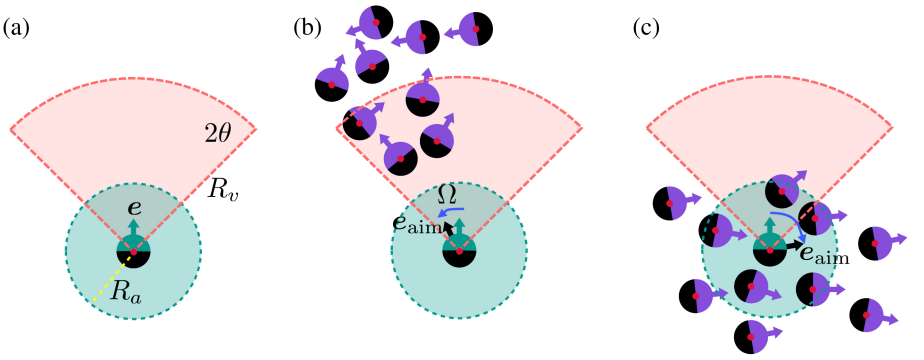


Figure 2. (a) Sensing ranges of a microswimmer (petrol/black circle) with propulsion direction $\boldsymbol{e}$ for alignment (green circular area with radius R_a, see Eq. 6) and visual perception (horizontally symmetric magenta cone with radius R_v and central angle θ, see Eq. 7). Illustration of conformations, where in (b) visual sensing and in (c) alignment dominates the resultant cognitive signal given in Eq. 8. The corresponding aiming vectors $\boldsymbol{e}_{\mathrm{aim}}$ and reorientation of the propulsion directions with maneuverability Ω are depicted by thick black and thin blue arrows, respectively.

between particles is modelled via the signal strength

$$\boldsymbol{s}_i^a = \frac{1}{N_{a,i}} \sum_{j \in PA} \boldsymbol{e}_j. \tag{6}$$

Here, PA indicates the polar-alignment sphere with radius R_a, and $N_{a,i}$ is the number of neighbours in PA of the i-th microswimmer. In order to facilitate swarm cohesion, it is also necessary for individuals to join larger groups. This is achieved by directional sensing – such as visual perception[14] –, inspired by the behavioural zonal model[15], mediated via a cognitive signal

$$\boldsymbol{s}_i^v = \frac{1}{N_{v,i}} \sum_{j \in VC} e^{-r_{ij}/R_0} \frac{\boldsymbol{r}_{ij}}{r_{ij}}, \tag{7}$$

where VC is a 'vision cone' with vision range $R_v \equiv 4R_0$ and vision angle θ around the self-propulsion direction $\boldsymbol{e}_i$, and $N_{v,i}$ the number of microswimmers in the vision cone of the i-th microswimmer (Fig. 2(b)). It is important to note that both interactions are *non-additive*, due to the normalisation by the number of particles in the interaction range[16].

Combining these two signals with the ratio ζ between alignment and directional maneuverability, we employ an overall cognitive signal strength

$$\boldsymbol{e}_{\mathrm{aim},i} = \boldsymbol{s}_i^a + \zeta \boldsymbol{s}_i^v, \tag{8}$$

which determines the surface slip velocity of self-steering squirmers. It is important to note that visual perception for vision angle $\theta < \pi$ is *non-reciprocal*, as one particle may be within the vision cone of another, but not *vice versa*. Also, non-additivity contributes to the non-reciprocality of the interactions.

2.3 Implementation and Parameters

The fluid dynamics is modelled by the multiparticle collision dynamics (MPC) approach[17], where the interactions and momentum exchange between fluid particles occur locally in collision cells, making the algorithm highly parallelisable[5]. Our plugin-based GPU/CPU

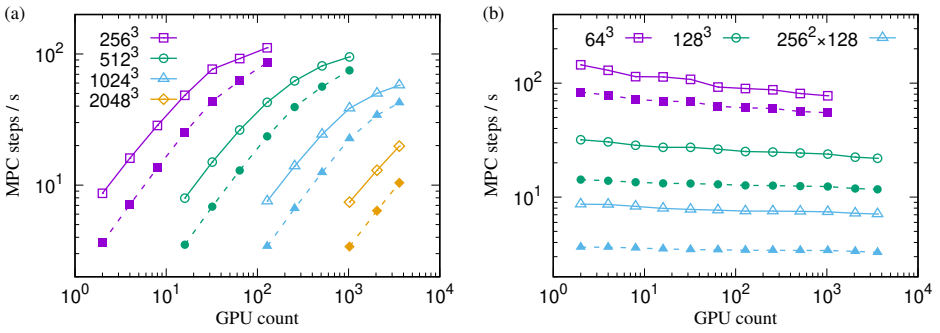


Figure 3. (a) Strong and (b) weak scaling of the MPC code for the indicated sizes $(L/a)^3$ of periodic cubic systems. The number of fluid particles per collision cell is $\langle N_c \rangle = 50$. Adapted from Westphal *et al.*[6].

code, which has been tested for up to 1.5 trillion MPC fluid particles[6], shows good strong and weak scaling behaviour with speedup up to a factor of ten, as shown in Fig. 3.

In simulations, we use the MPC fluid density (particles per collision cell) $\langle N_c \rangle = 20$, the collision time $h = 0.02a\sqrt{m/(k_BT)}$, with side length a of collision cells, and the rotation angle $\alpha = 130°$, which yields the fluid viscosity $\eta = 42.6\sqrt{mk_BT}/a^2$. For squirmers, we consider the radius $R_{\mathrm{sq}} = 3a$, the swimming speed $v_0 = 0.047\sqrt{k_BT/m}$ – resulting in the Péclet number $\mathrm{Pe} = 128$ –, and the sensing ranges $R_a = 4R_{\mathrm{sq}}$ and $R_0 = 2R_{\mathrm{sq}}$.

3 Swarming of Aligning Microswimmers

We first provide a brief overview of the emergent dynamics of aligning squirmers, i.e., $\zeta = 0$. We refer to Goh *et al.*[18] for more details.

Pushers. Aligning pushers feature active turbulence for a maneuverability $\Omega \geq 512$, best characterised by the scaling behaviour in the kinetic energy spectrum, which displays a power-law decay $E(k) \sim k^{-\alpha}$ as a function of wave number k. The exponents α extracted from our simulations show a non-universal behaviour in the range $2.8 \lesssim \alpha \lesssim 4.0$, and increase as the maneuverability increases. The typical size of vortices as well as the peak kinetic energy also increase with increasing maneuverability. A typical fluid profile with pronounced vortical structures is presented in Fig. 4(a). Squirmers are homogeneously distributed across the system, while their propulsion directions are predominantly aligned with the local flow direction of the ambient fluid. Examination of the mean-square displacement (Fig. 4(c)), together with the peak value of the kinetic energy spectrum, reveals that the collective advection of microswimmers is much faster than their intrinsic swim speed, which reflects strong hydrodynamic effects, and implies that the generated fluid flows are faster than the self-propulsion speed.

Pullers. A suspension of aligning pullers exhibits an enhanced clustering tendency (right panel in Fig. 1) due to hydrodynamic interactions, possibly with an additional peak in the local density distribution at higher density, $\rho_{\mathrm{loc}} \approx 0.5$. Still, the speed of such collective advection is significantly faster than the self-propulsion speed. Remarkably, strong alignment occurs for $\Omega \geq 2048$ and consequently polar ordering within a cluster generates a fluid jet, as the pullers collectively pull the fluid in front, which in turn gives rise to a vortex-ring structure in the fluid as shown in Fig. 4(b). We also note that the swimming direction of pullers ($\boldsymbol{v}/|\boldsymbol{v}|$) does not necessarily coincide with that of the orientation ($\boldsymbol{e}$), as demonstrated by negative values of the inner product between them (Fig. 4(b)-IV). The corresponding dynamics is again chaotic, exhibiting a power-law decay in the energy spectrum with a universal exponent of $\alpha = 11/3$. However, the distribution of the squirmer velocity deviates from a Gaussian, which is typically observed in active turbulence[19], with fat tails at higher velocities, see Fig. 4(d). This indicates that the collective dynamics of self-steering aligning pullers is a new type of self-organisation.

4 Directional Sensing

The formation of motile swarms, like bird flocks, fish schools, and animal herds, where the whole ensemble displays some coherent motion (in contrast to some insects swarms, which

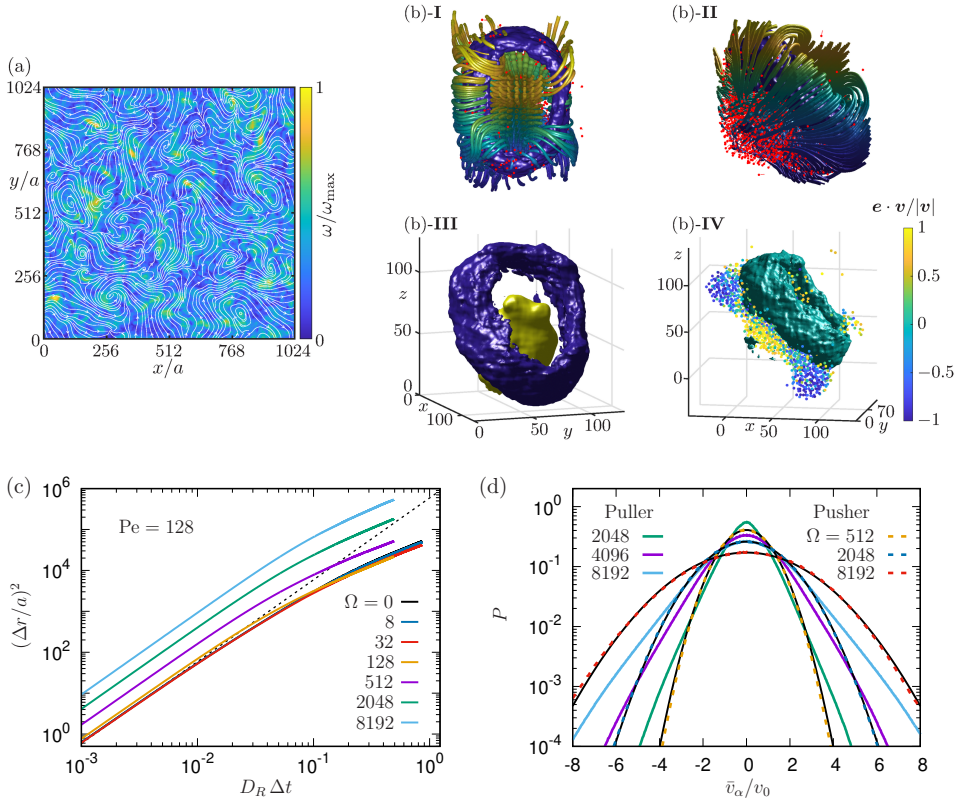


Figure 4. (a) Fluid velocity field (white streamlines) and the magnitude of the vorticity (heat map) for active turbulence of pushers. (b) Fluid streamlines (thick lines), fluid jet (yellow surface in III), and vortex-ring (torus) emerging in a system of aligning pullers (bullets). (c) Mean-square displacement $(\Delta r)^2 \equiv \langle (\boldsymbol{r}_i(t+\Delta t) - \boldsymbol{r}_i(t))^2 \rangle$ of pushers for various Ω. The black-dashed line represents the ballistic dynamics of $(\Delta r)^2 = v_0^2 (\Delta t)^2$. (d) Distribution of the Cartesian squirmer velocity component $\bar{v}_a$, averaged over the three coordinate directions. Black solid lines indicate Gaussian distributions. Adapted from Goh *et al.*[18].

can be quite stationary), requires the simultaneous presence of alignment, directional sensing, and "joining-the-group" behaviour. Thus, we study the model described in Sec. 2.2, with vision-alignment ratio $\zeta > 0$ in Eq. 8. Figs. 5 and 6 show snapshots of the dynamics for two different volume fractions. For the lower volume fraction, Fig. 5 also presents directional auto-correlation functions, which provide information about the persistence of motion of a swarm. For the higher volume fraction, Fig. 6 displays local density distributions of squirmers extracted from a Voronoi construction and the kinetic energy spectra in Fourier space for various vision-alignment ratios ζ.

Pushers. Directional sensing strongly affects conformations of microswimmers at low squirmer volume fractions, as it gives rise to elongated, worm-like swarms, because microswimmers naturally follow other microswimmers in front. Formation of such elongated clusters is well captured in the simulations at low squirmer density, as shown in

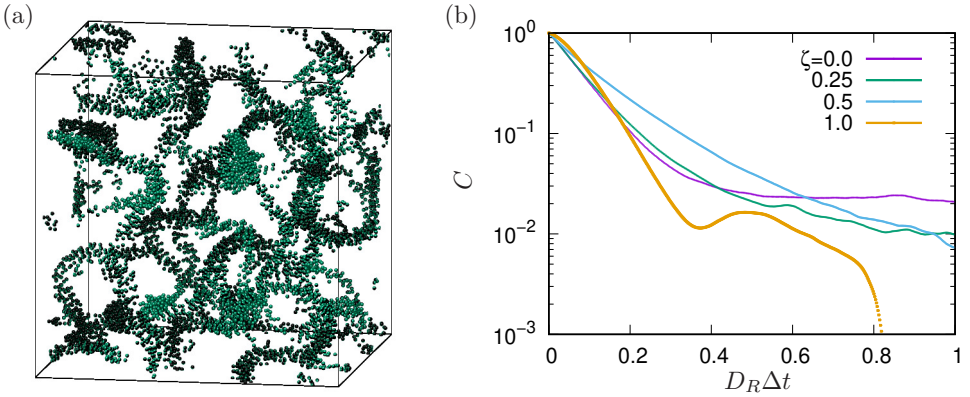


Figure 5. Self-organisation of an ensemble of pushers at low squirmer volume fraction 0.0067, with wide vision angle $\theta = \pi$. (a) Simulation snapshot for $\zeta = 1.0$. (b) Directional autocorrelation function $C(\Delta t) \equiv \langle \boldsymbol{e}_i(t) \cdot \boldsymbol{e}_i(t + \Delta t) \rangle$ for various values of ζ as indicated. The other simulation parameters are the Péclet number $\mathrm{Pe} = 128$, the maneuverability $\Omega = 8192$, and the system size $L/a = 512$.

Fig. 5(a). The directional auto-correlation function $C(\Delta t)$, displayed in Fig. 5(b), decays quite rapidly, with a typical time scale of only about $0.2/D_R$. This is partially due to the low, but not ultra-low, volume fraction, where local swarms strongly interact and often collide, see Fig. 5(a). Also, clusters are significantly bent and twisted due to hydrodynamic interactions between pushers, in contrast to dry systems[11]. This implies that $C(\Delta t)$ decays faster at $\zeta = 1$ than for weaker directional sensing, where alignment dominates the dynamics and the formation of worm-like swarms is not noticeable. We also note that transient torus-like structures with rotational motion emerge, depending on the parameters. Such circulating behaviour is often observed in fish schools, compare Fig. 1, but also dry systems, such as reindeer herds and fire ant groups.

For higher squirmer volume fractions, pushers again feature active turbulence, as for $\zeta = 0$. However, as the strength of directional sensing increases, squirmers tend to form aggregates (see snapshot in Fig. 6(a)), particularly for $\zeta \geq 1.0$, as demonstrated by the upward shift in the local density distribution at high densities (Fig. 6(c)). In the corresponding energy spectra, Fig. 6(e), the peak values of the kinetic energy $|E(k)|$ are reduced as ζ increases ($1 < \zeta \lesssim 2$), but the scaling exponents are not altered. This implies that the impact of directional sensing on the dynamics is mainly a slow-down of advection. For the largest investigated value $\zeta = 4.0$, even a new density peak appears at $\rho_{\mathrm{loc}} \approx 0.5$, which indicates a pronounced clustering tendency. In this case, the corresponding scaling exponent deviates significantly from those for smaller ζ. Moreover, the maximum energy is substantially reduced. Still, the scaling regime in the energy spectrum is broad, suggesting that the dynamics is chaotic. However, more careful simulations with higher MPC fluid density seem necessary to rule out potential artifacts due to fluid compressibility[20], which may be the origin of the third peak at $\rho_{\mathrm{loc}} = 0.58$ in Fig. 6(c).

Pullers. Pullers exhibit a strong clustering tendency for $\Omega \geq 128$, regardless of ζ. Surprisingly, however, the effect of additional directional sensing turns out to be non-monotonic. For large ratios of $\zeta \geq 2$, the height of the second peak in the local density distribution

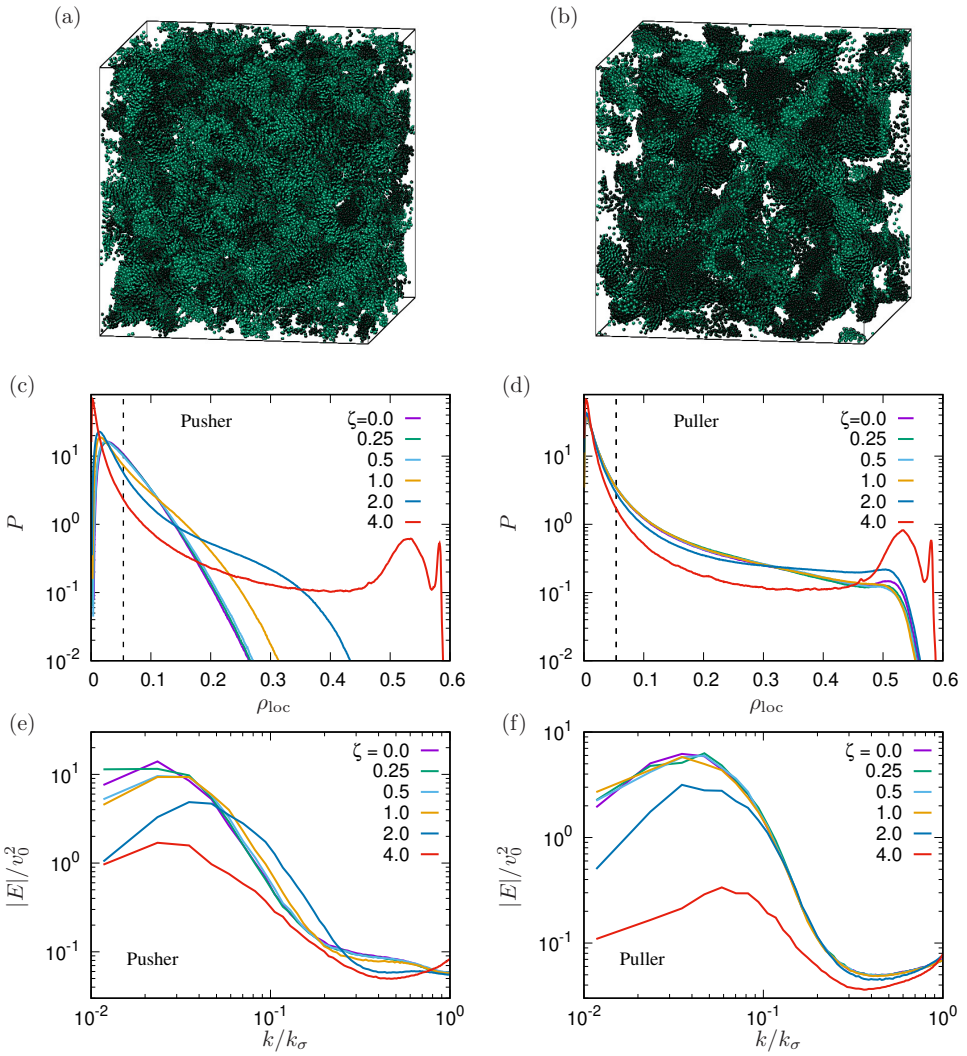


Figure 6. Self-organised structures in systems of pushers (a,c,e) and pullers (b,d,f) with alignment and visual perception for various vision-alignment ratios ζ. (a,b) Snapshots of squirmers for $\zeta = 2$ and the squirmer volume fraction 0.054, (c,d) local density distribution, and (e,f) fluid energy spectrum. The vision angle is $\theta = \pi/2$ for pushers and $\pi/12$ for pullers. In both cases, $\mathrm{Pe} = 128$, $\Omega = 2048$, and $L/a = 512$.

is higher than that at $\zeta = 0$, as shown in Fig. 6(d), which agrees with the expectation that directional sensing will result in a cohesive behaviour between microswimmers. In contrast, the height of the second peak for $\zeta \leq 1.0$ is slightly lower than that for $\zeta = 0$, indicating that a rearrangement of pullers due to directional sensing weakens hydrodynamic attraction, which may suppress aggregate formation, though weakly. In the regime $\zeta \lesssim 1$, the energy spectrum shows rather universal behaviour, as shown in Fig. 6(f). When a pronounced second peak develops in the local density distribution for $\zeta \geq 2$, a pronounced

shift in the peak as well as a significant shrinkage of the scaling regime in the energy spectrum are also observed. This indicates that the resultant dynamics may no longer be chaotic for $\zeta = 4$, but instead rather stable clusters are forming. Note that fluid-compressibility effects are usually more pronounced for pullers than for pushers.

The weak dependence of the puller self-organisation on ζ at higher volume fractions can be understood from the increased crowding in the vision cone. When a microswimmer senses a nearly homogeneous density distribution in the vision cone, then visual information no longer provides a clue for selecting the direction of self-propulsion.

5 Concluding Remarks

We have demonstrated that alignment and visual sensing may lead to fascinating collective behaviours of intelligent active particles in a hydrodynamic environment, confirming the rich interplay between cognitive sensing, self-propulsion, self-steering, and hydrodynamic interactions. The observed self-organisation and dynamical behaviour includes the emergence of clustering, formation of jets, swirls, and vortices, as well as generation of fast fluid flows. The possibility of performing large-scale simulations with massively parallel, GPU-based implementations on supercomputers like JUWELS are essential for unravelling self-organisation across a multitude of length scales. It is important to realise that the forms of sensing and cognitive self-steering are based on rather simple rules so far. This is necessary to gain an understanding of basic mechanisms. However, biological systems and potentially microrobotic systems can be much more complex. It will thus be very interesting to study more complex cognitive systems in the future.

Acknowledgements

The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (`www.gauss-centre.eu`) for funding this project by providing computing time on the GCS Supercomputer JUWELS[7] at Jülich Supercomputing Centre (JSC).

References

1. G. Gompper *et al.*, *The 2020 motile active matter roadmap*, J. Phys.: Condens. Matter **32**, 193001, 2020.
2. S. Goh, R. G. Winkler, and G. Gompper, *Noisy pursuit and pattern formation of self-steering active particles*, New J. Phys. **24**, 093039, 2022.
3. J. Elgeti, R. G. Winkler, and G. Gompper, *Physics of microswimmers–single particle motion and collective behavior: A review*, Rep. Prog. Phys. **78**, 056601, 2015.
4. M. R. Shaebani, A. Wysocki, R. G. Winkler, G. Gompper, H. Rieger, *Computational models for active matter*, Nat. Rev. Phys. **2**, 181, 2020.
5. E. Westphal, S. P. Singh, C. C. Huang, G. Gompper, and R. G. Winkler, *Multiparticle collision dynamics: GPU accelerated particle-based mesoscale hydrodynamic simulations*, Comput. Phys. Commun. **185**, 495-503, 2014.
6. E. Westphal, S. Goh, R. G. Winkler, and G. Gompper, *HTMPC: A heavily templated C++ library for large scale particle-based mesoscale hydrodynamics simulations using multiparticle collision dynamics*, 2024, arXiv:2406.15236.

7. Jülich Supercomputing Centre, *JUWELS cluster and booster: Exascale pathfinder with modular supercomputing architecture at Jülich Supercomputing Centre*, J. Large-Scale Res. Fac. **7**, A138, 2021.
8. O. S. Pak, and E. Lauga, *Generalized squirming motion of a sphere*, J. Eng. Math. **88**, 1-28, 2014.
9. S. Goh, R. G. Winkler, and G. Gompper, *Hydrodynamic pursuit by cognitive self-steering microswimmers*, Commun. Phys. **6**, 310, 2023.
10. M. Gassner, S. Goh, G. Gompper, and R. G. Winkler, *Noisy pursuit by a self-steering active particle in confinement*, Europhys. Lett. **142**, 21002, 2023.
11. R. S. Negi, R. G. Winkler, and G. Gompper, *Collective behavior of self-steering active particles with velocity alignment and visual perception*, Phys. Rev. Res. **6**, 013118, 2024.
12. T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet, *Novel type of phase transition in a system of self-driven particles*, Phys. Rev. Lett. **75**, 1226, 1995.
13. H. Chaté, *Dry aligning dilute active matter*, Annu. Rev. Condens. Matter Phys. **11**, 189, 2020.
14. L. Barberis and F. Peruani, *Large-scale patterns in a minimal cognitive flocking model: incidental leaders, nematic patterns, and aggregates*, Phys. Rev. Lett. **117**, 248001, 2016.
15. I. D. Couzin, J. Krause, N. R. Franks, and S. A. Levin, *Effective leadership and decision-making in animal groups on the move*, Nature **433**, 513, 2005.
16. O. Chepizhko, D. Saintillan, and F. Peruani, *Revisiting the emergence of order in active matter*, Soft Matter **17**, 3113-3120, 2021.
17. G. Gompper, T. Ihle, D. M. Kroll, and R. G. Winkler, *Multi-particle collision dynamics: A particle-based mesoscale simulation approach to the hydrodynamics of complex fluids*, Adv. Polym. Sci. **221**, 1-87, 2009.
18. S. Goh, E. Westphal, R. G. Winkler, and G. Gompper, *Alignment-induced self-organization of autonomously steering microswimmers: Turbulence, vortices, and jets*, 2024, arXiv:2406.17704.
19. K. Qi, G. Gompper, and R. G. Winkler, *Emergence of active turbulence in microswimmer suspensions due to active hydrodynamic stress and volume exclusion*, Commun. Phys. **5**, 49, 2022.
20. M. Theers, E. Westphal, K. Qi, R. G. Winkler, and G. Gompper, *Clustering of microswimmers: interplay of shape and hydrodynamics*, Soft Matter **14**, 8590, 2018.

Neural Wave Function for Superfluids

Wan Tong Lou[1], Halvard Sutterud[1], Gino Cassella[1], W. Matthew C. Foulkes[1], Johannes Knolle[1,2,3], David Pfau[1,4], and James S. Spencer[4]

[1] Department of Physics, Imperial College London,
South Kensington Campus, London SW7 2AZ, United Kingdom
E-mail: {w.lou21, h.sutterud21, g.cassella20, wmc.foulkes}@imperial.ac.uk

[2] Department of Physics TQM, Technische Universität München,
James-Franck-Straße 1, 85748 Garching, Germany
E-mail: j.knolle@tum.de

[3] Munich Center for Quantum Science and Technology (MCQST), 80799 Munich, Germany

[4] Google DeepMind, 6 Pancras Square, London N1C 4AG, United Kingdom
E-mail: {pfau, jamessspencer}@google.com

Understanding exotic quantum phases of matter remains a major goal of condensed matter physics. Here we address this challenge by simulating superfluids using the recently developed Fermionic neural network (FermiNet) approach [a] and the variational Monte Carlo algorithm. We study a paradigmatic strongly-correlated quantum system, the unitary Fermi gas, which has been known to possess a superfluid ground state but is difficult to describe quantitatively. We find limitations of the original FermiNet Ansatz in studying superfluidity and propose an improved Ansatz based on the idea of an antisymmetric geminal power singlet (AGPs) wave function. The results obtained using our new Ansatz are consistent with experiment and more accurate than previous benchmarks obtained using state-of-the-art fixed-node diffusion Monte Carlo simulations. We prove mathematically that the new Ansatz is a strict generalisation of the original FermiNet architecture, despite the use of fewer parameters. Our approach shares several advantages with the original FermiNet: the use of a neural network removes the need for an underlying basis set; and the flexibility of the network yields extremely accurate results within a variational quantum Monte Carlo framework that provides access to unbiased estimates of arbitrary ground-state expectation values. We discuss how the method can be extended to study other superfluids.

1 Introduction

Solving the many-body Schrödinger equation analytically is intractably difficult for systems of more than a few particles, although mean-field-like (Hartree-Fock and density functional) approaches, which treat particles as independent entities by averaging over the interactions between them, can often provide sufficient physical insights and qualitative results. However, in strongly correlated quantum systems, where particle interactions dominate, mean-field descriptions are often insufficient and more sophisticated numerical methods are needed to obtain qualitatively correct and quantitatively accurate results. Quantum Monte Carlo (QMC) methods[1], which use Monte Carlo integration to determine the properties of quantum many-body systems, are among the leading tools for studying strongly correlated quantum systems beyond the mean-field level. Despite the success of QMC methods, the accuracy of the results is often limited by the quality of the trial wave function, which is used to approximate the ground state wave function of the system.

[a] D. Pfau et al., doi:10.1103/PhysRevResearch.2.033429, Phys. Rev. Res. 2, 033429 (2020).

Recent years have seen the introduction of a new class of variational wave functions, known as neural wave functions or neural network quantum states, which utilise neural networks to approximate the ground state and sometimes also low-lying excited states[2]. This novel approach has been applied to a wide range of systems in condensed matter physics, such as spin and lattice models[2,3], molecules[4–7], and systems with quantum phase transitions[8,9], often achieving state-of-the-art results and outperforming other methods.

In this proceeding, which is based on Lou *et al.*[10], we apply the neural wave function approach to study superfluidity, one of the most famous macroscopic quantum phenomena of many-body systems. Specifically, our work focuses on the unitary Fermi gas (UFG), a paradigmatic example of a strongly correlated quantum system, which is known to possess a superfluid ground state, and is difficult to describe quantitatively. We use the FermiNet – a neural network architecture specifically designed to represent many-fermion wave functions[5] – to study the properties of the UFG. We demonstrate key limitations of the FermiNet Ansatz in studying the UFG and propose an improved Ansatz based on the idea of an antisymmetric geminal power singlet (AGPs) wave function[11,12].

2 The Unitary Fermi Gas

The unitary Fermi gas (UFG) is a strongly interacting system of two-component fermions that exhibits superfluidity in the crossover region between a Bardeen-Cooper-Schrieffer (BCS) superconductor and a Bose-Einstein condensate[13]. The effective range of the interaction is zero and the s-wave scattering length diverges (the "unitarity limit"), so the UFG has no intrinsic length scale. The only remaining length is the inverse of the Fermi wavevector $1/k_F$, on which all thermodynamic quantities depend. For example, for a given particle density, the ground-state energy per particle of an UFG can be written as

$$E = \xi E_{FG} = \xi \frac{3}{5} \frac{\hbar^2 k_F^2}{2m}, \tag{1}$$

where E_{FG} is the energy per particle of a non-interacting Fermi gas of the same density. The dimensionless constant ξ is known as the Bertsch parameter[14].

Because of the universality of the UFG model, it can be used to describe many real physical systems at different scales, such as the neutron matter in the inner crust of a neutron star[15] or the quantum criticality of an s-wave atomic superfluid[16]. The size of the pairs in the UFG is comparable to the inter-particle spacing, which is also a feature of many high-T_c superconductors[17]. As a result, the UFG has been studied extensively. Although the UFG is an idealised model, it can be accurately realised in the laboratory using ultracold atomic gases in which the interactions have been tuned by using an external magnetic field to drive the system across a Feshbach resonance.

The UFG has been studied for decades, but it remains a challenge to calculate its ground-state properties accurately. Mean-field treatments such as BCS theory give good results for systems with weak interactions, but fail in the strongly interacting regime. As a result, various quantum Monte Carlo (QMC) methods[1,18] have been used to simulate the properties of the UFG to high accuracy at zero and finite temperature. Methods used include variational Monte Carlo (VMC), fixed-node diffusion Monte Carlo (FN-DMC), fixed-node Green's function Monte Carlo, auxiliary field Monte Carlo, and diagrammatic

Monte Carlo[19–22]. However, a full quantitative description remains an open and challenging problem. In our work, we combine the FermiNet Ansatz with the simplest QMC scheme, the VMC method, to study the UFG.

3 Variational Monte Carlo

Despite being the conceptually simplest QMC method, VMC is a powerful tool for studying quantum many-body systems. It is based on a well-known variational principle, which states that the expectation value of the energy of a quantum system for a given trial wave function $\Psi_T(\mathbf{r}_1, \alpha_1, \mathbf{r}_2, \alpha_2, \ldots, \mathbf{r}_N, \alpha_N)$, where $\mathbf{r}_n$ and $\alpha_n \in \{\uparrow, \downarrow\}$ are the position and spin projection of particle n, is always greater than or equal to the ground-state energy E_0:

$$\left\langle \hat{H} \right\rangle = \frac{\langle \Psi_T | \hat{H} | \Psi_T \rangle}{\langle \Psi_T | \Psi_T \rangle} \geq E_0. \tag{2}$$

This provides the theoretical foundation for VMC, in which the adjustable parameters on which Ψ_T depends are chosen to minimise the energy expectation value. The high-dimensional integral that defines the expectation value cannot be evaluated analytically but can be estimated using Monte Carlo integration,

$$\left\langle \hat{H} \right\rangle = \frac{\int d\mathbf{R} \Psi_T^*(\mathbf{R}) \hat{H} \Psi_T(\mathbf{R})}{\int d\mathbf{R} |\Psi_T(\mathbf{R})|^2} = \frac{\int d\mathbf{R} |\Psi_T(\mathbf{R})|^2 \frac{\hat{H}\Psi_T(\mathbf{R})}{\Psi_T(\mathbf{R})}}{\int d\mathbf{R} |\Psi_T(\mathbf{R})|^2} \approx \frac{1}{M} \sum_{i=1}^{M} \frac{\hat{H}\Psi_T(\mathbf{R}_i)}{\Psi_T(\mathbf{R}_i)}, \tag{3}$$

where $\mathbf{R}$ is shorthand for $(\mathbf{r}_1, \alpha_1, \mathbf{r}_2, \alpha_2, \ldots, \mathbf{r}_N, \alpha_N)$ and the points $\mathbf{R}_i$ are sampled from the probability density $|\Psi_T(\mathbf{R})|^2 / \int d\mathbf{R} \, |\Psi_T(\mathbf{R})|^2$.

4 Neural Wave Functions

The VMC method is a powerful tool for studying quantum many-body systems, but the accuracy of the results is limited by the quality of the trial wave function. In conventional VMC, this is usually constructed using Slater determinants of Hartree-Fock orbitals, multiplied by a Jastrow factor to account for the electron-electron correlation and cusp conditions. Because of the limitations of the trial wave functions used, VMC alone is often unable to provide results accurate enough to throw much light on interesting chemical and materials physics problems.

Recently, a new class of trial wave functions, known as neural wave functions or neural network quantum states, has been introduced, utilising neural networks to represent the trial wave function[2]. The neural network takes the coordinates of the particles $(\mathbf{r}_i, \alpha_i)$ as input and outputs a set of latent space vectors $\mathbf{h}_i^{L\alpha} \equiv \mathbf{h}^{L\alpha}(\mathbf{r}_i^\alpha; \{\mathbf{r}_{/i}^\alpha\}; \{\mathbf{r}^{\bar{\alpha}}\}) \in \mathbb{R}^{n_L}$, with n_L being the size of the final layer L of the network. Here $i \in \{1, 2, \ldots, N^\alpha\}$, $\alpha \in \{\uparrow, \downarrow\}$, and $\bar{\alpha}$ the is spin projection opposite to α. The latent space vectors are then used to compute many-particle orbitals

$$\phi_i^\alpha(\mathbf{r}_j^\alpha; \{\mathbf{r}_{/j}^\alpha\}; \{\mathbf{r}^{\bar{\alpha}}\}) = \left[\mathbf{w}_i^\alpha \cdot \mathbf{h}_j^{L\alpha}\right] \chi_i^\alpha(\mathbf{r}_j^\alpha), \tag{4}$$

where $\mathbf{w}_i^\alpha$ is the weight vector of the i-th orbital, and $\chi_i^\alpha(\mathbf{r}_j^\alpha)$ is an envelope function to enforce the boundary conditions of the system. Note that the vectors $\mathbf{h}_j^{L\alpha}$ are invariant

under permutations of all particle coordinates except for the j-th particle, which means that the output orbitals have the same property. This is indicated by the use of (unordered) set notation for $\{\mathbf{r}^{\alpha}_{/j}\}$ and $\{\mathbf{r}^{\bar{\alpha}}\}$.

The full wave function is expressed as a determinant of the many-particle orbitals

$$\begin{aligned}\Psi_{\text{Slater FermiNet}}(\mathbf{R}) &= \det\left[\phi^{\alpha}_{i}(\mathbf{r}^{\alpha}_{j};\{\mathbf{r}^{\alpha}_{/j}\};\{\mathbf{r}^{\bar{\alpha}}\})\right] \\ &= \det\left[\phi^{\uparrow}_{i}(\mathbf{r}^{\uparrow}_{j};\{\mathbf{r}^{\uparrow}_{/j}\};\{\mathbf{r}^{\downarrow}\})\right]\det\left[\phi^{\downarrow}_{i}(\mathbf{r}^{\downarrow}_{j};\{\mathbf{r}^{\uparrow}\};\{\mathbf{r}^{\downarrow}_{/j}\})\right], \end{aligned} \tag{5}$$

where the second step followed because we assigned the spins and factorised the single determinant into two determinants, one for each spin. The parameters of the neural wave function (*i.e.*, the weights and biases of the neural network) are optimised by gradient descent to minimise the energy expectation value as in conventional VMC. Note that there is no limit on the number of many-particle orbitals that can be generated. Thus, multiple sets of orbitals are often used to construct a wave function with multiple determinants:

$$\Psi^{D}_{\text{Slater FermiNet}}(\mathbf{R}) = \sum_{k}^{D}\det\left[\phi^{k\uparrow}_{i}(\mathbf{r}^{\uparrow}_{j};\{\mathbf{r}^{\uparrow}_{/j}\};\{\mathbf{r}^{\downarrow}\})\right]\det\left[\phi^{k\downarrow}_{i}(\mathbf{r}^{\downarrow}_{j};\{\mathbf{r}^{\uparrow}\};\{\mathbf{r}^{\downarrow}_{/j}\})\right], \tag{6}$$

which usually improves the accuracy of the results. The normalisations of the FermiNet determinants are learned during the optimisation, so there is no need to include expansion coefficients.

To study the UFG, we employ the FermiNet neural network architecture[5], which implements the totally antisymmetric multi-determinantal Ansatz described above. FermiNet has achieved state-of-the-art results in various quantum many-body systems, including atoms, molecules[5,6] and solids[23], and has shown itself able to discover quantum phase transitions in the homogeneous electron gas[9]. However, as we will show in the Results section, it does not describe superfluids accurately. This leads us to introduce a modified Ansatz based on the idea of the antisymmetric geminal power singlet (AGPs) wave function. We show that the new Ansatz is a strict generalisation of the FermiNet architecture, despite the use of fewer parameters.

4.1 Antisymmetric Geminal Power Singlet Wave Function

A conventional antisymmetric geminal power singlet (AGPs) wave function is a fixed particle-number analogue of the Bardeen-Cooper-Schrieffer (BCS) wave function. It replaces the single-particle orbitals that appear in conventional Slater determinants with pairing orbitals (geminals), which are functions of the coordinates of two particles instead of one. This much improves the description of paired systems. To adapt the original FermiNet – referred to as the Slater FermiNet in this context – for superfluid systems, we propose a modification to its architecture, incorporating a generalisation of the AGPs form.

In the original Slater FermiNet, the many-particle orbitals are constructed by taking the dot product between each latent space vector and a set of weights, as shown in Eq. 4. To build an AGPs wave function with FermiNet, we first need to construct a set of geminals:

$$\varphi^{k}(\mathbf{r}^{\alpha}_{i},\mathbf{r}^{\bar{\alpha}}_{j};\{\mathbf{r}^{\alpha}_{/i}\};\{\mathbf{r}^{\bar{\alpha}}_{/j}\}) = \left[\mathbf{w}^{k}\cdot\left(\mathbf{h}^{L\alpha}_{i}\odot\mathbf{h}^{L\bar{\alpha}}_{j}\right)\right]\chi^{k\alpha}(\mathbf{r}^{\alpha}_{i})\chi^{k\bar{\alpha}}(\mathbf{r}^{\bar{\alpha}}_{j}), \tag{7}$$

where $\odot$ denotes an element-wise product. The AGPs FermiNet wave function is obtained by taking determinants of these many-particle geminals and summing the determinants in

a manner analogous to Eq. 6:

$$\Psi^{D}_{\text{AGPs FermiNet}}(\mathbf{R}) = \sum_{k}^{D} \det\Big[\varphi^{k}(\mathbf{r}^{\alpha}_{i}, \mathbf{r}^{\bar{\alpha}}_{j}; \{\mathbf{r}^{\alpha}_{/i}\}; \{\mathbf{r}^{\bar{\alpha}}_{/j}\})\Big]. \tag{8}$$

A schematic diagram of the difference between the Slater FermiNet and the AGPs FermiNet can be found in Fig. 1.

In a recent paper[10], we showed mathematically that the Slater FermiNet is a limiting case of the AGPs FermiNet, and that the latter is a strict generalisation of the former.

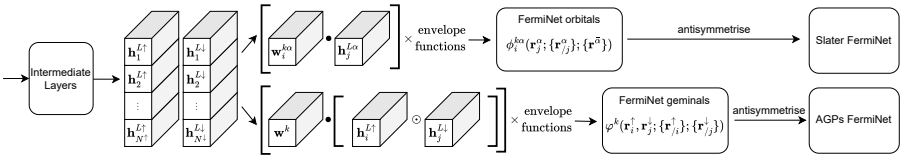


Figure 1. Schematic diagram of the difference between the Slater FermiNet and the AGPs FermiNet. The Slater FermiNet constructs many-particle orbitals by taking the dot product between each latent space vector and a set of weights. The AGPs FermiNet constructs many-particle geminals by taking the dot product between the element-wise product of two latent space vectors and a set of weights.

5 Results

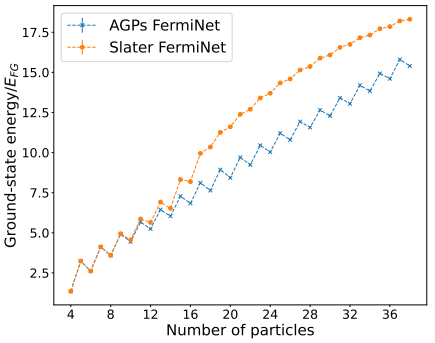
We show the power of the AGPs FermiNet Ansatz by studying the UFG. The Hamiltonian is given by

$$\hat{H} = -\frac{1}{2}\sum_{i}^{N}\nabla_{i}^{2} + \sum_{ij}^{N^{\uparrow}N^{\downarrow}} U(\mathbf{r}^{\uparrow}_{i} - \mathbf{r}^{\downarrow}_{j}), \quad \text{where} \quad U(\mathbf{r}) = -\frac{2v_{0}\mu^{2}}{\cosh^{2}(\mu r)} \tag{9}$$

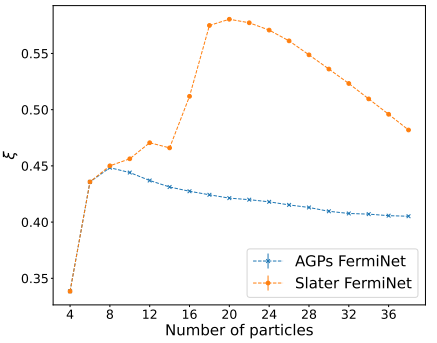
is the modified Pöschl-Teller potential, which is widely used in variational and diffusion QMC simulations[19–22] to model a delta function interaction. The s-wave scattering length of the Pöschl-Teller potential diverges when $v_0 = 1$. By changing the value of μ at fixed $v_0 = 1$, it is possible to vary the effective range of the interaction, $r_e = 2/\mu$, whilst holding the s-wave scattering length infinite.

A comparison of the ground-state energy expectation values given by the two Ansätze is shown in Fig. 2(a). The Slater FermiNet, which consists of a linear combination of block-diagonal determinants of FermiNet orbitals, performs well when the number of particles N is smaller than around 10, but the AGPs FermiNet is superior for larger systems. It is clear that the Slater FermiNet Ansatz has difficulties learning the ground states of large paired systems

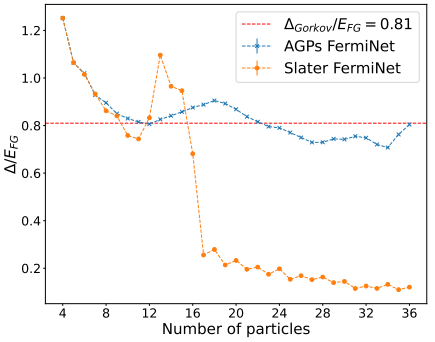
Another comparison between the two Ansätze is shown in Fig. 2(b), which depicts the ratio of the interacting and non-interacting energies per particle, known as the Bertsch parameter[14] and defined in Eq. 1, as a function of N. All FermiNet energies are variational, so the AGPs FermiNet, for which the Bertsch parameter is lower by up to around 30%, is much the better of the two Ansätze.



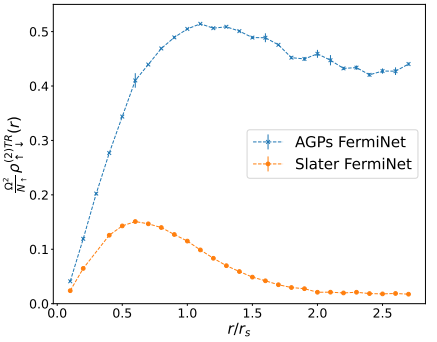
(a) The total energy of the UFG simulation cell, measured in units of the free Fermi gas energy E_{FG}. The Slater FermiNet Ansatz begins to fail when $N \gtrapprox 10$.



(b) The Bertsch parameter ξ (the ratio of the interacting and non-interacting ground-state energies per particle) against the number of particles N.



(c) Pairing gaps against the numbers of particles N.



(d) The TBDM estimators with $N = 38$ particles, plotted against the pair-separation.

Figure 2. Comparison between results obtained using the AGPs FermiNet and the Slater FermiNet for different numbers of particles, N, with $r_s = 1$ and $\mu = 12$, except for Fig. 2(d), where the number of particles is fixed to be $N = 38$. All simulations used 32 determinants, 300,000 optimisation steps, and the same hyperparameters. Most of the error bars are so small that they are obscured by the symbols. All figures adapted from Lou *et al.*[10]

To verify the accuracy of our AGPs FermiNet results, we compared them[10] with the state-of-the-art FN-DMC simulations of Forbes *et al.*[21] for the case $k_F = 1$ and $\mu = 12$. The AGPs FermiNet achieved a lower energy per particle than FN-DMC for all except tiny systems with $N = 4$ and $N = 6$ particles. The dependence of the Bertsch parameter on system size was also smoother when calculated with the AGPs FermiNet.

The pairing gap may be found using the approximation formula[18]

$$\Delta = (-1)^N \left[E(N+1) - \frac{1}{2}\left[E(N) + E(N+2)\right] \right], \tag{10}$$

where N is the total number of particles in the box. The results from $N = 4$ to $N = 36$ are shown in Fig. 2(c). The striking collapse of the pairing gap with increasing system

size shows that the Slater FermiNet Ansatz struggles to describe paired states in systems of more than 10 particles. The AGPs FermiNet Ansatz behaves much better, although we expect significant finite-size errors to remain even for the largest systems simulated. Also shown is the thermodynamic ($N \to \infty$) limit of the BCS pairing gap including Gorkov's polarisation correction: $\Delta_{\text{Gorkov}} = 0.815 E_{\text{FG}}$, where $E_{\text{FG}} = \frac{3}{5}\frac{\hbar^2 k_F^2}{2m}$ is the average energy per particle of an unpolarised non-interacting Fermi gas. The UFG is a strongly coupled system, so the Gorkov estimate of the gap need not be accurate.

Another signature of fermionic superfluidity is the presence of off-diagonal long-ranged order in the two-body density matrix (TBDM), $\rho^{(2)}_{\uparrow\downarrow}(\mathbf{r}_1, \mathbf{r}_2; \mathbf{r}'_1, \mathbf{r}'_2) = \langle \hat{\psi}^\dagger_\uparrow(\mathbf{r}_1)\hat{\psi}^\dagger_\downarrow(\mathbf{r}_2)\hat{\psi}_\downarrow(\mathbf{r}'_2)\hat{\psi}_\uparrow(\mathbf{r}'_1)\rangle$, the largest eigenvalue of which diverges as the number of particles N tends to infinity. The superfluid condensate fraction c may be obtained by evaluating[24]

$$c = \lim_{r\to\infty} \tilde{c}(r) = \lim_{r\to\infty} \frac{1}{4\pi r^2 N_\uparrow} \int \rho^{(2)}_{\uparrow\downarrow}(\mathbf{r}_1, \mathbf{r}_2; \mathbf{r}_1 + \mathbf{r}', \mathbf{r}_2 + \mathbf{r}')\delta(|\mathbf{r}'| - r) d\mathbf{r}_1 d\mathbf{r}_2 d\mathbf{r}', \quad (11)$$

where $N_\uparrow$ is the number of spin-up particles. As shown in Fig. 2(d), the TBDM of the 38-particle system calculated using the Slater FermiNet approaches zero in the large pair-separation limit, showing that the neural network wave function does not describe a superfluid. The same quantity for the AGPs FermiNet approaches a finite value, yielding a condensate fraction $c \approx 0.44(1)$. This is consistent with the value we obtained by simulating a larger 66-particle system, with experimental estimates, and with recent AFMC results[22]. The data are summarised in Tab. 1.

Method	Value
Our estimate for $N = 38$ at $k_F r_e = 0.32$	0.44(1)
Our estimate for $N = 66$ at $k_F r_e = 0.32$	0.42(1)
Our estimate for $N = 66$ at $k_F r_e = 0.17$	0.52(1)
FN-DMC for $N = 38$ at $k_F r_e = 0.03$[25]	0.61(2)
FN-DMC for $N = 66$ at $k_F r_e = 0.03$[25]	0.57(2)
FN-DMC for $N = 128$ with VMC extrapolation at $k_F r_e = 0.32$[20]	0.51
FN-DMC with $k_F r_e \to 0$ extrapolation for $N = 66$[26]	0.56(1)
AFMC with $k_F r_e \to 0$ extrapolation for $N = 66$[22]	0.43(2)
Experiment[27]	0.46(7)
Experiment[28]	0.47(7)

Table 1. Estimates of the superfluid condensate fraction at unitarity using various methods. The quantity $k_F r_e$ is a dimensionless number, indicating the deviation of the simulated system from a perfect UFG with zero-range interaction.

6 Discussion

We have used neural wave functions to study the superfluidity of the paradigmatic UFG. We showed that the Slater FermiNet Ansatz has difficulties in describing paired systems with

strong, short-ranged, attractive interactions between particles of opposite spin. This led us to improve the variational Ansatz by using determinants of FermiNet geminals, a drastic generalisation of a conventional AGPs or BCS wave function. We showed mathematically that the Slater FermiNet is a limiting case of the AGPs FermiNet despite the use of fewer parameters in the latter. It follows that any FermiNet wave function can in principle be written as an AGPs FermiNet wave function.

The inability of the Slater FermiNet Ansatz to accurately describe the UFG ground state came as something of a surprise because the original FermiNet paper[5] showed that any many-body fermionic wave function could be represented as a single determinant of FermiNet orbitals. However, the mathematical argument relies on the construction of FermiNet orbitals with unphysical discontinuities. Whether or not any wave function can be represented as a single determinant of FermiNet orbitals of the type used in practice, which are differentiable everywhere except at electron-electron and electron-nuclear coalescence points, remains an open question.

Another limitation is that the architecture of the FermiNet neural network, which is rather simple, may not be able to represent an arbitrary many-electron FermiNet orbital. Even if a single-determinant Slater FermiNet wave function is general in principle, there is no guarantee that it is equally easy to represent all wave functions. It may be that producing an accurate representation of a paired wave function requires the width and number of layers in the neural network to increase rapidly with system size. Equally, if a network of fixed size is used, it may be necessary to increase the number of Slater FermiNet determinants rapidly as the system size increases. The observation that the Slater FermiNet works well when $N \lessapprox 10$ but that the quality of the results degrades rapidly for larger systems, along with the scaling study described by Lou *et al.*[10], suggest that this is, in fact, the case. Whilst most work on neural wave functions has focused on improving the neural network part of the Ansatz, our work suggests that the method of antisymmetrisation is also crucial for the accuracy of the results. Our AGPs-inspired approach is not limited to the FermiNet neural network and can be readily adapted to use more recent architectures such as the Psiformer[29], GLOBE and MOON[30], and DeepErwin[7].

Finally, the AGPs FermiNet introduced here has a straightforward Pfaffian extension and can thus be applied to non-s-wave and triplet pairing. We expect it to become a powerful tool for understanding strongly correlated non-s-wave superfluid and superconducting systems such as Helium-3 or high-T_c and p-wave superconductors.

Acknowledgements

We thank Stefano Gandolfi and Michael M. Forbes for providing the diffusion Monte Carlo data. We gratefully acknowledge the European Union's PRACE program for the award of computing resources on the JUWELS Booster supercomputer in Jülich; the HPC RIVR consortium and EuroHPC JU for resources on the Vega high performance computing system at IZUM, the Institute of Information Science in Maribor; and the United Kingdom Engineering and Physical Sciences Research Council for resources on the Baskerville Tier 2 HPC service. Baskerville was funded by the EPSRC and UKRI through the World Class Labs scheme (EP/T022221/1) and the Digital Research Infrastructure program (EP/W032244/1) and is operated by Advanced Research Computing at the University of Birmingham. We also gratefully acknowledge the Gauss Centre for Super-

computing e.V. for funding this project by providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS Supercomputer JUWELS at Jülich Supercomputing Centre (JSC). J. K. is part of the Munich Quantum Valley, which is supported by the Bavarian state government with funds from the Hightech Agenda Bayern Plus. W. T. L. is supported by an Imperial College President's Ph.D. Scholarship; H. S. is supported by the Aker Scholarship; and G. C. is supported by the United Kingdom Engineering and Physical Sciences Research Council (EP/T51780X/1). We also acknowledge the support of the Imperial-TUM flagship partnership.

References

1. W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, *Quantum Monte Carlo simulations of solids*, Rev. Mod. Phys., **73**, 33-83, Jan. 2001.
2. G. Carleo and M. Troyer, *Solving the quantum many-body problem with artificial neural networks*, Science, **355**, no. 6325, 602-606, 2017.
3. D. Luo and B. K. Clark, *Backflow Transformations via Neural Networks for Quantum Many-Body Wave Functions*, Phys. Rev. Lett., **122**, 226401, June 2019.
4. J. Hermann, Z. Schätzle, and F. Noé, *Deep-neural-network solution of the electronic Schrödinger equation*, Nat. Chem., **12**, no. 10, 891-897, Oct. 2020.
5. D. Pfau, J. S. Spencer, A. G. D. G. Matthews, and W. M. C. Foulkes, *Ab initio solution of the many-electron Schrödinger equation with deep neural networks*, Phys. Rev. Res., **2**, 033429, Sep. 2020.
6. J. S. Spencer, D. Pfau, A. Botev, and W. M. C. Foulkes, *Better, faster fermionic neural networks*, 2020, arXiv:2011.07125.
7. M. Scherbela, R. Reisenhofer, L. Gerard, P. Marquetand, and P. Grohs, *Solving the electronic Schrödinger equation for multiple nuclear geometries with weight-sharing deep neural networks*, Nat. Comput. Sci., **2**, no. 5, 331-341, May 2022.
8. M. Wilson, S. Moroni, M. Holzmann, N. Gao, F. Wudarski, T. Vegge, and A. Bhowmik, *Neural network ansatz for periodic wave functions and the homogeneous electron gas*, Phys. Rev. B, **107**, 235139, June 2023.
9. G. Cassella, H. Sutterud, S. Azadi, N. D. Drummond, D. Pfau, J. S. Spencer, and W. M. C. Foulkes, *Discovering Quantum Phase Transitions with Fermionic Neural Networks*, Phys. Rev. Lett., **130**, 036401, Jan. 2023.
10. W. T. Lou, H. Sutterud, G. Cassella, W. M. C. Foulkes, J. Knolle, D. Pfau, and J. S. Spencer, *Neural wave functions for superfluids*, Physical Review X, **14**, no. 2, 021030, 2024.
11. M. Casula and S. Sorella, *Geminal wavefunctions with Jastrow correlation: a first application to atoms*, J. Chem. Phys., **119**, no. 13, 6500-6511, Oct. 2003.
12. M. Bajdich, *Generalized pairing wave functions and nodal properties for electronic structure quantum monte carlo*, 2007, arXiv:0712.3066.
13. A. J. Leggett, *Diatomic molecules and cooper pairs*, in: Modern Trends in the Theory of Condensed Matter, A. Pekalski and J. A. Przystawa (Eds.), Springer Berlin Heidelberg, 13-27, 1980.
14. T. Papenbrock and G. F. Bertsch, *Pairing in low-density Fermi gases*, Phys. Rev. C, **59**, 2052-2055, Apr. 1999.

15. S. Gandolfi, A. Gezerlis, and J. Carlson, *Neutron Matter from Low to High Density*, Annu. Rev. Nucl. Part. Sci., **65**, no. 1, 303-328, 2015.
16. P. Nikolić and S. Sachdev, *Renormalization-group fixed points, universal phase diagram, and 1/ N expansion for quantum liquids with interactions near the unitarity limit*, Phys. Rev. A, **75**, no. 3, 033608, 2007.
17. M. Randeria, J.-M. Duan, and L.-Y. Shieh, *Superconductivity in a two-dimensional Fermi gas: Evolution from Cooper pairing to Bose condensation*, Phys. Rev. B, **41**, 327-343, Jan. 1990.
18. J. Carlson, S. Gandolfi, and A. Gezerlis, *Superfluid pairing in neutrons and cold atoms*, in: Fifty Years of Nuclear BCS, WORLD SCIENTIFIC, 348-359, Mar. 2013.
19. J. Carlson, S.-Y. Chang, V. R. Pandharipande, and K. E. Schmidt, *Superfluid Fermi Gases with Large Scattering Length*, Phys. Rev. Lett., **91**, 050401, July 2003.
20. A. J. Morris, P. López Ríos, and R. J. Needs, *Ultracold atoms at unitarity within quantum Monte Carlo methods*, Phys. Rev. A, **81**, no. 3, 033619, Mar. 2010.
21. M. McNeil Forbes, S. Gandolfi, and A. Gezerlis, *Resonantly Interacting Fermions in a Box*, Phys. Rev. Lett., **106**, 235303, June 2011.
22. R. He, N. Li, B.-N. Lu, and D. Lee, *Superfluid condensate fraction and pairing wave function of the unitary Fermi gas*, Phys. Rev. A, **101**, 063615, June 2020.
23. X. Li, C. Fan, W. Ren, and J. Chen, *Fermionic neural network with effective core potential*, Phys. Rev. Res., **4**, 013021, Jan. 2022.
24. R. J. Needs, M. D. Towler, N. D. Drummond, P. López Ríos, and J. R. Trail, *Variational and diffusion quantum Monte Carlo calculations with the CASINO code*, J. Chem. Physics, **152**, no. 15, 154106, 2020.
25. G. E. Astrakharchik, J. Boronat, J. Casulleras, and S. Giorgini, *Momentum Distribution and Condensate Fraction of a Fermion Gas in the BCS-BEC Crossover*, Phys. Rev. Lett., **95**, 230405, Dec. 2005.
26. X. Li, J. Kolorenč, and L. Mitas, *Atomic Fermi gas in the unitary limit by quantum Monte Carlo methods: Effects of the interaction range*, Phys. Rev. A, **84**, 023615, Aug. 2011.
27. M. W. Zwierlein, C. H. Schunck, C. A. Stan, S. M. F. Raupach, and W. Ketterle, *Formation Dynamics of a Fermion Pair Condensate*, Phys. Rev. Lett., **94**, 180401, May 2005.
28. W. J. Kwon, G. Del Pace, R. Panza, M. Inguscio, W. Zwerger, M. Zaccanti, F. Scazza, and G. Roati, *Strongly correlated superfluid order parameters from dc Josephson supercurrents*, Science, **369**, no. 6499, 84-88, 2020.
29. I. von Glehn, J. S. Spencer, and D. Pfau, *A self-attention ansatz for ab-initio quantum chemistry*, 2023, arXiv:2211.13672.
30. N. Gao and S. Günnemann, *Generalizing Neural Wave Functions*, in: Proceedings of the 40th International Conference on Machine Learning, **vol. 202**, PMLR, 10708-10726, 23-29 July 2023.

Astrophysics

Astrophysics

Rolf Kuiper

Faculty of Physics, University of Duisburg-Essen, Lotharstraße 1, 47057 Duisburg, Germany
E-mail: rolf.kuiper@uni-due.de

Astronomy is a discovery-driven science, and our understanding of objects and processes in the universe grows with the advent of new observational techniques and instruments. State-of-the-art numerical simulations of the systems under study are required to understand the observational data, derive reliable implications, and decompose the overall results into their physical meaning. Here, results on the evolution of high-mass star-forming regions and the observational signatures of kilonovae are reported.

Introduction

In theoretical astrophysics, computational resources are typically required for numerical models of N-body dynamics, hydrodynamics, and radiative transfer. Each of these fundamental computations often involves multi-physics aspects such as magnetic fields, turbulence, phase transitions, chemical evolution, or nucleosynthesis. Post-processing of the simulation data often involves computationally intensive steps to derive synthetic observational data cubes for direct comparison with state-of-the-art observational surveys and individual studies.

In 2024, most of the computational time was devoted to the study of the formation and feedback of high-mass stars and the physics of kilonovae from neutron star-neutron star mergers to observed light curves. Both projects are based on multi-physics hydrodynamical simulations and the production of synthetic observations.

High-Mass Star Formation

The field of high-mass star formation research is currently being revolutionised by statistically powerful observational surveys such as the ALMAGAL initiative, which is collecting data on more than 1000 star-forming clumps (containing more than 6000 pre- to proto-stellar cores) at different evolutionary stages. To make the most of this fascinating data, numerical simulations of the evolution of high-mass star-forming regions are being performed. Even a single one of these simulations is computationally demanding due to the multi-physical and multi-spatial scales involved. Furthermore, the evolution of these regions needs to be studied as a function of a variety of environmental parameters, such as their mass distribution, turbulence level, overall dynamical state (angular momentum and/or converging flows), magnetic field strength, and metallicity. Meaningful sampling of this large parameter space seems challenging. For an example of such an attempt, please see the contribution of Birka Zimmermann and Stefanie Walch.

Kilonovae

Kilonovae are intense bursts of light that occur when neutron stars collide. The collision creates extremely neutron-rich conditions that trigger the rapid neutron capture (r-process).

This process plays a key role in the formation of many of the heaviest elements in the Universe. Kilonovae provide a valuable opportunity to explore the mechanisms behind r-process nucleosynthesis and to study the properties of matter at extreme densities, such as our still incomplete knowledge of the equation of state of neutron stars.

The new 2024 calculations by Christine E. Collins and her collaborators demonstrate the need for three-dimensional models and show the sensitivity of the results to accurate atomic data. Their modelling pipeline extends from hydrodynamical simulations of neutron star-neutron star mergers to nucleosynthesis calculations to radiative transfer to synthetic light curves. These light curves can then be directly compared with existing observational data of kilonovae bursts.

Outlook

In the field of theoretical astrophysics, the demand for computational resources has increased in the recent past. As a result, it is essential to develop software adapted to new hardware – and in some cases to develop it from scratch. This will undoubtedly be one of the most important challenges in the field for the next decade. If numerical methods can make the most of the available high-performance computing resources, we will continue to gain fascinating insights into the complex physics of the universe.

Feedback and Star Formation Efficiency in High-Mass Star-Forming Regions

Birka Zimmermann[1] **and Stefanie Walch**[1,2]

[1] I. Physikalisches Institut, Universität zu Köln, Zülpicher Str. 77, 50937 Köln, Germany
E-mail: {zimmermann, walch}@ph1.uni-koeln.de

[2] Center for data and simulation science, University of Cologne

The formation of stars, and especially high-mass stars, is a highly complex and dynamical process involving a large number of physical mechanisms. High-mass stars determine the evolution of galaxies due to their energetic feedback, such as (ionising) radiation, stellar winds, and supernovae. In order to better interpret real-world star-forming regions, simulations of collapsing clouds are used. We performed simulations of the gravitational collapse of isolated, parsec-scale, turbulent clouds to study the formation and evolution of massive stars as well as the impact of their highly energetic feedback. The initial conditions are physically motivated by real observations. A parameter study with different initial conditions is performed to obtain a statistical sample of simulations to compute synthetic telescope images which may be compared to observations made with modern telescopes like ALMA.

1 Introduction

Star formation is a highly active and rapidly developing topic in modern astrophysics. It is a fundamental process, shaping both the large and small astronomical scales and simultaneously influencing galactic dynamics[2] and planet formation. This impact is due to the star formation process being highly energetic, because of the intense feedback from newly born massive stars by protostellar jets, stellar winds, radiation and supernovae. In addition to the large dynamic ranges in spatial scale and density, these processes also make the star formation process difficult to simulate numerically.

One of the largest unresolved problems in modern star formation is that of the formation of massive stars, i.e. stars with a mass larger than eight times the mass of the Sun ($>8\,\mathrm{M}_\odot$). Lower mass stars can primarily be explained due to the interplay of gravity, turbulence and thermal pressure leading to quasi-Jeans mass fragmentation; however, additional processes are necessary for the formation of higher mass stars in terms of magnetic fields[3] and radiative feedback[4]. High mass stars are also thought to evolve faster and start nuclear burning before the mass accretion process is finished, thus feedback and accretion happen at the same time.

Due to the difficulty of studying high-mass star formation in both observations and in simulations, many open questions remain to date. In observations, young high-mass stars are difficult to detect primarily due to their rarity, which leads to them lying at considerably larger distances (on average) than lower mass stars, which makes it harder to fulfil the need for high-resolution observations. Additionally, high-mass stars typically form in the densest and most embedded environments, meaning that they are heavily obscured at most wavelengths and highly sensitive observations are needed, and even those are limited if the region is optically thick. Moreover, due to their high accretion rates, the formation process of high-mass stars is greatly accelerated compared to that of lower mass stars,

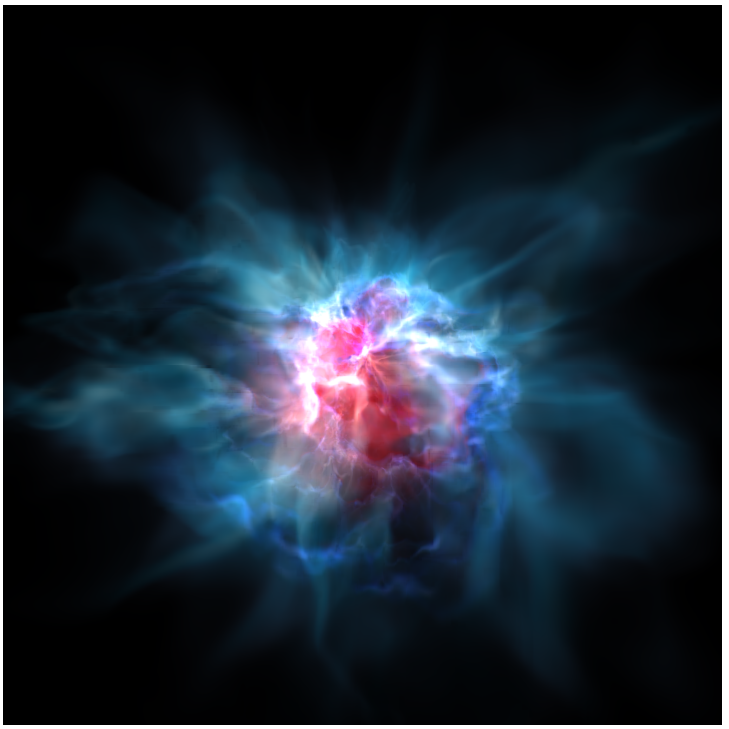


Figure 1. A look into the heart of a massive star-forming cloud. Shown is a simulation of the cold gas distribution of molecular hydrogen (white). Within the dense regions massive stars are formed which causes the ejection of atomic hydrogen (blue) and an expanding bubble of ionised hydrogen (red). The plotted volume has a side length of 2 pc[1].

lessening the time over which the process may be observed. In simulations, high-mass star formation is difficult to model due to the large number of coupled physical processes which are involved. Apart from magneto-hydrodynamics they include, but are not limited to, large magnetic field strengths, radiative heating and radiation pressure (RP), thermal pressure, ionising radiation, and stellar winds. It is difficult to construct models which are numerically stable when coupling these processes, and they are computationally expensive to run.

We simulate the collapse of isolated cores with different initial conditions in order to study massive star formation in a statistically relevant sample. We confront the numerical simulations and synthetic observations which eventually may be compared directly to the real-life observations.

This paper is structured as follows. First, we explain the numerical methods and initial conditions of the simulations in Sec. 2 and 3, respectively. In Sec. 4 we introduce a fiducial run, and show the importance of ionising radiation and RP as well as the impact of the numerical resolution on the cloud evolution. Moreover, we investigate the results of the parameter study and outline the importance of comparing simulations and synthetic observations. We conclude our results in Sec. 5.

2 Methods

We perform simulations of the gravitational collapse of isolated, parsec-scale, turbulent cores with the MPI-parallel, adaptive mesh refinement (AMR) code FLASH 4.7[5].

We include an entropy stable MHD solver[6] and compute self-gravity with an OctTree solver[7], which can also calculate the local shielding of the gas. We also model the radiative transfer of ionising radiation with our new radiative transfer scheme TreeRay[8], which can treat ionising and non-ionising (infrared) radiation and RP on dust[9]. We employ a local non-equilibrium chemistry network, which tracks the evolution of 7 species (H, H_2, H^+, C^+, CO, O, and free electrons)[10,11], and which is combined with the radiative transfer module[12,13]. Dust and gas temperatures are calculated separately and heating and cooling by dust is included. Stars are modelled with the use of sink particles, which are evolved with a $4^{\rm th}$ order Hermite integrator[14]. We model the evolution of individual stars with a protostellar model[15].

We use the JUWELS cluster module with the Intel Xenon Platinum 8168 CPU. Each simulation requires a computational time of around 1–2 Mio. core hours, and uses on average 500 cores (and up to 1200 cores) simultaneously. We produce data files to analyse and visualise the time evolution of the simulated core collapses. The simulations produce around 800 files each, necessitating a disk space of 80 TB.

3 Simulation Details

The cores are set up such that they are guaranteed to form massive stars under the evolution of self-gravity. The stars will further grow through the accretion of mass until they prevent accretion onto themselves through their own stellar feedback and reach a final mass.

The initial conditions of the simulated cores are formed by the parameter space covered by the ALMA large-scale program ALMAGAL. The core radius is kept at 1 pc while the side length of the cubic box is 4 pc with a diode boundary condition. The initial core gas temperature is 20 K. The initial dust temperature of 2.7 K is immediately adjusted to the thermal equilibrium value in the first time step.

The free-fall time, $t_{\rm ff}$, is the characteristic time a core with a uniform density $\bar{\rho}$ would take to collapse purely under its own gravity, and can be calculated by:

$$t_{\rm ff} = \sqrt{\frac{3\pi}{32 G \bar{\rho}}}\,. \tag{1}$$

The free-fall time in our simulations is 526,000 yr. The parameters we vary are the density profile, the virial parameter, and the metallicity. We are using a Plummer-like density profile, which is given by:

$$\rho(r) = \frac{\rho_0}{1 + (\frac{r}{r_0})^w}\,, \tag{2}$$

where r is the core radius, w the density exponent, ρ_0 the central density and r_0 the scale radius (~ 0.15 pc). We use three different density exponents, $w = 2$, $w = 1.5$, and $w = 0$, where $w = 0$ results in a constant density profile $\rho = \rho_0$. The central density changes with the different density exponents to keep the core mass at $1000\,{\rm M}_\odot$ for each simulation. The simulations are called FIDUCIAL, SHALLOW and FLAT, respectively.

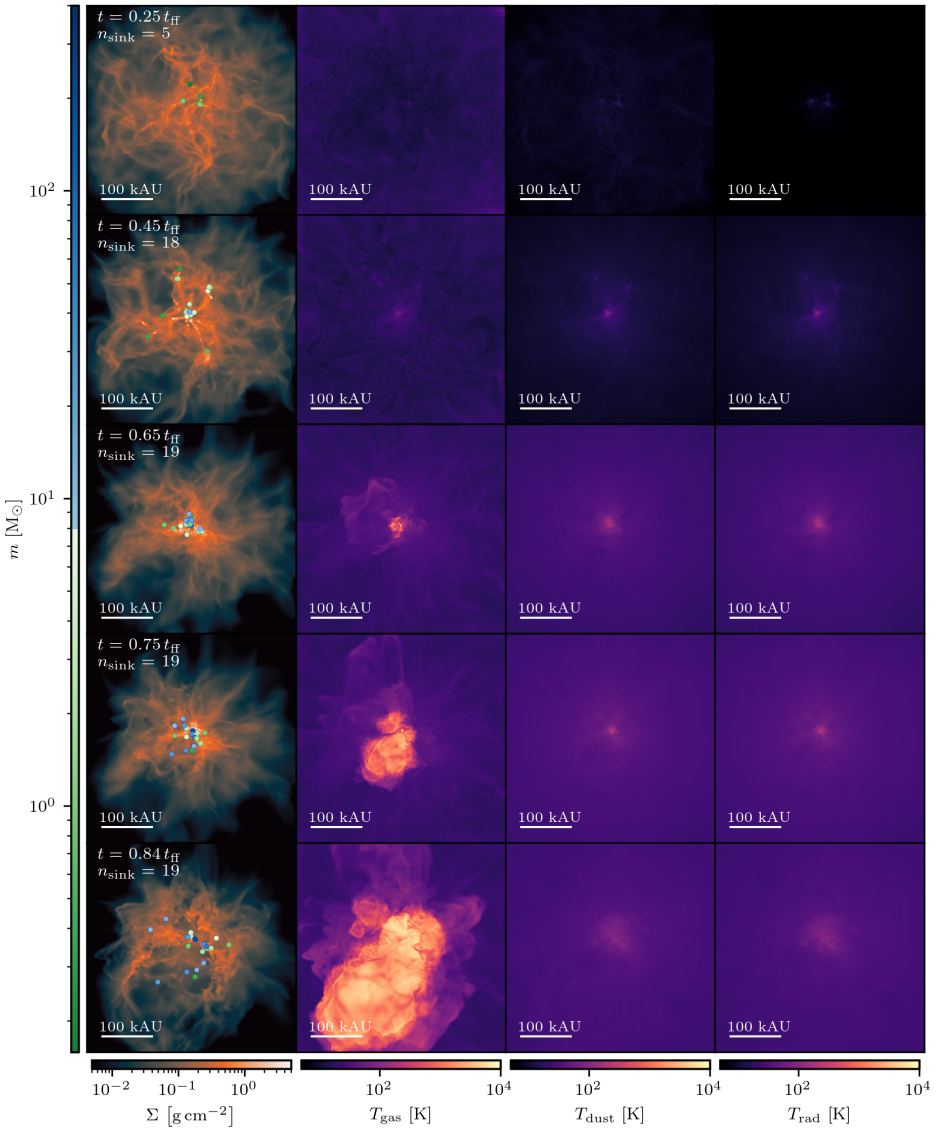


Figure 2. Time evolution of the fiducial run (from top to bottom). From left to right we show the projection in the z-direction of the column density, Σ, and the mass-weighted temperatures of gas, $T_{\rm gas}$, dust, $T_{\rm dust}$, and radiation, $T_{\rm rad}$. Small circles indicate sink particles, which represent stars. A green colour scheme represents lower-mass stars, while a blue colour scheme shows more massive sinks ($>8\,{\rm M}_\odot$). After $\sim 0.45\,t_{\rm ff}$ (where $t_{\rm ff}$ corresponds to 0.526 Myr) massive sink particles are formed which drive an expanding bubble of ionised hydrogen.

In order to study the impact of turbulent fields we change the virial parameter, $\alpha_{\rm vir}$, which is determined by the ratio of kinetic and gravitational energy:

$$\alpha_{vir} = \frac{5r\sigma^2}{GM}, \tag{3}$$

where σ is the velocity dispersion and M the total core mass. The FIDUCIAL run has a virial parameter of 0.6. We change the virial parameter from sub-virial with a very low velocity dispersion (Run αLOW with $\alpha_{\rm vir} = 0.2$) to supervirial (Run αHIGH with $\alpha_{\rm vir} = 1.2$).

To change the metallicity, Z, we reduce or increase the abundances of metals to mimic the conditions of massive star forming cores near the Galactic Centre and towards the outer Milky Way disc. The FIDUCIAL run has a solar metallicity of $1\,{\rm Z}_\odot$. We change the metallicity to $0.5\,{\rm Z}_\odot$ (ZLOW) and $2\,{\rm Z}_\odot$ (ZHIGH).

4 Results

4.1 Fiducial Run

We start from a spherical cloud that begins to collapse under its own gravity. Fig. 2 shows the time evolution of the FIDUCIAL run. Substructures which look like filaments emerge in the process due to turbulence in the cloud. The filaments converge centrally and feed the central hub. Star formation first takes place within the central region of the simulated box, but later on extends to the outer regions of the filaments as well.

At first, the dust temperature follows the morphology of the gas temperature. As soon as sink particles are formed the dust is heated by the radiative feedback of the sinks. After $\sim 0.45\,t_{\rm ff}$ (0.237 Myr), when massive sink particles are present, the dust temperature is dominated by the radiation temperature. In the inner parts, radiation, dust, and gas temperatures are mostly in equilibrium. In the outer parts, the gas temperature is still higher due to shock heating. Later on, ionising feedback from massive sink particles heats the gas, and as soon as the bubble full of ionised hydrogen expands, the gas temperature increases significantly. The pressure transferred from stellar radiation helps the bubble to grow. As a consequence, atomic hydrogen is expelled outwards radially and the core becomes dispersed. After the simulated time, only $50\,\%$ of the formed sink particles remain in a gravitationally bound cluster.

4.2 Resolution Study

Starting from the FIDUCIAL run with an effective spatial resolution of $\Delta x = 400\,{\rm AU}$ at refinement level 9, we increase the maximum refinement level in different simulations for the same initial conditions. Refinement level 9 corresponds to a net maximum resolution of $(2048)^3$ cells. We increase the refinement level to 10, 11, and 12 which corresponds to a resolution of $\Delta x = 200\,{\rm AU}$, $\Delta x = 100\,{\rm AU}$, $\Delta x = 50\,{\rm AU}$, respectively. The minimum refinement level is always set to 5, i.e. corresponding to a 128^3 base grid and a base resolution of $\sim$3,200 kAU.

The stars in our simulations are modelled with sink particles. These are checked against several criteria before they are formed or allowed to further accrete gas. One condition

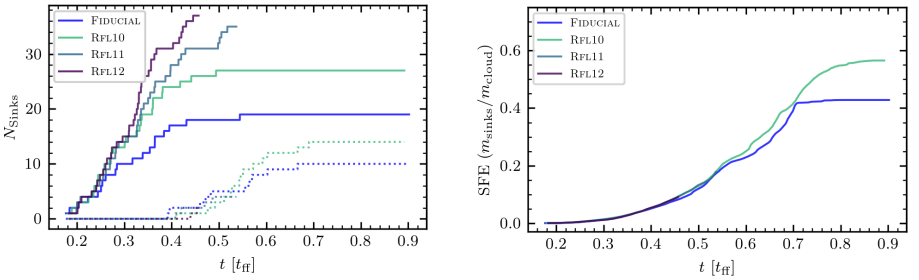


Figure 3. `Left`: Number of sink particles, while the dotted line shows the number of high mass sinks with $\geq 8\,\mathrm{M}_\odot$. The higher the resolution the more sink particles are formed. `Right`: Evolution of the star formation efficiency. The resolution has a minor influence on the general core evolution, but affects the time period where sink particles can accrete mass.

for the formation of sink particles is that the gas density is greater than a certain density threshold, $\rho_{\rm thresh}$.

$$\rho_{\rm thresh} = \frac{\pi c_{\rm s}^2}{G\lambda_{\rm J}^2} = \frac{\pi c_{\rm s}^2}{G(4\Delta x)^2}\,, \tag{4}$$

where G is the gravitational constant and $c_{\rm s}$ is the speed of sound.

It is related to the smallest resolvable Jeans length on the maximum refinement level and is set to $\lambda_{\rm J} = 4\Delta x$. The density threshold increases with increasing density.

During the initial collapse phase, the core cools efficiently via dust thermal emission and behaves isothermally. When the collapsing gas reaches densities $\rho \gtrsim 10^{-13}\mathrm{g\,cm^{-3}}$, it becomes optically thick to infrared radiation, cannot cool efficiently anymore, and behaves almost adiabatically. For $\rho > 10^{-13}\mathrm{g\,cm^{-3}}$, the Jeans mass therefore increases again with increasing density, and no further fragmentation should occur. Hence, a sink particle would represent a single star only at such high densities. In our simulations we reach densities starting from the lowest resolution $\rho = 10^{-17}\mathrm{g\,cm^{-3}}$ to the highest resolution $\rho = 10^{-15}\mathrm{g\,cm^{-3}}$. Therefore, for any maximum refinement level used in this work, sinks could harbour single stars, binaries, or higher order systems.

In the regime we resolve, it is expected that the number of sink particles increases with higher spatial resolution, as is indeed the case (see Fig. 3, left panel). We conclude that cloud fragmentation greatly depends on the resolution. For a detailed analysis of the fragmentation process and the resulting sink mass distribution, it would be crucial to run simulations at a higher refinement level.

On the other hand, we find that the amount of mass which is converted into stars, the so called the star formation efficiency (SFE), is comparable during the collapse phase until $\sim 0.7\,t_{\rm ff}$ (see Fig. 3, right panel). However, while the mass growth of the FIDUCIAL run decreases significantly after this time, RFL10 still accretes at a higher rate. In both cases, the SFE stays constant toward the end of the simulations. However, RFL10 has a higher final SFE of 0.56 (after $\sim 0.85\,t_{\rm ff}$) compared to the FIDUCIAL run, which stops at a SFE of 0.42. At a higher resolution we resolve higher densities and the stars are more deeply embedded thus the feedback is less efficient at dispersing the core and stopping further mass accretion.

Nevertheless, the general evolution of the collapse and core dispersal is similar for different resolutions. In the following, $l_{\mathrm{ref}} = 9$ is therefore used for a parameter study. At this resolution, it is possible to carry out a statistically relevant sample of simulations.

4.3 Parameter Study

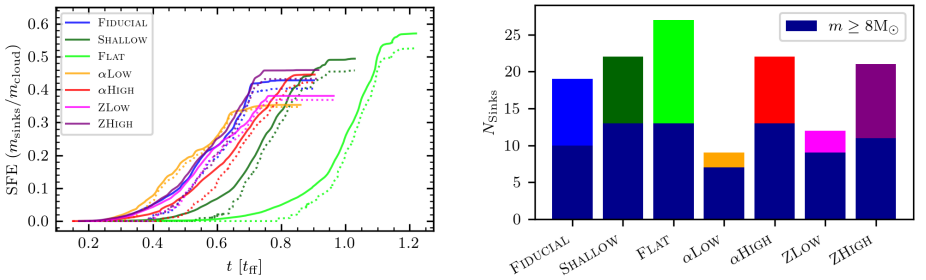


Figure 4. `Left`: Evolution of the star formation efficiency. The dotted lines represent the SFE of massive stars only, which contain most of the mass. `Right`: Number of formed sink particles for the different runs. The dark blue parts indicate the number of high mass stars. Runs with flatter density profiles collapse more slowly, but in the end they produce more fragments resulting in a higher SFE. Low virial parameter and low metallicities lead to fewer sink particles with reduced SFE. A high virial parameter and high velocity dispersion result in a slightly higher SFE and more fragmentation.

The evolution of the collapsing core and the fragmentation process depend on the initial conditions. Each simulation box contains a predefined amount of mass. How much of it is converted to stars (the SFE) is also affected by the initial conditions.

In the FIDUCIAL run the SFE reaches 0.42, which means that 42 % of the available mass is converted into sink particles (see Fig. 4, left panel). In total 19 sinks are formed, 10 of which grow into massive ones (see Fig. 4, right panel).

The *flatter* the initial density profile, the slower the core collapse. Due to the slower evolution, turbulence has more time to interact before the cloud collapses under gravity and more sub-structures are formed which leads to a higher number of formed sink particles (see Fig. 4, right panel). The slower core collapse initially leads to a slower increase in the mass accretion but also extends the time period over which sink particles can accrete mass (see Fig. 4, left panel). Thus, the increase in the SFE is time delayed in the runs FLAT and SHALLOW, but reaches even higher final numbers, 0.57 and 0.49, respectively, than in the FIDUCIAL run.

The virial parameter relates the kinetic and gravitational energies of the initial core. A low virial parameter indicates that the gravitational core collapse is less disturbed by kinetic motion which leads to less substructure formation during the collapse phase and *vice versa*. A low virial parameter leads to fewer (in total 9) but very massive sink particles; however, the SFE decreases to 0.35. A higher virial parameter leads to more fragmentation and 22 sink particles are formed. The SFE reaches 0.45 and ends up slightly higher than in the FIDUCIAL run (see Fig. 4).

The amount of metals in molecular clouds can impact the evolution of massive clouds. Metals are expected to effectively cool gas which may result in a suppression of fragmen-

tation in low metallicity environments. Starting with a lower initial metallicity, Run αLOW shows less fragmentation and only 12 sink particles are formed. The SFE decreases to 0.38. However, the core with a higher initial metallicity (Run ZHIGH) forms 22 sink particles and the SFE increases slightly to 0.45.

4.4 Comparing Simulations and Synthetic Observations

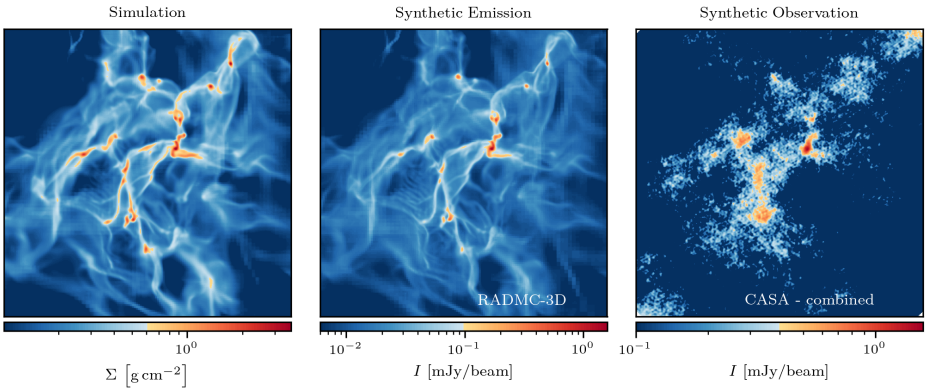


Figure 5. What would a simulation look like if we were to observe it in the sky with a real-world instrument like ALMA? In the left and middle panel the density and the dust emission, respectively, along the line of sight of the simulated image is shown. The right panel shows the same region but seen through from the perspective of the ALMA telescope[1].

From these simulations, synthetic observations are derived (see Fig. 5). The density profile is used to produce the emission of the dust continuum at 1362 μm with the radiative transfer code RADMC-3D[16] (see Fig. 5, middle panel). Furthermore, the software CASA makes it possible to simulate the effects of the instrumental limitations of the ALMA telescope (Atacama Large Millimeter/Submillimeter Array) in order to produce an image of a synthetic observation (see Fig. 5, right panel). The imperfect resolution of the telescope leads to a loss of the filamentary substructures so that only the brightest cores are still detectable. Comparing simulations and observations helps to understand the limitations of telescopes and how observational data can be interpreted.

5 Concluding Remarks

This work presents state-of-the-art numerical simulations in order to study massive star formation. These simulations are only possible on the most powerful supercomputers, and without we would miss crucial theoretical understanding of astrophysical processes. We simulate isolated core-collapse scenarios of cores with 1000 $M_\odot$ within a radius of 1 pc. A novel scheme to treat the radiative transfer of ionising and non-ionising radiation as well as radiation pressure on dust and gas is included. We investigate the formation and early

evolution of massive stars and their host cores up to the point where a bubble full of ionised hydrogen is established.

Sink particles are treated as single stars; while the resolution allows them to represent single stars but also groups of stars and their accretion disk or multiple-order systems. With higher resolution the number of formed sink particles is increased. However, the general trend of the initial core evolution is similar.

The impact of different initial conditions are investigated to produce a statistically relevant sample. The flatter the density profile, the slower the collapse, and the number of sink particles increases, as well as the SFE. With a low virial parameter, gravity is more dominant than turbulence, which leads to a faster core collapse that produces fewer but very massive sink particles. An initial high virial parameter delays the core collapse. Turbulence is more dominant, thus more substructures are produced. A lower metallicity reduces dust cooling, suppressing fragmentation. In this case fewer, but massive sink particles are formed. A higher metallicity leads to efficient dust cooling during the initial collapse phase, resulting in more fragmentation and a higher SFE.

From these simulations we derive synthetic observations while simulating the telescope effects of the ALMA telescope. With the limited resolution of the telescope the brightest cores can be seen while most of substructures become invisible. The comparison between simulations and synthetic observations supports the interpretation of real-world telescope data. This may guide the analysis of existing observations as well as planning of future observations of star-forming regions.

Acknowledgements

The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (`www.gauss-centre.eu`) for funding this project by providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS Supercomputer JUWELS at Jülich Supercomputing Centre (JSC). The software used in this work was in part developed by the DOE NNSA-ASC OASCR Flash Center at the University of Chicago. The authors gratefully acknowledge funding via the Collaborative Research Center 1601 (sub-project A5) funded by the German Science Foundation (DFG).

References

1. `https://www.gauss-centre.eu/results/astrophysics/towards-high-mass-star-formation-confronting-simulations-and-observations`
2. A. D. Bolatto, S. R. Warren, A. K. Leroy, F. Walter, S. Veilleux, E. C. Ostriker, J. Ott, M. Zwaan, D. B. Fisher, A. Weiss, E. Rosolowsky, and J. Hodge, *Suppression of star formation in the galaxy NGC 253 by a starburst-driven molecular wind*, Natur **499**, 450-453, 2013.
3. B. Commerçon, P. Hennebelle, and T. Henning, *Collapse of massive magnetized dense cores using radiation magnetohydrodynamics: early fragmentation inhibition*, The Astrophysical Journal Letters **742**, 1, 2011.
4. M. R. Krumholz, R. I. Klein, C. F. McKee, S. S. R. Offner, and A. J. Cunningham, *The Formation of Massive Star Systems by Accretion*, Science **323**, 754-757, 2009.

5. B. Fryxell, K. Olson, P. Ricker, F. X. Timmes, M. Zingale, D. Q. Lamb, P. MacNeice, R. Rosner, J. W. Truran, and H. Tufo, *FLASH: An Adaptive Mesh Hydrodynamics Code for Modeling Astrophysical Thermonuclear Flashes*, ApJ **131**, 273-334, 2000.
6. D. Derigs, A. R. Winters, G. J. Gassner, S. Walch, and M. Bohm, *Ideal GLM-MHD: About the entropy consistent nine-wave magnetic field divergence diminishing ideal magnetohydrodynamics equations*, JCP **364**, 420-467, 2018.
7. R. Wünsch, S. Walch, F. Dinnbier, and A. Whitworth, *Tree-based solvers for adaptive mesh refinement code FLASH - I: gravity and optical depths*, MNRAS **475**, 3393-3418, 2018.
8. R. Wünsch, S. Walch, D. František, D. Seifried, S. Haid, A. Klepitko, A. Whitworth, and J. Palouš, *Tree-based solvers for adaptive mesh refinement code FLASH - II: radiation transport module TreeRay*, MNRAS **505**, 3730-3754, 2021.
9. A. Klepitko, S. Walch, R. Wünsch, D. Seifried, F. Dinnbier, and S. Haid, *Tree-based solvers for adaptive mesh refinement code FLASH - III: a novel scheme for radiation pressure on dust and gas and radiative transfer from diffuse sources*, MNRAS **475**, 3393-3418, 2018.
10. S. C. O. Glover and M.-M. Mac Low, *Simulating the Formation of Molecular Clouds. I. Slow Formation by Gravitational Collapse from Static Initial Conditions*, ApJS **169**, 239-268, 2007.
11. S. C. O. Glover, C. Federrath, M. M. Mac Low, and R. S. Klessen, *Modelling CO formation in the turbulent interstellar medium*, MNRAS **404**, 2-29, 2010.
12. S. Walch, P. Girichidis, T. Naab, A. Gatto, S. C. O. Glover, R. Wünsch, R. S. Klessen, P. C. Clark, T. Peters, D. Derigs, and C. Baczynski, *The SILCC (SImulating the LifeCycle of molecular Clouds) project - I. Chemical evolution of the supernova-driven ISM*, MNRAS **454**, 238-268, 2015.
13. S. Haid, S. Walch, D. Seifried, R. Wünsch, F. Dinnbier, and T. Naab, *The relative impact of photoionizing radiation and stellar winds on different environments*, MNRAS **478**, 4799-4815, 2018.
14. F. Dinnbier and S. Walch, *How fast do young star clusters expel their natal gas? Estimating the upper limit of the gas expulsion time-scale*, MNRAS **499**, 748-767, 2020.
15. M. Klassen, R. E. Pudritz, and T. Peters, *Simulating protostellar evolution and radiative feedback in the cluster environment*, MNRAS **421**, 2861-2871, 2012.
16. C. P. Dullemond, A. Juhasz, A. Pohl, F. Sereshti, R. Shetty, T. Peters, B. Commercon, and M. Flock, *RADMC-3D: A multi-purpose radiative transfer tool*, Astrophysics Source Code Library **ascl**, 1202, 2012.

Understanding the Origin of the Heaviest Elements in the Universe with Kilonova Radiative Transfer Simulations

Christine E. Collins[1,2], Luke J. Shingles[1], Fiona McNeill[3], Andreas Bauswein[1], Gabriel Martínez-Pinedo[1] and Stuart A. Sim[3]

[1] GSI Helmholtzzentrum für Schwerionenforschung, Planckstraße 1, 64291 Darmstadt, Germany
E-mail: c.collins@gsi.de

[2] School of Physics, Trinity College Dublin, The University of Dublin, Dublin 2, Ireland

[3] Astrophysics Research Centre, School of Mathematics and Physics, Queen's University Belfast, Belfast BT7 1NN, UK

Kilonovae are the explosive bursts of light resulting from neutron star collisions. The extreme, neutron-rich conditions during the collision allow the rapid neutron-capture process (r-process) to take place, which is responsible for producing many of the heaviest elements in the Universe. Kilonovae provide the opportunity to understand r-process nucleosynthesis, as well as to constrain high-density matter physics. We have established a self-consistent modelling pipeline that allows us to compare kilonova simulations directly to observations, enabling the interpretation of these events. Our work has highlighted the importance of accurate atomic data for modelling kilonovae, as well as the importance of 3D simulations.

1 Introduction

When neutron stars collide, a bright, fast-evolving kilonova transient is produced. In 2017 a kilonova was observed (AT2017gfo[1,2]) following a gravitational-wave signal (GW170817[3]), igniting the field of multi-messenger astronomy. Understanding these cataclysmic events is the key to determining the origin of the heavy elements in our Universe, including gold, platinum and uranium. Kilonovae provide promising opportunities to study matter under extreme conditions offering a window into the dynamics of extremely dense nuclear matter.

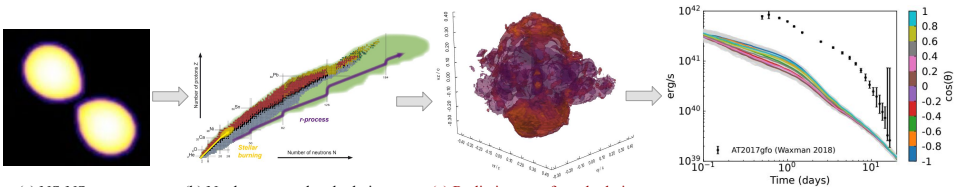
The astrophysical site where around half of all elements heavier than iron are synthesised by the rapid neutron-capture process (r-process), has long been debated and observations of AT2017gfo have strongly supported the ejecta of binary neutron star mergers as the primary sites for the r-process[4]. The observations of AT2017gfo have provided a powerful set of constraints for testing theoretical models of binary neutron star mergers, the incompletely known Equation of State (EoS) of dense nuclear matter, r-process nucleosynthesis, and radiative transfer for kilonovae. However, accurate theoretical models are required to link these observations back to information about the underlying physical conditions. To identify specific elements produced by binary neutron star mergers, we must interpret kilonova observations, and for this we need radiative transfer simulations.

The strongest feature in the observed spectra of AT2017gfo has been suggested to be Sr II[5–7]. However, most studies identifying this feature have used simplified radiative transfer methods, such as a backwards modelling approach (starting with observations and selecting a composition that produces matching spectra)[5], or parameterised ejecta models[6,7] rather than a forwards modelling approach using advanced multidimensional simulated binary neutron star merger ejecta as a basis for radiative transfer calculations. Most kilonova

radiative transfer simulations have been carried out in 1D or 2D[8–11] with only a handful of 3D calculations[12,13], which includes our 3D forwards modelling simulations of Collins et al. (2023a)[14] and Shingles et al. (2023)[15].

The aim of our project is to increase our understanding of kilonovae and their role in the production of heavy elements by performing calculations with our advanced radiative transfer method in three dimensions based on binary neutron star merger ejecta from numerical models. This pipeline is vital for the detailed interpretation of the spectra of AT2017gfo and of future kilonovae.

2 Methods



(a) NS-NS merger simulation (b) Nuclear network calculation (c) Radiative transfer calculation based on ejecta snapshot (d) Compare to observations

Figure 1. Pipeline to self-consistently simulate kilonovae from binary neutron star merger ejecta. A hydrodynamical neutron star merger simulation is carried out. Following this, r-process nuclear network calculations are carried out based on the merger simulation to calculate the nucleosynthetic abundances (image credit: EMMI, GSI/Different Arts). A snapshot of the merger simulation and the nucleosynthetic abundances are input to the radiative transfer calculation (shown is the density structure of the ejecta snapshot). The radiative transfer simulation produces light curves and spectra that can be directly compared to kilonova observations. Shown are simulated light curves compared to AT2017gfo (figure adapted from Ref. 14).

We use a multi-disciplinary pipeline to carry out our simulations, allowing us to self-consistently model kilonovae starting with binary neutron star merger simulations through to comparing synthetic observables directly to observations (see Fig. 1). State-of-the-art, relativistic 3D simulations of the ejecta from binary neutron star mergers are input to our radiative transfer simulations. The energy released during the merger (from β-decays, α-decays and fission fragments) and the nuclear abundances are obtained from nuclear network calculations (as in Mendoza et al.[16]). Using the simulated merger ejecta and nuclear network calculations as input, radiative transfer simulations are carried out for a snapshot of the merger ejecta, using the radiative transfer code ARTIS to follow the subsequent expansion, radioactive decay, and radiative transfer to to produce synthetic light curves and spectra. These can be directly compared to observations, linking the observations back to the underlying merger ejecta.

ARTIS is a multi-dimensional, state-of-the-art radiative transfer code. Shingles et al. (2023)[15] have enabled ARTIS to carry out simulations of kilonovae with line-by-line opacities for millions of bound-bound transitions of r-process elements for the first time in 3D. Importantly, this allows us to directly associate spectral features with specific elements. The main advancements include the use of a relativistic Doppler shift (for the

rapidly expanding ejecta), numerical improvements for handling of dense line lists (millions of transitions), an extended input snapshot that specifies each cell's nuclear abundances and energy release from prior reactions, and handling of α- and β-decays with a consistently evolving composition and time-dependent thermalisation of the emitted decay products.

2.1 Numerical Methods

ARTIS uses Monte Carlo methods[17–19] to simulate the complete radiation transport problem from energy injection all the way to the eventual escape of radiation from the ejecta. In kilonovae, energy is released from the radioactive decays of r-process material synthesised during the merger. Shingles et al. (2023) have greatly expanded the range of decays handled by ARTIS to include α and β decays, and added a new non-instantaneous thermalisation treatment. By leveraging a set of detailed network calculations performed on the 3D hydrodynamic trajectories (for the first few minutes of r-process reactions) and following each individual nuclear decay with its associated γ-ray spectrum and particle thermalisation conditions, we can self-consistently model the energy released.

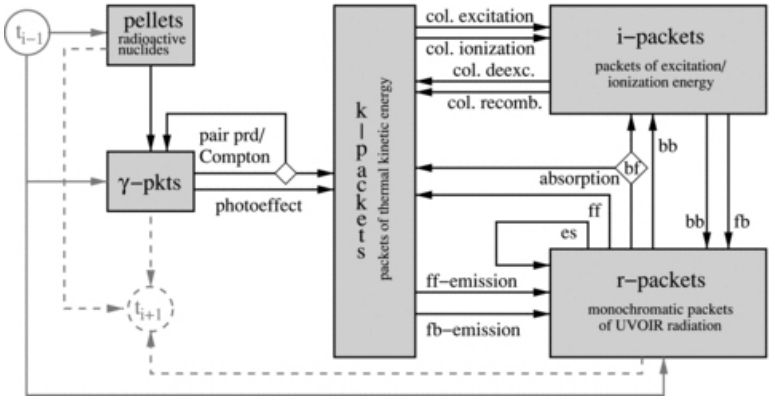


Figure 2. Flow chart outlining the mode of operation of the radiative transfer code ARTIS, and how physical processes (absorption, emission and scattering) are modelled within the framework of ARTIS. Figure taken from Kromer et al. (2009)[20].

At the beginning of a simulation, energy "pellets" are placed (and their decay-times set) within the ejecta according to the spatial and temporal distribution of radioactive decay energy throughout the selected time range. The pellets are activated according to the radioactive decays and become Monte Carlo "packets" of indivisible energy, which are then propagated through the expanding ejecta in three-dimensions, with transitions between packet types (e.g., γ-ray, kinetic energy, or optical energy, see Fig. 2) representing absorption, emission and scattering processes according to detailed Monte Carlo statistics, using the macro-atom formalisation[17,18]. Our line-by-line treatment enables each transition to be treated individually using the Sobolev approximation[21,22]. By considering individual lines, we can track the species responsible for transitions, thereby directly associating

spectral features with the underlying physical process. The outgoing packets of radiation (photons) are binned according to their direction, time of escape, and wavelength, allowing for synthetic light curves and spectra to be produced as a function of observer direction and time. This allows us to compute synthetic observables from multidimensional binary neutron star merger models in a self-consistent, time-dependent manner, giving significant predictive power to the merger simulation and allowing its outcome to be properly tested against observations.

Such sophisticated radiative transfer simulations are numerically expensive owing to the large numbers of Monte Carlo quanta that must be propagated. Fortunately, since the propagation of each quantum is independent of the others, the Monte Carlo scheme is extremely well-suited to parallelisation across very large numbers of cores. The code is fully parallelised with both MPI and OpenMP.

3 Scientific Results

We now discuss our first results produced for this project which have been published[14, 15, 23]. For these studies we chose a merger simulation of equal-mass 1.35 $M_\odot$ neutron stars as input to our radiative transfer simulations. Our ongoing work is to consider a broader range of neutron star merger simulation ejecta models in our radiative transfer calculations.

3.1 Neutron Star Merger Simulation

The merger simulation considered here (described by Collins et al.[14]) uses the SFHo[24] equation of state. It was carried out with a 3D general relativistic smoothed-particle hydrodynamics (SPH) code[25–27] and included an advanced neutrino leakage treatment, ILEAS (Improved leakage-equilibration-absorption scheme)[28]. The hydrodynamical simulation followed the evolution until 20 ms after the merger, and therefore only includes "dynamical" ejecta, i.e., the material becoming unbound during the early postmerger phase. The total mass of ejecta from this simulation is 0.005 $M_\odot$. It is expected that matter ejection will continue after this simulation was stopped, however, long-term evolution simulations are required to follow the hydrodynamics beyond this time (e.g., see Just et al.[29] or Kawaguchi et al.[30]).

3.2 Comparison to AT2017gfo

The spectra predicted by our radiative transfer simulations for this neutron star merger model show remarkable agreement with the observed evolution of the kilonova AT2017gfo, considering that we in no way tuned our model to try to match the observations. A comparison between the simulated and observed spectra is shown in Fig. 3. The simulation predicts a structure similar to the strong feature observed in AT2017gfo identified as Sr II. In our simulation this feature is predominantly due to Sr II, however, it also contains contributions from Y II and Zr II. With our line-by-line opacity treatment, we can identify the species forming spectral features. An example of this is shown by the colour coding in Fig. 3, which indicates the relative contributions of specific ions to the emitted spectra. Specifically, each Monte Carlo packet of radiation escaping the simulation is tagged with

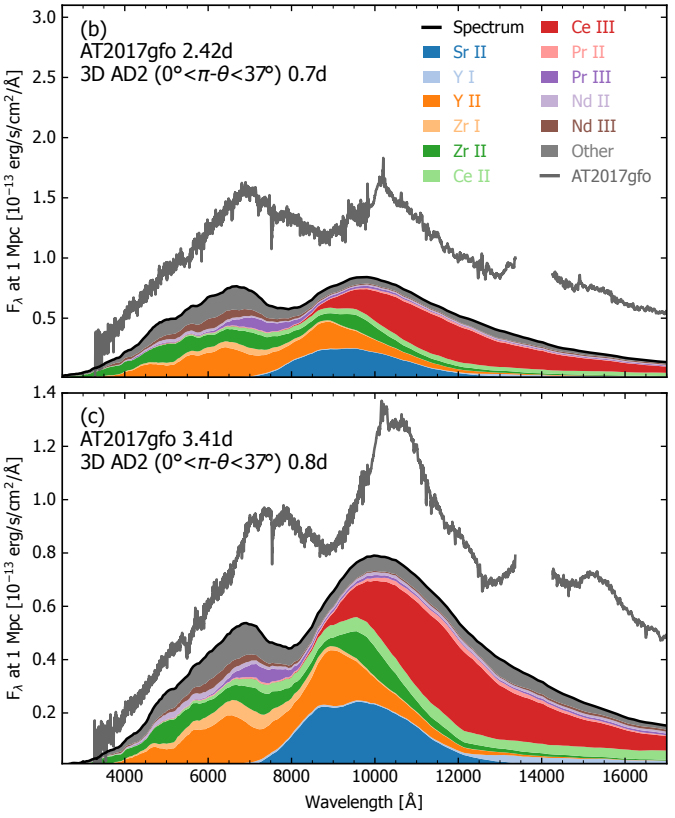


Figure 3. Spectra in the polar direction for model 3D AD2 from Shingles et al. (2023), which show remarkable similarities to the spectra of AT2017gfo, considering that we did not tune the model in any way to try to match the data. The colour coding indicates the species contributing to the formation of the spectrum, thereby directly associating spectral features with the ions responsible. Figure taken from Shingles et al. (2023)[15].

the last interaction underwent by the packet. For each wavelength bin in the synthetic spectrum, the area under the spectrum is colour coded in proportion to the energy carried by packets in that wavelength bin whose last interaction was with each of the ions considered. This analysis shows that Sr II makes significant contributions in forming the simulated spectra. Other significant contributions come from Y II, Zr II and Ce III, as can be seen in Fig. 3. Therefore, our simulations strongly support the interpretation of Sr being present in the ejecta of AT2017gfo and thus solidifies the conclusion that the r-process took place in the outflow of GW170817.

Although the evolution of the spectra is similar to that observed in AT2017gfo, our simulated spectra evolve too quickly compared to the observations. In Fig. 3, the simulated spectra are plotted at 0.7 and 0.8 days after the merger, and resemble the observations at 2.4 and 3.4 days. It is likely that the fast evolution is due to the lower ejecta mass in the merger model we selected for this study. The simulated ejecta mass is 0.005 $M_{\odot}$, which is around ten times lower than the mass inferred for AT2017gfo[1]. This motivates our aim

with this project to investigate the kilonovae predicted for a range of ejecta models with varying masses, and particularly models with higher ejecta masses more similar to the inferred mass of AT2017gfo.

3.3 Importance of Accurate Atomic Data

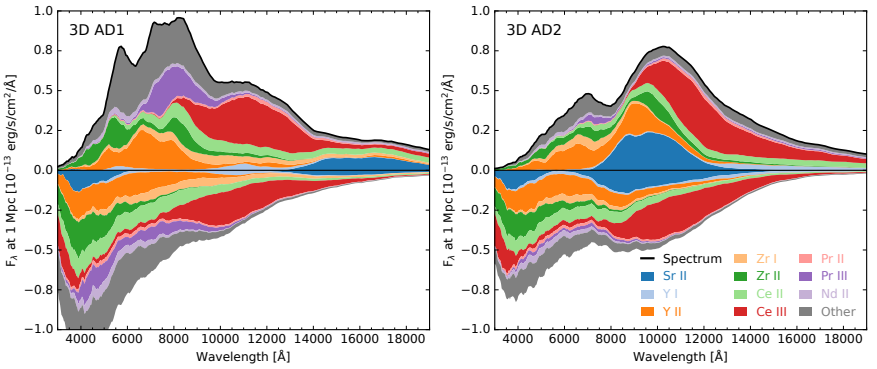


Figure 4. Simulated spectra at 0.8 days after the merger using the AD1 or AD2 atomic dataset. The only difference between AD1 and AD2 is that in AD2 the atomic data for Sr, Y and Zr has been replaced with calibrated atomic data instead of the theoretically calculated data in AD1. Using improved atomic data for only these elements has a dramatic effect on the predicted spectra, both in the spectral features predicted and in the overall spectral energy distribution. The colour coding indicates the relative contributions of specific ions to shaping the spectra. Above the axis indicates the last ion responsible for emitting or scattering an escaping packet of radiation. Beneath the axis indicates the absorption processes that last prevented a packet of radiation from escaping, thus indicating the species responsible for absorption. Figure adapted from Shingles et al. (2023)[15].

We have carried out a study into the importance of accurate atomic data. To do this, we carried out radiative transfer simulations using the same ejecta model, but different atomic datasets. The majority of our atomic data is sourced from the Japan-Lithuania Opacity Database for kilonovae[31], which is theoretically calculated and not calibrated to experimentally known values. We refer to this dataset as AD1. To test the importance of accurate atomic data, we replace the atomic data for Sr, Y and Zr in AD1 with experimentally calibrated data sourced from the Kurucz[32] extended line list. We refer to this dataset as AD2, which is the dataset used to produce the spectra in Fig. 3.

The difference that results from changing only the atomic data for Sr, Y and Zr is shown in Fig. 4. The spectral features predicted as well as the overall spectral energy distribution changes significantly when the calibrated atomic data is included. This highlights the need for accurate atomic data in radiative transfer simulations.

3.4 Importance of 3D Simulations

The importance of 3D radiative transfer simulations was also tested by comparing the 3D simulation to a 1D simulation based on the spherical average of the ejecta model. The comparison of the light curves from these simulations is shown in Fig. 5. The 1D

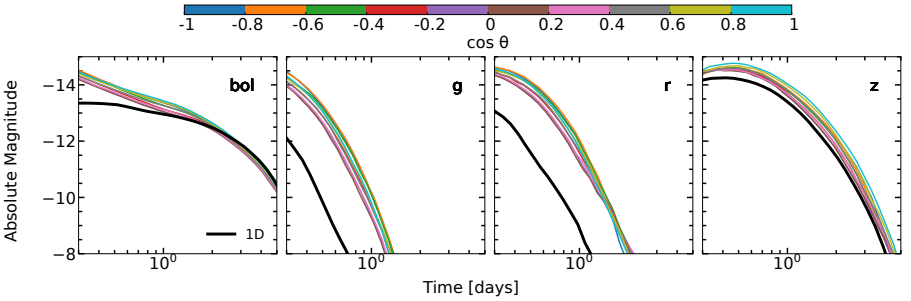


Figure 5. Simulated bolometric and grz–band light curves from the 3D simulation and from a 1D simulation using a spherical average of the ejecta model. The colour bar indicates the polar angle of each observer direction in the 3D model while the black lines show the light curves of the 1D model.

simulation is unable to reproduce the light curves in any line of sight of the 3D simulation. In particular, the g and r band light curves from the 1D simulation are fainter than all directions of the 3D simulation. This shows that 1D simulations may lead to overestimating the opacity at red wavelengths. Even the bolometric light curve is unable to match the 3D simulation, except at late times when the ejecta have become optically thin and the light curve follows the energy deposition rate. This demonstrates that 3D simulations are important for kilonovae modelling.

4 Concluding Remarks

We have established a kilonova modelling pipeline to self-consistently describe kilonovae, to directly compare simulations to observations. Our initial study has demonstrated the ability of these simulations to produce synthetic observables comparable to AT2017gfo. This work has highlighted the importance of accurate atomic data in radiative transfer simulations, not only for predicting specific spectral features, but also to predict the correct overall spectral energy distribution. We have also highlighted the need for 3D kilonova simulations. The 1D spherically averaged model does not reproduce any observer direction in the 3D simulation. Our project will continue to exploit our self-consistent modelling pipeline to investigate a range of neutron star merger models, allowing us to investigate the potential variability of kilonovae, and to place constraints on the underlying physics.

Acknowledgements

The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (`www.gauss-centre.eu`) for funding this project by providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS Supercomputer JUWELS at Jülich Supercomputing Centre (JSC). This project has received funding from the European Union's Horizon Europe research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 101152610. This work is funded/co-funded by the European Union (ERC, HEAVYMETAL, 101071865). Views and opinions

expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. LJS and GMP acknowledge support by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (ERC Advanced Grant KILONOVA No. 885281). AB, CEC, GMP and LJS acknowledge support by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project-ID 279384907 - SFB 1245 and MA 4248/3-1. The work of SAS was supported by the Science and Technology Facilities Council [grant numbers ST/P000312/1, ST/T000198/1, ST/X00094X/1].

References

1. S. J. Smartt et al., *A kilonova as the electromagnetic counterpart to a gravitational-wave source*, Nature, **551**, no. 7678, 75-79, Nov. 2017.
2. V. A. Villar, J. Guillochon, E. Berger, B. D. Metzger, P. S. Cowperthwaite, M. Nicholl, K. D. Alexander, P. K. Blanchard, R. Chornock, T. Eftekhari, W. Fong, R. Margutti, and P. K. G. Williams, *The Combined Ultraviolet, Optical, and Near-infrared Light Curves of the Kilonova Associated with the Binary Neutron Star Merger GW170817: Unified Data Set, Analytic Models, and Physical Implications*, ApJ, **851**, no. 1, L21, Dec. 2017.
3. B. P. Abbott et al., *GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral*, Phys. Rev. Lett., **119**, 161101, Oct 2017.
4. J. J. Cowan, F.-K. Thielemann, and J. W. Truran, *The R-process and nucleochronology*, Phys. Rep., **208**, no. 4-5, 267-394, Nov. 1991.
5. D. Watson, C. J. Hansen, J. Selsing, A. Koch, D. B. Malesani, A. C. Andersen, J. P. U. Fynbo, A. Arcones, A. Bauswein, S. Covino, A. Grado, K. E. Heintz, L. Hunt, C. Kouveliotou, G. Leloudas, A. J. Levan, P. Mazzali, and E. Pian, *Identification of strontium in the merger of two neutron stars*, Nature, **574**, no. 7779, 497-500, Oct. 2019.
6. N. Domoto, M. Tanaka, S. Wanajo, and K. Kawaguchi, *Signatures of r-process Elements in Kilonova Spectra*, ApJ, **913**, no. 1, 26, May 2021.
7. N. Domoto, M. Tanaka, D. Kato, K. Kawaguchi, K. Hotokezaka, and S. Wanajo, *Lanthanide Features in Near-infrared Spectra of Kilonovae*, ApJ, **939**, no. 1, 8, Nov. 2022.
8. D. Kasen, R. Fernández, and B. D. Metzger, *Kilonova light curves from the disc wind outflows of compact object mergers*, MNRAS, **450**, no. 2, 1777-1786, June 2015.
9. K. Kawaguchi, S. Fujibayashi, M. Shibata, M. Tanaka, and S. Wanajo, *A Low-mass Binary Neutron Star: Long-term Ejecta Evolution and Kilonovae with Weak Blue Emission*, ApJ, **913**, no. 2, 100, June 2021.
10. O. Just, I. Kullmann, S. Goriely, A. Bauswein, H. T. Janka, and C. E. Collins, *Dynamical ejecta of neutron star mergers with nucleonic weak processes - II: kilonova emission*, MNRAS, **510**, no. 2, 2820-2840, Feb. 2022.
11. K. Kawaguchi, S. Fujibayashi, K. Hotokezaka, M. Shibata, and S. Wanajo, *Electromagnetic Counterparts of Binary-neutron-star Mergers Leading to a Strongly Magnetized Long-lived Remnant Neutron Star*, ApJ, **933**, no. 1, 22, July 2022.

12. M. Tanaka and K. Hotokezaka, *Radiative Transfer Simulations of Neutron Star Merger Ejecta*, ApJ, **775**, no. 2, 113, Oct. 2013.
13. A. Neuweiler, T. Dietrich, M. Bulla, S. V. Chaurasia, S. Rosswog, and M. Ujevic, *Long-term simulations of dynamical ejecta: Homologous expansion and kilonova properties*, Phys. Rev. D, **107**, no. 2, 023016, Jan. 2023.
14. C. E. Collins, A. Bauswein, S. A. Sim, V. Vijayan, G. Martínez-Pinedo, O. Just, L. J. Shingles, and M. Kromer, *3D radiative transfer kilonova modelling for binary neutron star merger simulations*, MNRAS, **521**, no. 2, 1858-1870, May 2023.
15. L. J. Shingles, C. E. Collins, V. Vijayan, A. Flörs, O. Just, G. Leck, Z. Xiong, A. Bauswein, G. Martínez-Pinedo, and S. A. Sim, *Self-consistent 3D Radiative Transfer for Kilonovae: Directional Spectra from Merger Simulations*, ApJ, **954**, no. 2, L41, Sept. 2023.
16. J. d. J. Mendoza-Temis, M.-R. Wu, K. Langanke, G. Martínez-Pinedo, A. Bauswein, and H.-T. Janka, *Nuclear robustness of the r process in neutron-star mergers*, Phys. Rev. C, **92**, no. 5, 055805, Nov. 2015.
17. L. B. Lucy, *Monte Carlo transition probabilities*, A&A , **384**, 725-735, Mar. 2002.
18. L. B. Lucy, *Monte Carlo transition probabilities. II.*, A&A , **403**, 261-275, May 2003.
19. L. B. Lucy, *Monte Carlo techniques for time-dependent radiative transfer in 3-D supernovae*, A&A , **429**, 19-30, Jan. 2005.
20. M. Kromer and S. A. Sim, *Time-dependent three-dimensional spectrum synthesis for Type Ia supernovae*, MNRAS, **398**, 1809-1826, Aug. 2009.
21. H. J. G. L. M. Lamers and J. P. Cassinelli, *Introduction to Stellar Winds*, Cambridge University Press, 1999.
22. I. Hubeny and D. Mihalas, *Theory of Stellar Atmospheres*, Princeton Series in Astrophysics, 2014.
23. C. E. Collins, L. J. Shingles, A. Bauswein, S. A. Sim, T. Soultanis, V. Vijayan, A. Flörs, O. Just, G. Leck, G. Lioutas, G. Martínez-Pinedo, A. Sneppen, D. Watson, and Z. Xiong, *Towards inferring the geometry of kilonovae*, MNRAS, **529**, no. 2, 1333-1346, Apr. 2024.
24. A. W. Steiner, M. Hempel, and T. Fischer, *Core-collapse Supernova Equations of State Based on Neutron Star Observations*, ApJ, **774**, no. 1, 17, Sept. 2013.
25. R. Oechslin, S. Rosswog, and F.-K. Thielemann, *Conformally flat smoothed particle hydrodynamics application to neutron star mergers*, Phys. Rev. D, **65**, no. 10, 103005, May 2002.
26. R. Oechslin, H. T. Janka, and A. Marek, *Relativistic neutron star merger simulations with non-zero temperature equations of state. I. Variation of binary parameters and equation of state*, A&A , **467**, no. 2, 395-409, May 2007.
27. A. Bauswein, R. Oechslin, and H.-T. Janka, *Discriminating strange star mergers from neutron star mergers by gravitational-wave measurements*, Phys. Rev. D, **81**, no. 2, 024012, Jan. 2010.
28. R. Ardevol-Pulpillo, H. T. Janka, O. Just, and A. Bauswein, *Improved leakage-equilibration-absorption scheme (ILEAS) for neutrino physics in compact object mergers*, MNRAS, **485**, no. 4, 4754-4789, June 2019.
29. O. Just, V. Vijayan, Z. Xiong, S. Goriely, T. Soultanis, A. Bauswein, J. Guilet, H. Th Janka, and G. Martínez-Pinedo, *End-to-end Kilonova Models of Neutron Star Mergers with Delayed Black Hole Formation*, ApJ, **951**, no. 1, L12, July 2023.

30. K. Kawaguchi, N. Domoto, S. Fujibayashi, K. Hayashi, H. Hamidani, M. Shibata, M. Tanaka, and S. Wanajo, *Three dimensional end-to-end simulation for kilonova emission from a black-hole neutron-star merger*, Apr. 2024, arXiv:2404.15027.
31. D. Kato, I. Murakami, M. Tanaka, S. Banerjee, G. Gaigalas, L. Radžiūtė, and P Rynkun, Japan-Lithuania Opacity Database for Kilonova (version 1.1), 2021.
32. R. L. Kurucz, *Including All the Lines: Data Releases for Spectra and Opacities through 2017*, in: Workshop on Astrophysical Opacities, ASP Conference Series, **515**, 47-54, Aug. 2018.

Theoretical Chemistry

Theoretical Chemistry

Christine Peter

Theoretical and Computational Chemistry, University of Konstanz, 78457 Konstanz, Germany
E-mail: christine.peter@uni-konstanz.de

In the proceedings of the NIC Symposium 2025 both contributions in the chemistry section are devoted to advancing methodologies towards the characterisation or the design of materials.

In *GPU Acceleration of Three-Center Coulomb Integral Evaluation with Numeric Atom-Centered Orbitals* Francisco Delesma, Moritz Leucke, Ramón Panadés-Barrueta, and Dorothea Golze present their recent work to accelerate quantum chemical calculations by adapting the time consuming three-center Coulomb integral (3c-CI) evaluation to modern CPU/GPU high performance computing architectures. This work is carried out in the context of the development of highly accurate GW-based methods that are, for example, used for the prediction of X-ray spectroscopic data of materials systems. In general, 3c-CI are crucial for approximating four-center two-electron Coulomb integrals (4c-CIs) in many quantum chemical methods, including Hartree-Fock (HF), coupled cluster, and also the GW approximations. Direct computation of 4c-CIs is computationally expensive, thus the resolution-of-the-identity (RI) approach is a popular method to reduce this cost. The RI technique is primarily used with localised basis sets, where numeric atom-centered orbitals (NAOs), which are evaluated on numerical grids, have emerged as a promising and highly flexible alternative to other basis set approaches. NAOs are used in the FHI-aims package, the solid-state all-electron software package used by the authors. Previously, the Golze group had successfully made efforts to accelerate and reduce the scaling of core-level GW steps. This had resulted in the 3c-CI evaluation emerging as a remaining major computational bottleneck even for comparatively large systems of more than 100 atoms. Here, the authors present their impressive algorithmic advances and implementation of the acceleration of the computation of 3c-CIs based on CUDA for CPU/GPU HPC platforms as well as the benchmarking on JUWELS Booster. For medium-sized basis sets, a two-fold speedup, for larger basis sets, speedups of up to six-fold are achieved. Moreover, an important step forward has been made towards a full GPU implementation of these quantum chemical calculations which will be instrumental for the later use on GPU-based exascale machines.

In *Machine Learning for Accelerated Discovery and Design of Functional Energy Materials* Mohammad Eslamibidgoli, Max Dreger, Andre Colliard-Granero, Fabian Tipp, Michael Eikerling, and Kourosh Malek present a project performed on the JUWELS machine where they produce a data set of imidazolium-based compounds and their alkaline stability. This alkaline stability is one of the major factors in the development of improved anion exchange membranes (AEMs) for hydrogen fuel cells or water electrolysis. First, the authors computed degradation pathways with the help of density functional theory and coupled cluster approaches for a few compounds for which experimental reference data were available. From this data the free energy difference related to a hydroxide attack on the imidazolium ring was determined as a reliable descriptor of alkaline stability. This descriptor

was then evaluated computationally on JUWELS for a library of about 5800 imidazolium-based structures. The so-obtained data also serve as a basis for subsequent training of machine learning models. In the article, the authors further outline, how this project is embedded in a much wider framework of data-driven and machine learning methodologies to aid the development of sustainable energy technologies as part of the European Materials Modelling Ontology project.

These two articles give a very good impression of the breadth of methodological developments in theoretical materials chemistry: they span from the improvement of sophisticated and highly efficient electronic structure theory calculations to the machine-learning based design of new materials and the required generation of extensive electronic structure datasets.

GPU Acceleration of Three-Center Coulomb Integral Evaluation with Numeric Atom-Centered Orbitals

Francisco A. Delesma*, Moritz Leucke*,
Ramón L. Panadés-Barrueta, and Dorothea Golze

Faculty of Chemistry and Food Chemistry, Technische Universität Dresden,
01062 Dresden, Germany
E-mail: {francisco_antonio.delesma_diaz, moritz.leucke, dorothea.golze}@tu-dresden.de

In electronic structure theory calculations, the choice of basis set for expanding the wave function or electronic density is crucial for achieving both accuracy and performance. Numeric atom-centered orbitals have become popular due to their compact and localised nature, which enables accurate and efficient calculations for large molecules and solid-state systems. However, the numerical evaluation of three-center Coulomb integrals (3c-CIs), which appear in Hartree-Fock and correlated methods within the resolution-of-the-identity approach, can become a bottleneck in practical calculations. In this work, we detail and benchmark our reimplementation of the 3c-CI evaluation, leveraging graphical processing units (GPUs) to accelerate the calculations. For medium-sized basis sets, we achieve a 2x speedup, while for larger basis sets, speedups of 4x to 6x can be obtained for the 3c-CI evaluation.

1 Introduction

The four-center two-electron Coulomb integrals (4c-CIs) are present in various methods within electronic structure theory, including hybrid functionals[1], second-order Møller-Plesset (MP2) perturbation theory[2], coupled cluster methods[3,4], the Random Phase Approximation (RPA)[5,6], and the GW approximation[7]. The computation of the 4c-CIs scales $\mathcal{O}(N^4)$ with respect to system size N. The resolution-of-the-identity (RI) approach[8,9] is a popular method to reduce the computational cost for the evaluation of the 4c-CIs. The RI method refactors the 4c-CIs in products of two-center and three-center integrals. Some of the two-center integrals are Coulomb integrals, while the three-center integrals, as well as additional two-center integrals, can involve different interaction potentials depending on the specific RI flavour. In this work, the two- and three-center integrals are also Coulomb integrals (2c-CIs and 3c-CIs). When employing the RI approximation, the computation of the 3c-CIs dominates the integral evaluation with a computational complexity of $\mathcal{O}(N^2)$ to $\mathcal{O}(N^3)$.

In electronic structure theory, the wave function or electronic density is expanded in a basis. Broadly, we distinguish plane wave basis sets and localised basis sets. The latter are confined to certain regions in space, typically around an atom, while plane wave basis sets spread out over the entire structure. As demonstrated in the Supporting Information of Ref. 10, the RI reformulation of the 4c-CIs within a plane-wave framework reduces to trivial expressions, unlike for localised basis sets. Therefore, RI techniques and their various flavours are primarily used in the context of localised basis sets. While plane-wave basis sets are more commonly used in the solid-state community, localised basis sets

*These authors contributed equally to this work.

are prevalent in quantum chemistry codes due to their primary application to molecular systems. However, localised basis sets are also employed in solid-state codes, such as FHI-aims[11], the all-electron software package used in this project.

Localised basis sets may have an analytic form, such as Gaussian-type orbitals (GTOs) or Slater-type orbitals (STOs). An alternative are numeric atom-centered orbitals (NAOs), which are evaluated on numerical grids and which are used in FHI-aims. GTOs and STOs can be considered as a special case of an NAO. The NAOs are defined as:

$$\phi_\mu(\mathbf{r}) = \frac{u_\mu(r)}{r} Y_{lm}(\hat{\mathbf{r}}) \tag{1}$$

where u_μ are radial functions and $Y_{lm}(\hat{\mathbf{r}})$ spherical harmonics. The radial component of an NAO is entirely flexible and not constrained to any specific form. u_μ is the solution to the radial Schrödinger equation calculated on a dense logarithmic grid, while $Y_{lm}(\hat{\mathbf{r}})$ is evaluated on angular grids.

Compared to GTOs and STOs, NAOs are more flexible because they are not constrained to a predefined analytical shape. This is an advantage in solid-state systems, where atomic environments can vary significantly[12]. Additional to FHI-aims, packages such as ABACUS[13], OpenMX[14], SIESTA[15], and among others[16,17], have adopted the NAO methodology. Furthermore, NAOs can be numerically adjusted to describe both core and valence electron behaviour well. Compared to NAOs, GTO basis sets require many functions to describe deep core electrons correctly due to the incorrect description of the cusp behaviour at the nucleus. NAO basis sets are thus usually smaller than GTOs and can achieve higher accuracy with less basis functions[18].

2 Motivation

While NAOs are typically more compact than, for example, GTOs, the evaluation of the integrals is more challenging. Analytical techniques are available for GTOs[19–21], while NAO integrals are computed numerically on grids. The numerical integration has a larger computational prefactor than an analytical evaluation and substantially contributes to the overall computational cost. For example, we demonstrated that in core-level GW calculations, the 3c-CI computation over NAOs is the computationally most expensive step for system sizes up to 40-50 atoms[22], despite its scaling being only $O(N^2)$ to $O(N^3)$. The higher scaling GW-specific steps, with $O(N^4)$ and $O(N^5)$ complexity, dominate the computational cost only for larger systems. Our recent efforts[23] to reduce the scaling of the GW steps have resulted in the 3c-CI evaluation contributing to half of the total computational time, even for systems larger than 100 atoms. Therefore, our goal is to accelerate the computation of 3c-CIs by utilising new hybrid architectures that combine CPUs and graphical processing units (GPU). We present here our latest algorithmic advances based on CUDA for CPU/GPU high performance computing (HPC) platforms, including preliminary benchmark results.

3 Theory

The 4c-CIs, in Mulliken notation, are defined as

$$(ij|kl) = \iint \frac{\psi_i(\mathbf{r})\psi_j(\mathbf{r})\psi_k(\mathbf{r}')\psi_l(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}d\mathbf{r}', \tag{2}$$

where $\psi_i(\mathbf{r})$ are the molecular orbital (MOs). The MOs are expanded in the local basis set as

$$\psi_i(\mathbf{r}) = \sum_\mu c_{\mu i}\phi_\mu(\mathbf{r}) \tag{3}$$

where $c_{\mu i}$ are the MO coefficients and $\phi_\mu(\mathbf{r})$ are the atomic orbitals (AOs), which are in our case NAOs. Inserting Eq. (3) into Eq. (2) we obtain

$$(ij|kl) = \sum_{\mu\nu\sigma\tau} c_{\mu i}c_{\nu j}c_{\sigma k}c_{\tau l}(\mu\nu|\sigma\tau) \tag{4}$$

where the 4c-CIs are now computed in the AO representation as follows

$$(\mu\nu|\sigma\tau) = \iint \frac{\phi_\mu(\mathbf{r})\phi_\nu(\mathbf{r})\phi_\sigma(\mathbf{r}')\phi_\tau(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}d\mathbf{r}'. \tag{5}$$

The calculation of these integrals shows a scaling to the fourth power $\mathcal{O}(N^4)$ with the number of basis functions N.

The RI approach expands the product of two AOs, $\phi_\mu\phi_\nu$, in terms of a set of auxiliary basis functions (ABFs), $\{\varphi_P\}$, which are in our setup also NAOs

$$\rho_{\mu\nu}(\mathbf{r}) = \phi_\mu(\mathbf{r})\phi_\nu(\mathbf{r}) \approx \sum_P A^P_{\mu\nu}\varphi_P(\mathbf{r}) = \tilde{\rho}_{\mu\nu}(\mathbf{r}). \tag{6}$$

$A^P_{\mu\nu}$ denotes the RI expansion coefficients. Several methods to obtain $A^P_{\mu\nu}$ are available in the literature. In this work we use the Coulomb metric, which implies minimising the Coulomb repulsion of the density residual, $(\rho_{\mu\nu} - \tilde{\rho}_{\mu\nu}|\rho_{\mu\nu} - \tilde{\rho}_{\mu\nu})$, yielding[8,9,24,25]

$$A^P_{\mu\nu} = \sum_Q (\mu\nu|Q)V^{-1}_{QP} \tag{7}$$

where the three-center Coulomb integrals in Eq. (7) are defined as

$$(\mu\nu|Q) = \iint \frac{\phi_\mu(\mathbf{r})\phi_\nu(\mathbf{r})\varphi_Q(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}d\mathbf{r}' \tag{8}$$

and the two-center Coulomb integrals (2c-CI) are given by

$$V_{PQ} = \iint \frac{\varphi_P(\mathbf{r})\varphi_Q(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}d\mathbf{r}'. \tag{9}$$

Computing the expansion via Eq. (7) is known as the RI-V approach. Inserting Eq. (6) into Eq. (3) yields the expression

$$(\mu\nu|\sigma\tau)_{\text{RI-V}} \equiv \sum_{PQ} A^P_{\mu\nu}V_{PQ}A^Q_{\sigma\tau} = \sum_{PQ}(\mu\nu|P)V^{-1}_{PQ}(\sigma\tau|Q) \tag{10}$$

4 Methodology

4.1 Numerical Integration Techniques

For the 2c-CIs defined in Eq. (9), two strategies can be employed to compute the integrals numerically. The first strategy involves using logarithmic spherical Bessel transforms[26,27] to evaluate the integral in Fourier space. This approach is efficient because, in Fourier space, the integral simplifies to a one-dimensional integral over the radial part of the ABFs. This is the default way of integrating the 2c RI integrals in FHI-aims[28].

The other integration strategy employs atom-centered, spherical real-space grids. This method involves a two-step procedure: first, the Coulomb field $\Omega_P(\mathbf{r})$ of the auxiliary function $\varphi_P(\mathbf{r})$ is computed:

$$\Omega_P(\mathbf{r}) = \int \frac{\varphi_P(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}'. \tag{11}$$

The second step is to evaluate V_{PQ} as

$$V_{PQ} = \int \varphi_P(\mathbf{r})\Omega_Q(\mathbf{r}) d\mathbf{r} \tag{12}$$

$$= \sum_{\mathbf{r}} w(\mathbf{r})\varphi_P(\mathbf{r})\Omega_Q(\mathbf{r}). \tag{13}$$

Eq. (12) takes now the form of an overlap integral that can be discretised over the three-dimensional spatial grids as shown in Eq. (13). $w(\mathbf{r})$ is the weight of an integration grid point $\mathbf{r} = \mathbf{r}(a,s,t)$ that is uniquely determined by the atomic center a, the radial shell number s, and the angular point t. Both of these techniques for evaluating two-center integrals are computationally efficient, as the number of two-center integrals scales with $\mathcal{O}(N^2)$. The second strategy might be computationally slightly more expensive than the first one, but is commonly used by NAO codes in other contexts, such as the evaluation of the exchange correlation matrix in density functional theory (DFT)[29,30,11,31].

Turning now to the 3c-CIs, it is possible to solve the three-center integrals in Fourier space as well but the additional integration center leads to large multipole sums[32,33]. In FHI-aims the three-center integration was implemented by Ren *et al.*[28], building on the real-space grid integration (second strategy) for the 2c-CIs. The 3c-CIs are discretised over three-dimensional grids

$$(\mu\nu|P) = \int \phi_\mu(\mathbf{r})\phi_\nu(\mathbf{r})\Omega_P(\mathbf{r}) d\mathbf{r} \tag{14}$$

$$= \sum_{\mathbf{r}} w(\mathbf{r})\phi_\mu(\mathbf{r})\phi_\nu(\mathbf{r})\Omega_P(\mathbf{r}) = M_{\rho P}. \tag{15}$$

To simplify the computation of the 3c-CIs we combine the indices μ and ν into a single index ρ, resulting in the matrix $M_{\rho P}$, where ρ is the index of a unique pair formed from the primary AOs.

Unlike two-center integrals, the computation of three-center integrals is computationally expensive. Accelerating their computation significantly reduces the prefactor in calculations using the RI method. GPUs are a natural choice for this acceleration, as the large sums over real-space points in Eq. (15) are independent and can benefit greatly from the massive parallelism that GPUs provide.

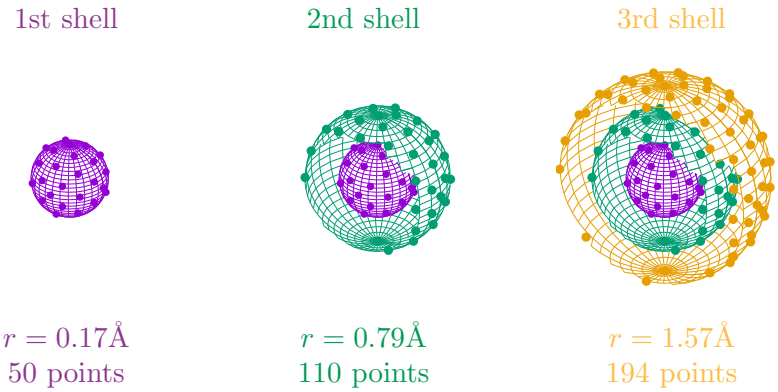


Figure 1. Construction of the atomic grids, with the innermost shell (violet) at $r = 0.17$ Å containing 50 points. The next shell (green) at $r = 0.79$ Å holds 110 points, and the third shell (yellow) at $r = 1.57$ Å contains 194 points.

4.2 Real Space Grid Generation

To perform the numerical integration of the 3c-CIs, we construct a discrete set of real-space grid points for each atom, defining a series of radial points, or shells, at different distances from the atoms' center. To arrange these shells, FHI-aims uses a logarithmic spacing, which places points more densely near the center and distributes them more widely as the distance increases[34]. The location of the radial shells is given by[34]

$$r(i) = r_{\text{outer}} \frac{\log\{1 - [i/(N_{\text{rad}} + 1)]^2\}}{\log\{1 - [N_{\text{rad}}/(N_{\text{rad}} + 1)]^2\}}. \tag{16}$$

r_{outer} is the outermost radius, which is set to 7.0 Å in the tight default settings of FHI-aims. N_{rad} is the number of radial shells. The value N_{rad} was empirically determined by Baker *et al.* and is dependent on the atomic number[34]. After the locations of the shells are determined, they are filled with angular points, starting from the innermost to the outermost shell. For the angular integration we use Lebedev grids[35]. Lebedev grids represent the quadrature on the surface of a unit sphere, featuring relatively simple point distributions with octahedral symmetry. In FHI-aims the angular grids with 50, 110, 194, 302, 434, 590, 770, 974 and 1202 points are available.

Fig. 1 illustrates the construction of the atomic grids, combining the radial point distribution and the angular grids: the first shell is filled with 50 angular points. The next shell employs an angular grid with 110 points, followed by another angular grid for the subsequent shell that contains 194 points. This process is repeated for each radial shell, systematically filling all shells with an increasingly larger number of angular points until the outermost shell is reached. For NAOs, using 434 points in the outermost shell typically offers a good balance between accuracy and computational time[36]. Generally, there are far more than three shells. For example, with the tight default settings, there would be a total of 69 shells for the carbon atom.

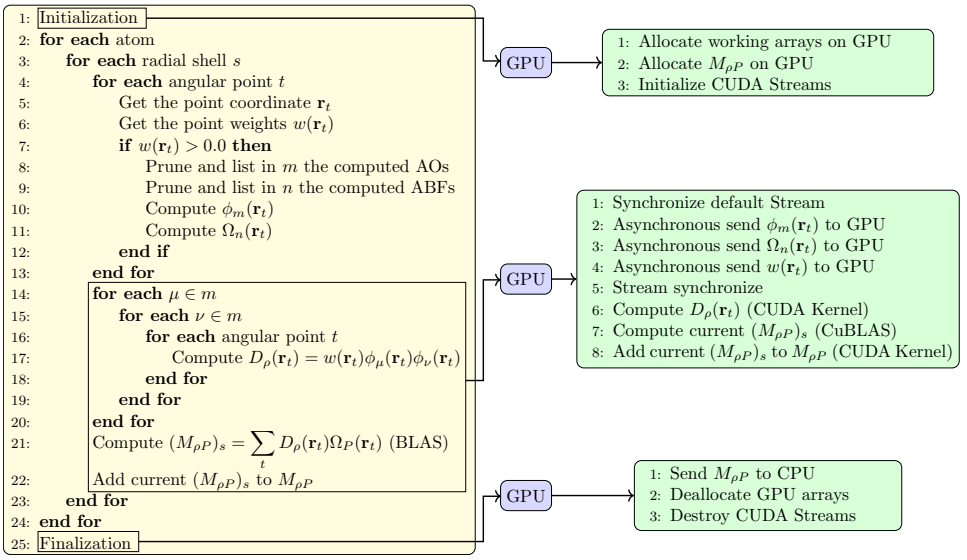


Figure 2. Pseudocode for the numerical integration of the 3c-CIs. The left panel (highlighted in yellow) shows the CPU operations, while the right panel (in green) displays the corresponding GPU operations.

5 Implementation Details

The pseudocode for the numerical integration of the 3c-CIs (Eq. (15)) is shown in Fig. 2 and it describes the operations performed in one CPU core. The 3c-CIs are parallelised via the message-passing-interface (MPI), distributing the set of ABFs $\{\varphi_P\}$ over the CPU cores. The numerical integration is performed sequentially for each atom a, and then for each radial shell s of a, following the same logic used in constructing the numerical grid shown in Fig. 1. For each angular point t, i.e., each point on the radial shell s, we retrieve the spatial coordinate $\mathbf{r}_t = (x_t, y_t, z_t)$ and the corresponding weight $w(\mathbf{r}_t)$, which is given by

$$w(\mathbf{r}) = p_3(\mathbf{r}, a) w_{\text{rad}}(s) w_{\text{ang}}(t), \tag{17}$$

where $w_{\text{rad}}(s)$ is the integration weight for the logarithmic radial grid and $w_{\text{ang}}(t)$ is the angular weight originating from the Lebedev grids. $p_3(\mathbf{r}, a)$ is the three-center partition function, which divides the full three-dimensional integrals into effective atom-by-atom components, as the integration shells overlap with each other.

If $w(\mathbf{r}_t) > 0$, meaning that the current point $\mathbf{r}_t$ contributes to the integral, a pruning step is performed. This step reduces the number of primary AO basis functions ϕ_μ and ABFs φ_P by checking, for each ϕ_μ and for each φ_P, whether their spatial extension includes $\mathbf{r}_t$. The indices of the contributing primary AO functions and ABFs are stored in lists m and n, respectively. Generally, the number of primary AOs and ABFs stored in n and m is smaller than the total number of primary and auxiliary functions available. For $\mathbf{r}_t$, we then tabulate all primary functions in list m and compute all Coulomb fields Ω_n for the ABFs in list n.

Next, we compute the density-matrix-like quantity $D_\rho(\mathbf{r}_t)$ as the product of two primary AO functions, multiplied by $w(\mathbf{r}_t)$. This is outlined in pseudocode steps 14–20, where ρ denotes the basis pair index, as defined in Eq. (15). $(M_{\rho P})_s$ is computed by contracting $D_\rho(\mathbf{r}_t)$ and $\Omega_P(\mathbf{r}_t)$ over the number of angular points t in each shell s. For this step, we employ the BLAS3 routines (`dgemm`) . Finally, each shell contribution to the 3c-CIs is added to the final quantity $M_{\rho P}$ as defined in Eq. (15). The described algorithm, corresponding to the left panel of Fig. 2, was previously implemented by Ren *et al.*[28] and served as starting point for this project.

The steps 14–20 in the left panel of Fig. 2 are the computational bottleneck in the evaluation of the 3c-CIs. To address this, we accelerated the computation of $(M_{\rho P})_s$ by leveraging GPUs. In the right panel of Fig. 2, highlighted in green, we present the pseudocode for the GPU implementation. Since the computation of the 3c-CIs also requires a substantial amount of memory, we base our implementation on CUDA to gain fine-grained control over the GPU, including memory management. Additionally, access to libraries such as cuBLAS and features like asynchronous execution provide greater flexibility for optimising computational performance.

The process begins by allocating the arrays $(M_{\rho P})_s$ and the full array $M_{\rho P}$ on the GPU, followed by the initialisation of the CUDA streams. These streams can be used for asynchronous execution of memory copies and kernel computations. Next, the weights $w(\mathbf{r}_t)$, primary basis functions $\phi_m(\mathbf{r}_t)$ and the Coulomb potential of the ABFs $\Omega_n(\mathbf{r}_t)$ are evaluated on the CPU and asynchronously transferred to the GPU. These asynchronous data transfers help mitigate latency, as memory operations are typically slow. By overlapping data transfers with computations - a technique known as latency hiding - we significantly improve overall performance. Using a CUDA kernel, we compute the $D_\rho(\mathbf{r}_t)$ elements and perform a matrix multiplication with $\Omega_P(\mathbf{r}_t)$ using the cuBLAS library. The s-shell contribution $(M_{\rho P})_s$ is then added to the full array $M_{\rho P}$. After collecting the contributions from each radial shell s, the GPU operations are finalised by transferring the array $M_{\rho P}$ back to the CPU and deallocating all arrays.

6 Hardware Considerations

We performed the validation and benchmark calculations on JUWELS Booster. Each JUWELS Booster node consists of 2 AMD EPYC Rome 7402 CPUs (48 cores in total) and 4 NVIDIA A100 GPUs. Each node contains 160 GB of total GPU memory and 512 GB of CPU RAM. As mentioned before, FHI-aims uses MPI for parallelisation, assigning one core to each task. Since a Booster node has more CPU cores than GPUs, multiple MPI tasks share a GPU. To enable this, we use NVIDIA's Multi-Process Service (MPS) for efficient GPU resource sharing across MPI tasks. Further speedup was achieved by binding tasks to specific CPU cores and configuring each task's `CUDA_VISIBLE_DEVICES` variable to target the GPU physically closest to the assigned core, minimising data transfer latency.

7 Benchmarks

To benchmark the computational performance of our implementation, we carried out hybrid DFT calculations using the PBE0 functional[37,38]. The computations were performed

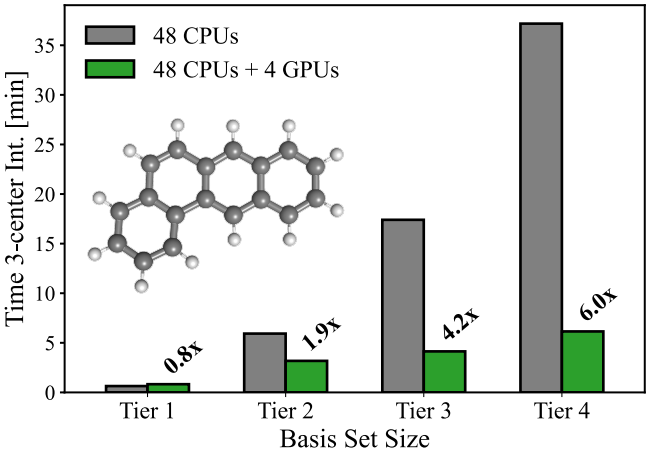


Figure 3. Walltime [min] for the evaluation of the 3c-CIs using CPU only (in gray) and CPU+GPU (in green) shown as a function of increasing basis size. The calculations were performed with the Tier 1, Tier 2, Tier 3, and Tier 4 NAO basis sets, using the benzanthracene molecule ($C_{18}H_{12}$) as test system, which consists of 30 atoms in total.

	Time		Three-center integrals		Fock
	CPU only	CPU+GPU	CPU only	CPU+GPU	exchange
Tier 1	0.98	1.14	0.64	0.83	0.03
Tier 2	8.30	6.07	5.93	3.18	2.05
Tier 3	28.73	15.71	17.40	4.14	9.75
Tier 4	68.43	37.60	37.18	6.15	26.73

Table 1. Total computational timings [min] for the PBE0 calculation of the benzanthracene molecule, including three-center Coulomb integrals and Fock exchange. Timings were obtained using the Tier 1, Tier 2, Tier 3 and Tier 4 basis set.

on one AMD EPYC Rome node, as detailed in Sec. 6. The benzanthracene molecule ($C_{18}H_{12}$), containing 30 atoms, was selected as benchmark system. Tab. 1 reports the total computational time (in minutes), along with the 3c-CI and Fock exchange timings as a function of increasing basis set size. Both CPU-only and CPU+GPU timings are reported, with GPU acceleration applied exclusively to the evaluation of the 3c-CIs. Considering the largest Tier 4 NAO basis set, Tab. 1 shows that the 3c-CIs account for 54% of the total time in the CPU-only implementation, while in the CPU+GPU implementation, they now contribute approximately 16% of the total time. Our CPU+GPU implementation removes the 3c-CIs as the computational bottleneck. The computationally most expensive step is now the calculation of the Fock exchange matrix, accounting for 71% of the total run time.

As reported in Tab. 1 and illustrated in Fig. 3, the GPU acceleration of the 3c-CI evaluation exhibits a distinct behaviour. With the small Tier 1 basis set, no acceleration is observed. However, as the basis set size increases, we achieve speedups of 1.8x, 3.5x, and

6.0x for the Tier 2, Tier 3, and Tier 4 basis sets, respectively. We attribute this behaviour to the fact that increasing the basis set size enlarges the arrays that the GPU is working on, thereby improving GPU utilisation. Correlated methods are known to converge slowly with respect to basis set size[39] and typically require basis sets larger than Tier 2 to achieve convergence[36, 39]. This aspect is highly advantageous for our CPU+GPU implementation, significantly accelerating production-run calculations.

8 Concluding Remarks

In this article, we presented a GPU implementation for computing 3c-CIs over NAOs in the FHI-aims program package, utilising low-level CUDA APIs. We demonstrated that our implementation can accelerate the evaluation by up to a factor of six when comparing CPU-only to CPU+GPU execution times. Additionally, we observed that the speed-up generally increases with larger basis sets. This represents a significant reduction in the time required for electronic structure calculations at the hybrid DFT and beyond-DFT levels, enabling computations on larger systems. Additionally, the 3c-CIs are central to the low-scaling RPA and GW algorithms currently under development in FHI-aims[40]. Our work paves the way for a full GPU implementation of these algorithms.

Acknowledgements

The authors acknowledge funding by the Emmy Noether Program of the German Research Foundation (Project No. 453275048). Furthermore, this project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 101118139. The JU receives support from the European Union's Horizon Europe Programme. The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) for funding this project by providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS Supercomputer JUWELS[41] at Jülich Supercomputing Centre (JSC).

References

1. A. D. Becke, *A new mixing of Hartree–Fock and local density–functional theories*, J. Chem. Phys., **98**, 1372-1377, 1993.
2. C. Møller and M. S. Plesset, *Note on an approximation treatment for many-electron systems*, Phys. Rev., **46**, 618-622, 1934.
3. F. Coester and H. Kümmel, *Short-range correlations in nuclear wave functions*, Nucl. Phys., **17**, 477-485, 1960.
4. J. Čížek, *On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods*, J. Chem. Phys., **45**, 4256-4266, 1966.
5. D. Bohm and D. Pines, *A Collective Description of Electron Interactions: III. Coulomb Interactions in a Degenerate Electron Gas*, Phys. Rev., **92**, 609-625, 1953.

6. M. Gell-Mann and K. A. Brueckner, *Correlation Energy of an Electron Gas at High Density*, Phys. Rev., **106**, 364-368, 1957.
7. L. Hedin, *New Method for Calculating the One-Particle Green's Function with Application to the Electron-Gas Problem*, Phys. Rev., **139**, A796-A823, 1965.
8. J. L. Whitten, *Coulombic potential energy integrals and approximations*, J. Chem. Phys., **58**, 4496-4501, 1973.
9. B. I. Dunlap, J. W. D. Connolly, and J. R. Sabin, *On some approximations in applications of Xα theory*, J. Chem. Phys., **71**, 3396-3402, 1979.
10. J. Wilhelm, P. Seewald, and D. Golze, *Low-Scaling GW with Benchmark Accuracy and Application to Phosphorene Nanosheets*, J. Chem. Theory Comput., **17**, 1662-1677, 2021.
11. V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter, and M. Scheffler, *Ab initio molecular simulations with numeric atom-centered orbitals*, Comput. Phys. Commun., **180**, 2175-2196, 2009.
12. X. Qin, J. Chen, Z. Luo, L. Wan, J. Li, S. Jiao, Z. Zhang, Q. Jiang, W. Hu, H. An, and J. Yang, *High performance computing for first-principles Kohn-Sham density functional theory towards exascale supercomputers*, CCF Trans. High Perform. Comput., **5**, 26-42, 2023.
13. P. Li, X. Liu, M. Chen, P. Lin, X. Ren, L. Lin, C. Yang, and L. He, *Large-scale ab initio simulations based on systematically improvable atomic basis*, Comput. Mater. Sci., **112**, 503-517, 2016.
14. T. Ozaki, *Variationally optimized atomic orbitals for large-scale electronic structures*, Phys. Rev. B, **67**, 155108, 2003.
15. J. M. Soler, E. Artacho, J. D. Gale, A. García, J. Junquera, P. Ordejón, and D. Sánchez-Portal, *The SIESTA method for ab initio order-N materials simulation*, J. Phys. Condens. Matter, **14**, 2745, 2002.
16. B. Delley, *From molecules to solids with the DMol3 approach*, J. Chem. Phys., **113**, 7756-7764, 2000.
17. K. Koepernik and H. Eschrig, *Full-potential nonorthogonal local-orbital minimum-basis band-structure scheme*, Phys. Rev. B, **59**, 1743-1757, 1999.
18. S. R. Jensen, S. Saha, J. A. Flores-Livas, W. Huhn, V. Blum, S. Goedecker, and L. Frediani, *The Elephant in the Room of Density Functional Theory Calculations*, J. Phys. Chem. Lett., **8**, 1449-1457, 2017.
19. S. Obara and A. Saika, *Efficient recursive computation of molecular integrals over Cartesian Gaussian functions*, J. Chem. Phys., **84**, 3963-3974, 1986.
20. R. Ahlrichs, *A simple algebraic derivation of the Obara-Saika scheme for general two-electron interaction potentials*, Phys. Chem. Chem. Phys., **8**, 3072-3077, 2006.
21. D. Golze, N. Benedikter, M. Iannuzzi, J. Wilhelm, and J. Hutter, *Fast evaluation of solid harmonic Gaussian integrals for local resolution-of-the-identity methods and range-separated hybrid functionals*, J. Chem. Phys., **146**, 034105, 2017.
22. D. Golze, J. Wilhelm, M. J. van Setten, and P. Rinke, *Core-Level Binding Energies from GW: An Efficient Full-Frequency Approach within a Localized Basis*, J. Chem. Theory Comput., **14**, 4856-4869, 2018.
23. R. L. Panadés-Barrueta and D. Golze, *Accelerating Core-Level GW Calculations by Combining the Contour Deformation Approach with the Analytic Continuation of W*, J. Chem. Theory Comput., **19**, 5450-5464, 2023.

24. J. W. Mintmire, J. R. Sabin, and S. B. Trickey, *Local-density-functional methods in two-dimensionally periodic systems. Hydrogen and beryllium monolayers*, Phys. Rev. B, **26**, 1743-1753, 1982.
25. O. Vahtras, J. Almlöf, and M. W. Feyereisen, *Integral approximations for LCAO-SCF calculations*, Chem. Phys. Lett., **213**, 514-518, 1993.
26. J. D. Talman, *Numerical calculation of four-center Coulomb integrals*, J. Chem. Phys., **80**, 2000-2008, 1984.
27. J. D. Talman, *Numerical methods for multicenter integrals for numerically defined basis functions applied in molecular calculations*, Int. J. Quantum Chem., **93**, 72-90, 2003.
28. X. Ren, P. Rinke, V. Blum, J. Wieferink, A. Tkatchenko, A. Sanfilippo, K. Reuter, and M. Scheffler, *Resolution-of-identity approach to Hartree–Fock, hybrid density functionals, RPA, MP2 and GW with numeric atom-centered orbital basis functions*, New J. Phys., **14**, 053020, 2012.
29. A. D. Becke, *A multicenter numerical integration scheme for polyatomic molecules*, J. Chem. Phys., **88**, 2547-2553, 1988.
30. B. Delley, *An all-electron numerical method for solving the local density functional for polyatomic molecules*, J. Chem. Phys., **92**, 508-517, 1990.
31. V. Havu, V. Blum, P. Havu, and M. Scheffler, *Efficient O (N) integration for all-electron electronic structure calculation using numeric basis functions*, J. Comp. Phys., **228**, 8367-8379, 2009.
32. J. D. Talman, *Multipole expansions for numerical orbital products*, Int. J. Quantum Chem., **107**, 1578-1584, 2007.
33. J. D. Talman, *Multipole expansions of orbital products about an intermediate center*, Int. J. Quantum Chem., **111**, 2221-2227, 2011.
34. J. Baker, J. Andzelm, A. Scheiner, and B. Delley, *The effect of grid quality and weight derivatives in density functional calculations*, J. Chem. Phys., **101**, 8894-8902, 1994.
35. V. I. Lebedev and D. N. Laikov, *A quadrature formula for the sphere of the 131st algebraic order of accuracy*, Dokl. Math., **59**, 477-481, 1999.
36. I. Y. Zhang, X. Ren, P. Rinke, V. Blum, and M. Scheffler, *Numeric atom-centered-orbital basis sets with valence-correlation consistency from H to Ar*, New J. Phys., **15**, 123033, 2013.
37. C. Adamo and V. Barone, *Toward reliable density functional methods without adjustable parameters: The PBE0 model*, J. Chem. Phys., **110**, 6158-6170, 1999.
38. M. Ernzerhof and G. E. Scuseria, *Assessment of the Perdew-Burke-Ernzerhof exchange-correlation functional*, J. Chem. Phys., **110**, 5029-5036, 1999.
39. D. Golze, M. Dvorak, and P. Rinke, *The GW compendium: A practical guide to theoretical photoemission spectroscopy*, Front. Chem., **7**, 377, 2019.
40. F. A. Delesma, M. Leucke, D. Golze, and P. Rinke, *Benchmarking the accuracy of the separable resolution of the identity approach for correlated methods in the numeric atom-centered orbitals framework*, J. Chem. Phys., **160**, 024118, 2024.
41. Jülich Supercomputing Centre, *JUWELS Cluster and Booster: Exascale Pathfinder with Modular Supercomputing Architecture at Juelich Supercomputing Centre*, Journal of large-scale research facilities, **7**, no. A138, 2021.

Machine Learning for Accelerated Discovery and Design of Functional Energy Materials

Mohammad J. Eslamibidgoli[1,2], Max Dreger[1,2], Andre Colliard-Granero[1,2], Fabian Tipp[1,2], Michael H. Eikerling[1,2], and Kourosh Malek[1,2]

[1] Institute of Energy Technologies, Theory and Computation of Energy Materials (IET-3), Forschungszentrum Jülich, 52425 Jülich, Germany
E-mail: {m.eslamibidgoli, m.eikerling, k.malek}@fz-juelich.de

[2] Centre for Advanced Simulation and Analytics (CASA), Simulation and Data Science Lab for Energy Materials (SDL-EM), Forschungszentrum Jülich GmbH, 52425 Jülich, Germany

We employ data-driven methodologies and machine learning with materials science approaches to accelerate the development of sustainable energy technologies. Our focus covers the entire materials workflow, from modelling and simulations for discovery, correlative diagnostics and property prediction, and inverse molecular design, to the extraction, management, and analytics of materials science data, as well as automated image analysis. Challenges such as diverse data types, limited training sets, and the complex, multi-scale nature of materials demand a synergy between data representation and management, machine learning models and domain expertise. We utilise atomistic simulations using density functional theory (DFT) calculations, integrated with artificial intelligence (AI), to efficiently screen the parametric space governing the life-cycle performance of anion exchange membrane materials. Our focus is on optimising these materials for enhanced stability and transport properties. Additionally, we are developing deep learning-based techniques to automate image analysis and characterisation across various applications in functional energy materials. Furthermore, we introduce the development of native graph databases by integrating standardised materials ontologies with knowledge graphs, facilitating flexible representations of the materials-to-device workflow.

1 Introduction

Clean energy technologies, such as fuel cells, hydrogen storage devices and solar technologies, heavily depend on energy materials that exhibit high performance in terms of activity and stability[1]. The accelerated development of such materials plays a pivotal role in driving the sustainable technologies forward. Despite their critical importance, the process of developing new materials from laboratory to the marketplace remains significantly lengthy, often taking from 10 to 20 years[2]. This prolonged process is mainly related to the multi-component nature of the devices and the complexity of the multi-step processes required to ensure the materials meet the cost, performance and scalability needs.

Traditional materials development approach often involves a sequential and largely empirical workflow which generally includes four primary steps: experiment planning based on chemical intuition, synthesis and characterisation of materials, data analytics to assess performance and lifetime, followed by an iterative process of repeating these steps to optimise materials properties[3]. As this approach is inherently slow and resource-intensive, researchers are increasingly turning to data-driven approaches, particularly those involving artificial intelligence (AI) and machine learning (ML), as means of accelerating discovery, design and integration of functional energy materials[4].

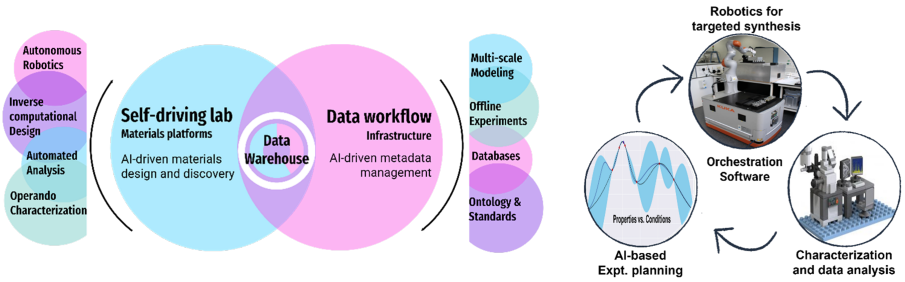


Figure 1. Data workflow and various in-line and off-line data sources involved in self-driving and semi-automated labs; autonomous optimisation workflow in self-driving labs.

In particular a shift has occurred toward the use of Materials Acceleration Platforms (MAPs). These platforms leverage ML and AI and high-throughput experimentation techniques to automate various steps of the traditional workflow[5]. MAPs utilise AI-driven models to rapidly explore the large and complex chemical spaces of materials and generate predictions about material properties without the need for extensive experimental work. For instance, autonomous robotic systems have been developed to assist in exploring chemical spaces for materials such as hydrogen production photocatalysts[6]. Such platforms have been also established to optimise thin films' optical and electronic properties, which are critical components in devices like solar cells and sensors[7]. Additionally, they can perform high-throughput characterisation for advanced materials[8].

Despite the significant potential of MAPs, several challenges remain. Current design of the workflow orchestration software is hardware-centric and targeted at the actual instruments[9]. This causes bottlenecks between theory and experimentation and a completely overlooked problem of automatic data analysis and data lineage tracking. An ideal data workflow should be capable of ingesting data from both in-line and off-line experiments, modelling, and simulations. Storage and retrieval of the integrated data from various sources, as depicted in Fig. 1, enable an effective development of self-serve online analytical tools for automated data analysis, improving the ability to connect to the analytical tools and the responsiveness to semantic and integrated queries, and the data access performance, all features that lie beyond the capabilities of current disjoint, heterogenous and often transaction-oriented databases and data infrastructures in the energy materials domain.

Another major challenge lies in the optimisation algorithms that guide materials design. These algorithms need to be scalable and robust enough to manage the noise and variability inherent in experimental conditions. Furthermore, materials design often requires multi-objective optimisation, where several competing properties – such as stability, conductivity, and cost – must be balanced. Developing algorithms capable of efficiently navigating this complex, multi-dimensional design space remains a significant hurdle[10].

This contribution addresses three key areas in the development of functional energy materials. First, we focus on creating a flexible data extraction and management system, leveraging ontology and graph databases to enable more efficient data handling and integration across the materials-to-device workflow. Second, we employ computer vision

and deep learning techniques for automated characterisation and image analysis, aiming to streamline and enhance the accuracy of material characterisation processes. Third, we integrate high-throughput density functional theory (DFT) simulations with inverse molecular design to accelerate the development of anion exchange membrane materials, optimising their stability and transport properties for energy applications.

2 Synergising Ontologies and Graph Databases for Highly Flexible Materials-to-Device Workflow Representations

In recent years, the increasing demand for sustainable energy solutions has led to a significant focus on the development of advanced materials, particularly those used in energy systems. The extraction and management of data related to these materials are crucial for optimising their properties and performance[11]. However, the field faces challenges due to the heterogeneity of data sources and the lack of standardised methods for data representation. To address these challenges, a flexible data extraction and management system is necessary, one that can handle the complexities of energy materials and their associated data.

2.1 Objective 1: Extension of the European Materials Modelling Ontology (EMMO) for Standardised Data Representation

The European Materials Modelling Ontology (EMMO) provides a structured framework for categorising and defining concepts within the materials science domain. However, its current structure may not fully encompass the specific needs of energy materials research[12]. Therefore, extending the EMMO to include these specialised concepts is essential for achieving standardised data representations across various research projects[13]. This extension will facilitate interoperability between different data sources and research institutions, enabling more efficient data sharing and collaboration.

In a previous publication, we demonstrated the utility of ontologies in enhancing data management in materials science, particularly through the integration of graph databases (see Fig. 2). Our approach focused on transforming non-standardised tabular data into knowledge graphs, which adhere to a defined ontology. This method not only improves data accessibility but also ensures that the data is semantically enriched, allowing for more sophisticated queries and analyses[14].

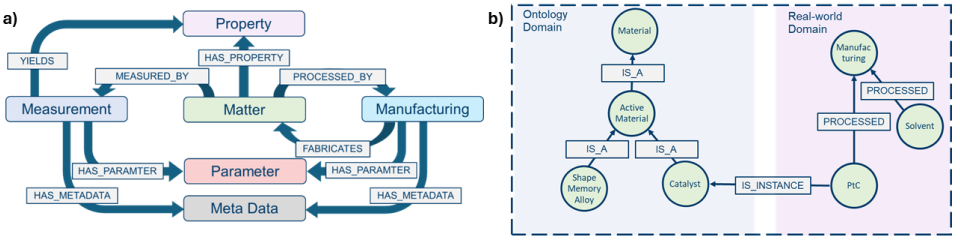


Figure 2. Schematic overview of the graph datamodel (a) and the node labelling system[14].

2.2 Objective 2: Creation of a New Graph Data Model for Workflows, Measurements, and Simulation Data

Energy materials research often involves complex workflows that integrate experimental measurements with simulation data. To manage this complexity, we propose the creation of a new graph data model that captures the relationships between different elements of the research process, such as materials, measurements, and simulation results[15]. This model will be designed to accommodate the diverse nature of energy materials research, ensuring that it is flexible enough to handle various types of data and workflows.

Our work on knowledge graph extraction from tabular data has laid the groundwork for this objective. By utilising large language models (LLMs) and rule-based feedback loops, we developed a semi-automated pipeline that transforms R&D tables into connected knowledge graphs. These graphs are highly adaptable, capable of representing the intricate relationships between different data points in a research workflow. The use of LLMs enables the system to infer context and meaning from data, ensuring that the resulting graphs are both accurate and meaningful.

2.3 Objective 3: Use of Neo4j Database and Django Framework for Cloud-Based Application Integration

To support the integration of the new graph data model into cloud-based applications, we plan to utilise the Neo4j graph database in conjunction with the Django framework. Neo4j's capabilities in handling complex relationships and its compatibility with various ontologies make it an ideal choice for this project. The Django framework will provide a robust platform for developing user interfaces and managing interactions with the database.

In our previous work, we demonstrated the effectiveness of integrating graph databases with semantic search capabilities within a Django application[15]. This integration allows for intuitive data management, enabling researchers to store, retrieve, and analyse data with ease. By building on this foundation, we aim to create a cloud-based system that facilitates the management of energy materials data, supporting research efforts across multiple institutions and projects.

In conclusion, the development of a flexible data extraction and management system for energy materials is a critical step towards advancing research in this field. By extending the EMMO, creating a new graph data model, and leveraging the capabilities of Neo4j and Django, we can provide a standardised, interoperable, and scalable solution that meets the unique needs of energy materials research. This system will not only enhance data accessibility and usability but also pave the way for new discoveries in the development of sustainable energy solutions.

3 Computer Vision and Deep Learning for Material Characterisation

The recent advances in computer vision and deep learning techniques have opened new horizons in the field of material characterisation[16]. These techniques allow for the automated analysis of large, complex imaging datasets, facilitating the study of intricate material structures and behaviours which otherwise would remain inaccessible[17].

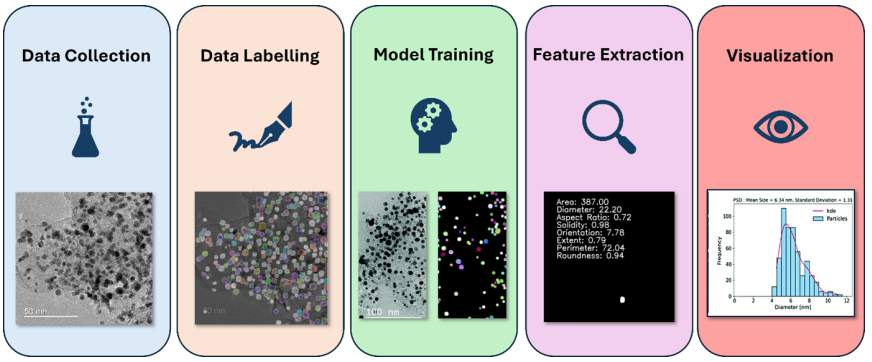


Figure 3. Schematic workflow employed for the development of deep learning workflows in the field of image analysis. The experimentalist collects the data, then the regions of interest are annotated. The dataset is employed to train DL models for prediction on unseen data. Finally, using computer vision, the features of interest are extracted and plotted properly for visualisation[20,22,23].

3.1 High-Throughput Analysis of Particle Size Distribution from Transmission Electron Microscopy (TEM) Images:

In the realm of nanomaterials, understanding particle size distribution is crucial for optimising the performance and durability of various materials. Especially, the analysis of platinum nanoparticles on a carbon support for PEMFC applications is sensitive to those parameters[18,19]. Deep learning methods, particularly CNNs, have been employed to automate the analysis of TEM images for this purpose. A notable application is the use of the StarDist model[20,21], which employs a U-Net architecture to perform instance segmentation of nanoparticles. This model has been effectively used to analyse high-resolution TEM images, providing accurate particle size distribution data with minimal human intervention. The ability of deep learning models to handle overlapping particles and varied shapes significantly improves the reliability of the analysis. Furthermore, in this work, a workflow for the automatic extraction of features of interest from the segmented images was developed. This allows for rapid characterisation of advanced metrics, such as morphological analysis of individual particles and their distributions.

3.2 Screening of Catalyst Layers and Ink Structural Characterisation for Polymer Electrolyte Fuel Cells:

The performance of PEMFCs is heavily influenced by the microstructure of catalyst layers, which are often formed from catalyst inks. Deep learning techniques have been applied to the high-throughput screening of these catalyst layers, using TEM images to identify and characterise structural features. Convolutional neural networks have also shown exceptional capability in distinguishing between different structural components of catalyst layers, classifying inks based on visual clues critical to the efficiency of fuel cells. By automating the screening process, these deep learning models facilitate the rapid evaluation and optimisation of catalyst materials, thereby accelerating the development of more efficient and cost-effective fuel cells[22,23].

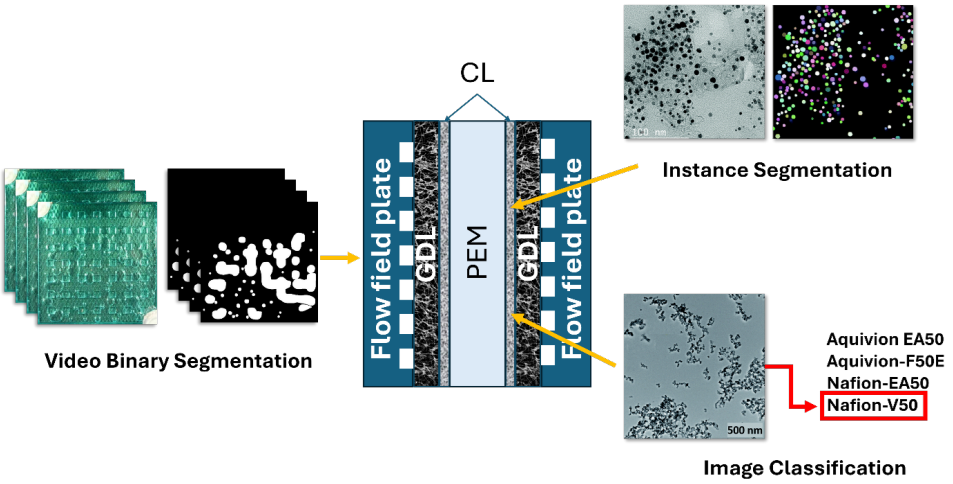


Figure 4. Schematic depiction of a PEMFC and the three different tools developed in this project: bubble dynamics analysis in videos (left), instance segmentation of nanoparticles in the catalyst layer (upper right), and catalyst ink image classification (bottom right)[24].

3.3 Automated Analysis of Bubble Dynamics in Proton Exchange Membrane Water Electrolysers:

In proton exchange membrane water electrolysers (PEMWEs), the dynamics of gas bubbles play a crucial role in determining overall system efficiency. The formation, growth, and detachment of bubbles can significantly impact the two-phase flow dynamics within the electrolyser. Deep learning models have been employed to analyse bubble dynamics using optical imaging data. These models are capable of segmenting and, with computer vision algorithms, tracking bubbles over time, providing detailed insights into bubble size, distribution, and behaviour under different operating conditions. Such automated analysis not only enhances our understanding of bubble dynamics but also aids in the optimisation of electrolyser design and operation, leading to improved performance[24].

The application of deep learning techniques in material characterisation has led to substantial improvements in both the performance, velocity, and reliability of segmentation models and the accuracy of characterisation outcomes. The ability to process large volumes of imaging data quickly and accurately enables researchers to obtain insights that were previously difficult to obtain from non-AI approaches. The implementation of deep learning workflows has demonstrated significant reductions in analysis time while simultaneously increasing the precision of measurements, such as particle size and morphological distributions, bubble dynamics, and ink classification. These advancements highlight the transformative potential of deep learning in the field of material science, paving the way for more efficient and insightful research methodologies.

These innovations provide a robust foundation for future developments, as the integration of artificial intelligence into material characterisation continues to evolve, offering new tools and techniques for understanding and optimising material properties at the macro, micro, and nanoscale.

4 Computational Exploration and Design of Electrochemical Materials

In the transition towards a green hydrogen economy, anion exchange membranes (AEMs) are emerging as a promising technology. Fuel cells and water electrolysers based around AEMs operate in a strongly alkaline environment and may function without the need for costly and rare platinum group metals (PGMs), which is crucial for the widespread adoption needed for a green hydrogen-based economy[25,26] However, currently a mayor challenge in the utilisation of AEMs is the significant degradation of many AEM materials in the harsh alkaline environments they are exposed to in application[27].

An AEM is generally constructed from a polymer backbone to give structural support and an organic cationic moiety to facilitate anion conductivity. A diverse set of structures for the cationic moieties have been studied in the literature with those based on imidazolium being especially prominent. The imidazolium structure can be augmented easily by adding various substituents to the five atoms making up the imidazolium heterocycle, with the resulting chemical structure having a significant impact on the molecule's resistance towards alkaline degradation[28].

In this project, the extensive computational capabilities of the JUWELS supercomputer are aimed towards the discovery of imidazolium-based molecules that possess a high alkaline stability. To achieve this goal, a multi-step approach is taken, which is summarised in the following:

1. Identify a computational descriptor of a given molecule's alkaline stability.
2. Automate the computation of the identified stability descriptor and apply it to a molecular dataset of a few thousand compounds.
3. Identify promising compounds from the dataset and test them experimentally to verify the found descriptor.
4. Train suitable machine learning models on the computed compounds to further accelerate the search for novel alkaline stable molecules.

To identify a reliable computational descriptor for the alkaline stability of imidazolium-based compounds, several molecules were identified for which experimental alkaline stability measurements have been performed in literature. For these compounds, the main degradation pathways were modelled at the Density Functional Theory (DFT) and Coupled Cluster (CC) level. Through comparison of the computed degradation energetic and experimental stability data, it could be identified that the free energy change of a hydroxide attack on the imidazolium ring is a reliable descriptor of alkaline stability[29].

In order to gain insights into the structure-stability relationship of imidazolium-based compounds, the descriptor was computed for a library of about 5800 structures. To achieve this, the computation of the stability descriptor was automated, and a diverse set of imidazolium-based structures were systematically generated. The JUWELS supercomputer[30] was then utilised to perform the stability prediction for all generated structures. From the generated dataset, diverse insights could be gained into the factors contributing to a high alkaline stability. Additionally, promising structures could be identified from the dataset and a selection of five compounds were chosen to be synthesised in the lab and

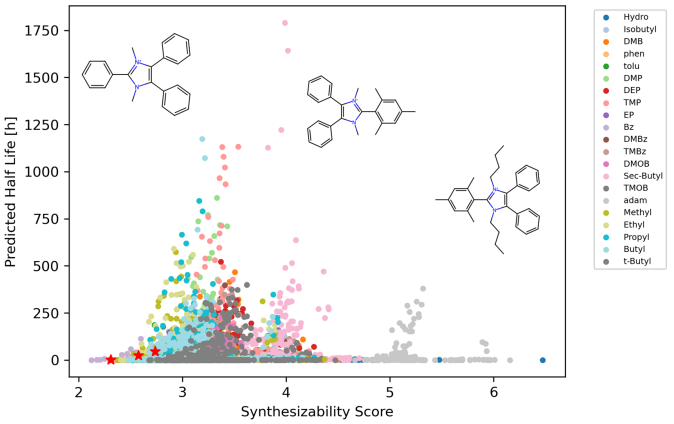


Figure 5. The predicted half life of imidazolium-based compounds for various substituents estimated through automated DFT calculations vs. their synthesisability score[29].

their alkaline stability was measured experimentally. The experimentally measured stability was shown to be in good agreement with the stability predicted with the computational descriptor and an especially stable compound was found.

The extensive molecular library that was generated is also sufficiently large to effectively train machine learning models to generate the stability descriptor, with graph neural networks being especially suitable for this task. Different machine learning models are being trained at the moment, reaching accuracies of about 1 kcal/mol while being multiple orders of magnitude faster than performing DFT calculations. Additionally, the dataset is ideal to benchmark various active learning approaches.

5 Concluding Remarks

In conclusion, our work integrates cutting-edge methodologies in data management, deep learning, and computational modelling to accelerate the discovery and optimisation of sustainable energy materials. We have extended the European Materials Modelling Ontology and developed a flexible graph data model for managing diverse materials workflows, enabling more efficient data representation and analysis. Additionally, the application of deep learning techniques has significantly advanced high-throughput image analysis, with applications ranging from particle size distribution in nanomaterials to catalyst layer characterisation and bubble dynamics in electrochemical systems. These innovations have not only streamlined material characterisation but also enhanced precision and scalability. Moreover, our computational pipeline for the design of alkaline-stable imidazolium-based compounds has accelerated the search for novel anion exchange membranes by leveraging density functional theory calculations and machine learning models. Together, these contributions demonstrate the potential of integrating ontologies, deep learning, and computational simulations to revolutionise materials discovery and accelerate the transition towards more efficient, sustainable energy technologies.

Acknowledgements

The authors acknowledge the Gauss Centre for Supercomputing e.V. (https://www.gauss-centre.eu) for funding this project by providing computing time on the GCS Supercomputer JUWELS at Jülich Supercomputing Centre (JSC) of Forschungszentrum Jülich[30].The authors also acknowledge the financial support from the Helmholtz Imaging Platform HIP (a platform of the Helmholtz Incubator on Information and Data Science) grant number DB002249, as well as the Federal Ministry of Science and Education (BMBF) under the German-Canadian Materials Acceleration Centre (GC-MAC) grant number 01DM21001A.

References

1. M. J. Eslamibidgoli, J. Huang, T. Kadyk, A. Malek, and M. Eikerling, *How theory and simulation can drive fuel cell electrocatalysis*, Nano Energy **29**, 334-361, 2016.
2. G. M. Whitesides and G. W. Crabtree, *Don't forget long-term fundamental research in energy*, Science **315 (5813)**, 796-798, 2007.
3. R. Pollice, G. dos Passos Gomes, M. Aldeghi, R. J. Hickman, M. Krenn, C. Lavigne, M. Lindner-D'Addario et al., *Data-driven strategies for accelerated materials design*, Accounts of Chemical Research **54 (4)**, 849-860, 2021.
4. M. Abolhasani and E. Kumacheva, *The rise of self-driving labs in chemical and materials sciences*, Nature Synthesis **2 (6)**, 483-492, 2023.
5. Z. Yao, Y. Lum, A. Johnston, L. M. Mejia-Mendoza, X. Zhou, Y. Wen, A. Aspuru-Guzik, E. H. Sargent, and Z. W. Seh, *Machine learning for a sustainable energy future*, Nature Reviews Materials **8 (3)**, 202-215, 2023.
6. E. Stach, B. DeCost, A. G. Kusne, J. Hattrick-Simpers, K. A. Brown, K. G. Reyes, J. Schrier et al., *Autonomous experimentation systems for materials development: A community perspective*, Matter **4 (9)**, 2702-2726, 2021.
7. M. Umehara, H. S. Stein, D. Guevarra, P. F. Newhouse, D. A. Boyd, and J. M. Gregoire, *Analyzing machine learning models to accelerate generation of fundamental materials insights*, npj Computational Materials **5 (1)**, 1-9, 2019.
8. G. Tom, S. P. Schmid, S. G. Baird, Y. Cao, K. Darvish, H. Hao, S. Lo et al., *Self-driving laboratories for chemistry and materials science*, Chemical Reviews **124 (16)**, 9633-9732, 2024.
9. F. Häse, M. Aldeghi, R. J. Hickman, L. M. Roch, M. Christensen, E. Liles, J. E. Hein, and A. Aspuru-Guzik, *Olympus: a benchmarking framework for noisy optimization and experiment planning*, Machine Learning: Science and Technology **2 (3)**, 035021, 2021.
10. F. Biscani and D. Izzo, *A parallel global multiobjective framework for optimization: pagmo*, Journal of Open Source Software **5 (53)**, 2338, 2020.
11. J. Kimmig, S. Zechel, and U. S. Schubert, *Digital transformation in materials science: A paradigm change in material's development*, Advanced Materials **33 (8)**, 2004940, 2021.
12. The Elementary Multiperspective Material Ontology (EMMO), https://emmo-repo.github.io/.
13. S. Clark et al., *Toward a unified description of battery data*, Advanced Energy Materials **12 (17)**, 2102702, 2022.

14. M. Dreger, M. J. Eslamibidgoli, and K. Malek, *Synergizing ontologies and graph databases for highly flexible materials-to-device workflow representations*, Journal of Materials Informatics **3**, 2, 2023.
15. D. Mrdjenovich et al., *Propnet: a knowledge graph for materials science*, Matter **2 (2)**, 464-480, 2020.
16. K. He, X. Zhang, S. Ren, and J. Sun, *Deep Residual Learning for Image Recognition*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition **1**, 770-778, 2016.
17. J.-N. Rouzaud and C. Clinard, *Characterization of carbon materials by surface area and pore size distribution analysis*, Fuel Processing Technology **77**, 229-235, 2002.
18. R. L. Borup, A. Kusoglu, K. C. Neyerlin, R. Mukundan, R. K. Ahluwalia, D. A. Cullen, K. L. More, A. Z. Weber, and D. J. Myers, *Scientific challenges for the next generation of fuel cells*, Current Opinion in Electrochemistry **21**, 192-200, 2020.
19. P. Jovanovič, A. Pavlišič, V. S. Šelih, M. Šala, N. Hodnik, M. Bele, S. Hočevar, and M. Gaberšček, *Bimetallic catalysts for selective hydrogenation reactions*, ChemCatChem **6**, 449-455, 2014.
20. A. Colliard-Granero, J. Jitsev, M. H. Eikerling, K. Malek, and M. J. Eslamibidgoli, *UTILE-Gen: Automated image analysis in nanoscience using synthetic dataset generator and deep learning*, ACS Nanoscience Au **3 (5)**, 398-407, 2023.
21. U. Schmidt, M. Weigert, C. Broaddus, and G. Myers, *Cell detection with nuclei segmentation in histopathological images*, International Conference on Medical Image Computing and Computer-Assisted Intervention **1**, 265-273, 2018.
22. A. Colliard-Granero, M. Batool, J. Jankovic, J. Jitsev, M. H. Eikerling, K. Malek, and M. J. Eslamibidgoli, *Deep learning for the automation of particle analysis in catalyst layers for polymer electrolyte fuel cells*, Nanoscale **14 (1)**, 10-18, 2022.
23. M. J. Eslamibidgoli, F. P. Tipp, J. Jitsev, J. Jankovic, M. H. Eikerling, and K. Malek, *Convolutional neural networks for high throughput screening of catalyst layer inks for polymer electrolyte fuel cells*, RSC Advances **11 (51)**, 32126-32134, 2021.
24. A. Colliard-Granero, K. A. Gompou, C. Rodenbücher, K. Malek, M. H. Eikerling, and M. J. Eslamibidgoli, *Deep learning-enhanced characterization of bubble dynamics in proton exchange membrane water electrolyzers*, Physical Chemistry Chemical Physics **26 (20)**, 14529-14537, 2024.
25. G. Das, *Anion exchange membranes for fuel cell application: a review*, Polymers **14 (6)**, 1197, 2022.
26. N. Du et al., *Anion-exchange membrane water electrolyzers*, Chemical reviews **122 (13)**, 11830-11895, 2022.
27. W. You, K. J. T. Noonan, and G. W. Coates, *Alkaline-stable anion exchange membranes: A review of synthetic approaches*, Progress in Polymer Science **100**, 101177, 2020.
28. U. Salma and N. Shalahin *A mini-review on alkaline stability of imidazolium cations and imidazolium-based anion exchange membranes*, Results in Materials **17**, 100366, 2023.
29. F. P. Tipp et al., *Stability Descriptors for (Benz) imidazolium-Based Anion Exchange Membranes*, Macromolecules **57 (4)**, 1734-1743, 2024.
30. D. Krause, *JUWELS: Modular Tier-0/1 supercomputer at the Jülich supercomputing centre*, Journal of Large-Scale Research Facilities (JLSRF), **5**, A135-A135, 2019.

Elementary Particle Physics

Elementary Particle Physics

Szabolcs Borsányi

Department of Physics, University of Wuppertal, Gaußstr 20, 42119 Wuppertal, Germany
E-mail: borsanyi@uni-wuppertal.de

One of the most enduring yet imperfect theories in high-energy physics is the Standard Model of particle physics. While it successfully describes electrodynamic, strong, and weak forces without an underlying unifying concept, its further limitations become apparent when considering dark matter or baryon asymmetry. Despite these flaws, the model has consistently passed experimental tests, demonstrating its remarkable accuracy and predictive power.

One notable example of the Standard Model's experimental validation is the ongoing effort at Fermilab to precisely measure the magnetic moment of the muon. At first glance, it may seem counterintuitive that a particle as light as 105 MeV could be used to search for heavy, unseen new particles. However, the key ingredient here is the experiment's extraordinary precision. Today, scientists can measure this quantity with an astonishing relative error of 10^{-10}.

This feat is particularly impressive given that the magnetic moment receives radiative corrections from fluctuating quantum fields, as demonstrated by Julian Schwinger in 1948. Since then, the accuracy of his theory has been rigorously tested and confirmed to as many as twelve digits in the case of the electron.

Schwinger's theory provides a solid foundation for understanding radiative corrections on the muon's magnetic moment when considering electromagnetic and weak interactions. However, strong interactions pose a challenge to standard quantum field theory methods. To overcome this hurdle, scientists sought another process where the same contribution could be isolated and measured in an unrelated experiment. The hadronic R ratio from e+e- scattering experiments offered a solution, allowing for the extraction of the missing piece with smaller error than any theoretical calculation before. This 'data-driven' approach led to a result that fell short of the direct measurements by over five standard deviations, sufficient to claim the discovery of a new interaction. Unfortunately, discrepancies exist among the individual R ratio experiments.

Numerical evaluation of the partition function of Quantum Chromodynamics (QCD), which governs strong interactions, were also converging on a precise measurement of the strong contribution of the muon's magnetic moment. Notably, many of these first-principles calculations from QCD exhibit agreement with several experimental and theoretical results of the Standard Model. In response to the challenge related to the muon's magnetic moment, leading theory groups dedicated themselves to refining their numerical approaches with the goal of achieving the experiment's level of precision.

One key step in this process was to break down the result into short-range, intermediate-range, and long-range terms, allowing each term to be addressed individually. A contribution by C. Lehner in this volume presents a thorough analysis of the agreement among major theory groups' numerical results. This also provides readers with insight into the

intricacies of the simulations, including their limitations and the meticulous attention required to ensure reliability at such high levels of precision.

Another, somewhat unrelated significant challenge to the Standard Model is complexity. Although we can reasonably control the behaviour of individual elementary particles, many-body problems pose a formidable test for both the theory and our methods to solve them. Nuclear physics, in particular, exemplifies this challenge.

A major milestone was reached over a decade ago with the successful computation of the mass of the proton and neutron. This achievement required a precise accounting for the gluon fields, which are the carriers of the strong force. In fact, much of the mass of these nucleons is attributed to the energy of the interacting fields.

The dynamics of how these nucleons interact with each other presents a new chapter in the physics of strong interactions. For instance, the binding energies of light atomic nuclei are extremely small, on the order of per-mill relative to the mass of the nuclei, posing a significant challenge to theoretical predictions. However, the probably greatest challenge facing this field is understanding the largest nuclei in nature, the core of neutron stars. These extreme environments pose unique difficulties for theoretical models, requiring innovative approaches and cutting-edge computational capabilities.

Today, readers have access to a wealth of recent discoveries in nuclear interactions. This chapter begins with a concise review of cutting-edge nuclear simulations by U.-G. Meißner. The simulation results include the spectrum and charge radii of light nuclei, as well as the lifetime of the triton, among other examples. Unlike the other contributions in this edition, Meißner's approach avoids the use of gluons and quarks to model the strong force. Instead, an effective field theory is employed, treating nucleons as point-like particles interacting with auxiliary fields.

Thanks to recent advances in algorithm development, sophisticated methods have emerged for describing nucleons. To accurately simulate the densities found in neutron stars, however, it is necessary to account for hyperons – essentially nucleons with a single light quark replaced by a strange quark. Initial attempts to incorporate hyperons yielded a pressure-density relation (equation of state) that was too soft, potentially leading to the collapse of heavy observed neutron stars into black holes. Fortunately, repulsive three-body forces can stiffen the equation of state, resolving what is known as the hyperon puzzle and ensuring the stability of these massive celestial bodies.

An alternative approach to quantifying binding energies in light nuclei is to simulate full Quantum Chromodynamics (QCD) without relying on effective field theories. The second contribution in this chapter takes this route, computing the binding energy of the H dibaryon – a conjectured bound state of two hyperons – using J. R. Green et al.'s methodology. A crucial aspect of such simulations is the treatment of discretisation artefacts, which can introduce significant errors. To mitigate these effects, a continuum extrapolation is essential. This process is discussed in the context of renormalisation by C. Alexandrou in a separate contribution. Renormalisation plays a key role in connecting observables in a discretised theory to those in the real world, enabling physical predictions and allowing us to test the Standard Model further. One of the ultimate goals is to compute form factors and QCD matrix elements – essential components from the strong force that can be combined with electromagnetic and weak physics. By achieving this and combining the results with experimental findings, we can put the Standard Model under new scrutiny.

Nuclear Lattice Effective Field Theory: Status A.D. 2024

Ulf-G. Meißner

[1] Helmholtz-Institut für Strahlen- und Kernphysik and Bethe Center for Theoretical Physics, Universität Bonn, 53115 Bonn, Germany
E-mail: meissner@hiskp.uni-bonn.de

[2] Institute for Advanced Simulation (IAS-4), Forschungszentrum Jülich, 52425 Jülich, Germany

I discuss recent developments in nuclear lattice effective field theory, which is a premier tool in the theory of nuclear structure and reactions. Topics include the wavefunction matching method as a new tool for quantum many-body theory, allowing for accurate calculations at N3LO in the chiral expansion, as well as applications of the precise forces in few- and many nucleon systems. I also discuss a first step for precision hypernuclear physics, describe the solution of a puzzle related to the ^{4}He monopole transition form factor, discuss the calculation of hyper-neutron matter and give new insights how Big Bang nucleosynthesis constrains fundamental parameters of the Standard Model.

1 Introduction

Understanding the formation of strongly interacting systems such a atomic nuclei from first principles calculations is still one of the biggest challenges within contemporary theoretical physics. While the theory of the strong interactions, Quantum Chromodynamics (QCD), is well tested in many processes, the matter that leads to life in our Universe is based on nuclei, which are self-bound systems of nucleons (protons and neutrons). As the nucleons themselves consist of quarks and gluons, and hence are not fundamental degrees of freedom, the forces between nucleons are not completely given in terms of two-body interactions, but include three-body and higher order interaction terms. Much progress in the understanding of the structure and dynamics of nuclei has been made in the context of Nuclear Lattice Effective Field Theory (NLEFT)[1], which combines the so successful low-energy chiral effective field theory of QCD with stochastic methods (Monte Carlo simulations). While direct calculations of nuclei based on quarks and gluons in the framework of lattice QCD are essentially impossible due to the severe sign problem, formulating the nuclear forces in terms of protons, neutrons and pions is not only more appropriate, but also comes with the added value of the approximate Wigner SU(4) (spin-isospin) symmetry of the underlying nuclear interactions. This symmetry in fact suppresses the sign oscillations strongly, and in the limit of an exact Wigner SU(4) symmetry, spin-isospin saturated nuclei like e.g. ^{4}He are free of any sign oscillation. In NLEFT simulations, Euclidean space-time is discretised on a torus of volume $L^3 \times L_t$, where L is the side length of the spatial dimension, and L_t denotes the extent of the Euclidean time dimension. The lattice spacing in the spatial (temporal) dimensions is a (a_t). The maximal momentum on the lattice is $p_{\rm max} \equiv \pi/a$, which serves as the UV regulator of the theory. Nucleons are point-like particles on the lattice sites, and the interactions between nucleons (pion exchanges and contact terms) are treated as insertions on the nucleon world lines via auxiliary-field representations. Properties of multi-nucleon systems are computed by means of the projection Monte Carlo (MC) method. Each nucleon is treated as a single particle propagating in a

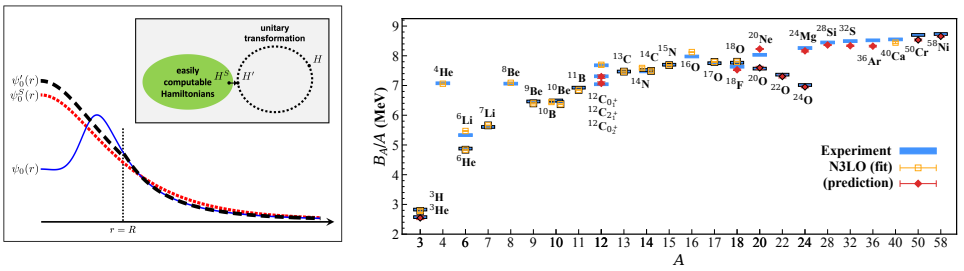


Figure 1. Left panel: Pictoral representation of wavefunction matching. The simple Hamiltonian H^S is an easily computable Hamiltonian while the high-fidelity Hamiltonian H is not. A unitary transformation on the two-nucleon interaction with finite range R is used to produce a new Hamiltonian H' that is close to H^S. In each two-body channel, the ground state wave function of H' matches the ground state wave function of H for $r > R$ and is proportional to the ground state wave function of H^S for $r < R$. Right panel: Results for nuclear binding energies using wavefunction matching. Calculated ground state and excited state energies of some selected nuclei with up to $A = 58$ at N3LO in chiral EFT and comparison with experimental data. The nuclei used in the fit of the higher-order three-nucleon interactions are labelled with open squares, while the other nuclei are predictions denoted with filled diamonds.

fluctuating background of pion and auxiliary fields. Both local and non-local smearings are applied to the nucleon creation and annihilation operators. Euclidean time projection is started from some initial state Ψ_A for Z protons and N neutrons (with $A = Z + N$). One calculates the ground state energy and other properties from the correlation function $Z(t) \equiv \langle \Psi_A | \exp(-tH) | \Psi_A \rangle = \mathrm{Tr}\{M^{L_t}\}$, in the limit of large Euclidean projection time t, with M the normal-ordered transfer-matrix operator and L_t the number of Euclidean time steps. Higher-order contributions are computed as perturbative corrections to the LO results. A much more detailed description is given in the monograph[1].

2 Wavefunction Matching

Quantum Monte Carlo (QMC) simulations are a powerful and efficient method that can describe strong correlations in quantum many-body systems. No assumptions about the nature of the system are necessary, and the computational effort grows only as a low power of the number of particles. For many problems of interest, a simple Hamiltonian H^S can be found that describes the energies and other observables of the many-body system in fair agreement with empirical data and is easily computable using MC methods. On the other hand, realistic high-fidelity Hamiltonians usually suffer from severe sign problems with positive and negative contributions to the averages cancelling each other, so that Monte Carlo calculations become impractical. In Ref. 2, this problem was solved introducing a new approach called wavefunction matching (WFM). While keeping the observable physics unchanged, wavefunction matching creates a new high-fidelity Hamiltonian H' such that wave functions at short distances match that of a simple Hamiltonian H^S which is easily computed. This allows for a rapidly converging expansion in powers of the difference $H' - H^S$. WFM can be used with any computational scheme. In the following analysis, we focus on the case of QMC simulations, where the method presents a promising and practical strategy for evading the sign problem in realistic calculations of nuclear quantum many-body systems. The basic idea of WFM is easily described. Starting from a

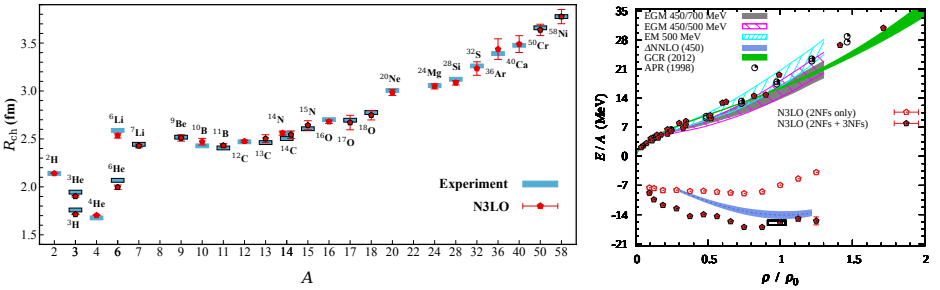


Figure 2. Left panel: Predictions for charge radii of nuclei up to $A = 58$ at N3LO in chiral EFT in comparison to the experimental data. Right panel: Predictions for pure neutron matter energy per neutron and symmetric nuclear matter energy per nucleon as a function of density at N3LO in chiral EFT. For comparison we show the results Refs. 5–8. The empirical saturation point is labelled with a black rectangular box.

realistic high-fidelity Hamiltonian H, WFM defines a new Hamiltonian

$$H' = U^\dagger H U \ , \tag{1}$$

where U is a unitary transformation and $U^\dagger$ its Hermitian conjugate. More precisely, we consider the two-nucleon interaction here. In each two-body angular momentum channel, the unitary transformation U is active only when the separation distance between two particles is less than some chosen distance R. Let us write $\psi_0(r)$, $\psi'_0(r)$, and $\psi^S_0(r)$ for the ground state wave functions of H, H', and the simple Hamiltonian H^S, respectively. The transformation U is defined such that $\psi'_0(r)$ is proportional to $\psi^S_0(r)$ for $r < R$. The simple Hamiltonian is chosen so that the constant of proportionality is close to 1. For $r > R$, however, U is not active and so $\psi'_0(r)$ remains equal $\psi_0(r)$. This is illustrated in the left panel of Fig. 1. In the NLEFT application discussed here, the simple Hamiltonian consists of SU(4) invariant two-body forces as well as one-pion exchange. All higher order corrections are treated in first order perturbation theory. The short-distance parts of the chiral three-nucleon (3N) interactions are then tuned to minimise errors in the binding energies of selected light and medium-mass nuclei. A total of six additional 3N parameters are adjusted, and it is found that with just one parameter, the root-mean-square-deviation (RMSD) for the energy per nucleon drops from 1.2 MeV down to 0.4 MeV. With the addition of a few additional parameters, the RMSD per nucleon drops further to about 0.1 MeV, as shown in the right panel of Fig. 1. Having fixed all parameters, we can now make predictions. In the left panel of Fig. 2, the results for the charge radii of nuclei with up to $A = 58$ nucleons are shown. No charge radii data were used to fit any interaction parameters. The one-standard-deviation point estimate error bars represent computational uncertainties due to MC errors, infinite volume extrapolation, and infinite time extrapolation. The agreement with empirical results is quite good, with an RMSD of about 0.03 fm. Note that the larger errors for the heaviest nuclei are statistical and can be decreased by utilising greater computational resources. This solves the long-standing "radius puzzle" observed in all continuum *ab initio* many-body calculations, namely that when getting the correct binding energies, the nuclear radii come out too small, see e.g. Refs. 3,4. In the right panel of Fig. 2, lattice results for the energy per nucleon versus density for pure neutron matter and symmetric nuclear matter are shown. None of the neutron matter and symmetric nu-

clear matter data were used to fit any interaction parameters. The density is expressed as a fraction of the saturation density for nuclear matter, $\rho_0 = 0.17$ fm^{-3}. For the neutron matter calculations, we consider 14 to 80 neutrons in periodic box lengths ranging from 6.58 fm to 13.2 fm. For the symmetric nuclear matter calculations, we use system sizes from 12 to 160 nucleons in a periodic box of length 9.21 fm. We see that the neutron matter calculations agree well with previous calculations. Within the uncertainties due to finite system size corrections, the symmetric nuclear matter calculations show saturation at an energy and density consistent with the empirical saturation point labelled with the black rectangular box. The relative uncertainties due to finite system size are at the 10% level for the energy.

3 Testing the High-Fidelity Interactions

Next, the so determined N3LO high-fidelity interactions are tested in a number of calculations, which we discuss briefly.

3.1 Structure Factors for Hot Neutron Matter

In Ref. 9 the first *ab initio* lattice calculation of spin and density correlations in hot neutron matter using the high-fidelity interactions at N3LO in chiral EFT was done. These correlations have a large impact on neutrino heating and shock revival in core-collapse supernovae and are encapsulated in functions called structure factors. Unfortunately, calculations of structure factors using high-fidelity chiral interactions were well out of reach using existing computational methods. To solve the problem, a computational approach called the rank-one operator (RO) method is introduced. The RO method is a general technique with broad applications to simulations of fermionic many-body systems. It solves the problem of exponential scaling of computational effort when using perturbation theory for higher-body operators and higher-order corrections. Using the RO method, we compute the vector and axial static structure factors for hot neutron matter as a function of temperature and density given by:

$$\begin{aligned} S_{\rm v}(\boldsymbol{q}) &= \frac{1}{L^3}\sum_{\boldsymbol{n}\boldsymbol{n}'} \mathrm{e}^{-i\boldsymbol{q}\cdot\boldsymbol{n}} \left[\langle\hat{\rho}(\boldsymbol{n}+\boldsymbol{n}')\hat{\rho}(\boldsymbol{n}')\rangle - (\rho^0)^2\right], \\ S_{\rm a}(\boldsymbol{q}) &= \frac{1}{L^3}\sum_{\boldsymbol{n}\boldsymbol{n}'} \mathrm{e}^{-i\boldsymbol{q}\cdot\boldsymbol{n}} \left[\langle\hat{\rho}_z(\boldsymbol{n}+\boldsymbol{n}')\hat{\rho}_z(\boldsymbol{n}')\rangle - (\rho_z^0)^2\right], \end{aligned} \tag{2}$$

where $\hat{\rho}$ and $\hat{\rho}_z$ are the density and the spin-density operators, respectively, and $\boldsymbol{n}, \boldsymbol{n}'$ represent coordinates on the L^3 cubic lattice. The *ab initio* lattice results are in good agreement with virial expansion calculations at low densities but are more reliable at higher densities, see the left panel of Fig. 3. Random phase approximation codes used to estimate neutrino opacity in core-collapse supernovae simulations can now be calibrated with these precise *ab initio* lattice calculations.

3.2 Nuclear Charge Radii of Silicon Isotopes

The next test of the N3LO forces was done in collaboration with experimentalists from FRIB[10]. They determined the nuclear charge radius of ^{32}Si using collinear laser spectroscopy, leading to $R_{\rm ch}(^{32}\mathrm{Si}) = 3.153(12)\,\mathrm{fm}$. The experimental result was confronted

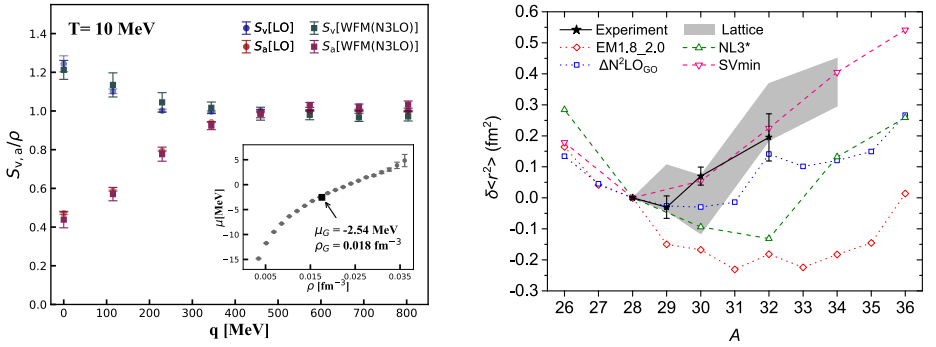


Figure 3. Left panel: Calculated momentum dependent neutron matter structure factors $S_{\rm v}$ and $S_{\rm a}$ at $T = 10$ MeV. WFM(N3LO) represents the NLEFT calculations with the WFM N3LO interaction. The insert figure shows calculated chemical potentials of canonical ensemble systems which are used for the construction of grand canonial ensemble at the chemical potential $\mu_G = -2.54$ MeV and the density $\rho_G = 0.01758(4)$ fm^{-3}. Right panel: Experimental and theoretical differential mean square charge radii of Si. The NLEFT calculation provided an uncertainty which is plotted as the gray band.

with *ab initio* NLEFT, valence-space in-medium similarity renormalisation group, and mean field calculations, highlighting important achievements and challenges of modern many-body methods. The lattice simulations for the charge radii are new calculations based upon the N3LO chiral interactions described in Ref. 2 with two additional improvements made. Rather than a global fit to all nuclei, we fit the three-nucleon coefficients $c_E^{(l)}$ and $c_E^{(t)}$ to ensure good agreement with the binding energies of the silicon isotopic chain. We also use the rank-one operator method introduced in Ref. 9 to compute the charge radii. As shown in the right panel of Fig. 3, the NLEFT results are in good agreement with the measured charge radii along the Si isotope chain from $A = 28$ to $A = 32$. The charge radius of ^{32}Si completes the radii of the mirror pair ^{32}Ar - ^{32}Si, whose difference is correlated to the slope L of the symmetry energy in the nuclear equation of state. The NLEFT result for L was determined from the calculations of pure neutron matter in Ref. 2, giving $L = 55(7)$ MeV, which agrees with complementary observables.

3.3 The Triton Lifetime

Nuclear β and double-β decays are fine probes of the weak interactions in their interplay with the strong force. Arguably the best example is the extraction of the CKM matrix element V_{ud} from superallowed β decays[11]. Triton β-decay is the process where ^{3}H decays into ^{3}He, an electron, and an electron antineutrino, $^3\mathrm{H} \rightarrow {}^3\mathrm{He} + e^- + \bar{\nu}_e$. The matrix elements of the weak transition are crucial to understanding this decay process. Thus, this decay serves as a benchmark for calculating weak nuclear decays. In addition, it is known that triton β-decay, that is the triton lifetime, together with the binding energies in the $A = 3$ system can lead to a robust determination of the low-energy constants c_D and c_E parameterising the leading three-nucleon forces in chiral EFT[12]. The triton lifetime is given in terms of two matrix elements (MEs), referred to as the Fermi and the Gamow-

Teller MEs,

$$\langle F\rangle = \sum_{n=1}^{3}\langle {}^3\mathrm{He}\|\tau_{n,+}\|{}^3\mathrm{H}\rangle \ , \quad \langle GT\rangle = \sum_{n=1}^{3}\langle {}^3\mathrm{He}||\sigma_n\tau_{n,+}||{}^3\mathrm{H}\rangle \ , \tag{3}$$

in order. Here, $\tau_{n,+}$ is the isospin-raising operator and the σ_n are the nucleon spin matrices. Despite the success of the WFM method in improving theoretical precision, the calculations in Ref. 2 were carried out using first-order perturbation theory. Since first-order perturbation theory only provides corrections to the energy and not to the wave functions, triton β-decay calculations at N3LO, requiring higher-order perturbative corrections for realistic wave functions, cannot be directly performed using the methods from Ref. 2. One potential solution to this challenge is to extend the calculations to second-order perturbation theory. Recent advances in perturbative QMC methods, as detailed in Ref. 13, provide an effective framework for incorporating higher-order perturbative corrections, making it particularly well-suited for applications to heavier nuclei. Alternatively, fully non-perturbative methods can be applied to light nuclear systems to generate realistic wave functions at N3LO, as required for triton β-decay calculation. This was done in Ref. 14. Performing this non-perturbative calculation, the Fermi ME as well as the Gamow-Teller ME are simultaneously obtained,

$$\langle F\rangle = 0.99949(11) \ , \quad \langle GT\rangle = 1.6743(58) \ , \tag{4}$$

where the uncertainty stems from the large L_t extrapolation and the variation of the strengths of the one-pion exchange and contact term topologies of the three-nucleon forces. These results are consistent with earlier theoretical calculations, confirming the robustness of our approach. The corresponding lifetime is given by $(1+\delta_R)\,t_{1/2}\,f_V = 1105.1(74)$ s, consistent with the empirical determinations, $(1+\delta_R)\,t_{1/2}\,f_V = 1132.1(25)$ s. The remaining discrepancies are due to the fact that the corrections to the pion exchange currents have not yet been included. This study marks a significant advancement in the systematic application of NLEFT to nuclear β-decay processes, paving the way for future high-precision calculations in more complex nuclear systems, such as neutrinoless double-β decay in ^{48}Ca or ^{76}Ge.

4 Towards Hypernuclei from NLEFT

Understanding the strong interactions in the light quark sector is crucial for a comprehensive description of baryonic systems such as nuclei and hypernuclei. The study of hypernuclei provides valuable insights into the baryon-baryon interactions, and an accurate description of the properties of hypernuclei requires a systematic formulation of interactions between hyperons and nucleons, as well as constraining their low-energy constants (LECs). The great success of both phenomenological potential models and chiral EFT for nucleons is based on rich and precise NN-scattering data and nuclear binding energies. However, due to the scarcity of hyperon-nucleon and hyperon-hyperon scattering data, the spectra of hypernuclei are pivotal in constraining the hyperon-nucleon and hyperon-hyperon interactions, deepening our understanding of SU(3) flavour symmetry breaking and charge symmetry breaking in strong interactions. In Ref. 15 we calculated the ground state and excited state energies of hypernuclei up to $A = 16$. Our calculations employ the

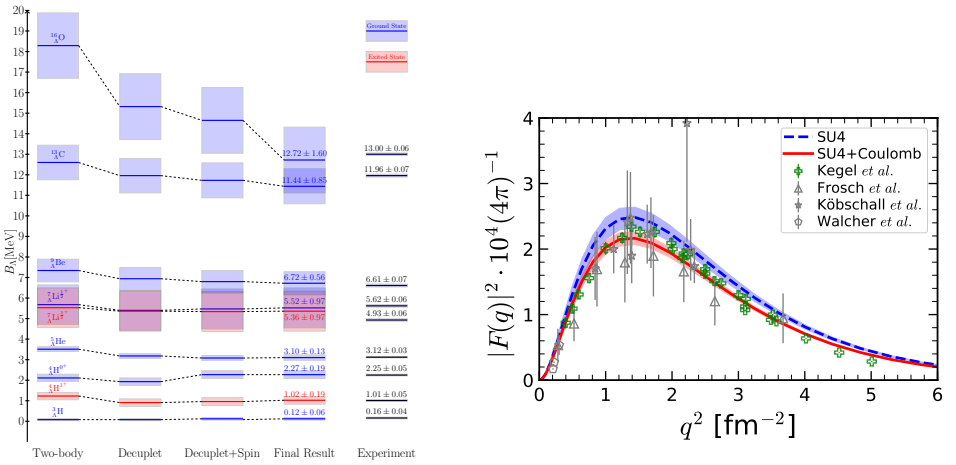


Figure 4. Left panel: Λ separation energies for different YNN forces (decuplet saturation: $C_1 = C_3$ and $C_2 = 0$, decuplet plus spin: $C_1 = C_3$ and $C_2 \neq 0$, final result: $C_1 \neq C_2 \neq C_3 \neq 0$). The large improvement resulting from the introduction of the spin-dependent three-body force $\sim C_2$ is clearly visible. The experimental values are taken from Ref. 16, where we averaged the four-body systems. Ground states are depicted in blue, excited states in red. The uncertainties are indicated by the shaded areas. Right panel: Calculated monopole form factor of the $0_2^+ \to 0_1^+$ transition in ^{4}He compared to the recent data from Mainz[18] (green squares) and the older data (grey symbols). Blue dashed line: SU(4) symmetric strong interaction with all parameters determined in Ref. 20. Red solid line: adding the Coulomb interaction perturbatively. The uncertainty bands in the lattice results include stochastic errors and uncertainties in the Euclidean time extrapolation.

high-fidelity chiral interactions at N3LO for nucleons from Ref. 2 and the leading-order S-wave hyperon-nucleon (YN) interactions are given by

$$V_{\rm YN} = \frac{1}{4} C_{YN}^S (\mathbb{1} - \boldsymbol{\sigma_1} \cdot \boldsymbol{\sigma_2}) + \frac{1}{4} C_{YN}^T (3 + \boldsymbol{\sigma_1} \cdot \boldsymbol{\sigma_2}) \,. \tag{5}$$

The LECs $C_{YN}^{S,T}$ are determined by a fit to the unpolarised $\Lambda p \to \Lambda p$ cross section and the hypertriton binding energy. The hyperon-nucleon-nucleon (YNN) interactions are given by

$$\begin{aligned} V_{\rm YNN} = {} & C_1 (\mathbb{1} - \boldsymbol{\sigma}_2 \cdot \boldsymbol{\sigma}_3)(3 + \boldsymbol{\tau}_2 \cdot \boldsymbol{\tau}_3) + C_2 \, \boldsymbol{\sigma}_1 \cdot (\boldsymbol{\sigma}_2 + \boldsymbol{\sigma}_3)(\mathbb{1} - \boldsymbol{\tau}_2 \cdot \boldsymbol{\tau}_3) \\ & + C_3 (3 + \boldsymbol{\sigma}_2 \cdot \boldsymbol{\sigma}_3)(\mathbb{1} - \boldsymbol{\tau}_2 \cdot \boldsymbol{\tau}_3) \,, \end{aligned} \tag{6}$$

and the LECs $C_{1,2,3}$ are determined by hypernuclear systems with $A = 4$ and 5. For the YNN interactions, we consider all possible forms of short-distance smearing. In our analysis, we calculate the RMSD over all calculated hypernuclear separation energies with $A \geq 4$, which are used to assess the accuracy of the YNN interactions in describing hypernuclei. In the left panel of Fig. 4 we show results for hypernuclei from $^3_\Lambda$He to $^{16}_\Lambda$O, where the hypernuclei with $A \leq 5$ shown here are included in the fit, while the other hypernuclei are predictions. We find that, within stochastic uncertainties of the MC simulations, our Hamiltonian can accurately describe hypernuclear systems. Clearly, improvements in the considered interactions here should be performed. We recommend including pion exchange forces in both the two-body and three-body sector. These forces not only allow

for an automatic inclusion of higher momentum contributions but also make excited states available in typical multichannel calculations, Additionally, this approach enables the inclusion of higher orders in the chiral expansion, which are necessary for better phase shift descriptions at higher orders that will also improve the description of the hypernuclei.

5 The Puzzling ^{4}He Transition Form Factor

The ^{4}He nucleus, the α-particle, is considered to be a benchmark nucleus for our understanding of the nuclear forces and the few-body methods to solve the nuclear A-body problem[17]. The attractive nucleon-nucleon interaction makes this highly symmetric four-nucleon system enormously stable. Furthermore, its first excited state has the same quantum numbers as the ground state, $J^P = 0^+$ with $J(P)$ the spin (parity), but is located about 20 MeV above the ground state. This large energy of the first quantum excitation makes the system difficult to perturb. This isoscalar monopole resonance of the ^{4}He nucleus presents a challenge to our understanding of nuclear few-body systems and the underlying nuclear forces. The recent precision measurement of the corresponding transition form factor of the first excited state to the ground state at the Mainz Microtron MAMI[18] compared with *ab initio* calculations based on the Lorentz-integral transformation method using phenomenological potentials as well as potentials based on chiral EFT revealed sizeable discrepancies as shown in Fig. 3 of Ref. 18. We addressed this issue in Ref. 19 within the framework of the minimal nuclear interaction that reproduces the ground state properties of light nuclei, medium-mass nuclei, and neutron matter simultaneously with no more than a few percent error in the energies and charge radii[20,21]. The transition form factor $F(q)$ of the monopole transition is related to the transition density $\rho_{\rm tr}(r)$ by

$$F(q) = \frac{4\pi}{Z}\int_0^\infty \rho_{\rm tr}(r) j_0(qr) r^2 dr = \frac{1}{Z}\sum_{\lambda=1}^{\infty} \frac{(-1)^\lambda}{(2\lambda+1)!} q^{2\lambda} \langle r^{2\lambda}\rangle_{\rm tr} \,, \tag{7}$$

with Z the charge of the nucleus under consideration. Here $Z = 2$, and $\rho_{\rm tr}(r) = \langle 0_1^+|\hat{\rho}(\vec{r})|0_2^+\rangle$ is the matrix element of the charge density operator $\hat{\rho}(\vec{r})$ between the ground state 0_1^+ and the first excited 0_2^+ state. We also display the expansion in moments in Eq. 7. The first excited state of ^{4}He is a resonance that sits just above the ^{3}H+p threshold. In order to study this continuum state, we perform calculations using three different cubic periodic boxes with lengths $L = 10, 11, 12$ in lattice units, corresponding to $L = 13.2$ fm, 14.5 fm, 15.7 fm. The corresponding ground and first excited state energies are $E(0_1^+) = -28.30(3)$ MeV and $E(0_2^+) = -7.96(9)$ MeV that compare well with the experimental values of -28.30 MeV and -8.09 MeV, respectively. Next, we turn to the analysis of the transition form factor, denoted as $F(q)$. In the framework of NLEFT, observables such as nucleon density distributions, charge radii and form factors can be computed using the pinhole algorithm. First, we consider the SU(4)-symmetric interactions without Coulomb. The resulting form factor is depicted by the blue dashed line in the right panel of Fig. 4. It somewhat overshoots the data, although the error band associated with stochastic errors and the large L_t extrapolation almost encompasses the data. Including the Coulomb interaction leads to an overall reduction of the transition form factor as shown by the red solid line in the right panel of Fig. 4. Overall, we achieve a good reproduction of the data and the uncertainty band is also somewhat reduced. This is due to the fact that inclusion of the Coulomb interaction leads to smaller fluctuations in the

Monte Carlo data when extrapolating to large L_t. Consequently, we find that the nuclear interaction defined in Ref. 20, which has already been shown to reproduce the essential elements of nuclear binding, also leads to a good description of the α-particle transition $0_2^+ \to 0_1^+$ form factor without adjusting any parameters. Thus, the nuclear forces relevant to this system are under good control, and we do not find the puzzle mentioned in Ref. 18, see also Refs. 22, 23.

6 *Ab initio* Calculation of Hyper-Neutron Matter

The equation of state (EoS) of neutron matter plays a decisive role to understand the neutron star properties and the gravitational waves from neutron star mergers. At sufficient densities, the appearance of hyperons generally softens the EoS, leading to a reduction in the maximum mass of neutron stars well below the observed values of about 2 solar masses. Even though repulsive three-body forces are known to solve this so-called "hyperon puzzle", see e.g. Refs. 24, 25, so far performing *ab initio* MC calculations with a substantial number of hyperons has remained elusive. We addressed this challenge by employing NLEFT in Ref. 26. First, we had to develop an algorithm that allows to go to densities beyond twice nuclear matter densities reached so far in QMC simulations which is not sufficient for the description of neutron stars. To achieve that, we combine the smeared nucleon operator with the operator representing the Λ hyperon, as detailed in Ref. 26. This enables simulations of systems consisting of both arbitrary number of nucleons and arbitrary number of Λ hyperons with a single auxiliary field. Second, we work with smeared contact interactions only, which allows to include all possible interactions, that is NN, NY, YY, NNN, NNY and NYY, which was never done in a QMC simulations before. The NN and NNN LECs are determined from a combined fit to the S-wave phase shifts and the saturation properties of nuclear matter, with $\rho_0 = 0.17\,\mathrm{fm}^{-3}$ the nuclear matter density. This calculation generates a very stiff neutron matter EoS as shown in the left panel of Fig. 5, and it required up to 232 nucleons in the finite volume to achieve densities of $5\rho_0$ as in the interior of neutron stars. Next, we show three different EoS when hyperons are included. The NNΛ and N$\Lambda\Lambda$ forces are constrained by the separation energies of single- and double-Λ hypernuclei, spanning systems from ${}^{5}_{\Lambda}$He to ${}^{6}_{\Lambda\Lambda}$Be, denoted as HNM(I). It is difficult to probe the behaviour of the EoS at high densities encountered in neutron stars in terrestrial laboratories, and various phenomenological schemes and microscopical models suggest that hyperons emerge in the inner core of neutron stars at densities around $\rho \approx (2-3)\rho_0$. Similar to using the saturation properties of symmetric nuclear matter to pin down the three-nucleon forces, we determined the NNΛ and N$\Lambda\Lambda$ forces by using the separation energies of hypernuclei and the Λ threshold densities $\rho_\Lambda^{\mathrm{th}}$ around $(2-3)\rho_0$ simultaneously in HNM(II) and HNM(III). We set $\rho_\Lambda^{\mathrm{th}}$ = 0.398(2)(5) fm^{-3} and 0.520(2)(6) fm^{-3} for HNM(II) and HNM(III), respectively. The corresponding EoSs are also shown in the left panel of Fig. 5. To fulfil the equilibrium condition for the chemical potentials, $\mu_n = \mu_\Lambda$, we needed 102, 92, and 32 Λs for HNM(I), HMN(II) and HNM(III), in order. The EoS becomes stiffer at higher densities for these variants, indicating the inclusion of more repulsion in the three-body hyperon-nucleon interactions. As anticipated, the inclusion of hyperons results in a softer EoS and HNM(III) is the stiffest EoS when hyperons are included. The squared speed of sound, c_s^2, is also shown in the inset in the left panel of Fig. 5. It is observed that the causality limit ($c_s^2 < 1$) is fulfilled for both PNM and HNM. The EoS characterised by nucleonic degrees of freedom exclusively demonstrate a

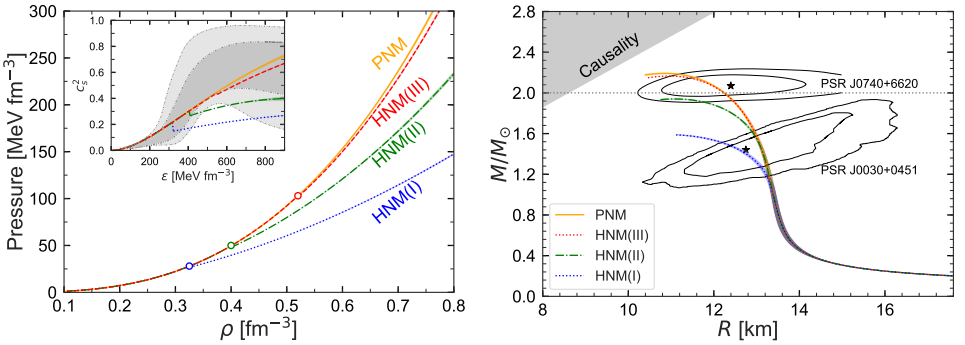


Figure 5. Left panel: EoS for PNM and HNM. The orange solid curve denotes pure neutron matter, obtained from the NN and NNN interactions. The red dashed line represents the EoS of HNM with hyperons interacting via the two-body interactions (NΛ and $\Lambda\Lambda$) and the third set of three-body hyperon-nucleon interaction (NNΛ and N$\Lambda\Lambda$). The blue dotted curve and the green dot-dashed curve are calculated with the first and second sets of three-body hyperon-nucleon interactions. The Λ threshold densities $\rho_\Lambda^{\rm th}$ are marked by open circles. In the inset, the speed of sound corresponding to the PNM and HNM EOSs is shown. The gray shaded regions are the inference of the speed of sound for neutron star matter in view of the recent observational data[27]. Right panel: Neutron star mass-radius relation. The legend is the same as in the left panel. The gray horizontal dotted line represents $2M_\odot$. The inner and outer contours indicate the allowed area of mass and radius of neutron stars by NICER's analysis of PSR J0030+0451[30] and PSR J0740+6620[31]. The excluded causality region is also shown by the grey shaded region[32].

monotonic increase in c_s^2 with increasing energy density. The appearances of Λ hyperons, however, induces changes in this behaviour, leading to non-monotonic curves that signify the incorporation of additional degrees of freedom. The onset of Λ hyperons precipitates a sharp reduction in the speed of sound, marking a significant transition in the stiffness of the EoS. For comparison, the constraints on c_s^2 within the interiors of neutron stars inferred by a Bayesian inference method are also shown[27]. The "holy grail" of neutron-star structure, the mass-radius (MR) relation, is displayed in the right panel of Fig. 5. These relations for PNM and HNM are obtained by solving the Tolman-Oppenheimer-Volkoff (TOV) equations[28,29] with the EoSs of Fig. 5 (left panel). The appearance of Λ hyperons in neutron star matter remarkably reduces the predicted maximum mass compared to the PNM scenario. The maximum mass for PNM, HNM(I), HNM(II), and HNM(III) is 2.19(1)(1) $M_\odot$, 1.59(1)(1) $M_\odot$, 1.94(1)(1) $M_\odot$, and 2.17(1)(1) $M_\odot$, respectively. Note that three neutron stars have been measured to have gravitational masses close to $2M_\odot$ that significantly constrain the EoS of dense nuclear matter. Our results show that the inclusion of the NNΛ and N$\Lambda\Lambda$ interaction in HNM(III) leads to an EoS stiff enough such that the resulting neutron star maximum mass is compatible with the three mentioned measurements of neutron star masses. Therefore, the repulsion introduced by the hyperonic three-body interactions plays a crucial role, since it substantially increases the value of the Λ threshold density. Note that Ref. 26 also contains the first *ab initio* calculation of the universal I-Love-Q relations, which connect the moment of inertia I, tidal deformability Λ, and the quadrupole moment Q in a slow rotation approximation. In the next steps, one should include the proton fraction, other hyperons of the baryon octet, and make use of the recently developed high-fidelity chiral interactions at N3LO[2], though this will pose a formidable computational challenge.

7 Big Bang Nucleosynthesis as a Probe of Fundamental Constants

Element generation in Big Bang nucleosynthesis (BBN) is a fine laboratory to study the possible variations of the fundamental parameters of the Standard Model, such as the quark masses or the electromagnetic fine-structure constant $\alpha_{\rm EM}$, see e.g. Ref. 33. However, the reaction network is also very sensitive to the nuclear physics input, which so far has not been studied systematically. In Ref. 34 we investigated the dependence of primordial nuclear abundances on fundamental nuclear observables such as binding energies, scattering lengths, neutron lifetime, etc. by varying these quantities. The numerical computations were performed with four publicly available codes, thus facilitating an investigation of the model-dependent (systematic) uncertainties on these dependences. Indeed deviations of the order of a few percent are found. Moreover, accounting for the temperature dependence of the sensitivity of the rates to some relevant parameters leads to a reduction of the sensitivity of the final primordial abundances, which in some cases is appreciable. These effects are considered to be relevant for studies of the dependence of the nuclear abundances on fundamental parameters such as quark masses or couplings underlying the nuclear parameters studied here. Based on that work, we studied in Ref. 35 the dependence of the primordial nuclear abundances as a function of $\alpha_{\rm EM}$, keeping all other fundamental constants fixed. We updated the leading nuclear reaction rates, in particular the electromagnetic contribution to the neutron-proton mass difference pertinent to β-decays, and went beyond certain approximations made in the literature. In particular, we included the temperature-dependence of the leading nuclear reactions rates and assessed the systematic uncertainties by using four different publicly available codes for BBN. Disregarding the unsolved so-called lithium-problem, we find that the current values for the observationally based ^{2}H and ^{4}He abundances restrict the fractional change in $\alpha_{\rm EM}$ to less than 2% , which is a tighter bound than found in earlier works on the subject. Further, in Ref. 37, we presented an improved calculation of the light element abundances in the framework of BBN as a function of the Higgs vacuum expectation value v. We improved and corrected the recent calculation of Ref. 36 and earlier works on this topic by combining up-to-date lattice data on the nucleon mass, the axial-vector coupling, etc, with chiral EFT methods. The PDG result for the ^{4}He abundance can be explained within 2σ by $0.004 \leq \delta v/v \leq 0.007$.For deuterium we find the constraint $-0.0007 \leq \delta v/v \leq -0.0002$.These bounds are more stringent than what was found earlier, and, in particular, the tightest bound is now set by deuterium, not ^{4}He any more (as in all earlier works). This is a significant step in the quest for finding the habitable universes as constrained by fundamental parameter variations in nuclear structure and reactions.

Acknowledgements

I thank my NLEFT colleagues for their contributions to the results presented here. The work reported here is part of the ERC AdG EXOTIC supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 101018170). This work was also supported in part by the DFG (DFG Project ID 196253076 - TRR 110) through the funds provided to the Sino-German CRC 110 "Symmetries and the Emergence of Structure in QCD", by the Chinese Academy of Sciences (CAS) President's International Fellowship Initiative (PIFI)

(grant no. 2025PD0022), by the MKW NRW under the funding code NW21-024-A and by VolkswagenStiftung (grant no. 93562). The computational resources were provided by the Jülich Supercomputing Centre (JSC) at Forschungszentrum Jülich. Special resources on the JURECA-DC supercomputer at the JSC are particularly acknowledged.

References

1. T. A. Lähde and U.-G. Meißner, Lect. Notes Phys. **957**, 1-396, 2019.
2. S. Elhatisari et al., Nature **630**, no.8015, 59-63, 2024.
3. A. Cipollone, C. Barbieri, and P. Navrátil, Phys. Rev. C **92**, no.1, 014306, 2015.
4. P. Maris et al. [LENPIC], Phys. Rev. C **106**, no.6, 064002, 2022.
5. A. Akmal et al., Phys. Rev. C **58**, 1804-1828, 1998.
6. S. Gandolfi, J. Carlson, and S. Reddy, Phys. Rev. C **85**, 032801, 2012.
7. I. Tews et al., Phys. Rev. Lett. **110**, no.3, 032504, 2013.
8. A. Ekström et al., Phys. Rev. C **97**, no.2, 024332, 2018.
9. Y. Z. Ma et al., Phys. Rev. Lett. **132**, no.23, 232502, 2024.
10. K. König et al., Phys. Rev. Lett. **132**, no.16, 162502, 2024.
11. J. C. Hardy and I. S. Towner, Phys. Rev. C **91**, no.2, 025501, 2015.
12. D. Gazit et al., Phys. Rev. Lett. **103**, 102502, 2009.
13. B. N. Lu et al., Phys. Rev. Lett. **128**, no.24, 242501, 2022.
14. S. Elhatisari, F. Hildenbrand, and U.-G. Meißner, Phys. Lett. B **859**, 139086, 2024.
15. F. Hildenbrand, S. Elhatisari, Z. Ren, and U.-G. Meißner, Eur. Phys. J. A **60**, 215, 2024.
16. P. Eckert, P. Achenbach and others, 2021, `https://hypernuclei.kph.uni-mainz.de`.
17. H. Kamada et al., Phys. Rev. C **64**, 044001, 2001.
18. S. Kegel et al., Phys. Rev. Lett. **130**, no.15, 152502, 2023.
19. U.-G. Meißner et al., Phys. Rev. Lett. **132**, no.6, 062501, 2024.
20. B. N. Lu et al., Phys. Lett. B **797**, 134863, 2019.
21. S. Shen et al., Nature Commun. **14**, no.1, 2777, 2023.
22. N. Michel et al., Phys. Rev. Lett. **131**, no.24, 242502, 2023.
23. M. Viviani et al., Few Body Syst. **65**, no.3, 74, 2024.
24. J. Schaffner-Bielich, Nucl. Phys. A **804**, 309-321, 2008.
25. D. Lonardoni et al., Phys. Rev. Lett. **114**, no.9, 092301, 2015.
26. H. Tong, S. Elhatisari, and U.-G. Meißner, 2024, arXiv:2405.01887 [nucl-th].
27. L. Brandes, W. Weise, and N. Kaiser, Phys. Rev. D **108**, no.9, 094014, 2023.
28. R. C. Tolman, Phys. Rev. **55**, 364-373, 1939.
29. J. R. Oppenheimer and G. M. Volkoff, Phys. Rev. **55**, 374-381, 1939.
30. M. C. Miller et al., Astrophys. J. Lett. **887**, no.1, L24, 2019.
31. T. E. Riley et al., Astrophys. J. Lett. **918**, no.2, L27, 2021.
32. J. M. Lattimer and M. Prakash, Phys. Rept. **442**, 109-165, 2007.
33. F. C. Adams, Phys. Rept. **807**, 1-111, 2019.
34. U.-G. Meißner and B. C. Metsch, Eur. Phys. J. A **58**, no.11, 212, 2022.
35. U.-G. Meißner, B. C. Metsch, and H. Meyer, Eur. Phys. J. A **59**, no.10, 223, 2023.
36. A. K. Burns et al., Phys. Rev. D **109**, no.12, 123506, 2024.
37. H. Meyer and U.-G. Meißner, JHEP **06**, 074, 2024, [Erratum: JHEP **01**, 033, 2025].

Hadron-Hadron Interactions from Lattice QCD

Jeremy R. Green[1], **Andrew D. Hanlon**[2], **Parikshit M. Junnarkar**[3,4], **Nolan B. Miller**[5,6], **Madanagopalan Padmanath**[7,8], **Srijit Paul**[9], **and Hartmut Wittig**[5,6]

[1] Deutsches Elektronen-Synchrotron DESY, Platanenallee 6, 15738 Zeuthen, Germany

[2] Department of Physics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA

[3] Research Centre for Nuclear Physics, Osaka University, 10-1 Mihogaoka, Ibaraki 567-0047, Japan

[4] Interdisciplinary Theoretical and Mathematical Sciences Program (iTHEMS), RIKEN, Wako 351-0198, Japan

[5] PRISMA$^+$ Cluster of Excellence and Institute for Nuclear Physics, Johannes Gutenberg-Universität Mainz, 55099 Mainz, Germany
E-mail: hartmut.wittig@uni-mainz.de

[6] Helmholtz Institute Mainz, GSI Helmholtz Centre for Heavy Ion Research, Staudingerweg 18, 55128 Mainz, Germany

[7] Institute of Mathematical Sciences, CIT Campus, Chennai 600113, India

[8] Homi Bhabha National Institute, Training School Complex, Anushaktinagar, Mumbai 400094, India

[9] Maryland Center for Theoretical Physics, University of Maryland, College Park, USA

Baryon-baryon interactions play a central role for a broad variety of phenomena in Nature, ranging from the formation of light nuclei to the physics of neutron stars. In our project we employ lattice QCD calculations and the formalism of finite-volume quantisation to perform a detailed study of baryon-baryon interactions. Specifically we focus on the H dibaryon, a conjectured bound state of two Λ hyperons, and nucleon-nucleon interactions, initially in three-flavour QCD. Our findings indicate that a detailed investigation of the continuum limit of lattice QCD is indispensable to obtain reliable results for binding energies. We observe a weakly bound H dibaryon with a binding energy of nearly 5 MeV. In the conceptually much more complex case of nucleon-nucleon interactions, higher partial waves as well as partial wave mixing must be included before firm conclusions can be drawn.

1 Introduction

Many open problems in nuclear physics are now being tackled through numerical simulations of the gauge theory that underlies the strong interaction, Quantum Chromodynamics (QCD), discretised on a Euclidean space-time lattice. One particular example is nucleon-nucleon interactions that provide key information on the formation of light nuclei. The small binding energy of the deuteron of 2.2 MeV has important consequences for Big Bang nucleosynthesis and the abundances of light elements in the universe. The question of how strongly the deuteron's binding energy depends on the mass of the light quarks is not merely an academic one but tells us how fine-tuned the universe actually is.

Moreover, baryons containing strange quarks (i.e. hyperons) are essential for our understanding of the physics of neutron-rich matter and neutron stars. A long-standing problem in this context is the question whether a bound state of two Λ hyperons – the so-called H dibaryon – could exist, and if so, what binding energy it would have. Thirdly, a detailed understanding of hadronic interactions is indispensable for investigating "exotic" hadronic states that cannot be accommodated within the standard three-quark or quark-antiquark framework of the quark model. Experimental evidence for penta- and tetra-quarks at the LHC and B factories has spawned a vigorous research effort in this direction.

Hadronic interactions and resonances can be treated in lattice QCD via the finite-volume quantisation formalism pioneered by Lüscher[1,2]. In this contribution, we report on our ongoing effort to study the H dibaryon and nucleon-nucleon interactions using this elegant formalism. As we shall see, it is essential to remove the effects arising from the discretisation of the QCD action in order to obtain reliable information on binding energies and scattering lengths.

2 Finite-Volume Quantisation Formalism

Lattice QCD calculations are performed in a finite volume of Euclidean space-time with spatial and temporal extent L and T, respectively. The standard method to extract hadron masses proceeds by isolating the exponential fall-off in the correlation function of suitably chosen interpolating operators at large Euclidean times. However, in the case of multi-hadron systems, the fall-off alone does not contain any information on hadronic interactions. This is the domain of finite-volume quantisation which we describe below.

For $2 \to 2$ scattering processes considered here, finite-volume quantisation is based on an exact relation between the scattering amplitude and the energy levels of two-particle states in a finite volume[1,2]. The generic form of the two-particle quantisation condition is[3]

$$\det\left[\tilde{K}^{-1}(p^2) - \mathcal{F}(p^2)\right] = 0, \tag{1}$$

where the matrix $\tilde{K}$ contains the scattering amplitude. For a system of two particles with individual momenta $\vec{p}_1$ and $\vec{p}_2$, the function $\mathcal{F}$ depends on the scattering momentum p^2, as well as on the volume, frame and irreducible representation (irrep) of the little group for total momentum $\vec{P} = \vec{p}_1 + \vec{p}_2$, which replaces the total angular momentum as a conserved quantum number. Each irrep contains a tower of different angular momenta J. The rows (or columns) of $\tilde{K}$ and $\mathcal{F}$ correspond to the different coupled channels and partial waves that can scatter, as well as different values of J.

In order to determine the K-matrix from Eq. 1 we must determine the scattering momentum p^2 which is related to the energy spectrum of the state of interest, e.g. the H dibaryon, which can be extracted from the correlation function. As our basis of interpolating operators, $\{O_i\}$, we choose products of two single-baryon operators, projected onto momenta $\vec{p}_1$ and $\vec{p}_2$, i.e.

$$\begin{aligned} B_\alpha &\equiv [rst]_\alpha = \epsilon_{ijk}\left(s^i C\gamma_5 t^j\right) r_k^\alpha, \\ (BB)_\Gamma(\vec{P},t) &= \sum_{\vec{x}} e^{i\vec{p_1}\cdot\vec{x}} B_1(\vec{x},t) C\Gamma P_+ \sum_{\vec{y}} e^{i\vec{p_2}\cdot\vec{y}} B_1(\vec{y},t)\,. \end{aligned} \tag{2}$$

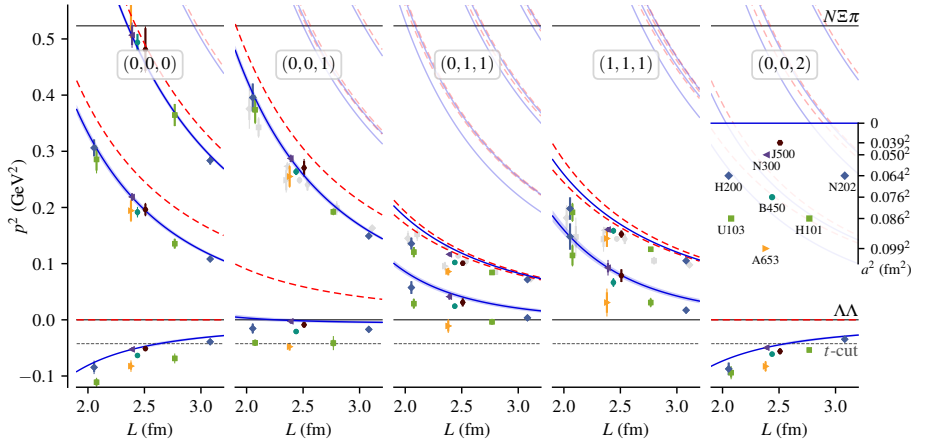


Figure 1. Finite-volume spectrum of the H dibaryon, plotted in terms of the centre-of-mass scattering momentum p^2 versus lattice extent L. Solid horizontal lines show the two- and three-particle thresholds, while dashed horizontal lines represent the t-channel cut of $-m_\pi^2/4$. The non-interacting spectrum is denoted by red dashed curves, and solid blue curves show the interacting spectrum determined in the continuum. The inset serves as a legend showing L and lattice spacings a^2 for the ensembles used in our calculation.

Here, r, s, t denote generic flavour indices, and the Dirac matrix Γ serves to describe different spin states. It is understood that a subsequent projection onto the desired representation of the flavour group is performed.

To compute correlation functions of our basis $\{O_i\}$ of interpolating operators, we use the so-called "distillation" technique[4] and Laplacian-Heaviside smearing[5], which allows us to compute "timeslice-to-all" propagators that are subsequently contracted to form hermitian correlator matrices $C_{ij} \equiv \langle O_i(t) O_j^\dagger(0) \rangle$ in all frames characterised by the total momentum $\vec{P}$. These numerical methods are key ingredients to alleviate the notorious problem of exponentially increasing statistical noise at large Euclidean times t[6,7].

The finite-volume energy levels that enter the quantisation condition are determined in each frame by solving a generalised eigenvalue problem[8–10] i.e.

$$C(\tau_D) v_n = \lambda_n \, C(\tau_0) v_n \tag{3}$$

for fixed timeslices τ_0 and τ_D. Subsequently, the eigenvectors v_n are used to transform the correlator matrix $C(t)$ into an approximately diagonal correlator $\tilde{C}(t)$ according to

$$\tilde{C}_{nm}(t) \equiv v_n^\dagger C(t) v_m . \tag{4}$$

Asymptotically, i.e. for $t \gg 0$, the diagonal elements of $\tilde{C}(t)$ fall off exponentially with a rate proportional to the finite-volume energy levels, E, for the given frame $\vec{P}$. Combining this information with the energy of two non-interacting baryons at rest, $2m_B$, yields the (squared) scattering momentum p^2 via

$$p^2 = \tfrac{1}{4}(E^2 - \vec{P} \cdot \vec{P}) - m_B^2 \, . \tag{5}$$

As a concrete example, we show in Fig. 1 the scattering momenta determined in different frames for the H dibaryon for different volumes and lattice spacings[11].

3 The H Dibaryon in Three-Flavour QCD

The H dibaryon is a spinless flavour-singlet state, originally proposed by Jaffe[12] in 1977 as a deeply bound system of six quarks with flavour content $uuddss$ ("hexaquark"). It has been speculated that a deeply bound scalar $uuddss$ hexaquark might qualify as a dark matter candidate, provided that its binding energy is of the order of a few hundred MeV[13–20]. An alternative interpretation of a possible H dibaryon is that of a weakly bound state of two Λ hyperons, in which case one expects the binding energy to be of order 10 MeV. There is a long history of calculations studying whether the H dibaryon is a prediction of QCD. Results for the binding energy B_H from the more recent calculations with dynamical quarks[21–27,11] vary considerably, with estimates ranging from a few MeV up to 75 MeV, depending on the methodology and/or the value of the pion mass. Recently, employing near-physical pion and kaon masses, the HAL QCD Collaboration reported that the Λ-Λ interaction is only weakly attractive and does not sustain a bound or resonant dihyperon[27].

We have performed an extensive investigation of the H dibaryon, focusing initially on the question whether a bound state exists at the SU(3)-symmetric situation with degenerate up, down and strange quarks, corresponding to $m_\pi = m_K \simeq 415\,\mathrm{MeV}$[11]. To this end we have computed the energy levels in the H dibaryon channel on eight different gauge ensembles of O(a)-improved Wilson fermions generated by the CLS effort[28]. Results for the scattering momenta in different frames are shown in Fig. 1. In order to determine the scattering amplitude, we truncate the finite-volume quantisation condition to its simplest form, with a single channel and keeping only S-wave, neglecting D-wave and higher partial waves. Eq. 1 then becomes

$$p\cot\delta_0(p^2) = \frac{2}{\sqrt{\pi}L\gamma} Z_{00}^{\vec{P}L/(2\pi)}\left(1, \left(\frac{pL}{2\pi}\right)^2\right), \tag{6}$$

with $\delta_0(p^2)$ the scattering phase shift, and $Z_{00}^{\vec{D}}$ a generalised zeta function. When performing global fits to the spectra from all of the ensembles with different lattice spacings and different volumes, we have used a fit *ansatz* which assumes that $p\cot\delta_0(p^2)$ can be described by a polynomial in p^2 with coefficients that are affine functions of the squared lattice spacing a^2:

$$p\cot\delta_0(p^2) = \sum_{i=0}^{N-1} c_i p^{2i}, \quad c_i = c_{i0} + c_{i1}a^2. \tag{7}$$

This fit provides a good description of our finite-volume energy levels, with the exception of the excited states in frames $(0,1,1)$ and $(1,1,1)$, which are strongly affected by the neglected D wave. Sending $a \to 0$, we obtain the continuum finite-volume energy levels shown as the blue curves in Fig. 1. The binding energy is determined from the poles of the scattering amplitude below threshold, which can be found as solutions to $p\cot\delta(p) = -\sqrt{-p^2}$.

In our 2021 publication[11] we reported the existence of a bound H dibaryon and, at the same time, a strong dependence of its binding energy B_H on the lattice spacing, with estimates varying between 35 MeV at the coarsest lattice spacing and a few MeV in the continuum limit. Our final result of $B_H = 4.56 \pm 1.13 \pm 0.63\,\mathrm{MeV}$ is significantly smaller than what most other lattice calculations observed[21–25]. These findings rule out a deeply

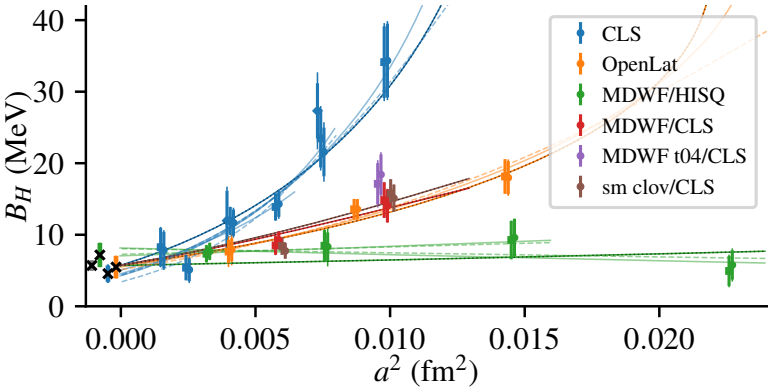


Figure 2. Continuum extrapolations of the H dibaryon's binding energy computed for different combinations of sea and valence quark actions. Blue data points show the results from our published calculation based on the CLS ensembles with O(a)-improved Wilson fermions both in the sea and valence sectors[11]. All data were computed in three-flavour QCD with $m_\pi = m_K \simeq 415\,\text{MeV}$.

bound H dibaryon as envisaged in Jaffe's paper, provided that the pion mass dependence and effects from SU(3) breaking are small.

The unexpectedly strong dependence of B_H on the lattice spacing prompted us to perform a detailed scaling study, by repeating the calculation for different discretisations of the QCD action. To this end we have used gauge ensembles that were generated by the OpenLat effort[29] with an exponentiated Clover term[30] added to the Wilson-Dirac operator. In addition, we explored several mixed action setups that combine different discretisations in the sea and valence sectors. For this purpose, we mainly used Möbius domain wall valence quarks together with gauge ensembles based on rooted staggered (HISQ) or, alternatively, the original CLS ensembles with varying degrees of Wilson flow applied to the gauge links. In order to study the effect that smeared gauge links may have on the scaling behaviour, we also used O(a)-improved Wilson valence quarks on flowed CLS ensembles. The resulting extrapolations to the continuum limit for each combination is shown in Fig. 2. While our original choice of lattice action shows the largest discretisation effects, the binding energies determined from the mixed action setups extrapolate to a consistent set of results in the continuum limit. This corroborates the finding in our original paper[11].

4 Nucleon-Nucleon Interactions

In Nature, there is one nucleon-nucleon bound state: the deuteron, bound by just 2.2 MeV. Understanding this and nucleon-nucleon scattering from *ab initio* QCD are essential steps in deriving nuclear physics from the Standard Model of particle physics. As a precursor to this challenging problem, we have been studying two-nucleon systems at the SU(3)-symmetric point, where the signal-to-noise problem is less severe and the heavier pion mass permits the use of smaller volumes.

Nucleon-nucleon systems at heavier-than-physical pion masses have been studied on the lattice for over a decade; however, it is now understood that the earlier calculations that found deeply bound states used unreliable methods to determine the finite-volume

spectrum[31,32]. Our use of a matrix of correlation functions together with the generalised eigenvalue problem allows us to avoid this problem by simultaneously extracting all of the low-lying elastic nucleon-nucleon states. As we also control discretisation effects, our data set allows us to obtain reliable results for QCD at our unphysical point.

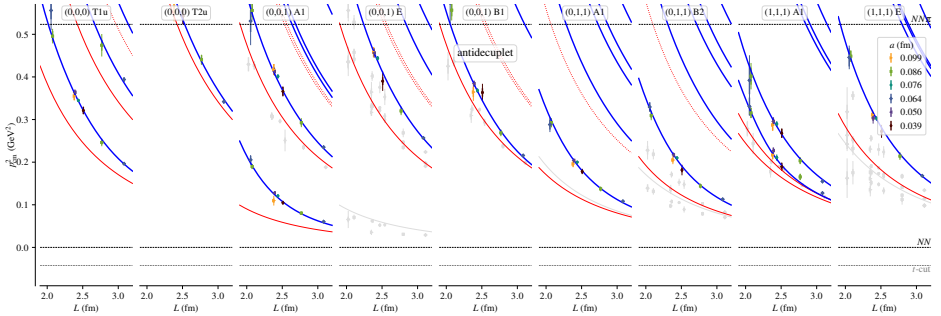


Figure 3. Finite-volume spectrum of two nucleons with $I = 0$: energy levels in little-group irreps relevant for the spin-zero odd partial waves ${}^1P_1$ and ${}^1F_3$. See the caption of Fig. 1. The gray points and noninteracting-level curves indicate levels with the same finite-volume quantum numbers that primarily couple to spin-one interpolating operators.

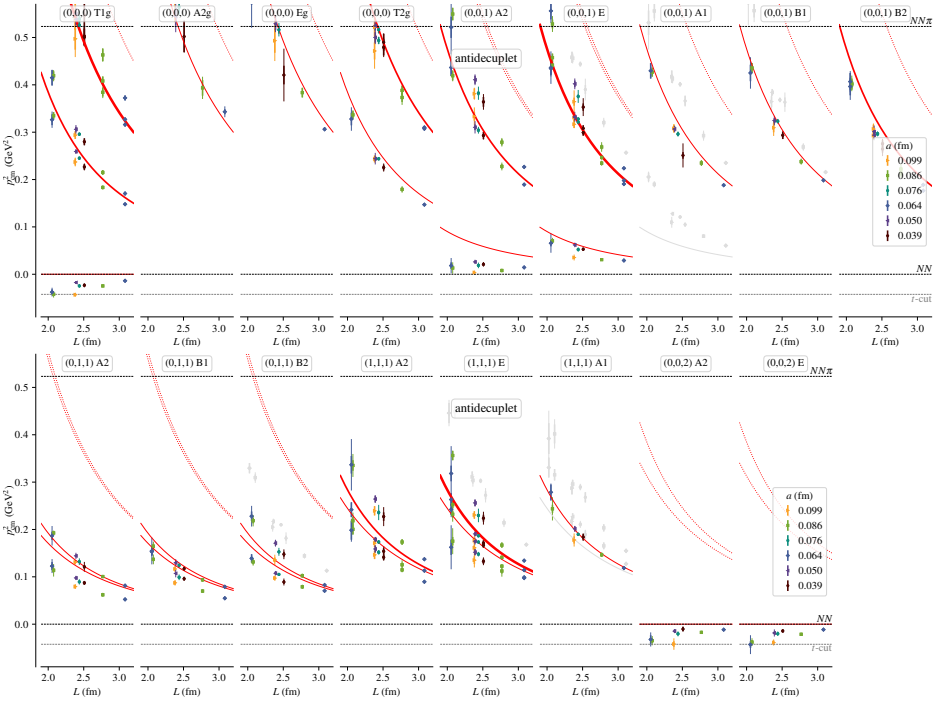


Figure 4. Finite-volume spectrum of two nucleons with $I = 0$: energy levels in little-group irreps relevant for the spin-one even partial waves ${}^3S_1$, ${}^3D_1$, ${}^3D_2$, and ${}^3D_3$. See the caption of Fig. 1. The thickness of the red curves indicating noninteracting levels is proportional to the degeneracy of that level.

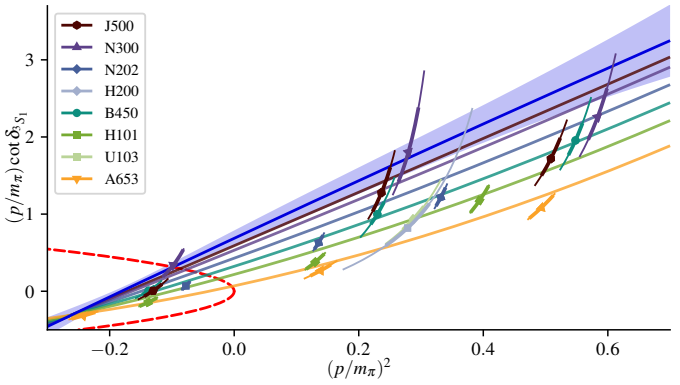


Figure 5. Phase shift for $NN\,{}^3S_1$, in the approximation that D waves and partial wave mixing vanish: $p\cot\delta(p)$ versus p^2, in units of the pion mass. Darker and more purple points and curves correspond to finer lattice spacings and pale points indicate small-volume ensembles. The blue curve with error band is the continuum limit of the fit. An intersection with the upper branch of the red dashed curve corresponds to a virtual state pole.

Partial-wave scattering states of two nucleons are characterised by the total spin S, orbital angular momentum ℓ, and total angular momentum J, which are denoted using the spectroscopic notation ${}^{2S+1}\ell_J$. As isospin I is a good symmetry in our calculation, the overall antisymmetry implies $S+\ell+I$ must be an odd number, and partial waves with spin zero do not couple to those with spin one. Although the finite-volume symmetries in moving frames allow mixing between states with different parities, the fact that both the scattering amplitude and the finite-volume matrix $\mathcal{F}$ are diagonal in spin implies that the determinant in Eq. 1 factorises and (once the spins of the energy levels have been identified) spin zero can be analysed separately from spin one. Thus, four sets of energy levels can be analysed independently: those with $(I,S)=(0,0)$, $(0,1)$, $(1,0)$, and $(1,1)$.

The simplest analysis is in the spin-zero sector (which has fewer partial waves and no partial wave coupling) and odd partial waves (which are weaker and less affected by discretisation effects than the S-wave). These energy levels are shown in Fig. 3. The positive energy shifts indicate both the ${}^1P_1$ and ${}^1F_3$ partial waves are repulsive. Modelling these two phase shifts with two parameters each and applying finite-volume quantisation, we are able to obtain a good fit quality (blue curves) assuming no discretisation effects.

The spin-one, even partial wave sector that potentially contains a bound deuteron is more complex, containing three different D waves, and mixing between the ${}^3S_1$ and ${}^3D_1$ partial waves. Fig. 4 shows the relevant energy levels, which number more than three hundred. In many of the lowest-lying levels, a pattern of discretisation effects similar to the case of the H dibaryon is visible.

A preliminary simplified analysis of the ${}^3S_1$ partial wave, based on the helicity average of ground-state levels[33] and neglecting all higher partial waves, is shown in Fig. 5. This simple fit describes the broad features of our data, although the fit quality could be improved. The interaction is attractive but insufficient to produce a bound state: instead, we find a virtual state pole that is near threshold on our coarsest lattice spacing but moves further below the threshold as the continuum limit is approached. A more sophisticated analysis including partial-wave mixing and higher partial waves is ongoing.

5 Conclusions

Finite-volume quantisation is a powerful formalism that allows for the *ab initio* study of baryon-baryon interactions. In this contribution we have reported on our ongoing project focusing on Λ-Λ and nucleon-nucleon interactions. Even though our results have been obtained in three-flavour QCD and at unphysically heavy pion mass of about 415 MeV, our calculations have consistently shown that the scenario of a deeply bound H dibaryon is highly unlikely. The spin-one even partial wave sector in nucleon-nucleon interactions (which contains the deuteron) is considerably more complex, and our preliminary analysis that ignores higher partial waves and partial wave mixing does not yet allow us to draw firm conclusions. These issues will be addressed in future work which is also focused on lowering the pion mass towards its physical value. In addition, we have applied finite-volume quantisation also to the case of charmed tetra-quarks that have generated a lot of interest recently.

Acknowledgements

We thank our colleagues within the BaSc Collaboration, in particular André Walker-Loud and Jamie Hudspith, for their participation in the scaling study for the H dibaryon. Calculations have been performed on JUWELS and JUWELS Booster at JSC as part of project HINTSPEC. This research was partly supported by DFG through the Cluster of Excellence "Precision Physics, Fundamental Interactions and Structure of Matter" (PRISMA+ EXC 2118/1) funded within the German Excellence Strategy (Project ID 39083149), as well as the Collaborative Research Centers SFB 1044 "The low-energy frontier of the Standard Model" and CRC-TR 211 "Strong-interaction matter under extreme conditions" (Project ID 315477589). MP gratefully acknowledges support from the Department of Science and Technology, India, SERB Start-up Research Grant No. SRG/2023/001235 and Department of Atomic Energy, India. SP is supported by Los Alamos National Lab and US Department of Energy under the project "Accelerating simulations of lattice QCD and neutrino scattering using AI/ML". We are grateful to our colleagues within the CLS and OpenLat initiatives for sharing ensembles.

References

1. M. Lüscher, *Two particle states on a torus and their relation to the scattering matrix*, Nucl. Phys., **B354**, 531-578, 1991.
2. M. Lüscher, *Signatures of unstable particles in finite volume*, Nucl. Phys., **B364**, 237-254, 1991.
3. C. Morningstar, J. Bulava, B. Singha, R. Brett, J. Fallica, A. Hanlon, and B. Hörz, *Estimating the two-particle K-matrix for multiple partial waves and decay channels from finite-volume energies*, Nucl. Phys. B, **924**, 477-507, 2017.
4. M. Peardon, J. Bulava, J. Foley, C. Morningstar, J. Dudek, R. G. Edwards, B. Joó, H.-W. Lin, D. G. Richards, and K. J. Juge, *Novel quark-field creation operator construction for hadronic physics in lattice QCD*, Phys. Rev. D, **80**, 054506, 2009.

5. C. Morningstar, J. Bulava, J. Foley, K. J. Juge, D. Lenkner, M. Peardon, and C. H. Wong, *Improved stochastic estimation of quark propagation with Laplacian Heaviside smearing in lattice QCD*, Phys. Rev. D, **83**, 114505, 2011.
6. H. W. Hamber, E. Marinari, G. Parisi, and C. Rebbi, *Considerations on Numerical Analysis of QCD*, Nucl. Phys. B, **225**, 475, 1983.
7. G. P. Lepage, *The Analysis of Algorithms for Lattice Field Theory*, in: Theoretical Advanced Study Institute in Elementary Particle Physics, June 1989.
8. C. Michael, *Adjoint Sources in Lattice Gauge Theory*, Nucl. Phys. B, **259**, 58-76, 1985.
9. M. Lüscher and U. Wolff, *How to Calculate the Elastic Scattering Matrix in Two-dimensional Quantum Field Theories by Numerical Simulation*, Nucl. Phys. B, **339**, 222-252, 1990.
10. B. Blossier, M. Della Morte, G. von Hippel, T. Mendes, and R. Sommer, *On the generalized eigenvalue method for energies and matrix elements in lattice field theory*, JHEP, **04**, 094, 2009.
11. J. R. Green, A. D. Hanlon, P. M. Junnarkar, and H. Wittig, *Weakly bound H dibaryon from SU(3)-flavor-symmetric QCD*, Phys. Rev. Lett., **127**, 242003, 2021.
12. R. L. Jaffe, *Perhaps a Stable Dihyperon*, Phys. Rev. Lett., **38**, 195-198, 1977, [Erratum: Phys. Rev. Lett., **38**, 617, 1977].
13. G. R. Farrar, *Stable Sexaquark*, Aug. 2017, arXiv:1708.08951.
14. C. Gross, A. Polosa, A. Strumia, A. Urbano, and W. Xue, *Dark Matter in the Standard Model?*, Phys. Rev. D, **98**, no. 6, 063005, 2018.
15. G. R. Farrar, *A precision test of the nature of Dark Matter and a probe of the QCD phase transition*, May 2018, arXiv:1805.03723.
16. E. W. Kolb and M. S. Turner, *Dibaryons cannot be the dark matter*, Phys. Rev. D, **99**, no. 6, 063519, 2019.
17. S. D. McDermott, S. Reddy, and S. Sen, *Deeply bound dibaryon is incompatible with neutron stars and supernovae*, Phys. Rev. D, **99**, no. 3, 035013, 2019.
18. J. P. Lees et al., *Search for a Stable Six-Quark State at BABAR*, Phys. Rev. Lett., **122**, no. 7, 072002, 2019.
19. K. Azizi, S. S. Agaev, and H. Sundu, *The Scalar Hexaquark $uuddss$: a Candidate to Dark Matter?*, J. Phys. G, **47**, no. 9, 095001, 2020.
20. G. R. Farrar, Z. Wang, and X. Xu, *Dark Matter Particle in QCD*, July 2020, arXiv:2007.10378.
21. S. R. Beane et al., *Evidence for a Bound H-dibaryon from Lattice QCD*, Phys. Rev. Lett., **106**, 162001, 2011.
22. S. R. Beane et al., *Present Constraints on the H-dibaryon at the Physical Point from Lattice QCD*, Mod. Phys. Lett. A, **26**, 2587-2595, 2011.
23. S. R. Beane, E. Chang, S. D. Cohen, W. Detmold, H. W. Lin, T. C. Luu, K. Orginos, A. Parreño, M. J. Savage, and A. Walker-Loud, *Light Nuclei and Hypernuclei from Quantum Chromodynamics in the Limit of SU(3) Flavor Symmetry*, Phys. Rev. D, **87**, 034506, 2013.
24. T. Inoue, N. Ishii, S. Aoki, T. Doi, T. Hatsuda, Y. Ikeda, K. Murano, H. Nemura, and K. Sasaki, *Bound H-dibaryon in Flavor SU(3) Limit of Lattice QCD*, Phys. Rev. Lett., **106**, 162002, 2011.

25. T. Inoue, S. Aoki, T. Doi, T. Hatsuda, Y. Ikeda, N. Ishii, K. Murano, H. Nemura, and K. Sasaki, *Two-Baryon Potentials and H-Dibaryon from 3-flavor Lattice QCD Simulations*, Nucl. Phys. A, **881**, 28-43, 2012.
26. A. Francis, J. R. Green, P. M. Junnarkar, Ch. Miao, T. D. Rae, and H. Wittig, *Lattice QCD study of the H dibaryon using hexaquark and two-baryon interpolators*, Phys. Rev., **D99**, no. 7, 074505, 2019.
27. K. Sasaki et al., $\Lambda\Lambda$ *and* $N\Xi$ *interactions from lattice QCD near the physical point*, Nucl. Phys. A, **998**, 121737, 2020.
28. M. Bruno et al., *Simulation of QCD with* $N_f = 2 + 1$ *flavors of non-perturbatively improved Wilson fermions*, JHEP, **02**, 043, 2015.
29. A. Francis, F. Cuteri, P. Fritzsch, G. Pederiva, A. Rago, A. Shindler, A. Walker-Loud, and S. Zafeiropoulos, *Progress in generating gauge ensembles with Stabilized Wilson Fermions*, in: Proceedings of The 40th International Symposium on Lattice Field Theory – PoS(LATTICE2023), **453**, 048, 2023.
30. A. Francis, P. Fritzsch, M. Lüscher, and A. Rago, *Master-field simulations of O(a)-improved lattice QCD: Algorithms, stability and exactness*, Comput. Phys. Commun., **255**, 107355, 2020.
31. T. Iritani et al., *Mirage in Temporal Correlation functions for Baryon-Baryon Interactions in Lattice QCD*, JHEP, **10**, 101, 2016.
32. T. Iritani, S. Aoki, T. Doi, T. Hatsuda, Y. Ikeda, T. Inoue, N. Ishii, H. Nemura, and K. Sasaki, *Are two nucleons bound in lattice QCD for heavy quark masses? Consistency check with Lüscher's finite volume formula*, Phys. Rev., **D96**, no. 3, 034521, 2017.
33. R. A. Briceño, Z. Davoudi, T. Luu, and M. J. Savage, *Two-nucleon systems in a finite volume. II.* ${}^3S_1 - {}^3D_1$ *coupled channels and the deuteron*, Phys. Rev. D, **88**, 114507, 2013.

Non-Perturbative Renormalisation of Gluon and Quark Flavour Singlet Operators

Constantia Alexandrou[1,2], Simone Bacchio[2], Martha Constantinou[3], Jacob Finkenrath[4], Karl Jansen[2,5], Giannis Koutsou[2], Bhavna Prasad[2], and Gregoris Spanoudes[1]

[1] Department of Physics, University of Cyprus, P.O. Box 20537, 1678 Nicosia, Cyprus

[2] Computation-based Science and Technology Research Center, The Cyprus Institute, 20 Konstantinou Kavafi Street, 2121 Aglantzia , Cyprus

[3] Department of Physics, Temple University, Philadelphia, PA 19122 - 1801, USA

[4] European Organization for Nuclear Research, CERN, CH-1211 Genève 23, Switzerland
E-mail: j.finkenrath@cern.ch

[5] CQTA, Deutsches Elektronen-Synchrotron DESY, Platanenallee 6, 15738 Zeuthen, Germany

Computations on state-of-the-art supercomputers, such as the cluster JUWELS at the Jülich Supercomputing Centre enables us to study fundamental building blocks of our universe. The allocation was used for computation needed for our effort on non-perturbative renormalisation, an integral part of our precision study of hadronic structure using lattice QCD. State-of-the-art lattices with two degenerate light, a strange, and a charm quark (N_f=2+1+1) with masses tuned to their physical values (physical point), are being used to study hadron structure at unprecedented accuracy. The project is enabling their non-perturbative renormalisation, a crucial component for relating lattice results to continuum quantities. To this end, we have carried out dedicated simulations with four degenerate quarks (N_f=4), used to compute renormalisation coefficients with lattice spacing a =0.093, 0.087, 0.08, 0.068, 0.057, and 0.049 fm, at multiple values of the quark mass allowing us to take the chiral limit of the renormalisation factors. The renormalisation factors are in turn being used to obtain the continuum limit of a series of quantities important in hadron structure, including: i) nucleon charges and moments of nucleon Parton Distribution Functions (PDFs), ii) the gluon momentum fraction renormalised non-perturbatively, and iii) nucleon form factors.

1 Motivation

One of the fundamental goals of research in the field of Hadron Physics is the quantitative description of the rich internal structure of hadrons, which are the building blocks of the visible Universe. State-of-the-art numerical simulations of Quantum Chromodynamics (QCD) formulated on a Euclidean 4-dimensional lattice (lattice QCD) provide a rigorous approach for an *ab initio* nonperturbative study of hadron structure, which captures the full dynamics and interactions of the constituent particles. Hadron form factors and distribution functions are key observables that can be determined directly from the evaluation of matrix elements in lattice QCD. The accurate computation of these key quantities not only provides input to phenomenological models and ongoing experiments but also gives predictions on observables that are not easily accessible experimentally, such as nucleon σ-terms, strange electromagnetic and axial form factors, polarised parton distributions[1], and properties of unstable particles, e.g., pion and kaon[2].

Significant progress has been accomplished in lattice QCD during the last decade due to algorithmic improvements in combination with access to more powerful computers. A major milestone is the access to simulations at quark masses tuned to their physical values. This enables the extraction of physical quantities free of uncontrolled systematic errors from chiral extrapolations. However, there are still many challenges that need to be addressed, such as the study of lattice systematic errors related to finite lattice spacing and volume, excited states, operator mixing and renormalisation, reaching higher statistical accuracy and the computation of more challenging observables, including transition form factors and properties of resonances. The computational project on JUWELS allows us to address operator mixing and the renormalisation of several quantum composite operators, which are relevant to the hadron structure investigations.

We focus on first-principle investigation of a family of quark and gluon distribution functions (DFs), which encode the momentum and spin decomposition of hadrons: parton distribution functions (PDFs), generalised parton distribution functions (GPDs) and transverse-momentum dependent parton distribution functions (TMDs). All three types of DFs are necessary in order to unravel the three-dimensional hadron picture. The studies of DFs are at the forefront of international activity both theoretically and experimentally. Novel methods for extracting full DFs on the lattice have been employed in recent years by calculating matrix elements of nonlocal operators[3]. Accurate determinations of these quantities from lattice QCD can significantly complement the experimental investigations by providing input into the analysis and interpretation of the experimental data.

Furthermore, Mellin moments of distribution functions are fundamental point of contact between experiment and theory, and lattice QCD is an ideal framework of determining these quantities nonperturbatively[4]. The study of such moments on the lattice entails the calculation of matrix elements for a large variety of higher-derivative gluon and quark local operators.

We target to extract nonperturbative renormalisation functions and mixing coefficients for a variety of quark and gluon local and nonlocal operators entering the investigations of DFs and their moments. Renormalisation is an essential ingredient for obtaining reliable theoretical predictions from lattice simulations matched to the physical values of hadron observables. A particular goal is the renormalisation of "disconnected" contributions, which are typically more noisy and thus, large statistics are needed. For this purpose, we conducted long Markov chain Monte Carlo simulations of relatively small lattice sizes at various lattice spacings. Our simulation setup combined with the features of the JUWELS Cluster module are ideal for generating such ensembles, as it is evident from previous allocations of our group on JUWELS. Resources from this allocation also covered the calculation of nonperturbative renormalisation functions for gluon and quark momentum fractions, nucleon charges, and quark PDFs.

2 Scientific Results

State-of-the-art simulations of lattice Quantum Chromodynamics (QCD), are being carried out at physical values of the up and down, strange, and charm quarks[5–7], and are being used to obtain hadronic matrix elements at ever increasing statistical precision. Using the twisted mass fermion formulation of lattice QCD, we are currently producing state-of-the-art results for hadronic matrix elements, such as the nucleon axial, scalar, and tensor

charges[8–10], moments of parton distribution functions (PDFs)[11], quark and gluon contributions to hadronic momentum and spin[12,13], and the x-dependence of the nucleon[14,15] and Δ-baryon[16] PDFs. For obtaining such high-accuracy results, a crucial component is the non-perturbative renormalisation of the lattice matrix elements, needed to relate them with physical, continuum quantities.

Within the multi-year computational project `renormglue` on JUWELS we are computing renormalisation factors non-perturbatively to a high precision. The full renormalisation program involves calculating the factors for multiple quark masses enabling us to take the chiral limit. With the renormalisation factors available at multiple lattice spacings, this allows for the continuum limit to be taken for all hadronic quantities available to us. Due to the computation provided by JUWELS we made significant progress in obtaining the renormalisation factors non-perturbatively at multiple lattice spacing.

2.1 Dedicated N_f=4 Simulations

Our renormalisation program uses the $\mathrm{RI}'_{\mathrm{MOM}}$ scheme[17] as explained in Ref. 18. Dedicated N_f=4 ensembles were generated using JUWELS at the same values of the coupling β as used for the ensembles with which the matrix elements to be renormalised were obtained on. Multiple values of the quark masses are used to extrapolate the renormalisation factors to the chiral limit. The complete set of ensembles available with N_f=4 are shown in Fig. 1. During allocation, resources from `renormglue` were used to generate the ensembles at several lattice spacing, $a \simeq$0.086 0.069 0.058 0.05 fm, labelled as "A′", "C", "D" and "E".

Our renormalisation program improves on the non-perturbative estimates of the renormalisation factors by subtracting finite lattice effects[19–21]. The latter are computed to one-loop in perturbation theory and to all orders in the lattice spacing, $\mathcal{O}(g^2\, a^\infty)$. These ar-

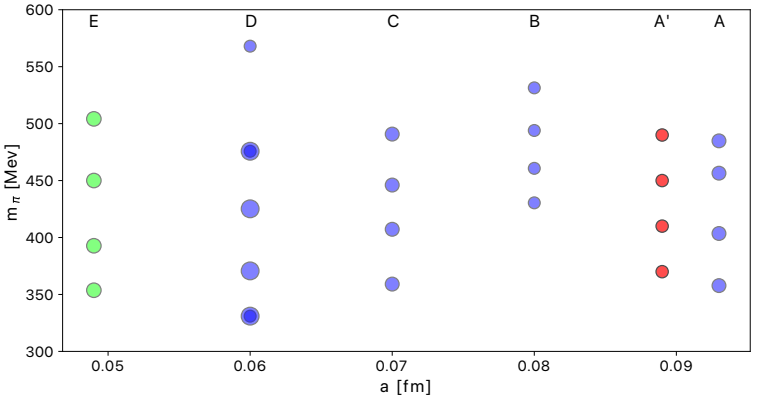


Figure 1. Parameters of the N_f=4 ensembles used to compute the renormalisation factors non-perturbatively. The size of the circles indicates the physical spatial extent L between 2.3 fm and 3.8 fm. We indicate at the top of the figure the labels used to refer to each lattice spacing. The "E" ensembles, shown in green, were generated on Juwels Cluster within this project during the past allocation period. The "A′" ensembles, show in red, were generated in the last allocation.

tifacts are present in the non-perturbative vertex functions of the fermion propagator and fermion operators under study. Furthermore, we note that we compute both non-singlet and singlet renormalisation functions, the latter of which require the calculation of disconnected contributions in addition to the connected, and are therefore more computationally demanding.

Within the allocation renormalisation factors have been extracted for the non-singlet quark bilinear operators, namely scalar (Z_S), pseudoscalar (Z_P), vector (Z_V), axial-vector (Z_A) and tensor (Z_T), as well as of the quark field (Z_q), on the generated N_f=4 ensembles. Here, we show factors extracted from the N_f=4 ensembles at the E lattice with twisted quark mass $a\mu_q = 0.0035$ and 0.0056. In Fig. 2, we indicatively plot the renormalisation factors of the vector and axial currents, as a function of the initial renormalisation momentum scale p^2. Here, we apply our program to reduce systematic errors related to discretisation effects by subtracting artifacts calculated in one-loop lattice perturbation theory. Chiral extrapolations are then performed with at least three μ values, followed by a conversion to the reference scheme of $\overline{\text{MS}}$ and evolution to the reference scale of 2 GeV.

A dedicated analysis of the renormalisation factors for all non-singlet quark bilinear operators is ongoing using all generated ensembles from the current and past allocations. This includes data from simulations of all six lattice spacings. The current status of the calculation is: The simulations of the $N_F = 4$ "A", "B","C", "D", and "E" ensembles have been completed; for the "A′", three out of four ensembles of different quark masses have been finished. The production of the remaining ensemble is ongoing. Given the plethora of data coming from several ensembles, a more sophisticated analysis has been initiated with the goal of further improving the accuracy of the extracted values. First results from this analysis have been applied in Ref. 10. The full analysis will be provided in a forthcoming publication. Further ongoing analyses by our group applying the generated $N_f = 4$ ensembles include the renormalisation of the singlet quark bilinear operators and of higher derivative operators.

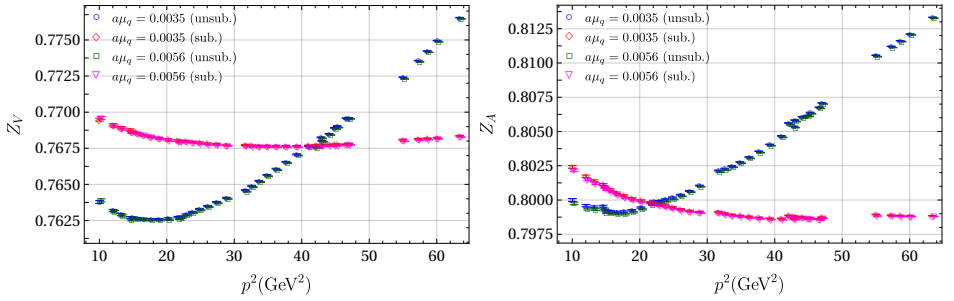


Figure 2. Renormalisation factors of the vector operator (Z_V, left) and axial-vector operator (Z_A, right) vs p^2 with (sub.) or without (unsub.) subtracting one-loop lattice artifacts for the twisted mass parameter $a\mu_q$ as indicated in the legend.

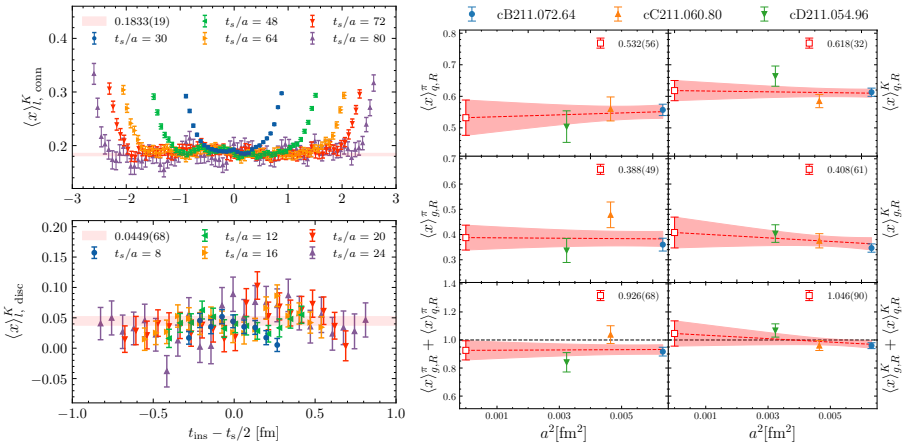


Figure 3. On the left panel we show the bare ratios of the light-quark momentum fraction in the kaon dependence on the sink-source time separation t_{sink} for the ensemble at intermediate lattice spacings. In the upper left panel, we show the ratio for the connected contribution and in the lower panel, the ratio for the disconnected contribution. We show results for several values of t_{sink}, from $t_{sink}/a = 30$ to $t_{sink}/a = 80$. The red bands show the result after model averaging of fits to a constant when varying the ranges of t_{sink} used. Continuum limit extrapolation are showing on the right panel for the pion (left part) and the kaon (right part). We present our results for the total quark and gluon contributions, as well as the momentum sum rule. The blue filled circles are results for ensemble B[13], the orange upwards triangles for ensemble C and the green downwards triangles for ensemble D. The open symbol is the result after model averaging between constant and linear fit. Further details can be found in the Supplemental Material.

2.2 Quark and Gluon Momentum Fraction in the Pion and Kaon

The dedicated program to calculate renormalisation factors, using the ensemble generated on JUWELS, enabled us to calculate the first complete momentum decomposition for both the pion and the kaon. This was done in terms of their quark and gluon constituents, performed within lattice QCD at the physical point[22].

Here, we used three ensembles with $N_f = 2 + 1 + 1$ quark flavours with their masses tuned to reproduce the physical light, strange and charm quark masses.

We use a model average for the continuum extrapolation performed using both a constant and a linear fit in the squared lattice spacing, a^2, see Fig. 3. Our results for $\langle x \rangle_g^{\pi,K}$ indicate a similar momentum fraction carried by gluons in the kaon and the pion. We find that the total momentum fraction carried by quarks is $0.532(56)$ and $0.618(32)$ and by gluons $0.388(49)$ and $0.408(61)$ in the pion and in the kaon, respectively, in the $\overline{\mathrm{MS}}$ scheme and at the renormalisation scale of 2 GeV.

The gluon momentum fraction have larger errors than the corresponding quark momentum fraction $\langle x \rangle_q^{\pi,K}$, which tends to be smaller in the pion. This indicates a possible larger momentum fraction carried by gluons in the pion as compared to the kaon.

The work presents a remarkable achievement as we were able to conduct a fully consistent theoretical calculation that allows us to directly compare the pion and the kaon at the level of their quark and gluon structure. In particular, the momentum sum rule can be tested by computing all components from first principles. Namely, we found that the momentum sum is 0.926(68) for the pion and 1.046(90) for the kaon, verifying the momentum

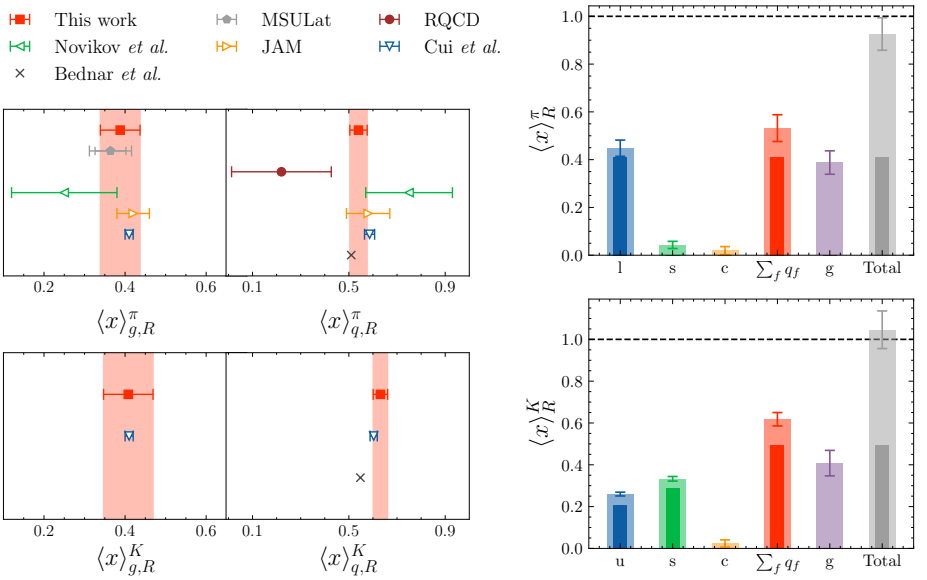


Figure 4. The left panel shows comparison of the results of this work, with other available data, both from phenomenology and from lattice QCD. All results are given in the $\overline{\text{MS}}$ scheme at the scale of $\mu = 2$ GeV. The upper panels show the results for the pion gluon (left) and quark (right) momentum fractions, $\langle x \rangle^{\pi}_{\text{g,R}}$ and $\langle x \rangle^{\pi}_{\text{q,R}}$, respectively. The lower panels show the corresponding results for the kaon. The red filled squares show the results of this work with the red band the associated error band. Recent results from phenomenological analyses of PDFs data are given by open symbols: left green triangle (Novikov *et al.*[23]) and right orange triangle (JAM Collaboration[24]). The result while the result based on the LFWF is represented by the down blue triangle (Cui *et al.*[25]), while using the DSE[26] by the black cross, where no error is provided. Recent lattice QCD results extrapolated to the continuum limit are given by the brown filled circle (RQCD[27]) and the gray pentagon (MSULat[28]). The right panel shows the quark and gluon momentum fractions for the pion (upper panel) and kaon (lower panel) obtained in the continuum. Inner bars represent only the connected contributions, while the outer bars show the total, including disconnected contributions.

sum rule. We stress that prior to the current work, there was no such decomposition into quark and gluon parts available of $\langle x \rangle^K$ using a first-principles calculation.

Moreover we demonstrated that the momentum fraction carried by gluons and sea quarks (the disconnected contributions) are important components of the two lightest pseudo-Goldstone bosons, and added together with the valence contributions verify the momentum sum rule for both cases.

2.3 Nucleon Charges and Form Factors

Results for other key quantities, such as the electromagnetic and axial form factors of the nucleon[29–32, 10] have been extracted using the renormalisation factors computed within this project. The allocation allowed us to extend previous calculations by including three physical point ensembles with lattice spacings 0.080 fm, 0.068 fm and 0.057 fm, and spatial sizes 5.1 fm, 5.44 fm, and 5.47 fm, respectively, yielding $m_\pi L > 3.6$. This enables us to take the continuum limit of the axial (G_A), induced pseudoscalar (G_P), and pseudoscalar (G_5) form factors, as shown in Fig. 5. Errors include statistical and systematics after a

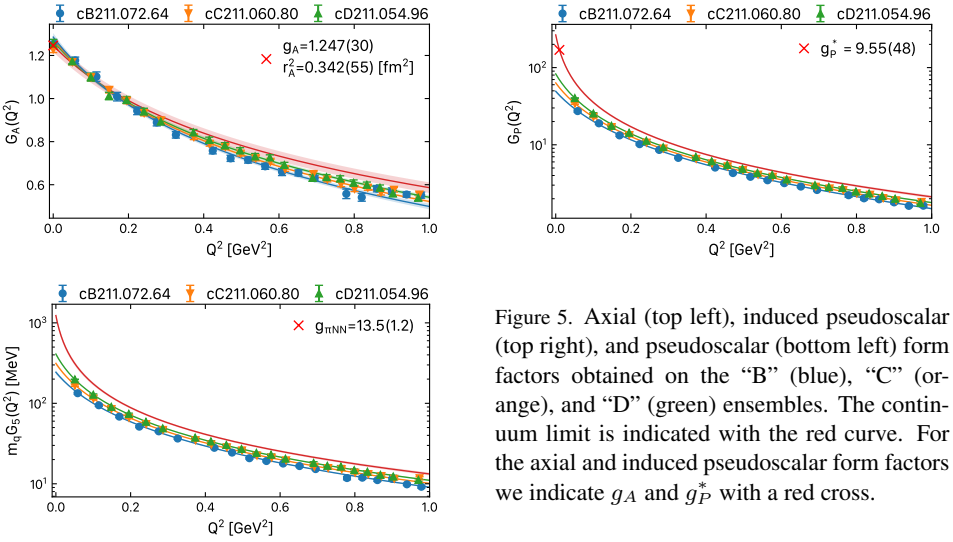


Figure 5. Axial (top left), induced pseudoscalar (top right), and pseudoscalar (bottom left) form factors obtained on the "B" (blue), "C" (orange), and "D" (green) ensembles. The continuum limit is indicated with the red curve. For the axial and induced pseudoscalar form factors we indicate g_A and g_P^* with a red cross.

detailed analysis of excited state effects and fit models to the Q^2-dependence of the form factors which includes dipole and z-expansion.

We find for the nucleon axial charge $g_A = 1.245(28)(14)$, for the axial radius $\langle r_A^2 \rangle = 0.339(48)(06)$ fm^2, for the pion-nucleon coupling constant $g_{\pi NN} \equiv \lim_{Q^2 \to -m_\pi^2} G_{\pi NN}(Q_2) = 13.25(67)(69)$ and for $G_P(0.88 m_\mu^2) \equiv g_P^* = 8.99(39)(49)$. The results on $G_A(Q^2)$ are in good agreement with other recent lattice QCD studies. Note, that by taking the continuum limit using ensembles at the physical-point mass, we avoid a chiral extrapolation that in the nucleon sector can lead to an uncontrolled systematic error. This allows to directly access cut-off effects. Namely for $G_A(Q^2)$, we found that the effects for the range of lattice spacings used are mild, ranging from not detectable within our errors at low Q^2 to slightly positive at high Q^2. On the other hand, the induced pseudoscalar and the pseudoscalar form factors exhibit similar large cut-off effects that can be traced back to the known $O(a^2)$ artifacts on the pion mass pole, which can be parameterised and addressed in our continuum extrapolation fits. An important outcome of this study is that in the continuum limit, all cut-off effects are safely eliminated as expected.

3 Conclusion

A comprehensive computation of nonperturbative renormalisation functions for a diverse array of quantum-field composite operators entering hadron observables, have been performed using resources from allocations on JUWELS. Our calculations contributed to the elimination of a major source of systematic uncertainty in lattice QCD simulations coming from the process of renormalising hadron matrix elements, which are essential for making contact to physical measurable observables. By leveraging our generated $N_f = 4$ ensembles across multiple lattice spacings and quark masses, we have conducted chiral and continuum extrapolations that are crucial for minimising systematic errors inherent in

lattice calculations. Looking ahead, we aim to extend our efforts to extract the renormalisation of higher moments of parton distribution functions, specifically the 2nd and 3rd moments, which present additional challenges due to operator mixing and the complexity introduced by numerous covariant derivatives. With the computational capacity of the next generation of supercomputers, such as JUPITER will deliver, this can be addressed by enhancing statistical outcomes and further deepening our understanding of the building blocks of our universe, the structure of hadrons.

Acknowledgements

We thank all members of ETMC for the most enjoyable collaboration. C. A. acknowledges partial support by the project 3D-nucleon, id number EXCELLENCE/0421/0043, co-financed by the European Regional Development Fund and the Republic of Cyprus through the Research and Innovation Foundation and by the European Joint Doctorate AQTIVATE that received funding from the European Union's research and innovation program under the Marie Skłodowska-Curie Doctoral Networks action and Grant Agreement No 101072344. S. B. is funded by the project QC4LGT, id number EXCELLENCE/0421/0019, co-financed by the European Regional Development Fund and the Republic of Cyprus through the Research and Innovation Foundation. J. F. acknowledges support by the German Research Foundation (DFG) research unit FOR5269 "Future methods for studying confined gluons in QCD" . S. B. and J. F. also acknowledge funding from the EuroCC project (grant agreement No. 951740) and from the Inno4scale project (grant agreement No. 101118139). G.K. acknowledges partial support by the project NiceQuarks, id number EXCELLENCE/0421/0195, co-financed by the European Regional Development Fund and the Republic of Cyprus through the Research and Innovation Foundation. M. C. acknowledges financial support from the U.S. Department of Energy, Office of Nuclear Physics, Early Career Award under Grant No. DE-SC0020405. G. S. acknowledges financial support from the European Regional Development Fund and the Republic of Cyprus through the Research and Innovation Foundation under contract number EXCELLENCE/0421/0025. This research was supported in part by grant NSF PHY-1748958 to the Kavli Institute for Theoretical Physics (KITP).

The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (`www.gauss-centre.eu`) for funding *renormglue* by providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS Supercomputer JUWELS at Jülich Supercomputing Centre (JSC).

Part of the results were created within the EA program of JUWELS Booster also with the help of the JUWELS Booster Project Team (JSC, Atos, ParTec, NVIDIA). Some of computations were supported by grants from the Swiss National Supercomputing Centre (CSCS) under projects with ids s702 and s1174. We also acknowledge PRACE for awarding us access to Piz Daint, hosted at CSCS, Switzerland, and Marconi100, hosted at CINECA, Italy. The authors also acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing HPC resources that have contributed to the research results. We thank the developers of the QUDA[34–36] library for their continued support. Ensemble production for this analysis made use of tmLQCD[37,38], DD-αAMG[39–41].

References

1. E. C. Aschenauer, I. Balitsky, L. Bland, S. J. Brodsky, M. Burkardt, V. Burkert, J. P. Chen, A. Deshpande, M. Diehlr, L. Gamberg et al., Eur. Phys. J. A **53**, no.4, 71, 2017, doi:10.1140/epja/i2017-12251-4.
2. S. J. Brodsky, A. L. Deshpande, H. Gao, R. D. McKeown, C. A. Meyer, Z. E. Meziani, R. G. Milner, J. Qiu, D. G. Richards, and C. D. Roberts, 2015, arXiv:1502.05728 [hep-ph].
3. X. Ji, Phys. Rev. Lett. **110**, 262002, 2013, doi:10.1103/PhysRevLett.110.262002.
4. C. Alexandrou et al. [ETM], Phys. Rev. D **104**, no.5, 054504, 2021, doi:10.1103/PhysRevD.104.054504.
5. J. Finkenrath, C. Alexandrou, S. Bacchio, P. Charalambous, P. Dimopoulos, R. Frezzotti, K. Jansen, B. Kostrzewa, G. Rossi, and C. Urbach, EPJ Web Conf. **175**, 02003, 2018, doi:10.1051/epjconf/201817502003.
6. C. Alexandrou, S. Bacchio, P. Charalambous, P. Dimopoulos, J. Finkenrath, R. Frezzotti, K. Hadjiyiannakou, K. Jansen, G. Koutsoui, B. Kostrzewa et al., Phys. Rev. D **98**, no.5, 054518, 2018, doi:10.1103/PhysRevD.98.054518.
7. J. Finkenrath, C. Alexandrou, S. Bacchio, M. Constantinou, P. Dimopoulos, R. Frezzotti, K. Jansen, B. Kostrzewa, G. Koutsou,G. Rossi et al., PoS **LATTICE2021**, 284, 2022, doi:10.22323/1.396.0284.
8. C. Alexandrou, S. Bacchio, M. Constantinou, J. Finkenrath, K. Hadjiyiannakou, K. Jansen, G. Koutsou, and A. Vaquero Aviles-Casco, Phys. Rev. D **102**, no.5, 054517, 2020, doi:10.1103/PhysRevD.102.054517.
9. C. Alexandrou, S. Bacchio, M. Constantinou, P. Dimopoulos, J. Finkenrath, R. Frezzotti, K. Hadjiyiannakou, K. Jansen, B. Kostrzewa, G. Koutsou et al., Phys. Rev. D **107**, no.5, 054504, 2023, doi:10.1103/PhysRevD.107.054504.
10. C. Alexandrou et al. [Extended Twisted Mass], Phys. Rev. D **109**, no.3, 034503, 2024, doi:10.1103/PhysRevD.109.034503.
11. C. Alexandrou, S. Bacchio, M. Constantinou, P. Dimopoulos, J. Finkenrath, R. Frezzotti, K. Hadjiyiannakou, K. Jansen, B. Kostrzewa, G. Koutsou et al., Phys. Rev. D **101**, no.3, 034519, 2020, doi:10.1103/PhysRevD.101.034519.
12. C. Alexandrou, S. Bacchio, M. Constantinou, J. Finkenrath, K. Hadjiyiannakou, K. Jansen, G. Koutsou, H. Panagopoulos, and G. Spanoudes, Phys. Rev. D **101**, no.9, 094513, 2020, doi:10.1103/PhysRevD.101.094513.
13. C. Alexandrou et al. [Extended Twisted Mass], Phys. Rev. Lett. **127**, no.25, 252001, 2021, doi:10.1103/PhysRevLett.127.252001.
14. C. Alexandrou, K. Cichy, M. Constantinou, K. Hadjiyiannakou, K. Jansen, A. Scapellato, and F. Steffens, Phys. Rev. D **99**, no.11, 114504, 2019, doi:10.1103/PhysRevD.99.114504.
15. C. Alexandrou, K. Cichy, M. Constantinou, K. Jansen, A. Scapellato, and F. Steffens, Phys. Rev. D **98**, no.9, 091503, 2018, doi:10.1103/PhysRevD.98.091503.
16. Y. Chai, Y. Li, S. Xia, C. Alexandrou, K. Cichy, M. Constantinou, X. Feng, K. Hadjiyiannakou, K. Jansen, G. Koutsou et al., Phys. Rev. D **102**, no.1, 014508, 2020, doi:10.1103/PhysRevD.102.014508.
17. G. Martinelli, C. Pittori, C. T. Sachrajda, M. Testa, and A. Vladikas, Nucl. Phys. B **445**, 81-108, 1995, doi:10.1016/0550-3213(95)00126-D.

18. C. Alexandrou, M. Constantinou, T. Korzec, H. Panagopoulos, and F. Stylianou, Phys. Rev. D **86**, 014505, 2012, doi:10.1103/PhysRevD.86.014505.
19. M. Constantinou, R. Horsley, H. Panagopoulos, H. Perlt, P. E. L. Rakow, G. Schierholz, A. Schiller, and J. M. Zanotti, Phys. Rev. D **91**, no.1, 014502, 2015, doi:10.1103/PhysRevD.91.014502.
20. C. Alexandrou et al. [ETM], Phys. Rev. D **95**, no.3, 034505, 2017, doi:10.1103/PhysRevD.95.034505.
21. G. Spanoudes, C. Alexandrou, J. Finkenrath, K. Hadjiyiannakou, H. Panagopoulos, and S. Yamamoto, PoS **LATTICE2022**, 125, 2023, doi:10.22323/1.430.0125.
22. C. Alexandrou, S. Bacchio, M. Constantinou, J. Delmar, J. Finkenrath, B. Kostrzewa, M. Petschlies, L. A. R. Chacon, G. Spanoudes, F. Steffens et al., 2024, arXiv:2405.08529 [hep-lat].
23. I. Novikov, H. Abdolmaleki, D. Britzger, A. Cooper-Sarkar, F. Giuli, A. Glazov, A. Kusina, A. Luszczak, F. Olness, P. Starovoitov et al., Phys. Rev. D **102**, no.1, 014040, 2020, doi:10.1103/PhysRevD.102.014040.
24. P. C. Barry et al. [Jefferson Lab Angular Momentum (JAM)], Phys. Rev. Lett. **127**, no.23, 232001, 2021, doi:10.1103/PhysRevLett.127.232001.
25. Z. F. Cui, M. Ding, F. Gao, K. Raya, D. Binosi, L. Chang, C. D. Roberts, J. Rodríguez-Quintero, and S. M. Schmidt, Eur. Phys. J. C **80**, no.11, 1064, 2020, doi:10.1140/epjc/s10052-020-08578-4.
26. K. D. Bednar, I. C. Cloët, and P. C. Tandy, Phys. Rev. Lett. **124**, no.4, 042002, 2020, doi:10.1103/PhysRevLett.124.042002.
27. M. Löffler et al. [RQCD], Phys. Rev. D **105**, no.1, 014505, 2022, doi:10.1103/PhysRevD.105.014505.
28. W. Good, K. Hasan, A. Chevis, and H. W. Lin, Phys. Rev. D **109**, no.11, 114509, 2024, doi:10.1103/PhysRevD.109.114509.
29. C. Alexandrou, S. Bacchio, M. Constantinou, J. Finkenrath, K. Hadjiyiannakou, K. Jansen, and G. Koutsou, Phys. Rev. D **101**, no.3, 031501, 2020, doi:10.1103/PhysRevD.101.031501.
30. C. Alexandrou, S. Bacchio, M. Constantinou, P. Dimopoulos, J. Finkenrath, K. Hadjiyiannakou, K. Jansen, G. Koutsou, B. Kostrzewa, T. Leontiou et al., Phys. Rev. D **103**, no.3, 034509, 2021, doi:10.1103/PhysRevD.103.034509.
31. C. Alexandrou, S. Bacchio, M. Constantinou, J. Finkenrath, K. Hadjiyiannakou, K. Jansen, G. Koutsou, and A. Vaquero, PoS **LATTICE2021**, 250, 2022, doi:10.22323/1.396.0250.
32. C. Alexandrou, S. Bacchio, M. Constantinou, K. Hadjiyiannakou, K. Jansen, and G. Koutsou, Phys. Rev. D **104**, 074503, 2021, doi:10.1103/PhysRevD.104.074503.
33. C. Alexandrou, G. Koutsou, Y. Li, M. Petschlies, and F. Pittler, Phys. Rev. D **110**, 094514, 2024, doi:10.1103/PhysRevD.110.094514.
34. M. A. Clark et al. [QUDA], Comput. Phys. Commun. **181**, 1517-1528, 2010, doi:10.1016/j.cpc.2010.05.002.
35. R. Babich et al. [QUDA], SC'11, doi:10.1145/2063384.2063478.
36. M. A. Clark et al. [QUDA], SC'16, doi:10.5555/3014904.3014995.
37. K. Jansen and C. Urbach, Comput. Phys. Commun. **180**, 2717-2738, 2009, doi:10.1016/j.cpc.2009.05.016.

38. B. Kostrzewa et al. [ETM], PoS **LATTICE2022**, 340, 2023, doi:10.22323/1.430.0340.
39. C. Alexandrou, S. Bacchio, J. Finkenrath, A. Frommer, K. Kahl, and M. Rottmann, Phys. Rev. D **94**, no.11, 114509, 2016, doi:10.1103/PhysRevD.94.114509.
40. S. Bacchio, C. Alexandrou, and J. Finkerath, EPJ Web Conf. **175**, 02002, 2018, doi:10.1051/epjconf/201817502002.
41. C. Alexandrou, S. Bacchioi, and J. Finkenrath, Comput. Phys. Commun. **236**, 51-64, 2019, doi:10.1016/j.cpc.2018.10.013.

High-Precision Calculation of the Muon Anomalous Magnetic Moment with Chiral Fermions

Christoph Lehner for the RBC/UKQCD collaborations

Universität Regensburg, 93053 Regensburg, Germany
E-mail: christoph.lehner@ur.de

One of the fundamental properties of quantum mechanics is the uncertainty principle, which limits the degree to which canonically conjugate variables such as position and momentum can simultaneously be known. An intriguing consequence of this principle is the contribution of virtual particles to physical processes. Concretely, heavy particles for which we do not have sufficient energy to produce them at our largest experiments can still contribute in this manner to processes that we can measure at lower energies. To successfully identify such a contribution, high-precision measurements and theory calculations are needed since these virtual effects tend to be suppressed. Interestingly, the coupling of muons, the heavier cousins of electrons, to a magnetic field provides such an opportunity since it is sensitive to the presence of heavier states, it can be experimentally measured at a relative precision of 10^{-10}, and a similarly precise theory calculation is possible. In order to achieve such a theory calculation, more precise results for contributions from known hadronic matter are needed. In this project, we work on matching the experimental precision of the new Fermilab E989 experiment by a high-precision calculation of the hadronic contributions using lattice QCD with chiral fermions.

1 Introduction

The anomalous magnetic moment of the muon $a_\mu = (g_\mu - 2)/2$ is a particularly important quantity. Since the early experiments by Stern and Gerlach in the 1920s for the electron case, it has played a pivotal role in establishing the foundations of our quantum-field-theoretical understanding of nature at ever higher precision. Major experimental efforts are underway at Fermilab and planned at J-PARC (in Japan) to reduce the experimental uncertainties.

The Fermilab experiment has in fact released first results in 2021, improving the previously best experimental uncertainty of 0.54 ppm[1] to 0.46 ppm[2]. Recently, Fermilab has further reduced the experimental uncertainty by a factor of 2.2. Over the next year, the Fermilab experiment aims to reduce the uncertainty further to approximately 0.14 ppm.

The current precision of a standard model theory prediction needs to be improved significantly in order to match the experimental precision and fully utilise its substantial progress. There is a community-wide effort underway by the Muon g-2 Theory Initiative[3] to establish such a high-precision result. Over the recent years, it has been demonstrated using the Euclidean windows introduced by our collaboration[4] that the methodology used so far for the hadronic vacuum polarisation (HVP) contribution from hadronic e^+e^- decays (R-ratio) needs further scrutiny and is currently not allowing for the needed precision of approximately 2/1000, see Fig. 1. First-principles calculations from lattice QCD, however, have made rapid progress and are a promising way towards matching the experimental precision.

We build on a multi-year effort to calculate the hadronic vacuum polarisation (HVP) contributions to a_μ to high precision. In addition, the data generated in this effort is beneficial for a wide range of additional physics projects that are limited by the so-far mostly

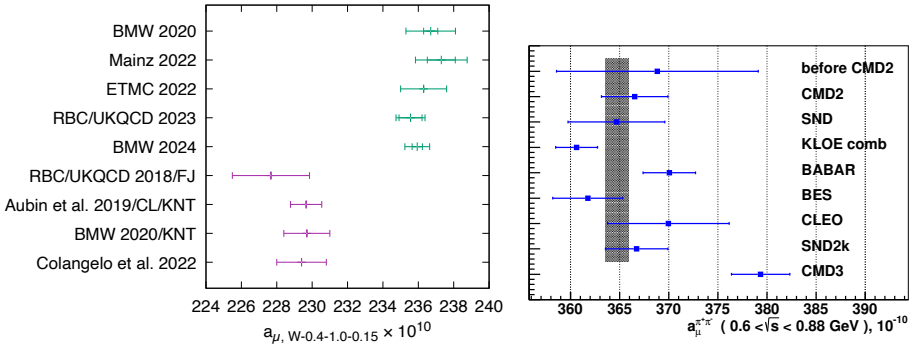


Figure 1. Using the intermediate window quantity defined in Ref. 4, the community has now established a clear tension between first-principles lattice QCD+QED results and data-driven R-ratio determination (minus the recent CMD3 result). This is shown in the left figure. In the right figure we show tensions in the R-ratio data sets that are so-far unresolved. There is a clear desire to have full first-principles lattice QCD result with competitive precision.

small volumes $m_\pi L \approx 4$ available for the chiral domain-wall ensembles used by the RBC/UKQCD collaborations. Since these ensembles also are widely used in the community, we expect this project to be of broad benefit outside of RBC/UKQCD as well. The HVP project has so far produced results at the $\approx 2\%$ accuracy[4] for the total contribution but has reached the needed target precision for the short-distance and intermediate-distance sub-quantities in the isospin symmetric limit[5] in 2023. At the Lattice 2024 Symposium in Liverpool we have presented a new result for the isospin-symmetric long-distance window contribution matching a 7/1000 precision.

In establishing the standard model theory result again at ever higher precision using first-principles lattice QCD+QED it is of utmost importance to have several independent lattice-QCD calculations of commensurate precision. We need detailed comparisons between precise lattice QCD calculations using different methods and with different systematic effects to test all aspects of the results thoroughly. Hence, independent and distinct theoretical calculations are both necessary and complementary. In the following, we explain our progress that was crucially enabled in part by the compute infrastructure provided in Jülich.

2 Gauge Ensemble Generation

In order to control the extrapolation to zero lattice spacing and infinite physical lattice volume better, we generated additional ensembles in the last two years. To address the finite lattice spacing uncertainty, we have generated new ensembles for $m_\pi = 280$ MeV with lattice cutoff of 2.7 GeV, 3.5 GeV, and first few configurations at 4.6 GeV. These ensembles are generated using the HMC algorithm made exact by a Metropolis accept-reject step.

In Fig. 2, we show the thermalisation of the three ensembles for the Wilson-flow scale $\sqrt{t_0}$. The thermalisation for the finest ensemble was first performed on smaller volumes

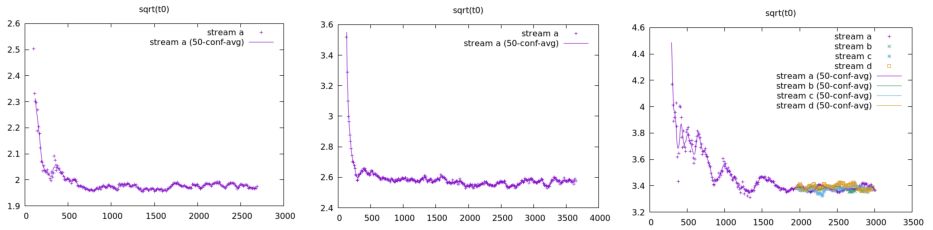


Figure 2. Thermalisation and ensemble generation as performed in last-years allocation for the 2.7 GeV, 3.5 GeV, and 4.6 GeV ensembles from left to right. The coarsest two ensembles were completely generated on Booster, the 4.6 GeV ensemble was thermalised for a few hundred trajectories on Juwels Booster.

and only at around molecular-dynamics (MD) unit 1500 we switched to the final large volume. Around MD unit 2000, we switched to four parallel streams.

In order to also improve the control over slight mistunings of the quark masses and over the large-volume limit, additional ensembles were created in recent years. In this context it was crucial to also generate lattice ensembles for which the pion Compton wavelength fits 8 times in the physical volume. At such large volumes the finite-volume corrections that usually need to be performed with limited systematic precision, are small compared to the statistical uncertainty. The corresponding large set of ensembles that are currently in use for the calculation of the muon anomalous magnetic moment is summarised in Fig. 3.

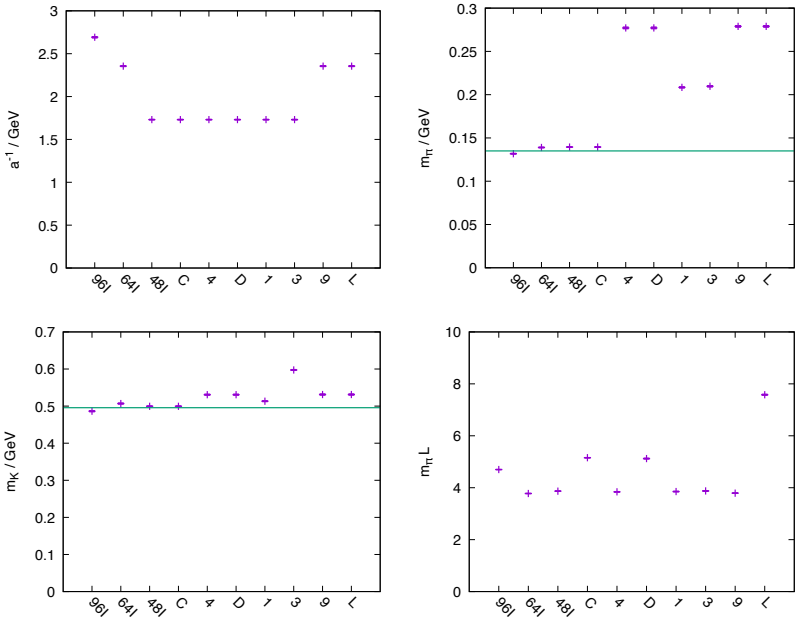


Figure 3. The most important ensemble properties are summarised in this figure. The uncertainties are shown but are typically in the per-mille range.

3 Consolidating Short and Intermediate Distances

Our project first focused on consolidating the short and intermediate distance Euclidean window contributions. To this end, we have generated new Dirac low-modes in a locally-coherent representation and have measured the vector-vector correlator and standard ensemble parameters with great precision. We have made codes for both sub-projects publicly available at `https://github.com/lehner/gpt`.

We have published an update for the short-distance and intermediate-distance windows[5] in which the crucial improvement over previous work[4] was the improved control over the zero-lattice spacing extrapolation. We added a finer lattice spacing at physical pion mass and studied two variations of discretising the muon-photon system combined with two variations of discretising the coupling of the photon to the quarks. We show the continuum limit including all of these variations for the intermediate window in Fig. 4. The extrapolations of different discretisations agree well in the continuum limit. The remaining spread was used as an uncertainty estimate for the residual continuum limit errors. With the current setup this uncertainty is smaller than the statistical uncertainty.

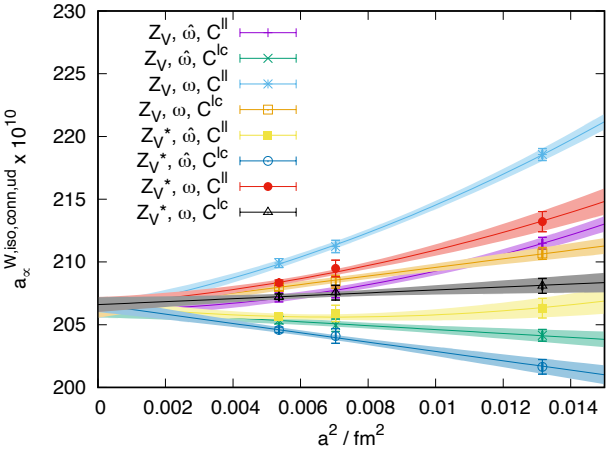


Figure 4. Continuum extrapolation of the window quantity with three lattice spacings, local-local (ll) and local-conserved (lc), as well as continuum (p) and lattice (phat) photon momenta.

Due to the importance of the result in resolving the emerging tension, we conducted our analysis in a blinded manner and had five analysis groups independently analysing the data. In Fig. 5, we compare our results against results of other groups. We note that both the short and intermediate distance window contribution are now consolidated at the precision needed to match the final Fermilab E989 experimental precision.

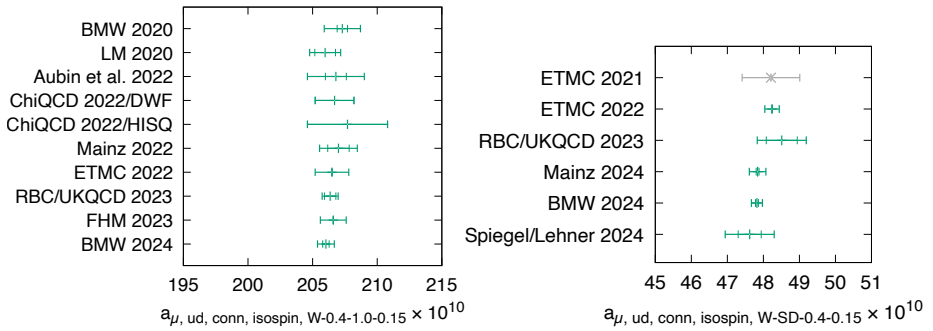


Figure 5. Comparison of short-distance and intermediate-distance window results.

4 Improving the Long-Distance Precision

The next step in our program was a significant improvement in precision of the long-distance contribution of the light quark connected contribution compared to earlier work[4]. This required the introduction of new methodology. We employed the improved bounding method developed by our team[6,7], which uses the information of the discrete finite-volume intermediate states to reconstruct the long-distance contribution. To this end one needs to consider a large operator basis of N operators transforming in the proper irreducible representation of the finite-volume and isospin symmetry group and the corresponding possible N^2 two-point correlation functions. This then allows for a high-precision spectral reconstruction.

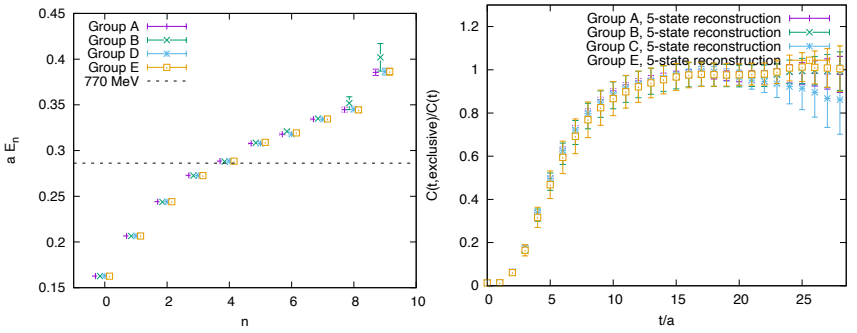


Figure 6. We show the spectrum on our 96I ensemble as obtained by four of our analysis groups on the left as well as the ratio of the reconstructed two-point correlator $C(t)$ divided by the fully inclusive one on the right for our 64I ensemble.

In Fig. 6, we show the numerical tests performed between our analysis groups. The saturation of the finite-volume reconstruction can clearly be seen in the right panel. This gives confidence that all necessary states have been captured.

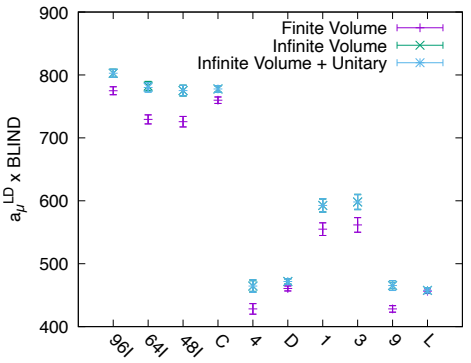


Figure 7. Results for the isospin symmetric long-distance window for all ensembles that entered the recently unblinded result at Lattice 2024. The data for ensembles 4, D, 1, 3, 9, and L was generated using last year's allocation and played a crucial role in the success of the project.

In Fig. 7, we show the results of this methodology applied to a large set of ensembles for which the parameters were shown in Fig. 3 both with and without finite-volume corrections applied. For the largest physical volume the finite-volume corrections were small compared to the statistical uncertainty. These individual lattice results are then combined to perform a continuum and infinite-volume extrapolation. The result of which is shown in Fig. 8. In this figure we show the results that were unblinded for the Lattice 2024 Symposium and compare it to other results in the literature. We notice that the tension between lattice QCD and the e^+e^- experimental average is even more pronounced in the long-distance contribution compared to the intermediate distance contribution. The computational resources provided by Jülich were a crucial component of this result. A publication containing these results is currently in preparation.

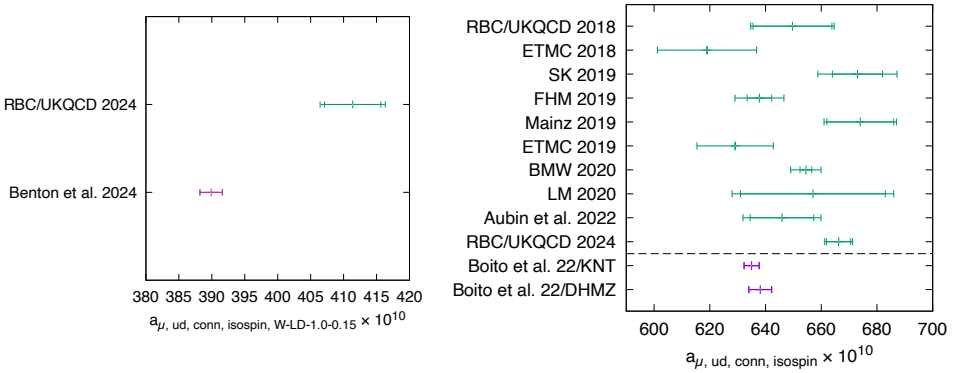


Figure 8. Using the improved bounding method developed by some of the investigators[6,7], we have recently unblinded a new result for the long-distance window and the complete isospin symmetric contribution. Our result is in strong tension with the R-ratio result and it is also clear that some consolidation of lattice QCD results at this higher precision level is needed.

5 Next Steps and Concluding Remarks

In order to achieve our final precision goal further steps are needed. First, we continue to increase statistics for the finest ensemble (96I) and have started generating a physical pion mass ensemble with $m_\pi L \approx 5$ and $a^{-1} \approx 3.5$ GeV to achieve the targeted final precision for the long-distance contribution in the next years. Another crucial improvement over our previous complete result[4] is an improved calculation of the quark disconnected contribution. To this end, we have re-used the distillation data sets generated for the long-distance project that can be applied to this case with small numerical effort. Finally, we need to compute the QED and strong-isospin-breaking corrections at improved precision. In order to achieve this, we have adopted the strategy that we have previously successfully used to compute the hadronic light-by-light scattering contributions to the muon anomalous magnetic moment and are currently in the process of performing the next analysis.

In recent years lattice QCD methodology has evolved and now allows for complete first-principles calculations at a high precision that is currently limited by the so-far acquired statistics of the Markov chains. The short and intermediate distance contributions are now computed at the needed precision for the Fermilab E989 experiment. The long-distance contribution requires additional effort but with continued computing support, we can match the precision and therefore fully utilise the new Fermilab E989 results not too long after their final result is published in 2025.

Acknowledgements

We gratefully acknowledge computing time allocated through the Gauss Centre for Supercomputing at JUWELS Booster in Jülich, through EuroHPC at LUMI-G and Leonardo, and through ALCC and INCITE programs of the US DOE at OakRidge and Argonne.

References

1. G. W. Bennett et al., *Final Report of the Muon E821 Anomalous Magnetic Moment Measurement at BNL*, Phys. Rev., **D73**, 072003, 2006.
2. B. Abi et al., *Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm*, Phys. Rev. Lett., **126**, no. 14, 141801, 2021.
3. T. Aoyama et al., *The anomalous magnetic moment of the muon in the Standard Model*, Phys. Reports, **887**, 1-166, 2020.
4. T. Blum, P. A. Boyle, V. Gülpers, T. Izubuchi, L. Jin, C. Jung, A. Jüttner, C. Lehner, A. Portelli, and J. T. Tsang, *Calculation of the hadronic vacuum polarization contribution to the muon anomalous magnetic moment*, Phys. Rev. Lett., **121**, 022003, 2018.
5. T. Blum et al., *Update of Euclidean windows of the hadronic vacuum polarization*, Phys. Rev. D, **108**, no. 5, 054507, 2023.
6. C. Lehner, *Status of HVP calculation by RBC/UKQCD*, in: 36th International Symposium on Lattice Field Theory (LATTICE2018), 2018.
7. M. Bruno, T. Izubuchi, C. Lehner, and A. S. Meyer, *Exclusive Channel Study of the Muon HVP*, in: 37th International Symposium on Lattice Field Theory (LATTICE2019), 2019.

Materials Science

Materials Science

Gustav Bihlmayer

Peter Grünberg Institut (PGI-1), Forschungszentrum Jülich, 52425 Jülich, Germany
E-mail: g.bihlmayer@fz-juelich.de

In our daily life, we are surrounded by materials designed for specific tasks, for example, the constituents of batteries and microelectronic circuits, magnets or lubricants. Often it is not a single physical process that leads to the performance of a material, it is the balance between many possible paths a system can take that matters. The art of computational materials science lies in cleverly sampling these possibilities, often guided by dedicated experiments. Modern supercomputing resources allow for studies that give us an increasingly complete picture of the real world. The following three articles are state-of-the-art examples of what can be achieved with the help of large computational resources, as provided by the John von Neumann Institute for Computing.

The first article gives a very educational example of complexity in an applied topic, the use of graphite as a solid lubricant. While traditionally the lubrifying properties of graphite are attributed to the easy gliding of parallel graphene layers, like cards of a card stack that are easily moved, this picture cannot account for the experimental observation that a certain humidity is necessary to sustain the desired properties. Klemenz and Moseler use a Density-Functional Tight-Binding (DFTB) method to simulate graphite under given pressures and sliding conditions. Different amounts of water molecules are intercalated in the gaps between the lamellae that can be parallel or perpendicular to the applied stress. Although there is no simple picture evolving, statistically it can be seen that a certain H_2O concentration is needed to avoid "cold welding" of graphene lamellae that happen to be oriented perpendicular to the sliding direction. A passivating water film, however, ensures an easy motion of the layers. High-resolution transmission electron microscopy images confirm experimentally the situation observed in the DFTB simulations. These insights might help to design graphite-based lubricants that can be used under dry conditions, e.g. for applications in vacuum.

A second example, of how reality can be more complex than traditional textbook knowledge, is given by Dabrowski *et al.*, who study the catalytic growth process of graphene and its insulating analog, hexagonal boron nitride (hBN), on germanium surfaces. While in usual catalytic reactions of molecules, the catalyst should not undergo structural changes in the process, here, where large two-dimensional layers are formed, also the underlying Ge surface gets modified: Density functional theory (DFT) calculations reveal that the Ge(110) surface tends to flatten during the graphene growth process as hydrogen from the molecular precursors removes Ge in the form of its hydride from steps and small islands. In contrast, on the Ge(001) surface a faceting process can be observed. The growth process of hBN on Ge surfaces is again very different, a fact that can at least partially be attributed to the different properties of the hydrogen-passivated edges. While graphene growth stops after a single layer, hBN easily forms multi-layer stacks. High-quality graphene and hBN layers are not only promising materials for future micro-

electronics but also explored for their catalytic properties. Studies in these directions on defect structures in bilayer graphene are ongoing.

While the first two contributions report on atomistic studies on different lengths and time scales, the third article goes into even smaller details of the electronic structure of a certain class of compounds, looking at the fine interplay of transition-metal (TM) and ligand orbitals in nickelates. Here, Lechermann is interested in superconducting layered compounds that bear some resemblance to the celebrated cuprates showing high-temperature superconductivity. Also in nickelates, hole-doping is essential to get the desired properties but the question, of where the holes localise is still debated. It is generally agreed that correlation effects in these oxides make a treatment beyond DFT necessary. Here the dynamical mean-field theory (DMFT) is applied to the d-orbitals of Ni. But in this study, special attention is paid to the oxygen p-states, which are usually just considered as ligand states to the TM atom. A more refined treatment, consisting of a combination of self-interaction corrected DFT with DMFT, is applied to two classes of nickelates that give detailed insights into the hole distribution in these materials. Although the connection to the superconducting properties is still an area to explore, the proposed method might provide a way to cope with materials that are on the border between charge-transfer and Mott-Hubbard insulators.

These three examples nicely illustrate how enhanced computational resources can lead to deeper insights in the field of materials physics and chemistry – not only due to larger length- and time-scales that can be simulated but also the increased sophistication of the the applied methods. On the one hand, computational materials science thus can help to optimise or improve industrially relevant processes, on the other hand, it also can uncover qualitatively new physics.

Quantum Molecular Dynamics Simulations Elucidate the Tribochemistry of Graphene-Based Materials

Andreas Klemenz[1] **and Michael Moseler**[1,2,3]

[1] Fraunhofer IWM, MicroTribology Center μTC, Wöhlerstraße 11, 79108 Freiburg, Germany
E-mail: {andreas.klemenz, michael.moseler}@iwm.fraunhofer.de

[2] Institute of Physics, University of Freiburg, Hermann-Herder-Straße 3, 79104 Freiburg, Germany

[3] Freiburg Materials Research Center, University of Freiburg,
Stefan-Meier-Str. 21, 79104 Freiburg, Germany

Graphite is a widely used solid lubricant. However, its lubricating properties are highly dependent on ambient humidity. In very dry environments, graphite lubrication fails, severely limiting its applicability. Traditional models to explain the lubrication properties of graphite do not fully explain this phenomenon. The objective of this research project was to investigate the mechanisms behind graphite lubrication under varying humidity and high mechanical load on the nanoscale. Through experiments and molecular dynamics simulations, the study demonstrated that two key mechanisms govern friction in graphite-lubricated contacts: at low humidity, cold welding occurs at the nanoscale, leading to high friction, while at humidities typical for laboratory air, water films form, allowing smooth sliding and reduced friction. In addition, we found the formation of turbostratic carbon at the sliding interfaces, a phenomenon not included in any current model of graphite lubrication. Our results provide a deeper understanding of the atomic scale mechanisms of graphite lubrication and are summarised in a simple, instructive model.

1 Introduction

Liquid lubricants or greases are typically used to reduce friction and wear in technical systems. However, in some applications, this is not possible. At particularly high temperatures, for example in metalworking, liquid lubricants would evaporate or chemically decompose. Solid lubricants are therefore often used in such applications. Graphite is one of the oldest and most effective representatives of this category[1].

X-ray diffraction experiments in the 1920s revealed that graphite consists of parallel graphene lamellae bound by weak van der Waals interactions[2]. Based on this discovery, the lubricating effect of graphite is usually explained using the so-called deck-of-cards model[3–5] (Fig. 1a). This assumes that the graphene layers can easily slide against each other due to the weak interaction, analogous to playing cards in a deck. This model provides a simple and intuitive explanation for the low friction in graphite-lubricated sliding contacts and is still widely presented in most textbooks today[6]. However, the first doubts about this approach were raised as early as the 1930s[7]. At this time, the first aeroplanes capable of reaching high altitudes were constructed. It was observed that the graphite contacts in the electrical onboard generators underwent rapid wear under those conditions[8]. This was later confirmed in controlled experiments, which showed that these observations were due to the low humidity at high altitudes[9].

The deck-of-cards model offers no explanation for this effect. Therefore, it was initially speculated that water molecules could intercalate between the graphene layers. It was assumed that the water would increase the distance between the layers, reducing the

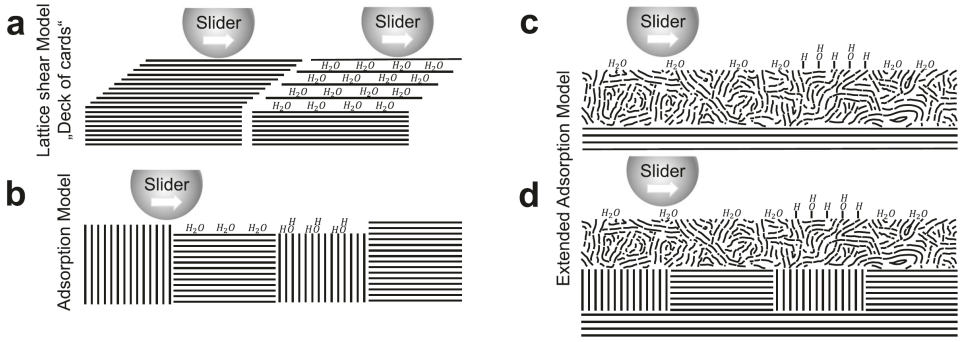


Figure 1. (a) Deck-of-cards model for graphite lubrication with (right) and without water intercalation (left). (b) Adsorption model. (c,d) Adsorption model extended by the shear-induced formation of turbostratic carbon for high loads (c) and low loads (d). (Figure adapted from Morstein, Klemenz et al.[18] CC BY 4.0)

interaction between neighbouring layers[10] (Fig. 1a). However, this was disproved experimentally[11,12]. Since then, the scientific literature has been dominated by various versions of the so-called adsorption model[9] (Fig. 1b). It is assumed that water molecules attach to dangling bonds on defects and edges of graphite crystals and passivate them[13–16]. This prevents the formation of covalent bonds between the graphene layers, reducing friction when the crystals slide against each other. The adsorption model has been tested in the past for applications with low mechanical stress, such as electric contacts, with successful results[9]. It has not been established whether this model can also be used to describe highly loaded systems such as roller bearings.

The objective of this research project was therefore to investigate the mechanisms of graphite lubrication in highly loaded contacts in detail. A particular focus was placed on the influence of humidity on friction. The work was carried out in close collaboration with an experimental group at the Karlsruhe Institute of Technology (KIT), which conducted accompanying experiments.

2 Influence of Humidity on Graphite Lubrication

In the experiments, iron platelets were coated with graphite. A linear reversing tribometer was used to measure the friction in an atmosphere with controlled humidity. The nanoscale structure of the coatings was examined using high resolution transmission electron microscopy (HRTEM)[17,18].

Varying the humidity resulted in the expected behaviour. Measurements showed that high friction and high wear occurred at very low humidities. As humidity increased, these values decreased considerably. At 10-30 % rH, which corresponds to typical ambient air, friction coefficients of 0.1-0.15 were measured. The HRTEM investigations also revealed a surprising change in the structure of the sliding interfaces. While the typical graphite structure with extended, parallel graphene layers could be observed in the initial state (Fig. 2a), so-called turbostratic carbon formed as a result of the tribological load (Fig. 2b). Like graphite, this is a material consisting of graphene lamellae. However, in contrast to graphite, these are twisted and shifted against each other, and no ordered structure can

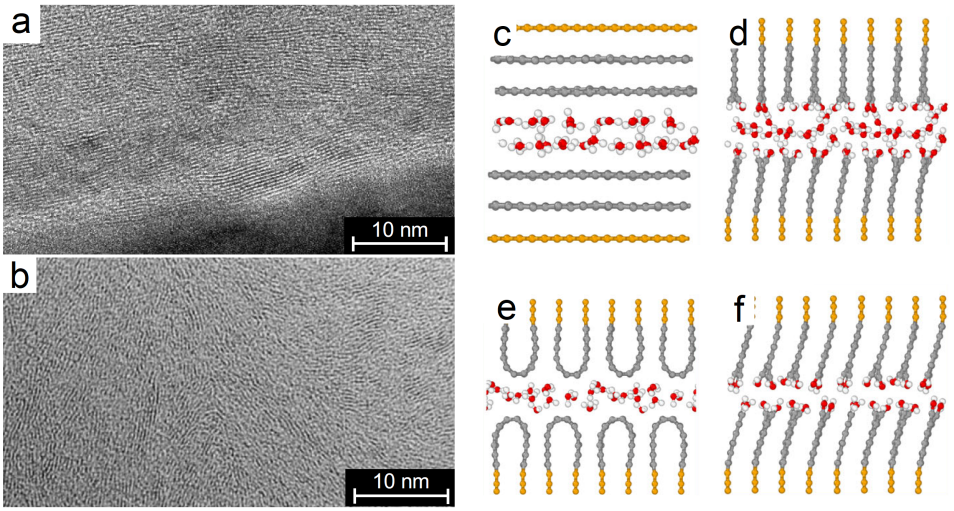


Figure 2. HRTEM images of a graphite layer prior to tribological load (a) and after a tribometer experiment under a load of 1 GPa at 24 % RH (b). Examples of setups for atomistic simulations with graphene lamellae with parallel (c) and perpendicular orientation to the sliding direction with water between the surfaces (d,e) and in dry contact (f). For perpendicular orientation, mixed H/OH passivated (d,f) and unpassivated (e) graphene edges were considered. The orange atoms are coupled to the barostat for pressure control. Please note that unterminated graphene layers tend to form hairpin-shaped structures that connect neighbouring graphene layers (e). This phenomenon can also be observed when graphite powder is heated in vacuum[24]. (Figure adapted from Morstein, Klemenz et al.[18] CC BY 4.0)

be observed on larger length scales. In friction experiments with low normal loads, turbostratic carbon formed only near the surface of the coatings. In contrast, a thick layer formed under high loads. These structures were not anticipated and are not predicted by the current models for the lubrication properties of graphite. In none of the experiments conducted, graphene lamellae were aligned parallel to the direction of sliding, contrary to the predictions of the deck-of-cards model.

Molecular dynamics simulations[19,20] were carried out to investigate the formation of the turbostratic carbon at the sliding interface observed in the experiments. The Density-Functional Tight-Binding (DFTB) method[21,22] was used to model the interatomic interactions and guarantee an adequate reproduction of chemical reactions. This method is fast enough to simulate systems consisting of a few hundred atoms over periods of a few hundred picoseconds. The HRTEM investigations of the experimental graphite contacts prior to tribological loading revealed areas in which the graphene lamellae were oriented parallel to the sliding direction as well as areas with perpendicular orientation to the sliding interface (Fig. 2a). Therefore, systems with parallel and perpendicular orientation were also considered in the simulations (Fig. 2c-f). The pressure was applied using a barostat[23] and a constant temperature of 300 K was set by coupling the systems to a Langevin thermostat[19]. A sliding velocity of 100 ms^{-1} was applied for a period of 300 ps. The number of water molecules in the contacts was varied between 0 and 32, and normal pressures between 500 MPa and 5 GPa were considered.

No chemical reactions could be observed in any combination of normal pressure and

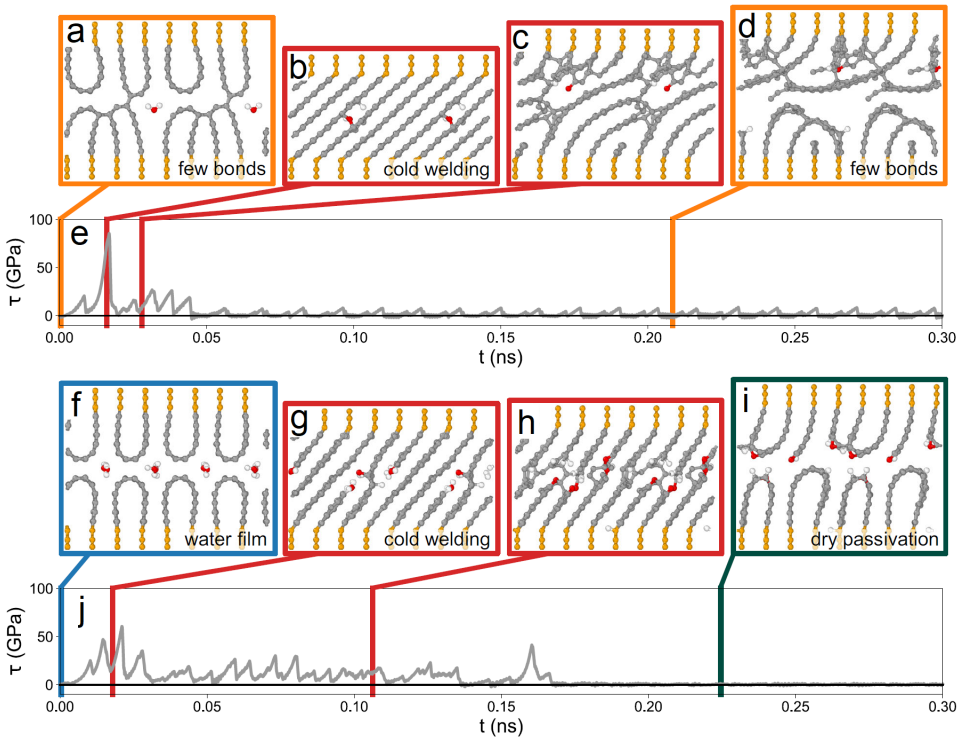


Figure 3. Structural evolution of a graphite slab during sliding in atomistic simulations. During sliding of two graphite blocks in contact with one water molecule (a) at a normal pressure of 1 GPa, the surfaces cold weld initially (b). After a few ps of sliding, an aromatic structure forms (c) and the shear stresses τ decrease (e). However, the upper and the lower graphite blocks remain in contact by a few bonds which frequently form and break during sliding (d). In a system with 5 water molecules (f), the surfaces also cold weld initially (g,h). After 0.17 ns of sliding, passivated surfaces without free water form and the upper and lower graphite blocks separate (i). The separation leads to a significant reduction of the shear stress τ (j). (Figure adapted from Morstein, Klemenz et al.[18] CC BY 4.0)

number of water molecules in contact with graphene lamellae orientated parallel to the sliding direction. However, if the lamellae were orientated perpendicular to the sliding direction, various reactions and structural changes could be observed in the systems. For the formation of turbostratic carbon at the friction interfaces, the systems with graphene layers perpendicular to the sliding direction are therefore of particular importance and the majority of the simulations therefore applied these setups (Fig. 2e).

The introduction of a large amount of water into the contact resulted in the formation of continuous water films in the systems, effectively separating the upper and lower graphite crystals from each other (Fig. 2e). Only small forces were required to shear these systems. In contrast, the simulation of dry contacts and those containing only a small number of water molecules showed cold welding of the surfaces in most cases (Fig. 3). The reduction in water required to observe cold welding was dependent on the normal pressure in the system. A smaller number of water molecules is sufficient for the formation of a water

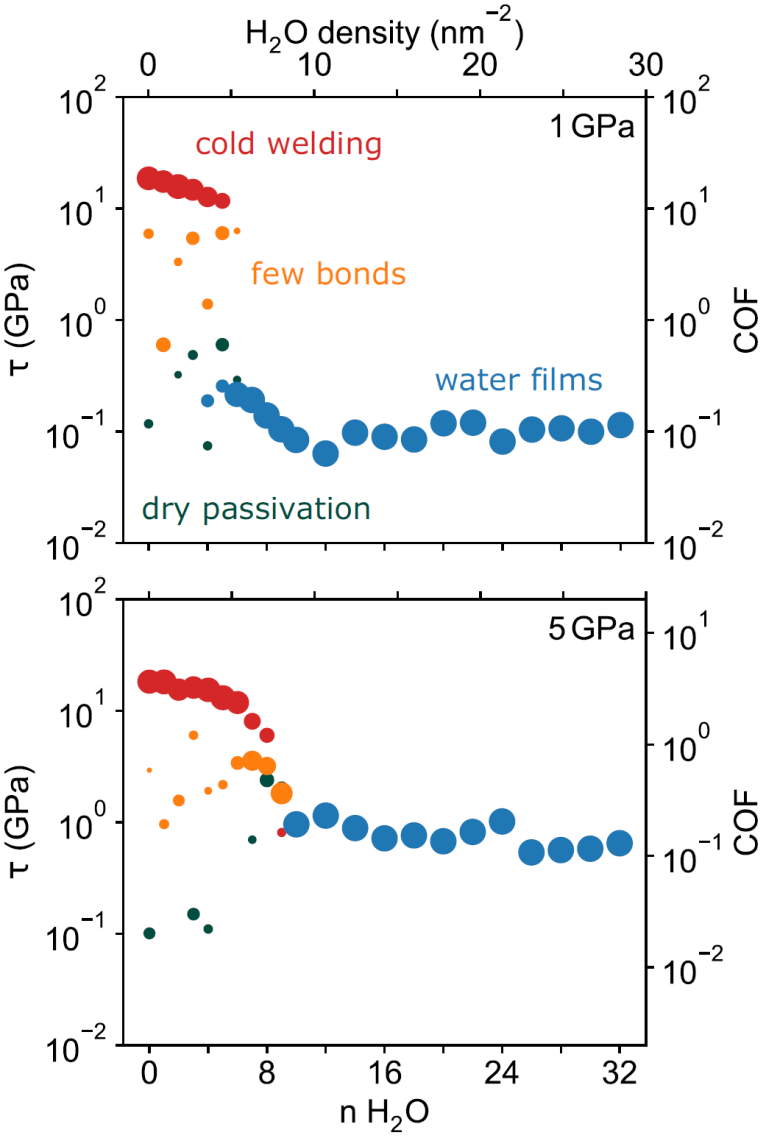


Figure 4. Average shear stresses in the different regimes for normal pressures of 1 and 5 GPa. The colours of the markers distinguish the different regimes, the size of the marker indicates the probability of a regime. (Figure adapted from Morstein, Klemenz et al.[18] CC BY 4.0)

film at low loads than at high loads. When cold welding occurred, a layer of amorphous material formed (Fig. 3c). Consequently, shearing the graphite crystals against each other required higher forces (Fig. 3e). Occasionally, after some shearing of the systems, aromatic structures formed and the upper and lower parts of the system separated from each other (Fig. 2d). In some cases, the separation was complete. In others, the upper and

lower parts of the system remained connected by a few bonds. This effect was associated with a significant reduction in friction forces (Fig. 3e). It turned out that this was not a deterministic effect. If spontaneous passivation was observed in a certain combination of water quantity and normal pressure, it was possible that it did not occur in a repetition with slightly varied initial conditions. Overall, it was not possible to predict under which conditions the separation would occur or how long shearing would have to be carried out before it occurred. However, it was possible to calculate the probability of the systems occurring in each configuration. It could be shown that cold welding at low water densities and the formation of water films at high water densities on the surfaces are the dominant effects in tribologically loaded graphite contacts.

To be able to make statements about the behaviour of macroscopic graphite layers in tribological contacts, it can be estimated that a monolayer of water corresponds to a density of approximately 12-13 nm^{-2} water molecules. A coating with a single monolayer should easily form under normal laboratory conditions due to condensation from ambient humidity. It can therefore be assumed that graphite contacts are typically in the water film regime in ambient conditions, which provides an explanation for the good lubricating properties of graphite. However, the water density only needs to decrease slightly to reach the cold welding regime, especially at high pressures. It can therefore be assumed that localised cold welding of the surfaces frequently occurs due to localised dry running of the contacts. This provides an explanation for the experimentally observed formation of turbostratic graphite at the sliding interface. The findings described are not included in any currently existing model to explain the lubrication properties of graphite. However, the known adsorption model can easily be extended to include the formation of turbostratic carbon at the shear interface[18] (Fig. 1c,d).

3 Concluding Remarks

These new findings provide an important contribution to a deeper understanding of the lubricating effect of graphite. They also offer promising starting points for further investigations. One possible application of these findings could be the development of novel, graphite-based solid lubricants for use in vacuum environments, such as those found in aero-space applications or in highly specialised industrial processes. The knowledge gained in this project therefore has the potential to provide important inspiration for future research and drive forward new technological developments.

Acknowledgements

We gratefully acknowledge the computing time granted by the John von Neumann Institute for Computing (NIC) and provided on the supercomputer JUWELS at Jülich Supercomputing Centre (grant HFR19). This work was funded by the Deutsche Forschungsgemeinschaft (DFG). The project “Mechanisms of Graphite Lubrication in Rolling Contacts” (“Mechanismen der Graphitschmierung in Wälzkontakten”) is part of the priority program SPP2074 (grant number MO879/20-1).

References

1. H. E. Sliney, *Solid Lubricants*, NASA Technical Memorandum 103803, NASA, Cleveland, Ohio, USA, 1991.
2. W. H. Bragg, *An introduction to Crystal Analysis*, G. Bell and Sons Ltd., London, UK, 1928.
3. R. M. Mortier, M. F. Fox, and S. T. Orszulik, *Chemistry and Technology of Lubricants*, Springer, Dordrecht, The Netherlands, 2010.
4. B. K. Yen, B. E. Schwickert, and M. F. Toney, *Origin of low-friction behavior in graphite investigated by surface x-ray diffraction*, Appl. Phys. Lett. **84**, 4702-4704, 2004.
5. T. W. Scharf and S. V. Prasad, *Solid lubricants: a review*, J. Mater. Sci. **48**, 511-531, 2013.
6. B. Bhushan, *Principles and Applications of Tribology*, 2nd Edition, John Wiley & Sons, Chichester, West Sussex, UK, 2013.
7. R. M. Baker, *Sliding contacts- electrical characteristics*, Trans. AIEE **55**, 94-100, 1936.
8. D. Ramadanoff and S. W. Glass, *High-altitude brush problem*, Trans. AIEE **63**, 825-830, 1944.
9. R. H. Savage, *Graphite lubrication*, J. Appl. Phys. **19**, 1-10, 1948.
10. G. W. Rowe, *Some observations on the frictional behaviour of boron nitride and of graphite*, Wear **3**, 274-285, 1960.
11. R. D. Arnell and D. G. Teer, *Lattice parameters of graphite in relation to friction and wear*, Nature **218**, 1155-1156, 1968.
12. J. Skinner, N. Gane, and D. Tabor, *Micro-friction of graphite*, Nat. Phys.Sci. **232**, 195i-196, 1971.
13. J. K. Lancaster and J. R. Pritchard, *The influence of environment and pressure on the transition to dusting wear of graphite*, J. Phys. D: Appl. Phys. **14**, 747-762, 1981.
14. P. Restuccia, M. Ferrario, and M. C. Righi, *Monitoring water and oxygen splitting at graphene edges and folds: insights into the lubricity of graphitic materials*, Carbon **156**, 93-103, 2020.
15. M. Dienwiebel, G. S. Verhoeven, N. Pradeep, J. W. M. Frenken, J. A. Heimberg, and H. W. Zandbergen, *Superlubricity of graphite*, Phys. Rev. Lett. **92**, 126101, 2004.
16. F. Bonelli, N. Manini, E. Cadelano, and L. Colombo, *Atomistic simulations of the sliding friction of graphene flakes*, Eur. Phys. J. B **70**, 449-459, 2009.
17. C. E. Morstein and M. Dienwiebel, *Graphite lubrication mechanisms under high mechanical load*, Wear **477**, 203794, 2021.
18. C. E. Morstein, A. Klemenz, M. Dienwiebel, and M. Moseler, *Humidity-dependent lubrication of highly loaded contacts by graphite and a structural transition to turbostratic carbon*, Nat. Commun. **13**, 5958, 2022.
19. D. Frenkel and B. Smit, *Understanding Molecular Simulations: From Algorithms to Applications*, 2nd Edition, Academic Press, San Diego, California, USA, 1996.
20. L. Pastewka, *Atomistica interatomic potentials library*, `https://github.com/Atomistica/atomistica`, Accessed 2022-08-25.
21. M. Elstner et al., *Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties*, Phys. Rev. B **58**, 7260-7268, 1998.

22. B. Aradi, *The DFTB website*, `https://dftb.org`, Accessed 2022-08-25.
23. L. Pastewka, M. Moser, and M. Moseler, *Atomistic insights into the running-in, lubrication, and failure of hydrogenated diamond-like carbon coatings*, Tribol. Lett. **39**, 49-61, 2010.
24. Z. Liu, K. Suenaga, P. J. F. Harris, and S. Iijima, *Open and closed edges of graphene layers*, Phys. Rev. Lett. **102**, 015501, 2009.

Ab Initio Simulation of Materials for Environmentally Friendly Technologies

Jarek Dąbrowski, Fatima Akhtar, Max Franck, and Mindaugas Lukosius

IHP – Leibniz-Institut für innovative Mikroelektronik,
Im Technologiepark 25, 15236 Frankfurt (Oder), Germany
E-mail: dabrowski@ihp-microelectronics.com

An *ab initio* DFT study of nucleation and growth of graphene and hBN for high-performance environmentally friendly microelectronics and of N and O co-doping of activated carbon for metal-free catalysis is reported. As fringe benefit, a new Ge(110) surface structure was revealed.

1 Introduction

The need for environmentally friendly technologies in response to global climate change necessitates innovation in materials. Simulations support these innovations by providing insights into complex physical processes and guiding experimental efforts. When the approximations used are well-founded in physics, simulations reduce intuitive bias while fostering reliable physical intuition comparable to insights gained from experiments. *Ab initio* density functional theory (DFT) falls into this category, as it operates at the quantum-mechanical level of atomic interactions. Although computationally intensive (which limits its use), it is indispensable for modern technology development, especially in the design and optimisation of materials at the atomic scale. By integrating theoretical and experimental insights, DFT simulations contribute to breakthroughs in material design, paving the way for greener solutions. This includes microelectronics, where innovative ideas applied on the atomistic level can help mitigate environmental impacts associated with cost-driven performance improvements and miniaturisation, while also reducing energy consumption.

For instance, Cu ion contamination is a concern in water pollution caused by the semiconductor industry[1,2]. The costs of removing Cu from wastewater are an important factor[3]. With the growing adoption of graphene (sheets of sp^2-bonded C atoms), which is typically synthesised on Cu, this issue is expected to intensify, as copper released during etching and transfer[4–6] increases filtration costs and environmental risks. Cu from graphene production may also contaminate devices and fabrication lines[7,8], compounding the financial and environmental pressures for innovative solutions. To address these challenges, the IHP is exploring the option of growing graphene on semiconductor substrates where feasible[9–14].

This work continues our previous studies[15–19] done at the IHP with the support of the John von Neumann Institute for Computing. We present selected results of our recent *ab initio* DFT calculations run on the JUWELS cluster. The discussion focuses on the nucleation and growth mechanisms of two-dimensional (2D) films for modern microelectronics: graphene[9–13] (Sec. 2) and hexagonal boron nitride (hBN)[20–24] (Sec. 3). In addition, the role of N and O co-doping in adsorption processes contributing to catalytic activity of activated carbon[25] (Sec. 4.1) is addressed and a concept of low-energy Ge(110) surface reconstruction reconciling two opposing models[26,27] (Sec. 4.2) is introduced.

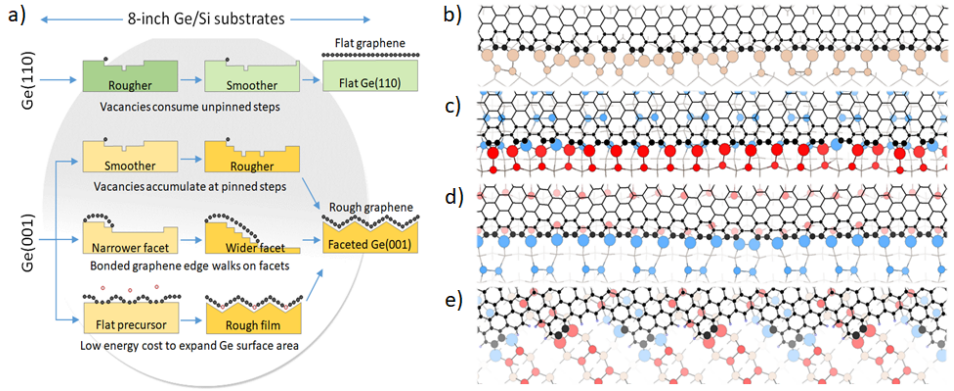


Figure 1. (a) The factors contributing to the response of Ge(110) and Ge(001) surfaces to graphene growth[12]. (b-e) Graphene bonded to straight steps on: (b) Ge(110) and (c-d) on Ge(001), and e) to a flat Ge(001) with some Ge dimer vacancy clusters attracted to the bonding front. The red and blue circles in the Ge(001) panels indicate the atoms of the two Ge dimerisation domains: the dimers in domain A are perpendicular to those in domain B. The elevation difference between the domains corresponds to a monatomic step height and equals to the depth of a dimer vacancy. The panel (e) illustrates the initial stage of roughening by (107) facet formation.

2 Graphene on Germanium Surfaces

Graphene is a sheet of sp^2-bonded carbon atoms. Among other applications, it is considered to be the fundamental material to enable technological breakthrough in microelectronics through synergistic combination of multiple 2D films and silicon technology[28]. The IHP evaluates its potential for usage in high-frequency transistors, in optical interconnects, and in sensors. The Materials Research department of IHP investigates the growth of graphene on Ge wafers and on Si wafers covered by Ge layers (Ge/Si), in the lab as well as under conditions relevant to mass production[12–14]. For this study, graphene layers were grown by chemical vapour deposition (CVD) on Ge from CH_4 and H_2 mixture at about 850°C. This is by roughly 150°C less than when grown on Cu, which may reduce the energy consumption[6], making this technology more friendly to the environment than the usual approach.

Ge(110) and Ge(001) surfaces respond differently to the same growth conditions: Ge(110) flattens and Ge(001) roughens. Fig. 1a collects the key factors causing this difference, as revealed by our work. It builds on the combination of extensive experimental work and the results of numerous DFT calculations for the total energies of atomic structures, for the energy barriers between the most important structures, the kinetic paths to their formation, and the experimental conditions. The analysis required knowledge on the surface energy differences, which are difficult to obtain from experiment (Sec. 4.2). Furthermore, the information gathered by us for these systems previously[9,29,10] was accounted for. This includes the energetics and kinetics of CH_4 adsorption and decomposition and the hypotheses on the formation mechanism of the (107) facets on Ge(001). The following discussion illustrates some of the aspects involved.

A clean Ge is surface is reactive against CH_4, causing the impinging molecules to dissociate step by step down to single C and H atoms. When exposed to CH_4, the surface

becomes initially covered by a mixture of CH_4 fragments and their polymers. The composition and concentration of this mixture depends on the substrate temperature, on the partial pressures of CH_4 and H_2, and – locally – on the size of the graphene nucleus that by chance has been formed within about a micrometer from the site of interest, *i. e.* , within the distance covered by an average C atom before it eventually dissolves in the bulk.

The smallest of these nuclei contain few C atoms and are mobile. They combine into larger, already immobile flakes. Van der Waals forces orient them parallel to the surface. Such a flake prefers to be attached to a surface step (Fig. 1b-d) rather than to the flat surface (Fig. 1e) because the in latter case it must be bent along the edge, which costs energy.

Being immobile, however, the larger flakes appear at random sites, not necessarily at steps. Yet the bonded Ge(001)-graphene boundary attracts mobile native defects (Ge adatoms, ad-dimers, surface vacancies). This creates new surface steps along the flakes. Furthermore, the flakes on Ge(001) can minimise their lateral strain as well. They achieve this by a slight rotation with respect to Ge dimer rows (Fig. 1e), which happens to orient the bonded boundary along the line of crossing between the Ge(001) and Ge(107) planes. The surface energy of Ge(107) is relatively low, so a (107) facet can extend away from the boundary. We verified that the flake can expand itself onto the facet simply by enlarging its "fingers" visible in Fig. 1e and then filling the regions between them with graphene. Eventually, the Ge(001) surface develops a system of (107) nano-facets (Fig. 1a), so that more CH_4 can convert into low-strained graphene than it would be possible on a flat Ge(001).

This is in opposition to what happens on Ge(110). While the reconstruction of Ge(001) is simple (just dimerisation of the surface atoms), Ge(110) has an intricate structure with sizeable building blocks (Sec. 4.2), which hinders the formation of graphene-induced facets. The alternative, preferred scenario is then that instead of accumulating in front of the bonded flake to form a new terrace (which would lead to surface roughening) the Ge vacancies produced by GeH_4 desorption[a] consume Ge(110) step edges, gradually removing the existing islands and making the surface smoother and smoother (Fig. 1a).

3 Hexagonal BN

Monolayer hBN is a 2D wide band-gap insulator with hexagonal lattice and atomic structure that resembles that of graphene (which is a 2D metal). Both materials have the form of a 2D honeycomb built of hexagonal atomic rings (C_6 in graphene, B_3N_3 in hBN). The lattice constant of hBN is by 1.7% larger than that of graphene.

Apart from optimising the graphene/Ge/Si growth recipe, we attempted the growth of heterostructures on Ge/Si and on Si, consisting of monolayer graphene and multilayer hBN films. Here we discuss the growth of the first hBN layer on Ge(001) and then the growth of subsequent hBN layers, that is, the growth of hBN on hBN.

The purpose of the calculations was to reveal the reactions of $B_xN_yH_z$ species with the substrate, the nucleation and expansion of seeds, and the kinetics of generation and annihilation of defects[20–24]. To address the posed questions one must treat a huge number of structures and structural transitions (our associated database contains by now more than 30k items), of size varying from isolated atoms and molecules, through isolated or periodic clusters of 100-200 atoms, up to periodic clusters consisting of more than a thousand atoms. The access to the JUWELS cluster was therefore crucial for this project.

[a] Hydrogen comes from CH_4 and from H_2

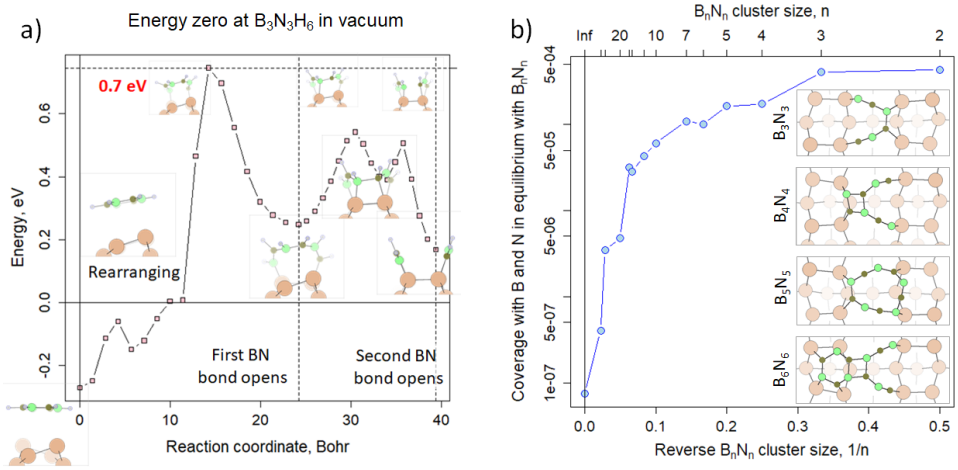


Figure 2. (a) Dissociative adsorption of $B_3N_3H_6$ on Ge(001): the path by direct BN ring splitting. The molecule in the leftmost insets is physisorbed, the rightmost insets show $B_3N_3H_6$ split into BN_2H_3 and B_2NH_3. B is brown, N is green, H is blue, Ge is sepia. (b) Ge(001) coverage with B and N atoms in equilibrium (900°C) with B_nN_n.

3.1 Nucleation and Growth of hBN from $B_3N_3H_6$ on Ge(001)

$B_3N_3H_6$ physisorbs on Ge(001) with no barrier, gaining 0.3 eV (Fig. 2a, leftmost insets). At the processing conditions, the barriers for subsequent chemisorpion are surmounted readily. For instance, when a H atom jumps from $B_3N_3H_6$ onto Ge, a pair of chemisorbed H and $B_3N_3H_5$ appears. The barrier depends on whether the dehydrogenated atom is B (0.4 eV, GeB bond formed) or N (0.7 eV, GeN bond formed). Alternatively, the physisorbed molecule splits into BN_2H_3 and B_2NH_3 (the barrier is 0.7 eV, Fig. 2a) and these fragments separate with a barrier of 1.8 eV (*i. e.* , within the range typical for objects mobile on Ge).

The adsorbed fragments are mobile and unstable: they collect H and desorb, or lose H and become more strongly bonded, or polymerise. $B_3N_3H_5$ can split by an H-assisted reaction similar to that shown in Fig. 2a. It is unclear if further opening of BN bonds is viable (although a BN dimer splits into B and N atoms), but what is crucial for the monolayer grow is that precursors with split BN rings are available. Namely, closed rings can produce only stoichiometric hBN with armchair edges, while with half-ring precursors any shape and any chemically realistic deviation from stoichiometry can be achieved.

The flake edges are strongly bonded to the substrate. For example, nearly all B and N atoms of a flake with its armchair edge perpendicular to dimer rows on flat Ge(001) make bonds with Ge atoms. At 900°C and H_2 partial pressures around 10^{-3} mbar (the processing conditions), H occupies about 0.1% of all edge atoms. Solely at H_2 pressures in the atmospheric regime (*e. g.* , at 100 mbar) this occupation may reach a few per cent.

The hBN flakes grow in a manner similar to that of graphene (Sec. 2): a cluster nucleated by chance collects material from its surrounding. Fig. 2b illustrates this on the example of the growth from atomic B and N. As the flake grows, the concentration of B and N in the surrounding decreases, reflecting the increasing stability of the flake. B_nN_n clusters with n equal to or barely exceeding 6 do not form sixfold rings (Fig. 2, insets).

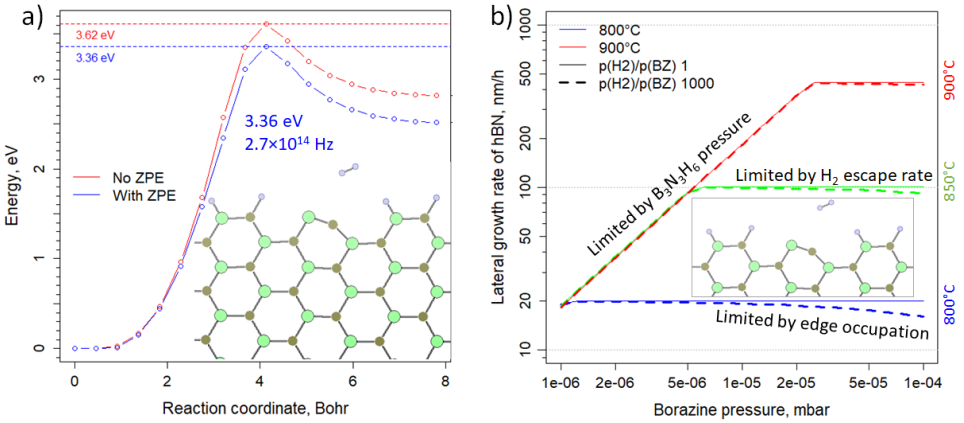


Figure 3. (a) Spontaneous H_2 emission from the hBN armchair edge, with and without Zero Point Energy (ZPE) correction. In the inset, N is green, B is brown, and H is blue. (b) Step flow velocity during growth of multilayer hBN. The hydrogenated sites are passive, only the H-free sites can adsorb $B_3N_3H_6$ from the gas phase. The attempt rate for spontaneous H_2 emission from armchair hBN edges was estimated from the transition state theory. The steric factor for $B_3N_3H_6$ attachment to an H-free site on the step edge is assumed to be 1.

This is an edge effect: such short or open rings occur on edges of even large flakes.

3.2 Growth of Multilayer hBN Films

Instead of stopping when the catalyst is fully covered by a single layer of the passive film, the hBN growth continues indefinitely. But no multilayer graphene would grow at comparable conditions from C_6H_6 (equivalent of $B_3N_3H_6$), and not even from more reactive CH_4, unless the precursors have been activated. Activation is thus expected also during hBN growth. We found that it must occur on hBN edges (Fig. 3a). In contrast to what happens on graphene, H_2 emission can keep the hBN growth front free of hydrogen even below 850°C (Fig. 3b). The sticking coefficient of $B_3N_3H_6$ to a hydrogen-free BN dimer is close to 1 because the energy barrier for this attachment is zero within the accuracy of the calculations and the molecule is pre-oriented by preliminary physisorption on the hBN surface. The armchair edge is free of hydrogen when the partial pressure of H_2 is 10^{-6} mbar; even when it is increased to 10^{-2} mbar, only a few per cent of the edge sites are passivated. At the $B_3N_3H_6$ partial pressure of 10^{-6} mbar the time to reach this equilibrium is only marginally longer than the average time that elapses between two subsequent collisions of $B_3N_3H_6$ with the edge. The attempt rate for H_2 escape from the dimer was estimated as $3{\cdot}10^{14}$ Hz. The growth rate obtained from the computed step flow velocity (Fig. 3b) and from the step-step distance estimated from TEM images agrees within the numerical and experimental uncertainties with the observed growth rates.

We have also tested the hypothesis that another process is responsible for the activation. We considered the possibility of H out-diffusion along screw dislocations in hBN from the growth front onto Ge, of catalytic activation of $B_3N_3H_6$ to $B_3N_3H_4$ or of H_2 to atomic H on hot elements in the chamber, and of catalytic activation of $B_3N_3H_6$ on orientational grain boundaries in hBN. The mechanism on the second place in the plausibility ranking is the

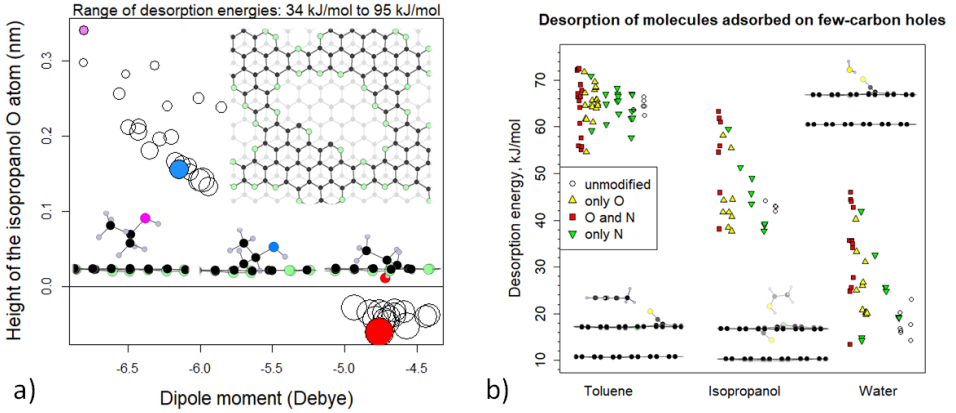


Figure 4. (a) Correlation between dipole moment, oxygen penetration depth, and desorption energy (circle diameter) for a hole in the top layer of bilayer graphene (inset, top view). The hole edge is saturated by pyridine N. (b) Overview of desorption energies computed for toluene, isopropanol, and water adsorbed on various few-atom holes with various N and O functional groups. The insets illustrate the strongest adsorption cases (side view).

activation on hot elements, but it appears to be incompatible with the possibility to grow hBN at temperatures and pressures well below the limits estimated from the calculations. Therefore, spontaneous H_2 emission (Fig. 3) comes out as the only known realistic answer.

4 Examples of Work in Progress

4.1 Catalytic Activity of 2D Films

Activated carbons have enormous application potential as catalysts[30], *e. g.* in the oxidative dehydrogenation of hydrocarbons[31] or in the oxidation of SO_2 to SO_3[32]. Defects and anchors of functional groups influence the surface interactions with the molecules from the liquid or gas phase and their activation towards catalytic transformations[33].

Catalytic activity of a surface begins with adsorption of the participating species. We currently investigate the influence of N and O functional groups on adsorption of test molecules on openings in bilayer graphene (Fig. 4). The model substrate is periodic to assure that its density of states is metallic, as in our experiments[25], in which structurally comparable samples of porous activated carbon are prepared by pyrolysis of sucrose ($C_{12}O_{11}H_{22}$), pure or with urea ($CO(NH_2)_2$) admixtures. The calculations show that the desorption energy of isopropanol (C_3H_7OH, a polar test molecule that is able to form hydrogen bonds through its OH group) correlates positively with the reduction of the dipole moment in the adsorbate-substrate system; this corresponds to the molecule being partially drawn into the opening (Fig. 4a). Nitrogen doping of an O-containing substrate has the tendency to increase the desorption energies of isopropanol and water even when the openings are only of the size of a few C atoms, but the effect on toluene ($C_6H_5CH_3$, non-polar test molecule that forms no hydrogen bonds) is practically non-existent (Fig. 4b).

Experimental verification of the computed desorption energies by temperature programmed desorption (TPD) measurements turned out to be difficult. The activation en-

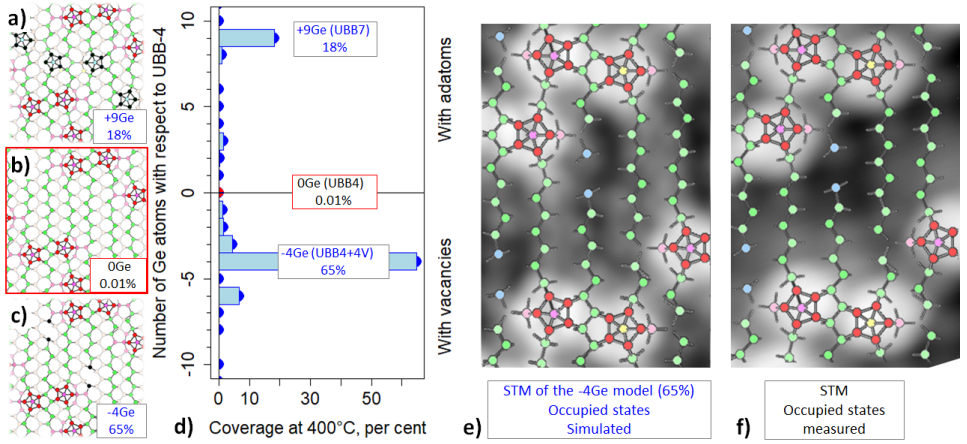


Figure 5. Ge(110) reconstructions. (a-c) Representative 8×10 reconstruction models. Pentamers are red, ordinary (110) zig-zag atoms are green. (a) The densified UBB7 model (+9Ge, 18% at 400°C), the additional pentamers are black. (b) The original UBB4 model[27] (0Ge, abundance 0.01%). (c) The UBB4+4V model (-4Ge, 65%) is on the bottom, the vacancy neighbours are black. (d) Approximate abundance of structures with various net balance of Ge atoms. (e-f) STM images of occupied states. (e) simulated for UBB4+4V, (f) measured.

ergies extracted from the heating rate dependence of the TPD spectra were unphysically low, which indicates that the spectra were dominated by desorption from long nano-pores, where re-adsorption and bottlenecks in the gas flow play the dominant role. Experiments on samples prepared by a modified procedure are in progress.

4.2 Ge(110) Surface Reconstruction

Knowing the atomic structure of a surface is prerequisite for reliable simulation and in-depth explanation of chemical and atomic-scale surface processes (Sec. 2). Clean Ge(110) is covered by objects visible in STM images as pentamers of atomic-sized spots. After long anneals these pentamers tend to arrange in rows separated from one another by strips of apparently flat surface (Fig. 5f). The atomic origin of the pentamers has been the subject of a long debate. We reconciled two structural models recently proposed for these pentamers: the adatom-based Universal Building Block model (UBB)[27] (Fig. 5b, 0Ge or UBB4) and the vacancy-based Tetramer Heptagonal- and Tetragonal Ring model (THTR)[26]. The search for a new model was motivated by our discovery that in contrast to the expectations from experiments, the UBB4 Ge(110) 8×10 surface has a tendency to acquire additional UBB pentamers (Fig. 5a, +9Ge or UBB7). This behaviour is suppressed and the surface energy is reduced when the UBB4 structure is augmented by two vacancy pairs (Fig. 5c, -4Ge or UBB4+4V), each rebonded as in THTR. The STM images simulated for this model are compatible with experiment (Fig. 5e-f).

5 Approach

Pseudopotential plane-wave DFT calculations were done with Quantum Espresso[34] using the PBE functional[35] and in special cases by the hybrid B3LYP functional[36,37]. Van der

Waals forces were treated as non-local DFT (rVV10)[38]. Reaction paths were determined with the nudged elastic band - climbing image (NEB-CI) algorithm[39]. The accuracy of pseudopotential calculations for isolated molecules was verified by all-electron calculations with NWChem[40], using the double-hybrid B2PLYP[41] functional. Vibrational spectra were computed within the density functional perturbation theory (DFPT)[42].

6 Concluding Remarks

Insight into problems pertinent to materials development for environmentally friendly technologies was obtained from DFT simulations coupled to experiment. The difference in the surface evolution of Ge(001) and Ge(110) in response to graphene growth by CVD was explained (Sec. 2) and, as a fringe benefit, a structural model reconciling two dissimilar concepts of Ge(110) reconstructions was formulated (Sec. 4.2). The growth mechanism of seeding (on Ge) and growth of multilayer hBN films was proposed and positively verified against the measured growth rates (Sec. 3). The role of N/O co-doping of activated carbon in adsorption of polar and non-polar molecules was addressed (Sec. 4.1).

Acknowledgements

The calculations were performed with a grant of computer time provided by the John von Neumann Institute for Computing on the JUWELS cluster at the Jülich Supercomputing Centre (JSC) in the compute projects IHPms21 and IHPms24. J. D., F. A., M. F., and M. L. gratefully acknowledge funding from the EU FLAG-ERA project 2DHetero[24]. Also the financial support by the German Research Foundation (DFG; grant Nr. KL 1202/17-1)[25] of the experimental work on the activated carbon is highly appreciated. We thank Andreas Becker for STM measurements.

References

1. C. L. Lai and S. H. Lin, *Treatment of chemical mechanical polishing wastewater by electrocoagulation: system performances and sludge settling characteristic*, Chemosphere, **54**, 235, 2004.
2. E. A. Noman et al., *An insight into microelectronics industry wastewater treatment, current challenges, and future perspectives: a critical review*, Applied Water Science, **14**, 64, 2024.
3. Y. Liu, H. Wang, Y. Cui, and N. Chen, *Removal of Copper Ions from Wastewater: A Review*, International Journal of Environmental Research and Public Health, **20**, no. 5, 2023.
4. X. Li, H. Jin, Y. Chan, H. Guo, and W. Ma, *Environmental impacts of graphene at industrial production scale and its application in electric heating technology*, Resources, Conservation and Recycling, **199**, 107250, 2023.
5. Y. Chen, X.-L. Gong, and J.-G. Gai, *Progress and Challenges in Transfer of Large-Area Graphene Films*, Advanced Science, **3**, 1500343, 2017.

6. R. Arvidsson, D. Kushnir, S. Molander, and B. A. Sandén, *Energy and resource use assessment of graphene as a substitute for indium tin oxide in transparent electrodes*, Journal of Cleaner Production, **132**, 289, 2016.
7. A. Ambrosi and M. Pumera, *The CVD graphene transfer procedure introduces metallic impurities which alter the graphene electrochemical properties*, Nanoscale, **6**, 72, 2014.
8. G. Lupina, J. Kitzmann, I. Costina, M. Lukosius, C. Wenger, A. Wolff, S. Vaziri, M. Östling, I. Pasternak, A. Krajewska, W. Strupinski, S. Kataria, A. Gahoi, M. C. Lemme, G. Ruhl, G. Zoth, O. Luxenhofer, and W. Mehr, *Residual metallic contamination of transferred chemical vapor deposited graphene*, ACS nano, **9**, 4776, 2015.
9. G. Lippert, J. Dabrowski, T. Schroeder, Y. Yamamoto, F. Herziger, J. Maultzsch, J. Baringhaus, C. Tegenkamp, M. C. Asensio, J. Avila, and G. Lupina, *Graphene on Ge(001) from atomic source*, Carbon, **75**, 104, 2014.
10. J. Dabrowski, G. Lippert, J. Avila, J. Baringhaus, I. Colambo, Yu. S. Dedkov, F. Herziger, G. Lupina, J. Maultzsch, T. Schaffus, T. Schroeder, M. Sowinska, C. Tegenkamp, D. Vignaud, and M.-C. Asensio, *Understanding the growth mechanism of graphene on Ge/Si(001) surfaces*, Sci. Rep., **6**, 31639, 2016.
11. M. Lukosius et al., *Metal-Free CVD Graphene Synthesis on 200 mm Ge/Si(001) Substrates*, ACS Appl. Mater. Interfaces, **8**, 33786, 2016.
12. F. Akhtar, J. Dabrowski, R. Lukose, C. Wenger, and M. Lukosius, *Chemical Vapor Deposition Growth of Graphene on 200 mm Ge(110)/Si Wafers and Ab Initio Analysis of Differences in Growth Mechanisms on Ge(110) and Ge(001)*, ACS Appl. Mater. Interfaces, **15**, 36966, 2023.
13. F. Akhtar, *Graphene synthesis under Si-CMOS compatible conditions*, PhD thesis, BTU Cottbus, April 2022.
14. A. Becker, Ch. Wenger, and J. Dabrowski, *Influence of temperature on growth of graphene on germanium*, J. Appl. Phys., **128**, 2020.
15. J. Dabrowski and V. Zavodinsky, *Ab initio study of Pr oxides for CMOS technology*, NIC Series, **20**, 171, 2004.
16. J. Dabrowski, A. Fleszar, G. Lippert, and G. Lupina, *Ab initio modelling of growth of graphene for Si-compatible microelectronics*, NIC Series, **47**, 207, 2014.
17. J. Dabrowski, G. Lippert, and G. Lupina, *Nucleation of graphene on the Ge(001) p2× 2 surface*, NIC Symposium, **8**, MAT 6, 2016.
18. J. Dabrowski, G. Kissinger, G. Lipptert, G. Lupina, M. Lukosius, P. Sana, and T. Schroeder, *Modelling of materials for silicon-compatible microelectronics*, NIC Series, **49**, 239, 2018.
19. J. Dabrowski, F. Reichmann, M. Franck, and M. Lukosius, *Ab initio simulations for microelectronics: $ZnGa_2O_4$ and hexagonal BN*, NIC Series, **51**, 245, 2022.
20. M. Franck, J. Dabrowski, M. A. Schubert, C. Wenger, and M. Lukosius, *Towards the Growth of Hexagonal Boron Nitride on Ge(001)/Si*, Nanomaterials, **12**, 3260, 2022.
21. M. Franck, J. Dabrowski, M. A. Schubert, D. Vignaud, M. Achehboune, J.-F. Colomer, L. Henrard, C. Wenger, and M. Lukosius, *Investigating impacts of local pressure and temperature on CVD growth of hexagonal boron nitride on Ge(001)/Si*, Advanced Materials Interfaces, **12**, 2400467, 2024.

22. W. Batista-Pessoa, M. Franck, N. Nuns, J. Dabrowski, M. Achehboune, J.-F. Colomer, L. Henrard, M. Lukosius, X. Wallart, and D. Vignaud, *Optimized two-step growth of large surface two-dimensional boron nitride on Ge(001) films by molecular beam epitaxy*, submitted to Applied Surface Science, 2024.
23. M. Achehboune, K. Zhour, J. Dabrowski, D. Vignaud, M. Franck, M. Lukosius, J.-F. Colomer, and L. Henrard, *Atomistic insights into the nucleation and growth of hexagonal boron nitride and graphene heterostructures*, Phys. Chem. Chem. Phys., **26**, 28198, 2024.
24. "2DHetero: hBN/graphene 2D Heterostructures: from scalable growth to integration", EU FLAG-ERA project.
25. O. Klepel, "Interplay of nitrogen- and oxygen functionalities of carbon based catalysis in redox reactions", DFG, KL1202/17-1, since 2019.
26. T. Yamasaki, K. Kato, T. Uda, and T. Yamamoto, *First-principles theory of Si(110)-(16 × 2) surface reconstruction for unveiling origin of pentagonal scanning tunneling microscopy images*, Applied Physics Express, **9**, 035501, 2016.
27. R. A. Zhachuk and A. A. Shklyaev, *Universal building block for (110)-family silicon and germanium surfaces*, Applied Surface Science, **494**, 46, 2019.
28. D. Akinwande, C. Huyghebaert, C.-H. Wang, M. I. Serena, S. Goossens, L.-J. Li, H-S. P. Wong, and F. H. L. Koppens, *Graphene and two-dimensional materials for silicon technology*, Nature, **573**, 507, 2019.
29. J. Dabrowski, G. Lippert, and G. Lupina, *Initial state of graphene growth on Ge(001) surfaces*, ECS Transactions, **69**, 345, 2015.
30. X. Liu, A. Cheruvathur, and R. Sharpe, *Carbon-based metal-free catalysis for dehydrogenation of hydrocarbons*, in: Metal-free Functionalized Carbons in Catalysis: Synthesis, Characteris, and Applications, The Royal Soc. Chem., 196, 2018.
31. I. Szewczyk et al., *Electrochemical Denitrification and Oxidative Dehydrogenation of Ethylbenzene over N-doped Mesoporous Carbon: Atomic Level Understanding of Catalytic Activity by 15N NMR Spectroscopy*, Chem. Mater., **32**, 7263, 2020.
32. E. Raymundo-Piñero, D. Cazorla-Amorós, C. Salinas-Martinez de Lecea, and A. Linares-Solano, *Factors controling the SO_2 removal by porous carbons: relevance of the SO_2 oxidation step*, Carbon, **38**, 335, 2000.
33. E. Vottero et al., *Assessing the functional groups in activated carbons through a multi-technique ap-proac*, Catal. Sci. Technol., **12**, 1271, 2022.
34. P. Giannozzi et al., *QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials*, J. Phys. Cond. Mat., **21**, 395502, 2009, see also: What Can Quantum Espresso Do? `http://www.quantum-espresso.org/whatcanqedo.php`.
35. J. P. Perdew, K. Burke, and M. Ernzerhof, *Generalized gradient approximation made simple*, Phys. Rev. Lett., **77**, 3865, 1996.
36. A. D. Becke, *Density-functional thermochemistry. III. The role of exact exchange*, J. Chem. Phys., **98**, 5648, 1993.
37. P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, *Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields*, J. Phys. Chem., **98**, 11623, 1994, and references therein.
38. R. Sabatini, T. Gorni, and S. de Gironcoli, *Nonlocal van der Waals density functional made simple and efficient*, Phys. Rev. B, **87**, 041108(R), 2013.

39. G. Henkelman, B. P. Uberuaga, and H. Jónsson, *A climbing image nudged elastic band method for finding saddle points and minimum en-ergy paths*, J. Chem. Phys., **113**, 9901, 2000.
40. E. Apra et al., *NWChem: Past, present, and future*, J. Chem. Phys, **152**, 184102, 2020.
41. S. Grimme, *Semiempirical hybrid density functional with perturbative second-order correlation*, J. Chem. Phys., **124**, 34108, 2006.
42. S. Baroni et al., *Phonons and related properties of extended systems from density-functional perturbation theory*, Rev. Mod. Phys., **73**, 515, 2001.

On the Oxygen p States in Superconducting Nickelates

Frank Lechermann

Institut für Theoretische Physik III, Ruhr-Universität Bochum, 44780 Bochum, Germany
E-mail: frank.lechermann@rub.de

While key attention in transition-metal oxides is usually devoted to the d states of the transition-metal ion, the O$(2p)$ states usually also carry important physics. We here examine these p states in representatives of the novel superconducting nickelates, as described in realistic dynamical mean-field theory. Since the materials are located on the boundary between Mott-Hubbard and charge-transfer systems, the role of oxygen is expectedly subtle. Strong reduction of doped holes on oxygen and first asymmetry effects are featured in infinite-layer nickelates. A pronounced nature of bridging p_z orbitals is identified in the $La_3Ni_2O_7$ system.

1 Introduction

In late summer 2019, the long-sought discovery of superconductivity in nickel oxides finally opened up a new research field in condensed matter physics. Strontium doping of thin films of so-called infinite-layer $NdNiO_2$ on $SrTiO_3$ substrates leads to a superconducting phase below $T_c \sim 15\,\mathrm{K}$[1]. Early follow-up works revealed further similar superconducting scenarios in Sr-doped (La,Pr)NiO_2, as well as in the stochiometric multilayer compound $Nd_6Ni_5O_{12}$[2]. All these low-valence nickelate materials with Ni$(3d^{9-\delta})$ filling share the fact that the apical oxygens of the basic NiO_6 octahedra are missing.

In early spring 2024, a second class of superconducting nickelates emerged. The bilayer $La_3Ni_2O_7$ was reported superconducting under high pressure $p > 14\,\mathrm{GPa}$ with a much higher $T_c \sim 80\,\mathrm{K}$[3]. Soon after, the trilayer compound $La_4Ni_3O_{10}$ was proven to also show superconductivity in a similar pressure regime, but with an again lower $T_c \sim 20\,\mathrm{K}$[4]. Those nickelates with Ni$(3d^{8-\delta})$ filling have intact NiO_6 octahedra.

From the start, the superconducting nickelates were compared to high T_c cuprates, because of their proximity in the periodic table and the akin building block of square-lattice transition-metal (TM) oxide planes. However the degree of similarity in terms of physical properties and superconducting nature is still under heavy debate. For detailed representations of the known and discussed features from experiment and theory, we here refer to available early review articles, e.g. Refs. 5–8. One key issue concerns the number of relevant Ni$(3d)$ frontier orbitals. While there is strong consensus that only the $d_{x^2-y^2}$ TM orbital is of crucial importance in high-T_c cuprates, in general electronic properties of nickelates the complete e_g subshell $\{d_{z^2}, d_{x^2-y^2}\}$ has to be taken into account. At least for the $d^{8-\delta}$ bilayer and trilayer superconducting compounds, a pure single-orbital cuprate(-like) physics seems very unrealistic both from theory and experiment.

Yet the present work does not focus on the detailed characteristics of the Ni$(3d)$ degrees of freedom. Instead, it takes a deeper look into the behaviour of the O$(2p)$ states in the normal state of superconducting nickelates. While the impact of those p states is more subtle, some interesting aspects may still be revealed and learned from first-principles many body theory. We show that the ligand-hole concept, i.e. deviation from the simplistic purely-ionic O^{2-} picture, is a steady companion in these nickel oxides. It asks to be properly

considered, weighed and addressed, both in the low- and high-energy regime. Noteworthy differences in the energetics and the occupation between p_z and $p_{x,y}$ orbitals are observed.

2 Theoretical Background and Approach

The physics of strong electron correlation in materials such as the superconducting nickelates asks for a proper treatment of both, the realistic band theoretical aspect as provided by the chemical bond in the solid, as well as the explicit many-body aspect of interacting electrons. The state-of-the-art approach to realise such a faithful description is given by the hybrid scheme of combining density functional theory (DFT) with dynamical-mean field theory (DMFT), i.e. the so-called DFT+DMFT method (see, e.g. Ref. 9 for the basics). There, usually the transition-metal (TM) sites are the centre of attention, serving as DMFT impurities, while the description of the ligand states remains on the DFT level (albeit surely coupled to the strongly correlated TM sites). However, in nickelates, we are most often dealing with possible low-energy states of strongly hybridised TM($3d$)-O($2p$) character. This is different compared to early TM oxides, e.g. titanates or vanadates, were the O($2p$) states only weakly hybridise into the low-energy physics and the systems reside more robustly in the so-called Mott-Hubbard regime of correlated materials[10]. There, the effective Hubbard U_{eff} is smaller than the charge-transfer energy $\Delta = \varepsilon_d - \varepsilon_p$, with $\varepsilon_{d,p}$ describing the average onsite energy of the TM($3d$) and O($2p$) states.

From a basic quantum-scattering perspective, TM($3d$) states are difficult to handle by sole band theory. Because the $3d$ orbitals are not orthogonalised on lower lying d orbitals, their electrons approach the ion core region quite easily. This leads to the observed strong competition between the electrons' itinerant vs. localised character in the solid state, giving rise to the plethora of correlation effects emerging from TM($3d$) compounds. Notably, also O($2p$) is not orthogonalised on lower lying p states and therefore the electrons are also more localised than $3p$ or even higher p electrons. While due to the high electronegativity, smaller orbital Hilbert space and strong bonding tendencies, the quantum-fluctuating character of O($2p$) is much less pronounced than TM($3d$), correlation effects beyond DFT can still matter also from that perspective. This may especially be true when O($2p$) plays a more prominent role in the low-energy regime of late TM oxides. Therefore, on the methodological level, the differentiation in the correlation treatment of TM($3d$) and O($2p$) states may be too severe in standard DFT+DMFT for certain classes of materials. And understandably, issues such as the pd splitting, p-induced renormalisation effects or intriguing ligand-mediated exchange might not be well described. There are several ideas in going beyond this strong-differentiation treatment, for instance by combining DMFT with the GW method[11–13] or with screened-exchange formalisms[14]. Our choice builds up on introducing the self-interaction correction (SIC) scheme for oxygen orbitals on the pseudopotential level[15], to be utilised in a complete charge-selfconsistent DFT+DMFT framework[16]. This so-called DFT+sicDMFT scheme[17] improves the pd-splitting description and handles p-induced correlation effects (e.g. explicit oxygen-mediated band narrowing). The SIC naturally enables correlation effects free from the need of symmetry breakings, but is numerically much less demanding as, e.g. GW+DMFT.

In the present context of superconducting nickelates, the Ni sites act as DMFT quantum impurities and Coulomb interactions on oxygen enter by the SIC-modified pseudopotentials. The DFT part consists of a mixed-basis pseudopotential code[18–20] and SIC is applied

to the O$(2s, 2p)$ orbitals via weight factors w_p. While the $2s$ orbital is fully corrected with $w_p = 1.0$, the choice[15,17,21] $w_p = 0.8$ is used for $2p$ orbitals. Continuous-time quantum Monte Carlo in hybridisation-expansion scheme[22] as implemented in the TRIQS code[23,24] solves the DMFT problem. A five-orbital Slater-Hamiltonian, parameterised by Hubbard $U = 10\,\mathrm{eV}$ and Hund exchange $J_\mathrm{H} = 1\,\mathrm{eV}$[21], governs the correlated subspace defined by Ni$(3d)$ projected-local orbitals[25]. Crystallographic data are taken from experiment.

3 General Remarks on Oxygen p States in Nickelates

Nickelates are special in their correlation physics. While cuprates are usually located in the strong charge-transfer regime, i.e. the effective Hubbard U_eff is way larger than the charge-transfer energy Δ, the nickelates are closer to a competing regime of Mott-Hubbard versus charge-transfer dominance. Note that the calculational $U > U_\mathrm{eff}$ in DFT+sicDMFT, since further screening processes occur in the charge-selfconsistent many-body approach. The levelling of U_eff and Δ leads to a subtle role of O$(2p)$ regarding the interplay of its low- and high-energy electronic character. In a Mott-Hubbard system, the oxygen p states reside in their localised high-energy being, whereas their itinerant nature is most vital in a charge-transfer system. The subtlety is also reflected in the ligand-hole physics of in fact many nickelates. Fig. 1 displays the evolution of the charge-transfer energy Δ and the amount of holes in the oxygen $2p$ shell (per single O ion) h_p from high to low valence of Ni in given Ruddlesden-Popper(-like) nickelates. The data for Δ is computed from DFT+sic, i.e. a conventional Kohn-Sham calculation but using the SIC-modified oxygen pseudopotential. As the result of the fully-interacting problem, the data for h_p is obtained from DFT+sicDMFT.

From formal Ni^{3+} to formal Ni^{+}, the value of Δ rises more or less monotonically, as expected from the ionic physics of these elements. This means, O$(2p)$ and Ni$(3d)$ split energetically stronger when the Ni valence becomes smaller. The filling of the O$(2p)$ shell

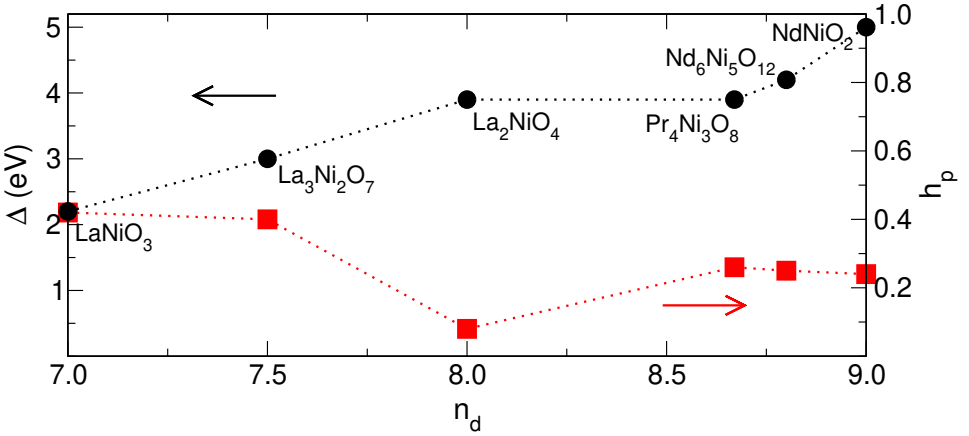


Figure 1. Charge-transfer energy Δ and O$(2p)$ hole content h_p for selected standard ($n_d \leq 8$) and reduced ($n_d > 8$) Ruddlesden-Popper nickelates with formal Ni$(3d)$ count n_d. Note the plateau-like region in Δ for $8 < n_d < 8.67$.

is set by Δ, hoppings as well as by the Coulomb interactions in the system, most notably the on-site Coulomb U_{pp} and the inter-site Coulomb U_{pd}. In other words, the standard identification of O^{2-} in oxides is not necessarily always true. For instance, this ionic configuration amounts to a large Coulomb penalty within O($2p$), which should be properly considered within the full group of mechanisms on the given lattice that eventually fix h_p. Intuitively, we expect for small Δ a larger h_p since electrons can more easily be transferred between sites, i.e. the degree of covalency is increased. Indeed, for formal Ni^{3+} the hole content on oxygen is substantial, amounting to roughly one charge unit within the unit cell. This ligand-hole $3d^8\underline{L}$ state is well-known for nickelates with nominal higher valence than the ideal $2+$ oxidation in the solid[26]. However, as shown in Fig. 1 this hole content does not fall to zero for less than nominal Ni^{2+}. The situation is more intriguing due to the various electronic mechanisms at play, leading to quite some finite covalency also for the lower-valence nickelate. Again, the O^{2-} fully-ionic picture is usually fine as a first guess, but by no means always the full truth in a complex realistic many-body compound. This limit is here only reasonably-well reached for the Mott-charge-transfer insulator La_2NiO_4.

While the existing basic Mott vs. charge-transfer schemes provide good principle guidance, the actual charge distribution in a concrete material is quite sophisticated, building up on the interplay of a multitude of physical processes (in a Hilbert space enclosing usually more than only $3d$ and $2p$ states). Boiling down the full physics to $U_{\rm eff}$ and Δ only, often appears too narrow. In this regard, also a more detailed high- vs- low-energy differentiation appears of vital importance.

4 Infinite-Layer $NdNiO_2$ System

We focus on the superconducting nickelates and first analyse the O($2p$) states in low-valence $NdNiO_2$ at stoichiometry and with finite hole doping. To set the stage, Fig. 2 summarises the established **k**-resolved picture within DFT+sicDMFT[21,27]. At stoichometry, Ni-$d_{x^2-y^2}$ is (nearly) Mott-insulating and finite conductivity is (mainly) carried by weakly-filled self-doping bands (cf. Fig. 2a). The latter have mixed character of Nd($5d$) as well as Ni-d_{z^2} (around Γ) and Ni-$d_{xz,yz}$ (around A) (see Fig. 2b,e). With hole doping, the flat-band part of Ni-d_{z^2} for $k_z = 1/2$ is shifted towards the Fermi level (cf. Figs. 2d), presumably playing a crucial role in the emergence of superconductivity. This prominent role of the Ni-d_{z^2} flat-band part is similarly revealed in GW+DMFT[28]. Note that standard DFT+DMFT studies without correlations on oxygen result in weaker correlations for Ni-$d_{x^2-y^2}$, enabling a more cuprate(-like) picture for superconductivity (e.g. Refs. 29,30).

Fig. 1 reveals that even with a quite large $\Delta = 5.0\,$eV, the hole content h_p amounts to 0.24. The resulting **k**-resolved character of the O($2p$) states is depicted in Fig. 2c,f. In the stoichiometric case, O($2p$) dominated dispersions are easily visible in the energy window $\sim [-10, -5]\,$eV. With hole doping, the oxygen p states mix-in more strongly near the Fermi level and into the shifted self-doping bands above. For instance, around the R point, the flat band at $\varepsilon_{\rm F}$ is a strong mixture of Ni-$d_{x^2-y^2}$, Ni-d_{z^2} and O($2p$). Interestingly, the weak Ni-$d_{x^2-y^2}$ low-energy weight in the $k_z = 0$ plane with Γ, X, M is not sizeably hybridised with the p states. This underlines the high-Δ character with seemingly weak Zhang-Rice nature[31] of it's low-energy physics upon hole doping.

Weak Zhang-Rice character is also observable in the low-energy part of the **k**-integrated spectra, which is displayed in a larger energy range in Figs. 3(a-c). While the

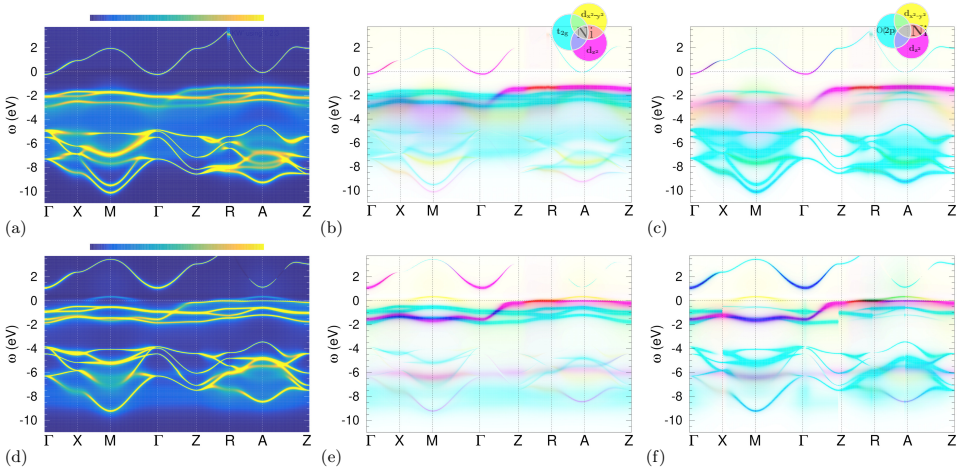


Figure 2. **k**-resolved spectral function for stoichiometric $NdNiO_2$ (a-c) and with 15% hole doping (d-f). (a,d) full $A(\mathbf{k}, \omega)$. (b,e) Fatspec representation resolving Ni-$d_{x^2-y^2}$ (yellow), Ni-d_{z^2} (pink) and Ni-t_{2g} (cyan). (c,f) Same as (b,e) but Ni-t_{2g} replaced by O($2p$).

small p weight is comparable to the small d weight at stoichiometry $\delta = 0$ close to the Fermi level, the low-energy d weight strongly dominates over p for finite hole doping δ. Notably, as drawn from the integrated p_z and $(p_x + p_y)$ spectral weight in Figs. 3(d-f), the overall oxygen hole content only increases marginally with substantial hole doping from replacing La by Sr. In other words, most doped holes do not localise on oxygen as in cuprates, but elsewhere, in agreement with electron energy-loss spectroscopy[32]. In the

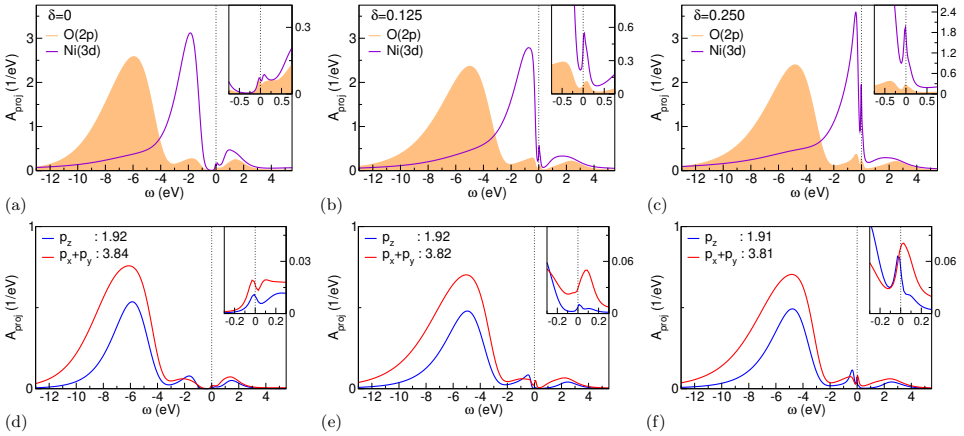


Figure 3. **k**-integrated spectral function for stoichiometric $NdNiO_2$ (a,d), with 12.5% (b,e) and 25% (c,f) hole doping. (a-c) Projected O($2p$) and Ni($3d$) spectrum. (d-f) Projected O-p_z and O-$(p_x + p_y)$ spectrum, with numbers providing the respective filling.

DFT+sicDMFT calculations, doped holes are mostly hosted in the Ni-d_{z^2} orbital[21]. Besides the total O($2p$) occupation, a possible asymmetry between p_z and $(p_x + p_y)$ filling may have impact[33]. But the data shown in Figs. 3(d-f) show only rather weak asymmetry for the *total* p filling. Yet there is some asymmetry in the low-energy weight, especially with hole doping. For $\delta = 0.125$, the p_z weight close to $\varepsilon_{\rm F}$ is quite small, while for $\delta = 0.25$ it reaches nearly the same value as the one for $(p_x + p_y)$. This may be understandable from the special role of p_z, not meeting a Ni atom in nearest-neighbour distance. For larger δ, low-energy p_z weight becomes more easily coherent with sizeable Ni-d_{z^2} content in the same energy range.

5 $La_3Ni_2O_7$ System

The present challenges of the $La_3Ni_2O_7$ system, representing here the Ni($3d^{8-\delta}$) superconducting-nickelate class, are not linked to missing apical oxygens or additional doping features. Here, the sophistication lies in complexities of structural kind and the subtle modifications realised in the high-pressure regime. Canonically, the $La_3Ni_2O_7$ compound is identified as the bilayer, so-called '2222', representant of the m-layered Ruddlesden-Popper series $La_{m+1}Ni_mO_{3m+1}$. Recently, an alternation of mono- and trilayers, so-called '1313', has been found as a competing structural motif[34–36]. The distinct impact of high pressure on the electronic structure is heavily debated. At presence, it seems that the (non-)appearance of a Ni-d_{z^2}-based flat band depends on pressure[8].

Here, we again want to focus on the O($2p$) degrees of freedom and their main characteristics in the correlated electronic structure. Nominally, a $Ni^{7.5}$ filling results from an ionic-limit picture assuming O^{2-} and La^{3+}. However, the charge-transfer energy $\Delta = 3.0$ is comparatively low, and correspondingly, the ligand-hole content $h_p = 0.40$ rather high (see Fig. 1). In general, nickelates with formal oxidation states higher than $Ni^{2+}(d^8)$ form ligand holes on oxygen to keep an effective Ni^{2+} configuration. From DFT+sicDMFT, this is also true for $La_3Ni_2O_7$ with near d^8 filling and one electron/hole in the d_{z^2} and the $d_{x^2-y^2}$ orbital, respectively. Recent experiments[37] confirm the theoretical predictions of existing ligand holes. Figs. 4(a-c) show that also the low-energy p-weight is somewhat enhanced compared to the infinite-layer case, though a very strong Zhang-Rice picture as in cuprates does still not emerge. Furthermore, applied pressure appears effective to shift both, main Ni($3d$) and main O($2p$) peak to slightly deeper energies in the occupied spectrum.

It proves informative to plot the p_z- and $(p_x + p_y)$-resolved spectral weight for the different symmetry-inequivalent oxygen sites in the given primitive cells, as done in Figs. 4(d-f). For instance, it may be observed that for all three discussed cases here, i.e. ambient-pressure 2222, high-pressure 2222 and high-pressure 1313, the $p_{x,y}$ orbitals in the LaO fluorite block separating the NiO_2 multilayers are responsible for the high-energy ligand-hole peak in the unoccupied spectrum (dashed dark curves from the dark-oxygen-ion symmetry class). One also realises that the apical oxygens connecting NiO_2 (red-coloured ions), here termed bridging (BR) oxygens, show generally the largest energy splitting between p_z and $p_{x,y}$ orbitals. The most interesting behaviour is attributed to the low-energy region around the Fermi level, where the p_z orbital of the BR oxygens has apparently a prominent role. Especially for ambient-pressure 2222, the corresponding BR p_z orbital has the strongest and peaked spectral weight at $\varepsilon_{\rm F}$, and furthermore a dominant low-

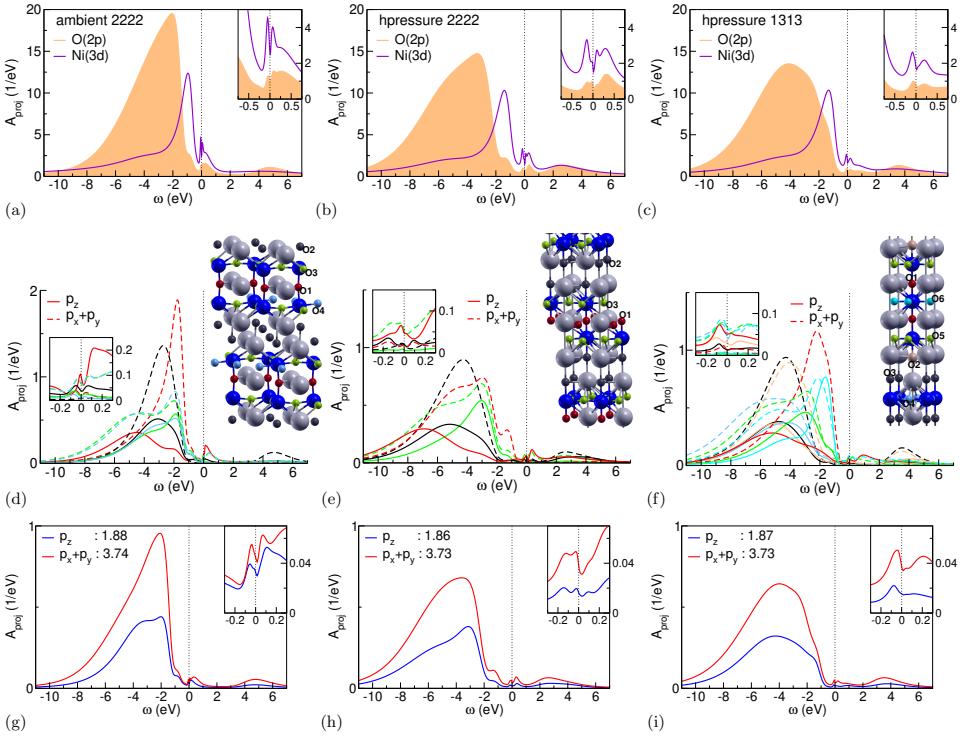


Figure 4. **k**-integrated spectral function for $La_3Ni_2O_7$, with La: grey, Ni: darkblue. (a,d,g) 2222 structural motif at ambient pressure and (b,e,f) at high pressure, as well as (c,f,i) 1313 structural motif at high pressure. (a-c) Projected O($2p$) and Ni($3d$) spectrum. (d-f) Projected O-p_z and O-($p_x + p_y$) spectrum for symmetry-inequivalent oxygen sites, respectively. (g-i) As (d-f), but summed over all oxygen sites and with numbers providing the respective filling. Note that possible small asymmetries between $p_{x,y}$ are averaged in the plots.

energy ligand-hole weight up to $\sim$ 0.3 eV above the Fermi level (see Fig. 4d). Second in low-energy weight are the in-plane $p_{x,y}$ orbitals (green- and lightblue-coloured ions) of the NiO_2 planes, in line with experiment[37]. Hence there is a quite substantial low-energy asymmetry in favour of p_z within the available O($2p$) states. This finding may by linked to the spin-density-wave transition in ambient-pressure $La_3Ni_2O_7$, which from experiment seems majorly connected to Ni-d_{z^2} involvement[38]. The p-asymmetry qualitatively still holds at high pressure right at the Fermi level, but is shifted to the higher unoccupied region above ε_F (see Figs. 4e,f). For the 1313 structural motif at high pressure, the p_z orbital of the BR oxygen ion connecting inner at outer layer of the trilayer segment is more low-energy dominant than the one connecting the trilayer with the monolayer segment. When integrating over all symmetry-inequivalent O($2p$) degrees of freedom in Figs. 4(g-i), the discussed asymmetries are mostly evened out, albeit an enhanced p_z low-energy weight is still easily observable for ambient-pressure 2222. A minor slight increasement of the total hole content on O($2p$) with pressure may be additionally read off from the data.

6 Concluding Remarks

A detailed inspection of O($2p$) degrees of freedom proves useful and necessary for superconducting nickelates, because of their intriguing placement between strong Mott-Hubbard and strong charge-transfer character. Mild orbital asymmetry and lack of significant further oxygen holes upon additional doping are revealed for the infinite-layer systems. On the other hand, a pronounced low-energy p-orbital asymmetry towards the briding p_z orbital is encountered in the $La_3Ni_2O_7$ system. While some of these observations sound rather subtle, they still may have relevant impact on the low-energy superconductivity phenomenon in these materials. Further work in weighing the relevance of the various site- and orbital sectors in these challenging materials is required.

Acknowledgements

The author thanks N. Poccia, I. M. Eremin and S. Bötzel for helpful discussions. Computations were performed at the Ruhr-University Bochum and the JUWELS Cluster of the Jülich Supercomputing Centre (JSC) under project miqs.

References

1. D. Li, K. Lee, B. Y. Wang, M. Osada, S. Crossley, H. R. Lee, Y. Cui, Y. Hikita, and H. Y. Hwang, *Superconductivity in an infinite-layer nickelate*, Nature, **572**, 624, 2019.
2. G. A. Pan, D. F. Segedin, H. LaBollita, Q. Song, E. M. Nica, B. H. Goodge, A. T. Pierce, S. Doyle, S. Novakov, D. C. Carrizales, A. T. N'Diaye, P. Shafer, H. Paik, J. T. Heron, J. A. Mason, A. Yacoby, L. F. Kourkoutis, O. Erten, C. M. Brooks, A. S. Botana, and J. A. Mundy, *Superconductivity in a quintuple-layer square-planar nickelate*, Nature Materials, **21**, 160, 2021.
3. H. Sun, M. Huo, X. Hu, J. Li, Y. Han, L. Tang, Z. Mao, P. Yang, B. Wang, J. Cheng, D.-X. Yao, G.-M. Zhang, and M. Wang, *Signatures of superconductivity near 80 K in a nickelate under high pressure*, Nature, **621**, 493, 2023.
4. Y. Zhu, D Peng, E. Zhang, B Pan, X. Chen, L. Chen, H. Ren, F. Liu, Y. Hao, N. Li, Z. Xing, F. Lan, J. Han, J. Wang, D. Jia, H. Wo, Y. Gu, Y. Gu, L. Ji, W. Wang, H. Gou, Y. Shen, T. Ying, X. Chen, W. Yang, H. Cao, C. Zheng, Q. Zeng, J.-g. Guo, and J. Zhao, *Superconductivity in pressurized trilayer $La_4Ni_3O_{10-\delta}$ single crystals*, Nature, **631**, 531, 2024.
5. A. S. Botana, F. Bernardini, and A. Cano, *Nickelate Superconductors: An Ongoing Dialog between Theory and Experiments*, J. Exp. Theor. Phys., **132**, 618, 2021.
6. H. Chen, A. Hampel, J. Karp, F. Lechermann, and A. Millis, *Dynamical mean field studies of infinite layer nickelates: Physics results and methodological implications*, Front. Phys., **10**, 835942, 2022.
7. B. Y. Wang, K. Lee, and B. H. Goodge, *Experimental Progress in Superconducting Nickelates*, Annu. Rev. Condens. Matter Phys., **15**, 305, 2024.
8. M. Wang, H.-H. Wen, T. Wu, D.-X. Yao, and T. Xiang, *Normal and Superconducting Properties of $La_3Ni_2O_7$*, Chinese Phys. Lett., **41**, 077402, 2024.

9. G. Kotliar, S. Y. Savrasov, K. Haule, V. S. Oudovenko, O. Parcollet, and C. A. Marianetti, *Electronic structure calculations with dynamical mean-field theory*, Rev. Mod. Phys., **78**, 865-951, 2006.
10. J. Zaanen, G. A. Sawatzky, and J. W. Allen, *Band gaps and electronic structure of transition-metal compounds*, Phys. Rev. Lett., **55**, 418, 1985.
11. S. Biermann, F. Aryasetiawan, and A. Georges, *First-Principles Approach to the Electronic Structure of Strongly Correlated Systems: Combining the GW Approximation and Dynamical Mean-Field Theory*, Phys. Rev. Lett., **90**, 086402, 2003.
12. L. Boehnke, F. Nilsson, F. Aryasetiawan, and P. Werner, *When strong correlations become weak: Consistent merging of GW and DMFT*, Phys. Rev. B, **94**, 201106, 2016.
13. S. Choi, A. Kutepov, K. Haule, M. van Schilfgaarde, and G. Kotliar, *First-principles treatment of Mott insulators: linearized QSGW+DMFT approach*, npj Quantum Materials, **1**, 16001, 2016.
14. A. van Roekeghem, T. Ayral, J. M. Tomczak, M. Casula, N. Xu, H. Ding, M. Ferrero, O. Parcollet, H. Jiang, and S. Biermann, *Dynamical Correlations and Screened Exchange on the Experimental Bench: Spectral Properties of the Cobalt Pnictide* $\mathrm{BaCo_2As_2}$, Phys. Rev. Lett., **113**, 266403, 2014.
15. W. Körner and C. Elsässer, *First-principles density functional study of dopant elements at grain boundaries in ZnO*, Phys. Rev. B, **81**, 085324, 2010.
16. D. Grieger, C. Piefke, O. E. Peil, and F. Lechermann, *Approaching finite-temperature phase diagrams of strongly correlated materials: A case study for* V_2O_3, Phys. Rev. B, **86**, 155121, 2012.
17. F. Lechermann, W. Körner, D. F. Urban, and C. Elsässer, *Interplay of charge-transfer and Mott-Hubbard physics approached by an efficient combination of self-interaction correction and dynamical mean-field theory*, Phys. Rev. B, **100**, 115125, 2019.
18. C. Elsasser, N. Takeuchi, K. M. Ho, C. T. Chan, P. Braun, and M. Fahnle, *Relativistic effects on ground state properties of 4d and 5d transition metals*, Journal of Physics: Condensed Matter, **2**, no. 19, 4371-4394, May 1990.
19. F. Lechermann, F. Welsch, C. Elsässer, C. Ederer, M. Fähnle, J. M. Sanchez, and B. Meyer, *Density-functional study of* Fe_3Al*: LSDA versus GGA*, Phys. Rev. B, **65**, 132104, 2002.
20. B. Meyer, C. Elsässer, F. Lechermann, and M. Fähnle, *FORTRAN 90 Program for Mixed-Basis-Pseudopotential Calculations for Crystals*, Max-Planck-Institut für Metallforschung, Stuttgart, 1998.
21. F. Lechermann, *Late transition metal oxides with infinite-layer structure: Nickelates versus cuprates*, Phys. Rev. B, **101**, 081110, 2020.
22. P. Werner, A. Comanac, L. de' Medici, M. Troyer, and A. J. Millis, *Continuous-Time Solver for Quantum Impurity Models*, Phys. Rev. Lett., **97**, 076405, 2006.
23. O. Parcollet, M. Ferrero, T. Ayral, H. Hafermann, I. Krivenko, L. Messio, and P. Seth, *TRIQS: A toolbox for research on interacting quantum systems*, Computer Physics Communications, **196**, 398-415, 2015.
24. P. Seth, I. Krivenko, M. Ferrero, and O. Parcollet, *TRIQS/CTHYB: A continuous-time quantum Monte Carlo hybridisation expansion solver for quantum impurity problems*, Comput. Phys. Commun., **200**, 274-284, 2016.

25. B. Amadon, F. Lechermann, A. Georges, F. Jollet, T. O. Wehling, and A. I. Lichtenstein, *Plane-wave based electronic structure calculations for correlated materials using dynamical mean-field theory and projected local orbitals*, Phys. Rev. B, **77**, 205112, 2008.
26. V. Bisogni, S. Catalano, R. J. Green, M. Gibert, R. Scherwitzl, Y. Huang, V. N. Strocov, P. Zubko, S. Balandeh, J.-M. Triscone, G. Sawatzky, and T. Schmitt, *Ground-state oxygen holes and the metal–insulator transition in the negative charge-transfer rare-earth nickelates*, Nature Communications, **7**, 13017, 2016.
27. F. Lechermann, *Doping-dependent character and possible magnetic ordering of $NdNiO_2$*, Phys. Rev. Mater., **5**, 044803, 2021.
28. F. Petocchi, V. Christiansson, F. Nilsson, F. Aryasetiawan, and P. Werner, *Normal State of $\mathrm{Nd}_{1-x}\mathrm{Sr}_x\mathrm{NiO}_2$ from Self-Consistent $GW + \mathrm{EDMFT}$*, Phys. Rev. X, **10**, 041047, 2020.
29. M. Kitatani, L. Si, O. Janson, R. Arita, Z. Zhong, and K. Held, *Nickelate superconductors-a renaissance of the one-band Hubbard model*, npj Quantum Mater., **5**, 59, 2020.
30. J. Karp, A. S. Botana, M. R. Norman, H. Park, M. Zingl, and A. Millis, *Many-Body Electronic Structure of $NdNiO_2$ and $CaCuO_2$*, Phys. Rev. X, **10**, 021061, 2020.
31. F. C. Zhang and T. M. Rice, *Effective Hamiltonian for the superconducting Cu oxides*, Phys. Rev. B, **37**, 3759, 1988.
32. B. H. Goodge, D. Li, K. Lee, M. Osada, B. Y. Wang, G. A. Sawatzky, H. Y. Hwang, and L. F. Kourkoutis, *Doping evolution of the Mott–Hubbard landscape in infinite-layer nickelates*, PNAS, **118**, e2007683118, 2021.
33. A. Bianconi, M. De Santis, A. Di Cicco, A. M. Flank, A. Fontaine, P. Lagarde, H. Katayama-Yoshida, A. Kotani, and A. Marcelli, *Symmetry of the $3d^9$ ligand hole induced by doping in $\mathrm{YBa}_2\mathrm{Cu}_3\mathrm{O}_{7-\delta}$*, Phys. Rev. B, **38**, 7196, 1988.
34. X. Chen, J. Zhang, A. S. Thind, S. Sharma, H. LaBollita, G. Peterson, H. Zheng, D. P. Phelan, A. S. Botana, R. F. Klie, and J. F. Mitchell, *Polymorphism in the Ruddlesden–Popper Nickelate $La_3Ni_2O_7$: Discovery of a Hidden Phase with Distinctive Layer Stacking*, J. Am. Chem. Soc., **146**, no. 6, 3640, 2024.
35. P. Puphal, P. Reiss, N. Enderlein, Y.-M. Wu, G. Khaliullin, V. Sundaramurthy, T. Priessnitz, M. Knauft, A. Suthar, L. Richter, M. Isobe, P. A. van Aken, H. Takagi, B. Keimer, Y. E. Suyolcu, B. Wehinger, P. Hansmann, and M. Hepting, *Unconventional Crystal Structure of the High-Pressure Superconductor $\mathrm{La}_3\mathrm{Ni}_2\mathrm{O}_7$*, Phys. Rev. Lett., **133**, 146002, 2024.
36. H. Wang, L. Chen, A. Rutherford, H. Zhou, and W. Xie, *Long-Range Structural Order in a Hidden Phase of Ruddlesden–Popper Bilayer Nickelate $La_3Ni_2O_7$*, Inorganic Chemistry, **63**, no. 11, 5020–5026, 2024.
37. Z. Dong, M. Huo, J. Li, J. Li, P. Li, H. Sun, L. Gu, Y. Lu, M. Wang, Y. Wang, and Z. Chen, *Visualization of oxygen vacancies and self-doped ligand holes in $La_3Ni_2O_{7-\delta}$*, Nature, **630**, 847, 2024.
38. X. Chen, J. Choi, Z. Jiang, J. Mei, K. Jiang, J. Li, S. Agrestini, M. Garcia-Fernandez, X. Huang, H. Sun, D. Shen, M. Wang, J. Hu, Y. Lu, K.-J. Zhou, and D. Feng, *Electronic and magnetic excitations in $La_3Ni_2O_7$*, Nature Communications, **15**, 9597, 2024.

Condensed Matter Theory

Condensed Matter Theory

Frithjof B. Anders

Condensed Matter Theory, Department of Physics, TU Dortmund University,
44221 Dortmund, Germany
E-mail: frithjof.anders@tu-dortmund.de

Although the fundamental equation for describing the dynamics of many-body solid state systems are known since the advent of quantum mechanics, an accurate numerical solution of the equation is impossible due to macroscopically large number of elementary degrees of freedom. Therefore, successful approaches are based on the identification of the relevant effective degrees of freedom and their interactions to understand the physical properties of a system. These effective low-energy theories, such as a Heisenberg or a Hubbard model, might appear simple in its structure, but the quantum mechanical entanglement can cause a rich variety of different competing macroscopic phases. Statistical mechanics is required in combination with adequate approximations to include thermal or quantum fluctuations of the problem for determining the correct state of matter.

The effective degrees of freedom are typically either localised spins in insulating materials or quasiparticles with bosonic or fermionic nature describing elementary excitations above the ground state. Collective phenomena such as superconductivity, superfluidity, or different types of magnetic order can emerge from the interactions between these effective degrees of freedom. It has been shown that such a low energy effective description can also be engineered in ultra-cold atomic gases: Optical lattices allow to realise model system and very precisely tailor their interactions such that collective phases in 1D and 2D lattice models can be studied by tuning its parameters. In designing artificial matter, one can also imagine shaping local superconducting islands that are connected by a Josephson tunnelling leading to an Josephson junction array.

Magnetic order comes in a rich variety of different phases such as ferro- and ferrimagnetisms, antiferromagnetism, altermagneticm, commensurable and incommensurable spin-density wave phases as typical example of itinerant magnetism. Much debated are frustrated lattices where competing interactions can suppress the magnetic order and lead to a so-called disordered spin liquid phase. Burkard *et al.* start from localised spins as elementary degrees of freedom. Spins have a simple algebra but they fulfil neither fermionic or boson commutation relations. To make spin models accessible to a very successful field theoretical approach developed for fermions, the authors use a pseudo-majorana representation of each spin and apply the functional renormalisation group approach to the problem. Solving the spin dynamics by numerically integrating coupled differential flow equations for different system sizes allows access to temperature dependent phase diagram for antiferromagneticaly coupled spins on different lattice geometries. Since the approach is able handle even long-range dipolar XY models, the authors report on their simulation for a realisation of such a model by a 2D ultra-cold gas experiment with Rubidium atoms in highly excited Rydberg states localised in an optical lattice. The agreement between the experimental data and calculated values for the spin-spin correlation functions is impressive.

Tausendpfund *et al.* apply a variational ansatz to quantum many-body systems. Tensor network methods are highly successful for determining the ground state properties of complicated pseudo-1D systems. White's density matrix renormalisation group approach can be embedded in this type of approach where the quantum state of a many-body system is approximated by a decomposition in connected smaller tensors. Their tensor elements are typically optimise by minimising an energy function. These connections represent the appropriate entanglement structure in a given quantum many-body state. Such approaches are very versatile and successful in problems with local entanglement. Their limitations are the dimension of interconnecting bonds which can be viewed as a measure for entanglement. The big advantage of such a decomposition is the numerically rather effective evaluation of expectation values and the optimisation of the many-body state by local matrix multiplications and matrix decomposition for which highly efficient numerically library algorithms have been developed in the last 50 year and tuned by the hardware manufactures. Tausendpfund *et al.* present results of the order parameter for a realisation of a tri-critical Ising model in a Josephson junction array as well as addressing bosonic fractional Hall states in Bose-Hubbard model subject to a gauge field. Although tensor network approaches are originally developed in the context of quantum many-body physics, it has growing into different areas such as quantum chemistry, plasma physics and is currently also branching out into the machine learning.

While two of the presented papers are concerned with such fundamental questions of many-body systems, the contribution by Willsch *et al.* focuses on the simulation of Shor's factoring algorithm for quantum computers on a classical supercomputer using a spin representation for each qubit. Present implementations of digital quantum computer suffer from dephasing by environmental noise. Therefore, the term noisy intermediate-scale quantum (NISQ) era was coined for the present era where the fidelity of implemented quantum algorithms on digital quantum computers suffers from the environmental noise. The authors report on the influence of such noise induced errors on the quality of Shor's algorithm for 20 and 30 qubits. While in a digital quantum computer qubits are individually controllable, in analog quantum computers such as quantum annealers the coupled quantum system undergoes a time-dependent evolution. Such systems are designed for efficiently solving optimisation problems. In the second part of their paper, Willsch *et al.* employed a D-Wave Advantage QPU for their three different factorisation methods and present a comparison.

Progress in condensed matter theory has been achieved in these three rather different fields by combining development of novel and very efficient algorithms in combination with the performance of top class supercomputers as provided by the Gauss Centre for Supercomputing (GCS)/John von Neumann Institute for Computing (NIC). One of the papers even employed an analog quantum computer from D-Wave at the JSC to complement their study of Shor's factoring algorithm on a simulated digital quantum computer implemented on the JUWELS Booster.

Pseudo Majorana Functional Renormalisation Group for Quantum Spin Systems

Ruben Burkard[1], Anna Fancelli[2,3], Matías Gonzalez[2,3], Nils Niggemann[2,3], Vincent Noculak[2,3], Johannes Reuther[2,3], Björn Sbierski[1], Yannik Schaden[2,3], and Benedikt Schneider[4]

[1] Institut für Theoretische Physik, Universität Tübingen, Auf der Morgenstelle 14, 72076 Tübingen, Germany

[2] Freie Universität Berlin, Physics Department, Arnimallee 14, 14195 Berlin, Germany
E-mail: johannes.reuther@fu-berlin.de

[3] Helmholtz-Zentrum Berlin für Materialien und Energie, Hahn-Meitner-Platz 1, 14109 Berlin, Germany

[4] Ludwig-Maximilians-Universität München, Theresienstr. 37, 80333 München, Germany

Frustrated quantum spin systems are well known for their rich phenomenology including exotic magnetic orders or quantum spin liquid behaviour. On the other hand, theoretical approaches to these systems in terms of numerical simulations are notoriously difficult even if modern large-scale computational resources are applied. The combination of many interacting spins on high-dimensional frustrated lattices renders many standard techniques, like exact diagonalisation, tensor-networks or quantum Monte-Carlo inapplicable. The recently developed pseudo-Majorana functional renormalisation group (PMFRG) represents a finite-temperature extension of the established pseudofermion (PF)FRG method, sharing its strengths through its unbiased field theoretical formulation. Motivated by promising benchmark calculations in an initial publication, we show here different applications of PMFRG to systems that are of current interest to the scientific community. To extend our method's applicability down to lower temperatures we further investigate the effects of two-loop corrections.

1 Introduction

Developing numerical methods to approximately solve the quantum mechanical many-body problem is of central importance in modern condensed matter theory. Spin systems such as Heisenberg models realised in strongly correlated quantum materials play a particularly prominent role in these endeavours as they harbour some of the most fascinating phases currently known. An outstanding example is the quantum spin liquid[1] which is characterised by the absence of broken symmetries or long-range magnetic order but features long-range entanglement, fractional quasiparticle excitations and topological order. Besides being of interest in fundamental research, exotic states of quantum matter also promise technological applications through a better understanding of high-temperature superconductivity, or through the exploitation of topological protection in future quantum information applications[2].

Unravelling the ground state or low-temperature properties of a generic quantum spin model using numerical techniques, however, is notoriously challenging. A variety of complementary approaches exists, such as exact diagonalisation, tensor network methods, quantum Monte Carlo and coupled cluster methods. However, each of them comes along with particular difficulties and limitations in system size, dimensionality, entanglement or

types of interactions and one may still conclude that the field suffers from a significant lack of techniques. Therefore, it is of paramount importance to develop and apply new and promising methods.

More than a decade ago, the pseudofermion functional renormalisation group (PFFRG) method has been introduced as a novel technique to calculate ground-state properties of quantum spin systems[3]. It takes a very different starting point than the aforementioned techniques as it maps spin operators onto auxiliary fermions. This strongly interacting fermionic problem is then treated within the framework of functional renormalisation[4]. Compared to other methods, its flexibility is a clear advantage since it allows the treatment of arbitrary models with two-body spin interactions and unfolds its full strength in higher dimensions (particularly for three dimensional systems) where other techniques are severely challenged. On the other hand, the PFFRG is plagued with some conceptual limitations, for example, the mapping onto auxiliary fermions introduces unphysical states which represent a potential source of errors.

Recently, a new path to functional renormalisation for spin systems has been introduced[5,6]: Instead of (complex) fermionic auxiliary particles, a spin representation in terms of Majorana (real) fermions is employed[7]. This modification resolves the problem of unphysical Hilbert space sectors. The resulting method – dubbed pseudo-Majorana functional renormalisation group (PMFRG) in one-[5] and two-loop implementation[6] - has shown large potential for future applications: It combines the strengths of PFFRG, i.e. flexibility, applicability to complex frustrated, three dimensional, and long-range coupled systems, and at the same time leverages these features to finite temperatures. Moreover, finite temperature also offers rigorous control of the truncations necessarily involved in the numerical solution of the FRG flow equations. In summary, the PMFRG allows for the computation of experimentally relevant temperature-dependent observables such as susceptibilities or the specific heat that were previously inaccessible with the PFFRG.

In a first publication[5], we have already demonstrated a proof of concept for the applicability of the PMFRG to simple spin systems. In a second work[6], a study of more complex and three-dimensional frustrated and unfrustrated spin systems indicates a surprisingly good quantitative agreement with other numerical methods like quantum Monte Carlo, for example in the context of resolving critical temperatures and scaling behaviour. In this work, we will give an overview of several different examples in which the PMFRG method can be applied to obtain valuable information from different spin systems.

2 Method

The PMFRG method developed in Ref. 5 rests on the SO(3) pseudo-Majorana representation of spin-1/2 operators[7],

$$S_i^x = -i\eta_i^y\eta_i^z, \quad S_i^y = -i\eta_i^z\eta_i^x, \quad S_i^z = -i\eta_i^x\eta_i^y. \tag{1}$$

Per spin $\mathbf{S}_i$, three Majorana fermions η_i^α with $\alpha = x, y, z$ are used, which increases the local Hilbert space dimension by a factor $\sqrt{2}$ as compared to the physical spin-$1/2$ case. In contrast to the complex fermion representation employed by the PFFRG, this Majorana representation avoids the appearance of unphysical energies in the spectrum of the fermionic model. The enlargement of the Hilbert space merely causes a degeneracy of spin eigenstates.

This enlargement can be trivially accounted for in the calculation of the physical spin partition function, for correlators it has no effect at all. Since in the Majorana representation unphysical states are absent in the entire spectrum, finite temperatures can be treated. Using the representation in Eq. 1, a Heisenberg Hamiltonian,

$$H = \sum_{(i,j)} J_{i,j} \sum_{\alpha=x,y,z} S_i^\alpha S_j^\alpha, \tag{2}$$

is converted to an interacting Hamiltonian of (pseudo-)Majoranas. As a necessary preparation for the subsequent treatment using functional renormalisation, one-loop flow equations for generic interacting Majorana Hamiltonians have been derived[5]. The symmetries for the specific pseudo-Majorana representation of a Heisenberg model restrict the irreducible four-point vertices to only a few distinct types and the local $\mathbb{Z}_2$ gauge symmetry of the representation in Eq. 1 ensures a bilocal structure in real space. We solve the flow equations for finite temperature using a Matsubara frequency cutoff in the bare propagator. At the end of the flow, physical observables like the static spin structure factor and - via its own flow equation - the free energy are obtained. The first quantity allows us to compare with neutron scattering experiments but susceptibility or the correlation length can further be used to determine finite-temperature phase boundaries below which magnetic order sets in. On the other hand, access to the free energy allows for a computation of several other experimentally relevant thermodynamic quantities like heat capacity.

However, this improvement from the PFFRG to the PMFRG came at a considerable numerical cost: The zero-temperature PFFRG approach typically requires only a single run. To detect magnetic order within that method, one commonly analyses features in the flow of the magnetic susceptibility as a function of the infrared cutoff Λ. On the other hand, to make use of the unbiased nature of finite size scaling in PMFRG, it is necessary to run many independent simulations at a dense grid of temperatures around the expected critical temperature T_c, each for several system sizes L.

To combine the advantages of the zero-temperature and the finite-temperature approaches, a key observation is that the infrared cutoff Λ, an artificial parameter that sup-

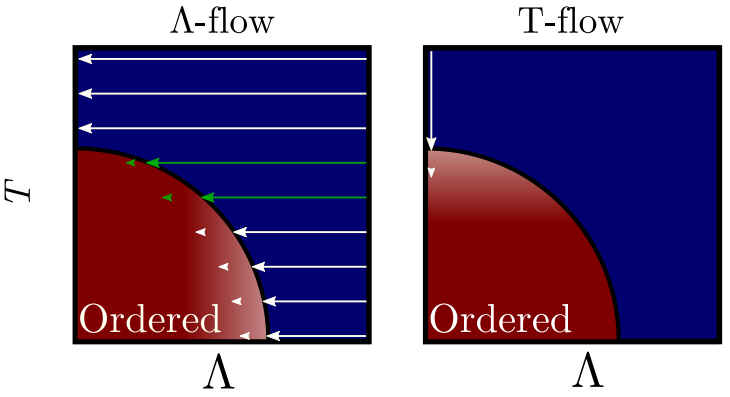


Figure 1. Schematic picture of the difference between PMFRG in Λ-flow (left) and T-flow (right) scheme. Green arrows indicate FRG runs which need to cross through the phase boundary to resolve magnetic order. Reproduced from Ref. 8.

presses fermionic propagation at (Matsubara) frequencies, has an effect that is similar to a finite temperature. In Ref. 8, we built on this insight and proposed the temperature flow scheme for PMFRG, previously used for the fermionic Hubbard model in Ref. 9. The key differences between the two flow schemes are summarised in Fig. 1: Both the physical temperature T and Λ serve to suppress magnetic ordering. The main advantage of the temperature flow is that one is typically only interested in the results at $\Lambda = 0$, so that only one FRG run needs to be calculated. Moreover, the T-flow does not require to flow through a critical region near a phase boundary to detect magnetic order. This allows us to better resolve magnetic order, in particular, at much lower critical temperatures than possible with the Λ-flow.

3 Applications

3.1 Finite-Temperature Phase Transition and Finite-Size Scaling in the Cubic Lattice Antiferromagnet

As anticipated, the PMFRG method can be used to detect magnetic long-range order. This can be illustrated in a simple and well-known system, such as the antiferromagnetic nearest neighbour Heisenberg model on the simple cubic lattice. Despite the simplicity of the lattice, not many methods are capable of finding an accurate value for the critical temperature that separates the paramagnetic and Néel ordered regimes. However, since the lattice is unfrustrated, the quantum Monte Carlo method can be applied, giving us a solid testing ground for the PMFRG method.

More specifically, to determine the critical temperature T_c the finite-size scaling behaviour of the correlation length ξ is investigated. Here, ξ is determined from the susceptibility $\chi(\mathbf{k}_N)$ at the ordering wave vector $\mathbf{k}_N$ via the so-called correlation ratio

$$\xi/L = \frac{1}{2\pi}\sqrt{\frac{\chi(\mathbf{k}_N)}{\chi(\mathbf{k}_N + \frac{2\pi}{L}\mathbf{e}_x)} - 1}. \tag{3}$$

As shown in Fig. 2(a) when plotting ξ/L over the temperature T, the curves for different system sizes L cross at a temperature $T_c = 0.92J$. This determines the critical temperature

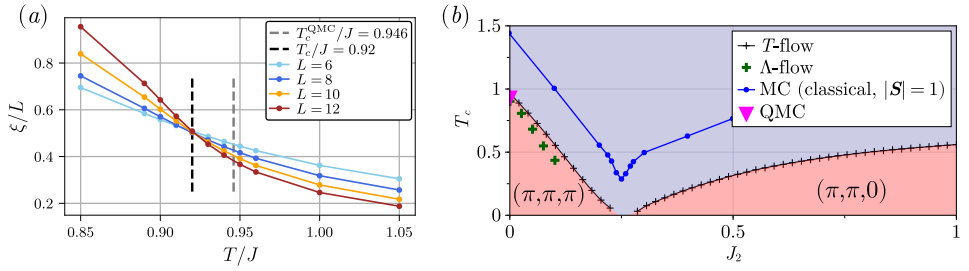


Figure 2. (a) PMFRG results for the correlation ratio ξ/L as a function of temperature for different system sizes L of the antiferromagnetic nearest neighbour Heisenberg model on the simple cubic lattice. The dashed black line highlights the crossing point which is the PMFRG estimate of the critical temperature while the grey dashed line corresponds to the quantum Monte Carlo result[10]. Figure reproduced from Ref. 11. (b) Phase diagram of the J_1-J_2 Heisenberg cubic antiferromagnet. Reproduced from Ref. 8.

to magnetic Néel order which is in very good agreement with the quantum Monte Carlo result $T_c = 0.946(1)$[10].

Adding the next-nearest neighbour coupling J_2 frustrates the Néel order, leading to a suppression of the critical temperature. In this case, the quantum Monte Carlo method is no longer applicable due to the sign problem. However, PMFRG can still perform without any extra computational effort. In Fig. 2(b), we show the J_2-T phase diagram of the antiferromagnetic Heisenberg model on the simple cubic lattice with first (J_1) and second (J_2) neighbour interactions. As stated before, our result for the critical temperature at $J_2 = 0$ is in excellent agreement with the value obtained by quantum Monte Carlo. But now we can extend these calculations to finite values of J_2 which suppress the critical temperature and might even give room to a small region with a non-magnetic ground state.

3.2 Quantum Paramagnetism in the Decorated Square-Kagome Antiferromagnet $Na_6Cu_7BiO_4(PO_4)_4Cl_3$

As the search for candidate materials for quantum spin liquids continues, numerical investigations of experimentally available compounds are of paramount importance. The recent synthesis of $Na_6Cu_7BiO_4(PO_4)_4Cl_3$ and its experimentally observed absence of order down to the lowest temperatures[12] constitutes an ideal scenario to apply the PMFRG method. The compound realises a highly frustrated square kagome lattice, on which we can perform PMFRG[13]. The microscopic couplings were obtained via DFT energy mapping[14] (see Fig. 3 a). The so-obtained couplings form a set of decoupled 2D square kagome layers (also referred to as *shuriken* lattice) with added sites Cu(3) at the centre of each shuriken (see Fig. 3 b). This gives place to a very complicated Heisenberg Hamiltonian with five different types of exchange couplings.

Our results are shown on panel (c) of Fig. 3, where we plot a comparison of different numerical approaches for the scaling of the dominant peak in the spin structure factor. While the methods differ in the implementation of lattices (unlike other methods, PMFRG

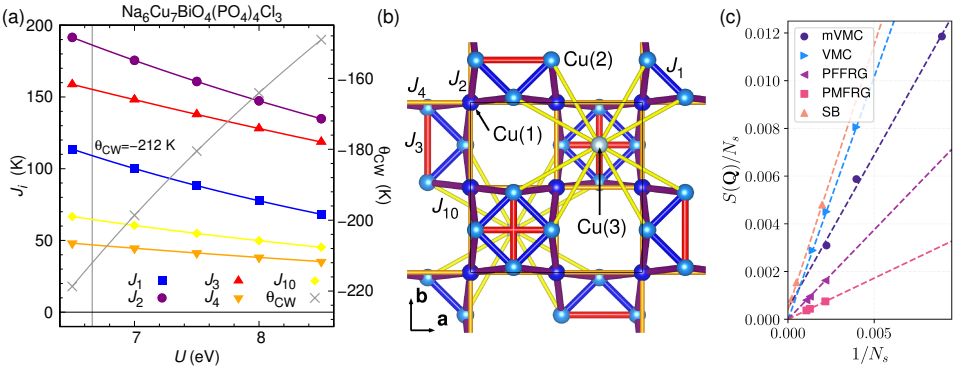


Figure 3. (a): Spin couplings of $Na_6Cu_7BiO_4(PO_4)_4Cl_3$ obtained via DFT energy mapping. The vertical line indicates the coupling strength through equivalence with the experimental Curie-Weiss temperature Θ_{CW}. (b) Lattice structure and non-negligible couplings of the compound viewed from the z-direction, resulting in the effective realisation of a square kagome lattice. (c): Finite-size scaling of the structure factor at the dominant wavevector $\mathcal{S}(\mathbf{Q})/N_s$ divided by the number of sites for different methods. Figure reproduced from Ref. 14.

and PFFRG do not implement finite-size clusters but rather fully translationally invariant lattices with restricted range of correlations), all approaches agree on an extrapolation of vanishing $\mathcal{S}(\mathbf{Q})/N_s \to 0$. This indicates the absence of magnetic order in agreement with the experimental observation.

While the system lacks magnetic order, or results identify valence-bond order, in which neighbouring sites form singlets that break the translational symmetries of the lattice but maintain the global spin rotation symmetry. To aid comparison with future neutron scattering experiments, predictions for the spin structure factor as well as its powder average based on the real, three-dimensional crystallographic structure of the compound $Na_6Cu_7BiO_4(PO_4)_4Cl_3$ were made using PMFRG[14].

3.3 Analog Quantum Simulation: Magnetism in the 2d Dipolar XY Model

The basic idea in the field of analog quantum simulation is to use synthetic and well controlled quantum systems as models to simulate the essential behaviour of other quantum systems which are out of reach for direct numerical computation on classical (super-) computers. Often, the synthetic quantum system is realised with cold atoms as elementary building blocks which offer an exquisite level of control and measurement opportunities.

Certainly, frustrated quantum spin Hamiltonians are one of the main targets for quantum simulations. In a recent experiment, the quantum spins are represented by a pair of highly excited ("Rydberg") states of Rubidium atoms[15]. With optical tweezers, these atoms were arranged on a 10×10 square-lattice array and the interactions between atoms give rise to an effective long-range dipolar XY model ($J_{ij} \propto \pm 1/|\mathbf{r}_i - \mathbf{r}_j|^3$) where both ferromagnetic (FM) and antiferromagnetic (AFM) exchange was experimentally realised. The experimental observables are site-resolved equal-time spin-spin correlation functions as shown by the dots in Fig. 4. While the FM case was simulated with quantum Monte Carlo

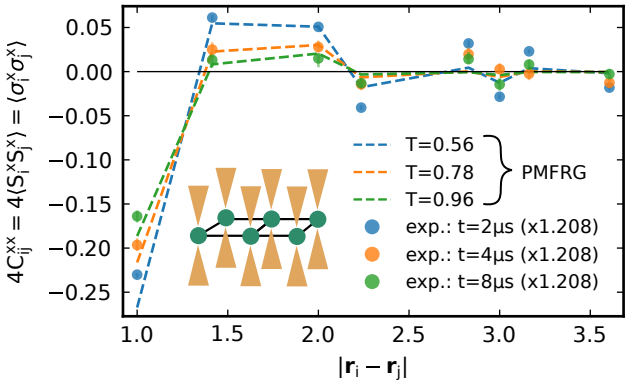


Figure 4. Experimental spin correlation function of an AFM-dipolar XY model realised in a Rydberg atom array shown schematically in the inset. The data was measured after $t = 2\mu s$ (blue dots), $t = 4\mu s$ (orange dots) and $t = 8\mu s$ (green dots) and is taken from Ref. 15. The measured data was multiplied with a factor of 1.208 to take into account measurement errors and temperature is given in units of the nearest-neighbour coupling. The data for small distances are well reproduced by thermal PMFRG simulations (dashed line, infinite system) at temperatures shown in the labels. Figure reproduced from Ref. 16.

techniques, the PMFRG can be used to study the correlation functions expected from the AFM Hamiltonian.

To numerically study the Rydberg atom array, it was necessary to generalise the PMFRG method beyond $SU(2)$ Heisenberg models to the XXZ case with a $U(1)$ symmetry. Another generalisation within this project was the simulation of frustrated long-range (power-law) interactions which are rarely relevant in solid-state quantum magnets but ubiquitous in atomic physics. Our results in Ref. 16 were used to provide thermometry for the experiment, i.e. to find a temperature that reproduces the measured correlation profile. Importantly, this assumes that the system had time to equilibrate to a thermal state, an assumption far from trivial in quantum optical setups. As shown in Fig. 4 by the dashed lines, this was indeed successful and provided a much better agreement to the data than zero-temperature density-matrix renormalisation group results offered in the original experimental article[15]. Even the progressive self-heating of the Rydberg array under spontaneous atomic decay events can be clearly followed from the extracted temperatures.

4 Concluding Remarks

We have shown that the PMFRG method is a powerful tool for studying a wide variety of different spin systems even in the case of very complex microscopic settings of spin interactions. The approach allows the calculation of critical temperatures, phase transitions and spin correlation functions in models where other methods fail, providing access to novel phase diagrams with paramagnetic regions that could be candidates for quantum spin liquids or other exotic phases. The method also allows direct comparison with experiments on complex compounds or analog quantum simulations in optical lattices, opening up a wide range of applications.

In this article, we have presented three different applications that illustrate the benefits of PMFRG. Specifically, we discussed applications to frustrated and unfrustrated three-dimensional lattices of Heisenberg spins, complex two-dimensional lattices with many symmetry inequivalent spins and interactions as realised in the recently synthesised quantum magnet $Na_6Cu_7BiO_4(PO_4)_4Cl_3$, and dipolar systems relevant to Rydberg atom arrays in which the interactions are long-ranged.

We expect that new developments and improvements will continue to expand the horizons of the PMFRG method in the future. For example, our currently most advanced PMFRG code can treat anisotropic spin interactions with $U(1)$ symmetry and magnetic fields pointing along the symmetry axis. However, the case of general anisotropic two-body spin interactions without continuous spin rotation symmetry has not yet been implemented. The latter generalisation, which is expected within the next few years, will further increase the applicability and flexibility of the PMFRG.

Acknowledgements

The authors would like to acknowledge helpful discussions with Frederic Bippus, Yasir Iqbal, Dominik Kiese, Tobias Müller, Ronny Thomale, and Simon Trebst. We would also like to gratefully acknowledge usage of the JUWELS cluster at the Forschungszentrum Jülich for the computer time provided, without which the development, testing and application of the PMFRG method would not have been possible.

References

1. L. Savary and L. Balents, *Quantum spin liquids: a review*, Rep. Prog. Phys. **80**, 016502, 2017.
2. A. Yu. Kitaev, *Fault-Tolerant Quantum Computation by Anyons*, Annals of Physics **303**, 2-30, 2003.
3. J. Reuther and P. Wölfle, *J_1-J_2 frustrated two-dimensional Heisenberg model: Random phase approximation and functional renormalization group*, Phys. Rev. B **81**, 144410, 2010.
4. C. Wetterich, *Exact evolution equation for the effective potential*, Phys. Lett. B **301**, 1, 1993.
5. N. Niggemann, B. Sbierski, and J. Reuther, *Frustrated quantum spins at finite temperature: Pseudo-Majorana functional renormalization group approach*, Phys. Rev. B **103**, 104431, 2021.
6. N. Niggemann, J. Reuther, and B. Sbierski, *Quantitative functional renormalization for three-dimensional quantum Heisenberg models*, SciPost Phys. **12**, 156, 2022.
7. A. M. Tsvelik, *New fermionic description of quantum spin liquid state*, Phys. Rev. Lett. **69**, 2142, 1992.
8. B. Schneider, J. Reuther, M. G. Gonzalez, B. Sbierski, and N. Niggemann, *Temperature flow in pseudo-Majorana functional renormalization for quantum spins*, Phys. Rev. B **109**, 195109, 2024.
9. C. Honerkamp and M. Salmhofer, *Temperature-flow renormalization group and the competition between superconductivity and ferromagnetism*, Phys. Rev. B **64**, 184516, 2001.
10. A. W. Sandvik, *Critical Temperature and the Transition from Quantum to Classical Order Parameter Fluctuations in the Three-Dimensional Heisenberg Antiferromagnet*, Phys. Rev. Lett. **80**, 5196, 1998.
11. T. Müller, D. Kiese, N. Niggemann, B. Sbierski, J. Reuther, S. Trebst, R. Thomale, and Y. Iqbal, *Pseudo-fermion functional renormalization group for spin models*, Rep. Prog. Phys. **87**, 036501, 2024.
12. M. Fujihala, K. Morita, R. Mole et al., *Gapless spin liquid in a square-kagome lattice antiferromagnet*, Nat. Commun. **11**, 3429, 2020.
13. N. Astrakhantsev, F. Ferrari, N. Niggemann, T. Müller, A. Chauhan, A. Kshetrimayum, P. Ghosh, N. Regnault, R. Thomale, J. Reuther, T. Neupert, and Y. Iqbal, *Pinwheel valence bond crystal ground state of the spin-$\frac{1}{2}$ Heisenberg antiferromagnet on the shuriken lattice*, Phys. Rev. B **104**, L220408, 2021.
14. N. Niggemann, N. Astrakhantsev, A. Ralko, F. Ferrari, A. Maity, T. Müller, J. Richter, R. Thomale, T. Neupert, J. Reuther, Y. Iqbal, and H. O. Jeschke, *Quantum paramagnetism in the decorated square-kagome antiferromagnet $Na_6Cu_7BiO_4(PO_4)_4Cl_3$*, Phys. Rev. B **108**, L241117, 2023.
15. C. Chen, G. Bornet, M. Bintz et al., *Continuous symmetry breaking in a two-dimensional Rydberg array*, Nature **616**, 691-695, 2023.
16. B. Sbierski, M. Bintz, S. Chatterjee, M. Schuler, N. Y. Yao, and L. Pollet, *Magnetism in the two-dimensional dipolar XY model*, Phys. Rev. B **109**, 144411, 2024.

Variational Tensor Network Methods for Quantum Many-Body Systems

Niklas Tausendpfund[1], Erik Weerda[2], and Matteo Rizzi[1,2]

[1] Institute of Quantum Control (PGI-8), Forschungszentrum Jülich, 52425 Jülich, Germany
E-mail: {n.tausendpfund, m.rizzi}@fz-juelich.de

[2] Institute for Theoretical Physics, University of Cologne, 50937 Köln, Germany
E-mail: weerda@thp.uni-koeln.de

We briefly revise quantum-information inspired tensor network numerical methods for infinite-size quantum systems in spatial dimension one and two. We discuss two concrete applications to quantum simulator platforms, namely quasi-1D ladders of Josephson Junction Arrays and cold bosonic atoms in 2D optical lattices with synthetic magnetic flux. We describe our findings about a tri-critical Ising conformal field theory in the first, and about fractional quantum Hall bosonic states in the latter. We conclude with an outlook on future developments, both technical and topical, since tensor networks are rapidly extending their field of use outside of quantum many-body physics.

1 Introduction

In the last three decades, Tensor Networks (TN) have emerged as powerful theoretical and numerical versatile tools to simulate complex quantum systems on a classical computer[1,2]. From a numerical point of view, indeed, the simulation of quantum matter constitutes a formidable challenge: the quantum wave-function is an element of a Hilbert space whose dimension grows exponentially with the number of system constituents. In a nutshell, TN prune down the description of many-body systems to polynomially many coefficients by making profit of quantum information insights on the entanglement structure of the typical wavefunctions of interest. As such, they are natural candidates to guide and benchmark the development of quantum technologies (QT) and to assess quantum supremacy[3,4], if any attainable. More recently, the applications of TN numerical tools have also spread over interdisciplinary areas, ranging from applied mathematics as solvers of complex optimisation problems, partial differential equations, high-energy theoretical physics, to AI as a valid alternative machine learning tool[5–7a].

Here, in Sec. 2, we offer a brief overview of a couple of TN Ansätze, in spatial dimension one and two, which work directly in the thermodynamic limit and are thus best suited to elucidate phase diagrams of many-body quantum systems and to validate implementation schemes of quantum simulators.

Quantum simulators are ground-breaking experimental platforms that owe their name to the possibility to experimentally program specific models of matter in- and out-of-equilibrium – similar to numerical simulations. Such platforms leverage on the ability to control and observe the individual quantum degrees of freedom (e.g., in neutral atoms, ions, superconducting circuits, lattice defects, etc.) and their interactions with the highest precision, and on the possibility of resolving their inherent time-scales. Their purpose follows Feynman's original vision to circumvent the *in*famous many-body curse of dimensionality

[a] Compared to deep neural networks, TN - as a completely different method – offer more transparency in complex AI applications due to their though high-dimensional but linear processing capabilities of big data.

by implementing the simulation itself in a quantum mechanical system. Ultimately, the goal is to develop quantum simulators that allow us to investigate increasingly complex models of quantum matter, to address pivotal problems in quantum many-body physics and quantum chemistry, e.g., high-temperature superconductivity or chemical reaction dynamics. The availability of early-stage quantum simulators, together with sophisticated numerical techniques that keep pushing the classical computational capabilities, puts us today in the exciting situation that both rely on each other for mutual certification while entering otherwise uncharted terrain.

Here, we review some recent results of ours concerning in Sec. 3.1 a scheme for quantum simulation of peculiar field theories (more precisely, tricritical Ising conformal field theory) in platforms of tunable Josephson junction arrays (JJAs), and in Sec. 3.2 a long-sought description of fractional quantum Hall (FQH) states via genuinely 2D tensor networks, that could also have practical implications for quantum simulation schemes with cold atoms in optical lattices.

2 Tensor Networks for Translation Invariant Systems

Tensor Networks (TN) consist of a convenient decomposition of the system wave-function into smaller tensors, whose interconnection bonds carry the entanglement structure at the heart of quantum effects. Such a rewriting brings along an increased efficiency (from exponential to polynomial, typically) in dealing with ground-state searches and extraction of many relevant quantities (e.g., local observables, long-range and edge-to-edge correlations, entanglement spectra, etc.). Noticeably, TN methods also give direct access to the entanglement spectrum[8,9], a quantity whose importance in theoretical condensed matter has exploded in the last decade, and which has recently become experimentally measurable, at least in some cases[10].

Another property of tensor networks is that they are particularly suitable to treat physical systems directly in the thermodynamic limit, which is advantageous in many respects. Typically, indeed, phases of a model are strictly defined only in the thermodynamic limit and finite-size effects can lead to shadowing of the true ground state, especially if qualitatively different states are closely competing. A numerical method directly operating in such limit overcomes these problems by neglecting (at first) the boundary conditions, and assuming the ground state to be translational invariant. In the case of tensor networks, this allows one to approximate the ground-state of such a translational invariant Hamiltonian by a small unit-cell of tensors (1D or 2D) periodically repeated in any spacial direction of the system. Applying this idea to one dimensional systems, a particularly powerful declination of these ansatz states is offered by Variational Uniform Matrix Product States (**VUMPS**)[11][b], while the two-dimensional corresponding network is dubbed Infinite Projected Entangled-Pair States (**iPEPS**)[14]. Against the naive intuition, the infinite ansatz allows to simplify calculations, as one can make use of power methods to obtain effective tensors approximating semi-infinite parts of the translation-invariant networks. The calculation of expectation values of observables and correlation functions is then reduced to the contraction of relatively small networks. Tensor network ansatz states are mainly limited in their expressive power by the dimension of the interconnecting bonds, called the bond dimension. Having a finite bond dimension introduces finite entanglement effects, such

[b] Other closely related versions are those of infinite DMRG[12] and infinite MPS[13].

as effective finite correlation lengths in critical one dimensional models. Therefore it is crucial to extrapolate observables to infinite bond dimension.

2.1 VUMPS and iPEPS: Tensor Network Structures

Variational Uniform Matrix Product States (**VUMPS**)[11] represents a class of one dimensional ansatz states describing one dimensional chain systems in the thermodynamic limit. To ensure the correct thermodynamic properties, the VUMPS is constructed in a way to enforce translation invariance of the state. This is achieved by periodically repeating the same unit-cell formed by the set of matrices $\{A_1^{\sigma_1}, A_2^{\sigma_2}, \dots\}$ along a one dimensional line, see Fig. 1. Here, σ_j enumerates the physical states at position j. Analogously to the structure above we can periodically repeat the same higher rank tensors in a two-dimensional structure, in what are commonly referred to as infinite projected entangled-pair states (**iPEPS**)[14]. This can be done with different arrangements of the tensors in a periodically repeating unit cell, as illustrated in Fig. 2. Note that every tensor-network ansatz has a

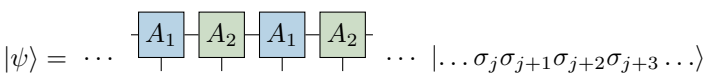


Figure 1. Tensor network representation of a VUMPS with a two-site unit-cell.

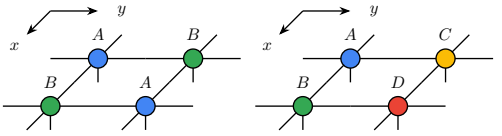


Figure 2. Example of different unit cell configurations for iPEPS. Taken from Ref. 15.

natural gauge degree of freedom: One can always insert a unity $\mathbb{1} = GG^{-1}$ on the virtual links, where G is an element of the general linear group. Absorbing G and G^{-1} into the neighbouring tensors does change the matrix elements and thus the representation of the physical state, but not the physical state itself, see Fig. 3 for a sketch. For a one dimensional model, one can actually profit from this gauge degree of freedom to define a representation rendering most of the calculations trivial[1], also known as the canonical gauge. For the iPEPS, such simplifications are not possible due to the presence of loops in the network.

The translation invariance of the infinite ansatz state allows for an efficient calculation of local observables. In both cases, VUMPS or iPEPS, this amounts to calculate so-called effective environments, representing the semi-infinite contracted tensor network. We will demonstrate this in the conceptually simpler case of a VUMPS with a unit cell consisting of a single tensor. For this, consider the expectation value of single site operator[c],

[c]Strictly speaking one has to calculate $\langle\hat{O}\rangle = \langle\psi|\hat{O}|\psi\rangle/\langle\psi|\psi\rangle$ for a general VUMPS state. However, here we assume $\langle\psi|\psi\rangle = 1$, which is always possible to achieve by a simple rescale of the tensors in the unit cell, see caption of Fig. 4.

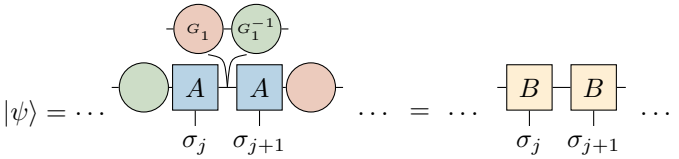


Figure 3. Example of using the gauge degree of freedom to transform one VUMPS representation $\{A^\sigma\}$ into new one $\{B^\sigma\}$ with $B^\sigma = GA^\sigma G^{-1}$ for each σ. Both, A^σ and B^σ, are describing the same physical state $|\psi\rangle$.

$\langle\hat{O}\rangle = \langle\psi|\hat{O}|\psi\rangle$. The diagrammatic representation for this contraction is shown in Fig. 4a). Left and right of the insertion of the local operator $\hat{O}$, there appear an infinite product of transfer-matrices formed by contracting the VUMPS tensor A^σ with its conjugate $\bar{A}^\sigma$ over the physical index σ. Because of this infinite product, one can simply replace the semi-infinite chains of transfer-matrices left and right of the local operator $\hat{O}$ by the dominant left and right eigenvector of this transfer-matrix, L and R. It follows that the expectation value collapses to a relatively small diagram one has to compute.

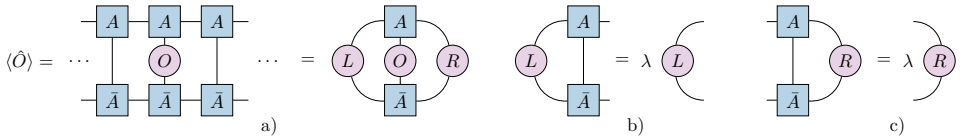


Figure 4. a) Diagram for the expectation value of a local observable with a normalised VUMPS $\langle\psi|\psi\rangle = 1$. The left and right environments are determined as the dominant left and right eigenvectors of the transfer-matrix, see b) and c). The normalisation condition is equivalent to $\lambda = 1$, which can always be achieved by a simple renormalisation of the VUMPS tensor $\tilde{A}^\sigma = 1/\sqrt{\lambda}A^\sigma$.

In the case of iPEPS, the transfer-matrix in any direction would contain already an infinite amount of tensors along the transverse direction. One thus further approximates the transfer-matrix itself. The procedure to obtain the effective environments is however similar to the VUMPS case, as one iteratively absorbs transfer-matrices in all directions into the boundary vectors. However, due to the additional approximation in the transfer-matrix itself, an additional renormalisation step is necessary. In Fig. 5 we sketch one possible way of obtaining effective environments, namely the so-called corner-transfer-matrix renormalisation group (CTMRG)[16][d]. Equipped with well-converged environments, one can now easily calculate local observables. Moreover, one can also reuse these environments to calculate any n-body operator expectation value (such as the Hamiltonian) and also correlation functions at a distance. The latter can be extracted by investigating the sub-dominant eigenpairs of the transfer-matrix. Scalar products (fidelities), or better said their intensive log-version, can be also extracted from mixed transfer matrices in a similar fashion.

[d]Another possibility is the usage of VUMPS themselves to approximate the boundaries strip-wise by a one dimensional tensor network[17].

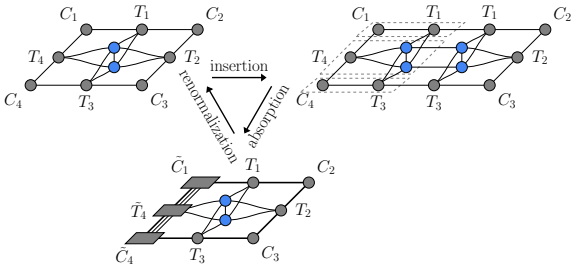


Figure 5. Illustration of the basic strategy of the CTMRG algorithm. Absorbtion of local tensors into the environment, increasing the environment-bond dimension. The bond dimension is then normalised to a fixed value. This procedure is iterated in all different direction until the environment tensors are converged. Taken from Ref. 15.

2.2 Variational Optimisation of Translation-Invariant Tensor Networks

Given a particular translationally invariant ansatz for our wavefunction, as described in the above section, it is crucial to find the lowest energy state within this ansatz class. The ground state search within the manifold of states defined by the translation-invariant tensor network ansatzes is fundamentally build on the variational principle:

$$E_{\mathrm{gs}} \leq \min_A \frac{\langle\psi(A)|\,\hat{H}\,|\psi(A)\rangle}{\langle\psi(A)|\psi(A)\rangle}, \tag{1}$$

where A represents the variational parameters (i.e., the tensor entries) of the ansatz in question. In the state-of-the-art algorithms (VUMPS (1d), gradient-based optimisation (2d)) we update all variational parameters of the ansatz simultaneously and are staying within the variational manifold during the optimisation procedure. This is part of the reason that these optimisation schemes have proven to be advantageous if compared to alternative approaches, like those based on imaginary-time evolution[18]. In particular, for frustrated systems, it has been shown that having the full gradient at disposal helps to avoid getting stuck in false extremal points. Note that the gradient that is used in the gradient-based optimisation in the iPEPS case can be obtained using automatic differentiation and fixed point methods - a set of tools already enjoying widespread in the machine learning community[15,19]. Although it is also possible to use AD-techniques together with a gradient based minimisation for the VUMPS ansatz, one can also exploit the gauge degree of freedom to reduce the problem to find the lowest state of an effective Hamiltonian[20].

3 Applications

3.1 Guiding the Realisation of the Tri-Critical Ising Model in Josephson-Junction Arrays Using VUMPS

As an illustration of the VUMPS technique, we revise here a recent work of ours[21], where we used this to corroborate the use of tunable Josepshon Junction arrays (JJAs) as quantum simulator platforms to realise the tri-critical Ising (TFI) conformal field theory (CFT). The TFI CFT is of special interest, as a certain class of the excitations share similar fusion

properties as the Fibonacci anyon, a possible platform to realise topological quantum computing[22]. Modern JJAs provide a perfect platform for realising interacting bosonic quantum field theories[23], as recent hybrid semi-/super-conducting devices allow engineering arrays with high tunability of their internal parameters[24].

The building blocks of our envisioned device are made up of two E-shaped superconducting island connected by three parallel junctions with tunable transparencies (T_1, T_2, T_1) and pierced by a magnetic flux Φ. Cooper pairs encircling the latter will pick up an additional geometric phase. The resulting potential for the phase difference between the two superconductors φ can be expanded in (leading) harmonics as

$$V_J(\varphi) = \mu_1 \cos(\varphi) + \mu_2 \cos(\varphi) + \mu_3 \cos(\varphi), \tag{2}$$

where the three coupling constants $\mu_j(\mathbf{X})$ are functions of the three free parameters, summarised via the coordinate $\mathbf{X} = (T_1 \cos(\Phi), T_1 \sin(\Phi), T_2)$. From a preliminary semi-classical analysis, the model has three potential phases, labelled as I, II and III. Both phases I and III are unordered phases with no local order parameter present: the unique minimum sits at $\varphi = 0$ or $\varphi = \pi$, respectively. When a transition between them takes place, it must be of first-order character. Phase II is instead an ordered phase characterised by a finite local single particle current $J_{\perp}^{(2e)} = \sin\left(\sqrt{2}\varphi\right) \neq 0$. Phase II and III are separated only by a second-order transition of the Ising universality class. On the other hand, the second-order (Ising) transition line separating phase II and I terminates into a first-order line. At exactly this termination point, we expect the TCI universality class to appear: it turns out that for all choices of the transparency T_2 there exists such a TCI point, $(T_1, \Phi)_c$.

By arranging many copies of the above building block in a 1D ladder geometry, we promote the phase difference φ of a single triple Josephson junction to a position dependent quantum field, $\hat{\varphi}(x)$. Including also intra- and inter-leg charging effects (due to capacitance), we arrive to the Hamiltonian

$$\begin{aligned}\widehat{H} = \sum_{j=0}^{L-1} \Bigg[\sum_{\alpha=\mathsf{a},\mathsf{b}} \Big(E_C \widehat{N}_{\alpha,j}^2 - E_J \cos\left(\hat{\varphi}_{\alpha,j+1} - \hat{\varphi}_{\alpha,j}\right) \Big) \\ + V_{\perp}\, \widehat{N}_{\mathsf{a},j} \widehat{N}_{\mathsf{b},j} + V_J\left(\hat{\varphi}_{\mathsf{a},j} - \hat{\varphi}_{\mathsf{b},j}\right) \Bigg],\end{aligned} \tag{3}$$

where $\hat{\varphi}_{\alpha,j}$ now represents the phase operator of the j-th island on the leg $\alpha \in \{\mathsf{a}, \mathsf{b}\}$. The charging effects are incorporated by the charge operator $\widehat{N}_{\alpha,j}$, canonically conjugated to the SC phases, $[\widehat{N}_{\alpha,j}, e^{i\hat{\varphi}_{\alpha,j}}] = -e^{i\hat{\varphi}_{\alpha,j}}$. It can be shown, by means of bosonization, that the low energy sector of the model 3 flows to the correct multi-frequency sine-Gordon model hosting a TCI CFT point[21]. However, this bosonization study also shows that the model of Eq. 3 possesses a second decoupled mode in the low energy spectrum which is always gapless. This gapless mode describes collective charge excitations, very similar to the charge-spin separation in interacting spinful fermionic chains in one dimensions[25].

For numerical simulations, we expressed the model in Eq. 3 in the charge basis. In this basis, the operator $\widehat{N}_{\alpha,j}$ assumes a diagonal form, i.e. $\widehat{N}_{\alpha,j} |n_{\alpha,j}\rangle = n_{\alpha,j} |n_{\alpha,j}\rangle$. The entries of $\widehat{N}_{\alpha,j}$ counts how much the number of Cooper pairs differs from the average

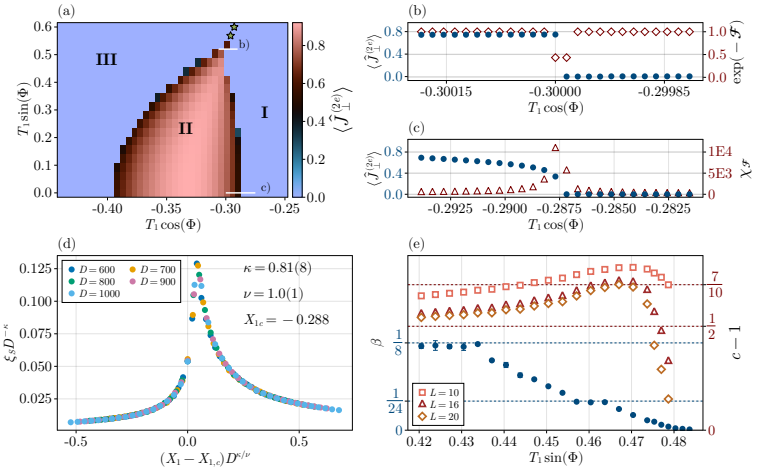


Figure 6. Numerical results obtained with fixed $T_2 = 0.6$. (a): Expectation value of the order parameter $\hat{J}_\perp^{(2e)}$. Although, the order parameter is zero crossing the location of the green stars, we observe a first-order phase transition indicated by a discontinuity of the fidelity density. Thus we identify the left and right part of the phase diagram as the phases I and III, consistent with the semi-classical picture. (b): First-order phase transition discontinuity of the fidelity density $\exp\{(-\mathcal{F})\}$ and the order parameter $\langle \hat{J}_\perp^{(2e)} \rangle$ between phases II and I at $X_2 = 0.52$ [cut b) in panel (a)]. (c): Identifying the second-order phase transition along the cut c) in panel a) at $X_2 = 0$ by the singular behaviour of the fidelity susceptibility $\chi_\mathcal{F}$ and order parameter. (d): Data collapse of the correlation length ξ_s at $X_2 = 0$ for five values of the bond dimension D by employing a finite-entanglement scaling[26]. (e): Critical exponent β obtained by fitting $\langle \hat{J}_\perp^{(2e)} \rangle$ as a function of X_1 for $0.42 < X_2 < 0.49$ and bond dimension $D = 600$ (blue dots). Two plateau appear close to the Ising ($\beta_{\mathrm{IS}} = 1/8$) and TCI ($\beta_{\mathrm{TCI}} = 1/24$) predictions. The central charge (empty symbols) where obtained using finite size DMRG simulations and increases from $c \simeq 1 + 1/2$ to $c \simeq 1 + 7/10$ before dropping to $c \simeq 1$. Reprinted with permission from Ref. 21.

occupation ($n_{\alpha,j} = 0$) on the island (α, j):

$$\widehat{N}_{\alpha,j} = \mathrm{diag}\left(\ldots, -2, -1, 0, 1, 2, \ldots\right) .$$

It is easy to show that the operator $e^{i\hat{\varphi}_{\alpha,j}}$ must to be of the form

$$(e^{i\hat{\varphi}_{\alpha,j}})_{j,j+1} = \delta_{j,j+1}$$

for the canonical commutator $[\widehat{N}_{\alpha,j}, e^{i\hat{\varphi}_{\alpha,j}}] = -e^{i\hat{\varphi}_{\alpha,j}}$ to hold. Furthermore, in order to fit the model into a computer, it is crucial to truncate the possible number of charge excitations, $|n| < N_{\max}$. This leads to an truncated local Hilbert-space of dimension $2N_{\max} + 1$ per each SC island. We explicitly checked convergence with respect to the included charge excitations $N_{\max}$ and observed that $N_{\max} = 6 - 9$ was sufficient.

In our numerical simulations, we constructed a VUMPS ansatz with a two-site unit cell, where the first position inside the unit cell represents the upper leg and the second position represents the lower leg of the ladder. By minimising the Hamiltonian in Eq. 3 for various choices of the parameters, we where able to map out the phase diagram as shown in Fig. 6. We identified the different phases of the model by using the local order parameter $\hat{J}_\perp^{(2e)}(x) = \sin\left(\sqrt{2}\hat{\varphi}(x)\right)$ to distinguish the ordered phase II from the disordered ones (I

and III), see Fig. 6a). To characterise the phase transitions, we used the correlation length and its scaling properties with an increasing bond dimension. Indeed, since a finite bond dimension induces a spurious length-scale, it is possible to use similar machinery to the usual finite-size scaling and to perform a data collapse to extract critical exponents. At sufficiently low values of the tuning flux Φ, this confirms the Ising nature of the transition between II and I, see Fig. 6d). We continued to classify the phase-transitions by using either fidelity measures, Fig. 6b) and c), or identifying the excitation gap in the model. We successfully showed that the second-order phase transition between phase II and I terminates into a first-order phase transition between the same two phases. By measuring the critical exponent of the local order parameter $\hat{J}_{\perp}^{(2e)}(x)$, together with measurements of the central charge from finite-size DMRG[e], we where able to demonstrate that at the point where the first-order transition merges into the second-order transition, a TCI point emerges as expected, see Fig. 6e).

3.2 Describing Bosonic Fractional Hall States with iPEPS

One of the most actively pursued goals in the cold atom community in recent years is the preparation and manipulation of fractional Hall states with bosons. Several recent experimental breakthroughs in this direction have been achieved with small number of particles[27,28]. An intriguing question in this line of research is which of the fractional Hall states in these experiments, which are typically prepared as ground state of interacting Harper-Hofstadter Hamiltonians, would survive in large-scale setups and which instead are stabilised only by the finite size effect of the small scale experiment. The Hamiltonian in question is

$$\hat{H} = -t \sum_{\langle ij \rangle} (e^{iA_{ij}} \hat{a}_i^\dagger \hat{a}_j + h.c.) - \mu \sum_i \hat{n}_i + \frac{U}{2} \sum_i \hat{n}_i(\hat{n}_i - 1). \tag{4}$$

which is a hopping Hamiltonian in two dimensions, with a chemical potential μ and on-site repulsive interaction U. The hopping parameter t is modified (via Peierls substitution) with a local phase such that a particle moving around a plaquette of the lattice would acquire a phase corresponding to the desired magnetic flux, $\phi = \sum_\square A_{ij}$.

In this work[29], which we summarise in this section, we approached this two-dimensional model with the state-of-the-art iPEPS machinery[15,19] in order to get access to the ground state in the thermodynamic limit. In this way we can determine which phases should be in fact stable in very large scale experiments. It should be noted that – as we discuss in our paper – this kind of chiral gapped states was for long time elusive to numerical investigations with traditionally (i.e., non-variationally) optimised iPEPS: therefore, a first crucial task was to prove at all that our algorithms were suitable for the scope. To this end, we first consider the hardcore-limit, $U \to \infty$, in which the maximum number of bosons per site is at most one. For this case, we varied the chemical potential μ to increase the local density $\langle n \rangle$ in the ground state. As shown in the left panel of Fig. 7 we observe an

[e] One could try to extrapolate the central charge from a bond-dimension scaling of the VUMPS data, too. However, the presence of a background gapless (charge) mode makes such an extrapolation extremely difficult in practise. Using the data collapse, we estimated the necessary bond dimension to be of order 10^5, far out of reach for any numerical simulation. Hence we resorted to more standard extrapolations from finite system bipartitions in this case. See Ref. 21 for more details.

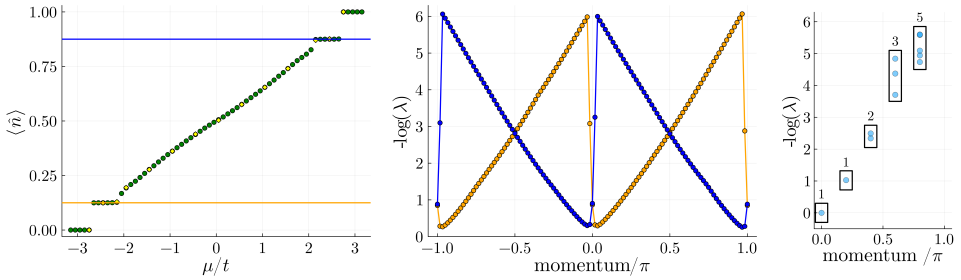


Figure 7. (left): Local bosonic occupation $\langle n \rangle$ as a function of the chemical potential. The yellow line indicates the filling factor for the Laughlin state while the blue line indicates its hole analog. (centre): The low lying part of the spectrum on the edge of state on the incompressible plateaus are shown. We find chiral spectra for both states on the plateaus. We notice that the chirality is reversed for the two plateaus. (right): Degeneracy counting of the lowest branch of the edge spectrum of the fractional Hall state. Taken from Ref. 29.

incompressible plateau ($\frac{d\langle n \rangle}{d\mu} = 0$) at the values of the filling factor $\nu = \frac{2\pi\langle \hat{n} \rangle}{\phi}$ that correspond to those expected in a Laughlin-state of bosons as well as the Laughlin-state of holes. The Laughlin state is the paradigmatic states for the fractional quantum Hall phenomenon. It should be noted that we also investigated the case of finite interaction U in our study and found other stable plateaus, see Ref. 29 for more details. In order to confirm that the states of the incompressible plateaus at the filling $\langle n \rangle$ indeed correspond to the Laughlin-states of bosons we investigate the physics of the edge-spectrum of this state. This edge spectrum is of chiral nature and has properties that can act as a smoking gun for fractional Hall state in question. Even though we have chosen a translationally invariant wave function with our iPEPS we can still access the physics of the edge by what is known as a *bulk-boundary correspondence* for the iPEPS[30,31]. This technique utilises the direct accessibility of entanglement quantitites, like the entanglement spectrum, to get information on the edge spectrum of our system via the well know conjecture relating edge- and entanglement spectra[8]. Using this technique, we can access the low lying part of the spectrum on the edge of a semi-infinite plane. As shown in the left panel of Fig. 7, we can clearly find a very chiral spectrum for both the Laughlin states, and its hole analog. Their respective chirality is reversed as expected. Additionally, we can also look into the spectrum of the edge of a semi-infinite cylinder of finite circumference. Due to this finite circumference of the cylinder the momentum along this direction is discrete. This allows us to investigate the degeneracy of the lowest momentum modes, which are determined by the Conformal Field Theory (CFT) associated with the edge of a fractional Hall state. In the case of a Laughlin state we have a chiral bosonic theory on the edge, which manifests in a counting following the partition of integers. We observe precisely this counting as shown in the right panel of Fig. 7. With these findings, we were able to show convincingly the presence of fractional Hall states as thermodynamic phases in the bosonic Harper-Hofstadter model and by extension large-scale cold atom experiments of the future. On the technical side, we were able to put to rest doubts of the tensor-network community about the applicability of the iPEPS ansatz for chiral topological states, like the Laughlin-state. We showed that upon variational optimisation of the iPEPS one can successfully determine the ground

state of a model of strong experimental relevance, hosting these chiral topological states as ground states.

4 Conclusions & Outlook

In this brief contribution, we have offered a quick overview of a couple of tensor-network algorithms directly tackling the thermodynamic limit of quantum many-body systems and reviewed a couple of recent studies of ours, both for quantum simulation and for genuinely theoretical purposes. We refer the interested readers to the full publications[21,29,15] and to references therein for more details. We did not discuss quantum dynamics, though this is attainable via tensor network algorithms both in closed and open quantum systems, as long as entanglement spreading allows for a faithful representation of the underlying state. Again, we refer the reader to general reviews for further details.

Before concluding, we want also to stress that TN techniques are quickly growing out of their original field (quantum many-body physics, indeed) to reach a very diverse range of fields from the solution of complex integro-differential equations (such as the Dyson or Parquet equation of quantum field theory)[32] to quantum chemistry[33], from classical hydrodynamics[5] to plasma physics[34], and stretching out to financial option pricing[7] and machine learning[6]. These techniques leverage the existence of an underlying mathematical structure to provide spectacular speed-ups (up to exponential) for some applications. In the last two-three years, it became increasingly more obvious that many common problems in applied mathematics and physics actually possess this underlying structure and could therefore benefit from accelerated simulations. We are still in the early days of these developments and most researchers still limit themselves to proof-of-concepts: we can be confident to see developments in near- to mid-term future.

Noticeably – despite the existence of several community codes for tensor manipulation[35–37] – no standard, fully scalable HPC software, accessible to the fast-growing community of potential users exists to this date. By this we mean that the ubiquitous multi-linear algebra tasks are still handled via the somewhat naive workflow of flattening and reshaping them into standard linear algebra operations typically designed for matrix and vector manipulations and codified into BLAS and LAPACK libraries. Designing and developing smarter, hardware-aware, low-level primitives for typical tensor contractions remains a formidable challenge with an high prize at stake, especially in terms of unleashing the upcoming Exascale computing facilities, which we hope to witness in future NIC Symposia.

Acknowledgements

We acknowledge support from the Deutsche Forschungsgemeinschaft (DFG) project Grant No. 277101999 within the CRC network TR 183, and under Germany's Excellence Strategy - Cluster of Excellence Matter and Light for Quantum Computing (ML4Q) EXC 2004/ 1 – 390534769. The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (`www.gauss-centre.eu`) for funding this project by providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS Supercomputer JUWELS[38] at the Jülich Supercomputing Centre (JSC) (Grant NeTeNeSyQuMa) and the Forschungszentrum Jülich for JURECA[39] (institute project PGI-8).

References

1. R. Orús, *A practical introduction to tensor networks: Matrix product states and projected entangled pair states*, Annals of Physics, **349**, 117-158, 2014.
2. P. Silvi et al., *The Tensor Networks Anthology: Simulation techniques for many-body quantum lattice systems*, SciPost Phys. Lect. Notes, 8, 2019.
3. T. Ayral et al., *Density-Matrix Renormalization Group Algorithm for Simulating Quantum Circuits with a Finite Fidelity*, PRX Quantum, **4**, 020304, 2023.
4. J. Tindall et al., *Efficient Tensor Network Simulation of IBM's Eagle Kicked Ising Experiment*, PRX Quantum, **5**, 010308, 2024.
5. N. Gourianov et al., *A quantum-inspired approach to exploit turbulence structures*, Nat. Comput. Sci., **2**, no. 1, 30-37, 2022.
6. C. Roberts et al., *TensorNetwork: A Library for Physics and Machine Learning*, 2019, arXiv:1905.01330.
7. M. Kastoryano and N. Pancotti, *A highly efficient tensor network algorithm for multi-asset Fourier options pricing*, 2022, arXiv:2203:02804.
8. H. Li and F. D. M. Haldane, *Entanglement Spectrum as a Generalization of Entanglement Entropy: Identification of Topological Order in Non-Abelian Fractional Quantum Hall Effect States*, Phys. Rev. Lett., **101**, 010504, 2008.
9. F. Pollmann, A. M. Turner, E. Berg, and M. Oshikawa, *Entanglement spectrum of a topological phase in one dimension*, Phys. Rev. B, **81**, 064439, 2010.
10. M. Dalmonte et al., *Entanglement Hamiltonians: From Field Theory to Lattice Models and Experiments*, Annalen der Physik, **534**, no. 11, 2200064, 2022.
11. J. Haegeman et al., *Unifying time evolution and optimization with matrix product states*, Phys. Rev. B, **94**, 165116, 2016.
12. I. P. McCulloch, *Infinite size density matrix renormalization group, revisited*, 2008, arXiv:0804.2509.
13. G. Vidal, *Classical Simulation of Infinite-Size Quantum Lattice Systems in One Spatial Dimension*, Phys. Rev. Lett., **98**, 070201, 2007.
14. F. Verstraete and J. I. Cirac, *Renormalization algorithms for Quantum-Many Body Systems in two and higher dimensions*, 2004, arXiv:cond-mat/0407066.
15. J. Naumann, E. L. Weerda, M. Rizzi, J. Eisert, and P. Schmoll, *An introduction to infinite projected entangled-pair state methods for variational ground state simulations using automatic differentiation*, SciPost Phys. Lect. Notes, 86, 2024.
16. R. Orús and G. Vidal, *Simulation of two-dimensional quantum systems on an infinite lattice revisited: Corner transfer matrix for tensor contraction*, Phys. Rev. B, **80**, 094403, 2009.
17. J. Jordan et al., *Classical Simulation of Infinite-Size Quantum Lattice Systems in Two Spatial Dimensions*, Phys. Rev. Lett., **101**, 250602, 2008.
18. P. Corboz, *Variational optimization with infinite projected entangled-pair states*, Phys. Rev. B, **94**, 035133, 2016.
19. J. Naumann and E. L. Weerda, *variPEPS – a versatile tensor network library for variational ground state simulations in two spatial dimensions*, `https://github.com/variPEPS`.
20. V. Zauner-Stauber et al., *Variational optimization algorithms for uniform matrix product states*, Phys. Rev. B, **97**, 045145, 2018.

21. L. Maffi, N. Tausendpfund, M. Rizzi, and M. Burrello, *Quantum Simulation of the Tricritical Ising Model in Tunable Josephson Junction Ladders*, Phys. Rev. Lett., **132**, 226502, 2024.
22. R. S. K. Mong et al., *Universal Topological Quantum Computation from a Superconductor-Abelian Quantum Hall Heterostructure*, Phys. Rev. X, **4**, no. 1, 011036, 2014.
23. A. Roy et al., *The quantum sine-Gordon model with quantum circuits*, Nucl. Phys. B, **968**, 115445, 2021.
24. J. Shabani et al., *Two-dimensional epitaxial superconductor-semiconductor heterostructures: A platform for topological superconducting networks*, Phys. Rev. B, **93**, no. 15, 155402, 2016.
25. T. Giamarchi, *Quantum Physics in One Dimension*, Oxford University Press, 2003.
26. L. Tagliacozzo et al., *Scaling of entanglement support for matrix product states*, Phys. Rev. B, **78**, no. 2, 024410, 2008.
27. J. Léonard et al., *Realization of a fractional quantum Hall state with ultracold atoms*, Nature, **619**, 495, 2023.
28. P. Lunt et al., *Realization of a Laughlin state of two rapidly rotating fermions*, 2024, arXiv:2402.14814.
29. E. L. Weerda and M. Rizzi, *Fractional quantum Hall states with variational projected entangled-pair states: A study of the bosonic Harper-Hofstadter model*, Phys. Rev. B, **109**, L241117, 2024.
30. J. I. Cirac et al., *Entanglement spectrum and boundary theories with projected entangled-pair states*, Phys. Rev. B, **83**, 245134, 2011.
31. H. Li and F. D. M. Haldane, *Entanglement Spectrum as a Generalization of Entanglement Entropy: Identification of Topological Order in Non-Abelian Fractional Quantum Hall Effect States*, Phys. Rev. Lett., **101**, 010504, 2008.
32. Y. Núñez Fernández et al., *Learning Feynman Diagrams with Tensor Trains*, Phys. Rev. X, **12**, 041018, 2022.
33. N. Nakatani and G. K.-L. Chan, *Efficient tree tensor network states (TTNS) for quantum chemistry: Generalizations of the density matrix renormalization group algorithm*, The Journal of Chemical Physics, **138**, no. 13, 134113, 2013.
34. E. Ye and N. Loureiro, *Quantized tensor networks for solving the Vlasov-Maxwell equations*, 2023, arXiv:2311.07756.
35. J. Haegeman et al., "Jutho/TensorKit.jl: v0.12.2", Zenodo, 2024, `https://doi.org/10.5281/zenodo.10574897`.
36. M. Fishman, S. R. White, and E. M. Stoudenmire, *The ITensor Software Library for Tensor Network Calculations*, SciPost Phys. Codebases, 4, 2022.
37. J. Hauschild and F. Pollmann, *Efficient numerical simulations with Tensor Networks: Tensor Network Python (TeNPy)*, SciPost Phys. Lect. Notes, 5, 2018.
38. Jülich Supercomputing Centre, *JUWELS Cluster and Booster: Exascale Pathfinder with Modular Supercomputing Architecture at Juelich Supercomputing Centre*, Journal of large-scale research facilities, **7**, A138, 2021.
39. Jülich Supercomputing Centre, *JURECA: Data Centric and Booster Modules implementing the Modular Supercomputing Architecture at Jülich Supercomputing Centre*, Journal of large-scale research facilities, **7**, A182, 2021.

The State of Factoring on Quantum Computers

Dennis Willsch[1,2], Philipp Hanussek[1,2], Georg Hoever[2], Madita Willsch[1,3], Fengping Jin[1], Hans De Raedt[1], and Kristel Michielsen[1,3,4]

[1] Jülich Supercomputing Centre, Institute for Advanced Simulation, Forschungszentrum Jülich, 52425 Jülich, Germany
E-mail: {d.willsch, m.willsch, f.jin, h.de.raedt, k.michielsen}@fz-juelich.de

[2] FH Aachen University of Applied Sciences, 52066 Aachen, Germany
E-mail: philipp.hanussek@alumni.fh-aachen.de, hoever@fh-aachen.de

[3] AIDAS, 52425 Jülich, Germany

[4] RWTH Aachen University, 52056 Aachen, Germany

We report on the current state of factoring integers on both digital and analog quantum computers. For digital quantum computers, we study the effect of errors for which one can formally prove that Shor's factoring algorithm fails. For analog quantum computers, we experimentally test three factorisation methods and provide evidence for a scaling performance that is absolutely and asymptotically better than random guessing but still exponential. We conclude with an overview of future perspectives on factoring large integers on quantum computers.

1 Introduction

The integer factorisation problem (IFP) is one of the oldest and most fascinating problems in mathematics[1,2]. It is defined as the problem of finding a non-trivial divisor of a composite integer N. Besides its historical significance, the IFP is of central importance to everyday data and communication security, in the sense that the security of common encryption systems and protocols in use is based on the difficulty of solving the IFP for large integers. The latest record is the factorisation of the 829-bit number RSA-250 from the RSA factoring challenge[3], involving a 32M-hour allocation on the JUWELS supercomputer. The best-known algorithms[3–5] to solve the IFP on conventional computers scale (sub)exponentially in the number of bits of the integer N. For this reason, cryptosystems like RSA[6] – currently using integers N with 1024, 2048, or 4096 bits – are still secure.

Quantum computers (QCs) are an emerging technology that promise a breakthrough in the solution of the IFP. We distinguish between **digital** and **analog** QCs. On an ideal digital QC, Shor's algorithm[7–9] can solve the IFP with time and space complexity that is polynomial – not exponential – in the number of bits of N. However, so far only very small integers $N \leq 35$ have been successfully factored[10–13] with Shor's algorithm on a digital QC[a]. By executing Shor's algorithm on a QC simulator using 2048 GPUs of JUWELS Booster, the largest integer that could be factored is the 39-bit number $N = 549\,755\,813\,701 = 712\,321 \times 771\,781$[15] (see Tab. 3 in Ref. 16 for an overview).

[a] Note that there exist many claims of factoring larger integers on digital QCs, but the underlying experiments often rely on a certain kind of oversimplification[14] that makes them equivalent to coin flipping. Even for $N = 15, 21, 35$, one can argue that the explicitly compiled quantum circuits might not have been found without previous knowledge about the answer to the IFP.

For analog QCs, several alternative approaches to solving the IFP exist[17–25]. Analog QCs hold the current record of the largest non-trivial integer factorisation by QC hardware, namely the 23-bit integer $N = 8\,219\,999 = 251 \times 32\,749$ factorised by a D-Wave quantum annealer[25]. Although a polynomial scaling of factorisation by analog QCs has been suggested numerically[17], an exponential scaling is considered more likely. In this article, we show experimental evidence for the latter.

This article is structured as follows. Sec. 2 focuses on solving the IFP with digital QCs. We review the main ideas of Shor's algorithm, its large-scale simulation on JUWELS, and future perspectives of factoring on digital QCs. Sec. 3 discusses three methods of factoring on analog QCs. In this section, we also present results of implementing these methods on quantum annealers. Sec. 4 contains our conclusions.

2 Digital Quantum Computers

A digital QC – also known as a *gate-based* or *universal* QC[26] – is a machine consisting of individually controllable *quantum bits* (qubits). A qubit is defined as a superposition of the classical-bit states "0" and "1" and is commonly written as $\alpha|0\rangle + \beta|1\rangle$ with $\alpha, \beta \in \mathbb{C}$. Crucially, when a qubit is measured at the end of a computation, one always obtains one of the two classical-bit states, namely either "0" with probability $|\alpha|^2$ or "1" with probability $|\beta|^2$. Hence, every QC is a *probabilistic* machine. In a digital QC, each individual qubit (and certain combinations of multiple qubits) are individually operable, and these operations are called *quantum gates*. A digital QC is called *universal*, because in principle, each program for conventional computers can be mapped to a combination of quantum gates with only polynomial overhead (note that this does not imply that everything will run faster on a QC – currently, only a few algorithms with a proven speedup are known).

The current three most promising technologies for digital quantum processing units (QPUs) are superconducting circuits, neutral atoms, and trapped ions. With superconducting circuits, IBM has manufactured a 1121-qubit QPU[27], and Google has demonstrated *quantum error correction* below the surface code threshold on a 105-qubit QPU[28]. With neutral atoms, QuEra has built a logical QPU with 280 physical qubits[29]. Finally, trapped ion QPUs produced by Quantinuum have achieved the best gate performance and an all-to-all connectivity[30,31]. Pioneering European companies producing digital QPUs are IQM focusing on superconducting circuits[32] and eleQtron focusing on trapped ions[33], both of which are being installed for provision at JSC. However, it is important to realise that all existing digital QPUs are still noisy prototypes, meaning that they can usually not compete with conventional (super)computers for most application problems.

2.1 Shor's Factoring Algorithm

Peter Shor proposed an algorithm to solve the IFP with an exponential speedup on an ideal digital QC in 1994[7,9], a result which arguably sparked most of the community's interest to build a digital QC until the present moment. To explain Shor's factoring algorithm, we consider the factorisation of a semiprime $N = p \times q$, i.e., a composite integer N with two unknown, non-trivial prime factors $p, q > 2$. The algorithm consists of four steps that are schematically shown in Fig. 1:

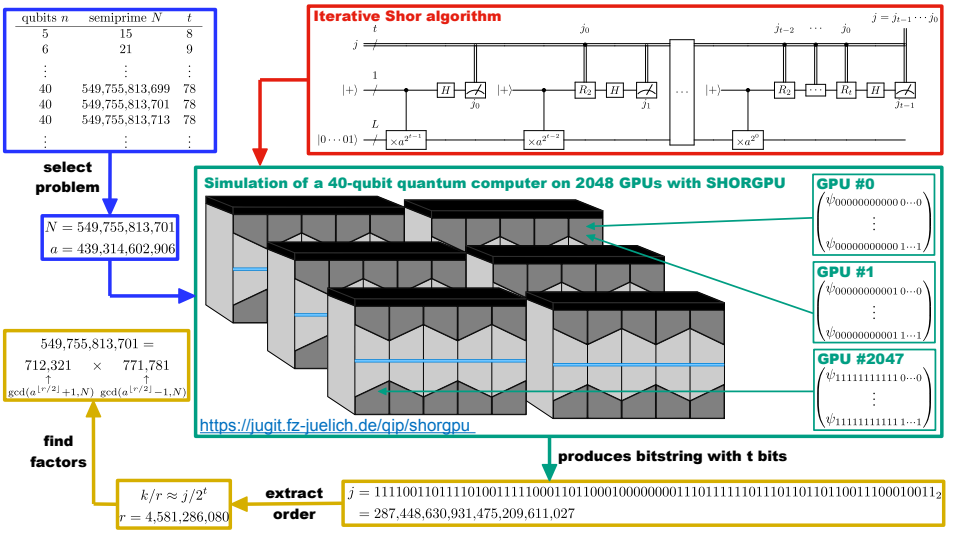


Figure 1. Schematic of performing Shor's factoring algorithm. First, one selects an L-bit semiprime N to factor and a random a (blue). Then, one executes the quantum gates of Shor's algorithm (red), using either a working digital QC or a large-scale simulation on multiple GPUs with `shorgpu`[34], which yields a bitstring $j_0 j_1 \cdots j_{t-1}$ (green). Note that the iterative quantum circuit needs $L+1$ qubits to simulate an L-bit factoring scenario. Finally, the integer j corresponding to the bitstring is post-processed, which yields with high probability a factor of the semiprime N (yellow). Further details are given in Ref. 15.

1. **Parameter Selection** (blue): Choose a random integer a with $2 \leq a < N$ and greatest common divisor $\gcd(a, N) = 1$.[b]

2. **Quantum Algorithm** (red + green): Execute the quantum gates of Shor's algorithm on a digital QC. The result of the QC are t bits $j_0 j_1 \cdots j_{t-1}$, which make up the binary representation of an integer j. The number of bits t is usually twice as large as the number of bits in N. Note that, in principle, the green "simulation" part in Fig. 1 can be completely replaced by a real digital QC once available and working.

3. **Classical Post-Processing** (yellow): Find the largest denominator $r < N$ such that $j/2^t \approx k/r$ using a continued fraction expansion[c].

4. **Factor Extraction** (yellow): Compute $\gcd(a^{\lfloor r/2 \rfloor} \pm 1, N)$[d] which will – with sufficiently high probability – yield one of the factors p or q.

The proof why the algorithm works is beyond the scope of this article. However, it is important to understand that there is a certain probability that Shor's algorithm fails (even

[b] Note that if the greatest common divisor is not 1, it would have to be either p or q, and the problem would have been solved by accident – which is very unlikely for large N. The greatest common divisor can be computed efficiently with the Euclidean algorithm.

[c] The continued fraction expansion is a systematic method that yields successive approximations $k_0/r_0, k_1/r_1, \ldots$ with increasing denominators $r_0 < r_1 < \cdots$ to an arbitrary real number (cf. e.g. Ref. 8).

[d] We note that this expression can be computed efficiently classically, because $\gcd(y, N) = \gcd(y \bmod N, N)$ for all y and the modular exponentiation $a^x \bmod N$ can be computed with the square-and-multiply algorithm.

on ideal digital QCs, in part also due to the probabilistic nature of the QC model itself, as is the case for many other standard quantum algorithms[26]; see Appendix A.2 of Ref. 15 for more information). This motivates us to perform large-scale simulations of Shor's factoring algorithm on JUWELS Booster, to obtain a practical estimate of the success probability – i.e., what *sufficiently high probability* in point 4 above means (see also Refs. 35–40 for related endeavours).

2.2 Large-Scale Simulations

Theoretical estimates of the success probability for Shor's factoring algorithm (as described in Sec. 2.1) are usually very pessimistic and amount to only a few percent[15]. We have designed a digital QC simulation[34] of Shor's algorithm (see Fig. 1, green part) to evaluate the practical performance for over 60 000 factoring problems, with a surprising result: There are many so-called **lucky cases** in which the factorisation is successful, even though Shor's algorithm is, according to theory, not expected to work. Furthermore, we have posed the challenge of factoring, on a real QPU, a non-trivial semiprime larger than the number $N = 549\,755\,813\,701 = 712\,321 \times 771\,781$ that we have factored by executing Shor's algorithm on a simulated QC.

We remark that the wall-clock time that this simulation takes actually grows only linearly with the number of qubits, due to the high degree of parallelism. However, the space complexity is exponential, as simulating the $L+1$-qubit quantum computer requires at least $16 \times 2^{L+1}$ bytes of memory[44,45]. Specifically, `shorgpu`[34], as well as universal QC simulators like JUQCS–G[46], require doubling the number of GPUs with every additional qubit.

A nice benefit of a large-scale QC simulation is that it allows the study of **classical**

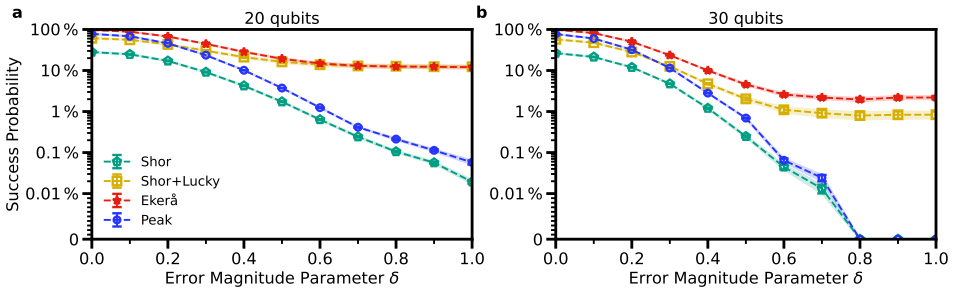


Figure 2. Success probability of factoring on digital QCs as a function of the error magnitude δ for **a**, 20 qubits factoring 19-bit semiprimes, and **b**, 30 qubits factoring 29-bit semiprimes. Markers denote the mean success probability over **a** 500 and **b** 1000 simulated factoring problems for each δ in four different cases: (i) `Shor` (green pentagons) corresponds to Shor's original factoring procedure[7,9], (ii) `Shor+Lucky` (yellow squares) includes the unexpected lucky cases, in which the factorisation works in practise even though the theoretical requirements[15] are not met, (iii) `Ekerå` (red stars) denotes the success probability when using the best-known classically efficient post-processing procedures[41,42] on the measured bitstrings, (iv) `Peak` (blue circles) indicates only the probability to observe a peak in Shor's bitstring-output distribution[15] that is the actual theoretical quantity studied in Cai's proof[43]. At $\delta = 0$, the success probabilities are between 25–100 %, in agreement with Ref. 15 for the no-error scenarios. Note the change from linear to logarithmic scale at 0.01 % on the vertical axes. Shaded areas and error bars indicate the unbiased standard error of the mean. Lines are guides to the eye.

and quantum errors, which affect any QPU device with various orders of magnitude. Of particular interest is an error model proposed by Cai in Ref. 43, for which one can formally prove that Shor's factoring algorithm fails. This error model is expressed in terms of an error magnitude parameter δ[e]. Cai's proof can be seen as formal support for the common viewpoint that for large-scale factoring on a digital QC to work, quantum error correction[26] would be required.

Fig. 2 shows results for the success probability as a function of the error magnitude δ. We see that from 20 to 30 qubits (panels **a** and **b**, respectively), the success probability for Shor's algorithm (green diamonds) indeed drops towards zero for errors with $\delta \geq 0.8$. Interestingly, however, when including the lucky cases (yellow squares), the success probability converges to a non-negligible, finite value. Even though this finite value might decrease exponentially when increasing the number of qubits – in agreement with Cai's proof – it is thus conceivable that the challenge of *limited quantum speedup*[47] posed in Ref. 15 may be met without the above-mentioned requirement for quantum error correction.

2.3 Future Perspectives

The quantum circuit in Fig. 1 needs $L + 1$ qubits to factor an L-bit semiprime. However, on a digital QPU, the individual quantum gates usually need to be compiled into realisable one- and two-qubit gates. This is expected to yield quantum circuits with $2L$ to $2L + 3$ qubits[48–52] (or $1.5L$ qubits with a trick[53]). As these qubits need to perform almost perfectly, a quantum error correction overhead can raise the required number of physical qubits dramatically. For instance, for the factorisation of 2048-bit RSA integers, *several millions of physical qubits* are currently anticipated[54].

Hence, over the past decades, there have been many algorithmic developments and alternative ideas to solve the IFP on digital QCs, often preserving the theoretical concept of an exponential speedup over current algorithms. In particular, the Ekerå-Håstad scheme[55] makes use of another algorithm invented by Shor, namely the discrete logarithm quantum algorithm[7,9]. The advantage of this scheme is that it yields a roughly $75\,\%$ shorter quantum circuit[f]. These optimisations, however, do not directly reduce the number of qubits.

Fascinatingly, Chevignard, Fouque, and Schrottenloher managed[59] to combine the Ekerå-Håstad scheme with a hash function technique[60] to obtain quantum circuits using between $0.5L$ and *less than* L qubits to factorise L-bit RSA integers (see Table 3 in Ref. 59). It is exciting to see what further research along these lines can bring.

[e]The error parameter δ used here and in Ref. 34 corresponds to the *global magnitude parameter* ϵ in Ref. 43, which expresses Gaussian noise on each rotation gate R in the quantum circuit of Shor's algorithm (see Refs. 15, 34, 43 for more information). Specifically, the faulty rotation gate is defined as $\widetilde{R_k} = \mathrm{diag}(1, e^{2\pi i(1+\delta r)/2^k})$ where r is a normally distributed random number.

[f]This means that $t \approx 1.5L$ in Fig. 1 would suffice instead of $t \approx 2L$. In this context, it is also worth mentioning Regev's multidimensional variants of Shor's algorithm[56–58], which also yield an asymptotically shorter quantum circuit.

3 Analog Quantum Computers

Like a digital QC, an analog QC is a machine consisting of individual qubits. However, in contrast to digital QCs, the qubits in an analog QC are not individually and arbitrarily controllable. Instead, after programming an analog QC, the qubits typically undergo a natural evolution for a certain time, and at the end each qubit is measured, yielding a classical-bit state. Note that this does not mean that analog QCs cannot be universal – in fact, one can prove a polynomial equivalence[61–63] to universal digital QCs[g].

Analog QCs are easier to manufacture than digital QCs, mostly due to the more relaxed requirements on individual qubit control. Therefore, much larger analog QPU systems have been built to date. D-Wave has manufactured superconducting quantum annealers with over 5600 qubits, one of which is located in Europe – the JUPSI system hosted at JSC – and a QPU with over 7000 qubits is in development[64]. The companies Pasqal and QuEra build analog QCs based on neutral atoms, with qubit numbers ranging from 196[65] to 256[66] up to 828[67], and the California Institute of Technology reports 6100 coherent atomic qubits[68].

3.1 Factorisation on Quantum Annealers

Quantum annealers are designed to solve optimisation problems. In particular, the D-Wave Advantage QPU addresses the Quadratic Unconstrained Binary Optimisation (QUBO) problem, defined as the minimisation $\min_{x_i=0,1} E(x_0, x_1, \ldots, x_{n-1})$ of the cost function

$$E(x_0, x_1, \ldots, x_{n-1}) = \sum_{i=0}^{n-1} a_i x_i + \sum_{i<j}^{n-1} b_{ij} x_i x_j \,. \tag{1}$$

Here, n is the number of qubits, $x_i = 0, 1$ are the binary problem variables that are represented by qubits on the QPU, and a_i and b_{ij} are the real-valued programmable biases and couplers of the qubits, respectively.

To solve the IFP on quantum annealers, we therefore have to represent the solution to the IFP as the minimum of Eq. 1. The most common approach is to use the qubits $x_0, x_1, \ldots \in \{0, 1\}$ to represent the unknown bits of the factors p and q. We express p (q) using l_p (l_q) bits[h]. Since N is odd – otherwise finding a factor would be trivial – we know that the least significant bits of p and q are 1. Furthermore, since l_p and l_q are fixed, we can set the most significant bits to one. The binary encoding of p and q thus reads

$$p = 1 p_{l_p^*} p_{l_p^*-1} \cdots p_2 p_1 1 \,, \tag{2}$$

$$q = 1 q_{l_q^*} p_{l_q^*-1} \cdots q_2 q_1 1 \,, \tag{3}$$

[g] However, this polynomial equivalence cannot be implemented on most currently existing analog QCs due to technical limitations. For instance, the analog QC would need to support 3-local terms or 6-dimensional quantum digits[61], other so-called *non-stoquastic* properties (cf. Ref. 62 for a comprehensive review), or successive back-and-forth annealing[63].

[h] Since l_p and l_q have to be fixed, one would usually start with $l_p \approx l_q \approx L/2$ (where L is the bit length of the semiprime N to factor) and then start decreasing l_p and increasing l_q until a factor is found. Note that this only incurs a polynomial overhead.

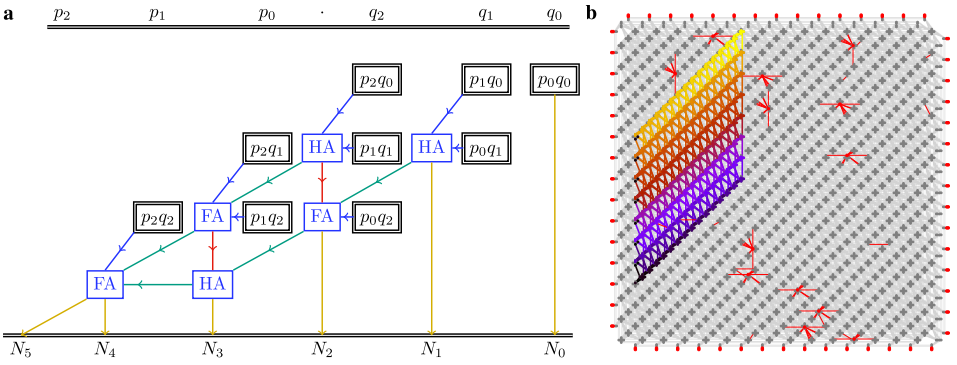


Figure 3. Visualisation of two factorisation methods on a D-Wave quantum annealer. **a**, 3×3-bit multiplication table for the MC method. The bits of p and q are connected with Boolean AND, HA, and FA logic gates. Arrows indicate immediate ancilla qubits representing `and` (blue), `sum` (red), and `carry` (green) bits. **b**, Embedding of a 15×8-bit multiplier onto the D-Wave Advantage 4.1 QPU. Nodes (edges) represent the 5627 physical qubits (40277 couplers) on the QPU. Red lines indicate qubits and couplers that exist in the full underlying Pegasus graph but not on the QPU (by construction or due to fabrication defects). When this multiplier is used for the factorisation of e.g. $N = 3\,548\,021$; the bits of the factor 101010100000111_2 (10100011_2) correspond to the vertically (diagonally) connected unit cells.

where $l_p^* = l_p - 2$ and $l_q^* = l_q - 2$ count the number of unknown bits $l = l_p^* + l_q^*$. Given this encoding, we consider three methods to obtain a QUBO cost function $E(x_0, x_1, \ldots, x_{n-1})$ with n qubits $(x_0, x_1, \ldots, x_{n-1}) = (p_1, \ldots, p_{l_p^*}, q_1, \ldots, q_{l_q^*}, \ldots)$:

1. **Direct Method**[17,18,21]: An obvious cost function to minimise is $f(p,q) = (N - pq)^2$, as its minimum $f(p,q) = 0$ is attained if and only if $N = p \times q$. However, when inserting the binary representations in Eqs. 2 and 3 into this cost function, one obtains higher-than-quadratic terms between the qubits. To solve this problem, one uses a reduction technique that yields $n_{\text{reduction}}$ additional so-called ancilla qubits to obtain a cost function of the form of Eq. 1[i]. We thus need $n = l + n_{\text{reduction}}$ qubits.

2. **Multiplication Circuit Method**[19] (MC Method): A complimentary approach is to write out the binary product $1 p_{l_p^*} p_{l_p^*-1} \cdots p_2 p_1 1 \times 1 q_{l_q^*} p_{l_q^*-1} \cdots q_2 q_1 1$ in a *long multiplication table*. Between all unknown bits, one can then identify Boolean AND, half-adder (HA), and full-adder (FA) gates (see Fig. 3a). For each such gate, one can find a QUBO cost function that attains its minimum if and only if the Boolean logic gate is satisfied. The sum of all these cost functions then yields the final cost function Eq. 1. We remark that also this method incurs additional ancilla qubits representing intermediate "and", "sum" and "carry" bits such that $n = l + n_{\text{and}} + n_{\text{sum}} + n_{\text{carry}}$.

3. **Controlled Full-Adder Method**[25] (CFA Method): Both direct and MC methods need many couplers b_{ij} between the qubits in Eq. 1. However, on the D-Wave Advantage

[i] In Ref. 69, the direct method is equal to the *Modified Multiplication Table (MMT) method*[21] in the limit of maximum block size, where there are no carry variables. For smaller block sizes, the ancilla qubits would consist of both $n_{\text{reduction}}$ qubits from the quadratic reductions and n_{carry} qubits for carry bits in the multiplication table. In this article, we only consider the direct method as its performance was found to be superior to MMT with smaller block sizes[69].

QPU, one qubit is only coupled to 15 other qubits on average. When more connections between qubits are required than physically exist on the QPU, one has to perform a heuristic *embedding*[70] step, by which multiple physical qubits are connected to represent a single logical problem variable. Such an embedding step is often found to hamper the performance of analog QCs[71]. The CFA method is a clever extension of the MC method, in which each Boolean logic gate can be directly embedded onto the qubits of the QPU (see Fig. 3b). Finding such *custom embeddings* is very often the key to successfully solve larger problems on analog QCs.

Further details about each method are given in Ref. 69 and supporting data and open-source code can be found in Ref. 72.

3.2 Results

We have evaluated each of the three factorisation methods for 337 randomly generated factoring problems with up to $l = 22$ unknown bits. The largest factored semiprimes N and the corresponding success frequencies f for the three methods were:

1. **Direct Method**: $N = 1\,042\,441$ with $f = 3.72\,\%$,
2. **MC Method**: $N = 1\,042\,441$ with $f = 0.01\,\%$,
3. **CFA Method**: $N = 3\,844\,417$ with $f = 0.01\,\%$.

The results are shown in Fig. 4a. All methods show larger average success frequencies than random guessing, but the results still suggest an exponential scaling as a function of l.

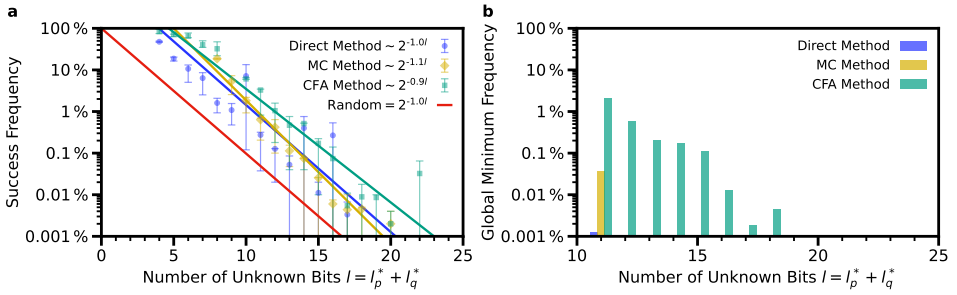


Figure 4. Performance of three factoring methods on analog QCs. **a**, Observed success frequencies as a function of the problem size, given by the number of unknown bits $l = l_p^* + l_q^*$ in the two factors p and q. Markers represent the median success frequencies and error bars denote the 25 % and 75 % percentile for the three methods (see legend). The corresponding lines represent exponential fits $\sim 2^{bl}$ to the data, with the resulting scaling exponents b given in the legend. The red line represents the probability 2^{-l} of randomly guessing the unknown bits. **b**, Frequency of the global minimum of the QUBO cost function, i.e., the sample in which not only the qubits representing p and q but also all additional ancilla qubits are correct (note that by construction of the QUBO, the solution bits representing p and q can be correct even though immediate carry bits are wrong). All results have been obtained on the D-Wave Advantage QPUs 5.4 (direct method, MC method) and 4.1 (CFA method) with ~ 10000 samples for each N and about 10 randomly selected semiprimes N for each l.

Interestingly, the MC method with a fitted scaling exponent of -1.1 seems to perform asymptotically worse than random guessing[j]. We conjecture that this result is due to the requirement for additional physical qubits e.g. from the embedding procedure.

In contrast, the custom-embedded CFA method shows a performance that seems asymptotically better than random guessing[k]. Also, the global minimum of the cost function was found for much larger problems (see Fig. 4b). Although the scaling still seems exponential, a sufficiently small exponent might actually allow analog QCs to first succeed in the near-term factoring challenge posed in Ref. 15. It is certainly interesting to see how the future Advantage2 QPU that is expected to have over 7000 qubits with most qubits coupled to 20 others[64] – which is larger than 15 on the current JUPSI QPU – will cope with the difficult problem of factoring integers.

4 Conclusions and Outlook

In this article, we have studied the problem of factoring integers – one of the key problems that has fuelled the interest in quantum computing – on both digital and analog QCs. For digital QCs, we have analysed an error model for which Shor's factoring algorithm[7,9] can be proven to fail[43], and we have found that unexpected "lucky" factorisations[15] and sophisticated post-processing procedures[42,41] can mitigate this effect.

For analog QCs, we have performed experiments on a quantum annealer. Among three studied factorisation methods, we found evidence that the custom-embedded CFA method[25] performs absolutely and asymptotically better than random guessing, although the data still suggests an exponential scaling as a function of problem size.

Although our results suggest that either error correction on digital QCs or a new method on analog QCs would be necessary, we believe that the factorisation challenge posed in Ref. 15 might be solvable in the near term, and it will be very interesting to see whether it can first be met on a digital or an analog QC. It is conceivable – should the IFP ever be practically solvable with polynomial resources for large integers – that maybe also a triple-hybrid use[73] along with conventional supercomputers may be successful.

Acknowledgements

The authors thank Jaka Vodeb, Paul Warburton, Jin-Yi Cai, and Martin Ekerå for comments and discussions.

D. W., M. W., F. J., and K. M. gratefully acknowledge the Gauss Centre for Supercomputing e.V. (`www.gauss-centre.eu`) for funding this project by providing computing time on the GCS Supercomputer JUWELS[74] at Jülich Supercomputing Centre (JSC).

The authors gratefully acknowledge the Jülich Supercomputing Centre (`https://www.fz-juelich.de/ias/jsc`) for funding this project by providing computing time on the D-Wave Advantage™ System JUPSI through the Jülich UNified Infrastructure for Quantum computing (JUNIQ).

[j] This is possible also for ideal, theoretical quantum annealing, if the annealing time is so short or the energy gap is so small that the annealing process systematically produces excited states with the wrong factors.

[k] We note that even though the fitted CFA scaling given in Fig. 4a is better than random guessing, further results on potentially larger QPUs might be necessary to make a statistically robust statement.

D. W. and M. W. acknowledge support from the project JUNIQ that has received funding from the German Federal Ministry of Education and Research (BMBF) and the Ministry of Culture and Science of the State of North Rhine-Westphalia.

References

1. D. M. Bressoud, Springer, New York, USA, 1989.
2. R. S. Lehman, Math. Comput., **28**, 637, 1974.
3. F. Boudot, P. Gaudry, A. Guillevic, N. Heninger, E. Thomé, and P. Zimmermann, in: Advances in Cryptology – CRYPTO 2020, D. Micciancio and T. Ristenpart (Eds.), Springer International Publishing, Cham, 62, 2020.
4. A. K. Lenstra and H. W. Lenstra, Lecture Notes in Mathematics, Springer, Berlin, Heidelberg, 1993.
5. T. Kleinjung, K. Aoki, J. Franke, A. K. Lenstra, E. Thomé, J. W. Bos, P. Gaudry, A. Kruppa, P. L. Montgomery, D. A. Osvik, H. te Riele, A. Timofeev, and P. Zimmermann, in: Advances in Cryptology – CRYPTO 2010, T. Rabin, (Ed.), Springer, Berlin, Heidelberg, 333, 2010.
6. R. L. Rivest, A. Shamir, and L. Adleman, Commun. ACM, **21**, 120, 1978.
7. P. W. Shor, in: Proceedings 35th Annual Symposium on Foundations of Computer Science, Santa Fe, NM, USA, 124, 1994.
8. A. Ekert and R. Jozsa, Rev. Mod. Phys., **68**, 733, 1996.
9. P. W. Shor, SIAM J. Comput., **26**, 1484, 1997.
10. L. M. K. Vandersypen, M. Steffen, G. Breyta, C. S. Yannoni, M. H. Sherwood, and I. L. Chuang, Nature, **414**, 883, 2001.
11. E. Martín-López, A. Laing, T. Lawson, R. Alvarez, X.-Q. Zhou, and J. L. O'Brien, Nat. Photonics, **6**, 773, 2012.
12. T. Monz, D. Nigg, E. A. Martinez, M. F. Brandl, P. Schindler, R. Rines, S. X. Wang, I. L. Chuang, and R. Blatt, Science, **351**, 1068, 2016.
13. M. Amico, Z. H. Saleem, and M. Kumph, Phys. Rev. A, **100**, 012305, 2019.
14. J. A. Smolin, G. Smith, and A. Vargo, Nature, **499**, 163, 2013.
15. D. Willsch, M. Willsch, F. Jin, H. De Raedt, and K. Michielsen, Mathematics, **11**, 4222, 2023.
16. T. L. Scholten, C. J. Williams, D. Moody, M. Mosca, W. Hurley, W. J. Zeng, M. Troyer, and J. M. Gambetta, 2024, arXiv:2401.16317.
17. X. Peng, Z. Liao, N. Xu, G. Qin, X. Zhou, D. Suter, and J. Du, Phys. Rev. Lett., **101**, 220405, 2008.
18. G. Schaller and R. Schützhold, Quantum Inf. Comput., **10**, 109, 2010.
19. E. Andriyash, Z. Bian, F. Chudak, M. Drew-Brook, A. D. King, W. G. Macready, and A. Roy, Tech. Rep., D-Wave Systems Inc., Burnaby, BC, Canada, 14-1002A-B, 2016.
20. R. Dridi and H. Alghassi, Sci. Rep., **7**, 43048, 2017.
21. S. Jiang, K. A. Britt, A. J. McCaskey, T. S. Humble, and S. Kais, Sci. Rep., **8**, 17667, 2018.
22. W. Peng, B. Wang, F. Hu, Y. Wang, X. Fang, X. Chen, and C. Wang, Sci. China Phys. Mech. Astron., **62**, 60311, 2019.
23. R. Mengoni, D. Ottaviani, and P. Iorio, 2020, arXiv:2005.02268.
24. B. Wang, F. Hu, H. Yao, and C. Wang, Sci. Rep., **10**, 7106, 2020.

25. J. Ding, G. Spallitta, and R. Sebastiani, Sci. Rep., **14**, 1, 2024.
26. M. A. Nielsen and I. L. Chuang, Cambridge University Press, New York, 2010.
27. D. Castelvecchi, Nature, **624**, 238, 2023.
28. Google Quantum AI, 2024, arXiv:2408.13687.
29. D. Bluvstein, S. J. Evered, A. A. Geim, S. H. Li, H. Zhou, T. Manovitz, S. Ebadi, M. Cain, M. Kalinowski, and D. Hangleiter et al., Nature, **626**, 58, 2024.
30. M. P. da Silva, C. Ryan-Anderson, J. M. Bello-Rivas, A. Chernoguzov, J. M. Dreiling, C. Foltz, F. Frachon, J. P. Gaebler, T. M. Gatterman, L. Grans-Samuelsson et al., 2024, arXiv:2404.02280.
31. M. DeCross, R. Haghshenas, M. Liu, E. Rinaldi, J. Gray, Y. Alexeev, C. H. Baldwin, J. P. Bartolotta, M. Bohn, and E. Chertkov et al., 2024, arXiv:2406.02501.
32. J. Rönkkö, O. Ahonen, V. Bergholm, A. Calzona, A. Geresdi, H. Heimonen, J. Heinsoo, V. Milchakov, S. Pogorzalek, M. Sarsby, M. Savytskyi, S. Seegerer, F. Šimkovic, P. V. Sriluckshmy, P. T. Vesanen, and M. Nakahara, EPJ Quantum Technol., **11**, 1, 2024.
33. C. Piltz, T. Sriarunothai, A. F. Varón, and C. Wunderlich, Nat. Commun., **5**, 1, 2014.
34. D. Willsch, 2023, `https://jugit.fz-juelich.de/qip/shorgpu.git`.
35. A. G. Fowler and L. C. L. Hollenberg, Phys. Rev. A, **70**, 032329, 2004.
36. Y. S. Nam and R. Blümel, Phys. Rev. A, **86**, 044303, 2012.
37. Y. S. Nam and R. Blümel, Phys. Rev. A, **87**, 032333, 2013.
38. Y. S. Nam and R. Blümel, Phys. Rev. A, **87**, 060304, 2013.
39. Y. S. Nam and R. Blümel, Phys. Rev. A, **88**, 062310, 2013.
40. Y. S. Nam and R. Blümel, Phys. Rev. A, **97**, 052311, 2018.
41. M. Ekerå, ACM Trans. Quantum Comput., **5**, 11, 2024.
42. M. Ekerå, Quantum Inf. Process., **20**, 205, 2021.
43. J.-Y. Cai, Sci. China Inf. Sci., **67**, 1, 2024.
44. K. De Raedt, K. Michielsen, H. De Raedt, B. Trieu, G. Arnold, M. Richter, Th. Lippert, H. Watanabe, and N. Ito, Comput. Phys. Commun., **176**, 121, 2007.
45. H. De Raedt, F. Jin, D. Willsch, M. Willsch, N. Yoshioka, N. Ito, S. Yuan, and K. Michielsen, Comput. Phys. Commun., **237**, 47, 2019.
46. D. Willsch, M. Willsch, F. Jin, K. Michielsen, and H. De Raedt, Comput. Phys. Commun., **278**, 108411, 2022.
47. T. F. Rønnow, Z. Wang, J. Job, S. Boixo, S. V. Isakov, D. Wecker, J. M. Martinis, D. A. Lidar, and M. Troyer, Science, **345**, 420, 2014.
48. S. Beauregard, Quantum Inf. Comput., **3**, 175, 2003.
49. Y. Takahashi and N. Kunihiro, Quantum Inf. Comput., **6**, 0184, 2006.
50. T. Häner, M. Roetteler, and K. M. Svore, Quantum Inf. Comput., **17**, 0673, 2017.
51. C. Gidney, 2018, arXiv:1706.07884.
52. G. D. Kahanamoku-Meyer and N. Y. Yao, 2024, arXiv:2403.18006.
53. C. Zalka, 2006, arXiv:quant-ph/0601097.
54. C. Gidney and M. Ekerå, Quantum, **5**, 433, 2021.
55. M. Ekerå and J. Håstad, in: Post-Quantum Cryptography, T. Lange and T. Takagi (Eds.), Springer International Publishing, Cham, 347, 2017.
56. O. Regev, 2024, arXiv:2308.06572.
57. S. Ragavan and V. Vaikuntanathan, in: Advances in Cryptology – CRYPTO 2024, L. Reyzin and D. Stebila (Eds.), Springer Nature Switzerland, Cham, 107, 2024.

58. M. Ekerå and J. Gärtner, in: Post-Quantum Cryptography, M.-J. Saarinen and D. Smith-Tone (Eds.), Springer Nature Switzerland, Cham, 211, 2024.
59. C. Chevignard, P.-A. Fouque, and A. Schrottenloher, Cryptology ePrint Archive, Paper 2024/222, 2024.
60. A. May and L. Schlieper, IACR Trans. Symmetric Cryptol., **2022**, 183, 2022.
61. D. Aharonov, W. van Dam, J. Kempe, Z. Landau, S. Lloyd, and O. Regev, SIAM J. Comput., **37**, 166, 2007.
62. T. Albash and D. A. Lidar, Rev. Mod. Phys., **90**, 015002, 2018.
63. T. Imoto, Y. Susa, R. Miyazaki, T. Kadowaki, and Y. Matsuzaki, 2024, arXiv:2402.19114.
64. C. McGeoch, Pau Farré, and K. Boothby, Tech. Rep., D-Wave Systems Inc, Burnaby, BC, Canada, 14-1063A-A, 2022.
65. P. Scholl, M. Schuler, H. J. Williams, A. A. Eberharter, D. Barredo, K.-N. Schymik, V. Lienhard, L.-P. Henry, T. C. Lang, T. Lahaye, A. M. Läuchli, and A. Browaeys, Nature, **595**, 233, 2021.
66. J. Wurtz, A. Bylinskii, B. Braverman, J. Amato-Grill, S. H. Cantu, F. Huber, A. Lukin, F. Liu, P. Weinberg, J. Long, S.-T. Wang, N. Gemelke, and A. Keesling, 2023, arXiv:2306.11727.
67. G. Pichard, D. Lim, E. Bloch, J. Vaneecloo, L. Bourachot, G.-J. Both, G. Mériaux, S. Dutartre, R. Hostein, J. Paris, B. Ximenez, A. Signoles, A. Browaeys, T. Lahaye, and D. Dreon, 2024, arXiv:2405.19503.
68. H. J. Manetsch, G. Nomura, E. Bataille, K. H. Leung, X. Lv, and M. Endres, 2024, arXiv:2403.12021.
69. P. Hanussek, 2024, `https://doi.org/10.34734/FZJ-2024-05254`.
70. V. Choi, Quantum Inf. Process., **7**, 193, 2008.
71. D. Willsch, M. Willsch, C. D. Gonzalez Calaza, F. Jin, H. De Raedt, M. Svensson, and K. Michielsen, Quantum Inf. Process., **21**, 141, 2022.
72. P. Hanussek, 2024, `https://jugit.fz-juelich.de/qip/jupsifactoring`.
73. M. S. Jattana, Phys. Scr., **99**, 095117, 2024.
74. Jülich Supercomputing Centre, J. of Large-Scale Res. Facil., **7**, A183, 2021.

Computational Soft Matter Science

Computational Soft Matter Science

Kurt Kremer

Max Planck Institute for Polymer Research, Ackermannweg 10, 55128 Mainz Germany
E-mail: kremer@mpip-mainz.mpg.de

There are three contributions from Soft Matter Science in this year's NIC Symposium volume, which in a very nice way cover the huge thematic width of the field. The topics range from structure-process relations in membranes through evolutionary molecular dynamics (MD) for biomolecular mechanisms to deformation and failure in glasses. These topics also reveal another important aspect, namely the major switch of research focus from equilibrium to non-equilibrium properties. Exactly this is the area, where Soft Matter can play a prototypical role in statistical physics, because the dynamics of processes is adjustable through molecular parameters and thus direct experimental observation is more easily in reach than for many other materials.

Nano- and microporous membranes play an important role in may fields of modern technology ranging from gas separation and water purification to biomedical applications. Despite of this general relevance processes needed to produce well controlled systems are quite involved and usually not fully understood. The Göttingen group in the first contribution by Blagojevic *et al.* analyses in a comprehensive simulation approach an important block copolymer based method, which can lead to the desired membranes. By combining AB blockcopolymers, where the different blocks eventually phase segregate, with evaporating and nonevaporating (partially poor) solvents, the process of AB phase segregation and collapse of one component can lead to well defined, stable nanoporous structures. This combination is called self-assembly and non-solvent induced phase separation (SNIPS) combined with evaporation-induced self-assembly (EISA). The trick is to combine and optimise the different physical phenomena such as solvent evaporation, self-assembly, microphase separation, and vitrification in SNPIS with and through the EISA process. This defines a path through a high dimensional parameter space and thus is experimentally extremely challenging. Here simulations of model systems can provide truly valuable general insight. The authors combine in a top down approach continuum and particle based simulations, which allow them to relate process properties to a ternary (polymer, volatile and non-volatile solvent) phase diagram. This top down combination of methods and the use of highly optimised analysis and simulation codes on e.g. JUWELS allowed them to investigate the SNIPS process of integral-asymmetric, isoporous block copolymer membranes at micrometer length and minute time scales, providing important new guidelines for experiment.

The second contribution by Methorst *et al.* explores completely different kind of molecular processes, namely biomolecular processes. Often, biological function is related to calculating free energies of e.g. binding of molecules. Such simulations explore specific reaction paths and are thought to identify most likely pathways between known (relative) free energy minima. Of course, such a selection of pathways is based on a specific scientific hypothesis. Alternatively, one might try to sample many possible molecular arrangements and reconstruct possible mechanisms out of this data set, i.e. similar to a physics based

inverse design. Such an approach is followed here for the selective binding and recognition of lipid membrane specificity. There the specific interaction of the lipid membrane with proteins is governed by unique attributes like curvature, lipid composition, and the organisation of lipids into ordered or disordered regions. Here the focus is on using evolutionary algorithms and Martini force field coarse-grained MD, here called Evo-MD, to design "engineered" short peptide sequences, which in their interaction are especially sensitive to membrane curvature. Starting from short random sequences the peptide sequences are allowed to evolve in their respective environment. To accelerate the simulations, i.e. making them possible at all, the authors assume that neighbouring sequences are also close in their interaction with the membrane and thus only short intermediate runs are needed to decide about acceptance of the newly proposed amino acid sequence. By using such an approach in an optimised way they were able to propose specific short sequences, which might be part of larger peptides, that facilitate taylored interactions with membrane curvature. In a next step this has been combined with machine learning where the authors developed a surrogate model trained on a broad spectrum of relative membrane surface binding free energies for amphiphilic peptides. This demonstrated the use of Evo-MD data to train neural networks.

The third contribution is not so much from the field of soft matter itself. However, the problems to tackle and the mechanisms to investigate often hold for soft matter as well. This contribution deals with deformation and failure of glasses, in this specific case of bulk metallic glasses (BMGs). They, for instance, bear many close similarities with colloidal glasses. Metallic glasses display excellent mechanical properties and might be good candidates for auxetic materials, i.e. materials with a negative Poisson ratio. For instance, they have the highest known damage tolerance, which is the product of fracture toughness and yield strength. At the down side BMGs exhibit strain softening, rendering undesired surface structures potentially leading to cracks. In addition, as for almost all glasses, these properties also depend on the cooling rates applied. To approach the properties of such materials Atila *et al.* performed large scale MD simulations focusing of three problems, (i) the connections of brittleness with the fragile (i.e. non Arrhenius) to non-fragile ("strong") transition close to the glass transition, (ii) the origin of high friction and (iii) the deformation behaviour of auxetic materials. In the first part huge 23 Mio atom systems were, for simulation time scales, slowly cooled and then subjected to nanoindentation simulations. Induced strain curves displayed qualitatively different shear band patterns in the strong and fragile regime. Despite these differences, some similarities are observed, which led the authors to the hypothesis of some "hidden order" in the melt. This hypothesis still is under evaluation. In a second set of simulations the focus was on the low internal friction which can be seen as in contrast to the very large sliding friction. For that internal deformation was induced by a numerical scratching experiment. The atomistic simulations reveal a special feature, namely that an increased softness on the local atomistic scale can results in harder and more elastic systems on large scales. Altogether BMGs seem to outperform the mechanical properties of the respective crystalline systems. Though details of the interactions on atomistic scale determine the results quantitatively, the general qualitative results most probably could apply to other glasses, such as colloidal systems, as well.

These three contributions demonstrate the power of advanced modelling for, here, soft matter science. Improved algorithms and models together with advances in hardware design, allow us to go on from equilibrium properties to molecular processes and from more idealised model systems to specific materials.

Membrane Fabrication via EISA and NIPS: Insights into the Spatiotemporal Evolution from Simulations

Niklas Blagojevic, Shibananda Das, Gregor Häfner, Jiayu Xie, and Marcus Müller

Institute for Theoretical Physics, Georg-August-University, 37077 Göttingen, Germany
E-mail: marcus.mueller@uni-goettingen.de

SNIPS, a combination of evaporation-induced self-assembly (EISA) and nonsolvent-induced phase separation (NIPS) of copolymer solutions, offers a bottom-up approach for fabricating integral-asymmetric, isoporous block copolymer membranes. During EISA, a self-assembled top layer of perpendicular cylindrical domains forms, imparting selectivity for ultrafiltration and water purification. Upon immersion in a nonsolvent bath, NIPS creates a spongy, macroporous support structure from the same material that provides mechanical stability.
Designing membranes with desired characteristics (e.g., copolymer chemistry, isoporous layer thickness, thermal and mechanical stability) for specific applications remains a challenge due to the nonequilibrium nature of the SNIPS process. This process is driven by complex physical phenomena, including solvent evaporation, self-assembly, solvent-nonsolvent exchange, macrophase separation, and glassy arrest.
To optimise permeability and selectivity and guide rational design, we employ GPU-accelerated particle-based and continuum simulations to model the entire SNIPS process. These simulations identify a process window for successful membrane fabrication and elucidate the interplay between structural, thermodynamic, kinetic, and process variables. We find that (i) minor incompatibility between the copolymer's matrix-forming block and the nonsolvent, (ii) glassy arrest at lower polymer concentrations, or (iii) greater dynamic contrast between polymer and solvent lead to a spongy, tortuous substructure. This simulation approach offers a platform for rational membrane design and guides experimental efforts to optimise permeability and selectivity.

1 Introduction

Membrane technology plays a crucial role in addressing some of the most pressing global challenges. As the world population grows, ensuring access to clean water, reducing carbon footprints, and treating waste streams become increasingly critical. Membranes offer an energy-efficient solution for separating a wide variety of gaseous or liquid mixtures, often outperforming traditional methods like distillation in terms of energy consumption[1,2]. Additionally, membrane processes can operate under mild conditions, making them particularly suitable for sensitive applications, such as blood hemodialysis[3] or protein separation[4,5].

The self-assembly and non-solvent induced phase separation (SNIPS) process for block copolymers and solvents enables the fabrication of integral-asymmetric, isoporous membranes[6]. An isoporous top layer, formed by Evaporation-Induced Self-Assembly (EISA), imparts selectivity for ultrafiltration of functional macromolecules or for water purification. This selective layer is supported by a macroporous bottom structure, created through Nonsolvent-Induced Phase Separation (NIPS), which provides mechanical stability. This combination optimises the permeability/selectivity tradeoff.

The SNIPS fabrication process involves various physical phenomena – such as solvent evaporation, self-assembly, macrophase separation, and vitrification – controlled by structural, thermodynamic, kinetic, and process parameters. As illustrated in Fig. 1, the

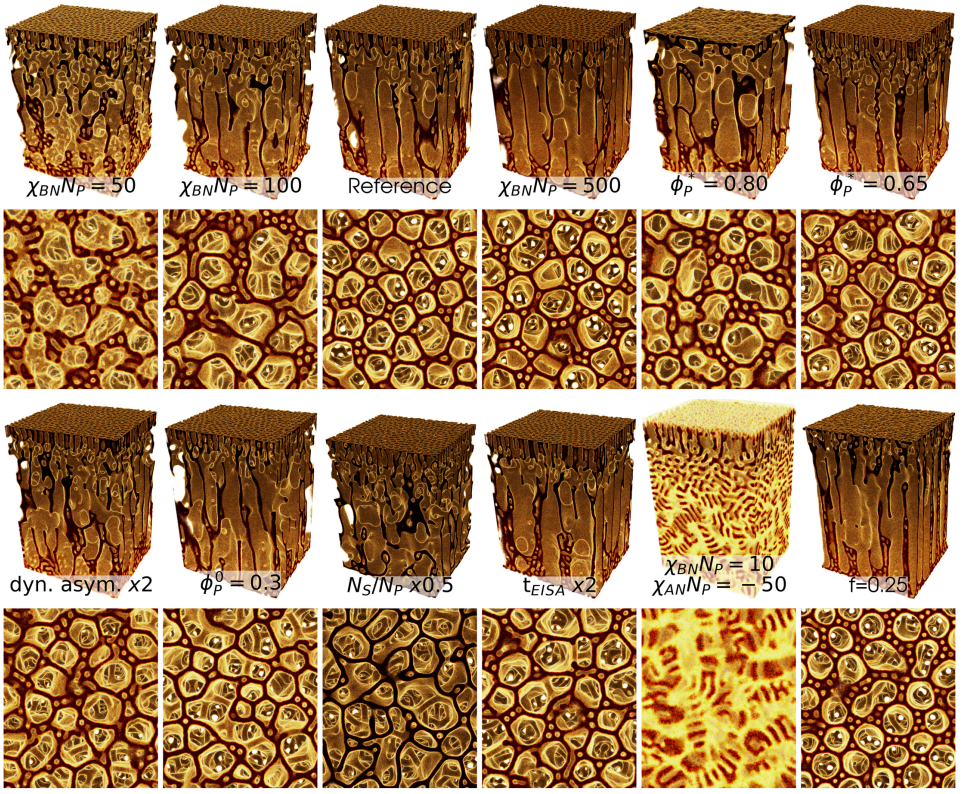


Figure 1. Snapshots of particle-based SNIPS simulations, depicting the majority-block concentration, ϕ_B, for different parameter variations with respect to the reference system. The reference system employs $\chi_{BN}N_P = 150$, $\chi_{AN}N_P = 10$, $\phi_P^* = 0.72$, $\phi_P^0 = 0.33$ and $f = 0.3125$. Top image: 3D image of the concentration of the matrix-forming component, B, of the diblock copolymer. Bottom image: view from the bottom of the membrane at $z = 40R_e$ pointing toward the film surface, $z = 0R_e$. From Ref. 7.

variation in membrane morphology is influenced by factors such as the incompatibility, $\chi_{BN}N_P$, between the matrix-forming block and the nonsolvent, the density, ϕ_P^*, at which the polymer vitrifies, the initial polymer density, ϕ_P^0, in the casting solution, the dynamical asymmetry between the polymer and solvents, and the duration, t_{EISA}, of the EISA process.

Optimising membrane properties and designing fabrication processes rationally is challenging due to the high-dimensional parameter space. However, particle-based and continuum simulations offer valuable insights by investigating how structural, thermodynamic, kinetic, and process parameters influence the final membrane morphology. These "digital twins" of the experimental fabrication process allow independent variation of parameters that may be difficult to achieve experimentally. For example, altering the incompatibility between the majority component, B, of the diblock copolymer and the nonsolvent by changing the chemical structure of the B-block also impacts other factors, such as

the copolymer block incompatibility, $\chi_{AB}N_P$, and the glass-transition dependence of the matrix-forming block on solvent concentration. Furthermore, simulations provide comprehensive data not only on the final membrane morphology but also on concentration profiles and fluxes throughout the entire SNIPS process.

To effectively contribute to the rational design of the fabrication process, simulations must capture the large length and time scales – on the order of micrometers and minutes – characteristic of the experimental SNIPS process. This demands both, a careful selection of models and simulation techniques, and High Performance Computing (HPC) capabilities.

2 Coarse-Grained Top-Down Model: Particle-Based *vs* Continuum Simulations

The SNIPS process involves an AB diblock copolymer, a volatile solvent, S, a nonvolatile solvent, C, and a nonsolvent, N. Instead of modelling a compressible five-component mixture, we simplify the system by treating it as incompressible while explicitly introducing gas particles, G, that are incompatible with the other components, separating the polymer film from the gas (vapour) phase[8].

The large time and length scales necessitate a highly coarse-grained model that captures only the relevant interactions, essential for the SNIPS process. These include (i) the molecular connectivity along the backbone, which determines the copolymer's molecular size, R_e, and dictates the length scale of microphase separation, and (ii) the binary interactions, parameterised by the Flory-Huggins parameters, $\chi_{\alpha\beta}N_P$, where Greek indices denote different particle species, and N_P represents the number of coarse-grained segments of a copolymer. In this highly coarse-grained top-down model, the structure and thermodynamics are governed by just a small number of key parameters – such as R_e and $\chi_{\alpha\beta}N_P$ – which are directly linked to experimental characteristics.

The final membrane morphology represents a nonequilibrium structure. If the system was allowed to fully equilibrate, the result would be a thin, dense, self-assembled copolymer film. However, as the plasticising solvents, S and C, leave the polymer-rich domains, the polymer arrests in a glassy state. This vitrification halts the macrophase separation between the polymer and nonsolvent, ultimately determining the scale of the macroporous substructure. To capture the plasticising effect of the solvents, we use concentration-dependent mobilities[9].

In a particle-based model, the coordinates of the individual coarse-grained particles are the dynamic degrees of freedom. In our simulation, we use forced-biased Monte-Carlo (MC) moves to propagate the system configuration, closely mimicking overdamped Rouse dynamics of unentangled polymers without hydrodynamics. The concentration-dependent mobilities are incorporated by modifying the MC acceptance probability, ensuring detailed balance is maintained. The simulations utilise the Single-Chain-in-Mean-Field (SCMF) algorithm[10], which temporarily replaces weak nonbonded interactions with quasi-instantaneous external fields on a collocation grid, mimicking the interactions between a particle and its surroundings. This approach enables efficient parallelisation, implemented in the GPU-accelerated program SOft coarse grained Monte-carlo Acceleration (SOMA)[11].

The particle-based model enables the description of various molecular architectures. For example, triblock copolymers have been employed to reduce the brittleness of the

final membrane material. Moreover, this particle-based approach directly links molecular dynamics to the kinetics of concentration fields, providing an accurate representation of the structure formation and transformation.

Instead of tracking the coordinates of individual particles, a continuum model represents the system using concentration fields for the different species as the dynamic degrees of freedom. The free energy of a given concentration-field configuration is determined by the Uneyama-Doi free energy functional (UDM)[12], characterised by parameters similar to those in the particle-based model, such as R_e and $\chi_{\alpha\beta}N_P$. However, the connection between these parameters and the underlying molecular characteristics is less direct. Since concentrations are locally conserved, the concentration fields evolve according to model-B[13] or Cahn-Hilliard dynamics, including thermal noise. Gradients in the chemical potential drive concentration fluxes, proportional to an Onsager coefficient, Λ, though accurately determining Λ remains an active area of research. The continuum model represents a higher level of coarse-graining, as concentration fields can be constructed from explicit particle configurations, but multiple particle arrangements can correspond to the same set of concentration fields.

The higher degree of coarse-graining speeds up the computation, and our implementation of the continuum model is about an order of magnitude faster than the particle-based simulations with SOMA. Additionally, the continuum description potentially allows to capture hydrodynamic flow via model-H dynamics[13] and generalisations to viscoelastic phase separation[14] can be envisioned.

3 Thermodynamic Considerations

The casting solution of the reference system is composed of equal parts polymer, $P = AB$, volatile solvent, S, and nonvolatile solvent, C, positioned at the centre of the Gibbs triangle, as shown in Fig. 2. During the process of evaporation-induced self-assembly (EISA), the volatile solvent, S, evaporates, causing a decrease in the volume fraction of solvent, ϕ_S, at the film surface. As a result, the system approximately follows the isopleth $\phi_P = \phi_C = \frac{1-\phi_S}{2}$, indicated by the arrow in the figure. When the surface concentration crosses the critical micelle concentration (CMC), nucleation of isolated A-core micelles can occur. A slight further reduction in ϕ_S triggers an instability leading to microphase separation, represented by the black dashed line. This process ultimately results in the formation of hexagonally ordered cylinders at the film surface.

A basic requirement for SNIPS is the stability of the initial homogeneous casting solution. Additionally, premature nucleation of micelles, well before reaching the spinodal for microphase separation during EISA, could reduce the degree of order in the self-assembled top layer.

The non-black lines in Fig. 2 mark the stability limits of the homogeneous solution in the presence of nonsolvent. Already, a minuscule amount of nonsolvent, $\phi_N \approx 0.2\%$, suffices to induce spontaneous macrophase separation in the system, shown by the blue solid lines. Further increasing ϕ_N rapidly broadens the concentration range in the Gibbs triangle where macrophase instability occurs, as indicated by the significant expansion of the solid lines.

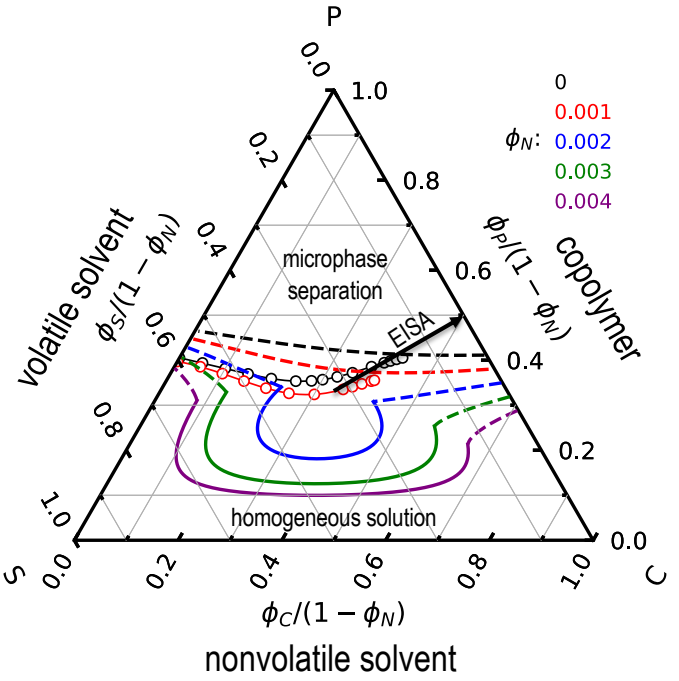


Figure 2. Thermodynamic stability of the homogeneous casting solution is depicted on the ϕ_P-ϕ_S-ϕ_C Gibbs triangle. Dashed and solid lines represent the spinodals for microphase and macrophase separations, respectively, while open-circle lines approximate the critical micelle concentration (CMC) for A-core micelle formation. Curves with different nonsolvent concentrations, ϕ_N, are distinguished by colour, with the black curves representing the initial casting solution, $\phi_N = 0$. The arrow marked "EISA" indicates the evaporation of the volatile solvent S. Adapted from Ref. 7.

4 The EISA Process in Particle-Based and Continuum Simulations

During EISA, the film surface retracts, causing the slower-moving copolymers to accumulate at the descending surface, where they self-assemble into a perpendicular cylindrical morphology. In Fig. 3, we compare the surface retraction and the downward growth of A-cylinders between the particle-based simulation and the continuum model (Uneyama-Doi Model (UDM)), using identical structural and thermodynamic parameters, R_e and $\chi_{\alpha\beta}N_P$.

We simulate the EISA process using the particle-based model[9,7] and extract the concentration fields, $\phi_\alpha(\mathbf{r}, t_0)$, at various times, $t_0/\tau_R = 1.6$, 3.2, and 4.8. These concentration fields are then used to initialise the UDM, allowing for a quantitative comparison of the subsequent time evolution between the two models. To determine the timescale factor, $\tau_R\lambda_0$, we examine the mean concentration deviation, $\Delta\phi(t) \equiv \frac{1}{5V}\int \mathrm{d}\mathbf{r} \sum_\alpha |\phi_\alpha^{\mathrm{SOMA}}(\mathbf{r}, t) - \phi_\alpha^{\mathrm{UDM}}(\mathbf{r}, t)|$. Fig. 3a shows this deviation for different initialisation times as a function of $\tau_R\lambda_0$. The best match is achieved at $\tau_R\lambda_0 = 0.76$, which is used in subsequent analysis.

Fig. 3b shows the time evolution of the mean concentration deviation for this value. Following initialisation, the deviation rapidly increases to $\Delta\phi \approx 7 \cdot 10^{-2}$, where it sta-

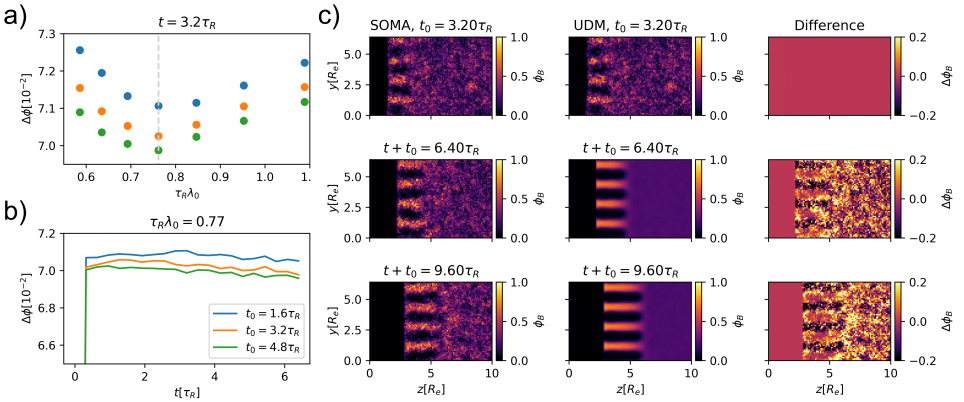


Figure 3. Time-scale matching and concentration comparisons between the two models for the EISA process. a) Concentration deviation, $\Delta\phi$, between the particle-based and continuum models for different initialisation times, t_0, (as indicated in the legend of panel b) after fixed run time of $t = 3.2\tau_R$. The minimal deviation occurs at $\tau_R\lambda_0 = 0.76$, marked by the dashed vertical line. b) Dependence of concentration deviation, $\Delta\phi$, on the run length, t, using the optimal time-scale matching $\tau_R\lambda_0 = 0.76$. c) Time evolution of B-concentration, ϕ_B, shown for a 2D slice near the film surface, for an initialisation time of $t_0 = 3.2\tau_R$. Panels show particle-based simulation (left) and continuum simulation (centre), along with the deviation, $\Delta\phi_B = \phi_B^{\mathrm{SOMA}} - \phi_B^{\mathrm{UDM}}$ (right). Simulation time increases from top to bottom.

bilises, suggesting that the plateau is primarily influenced by thermal fluctuations because the segregation within the self-assembled cylinders is less pronounced in the particle-based simulation compared to the continuum model.

Panel c provides a spatially resolved comparison between the particle-based and continuum simulations of the EISA process. As anticipated, the qualitative time evolution is consistent between the two models, including the surface retraction speed and the growth of cylinder length.

5 Particle-Based Simulation of the Entire SNIPS Process

In Fig. 4, we present 3D images showing the concentration of the matrix-forming block, B, during SNIPS at various times. By $t_{\mathrm{EISA}} = 16\tau_R$, a well-ordered layer of perpendicular cylinders with a thickness of approximately $4R_e$ has formed. At this point, the gas is exchanged with the nonsolvent, N, initiating the NIPS process. The nonsolvent is incompatible with the polymer but miscible with the two solvents, S and C. As the remaining solvent leaves the polymer skin, the polymer concentration exceeds the vitrification threshold, ϕ_P^*, arresting the self-assembled structure in a solid, glassy state. Because the nonvolatile solvent, C, is enriched in the A cylinders, and the nonsolvent, N, is more incompatible with the matrix-forming block, B, the nonsolvent migrates toward the disordered polymer solution beneath the glassy, self-assembled top layer through the A cylinders. Due to the strong incompatibility, $\chi_{BN}N_P$, with the B block, polymer and nonsolvent undergo macrophase separation. The nonsolvent-filled macrovoids are larger than the domains generated by the

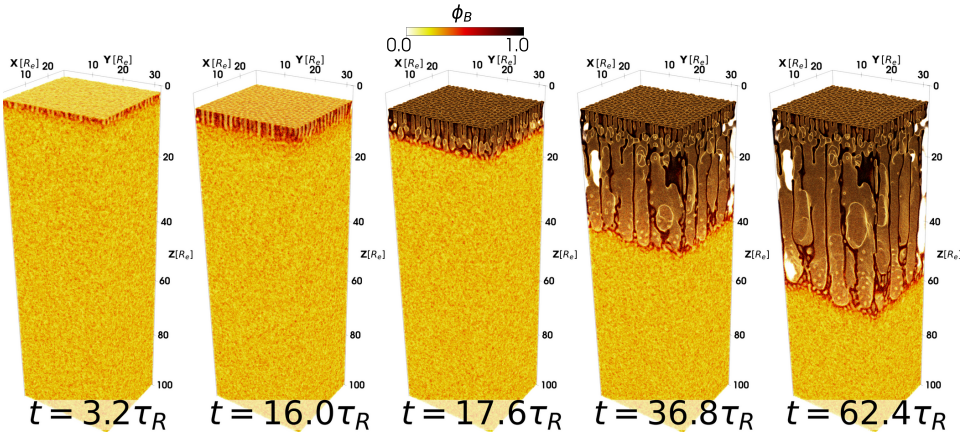


Figure 4. 3D images showing the B-concentration during SNIPS. The transition from EISA to NIPS occurs at $t_{\text{EISA}} = 16\tau_R$. The system's geometry is $27.6 \times 32 \times 100 R_e^3$. From Ref. 7.

diblock copolymer self-assembly. As the macrovoids grow, they further deplete the surrounding polymer-rich regions of solvents, leading these regions to vitrify and halt further macrophase coarsening.

In the top panel of Fig. 5, we quantify the structure using four characteristic lateral xy cross-sections at $t = 62.4\tau_R$. A gradual qualitative change in morphology is observed in the cross-sections as depth increases.

The bottom panel of Fig. 5 shows the structure factor of the polymer concentration, which characterises the macrophase separation between the polymer and nonsolvent:

$$S_P(q_\parallel, z) \propto \left\langle \left| \int \mathrm{d}x\mathrm{d}y \; [\phi_A(x,y,z) + \phi_B(x,y,z)] \; e^{i(q_x x + q_y y)} \right|^2 \right\rangle \tag{1}$$

with $q_\parallel^2 = q_x^2 + q_y^2$. The position of the structure-factor peak, $q_{\max}$, determines the lateral domain size, $d = 2\pi/q_{\max}$, which increases with depth. However, $S_P(q_\parallel, z)$ does not capture the qualitative differences in domain topology between the xy cross-sections at $z = 20R_e$ and $52R_e$.

We complement the layer-resolved structure factor, $S_P(q_\parallel, z)$, with the Euler characteristic combined with morphological erosion to characterise the topology of lateral structures and quantify their characteristic sizes. To achieve this, we convert the polymer concentration in a lateral xy cross-section into a binary black-and-white map using the threshold $\phi_P = 0.26$, where white regions represent polymer-rich areas and black regions represent nonsolvent-rich areas. By counting the number of distinct, unconnected polymer domains, n_{S_P}, and nonsolvent domains, n_{S_N}, we calculate the Euler characteristic as the difference between the two: $\chi_{\text{E}} = n_{S_P} - n_{S_N}$ [15,16] A negative Euler characteristic indicates a topology with multiple nonsolvent-rich domains dispersed within a polymer-rich matrix, as seen in the lateral cross-section of the hexagonally self-assembled polymer skin at $z = 5R_e$ in the leftmost top panel of Fig. 5.

Calculating the Euler characteristic after applying "morphological erosion" to the

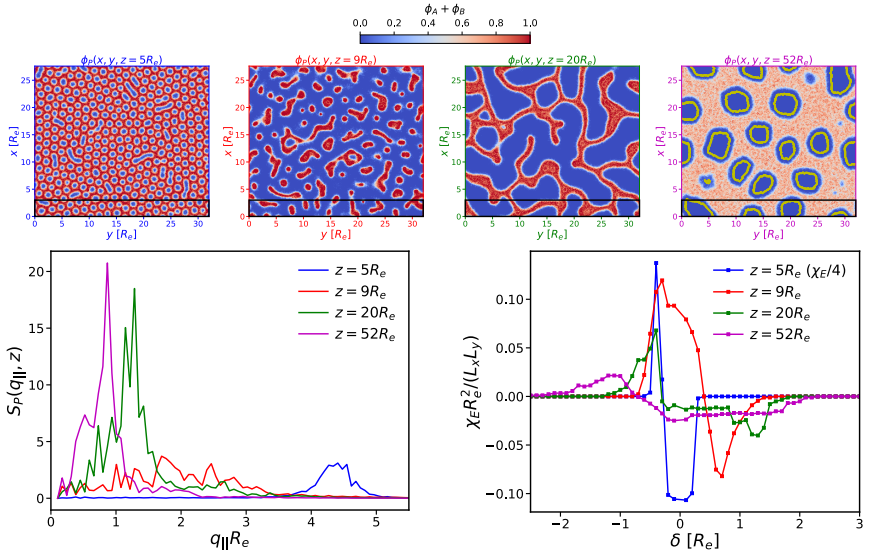


Figure 5. Top row: Lateral xy-cross-sections at depths $z/R_e = 5$, 9, 20, and 52. The yellow contours in the cross-section at $z/R_e = 52$ trace the polymer and nonsolvent interface after erosion of the nonsolvent domains by parallely shifting the interfaces by $\delta = 0.8R_e$. Bottom row: Structure factor, $S_P(q_\parallel, z)$, of polymer concentration (left). Euler characteristic, $\chi_E(\delta)$, as a function of the erosion distance, δ, of the domains (right). Colours indicate the different lateral xy-cross-sections. From Ref. 7.

black-and-white map provides additional insights into the lateral length scales. This process involves shifting each point along the interfaces between the black and white domains in a direction perpendicular to the interface contour by a distance δ. A positive δ expands the polymer-rich domains, effectively eroding the nonsolvent-rich regions, whereas a negative δ erodes the polymer-rich domains. The contours resulting from a parallel shift of $\delta = 0.8R_e$ are shown in the rightmost top panel of Fig. 5 for the xy cross-section at $z/R_e = 52$.

At $z = 9R_e$, just below the polymer skin, the 2D cross-section reveals isolated polymer-rich domains dispersed within the nonsolvent, yielding a positive Euler characteristic, $\chi_E(0) \approx 74 > 0$. The broad size distribution of these polymer domains results in a structure factor with a wide peak, corresponding to a characteristic length scale of $d \approx 3.7R_e$. The positive Euler characteristic, $\chi_E(0) \approx 74$, suggests a similar average distance between polymer-rich domains, estimated as $d = \sqrt{L_x L_y/\chi_E(0)} \approx 3.5R_e$. As the polymer domains expand and merge, χ_E reaches a minimum at $\delta \approx 0.7R_e$, indicating that the typical width of a nonsolvent-rich domain is approximately $1.4R_e$. Conversely, when the polymer-rich domains undergo erosion, they fragment into smaller pieces, signalled by a maximum in χ_E at $\delta \approx -0.3R_e$, corresponding to a typical diameter of approximately $0.6R_e$ for the polymer-rich domains.

In the later stages of NIPS, three qualitatively different layers emerge, as shown in the snapshots at $t = 36.8\tau_R$ and $62.4\tau_R$ in Fig. 6:

(i) *Arrested phase separation*: The first layer, extending from the hexagonally self-assembled polymer skin at the film surface to the upper part of the macroporous mem-

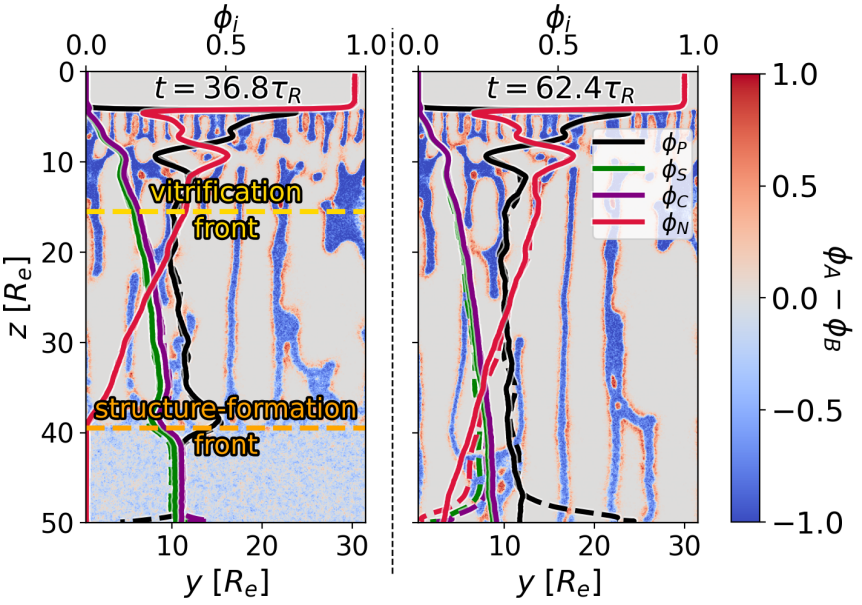


Figure 6. Density profiles at $t/\tau_R = 36.8$ (left) and 62.4 (right). The background presents a 2D slice of $\phi_A - \phi_B$. The vitrification and the structure formation front are indicated in the left panel. Adapted from Ref. 7.

brane substructure, is characterised by minimal temporal changes in polymer concentration. Within this layer, the low solvent concentration in the polymer-rich domains leads to the vitrification of the polymer morphology, particularly of the matrix-forming block B, as the polymer concentration surpasses the glass-transition threshold, $\phi_P \gtrsim \phi_P^*$ during NIPS. In Fig. 6 at $t = 36.8\tau_R$, this vitrification front is marked by a dashed yellow line. The dynamics are primarily governed by the exchange of nonsolvent, N, with the solvents, S and C, which can be described as transport through a porous medium with a complex but static geometry.

(ii) *NIPS-formation*: Proper structure formation occurs in the subsequent layer, which extends from the end of the vitrified layer to the structure-formation front. In this region, the nonsolvent concentration is sufficiently high, $\phi_N > 0.002$ c.f. Fig. 2, to induce macrophase separation between the polymer and nonsolvent, and/or the solvent concentration within the polymer is low enough to trigger self-assembly. For the chosen SNIPS parameters, these macro- and microphase separation fronts coincide. The structure-formation front is indicated by a vertical orange dashed line in Fig. 6 at $t = 36.8\tau_R$.

(iii) *Homogeneous solution*: Deeper within the film, the solution becomes laterally homogeneous, with neither macro- nor microphase separation occurring and only shallow concentration gradients present. In this region, the dynamics can be described as one-dimensional (1D) transport through an unstructured, homogeneous medium, involving the exchange between nonsolvent, N, and solvents, S and C. Despite its simplicity, this region is important to consider to avoid finite-size artifacts due to the bottom boundary.

6 Outlook

The use of highly coarse-grained top-down models, advanced analysis techniques, and GPU-based supercomputers – such as JUWELS at Jülich Supercomputing Centre (JSC) – allows us to investigate the SNIPS process of integral-asymmetric, isoporous block copolymer membranes at micrometer and minute scales. These simulations offer extensive insights into the spatiotemporal evolution of the structure. However, particle-based simulations involving approx. 10^9 highly coarse-grained segments are computationally intensive. To systematically explore the high-dimensional space of structural, thermodynamic, kinetic, and process parameters and rationally optimise the fabrication process, it is crucial to minimise computational costs. This can be achieved by integrating particle-based models with less detailed but computationally faster continuum models. The three-layer structure illustrated in Fig. 6 suggests that detailed particle-based simulations are most necessary near the structure-formation front, whereas the dynamics in the zones of arrested phase separation and homogeneous solutions can be adequately described using continuum models.

Acknowledgements

Financial support has been provided by the Bundesministerium für Bildung und Forschung (BMBF) within the project 16ME0658K MExMeMo and European Union – NextGenerationEU. The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (`www.gauss-centre.eu`) for providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS Supercomputer JUWELS at Jülich Supercomputing Centre (JSC).

References

1. E. Drioli, A. Brunetti, G. Di Profio, and G. Barbieri, *Process intensification strategies and membrane engineering*, Green Chem., **14**, 1561-1572, 2012.
2. D. S. Sholl and R. P. Lively, *Seven chemical separations to change the world*, Nature, **532**, 435-437, 2016.
3. C. Ronco and W. R. Clark, *Haemodialysis membranes*, Nat. Rev. Nephrol., **14**, 394-410, 2018.
4. A. Saxena, B. P. Tripathi, M. Kumar, and V. K. Shahi, *Membrane-based techniques for the separation and purification of proteins: An overview*, Adv. Colloid Interface Sci., **145**, 1-22, 2009.
5. R. van Reis and A. Zydney, *Membrane separations in biotechnology*, Curr. Opin. Biotechnol., **12**, 208-211, 2001.
6. V. Abetz, *Isoporous Block Copolymer Membranes*, Macromolecular Rapid Communications, **36**, 10-22, 2015.
7. N. Blagojevic, S. Das, J. Xie, O. Dreyer, M. Radjabian, M. Held, V. Abetz, and M. Müller, *Toward Predicting the Formation of Integral-Asymmetric, Isoporous Diblock Copolymer Membranes*, Adv. Mater., **36**, 2404560, 2024.

8. O. Dreyer, G. Ibbeken, L. Schneider, N. Blagojevic, M. Radjabian, V. Abetz, and M. Müller, *Simulation of Solvent Evaporation from a Diblock Copolymer Film: Orientation of the Cylindrical Mesophase*, Macromolecules, **55**, 7564-7582, 2022.
9. N. Blagojevic and M. Müller, *Simulation of Membrane Fabrication via Solvent Evaporation and Nonsolvent-Induced Phase Separation*, ACS Appl. Mater. Interfaces, **15**, 57913-57927, 2023.
10. K. C. Daoulas and M. Müller, *Single Chain in Mean Field simulations: Quasi-instantaneous field approximation and quantitative comparison with Monte Carlo simulations*, J. Chem. Phys., **125**, 184904, 2006.
11. L. Schneider and M. Müller, *Multi-Architecture Monte-Carlo (MC) Simulation of Soft Coarse-Grained Polymeric Materials: SOft coarse grained Monte-carlo Acceleration (SOMA)*, Comp. Phys. Comm., **235**, 463, 2019.
12. T. Uneyama and M. Doi, *Density Functional Theory for Block Copolymer Melts and Blends*, Macromolecules, **38**, 196-205, 2005.
13. P. C. Hohenberg and B. I. Halperin, *Theory of dynamic critical phenomena*, Rev. Mod. Phys., **49**, 435-479, 1977.
14. H. Tanaka, *Viscoelastic phase separation*, J. Phys.: Condens. Mat., **12**, R207, 2000.
15. K. Michielsen and H. De Raedt, *Morphological image analysis*, Comput. Phys. Commun., **132**, 94-103, 2000.
16. J. Wang and M. Müller, *Microphase Separation of Diblock Copolymer Brushes in Selective Solvents: Single-Chain-in-Mean-Field Simulations and Integral Geometry Analysis*, Macromolecules, **42**, 2251-2264, 2009.

Exploring the Potential of Evolutionary Molecular Dynamics (Evo-MD) in Uncovering and Controlling Biomolecular Mechanisms

Jeroen Methorst[1,2], Niek van Hilten[3], Kai Steffen Stroh[4], Sebastian Lütge[1], Max Krebs[1], Maria Kelidou[1], and Herre Jelger Risselada[1]

[1] Department of Physics, Technische Universität Dortmund, 44227 Dortmund, Germany
E-mail: jelger.risselada@tu-dortmund.de

[2] Leiden Institute of Chemistry, Leiden University, 2333 CC Leiden, The Netherlands

[3] Department of Pharmaceutical Chemistry, University of California, San Francisco, USA

[4] Laboratory of biology and modelling of cells, ENS de Lyon, France

1 Introduction

Traditional biomolecular research mainly explores how peptides and proteins perform specific functions, often linked to various health issues. Much of this research concentrates on understanding molecular mechanisms because it's traditionally believed that grasping these mechanisms is crucial for controlling function, which remains the ultimate goal.

Biomolecular simulations tackle the molecular mechanisms in biological systems by calculating free energies. This helps to assess the plausibility of hypothesised reactions and how natural proteins overcome related free energy barriers (e.g, Ref. 1). These simulations explore the phase-space along the proposed reaction coordinates, which outline the mechanism, or identify the most likely reaction pathways connecting known intermediates (free energy minima).

An additional, novel approach for understanding biomolecular mechanisms emerges from physics-based inverse design. Unlike traditional approaches that aim to validate existing hypotheses about biomolecular processes, this innovative strategy concentrates on autonomously replicating the biomolecule's functions – such as selective binding – via the automated generation and optimisation of artificial molecular constructs. It achieves this by simulating artificial evolution within biomolecular simulations[2]. By adopting such strategy, we can discover innovative solutions that may not exist in nature but offer vital insights into the individual molecular components and thermodynamic driving forces that mediate specific biological functions. This approach helps to develop a deep understanding of how things work by directly teaching us how functionality is controlled[2].

In this short perspective, we delve into one of the most fundamental yet intricate processes in living organisms: the selective binding and recognition of lipid membrane specificity[3,4]. Lipid membranes, characterised by a thin layer of fats (a fluid-fluid interface), serve as barriers that separate cells from their surroundings and differentiate various internal components, including organelles. Proteins interacting with various cell membranes utilise specific mechanisms to distinguish between them based on unique attributes like curvature, lipid composition, and the organisation of lipids into ordered or disordered phases[3,5,4].

We investigate the application of physics-based inverse design techniques[2], incorporating evolutionary algorithms and coarse-grained molecular simulations, to facilitate the reverse engineering of short peptide sequences capable of recognising these distinctive properties of biological lipid membranes, especially curvature. These reverse-engineered sequences could uncover previously unknown functional domains within larger peripheral membrane proteins. Crucially, this advancement also lays the groundwork for the direct development of peptide drugs that specifically target the unique lipid membranes of viruses, bacteria, and cancer cells. To achieve this goal, we propose a novel method, Evolutionary Molecular Dynamics (Evo-MD) simulations, which has been implemented in our in-house software suite coined EVO-MD, which merges evolutionary algorithms with coarse-grained molecular dynamics simulations. We specifically employ building-block coarse-grained models like the Martini force field[6,7], which condenses multiple atoms into a single interaction site, thereby enhancing computational efficiency by up to 100 times or more. The significant computational boost is essential due to the iterative nature of evolutionary algorithms, which evolve through numerous generations of large genetic populations – comprising hundreds of individuals (simulation systems) – all processed in parallel. Furthermore, the versatility of building block force-fields, which allows for the integration of all interaction types specified within the force-field, is crucial in the evolutionary development of new molecules. This integration ensures that these newly generated molecules are accurately depicted, leveraging a fundamental concept known as force-field transferability[8]. This principle underpins the reliable forecasting of molecular behaviour across varied conditions and system compositions.

The Evo-MD approach guides the evolutionary process from random amino acid sequences towards peptides capable of selectively interacting with complex fluid phases, including biological lipid membranes[9,2,10,11]. We underscore that this method holds great promise for developing peptide-based sensors and therapeutics since it can be customised to identify or selectively target specific characteristics such as membrane curvature, lipid composition, membrane phase (e.g., liquid ordered phases), and protein-fluid phases. While the optimised solutions may not always match biological standards precisely, physics-based inverse design excels at isolating physicochemical principles and thermodynamic drivers behind selective protein-membrane interactions thereby successfully uncovering the signatures ('the design rules') of evolutionary optimisation in nature. Furthermore, we highlight the distinctive capability of the Evo-MD methodology to generate pivotal training datasets for predictive neural network models, strategically covering the relevant physicochemical spectrum within peptide space. This development has now led to the introduction of a publicly available Protein Membrane Interaction prediction (PMIpred) server[11]. This server offers quantitative assessments of membrane binding tendencies and the membrane binding strength of individual amino acid within native proteins, significantly enhancing our understanding and prediction capabilities in the fields of peripheral membrane proteins.

2 Methods and Implementation

We introduce EVO-MD[9] as an example implementation of the physics-based inverse design concept, which we have applied in the development of membrane-interacting peptides[2]. EVO-MD integrates molecular simulations based on building-block coarse-grained

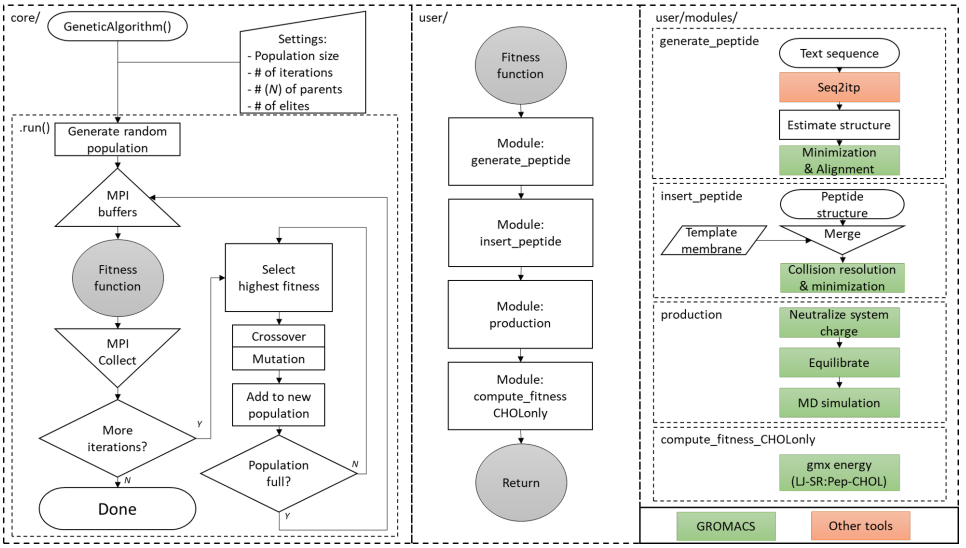


Figure 1. Flowchart of the EVO-MD software build on top of the GROMACS engine for molecular dynamics simulations[12]. Core/ describes the flow of the genetic algorithm. User/ implements the simulation software into various modules that are executed sequentially as part of the fitness function. Figure adopted from Ref. 9.

force-fields, such as the Martini model, into a custom genetic algorithm wrapper program build upon the highly parallel Gromacs engine[12], allowing for the automated setup, production, and subsequent analysis of MD simulations based on candidate sequences selected by the genetic algorithm (see Fig. 1).

The concept of EVO-MD is inspired by the work on virtual creatures performed by Karl Sims in the 1990s[13,14]. Virtual creatures, as simulated by Karl Sims, involve the evolution of virtual block creatures in a simulated dynamic environment. These creatures are created within a computer and undergo a process of variation and selection to improve their ability to perform specific tasks, such as swimming or walking, or even competition for food. The goal is to create creatures with successful behaviours through the evolution of their virtual genes. At its heart, EVO-MD uses the idea of virtual creatures to guide the evolution of biomolecules within a molecular dynamics environment. It harnesses the laws of physics and thermodynamic forces to shape biomolecules starting from completely random sequences[9,10].

The heavy computational and time requirements associated with physics-based inverse design, particularly when considering both large search spaces (e.g. exceeding 20^{24} combinations) and random initialisation, are mitigated using strategies that maximise information extraction from the available simulation data. The largest gain in efficiency follows from the use of relatively short coarse-grained molecular dynamics (MD) simulations for the fitness evaluations. While the simulations are not yet converged within these time frames and the measured observable(s) are therefore far from accurate, genetic algorithms do not require the absolute value of an observable. All that is required for evolution to proceed is an estimation of the relative *ranking* of the solutions within a population as this is the sole criterion on which selection is based. As long as a 'better' solution relatively outperforms

most other solutions, evolution proceeds in the proper direction. However, the closer we approach the (global) optimum in evolution, the smaller the spread in fitness within the population pool and thus the more relevant robust sampling becomes.

Undersampling of simulation observables does pose a problem for the selection step, as outliers – being excessively overestimated observables with respect to their actual value due to undersampling – would almost fully constitute a selection pool after evaluation. We devised several approaches to combat this: (i) We assume that the outliers are due to undersampling and that the distributions of the mean values are (mostly) sequence independent. From this follows that outliers of better solutions (i.e. with high 'true' values for their observable(s)) are more likely to exceed the outliers of worse results, and therefore allows us save time in our simulations by intentionally undersampling (within reason). (ii) We optimise the time that we allocate for sampling through the use of simulation replicas. This can provide a more efficient and accurate estimation of the observable(s) compared to a single, longer simulation – in particular when the largest relaxation time in the system of interest is in the order of the time-scale of the simulation[15,16]. (iii) We eliminate outliers by verifying the best results of each iteration, and maintaining high-performing solutions between iterations. The usage of elitism – where the best performing solutions in a population are directly copied over to the new population – leads to solution occurring which have already been evaluated. To verify the best performers, the corresponding observable(s) are again estimated using independent simulations. The new value is then computed from the weighted average of previous estimations and the new simulation result. This process effectively adds additional replicas into a solution's evaluation, increasing accuracy and thereby removing outliers from the selection pool. Sequences that maintain their high-performing status are retained through a second elitism procedure, to ensure that high-performing solutions are not lost.

Notably, we have recently shifted from a synchronous to an asynchronous genetic algorithm[17]. This change is pivotal because it enhances computational performance and efficiency. Unlike the synchronous version, where CPUs might remain idle up to the exchange of fitness values, the new asynchronous GA ensures continuous utilisation of CPU resources. With the asynchronous approach, once a CPU completes its computation, it immediately communicates with the main process to receive the next task. This immediate feedback loop prevents downtime and maximises resource usage, leading to faster and more efficient processing. Furthermore, adopting an asynchronous method significantly enhanced the computational stability and reliability during large-scale parallel simulations. This improvement was particularly notable in reducing the likelihood of system crashes due to individual replica simulations failing, especially under conditions involving numerous processes.

Finally, since EVO-MD hinges on the efficacy and precision of coarse-grained force fields[8] in accurately depicting lipids, amino acids, and their interactions, the availability of accurate, representative coarse-grained models of these molecules is critical in its success. Notably, optimising force fields resembles a physics-based inverse design challenge. Adjusting interaction parameters translates into a sophisticated optimisation task, assessed through molecular simulations with experimental trial force fields. The goal is to mirror established structural features (such as Radial Distribution Functions, RDFs) and thermodynamic attributes (including phase-transition temperatures) of target molecules. Recognising the Evo-MD method's reliance on precise coarse-grained molecule models,

we have simultaneously explored another avenue of research within the context of search space optimisation. This research employs a swarm intelligence based method coined *GCCompiler*[18] to automate the parameterisation/construction of coarse-grained molecular models within building block force fields, like the Martini model. Readers interested in more details about this work are directed to Ref. 18.

2.1 Results

2.1.1 Optimisation of Amino Acid Sequences Selectively Interacting with Specific Lipids and Lipid Membranes

Initially, we employed EVO-MD to address the challenge of identifying transmembrane domain sequences that exhibit maximum cholesterol attraction[9]. Subsequently, we extended its application to the inverse design of membrane binding peptides (see Fig. 2), focusing on those that demonstrate enhanced binding affinity for positively curved membrane surfaces compared to tension-less, non-curved ones[19,10,11]. These studies illustrated the convergence of evolutionary processes towards a global thermodynamic optimum, even within a vast search space required to cover biological relevant peptide motif lengths, characterised by a dimensionality of (20^{24}) sequences, as evidenced by repetitive recovery of the same solution space when starting from different random peptide sequences (different initial genetic pools). This emphasises the EVO-MD methodology's capacity to provide a unique viewpoint on the thermodynamic forces and concomitant chemical features that drive selective interactions between proteins and lipids, as well as proteins and membranes.

Additionally, unlike data-driven approaches that frequently demand simplification via latent spaces for molecular optimisation, physics-based inverse design avoids the need to reduce system complexity to predefined descriptors[20]. It also eliminates the requirement

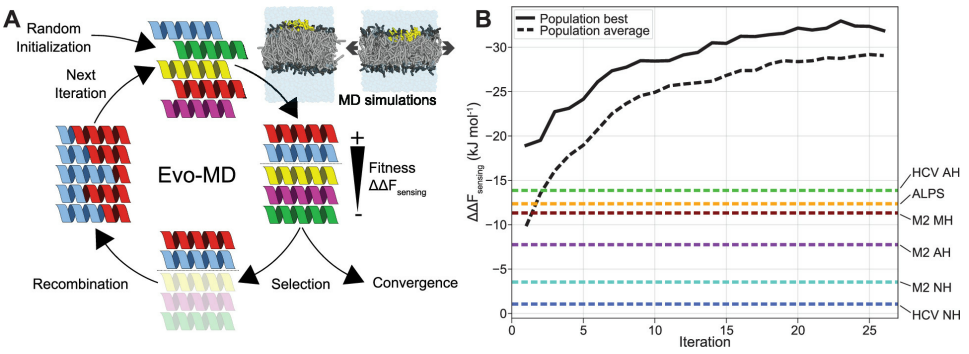


Figure 2. Example of the basic concept of evolutionary molecular dynamics (Evo-MD) in the optimisation of peptide sequences. (A) The Evo-MD scheme for optimising a peptide's affinity toward lipid packing defects associated with positive membrane curvature. Figure adapted from Ref. 9. Positive curvature is modelled using a membrane under tension. Generated peptides (starting from a population of random sequences) are iteratively ranked on their "fitness" (the relative binding free energy, $\Delta\Delta F$), as determined by an end-state free energy calculation method described in Ref. 19. Best sequences are picked and recombined to produce the next generation, leading to gradual evolution toward the optimal lipid packing sensing peptides. (B) Within 25 iterations of EVO-MD, we observe convergence with a fitness that far exceeds the values of that of known lipid packing defect sensors (e.g., HCV AH, ALPS, and M2 MH). Figure adopted from Ref. 19.

for concurrent variational autoencoders/decoders[20]. Consequently, the full complexity of the system is preserved throughout the optimisation process. Furthermore, the insights derived from physics-based design – specifically, the molecular design rules – are naturally comprehensible to humans since they emerge from the observable physical or chemical characteristics of molecules as they progress towards an optimal state. Focusing on these tangible, evolving attributes simplifies the understanding of the thermodynamic forces in action, providing a clear and thorough insight into the mechanisms which facilitate functionality[2].

Moreover, by determining the theoretical optimum from randomly initiated sequences, the EVO-MD methodology facilitates the thorough exploration of the entire search space[2,10]. This approach allows for the strategic generation of data across the full spectrum of achievable fitness values[10,11]. In contrast, existing biological data on functionality is often concentrated due to the common origin of various peptide sequences from the same or related protein families. This strategic data generation capability is particularly advantageous for using EVO-MD data to train neural network models for rapid and cost-efficient fitness predictions often being referred to as 'surrogate models'. Surrogate models serve as valuable tools, for example, within the domain of synthetic polymer design, utilising data produced through rigorous molecular simulations[21]. Essentially, surrogate models serve as substitutes for molecular simulations in predicting fitness. Training surrogate models with EVO-MD data has the potential to reveal alternative viable solutions that might go unnoticed when relying solely on data derived from known biological inputs. This principle is equivalent to fitting an unknown function to data points that are well spaced over the whole range of the applicability domain versus data points that are only clustered within a narrow window[10,2]. Particularly, precise knowledge of the maxima (and minima) of a function – which a physics-based optimisation is able to resolve – will benefit the quality of a fit or model, also within the biologically relevant domain of the search space if the model possesses adequate generalisation capabilities[11]. Following the training of a surrogate model for fitness prediction, it can act as a replacement for coarse-grained MD simulations in the molecular optimisation procedure directed by genetic algorithms. This approach can accommodate retrospective integration of additional practical constraints, such as overall peptide hydrophobicity, into the optimisation process, effectively negating the necessity for recurring, resource-demanding EVO-MD simulations[2].

2.1.2 The Powerful Synergy between EVO-MD and Machine Learning

A compelling illustration of the powerful synergy between the Evo-MD method and machine learning is demonstrated in a recent study, where we developed a surrogate model trained on a broad spectrum of relative membrane surface binding free energies for amphiphilic peptides. This model aimed to forecast curvature sensing behaviour, a characteristic identified in peptide sequences that selectively adhere to vesicles measuring less than approximately 100 nm in diameter or remain soluble[22]. Currently, only around ten sequences, originating from two distinct protein families (α-synuclein and ADP-ribosylation factor GTPase-activating protein[22]), are recognised for possessing this trait. Consequently, there remains a significant gap in the availability of data for predicting curvature sensing sequences.

To address the challenge of data insufficiency, we developed a statistical mechanical

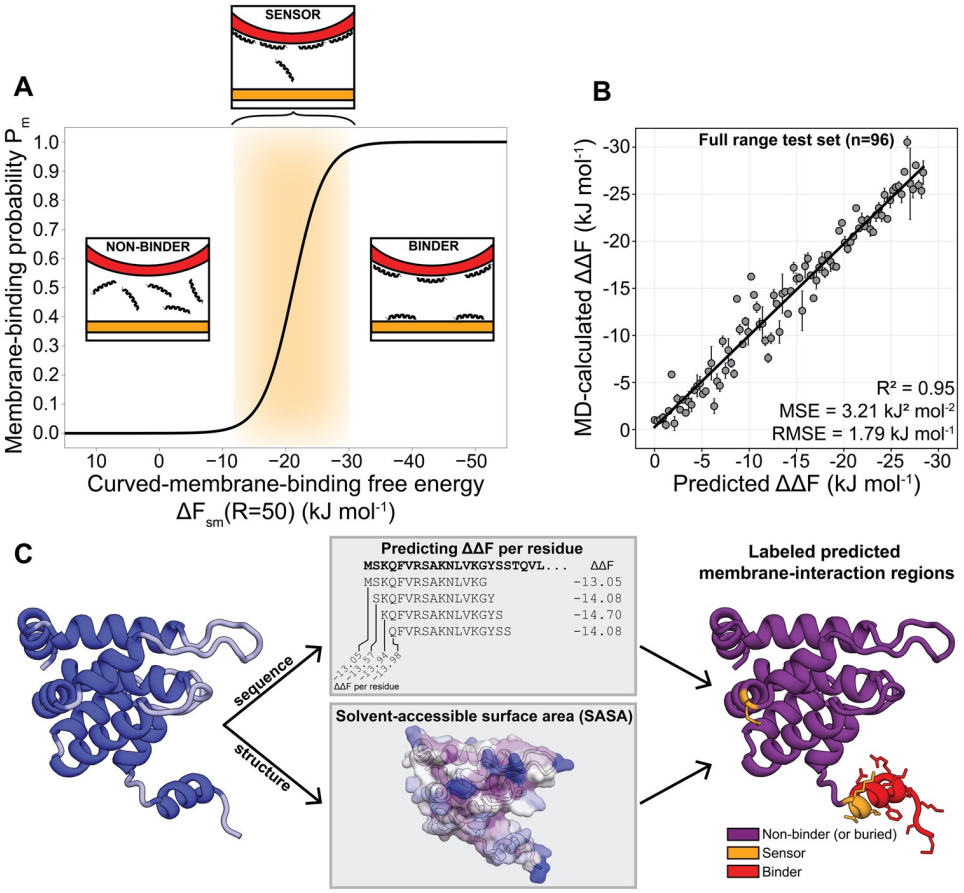


Figure 3. Synergy Between Evo-MD and Machine Learning in Curvature Sensing Sequence Prediction (A) Membrane-Binding Probability and Free Energy: Statistical mechanical model on the likelihood of membrane binding as a function of the absolute membrane-binding free energy. It highlights a critical regime marked by a sudden shift in likelihood. The free energy values associated with sequences that specifically adhere to small liposomes – acting as curvature sensors – are found within or close to this pivotal range. (B) Transformer Model's Predictive Power: Demonstration of the predictive capability of the Transformer-based neural network in forecasting (relative) binding free energies. This model was trained on EVO-MD generated sequences, covering the full spectrum of thermodynamic possibilities. (C) Predicting Membrane Binding in Peripheral Proteins: Shows how machine learning can predict interactions between molecules and membranes in larger, naturally occurring peripheral membrane proteins. This is achieved by analysing the sequence and incorporating knowledge about the structural accessibility of individual amino acids regarding their exposure to solvents (membranes). Figure adopted from Ref. 10.

model to predict the precise range of relative binding free energy associated with curvature sensing (see Fig. 3A). Leveraging EVO-MD as a physics-based generative model, we generated a substantial volume of data strategically encompassing the hypothesised curvature sensing regime. This enriched dataset served as the foundation for training a convolutional neural network (CNN) model, designed to predict the relative binding free energies of authentic peptide sequences (Fig. 3B).

By integrating this physics-enriched CNN model with our theoretical model on the location of the curvature sensing regime illustrated in Fig. 3A, we achieved successful independent classification of the binding behaviours of known peptides, categorising them as soluble, curvature sensing, or exhibiting aspecific membrane binding, with good precision on established datasets. This accomplishment was achieved even without the need for prior data on the limited number of curvature sensors already discovered.

Moreover, the observed accuracy and wide applicability of neural networks trained by Evo-MD data has motivated us to launch the Protein-Membrane Interaction prediction (PMIpred) server (`pmipred.fkt.physik.tu-dortmund.de`)[11], which utilises a transformer model trained by physics-based generation (EVO-MD) of over 54,000 curvature-sensing peptide sequences to quantitatively predict the membrane affinity of peptide sequences. PMIpred is designed to analyse the interaction of large peripheral membrane proteins, leveraging its predictive capabilities for shorter peptide sequences (Fig. 3C). It employs a sliding window approach on protein sequences to compute an average free-energy contribution for each amino acid residue. This analysis is then linked with the solvent-accessible surface area (SASA) of each residue, allowing the module to accurately map and visualise the anticipated membrane-interaction activities within the protein structure. We have tested PMIpred using a broad and varied dataset of known Peripheral Membrane Proteins (PMPs). Our findings indicate that PMIpred performs comparably to leading-edge tools such as DREAMM, PPM3, and MODA[11]. However, what sets PMIpred apart is its much wider applicability and ability to offer quantitative predictions regarding membrane affinities, which are straightforward to interpret in biological terms. Additionally, PMIpred uniquely empowers users to differentiate between curvature-sensing and membrane-binding motifs, providing a more nuanced understanding of protein-membrane interactions than hitherto possible.

3 Concluding Remarks

We have outlined how evolutionary molecular dynamics (Evo-MD) in conjunction with the in-house developed software suite called EVO-MD can open a promising new avenue for the understanding of selective interactions of protein with lipids, lipid membranes, or other relevant fluid-fluid interfaces such as protein fluid phases. The Evo-MD methodology offers an intuitive and comprehensive way to explore and optimise molecular systems, enhancing our ability to understand, predict and control their behaviour without the need of simplification into a latent space. This methodology could introduce a quantum leap in the development of bio-molecular sensors and peptide drugs that either recognise or selectively target membrane curvature, membrane lipid composition or membrane phase (e.g. lipid rafts), and even protein condensates. In addition, we have outlined how it can excel at creating strategic training data for predictive neural network models under circumstances where genuine data is limited.

4 Acknowledgements

We thank Art Hoti, Sebastian Lütge, Nino Verwei, Maria Kelidou, Alireza Soleimani, and Max Krebs for fruitful discussions. H. J. R., N. v. H. and J. M. thank the NWO Vidi

Scheme, The Netherlands, (project number: 723.016.005). HJR thanks also the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) for funding this work under Germany's Excellence Strategy - EXC 2033 - 390677874 - RESOLV.

The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (`www.gauss-centre.eu`) for funding this project by providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS Supercomputer JUWELS at Jülich Supercomputing Centre (JSC).

References

1. Y. G. Smirnova, H. J. Risselada, and M. Müller, *Thermodynamically reversible paths of the first fusion intermediate reveal an important role for membrane anchors of fusion proteins*, Proceedings of the National Academy of Sciences, **116**, no. 7, 2571-2576, 2019.
2. J. Methorst, N. van Hilten, A. Hoti, K. S. Stroh, and H. J. Risselada, *When Data Are Lacking: Physics-Based Inverse Design of Biopolymers Interacting with Complex, Fluid Phases*, Journal of Chemical Theory and Computation, **20**, no. 5, 1763-1776, 2024.
3. M. A. Lemmon, *Membrane recognition by phospholipid-binding domains*, Nature Reviews Molecular Cell Biology, **9**, no. 2, 99-111, 2008.
4. J. Bigay and B. Antonny, *Curvature, lipid packing, and electrostatics of membrane organelles: defining cellular territories in determining specificity*, Developmental Cell, **23**, no. 5, 886-895, 2012.
5. E. Sezgin, I. Levental, S. Mayor, and C. Eggeling, *The mystery of membrane organization: composition, regulation and roles of lipid rafts*, Nature Reviews Molecular Cell Biology, **18**, no. 6, 361-374, 2017.
6. S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries, *The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations*, J. Phys. Chem. B., **111**, 7812-7824, 2007.
7. P. C. T. Souza, R. Alessandri, J. Barnoud, S. Thallmair, I. Faustino, F. Grünewald, I. Patmanidis, H. Abdizadeh, B. M. H. Bruininks, T. A. Wassenaar, P. C. Kroon, J. Melcr, V. Nieto, V. Corradi, H. M. Khan, J. Domański, M. Javanainen, H. Martinez-Seara, N. Reuter, R. B. Best, I. Vattulainen, L. Monticelli, X. Periole, D. P. Tieleman, A. H. de Vries, and S. J. Marrink, *Martini 3: a general purpose force field for coarse-grained molecular dynamics*, Nat. Methods, **18**, no. 4, 382-388, Mar. 2021.
8. H. J. Risselada, *Martini 3: a coarse-grained force field with an eye for atomic detail*, Nat. Methods, **18**, no. 4, 342-343, Mar. 2021.
9. J. Methorst, N. van Hilten, and H. J. Risselada, *Inverse design of cholesterol attracting transmembrane helices reveals a paradoxical role of hydrophobic length*, bioRxiv, July 2021, doi:10.1101/2021.07.01.450699 (accessed Dec 8, 2023).
10. N. van Hilten, J. Methorst, N. Verwei, and H. J. Risselada, *Physics-based generative model of curvature sensing peptides; distinguishing sensors from binders*, Science Advances, **9**, no. 11, eade8839, 2023.
11. N. van Hilten, N. Verwei, J. Methorst, C. Nase, A. Bernatavicius, and H. J. Risselada, *PMIpred: a physics-informed web server for quantitative protein–membrane interaction prediction*, Bioinformatics, **40**, no. 2, btae069, 2024.

12. M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, *GROMACS: High Performance Molecular Simulations Through Multi-Level Parallelism from Laptops to Supercomputers*, Soft. X, **1-2**, 19-25, 2015.
13. K. Sims, *Evolving virtual creatures*, in: Proceedings of the 21st annual conference on Computer graphics and interactive techniques - SIGGRAPH '94, SIGGRAPH '94, ACM Press, 1994, doi:10.1145/192161.192167 (accessed Dec 8, 2023).
14. K. Sims, *Evolving 3D morphology and behavior by competition*, Artificial Life, **1**, no. 4, 353-372, 1994.
15. P. V. Coveney and S. Wan, *On the calculation of equilibrium thermodynamic properties from molecular dynamics*, Phys. Chem. Chem. Phys., **18**, no. 44, 30236-30240, 2016.
16. A. P. Bhati, A. Hoti, A. Potterton, M. K. Bieniek, and P. V. Coveney, *Long Time Scale Ensemble Methods in Molecular Dynamics: Ligand–Protein Interactions and Allostery in SARS-CoV-2 Targets*, J. Chem. Theory Comput., **19**, no. 11, 3359-3378, May 2023.
17. E. Alba and J. M. Troya, *Analyzing synchronous and asynchronous parallel distributed genetic algorithms*, Future Generation Computer Systems, **17**, no. 4, 451-465, 2001.
18. K. S. Stroh, P. C. T. Souza, L. Monticelli, and H. J. Risselada, *CGCompiler: Automated Coarse-Grained Molecule Parametrization via Noise-Resistant Mixed-Variable Optimization*, Journal of Chemical Theory and Computation, **19**, no. 22, 8384-8400, 2023.
19. N. van Hilten, K. S. Stroh, and H. J. Risselada, *Efficient quantification of lipid packing defect sensing by amphipathic peptides: Comparing Martini 2 and 3 with CHARMM36*, Journal of Chemical Theory and Computation, **18**, no. 7, 4503-4514, 2022.
20. B. Sanchez-Lengeling and A. Aspuru-Guzik, *Inverse molecular design using machine learning: Generative models for matter engineering*, Science, **361**, no. 6400, 360-365, July 2018.
21. M. A. Webb, N. E. Jackson, P. S. Gil, and J. J. de Pablo, *Targeted sequence design within the coarse-grained polymer genome*, Sci. Adv., **6**, no. 43, Oct. 2020.
22. B. Antonny, *Mechanisms of membrane curvature sensing*, Annu. Rev. Biochem., **80**, 101-123, 2011.

Deformation and Failure Mechanisms of Bulk Metallic Glasses

Achraf Atila, Sergey Sukhomlinov, and Martin Müser

Department of Material Science and Engineering, Saarland University,
66123 Saarbrücken, Germany
E-mail: martin.mueser@mx.uni-saarland.de

Bulk metallic glasses (BMGs) are known for their excellent mechanical properties, including high tensile and yield strengths. However, they experience strain softening, which limits their technological applications, as it leads to undesirable shear banding during mechanical loading. Using molecular dynamics simulations, we explored the following questions: How does the atomic structure at which the liquid falls out of equilibrium during cooling influence shear banding and, thereby, plastic deformation under loading? Does strain softening prevent BMGs from being viable candidates for use as auxetic materials and low-friction coatings? We observe a quasi-discontinuity in the plastic response of BMGs, depending on whether the material fell out of thermodynamic equilibrium above or below the so-called fragile-to-strong transition temperature T^*. Specifically, when the melt is quenched from $T > T^*$, i.e., from the fragile liquid phase, the material is significantly less prone to strain localisation and shear banding compared to quenching from an equilibrium strong melt ($T < T^*$). Additionally, we found a relatively large friction coefficient for our BMG, even when the indentation had not yet caused shear band formation. This increase is attributed to repeated small-scale plastic deformations occurring as the material is scratched along the same wear track. Finally, we clarified that the negative Poisson's ratio observed in auxetic structures is a large-displacement effect, showing similar strain-dependent behaviour of the Poisson's ratio in BMGs and (poly-)crystalline metals. Since the BMGs yield at substantially larger strains than (nano-) crystalline metals they are promising candidates for the design of auxetic structures.

1 Introduction

Bulk metallic glasses (BMGs) exhibit several excellent mechanical properties compared to crystalline and polycrystalline metals. These include an order-of-magnitude higher elastic strain limit, as well as similarly enhanced tensile[1] and yield strengths, while also being tough[2]. In fact, BMGs have the highest known damage tolerance[3–5,1], defined as the product of fracture toughness and yield strength. On the downside, BMGs exhibit strain softening, which makes them prone to localised deformations in the form of shear bands[3–5,1], leading to undesirable surface markings that can ultimately evolve into cracks. This behaviour limits the moldability of BMGs and their broader technological applicability.

Unravelling the interplay of atomic structure, shear bands, and thermal history – i.e., how a BMG was prepared by quenching from a given melt – has a 70-year-long history and remains a subject of ongoing research[6–11]. It is well-established both experimentally and through simulations that (bulk metallic) glasses produced with slower cooling rates are stronger but also more prone to shear banding than those quenched more quickly[6–11]. These effects are usually studied with continuously varying cooling rates, which smooth out any potential discontinuities that might arise depending on whether the glass's microstructure resembles that of a fragile or a strong glass-forming liquid. This complicates efforts to clearly understand how atomic structure, nanostructure, and thermal history influence the

deformation behaviour and shear band formation in BMGs. Moreover, it remains unclear how the plastic properties of BMGs influence their failure behaviour under different types of mechanical loading, such as nano-indentation, scratching, or tensile strain – particularly in the case of auxetic materials, whose meta-structure results in a negative Poisson's ratio under finite stress. A negative Poisson's ratio means that a material expands in the direction perpendicular to an applied tensile strain, and when compressed, it shrinks laterally instead of expanding.

Here, we report the results of large-scale molecular dynamics (MD) simulations that were conducted to investigate three issues related to the large-strain deformation behaviour of BMGs: (I) the hypothesis that the degree of brittleness in BMGs is closely tied to the fragile-to-strong (FTS) transition during melt quenching. Glass-forming liquids tend to undergo a rather sharp transition from so-called fragile liquids at high temperatures (where the specific heat clearly exceeds that of harmonic solids, and transport properties exhibit non-Arrhenius temperature dependence) to so-called strong liquids (where the specific heat is close to that of harmonic solids, and transport properties reveal Arrhenius-like dependencies) at lower temperatures[12]. (II) The reason for the high friction exhibited by many BMGs, which is counterintuitive given their potentially high coefficients of restitution, is also investigated. (III) The deformation behaviour and mechanisms of a nanostructured auxetic material[13] made of a strong BMG at atomic scales.

2 Atomic-Scale Deformation Mechanisms of Fragile and Strong BMG

The first part of the study aimed to provide an understanding of the differences in the atomic-scale deformation mechanisms of fragile and strong BMGs during nanoindentation and stress release. In addition to the studies mentioned above, we also performed simulations to understand the propagation and interaction of the shear bands in BMGs. In each of these simulations, we investigated $Zr_{0.6}Cu_{0.3}Al_{0.1}$ as a model for a generic but also commercially used BMG[14]. To this end, a potential designed for Zr-Cu-Al ternaries was employed[15].

The first part of the project was achieved by performing MD simulations of our model BMG subjected to nanoindentation simulations using glasses prepared with the slowest effective cooling rate (10^7 K/s) using natural dynamics[16] and the largest system size (23 million atoms) in the literature. A rigid cylindrical indenter with radius $R = 100$ nm was used to perform the nanoindentation at room temperature. The simulation setup is shown in Fig. 1(a).

Fig. 1(b) depicts the force-displacement curves $F(d)$ of strong and fragile BMGs. Differences in the elastic regime ($d \lesssim 3$ nm) occur but are too small to be visible to the naked eye. In the incipient plastic regime, 3 nm $\lesssim d \lesssim$ 12 nm, differences become clearly noticeable between the "strong" and the two "fragile" glasses. While the transition to the stress-softening regime occurs at similar indentation depths near $d = 12$ nm in all samples, the maximum indentation force of the strong glass is much in excess of that of the fragile glasses. Moreover, the stress softening regime is noticeably shortened for the strong glass compared to the fragile ones. Finally, the forces in the saturation regime are almost identical for the fragile systems and clearly enhanced in the strong glass. Increasing T_1 to above 1,000 K does not lead to significant changes in $F(d)$ curves compared to those labelled as

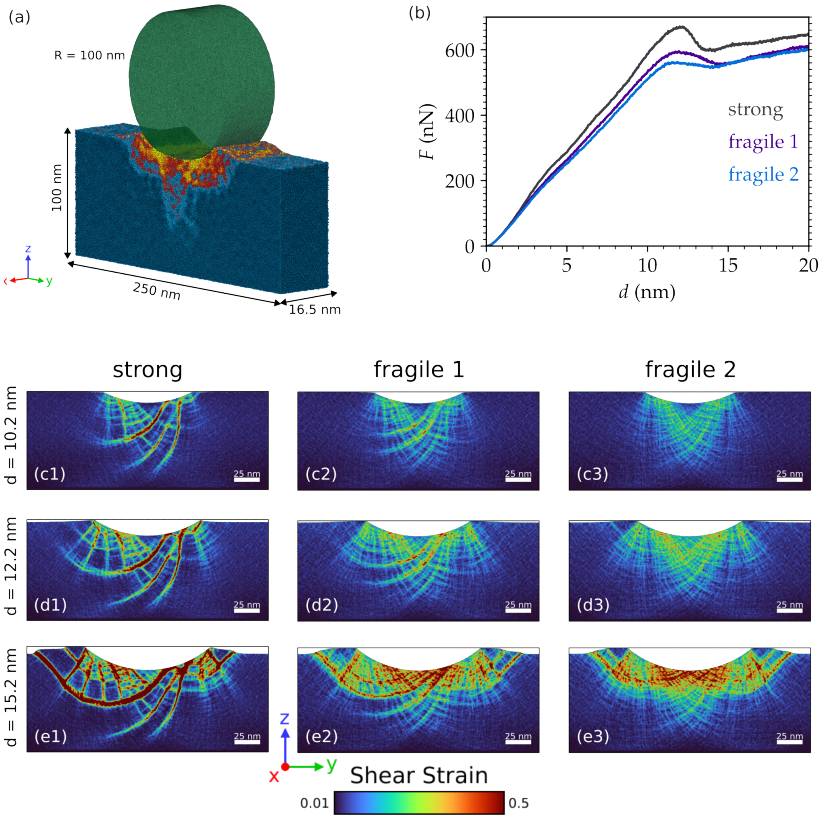


Figure 1. (a) Setup used for the nanoindentation simulations. (b) Force-displacement curves during loading for a glass produced from a strong liquid (quenched from equilibrium at $T_1 = 812$ K) and from two fragile melts (fragile 1, $T_1 = 850$ K and fragile 2, $T_1 = 887$ K), which had all been quenched to $T = 300$ K. (c - e) Local shear strain maps for the strong and fragile glasses loaded to an indentation depth of (c) $d = 10.2$ nm, (d) $d = 12.2$ nm, and (e) $d = 15.2$ nm.

fragile 2. Therefore, a change of $\Delta T_1 = 25$ K near T^* affects the plasticity of our BMG more than a 200 K change above T^*[17].

Not only do the $F(d)$ relations reveal differences, but the plastically deformed structures themselves also disclose clear distinctions between strong and fragile glasses. To reveal these dissimilarities, Fig. 1(c - e) shows the shear strain deduced from atomistic configurations that correspond to three different regimes in the load-displacement curves, that is, just before the maximum in $F(d)$ in Fig. 1(c), at the maximum in Fig. 1(d), and after the force drop in Fig. 1(e). The strong glass simulated here shows visible radially shaped shear bands with an average width of approximately 8 nm. We note that the intrinsic thickness of the mature shear bands in metallic glasses, where the plastic strains are accommodated, has been widely accepted as ≈ 10 nm[18], comparable with our observations for the strong glass. These shear bands are asymmetric in the strong glass case and essentially symmetric for the fragile ones. Comparing the strong glass with the fragile glasses,

there are apparent differences in the shear bands formed in these samples. The deformation is much more localised in the strong glass, less localised, and more symmetric for the fragile glasses. The homogeneity of the shear strain distribution increases when further increasing T_1 above T^*[17].

Some of the detailed features observed during compression in the fragile 1 sample seem to appear also in the strong glass, although the strong melt had been equilibrated several hundred times the energy- and density-autocorrelation times after the fragile melt had been quenched to 300 K. The most plausible explanation of this phenomenon is that the shear-band features are deterministic (if stresses are non-isotropic) despite their erratic appearance. However, when rotating the sample from the slow cooling run by 90° (and/or when shifting it by half the cell dimension) before duplicating it numerous times so that it fills the volume needed for nanoindentation, shear bands look clearly different from case to case. This led us to pursue the hypothesis that there might be some long-lasting "hidden order" in the melt. Thus, to better understand to what degree the shear-band topology is deterministic or affected by random disorder, numerous simulations like the ones depicted in Fig. 1 were repeated (including decompression from varying penetration depth in selected cases). The pertinent data is still being analysed. Recently, we moved from more traditional analysis correlating "traditional measures" in time (density of certain local bonding features in particular octahedral and icosahedral clusters, but also partial densities of individual atom types) to image-based analyses, specifically Contrastive Language–Image Pre-training (CLIP)[19]. It allows the degree of similarity to be evaluated without human bias and without having to construct traditional order parameters that quantify the similarity of shear bands, which seemed to be an infeasible task.

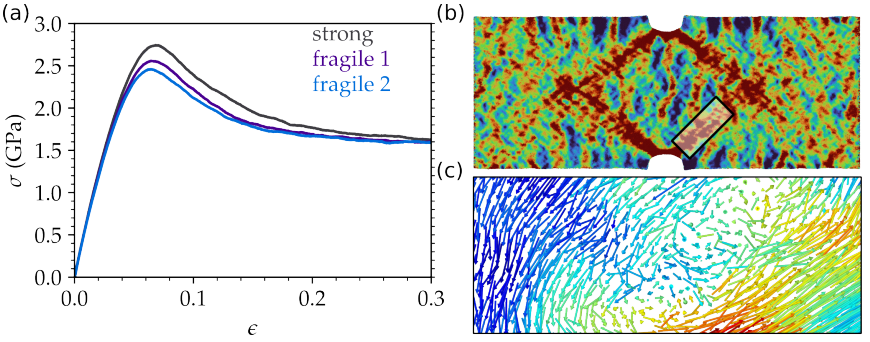


Figure 2. (a) Stress-strain curves $\sigma(\varepsilon)$ for bulk analogues of those studied in Fig. 1. Panel (b) displays the simulation setup for studying shear band interaction and also shows the local shear strain at 30% tensile strain, while panel (c) depicts the local atomic displacements in the rectangular region highlighted in panel (b).

Similar studies, in addition to the just-reported nanoindentation simulations, were conducted to explore the effect of the fragility of the parent metallic liquid on the plasticity in bulk glass samples. The tensile stress-strain curves of strong and fragile glasses are shown in Fig. 2(a) and reveal similar behaviour as seen during the nanoindentation, albeit with a more extended stress-softening regime. The goal of these simulations was to study the interaction of shear bands. To this end, two opposed notches were introduced into a

BMG slab. They will be sources of stress localisation and thus will serve as shear-band initiation sites. Since the shear band will form an angle of approximately 45° with respect to the loading axis, the shear band initiated from the two notches will eventually interact as shown in Fig. 2(b). Fig. 2(c) is a zoom-in of the region highlighted by the black box in Fig. 2(b), which shows the displacement of atoms in a vortex shape. This corroborates the view that the shear band in metallic glasses is made of an alignment of Eshelby-like quadrupoles[20].

3 Atomic-Scale Mechanisms of Friction and Wear of BMGs

In this section, we address the question why BMGs appear to have extremely small internal friction but unfavourable tribological properties like large wear and large sliding friction[21]. The small internal friction makes small BMG spheres bounce off surfaces like rubber balls. Specifically, their coefficient of restitution has been reported to be up to 0.99, making BMG-made golf clubs "the Ferrari of golf clubs world", thus surpassing the 0.86 allowance of the United States Golf Association. This performance is commonly attributed to the lack of dislocations, whose impact-induced motion dissipates much energy. But then, why is their sliding friction high? To address this question, we scratched a BMG surface with an indenter consisting of a perfectly smooth, repulsive indenter, also known as a mathematical wall. Using such a smooth indenter suppresses all explicit (static) friction at the surface, since the indenter is translationally invariant. Coulomb friction can then only arise when sliding induces instabilities inside the solid. The friction-simulation set-up is shown in Fig. 3(a). It is equivalent to a pin-on-disk tribometer due to the use of periodic-boundary conditions.

Fig. 3(a) shows that our BMG "runs in" as regular metal surfaces do, that is, with each pass of the surface, the friction coefficient μ decreases with the first few strides but later levels off, in our case near $\mu = 0.1$. While $\mu = 0.1$ is a relatively small friction coeffi-

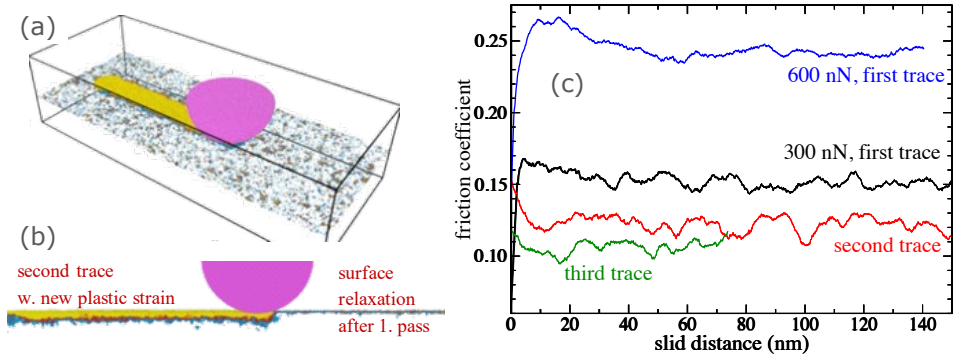


Figure 3. (a) Geometry of the set-up of a "mathematical" tip (purple) scratching over a surface. Bright colours in the BMG substrate indicate large plastic shear strain. (b) Side-view of the sliding process during the second pass at a normal load of $L = 300$ nN, which translates to a mean contact pressure of $p_{33} \approx 300$ MPa, assuming the contact area to be $1,000$ nm^2. (c) Friction coefficient as a function of sliding distance at a normal load of 300 nN (three traces) and 600 nN (one trace).

cient, it must be kept in mind that μ would instantly double if the counter face were also composed of a BMG and supposedly more than triple if the scratching tip were atomically rather than perfectly smooth. Adding realistic (random) roughness to the indenter and/or adhesive interactions would supposedly lead to a further increase in friction making μ clearly exceed that of metals with nanocrystalline coatings. Increasing the load from 300 nN to 600 nN makes the friction more than triple, i.e., μ increases by 65%, while metals with nanocrystalline coatings obey Amontons' law stating linearity between friction and load even at small scales reasonably well[22].

The results on BMG friction can be rationalised to a significant degree by the results on plasticity presented earlier. First, BMG work soften. This is already problematic before effects like unstable friction, increased material transfer or wear set in: Massive, energy-consuming plasticity occurs each time even when the tip scratches over an existing "wear track", in contrast to what would happen on a work-hardening surface. Second, due to plasticity happening in localised zones, the scale dependence of plasticity in BMGs is relatively minor. While metals become stronger at smaller scales (the size of crystallites in a metal increases with distance from the rubbed surface), BMGs appear to have at best a weak scaling of hardness with scale. This makes them softer than crystals at very small scales and thus have relatively large friction, but harder and more elastic on large scales, i.e., on the scales that matter when impinging on an object with radii of curvature in the cm range (as golf balls), while local roughness with local nanometre-scale curvatures determines frictional processes.

4 Atomic-Scale Deformation and Failure Mechanisms of Metallic-Based Auxetic Structure

An interesting type of metamaterials is auxetic metamaterials, which have a specific structure that leads to a negative Poisson's ratio. We devote this section of our work to clarifying the differences in the deformation and failure mechanisms between different auxetics as a function of materials and scale.

Auxetic materials achieve their negative Poisson's ratio ν through the structuring of rigid and soft entities into a network, rather than by the design of a homogeneous material. Their fundamental principle can be rationalised using simple bead-spring models, as evidenced in Fig. 4(a).

This figure superimposes an optical image of a regular alloy provided by the group of Prof. Dr. Ing. Stefan Diebels at Saarland University with a model we produced: Atoms are placed on a triangular lattice, with harmonic springs connecting two adjacent atoms such that the springs are relaxed at the ideal nearest-neighbour distance. All atoms are then deleted, resulting in a void where the real object would be.

By design, this bead-spring model falls into the small-strain, large-displacement category. Since the unit cell of the undeformed solid has sharp corners, stress singularities arise at these corners, reducing the convergence radius of the linear elasticity of the unit cell to zero. This condition leads to (a) the need to account for large displacements even under small forces, and (b) ν depending on the loading condition, even at low external forces.

One benefit of auxetic materials is their ability to withstand large strain without failing. That limit cannot be probed using an elastic system, which, nonetheless yields a similar $\nu(\varepsilon_{xx})$ dependency as "real materials" do, see Fig. 4(h). To contrast the performance

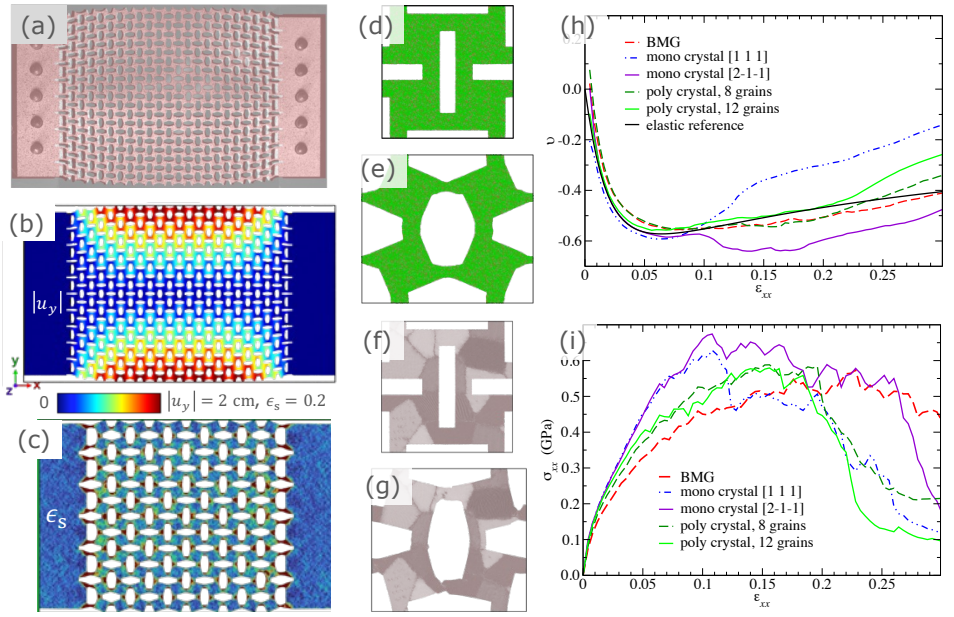


Figure 4. (a) Superimposition of the auxetic structure obtained experimentally (courtesy of Prof. Stefan Diebels, Saarland University) with that of a bead-spring model. (b) Absolute displacement in y-direction $|u_y|$ of the elastic reference and (c) von Mises shear-strain ϵ_s in the small-scale BMG auxetic material. Unit cells of a BMG (d,e) and the 12-grain aluminum poly crystal (f,g) at zero strain (d,f) as well as at 30% (e) and at 20% (g) strain, the latter being just before failure. (h) Poisson's ratio ν and (i) longitudinal stress σ_{xx} both as a function of the strain ε_{xx}.

of BMGs to crystalline materials, the same relaxed configuration as that used in section 2-1 was recycled again. Specifically, the strong-glass sample produced in Ref. 16 was duplicated numerous times. After applying the mask and relaxing the structure again, it was strained in x direction, while having open boundary conditions along y and being periodic parallel to z. In addition, we simulated (hypothetical) single-crystalline aluminum with two different surfaces parallel to the (x-) direction of tensile loading, i.e., [111] and [$2\bar{1}\bar{1}$] as well as poly-crystalline aluminum with either 8 or 12 crystallites per unit cell. An embedded atom potential[23], which is frequently used for the simulation of mono-atomic systems, was employed to that end.

The overall trend for the $\nu(\varepsilon_{xx})$ matches that of experiments and bead-spring model in all cases, i.e., with a Poisson ratio of $\nu \gtrsim -0.6$ at 5% strain. However, all-atom simulations produce a slightly positive Poisson's ratio at small strain, in contrast to the elastic model, as revealed in Fig. 4(h). This is supposedly because the stress singularities at sharp edges induce plasticity in the *in silico* samples, which in turn suppresses large-displacement effects causing the studied auxetic material to have a negative ν. We will redo the very initial parts of these curves (probably with cheaper minimisation rather than molecular dynamics) to further explore the origin of the positive ν at small ε_{xx}.

Although initially the softest of all analysed materials, see Fig. 4(i), the BMG clearly outperforms the tensile loading abilities of all other crystalline structures, because it is

the only sample that has not yet yielded at 20% strain. The microscopic configuration at that strain is shown in Fig. 4(e), while that of the most readily failing sample, i.e., polycrystalline aluminum, is depicted near 20% strain in panel (g), which is just before its sudden drop in stress.

5 Concluding Remarks

In conclusion, we discovered that the plasticity of bulk metallic glasses (BMGs) is significantly influenced by the extent to which the melt was out of equilibrium when passing through the fragile-to-strong transition temperature, T^*. Specifically, when a liquid was quenched from a temperature T_q just a few Kelvin above T^*, the sample did not exhibit mature shear bands during nano-indentation. In fact, nano-indentation curves for all $T_\mathrm{q} > T^*$ displayed similar characteristics, including those not explicitly shown in this article. However, pronounced shear bands formed once T_q fell a few Kelvin below T^*. While this observation aligns with experimental findings regarding lower cooling rates, we revealed that yield strength and brittleness increase quasi-discontinuously when T_1 is below T^*, whereas changes with T_1 are small and smooth for $T_1 > T^*$. We speculate that other glass formers may exhibit similar behaviour; however, BMGs appear to be the only broader class of glass formers that allow for the liquid-to-strong transition to be analysed both experimentally and through simulations, while also offering the possibility to design compositions that can be frozen into either a fragile or a strong melt. More details can be found in Ref. 24.

Furthermore, we revealed that a bulk metallic glass (BMG) sample in contact with a smooth, sliding counterbody undergoes repeated plastic deformation when scratched multiple times. We attribute this behaviour to strain-softening. Interestingly, when the load is doubled, the friction coefficient also increased by a factor of ≈ 1.7, indicating that the friction itself triples. This behaviour presents a clear violation of Amontons's law, which states that friction is proportional to load. However, this phenomenon at the microscopic scale does not necessarily lead to the breakdown of Amontons's law at the macroscopic scale, even if this friction mechanism is the dominant one. Thus, while (strain-softening) BMGs may not be suitable for low-friction coatings, we found them to be highly appropriate for use in auxetic materials. Their strain-stress curves were closer to those of ideally elastic references than any (poly-) crystalline solids studied, and they also yielded at significantly larger strains than any other simulated materials molded into the same auxetic shape. Details on these two threads of work are still in the process of being written up in greater detail.

Acknowledgements

The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (`www.gauss-centre.eu`) for funding this project by providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS Supercomputer JUWELS at Jülich Supercomputing Centre (JSC).

References

1. A. Inoue, B. Shen, H. Koshiba, H. Kato, and A. R. Yavari, *Cobalt-based bulk glassy alloy with ultrahigh strength and soft magnetic properties*, Nature Materials, **2**, no. 10, 661-663, Sep. 2003.
2. R. O. Ritchie, *The conflicts between strength and toughness*, Nature Materials, **10**, no. 11, 817-822, Oct. 2011.
3. J. J. Kruzic, *Bulk Metallic Glasses as Structural Materials: A Review*, Advanced Engineering Materials, **18**, no. 8, 1308-1331, May 2016.
4. H. A. Bruck, T. Christman, A. J. Rosakis, and W. L. Johnson, *Quasi-static constitutive behavior of $Zr_{41.25}Ti_{13.75}Ni_{10}Cu_{12.5}Be_{22.5}$ bulk amorphous alloys*, Scripta Metallurgica et Materialia, **30**, no. 4, 429-434, Feb. 1994.
5. X. J. Gu, A. G. McDermott, S. J. Poon, and G. J. Shiflet, *Critical Poisson's ratio for plasticity in Fe–Mo–C–B–Ln bulk amorphous steel*, Applied Physics Letters, **88**, no. 21, 211905, May 2006.
6. Y. Wu, D. Cao, Y. Yao, G. Zhang, J. Wang, L. Liu, F. Li, H. Fan, X. Liu, H. Wang, X. Wang, H. Zhu, S. Jiang, P. Kontis, D. Raabe, B. Gault, and Z. Lu, *Substantially enhanced plasticity of bulk metallic glasses by densifying local atomic packing*, Nature Communications, **12**, no. 1, 6582, Nov 2021.
7. X. Mu, M. R. Chellali, E. Boltynjuk, D. Gunderov, R. Z. Valiev, H. Hahn, C. Kübel, Y. Ivanisenko, and L. Velasco, *Unveiling the Local Atomic Arrangements in the Shear Band Regions of Metallic Glass*, Advanced Materials, **33**, no. 12, 2007267, Feb. 2021.
8. Z. Y. Liu, Y. Yang, and C. T. Liu, *Yielding and shear banding of metallic glasses*, Acta Materialia, **61**, no. 16, 5928-5936, Sept. 2013.
9. Y. Yokoyama and A. Inoue, *Compositional Dependence of Thermal and Mechanical Properties of Quaternary Zr-Cu-Ni-Al Bulk Glassy Alloys*, Materials Transactions, **48**, no. 6, 1282-1287, 2007.
10. C. L. Qin, W. Zhang, Q. S. Zhang, K. Asami, and A. Inoue, *Chemical characteristics of the passive surface films formed on newly developed Cu–Zr–Ag–Al bulk metallic glasses*, Journal of Materials Research, **23**, no. 8, 2091-2098, Aug. 2008.
11. K. F. Kelton, *A perspective on metallic liquids and glasses*, Journal of Applied Physics, **134**, no. 1, July 2023.
12. P. Lucas, *Fragile-to-strong transitions in glass forming liquids*, Journal of Non-Crystalline Solids: X, **4**, 100034, Dec. 2019.
13. X. Ren, R. Das, P. Tran, T. D. Ngo, and Y. M. Xie, *Auxetic metamaterials and structures: a review*, Smart Materials and Structures, **27**, no. 2, 023001, Jan. 2018.
14. M. Stolpe, I. Jonas, S. Wei, Z. Evenson, W. Hembree, F. Yang, A. Meyer, and R. Busch, *Structural changes during a liquid-liquid transition in the deeply undercooled* $Zr_{58.5}Cu_{15.6}Ni_{12.8}Al_{10.3}Nb_{2.8}$ *bulk metallic glass forming melt*, Phys. Rev. B, **93**, 014201, Jan. 2016.
15. Y. Q. Cheng, E. Ma, and H. W. Sheng, *Atomic level structure in multicomponent bulk metallic glass*, Phys. Rev. Lett., **102**, no. 24, 245501, June 2009.
16. S. V. Sukhomlinov and M. H. Müser, *Quasidiscontinuous change of the density correlation length at the fragile-to-strong transition in a bulk-metallic-glass forming melt*, Phys. Rev. Mater., **2**, 115604, Nov. 2018.

17. A. Atila, S. V. Sukhomlinov, M. J. Honecker, and M. H. Müser, *Brittleness of metallic glasses dictated by their state at the fragile-to-strong transition temperature*, 2024, arXiv:2408.00536.
18. Y. Zhang and A. L. Greer, *Thickness of shear bands in metallic glasses*, Applied Physics Letters, **89**, no. 7, 071907, 08 2006.
19. A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever, *Learning Transferable Visual Models From Natural Language Supervision*, 2021, arXiv:2103.00020.
20. D. Şopu, *STZ-Vortex model: The key to understand STZ percolation and shear banding in metallic glasses*, Journal of Alloys and Compounds, **960**, 170585, Oct. 2023.
21. N. W. Khun, H. Yu, Z. Z. Chong, P. Tian, Y. Tian, S. B. Tor, and E. Liu, *Mechanical and tribological properties of Zr-based bulk metallic glass for sports applications*, Materials &; Design, **92**, 667–673, Feb. 2016.
22. M. Chandross and N. Argibay, *Friction of Metals: A Review of Microstructural Evolution and Nanoscale Phenomena in Shearing Contacts*, Tribology Letters, **69**, no. 4, Aug. 2021.
23. Y. Mishin, D. Farkas, M. J. Mehl, and D. A. Papaconstantopoulos, *Interatomic potentials for monoatomic metals from experimental data and ab initio calculations*, Physical Review B, **59**, no. 5, 3393–3407, Feb. 1999.
24. A. Atila, M. Kbirou, S. Ouaskit, and A. Hasnaoui, *On the presence of nanoscale heterogeneity in $Al_{70}Ni_{15}Co_{15}$ metallic glass under pressure*, Journal of Non-Crystalline Solids, **550**, 120381, Dec. 2020.

Earth and Environment

Earth and Environment

Patrick Jöckel

Deutsches Zentrum für Luft- und Raumfahrt (DLR) e.V., Institut für Physik der Atmosphäre,
Münchner Str. 20, Oberfaffenhofen, 82234 Wessling, Germany
E-mail: patrick.joeckel@dlr.de

Ever since Jule Charney, Ragnar Fjørtoft, and John von Neumann performed the first weather forecast on an electronic computer almost three quarters of a century ago (in 1950), numerical weather prediction was, and still is, at the forefront of scientific disciplines utilising the largest and powerful electronic computers. Over time, numerical simulations became the important third pillar, besides experiment / observation and theory, in almost all disciplines of Earth system sciences. The transition from classical central processing units (CPUs) to hybrid systems utilising accelerators, such as graphics processing units (GPUs), is still ongoing and has now reached the exascale era. For this, many (large) source codes have to be re-factored and even redesigned with new algorithms as a prerequisite to allow for the full exploitation of the increasing computational power. Today, the availability of large high-performance computing (HPC) systems, as those operated and made available for public research by NIC, is an essential prerequisite for scientific progress in all Earth system sciences, in particular given that society is facing large challenges caused by the rapid change of climate and environment. A common challenge in all Earth system sciences, when it comes to numerical simulation of real systems, is the need to represent a wide range of spatial and temporal scales, because processes on the different scales are non-linearly coupled, and thus interact with each other. Even on exascale systems, this challenge cannot be entirely overcome - for a multitude of reasons. In many cases, grid resolution, i.e. the ability to simulate finer scales directly, has to be sacrificed for longer simulated time spans, for larger ensemble sizes, or for additionally added, computationally expensive processes. In consequence, additional methods are required, with which small scale processes in coarser resolved model simulations can be taken into account.

Bastian Waldowski and colleagues examine, how the reliability of water availability forecasts with numerical models, e.g. for water resources management, can be improved. In their case, the model grid resolution needs to be reduced for the sake of a sufficiently large ensemble size of perturbed simulations. The calculation of the forecasts, however, relies on the assimilation of observed data. This data assimilation is used for an objective state estimation by optimally combining the model state with observations. Applied on a coarser grid, it introduces biases, in this case because the topographical information is smoothed on a coarsened grid. Methods exist to mitigate or compensate such biases, and one of these methods is assessed by the authors.

In numerical weather prediction and for regional climate projections, from both of which local information is desired, the classical method of statistical downscaling has been established. This method utilises the statistical relationships (e.g., based on historical data) between the large scale situation (weather pattern) and the local weather phenomena (e.g. local precipitation). Ankit Patnala and colleagues examine a new, machine learning based

method, for its applicability for such statistical downscaling. Such methods are based on the assumption that the trained neural network contains the knowledge about the relationships between small scale and large scale information contents. The authors further assess the applicability of the same method for data compression, with which fine scale weather information is reconstructed by the neural network from artificially reduced or coarsened data.

Last, but not least, Bernd Schalge and colleagues address advanced concepts for the generation of reanalysis data, focussing on the hydrological cycle between land-surface and atmosphere on the continental scale. Reanalysis is a method to objectively combine observational data by means of data assimilation (i.e., statistical methods) into a numerical model, with the aim to provide an internally consistent (long-term) historical time-series of domain-covering (regional or global) gridded data about the state of the considered system. As such, reanalysis data provide the most complete picture currently possible of the past states in the Earth system. The authors examine a new data assimilation technique and a combined approach for atmosphere and land-surface, which also takes into account anthropogenic water use through irrigation.

All three examples show how the utilisation of modern HPC systems can boost our scientific understanding of the Earth system. But at the same time, they also show that further challenges need to be tackled and that the exascale era has just begun – with a multitude of new possibilities for numerical Earth system sciences.

The Role of Biases Due to Coarse Resolution for Data Assimilation in Terrestrial Systems

Bastian Waldowski[1], **Harrie-Jan Hendricks-Franssen**[2,3], **Johannes Keller**[2,3], **and Insa Neuweiler**[1]

[1] Institute of Fluid Mechanics and Environmental Physics in Civil Engineering, Leibniz University Hannover, Germany
E-mail: {waldowski, neuweiler}@hydromech.uni-hannover.de

[2] Institute of Bio- and Geosciences- Agrosphere (IBG-3), Forschungszentrum Jülich, 52425 Jülich, Germany
E-mail: {h.hendricks-franssen, jo.keller}@fz-juelich.de

[3] Centre for High Performance Scientific Computing in Terrestrial Systems (HPSC TerrSys) (Geoverbund ABC/J), 52425 Jülich, Germany

Accurate forecasts of water availability, both in the upper soil layers as well as stored in aquifers, are needed for water resources management and other purposes. Soil moisture and groundwater tables can be predicted by a fully coupled numerical model that represents land surface and subsurface processes in a physically based manner. The forecasts of these models are tied to many uncertainties (such as e.g. uncertain soil properties). Data assimilation (DA) is a tool to reduce such uncertainties by utilising information from measurements, such as groundwater heads. This requires multiple model runs of an expensive forward model, so the numerical grid (i.e. the resolution) of the model often needs to be chosen coarser than required by model quality criteria. This coarsening introduces a bias. Such biases lead to a violation of core assumptions of the most common DA approaches. There are several methods to mitigate such biases. In this work, we highlight two DA experiments. One where we compensated the coarsening bias preemptively and one where we did not. We find that compensating the coarsening bias allows DA to improve the root zone soil moisture forecast. In the experiment without compensation, we see that the predictions using DA deteriorate more extensively and severely the predictions at locations where no information from measurements was utilised. We thus find that preemptively compensating for the coarsening bias increases the spatial extent in which DA improves the forecast and mitigates deteriorations by DA.

1 Introduction

To evaluate water availability and to assess risks, like flooding or drought, a reliable forecast of the hydrologic state of the terrestrial system is needed. Such forecasts predict water availability in soils, aquifers, and rivers, for example. To a certain extent, water availability can be measured. However, measurement capabilities are generally limited, and physically based numerical models of water flow and storage are often used to generate additional information. Such models make forecasts based on equations derived from physical relationships. Making such forecasts is complicated by the interconnected nature of the hydrologic cycle. Different compartments such as groundwater, the vadose zone (the subsurface above the groundwater), and rivers generally influence each other in a two-directional way. For instance, a higher water level in a river can cause the groundwater table to rise, which then increases flow from the groundwater into the river at a later time, which then again influences the groundwater, and so on and so forth. Such two-way feedback is often not

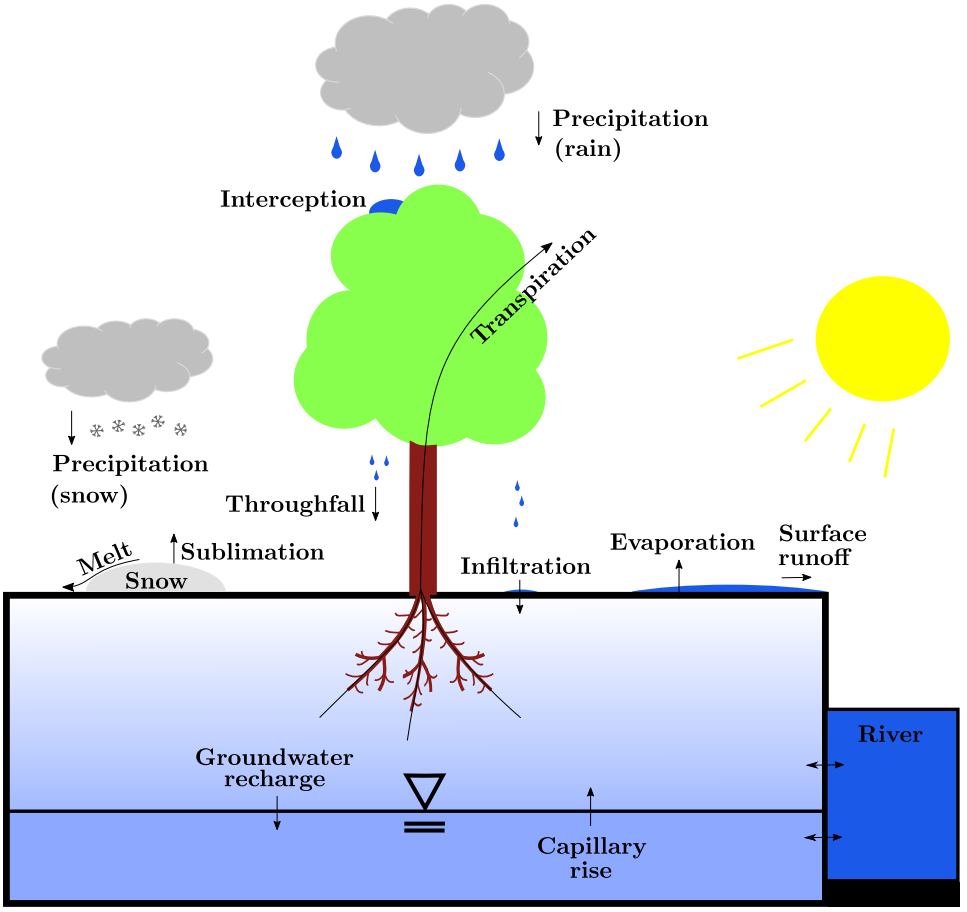


Figure 1. Illustration of the decisive processes represented by a fully coupled numerical model, such as the one used in this project (TSMP with ParFlow-CLM).

represented well if the compartments are considered separate. Fully coupled numerical models are designed to account for such two-way feedback and simulate the hydrologic system in an integrated manner, which covers the processes shown in Fig. 1. They can thus generate a lot of information about hydrological states in addition to measurements and give important insights into the interrelations between the different compartments.

However, this comes at a cost of a rather high computational demand. Fully coupled models also require a lot of information that is difficult to acquire accurately. For example, detailed information about highly conductive flow paths within a soil is usually not available. The inputs of fully coupled models (such as soil properties or precipitation rates and patterns) are highly uncertain, which can compromise the quality of the forecast of these models notably. Data assimilation (DA) uses measurements of hydrologic states (such as soil moisture or the depth to groundwater table) to reduce this uncertainty. Often, a stochastic framework is applied. However, this means that the uncertainty of the numerical model

needs to be represented. Often this is done using Monte Carlo methods, where sampling is done with an ensemble of model simulations (multiple simulations with different inputs), which needs to be large enough to be a representative sample. This means that many evaluations of a very computationally expensive model are needed, which can even exceed the computing capacity of a supercomputer. Fully coupled numerical models approximate the fluxes and states of the hydrologic cycle on a numerical grid, with solutions only at discrete locations. The finer the grid, the more computationally expensive the model. Such models are thus generally run on grid resolutions that are lower than required for adequate process representation. This leads to a smoothing of parameters and topography, which usually introduces a bias into the model predictions. Biases are generally problematic in the context of data assimilation, as data assimilation approaches usually assume unbiased errors.

Within this work we look at a test example to i.) assess the role of the coarsening bias for data assimilation and ii.) test preemptively compensating this bias by adjusting parameters to mitigate coarsening effects. To directly tackle this issue, using measurements from a real catchment would be not suitable. Using real measurements, many more biases would be present (e.g. due to simplifying assumptions of the numerical model), which would make clear conclusions on the effects of the grid coarsening biases very difficult. We will thus take measurements from a virtual reality (VR) reference model that has a higher resolution of the numerical grid than the ensemble used for DA. The test case that is used within this project is artificial but realistic.

2 Test Case

For modelling water fluxes in the system, we use TSMP-PDAF[1–4], coupling[5,6] the fully integrated subsurface/surface flow model ParFlow[7–9] with the land surface model CLM[10] and the Parallel Data Assimilation Framework (PDAF). Parallelisation of all forward codes

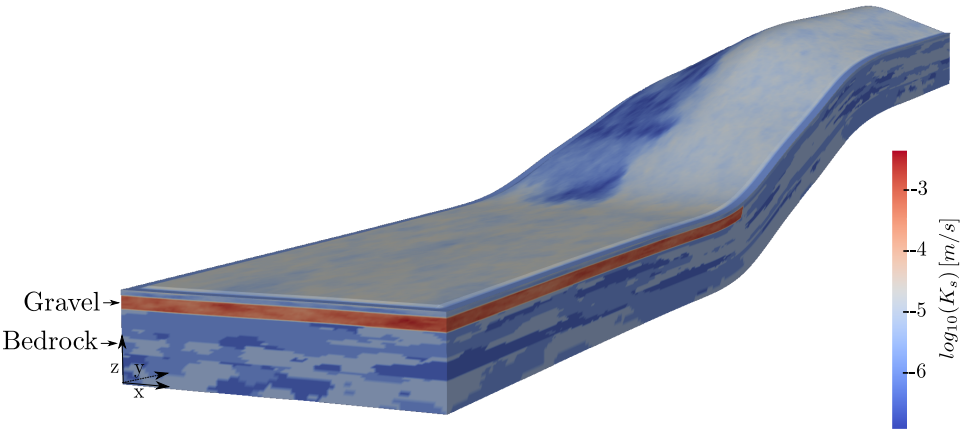


Figure 2. Domain of our test case. The z-axis is scaled by a factor of 10 in this plot. The colour scheme illustrates the hydraulic conductivity of the soil from red (high permeability) to blue (low permeability).

for TSMP-PDAF is implemented in pure MPI. PDAF is especially well suited for HPC environments, as it does not write or read files during the assimilation process. Furthermore, it has been tested and run extensively on JUWELS in the past. The model domain covers a rectangular area of $1\,km$ x $5\,km$, with a uniform depth of $50\,m$. It is characterised by a flood plain with an adjacent hillslope, enclosed by three rivers (see Fig. 2). The hill is tilted in the x-direction, meaning that the river along the "right" side (referring to Fig. 2) is at a higher elevation than the river along the "left" side. This causes an offset between the water divide and the crest of the hill. The water divide is moved more towards the "right" boundary, which can cause lateral fluxes against the direction of the topography. While such processes occur in natural landscapes, they are usually neglected, as water divides are generally directly derived from the topography.

The grid has a horizontal resolution of $10\,m$ (very high resolution in the context of fully coupled modelling) and vertical layers of variable thickness between $2\,m$ and $2\,cm$. The subsurface material is divided into topsoil, gravel, bedrock, and riverbed units. Each of these units is characterised by unique spatially heterogeneous parameter fields. At the land surface, different types of vegetation are considered in different parts of the domain. Atmospheric forcings are taken from a previous simulation of the Neckar catchment[11].

2.1 Specifications for Data Assimilation

We use the Localised Ensemble Kalman Filter[12,13] with 93 ensemble members. Each ensemble member has a horizontal resolution of $\Delta x, y = 40\,m$, which reduces the number of cells by a factor of 16 compared to the virtual reality (reducing the order of magnitude from 10^6 to 10^5). Each ensemble member runs on 40 cores, 36 of them used by ParFlow and 4 by CLM. This means that each ensemble simulation uses 3720 cores. The highly resolved VR used 228 cores and was notably slower than the coarse models. In the coarse models used for DA, the highly resolved properties of the VR are smoothed. From the

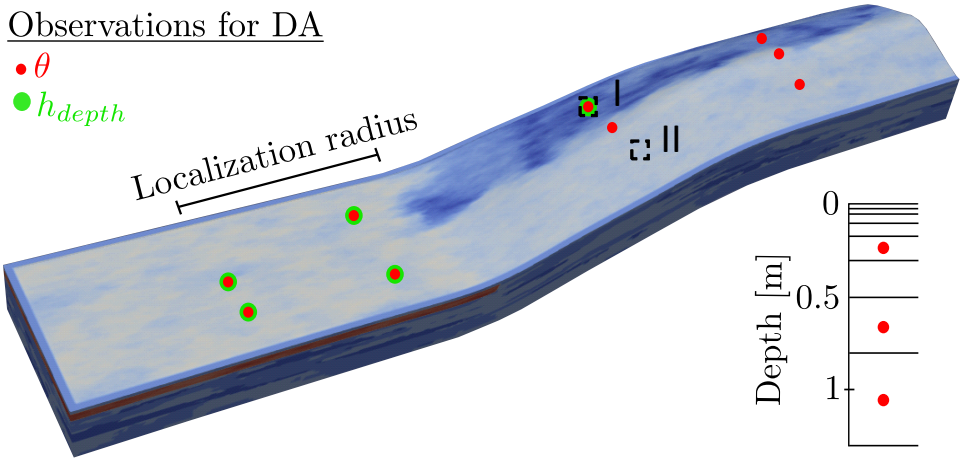


Figure 3. Setup for DA experiments. Observation locations are marked in red for soil moisture and green for groundwater table depth. Locations I and II will be referred to in Fig. 4. In the bottom right, measurement depths for soil moisture are shown.

hillslope to the continental scale, topographical smoothing is often found to result in reduced drainage out of the system[9,14–18]. In our test case, this topographical smoothing bias leads to a notably higher groundwater table at the hill compared to that of the VR. As a compensation strategy, we increase the lateral hydraulic conductivity in the cells of the hill, which allows for more efficient drainage. This is similar to other approaches[17–19] that have been used in the past. However, the impact of such compensation on data assimilation has not been studied before. We conduct DA experiments with two different ensembles. One where we compensate topographical smoothing preemptively and one where we do not. Fig. 3 shows the locations where observations were taken for DA. Soil moisture measurements are taken from three different depths and nine horizontal locations (marked in red) and groundwater table measurements are used at five different locations (green in the figure). Locations I and II are labelled for later analysis. Both are located on the hill, as this is where the topographical smoothing bias appears and where compensation has an effect. At location I, both soil moisture and groundwater table measurements are used for DA, so the direct impacts of DA can be seen. Location II is used for validation, which is particularly challenging at that location, as the majority of observation locations on the hill are on the other side of the water divide. In addition to the states (soil moisture and groundwater tables), we also change the soil hydraulic parameters by assimilation, as they have a strong influence on the prediction uncertainty. The states are updated daily and the parameters weekly with a damping factor[20] of 0.1. Localisation, which limits updates of DA to a certain area to counteract "wrong" updates due to spurious correlations, is applied for each observation within the radius indicated in Fig. 3.

3 Results

We define an error of our coarsened simulations with and without DA by comparing results of root zone soil moisture θ_{rz} (relevant for evapotranspiration) and groundwater table depth h_{depth} with the VR. Comparing errors made in a simulation without DA to those made in a simulation with DA allows a conclusion on how much DA improved (or deteriorated) the

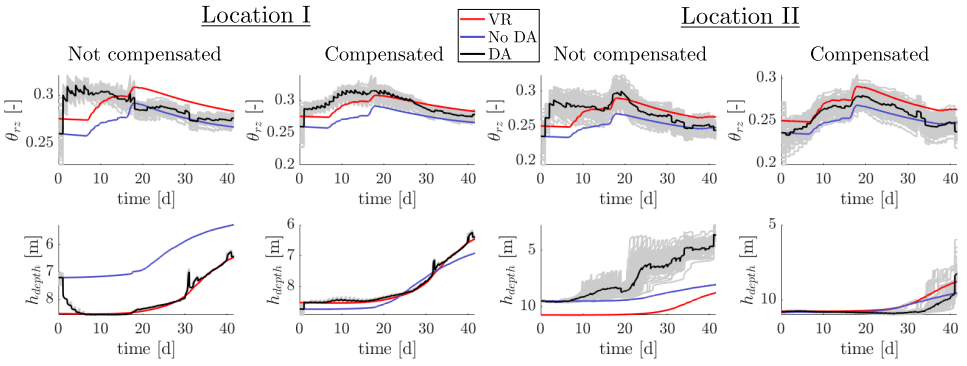


Figure 4. Root zone soil moisture and groundwater table depth over time at locations I and II for both DA experiments. The VR is shown in red, the ensemble without DA in blue, and the DA experiments are in black, with values for each ensemble member shown in grey.

respective prediction. In this work, this is quantified with δ, which expresses the difference between the prediction error of a simulation without DA and a simulation with DA. Positive values of δ directly indicate an improvement by DA. For the compensated experiments, both the simulation without DA and with DA are compensated, such that δ only captures improvement by DA itself.

Fig. 4 shows model predictions of the root zone soil moisture and groundwater table depth at locations I and II, which were shown in Fig. 3, for both DA experiments. The DA experiments are plotted in black, red is the VR, and blue is the reference simulation without DA. At location I, the prediction of h_{depth} can be notably improved by DA, both in the compensated experiment, as well as in the experiment without compensation. For the root zone soil moisture, the DA experiment without compensation does not make a good forecast. The deviations between predictions and observations are similar to those of the simulation without DA and the temporal evolution is poorly met. If compensation is applied, the root zone soil moisture at location I is much closer to the truth (VR). At the other side of the hill at location II, trends for root zone soil moisture look similar to location I, with the compensated simulations being better able to reproduce the temporal evolution of the truth. For the groundwater table depth at location II, the forecast of the DA experiment without compensation is not well matched and worsens over time. For the compensated DA experiment, the forecast is also worse than that of the simulation without DA, but in general much better than for the DA experiment without compensation. It should also be noted that the simulation of the compensated ensemble without DA (blue line) is very close to the truth already at that location, which makes it hard to improve.

To get further insights into the impacts of DA all over the domain, spatial distributions of the time-averaged δ for the root zone soil moisture and the groundwater table depth are shown for both the DA experiment without compensation (see Fig. 5) and with compensation (see Fig. 6). Fig. 5 shows that DA without compensation deteriorates forecasts at the hill, especially for non-observed locations. Comparing these results to Fig. 6, it can be seen that the compensation can avoid and mitigate much of the deterioration that is present in the DA experiment without compensation. DA works better with the compensated model.

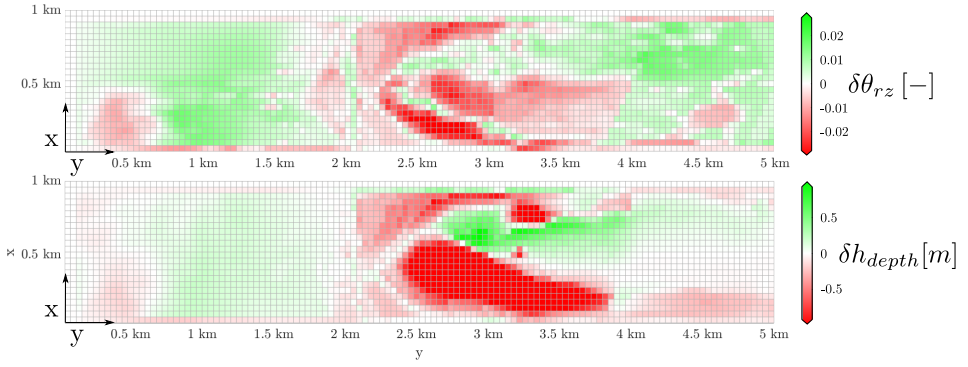


Figure 5. Temporal average of the improvement of root zone soil moisture and groundwater table depth characterisation for the DA experiment without compensation.

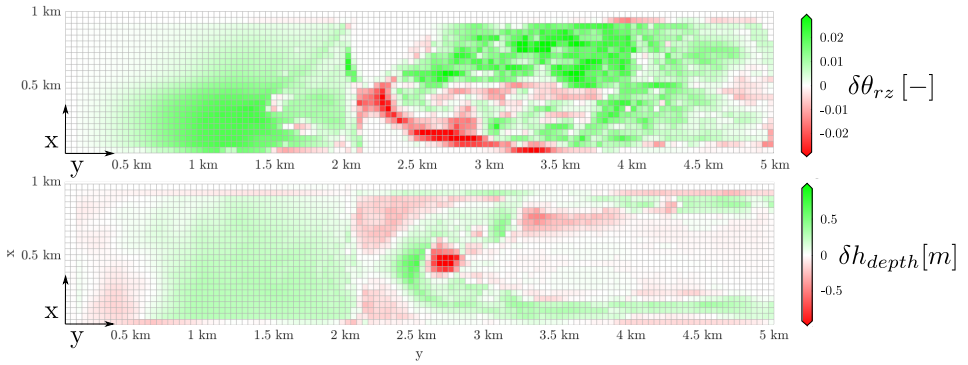


Figure 6. Temporal average of the improvement of root zone soil moisture and groundwater table depth characterisation for the DA experiment with compensation.

4 Concluding Remarks

Within this work, we have investigated the effects of biases due to topographical smoothing on DA forecasts with fully coupled numerical models. We did that with a highly resolved topographically challenging synthetic test case. In general, DA was able to improve forecasts at locations where measurements were available, but spatial validation was poor if the model was biased. We have found that preemptive compensation of the bias due to topographical smoothing yields great potential for improving forecasts at locations where no measurements are available. As DA experiments with fully coupled numerical models, such as TSMP used in this work, are generally conducted on too-coarse grids without any compensation for the topographical smoothing, this is an important insight. Still, we are aware that this is only one step, as there are many other sources of biases between the numerical simulations and real measurements that can also cause problems when conducting DA experiments with such models.

Acknowledgements

Part of this research was funded by the Deutsche Forschungsgemeinschaft (DFG, FOR2131: "Data Assimilation for Improved Characterization of Fluxes across Compartmental Interfaces", grants: NE 824/12-2 and HE 6239/4-2). The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (`www.gauss-centre.eu`) for funding this project by providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS Supercomputer JUWELS at Jülich Supercomputing Centre (JSC).

References

1. L. Nerger and W. Hiller, *Software for ensemble-based data assimilation systems – Implementation strategies and scalability*, Computers & Geosciences, **55**, 110-118, 2013.

2. P. Shrestha, M. Sulis, M. Masbou, S. Kollet, and C. Simmer, *A scale-consistent terrestrial systems modeling platform based on COSMO, CLM, and ParFlow*, Monthly weather review, **142**, no. 9, 3466-3483, 2014.
3. F. Gasper, K. Gorgen, P. Shrestha, M. Sulis, J. Rihani, M. Geimer, and S. Kollet, *Implementation and scaling of the fully coupled Terrestrial Systems Modeling Platform (TerrSysMP v1. 0) in a massively parallel supercomputing environment – a case study on JUQUEEN (IBM Blue GeneQ)*, Geosci. Model Dev, **7**, no. 5, 2531-2543, 2014.
4. W. Kurtz, G. He, S. J. Kollet, R. M. Maxwell, H. Vereecken, and H.-J. Hendricks Franssen, *TerrSysMP-PDAF (version 1.0): a modular high-performance data assimilation framework for an integrated land surface-subsurface model*, Geoscientific Model Development, **9**, no. 4, 1341-1360, 2016.
5. S. J. Kollet and R. M. Maxwell, *Capturing the influence of groundwater dynamics on land surface processes using an integrated, distributed watershed model*, Water Resources Research, **44**, no. 2, 2008.
6. S. Valcke, *The OASIS3 coupler: A European climate modelling community software*, Geoscientific Model Development, **6**, no. 2, 373-388, 2013.
7. J. E. Jones and C. S. Woodward, *Newton-Krylov-multigrid solvers for large-scale, highly heterogeneous, variably saturated flow problems*, Advances in Water Resources, **24**, no. 7, 763-774, 2001.
8. S. J. Kollet and R. M. Maxwell, *Integrated surface-groundwater flow modeling: A free-surface overland flow boundary condition in a parallel groundwater flow model*, Advances in Water Resources, **29**, no. 7, 945-958, 2006.
9. R. M. Maxwell, *A terrain-following grid transform and preconditioner for parallel, large-scale, integrated hydrologic modeling*, Advances in Water Resources, **53**, 109-117, 2013.
10. K. W. Oleson, Y. Dai, G. Bonan, M. Bosilovich, R. Dickinson, P. Dirmeyer, F. Hoffman, P. Houser, S. Levis, G.-Y. Niu et al., *Technical description of the community land model (CLM)*, Tech. Note NCARTN-461+ STR, 2004.
11. B. Schalge, J. Rihani, G. Baroni, D. Erdal, G. Geppert, V. Haefliger, B. Haese, P. Saavedra, I. Neuweiler, H.-J. Hendricks Franssen et al., *High-resolution virtual catchment simulations of the subsurface-land surface-atmosphere system*, Hydrology and Earth System Sciences Discussions, 1-44, 2016.
12. G. Evensen, *Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics*, Journal of Geophysical Research: Oceans, **99**, no. C5, 10143-10162, 1994.
13. P. L. Houtekamer and H. L. Mitchell, *Data assimilation using an ensemble Kalman filter technique*, Monthly weather review, **126**, no. 3, 796-811, 1998.
14. R. N. Armstrong and L. W. Martz, *Topographic parameterization in continental hydrology: a study in scale*, Hydrological Processes, **17**, no. 18, 3763-3781, 2003.
15. P. Shrestha, M. Sulis, C. Simmer, and S. Kollet, *Impacts of grid resolution on surface energy fluxes simulated with an integrated surface-groundwater flow model*, Hydrology and Earth System Sciences, **19**, no. 10, 4317-4326, 2015.
16. P. Shrestha, M. Sulis, C. Simmer, and S. Kollet, *Effects of horizontal grid resolution on evapotranspiration partitioning using TerrSysMP*, Journal of hydrology, **557**, 910-915, 2018.

17. W.-L. Kuo, T. S. Steenhuis, C. E. McCulloch, C. L. Mohler, D. A. Weinstein, S. D. DeGloria, and D. P. Swaney, *Effect of grid size on runoff and soil moisture for a variable-source-area hydrology model*, Water Resources Research, **35**, no. 11, 3419-3428, 1999.
18. M. Niedda, *Upscaling hydraulic conductivity by means of entropy of terrain curvature representation*, Water Resources Research, **40**, no. 4, 2004.
19. L. M. Foster and R. M. Maxwell, *Sensitivity analysis of hydraulic conductivity and Manning's n parameters lead to new method to scale effective hydraulic conductivity across model resolutions*, Hydrological Processes, **33**, no. 3, 332-349, 2019.
20. H.-J. Hendricks Franssen and W. Kinzelbach, *Real-time groundwater flow modeling with the ensemble Kalman filter: Joint estimation of states and parameters and the filter inbreeding problem*, Water Resources Research, **44**, no. 9, 2008.

Applying AtmoRep for Diverse Weather Applications

Ankit Patnala[1], Belkis Asma Semcheddine[1], Michael Langguth[1], Martin G. Schultz[1,2], Christian Lessig[3], and Ilaria Luise[4]

[1] Jülich Supercomputing Centre, Forschungszentrum Jülich, 52428 Jülich, Germany
E-mail: {a.patnala, a.semcheddine, m.langguth, m.schultz}@fz-juelich.de

[2] University of Cologne, Department of Computer Science, Cologne, Germany

[3] European Centre for Medium-Range Weather Forecasts, 53175 Bonn, Germany
E-mail: christian.lessig@ecmwf.int

[4] CERN, Geneva, Switzerland
E-mail: ilaria.luise@cern.ch

Machine learning has recently seen a rapid wide-spread adoption across various fields of science including atmospheric and weather research. The emergence of foundation models has marked a transformation in the science of machine learning. These foundation models are general-purpose models trained on huge amounts of data using self-supervised methods, eliminating the need for labelled data. Once trained, the parameters of these models can be utilised as a starting point for a range of domain-specific tasks. This approach is advantageous in terms of both cost and performance, as it minimises the reliance on annotated data compared to models trained from scratch. Motivated by this, our study explores the foundational capabilities of AtmoRep, a stochastic atmospheric foundation model, for two distinct weather-related applications, data compression and statistical downscaling. The training of the 3.5 billion parameter AtmoRep model consumed about a few weeks of compute time on 32 JUWELS Booster nodes.

1 Introduction

Local weather is characterised by atmospheric variables such as temperature, specific humidity, and wind speed at a given location, time, and altitude. In meteorological studies, weather typically refers to time scales ranging from hours to several days[1]. Accurate weather predictions are crucial for mitigating severe weather impacts like high winds and flooding[2] and they are relevant for many planning purposes. Understanding weather patterns requires studying complex interactions among atmospheric variables. The physical laws describing these interactions are primarily derived from fluid dynamics and radiative transfer. They are governed by conservation laws of mass, momentum, and energy[3,4]. Numerical Weather Prediction (NWP) models forecast intricate weather patterns[5,6], utilising preprocessed observational data to estimate the initial conditions[7]. The NWP models employ discretisation in space and time with current operational models typically achieving resolutions of around 10 km in longitude and latitude for global forecasts. The output from NWP models is often post-processed with statistical tools, for example, to achieve bias correction and to further increase the spatial resolution with statistical downscaling[8]. Despite continuous improvements over decades and generally good predictive skills, NWP models suffer from inherent biases, limited spatial resolution, and structural errors[9], along with high computational costs.

Recently, advanced machine learning (ML) models have transformed weather forecasting. These AI-driven approaches have emerged as strong competitors to traditional NWP

models, offering better predictions at a fraction of the computational cost[10,11]. Although purely data-driven and lacking explicit physics information, these models effectively capture complex interactions among atmospheric state variables and their spatio-temporal patterns[12]. ML models also offer enhanced flexibility and can be trained to directly predict the atmospheric state several hours into the future, unlike NWP models, which are constrained by the Courant-Friedrichs-Lewy (CFL) condition[14]. Additionally, advanced ML models can exploit the added value from multiple datasets with varying resolutions and they are able to provide efficient ensemble predictions, thus offering confidence intervals for understanding forecasting uncertainty[16,17].

The emergence of foundation models has enabled a new revolution in machine learning. These models are trained on vast datasets using unsupervised and self-supervised techniques, allowing adaptation for various tasks with minimal additional training. Foundation models are also making their way into the field of weather forecasting; one such model is AtmoRep[17]. The training on a large subset of data from the 5$^{\text{th}}$ European reanalysis (ERA5[23]) enables AtmoRep to learn comprehensive representations of atmospheric dynamics. The pretrained AtmoRep exhibits skilful capabilities for various tasks such as forecasting, temporal interpolation and counterfactuals. Through fine-tuning, the performance of AtmoRep can be further improved achieving state-of-the-art results (e.g. forecasting) or applied to other downstream tasks (e.g. statistical downscaling). In this paper, we explore the capabilities of AtmoRep for two downstream tasks: data compression and downscaling for 2 m temperature.

In the following, we first provide an overview of the core AtmoRep model, focusing on the processing pipeline of the atmospheric variables and the employed training methodology. We then describe the two downstream tasks utilising AtmoRep and discuss the results from initial sets of experiments. At the end, we conclude with a summary of our findings and future research directions.

2 The AtmoRep Model

AtmoRep[17] is a stochastic, generative neural network model for atmospheric dynamics, utilising large-scale representation learning to identify patterns within the high-dimensional state space of atmospheric data. The inherently stochastic nature of the model is crucial to capture the inherent statistical nature of atmospheric dynamics. The model has been trained with ERA5 reanalysis data from 1980 to 2017 and evaluated on data of the year 2018, similar to other ML studies on weather forecasting. The architecture of AtmoRep is inspired by established transformer models[18] and Vision Transformer (ViT)[20], which have demonstrated remarkable success in natural language processing and computer vision, respectively. AtmoRep's training strategy has been adopted from BERT (Bidirectional Encoder Representations from Transformers,[19]). The model can be flexibly configured with respect to the variables and vertical levels.

The flexibility with respect to the variables is achieved through a two-step training process: In a first step, independent transformer models, termed *singleformers*, are trained separately for each atmospheric variable. In a second step, these per-variable transformers are combined with cross-attention heads added to the encoder to enable interaction between variables in the resulting multi-variable transformer model (termed *multiformer*). This

approach proves efficient, since it significantly reduces the training time needed for a high-performing AtmoRep model compared to training a multi-variable model from scratch.

Various pre-trained configurations singleformers and multiformers are publicly available from
`https://datapub.fz-juelich.de/atmorep/trained-models.html`.
All the available models were trained on 5 model levels (96, 105, 114, 123, 137), ranging from the Earth's surface to about 5 km altitude. The downstream applications discussed in this work employ the `singleformer-t` configuration for temperature and the `multi3-uv` configuration trained on temperature and wind vector components.

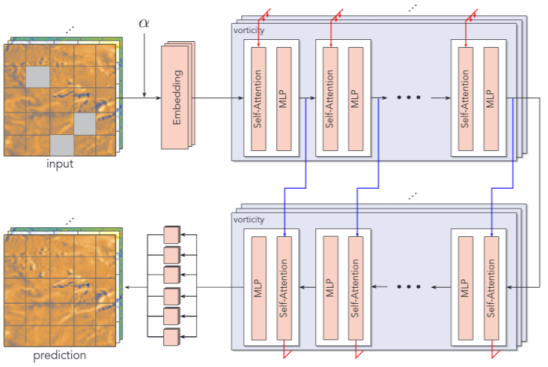


Figure 1. Schematic of AtmoRep's model architecture and training scheme[17].

Fig. 1 illustrates the AtmoRep model architecture. Transformer models implement three main concepts: tokenization, embedding, and attention mapping. In AtmoRep, the tokenization process consists of dividing the randomly selected subset of gridded ERA5 data ($36 \times 54 \times 108$ in $time \times lat \times lon$ dimensions, respectively) into several 3-dimensional cubes known as patches or tokens. The standard configuration of these tokens are variable-dependent. In general, variables with higher modes of variability (e.g., vorticity and divergence) are cut into smaller tokens than variables with less high-frequency variations (e.g., temperature). For temperature, the standard token size is $3 \times 27 \times 27$.

Subsequently, the tokens are embedded into high-dimensional vectors. Because the attention mechanism is position-independent, relative positional encoding is added to the tokens. Furthermore, latitude, longitude, model level, year, day-of-year, and time-of-day are added as auxiliary information to encode external forcings such as the seasonal cycle that is determined by the planetary motion of the Earth. The combined embeddings of positional encoding and tokens are subsequently processed by the attention blocks of the encoder network in the AtmoRep model. Self-attention is used to identify relations between patches of one variable, while cross-attention emphasises correlations across variables. The output from the encoder encompasses an abstract, feature-rich representation of atmospheric dynamics. The purpose of the decoder is then to reconstruct physical fields based on this abstract representation. The final network layer consists of a tail network

with multiple prediction heads that draw individual samples from the learned probability distribution of the atmospheric state.

To train the AtmoRep model, the principles of the BERT[19] protocol are adopted. In this framework, some tokens are randomly masked or modified during training. The model then learns to reconstruct the masked tokens based on contextual information provided by unmasked tokens. AtmoRep's training protocol is formulated as $p_\theta(y|x, \alpha)$ where x refers to the masked weather data, α refers to the auxiliary information, and y refers to the reconstructed tokens. The loss function employed to optimise the model's parameters combines Mean Squared Error (MSE) loss with a novel statistical loss that takes into account the first two statistical moments of the predicted ensemble. We refer to the original AtmoRep paper[17] for more details about the model architecture and the training process.

When the pre-trained model is applied to weather-related tasks without further fine-tuning, this is called zero-shot inference. In AtmoRep[17] zero-shot performance is evaluated for forecasting, bias correction, data interpolation, and counterfactual experiments. Here, we add results from the data (de)compression task and provide an update on 2 m temperature downscaling. For the latter, the AtmoRep core model is extended with a downscaling tail network to account for the increased output dimension. In contrast to zero-shot applications, this extension requires fine-tuning of the task-specific AtmoRep model application.

3 Downstream Tasks

3.1 Data Compression and Reconstruction

The output of climate model simulations has been growing substantially due to increased model resolution and the increased demand for detailed and high-frequency output of a comprehensive set of variables[24,25]. The storage of climate model data is therefore becoming a fundamental bottleneck limiting the possible applications of climate simulations. Data compression is one way to potentially alleviate this issue. Here, we explore how we can use the rich representation of atmospheric dynamics learned by AtmoRep to reconstruct climate data from subsets of the original fields. In principle, AtmoRep should allow for the faithful reconstruction of variables even when large portions of the data are missing, since the model was trained with randomly masked data. In this section, we investigate how well AtmoRep can reconstruct data when certain systematic masking patterns are applied.

Fig. 2 illustrates different masking patterns we employed to assess the reconstruction capabilities of AtmoRep. Our tests were constructed to assess the reconstruction quality along individual dimensions, whereas longitude and latitude were combined into a "geographic" masking pattern. The compression ratios varied from 1.42 to 4 (see Tab. 1), which means that up to 75% of the data is being omitted (i.e. masked). It should be emphasised that we tested the data reconstruction in a zero-shot setting, i.e. using the pre-trained `singleformer-t` AtmoRep configuration without any fine-tuning.

The space-time tokenization was set to $3 \times 27 \times 27$, and the neighbourhood was selected for each batch as $12 \times 2 \times 4$. The masking patterns applied are summarised in Tab. 1. For every configuration, we randomly sampled 100 days from the test year 2021 (starting from December 2020). In the first configuration (A), a "checkerboard pattern" was applied at each model level and for all time steps: every second token in longitude

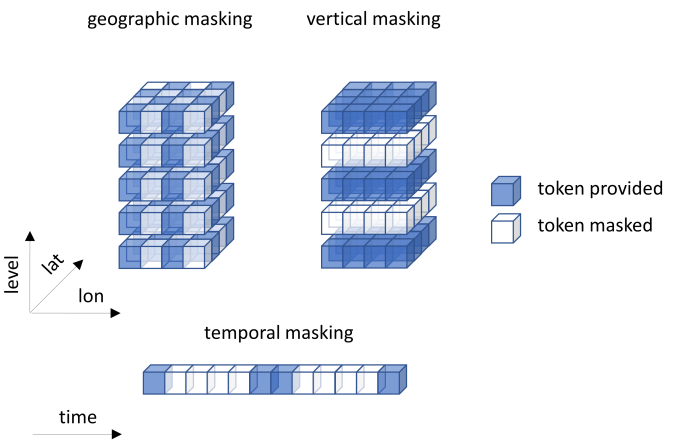


Figure 2. Illustration of masking patterns applied in the data reconstruction/decompression task.

and latitude dimension was masked resulting in a compression ratio of 2. The resulting reconstructions look physically plausible. The mean root mean square error (RMSE) ranges from 0.45 to 0.85 K across the vertical levels, with smaller errors near the surface (see Tab. 1). Configuration B explored temporal masking. In configuration B1, every second time step was masked, while configuration B2 explored a higher compression ratio of 3. Ablation studies on temporal masking indicated that the reconstruction results are best if data points at the beginning and end and at the centre of the 36-hour time window are retained. This is the motivation for applying the pattern depicted at the bottom of Fig. 2. Compared to geographic masking the reconstruction accuracy of temporal masking is slightly better. It is interesting to observe that a larger compression ratio has little influence on the reconstruction error near the surface and actually leads to smaller errors at higher model levels. The reasons for this behaviour are not fully understood. In config-

run	compression ratio	level-wise RMSE				
A	2	0.4451	0.4878	0.6748	0.7374	0.8465
B1	2	0.3765	0.4117	0.5514	0.6568	1.2300
B2	4	0.4568	0.4726	0.6119	0.6551	0.9225
AB	4	0.4009	0.4318	0.5939	0.6379	0.7287
C1	2.5	NA	1.6845	NA	1.4229	NA
C2	1.42	NA	0.3996	NA	0.4506	NA

Table 1. Configuration and accuracy (RMSE) of the data reconstruction experiments. For explanations of the masking patterns, see Fig. 2 and text. The compression ratio is defined as the ratio of available input tokens to the full number of tokens of the reconstructed field. Further information on the masking patterns is given in the text. RMSE values are given per model level with level indices (from left to right) 137, 123, 114, 105, and 96.

uration AB we combine geographic and temporal masking, thus achieving a compression ratio of 4. The results show a slight improvement in RMSE values.

The third set of configurations explores vertical masking, i.e. leaving out data from specific vertical levels and asking the model to interpolate vertically. We found that masking entire levels of temperature data resulted in poor reconstruction accuracy (configuration C1 in Tab. 1). Therefore, we tested a second variant where we applied the temporal masking pattern on the intermediate levels, while leaving the other levels complete (configuration C2). As expected, the results with C2 are significantly better than with C1. However, the added value compared to the geographical and temporal masking patterns (configurations A to AB) is small, especially in light of the much smaller deployed compression ratio (more input information available).

As shown above, the data reconstruction is in principle possible, but further work is needed to achieve the desired level of accuracy (e.g., $RMSE < 0.1$ K) and computational performance (e.g., reconstruction time $<$ 1s). In any case, the experiments revealed interesting aspects of the model behaviour. Provided that it is possible to solve the issues described above, this novel data compression approach offers a lot of potential, because it would enable very high compression ratios (up to 100 or more) with relatively little dependence of the reconstruction quality on the masking ratio (since most of the information is stored in the model weights). We anticipate that proper fine-tuning and the use of multivariate information will further improve the results.

3.2 Statistical Downscaling

Localised and regional meteorological data is highly relevant for society, agriculture, and several industrial sectors, such as renewable energies. This particularly holds for regions with complex terrain which introduces significant spatial variability in key meteorological variables such as precipitation, wind speed, or the near surface temperature. The ERA5 reanalysis, which has been utilised to train AtmoRep, operates at a resolution of $\Delta x_{ERA5} \simeq 30$ km, which is clearly insufficient to reproduce orographic features. While ERA5 provides a comprehensive estimate of the atmospheric state[23], it has well-documented limitations in mountainous regions, such as the Alpine region in Central Europe. Even though there are ongoing efforts to generate meteorological data on the scale of $1 - 2$ km with numerical models, these constitute a major computational challenge. Therefore, several weather centres developed statistical models to create higher-resolution information from coarser-resolution model output. ML models can be applied to this task with great efficiency and equal to better quality.

To demonstrate AtmoRep's adaptability for downstream applications, we applied it to perform statistical downscaling of T2m data to a resolution of approximately 6 km. For this purpose, we selected the COSMO REA6 reanalysis[26] as target dataset. COSMO REA6 provides much more accurate information than ERA5, especially over the Alpine region[22]. While the downscaling application has already been introduced in AtmoRep[17], we extend this analysis to further demonstrate the model's effectiveness for this task. This includes a more detailed analysis of spatial error patterns and of the spatial variability in the downscaled T2m field. To substantiate our findings, we compare AtmoRep's performance with an Wasserstein Generative Adversarial Network (WGAN[27]), offering a more advanced benchmark than previously used in AtmoRep[17].

The downscaling application utilises the `multi3-uv` configuration of AtmoRep which has been pre-trained on temperature and the horizontal wind components. Note that AtmoRep does not require input of high-resolution topography as many other downscaling models; it can extract the high-resolution features from the dynamic variables alone. For the downscaling task, we extended the core model with a tail network of 6 transformer blocks that is connected to the last transformer block of AtmoRep's decoder. Each block comprises a self-attention layer with 16 attention heads and a multilayer perceptron with two layers. To achieve the desired resolution of the data, the output token size of the downscaling network is enhanced by a factor of 4. The increased token size necessitates an increased embedding dimension for the temperature data achieved with a linear layer at the beginning of the downscaling network. Accordingly, the local position encoding is updated. Again, an ensemble tail is deployed to provide a probablistic downscaling output. However, a small ensemble member size of 4 was chosen due to computational constraints. During fine-tuning, the network parameters of both the core model and the tail network were optimised, resulting in about 1.85B trainable parameters. For optimal hardware utilisation, we employed both data and model parallelism. The downscaling network has been trained for three days on 8 nodes on JUWELS Booster.

Fig. 3 showcases a sample from the test year 2018, demonstrating that the downscaling not only generates super resolution output, but also achieves a bias correction of the input data.

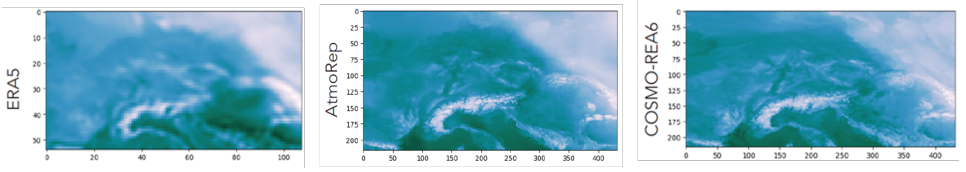


Figure 3. Downscaling sample from 2018 with an air mass boundary (AMB) in the north-eastern part of the domain. The AMB in ERA5 (left) is located further north-east compared to COSMO REA6 (right). The AtmoRep downscaling result (centre) demonstrates that the location of the AMB is corrected towards the ground truth data.

To assess the potential benefit of using AtmoRep for downscaling, we compare our results with those from a WGAN. The WGAN utilises a U-Net generator with 3.5 million trainable parameters that has been adopted from the 2 m temperature downscaling study of[28] and a convolutional critic network with 1.5 M trainable parameters. In analogy to AtmoRep, the generator is informed with temperature and wind information from several model levels. Additionally, it also inputs coarse- and high-resolved surface topography data to support the resolution mapping. The generator and critic components are trained adversarially for 40 epochs on a single A100 GPU requiring about 20 hours. No noise injection is performed in the generator, resulting in a determinstic WGAN downscaling model. To reduce the memory requirements during training, a smaller target region is chosen for the WGAN.

Fig. 4a shows the diurnal cycle of the space-time averaged RMSE over the complete test year 2018 for both models. With an ensemble-averaged RMSE of $0.989\,\mathrm{K}$, the AtmoRep downscaling model clearly outperforms the WGAN ($\overline{RMSE} = 1.163\,\mathrm{K}$). The

margin is largest for the afternoon and evening hours and can mainly be attributed to lower errors over the Alpine region. As depicted in Fig. 4b, the spatial RMSE distribution is rather uniform with AtmoRep, whereas the WGAN exhibits RMSE values up to 3 K over the Alpine region. This clearly documents the superiority of AtmoRep for the downscaling task and its ability to fill in realistic orographic features in complex terrain even without explicit topographic information.

In contrast to the conclusion above, the WGAN model is slightly better in reproducing the spatial variability of the downscaled T2m field compared to AtmoRep (not shown). Power spectrum analysis, along with comparisons of the domain-averaged horizontal T2m gradient against the COSMO REA6 ground truth, indicates that AtmoRep underestimates small-scale spatial variability by approximately 10 % (not shown). This is not entirely surprising since we are evaluating the ensemble mean state of AtmoRep, which will decrease variability. When we look at individual ensemble members, the underestimation of variability is slightly reduced, but differences to COSMO REA6 remain. A possible reason for this could be the very small ensemble size of 4 members. An increased ensemble size would require a more efficient model configuration. Strategies for this include freezing portions of AtmoRep's encoder-decoder weights during fine-tuning or implementing a more light-weight tail network, for instance with Swin Transformers[29] or Perceiver IO-modules[30].

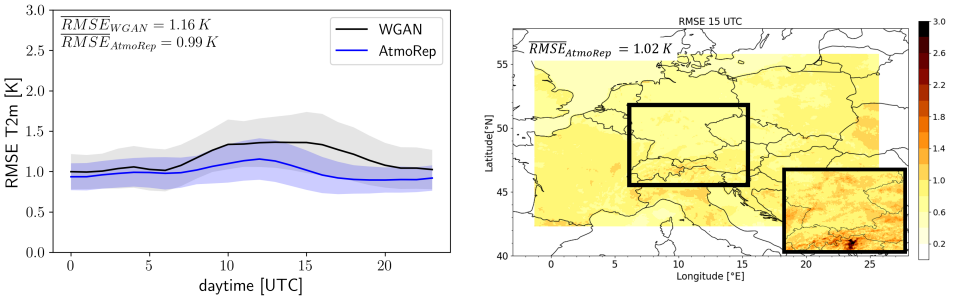


Figure 4. (a) Diurnal cycle of the domain-averaged RMSE for AtmoRep and the baseline WGAN downscaling model over the test data from 2018. The shaded area shows the standard deviation. (b) Spatial distribution of the RMSE with the AtmoRep downscaling model averaged over the test data at 15 UTC. The evaluation region is rendered in black. Additionally, the corresponding results of the WGAN model are displayed in the lower-right corner.

4 Summary and Outlook

AtmoRep is one of the first foundation models for weather and climate applications that fully exploits modern concepts of generative machine learning. In the 3 years since its conception, the model has demonstrated very good skills at a variety of meteorological tasks which had not been part of the original (pre-)training schedule. AtmoRep's capabilities for high-quality short-term forecasting, model correction, statistical downscaling, and counterfactual experiments have been demonstrated in Ref. 17. Here, we extended

the evaluation of AtmoRep by exploring its use as a data compression engine in a zero-shot setting and by further analysing the downscaling application including a comparison against a competitive Wasserstein GAN model.

The data (de)compression application explored a scenario where humongous amounts of climate data could be reduced by storing only every n-th grid box, k-th time step, or m-th model level. While this task has a lot of similarities with the pre-training task of random masking, our results nevertheless show that the systematic masking of specific patterns along the horizontal, vertical, or time dimension can introduce systematic biases in the reconstructed fields. RMSE values of reconstructed temperature fields range from about $0.4\,\mathrm{K}$ on the lowest model level to slightly higher values at the top level of $5\,\mathrm{km}$ altitude using compression ratios between 1.42 and 4. Although this is worse than the reconstruction quality of standard compression algorithms (e.g., JPEG), the advantage of AtmoRep is that it does allow for much larger compression ratios (combination of patterns) with relatively little degradation in performance. Furthermore, due to its probabilistic nature, AtmoRep can generate entire ensembles based on the compressed input of a single field. It can be expected that the reconstruction quality further improves when the model is fine-tuned and when we exploit cross-variable correlations.

Concerning the downscaling application, AtmoRep has demonstrated its superiority over a leading competitor model based on a WGAN. Although it failed to fully capture the enhanced variability of high-resolution temperature fields in complex terrain, it achieved very good scores in terms of absolute error and RMSE and generated credible high-resolution patterns following the complex orography over the Alps, even though no topographic information was provided to the model. Initial results suggest that the downscaling concept also works for other variables, in particular precipitation, which is most challenging but also highly relevant. In the current configuration, the ensemble size is very limited so that a robust assessment of the uncertainty of the downscaled field is not possible. Various approaches to overcome these limitations have been discussed above and are currently being explored. Already now, AtmoRep establishes a new state-of-the-art with respect to temperature downscaling and we are confident that this will also apply to other variables and regions.

The research on AtmoRep presented in this paper has been carried out with very little specific funding. Only recently, several projects that aim to further develop AtmoRep into a versatile model for weather and climate applications have been granted and the AtmoRep consortium continues to grow. While foundation models for weather and climate are still in their infancy, AtmoRep already allows some glimpses into what may become possible with such tools. It can be expected that foundation models for weather and climate will at some point replace classical numerical models in many different application areas as they are substantially faster and often better. However, there are still several fundamental questions to be solved and various technical challenges to be overcome. The evolution of supercomputing centres to provide more dedicated support for AI applications is one important cornerstone for building a bright future for weather and climate AI.

Acknowledgements

JSC, NIC and GCS for the provision of computing time and storage under the projects deeprain, deepacf, and atmorep. ECMWF for producing ERA5 and making it available. Several colleagues from JSC for technical support and advice. We also acknowledge the

efforts of Nils Nobre Witter who conducted few data compression tests as a part of his bachelor thesis. We acknowledge the CERN Knowledge transfer fund for the support in the development of the AtmoRep core model.

References

1. Freie Universität Berlin, *Definition of Weather and Climate*, `https://www.geo.fu-berlin.de/en/v/iwm-network/learning_content/environmental-background/basics_climategeography/definitions/index.html` (Accessed: 2024-10-11).
2. WMO, *WMO Atlas of Mortality and Economic Losses from Weather, Climate and Water Extremes (1970–2019)*, WMO Publication No. 1267, 2021.
3. D. D. Holm, J. E. Marsden, and T. S. Ratiu, *The Euler-Poincaré Equations in Geophysical Fluid Dynamics*, Geophysical Fluid Dynamics **251-300**, Cambridge University Press, 2002.
4. D. R. Durran, *Numerical methods for fluid dynamics: with applications to geophysics*, Springer, 2010.
5. P. Bauer, A. Thorpe, and G. Brunet, *The quiet revolution of numerical weather prediction*, Nature **525**, 47-55, 2015.
6. J. Rockström et al., *Safe and just earth system boundaries*, Nature **619**, 102-111, 2023.
7. T. N. Palmer, *Stochastic weather and climate models*, Nature Reviews Physics **1**, 463-471, 2019.
8. D. Maraun and M. Widmann, *Statistical Downscaling Concepts and Methods*, In: Statistical Downscaling and Bias Correction for Climate Research, Cambridge University Press, 133-134, 2018.
9. T. Palmer and B. Stevens, *The scientific challenge of understanding and estimating climate change*, Proceedings of the National Academy of Sciences **116**, 24390-24395, 2019.
10. S. Karthik Mukkavilli, D. Salles Civitarese, J. Schmude, J. Jakubik, A. Jones, N. Nguyen, C. Phillips, S. Roy, S. Singh, C. Watson, R. Ganti, H. Hamann, U. Nair, R. Ramachandran, and K. Weldemariam, *AI Foundation Models for Weather and Climate: Applications, Design, and Implementation*, 2023, arXiv:2309.10808.
11. Z. Ben-Bouallegue, M. C. A. Clare, L. Magnusson, E. Gascon, M. Maier-Gerber, M. Janousek, M. Rodwell, F. Pinault, J. S. Dramsch, S. T. K. Lang, B. Raoult, F. Rabier, M. Chevallier, I. Sandu, P. Dueben, M. Chantry, and F. Pappenberger, *The rise of data-driven weather forecasting*, 2023, arXiv:2307.10128.
12. G. J. Hakim and S. Masanam, *Dynamical Tests of a Deep Learning Weather Prediction Model*, Artificial Intelligence for the Earth Systems **3**, e230090, 2024.
13. K. Bi, L. Xie, H. Zhang, X. Chen, X. Gu, and Q. Tian, *Pangu-Weather: A 3D High-Resolution Model for Fast and Accurate Global Weather Forecast*, 2022, arXiv:2211.02556.
14. C. Bodnar, W. P. Bruinsma, A. Lucic, M. Stanley, J. Brandstetter, P. Garvan, M. Riechert, J. Weyn, H. Dong, A. Vaughan, J. K. Gupta, K. Tambiratnam, A. Archibald, E. Heider, M. Welling, R. E. Turner, and P. Perdikaris, *Aurora: A Foundation Model of the Atmosphere*, 2024, arXiv:2405.13063.

15. T. Nguyen, R. Shah, H. Bansal, T. Arcomano, S. Madireddy, R. Maulik, V. Kotamarthi, I. Foster, and A. Grover, *Scaling transformer neural networks for skillful and reliable medium-range weather forecasting*, 2023, arXiv:2312.03876.
16. I. Price, A. Sanchez-Gonzalez, F. Alet, T. R. Andersson, A. El-Kadi, D. Masters, T. Ewalds, J. Stott, S. Mohamed, P. Battaglia, R. Lam, and M. Willson, *GenCast: Diffusion-based ensemble forecasting for medium-range weather*, 2023, arXiv:2312.15796.
17. C. Lessig, I. Luise, B. Gong, M. Langguth, S. Stadtler, and M. Schultz, *AtmoRep: A stochastic model of atmosphere dynamics using large scale representation learning*, 2023, arXiv:2308.13280.
18. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, *Attention Is All You Need*, 2017, arXiv:1706.03762.
19. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*, 2018, arXiv:1810.04805.
20. A. Dosovitskiy, H. Beyer, A. Kolesnikov, D. Weissenborn, and X. Zhang, *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*, 2010, arXiv:2010.11929.
21. S. Rasp and S. Lerch, *Neural networks for postprocessing ensemble weather forecasts*, Monthly Weather Review **146**, 3885-3900, 2018.
22. S. C. Scherrer, *Temperature monitoring in mountain regions using reanalyses: lessons from the Alps*, Environmental Research Letters **15.4**, 044005, 2020.
23. H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Munoz-Sabater et al.,*The ERA5 global reanalysis*, Quarterly Journal of the Royal Meteorological Society **146.730**, 1999-2049, 2020.
24. P. Bauer, P. D. Dueben, T. Hoefler et al., *The digital revolution of Earth-system science*, Nature Computational Science **1**, 104–113, 2021.
25. M. Govett et al., *Exascale Computing and Data Handling: Challenges and Opportunities for Weather and Climate Prediction*, Bulletin of the American Meteorological Society, in press, 2024, doi:10.1175/BAMS-D-23-0220.1.
26. C. Bollmeyer, J. D. Keller, C. Ohlwein, S. Wahl, S. Crewell, P. Friederichs, A. Hense, J. Keune, S. Kneifel, I. Pscheidt, S. Redl, and S. Steinke, *Towards a high-resolution regional reanalysis for the European CORDEX domain*, Quarterly Journal of the Royal Meteorological Society **141**, 1–15, 2015.
27. M. Arjovsky, S. Chintala, and L. Bottou, *Wasserstein GAN*, 2017, arXiv:1701.07875.
28. Y. Sha, D. J. Gagne II, G. West, and R. Stull, *Deep-Learning-Based Gridded Downscaling of Surface Meteorological Variables in Complex Terrain. Part I: Daily Maximum and Minimum 2-m Temperature*, Journal of Applied Meteorology and Climatology **59**, 2057-2073, 2020.
29. Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, *Swin Transformer: Hierarchical Vision Transformer using Shifted Windows*, 2021, arXiv:2103.14030.
30. A. Jaegle, S. Borgeaud, J. B. Alayrac, C. Doersch, C. Ionescu, D. Ding, S. Koppula, D. Zoran, A. Brock, E. Shelhamer, O. Hénaff, M. M. Botvinick, A. Zisserman, O. Vinyals, and J. Carreira, *Perceiver IO: A General Architecture for Structured Inputs & Outputs*, 2021, arXiv:2107.14795.

High-Resolution Limited Area Reanalysis and Irrigation Impacts in Europe

Bernd Schalge[1], Jane Roque[2], Haojin Zhao[3], Jan D. Keller[1] Harrie-Jan Hendricks-Franssen[3], and Arianna Valmassoi[1]

[1] Deutscher Wetterdienst, Abteilung FE1, 63067 Offenbach, Germany
E-mail: {Bernd.Schalge, Jan.Keller, Arianna.Valmassoi}@dwd.de

[2] Institute of Geosciences, University of Bonn, Bonn, Germany
E-mail: jroquema@uni-bonn.de

[3] Institute of Bio- and Geosciences: Agrosphere (IBG-3), Forschungszentrum Jülich, 52425 Jülich, Germany
E-mail: {h.zhao, h.hendricks-franssen}@fz-juelich.de

In this work advanced concepts for reanalyses are presented. This includes the detailed consideration of the land-surface and subsurface as well as an investigation of the impact of human water use on the system. Three distinct setups are presented, each focusing on a specific part of the Earth-System. The basis is a convection-permitting continental scale traditional reanalysis with focus on the atmosphere. From there, a detailed land-surface version is created as well as a version with explicit consideration of anthropogenic water use, here as irrigation. Some preliminary results from all three versions are presented and advantages compared to traditional and existing products are discussed.

1 Introduction

Atmospheric reanalysis combines information from different observing networks into a gridded climatological data set using a numerical (weather prediction) model and a corresponding data assimilation (DA) scheme. They are used to provide consistent spatial information on the past spatio-temporal state of the atmosphere[10]. Ongoing efforts in the community point toward increasing the spatio-temporal resolution as well as including more Earth System components, i.e. moving from atmosphere-centric products to coupled ones[20]. Only a few approaches exclusively for global reanalysis have been made, often focusing on the coupling of complex atmospheric and ocean models[20]. The land surface model in reanalyses is mostly rudimentary, especially with respect to the representation of the subsurface. Further, due to the coarse resolution, the resolved scales do not reflect the spatial heterogeneity of land- and subsurface and the associated processes[5,8,9], thus also affecting the feedback effects between the compartments, of which the water cycle is a crucial part. While the introduction of high-resolution (atmospheric) regional reanalysis has been found to considerably improve the terrestrial water cycle estimates from global reanalyses[16], it can still only be seen as the first step to a more comprehensive reanalysis approach. Therefore, the implementation of an Earth system reanalysis is an important goal to enhance the consistency in representing water and energy budgets. As an increasing number of remote sensing observations and *in situ* measurements become available, it is important to develop algorithms that can make optimal use of the different types of observations in order to improve the accuracy of land surface models and generate a highly

accurate reanalysis product. We propose the application of iterative Ensemble Smoothers with multiple data assimilation (ES-MDA)[23] for the updating of a comprehensive set of parameters related to hydrological cycles and ecosystems within the land surface model on a continental scale. This algorithm excels in efficiently incorporating information from various data sources simultaneously. A further advantage of this algorithm is that a long time period can be assimilated at once, instead of having short sequential assimilation windows. Each model parameter can be informed by a comprehensive set of observations that includes temporal and spatial dynamics. By updating the model parameters rather than the model state vector, systematic deviations can also be minimised for longer-term simulations, without introducing imbalances in the water budget systems or numerical instabilities[22]. The objective is to develop and evaluate the potential of the ES-MDA to constrain land surface model parameters, and to assess its applicability in generating long term reanalysis products. While such a comprehensive modelling and DA framework is expected to lead to a considerable enhancement in representing the terrestrial water cycle, at least one major source of uncertainty would remain. Human activities can have a significant impact on the water budget, depending on time and location. Irrigation in particular plays an important role in the anthropogenic redistribution of water in the terrestrial system as it accounts for 80-90% of the global freshwater withdrawal[20]. Yet, only recently, modelling studies have focused on the explicit representation of irrigation in case study experiments. Possible reasons for this are (a) the lack of comprehensive and extensive knowledge and measurements as a basis for parameterisations and (b) the complexity of representing such processes in a numerical model. In this respect, we are not aware of any reanalysis that, to date, includes an explicit representation for irrigation despite being identified as an important process for soil moisture and near-surface representation[20]. While one could argue that the lack of irrigation in reanalyses can be corrected with the DA step, there are two main caveats to this approach. The first is the lack of high-resolution horizontal soil moisture or surface water observations to capture the effect of irrigation. Second, the vertical redistribution of water from the surface or soil into the atmosphere or onto the canopy is difficult to reproduce realistically in a DA scheme. Therefore, neglecting irrigation leads to significant biases in reanalysis data sets[17], and an explicit representation of irrigation can circumvent these shortcomings. In this respect, irrigation can also have a significant impact on feedback between the different compartments, such as potential evapotranspiration, atmospheric local and synoptic circulation and precipitation[19]. Further, irrigation leads to an increase of soil moisture and a decrease of the average diurnal surface temperature (except under special circumstances[20]), with a strong asymmetric seasonal and inter-annual variability[15]. Further, irrigation can play an important role in shaping hydrometeorological extremes[13], which in turn have feedback on the terrestrial water cycle. Given these potential effects, the analysis of an explicit representation of irrigation in a fully coupled reanalysis framework has to be thoroughly assessed, including the inherent uncertainties.

2 Motivation

As previously discussed, reanalyses are reference datasets for the past climate state, as the most up-to-date models and DA systems are used. This is why reanalyses will regularly be re-done after enough progress either on the DA or model side (or both) happened, such that higher quality can be expected. More recently, reanalyses gained further relevance as train-

ing dataset for AI-models for weather forecast, highlighting the need of high-resolution and high quality data. As for this reanalysis project, we mostly aim to get the water cycle represented as best as possible. As such, we have three separate reanalysis projects that focus on different aspects:

- The atmospheric reanalysis focuses on delivering the best possible atmospheric reference, with precipitation being a very important factor for the water cycle, followed by other near-surface variables that dictate evapotranspiration.
- The land-surface-subsurface reanalysis focuses on representing the surface/subsurface part of the water cycle as accurately as possible. This includes representation of plants, the soil an deeper subsurface, as well as overland flow in rivers.
- The atmospheric-land-surface reanalysis with additional focus on anthropogenic water use (mostly irrigation) is designed to find differences that arise from said anthropogenic water use. This reanalysis can be compared to our first one to assess sensitivity with respect to human water usage.
- The final reanalysis product would be a combined, coupled version of all previous ones.

Creating a reanalysis is a complex challenge, not only because of immediate issues and choices to make, but also with the prerequisites that have to be fulfilled before an attempt can even be made. These prerequisites include:

- A robust numerical model that has shown to perform very well when compared to competing models, especially important for local area models that needs tuning to work for certain regions of the globe. Fortunately, we have access to the models ICON, eCLM and ParFlow, all of which will be presented in more detail below.
- A state-of-the-art DA system is required. This DA system must be able to assimilate all the various observation types chosen for the reanalysis as well as use modern techniques to create the analysis to ensure best fits between observations and models.
- A database of quality-controlled observations with proper meta-data is key. If observations are of poor quality or are faulty, this will likely cause issues with the DA system. Further, we must ensure that all observations have proper metadata.
- A computing facility able to run the models, store the input and output data (hundreds of terabytes) and handle data transfer requests for analysis or visualisation purposes.

3 Methodology

In order to realise the Reanalysis we use the Terrestrial Systems Modelling Platform (TSMP) with their component models as well as the DA suites that accompany them. For the atmosphere this is the ICON model[21] with the KENDA DA system. ICON is a very versatile model, able to run global simulation as well as highly resolved local area versions. Even large eddy setups are possible, but still experimental at this moment. ICON has been

used operationally for several years now at the German Weather Service (Deutscher Wetterdienst, DWD) for both the global system as well as for a high-resolution (2km) local area mode for central Europe (ICON-D2). They have shown to be among the leading NWP models making them suitable for reanalysis purposes. For DA ICON uses the KENDA system, which is also in use operationally at DWD to create the analysis, a requirement for any forecast run. For limited area assimilation, the LETKF (Local Ensemble Transform Kalman Filter) is the default option for DA with ICON.

For the land surface the Community Land Model (CLM) in version 5.0 is used[11]. CLM has seen many iterations over the years and has come to be known a very efficient and detailed land model. Able to run in various different modes, it is used to tackle a wide variety of problems regarding the land surface. When coupled to ICON it replaces the default land surface scheme that is included with ICON. CLM is more detailed and uses more types of vegetation and has a much finer spectrum for soil types. The hydrological part of the simulation is covered by ParFlow[12]. While CLM is able to simulate soil moisture and overland routing of rivers, ParFlow has the advantage of being able to close the water balance, as all fluxes across any domain border can be exactly calculated since it is uses the full 3D Richards equations for subsurface flow (shallow water equations for water flow in the land surface). In addition, ParFlow uses Van-Genuchten parameters to specify soil properties. A very detailed knowledge of soil and subsurface properties is needed for best results. All these models are coupled using the external OASIS coupler. For CLM and Parflow DA is covered by PDAF[14], which features a wide variety of options for the DA options also including the LETKF.

The specific setup of these models depend a lot on the goal of the simulation, which can vary quite a bit. We now describe in more detail the setup of our experiments, starting from the ICON stand-alone reanalyses, following with the land surface reanalysis, and finishing with the inclusion of the irrigation parameterisation in the ICON model.

The first is an ICON standalone reanalysis. We will show results from an exploratory simulation with the year 2022. The final reanalysis will follow the same procedure an additional observation type (atmospheric motion vectors or AMVs) will be added. For the basic setup we start with the operational ICON-D2 setup. It has been shown that this setup is optimal for central Europe at 2km resolution. However, we are using a 3km resolution and the EURO-CORDEX domain is much larger than the ICON-D2 domain. For optimal tuning, a lengthy process of sensitivity experiments would be needed. But even then we will likely end up with a system that still performs worse than any properly tuned system on any subset of our domain. So given that the ICON-D2 domain is right at the centre of the EURO-CORDEX domain and the difference in resolution is still small, we remained with the ICON-D2 tuning, knowing that it will not be optimal in other parts of the world, such as the eastern Mediterranean. We did increase the vertical extent from default 60 to 75 layers up to 33km, to allow for deeper troposphers, as can be found in the southern regions in summer. The boundary conditions for our experiment are provided by ERA-5, which especially for the ensemble are at coarse ($\sim$80km) so near the boundary of the domain larger deviations are expected. As these regions are mostly ocean (western/norther boundary), desert (southern boundary) or sparsely populated areas with few observing stations (eastern boundary, with some exceptions), this should overall have negligible impact on the results. For the DA system we also use ICON-D2 as a basis, with some key differences. First, we can only afford to run 10 ensemble members, due to the computational

cost and the storage requirements. Another change is the removal of observation types that require huge amounts of disk storage and memory. This includes most satellite data as well as radar observations. We still use all conventional observations, which includes classic weather stations and radiosondes, but also wind profilers and aircraft data. The latter is only available in a greatly reduced quantity though. The results obtained from this reanalysis will be compared to other reanalysis projects, notably ICON-DREAM, another ICON reanalysis run globally at 13km resolution with a 6,5km nest over Europe with almost the same extent as the EURO-CORDEX domain. ERA-5 is the other reanalysis we compare to, as it is the most widely used and also the source of our boundary data, meaning any improvements would be due to our system rather than a better quality forcing.

For the land surface reanalysis, we started with the eCLM and implemented with the ES-MDA algorithm. Compared to the traditional EnKF, the ES-MDA assimilates a longer time period instead of short sequential windows that can better capture slow processes and associated parameters. The simulations were performed on an extended EURO-CORDEX domain with a spatial resolution of 0.11° driven by ERA5 reanalysis dataset. A number of studies[25,24] have investigated the sensitivity of parameters (hard-coded) in CLM 5.0. Based on this, we chose a set of input variables and parameters that affect the hydrological and biophysical processes.

Considering the uncertainty in the soil texture, we introduced depth-invariant spatial perturbation fields to the percentage sand, percentage clay and percentage soil carbon content. As CLM 5.0 calculates the porosity, saturated hydraulic conductivity and soil matric potential based on pedo-transfer functions for both mineral[26]and organic soil, the slope and intercept parameters in the empirical regressions were perturbed. The runoff generation process has a strong influence on soil moisture[27], consequently, we also perturbed the decay factor that links soil moisture and the fractional saturated area in the model. On the other hand, the evapotranspiration process is sensitive to the stomatal conductance and photosynthesis parameters[24]. To account for this, we perturbed the parameters in the stomatal conductance model, the Medlyn model[28] for each Plant Functional Type (PFT) and parameters in the photosynthesis model. At last, considering the transpiration is largely regulated by the plant hydraulic conductivity, we further perturbed the maximum conductance and maximum canopy water storage. For the observations, we used the SMAP enhanced Level 3 soil moisture product operating at 9 kilometre resolution[29]. The observation error was set to 0.04 $mm^3\ mm^{-3}$, which is consistent with the target accuracy under favourable conditions[29]. Additionally, we used the evapotranspiration measured by eddy covariance (EC) from the Integrated Carbon Observation Systems (ICOS) over 70 sites. In order to evaluate the impact of assimilation type, tapering and inflation factors and the number of iterations, a series of sensitivity tests were conducted. The tapering factor was set to 0.4 to achieve a balance between model mismatch for SMAP and ICOS observations. Meanwhile, the inflation factor was adjusted to 1.05 to maintain ensemble spread and better represent model uncertainty. The results demonstrated that the optimal performance is achieved through the assimilation of both SMAP and ICOS data with 5 iterations with 64 ensemble members. Finally, we used the aforementioned settings and assimilated data for the year 2019 and validated the results for 2020.

The third experiment, the atmospheric-land-surface reanalysis including the representation of irrigation started with implementing the irrigation parameterisation in ICON. This parameterisation is an adaptation of the CHANNEL parameterisation developed originally

for the Weather Research and Forecast model (WAF)[18]. We introduce this adaptation in the interface between the land surface scheme and the atmosphere. Therefore, this parameterisation adds irrigation water into the grid-scale precipitation before this variable goes to TERRA, which is the default Land Surface Model coupled with ICON-nwp. Even though the precipitation amount changes with the irrigation water, canopy interception is not possible as it does not exist a canopy layer in this land surface scheme. At the moment, the irrigation water (W_1) is calculated with the same equation (Eq. 1[18]), and it depends on the daily irrigation amount (V_1 in mm d^{-1}), the number of irrigation hours in seconds (h_1), and the irrigation interval (T_1 in absolute number of days). We decided to consider a fixed irrigation water amount for the free simulations, and we choose three of the national-reported values of water abstractions from Eurostat[7] and we apply them to the whole domain, i.e. daily values of 2.6 mm (France MFR hereafter), 6.7 mm (Spain MSP hereafter), 11.1 mm (Italy MIT hereafter). Besides the irrigation amounts from the Eurostat, we included two other experiments, fixed soil moisture at field capacity and saturation for the top six layers where TERRA calculates the water balance. Other studies also considered including or limiting irrigation water directly to the soil moisture[2].

Regarding the other irrigation settings, we concentrate on the summer season as the effects of irrigation on the atmospheric component are more evident when atmospheric conditions are warm and dry[20]. Concerning the irrigation frequency and length, we opted for irrigating the whole day (24 h) with a daily frequency. For the free simulations we used ICON global operational initial and boundary conditions, but similar settings as the ICON standalone reanalysis, and we run our experiments from March 01, 2022 until the end of August.

4 Results

The first part of the results are from the ICON standalone reanalysis without any changes to irrigation. Here, the focus is on the comparison to other existing reanalysis products. As discussed in the previous section where the setup is presented, the main strength of this reanalysis is the high resolution, which helps specifically with precipitation and for stations in complex terrain, which are often be filtered out in more coarse models. This results in a counter-intuitive effect that this reanalysis features some stations with more

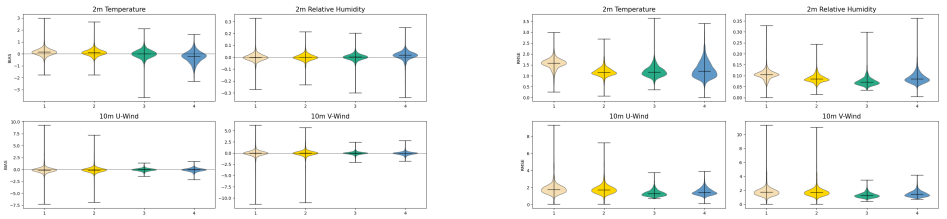


Figure 1. Comparison of 2m temperature, humidity and wind components BIAS and RMSE between different reanalysis products for 2022. The distribution shows all actively assimilated stations for each product. Light yellow is the first guess for our reanalysis, bright yellow the actual reanalysis, green is ICON-DREAM and blue ERA-5.

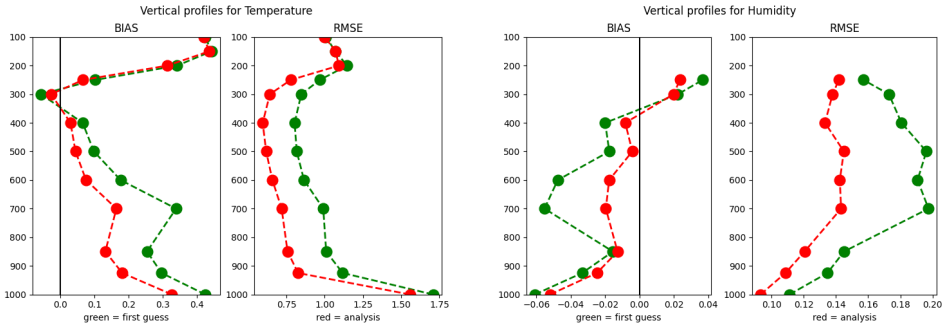


Figure 2. Comparison of temperature and humidity profiles of BIAS and RMSE. The green line is the first guess, the red line the analysis.

extreme values with regards to biases and RMSE, as can bee seen in Fig. 1. However, there are only very few such stations, so for mean values these do not play any significant role. Overall we see that we are competitive in quality with other reanalysis products despite the many compromises we had to make. We see similar performance as ICON-DREAM and better performance than ERA-5 (except for Wind).

Unfortunately, we do not yet have a comparison of vertical profiles for all reanalysis products, instead we use the vertical profiles of Fig. 2 to highlight the performance of the DA system itself. First of, the shown biases and RMSE values are consistent with operational ICON-D2 results, albeit with slightly larger magnitude, due to all the compromises discussed in the previous section. But what is very obvious and important is that the assimilation leads to a better fit to observation in all elevations of the analysis compared to the first guess, despite the rather sparse coverage at higher altitudes.

The results of the data assimilation with the land surface model eCLM and the ESMDA approach are illustrated in Fig. 3. The eCLM open loop runs, i.e. the run without the Data

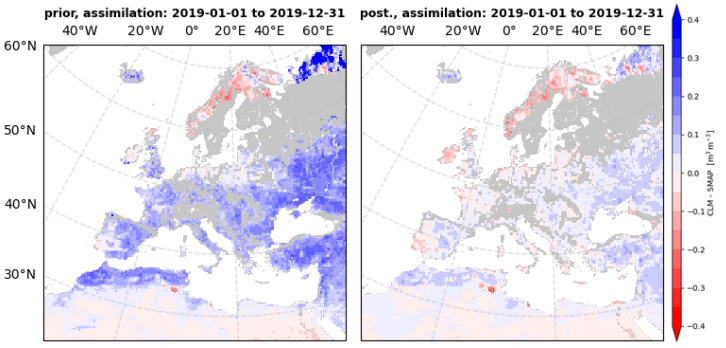


Figure 3. Spatial distribution of mismatch between the SMAP soil moisture product and CLM 5.0 modelled soil moisture for open loop simulations and data assimilated runs. The left column shows open loop results and the right column DA results.

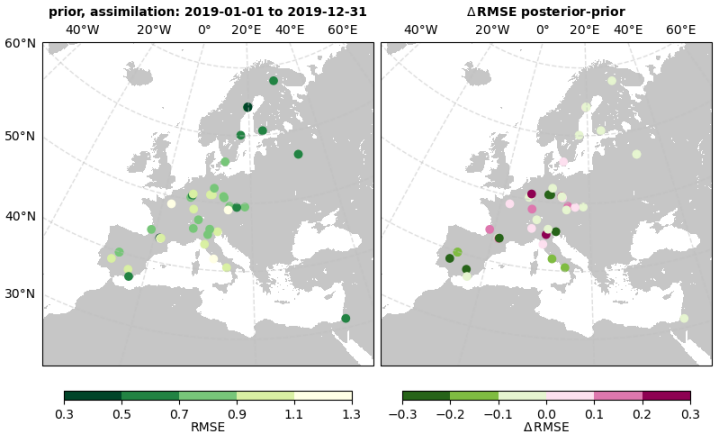


Figure 4. RMSE between the evapotranspiration (ET) modelled by CLM 5.0 and measured form ICON sites (in mm d^{-1}), same as Fig. 3.

Assimilation, exhibit an overestimation of soil moisture across the majority of Central and Eastern Europe, as well as Northern Russia, and an underestimation in the Nordic regions when compared to the SMAP retrievals. The results indicate that these discrepancies between modelled soil moisture and SMAP soil moisture retrievals become less pronounced for both the assimilation period (2019) and the validation period (2020, not shown), suggesting that structural improvements are made with respect to the overall modelled soil moisture.

We also present the RMSE between the modelled and observed evapotranspiration estimates in Fig. 4, most of them lie below 1 mm d^{-1} for the open loop simulations. It is noteworthy that the modelled ET shows promising results in Nordic needle leaf forests, exhibiting an RMSE of 0.5 mm d^{-1}. The right panel shows the reduction or increase in RMSE after the data assimilation. A limited number of stations demonstrate a notable reduction in RMSE, predominantly from cropland and grassland sites in Germany and France. Conversely, a number of stations exhibit little changes, particularly those dominated by deciduous broadleaf forests. This may indicate that the use of a single plant functional type to represent deciduous broadleaf forests across Europe may be inadequate.

The sensitivity tests using the irrigation parameterisation including different irrigation amounts show that soil moisture increased in different magnitudes over all experiments. After subtracting the CTRL from the irrigation runs, the mean values for the top soil layer (0 - 9 cm) over all tests for May and JJA in irrigated areas are 3.4 and 3.7 kgm^{-2} respectively. Fig. 5 shows the soil moisture differences for two irrigation experiments: soil moisture fix to field capacity (FC) and the experiment with a mean irrigation amount from France (MFR). Soil moisture increased on average 2.9 kgm^{-2} and 3.6 kgm^{-2} for FC and MFR respectively. The changes in soil moisture influenced the partitioning of energy fluxes, as shown already by other studies[15], by increasing the Latent Heat Flux (LHF) and decreasing the Sensible Heat Flux (SHF). The average LHF increase over all experiments in irrigated areas is 64.7 Wm^{-2}, and the average SHF decrease over all experiments is 52 Wm^{-2}. This Figure also demonstrates the close proximity between the results from differ-

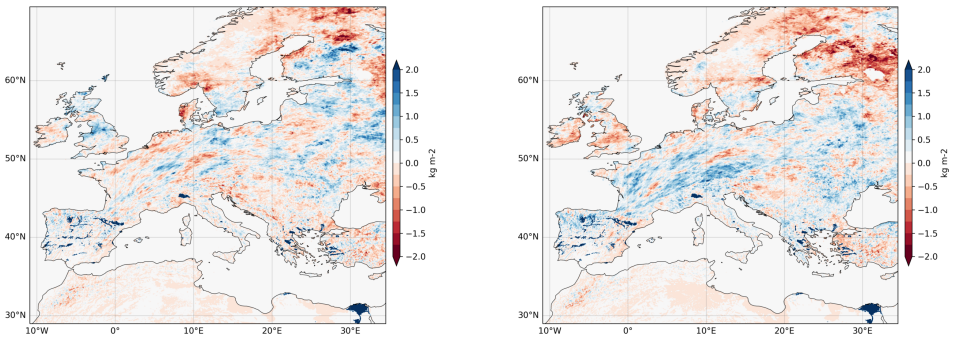
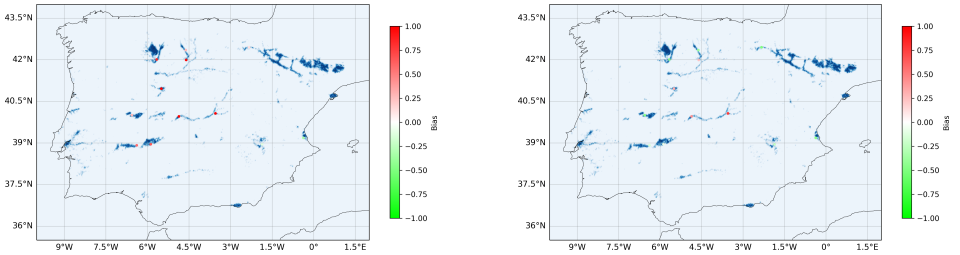


Figure 5. Differences between irrigation experiments and control for soil moisture (0-9 cm), JJA. Left: FC experiment. Right: MSP experiment.

ent experiments, as there is an overlap of averaged values. This suggests a lower sensitivity between applying different irrigation amounts, at least with values between 2.6 and 11.1 mmd^{-1} in the irrigated fraction of GlobCover 2009. Even though SAT and FC overlap, their influence in surface variables is lower than the influence of the other experiments (MFR, MSP and MIT), as their values were directly forced in the first six soil layers of TERRA.

The changes in the energy flux partitioning indicate that less energy is used to heat the surface. As a result, we expected a decrease in temperature values. Our sensitivity simulations depict this scenario in relation to 2 m temperature (T2m). In this case, all simulations obtained negative values in the irrigated areas after subtracting the CTRL run. The average temperature in irrigated areas decreased by 1.0 K and 1.4 K in May and JJA respectively, with a stronger cooling effect during day-time (decrease of 2.3 K) than during night-time (decrease of 0.5 K). We compared these temperature values with observations retrieved by the first experiment (reanalysis). The location with more observations available close or in irrigated areas for the month of June was Spain. The average bias from FC and MSP experiments is lower compared with the control. The control run reaches a bias of 0.66, while FC and MSP have a bias of -0.08 and -0.29 respectively (Fig. 6).



(a) CTRL (b) FC

Figure 6. Bias from irrigation experiments in comparison with observations for Spain, June 2022.

5 Concluding Remarks

We have shown preliminary results from various setups that will eventually lead to a fully-coupled and unified reanalysis framework, since each setup is designed to test specific aspects.

For the atmospheric baseline setup we were able to show good performance compared to other reanalysis despite using a very small ensemble size and a limited set of observations. The reanalyses period is currently being extended and evaluated against the other reanalyses.

The ES-MDA, used for land surface data assimilation, is proved to be an efficient algorithm for the assimilation of spatial and temporal data from different sources. The tuning of the different pedotransfer function parameters have been estimated at a continental scale. The ecosystem parameters are estimated using measurements from ICOS. The wet biases in soil moisture have been further reduced through assimilation while the estimation of ET has seen only a slight improvement. Therefore, the next stage of the process involves running the eCLM model in biogeochemical mode and assimilating the Leaf Area Index (LAI) data retrieved from MODIS (the Moderate Resolution Imaging Spectroradiometer).

The irrigation sensitivity experiments demonstrated that simulations with ICON-NWP exhibit minimal sensitivity to different irrigation amounts. However, the results differ when the irrigation water is included as grid precipitation or directly forced in the soil. Also, although we found a bias reduction in T2m for the irrigation experiments, there is still room for improvement by adjusting certain irrigation settings. Nevertheless, when comparing results with observations, we should represent irrigation more realistically. Therefore, further testing and tuning is planned.

Acknowledgements

The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. for funding this project by providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS Supercomputer JUWELS[1] at Jülich Supercomputing Centre (JSC). Further, we acknowledge the help and support from within the DETECT research cluster Z03-Z04, notably Klaus Görgen, Olaf Stein, Stefan Poll, Daniel Caviedes Voullieme, and Marco van Hulten for both technical support and data management issues. Furthermore we acknowledge support from the FE1 department at DWD regarding the ICON model and KENDA DA Framework.

References

1. Jülich Supercomputing Centre, *JUWELS Cluster and Booster: Exascale Pathfinder with Modular Supercomputing Architecture at Juelich Supercomputing Centre*, Journal of large-scale research facilities **7**, A183, 2021.
2. C. Asmus, P. Hoffmann, J. P. Pietikäinen, J. Böhner, and D. Rechid, *Modeling and evaluating the effects of irrigation on land-atmosphere interaction in southwestern europe with the regional climate model REMO2020-iMOVE using a newly developed parameterization*, Geoscientific Model Development **16**, 7311-7337, 2023.

3. R. Avissar and R. A. Pielke, *A parameterization of heterogeneous land surfaces for atmospheric numerical models and its impact on regional meteorology*, Monthly Weather Review **117**, 2113-2136, 1989.
4. C. Bollmeyer, J. D. Keller, C. Ohlwein, S. Wahl, S. Crewell, P. Friederichs, A. Hense, J. Keune, S. Kneifel, I. Pscheidt, S. Redl, and S. Steinke, *Towards a high-resolution regional reanalysis for the European CORDEX domain*, Q. J. Roy. Meteor. Soc. **141**, 1-15, 2015.
5. E. Bou-Zeid, W. Anderson, G. G. Katul et al., *The Persistent Challenge of Surface Heterogeneity in Boundary-Layer Meteorology: A Review*, Boundary-Layer Meteorol **177**, 227-245, 2020.
6. A. Craig, S. Valcke, and L. Coquart, *Development and performance of a new version of the OASIS coupler, OASIS3-MCT_3.0*, Geoscientific Model Development **10**, 3297-3308, 2017.
7. Eurostat, *Irrigation: number of farms, areas and equipment by size of irrigated area and nuts 2 regions*, Publications Office of the European Union, Luxembourg, 2018.
8. E. R. Freund, M. Zappa, and J. W. Kirchner, *Averaging over spatiotemporal heterogeneity substantially biases evapotranspiration rates in a mechanistic large-scale land evaporation model*, Hydrology and Earth System Sciences **24**, 5015-5025, 2020.
9. F. Giorgi and R. Avissar, *Representation of heterogeneity effects in Earth system modeling: Experience from land surface modeling*, Rev. Geophys. **35**, 413-437, 1997.
10. E. Kalnay, M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, A. Leetmaa, B. Reynolds, R. Jenne, and D. Joseph, *The NCEP/NCAR 40-Year Reanalysis Project*, Bulletin of the American Meteorological Society **77**, 437-472, 1996.
11. D. M. Lawrence, R. A. Fisher, C. D. Koven, K. W. Oleson, S. C. Swenson, and G. Bonan, *The Community Land Model version 5: Description of new features, benchmarking, and impact of forcing uncertainty*, Journal of Advances in Modeling Earth Systems **11**, 4245-4287, 2019.
12. R. M. Maxwell, F. K. Chow, and S. J. Kollet, *The groundwater-land-surface-atmosphere connection: Soil moisture effects on the atmospheric boundary layer in fully-coupled simulations*, Advances in Water Resources **30**, 2447-2466, 2007.
13. D. G. Miralles, P. Gentine, S. I. Seneviratne, and A. J. Teuling, *Land-atmospheric feedbacks during droughts and heatwaves: state of the science and current challenges*, Ann. N.Y. Acad. Sci. **1436**, 19-35, 2019.
14. L. Nerger and W. Hiller, *Software for ensemble-based data assimilation systems – Implementation strategies and scalability*, Computers and Geosciences **55**, 110-118, 2013.
15. M. J. Puma and B. I. Cook, *Effects of irrigation on global climate during the 20th century*, Journal of Geophysical Research Atmospheres **115**, D16120, 2010.
16. A. Springer, A. Eicker, A. Bettge, J. Kusche, and A. Hense, *Evaluation of the Water Cycle in the European COSMO-REA6 Reanalysis Using GRACE*, Water **9**, 289, 2017.
17. O. A. Tuinenburg and J. P. R. de Vries, *Irrigation patterns resemble ERA-Interim reanalysis soil moisture additions*, Geophysical Research Letters **44**, 10,341-10, 348, 2017.

18. A. Valmassoi, J. Dudhia, S. Di Sabatino, and F. Pilla, *Evaluation of three new surface irrigation parameterizations in the WRF-ARW v3. 8.1 model: the Po Valley (Italy) case study*, Geoscientific Model Development **13**, 3179, 2020.
19. A. Valmassoi and J. D. Keller, *A review on irrigation parameterizations in Earth system models*, Frontiers in Water **4**, 906664, 2022.
20. A. Valmassoi, J. D. Keller, D. T. Kleist, S. English, B. Ahrens, I. Bašták Ďurán, E. Bauernschubert, M. G. Bosilovich, M. Fujiwara, H. Hersbach, L. Lei, U. Löhnert, N. Mamnun, C. R. Martin, A. Moore, D. Niermann, J. José Ruiz, and L. Scheck, *Current challenges and future directions in data assimilation and reanalysis*, Bulletin of the American Meteorological Society **104**, E756-E767, 2022.
21. G. Zängl, D. Reinert, P. Rípodas, and M. Baldauf, *The ICON (ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core*, Q.J.R. Meteorol. Soc. **141**, 563-579, 2015.
22. G. J. M. De Lannoy, M. Bechtold, C. Albergel, L. Brocca, J.-C. Calvet, A. Carrassi, W. T. Crow, P. de Rosnay, M. Durand, B. Forman, G. Geppert, M. Girotto, H.-J. Hendricks Franssen, T. Jonas, S. Kumar, H. Lievens, Y. Lu, C. Massari, V. R. N. Pauwels, R. H. Reichle, and S. Steele-Dunne, *Perspective on satellite-based land data assimilation to estimate water cycle components in an era of advanced data availability and model sophistication*, Frontiers in Water **4**, 981745, 2022.
23. A. A. Emerick and A. C. Reynolds, *Ensemble smoother with multiple data assimilation*, Comput Geosci. **55**, 3-15, 2013.
24. K. Dagon, B. Sanderson, R. Fisher, and D. Lawrence, *A machine learning approach to emulation and biophysical parameter estimation with the Community Land Model, version 5*, Advances In Statistical Climatology, Meteorology And Oceanography **6**, 223-244, 2020.
25. H. Yan, N. Sun, H. Eldardiry, T. Thurber, P. Reed, K. Malek, R. Gupta, D. Kennedy, S. Swenson, Z. Hou et al., *Large ensemble diagnostic evaluation of hydrologic parameter uncertainty in the Community Land Model Version 5 (CLM5)*, Journal Of Advances In Modeling Earth Systems **15**, e2022MS003312, 2023.
26. R. Clapp and G. Hornberger, *Empirical equations for some soil hydraulic properties*, Water Resources Research **14**, 601-604, 1978.
27. H. Yan, N. Sun, H. Eldardiry, T. Thurber, P. Reed, K. Malek, R. Gupta, D. Kennedy, S. Swenson, L. Wang et al., *Characterizing uncertainty in Community Land Model version 5 hydrological applications in the United States*, Scientific Data **10**, 187, 2023.
28. B. Medlyn, R. Duursma, D. Eamus, D. Ellsworth, I. Prentice, C. Barton, K. Crous, P. De Angelis, M. Freeman, and L. Wingate, *Reconciling the optimal and empirical approaches to modelling stomatal conductance*, Global Change Biology **17**, 2134-2144, 2011.
29. D. Entekhabi, E. Njoku, P. O'neill, K. Kellogg, W. Crow, W. Edelstein, J. Entin, S. Goodman, T. Jackson, J. Johnson et al., *The soil moisture active passive (SMAP) mission*, Proceedings Of The IEEE, **98**, 704-716, 2010.

Computer Science and Numerical Mathematics

Computer Science and Numerical Mathematics

Thomas D. Kühne

[1] Center for Advanced Systems Understanding, Helmholtz-Zentrum Dresden-Rossendorf, 02826 Görlitz, Germany

[2] Institute of Artificial Intelligence, Chair of Computational System Sciences, Technische Universität Dresden, 01187 Dresden, Germany
E-mail: tkuehne@cp2k.org

With the rapid rise of artificial intelligence within the computational sciences, so has their share in terms of computing time made available on supercomputing centres. The latter is particularly true for the area of computer science and numerical mathematics, where beside state-of-the-art simulations an ever increasing part of nowadays allocation is dedicated to the development of basic techniques, assessment and training of foundation models.

A particular good example in this regard is the contribution of Georg Rehm and coworkers on the OpenGPT-X large language model (LLM), which is a collaborative initiative between academic and industrial partners to develop a local, open and freely available multilingual LLM to widespread used commercial services such as ChatGPT. Specifically, they have demonstrated a multitude of techniques to optimise the training of large-scale LLMs on JUWELS Booster with an emphasis on the pre-training of neural networks based on the transformer architecture. Extending the existing Megatron-LM, a highly distributed and optimised PyTorch codebase to train large-scale transformer models, they demonstrated an outstanding scalability within the training of small- and large-scale LLM ranging from 80M to 13B parameters on up to 1024 Nvidia A100 GPUs. A very important side aspect of their effort is that they provide a vast amount of data such as the selection process of the employed model architecture and corresponding hyperparameters including the required computational costs – details that otherwise often remain in the fog due to the proprietary nature of commercial alternatives.

Along similar lines, though from the area of video and image segmentation, is at the core of the contribution of Alexander Hermans *et al.* from RWTH Aachen's Visual Computing Institute. Therein, they studied how to best exploit the GPU-accelerated JUWELS Booster within the segmentation of inputs at the pixel-level by assigning them to specific class or objects instances to understand visual scenes from image and video data. Contrary conventional deep learning task such as classification and detection, segmentation requires pixel-level annotations that are rather time-consuming to obtain, thereby rendering the availability of high-quality, large-scale datasets as a major bottleneck in advancing the field. This is even more stringent for videos, where dense, frame-by-frame annotations are required. To ameliorate the latter, the focus of the group of Bastian Leibe was on the development of TarViS, a unified approach for target-based video segmentation, an holistic approach to perform multiple video segmentation tasks using a single trained model only. As before, using the PyTorch library, which provides a `DistributedDataParallel` API for seamlessly parallelising network training across multiple nodes, a parallel efficiency of approximately 70 % had been achieved using all GPUs of multiple JUWELS Booster nodes.

At variance, the final contribution of Andreas Frommer *et al.* deals with novel adaptive algebraic multigrid solvers in state-of-the-art lattice quantum chromodynamics (QCD) simulations. One of the major tasks of lattice QCD is the solution of linear systems such as $Dx = b$, where D is the discretised Dirac operator that, to make matters worse, is typically ill-conditioned. In their work, so called domain decomposition aggregation-based algebraic multigrids (DD-αAMG) are employed, which beside accelerating the performance to linear solvers allows to conduct partial eigendecompositions of D and to approximate traces of the form $\mathrm{Tr}(D^{-1})$. Offloading some of the critical sections of the DD-αAMG method via CUDA and HIP to GPUs, permits the effective usage of heterogenous computing architectures. In the numerical experiments conducted on JUWELS Booster, a speed-up of a factor 8 had been observed when using their GPU implementation of DD-αAMG with respect to the CPU-only variant.

Multilevel Approaches in Lattice QCD Simulations

Andreas Frommer[1], **Jose Jimenez-Merchan**[1], **Karsten Kahl**[1], **and Gustavo Ramirez-Hidalgo**[2]

[1] School of Mathematics & Natural Sciences, IMACM, University of Wuppertal, 42119 Wuppertal, Germany
E-mail: {frommer, jimenezmerchan, kkhal, ramirezhidalgo}@uni-wuppertal.de

[2] Jülich Supercomputing Centre, Forschungszentrum Jülich, 52425 Jülich, Germany
E-mail: g.ramirez.hidalgo@fz-juelich.de

Adaptive algebraic multigrid methods have meanwhile established themselves as standard solvers in lattice QCD computations on current supercomputers. In this paper we review recent additions and recent new use cases for the solvers, in particular our DD-αAMG method and report performance results obtained on JUWELS at JSC.

1 Introduction

In lattice Quantum Chromodynamics (QCD), a discretisation of the physical theory describing the strong force between the quarks as constituents of matter on the lattice, simulations take place on state of the art grids of sizes 128^4 and larger[1]. At such large scales, the memory and arithmetic demands of the computations force the use of supercomputers. Moreover, as simulations on the lattice approach the physical continuum, advanced algorithmic techniques are necessary. In particular, when solving linear systems, the matrix of coefficients, known as the (discretised) Dirac operator D, is very ill-conditioned, and multigrid (MG) methods[2,3] have become the state of the art in lattice QCD.

Using an algebraic multigrid (AMG) method rather than a geometric one is a necessity for solving $Dx = b$. The current status is the result of twenty years of continuous numerical developments, leading to AMG constructions for most of the discretisations in lattice QCD[3–6]. One realisation of an AMG solver for the Wilson-Dirac discretisation of the Dirac equation is the aggregation-based adaptive domain decomposition multigrid method[3] (DD-αAMG), with an extension for twisted mass fermions available[7,8].

The multigrid hierarchy built by DD-αAMG allows one not only to boost the performance of linear solves but also of methods for other problems arising in lattice QCD simulations, e.g., partial eigendecompositions of D and the approximation of traces such as $\mathrm{tr}(D^{-1})$. To develop and test those methods, the resources and support provided by the Jülich Supercomputing Centre have been crucial.

This paper is structured as follows. We review the basics of the DD-αAMG framework in Sec. 2 along with some recent improvements of the corresponding library, followed by a discussion in Sec. 3 on how trace estimation for $\mathrm{tr}(D^{-1})$ can make use of a multigrid hierarchy for variance reduction. Sec. 4 then presents the overlap discretisation and outlines methods that can be used for speeding up linear solves in that context. We close with a collection of some numerical results in Sec. 5.

2 Aggregation-Based Domain Decomposition MG for Lattice QCD

We start this section by giving a brief introduction to AMG, from which we then are able to give a general overview of the DD-αAMG framework. Then, more recent improvements of the library, namely coarsest-level improvements and GPU portations, are discussed.

2.1 Algebraic Multigrid

Multigrid solvers rely on the interplay of a *smoother* and a *coarse grid correction*. Smoothers are methods such as Gauss-Seidel or GMRES[9]. Those are generally very good at suppressing the part of the error that is connected to the eigenvalues of D with largest magnitude but they start plateauing in their error reduction after a few iterations. The reason for this stagnation is that components of the error in the direction of the eigenvectors whose eigenvalues are the smallest in magnitude are largely unaffected. The task of the coarse grid correction is thus to reduce the low modes of the error. Multigrid methods construct a hierarchy of coarse grid operators D_ℓ at different levels ℓ, prolongation and restriction operators P_ℓ and R_ℓ between the levels. The coarse grid correction then approximates the error at level ℓ as a prolongation of the equation for the restricted residual at level $\ell + 1$. Typically, one uses a Galerkin construction, i.e. $D_{\ell+1} = R_{\ell+1} D_\ell P_{\ell+1}$, and the range of P_ℓ should well approximate small eigenmodes of D_ℓ.

2.2 DD-αAMG

Domain decomposition aggregation-based algebraic multigrid (DD-αAMG)[3] is both an algorithmic framework and a code[10] for solving linear systems in lattice QCD. It targets the Wilson-Dirac discretisation with a clover improvement and implements an *aggregation-based* AMG[11,12]. The aggregation-based construction relies on the concept of *local coherence*[13], which states that many low modes of D can be approximately obtained from just a few low modes by looking at the local behaviour of those few modes. This construction allows us to have many approximate low modes of $D = D_1$ as columns of the prolongator P_1, with a sparse structure in P_1 leading to a coarse-grid Dirac operator D_2 resembling D_1 in its nearest-neighbour structure, and this is repeated on the further levels.

The DD-αAMG solver has *setup* and *solve* phases. During the setup phase, it builds the multigrid hierarchy, i.e. the operators D_ℓ, P_ℓ and $R_\ell = P_\ell^H$. The solve phase then makes use of this hierarchy to solve one or more linear systems, i.e., for one or more right-hand sides in a sequence.

During the solve phase, the cycling strategy is K-cycles[14], i.e. we use flexible GMRES (FGMRES)[15] to wrap every level of the multigrid solver. In the overall method, this K-cycle multigrid is then used as a preconditioner of an outer iteration which is, again, FGMRES. The outer preconditioned FGMRES is solved up to the desired relative residual tolerance, e.g., 10^{-12}. On the coarser levels of the K-cycle, the linear systems for the error representation at that level are solved to a relative residual tolerance of 10^{-1}.

The smoother in DD-αAMG could originally be one of either GMRES or SAP, but more recently this has been extended to include GCR and modified Richardson[16]. If the parameters of the multigrid solver are chosen appropriately, the number of iterations for the finest-level FGMRES to reach the desired tolerance will remain essentially constant

as the conditioning $\kappa(D)$ of D becomes larger. When $\kappa(D)$ increases, the condition on the coarser levels will also increase, which impacts in particular the coarsest level. If the coarsest level is solved with a standard iterative method, this requires many iterations for each coarse level solve and results in increased execution times.

DD-αAMG multigrid hierarchy is also useful for other tasks than linear systems. For example, an eigensolver was devised based on Generalized Davidson where the correction equation is solved with DD-αAMG[19]. We discuss in Sec. 3 more recent work where we use a multilevel decomposition in the stochastic estimation of the trace of the inverse of D.

2.3 Coarsest Level Improvements

When a multigrid method for a certain problem is built, the standard recipe is to do the coarse-grid correction via a recursive call of the same two-level method but on $D_c x_c = Rr$, and to continue this recursion until the coarse-grid operator is small enough so that a direct method, e.g., an LU factorisation can be used. The K-cycles used in lattice QCD multigrid solvers render having too many levels counter-efficient, with some reasons for this being thrashing, i.e., a lack of cash reusage, having idle processes when using many nodes, and the dominance of communications at some point when going for strong scaling. Hence, having two or three levels has become the standard in this field.

When three levels are used and as $\kappa(D)$ grows, the coarsest level can become very ill-conditioned; this can be particularly extreme for twisted mass fermions[8]. In recent work[17] the coarsest-level solver in the twisted mass version of DD-αAMG has been extended from restarted GMRES to preconditioned GCRO-DR. The latter method uses a polynomial preconditioner inside GCRO-DR[18], which in turn consists of an inner-outer method where deflation is used and the deflation subspace is updated via a recycling strategy, implying that sequences of linear systems can be solved, i.e., one can transfer deflation data from one linear system to the next. In Sec. 5 we show some of the most important results stemming from those coarsest-level improvements.

2.4 DD-αAMG on GPUs

Recent work[16,20] has extended DD-αAMG, via `CUDA`[21], to offload some of its critical sections to GPUs. This allows running the solver on heterogeneous machines with either Nvidia or AMD accelerators, with the latter being possible due to `HIP`[22]. More extensions also include a broader set of smoothers[16], where in particular Richardson iteration is shown to be competitive against the best smoother (that is, SAP) and with its additional benefits of low-memory requirements and simplicity of implementation. We discuss the current status of the CPU+GPU version of DD-αAMG in Sec. 5.

3 Multilevel Methods for Trace Estimation

The problem of estimating the trace $\mathrm{tr}(f(D))$ of a matrix function stochastically, such as $f(D) = D^{-1}$ with $D \in \mathbb{C}^{n\times n}$, arises in various fields including machine learning, network analysis, and particularly in lattice QCD where D is the Dirac operator or a composition of the Dirac operator with additional discrete operators. Directly computing the trace by solving n linear systems is infeasible for large matrices due to the

excessive computational cost. A common alternative is Hutchinson's method, where Rademacher (Z_2) or more generally Z_k random vectors x_i are used to approximate the trace as $\mathrm{tr}(D^{-1}) \approx \frac{1}{N}\sum_{i=1}^{D} x_i^* D^{-1} x_i$. The variance of this stochastic estimator decreases only as $O(1/N)$, making it computationally expensive for achieving high precision. We now discuss techniques for variance reduction and in particular we explain how the multigrid hierarchy built in DD-αAMG can be used for such purpose.

3.1 Variance Reduction Techniques

Variance reduction techniques makes use of the connection between the variance of the stochastic estimator and the Frobenius norm, which for Rademacher vectors is given by

$$\mathbb{V}[x^* D^{-1} x] \propto \|\mathrm{offdiag}(D^{-1})\|_F^2,$$

where the $\|\cdot\|_F$ is the Frobenius norm. Since $\|\mathrm{offdiag}(D^{-1})\|_F^2 = \sum_{i=1}^{n} \sigma_i^{-2} - \sum_{i=1}^{n} |(D^{-1})_{ii}|^2$, the small singular values of D dominate the variance of the stochastic estimator. By deflating those, the variance can be significantly reduced. The deflation scheme achieves this by constructing a projector Π with range approximating the small singular modes of D[23,24] and using the decomposition $\mathrm{tr}(D^{-1}) = \mathrm{tr}(D^{-1}(I-\Pi)) + \mathrm{tr}(\Pi D^{-1})$.

3.2 Multigrid Multilevel Monte Carlo

A different approach for variance reduction in the stochastic estimation of a variable f is the Multilevel Monte Carlo (MLMC) method[25]. It reduces variance by splitting the random variable f into contributions across multiple levels: the random variable is decomposed as $f = \sum_{\ell=1}^{L} g_\ell$, where each g_ℓ represents the contribution at level ℓ. The expected value $\mathbb{E}[f]$ is approximated by summing these contributions across all levels, with the number of samples N_ℓ at each level being chosen to balance the variance $\mathbb{V}(g_\ell)$ and the cost C_ℓ. This leads to an efficient overall estimator that reduces the computational cost compared to standard Monte Carlo methods. MLMC methods based on polynomial sequences and frequency splitting have been recently used on the problem of trace estimation for the inverse Dirac operator[26,27].

In 2022[28], we suggested the multigrid multilevel Monte Carlo (MGMLMC) method, a multilevel technique that makes use of the multigrid hierarchy for linear solvers as described in Sec. 2. A multilevel decomposition for $\mathrm{tr}(D^{-1})$ is obtained as

$$\mathrm{tr}(D^{-1}) = \sum_{\ell=1}^{L} \mathrm{tr}\left(D_\ell^{-1} - P_{\ell+1} D_{\ell+1}^{-1} R_{\ell+1}\right) + \mathrm{tr}(D_L^{-1}),$$

where we made use of the fact that the prolongations and restrictions are unitary in the QCD context, since they are obtained via aggregation.

Each term in the above sum is now estimated stochastically, using random vectors of increasingly smaller size as ℓ increases. Taking into account that by the algebraic multigrid construction the range of $P_{\ell+1}$ contains or approximates many of the low modes of D_ℓ we effectively produce an (almost) cancellation of those modes in the term $D_\ell^{-1} - (P_{\ell+1} D_{\ell+1}^{-1} R_{\ell+1})$ and thus reduce the variance of the stochastic estimator for this term.

3.3 Multigrid Multilevel Monte Carlo for a Lattice with a Displacement

In 2023 we built upon the MGMLMC method and incorporated a displacement matrix $\widetilde{P}$ into the estimation of the trace[29]. These traces emerge, e.g., in lattice QCD when computing helicity parton distribution functions of the proton[30]. The problem involves addressing the off-diagonal elements that emerge due to the displacement along one of the spacetime dimensions of the lattice. This leads to the estimation of the trace of the displaced operator, expressed as $\mathrm{tr}(D^{-1}\widetilde{P}^H)$, where $\widetilde{P}$ represents a permutation matrix. When applying two-level MGMLMC to this problem, we obtain

$$\mathrm{tr}(D^{-1}\widetilde{P}^H) = \mathrm{tr}\left(\left(D_1^{-1} - P_1 D_2^{-1} P_1^H\right)\widetilde{P}^H\right) + \mathrm{tr}(D_2^{-1} P_1^H \widetilde{P}^H P_1).$$

The application of this two-level method is natural, as $\widetilde{P}$ is unitary and therefore $D^{-1}\widetilde{P}$ and D^{-1} share the same singular values. But the effectiveness of a three-level method is not immediately evident. We have found[29] that, not only Hutchinson on $\mathrm{tr}((D_2^{-1} - P_2 D_3^{-1} P_2^H) P_1^H \widetilde{P}^H P_1)$ experiences a significant variance reduction compared to $\mathrm{tr}(D_2^{-1} P_1^H \widetilde{P}^H P_1)$, but also $P_1^H \widetilde{P}^H P_1$ serves as a deflation factor in the two-level decomposition above, i.e., Hutchinson on $\mathrm{tr}(D_2^{-1} P_1^H \widetilde{P}^H P_1)$ has a lower variance compared to $\mathrm{tr}(D_2^{-1})$. Hence, a multilevel MGMLMC is well suited for this problem, at least from the variance reduction point of view. The construction of a cost model to assess whether this multilevel computation leads to a total cost reduction is not yet finished and part of current work.

4 Sign Function in the Overlap Discretisation

The *overlap Dirac operator*[31,32] at nonzero chemical potential μ, $D_{ov}(\mu)$, preserves a form of chiral symmetry on the lattice, an important physical property, while other discretisations as, e.g., the Wilson-Dirac operator do not. The overlap Dirac operator takes the form

$$D_{ov}(\mu) = \rho I + \Gamma_5 \mathrm{sign}(\underbrace{\Gamma_5 D(m_w, \mu)}_{=:Q(m_w,\mu)}).$$

Here, Γ_5 is a simple diagonal matrix which acts as the identity on spinor components belonging to spins 1 and 2 and as the negative identity on those belonging to spins 3 and 4, and $\rho \in (0, 1)$ is a mass parameter, typically close to 1.

In the argument of the sign function, $D(m_w, \mu)$ is the massless Dirac-Wilson operator with a shift $m_w \in (-2, 0)$ chosen to improve its locality, and a chemical potential μ. It is the presence of $\mu \neq 0$ that makes $Q(m_w, \mu)$ non-Hermitian; see Ref. 33. For notational simplicity, we abbreviate $Q(m_w, \mu)$ as Q_μ and $D(m_w, \mu)$ as D_μ from now on.

The sign function $\mathrm{sign}(Q_\mu)$ of Q_μ is defined in the usual matrix function sense. Although Q_μ is sparse, $\mathrm{sign}(Q_\mu)$ is a full matrix and therefore cannot be computed directly. Rather, in an iterative solver for the overlap operator, one has to compute the action of the sign function on a new vector in each iterative step. One can express $\mathrm{sign}(Q)$ as $Q(Q^2)^{-1/2}$, so that the computational burden in $\mathrm{sign}(Q)b$ resides in $(Q^2)^{-1/2}b$, and using this has become standard in overlap computations. The action of the inverse square root (of Q_μ^2) has now to be computed in an further, *inner* iterative procedure, typically based

on the Arnoldi process. To keep computational cost of overlap simulations at a bearable level, it is therefore of primordial interest to compute these actions of the sign function as efficiently as possible.

We recently introduced polynomial preconditioning[34] as a novel technique to accelerate sign function computations. The idea is to first cheaply build a polynomial $q(Q_\mu^2)$ which approximates $(Q_\mu^2)^{-1/2}$ and then preconditioning the sign function application by using $\text{sign}(Q_\mu)b = Q_\mu q(Q_\mu^2)(q^2(Q_\mu^2)Q_\mu^2)^{-1/2}b$. Polynomial preconditioning allows to avoid cumbersome restart techniques within the Arnoldi process, it limits the work spent in full orthogonalisations required within the Arnoldi process and, in a parallel environment, the number of global reductions due to dot products is kept low.

5 Numerical Experiments

The results in this section were obtained on JUWELS at JSC. Different lattice QCD configurations were used for the different numerical experiments, which is indicated in more detail within each of the following subsections.

5.1 Effect of Coarsest-Level Improvements on Solver Performance

We upgraded the coarsest-level solver in DD-αAMG from a restarted GMRES to preconditioned GCRO-DR with a polynomial preconditioner, improving with this the convergence of the coarsest-level solver and reducing the impact of global communication on it[17]. Fig. 1 illustrates the effect of these improvements on the Wilson-Dirac operator on a 128×64^3 lattice, generated by the CLS collaboration[36]. We compare the total execution times of the original and improved DD-αAMG solvers as m_0 approaches its critical value. The solver with coarsest-level preconditioning and deflation maintains a stable performance across the entire range of m_0 values. This indicates that the new techniques mitigate critical slowing down, resulting in a nearly constant execution time even in highly ill-conditioned regimes, a highly sought after feature in a multigrid solver.

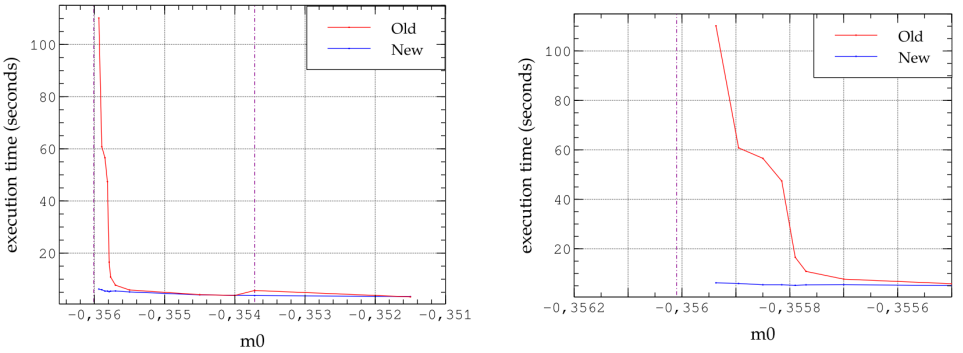


Figure 1. Total execution time of the solve phase in DDαAMG as the system becomes more ill-conditioned as a function of m_0. Comparison between restarted GMRES on coarsest level (Old) and GCRO-DR with polynomial preconditioning and deflation (New)[17].

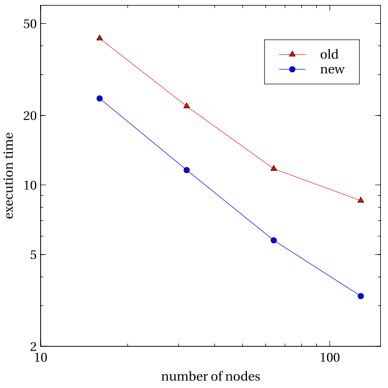


Figure 2. Strong scaling for twisted mass fermions. Old: $\delta = 8.0$, with restarted GMRES as the coarsest-level solver. New: $\delta = 1.0$, polynomial preconditioner and deflation with GCRO-DR on coarsest-level. 32 MPI processes per node and 1 thread per process[17].

Fig. 2 extends this to the twisted mass discretisation and shows how strong scaling can improve in the twisted mass DD-αAMG solver when using the coarsest-level improvements. The figure highlights that, while the old solver exhibits limited scalability due to poor convergence at the coarsest level, our new solver achieves a speedup of up to 2.6 in total execution time when increasing the node count up to 128. This improvement is particularly noteworthy because it enabled us to use the original twisted mass parameter μ at the coarsest level instead of the value $\mu_c = \delta \cdot \mu$ that was artificially introduced in the old solver to cope with the severe ill-conditioning in this discretisation. Restoring μ simplifies the setup and leads to a more accurate representation of the twisted mass discretisation across all levels. By employing a polynomial preconditioner and deflation through GCRO-DR, we reduced the sensitivity of the coarsest-level solver to ill-conditioning, enabling better strong scaling behaviour and overall efficiency in solving twisted mass systems.

5.2 DD-αAMG on GPUs

The original code of DD-αAMG[10] has been extended[20] to include GPU developments. The SAP smoother has been fully ported via `CUDA C` to run on GPUs[38]. Time comparisons are shown in Tab. 1, which come from solving linear systems in DD-αAMG with the E250 configuration from the CLS collaboration[36]. The local (per node) lattice is different for CPU and GPU in Tab. 1, which reflects the typical use of strong scaling in CPU executions

	lattice per node	time (seconds)
GPU	$16 \times 64 \times 96^2$	0.048
CPU	$8 \times 32 \times 96^2$	0.098

Table 1. Average execution time of a single application of finest-level SAP. The CPU run was done on JUWELS Cluster and the GPU one on JUWELS Booster. The speedup of CPU to GPU is 8.17.

n_{defl}	**Hutch.** ($\times 10^3$)	**MGMLMC** ($\times 10^3$)
0	145.5	110.7
32	131.8	109.8
128	120.5	105.5

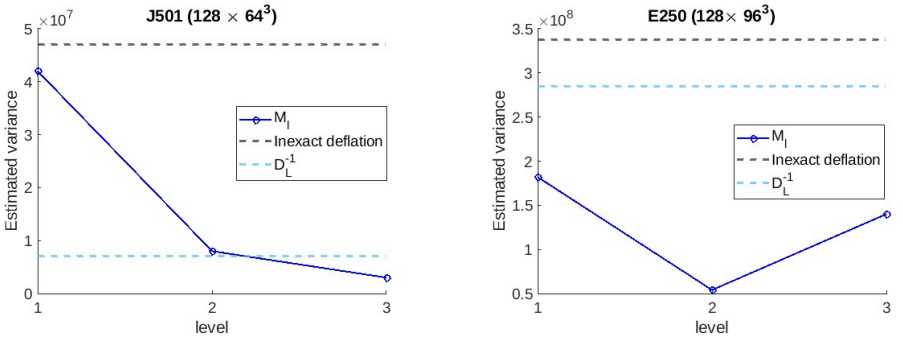
Table 2. Variance of $\Pi_D D_1^{-1} \widetilde{P}^H$ (middle column) and $\Pi_M M_1 \widetilde{P}^H = \Pi_M (D_1^{-1} - P_1 D_2^{-1} P_1^H) \widetilde{P}^H$ (rightmost column), with Π_D deflating approximate left singular vectors of D_1^{-1} and Π_M approximate left singular vectors of $M_1 = D_1^{-1} - P_1 D_2^{-1} P_1^H$. The displacement on the lattice is of five sites along the z dimension. Taken from Ref. 29.

versus the use of considerably lesser nodes when using GPUs, as the latter do not scale well when local lattices become smaller. The GPU code is around 8x faster in this comparison, with one application of SAP on four times the local volume running in roughly half the time.

Building on the SAP `CUDA` code, DD-αAMG has been further extended[16,37]. In particular, we have found that odd-even Richardson as a smoother leads to a multigrid solver which is only 10% slower compared to using SAP, with the former smoother being significantly simpler in terms of implementation and with minimal memory requirements.

5.3 Comparison of Multigrid MLMC and Deflation Methods for the Trace

Our first application of MGMLMC in lattice QCD was done in the presence of a displacement on the lattice[29] (see Sec. 3.3). For that case, using a configuration of size 16^4, Tab. 2 shows a comparison in variance reduction of the Hutchinson estimator for the first level-difference in MGMLMC and of deflated Hutchinson. In there, n_{defl} is the number of (inexactly) deflated vectors for both methods, as deflation has also been applied on top of MGMLMC. The MGMLMC method achieves consistently lower variance compared to the standard deflated Hutchinson approach for the same n_{defl}. Even without



(a) (b)

Figure 3. Variances for the operator differences $M_\ell = D_\ell^{-1} - P_\ell D_{\ell+1}^{-1} P_\ell^H$ at the various levels ℓ and coarsest level operator D_L^{-1}, compared against the variance observed with inexact deflation.

deflation, MGMLMC provides a notable variance reduction, and deflated Hutchinson with $n_{\text{defl}} = 128$ does not reach the variance reduction that MGMLMC with $n_{\text{defl}} = 0$ does.

We also applied MGMLMC for the estimation of $\text{tr}(D^{-1})$ on larger and more ill-conditioned CLS[36] configurations, namely J501 (128×64^3) and E250 (128×96^3). To manage this, we work on the JUWELS Cluster, using up to 27 nodes, each with the 2×24 available cores. Improvements at the coarsest level of MGMLMC play an important role, as they accelerate the many solves required at every level, especially in the case of the E250 lattice. Fig. 3 shows results for the variance on those lattices using a 4-level multigrid hierarchy. We see that the variance at the finest difference is smaller than that observed with inexact deflation, providing a clear advantage as MGMLMC bypasses the need for computing approximate singular vectors. For the J501 configuration (left panel), the variance progressively decreases at coarser levels, significantly improving the overall efficiency of the multilevel approach. Each solve on coarser grids is approximately eight times faster than on the preceding finer level, resulting in a substantial reduction in computational costs of samples as the levels progress. In the case of the more ill-conditioned E250 configuration (right panel), the majority of the variance shifts to the coarsest levels, where the computational cost is minimal.

5.4 Sign Function in the Overlap Discretisation

We present results for a 64×32^3 lattice on JUWELS coming from a physically relevant ensemble that was provided by the lattice QCD group at the University of Regensburg.

The left part of Fig. 4 displays the relative error as a function of the iteration counts, and Tab. 3 gives operations counts and timings, both for approximating $(Q_\mu^2)^{-1/2}b$ using various polynomial preconditioning degrees up to a relative error of $4.0{\cdot}10^{-5}$. The results illustrate the significant reduction in iteration count achieved by using polynomial preconditioning compared to the unpreconditioned Arnoldi method (first row). Increasing the polynomial degree d results in a higher number of matrix-vector multiplications (mvms) per iteration in exchange for a reduction of the number of inner products, leading to a sub-

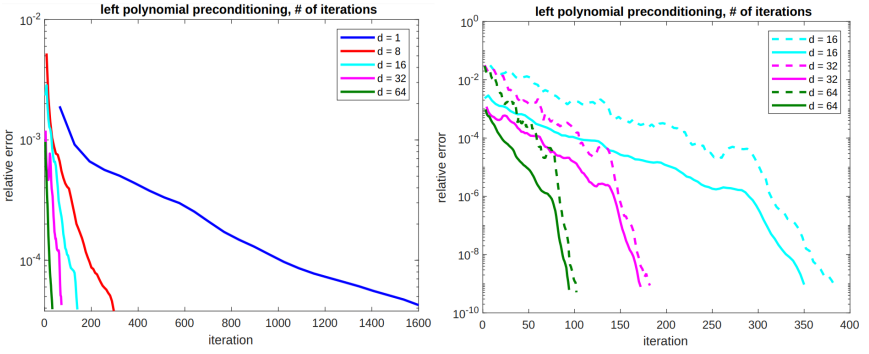


Figure 4. Results for approximating $(Q_\mu^2)^{-1/2}b$ using Arnoldi-preconditioning polynomials of various degrees $d-1$. **Left**: Relative error as a function of the Arnoldi basis size. **Right**: close up for degrees 16 to 64. Solid lines: Relative error $\|f_m - f^*\|/\|f^*\|$. Dashed lines: Error measure $\|f_m - f_{m+k}\|/\|f_{m+k}\|$ with $k = 64/d$. Taken from Ref. 34.

d	Iterations	MVMS	Inner Prods	Time (64 nodes) [s]	Time (256 nodes) [s]
1	1600	3200	1,279,200	127.8	105.8
8	296	8910	42,510	25.7	9.8
16	140	8742	9,991	12.3	7.8
32	72	9198	3,125	11.6	7.4
64	33	8636	2,578	10.6	5.5

Table 3. Timings and operation counts for approximating $((Q_\mu)^2)^{-1/2}b$ with Arnoldi-preconditioning polynomials q of various degrees $d-1$. Taken from Ref. 34.

stantial reduction of the overall execution time due to much less orthogonalisation times. The table also shows that for $d = 64$, the preconditioned method achieves the fastest execution time, being approximately 12 times faster than the unpreconditioned method on 64 nodes, and 19 times faster on 256 nodes.

The right panel in Fig. 4 lowers the relative error down to $3.0{\cdot}10^{10}$. The unpreconditioned method takes 6,000 iterations, hence a comparison of the preconditioned methods against the $d = 0$ case is skipped in right side of Fig. 4. The results emphasise the efficiency of polynomial preconditioning in reducing the computational cost of orthogonalisation, the dominant factor when a large number of iterations is required.

Acknowledgements

The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (`www.gauss-centre.eu`) for funding this project by providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS Supercomputer JUWELS at Jülich Supercomputing Centre (JSC), under the projects with id CHWU29 and MUL-TRA. This work is partially supported by the German Research Foundation (DFG) research unit FOR5269 "Future methods for studying confined gluons in QCD".

References

1. J. Finkenrath, *Review on algorithms for dynamical fermions*, 2024, arXiv:2402.11704.
2. R. Babich et al., *Adaptive multigrid algorithm for the lattice Wilson-Dirac operator*, Physical Review Letters **105**, 201602, 2010.
3. A. Frommer et al., *Adaptive aggregation-based domain decomposition multigrid for the lattice Wilson-Dirac operator*, SIAM J. Sci. Comput. **36**, A1581-A1608, 2014.
4. J. Brannick et al., *Adaptive multigrid algorithm for lattice QCD*, Phys. Rev. Lett. **100**, 041601, 2008.
5. R. C. Brower et al., *Multigrid for chiral lattice fermions: Domain wall*, Phys. Rev. D **102**, 094517, 2020.
6. J. C. Osborn at al., *Multigrid solver for clover fermions*, 2010, arXiv:1011.2775.
7. S. Bacchio, *DDalphaAMG (twisted mass): A MATLAB implementation of the adaptive multigrid algorithm for domain decomposition*, GitHub repository, 2023, `https://github.com/sbacchio/DDalphaAMG`.

8. S. Bacchio, *Simulating maximally twisted mass fermions at the physical point with multigrid methods*, PhD thesis, University of Wuppertal, 2019, https://elekpub.bib.uni-wuppertal.de/urn/urn:nbn:de:hbz:468-20190703-114133-8.
9. Y. Saad, *Iterative methods for sparse linear systems*, SIAM, 2003.
10. M. Rottmann et al., *DD-αAMG*, GitHub repository, 2016, https://github.com/mrottmann/DDalphaAMG.
11. F. Chatelin and W. L. Miranker, *Acceleration by aggregation of successive approximation methods*, Linear Algebra Appl. **43**, 17-47, 1982.
12. M. Brezina et al., *Adaptive smoothed aggregation (α SA)*, SIAM J. Sci. Comput. **25**, 1896-1920, 2004.
13. M. Lüscher, *Local coherence and deflation of the low quark modes in lattice QCD*, J. High Energy Phys. **2007**, 081, 2007.
14. Y. Notay and P. S. Vassilevski, *Recursive Krylov-based multigrid cycles*, Numer. Linear Algebra Appl. **15**, 473-487, 2008.
15. Y. Saad, *A flexible inner-outer preconditioned GMRES algorithm*, SIAM J. Sci. Comput. **14**, 461-469, 1993.
16. L. He, G. Ramirez-Hidalgo, and K.-L. Zhang, *Extending DD-αAMG on heterogeneous machines*, 2024, arXiv:2407.08092.
17. J. Espinoza-Valverde et al., *Coarsest-level improvements in multigrid for lattice QCD on large-scale computers*, Comput. Phys. Commun. **292**, 108869, 2023.
18. M. L. Parks et al., *Recycling Krylov subspaces for sequences of linear systems*, SIAM J. Sci. Comput. **28**, 1651-1674, 2006.
19. A. Frommer et al., *A multigrid accelerated eigensolver for the Hermitian Wilson-Dirac operator in lattice QCD*, Comput. Phys. Commun. **258**, 107615, 2021.
20. M. Rottmann et al., *CPU+GPU DD-αAMG*, GitHub repository, 2023, available at: https://github.com/Gustavroot/DDalphaAMG. Commit: develop.
21. NVIDIA, *NVIDIA CUDA - NVIDIA Docs*, 2007, available at: https://docs.nvidia.com/cuda/doc/index.html, accessed 29 June 2024.
22. Advanced Micro Devices, Inc., *HIP*, GitHub repository, 2008, Available at: https://github.com/ROCm/HIP, Commit: develop.
23. A. S. Gambhir, A. Stathopoulos, and K. Orginos, *Deflation as a method of variance reduction for estimating the trace of a matrix inverse*, SIAM J. Sci. Comput. **39**, A532-A558, 2017.
24. E. Romero, A. Stathopoulos, and K. Orginos, *Multigrid deflation for lattice QCD*, J. Comp. Phys. **409**, 109356, 2020.
25. M. B. Giles, *Multilevel Monte Carlo methods*, Acta Numerica **24**, 259-328, 2015.
26. P. Lashomb et al., *Multipolynomial Monte Carlo for trace estimation in lattice QCD*, Comput. Phys. Commun. **300**, 109163, 2024.
27. T. Whyte et al., *Optimizing shift selection in multilevel Monte Carlo for disconnected diagrams in lattice QCD*, Comput. Phys. Commun. **294**, 108928, 2024.
28. A. Frommer, M. Khalil Nasr, and G. Ramirez-Hidalgo, *A multilevel approach to variance reduction in the stochastic estimation of the trace of a matrix*, SIAM J. Sci. Comput. **44**, A2536-A2556, 2022.

29. G. Ramirez-Hidalgo, *Multigrid multilevel Monte Carlo for a lattice with a displacement*, PAMM **23**, 202300242, 2023.
30. C. Alexandrou et al., *Flavor decomposition for the proton helicity parton distribution functions*, Physical Review Letters **126**, 102003, 2021.
31. H. Neuberger, *Exactly massless quarks on the lattice*, Phys. Lett. B **417**, 141-144, 1998.
32. P. Hernandez, K. Jansen, and M. Lüscher, *Locality properties of Neuberger's lattice Dirac operator*, Nucl. Phys. B **552**, 363-378, 1999.
33. J. Bloch and T. Wettig, *Overlap Dirac operator at nonzero chemical potential and random matrix theory*, Phys. Rev. Lett. **97**, 012003, 2006.
34. A. Frommer et al., *Polynomial preconditioning for the action of the matrix square root and inverse square root*, Electron. Trans. Numer. Anal. **60**, 381-404, 2024.
35. J. Brannick et al., *Multigrid preconditioning for the overlap operator in lattice QCD*, Numer. Math. **132**, 463-490, 2016.
36. D. Mohler, S. Schaefer, and J. Simeth, *CLS 2+1 flavor simulations at physical light- and strange-quark masses*, EPJ Web Conf. **175**, 02010, 2018.
37. T. Matthaei, *Accelerating lattice QCD simulations using GPUs*, 2024, arXiv:2407.00041.
38. G. A. Ramirez Hidalgo, *Multilevel algorithms in Lattice QCD for Exascale Machines*, Ph.D. thesis, Bergische Universität Wuppertal, 2022, `https://elekpub.bib.uni-wuppertal.de/doi/10.25926/BUW/0-21`.

OpenGPT-X: Leveraging GCS Infrastructure for European Large Language Models

Jan Ebert[1], Mehdi Ali[2,3], Michael Fromm[2,3], Klaudia Thellmann[4], Alexander Arno Weber[2,3], Richard Rutmann[2,3], Charvi Jain[2,3], Max Lübbering[2,3], Daniel Steinigen[2], Johannes Leveling[2], Katrin Klug[2], Jasper Schulze Buschhoff[2], Lena Jurkschat[4], Hammam Abdelwahab[2], Benny Jörg Stein[2], Karl-Heinz Sylla[2], Pavel Denisov[2], Nicolo' Brandizzi[2], Qasid Saleem[2], Anirban Bhowmick[2], Chelsea John[1], Pedro Ortiz Suarez[5], Malte Ostendorff[5], Alex Jude[2], Lalith Manjunath[4], Samuel Weinbach[7], Carolin Penke[1], Oleg Filatov[1], Shima Asaadi[6], Fabio Barth[5], Rafet Sifa[2], Fabian Küch[6], Andreas Herten[1], René Jäkel[4], Stefan Kesselheim[1], Joachim Köhler[2], Nicolas Flores-Herr[2], and Georg Rehm[5]

[1] Jülich Supercomputing Centre, Forschungszentrum Jülich, 52425 Jülich, Germany
E-mail: s.kesselheim@fz-juelich.de

[2] Fraunhofer Institute for Intelligent Analysis and Information Systems, 53757 Sankt Augustin, Germany

[3] Lamarr Institute for Machine Learning and Artificial Intelligence, 44227 Dortmund , Germany

[4] Technische Universität Dresden, 01069 Dresden, Germany

[5] Deutsches Forschungszentrum für Künstliche Intelligenz, 10559 Berlin, Germany

[6] Fraunhofer-Institut für Integrierte Schaltungen, 91058 Erlangen, Germany

[7] Aleph Alpha, 69123 Heidelberg, Germany

In high performance computing, the training of large language models (LLMs) has become a prevalent workload, accompanied by an increasing public interest since the release of services such as ChatGPT. The OpenGPT-X project, a collaborative initiative between academic and industrial partners, represents one of the first German efforts to develop an open, freely available and multilingual LLM solution. This paper presents technical details describing the efficient use of Gauss Centre for Supercomputing infrastructure, in particular the supercomputer JUWELS Booster, for the task of pre-training a neural network based on the transformer architecture. We analyse the training's scaling behaviour in terms of throughput, and examine the impact of hyperparameters related to model parallelisation across different dimensions. Key training decisions, including the selection of model activation function, optimiser, positional embedding, and normalisation layers, are evaluated through ablation studies in the compute-equivalent setting and recommendations provided.

1 Introduction

Large language models (LLMs) are transforming the way humans manage and process information. This trend has been apparent at least since the US-based company OpenAI published GPT-3, a closed-source language model, in May 2020[1]. While this was still a long way from the end-user friendly chat systems provided today, the surprising capabilities in generating natural language struck scientists. However, the developments at OpenAI and other competitors remained behind closed curtains. In response to both the

successes and concerns about transparency, the OpenGPT-X project was initiated, funded by the German Federal Ministry for Economic Affairs and Climate Action (BMWK). Since early 2022, OpenGPT-X has accepted the challenge to compete with big industrial players. The project aims to train and publish LLMs with a focus on European languages, providing full transparency of the training process. Alongside expertise and curated datasets, large-scale computational resources are a crucial ingredient for the success of the endeavour. As a member of the consortium, the Jülich Supercomputing Centre (JSC) provides access to its computational infrastructure, supports its efficient use, and implements optimisations to enhance performance. The JUWELS (Jülich Wizard for European Leadership Science) supercomputer, Jülich's flagship system, was employed, particularly the Booster module as it contains numerous powerful A100 Nvidia GPUs, that are well-suited for model training.

During the project, critical decisions were required to define the training process, including model architecture, tokenisation, training paradigm and dataset composition, each having a potential impact on training efficiency and model performance. Given the substantial computational resources required for training an LLM, it was essential to make informed choices to avoid wasting resources. To support these decisions, we conducted several studies to assess the impact of each choice. In line with our commitment to transparency, the findings from these studies have been published in independent research papers. We have shown that the tokenisation procedure, the first step in the processing chain, needs to be adjusted to the language composition of the considered dataset[2]. To gain this insight, a series of smaller models was trained. Furthermore, we investigated the interplay between instruction tuning and multilinguality[3]. A major outcome of the OpenGPT-X project is the Teuken model family. In a detailed analysis, we compare its performance to other open sources models[4].

The design space to define the training of large machine learning (ML) models contains parameters that mainly affect the training performance, in terms of throughput and GPU utilisation, as well as parameters, that have a large influence on the quality of the final model. Due to the size of the computational endeavour, certain decisions need to be made on a heuristic basis, relying on experience and community best practices. General architectural decisions, e.g., for a GPT-style decoder-only model (see below), could be readily made based on the planned application of the models. Relatively small changes in model architecture or hyperparameters, such as the choice of positional embeddings, can have a substantial influence on the overall model performance[5]. This paper highlights a series of experiments and provides quantitative results as a sound basis to guide rational decision-making when presented with these influential architecture choices. Parameters relevant for the training performance are studied by throughput-related benchmarks. Parameters affecting the final model quality were analysed systematically by training smaller models and evaluating their performance on select, well-established downstream tasks. These experiments, dubbed *ablations* in our project, have guided our architectural and hyperparameter choices and are worth being put forward for future reference. In this paper, we provide background information on LLMs (Sec. 2.1) and describe the computational environment (Sec. 2.2). An analysis of influential factors on the training performance is provided based on throughput measurements (Sec. 5), followed by an analysis of factors relevant for model quality in the compute-equivalent setting (Sec. 3). Furthermore, we summarise key findings regarding the trained models (Sec. 4). We conclude with a summary and outlook (Sec. 5).

2 Background

2.1 Transformer-Based LLMs

Transformers[6] are deep neural networks designed for sequence processing. By combining best practices such as an architecture designed for large-scale parallel distributed training,various methods for stabilising training, and information-rich embeddings, they have achieved popularity and managed to achieve state-of-the-art results in various domains such as text, image, or audio processing[7,1,8,9]. Their Self-Attention mechanism compares each sequence element to each other sequence element to generate an importance weight. This allows these networks to build information-rich graphs over long sequences, which are used to transmit information between arbitrarily distant sequence elements. While the Attention operation has a $\mathcal{O}(n^2)$ runtime complexity in the sequence length n, it is key to the Transformers' ability to efficiently solve various classes of problems. By learning and weighting multiple different representations in the latent Attention space (called multi-head Attention), Transformers expand their sequence interaction space and allow for different "perspectives" on a sequence to be combined. Recent research in the field of mechanistic interpretability has shown that Transformers are able to learn highly efficient algorithms, e.g., a dynamic programming algorithm for parsing context-free grammars[10].

With the advent of ChatGPT, Transformers have received another increase in popularity, this time for practical, user-facing language processing applications. The architecture used by ChatGPT and most other modern user-facing language models is a decoder-only Transformer. By not allowing the model to attend to subsequent sequence elements when predicting a sequence element (during training, all elements of a given sequence are predicted at once), i.e., by using only "causal" attention, the architecture becomes especially useful for efficient autoregressive generation tasks. This is because of the ability to cache parts of resource-intensive Attention calculations for earlier sequence elements. In this work, we concern ourselves with decoder-only Transformer models, but also use techniques originally applied on encoder-decoder Transformer models (i.e., the original architecture).

2.2 Computational Infrastructure

The goal of training a neural network is to find weights that encode an input-output mapping in the training data. During the forward pass, these weights are applied to the input data through a series of matrix multiplications. The weights are updated using a variant of stochastic gradient descent (the *optimiser*), with gradients computed via backpropagation[11]. Matrix multiplications are inherently parallel operations, making them particularly well-suited for massively parallel architectures such as GPUs. Consequently, GPUs are generally preferred over CPUs for both neural network training and inference. Although Nvidia currently holds a dominant position in the market, competitive GPUs from other manufacturers are also available. Moreover, accelerators based on the dataflow paradigm are emerging as alternatives, promising to be even better optimised for AI workloads[12].

Training an LLM necessitates vast amounts of text data, which is processed by the model during training, leading to substantial computational demands. These demands can only be met by highly distributed computing systems, such as supercomputers or extensive cloud resources. Efficient scaling to leverage these systems requires exploiting higher

levels of parallelism. Given the growing importance of this workload, newly procured exascale HPC systems are now being designed to optimally meet its requirements[13].

The model training and relevant experiments were conducted on JUWELS Booster. This cluster was installed in 2020, ranking 7th on the Top500 list at the time, making it the fastest supercomputer in Europe. With 936 GPU nodes and a high bandwidth network, it is ideally suited for training large-scale neural networks[14]. Each GPU node contains 4 NVIDIA A100 GPUs with 40GB memory and 4 NVIDIA Mellanox InfiniBand HDR200 adapters, with one adapter available per GPU. Within the DragonFly+ topology of the high-speed interconnect, 48 nodes make up a *cell*.

3 Training Ablation Experiments

In this section, we present an extended version of the training ablation results published in Ali *et al.* (2024)[4]. We performed medium-scale (Chinchilla-optimal[15] training of a 2.6B parameter model) ablation runs for various variables during training, such as the model's position embedding, the optimiser, or the learning rate. Our goal with these ablations was not to optimise the model in a vacuum. Instead, our goal was finding improvements in the compute-equivalent setting while confirming whether proposed modifications transfer to our codebase, which is not necessarily the case[16]. A summary of the results can be found in Tab. 1 .

Change	ARC Easy	HellaSwag	LAMBADA	Interpretation
-	0.535	0.355	0.503	0
SwiGLU	0.527	0.361	0.507	+
Untied in/out embedding	0.524	0.355	0.498	-
No `Linear` biases (33k st.)	-	-	-	+
No GPT-like weight init (39k st.)	-	-	-	-
Head scaling	0.527	0.356	0.493	-
No dropout (21k st.)	*0.492*	*0.334*	*0.414*	+?
RMSNorm	0.530	0.358	0.502	0
NoPE	0.516	0.351	0.486	-
ALiBi	0.527	0.349	0.486	-
GQA (2 groups)	0.513	0.346	0.459	?
Adan (4× base LR)	0.544	**0.374**	**0.522**	+?
2× learning rate	0.540	0.369	0.514	+
4× learning rate (48k st.)	***0.545***	*0.371*	*0.517*	+

Table 1. Selected evaluation results. Bold is best, underlined is better than baseline. Italic means the run was evaluated before finishing. Ablations without values were not evaluated (see the listed ablations below). The rightmost column contains a subjective interpretation/recommendation, where "+", "-", "0" indicate a positive, negative, and neutral interpretation, respectively. "?" indicates it is difficult to make a conclusive statement.

Our baseline for the ablation experiments was the best-performing 2.6B parameter model from Ali *et al.* (2023)[2]. To make experiments comparable, the training ablations

used the same training setup. For evaluation, we focused on 3 tasks: (1) ARC Easy[17], (2) HellaSwag, and (3) LAMBADA. The models often achieved better-than-random performance in these evaluation tasks, and because the tasks additionally test different abilities of the model, they were deemed a good-enough proxy measure of general downstream improvements. Many of the ablations required implementation contributions. For example, we implemented tensor-model-parallel SwiGLU layers, per-head scaling, or fused RMSNorm into the codebase, to name a few. Some of these have since become de-facto standard parts of many Transformer libraries, including upstream Megatron-LM. Factors that influenced the results of the ablations include that (1) job allocations on JUWELS Booster vary in determined layout, and (2) the parallel file system means I/O throughput varies. Average throughput may vary around up to a percent depending on system load and/or job layout. Since throughput was crucial to asserting compute-equivalence, it is important that this variance is minimal.

Due to hardware failures and resource constraints, we were unable to finish all runs and do fair evaluation comparisons. Since we believe our findings to be meaningful nonetheless, unfinished ablations are also included in our results.

To compare compute-equivalently, we used throughput-normalised times with regard to the average iterations per second of the baseline model's training to compare loss at given times. Specifically, we assume that every training iteration takes the average amount of time and thus apply the same proportional normalisation to every step. For the ablations conducted, this assumption is fine because there is no theoretical throughput variance between steps. We normalise across time using the following formula:

$$t_i(s) = s\frac{1}{\bar{T}_i}, \quad s'_i = \text{time-normalize}_i(s) = s\frac{t_i(s)}{t_{\text{baseline}}(s)},$$

where t is a time (e.g., in seconds), i is one of the ablation experiments (including the baseline), s is the iteration (step) to normalise time until, $\bar{T}$ are averaged throughput values (e.g., iterations per second), and s'_i is a time-normalised step for experiment i. This normalisation allows us to compare the training and validation loss at each point in time and express that a method with lower loss at the same point in time as the baseline is more compute-efficient up to that point in time with regard to that loss.

The following lists all ablated changes and additional information where appropriate:

1. **SwiGLU:** replace the first layer of the MLP part of the Transformer with a T5-style (i.e., without biases)[18,19] Swish-activated[20] gated linear unit layer[21,22].
2. **Untied in/out embedding:** learn separate weights for the input embedding and output "unembedding" layers[19].
3. **No `Linear` biases:** remove all bias terms in `Linear` layers[19]. The ablation ran until 33 000 steps and was not evaluated due to being deemed too far from completion.
4. **No GPT-like weight init:** whether to scale weight initialisation in layers that a residual path leads into as a function of depth[23]. The ablation ran until 39 000 steps and was not evaluated due to being deemed too far from completion.
5. **Head scaling:** multiply each Attention head's output by a learned scalar factor[24].
6. **No dropout:** disable dropout[25] in all layers[19]. This ablation only ran until 24 000 steps, the latest checkpoint used for evaluation being at 21 000 steps. While

it showed a strongly monotonic improvement over the baseline, it is especially hard to interpret improvement in training loss in the dropout vs. no dropout setting. We evaluated this change even though it was far from finishing training because its improvements were so drastic.

7. **RMSNorm:** replace LayerNorm normalisation layers[26] with root mean square layer normalisation layers[27].

8. **NoPE:** no position embedding[5]; completely remove position embeddings.

9. **ALiBi:** replace the Rotary position embedding[28] with the Attention with linear biases position embedding[29]. Note that we were not able to use our ALiBi kernelwe developed for this ablation due to the baseline model using Attention dropout, which we had not implemented support for in the ALiBi kernel. Since we would rather use the kernel than Attention dropout during an actual training (especially considering the "No dropout" ablation results), throughput-normalisation was not considered for this ablation out of fairness.

10. **GQA (2 groups):** replace multi-head Attention with grouped-query Attention[30] with 2 groups (i.e., 2 key/value heads).

11. **Adan (4× base LR):** replace the AdamW optimiser with the Adan optimiser[31], using the baseline's learning rate multiplied by 4. The increased learning rate was chosen based on previous small-scale experiments.

12. **2×/4× base LR:** use the baseline's learning rate multiplied by 2 or 4. The ablation with a factor 4 increase did not run until completion but was discontinued after 51 000 steps. The checkpoint used for evaluation was saved at 48,000 steps, 5 100 steps before the end of training, and was deemed "close enough" to completion to provide a fair evaluation comparison, especially due to its noticeable training loss improvements.

We decided to implement most of the "free lunch" improvements and some neutral results based around current research results at that point in time while disregarding some findings that were deemed too experimental and/or risky. Notably, we decided to use neither Adan nor 4× the learning rate despite these changes yielding the best results. Adan was not chosen due to its learning rate requiring further ablations and it carrying risks for a large-scale training. This is because there is not a lot of empirical evidence for Adan's performance at a large scale. Similarly, the higher learning rate was not used, and instead the value for the 7B training was taken from Llama-2[32]. Due to the difficulty in finding an optimal learning rate for both an increased parameter amount and training horizon, it is hard to judge whether using a scale-adjusted larger learning rate would have been more optimal[33]. Due to our inexperience at training at this scale, we were also worried about possible convergence problems if done incorrectly. Grouped-query attention underperformed in terms of the compute-equivalent setting, but we decided to use it for the large-scale training because of its significant benefits for model inference. That is, by including this change, we optimised for the post-training setting.

In summary, the chosen changes are: SwiGLU, no biases, no dropout during pretraining, RMSNorm, GQA (with 2 groups).

4 The Teuken Model

The previous ablations defined the path towards Teuken-7B-Base[4], a 7B parameter LLM that was trained on 4T tokens, covering *all official 24 European languages*. We additionally trained a multilingual instruction-tuned version Teuken-7B-Instruct[4] that allows users to easily interact with our model.

Our models are addressing the limitations of English-centric models by incorporating more than 60% non-English documents. The training data comprises curated and web-crawled datasets, with a strong emphasis on non-English European languages[34]. Teuken-7B-Base was built with a custom tokeniser similar to Ali *et al.* (2023)[2] to reduce the fragmentation of text and improve efficiency across all 24 languages. In order to evaluate the multilingual capabilities of Teuken-7B-Base and Teuken-7B-Instruct we translate four well-known datasets, ARC[35], HellaSwag[36], MMLU[37], TruthfulQA[38] from English into 20 additional European languages and created a new European benchmark dataset[39].

Our LLMs demonstrate competitive performance on our European multilingual benchmarks, offering a significant step toward creating European-centric LLMs. Evaluations can be found in the European LLM Leaderboard[a]. Our research highlights the limitations of existing open-source models, such as their focus on high-resource languages and lack of transparency in model and data development. The Teuken-7B models aim to democratise this technology by providing insights into the machinery, e.g., the data preprocessing, the model design, the challenging training process, as well as the evaluations and instruction tuning. By covering the complete pipeline, we support further model developments and fine-tunings across European languages.

5 Conclusion

We demonstrated various techniques used for optimising a large-scale LLM training, with our focus being on best use of the available resources. While various papers publish their selected hyperparameters and model architecture, the process by which they are selected is often omitted. Similarly, vast amounts of papers publish only benchmark results of their technique, but do not properly observe the compute that the technique costs. We hope that publishing our results of medium-scale, compute-equivalent ablations enables more trust in these hyperparameters and a renewed perspective on how to conduct machine learning research for practical applicability. With regard to our codebase (an extremely popular one for large-scale training) and setup, both our positive and negative results confirm whether individual changes bring practical benefits – or do not.

Acknowledgements

This research was conducted as part of the OpenGPT-X project, which is funded by the German Federal Ministry for Economic Affairs and Climate Action (BMWK). The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V.[b] for providing computing time on the GCS Supercomputer JUWELS[40,14] at Jülich Supercomputing Centre (JSC).

[a] https://huggingface.co/spaces/openGPT-X/european-llm-leaderboard
[b] www.gauss-centre.eu

References

1. T. B. Brown, B. Mann, N. Ryder, M. Subbiah et al., *Language models are few-shot learners*, 2020, arXiv:2005.14165.
2. M. Ali, M. Fromm, K. Thellmann, R. Rutmann et al., *Tokenizer Choice For LLM Training: Negligible or Crucial?*, in: Findings of the Association for Computational Linguistics: NAACL 2024, Kevin Duh, Helena Gomez, and Steven Bethard, (Eds.), 3907-3924, Association for Computational Linguistics, Mexico City, Mexico, June 2024.
3. A. A. Weber, K. Thellmann, J. Ebert, N. Flores-Herr et al., *Investigating multilingual instruction-tuning: Do polyglot models demand for multilingual instructions?*, 2024, arXiv:2402.13703.
4. M. Ali, M. Fromm, K. Thellmann, J. Ebert et al., *Teuken-7b-base & teuken-7b-instruct: Towards european llms*, 2024, arXiv:2410.03730.
5. A. Kazemnejad, I. Padhi, K. Natesan Ramamurthy, P. Das, and S. Reddy, *The impact of positional encoding on length generalization in transformers*, Advances in Neural Information Processing Systems, **36**, 2024.
6. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit et al., *Attention is All you Need*, in: Advances in Neural Information Processing Systems, I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach et al. (Eds.), Vol. 30, Curran Associates, Inc., 2017.
7. J. Devlin, *Bert: Pre-training of deep bidirectional transformers for language understanding*, 2018, arXiv:1810.04805.
8. A. Dosovitskiy, *An image is worth 16x16 words: Transformers for image recognition at scale*, 2020, arXiv:2010.11929.
9. A. Radford, J. W. Kim, T. Xu, G. Brockman et al., *Robust speech recognition via large-scale weak supervision*, in: International conference on machine learning, PMLR, 28492-28518, 2023.
10. Z. Allen-Zhu and Y. Li, *Physics of language models: Part 1, learning hierarchical language structures*, 2023, arXiv:2305.13673.
11. S. Linnainmaa, *Taylor expansion of the accumulated rounding error*, BIT Numerical Mathematics, **16**, no. 2, 146-160, 1976.
12. C. M. John, S. Nassyr, C. Penke, and A. Herten, *Performance and Power: Systematic Evaluation of AI Workloads on Accelerators with CARAML*, in: Proceedings of the SC '24 Workshops of The International Conference on High Performance Computing, Network, Storage, and Analysis, SC-W '24, Association for Computing Machinery, New York, NY, USA, 2024, to appear.
13. A. Herten, S. Achilles, D. Alvarez, J. Badwaik et al., *Application-Driven Exascale: The JUPITER Benchmark Suite*, in: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, SC '24, Association for Computing Machinery, New York, NY, USA, 2024, to appear.
14. S. Kesselheim, A. Herten, K. Krajsek, J. Ebert et al., *JUWELS Booster – A Supercomputer for Large-Scale AI Research*, in: High Performance Computing, Springer International Publishing, Cham, 453-468, 2021.
15. J. Hoffmann, S. Borgeaud, A. Mensch, E. Buchatskaya et al., *Training compute-optimal large language models*, 2022, arXiv:2203.15556.

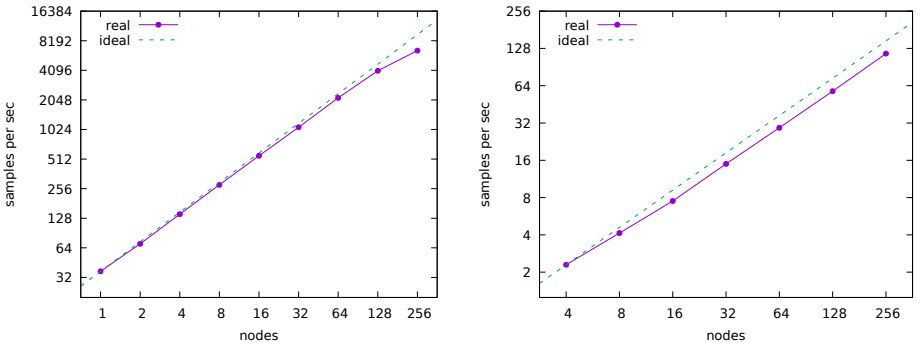
16. S. Narang, H. W. Chung, Y. Tay, W. Fedus et al., *Do transformer modifications transfer across implementations and applications?*, 2021, arXiv:2102.11972.
17. P. Clark, I. Cowhey, O. Etzioni, T. Khot et al., *Think you have solved question answering? try arc, the ai2 reasoning challenge*, 2018, arXiv:1803.05457.
18. C. Raffel, N. Shazeer, A. Roberts, K. Lee et al., *Exploring the limits of transfer learning with a unified text-to-text transformer*, Journal of machine learning research, **21**, no. 140, 1-67, 2020.
19. Google, "T5 v1.1", 2020.
20. P. Ramachandran, B. Zoph, and Q. V. Le, *Searching for activation functions*, 2017, arXiv:1710.05941.
21. Y. N. Dauphin, A. Fan, M. Auli, and D. Grangier, *Language modeling with gated convolutional networks*, in: International conference on machine learning, PMLR, 933-941, 2017.
22. N. Shazeer, *Glu variants improve transformer*, 2020, arXiv:2002.05202.
23. A. Radford, J. Wu, R. Child, D. Luan et al., *Language models are unsupervised multitask learners*, OpenAI, 2019.
24. S. Shleifer, J. Weston, and M. Ott, *Normformer: Improved transformer pretraining with extra normalization*, 2021, arXiv:2110.09456.
25. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, *Dropout: a simple way to prevent neural networks from overfitting*, The journal of machine learning research, **15**, no. 1, 1929-1958, 2014.
26. J. Lei Ba, J. R. Kiros, and G. E. Hinton, *Layer normalization*, 2016, arXiv:1607.06450.
27. B. Zhang and R. Sennrich, *Root mean square layer normalization*, Advances in Neural Information Processing Systems, **32**, 2019.
28. J. Su, M. Ahmed, Y. Lu, S. Pan et al., *Roformer: Enhanced transformer with rotary position embedding*, Neurocomputing, **568**, 127063, 2024.
29. O. Press, N. A. Smith, and M. Lewis, *Train short, test long: Attention with linear biases enables input length extrapolation*, 2021, arXiv:2108.12409.
30. J. Ainslie, J. Lee-Thorp, M. de Jong, Y. Zemlyanskiy et al., *Gqa: Training generalized multi-query transformer models from multi-head checkpoints*, 2023, arXiv:2305.13245.
31. X. Xie, P. Zhou, H. Li, Z. Lin, and S. Yan, *Adan: Adaptive nesterov momentum algorithm for faster optimizing deep models*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024.
32. H. Touvron, L. Martin, K. Stone, P. Albert et al., *Llama 2: Open foundation and fine-tuned chat models*, 2023, arXiv:2307.09288.
33. O. Filatov, J. Ebert, J. Wang, and S. Kesselheim, *Time Transfer: On Optimal Learning Rate and Batch Size In The Infinite Data Limit*, 2024, arXiv:2410.05838.
34. N. Brandizzi, H. Abdelwahab, A. Bhowmick, L. Helmer et al., *Data processing for the opengpt-x model family*, 2024, arXiv:2410.08800.
35. P. Clark, I. Cowhey, O. Etzioni, T. Khot et al., *Think you have Solved Question Answering? Try ARC, the AI2 Reasoning Challenge*, CoRR, 2018, arXiv:1803.05457.
36. R. Zellers, A. Holtzman, Y. Bisk, A. Farhadi, and Y. Choi, *HellaSwag: Can a Machine Really Finish Your Sentence?*, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 4791-4800, 2019.

37. D. Hendrycks, C. Burns, S. Basart, A. Zou et al., *Measuring Massive Multitask Language Understanding*, CoRR, 2020, arXiv:2009.03300.
38. S. Lin, J. Hilton, and O. Evans, *TruthfulQA: Measuring How Models Mimic Human Falsehoods*, in: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), S. Muresan, P. Nakovi, and A. Villavicencio (Eds.), 3214-3252, Dublin, Ireland, May 2022.
39. K. Thellmann, B. Stadler, M. Fromm, J. S. Buschhoff et al., *Towards cross-lingual llm evaluation for european languages*, 2024, arXiv:2410.08928.
40. Jülich Supercomputing Centre, *JUWELS Cluster and Booster: Exascale Pathfinder with Modular Supercomputing Architecture at Juelich Supercomputing Centre*, Journal of large-scale research facilities, **7**, no. A138, 2021.
41. M. Shoeybi, M. Patwary, R. Puri, P. LeGresley et al., *Megatron-lm: Training multi-billion parameter language models using model parallelism*, 2019, arXiv:1909.08053.
42. D. Narayanan, M. Shoeybi, J. Casper, P. LeGresley et al., *Efficient large-scale language model training on gpu clusters using megatron-lm*, in: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, 1-15, 2021.
43. H. Liu, M. Zaharia, and P. Abbeel, *Ring attention with blockwise transformers for near-infinite context*, 2023, arXiv:2310.01889.
44. S. Rajbhandari, J. Rasley, O. Ruwase, and Y. He, *Zero: Memory optimizations toward training trillion parameter models*, in: SC20: International Conference for High Performance Computing, Networking, Storage and Analysis, IEEE, 1-16, 2020.
45. A. Paszke, S. Gross, F. Massa, A. Lerer et al., *Pytorch: An imperative style, high-performance deep learning library*, Advances in neural information processing systems, **32**, 2019.
46. V. A. Korthikanti, J. Casper, S. Lym, L. McAfee et al., *Reducing activation recomputation in large transformer models*, Proceedings of Machine Learning and Systems, **5**, 341-353, 2023.

Appendix

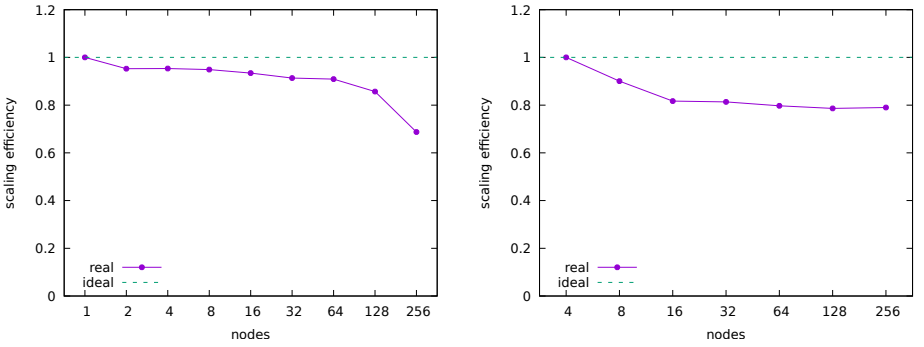
A Performance Analysis

For training our models, we created a fork of Megatron-LM[c], a highly distributed (using 4-dimensional data, tensor model[41], pipeline model[42], and context[43] parallelism, as well as optimiser sharding[44]) and optimised PyTorch[45] codebase for large-scale training of various Transformer models.



(a) 80M parameters (b) 13B parameters

Figure 1. Scaling behaviour of an (a) 80M and (b) 13B parameter language model; 4 GPUs per node.



(a) 80M parameters (b) 13B parameters

Figure 2. Scaling efficiency of an (a) 80M and (b) 13B parameter language model; 4 GPUs per node.

To ensure the selected codebase was a good fit for large-scale training of both small and large language models, we measured the scaling behaviour of our codebase on two

[c]Our fork can be found at `https://github.com/OpenGPTX/Megatron-LM`, the original codebase at `https://github.com/NVIDIA/Megatron-LM`.

model sizes, containing 80M and 13B parameters, respectively, on up to 1024 GPUs. We were interested in efficiency of small models to enable a quick feedback loop during experimental phases. The results are visualised in Figs. 1 and 2. These scaling runs were performed on the BigScience fork of the Megatron-DeepSpeed codebase (see Sec. 5), our initially chosen codebase.

Even minor throughput improvements greatly accumulate across a long-time large-scale training. In order to maximise the training throughput of our model, we conducted several benchmarks on optimal parameters for the model's parallelisation. The results are displayed in Tab. 2. The model configuration used selective activation checkpointing and sequence parallelism in normalisation layers[46]. The chosen parallelisation setting was tensor parallelism of degree 2, pipeline parallelism of degree 1, batch size of 1, and not constraining our jobs to a single InfiniBand cell (since this would result in longer queue times but did not make a significant difference in throughput).

GPUs	TP	PP	batch size	constrained	avg secs/iter	avg samples/sec
64	2	1	1	y	**19.9562**	**51.3131**
64	2	2	2	n	20.5169	49.9102
64	2	2	2	y	20.4896	49.9767
128	2	1	1	n	**10.1913**	**100.4780**
128	2	1	1	y	**12.2897**	**95.3088**
128	2	1	2	-	-	-
128	2	2	2	n	10.6927	95.7662
128	2	2	2	y	10.6420	96.2223
128	2	2	4	-	-	-
128	2	4	2	n	20.9229	69.3918
128	2	4	2	y	12.1837	86.9394
128	2	4	4	-	-	-
128	2	8	4	-	-	-

Table 2. Results of the parallelism layout benchmarks. One run per row/benchmark run, TP is degree of tensor model parallelism, PP is degree of pipeline model parallelism. Warmup of 3 steps and took the average of the rest of the values, with each run being given 15 minutes of wall-clock run time (including job initialisation). *Constrained* means the run executed on a single InfiniBand cell; otherwise, it was arbitrarily spread apart across at least 2 InfiniBand cells. Bold values mark the best-performing setting for that number of GPUs. The first row with bold values only contained 4 samples after warmup (a node failure killed the job early). For the second row with bold values, the values seemed similar to those in the row above it, but the average was skewed due to outliers. Due to the uncertainty in the first bold row's values, both bold results are hard to interpret. Rows without numbers resulted in out-of-memory errors.

Advancing Architectures for Video and Image Segmentation

Alexander Hermans, Ali Athar, Sabarinath Mahadevan, Idil Esen Zulfikar, and Bastian Leibe

Visual Computing Institute – Computer Vision, RWTH Aachen University,
52078 Aachen, Germany
E-mail: lastname@vision.rwth-aachen.de

In this report, we provide an overview of how high-performance computing resources were utilised to advance image and video segmentation through the development of three innovative architectures: TarViS, DynaMITe, and PointVOS. TarViS reformulates several video segmentation tasks, unifying four distinct tasks into a single, more generic architecture, trained jointly across all tasks. With DynaMITe, we address interactive image segmentation for fast annotation, demonstrating advantages in handling multiple objects simultaneously, unlike previous methods that focused on segmenting one object at a time. Finally, Point-VOS offers an efficient solution through sparse, point-based supervision, significantly reducing the need for large, fully labelled datasets in video object segmentation. We show how extensive experiments performed on HPC systems, particularly the GPU-accelerated JUWELS Booster, allowed us to improve training speeds and gain notable new insights in the area of video and image segmentation.

1 Introduction

Understanding visual scenes from images and videos is a fundamental challenge in the field of computer vision, with use-cases such as autonomous vehicles and robots interacting with their environments. An important sub-task of this understanding is the segmentation of inputs at a pixel level, assigning them to a specific class or object instance.

Traditional methods[4,5] for these tasks paved the way for modern advancements, but they were often limited in accuracy, flexibility, and scalability. The deep learning revolution transformed the segmentation landscape, significantly boosting performance[6–10].

However, these advances have come with new challenges. A key issue is the requirement for large amounts of (labelled) data for training high-performing models. Unlike tasks such as classification or detection, segmentation requires pixel-level annotations that are more time-consuming and expensive to obtain. This demand for high-quality, large-scale datasets presents a bottleneck in advancing the field, which is especially the case for video data, where dense, frame-by-frame annotations are even more costly to obtain.

Additionally, the domain of segmentation is highly fragmented, with various sub-tasks having their own distinct objectives, benchmarks, and metrics. Several segmentation tasks have emerged, including semantic segmentation, which classifies every pixel into categories; instance segmentation, which differentiates between individual objects; and panoptic segmentation, which seeks to unify the two by segmenting all objects and background classes in a scene. Recently, a stronger focus has also been placed on class-agnostic segmentation, with models such as SAM[11], that segment images into separable entities, with-

Figures and some text passages have been re-used from previous papers[1–3] and reports.

out assigning objects to a predefined set of classes. These segmentation tasks are also performed on videos, however here the fragmentation is even larger, with additional tasks such as Video Object Segmentation[12] (VOS), where the user provides an initial mask for an arbitrary object that should be segmented across the video. While significant progress has been made for all of these tasks, current methods are often tailored narrowly to the requirements of a single problem. As a result, they lack the flexibility to generalise across tasks, resulting in redundant developments. This also means that datasets require task-specific annotations, resulting in many smaller datasets for each of the tasks, that can often not be re-used or merged into a larger training dataset.

The motivation behind three of our recent contributions is to tackle different challenges in the segmentation community. Firstly, we introduce an approach that unifies several video segmentation datasets into a single model[13]. This allows us to jointly train on multiple tasks-specific datasets, revealing interesting synergies. Secondly, we discuss our model for interactive segmentation of images[2]. This method aids a user in the dense annotation of images with a small set of clicks on objects, greatly increasing the annotation speed. Finally, we investigate whether video object segmentation requires dense annotations in space and time and show that also with sparse annotations, VOS approaches can be trained[3].

The deep learning models used for segmentation, especially those operating on large-scale video datasets, require significant computational resources for training. In today's fast-evolving field of computer vision, where entirely new research directions can emerge within months, being able to conduct experiments quickly is crucial. By leveraging the GPU resources of the JUWELS supercomputer at the Jülich Supercomputing Centre (JSC), we significantly accelerated our research, allowing us to publish three papers at top-tier computer vision conferences (two at CVPR and one at ICCV). The compute resources were essential for performing experiments at the required depth and breadth within a reasonable time-frame. Performing multi-GPU and multi-node training significantly improved experimental speed, but running multiple parallel experiments also proved invaluable.

The remainder of this report provides a more in-depth look at our three papers (see Sections 2, 3, and 4), each addressing a specific problem in the image and video segmentation domain. Furthermore, in Section 5, we briefly discuss how we utilised the GPU compute resources at the JSC.

2 TarViS: A Unified Approach for Target-Based Video Segmentation

Our first focus is on the fragmented nature of video segmentation tasks. Several public benchmarks and tasks involve object tracking and segmentation in videos; this includes Multi-object Tracking and Segmentation (MOTS)[14], Video Instance Segmentation (VIS)[15,16], Video Object Segmentation[17], *etc*. These benchmarks were created by different research groups, at different points in time for different motivations, and over time each of them have spawned their own research sub-communities. As a result, most existing approaches are task-specific, even though the tasks themselves are highly overlapping and share common characteristics, *e. g.* , all of these tasks involve learning temporally consistent features for the video frames. With TarViS[1] (Target-based Video Segmentation) we introduce a novel, unified architecture to perform multiple of these tasks with a single trained model. To develop a unified architecture for these tasks, we first unified the task definitions on a conceptual level. To explain this, we will briefly describe the key video

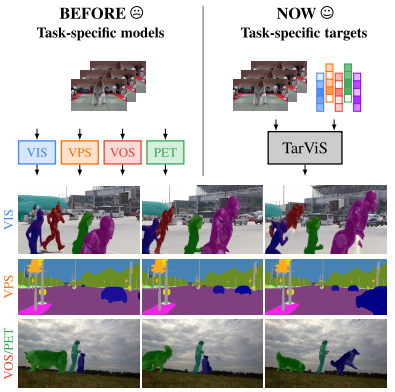


Figure 1. **Conceptual overview of TarViS.** Instead of heaving several task-specific models for segmentation tasks, we unify tasks into a single model where we specify task-specific targets to be tracked and segmented.

segmentation tasks:

- Video Instance Segmentation (VIS): This task requires all objects belonging to a set of predefined object classes to be segmented and tracked.
- Video Panoptic Segmentation (VPS): This can be seen as a super-set of the VIS task above. In addition to segmenting and tracking objects, we also have to assign a class label for points belonging non-instantiable *stuff* classes *e. g. wall*, *vegetation*, *road*, *etc*.
- Video Object Segmentation (VOS): Here the ground-truth masks for a certain set of objects in the first frame in which they appear is provided. The task is to then segment and track these objects in the remaining video.
- Point Exemplar-guided Tracking (PET): This can be seen as a more difficult variant of VOS where, instead of being given the full first-frame mask, we are only given the (x, y) coordinates of a single point on the object in the first frame in which it occurs.

Based on these definitions, it is clear that all of these tasks involve segmenting and tracking a set of *targets* in a video, with the difference being how these targets are defined. They are either objects belonging to a set of predefined categories (as in VIS and VPS) or a specific set of objects for which we are given some form of explicit guidance (VOS and PET). Based on this understanding, we designed TarViS as a generalised method capable of segmenting and tracking any set of targets. We achieve this by separating the task definition from the network architecture, as shown in Fig. 1.

Specifically, we model the task-specific segmentation targets as a set of concise *queries* which are fed into the network together with the video sequence. The network comprises a (1) backbone network which learns multi-scale feature maps for the video, and (2) a transformer-based decoder that accepts these feature maps as well as the target queries as input, and outputs a set of refined queries. The transformer decoder employs several layers of multi-head attention[18] to refine the feature representation of the queries. Specifically, the

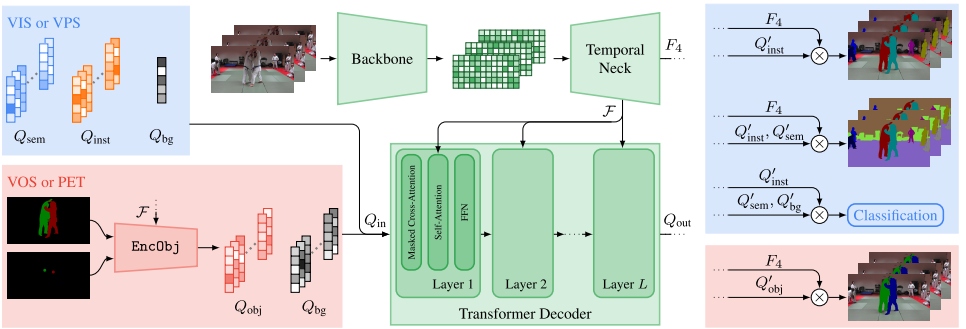


Figure 2. **Detailed network architecture for TarViS.** Segmentation targets are encoded as queries Q, and together with features F extracted from the backbone fed into a transformer decoder. Segmentation masks are then extracted based on dot products between the refined queries and the video features.

Setting	Video Training Data							VIS		VPS			VOS	PET
	YTVIS	OVIS	KITTI	C-VPS	VIPSeg	DAVIS	BURST	YTVIS (mAP)	OVIS (mAP)	KITTI (STQ)	C-VPS (VPQ)	VIPSeg (VPQ)	DAVIS ($\mathcal{J}\&\mathcal{F}$)	BURST (HOTA)
1. VIS	✓	✓						46.3	31.5	-	-	-	-	-
2. VPS			✓	✓	✓			-	-	0.70	49.7	32.4	-	-
3. VOS + PET						✓	✓	-	-	-	-	-	81.1	34.7
Final	✓	✓	✓	✓	✓	✓	✓	48.3	31.1	0.70	53.3	33.5	82.0	30.9

Table 1. TarViS trained on different dataset and task combinations using a ResNet-50[19] backbone. It can clearly be seen that in most cases the unified model outperforms the model trained for the specific task or performs competitively. C-VPS: CityscapesVPS, YTVIS: YouTube-VIS, KITTI: KITTI-STEP.

queries attend to each other (self-attention) and to the video feature maps (cross-attention). To obtain the final segmentation masks, we simply compute the dot-product between the refined queries output by the decoder, and the video features, followed by thresholding at 50% confidence. A detailed illustration of the architecture is given in Fig. 2. This approach enables us to train our model on a collection of datasets spanning multiple tasks. During inference, we can tackle any of these tasks by simply hot-swapping the query inputs without requiring any task-specific re-training.

To evaluate the efficacy of our approach, we applied it to seven different benchmarks spanning all four tasks listed above, each evaluated with their respective metrics. Tab. 1 shows that we can train this architecture on the separate tasks, but we can also jointly train TarViS on all seven considered datasets[15–17,20–23], showing that in many cases we can outperform the baseline or obtain similar performances. Compared to other existing methods, TarVis achieved state-of-the-art results for five out of seven benchmarks and performed competitively on the remaining two.

TarViS showcases how a unified model can effectively address diverse video segmentation tasks with minimal adaptation, underscoring the potential for flexible, task-agnostic approaches. This opens the door to further investigating shared principles across seemingly distinct challenges in video analysis. Further details can be found in the TarViS paper[1].

3 DynaMITe

While with TarViS we can merge several video segmentation tasks into a single model, we require annotated training datasets. These datasets can be annotated manually or with the help of interactive segmentation tools, where a user interactively clicks on an object in an image or segmentation mistake, guiding the model to segment a certain object. This allows a prediction of a model to be iteratively refined by additional user clicks, correcting mistakes of the model. Previous interactive segmentation methods[24–28] only predict binary segmentation masks, though, and can thus only be used to annotate one object at a time. As such, for every foreground object, the remaining objects are considered as background, forcing the user to perform many redundant clicks in order to annotate scenes containing multiple objects. Additionally, existing models process user inputs in such a way that requires the complete model to be executed for each click, thereby limiting their network sizes to achieve a good runtime performance, which is critical in interactive settings.

We developed an interactive image segmentation network called DynaMITe[2] (Dynamic Query Bootstrapping for Multi-object Interactive Segmentation Transformer) which addresses these two shortcomings. DynaMITe models multiple objects at the same time, while only extracting features from the image with a strong backbone once. To this end we formulate the user clicks as a spatio-temporal sequence of data and translate them into queries that are processed by our Interactive Transformer module. Inside the Interactive Transformer module, these queries can interact between each other which enables a common background modelling, thereby reducing redundancy in background clicks.

Fig. 3 shows the overall architecture of DynaMITe which takes an input image and the corresponding set of user clicks as inputs. The user clicks can either be a positive click representing a foreground object or a negative click representing the common background. Based on the current prediction, a user can then add the next click. Unlike previous works, DynaMITe can handle multiple instances at once and hence the positive clicks can belong to different foreground objects, which are grouped based on the object ID assigned during the click. The backbone processes the image and extracts low-level features, which are

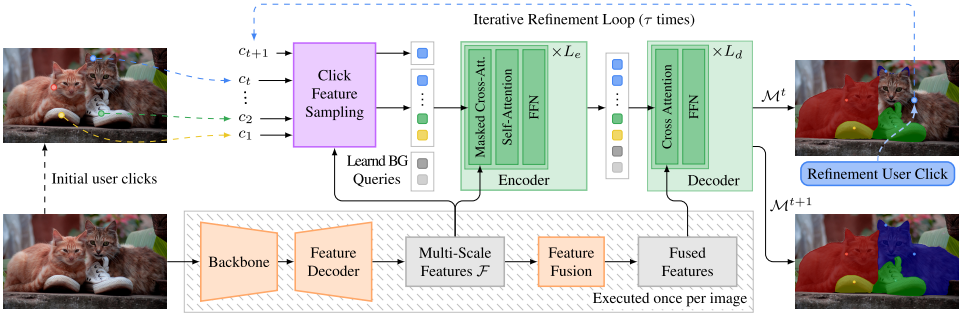


Figure 3. **DynaMITe architecture overview.** DynaMITe consists of a backbone, a feature decoder, and an interactive Transformer. Point features at click locations at time t are translated into queries which are processed by a Transformer encoder-decoder structure to generate a set of output masks $\mathcal{M}^t$ for all the relevant objects. Based on $\mathcal{M}^t$, the user provides a new input click which is in turn used by the interactive Transformer to generate a new set of updated masks $\mathcal{M}^{t+1}$. This process is iterated until the masks reach the desired quality.

Method	Backbone	COCO				SBD				DAVIS17			
		NCI↓	NFO↓	NFI↓	IoU↑	NCI↓	NFO↓	NFI↓	IoU↑	NCI↓	NFO↓	NFI↓	IoU↑
FocalClick[27]	Segformer-B0[33]	7.31	19422	3004	73.7	4.26	1115	599	87.3	4.6	802	562	84.6
DynaMITe	Segformer-B0[33]	6.04	12986	2431	84.9	2.76	528	313	90.6	3.27	549	356	87.9
DynaMITe	Swin-T[34]	6.00	**12710**	2401	**85.1**	2.69	**510**	303	**90.7**	**3.16**	514	**338**	**88.0**

Table 2. Results on our Multi-Instance Segmentation Task. Segmentation quality is iterated until the mask Intersection over Union (IoU) reaches 85%. NCI: normalised clicks per image, NFO: number of failed objects, NFI: number of failed images.

then fed to the Interactive Transformer along with the associated user interactions. The Interactive Transformer has an encoder-decoder structure, where the encoder closely follows the transformer decoder architecture from Mask2Former[9]. The main task is to update the queries that represent the different user clicks based on the image features. The decoder does the opposite, using the updated click queries in order to update the image features via cross-attention, before finally a prediction is extracted based on a dot product between the queries and the image features, very similar as also done in TarViS.

We evaluate DynaMITe on a range of datasets across two interactive segmentation task settings: (i) the well established single-instance setting using small-scale datasets mostly containing one object instance per image such as GrabCut[4], Berkeley[29], COCO MVal (a small subset of COCO[30]), and DAVIS[31] and (ii) our novel multi-instance segmentation task (MIST) on larger multi-instance datasets like COCO[30], DAVIS17[17], and SBD[32]. For the single instance setting, DynaMITe performs competitively compared with previous state-of-the art methods, outperforming them on many datasets when using the same backbone network (see Table 1 of the main paper[2]). More relevant to us is our novel Multi-instance Interactive Segmentation Task. While for the automatic evaluation in the single-instance case it is fairly easy to simulate where a user might click next given a prediction, such a clicking strategy is more ambiguous when it comes to multiple objects. For example, a user might complete one object before moving to the next, focus on the largest errors across all objects, or use some other strategy. For automatic evaluation: (i) we place an initial positive click in the centre of each object and generate an initial prediction, (ii) choose a random object from this prediction set that has not achieved the required segmentation quality, and (iii) place a click on the largest error region for the chosen object. The sampling process is repeated until either the entire image is segmented, or until we exhaust an image-level click budget (10 times to number of objects in the image). We use a new metric called Normalised Clicks per Image (NCI), which is obtained by normalising the total number of clicks used for an input image by the number of foreground objects in that image. Additionally, we also mark the average number of failed objects (NFO), number of failed images (NFI) which could not obtain a desired segmentation quality within the click budget, and the average IoU of the final segmentation. As a baseline, we adapt the state-of-the-art single-instance FocalClick[27] to our MIST setting by processing each object in parallel and choosing the object with the largest error to be refined by the next click.

Tab. 3 compares the performance of DynaMITe against the adapted FocalClick. Our method outperforms the baseline on all metrics for this task and achieves a significantly better final segmentation quality as shown by the IoU values, highlighting the benefit of jointly predicting multiple instances for interactive segmentation.

4 Point-VOS: Pointing Up Video Object Segmentation

Finally, we investigate whether video object segmentation can be made more efficient by reducing the annotation burden. Rather than requiring dense annotations for every frame, we explore the potential of sparse annotation schemes where only key points across selected frames are annotated, turning it into a form of weak supervision. Such an approach can drastically cut down on the time and cost of annotation while still allowing models to achieve results comparable to those trained on fully annotated datasets.

The conventional (VOS) task utilises dense segmentation masks for each frame during training and initialises the first-frame reference with dense masks during inference. In contrast, we propose to change this paradigm with our new Point-VOS task, where we use only spatially sparse point annotations on a sparse subset of frames during training, and only a few points for the first-frame reference initialisation, as seen in Fig. 4. With this we address the annotation cost problem in the conventional VOS task by proposing an entirely point-based framework. Point-VOS moves away from using full mask supervision and instead relies on spatio-temporal sparse point annotations as weak supervision signals.

To study the effect of training and initialising with spatio-temporal sparse points, we performed a series of experiments with STCN[10], a state-of-the-art VOS approach, modified such that it can be trained on sparse points. First, we analysed the number of points required for training supervision and test-time initialisation (see Fig. 5 left). It can clearly be seen that there is a diminishing return with additional training points. A similar effect can be seen with the number of points used to mark the reference object. We also analysed the number of frames required for training on sampled points per frame per object (see Fig. 5 right). Overall, we find 10 points per object on 10 frames to be a good trade-off between additional performance gain and additional annotation costs for our Point-VOS task.

Based on these findings, an in order to facilitate our proposed Point-VOS task, we annotate two large-scale video datasets, Point-VOS Oops[35] (PV-Oops) and Point-VOS Kinetics[36] (PV-Kinetics), with altogether 19M points for 133K objects in 32K videos. These datasets contain significantly more videos and objects than the previously largest existing

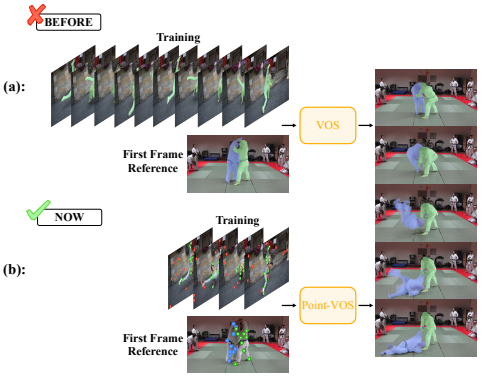


Figure 4. **Comparison of the conventional VOS task (a) with our new Point-VOS task (b).** Green and blue dots represent foreground points and red dots background points. In both cases we train a network to predict dense masks based on an initial reference segmentation.

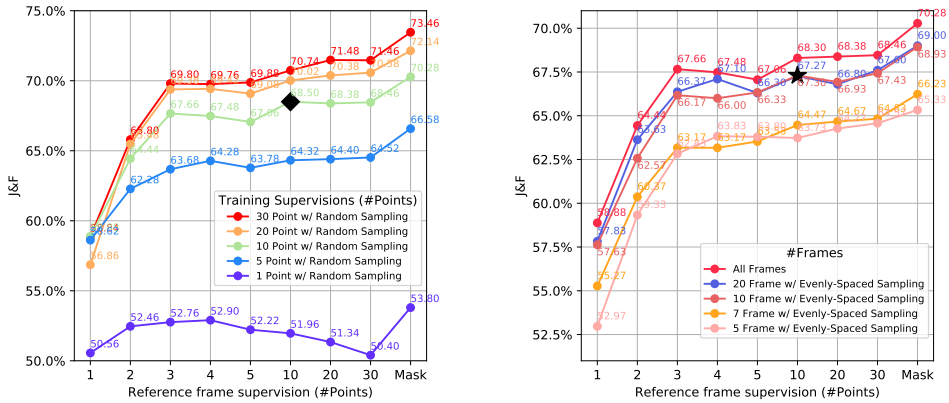


Figure 5. **(Left)** Training vs. test-time point supervision results. **(Right)** Results for varying temporal sparsity, when trained on 10 randomly sampled points per frame per object. ♦ and ★ represent our chosen setting, *i. e.* 10 points for training supervision across 10 frames per video and 10 points for test-time reference frames. We train an STCN[10] model five times for each configuration and report the mean score on the DAVIS[17] validation set.

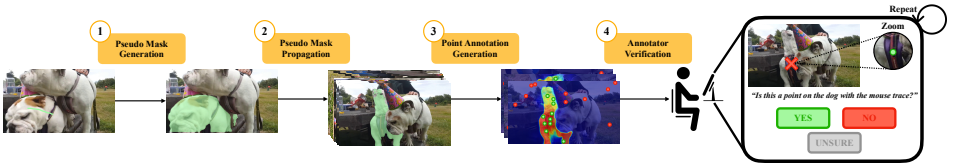


Figure 6. **Semi-automatic annotation pipeline.** We extract pseudo masks from initial mouse trace annotations, these masks are propagated across the video, and finally, after sampling points from the masks, an annotator manually verifies their correctness. Green circles represent foreground points and red circles background points.

VOS datasets VISOR[37] (7.8K videos) and BURST[23] (2.9K videos and 16K objects). The PV-Oops and PV-Kinetics datasets are multi-modal, *i. e.* , they include language annotations, which opens up additional interesting use-cases.

To create these two datasets, we developed a semi-automatic annotation scheme, as shown in Fig. 6. In our annotation pipeline, we first extract a mouse trace segment for each noun in the VidLN[38] captions and convert it into a pseudo mask using a slightly modified version of DynaMITe. We then propagate the pseudo-mask across the video using STCN[10]. We use the output probability maps to sample sparse point annotations and let annotators verify them manually. We measured the speed of annotating sparse spatial-temporal points to show the efficiency of our Point-VOS task. Annotation on average takes 0.95 seconds per point which is 40 times faster than the dense mask annotation scheme. While the resulting annotations are sparse, the Point-VOS concept enables the annotation of significantly larger datasets, boosting the variety of objects in training sets.

Based on these datasets we define a new Point-VOS benchmark. We propose two variants, training and testing purely on points, or based on pseudo masks obtained by using the points to prompt an interactive image segmentation method like DynaMITe. The two

versions of our benchmark are evaluated on Point-VOS versions of DAVIS[17], YouTube-VOS[39], and our Point-VOS Oops dataset. Detailed results of initial baselines on this benchmark, as well as further interesting experiments can be found in our Point-VOS paper[3].

5 Hardware and Software Configuration

For all three projects, we follow a similar approach for code parallelisation across multiple GPUs and across different nodes. Specifically, we use the PyTorch library[40] which provides a `DistributedDataParallel` (DDP) API for seamlessly parallelising network training across multiple nodes. Fig. 7 shows the speedup we can gain using this strategy for our TarViS code base, where the other two projects result in similar curves. Further details can be found in the respective GitHub repositories[a].

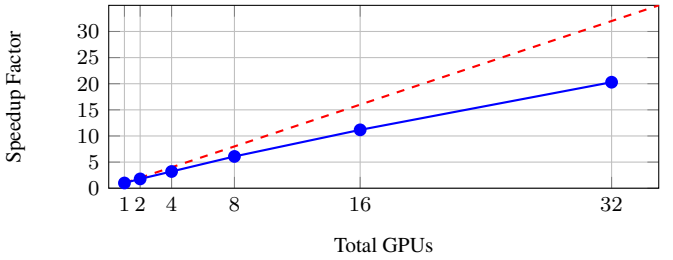


Figure 7. **Parallelisation speedup for TarViS** when trained on multiple GPUs and nodes.

While the speedup is sublinear, experiments run significantly faster[b]. Additionally, a multi-GPU setup allows us to use significantly larger batch sizes during training, which is often crucial for good performance of deep learning models. Orthogonal to the parallelisation of our code, the cluster allows us to run several jobs in parallel, which results in an additional linear speedup. Using these two approaches we were able to significantly increase the speed of our experiments, resulting in deeper insights.

6 Concluding Remarks

Using the compute resources provided at the Jülich Supercomputing Centre, we explored three important aspects of segmentation. With TarViS we unified several video segmentation tasks into a single network architecture by unifying the task descriptions. This allowed us to utilise many datasets at once, showing synergies between the tasks and resulting in an overall very strong performance across the different considered tasks. Based on our DynaMITe approach, we showed that an interactive multi-instance segmentation approach can reduce the number of click interactions a user needs to perform in order to annotate

[a] TarViS: `https://github.com/Ali2500/TarViS`
DynaMITe: `https://github.com/amitrana001/DynaMITe`
Point-VOS: `https://github.com/idilesenzulfikar/PointVOS`

[b] In more recent projects we also utilised DeepSpeed[41], resulting in even larger speedups.

images with multiple objects. At the same time the resulting annotations reach a higher accuracy, which is crucial when such an interactive approach is used to annotate training data. Finally, we introduced the new Point-VOS task, a sparse version of the video object segmentation task and showed that with spatially and temporally sparse point annotations, we are able to train video object segmentation networks that produce dense masks, further reducing the annotation burden.

The GPU compute resources were critical in the development of these novel approaches. In the future we plan to investigate further aspects of video segmentation. We will extend our interactive segmentation to videos, both for 2D videos, but also for temporal sequences of 3D point clouds. We will also investigate the use of large foundation modals such as DINOv2[42] or SAM[11], merging them with strong large language models to enable video segmentation. All of these endeavours require significant amounts of GPU compute and here supercomputing Centres will hopefully remain a strong cornerstone, enabling important AI research within academia.

Acknowledgements

The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (`www.gauss-centre.eu`) for supporting the underlying project by providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS Supercomputer JUWELS at Jülich Supercomputing Centre (JSC). Funding was provided, in parts, by the ERC Consolidator Grant DeeVise (ERC-2017-COG- 773161) and the BMBF project NeuroSys-D (03ZU1106DA). Experiments were partially also performed using the computing resources granted by RWTH Aachen University under project `rwth1239` and `supp0003`.

References

1. A. Athar, A. Hermans, J. Luiten, D. Ramanan, and B. Leibe, *TarViS: A Unified Architecture for Target-based Video Segmentation*, in: CVPR, 2023.
2. A. Rana, S. Mahadevan, A. Hermans, and B. Leibe, *DynaMITe: Dynamic Query Bootstrapping for Multi-object Interactive Segmentation Transformer*, in: ICCV, 2023.
3. S. Mahadevan, I. E. Zulfikar, P. Voigtlaender, and B. Leibe, *Point-VOS: Pointing Up Video Object Segmentation*, in: CVPR, 2024.
4. C. Rother, V. Kolmogorov, and A. Blake, *"GrabCut": Interactive foreground extraction using iterated graph cuts.*, in: SIGGRAPH, 2004.
5. P. Krähenbühl and V. Koltun, *Efficient inference in fully connected crfs with gaussian edge potentials*, in: NeurIPS, 2011.
6. J. Long, E. Shelhamer, and T. Darrell, *Fully Convolutional Networks for Semantic Segmentation*, in: CVPR, 2015.
7. L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, *Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs*, TPAMI, 2017.
8. K. He, G. Gkioxari, P. Dollár, and R. Girshick, *Mask R-CNN*, in: ICCV, 2017.
9. B. Cheng, I. Misra, A. G. Schwing, A. Kirillov, and R. Girdhar, *Masked-attention Mask Transformer for Universal Image Segmentation*, in: CVPR, 2022.

10. H. K. Cheng, Y.-W. Tai, and C.-K. Tang, *Rethinking Space-Time Networks with Improved Memory Coverage for Efficient Video Object Segmentation*, in: NeurIPS, 2021.
11. A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo et al., *Segment anything*, in: ICCV, 2023.
12. S. Caelles, K.-K. Maninis, J. Pont-Tuset, L. Leal-Taixé, D. Cremers, and L. Van Gool, *One-Shot Video Object Segmentation*, in: CVPR, 2017.
13. A. Athar, J. Luiten, A. Hermans, D. Ramanan, and B. Leibe, *HODOR: High-level Object Descriptors for Object Re-segmentation in Video Learned from Static Images*, in: CVPR, 2022.
14. P. Voigtlaender, Y. Chai, F. Schroff, H. Adam, B. Leibe, and L.-C. Chen, *Feelvos: Fast end-to-end embedding learning for video object segmentation*, in: CVPR, 2019.
15. L. Yang, Y. Fan, and N. Xu, *Video instance segmentation*, in: ICCV, 2019.
16. J. Qi, Y. Gao, Y. Hu, X. Wang, X. Liu, X. Bai, S. Belongie, A. Yuille, P. Torr, and S. Bai, *Occluded Video Instance Segmentation: A Benchmark*, IJCV, 2022.
17. J. Pont-Tuset, F. Perazzi, S. Caelles, P. Arbeláez, A. Sorkine-Hornung, and L. Van Gool, *The 2017 DAVIS Challenge on Video Object Segmentation*, 2017, arXiv:1704.00675.
18. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, *Attention is all you need*, in: NeurIPS, 2017.
19. K. He, X. Zhang, S. Ren, and J. Sun, *Deep residual learning for image recognition*, in: CVPR, 2016.
20. M. Weber, J. Xie, M. Collins, Y. Zhu, P. Voigtlaender, H. Adam, B. Green, A. Geiger, B. Leibe, D. Cremers, A. Ošep, L. Leal-Taixé, and L.-C. Chen, *STEP: Segmenting and Tracking Every Pixel*, in: NeurIPS, 2021.
21. D. Kim, S. Woo, J.-Y. Lee, and I. S. Kweon, *Video panoptic segmentation*, in: CVPR, 2020.
22. J. Miao, X. Wang, Y. Wu, W. Li, X. Zhang, Y. Wei, and Y. Yang, *Large-Scale Video Panoptic Segmentation in the Wild: A Benchmark*, in: CVPR, 2022.
23. A. Athar, J. Luiten, P. Voigtlaender, T. Khurana, A. Dave, B. Leibe, and D. Ramanan, *BURST: A Benchmark for Unifying Object Recognition, Segmentation and Tracking in Video*, in: WACV, 2023.
24. S. Mahadevan, P. Voigtlaender, and B. Leibe, *Iteratively Trained Interactive Segmentation*, in: BMVC, 2018.
25. K. Sofiiuk, I. Petrov, and A. Konushin, *Reviving Iterative Training with Mask Guidance for Interactive Segmentation*, 2021, arXiv:2102.06583.
26. K. Sofiiuk, I. Petrov, O. Barinova, and A. Konushin, *f-brs: Rethinking backpropagating refinement for interactive segmentation*, in: CVPR, 2020.
27. X. Chen, Z. Zhao, Y. Zhang, M. Duan, D. Qi, and H. Zhao, *FocalClick: Towards Practical Interactive Image Segmentation*, in: CVPR, 2022.
28. Q. Liu, M. Zheng, B. Planche, S. Karanam, T. Chen, M. Niethammer, and Z. Wu, *PseudoClick: Interactive Image Segmentation with Click Imitation*, in: ECCV, 2022.
29. K. McGuinness and N. E. O'connor, *A comparative evaluation of interactive segmentation algorithms*, Pattern Recognition **43**, 434-444, 2010.
30. T.-Y. Lin, M. Maire, S. J. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, *Microsoft COCO: Common Objects in Context*, in: ECCV, 2014.

31. F. Perazzi, J. Pont-Tuset, B. McWilliams, L. Van Gool, M. Gross, and A. Sorkine-Hornung, *A Benchmark Dataset and Evaluation Methodology for Video Object Segmentation*, in: CVPR, 2016.
32. B. Hariharan, P. Arbeláez, L. Bourdev, S. Maji, and J. Malik, *Semantic contours from inverse detectors*, in: ICCV, 2011.
33. E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, *SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers*, in: NeurIPS, 2021.
34. Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, *Swin transformer: Hierarchical vision transformer using shifted windows*, in: ICCV, 2021.
35. D. Epstein, B. Chen, and C. Vondrick, *Oops! Predicting Unintentional Action in Video*, in: CVPR, 2020.
36. W. Kay, J. Carreira, K. Simonyan, B. Zhang, C. Hillier, S. Vijayanarasimhan, F. Viola, T. Green, T. Back, P. Natsev et al., *The kinetics human action video dataset*, 2017, arXiv:1705.06950.
37. A. Darkhalil, D. Shan, B. Zhu, J. Ma, A. Kar, R. Higgins, S. Fidler, D. Fouhey, and D. Damen, *EPIC-KITCHENS VISOR Benchmark: VIdeo Segmentations and Object Relations*, in: NeurIPS, 2022.
38. P. Voigtlaender, S. Changpinyo, J. Pont-Tuset, R. Soricut, and V. Ferrari, *Connecting Vision and Language with Video Localized Narratives*, in: CVPR, 2023.
39. N. Xu, L. Yang, Y. Fan, J. Yang, D. Yue, Y. Liang, B. L. Price, S. D. Cohen, and T. S. Huang, *YouTube-VOS: Sequence-to-Sequence Video Object Segmentation*, in: ECCV, 2018.
40. A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein et al., *Pytorch: An imperative style, high-performance deep learning library*, in: NeurIPS, 2019.
41. J. Rasley, S. Rajbhandari, O. Ruwase, and Y. He, *DeepSpeed: System Optimizations Enable Training Deep Learning Models with Over 100 Billion Parameters*, in: SIGKDD, 2020.
42. M. Oquab, T. Darcet, T. Moutakanni, H. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. Haziza, F. Massa et al., *DINOv2: Learning robust visual features without supervision*, 2023, arXiv:2304.07193.

Fluid Mechanics

Fluid Mechanics

Christian Stemmer

Lehrstuhl für Aerodynamik und Strömungsmechanik, Technische Universität München,
Boltzmannstraße 15, 85748 Garching b. München, Germany
E-mail: Christian.Stemmer@tum.de

The three contributions to this year's NIC Symposium deal with combustion simulations including hydrogen combustion and machine-learning applications. The ongoing change from a carbon-based to a hydrogen-based economy in terms of energy supply mandates intensive research into hydrogen combustion mechanism. They pose huge challenges for the modelling and simulation engineers mainly because reaction time scales are much shorter compared to carbon-based combustion. Flame-holding or flame extinction are just two challenging features. Increasing computer power including massive GPU-usage poses a challenge in the effective use of the available resources but opens up the potential for high-fidelity simulations and research to supply high-resolution data for lower-order modelling efforts in the context of engineering design methodologies. As combustion and turbulence pose a specific challenge for simulations, it is to no surprise that this year's contributions circle around theses hot topics and ideas using machine-learning techniques to reduce the effort for future designs and development for everyday engineers.

The first paper lays out the computational challenges for the use of the new massive-GPU based exascale JUPITER system at the JSC. The RWTH Aachen research group at ITV simulated lean premixed turbulent hydrogen/air flames with the PeleLMeX code. Adaptive mesh refinement is a must in combustion simulations as the flame is very thin relative to the entire computational domain requiring a considerable hierarchy of local mesh refinements also influencing the applicable time-step limit. The parallel-computation approach uses hybrid MPI+X methodologies enabling the use of CUDA-based load distribution to the massive GPUS available on JUPITER. The detailed simulations show the development of thermodiffusive instabilities with increasing pressures necessary for clean combustion processes.

The second paper in the fluid dynamics section is from the same discipline underlining the importance of computational high-resolution simulations providing accurate data for low-order modelling. With the help of the nekCRF code, data is generated which supports the formulation of LES subgrid-scale models using the potential advantages of machine-learning algorithms. The methodology to develop highly accurate LES models with the help of machine-learning methods is presented. With the support of a large number of GPUs, these models can be used to efficiently investigate accurate hydrogen combustion cases for the application to design processes and evaluations of efficient usage of hydrogen fuels.

The third contribution focuses on the small-scale nature of turbulence and the enhanced insights to its development accessible through high-performance computing results. The highly resolved flow structures are exploited, feeding deep-learning algorithms with the velocity gradient tensor. The resulting model is tested over a range of Reynolds numbers of two orders of magnitude.

Towards Exascale Simulations of Carbon-Free Combustion Systems

Thomas L. Howarth[1], **Terence Lehmann**[1], **Michael Gauding**[1], **Marcus S. Day**[2], **and Heinz Pitsch**[1]

[1] Institute for Combustion Technology, RWTH Aachen University, 52056 Aachen, Germany
E-mail: {t.howarth, t.lehmann, m.gauding, h.pitsch}@itv.rwth-aachen.de

[2] Computational Science Center, National Renewable Energy Laboratory, Golden, CO 80401-3305, USA
E-mail: marcus.day@nrel.gov

Recent advances in computational power, particularly the transition to GPU-based HPC systems, have allowed for larger and more complex combustion system simulations. It is now possible to run simulations at higher, and more technically relevant, Reynolds numbers, pressures and temperatures. This enables the direct numerical simulation of experiments at relevant laboratory conditions. Here, the capability of the reacting flow solver PeleLMeX, developed under the US Department of Energy's Exascale Computing Project (ECP), is demonstrated on the early access module (JEDI) of the exascale machine JUPITER and the pre-exascale JUWELS Booster and Cluster machines. A series of simulations is also presented to demonstrate some of the physics that can be explored using PeleLMeX.

1 Introduction

Global carbon emissions are driving governments and research institutions to develop policies and technology to allow for the ongoing transition away from fossil-fuelled devices. One option in this transition is the utilisation of hydrogen-based fuels. This can involve burning hydrogen or ammonia in air directly or using ammonia as a hydrogen carrier that can be (partially-)cracked into hydrogen. From a combustion standpoint, hydrogen and ammonia present unique challenges[1]. Hydrogen burns efficiently but also has the potential for both thermodiffusive and thermoacoustic combustion instabilities. Ammonia, on the other hand, burns very slowly but is far easier to produce and transport. Understanding how these fuels burn independently and as a blend is key to both the adaptation of existing, and the development of new combustion technology.

Direct numerical simulations (DNS) of reacting flows resolve the flow and chemistry to the degree that the physics does not rely on the use of subgrid models. While simulations of this nature allow for a direct inspection of the physics, such high resolutions can make simulations extremely computationally expensive. This is particularly true in reacting flows, where the use of detailed chemistry creates challenges in terms of the number of equations to solve and the range of timescales introduced. Databases generated using DNS can be used to explore physics at a scale not possible in the laboratory or to develop closures for models to be used in low-fidelity models that can be used to rapidly design and prototype new technology. In either case, performing simulations at larger ranges of scales than ever is crucial. The Pele suite of solvers offers the opportunity to utilise emerging exascale computing systems to perform these simulations and develop carbon-free combustion technology.

2 Software and Algorithms

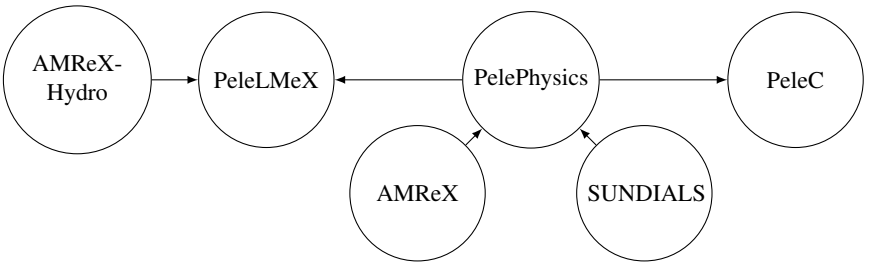


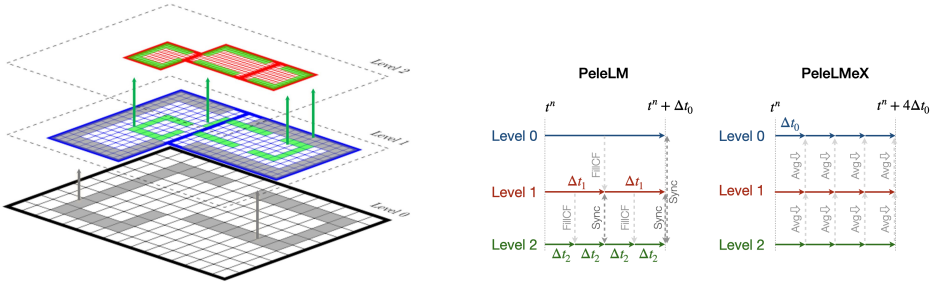
Figure 1. Modular structure of AMReX-Combustion codes.

The Pele suite of codes is an open-source group of solvers[2] written in C++ for the simulation of reacting flow, with two main solvers (PeleC[3] and PeleLMeX[4]) and a support library (PelePhysics) that provides common models and code between the two. PeleC solves the equations using a compressible formulation, suitable for high-speed flows where the effects of propagating pressure waves in the domain are non-negligible. PeleLMeX solves the governing equations in the low Mach number limit, where pressure waves are instead assumed to propagate at infinite speed. Both codes are built on a modular hierarchical structure (shown in Fig. 1) and use the same models concerning transport, thermodynamics and chemical mechanisms, which are contained within PelePhysics. Additionally, they also both use the AMReX[5] library that supports block-structured adaptive mesh refinement as well as tools to support a variety of parallel programming strategies. In this section, more details are provided about AMReX and PeleLMeX, as used in the work that follows.

2.1 AMReX: Block-Structured Adaptive Mesh Refinement Framework

AMReX is a software framework that allows for the development of partial differential equation solvers in multi-dimensional domains that benefit from the use of adaptive mesh refinement (AMR). PeleLMeX is only one of the many suites of software developed using the AMReX library, with many other solvers created for applications in astrophysics[6], particle-laden flows[7], atmospheric boundary layers[8] and geophysical fluid dynamics[9].

A typical simulation using AMReX employs one or more levels of AMR, which are situated directly on top of each other (Fig. 2(a)). The "base grid" (level 0 in Fig. 2(a)) is fixed and covers the whole computational domain, with the additional levels added dynamically. In a reacting flow simulation, individual cells at a coarse level can be identified for additional refinement based on concentrations of intermediate species, for example. These cells are grouped into contiguous boxes, refined by a factor of two or four, and compose the next finer AMR level. During the simulation, the size of each level will vary depending on the criteria given, and the simulation will be analysed for potential regridding at an interval determined by the user. Timestepping these levels of data can either be done with or without temporal subcycling, typically to enforce that each cell/level is advanced with a uniform CFL number (see Fig. 2(b)). PeleLMeX is the non-subcycling version of its predecessor PeleLM. In PeleLM, the coarsest level (level 0) is advanced by using a timestep



(a) Hierarchy of AMR levels.

(b) Subcycling across levels of AMR.

Figure 2. AMReX features.

determined by stability, and subsequent levels of refinement are advanced with sequentially halved timesteps. Whether or not subcycling is advantageous depends strongly on the configuration. When adjacent levels in the AMR hierarchy reach the same simulation time, a synchronisation procedure is used to enforce desirable properties of the algorithm, such as conservation and elliptic regularity. The algorithmically simpler non-subcycling algorithm used in PeleLMeX is constructed to also satisfy local conservation and elliptic regularity but does so by advancing the entire AMR hierarchy at the time step dictated by stability constraints of the finest level. The non-subcycling variant additionally supports combustion in closed chambers, where heat release and transport can lead to pressurisation of the system during evolution.

Data may be held in cell-centred, face-centred, edge-centred or nodal forms, depending on the needs of the numerical scheme. In addition to structured mesh data, AMReX also supports the usage of particle data, which in the context of Pele allows for the modelling of soot particles and spray droplets in combustion applications using a Lagrangian formulation that couples to the Eulerian field quantities through inter-phase transfer terms. Complex geometry can also be represented through a multilevel embedded boundary (EB) representation (i.e. cut-cell). As well as the underlying data structures, AMReX also provides linear solvers for cell-centred and nodal data and a native file format that is widely readable by many visualisation software packages (e.g. ParaView, VisIt, yt).

AMReX adopts an MPI+X approach to the parallelisation, where X can be OpenMP in the case of multithreading CPUs, and combined with any of CUDA/HIP/SYCL for NVIDIA, AMD or Intel GPUs, respectively. The complexities of each parallelisation strategy are exposed to developers through C++ lambda expressions. An example of a typical ParallelFor loop is given in Listing 1. Each for loop works on the collection of boxes (i.e. subdomains) of data that have been allocated to a particular MPI process.

Listing 1. Example of a ParallelFor loop in an AMReX-based application.

```
auto const& a = mfa.arrays();
auto const& b = mfb.const_arrays();
ParallelFor(mfa,
    [=] AMREX_GPU_DEVICE (int box, int i, int j, int k) {
    a[box](i,j,k) = 2*b[box](i,j,k);
});
```

2.2 PeleLMeX: Low-Mach Number Reacting Flow Solver

In this section, the governing equations, numerical methods and algorithm used in PeleLMeX are outlined.

PeleLMeX solves the reacting Navier-Stokes equations in the low-Mach limit, given by

$$\frac{\partial \rho \mathbf{u}}{\partial t} + \nabla \cdot (\rho \mathbf{u}\mathbf{u}) = -\nabla \pi + \nabla \cdot \boldsymbol{\tau} \tag{1}$$

$$\frac{\partial \rho Y_k}{\partial t} + \nabla \cdot (\rho Y_k \mathbf{u} + \mathbf{F}_k) = \dot{\omega}_k, \quad k = 1, \ldots, N \tag{2}$$

$$\frac{\partial \rho h}{\partial t} + \nabla \cdot (\rho h \mathbf{u} + \mathbf{Q}) = 0 \tag{3}$$

where $\mathbf{u}$ is the velocity field, ρ is the density, π is the dynamic pressure, $\boldsymbol{\tau}$ is the viscous tensor, Y_k, $\mathbf{F}_k$ and $\dot{\omega}_k$ are the mass fraction, diffusive flux and chemical source term for species k, h is the specific enthalpy and $\mathbf{Q}$ is the heat flux. This equation set is supplemented by an equation of state that expresses the thermodynamic pressure, p_0, as a function of the other state variables. Note that in the low Mach limit, p_0 is assumed to be spatially uniform, and $\pi/p_0 \sim O(\mathrm{Ma}^2)$, where Ma is the local Mach number. For example, the ideal gas law

$$p_0 = \rho \mathcal{R} T \sum_k \frac{Y_k}{W_k}, \tag{4}$$

where $\mathcal{R}$ is the universal gas constant, T is the temperature and W_k is the molecular weight of species k. In this model, the thermodynamic pressure p_0 is decoupled from the dynamic pressure π. Pressure waves are assumed to propagate infinitely fast, and a divergence constraint can be derived for the velocity field by substituting Eqs. 1-3 into Eq. 4 and setting $Dp_0/Dt = 0$, given by[10]

$$\nabla \cdot \mathbf{u} = \frac{1}{T}\frac{DT}{Dt} + W \sum_k \frac{1}{W_k}\frac{DY_k}{Dt} = S. \tag{5}$$

Note that S is non-zero in regions close to the flame; a simple generalisation of this procedure can accommodate time-varying p_0[11]. By formulating the problem in the low Mach limit, rather than compressible, much larger timesteps can be taken and the stability condition for timestepping is now given by

$$\Delta t_{LM} = \frac{CFL\Delta x}{||\mathbf{u}||_\infty} \gg \Delta t_C = \frac{CFL\Delta x}{|||\mathbf{u}| + c||_\infty}, \tag{6}$$

where c is the local speed of sound. Hence, $\Delta t_{LM}/\Delta t_C \sim Ma^{-1}$ and for many combustion applications $Ma < 0.1$, providing at least 10 times larger timesteps.

For the diffusive fluxes, the following model is used

$$\mathbf{F}_k = -\rho D_{k,mix}\frac{\overline{W}}{W_k}\nabla Y_k - \rho D_{k,mix}\frac{Y_k}{W_k}\nabla \overline{W} - \rho D_{k,mix}\chi_k \frac{\nabla T}{T}, \tag{7}$$

where $D_{k,mix}, W_k$, and χ_k are the mixture-averaged diffusion coefficient, molecular weight and thermal diffusion ratio, respectively, of species k and $\overline{W}$ is the mean molecular

weight. For the chemical source terms in Eq. 2, each is represented by a sum of fundamental Arrhenius reactions, given by

$$\dot{\omega}_k = \sum_{i=1}^{R} (\nu_{k,i}^{(f)} - \nu_{k,i}^{(b)}) \Omega_i \tag{8}$$

$$\Omega_i = k_i^{(f)} \prod_{l=1}^{N} \left(\frac{\rho Y_l}{W_l} \right)^{\nu_{k,i}^{(f)}} - k_i^{(b)} \prod_{l=1}^{N} \left(\frac{\rho Y_l}{W_l} \right)^{\nu_{k,i}^{(b)}} \tag{9}$$

$$k_i^{(f)} = A_i T^{\beta_i} \exp\left\{ \left(-\frac{E_i}{\mathcal{R}T} \right) \right\} \tag{10}$$

where R is the number of reactions, $\nu_{k,i}^{(f/b)}$ are the forward/backward stoichiometric coefficients of the kth species in the ith reaction, A_i, β_i and E_i are the pre-exponential factor, pre-exponential temperature exponent and activation energy of the ith reaction. The backwards rate $k_i^{(b)}$ is determined through an equilibrium assumption. In many combustion reaction systems, these simple expressions are modified to include third-body accelerations and pressure dependencies.

PeleLMeX discretises each of the equations above using a second-order Godunov (finite-volume) approach on a Cartesian grid with constant grid spacing on each level of mesh refinement. The terms relating to advection are treated explicitly, with several advection schemes available. In particular, the bound-preserving BDS scheme[12] prevents scalar overshoots or undershoots that would otherwise represent unphysical states. The diffusion and reaction terms are both treated implicitly. The linear system arising from the implicit diffusion is solved using the in-built linear solvers provided by AMReX. For the chemical source term, a sub-cycled solve is performed in each cell by solving a local system of ordinary differential equations which are offloaded to a suite of integrators, available via SUNDIALS[13]. To evolve the velocity field, constrained by condition 4, a projection method is used[14], again using the linear solvers provided by AMReX.

The most complex part of the algorithm is the coupling between processes of advection, diffusion and reactions, which typically exhibit widely varying timescale. PeleLMeX uses an iterative spectral deferred correction (SDC) approach[11], rather than a more typical Strang splitting[15] where large operator splitting errors can occur when either the diffusion or reactions exhibit timescales much shorter than those of advection. Unlike Strang splitting, where each term is integrated sequentially in a somewhat decoupled fashion, each SDC iteration provides a lagged coupling of all terms in each process.

To evaluate the physical properties required for the diffusion and chemistry, PelePhysics provides a recasting software called CEPTR (Chemistry Evaluation for Pele Through Recasting) that converts standard YAML format used for Cantera[16] into C++ code that can be readily compiled.

3 Code Performance

In this section, the performance of the code with respect to both scaling and GPU acceleration of the code is presented.

3.1 Scaling

PeleLMeX has been used extensively on the JUWELS Booster module and is also being tested through the JUREAP program to demonstrate suitability on the early access module (JEDI) of Europe's first exascale machine, JUPITER. Prior to tests in Europe, PeleLMeX has also been extensively tested on HPC systems in the US, including the exascale computing system Frontier. For both weak and strong scaling a slot jet configuration is used, similar to the simulations presented in Sec. 4. An ammonia/hydrogen/air mixture with an equivalence ratio of 0.6, temperature of 575K and hydrogen blend fraction of 40% is used at the inlet, and the thermodynamic pressure is set to 10 atm. A thermodynamic, transport and chemistry model containing 30 species and 243 reactions[17] is employed. Such models involving ammonia chemistry are typically numerically stiff (that is, exhibit a broad range of time scales, many much shorter than those of advection) due to some reactions containing negative activation energies. However, the CVODE integrator provided through SUNDIALS, was found to handle these cases efficiently, particularly when using the modified Newton solver strategy, where the underlying linear solves are performed using the MAGMA linear algebra package[18].

# of nodes	Base grid	Effective grid	# of cells
1	$512 \times 16 \times 512$	$4096 \times 128 \times 4096$	2.24×10^7
2	$512 \times 32 \times 512$	$4096 \times 256 \times 4096$	4.48×10^7
4	$512 \times 64 \times 512$	$4096 \times 512 \times 4096$	8.96×10^7
8	$512 \times 128 \times 512$	$4096 \times 1024 \times 4096$	1.79×10^8
16	$512 \times 256 \times 512$	$4096 \times 2048 \times 4096$	3.58×10^8
32	$512 \times 512 \times 512$	$4096 \times 4096 \times 4096$	7.17×10^8

Table 1. Weak scaling cases on JEDI.

# of nodes	Advection (s)	Diffusion (s)	Chemistry (s)	Pressure (s)	Total (s)
1	0.208	1.48	5.11	1.48	8.85
2	0.144	1.64	5.76	1.14	9.38
4	0.198	1.79	5.83	1.04	9.61
8	0.155	1.93	6.07	1.03	9.99
16	0.211	1.98	6.07	0.995	10.0
32	0.196	2.37	6.26	1.08	11.0

Table 2. Timing of the weak scaling cases on JEDI.

The different cases for weak scaling for JEDI are shown in Tab. 1. The solver is run for 100 timesteps and the average timings for each section at each timestep of the code at the specified number of nodes is given in Tab. 2; the parallel efficiency is plotted in Fig. 3.Very good weak scaling is seen up to a simulation size of 717 million cells with less than a 20% loss in efficiency when scaling up from 20 million cells on 1 node. Without the AMR, this

simulation would have an effective resolution of over 68 billion cells to obtain the same resolution at the flame.

For the strong scaling, a single case was devised to fit on 1 node of JEDI. The strong scaling timing results on JEDI are shown in Tab. 3. A more comprehensive scaling was also performed on JUWELS Booster as part of the JUREAP program, and the weak and strong scaling from this is shown in Fig. 3. As can be seen, the portions of the solver that are highly-local, i.e. the explicit advection and subcycled chemistry solve, exhibit good strong scaling. However, portions of the solver that require significant communication due to the implicit or elliptic solve, i.e. the diffusion and pressure portions, do not scale as well in the strong sense. Such behaviour has been noted in other CFD solvers on heterogeneous architectures[19] regardless of computational scheme and is associated with communication costs among GPUs.

# of nodes	Advection (s)	Diffusion (s)	Chemistry (s)	Pressure (s)	Total (s)
1	0.198	1.51	6.51	0.754	9.49
2	0.0785	1.18	3.56	0.615	5.86
4	0.0436	0.933	2.12	0.540	3.90
8	0.0285	0.744	1.33	0.451	2.74
16	0.0193	0.705	0.830	0.460	2.16
32	0.0151	0.680	0.791	0.465	2.09

Table 3. Timing of the strong scaling case on JEDI.

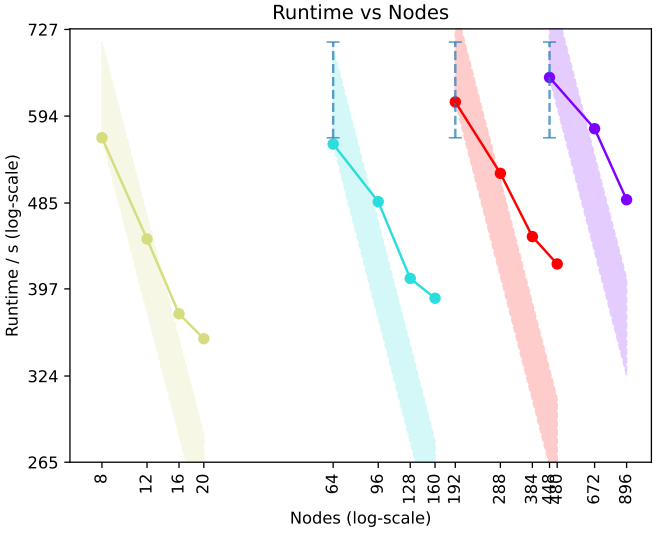


Figure 3. Combined weak and strong scaling plot for the JUWELS Booster, extending up to the full machine (896 nodes).

3.2 CPU vs. GPU

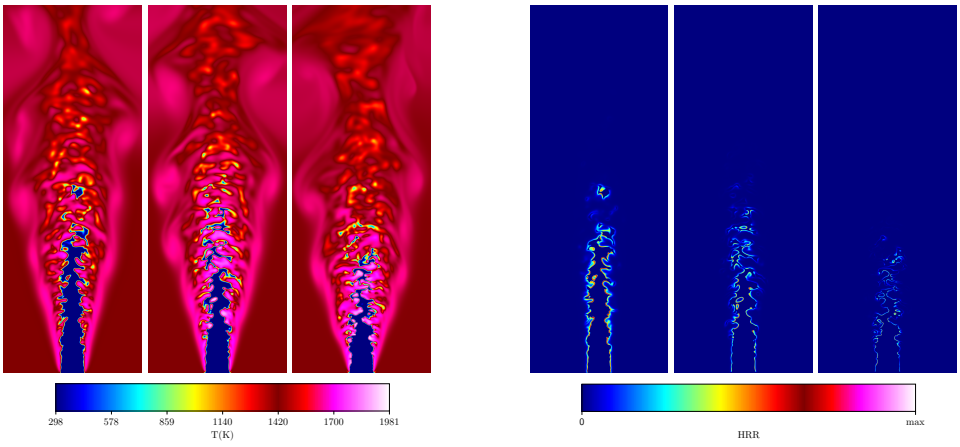
In addition to scaling on the GPUs, a single node comparison between a CPU run and GPU run has been made on the JUWELS Cluster and JUWELS Booster. Tab. 4 profiles each part of the solver running on either CPUs or GPUs. On a per-compute-node basis, where each Cluster compute node has 48 CPUs, and each Booster compute node has 4 GPUs, usage of GPU resources results in a time-to-solution speedup of 13.8 times. On a per-MPI-process basis, this results in an approximate speedup of 166 times. It should be noted that these results are sensitive to both the problem configuration and machine infrastructure. Nevertheless, this demonstrates the tremendous potential for much larger simulations at reduced cost.

	Advection (s/-)	Diffusion (s/-)	Chemistry (s/-)	Pressure (s/-)	Total (s/-)
CPU	4.01	19.1	57.2	5.79	109
GPU	0.0979	1.52	4.96	0.738	7.90
Ratio	41.0	12.6	11.5	7.85	13.8

Table 4. Comparison between single node CPU/GPU run on JUWELS Cluster/Booster

4 Simulations

To showcase the application of PeleLMeX and the physics that can be explored, a series of DNS of lean premixed turbulent hydrogen/air flames are presented in Fig. 4. In the slot burner configuration, a homogeneous cold ($T_\mathrm{u} = 298$ K) mixture of hydrogen and air with equivalence ratio $\phi = 0.4$ is entering the domain in the centre at a bulk velocity of



(a) Temperature

(b) Heat release rate

Figure 4. Lean premixed H_2/air jet flame at a $Re = 11000$ and $p_0 = 1, 5, 10$atm.

$u = 21\mathrm{m/s}$ with a prescribed turbulent velocity field obtained from an auxiliary periodic channel DNS. This central jet is surrounded by a slower laminar co-flow, which consists of hot ($T_{\mathrm{coflow}} = 1423$ K) combustion products. The temperature of the co-flow resemble the adiabatic flame temperature and the composition of the final combustion state of the combustible mixture in a 1D case. For the simulations shown, the spatially uniform thermodynamic pressure is varied from 1 to 10atm. These simulations are used to investigate thermodiffusive instabilities (TDI), which arise from the high diffusivity of hydrogen relative to the thermal diffusion, leading to an enthalpy accumulation in the flame areas convexly curved to the unburned. As a result, local temperatures rise above the adiabatic flame temperature and local burning rates increase strongly. Since TDI can easily increase the global fuel consumption speed by a factor of 4 or more and can dramatically increase post-flame gas temperatures where NOx emissions are produced, understanding them is highly relevant for model development and real-world applications. In the simulations presented in Fig. 4(a), TDI can be recognised by the local temperature overshoots compared to the co-flow. TDI ten to act on the smallest scales, forming structures which then grow onto larger scales. As a result, the resolution requirements are high since both the smallest scales of the flow field and the flame structure need to be resolved simultaneously.

Investigating high-pressure configurations is important and challenging at the same time. On the one hand, most real-world combustion applications are operating at elevated pressure. On the other hand, these simulations are especially challenging from a computational perspective, since the flame thickness decreases significantly with pressure, as visualised by the local heat release rate in Fig. 4(b). Since resolving the flame in the simulation is a key aspect of DNS, this leads to a higher resolution criterion and hence increases computational costs with increasing pressure. Here, PeleLMeX with its adaptive mesh refinement comes in handy, as the higher resolution needs only be applied to the relatively small flame area, while large portions of the domain, such as the co-flow, are adequately resolved with larger grid spacing. Investigating these cases in Fig. 4(a) reveals that instabilities become stronger, featuring higher temperature overshoots and an overall shorter flame.

5 Concluding Remarks

PeleLMeX is a reacting flow solver capable of running on heterogeneous HPC architectures by leveraging the AMReX library. Here, the models and numerical methods employed by the solver have been reviewed, and the performance of the code on JEDI and JUWELS has been demonstrated, with excellent weak scaling and satisfactory strong scaling performance shown. A direct comparison of a single node CPU run on the JUWELS Cluster module versus a single node GPU run on the JUWELS Booster module has also been shown, and it was found that, on a per node basis, there was an approximately 14 times speedup. Finally, three cases of pressurised turbulent hydrogen jet flames were presented to demonstrate the physics that can be explored. The results shown here demonstrate the potential for PeleLMeX to be used on exascale computing systems for the simulation and development of carbon-free combustion technology.

Acknowledgements

TLH, TL, MG and HP acknowledge funding from the European Union (ERC, HYDROGENATE, 101054894). The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) for funding this project by providing computing time on the GCS Supercomputer JUWELS at Jülich Supercomputing Centre (JSC) under the projects h2ex and instahype-2. Additionally, time was provided on JEDI at JSC through the JUREAP program under project jureap34. This work was authored in part by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

References

1. H. Pitsch, *The transition to sustainable combustion: Hydrogen- and carbon-based future fuels and methods for dealing with their challenges*, Proc. Comb. Inst. **40**, 1-4, 2024.
2. AMReX-Combustion, https://amrex-combustion.github.io.
3. M. Henry de Frahan, J. Rood, M. Day, H. Sitaraman, S. Yellapantula, B. Perry, R. Grout, A. Almgren, W. Zhang, J. Bell, and J. Chen, *PeleC: An adaptive mesh refinement solver for compressible reacting flows*, The International Journal Of High Performance Computing Applications **37**, 115-131, 2022.
4. L. Esclapez, M. Day, J. Bell, A. Felden, C. Gilet, R. Grout, M. Henry de Frahan, E. Motheau, A. Nonaka, L. Owen, B. Perry, J. Rood, N. Wimer, and W. Zhang, *PeleLMeX: an AMR Low Mach Number Reactive Flow Simulation Code without level sub-cycling*, Journal Of Open Source Software **8**, 5450, 2023.
5. W. Zhang, A. Almgren, V. Beckner, J. Bell, J. Blaschke, C. Chan, M. Day, B. Friesen, K. Gott, D. Graves et al., *AMReX: a framework for block-structured adaptive mesh refinement*, The Journal Of Open Source Software **4**, 1370, 2019.
6. A. Almgren, M. Sazo, J. Bell, A. Harpole, M. Katz, J. Sexton, D. Willcox, W. Zhang, and M. Zingale, *CASTRO: A Massively Parallel Compressible Astrophysics Simulation Code*, Journal Of Open Source Software **5**, 2513, 2020.
7. J. Musser, A. Almgren, W. Fullmer, O. Antepara, J. Bell, J. Blaschke, K. Gott, A. Myers, R. Porcu, D. Rangarajan, M. Rosso, W. Zhang, and M. Syamlal, *MFIX-Exa: A path toward exascale CFD-DEM simulations*, The International Journal Of High Performance Computing Applications **36**, 40-58, 2022.
8. A. Sharma, M. Brazell, G. Vijayakumar, S. Ananthan, L. Cheung, N. DeVelder, M. Frahan, N. Matula, P. Mullowney, J. Rood et al., *ExaWind: Open-source CFD for hybrid-RANS/LES geometry-resolved wind turbine simulations in atmospheric flows*, Wind Energy **27**, 225-257, 2024.
9. A. Almgren, A. Lattanzi, R. Haque, P. Jha, B. Kosovic, J. Mirocha, B. Perry, E. Quon, M. Sanders, D. Wiersema et al., *ERF: energy research and forecasting*, The Journal Of Open Source Software **8**, 5202, 2023.

10. M. Day and J. Bell, *Numerical simulation of laminar reacting flows with complex chemistry*, Combustion Theory And Modelling **4**, 535, 2000.
11. A. Nonaka, M. Day, and J. Bell, *A conservative, thermodynamically consistent numerical approach for low Mach number combustion. Part I: Single-level integration*, Combustion Theory And Modelling **22**, 156-184, 2018.
12. J. Bell, C. Dawson, and G. Shubin, *An unsplit, higher order Godunov method for scalar conservation laws in multiple dimensions*, Journal Of Computational Physics **74**, 1-24, 1988.
13. C. Balos, D. Gardner, C. Woodward, and D. Reynolds, *Enabling GPU accelerated computing in the SUNDIALS time integration library*, Parallel Computing **108**, 102836, 2021.
14. A. Almgren, J. Bell, P. Colella, L. Howell, and M. Welcome, *A conservative adaptive projection method for the variable density incompressible Navier–Stokes equations*, Journal Of Computational Physics **142**, 1-46, 1998.
15. G. Strang, *On the Construction and Comparison of Difference Schemes*, SIAM Journal On Numerical Analysis **5**, 506-517, 1968.
16. D. Goodwin, H. Moffat, I. Schoegl, R. Speth, and B. Weber, *Cantera: An Object-oriented Software Toolkit for Chemical Kinetics, Thermodynamics, and Transport Processes*, Version 3.0.0, 2023, `https://www.cantera.org`.
17. X. Zhang, S. Moosakutty, R. Rajan, M. Younes, and S. Sarathy, *Combustion chemistry of ammonia/hydrogen mixtures: Jet-stirred reactor measurements and comprehensive kinetic modeling*, Combustion And Flame **234**, 111653, 2021.
18. A. Abdelfattah, N. Beams, R. Carson, P. Ghysels, T. Kolev, T. Stitt, A. Vargas, S. Tomov, and J. Dongarra, *MAGMA: Enabling exascale performance with accelerated BLAS and LAPACK for diverse GPU architectures*, The International Journal Of High Performance Computing Applications **38**, 468-490, 2024.
19. M. Min, M. Brazell, A. Tomboulides, M. Churchfield, P. Fischer, and M. Sprague, *Towards exascale for wind energy simulations*, The International Journal Of High Performance Computing Applications **38**, 337-355, 2024.

Driving Green Combustion Innovation: High-Fidelity Direct Numerical Simulations and Large-Scale Machine Learning

Mathis Bode[1], **Driss Kaddar**[2], **Hendrik Nicolai**[2], **and Christian Hasse**[2]

[1] Jülich Supercomputing Centre (JSC), Forschungszentrum Jülich, 52425 Jülich, Germany
E-mail: m.bode@fz-juelich.de

[2] Simulation of Reactive Thermo-Fluid Systems, TU Darmstadt, 64277 Darmstadt, Germany
E-mail: {kaddar, nicolai, hasse}@stfs.tu-darmstadt.de

The climate crisis is arguably the most pressing challenge of our time. To create effective solutions, such as gas turbines fuelled by green hydrogen, it is crucial to fully understand complex physical phenomena. This understanding is key to developing models that can accurately predict flame dynamics and pollutant formation in these innovative systems. In addition to costly and time-consuming experimental investigations, the urgent demand for carbon-free fuels requires highly accurate direct numerical simulations to improve physical understanding and support model development.

Over the last three years, we have developed a highly scalable simulation code called nekCRF at the Center of Excellence in Combustion (CoEC), which is based on nekRS, and created relevant direct numerical simulation (DNS) setups for simulating turbulent hydrogen jet flames. The results of these large-scale simulations presented here enable the analysis of the complex physics of hydrogen combustion and the development of the next generation of combustion models. As we transition to upcoming exascale systems, these simulations will push the boundaries of hydrogen mixture predictions and high-pressure simulations well beyond current limits, paving the way for new industrial technologies.

In addition, an innovative machine learning framework has been developed to quickly develop highly accurate simulation models from the generated data. This framework uses highly parallelised processes to develop accurate models for large-eddy simulations (LESs), which can be used for the final optimisation of relevant industrial processes.

Overall, our work is an example of how the interplay of highly accurate simulations and large-scale machine learning can help solve societal problems such as the green energy transition using the latest supercomputers. The efficient parallelisation of the entire workflow makes it possible to benefit optimally from the ever more powerful supercomputers and to significantly reduce time-to-innovation.

1 Introduction

Addressing climate change requires transitioning away from fossil fuels in our energy system. Gas turbines, central to power generation and aviation, have traditionally relied on hydrocarbon fuels. However, new turbine technologies can be developed to efficiently use zero-carbon fuels, supporting a sustainable energy future. In low-carbon energy systems, with intermittent renewables like solar and wind, hydrogen (H_2) has emerged as a carbon-free option for large-scale energy storage. Its chemical simplicity and lack of carbon enable combustion for power and heat generation with zero CO_2 emissions and minimal pollutants like NOx.

The direct use of hydrogen presents significant challenges: lean hydrogen flames tend to become unstable, with turbulent burning rates far exceeding those of conventional hydrocarbon fuels. Hydrodynamic Darrieus-Landau (DL) instabilities and intense thermo-diffusive (TD) instabilities, driven by differential diffusion, have a pronounced effect on H_2 and H_2-enriched flames. These instabilities, depicted in Fig. 1 (left), arise from the large differences in diffusion rates between hydrogen, heat, and other species. They can destabilise flames, increasing the risk of flashback and raising NOx emissions due to super-adiabatic temperatures or hot spots. These effects are amplified under the high pressures typical of practical systems. Currently, our understanding of burning rates, stability limits, and emission formation in high-pressure lean H_2 flames is based largely on empirical data, which poses a significant challenge for industrial applications. Closing this knowledge gap is essential for developing clean, efficient, and reliable hydrogen-fired gas turbines.

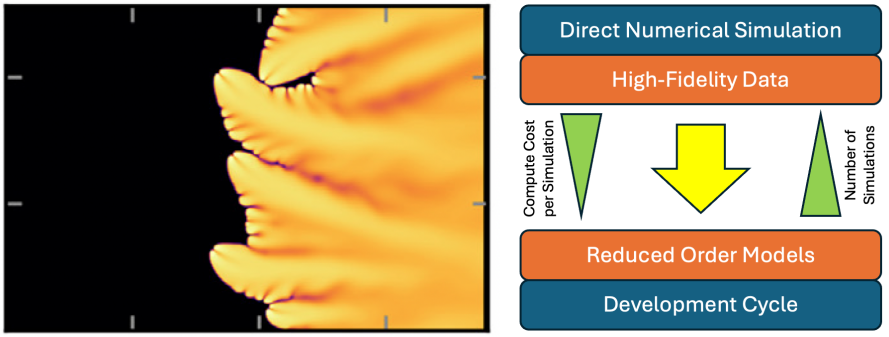


Figure 1. DNS of a laminar hydrogen flame exhibiting flame finger formation due to thermo-diffusive instabilities[1] (left) and sketch of the innovation process (right).

An established method for developing new energy devices is the combination of high-resolution direct numerical simulations (DNSs) and the reduced order models (ROMs) derived from them. DNSs can be assumed to be the ground-truth of individual physical and chemical sub-processes and thus contribute to the understanding of these processes under well-defined boundary conditions. However, they are limited in two ways: On the one hand, it is sometimes very challenging to consider all important physical and chemical sub-processes simultaneously based on first principles. On the other hand, DNSs of industrially relevant conditions quickly become very computationally intensive and can therefore only be carried out in individual cases. It is usually not possible to carry out DNSs iteratively more often within a development cycle. Therefore, computationally less expensive ROMs must usually be used here, and deriving predictive ROMs from DNS data effectively becomes another key challenge (in addition to the DNS data generation itself). This innovation process is sketch in Fig. 1 (right).

In the realm of turbulent combustion modelling, large-eddy simulation (LES) has emerged as the primary research tool[2]. LES is favoured because it captures the larger turbulent scales, which includes the majority of turbulent kinetic energy and exhibit a strong dependence on the device geometry. Advancements in turbulence-chemistry interaction (TCI) models have been relatively modest during this period. However, due to the unique

characteristics of H_2 as a fuel, LES faces new challenges[3, 1, 4]. Firstly, H_2's high reactivity results in thinner reaction fronts, counteracting the benefits of increased computational power for spatial resolution. Secondly, issues like DL and TD instabilities occur at subgrid scales and cannot be predicted by existing TCI models.

The process of deriving predictive models for LES from DNS data can be very time consuming. Conventionally, one tries to derive analytical relations by which unknown or generally subgrid quantities can be determined. More recently, this manual process has been attempted to be replaced by data-driven methods, such as machine learning (ML), deep learning, and artificial intelligence. In this paper, we describe how such a DNS of a relevant flame can be calculated using nekCRF, and how highly accurate LES models can be derived from the data using the data-driven framework JuLES. We demonstrate that the whole process can be efficiently parallelised on thousands of GPUs, thus significantly reducing the time to innovation.

2 DNS Methods and Setup

2.1 Setup

The configuration of the DNS is based on a turbulent jet burner that has also been experimentally investigated. This type of burner is known as a McKenna burner. A schematic setup is shown in Fig. 2. The burner has three separate gas outlets: an outer ring made of sintered bronze through which shielding gas flows, an inner ring made of sintered steel for the pilot fuel-air mixture, and a central tube from which the fuel mixture for the main flame is emitted.

The outer ring made of sintered bronze, referred to as the co-flow, is supplied with inert nitrogen gas during operation. This creates a gas stream that forms a protective layer around the inner flames, preventing ambient air pressure fluctuations from reaching the flames and ensuring uniform conditions. The inner sintered steel matrix, referred to as the pilot, is operated with a mixture of ammonia, hydrogen, nitrogen, and air. Igniting

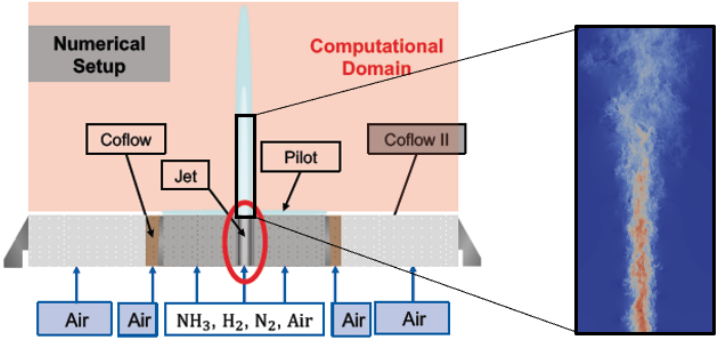


Figure 2. Schematic depiction of the McKenna burner configuration. Zoom shows the non-reactive DNS of the jet-flow field.

this mixture generates a flat, laminar flame just above the matrix, preventing cold ambient air from entering the main flame. Additionally, the pilot flame and the hot exhaust gases provide thermal energy to the reaction zone of the main flame, ensuring a stable flame within the pilot's influence zone over a wide range of operating conditions. The central burner tube, known as the jet, is also supplied with a mixture of hydrogen and air. When ignited, this mixture forms the main flame of the burner, whose properties are examined in this work by varying different mixture parameters. Depending on the operating conditions, the flame ranges from laminar to highly turbulent. The fuel gases for both the pilot and the jet are premixed by combining the gas supply lines, ensuring that the mixture is already homogeneous before entering the burner.

For the numerical setup, a domain similar to the experimental configuration is used. However, in this study, only the jet stream and the flue gases from the pilot stream are considered. The domain has a length of 100 mm. The mesh is constructed using multiple O-grids, allowing for varying refinement levels to ensure proper resolution of the inlet pipe. High resolution near the inlet and flame zone is crucial for accurately capturing turbulence and combustion properties, while the mesh becomes coarser towards the outlet and the edges of the domain. The final mesh consists of 800 million unique grid points.

2.2 Numerical Methods

As simulation framework, nekRS[5] is used in combination with the chemistry plugin nekCRF[6] developed as part of Center of Excellence in Combustion (CoEC). It employs high-order spectral elements in which the solution, data, and test functions are represented as locally structured Nth-order tensor product polynomials on a set of E globally unstructured conforming hexahedral brick elements. Time integration in nekRS employs a semi-implicit splitting scheme, utilising kth-order (k up to three) backward differences (BDFk) to approximate the time derivative, resulting in an implicit treatment of the viscous and pressure terms, and kth-order extrapolation (EXTk) for the advection and forcing terms.The discretisation leads to a sequence of symmetric positive definite linear systems for pressure, velocity and temperature.

NekRS is written in C++ and the kernels are implemented using the portable Open Concurrent Compute Abstraction (OCCA) library[7] in order to abstract between different parallel languages. OCCA enables the implementation of the parallel kernel code in the slightly decorated C++ language OKL[7].

In this way, the MPI+X hybrid parallelism can support seamlessly CUDA, HIP, OpenCL as well as CPUs. The domain is partitioned on MPI ranks and the discretised equations are advanced in time using iterative solvers to solve the elliptic subproblems for the velocity, temperature, and pressure.

The chemistry plugin nekCRF is fully integrated into the programming approach of nekRS. The thermochemistry (energy and species equations) is treated with this highly optimised chemistry plugin that generates optimised kernels for the source term, thermodynamic and transport properties evaluation for GPUs. It also provides consistent advection and diffusion transport operators acting efficiently on multiple scalars. The resulting large system is integrated without further splitting of the convection, diffusion and reaction term using CVODE[8]. In addition, important features for high-performance GPU computing of reactive flow simulations, such as an approximate Jacobian-vector product, a compressed

basis GMRES solver using a lower precision (FP32) Krylov basis, and overlap MPI communication for halo exchange with local computation are available.

2.3 Simulation Results

Fig. 3 shows the instantaneous temperature field. Characteristic for a lean premixed hydrogen/air mixture, the flame locally exhibits super-adiabatic temperature indicating the persisting influence of Lewis number effects under turbulent conditions.

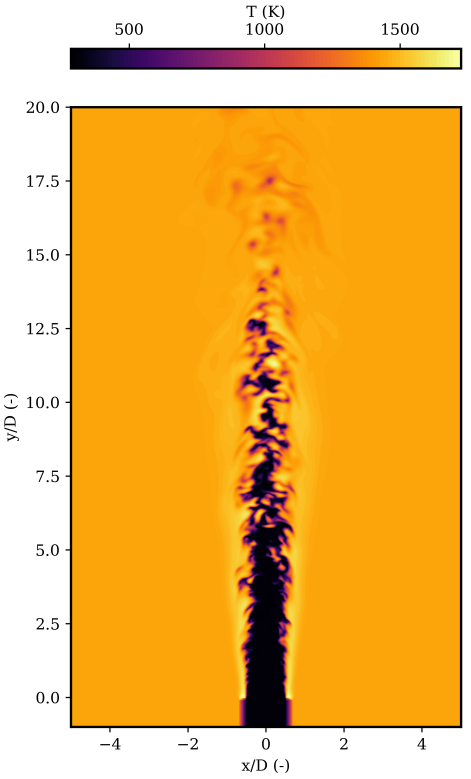


Figure 3. Contour of the instantaneous temperature field.

The influence of molecular transport effects are further highlighted in Fig. 4. Here, the local equivalence ratio exhibits strong variations across the turbulent flame front. Additionally, high sensitivity of the equivalence ratio on the flame curvature is observed. Correspondingly, the OH mass fraction indicates the increased reactivity in the fuel-rich regions leading to super-adiabatic temperatures.

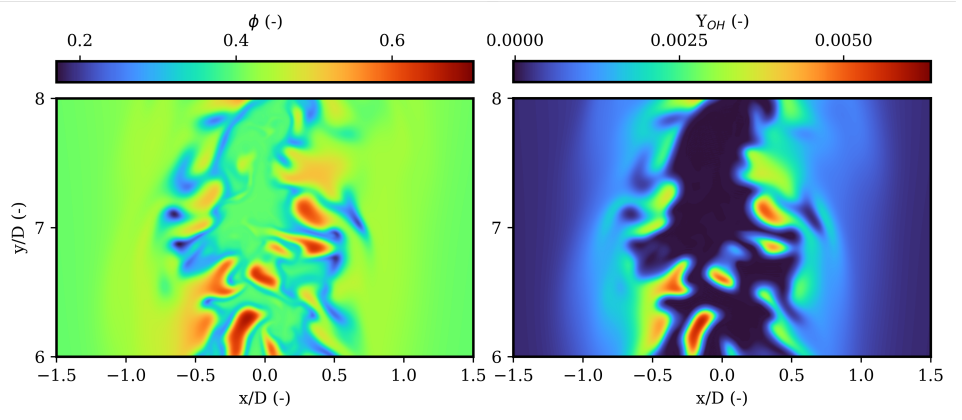


Figure 4. Zoomed view of the contours of the instantaneous local equivalence ratio (left) and OH mass fraction (right).

3 Machine Learning and Application

3.1 Application

ML has the potential to significantly accelerate the process of LES-subgrid development. The combination of high-resolution simulations and cutting-edge ML also has the potential to derive much more general and predictive subgrid models. The idea is to filter highly accurate DNS data and to train an ML network using the data pairs generated in this way (or a data collection if several different filter widths are applied), which is able to add the subgrid contributions to the filtered data. This can be done in a post-processing step or directly in parallel with data generation. The parallel approach has the advantage of minimising "expensive" writing of data to storage and also avoids further data transfer during processing. A disadvantage can be that if not all intermediate steps are stored, the training is not completely reproducible.

On the user side, the combination of simulation and ML adds another layer of complexity. To make the approach easier for users to handle, we have developed the JuLES framework. JuLES enables both training in a post-processing step and parallel training using an $n : m$ data bridge. The framework is integrated into the compute infrastructure of the Jülich Supercomputing Centre (JSC); this includes both the supercomputers and the interactive JupyterLab interface, as well as experimental access to the quantum computer. Features for simulation monitoring and control, such as JuMonC[9], are supported as well. JuLES is shown in Fig. 5.

3.2 Methods

The so-called physics-informed enhanced super-resolution GAN (PIESRGAN)[10] is used as the ML network in JuLES, which extends and modifies ESRGAN for flow LES subgrid modelling and follows a hybrid approach. The subgrid modelling is done by PIESRGAN, but the simulation as a whole is advanced classically with the filtered equations, i. e., the

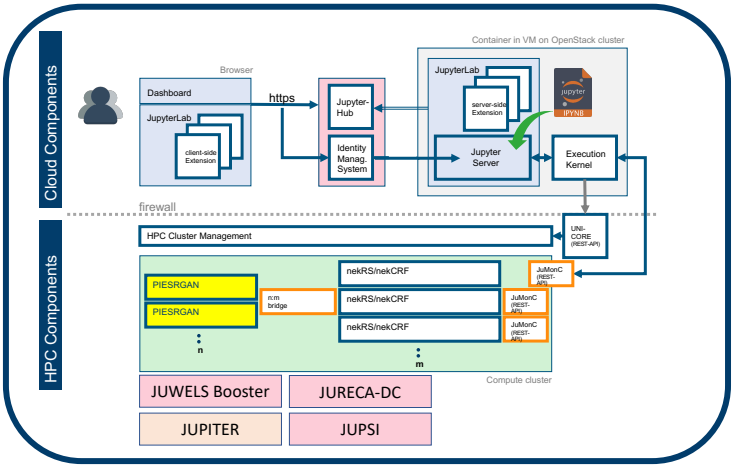


Figure 5. Overview of the framework JuLES.

time integration is not incorporated into the neural network as, e. g., done in Ref. 11. An important advantage of keeping time integration and subgrid modelling separate is the universality with respect to, e. g., different geometries and setups. In addition, the equations for time integration are well known and much experience is available. In PIESRGAN, physically motivated conditions are enforced as part of the loss function, i.e., the objective function that is minimised during training of the network. PIESRGAN has been widely applied to various physical flow simulations and results are available in the literature[12, 10, 13–20].

The physically motivated loss term is very important for the application of PIESRGAN to flow problems. If the conservation laws are not very well fulfilled, the simulations tend to explode rapidly, which is an important difference to super-resolution in the context of images. Errors that may be acceptable there can easily be too large for use as a subgrid model[10].

3.3 Prediction Results

The prediction accuracy of a subgrid model that was trained in parallel by JuLES was evaluated on the basis of LES of hydrogen flames with different filter widths. The maximum relative error of all thermo-chemical variables was analysed, as shown in Tab. 1. Regardless of the filter width and even with a flexible filter width, the maximum relative error is always below 1.37 %.

4 Conclusions

Both nekCRF and JuLES are powerful tools for generating highly accurate DNS data from flames and automatically developing predictive LES models from this data. Both are es-

Filterwidth	$2\mathrm{d}x$	$4\mathrm{d}x$	$8\mathrm{d}x$	$2\mathrm{d}x < \{\cdot\} < 8\mathrm{d}x$
Error [%]	0.91	1.37	0.85	1.14

Table 1. Prediction quality of ML-trained subgrid models.

sential to advancing the green energy transition as quickly as possible and to becoming climate neutral.

The degree of acceleration depends on the ability of the tools to be accelerated in parallel. Both nekCRF and JuLES use GPUs for simulation and training, respectively. This is already a significant acceleration compared to CPU-based solutions. Furthermore, Fig. 6 demonstrates that nekCRF, but also the coupled workflow, scale effectively on thousands of GPUs. The time to solution and time to innovation can thus be further reduced by adding more and more computing power. We have successfully deployed both tools on up to 4000 GPUs in parallel on JUWELS Booster for computing green hydrogen flames and are optimistic that we will be able to further increase the number of applications as well as the degree of acceleration with JUPITER.

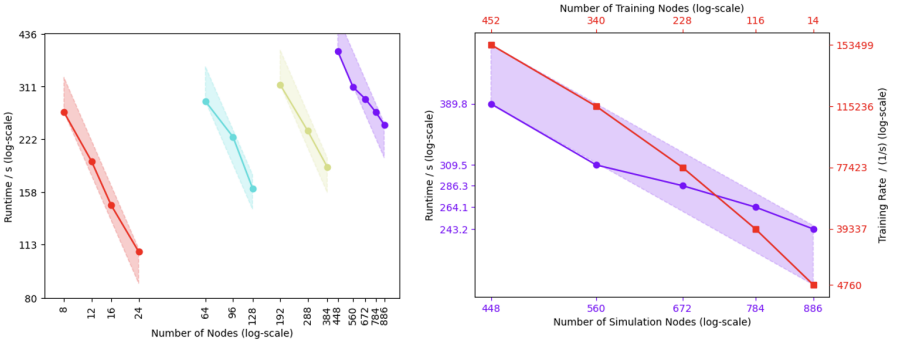


Figure 6. Scaling of nekCRF (left) and scaling/performance of coupled nekCRF/JuLES workflow (right). All numbers were evaluated on JUWELS Booster featuring four NVIDIA A100 per node.

Acknowledgements

This work was supported by the European Union's Horizon 2020 research and innovation program under grant agreements No. 952181 (Center of Excellence in Combustion) and has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No. 101118139 (Inno4Scale). The JU receives support from the European Union's Horizon Europe Program. The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (`www.gauss-centre.eu`) for funding this project by providing computing time on the GCS Supercomputer JUWELS at Jülich Supercomputing Centre (JSC) and by providing computing time through the John von Neumann Institute for Computing (NIC).

References

1. V. Schuh, C. Hasse, and H. Nicolai, *An extension of the artificially thickened flame approach for premixed hydrogen flames with intrinsic instabilities*, Proceedings of the Combustion Institute, **40**, no. 1-4, 105673, 2024.
2. B. Fiorina, T. P. Luu, S. Dillon, R. Mercier, P. Wang, L. Angelilli, P. P. Ciottoli, F. E. Hernández-Pérez, M. Valorani, H. G. Im et al., *A joint numerical study of multi-regime turbulent combustion*, Applications in Energy and Combustion Science, **16**, 100221, 2023.
3. H. Pitsch, *The transition to sustainable combustion: Hydrogen-and carbon-based future fuels and methods for dealing with their challenges*, Proceedings of the Combustion Institute, **40**, no. 1-4, 105638, 2024.
4. H. Böttler, D. Kaddar, T. J. P. Karpowski, F. Ferraro, A. Scholtissek, H. Nicolai, and C. Hasse, *Can flamelet manifolds capture the interactions of thermo-diffusive instabilities and turbulence in lean hydrogen flames? – An a-priori analysis*, International Journal of Hydrogen Energy, **56**, 1397-1407, 2024.
5. P. Fischer, S. Kerkemeier, M. Min, Y.-H. Lan, M. Phillips, T. Rathnayake, E. Merzari, A. Tomboulides, A. Karakus, N. Chalmers et al., *NekRS, a GPU-accelerated spectral element Navier-Stokes solver*, Parallel Computing, **114**, 102982, 2022.
6. S. Kerkemeier, C. E. Frouzakis, A. G. Tomboulides, P. Fischer, and M. Bode, *nekCRF: A next generation high-order reactive low Mach flow solver for direct numerical simulations*, 2024, arXiv:2409.06404.
7. D. S. Medina, A. St-Cyr, and T. Warburton, *OCCA: A unified approach to multi-threading languages*, 2014, arXiv:1403.0968.
8. S. D. Cohen, A. C. Hindmarsh, P. F. Dubois et al., *CVODE, a stiff/nonstiff ODE solver in C*, Computers in physics, **10**, no. 2, 138-143, 1996.
9. C. Witzler, F. Souza Mendes Guimarães, D. Mira, H. Anzt, J. H. Göbbert, W. Frings, and M. Bode, *JuMonC: A RESTful Tool for Enabling Monitoring and Control of Simulations at Scale*, Future Generation Computer System, **164**, 107541, 2025.
10. M. Bode, M. Gauding, Z. Lian, D. Denker, M. Davidovic, K. Kleinheinz et al., *Using physics-informed enhanced super-resolution generative adversarial networks for sub-filter modeling in turbulent reactive flows*, Proceedings of the Combustion Institute, **38**, 2617-2625, 2021.
11. K. Fukami, R. Maulik, N. Ramachandra, K. Fukagata, and K. Taira, *Global field reconstruction from sparse sensors with Voronoi tessellation-assisted deep learning*, Nature Machine Intelligence, **3**, 945-951, 2021.
12. M. Bode, M. Gauding, K. Kleinheinz, and H. Pitsch, *Deep learning at scale for sub-grid modeling in turbulent flows: regression and reconstruction*, Lecture Notes in Computer Science, **11887**, 541-560, 2019.
13. M. Bode, *Applying physics-informed enhanced super-resolution generative adversarial networks to large-eddy simulations of ECN Spray C*, SAE International Journal of Advances and Current Practices in Mobility, **4**, 2211-2219, 2022.
14. M. Bode, *Applying physics-informed enhanced super-resolution generative adversarial networks to finite-rate-chemistry flows and predicting lean premixed gas turbine combustors*, 2022, arXiv:2210.16219.

15. M. Bode, M. Gauding, D. Goeb, T. Falkenstein, and H. Pitsch, *Applying physics-informed enhanced super-resolution generative adversarial networks to turbulent premixed combustion and engine-like flame kernel direct numerical simulation data*, Proceedings of the Combustion Institute, **39**, 5289-5298, 2023.
16. M. Bode, *Applying physics-informed enhanced super-resolution generative adversarial networks to turbulent non-premixed combustion on non-uniform meshes and demonstration of an accelerated simulation workflow*, 2022, arXiv:2210.16248.
17. M. Bode, *AI super-resolution: Application to turbulence and combustion*, in: Machine learning and its application to reacting flows, Lecture Notes in Energy 44, N. Swaminathan and A. Parente, (Eds.), Springer, 2023.
18. M. Bode, *AI super-resolution-based subfilter modeling for finite-rate-chemistry flows: A jet flow case study*, SAE Technical Paper 2023-01-0200, 2023.
19. M. Bode, *AI super-resolution subfilter modeling for multi-physics flows*, in: Platform for Advanced Scientific Computing Conference (PASC '23), ACM, 2023.
20. M. Bode and J. H. Göbbert, *Acceleration of complex high-performance computing ensemble simulations with super-resolution-based subfilter models*, Computers and Fluids, **271**, 106150, 2024.

Deep Learning for Small Scale Dynamics of Turbulence

Dhawal Buaria

[1] Max Planck Institute for Dynamics and Self-Organization, 37077 Göttingen, Germany
E-mail: dhawal.buaria@ds.mpg.de

[2] Department of Mechanical Engineering, Texas Tech University, Lubbock, TX 79409, USA

Turbulent flows are characterised by a wide range of scales, with the scale range growing as some power of the non-dimensional flow parameter Reynolds number. Consequently, a faithful simulation of turbulence at high Reynolds numbers, as routinely encountered in nature and engineering, remains prohibitively expensive. A substantial computational burden comes from resolving the small-scale motions; thus, major emphasis is placed on understanding their universal aspects and thereafter exploiting it for modelling. Here, by leveraging physics-informed deep learning, we present a novel framework to capture and predict the small scale dynamics of turbulence, via the velocity gradient tensor. We consider the evolution equation of velocity gradients and obtain a functional closure of unclosed terms using deep neural networks. A massive simulation database, spanning two orders of magnitude in Reynolds number, is then utilised for training and validation. The model learns from low to moderate Reynolds numbers and successfully predicts statistics at both seen and higher unseen Reynolds number, demonstrating the viability of our approach over traditional modelling in capturing and predicting small-scale dynamics of turbulence.

1 Introduction

Turbulent flows, ubiquitous in both natural and technological applications, are characterised by strong and chaotic fluctuations spanning a wide range of interacting scales in space and time. These multiscale interactions are highly nonlinear, rendering the governing equations mathematically intractable. Consequently, turbulence has defied an adequate framework despite a sustained effort in physics, mathematics and engineering, and our present understanding remains incomplete, often relying on phenomenological approaches[1]. An essential notion in this regard is that of small scale universality[2], which forms the backbone of turbulence theories and models. It stipulates that, while the large scales are non-universal because of their dependence on flow geometry and energy injection mechanisms, such dependencies become progressively weaker as energy cascades to smaller scales, ultimately endowing them with some form of universality that depends only on a few parameters of the flow.

It follows that universality requires sufficiently large separation between the scales at which the energy is injected and those at which it is dissipated into molecular motion. This scale separation is determined by the non-dimensional parameter Reynolds number, Re; thus, investigating universality requires data at high Re. While such high Reynolds numbers are attainable in some laboratory flows and all geophysical flows, most quantities pertaining to small scales are still very difficult to measure[3]. Alternatively, direct numerical simulations (DNS) of the governing equations, where the entire range of scales is resolved on a computational mesh[4], provide information on every quantity desired. But, DNS is extremely expensive, with recent studies showing that its cost scales even faster than the traditional estimate of Re^3 (see e.g. Refs. 5–7). Thus, despite the rapid advances in high

performance computing, high-Re DNS, representative of natural and engineering flows, remains unattainable for decades to come.

Motivated by these considerations, we devise here an alternative approach based on machine learning techniques to characterise the small scales of turbulence. In recent years, the use of machine learning, especially deep learning, has ushered in a new paradigm in various disciplines[8]. The field of turbulence is no different and there has been a flurry of machine learning methods to improve turbulence modelling[9]. A vast majority of them utilise the framework of supervised learning[10], where neural networks are trained on input data against labelled output data, although other paradigms have also be used[11]. The learning is also often "physics informed", i.e., neural networks are designed to satisfy some physical constraints, enabling efficient learning including significantly improved accuracy and stability. The approach utilised here follows a broadly similar paradigm, but in a newly developed framework, specifically suited for small scales of turbulence. In particular, we capture the small scale dynamics of turbulence by training deep neural networks on existing DNS data at low and moderate Re, and demonstrate the capability for predicting their dynamics at both seen and higher unseen Re, with important consequences for turbulence simulations.

2 Governing Equations and Modelling Framework

2.1 Governing Equations

To characterise the small scales of turbulence, we will focus on the velocity gradient tensor $\mathbf{A} = \nabla \mathbf{u}$, where $\mathbf{u}$ is the turbulent velocity field. The tensor $\mathbf{A}$ encodes various structural and statistical properties of turbulence, which are known to be universal; for instance, the non-Gaussianity of its fluctuations and associated extreme events[12,13,6], the negative skewness of the longitudinal (or diagonal) components associated with the energy cascade[14], the preferential alignment of vorticity with the intermediate strain eigenvector[15,16]. The evolution equation for $\mathbf{A}$, obtained by taking the gradient of Navier-Stokes equations, is given as:

$$\frac{D\mathbf{A}}{Dt} = -\mathbf{A}^2 - \mathbf{H} + \nu \nabla^2 \mathbf{A}, \tag{1}$$

where D/Dt is the material (or Lagrangian) derivative, $H = \nabla\nabla P$ is the Hessian tensor of the kinematic pressure P and ν the kinematic viscosity. The above equation dictates that the velocity gradient tensor changes along a fluid element according to quadratic nonlinearity, the pressure effects and viscous diffusion. Since incompressibility give trace $\mathrm{Tr}(\mathbf{A}) = 0$, it follows that

$$\nabla^2 P = \mathrm{Tr}(\mathbf{H}) = -\mathrm{Tr}(\mathbf{A}^2)\,, \tag{2}$$

i.e., the pressure field is related to $\mathbf{A}$ through a Poisson equation, implying that the pressure Hessian is non-local, essentially coupling all scales of the flow.

In direct numerical simulations (DNS), the Navier-Stokes equations are directly solved on a large computational mesh by resolving all dynamically relevant scales[4]. Whereas other simulation paradigms resolve only a range of scales. For instance, in large-eddy simulation (LES), the large-scales are resolved on a mesh and the effects of small scales

are directly modelled. In contrast, our approach here is to directly develop a reduced-order closure model for $\mathbf{A}$. This can be accomplished by modelling the pressure Hessian and the viscous Laplacian terms explicitly in terms of $\mathbf{A}$, leading to a fully local description[17], i.e., the dynamics of $\mathbf{A}$ can be modelled by an ordinary differential equation (ODE), whereby statistical quantities of interest can be obtained, for example, by running Monte Carlo simulations of the ODE with arbitrary initial conditions, providing massive cost savings with respect to DNS.

2.2 Tensor Bases for Modelling

To obtain a functional closure, the pressure Hessian and viscous Laplacian terms in Eq. 1 have to be specified as tensor functions of $\mathbf{A}$. This can be most generally achieved using tensor representation theory[18], which allows us to express any desired (second order) tensor as a function of $\mathbf{A}$. This is achieved by expressing the desired tensor as a linear combination of tensors in an appropriate tensor basis constructed from $\mathbf{A}$, with the coefficients that are functions of the scalar basis of $\mathbf{A}$. To obtain the tensor and scalar bases the first step is to decompose $\mathbf{A}$ into its symmetric and skew-symmetric parts, which are the strain-rate and rotation rate tensors, respectively:

$$\mathbf{S} = \frac{1}{2}(\mathbf{A} + \mathbf{A}^{\mathrm{T}}) \,, \quad \mathbf{R} = \frac{1}{2}(\mathbf{A} - \mathbf{A}^{\mathrm{T}}) \,. \tag{3}$$

Using $\mathbf{S}$ and $\mathbf{R}$, a general bases for tensors and scalars can be constructed:

$$\begin{aligned}
&\mathbf{T}^{(1)} = \mathbf{S} \,, \;\; \mathbf{T}^{(2)} = \mathbf{SR} - \mathbf{RS} \,, \;\; \mathbf{T}^{(3)} = \mathbf{S}^2 - \frac{1}{3}\,\mathrm{Tr}(\mathbf{S}^2)\mathbf{I} \,, \\
&\mathbf{T}^{(4)} = \mathbf{R}^2 - \frac{1}{3}\,\mathrm{Tr}(\mathbf{R}^2)\mathbf{I} \,, \;\; \mathbf{T}^{(5)} = \mathbf{RS}^2 - \mathbf{S}^2\mathbf{R} \,, \\
&\mathbf{T}^{(6)} = \mathbf{SR}^2 + \mathbf{R}^2\mathbf{S} - \frac{2}{3}\,\mathrm{Tr}(\mathbf{SR}^2)\mathbf{I} \,, \\
&\mathbf{T}^{(7)} = \mathbf{RSR}^2 - \mathbf{R}^2\mathbf{SR} \,, \;\; \mathbf{T}^{(8)} = \mathbf{SRS}^2 - \mathbf{S}^2\mathbf{RS} \,, \\
&\mathbf{T}^{(9)} = \mathbf{R}^2\mathbf{S}^2 + \mathbf{S}^2\mathbf{R}^2 - \frac{2}{3}\,\mathrm{Tr}(\mathbf{S}^2\mathbf{R}^2)\mathbf{I} \,, \;\; \mathbf{T}^{(10)} = \mathbf{RS}^2\mathbf{R}^2 - \mathbf{R}^2\mathbf{S}^2\mathbf{R}
\end{aligned} \tag{4}$$

$$\begin{aligned}
&\mathbf{B}^{(1)} = \mathbf{R} \,, \;\; \mathbf{B}^{(2)} = \mathbf{SR} + \mathbf{RS} \,, \;\; \mathbf{B}^{(3)} = \mathbf{S}^2\mathbf{R} + \mathbf{RS}^2 \,, \\
&\mathbf{B}^{(4)} = \mathbf{R}^2\mathbf{S} - \mathbf{SR}^2 \,, \;\; \mathbf{B}^{(5)} = \mathbf{R}^2\mathbf{S}^2 - \mathbf{S}^2\mathbf{R}^2 \,, \;\; \mathbf{B}^{(6)} = \mathbf{SR}^2\mathbf{S}^2 - \mathbf{S}^2\mathbf{R}^2\mathbf{S}
\end{aligned} \tag{5}$$

$$\lambda_1 = \mathrm{Tr}(\mathbf{S}^2) \,, \;\; \lambda_2 = \mathrm{Tr}(\mathbf{R}^2) \,, \;\; \lambda_3 = \mathrm{Tr}(\mathbf{S}^3) \,, \;\; \lambda_4 = \mathrm{Tr}(\mathbf{R}^2\mathbf{S}) \,, \;\; \lambda_5 = \mathrm{Tr}(\mathbf{S}^2\mathbf{R}^2) \tag{6}$$

The ten $\mathbf{T}^{(i)}$ in Eq. 4 form the basis for symmetric tensors, and the six $\mathbf{B}^{(i)}$ in Eq. 5 for skew-symmetric tensors; λ_i in Eq. 6 is the basis of scalar invariants required to determine the necessary coefficients. Since incompressibility gives $\mathrm{Tr}(\mathbf{S}) = 0$, it is easy to show $\mathrm{Tr}(\mathbf{T}^{(i)}) = 0$.

2.3 Non-Dimensionalisation and Reynolds Number Dependence

Before using the tensor framework with neural networks, it is important to non-dimensionalise all quantities, since it would facilitate efficient learning[10]. Additionally,

non-dimensionalisation will also allow us to appropriately introduce Re as a parameter, whereby the model system can be run at any chosen Re to obtain desired (non-dimensional) statistics of velocity gradients[19].

Since velocity gradients characterise small scales, a natural choice for non-dimensionalisation is to utilise the Kolmogorov time and length scales given as

$$\tau_K = (\nu/\langle\epsilon\rangle)^{1/2}\,, \qquad \eta_K = (\nu^3/\langle\epsilon\rangle)^{1/4}. \tag{7}$$

Here, $\epsilon = 2\nu S_{ij}S_{ij}$ is the energy dissipation rate and $\langle\cdot\rangle$ denotes averaging over space and time. In homogeneous turbulence, we have $\langle S_{ij}S_{ij}\rangle = \langle R_{ij}R_{ij}\rangle = \langle A_{ij}A_{ij}\rangle/2$. Thus, $\langle\epsilon\rangle = \nu\langle A_{ij}A_{ij}\rangle$, giving $\langle A_{ij}A_{ij}\rangle\tau_K^2 = 1$, implying that $1/\tau_K$ quantifies the root-mean-square amplitude of $\mathbf{A}$, which also allows us to impose constraints on running the model system.

We use the following non-dimensionalisation: $t^* = t/\tau_K$, $\mathbf{x}^* = \mathbf{x}/\eta_K$ (i.e., $\nabla^* = \eta_K\nabla$), $\mathbf{A}^* = \mathbf{A}\tau_K$, $\mathbf{H}^* = \mathbf{H}\tau_K^2$ (and $\mathbf{H}_{\rm d}^* = \mathbf{H}_{\rm d}\tau_K^2$), and then obtain the following equation for $\mathbf{A}^*$:

$$\frac{D\mathbf{A}^*}{Dt^*} = -(\mathbf{A}^{*2} - \frac{1}{3}\,\mathrm{Tr}(\mathbf{A}^{*2})\mathbf{I}) - \mathbf{H}_{\rm d}^* + \nabla^{*2}\mathbf{A}^*\,. \tag{8}$$

where we have isolated the deviatoric part of the pressure Hessian: $\mathbf{H}_{\rm d} \equiv \mathbf{H} - \frac{1}{3}\,\mathrm{Tr}(\mathbf{H})\mathbf{I}$ which is trace-free (thus facilitating the use of symmetric trace-free basis in Eq 4).

The terms $\mathbf{H}_{\rm d}^*$ and $\nabla^{*2}\mathbf{A}^*$ can now be modelled in terms of $\mathbf{A}^*$ using the tensor bases, appropriately replaced by their non-dimensional counterparts, i.e., $\mathbf{T}^{*(i)}$ and $\mathbf{B}^{*(i)}$, and the coefficients $c_1^{(i)}$, $c_2^{(i)}$, $c_3^{(i)}$ are dimensionless. It can be seen immediately that the above system does not have any Reynolds number dependence. This is not surprising since we utilised Kolmogorov scales for non-dimensionalisation, temporarily choosing to ignore intermittency. Thus, the Reynolds number dependence has to be reintroduced by hand (and validated *a posteriori*). Previous modelling attempts do not recognise this aspect and have consequently not captured the Reynolds number dependence; see e.g., Refs. 17,20,21.

We devise the following pragmatic way to introduce the Reynolds number dependence and intermittency effects. We maintain the non-dimensionalisation by Kolmogorov variables and rescale the tensor bases as

$$\mathbf{H}_{\rm d}^* = \sum_{i=1}^{10} c_1^{(i)} R_\lambda^{-\beta_1^{(i)}} \mathbf{T}^{*(i)}\,, \tag{9}$$

$$\nabla^{*2}\mathbf{A}^* = \sum_{i=1}^{10} c_2^{(i)} R_\lambda^{-\beta_3^{(i)}} \mathbf{T}^{*(i)} + \sum_{i=1}^{6} c_3^{(i)} R_\lambda^{-\beta_3^{(i)}} \mathbf{B}^{*(i)}\,. \tag{10}$$

Here, R_λ is the Reynolds number based on Taylor length scale (note that $R_\lambda \sim Re^{1/2}$) and the exponents $\beta_1^{(i)}$, $\beta_2^{(i)}$ and $\beta_3^{(i)}$ are additional model parameters which will be determined. This choice is motivated by two main reasons. Firstly, the well-known multifractal description of turbulence suggests that velocity gradient statistics scale as power laws (or combinations of power laws) in Reynolds[1]. Secondly, the tensors in the bases span various orders of $\mathbf{A}$, all of which feel the intermittency effects differently. Thus, the Reynolds number factors can rescale them to the same order, allowing for more efficient learning, essentially acting as additional physics-informed constraints to accommodate intermittency.

3 ReS-TBNN: Reynolds-Number-Scaled Tensor-Based Neural Network

We now consider the neural network architecture utilised to model the unclosed terms. The tensor-based neural network (TBNN), utilising only the symmetric basis from Eq. 4, was first proposed by Ling et al.[22] for turbulence modelling of the Reynolds stress tensor. More recently, it was extended to modelling the pressure Hessian in[21], but without taking Reynolds number into consideration. Unlike traditional neural networks, TBNN utilises two input layers. The architecture to model the pressure Hessian is shown in Fig. 1. The first input layer uses the scalar basis λ_i, which are then fed forward to multiple hidden layers to obtain the scalar coefficients $c_1^{(i)}$ in the first output layer. The essence of this step is to model the scalar coefficients as strongly nonlinear functions of the scalar basis – an exercise traditionally performed via "human learning"[17]. This is precisely the step where deep neural networks are advantageous. The second input layer uses the rescaled tensor basis as input, for instance $R_\lambda^{-\beta_1^{(i)}}\mathbf{T}^{(i)}$, for the pressure Hessian. This second input layer is contracted with the first output layer to obtain the predicted pressure Hessian tensor in the final output layer, in accordance with Eq. 9. We reiterate that only the deviatoric part of the pressure Hessian needs to be modelled. The architecture for the viscous Laplacian is essentially identical to that shown in Fig. 1, with the difference that the first output layer and the second input layer have both 16 nodes, corresponding to the coefficients $c_2^{(i)}$

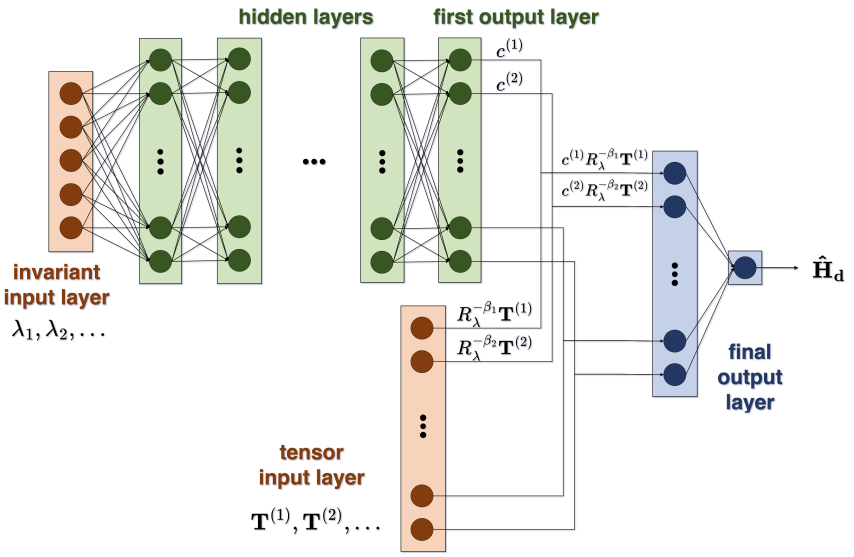


Figure 1. Reynolds number scaled tensor based neural network (ReS-TBNN) architecture utilised for modelling the deviatoric pressure Hessian, based on Eq. 9. The first output layer and the second input layer both have 10 nodes. Note that the exponents $\beta^{(i)}$s are not fixed inputs, but obtained from network training using standard backpropagation. A similar network is utilised for the viscous Laplacian, utilising both the symmetric and skew-symmetric tensor bases, as mentioned in Eq. 10. The first output layer and second input layer have both 16 nodes in this case.

and $c_3^{(i)}$ and the tensors $\mathbf{T}^{(i)}$ and $\mathbf{B}^{(i)}$, with appropriate pre factors corresponding to the Reynolds number scaling.

3.1 DNS Data (Ground Truth)

To train the ReS-TBNN model, the "ground truth" data are obtained from a massive DNS database corresponding to forced stationary isotropic turbulence in a periodic domain[6]. The simulations were performed using Fourier pseudo-spectral methods, allowing us to obtain the data with the highest accuracy practicable. A key aspect of our data is that we have simultaneously achieved a wide range of Reynolds numbers and the necessary small-scale resolution to accurately resolve extreme events[6,7]. Both of these conditions are indispensable for successful model development. The Taylor-scale based Reynolds number R_λ of our database ranges from $140 - 1300$. The data have been utilised and validated in several recent studies[6,16,23,24], where one can also find more detailed account of the DNS methods and database. In order to train our network, only the data for $R_\lambda = 140 - 650$ are utilised; subsequently, we will demonstrate that the trained network can successfully predict statistics at higher (unseen) $R_\lambda = 1300$. Though this R_λ is only twice as large as the largest one used in the training, its usefulness should be assessed in the context of the computational expense of DNS, which would be easily 100 times more. This is because the cost of DNS increases at least as strongly as R_λ^6, going up to R_λ^8 in the limit of large R_λ, to accurately resolve the smallest scales[5,7].

3.2 Training of the ReS-TBNN Model

The training of the ReS-TBNN model is implemented in FORTRAN using a massively parallel in-house deep-learning library. To update the parameters of the neural network, the quadratic loss function is minimised using the standard backpropagation algorithm[10]. For example, for the pressure Hessian tensor, the loss function is given by

$$\mathcal{L} = \frac{1}{2N_{\text{data}}} \sum_{m=1}^{N_{\text{data}}} ||\hat{\mathbf{H}}_{\text{d}}^{(m)} - \mathbf{H}_{\text{d}}^{(m)}||_F^2. \tag{11}$$

where $\hat{\mathbf{H}}_d$ is the model output and $||\cdot||_F$ denotes the Frobenius norm. A similar loss function can also be written for the viscous Laplacian term. The network weights and biases, as well as the exponents $\beta^{(i)}$ in Eqs. 9-10, are simply updated using gradient descent: $x = x - \alpha(\partial\mathcal{L}/\partial x)$, where x is the variable being updated and α is the learning rate. For precise details pertaining to training parameters and behaviour of loss function, we refer the reader to the recent work[19].

4 Comparison of the ReS-TBNN Model with DNS

The effectiveness of the trained ReS-TBNN model will now be evaluated by comparing its outcome with DNS results. We first focus on the Reynolds number trend of velocity gradient statistics (this being a key contribution of the model). We particularly consider the probability density functions (PDFs) that display increasingly non-Gaussian tails with

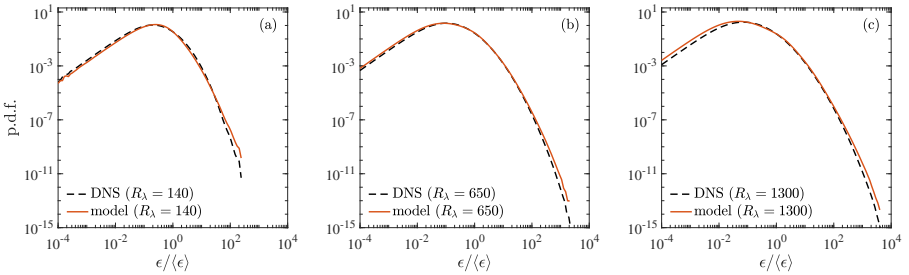


Figure 2. Comparisons of probability density functions (PDFs) of the energy dissipation rate, non-dimensionalised by the mean value, as obtained from network model and DNS. Panels a-c show the comparison at $R_\lambda = 140$, 650 and 1300, respectively, on log-log scales. We reiterate that $R_\lambda = 1300$ is never seen by the model. For a more comprehensive comparison, see Ref. 19.

increasing Reynolds number, because of intermittency. All components of $\mathbf{A}$ exhibit intermittency, but it is convenient to consider scalar quantities of direct physical significance, such as the energy dissipation rate, whose mean value is the net energy flux from large to small scales. As is well known, the instantaneous energy transfers are highly intermittent, leading to extreme dissipation events.

Fig. 2 shows comparisons of the PDFs of the energy dissipation rate, normalised by its mean value, from DNS and the model. Panels a-c illustrate the comparison on log-log scales at $R_\lambda = 140$, 650 and 1300, respectively, showing excellent agreement between the two results. The ReS-TBNN model has been trained only up to $R_\lambda = 650$ and has not seen any data for $R_\lambda = 1300$. For a closer inspection, panels d-e show the same comparisons on linear-log scales for all R_λ available. The model captures the intermittent tails qualitatively well, though the extreme events are overpredicted (see below). We believe that this over-prediction occurs because of the Reynolds number scaling of the tensor bases; essentially, the rescaling serves to normalise the tensor and the extreme events have slightly stronger influence on the weights and biases. Note that, similar to the dissipation rate, one can also consider other scalar measures derived from $\mathbf{A}$, such as enstrophy $\Omega = \omega_i\omega_i$, where $\omega_i = \epsilon_{ijk}A_{jk}$ is the vorticity vector (with ϵ_{ijk} being the Levi-Civita symbol). Although not shown here, the agreement observed for enstrophy is similar.

The over prediction by the model occurs principally for events with probability less than about 10^{-9}. Such events are obviously very important for high order moments, but the reliability of such very large moments is not quite assured for the DNS data itself. For example, suppose we compute the sixth moment of the energy dissipation rate. This is equivalent to obtaining the 12-th order moment of velocity gradients, which would be stretching one's credulity even for the large size of the present database. To better understand how well the model and the DNS agree, it is more useful to directly compare some moments from the PDFs. For this comparison and accompanying discussion, we refer the reader to recently published work[19].

In addition to the PDFs and moments, it is also important to capture the structure of the velocity gradient tensor. We examine a well known universal results to this end, the alignment of the vorticity vector with the eigenvectors of strain tensor, shown in Fig. 3. Panel a shows the PDFs of the cosines of the alignment from DNS. Consistent with the well known result from the literature[15], the vorticity preferentially aligns with the second

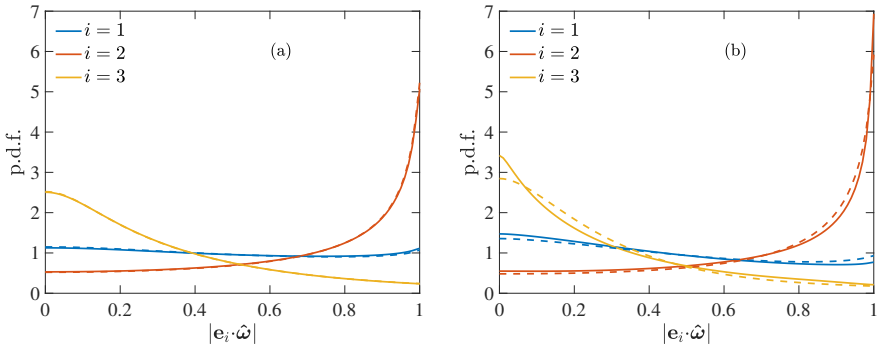


Figure 3. Comparison of the PDFs of the cosine of the angles between the vorticity unit vector $\hat{\boldsymbol{\omega}}$ and the eigenvectors of the strain tensor $\mathbf{e}_i$, corresponding to eigenvalues λ_i, where $\lambda_1 \geq \lambda_2 \geq \lambda_3$. Panel a shows the result from DNS corresponding to $R_\lambda = 1300$ in solid lines and $R_\lambda = 140$ in dashed lines. Panel b shows the result from the network model corresponding to same R_λ values in solid and dashed lines. The alignment PDFs exhibit no R_λ dependence in DNS; the model shows similarly negligible dependence.

eigenvector of strain, and is weakly orthogonal to third eigenvector; whereas there is no preferential alignment with the first eigenvector. There is virtually no R_λ-dependence of these PDFs (as noted in Ref. 16). In Fig. 3b the corresponding result is shown from the ReS-TBNN model. The model captures the trends very well, with slight enhancement of the respective alignments. This trend is consistent with the result in Fig. 2 where the model slightly overpredicts extreme events (note that the alignments are enhanced when considering extreme events[16]). We also note that the model shows only very weak Reynolds number dependence of the alignment properties, which is inconsequential for all practical purposes.

5 Concluding Remarks

In fluid dynamics and turbulence, the DNS of Navier-Stokes equations on massive supercomputers is now an established area for gaining a fuller understanding of flow physics, leading to more reliable predictions. However, both theoretical and practical needs demand ever-increasing size of computations, so that fluid turbulence will always remain, for the foreseeable future, as one of the frontier computational problems, no matter how large the supercomputers become.

In this regard, the major bottleneck is the need to simulate small scales of turbulence with high fidelity. To make progress on real problems, one needs to model small scales well, for instance in large-eddy simulations (LES) where large scales are resolved, but small scales are modelled assuming a degree of universality. This modelling approach has been largely guided by "human learning", often resulting in *ad hoc* considerations depending on the flow. As modern methods of deep machine learning have expanded, it appears possible for them to aid in modelling by directly learning from vast amount of high-fidelity data that is already available over some range of Reynolds numbers. In this scenario, deep neural networks are allowed to do the fitting at a deeper level of instantaneous data, in the process satisfying a substantially larger set of constraints than possible by "human learning".

In this paper, we have made an attempt towards our stated goal. We have demonstrated that the small scale dynamics of turbulence, as captured by velocity gradients, can be modelled reasonably well using deep neural networks. The deep neural networks are set up to functionally model the nonlocal pressure and viscous contributions to velocity gradient dynamics. The networks are then trained on a range of Reynolds numbers available from DNS, and the training is leveraged to predict results at higher Reynolds numbers, whose properties the network does not know in advance. The effort is very encouraging not only in predicting the intermittency of velocity gradients with increasing Reynolds number, but also various signature topological properties of the velocity gradient tensor, such as alignment of the vorticity with strain rate eigenvectors, and the joint PDFs of invariants displaying a tear-drop shape. Overall, the modelling effort developed here provides a substantial improvement upon the prior work, especially with respect to the robustness of the local functional modelling of pressure and viscous terms across a range of Reynolds numbers.

There are certain shortcomings of the trained model when considering truly extreme events; fortunately, they contribute significantly only to very high order moments. Nevertheless, it should be possible to further improve this aspect by incorporating the current deep learning approach in alternative frameworks for velocity gradient dynamics, which incorporate Reynolds number dependencies more naturally. Finally, it would also be worth expanding the current effort in a more concerted way to other modelling paradigms such as LES[25] – allowing one to tackle more complex turbulent flows at Reynolds numbers of practical interest in nature and engineering. Such an extension can be accomplished, for instance, by considering filtered velocity gradient tensors, which would be amenable to the same tensor framework as utilised here[22,21]. Likewise, the framework developed here can also be extended to study the dynamics of scalars in turbulent mixing problems especially in the high Schmidt number regime[26].

Acknowledgements

We gratefully acknowledge the Gauss Centre for Supercomputing e.V. (`www.gauss-centre.eu`) for providing computing time on the supercomputers JUQUEEN and JUWELS at Jülich Supercomputing Centre (JSC), where the simulations and analyses reported in this paper were primarily performed.

References

1. U. Frisch, *Turbulence: the legacy of Kolmogorov*, Cambridge University Press, Cambridge, 1995.
2. A. N. Kolmogorov, *The local structure of turbulence in an incompressible fluid for very large Reynolds numbers*, Dokl. Akad. Nauk. SSSR, **30**, 299-303, 1941.
3. J. M. Wallace, *Twenty years of experimental and direct numerical simulation access to the velocity gradient tensor: What have we learned about turbulence?*, Phys. Fluids, **21**, 021301, 2009.
4. P. Moin and K. Mahesh, *Direct numerical simulation: a tool in turbulence research*, Annu. Rev. Fluid Mech., **30**, 539-578, 1998.
5. V. Yakhot and K. R. Sreenivasan, *Anomalous scaling of structure functions and dynamic constraints on turbulence simulation*, J. Stat. Phys., **121**, 823-841, 2005.

6. D. Buaria, A. Pumir, E. Bodenschatz, and P. K. Yeung, *Extreme velocity gradients in turbulent flows*, New J. Phys., **21**, 043004, 2019.
7. D. Buaria and A. Pumir, *Vorticity-strain rate dynamics and the smallest scales of turbulence*, Phys. Rev. Lett., **128**, 094501, 2022.
8. Y. LeCun, Y. Bengio, and G. Hinton, *Deep learning*, Nature, **521**, 436-444, 2015.
9. K. Duraisamy, G. Iaccarino, and H. Xiao, *Turbulence modeling in the age of data*, Annu. Rev. Fluid Mech., **51**, 357-377, 2019.
10. I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*, MIT Press, 2016.
11. H. Kim, J. Kim, S. Won, and C. Lee, *Unsupervised deep learning for super-resolution reconstruction of turbulence*, J. Fluid Mech., **910**, A29, 2021.
12. B. W. Zeff, D. D. Lanterman, R. McAllister, R. Roy, E. H. Kostelich, and D. P. Lathrop, *Measuring intense rotation and dissipation in turbulent flows*, Nature, **421**, 146-149, 2003.
13. J. Schumacher, J. D. Scheel, D. Krasnov, D. A. Donzis, V. Yakhot, and K. R. Sreenivasan, *Small-scale universality in fluid turbulence*, Proc. Natl. Acad. Sci., **111**, 10961-10965, 2014.
14. R. M. Kerr, *Higher-order derivative correlations and the alignment of small-scale structures in isotropic numerical turbulence*, J. Fluid Mech., **153**, 31-58, 1985.
15. W. T. Ashurst, A. R. Kerstein, R. M. Kerr, and C. H. Gibson, *Alignment of vorticity and scalar gradient with strain rate in simulated Navier-Stokes turbulence*, Phys. Fluids, **30**, 2343-2353, 1987.
16. D. Buaria, E. Bodenschatz, and A. Pumir, *Vortex stretching and enstrophy production in high Reynolds number turbulence*, Phys. Rev. Fluids, **5**, 104602, 2020.
17. C. Meneveau, *Lagrangian dynamics and models of the velocity gradient tensor in turbulent flows*, Annu. Rev. Fluid Mech., **43**, 219-245, 2011.
18. G. F. Smith, *On isotropic functions of symmetric tensors, skew-symmetric tensors and vectors*, International Journal of Engineering Science, **9**, 899-916, 1971.
19. D. Buaria and K. R. Sreenivasan, *Forecasting small-scale dynamics of fluid turbulence using deep neural networks*, Proc. Natl. Acad. Sci., **120**, e2305765120, 2023.
20. P. L. Johnson and C. Meneveau, *A closure for Lagrangian velocity gradient evolution in turbulence using recent-deformation mapping of initially Gaussian fields*, J. Fluid Mech., **804**, 387-419, 2016.
21. Y. Tian, D. Livescu, and M. Chertkov, *Physics-informed machine learning of the Lagrangian dynamics of velocity gradient tensor*, Phys. Rev. Fluids, **6**, no. 9, 094607, 2021.
22. J. Ling, A. Kurzawski, and J. Templeton, *Reynolds averaged turbulence modelling using deep neural networks with embedded invariance*, J. Fluid Mech., **807**, 155-166, 2016.
23. D. Buaria and K. R. Sreenivasan, *Intermittency of turbulent velocity and scalar fields using three-dimensional local averaging*, Phys. Rev. Fluids, **7**, L072601, 2022.
24. D. Buaria and K. R. Sreenivasan, *Scaling of acceleration statistics in high Reynolds number turbulence*, Phys. Rev. Lett., **128**, 234502, 2022.
25. A. Beck, D. Flad, and C.-D. Munz, *Deep neural networks for data-driven LES closure models*, J. Comput. Phys., **398**, 108910, 2019.
26. D. Buaria, M. P. Clay, K. R. Sreenivasan, and P. K. Yeung, *Turbulence is an ineffective mixer when Schmidt numbers are large*, Phys. Rev. Lett., **126**, 074501, 2021.

Plasma Physics and Charged Particle Dynamics

Numerical Simulations of Plasmas in Support of Facility Development

Maria Elena Innocenti

Institute for Theoretical Physics, Ruhr University Bochum,
Universitätsstraße 150, 44801 Bochum, Germany
E-mail: mariaelena.innocenti@rub.de

Numerical simulations of plasmas are crucial to the design and development of techniques and facilities aimed at furthering our understanding of the Universe and of the interaction between lasers and plasmas. To do so effectively, the large scale separation that characterises plasmas has to be overcome, both using new computational techniques and HPC resources. Here, we will see two examples of such cutting-edge simulations.

Numerical simulations constitute an irreplaceable support to the design and development of techniques and facilities aimed at furthering our understanding of the Universe and of the interaction between lasers and plasmas. These two contributions highlight how innovative numerical techniques and HPC resources such as the ones provided at the Jülich Supercomputing Centre have to be leveraged together to simulate new facilities and scenarios.

Plasma accelerators

Acceleration of particles to the extreme energies (TeV) required to drive discoveries in particle physics can be achieved with extreme large ($\sim$ hundreds of kilometres) linear colliders. Plasma accelerators are considered a promising technology to deliver extreme high energy particles using more compact facilities delivering higher acceleration gradients with respect to linear colliders. Designing an effective plasma accelerator requires to be able to accurately model the interaction of a driver beam (either a charged-particle or a laser beam) with the electrons and ions in a plasma, overcoming the large scale separation between system scales (kilometres) and the transverse size of the colliding beam (10s of nanometres). Thévenet *et al.* describe how to do so using different descriptions for beam and plasma macro-particles, adaptive mesh refinement and a code optimised for GPU clusters.

Free Electron Lasers

Coherent light amplification and laser technology have transformed fields such as material science, medicine and industry, which all rely on the production of stable, coherent and controlled light sources. Free Electron Lasers (FELs) use beams of electrons travelling at relativistic speeds through magnetic structures called undulators to generate coherent light; seeding techniques can be used for improving the coherence and quality of the beams. Niknejadi and Schaper highlight the importance of end-to-end simulations of FEL facilities such as the FLASH facility at DESY, which is being upgraded and will soon feature two independently tuneable FEL lines. They show how their simulations are instrumental to the improvement of coherence and stability of seeded FELs, and how the incorporating of synthetic datasets has lowered computational demands and facilitated parameter scans.

Optimising Coherence and Stability in Seeded Free Electron Lasers through Synthetic Simulation Datasets

Pardis Niknejadi and Lucas Schaper

Deutsches Elektronen-Synchrotron DESY, 22607 Hamburg, Germany
E-mail: pardis.niknejadi@desy.de

Improvements in coherence, stability, and control of Free Electron Lasers (FELs) have advanced research and development in both FEL and computational approaches. To improve coherence, methods such as external seeding of FELs are explored where external laser pulses are leveraged to initiate seeded pre-bunching for coherent emission and amplification. Externally seeded approaches, combined with finely tuned control mechanisms, enable FELs to achieve high coherence, precision, and stability, meeting the advanced demands of the photon user community. To support these initiatives, we first performed comprehensive start-to-end (S2E) simulations of the linac-based FEL FLASH1, which is on track to upgrade to a seeded FEL. Next, we generated synthetic datasets requiring fewer computational resources, which allowed us to explore a broader parameter space efficiently. This approach enabled us to perform multidimensional optimisation of seeded FEL, identifying stable operational points in the presence of machine jitter, which was made possible by the high-performance computing resources of JUWELS. We expect the extension of this work to enhance FEL performance and coherence at the upgraded facility.

1 Introduction

In just over sixty years, the development of coherent light amplification and laser technology has transformed fields like material science, medicine, and industry. The world's first laser was constructed using a spiral flashlamp surrounding a ruby rod, approximately 1 cm in diameter, built by Theodore Maiman in 1960[1]. This design was based on earlier groundbreaking work by Townes and Schawlow[2]. In this laser, the spiral flashlamp provides an intense burst of light to excite the atoms in the ruby rod, which is the laser medium. When these atoms absorb energy, they enter an "excited" state and later emit photons in a process known as stimulated emission. The laser requires a mechanism to amplify and direct the energy for the emitted photons to form a coherent beam. This amplification was achieved through the use of a laser cavity. The flashlamp can only provide short bursts of light, limiting the laser's operational efficiency. In addition, factors like the gain medium's inability to support population inversion at high photon energies and limitations of optical cavities – composed of mirrors that reflect and amplify light – prevent these lasers from operating at the shorter-wavelength end of the spectrum, such as in the vacuum ultraviolet (VUV) and X-ray regions.

A few years later, it was shown that lasers could generate ultrashort pulses by locking the phases of different light modes in the cavity, yielding a single intense pulse[3]. This advancement facilitated applications requiring short, high-energy light pulses, like precision spectroscopy. Another critical development was Chirped Pulse Amplification (CPA), which amplifies ultrashort pulses to high intensities without damaging the laser medium by stretching the pulse before amplification and then compressing it back[4]. This technique has led to significant increase in laser intensities, supporting applications such as laser-plasma

accelerators. Additionally, High Harmonic Generation (HHG) enabled the conversion of laser light into high-frequency harmonics, advancing laser applications into the extreme ultraviolet and soft X-ray regions, crucial for attosecond science to observe rapid processes like electron dynamics in atoms[5].

Another paradigm shift was the successful demonstration of Free Electron Lasers (FELs) in 1977[6], which operated without a traditional lasing medium. FELs use beams of electrons travelling at relativistic speeds through magnetic structures called undulators to generate coherent light. Therefore, various problems induced by conventional gain media (e.g. thermal and damage issues or accessible wavelength) can be overcome in an FEL, resulting in a drastically expanded wavelength range and producing high-energy, tunable pulses in the X-ray range. Also, some FEL lasing methods/schemes do not require a cavity. One method of generating these pulses is Self-Amplified Spontaneous Emission (SASE)[7], where the spontaneous emission of the electron beam is amplified as it passes through the undulator, producing a highly intense, coherent beam. However, SASE lasers suffer from limitations such as relatively poor temporal coherence due to the random nature of the initial emission, which limits their output applications.

Seeding techniques[8,9] are a significant research and development topic for improving the coherence and quality of FEL beams. Generally speaking, in a seeded FEL, as shown in Fig. 1, external coherent laser pulses are used as "seed" and injected into the FEL, which synchronises the electron beam's emission, producing a more coherent and stable output. In principle, this approach allows finer control over the final FEL output properties, making it ideal for high-precision materials science, chemistry, biology, and many more applications. In practice, optimising the performance of a Free Electron Laser and tailoring the FEL output involves a deep understanding of the physical principles governing its many components – such as the injector, linac, and undulators. Each element is modelled by specialised simulation tools used to predict how they interact and influence other beamline components, especially when considering coupled effects, like how the variation of energy along the longitudinal profile of the beam affects optimisation and photon output. Nonlin-

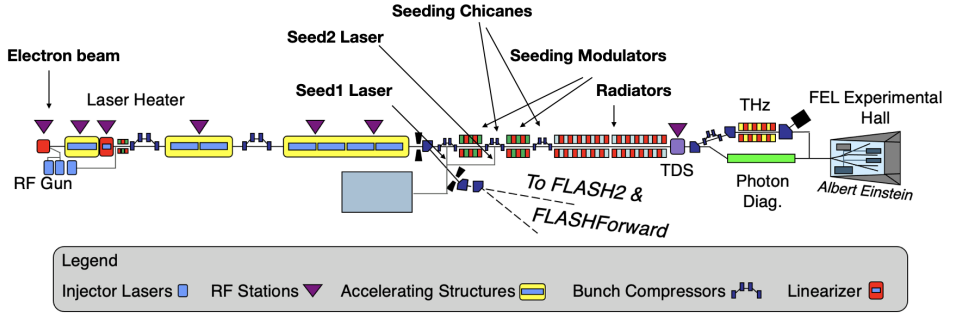


Figure 1. Schematic of Seeded FLASH1: Unlike a SASE FEL (i.e. FLASH2) setup, where the electron beam, after acceleration, passes directly through the radiators, a seeded FEL setup involves several additional components. These include modulators, dispersive sections, and synchronised seed lasers, which initiate seeded pre-bunching for coherent light amplification. S2e simulations for seeded FELs are complex and need careful interfacing between multiple simulation tools to model the beam dynamics and interactions throughout the system.

ear effects also need to be considered; however, thresholds must be established to balance precision and computational complexity. Implementing robust control systems requires advanced algorithms and computational tools to fine-tune FEL performance for various applications.

The following sections address key computational challenges in modelling FEL systems. We introduce the FLASH facility to illustrate the complexity of simulating interactions across various beamline components and emphasise the need for efficient data exchange among simulation tools. Next, we discuss how utilising the JUWELS[10] supercomputer has helped us overcome specific challenges in S2E simulations, leading to more accurate and faster computations. We also compare the computational demands of synthetic datasets with those of S2E simulations and showcase an example of our multi-dimensional optimisation results. Finally, we provide a brief outlook on future research directions to enhance the performance and capabilities of FEL systems further.

2 Computational Modelling

Accelerator-based light sources have been modelled through specialised simulation tools that address beam dynamics from the photoinjector stage to the undulator, as outlined in Ref. 11 for SASE FELs. These simulations capture detailed interactions at each beamline stage. Tools like PARMELA[12] or OPAL[13] simulate the photoinjector, while IMPACT-Z[14] and `elegant`[15] handle linear accelerator (linac) dynamics and compression, respectively, and GENESIS[16] models the FEL section, each focusing on specific aspects of beam dynamics such as wakefields, coherent synchrotron radiation, and phase space evolution. The setup is inherently more complex for seeded FEL systems. As shown in Fig. 1, a seeded FEL beamline includes additional components such as synchronised seed lasers, modulators, and dispersive sections, which are critical for initiating the seeded pre-bunching for coherent emission. Each element – from lasers to modulators, dispersive sections, and radiators – requires precise tuning to achieve optimal performance. This level of detail and interdependence makes S2E simulations for seeded FELs far more intricate than SASE systems.

Simulation tools for accelerator-based light sources often produce outputs in unique formats, making compatibility between different stages challenging. To bridge these gaps, we developed in-house libraries[17] that standardise data transfer, converting formats between tools to ensure seamless integration. Fig. 2a-c show the different stages of seeded FEL simulation. Fig. 2d shows the colour-matched codes for each element of the S2E simulation workflow. Furthermore, many of these codes were originally developed in different computing eras, often transitioning from 32-bit to 64-bit architectures. Benchmarking these tools for efficiency is crucial to minimising computational resources while maintaining accuracy. We prioritise using tools that yield the same insights with reduced resource consumption. Given the resource-intensive nature of these simulations, creating reusable datasets that remain informative for future research has been central to our approach. This strategy enables us to share comprehensive, high-quality data across teams, supporting ongoing advancements while optimising resource use responsibly.

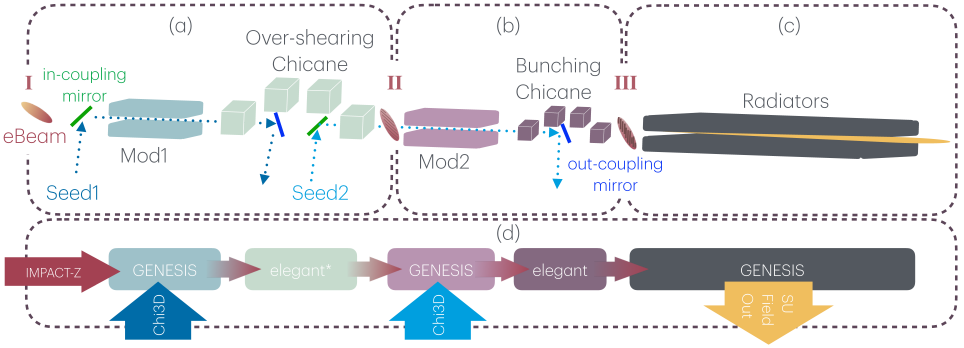


Figure 2. Layout for simulation and optimisation. In the first modulator-chicane stage, an electron beam (eBeam) interacts with Seed1 laser in (Mod1) and then experiences strong dispersion (a). In the second modulator-chicane, the fine bunching of eBeam is optimised (b). The radiation from the seeded pre-bunched beam is amplified in the radiator section (c). The start-to-end (S2E) simulation workflow is shown in (d), where the tool used for each element is shown in a matching colour. Horizontal arrows correspond to eBeam handover, and vertical arrows correspond to field handover.

2.1 FLASH: A High Repetition Rate Seeded FEL

The FLASH Facility at DESY is being upgraded and will soon feature two independently tunable FEL lines – FLASH1 operating in seeded mode and FLASH2 in SASE mode[18]. This would allow the utilisation of various schemes such as High-Gain Harmonic Generation (HGHG) as detailed in Ref. 8 and Echo-Enabled Harmonic Generation (EEHG) as detailed in Ref. 9, covering ultrashort single cycle, long pulse mode, and a spectral range from 60 nm to 4 nm, including the oxygen K-edge. Specifically, FLASH1 will offer high-quality pulses with precise polarisation and coherence control. As a Superconducting Radio Frequency machine, FLASH will be able to deliver long bursts of up to 5000 highly coherent FEL pulses per second at bunch repetition rates of 1 MHz.

2.2 Objectives of Simulation Studies in Seeded FELs

Coherent amplification is the cornerstone of FEL technology, enabling it to generate light that, like conventional lasers, is highly monochromatic (narrow bandwidth centred on a single wavelength), highly directional, and comprised of photons oscillating in unison – making FELs ideal probes with exceptional spectral and temporal resolution. Furthermore, the high repetition rate of FLASH means that probing light pulses with precise spectral and temporal resolution are available in large numbers, enabling high-statistics data collection comparable to that of synchrotron facilities. Thus, the primary goal of our simulations, analyses, and optimisations is to develop a stable, coherent source.

As discussed, the seeded FEL setup leverages external lasers as seeds. For FLASH1, we use seed lasers that employ a variation of CPA involving an optical parametric amplifier (OPA) for laser pulse amplification, providing both broad tunability and high peak powers[19]. The OPA relies on nonlinear optical processes, where a high-power "pump" laser transfers energy to a weaker "signal" laser, simultaneously generating an "idler" beam. This setup results in an Optical Parametric Chirped-Pulse Amplified beam with high-power

pulses in the ultraviolet (UV) spectrum (290-317 nm specifically), delivering Fourier limited pulses of approximately 50 femtoseconds. The seeded FEL then produces even shorter pulses by a factor of $n^{-1/3}$ (where n is the harmonic number)[20,21], achieving higher energies and shorter wavelengths, with $\lambda_{\text{FEL}} \propto \lambda_{\text{seed}}/n$. The system's stability depends on both the seed laser and electron beam parameters. Fig. 2 illustrates the two seeding and final amplification segments in the seeded S2E simulation setup that can be independently tuned and optimised. Specifically, for SASE FELs, only section "c" (the undulators/radiators) requires modelling; for HGHG, both sections "b" (dispersive section and modulator) and "c" are involved; and for EEHG, sections "a", "b", and "c" are essential. These additional segments add complexity to data transfer or "handshaking"; each stage must pass precise beam properties (i.e. beam at the beamline's "I", "II", and "III" positions) to the next simulation step.

2.3 Challenges Addressed by JUWELS Resources

Compared to older HPC architectures, such as JURECA with Intel Xeon CPUs, JUWELS' AMD EPYC CPUs provide higher core density, greater memory bandwidth, and better energy efficiency, making them ideal for large-scale, CPU-based sequential simulations. For instance, mem192 cores are suitable for efficiently handling memory-intensive tasks such as beam conditioning and interfacing between different software, ensuring good input/output performance. Fig. 3 shows such a processed beam at the location "I" of the beamline. The uniform CPU architecture has also helped minimise simulation artefacts

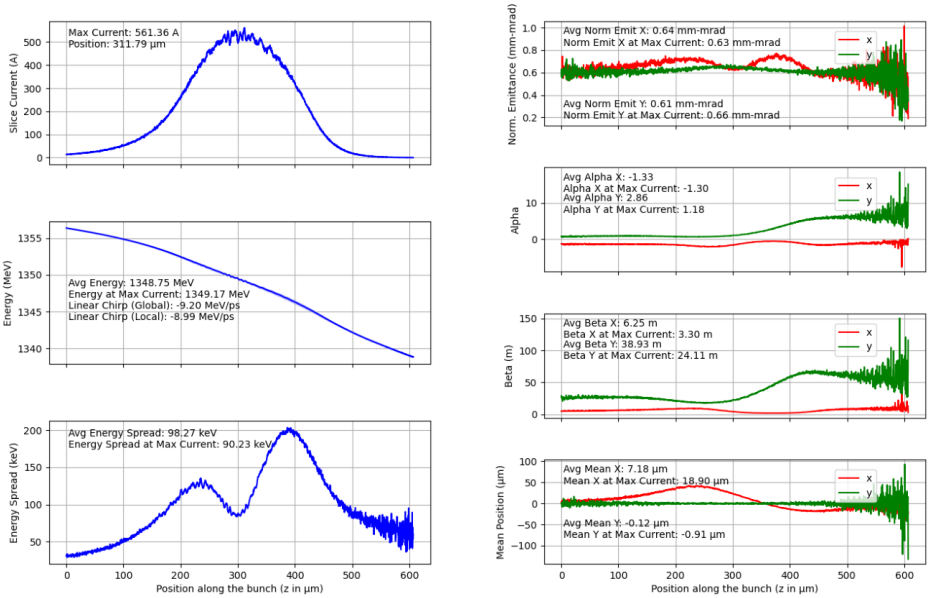


Figure 3. Slice parameters of the beam from the S2E simulation at the "I" location. This beam, generated by IMPACT-Z after multiple optimisation runs in the acceleration and compression sections, is processed as input for GENESIS, which models the interaction within the modulator.

and support high data fidelity. Additionally, the large number of available core hours on JUWELS enables the generation of a comprehensive dataset for benchmarking and ensuring cross-code consistency, which is essential during and after the upgrade of FLASH (i.e. when the experimental campaigns become possible on the machine).

Moreover, the JUWELS Booster GPU module, added in 2020, led to faster optimisation algorithms, making it possible to develop new and faster studies leveraging existing datasets. This approach reuses existing data and accelerates computational workflows to facilitate the creation of better synthetic datasets for future simulations and studies.

3 Simulation and Optimisation Workflow and Stability Assessments

Initially, parameters for seeding sections "a" and "b" (highlighted in Fig. 2) are optimised using a simplified model focused on achieving the best bunching at the target harmonic. However, simplified models are insufficient for precise optimisation and stability checks. For instance, using a beam with Gaussian current and energy spread distribution, a flat energy profile, and uniform beam size and offset for all slices would only reveals the effects of current variation during relative electron and seed laser delay scans, which has minimal implications for the seeding sections. This highlights the need for S2E simulation. In such simulation, electron beam acceleration and compression are modelled in IMPACT-Z, building on previous work including longitudinal space charge and other collective effects, analogous to those discussed in Ref. 11 and 22. Seed lasers are simulated in Chi3D[23] and converted into a format compatible with GENESIS. Both the electron beam and seed laser are then used to simulate the modulator interaction in GENESIS, with the output further converted for compatibility with `elegant` to account for coherent synchrotron radiation effects in dispersive segments (chicanes). Expanding on previous work, we compared the effects of boundary conditions on the bandwidth of seeded pre-bunched beams. This study is submitted for publication[24].

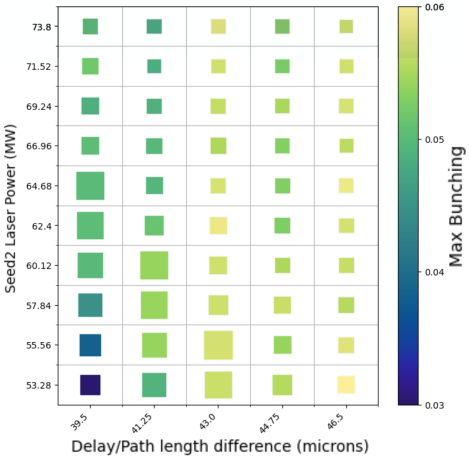


Figure 4. 2D heatmap of bunching and bunching bandwidth, with square size indicating the normalised bunching bandwidth at beamline position "III".

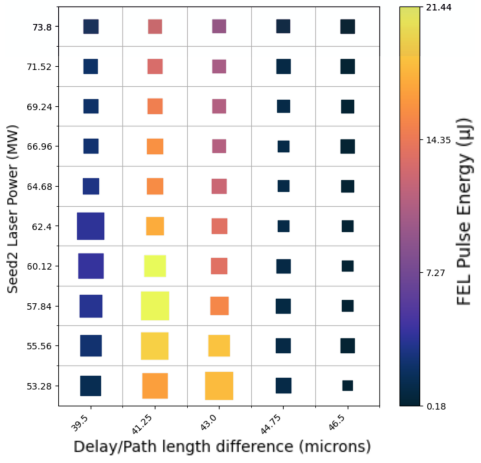


Figure 5. 2D heatmap of FEL pulse energy and length after six radiators for the 4 nm working point. Square size indicates the normalised pulse length.

An effective optimisation strategy requires a multi-objective approach targeting key parameters that drive system performance. The heat maps in Fig. 4 and Fig. 5 are typical visualisation tools to illustrate the outcomes of such studies.

In sections “a” and “b”, the amplitude of energy modulation and the dispersion strength, represented by the electron delay or path difference, are critical for controlling optimal electron bunching at the desired harmonic. Analytical methods from Ref. 25 are then used to predict the seeded pre-bunched beam’s performance in the radiator. However, this serves as an approximation to reduce the scan parameter space, as discussed by the authors in Ref. 26. Finally, iterative tapering is employed to enhance energy extraction in FELs with tapered sections, maximising radiation power.

We learnt that maximising power can affect the spectral purity of the final FEL beam. For extended studies, we turned to synthetic simulations that replicate some or all characteristics of S2E generated beams while offering smoother profiles and reduced noise. S2E simulations typically require close to 2000 CPU cores to track all particles in the beam. With handshaking and checks at each interface step, a full S2E simulation takes at least 5-6 hours and uses 10,000-30,000 core hours. In contrast, synthetic simulations can be run in less than an hour and require fewer cores (needing less than 1,000 core hours per run) due to smoothed-out beam tails and heads. This reduction enabled us to perform multiple multi-objective optimisations, as described in Ref. 22, exploring trade-offs between optimising for short pulse length, high power, or narrow bandwidth while keeping the total carbon footprint of our project low[27].

In Figs. 4 and 5, for example, we observe that bunching and bunching bandwidth is optimised for the 43-micron delay and seed laser power of 62.4 MW. Still, a working point in the amplification stage that yields a shorter FEL pulse is possible with a shorter delay and less laser power. We chose this as a more stable point since the seed laser has less power fluctuation at lower power. Finally, Fig. 6 shows a timing scan that captures essential features of the electron beam from S2E simulation, such as energy variation and position offset along the bunch. For this scan, we start with the seed laser waist overlapping with the electron beam, where the offset is minimal. We then vary the delay between these

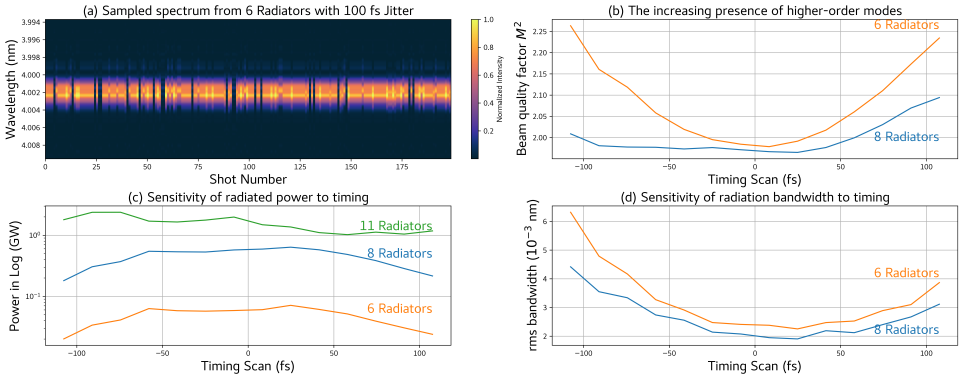


Figure 6. The effect of timing jitter on an optimised working point for an electron beam modelled based on S2E beams. In (a), 100 random samples from 15 timing points in parts (b-d) mimics data taken from the spectrum during commissioning or setup. (b) Impact on quality of the beam: as the quality factor increases, the FEL beam can diverge more rapidly downstream. (c) and (d) illustrate the effect on power and bandwidth.

two beams and observe changes in FEL power, bandwidth, and quality factor. This study is representative of impact of timing jitter in machine. A comprehensive set of similar studies is being prepared for publication.

4 Concluding Remarks

Utilising the high-performance computing resources available at JUWELS, we have enhanced our S2E simulations by incorporating synthetic datasets. This approach has significantly lowered computational demands and has facilitated thorough parameter scans. Our established simulation and optimisation workflow aims to improve the coherence and stability of seeded FELs, focusing on upgrading FLASH1. Our approach allows multi-objective optimisation and investigates trade-offs between tuning for short pulse length, high power, or best spectral purity. This work aims to support operation of FELs and meeting the needs of the photon user community. Future studies will extend the current analyses to enhance the FLASH facility's control software and online optimisation.

Acknowledgements

We are grateful to the seeding simulation team at DESY, especially D. Samoilenko, E. Ferrari, G. Paraskaki, T. Lang, and F. Pannek, for their invaluable work and expertise, which have greatly enriched the development of this research. The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (`www.gauss-centre.eu`) for funding this project by providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS Supercomputer JUWELS at Jülich Supercomputing Centre (JSC).

References

1. T. Maiman, *Stimulated Optical Radiation in Ruby*, Nature **187**, 493-494, 1960.
2. A. L. Schawlow and C. H. Townes, *Infrared and Optical Masers*, Phys. Rev. **112**, 1940-1949, 1958.
3. L. E. Hargrove et al., *Locking of He–Ne Laser Modes Induced by Synchronous Intracavity Modulation*, Appl. Phys. Lett. **5**, 4-5, 1964.
4. D. Strickland and G. Mourou, *Compression of Amplified Chirped Optical Pulses*, Opt. Commun. **56**, 219-221, 1985.
5. A. McPherson et al., *Studies of Multiphoton Production of Vacuum-Ultraviolet Radiation in the Rare Gases*, J. Opt. Soc. Am. B **4**, 595, 1987.
6. D. A. G. Deacon et al., *First Operation of a Free-Electron Laser*, Phys. Rev. Lett. **38**, 892-894, 1977.
7. A. M. Kondratenko and E. L. Saldin, *Generation of Coherent Radiation by a Relativistic Electron Beam in an Undulator*, Part. Accel. **10**, 207-216, 1980.
8. L. H. Yu, *Generation of Intense UV Radiation by Subharmonically Seeded Single-Pass Free-Electron Lasers*, Phys. Rev. A **44**, 5178, 1991.
9. G. Stupakov, *Using the Beam-Echo Effect for Generation of Short-Wavelength Radiation*, Phys. Rev. Lett. **102**, 074801, 2009.

10. D. Alvarez, *JUWELS Cluster and Booster: Exascale Pathfinder with Modular Supercomputing Architecture at Juelich Supercomputing Centre*, J. Large-Scale Res. Facil. **7**, A183, 2021.
11. M. Borland et al., *Start-to-End Simulation of Self-Amplified Spontaneous Emission Free Electron Lasers from the Gun Through the Undulator*, Nucl. Instrum. Methods Phys. Res., Sect. A **483**, 268-272, 2002.
12. L. Young and J. Billen, *The Particle Tracking Code PARMELA*, Proceedings of the 2003 Particle Accelerator Conference, pp. 3521-3523, Los Alamos National Laboratory, Los Alamos, NM, USA, IEEE, 2003.
13. A. Adelmann et al., *The Object Oriented Parallel Accelerator Library (OPAL)*, Proceedings of the Particle Accelerator Conference (PAC 09), paper FR5PFP065, 2010.
14. J. Qiang et al., *An Object-Oriented Parallel Particle-in-Cell Code for Beam Dynamics Simulation in Linear Accelerators*, J. Comput. Phys. **163**, 434-451, 2000.
15. M. Borland, *ELEGANT: A Flexible SDDS-Compliant Code for Accelerator Simulation*, Proceedings of International Computational Accelerator Physics Conference ICAP 2000, Darmstadt, Germany, Sep. 2000.
16. S. Reiche, *GENESIS 1.3: A Fully 3D Time-Dependent FEL Simulation Code*, Nucl. Instrum. **429**, 243-248, 1999.
17. `https://gitlab.desy.de/xseed/geist`[a]
18. L. Schaper et al., *Flexible and Coherent Soft X-ray Pulses at High Repetition Rate: Current Research and Perspectives*, Appl. Sci. **11**, 9729, 2021.
19. I. N. Ross et al., *Optical Parametric Chirped Pulse Amplifiers for the Generation of Extremes in Power, Intensity and Pulse Duration*, Proceedings of Conference on Lasers and Electro-Optics-Europe, in Technical Digest Series, Optica Publishing Group, paper CTuI77, 1998.
20. E. Hemsing, *Minimum Spectral Bandwidth in Echo Seeded Free Electron Lasers*, Front. Phys. **7**, 35, 2019.
21. P. Finetti et al., *Pulse Duration of Seeded Free-Electron Lasers*, Phys. Rev. X **7**, 021043, 2017.
22. R. Bartolini et al., *Multiobjective Genetic Algorithm optimisation of the Beam Dynamics in Linac Drivers for Free Electron Lasers*, Phys. Rev. ST Accel. Beams **15**, 030701, 2012.
23. T. Lang, *Chi3D*, Accessed on: Sep. 22, 2024, `http://www.chi23d.com`.
24. D. Samoilenko et al., *Effects of Boundary Conditions on Coherent Synchrotron Radiation in Echo-Enabled Harmonic Generation*, submitted to Phys. Rev. Accel. Beams, 2024.
25. L. Giannessi, *Seeding and Harmonic Generation in Free-Electron Lasers*, in: Synchrotron Light Sources and Free-Electron Lasers, Springer, 119-147, 2020.
26. A. Mak et al., *Model-Based optimisation of Tapered Free-Electron Lasers*, Phys. Rev. ST Accel. Beams **18**, 040702, 2015.
27. V. Lang et al., *Know your footprint – Evaluation of the Professional Carbon Footprint for Individual Researchers in High Energy Physics and Related Fields*, 2024, arXiv:2403.03308[b].

[a] For access, please contact the corresponding author.

[b] Submitted to Nature Partner Journals Climate Action.

Small-Scale Particle Accelerators for Large-Scale Science thanks to High-Performance Computing

Maxence Thévenet[1], Severin Diederichs[2,1], Axel Huebl[3], and Alexander Sinn[1]

[1] Deutsches Elektronen-Synchrotron DESY, Notkestraße 85, 22607 Hamburg, Germany
E-mail: maxence.thevenet@desy.de, alexander.sinn@desy.de

[2] CERN, Esplanade des Particules 1, 1211 Geneva, Switzerland
E-mail: severin.diederichs@cern.ch

[3] Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720, USA
E-mail: axelhuebl@lbl.gov

Particle colliders, essential for advancing our understanding of the universe, are among the largest research facilities, often spanning many kilometres. Plasma acceleration – a cutting-edge technique that harnesses the immense electric and magnetic fields within plasma – offers a promising solution to reduce both the cost and size of these large-scale facilities. Recent advancements in the open-source, GPU-capable code HiPACE++ have enabled the exploration of this highly demanding plasma acceleration regime on the JUWELS Booster supercomputer. Notably, it was found that accelerating flat beams, commonly used in particle colliders, presents new and unique challenges.

1 Introduction

Particle accelerators have long been essential to scientific progress, driving discoveries in particle physics that lay the foundation of our understanding of matter. Despite many open questions in the Standard Model of particle physics, such as the nature of dark matter, dark energy, and the asymmetry between matter and antimatter, no major breakthrough has been achieved since the discovery of the Higgs Boson in 2012[1,2]. Two complementary approaches are actively discussed in the particle physics community to explore the unsolved mysteries of our universe: a discovery collider with a centre-of-mass per parton energy of 10 Tera-electronVolt (TeV)[3]; or the precise measurement of the properties of the Higgs Boson with a so-called Higgs factory, which has been considered highest priority by the European Strategy for Particle Physics[4]. The construction and operation costs associated with a linear collider in this regime using common radio-frequency-based accelerator technology is enormous, even for a world-wide collaboration effort. This is because **the accelerating gradient with these technologies is limited to ~100 MeV/m, such that TeV-scale energies requires acceleration over a distance on the order of, or surpassing, 100 km**.

Plasma accelerators[5] are a promising technology to make these facilities orders of magnitude more compact and more affordable due to their ultra-high accelerating gradients of 1-100 GeV/m. Plasma is the fourth state of matter, in which the light electrons are separated from their heavy ions. In the so-called bubble regime of a plasma accelerator, a drive beam (either a charged-particle or laser beam) excites a strong plasma wake, by pushing away the light electrons. The remaining ions pull back the electrons, such that the electrons return to the propagation axis at some distance behind the driver, giving the

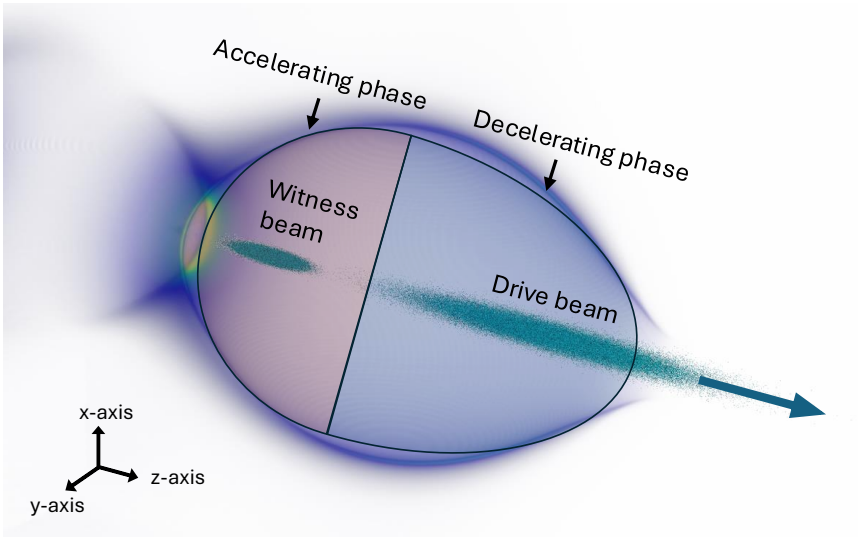


Figure 1. Illustration of a plasma accelerator. A drive beam (here: a particle driver, cloud of blue dots, on the right) propagates near the speed of light to the right through a plasma and excites a bubble-shaped plasma wake. In the plasma wake, strong electromagnetic fields decelerate the drive beam (blue) and accelerate the witness beam (fields in red, witness beam as cloud of blue dots) at the back of the plasma bubble. The arrow shows the propagation direction, assumed to be z by convention.

wake a bubble-like shape. In this plasma wake, extreme electromagnetic fields decelerate the drive beam and accelerate the witness beam, as illustrated in Fig. 1. The bubble size is typically smaller than a millimetre in all directions, and the beams can be even much thinner than this in the transverse (x and y) directions, as is the case on Fig. 1.

Recent designs propose plasma-based colliders both for Higgs factories[6,7] and for the 10-TeV-range discovery machines with length below $10\,\mathrm{km}$[8]. To overcome many technological challenges, a mature design of such a collider requires realistic simulations. However, already **modelling a plasma-based collider at this energy is extremely challenging due to the immense disparity of length scales**: while the full accelerator has a length on the order of kilometres, the transverse beam size of the colliding beams is only on the order of 10s of nanometres, an astounding difference of 11 orders of magnitude!

In this article, we show that by separating the time and length scales, and adding a novel mesh refinement algorithm, **we can drastically advance the state-of-the-art of full 3D simulations of plasma-accelerators for collider applications, using the computing resources of the JUWELS Booster provided by the Forschungszentrum Jülich**. These improvements enabled us to perform realistic and fully converged simulations of most challenging scenarios and unveil new effects for the acceleration of flat beams.

2 Numerical Methods

Depending on temperature and size, a plasma can exhibit very different behaviours. In the Sun, for example, the plasma can be described as a *fluid*. This is not the case for plasma acceleration, where the driver triggers a violent plasma response where, in princi-

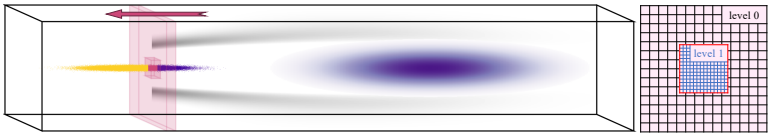


Figure 2. Illustration of one HiPACE++ time step. At the beginning, only the driver (dark purple colourmap on the right of the 3D domain, in this case a laser driver) and the accelerated beam (cloud of dots on the left of the 3D domain) at time step n are known. The beams and the simulation window propagate towards the right near the speed of light. A 2D slab of plasma containing both electromagnetic field and plasma macro-particles is initialised on the rightmost longitudinal cell (in the ζ direction), or *slice*, of the domain, and advanced slice by slice to the left to compute the plasma response (grey-scale colourmap) from the head to the tail of the domain. The current slice being computed is shown in pink, beam quantities at time step n are shown in yellow, those already pushed to time step $n+1$ are shown in dark purple. The beam quantities in the current slice are advanced from n to $n+1$ during the swipe. Mesh refinement of the slice being computed can be used to resolve fine structures around the beam: this is illustrated in the zoom on the right (ghost cells of level 1 are shown in red).

ple, the position and momentum of each particle must be resolved to capture the dynamics. **Particle-in-cell (PIC) is a common and affordable method for this *kinetic* dynamics, where a grid of a 3D domain is used to describe the electric and magnetic fields – the 3 components of each field, and other quantities appearing in Maxwell's equations, are stored in every grid cell – and a collection of macro-particles – each representing a set of physical particles, moving freely in the simulation domain, to describe the plasma dynamics**. A time iteration of a standard PIC loop contains four steps to advance self-consistently the fields on the grid and the plasma macro-particles: (i) the electric and magnetic fields on the grid are advanced by one time step; (ii) macro-particles gather the fields from their neighbouring cells; (iii) macro-particles are advanced by 1 time step; (iv) they deposit their charge and current densities on the neighbouring cells.

In a plasma accelerator in particular, the beams can most of the time be assumed to evolve on a much slower time scale than the plasma response. This assumption, called the quasi-static approximation, allows for a treatment of separate time scales for beams and plasma and is used in HiPACE++[9,10]. **By enabling a different handling of beam and plasma macro-particles, the quasi-static approximation enables much larger time steps for the beam propagation, speeding up the simulations** by a factor depending on the beam energy, typically 50 to 200 for beams with an energy of 1 to 10 GeV. Fig. 2 describes one time step of a HiPACE++ simulation, see caption for algorithmic details. In the head-to-tail swipe over the longitudinal *co-moving* variable $\zeta = z - ct$, the algorithm to advance the plasma slab from slice i ($\zeta = i\Delta\zeta$) to slice $i-1$ ($\zeta = (i-1)\Delta\zeta$) is shown below. In HiPACE++, each of these operations is computed on a single GPU, keeping the data on the GPU memory, to avoid time- and energy-consuming communications and benefit from optimisations[11]:

- Deposit beam and plasma density on the grid of level 0;
- Solve for fields on the grid of level 0;
- Interpolate fields to ghost cells of level 1;
- Solve for fields on the grid of level 1;
- Advance plasma particles by 1 slice ($-=\Delta\zeta$);
- Advance beam particles in current slice by 1 time step ($+=\Delta t$);
- Advance laser cells in current slice by 1 time step ($+=\Delta t$);

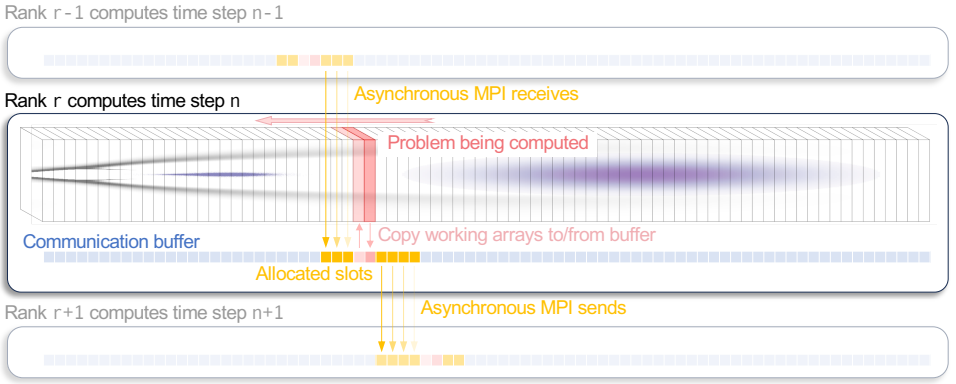


Figure 3. HiPACE++ parallelisation pipeline. On the current rank `r`, the communication buffer (stored in GPU or host memory) contains the beam (particle and laser) data and handles fully asynchronous exchanges with the neighbouring ranks. The data of the current slice is loaded to the compute arrays (in GPU memory), where the slice is computed and the resulting beam data is sent back to the buffer. An asynchronous MPI send is triggered, and an adjustable number of allocated slots allows for hiding load imbalance between ranks. Particles travelling significantly slower than the speed of light may slip backwards (to the left), and are then stored in the appropriate buffer slot.

The mesh refinement can be used to resolve small regions of interest that require high resolution, e.g., a tiny patch around the transversely small witness beam. In our setups, the refinement ratio from the refined level 1 to the base level 0 is typically on the order of 10 to 100 per transverse direction, leading to a speedup of 10^2 to 10^4 in comparison to a full simulation at high resolution. The computational steps specific to mesh refinement are shown in red and blue in the bullet points above, where the colours refer to the right schematic in Fig. 2. Furthermore, it should be noted that a full simulation at high resolution is often not possible due to memory constraints. Finally, as illustrated on Fig. 2, the computation of one time step for the full 3D domain in HiPACE++ is serial as it consists of a loop over longitudinal slices from the head of the box (on the right) to the tail. This makes standard domain decomposition inapplicable, at least in the longitudinal direction. Nevertheless, the part of the domain already computed (purple on Fig. 2) has already been advanced to the next time step and can be sent to the next rank for computation. This feature is fully capitalised on in HiPACE++ through an advanced pipeline algorithm illustrated on Fig. 3, where the beam (laser or particle) data of each slice is sent to the next rank as soon as the slice is computed. Thus, **the combination of mesh refinement and the novel pipeline provides drastic speedup to modelling plasma accelerators with quasi-static PIC codes**.

3 Steps Towards a Plasma-Based Particle Collider

To understand the problem we are about to discuss, we need to review some basic principles of a linear electron-positron collider. In a linear collider, two high-energy particle bunches are brought to head-on collision at the interaction point. To interpret the studied physics processes correctly, the kinetic energy of the colliding particles needs to be precisely known. Therefore, the energy spread in the particle bunch is an important parameter

of an accelerator for a collider. **The performance of a collider is measured by the event rate of the desired physics process, which scales with another important parameter, the luminosity**. The luminosity can be calculated from geometric considerations (for the interested reader, a good overview is given in Ref. 12). For beams with a Gaussian particle distribution, the luminosity scales inversely with the product of the root-mean-square (rms) beam size in the transverse planes $1/(\sigma_x\sigma_y)$ at the interaction point. The tighter the beams are focused at the collision point, the higher the event rate.

Rather than the beam transverse size at the interaction point, **domain scientist define a new quantity to describe the beam's quality: the emittance**. While strongly related to the beam size, it has the advantage of being at best preserved during acceleration, and growing in non-ideal conditions, such that tracking its evolution in a beamline helps scientist to quickly identify and address causes of emittance degradation (growth) in the beamline. The achievable beam size in each plane at the interaction point is proportional to the beam emittance $\epsilon_{[x,y]}$. Hence, the event rate scales inversely with $1/\sqrt{\epsilon_x\epsilon_y}$, which highlights why preserving the emittance as low as possible during acceleration is paramount. In particular, in order to preserve the emittance of a Gaussian particle beam, linear focusing fields are required. Nonlinear fields can lead to a finite emittance growth, that depends on the shape and strength of the nonlinearity.

The desired effects at the interaction point come from *individual interactions*, collisions involving a single pair of particles (one of each colliding beam). Unfortunately, the beams are also affected by *collective interactions*, in particular beamstrahlung where the space-charge force of each beam focuses the other beam, leading to significant deflection of trajectories and the emission of synchrotron radiation and loss of energy for the particles. Interestingly, beamstrahlung scales inversely with the sum of the rms beam sizes $1/(\sigma_x + \sigma_y)$[13] and hence $1/(\epsilon_x + \epsilon_y)$. **To maximise luminosity** $\propto 1/\sqrt{\epsilon_x\epsilon_y}$ **while minimising beamstrahlung** $\propto 1/\sqrt{\epsilon_x + \epsilon_y}$**, the community operates with flat beams, much bigger in the** x **direction than in** y**, namely** $\epsilon_x \gg \epsilon_y$. Such beams break the axial symmetry typically assumed for plasma accelerators and motivate the use of a 3D code like HiPACE++, which was prohibitively expensive up to now.

We were able to simulate the HALHF collider[6], a plasma-based Higgs factory proposal that received considerable attention in plasma and radio-frequency accelerator communities, including all relevant physics and fully resolving the small, flat witness beam with mesh refinement. A previously unknown effect was uncovered: the emittance mixing of flat beams in plasma accelerators due to transversely coupled wakefields[14]. Just like two mechanical oscillators can affect each other if there is a coupling term, as largely studied with pendulums[15–17], **the** x **and** y **dynamics of each beam particle can be coupled in the plasma accelerator in the presence of non-linearity in the focusing force causing a resonance**. In the worst case, the coupling causes the flat beam to become round (its large emittance in x is transferred to y), reducing the luminosity by orders of magnitude.

This is shown in Fig. 4, where the solid blue line shows the emittance in x and the dashed blue line the emittance in y. While the emittance in x decreases, the emittance in y increases drastically, leading to a significant increase of their geometric average. A resonance condition was identified as responsible for the worst-case scenario. This led to a possible solution: **the resonance can be broken by using a flat drive beam** (orange lines), leading to a suppression of the emittance exchange, as shown by the constant emittance in x and the resulting much smaller increase of the emittance in y. Note that the remaining

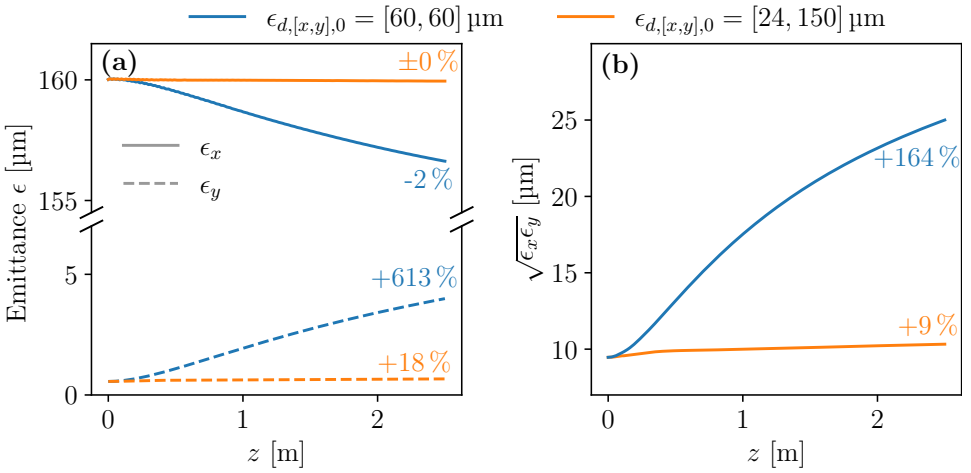


Figure 4. Emittance mixing in the first plasma stage of the HALHF collider. (a) For a round drive beam (blue lines) the emittance of the witness beam in x (solid lines) decreases by 2 %, while the emittance in y (dashed line) increases by 613 %. For a flat drive beam (orange lines) the emittance in x remains the same, while the emittance in y increases by 9 % due to nonlinear mismatch (b) The corresponding geometric average of the emittance for both cases. The emittance mixing with the round beam leads to a severe increase in $\sqrt{\epsilon_x \epsilon_y}$, while the nonlinear mismatch increases it only slightly.

emittance growth is caused by the still-present nonlinearity and is expected, but could in principle be prevented by advanced preparation of the witness bunch[18].

As it was shown, modelling plasma accelerators for colliders with ultra-high resolution can uncover effects that have major impact. **The emittance mixing described here has severe implications for any plasma-based collider using flat beams and significantly affects their design**. Besides uncovering emittance mixing in plasma accelerators, the new mesh refinement capabilities were also used to simulate positron acceleration in a warm plasma at unprecedented resolution, revealing that a finite temperature can improve the beam quality during acceleration[19,20]. In the future, HiPACE++ can be used to study other relevant phenomena for plasma-based colliders, such as staging, or the impact of radiation reactions of the beam particles.

4 Perspectives

The implementation of mesh refinement and our pipeline algorithm reduced considerably the simulation cost, not only allowing for previously unattainable simulation setups, but also making it possible to run large ensembles of simulations. This enables the use of **advanced optimisation methods, like multi-task Bayesian optimisation[21] with our library Optimas[22]** shown in Fig. 5, to propose optimised solutions for the future challenges of plasma acceleration for a collider application.

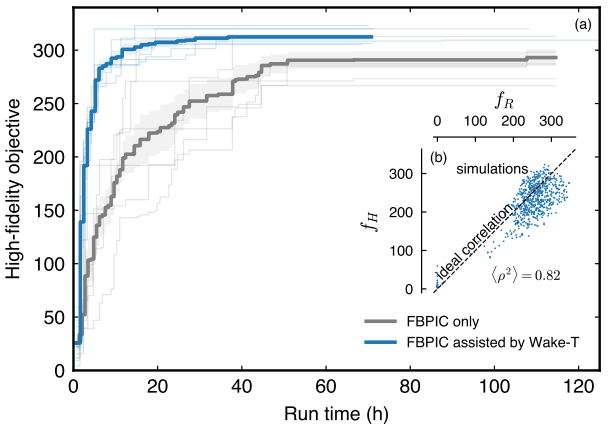


Figure 5. Multi-task Bayesian optimisation uses many low-fidelity simulations (low resolution and/or reduced model) to find an optimum with fewer production, and therefore more expensive, simulations. In this case, the production simulations are done with FBPIC[23], the reduced model uses Wake-T[24], and the time-to-convergence is considerably reduced provided the two models are sufficiently correlated. Reprinted figure with permission from [A. Ferran Pousa et al., Phys. Rev. Accel. Beams 26, 084601 (2023)] Copyright 2023 by the American Physical Society.

Acknowledgements

The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (`www.gauss-centre.eu`) for funding this project by providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS Supercomputer JUWELS [57] at Jülich Supercomputing Centre (JSC), through the allocation plasmabbq. This material is based upon work supported by the CAMPA collaboration, a project of the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research and Office of High Energy Physics, Scientific Discovery through Advanced Computing (SciDAC) program.

References

1. G. Aad et al., *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Physics Letters B, **716**, no. 1, 1-29, 2012.
2. S. Chatrchyan et al., *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Physics Letters B, **716**, no. 1, 30-61, 2012.
3. H. Murayama, S. Asai, K. Heeger, A. Ballarino, T. Bose, K. Cranmer, F.-Y. Cyr-Racine, S. Demers, C. Geddes, Y. Gershtein, B. Heinemann, J. Hewett, P. Huber, K. Mahn, R. Mandelbaum, J. Maricic, P. Merkel, C. Monahan, P. Onyisi, M. Palmer, T. Raubenheimer, M. Sanchez, R. Schnee, S. Seidel, S.-H. Seo, J. Thaler, C. Touramanis, A. Vieregg, A. Weinstein, L. Winslow, T.-T. Yu, and R. Zwaska, *Exploring the Quantum Universe: Pathways to Innovation and Discovery in Particle Physics*, 2024, arXiv:2407.19176.

4. European Strategy Group, *2020 Update of the European Strategy for Particle Physics*, Tech. Rep., Geneva, 2020.
5. E. Esarey, C. B. Schroeder, and W. P. Leemans, *Physics of laser-driven plasma-based electron accelerators*, Rev. Mod. Phys., **81**, 1229-1285, 8 2009.
6. B. Foster, R. D'Arcy, and C. A. Lindstrøm, *A hybrid, asymmetric, linear Higgs factory based on plasma-wakefield and radio-frequency acceleration*, New Journal of Physics, **25**, no. 9, 093037, Sep. 2023.
7. J. P. Farmer, A. Caldwell, and A. Pukhov, *Preliminary investigation of a higgs factory based on proton-driven plasma wakefield acceleration*, 2024, arXiv:2401.14765.
8. T. Roser, R. Brinkmann, S. Cousineau, D. Denisov, S. Gessner, S. Gourlay, P. Lebrun, M. Narain, K. Oide, T. Raubenheimer, J. Seeman, V. Shiltsev, J. Strait, M. Turner, and L.-T. Wang, *Report of the snowmass 2021 collider implementation task force*, Journal of Instrumentation, **18**, P05018, 2023.
9. `HiPACE++`, `https://github.com/Hi-PACE/hipace`.
10. S. Diederichs, C. Benedetti, A. Huebl, R. Lehe, A. Myers, A. Sinn, J.-L. Vay, W. Zhang, and M. Thévenet, *HiPACE++: A portable, 3D quasi-static particle-in-cell code*, Computer Physics Communications, **278**, 108421, 2022.
11. A. Myers, W. Zhang, A. Almgren, T. Antoun, J. Bell, A. Huebl, and A. Sinn, *AMReX and pyAMReX: Looking beyond the exascale computing project*, The International Journal of High Performance Computing Applications, **38**, no. 6, 599-611, 2024.
12. W. Herr and B. Muratori, *Concept of luminosity*, CERN Accelerator School Proceedings, 2006.
13. C. B. Schroeder, C. Benedetti, S. S. Bulanov, D. Terzani, E. Esarey, and C. G. R. Geddes, *Beam dynamics challenges in linear colliders based on laser-plasma accelerators*, Journal of Instrumentation, **17**, no. 05, P05011, 5 2022.
14. S. Diederichs, C. Benedetti, A. Ferran Pousa, A. Sinn, J. Osterhoff, C. B. Schroeder, and M. Thévenet, *Resonant emittance mixing of flat beams in plasma accelerators*, Phys. Rev. Lett., **133**, 265003, Dec. 2024.
15. A. Kovaleva and L. I. Manevitch, *Resonance energy transport and exchange in oscillator arrays*, Phys. Rev. E, **88**, 022904, Aug. 2013.
16. R. E. Berg and T. S. Marshall, *Wilberforce pendulum oscillations and normal modes*, American Journal of Physics, **59**, no. 1, 32-38, 1991.
17. L. R. Wilberforce, *XLIV. On the vibrations of a loaded spiral spring*, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, **38**, no. 233, 386-392, 1894.
18. C. Benedetti, C. B. Schroeder, E. Esarey, and W. P. Leemans, *Emittance preservation in plasma-based accelerators with ion motion*, Phys. Rev. Accel. Beams, **20**, 111301, 11 2017.
19. S. Diederichs, C. Benedetti, E. Esarey, M. Thévenet, A. Sinn, J. Osterhoff, and C. B. Schroeder, *Temperature effects in plasma-based positron acceleration schemes using electron filaments*, Physics of Plasmas, **30**, no. 7, 073104, 07 2023.
20. S. Diederichs, C. Benedetti, E. Esarey, A. Sinn, J. Osterhoff, C. B. Schroeder, and M. Thévenet, *Emittance-preserving acceleration of high-quality positron beams using warm plasma filaments*, 2023, arXiv:2311.07402.
21. A. Ferran Pousa, S. Jalas, M. Kirchen, A. Martinez de la Ossa, M. Thévenet, S. Hudson, J. Larson, A. Huebl, J.-L. Vay, and R. Lehe, *Bayesian optimization of laser-*

plasma accelerators assisted by reduced physical models, Phys. Rev. Accel. Beams, **26**, 084601, Aug. 2023.
22. `Optimas`, `https://github.com/optimas-org/optimas`.
23. R. Lehe, M. Kirchen, I. A. Andriyash, B. B. Godfrey, and J.-L. Vay, *A spectral, quasi-cylindrical and dispersion-free Particle-In-Cell algorithm*, Computer Physics Communications, **203**, 66-82, 2016.
24. A. Ferran Pousa, R. Assmann, and A. Martinez de la Ossa, *Wake-T: a fast particle tracking code for plasma-based accelerators*, Journal of Physics: Conference Series, **1350**, no. 1, 012056, Nov. 2019.

John von Neumann Institute for Computing
NIC Series

NIC Series Volume 39
NIC Symposium 2008 - Proceedings
20 - 21 February 2008, Jülich, Germany
edited by G. Münster, D. Wolf, M. Kremer (2008), iv, 358 pages
ISBN: 978-3-9810843-5-1

NIC Series Volume 40
From Computational Biophysics to Systems Biology (CBSB08)
Proceedings
edited by U.H.E. Hansmann, J.H. Meinke, S. Mohanty, W. Nadler,
O. Zimmermann (2008), viii, 430 pages
ISBN: 978-3-9810843-6-8

NIC Series Volume 41
Multigrid methods for structured grids and their application in particle simulation
by M. Bolten (2008), viii, 132 pages
ISBN: 978-3-9810843-7-5

NIC Series Volume 42
Multiscale Simulation Methods in Molecular Sciences - Lecture Notes
Winter School, 2 - 6 March 2009, Forschungszentrum Jülich
edited by J. Grotendorst, N. Attig, S. Blügel, D. Marx (2009), vi, 576 pages
ISBN: 978-3-9810843-8-2

NIC Series Volume 43
Towards the Confirmation of QCD on the Lattice
Improved Actions and Algorithms
by S. F. Krieg (2008), vi, 89 pages
ISBN: 978-3-9810843-9-9

NIC Series Volume 44
NIC Symposium 2010 – Proceedings
24 - 25 February 2010 | Jülich, Germany
edited by G. Münster, D. Wolf, M. Kremer (2012), v, 395 pages
ISBN: 978-3-89336-757-3

NIC Series Volume 45
NIC Symposium 2012 – Proceedings
25 Years HLRZ / NIC
7 - 8 February 2012 | Jülich, Germany
edited by K. Binder, G. Münster, M. Kremer (2012), v, 400 pages
ISBN: 978-3-89336-758-0

NIC Series Volume 46
Hybrid Particle-Continuum Methods in Computational Materials Physics Proceedings
4 - 7 March 2013 | Jülich, Germany
edited by M. H. Müser, G. Sutmann, R. G. Winkler (2013), ii, 232 pages
ISBN: 978-3-89336-849-5

NIC Series Volume 47
NIC Symposium 2014 - Proceedings
12 - 13 February 2014 | Jülich, Germany
edited by K. Binder, G. Münster, M. Kremer (2014), vi, 434 pages
ISBN: 978-3-89336-933-1

NIC Series Volume 48
NIC Symposium 2016 - Proceedings
11 - 12 February 2016 | Jülich, Germany
edited by K. Binder, M. Müller, M. Kremer, A. Schnurpfeil (2015), vi, 418 pages
ISBN: 978-3-95806-109-5

NIC Series Volume 49
NIC Symposium 2018 - Proceedings
22 - 23 February 2018 | Jülich, Germany
edited by K. Binder, M. Müller, A. Trautmann (2018), vi, 448 pages
ISBN: 978-3-95806-285-6

NIC Series Volume 50
NIC Symposium 2020 - Proceedings
27 - 28 February 2020 | Jülich, Germany
edited by M. Müller, K. Binder, A. Trautmann (2020), v, 424 pages
ISBN: 978-3-95806-443-0

NIC Series Volume 51
NIC Symposium 2022 - Proceedings
29 - 30 September 2022 I Jülich, Germany
edited by M. Müller, Ch. Peter, A. Trautmann (2022), v, 450 pages
ISBN: 978-3-95806-646-5

NIC Series Volume 52
NIC Symposium 2025 - Proceedings
6 - 7 March 2025 | Jülich, Germany
edited by Ch. Peter, M. Müller, A. Trautmann (2025), v, 423 pages
ISBN: 978-3-95806-793-6

The John von Neumann Institute for Computing (NIC) is a joint foundation of the three Helmholtz Centres Forschungszentrum Jülich, Deutsches Elektronensynchrotron DESY, and GSI Helmholtzzentrum für Schwerionenforschung as well as the Goethe University Frankfurt. It is a long term supporter of computational science in Germany and Europe.

The core task of the NIC is the peer-reviewed allocation of supercomputing resources to computational science projects in Germany and Europe. The NIC partners also support supercomputer-aided research in science and engineering through a three-way strategy:

- Provision of supercomputing resources for projects in science, artificial intelligence, research, and industry.
- Supercomputer-oriented research and development by research groups in selected fields of physics and natural sciences.
- Education and training in all areas of supercomputing by symposia, workshops, summer schools, seminars, courses, and guest programmes for scientists and students.

The NIC Symposium is held biennially to give an overview on activities and results obtained by the NIC projects in the last two years. The contributions for this 12th NIC Symposium are from projects that have been supported by the supercomputer JUWELS in Jülich. They cover selected topics in the fields of Astrophysics, Biophysics, Chemistry, Elementary Particle Physics, Materials Science, Theoretical Condensed Matter, Soft Matter Science, Earth and Environment, Computer Science and Numerical Mathematics, Fluid Mechanics and Engineering, Plasma Physics and Charged Particle Dynamics.

NIC Series Volume 52

ISBN 978-3-95806-793-6