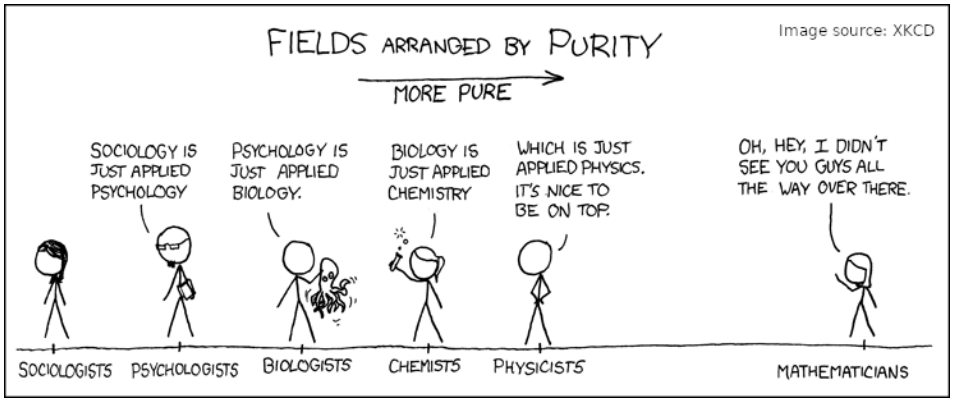
Machine Learning Multiscale Processes

An ICLR 2025 workshop proposal

Nikita Kazeev, Eléonore Vissol–Gaudin, Mengyi Chen, Isabelle Guyon and Andrey Ustyuzhanin

1 Scientific summary



Fundamental laws of Nature, Standard Model of Physics, and the most applied part of it, quantum mechanics, are well established. Theoretically, the dynamics of anything starting from a hydrogen atom and all the way to Earth's climate follow those equations. The problem is complexity Dirac [1929]. An exact computation of a modest system containing 100 atoms is still beyond the capability of modern computers.

Some of the greatest scientific achievements resulted from breakthroughs in scale transitions: renormalization [Nobel Prize 1982], density functional theory [Nobel Prize 1998], Higgs boson [Nobel Prize 2013], multiscale models for complex chemical systems [Nobel Prize 2013], climate modeling [Nobel Prize 2021], protein folding [Nobel Prize 2024]. Those achievements are highly regarded because they are impactful – but also unique and can't be readily applied to different systems.

Affiliation Omniscale Intelligence Initiative
Correspondence kna@nus.edu.sg
Version October 20, 2024

Encouraged by the recent successes, this workshop aims to enable the development of *universal* AI methods that would be able find efficient and accurate approximations, and use them for some of the most pressing and high-impact scientific problems that have computational complexity as the limiting factor to an in silico solution, such as:

1. High-temperature superconductivity
2. Fusion power
3. Weather prediction
4. Catalysts
5. Brain and consciousness

If we solve scale transition, we solve science.

1.1 Workshop themes and goals

In order to meet its objective, at the workshop we will work towards it in two orthogonal projections.

Methodology Develop the mathematical and algorithmic foundation for different approaches to multiscale machine learning:

1. Data–driven, e. g. dimensionality reduction Xie et al. [2022], Yang et al. [2022], manifold learning Brehmer and Cranmer [2020], Belyshev et al. [2024]
2. Assumption– and law–driven, e. g. machine learning interatomic interaction potential Korotaev et al. [2019], Wen et al. [2022], Kazeev et al. [2023], Hamiltonian Li et al. [2022] & ODE Chen et al. [2024] learning; explicit multiscale representation Somnath et al. [2021]
3. Reasoning– and code– driven, e. g. Gukov et al. [2024], He [2024], Rosas et al. [2024]
4. Tuning computational models with experimental data Bartók et al. [2013], Faber et al. [2017], Seljak et al. [2017]

Applications Collect **datasets**, **problem statements**, and ultimately **accelerate the innovation process** at every scale level:

1. Quantum, 10^{-10} m
2. Materials and Chemistry , 10^{-8} m
3. Life, 10^{-9} – 10^{0} m
4. Hydrodynamics, 10^{-3} – 10^{3} m
5. Climate, 10^{4} m
6. Astrophysics, 10^{4} – 10^{26} m

7. Artificial and social complex systems

1.2 Place in the ML conference workshop ecosystem

AI for science is a burgeoning field, with most recent NeurIPS, ICML, and ICLR conferences featuring workshops on physics, materials, climate, mathematics – or just generic AI for science. In contrast to them, our workshop is focused on a single question: *"Given low-level theory and computationally-expensive simulation code, how can we model complex systems on a useful time scale?"*. We will bring in people who work on different aspects of scale transition from different angles, normally belong to different fields – and rarely talk to each other.

2 Format

2.1 Submissions

We will solicit long papers (6 pages), and short paper (4 pages) with lower requirements, quoting the ICLR workshop guidance: less–than–full–conference papers that, for example, present an implementation and evaluation of an unpublished but simple idea, a modest but self-contained theoretical result, a follow–up experiment or re–analysis of a previously published paper, or a fresh perspective on an existing publication. Publication of code and data will be encouraged, but not mandatory. We will reject work previously published in archiving venues. We will however accept papers that are under review at any venue at the time of submission. Thematically, there will be three tracks:

1. New scientific result
2. Dataset or benchmark
3. Findings and open challenges

2.2 Review

In addition to internal resources we will use the channels listed in section 4.2 to solicit a diverse body of reviewers. We will use OpenReview to detect and avoid conflicts of interest following the ICLR policy. Organizers will be treated the same way as any other reviewers, and we will additionally ensure that they are not involved in the assessment of a submission from someone within the same organization. Reviews will be double-blind, with a designated organizer who will manage conflicts of interest, but won't participate in the scientific review.

Experiment to enhance reproducibility We will provide the submitters access to a cloud computing platform from our industrial partner offering a fully–

customizable Docker Merkel et al. [2014] environment to deploy their code and data. Reviewers will be able to run the code and analyze the data with minimum friction, using Visual Studio Code and Jupyter notebooks Kluyver et al. [2016] inside their web browser, without the need to set up or download anything, thus enchanting the review quality and reviewer experience. This will be voluntary, and participants will be able to propose another cloud platform, provided they ensure the double-blindness.

2.3 Tentative schedule

The event will feature 5 keynote talks, and a mixture of contributed talks and posters. Poster presenters will be given an opportunity to do a 2–3 minutes spotlight talk. The duration of the poster and spotlight sessions will be adjusted based on the number of posters, taking time from panel discussions and keynotes if necessary. Posters will hang for the whole day allowing people to engage with them during the coffee breaks. To foster debate, oral talks will be grouped in blocks, followed by a panel discussion. We will hold a competition for the best poster, and its author will be invited to join the final panel. When authors create their submissions, we will ask them to optionally leave questions to the panels. Finally, at the end of the workshop we will feature a pitching session where participants will advertise collaboration and job opportunities.

Tentative Timetable of the Day

Morning

8:00–8:45 Poster setup
8:45–09:00 Opening remarks
30m Keynote: Kostya Novoselov
9:30–12:30 Methodology:
30m Keynote: Daniel Polani
15m Contributed talk 1
15m Contributed talk 2
30m *Discussions and Coffee break*
30m Keynote: Sergei Gukov
15m Contributed talk 3
15m Contributed talk 4
30m Panel Discussion
1h Lunch

Afternoon

13:30–15:30 Poster Session:
30m spotlight talks
1h30m Poster session
30m *Discussions and Coffee break*
16:00–17:30 Applications:
30m Keynote: Charlotte Bunne
30m Keynote: Qianxiao Li
30m Panel Discussion
17:30–17:50 Collaboration pitches
17:50–18:00 Closing remarks

Tentative Schedule of Paper Submission We will follow the suggested dates:

- Workshop paper submission deadline: February 3, 2025
- Workshop paper notification date: March 3, 2025
- Camera-ready deadline: April 3, 2025

3 Invited Speakers

Kostya Novoselov (confirmed) Nobel Prize in Physics 2010, *inter alia*. Director of the Institute of Functional Intelligent Materials at the National University of Singapore, where he leads the *multiscale* endeavor to develop a new class of smart materials that actively respond to external stimulus – and can't be simulated with the existing methods.

Daniel Polani (confirmed) Professor of Artificial Intelligence at the Department of Computer Science, Director of the Centre for Artificial Intelligence and Robotics Research (CAIRR) and Head of the Adaptive Systems Research Group at the University of Hertfordshire. His research interests concentrate on the understanding and modeling of collective complex systems and intelligent decision-making, especially in the context of cognition in artificial and biological agents. His research ranges from fundamental questions, such as the role of embodiment, intrinsic motivations, taskless utilities, self-organization and Artificial Life, to questions from cognitive science, psychology, social science, and biology.

Sergei Gukov (confirmed) Director of Merkin Center for Pure and Applied Mathematics; Consulting Director of American Institute of Mathematics; John D. MacArthur Professor of Theoretical Physics and Mathematics at California Institute of Technology. Sergei is a member of the Scientific Board of the American Institute of Mathematics (AIM) and a member of the International Advisory Board of the Centre for Quantum Mathematics (QM). He has served on numerous other scientific committees and advisory boards. He is editor of the journal Communications in Mathematical Physics, Journal of Knot Theory and Its Ramifications, and Letters in Mathematical Physics. Known for Gukov-Vafa-Witten superpotential, Gukov-Witten surface operators, and Gukov-Pei-Putrov-Vafa (GPPV) invariants. Sergei's expertise is uniquely positioned at the intersection of theoretical physics, mathematics and machine learning.

Charlotte Bunne (confirmed) Assistant professor at EPFL in the School of Computer and Communication Sciences (IC) and School of Life Sciences (SV) and a

member of the Swiss Institute for Experimental Cancer Research (ISREC). Before, she was a PostDoc at Genentech and Stanford with Aviv Regev and Jure Leskovec and in 2023 completed a PhD in Computer Science at ETH Zurich working with Andreas Krause and Marco Cuturi. During her graduate studies, she was a visiting researcher at the Broad Institute of MIT and Harvard hosted by Anne Carpenter and Shantanu Singh and worked with Stefanie Jegelka at MIT. Her research aims to advance personalized medicine by utilizing machine learning and large-scale biomedical data. Charlotte Bunne's interdisciplinary research has won several (best paper) awards. Charlotte has been a Fellow of the German National Academic Foundation and is a recipient of the ETH Medal.

Qianxiao Li (confirmed) Assistant professor in the Department of Mathematics, National University of Singapore. He graduated with a BA in mathematics from the University of Cambridge and a PhD in applied mathematics from Princeton University in 2016. His research interests include the interplay of machine learning and dynamical systems, control theory, stochastic optimization algorithms and data-driven methods for science and engineering. He is a recipient of the PSTA Young Scientist Award, National Research Foundation Singapore. Qianxiao has published in top material science and machine learning venues, such as Journal of Machine Learning Research, International Conference on Machine Learning, and Matter.

4 Audience

4.1 Size estimation

As a benchmark we looked through some recent related workshops:

- ML4PS @ NeurIPS2023 (250 papers)
- AI4DifferentialEquations @ ICLR 2024 (87 papers)
- ML4LMS @ ICML 2024 (64 papers)

A keynote by a Nobel Laureate might attract attendees in addition to the paper submitters. Singapore has a robust AI for science ecosystem to which the organizing team is well-connected. Final projection: 100 – 200.

4.2 Promotion

1. Mailing lists of the conferences previously organized by our team and friends: AI4SCI/NTCI, AI4AM, Accelerate 2024.
2. Social media of the organizers and speakers
3. Institutional resources and mailing lists of the organizers and speakers

4. We'll ask the AI for science communities that won't organize workshops at ICLR 2025 to disseminate our announcement; we'll offer cross-promotion to those who do
5. Resources of industrial partners
6. Local Asian scientific organizations and universities
7. Affinity groups

5 Virtual access to workshop materials and outcome

We strictly aim for hosting an in-person event. If a speaker is unable to attend due to exceptional circumstances (a good criterion will be that they have registered for the conference), organizers can display posters on their behalf, and include a video in the website materials. Accepted papers and posters will be published on the website in a non-archiving way. In order to facilitate the choice of attendance based on content, the accepted paper titles will be put on the website at the moment of acceptance, not at the camera–ready deadline. Subject to resource availability the workshop will be streamed online and recorded.

5.1 Experimental knowledge model

We will deploy an LLM-driven knowledge model made by Constructor. Its core is formed by retrieval-augmented generation Lewis et al. [2020] over accepted papers. The model will present each participant with a targeted recommendation of papers and people – this will facilitate discussion, networking and mentorship. The model also has a chat interface allowing free–form queries to the collective knowledge base of the workshop and individual papers within it; this will improve the interdisciplinary dialog by helping people understand the work from others' domains, and will be particularly useful for those unable to attend and discuss in person.

6 Diversity commitment

We are especially proud in the diversity of approaches and expertise presented by our speakers, drawn from the different natural science domains and AI paradigms.

Organizing team & invited speakers body is international and demographically diverse, it contains males and females; Western European, Slavic, and Asian ethnicities. The organizing team includes a PhD student, postdocs, and senior researchers. The 5 invited speakers come from 4 countries on 3 continents. The most junior invited speaker got PhD in 2023, the most senior got a Nobel Prize.

We are committed to reaching under-represented groups by ensuring invitations and promotional materials are circulated through affinity groups (WIML, Black in AI, and others).

We have secured an initial sponsorship offer of US$6000 and in-kind from Constructor Technology. It will enable us to provide financial assistance. While the award will be need–based, we will focus outreach on the nearby underrepresented developing countries in South-East Asia, where such aid will make the most impact. Sponsorship will also allow us to organize lunch and dinner accessible to all participants to facilitate networking and discussions. We will work to secure additional sponsorship.

7 Organizers

Most of the organizing team is based in Singapore. This ensures a deep understanding of the local research community, strong ties with Singapore's thriving AI and machine learning ecosystem, and easier collaboration with local institutions and universities. This local presence will also allow us to efficiently manage workshop logistics and leverage regional networks to attract a diverse range of participants from underrepresented groups in South–East Asia.

Nikita Kazeev Research Fellow in the National University of Singapore. In 2020 he got a double PhD degree in Computer Science from HSE University & in Physics from Sapienza University of Rome. Nikita has worked on a diverse range subjects in the area of on machine learning for science, from particle physics and materials to spacecraft evasive maneuvers; mentored students and interns; promoted science in lectures and article for general public. He reviewed for AI4Mat @ [NeurIPS 2024 & 2023, Vienna 2024].

Eléonore Vissol–Gaudin Research Fellow in the Department of Materials Science and Engineering at Nanyang Technological University and member of the Hip lab. She got her PhD from Durham University in the UK, working on the development of unconventional computing devices. Her current research focuses on data-driven modelling of dynamical systems and integrating machine learning into experimental workflows. She has worked on the organisation of the AI for Science and Nobel Turing Challenge Conference 2024, along with 14 domain-specific AI for Science workshops in Singapore. She has been associate editor and reviewer for the IEEE Nanotechnology Council flagship conference IEEE NANO 2024 and chaired sessions at IEEE NANO2024 and the Singapore MRS' ICMAT2023.

Mengyi Chen PhD student in the Department of Mathematics, National University of Singapore. Her research interests focus on leveraging machine learning to explore the dynamics of complex systems. Mengyi has published in top machine learning venues such as the Conference on Neural Information Processing Systems (NeuRIPS).

Isabelle Guyon Director, Research Scientist, Google DeepMind. Chaired Professor of Artificial Intelligence (PR EX1) and INRIA researcher, University Paris-Saclay. Co-program chair of NeurIPS 2016 and co-general chair of NeurIPS 2017; then NeurIPS board member. AMIA and an ELLIS fellow. Action editor at JMLR, and CiML Springer series editor. BBVA award recipient (2020). Since 2003 machine learning challenge organizer as a means of directing research in domains including causality, computer vision, automatic machine learning, and high energy physics.

Andrey Ustyuzhanin Director of AI/ML Research at Acronis, and an Adjunct Professor at Constructor University Bremen. Founder of the Omniscale Intelligence Initiative. His main motivation is to achieve a better understanding of how the universe and consciousness work together. His main area of expertise includes applying artificial intelligence techniques and methods to various scientific domains. With over 20 years of active research and an impressive h-index above 130, Andrey is driven to push the boundaries of computational science through innovative simulation techniques, the design of generative models, the application of optimization algorithms, and the use of interpretable machine learning processes to accelerate scientific discovery and promote human evolution. Andrey has extensive organizational experience, including 7 years of organizing the international summer school "Machine Learning for High Energy Physics", data challenges, ALEPH Workshop @ NIPS 2015, TrackML Competition @ NeurIPS 2018, and multiple local workshops on AI and Physics.

References

Albert P Bartók, Michael J Gillan, Frederick R Manby, and Gábor Csányi. Machine-learning approach for one-and two-body corrections to density functional theory: Applications to molecular and condensed water. *Physical Review B—Condensed Matter and Materials Physics*, 88(5):054104, 2013.

Antonii Belyshev, Alexander Kovrigin, and Andrey Ustyuzhanin. Beyond dynam-

ics: learning to discover conservation principles. *Machine Learning: Science and Technology*, 5(2):025055, 2024.

Johann Brehmer and Kyle Cranmer. Flows for simultaneous manifold learning and density estimation. *Advances in neural information processing systems*, 33: 442–453, 2020.

Xiaoli Chen, Beatrice W Soh, Zi-En Ooi, Eleonore Vissol-Gaudin, Haijun Yu, Kostya S Novoselov, Kedar Hippalgaonkar, and Qianxiao Li. Constructing custom thermodynamics using deep learning. *Nature Computational Science*, 4(1): 66–85, 2024.

Paul Dirac. Quantum mechanics of many-electron systems. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 123(792):714–733, 1929.

Felix A Faber, Luke Hutchison, Bing Huang, Justin Gilmer, Samuel S Schoenholz, George E Dahl, Oriol Vinyals, Steven Kearnes, Patrick F Riley, and O Anatole Von Lilienfeld. Prediction errors of molecular machine learning models lower than hybrid dft error. *Journal of chemical theory and computation*, 13(11):5255–5264, 2017.

Sergei Gukov, James Halverson, and Fabian Ruehle. Rigor with machine learning from field theory to the poincaré conjecture. *Nature Reviews Physics*, pages 1–10, 2024.

Yang-Hui He. Ai-driven research in pure mathematics and theoretical physics. *Nature Reviews Physics*, pages 1–8, 2024.

Nikita Kazeev, Abdalaziz Rashid Al-Maeeni, Ignat Romanov, Maxim Faleev, Ruslan Lukin, Alexander Tormasov, AH Castro Neto, Kostya S Novoselov, Pengru Huang, and Andrey Ustyuzhanin. Sparse representation for machine learning the properties of defects in 2d materials. *npj Computational Materials*, 9(1):113, 2023.

Thomas Kluyver, Benjamin Ragan-Kelley, Fernando Pérez, Brian Granger, Matthias Bussonnier, Jonathan Frederic, Kyle Kelley, Jessica Hamrick, Jason Grout, Sylvain Corlay, et al. Jupyter notebooks–a publishing format for reproducible computational workflows. In *Positioning and power in academic publishing: Players, agents and agendas*, pages 87–90. IOS press, 2016.

Pavel Korotaev, Ivan Novoselov, Aleksey Yanilkin, and Alexander Shapeev. Accessing thermal conductivity of complex compounds by machine learning interatomic potentials. *Physical Review B*, 100(14):144308, 2019.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33:9459–9474, 2020.

He Li, Zun Wang, Nianlong Zou, Meng Ye, Runzhang Xu, Xiaoxun Gong, Wenhui Duan, and Yong Xu. Deep-learning density functional theory hamiltonian for efficient ab initio electronic-structure calculation. *Nature Computational Science*, 2(6):367–377, 2022.

Dirk Merkel et al. Docker: lightweight linux containers for consistent development and deployment. *Linux j*, 239(2):2, 2014.

Fernando E Rosas, Bernhard C Geiger, Andrea I Luppi, Anil K Seth, Daniel Polani, Michael Gastpar, and Pedro AM Mediano. Software in the natural world: A computational approach to emergence in complex multi-level systems. *arXiv preprint arXiv:2402.09090*, 2024.

Uroš Seljak, Grigor Aslanyan, Yu Feng, and Chirag Modi. Towards optimal extraction of cosmological information from nonlinear data. *Journal of Cosmology and Astroparticle Physics*, 2017(12):009, 2017.

Vignesh Ram Somnath, Charlotte Bunne, and Andreas Krause. Multi-scale representation learning on proteins. *Advances in Neural Information Processing Systems*, 34:25244–25255, 2021.

Tongqi Wen, Linfeng Zhang, and Han Wang. Weinan e, david j srolovitz. deep potentials for materials science. *Materials Futures*, 1(2):022601, 2022.

Xiaoyu Xie, Arash Samaei, Jiachen Guo, Wing Kam Liu, and Zhengtao Gan. Data-driven discovery of dimensionless numbers and governing laws from scarce measurements. *Nature communications*, 13(1):7562, 2022.

Tsung-Yen Yang, Justinian Rosca, Karthik Narasimhan, and Peter J Ramadge. Learning physics constrained dynamics using autoencoders. *Advances in Neural Information Processing Systems*, 35:17157–17172, 2022.