FRONTIERS IN PROBABILISTIC INFERENCE: LEARNING MEETS SAMPLING

HTTPS://SITES.GOOGLE.COM/VIEW/FPIWORKSHOP/ABOUT

ABSTRACT

Probabilistic inference, particularly through the use of sampling-based methods, is a cornerstone for modeling across diverse fields, from machine learning and statistics to natural sciences such as physics, biology, and chemistry. However, many challenges exist, including scaling, which has resulted in the development of new machine learning methods. In response to these rapid developments, we propose a workshop, FRONTIERS IN PROBABILISTIC INFERENCE: LEARNING MEETS SAMPLING (FPI-2025), to foster collaboration between communities working on sampling and learning-based inference. The workshop aims to center community discussions on (i) key challenges in sampling, (ii) new sampling methods, and (iii) their applications to natural sciences and uncertainty estimation. We have assembled an exciting speaker list with diverse perspectives; our goal is that attendees leave with a deeper understanding of the latest advances in sampling methods, practical insights into their applications, and new connections to collaborate on future research endeavors.

1 WORKSHOP SUMMARY

Motivation. Probabilistic inference (Onsager & Machlup, 1953; Black & Scholes, 1973), particularly through the application of sampling-based methodologies, serves as a foundational framework for a wide range of applications and modeling approaches across diverse fields, including machine learning, statistics, and the natural sciences such as physics, biology, and chemistry. Using techniques such as deep generative models (Kingma, 2013; Rezende et al., 2014; Goodfellow et al., 2020), Markov Chain Monte Carlo (MCMC) (Chen et al., 2014; Doucet et al., 2001) and Bayesian posterior sampling (Blundell et al., 2015; Welling & Teh, 2011; Izmailov et al., 2021), researchers can effectively incorporate uncertainty and variability into their models. In physics, these methods assist in interpreting experimental data and simulating complex systems (Adam et al., 2022; Levy et al., 2024), while in chemistry, they help predict molecular behavior and reaction outcomes by exploring a vast parameter space efficiently (Aranganathan et al., 2024; Deringer et al., 2019). This ability to sample from probability distributions not only enhances the accuracy of predictions but also provides valuable insights into underlying processes, making probabilistic inference and sampling essential for addressing real-world challenges across diverse domains.

While sampling approaches have gained significant traction in data-rich domains (Kingma, 2013; Rezende et al., 2014; Goodfellow et al., 2020; Ho et al., 2020), modern probabilistic inference methods must address the pressing and underexplored challenge of sampling proportional to unnormalized densities—an area that presents unique difficulties due to limited or non-existent data availability (Watson et al., 2023; Duan et al., 2023). Indeed, sampling in itself is a fundamental question in machine learning and specifically sampling according to unnormalized densities encompasses a wide array of complex and timely problems, including molecular dynamics simulations (Aranganathan et al., 2024; Deringer et al., 2019), Bayesian posterior inference and inverse problems (Adam et al., 2022; Levy et al., 2024), and generation from conditional or product densities (such as in diffusion models and large language models), where pre-trained model's density is weighted by a known target density (e.g., finetuning and inference-time alignment) (Domingo-Enrich et al., 2024b; Zhao et al., 2024). As a result, core advances in such sampling methodologies have the potential to impact several problems of interest simultaneously.

Despite being a long-standing challenge in science and statistics, new promises have been made from the connections between *learning* deep probabilistic models (e.g. diffusion, flow models) and dynamical systems, thermodynamics, stochastic optimal control, and optimal transport (Sohl-Dickstein et al., 2015; Song et al., 2021; Lipman et al., 2023; Tong et al., 2024). To scale to large problems, it is expected that further advances in sampling can be spurned by embracing learning as a core component that bolsters or even supplants existing local search methods such as Markov chain

Monte Carlo (MCMC). Indeed, new learning-based samplers have already provided deeper theoretical insights without sacrificing their importance for use in practical applications (Chen et al.; Noé et al., 2019). This new surge of interest is in particular driven by advances in generative modeling which have the potential to unlock some of the challenges in scalability faced by more traditional techniques. These modern approaches include, but are not limited to, combining Monte Carlo techniques, (Du et al., 2024; Huang et al., 2024; Grenioux et al., 2024; Midgley et al., 2023), stochastic optimal control (Holdijk et al., 2023; Domingo-Enrich et al., 2024a), connecting optimal transport and variational inference (Vargas et al., 2024), diffusion-based samplers (Richter et al., 2023; Vargas et al., 2023; Zhang & Chen, 2022) or generative flow networks (Venkatraman et al., 2024).

In order to support the burgeoning interest in sampling through learning-based approaches, we propose a full-day workshop called FRONTIERS IN PROBABILISTIC INFERENCE: LEARNING MEETS SAMPLING (FPI-2025). The primary goal of this workshop is to facilitate community building and to work towards the identification of key challenges in comparison to classical sampling-based approaches, along with techniques to overcome these challenges. Specifically, we aim to create a space where different communities in the periphery of sampling can come together and harmonize a new set of tools and ideas that can be used to tackle problems in probabilistic inference. We have found that there is a dire lack of space or a workshop dedicated to examining the interchange of tools that are used in the sampling community with the new techniques in the fast-growing field of generative modeling. Bringing together these communities to discuss how to combine amortized approaches to solving real-world probabilistic inference problems has significant potential and can lead to breakthroughs beyond classical machine learning.

FPI-2025 Technical themes. We will center the workshop discussions around the following topics/questions:

- **Topic 1: Design of new sampling methods using optimal control, variational calculus, and partial differential equations**. To explore this, we have invited the following speakers with relevant works on diffusion models and Schrodinger bridges—Sitan Chen, Francisco Vargas, Grant M. Rotskoff, and Rianne van den Berg.
- **Topic 2: Probabilistic inference in applications to natural sciences and uncertainty estimation** We invite the following speakers based on their works on molecular dynamics, protein generation and Bayesian posterior inference—Marylou Gabrié, Rianne van den Berg and Mohammad Emtiyaz Khan.

To address these questions and challenges, we have invited a selection of speakers and panelists renowned for their pioneering contributions to the field of sampling. They will approach this topic from multiple perspectives, including theoretical advancements, methodological innovations, and practical applications in statistical physics, biology, and chemistry. We aim to both uncover and highlight the promising recent advancements, as well as the remaining and emerging challenges. The talks by the speakers will foster deeper discussions around the methodological challenges of this field. Moreover, they will highlight how tools and methods from various subfields, which have traditionally progressed independently, can be harnessed to address broader challenges in sampling and probabilistic inference. Our panel will also provide a collaborative and interactive setting in which we dig deeper into new questions about using learned samplers as opposed to traditional sampling methods, as well as how they unlock potential scientific applications in the industry.

FPI-2025 Community engagement.

- **Call for contributed work.** We will welcome contributed work across three different tracks to encourage a diverse array of submissions in this newly growing field: (1) the *Research papers* track (4-8 pages) for any original research work (2) the *Challenges and Reflections* track (4-8 pages), dedicated to exploring setbacks, unexpected outcomes, and the valuable lessons learned from methods that didn't achieve their intended goals, (3) the *Benchmarks and Datasets* track (4-8 pages) for new and interesting settings for researchers to test their models on, and (4) the *Tiny Papers* track (max. 2 pages) in response to ICLR's DEI initiative to encourage research from under-represented, under-resourced and budding researchers. Listed page limits do not include references and appendices. Papers will use the ICLR 2025 template and will be non-archival; the titles of accepted works will be posted on our website.

- **Travel Grants**. High travel costs can pose obstacles for interested participants. If we are able to obtain corporate sponsorshipg (see Section 4.5), we will provide travel grants to marginalized and under-represented communities to reduce the barrier for them to attend the workshop and providing them a platform to be a part of the larger sampling community. Before the workshop, we will send out forms to workshop contributors so that they can apply for these grants if we are able to obtain funding.
- **Networking opportunity & social**. To encourage discussion and collaboration among researchers and speakers from diverse technical and cultural backgrounds as in the sampling community, we will have a networking opportunity and dinner to conclude the workshop at a local nearby restaurant. All speakers and attendees will be invited. We will try to procure funding to subsidize the cost of this (see Section 4.5) for workshop contributors. Our goal with this is to provide an opportunity for both young and under-represented researchers to connect with the larger community and spark new ideas and collaborations.

2 TENTATIVE SCHEDULE

This one-day workshop will consist of 6 invited talks, 6 spotlight talks, 1 panel discussion and 2 poster sessions. We show a tentative schedule that we envision for the workshop day in Table 1.

Invited talks (6 $\times$ 30 mins (25 mins talk, 5 mins Q&A)). We will host leading experts in the field of sampling and probabilistic inference who will share their insights on the recent advancements and the challenges in this field. The speakers are from a wide range of institutions and backgrounds. The speakers will share their views on different perspectives of the sampling problem: theoretical discussions, variational inference and Bayesian approximations, applications to biophysics and molecular simulations, machine learning methods, and in particular generative modeling for sampling. **All of the speakers have been contacted and confirmed their interest in participating in the workshop.**

- **Marylou Gabrié: Assistant Professor @ Physics Department of École Normale Supérieure**. Her research involves developing ML methods to solve computational problems in statistical physics. Her recent work focuses on developing deep generative models for sampling, with applications to complex multi-modal distributions in molecular modeling and statistical physics. She is, therefore, uniquely well-positioned to discuss scalable learned sampling algorithms, which are rapidly gaining traction in both physics and machine learning communities.
- **Grant M. Rotskoff: Assistant Professor @ Chemistry Department of Stanford University**. His research interests lie in the intersection between statistical mechanics and ML, from sampling protein conformation ensembles to understanding diffusion models from nonequilibrium thermodynamics. Grant's works shed light on one of the key applications of sampling and he can highlight some of the real-world applications of sampling and the challenges associated with it.
- **Sitan Chen: Assistant Professor @ Computer Science Department of Harvard University**. He is interested in designing algorithms with provable guarantees for fundamental problems in data science, with a focus on generative modeling (e.g. understanding diffusion models from sampling and theoretical perspective). He can bring his unique expertise in understanding sampling and its applications from the theoretical and algorithmic perspectives.
- **Emtiyaz Khan: Team Leader @ RIKEN Center for Advanced Intelligence Project (AIP) in Tokyo**. He specializes in deterministic Bayesian approximations, such as variational inference, and enhances them by applying optimization tools and concepts from information geometry. Within the field of deep learning, his focus encompasses supervised learning, Variational Auto-Encoders, and, more recently, model merging, continual learning, and general model adaptation.
- **Rianne van den Berg: Principal Research Manager @ Microsoft Research Amsterdam**. She works on the intersection of deep probabilistic models and computational chemistry and physics for molecular simulation. She has an extensive body of research on generative modeling with applications to proteins and molecular dynamics and, therefore, is uniquely positioned to highlight how tools and techniques from generative modeling can be applied to sampling for real-world applications.

- **Francisco Vargas: ML Research Scientist @ Xaira Therapeutics, PhD Student @ Cambridge**. He has made several important and pioneering contributions at the intersection of transport and stochastic control, focusing on free energy computations with applications in sampling and conditional generative modeling. He is therefore an ideal speaker to discuss the methodological advancements of the field, as well as its applications in the drug design industry.

Spotlight Talks (6×10 mins). We will have spotlight talks from the top submissions to the workshop, allowing the participants to showcase their research in more detail.

Poster Sessions (2×60 mins). Throughout the day, we will host two poster sessions where authors of accepted submissions can present their work in an interactive manner. The accepted submissions will be divided into these two groups beforehand. We will encourage active participation from the audience by holding a vote and a prize for the best poster.

Panel Discussion (1×60 mins). The panel discussion will be an engaging discussion where we will delve deeper into the topic of how sampling helps learning and how learning can help sampling. Some of the questions we aim to discuss in this panel are:

- Scalability challenges in traditional probabilistic inference and how learning-based samplers might address them.
- The interplay between local search techniques and amortized inference in emerging methods.
- Theoretical guarantees vs. empirical evidence: Can learned samplers completely replace traditional methods?
- Shortcomings of current learned samplers: open questions and future directions.
- Applications where combining sampling and learning is essential, such as climate simulations and drug discovery.
- Can the sampling perspective lead to novel methods for controllable generation of generative models?

We also plan to curate a list of questions with additional questions from the audience to facilitate discussions. For our panel, we have selected a subset of invited speakers and invited a new one to balance previously-discussed perspectives while introducing new takes across various axes. **Ricky T. Q. Chen** brings expertise in real-world applications, particularly in discrete space, fine-tuning, and inference. **Rianne van den Berg** contributes her knowledge of chemistry and biology applications. **Marylou Gabrié** offers insights into novel sampling methods with applications to physics, and **Sitan Chen** for his insights on theoretical aspects of learned samplers.

3 Diversity commitment

A key priority for FPI-2025 is to ensure that the diverse viewpoints of the entire community are reflected during our workshop. To this point, we have selected speakers and panelists for diversity along several axes, including gender (2 female-identifying, 5 male-identifying), geography (Asia, Europe, North America), affiliations (4 academia, 2 industry, 1 joint academia and industry), seniority (2 senior and 5 junior), ethnicity, and expertise (including Biology, Physics, Bayesian Deep Learning; we visualize them in the last panel of Figure 1). We have also constructed our organizing committee to uphold a diverse set of perspectives in terms of seniority (4 PhD students, 2 postdoctoral fellows, 5 professors), ethnicity (we have identified 12 cultural backgrounds in our committee: Iranian, Chinese, Indian, American, Canadian, Polish, Ukrainian, Dutch, French, Russian, Israeli, Finnish), gender (2 female-identifying, 9 male-identifying), geography (North America, Europe), and affiliations (7 academic institutions, 4 companies). Furthermore, two individuals on our committee are first-time organizers. We visualize distributions of institution location, career stage, and diverse research interests in Figure 1.

We aim to phrase our call for submissions to be welcoming of all technical and demographic backgrounds, which is also supported by having multiple submission tracks.

Time	Description	Speakers
08:00 – 08:10	Introduction & opening remarks	*Organizing committee*
08:10 – 08:40	Invited talks, Topic 1: *Sampling methods and links to differential equations & optimal control*	**Sitan Chen**
8:40 – 9:40	Poster session I	Accepted works
9:40 – 10:10	Spotlight talks (x3)	TBD - selected from contributed works
10:10 – 10:30	Coffee break	Everyone
10:30 – 11:00 11:00 – 11:30	Invited talks, Topic 1 cont'd	**Francisco Vargas** **Grant M. Rotskoff**
11:30 – 12:30	Panel	*Moderator (organizing committee member)* & Panelists: **Ricky T.Q. Chen** **Rianne van den Berg** **Marylou Gabrié** **Sitan Chen**
12:30 – 13:30	Lunch break	Everyone
13:30 – 14:00	Spotlight talks (x3)	TBD - selected from contributed works
14:00 – 14:30	Invited talks, Topic 2: *Applications of sampling to natural sciences & uncertainty estimation*	**Marylou Gabrié**
14:30 – 15:30	Poster session II	Accepted works
15:30 – 15:50	Coffee break	Everyone
15:50 – 16:20 16:20 – 16:50	Invited talks, Topic 2 cont'd	**Rianne van den Berg** **Mohammad Emtiyaz**
16:50–17:00	Closing remarks & Best paper announcement	*Organizing committee*
17:00	Networking opportunity	Everyone (*at nearby restaurant*)

Table 1: Proposed schedule for FPI-2025. All times are in local Singapore Standard Time. Speakers and panelists who have confirmed their participation are highlighted in **bold**. Start time, end time, coffee breaks and lunch are based on times from ICLR-2024 workshops; these will be adjusted when the ICLR-2025 schedule is released.

4 Organizational details

4.1 Timeline

We aim to strictly adhere to the suggested timeline provided by the ICLR workshop committee, also allowing a few buffer days before mandatory deadlines in case anything unexpected arises.

- Workshop acceptance notification: 02 December, 2024
- Website release & advertising: 10 December, 2024
- Submission portal open: 06 January, 2025
- Submission deadline: 03 February, 2025
- Acceptance notification: 02 March, 2025

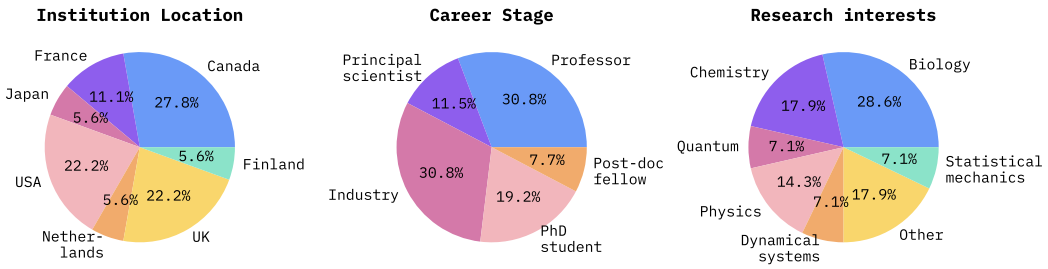

Figure 1: Distributions of institution location, career stage, and primary research interests of our combined organizing committee, invited speakers, and panelists.

- Workshop: 27 or 28 April, 2025

4.2 Call for submissions and review process

We will host all papers using the OpenReview platform and allow comments and questions prior to the workshop. Through this, we efficiently handle conflicts of interest while facilitating an easy forum for concentrated discussion on any given paper that will remain accessible after the workshop. Each of our submission tracks is detailed in §1 and consists of calls for original unpublished contributed work and a tiny papers' track that aims to improve accessibility to this growing subfield.

Details of the review process. The review process, including all communications, will also be done entirely using OpenReview. We anticipate receiving approximately 50 to 100 submissions based on the increasing interest in sampling and the record from previous workshops that bear some overlap with probabilistic inference. To ensure a high-quality peer-review process, each paper will be reviewed by at least 3 reviewers, and our goal is for each reviewer to review at most 3 papers. Due to the rich organizing experience of our organizing team, we have already recruited a list of more than 100 reviewers with diverse expertise and interests in probabilistic inference and applications in science (listed in Section 7), and we will continue recruiting reviewers after the acceptance of the workshop. To prevent any conflicts of interest between reviewers and submission authors, we will use the OpenReview built-in matching algorithm that takes into account author OpenReview profiles and institution histories when assigning papers to reviewers. The organizing committee will parse the reviews and make the final decisions on which works are accepted. If a member of the organizing committee has a conflict of interest with a work, they will not be involved in making the decision for that work. FPI-2025 has no target acceptance rate; rather, our goal is to accept all high-quality work that we think the community would benefit from, as determined by the peer-review system.

4.3 Workshop attendance

Expected size. The workshop topic is intentionally focused and seeks to bring together a community of practitioners building probabilistic inference techniques for sampling and their applications in domains such as statistics, machine learning, and the physical sciences. Based on the growing interest in this area, the growth of ICLR attendance, the organizers' past experiences in contributing to workshops, and increased inclusivity efforts in ICLR workshops, we expect between 100-200 participants. We will also accommodate online attendees following the ICLR livestreaming guidelines.

Details on the panel session. Authors will be asked to submit potential panel questions through Slido. In addition to live-audience questions, we will also solicit questions via X/Twitter. During the panel itself, we will begin the discussion with questions we had previously collected and transition to taking questions from the audience.

Audience Engagement. We will encourage an interactive and engaging workshop in various ways. We highlight and reiterate some of these strategies below.

- We have 6 spotlight talks from selected accepted submissions. This will allow the participants to further showcase their research at the workshop.

- We will have a panel discussion where part of the discussion will be centered around questions from the participants. We will solicit these questions before the panel discussion using social media channels and will also take questions from the audience during the panel.
- We will have a "best poster vote" and a prize, where we encourage participants to be more engaged during the poster session and to submit their vote for a poster of their choice from the two poster sessions at the workshop.
- We will have a networking opportunity at the end of the workshop day to encourage further discussion and collaboration among the participants.

4.4 Virtual access

Titles of accepted works will be posted on our website shortly after author notification, and spotlight talks will be marked with a special symbol.

While we are planning for the workshop to be held primarily in person, we appreciate that exceptional circumstances such as visa issues can arise. For invited and spotlight speakers, we will ask them to submit a recording of their talk beforehand if they are suddenly unable to attend. If ICLR regulations permit, we will live-stream them over Zoom to deliver the talk in real-time; otherwise, we will play the recording and provide a forum for attendees to submit questions. For poster presenters, we will display their posters on our website and give the option of recording a short 2-minute presentation of the poster for those participants.

4.5 Funding

NVIDIA has generously partnered with us as a corporate sponsor for FPI-2025; they will provide NVIDIA RTX A6000 - 48GB GPUs for the best papers, as determined by the organizing committee, and the best posters, as determined by audience vote. We are in the final stages of discussions with other companies—who have positively responded to this workshop—so that additional corporate sponsorship may be secured. Acquired funds will be used in the following ways, from highest to lowest priority:

1. Travel grants for participants from minority demographics including first-time attendees.
2. Post-workshop social gathering at a nearby restaurant.
3. Additional awards for top reviewers and best contributors.

5 Organizers

We believe that our organizing team is well-suited to execute this workshop, owing to the long-standing involvement of our members with the community and their track record of putting together large-scale events. Furthermore, the composition of our team provides a diverse set of perspectives from many institutions, which is also important for handling any conflicts of interest that arise. The main contact of the workshop is Tara.

Tara Akhound-Sadegh* (tara.akhoundsadegh@mila.quebec) is a PhD student at McGill University and Mila. She is interested in Sampling and Generative Modelling, particularly with applications in physics and biology, as well as Geometric Deep Learning (symmetries and equivariant models). Previously, she was a research scientist in residence at DreamFold. She is also a lead organizer of the Geometric Deep Learning Reading Group at Mila and has experience in organizing events as a member of the EDI committee at Mila and as a previous volunteer member of the non-profit organization initiasciences.

Marta Skreta is a PhD student at the University of Toronto and Vector Institute working with Alán Aspuru-Guzik under the Canada Graduate Scholarship. Her research focuses on molecular discovery using generative modelling and natural language, as well as automating chemistry experiments in self-driving labs. Previously, Marta completed internships at Apple and Mila AI for Humanity. She has experience organizing large-scale events, including coding workshops for 1000+ Canadian high school women (HER CODE CAMP). Marta is a founding member of the AI4Materials workshop and has organized it at NeurIPS in 2022, 2023, and 2024 (upcoming).

Yuanqi Du is a PhD student at Cornell. His research focuses on probabilistic models and geometric deep learning for physical sciences and sustainability. He has been the lead organizer and (co)founder for a series of events related to probabilistic machine learning (DGM4HSD-ICLR22, SPIGM-ICML23, SPIGM-ICML24 workshops), AI for Science (workshops, online seminars and symposiums), and geometric deep learning (Learning on Graphs conference and local meetups).

Sarthak Mittal is a PhD student at Université de Montréal and Mila and a student researcher at Meta. His prior research focused on neuro-inspired deep learning architectures and the investigation of inductive biases to enhance generalization capabilities. Currently, he is exploring generative models, including diffusion and flow matching, and their applications in Bayesian posterior estimation to facilitate efficient and scalable inference.

Joey Bose is a Post-Doctoral fellow at the University of Oxford, a Distinguished Research Scientist at Dreamfold, and a Mila Affiliate member and holds a PhD in Computer Science from McGill/Mila. His research interests span Generative Modelling, Differential Geometry for Machine Learning, and Equivariant Machine Learning with a current emphasis on the foundations of geometry-aware generative models. In addition, he was the lead organizer for the Differential Geometry meets Deep Learning (DiffGeo4DL) workshop at NeurIPS2020 and also the primary instructor for a course on Geometric Generative Models given at McGill University/Mila in the Fall of 2022.

Alex Tong is an incoming assistant professor at Duke University, a Post-Doctoral fellow at Université de Montréal, cofounder and CTO of Dreamfold, and holds a PhD in computer science from Yale University. He co-organized a Banff workshop on "Deep Exploration of non-Euclidean Data with Geometric and Topological Representation Learning" in 2022.

Kirill Neklyudov is an Assistant Professor at the University of Montreal and a Core Academic Member at Mila - Quebec AI Institute. His research focuses on developing novel methods in generative modelling, Monte Carlo methods, Optimal Transport, and applications of these methods to fundamental problems in natural sciences, e.g. finding eigenstates of the many-body Schrodinger equation, simulating molecular dynamics, predicting the development of biological cells, conformational sampling, and protein folding. His community service includes organizing the SPIGM-ICML24 workshop and serving as an area chair at ICLR 2025.

Max Welling is a full professor and research chair in machine learning at the University of Amsterdam and a Merkin distinguished visiting professor at Caltech. He is co-founder and CAIO of the startup CuspAI in Materials Design. He is a fellow at the Canadian Institute for Advanced Research (CIFAR) and the European Lab for Learning and Intelligent Systems (ELLIS) where he served on the founding board. His previous appointments include Partner and VP at Microsoft Research, VP at Qualcomm Technologies, and professor at UC Irvine. Max Welling has served as associate editor-in-chief of IEEE TPAMI from 2011-2015, and on the advisory board of the Neurips Foundation since 2015. He is co-founder of the ELLIS and served on its board until 2021; he was the program chair and general chair of Neurips in 2013 and 2014 respectively. He was also program chair of AISTATS in 2009 and ECCV in 2016 and general chair and cofounder of MIDL 2018. Max Welling is the recipient of the ECCV Koenderink Prize in 2010, and the 10-year Test of Time awards at ICML in 2021 and ICLR in 2024.

Michael Bronstein is the DeepMind Professor of AI at the University of Oxford and Scientific Director at Aithyra. He was previously Head of Graph Learning Research at Twitter, a professor at Imperial College London and held visiting appointments at Stanford, MIT, and Harvard. He has been affiliated with three Institutes for Advanced Study (at TUM as a Rudolf Diesel Fellow (2017-2019), at Harvard as a Radcliffe fellow (2017-2018), and at Princeton as a short-time scholar (2020)). Michael received his PhD from the Technion in 2007. He is the recipient of the EPSRC Turing AI World Leading Research Fellowship, Royal Society Wolfson Research Merit Award, Royal Academy of Engineering Silver Medal, five ERC grants, two Google Faculty Research Awards, and two Amazon AWS ML Research Awards. He is a Member of the Academia Europaea, a Fellow of IEEE, IAPR, BCS, and ELLIS, ACM Distinguished Speaker, and a World Economic Forum Young Scientist. In addition to his academic career, Michael is a serial entrepreneur and founder of multiple startup companies, including Novafora, Invision (acquired by Intel in 2012), Videocites, and Fabula AI (acquired by Twitter in 2019).

Arnaud Doucet is a Senior Staff Research Scientist at Google DeepMind and a professor of statistics at the University of Oxford. He has also held academic positions at Cambridge University, Mel-

bourne University, The Institute of Statistical Mathematics in Tokyo and the University of British Columbia where her was a Canada Research Chair in Stochastic Computation. He was also an Institute of Mathematical Statistics Medallion (IMS) Lecturer in 2016, was elected IMS Fellow in 2017 and was awarded the Guy Silver Medal from the Royal Statistical Society in 2020.

Aapo Hyvarinen studied undergraduate mathematics at the universities of Helsinki (Finland), Vienna (Austria), and Paris (France), and obtained a Ph.D. degree in Information Science at the Helsinki University of Technology in 1997. After post-doctoral work at the Helsinki University of Technology, he moved to the University of Helsinki, where he was appointed Professor in 2008, at the Department of Computer Science. From 2016 to 2019, he was a Professor of Machine Learning at the Gatsby Computational Neuroscience Unit, University College London, UK. Aapo Hyvarinen is the main author of the books "Independent Component Analysis" (2001), "Natural Image Statistics" (2009), and "Painful Intelligence" (2022). He is Action Editor at the Journal of Machine Learning Research and Neural Computation and has worked as Area Chair at ICML, ICLR, AISTATS, UAI, ACML and NeurIPS.

6 Previous Related Workshops and Points of Difference

The first group of previous workshops related to us includes diffusion models[1], SPIGM[2] and LCDS[3]. However, each of them covers specific topics under probabilistic inference. For example, the workshop on diffusion models specifically discusses one class of generative models: diffusion models; SPIGM highlights the use of structure in a wide range of probabilistic inference problems; and LCDS focuses on the dynamical system and control perspective of probabilistic inference. However, the main focus of our workshop is on the sampling problem, where no data is available, and we only have access to an unnormalized density function.

As probabilistic inference problems broadly appear in scientific applications, another group of previous workshops related to us is the general AI for Science workshop[4] and the individual workshops that focus on the intersection of AI and different areas of sciences, e.g. ML4PS[5], MLSB[6], AI4Mat[7], etc. However, none of the workshops focus specifically on modern approaches to the sampling problem and the synergies between science and sampling are not widely discussed (e.g. sampling and molecular dynamics simulation).

7 Program Committee

The program committee members will help review the workshop submissions. We aim to have 3 reviewers per submission and assign a maximum of 3 papers to each of the reviewers. We already have a list of more than 100 researchers who have tentatively confirmed to be reviewers. Each of the organizers has committed to recruiting more reviewers, so we have a more than adequate pool of reviewers based on our estimation of likely submissions. Given that the organizers are from various labs and have different connections, we expect to have a rich and diverse pool of reviewers for the workshop.

The tentative list of the workshop's program committee is:

Vikash Sehwag, Maxence Noble, Yeongmin Kim, Austin J Stromme, Tanya Veeravalli, Tao Hu, Changwen Xu, Ajil Jalal, Yulai Zhao, Weitao Du, Chen Xu, Nikita Durasov, Sikun Yang, Rumen Dangovski, Siddharth Mishra-Sharma, Manan Gandhi, Hengyi Wang, Viktor Nilsson, Jonas Köhler, Tejas Jayashankar, Ilia Igashov, Kirty Vedula, Xinyu Yuan, Fanwang Meng, Wenhao Chai, Zhuoran Qiao, Peter Y. Lu, Xiaolin Hu, James Thornton, Ali Siahkoohi, Smita Krishnaswamy, Francisco Vargas, Wenhao Gao, Cong Fu, Mohammadreza Mofayezi, Li Du, Xinyao Zhang, Zirui Yan, Linlin Yu, Sungsoo Ahn, Yinkai Wang, Augustinos D Saravanos, Sanketh Vedula, Daniel Levy, Dimitra Maoutsa, Yash Pote, Alexandra Volokhova, Florentin Coeurdoux, Siddarth Venkatraman,

[1] https://diffusionworkshop.github.io/
[2] https://spigmworkshop2024.github.io/
[3] https://frontiers4lcd.github.io/
[4] https://ai4sciencecommunity.github.io/
[5] https://ml4physicalsciences.github.io/2024/
[6] https://www.mlsb.io/
[7] https://sites.google.com/view/ai4mat/home

Sahil Goyal, Yiqiao Jin, Guillaume Huguet, Aneesh Komanduri, Xiaoxiao He, Kulin Shah, Ya-Ping Hsieh, Ziyi Wang, Andrew K Tan, Moksh Jain, Austin Tripp, Ryan-Rhys Griffiths, Qunxi Zhu, Alexandre Adam, Manish Prajapat, Longkun Xu, Aram-Alexandre Pooladian, Justin T Chiu, Lingkai Kong, Karn Tiwari, Qi Li, Song Wei, Alex Morehead, Hao-Lun Hsu, Alexander V Nikitin, Asic Q Chen, Bolian Li, Xi Chen, Minsu Kim, Samuel Power, Gautheir Gidel, Jiecong Lin, Xianjun Yang, Haowei Lin, Wujie Wang, Pierre-Andre Noel, Chao-Han Huck Yang, Haoxuan Chen, Diego Mesquita, Théo Uscidda, Oswin So, Artur Bekasov, Bhavya Sukhija, Reza Akbarian Bafghi, Ziqing Yang, Shenao Zhang, Kyurae Kim, Alex Oshin, Zijing Ou, Bruce Wingo, Junyu Xuan, Titas Anciukevičius, Hehuan Ma, Wenbo Chen, Jarrid Rector-Brooks, Mohsin Hasan, Marcin Sendera, Pablo Lemos.

REFERENCES

Alexandre Adam, Adam Coogan, Nikolay Malkin, Ronan Legin, Laurence Perreault-Levasseur, Yashar Hezaveh, and Yoshua Bengio. Posterior samples of source galaxies in strong gravitational lenses with score-based priors. 2022.

Akashnathan Aranganathan, Xinyu Gu, Dedi Wang, Bodhi Vani, and Pratyush Tiwary. Modeling boltzmann weighted structural ensembles of proteins using ai based methods. *ChemRxiv*, Sep 2024. URL https://chemrxiv.org/engage/chemrxiv/article-details/66edf5f412ff75c3a1f638b9.

Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *Journal of political economy*, 81(3):637–654, 1973.

Charles Blundell, Julien Cornebise, Koray Kavukcuoglu, and Daan Wierstra. Weight uncertainty in neural network. In *International conference on machine learning*, pp. 1613–1622. PMLR, 2015.

Sitan Chen, Sinho Chewi, Jerry Li, Yuanzhi Li, Adil Salim, and Anru Zhang. Sampling is as easy as learning the score: theory for diffusion models with minimal data assumptions. In *The Eleventh International Conference on Learning Representations*.

Tianqi Chen, Emily Fox, and Carlos Guestrin. Stochastic gradient hamiltonian monte carlo. In *International conference on machine learning*, pp. 1683–1691. PMLR, 2014.

Volker L Deringer, Miguel A Caro, and Gábor Csányi. Machine learning interatomic potentials as emerging tools for materials science. *Advanced Materials*, 31(46):1902765, 2019.

Carles Domingo-Enrich, Michal Drozdzal, Brian Karrer, and Ricky T. Q. Chen. Adjoint matching: Fine-tuning flow and diffusion generative models with memoryless stochastic optimal control. 2024a. URL https://arxiv.org/abs/2409.08861.

Carles Domingo-Enrich, Michal Drozdzal, Brian Karrer, and Ricky TQ Chen. Adjoint matching: Fine-tuning flow and diffusion generative models with memoryless stochastic optimal control. *arXiv preprint arXiv:2409.08861*, 2024b.

Arnaud Doucet, Nando De Freitas, and Neil Gordon. An introduction to sequential monte carlo methods. *Sequential Monte Carlo methods in practice*, pp. 3–14, 2001.

Yilun Du, Conor Durkan, Robin Strudel, Joshua B. Tenenbaum, Sander Dieleman, Rob Fergus, Jascha Sohl-Dickstein, Arnaud Doucet, and Will Grathwohl. Reduce, reuse, recycle: Compositional generation with energy-based diffusion models and mcmc. 2024.

Chenru Duan, Yuanqi Du, Haojun Jia, and Heather J Kulik. Accurate transition state generation with an object-aware equivariant elementary reaction diffusion model. *Nature Computational Science*, 3(12):1045–1055, 2023.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.

Louis Grenioux, Maxence Noble, Marylou Gabrié, and Alain Oliviero Durmus. Stochastic localization via iterative posterior sampling. 2024.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. 2020.

Lars Holdijk, Yuanqi Du, Ferry Hooft, Priyank Jaini, Berend Ensing, and Max Welling. Stochastic optimal control for collective variable free sampling of molecular transition paths. *Advances in Neural Information Processing Systems*, 36, 2023.

Xunpeng Huang, Hanze Dong, Yifan Hao, Yi-An Ma, and Tong Zhang. Reverse diffusion Monte Carlo. *International Conference on Learning Representations*, 2024.

Pavel Izmailov, Sharad Vikram, Matthew D Hoffman, and Andrew Gordon Gordon Wilson. What are bayesian neural network posteriors really like? In *International conference on machine learning*, pp. 4629–4640. PMLR, 2021.

Diederik P Kingma. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

Axel Levy, Eric R. Chan, Sara Fridovich-Keil, Frédéric Poitevin, Ellen D. Zhong, and Gordon Wetzstein. Solving inverse problems in protein space using diffusion-based priors. 2024.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, 2023.

Laurence Illing Midgley, Vincent Stimper, Gregor NC Simm, Bernhard Schölkopf, and José Miguel Hernández-Lobato. Flow annealed importance sampling bootstrap. *International Conference on Learning Representations (ICLR)*, 2023.

Frank Noé, Simon Olsson, Jonas Köhler, and Hao Wu. Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science*, 365(6457):eaaw1147, 2019.

Lars Onsager and Stefan Machlup. Fluctuations and irreversible processes. *Physical Review*, 91(6): 1505, 1953.

Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic backpropagation and variational inference in deep latent gaussian models. In *International conference on machine learning*, volume 2, pp. 2, 2014.

Lorenz Richter, Julius Berner, and Guan-Horng Liu. Improved sampling via learned diffusions. *arXiv preprint arXiv:2307.01198*, 2023.

Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. 2015.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021.

Alexander Tong, Kilian FATRAS, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. Expert Certification.

Francisco Vargas, Will Grathwohl, and Arnaud Doucet. Denoising diffusion samplers. *International Conference on Learning Representations (ICLR)*, 2023.

Francisco Vargas, Shreyas Padhy, Denis Blessing, and Nikolas Nüsken. Transport meets variational inference: Controlled monte carlo diffusions. In *The Twelfth International Conference on Learning Representations*, 2024.

Siddarth Venkatraman, Moksh Jain, Luca Scimeca, Minsu Kim, Marcin Sendera, Mohsin Hasan, Luke Rowe, Sarthak Mittal, Pablo Lemos, Emmanuel Bengio, Alexandre Adam, Jarrid Rector-Brooks, Yoshua Bengio, Glen Berseth, and Nikolay Malkin. Amortizing intractable inference in diffusion models for vision, language, and control. 2024.

Joseph L Watson, David Juergens, Nathaniel R Bennett, Brian L Trippe, Jason Yim, Helen E Eisenach, Woody Ahern, Andrew J Borst, Robert J Ragotte, Lukas F Milles, et al. De novo design of protein structure and function with rfdiffusion. *Nature*, 620(7976):1089–1100, 2023.

Max Welling and Yee W Teh. Bayesian learning via stochastic gradient langevin dynamics. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pp. 681–688. Citeseer, 2011.

Qinsheng Zhang and Yongxin Chen. Path integral sampler: a stochastic control approach for sampling. *International Conference on Learning Representations (ICLR)*, 2022.

Stephen Zhao, Rob Brekelmans, Alireza Makhzani, and Roger Baker Grosse. Probabilistic inference in language models via twisted sequential monte carlo. In *Forty-first International Conference on Machine Learning*, 2024.