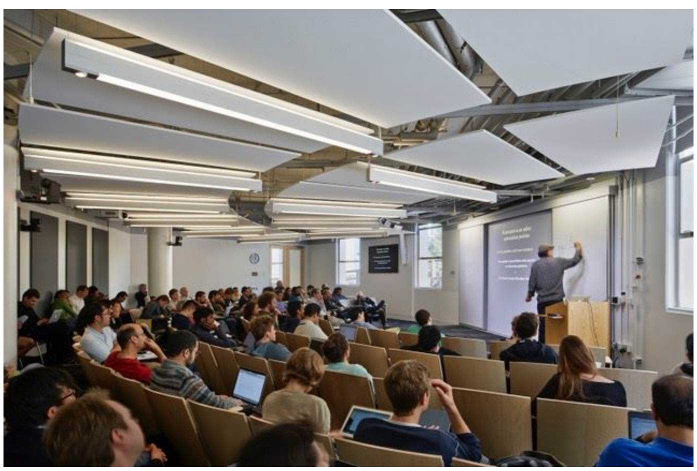
Simons Institute for the Theory of Computing

6th Annual Industry Day

Thursday, November 5, 2020



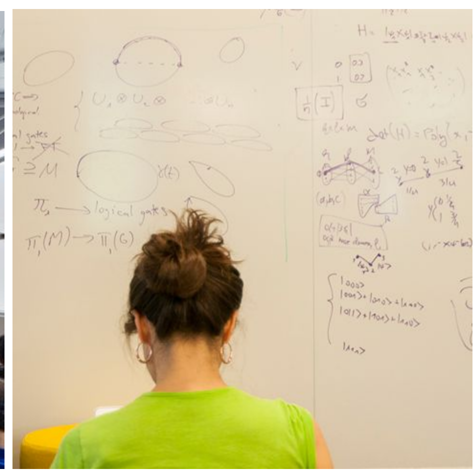


Table of Contents

Statistical Analysis of Privacy Protected Data, *Facebook*

Welcome

Dear friends,

It is with sincere enthusiasm and gratitude that we welcome you to the 6th Annual Simons Institute Industry Day.

So much has changed since our last industry day and through it all our community of scientists and scholars has remained steadily involved in our collaborative research platform. We are encouraged by tremendously high participation in our virtual workshops and other activities this year, with some events welcoming five times the average number of attendees. This aligns with our mission to enhance public understanding of the increasing prominence of algorithms in every aspect of human life and society.

Today's event is once again a chance for our research fellows and industry scientists to take the stage, to share their work, and to uncover new connections and synergies in research.

Our industry partnerships are vital to the relevance of our research. They represent a variety of industries including computing software and hardware, social media, computing security, financial services, risk and reinsurance. They often provide a sounding board for thoughts and reveal new perspectives and directions for research. The theory of computing is critical and ever changing as the demand grows for technology at scale. We applaud our industry partners for supporting and being involved closely with core science. We see our shared pursuits in research as responsible actions. As the world becomes more complex and in some cases more fragile, together we can improve the human condition through joint discovery.

We have had much to celebrate this year, with the announcement of a second generous grant from the Simons Foundation and several groundbreaking grants from the National Science Foundation in the areas of Data Science, Deep Learning and Quantum. These, along with the support we receive from our industry partners, and private individuals and foundations, ensure for the Institute an active and productive second decade of research, discovery and connections.

To all, thank you for your visionary partnership, and your dedication to this important work.

Yours,

Shafi Goldwasser, Director
Peter Bartlett, Associate Director
Prasad Raghavendra, Senior Scientist

Return to Top

Industry Day Agenda

8:45 a.m.	Host asks participants to add their affiliations next to their names in Zoom
9:00 a.m.	Introduction - (Shafi Goldwasser and Peter Bartlett) 5 mins
9:05 a.m.	Program presentations (6 presentations @ 6 mins/each =) 36-40 mins Satisfiability: Theory, Practice, and Beyond: *Jan. 12–May 14, 2021* *Antonina Kolokolova* Theoretical Foundations of Computer Systems: *Jan. 12–May 14, 2021* *Moshe Vardi* Computational Innovation and Data-Driven Biology: *July 6–Aug. 6, 2021* *Ron Shamir* Interpretable Machine Learning: *June 28–Aug. 6, 2021* *Shai Ben-David* Computational Complexity of Statistical InferenceI: *Aug. 18–Dec. 17, 2021* *Guy Bresler* Causality: *Jan. 12–May 14, 2022* *Frederick Eberhardt*
9:45 a.m.	Break - 7 mins
9:52 a.m.	Lightning talks research fellows (13 @ 4 mins/each =) 48-60 mins
10:42 a.m.	"Meet the Research Fellows" - 15 mins (curated icebreaker via Zoom) 2-3 RF are assigned to each room, participants choose rooms in real time Room 1: Jalaj Bhandari, Zhuoran Yang Room 2: Ashwin Pananjady, Mohamad Kazem Shirani Faradonbeh Room 3: Ahmed El Alaoui, Galyna Livshyts Room 4: Vidya Muthukumar, Christina Yu Room 5: Zhaoran Wang, Lin Yang Room 6: Michael Kim, Erik Waingarten Room 7: Lin Chen, Aditya Grover
10:57 a.m.	Lightning talks industry partners - Part 1 (4 @ 10 mins/each =) 40 mins VMware, *Parikshit Gopalan* NTT, *Kazuhiro Gomi* Swiss Re, *Jeffrey Bohn* Apple, *Kunal Talwar*
11:37 a.m.	Brief tutorial for Gather.Town and then networking break-out - 15 mins (casual via Gather.Town)
11:52 a.m.	Lightning talks industry partners - Part 2 (4 @ 10 mins/each =) 40 mins JP Morgan, *Sumitra Ganesh* Google, *Guru Guruganesh* Microsoft Research, *Sébastien Bubeck* Novi, *Alberto Sonnino - 5 mins*

Facebook, Runchao Jiang - *5 mins*

12: 32 p.m. Wrap-up - (Shafi) 3 mins

12:35 p.m. Final networking break-out - 10 mins (casual via Gather.Town)

Return to Top

Preparation for Industry Day 2020

Connecting

Zoom information: https://berkeley.zoom.us/j/99066144668?pwd=Qld1c0RQZTBvS3BHSDFaM0hCMVdhUT09
Password: 494019
Gather.Town information: https://gather.town/app/CNswn3WtnQqQth8h/industrydayconference

Tech Support - day of event: Christine Wang, Science & Outreach Administrator
voice or text: +1.562.716.3922 or email: chris24@berkeley.edu

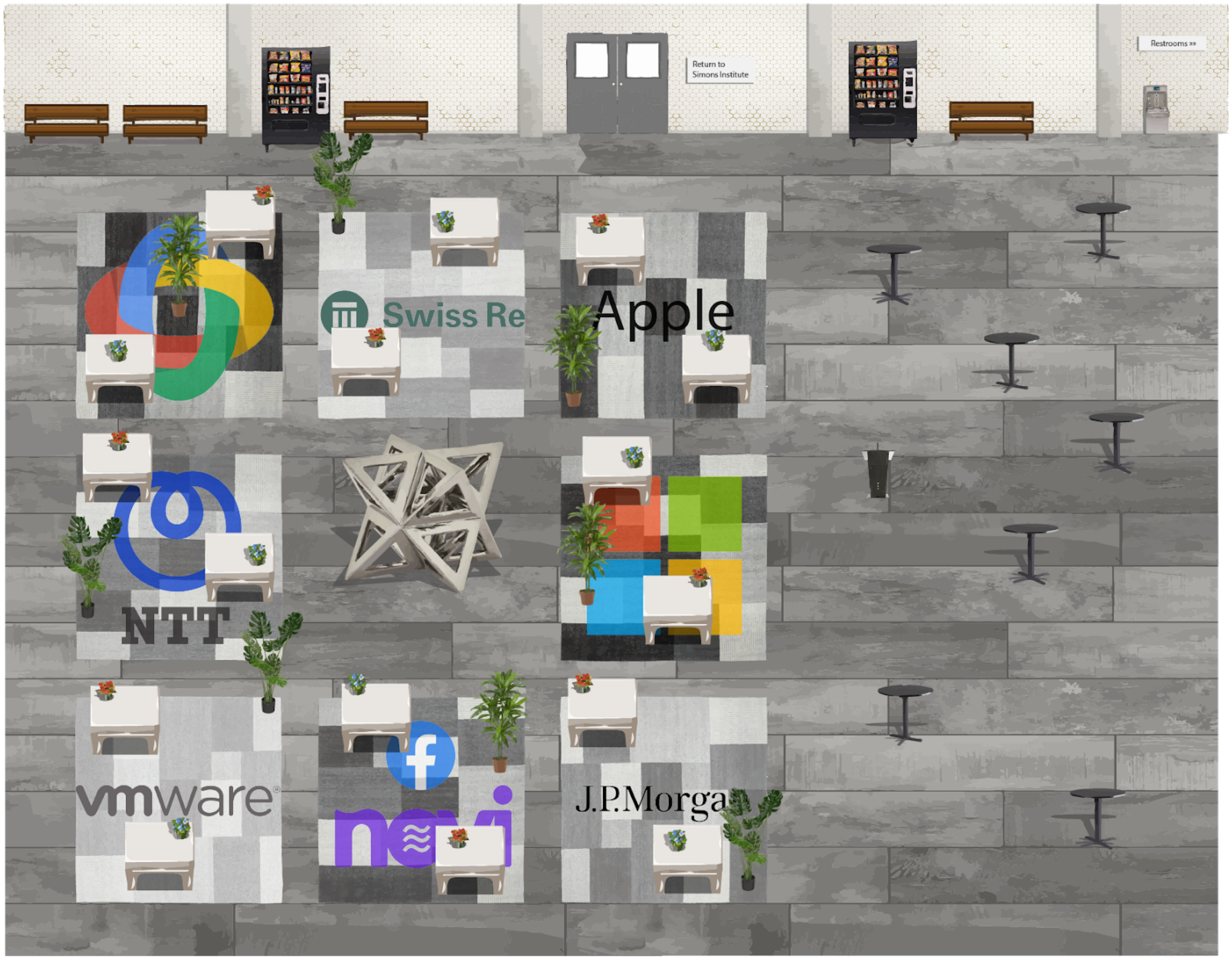
*** PLEASE UPGRADE to Zoom 5.3.0. or higher before the event. ***

Conference Format

The main event will be via Zoom meeting to allow participants to engage fully in the activities. Participants are asked to remain muted with their videos on during talks. The chat feature will be unrestricted between speakers and participants.

The networking portion of the event will be as follows.

1) *Meet the Research Fellows* via Zoom meeting rooms. We will have in place meeting rooms with 2-3 research fellows assigned. Participants will be able to select which room to enter and be free to move in and out of meeting rooms during the break.
2) *Informal Networking* (two sessions) via Gather.Town. Gather is a virtual room, which in this case will be set up as a large conferencing space with tables logoed for each industry partner, as shown in the screenshot below. Participants will be able to navigate the space as avatars, moving next to people with whom they wish to speak.



Emcee

Peter Bartlett, Simons Institute Associate Director, will run the show and manage the talks by program organizers and industry partners.

Moderator

Prasad Raghavendra, Simons Institute Senior Scientist, will manage the talks by research fellows.

Speakers

- **For connectivity tech support, Christine Wang - voice or text: 562.716.3922 or email: chris24@berkeley.edu**
- Slides and other visuals:
  - Program Organizers and Industry Partners will **manage their own visuals** and share their screens when they are speaking.
  - Research Fellows' visuals will be managed by the administrator, whom they will instruct to progress slides.
- Speakers will be allocated a precise amount of time:
  - Program Organizers - 6 minutes
  - Research Fellows - 4 minutes
  - Industry Partners - 10 minutes *(except Novi+Facebook: two speakers @ 5 mins each)*
- Keeping time:
  - Program Organizers and Industry Partners will manage their own time, and hear "*one-minute remaining*" and

    "*time's up*" from our time-keeper, Raquel Romero.
    - Research Fellows will have a stopwatch counter on their screen.
- We suggest speakers **leave the final minute** for questions from the floor.
- **Questions will be fielded** by Peter Bartlett for program organizers and industry partners, and by Prasad Raghavendra for research fellows.
- If your **connection is faulty**, we may move down the queue of speakers until you are ready.
- Speakers may **sign up for a technical run-through** with Drew Mason (dmason@berkeley.edu) either Monday, November 2, Tuesday, November 3 or Wednesday November 4 @ 8:30am-10:00am. Please contact him directly to choose your time.

Recording Talks

- Speakers who have signed release forms will be recorded, and their talks will be available after the event on the Simons Institute YouTube channel. If you would like to be recorded but have not yet received a release form, please contact Amy Ambrose amyambrose@berkeley.edu

Managing Q & A

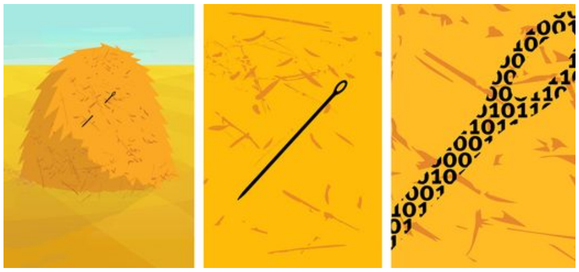
- Participants are asked to **remain muted during the talks**, and when called on, to unmute to ask their question.
- Participants may use the **raise hand function** or pose questions directly in the **chat**.
- At the end of each talk, Peter or Prasad will either call on one raised hand or select a question from the chat and ask the participant to **unmute to speak**.
- Participants are encouraged to continue asking questions in the networking sessions.

Points of Contact – Before & During the Event

Christine Wang Science & Outreach Admin chris24@berkeley.edu +1.562.716.3922	Drew Mason Information Systems Analyst dmason@berkeley.edu	Jesse Gil Associate Event Coordinator simonsevents@berkeley.edu +1.714.654.3247	Amy Ambrose Sr. Development Director amyambrose@berkeley.edu +1.510.944.6674

Return to Top

Program Presentations



Satisfiability: Theory, Practice, and Beyond
Jan. 12–May 14, 2021

https://simons.berkeley.edu/programs/sat2021

In an age of ubiquitous computing, computational complexity theory is the science that studies what problems can be efficiently solved by computation. Since the founding work of the 1970s, an influential line of research has zoomed in on NP-complete problems, with the satisfiability problem for Boolean logic formulas (SAT) at its head, which turned out to be exactly the right notion to capture literally thousands of important applied problems in different fields. Based on the assumption of the hardness of NP, the validity of which is one of the famous *Millennium Prize Problems*, a rich mathematical theory has been developed for establishing conditional results that state that all these problems are infeasible to solve, in the worst case.

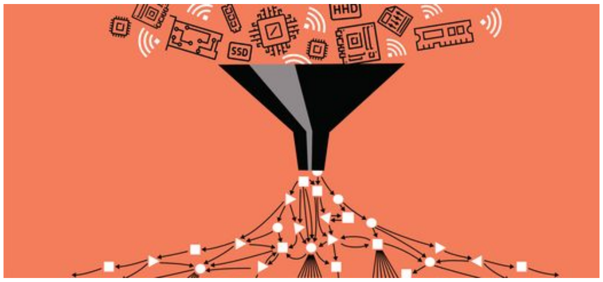
The trouble is that real problems are not worst-case. The last two decades have seen the development of exceedingly efficient algorithms for many of these problems, perhaps most impressively in the form of so-called SAT solvers for logic formulas. Traditional complexity analysis claiming exponential lower bounds is arguably not very relevant in this setting, but this also means that we lack tools to understand how these algorithms can perform so well and why they sometimes spectacularly fail.

This program will bring together leading theoreticians and practitioners that work on SAT and its generalizations, and approach it from all possible angles: complexity theory, logic, computational algebra, optimization, SAT solving, constraint programming, random instances, SAT modulo theories (SMT), etc. The goal is to develop a better mathematical understanding of real-world efficient computation, and to work towards further algorithmic progress on problems that are currently beyond reach. So far, theoretical and applied research in these areas has developed mostly separately, but several joint workshops in recent years showed that the communities are eager for more interaction.

This program is an opportunity for a long-term collaboration and exchange of ideas between theoreticians and practitioners, that has potential for significant long-term impact in mathematics, computer science, and industry. This semester program could provide a decisive contribution in building a joint community of researchers working on constructing efficient algorithms and analyzing the computational complexity of applied problems.

Antonina Kolokolova is associate professor in computer science at the Memorial University of Newfoundland in Canada. She obtained her PhD in 2005 from the University of Toronto, under the supervision of Stephen Cook. Before taking her current position, Antonina was a postdoctoral researcher at the Mathematical Institute of the Academy of Sciences in the Czech Republic, and at Simon Fraser University in Canada. Her main research interests are in theoretical computer science, in particular, computational complexity theory, mathematical logic and proof complexity. Antonina is also interested in the connections between theory and neuroscience.

Return to Top



Theoretical Foundations of Computer Systems
Jan. 12–May 14, 2021

https://simons.berkeley.edu/programs/tfcs2021

This program aims at the development of the theoretical foundations of computer systems (TFCS). This field of research was intensively developed over the last three decades, yielding major improvements in model checking techniques as well as in satisfiability solving. The major challenge in the field is the need for scalability. The program aims at bringing together leading researchers on these themes. TFCS is collocated with a closely related program on satisfiability (SAT). The program will focus on the following aspects:

New developments in logic: Logic is often used in TFCS as a specification formalism, describing in a formal and rigorous way the requirements that a system under design or under verification is expected to satisfy.

New developments in automata: Automata are used in TFCS both as a modeling formalism, for example, transducers are used to model reactive systems, and as reasoning tools, as in the automata-theoretic approach to temporal model checking.

Probabilistic modeling in the analysis of systems: Probabilistic methods are used in the modeling and analysis of systems that exhibit probabilistic behavior, from randomized algorithms to biological systems.

The use of games and their equilibria: Games are used in TFCS both as an algorithmic construct, for example, in the usage of alternating automata in temporal model checking, and as a modeling construct, for example, in the design of reactive systems it is convenient to consider the setting as a game between the system and its environment. Once multi-agent systems are considered, equilibria enter in a natural way.

Techniques for the analysis of cyber-physical systems: Many features of real systems, for example time and energy usage, are quantitative, and to express these one needs continuous, real-valued functions and suitable hybrid combinations of discrete and quantitative constructs.

Moshe Y. Vardi is university professor, Karen Ostrum George Distinguished Service Professor in Computational Engineering at Rice University, where he is leading the university's Initiative on Technology, Culture, and Society. His interests focus on automated reasoning, a branch of artificial intelligence with broad applications to computer science, including machine learning, database theory, computational-complexity theory, knowledge in multi-agent systems, computer-aided verification, and teaching logic across the curriculum. He is also a faculty scholar at the Baker Institute for Public Policy at Rice University.

Return to Top

Computational Innovation and Data-Driven Biology

July 6–Aug. 6, 2021

https://simons.berkeley.edu/programs/bio2021

The rapid growth of biological and medical data offers challenges and opportunities. It can lead to faster discoveries and deeper understanding of biological systems and disease, but it requires novel advanced algorithms, statistics, and mathematics to make sense of the data. The goal of this summer program is to bring together researchers in bioinformatics — both computational method developers and experimentalists — to discuss key challenges and collaborate on approaches for their solution.

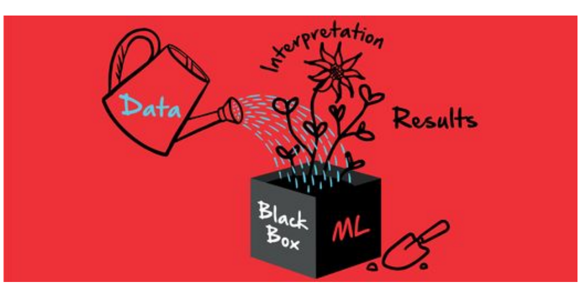
This five-week summer program is part of the Koret-UC Berkeley-Tel Aviv University (TAU) Initiative in Computational Biology and Bioinformatics. The program will bring together a broad international community of researchers, including UC Berkeley and TAU researchers.

In addition to ongoing activities throughout the program, there will be two four-day workshops bringing together a broader group of researchers for a shorter period. The two workshops will aim to attract both computational and applied scientists in bioinformatics. In this way, each will have both a theoretical and algorithm development component and a methodological, biotechnology, and data analysis component.

This program is supported in part by the Koret Foundation.

Ron Shamir is a professor of computer science and head of the Edmond J. Safra Center for Bioinformatics at Tel Aviv University. He develops methods and tools in computational genomics, specializing in gene expression, gene regulation, molecular networks and disease analysis. Software tools developed by Shamir's group are in use by hundreds of laboratories around the world. Shamir was among the founders of bioinformatics both internationally (RECOMB) and in Israel (ISBCB and IBS). He received the Michael Landau Prize in Bioinformatics and is a fellow of ISCB and ACM.

Return to Top



Summer Cluster: Interpretable Machine Learning

June 28–Aug. 6, 2021

https://simons.berkeley.edu/programs/iml2020

This cluster will convene an interdisciplinary group of scholars to develop firm theoretical and philosophical foundations for addressing some major issues concerning interpretability of machine learning-based models. Program participants include experts in theoretical computer science, machine learning, statistics, causal inference, and fairness and the present community on interpretability. We aim to address the following fundamental questions about how to best use machine learning for real-life tasks:

1. What notions of interpretability might render models that are more amenable to monitoring by regulators?
2. How can the quality and usefulness of interpretation in a given context (such as a particular human audience and for a particular domain problem) be evaluated both empirically and theoretically?
3. For which of the desiderata that interpretability purports to address must we sacrifice predictive accuracy?
4. Are there any feasibly measurable properties of neural networks that can yield significant insights into their input-output functionality? More generally, are there sound theoretical principles under which today's deep learning tools can be leveraged to confer insights beyond their predictive accuracy?
5. What role, if any, do various interpretation or explanation techniques have to offer the discourse on algorithmic fairness and discrimination? Are there any inherent trade-offs between notions of interpretability and fairness?

The cluster will address a variety of perspectives on defining and developing tools for achieving these goals in automated decision-making systems.

Shai Ben-David is professor in computer science at the University of Waterloo. He holds a PhD in mathematics from the Hebrew University in Jerusalem. He has held postdoctoral positions at the University of Toronto in both the Mathematics and Computer Science departments. He was a professor of computer science at Technion - Institute of Technology in Haifa, Israel. Ben-David has held visiting positions at the Australian National University and Cornell University, and since 2004 has been a professor of computer science at the University of Waterloo in Canada. His research interests are in CS theory and machine learning. He has been program chair for COLT and ALT (the two main annual machine learning conferences), and also served on their steering committees. Ben-David has also been an area chair and senior program committee member for ICML and NIPS (the two major general machine learning conferences).

Return to Top

Computational Complexity of Statistical Inference

Aug. 18–Dec. 17, 2021

https://simons.berkeley.edu/programs/si2021

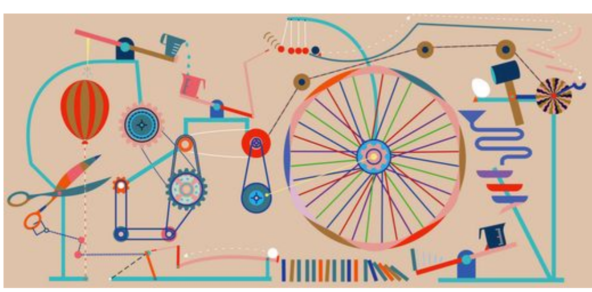
The two basic lines of inquiry in statistical inference have long been: (i) to determine fundamental statistical (i.e., information-theoretic) limits; and (ii) to find efficient algorithms achieving these limits. However, for many structured inference problems, it is not clear if statistical optimality is compatible with efficient computation. Statistically optimal estimators often entail an infeasible exhaustive search. Conversely, for many settings the computationally efficient algorithms we know are statistically suboptimal, requiring higher signal strength or more data than is information-theoretically necessary. This phenomenon is both fascinating and unsettling. It suggests that the information-theoretic limit on the signal-to-noise ratio (or the amount of data) for these problems, as studied since the beginning of mathematical statistics, is not the practically relevant benchmark for modem high-dimensional settings. Instead, the practically relevant benchmark is the fundamental statistical limit for *computationally efficient* algorithms.

By now dozens of fundamental high-dimensional statistical estimation problems are conjectured to have different computational and statistical limits. These problems (for example, sparse linear regression or sparse phase retrieval) are ubiquitous in practice and well-studied theoretically, yet the central mysteries remain: What are the fundamental data limits for computationally efficient algorithms? How do we find optimal efficient algorithms? At a more basic level, are these statistical-computational gaps in various problems appearing for a common reason? Is there hope of building a widely applicable theory describing and explaining statistical-computational trade-offs?

The objective of the program is to advance the methodology for reasoning about the computational complexity of statistical estimation. Over the last decade several disparate communities and lines of work have begun to make progress on these questions. This program aims to stimulate work towards developing a deeper understanding and building a coherent theory by forming new collaborations between researchers in complexity theory, algorithms, statistics, learning theory, probability, and information theory.

Guy Bresler is associate professor in the Department of Electrical Engineering and Computer Science at Massachusetts Institute of Technology, and a member of LIDS, Center for Statistics, and IDSS. Previously, he was a postdoc at MIT. Before that Guy received his PhD from the EECS Department at the University of California Berkeley. His undergraduate degree is from the University of Illinois at Urbana-Champaign. In the last several years, his research has focused on the interface between computation and statistics with the aim of understanding the relationship between combinatorial structure and computational tractability of high-dimensional inference.

Return to Top



Causality
Jan. 12–May 14, 2022
https://simons.berkeley.edu/programs/Causality2022

This program aims to integrate advances and techniques from theoretical computer science into methods for causal inference and discovery.

Although attempts to characterize causal relations can be found in some of the oldest written records, the history of the usage of causal concepts within scientific discussions over the past 100 years has been rocky, varying from the outright denial of any role of causality in mature scientific theories to a disingenuous usage of ambiguous terms that obscure the role of cause and effect (e.g., "link," "connection," etc.).

A substantive development of new formal approaches to causality in the 1970s and 1980s precipitated a change in attitude toward the scientific investigation of causal questions. The change was led by the development of two largely intertranslatable mathematical frameworks: the potential outcome framework and the causal graphical models framework. These frameworks integrated three concepts central to the notion of causation: (i) the connection between the underlying causal relations and observed data; (ii) the difference that interventions can make to a causal system; and (iii) counterfactual statements about a system. All of these aspects of causality play a central role in scientific testing and explanation, often constituting the goal of scientific inquiry itself.

The mathematization of questions of causality has resulted in the development of inference techniques and learning methods to infer causal relations from data. These formal approaches are now starting to spread throughout the applied sciences, where just about any field of study is seeing a renewed and explicit interest in tackling causality.

Broad application of these theoretical frameworks in scientific domains requires not only conceptual clarity and "in principle" methods, but a detailed understanding of how the methods behave in practice, how to scale and approximate the ideally desired computations, and how to optimize methods for the particular constraints present in a domain.

This program will bring together theoretical and applied researchers from a broad variety of domains with the goal of understanding the complexity, optimizations, and possible approximation regimes required to turn the methods of causal inference into a broadly applicable scientific toolbox.

Frederick Eberhardt is professor of philosophy in the Division of the Humanities and Social Sciences at the California Institute of Technology. Before coming to Caltech he was assistant professor in the philosophy-neuroscience-psychology (PNP) program and the Department of Philosophy at Washington University in St. Louis. In 2011 he had a two-year research leave to work on causal discovery methods at Carnegie Mellon University with a grant from the James S. McDonnell Foundation. Before going to St. Louis, he was a McDonnell postdoc at the Institute of Cognitive and Brain Sciences at the University of California Berkeley. Frederick completed his PhD in the Philosophy Department at Carnegie Mellon University. His research interests lie at the formal end of the philosophy of science, the machine learning end of statistics and computer science, and the learning and modeling end of psychology and cognitive science. His work has focused primarily on methods for causal discovery from statistical data, the use of experiments in causal discovery, the integration of causal inferences from different data sets, and the philosophical issues at the foundations of causality and probability. He has done some work on computational models in cognitive science and some historical work on the philosophy of Hans Reichenbach, especially his frequentist interpretation of probability.

Return to Top

Research Fellows

Global Optimality Guarantees for Policy Gradient Methods

Many recent successes in reinforcement learning have been driven by a class of algorithms called policy gradient methods, which apply to complex, poorly understood, control problems by performing stochastic gradient descent over a parameterized class of polices. Unfortunately, even for simple control problems solvable by standard dynamic programming techniques, policy gradient algorithms face non-convex optimization problems and are widely understood to converge only to a stationary point. We identify structural properties — shared by several classic control problems — which guarantee that policy gradient objective function has no suboptimal stationary points despite being non-convex. Strengthening these conditions shows how the policy gradient objective satisfies a gradient dominance condition yielding convergence rates.

Jalaj Bhandari
Swiss Re Research Fellow
Program: Theory of Reinforcement Learning
Dates of Visit: Aug. 19–Dec. 18, 2020

Jalaj is a PhD student in operations research at Columbia University, working with Professor Daniel Russo. His thesis work explores foundations of modern reinforcement learning (RL) algorithms using ideas from optimization theory. In the past, he has also done research work on Bayesian Machine learning methods, specifically in designing computationally efficient Markov Chain Monte Carlo (MCMC) algorithms for posterior sampling. He is excited about applying RL and machine learning methods to problems of practical interest, for example in the areas of healthcare, neuroscience, autonomous systems, personalized Ads and more. Prior to Columbia, Jalaj graduated from the Indian Institute of Technology Delhi in 2012 with a BTech in industrial engineering and operations research.

Return to Top

Deep Neural Tangent Kernel and Laplace Kernel Have the Same RKHS

We prove that the reproducing kernel Hilbert spaces (RKHS) of a deep neural tangent kernel and the Laplace kernel include the same set of functions, when both kernels are restricted to the sphere \$S^{d–1}\$. Additionally, we prove that the exponential power kernel with a smaller power (making the kernel more non-smooth) leads to a larger RKHS, when it is restricted to the sphere \$S^{d–1}\$ and when it is defined on the entire \$R^d\$.

Lin Chen
JP Morgan Research Fellow
Program: Theory of Reinforcement Learning
Dates of Visit: Aug. 19–Dec. 18, 2020

Lin Chen received his PhD from Yale University in 2020, under the supervision of Professor Amin Karbasi. He received his BS from Peking University in 2014. His research focuses on theoretical machine learning, including online optimization, submodular optimization, and adversarial robustness.

Return to Top

Optimization of Random Functions

Minimizing a non-convex function that depends on a large number of variables is a computationally challenging problem that has a large number of applications in science and engineering. While from a worst case perspective this task is intractable, such problems are often solved to near-optimality using a variety of heuristics.
Focusing on random functions — in which this hidden simplicity cannot be explained by any obvious structural property — El Alaoui will describe a new class of optimization algorithms, which are often able to find near optima.

Ahmed El Alaoui
RM Karp Research Fellow
Program: Probability, Geometry, and Computation in High Dimensions
Dates of Visit: Aug. 19–Dec. 18, 2020

Ahmed El Alaoui received his PhD in 2018 in electrical engineering and computer sciences from the University of California Berkeley, under the supervision of Michael I. Jordan. He was a postdoctoral researcher from Sept. 2018 to July 2020 at Stanford University, hosted by Andrea Montanari. His research interests revolve around high-dimensional statistics, probability theory, statistical physics, algorithms, and problems where these areas meet. El Alaoui will join Cornell University as an assistant professor in the Department of Statistics and Data Science in January 2021.

Return to Top

Fair Generative Modeling via Weak Supervision

Real-world datasets are often biased with respect to key demographic factors such as race and gender. Due to the latent nature of the underlying factors, detecting and mitigating bias is especially challenging for unsupervised machine learning. We present a weakly supervised algorithm for overcoming dataset bias for deep generative models. Our approach requires access to an additional small, unlabeled reference dataset as the supervision signal, thus sidestepping the need for explicit labels on the underlying bias factors. Using this supplementary dataset, we detect the bias in existing datasets via a density ratio technique and learn generative models that efficiently achieve the twin goals of: i) data efficiency by using training examples from both biased and reference datasets for learning; and ii) data generation close in distribution to the reference dataset at test time. Empirically, we demonstrate the efficacy of our approach, which reduces bias with respect to latent factors by an average of up to 34.6% more than baselines for comparable image generation using generative adversarial networks.

Aditya Grover
Research Scientist, Facebook AI Research
Google Research Fellow
Program: Probability, Geometry, and Computation in High Dimensions
Dates of Visit: Aug. 19–Dec. 18, 2020

Aditya Grover is a research scientist at Facebook AI Research, a visiting postdoctoral researcher at the University of California Berkeley, and an incoming assistant professor of computer science at UCLA (starting Fall 2021). His research focuses on probabilistic modeling for representation learning and reasoning in high dimensions, and is grounded in applications in science and sustainability, such as weather forecasting and electric batteries. Aditya's research has been published in top machine learning and scientific venues including *Nature*, covered by various media outlets, included in widely used open-source software, and deployed into production at major technology companies. He has won several awards, including a best paper award (StarAI); a best undergraduate thesis award; a Stanford Centennial Teaching Award; a Stanford Data Science Scholarship; a Lieberman Fellowship; and a Microsoft Research PhD Fellowship. Aditya received his PhD and masters in computer science from Stanford University in 2020 and his bachelor degree in computer science from the Indian Institute of Technology Delhi in 2015.

Return to Top

A Complexity-Theoretic Perspective on Fairness

Algorithms make predictions about people constantly. The spread of such prediction systems---from precision medicine to targeted advertising to predictive policing---has raised concerns that algorithms may perpetrate unfair discrimination, especially against individuals from minority groups. While it's easy to speculate on the risks of unfair prediction, devising an effective definition of algorithmic fairness is challenging. Most definitions tend towards one of two extremes---individual fairness notions provide theoretically-appealing protections but present practical challenges at scale, whereas group fairness notions are popular in practice but offer marginal protections. In this talk, we present a novel framework for approaching fairness, called multi-calibration, that strengthens the guarantees of group fairness while maintaining practical viability. Parameterized by a computational class (e.g., halfspaces, decision trees, neural networks), multi-calibration enforces a rich set of consistency constraints that can be audited within the computational class given a small set of data, setting a rigorous standard for what we should expect from a machine-learned predictor.

Michael P. Kim
Miller Postdoctoral Fellow, UC Berkeley

Michael P. Kim is a Miller Postdoctoral Fellow at UC Berkeley, working with Shafi Goldwasser. Prior to this, he completed his Ph.D. in the Stanford Theory Group under the guidance of Omer Reingold. Michael's research investigates foundational questions about responsible machine learning. Much of this work aims to (1) identify ways in which machine-learned predictors can exhibit unfair discrimination and (2) develop algorithmic tools that provably mitigate such forms of discrimination. More broadly, he is interested in how the computational lens (i.e. algorithms and complexity theory) can help tackle emerging societal and scientific challenges.

Return to Top

On Inhomogeneous Random Matrices

We shall discuss briefly the invertibility of inhomogeneous random matrices, and a discretization procedure that is helpful in this question.

Galyna Livshyts
Assistant Professor, Georgia Institute of Technology
Research Fellow
Program: Probability, Geometry, and Computation in High Dimensions
Dates of Visit: Aug. 19–Dec. 18, 2020

Galyna Livshyts completed her undergraduate studies in Kharkiv, Ukraine. She obtained her PhD from Kent State University in Ohio in 2015 under the supervision of Artem Zvavitch. Since 2015, Galyna has been an assistant professor at the School of Mathematics, Georgia Institute of Technology. In Fall 2017, she was a postdoc in the geometric asymptotic analysis and applications program at the Mathematical Sciences Research Institute in Berkeley. Galyna is interested in high-dimensional probability and convexity, as well as asymptotic analysis and random matrix theory.

Return to Top

Classification vs Regression in Overparameterized Models: Does the Loss Function Matter?

Recent years have seen substantial interest in a first-principles theoretical understanding of the behavior of overparameterized models that interpolate noisy training data, based on their surprising empirical success. Of particular interest are solutions with a minimum-norm inductive bias. Research in the last year has shown that minimum-norm interpolations in the overparameterized linear model can generalize well for the least-squares regression problem as long as the distribution of features navigates a delicate tradeoff between being able to preserve signal and absorb noise.

In this talk, I compare classification and regression tasks in the overparameterized linear model. On the one hand, we show that with sufficient overparameterization all training points are support vectors: Solutions obtained by least-squares minimum-norm interpolation, typically used for regression, are identical to those produced by the hard-margin support vector machine (SVM) that minimizes the hinge loss, typically used for training classifiers. On the other hand, we show that there exist regimes where these solutions are near-optimal when evaluated by the 0–1 test loss function, but do not generalize if evaluated by the square loss function, i.e. they achieve the null risk. Our results demonstrate that: i) different loss functions at the training (optimization) phase could yield similar solutions; and ii) a significantly higher level of effective overparameterization admits good generalization in classification tasks as compared to regression tasks.

Joint work with Misha Belkin, Daniel Hsu, Adhyyan Narang, Anant Sahai, Vignesh Subramanian and Ji Xu, based on the preprints: https://arxiv.org/abs/2005.08054 and https://arxiv.org/abs/2009.10670.

Vidya Muthukumar
PhD candidate, UC Berkeley
Google Research Fellow
Program: Theory of Reinforcement Learning
Dates of Visit: Aug. 19–Dec. 18, 2020

Vidya Muthukumar is a final-year PhD candidate at the University of California Berkeley, advised by Anant Sahai. Her broad interests are in online decision-making and game theory. She is particularly interested in designing learning algorithms that provably adapt in limited-information feedback, Markov decision processes and strategic environments. Additional interests include fundamental properties of overparameterized models, and fairness, accountability, and transparency in machine learning.

After her stint at the Simons Institute, Vidya will join Georgia Tech's departments of Electrical Engineering and Industrial and Systems-Engineering as a tenure-track assistant professor.

Return to Top

Towards Instance Optimality in Reinforcement Learning: From Lower Bounds to Algorithms

The paradigm of reinforcement learning has now made inroads in a wide range of applied problem domains. This empirical research has revealed the limitations of our theoretical understanding—popular RL algorithms exhibit a variety of behavior across domains and problem instances, and existing theoretical bounds, which are generally based on worst-case assumptions, fail to capture this variety and can often produce pessimistic predictions. An important theoretical goal is to develop instance-specific analyses that help to reveal what aspects of a given problem make it "easy" or "hard", and allow distinctions to be drawn between ostensibly similar algorithms in terms of their (finite-sample) performance profiles.

Taking an approach grounded in nonparametric statistics, we study this question for the policy evaluation problem. We show via information-theoretic lower bounds that many popular variants of temporal difference (TD) learning algorithms *do not* exhibit the optimal instance-specific performance in the finite-sample regime. On the other hand, well-chosen modifications to these algorithms result in automatic adaptation to the intrinsic difficulty of the problem. When there is function approximation involved, our lower bounds also characterize the optimal tradeoff between "bias" and "variance" that is, notably, distinct from analogous bounds for supervised learning. This characterization reveals, once again, that some popular TD algorithms operate at a suboptimal point on the tradeoff. I will also briefly touch upon what this perspective has to offer for the policy optimization problem.

Joint work with Koulik Khamaru, Wenlong Mou, Feng Ruan, Martin Wainwright, and Michael Jordan.

Ashwin Pananjady
PhD candidate, UC Berkeley
Swiss Re Research Fellow
Program: Probability, Geometry, and Computation in High Dimensions
Dates of Visit: Aug. 19–Dec. 18, 2020

Ashwin Pananjady is a (soon to be graduating) PhD student in the Department of Electrical Engineering and Computer Sciences at the University of California Berkeley, advised by Martin Wainwright and Thomas Courtade. His interests lie broadly in statistics, machine learning, information theory, and optimization, and include ranking and permutation estimation, high-dimensional and non-parametric statistics, high-dimensional probability, and reinforcement learning. He is a recipient of the inaugural Lawrence D. Brown PhD Student Award from the Institute of Mathematical Statistics, the David J. Sakrison Memorial Prize for his dissertation research (EECS, UC Berkeley), an Outstanding Graduate Student Instructor Award from UC Berkeley, and the Governor's Gold Medal from the Indian Institute of Technology Madras.
After a stint at the Simons Institute, Ashwin will be joining Georgia Tech as an assistant professor, with a joint appointment between the School of Industrial and Systems Engineering and the School of Electrical and Computer Engineering.

Return to Top

Linear-Quadratic Reinforcement Learning (LQRL)

Linear-Quadratic (LQ) models are classical for decision-making in unknown environments. In these models, state vectors of the environment evolve according to an unknown linear stochastic transition, and the reward is a quadratic function of state and action. Because of uncertainty in the state transition, reinforcement learning (RL) algorithms are needed to rapidly learn optimal actions. The fundamental challenge is balancing the contradictory objectives of learning and earning. That is, the trade-off between exploring to learn the unknown environment versus exploiting the available information to earn the most. We present RL algorithms with provable performance guarantees to learn accurately and earn optimally at the same time.

Mohamad Kazem Shirani Faradonbeh
Postdoctoral Researcher, University of Florida
Research Fellow
Program: Theory of Reinforcement Learning
Dates of Visit: Aug. 19–Dec. 18, 2020

Mohamad Kazem Shirani Faradonbeh received a PhD degree in statistics from the University of Michigan, Ann Arbor in 2017, and a BS degree in electrical engineering from Sharif University of Technology, Tehran in 2012. He is currently a postdoctoral fellow with the Informatics Institute and the Department of Statistics at the University of Florida, and will join the Statistics Department at the University of Georgia as an assistant professor, starting Fall 2020.

Return to Top

Approximate Nearest Neighbor Search in Non-Euclidean Spaces

The (approximate) nearest neighbor (ANN) search problem asks: given a dataset of n vectors in R^d, endowed with some norm ||.||, design a data structure, which on a query vector q in R^d outputs a point p from the dataset that is approximately the nearest neighbor within the dataset. Much of the progress in ANN stems from the development of locality-sensitive hashing (LSH), and while these techniques give good approximations for normed spaces like ell_1 and ell_2, the landscape quickly blurs beyond these. We discuss recent progress for ell_p spaces (with p > 2) resulting from certain space decompositions which go beyond LSH.

Erik Waingarten
Postdoctoral Researcher, Stanford University
Research Fellow
Program: Probability, Geometry, and Computation in High Dimensions
Dates of Visit: Aug. 19–Dec. 18, 2020

Erik Waingarten received his PhD in 2020 from Columbia University, advised by Rocco Servedio and Xi Chen. His primary research interest is in sublinear time and space algorithms.

Return to Top

Principled Deep Reinforcement Learning for Societal Systems

The goal of this research is to develop a new generation of data-driven decision-making algorithms, theory, and software to address pressing challenges in societal systems. (i) Specifically, it aims to break various barriers that prohibit principled applications of deep reinforcement learning (RL) in critical domains, e.g., healthcare, transportation, power grid, financial network, and supply chain. (ii) Also, it aims to initiate a new subfield, namely societal deep RL, by connecting deep RL with multiple fields, e.g., nonconvex optimization, nonparametric statistics, causal inference, stochastic game, and social science. (iii) Meanwhile, it aims to train future leaders of academia, industry, and government by equipping them with fundamental skills in data science and artificial intelligence, which are needed to make accountable decisions with positive and progressive social impacts.

Zhaoran Wang
Assistant Professor, Northwestern University
JP Morgan Research Fellow
Program: Theory of Reinforcement Learning
Dates of Visit: Aug. 19–Dec. 18, 2020

Zhaoran Wang is an assistant professor at Northwestern University, working at the interface of machine learning, statistics, and optimization. He is the recipient of the AISTATS (Artificial Intelligence and Statistics Conference) notable paper award; ASA (American Statistical Association) best student paper in statistical learning and data mining; INFORMS (Institute for Operations Research and the Management Sciences) best student paper finalist in data mining; and the Microsoft fellowship

Return to Top

Provably Efficient Reinforcement Learning with Value Function Approximation

Value function approximation (VFA) has demonstrated phenomenal empirical success in reinforcement learning (RL). Nevertheless, the understanding of function approximation schemes is still largely missing. In this talk, we will discuss recent progress on understanding efficient RL algorithms with function approximation. In particular, we show tight characterizations of RL with VFA about when it will be hard or easy. We show exponential sample lower bound for hard instances and also provably efficient algorithms for general function approximation under appropriate assumptions.

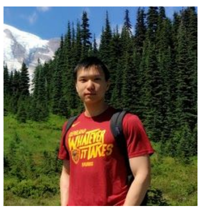
Lin Yang
Assistant Professor, UCLA
Facebook/Novi Research Fellow
Program: Theory of Reinforcement Learning
Dates of Visit: Aug. 19–Dec. 18, 2020

Dr. Lin Yang is an assistant professor in the Department of Electrical and Computer Engineering at the University of California, Los Angeles. His current research focus is on reinforcement learning theory and applications, learning for control, non-convex optimization, and streaming algorithms. Previously, he was a postdoc at Princeton University working with Professor Mengdi Wang. He obtained two PhD degrees (in computer science and in physics and astronomy) simultaneously, from Johns Hopkins University. Prior to that, he obtained a bachelor degree in mathematics and physics from Tsinghua University. He was a recipient of the Dean Robert H. Roy Fellowship at Johns Hopkins.

Return to Top

Provably Efficient Exploration in Reinforcement Learning: An Optimistic Approach

Modern Reinforcement Learning (RL) is commonly applied to practical problems with an enormous number of states, where function approximation such as deep neural networks must be deployed to approximate either the value function or the policy. The introduction of function approximation raises a fundamental set of challenges involving computational and statistical efficiency, especially under the online setting with active data acquisition. As a result, a core RL question remains open: How can we design provably efficient RL algorithms that incorporate possibly nonlinear function approximation?
In this talk, I will introduce the first generation of efficient value-based and policy-based RL algorithms under the setting where both the value function and policy are represented by powerful function approximators such as the kernel and neural network functions. The proposed algorithms highlight a systematic integration of the "optimism under the face of uncertainty" principle into algorithm design and are shown to enjoy both polynomial runtime and polynomial sample complexity. Finally, as an initial attempt to study multi-agent reinforcement learning, I will show how the value-based algorithm can be modified for solving zero-sum stochastic games with efficiency.



Zhuoran Yang
Graduate Student, Princeton University
VMware Research Fellow
Program: Theory of Reinforcement Learning
Dates of Visit: Aug. 19–Dec. 18, 2020

Zhuoran Yang is a PhD candidate in the Department of Operations Research and Financial Engineering at Princeton University advised by professors Jianqing Fan and Han Liu. Prior to attending Princeton, he obtained a bachelor of mathematics degree from Tsinghua University.
His research interests lie in the interface between machine learning, statistics and optimization. The primary goal of his research is to design efficient learning algorithms for large-scale decision-making problems that arise in reinforcement learning and stochastic games, with both statistical and computational guarantees. In addition, he is also interested in the applications of reinforcement learning, such as computer games and robotics.

[Return to Top](#)

Adaptive Discretization For Reinforcement Learning

We introduce the technique of adaptive discretization to design efficient model-free and model-based episodic reinforcement learning algorithms in large (potentially continuous) state-action spaces. We provide worst-case regret bounds for our algorithms, which are competitive compared to the state-of-the-art algorithms. Our algorithms have lower storage and computational requirements due to maintaining a more efficient partition of the state and action spaces. We illustrate this via experiments on several canonical control problems, which show that our algorithms empirically perform significantly better than fixed discretization in terms of both faster convergence and lower memory usage.

Christina Yu
Assistant Professor, Cornell University
Research Fellow
Program: Theory of Reinforcement Learning
Dates of Visit: Aug. 19–Dec. 18, 2020

Christina Lee Yu is an assistant professor at Cornell University in the School of Operations Research and Information Engineering. Prior to Cornell, she was a postdoc at Microsoft Research New England. She received her PhD in 2017 and MS in 2013 in electrical engineering and computer science from Massachusetts Institute of Technology in the Laboratory for Information and Decision Systems. She received her BS in computer science from California Institute of Technology in 2011. She received honorable mention for the 2018 INFORMS Dantzig Dissertation Award. Her recent interests include matrix and tensor estimation, multi-arm bandits, and reinforcement learning.

Return to Top

Industry Partners

Multi Calibration for Importance Weights

Consider a scenario where we are expecting to see samples from a prior distribution, but in reality we see a different distribution. How do we update our prior belief to model reality? This is a ubiquitous problem in statistics and machine learning, arising in anomaly detection and domain adaptation for example. A classical solution is to reweight the prior using importance weights to better model reality. However, computing exact importance weights from random samples is intractable for statistical and computational reasons. Given this, how should we define what makes a set of weights good, and how do we compute good weights?

In this work, we introduce the notion of multi calibration for importance weights, inspired by recent work on fairness in supervised learning. We will explain what multi calibration should mean in this setting, why it is a desirable notion, and how one can achieve it. Based on joint work with Omer Reingold (Stanford), Vatsal Sharan (Stanford/MIT) and Udi Wieder (VMware).

Dr. Parikshit Gopalan is senior researcher in the VMware research group. Prior to this, he was a researcher at Microsoft Research (Silicon Valley and Redmond), and a postdoc at the University of Texas Austin and the University of Washington. Dr. Gopalan received his B Tech in computer science from the Indian Institute of Technology Bombay in 2000 and his PhD in computer science from Georgia Institute of Technology in 2006. His current research interests include algorithms, coding and information theory, and their applications to machine learning, big data and distributed storage.

Return to Top

Upgrade Reality: Overview of NTT Research Activities at Simons Institute

NTT Research opened its Palo Alto offices in July 2019 as a new Silicon Valley startup to conduct basic research and advance technologies that promote positive change for humankind. NTT Research is part of NTT, a global technology and business solutions provider with an annual research and development budget of $3.6 billion. Currently, three labs are housed at NTT

Research: the PHI Lab (Physics and Informatics), the CIS Lab (Cryptography and Information Security), and the MEI Lab (Medical and Health Informatics). This talk will focus on activities to upgrade reality in three areas: i) quantum information, neuro-science and photonics; ii) cryptographic and information security; and iii) medical and health informatics.

Kazuhiro “Kazu” Gomi is the president and CEO of NTT Research Inc. Mr. Gomi has held several key roles within NTT, including VP in charge of global IP-network business, and president and CEO of NTT America Inc. Mr. Gomi received his MS in electrical engineering from Tokyo’s Keio University in 1985 and his MS in industrial engineering from University of Illinois Urbana-Champaign in 1992.

Return to Top

Hybridizing Machine Learning Algorithms to Explore Ways to Transform Insurance Enterprises with Machine Intelligence.

Swiss Re

Over the past decade, re/insurers have invested heavily in experimenting with a wide range of machine intelligence (MI). With a few exceptions, most of the investment in newer MI such as machine learning (ML) and artificial intelligence (AI) has not substantively transformed the insurance industry (sigma 5/2020 – Machine intelligence in insurance). One promising path to changing this outcome involves hybridizing ML algorithms with physics-based models, systematized subject-matter expertise, causal-inference techniques, and simulation. In this presentation, Dr. Bohn will provide an overview of the challenges with enterprise-MI deployment in the insurance industry and present initial ideas for how hybridized ML algorithms can realize the transformative potential of MI.

Dr. Jeffrey R. Bohn is the chief research and innovation officer and head of research and engagement at the Swiss Re Institute. Most recently, he served as chief science officer and head of GX Labs at State Street Global Exchange in San Francisco. Before moving back to California, he established the portfolio analytics and valuation department within State Street Global Markets Japan in Tokyo. Dr. Bohn often conducts seminars on topics ranging from credit instrument valuation and portfolio management to machine learning. He has published widely in the area of credit risk. He co-authored with Roger Stein *Active Credit Portfolio Management in Practice* (Wiley, 2009). His recent research focuses on socially responsible machine intelligence, causal inference to improve machine-learning interpretability, natural-language processing to improve risk modeling, and on a range of machine-intelligence-enabled tools to assess company and urban risk plus resilience in environmental, social, and governance areas. Dr. Bohn received his MS in 1997 and PhD in 1999 in finance from UC Berkeley Haas School of Business.

Return to Top

Differential Privacy for Crypto (for Differential Privacy)

Apple

Differential privacy (DP) addresses the question of what functions we can compute while ensuring privacy of the data we compute on. When the data is distributed across users, traditional cryptographic tools such as secure multiparty computation

(SMC) can ensure privacy during the process of the computation. These methods provide a strong notion of security, but can be expensive in terms of communication and computation.

In this talk, I will discuss the potential benefits of relaxing the security guarantee in SMC to differential privacy. I will show how this relaxation leads to a more efficient protocol for computing the (noisy) sum of real-valued vectors. The protocol can be made robust to data-poisoning attacks. This work raises the possibility of a more efficient and practical protocol for other common cryptographic tasks in settings in which a DP-type security guarantee suffices.

Kunal Talwar is a research scientist at Apple, working in the areas of differential privacy, machine learning, algorithms and data structures. He has previously worked at Google Brain and at Microsoft Research SVC. He received a bachelor degree from the Indian Institute of Technology Delhi and a PhD from University of California Berkeley. He has won the PET Award for Outstanding Research in Privacy Enhancing Technologies for inventing the exponential mechanism in differential privacy, and the 2017 Best Paper Award at the International Conference on Learning Representations (ICLR).

Return to Top

Multi-agent reinforcement learning for agent-based modeling of markets

This talk will cover a specific research problem, agent-based modeling of markets and related open problems. The talk will conclude with a sampling of other research problems important to JP Morgan, and across the financial services domain.

Dr. Sumitra Ganesh is a Research Director at J.P.Morgan AI Research. Her current research is focused on agent-based modeling of complex systems (e.g. markets), and developing learning algorithms that can work efficiently and safely in the real world and in the presence of other strategic agents. Sumitra previously led the Cross-asset Client Intelligence team in the Corporate and Investment Bank at J.P. Morgan, where she was instrumental in improving client experience through several ML products – her team built the first personalization engine for J.P. Morgan Markets digital platform, a virtual assistant platform for client support and several predictive tools to empower sales. Prior to joining J.P. Morgan, Sumitra worked at Goldman Sachs and as a researcher at the Kellogg School of Management. Sumitra holds a Ph.D. in EECS from U.C. Berkeley where her research focused on modeling and recognition of human actions from 3D visual data.

Return to Top

Convex Function Chasing

We consider the problem of chasing convex functions, where functions arrive over time. The player takes actions after seeing the function, and the goal is to achieve a small function cost for these actions, as well as a small cost for moving between actions. We show an O(d)-competitive algorithm for the general problem when the functions are from R^d to R.

Dr. Guru Guruganesh is a research scientist at the Market Algorithms Group at Google Research. He received his PhD from Carnegie Mellon University in 2018. His research interests include approximation algorithms, online algorithms, and algorithmic game theory.

[Return to Top](#)

A Law of Robustness for Neural Networks

Microsoft Research

The talk will present a precise mathematical conjecture that overparameterization is necessary for adversarial robustness in neural networks.

Dr. Sébastien Bubeck manages the Machine Learning Foundations group at Microsoft Research (MSR), which spans a broad range of topics in theoretical computer science and machine learning. His current main research directions are adversarial examples, multi-player multi-armed bandits, and smooth non-convex optimization in low dimensions. His prior works on online decision making and convex optimization have won several best paper and best student paper awards at machine learning conferences (NeurIPS 2018, ALT 2018, COLT 2009 and 2016). He was also a 2015 Alfred P. Sloan Research Fellow in Computer Science.

[Return to Top](#)

FastPay: High-Performance Byzantine Fault Tolerant Settlement

novi

FastPay allows a set of distributed authorities, some of which are Byzantine, to maintain a high-integrity and availability settlement system for pre-funded payments. It can be used to settle payments in a native unit of value (crypto-currency), or as a financial side-infrastructure to support retail payments in fiat currencies. FastPay is based on Byzantine consistent broadcast as its core primitive, foregoing the expenses of full atomic commit channels (consensus). The resulting system has low-latency for both confirmation and payment finality. Remarkably, each authority can be sharded across many machines to allow unbounded horizontal scalability. Our experiments demonstrate intra-continental confirmation latency of less than 100ms, making FastPay applicable to point of sale payments. In laboratory environments, we achieve more than 80,000 transactions per second with 20 authorities — surpassing the requirements of current retail card payment networks, while significantly increasing their robustness.

Alberto Sonnino was a co-founder and researcher at chainspace.io, which builds a scalable smart contract platform. He is now a research scientist at Facebook Novi, based in London. His research interests are in distributed systems, blockchains, and privacy enhancing technologies. He has a special interest in cryptography. In the past, he worked on numerical calculations and simulations for physics-related problems (plasma physics, category theory, symmetry groups).

Return to Top

Statistical Analysis of Privacy Protected Data

In early 2019, Facebook released, as far as we know, the largest social science research dataset ever. The dataset covers millions of URLs that had been shared publicly on Facebook by at least 100 unique Facebook users. Differential privacy techniques were used to protect the privacy of individuals in this dataset. Carefully calibrated noise was ingested into data cells, where the distribution of the noise was also released. The ingested noise introduced measurement error. The naive estimators can be asymptotically biased when the predictors are subject to measurement error. In linear regression, we use methods of moments to debias the coefficient estimators and use simulation to estimate the variance-covariance matrix, following the proposal by OpenDP researchers. In logistic regression, we are exploring the use of the conditional score estimator for the coefficient and use m-estimator for the variance-covariance matrix. Computational efficiency is critical since most of the privacy protected datasets are very large. We open sourced a python package, svinfer, on Github, which provides convenient tools to analyze the privacy protected dataset either in memory or in databases.

Dr. Runchao Jiang is a software engineer at Facebook, Inc. Her current work focuses on privacy protection techniques and their application to the industry scale big data. Her work also involves statistical inference on privacy protected data. She received her PhD in statistics from North Carolina State University. Her paper on personalized medicine was selected for the 2015 ENAR Distinguished Student Paper Award. She joined Facebook after graduation as a quantitative researcher before switching to her current role.

Return to Top

Thank you for your partnership!

Please join us for these coming programs:

Spring 2021

Satisfiability: Theory, Practice, and Beyond	Jan. 12–May 14, 2021
Theoretical Foundations of Computer Systems	Jan. 12–May 14, 2021

Summer 2021

Summer Cluster: Interpretable Machine Learning	June 28–Aug. 6, 2021
Computational Innovation and Data-Driven Biology	July 6–Aug. 6, 2021

Fall 2021

Computational Complexity of Statistical Inference	Aug. 18–Dec. 17, 2021

Spring 2022

Causality	Jan. 12–May 14, 2022

Simons Institute contacts:

Shafi Goldwasser, Director and contact information, simonsdirector@berkeley.edu

Peter Bartlett, Associate Director and contact information, simonsassociatedirector@berkeley.edu

Christine Wang, Science and Outreach Administrator, chris24@berkeley.edu

Amy Ambrose, Sr. Development Director amyambrose@berkeley.edu