

FRONTIERS
IN COMPUTING
+ MATHEMATICAL
SCIENCES

FEBRUARY 29
AND MARCH 7 **2016**

Caltech

ALL TALKS ARE ONE HOUR LONG, AND WILL BE HELD IN ANNENBERG 105.
COFFEE BREAKS WILL BE HELD IN THE ANNENBERG LOBBY.

FEBRUARY 29

MARCH 7

9:20am	Introduction	Joel Tropp Caltech	9:20am	Introduction	Joel Tropp Caltech
9:30am	Fully Verified Outsourced Computation	Bryan Parno Microsoft Research	9:30am	From Communication to Sensing and Learning: Information Theory at the Heart of Data Science	Hamed Hassani ETH Zürich
10:30am	Coffee Break		10:30am	Coffee Break	
11:00am	Algorithms as Multilinear Tensor Equations	Edgar Solomonik ETH Zürich	11:00am	Bringing the Machine into the Loop of Machine Learning	Azalia Mirhoseini UC San Diego
1:00pm	Less Talking, More Learning: Avoiding Coordination In Parallel Machine Learning Algorithms	Dimitris Papailiopoulos UC Berkeley	1:00pm	Network Analysis: Community Detection and Graphon Estimation	Chao Gao Yale
2:00pm	Coffee Break		2:00pm	Coffee Break	
2:30pm	Computation in a Strategic World	Yang Cai McGill University	2:30pm	Learning in Strategic Environments: Theory and Data	Vasilis Syrgkanis Microsoft Research
3:30pm	Coffee Break		3:30pm	Coffee Break	
4:00pm	Middleboxes as a Cloud Service	Justine Sherry UC Berkeley	4:00pm	From Compressed Sensing to Coding Theory	Mary Wootters Carnegie Mellon

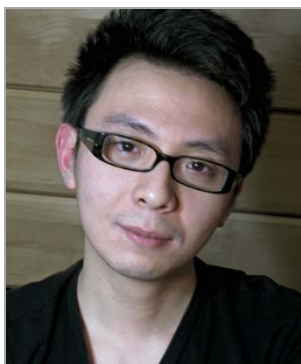
Computation in a Strategic World

Yang Cai
McGill University

Engineering complex systems become even more challenging in the presence of strategic agents whose behavior in the system is guided by their own incentives. Examples include such important applications as the design of road networks, spectrum auctions, and online markets. How does one design a system so that the designer's objective is achieved robustly despite the existence of strategic interactions inside the system?

From an optimization standpoint, this question can be formalized as optimizing an objective function over inputs supplied by strategic agents who have a keen interest in the output of the optimization and may therefore purposefully misreport their inputs to manipulate the output. I will present a computationally efficient transformation that allows optimizing any objective function in the strategic setting by optimizing a perturbed version of the same objective function in the traditional, non-strategic setting. As an application of the transformation, I will provide an algorithmic solution to a central open problem in Economics, generalizing Myerson's celebrated revenue-optimal single-item auction to multi-item settings.

I will overview the combinatorial optimization techniques underlying this result, including computationally efficient variants and generalizations of Border's characterization result. I will also touch upon other work motivated by the study of strategic settings from a computational standpoint, including a multi-player generalization of the min-max theorem, new analysis of the combinatorial clock auction for selling spectrum, and the development of new extreme value theorems.



Biography:

Yang Cai is an assistant professor of Computer Science at McGill University. He received his PhD in computer science from MIT, advised by Costis Daskalakis. He was a postdoctoral researcher in UC Berkeley. Yang's research interests lie in the area of theoretical computer science, in particular algorithmic game theory, applied probability, online algorithms and logic. His Ph.D. thesis was honored with George M. Sprowls dissertation award (honorable mention) and the SIGEcom Dissertation Award (runner-up).

Network Analysis: Community Detection and Graphon Estimation

Chao Gao
Yale University

Network data analysis is an important area in modern statistics. How to do community detection and parameter estimation in provably optimal ways are two key questions in analyzing network data. In this talk, I will present results for both problems. I will first discuss a real data example that motivates the setting of a degree corrected block model. Then I will introduce an efficient two-step polynomial-time algorithm that can achieve the optimal misclassification error for community detection in this setting. The procedure consists of a novel spectral initialization step and a majority voting refinement step. I will then formulate the problem of network parameter estimation as nonparametric graphon estimation, and establish its link to nonparametric regression without observing design. The minimax rate of graphon estimation consists of two parts: the nonparametric part and the clustering part. An interesting implication is that the smoothness of the graphon does not affect the rate once it is greater than 1.



Biography:

Chao Gao is a 5th-year PhD student in statistics at Yale University. His advisor is Harry Zhou. His research lies in nonparametric and high-dimensional statistics, network analysis, Bayes theory and robust statistics.

From Communication to Sensing and Learning: Information Theory at the Heart of Data Science

Hamed Hassani
ETH Zürich

We are witnessing a new era of science — ushered in by our ability to collect massive amounts of data and by unprecedented ways to learn about the physical world. Beyond the challenges of storage and communication, there are new frontiers in the acquisition, analysis and exploration of data. In this talk, I will view these frontiers through the lens of information theory. I will argue that information theory lies at the center of data science, offering insights beyond its classical applications. As a concrete example, I will consider the problem of optimal data acquisition, a challenge that arises in active learning, optimal sensing and experimental design. Based on information theoretic foundations, and equipped with tools from submodular optimization theory, I will present a rigorous analysis of the widely-used sequential information maximization policy (also known as the information-gain heuristic). Our analysis establishes conditions under which this policy provably works near-optimally and identifies situations where the policy fails. In the latter case, our framework suggests novel, efficient surrogate objectives and algorithms that outperform classical techniques.



Biography:

Hamed Hassani is a post-doctoral scholar at the Institute for Machine Learning at ETH Zurich. He received a Ph.D. degree in Computer and Communication Sciences from EPFL, Lausanne. Prior to that, he received a B.Sc. degree in Electrical Engineering and a B.Sc. degree in Mathematics from Sharif University of Technology, Tehran. Hamed's fields of interest include machine learning, coding and information theory as well as theory and applications of graphical models. He is the recipient of the 2014 IEEE Information Theory Society Thomas M. Cover Dissertation Award. His co-authored paper at ISIT 2015 received the IEEE Jack Keil Wolf ISIT Student Paper Award.

Bringing the Machine into the Loop of Machine Learning

Azalia Mirhoseini

UC San Diego

Contemporary analytical algorithms are often focused on functionality and accuracy with system performance as an afterthought. As their use/scale grows and the computing platforms become diverse, spanning from servers and desktops to smartphones and Internet of Things (IoT) devices, functionality is not just about algorithmic efficiency and accuracy, but also practicality on real-world computing machines. One-size fits all solutions will not meet the physical needs of emerging analytical application scenarios.

In this talk, I will present my research on novel computing frameworks that bring hardware into the loop of designing scalable inference algorithms and learning systems. I will describe how a multi-faceted design that holistically considers the computing domain parameters, namely data, algorithm, and machine, introduces game changing performance gains across the board, including runtime, energy, memory, and network bandwidth. I will then describe my tools which enable automatic end-to-end adoption of the proposed frameworks in a wide range of data inference application scenarios. On the theoretical side, I show how my new solutions reach the target machine's computation/communication bounds. On the practical side, I present customized approaches for a range of algorithms and applications (e.g., penalized regression, classification, and deep neural networks), datasets (e.g., visual and sensing), and machines (e.g., GPU, FPGA, CPU clusters, and heterogeneous architectures). I also demonstrate my approach towards enabling single-pass streaming learning problems. Finally, I discuss how lessons learned in the context of my holistic frameworks can bring new directions in the design of broader analytical scenarios such as privacy preserving and just-in-time computing.



Biography:

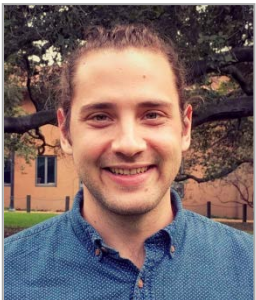
Azalia Mirhoseini is a postdoctoral researcher in the department of Electrical and Computer Engineering (ECE) at the University of California, San Diego and Rice University. She received her Ph.D. from Rice University where she worked on algorithms and architectures for performance efficient data analytics. Azalia has received multiple awards, including the best 2015 Ph.D. thesis award at Rice ECE department, gold medal in the national math olympiad in Iran, and fellowships from IBM, Schlumberger, and Microsoft Research.

Less Talking, More Learning: Avoiding Coordination In Parallel Machine Learning Algorithms

Dimitris Papailiopoulos
UC Berkeley

The recent success of machine learning (ML) in both science and industry has generated an increasing demand to support ML algorithms at scale. In this talk, I will discuss strategies to gracefully scale machine learning on modern parallel computational platforms. A common approach to such scaling is coordination-free parallel algorithms, where individual processors run independently without communication, thus maximizing the time they compute. However, analyzing the performance of these algorithms can be challenging, as they often introduce race conditions and synchronization problems.

In this talk, I will introduce a general methodology for analyzing asynchronous parallel algorithms. The key idea is to model the effects of core asynchrony as noise in the algorithmic input. This allows us to understand the performance of several popular asynchronous machine learning approaches, and to determine when asynchrony effects might overwhelm them. To overcome these effects, I will propose a new framework for parallelizing ML algorithms, where all memory conflicts and race conditions can be completely avoided. I will discuss the implementation of these ideas in practice, and demonstrate that they outperform the state-of-the-art across a large number of ML tasks on gigabyte-scale data sets.



Biography:

Dimitris Papailiopoulos is a postdoctoral researcher in the Department of Electrical Engineering and Computer Sciences at UC Berkeley and a member of the AMPLab. His research interests span machine learning, coding theory, and parallel and distributed algorithms, with a current focus on coordination-free parallel machine learning, large-scale data and graph analytics, and the use of codes to speed up distributed computation.

Dimitris completed his Ph.D. in electrical and computer engineering at UT Austin in 2014. At Austin he worked under the supervision of Alex Dimakis. In 2015, he received the IEEE Signal Processing Society, Young Author Best Paper Award.

Fully Verified Outsourced Computation

Bryan Parno
Microsoft Research

Frequent headline-grabbing data breaches suggest that current best practices for safeguarding personal data are woefully inadequate. To try to move beyond the cycle of attacks and patches we see today, I design and build systems with formal end-to-end guarantees. For example, to provide strong guarantees for outsourced computations, I developed a new cryptographic framework, verifiable computation, which allows clients to outsource general computations to completely untrusted services and efficiently verify the correctness of each returned result. Through improvements to the theory and the underlying systems, we reduced the costs of verification by over twenty orders of magnitude. As a result, verifiable computation is now a thriving research area that has produced several startups, as well as enhancements to the security and privacy of X.509, MapReduce, and Bitcoin.

While verifiable computation provides strong guarantees, even the best cryptographic system is useless if implemented badly, applied incorrectly, or used in a vulnerable system. Thus, I have led a team of researchers and engineers in the Ironclad project, working to expand formal software verification to provide end-to-end guarantees about the security and reliability of complex systems. By creating a set of new tools and methodologies, Ironclad produced the first complete stack of verified-secure software. We also recently developed the first methodology for verifying both the safety and liveness of complex distributed systems implementations. While interesting challenges remain, I expect that verification will fundamentally improve the software that underpins our digital and physical infrastructure.



Biography:

Bryan Parno is a Researcher in the Security and Privacy Group at Microsoft Research. After receiving a Bachelor's degree from Harvard College, he completed his PhD at Carnegie Mellon University, where his dissertation won the 2010 ACM Doctoral Dissertation Award. He formalized and worked to optimize verifiable computation, receiving a Best Paper Award at the IEEE Symposium on Security and Privacy his advances. He coauthored a book on Bootstrapping Trust in Modern Computers, and his work in that area has been incorporated into the latest security enhancements in Intel CPUs. His research into security for new application models was incorporated into Windows and received a Best Paper Awards at the IEEE Symposium on Security and Privacy and the USENIX Symposium on Networked Systems Design and Implementation. He has recently extended his interest in bootstrapping trust to the problem of building practical, formally verified secure systems. His other research interests include user authentication, secure network protocols, and security in constrained environments (e.g., RFID tags, sensor networks, or vehicles).

Middleboxes as a Cloud Service

Justine Sherry
UC Berkeley

Today's networks do much more than merely deliver packets. Through the deployment of middleboxes, enterprise networks today provide improved security -- e.g., filtering malicious content -- and performance capabilities -- e.g., caching frequently accessed content. Although middleboxes are deployed widely in enterprises, they bring with them many challenges: they are complicated to manage, expensive, prone to failures, and challenge privacy expectations.

In this talk, we aim to bring the benefits of cloud computing to networking. We argue that middlebox services can be outsourced to cloud providers in a similar fashion to how mail, compute, and storage are today outsourced. We begin by presenting APLOMB, a system that allows enterprises to outsource middlebox processing to a third party cloud or ISP. For enterprise networks, APLOMB can reduce costs, ease management, and provide resources for scalability and failover. For service providers, APLOMB offers new customers and business opportunities, but also presents new challenges. Middleboxes have tighter performance demands than existing cloud services, and hence supporting APLOMB requires redesigning software at the cloud. We re-consider classical cloud challenges including fault-tolerance and privacy, showing how to implement middlebox software solutions with throughput and latency 2-4 orders of magnitude more efficient than general-purpose cloud approaches. Some of the technologies discussed in this talk are presently being adopted by industrial systems used by cloud providers and ISPs.



Biography:

Justine Sherry is a computer scientist and doctoral candidate at UC Berkeley. Her interests are in computer networking; her work includes middleboxes, networked systems, measurement, cloud computing, and congestion control. Justine's dissertation focuses on new opportunities and challenges arising from the deployment of middleboxes -- such as firewalls and proxies -- as services offered by clouds and ISPs. Justine received her MS from UC Berkeley in 2012, and her BS and BA from the University of Washington in 2010. She is an NSF Graduate Research Fellow, has won paper awards from both USENIX NSDI and ACM SIGCOMM, and is always on the lookout for a great cappuccino.

Algorithms as Multilinear Tensor Equations

Edgar Solomonik
ETH Zürich

Tensors provide a powerful abstraction for expressing algorithms on sparse or dense datasets in their natural dimensionality. Graph algorithms such as betweenness centrality, as well as recursive algorithms such as FFT and bitonic sort, can be succinctly written as tensor operations over a suitable algebraic structure. I will introduce communication and synchronization cost lower bounds for a general class of tensor algorithms, including sparse iterative methods and matrix factorizations. Then, I will present parallel algorithms that achieve minimal cost with respect to these bounds and obtain improved scalability on supercomputers. Additionally, I will describe new innovations in handling symmetry and sparsity in tensors. Some of the proposed algorithms are deployed in a massively-parallel tensor framework, whose development has been driven by applications in quantum chemistry. I will show the performance of the framework for algorithmic benchmarks as well as for coupled cluster methods, which model electronic correlation by solving tensor equations.



Biography:

Edgar Solomonik is a postdoctoral fellow at ETH Zürich working in the field of parallel algorithms. His research introduced more communication-efficient algorithms for numerical linear algebra and his software for tensor computations has been widely adopted in the field of electronic structure calculations. He obtained his BS from the University of Illinois, Urbana-Champaign and his PhD from the University of California, Berkeley, both in Computer Science. He was the recipient of the DOE Computational Science Graduate Fellowship, the David J. Sakrison Memorial Prize, and the ACM-IEEE George Michael HPC Fellowship.

Learning in Strategic Environments: Theory and Data

Vasilis Syrgkanis

Microsoft Research

The strategic interaction of multiple parties with different objectives is at the heart of modern large scale computer systems and electronic markets. Participants face such complex decisions in these settings that the classic economic equilibrium is not a good predictor of their behavior. The analysis and design of these systems has to go beyond equilibrium assumptions. Evidence from online auction marketplaces suggests that participants rather use algorithmic learning. In the first part of the talk, I will describe a theoretical framework for the analysis and design of efficient market mechanisms, with robust guarantees that hold under learning behavior, incomplete information and in complex environments with many mechanisms running at the same time. In the second part of the talk, I will describe a method for analyzing datasets from such marketplaces and inferring private parameters of participants under the assumption that their observed behavior is the outcome of a learning algorithm. I will give an example application on datasets from Microsoft's sponsored search auction system.



Biography:

Vasilis Syrgkanis is a postdoctoral researcher at Microsoft Research NYC, where he is a member of the algorithmic economics and machine learning groups. He received his Ph.D. in Computer Science from Cornell University in 2014, under the supervision of Prof. Eva Tardos. His research addresses problems at the intersection of theoretical computer science, machine learning and economics. His work received best paper awards at the 2015 ACM Conference on Economics and Computation (EC'15) and at the 2015 Annual Conference on Neural Information Processing Systems (NIPS'15). He was the recipient of the Simons Fellowship for graduate students in theoretical computer science 2012-2014.

From Compressed Sensing to Coding Theory

Mary Wootters
Carnegie Mellon University

I'll discuss two problems, which on the surface seem quite different. The first, which comes up in signal processing and in algorithm design, is the problem of coming up with linear, geometry-preserving maps which are efficient to store and manipulate. The second, which comes up in coding theory and theoretical computer science, is the problem of establishing the list-decodability -- a combinatorial property -- of error correcting codes. I'll establish a connection between these two problems, and discuss how techniques from high-dimensional probability can be used to handle both. Punchlines include improved fast Johnson-Lindenstrauss transforms and structured RIP matrices, and the answer to some longstanding open combinatorial questions in coding theory.



Biography:

Mary Wootters is an NSF postdoctoral fellow in the Computer Science Department at Carnegie Mellon University. She received her Ph.D. in Mathematics from the University of Michigan in 2014, and her B.A. in math and computer science from Swarthmore College in 2008. Her research focuses on randomized algorithms, for problems in signal processing, communication, and for dealing with high-dimensional data.