General information: Meetings now usually take place on Thursday at 3pm in Toluca (Building 53, Room 4002) at SLAC, but please check the schedule. You can join the mailing list list-serv AI-AT-SLAC at listserv.slac.stanford.edu. Please contact Kazuhiro Terao (kerao@slac.stanford.edu) if you are interested to give a talk!

- Upcoming Seminars
  - Applying Convolutional Neural Networks to MicroBooNE
- Past Seminars
  - Machine Learning, Datascience and Neutrino Physics at Argonne’s Leadership Computing Facility
  - Deep Learning for Particle Track Finding in High Energy Physics
  - Machine Learning for medical applications of Physics
  - Machine Learning synthetic data, scanning probe data, and reciprocal space data on quantum materials
  - Local-to-Global Methods for Topological Data Analysis
  - Experience with a Virtual Multi-Slit Phase Space Diagnostic at Fermilab’s FAST Facility
  - A Novel Approach - IoT Device Virtualization using ML
  - Sparsity/Undersampling Tradeoffs in Compressed Sensing
  - Learned predictive models: integrating large-scale simulations and experiments using deep learning
  - Rapid Gaussian Process Training via Structured Low-Rank Kernel Approximation of Gridded Measurements
  - Machine learning applications for hospitals
  - Beyond Data and Model Parallelism for Deep Neural Networks
  - X-ray spectrometer data processing with unsupervised clustering (Sideband signal seeking)
  - Experience with FEL taper tuning using reinforcement learning and clustering
  - Statistical Learning of Reduced Kinetic Monte Carlo Models of Complex Chemistry from Molecular Dynamics
  - Machine Learning for Jet Physics at the Large Hadron Collider
  - In situ visualization with task-based parallelism
  - Data Reconstruction Using Deep Neural Networks for Liquid Argon Time Projection Chamber Detectors
  - Towards a cosmology emulator using Generative Adversarial Networks
  - Optimal Segmentation with Pruned Dynamic Programming
  - Fast automated analysis of strong gravitational lenses with convolutional neural networks
  - MacroBase: A Search Engine for Fast Data Streams
  - Object-Centric Machine Learning
  - Reconstruction Algorithms for Next-Generation Imaging: Multi-Tiered Iterative Phasing for Fluctuation X-ray Scattering and Single-Particle Diffraction
  - Exploratory Studies in Neural Network-based Modeling and Control of Particle Accelerators
  - Estimating behind-the-meter solar generation with existing measurement infrastructure
  - Development and Application of Online Optimization Algorithms
  - Machine Learning at NERSC: Past, Present, and Future
  - Optimization for Transportation Efficiency
  - Detecting Simultaneous Changepoints Across Multiple Data Sequences
  - Low Data Drug Discovery with One-Shot Learning
  - Learning Particle Physics by Example: Location-Aware Generative Adversarial Networks for Physics Synthesis
  - Models and Algorithms for Solving Sequential Decision Problems under Uncertainty
  - Data Programming: A New Framework for Weakly Supervising Machine Learning Models
  - ProxImaL: Efficient Image Optimization using Proximal Algorithms
  - Energy-efficient neuromorphic hardware and its use for deep neural networks
  - Locating Features in Detector Images with Machine Learning
  - Tractable quantum leaps in battery materials and performance via machine learning
  - Deep Learning and Computer Vision in High Energy Physics
  - Label-Free Supervision of Neural Networks with Physics and Domain Knowledge
  - Can machine learning teach us physics? Using Hidden Markov Models to understand molecular dynamics.
  - On-the-fly unsupervised discovery of functional materials
  - Machine Learning and Optimization to Enhance the FEL Brightness
  - Automated tuning at LCLS using Bayesian optimization
  - Using Deep Learning to Sort Down Data

Upcoming Seminars

Applying Convolutional Neural Networks to MicroBooNE

Date: June 6th 3:00pm at Bldg 53 Rm 4002
Speaker: Taritree Wongjirad

The MicroBooNE experiment consists of liquid argon time projection chamber (LArTPC) situated in the path of the Booster Neutrino Beam (BNB) at Fermilab. The goals of the experiment are to (1) investigate the observation of an excess of a possible electron-neutrino and anti-neutrino events by the MiniBooNE experiment, (2) measure argon-nucleus cross sections, and (3) perform R&D for LArTPCs. The data from MicroBooNE, and other LArTPCs, can be naturally arranged as high-resolution images of particle tracks traversing the detector. This has spurred effort on
MicroBooNE towards applying convolutional neural networks (CNNs), a type of deep learning algorithm shown to be extremely effective in numerous computer vision problems, to our data. I’ll talk about the ways in which MicroBooNE uses CNNs with a focus on recent results demonstrating their performance on real data. I’ll also discuss future directions MicroBooNE is exploring to further apply CNNs.

Past Seminars

Machine Learning, Data science and Neutrino Physics at Argonne’s Leadership Computing Facility
**Date:** February 28th 3:00pm at Bldg 53 Rm 4002  
**Speaker:** Corey Adams  
**Abstract:** As machine learning and deep learning have become important tools for science, both in analysis and simulation of data, high performance computing workflows have expanded to target learning applications. The Leadership Computing Facility at Argonne National Laboratory (ALCF) will be the home to the first US Exascale computer in 2021, Aurora, and the data science group at ALCF works to scale applications in simulation and learning to current and future systems across broad scientific domains. In this talk, I will describe the ALCF center and its suite of data science projects, and as a neutrino physicist I will describe the neutrinos-on-HPC deep learning applications at ALCF. In particular, I will cover many tools and best practices for getting good performance of machine learning workflows on HPC systems.

Deep Learning for Particle Track Finding in High Energy Physics
**Date:** February 21st 3:00pm at Bldg 53 Rm 4002  
**Speaker:** Steve Farrell  
**Abstract:** Particle track reconstruction is a challenging but critical component of data analysis in high energy physics experiments such as those at the Large Hadron Collider. In the high-luminosity LHC era, experiments will need to identify O(10k) particle tracks in every proton-proton collision event using detectors with O(100M) readout channels. Today's solutions use Kalman Filters and combinatorial searches which are inherently serial and are expected to consume a substantial amount of compute resources at the HL-LHC. Meanwhile, deep learning methodologies are enjoying increasing success in many domains in industry and science. These approaches provide the means to learn complex representations of high-dimensional data for solving complex tasks while also efficiently using modern parallel hardware devices such as GPUs. The HEP.TrkX project has therefore been exploring deep learning solutions using convolutional, recurrent, and graph neural networks for image-like and spacepoint representations of tracking detector data. In this presentation I will describe the approaches that have been studied in the HEP.TrkX project, including an in-depth focus on a graph neural network (GNN) model which operates on a graph-structured representation of tracking detector data. This GNN model is able to uncover particle trajectories in entire collision events by resolving pairwise associations between position measurements in the detector and has shown promising results on simulated data from the Kaggle tracking machine learning challenge. I will also discuss ongoing and future directions for this line of research.

Machine Learning for medical applications of Physics
**Date:** January 16th at 12:30pm at Bldg 53 Rm 4002  
**Speaker:** Carlo Mancini (INFN, Rome)  
**Abstract:** Deep Neural Networks (DNNs) techniques are applied to a vast number of cases, such as human face recognition, image segmentation, self-driving cars, and even playing Go. In this talk, I present our first steps in using DNNs in medical applications, i.e.: to segment Magnetic Resonance (MR) images and to reproduce the final state of a low energy nuclear interaction model, BLOB (Boltzmann Langevin One Body). The first application tries to give an answer to the necessity, expressed by clinicians, of identifying rectal cancer patients who do not need radical surgery after the chemo-radiotherapy prescribed by the clinical protocol. The second one aims at exploring the possibility of using a Variational...
Auto Encoder (VAE) to simulate accurately low energy nuclear interactions in order to reduce the computation time with respect to the full model. Once trained, the VAE could be used in Monte Carlo simulation of patients’ treatments with ion beams.

**Machine Learning synthetic data, scanning probe data, and reciprocal space data on quantum materials**

*Date: October 19 at 1pm (note change in time!)*
*Speaker: Eun-Ah Kim (Cornell)*

**Abstract:** The scientific questions in the field of electronic quantum matter require fundamentally new approaches to data science for two reasons: (1) quantum mechanical imaging of electronic behavior is probabilistic, (2) inference from data should be subject to fundamental laws governing microscopic interactions. In this talk, I will discuss our approaches to synthetic data, scanning probe data and comprehensive X-ray data.

---

**Local-to-Global Methods for Topological Data Analysis**

*Date: October 1, 2018 at 3pm*
*Speaker: Brad Nelson*

**Abstract:** Topological data analysis (TDA) seeks to understand and utilize the shape of data. An important task in TDA is the design and computation of topological invariants of sampled data, which can be transformed into features for machine learning, or used to inform an exploratory modeling process. This talk will begin with an introduction to the core methods of TDA and some scientific applications. We will then present our recent work on the use of hierarchical models to investigate larger and more complex data sets. Video of presentation: [https://stanford.zoom.us/recording/share/E0gJd-JheyJin0gCRNOeWsVWLpB4q4pJjOL_pNM4mnGwlumeKZizMw?startT=153843132700](https://stanford.zoom.us/recording/share/E0gJd-JheyJin0gCRNOeWsVWLpB4q4pJjOL_pNM4mnGwlumeKZizMw?startT=153843132700)

---

**Experience with a Virtual Multi-Slit Phase Space Diagnostic at Fermilab’s FAST Facility**

*Date: August 27, 2018 at 3pm*
*Speaker: Auralee Edelen*

**Abstract:** We discuss an ML-based virtual diagnostic for multi-slit beam measurements at the Fermilab Accelerator Science and Technology...
Facility (FAST). At FAST, a multi-slit diagnostic setup is used to obtain information about the transverse electron beam phase space. In addition to being a destructive measurement, the process of conducting this measurement at multiple points along the machine (e.g. before and after the chicane) is time-consuming. Here, we use a learned representation of the low-energy portion of the machine to provide estimates of the multi-slit measurement after the second superconducting capture cavity. The result is a fast-executing tool that uses present machine settings to give an online prediction of the resultant multi-slit measurement without the need to repeatedly insert the intercepting diagnostic. This work is particularly relevant to ongoing efforts at FACET-II and LCLS/LCLS-II to create similar kinds of virtual diagnostic tools for prediction and, ultimately, control of the longitudinal phase space.

A Novel Approach - IoT Device Virtualization using ML

Date: July 19, 2018 at 11am (Note special time!)
Speaker: Knowthings
Abstract: Developing and testing IoT solutions is challenging today due to many factors: complexity around available platforms, heterogeneous environments, voluminous device requirements, multiple protocols, and communication synchronization. In this session we'll find out how IoT Device Virtualization can help expedite development and testing of IoT applications, and how machine learning can play a very important role in that. We'll see how sequence alignment techniques and data mining methods, like Needleman Wunsch algorithm, enable the means of learning device behavior, and generate a data model out of it. This model can then be played back at runtime to behave like an actual IoT device.

Sparsity/Undersampling Tradeoffs in Compressed Sensing

Date: July 9, 2018 at 3pm
Speaker: Hatef Monajemi (Stanford)
Abstract: Compressed sensing (CS) is a sampling technique that exploits sparsity to speed up acquisition. An important question in CS is "how much undersampling is allowed for a given level of sparsity?" This question has been answered when sampling is done using Gaussian random matrices. Unfortunately, the theories for Gaussian matrices are not directly applicable to certain real life applications such as magnetic resonance imaging/spectroscopy where unique experimental considerations may impose extra constraints on the sampling matrix. In this talk, we will review the literature on sparsity/undersampling tradeoffs for Gaussian matrices and then present new predictions that are applicable to MR spectroscopy/imaging.

Learned predictive models: integrating large-scale simulations and experiments using deep learning

Date: June 25, 2018 at 3pm
Speaker: Brian Spears (LLNL)
Abstract: Across scientific missions, we regularly need to develop accurate models that closely predict experimental observation. Our team is developing a new class of model, called the learned predictive model, that captures theory-driven simulation, but also improves by exposure to experimental observation. We begin by designing specialized deep neural networks that can learn the behavior of complicated simulation models from exceptionally large simulation databases. Later, we improve, or elevate, the trained models by incorporating experimental data using a technique called transfer learning. The training and elevation process improves our predictive accuracy, provides a quantitative measure of uncertainty, and helps us cope with limited experimental data volumes. To drive this procedure, we have also developed a complex computational workflow that can generate hundreds of thousands to billions of simulated training examples and can steer the subsequent training and elevation process. These workflow tasks require a heterogeneous high-performance computing environment supporting computation on CPUs, GPUs, and sometimes specialized, low-precision processors. We will present a global view of our deep learning efforts, our computational workflows, and some implications that this computational work has for current and future large-scale computing platforms.
Rapid Gaussian Process Training via Structured Low-Rank Kernel Approximation of Gridded Measurements

Date: June 4, 2018 at 3pm
Speaker: Franklin Fuller

Abstract: The cubic scaling of matrix inversion with the number of data points is the main computational cost in Gaussian Process (GP) regression. Sparse GP approaches reduce the complexity of matrix inversion to linear complexity by making an optimized low rank approximation to the kernel, but the quality of the approximation depends (and scales with) how many “inducing” or representative points are allowed. When the problem at hand allows the kernel to be decomposed into a kronecker product of lower dimensional kernels, many more inducing points can be feasibly processed by exploiting the kronecker factorization, resulting in a much higher quality fit. Kronecker factorizations suffer from exponentially scaling with the dimension of the input, however, which has limited this approach to problems of only a few input dimensions. It was recently shown how this problem can be circumvented by making an additional low-rank approximation across input dimensions, resulting in an approach that scales linearly with both data points and the input dimensionality. We explore a special case of this recent work wherein the observed data are measured on a complete multi-dimensional grid (not necessarily uniformly spaced), which is a very common scenario in scientific measurement environments. In this special case, the problem decomposes over axes of the input grid, making the cost linearly scale mainly with the largest axis of the grid. We apply this approach to deconvolve linearly mixed spectroscopic signals and are able to optimize kernel hyperparameters on datasets containing billions of measurements in minutes with a laptop.

Machine learning applications for hospitals

Date: May 21, 2018 at 3pm
Speaker: David Scheinker

Abstract: Academic hospitals and particle accelerators have a lot in common. Both are complex organizations; employ numerous staff and scientists; deliver a variety of services; research how to improve the delivery of those services; and do it all with a variety of large expensive machines. My group focuses on helping the Stanford hospitals, mostly the Children’s Hospital, seek to improve: throughput, decision-support, resource management, innovation, and education. I’ll present brief overviews of a variety of ML-based approaches to projects in each of these areas. For example, integer programming to optimize surgical scheduling and Neural Networks to interpret continuous-time waveform monitor data. I will conclude with a broader vision for how modern analytics methodology could potentially transform healthcare delivery. More information on the projects to be discussed is available at surf.stanford.edu/projects

Beyond Data and Model Parallelism for Deep Neural Networks

Date: May 7, 2018 at 3pm
Speaker: Zhihao Jia

Abstract: Existing deep learning systems parallelize the training process of deep neural networks (DNNs) by using simple strategies, such as data and model parallelism, which usually results in suboptimal parallelization performance for large scale training. In this talk, I will first formalize the space of all possible parallelization strategies for training DNNs. After that, I will present FlexFlow, a deep learning framework that automatically finds efficient parallelization strategies by using a guided random search algorithm to explore the space of all possible parallelization strategies. Finally, I will show that FlexFlow significantly outperforms state-of-the-art parallelization approaches by increasing training throughput, reducing communication costs, and achieving improved scalability.

X-ray spectrometer data processing with unsupervised clustering (Sideband signal seeking)

Date: April 9, 2018 at 3pm
Speaker: Guanqun Zhou
Abstract: Online spectrometer plays an important role in the characterization of the free-electron laser (FEL) pulse spectrum. With the help of beam synchronization acquisition (BSA) system, the spectrum of independent shot can be stored, which helps the downstream scientific researchers a lot. However, because of spontaneous radiation, FEL intrinsic fluctuations and other stochastic effects, the data from spectrometer cannot be fully utilized. A specific case is sideband signal resolution in hard-xray self-seeding experiment. During the seminar, I will present my exploration of employing unsupervised clustering algorithm to mine the latent information in the spectrometer data. In this way, sideband signal starts to appear.

Experience with FEL taper tuning using reinforcement learning and clustering
Date: April 2, 2018 at 3pm
Speaker: Juhao Wu
Abstract: LCLS, world’s first hard X-ray Free Electron Laser (FEL) is serving multiple users. It commonly happens that different scientific research requires very different parameters of the X-Ray pulses, therefore setting up the system in a timing fashion meeting these requests is a nontrivial task. Artificial intelligence is not only very helpful to conduct well defined task towards definitive goal, it also helps to find new operating regime generating unexpected great results. Here in this talk, we will report experience with FEL taper tuning using reinforcement learning and clustering. Such study opens up novel taper configuration such as a zig-zag taper which takes full advantages of the filamentation of the electron bunch phase space in the deep saturated regime.

Statistical Learning of Reduced Kinetic Monte Carlo Models of Complex Chemistry from Molecular Dynamics
Date: Feb. 26, 2018 at 3pm
Speaker: Qian Yang (Stanford)
Abstract: Complex chemical processes, such as the decomposition of energetic materials and the chemistry of planetary interiors, are typically studied using large-scale molecular dynamics simulations that can run for weeks on high performance parallel machines. These computations may involve thousands of atoms forming hundreds of molecular species and undergoing thousands of reactions. It is natural to wonder whether this wealth of data can be utilized to build more efficient, interpretable, and predictive models of complex chemistry. In this talk, we will use techniques from statistical learning to develop a framework for constructing Kinetic Monte Carlo (KMC) models from molecular dynamics data. We will show that our KMC models can not only extrapolate the behavior of the chemical system by as much as an order of magnitude in time, but can also be used to study the dynamics of entirely different chemical trajectories with a high degree of fidelity. Then, we will discuss a new and efficient data-driven method using L1-regularization for automatically reducing our learned KMC models from thousands of reactions to a smaller subset that effectively reproduces the dynamics of interest.
Machine Learning for Jet Physics at the Large Hadron Collider

Date: February 12, 2018 at 3pm  
Speaker: Ben Nachman (CERN)

Abstract: Modern machine learning (ML) has introduced a new and powerful toolkit to High Energy Physics. While only a small number of these techniques are currently used in practice, research and development centered around modern ML has exploded over the last year(s). I will highlight recent advances with a focus on jets: collimated sprays of particles resulting from quarks and gluons produced at high energy. Themselves defined by unsupervised learning algorithms, jets are a prime benchmark for state-of-the-art ML applications and innovations. For example, I will show how deep learning has been applied to jets for classification, regression, and generation. These tools hold immense potential, but incorporating domain-specific knowledge is necessary for optimal performance. In addition, studying what the machines are learning is critical for robustness and may even help us learn new physics!

Tutorial – Implementation Practice of Deep Neural Network Technique into Our Research

Date: Wednesday Jan. 24, 2018 at 3pm-5pm in Mammoth B53-3036 (Note time and place!)  
Speaker: Kazuhiro Terao

Abstract: The progress of machine learning techniques in recent years has been impactful in many fields of research including physics science. While learning more about machine learning subject exciting, an implementation in our research can face a huge learning curve and take time. In this first AI seminar of 2018, I will show our implementation of convolution neural network for analyzing liquid argon time projection chamber (LArTPC) detector data. In particular, we will look at an instance of convolutional neural network trained for semantic-segmentation task (i.e. pixel-level object classification task). This seminar will be in an interactive tutorial style using Jupyter python notebook. In the first part of the seminar, I will give an overview of software stacks and workflow. Then I will demonstrate training the algorithm. Both the software and data will be made available in advance, and the audience is welcome to participate in training the algorithm (which may require an NVIDIA GPU enabled Linux interactive shell such as SLAC computing or AWS service).
In situ visualization with task-based parallelism

Date: Nov. 27, 2017 at 3pm  
Speaker: Alan Heirich

Abstract: This short paper describes an experimental prototype of in situ visualization in a task-based parallel programming framework. A set of reusable visualization tasks were composed with an existing simulation. The visualization tasks include a local OpenGL renderer, a parallel image compositor, and a display task. These tasks were added to an existing fluid-particle-radiation simulation and weak scaling tests were run on up to 512 nodes of the Piz Daint supercomputer. Benchmarks showed that the visualization components scaled and did not reduce the simulation throughput. The compositor latency increased logarithmically with increasing node count.

Data Reconstruction Using Deep Neural Networks for Liquid Argon Time Projection Chamber Detectors

Date: Oct. 16, 2017 at 3pm  
Speaker: Kazuhiro Terao

Deep neural networks (DNNs) have found a vast number of applications ranging from automated human face recognition, real-time object detection for self-driving cars, teaching a robot Chinese, and even playing Go. In this talk, I present our first steps in exploring the use of DNNs to the task of analyzing neutrino events coming from Liquid Argon Time Projection Chambers (LArTPC), in particular the MicroBooNE detector. LArTPCs consist of a large volume of liquid argon sandwiched between a cathode and anode wire planes. These detectors are capable of recording images of charged particle tracks with breathtaking resolution. Such detailed information will allow LArTPCs to perform accurate particle identification and calorimetry, making it the detector of choice for many current and future neutrino experiments. However, analyzing such images can be challenging, requiring the development of many algorithms to identify and assemble features of the events in order to identify and remove cosmic-ray-induced particles and reconstruct neutrino interactions. This talk shows the current status of DNN applications and our future direction.
Towards a cosmology emulator using Generative Adversarial Networks

Date: Oct 3, 2017 at 2pm
Speaker: Mustafa Mustafa

The application of deep learning techniques to generative modeling is renewing interest in using high dimensional density estimators as computationally inexpensive emulators of fully-fledged simulations. These generative models have the potential to make a dramatic shift in the field of scientific simulations, but for that shift to happen we need to study the performance of such generators in the precision regime needed for science applications. To this end, in this talk we apply Generative Adversarial Networks to the problem of generating cosmological weak lensing convergence maps. We show that our generator network produces maps that are described by, with high statistical confidence, the same summary statistics as the fully simulated maps.

Optimal Segmentation with Pruned Dynamic Programming

Date: Sept. 12, 2017 at 2pm
Speaker: Jeffrey Scargle (NASA)

Bayesian Blocks (1207.5578) is an O(N^2) dynamic programming algorithm to compute exact global optimal segmentations of sequential data of arbitrary mode and dimensionality. Multivariate data, generalized block shapes, and higher dimensional data are easily treated. Incorporating a simple pruning method yields a (still exact) O(N) algorithm allowing fast analysis of series of ~100M data points. Sample applications include analysis of X- and gamma-ray time series, identification of GC-islands in the human genome, data-adaptive triggers and histograms, and elucidating the Cosmic Web from 3D galaxy redshift data.
Fast automated analysis of strong gravitational lenses with convolutional neural networks

Date: Sept. 12, 2017 at 2pm

Speaker: Yashar Hezaveh

Strong gravitational lensing is a phenomenon in which the image of a distant galaxy appears highly distorted due to the deflection of its light rays by the gravity of a more nearby, intervening galaxy. We often see multiple distinct arc-shaped images of the background galaxy around the intervening (lens) galaxy, just like images in a funhouse mirror. Strong lensing gives astrophysicists a unique opportunity to carry out different investigations, including mapping the detailed distribution of dark matter, or measuring the expansion rate of the universe. All these great sciences, however, require a detailed knowledge of the distribution of matter in the lensing galaxies, measured from the distortions in the images. This has been traditionally performed with maximum-likelihood lens modeling, a procedure in which simulated observations are generated and compared to the data in a statistical way. The parameters controlling the simulations are then explored with samplers like MCMC. This is a time and resource consuming procedure, requiring hundreds of hours of computer and human time for a single system. In this talk, I will discuss our recent work in which we showed that deep convolutional neural networks can solve this problem more than 10 million times faster: about 0.01 seconds per system on a single GPU. I will also review our method for quantifying the uncertainties of the parameters obtained with these networks. With the advent of upcoming sky surveys such as the Large Synoptic Survey Telescope, we are anticipating the discovery of tens of thousands of new gravitational lenses. Neural networks can be an essential tool for the analysis of such high volumes of data.

MacroBase: A Search Engine for Fast Data Streams

Date: Sept. 5, 2017 at 2pm

Speaker: Sahaana Suri (Stanford)

While data volumes generated by sensors, automated process, and application telemetry continue to rise, the capacity of human attention remains limited. To harness the potential of these large scale data streams, machines must step in by processing, aggregating, and contextualizing significant behaviors within these data streams. This talk will describe progress towards achieving this goal via MacroBase, a new analytics engine for prioritizing attention in this large-scale "fast data" that has begun to deliver results in several production environments. Key to this progress are new methods for constructing cascades of analytic operators for classification, aggregation, and high-dimensional feature selection; when combined, these cascades yield new opportunities for dramatic scalability improvements via end-to-end optimization for streams spanning time-series, video, and structured data. MacroBase is a core component of the Stanford DAWN project (http://dawn.cs.stanford.edu/), a new research initiative designed to enable more usable and efficient machine learning infrastructure.
Object-Centric Machine Learning

Date: Aug. 29, 2017 at 2pm

Speaker: Leo Guibas (Stanford)

Deep knowledge of the world is necessary if we are to have autonomous and intelligent agents and artifacts that can assist us in everyday activities, or even carry out tasks entirely independently. One way to factorize the complexity of the world is to associate information and knowledge with stable entities, animate or inanimate, such as persons or vehicles, etc -- what we generally refer to as "objects."

In this talk I’ll survey a number of recent efforts whose aim is to create and annotate reference representations for (inanimate) objects based on 3D models with the aim of delivering such information to new observations, as needed. In this object-centric view, the goal is to learn about object geometry, appearance, articulation, materials, physical properties, affordances, and functionality. We acquire such information in a multitude of ways, both from crowd-sourcing and from establishing direct links between models and signals, such as images, videos, and 3D scans -- and through these to language and text. The purity of the 3D representation allows us to establish robust maps and correspondences for transferring information among the 3D models themselves -- making our current 3D repository, ShapeNet, a true network.

While neural network architectures have had tremendous impact in image understanding and language processing, their adaptation to 3D data is not entirely straightforward. The talk will also briefly discuss current approaches in designing deep nets appropriate for operating directly on irregular 3D data representations, such as meshes or point clouds, both for analysis and synthesis -- as well as ways to learn object function from observing multiple action sequences involving objects -- in support of the above program.

Reconstruction Algorithms for Next-Generation Imaging: Multi-Tiered Iterative Phasing for Fluctuation X-ray Scattering and Single-Particle Diffraction

Date: Aug. 15, 2017 at 2pm

Location: Tulare (BS3-4006) (NOTE CHANGE IN ROOM!)

Speaker: Jeffrey Donatelli (CAMERA, Berkeley)

Abstract: The development of X-ray free-electron lasers has enabled new experiments to study uncrystallized biomolecules that were previously infeasible with traditional X-ray sources. One such emerging experimental technique is fluctuation X-ray scattering (FXS), where one collects a series of diffraction patterns, each from multiple particles in solution, using ultrashort X-ray pulses that allow snapshots to be taken below rotational diffusion times of the particles. The resulting images contain angularly varying information from which angular correlations can be computed, yielding several orders of magnitude more information than traditional solution scattering methods. However, determining molecular structure from FXS data introduces several challenges, since, in addition to the classical phase problem, one must also solve a hyper-phase problem to determine the 3D intensity function from the correlation data.

In another technique known as single-particle diffraction (SPD), several diffraction patterns are collected, each from an individual particle. However, the samples are delivered to the beam at unknown orientations and may also be present in several different conformational states. In order to reconstruct structural information from SPD, one must determine the orientation and state for each image, extract an accurate 3D model of the intensity function from the images, and solve for the missing complex phases, which are not measured in diffraction images.

In this talk, we present the multi-tiered iterative phasing (M-TIP) algorithm for determining molecular structure from both FXS and SPD data. This algorithm breaks up the associated reconstruction problems into a set of simpler subproblems that can be efficiently solved by applying a series of projection operators. These operators are combined in a modular iterative framework which is able to simultaneously determine missing parameters, the 3D intensity function, the complex phases, and the underlying structure from the data. In particular, this approach is able to leverage prior knowledge about the structural model, such as shape or symmetry, to obtain a reconstruction from very limited data with excellent global convergence properties and high computational efficiency. We show results from applying M-TIP to determine molecular structure from both simulated data and experimental data collected at the Linac Coherent Light Source (LCLS).
Exploratory Studies in Neural Network-based Modeling and Control of Particle Accelerators

Date: Aug 1, 2017 at 2pm
Speaker: Auralee Edelen (CSU)

Particle accelerators are host to myriad control challenges: they involve a multitude of interacting systems, are often subject to tight performance demands, in many cases exhibit nonlinear behavior, sometimes are not well-characterized due to practical and/or fundamental limitations, and should be able to run for extended periods of time with minimal interruption. One avenue toward improving the way these systems are controlled is to incorporate techniques from machine learning. Within machine learning, neural networks in particular are appealing because they are highly flexible, they are well-suited to problems with nonlinear behavior and large parameter spaces, and their recent success in other fields (driven largely by algorithmic advances, greater availability of large data sets, and improvements in high performance computing resources) is an encouraging indicator that they are now technologically mature enough to be fruitfully applied to particle accelerators. This talk will highlight a few recent efforts in this area that were focused on exploring neural network-based approaches for modeling and control of several particle accelerator subsystems, both through simulation and experimental studies.

Estimating behind-the-meter solar generation with existing measurement infrastructure

Date: July 11, 2017 at 2pm
Speaker: Emre Kara

Real-time PV generation information is crucial for distribution system operations such as switching, state-estimation, and voltage management. However, most behind-the-meter solar installations are not monitored. Typically, the only information available to the distribution system operator is the installed
capacity of solar behind each meter; though in many cases even the presence of solar may be unknown. We present a method for disaggregating behind-the-meter solar generation using only information that is already available in most distribution systems. Specifically, we present a contextually supervised source separation strategy adopted to address the behind-the-meter solar disaggregation problem. We evaluate the model sensitivities to different input parameters such as the number of solar proxy measurements, number of days in the training set, and region size.

Development and Application of Online Optimization Algorithms

Date: June 27, 2017 at 3pm
Location: Kings River, B52-306 (Note change in time and place!)
Speaker: Xiabiao Huang

Automated tuning is an online optimization process. It can be faster and more efficient than manual tuning and can lead to better performance. It may also substitute or improve upon model based methods. Noise tolerance is a fundamental challenge to online optimization algorithms. We discuss our experience in developing a high efficiency, noise-tolerant optimization algorithm, the RCDS method, and the successful application of the algorithm to various real-life accelerator problems. Experience with a few other online optimization algorithms are also discussed.

Machine Learning at NERSC: Past, Present, and Future

Date: May 16, 2017 at 2pm
Speaker: Prabhat (NERSC)

Modern scientific discovery increasingly relies upon analysis of experimental and observational data. Instruments across a broad range of spatial
scales: telescopes, satellites, drones, genome sequencers, microscopes, particle accelerators, gather increasingly large and complex datasets. In order to ‘infer’ properties of nature, in light of noisy, incomplete measurements, scientists needs access to sophisticated statistics and machine learning tools. In order to address these emerging challenges, NERSC has deployed a portfolio of Big Data technologies on HPC platforms. This talk will review the evolution of Data Analytics tools (statistics, machine learning/deep learning) in the recent past, comment on current scientific use cases and challenges, and speculate on the future of AI-powered scientific discovery.

Optimization for Transportation Efficiency

Date: May 2, 2017 at 2pm
Location: Sycamore Conference Room (040-195)
Speaker: John Fox

Abstract: Plug-in hybrid and all-electric vehicles offer potential to transfer energy demands from liquid petroleum fuels to grid-sourced electricity. We are investigating optimization methods to improve the efficiency and resource utilization of Plug-in Hybrid Electric Vehicles (HEVs). Our optimization uses information about a known or estimated vehicle route to predict energy demands and optimally manage on-board battery and fuel energy resources to maximally use grid-sourced electricity and minimally use petroleum resources for a given route. Our convex optimization method uses a simplified car model to find the optimal strategy over the whole route, which allows for re-optimization on the fly as updated route information becomes available. Validation between the simplified model and a more complete vehicle technology model simulation developed at Argonne National Laboratory was accomplished by “driving” the complete car simulation with the simplified control model. By driving on routes with the same total energy demand but different demand profiles we show fuel efficiency gains of 5-15% on mixed urban/suburban routes compared to a Charge Depleting Charge Sustaining (CDCS) battery controller. The method also allows optimizing the economic lifetime of the vehicle battery by considering the stress on the battery from charge and discharge cycles in the resource optimization.

Detecting Simultaneous Changepoints Across Multiple Data Sequences

Date: April 25, 2017 at 3pm
Location: Kings River, 052-306 (NOTE DIFFERENT LOCATION)
Speaker: Zhou Fan

Abstract: Motivated by applications in genomics, finance, and biomolecular simulation, we introduce a Bayesian model called BASIC for changepoints that tend to co-occur across multiple related data sequences. We design efficient algorithms to infer changepoint locations by sampling from and maximizing over the posterior changepoint distribution. We further develop a Monte Carlo expectation-maximization procedure for estimating unknown prior hyperparameters from data. The resulting framework accommodates a broad range of data and changepoint types, including real-valued sequences with changing mean or variance and sequences of counts or binary observations. We use the resulting BASIC framework to analyze DNA copy number variations in the NCI-60 cancer cell lines and to identify important events that affected the price volatility of S&P 500 stocks from 2000 to 2009.
Low Data Drug Discovery with One-Shot Learning

Date: April 18, 2017 at 2pm
Speaker: Bharath Ramsundar
Location: Berryessa Conference Room (B53-2002) (NOTE DIFFERENT ROOM!)

Abstract: Recent advances in machine learning have made significant contributions to drug discovery. Deep neural networks in particular have been demonstrated to provide significant boosts in predictive power when inferring the properties and activities of small-molecule compounds. However, the applicability of deep learning has been limited by the requirement for large amounts of training data. In this work, we demonstrate how one-shot learning can be used to significantly lower the amounts of data required to make meaningful predictions in drug discovery applications. We introduce a new architecture, the iterative refinement long short-term memory, that, when combined with graph convolutional neural networks, significantly improves learning of meaningful distance metrics over small-molecules. Our models are open-sourced as part of DeepChem, an open framework for deep-learning in drug discovery and quantum chemistry.

Bio: Bharath Ramsundar received a BA and BS from UC Berkeley in EECS and Mathematics and was valedictorian of his graduating class in mathematics. He is currently a PhD student in computer science at Stanford University with the Pande group. His research focuses on the application of deep-learning to drug-discovery. In particular, Bharath is the creator and lead-developer of DeepChem, an open source package that aims to democratize the use of deep-learning in drug-discovery and quantum chemistry. He is supported by a Hertz Fellowship, the most selective graduate fellowship in the sciences.

Learning Particle Physics by Example: Location-Aware Generative Adversarial Networks for Physics Synthesis

Date: Mar. 28, 2017 at 2pm
Speaker: Michela Paganini
Location: Berryessa Conference Room (B53-2002)

Abstract: We provide a bridge between generative modeling in the Machine Learning community and simulated physical processes in High Energy Particle Physics by applying a novel Generative Adversarial Network (GAN) architecture to the production of jet images -- 2D representations of energy depositions from particles interacting with a calorimeter. We propose a simple architecture, the Location-Aware Generative Adversarial Network, that learns to produce realistic radiation patterns from simulated high energy particle collisions. The pixel intensities of GAN-generated images faithfully span over many orders of magnitude and exhibit the desired low-dimensional physical properties (i.e., jet mass, n-subjettiness, etc.). We shed light on limitations, and provide a novel empirical validation of image quality and validity of GAN-produced simulations of the natural world. This work provides a base for further explorations of GANs for use in faster simulation in High Energy Particle Physics.
Models and Algorithms for Solving Sequential Decision Problems under Uncertainty

Date: Mar. 21, 2017 at 2pm
Speaker: Mykel Kochenderfer
Location: Sycamore Conference Room (040-195)

Abstract: Many important problems involve decision making under uncertainty, including aircraft collision avoidance, wildfire management, and disaster response. When designing automated decision support systems, it is important to account for the various sources of uncertainty when making or recommending decisions. Accounting for these sources of uncertainty and carefully balancing the multiple objectives of the system can be very challenging. One way to model such problems is as a partially observable Markov decision process (POMDP). Recent advances in algorithms, memory capacity, and processing power, have allowed us to solve POMDPs for real-world problems. This talk will discuss models for sequential decision making and algorithms for solving them.

Data Programming: A New Framework for Weakly Supervising Machine Learning Models

Date: Mar. 7, 2017 at 2pm
Speaker: Alex Ratner
Location: Sycamore Conference Room (040-195)

Abstract: Today’s state-of-the-art machine learning models require massive labeled training sets—which usually do not exist for real-world applications. Instead, I’ll discuss a newly proposed machine learning paradigm—data programming—and a system built around it, Snorkel, in which the developer focuses on writing a set of labeling functions, which are just scripts that programmatically label data. The resulting labels are noisy, but we model this as a generative process—learning, essentially, which labeling functions are more accurate than others—and then use this to train an end discriminative model (for example, a deep neural network in TensorFlow). Given certain conditions, we show that this method has the same asymptotic scaling with respect to generalization error as directly-supervised approaches. Empirically, we find that by modeling a noisy training set creation process in this way, we can take potentially low-quality labeling functions from the user, and use these to train high-quality end models. We see this as providing a general framework for many weak supervision techniques, and at a higher level, as defining a new programming model for weakly-supervised machine learning systems.
ProxImaL: Efficient Image Optimization using Proximal Algorithms

Date: Feb. 28, 2017 at 2pm
Speaker: Felix Heide
Location: Truckee Room (B52-206) (NOTE ROOM CHANGE!)

Abstract: Computational photography systems are becoming increasingly diverse while computational resources, for example on mobile platforms, are rapidly increasing. As diverse as these camera systems may be, slightly different variants of the underlying image processing tasks, such as demosaicking, deconvolution, denoising, inpainting, image fusion, and alignment, are shared between all of these systems. Formal optimization methods have recently been demonstrated to achieve state-of-the-art quality for many of these applications. Unfortunately, different combinations of natural image priors and optimization algorithms may be optimal for different problems, and implementing and testing each combination is currently a time consuming and error prone process.

ProxImaL is a domain-specific language and compiler for image optimization problems that makes it easy to experiment with different problem formulations and algorithm choices. The language uses proximal operators as the fundamental building blocks of a variety of linear and nonlinear image formation models and cost functions, advanced image priors, and different noise models. The compiler intelligently chooses the best way to translate a problem formulation and choice of optimization algorithm into an efficient solver implementation. In applications to the image processing pipeline deconvolution in the presence of Poisson-distributed shot noise, and burst denoising, we show that a few lines of ProxImaL code can generate a highly-efficient solver that achieves state-of-the-art results. We also show applications to the nonlinear and nonconvex problem of phase retrieval.

Energy-efficient neuromorphic hardware and its use for deep neural networks

Date: Feb. 14, 2017 at 2pm
Speaker: Steve Esser (IBM)
Location: Sycamore Conference Room (040-195)

Abstract: Neuromorphic computing draws inspiration from the brain's structure to create energy-efficient hardware for running neural networks. Pursuing this vision, we created the TrueNorth chip, which embodies 1 million neurons and 256 million configurable synapses in contemporary silicon technology, and runs using under 100 milliwatts. Spiking neurons, low-precision synapses and constrained connectivity are key design factors in achieving chip efficiency, though they stand in contrast to today's conventional neural networks that use high precision neurons and synapses and have unrestricted connectivity. Conventional networks are trained today using deep learning, a field developed independent of neuromorphic computing, and are able to achieve human-level performance on a broad spectrum of recognition tasks. Until recently, it was unclear whether the constraints of energy-efficient neuromorphic computing were compatible with networks created through deep learning. Taking on this challenge, we demonstrated that relatively minor modifications to deep learning methods allows for the creation of high performing networks that can run on the TrueNorth chip. The approach was demonstrated on 8 standard datasets encompassing vision and speech, where near state-of-the-art performance was achieved while maintaining the hardware's underlying energy-efficiency to run at > 6000 frames / sec / watt. In this talk, I will present an overview of the TrueNorth chip, our methods to train networks for this chip and a selection of performance results.

http://science.sciencemag.org/content/345/6197/668
http://www.pnas.org/content/113/41/11441.abstract
Locating Features in Detector Images with Machine Learning

Date: Feb. 7, 2017 at 2pm
Speaker: David Schneider
Location: Sycamore Conference Room (040-195)

Abstract: Often analysis at LCLS involves image processing of large area detectors. One goal is to find the presence, and location of certain features in the images. We'll look at several approaches to locating features using machine learning. The most straightforward is learning from training data that includes feature locations. When location labels are not in the training data, techniques like guided back propagation, relevance propagation, or occlusion can be tried. We'll discuss work on applying these approaches. We'll also discuss ideas based on generative models like GAN's (Generative Adversarial Networks) or VAE's (Variational Auto Encoders).

Tractable quantum leaps in battery materials and performance via machine learning

Date: Jan. 17, 2017
Speaker: Austin Sendek

Abstract: The realization of an all solid-state lithium-ion battery would be a tremendous development towards remedying the safety issues currently plaguing lithium-ion technology. However, identifying new solid materials that will perform well as battery electrolytes is a difficult task, and our scientific intuition on whether a material is a promising candidate is often poor. Compounding on this problem is the fact that experimental measurements of performance are often very time- and cost intensive, resulting in slow progress in the field over the last several decades. We seek to accelerate discovery and design efforts by leveraging previously reported data to train learning algorithms to discriminate between high- and poor performance materials. The resulting model provides new insight into the physics of ion conduction in solids and evaluates promise in candidate materials nearly one million times faster than state-of-the-art methods. We have coupled this new model with several other heuristics to perform the first comprehensive screening of all 12,000+ known lithium-containing solids, allowing us to identify several new promising candidates.
Deep Learning and Computer Vision in High Energy Physics

Date: Dec 6, 2016
Speaker: Michael Kagan
Location: Kings River 306, B52

Abstract: Recent advances in deep learning have seen great success in the realms of computer vision, natural language processing, and broadly in data science. However, these new ideas are only just beginning to be applied to the analysis of High Energy Physics data. In this talk, I will discuss developments in the application of computer vision and deep learning techniques to the analysis and interpretation of High Energy Physics data, with a focus on the Large Hadron Collider. I will show how these state-of-the-art techniques can significantly improve particle identification, aid in searches for new physics signatures, and help reduce the impact of systematic uncertainties. Furthermore, I will discuss methods to visualize and interpret the high level features learned by deep neural networks that provide discrimination beyond physics derived variables, adding a new capability to understand physics and to design more powerful classification methods in High Energy Physics.

Links to papers discussed:
https://arxiv.org/abs/1511.05190
https://arxiv.org/abs/1611.01046

Label-Free Supervision of Neural Networks with Physics and Domain Knowledge
Abstract: In many machine learning applications, labeled data is scarce and obtaining more labels is expensive. We introduce a new approach to supervising neural networks by specifying constraints that should hold over the output space, rather than direct examples of input-output pairs. These constraints are derived from prior domain knowledge, e.g., from known laws of physics. We demonstrate the effectiveness of this approach on real world and simulated computer vision tasks. We are able to train a convolutional neural network to detect and track objects without any labeled examples. Our approach can significantly reduce the need for labeled training data, but introduces new challenges for encoding prior knowledge into appropriate loss functions.

Can machine learning teach us physics? Using Hidden Markov Models to understand molecular dynamics.

Abstract: Machine learning algorithms are often described solely in terms of their predictive capabilities, and not utilized in a descriptive fashion. This "black box" approach stands in contrast to traditional physical theories, which are generated primarily to describe the world, and use prediction as a means of validation. I will describe one case study where this dichotomy between prediction and description breaks down. While attempting to model protein dynamics using master equation models — known in physics since the early 20th century — it was discovered that there was a homology between these models and Hidden Markov Models (HMMs), a common machine learning technique. By adopting fitting procedures for HMMs, we were able to model large scale simulations of protein dynamics and interpret them as physical master equations, with implications for protein folding, signal transduction, and allosteric modulation.
On-the-fly unsupervised discovery of functional materials

Date: Aug 31, 2016
Speaker: Apurva Mehta

Abstract: Solutions to many of the challenges facing us today, from sustainable generation and storage of energy to faster electronics and cleaner environment through efficient sequestration of pollutants, is enabled by the rapid discovery of new functional materials. The present paradigm based on serial experimentation and serendipitous discoveries takes decades from initiation of a new search for a material to marketplace deployment of a device based on it. Major road-blocks in this process arise from heavy dependence on humans to transfer knowledge between interdependent steps. For example, currently humans look for patterns in current knowledge-bases, build hypotheses, plan and conduct experiments, evaluate results and extract knowledge to create the next hypothesis. The recent insight, emerging from the materials genome initiative, is that rapid transfer of information between hypothesis building, experimental testing and scale-up engineering can reduce the time and cost of material discovery and deployment by half. Humans, though superb at pattern recognition and complex decision making, are too slow and the major challenge in this new discovery paradigm is to reliably extract high-level actionable information from large and noisy data on-the-fly with minimal human intervention. In here, I will discuss some of the strategies and challenges involved in construction of unsupervised machines that perform these tasks on high throughput and large volume X-ray spectroscopic and scattering data sets.

Machine Learning and Optimization to Enhance the FEL Brightness

Date: Aug 17, 2016
Speakers: Anna Leskova, Hananiel Setiawan, Tanner M. Worden, Juhao Wu

Abstract: Recent studies on enhancing the FEL brightness via machine learning and optimization will be reported. The topics are tapered FEL and improved SASE. The existing popular machine learning approaches will be reviewed and selected based on the characteristics of different tasks. Numerical simulation and preliminary LCLS experiment results will be presented.
Automated tuning at LCLS using Bayesian optimization

Date: July 6, 2016
Speaker: Mitch McIntire
Location: Truckee Room, B52-206 T

Abstract: The LCLS free-electron laser has historically been tuned by hand by the machine operators. Existing tuning procedures account for hundreds of hours of machine time per year, and so efforts are underway to reduce this tuning time via automation. We introduce an approach for automated tuning using Bayesian optimization with statistical models called Gaussian processes. Initial testing has shown that this method can substantially reduce tuning time and is potentially a significant improvement on existing automated tuning methods. In this talk I'll describe Bayesian optimization and Gaussian processes and share some details and insights of implementation, as well as our preliminary results.

Using Deep Learning to Sort Down Data

Date: June 15, 2016
Speaker: David Schneider

Abstract:
We worked on data from a two color experiment (each pulse has two bunches at different energy levels). The sample reacts differently depending on which of the colors lased and the energy in the lasing. We used deep learning to train a convolutional neural network to predict these lasing and energy levels from the xtcav diagnostic images. We then sorted down the data taken of the sample based on these values and identified
differences in how the sample reacted. Scientific results from the experiment will start with an analysis of these differences. We used guided back propagation to see what the neural network identified as important and were able to obtain images that isolate the lasing portions of the xtcav images.