New York Scientific Data Summit 2020:

Data-Driven Discovery in Science and Industry

Organized by:

BROOKHAVEN NATIONAL LABORATORY

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

FLATIRON INSTITUTE Division of Simons Foundation

NYU

RUTGERS THE STATE UNIVERSITY OF NEW JERSEY

Stony Brook University

TEXAS The University of Texas at Austin

Agenda

Virtual Conference via Zoomgov.com Note: *All times are shown in EDT*

TIME	SESSIONS/SPEAKERS
TUESDAY, OCTOBER 20 - CLIMATE	
12:00 – 12:05 PM	*Opening,* **Meifeng Lin** (Brookhaven National Laboratory)
12:05 – 12:10 PM	*Welcome Remarks,* **Kerstin Kleese van Dam** (Director, Computational Science Initiative, Brookhaven National Laboratory)
Session #1, Session Chair: **Nathan Urban** (Brookhaven National Laboratory)	
12:10 – 1:10 PM	***Keynote:*** *Climate, Carbon, and Water: Tracking and Anticipating Human Impacts,* **Anna Michalak** (Carnegie Institution for Science)
1:10 – 1:55 PM	*Quantifying Parameter Uncertainty Within a Climate Mode*l, **Oliver Dunbar** (California Institute of Technology)
1:55 – 2:00 PM	*Virtual Coffee Break*
Session #2, Session Chair: **Frank Alexander** (Brookhaven National Laboratory)	
2:00 – 2:30 PM	*Lightning Talk Session #1*
2:30 – 3:15 PM	*Numerical Methods for Predicting Coastal Flooding with Uncertainty,* **Kyle Mandli** (Columbia University)
3:15 – 4:15 PM	*Panel Discussion* - Challenges and Opportunities for Climate and Environmental Research Moderator: **Allison McComiskey** (Brookhaven National Laboratory) Panelists: **Anna Michalak** (Carnegie Institute for Science), **Oliver Dunbar** (California Institute of Technology), **Kyle Mandli** (Columbia University)
4:15 – 5:00 PM	*Virtual Social Hour via Zoom*
WEDNESDAY, OCTOBER 21 - CRITICAL INFRASTRUCTURE/MANUFACTURING	
Session #3, Session Chair: **Lakshminarayanan Subramanian** (New York University)	
12:00 – 1:00 PM	***Keynote:*** *Risk in Power Grids,* **Daniel Bienstock** (Columbia University)
1:00 – 1:45 PM	*Cyber-Physical Systems for Smart Cities*: A Mobility Perspective, **Desheng Zhang** (Rutgers University)
1:45 – 2:00 PM	*Virtual Coffee Break*
Session #4, Session Chair: **Robert Harrison** (Stony Brook University)	
2:00 – 2:15 PM	*Lightning Talk Session #2*
2:15 – 3:00 PM	*Computational Modeling at GE*, **Richard Arthur** (General Electric Research)
3:00 – 4:00 PM	*Panel Discussion - Challenges and Opportunities in Critical Infrastructure and Manufacturing* Moderator: **Karen Willcox** (University of Texas at Austin) Panelists: **Daniel Bienstock** (Columbia University), **Desheng Zhang** (Rutgers University), **Richard Arthur** (General Electric Research)
4:00 – 5:00 PM	*Virtual Social Hour via Zoom*

Virtual Conference via Zoomgov.com Note: *All times are shown in EDT*

TIME	SESSIONS/SPEAKERS
THURSDAY, OCTOBER 22 – HEALTH AND MEDICINE	
Session #5, Session Chair: **Frank Alexander** (Brookhaven National Laboratory)	
12:00 – 1:00 PM	***Keynote:*** *Re-Engineering the Future of Health with Predictive Models*, **Grace C.Y. Peng** (National Institutes of Health)
1:00 – 1:45 PM	*Applications of AI in Cancer Research – Preparations, Progress and Predictions*, **Eric Stahlberg** (Frederick National Laboratory for Cancer Research)
1:45 – 2:00 PM	*Virtual Coffee Break*
Session #6, Session Chair: **Shantenu Jha** (Rutgers University/Brookhaven National Laboratory)	
2:00 – 2:15 PM	*Lightning Talk Session #3*
2:15 – 3:00 PM	*Data-driven Modeling of COVID-19: Lessons Learned*, **Ellen Kuhl** (Stanford University)
3:00 – 3:45 PM	*Computational Microscopy of SARS-CoV-2*, **Rommie Amaro** (University of California, San Diego)
3:45 – 4:00 PM	*Virtual Coffee Break*
4:00 – 5:00 PM	*Panel Discussion – Challenges and Opportunities in Computational Medicine/Health* Moderator: **Arvind Ramanathan** (Argonne National Laboratory) Panelists: **Grace C.Y. Peng** (National Institutes of Health), **Eric Stahlberg** (Frederick National Laboratory for Cancer Research), **Ellen Kuhl** (Stanford University)
FRIDAY, OCTOBER 23 – CROSS-CUTTING TOPICS	
Session #7, Session Chair: **Andrew Millis** (Flatiron Institute)	
12:00 – 1:00 PM	***Keynote:*** *Drug Design and Discovery for SARS-CoV2 by Integrating Artificial Intelligence and Physics-based Models,* **Arvind Ramanathan** (Argonne National Laboratory)
1:00 – 1:45 PM	*The Quantum Many-body Problem as a Challenge for Machine Learning Methods*, **Giuseppe Carleo** (EPFL, Switzerland)
1:45 – 2:00 PM	*Virtual Coffee Break*
Session #8, Session Chair: **Qiang Du** (Columbia University)	
2:00 – 2:15 PM	*Lightning Talk Session #4*
2:15 – 3:00 PM	*The Non-uniform FFT and its Applications,* **Leslie Greengard** (New York University)
3:00 – 3:45 PM	*HPC+AI: Pushing Molecular Dynamics Simulation with Ab Initio Accuracy to 100 Million Atoms*, **Lin Lin** (University of California, Berkeley)
3:45 – 4:00 PM	*Virtual Coffee Break*
4:00 – 5:00 pm	*Concluding Discussions* Moderator: **Frank Alexander** (Brookhaven National Laboratory)

Agenda

Virtual Conference via Zoomgov.com Note: *All times are shown in EDT*

OCTOBER 20, 2020

2:00 - 2:30 PM — *Lightning Talk Session #1*

1. *Artificial Intelligence for the Accuracy and Speed of Multiscale Modeling* (**Changnian Han, Peng Zhang, Jawaad Sheriff, Guojing Cong, Danny Bluestein, Yuefan Deng**)
2. *Exploring Sensitivity of ICF Outputs to Design Parameters in Experiments using Machine Learning* (**Julia Nakhleh, M. Giselle Fernández-Godino, Michael J. Grosskopf, Brandon M. Wilson, John Kline, Gowri Srinivasan**)
3. *Multitask Learning and Multi-Armed Bandit-Based Bayesian Optimization for High-Performance Computing Applications* (**Yang Liu, Xinran Zhu, Wissam M. Sid-Lakhdar, Osni A. Marques, Xiaoye S. Li, James W. Demmel**)
4. *Alzheimer's Disease Prognosis Using Graph Convolutional Neural Networks* (**Animesh Ghose, Shinjae Yoo, Ai Kagawa**)
5. *Examining graph topology using quantum walks* (**Raffaele Miceli, Michael McGuigan**)

OCTOBER 21, 2020

2:00 - 2:15 PM — *Lightning Talk Session #2*

1. *Identifying Complex Physics Relationships using Sparse Matrix Decomposition to Inform Plasma Fusion Design* (**M. Giselle Fernández-Godino, Julia B. Nakhleh, Michael J. Grosskopf, Brandon M. Wilson, John Kline, Gowri Srinivasan**)
2. *Using Unstructured Data to Improve Homelessness and Suicide Prediction* (**Rafael Zamora-Resendiz, Destinee Morrow, Silvia Crivelli**)
3. *e3nn: 3D Euclidean symmetry equivariant neural networks – Learning from the geometry and geometric tensors of physical systems* (**Tess E. Smidt, Mario Geiger, Benjamin Kurt Miller, Kostiantyn Lapchevskyi**)

Virtual Conference via Zoomgov.com Note: *All times are shown in EDT*

OCTOBER 22, 2020

2:00 – 2:15 PM, *Lightning Talk Session #3*

1. *Thermal Analysis of the SARS-CoV-2 Spike Glycoprotein by in silico and in vitro Experiments on the Supercomputers* (**Meichen Song, Fan Yang, Miriam Rafailovich, Marcia Simon, Yuefan Deng, Peng Zhang**)
2. *Millisecond Multiscale Simulations of Multi-Platelet Aggregation in Shear Flow on Supercomputers* (**Yicong Zhu, Changnian Han, Peng Zhang, Guojing Cong, Danny Bluestein, Yuefan Deng**)
3. *AI Meets HPC: Learning the Platelet Dynamics from In Vitro and In Silico Experiments* (**Ziji Zhang, Changnian Han, Peng Zhang, Guojing Cong, Jawaad Sheriff, Danny Bluestein, Yuefan Deng**)

OCTOBER 23, 2020

2:00 – 2:15 PM *Lightning Talk Session #4*

1. *The pH-varying Conformational States of SARS-CoV-2 Spike Glycoprotein* (**Ziyuan Niu, Yuefan Deng, Zhang Peng**)
2. *Simulating the Ground State Energies of Molecules Using IBM's Quantum Emulators* (**Mohammad Hassan, Michael McGuigan**)
3. *Quantum Computations of Dark Energy Models* (**Juan Varela, Michael McGuigan**)

Day 1 Talks: October 20, 2020

KEYNOTE SPEAKER: DR. ANNA MICHALAK, CARNEGIE INSTITUTION FOR SCIENCE

Climate, Carbon, and Water–Tracking and Anticipating Human Impacts

Abstract:

Human action, at the individual to the global scale, is inextricably linked to the world's climate, carbon, and water. In this talk, Dr. Michalak will provide examples from her group's recent work that fall under two areas within this broad topic. First, she will discuss statistical and computational tools for characterizing the cycling and emissions of greenhouse gases at urban to global scales that make it possible to track and anticipate changes in carbon emissions and uptake. Second, she will give an overview of recent progress in quantifying climate change impacts on water quality in freshwater and coastal systems in support of water sustainability goals. In both of these areas, highly data-driven approaches that account for spatiotemporal variability and can integrate disparate data types and sources are required for successful mechanistic understanding of the interactions among human activity, climate, carbon, and water.

Bio:

Dr. Anna M. Michalak is the Director of the Department of Global Ecology of the Carnegie Institution for Science and Professor in the Department of Earth System Science at Stanford University. Previously, she was the Frank and Brooke Transue Faculty Scholar and Associate Professor at the University of Michigan in Ann Arbor, with appointments in the Department of Civil and Environmental Engineering and Department of Atmospheric Oceanic and Space Sciences. She holds a doctorate and M.S. in Civil and Environmental Engineering from Stanford University and a B.Sc.(Eng.) in Environmental Engineering from the University of Guelph (Canada). She is the lead author of the U.S. Carbon Cycle Science Plan, Chair of the Scientific Advisory Board for the European Integrated Carbon Observation System, Member of the National Academies Committee on Earth Sciences and Applications from Space, Member of the NASA Earth Science Senior Review Subcommittee for operating missions, Member of the Orbiting Carbon Observatory 2 satellite Science Team, and a former Editor of the journal Water Resources Research. She is the recipient of the Presidential Early Career Award for Scientists and Engineers (nominated by NASA), the NSF CAREER award, and the Leopold Fellowship in environmental leadership.

SPEAKER: KYLE T. MANDLI, COLUMBIA UNIVERSITY

Numerical Methods for Predicting Coastal Flooding with Uncertainty

Abstract:
Coastal hazards related to strong storms are one of the most ubiquitous hazards to coastal communities throughout the world. In particular, storm surge, the rise of the sea surface in response to wind and pressure forcing from these storms, can have a devastating effect at the coastline. Changes to the climate only compound the need for predictive tools that can also handle the uncertainty inherent in climate predictions. Computational approaches are the "go-to" tool for dealing with these difficulties, but it is a far from trivial problem. The problem is inherently multiscale; the uncertainties difficult to represent; and the hyperbolic structure of the most well-used set of representative equations, shallow water equations, presents additional issues when looking for low-rank approximations. This talk will describe many of these difficulties, where they come from, and what research efforts are attempting to address them. This includes extensions to shallow water equations, techniques for representing the uncertainty and measuring sensitivity in the problem, and finally how reduced-order modeling may help to produce low-rank approximations to hyperbolic equations in general.

Bio:
Kyle Mandli is Associate Professor of Applied Mathematics in Columbia University's Applied Physics and Applied Mathematics department and affiliated with the Columbia Data Science Institute. Before Columbia, he was at the University of Texas at Austin, where he was a Research Associate at the Institute for Computational and Engineering Sciences in the computational hydraulics group. He received his doctorate in Applied Mathematics in 2011 from the University of Washington studying multi-layered flow as it applies to storm-surge simulation. His research interests involve the computational and analytical aspects of geophysical shallow mass flows, such as storm-surge, tsunamis, and other coastal flooding. This also includes the development of advanced computational approaches, such as adaptive mesh refinement; leveraging novel computational technologies, such as accelerators; and the application of good software development practices as applied more generally to scientific and engineering software.

SPEAKER: OLIVER R. DUNBAR, CALIFORNIA INSTITUTE OF TECHNOLOGY

Quantifying Parameter Uncertainty within a Climate Model

Abstract:

Current state-of-the-art climate models produce uncertain predictions, as evidenced by the variability in competing models. However, they typically are ill-equipped to quantify this uncertainty. Where does the uncertainty come from? Climate models necessarily contain simplified physical schemes used to represent small-scale dynamics or poorly understood physics. The schemes depend upon parameters that are calibrated (often by hand) to fit data, although there may be a range of parameters that feasibly produce a given piece of data. In climate models, this spread of parameters used for convection and turbulent parameterizations is the dominant cause in the spread of resulting decadal predictions. Therefore, it is essential to quantify it. Unfortunately, this task is far more computationally intensive than parameter calibration and, historically, has been out of reach of climate models. However, applying the new methodology Calibrate-Emulate-Sample (CES) to time-averaged climate statistics makes this possible. CES is based on three steps: a first Calibration step, which takes the climate model as a black box input and is well adapted to high performance computing architectures. A second Emulation step automates, smooths, and speeds up calculation of the black box climate model by several orders of magnitude by making use of Gaussian processes (a machine learning tool). A final Sampling step may be applied using standard methods from computational statistics to quantify uncertainty in the calibration. This talk will consider an idealized aquaplanet general circulation model (GCM) using CES to perform uncertainty quantification on the closure parameters for convection.

Bio:

Oliver R. Dunbar is a postdoctoral scholar based at Caltech, working with Tapio Schneider and Andrew Stuart as part of the CliMA (Climate Modelling Alliance) project. He works on data assimilation, uncertainty quantification, and optimal experimental design arising in climate modeling. In 2017, Dunbar earn a doctorate in Mathematics and Statistics from the University of Warwick with work on modeling multi-phase fluid flow. In 2018, he completed a postdoc at Warwick, working on inverse problems found in seismic tomography where his research interest moved toward Bayesian inverse problems and uncertainty quantification.

SPEAKER: DANIEL BIENSTOCK, COLUMBIA UNIVERSITY

Risk in Power Grids

Abstract:
Risk in power grids concerns the inability to safely deliver power to customers when it is needed. Moving forward, such risks will be exacerbated by several converging trends: (a) the very large cost of expanding grids and replacing aging transmission assets; (b) increasing reliance on variable power sources, such as renewables; and (c) an expanding role of smart loads, batteries, and other local or customer-based resources, such as solar panels and local generation. It is worth noting that (b) and (c) are, to some extent, a corollary of (a). In this talk, Dr. Bienstock will present a brief overview of current power grid operations and describe how risk can be detailed using precise mathematical terms. Finally, he will present current investigation trends that are likely to change power grid practices over the coming decade.

Bio:
Daniel Bienstock is Liu Family Professor of Operations Research at Columbia University with affiliate appointments in Applied Mathematics and Electrical Engineering. His work focuses on theory and computational aspects of optimization, particularly optimization under algebraic constraints. A parallel stream of research concerns mathematical problems related to power grid operations. He is an INFORMS Fellow and holds a doctorate from the Massachusetts Institute of Technology.

SPEAKER: DESHENG ZHANG, RUTGERS UNIVERSITY

Cyber-Physical Systems for Smart Cities: A Mobility Perspective

Abstract:

For the first time ever, more people are living in urban areas than in rural ones. Based on this inevitable urbanization, Dr. Zhang's research group is addressing sustainability challenges related to urban mobility (e.g., energy consumption and traffic congestion) by data-driven modeling and applications with a Cyber-Physical-Systems (CPS) approach in the vision of smart cities. In this talk, Dr. Zhang will focus on mobility modeling and resultant applications based on large-scale cross-domain CPS, e.g., cellular networks, payment systems, social networks, and transportation systems (including electric vehicles, taxis, buses, subway, private vehicles, and Ubers). He will show how cross-domain CPS systems can be collaboratively used to capture real-time urban mobility via a set of model integration techniques. Then, he will illustrate how the captured mobility can be used to design various urban mobile services to close the "loop," from urban-scale ridesharing to for-hire vehicle dispatching, electric toll collection management, electric-vehicle charging recommendation, and emergency response under mobility anomaly. Finally, Dr. Zhang will present some research challenges related to future cross-domain CPS in the context of the smart cities research.

Bio:

Desheng Zhang is an Assistant Professor in the Department of Computer Science at Rutgers University. Previously, he was offered the Senseable City Consortium Postdoctoral Fellowship from the Massachusetts Institute of Technology and awarded his doctorate in Computer Science from the University of Minnesota. He is broadly interested in mobile sensing, ubiquitous computing, and cyber-physical systems with a focus on sensing, measurement, and applications for cross-domain mobile systems including cellular networks, Wi-Fi networks, mobile payment systems, taxis, buses, subways, bikes, personal vehicles, electric vehicles, trucks, and social networks from a data-driven perspective (details: https://www.cs.rutgers.edu/~dz220/).

SPEAKER: RICHARD (RICK) ARTHUR, GENERAL ELECTRIC RESEARCH

Computational Modeling at GE

Abstract:
To compete in multiple global, high-tech industries, General Electric (GE) scientists and engineers employ computational methods for innovation, discovery, and optimization across the product life cycle. GE leverages scalable analytics, simulations, and machine learning to model critical infrastructure systems, machines, and services in areas such as healthcare, renewable and gas power generation, avionics, and jet propulsion.

Bio:
Rick Arthur leads pathfinding efforts in applying computational methods at GE Research, supporting products and services that span diverse industrial sectors, such as healthcare, air and rail transportation, media, finance, defense, and energy. He represents GE in government policy and project discussions, including as co-chair of the U.S. Council on Competitiveness Advanced Computing Roundtable and as a member of the Exascale Computing Project Industry Council technical advisors. He holds a B.S. and M. Eng. in Computer Engineering and an MBA and is a Senior Member of ACM.

SPEAKER: ROMMIE E. AMARO, UNIVERSITY OF CALIFORNIA, SAN DIEGO

Computational Microscopy of SARS-CoV-2

Abstract:
Dr. Amaro will discuss her lab's efforts working with collaborators to understand the SARS-CoV-2 virus in atomic detail. Their goals are to better understand molecular recognition of the virus and host cell receptors and antibody binding and design, as well as augment the search for novel therapeutics.

Bio:
Rommie E. Amaro holds the Distinguished Professorship in Theoretical and Computational Chemistry at the Department of Chemistry and Biochemistry at the University of California, San Diego (UCSD). She received her B.S. in Chemical Engineering (1999) and her doctorate in Chemistry (2005) from the University of Illinois at Urbana-Champaign. Dr. Amaro was a National Institutes of Health (NIH) postdoctoral fellow with Prof. J. Andrew McCammon at UCSD from 2005-2009 and started her independent lab in 2009. She is the recipient of an NIH New Innovator Award, the Presidential Early Career Award for Scientists and Engineers, the ACS COMP OpenEye Outstanding Junior Faculty Award, the ACS Kavli Foundation Emerging Leader in Chemistry, and the Corwin Hansch Award. Her scientific interests reside at the intersection of computer-aided drug discovery and biophysical simulation. Her scientific vision revolves around expanding the range and complexity of molecular constituents represented in such simulations, development of novel multiscale methods for elucidating their time-dependent dynamics, and discovery of novel chemical matter-controlling biological function.

SPEAKER: ERIC STAHLBERG, FREDERICK NATIONAL LABORATORY FOR CANCER RESEARCH

Applications of AI in Cancer Research–Preparations, Progress, and Predictions

Abstract:
The talk will focus on recent efforts to develop cancer research insights, accelerate development of treatments, and transform clinical care for cancer patients through the application of artificial intelligence (AI), data science, and advanced computing. The presentation will include discussion of important formative steps, as well as lessons learned, to realize the tremendous potential for transformative insight for predictive oncology at the convergence of growing volumes of cancer data, large-scale and advanced computing, and new avenues for collaboration. Examples will be drawn from collaborative efforts involving the National Cancer Institute, Frederick National Laboratory, the U.S. Department of Energy, and others. The talk also will look ahead to emerging and new opportunities at this exciting and rapidly evolving frontier in data science.

Bio:
Eric Stahlberg is the director of Biomedical Informatics and Data Science (BIDS) at the Frederick National Laboratory (FNL) for Cancer Research where he heads the directorate responsible for delivering cancer research data to the community, advancing biomedical computational science, and pursuing strategic initiatives in data science. His efforts fostering collaborations between the National Cancer Institute and the Department of Energy, including the Joint Design of Advanced Computing Solutions for Cancer and Accelerating Therapeutics for Opportunities in Medicine (ATOM), have led to significant progress in understanding opportunities to advance cancer research through the use of AI and high-performance computing. He also has led efforts to develop a scalable data management environment supporting rapidly growing volumes of cancer research data at FNL. Beginning his professional career with a post-doctoral appointment at Argonne National Laboratory, Dr. Stahlberg has worked in industry, academia, and non-profit sectors, focusing on collaborative initiatives involving life science applications at the intersection of data and computation. He continues these efforts as co-organizer of the Computational Approaches for Cancer workshop at the annual SuperComputing conference, co-lead for ATOM, and in seeking out new opportunities where technology, data, and computing open new avenues for scientific insight in biomedical science.

KEYNOTE SPEAKER: GRACE C.Y. PENG, NATIONAL INSTITUTES OF HEALTH

Re-Engineering the Future of Health with Predictive Models

Abstract:

The National Institute of Biomedical Imaging and Bioengineering (NIBIB)'s mission is to improve human health by leading the development and accelerating the application of biomedical technologies. NIBIB is committed to integrating the physical and engineering sciences with the life sciences to advance basic research and medical care. Since 2003, NIBIB has served in a leadership role to promote multiscale modeling of biomedical, biological, and behavioral systems through the Interagency Modeling and Analysis Group (IMAG) of program officers from more than 10 government agencies. In turn, funding initiatives from IMAG and its partnering agencies have supported and coordinated the Multiscale Modeling Consortium (MSM) since 2006. The IMAG MSM Consortium currently has over 100 multiscale-modeling-related projects, and its mission is to grow the field of multiscale modeling in biomedical, biological, and behavioral systems by: (1) promoting multidisciplinary scientific collaboration among multiscale modelers; (2) encouraging future generations of multiscale modelers; (3) developing accurate methods and algorithms to cross the interface between multiple spatiotemporal scales; (4) promoting model sharing and development of reusable multiscale models; and (5) disseminating models and insights arrived from them to the larger biomedical, biological, and behavioral research community. The MSM Consortium is actively addressing many pressing issues facing the multiscale modeling community (see the IMAG wiki for reference: https://www.imagwiki.nibib.nih.gov/). This talk will describe the importance of predictive modeling in biomedical research and current challenges for implementation, as well as introduce the latest modeling, simulation, and analysis initiatives of the NIH, NIBIB, IMAG and the MSM Consortium.

Bio:

Grace C.Y. Peng is the Director of Mathematical Modeling, Simulation and Analysis at the National Institute of Biomedical Imaging and Bioengineering (NIBIB) within the National Institutes of Health (NIH) in the U.S. Department of Health and Human Services (DHHS). In this capacity, she has programmatic oversight of extramural activities in these areas. Dr. Peng received a B.S. in electrical engineering from the University of Illinois and her M.S. and doctorate in biomedical engineering from Northwestern University. She performed postdoctoral and faculty research in the department of Neurology at the Johns Hopkins University. In 2000, she became the Clare Boothe Luce Professor of Biomedical Engineering at the Catholic University of America.

SPEAKER: ELLEN KUHL, STANFORD UNIVERSITY

Data-driven Modeling of COVID-19: Lessons Learned

Abstract:

Understanding the outbreak dynamics of COVID-19 through the lens of mathematical models is an elusive but significant goal. Within less than a year, the COVID-19 pandemic has resulted in more than 30 million reported cases across 188 countries with more than 1 million deaths worldwide. Unlike any other disease in history, COVID-19 has generated an unprecedented volume of data—well documented, continuously updated, and broadly available to the general public. Yet, the precise role of mathematical modeling in providing quantitative insight into the COVID-19 pandemic remains a topic of ongoing debate. Here, Dr. Kuhl will discuss the lessons learned from six month of modeling COVID-19, highlighting the early success of classical models for infectious diseases and showing why these models fail to predict the current outbreak dynamics of COVID-19. She will illustrate how data-driven modeling can integrate classical epidemiology modeling and machine learning to infer critical disease parameters from reported case data—in real time—to make informed predictions and guide political decision making. She will critically discuss questions that these models can and cannot answer and showcase controversial decisions around the early outbreak dynamics, outbreak control, and exit strategies. This summary will stimulate discussion within the modeling community and help provide guidelines for robust mathematical models to understand and manage the COVID-19 pandemic.

Bio:

Ellen Kuhl is the Robert Bosch Chair of Mechanical Engineering at Stanford University. She is a Professor of Mechanical Engineering and, by courtesy, Bioengineering. She received her doctorate from the University of Stuttgart in 2000 and her Habilitation from the University of Kaiserslautern in 2004. Her area of expertise is Living Matter Physics, the design of theoretical and computational models to simulate and predict the behavior of living structures. Dr. Kuhl has published more than 200 peer-reviewed journal articles and edited two books. She is an active reviewer for more than 20 journals at the interface of engineering and medicine and an editorial board member of seven international journals in her field. She is the current Chair of the U.S. National Committee on Biomechanics and a Member-Elect of the World Council of Biomechanics. She is a Fellow of the American Society of Mechanical Engineers and of the American Institute for Mechanical and Biological Engineering. She received the National Science Foundation Career Award in 2010, was selected as Midwest Mechanics Seminar Speaker in 2014, and received the Humboldt Research Award in 2016. Dr. Kuhl is an All-American triathlete on the Wattie Ink. Elite Team; a multiple Boston, Chicago, and New York marathon runner; and a Kona Ironman World Championship finisher.

KEYNOTE SPEAKER: ARVIND RAMANATHAN, ARGONNE NATIONAL LABORATORY

Drug Design and Discovery for SARS-CoV2 by Integrating Artificial Intelligence and Physics-based Models

Abstract:

Within the Data Science and Learning Division at Argonne National Laboratory, work addresses both the fundamental biological mechanisms of the SARS-CoV-2 virus and the disease while simultaneously targeting the entire viral proteome to identify potential therapeutics. Machine learning (ML), deep learning (DL), and artificial intelligence (AI) techniques are being designed to: (i) identify and build accurate three-dimensional structural models of the SARS-CoV-2 proteome by integrating experimental structural and systems biology datasets; (ii) accelerate adaptive conformational sampling of the viral proteins to potentially identify novel binding sites/pockets that can be targeted by compound libraries; (iii) rapidly filter, rank, and search for small molecules across widely available chemical libraries and to integrate virtual screening with experimental high-throughput screening; and (iv) characterize the virus' evolutionary "traits," including identification of epitopes and the viral genome that can be targeted for vaccine design. The immediate impact of current research is an ecosystem of open-source AI/ML tools and conventional physics-based simulations that can accelerate timely response for treating such pandemics. Argonne scientists have made significant progress across the aforementioned goals, including the release of more than 60 terabytes of machine-readable data for various open-source chemical compound libraries (https://2019-ncovgroup.github.io/data/), development of scalable AI/ML methods for rapidly filtering the chemical space that can bind specifically to viral protein targets, and adaptive conformational sampling using molecular dynamics (MD) simulations to quantify the stability and binding of AI-predicted compounds to specific targets. The outputs from physics-based models are used iteratively to improve the prediction capabilities of these AI/ML approaches, successively improving the overall yield of drug candidates that can be refined further using biochemical and biological assays. Together, this integrated approach provides insights into how the overall drug-design and discovery process can be improved for emerging pandemics.

Bio:

Arvind Ramanathan is a computational biologist in the Data Science and Learning Division at Argonne National Laboratory and a senior scientist at the University of Chicago Consortium for Advanced Science and Engineering (CASE). His research interests are at the intersection of data science, high-performance computing, and biological/biomedical sciences.

SPEAKER: LESLIE GREENGARD, NEW YORK UNIVERSITY

The Non-uniform FFT and its Applications

Abstract:
The nonuniform fast Fourier transform (NUFFT) arises in a variety of application areas, including medical imaging, astronomy, X-ray scattering, and the numerical solution of partial differential equations. In its most general form, it takes as input an irregular sampling of a function and seeks to compute its Fourier transform at a nonuniform sampling of frequency locations. Like the classical FFT, the amount of work required is of the order O(N log N), where N denotes the number of sampling points in both the physical and spectral domains. In this lecture, Dr. Greengard will present the essential ideas underlying the algorithm, describe recent improvements in both the algorithm and software implementation, and illustrate its application to problems in magnetic resonance imaging, wave scattering, and the solution of the time-dependent Schrödinger equation.

Bio:
Leslie Greengard received his B.A. degree in Mathematics from Wesleyan University in 1979 and his doctorate in Computer Science and M.D. degree from Yale University in 1987. He has been a member of the faculty at the Courant Institute of Mathematical Sciences, NYU since 1989 and was Director of the Institute from 2006-2011. He also is Professor of Electrical and Computer Engineering at NYU's Tandon School of Engineering and presently serves as Director of the Center for Computational Mathematics at the Flatiron Institute, a division of the Simons Foundation. Greengard, together with V. Rokhlin, developed the Fast Multipole Method (FMM) for problems in gravitation, electrostatics, and electromagnetics. For their work, they received the Steele Prize from the American Mathematical Society in 2001. Much of Greengard's research has been aimed at the development of high-order accurate integral equation-based methods for partial differential equations in complex geometry. He is a member of the National Academy of Sciences, National Academy of Engineering, and American Academy of Arts and Sciences.

SPEAKER: LIN LIN, UNIVERSITY OF CALIFORNIA, BERKELEY

HPC+AI: Pushing Molecular Dynamics Simulation with Ab Initio Accuracy to 100 Million Atoms

Abstract:

For 35 years, ab initio molecular dynamics (AIMD) has been the method of choice for modeling complex atomistic phenomena from first principles. Due to the high computational cost, most AIMD applications are limited to quantum systems with up to thousands of atoms. This talk will report that a machine-learning-based simulation protocol (Deep Potential Molecular Dynamics), while retaining ab initio accuracy, can simulate more than 1 nanosecond-long trajectory of over 100 million atoms per day using a highly optimized code (GPU DeePMD-kit) on the Summit supercomputer. The code can efficiently scale up to the entire Summit supercomputer, attaining 91 PFLOPS in double precision (45.5% of the peak) and 162/275 PFLOPS in mixed-single/half precision. This work opens the door to a range of molecular dynamics applications of unprecedented size and timescale with ab initio accuracy. Dr. Lin also will discuss some recent progress on using AI to improve the performance of exchange-correlation functionals in density functional theory.

Bio:

Lin Lin received his B.S. degree in Computational Mathematics from Peking University in 2007 and doctorate in Applied and Computational Mathematics from Princeton University in 2011 (advised by Professor Weinan E and Professor Roberto Car). His research focuses on the development of efficient and accurate numerical methods for electronic structure calculations with broad applications in quantum chemistry, quantum physics, and materials science. He is now an associate professor in the Department of Mathematics at the University of California, Berkeley, a faculty scientist with Berkeley Lab's Mathematics Group within the Computational Research Division, and a mathematician with Berkeley Lab's Center for Advanced Mathematics for Energy Research Applications (CAMERA). He has received the Sloan Research Fellowship (2015), National Science Foundation CAREER award (2017), Department of Energy Early Career award (2017), inaugural SIAM Computational Science and Engineering (CSE) Early Career award (2017), and Presidential Early Career Awards for Scientists and Engineers (PECASE) (2019).

SPEAKER: GIUSEPPE CARLEO, EPFL SWITZERLAND

The Quantum Many-body Problem as a Challenge for Machine Learning Methods

Abstract:
Many-body quantum theory is a fundamentally "data-abundant" branch of physics. In recent years, several new machine learning approaches have been developed in the field with the aim of taming the infamous quantum many-body problem. In this presentation, Dr. Carleo will discuss how a systematic, controlled machine learning of the many-body wave-function can be realized. This goal is achieved by a variational representation of quantum states based on artificial neural networks. He also will discuss applications in diverse domains, ranging from open problems in condensed matter physics to applications in quantum computing. Focusing on the latter case, he will show that there are relevant cases where machine learning techniques can be used to simulate useful quantum algorithms on purely classical hardware, effectively bypassing the need for specialized quantum resources.

Bio:
Giuseppe Carleo is a computational quantum physicist, best known for the introduction and development of novel machine learning techniques to study complex quantum phenomena. He earned a doctorate in 2011 at the International School for Advanced Studies (SISSA) in Italy. He has held several research and academic positions, including lecturer at ETH Zurich and staff research scientist at Flatiron Institute in New York City. Since 2020, he has been an assistant professor at the Institute of Physics of EPFL in Switzerland and serves as head of the Computational Quantum Science Lab.

BROOKHAVEN NATIONAL LABORATORY

U.S. DEPARTMENT OF ENERGY | Office of Science

Connect with Brookhaven Lab on Social Media

@BrookhavenLab