



# CASC Newsletter | Vol 11

## December 2021

### ***In This Issue:***

- [From the Director](#)
- [Collaborations: Optimization with Hybrid Learning for Jointly Solving Modern Power Grid Challenges](#)
- [Lab Impact: Ubiquitous Performance Analysis with Caliper and SPOT](#)
- [Advancing the Discipline: An Asymptotic-Preserving Algorithm for Magnetic Fusion](#)
- [Machine Learning & Applications: ML Utilized for Autonomous Multiscale Simulations for Cancer Biology](#)

### **From the Director**

Contact: [Jeff Hittinger](#)

“Every once in a while, a new technology, an old problem, and a big idea turn into an innovation.” – *Dean Kamen*

It's that time of the year, when people's thoughts turn to ghost and goblins, pumpkin spice, the approaching holidays, and, of course, SC. The press is abuzz with new hardware announcements from industry, and the latest Top500 and Green500 results are announced with much fanfare and, lately, intrigue. All of this focus on the technology is, of course, well-deserved, but we should never lose sight of the fact that innovative advances in software and algorithms further amplify these technological gains.

Reducing the computational complexity of a problem through clever algorithms can provide payoffs far beyond the speed-ups from hardware. In fact, that's really why the possibility quantum computing is so attractive: It's not that the hardware is faster per se; it's that the hardware could support algorithms that have polynomial complexity for problems that, classically, have super-polynomial cost (i.e., exact solutions are impractical to compute for all but the smallest problems). Furthermore, for as impressive as modern supercomputers are, when it comes to using them, they are rather opaque and finicky creatures lurking behind a seemingly simple command line prompt. It takes some serious effort to interrogate these devices to make sense of what, in fact, a supercomputer is doing with the thousands of lines of codes you asked it to execute.



In this spirit, we feature in this installment of the CASC Newsletter four vignettes of efforts to advance HPC software and algorithms in support of challenging problems of great interest to the nation. First, we have work led by Xiao Chen that has advanced the state of uncertainty quantification and risk analysis methods that will lead to a more robust, secure, and flexible power grid. Next, we feature the concept of and tools for ubiquitous performance analysis, as championed by David Boehme and Olga Pearce, as a better way for our simulation code teams to manage and maintain high performance across platforms—not as an afterthought, but throughout software development. In support of fusion simulation, Lee Ricketson writes about a new asymptotic-preserving, conservative algorithm that accurately models the complex boundary conditions in a magnetically confined plasma while gaining orders of magnitude in time-step size. Finally, the novel multiscale modeling approach to the RAS cancer initiation process, which leverages machine learning and represents a large breakthrough in the scale and speed of modeling the complex RAS protein/cell membrane interaction, is described by Harsh Bhatia. Taken together, these examples provide an incomplete but informative cross-section of the many efforts in CASC to increase the impact of HPC—through algorithms and software—on the mission of the Laboratory and on the nation.

## **Collaborations | Optimization with Hybrid Learning for Jointly Solving Modern Power Grid Challenges**

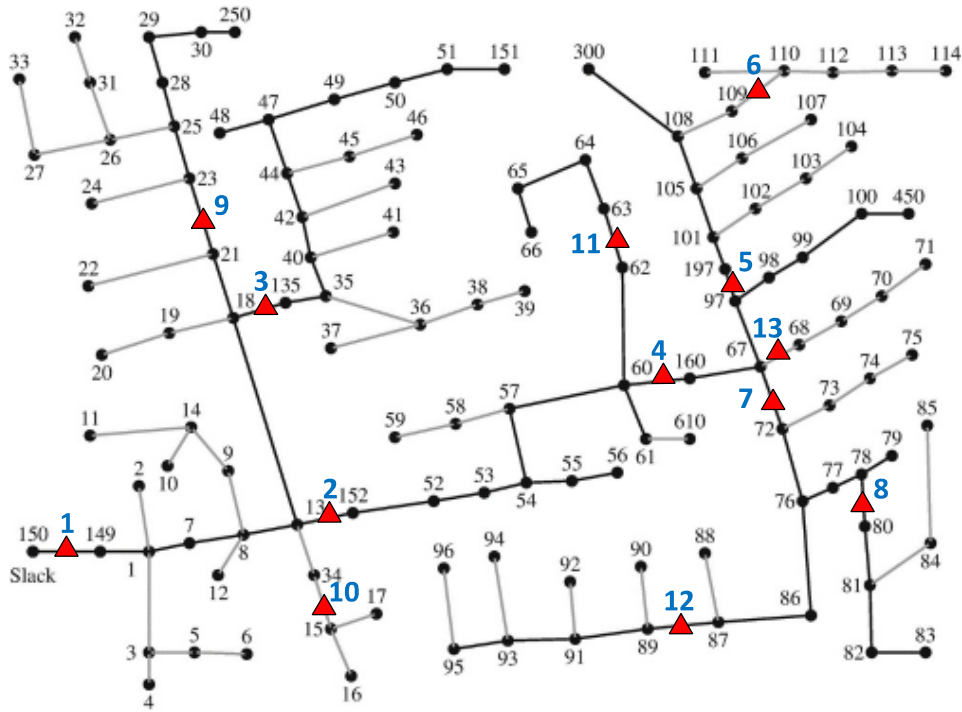
Contact: [Xiao Chen](#)

The electric power industry has undergone extensive changes over the past several decades. It has become increasingly important to ensure that the modernized grid is reliable, resilient, secure, affordable, sustainable, and flexible. As a result, CASC researchers in the areas of optimization, uncertainty quantification (UQ) [1, 6], risk analysis [2], decision support [3, 4], and machine learning (ML) are working to address the needs emerging within modern power grids. In one such effort, CASC researchers Xiao Chen (project lead) and Charles Tong collaborated with Engineering researchers Mert Korkali, Can Huang, Emma Stewart (now at National Rural Electric Cooperative Association), Liang Min (now at Stanford University), and Vaibhav Donde on an LLNL LDRD project to develop an HPC-enabled computational framework by combining optimization, UQ, and ML and demonstrated the framework on several power grid applications.

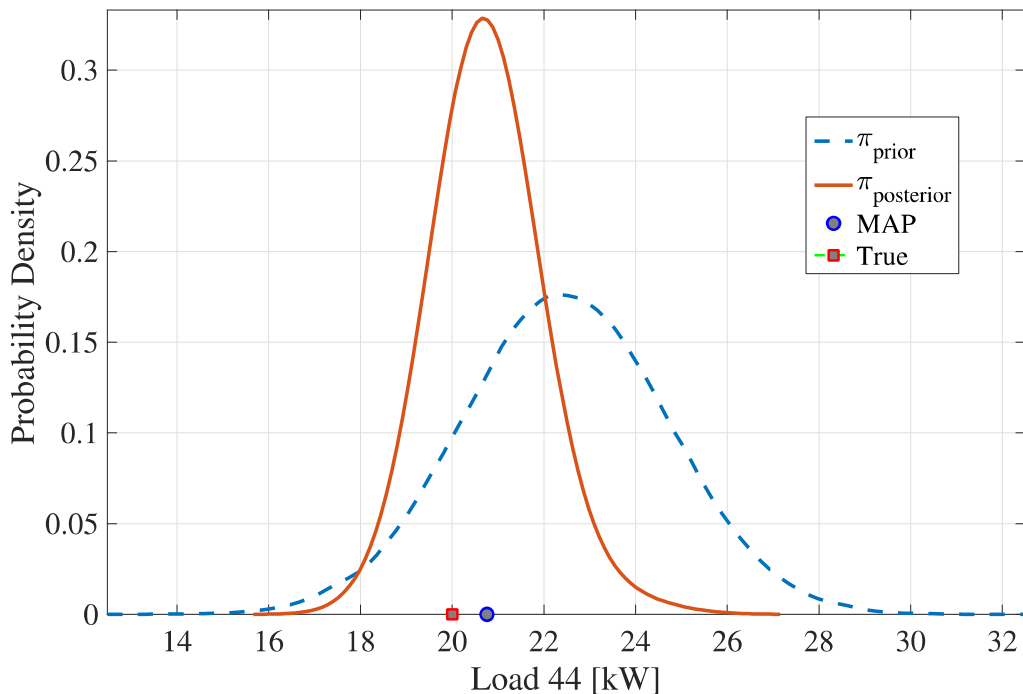
A modern grid must have improved reliability for everyday operations and superior flexibility to respond to uncertainties. To account for these uncertainties in inherently stochastic power grids, the research team developed a Bayesian approach [1] that can efficiently and robustly recover an unknown and a complex random field with associated uncertainty from high-dimensional models and multi-source datasets. As an application to the power grid, the research team developed a cost-effective, gradient-enhanced,



multi-surrogate method based on a multi-fidelity adaptive model for assessing risks [2] in probabilistic power-flow analysis under rare events, such as random outages of transmission lines or transformers.



**Figure 1:** (Top) Topology of the IEEE 123-bus system. (Bottom) Bayesian posterior for the active power of Load 44. The MAP estimate (blue circle) is at 20.6522 kW while the true value (red square) is at 20 kW.





The research team also collaborated with Virginia Tech and Missouri State University on a project supported by the DOE Advanced Grid Modeling program (AGM) under the DOE Office of Electricity (OE). The DOE OE/AGM program supports the nation's foundational capacity to analyze the electric power system using big data, advanced mathematical theory, and high performance computing to enable grid operators to optimize their decision-making in real-time, thus giving the industry sophisticated tools to substantially improve reliability, resiliency, and grid security.

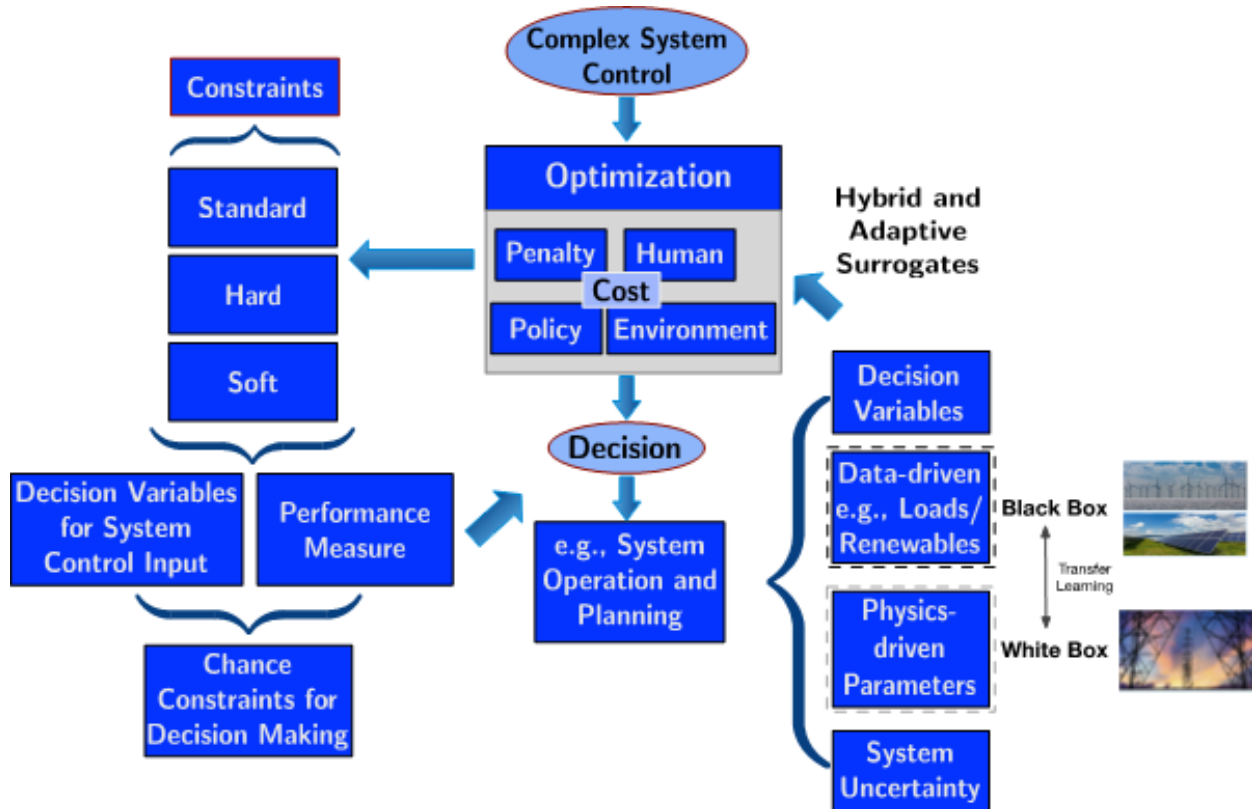
For example, this research team developed a Bayesian-inference framework that enables simultaneous estimation of the topology and the state of a three-phase, unbalanced power distribution system [3]. Specifically, by using a very limited number of measurements that are associated with the forecast load data, this method efficiently recovered the full Bayesian posterior distributions of the system topology under both normal and outage operation conditions. The inference is performed through an adaptive importance sampling procedure that greatly alleviates the computational burden of the traditional Monte Carlo sampling-based approach while maintaining good estimation accuracy. The simulations conducted on the IEEE 123-bus test system and an unbalanced 1,282-bus system revealed excellent performance of the new approach. Figure 1 shows the non-trivial topology of the IEEE 123-bus system and one of the framework results from estimating the Bayesian posterior.

Recently, this research team has developed several approaches to surrogate-based risk assessment of rare events, including response-surface-based, chance-constrained alternating current optimal power flow [4]; and, most recently, "hybrid learning" and optimization for transmission (white-box) and distribution (black-box) systems [5]. For the ongoing hybrid learning project, the research team is currently developing computational frameworks in which to optimize hybrid problems, which involve both incomplete/unknown information and mathematically known functions, in such a way that maximizes the interchange between the known (white-box) and unknown (black-box) components (Figure 2). Together, these new methods can drastically improve the decision making process under power system uncertainty in order to enhance grid security.

[1] Z. Hu, Y. Xu, M. Korkali, X. Chen, L. Mili, J. Valinejad, and C. Tong. "A Bayesian Approach for Estimating Uncertainty in Stochastic Economic Dispatch Considering Wind Power Penetration." *IEEE Transactions on Sustainable Energy*. 12(1):671–681, 2021.

[2] Y. Xu, M. Korkali, L. Mili, X. Chen, and L. Min. "Risk Assessment of Rare Events in Probabilistic Power Flow via Hybrid Multi-Surrogate Method." *IEEE Transactions on Smart Grid*. 11(2):1593–1603, 2020.

[3] Y. Xu, M. Korkali, L. Mili, X. Chen, J. Valinejad, and Z. Zheng. "An Adaptive-Importance-Sampling-Enhanced Bayesian Approach for Topology Estimation in an Unbalanced Power Distribution System." *IEEE Transactions on Power Systems*, 2021.



**Figure 2:** Hybrid and adaptive surrogates, in conjunction with a high-fidelity system model, for multi-objective optimization will seek desirable tradeoffs between multiple objectives while assessing and managing the risk of failures in power grids.

[4] Y. Xu, M. Korkali, L. Mili, J. Valinejad, T. Chen, and X. Chen. "An Iterative Response-Surface-Based Approach for Chance-Constrained AC Optimal Power Flow Considering Dependent Uncertainty." *IEEE Transactions on Smart Grid*. 12(3):2696–2707, 2021.

[5] Y. Xu, L. Mili, M. Korkali, X. Chen, J. Valinejad, and L. Peng. "A Surrogate-Enhanced Scheme in Decision Making under Uncertainty in Power Systems." In *Proceedings of the 2021 IEEE Power and Energy Society General Meeting (PESGM)*, 2021.

[6] C. Huang, C. Thimmisetty, X. Chen, E. Stewart, P. Top, M. Korkali, V. Donde, and L. Ming. "Power Distribution System Synchronphasor Measurements With Non-Gaussian Noises: Real-World Data Testing and Analysis." *IEEE Open Access Journal of Power and Energy*. 8:228-238, 2021.



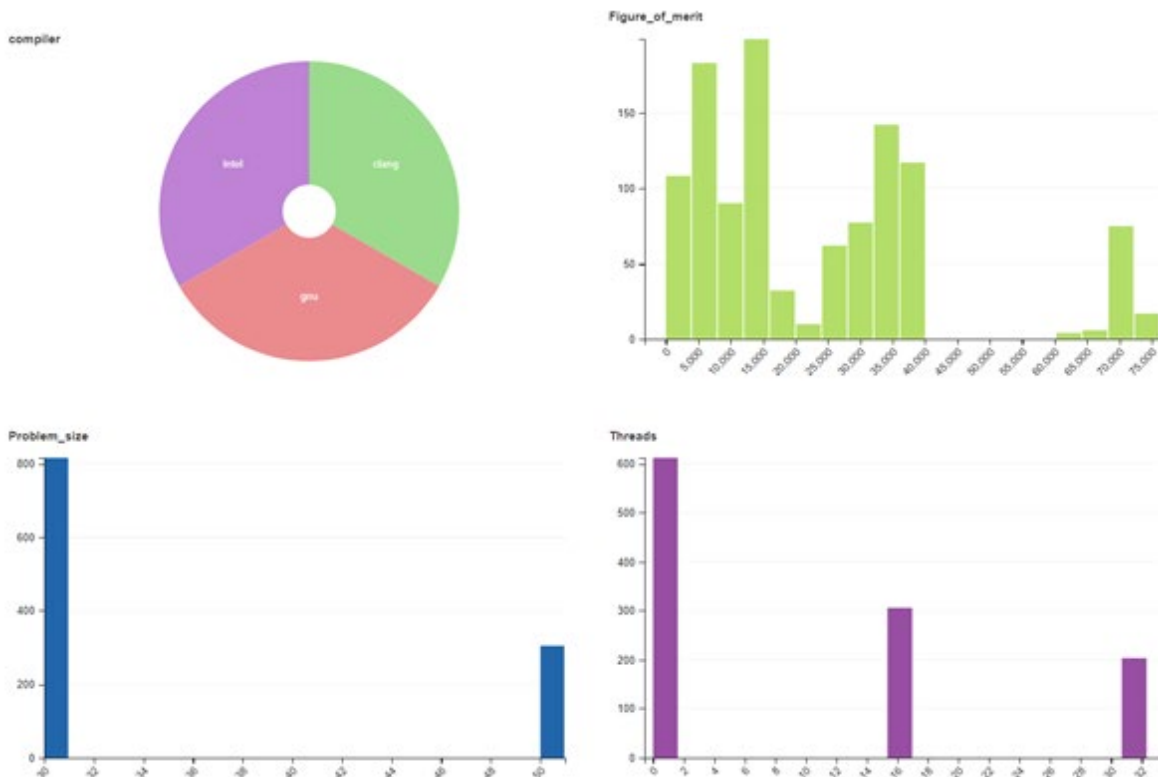
## Lab Impact | Ubiquitous Performance Analysis with Caliper and SPOT

Contact: [David Boehme](#) and [Olga Pearce](#)

Achieving optimal performance on today's and tomorrow's leading supercomputers remains a challenging task for LLNL application developers. To support them in this challenge, CASC researchers David Boehme and Olga Pearce along with Computing researchers Pascal Aschwanden, Kenneth Weiss, and Matthew LeGendre have been developing a new set of *ubiquitous performance analysis* workflows and tools that simplify and automate many common performance profiling tasks. This new approach represents a shift in how users view performance tracking in long-lived HPC codes: where performance profiling was once a sporadic activity that often required the help of experts, it is now a regular part of nightly unit testing. With the gathered data, developers can quickly identify *performance regressions* (loss of performance), track optimization goals, and compare performance between different hardware platforms or program configurations.

Ubiquitous performance analysis is built around [Caliper](#), an instrumentation and profiling library, and [SPOT](#), a web-based visualization and analysis frontend, both developed at LLNL. Caliper integrates performance measurement capabilities directly into the target codes, leveraging user-defined instrumentation points placed in the source code as a robust and efficient way to record performance data in complex codes. Once in place, users can enable Caliper performance measurements with a simple command-line switch in the application.

A new addition to the tool stack is [Adiak](#), a small library to record program metadata. This capability is extremely valuable because most use cases for ubiquitous performance analysis involve comparisons between different program executions; for example, what is the performance on different systems, or how does it change with different program configurations? To facilitate such performance investigations, a descriptive metadata about each run is needed, which includes information such as the machine on which the program was running, the launch date and time, program version, job configuration (e.g., MPI job size), as well as application-specific input and configuration parameters (e.g., problem size or enabled features). Adiak has built-in functions to record many common metadata values. In addition, developers can add application-specific data in *key:value* form. The Adiak metadata is automatically added to the Caliper performance profiles.

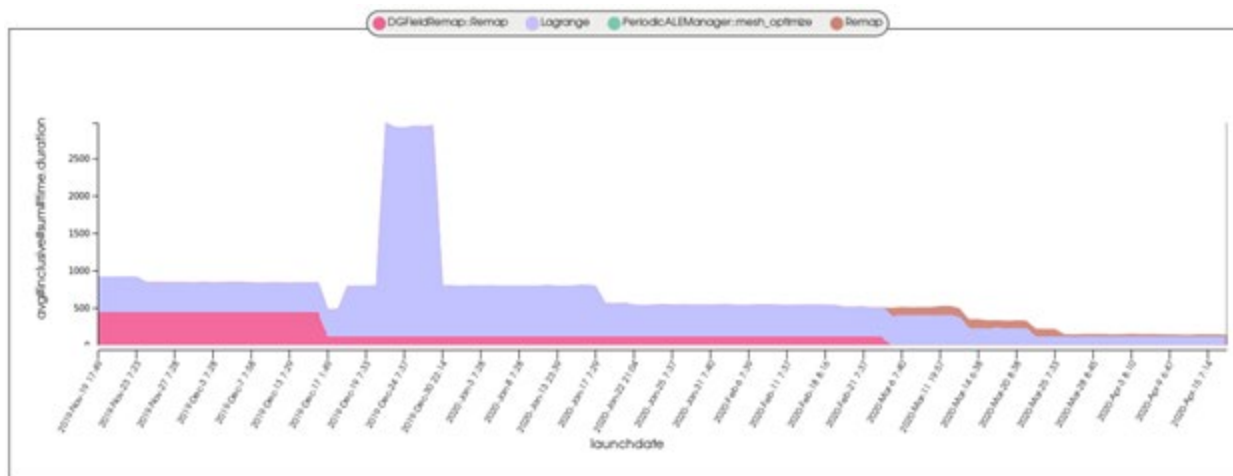


**Figure 3:** The SPOT landing page featuring four histogram charts that show the number of runs with certain metadata values.

Ubiquitous performance analysis often collects large amounts of data: The Marbl team, for example, collects around 80 performance datasets with each nightly test run. Helping with the analysis of this data is SPOT, a web front-end with filter and visualization tools specifically designed for the analysis of large collections of performance data. The SPOT *landing page* (Figure 3) is populated with interactive charts that show summary histograms for selected metadata attributes—for example, the runs performed by a particular user or runs that invoked a particular physics package. Users can select subsets of data to discover correlations between metadata variables—for example, to identify the program configurations that deliver the highest figure-of-merit.

Another powerful SPOT tool, the *comparison page*, helps with performance regression testing; the plot in Figure 4 shows the evolution of performance (time) in various code regions for one of Marbl’s nightly test cases over several months during its GPU port. Ubiquitous performance analysis made it easy for the Marbl team to detect a performance regression (in late December 2019) and can seamlessly handle changing annotation labels, such as when the “DGFieldRemap::Remap” annotation (pink) was renamed to “Remap” (brown) around March 2020. SPOT is available to all LC account holders via [LC’s web portal](#).





**Figure 4:** SPOT performance tracking for Marbl’s triple-point-3D problem on a 4 GPU compute node over the course of several months during its ongoing GPU port.

Recently, ubiquitous performance analysis has been integrated into several large LLNL codes. In Marbl, for example, ubiquitous performance analysis has been instrumental in helping the development team track and understand the code’s performance as they port the codebase to new architectures. Automatic performance capture also ensures that changes do not result in performance regressions on other platforms during the porting process. Similarly, ubiquitous performance analysis has simplified the set up node-to-node performance scaling studies in Marbl to compare the code’s performance across several HPC architectures including Intel- and ARM-based CPU clusters as well as a GPU-based cluster. A recent paper at ISC High Performance 2021 [1] provides a detailed overview of the ubiquitous performance analysis approach, and tutorial presentations at the 2021 ECP Annual Meeting and SC21 introduced our concept of ubiquitous performance analysis to the HPC community at large. Critically, to serve the mission needs of the laboratory, CASC and LC researchers continue to integrate Caliper and SPOT into additional LLNL application codes.

[1] D. Boehme, P. Aschwanden, O. Pearce, K. Weiss, and M. LeGendre. “Ubiquitous Performance Analysis.” In *Proceedings of ISC High Performance*, 2021.

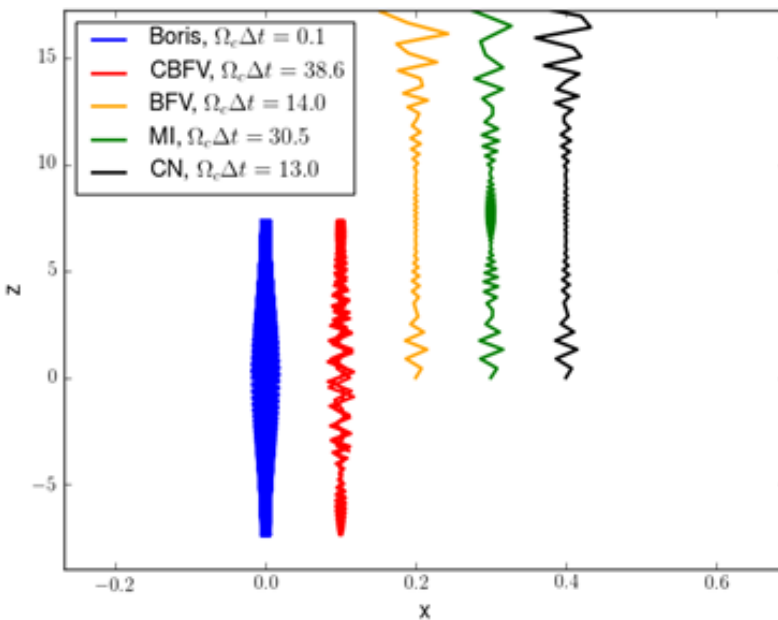




# Advancing the Discipline | An Asymptotic-Preserving Algorithm for Magnetic Fusion

Contact: [Lee Ricketson](#)

Magnetic confinement fusion reactors have long held the promise of abundant, renewable, clean energy. As the name suggests, fusion is achieved by using strong magnetic fields to confine charged particles while they are heated to the point that energy-releasing fusion reactions occur. Prominent among the many challenges in simulating such devices is the fact that such strong magnetic fields induce gyration in the particles with a frequency that is often much higher than the dynamics of interest.



**Figure 5:** Particle trajectories in the  $x$ - $z$  plane in a magnetic mirror. The new algorithm (red) accurately reproduces the trapped orbit of a finely-resolved scheme (blue) with a time-step 386 times as large. Prior efforts incorrectly predict that the particle is unconfined.

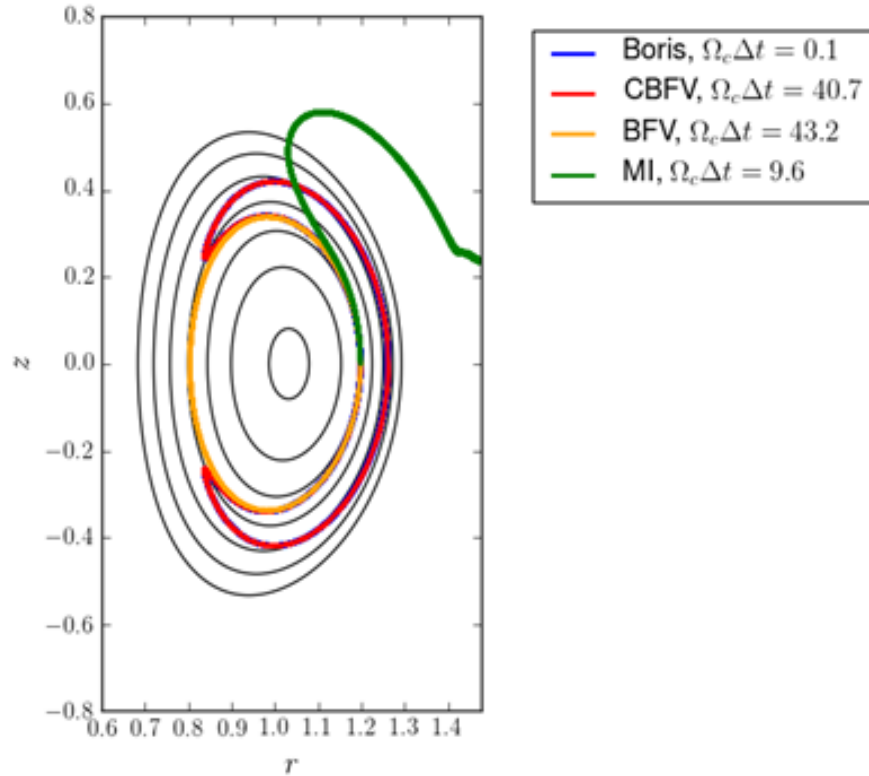
CASC researcher Lee Ricketson, in collaboration with colleagues Luis Chacón, Guangye Chen, and Oleksandr Koshkarov

at LANL, has developed a new method for accurately computing the motion of such charged particles while largely ignoring the fast gyration timescale. The existing state-of-the-art uses an asymptotic expansion to analytically eliminate the fast scale and has historically been successful, but it presents difficulties in circumstances when the gyration must be resolved—for instance, when particles interact with the walls of fusion devices. While such interactions are rare, accurately evaluating their effects is critical to determining device performance and longevity. In contrast, the new method proposed by Ricketson and colleagues recovers the full particle dynamics when gyration is resolved while still correctly recovering the relevant asymptotic limit when it is not.



This result is achieved by introducing a fictitious force to the differential equation that governs particle motion, which vanishes in the small time-step limit and helps recover the correct asymptotics for large time-steps. While previous efforts have pursued a similar approach, Ricketson et al.'s new method is the first to conserve particle energy, a critical property when determining particle confinement, as illustrated in Figures 5 and 6. In particular, Figure 6 demonstrates the scheme's ability to handle the complex magnetic geometry of the International Thermonuclear Experimental Reactor (ITER) being developed in France.

**Figure 6:** Particle “banana orbits” in the  $r$ - $z$  plane in ITER-like tokamak geometry, with magnetic flux surfaces shown in black. The new algorithm (red) again accurately reproduces the finely-resolved scheme (blue, barely visible under red) with a time-step 407 times as large. Prior efforts make large, qualitative errors in the particle trajectory.



The initial method was published in the *Journal of Computational Physics* [1] and was the subject of an invited talk at the Sherwood Fusion Theory Conference, a leading conference for the magnetic fusion energy community. A current LLNL LDRD, led by Justin Angus in Engineering, has proposed to use the method to study Z-Pinch plasmas, and implementation of the method in plasma simulation codes at LANL is underway. While increases in computer hardware capabilities have allowed fusion scientists to study ever more detailed processes in plasmas, algorithmic advances of this sort are critical to the modeling effort, as they provide significant reductions in computational complexity and cost rivaling and complementing the advances in HPC systems.

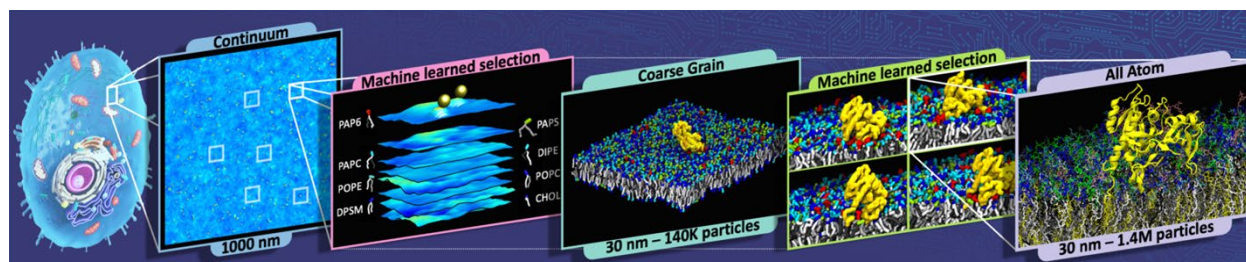
[1] L. F. Ricketson and L. Chacón. “An Energy-Conserving and Asymptotic-Preserving Charged-Particle Orbit Implicit Time Integrator for Arbitrary Electromagnetic Fields.” *Journal of Computational Physics*, 2020.



# Machine Learning & Applications | ML Utilized for Autonomous Multiscale Simulations for Cancer Biology

Contact: [Harsh Bhatia](#)

Multiscale modeling and simulations are well-accepted ways to bridge the length- and time-scales required for scientific studies. However, capturing coupled processes within multiscale simulations for systems of increasing size and complexity quickly exceeds the limits of contemporary computational capabilities. A good example of this is the DOE/National Cancer Institute Pilot2 project, which is studying the cancer initiation pathway mediated by the RAS and RAF oncogenes. RAS and RAF are important links in the cell's signaling system and, if malfunctioning, lead to uncontrolled cell growth. Nearly a third of all cancers diagnosed in the U.S. are driven by mutations in RAS genes and their protein products. Central to the Pilot2 study is the need to simulate these signaling interactions at multiple scales of resolutions (Figure 7) to foster new understanding of cancer initiation.



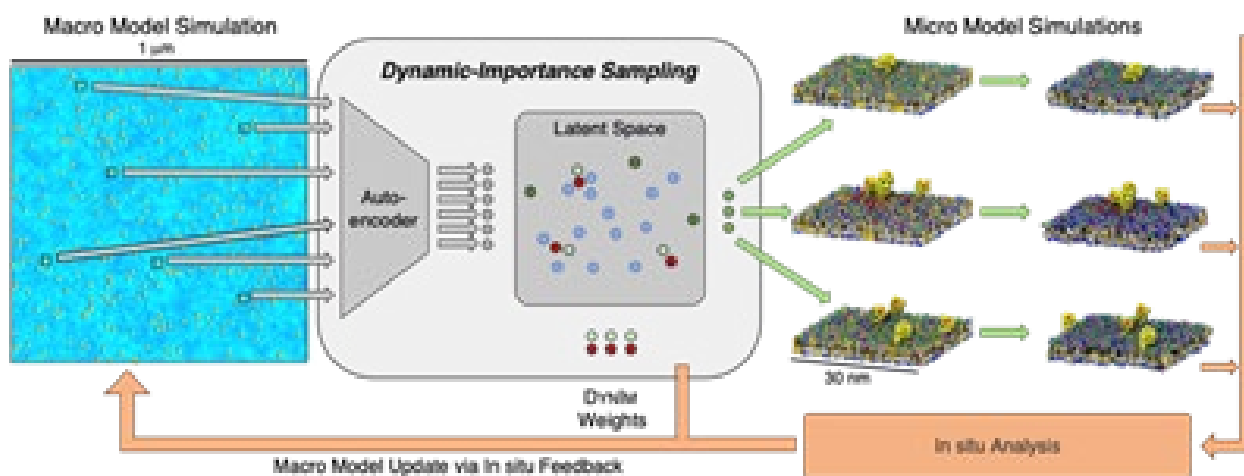
**Figure 7:** The three-scale MuMMI simulation campaign illustrates RAS-RAF protein interactions and membrane dynamics at continuum, coarse-grained, and all-atom scales of resolution. These scales are coupled through ML-based selection in a pairwise manner to promote important configurations from coarse to fine scales.

To address this challenge, CASC researchers Harsh Bhatia (technical lead), Shusen Liu, Brian Van Essen, and Timo Bremer, in collaboration with a team of researchers from Physical and Life Sciences, have developed a machine learning (ML) framework to facilitate a new style of autonomous multiscale simulation by capturing the inherent variability in macro configurations of interest. The proposed framework, called *dynamic-importance (DynIm) sampling* [1], makes two significant contributions to the field of multiscale simulations. First, DynIm is designed to maximize multiscale interrogation by connecting each macro configuration to a micro simulation that is sufficiently similar to serve as its statistical proxy. As a result, the multiscale simulation enables the exploration of macro length- and time-scales, but with the effective precision of the microscale model. Second, the dynamic nature of the sampling automates a feedback process in which microscale data is used to improve the parameterization of the macro



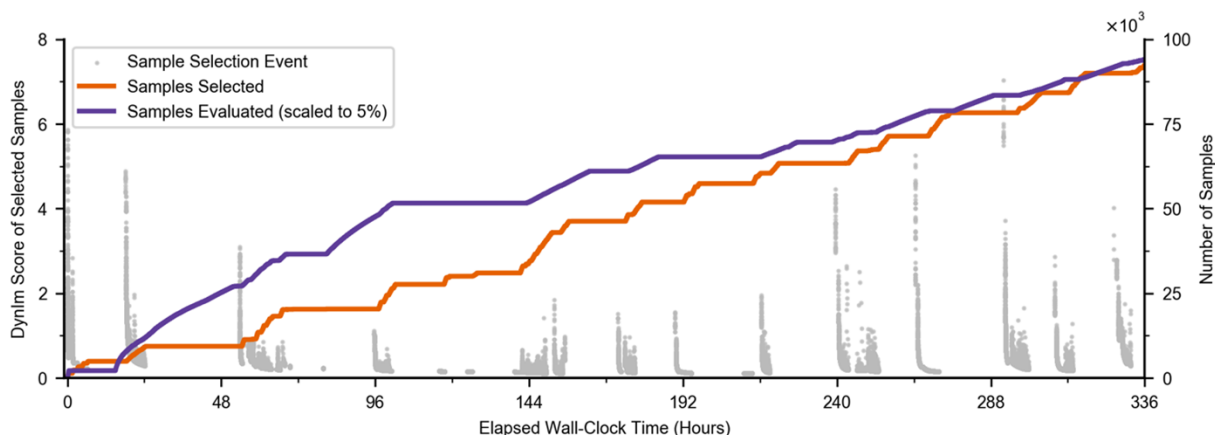
model while the simulation is running. As a result, the framework represents a self-healing paradigm in multiscale simulations.

As illustrated in Figure 8, DynIm uses a variational autoencoder (VAE) that can learn to map *patches*, which represent local neighborhoods of RAS proteins on the cell membrane, into a 15-dimensional latent space that understands how similar or dissimilar two such neighborhoods are. By encoding a patch into this low-dimensional vector and then decoding it back into the patch, a VAE learns how to preserve the most important characteristics of the given patch. DynIm then uses a farthest-point sampling approach that can identify a diverse subset of patches, which are most dissimilar to the ones already sampled. Compared to existing sampling methods, this framework does not require a complete input distribution for the sampling to be prespecified, but rather adapts its sampling choices dynamically based on an evolving input distribution and supports *ad hoc* requests for samples (Figure 9).



**Figure 8:** Using a variational autoencoder (VAE) to project the data onto a latent space, DynIm provides a mathematically relevant importance metric and a computationally feasible and scalable solution to massive multiscale simulations.

The DynIm framework [1] has been instrumental in enabling the Pilot2 studies [2] by providing a way to explore *continuum scale* (based on dynamic density functional theory) and to identify therein the “most important” configurations that are then promoted to *coarse-grained* molecular dynamics scale. DynIm was deployed as part of the Machine Learned Multiscale Modeling Infrastructure (MuMMI). In 2018, the research team conducted the first-ever two-scale simulation [3] of RAS-membrane interactions on Sierra, which led to new insights into understanding lipid fingerprints promoted by RAS aggregation. Since then, they have expanded DynIm through more-sophisticated ML model and have deployed it on Summit to simulate RAS-RAF-membrane interactions at three scales [4]. In general, this effort is geared towards improving the capabilities of large multiscale simulation campaigns to leverage modern, heterogeneous computing architectures, especially in anticipation of the upcoming exascale machines.



**Figure 9:** Event history of the DynIm sampling for a period of 14 days of wall-clock time. Each grey dot is a selection event with the score of the selected candidate mapped to the left axis, and the line plots show the cumulative number of patches evaluated and selected (mapped to the right axis).

- [1] Bhatia *et al.* “Machine Learning Based Dynamic-Importance Sampling for Adaptive Multiscale Simulations.” *Nature Machine Intelligence*, 3, 401–409, 2021.
- [2] Ingólfsson *et al.* “Machine Learning-driven Multiscale Modeling Reveals Lipid-Dependent Dynamics of RAS Signaling Proteins.” In *Proceedings of National Academy of Sciences (PNAS)*, 2021. In press.
- [3] Di Natale *et al.* “A Massively Parallel Infrastructure for Adaptive Multiscale Simulations: Modeling RAS Initiation Pathway for Cancer.” In *Proceedings of International Conference for High Performance Computing, Networking, Storage and Analysis (SC19)*, 2019. **Best Paper Award.**
- [4] Bhatia *et al.* “Generalizable Coordination of Large Multiscale Workflows: Challenges and Learnings at Scale.” In *Proceedings of International Conference for High Performance Computing, Networking, Storage and Analysis (SC21)*, 2021.

## CASC Newsletter Sign-Up

Was this newsletter link passed along to you? Or did you happen to find it on social media? [Sign up](#) to be notified of future newsletters.

*This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. LLNL-MI-830202. Edited by [Ming Jiang](#).*