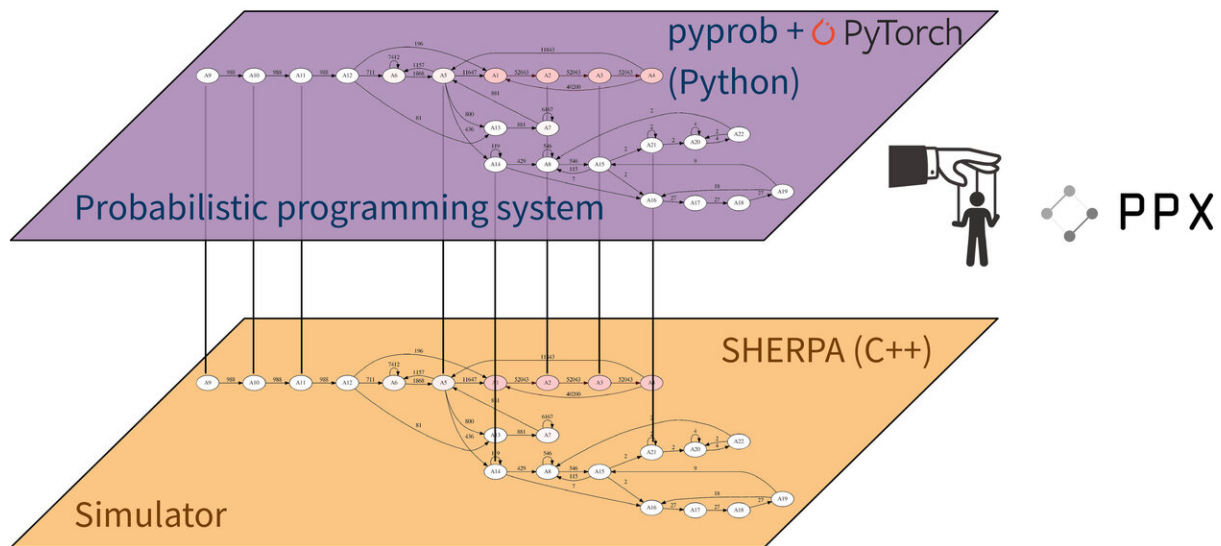# Etalumis 'reverses' simulations to reveal new science

November 13 2019, by Keri Troutman

Etalumis performs Bayesian inference--a method of statistical inference in which Bayes' theorem is used to update the probability for a hypothesis as more evidence or information becomes available--essentially inverting the simulator to predict input parameters from observations. This image provides an overview of the software framework. Credit: Wahid Bhimji, Lawrence Berkeley National Laboratory

Scientists have built simulations to help explain behavior in the real world, including modeling for disease transmission and prevention, autonomous vehicles, climate science, and in the search for the fundamental secrets of the universe. But how to interpret vast volumes

of experimental data in terms of these detailed simulations remains a key challenge. Probabilistic programming offers a solution—essentially reverse-engineering the simulation—but this technique has long been limited due to the need to rewrite the simulation in custom computer languages, plus the intense computing power required.

To address this challenge, a multinational collaboration of researchers using computing resources at Lawrence Berkeley National Laboratory's National Energy Research Scientific Computing Center (NERSC) has developed the first probabilistic programming framework capable of controlling existing simulators and running at large-scale on HPC platforms. The system, called Etalumis ("simulate" spelled backwards), was developed by a group of scientists from the University of Oxford, University of British Columbia (UBC), Intel, New York University, CERN, and NERSC as part of a Big Data Center project.

Etalumis performs Bayesian inference—a method of statistical inference in which Bayes' theorem is used to update the probability for a hypothesis as more evidence or information becomes available—essentially inverting the simulator to predict input parameters from observations. The team deployed Etalumis for the first time for the Large Hadron Collider (LHC) at CERN, bringing a new level of interpretability to data analysis from the LHC's high-energy physics detectors. A paper based on Etalumis has been selected as a finalist for Best Paper at SC19. The authors will speak about Etalumis at SC19 on Tuesday, November 19 at 4:30 p.m.

From Days to Minutes

Bayesian inference is used in virtually all scientific disciplines, according to Frank Wood, an Etalumis collaborator, Associate Professor of Computer Science at UBC, and one of the pioneers of probabilistic programming.

"I was particularly interested in applying Bayesian inference to an extremely complex physics problem, and high-energy physics detectors felt like the perfect proving ground for our group's seminal research," he says. "The Etalumis project provided a unique opportunity to combine a cutting-edge neural network based on an 'inference compilation' approach with a software framework (pyprob) to directly couple this inference engine to existing detailed particle physics simulators and run it on HPC-scale resources."
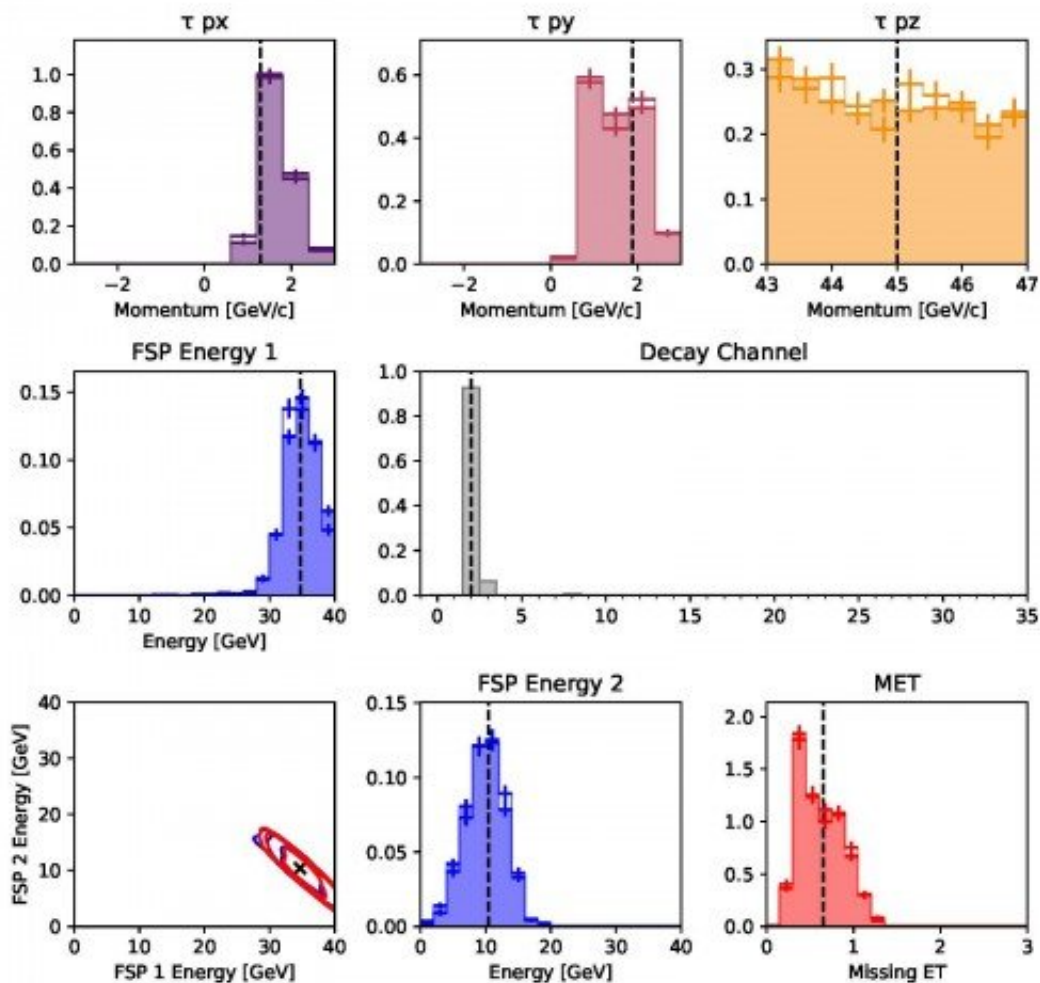
A comparison of some of the predictions from the Etalumis project's inference compilation approach (outline histograms), which can achieve the same levels of

precision as computationally intractable methods (filled histograms). Credit: Lawrence Berkeley National Laboratory

Scientists already have robust simulation software packages that model the physics and everything that occurs within the detector. Etalumis brings in probabilistic programming to couple with this existing software, essentially giving researchers the ability to say "We had this observation; how did we get there?"

"This project is exciting because it makes existing simulators across many fields of science and engineering subject to probabilistic machine learning," says Atilim Gunes Baydin, lead developer of the Etalumis project and lead author of the SC19 paper. Gunes is currently a postdoctoral researcher in machine learning at the University of Oxford. "This means the simulator is no longer used as a black box to generate synthetic training data, but as an interpretable probabilistic generative model that the simulator's code already specifies, in which we can perform inference.

"We need to be able to control the program to run down every possibility, so in this project we added this capability as a software layer," adds Wahid Bhimji, a Big Data Architect in the Data and Analytics Services team at NERSC. However, performing inference in such complex settings brings computational challenges. "Conventional methods for this kind of Bayesian inference are extremely computationally expensive," Bhimji adds. "Etalumis allows us to do in minutes what would normally take days, using NERSC HPC resources."

Deep Interpretability

For the LHC use case, the team trained a neural network to perform

inference, learning to come up with good proposals about what detailed chain of physics processes from the simulator might have occurred. This required improvements to the [PyTorch](#) deep-learning framework to train a complex dynamic [neural network](#) on more than 1,000 nodes (32,000 CPU cores) of the Cori supercomputer at NERSC. As a result, training that would take months with the original unoptimized software on a single node can now be completed in less than 10 minutes on Cori. Scientists thus gained an opportunity to study the choices that went into producing each outcome, giving them a greater understanding of the data.

"In many cases you know there's an uncertainty in determining the physics that occurred at an LHC collision but you don't know the probabilities of all the processes that could have given rise to a particular observation; with Etalumis, you get a model of that," Bhimji explains.

The deep interpretability that Etalumis brings to data analysis from the LHC could support major advances in the physics world. "Signs of new physics may well be hiding in the LHC data; revealing those signals may require a paradigm change from the classical algorithmic processing of the data to a more nuanced probabilistic approach," says Kyle Cranmer, an NYU physicist who was part of the Etalumis project. "This approach takes us to the limit of what is knowable quantum mechanically."

  More information: Etalumis: Bringing Probabilistic Programming to Scientific Simulators at Scale arXiv:1907.03382 [cs.LG] arxiv.org/abs/1907.03382

Provided by Lawrence Berkeley National Laboratory

Citation: Etalumis 'reverses' simulations to reveal new science (2019, November 13) retrieved 19

April 2024 from
https://phys.org/news/2019-11-etalumis-reverses-simulations-reveal-science.html