HYPERION RESEARCH

HPC User Forum Update

AI for Science

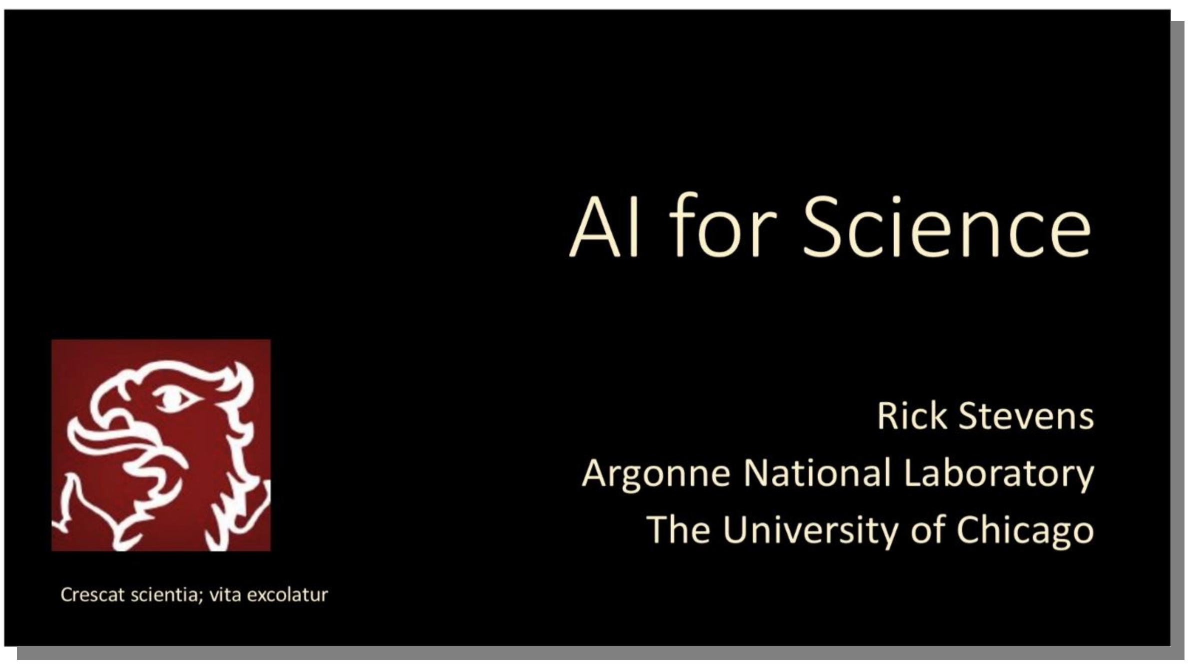
Bob Sorensen
November 2019

IN THIS UPDATE

The HPC User Forum was established in 1999 to promote the health of the global HPC industry and address issues of common concern to users. In September 2019, the 73rd HPC User Forum took place at Argonne National Laboratory in Illinois. This update summarizes a presentation from that meeting in the session entitled *AI for Science* by Rick Stevens from Argonne National Laboratory.

Stevens' talk described Argonne’s new AI for Science initiative, which is aimed at accelerating the development and adoption of artificial intelligence approaches in scientific and engineering domains. The goal is to drive research and development breakthroughs in energy, basic science, medicine, and national security, especially where there is significant volumes of data and relatively less developed theory.

"“AI is transforming our regular life. It hasn't yet transformed our scientific and engineering lives, but it will. AI methods allow us to discover patterns in data that can lead to experimental hypotheses and thus link data driven methods to new experiments and new understanding", according to Stevens.

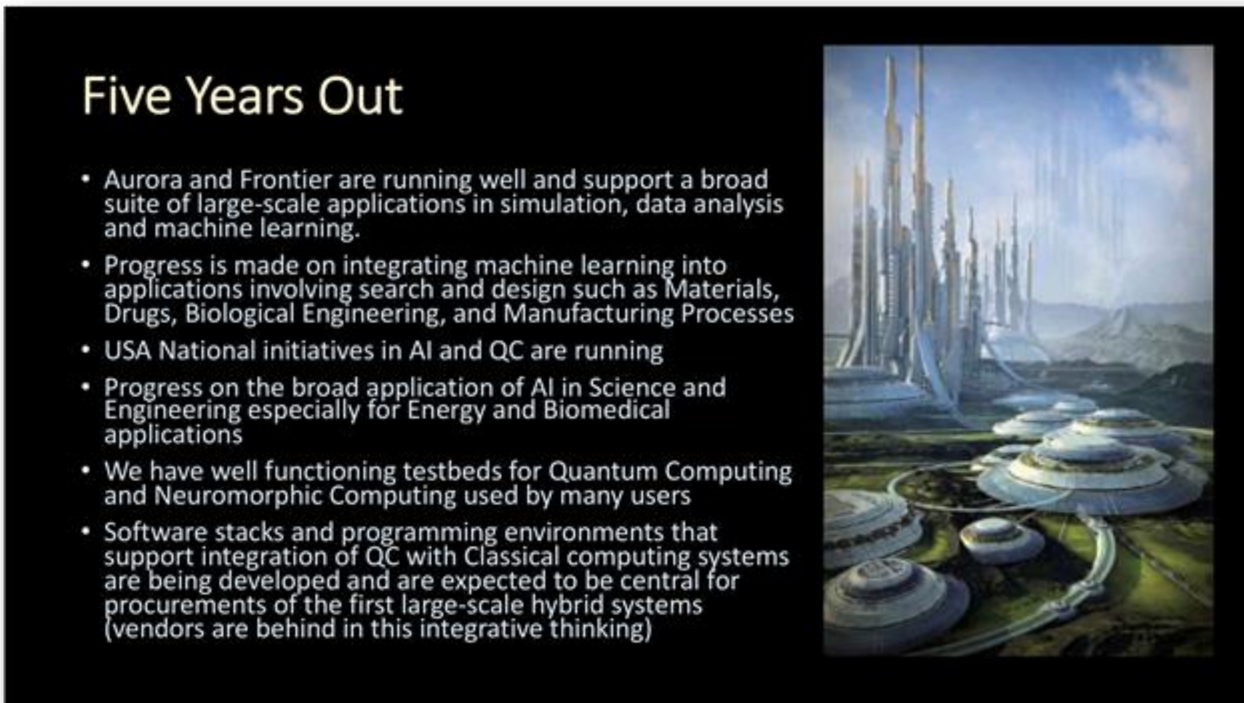


Source: Argonne National Laboratory and Hyperion Research, 2019

PRESENTATION: AI FOR SCIENCE, RICK STEVENS, ARGONNE NATIONAL LAB

Rick Stevens from Argonne National Laboratory described the lab's new AI for Science initiative. He began his talk defining what his group means by artificial intelligence and then laid out a set of predictions for how scientists will use AI five, ten, and fifteen years out from now. Together these predictions lay the foundation of an AI plan for the Department of Energy and potentially a roadmap for post-exascale computing.

FIGURE 1



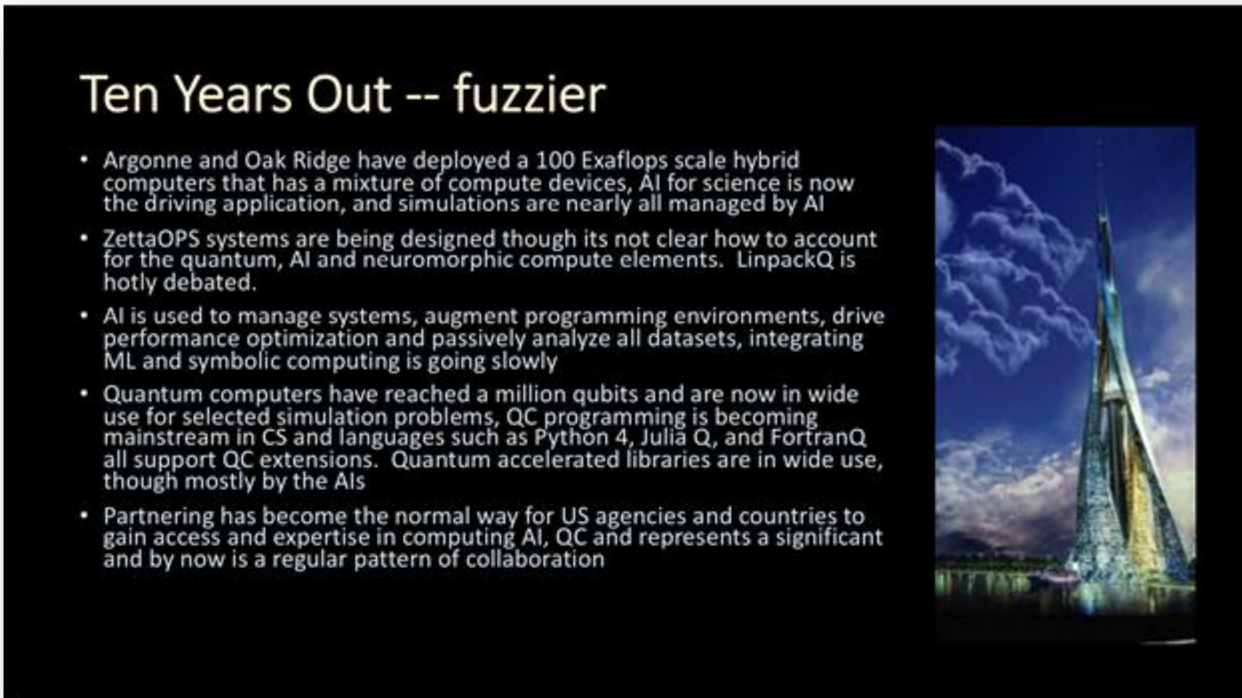
Source: Argonne National Laboratory and Hyperion Research, 2019

AI Five Years Out

- Stevens said that five years out, AI capabilities will be well on its way to integrating machine learning into applications involving search and design such as materials, drugs, biological engineering, and manufacturing processes. Fueled by US national initiatives in AI and quantum computing, researchers will also be making progress on the broad application of AI in science and engineering especially for energy and biomedical applications. In addition, Stevens predicted that there will be well-functioning testbeds for quantum and neuromorphic computing used by many users.

- Within five years, Stevens said there will be software stacks and programming environments that support integration of QC with classical computing and that this foundation will be central to procurements of the first large-scale hybrid (AI and QC) systems.

FIGURE 2



Source: Argonne National Laboratory and Hyperion Research, 2019

AI Ten Years Out

"In the Ten Year Outlook, AI for Science will be the driving application, and simulations are nearly all managed by AI. Things like biology will be done by people sitting in coffee shops, writing Python code that we'll execute in the robots in the bio cloud. Running simulations will be done the same way you write medical code to manage the simulations," according to Stevens.

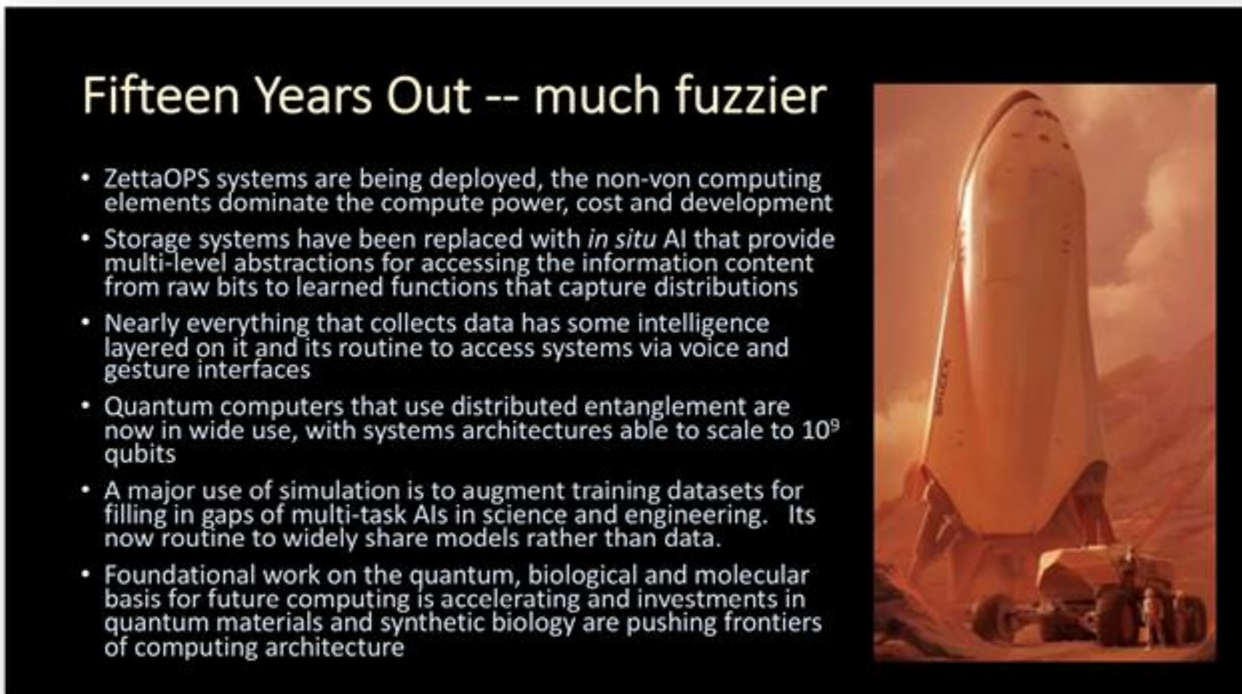
To fully power this capability, Argonne and Oak Ridge National Laboratories will have deployed 100 exaflops scale hybrid computers with a mixture of compute devices. To wrangle with the immense system scale and complexity, AI will be used to manage systems, augment programming environments, drive performance optimization, and passively analyze all datasets.

"Storage devices won't be static, just storing the bits. If you're going to write the bits out, you might as well do something with them in parallel with them sitting there. Right? So there will be passive analysis of data where the storage system tries to figure out the patterns in the data spontaneously. In this scenario, the system says, 'I found something', and you can decide if you care," according to Stevens.

In a similar fashion to the buildup to exascale, zettaOPS systems will be on the drawing board when 100 exaflops systems are deployed. Stevens noted that it is not clear how to account for the quantum, AI, and neuromorphic compute elements as they will each have their own unique developmental challenges.

Stevens forecasted that by 2029, quantum computers will have reached a million qubits and will be in wide use for selected simulation problems. Stevens added that QC programming will be moving into the mainstream in computer science and languages such as Python 4, Julia Q, and FortranQ will support QC extensions. Quantum accelerated libraries are in wide use, though mostly by the AI programs.

FIGURE 3



Source: Argonne National Laboratory and Hyperion Research, 2019

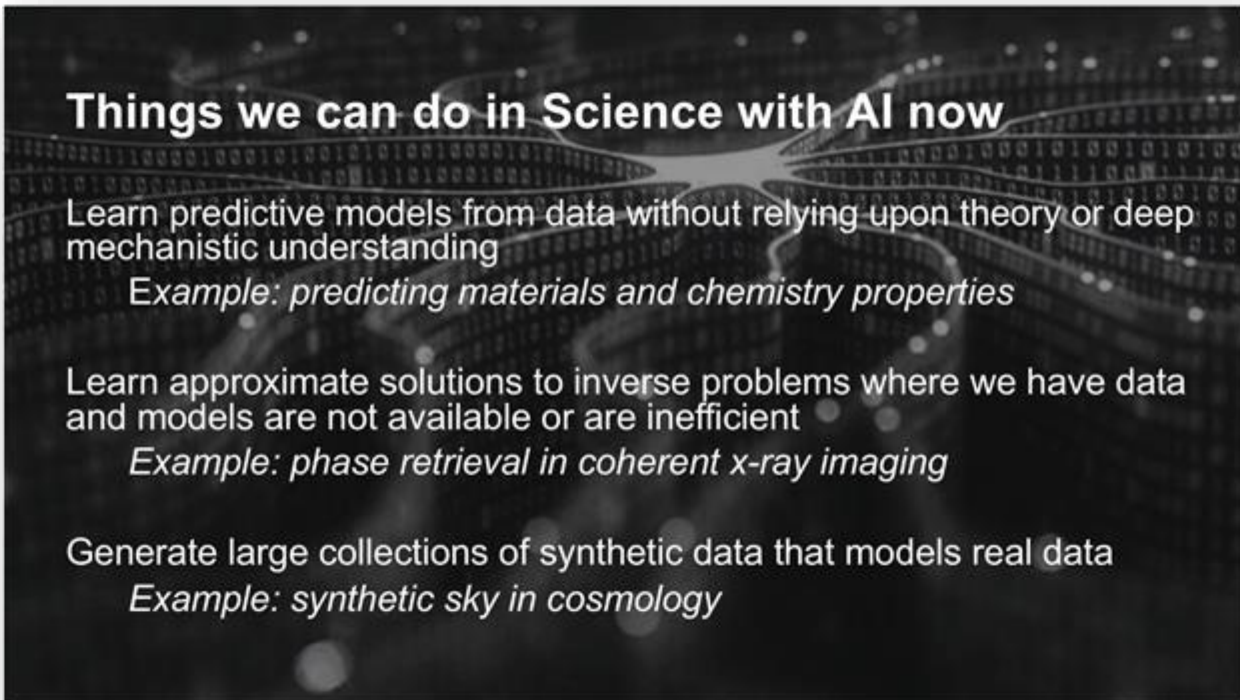
AI Fifteen Years Out

"In the fifteen year timeframe, zettaOPS systems are being deployed where the non-von computing elements dominate the compute power, cost and development. This will make it difficult to estimate compute cost as development on these things will be a challenge," according to Stevens.

In fifteen years, storage systems will be replaced by in-situ AI. Programmers will not write data out to a storage system, but instead hand it off to an AI that will learn a model and throw the data away. Then, when data is to be retrieved, the AI will tell you what the data was. The goal will be to have a system that automatically comes up with multilevel abstractions of your data. That's something that researchers don't exactly know how to do, but by then they should be able to figure out some solution.

"Instead of reproducing the raw bits of storage like we do today, the system will give you back some high level function that learns, meaning it captures the distribution on the data. Nearly everything that collects data, we'll have some intelligence in it. I think IOT is kind of the wrong concept because it talks about the Internet. In this context, it will be more like the Intelligence of Things or something like that," according to Stevens.

FIGURE 4



Source: Argonne National Laboratory and Hyperion Research, 2019

The remainder of Stevens' presentation looked at what is currently possible with AI for Science for a number of scientific domains.

"So right now, it's routinely possible to learn a model from a data set where in which we have no understanding of the mechanistic theory underneath that dataset. And to build a model that's pretty accurate (depending on how much data you have and how good the data is) it works fairly well. We can do this in chemistry and materials. We can do it in drug design and lots of other areas as matter of routine. As of now, most of the datasets aren't huge," according to Stevens.

One example AI programmers can do is learn approximate solutions to inverse problems, which represents a significant portion of HPC problems being solved today. For applications like oil and gas or fast Fourier analysis, developers can now use machine learning methods to learn and approximate. These methods will not only solve the problem, but often can solve it much faster and with less compute resource.

Conclusions and Issues Going Forward:

- AI is transforming our regular life. It hasn't yet transformed our scientific and engineering lives, but it will.
- AI methods allow us to discover patterns in data that can lead to experimental hypotheses and thus link data driven methods to new experiments and new understanding.
- Researchers are already on their way to integrating machine learning into applications involving search and design such as materials, drugs, biological engineering, and manufacturing processes.
- AI needs HPC. Pre-exascale systems like NERSC 9 (Perlmutter) and the three planned DOE exascale machines (Aurora, Frontier, El Capitan) will be excellent platforms for both simulation and for doing AI training. These accelerated systems will be good at learning and reasonably good at inference.
- Within 10 years, AI for Science will be the driving application, and nearly all simulations will be managed by AI.

For more information or to view this and other presentations given at HPC User Forums dating back to 2008, visit www.hpcuserforum.com.

About Hyperion Research, LLC

Hyperion Research provides data driven research, analysis and recommendations for technologies, applications, and markets in high performance computing and emerging technology areas to help organizations worldwide make effective decisions and seize growth opportunities. Research includes market sizing and forecasting, share tracking, segmentation, technology and related trend analysis, and both user & vendor analysis for multi-user technical server technology used for HPC and HPDA (high performance data analysis). We provide thought leadership and practical guidance for users, vendors and other members of the HPC community by focusing on key market and technology trends across government, industry, commerce, and academia.

Headquarters

365 Summit Avenue
St. Paul, MN 55102
USA
612.812.5798
www.HyperionResearch.com and www.hpcuserforum.com
