Exascale Computing Project Industry Council

Meeting Notes: April 18, 2018

Oak Ridge National Laboratory (ORNL)

Host Welcome, Safety talk, Agenda Review
Suzy Tichenor, Co-Executive Director, ECP Industry Council
David Martin, Co-Executive Director, ECP Industry Council

Suzy Tichenor and David Martin opened the meeting by welcoming the ECP Industry Council and ECP Leadership team and reviewing the goals for the day.

Welcome Talk
Dan Hoag, Federal Project Director, DOE Oak Ridge National Laboratory Site Office

The Exascale Computing Project (ECP) is a high-priority for the US Department of Energy (DOE), which is making a $1.7–2.1 billion investment through the 2022–2023 timeframe. The current administration has given ECP tremendous support.

Dan Hoag's role is federal project director for the DOE Office of Science. Hoag seeks active participation from the Industry Council and industry at large. One question he would like answered is how ECP and the Industry Council can evaluate and manage project progress.

Hoag knows industry needs to more narrowly focus its objectives and increase interaction both formally and informally to create ongoing collaborations.

Chair Transition from Michael McQuade to Dave Kepczynski
Michael McQuade, Outgoing ECP Industry Council Chair

This meeting (04/18/18) is Michael McQuade's last official meeting. He recently retired from United Technologies Corporation.

McQuade and ECP Director Doug Kothe decided the Industry Council needed a chair who is active in a company. Dave Kepczynski, Chief Information Officer at GE Global Research, has a great knowledge of what high-performance computing (HPC) can do in industrial applications, and he knows how the labs can work with all the industry partners.

Dave Kepczynski, Incoming ECP Industry Council Chair

General Electric (GE) believes that HPC is a key part of US competitiveness. While the computational infrastructure within GE is not at the scale operated by the US government, the

company derives enormous value from both its internal resources and from competitively-won time at DOE leadership computing facilities. All of industry should be looking at how it will be able to leverage exascale.

Industry needs align with DOE/government and academic trajectories for HPC hardware and software (applications and ecosystem): (1) scalable, affordable-at-scale, and validated; (2) flexible, portable, open and interoperable; (3) supported and user-friendly. Ultimately, what is being created will be something companies will bring in as part of their HPC environments.

US Exascale Computing Project: Director's Update—DOE Budget and ECP Timeline
Doug Kothe, Director, Exascale Computing Project

ECP wants to share its results, understand companies' problems they need to solve, get to know companies' business models, share ECP's risks and challenges, understand how ECP can implement a proactive and mutually-beneficial collaboration with independent software vendors (ISVs), solicit companies' formal help in reviewing ECP's project management approach, and ask if ECP's vision and mission are clear and compelling.

ECP has an important review in August, so it does need feedback about where it is and where it is headed.

At a high level, DOE and the Office of Management and Budget (OMB) and others want to know how they can track ECP progress. There are lots of activities within DOE and the Exascale Computing Initiative (ECI). What are industry's milestones? ECP is a project that's on a watch list, which means it's big, it's interesting, or it's in trouble. ECP is hopefully on the first two.

The budget that ECP received for 2018 will use the increased funds to mitigate risk. Meeting participants discussed the current budget and the outlook for FY19 and FY20.

On April 9, 2018, CORAL-2 (**C**ollaboration of **O**ak **R**idge, **A**rgonne and **L**ivermore) released a request for proposals (RFP) for 2-3 exascale systems. ECP is paving the road with different vendors and getting the ecosystem as rich as possible.

Staff is advancing ECP toward CD (Critical Decision)-2 approval by 4Q FY19. ECP's review of project management is in August. ECP is looking for Industry Council participation.

Applications Development Focus Area Update: Major Accomplishments since January Call, Outlook on Major Milestones and Deliverables
Andrew Siegel, Director, ECP Application Development

This entire process began by looking at the DOE mission. Which mission areas could see significant progress with faster computers? The ECP staff identified 20–25 applications that pinpointed, in theory, what they thought they could do with exascale. The objective of the

Application Development focus area is to be able to efficiently run and to do the science that was envisioned and promised in this initial analysis.

The five different application areas in ECP are: chemistry and materials, energy (energy systems like reactors), earth and space science (astrophysics, computational cosmology, and climate), data analytics and optimization (biology, urban systems, and CANDLE), and national security.

Ideally, ECP wants all 22 mission-critical applications to be able to run on an exascale machine. Second, ECP wants to be ready, in the first several months after system delivery, to be able to do science on the machine.

Every application project has defined a metric for its baseline and what it means to improve its performance over that baseline. The principal investigators (PIs) will need to pick an existing system and sprint on that system. It's a stretch for every application to run everywhere. Certain systems will be targeted.

Everyone had to get a baseline on existing machines.

The next machine is the intermediate: Summit. Summit is the only bridge to exascale. There is no other intermediate.

A21 will be the exascale system at Argonne. ECP involvement happens via Path Forward and the ECP/Argonne Leadership Computing Facility (ALCF) engagement plan.

There are power constraints and money constraints. It's not plausible to just take existing applications and scale them up larger. It isn't easy to transition already existing applications to every potential exascale system.

At a high level, the ECP staff have a pretty good idea of the requirements. They will have failed if these 22 applications can run only on one system.

Some applications only need to accelerate a small portion of the code.

ECP needs to engage with HPC vendors, become aware of emerging technologies, and drive new science and engineering opportunities and numerical approaches to key features of the hardware.

Driving Toward Exascale: What Industry, ISVs, and Their Customers are Doing to Prepare for Exascale Technology
Mark Meili, Procter & Gamble
Dipankar Choudhury, Ansys, Incorporated (Calling in)
Frank Ham, Cascade Technologies, Incorporated
Sam Mahalingam, Altair Engineering, Incorporated

The panelists gave an introduction of their company and roles and then discussed how ISVs are making the transition to exascale.

Frank: The message of compute and the system-level of compute was very powerful. That's what makes me optimistic. What makes me pessimistic is that I want to take this technology to the masses, but the software infrastructure needed to leverage those systems is just not there, either culturally or financially.

The clear majority of jobs are using less than 5% of the cores on big systems. It's a cultural thing—there are no mandated targets. And secondly, software development just isn't keeping pace with the hardware.

ECP Leadership Team Comment: To do science well, you must have a spectrum. The Department of Defense systems don't have the utilization for a larger system. That's why the Oak Ridge Leadership Computing Facility (OLCF) and Argonne Leadership Computing Facility (ALCF) push the large jobs.

Frank: Companies usually have a business imperative to solve a problem. At that point, it motivates the need to compute at a large scale.

ECP Industry Council Member Comment: We are analyzing and optimizing at sub and part levels rather than full and ecosystem levels.

Mark: We don't even try larger jobs, because a lot of times, they don't run faster. So, we just run more jobs. Hardware cost has dominated. Historically, how much money we could spend on hardware wasn't dependent upon the power bill. That's why these new architectures must work. Because if they don't, from an industrial level, this is as big as we are going to get. We won't be able to do these kinds of problems because they won't be financially viable. Most of our commercial scales don't leverage today's architectures, much less tomorrow's. If there isn't a trickle-down, we are going to have the haves and have-nots. There will be big problems we run in a few places, and then there will be everything else.

There also aren't a lot of administrative skills to run these kinds of machines and operations. Systems administration and code porting skills are in short supply.

Sam: In some use cases, the turnaround time is very important. The model size has grown but you still have the same turnaround. You need a lot more hardware on which the computation can take place.

Mark: I think many of our industries have built our work processes on this overnight, 16-hour worknight turnaround. What would be different if that were half an hour? The answer is: a lot.

ECP Leadership Team Comment: Your comment about people is probably a top 3 risk for us. Finding good people is harder than losing them. Performance engineers, people who love to get down in the codes and architecture are really hard to find. People are an important risk for us. The third point,

your laptop in 5 years might have 5,000 cores and accelerators. So, node-based performance is really important.

ECP Industry Council Member Comment: We want affordable, scalable, interactable, portable, robust. Have we targeted applications, have we begun development on refactoring for exascale? We have targeted five in this mission-critical bucket.

Dipankar: To reach the scalability we desire requires ruthless inspection and refactoring of almost everything in the code. Unfortunately, when it comes to our existing software in production, these projects we just talked about are very pointed. We are focusing on new code for scaling to exascale.

ECP Leadership Team Comment: The obvious problem is that all paths to exascale require significant progress. A large fraction of the performance of the code is concentrated in a relatively small number of lines or functions. People can focus on a very small piece and do effective tuning.

Frank: Sometimes the large simulations become so cumbersome, it's challenging to extract the real information from the data. The machine learning approaches can help solve this problem. We are definitely mapping everything we can right now.

ECP Leadership Team Comment: As a company, have you identified certain applications that are critical that you want to run at exascale? And if yes, are you taking steps with universities, etc. that will get you there?

Mark: So yes. We really don't write our own codes. We use either the commercial codes off the shelf or open source as it exists without a lot of modification. In terms of actively engaging, we are starting to do that. We are crawling before we walk, much less, run.

What would be the mechanism for some of the active research that's going on and how certain algorithms can be mapped? What's the way to trickle down that technology into the ISVs or other methods? Other than the ISVs smartly hiring people out of grad programs, is there another way that we can be more purposeful?

ECP Leadership Team Comment: We can work on your code. There are "funds-in" situations, and these must be approved by our institution. We are trying to push everything open source.

ECP Leadership Team Comment: DOE is very open to software vendors getting access to the architectures. There are user agreements industry must sign, but we have unique ways we can do

that to protect the software you're developing. Don't let lack of access to large systems be inhibitive for you. We can take that away.

Dipankar: The access to these machines is an extremely important piece, but to me it's a downstream activity. More valuable to us is simply knowledge. I was trying to expose the

adversity of methods and applications. If we had a workshop setting where we could come up to speed on the lessons learned in ECP would be helpful.

Some kind of launch event that would help us get that activation energy and expose what is out there and push this interaction and find out what kind of information is out there, would be helpful.

ECP Training and Workforce Development—Training Program & How to Get Connected
Ashley Barker, Principal Investigator, ECP Training and Productivity

There are several ways industry partners can stay informed about things ECP is doing. The first and foremost is the ECP public website. All the events are listed on this website. The slides are listed. Videos of the events and some documentation are listed as well. The OLCF's newsletter is also available. OLCF also has a blog post. The cleanest way is to go to the OLCF website and go to "Training," and you'll be able to see all the OLCF's events.
https://www.exascaleproject.org/training-and-resources/

ECP staff are trying their best to make everything available to anyone outside of the ECP project.

OLCF for example, also has a YouTube channel where it posts its training videos. So far OLCF training has been through tutorials, webinars, workshops, and deep dives. It has focused on the software technologies being developed inside of ECP and made sure everyone knows where they are, what they are, and what they are being used for.

Hardware and Integration Focus Area Update—Major Accomplishments since January Call, Outlook on Major Milestones and Deliverables
Susan Coghlan, Deputy Director, ECP Hardware & Integration

One of ECP's primary drivers in forming this group is that it needed a better partnership within the DOE computing facilities. The applications and software need machines to run on or they aren't useful.

ECP is focused on providing support to the facilities. A key goal for ECP was getting the facility engagement plan worked out with each of the facilities. Most of the work has been with the three Office of Science facilities: ALCF, the National Energy Research Scientific Computing Center - NERSC), and OLCF.

ECP wants to give the teams what they need for their specific application. The facilities utilize their funding to best work with those teams.

The next step is for ECP to start getting the application teams tied in with the facilities, get them connected.

Facilities have standard training they do and are experts at it. ECP wants to work together with them and leverage one another's training.

ECP's Julia White now is tracking ECP utilization of the facility computing resources in the ECP dashboard, and she's managing the allocations of compute time across the facilities for ECP projects. She does this through a Resource Allocation Council. The hours that are available for allocation to ECP projects were taken away from the facilities' other programs.

ECP also has an effort called the PathForward for funding hardware research and development. Public descriptions for the PathForward awards have been posted to the ECP website. You won't be able to find detailed information because it's NDA.

Hardware technologies focused in five main areas: node-level simulation, interconnect modeling, memory technologies, analytic modeling, abstract machines, and high-level system models.

Some of the challenges that exist are: funding for PathForward, building strong partnerships with the facilities, and shortened timeline due to late hardware and integration start.

Working Lunch—ORNL: Delivering Breakthrough Science for National Security and Economic Competitiveness
Michelle Buchanan, Deputy Laboratory Director for Science and Technology

Michelle Buchanan provided an overview of Oak Ridge National Laboratory and its strategic focus areas.

Software Technology Focus Area Update
Michael Heroux, Director, ECP Software Technology

ECP wants to extend technologies to exascale where possible. This includes primarily message-passing but also includes OpenMP and CUDA. ECP is also investing in capabilities that are disruptive to the way it must design its algorithms in codes.

ECP has 77 milestones/activities going on right now. 426 milestones are planned for FY18–FY19.

55 projects contribute to 89 unique software technology (ST) products. Spack is very important. 43 of the projects have Spack packages and 21 have Spack packages in progress. Of the 89 products, 81 can be built from source.

Software development kits (SDKs) allow for collections of similar products to be brought together so they can be built, put together seamlessly, and provided to the user base. SDKs provide seamlessness across a software stack.

The software review board allows for better-designed software because there are multiple reviews of the same software.

Vectorization improves with each generation of processors. The more parallelism that's visible in a code, the better the code will perform.

One of the things ECP is exploring is how to enable users to have a common application programming interface to on-node parallel resources. This is the first effort of transferring control from one execution environment to another.

Qthreads is a nice proxy for the next-generation of OpenMP. Users can study on-node parallelism underneath the message-passing interface by using Qthreads instead of waiting for OpenMP to be available.

- **Software Technology** Challenges: Qualitative; massive concurrency; multiscale; data-driven

- **Software Technology** Next Steps: Large collection of widely-used DOE products, fully tested and ready for use

ECP Project Management
Kathlyn Boudwin, Director, ECP Project Management

ECP Project Management's use of tools to assess progress merges traditional earned value management tools with Agile software management tools.

An independent review is scheduled for August 7–8. This review will be chaired by Robert Wunderlich. This review is sponsored internally and has to do with the way ECP has tailored the DOE order that's used to manage large projects. ECP is looking for independent eyes to look at how it has done its tailoring, especially regarding Agile project management.

ECP has Primavera (traditional earned value management [EVM]) that's used for formal reporting to DOE.

Reporting occurs on a quarterly basis rather than a monthly basis. This allows ECP to include more milestones and get a better idea of what milestones it is meeting.

Dashboard development is happening to allow for data collection from various sources with drill-down capability. A variety of interactive chart types and personalized views have been incorporated. It can help track progress and track progress towards the key performance parameters (KPPs). It helps track the system usage at facilities. The financial tables can be organized by work breakdown structure and by institution as well.

The goals of the August review are: (1) reviewing EVM tailoring, (2) reviewing ECP's merging of Agile software management tools and traditional management methodologies, (3) reviewing the project dashboard, and (4) reviewing the plan to achieve CD-2.

ECP is looking for suggestions for several Industry Council reviewers for the August reviews with expertise in project management with Agile expertise.

Member Roundtable Discussion

The meeting closed with a roundtable discussion among members about the progress made since the last meeting and opportunities moving forward. Comments included:

- I think we have made tremendous progress in ECP. When you look at the industries, I think, there are a lot that are in the 1–5 petaflop range of computing. We are investing significantly in our IT infrastructure. So clearly, we are marching up. We talked about companies having the hardware—we have the critical software vendors, but you're going to see more interaction.

- I think there is tremendous progress. Having said that, I also feel that there should be a structured form of collaboration for information and sharing. If we have information exchanges, we can learn from each other.

- I liked what Dave said about using these systems to test it out. We do things in a very repeatable way. How you tune or optimize your system for something that's constantly changing is very interesting—and how you're going to do these measurements, how you reach your KPPs. We really want to understand how you can optimize that, but whatever you figure out, we want to adopt that. Things like provisioning and automating a lot of the management of the system are going to be important.

- All of us need to think about how we can work with the facility here. We are users, not developers, of simulations. We certainly don't have the ability to develop code, so we are totally reliant on the ISVs. But I think we probably need to think about how we enable that to happen; otherwise, it's going to only be here.

- Oil and gas had an HPC-specific industry conference in March. We found that to be very useful as well as having a side meeting that was ECP-specific. The key constraint is talent.

We are very constrained. The easier you help us engage and streamline that interface, the better. There was another action around this leadership—I can see where that could help put a focus back on this. How do we make sure we don't lose our competitive edge in HPC?

- The sense of progress and confidence in the team is real…to me it seems as if people feel like this can be done. It's a lot of work, and I don't want to minimize that. When we started this, I heard a lot of people say that they don't know how this can be done.

  I think that dashboards are a waste of time if you're doing them manually. I would spend every moment I can finding dashboards that are automatically generated and keeping track of what the team is doing on a day-to-day database.

  For the ECP team, one of the things I think is important to look at is: how quickly will it take to get science done on Summit? What happens on Frontier? What happens on A21? Real science deliverable speed is something we really need to track. Are there mechanisms that allow people to engage over multiple generations?

- There's a lot more information available than I thought was available, but with that said, there's still likely some kind of intervention that we need to drive people to go get and use that. We need to figure out where we need to make an intervention. We also need to do that on the workforce side. If we don't make intervention happen, it's not going to happen. That last one was an offer.

- It's interesting to see where the power numbers are coming in. They're very reasonable, although huge. The key is to be able to leverage it. It's important for me to hear about the connections and ways you are reaching out into the community.

- The Industry Council can support this effort. In the Red Team software validations, we can be integrated into stressing the system, finding unknowns, etc.

  A lot of our baselining is what we know today. That may be a little bit of a concern. A real win-win would be if we can get much higher fidelity and performance, and the physics is the same and error rates have gone down. A lot of the proposals, or a portion of them, are in machine learning, deep learning, neural networks, or the AI space. "Legacy code–optimized" may not get us as big of a performance increase as starting over. We are a cross-section of all our industry segments. We can widen the audience and it can become a greater part of the exascale journey.

Engaging Industry Communities
David Martin, Co-Executive Director, ECP Industry Council
Suzy Tichenor, Co-Executive Director, ECP Industry Council

This was to be a report on a specially-organized workshop between ECP leadership and the oil and gas industry technical leaders at the Rice HPC Oil and Gas Conference.

In the interest of time, this discussion was postponed until the next quarterly call. David and Suzy will share their ideas for broader industry outreach and engagement.

Meeting Action Items

- Make all slides available.

- Issue consolidated meeting notes.

- Pick location and timing for next meeting, after the October IPR (Independent project Review). (Date forthcoming)

- Consider extended format for meeting.

- Convey ECP training opportunities to Industry Council.

- Develop a plan to broadly expose the Industry Council and their technical leaders to the large collection of software and applications being developed and "productized."

- Work with Industry Council members to identify industry experts from their firms (or elsewhere) to participate in August Agile Project Management/Earned Value Management Review for large software and application development projects.