Sandia National Laboratories

Exceptional service in the national interest

EXASCALE COMPUTING PROJECT

History and Lessons Learned

Douglas B. Kothe

Chief Research Officer

Associate Laboratories Director of Advanced Science & Technology

Sandia National Laboratories

HPC User Forum

Paris, France

October 7, 2025

"ECP provided the resources, time, mission, and structure for a large, multi-institution team of highly self-motivated, educated, and creative problem solvers to research, develop, and deliver software libraries, tools, and applications that represented a step-function improvement in computational science capabilities"

- Mike Heroux

U.S. DEPARTMENT of ENERGY NNSA

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

SAND2025-04777C

DISCLAIMER

Content expressed herein is solely the opinion of Doug Kothe, with all comments and any inferred associated blame aimed directly back at Doug.

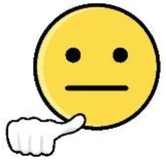
THE EXASCALE COMPUTING PROJECT STARTED ROCKY AND FINISHED WITH NO FORMAL NEXT PHASE

- In 2016-2017, a large amount of funds were transferred from the Advanced Scientific Computing Research (ASCR) Division to ECP HQ, leaving many staff engaged in ASCR research affected.
- ECP leadership made initial portfolio choices without full recognition of funding sustainment for affected staff.
- Some academic PIs for software technology were targeted, with no broad RFI; and private sector PI input was mostly limited to PathForward.
- DOE labs led application efforts, narrowing from 150 white papers to 25 selections, relying on DOE sponsors and external SME reviewers, with academia and private sector staff being subproject team members (and not PIs).
- ECP advocated for a transition to operations seeking sustained support similar to Advanced Simulation and Computing (ASC) (NNSA/NA-114).
- Advanced Scientific Computing Advisory Committee (ASCAC) provided valuable and actionable recommendations on the ECP Transition consistent with our thinking.
- Some grassroots efforts remain and look promising (e.g., High Performance Software Foundation - https://hpsf.io/). Software as a facility/consortium has promise. A single organization with connections to all other HPC entities would enable DOE S/W to be a first-class citizen in the HPC ecosystem.

ASCAC SUBCOMMITTEE ON ECP TRANSITION

Recommendations (Roscoe Giles, Subcommittee Chair September 2020)

- Establish a shared software stewardship program in ASCR.
- Address current and future software needs.
- Provide collaborative application support.
- Enhance engagement with industry and academia.
- Implement modern project management practices.
- Increase investment in ASCR research.
- Foster a stable environment for basic research.
- Distribute research software effectively.
- Encourage researchers' re-engagement in "blue sky research."
- Retain the existing workforce.
- Strengthen connections with universities and the ecosystem.
- Develop career paths for scientific software professionals.
- Maintain DOE/ASCR's leadership in advanced computing.
- Collaborate with national stakeholders in other agencies.



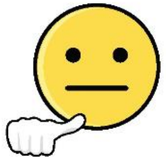
ECP’S SOFTWARE EFFORTS WERE POSITIVE, BUT THERE’S ALWAYS MORE TO DO . . .

Achievements/Successes

- Embraced emerging container technology despite ongoing candidate development.
- Spack gained attention and effectively addressed “Unified Build Theory”.
- Achieved integration and deployment for 70 products with adequate resources.
- ECP made initial strides in defining "performance portability" for exascale environments (i.e., KPP-3)
- Invested in reduced and mixed precision algorithms, needing further development for traditional ModSim.
- Moved LLVM forward for Fortran, C, C++, and MLIR.
- Software abstraction layers (Kokkos, Raja, OCCA) adopted by many apps & influenced language standards.

Challenges/Obstacles

- Energy efficiency and low-energy software approaches were not explicit metrics
- Contributed to the unconstrained explosion of advanced software workflow components.
- Missed opportunities in the evolving AI software stack.
- No scope for quantum software support and prepping for resiliency, fault tolerance, self-healing.
- Computational steering, now resurging (since 1990s) & more relevant with AI drivers, was not directly considered.
- Faced challenges in obtaining timely vendor-based hardware specs for debugging tools.
- Limited progress in task-based run-time operating systems and data-flow architectures. Well-defined project scope made us gun shy.



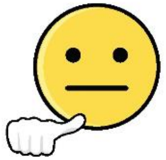
ECP WAS REQUIRED TO "PROJECTIZE" ITS R&D SCOPE

Achievements/Successes

- Formal project execution enabled co-design of a vertically integrated OS and software stack, leading to significant tech outcomes like E4S (e4s.io) and Spack.
- This amount of stable funding for this long is a game-changer, particularly for applications.
- A unified project community with quantified metrics (KPPs) incentivized collaboration and tracked progress, with dedicated CI/CD funding for software stack & apps perf at HPC facilities.
- Centralized tools (Jira, Confluence, Slack) ensured team alignment.
- Proper milestone constructs can focus efforts, drive teams with a sense of urgency, clarify progress measures.
- Enhanced funding provided greater purchasing power than typical DOE programs (no LDRD or SBIR tax – 10% savings).
- Project-centric approaches favored by the Hill and OMB, with measurable outcomes and ongoing cost/scheduled performance.
- ECP leadership was granted the agility to make scope and funding decisions (with a contingency reserve), enabling initiatives like resourcing broadening participation and pivoting during COVID.

Challenges/Obstacles

- Lost time and additional cost associated with unneeded DOE O 413.3B scope and process (~10% in cost and time?)
- Heavy review burden onerous – leadership team mostly AWOL for 2-3 months prior to each review (perhaps that's not bad?)
- Technical staff can become confused and frustrated by formal project management constructs
- Absorbing research programs can result in short-term focus at expense of longer-term goals
- Project "Mission Need Statement" (estd 2016) viewed as "Commandments" set in stone
- Milestone targets need be clear, but over-specified accomplishments can stifle innovation and agility
- Too process-heavy, too much reporting, too many meetings, etc. Many teams found having ECP funding painful early in the project



ECP'S PATHFORWARD AND HARDWARE INTEGRATION PROGRAM HAD IMPACT BUT NEEDS FURTHER SUSTAINMENT AND GROWTH

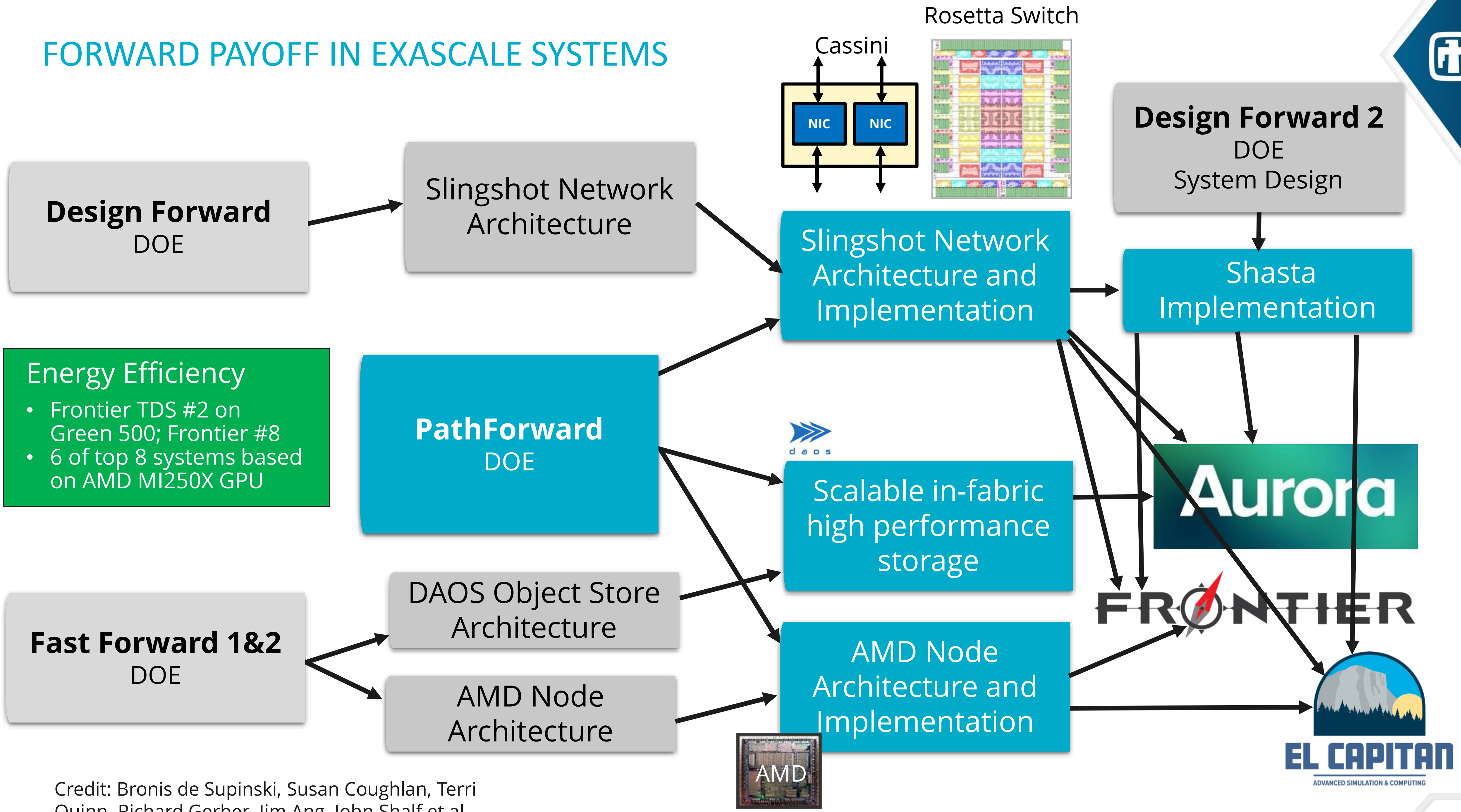
Achievements/Successes

- PathForward built on DesignForward and FastForward, maintaining momentum with U.S. hardware vendors for high-risk R&D aligned with DOE needs.
- It prepared U.S. industry for exascale procurements and aimed to enhance global competitiveness, ensuring DOE can acquire at least two exascale computers.
- The initiative involved over $280M in DOE investment (40-50% industry cost share) for six vendors (Intel, AMD, NVIDIA, IBM, HPE, Cray) over four years, with 267 milestones successfully delivered.
- HI accelerated access to testbeds and software readiness, addressing a critical need in ECP.
- Dedicated CI is essential in a separate environment from production resources.
- Training events and hackathons improve software development and performance.
- A common interface for integration testing is valuable for HPC developers.

Challenges/Obstacles

- Continuing a PathForward follow-on (PathForward II) with expanded scope for a S/W element (Software PathForward) was a missed opportunity. These investments need ongoing - now more than ever with AI and quantum.
- More agility in milestone commitments could have increased the ROI for both DOE and the vendor, e.g., with perhaps more focus on novel memory technologies, optical, etc.

FORWARD PAYOFF IN EXASCALE SYSTEMS



Credit: Bronis de Supinski, Susan Coughlan, Terri Quinn, Richard Gerber, Jim Ang, John Shalf et al.

ECP HAD LITTLE SCOPE IN ARTIFICIAL INTELLIGENCE AND HAD NO QUANTUM COMPUTING SCOPE

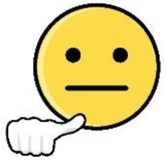
Achievements/Successes

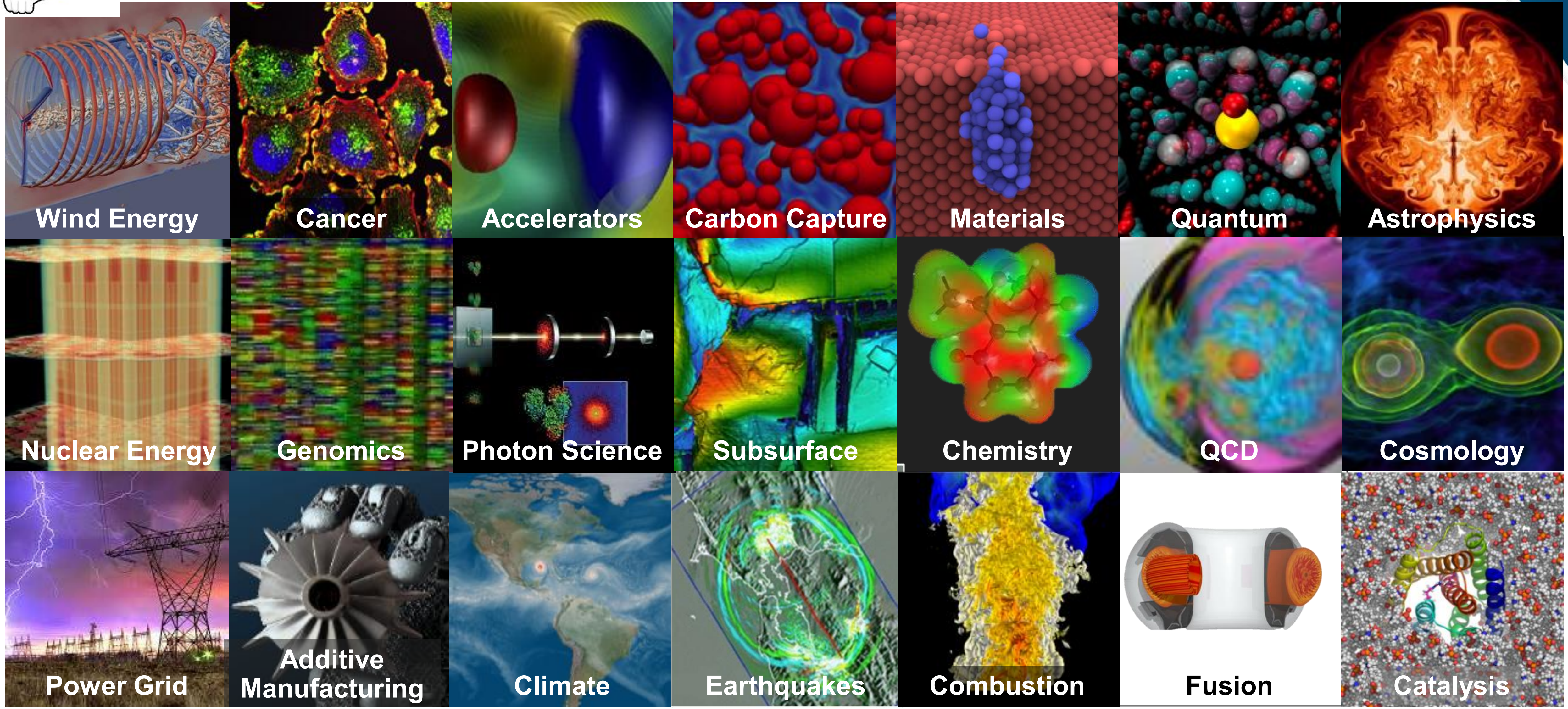
- Project formalism with clear scope and change management “helped” us stay focused in a rapidly changing environment.
- Took 6 months to convince sponsors to at least start *something* formally in AI – we recognized that doing something deep in AI (e.g., foundation models) was naïve but we felt strongly that some of our staff needed to be thinking and working 24x7 in AI to spin up and help augment / accelerate our existing workflows.
- For a quick start we set aside funds for 8 labs (equal $500K increments) and asked the team (ExaLearn) to “self organize and choose a leader”. It worked!
- Proposed simulating ~50+ logical qubits at exascale, but DOE advised us to "stay in our lane.” Probably good advice.

Challenges/Obstacles

- ECP arguably invested less than 5% in AI, despite the rise of AI during the latter part of ECP. We could have pivoted more aggressively, but not at the expense of descoping other efforts. Should we have pursued extending the project completion date and increased the Total Project Cost?
- Could ECP have tackled performance inefficiencies in AI training and helped to understand emergent behavior of 1T+ parameter models?
- Could ECP have been a hub for a quantum computing software stack, given the efforts of 5 DOE-funded National Quantum Initiative centers?
- Lean in more on *HPC for AI and AI for HPC*. Opportunity missed.

ECP DID NOT HAVE ONE “KILLER APP” BUT INSTEAD HAD A WHOLE GRAB BAG OF APPLICATIONS





Spread thin to cover 10 DOE program offices; haves and have nots; all DOE led – targeted call; one app per domain; quantified outcomes

EXASCALE APPLICATIONS: POTENTIAL OUTCOMES AND IMPACT

Will be far-reaching for decades to come?

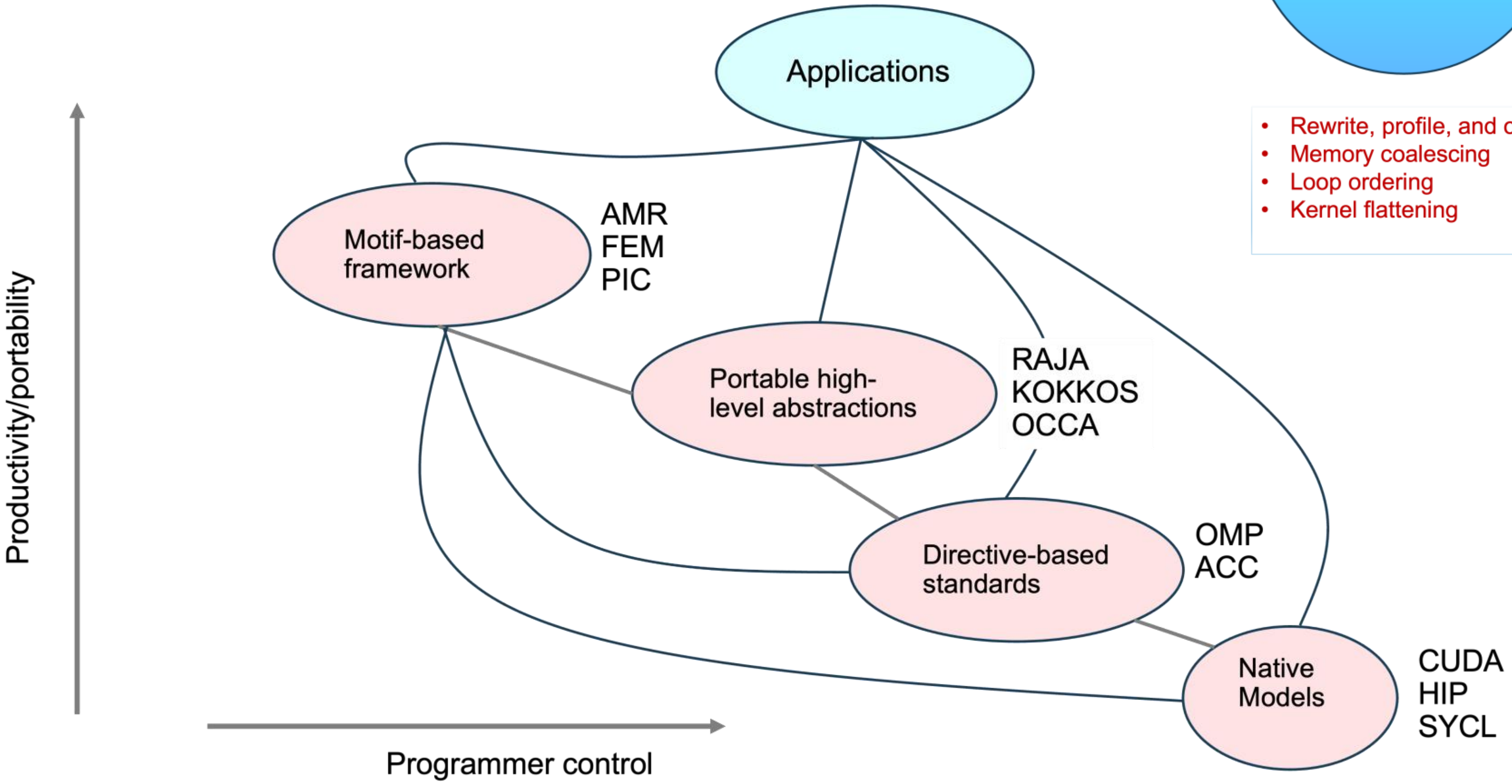
- Predictive microstructural evolution of novel **chemicals and materials for energy** applications.
- Robust and selective **design of catalysts** an order of magnitude more efficient at temperatures hundreds of degrees lower.
- Accelerate the widespread adoption of additive manufacturing by enabling the **routine fabrication of qualifiable metal alloy parts**.
- Design **next-generation quantum materials** from first principles with predictive accuracy.
- Predict **properties of light nuclei** with less than 1% uncertainty from first principles.
- **Harden wind plant design** and layout against energy loss susceptibility, allowing higher penetration of wind energy.
- Demonstrate commercial-scale transformational energy technologies that **curb fossil fuel plant CO2 emission** by 2030.
- Accelerate the **design and commercialization of small and micronuclear reactors**.
- Provide the foundational underpinnings for a **‘whole device’ modelling capability for magnetically confined fusion plasmas** useful in the design and operation of ITER and future fusion reactors.

EXASCALE APPLICATIONS: POTENTIAL OUTCOMES AND IMPACT

Will be far-reaching for decades to come?

- Address **fundamental science questions** such as the **origin of elements** in the universe, the **behavior of matter at extreme densities**, the source of **gravity waves**; and demystify key unknowns in the dynamics of the universe (**dark matter, dark energy and inflation**).
- Reduce the current major uncertainties in **earthquake hazard and risk assessments** to ensure the safest and most cost-effective seismic designs.
- Reliably guide safe long-term **consequential decisions about carbon storage and sequestration**.
- Forecast, with confidence, **water resource availability, food supply changes and severe weather probabilities in our complex earth system environment**.
- **Optimize power grid planning and secure operation** with very high reliability within narrow operating voltage and frequency ranges.
- Develop treatment **strategies and pre-clinical cancer drug response models** and mechanisms for RAS/RAF-driven cancers.
- Discover, through **metagenomics analysis**, knowledge useful for environment remediation and the manufacture of novel chemicals and medicines.
- Dramatically **cut the cost and size of advanced particle accelerators** for various applications impacting our lives, from sterilizing food of toxic waste, implanting ions in semiconductors, developing new drugs or treating cancer.

ECP DID NOT MANDATE A PROGRAMMING MODEL



Port Code

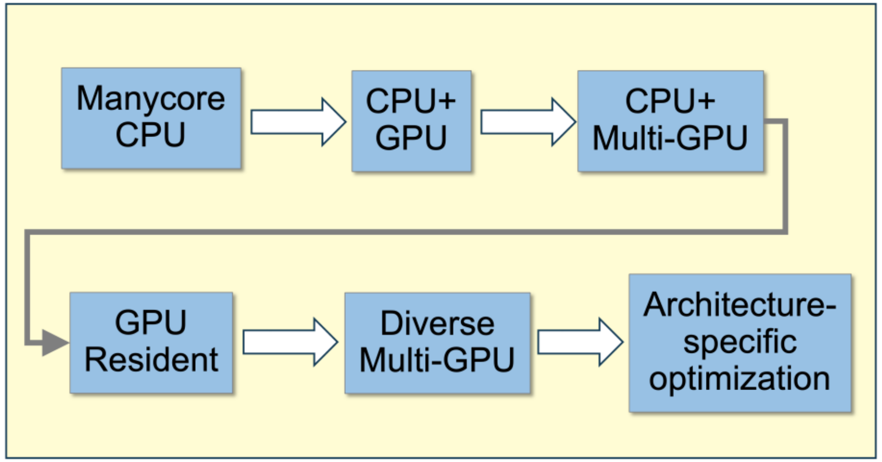
- Rewrite, profile, and optimize
- Memory coalescing
- Loop ordering
- Kernel flattening

Adapt Numerics

- Reduced synchronization
- Reduced precision
- Communication avoiding

Adapt Models

- Mathematical representation
- "On the fly" recomputing vs. lookup tables
- Prioritization of new physical models



Credit: Andrew Siegel, Erik Draeger

ECP HAD UNWAVERING SPONSOR MANAGEMENT AND STAKEHOLDER SUPPORT

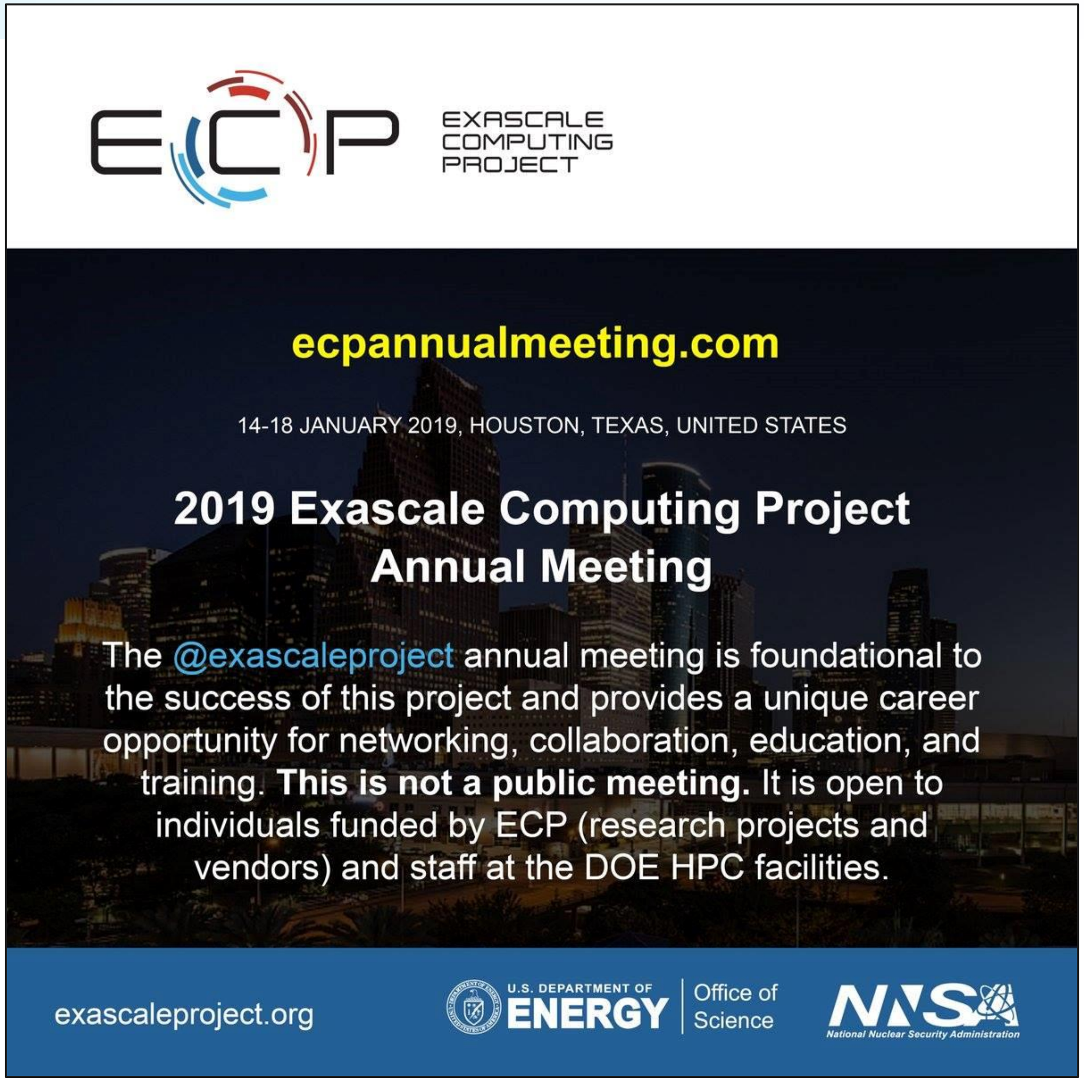
- ECP had only three federal (DOE) sponsors over its lifetime. ECP was fortunate to have a light Fed touch who empowered us, entrusted us, and also held us accountable. As a team we were all after the same thing – project success. ☺
- ECP's leadership, working with DOE sponsors, were able to put several best-practice constructs in place from the beginning: MOU between DOE SC & NNSA, MOA between 6 DOE labs responsible for leading ECP, Board of Directors, Board of Directors committee (the Laboratory Operations Task Force) overseeing execution, and the Industry & Agency council
- Direct consultations with ten DOE program offices ensured alignment with mission strategies and allowed for necessary adjustments.
- ECP staff engaged diverse audiences nationally and internationally, fostering advocacy and long-term collaborations.
- Positive feedback from Members of Congress and staffers on the Hill staff reinforced confidence in DOE lab capabilities, with ECP's achievements recognized in the Congressional Record. We are very proud of that. ☺
- Thanks to our DOE sponsors in ASCR, ASC, and DOE Lab site and field offices!

ECP EXPANDED, DEVELOPED, AND INTEGRATED OUR WORKFORCE

- **Project Management** : I hope we did not screw up thousands of researchers (especially early career) and hundreds of teams by exposing them to more rigorous project management and the use of community tools to help manage their subprojects. I'm going to claim "no".
- **Networking Opportunities**: ECP's support of all 17 DOE Labs and dozens of universities fostered valuable connections, precipitated by the Annual Meeting. (Thanks to Paul Messina and Barb Helland.)
- **Access to Tools**: Researchers gained new tools and technologies not otherwise possible. ECP drove this integration because of its project nature, causing a groundswell of support. Some of this integration was driven by ECP's PIs being competitive with each other. ECP has quantitative measures of integration
- **Long-term Benefits**: 7+ year ECP duration enabled students to complete degrees and early career staff to emerge as next generation leader in our community. What a great way to start a career!
- **Training Initiatives**: ECP's earnest effort in training (hackathons, etc.) acquainted staff with vendor and leading DOE experts in HPC, accelerating not only the learning process, but helping to build teams for the "next big thing". Vendors benefitted greatly from these relationships, not the least of which was staffing up their own RD&D efforts
- **Marketability of Computational Staff**: Evidence shows computational staff often become future leaders in DOE Labs (many are now LDs!), academia, and the private sector.
- **Looking Ahead**: Excited to see the achievements of ECP alumni!

ECP ANNUAL MEETING

A valuable resource for integration, and a "go-to" for the HPC community.



ECP BENEFITTED IMMEASURABLY FROM A HIGH-PERFORMING TEAM

https://www.tablegroup.com/product/dysfunctions/

EVIDENCE SHOWED ECP HAD A HIGH-PERFORMING TEAM

Thriving through team excellence!

Doug’s house (May 3, 2023)

DOE Secretary’s Honor Award
Ceremony (January 8, 2025)

QUESTIONS?

ECP'S MISSION NEEDS

Deliver a long-term, **sustainable software ecosystem** that can be used and maintained for years to come

- E4S deployed at HPC facilities around the US and the world
- 76 HPC products available for computing at all scales
- Performance portability tools developed and widely used

Promote the **health of the U.S. HPC industry**

- Six vendors funded under PathForward; outcomes realized in exascale systems
- Accelerator-based computing lowers cost of energy across the board
- The ECP Industry and Agency Council stimulates consumption of HPC resources

Ensure that exascale systems can be used to deliver **mission-critical applications**

- ECP applications demonstrate outstanding performance and capabilities at exascale
- Previously unattainable results in real-world challenge problems
- ECP lessons learned pave the way for many additional applications to leverage accelerator-based computing

Maintain **international leadership in HPC**

- Frontier is the world's first exascale machine – in part due to ECP/ECI investments
- The Aurora exascale system is accessible to all ECP teams
- 1000+ researchers trained in GPU computing

KEY PERFORMANCE PARAMETERS

KPP ID	Description of Scope	Threshold KPP	KPP Achieved at CD-4	CD-2 Threshold KPP Met or Exceeded?
KPP-1	**11 selected applications** demonstrate performance improvement for mission-critical problems	50% of selected applications achieve Figure of Merit improvement ≥50	**10 of 11 (91%)** applications met or exceeded FOM targets on their base challenge problem	✓ **Threshold exceeded**
KPP-2	**14 selected applications** broaden the reach of exascale science and mission capability	50% of selected applications can execute their challenge problem	**9 of 10 (90%)** DOE Science and Applied Energy and **3 of 4 (75%)** NNSA applications achieved their base challenge problem	✓ **Threshold exceeded**
KPP-3	**76 software products** selected to meet an aggregate capability integration score	50% of the weighted impact goals are met	**63.5 of 68.0 (93%)** points were achieved by the software products	✓ **Threshold exceeded**
KPP-4	Delivery of **267 vendor baselined milestones** in the PathForward element	Vendors meet 80% of all the PathForward milestones	**267 (100%)** of PathForward milestones met	✓ **Threshold exceeded**

INDUSTRY AND AGENCY COUNCIL

Effective Outreach

ECP-FUNDED TECHNOLOGIES AND PARTICIPANTS ARE LEADERSHIP CLASS

R&D 100 AWARDS

CANDLE (2023)
ZFP (2023)
Variorum (2023)
Flash-X (2022)
Mochi (2021)
SZ (2021)
Flux (2021)
Legion (2020)
Spack (2019)
UCX (2019)
Darshan (2018)
Swift/T (2018)
Nek5000 (2016)

HPC wire
People to Watch

Jeff Vetter (2021)
Bronis de Supinski (2021)
Lori Diachin (2019)
Doug Kothe (2018)

acm

Gordon Bell Finalist | ExaSMR (2023)
Gordon Bell | WarpX (2022)
Gordon Bell Special Prize | CANDLE (2022)
A. M. Turing Award | Jack Dongarra (2021)
ACM Fellow | Bronis de Supinski (2022), Rob Ross (2021), Rick Stevens (2020)
IEEE CS Ken Kennedy Award | Ian Foster (2022)
IEEE CS Sidney Fernbach Award | Salman Habib (2020)

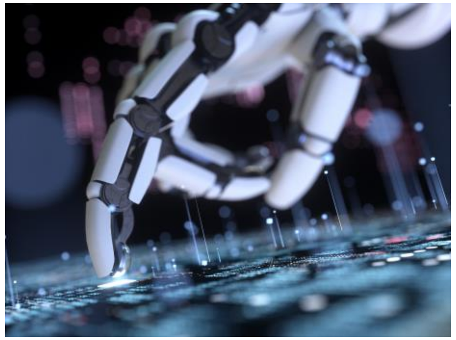
…and many more notables, including

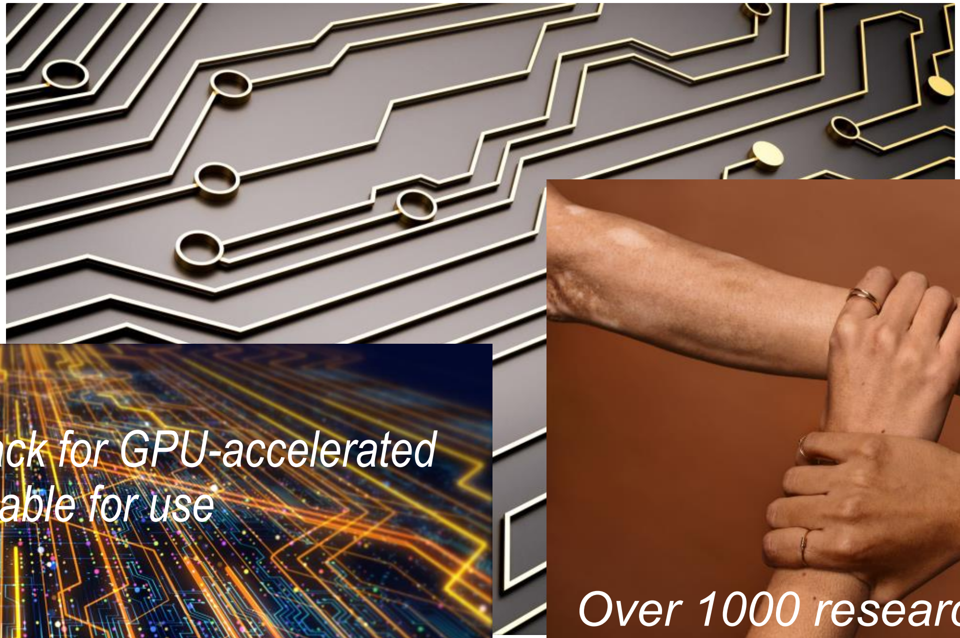
CRA Distinguished Service Award | Paul Messina (2018)
Best Open Source Contribution (For IPDPS 2023 paper) | ExaIO (2023)
SIAM/ACM prize in Computational Science and Engineering | SUNDIALS (2023)
The Ernest Orlando Lawrence Award | ExaStar and DataLib (2021)
SIAM Fellow | Mike Heroux (2019); IEEE Fellow | Rajeev Thakur (2022)

ECP’S IMPACT AND LEGACY

A suite of applications that will impact problems of national importance for decades to come

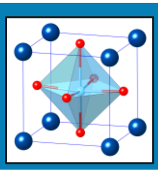
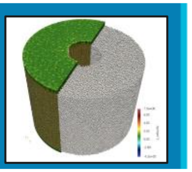
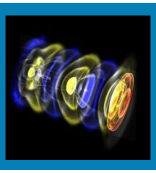
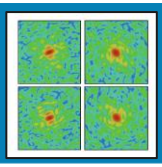
Best practices for running a large-scale software development RD&D 413.3B project





Best practices and lessons learned for thinking about how to program GPUs – moving the nation forward

ECP INTEGRATION EXAMPLES

ExaWind: advanced wind farm modeling		Sparse linear solvers optimized for strong scaling and GPU performance **from hypre and Trilinos**
QMCPACK: quantum Monte Carlo for materials		Batched dense linear algebra kernels significantly improved GPU performance **from SLATE**
ExaSGD: power grid optimization		Optimize sparse indefinite solvers developed and optimized for large-scale grid problems **from SuperLU and Ginkgo**
ExaSMR: small modular reactor modeling		Custom discretization designed and tuned for specific reactor assembly **from CEED**
WarpX: plasma wakefield accelerator design		Adaptive mesh data structures and solvers highly optimized for GPU performance **from AMReX**
ExaFEL: real-time light source analysis and reconstruction		Non-uniform FFTs designed to minimize data motion **from FFTX**

PROGRAMMING MODELS AND RUNTIMES

GPU-specific kernels

- Isolate the computationally-intensive parts of the code into CUDA/HIP/SYCL kernels.
- Refactoring the code to work well with the GPU is the majority of effort.

Loop pragma models

- Offload loops to GPU with OpenMP or OpenACC.
- Most common portability strategy for Fortran codes.

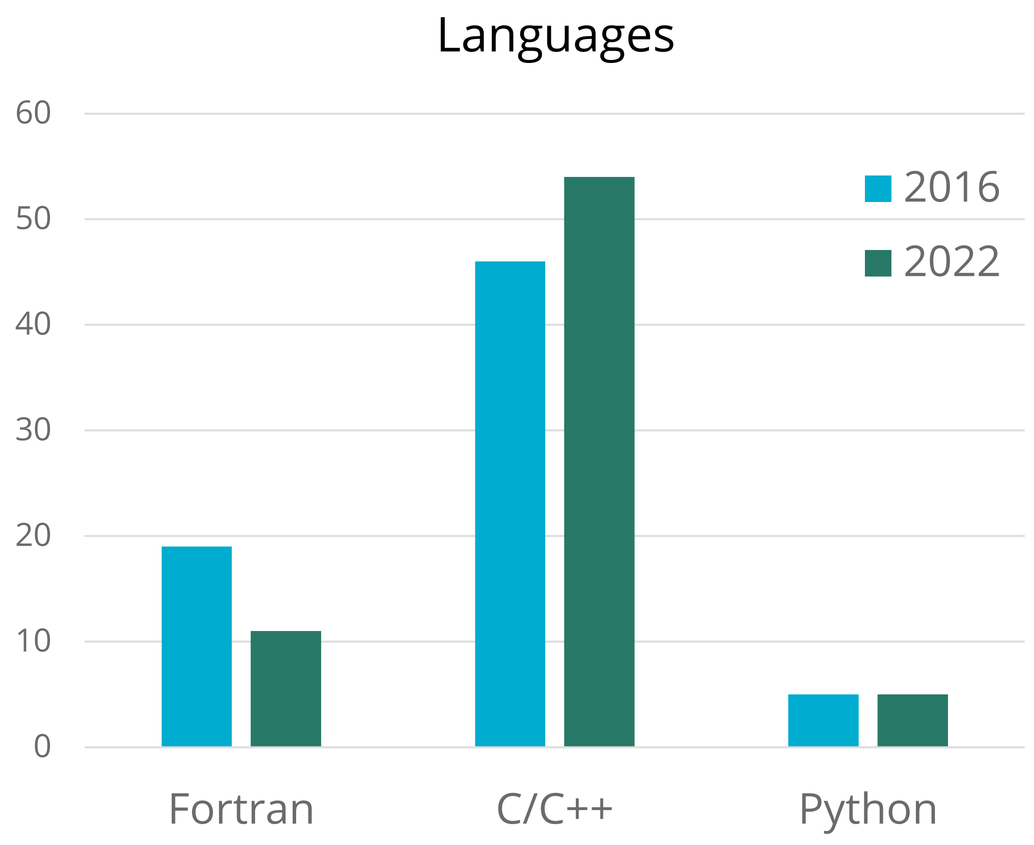
C++ abstractions

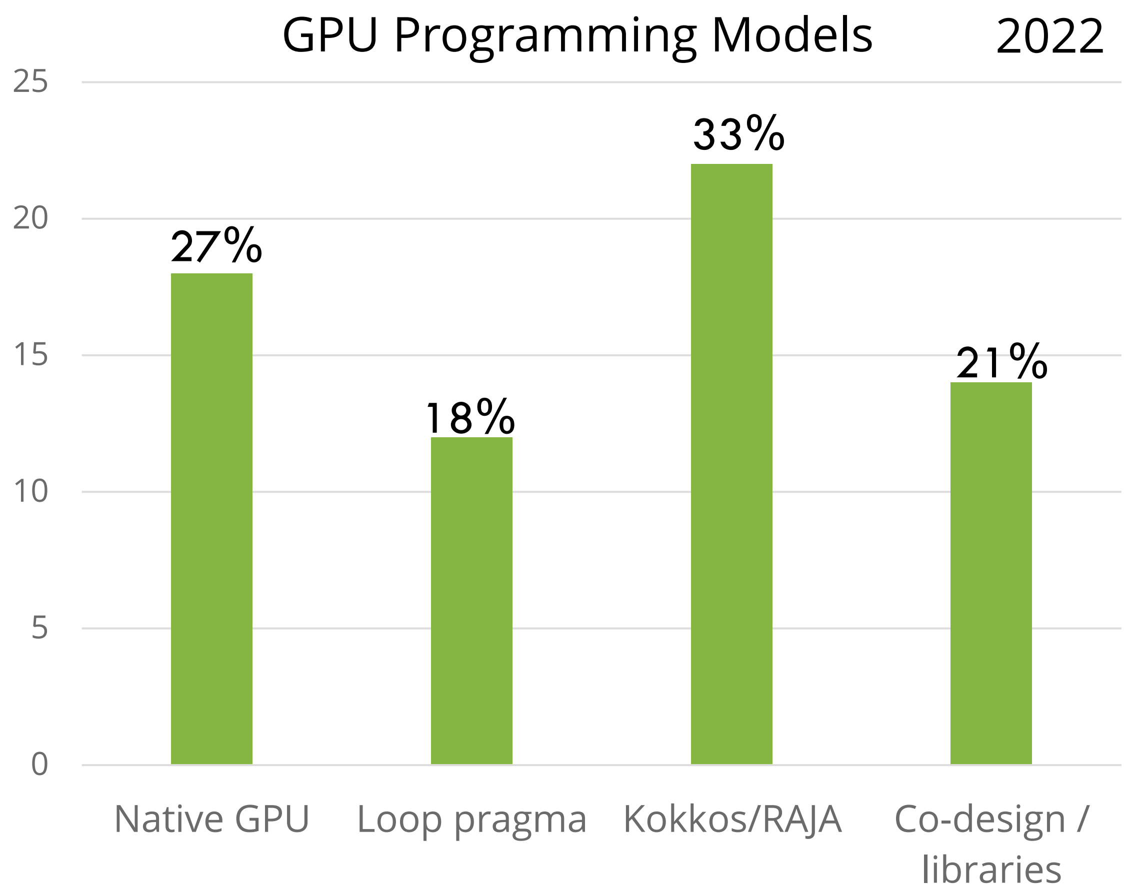
- Fully abstract loop execution and data management using advanced C++ features.
- Kokkos and RAJA developed by NNSA in response to increasing hardware diversity.

Co-design frameworks

- Design application with a specific motif to use common software components
- Depend on co-design code (e.g. CEED, AMReX) to implement key functions on GPU.

DISTRIBUTION OF ECP PROGRAMMING MODELS OVER TIME





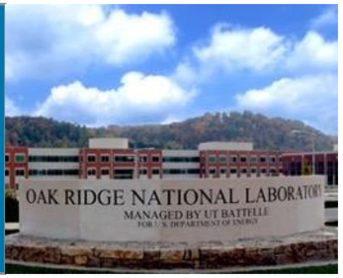
Many ECP applications started out using native GPU and loop pragma models before moving to C++ abstractions and co-design libraries

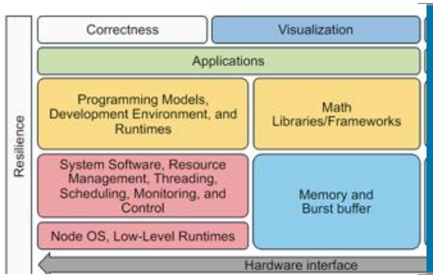
Evans TM, Siegel A, Draeger EW, *et al.* *"A survey of software implementations used by application codes in the Exascale Computing Project."* The International Journal of High Performance Computing Applications. 2022;36(1):5-12.

APPINT ENABLED APPLICATIONS AND IMPROVED THE EXCOSYSTEM FOR ALL

Collaborated via workshops, hackathons, & dungeon sessions to improve application readiness and performance on Aurora, Frontier & Perlmutter.

Collaborated with AD, ST, Facility staff, & vendors through the Facilities' Centers of Excellence on porting codes to exascale programming models.



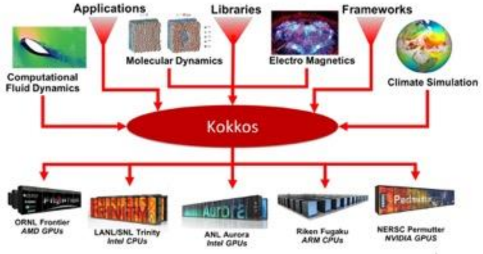


Tested applications with the exascale system software stacks and helped them perform at high performance at scale in each environment.

Created and tested proxy apps used as easy-to-run and test surrogates for key parts of larger codes.

Identified, reported, and tracked through resolution many bugs in the software stack, thus improving the environment for all users.

Created work-arounds for codes that enabled them to run on Frontier, Aurora, and proxy systems, while waiting for bugs or other issues to be addressed.



Helped implement and test portable programming paradigms like Kokkos, Raja, OpenMP.

Managed contracts with vendors to implement HIP for Aurora and SYCL on Frontier to enable application portability.