

NERSC Application Readiness Process and Strategy

Charlene Yang
Application Performance Group
July 1, 2019

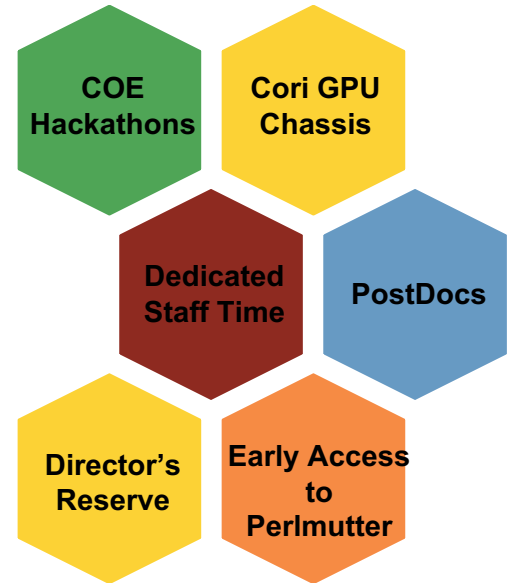
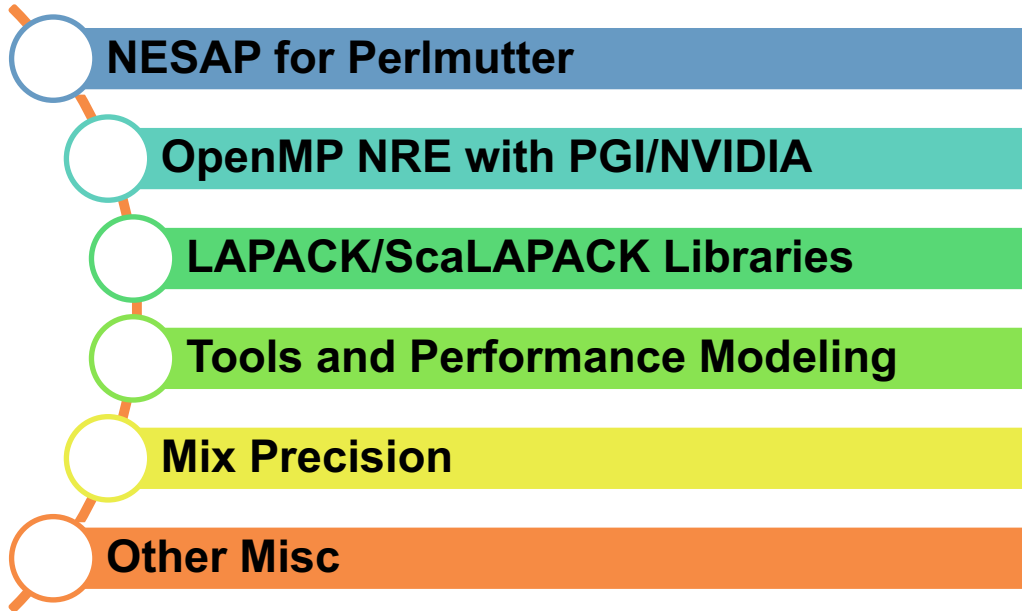


Our Grand Challenge

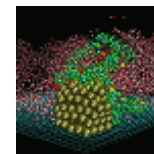
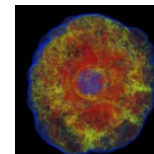
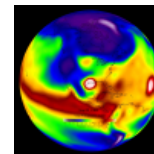
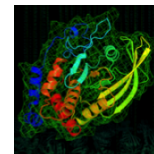
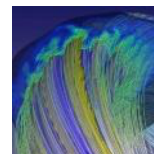
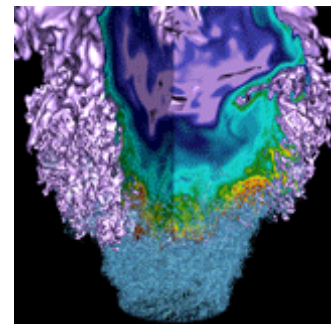


Enable a diverse community of ~7000 users and ~800 codes to run efficiently on advanced architectures such as Cori, **Perlmutter** and beyond

Our Solutions to it



NESAP for Perlmutter



NESAP for Perlmutter

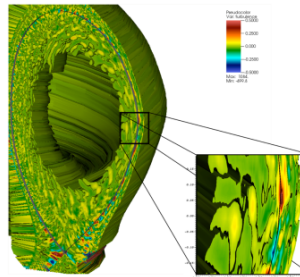
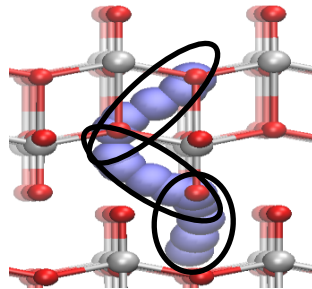
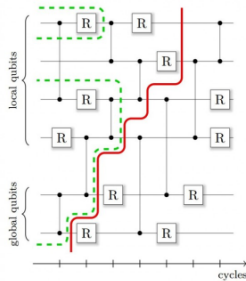
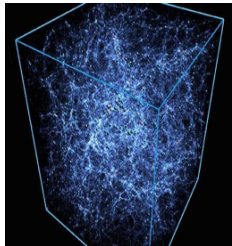
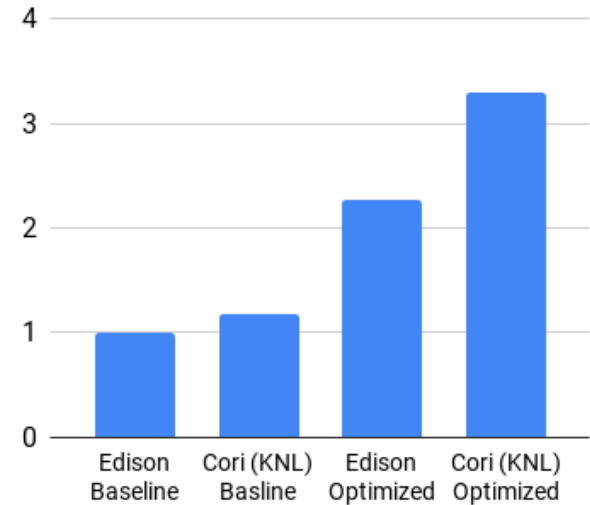


NESAP is NERSC's Application Readiness Program. Initiated with Cori; Continuing with Perlmutter.

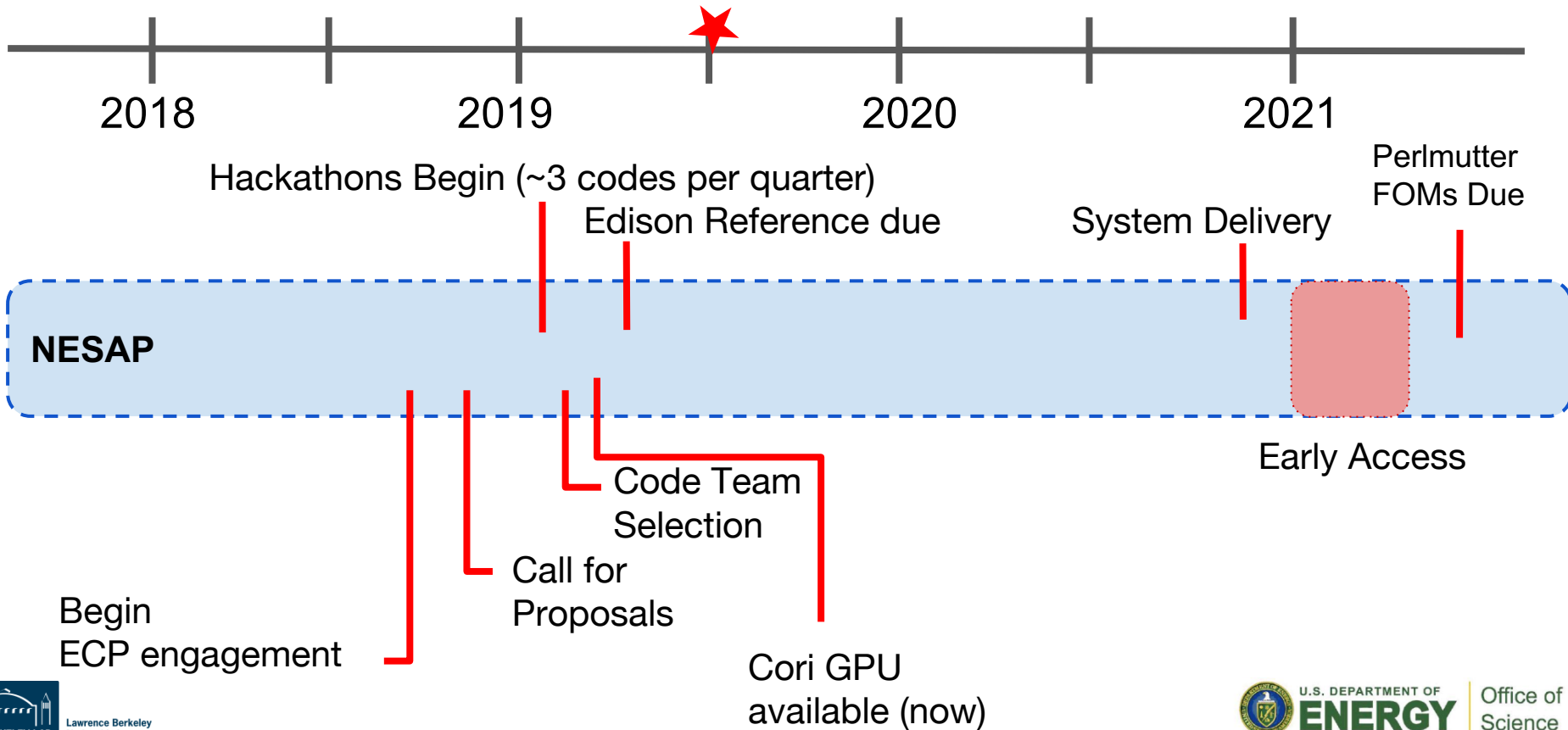
Strategy: Partner with app teams and vendors to optimize participating apps. Share lessons learned with with NERSC community via documentation and training.

We are really excited about working with you to accelerate science discovery on Perlmutter!

NESAP For Cori Speedups



NESAP Timeline



Application Selection

Simulation
~12 Apps

Data Analysis
~8 Apps

Learning
~5 Apps

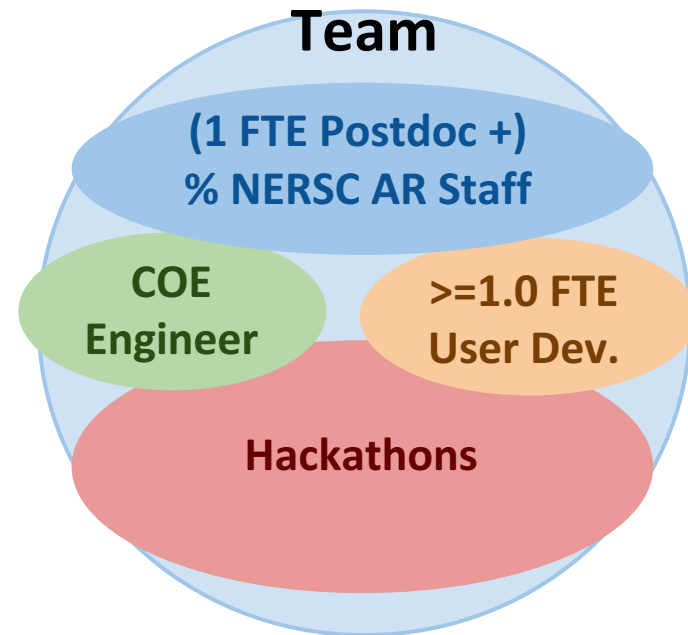
- 6 NESAP for Data apps continued
- 5 ECP Apps jointly selected (Participation funded by ECP)
- Open call for proposals
- Reviewed by a committee of NERSC staff, external reviewers and input from DOE PMs
- Multiple applications from each SC Office and algorithm area
- Beyond this 25 **Tier-1 apps**, additional applications selected for **Tier-2 NESAP**

Support for NESAP Teams



Benefit	Tier 1	Tier 2
Early Access to Perlmutter	yes	eligible
Hack-a-thon with vendors	yes	eligible
Training resources	yes	yes
Additional NERSC hours from Director's Reserve	yes	eligible
NERSC funded postdoctoral fellow	eligible	no
Commitment of NERSC staff assistance	yes	no

Target Application Team



Hack-a-Thons



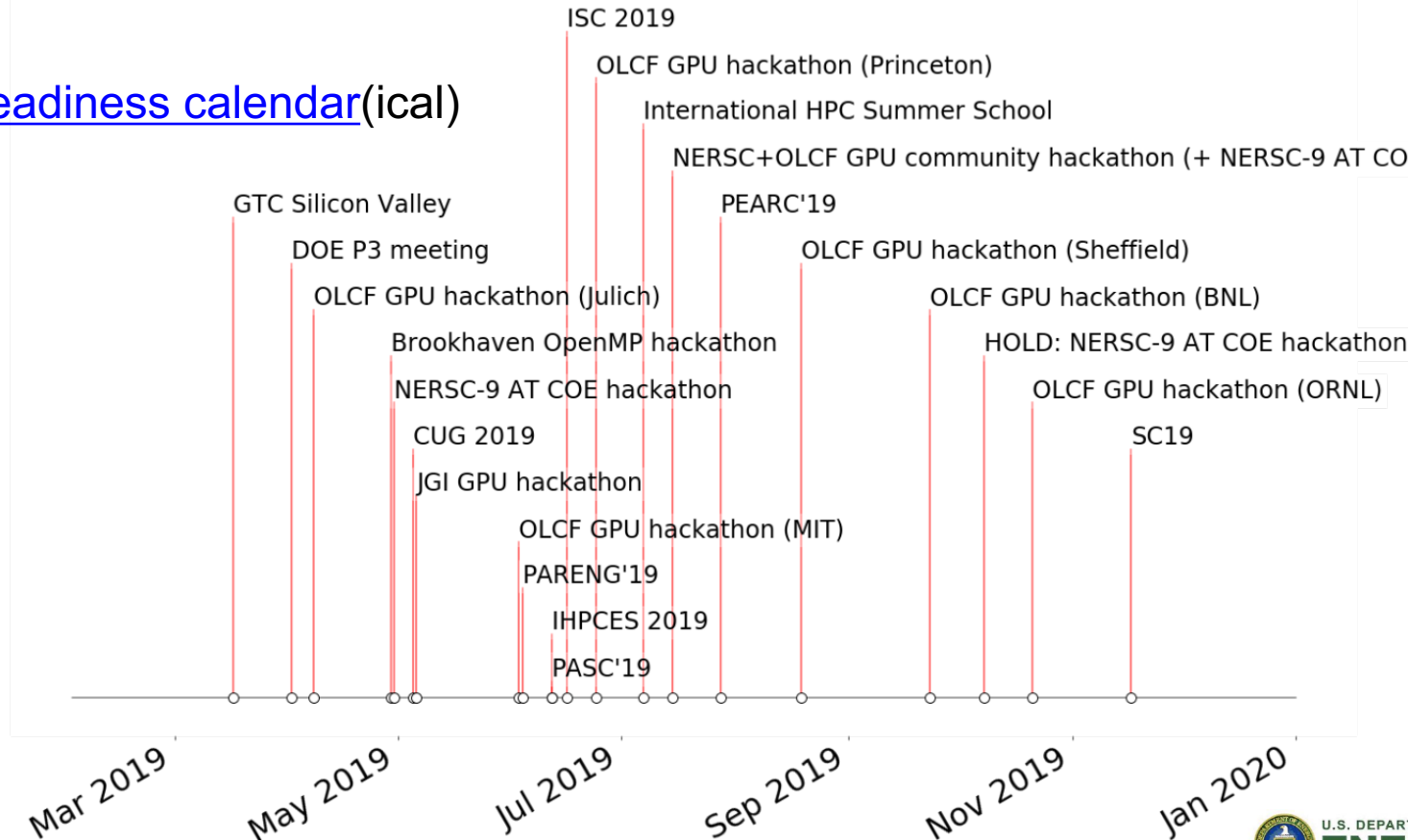
- Quarterly GPU hackathons from 2019-2021
- ~3 apps per hackathon
- 6-week prep with performance engineers, leading up to 1 week of hackathon
- Deep dives with experts from Cray, NVIDIA, NERSC
- Tutorials throughout the week on different topics
 - OpenMP/OpenACC, Kokkos, CUDA etc.
 - profiler techniques/advanced tips
 - GPU hardware characteristics, best known practices



Other Events



[App Readiness calendar\(ical\)](#)



NESAP Postdocs



NERSC plans to hire a steady-state of between 10-15 PostDocs to work with NESAP teams towards Perlmutter readiness.

Positions are non-traditional from most academic PostDocs. Project is mission driven (to optimize applications for Perlmutter).

Projects with a mix of Science, Algorithms and Computer Science are often most compelling/successful. **Need to be well connected w/ team.**

PostDocs sit at NERSC and collaborate closely with other NESAP staff but available to regularly travel to team location.

Previous NESAP Postdocs



Mathieu Lobet (WARP)
La Maison de la Simulation (CEA)
(Career)



Zahra Ronaghi (Tomopy)
NVIDIA (Career)



Brian Friesen (Boxlib/AMReX)
NERSC (Career)



Rahul Gayatri (Perf. Port.)
ECP/NERSC (Term)



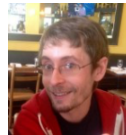
Tareq Malas (EMGEO)
Intel (Career)



Tuomas Koskela (XGC1)
Helsinki (Term)



Andre Ovsyanikov (Chombo)
Intel (Career)



Bill Arndt (E3SM)
NERSC (Career)

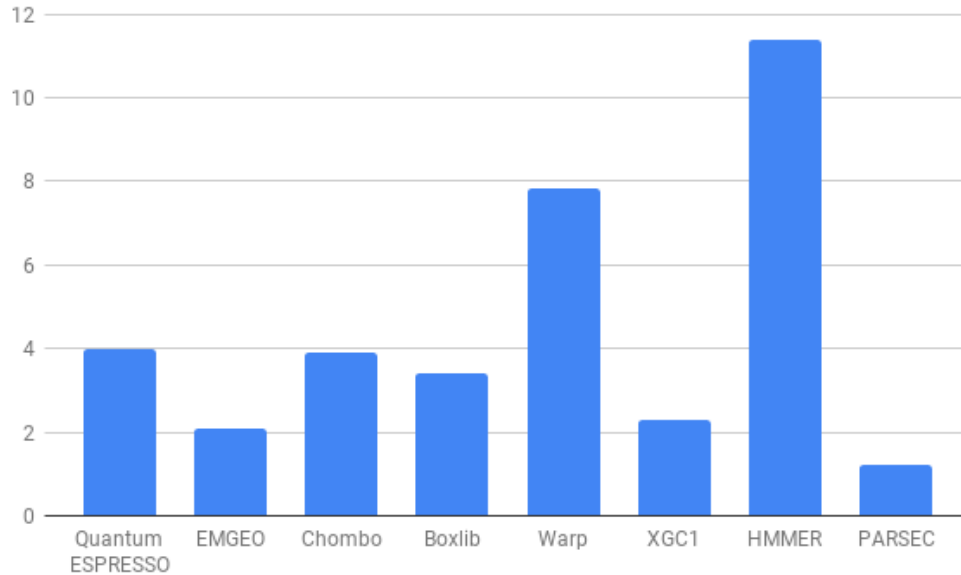


Taylor Barnes (Quantum
ESPRESSO)
MOLSSI (Career)



Kevin Gott (PARSEC)
ECP/NERSC (Term)

Postdoc Speedups for Cori



PostDocs made average of 4.5X SpeedUp in NESAP for Cori

Published 20+ Papers Along with NESAP Teams and Staff

We Need Your Help!



The best way to guarantee your project a PostDoc is to help us recruit one!

Encourage bright, qualified and eligible (must have less than 3 years existing PostDoc experience) candidates to apply (and email Jack Deslippe - jrdeslippe@lbl.gov)

We are interested in advertising in your domain mailing lists.

NESAP PostDoc Position:

<http://m.rfer.us/LBLRJs1a1>



NERSC Liaisons



NERSC has steadily built up a team of Application Performance experts who are excited to work with you.



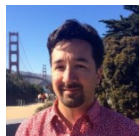
Jack Deslippe
Apps Performance Lead
NESAP LEAD



Brandon Cook
Simulation Area
Lead



Thorsten Kurth
Learning Area
Lead



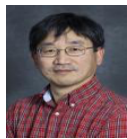
Rollin Thomas
Data Area
Lead



Brian Friesen
Cray/NVIDIA COE
Coordinator



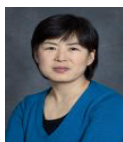
Charlene
Yang
Tools/Libraries
Lead



Woo-Sun
Yang



Doug
Doerfler



Zhengji
Zhao



Helen He



Stephen
Leak



Kevin Gott



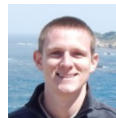
Lisa
Gerhardt



Jonathan
Madsen



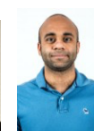
Rahul
Gayatri



Chris
Daley



Wahid
Bhimji



Mustafa
Mustafa



Steve
Farrell



Mario
Melara

What we can and can't help with:

Can:

- Help Facilitate Between Team and Vendors/NERSC
- Help Profile, Analyze Performance and Guide Optimization
- Get hands on with code, suggest patches for well contained regions
- Help guide PostDocs' progress and provide career advice

Can't (in most cases):

- Become Domain Experts in Your Field
- Redesign an application/algorithm from scratch
- Rewrite/Refactor large sections of your application
- Be the only point-of-contact a NESAP PostDoc has with team

Cori GPU Access



- 18 nodes in total, each node has:
 - 2 sockets of 20-core Intel Xeon Skylake processor
 - 384 GB DDR4 memory
 - 930 GB on-node NVMe storage
 - 8 NVIDIA V100 Volta GPUs with 16 GB HBM2 memory
 - Connected with NVLink interconnect
- CUDA, OpenMP, OpenACC support
- MPI support
- Access for NESAP Teams by request
 - Request form link will be sent to NESAP mailing list

Training, Case Studies and Documentation

- For those teams **NOT** in NESAP, there will be a robust training program
- Lessons learned from deep dives from NESAP teams will be shared through case studies and documentation



The screenshot shows the NERSC website with the following elements:

- NERSC Logo:** Powering Scientific Discovery Since 1974
- Navigation:** HOME, ABOUT, SCIENCE AT NERSC, SYSTEMS, FOR USERS (selected), NEWS & PUBLICATIONS, R & D, EVENTS, LIVE STATUS, TIMELINE
- Search Bar:** search... with a search icon and a share icon.
- FOR USERS Sidebar:**
 - Live Status
 - User Announcements
 - My NERSC
 - Getting Started
 - Connecting to NERSC
 - Accounts & Allocations
 - Computational Systems
 - Cori
 - Updates and Status
 - Cori Timeline
 - Configuration
 - Getting Started
 - Programming
 - Running Jobs
 - Burst Buffer
 - Cori Intel Xeon Phi Nodes
 - Application Porting and Performance
 - Getting Started and Optimization Strategy
 - Application Case Studies** (highlighted)
 - EMGEO Case Study
 - BerkeleyGW Case Study
 - QPhIX Case Study
 - WARP Case Study
 - MFDn Case Study
 - BoxLib Case Study
 - VASP Case Study
 - CESM Case Study
 - Chombo-Crunch Case Study
 - HIMMER3 Case Study
 - Early application case studies
 - ISCL16 IXPUG Performance Workshop
 - Quantum ESPRESSO Exact Exchange Case Study
 - XGCI Case Study
 - Profiling Your Application

- Main Content Area:**
- Home » For Users » Computational Systems » Cori » Application Porting and Performance » Application Case Studies
- ## APPLICATION CASE STUDIES
- NERSC staff along with engineers have worked with NESAP applications to prepare for the Cori-Phase 2 system based on the Xeon Phi "Knights Landing" processor. We document the several optimization case studies below.
- Our presentations at ISC 16 IXPUG Workshop can all be found: <https://www.ixpug.org/events/ixpug-isc-2016>
- Other pages of interest for those wishing to learn optimization strategies of Cori Phase 2 (Knights Landing):
 - [Getting Started](#)
 - [Measuring Arithmetic Intensity](#)
 - [Measuring and Understanding Memory Bandwidth](#)
 - [Vectorization](#)
- ### EMGEO Case Study »

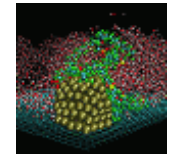
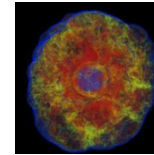
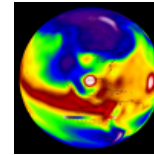
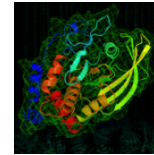
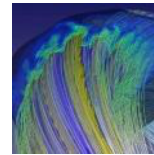
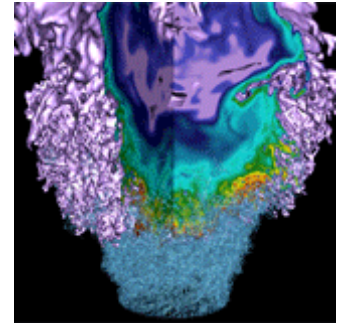
June 20, 2016
Early experiences working with the EMGeo geophysical imaging applications. [Read More »](#)
- ### BerkeleyGW Case Study »

Code Description and Science Problem BerkeleyGW is a Materials Science application for calculating the excited state properties of materials such as band gaps, band structures, absorption spectroscopy, photoemission spectroscopy and more. It requires as input the Kohn-Sham orbitals and energies from a DFT code like Quantum ESPRESSO, PARATEC, PARSEC etc. Like such DFT codes, it is heavily dependent on FFTs, Dense Linear algebra and tensor contraction type operations similar in nature to those... [Read More »](#)
- ### QPhIX Case Study »

June 20, 2016
Background QPhIX [1,2,3] is a library optimized for Intel(R) manycore architectures and provides sparse solvers and slash kernels for Lattice QCD calculations. It supports the Wilson dslash operator with and without clover term as well as Conjugate Gradient [4] and BiCGStab [5] solvers. The main task for QPhIX is to solve the sparse linear system where the Dslash kernel is defined by Here, U are complex, special unitary, 3x3 matrices (the so-called gauge links) which depend on lattice site x... [Read More »](#)
- ### WARP Case Study »

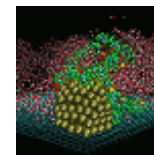
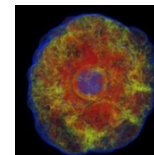
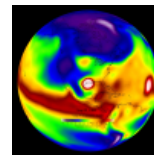
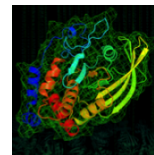
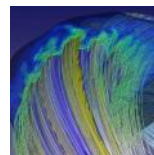
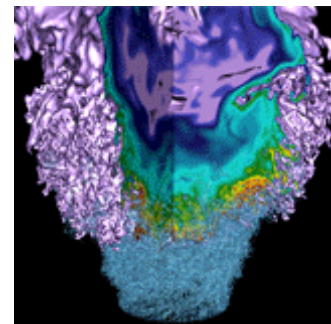
Update A more complete summary is now available at <https://picsar.net/> Background WARP is an accelerator code that is used

OpenMP NRE



- Add OpenMP GPU-offload support to PGI C, C++, Fortran compilers
 - Performance-focused subset of OpenMP-5.0 for GPUs
 - Compiler will be optimized for NESAP applications
- Early and continual collaboration will help us improve the compiler for you. Please
 - Strongly consider using OpenMP GPU-offload in your NESAP applications
 - Let us help you to use OpenMP GPU-offload
 - Share representative mini-apps and kernels with us
 - Experiment with the GPU-enabled OpenMP compiler stacks on Cori-GPU (LLVM/Clang, Cray, GNU)
 - Contact Chris Daley (csdaley@lbl.gov) and/or your NESAP project POC

(Sca)LAPACK Libraries



Lack of (Sca)LAPACK on GPUs



	Library	Support for NVIDIA GPUs
	cuSolver	Incomplete LAPACK (cuSolverDN, cuSolverSP, cuSolverRF)
	MAGMA	Incomplete LAPACK
Single GPU	Cray LibSci_ACC	Incomplete LAPACK and not promised/planned for Perlmutter
	PETSc	Certain subclasses ported using Thrust and CUSP
	Trilinos	Certain packages implemented using Kokkos
	SLATE	Ongoing ECP work, due to finish in 2021
Multiple GPUs (Distributed)	ELPA	Only support eigensolvers
	???	???

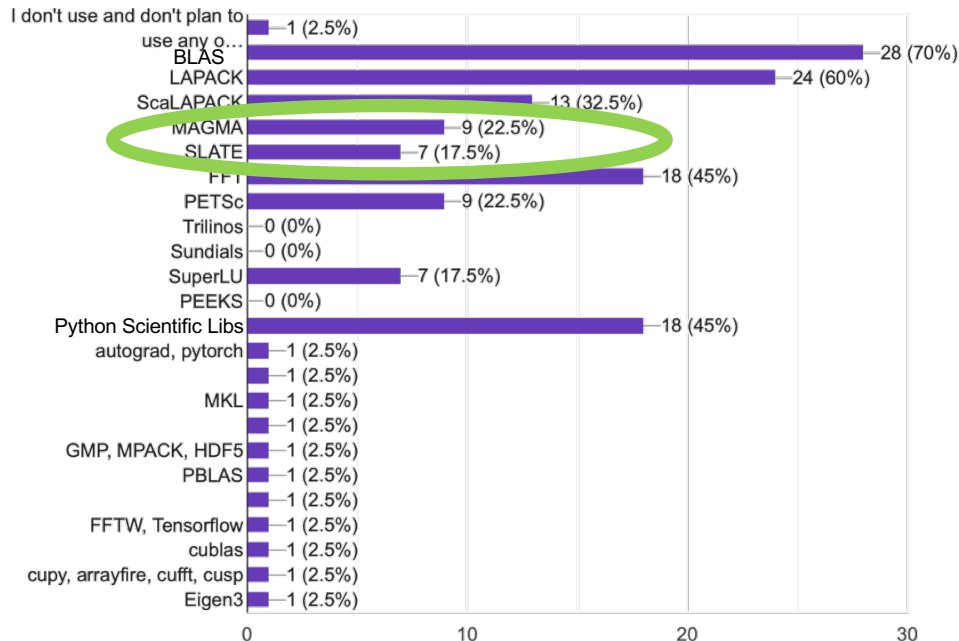
NESAP Survey



- April 10-30; 40 responses
- What libraries do you use?
- What routines in LAPACK?
- What routines in ScaLAPACK?
(% of runtime, matrix size, sdcz)
- More details at [Results](#)

What math libraries do you use or plan to use?

40 responses



SciLAPACK required by

- VASP
- Quantum Espresso
- NAMD
- CP2K
- BerkeleyGW
- NWChemEx
- WEST
- Qbox
- DFT-FE
- ExaSGD
- PARSEC
- M3DC1
- MFDn
- WDMApp

Even more for LAPACK...

$$\epsilon_{\mathbf{G}\mathbf{G}'}(\mathbf{q}; 0) = \delta_{\mathbf{G}\mathbf{G}'} - v(\mathbf{q}+\mathbf{G})\chi_{\mathbf{G}\mathbf{G}'}(\mathbf{q}; 0)$$

$$W_{\mathbf{G}\mathbf{G}'}(\mathbf{q}; 0) = \epsilon_{\mathbf{G}\mathbf{G}'}^{-1}(\mathbf{q}; 0)v(\mathbf{q}+\mathbf{G}')$$

$$(E_{\mathbf{c}\mathbf{k}}^{\text{QP}} - E_{\mathbf{v}\mathbf{k}}^{\text{QP}})A_{\mathbf{v}\mathbf{c}\mathbf{k}}^S + \sum_{\mathbf{v}'\mathbf{c}'\mathbf{k}'} \langle \mathbf{v}\mathbf{c}\mathbf{k} | K^{\text{eh}} | \mathbf{v}'\mathbf{c}'\mathbf{k}' \rangle = \Omega^S A_{\mathbf{v}\mathbf{c}\mathbf{k}}^S$$

Diagonalization and inversion of large matrices, e.g. 200k x 200k

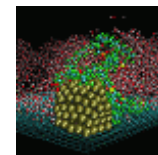
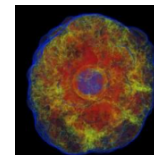
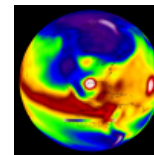
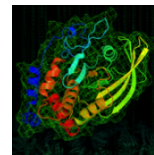
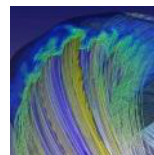
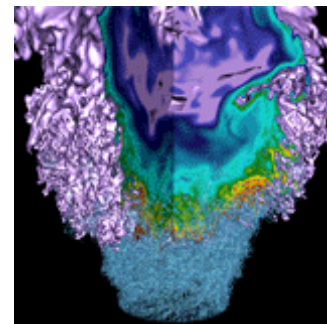


Collaboration with PGI/NVIDIA



- **A drop-in replacement for LAPACK/ScaLAPACK**
- **Support distributed memory systems with NVIDIA GPUs**
- **Possibly leverage SLATE and ELPA efforts**

Tools and Performance Models



Tools and Roofline



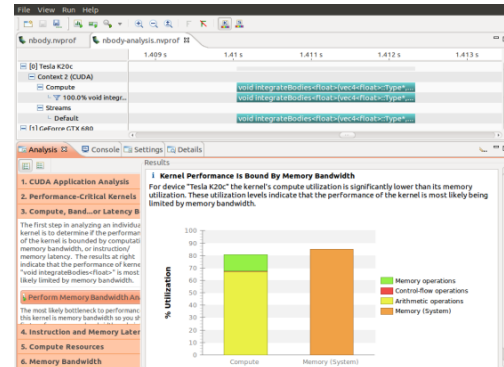
Profiling tools

- provide a rich set of features
- nvprof/nvvp, Nsight Systems, Nsight Compute
- TAU, HPC Toolkit

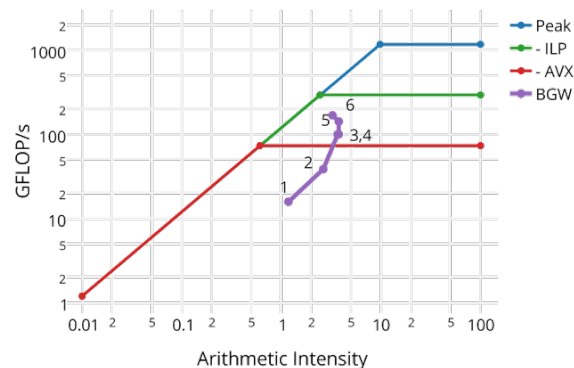
Roofline Performance Model:

- offers a holistic view of the application
- captures effects of bandwidth/latency, memory coalescing, instruction mix, thread divergence, *etc*

We are actively working with NVIDIA towards GPU Roofline analysis using nvprof/Nsight Compute.



Haswell Roofline Optimization Path



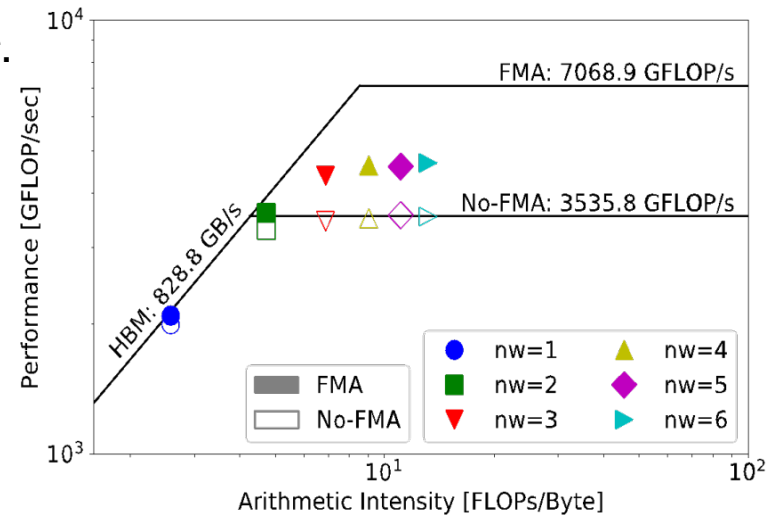
Roofline on GPUs

So far, we've been able to construct a hierarchical Roofline on NVIDIA GPUs

- nvprof metrics for runtime, FLOPs, and bytes
- memory hierarchy: L1/shared, L2, DRAM, etc.

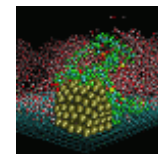
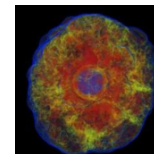
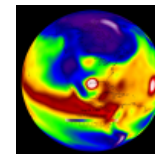
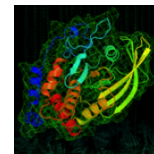
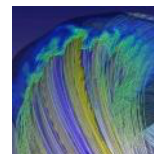
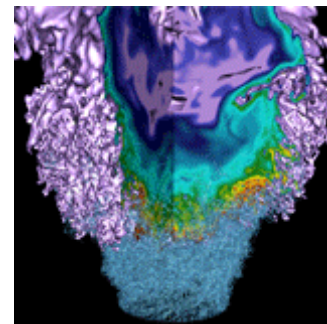
WorkFlow:

1. Use nvprof to collect application data (FLOPs, bytes, runtime)
2. Calculate Arithmetic Intensity (FLOPs/byte) and application performance (GFLOP/s)
3. Plot Roofline



GPP on V100

Mixed Precision



Benefits of reduced/mixed precision:

- From FP64 to FP32
 - 2x speedup due to bandwidth savings or compute unit availability
 - similar savings in network communication
- More modern architectures support efficient FP16 operations
 - speedup of about 15x possible compared to FP64 for certain operations
- Similar speedups are possible if most operations are done in lower precision

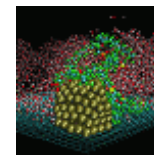
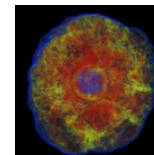
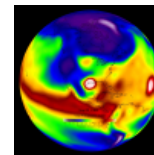
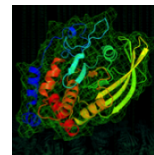
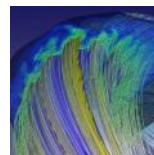
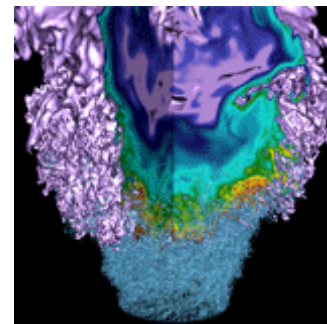
NESAP collaboration with CRD (Costin Iancu) and NVIDIA (Chris Newburn):

- Investigate the applicability of mixed precision arithmetic
- Extract general guidelines and rules of when it works when it doesn't
- Apply findings to some NESAP applications to improve performance

How can I get involved?

- Follow opportunities to follow on the NESAP mailing list

Other Work



Performance Portability

The screenshot shows the website's navigation structure. On the left, there are dropdown menus for 'Performance Portability', 'Measurements', and 'Measurement Techniques'. The main content area features a table of contents with sections like 'Measuring Portability' and 'Measuring Performance'. A search bar is located at the top right of the page.

OpenACC

Directives for Accelerators

NERSC now a member.



NERSC leading development of performanceportability.org



NERSC hosted 2016 C++ Summit and ISO C++ meeting on HPC.

NERSC leading 2019 DOE COE Perf. Port. Meeting





Thank You