Do The Right Thing!

Pursuing long-term sustainability (sanity) goals over immediate profits

Hai Ah Nam
Computational Physics & Methods (CCS-2)
LANL Center of Excellence

Presented to:
NERSC GPUs for Science Day
Berkeley, CA

July 2, 2019
One-Off like Roadrunner?

Not as bad as we thought
The ties that bind us

Pre-Exascale Landscape (2016-2019)

Many-core & Accelerators are a reality instead of a one-off

Los Alamos National Laboratory

7/2/2019 | 3
Pre-exascale (2016-2019) ~ 100s PF
The ties that bind us

Performance Portability
Portability across platforms: ~Single source code run with acceptable performance across platforms
Addressing the design flaw in the DOE HPC landscape

- Thou shall create HPC systems that “promotes a competition of ideas and technologies” and “promotes a rich and healthy HPC Ecosystem”

- Thou shall provide an HPC system that optimizes for power, space, cooling & performance = better, faster or more efficient than others that came before

- Thou shall be at the forefront of scientific discovery for your domain and pop out an highlight every quarter = productivity

- Oh, and do it all using cutting edge technology that is largely research itself
The Department of Energy (DOE) Performance, Portability and Productivity Annual Meeting is an opportunity to share ideas, progress, and challenges toward the goal of performance portability across DOE's current and future advanced supercomputers. The need for applications to run effectively on multiple vendor architecture solutions is pervasive across application teams within the DOE.

The two primary goals of this meeting are to:

- Inform application teams and tool developers of activities and methodologies...
- ... work with the vendors and tool providers on determining implementations and solutions that will meet their own performance criteria without inadvertently impairing performance results elsewhere.
‘Supported’ PPP Approaches

• MPI + 🚃 = still viable + has community support
  – Directives
    • OpenMP | OpenMP target
      – NERSC: NRE contract with NVIDIA to enhance the NVIDIA’s PGI C, C++ and Fortran compilers to enable OpenMP applications to run on NVIDIA GPUs.
    • OpenACC (limited vendor support, PGI)
  – Abstractions
    • Kokkos CPU | Kokkos GPU
      – Fortran Interoperability (Womeldorff & Gaspar, LANL)
    • RAJA CPU | RAJA GPU

• Good if you can get them to work for you
  – Libraries: portable if available and you can isolate computational intensity
  – DSLs: portable if available and your community supports it

• On the horizon?
  – OpenCL Framework (resurgence): Open Computing Language
  – SYCL: C++ Single-source Heterogeneous Programming for OpenCL
Evolving, focusing and broadening engagement

**2016**
- Performance Portability Definition (agree to disagree)
- Productivity matters
- Memory Hierarchy was a BIG concern (MCDRAM & UVM)
- Larger solution space of PP approaches
- Few applications, some libraries, many proxy explorations
- Pushing needs into procurement

**2017**
- Performance Portability Metrics (agree to disagree)
- Memory hierarchy is a moderate concern
- Solution space for C++ codes have narrowed (limited resources)
- Increasing # of real application experiences
- Standards committee (C++, OpenMP) involvement needed

**2019**
- +Productivity -COE in meeting target
- **PP&P metrics** (agree to disagree)
- Memory hierarchy not so bad
- Focus on Fortran ([Wolfe](#))
- Lots of systematic comparisons between approaches across many codes ([McIntosh-Smith](#), CUG paper)
- Pushing HPC needs into language standards

Los Alamos National Laboratory
Pre-exascale to Exascale Landscape (2020 – 2022)

The complexity is increasing

Crossroad
LANL/SNL
??? 2021

El Capitan
LLNL
? 2021

Perlmutter
LBNL
Cray/AMD/NVIDIA
GPU 2020

Aurora
ANL
Intel/Cray
Intel Xeon, Intel Xe 2021

Frontier
ORNL
Cray/AMD/AMD
GPU 2021

What about
???

Los Alamos National Laboratory
What does the future hold?

- **3 Primary System Types in Top500**
  - Commodity (e.g. Intel)
  - Commodity + accelerator (e.g. GPUs) (138 systems in Top500)
    - Harder to port (~months);
    - **Must work** to get performance but there is potential for large performance gains;
    - Backward portability is not guaranteed
  - Special purpose lightweight cores (e.g. IBM BG, Knights Landing, TaihuLight, Arm)
    - Easy to port (~days);
    - **Must work** to get comparable performance (vectorize!) and work harder to get modest performance improvements

---

### 2020

- Portability
  - Compilers, compilers, compilers
- Performance
  - Compilers, compilers, compilers
- Asynchronous Many Task Programming & Runtime Systems
- Speculation & hair-on-fire
Best Practices for Performance, Portability & Productivity

• Choose approaches and open (standard) programming languages that are supported by multiple vendors across multiple platforms with a community of interest: Strength in numbers

• Avoid over optimizing for one platform: Target 2 of the primary system types during optimizations for PPP

• Compiler diversity increasing: Be an early user/beta tester and provide feedback to the vendors

• PPP is highly dependent on compiler focus & maturity (esp. OpenMP)
  – Provide use cases, reproducers, and bugs to vendors
  – Constant feedback and communication with vendors (use your local COE-entity to have a voice)
Cohorts in Crime – PPP Committee

DOE National Laboratories
- Rob Neely (LLNL)
- Ian Karlin (LLNL)
- Hai Ah Nam (LANL)
- Charles Ferenbaugh (LANL)
- Doug Doerfler (LBL)
- Jack Deslippe (LBL)
- Rebecca Harman-Baker (LBL)
- Scott Parker (ANL)
- Nick Romero (ANL)
- Mike Glass (SNL)
- Rob Hoekstra (SNL)
- Phil Roth (ORNL)
- Tjerk Straatsma (ORNL)

Vendor Partners
- Bill Brantley (AMD)
- Chip Freitag (AMD)
- Abdullah (Apo) Kayi (IBM)
- John Levesque (Cray)
- CJ Newburn (Nvidia)
- John Pennycook (Intel)
- Jason Sewall (Intel)
- Arm
- HPE
- Others?
Resources (slides & reports available)

- **DOE COE Performance Portability Meeting 2016**, Glendale AZ
- **DOE COE Performance Portability Meeting 2017**, Denver, CO
  - International Workshop on Performance, Portability and Productivity in HPC (P3HPC) @ SC18
  - PerformancePortability.org (Office of Science site)
- **2019 Performance, Portability and Productivity Annual Meeting**, Denver, CO
  - 2019 ISC – Performance Portability and Productivity: Panel Discussion
  - 2nd International Workshop on Performance, Portability and Productivity in HPC (P3HPC)
    - Submissions due August 26, 2019
- **2020 PPP Annual Meeting**, Kansas City, MO(?) – lots of speculation to be had!!