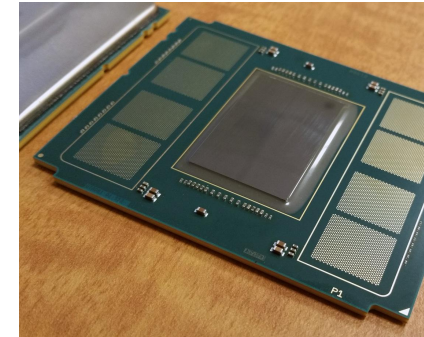
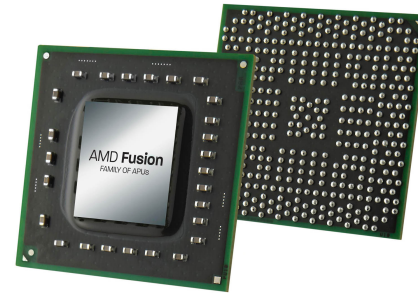
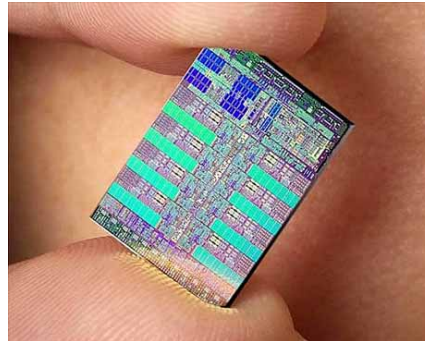


Exceptional service in the national interest

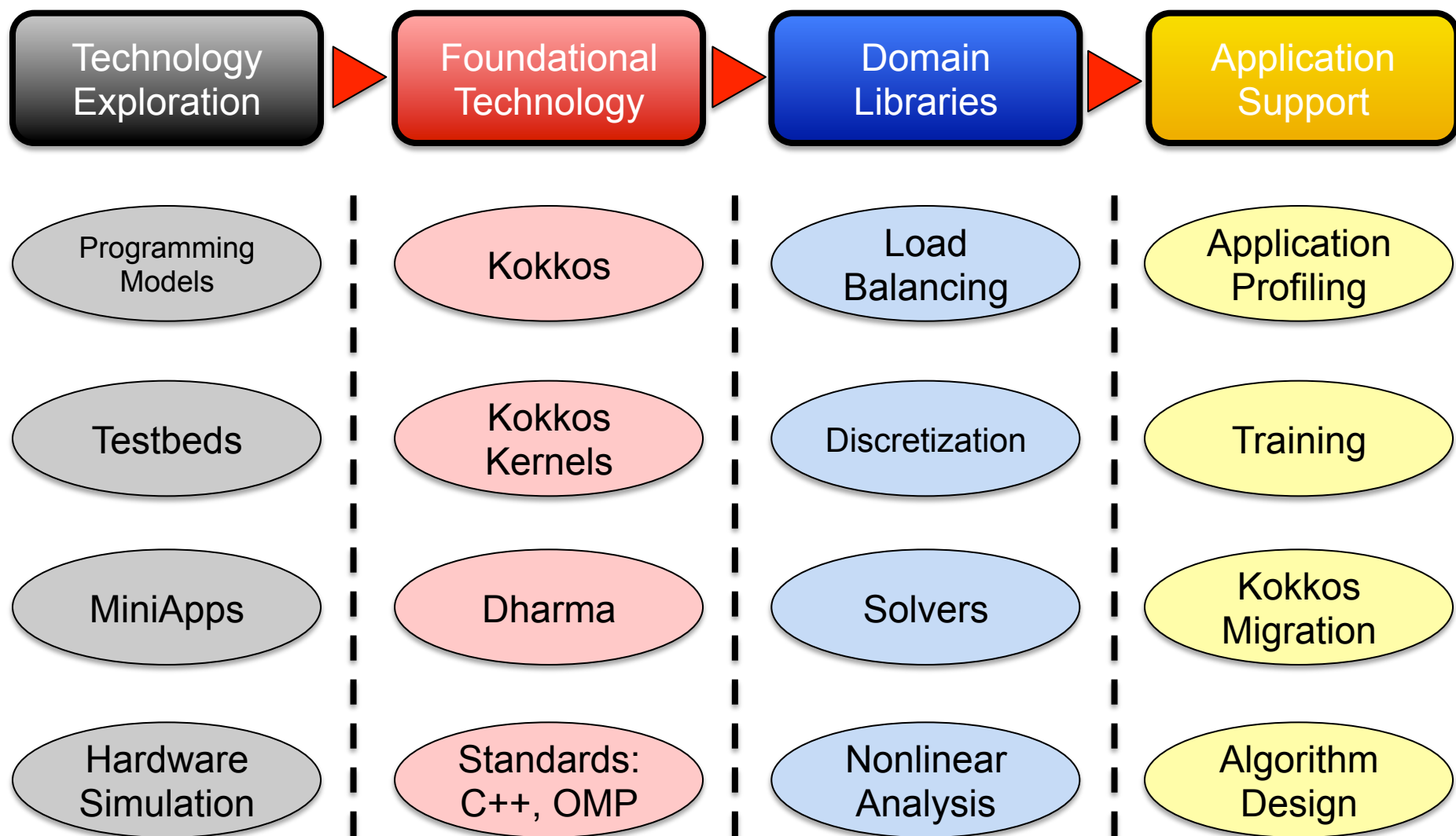


Pathfinding: From Technology Exploration to Application Support

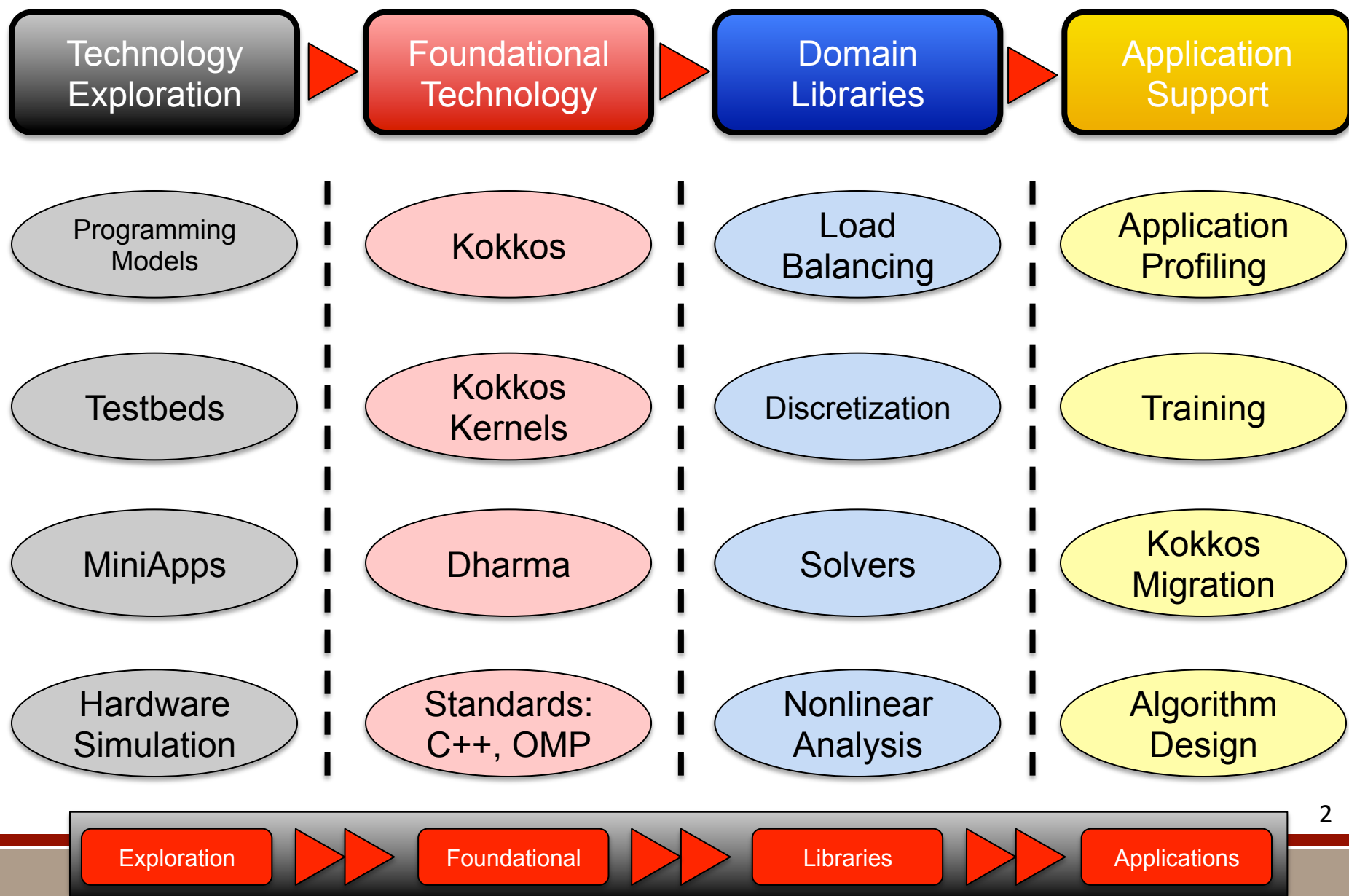
Christian Trott et.al.



Building a Next Generation Software Stack



Building a Next Generation Software Stack

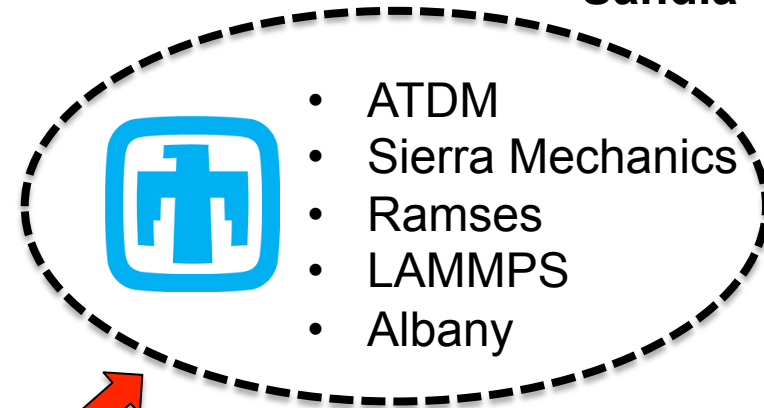


Connecting the Community

Laboratories



Sandia



Application
Support

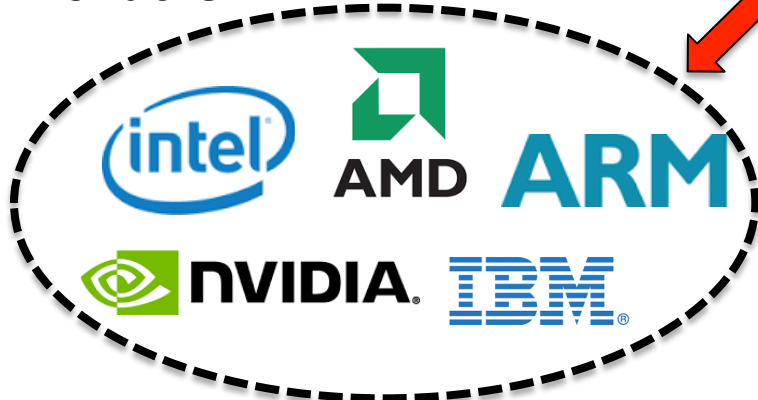
Foundational
Technologies



Solvers

Technology
Exploration

Vendors



Academia



Exploration

Foundational

Libraries

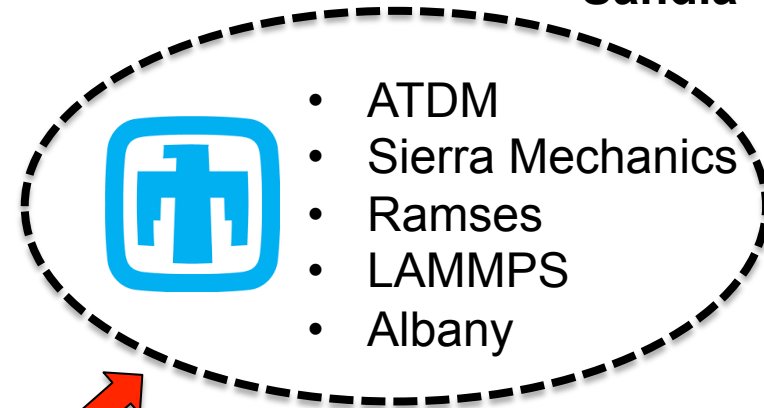
Applications

Connecting the Community

Laboratories



Sandia



Application
Support

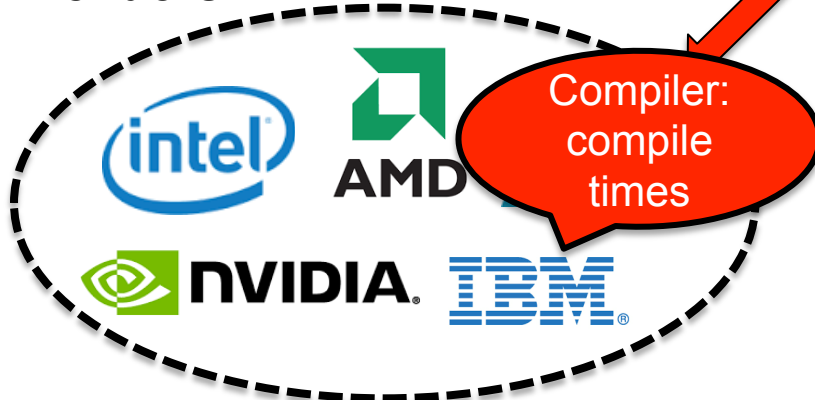
Foundational
Technologies



Solvers

Technology
Exploration

Vendors



Academia



Exploration

Foundational

Libraries

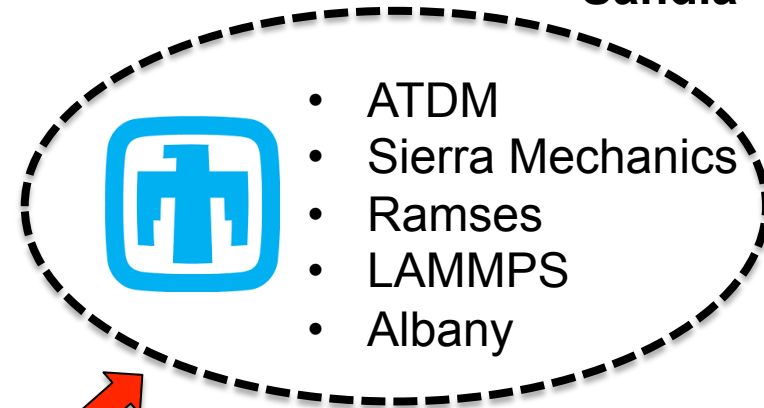
Applications

Connecting the Community

Laboratories



Sandia



Application
Support

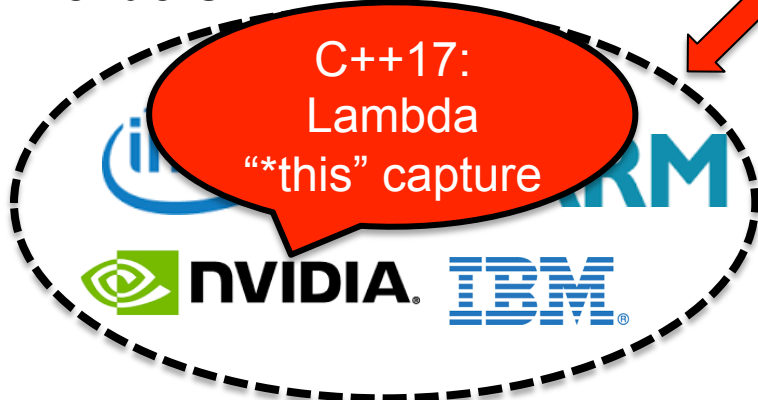
Foundational
Technologies



Solvers

Technology
Exploration

Vendors



Academia



Exploration

Foundational

Libraries

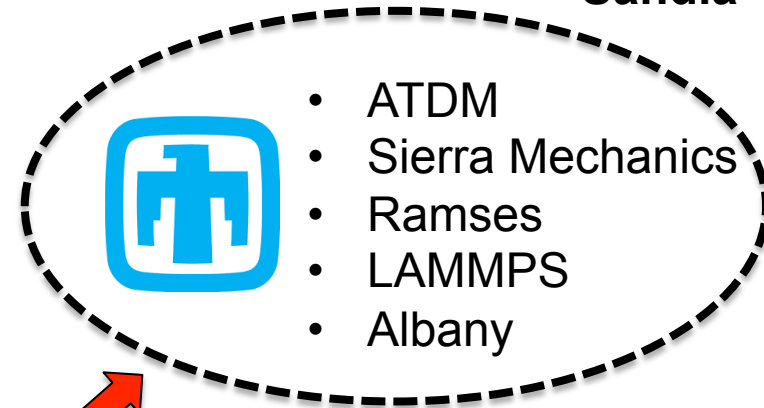
Applications

Connecting the Community

Laboratories



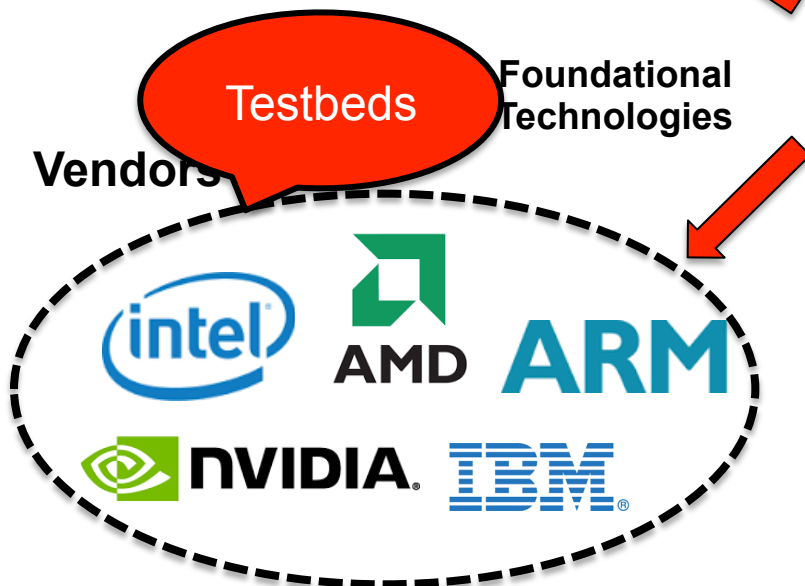
Sandia



Application
Support

Solvers

Vendors



Foundational
Technologies

Technology
Exploration

Academia



Exploration

Foundational

Libraries

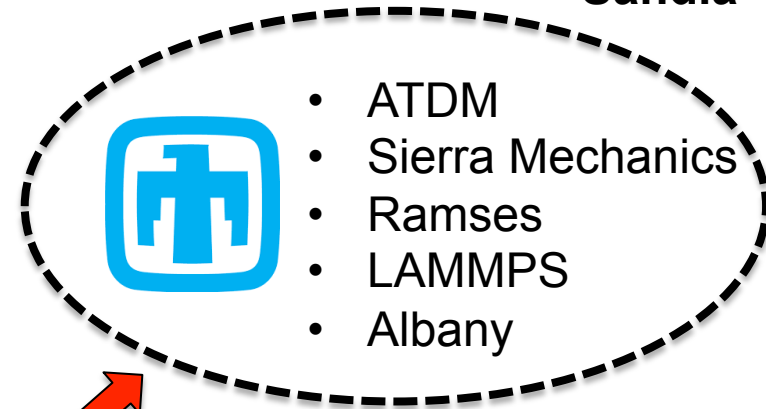
Applications

Connecting to Community

Laboratories

Kokkos +
Legion
Integration

Sandia



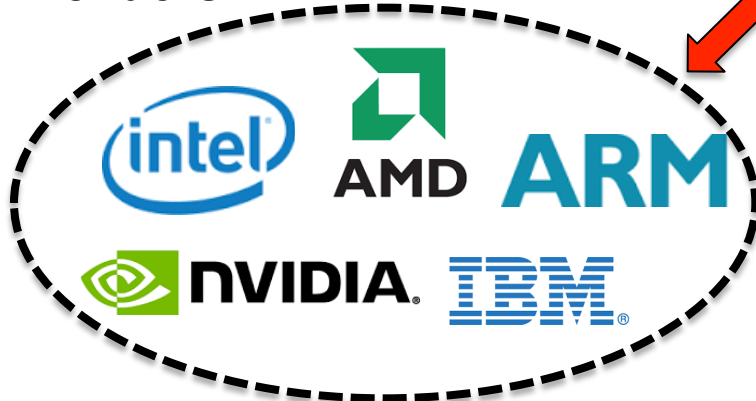
Application
Support

Foundational
Technologies

Solvers

Vendors

Academia



Technology
Exploration

Exploration

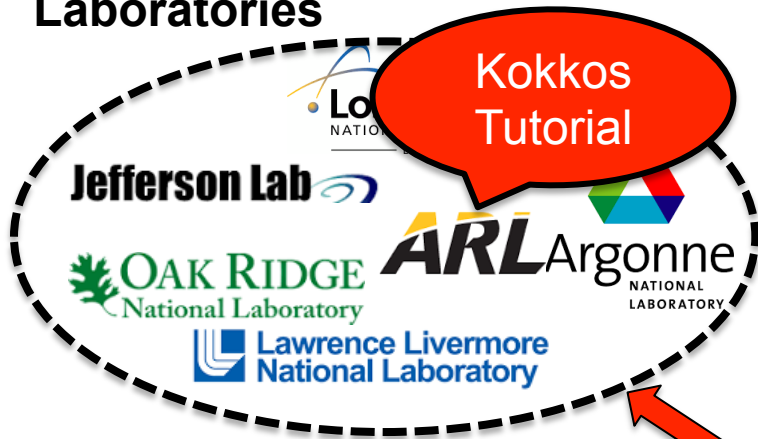
Foundational

Libraries

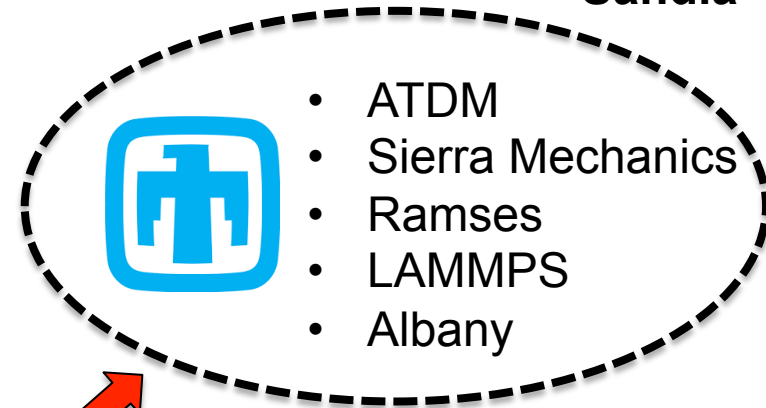
Applications

Connecting the Community

Laboratories



Sandia



Application
Support

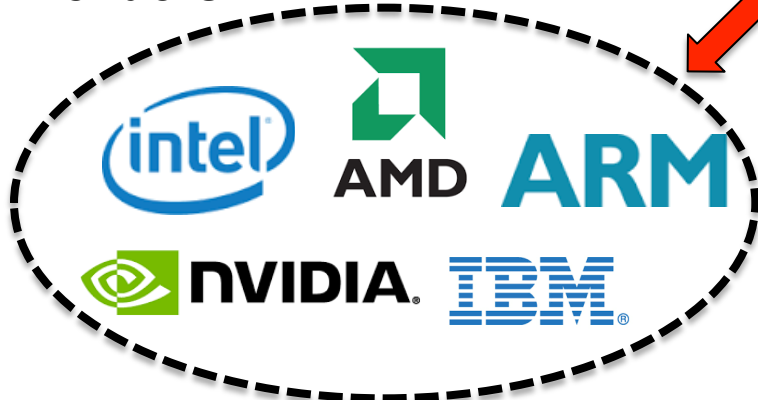
Foundational
Technologies



Solvers

Technology
Exploration

Vendors



Academia



Exploration

Foundational

Libraries

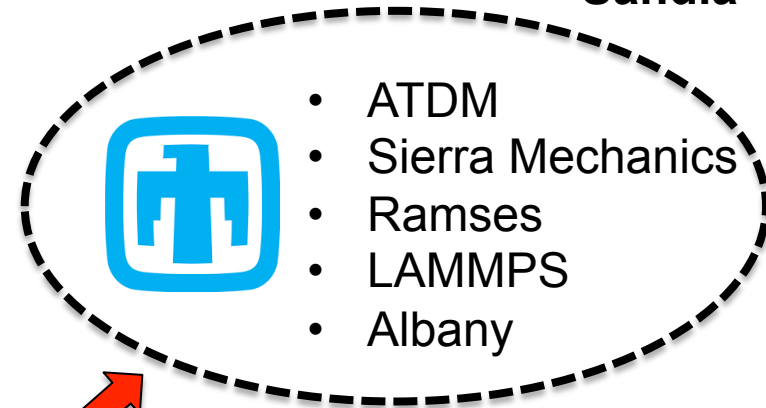
Applications

Connecting the Community

Laboratories



Sandia



Application
Support

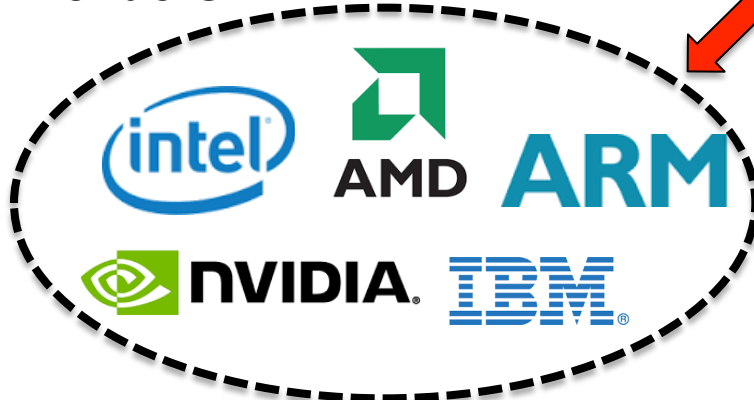
Foundational
Technologies



Solvers

Technology
Exploration

Vendors



Academia



Exploration

Foundational

Libraries

Applications

Connecting the Community

Laboratories

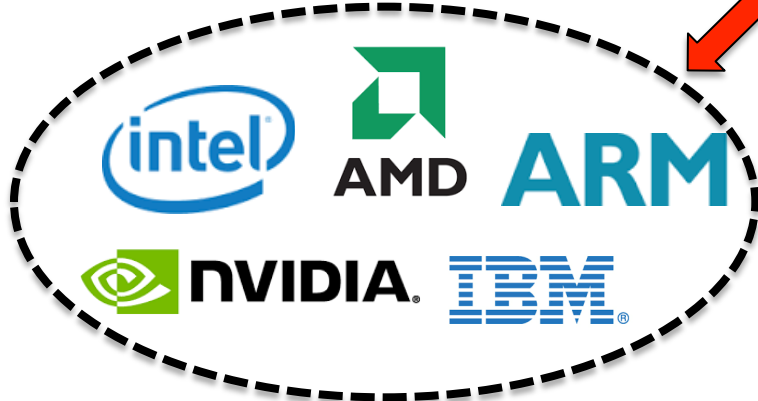


Application
Support



Foundational
Technologies

Vendors



Solvers

Technology
Exploration

Academia



Exploration

Foundational

Libraries

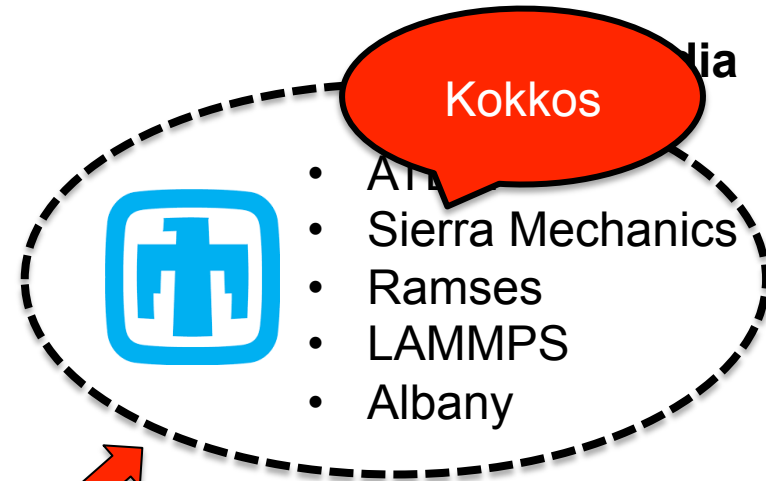
Applications

Connecting the Community

Laboratories

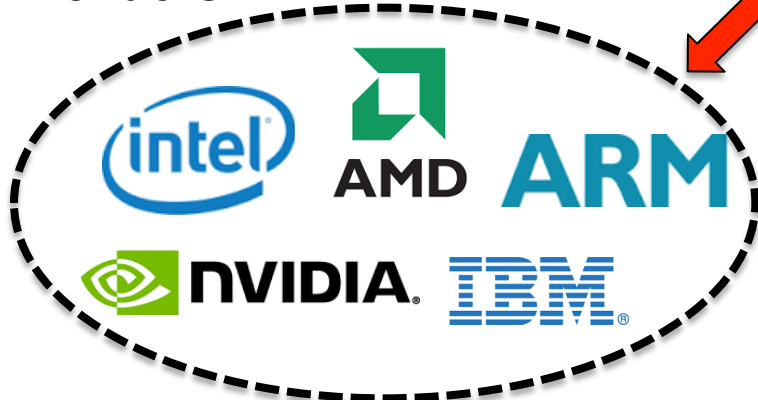


Application
Support



Foundational
Technologies

Vendors



Solvers

Technology
Exploration

Academia



Exploration

Foundational

Libraries

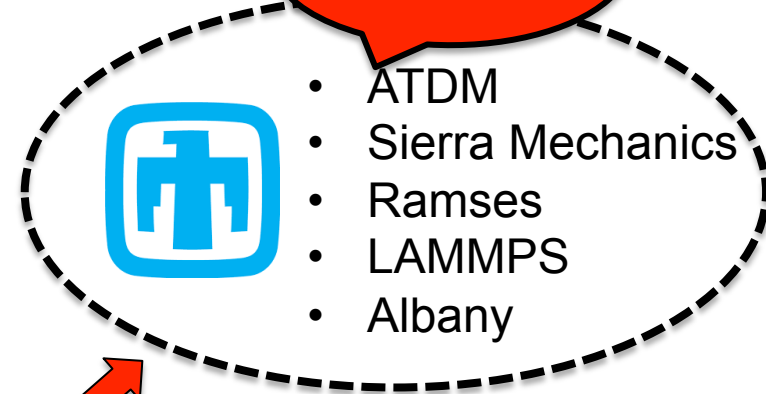
Applications

Connecting the Community

Laboratories



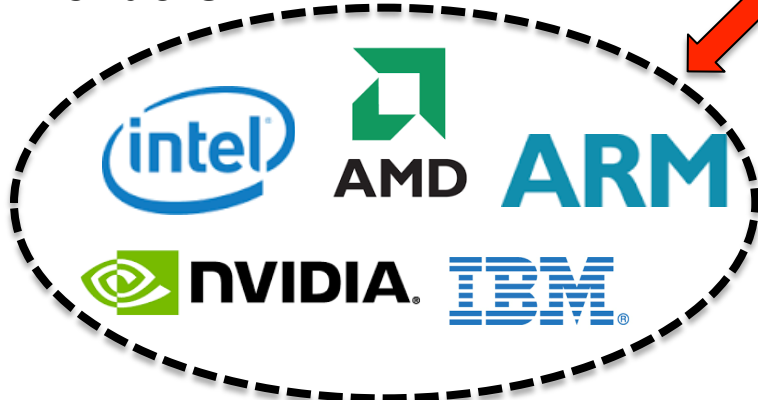
Dharma



Application Support

Foundational Technologies

Vendors



Solvers

Academia



Technology Exploration

Exploration

Foundational

Libraries

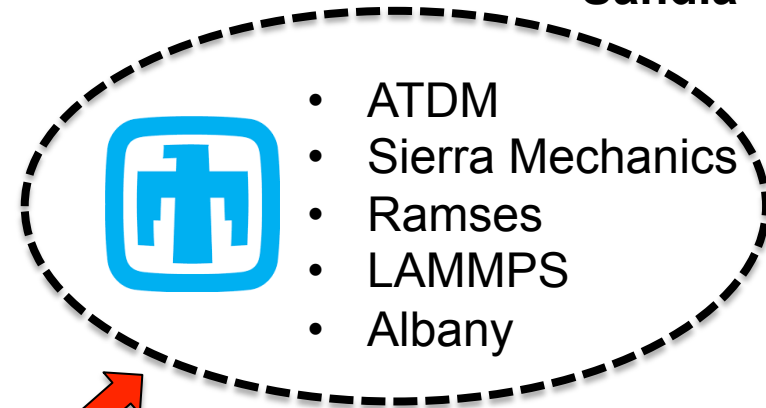
Applications

Connecting the Community

Laboratories



Sandia



Application
Support

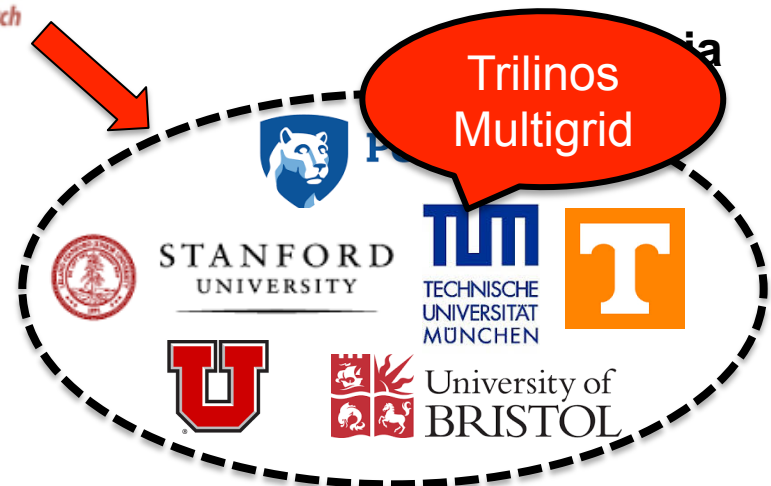
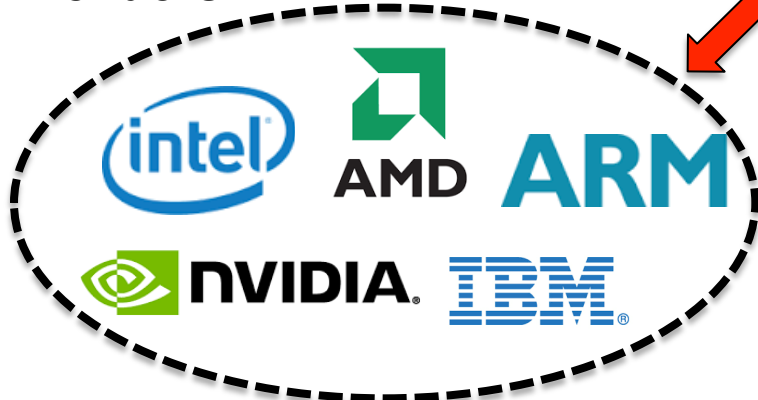
Foundational
Technologies



Solvers

Technology
Exploration

Vendors



Exploration

Foundational

Libraries

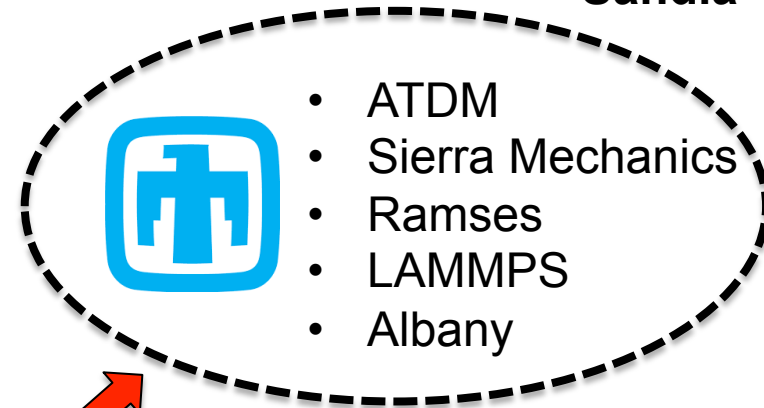
Applications

Connecting the Community

Laboratories



Sandia



Application
Support

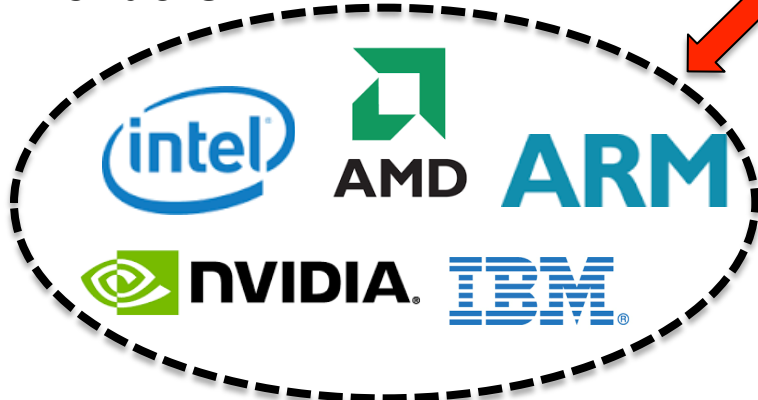
Foundational
Technologies

Solvers



Technology
Exploration

Vendors



Academia



Exploration

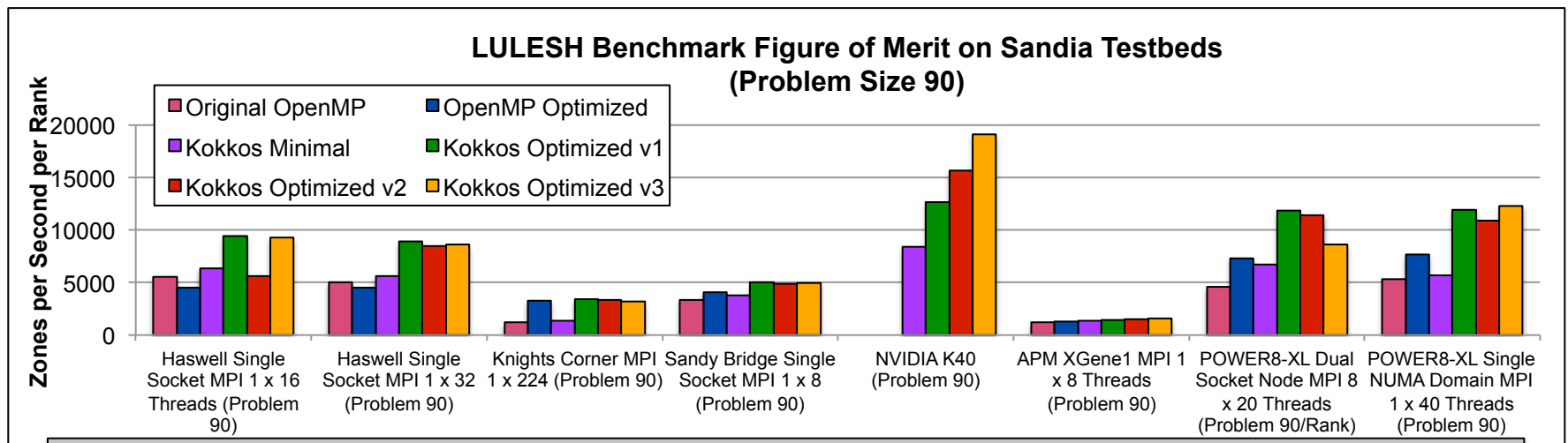
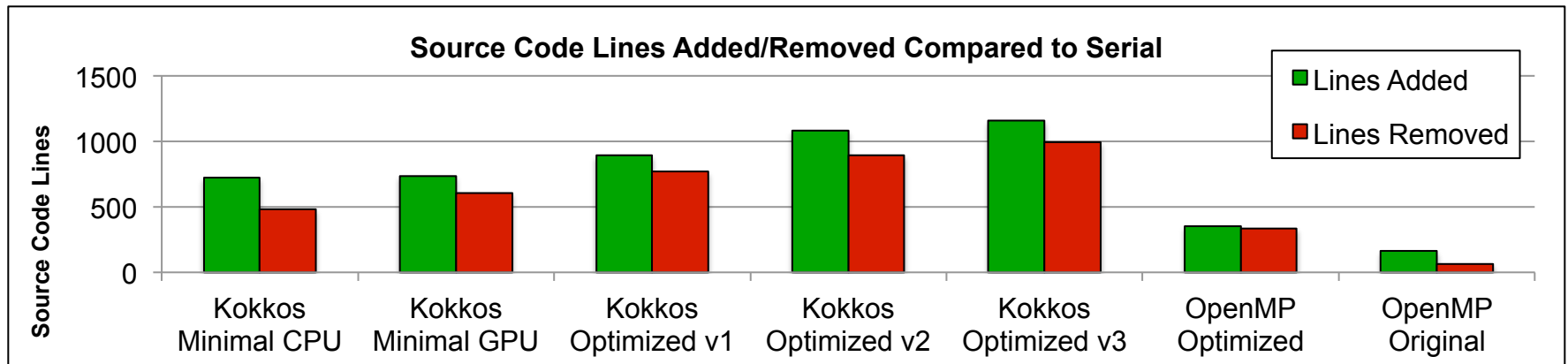
Foundational

Libraries

Applications

Technology Exploration: ASC L2 CoDesign 2015

- Evaluate the performance and productivity tradeoff when using the Kokkos C++ programming model
- Chose LLNL mini-app Lulesh to demonstrate broad applicability
- Report: <http://prod.sandia.gov/techlib/access-control.cgi/2015/157886.pdf>



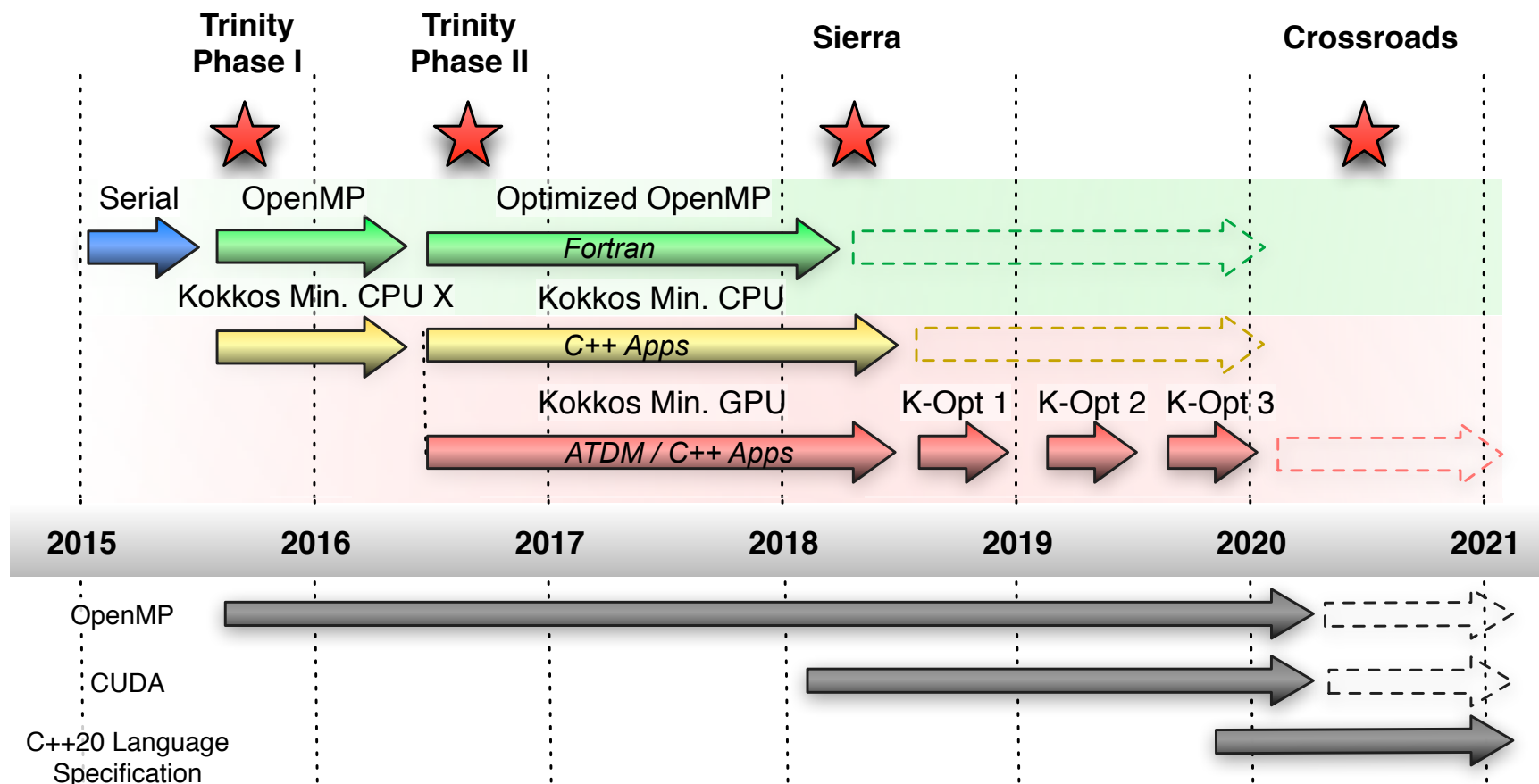
Exploration

Foundational

Libraries

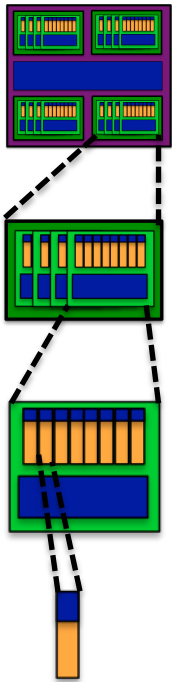
Applications

Developing a Migration Timetable



Foundational Technology: KokkosKernels

- Provide BLAS (1,2,3); Sparse; Graph and Tensor Kernels
- Kokkos based: Performance Portable
- Interfaces to vendor libraries if applicable (MKL, CuSparse, ...)
- Goal: Provide kernels for all levels of node hierarchy



Socket

- Thread Teams, Shared L3, e.g. Full Solve

Core

- Thread Parallelism, Shared L1/L2, e.g. Subdomain Solve

Hyper Thread

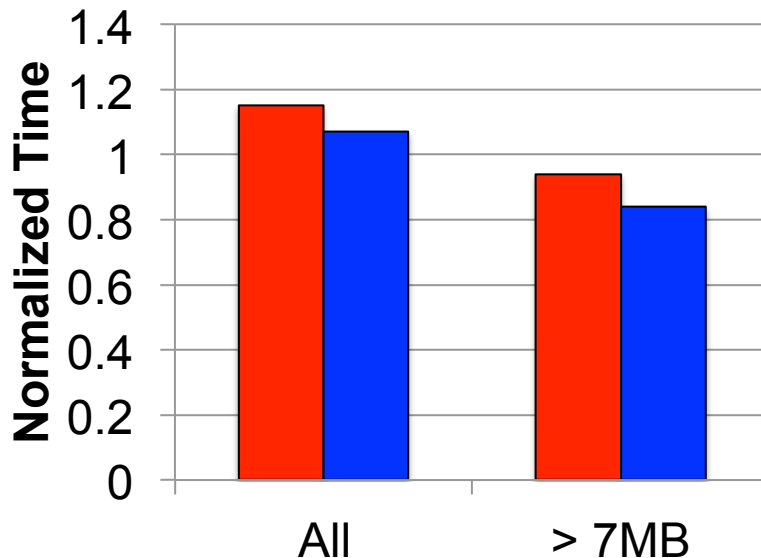
- Vector Parallelism, Synch free, e.g. Matrix Row x Vector

Vector Lane

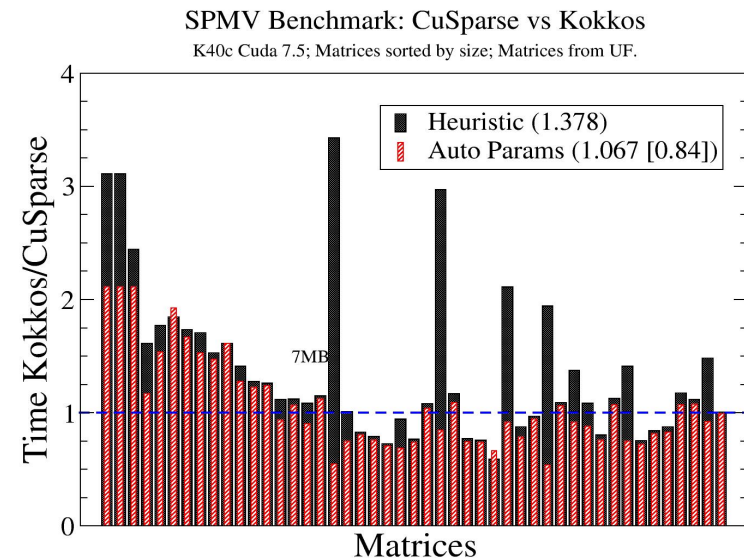
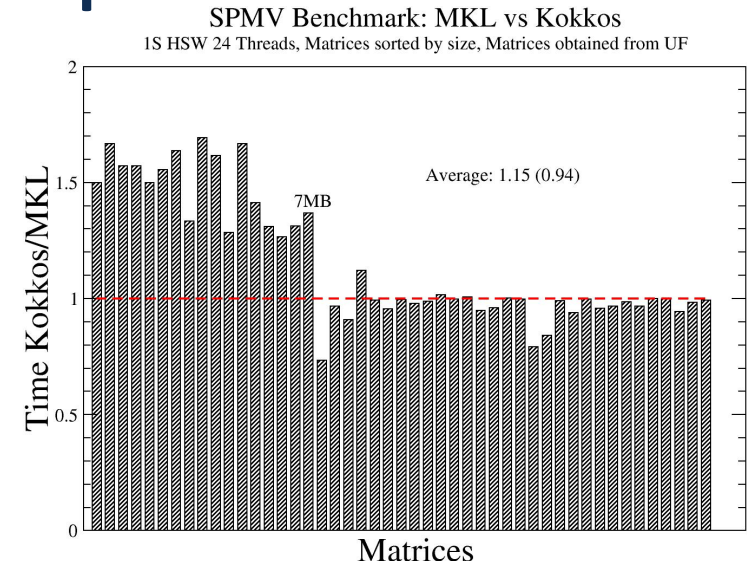
- Elemental Functions, Serial, e.g. 3x3 DGEMM

Performance Portability Example: SPMV

- Single Implementation
- Parameterized Hierarchical Parallelism
- Beats vendor libraries for relevant matrix sizes



■ Kokkos/MKL ■ Kokkos/CuSparse



Solver Stack: Preparing Trilinos for ATS-1 and 2

- Baseline Linear Algebra Done
 - Matvec, Vector Operations etc.
- Biggest Gap: Sparse Matrix-Matrix Multiply
 - BlockCRS capabilities host only right now
 - Triangular solve host only
- Some Threaded Algorithms Converge Differently
 - e.g., Gauss-Seidel, ILU, iterative triangular solve
 - some not integrated yet, others need to be requested explicitly
- Higher Level Iterative Solvers Mostly OK
 - Works if they call threaded baseline functions (e.g. CG Solve)
- Preconditioners OK if Composed of Baseline functions
 - E.g. Jacobi preconditioning works

Trilinos Working Across Platforms

Trinity

Sierra

Haswell

KNL

Power

Power + GPU

GCC 4.7.2:



GCC 5.1.0:



Intel 15:



Intel 16:



Intel 17:



Intel 17:



SON: Bowman
SRN: Morgan

GCC 4.9.2:



XL 13.1.3:



GCC 4.9.2:



CUDA 7.5:

GCC 4.9.2:



CUDA 8.0:

XL 13.1.3:



CUDA 7.5:

SON: Shepard
SRN: Mutrino

ARM

GCC 5.3.0:



SON: Sullivan

Pass Rate



>99%



>96%



<96%

SON: White
SRN: Ride

Exploration



Foundational



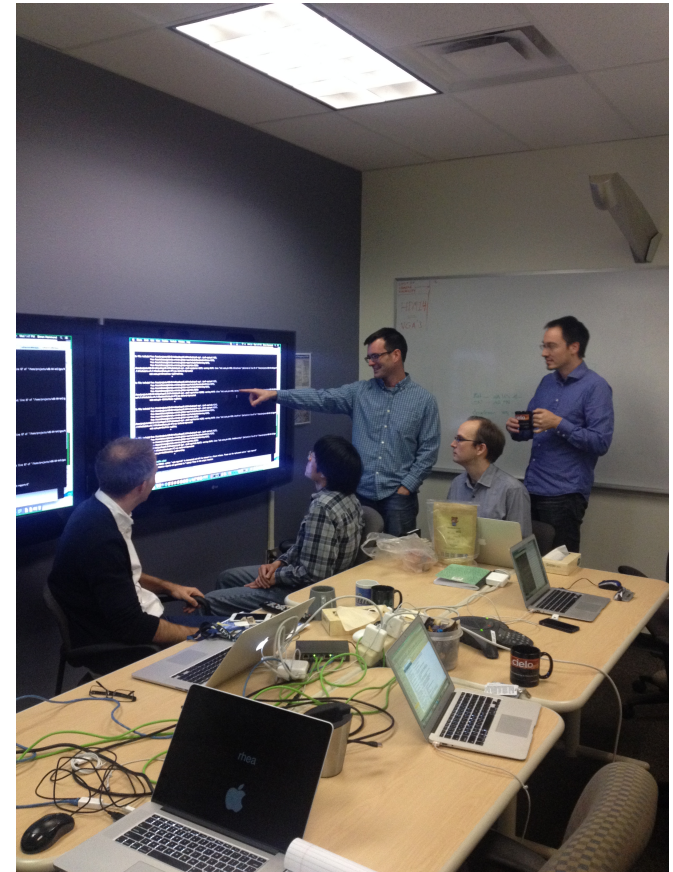
Libraries



Applications

Helping Applications Adopt Next Gen Technologies

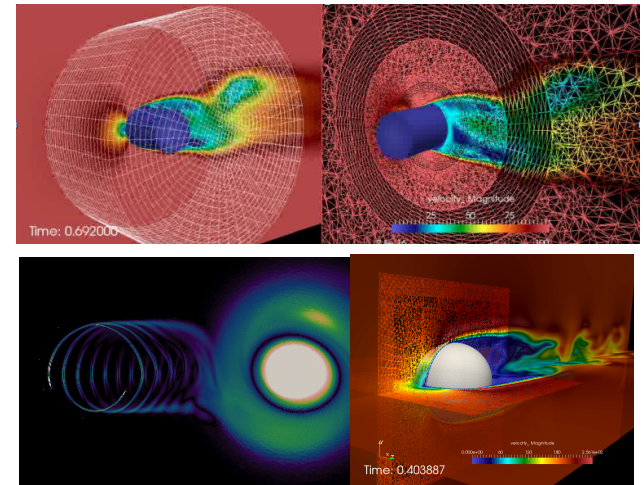
- Extensive Training Program
 - Programming Guide for Kokkos
 - Frequent Tutorials
- Integrated Codes started Migration to Kokkos
 - Exploration of algorithm choices for both KNL and Power/GPU
 - Adding initial threading support in applications
- ATDM Codes Using Kokkos from Beginning
 - Target KNL and Power+GPU from the start
 - All codes are running on GPUs at this point
- Support through code reviews as well as Team Room session
 - Sierra SM, Sierra TF, Sierra Toolkit bi-weekly half-day meetings
 - Discuss code changes, profile codes, implement prototypes



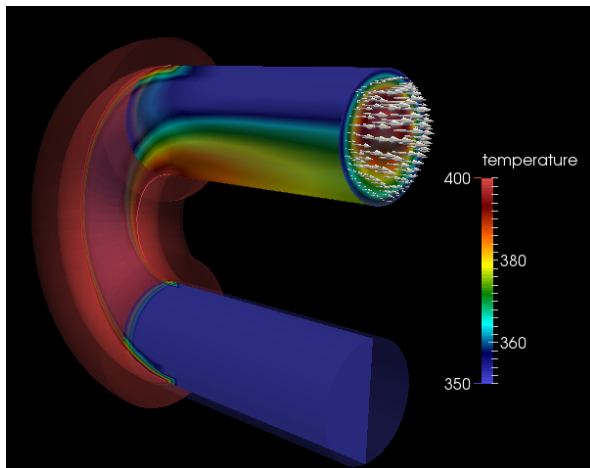
NALU: a CFD Application as Prototype

Nalu Production (Master)

- www.github.com/spdamin/Nalu
- Patterns, and data structures shared with IC production apps
- Uses Trilinos Solver stack
- Scaling demonstrated to >500k cores



2D/3D Sliding/Overset Mesh



MultiPhysics CHT

Migrating Nalu with Kokkos

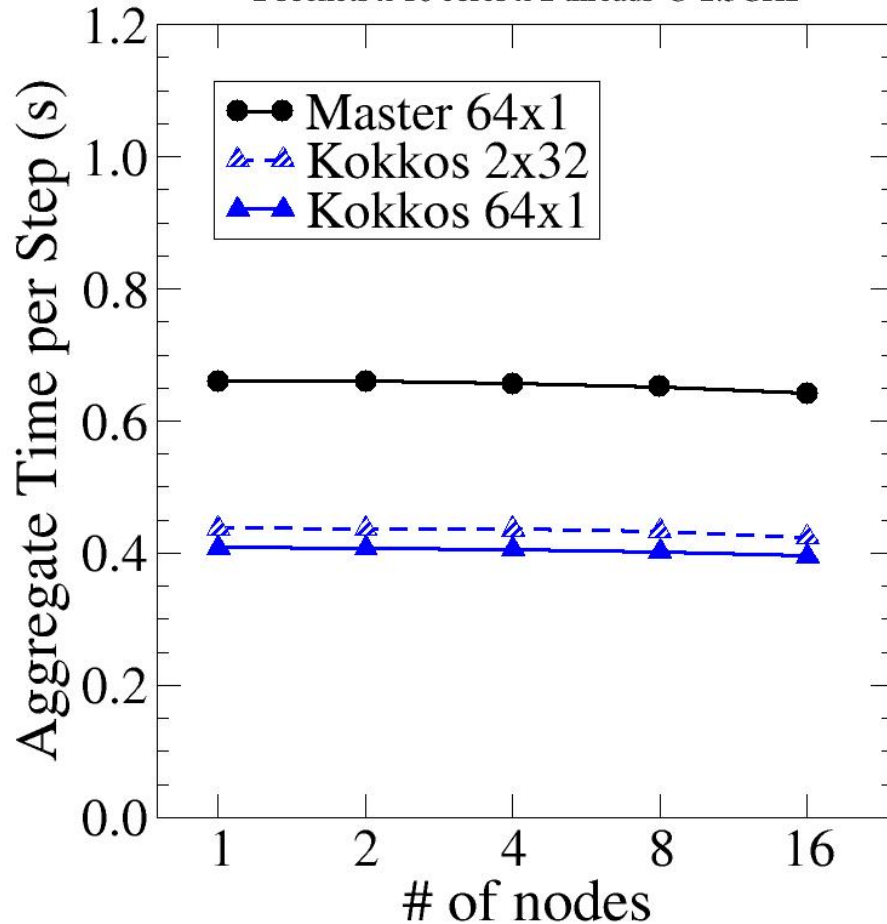
- www.github.com/crtrott/Nalu
- Migrating Nalu with Kokkos
- Prototype Kokkos assembly
- Used to harden threaded solver stack
- Get Nalu ready for Trinity Phase II

Making Matrix Assembly Thread Scalable

Simple Heat Conduction Problem, 2M elements, strong-scaling

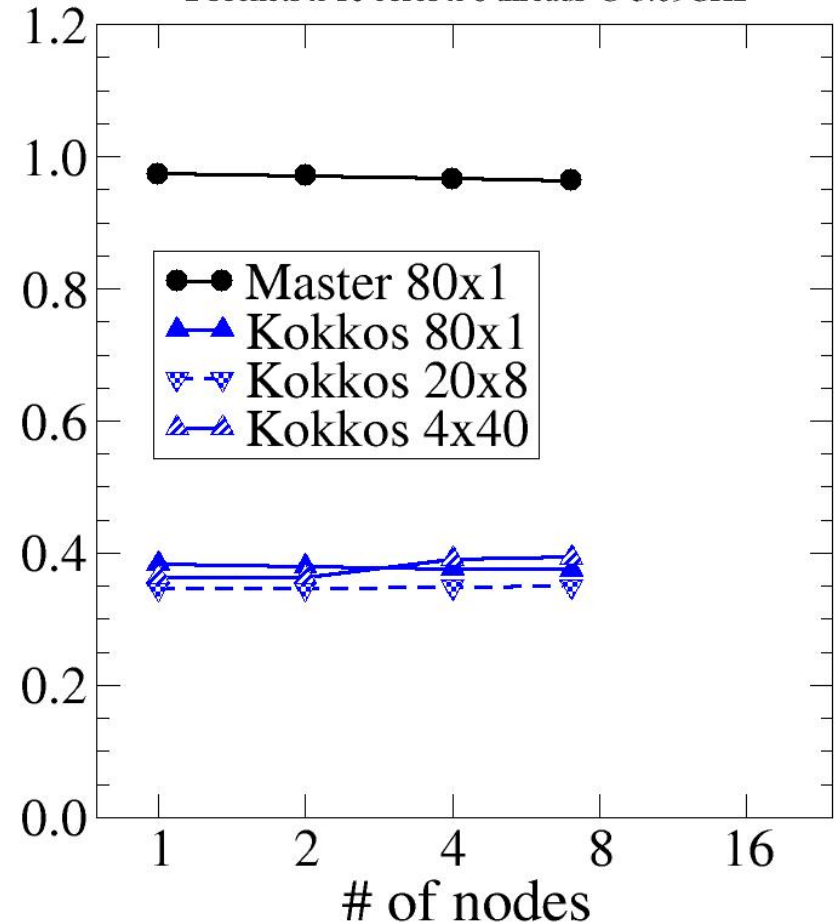
Intel Haswell

2 sockets x 16 cores x 2 threads @ 2.3GHz



IBM Power8

2 sockets x 10 cores x 8 threads @ 3.69GHz

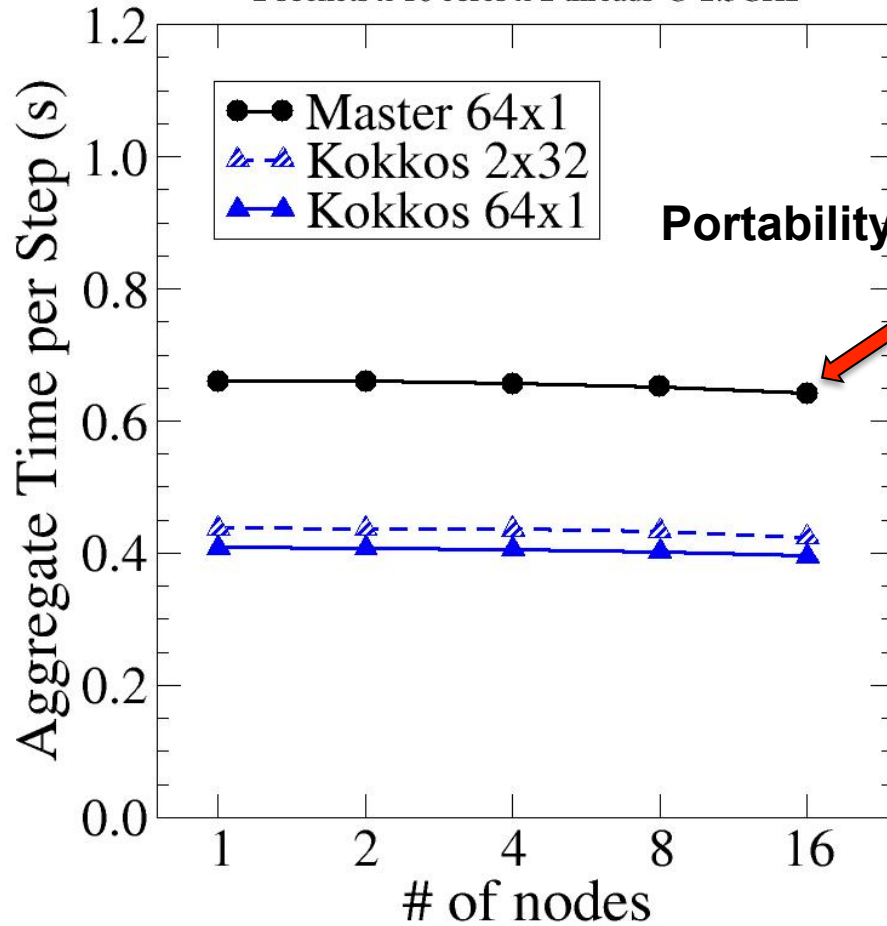


Making Matrix Assembly Thread Scalable

Simple Heat Conduction Problem, 2M elements, strong-scaling

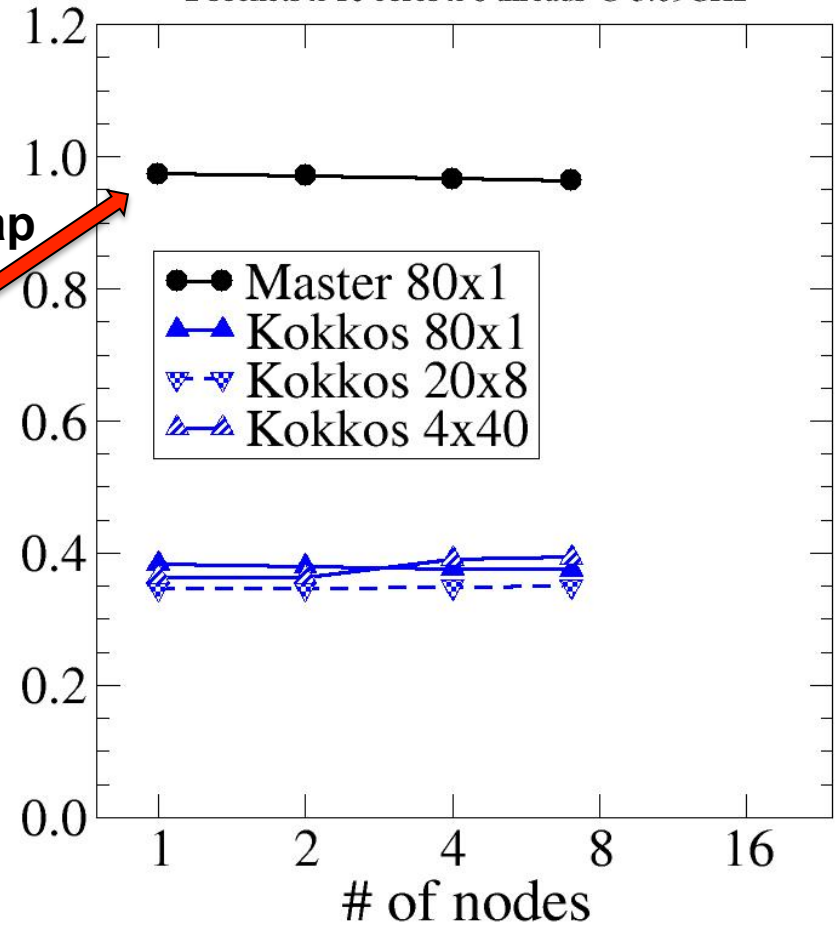
Intel Haswell

2 sockets x 16 cores x 2 threads @ 2.3GHz



IBM Power8

2 sockets x 10 cores x 8 threads @ 3.69GHz

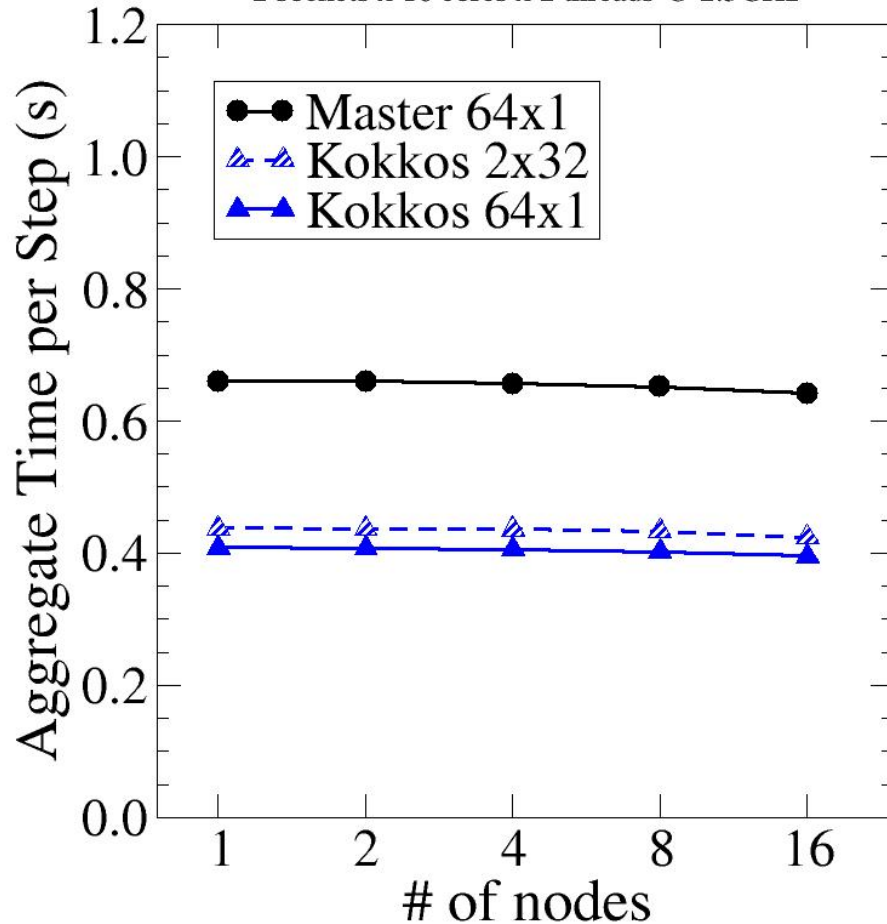


Making Matrix Assembly Thread Scalable

Simple Heat Conduction Problem, 2M elements, strong-scaling

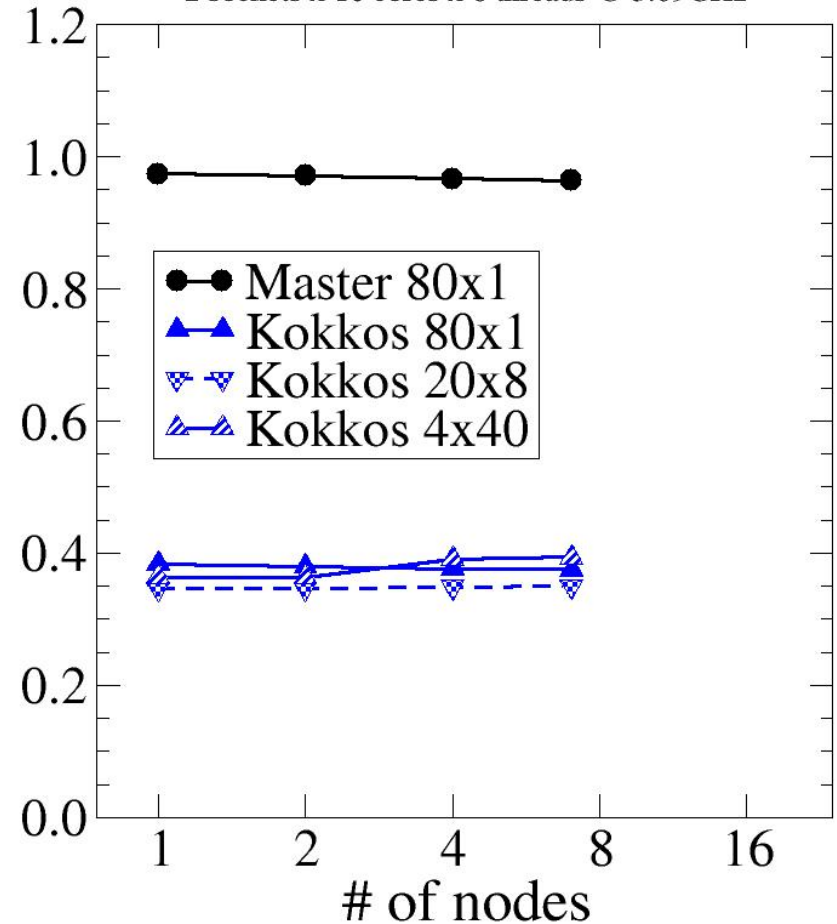
Intel Haswell

2 sockets x 16 cores x 2 threads @ 2.3GHz



IBM Power8

2 sockets x 10 cores x 8 threads @ 3.69GHz

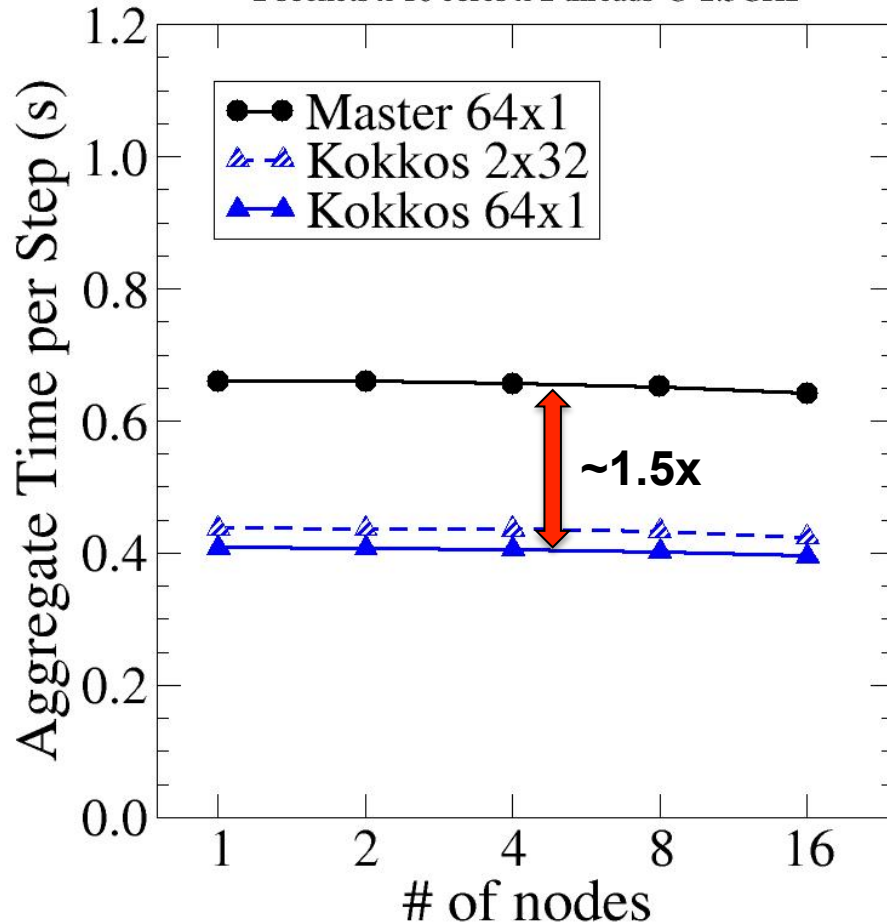


Making Matrix Assembly Thread Scalable

Simple Heat Conduction Problem, 2M elements, strong-scaling

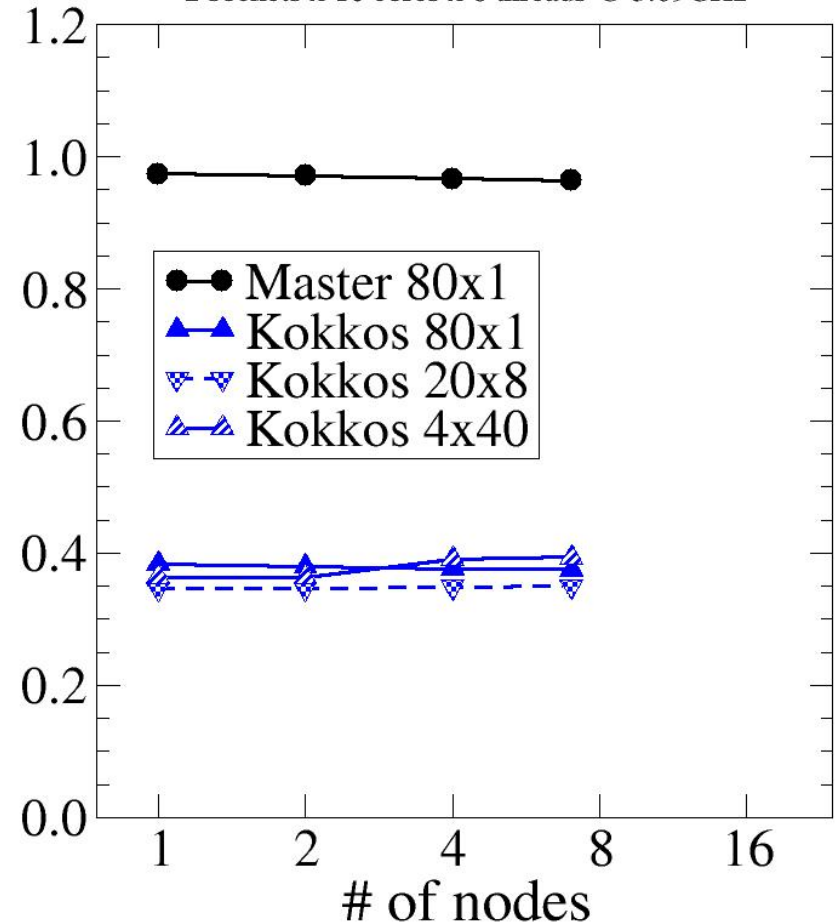
Intel Haswell

2 sockets x 16 cores x 2 threads @ 2.3GHz



IBM Power8

2 sockets x 10 cores x 8 threads @ 3.69GHz

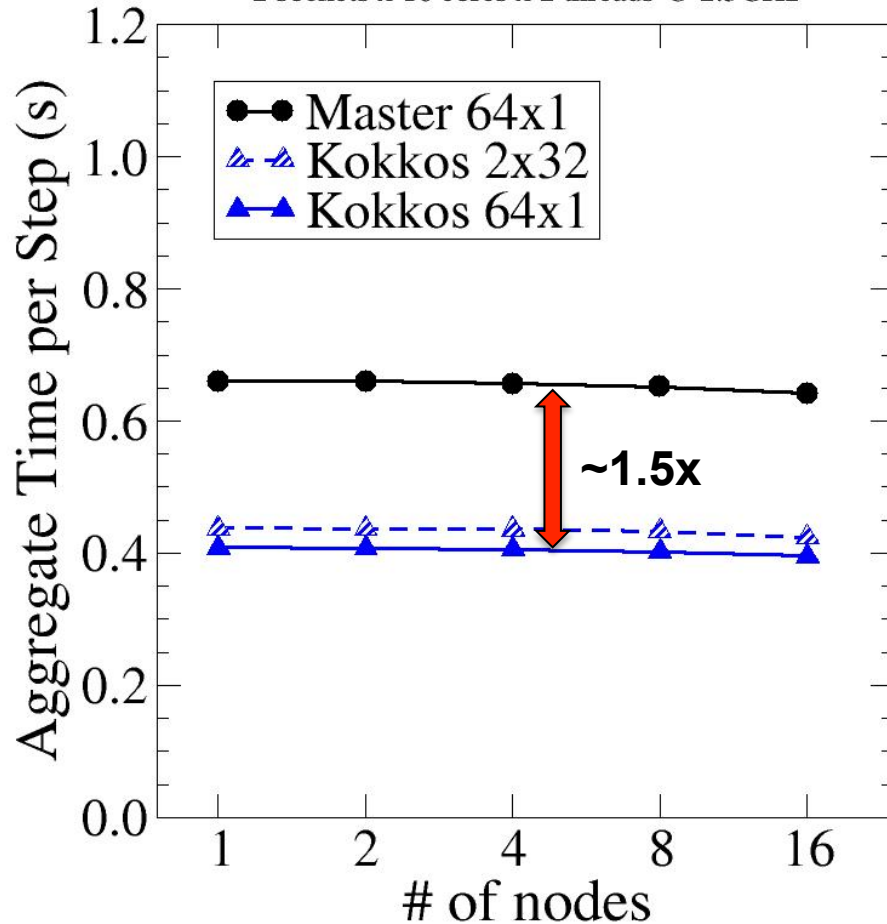


Making Matrix Assembly Thread Scalable

Simple Heat Conduction Problem, 2M elements, strong-scaling

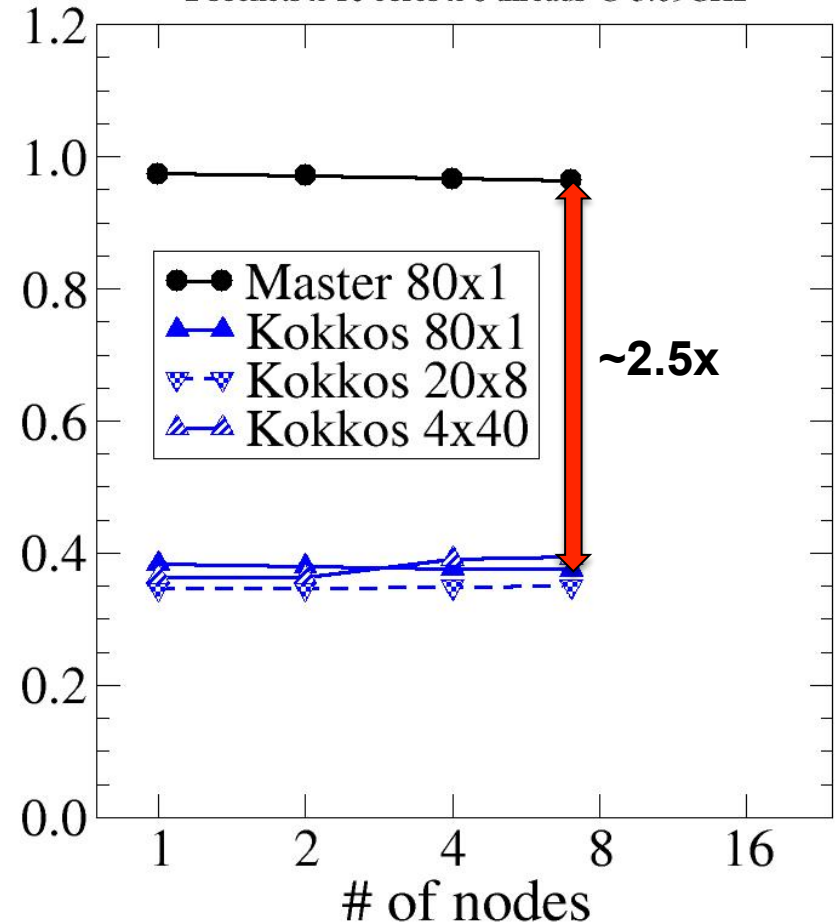
Intel Haswell

2 sockets x 16 cores x 2 threads @ 2.3GHz



IBM Power8

2 sockets x 10 cores x 8 threads @ 3.69GHz

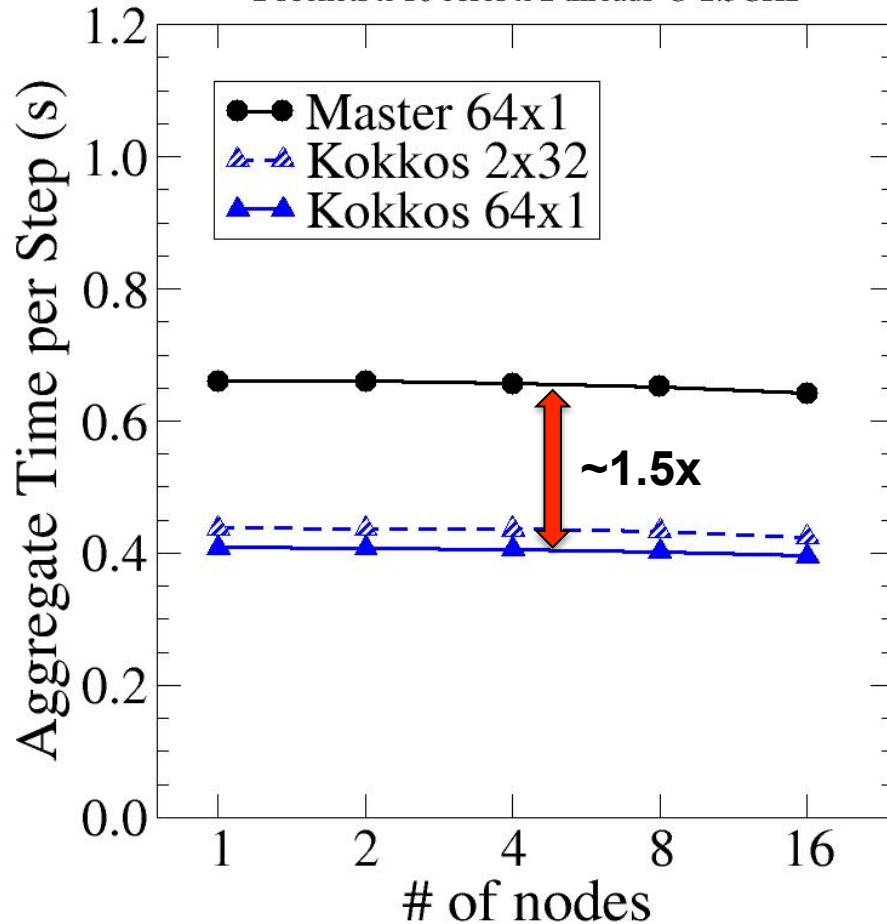


Making Matrix Assembly Thread Scalable

Simple Heat Conduction Problem, 2M elements, strong-scaling

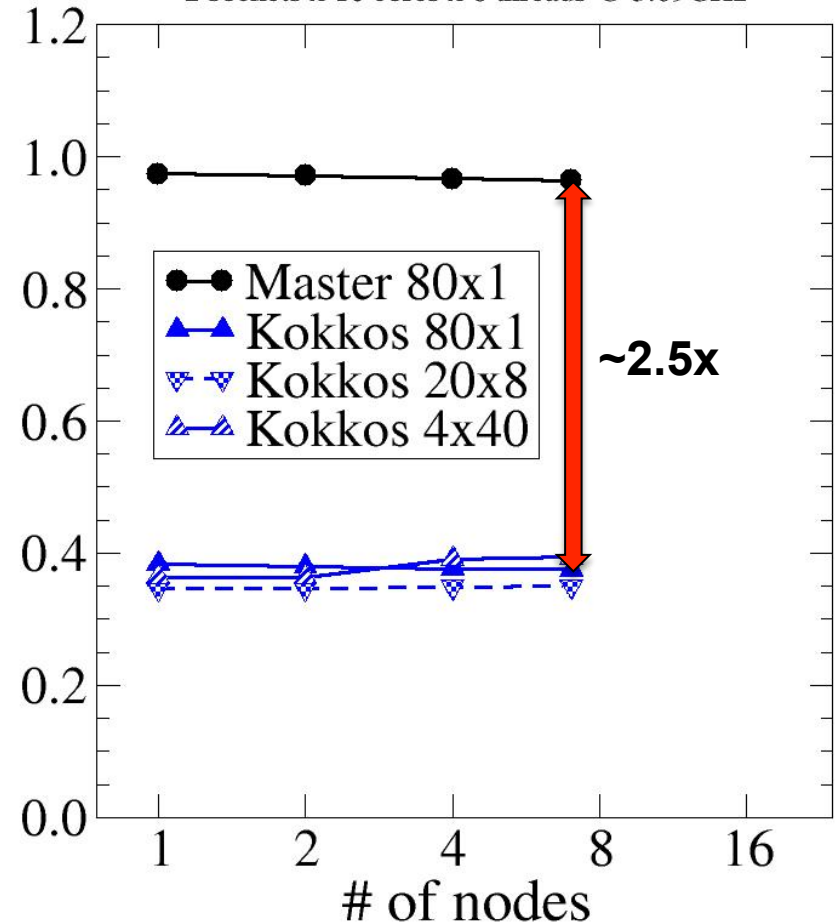
Intel Haswell

2 sockets x 16 cores x 2 threads @ 2.3GHz



IBM Power8

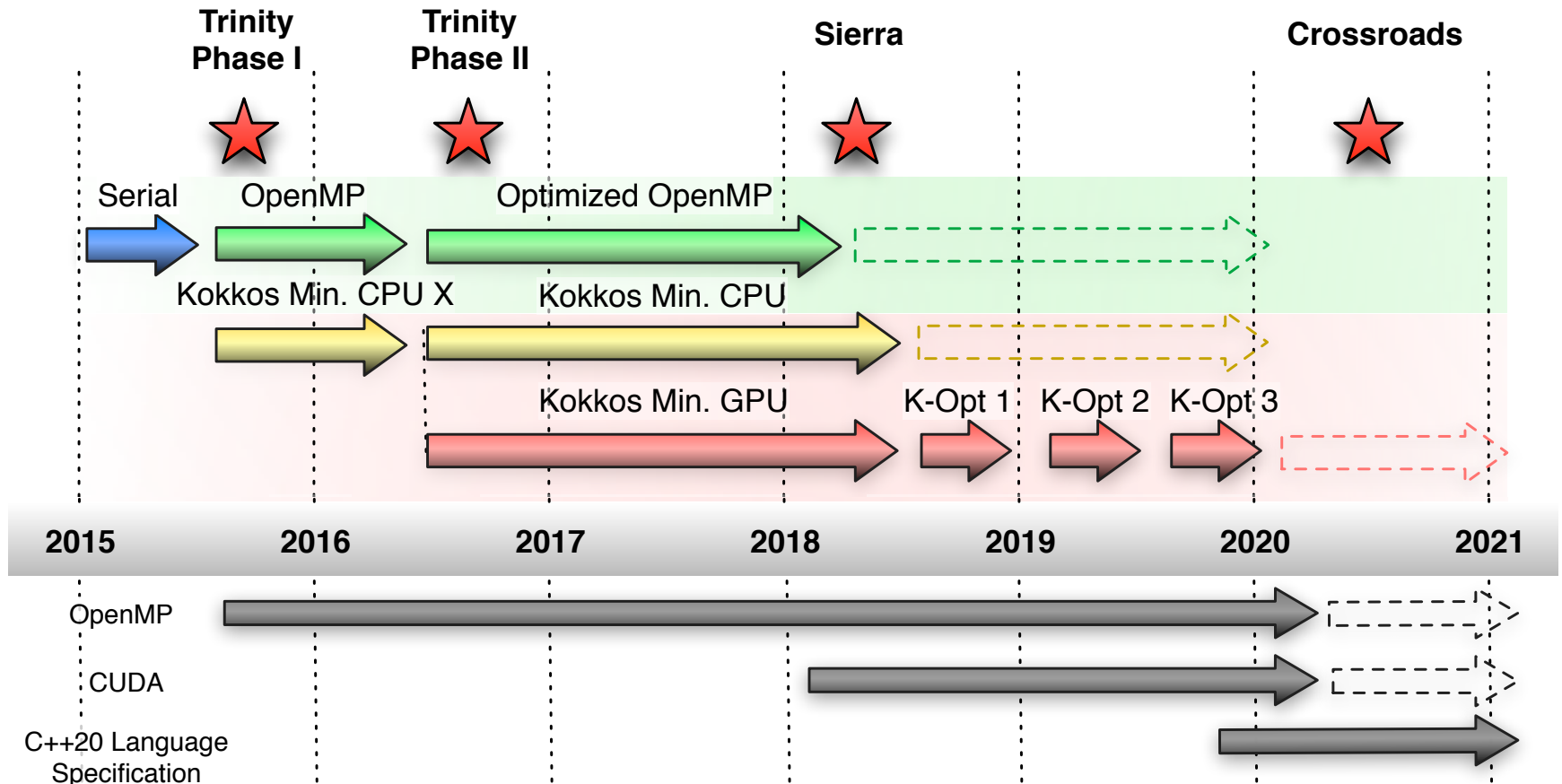
2 sockets x 10 cores x 8 threads @ 3.69GHz



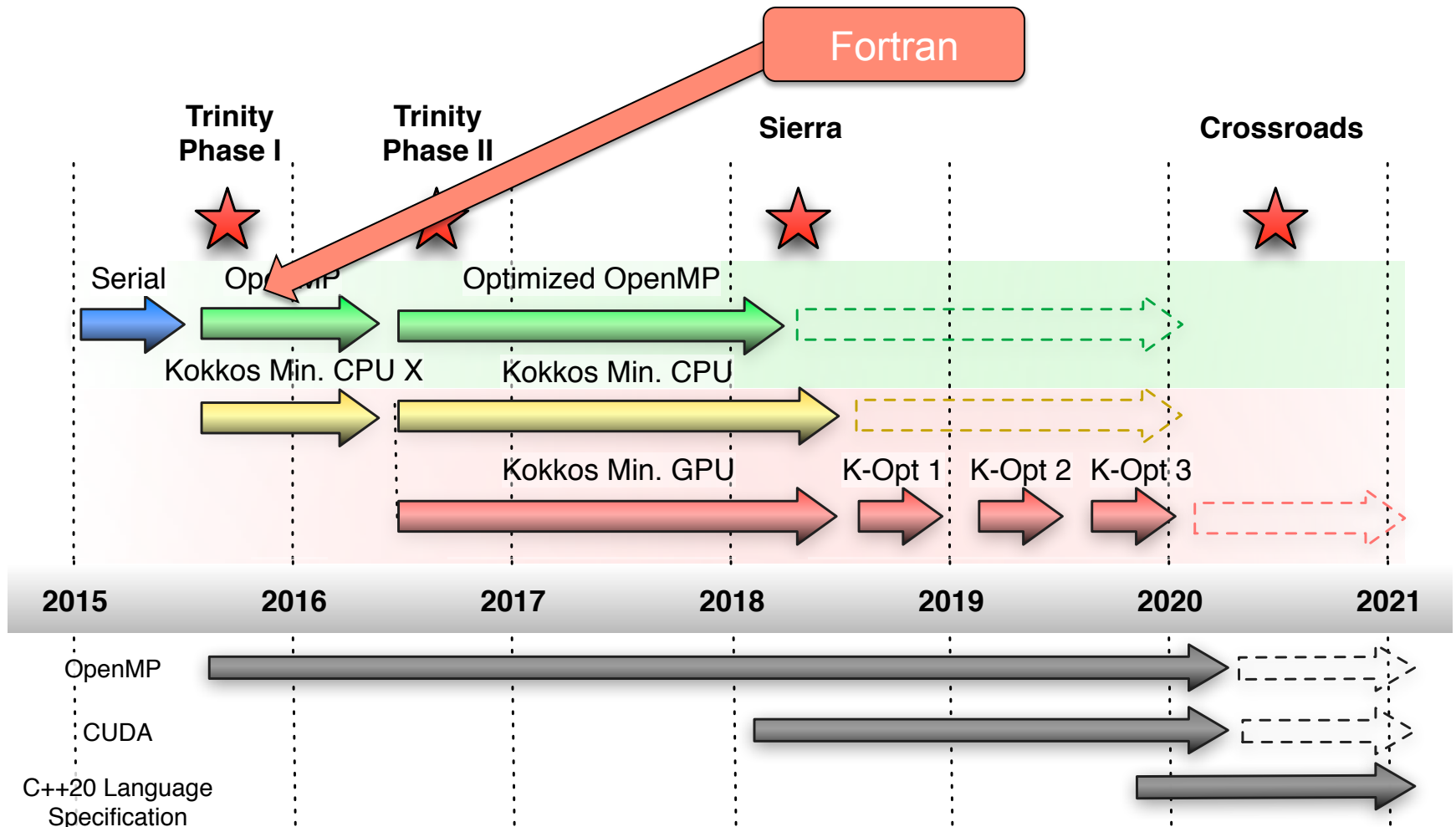
KNL: looks similar, data still under NDA

GPUs: data structures are not transitioned

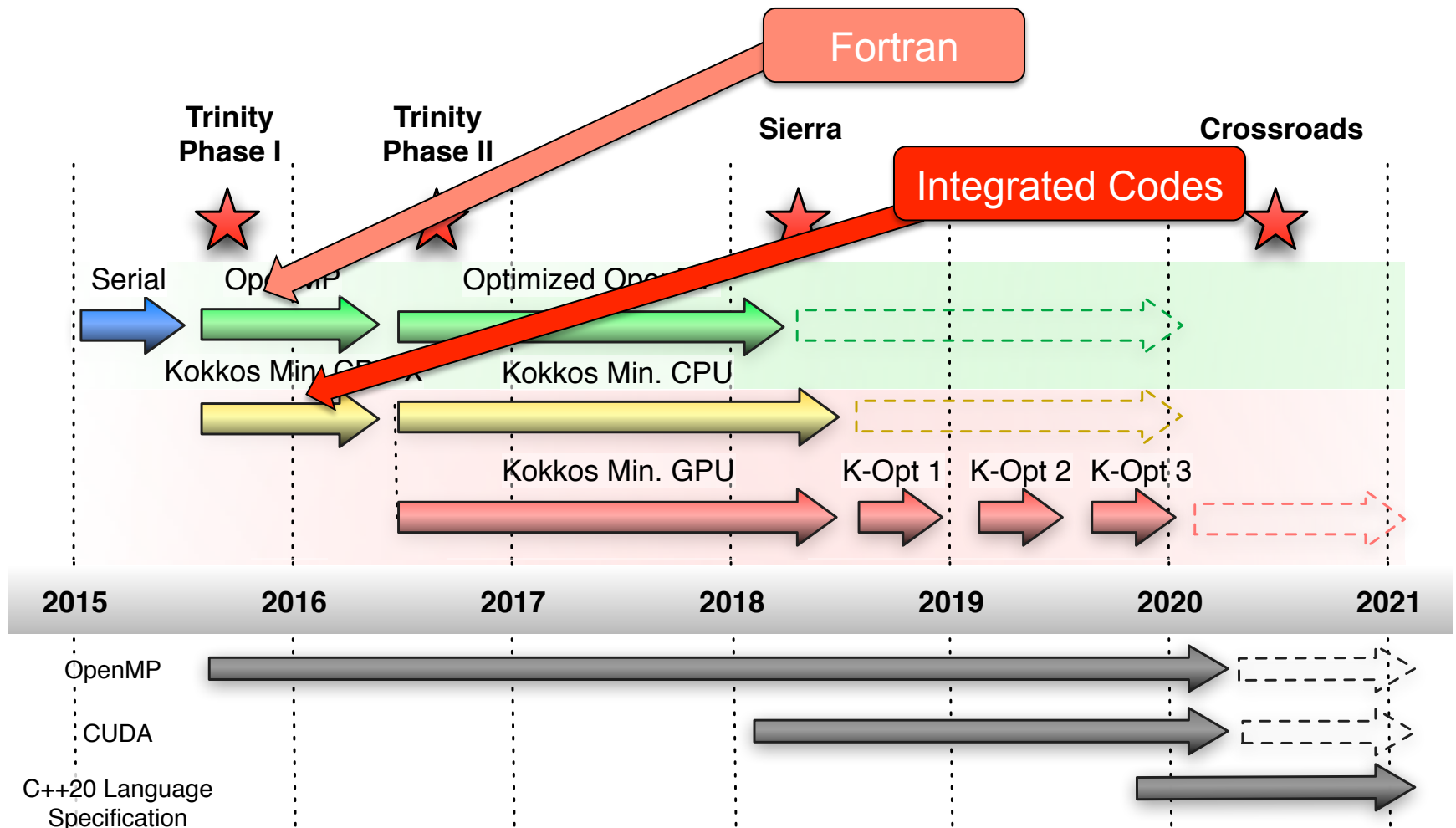
The Migration Plan: Current State



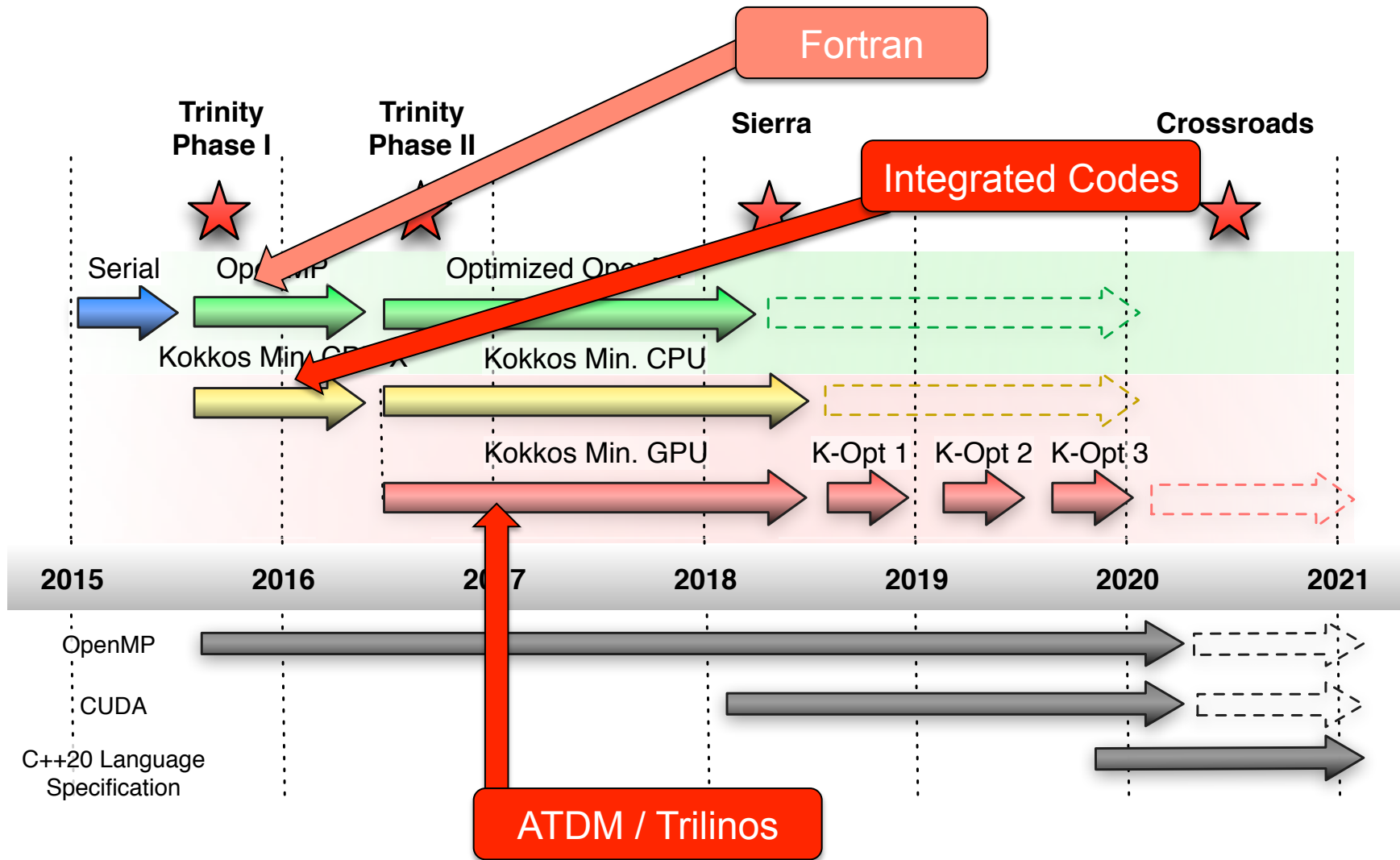
The Migration Plan: Current State

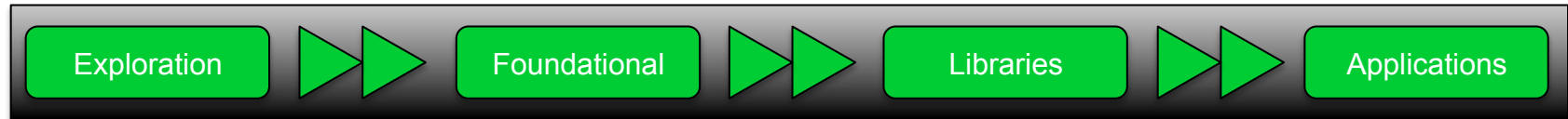


The Migration Plan: Current State



The Migration Plan: Current State





- **Sandia early adopter production codes on track for ASC capability machine deployments**
- CoDesign through the hardware and software stack has been **critical** to identifying problems and solutions
- Fundamental NGP technologies are transitioning from research to existing production apps and future simulation codes
- Majority of Trilinos running on all major HPC platforms
- Huge value to early access test beds and prototype software stacks



Sandia
National
Laboratories

Exceptional service in the national interest

Questions: crtrott@sandia.gov