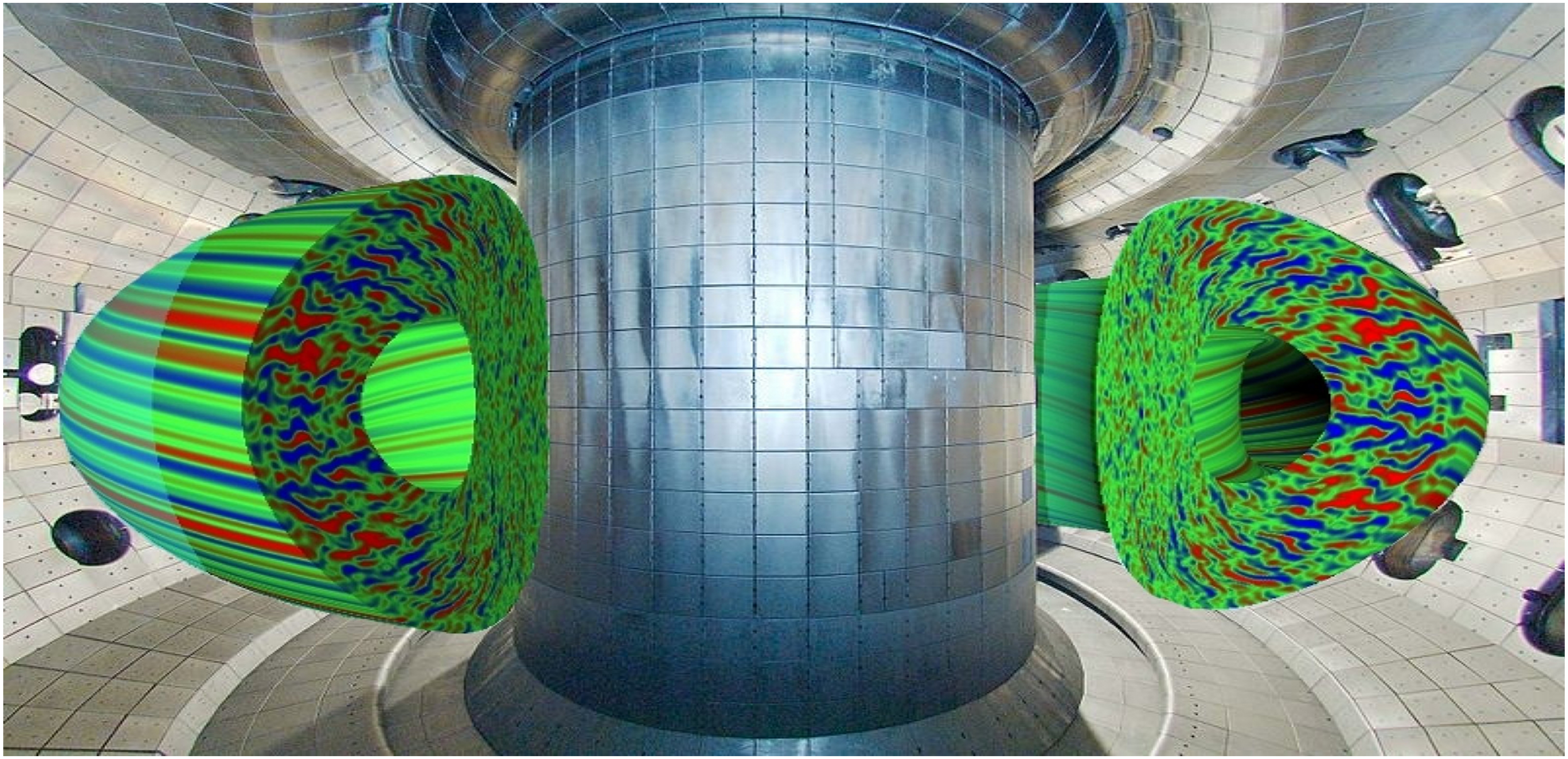
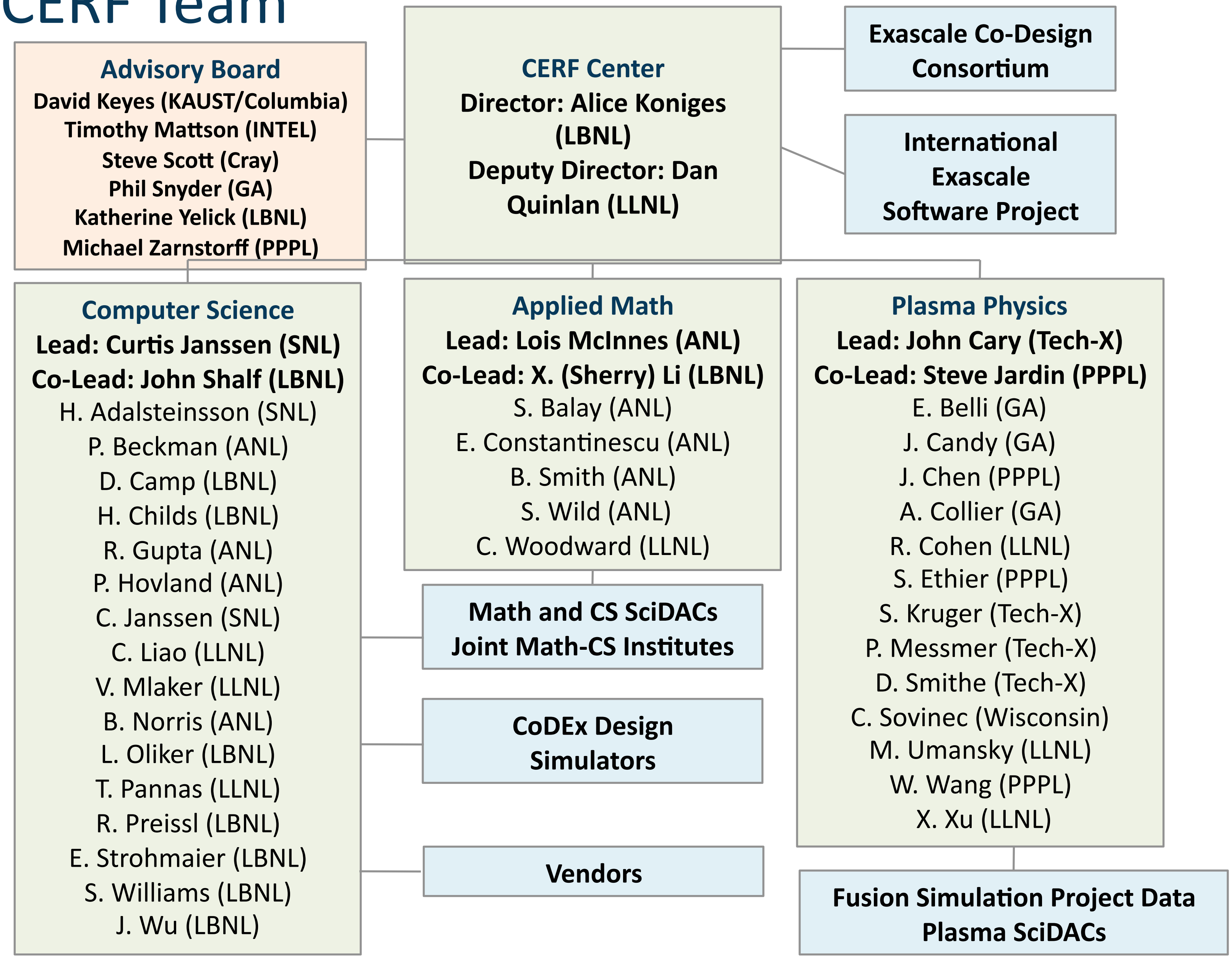


The CERF Center

Co-design for Exascale Research in Fusion

Director: Alice Koniges, Lawrence Berkeley National Laboratory, AEKoniges@lbl.gov

The CERF Team



Advisory Board
David Keyes (KAUST/Columbia)
Timothy Mattson (INTEL)
Steve Scott (Cray)
Phil Snyder (GA)
Katherine Yelick (LBNL)
Michael Zarnstorff (PPPL)

CERF Center
Director: Alice Koniges (LBNL)
Deputy Director: Dan Quinlan (LLNL)

Exascale Co-Design Consortium

International Exascale Software Project

Computer Science
Lead: Curtis Janssen (SNL)
Co-Lead: John Shalf (LBNL)
H. Adalsteinsson (SNL)
P. Beckman (ANL)
D. Camp (LBNL)
H. Childs (LBNL)
R. Gupta (ANL)
P. Hovland (ANL)
C. Janssen (SNL)
C. Liao (LLNL)
V. Mlaker (LLNL)
B. Norris (ANL)
L. Oliker (LBNL)
T. Pannas (LLNL)
R. Preissl (LBNL)
E. Strohmaier (LBNL)
S. Williams (LBNL)
J. Wu (LBNL)

Applied Math
Lead: Lois McInnes (ANL)
Co-Lead: X. (Sherry) Li (LBNL)
S. Balay (ANL)
E. Constantinescu (ANL)
B. Smith (ANL)
S. Wild (ANL)
C. Woodward (LLNL)

Math and CS SciDACs
Joint Math-CS Institutes

CoDEx Design Simulators

Vendors

Plasma Physics
Lead: John Cary (Tech-X)
Co-Lead: Steve Jardin (PPPL)
E. Belli (GA)
J. Candy (GA)
J. Chen (PPPL)
A. Collier (GA)
R. Cohen (LLNL)
S. Ethier (PPPL)
S. Kruger (Tech-X)
P. Messmer (Tech-X)
D. Smithe (Tech-X)
C. Sovinec (Wisconsin)
M. Umansky (LLNL)
W. Wang (PPPL)
X. Xu (LLNL)

Fusion Simulation Project Data
Plasma SciDACs

CERF Code Areas and Leads

Core Transport:

GYRO/NEO, J. Candy, E. Belli, A. Collier (GA)

Collisional Edge Plasma:

BOUT++, M. Umansky, R. Cohen, X. Xu (LLNL)

MHD:

M3D-C1, S. Jardin, J. Chen (PPPL);

NIMROD, C. Sovinec (Wisconsin), S. Kruger (Tech-X)

Explicit PIC Modeling:

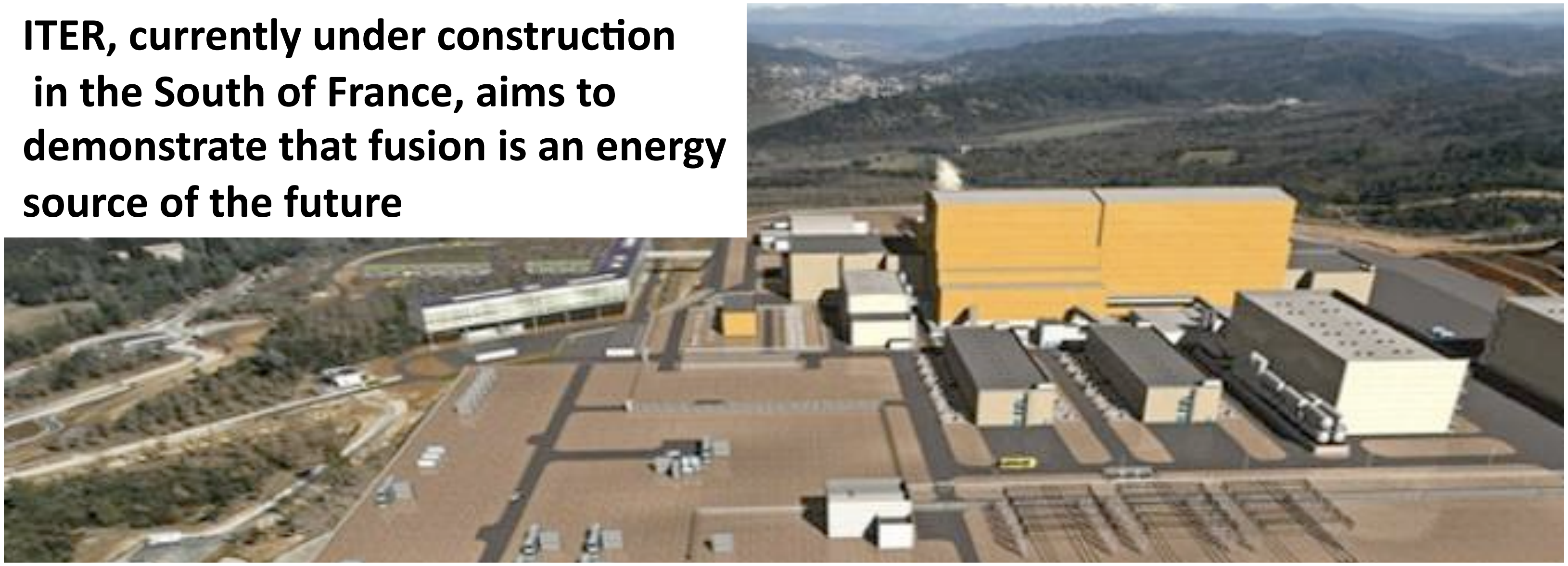
GTS, W. Wang, S. Ethier (PPPL);

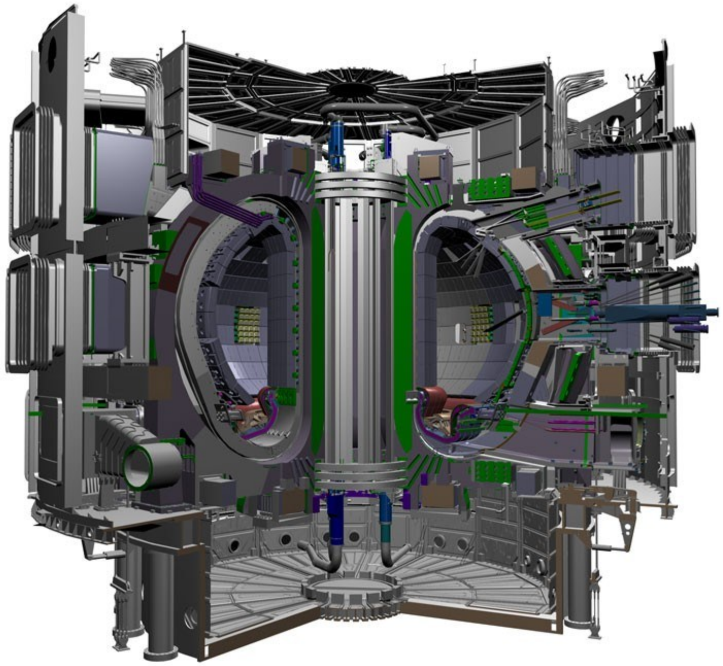
VORPAL, P. Messmer, D. Smithe (Tech-X)

Code Integration Framework:

FACETS, J. Cary, S. Kruger (Tech-X)

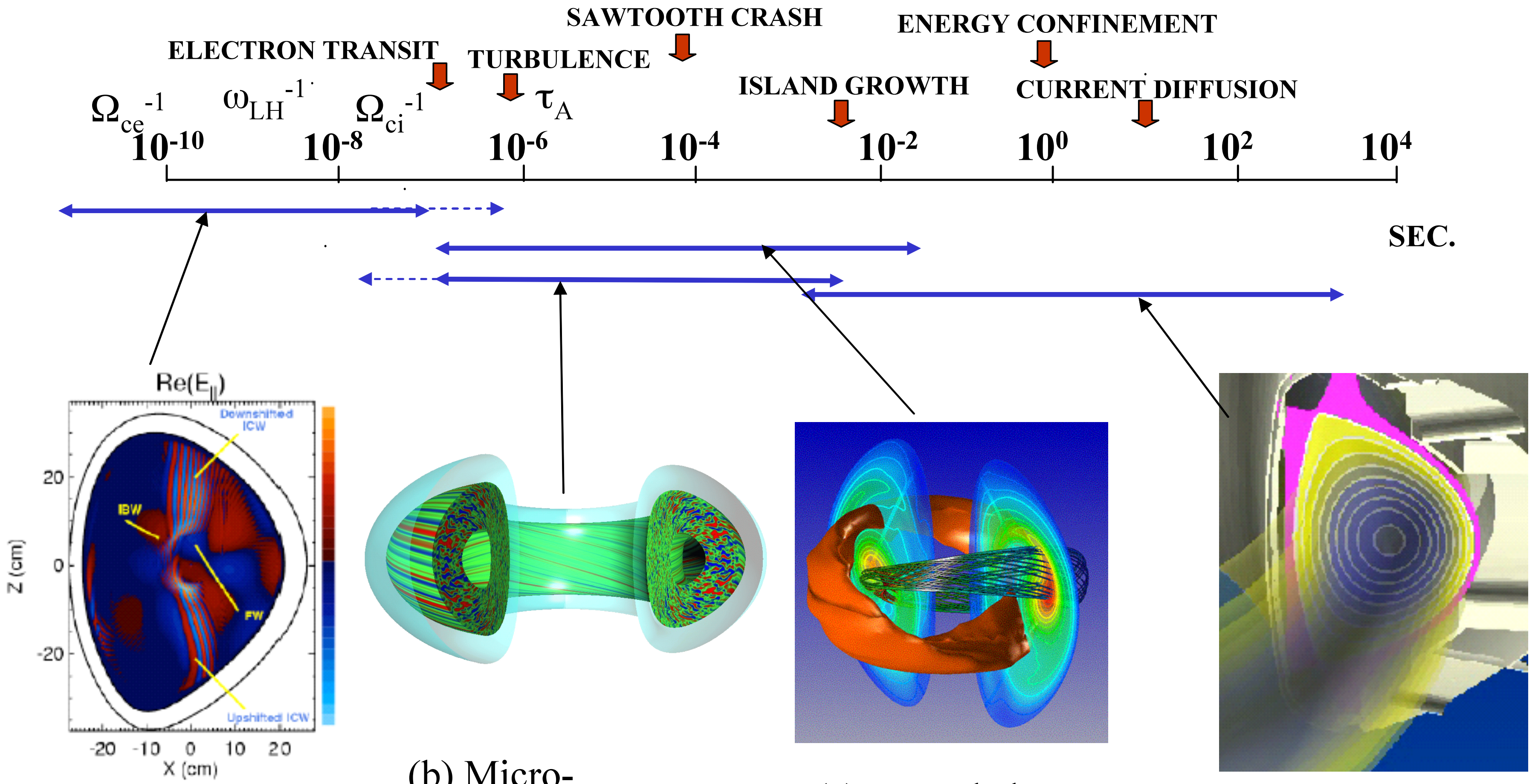
ITER, currently under construction in the South of France, aims to demonstrate that fusion is an energy source of the future





- **Top-to-bottom exascale computer design is essential for efficient design/ operation of large-scale experiments**
  - Typical ITER discharge can be estimated at 1M$

The magnetic fusion codes are challenged by the large range of temporal and spatial scales



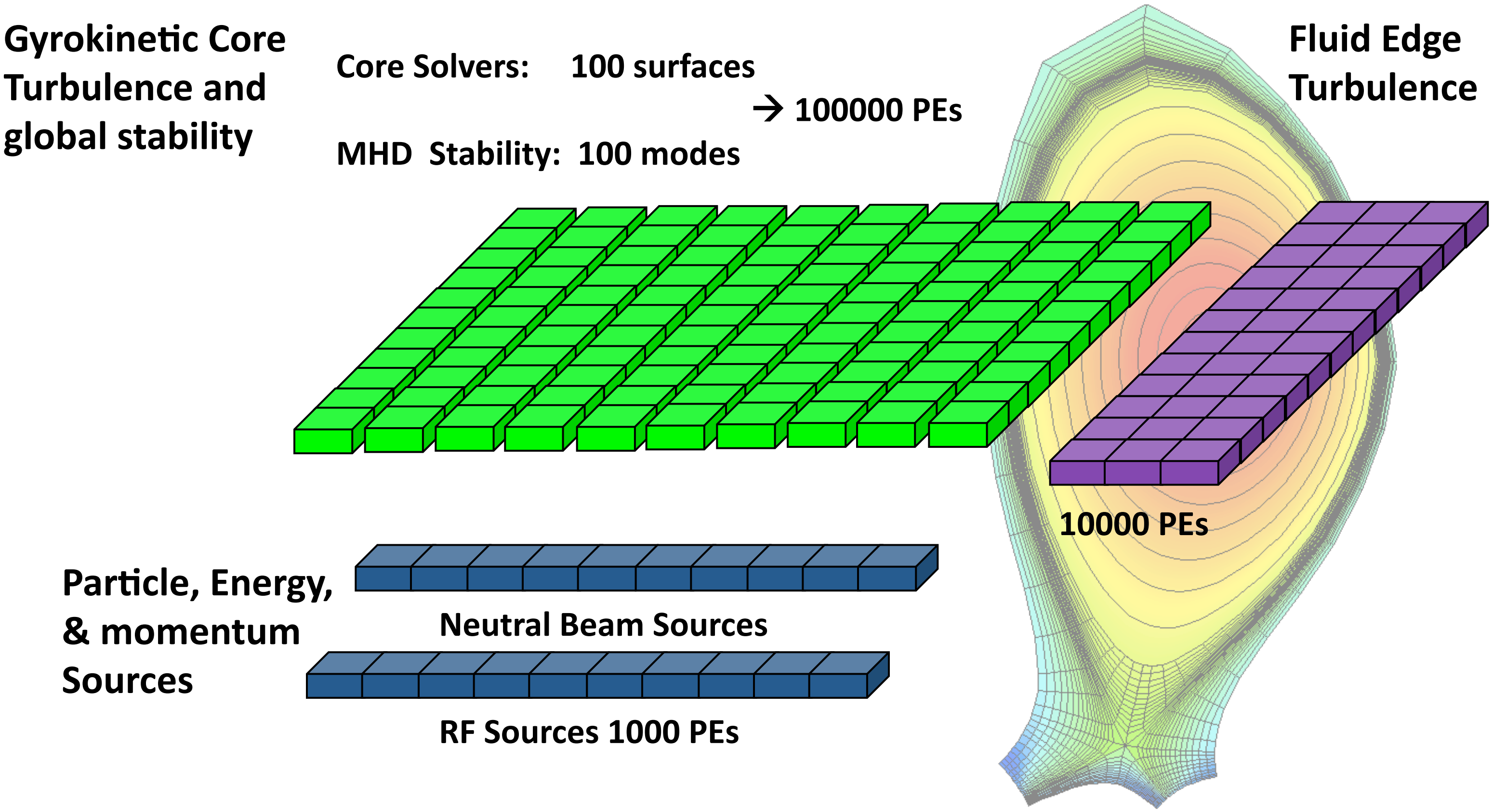
(a) RF codes

(b) Micro-turbulence codes

(c) Extended-MHD codes

(d) Transport Codes

What makes this effort unique:
CERF vision for an integrated exascale simulation



Gyrokinetic Core Turbulence and global stability

Core Solvers: 100 surfaces

→ 100000 PEs

MHD Stability: 100 modes

Fluid Edge Turbulence

10000 PEs

Particle, Energy, & momentum Sources

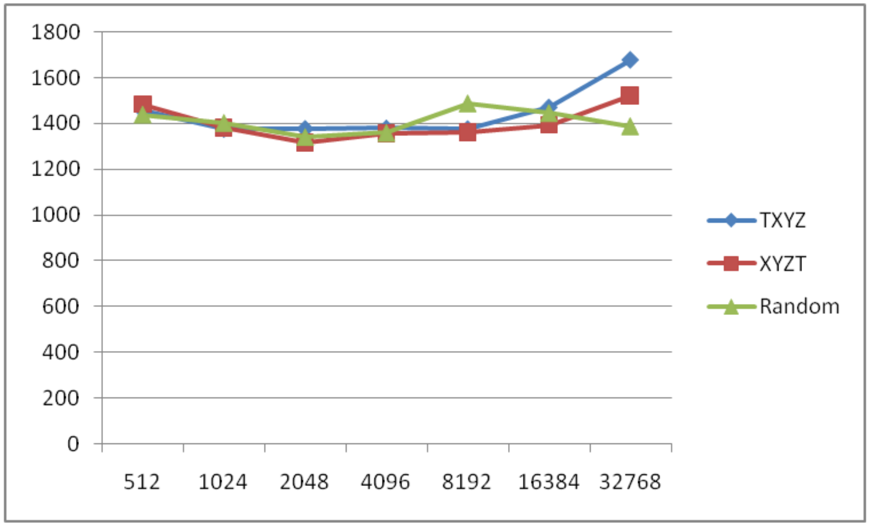
Neutral Beam Sources

RF Sources 1000 PEs

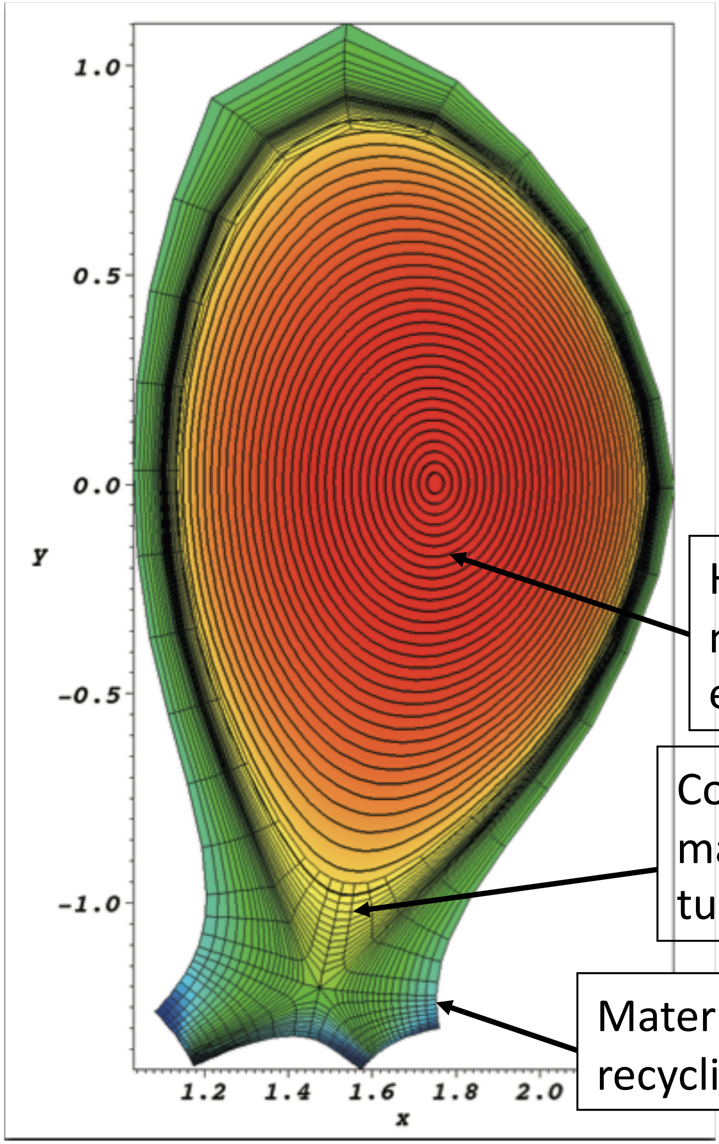
Categories of CERF Applications

- Group 1: Discretized fluid equations requiring implicit solutions: BOUT++, M3D-C1, NIMROD
  - incorporates global linear and nonlinear solvers to handle both advection and diffusion, including issues of global communication and sparse matrix operations
  - Different grid layouts and parallelization strategies
- Group 2: Particle-in-cell (PIC) approaches: GTS, VORPAL
  - gather/scatter issue of data misalignment, as the particles close in memory have to communicate with fields at disparate locations
- Group 3: Discretized kinetic equations with implicit/explicit solvers: GYRO, NEO
  - high-dimensional continuum discretizations combined with regular spatial discretizations, for yet another kind of data layout, a mix of communication patterns, and a rich space of strategies

The GLUE: FACETS – Highly scalable coupling framework for Plasma Simulations



BGP runs of FACETS-driven-Gyro scales on 32K cores

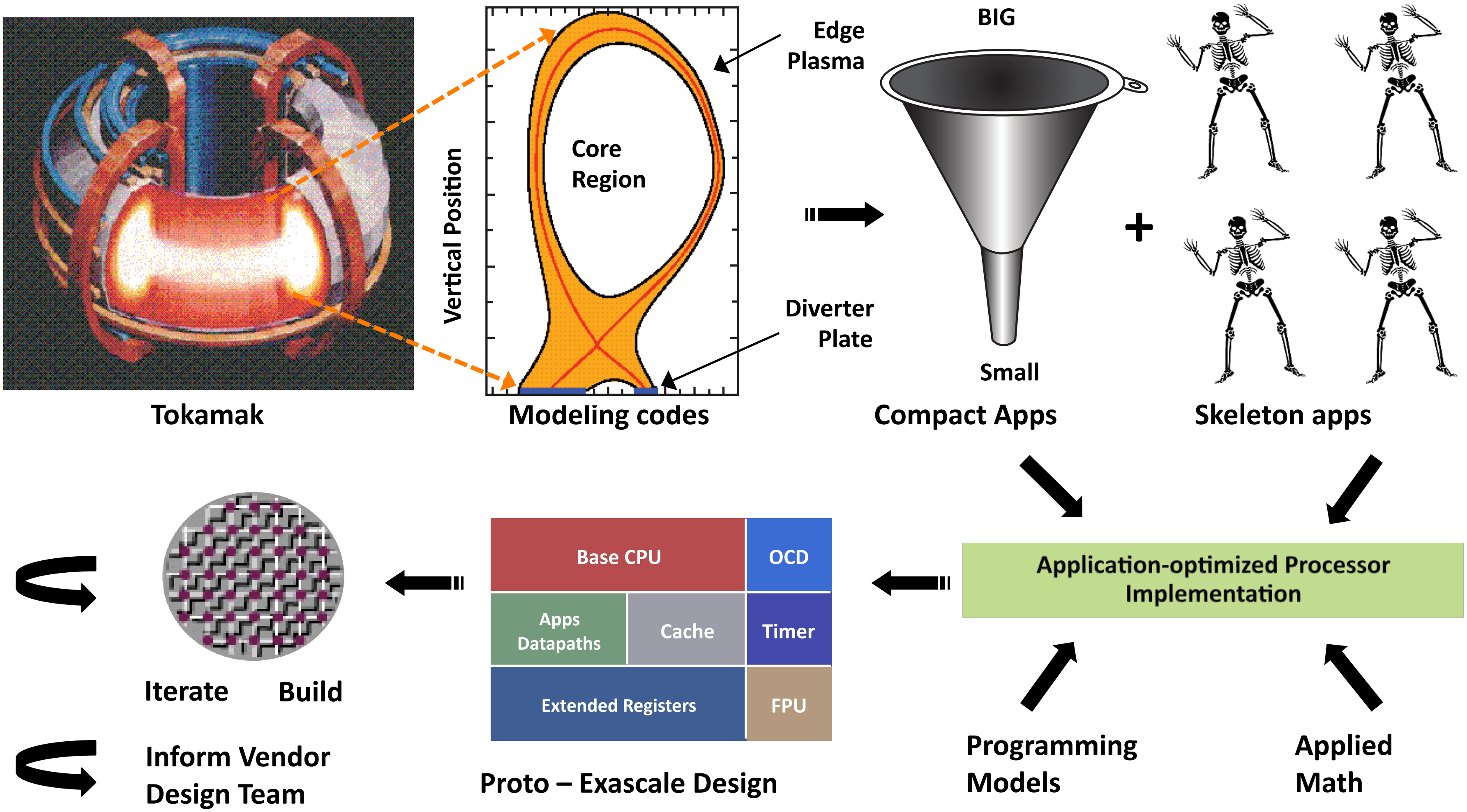


Hot central plasma: nearly completely ionized, magnetic lines lie on flux surfaces, 3D turbulence embedded in **1D** transport

Cooler edge plasma: atomic physics important, magnetic lines terminate on material surfaces, 3D turbulence embedded in **2D** transport

Material walls, embedded hydrogenic species, recycling

CERF methodology includes building skeleton and compact apps, math input, testing, build and vendors



Tokamak

Modeling codes

Compact Apps

Skeleton apps

Application-optimized Processor Implementation

Programming Models

Applied Math

Base CPU

OCD

Apps Datapaths

Cache

Timer

Extended Registers

FPU

Proto – Exascale Design

Iterate

Build

Inform Vendor Design Team

CERF Co-design process will be integrated with exascale associations and vendors

- Several high-level elements of the CERF co-design process are understood
  - – CERF will develop several co-design vehicles (CDVs)
    - Kernels to investigate node-level behavior
    - Skeletons to investigate large-scale behavior
    - Compact apps to provide a simplification of the full application
  - – Simulation will be used to predict performance
    - Hardware-based simulation for rapid turnaround of node designs
    - SST/macro: Coarse-grained network simulation to study large-scale behavior
    - Vendor simulators: will be used subject to IP issues and availability
  - – ROSE-based tools aid with automatic generation of skeletons from application code.

Co-design process is inherently iterative

- Develop realistic CDVs in collaboration with plasma physicists
- Vendors and CERF CS team analyze CDVs using simulation, prototypes, proxy hardware, etc.
- Programming models explored using the CDVs
- Simulate coupled-physics app using FACETS CDV to link other CDVs
- Vendors and CERF team examine the options for hardware and software alterations
- Iterate

What the Fusion Codes need for Exascale

- Whole application coupling using FACETS (or other) enabled by OS
- Math Research
  - Variety of TOPS scalable libraries (e.g., PETSc, SUNDIALS, SuperLU)
  - Linear solvers, nonlinear solvers, time integration
  - Communication-avoiding and latency-tolerant algorithms
- Programming Model Research – see examples of OpenMP tasking and mixed CAF/MPI/OpenMP code
- Enabling tools (e.g. ROSE for skeleton extraction and possible automatic hybridization)
- Architecture Research
  - HW simulators for exascale designs
- Tools that work with our mixed programming model codes
- UQ Analysis especially in connection with experimental data
- Data management and I/O support for full simulations and visualization

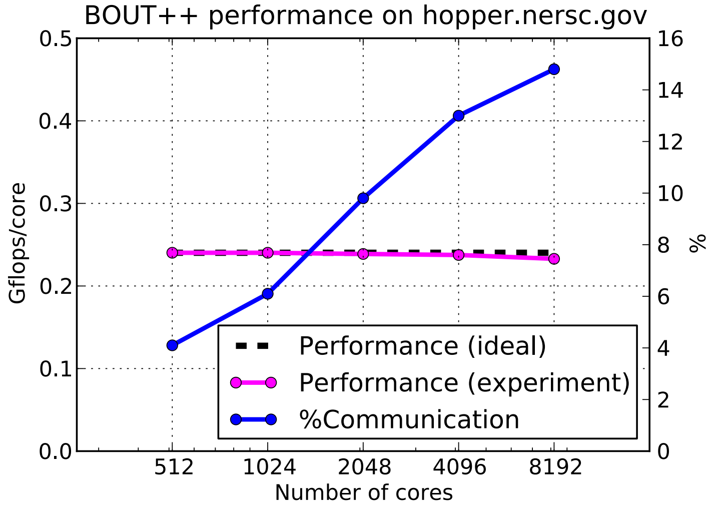
Preliminary results with DOE Co-design Center Planning Grant, ARRA, and existing SciDAC funding

- Performance analysis, bottleneck identification
  - Profiling of all component codes
  - Starting work on codes coupled with FACETS
  - Experimentation with different tools
- Programming Models
  - PGAS hybrid models on 130K cores
  - Advanced OpenMP (tasking model)
  - CUDA
- Data Analysis
- Algorithms—scalable replacements for existing solvers
- Auto-tuning

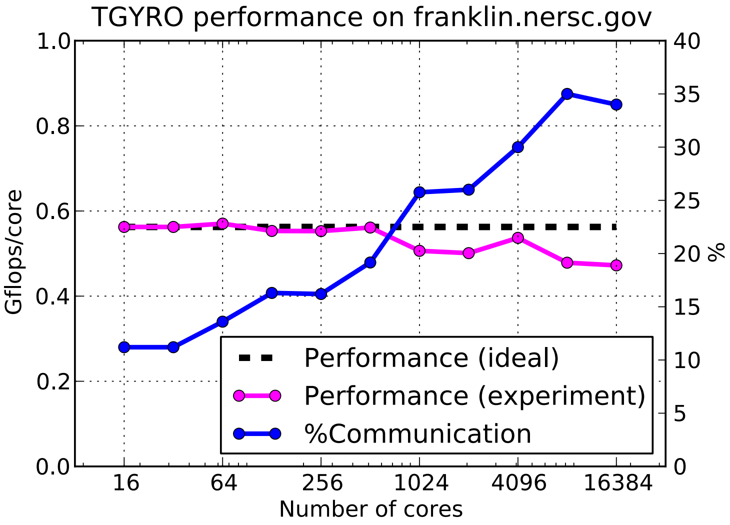
OUR Motto: Just do it …

Sample Performance results for each of the three major code groups using IPM, CrayPAT

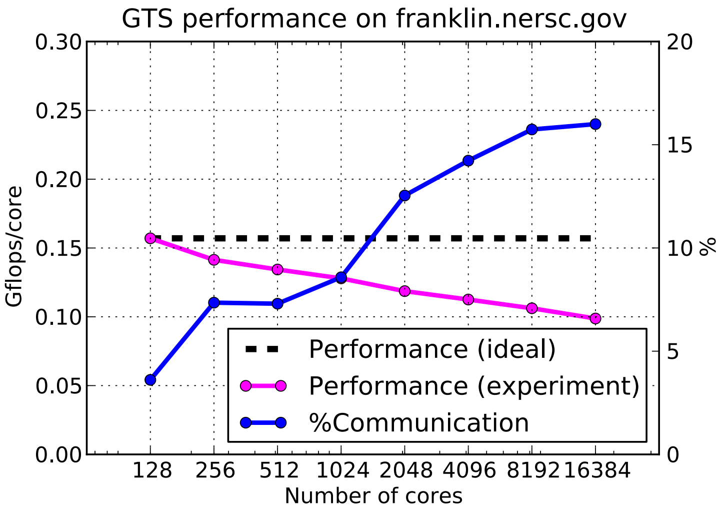
GROUP 1

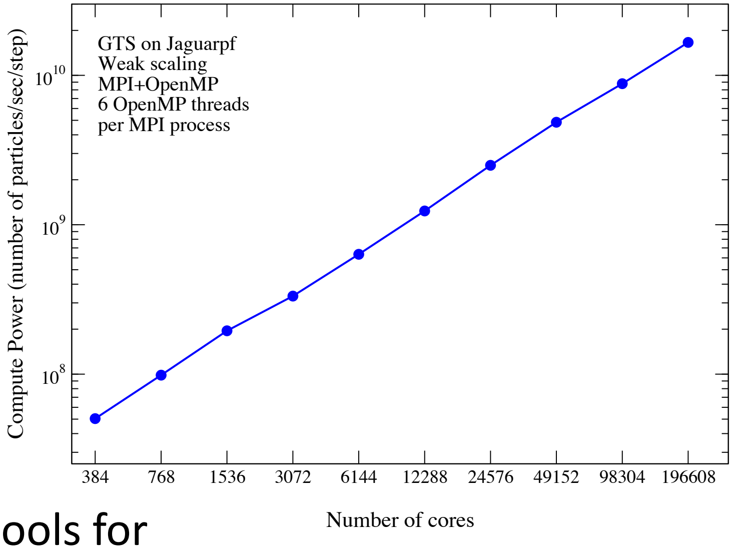


GROUP 2



GROUP 3



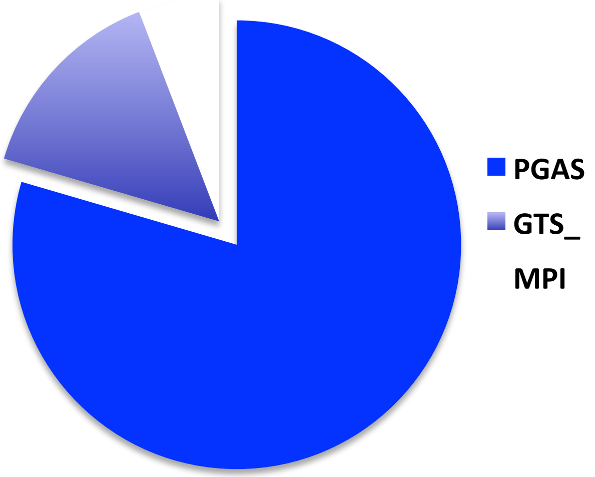


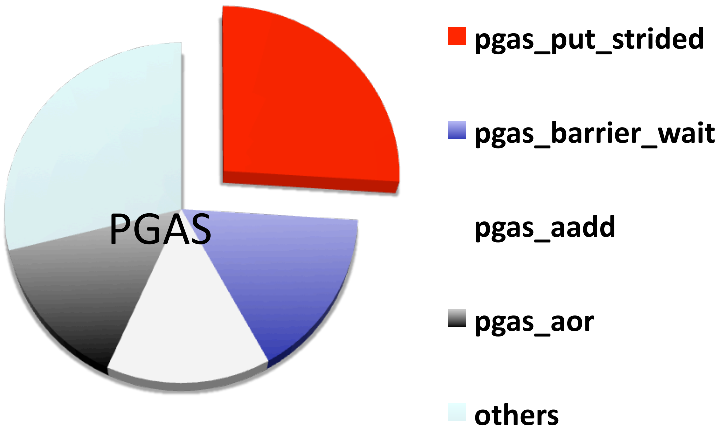
Needs: More standardization of tools, continued tools for scale, wider variety of programming language support

Postdoctoral Researcher: P. Narayanan

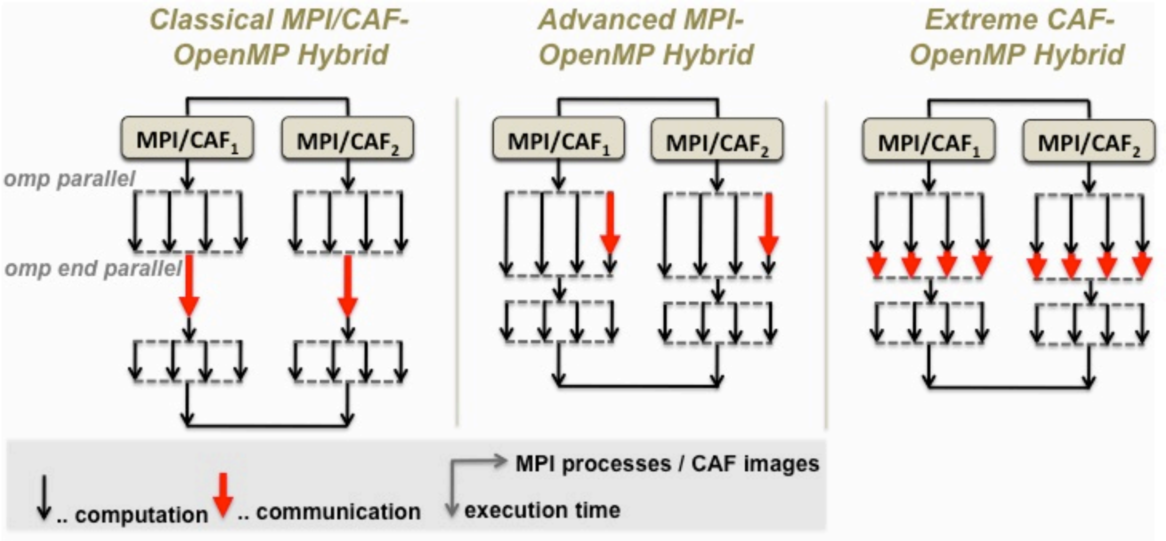
Performance monitoring of CERF codes

- Need performance monitoring tools to support advanced programming models in CERF codes
  - MPI+OpenMP+PGAS
  - MPI generally well supported, OpenMP and PGAS need work
  - PGAS challenging because of one-sided communication model
- Approach:
  - Build code and instrument it to monitor function groups of interest (MPI, CAF, OpenMP)
  - Run instrumented binary and process results
  - Identify potential bottlenecks and optimize code
- Output for sample CAF+MPI code
  - Code spends 80 % of time in CAF, 5 % in MPI
  - Breakup by walltime for individual functions in CAF and MPI available
    - Eg. Pgas_put_strided forms largest PGAS chunk
    - Moniter at high concurrency to catch scaling bottlenecks



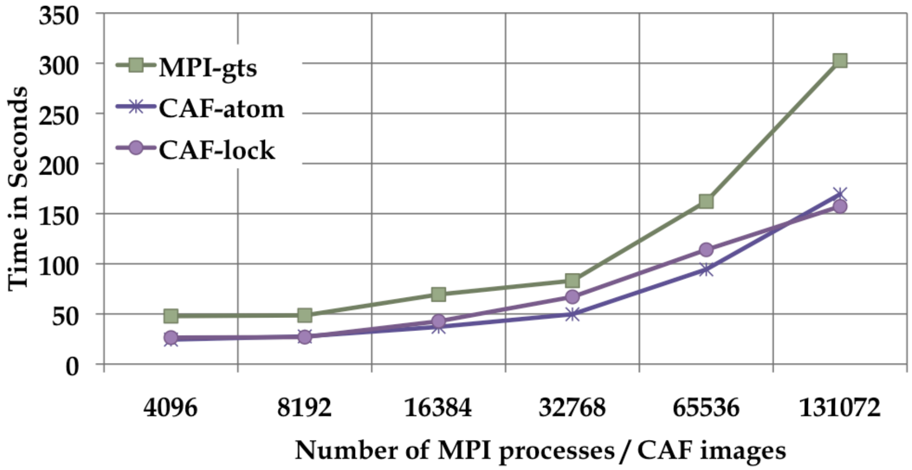


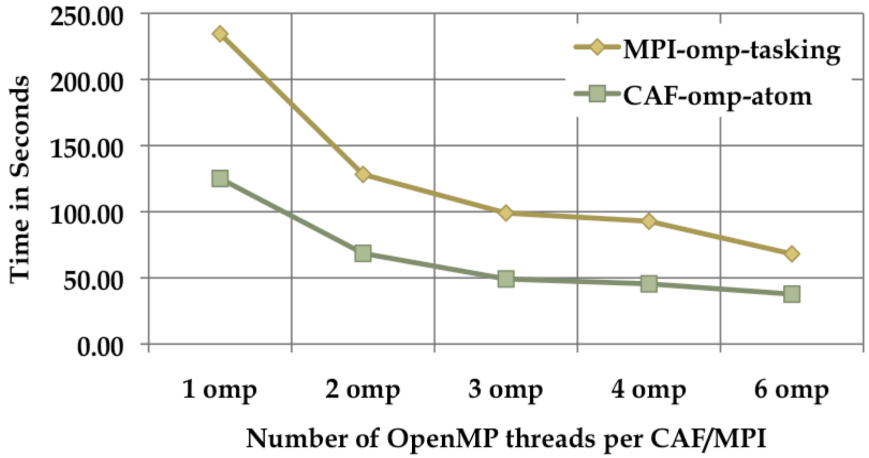
Advanced Programming Model Studies using Co-Array Fortran (CAF) in the GTS code



Classical MPI/CAF-OpenMP Hybrid
Advanced MPI-OpenMP Hybrid
Extreme CAF-OpenMP Hybrid
MPI/CAF$_1$
MPI/CAF$_2$
omp parallel
omp end parallel
.. computation
.. communication
MPI processes / CAF images
execution time

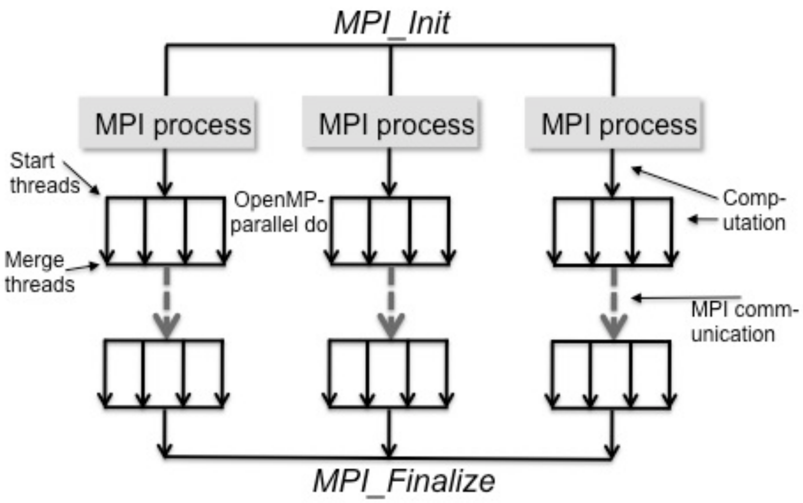
Extend the existing hybrid MPI/OpenMP communication model for better performance and investigate the applicability of new parallel programming models in the communication-intensive part of GTS, a plasma PIC code

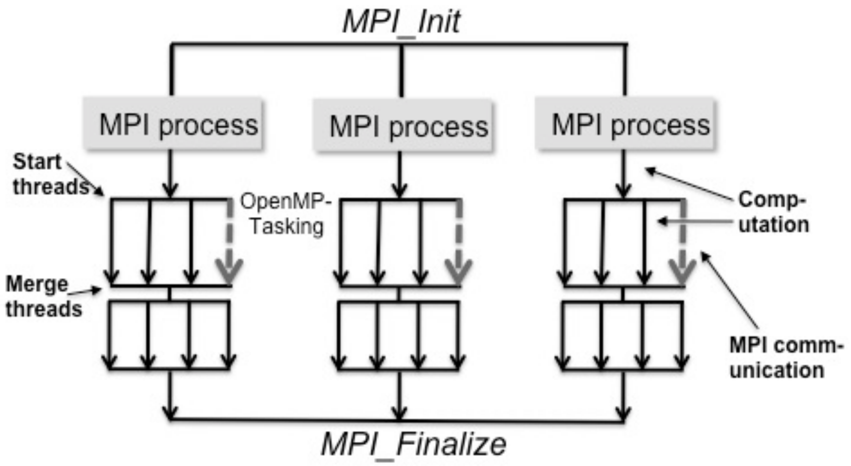




Postdoctoral Researcher: R. Preissl

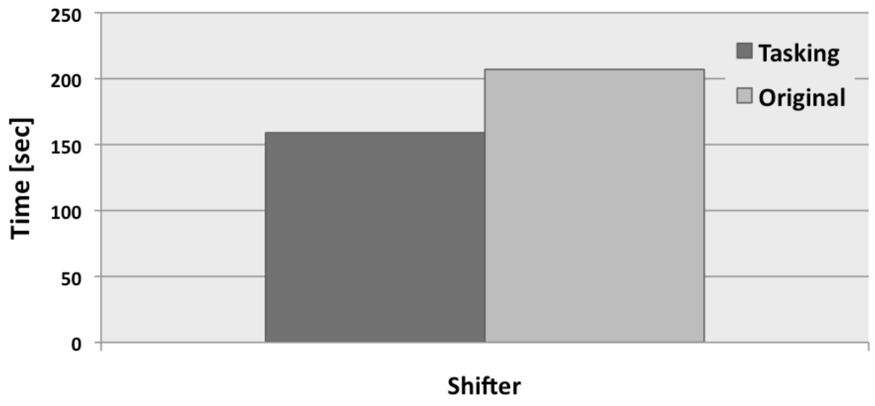
Newer versions of OpenMP include hybrid tasking model used to speed up the GTS particle code





Successfully overlapped MPI communication in the GTS particle shifter routine

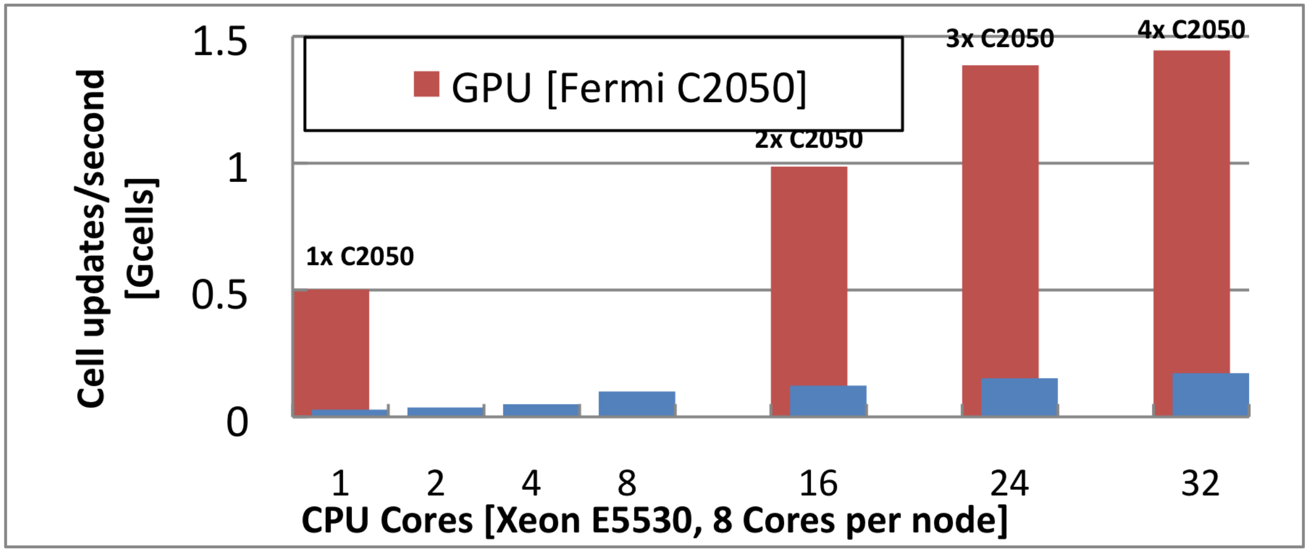
(4 OpenMP threads per MPI process in a 2048 MPI process run on Franklin Cray XT4)



Postdoctoral Researcher: R. Preissl

Preliminary Experiments with CUDA on GPUs for Vorpal PIC Code are promising

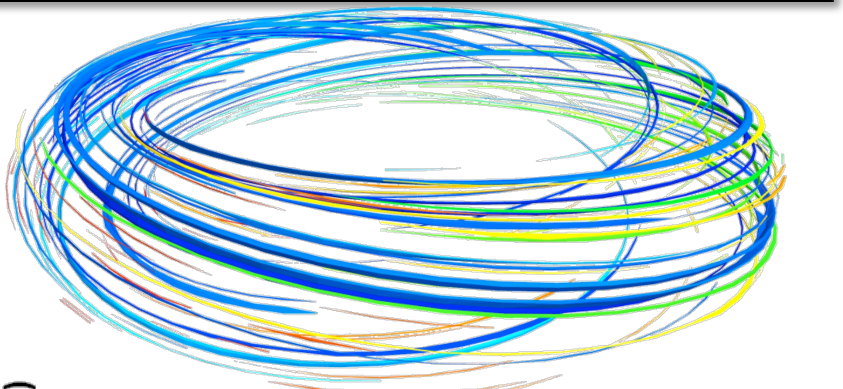
Embedded-boundary, explicit electromagnetics computations using Vorpal 5-8x faster on 1 GPU compared with 8-CPU-core node. GPU power only 1.5x higher.

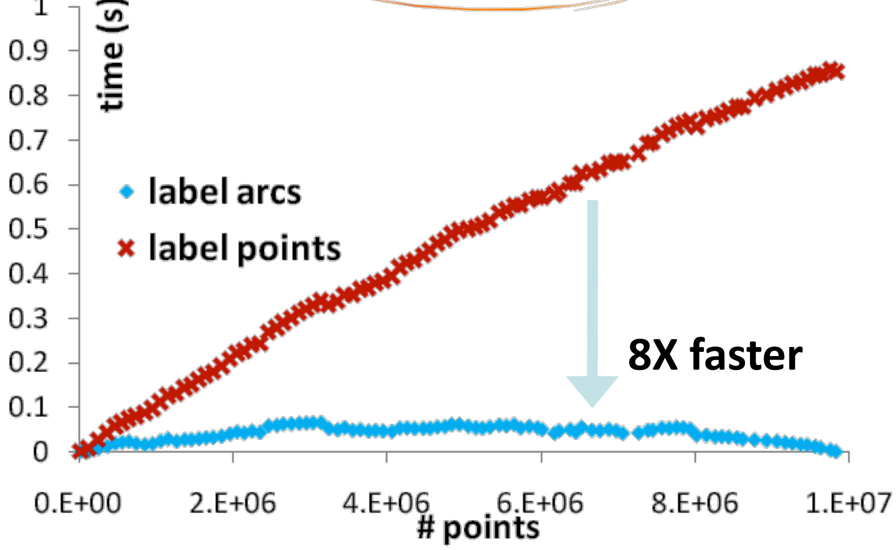


Data management issues towards exascale: Efficient Searching Algorithms

Using FastBit compressed data structures and an application specific coordinate system to significantly reduce the feature identification time

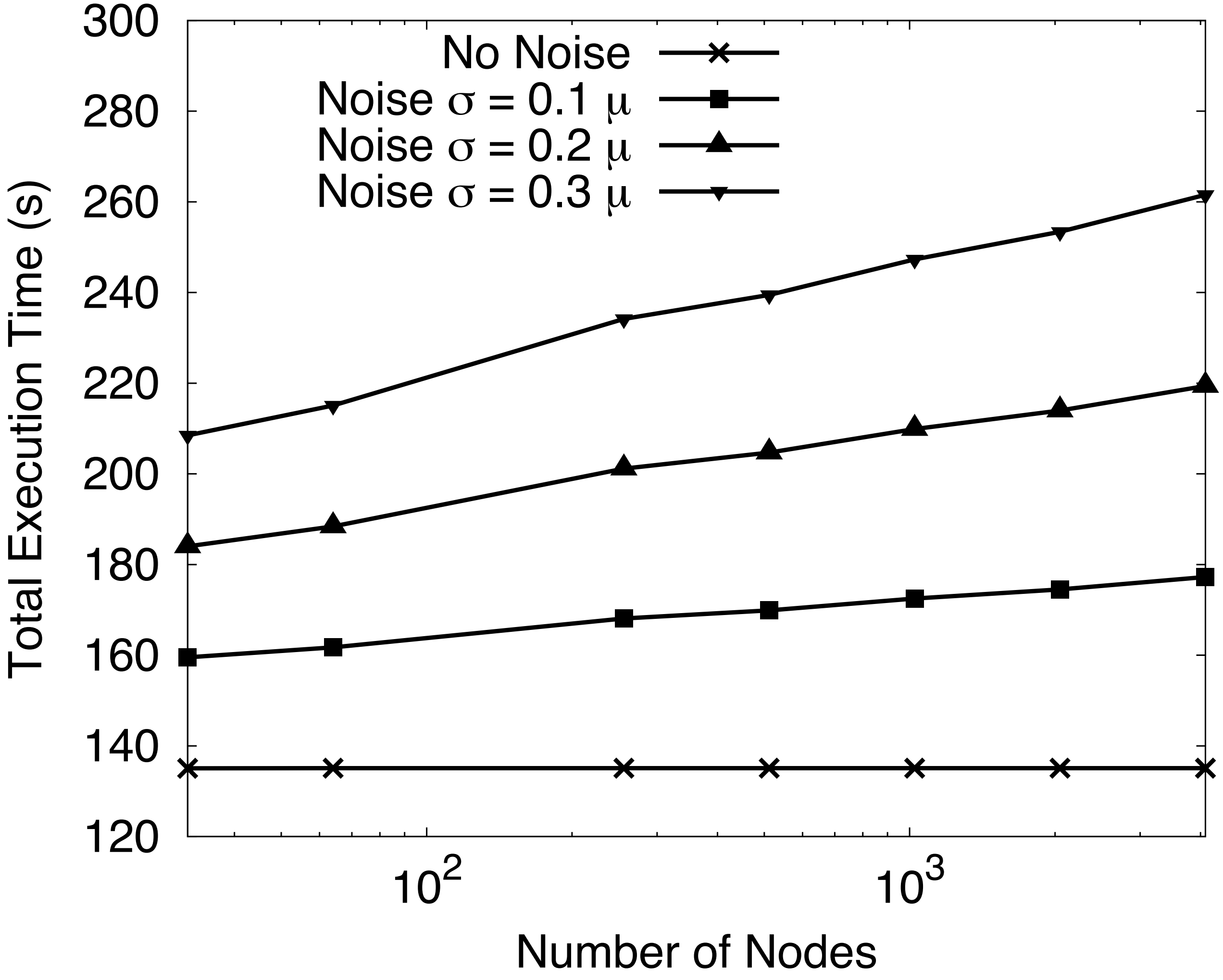
- identify features such as regions with high magnetic potential
- We break the task into two steps: (1) use FastBit index to find all points of interest; (2) assign a unique label to all connected points
- Working with groups of points (arcs) instead of individual points reduces execution time by 8X on average
- Using the magnetic coordinate system to connect the points further reduces the execution time



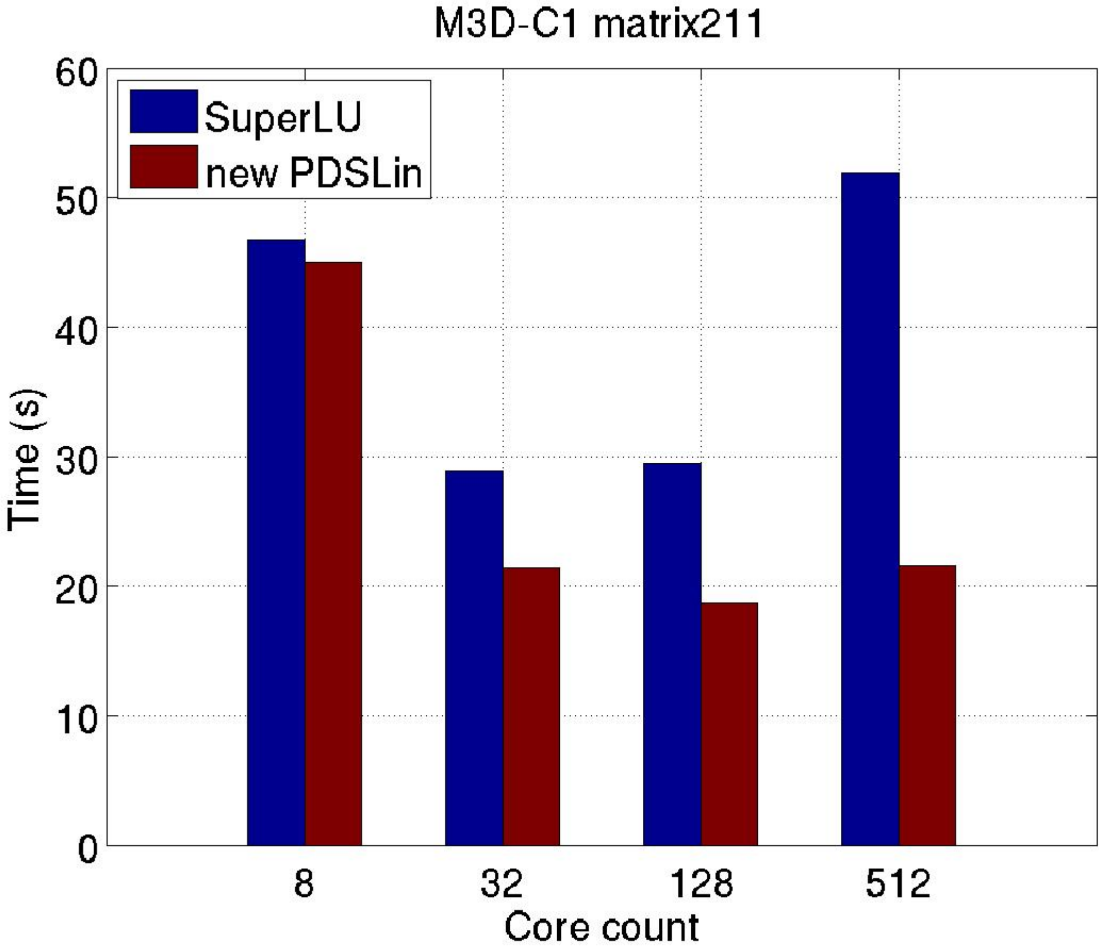


Contact: John Wu, LBNL (kwu@lbl.gov)

Example of how SST/macro can be used to understand performance of a CERF Skeleton Application



Algorithm Improvement being explored as needed – Here: 2D example for a developing CERF app

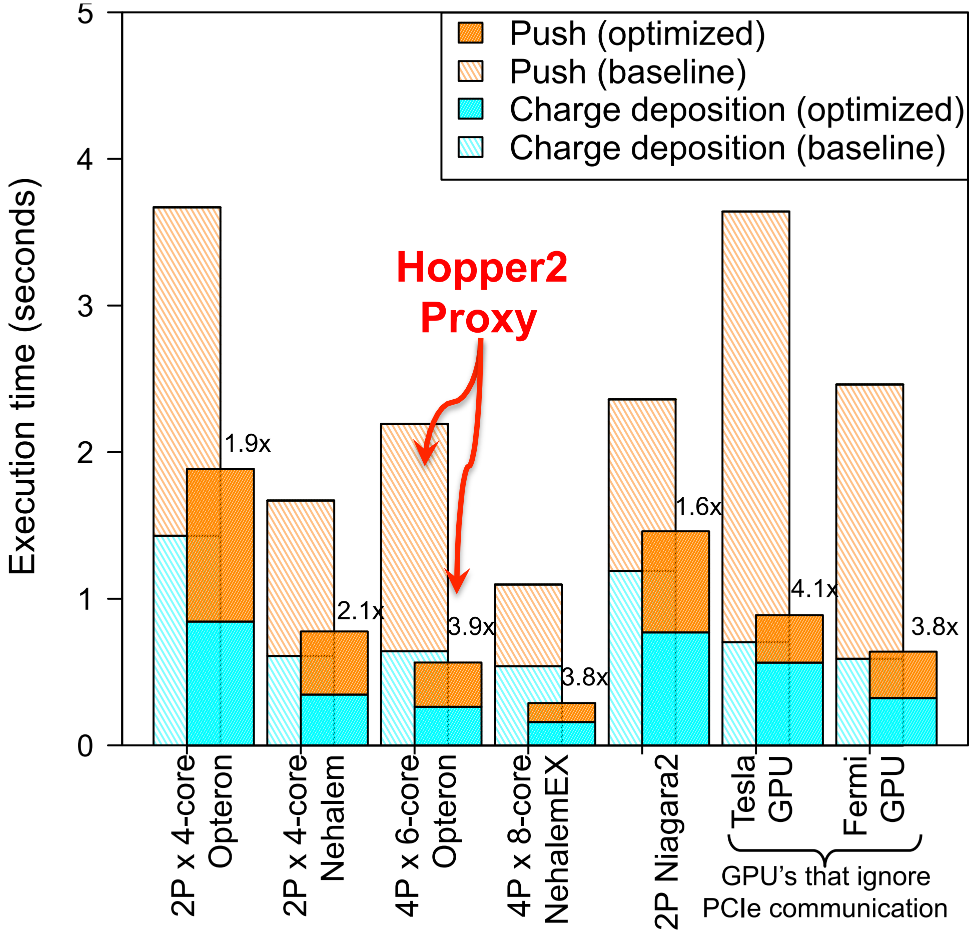


Improvement of the new domain decomposition hybrid solver over direct solver SuperLU for a M3D-C1 linear system. This result is for a 2D case – 800K dimension (since M3D-C1 is being developed), but the solver has been shown to scale on systems of dimension of ~ 60M.

Auto tuning experiments yield 4X potential speed-up in a particle code

- Performance evaluated on CPU's and GPU's
- Defined and explored a large optimization space including threading and floating-point atomics.

- Observed substantial speedups on all machines via optimization (nearly 4x on Hopper2 proxy)
- Although speedups on GPU's were larger, GPU performance and energy efficiency is at best comparable to top-of-the-line CPUs.


Hopper2 Proxy

IP issues with Vendor Interaction

- **Concern about IP issues**
  - Details of vendor architecture are closely guarded secrets
    - Its not co-design unless we can have a detailed 2-way conversation
    - But vendors cannot afford to have information bleed even accidentally through codesign interactions
  - Suggestions from Vendors
    - Have suggested having entire CoDesign teams bound to one "association"
    - Subset of team members dedicated to single vendor
    - One vendor has suggested there would be no issue if all exascale systems adopt their design
- **Open (non-proprietary) simulation capability for target architectures to support safe multi-vendor interaction for co-design cycle**
  - Create model that is structurally similar to target, but sufficiently different to not expose vendor IP
  - Provide concrete advice to vendors about impact of hardware changes on algorithm performance during concept phase of design cycle
    - *Enables rapid exploration of hardware/software design trade-offs that are cross-platform issues*
    - Vendor simulators come available later, when we can do fewer changes
    - We can shift to vendor-specific optimizations when vendor sims arrive
  - *Provide air-gap for sensitive vendor IP using proxy model*
- CERF also depends on advice from vendors on policy board for guidance (prior to selection of exascale associations)