Using Parallel Mesh Partitioning Strategies to Improve the Performance of Tau3P, an Electromagnetic Field Solver

Michael M. Wolf, University of Illinois, Urbana-Champaign; Ali Pinar, Lawrence Berkeley National Laboratory; Karen D. Devine, Sandia National Laboratories; Adam Guetz, Nate Folwell and Kwok Ko, Stanford Linear Accelerator Center

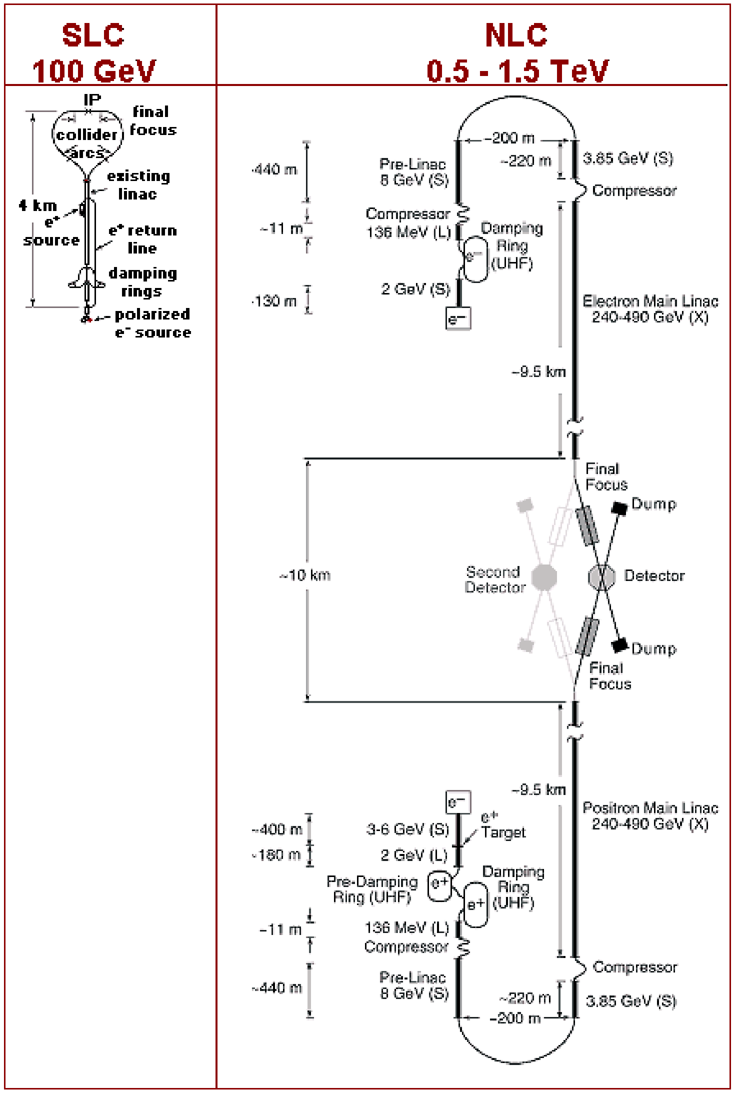
Outline

- Motivation
- Brief Description of Tau3P
- Tau3P Performance
- Partitioning Results
- Port Grouping
- Future Work

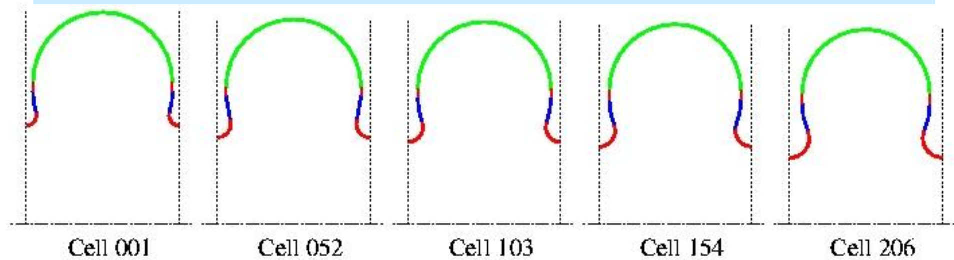
Challenges in E&M Modeling of Accelerators

- Accurate modeling essential for modern accelerator design
  - Reduces Design Cost
  - Reduces Design Cycle
- Conformal meshes (Unstructured grid)
- Large, complex electromagnetic structures
  - 100's of millions of DOFs
- Small beam size
  - Large number of mesh points
  - Long run time
- Parallel Computing needed (time and storage)

Next Linear Collider (NLC)



Cell to cell variation of order microns to suppress short range wakes by detuning

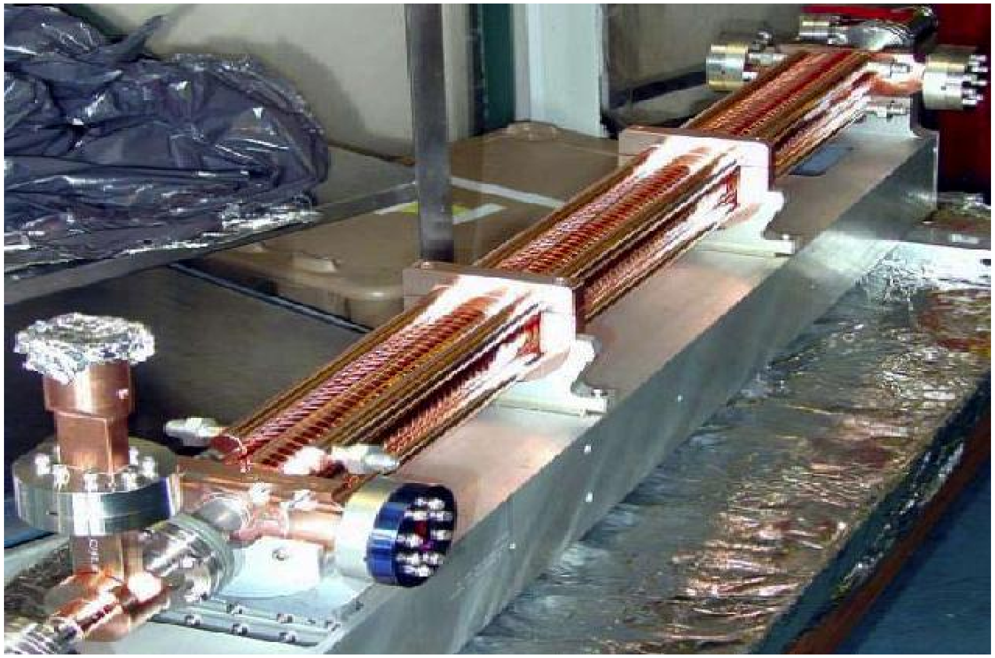


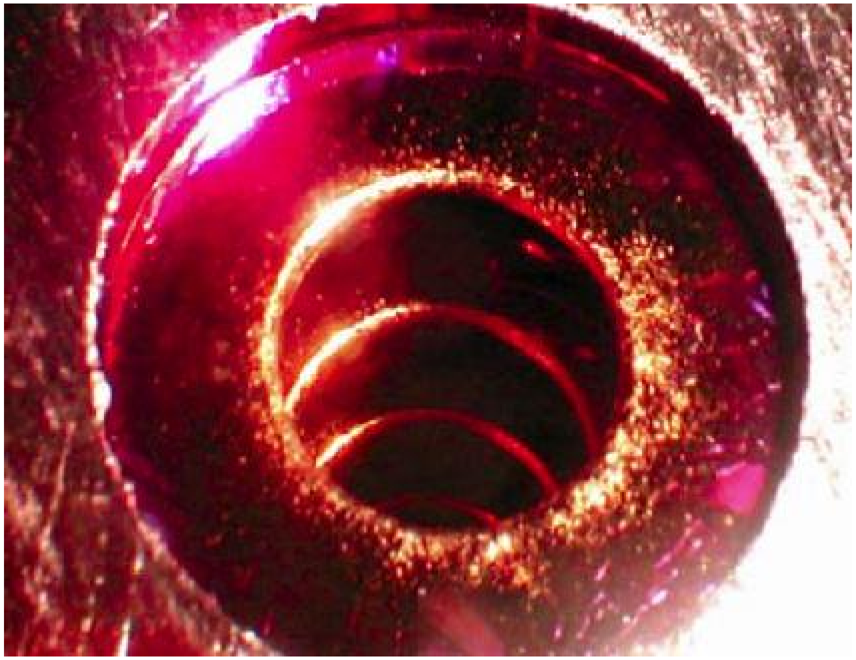
ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

SLAC

End-to-end NLC Structure Simulation

- NLC X-band structure showing damage in the structure cells after high power test
- Theoretical understanding of underlying processes lacking so realistic simulation is needed



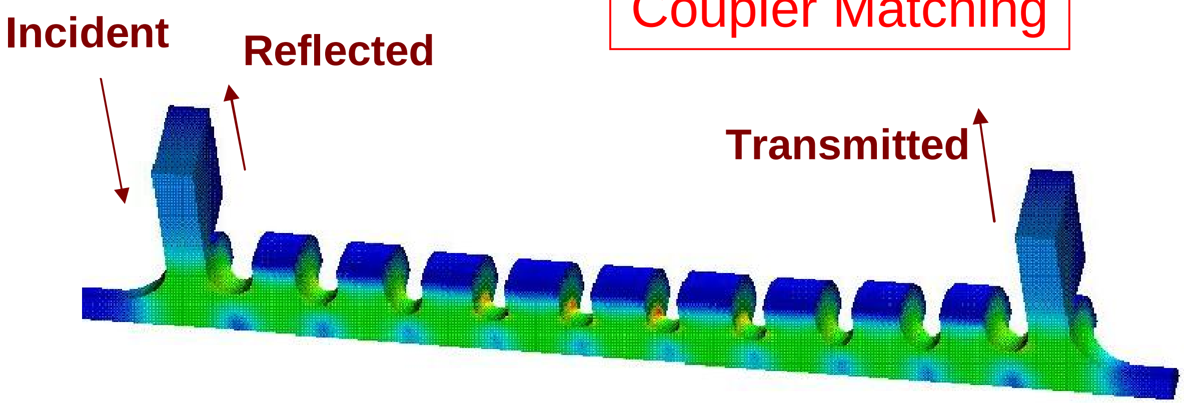


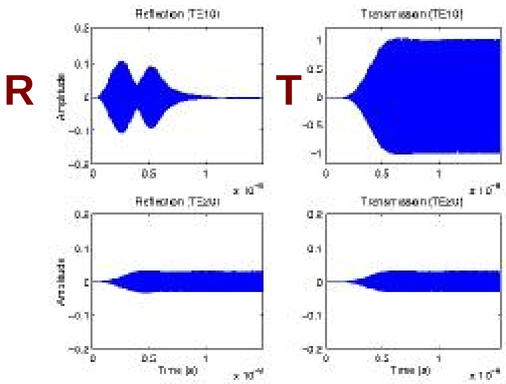
ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

SLAC

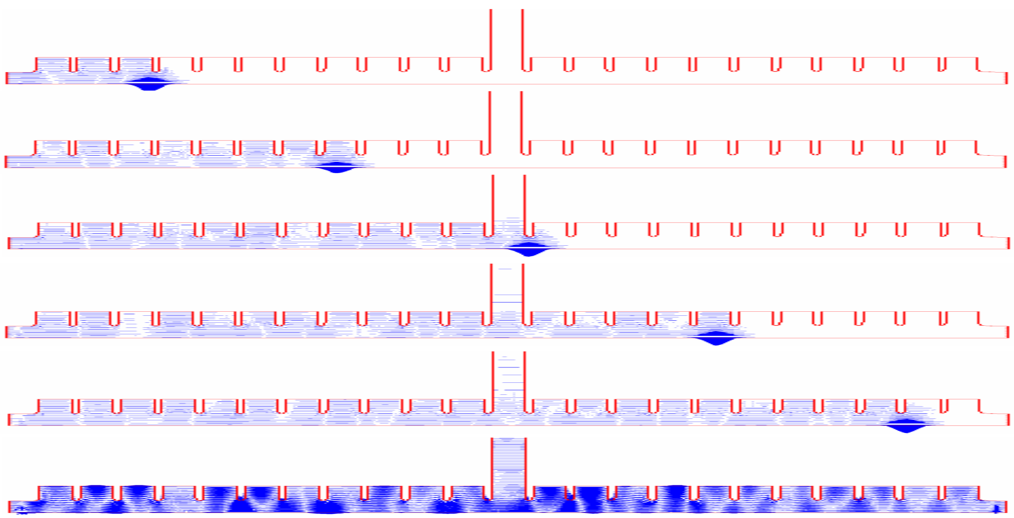
Parallel Time-Domain Field Solver – Tau3P

Coupler Matching

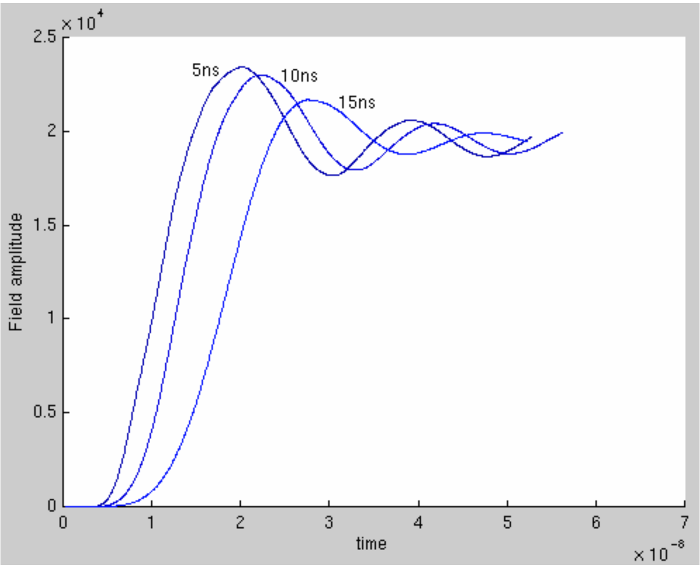




Wakefield Calculations



Rise Time Effects

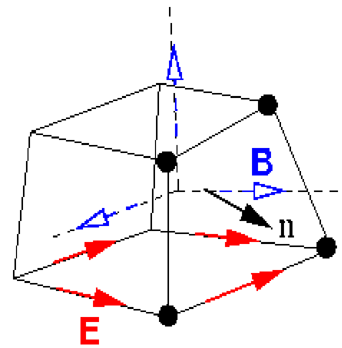


Parallel Time-Domain Field Solver – Tau3P

- Follows evolution of E and H fields inside accelerator cavity
- DSI method on non-orthogonal meshes

$$\oint E \bullet ds = -\iint \frac{\partial B}{\partial t} \bullet dA$$

$$\oint H \bullet ds^* = \iint \frac{\partial D}{\partial t} \bullet dA^* + \iint j \bullet dA^*$$



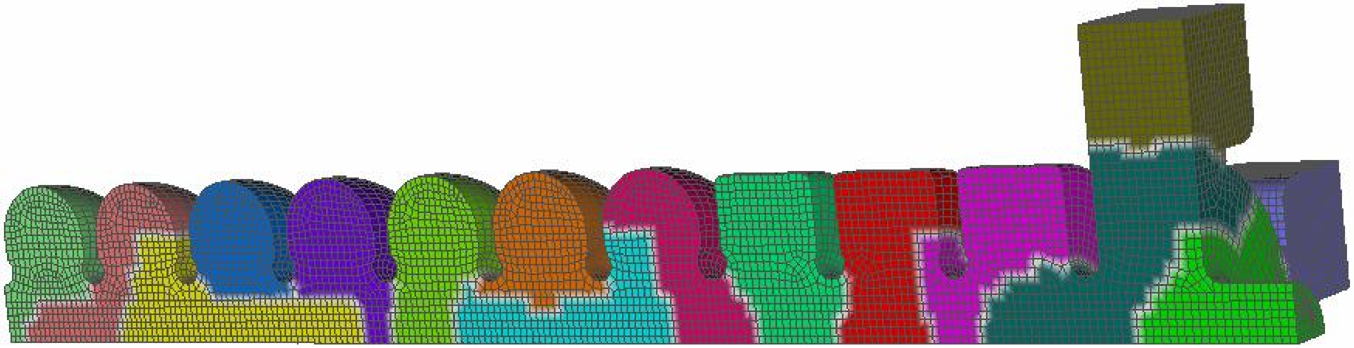
The DSI formulation yields:

$$\vec{e} += a \cdot A_H \cdot \vec{h}$$

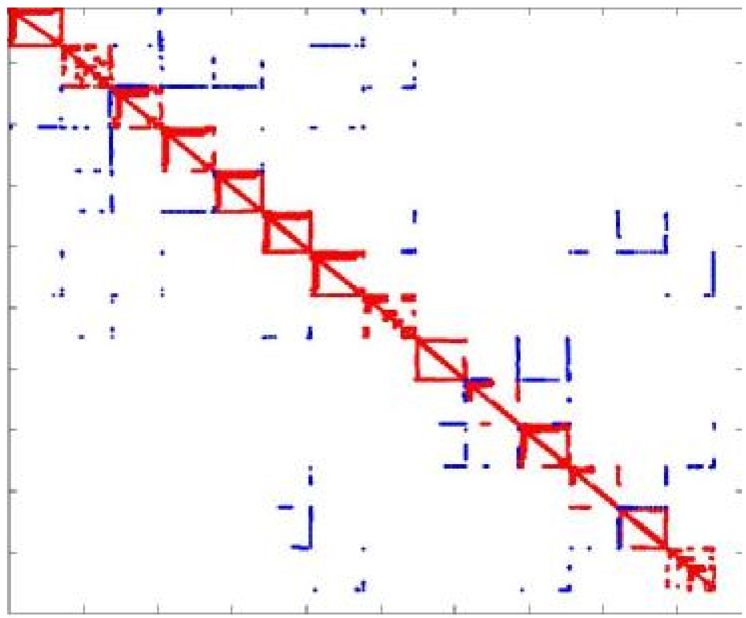
$$\vec{h} += b \cdot A_E \cdot \vec{e}$$

- α, β are constants proportional to dt
- A_H,A_E are matrices
- Electric fields on primary grid
- Magnetic fields on embedded dual grid
- Leapfrog time advancement
- (FDTD) for orthogonal grids

Tau3P Implementation



Example of Distributed Mesh



Typical Distributed Matrix

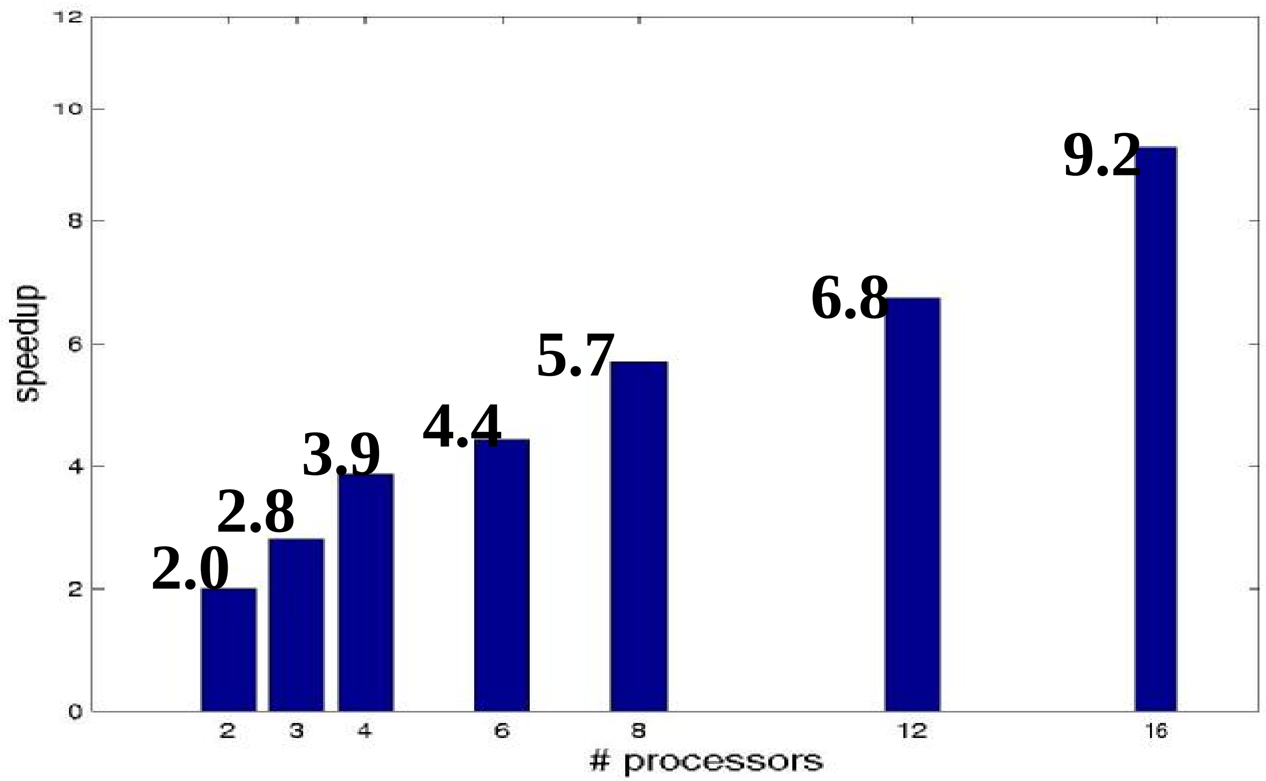
- Very Sparse Matrices
  - 4-20 nonzeros per row
- 2 Coupled Matrices (A_H,A_E)
- Nonsymmetric (Rectangular)

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Parallel Performance of Tau3P (ParMETIS)

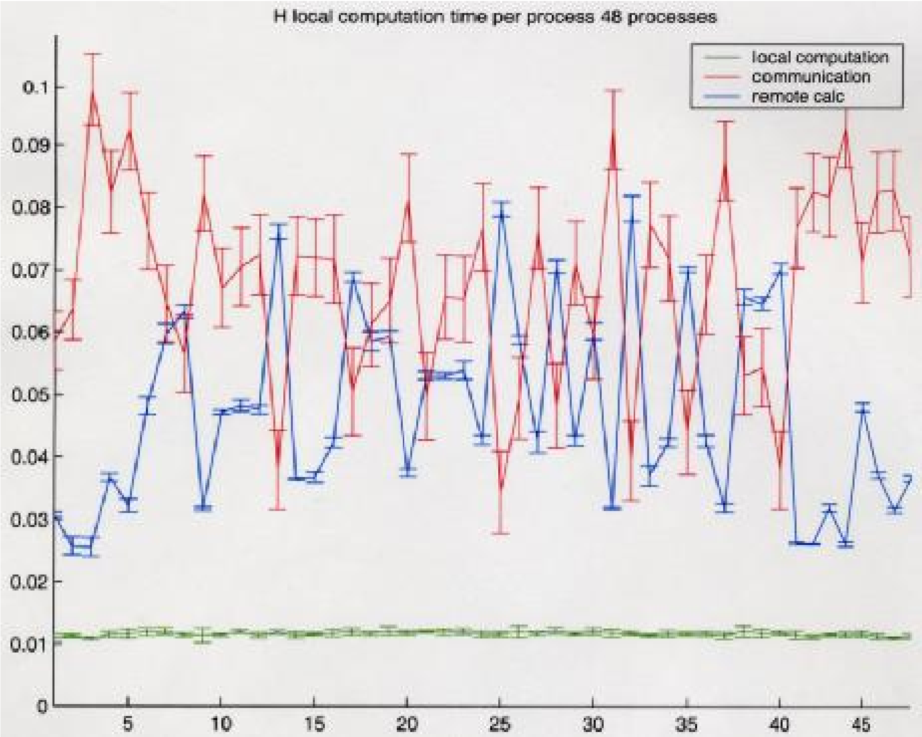
- 257K hexahedrons
- 11.4 million non-zeroes

Parallel Speedup

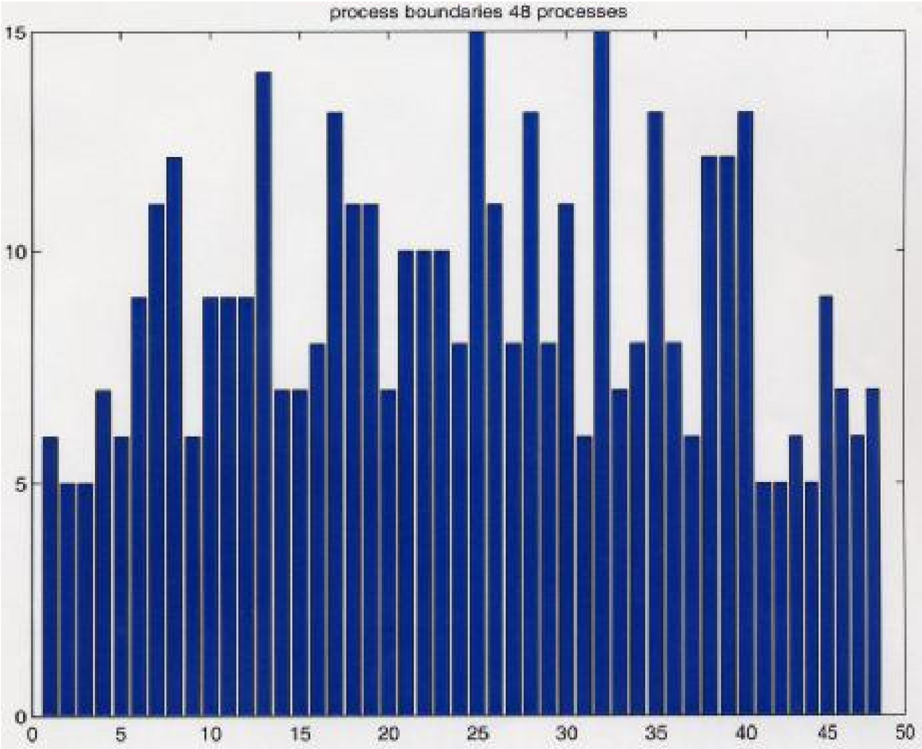


Communication in Tau3P (ParMETIS Partitioning)

Communication vs. Computation



Process Boundaries



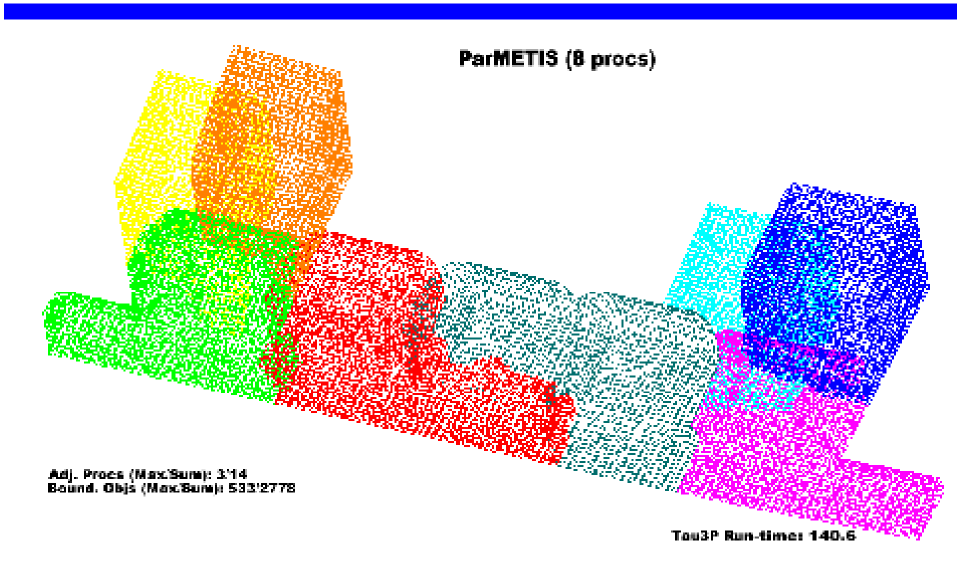
Flexibility in Tau3P Mesh Partitioning

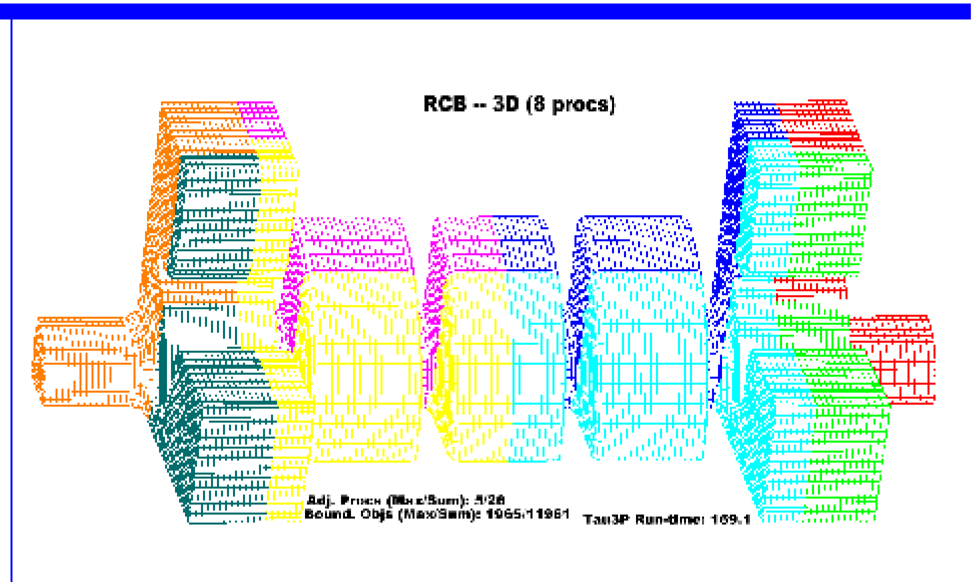
- Long simulation times
  - Tens of thousands of CPU hours
- Problem initialization short
- Most time spent in time advancement
  - Millions of time steps
- Static mesh partitioning
- Willing to pay HIGH price upfront for increased performance of solver

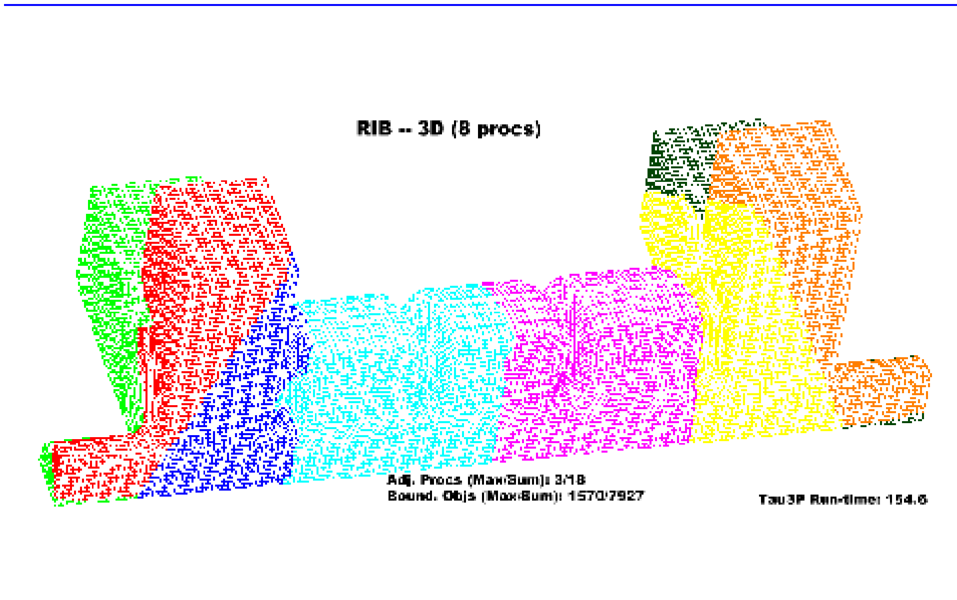
Partitioning Methods

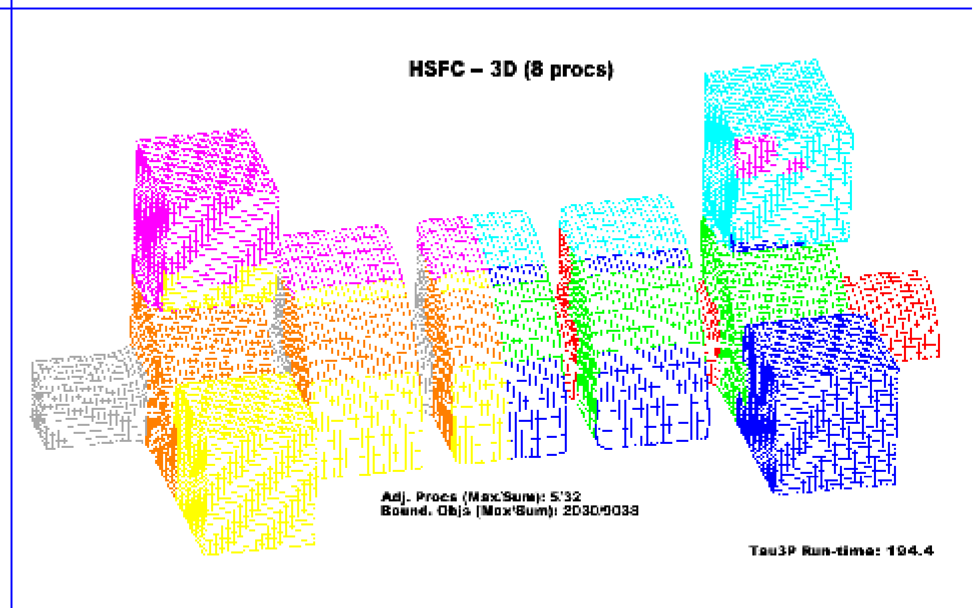
- Using Zoltan (Sandia National Laboratory)
- Tried Several Mesh Partitioning Methods:
  - Graph Partitioning Algorithms
    - ParMETIS
  - Geometric Partitioning Algorithms (1D/2D/3D)
    - Recursive Coordinate Bisection (RCB)
    - Recursive Inertial Bisection (RIB)
    - Hilbert Space-Filling Curve (HSFC)

Several Partitioning Methods



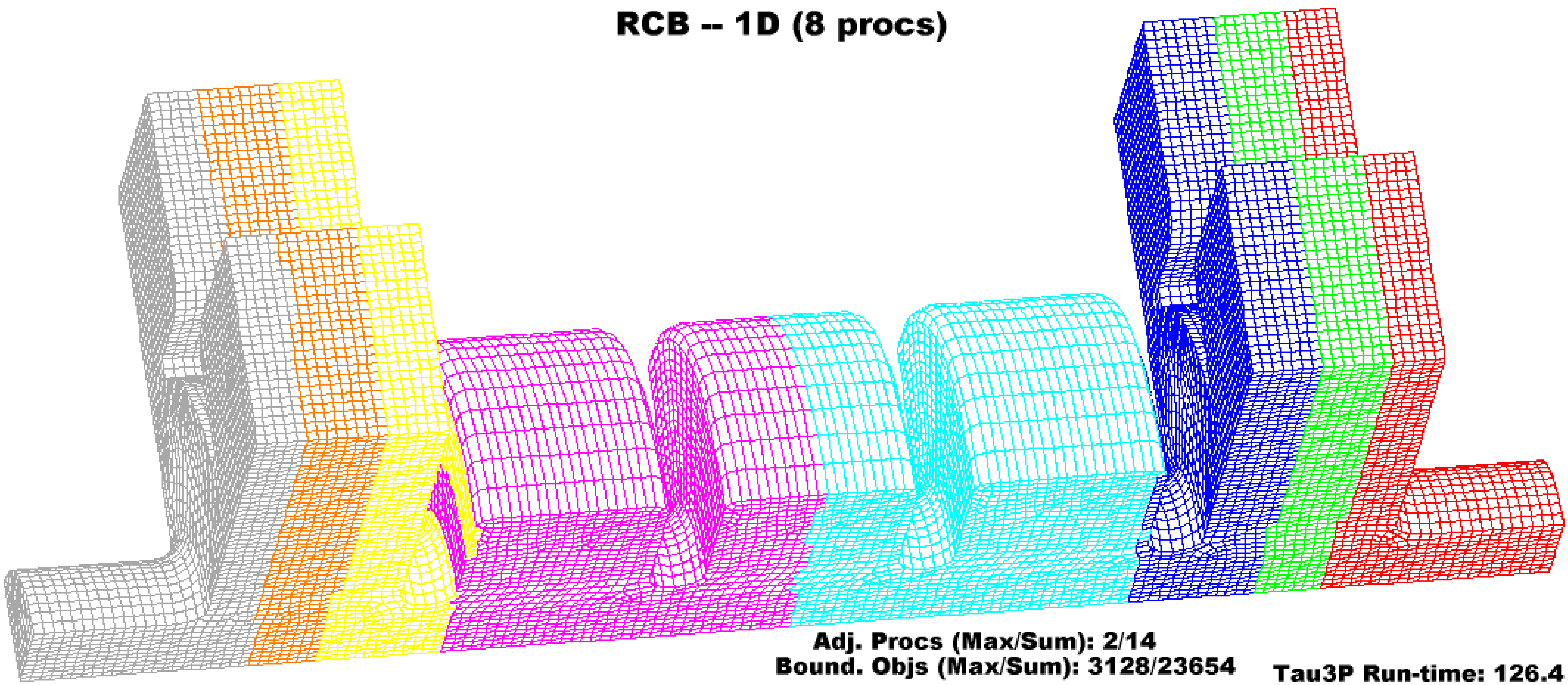






ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

1D (RCB-1D(z)) Partitioning



5 Cell RDDS (8 processors) Partitioning

	Tau3P Runtime	Max Adj. Procs	Sum Adj. Procs	Max Bound. Objs	Sum Bound. Objs
ParMETIS	288.5 s	3	14	585	2909
RCB-1D (z)	218.5 s	2	14	3128	14363
RCB-3D	343.0 s	5	26	1965	11961
RIB-3D	282.4 s	3	18	1570	7927
HSFC-3D	387.3 s	5	32	2030	9038

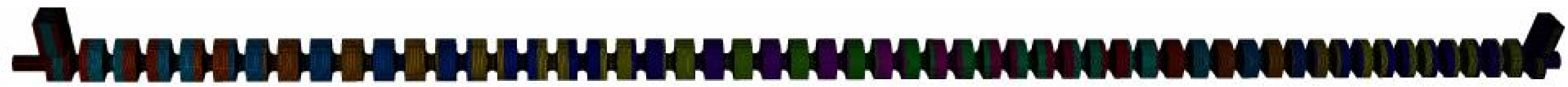
2.0 ns runtime
IBM SP3 (NERSC)

5 Cell RDDS (32 processors) Partitioning

	Tau3P Runtime	Max Adj. Procs	Sum Adj. Procs	Max Bound. Objs	Sum Bound. Objs
ParMETIS	165.5 s	8	134	731	16405
RCB-1D (z)	67.7 s	3	66	2683	63510
RCB-3D	373.2 s	10	208	1404	24321
RIB-3D	266.8 s	8	162	808	20156
HSFC-3D	272.2 s	10	202	1279	26684

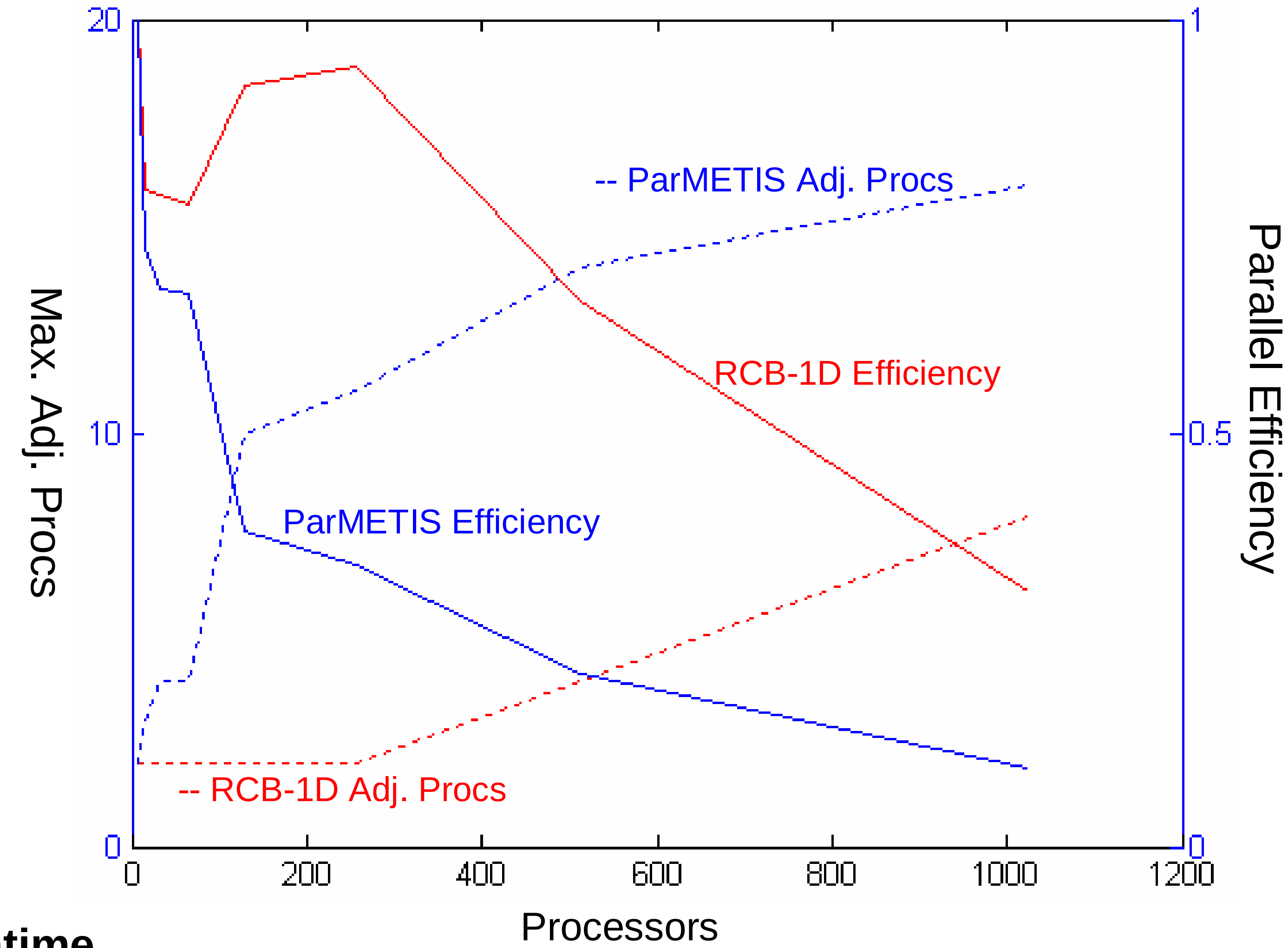
2.0 ns runtime
IBM SP3 (NERSC)

H60VG3 (“real” structure)



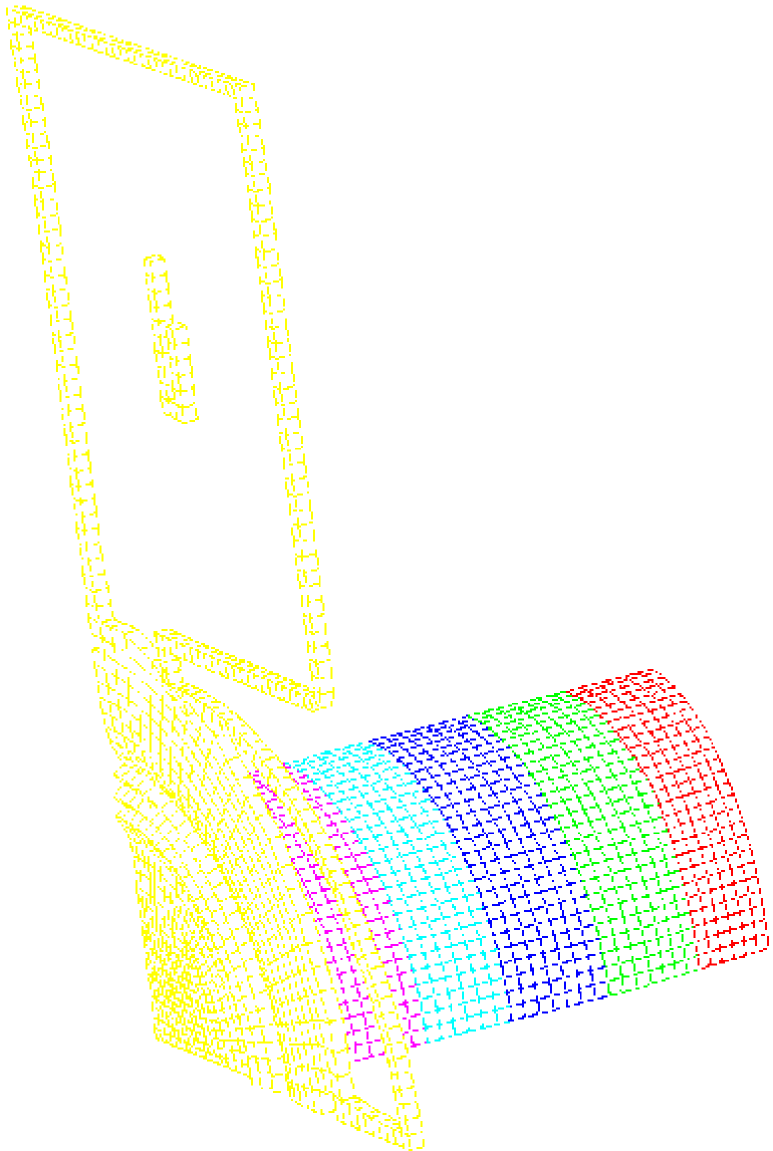
55 cells (w/ coupler)
1,122,445 elements

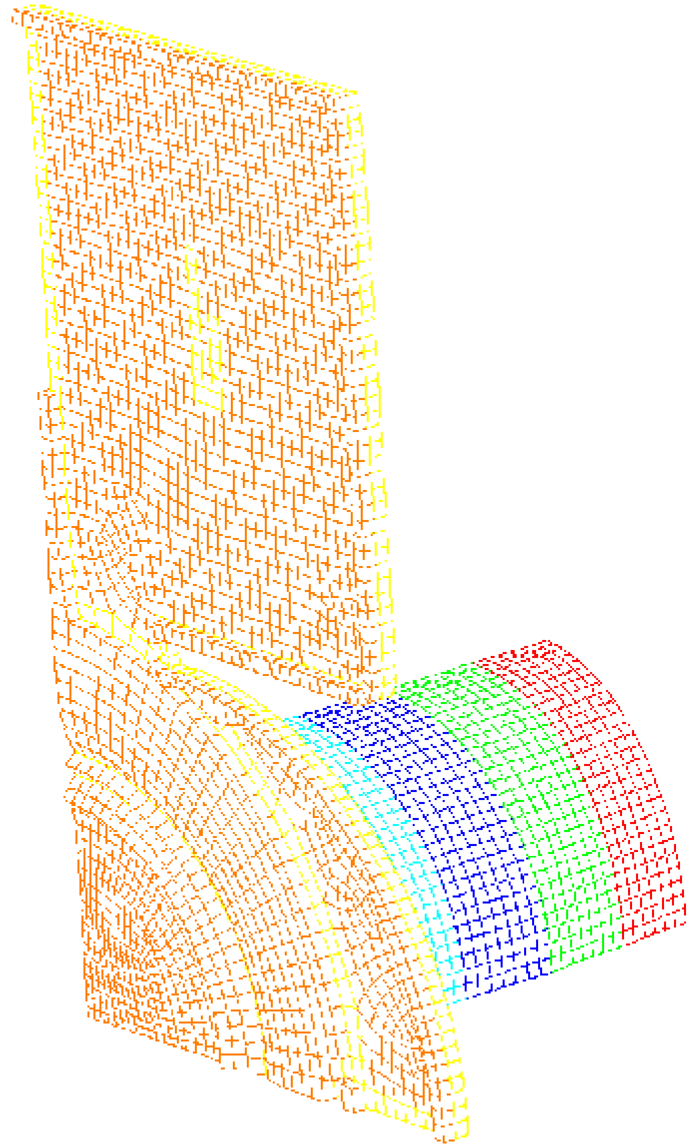
H60VG3 RDDS Partitioning (w/o port grouping)



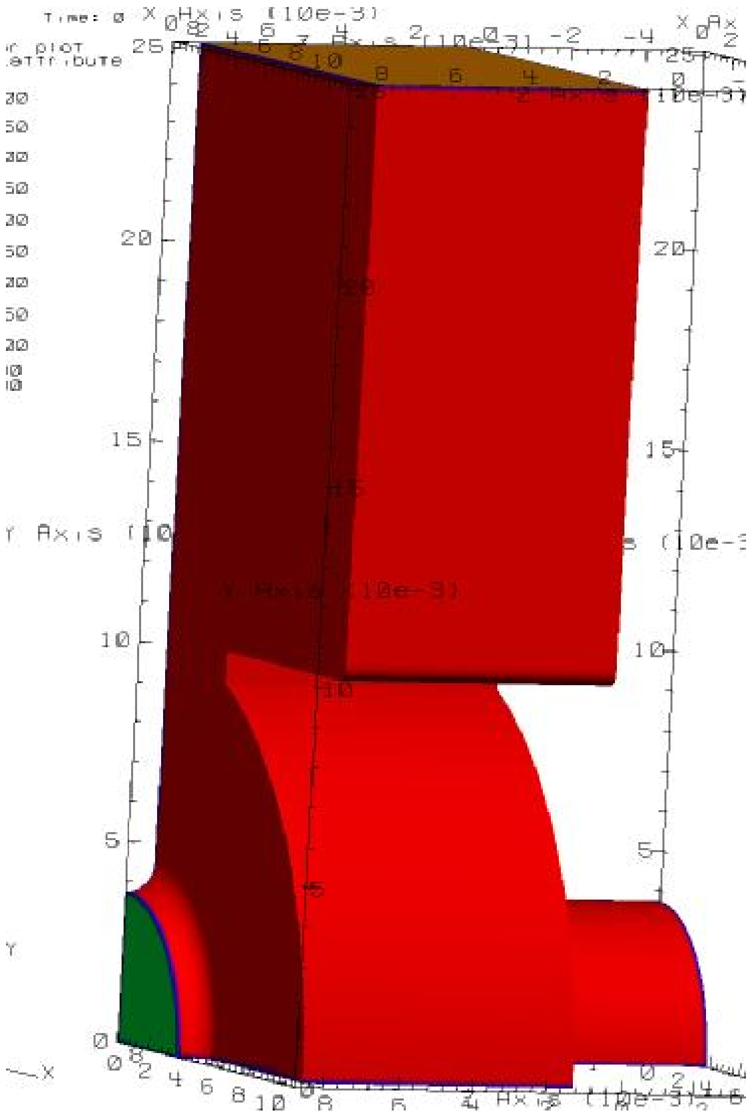
1.0 ns runtime
IBM SP3 (NERSC)

RCB-1D Scalability Leveling Off

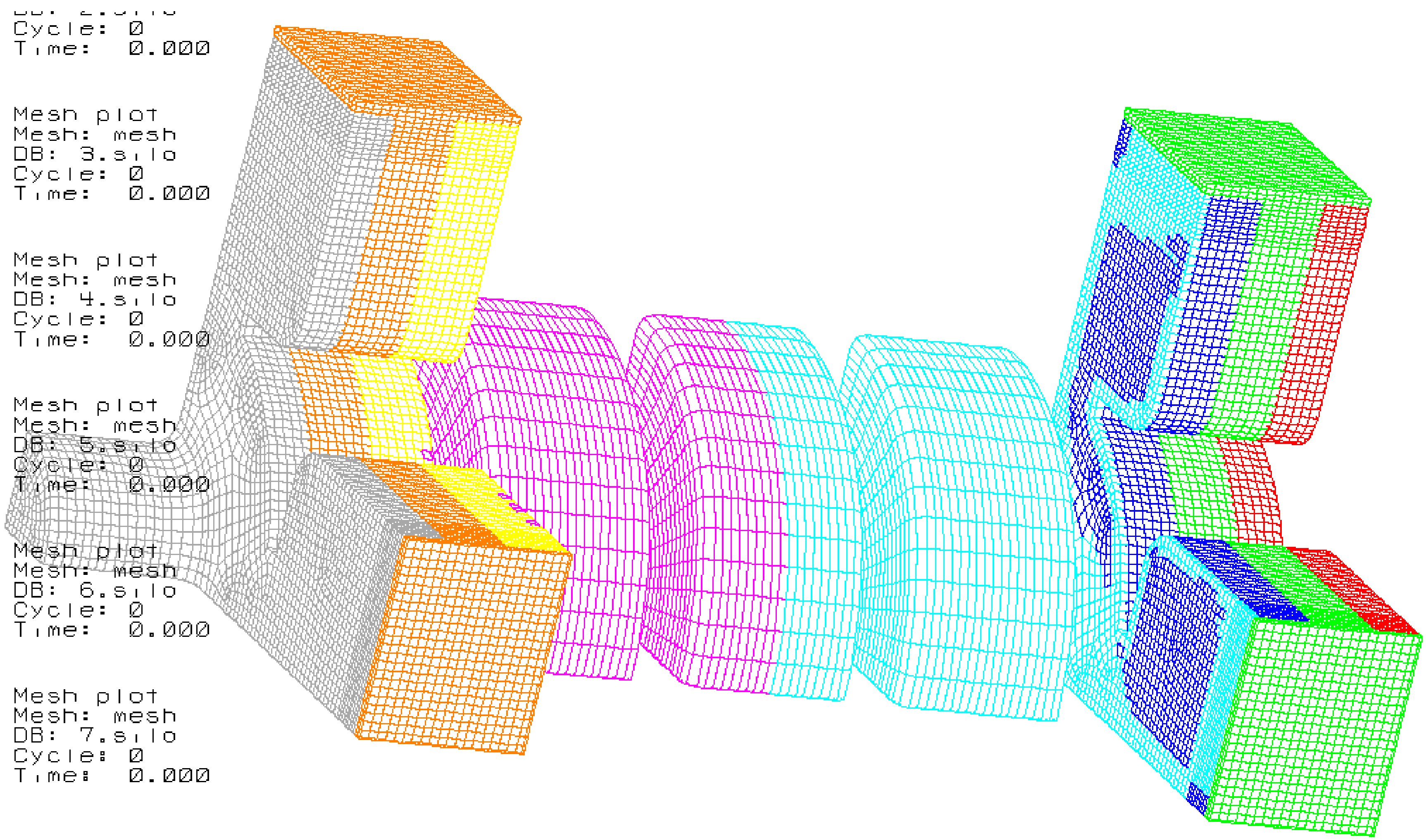




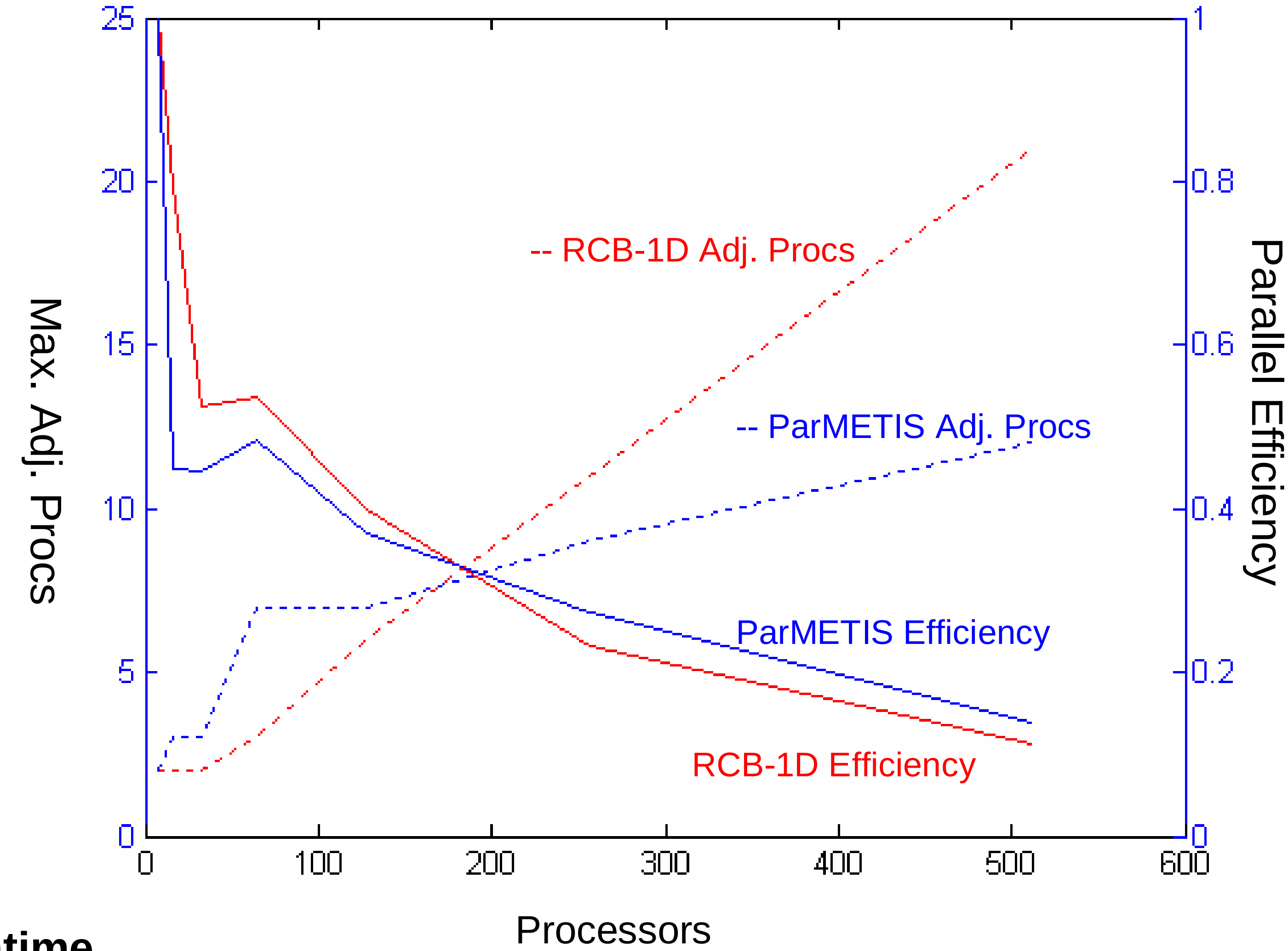
Coupler Port Grouping Complication



Coupler Port Grouping Complication

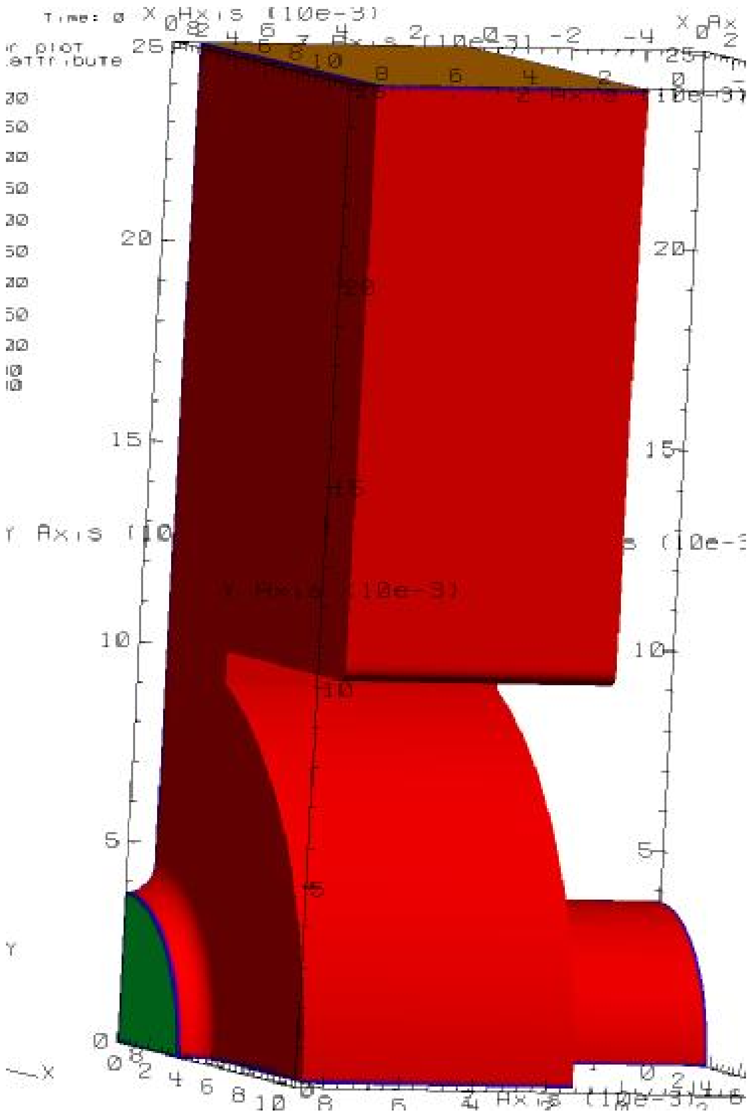


H60VG3 RDDS Partitioning (w/ coupler port grouping)



1.0 ns runtime
IBM SP3 (NERSC)

Constrained Mesh Partitioning



RDDS Coupler Cell Constrained Partition (16 procs)

Method	Max Adj. Procs
HSFC-3D	14
ParMETIS	8
RCB-1D-z	14
RCB-2D-xy	5
RCB-2D-xz	14
RCB-2D-yz	6
RCB-3D	8
RIB-2D-xy	6
RIB-2D-xz	14
RIB-2D-yz	5
RIB-3D	7

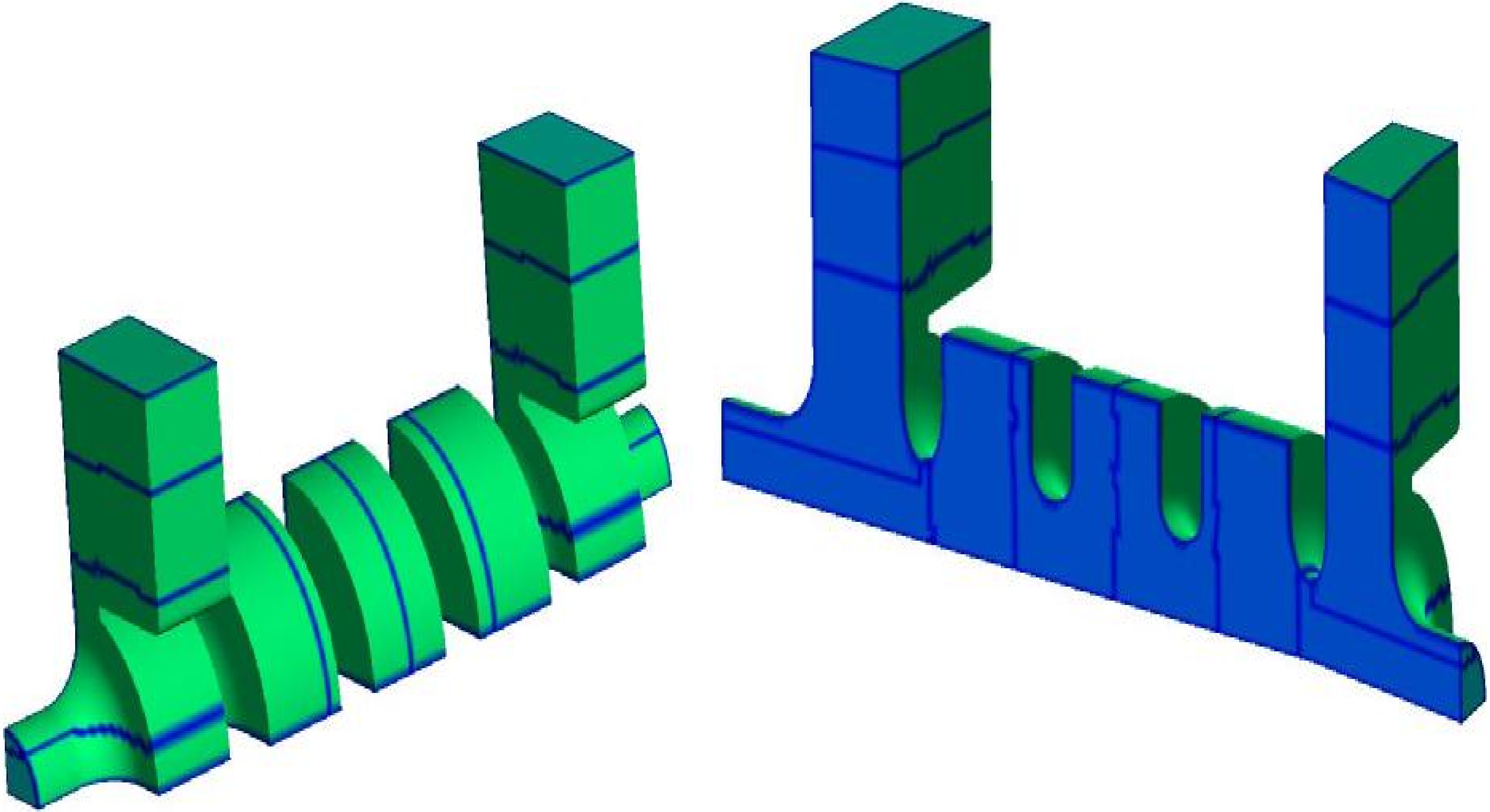
RDDS Coupler Cell Constrained Partition (32 procs)

Method	Max Adj. Procs
HSFC-3D	17
ParMETIS	14
RCB-1D-z	29
RCB-2D-xy	7
RCB-2D-xz	29
RCB-2D-yz	7
RCB-3D	11
RIB-2D-xy	7
RIB-2D-xz	29
RIB-2D-yz	6
RIB-3D	12

"U" Partitioning (Ongoing)

- RCB 1D Partitioning
- Remap coordinates
- Partition based on distance from curve.

“U” Partitioning (Ongoing)



Future Work

- “U” or “Fork” Partitioning
- Stitching Multiple Partitions Together
- Method Competition
- Connectivity into geometric methods
- Local partitioning
- Other methods

Acknowledgements

- SNL
  - Karen Devine, et al.
- LBNL
  - Ali Pinar
- SLAC (ACD)
  - Adam Guetz, Cho Ng, Nate Folwell, Lixin Ge, Greg Schussman, Kwok Ko

- Work supported by U.S. DOE ASCR & HENP Divisions under contract DE-AC0376SF00515

Load Balancing in Tau3P

- Load balancing inTau3P (NLC Input Coupler)
  - Unstructured meshes lead to matrices for which nonzero entries are not evenly distributed.
  - Complicates work assignment and load balancing in a parallel setting.
  - Originally used ParMETIS

NZ Distribution over 14 cpu's