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
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
Parallel Time-Domain Field Solver - Tau3P

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

\[ \oint E \cdot ds = -\iint \frac{dB}{dt} \cdot dA \]
\[ \oint H \cdot ds^* = \iint \frac{dD}{dt} \cdot dA^* + \iint j \cdot dA^* \]

The DSI formulation yields:

\[ \nabla e + = \alpha \cdot A_H \cdot \mathbf{h} \]
\[ \nabla h + = \beta \cdot A_E \cdot e \]

- \( \alpha, \beta \) 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)
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

ParMETIS (8 procs)
Adj. Procs (Max Sum): 3718
Bound. Obj. (Max Sum): 56707927
Tu3P Run-time: 140.6

RCB - 3D (8 procs)
Adj. Procs (Max Sum): 3718
Bound. Obj. (Max Sum): 56707927
Tu3P Run-time: 109.1

RIB - 3D (8 procs)
Adj. Procs (Max Sum): 3718
Bound. Obj. (Max Sum): 15707927
Tu3P Run-time: 154.6

HSFC - 3D (8 procs)
Adj. Procs (Max Sum): 532
Bound. Obj. (Max Sum): 20309238
Tu3P Run-time: 104.4
1D (RCB-1D(z)) Partitioning

RCB – 1D (8 procs)

Adj. Procs (Max/Sum): 2/14
Bound. Objs (Max/Sum): 3128/23654
Tau3P Run-time: 126.4
### 5 Cell RDDS (8 processors) Partitioning

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

2.0 ns runtime  
IBM SP3 (NERSC)
## 5 Cell RDDS (32 processors) Partitioning

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

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)

<table>
<thead>
<tr>
<th>Method</th>
<th>Max Adj. Procs</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSFC-3D</td>
<td>14</td>
</tr>
<tr>
<td>ParMETIS</td>
<td>8</td>
</tr>
<tr>
<td>RCB-1D-z</td>
<td>14</td>
</tr>
<tr>
<td>RCB-2D-xy</td>
<td>5</td>
</tr>
<tr>
<td>RCB-2D-xz</td>
<td>14</td>
</tr>
<tr>
<td>RCB-2D-yz</td>
<td>6</td>
</tr>
<tr>
<td>RCB-3D</td>
<td>8</td>
</tr>
<tr>
<td>RIB-2D-xy</td>
<td>6</td>
</tr>
<tr>
<td>RIB-2D-xz</td>
<td>14</td>
</tr>
<tr>
<td>RIB-2D-yz</td>
<td>5</td>
</tr>
<tr>
<td>RIB-3D</td>
<td>7</td>
</tr>
</tbody>
</table>
# RDDS Coupler Cell Constrained Partition (32 procs)

<table>
<thead>
<tr>
<th>Method</th>
<th>Max Adj. Procs</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSFC-3D</td>
<td>17</td>
</tr>
<tr>
<td>ParMETIS</td>
<td>14</td>
</tr>
<tr>
<td>RCB-1D-z</td>
<td>29</td>
</tr>
<tr>
<td>RCB-2D-xy</td>
<td>7</td>
</tr>
<tr>
<td>RCB-2D-xz</td>
<td>29</td>
</tr>
<tr>
<td>RCB-2D-yz</td>
<td>7</td>
</tr>
<tr>
<td>RCB-3D</td>
<td>11</td>
</tr>
<tr>
<td>RIB-2D-xy</td>
<td>7</td>
</tr>
<tr>
<td>RIB-2D-xz</td>
<td>29</td>
</tr>
<tr>
<td>RIB-2D-yz</td>
<td>6</td>
</tr>
<tr>
<td>RIB-3D</td>
<td>12</td>
</tr>
</tbody>
</table>
"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 in Tau3P (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