Parallel Mesh Partitioning at SLAC

Michael M. Wolf
Stanford Linear Accelerator Center
Stanford Linear Accelerator Center

- DOE laboratory managed by Stanford University
- Established 1962
- Located at Stanford University in Menlo Park, CA
- “Mission is to design, construct and operate state-of-the-art electron accelerators and related experimental facilities for use in high-energy physics and synchrotron radiation research.”
- 3 kilometers long (e-gun to start of rings)
- 3 Nobel Prize Winners
- Home of the first U.S. Website
Stanford Linear Accelerator Center

Courtesy Stanford Linear Accelerator Center
Stanford Linear Accelerator Center

SLAC/LBL/LLNL
SLAC-Based B Factory:
PEP-II and BaBAR

Both Rings Housed in Current PEP Tunnel

Courtesy Stanford Linear Accelerator Center
Stanford e+e- Linear Collider (SLC)

1/20,000,000,000 second later
(notice how far the bunches have moved)
Advanced Computations Department
# ACD Organization/Collaborators

## Advanced Computations Department

<table>
<thead>
<tr>
<th>Accelerator Modeling</th>
<th>Computational Mathematics</th>
<th>Computing Technologies</th>
</tr>
</thead>
<tbody>
<tr>
<td>V. Ivanov, A. Kabel, K. Ko, Z. Li, C. Ng, L. Stingelin (PSI)</td>
<td>Y. Liu, I. Malik, W. Mi, J. Scoville, K. Shah, Y. Sun (Stanford)</td>
<td>N. Folwell, L. Ge, A. Guetz, R. Lee, M. Wolf</td>
</tr>
</tbody>
</table>

## Collaborators

<table>
<thead>
<tr>
<th>LBNL (SCG)</th>
<th>Stanford (SCCM)</th>
<th>Sandia</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. Ng, P. Husbands, S. Li, A. Pinar</td>
<td>G. Golub, O. Livne,</td>
<td>P. Knupp, T. Tautges, L. Freitag, K. Devine</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>UCD (VGRG)</th>
<th>LLNL (CASC)</th>
<th>RPI (SCOREC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>K. Ma, G. Schussman</td>
<td>D. Brown, K. Chand, B. Henshaw</td>
<td>M. Shephard, Y. Luo</td>
</tr>
</tbody>
</table>
• Accurate modeling essential for modern accelerator design.
• Uncertainty in design greatly increases cost
• Accurate computer models reduce design costs and design cycle
• Need for accurate cavity design tools
• ACD develops these simulation codes
  • E&M field, resonant frequency, particle tracking simulations
  • Conformal meshes (Unstructured grid)
  • Parallel processing
• Codes: Omega3P, Tau3P, Track3P, S3P, Phi3P
Challenges in E&M Modeling of Accelerators

- Accurate geometry is important due to tight tolerance
  - needs unstructured grid to conform to curved surfaces

- Large, complex electromagnetic structures
  - large matrices after discretization (100's of millions of DOF), needs parallel computing for both problem storage and reduction of simulation time

- Small beam size ~ delta function excitation in time & space
  - **Time domain** - needs to resolve beam size leading to huge number of grid points, long run time & numerical stability issues
  - **Frequency domain** - wide, dense spectrum to solve for thousands of modes to high accuracy
Motivation for New EM Capability

Modeling RDDS Cell with standard accelerator code MAFIA using Structured Grid on Desktops demonstrates the need for MORE ACCURATE EM codes.

Graph showing the relationship between number of mesh points and DF (DF = F-11424) (MHz) with corresponding execution times:
- 0.01% ~ 1 MHz
- 1.2 hrs
- 3.4 hrs
- 5.4 hrs
- 7.0 hrs

The graph indicates a linear increase in execution time with an increase in the number of mesh points.
Typical ACD problems
Beam heating in the beamline complex near the IR limited the PEP-II from operating at high currents. Omega3P analysis helped in redesigning the IR for the upgrade.
Next Linear Collider (NLC)

<table>
<thead>
<tr>
<th></th>
<th>SLC</th>
<th>NLC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Center of Mass</td>
<td>100 GeV</td>
<td>500 GeV</td>
</tr>
<tr>
<td>Bunches per pulse</td>
<td>1</td>
<td>95</td>
</tr>
<tr>
<td>Operating Frequency</td>
<td>(S) 2.856 GHz</td>
<td>(X) 11.424 GHz</td>
</tr>
<tr>
<td>Number of Cavities</td>
<td>80,000</td>
<td>2 million</td>
</tr>
<tr>
<td>Post-Tuning</td>
<td>yes</td>
<td>No</td>
</tr>
</tbody>
</table>

Control of Long Range Wakefield crucial to multi-bunch operation

Wakefield Amp. (VpC/m/m/m) vs Distance Behind Bunch (m)
The NLC Accelerating Structure

206-Cell Round Damped Detuned Structure RDDS

- Needs accelerating frequency calculated to 0.01% accuracy to maintain structure efficiency
- Optimized design could save $100 million in machine cost

Cell optimized to increase shunt impedance (~14%) & minimize surface gradients

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

Manifold damping to suppress long range wakes

11 cavity dimensions
NLC Structure Design Requirements

- **RDDS Cell**: Design to 0.01% accuracy in accelerating frequency,
- **RDDS Section**: Model damping/detuning of dipole wakefields.
Particle Tracking in 5 Cell RDDS (Tau3P/Track3P)
Cyclotron COMET (Omega3P)

First ever detailed analysis of an entire cyclotron structure
- L. Stingelin, PSI

„Dee“: RF electrode
„Liner“: outer shell of RF cavity

Proton trajectories
Electric field in acceleration gap
Magnetic Field

Fig 1. The cyclotron COMET
Where we want to go
End-to-end NLC Structure Simulation

(J. Wang, C. Adolphsen - SLAC)

- 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
End-to-end NLC Klystron Simulation

Field and particle data estimated to be TB size
Tau3P
Parallel Time-Domain 3D Field Solver
Parallel Time-Domain Field Solver – Tau3P

Follows evolution of E and H fields on primary/dual meshes (hexahedral, tetrahedral and pentahedral elements) using leap-frog scheme in time (DSI scheme)

\[ \oint E \cdot ds = -\iint \frac{\partial B}{\partial t} \cdot dA \]

\[ \oint H \cdot ds^* = \iint \frac{\partial D}{\partial t} \cdot dA^* + \iint j \cdot dA^* \]
Discrete Surface Integral Method

- Electric fields on primary grid
- Magnetic fields on dual grid
- Primary and dual grids non-orthogonal
- Dual grid constructed by joining centers of primary cells
- Electric and magnetic fields advanced in time using the leapfrog algorithm
- Reduces to conventional finite difference time domain method (FDTD) for non-orthogonal grids
- Conforms to complicated geometry by appropriate choice of element types

Tau3P Applications
Time Domain Design & Analysis (Tau3P)

Matching NLC Input Coupler w/ Inline Taper

Dipole mode spectrum

Output Coupler loading on HOM modes at the RDDS output end

Dipole Excitation
Wakefield Calculation (Tau3P)

- Response of a 23-cell X-Band Standing Wave Structure w/ Input Coupler & Tapered Cells to a transit beam in Tau3P.
- Direct wakefield simulation of exact structure to verify approximate results from circuit model.

a: 4.663-4.875 mm, b: 10.796-10.879 mm, t: 2.541-2.684 mm
Determining Peak Fields (Tau3P)

- When and where Peak Fields occur during the pulse?
- Transient fields 20% higher than steady-state value due to dispersive effects

**Drive pulse**

**Electric field vs time**

Rise time = 10, 15, 20 ns

Steady-state Surface Electric field amplitude
15-Cell H90VG5 Model

- Peak field appears near the middle of the structure
- About 25% overshoot in peak field due to the narrower bandwidth

Overshoots for different rise times

Fields at different cell disks as a function of time

Rise time = 10ns
Electric Fields (Tau3P)
Tau3P Matrices
Discrete Surface Integral Formulation

\[ \oint E \cdot ds = - \oint \frac{\partial B}{\partial t} \cdot dA \]

\[ \oint H \cdot ds^* = \oint \frac{\partial D}{\partial t} \cdot dA^* + \oint j \cdot dA^* \]

• The DSI formulation yields:
  - efield += \( \alpha \cdot AH \cdot hfield \)
  - hfield += \( \beta \cdot AE \cdot efield \)
    • efield, hfield are vectors of field projections along edges/dual-edges
    • AH, AE are matrices
    • \( \alpha, \beta \) are constants proportional to \( dt \)
Tau3P Implementation

Example of Distributed Mesh

Typical AE Distributed Matrix
Tau3P Matrix Properties

• Very Sparse Matrices
  - 4-20 nonzeros per row
• 2 Coupled Matrices (AH, AE)
• Nonsymmetric (Rectangular)

Typical Distributed Matrix
Tau3P Performance Problems
Load Balancing Issues in Tau3P

  - Unstructured meshes lead to matrices for which nonzero entries are not evenly distributed.
  - Complicates work assignment and load balancing in a parallel setting.
  - Tau3P originally used ParMETIS to partition the domain to minimize communication.

NZ Distribution over 14 cpu’s

Parallel Speedup
Parallel Performance of Tau3P

- 257K hexahedrons
- 11.4 million non-zeroes

Parallel Speedup

SLAC PC Cluster
Communication in Tau3P (ParMetis Partitioning)
Communication in Tau3P (ParMetis Partitioning)
Improved Mesh Partitioning Schemes
Luxury in Tau3P Mesh Partitioning

- Long simulation times
  - Tens of thousands of CPU hours
- Long time spent in time stepping
  - Millions of time steps
- Problem initialization short
- Static (not dynamic) mesh partitioning
- Willing to pay HIGH price upfront for increased performance of solver
Zoltan Overview

- Developed at Sandia National Laboratory (NM)
- Collection of Data Management Services for Parallel, Unstructured, Adaptive, and Dynamic Applications
- Supports Several Load Balancing Methods:
  - Graph Partitioning Algorithms
    - ParMETIS
    - Jostle
  - Geometric Partitioning Algorithms (1D/2D/3D)
    - Recursive Coordinate Bisection (RCB)
    - Recursive Inertial Bisection (RIB)
    - Hilbert Space-Filing Curve (HSFC)
    - Octree Partitioning (various traversal schemes including HSFC)
    - Refinement Tree Based Partitioning (mesh refinement)
- Supports Dynamic Load Balancing/Data Migration
Zoltan Partitioning Methods
ParMETIS (Graph)
Recursive Coordinate Bisection (Geometric)
Hilbert Space Filling Curve (Geometric)
Hilbert Space Filling Curve (Geometric)
Tau3P Partitioning Results
RDDS (5 cell w/ couplers) ParMETIS Partition

ParMETIS (8 procs)

Adj. Procs (Max/Sum): 3/14
Bound. Objs (Max/Sum): 533/2778

Tau3P Run-time: 140.6
RDDS (5 cell w/ couplers) RCB-1D(z) Partition

RCB – 1D (8 procs)

Adj. Procs (Max/Sum): 2/14
Bound. Objs (Max/Sum): 3128/23654

Tau3P Run-time: 126.4
RDDS (5 cell w/ couplers) RCB-3D Partition

RCB -- 3D (8 procs)

Adj. Proc (Max/Sum): 5/26
Bound. Objs (Max/Sum): 1965/11961  Tau3P Run-time: 169.1
RDDS (5 cell w/ couplers) RIB-3D Partition

RIB – 3D (8 procs)

Adj. Procs (Max/Sum): 3/18
Bound. Objs (Max/Sum): 1570/7927

Tau3P Run-time: 154.6
RDDS (5 cell w/ couplers) HSFC-3D Partition

HSFC -- 3D (8 procs)

Adj. Procs (Max/Sum): 5/32
Bound. Obj. (Max/Sum): 2030/9038

Tau3P Run-time: 194.4
## 5 Cell RDDS (8 processors) Partitioning

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></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)

<table>
<thead>
<tr>
<th># of Procs</th>
<th>ParMETIS Max Adj. Procs</th>
<th>ParMETIS Speedup</th>
<th>ParMETIS Runtime</th>
<th>RCB-1D Runtime</th>
<th>RCB-1D Speedup</th>
<th>RCB-1D Max Adj. Procs</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>2</td>
<td>8/8</td>
<td>3930.7 s</td>
<td>3898.6 s</td>
<td>8/8</td>
<td>2</td>
</tr>
<tr>
<td>16</td>
<td>3</td>
<td>11.6/16</td>
<td>2703.3 s</td>
<td>2458.5 s</td>
<td>12.7/16</td>
<td>2</td>
</tr>
<tr>
<td>32</td>
<td>4</td>
<td>21.6/32</td>
<td>1455.0 s</td>
<td>1236.6 s</td>
<td>25.2/32</td>
<td>2</td>
</tr>
<tr>
<td>64</td>
<td>4</td>
<td>42.7/64</td>
<td>736.6 s</td>
<td>627.2 s</td>
<td>49.7/64</td>
<td>2</td>
</tr>
<tr>
<td>128</td>
<td>10</td>
<td>48.9/128</td>
<td>643.0 s</td>
<td>265.1 s</td>
<td>117.6/128</td>
<td>2</td>
</tr>
<tr>
<td>256</td>
<td>11</td>
<td>87.3/256</td>
<td>360.0 s</td>
<td>129.2 s</td>
<td>241.4/256</td>
<td>2</td>
</tr>
<tr>
<td>512</td>
<td>14</td>
<td>107.7/512</td>
<td>292.1 s</td>
<td>92.3 s</td>
<td>337.9/512</td>
<td>4</td>
</tr>
<tr>
<td>1024</td>
<td>16</td>
<td>96.0/1024</td>
<td>327.5 s</td>
<td>99.0 s</td>
<td>315.0/1024</td>
<td>8</td>
</tr>
</tbody>
</table>

**1.0 ns runtime**

**IBM SP3 (NERSC)**
RCB Scalability Leveling Off
RCB Scalability Leveling Off
Coupler Port Grouping Complication
Coupler Port Grouping Complication
H60VG3 RDDS Partitioning (w/ coupler port grouping)

<table>
<thead>
<tr>
<th># of Procs</th>
<th>ParMETIS Max Adj. Procs</th>
<th>ParMETIS Speedup</th>
<th>ParMETIS Runtime</th>
<th>RCB-1D Runtime</th>
<th>RCB-1D Speedup</th>
<th>RCB-1D Max Adj. Procs</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>2</td>
<td>8/8</td>
<td>3856.2 s</td>
<td>3826.5 s</td>
<td>8/8</td>
<td>2</td>
</tr>
<tr>
<td>16</td>
<td>3</td>
<td>7.2/16</td>
<td>4257.0 s</td>
<td>2405.4 s</td>
<td>12.7/16</td>
<td>2</td>
</tr>
<tr>
<td>32</td>
<td>3</td>
<td>14.3/32</td>
<td>2158.3 s</td>
<td>1820.7 s</td>
<td>16.8/32</td>
<td>2</td>
</tr>
<tr>
<td>64</td>
<td>7</td>
<td>31.0/64</td>
<td>995.1 s</td>
<td>889.3 s</td>
<td>34.4/64</td>
<td>3</td>
</tr>
<tr>
<td>128</td>
<td>7</td>
<td>47.5/128</td>
<td>649.6 s</td>
<td>599.0 s</td>
<td>51.1/128</td>
<td>6</td>
</tr>
<tr>
<td>256</td>
<td>9</td>
<td>69.7/256</td>
<td>442.9 s</td>
<td>516.5 s</td>
<td>59.3/256</td>
<td>11</td>
</tr>
<tr>
<td>512</td>
<td>12</td>
<td>70.4/256</td>
<td>438.1 s</td>
<td>531.1 s</td>
<td>57.6/512</td>
<td>21</td>
</tr>
</tbody>
</table>

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>
Future Work I Would Have Done

• Stitching Multiple Partitions Together
• Competition
• Onion Partition growing
• Dynamic Partitioning for Track3P
Acknowledgements

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

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