A Co-Design Study Of Fusion Whole Device Modeling Using Code Coupling

Jong Youl Choi, Jeremy Logan, Kshitij Mehta, Eric Suchyta, William Godoy, Nicholas Thompson, Lipeng Wan, Jieyang Chen, Norbert Podhorszki, Matthew Wolf, and Scott Klasky (ORNL), Julien Dominski and Choong-Seock Chang (PPPL)

The 5th International Workshop on Data Analysis and Reduction for Big Scientific Data (DRBSD-5) in conjunction with SC19
17 November 2019, Denver, Colorado
Multiscale and Multiphysics Applications in Exascale

- Science code is getting complex
  - Multi-scale, multi-physics
  - Multiple components
  - Multiple systems and H/Ws
- And, code coupling has been developed
- But, it is challenging to understand interactions and trade-offs between parameters and codes
- Therefore, we need to codesign study to investigate various trade-offs
CODAR: Online Data Analysis and Reduction

“online indicates a state of connectivity … offline indicates a disconnected state” [wikipedia]

(I. Foster et al., CODAR, 2019)
Can couple tasks via file system? 

Yes: Not our concern …

No: Too much data to output, store, or analyze offline. Must couple tasks online.

Which tasks?

Online reduction
- Application + Reduction

Online analysis
- Application + Analysis

Online coupling
- Application + Application

Many Applications
- Online aggregation

Which Applications?
- Many Applications
- Online reduction
  - Application + Reduction
  - ExaFEL: X-ray laser imaging
  - NWChemEx: Molecular dynamics
  - WDMApp: Fusion whole device model
  - CANDLE: Cancer deep learning

Hyperparam. optimization: $10^3$–$10^6$ training runs, each fitting many parameters.
What is Co-Design Study in CODAR?

- Cross-cutting technical challenges for which solutions must be developed and/or integrated

- **Identify the best data analysis and reduction algorithms** for different application classes, in terms of speed, accuracy, and resource requirements

- **Quantify tradeoffs** in data analysis accuracy, resource needs, and overall application performance among various data reduction methods. How do these tradeoffs vary with exascale hardware and software choices?

- **Effectively orchestrate** online data analysis and reduction to reduce associated overheads. How can exascale hardware and software help with orchestration?
Fusion Whole Device Model (WDM)

- Magnetic fusion plasma is governed by several multiscale multi-physics
  - Coupled simulation is necessary for high-fidelity

- Core and edge physics
  - Core obeys the near-thermal-equilibrium physics
  - Edge obeys the far-from-equilibrium physics: scale-inseparable multi-physics
  - Using a single-executable XGC-edge for a whole-device ITER turbulence solution would consume ~50 days of wall-clock time on 27 PF Titan
  - With a successful core-edge coupling, the wall-clock time can be reduced to ~5 days

(Credit: CS Chang, S. Klasky)
Building Coupling Workflow

• Monolithic design
  • One large application with one big communicator
  • Single MPI World communicator
  • Any failure can destroy whole workflow (weak resilience)
  • High complexity in development and testing

• A New (?) Approach
  • Many independent applications (including other science applications, services, plug-ins, etc)
  • Each owns MPI World communicator (if they are MPI-based applications)
  • Separation of concerns (sandbox approach)
  • Incremental testing/development process: file-based coupling ➔ in-memory coupling/in situ analysis
What is ADIOS

- An extendable **framework** that allows developers to **plug-in**
  - **I/O methods**: Aggregate, Posix, MPI
  - **Services**: Compression, Decompression
  - **Formats**: HDF5, netcdf, ADIOS-BP,…
  - **Plug-ins**: Analytic, Visualization
- Incorporates the “best” practices in the I/O middleware layer
- [https://csmd.ornl.gov/adios](https://csmd.ornl.gov/adios), [https://github.com/ornladios/ADIOS](https://github.com/ornladios/ADIOS), [https://github.com/ornladios/ADIOS2](https://github.com/ornladios/ADIOS2)
Coupling Methods in ADIOS

• Sustainable Staging Transport (SST)
  • In situ infrastructure for staging in a streaming-like fashion using RDMA, SOCKETS with “active” connect/disconnect

• InSituMPI
  • MPI-based staging for MPMD applications, for strong coupling

• DataMan
  • WAN transfers using sockets and ZeroMQ for EO data

• Inline
  • Synchronous in situ, direct pass through of data structures to analytics subroutine


(Credit: S. Klasky, N. Podhorszki)
Develop tools for support **complex, coupled workflows** consisting of independently running **simulation** and **analysis** applications.

- **Challenges**
  - Big data and performance challenge
  - Supporting In situ/online analysis
  - Managing complex workflow

- **Impact**
  - ECP whole device modeling demonstration and tutorials
  - CODAR co-design study
**Approaches to build WDM mini-app**

- We need a simplified application to test on various machines, Adios methods, placements, data reduction, etc.
- Use the same computational and communication kernels
- Only coupling parts have been mini-appified
- Can be less flexible
- But, it can be more precise to the real application.

<table>
<thead>
<tr>
<th>Type</th>
<th>Example</th>
<th>Pros/Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automatic generation</td>
<td>Skel, IOR</td>
<td>• Easy parameterization</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Flexible</td>
</tr>
<tr>
<td>Trace-based generation</td>
<td>APPrime, ScalBenGen</td>
<td>• Automatic generation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Replay based</td>
</tr>
<tr>
<td>Application Specific</td>
<td></td>
<td>• Close to the real application</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Application specific</td>
</tr>
</tbody>
</table>
WDM Mini-App Coupling Workflow

- Multiple WDM coupling scenarios:
  - XGC-X coupling, where X=GEM, GENE, XGC1, and XGCa
- 3 physics property to couple:
  - Fluid information (mesh data)
  - 5D distribution (5D f data)
  - Particles (particle data)
- XGC edge code runs with GPUs, while XGC core code runs only with remaining resources on Summit
- On Summit, we can run coupling codes to use separate nodes or shared nodes
WDM Mini-app Coupling Data

<table>
<thead>
<tr>
<th>Type</th>
<th>Shape</th>
<th>Size</th>
<th>Communication Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fluid</td>
<td>3D array</td>
<td>Small</td>
<td>One process per plane</td>
</tr>
<tr>
<td>5D distribution</td>
<td>5D array</td>
<td>Medium</td>
<td>Each process</td>
</tr>
<tr>
<td>Particle</td>
<td>Table</td>
<td>Large</td>
<td>Each process</td>
</tr>
</tbody>
</table>
Co-design Spaces and Parameters

- Process layouts
  - Shared node vs separate nodes
  - Shared resources
- Coupling ratios
  - CPU ratios
- Adios coupling methods
  - Files vs SST vs InSituMPI
- Data compression (future work)
  - Compression methods vs physics information
### Process Layout on Summit

**Summit shared node layout**

**Summit separate node layout**

<table>
<thead>
<tr>
<th>Layout</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
</table>
| Shared node layout      | • Shared memory  
                          • Minimize out-of-node communication | • No less than 1:6 ratio          |
| Separate node layout    | • Able to allocate large ratio                                      | • Out-of-node communication       |
CoDAR Study: Trade-offs on Summit

- WDM coupling workflow trade-offs
  - Run them in a shared mode vs run them in a separate node
  - Best process ratios for XGC core and XGC edge
  - Use GPFS vs NVME vs SST

### Computation Time

<table>
<thead>
<tr>
<th>Case</th>
<th>N1</th>
<th>N2</th>
<th>N3</th>
<th>N4</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (sec) SEPARATE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>12</td>
<td>18</td>
<td>20</td>
<td>5</td>
<td>7</td>
<td>9</td>
<td>11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Case</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (sec) SHARED</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5</td>
<td>7</td>
<td>9</td>
</tr>
</tbody>
</table>

- **Edge**
- **Core**

### Particle data coupling

<table>
<thead>
<tr>
<th>Coupling method</th>
<th>GPFS</th>
<th>NVME</th>
<th>SST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (sec) Main</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Time (sec) Read</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Time (sec) Write</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>
Summary

• WDM coupling workflow gives challenges
  • Data coordination
  • Workflow management
• CODAR is to co-design study to explore trade-offs between different system parameters
• WDM mini-app can help to conduct CODAR studies for WDM applications
Questions