SUPER Performance Autotuning

Khalid Ahmad, Mary Hall, Anand Venkat
University of Utah

Samuel Williams
Lawrence Berkeley National Laboratory

Ray Chen, Jeffrey Hollingsworth
University of Maryland

**SUPER Researchers Develop State-of-the-Art Performance Tools**

<table>
<thead>
<tr>
<th>Commercial Tools</th>
<th>SUPER Tools</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Target Workload</strong></td>
<td></td>
</tr>
<tr>
<td>General-purpose workloads (FP, CF, RE, SY)</td>
<td>Scientific simulation (FP, CF, RE, SY)</td>
</tr>
<tr>
<td><strong>Driver</strong></td>
<td>Performance, primarily</td>
</tr>
<tr>
<td>Programmer productivity, primarily</td>
<td>Performance, primarily</td>
</tr>
<tr>
<td><strong>Performance Measurement</strong></td>
<td></td>
</tr>
<tr>
<td>Focuses on execution time</td>
<td>Extensive, pinpoints opportunities for improvement</td>
</tr>
<tr>
<td><strong>Optimization</strong></td>
<td></td>
</tr>
<tr>
<td>Conservative (static, architecture independent)</td>
<td>Aggressive (dynamic, autotuning), architecture specific</td>
</tr>
<tr>
<td><strong>Riskiness</strong></td>
<td></td>
</tr>
<tr>
<td>Proven Technology</td>
<td>State-of-the-art Technology</td>
</tr>
</tbody>
</table>

**Performance Tool Integration**

- **LOBPCG and CSB Representation (2014)**
  - LOBPCG is a block eigensolver that restructures computation into a series of Sparse Matrix-Dense Matrix Multiplications (SpMM) where the dense matrix is tall and skinny (500,000 x 16 per process).
  - Once again, symmetric necessitates SpMV, T (transpose)
  - SpMM reuses matrix, reduces memory bandwidth requirements.
  - SUPER collaborated with FastMath to model, analyze, and optimize these operations.
  - Optimization Strategy employs Compressed Sparse Block (CSB) matrix representation.

- **MFDn Application (NUCLEI)**
  - Many Fermion Dynamics: Nuclear (MFDn) is used to calculate the properties of light atomic nuclei.
  - Nominally, this requires an eigensolver like Lanczos which requires applying the operator repeatedly.
  - MFDn forms the very large (half a billion nonzeroes per process) symmetric configuration interaction matrix explicitly.
  - Repeated application of the operator is bandwidth intensive.
  - Symmetry further complicates this as it necessitates high performance SpMV and SpMV, T (transpose)

**State-of-the-art tools are applied to SciDAC applications.**

- Application development informed by results derived by SUPER tools.
- Tool development informed by application requirements.

**Impact**

- Evolutionary algorithms favor exploration.
- Attempt to approximate the performance landscape.
- The ANGEL search algorithm favors exploitation.
- Attempts to move toward the goal with each evaluation.
- Mitigates the overhead of evaluating a bad configuration.
- Meets the demands of online autotuning.
- Priority and leeway specify the goal.
  - "Allow a 30% drop in my primary objective in order to improve my secondary objective."

**Penalty Function**

- ANGEL penalty function considers multiple objectives.
- Penalty in proportion to leeway violations of higher objectives.
- Prevents "needle-in-a-haystack" search surfaces
- Penalty effectively combines search surfaces.

**ANGEL: Multi-Objective Autotuning**

- ANGEL almost always performs better than others.
- When others are better, ANGEL results are close to best.

**ANGEL: Comparison with Testsuite**

- ANGEL selects solutions close to, but not violating the leeway.
- As a larger leeway is given, more energy is saved.
- ANGEL often finds the optimal performance.

**LULESH2 Kernel Time/Energy Results**

- When it does not, ANGEL remains close to the goal.

Support for this work was provided through the Scientific Discovery through Advanced Computing (SciDAC) program funded by the U.S. Department of Energy Office of Advanced Scientific Computing Research.