Intel® Xeon Phi™ Coprocessor NWP Application Experiences

Programming weather, climate, and earth-system models on multi-core platforms
September 2013
Mike Greenfield
Notice and Disclaimers

INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN INTEL’S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT.

A "Mission Critical Application" is any application in which failure of the Intel Product could result, directly or indirectly, in personal injury or death. SHOULD YOU PURCHASE OR USE INTEL'S PRODUCTS FOR ANY SUCH MISSION CRITICAL APPLICATION, YOU SHALL INDEMNIFY AND HOLD INTEL AND ITS SUBSIDIARIES, SUBCONTRACTORS AND AFFILIATES, AND THE DIRECTORS, OFFICERS, AND EMPLOYEES OF EACH, HARMLESS AGAINST ALL CLAIMS COSTS, DAMAGES, AND EXPENSES AND REASONABLE ATTORNEYS' FEES ARISING OUT OF, DIRECTLY OR INDIRECTLY, ANY CLAIM OF PRODUCT LIABILITY, PERSONAL INJURY, OR DEATH ARISING IN ANY WAY OUT OF SUCH MISSION CRITICAL APPLICATION, WHETHER OR NOT INTEL OR ITS SUBCONTRACTOR WAS NEGLIGENT IN THE DESIGN, MANUFACTURE, OR WARNING OF THE INTEL PRODUCT OR ANY OF ITS PARTS.

All products, computer systems, dates and figures specified are preliminary based on current expectations, and are subject to change without notice.

Intel may make changes to specifications and product descriptions at any time, without notice. Designers must not rely on the absence or characteristics of any features or instructions marked "reserved" or "undefined". Intel reserves these for future definition and shall have no responsibility whatsoever for conflicts or incompatibilities arising from future changes to them. The information here is subject to change without notice. Do not finalize a design with this information.

The products described in this document may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request. Contact your local Intel sales office or your distributor to obtain the latest specifications and before placing your product order.

Copies of documents which have an order number and are referenced in this document, or other Intel literature, may be obtained by calling 1-800-548-4725, or go to: http://www.intel.com/design/literature.htm

Intel, Intel Xeon, Intel Xeon Phi™ are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States or other countries.

*Other brands and names may be claimed as the property of others.

Copyright © 2013 Intel Corporation . All rights reserved.
<table>
<thead>
<tr>
<th>Optimization Notice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intel’s compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel.</td>
</tr>
<tr>
<td>Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice.</td>
</tr>
<tr>
<td>Notice revision #20110804</td>
</tr>
</tbody>
</table>
Agenda

• Intel HPC Platforms
• Overview of NWP projects and collaborations
• Scalability Challenge
• MIC Development considerations

Referenced in this presentation
• MIC = Intel® Many Integrated Core Architecture
• KNC, KNL = Knights Corner, Knights Landing, aka Intel® Xeon Phi™ Coprocessor
• SNB-EP = E5-2670 (2x8c, 2.7Ghz, 64GB 1600 DDR3 (Intel® Xeon® Processor)
• IVB-EP = E5-2697v2 (2x12C, 2.7GHz, 64GB 1600 (Intel® Xeon® Processor)
Intel’s Many Core and Multi-core Engines

**Intel® Xeon® processor:**
- Intel’s Foundation of HPC Performance
- Suited for full scope of workloads
- Industry leading performance/watt for serial & highly parallel workloads.

**Intel® Xeon Phi™ Coprocessor:**
- Optimized for highly parallelized compute intensive workloads
- Common programming model & S/W tools with Xeon processors, enabling efficient app readiness and performance tuning
- 22nm with up to 61 cores and high memory b/w to provide outstanding performance for highly parallel HPC uses
The Intel® Xeon Phi™ Family of Products

Knights Corner
1st Intel® Xeon Phi™ Coprocessor product
22nm process
Up to 61 Intel Architecture Cores
PCIe

Knights Landing
2nd generation Intel® Many Integrated Core Architecture
14nm process; Intel® AVX-512
Processor and Coprocessor
On-package High BW memory

Future Knights Products

Future options subject to change without notice.
MIC Applications reported at this conference

2012
- FIM
- WRF (WSM5, CHEM)

2013
- NIM and FIM
- WRF (WRF, WSM5, CHEM)
- HOMME, CAM, DGKERNEL
- HBM Myocv3
- GEOS5 (fyppm, cubesphere)
- ESMF, NEMS/NMMB
**WEATHER RESEARCH AND FORECASTING (WRF)**

- **Application:** WRF

- **Code Optimization:**
  - Approximately two dozen files with less than 2,000 lines of code were modified (out of approximately 700,000 lines of code in about 800 files, all Fortran standard compliant)
  - Most modifications improved performance for both the host and the co-processors

- **Performance Measurements:** V3.5 and NCAR supported CONUS2.5KM benchmark (a high resolution weather forecast)

- **Acknowledgments:** There were many contributors to these results, including the National Renewable Energy Laboratory and The Weather Channel Companies

---

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change in any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products.

Source: Intel or third party measured. Results as of September 2013. Configuration Details: Please reference slide speaker notes. For more information go to http://www.intel.com/performance Any difference in system hardware or software design or configuration may affect actual performance. Copyright © 2013, Intel Corporation. * Other names and brands may be claimed as the property of others.
Ref: DMI Technical reports TR12-11, TR12-20, Jacob Weismann Poulsen and Per Berg
- KNC evaluation of HBM myov3 underway
- Kit built and ran with no source changes
- Key challenges: cache blocking, vectorization, load imbalance
- Status: KNC 10% slower than E5-2670
- Next Steps: Several optimizations to be integrated, Host/Card operation, Multinode


Copyright Intel Corporation 2011-13
Performance Disambiguation

Setup | Kernel | Data xfer

Time | Kernel

?Precision | Ecc?

Original Problem (difficult Parts) vs Subset Problem

CPU: -00, -no || -noVec
GPU: -05, -full || -Vec

Parallel, Accelerators and Co-Processors challenge “Performance Discipline”

Adapted from Ten Ways to Fool the Masses When Giving Performance Results on GPUs, Scott Pakin, Los Alamos National Laboratory, December 2011
**Peak Performance:** The prize in the peak performance category is given to the entry demonstrating the highest performance achieved in terms of operations per second on a genuine application program.

**Intel® Xeon Phi™ product family:** Designed for **Highly Parallel** workloads

**Question:** Are the applications and workloads cited as past Gordon Bell “Peak Performance” prize award winners likely to run well on Clusters based on the Intel® Xeon Phi™ Coprocessor & Intel® Xeon® processor?
<table>
<thead>
<tr>
<th>Scalability Dimension</th>
<th>Performance Validation examples (Improves with..)</th>
<th>Critical Dependence?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fabric Scaling</td>
<td>increased node count</td>
<td>GB: Required</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MIC: Depends</td>
</tr>
<tr>
<td>Thread Scaling on shared memory</td>
<td>increased cores on coherent shared memory</td>
<td>GB: nice to have</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MIC: Essential</td>
</tr>
<tr>
<td>Vector Scaling</td>
<td># and length of registers</td>
<td>GB: nice to have</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MIC: Essential</td>
</tr>
<tr>
<td>Cache Scaling</td>
<td>effective blocking</td>
<td>GB: nice to have</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MIC: Essential</td>
</tr>
</tbody>
</table>

**GB = Gordon Bell**
## The Scalability Challenge

<table>
<thead>
<tr>
<th>Scalability Dimension</th>
<th>Suggested Experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fabric Scaling</td>
<td>Performance as $f(\text{node-count})$</td>
</tr>
<tr>
<td>Thread Scaling on shared memory</td>
<td>At some fixed large node count, vary # ranks or # threads from 1 to max cores. What is core scaling?</td>
</tr>
<tr>
<td>Vector Scaling</td>
<td>At some fixed node &amp; rank/thread count, vary vectorization. What is profitability?</td>
</tr>
<tr>
<td>Cache Scaling</td>
<td>What is range of performance improvement obtained from cache blocking (tuning workload to reduce memory bandwidth pressure and improve cache hit rate)?</td>
</tr>
</tbody>
</table>

How would your parallel applications look in these tests?
Application FOM?

Ref: “Hello World”

WRF: 2000 East Coast Winter Storm

<table>
<thead>
<tr>
<th>FOM</th>
<th>Scalability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fabric</td>
<td></td>
</tr>
<tr>
<td>Threads</td>
<td></td>
</tr>
<tr>
<td>Vectors</td>
<td></td>
</tr>
<tr>
<td>Cache</td>
<td></td>
</tr>
</tbody>
</table>

Necessary, but not sufficient
WRF CONUS12KM Example

64 Nodes: Performance vs # Ranks/node

Additional notes:
• RHS chart normalized to 64N 1rank/node
• LHS chart normalized to 1N 16 rank
• WRF V3.4

1->64 Nodes: Performance vs # Nodes

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change in any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products.

Source: Intel or third party measured. Results as of September 2013. Configuration Details: Please reference slide speaker notes. For more information go to http://www.intel.com/performance. Any difference in system hardware or software design or configuration may affect actual performance. Copyright © 2013, Intel Corporation. * Other names and brands may be claimed as the property of others.
WRF-CHEM Example

Active threads < 40

Available threads = 4*61=244

Expose Parallelism (required)
Development ROI

**Essentials:**
- Thread scaling
- Profitable Vectorization
- Ram/Cache Scaling

**Additional Optimizations:**
- Cache Blocking
- Memory Layout
- Large Pages
- Alignment
- Pre-fetching
- Reduced Precision

Re-compile?

Offload

Offload (persistence, threshold tuning)

Concurrency, Load Balance

Cilk Plus

Offload (asynchronous, persistence, threshold tuning)

Development Investment
KNC Application Retrospective

**FOM:**
- Thread Scaling
- Vectorization
- Cache Scaling

**Portability and Compatibility**

**Numerical Quality**

**Repeatability**

**Decomposition Strategies**
- MPI ranks are memory heavy
- MPI: Host > Co-Processor Perf
- Best: Minimize MPI ranks on Card, exploit parallelism w OMP

**Reveal Concurrency**

**Load Imbalance = Serialization**

**Big CORE** ≠ Small Core
KNC Application Retrospective, Part 2

Data Alignment
- Align Data
- Comprehension
- Exploitation

Decomposition Tuning:
- Platform Mapping
- Runtime Decomposition

Precision vs Sledgehammer Tuning (e.g. pre-fetch)

Timing Avalanche

Division expensive: replace with Multiplication by reciprocal where possible

Compilation Options:
- Conservative - Aggressive
Knights Corner
Where to learn more

http://software.intel.com/mic-developer

Intel® Xeon Phi™ Coprocessor

Parallel Processing
Architecture for Discovery

Productivity via architecture innovation coupled with familiar software. Intel® Xeon Phi™ coprocessor:

- Extends hardware support to higher degrees of parallelism with power savings
- Uses familiar and standard programming models to preserve investments
- Shares parallel programming with general purpose processor
Summary

• Dual Tune: Scale and vectorize to optimize

[Image of Intel Xeon and Intel Xeon Phi]

and Together!

• Parallel has many dimensions; Reveal great thread, vector and cache blocking scalability in your applications

• Intel® Xeon Phi™ Coprocessor provides delivers programmability and performance/watt for highly parallel applications