HPCG Results on IA:
What does it tell about architecture?

Jongsoo Park[*], Mikhail Smelyanskiy[*], Alexander Heinecke[*], Vadim Pirogov[*], Scott David[*], Carlos Rosales-Fernandez[#], Christopher Daley[$], Yutong Lu[%], Laurent Nguyen[&], and Ludovic Sauge[+]

[*]Intel, [#]TACC, [$]LBNL, [%]NUDT, [&]CEA, [+]BULL

SC14, HPCG BoF

HPCG performs well with

- High memory BW

HPCG performs well with

High memory BW:

mostly technology constrained

HPCG performs well with

High memory BW:

mostly technology constrained

What architecture can do?

- Efficient caching
- Fast core-to-core communication
- High-radix network

IA HPCG Results – Single Node

Computer	HPCG perf. (GFLOPS)	Single-node HPCG perf. (GFLOPS)	Single-node Efficiency w.r.t. STREAM
TH2[1)]	580,109	45.9	61%
SuperMUC[2)]	83,261	10.8	93%
Edison[3)]	78,644	15.0	100%
Occigen[4)]	45,475	21.9	112%
1/3 of Stampede[5)]	43,959	29.9	72%

1) Results from ISC14. STREAM BW 150 GB/s per Intel® Xeon Phi™ coprocessors. Intel® Xeon® processors were not used for computation, and their STREAM BW was excluded when computing BW efficiency.
2) Results from ISC14. STREAM BW 70 GB/s per node is used for BW efficiency.
3) STREAM BW 90 GB/s per node is used for BW efficiency.
4) STREAM BW 117 GB/s per node is used for BW efficiency.
5) Both Xeon processors and Xeon Phi coprocessors are used. 80 GB/s Xeon processor STREAM BW and 170 GB/s Xeon Phi coprocessor STREAM BW is used for BW efficiency

Very high Intel® Xeon® processor efficiency w.r.t. STREAM BW

Computer	HPCG perf. (GFLOPS)	Single-node HPCG perf. (GFLOPS)	Single-node Efficiency w.r.t. STREAM	
TH2[1)]	580,109	45.9		61%
SuperMUC[2)]	83,261	10.8	SNB	93%
Edison[3)]	78,644	15.0	IVB	100%
Occigen[4)]	45,475	21.9	HSW	112%
1/3 of Stampede[5)]	43,959	29.9		72%

1) Results from ISC14. STREAM BW 150 GB/s per Intel® Xeon Phi™ coprocessors. Intel® Xeon® processors were not used for computation, and their STREAM BW was excluded when computing BW efficiency.
2) Results from ISC14. STREAM BW 70 GB/s per node is used for BW efficiency.
3) STREAM BW 90 GB/s per node is used for BW efficiency.
4) STREAM BW 117 GB/s per node is used for BW efficiency.
5) Both Xeon processors and Xeon Phi coprocessors are used. 80 GB/s Xeon processor STREAM BW and 170 GB/s Xeon Phi coprocessor STREAM BW is used for BW efficiency

Very high Intel® Xeon® processor efficiency w.r.t. STREAM BW

Computer	HPCG perf. (GFLOPS)	Single-node HPCG perf. (GFLOPS)	Single-node Efficiency w.r.t. STREAM	
TH2[1)]	580,109	45.9		61%
SuperMUC[2)]	83,261	10.8	SNB	93%
Edison[3)]	78,644	15.0	IVB	100%
Occigen[4)]	45,475	21.9	HSW	112%
1/3 of Stampede[5)]	43,959	29.9		72%

GS + SpMV loop fusion optimization → exceed memory BW limit
Newer generations with large LLCs show high efficiency
Cache matters for HPCG!

1.5x single-node performance in Stampede compared to ISC14

Computer	HPCG perf. (GFLOPS)	Single-node HPCG perf. (GFLOPS)	Single-node Efficiency w.r.t. STREAM
TH2[1)]	580,109	45.9	61%
SuperMUC[2)]	83,261	10.8	93%
Edison[3)]	78,644	15.0	100%
Occigen[4)]	45,475	21.9	112%
1/3 of Stampede[5)]	43,959	29.9	72%
1/6 of Stampede (ISC)	16,100	20.8	50%

1) Results from ISC14. STREAM BW 180 GB/s per Intel Xeon Phi coprocessor. Intel Xeon processors were not used for computation, and their STREAM BW was excluded when computing BW efficiency.
2) Results from ISC14. STREAM BW 70 GB/s per node is used for BW efficiency.
3) STREAM BW 90 GB/s per node is used for BW efficiency.
4) STREAM BW 117 GB/s per node is used for BW efficiency.
5) Both Xeon processors and Xeon Phi coprocessors are used. 80 GB/s Xeon processor STREAM BW and 170 GB/s Xeon Phi coprocessor STREAM BW is used for BW efficiency

Hybrid parallelization: higher Xeon Phi performance

Important at coarser MG levels

Important at finer MG levels

	Parallelism	Convergence (# iterations)
Level-schedule	$O(N^2)$	**50** ☺
Block-coloring (B-size blocks)	**$O(N^3/B)$** ☺	56-66

Hybrid parallelization: higher Xeon Phi performance

Important at coarser MG levels

Important at finer MG levels

	Parallelism	Convergence (# iterations)
Level-schedule	$O(N^2)$	**50** ☺
Block-coloring (B-size blocks)	**$O(N^3/B)$** ☺	56-66

Architecture support for fine-grain synchronization → Level-scheduling → Keep convergence → Reduce all-reduce time

Find more technical details at our paper presentation on Thu 2:30pm

Efficient Shared-Memory Implementation of HPCG Benchmark and Its Application to Unstructured Matrices

Jongsoo Park*, Mikhail Smelyanskiy*, Karhikeyan Vaidyanathan*,
Alexander Heinecke*, Dhiraj D. Kalamkar*, Xing Liu+,
Md. Mosotofa Ali Patwary*, Yutong Lu#, and Pradeep Dubey*

*Parallel Computing Lab, Intel Corporation
+Georgia Institute of Technology, USA
#National University of Defense Technology, China

Multi-node Results Analysis

Computer	# of Nodes	# of MPI Ranks	MPI Parallel. Efficiency	All-reduce time[1)]	Halo wait time
TH2	15,360	46,080	82%	8.4%	~9.5%[2)]
SuperMUC	9,216	18,432	84%	12.9%	3.8%
Edison	5,555	11,110	95%	1.9%	3.4%
Curie thin[3)]	5,003	80,048	91%	11%	N/A
Occigen	2,082	4,164	N/A	5.9%	5.4%
1/3 of Stampede	2,048	6,144	72%	22.0%	4.7%

Very high parallelization efficiency from Cray Aries network with Dragonfly topology

1) Average all-reduce time. Includes load-imbalance
2) From a 7680-node run.
3) CEA implementation

Multi-node Results Analysis

Computer	# of Nodes	# of MPI Ranks	MPI Parallel. Efficiency	All-reduce time[1)]	Halo wait time
TH2	15,360	46,080	82%	8.4%	~9.5%[2)]
SuperMUC	9,216	18,432	84%	12.9%	3.8%
Edison	5,555	11,110	95%	1.9%	3.4%
Curie thin[3)]	5,003	80,048	91%	11%	N/A
Occigen	2,082	4,164	N/A	5.9%	5.4%
1/3 of Stampede	2,048	6,144	72%	22.0%	4.7%

High scalability using pure-MPI parallelization with Bullx MPI
Pure-MPI: better cache locality, easier to program

1) Average all-reduce time. Includes load-imbalance
2) From a 7680-node run.
3) CEA implementation

Discussion

- Penalize slower convergence (~10%) with more MPI ranks per processor
- Allow MPI-like decomposition in OpenMP
- Penalize prime number of nodes
- Flexible decomposition for heterogeneous computing (e.g., Stampede)
- Make input less regular
- Report halo exchange time
- Measure power and compare with HPL to see energy proportionality
- …

Summary: HPCG benefits from

- High BW memory
- Caches (with fusion optimization)
- Fast core-to-core communication
- High-radix low-diameter network

Intel Optimized Technology Preview for HPCG

https://software.intel.com/en-us/articles/intel-optimized-technology-preview-for-high-performance-conjugate-gradient-benchmark

Xeon and Xeon Phi versions

Notice and Disclaimers

INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL® PRODUCTS. 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. Intel products are not intended for use in medical, life saving, or life sustaining applications. Intel may make changes to specifications, product descriptions, and plans at any time, without notice.

All products, dates, and figures are preliminary for planning purposes and are subject to change 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.

Performance tests and ratings are measured using specific computer systems and/or components and reflect the approximate performance of Intel products as measured by those tests. Any difference in system hardware or software design or configuration may affect actual performance.

The Intel products discussed herein 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.

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 by visiting Intel's website at http://www.intel.com.

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

17

Notice and Disclaimers Continued …

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 to 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. For more information go to http://www.intel.com/performance

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. 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. Notice revision #20110804

Q&A