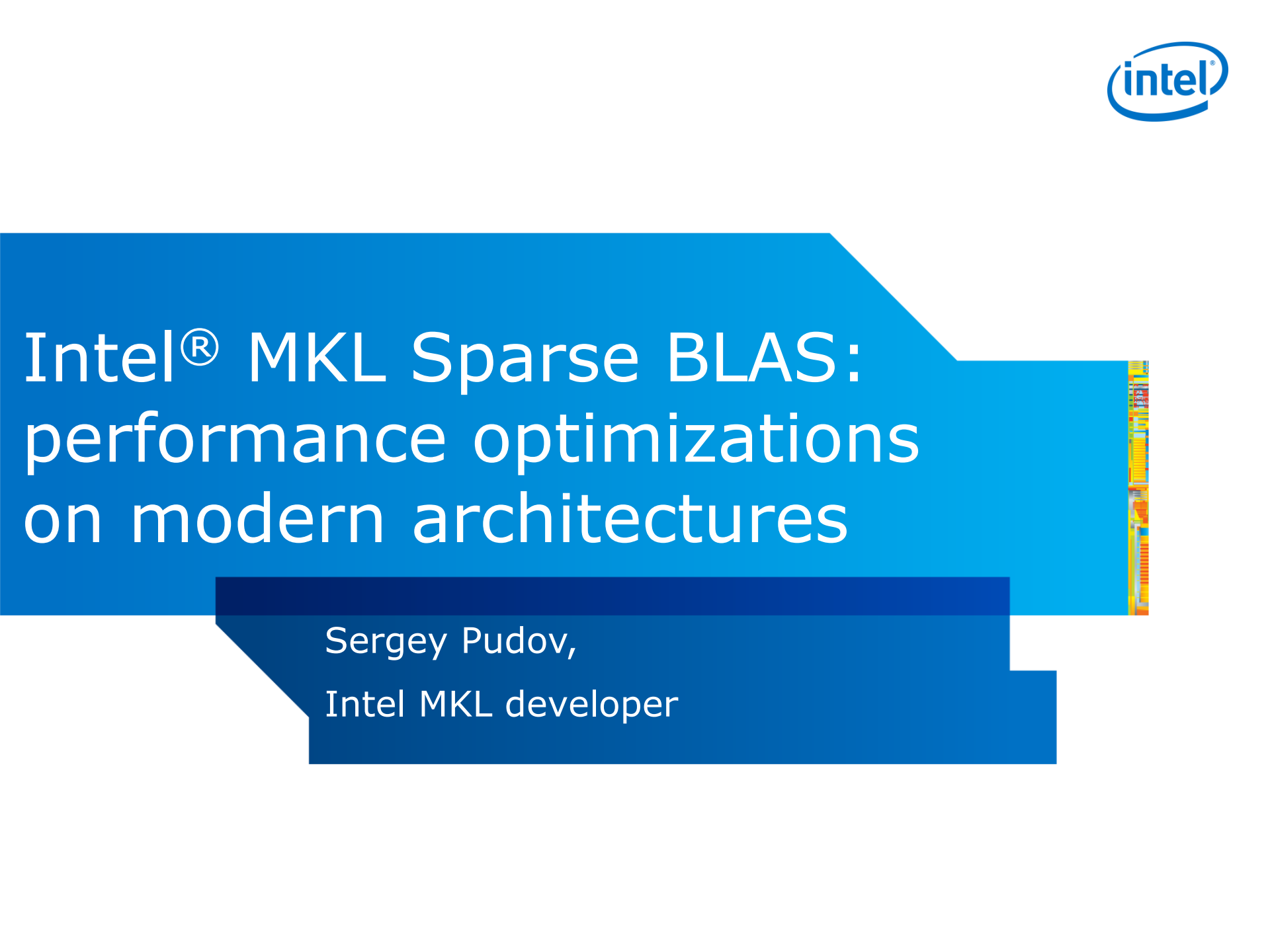


Intel® MKL Sparse BLAS: performance optimizations on modern architectures

Sergey Pudov,

Intel MKL developer

Intel® MKL Sparse BLAS: introduction

Intel MKL Sparse BLAS supports 6 sparse formats: CSR, CSC, BSR, DIA, COO, and SKY. It is designed mainly for applications where the computations are done a few times.

Every function calculates the result in a single call, which includes simple matrix analysis and execution steps.

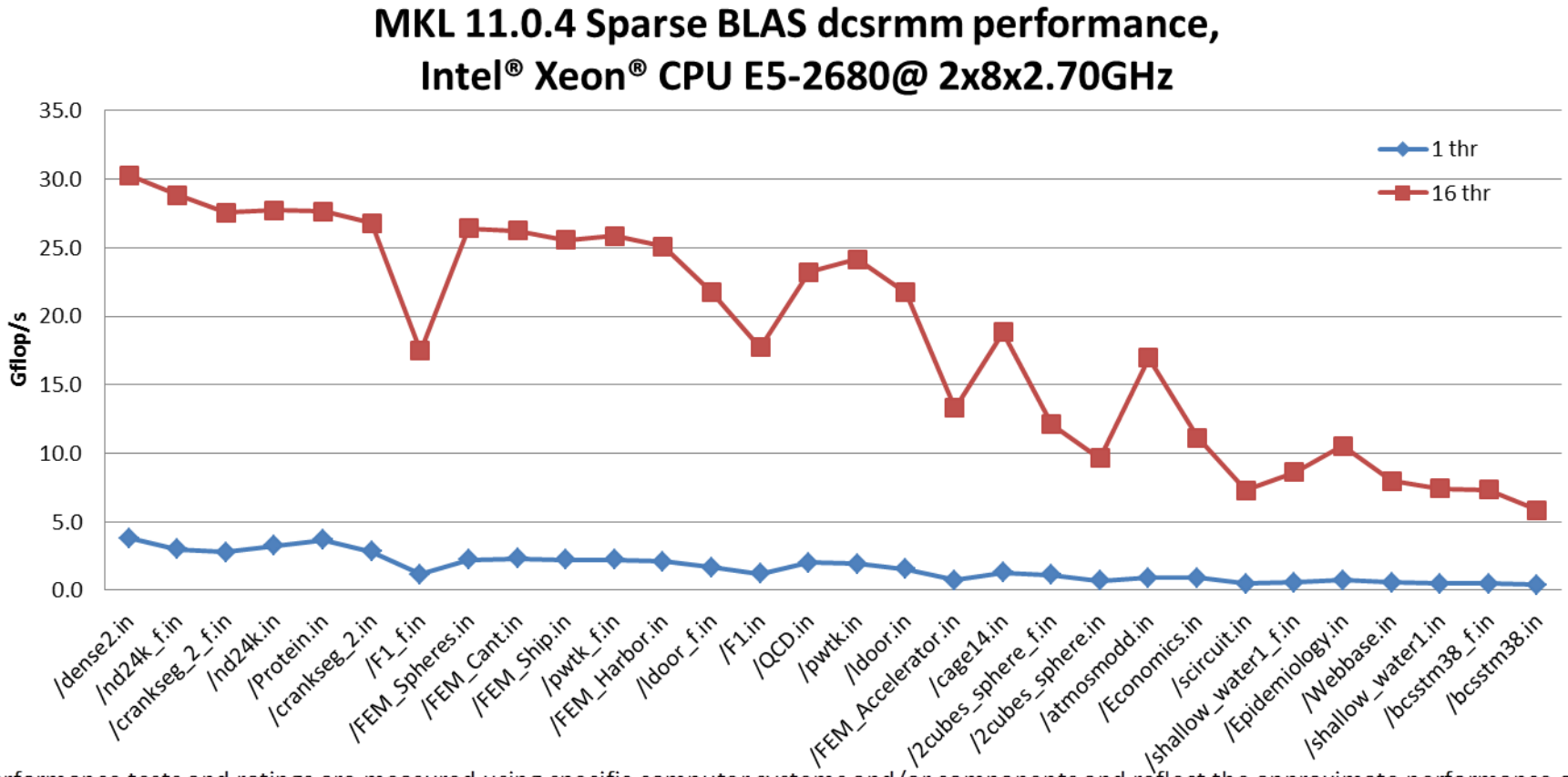
Deep investigation of the sparse matrix pattern is not performed because it is a time consuming operation that affects the performance.

Example:

mkl_?csrmv(&transa, **&m, &k**, &alpha, **matdescra**, **val, indx, pntrb, pntre**, x, &beta, y);

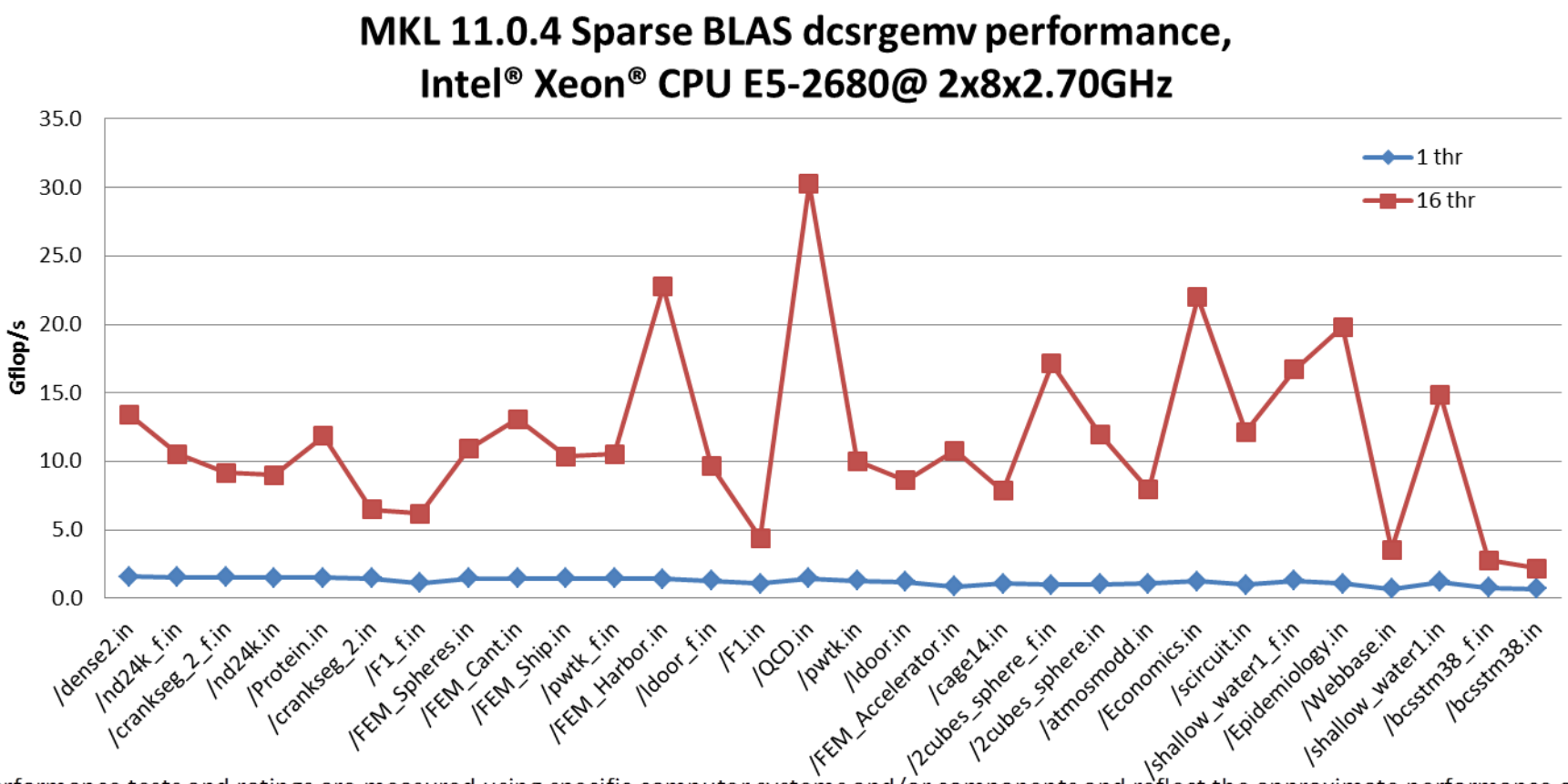
Note that **matdescra** in general does not describe the matrix, it only describes the way in which the matrix should be processed by the routines.

Performance of Intel® MKL CSR SpMM



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. Buyers should consult other sources of information to evaluate the performance of systems or components they are considering purchasing. For more information on performance tests and on the performance of Intel products, refer to
http://www.intel.com/content/www/us/en/benchmarks/resources-benchmark-limitations.html
Refer to our Optimization Notice for more information regarding performance and optimization choices in Intel software products at:
http://software.intel.com/en-ru/articles/optimization-notice/
*Other brands and names are the property of their respective owners.

Performance of Intel® MKL CSR SpMV



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. Buyers should consult other sources of information to evaluate the performance of systems or components they are considering purchasing. For more information on performance tests and on the performance of Intel products, refer to
http://www.intel.com/content/www/us/en/benchmarks/resources-benchmark-limitations.html
Refer to our Optimization Notice for more information regarding performance and optimization choices in Intel software products at:
http://software.intel.com/en-ru/articles/optimization-notice/
*Other brands and names are the property of their respective owners.

Two-step Computation Approach

It is known that for best performance, computational kernels and the workload balancing algorithm should depend on the structure of the matrix.

When multiple calls are expected with a particular sparse matrix pattern, it is better to organize computations in two steps:

- ***Analysis***, which chooses the best kernel and workload balancing algorithm for a given computer architecture.
- ***Execution***, where the information from the previous step is used to get high performance.

We can try to use this approach especially since the time required for a single analysis step is usually less than the overall performance benefit from multiple execution steps.

Limitations imposed by the single-call Sparse BLAS interfaces are mostly visible on modern architectures with multiple cores where even small workload imbalance may result in a significant performance deficiency.

Experimental API

Experimental library for an Intel® Xeon Phi™ coprocessor contains some SpMV functionality with two-step interface to investigate performance benefits of this approach. The library supports:

- Two formats: CSR and ESB*
- A couple of workload balancing algorithms

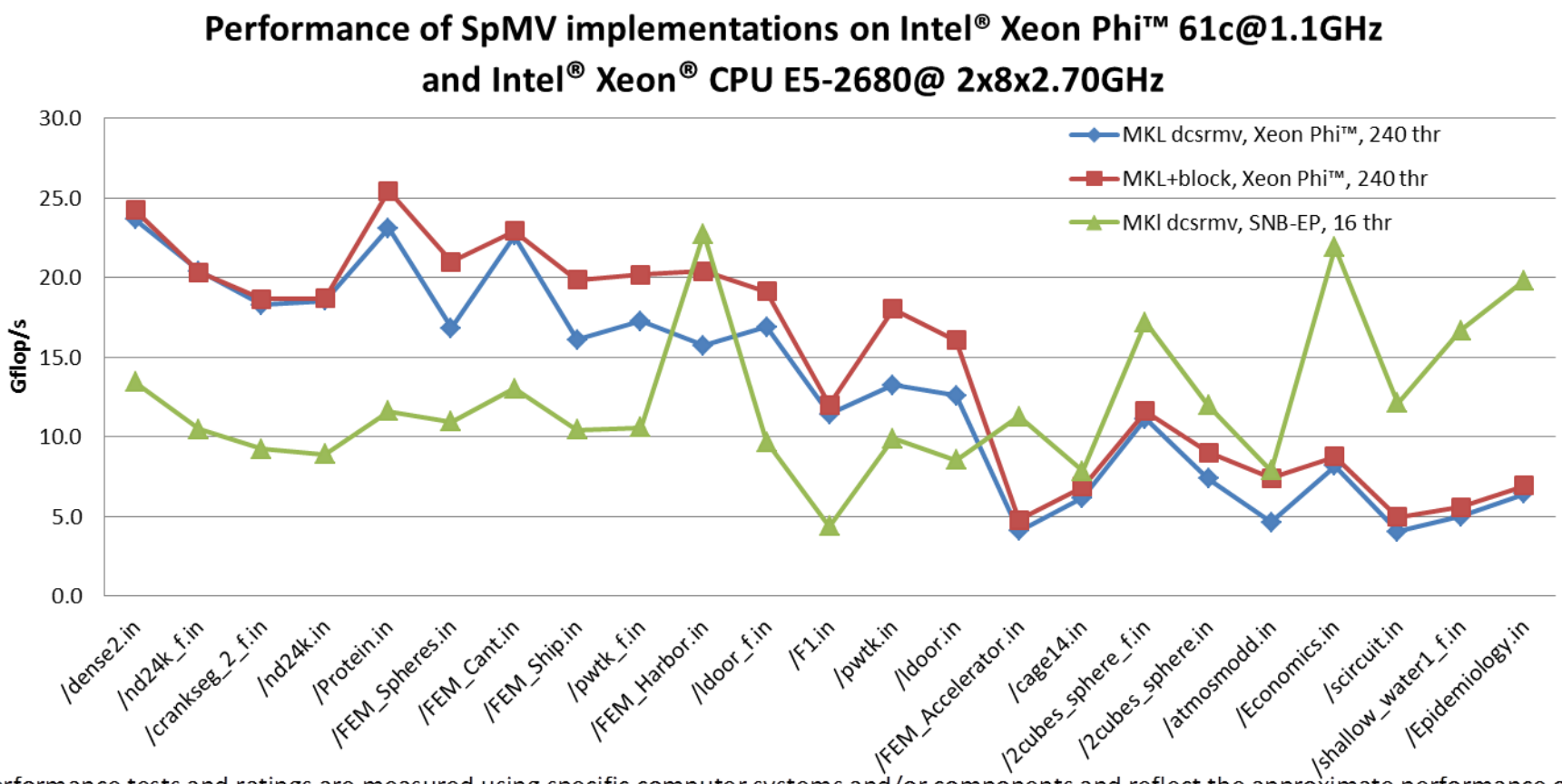
The goal of the experiment is to collect early feedback. The library will be available via request to intel.mkl@intel.com after release.

Please send your feedback to

intel.mkl@intel.com or sergey.g.pudov@intel.com

* Xing Liu, Mikhail Smelyanskiy, Edmond Chow and Pradeep Dubey. Efficient Sparse Matrix-Vector Multiplication on x86-based Many-core Processors. - ICS'13, June 10–14, 2013, Eugene, Oregon, USA.

Experimental library: performance



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. Buyers should consult other sources of information to evaluate the performance of systems or components they are considering purchasing. For more information on performance tests and on the performance of Intel products, refer to
http://www.intel.com/content/www/us/en/benchmarks/resources-benchmark-limitations.html
Refer to our Optimization Notice for more information regarding performance and optimization choices in Intel software products at:
http://software.intel.com/en-ru/articles/optimization-notice/
*Other brands and names are the property of their respective owners.

Thank You

Legal Disclaimer & Optimization Notice

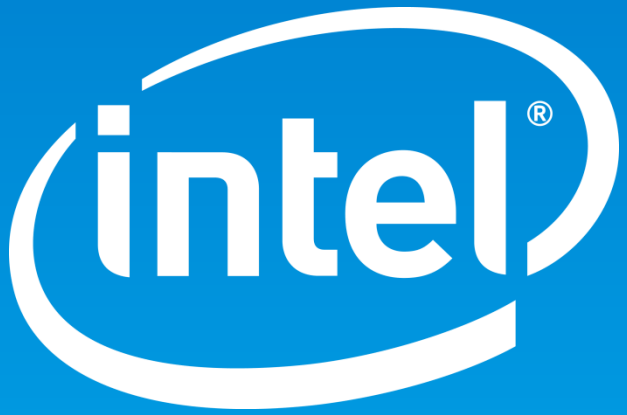
INFORMATION IN THIS DOCUMENT IS PROVIDED “AS IS”. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. INTEL ASSUMES NO LIABILITY WHATSOEVER AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO THIS INFORMATION INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT.

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.

Optimization Notice

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 SSSE3 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


intel