Health and Life Sciences at Intel

Marie-Christine Sawley
IPAG-EU
Intel Corporation
(with the help of Kristina Kermanshahche, Chief Architect, Health & Life Sciences)

Open Lab Healthcare workshop
CERN Nov. 11th, 2014
LEGAL 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.

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

- Relative performance is calculated by assigning a baseline value of 1.0 to one benchmark result, and then dividing the actual benchmark result for the baseline platform into each of the specific benchmark results of each of the other platforms, and assigning them a relative performance number that correlates with the performance improvements reported.

- Intel does not control or audit the design or implementation of third party benchmarks or Web sites referenced in this document. Intel encourages all of its customers to visit the referenced Web sites or others where similar performance benchmarks are reported and confirm whether the referenced benchmarks are accurate and reflect performance of systems available for purchase.

Intel® Turbo Boost Technology requires a Platform with a processor with Intel Turbo Boost Technology capability. Intel Turbo Boost Technology performance varies depending on hardware, software and overall system configuration. Check with your platform manufacturer on whether your system delivers Intel Turbo Boost Technology. For more information, see http://www.intel.com/technology/turboboost

Intel processor numbers are not a measure of performance. Processor numbers differentiate features within each processor series, not across different processor sequences. See http://www.intel.com/products/processor_number for details. Intel products are not intended for use in medical, life saving, life sustaining, critical control or safety systems, or in nuclear facility applications. All dates and products specified are for planning purposes only and are subject to change without notice

Intel product plans in this presentation do not constitute Intel plan of record product roadmaps. Please contact your Intel representative to obtain Intel's current plan of record product roadmaps. Product plans, dates, and specifications are preliminary and subject to change without 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
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 © 2014 Intel Corporation. All rights reserved. Intel, the Intel logo, Xeon, Xeon logo, Xeon Phi, and Xeon Phi logo are trademarks of Intel Corporation in the U.S. and/or other countries. All dates and products specified are for planning purposes only and are subject to change without notice.
Intel Corporation
The World’s Largest Semiconductor Manufacturer

- Leading Manufacturer of Computer, Networking & Communications Products
- Founded by Gordon Moore and Robert Noyce in 1968
- Headquartered in Santa Clara, California
- $52.7B in Annual Revenues - 25+ Consecutive Years of Positive Net Income
- 170 Sites in 66 Countries
- Over 107,000 Employees – 84,600 technical roles, 10,200 Masters in Science, 5,400 PhDs, 4,000 MBAs
- Named one of the Top Ten Most Valuable Brands in the World by Interbrand
- Ranked #42 on Fortune’s World’s Most Admired Companies
- Largest Voluntary Purchaser of Green Power in the United States for 6 years in a row
- Invests $100 Million Each Year in Education Across More than 100 Countries
- 4 Million Hours of Volunteer Service toward improving education over the past decade
Life Sciences: At the intersection of transformative forces

**HPC / Big Data**

10^18

Enabling exascale computing on massive datasets

**Cloud**

Helping enterprises build open interoperable clouds

**Commercial & Open Source**

Contributing code and fostering ecosystem
Trends & Challenges in Life Sciences

**Big Data in Life Sciences**
- Sequencer advances – 4x data in 50% less time .5TB/device/day
- 4D molecular imaging produces 2TB/device/day

**Burdens of Data Management**
- Store, manage, share, ingest and move PBs of research & clinical data
- Need to reliably 'snapshot' pipelines with archive to tiered storage

**Innovation Drives Change**
- Rapid iteration of algorithms far outpace IT, requiring flexibility, agility
- Most applications do not fully leverage available infrastructure

**Converged Infrastructure**
- Workloads converging between local and cloud-based HPC/Big Data
- Advanced orchestration required to maximize throughput & efficiency
Intel Assets for Life Sciences

**Intel Xeon E5**
- Up to 80% greater performance
- Up to 70% more energy efficiency
- Up to 30% less network latency
- Hardware-accelerated security (AES-NI)
- Broad industry adoption

Consistent Performance Gains each generation

**Intel Xeon Phi**
- Performance and programmability for highly-parallel workloads
- Programming continuity and scalable parallel programming models: common source code and software tools between multicore Intel® Xeon® and manycore Intel® Xeon Phi™
- Partner ecosystem continues growing and making progress

**Intel Software**
- Intel® Cluster Studio XE compilers, libraries, analysis tools, OpenMP and MPI
- Apache Hadoop*
- Intel Analytics Library
- Intel® Data Center Manager and Intel® Node Manager (NM)
- Intel® Expressway Service Gateway for Cloud usage models

**Intel Fabric**
- Intel® True Scale Fabric designed from the ground up for HPC
- QDR-40 and QDR-80 deliver performance that scales - high MPI message rates and end-to-end latency that stays low at scale
- Optimized support for Intel® Xeon® E5 and Xeon® Phi processors

**Intel Storage**
- Intel® Xeon® processors and platforms are enabled with beneficial storage optimizations
- Solid State Drives (SSD) and other NVM technologies improve storage performance
- Intel® Cache Acceleration Software
- Intel’s open source Lustre file-system support/development and Chroma management/provisioning tools

*Other names and brands may be claimed as the property of others.
Optimizing Top Applications and Pipelines
Intel working with industry experts worldwide

- Genomics, Molecular Dynamics and Molecular Imaging applications targeting both Intel® Xeon® processors and Xeon® Phi™ coprocessors
- Fine- and coarse-grained optimization at the node and cluster level
- Work with code authors to release optimizations, disseminate best practices

*Some names and brands may be claimed as the property of others.
Intel Genomics & Health Analytics Appliances

Scale through independent solutions, each targeting a different segment & usage model

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

*Other names and brands may be claimed as the property of others.
Dell Genomics Data Analytics Platform

- **Challenge**: Developing analysis pipelines for novel areas of cancer research takes weeks with current infrastructure, delays scientific discovery and clinical advances

- **Preliminary Results**:
  - Preprocessing stage reduced from 2 hrs to 20 mins
  - Custom genome-wide R statistical analysis pipeline reduced from 20 hrs to 3 hrs 40 min

- **Solution**: Dell* Genomics Data Analytics Platform
  - Single Rack Solution with balanced HPC compute, fabric and storage
  - 9 Teraflops of Intel® Xeon® E5 v2 processors
  - Intel® Enterprise Edition for Lustre*
  - Intel® Cluster Studio XE

- **Benefits**: Rapid iteration during early algorithm development improves scientific accuracy, drives innovation in cancer research

*Other names and brands may be claimed as the property of others.
Charité “Real-time” Cancer Analysis – Matching proper therapies to patients using in-memory techniques

• **Challenge:** Real-time analysis of cancer patients using in-memory SAP HANA* Oncolyzer database running on Intel® Xeon® family infrastructure.

• Up to 3.5M data points & 20TB data/patient

• **Solution:** Using structured and unstructured data to collect and analyze tables used to take up to two days -- now takes seconds

• **Benefits:** Improves medical quality in disruptive way for Patient, Doctor, Hospital, Research

*Other names and brands may be claimed as the property of others.*
Secure Healthcare Cloud: Medical Imaging

“The data-intensive nature of our cloud-based PACS service places great demands on our data center infrastructure. With the Intel® Xeon® processor E5 family, we can significantly boost the processing output that can be achieved with each server. This will help us continue to deliver the scalable, high-performance platform we need to meet rising user demand while maintaining the quality of our service.”

-Seffi Markov, Director of PACS R&D, Carestream Health

- Carestream Health – Cloud-based medical imaging service (PACS, VNA, Universal Imaging Viewer)
- **Challenge**: Create a powerful, scalable data center platform that can accommodate more users without compromising service quality
- **Solution**: Intel Xeon E5 processors – to handle more users and data-intensive workloads. Intel® Advanced Vector Extensions (AVX) for parallel processing and high memory availability enabled quicker image rendering
- **Benefits**:  
  - Process images >28% faster
  - Handle 24% more users


*Other names and brands may be claimed as the property of others.*
High Throughput Science: Large Scale Computational Chemistry Simulation

**Virtual Screening**

- **Challenge:** Sustaining access to 50,000+ cores for large scale computational chemistry simulation results in under a week. Ability to monitor and re-launch jobs, no additional capital expenditure with internal HPC already running at capacity.

- **Solution:** Novartis leveraged software from AWS partner, Cycle Computing, and MolSoft to provision a fully secured cluster of 30,000 CPUs, powered by the Intel® Xeon® processor E5 family.

  - Completed screening of 3.2 million compounds in approximately 9 hrs, compared to 4 - 14 days on existing resources.

*Other names and brands may be claimed as the property of others.
High Performance Scale-Out Storage

“If you need 10,000 cores to perform an extra layer of analysis in an hour, you have to scale a significant cluster to get answers quickly. You need a real solution that can address everything from very small to extremely large data sets.”

Dr. Tim Cutts
Acting Head of Scientific Computing
Wellcome Trust Sanger Institute

- **Challenge:** Delivering on extremely high service levels required to store, manage, access and archive massive amounts of research data
- **Solution:** 22PB DDN SFA® high-performance storage engine and EXAScaler™ Lustre® appliance, backed with Intel® Global Lustre* Support
- **Benefits:** Flexible scaling ensures that Sanger Institute has sufficient storage performance to support downstream analysis, which is difficult to predict and varies by workload and project

*Some names and brands may be claimed as the property of others.*
BLAST Bioinformatics Scale-out

**Problem Statement:**
Back in 2008 a genome research team faced compute and scalability issue in comparing all pairs of 4M proteins, the BLAST search results overwhelmed a single database table. Today they need to compare 14M proteins, **this requirement cannot be addressed with existing technology**

**Solution:** Apache Hadoop, Map Reduce, Hbase, Hive

**Benefits:** Ability to compare 14+M proteins, reducing processing time from days to hours

**Project Characteristics:**
- **Hadoop:** 5 nodes Cluster
- **Storage:** 16TB (Internal storage) per server
- **Servers:** Xeon E5 2 socket 8 cores, 64GB RAM
- **SLA:** reduced processing time from 30 days to less than a day and scale to 4x4 million samples comparison

**Data:** Multi-Terabyte database

*Other names and brands may be claimed as the property of others.*
Intel Exascale Labs — Europe

Strong Commitment To Advance Computing Leading Edge:

*Intel collaborating with HPC community & European researchers*

*4 labs in Europe - Exascale computing is the central topic*

- **ExaScale Computing Research Lab, Paris**
  - Performance and scalability of Exascale applications
  - Tools for performance characterization

- **ExaCluster Lab, Jülich**
  - Exascale cluster scalability and reliability

- **ExaScience Life Lab, Leuven**
  - HPC for Life Science
  - Genomics, Biostatistics

- **Intel and BSC Exascale Lab, Barcelona**
  - Scalable RTS and tools
  - New algorithms

*www.exascale-labs.eu*

*Some names and brands may be claimed as the property of others.*
GATE: Geant4 Application for Tomographic Emission

- **OpenGATE international collaboration**
  - 20 laboratories
  - ~60 scientists
  - Technical coordinator: Sébastien JAN - CEA
  - Spokesperson: Irène Buvat - INSERM

- **GATE software**
  - First developments: 2002
  - General simulation platform for emission tomography

  **First public release: 3 mai 2004**
  *Today: more than 900 users – 16 releases – version 6.0.0*

- **General configuration**
  - Isotope & Source Selection
  - Activity
  - Acquisition parameters
    - energy resolution
    - Detector & Source Movements
    - Time parameters
  - Detector modelling
  - Physical processes

---

Ziad EL BITAR – IPHC

Ref: http://indico.cern.ch/event/277160/session/2/contribution/7/material/slides/0.pdf
Focus on the ExaScience Life Lab
Focus on the ExaScience Life Lab

HIGH PERFORMANCE COMPUTING FOR BIG DATA ANALYTICS IN LIFE SCIENCES

- **Optimize Pharma Codes for Current and Future Architectures**
  - **Big Data Analytics ~> Personalized Medicine**
    - ChemoGenomics (Machine Learning)
    - DNA/RNA Analytics (Visual Analytics)
    - High Throughput Screening (Image Analysis)
    - BioStatistics (Bayesian Networks)

- **Using Current and Future (Disruptive) Technology**
  - HPC in the Cloud
    - Localized HPC Services
  - Accelerators
    - Many-Core Architectures (Xeon Phi)
    - Xeon & FPGA
  - Next Gen Memories (Persistent Memories, Memory Cubes)
Matching Technology and Research in Biology

<table>
<thead>
<tr>
<th>Life Sciences</th>
<th>DNA / RNA Sequencing</th>
<th>Adaptive Trial Design</th>
<th>Chemo Genomics</th>
<th>High Throughput Screening</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bio Data Mining</td>
<td>Real Time HPC Visual Analytics</td>
<td>Bio Statistical Models</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communication Avoiding Algorithms</td>
<td>HPC Machine Learning</td>
<td>HPC ODE/PDE Kernels</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Programming Language Composition</td>
<td>Dynamic Load Balancing</td>
<td>HPC Cloud Programming Models</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Architectural Simulator</td>
<td>Next Gen Architectures</td>
<td>Next Gen OffCore Architectures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lifescience specific ‘Si’-architecture</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The Computational Radiology Laboratory at Boston Children’s Hospital, An Intel Parallel Computing Center

Technical goals:

• Improve cache performance, vectorization performance and multi-threading performance for Xeon™ and Xeon Phi™

• Improved data structures

• Improved algorithms

• Open source implementations

⇒ Modernize medical image computing
Intel’s Vision for Personalized Medicine

- Payer
  - Fraud
  - Population risk management
- Provider
  - Imaging
  - Evidence-based clinical decision support
- Life Sciences
  - Genomics
  - Bioinformatics
  - Clinical Trials
- Patient
  - Wellness
  - Population risk management
- Accountability
- Decision Transparency
- Treatment Value
- Personalization
Summary

• Enabling ecosystem of partners to innovate into Health and Life Sciences

• Delivering hardware-enhanced capabilities and solutions to accelerate science, translate results, deliver today.

• **Looking for collaboration opportunities for Exascale class computing solution by 2020**

• **Contact:**
  - kristina.m.kermanshahche@intel.com
  - Claudio.Bellini@intel.com
  - Marie-Christine.sawley@intel.com