Big Data: An Opportunity Knocking on the Doors of Computing

Dr. Pradeep K. Dubey
Intel Fellow and Fellow of IEEE
Director, Parallel Computing Lab
Systems and Software Research, Intel Labs

April 22\textsuperscript{nd} 2014

\textit{Data-data everywhere, not a bit of sense!}
What You Will Hear From Me Today

- Applications of tomorrow will be increasingly about enabling complex analytics real-time

- There is an underlying *common* computational core that will become the key enabler for emerging needs of both HPC and Big Data

- Recent results:
  - Approximate Computing: An opportunity for architecture-aware co-design
Enabling A Data Driven Science

Traditional HPC

From Observations To Postulates

From Postulates To Physical Laws

From Physical Laws To Predictions

$F_d = 6\pi \mu R v$

Big Data HPC

Data Driven Model

HPC

TRANSFORMING DATA INTO KNOWLEDGE
“The great strength of computers is that they can reliably manipulate vast amounts of data very quickly. Their great weakness is that they don’t have a clue as to what any of that data actually means.”


A New Architectural Paradigm for the Tera Era ... Instead of a chip with a single execution unit, Intel is developing a “many-core” chip architecture (with potentially hundreds of compute threads). Imagine instead of one processor, there are 16, 32, 64, and so on, up to perhaps 256 processor cores on a single die.”
Decomposing Emerging Applications

Mining
Is it …?

Synthesis
What if …?

Evaluation
Data Mining
Web Mining
Semantic Search
Streaming Data Mining
Distributed Data Mining
Content-based Image Retrieval
Query By Humming
Video Mining
Intrusion Mining
Physical Simulation
Strategy Simulation
Audio Synthesis
Video/Image Synthesis
Clustering / Classification
Bayesian Network
Markov Model
Decision Trees
Forests of Trees
Neural Networks
Probabilistic Networks
Optimization-based models: Linear/Non-linear/Stochastic
Time-series models

Clustering / Classification
Bayesian Network
Markov Model
Decision Trees
Forests of Trees
Neural Networks
Probabilistic Networks
Optimization-based models: Linear/Non-linear/Stochastic
Time-series models

Modeling (Recognition)
What is …?

Execution

Mining or Synthesis Quality Determines Model Goodness
Models: A deeper look

- Model
  - Traditional HPC
  - Procedural or Analytical
    - Data-Driven
  - Big Data
    - Irregular Data Access
      - Such as: Unstructured Grids, Graphs ...
    - Sparse Solver

Big Data: Common Link Between Scientific and Enterprise Computing
Sparse Solvers: Common Link Between HPC and Big Data Computing
Compute Platform Abstraction

Edge Computing
(Clients)

Private data, sensory inputs, streams/feeds
immersive 3D graphics output, interactive visualization

Big Data Analytics
(Servers)

Intersection of massive data with massive compute
real-time analytics, massive data mining-learning

Architectural Implications Are Radical!
Big Data Computing: Opportunity

• Data-driven models are now tractable and usable
  - We are not limited to analytical models any more
  - No need to rely on heuristics alone for unknown models
  - Massive data offers new algorithmic opportunities
    - Many traditional compute problems worth revisiting

• Web connectivity significantly speeds up model-training

• Real-time connectivity enables continuous model refinement
  - Poor model is an acceptable starting point
  - Classification accuracy improves over time

Can compute tame the beast of massive, unstructured, dynamic datasets to enable continuous-time simulation and analytics?
What makes it hard and fun 😊

- Less regular to highly irregular computation
  - Data-movement dominated execution
- Very large working set arising from complex models
  - Often far exceeding small caches in modern processors
- Exploration of compute-communication tradeoff space
  - Communication-avoiding algorithms
- Exploration of relevant approximate algorithm space
  - Trading off lower time-complexity for increased irregularity and space complexity
- Integration of data-management with analytics
- Exploiting growing cores/threads/SIMD and NVM
Explosion of Connected “End Points”

Mobile Devices  Internet of Things  Sea of Sensors

50B by 2020
CONNECTED ENDPOINTS

Source: Cisco
Making sense of one petabyte

50x
To read in Library of Congress

13y
To view as HD Video

11s
To generate in 2012

Compute Doubles Every 1.5 Year
Data Doubles Every Year

Virtuous cycle of data

Richer data to analyze

CLOUD

Richer data from devices

INTELLIGENT SYSTEMS

Richer user experiences

CLIENTS
Supercomputing 2013: Intel® Xeon Phi™ based TH-2 becomes Top500 #1

More than 50PF Peak
Koomey’s Law and Exascale Goal

2011 checkpoint:
Top500 #1 machine: ~10PF, ~10MW → 1 GF/W

Achieving EF at 50GF/W →

As per Koomey’s Law, this should happen in 9+ years → by decade end

2013 checkpoint: ~ 2GF/W
Some Recent Results and Ongoing Research
Many Core makes sense for workloads with high enough “P“-parallel component - for simplicity, we call these Highly Parallel

\[ S = \frac{1}{(1 - P) + \frac{P}{N}} \]

For \( S \geq 1 \), \( P \geq \frac{N(K_N - 1)}{NK_N - 1} \)

S = speedup, P = parallel fraction, # of Cores = N, Kn = single thread performance (single core/multicore)
Next Intel® Xeon Phi™ Processor: 
Knights Landing

Designed using Intel’s cutting-edge 14nm process

Not bound by “offloading” bottlenecks

Standalone CPU or PCIe coprocessor

Leadership compute & memory bandwidth

Integrated on-package memory

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

Applications:
- Twitter Search
- Genomics
- Quantum Chemistry
- Path Planning

Complex Analytics:
- Machine Learning
  - Kmeans, LSH, DBN, SGD, DBSCAN...
- Numeric Computing
  - Regression, Covariance...

Data Management:
- Database Primitives
  - Join, Select, Sort...
- Graph algorithms
  - BFS, SSSP, BC...

Languages:
- Scripting (R or Python or ...) Vs. C/C++ with OpenMP

Frameworks:
- Postgres
- Column store
- SciDB
- Graphlab
- KDT Combblas
- Giraph
- Galois

Libraries:
- Intel MKL

Runtime:
- MPI+X Vs. SWARM/ParalleX/OCR

Hardware:
- Multicore
- Manycore
- MCDRAM
- Persistent Memory
- Fabric
- Accelerators
Encouraging New Results

FFT*  
QCD*

* See Backup slides for benchmark and system details

Order of Magnitude Speedup Only Happens at Algorithm-Level
Three Essentials: LCA (Locality, Communication, Approximation)

Intel® Xeon® and Xeon Phi™: both benefit from a common-set of manycore-friendly optimizations

* See Backup slides for benchmark and system details
Approximate Computing Using: Locality Sensitive Hashing*

- Problem: Find nearest neighbors in a high dimensional space
- Known technique for computing approximate nearest neighbors
- Idea: Use “locality sensitive” hash functions to probabilistically aggregate nearby points
Applying LSH: Searching Twitter

- 400 million+ tweets per day
- Response time should be in tens of ms
- Streaming data, Dynamic updates
- LSH reduces search complexity from $O(N)$ to $O(N^{0.5})^[*]$
  - Fewer data accesses, but less predictable
  - Super-linear memory requirements $O(N^{1.5})$

LSH $\rightarrow$ PLSH: When naïve parallelism is not enough

• Primary innovations:
  - Manycore-friendly implementation of an irregular algorithm like LSH that is naïvely parallel, but then memory-bound
  - A model that predicts for a given platform: # of hash tables and table size for optimal throughput, under accuracy-latency constraints:
    ➢ Accuracy controlled: $1 - \left(1 - p^k\right)^L \geq 1 - \delta$
  - Insert-optimized delta-table and merging heuristic

• Almost order-of-magnitude improvement in single-node performance
  - While maintaining the sub-second latency, accuracy of over 90%, and the highly dynamic nature of this application stream
PLSH: Performance*

- **Intel® Xeon® performance**
- Over 8x single-node improvement

Intel® Xeon Phi™ Offload speedup: 1.95x from one card

* "Streaming Similarity Search over one Billion Tweets using Parallel Locality Sensitive Hashing”. Narayanan Sundaram, Aizana Turmukhametova (MIT), Nadathur Satish, Todd Mostak (Harvard), Piotr Indyk (MIT), Samuel Madden (MIT), and Pradeep Dubey; To appear at VLDB 2014. System configuration details in ‘Backup’ section.

* See Backup slide 31 for benchmark and system details
Micro Array Data

$10^{8-10}$

Gene Identifiers  
Sample Type  
Sample Class Labels  
Normalized Gene Expression Levels
Consider: $10^{4-5}$ gene expressions per sample, up to $10^{8-10}$ samples (multiple samples/patient)

Implies a problem that scales to $10^{2-5}$ nodes of a large cluster with each node handling $10^{4-5}$ samples
Learning gene regulatory networks

Gene expression data – measurement of mRNA levels of all genes

Samples (e.g. experimental conditions, Human individuals)

Goal: Infer the gene regulatory network that controls gene expression

- Representation: Bayesian network
- Each node represents a gene $e_i$
- Edge indicates 2 genes are conditionally dependent
- The distribution on C is fully determined by C's parents (A and B)
  $\Rightarrow$ A and B regulate the expression of C

Potential GRN Challenge Problem

• Arabidopsis whole-genome network
  - Model plant organism that is heavily researched
  - 22,000+ genes (NP-hard on this dimension)
  - 10,000+ microarray experiments (large-scale, but still sparse data)
  - Each experiment has expression value of every gene (matrix of genes vs. experiments)
    - Over 220 million elements
Challenge: Extreme Complexity

Vertex view

Don’t have regularized structure
Overlapping communities
We are at an unprecedented convergence of massive compute with massive data ...

This confluence will have a lasting impact on both how we do computing and what computing can do for us!
Legal Information

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.

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.

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
All products, computer systems, dates, and figures specified are preliminary based on current expectations, and are subject to change without notice.
Intel processor numbers are not a measure of performance. Processor numbers differentiate features within each processor family, not across different processor families. Go to: http://www.intel.com/products/processor_number

Intel, processors, chipsets, and desktop boards 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.


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

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

Copyright © 2013, Intel Corporation. All rights reserved.
Back Up
Additional Information Regarding Reported FFT Performance Data

• For benchmark details, refer to:
  "Tera-Scale 1D FFT with Low-Communication Algorithm and Intel® Xeon Phi™ Coprocessors", Jongsoo Park, Ganesh Bikshandi, Karthikeyan Vaidyanathan, Daehyun Kim, Ping Tak Peter Tang, Pradeep Dubey; to appear at Supercomputing ‘13.

• System used: TACC Stampede Cluster
  - MPSS: Gold/Update3
  - Compiler: 13.1.0 Rev 2.146 Build 20130121
  - MPI: 4.1.0.030
  - Intel® Xeon Phi™ KNC card: 61c 1.1GHz SE10P
  - Host: Intel® Xeon®: E5-2680 2.7GHz 32GB, DDR3 1.6GHz
  - Fabric: Mellanox, FDR IB, 56 Gb/s
Additional Information Regarding Reported QCD Performance Data

• For benchmark details, refer to:
  “Lattice QCD on Intel® Xeon Phi™ coprocessors” Balint Joo (Jefferson Labs), Dhiraj Kalamkar, Karthikeyan Vaidyanathan, Mikhail Smelyanskiy, Kiran Pamnany, Victor Lee, Pradeep Dubey & William Watson (Jefferson Labs); International Super Computing Conference 2013 (ISC’13)

• System used: TACC Stampede Cluster
  - MPSS version is 2.1.6720-13 and the flash version is 2.1.02.0386.
  - Compiler: 13.1.0 Rev 2.146 Build 20130121
  - MPI: 4.1.0.030
  - Intel® Xeon Phi™ KNC card: 61c 1.1GHz ES2-P/A/X 1750
  - Host: Intel® Xeon® E5-2680 2.7GHz 32GB, DDR3 1.6GHz
  - Fabric: FDR Infiniband, Mellanox ConnectX host adapters with a maximum peak bandwidth of up to 7 GB/s.
  - OFED version is v1.0-ofed1.5.4
Additional Information Regarding Reported PLSH Performance Data

• For benchmark details, refer to:
  “Streaming Similarity Search over one Billion Tweets using Parallel Locality Sensitive Hashing”. Narayanan Sundaram, Aizana Turmakhametova (MIT), Nadathur Satish, Todd Mostak (Harvard), Piotr Indyk (MIT), Samuel Madden (MIT), and Pradeep Dubey; To appear at VLDB 2014.

• System used: Intel Endeavour Cluster
  • KNC: MPSS Version : 2.1.6720-15
  • Intel® Xeon Phi™ KNC card SKU: C0-7120P/7120X/7120
    Total No of Active Cores : 61, Frequency: 1100000 kHz, Card Memory: 16GB GDDR5
  • Host: Intel® Xeon®: E5-2680 0 @ 2.70GHz
  • Compiler: icpc version 13.1.1 (gcc version 4.4.6 compatibility)
  • MPI: Intel(R) MPI Library for Linux, 64-bit applications, Version 4.1 Build 20130116
Additional Information Regarding Reported GenBase Performance Data

• For benchmark details, refer to:
  http://istc-bigdata.org/index.php/genbase-a-benchmark-for-the-genomics-era/

System used: Single-node

- MPSS Version : 2.1.6720-19
- Intel® Xeon Phi™ KNC Card: Board SKU : B1PRQ-5110P/5120D
- Frequency : 1052631 kHz
- Host: Intel® Xeon®: CPU E5-2620 0 @ 2.00GHz
- Compiler: icpc (ICC) 14.0.0 20130728
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.

Copyright ©, Intel Corporation. All rights reserved. Intel, the Intel logo, Xeon, Xeon Phi, Core, VTune, and Cilk are trademarks of Intel Corporation in the U.S. and other countries.

**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