Data Intensive Computing

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

Nov 8h 2013

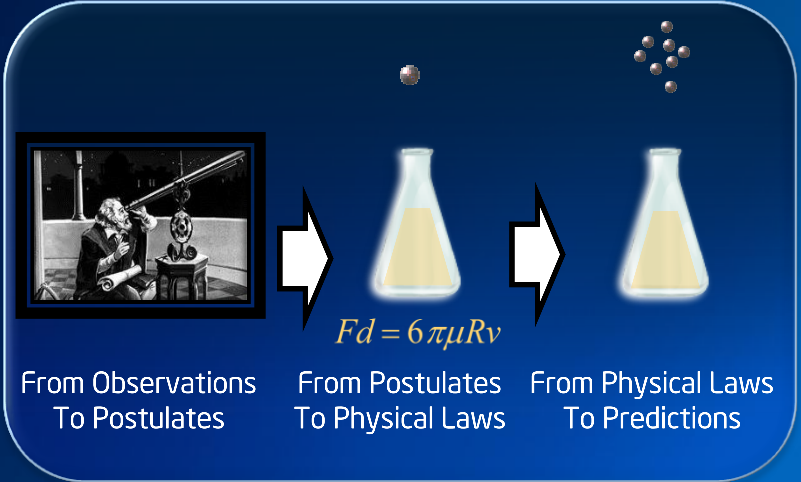
Data-data everywhere, not a bit of sense!

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
- Some early results on Xeon/Xeon Phi based systems

Enabling A Data Driven Science

Traditional HPC



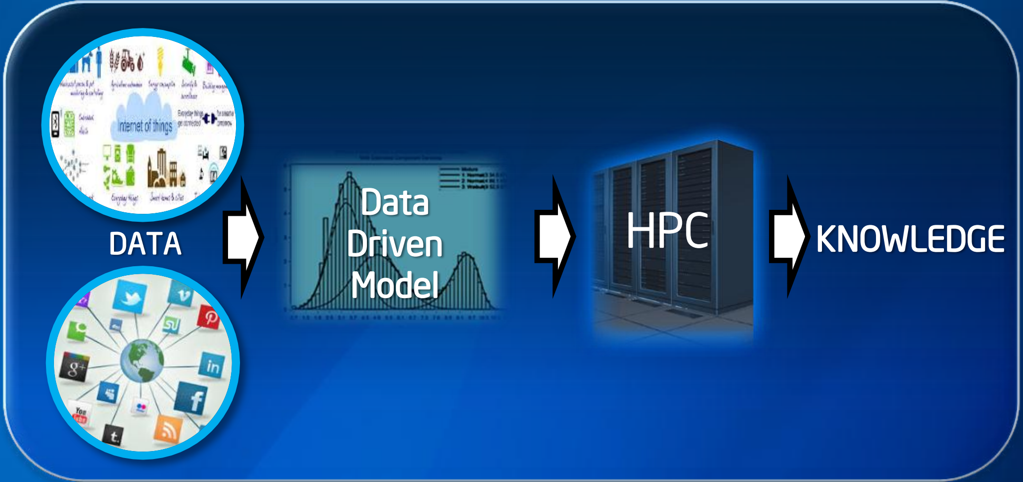
$Fd = 6\pi\mu Rv$

From Observations
To Postulates

From Postulates
To Physical Laws

From Physical Laws
To Predictions

Big Data HPC



DATA

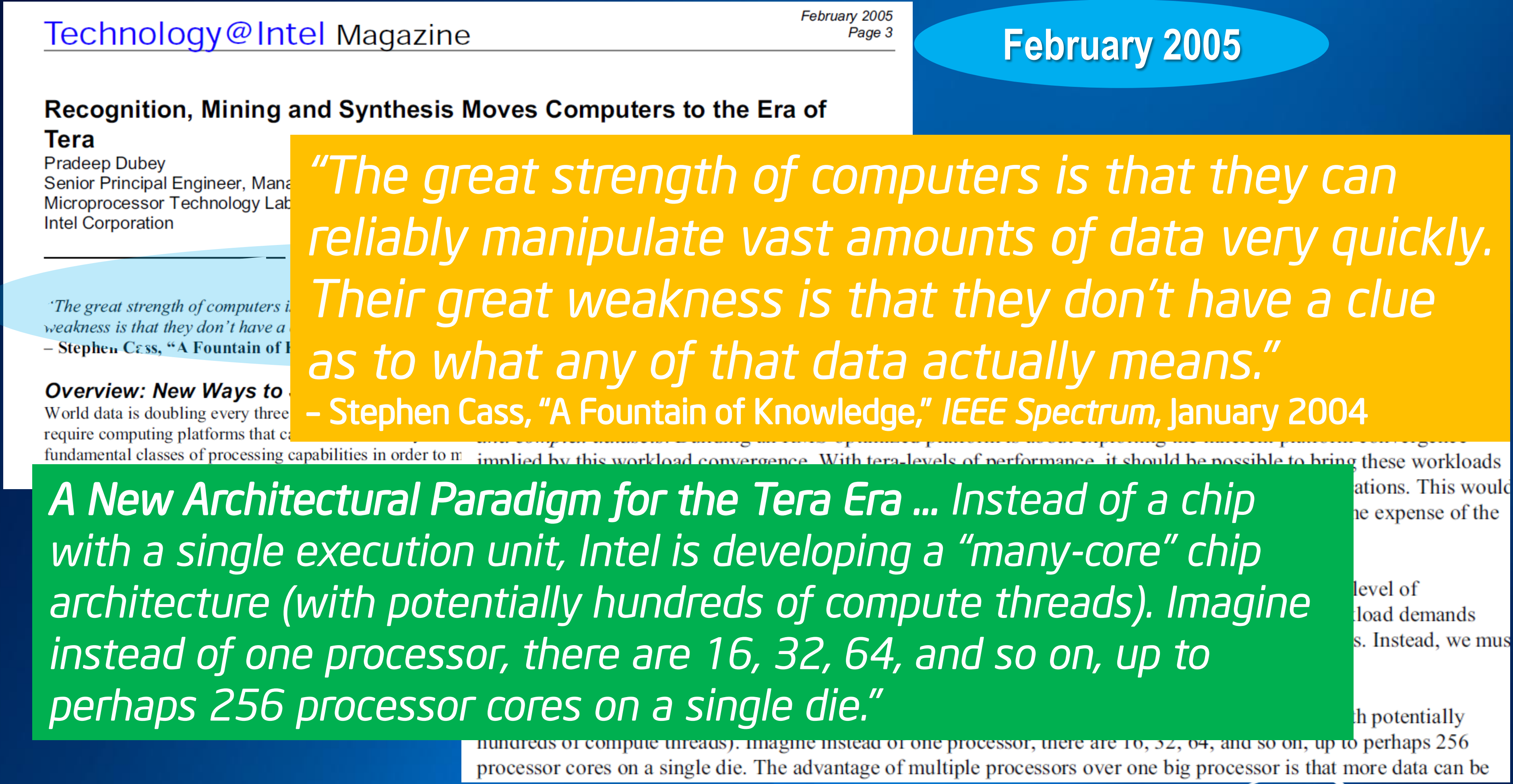
Data
Driven
Model

HPC

KNOWLEDGE

TRANSFORMING *DATA* INTO *KNOWLEDGE*

February 2005



Technology@Intel Magazine
February 2005
Page 3

Recognition, Mining and Synthesis Moves Computers to the Era of Tera

Pradeep Dubey
Senior Principal Engineer, Mana
Microprocessor Technology Lab
Intel Corporation

'The great strength of computers i
weakness is that they don't have a
– Stephen Cass, "A Fountain of K

Overview: New Ways to

World data is doubling every three
require computing platforms that ca
fundamental classes of processing capabilities in order to m

implied by this workload convergence. With tera-levels of performance, it should be possible to bring these workloads
ations. This would
ne expense of the

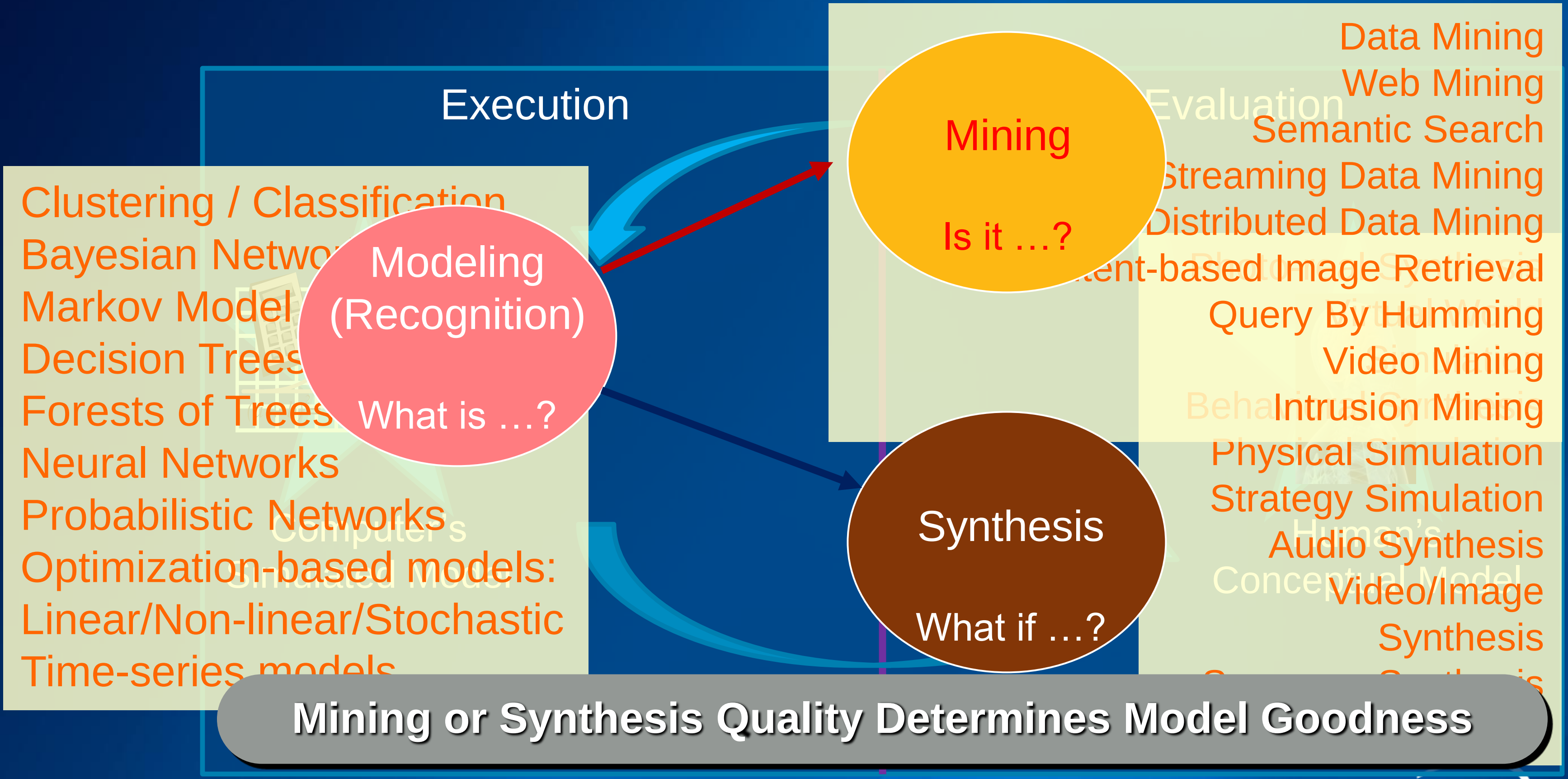
level of
load demands
s. Instead, we mus

h 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. The advantage of multiple processors over one big processor is that more data can be

> *"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."*
> – Stephen Cass, "A Fountain of Knowledge," *IEEE Spectrum*, January 2004

> ***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

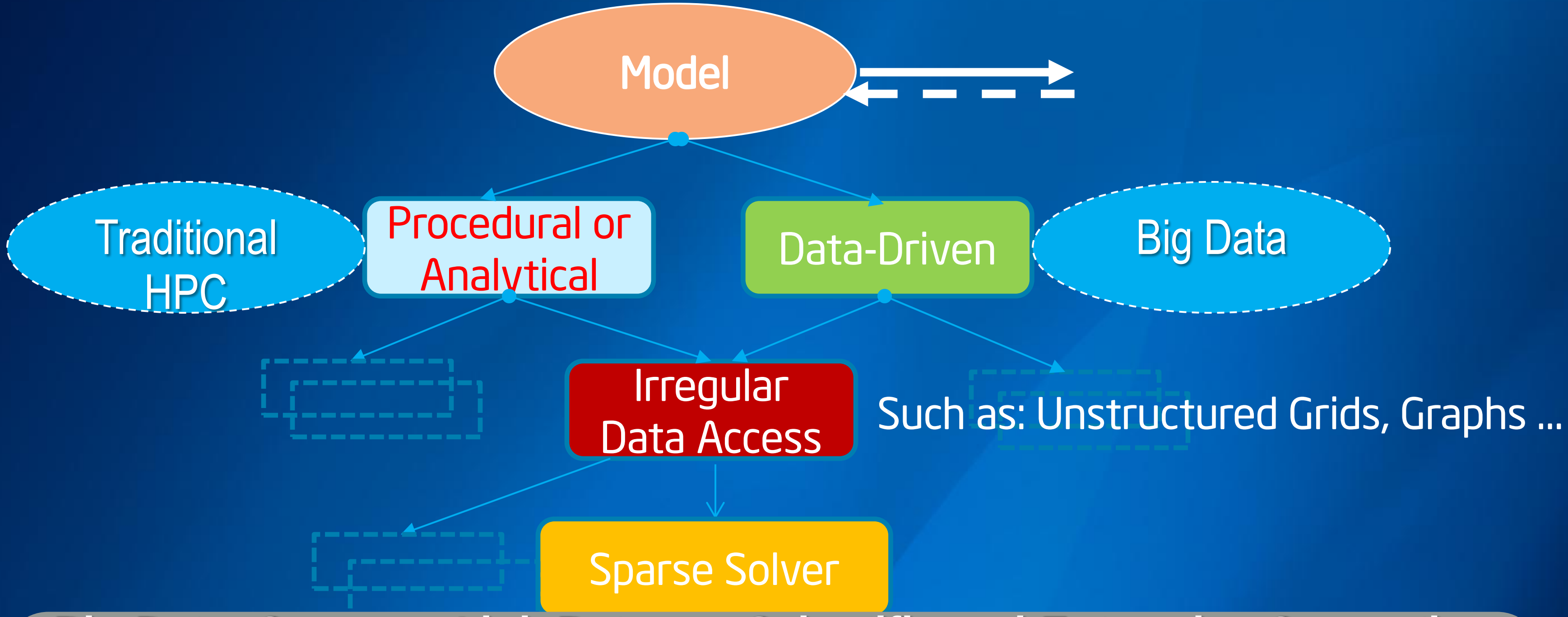


Clustering / Classification
Bayesian Network
Markov Model
Decision Trees
Forests of Trees
Neural Networks
Probabilistic Networks
Optimization-based models:
Linear/Non-linear/Stochastic
Time-series models

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

Mining or Synthesis Quality Determines Model Goodness

Models: A deeper look



Model

Traditional HPC

Procedural or Analytical

Data-Driven

Big Data

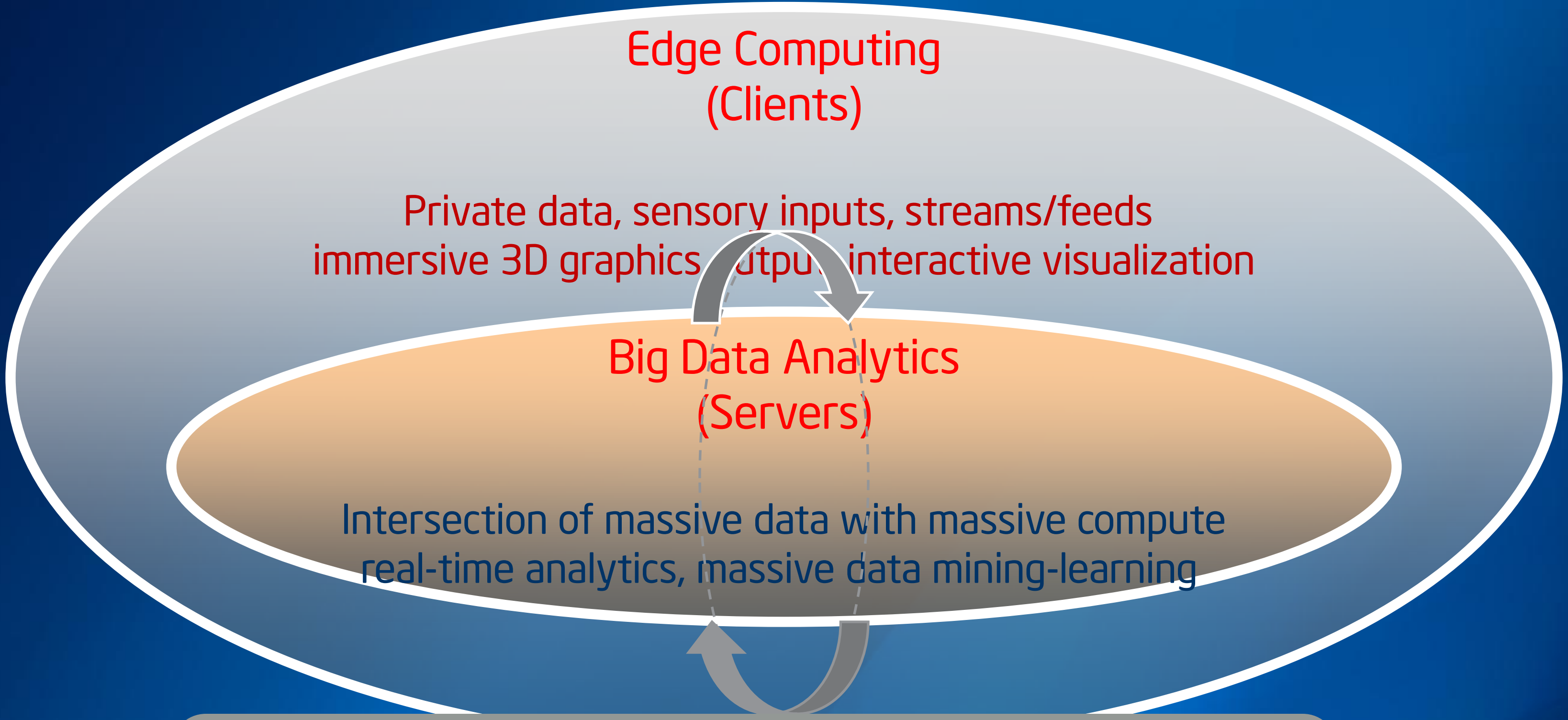
Irregular Data Access

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

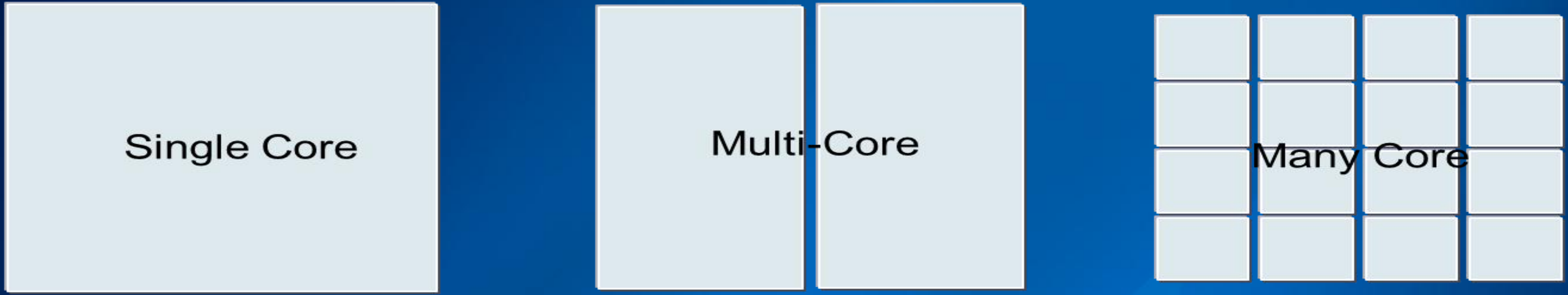
- Data-driven models are now tractable and usable
  - We are not limited to analytical models any more
  - No need to rely on *heuristics* alone for unknown models
  - Massive data offers new algorithmic opportunities
    - Many traditional compute problems worth revisiting
- Web connectivity significantly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

Multicore Versus Manycore



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

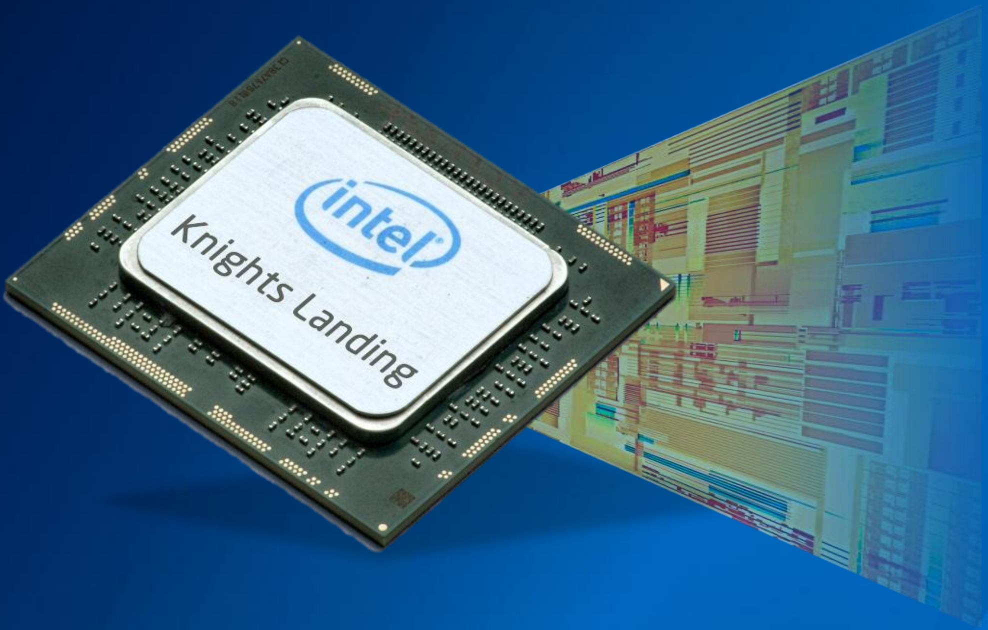
$$S = \frac{1}{(1-P)K_N + \frac{P}{N}}$$

Many Core makes sense for workloads with high enough "P "- parallel component - for simplicity, we call these Highly Parallel

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

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

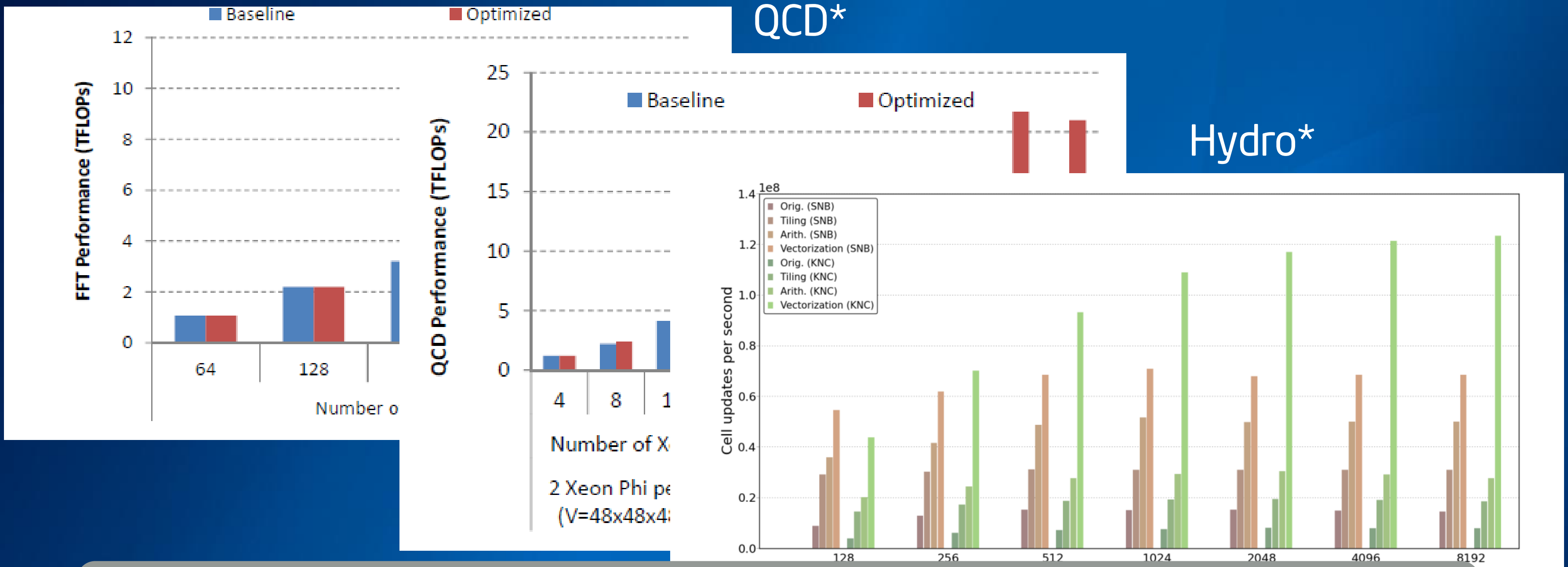
Some Recent Results

Encouraging New Results

FFT*

QCD*

Hydro*



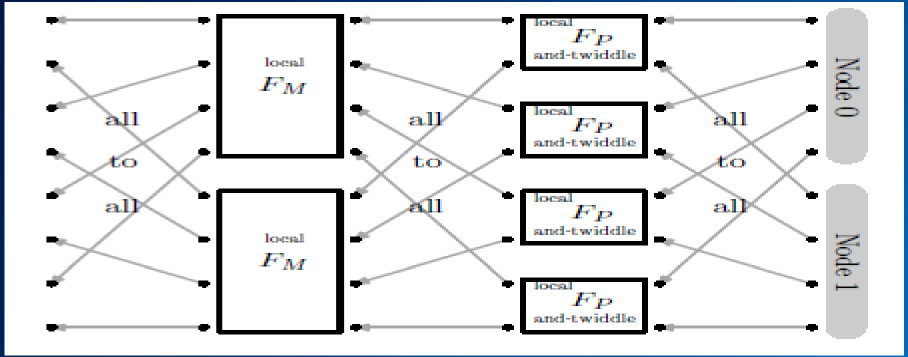
Intel® Xeon® and Xeon Phi™: both benefit from a common-set of manycore-friendly optimizations

* See Backup slides 28-30 for benchmark and system details

Optimization Notice

Look Inside.™

1D FFT: Segment-Of-Interest Algorithm*

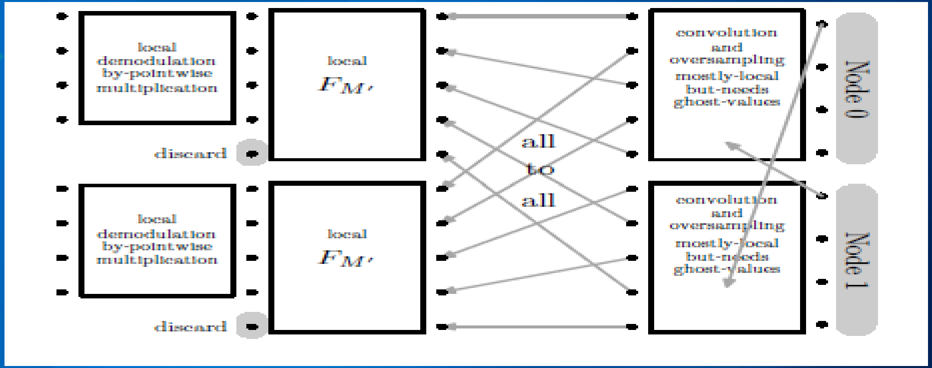


Motivation:

For big FFT problems running on large clusters, 50%-90% of time can be spent waiting for node-node data transfers.

Innovation:

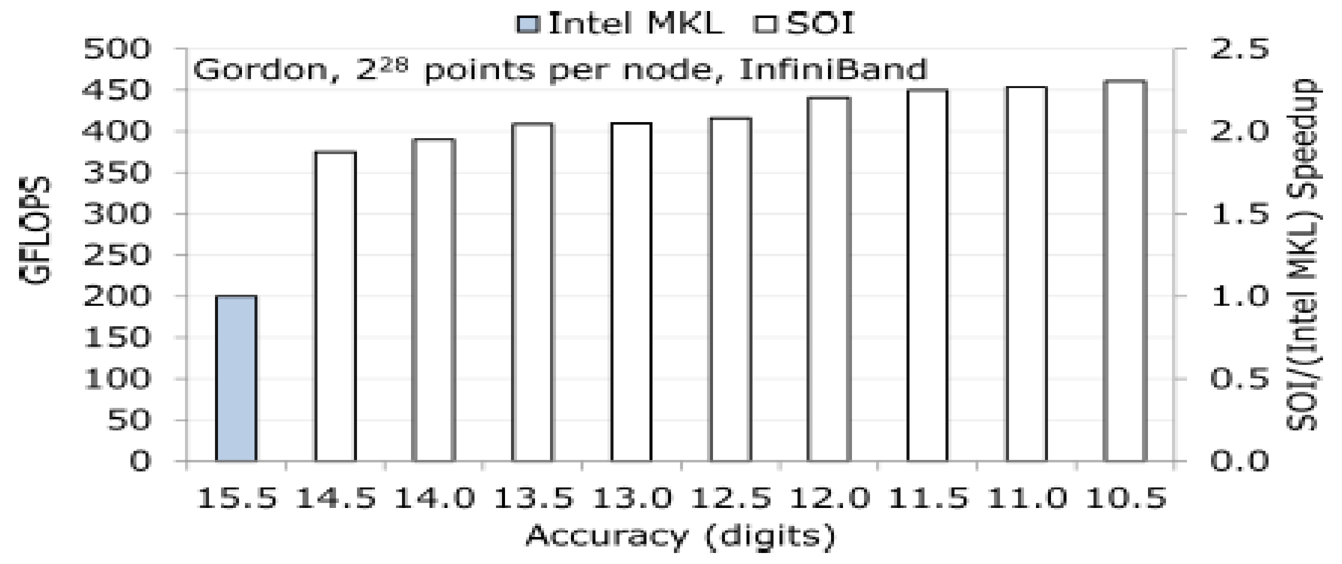
- Framework requiring single all-to-all communication
- For most factorizations: all factors well-approximated via sparse matrices



---

* A Framework for Low-Communication 1-D FFT", Ping Tak Peter Tang, Jongsoo Park, Daehyun Kim, Vladimir Petrov ; Supercomputing '12 (**Best Paper Award**)

Performance Vs. Accuracy

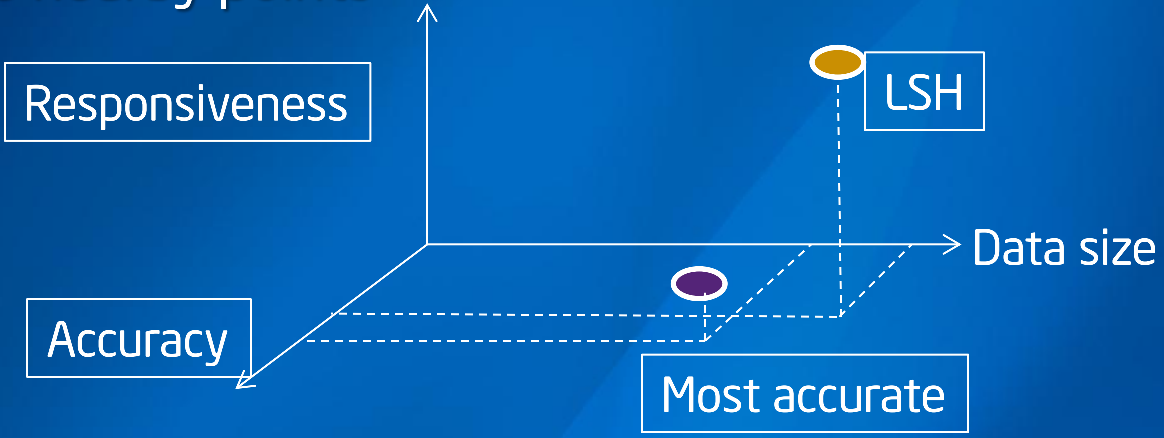


* See Backup slide 28 for benchmark and system details

Optimization Notice

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



Responsiveness

LSH

Data size

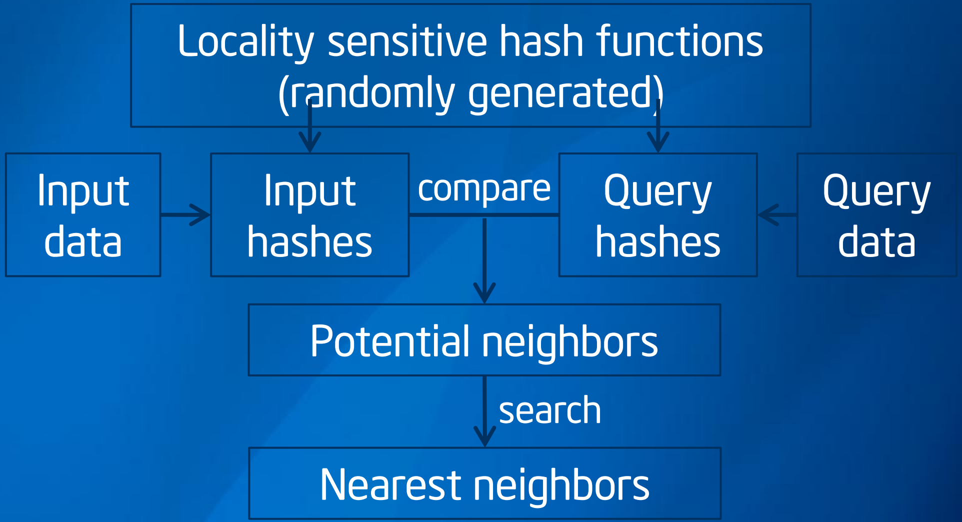
Accuracy

Most accurate

In Collaboration with ISTC-Big Data (MIT)

Applying LSH: Searching Twitter

- 400 million+ tweets per day
- Response time should be in 10s 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 $O(N^{1.5})$



Locality sensitive hash functions (randomly generated)

Input data

Input hashes

compare

Query hashes

Query data

Potential neighbors

search

Nearest neighbors

[*] Gionis, A.; Indyk, P., Motwani, R. (1999). , "Similarity Search in High Dimensions via Hashing". Proceedings of the 25th Very Large Database (VLDB) Conference.

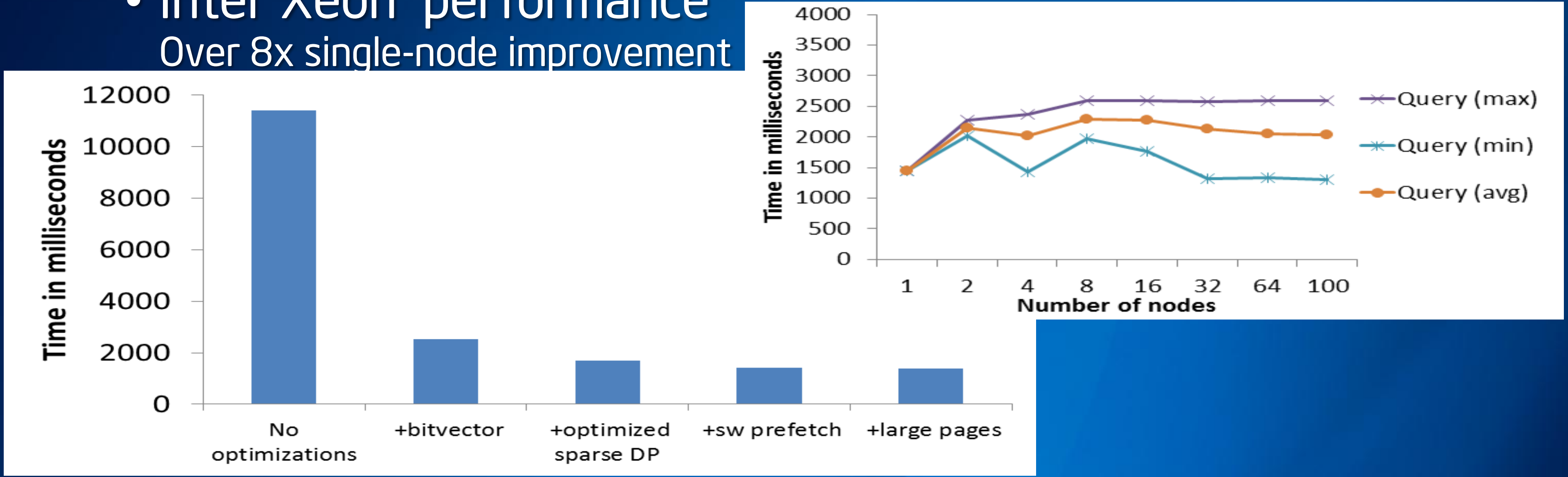
LSH → PLSH: When naïve parallelism is not enough

- 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

Optimization Notice

Look Inside.™

GenBase: A Complex Analytics Benchmark

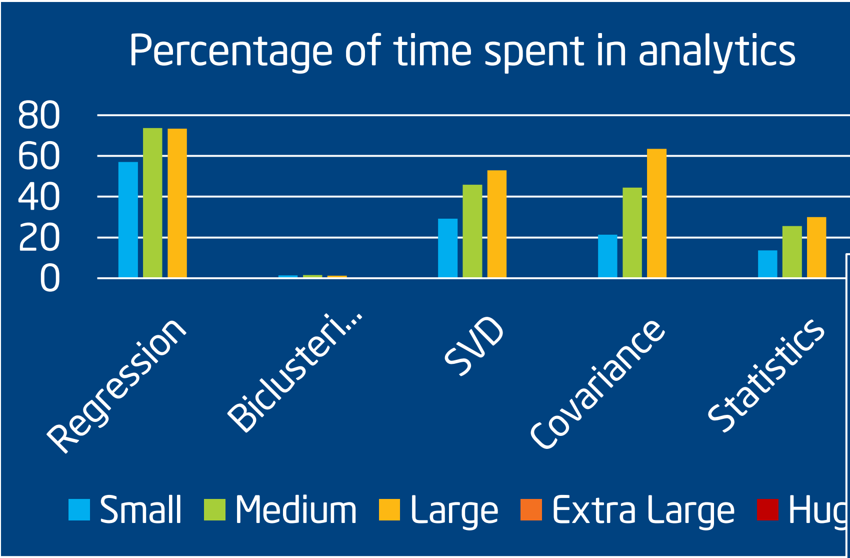
- Genomics:
  - Consider: 10^{4-5} gene expressions per sample, up to 10^{8-10} samples (multiple samples/patient)
  - Implies a problem that scales to10^{2-5} nodes of a large cluster with each node handling10^{4-5} samples

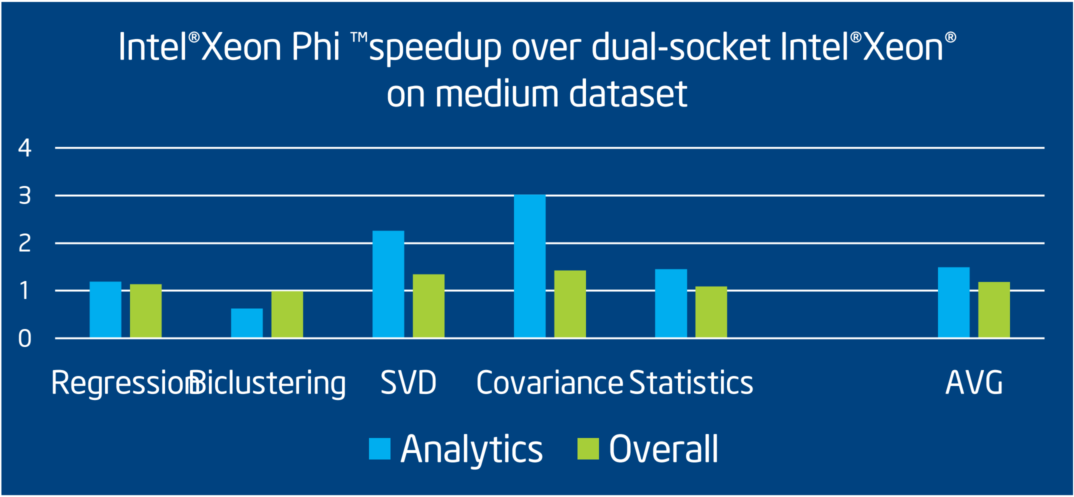
- GenBase:
  - Combination of DB ops + analytics
  - Analytics Queries: Regression, Covariance, Biclustering, SVD, Statistics

- Systems:
  - Systems: R, Postgres + R, Postgres + Madlib, Hadoop, SciDB, SciDB+ Xeon Phi, popular columnar DBMS + UDFs, columnar DBMS + R

In Collaboration with ISTC-Big Data (MIT)

GenBase: Preliminary Results


Percentage of time spent in analytics


Intel®Xeon Phi ™speedup over dual-socket Intel®Xeon® on medium dataset

* See Backup slide 32 for benchmark and system details

Optimization Notice

Apps ←→ HW

Twitter Search, Genomics, SAR, Billion-point correlation …

APPLICATIONS

Data Management

Complex Analytics

ALGORITHMS

- Database Primitives (Join, Select, Sort, …)
- Graph algorithms (BFS, SSSP, BC, …)
- Machine Learning (Kmeans, LSH, DBN, SGD, DBSCAN …)
- Numeric Computing (Regression, Covariance …)

FRAMEWORKS

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

LIBRARIES

Intel MKL

HARDWARE

Large Clusters

Multicore Manycore

Coprocessors

Announcing Intel® Parallel Computing Centers

Co-investing and collaborating to deliver modern parallel applications that are

☑ Open
☑ Standard
☑ Portable
☑ Scalable
☑ Greatest long-term return on investment

Join us to accelerating the next decade of discovery

Open call for Proposals

Submit your collaboration proposals by December 1st at: http://software.intel.com/academic

Summary

- Applications of tomorrow will increasingly be dominated by complex data-driven models
- Analytics will be increasingly complex, real-time, and continuous
- Imminent convergence of HPC with Big Data computing
- Density of manycore platforms, coupled with Locality-Communication-Approximation (LCA)-aware algorithms, will be key to efficiency

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, Intel Xeon, Intel Xeon Phi, Intel Hadoop Distribution, Intel Cluster Ready, Intel OpenMP, Intel Cilk Plus, Intel Threaded Buildiing blocks, Intel Cluster Studio, Intel Parallel Studio, Intel Coarray Fortran, Intel Math Kernal Library, Intel Enterprise Edition for Lustre Software, Intel Composer, the Intel Xeon Phi logo, the Intel Xeon logo and the Intel logo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

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 R 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 Hydro Performance Data

- For benchmark details, refer to:

  http://www.prace-project.eu/IMG/pdf/porting_and_optimizing_hydro_to_new_platforms.pdf
- System used: Single-node
  - MPSS Version : 2.1.3653-8
  - Intel® Xeon Phi™ KNC Card: Stepping: B0, Board SKU : ES2-P1750
  - Frequency : 1090909 kHz
  - Host: Intel® Xeon®: CPU E5-2680 0 @ 2.70GHz
  - Compiler: icpc (ICC) 13.1.3 20130607

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 Turmukhametova (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

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.

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