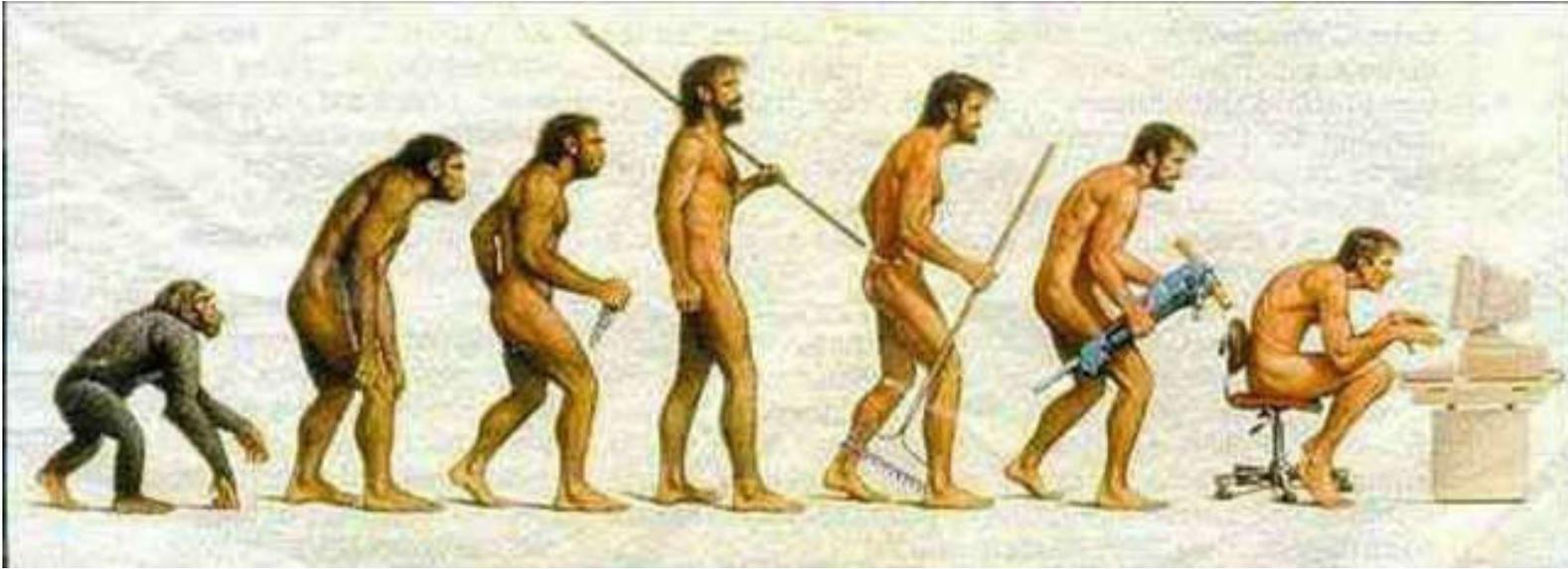




NEW COMPUTE FRONTIER = HPC + ML

Victor Lee
Data Center Group
Intel Corporate

Human and Tools



Tools help differentiate human from other animals

Computers are among the most amazing tools human created

What Will the Future of Computing Looks like

Better Hardware



Bigger Data



Transistor density doubles 18m
Computation / kwh doubles 18m
Cost / Gigabyte in 1995: \$1000.00
Cost / Gigabyte in 2015: \$0.03

Numbers: 5 KB / record
Text: 500 KB / record
Image: 1000 KB / picture
Audio: 5000 KB / song
Video: 5,000,000 KB / movie
High-Res: 50,000,000 KB / object

HPC is Transforming, accelerating Deeper Insight



**Fundamental
Discovery**

Advancing Science



**Business
Innovation**

High ROI: \$515
Return Per \$1 of HPC Invest¹



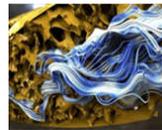
**New Analytical
Insight**

Data-Driven Analytics

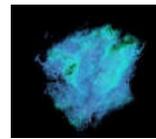
Diverse & New Workloads Driving Science & Industry



**Modeling &
Simulation**



Visualization



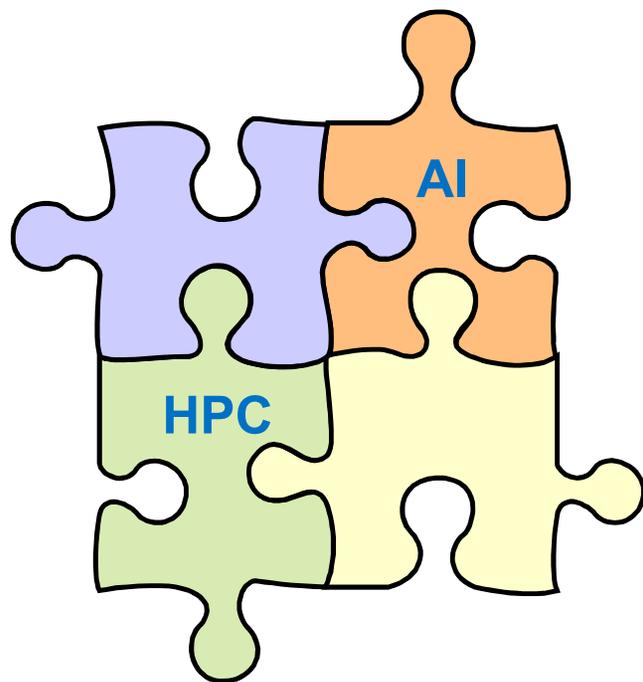
**High Performance
Data Analytics**



**Machine
Learning**

¹Source: IDC HPC and ROI Study Update (September 2015)

HPC vs. AI (Similarities)



Both are key building blocks of modern applications

Compute

Memory



Interconnect

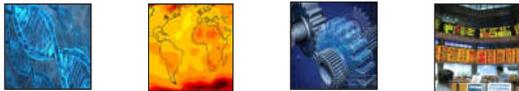
Both demands high performance compute, good memory subsystem and fast interconnect

HPC vs. AI (Differences)

HPC

AI

APPLICATIONS



PRODUCTIVITY FRAMEWORK

Intel® Nervana™ DL Studio



mxnet

Caffe



PERFORMANCE LIBRARY

Intel® Math Kernel Library (MKL)

Intel® Performance Primitives (IPP)

Intel® Python Distribution

Intel® Data Analytics Acceleration Library (DAAL)

Intel® Nervana™ Graph*
Intel® Math Kernel Library (MKL, MKL-DNN)

HARDWARE



Compute



Memory & Storage

+



Networking



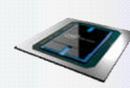
Canyon Vista



Intel® Movidius™ Myriad™ X



Intel® GNA*

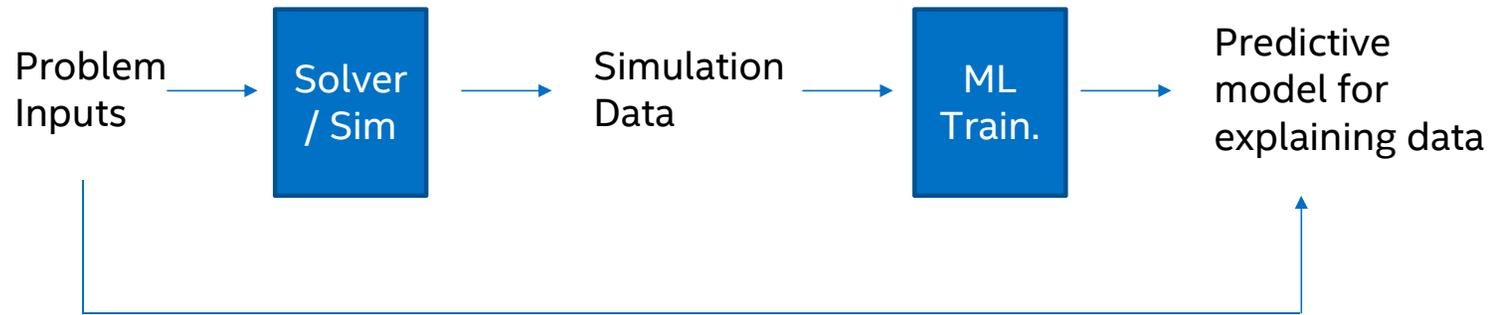


Intel® Nervana™ Neural Network Processor

*Future

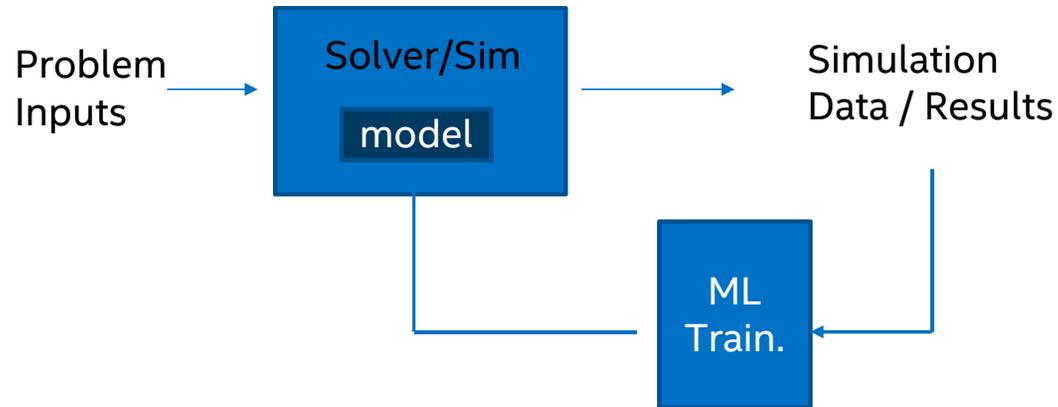
A Few Aspects

Ways in Combining HPC and ML (1)



- Machine Learning is applied in TANDEM to solver / simulation
- Could benefit from many years of simulations (e.g. Weather, molecular simulation, etc.)
- Data management (I/O) can become a major bottleneck

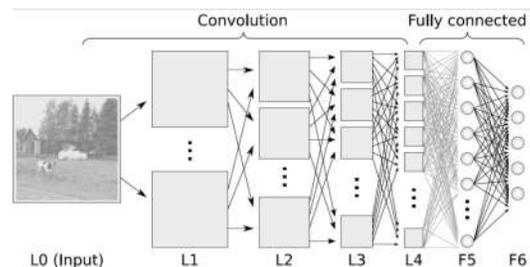
Ways in Combining HPC and ML (2)



- Rarely, Machine Learning module is embedded in the Solver / Simulation
- Potentially a new perspective to improve HPC (e.g. quantum material simulation)

Is DNN the best way to learn?

Convolutional Neural Network



Key operations: convolution, non-linear activation, batch normalization, etc.

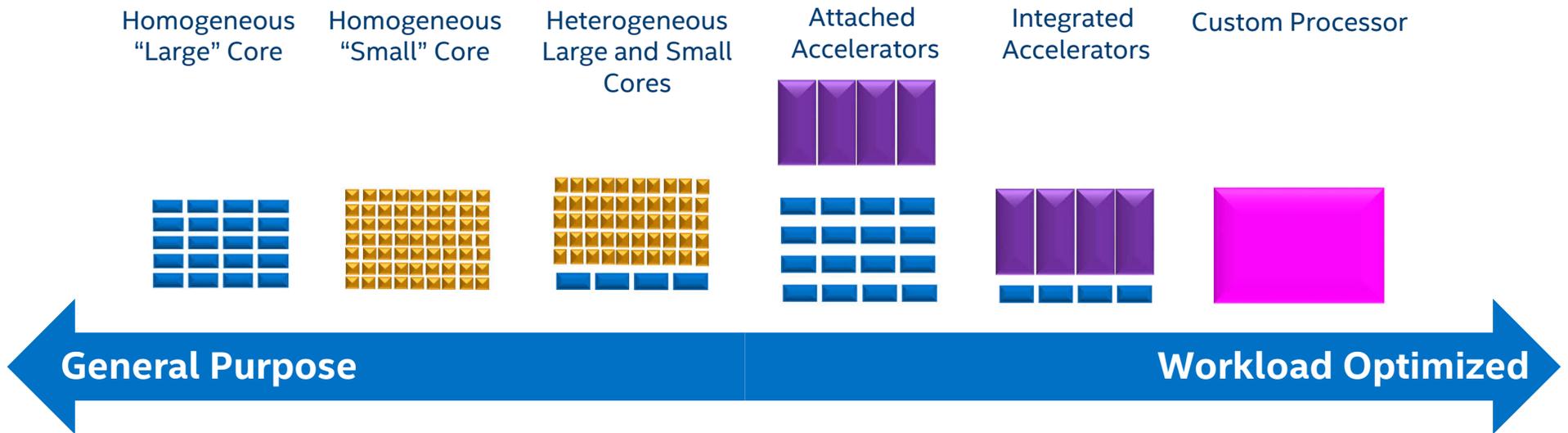
Computational requirements:

- Heavy demand on **regular** compute (100s PF – EF)
- Memory demand increases rapidly with data dimension (100s MB -> 100s GB)
- Heavy network communication for training (**Latency critical**)



Brute
Force

Diverse Architectural Approaches



What to use for HPC + ML?

Summary / Call to Action

Entering exciting compute era with abundant of compute and data

HPC+ML will be prevalent in scientific and industrial usages

Collaborate to innovate and discover



NOTICES AND DISCLAIMERS

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration.

No computer system can be absolutely secure.

Tests document performance of components on a particular test, in specific systems. Differences in hardware, software, or configuration will affect actual performance. For more complete information about performance and benchmark results, visit <http://www.intel.com/benchmarks>.

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 complete information visit <http://www.intel.com/benchmarks>.

Intel® Advanced Vector Extensions (Intel® AVX)* provides higher throughput to certain processor operations. Due to varying processor power characteristics, utilizing AVX instructions may cause a) some parts to operate at less than the rated frequency and b) some parts with Intel® Turbo Boost Technology 2.0 to not achieve any or maximum turbo frequencies. Performance varies depending on hardware, software, and system configuration and you can learn more at <http://www.intel.com/go/turbo>.

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.

Cost reduction scenarios described are intended as examples of how a given Intel-based product, in the specified circumstances and configurations, may affect future costs and provide cost savings. Circumstances will vary. Intel does not guarantee any costs or cost reduction.

Intel does not control or audit third-party benchmark data or the web sites referenced in this document. You should visit the referenced web site and confirm whether referenced data are accurate. Intel, the Intel logo, and Intel Xeon are trademarks of Intel Corporation in the U.S. and/or other countries.

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

© 2018 Intel Corporation.



THANK YOU!