LEGAL NOTICES & DISCLAIMERS

This document contains information on products, services and/or processes in development. All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest forecast, schedule, specifications and roadmaps.

Intel technologies’ features and benefits depend on system configuration and may require enabled hardware, software or service activation. Learn more at intel.com, or from the OEM or retailer. 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. Consult other sources of information to evaluate performance as you consider your purchase. For more complete information about performance and benchmark results, visit http://www.intel.com/performance.

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.

Statements in this document that refer to Intel's plans and expectations for the quarter, the year, and the future, are forward-looking statements that involve a number of risks and uncertainties. A detailed discussion of the factors that could affect Intel's results and plans is included in Intel's SEC filings, including the annual report on Form 10-K.

The products described 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.

No license (express or implied, by estoppel or otherwise) to any intellectual property rights is granted by this document.

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, Pentium, Celeron, Atom, Core, Xeon and others are trademarks of Intel Corporation in the U.S. and/or other countries. *Other names and brands may be claimed as the property of others.

© 2016 Intel Corporation.
THE COMING FLOOD OF DATA
By 2020…

The average internet user will generate
~1.5 GB OF TRAFFIC PER DAY

Smart hospitals will be generating over
3,000 GB PER DAY

Self driving cars will be generating over
4,000 GB PER DAY… EACH

A connected plane will be generating over
40,000 GB PER DAY

A connected factory will be generating over
1,000,000 GB PER DAY

RADAR ~10-100 KB PER SECOND
SONAR ~10-100 KB PER SECOND
GPS ~50 KB PER SECOND
LIDAR ~10-70 MB PER SECOND
CAMERAS ~20-40 MB PER SECOND
1 CAR 5 EXAFLOPS PER HOUR

All numbers are approximated

Design the Difference with Digital twin
ATC 2017
THE NEXT BIG WAVE OF COMPUTING

AI COMPUTE CYCLES WILL GROW 12X BY 2020

- MAINFRAMES
- STANDARDS-BASED SERVERS
- CLOUD COMPUTING

- DATA DELUGE
- COMPUTE BREAKTHROUGH
- INNOVATION SURGE

ARTIFICIAL INTELLIGENCE

Source: Intel
AI WILL Usher in a Better World
On the Scale of the Agricultural, Industrial and Digital Revolutions

ACCELERATE
Large scale solutions
- Cure Diseases
- Prevent Crime
- Unlock Dark Data

UNLEASH
Scientific Discovery
- Explore New Worlds
- Decode the Brain
- Uncover New Theories

EXTEND
Human Capabilities
- Personalize Learning
- Enhance Decisions
- Optimize Time

AUTOMATE
Undesirable Tasks
- Automate Driving
- Save Lives in Danger
- Perform Chores

Source: Intel
AI IS TRANSFORMING INDUSTRIES

**CONSUMER**
- Smart Assistants
- Chatbots
- Search
- Personalization
- Augmented Reality
- Robots

**HEALTH**
- Enhanced Diagnostics
- Drug Discovery
- Patient Care
- Research
- Sensory Aids

**FINANCE**
- Algorithmic Trading
- Fraud Detection
- Research
- Personal Finance
- Risk Mitigation

**RETAIL**
- Support
- Experience
- Marketing
- Merchandising
- Loyalty
- Supply Chain
- Security

**GOVERNMENT**
- Defense
- Data Insights
- Safety & Security
- Resident Engagement
- Smarter Cities

**ENERGY**
- Oil & Gas Exploration
- Smart Grid
- Operational Improvement
- Conservation

**TRANSPORT**
- Automated Cars
- Automated Trucking
- Aerospace
- Shipping
- Search & Rescue

**INDUSTRIAL**
- Efficiency Improvement
- Factory Automation
- Predictive Maintenance
- Precision Agriculture
- Field Automation

**OTHER**
- Advertising
- Education
- Gaming
- Professional & IT Services
- Telco/Media
- Sports

**EARLY ADOPTION**

Source: Intel forecast
A COMMON LANGUAGE FOR AI TODAY

ARTIFICIAL INTELLIGENCE

SENSE  REASON  ACT  ADAPT

REMEMBER

MACHINE LEARNING  REASONING SYSTEMS

DEEP LEARNING  CLASSIC ML  MEMORY BASED  LOGIC BASED
**WHAT IS MACHINE LEARNING?**

### CLASSIC ML

Using optimized functions or algorithms to extract insights from data

**Algorithms**
- Random Forest
- Support Vector Machines
- Regression
- Naïve Bayes
- Hidden Markov
- K-Means Clustering
- More...

### DEEP LEARNING

Using massive labeled data sets to train deep (neural) graphs that can make inferences about new data

**Step 1: Training**

Use massive labeled dataset (e.g. 10M tagged images) to iteratively adjust weighting of neural network connections

**Step 2: Inference**

Hours to Days in Cloud

Real-Time at Edge/Cloud

Form inference about new input data (e.g. a photo) using trained neural network

*Note: not all classic machine learning functions require training*
ARTIFICIAL INTELLIGENCE @ INTEL

Unleash Your Potential with Intel’s Complete AI Portfolio
END-TO-END USE CASE
Artificial Intelligence For Automated Driving

VEHICLE
- Driving Functions
- Environment Modeling
- Sensor Fusion

Anomaly Detection

NETWORK
- Captured Sensor Data
- Real Time Model, SW, FW Updates
- Data Formatting, Storage, Management, Traceability

CLOUD
- Model Training
- Model Inference
- Compress Model
- Universal Models
- Reasoning Systems

5G
DL LB

ML DL RB LB

Design the Difference with Digital twin

ATC 2017
COMPUTER VISION PORTFOLIO

Vital AI Capability… Involves Over 60% of Human Brain

- 3-in-1: 1080p HD camera, infrared camera and laser projector
- See depth and track motion like the human eye

- Myriad 2 vision processing units (VPUs)
- High-performance for complex neural nets
- Low power (less than ~1W)

INTEL COMPUTE

- Intel® Xeon Phi™ processors
- Intel® Xeon® processors
- Intel® Core™ processors
- Intel® Atom™ processors
- Intel® Joule™ processors
- Intel® Quark™ processors
INTEL’S INVESTMENTS

CUSTOMER COLLABORATIONS

RESEARCH SYSTEMS/POCs

UX DESIGN & DEVELOPMENT

* Other names and brands may be claimed as the property of others
AUTOMATED DRIVING SYSTEM

MODEL INFERENCE
TRAJECTORY ENUMERATION
PATH PLANNING
ENVIRONMENT MONITORING
SENSOR PROCESSING AND FUSION

REAL TIME HD MAP UPDATES
CAR-TO-CAR COMMUNICATION
ANOMALY NOTIFICATION
CROWD SOURCED DATA DISTRIBUTION
PUBLIC AND PRIVATE DATA SECURITY

ENDPOINT MANAGEMENT
BIG DATA AND STATISTICAL ANALYTICS
MODEL TRAINING
DATA FORMATTING
ANNOTATION
DATASET MANAGEMENT
DATA AGGREGATION
INTEL® NERVANA™ PORTFOLIO
Common Architecture for Machine & Deep Learning

INTEL® XEON® PROCESSORS
Most Widely Deployed Machine Learning Platform (>97%*)

INTEL® XEON PHI™ PROCESSORS
Higher Performance Machine Learning, General Purpose

INTEL® XEON® PROCESSOR + FPGA
Breakthrough Deep Learning Inference & Workload Flexibility

INTEL® XEON® PROCESSOR + LAKE CREST
Best-in-Class Neural Network Training Performance

TARGETED ACCELERATION

*Intel® Xeon® processors are used in 97% of servers that are running machine learning workloads today (Source: Intel)
INTEL® NERVANA™ PORFTOLIO (DETAIL)

TRAINING

BATCH
Train machine learning models across a diverse set of dense and sparse data

MANY BATCH MODELS
Train large deep neural networks
Train large models as fast as possible

INFEERENCE

BATCH
Infer billions of data samples at a time and feed applications within ~1 day

STREAM
Infer deep data streams with low latency in order to take action within milliseconds

EDGE
Power-constrained environments

Intel® Xeon™
OR
Intel® Xeon Phi™

LAKE CREST
*Future*

Option for higher throughput/watt

Required for low latency

Design the Difference with Digital twin
### Intel® Xeon® Processor Family

**Most Widely Deployed Machine Learning Platform (97% share)**

- **Lowest TCO With Superior Infrastructure Flexibility**
  - Standard server infrastructure
  - Open standards, libraries & frameworks
  - Optimized to run wide variety of data center workloads

- **Server Class Reliability**
  - Industry standard server features: high reliability, hardware enhanced security

- **Leadership Throughput**
  - Industry leading inference performance
  - Up to **18X** performance on existing hardware via Intel software optimizations

---

**Processor optimized for a wide variety of datacenter workloads, flexible infrastructure with low TCO**

---

Configuration details on slide: 30

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/performance. Source: Intel measured as of November 2016

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.
Removing IO and Memory Barriers

- Integrated Intel® Omni-Path fabric increases price-performance and reduces communication latency
- Direct access of up to 400 GB of memory with no PCIe performance lag (vs. GPU:16GB)

Breakthrough Highly Parallel Performance

- Up to 400X deep learning performance on existing hardware via Intel software optimizations
- Up to 4X deep learning performance increase estimated (Knights Mill, 2017)

Easier Programmability

- Binary-compatible with Intel® Xeon® processors
- Open standards, libraries and frameworks
INTEL® ARRIA® 10 FPGA
Breakthrough Deep Learning Inference & Workload Flexibility

Reconfigurable accelerator for enhanced inference needs and flexible workload acceleration

**Exceptional Power Efficiency**
- Up to 80% reduction in power consumption (vs. Intel® Xeon® processor), with inference up to 25 images/sec/watt on Caffe running Alexnet

**High-Throughput, Low Latency**
- Deterministic, real-time, inline processing of streaming data without buffering

**Infrastructure Flexibility**
- Future proof for new neural network topologies, arbitrary precision data types (FloatP32 => FixedP2, sparsity, weight sharing), inline & offload processing
- Reconfigurable accelerator can be used for variety of data center workloads with fast switching

---

Note: available as discrete or Xeon with integrated FPGA (Broadwell Proof of Concept)

Configuration details on slide 51

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/performance. Source: Intel measured as of November 2016

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
LAKE CREST

Best-in-Class Deep Learning Training Performance

Accelerator for unprecedented training compute density in deep learning centric environments

<table>
<thead>
<tr>
<th>Hardware for DL Workloads</th>
<th>Blazingly Fast Data Access</th>
<th>High Speed Scalability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Custom-designed for deep learning</td>
<td>32 GB of in package memory via HBM2 technology</td>
<td>12 bi-directional high-bandwidth links</td>
</tr>
<tr>
<td>Unprecedented compute density</td>
<td>8 Tera-bits/s of memory access speed</td>
<td>Seamless data transfer via interconnects</td>
</tr>
<tr>
<td>More raw computing power than today’s state-of-the-art GPUs</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Everything needed for deep learning and nothing more!
INTEL® OMNI-PATH ARCHITECTURE

World-Class Interconnect Solution for Shorter Time to Train

Fabric interconnect for breakthrough performance on scale-out apps like deep learning training

Building on some of Industry’s best technologies
- Highly leverage existing Aries & Intel True Scale fabrics
- Excellent price/performance \(\leftrightarrow\) price/port, 48 radix
- Re-use of existing OpenFabrics Alliance Software
- Over 80+ Fabric Builder Members

Breakthrough Performance
- Increases price performance, reduces communication latency compared to InfiniBand EDR\(^1\):
  - Up to 21% Higher Performance, lower latency at scale
  - Up to 17% higher messaging rate
  - Up to 9% higher application performance

Innovative Features
- Improve performance, reliability and QoS through:
  - Traffic Flow Optimization to maximize QoS in mixed traffic
  - Packet Integrity Protection for rapid and transparent recovery of transmission errors
  - Dynamic lane scaling to maintain link continuity

HFI Adapters
- Single port
  - x8 and x16

Edge Switches
- 1U Form Factor
  - 24 and 48 port

Director Switches
- QSFP-based
  - 192 and 768 port

Software
- Open Source
  - Host Software and Fabric Manager

Cables
- Third Party Vendors
  - Passive Copper
  - Active Optical

---

\(^{1}\)Intel® Xeon® Processor E5-2697A v4 dual-socket servers with 2133 MHz DDR4 memory, Intel® Turbo Boost Technology and Intel® Hyper Threading Technology enabled. BIOS: Early snoop disabled. Cluster on Die disabled, I/OU non-posted prefetch disabled, Snoop hold-off timer=9. Red Hat Enterprise Linux Server release 7.2 (Maipo). Intel® OPA testing performed with Intel Corporation Device 24f0 – Series 100 HFI ASIC (B0 silicon). OPA Switch: Series 100 Edge Switch – 48 port (B0 silicon). Intel® OPA host software 10.1 or newer using Open MPI 1.10.x contained within host software package. EDR IB\(^*\) testing performed with Mellanox EDR ConnectX-4 Single Port Rev 3 MCX455A HCA. Mellanox SB7700 - 36 Port EDR Infiniband switch. EDR tested with MLNX_OFED_LINUX-3.2.x, OpenMPI 1.10.x contained within MLNX HPC-X. Message rate claim: Ohio State Micro Benchmarks v. 5.0. osu_mbb_mr, 8 B message (uni-directional), 32 MPI rank pairs. Maximum rank pair communication time used instead of average time, average timing introduced into Ohio State Micro Benchmarks as of v3.9 (2/28/13). Best of default, MXM_TLS=slfrc, and -mca pml yalla tunings. All measurements include one switch hop. Latency claim: HPCC 1.4.3 Random order ring latency using 16 nodes, 32 MPI ranks per node, 512 total MPI ranks. Application claim: GROMACS version 5.0.4 ion_channel benchmark. 16 nodes, 32 MPI ranks per node, 512 total MPI ranks. Intel® MPI Library 2017.0.064. Additional configuration details available upon request.
Artificial intelligence (AI), the next big wave in computing, is an increasingly important source for competitive advantage that is already transforming industries.

Today is the ideal time to begin integrating AI into your products, services and business processes.

Intel has the complete AI portfolio, world-class silicon, and experience from successfully driving previous major computing transformations.

Get started with #IntelAI today!

- Use Intel’s performance-optimized libraries & frameworks
- Contact your Intel representative for help and POC opportunities
- Find out more at www.intel.com/ai & software.intel.com/ai
THANK YOU!