MIZUHO Mizuho Securities USA Inc.

Technology Research
Semiconductors
Industry Update

U.S. Equity Research January 18, 2017

Mi-Tech Vol.46: AI & Deep Learning; Primer to a Revolution

A Major Paradigm Shift in Computing for the Next Decade

Summary

Vol. 46 of the Mizuho Global Tech note is intended to be a Primer on the next big paradigm shift in computing for the next 20-50 years, a major shift from the last 20-50 years which was about computing and speed. We expect Deep Learning and AI to pervade all facets of the economy. We see NVDA as well-positioned for the next decade, and see opportunities for AMD and Intel with FPGA, new AI players and new, massively parallel computing platforms.

		Price	Rating		
Company	Symbol	(1/18)	Prior	Curr	PT
Advanced Micro Devices, Inc.	AMD	$9.88	–	Buy	$13.00
Intel Corporation	INTC	$36.76	–	Buy	$42.00
NVIDIA Corporation	NVDA	$102.95	–	Buy	$115.00
QUALCOMM Incorporated	QCOM	$65.13	–	Buy	$75.00

Source: Bloomberg and Mizuho Securities USA

Key Points

Welcome to Computing 2.0. We believe deep learning and AI with parallel processing is driving broad industry adoption as the enterprise segment looks to use available, real-time data to learn, predict, and prepare for contingencies better and faster. DL/ML/AI is being broadly adopted in healthcare, manufacturing, automotive, finance, insurance, banking, and retail.

How big can AI be? Global server market revenue is ~$50B according to Gartner and Intel has ~$12B of data center chip revenue annually. We believe as AI and deep learning start to permeate Enterprise, this could eventually grow to the size of the Server/Data Center market. Intel has noted only 1% of Servers today are dedicated to AI, and it is a market expected to grow 12x over the next four years. Also, to give some context, NVDA, one of the hardware leaders in DL, did all-in revenue of ~$800M for 2016, so we think this is going to be a big L-T secular market.

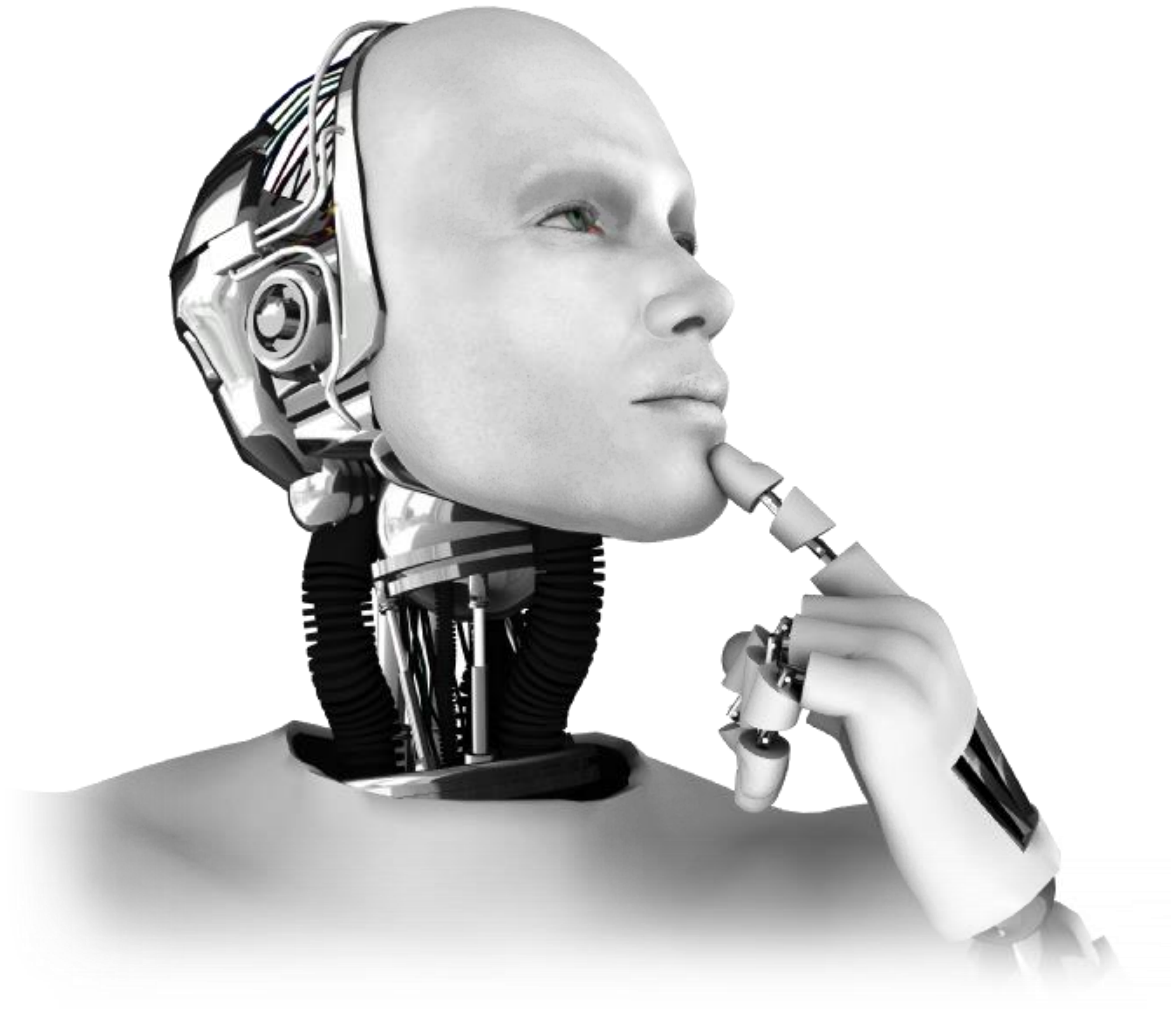
Will AI and Deep Learning be done on CPUs, GPUs, or FPGAs? We believe one of the key reasons for the move to multi-core with GPUs is because of the higher processing speeds that can be achieved. Traditional Von Neumann computing needs single threaded processors such as Skylake/Xeon, which are also hitting power/performance limits. AI frameworks need parallel processing more easily achieved by dedicating cores for training, driving the need to GPUs and NVDA, which deliver much higher speedup and frequencies. But there are also FPGAs for lower power and fixed logic around the corner, though we believe there are limitations.

Not to forget, we see Automotive ADAS as one of the biggest single industry DL implementations. We believe while machine learning will revolutionize traditional insurance processing, healthcare, hospitals, banking and retail, one of the largest revolutions will be a "learning car," as ADAS solutions such as Drive PX2 advance a learning car. **Much more on subsequent pages.**

Vijay Rakesh
Managing Director, Americas Research
+1 312 294 8682
Vijay.Rakesh@us.mizuho-sc.com

PLEASE REFER TO PAGE 30 OF THIS REPORT FOR IMPORTANT DISCLOSURE AND ANALYST CERTIFICATION INFORMATION. Mizuho Securities USA Inc. does and seeks to do business with companies covered in its research reports. As a result, investors should be aware that the Firm may have a conflict of interest that could affect the objectivity of this report. Investors should consider this report as only a single factor in making their investment decision.

Exhibit 1: The Road to the Future of Artificial Intelligence



Source: wccs14.org

Contents

Introduction to Artificial Intelligence, Machine Learning, and Deep Learning

While the idea of Artificial Intelligence has been around for quite some time, we are beginning to make another push into the world of machine learning. As we enter a new "spring" of AI after a number of winters in AI's 50+ years of existence, we take a look at some of the types of AI, solutions, end goals, and players in the updated and emerging game.

Some of the major players in Deep Learning and AI include not only the well-known Google, Facebook, Amazon, etc. but also Huawei as a key supplier of hardware architecture, and Baidu, which now has the founder of Deep Learning, Andrew Ing, leading its AI effort. Baidu also has its open source deep learning platform called Paddle (Parallel Distributed Deep Learning) AI which is broad in its implementation finding its way from the Amazon Alexa to Siri to Android Pixel phones and even Huawei Honor phones using AI and Deep Learning to adopt to its user's tastes and locations.

Deep Learning and AI are driving the next revolution as Softbank CEO Masayoshi San noted at the ARM Techcon 2016 that by 2020 there will be 1-trillion connected devices and noting that the collective intelligence of machines will exceed the collective intelligence of humans.

Artificial Intelligence

AI (Human Intelligence Exhibited by Machines) is a field of computer science that was created in the 1960s to solve tasks that are easy for humans but hard for computers. The term AI is used as a catch-all and includes machine learning (narrow AI) and Deep Learning (Strong AI) as its subsets. Specifically, Strong AI would be a system that could do anything a human can, such as planning, recognizing objects/sounds, speaking, translation, creative work, social or business transactions other than physical activity. "Narrow AI" technologies perform specific tasks better than humans can, for example image classification on Pinterest or face recognition on Facebook.

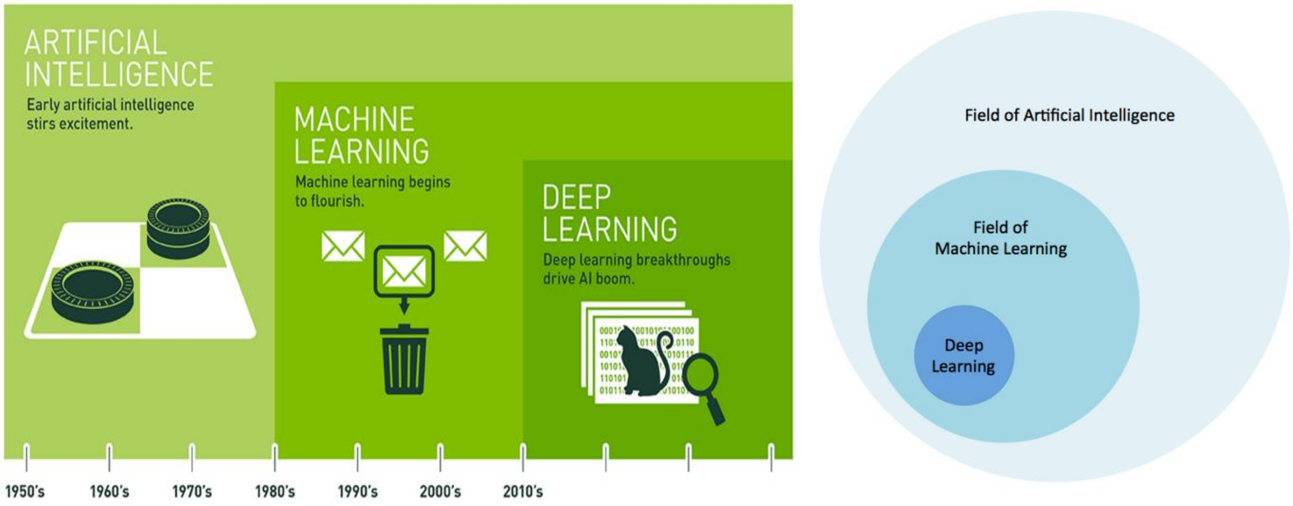
AI systems reason, and in particular, draw conclusions (inference or computing), and make guesses about what is going to happen next (predict and train). AI and Deep Learning exploded in 2015, especially with the availability of GPUs for parallel processing. AI involves 3 key steps: 1) Sensing with the infinite storage and flood of data, 2) Reasoning, and 3) Acting or predicting. AI involves sensing the data, including objects, faces, speech recognition or robotics. Reasoning involves connecting or relating words to what is known such as 1) Language processing, 2) Situation assessment, 3) Logic-based inference using CPUs and leading to learning and acting such as Natural Language processing, and speech and audio generation and robotic control. AI can be autonomous supervised or autonomous unsupervised learning.

Supervised learning or curated networks. This is when the machine is told what the correct answer is for a particular input, such as the system is shown an image of a car and told that the correct answer was "car." It is one of the most common training techniques for neural networks and other machine learning architectures. **The IBM Watson, we believe, is an example of a supervised learning or curated network framework.** Most commercial deep learning products use supervised learning, such to recognize a cat in a photo, a neural net will need to be trained with a set of labeled data of the different kinds of cats, sizes, colors etc. This tells the algorithm that there is a "cat" in the image, or there is not a "cat" in the photo. With enough images used to teach the neural network, it will learn to identify a "cat" in an image. Cognitive computing we believe is supervised learning and was a term coined we believe by IBM for the simulation of the human thought process in a computerized model.

Unsupervised learning/predictive learning or Neural Networks. Deep Learning uses unsupervised networks where the system learn to look at objects and understand, just as a normal human brain, or we as children grew up and learned by watching our parents. This is implemented to Frameworks (we look at them in a later section) to implement deep learning trying to mimic. **Unsupervised learning is used to discover new patterns and insights by approaching problems with little or no idea what our results should look like.** The important distinction between supervised and unsupervised learning is that a deep learning, unsupervised network has a feedback loop with a back propagation layer so the system can learn.

As we show below in Exhibit-2, AI is the catch all for all applications of AI/ML/DL.

Exhibit 2: Deep Learning; The Newest Driver of AI


ARTIFICIAL INTELLIGENCE
Early artificial intelligence stirs excitement.
MACHINE LEARNING
Machine learning begins to flourish.
DEEP LEARNING
Deep learning breakthroughs drive AI boom.
1950's
1960's
1970's
1980's
1990's
2000's
2010's
Field of Artificial Intelligence
Field of Machine Learning
Deep Learning

Source: Nvidia company website

Machine Learning

Machine Learning uses a suite of algorithms to go through data to make and improve the decision making process. Machine learning can be defined as:

1) the ability to learn without being explicitly programmed, as the systems uses algorithms to collect and sort data, and,

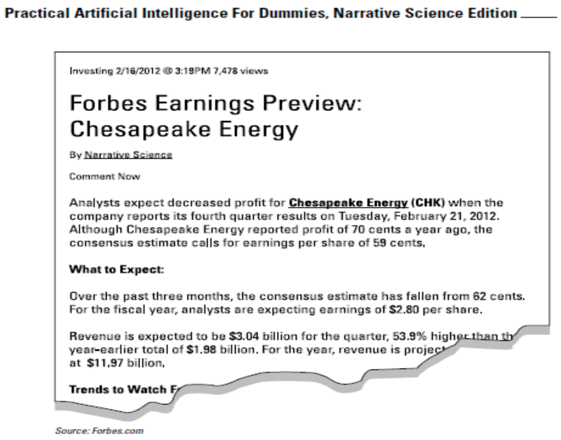
2) learn from past outcomes to determine a future prediction.

One of the most popular machine learning applications is image recognition, however the machine needs to be trained. A human must look at a number of pictures and tell the machine what they are, and after thousands of repetitions the machine is able to learn based on patterns. So the machine is trained with large amounts of data and algorithms to develop the capability to perform the function. Another form of Machine Learning is **Natural language processing (NLP)**, which is attempting to understand natural human communication, either written or spoken, and communicate in return with us using similar, natural language.

Use Cases: ML is used here to help machines understand the vast nuances in the human language and to learn to respond in a way that a particular audience is likely to comprehend – with typical applications in call centers, or to read millions of credit and mortgage originations and insurance documents. Other machine learning applications can include show recommendations on Netflix, Facebook newsfeeds, and featured Amazon products, all predictions based on patterns of existing data. Or for simple applications, such as when you make a typo in a Google search, it offers up "Did you mean?.."

We show below how ML can used to write up actual news pieces on the internet or even in an automated call center, telling customers locations or other data.

Exhibit 3: Machine Learning Applications in Research and Speech Recognition



Practical Artificial Intelligence For Dummies, Narrative Science Edition

Investing 2/16/2012 @ 3:19PM 7,478 views

Forbes Earnings Preview: Chesapeake Energy

By Narrative Science

Comment Now

Analysts expect decreased profit for **Chesapeake Energy** (CHK) when the company reports its fourth quarter results on Tuesday, February 21, 2012. Although Chesapeake Energy reported profit of 70 cents a year ago, the consensus estimate calls for earnings per share of 59 cents.

What to Expect:

Over the past three months, the consensus estimate has fallen from 62 cents. For the fiscal year, analysts are expecting earnings of $2.80 per share.

Revenue is expected to be $3.04 billion for the quarter, 53.9% higher than th year-earlier total of $1.98 billion. For the year, revenue is project at $11.97 billion.

Trends to Watch F

Source: Forbes.com

"There is a pizza restaurant called Gino's about three blocks from here."

Source: AI for Dummies

Deep Learning

Deep learning (a subset of AI and within ML) attempts to emulate brain functions with deep neural networks. An artificial neural network or neural net is a system that has been designed to process information in ways that are similar to the ways biological brains work. Neural nets are the basis of deep learning, and are designed to work the same way a human brain works. **Deep Learning can make sense of data using multiple layers of abstraction and hence the need for parallel or massively parallel processing architectures afforded by GPUs away from CPUs.** During the training process, a deep neural network learns to discover useful patterns in the digital representation of data, like sounds and images. The machine gathers and sorts unstructured data to generate a prediction, but with each further bit of data collection, the machines unsupervised capabilities improve.

Deep Learning is the ability for the system to work with 1) unstructured data, 2) ask untutored curated questions, 3) synthesize trends, and 4) predict out of raw data. A key differentiator for DL is a feedback loop or a backward propagation layer where the system learns or trains as new sets of data are received in real time.

Use Cases. Some applications of Deep Learning include how Facebook can automatically organize photos, identify faces, and suggest which friends to tag. Google can programmatically translate 103 languages with extreme accuracy, as well as Search, Gmail, and Maps. The Big 5 accounting houses track millions of global ledger entries in companies globally and across markets and industries daily, looking to find legitimate, logical, or illogical entries much faster than an army of accountants can.

Artificial Neural Network (ANN). ANNs use large datasets in order to produce a statistical outcome. These networks train by gathering large datasets from databases and perform tasks such as classification and image recognition. Like AI, ANNs are also limited to supervised training.

Neuromorphic or Cortical. Cortical solutions are based on biological premises and modeled after the human brain. It takes constant data streams that may or may not be labeled, and is able to predict outcomes, detect anomalies, and classify. Cortical machines continuously learn and are unsupervised, using one algorithm to complete multiple tasks. Numenta’s HTM (Hierarchical Temporal Memory) is an example of a cortical machine. One key player using Neuromorphic learning and NLP is Cortical, out of Austria.

Some Use Cases of AI and Deep Learning Across Markets; ’tis the Future!

Banking - Machine learning is being used at financial institutions in regards to loan originations. Large banks can originate over 500,000 loans per year, each of which needs to be looked over and approved by human eyes. This task is long and mundane, creating an environment prone to errors. Machines are able to sort through these loan summaries, eliminating errors and improving human error rates in origination and syndications.

Trading - AI programs are able to identify a number of anomalies when it comes to stocks. When anomalies are established, AI systems can pick out certain securities that are moving in ways that are typically not seen, either ahead of a spike up or a crash in the security prices, creating an opportunity for traders. The deep learning and AI algorithms sift through reams of data and correlations and combinations to learn and predict trends and upcoming anomalies.

Healthcare - one of the biggest areas AI is making an impact is in healthcare. In healthcare, AI and DL is used to approve the 1-2M insurance claims a month, such that the system looks at all available data and understands if the claim can be approved. In 90% of the cases, the claims can be approved without supervisory intervention as the system aggregates data from the customers' behavior, hospital visits, prior history, and medication much faster than a human operating a customer support center can. At hospitals, deep learning is being used to identify anomalies in x-rays in order to identify cancerous tumors 1mm in size, well before it reaches 1cm, which is significantly harder to treat. It's also being used in pharmacies such as Walgreens, where AI algorithms are used to predict Flu patterns, allowing pharmacies to have the correct amount of inventory of medications at a given time. DL can be used to scan through 10s of thousands of scans to identifying indicators for cancer in blood and tumors in MRI scans, better than can be detected by doctors and drive earlier attack of the malignant diseases.

Insurance - deep learning is also moving into the insurance industry. A key implementation is to look at claims approvals processing in home, flood, and property to speed up the claims process, again similar to healthcare where the system looks at all available data and understands if the claim can be approved. In 90% of the cases, the claims can be approved without supervisory intervention and much faster than humans. The insurance deep learning process can be used to predict spikes in claims and to make reserves for that or even to detect fraud in the system. For example, companies that offer auto insurance are able to insure lower rates based on the model driven as well as driving style. Accelerometers placed in the car can evaluate a person's driving style, recording data on acceleration, braking, intensity of turns, etc., rewarding more cautious drivers with a more favorable rate.

Customer Service - customer service is an area that should be able to find solutions on an expedited process with AI. While customers looking for a solution may have the same problem, it can be explained in number of different ways. This makes solving an issue for a customer more complicated even though a solution may be readily available. So, more and more AI and deep learning is being used to automate the customer support call center, from mundane customer support to deep technical support as the system is able to understand the nature of request looking at the volume of calls coming in and all available data in the system. By being able to narrow down the problem from different descriptions of the same problem, all potentially faster than current manual customer support centers, AI is able to use a number of "fingerprints" in data in order to identify similarities, reaching a solution in a far quicker timeframe versus traditional troubleshooting.

Security and Surveillance - one of deep learnings biggest strengths in computer vision. Cameras are able to be set up at multiple points of interest, such as an airport or busy downtown area, and able to recognize faces and security breaches. Deep learning is able to better identify legitimate threats. These solutions are able to help eliminate the over 1,000 false alarms per month at some facilities versus a less advanced system.

One the biggest single industry DL implementations, at the crossroad of ML and AI, is Automotive ADAS. One of the key verticals for AI Deep Learning is Automotive to reduce the 35-40,000 traffic fatalities and the millions of traffic injuries globally. Some AI systems onboard autos are able to learn frequently used routes after 25-30 trips and maneuver traffic patterns, driver habits, distractions and traffic speed limits and patterns. We believe processors such as MBLY but, more importantly, the Drive PX/PX2 from NVDA use multicore GPUs to drive deep learning in the car and learn from repetitive driving on the route, looking at lanes, traffic patterns and congestions, turns and traffic signals, and speed limits to make a more aware car and significantly safer driving experience. So, essentially what the Automotive ADAS DL network needs is training. It needs to see hundreds of thousands, even millions of images, until the weightings of the neuron inputs are tuned so precisely that it gets the answer right every time, under all weather conditions: fog, sun or rain.

So why AI and Deep Learning Now? A Coming of Age

Deep Learning has been in the making since the 1940-50s, but the building blocks came together more recently driving a broader underlying shift to AI and Deep Learning in multiple market segments from healthcare to insurance claims and approvals, mortgage applications, and fixed income debt markets, automotive to reduce traffic fatalities, and global trading patterns. **Key to deep learning is:**

1) Availability of historical and real-time data,

2) Connected data with the Cloud, Data Storage and Flash access,

3) Pervasive connectivity Wi-Fi .11ad/.ax,

4) Broad communication pipes with 4G, and last but not least,

5) Computing power with CPUs, accelerators, GPUs, FPGAs and massively parallel architectures. These drive a blossoming of knowledge from the last decade that was defined by computing, to the next decade defined by learning and predictive capabilities. What is driving AI today is basic Economics…

So what is driving AI today? Automation & Economics

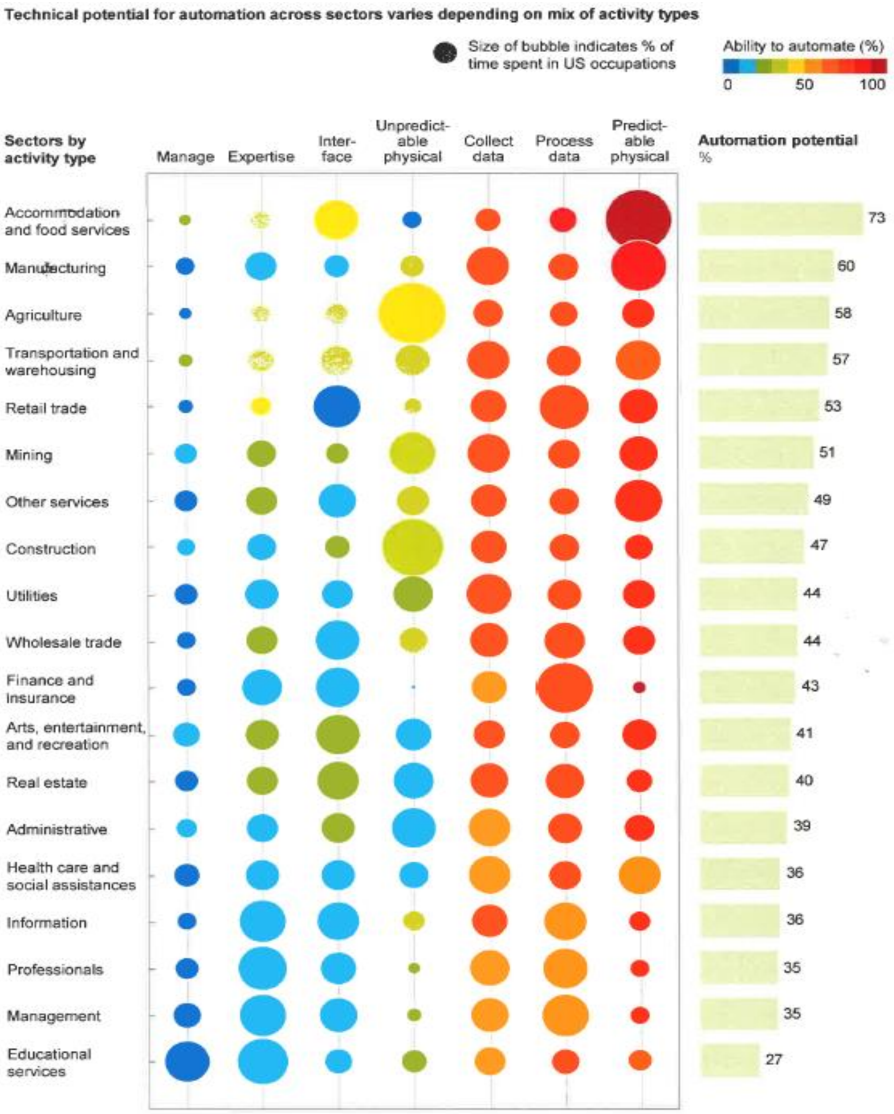
We believe one of the key reasons AI and Deep Learning is seeing broad adoption across multiple verticals and industry is the drive to increased levels of automation as machines can process the flood of pervasive big data and make interconnections between data to drive training and predictive capability much faster than humans.

This is especially true when it comes to laborious manual tasks as data matching in healthcare approvals, and insurance approvals. But machines can now be made smarter to analyze the flow of data and predict surges in hospital patients, pharmacy requirements across a network, or predict insurance losses with weather and historical trends. Any and all of these tasks can be performed, better, faster and significantly cheaper than an army of professionals with much less of the human intervention and downtime required.

AI and deep learning can also be used to train a system to look for fraud, intrusion and hacking, and also predict anomalies based on a confluence of multiple disparate factors even if they have not occurred before.

But as noted at the AI Futures Conference in January-2017 by McKinsey Consulting, broader adoption of Automation across industries is a key driver in the adoption of AI. As shown below, machine learning and deep learning can drive 30-50% further automation across multiple industry groups, redefining tomorrow’s workforce and driving significant productivity.

Exhibit 4: Automation….



Source: McKinsey Consulting

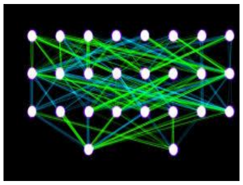
How big can AI get?

Only 1% of Servers today dedicated to AI. We believe Deep Learning (DL), machine Learning (ML) and AI are in its infancy with a long roadmap ahead of it. As INTC has noted only 1% of servers today are dedicated to AI, opening up significant opportunities for massively parallel implementation of processors to learn and predict from the data flow. Deep Learning and AI are expected to grow 12x by 2020E. The market for Neuromorphic chips (massively parallel CPU, GPU or FPGAs) for deep learning is expected to grow to ~25% of the Data Center market from 5-10% today or to put it different terms ~6x NVDA's current Data Center Revenues and presents an attractive TAM for incumbents with CPU, GPU, FPGAs, and new hardware entrants with massively parallel computing architectures.

Why the Architecture Change to GPUs? Why Can't CPUs do it?

Deep Learning requires multiple tasks to be performed in parallel when performing computing mimicking the brain. Our brain processes multiple sets of data in parallel while looking for patterns, linkages, and past trends before coming to a conclusion or making a prediction. But the GPUs and CPUs of today were made primarily for speed and computing than the highly parallel and logical thinking to drive predictive capabilities. We take a look at the Von Neumann computing architecture and new massively parallel multicore architectures below.

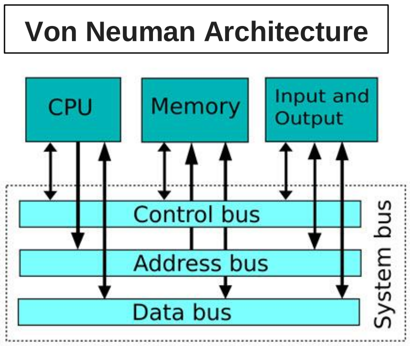
Most of the PC Server computing as we know today is based on the Von Neumann architecture where the focus is on high performance single threaded computing. And so, Moore's Law and Dennard's Scaling served the process to drive faster processing and higher frequencies. But as we show below, at the 16nm/14nm nodes, we hit the limits of Single Core performance as frequencies and speeds have peaked. So we now see the need for multicore processors to drive faster processing but also new parallel compute architectures driving platforms suited for distributed hyperthreaded processing or massively parallel computing architectures.

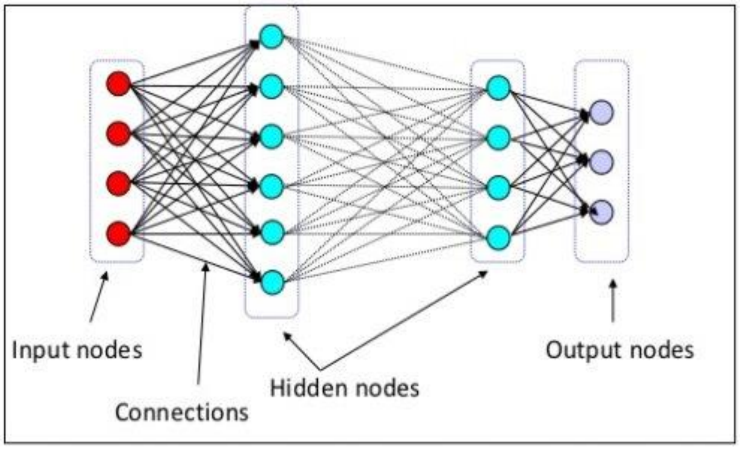


Parallel Neural Networks

GPUs are much more efficient than the traditional high performance single threaded Intel Xeon Cores, running almost 5-100x better performance of the leading edge Xeon in training or running parallel processing algorithms. Most Deep Neural Networks (DNNs) and Convolutional Neural Networks (CNNs) require massively parallel processing than single threaded performance. Google's simple "cat" image recognition neural network used 1 billion interconnections, 16,000 processors, and multiple reiterative learning before it could identify a cat as a cat. In parallel training environments, GPUs are significantly better than FPGAs and CPUs.

Exhibit 5: Today's Neural Networks Require Massively Parallel Architectures Rather than Single-threaded Performance in Von Neumann Architecture Computing Today





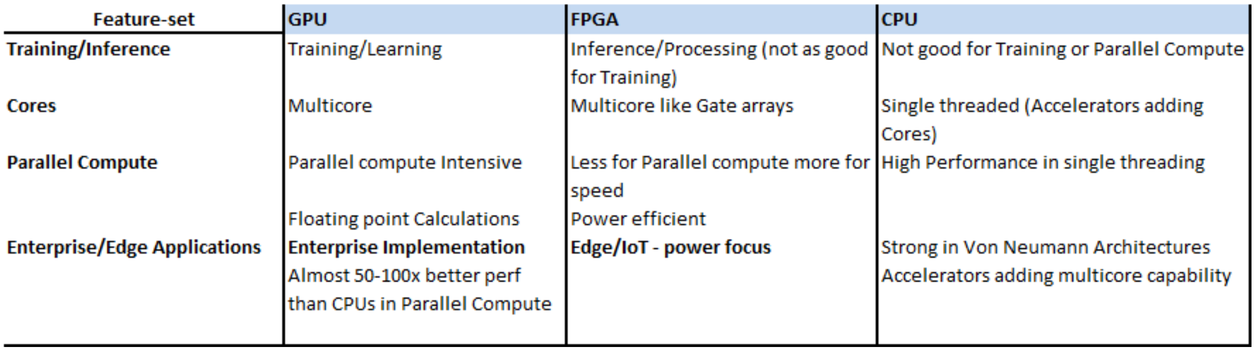
Source: KDnuggets.com; EdgeFxkits.com

But as AI becomes more prevalent in the next 5-10 years, we expect a bigger focus on power and energy efficient systems than the luxurious performance focused systems of today. **Deep Learning can be GPU-Lite, CPU-Lite or compute intensive requiring low power FPGAs to new, Massively Parallel architecture.** Deep Learning uses both computational and inferences using CPUs, and also training using parallel GPUs.

As we show there are many ways to skin the AI Deep Learning cat, but make no mistake, DL, ML, and AI are the future, as retail, healthcare, insurance and banking try to learn from clouds of mined data to predict our individual habits and predict future purchases, ailments, claims, finances and mortgage approvals.

The Role of GPUs, versus CPU and FPGAs. As we show in subsequent pages, DL, ML, and AI are being implemented with CPUs, accelerators, GPUs, and FPGAs as well as new massively parallel processors for new applications, but we believe training will be dominated by GPUs. As we show in Exhibit-6 below, GPUs are good for training offering 50-100x better performance versus CPUs. FPGAs are better for inference and offer much lower power consumption. Not to be left behind, CPUs with multicore gate arrays from FPGAs are also being implemented in deep learning.

Exhibit 6: GPU vs CPU vs FPGA: A Top-down Look



Feature-set	GPU	FPGA	CPU
Training/Inference	Training/Learning	Inference/Processing (not as good for Training)	Not good for Training or Parallel Compute
Cores	Multicore	Multicore like Gate arrays	Single threaded (Accelerators adding Cores)
Parallel Compute	Parallel compute Intensive	Less for Parallel compute more for speed	High Performance in single threading
	Floating point Calculations	Power efficient	
Enterprise/Edge Applications	**Enterprise Implementation** Almost 50-100x better perf than CPUs in Parallel Compute	**Edge/IoT - power focus**	Strong in Von Neumann Architectures Accelerators adding multicore capability

Source: Mizuho Securities research

AI …the Future with H2O …. New hardware with Graphcore...and Multicore Processors

The Hardware: Will AI and Deep Learning be done on CPUs, GPUs, FPGAs, or does the breakdown of Dennard's Scaling say time for Multicore?

For the past 20-25 years, computing has been about speed and lower power consumption. **As we show below, the power (P) consumed by a transistor is given by the product of the # of transistors (Q), the Frequency (F), the Area or Capacitance (C) of the transistor and threshold or operating Voltage (V), and any reduction in power by shrink overwhelmed by the Leakage currents:**

$$\boldsymbol{Transistor\ Power = QCFV^2 + VI_{leak}}$$

The power consumed by the transistor (as shown above) has been coming down as V has declined from 3V to 1V for the CPU and area of the transistors (C) have shrunk from 65nm to 10nm, in line with Moore's Law of the # of transistors doubling every 18-24 months. Also, according to Dennard's Scaling, the power consumed by the transistor decreases with the Voltage and current and so as power consumed declined with threshold voltage (**V**), OEMs have been able to raise operating Frequencies (**F**). And so, it all worked well until we hit a wall because, as the transistors came closer with the shrinks, the leakage current (I_{leak}) increased and overwhelmed any advantage from the reduction in Q, C, and V below.

A Breakdown of Dennard's Scaling drives the need to Multicore Processors

And so, the declining power consumption (**P**) of the transistor broke down at the 16nm/14nm node, as $\boldsymbol{I_{leak},}$ or the leakage current (the second term in the Equation above), between transistors started to overwhelm the shrink of transistor sizes put a base power wall or power consumption. Despite using Hafnium gates and 3D FinFET, frequencies of INTC CPUs have stayed at the 1-3GHz levels as leakage currents have increased. Now, OEMs with the breakdown of Dennard's Scaling and the inability to increase frequencies are moving to Multicore processors to improve performance and frequencies. Multicore processors provide increased speedup of the processor according to Gene Amdahl's law.

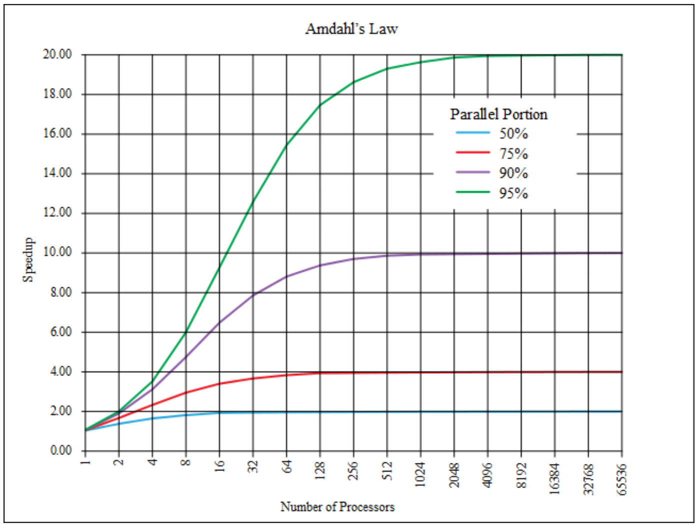
The Speedup (**S**) and performance of the processor in a multicore parallel processing environment is proportional to the number of cores (**N**):

Gene Amdahl's law with processor speed up linearly proportional to the number of cores…

$$\boldsymbol{S_{speedup(Amdahl)} = \frac{1}{(1-F)+F/N}\ or\ S_{speedup} \approx N}$$

Theoretically, it appears there is no limit to multicore performance scaling if applications with neural networks or parallel processing grows with the number of cores. **GPUs can provide multithreaded parallelism**, but also, as shown below, the 50-60 cores in GPUs today can drive faster processing and speedup in parallel architectures.

Exhibit 7: Amdahl's Law: Increasing Processor Speed up Proportional to the Number of Cores and the Amount of Parallel Processing Required



Source: Extremetech.com

The multicore processor can also be implemented with a large core processor CPU and multiple smaller cores, such as a GPU/Accelerator implementation. The CPU enables the inference portion of the application processing. FPGAs have fields of programmable gates that can be programmed into many parallel paths, to create task specific cores that run like parallel computing CPUs so the hardware allows for high throughput.

So Why GPUs? Are FPGAs taking over?

More than 12 years after IBM started into the age of multicore processors with the IBM Power4, the first commercial dual core processor chip, we believe there is more focus on multicore GPU platforms away from single threaded single core high performance compute.

Expect GPUs to Dominate Training or Learning in DL/ML/AI for a While…..

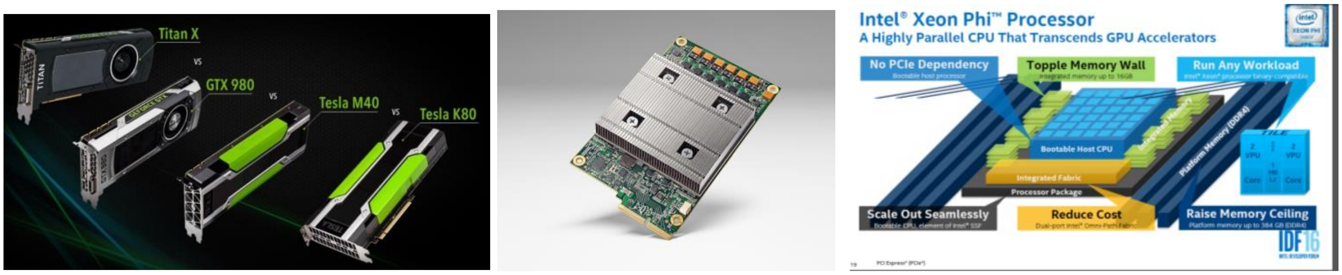
Today, we believe multicore GPUs are vital and key to parallel processing and training and the backbone of deep learning and AI. GPUs will provide the key backbone for training in deep learning and Neural networks, because of the programmability and flexibility. GPUs will also continue to dominate the core enterprise and PC/Server computing environment. **We believe in multiple deep learning and ML environments in healthcare, search, retail, social networking, automotive ADAS, finance, banking, and trading, we could see GPUs dominate initially as data scientists use raw unstructured logic to train the systems and could be a strong multi-year trend.** We see NVDA dominating the GPU trend as it also provides a CUDA platform. AMD is also starting to benefit from the GPU trend for deep learning but primarily because OEMs and hyperscale are looking for an alternative supplier to the dominant NVDA.

FPGAs are not good at training for deep learning algorithms. But, once the logic is frozen, which we believe is post a process of learning/training with the GPUs, we believe, FPGAs provide a cost-effective, high-performance and low-power solution. So FPGAs will potentially be used more in inference/calculations (than training). **Also as we noted while GPUs could dominate core enterprise, the Edge/IoT where new implementations are happening in industrial and medical could see FPGAs get design wins in the open architecture environment.**

FPGAs more for inference (not Training/Learning) but can help with better Performance and Power….

FPGAs have a limited number of gates and hence the number of nodes to be used for training. Training is also much more computationally intensive, requiring constant adjustment of parameters and making it harder to implement the algorithm architecturally in FPGA hardware. **But FPGAs consume much lower power versus GPUs (as much as 10x more power efficient) and so, one of the attractions with FPGAs with 2-4x better performance versus GPUs in inference (non-training environment when the logic has been fixed).** Notably once the logic is gleaned out and set, it is feasible to move to cheaper ASICs for deep learning implementation at the Edge/IoT and in some cases Enterprise.

Exhibit 8: A Look at GPUs, Xeon Phi CPUs, FPGAs, TPU and the new Graphcore



Intel® Xeon Phi™ Processor
A Highly Parallel CPU That Transcends GPU Accelerators

No PCIe Dependency
Bootable host processor

Topple Memory Wall
Integrated memory up to 16GB

Run Any Workload
Intel® Xeon® processor binary-compatible

Scale Out Seamlessly
Bootable CPU, element of Intel® SSF

Reduce Cost
Dual-port Intel® Omni-Path Fabric

Raise Memory Ceiling
Platform memory up to 384 GB (DDR4)

Source: DeepLearning4j.org, Google Blog, Intel company presentation

We should not write off the core CPU/Accelerator implementation for Deep Learning as we noted there are potential implementations with a large central core processor (Intel CPU) and multiple smaller parallel cores (FPGA Altera accelerators) to implement DL as with the Intel Xeon Phi. Intel has been trying to beef up its DL offerings with multicore and with acquisitions such as Nervana systems, and offering Deep Learning as a service in the Cloud. Given a strong data center presence with 96% market share, INTC has been focusing on deep learning in Cloud server computing.

Graphcore - New Hardware platform IPU

But, as we noted prior the speedup and higher frequencies achieved for parallel programming are a key feature in neural networks and could improve with higher cores. So we are now seeing new hardware platforms such as Graphcore's IPU (Intelligent processing Unit) and a neural network accelerator, which can have from 100 to 1000 to 3000 cores, with the first chips expected in 2H17. But as in all hardware platforms, there will be a big learning curve.

We would also note Google has been working on its own hardware architecture TPU (Tensor processing unit) that could be competition to GPUs, CPUs, accelerators, and FPGAs. Google's deep learning hardware is used across its Android platform from Street View to voice search, among others. **Google also used its TPU platform to teach its AlphaGo program the game Go, and compete in a match against world champion, Lee Sedol.** The Google AlphaGo TPU platform essentially played Go versus itself 1000's of times, learning and mastering the game before playing against the 18-time World Champion and beating him. The TPU platform is based on ASICs delivering better power performance in the data center for Google. But we believe deep learning adoption in the industry is so broad that it will be a tailwind to the GPU OEMs NVDA and others in the ecosystem. There are also other multicore architecture and logic suppliers such as Wave Computing with their Dataflow processors noting 16,000 independent elements/cores/processors that can speed up to 6.7-8GHz versus current leading edge Intel Xeon at ~3GHz.

And the Future. While today much of Neural Network processing on parallel hardware is conducted in the Cloud and on-premise in Enterprise, we could see parallel processing and DL/ML moving to the device on-to handsets, drones, and autonomous EV/HEV vehicles where power performance becomes important.

A Look at Some of the Major Players: New AI Algorithms and Hardware; H2O.ai

We believe many of the AI platforms below are being implemented on GPUs as the predominant and in some cases sole hardware platform. But also, CPU/Accelerators, FPGAs and DSPs are seeing traction as well. Some of the major AI startups include H2O we believe which is leading the space with deep learning and AI implementation in the Enterprise over broad sectors such as Insurance, Pharmacy, healthcare, auditing and banking. We would also note other players such as Loop AI, Cortical, Numenta (a startup founded by the founder of Palm and Handspring), and Clarifai, among others. We would also note new hardware platforms such as Graphcore, DeePhi, KnuPath and Brainchip.

Nvidia – The AI GPU King

We believe NVDA is one of the key AI hardware leaders with its Pascal GPU architecture purpose built for AI. The company's DGX-1 AI supercomputer is specifically designed for deep learning and analytics, fully integrated with hardware, DL software, developmental tools, and analytics applications. NVDA noting that the DGX-1's performance is equal to ~250 conventional servers. NVDA is exposed across virtually all AI segments as we believe AI companies see NVDA as the premier AI hardware offering. In the OctQ, NVDA data center revenues were up 193% y/y, we believe with significant strength driven by AI and deep learning demand.

Exhibit 9: Nvidia DGX-1 AI Supercomputer

Source: Nvidia company website

H2o.AI - we believe a Leader in Software

H2O.ai is focused on bringing AI software to the business world. The company offers an open source machine learning platform for smarter applications and data products. Applications for H2O include predictive maintenance, operational intelligence, security, fraud, auditing, churn, credit scoring, insurance, and ICU monitoring at customers including Capital One, Progressive, UnitedHealth, HCA, Zurich NA, Transamerica, Comcast, Macy's, and Walgreens, among others. H2O has also implemented Deep learning at eBay and for fraud detection at PayPal. We would note H2O's new open sourced Deep Water framework brings multiple learning frameworks such as Tensorflow, MXNet, and Caffe together. It has also shown training on complex multilayer artificial neural networks (ANNs) **using GPUs at 10-75x speed of CPUs**. Some of H2O's networks have 1000s of layers of data processing. We believe H2O is based primarily on a NVDA GPU/CUDA platform. H2O has been able to implement deep learning in manufacturing environment to detect and predict failure in wind turbines located in geographically inaccessible areas that could benefit from predictive maintenance capabilities.

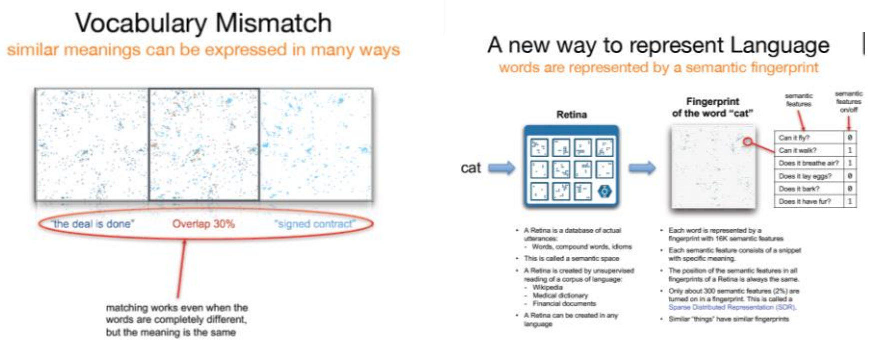
H2O is one of the leaders with an industry leading Training time and broad GPU-based Java/Scale implementation across Enterprise….in Insurance, Pharmacy, Banking, Healthcare Industrial

Loop AI

Loop AI was founded in 2012 and is focused on unlocking "dark data," meaning finding a way to process the 90% data that computers currently store but are unable to understand. The company uses its Loop Q platform to enable robotic process automation (RBA) to unlock 100% of all data, giving the computer human cognitive power. The Loop Q platform is based on principles inspired by the neocortex, the brains center for language and reasoning. With its own algorithms, Loop AI aims to integrate hardware and software without having to offer human guidance or labeling.

Loop AI is focused in sectors including aerospace and defense, automotive, banking, healthcare, media, oil and gas, power and utilities, technology, telecommunications, and retail.

Exhibit 10: New AI Implementations with Numenta and Cortical.ai …



Vocabulary Mismatch

similar meanings can be expressed in many ways

matching works even when the words are completely different, but the meaning is the same

A new way to represent Language

words are represented by a semantic fingerprint

Can it fly?	0
Can it walk?	1
Does it breathe air?	1
Does it lay eggs?	0
Does it bark?	0
Does it have fur?	1

- A Retina is a database of actual utterances:
  - Words, compound words, idioms
- This is called a semantic space
- A Retina is created by unsupervised reading of a corpus of language:
  - Wikipedia
  - Medical dictionary
  - Financial documents
- A Retina can be created in any language

- Each word is represented by a fingerprint with 16K semantic features
- Each semantic feature consists of a snippet with specific meaning.
- The position of the semantic features in all fingerprints of a Retina is always the same.
- Only about 300 semantic features (2%) are turned on in a fingerprint. This is called a Sparse Distributed Representation (SDR).
- Similar "things" have similar fingerprints

Source: Cortical.io

New Massively Parallel Architecture with Graphcore

Graphcore is a startup developing a machine learning processor, we believe funded by Samsung and Robert Bosch. Graphcore's platform, an intelligent processor unit (IPU), is built on 16nm FINFET from TSMC, massively parallel, low-precision floating-point compute and a much higher compute density than other solutions. Graphcore's first chip solutions expected in 2H17 aim to drive massively parallel compute to 1000-3000 cores (versus 70/100 cores in GPUs and 15-30 cores in Xeon). As in all new hardware solutions, there will be a learning curve. Graphcore's CEO is Nigel Toon.

BrainChip - an FPGA Variant with Autonomous Logic

BrainChip developed a Spiking Neuron Adaptive Processor (SNAP) that learns, evolves, and associates information automatically, similar to the human brain. We believe BrainChip currently focuses on supervised and unsupervised autonomous learning, mostly in security surveillance and gaming casino applications. BrainChip uses an FPGA platform and estimates that the Neuromorphic Chip Market is estimated to be ~$4.8B by 2022 with a CAGR of ~26% between 2016 and 2022. BrainChip's CEO is Louis DiNardo, who was previously at Exar.

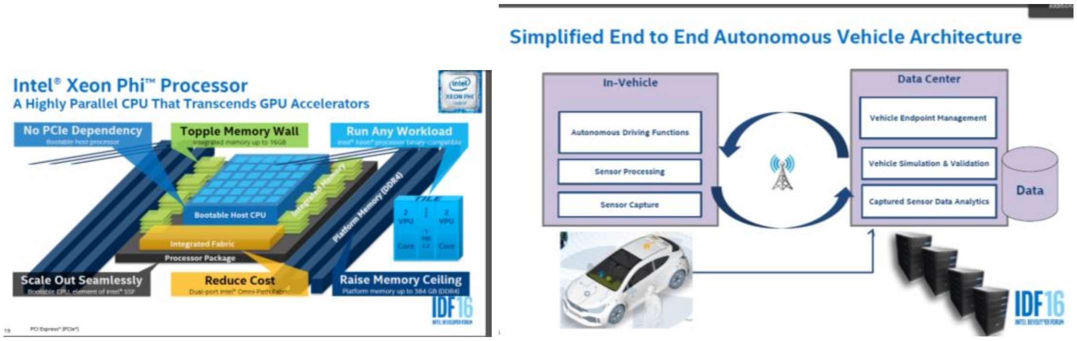
New Entrants from China – DeePhi with FPGA Variants

Another major startup is DeePhi, a collaboration with China's leading University Tsinghua. Overall, DeePhi has developed a complete automation flow of compression, compiling, and acceleration which achieves joint optimization between algorithm, software and hardware. A smaller, faster and more efficient deep learning processing unit (DPU) will eventually be released to public. DeePhi noting its FPGA based DPU platform is more power efficient for drones, surveillance, and image recognition.

Intel with Xeon Phi Accelerators

Intel has the multichip platform pairing a 14-nanometer Xeon E5-2600 v4 "Broadwell" with Arria10 field-programmable gate arrays (FPGAs) from Altera. Currently, Altera's FPGAs are used in the Microsoft Azure Cloud, but we believe mostly for inference rather than training. As we have noted in our hardware assessment, FPGAs have different capabilities and offer competitive value in different faces of ML/DL/AI.

Exhibit 11: Growth in the Deep Learning Software Market; Intel’s ADAS roadmap with DL


Intel® Xeon Phi™ Processor
A Highly Parallel CPU That Transcends GPU Accelerators

Simplified End to End Autonomous Vehicle Architecture

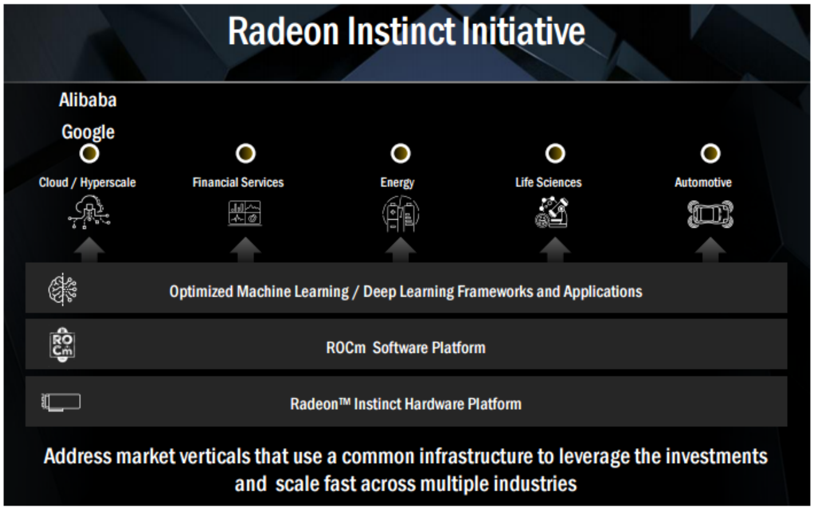
In-Vehicle
Autonomous Driving Functions
Sensor Processing
Sensor Capture
Data Center
Vehicle Endpoint Management
Vehicle Simulation & Validation
Captured Sensor Data Analytics
Data
IDF16
INTEL DEVELOPER FORUM

Source: Intel IDF Presentations 2016

AMD Radeon Instinct

AMD is jumping into the Machine Learning game with its recently announced Radeon Instinct GPU accelerators. The company sees opportunities across multiple end markets including cloud/hyper scale, financial services, energy, life sciences, and automotive. Radeon processors run on a ROCm (Radeon Open Compute) Software Platform, an open platform. AMD also recently announced that it will be partnering with Google to supply Radeon processors for Google Compute Engine and Google Cloud Machine Learning. **While we believe NVDA GPUs are currently the top choice for many Deep Learning and AI companies, AMD is well positioned for any increase in high performance GPUs as a key second source, especially as hyperscale and Enterprise look to beef up the supply chain.** AMD estimates the total addressable data center market is ~$18B, which we believe will be further driven by advances in machine learning driving the need for storage.

Exhibit 12: AMD Radeon Instinct Market Opportunities: Cloud, Financial, Energy, Automotive



Radeon Instinct Initiative

Alibaba
Google

Cloud / Hyperscale
Financial Services
Energy
Life Sciences
Automotive

Optimized Machine Learning / Deep Learning Frameworks and Applications

ROCm Software Platform

Radeon™ Instinct Hardware Platform

Address market verticals that use a common infrastructure to leverage the investments and scale fast across multiple industries

Source: AMD Company Presentation Dec 2016

CEVA

CEVA offers a number of neural network products, primarily DSPs (Digital signal processors) with imaging and computer vision processor IP designed to bring deep learning and artificial intelligence capabilities to lower power embedded systems in Automotive. The company also offers its Deep Neural Network (CDNN) Toolkit which is optimized for its XM family (XM4, XM6) of imaging and vision DSPs. The CDNN includes CEVA's software framework, network generator, and hardware accelerator all working together for better performance in the evolving landscape of machine learning. At CES 2017, CEVA announced that ON Semiconductor (ON, Buy) is licensing the company's imaging and vision platform for ADAS products.

Cortex A – Softbank / ARM

Not to be left behind, we believe Softbank (which now owns ARM) also has some hardware offerings in low power, multicore processing with its Cortex-A ARM processors. We believe at the ARM Techcon 2016, ARM has shown new Cortex processors for Machine vision and intelligence. Softbank CEO, Masayoshi San, saying there will be 1 trillion connected devices by 2020.

We would also note other hardware implementation with **KnuPath** which uses a DSP with the potential to integrate FPGA, and also others such as **Wave computing**.

How is AI Implemented? What are the major Frameworks?

DL, ML, and AI are implemented using training or learning frameworks in the Enterprise, cloud, or end devices, which could run on separate appliances or on top of current enterprise systems. The learning frameworks provide a backbone to drive efficient algorithms. The major frameworks include Caffe and others implemented on CPUs, GPUs, FPGAs, or TPU (Google's Tensor Processing Unit).

While there are multiple frameworks today, longer term we expect standardization of the Frameworks to drive further Deep Learning. Most of the major DL, ML and AI frameworks are based on either Python or some of the broader industry standards such as Java and can run on Hadoop/Spark. There are a lot of computational frameworks out there, many with few libraries or training models, and some without much commercial support and bare documentation. But we believe as deep learning gathers speed and broad adoption in the Enterprise expect a standardization of AI/ML/DL frameworks.

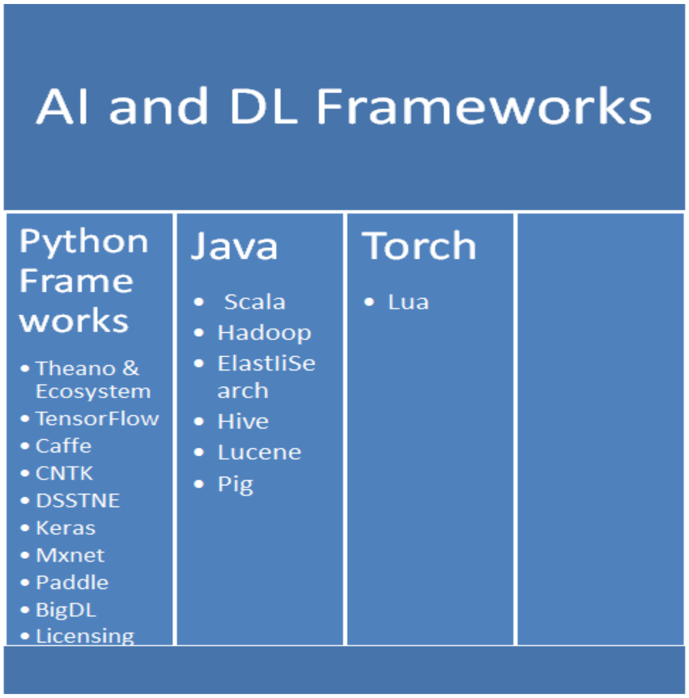
A look at some of the popular AI frameworks below:

Theano is one of the oldest deep learning frameworks. Theano runs mostly on GPUs and has noted almost 140x performance improvement in parallel compute using multicore GPUs versus CPUs. Theano is now being replaced with TensorFlow, another Google framework founded with Theano creator, Ian Goodfellow.

Caffe is another machine vision deep learning framework using C/C++ and was founded as a PhD project by Yangqing Jia at Berkeley. Caffe is widely used to write parallel compute layers for GPU and CUDA platforms.

Torch/Lua is a computational ML framework for deep learning used in some instances by Facebook and Twitter. Torch and Lua run on GPUs.

Exhibit 13: AI and Deep Learning Frameworks



AI and DL Frameworks

Python Frame works

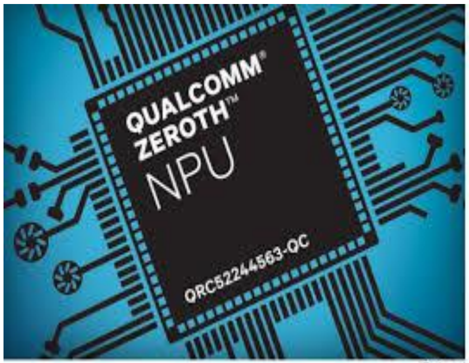
- Theano & Ecosystem
- TensorFlow
- Caffe
- CNTK
- DSSTNE
- Keras
- Mxnet
- Paddle
- BigDL
- Licensing

Java

- Scala
- Hadoop
- ElastliSearch
- Hive
- Lucene
- Pig

Torch

- Lua



(Qualcomm)

Source: DeepLearning4j.org, Mizuho Securities research

In 2016, Microsoft released its open source AI license with Computational Neural Toolkit (**CNTK**) which uses Python and C++ code.

DSSTNE - is Amazon's Deep Scalable Sparse Tensor Network Engine (DSSTNE) and follows the MXNet which is used in Amazon Web Services (AWS).

Baidu Paddle – There are also other open source frameworks such as Baidu's open source deep learning platform called Paddle (Parallel distributed Deep Learning).

Qualcomm has its Zeroth platform for on-handset deep learning using its Snapdragon processors.

We believe the key steps to implementing an AI framework in enterprise include

- **Using Existing Training Models such as AlexNET, AllCNN, GoogleNET or others with the Framework.** We believe part of the AI/DL implementation starts with implementing the basic training framework noted above.
- **Train with historical and real time Data.** Big data and available real time data is used to continuously train the system.
- **Evaluate the Hardware, Software and Libraries** using the Frameworks.

- **Or go to the Cloud: Deep Learning as a Service.** Though on-site is faster, we believe Intel with Nervana has been looking at implementing DL more in cloud computing environments.
- **And lastly, develop a Scalable solution, with the need for Ludicrous (not Light Speed) as noted by Caffe founder** Yangqing Jia at the AI futures conference earlier this year.

Conclusion: We hope this preliminary primer gives a brief bird's eye view of DL, ML, and AI, and the crucial paradigm shift in computing to massively parallel computing and what the future holds.

Vol.	Past Editions	Date
I	Connecting the Dots....Handset Markets into 2H15...4G..	June 2, 2015
II	NXPI, ON, CY Automotive-What is NCAP? A Look at ADAS L2/3	June 7, 2015
III	DRAM – Can We Expect Discipline into 2016?	June 15, 2015
IV	Global Automotive OEMs - Electrification and ADAS	June 30, 2015
V	3D-NAND – Are we there yet?	July 7, 2015
VI	Why 2016 Should be Another Year of Strong RF Growth	July 27, 2015
VII	A Look at the EMV Ramp in the U.S.	Aug. 6, 2015
VIII	Payments Takeaways - Strong 2016 with China, NFC, and U.S. EMV	Aug. 16, 2015
IX	Display OEMs Join In-Cell; SYNA Well Positioned	Sept. 13, 2015
X	Global Auto Sales through August; We Like NXPI, CY, ON	Sept. 17, 2015
XI	Why You Should Buy Integrated RF Suppliers AVGO, SWKS, and QRVO	Sept. 22, 2015
XII	Handset RF Outlook - A Changing Landscape	Nov. 13, 2015
XIII	Takeaways from our NAND Call: The Enterprise Roadmap	Nov. 13, 2015
XIV	SYNA and FPC: The Mobile FP and Biometrics Landscape	Nov. 22, 2015
XV	In NAND, 3D is Key as TLC becomes so "last year"	Nov. 29, 2015
XVI	Samsung 18nm DRAM in mid-2016?	Nov. 29, 2015
XVII	Fingerprints: A Call with the King	Dec. 3, 2015
XVIII	Strong November 4G Adds, China 5/6 mode 2016 Tailwinds	Dec. 28, 2015
XIX	Autos - A Tax Cut (China) and 2017 Hike (Japan)	Jan. 3, 2016
XX	ADAS Gets a Disruptive Quanergy Boost; Positive for NXPI, CY, ON	Jan. 21, 2016
XXI	A Look at the Challenged PC and Storage Markets	Jan. 25, 2016
XXII	Storage Weak - SSD/HDD and a Strong Yen	Feb. 15, 2016
XXIII	Quanergy Accelerates ADAS and 3D-Mapping	Feb. 21, 2016
XXIV	China 4G Transition from 2G Gaining Steam	Feb. 28, 2016
XXV	NXPI-Apple Pay with CUP Drives Mobile Payments	Feb. 28, 2016
XXVI	Fingerprints and NFC-China Driving Growth with Payments	Mar. 16, 2016
XXVII	A Look at the Virtual Reality Roadmap with AMD	Mar. 20, 2016
XXVIII	A 3D-NAND and SSD Industry Update, and the HDD Roadmap	Mar. 22, 2016
XXIX	Huawei - Key Handset & Carrier Takeaways	Mar. 31, 2016
XXX	A Preliminary Look at VR; Closing in on Gaming	May 11, 2016
XXXI	Magna Takeaways; Focus on NXPI, CY and ON	May 12, 2016
XXXII	Tailwinds from Automotive Sales and AEB	June 13, 2016
XXXIII	Previewing Huawei, a Global Telecom and Networking Juggernaut	June 15, 2016

XXXIV	***Updated Memory/HDD Model; Stable DRAM into 2017***	***June 20, 2016***
XXXV	***BREXIT - What's Next for Autos?***	***June 26, 2016***
XXXVI	***Data Center IT Survey & Key 2016 Storage Trends***	***July 26, 2016***
XXXVII	***ADAS - New Roadmaps...New Disruptors...***	***Aug. 18, 2016***
XXXVIII	***A Look at YE2016 Handset RF/4G Trends***	***Aug. 23, 2016***
XXXIX	***The Race to ADAS, LiDAR and Alliances***	***Sept. 16, 2016***
XL	***Huawei Class of 2017: Better Memory, TDDI, FP Diversification***	***Oct. 13, 2016***
XLI	***Autos, Infineon LiDAR and Disruptive Business Models***	***Oct. 17, 2016***
XLII	***Global YTD Auto Sales Strong in China; Positive for ON and CY***	***Oct. 23, 2016***
XLIII	***A 3D-NAND Crystal Ball; is 2018 the Peak?***	***Nov. 27, 2016***
XLIV	***Autos Leading with LiDAR; is Computing Power Hyped?***	***Nov. 29, 2016***
XLV	***ADAS 2017 Another Step Closer; LiDAR Comes of Age***	***Jan. 12, 2017***

Glossary

- ~ - approximately
- 1H/2H - first half/second half
- 2G/3G/4G - 2nd generation, third generation. 4th generation wireless
- 3G/4G - Third generation / Fourth generation
- 4G-LTE - Fourth generation, long term evolution
- ADAS - Automotive driver assist systems
- AI - artificial intelligence
- ANN - artificial neural network
- APU - accelerated processing unit
- ARM - a family of instruction set architectures used for processors for computers, servers, etc
- ASIC - application specific integrated circuits
- ASP - average selling price
- ATM - Asynchronous Transfer mode
- B - Billion
- BAW - bulk acoustic wave filters
- BiDi - Bi Directional
- BOM - bill of materials
- bps - basis points
- CA - carrier aggregation
- CAGR - compound annual growth rate
- CAPEX - capital expenditures
- CDMA - code division multiple access
- CEO/CFO - Chief Executive/Chief Financial
- CES - consumer electronics show
- CF - cash flow
- CFIUS - Committee on Foreign Investment in the United States
- COO - Chief Operation Officer
- CSP - chip scale packaging
- CY - calendar year
- D/E - debt to equity
- DCF - discounted cash flow
- DL - deep learning
- DoD - Department of Defense
- DRAM - dynamic random access memory
- DT - desktop
- e.MMC - Embedded managed NAND solution
- EBITDA - earnings before interest, taxes, depreciation and amortization
- eMCP - embedded multi-chip module using DRAM and NAND
- EMV - Europay, MasterCard and Visa, a payment consortium
- EOY - end of year
- EPS - earnings per share
- ET - envelope tracking
- ETD - Emerging Technologies Division
- EU - European Union
- EV - enterprise value
- EvDO - Evolution Data Only
- EVP - Executive Vice President
- F - Fiscal
- FASB - Financial Accounting Standards Board
- FBAR - film bulk acoustic resonator, a type of filter
- FBAR/BAW -Film Bulk acoustic resonator/Bulk acoustic wave Filters
- FCF - free cash flow
- FDD LTE - Frequency Division Duplex Long Term Evolution
- FP - finger print
- FPGA - field programmable gate arrays
- FT - force touch
- FTC - Federal Trade Commission
- FY - fiscal/full year
- GAAP - generally accepted accounting principles
- Gb/GB - gigabytes/Gigabit
- Gb/s - Gigabit per second
- GF - Global Foundries
- GHz - gigahertz
- GM - gross margin
- GPU - graphics processing unit
- GSM - Global System for Mobile Communication
- HDD - hard disk drive
- HSA - heterogeneous system architecture combining x86 and ARM
- HSD - high single digits
- I/O - input output operations
- IC - integrated circuits
- IoT - internet of things
- IP - intellectual property
- ISM - Institute for Supply Management
- ITU - International Telecommunication Union
- JPY - Japanese yen
- JV - joint venture
- K - Thousand
- Kbps/Mbps - Kilobit per second/Megabit per second bandwidth
- KGD - Known Good Die
- Kwpm - thousand wafer per month capacity
- LIBOR - London Interbank Offered Rate
- LQ - last quarter
- LSD - low single digits
- LT - long term
- LTE - long term evolution, a 4th generation wireless protocol
- LY - last year
- m - Meters
- M - Million
- M&A - mergers and acquisitions
- MB - megabyte
- Mbps - megabit per second
- MCU - micro controller unit
- MHz - megahertz
- MIIT - China Ministry of Industry and Information Technology
- MIMO - multiple in, multiple out
- ML - machine learning
- MLC - multi level cell
- MoE - merger of equals
- MOFCOM - Ministry of Commerce People's Republic of China
- MSD - mid single digits
- MSM - multi station modems, QCOM's QCT chips
- NAND - "not and," a type of memory
- NB - notebook
- NDRC - National Development and Reform Commission in China
- NFC - near field communications
- NLP - natural language processing
- nm - nanometer
- NN - neural network
- NOL - Net Operating Losses
- NOR - a type of non-volatile storage memory
- NPV - net present value
- NVMe - Non-volatile Memory Express
- NYSE - New York Stock Exchange
- ODM - original design manufacturer
- OEM - original equipment manufacturer
- OFN - optical finger navigation
- OLT - Optical Line Termination or Terminal
- OM - operating margin
- ONU - Optical Network Unit
- Opex - operating expenses
- P/B - price to book value ratio
- P/E - price to earnings
- P/S - price to sales
- PA - power amplifier
- PAD - power amplifier duplexer, essentially 2 filters and a Power amplifier
- PC - personal computer
- PCIe - Peripheral Component Interconnect Express
- PMI - Purchasing Managers' Index
- PMIC - Power Management Integrated Circuit
- PoE - Power over Ethernet
- PSD - Programmable systems division
- PSoC - programmable system on a chip
- PT - price target
- Q - quarter
- q/q - quarter over quarter
- QCT - Qualcomm chip technologies
- QTL - Qualcomm technology licensing
- R&D - research and development
- Rev - revenues
- RF - radio frequency
- ROI - return on investment
- RSP - Renesas Semiconductor products/Drivers
- RSU - Restricted Stock Units
- SAS - serial attached SCSI (small computer system interface)
- SAW - surface acoustic wave filters
- SDH - Synchronous Digital Hierarchy, mostly in Europe
- SG&A - Sales, General and Administrative
- SLAC - subscriber line audio-processing circuit
- SLIC - subscriber line interface circuit
- SMIC - Semiconductor Manufacturing International Corporation
- SoC - system on chip
- SONET - Synchronous Optical Network, used in North America
- SOX - Philadelphia semiconductor index
- SRAM - static random access memory
- SSD - solid state drive
- TAM - total available market
- TD - time division
- TDDI - touch display driver integration
- TDD-LTE - Time Division Duplex Long Term Evolution
- TD-SCDMA - Time Division Synchronous
- TLC - triple level cell
- Tx/Rx - Transmit / Receive
- USB - universal serial bus
- VCSEL - vertical-cavity surface-emitting laser
- wpm - wafer per month
- x86 - Intel based processor architecture
- XMC - Wuhan Xin Xin Semiconductor Manufacturing Corporation
- y/y - year over year
- YE - year end
- YTD - year to date

Price Target Calculation and Key Risks

Advanced Micro Devices, Inc.

Price Target: We have AMD at Buy with a $13 PT, based on 2.5x our F18E P/S. We believe AMD is well positioned to execute in the gaming, VR, and AI markets with multiple new product releases and high demand in each of these segments.

Risks: AMD competes in a cyclical, technologically intensive industry and sells to a concentrated customer base. Its ability to meet its own or our financial expectations and achieve future growth is subject to a number of risk factors, including, but are not limited to, the following:

- Demand for AMDs products is variable and could differ from expectations;

- Gross margin percentage could vary significantly;

- Competition and pricing pressure from other low-cost OEMs, ODMs and suppliers;

- AMD relies on third party manufacturing;

- Unexpected changes in legal and regulatory requirements, tariffs and exchange rates, political and economic stability, staffing and management issues, and potentially adverse tax consequences for its international operations;

- Potential loss of intellectual property, Commercialization of competing technologies;

- Litigation Risks;

- Adverse effects of potential possible future patent or other litigation

Intel Corporation

Our 12-month price target of $42 is based on ~15.3x our 2017 EPS estimate of $2.74 plus cash. In addition to competitive risks from a broad array of semiconductor and OEMs, macroeconomic risks and new product execution risks could impede the realization of our target price. INTC should see further LT upside from machine learning, deep learning, and AI. INTC has traded between a 10-16x forward P/E over the last five years.

Risks. We believe the risks to INTC are from a maturing PC market and limited traction in wireless and difficult comparables in tablets. Where we could be wrong is if INTC gets a significant foundry deal or makes a significant handset acquisition.

NVIDIA Corporation

Price Target: While NVDA's valuations are steep, we believe current street estimates are conservative, reflect licensing slowdown, so that improving PCs, gaming trends, VR, and datacenter position for upside to estimates. NVDA is also well positioned for the up and coming machine learning, deep learning, and AI markets. Our NVDA F17/F18(Jan) rev/EPS at $8.1B/$2.82 and $8.5B/$3.08 respectively. We have NVDA with a Buy-$115PT, ~37.3x P/E, at the higher end of its historical valuations.

Risks. NVDA competes in a cyclical, technologically intensive industry and sells to a concentrated customer base. Its ability to meet its own or our financial expectations and achieve future growth is subject to a number of risk factors, including, but are not limited to, the following:

- Demand for NVDA’s products is variable and could differ from expectations;

- Gross margin percentage could vary significantly;

•Competition and pricing pressure from other low-cost OEMs, ODMs, and suppliers;

•NVDA relies on third party manufacturing;

•NVDA has a very high valuation, and investors are risk averse having seen significant resets in equities trading at high valuations such as AMBA and MBLY. We believe where NVDA differs is a significantly diversified revenue base, low customer concentration and conservative estimates. Also its expected catalysts are near-term, compared to longer-term growth objectives that have technology and regulatory risks.

•Unexpected changes in legal and regulatory requirements, tariffs and exchange rates, political and economic stability, staffing and management issues, and potential adverse tax consequences;

•Seasonal fluctuations associated with consumer products and the PC market;

•Potential loss of intellectual property, commercialization of competing technologies;

•Adverse effects of potential possible future patent or other litigation;

•NVDA receives a significant amount of revenue from a limited number of customers

QUALCOMM Incorporated

We have QCOM with a Buy rating and a $75 PT, based on ~16.0x our F17E EPS of $4.70.

Risks: QCOM competes in a technologically intensive and cyclical industry and we believe the risks to QCOM continue to be a slowdown in the handset market, lower royalties, competition from China handset suppliers. Also increasing handset market share between Apple and Samsung has implied less merchant processor opportunity for QCOM.

Companies Mentioned (prices as of 1/18)

Advanced Micro Devices, Inc. (AMD- Buy $9.88)
NVIDIA Corporation (NVDA- Buy $102.95)
QUALCOMM Incorporated (QCOM- Buy $65.13)
Intel Corporation (INTC- Buy $36.76)
ON Semiconductor Corporation (ON- Buy $13.26)

IMPORTANT DISCLOSURES

The disclosures for the subject companies of this report as well as the disclosures for Mizuho Securities USA Inc. entire coverage universe can be found at https://msusa.bluematrix.com/sellside/Disclosures.action or obtained by contacting EQSupervisoryAnalystUS@us.mizuho-sc.com or via postal mail at Equity Research Editorial Department, Mizuho Securities USA Inc., 320 Park Avenue, 12th Floor, New York NY, 10022.

Investment Risks and Valuation Methods can be located in the following section of this research report - Price Target Calculation and Key Risks.

Ownership Disclosures and Material Conflicts of Interest or Position as Officer or Director

None

Receipt of Compensation

Mizuho Securities USA Inc. and or its affiliates makes a market in the following securities: NVIDIA Corporation, Advanced Micro Devices, Inc., Intel Corporation, QUALCOMM Incorporated and ON Semiconductor Corporation
The compensation of the research analyst writing this report, in whole or part, is based on MSUSA's annual revenue and earnings and is not directly related to any specific investment banking compensation. MSUSA's internal policies and procedures prohibit research analysts from receiving compensation from companies covered in the research reports.

Regulation Analyst Certification (AC)

I, Vijay Rakesh, hereby certify that the views expressed in this research report accurately reflect my personal views about any and all the subject companies. No part of my compensation was, is or will be, directly or indirectly, related to the specific recommendations or views expressed in this research report.

Rating Definitions

Mizuho Securities USA investment ratings are based on the following definitions. Anticipated share price change is based on a 6- to 12-month time frame. Return expectation excludes dividends.

Buy:	Stocks for which the anticipated share price appreciation exceeds 10%.
Neutral:	Stocks for which the anticipated share price appreciation is within 10% of the share price.
Underperform:	Stocks for which the anticipated share price falls by 10% or more.
RS:	Rating Suspended - rating and price objective temporarily suspended.
NR:	No Rating - not covered, and therefore not assigned a rating.

Rating Distribution

(As of 1/18)	**% of coverage**	**IB service past 12 mo**
Buy (Buy)	45.55%	42.19%
Hold (Neutral)	51.25%	36.81%
Sell (Underperform)	3.20%	44.44%

For disclosure purposes only (NYSE and FINRA ratings distribution requirements), our Buy, Neutral and Underperform ratings are displayed as Buy, Hold and Sell, respectively.

For additional information: Please log on to http://www.mizuhosecurities.com/us or write to Mizuho Securities USA Inc. 320 Park Ave, 12th FL, New York, NY 10020.

Disclaimers

This report has been prepared by Mizuho Securities USA Inc. (“MSUSA”), a subsidiary of Mizuho Americas LLC, solely for the purpose of supplying information to the clients of MSUSA and/or its affiliates to whom it is distributed. This report is not, and should not be construed as, a solicitation or offer to buy or sell any securities or related financial products.

This report has been prepared by MSUSA solely from publicly available information. The information contained herein is believed to be reliable but has not been independently verified. MSUSA makes no guarantee, representation or warranty, and MSUSA, MHSC and/or their affiliates, directors, employees or agents accept no responsibility or liability whatsoever as to the accuracy, completeness or appropriateness of such information or for any loss or damage arising from the use or further communication of this report or any part of it. Information contained herein may not be current due to, among other things, changes in the financial markets or economic environment. Opinions reflected in this report are subject to change without notice.

This report does not constitute, and should not be used as a substitute for, tax, legal or investment advice. The report has been prepared without regard to the individual financial circumstances, needs or objectives of persons who receive it. The securities and investments related to the securities discussed in this report may not be suitable for all investors, and the report is intended for distribution to Institutional Investors. Readers should independently evaluate particular investments and strategies, and seek the advice of a financial adviser before making any investment or entering into any transaction in relation to the securities mentioned in this report.

MSUSA has no legal responsibility to any investor who directly or indirectly receives this material. Investment decisions are to be made by and remain as the sole responsibility of the investor. Investment involves risks. The price of securities may go down as well as up, and under certain circumstances investors may sustain total loss of investment. Past performance should not be taken as an indication or guarantee of future performance. Unless otherwise attributed, forecasts of future performance represent analysts' estimates based on factors they consider relevant. Actual performance may vary. Consequently, no express or implied warranty can be made regarding future performance.

Any references in this report to Mizuho Financial Group, Inc. ("MHFG"), MHSC and/or its affiliates are based only on publicly available information. The authors of this report are prohibited from using or even obtaining any insider information. As a direct subsidiary of Mizuho Americas LLC and indirect subsidiary of MHFG, MSUSA does not, as a matter of corporate policy, cover MHFG or MHSC for investment recommendation purposes.

MSUSA or other companies affiliated with MHFG, Mizuho Americas LLC or MHSC, together with their respective directors and officers, may have or take positions in the securities mentioned in this report, or derivatives of such securities or other securities issued by companies mentioned in this report, for their own account or the accounts of others, or enter into transactions contrary to any recommendations contained herein, and also may perform or seek to perform broking and other investment or securities related services for the companies mentioned in this report as well as other parties generally.

Restrictions on Distribution

This report is not directed to, or intended for distribution to or use by, any person who is a citizen or resident of, or entity located in, any locality, territory, state, country or other jurisdiction where such distribution, publication, availability or use would be contrary to or restricted by law or regulation. Persons or entities into whose possession this report comes should inform themselves about and observe such restrictions.

United States: Mizuho Securities USA Inc., a subsidiary of Mizuho Americas LLC, 320 Park Avenue, 12th Floor, New York, NY 10022, USA, contact number +1-212-209-9300, distributes or approves the distribution of this report in the United States and takes responsibility for it. Any transaction by a US investor resulting from the information contained in this report may be effected only through MSUSA. Interested US investors should contact their MSUSA sales representative.

United Kingdom/European Economic Area: This report is distributed or has been approved for issue and distribution in the UK by Mizuho International plc ("MHI"), Mizuho House, 30 Old Bailey, London EC4M 7AU, a member of the MHSC Group. MHI is authorized and regulated by the Financial Services Authority and is a member of the London Stock Exchange. For the avoidance of doubt this report is not intended for retail clients. This report may be distributed in other member states of the European Union.

Japan: This report is distributed in Japan by Mizuho Securities Co., Ltd. ("MHSC"), Otemachi First Square Otemachi 1-chome, Chiyoda-ku, Tokyo 100-0004, Japan. Registered Financial Instruments Firm, No. 94 (Kinsho), issued by the Director, Kanto Local Finance Bureau. MHSC is a member of the Japan Securities Dealers Association, the Japan Securities Investment Advisers Association and the Financial Futures Association of Japan, and the Type II Financial Instruments Firms Association.

Singapore: This report is distributed or has been approved for distribution in Singapore by Mizuho Securities (Singapore) Pte. Ltd. ("MHSS"), a member of the MHSC Group, which is regulated by the Monetary Authority of Singapore. Any research report produced by a foreign Mizuho entity, analyst or affiliate is distributed in Singapore only to "Institutional Investors," "Expert Investors" or "Accredited Investors" as defined in the Securities and Futures Act, Chap. 289 of Singapore. Any matters arising from, or in connection with this material, should be brought to the attention of MHSS.

Hong Kong: This report is being distributed in Hong Kong by Mizuho Securities Asia Limited ("MHSA"), a member of the MHSC Group, which is licensed and regulated by the Hong Kong Securities and Futures Commission.

Australia: This report is being distributed in Australia by MHSA, which is exempted from the requirement to hold an Australian financial services license under the Corporation Act 2001 ("CA") in respect of the financial services provided to the recipients. MHSA is regulated by the Securities and Futures Commission under the laws of Hong Kong, which differ from Australian laws. Distribution of this report is intended only for recipients who are "wholesale clients" within the meaning of the CA.

If you do not wish to receive our reports in the future, please contact your sales person and request to be removed from receiving this distribution.