embedded VISION SUMMIT 2018

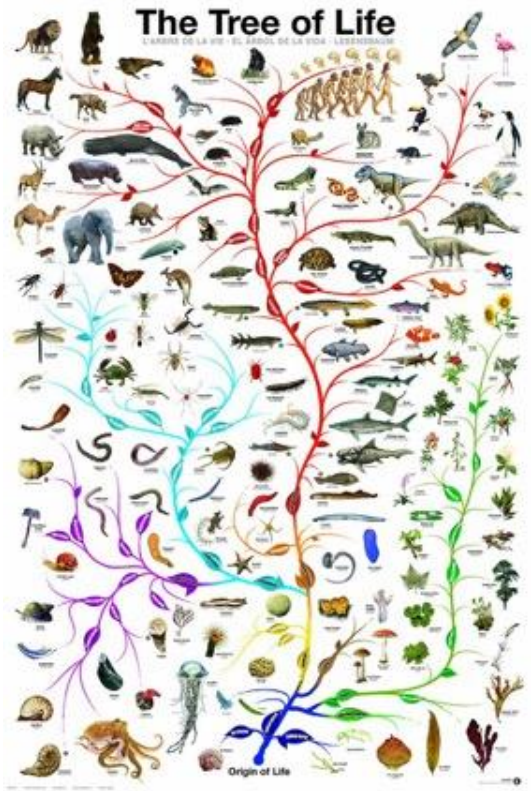
Energy-efficient Processors Enable the Era of Intelligent Devices

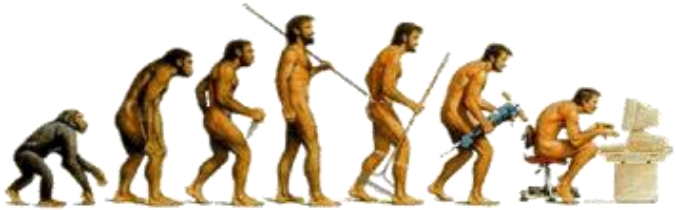
NOVUMIND
Making Things Think

Dr. Ren Wu

May 22, 2018

Evolutionary



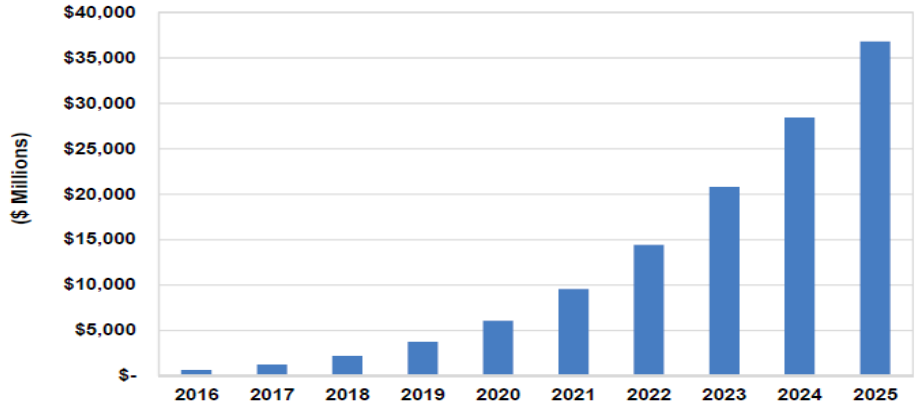


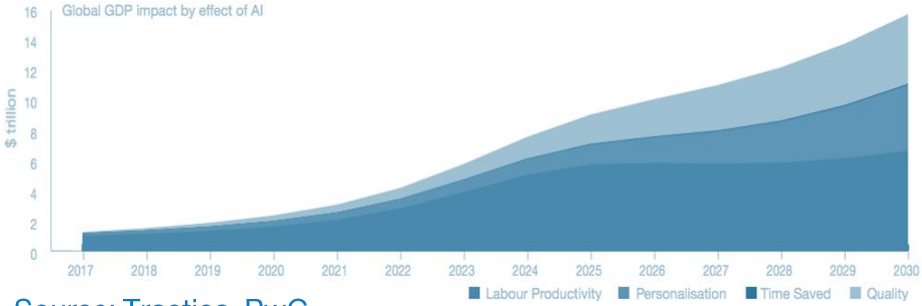
Intelligence is based on how efficient spices became at doing the things they need.

-- Charles Darwin

AI Technology Trend

Artificial Intelligence Revenue, World Markets: 2016-2025

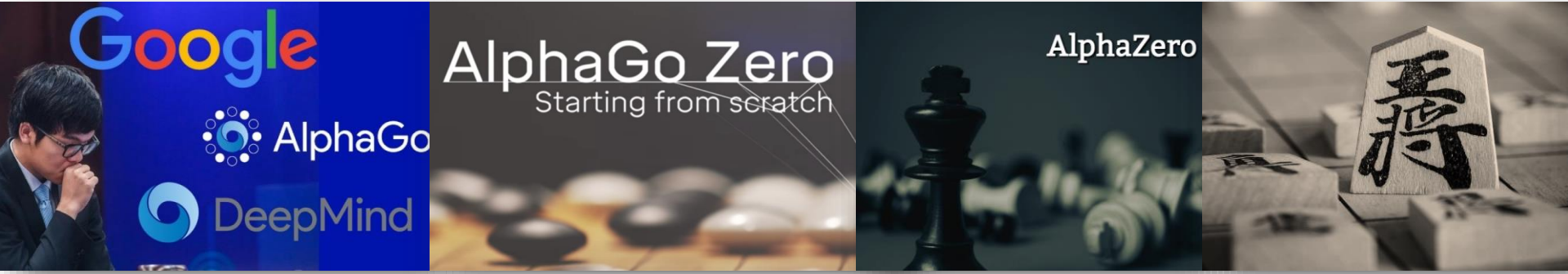




Source: Tractica, PwC

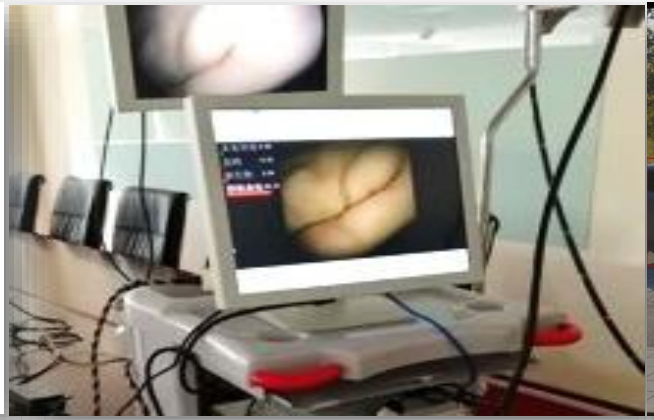
- The annual global revenue for Artificial Intelligence will grow from 643.7 million in 2016 to **$36.8 billion** by 2025, a **57-fold increase** over that time period.
- AI technology would contribute **$15.7 trillion** to global GDP in 2030.

Deep-Learning-Based AI is More Capable

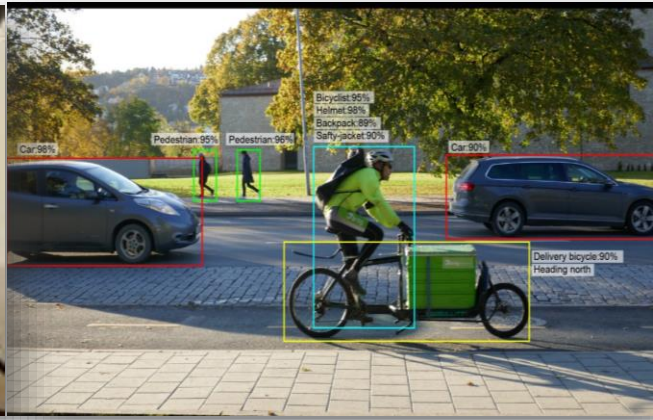


Breakthroughs in AI

Speech Recognition



Smart Health

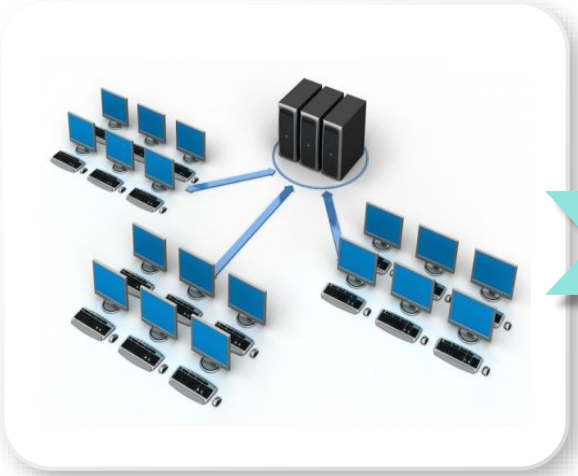


Green City

And More…

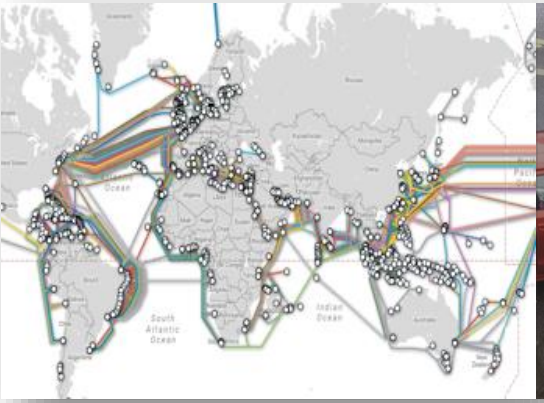
Trend and Opportunity

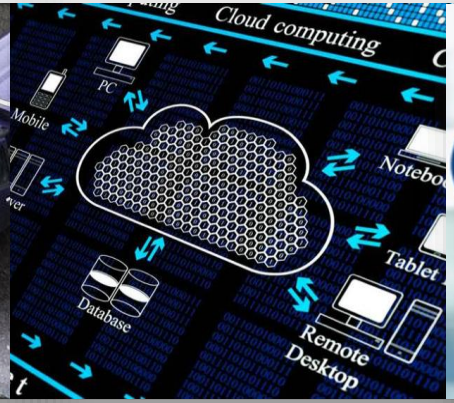
Internet → Mobile Internet → Internet of Things



Making “Things” Smarter -- Huge Opportunity!

Need for Smarter “Things”





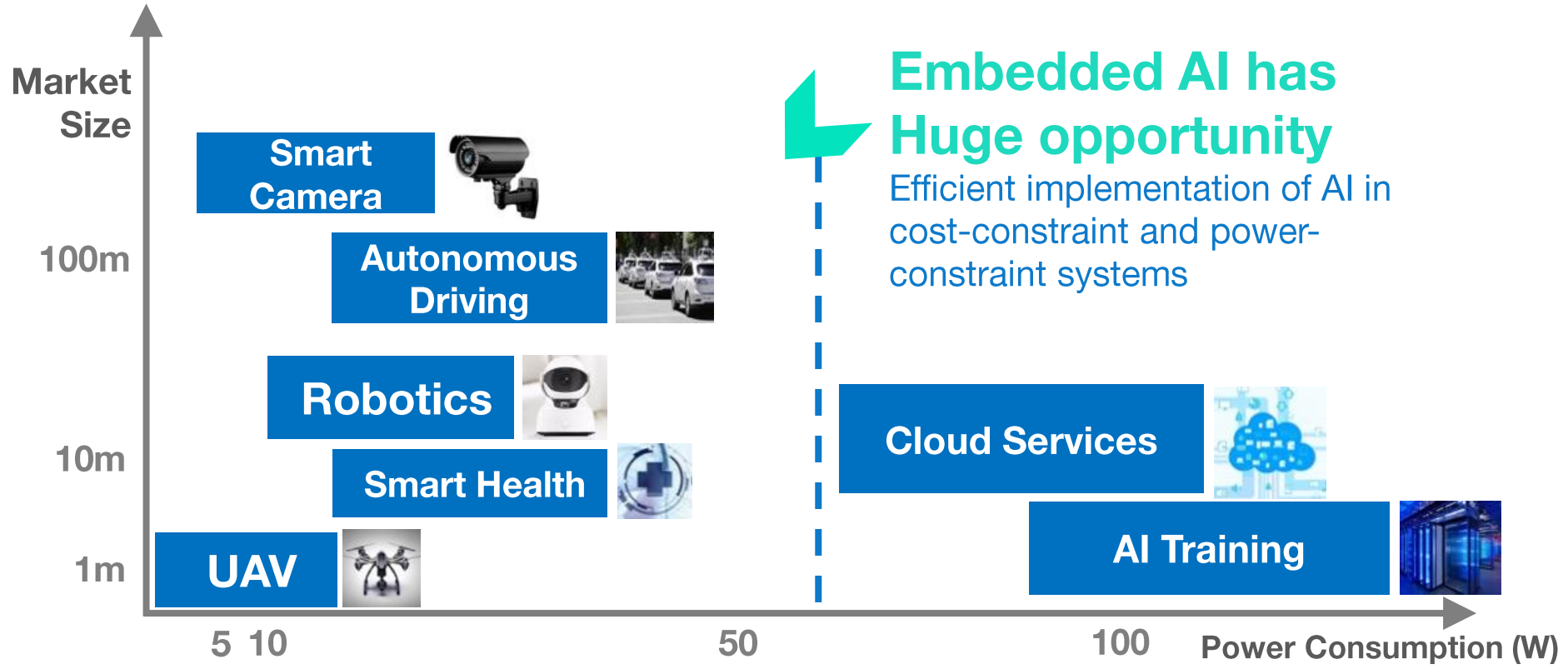
Bandwidth? Latency? Power Consumption? Privacy?

Cloud computing will not work for many applications!

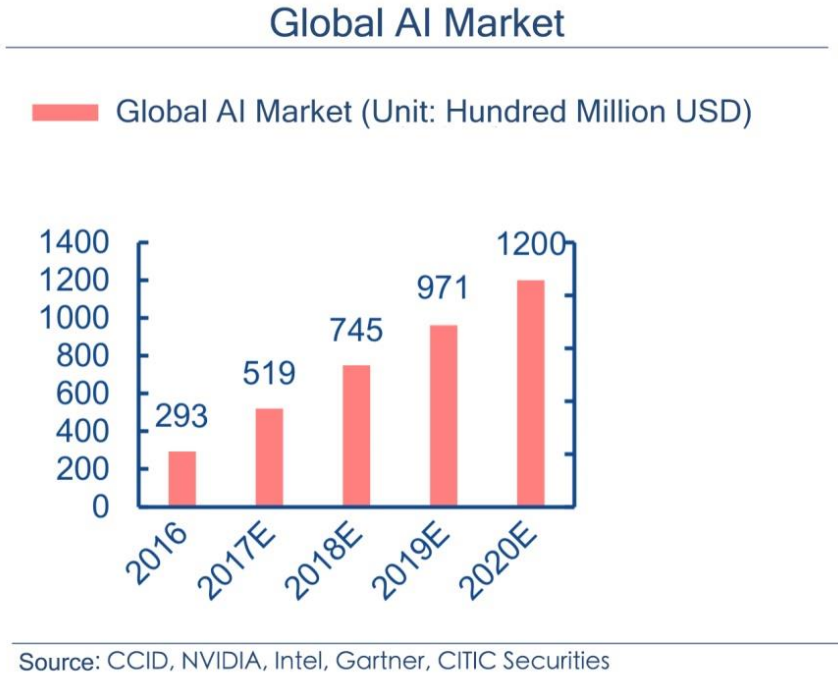
From Capable to Accessible

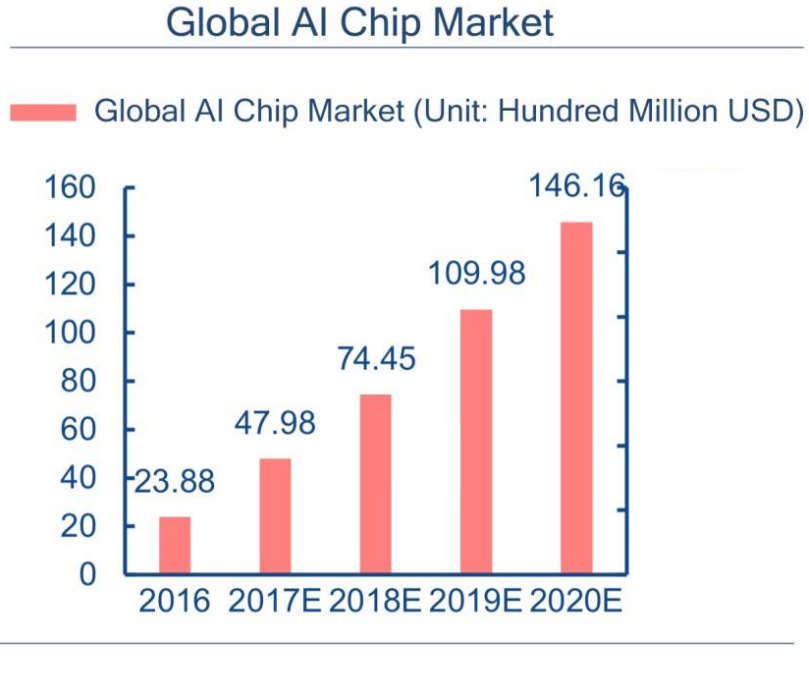
Cloud +  Device Computing !

Embedded AI: Huge Opportunity



AI Chips Market

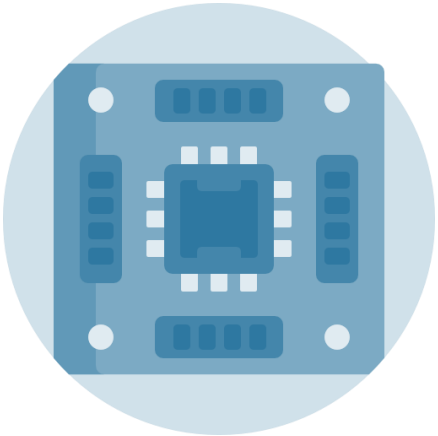




Source: CCID, NVIDIA, Intel, Gartner, CITIC Securities

Market size of AI Chips will reach **15 billion** US dollars, taking **12%** of the global AI market in 2020

AI Chips: the Key Battlefield



NVIDIA
Google Cloud TPU
Movidius
intel
nervana SYSTEMS
SAMSUNG
HUAWEI
QUALCOMM
cerebras
Graphcore
groq
Cambricon
Horizon Robotics 地平线机器人技术
BITMAIN
DEEPHI 深鉴科技

Shortcoming of Current AI Chips

Name	Power (W)	Performance (TOPS)	Price ($)
Movidius Myriad2	1	0.2	50
CEVA XM6*	2	1	N/A
QCOM S835	2.5	0.19	100
HUAWEI Kirin970	2	1.9	?
Cadence C5*	2	2	N/A
Nvidia TX2	10	1	399
Nvidia Xavier (2018)	30	30	1999?
Google TPU1	40	23	N/A
Google TPU2	250	180	N/A
Nvidia P100 GPU	250	10.6	5000
Nvidia V100 GPU	250	120	6000

Low Power
Low Performance

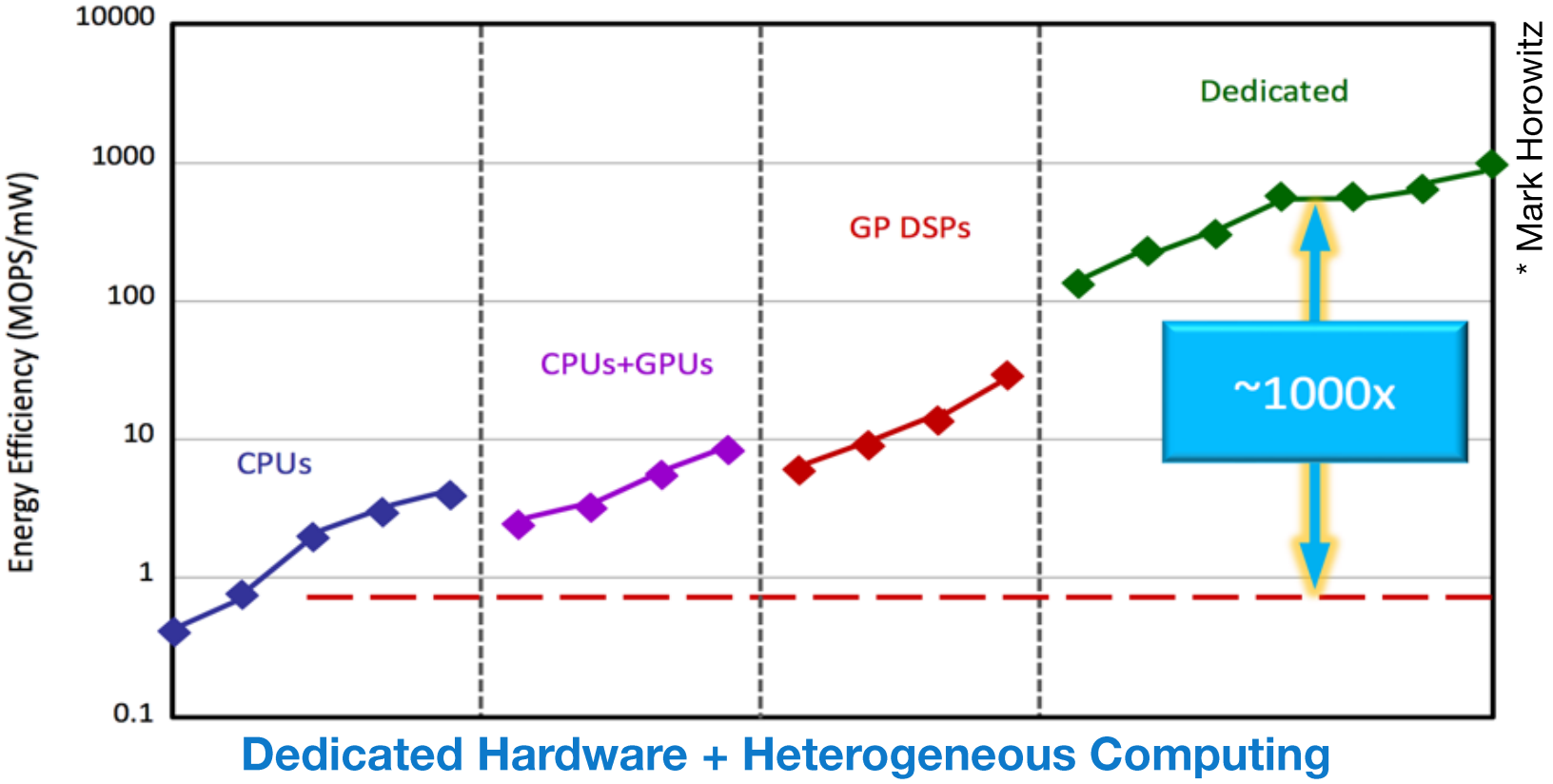
High Power
High Performance

* IP vendor

The Key to Embedded Intelligence

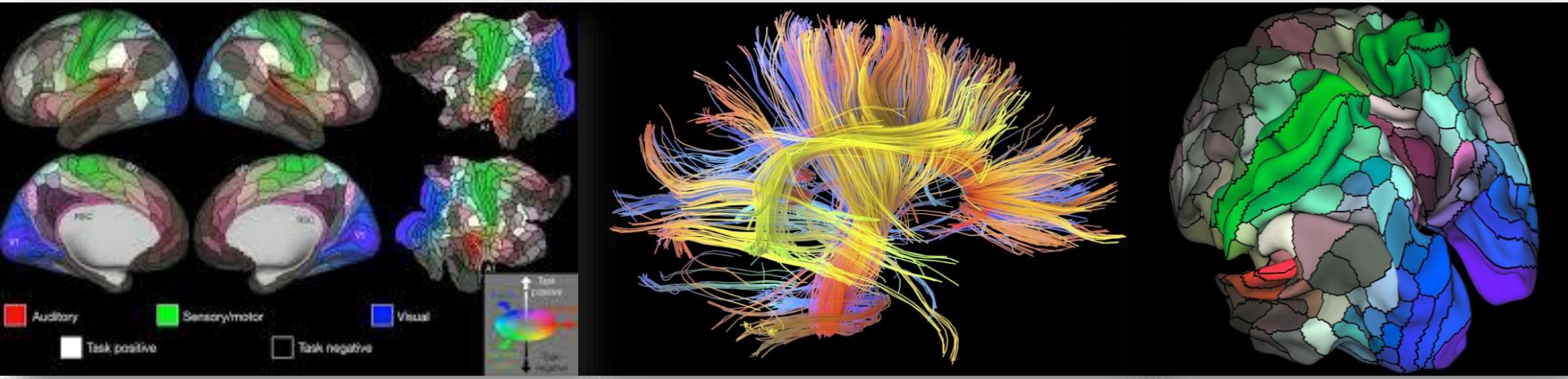
High Performance
Low Cost
Low Power Consumption

Proposed Solutions



Dedicated Hardware + Heterogeneous Computing

Heterogeneous Computing



John Hennessy – 2017 Turning Prize Winner

Domain Specific Architecture (DSAs) –

Achieve higher efficiency by tailoring the architecture to characteristics of the domain.

"The easy ride of software is over."

AI Chip – To Enable Embedded Intelligence

AI Chips	Power (W)	Performance (TOPS)	Perf/Watt (TOPS/W)
NovuTensor V2	**5**	**20**	**4**
Movidius Myriad2	1	0.2	0.2
CEVA XM6*	2	1	0.5
QCOM S835	2.5	0.19	0.08
HUAWEI Kirin970	2	1.9	0.95
Cadence C5*	2	2	1
Nvidia TX2	10	1	0.1
Nvidia Xavier (2018)	30	30	1
Google TPU1	40	23	0.58
Google TPU2	250	180	0.72
Nvidia P100 GPU	250	10.6	0.04
Nvidia V100 GPU	250	120	0.48

400%
better perf/watt

Low Power
Low Performance

High Power
High Performance

* IP vendor

Embedded Intelligence in Healthcare

- Real-time instantaneous feedback with accurate result will change medical diagnostic practice.
- Improve human's lives!

Embedded Intelligence in Green City

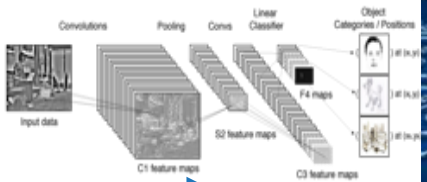
- **Greener** transportation, **greener** life.
- Smarter city without compromising privacy!

NovuMind – Full Stack AI Company

NovuStar

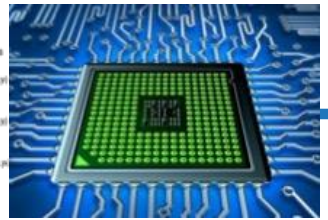
Supercomputer

Training



Neural Network Models

NovuTensor



AI Chip

NovuBrain

AI Module

Full Stack Solutions

NovuMind

Better Training Capacity

Intelligent & Efficient

User friendly

Turnkey Solutions

Big Data

Industries

Companies

Conclusion

Era of intelligent devices

I²oT

Powered by

NovuMind

Making people's life better

Allied Vision

Qualcomm intel

arm XILINX

NOVUMIND Making Things Think NXP

cadence AIMOTIVE

SYNOPSYS®

Meet us at booth #704