

Al at the Edge: Bringing Intelligence to Small Devices at the Network Edge

Kevin A. Shaw, Ph.D.
Chief Technology Officer / Cofounder



Technology Revolution: The Drive for Ubiquity



First Wave: Resources shared by many.



Second Wave: One-to-one.



Third Wave: Many to Many.

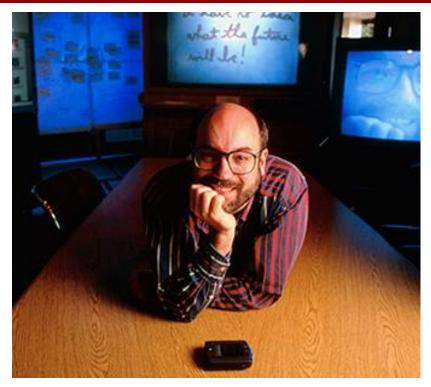
Technology



Ubiquity: Computing Everywhere

"The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it."

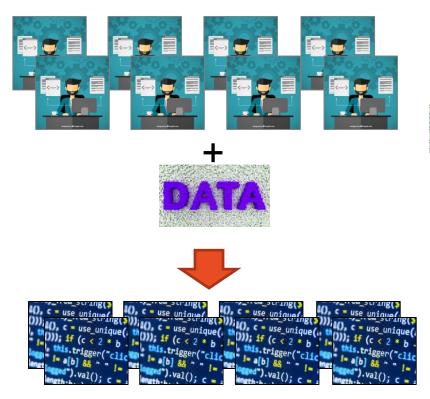
-- Mark Weiser, 1991, Scientific American



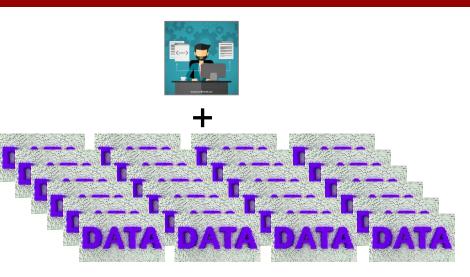
Mark Weiser, CTO at Xerox PARC, 1990

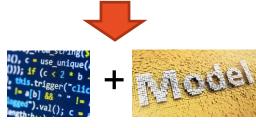


Algorithmic Revolution



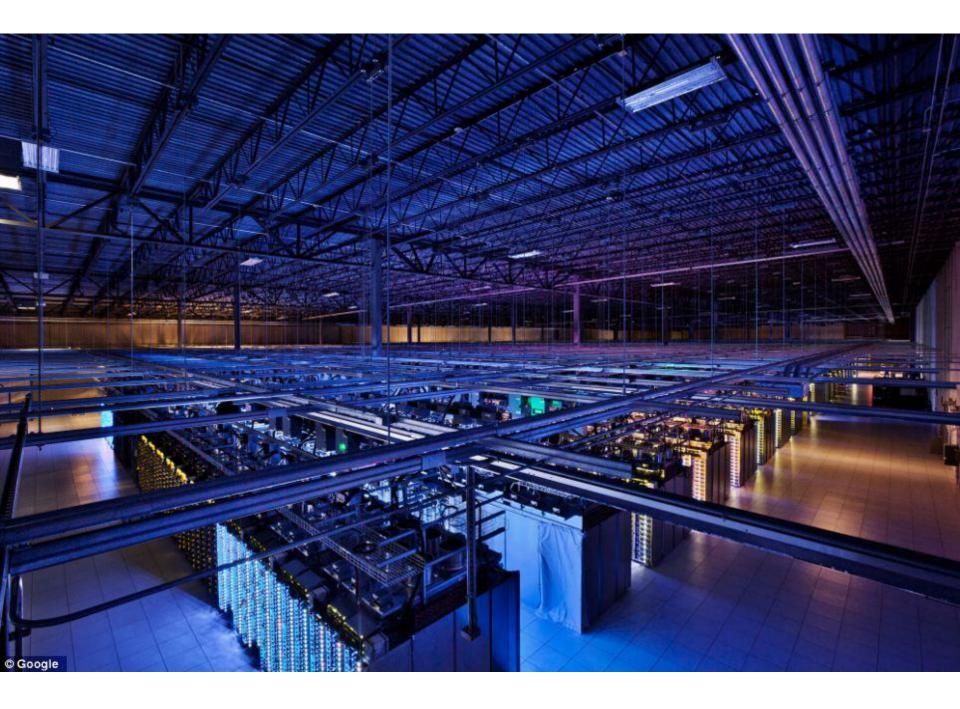
Traditional Programming Methodology





Machine Learning Methodology





Devices at the Edge











Devices at the Edge



Battery powered



Low maintenance





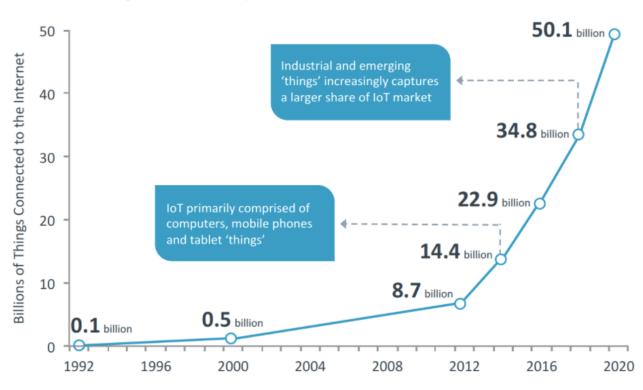




Devices at the Edge

Projecting the 'Things' Behind the Internet of Things

From 2014-2020, IoT grows at an annual compound rate of 23.1% CAGR



CompTIA.

Sources: Group SJR | Cisco | CompTIA



Example: Kitchen Faucet



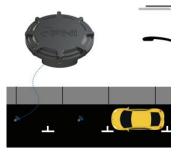




The World of IoT











Door Knob

Gas Meter

Parking Sensor

Thermostat

Doorbell



LED Lighting



Smoke Detectors



Waste Bins



Fitness Bands



Appliances

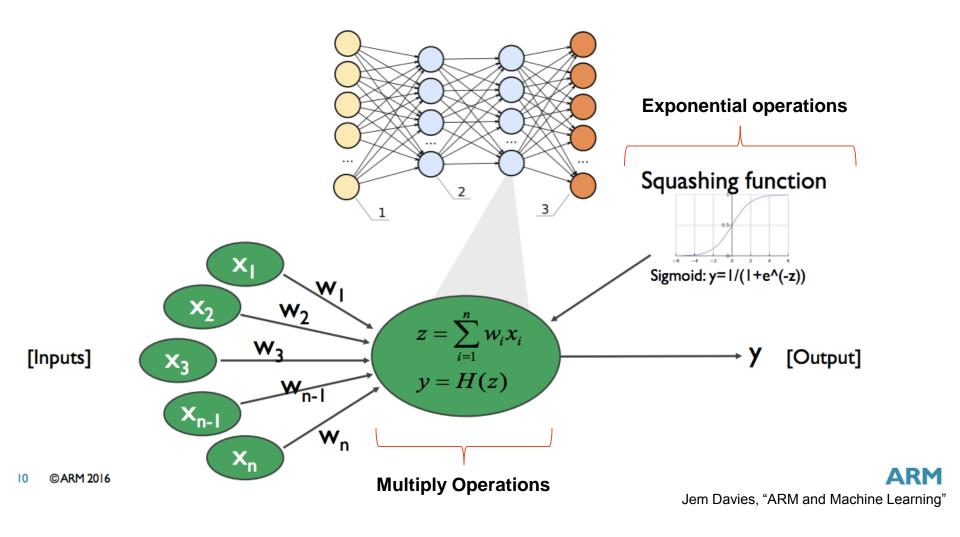


Compute and Sensing Requirements

- High Compute
 - Vision/Camera images: Face detection; object detection
 - Acoustic Processing: voice detection (Alexa, Siri)
 - Requirements: GPU or dedicated Vision Processor
- Low Compute
 - Motion: Steps and Activity
 - Environmental: humidity, temperature, gasses, particulates
 - Pressure: elevation
 - Location/Proximity: Beacons
 - Requirements: ARM Cortex M-class or similar



Basic Neural Network



Challenges: Hardware

- Memory Limitations
 - Memory is 64K to 512K in most MCU cores
- Power Limitations
 - Can't send raw data for processing
 - Network transport is very costly:
 - Compute often; transport rarely
- Math Operations
 - FPU (Floating Point Unit)
- Processor design
 - 32bit
 - Low power designs; peripherals can dominate power profile

Challenges: Software

- Software frameworks for DNN
 - Many available, but not for embedded
- DNN Training
 - Typically running on large machines with GPU support
 - Generates a model file for the inferencing engine
 - TensorFlow and others are good choices
- DNN Inferencing
 - Either full data needs to be uploaded to the cloud
 - Or, needs to run on the local processor
 - Many frameworks available, but not for embedded
 - Model files are often large, need to extract minimal data

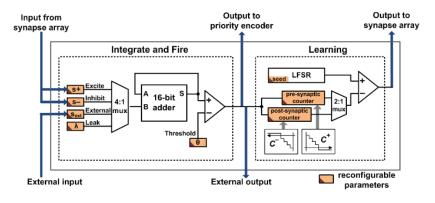
Processors: MCU/CPU based

- 32b core (MCU) + FPU
 - ARM M-class
 - Ambiq Apollo 2: Ultra low power ARM M4F
- DSPs or Custom variants:
 - Certain operations are hardened into gates for acceleration
 - Bosch Sensortec DSP "Fuser Core": BHI160
 - Greenwaves Tech: GAP8, RISC-V with HW NN
 - Custom instructions: Tensilica, ARC



Processors: Neuromorphic

- Building the nodes (neurons) directly in silicon
 - Analog approach vs digital approach
 - Parallel operation
- General Vision + Intel
- Qualcomm Zeroth NPU (Neural Processing Unit)
- IBM TrueNorth processor (1,000,000+ neurons)
- Toshiba TDNN (Time Domain Neural Network)





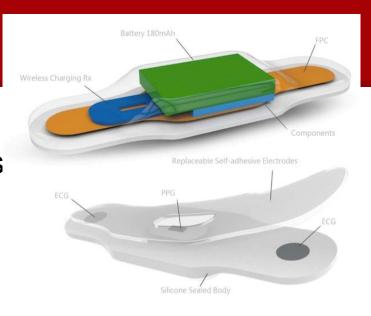
Conclusion

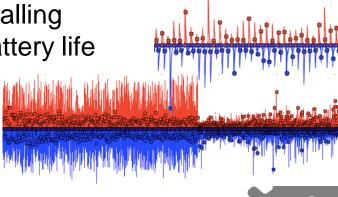
- New algorithmic schemas are pushing computation back to the cloud
- Drive for distributed computing is too great
 - mobile cores are too cheap and too capable
- ANNs are well suited for edge processing
- New dedicated hardware is coming to accelerate



Algorithmic Intuition Inc

- Building Machine-Learning algorithms for embedded sensors
 - Focus on activity recognition
- Ai² Active Living Monitor[©]
 - Products to track wellness in Aging
 - Software & Hardware Platform
 - Detect Vitals, Activities of Daily Living, Falling
 - On-body sensor computation for long-battery life
 - Hardware Platform reference design
 - Sensors, MCU, BLE and PMIC







Al at the Edge: Bringing Intelligence to Small Devices at the Network Edge

Kevin A. Shaw, Ph.D.
Chief Technology Officer / Cofounder

