Crossbar ReRAM

Sylvain Dubois

Vice President Business Development & Marketing

sylvain.dubois@crossbar-inc.com

April 9th, 2019

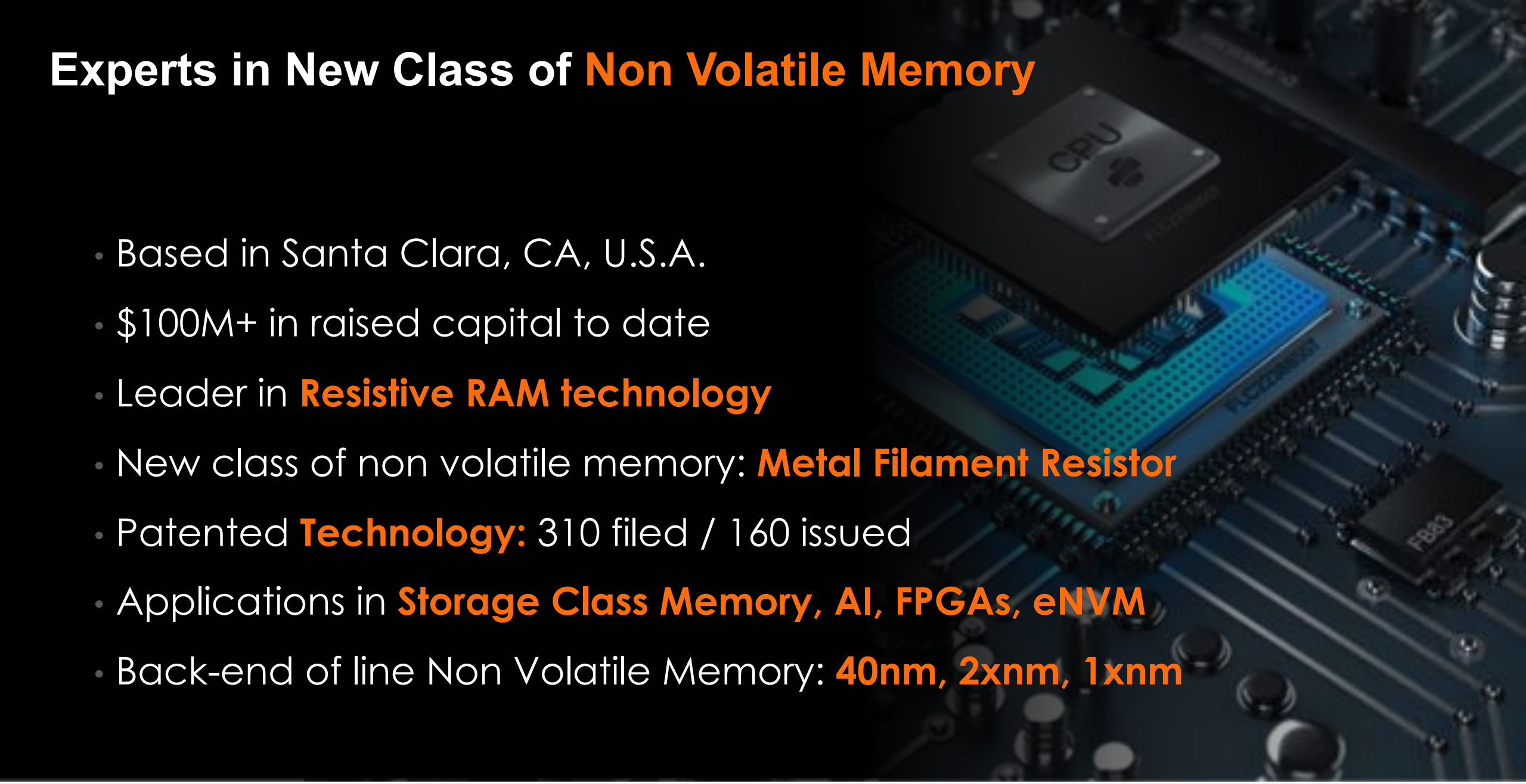
CROSSBAR

IP-SoC 2019

IP-SOC CONFERENCE & EXHIBITION

Santa Clara, California | April 9, 2019

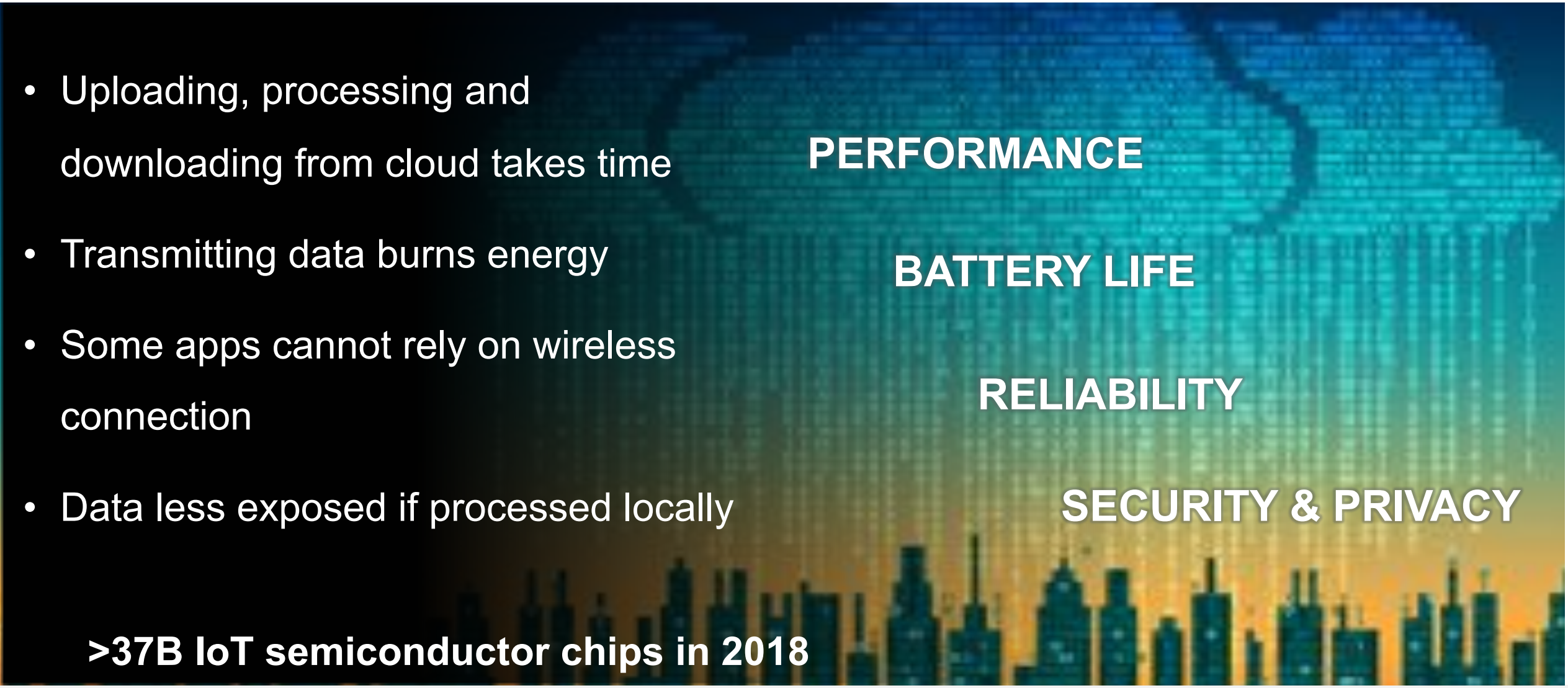
Experts in New Class of Non Volatile Memory



- Based in Santa Clara, CA, U.S.A.
- $100M+ in raised capital to date
- Leader in **Resistive RAM technology**
- New class of non volatile memory: **Metal Filament Resistor**
- Patented **Technology:** 310 filed / 160 issued
- Applications in **Storage Class Memory, AI, FPGAs, eNVM**
- Back-end of line Non Volatile Memory: **40nm, 2xnm, 1xnm**

CROSSBAR

AI is moving to the EDGE



- Uploading, processing and downloading from cloud takes time
- Transmitting data burns energy
- Some apps cannot rely on wireless connection
- Data less exposed if processed locally

PERFORMANCE

BATTERY LIFE

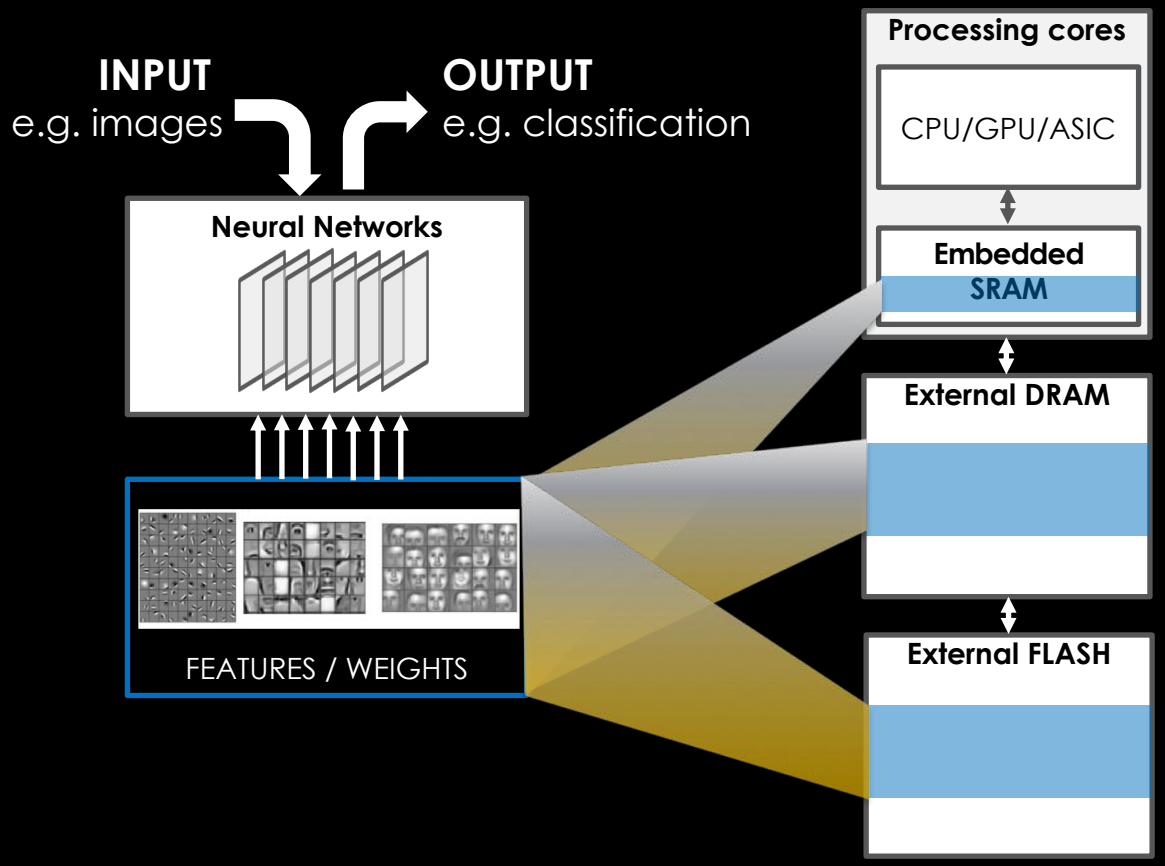
RELIABILITY

SECURITY & PRIVACY

>37B IoT semiconductor chips in 2018

CROSSBAR

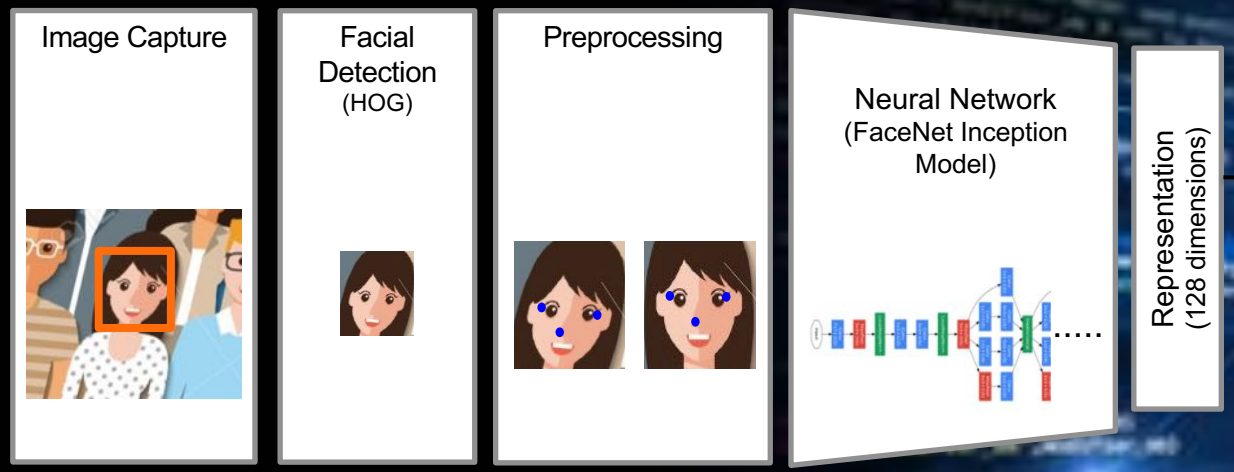
Problem: The Memory Bottleneck in AI

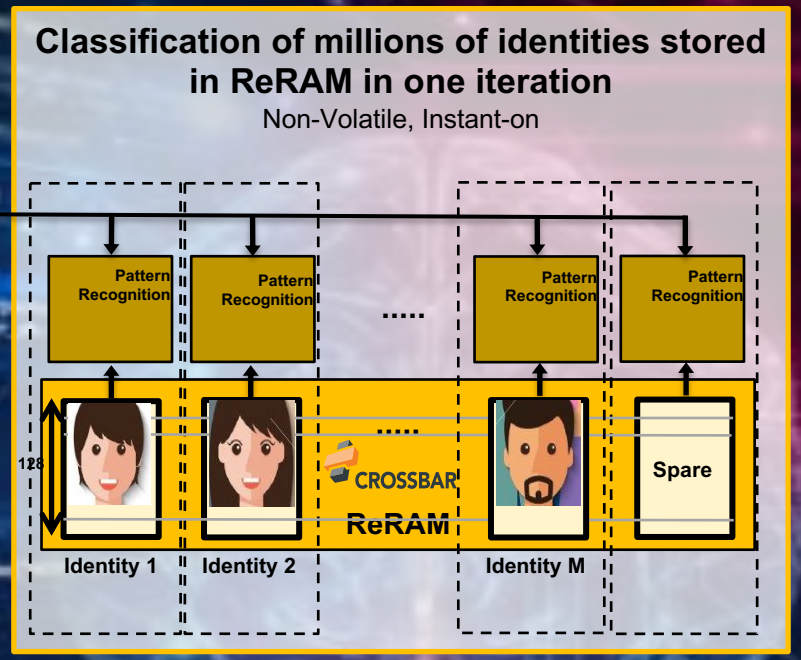


- AI performance is limited by memory bus bandwidth and latencies
- Energy wasted moving data from FLASH to DRAM to SRAM to processing cores
- Need to refresh/reload model in SRAM/DRAM at every wake-up

How to solve the problem? **Bring data and algorithms on same chip**

Solution: Hardware Acceleration with Crossbar ReRAM





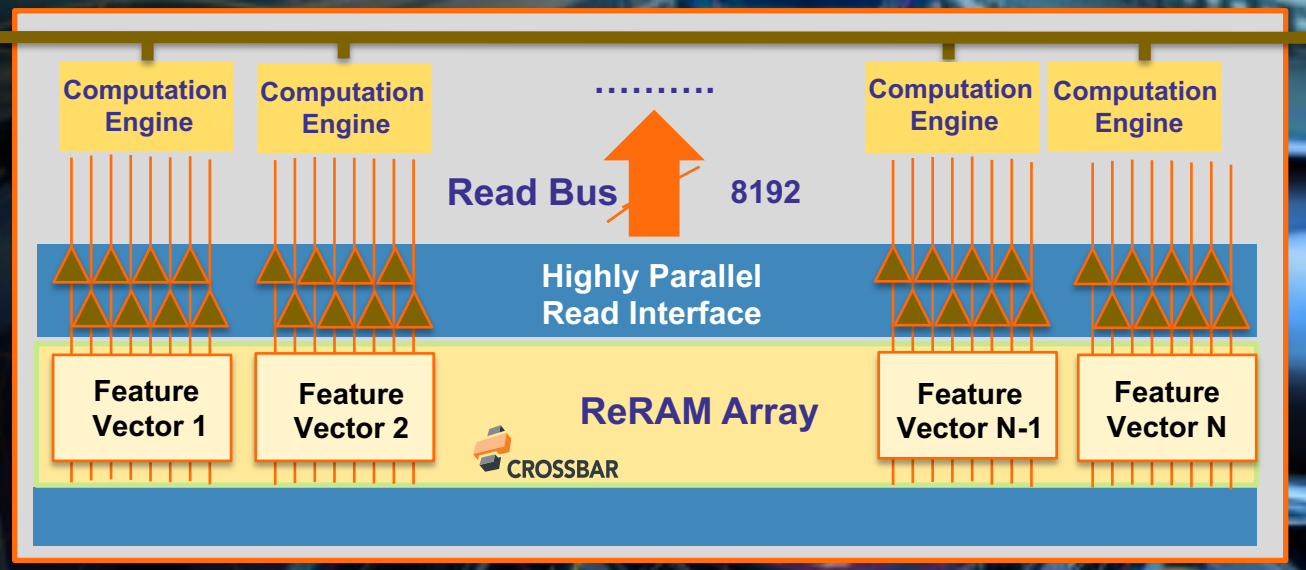
Simultaneous Processing with Deterministic Performance

- Parallel comparison against all identities
- If no match, new identity created (learning)
- Classification performed in one cycle independent of number of identities

ReRAM for Massive Search Hardware Acceleration

- Very wide Non-Volatile memory array
  - 50GB/s Read throughput (@8K wide)
- Applications
  - Massive Search
  - KNN
    - 1000's of Distance Calculators above HPM
    - Classifications
  - CNN, RNN, NLP Inference
    - Weights in ReRAM
    - Embedded MACs
    - Edge or Cloud
- Flexible architecture
  - Number of instances, 8-bit to binary.
  - 4 to 8 dies per chip, 25 chips per M.2, 4 M.2 per PCIe
- Spare memory enabling Learning at the Edge

Simultaneous Processing Deterministic Performance



400,000 search/sec across 1000 instances

vector length 128B

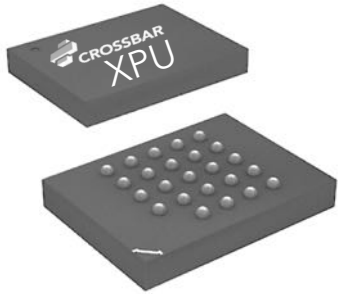
12.5 TOPS/W

Massive search acceleration scalable to PetaFLOPS

25 to 50M instances **(binary 128-deep)**
or 400,000 to 800,000 instances (8-bit 1K-deep)
PCIe card
4 units of M.2

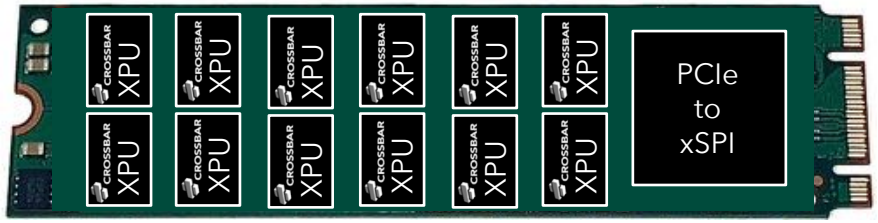
6.4 to 12.8M instances **(binary 128-deep)**
or 100K to 200K instances (8-bit 1K-deep)
M.2 cards 80mm
25 units of XPU

256,000 to 512,000 instances **(binary 128-deep)**
or 4,000 to 8,000 instances (8-bit, 1K-deep)
SPI package 6 mm X 8 mm
4 to 8 dies per XPU package



AT EDGE

51 TOPS @ 50MHz
per package



AT GATEWAY

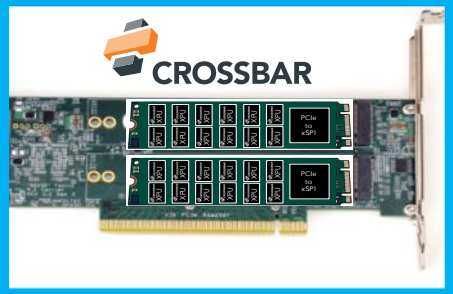
1,280 TOPS @ 50MHz
per M.2 card

IN CLOUD

5,000 TOPS @ 50MHz
per PCIe card

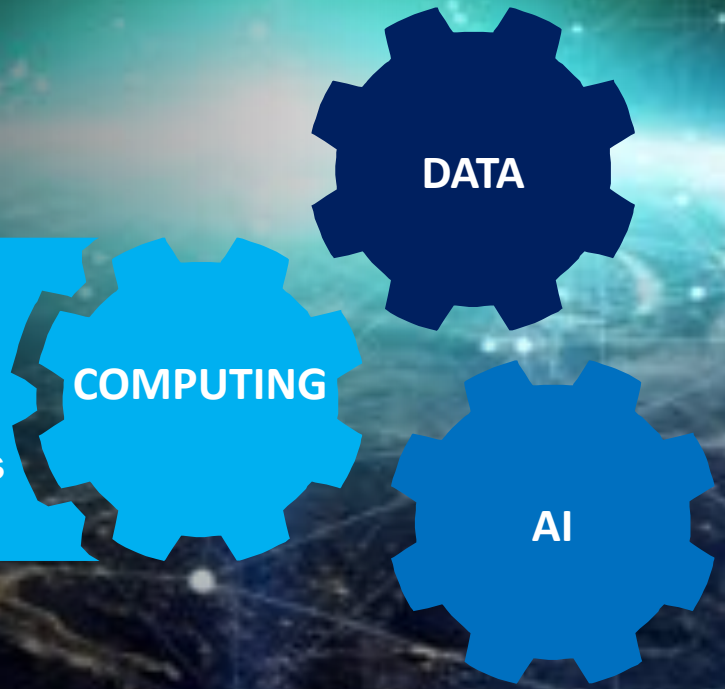
CROSSBAR: Enabling Data and Computing Acceleration

Silicon demonstration available



5 PETAFLOPS PCIe card

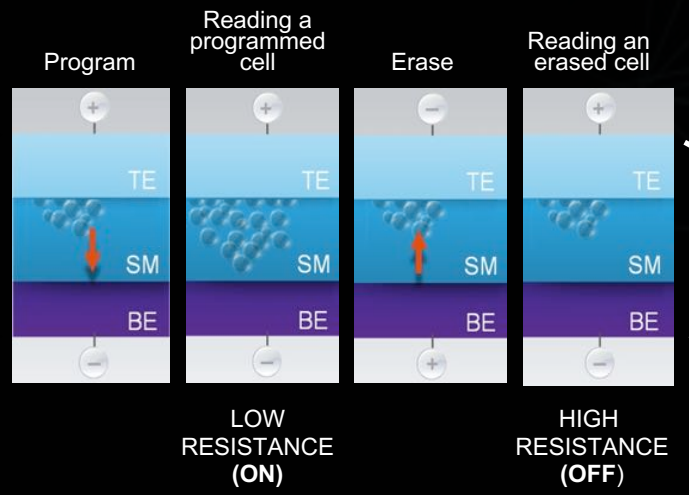
50M instances parallel search
50GB/s internal bus
400K search per second across 1000 instances
12.5 TOPS/W

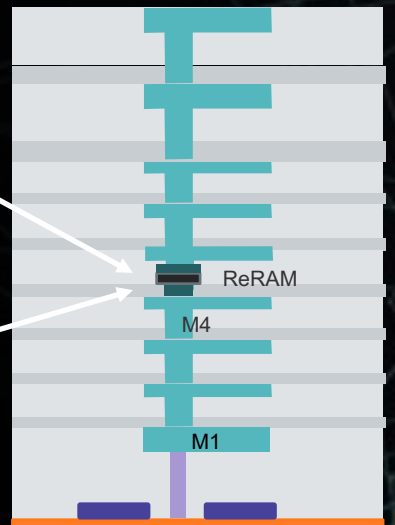


Crossbar ReRAM: Unique and superior memory technology

Sub-5nm filament

ReRAM cell





- 1000X ON/OFF ratio
- -40/+125C
- 1M+ write cycles
- 10 years retention
- 10ns read latency
- <<1uA/MHz/bit
- 40 nm / 2x nm / 1x nm

EMBEDDED

LOW COST

RELIABLE

HIGH PERFORMANCE

LOW ENERGY

SCAiLE.org

SCalable AI for Learning at the Edge

CROSSBAR

gyrfalcon technology

mtes Neural Networks

ROBOSENSING

Crossbar is a founding member of SCAiLE.org

SCAiLE.org - Press Coverage

Consortium seeks to scale artificial intelligence

February 19, 2019 // By Peter Clarke

eeNews AUTOMOTIVE

Four diverse companies – Crossbar, Gyrfalcon, mtes Neural Networks and Robosensing –have joined forces in a consortium called SCAiLE for the development and commercialization of artificial intelligence platforms.

SCAiLE stands for SCalable AI for Learning at the Edge and the consortium is working with the Japanese authorities to review opportunities for the 2020 Olympics, including video-based event detection and response capability.

"It is an open group and we have other companies expressing an interest to join," Sylvain Dubois, vice president of marketing and business development of Crossbar, told *eeNews Europe* .

Home » News » Business

By David Manners 20th February 2019

Electronics Weekly.com

AI start-ups huddle

AI start-ups Crossbar, Gyrfalcon Technology, Neural Networks Corporation and Robosensing are getting together to deliver an AI platform and standard for edge computing, gateways, cloud and data centers.

エッジでのAI活用へ日米企業連合 | ニュープロダクツ | リスク対策.com（リスク対策ドットコム） | 新建新聞社

risktaisaku.com

Emerging Memories Team Up With AI

Tom Coughlin Contributor

Enterprise & Cloud

Forbes

The use of emerging memories with artificial intelligence (AI) applications is heating up. This piece will cover a new emerging memory/AI consortium for edge-based machine learning applications as well as announcements from CEA-LETI and Stanford University of a proof of concept chip combining RRAM and processing on a single chip for AI edge applications.

Crossbar Inc., Gyrfalcon Technology Inc., mtes Neural Networks Corporation (mNN) and Robosensing Inc. announced an AI consortium that will deliver a vastly accelerated, power-saving AI platform and standard that enables new AI-rich capability for edge computing, gateways, cloud and data centers. This platform is based upon emerging memories, such as resistive RAM (RRAM) and magnetoresistive RAM (MRAM) technologies.

EETimes

HOME NEWS PERSPECTIVES DESIGNLINES VIDEOS RADIO EDUCATION

DESIGNLINES | MEMORY DESIGNLINE

AI Needs Memory to Get Cozier with Compute

By Gary Hilson, 03.12.19 2

Share Post Share on Facebook Share on Twitter in

TORONTO — Big data applications have already driven the need for architectures that put memory closer to compute resources, but artificial intelligence (AI) and machine learning are further demonstrating how hardware and hardware architectures play a critical role in successful deployments. A key question, however, is where the memory is going to reside.

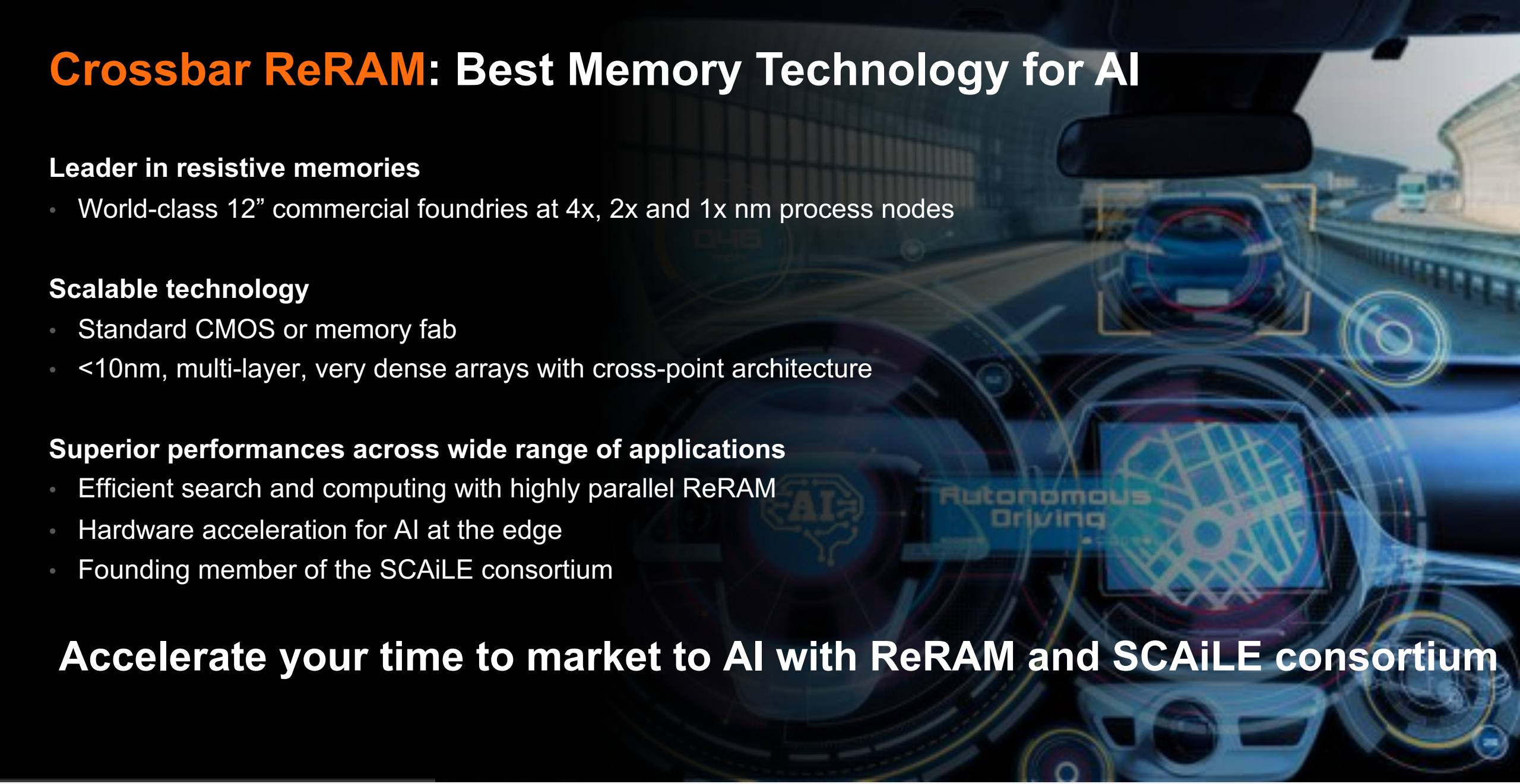
Research commissioned by Micron Technology found that 89% of respondents say it is important or critical that compute and memory are architecturally close together. The survey, carried out by Forrester Research, also found that memory and storage are the most commonly cited concerns regarding hardware constraints limiting AI and machine learning today. More than 75% of respondents recognize a need to upgrade or re-architect their memory and storage to limit architectural constraints.

AI compounds the challenges already unearthed by big data and analytics requirements because machine learning does multiple accumulation operations on a vast matrix of data over neural networks. These operations are repeated over and over as more results come in to produce an algorithm that is the best path and the best choice each time — it learns from working on the data.

https://www.electronicsweekly.com/news/business/ai-start-ups-huddle-together-2019-02/
https://www.forbes.com/sites/tomcoughlin/2019/02/20/emerging-memories-teams-up-with-ai/#2b929a5372bd
https://www.electronicsweekly.com/news/business/ai-start-ups-huddle-together-2019-02/
https://www.eetimes.com/document.asp?doc_id=1334421

CROSSBAR

Crossbar ReRAM: Best Memory Technology for AI



Leader in resistive memories

- World-class 12” commercial foundries at 4x, 2x and 1x nm process nodes

Scalable technology

- Standard CMOS or memory fab
- <10nm, multi-layer, very dense arrays with cross-point architecture

Superior performances across wide range of applications

- Efficient search and computing with highly parallel ReRAM
- Hardware acceleration for AI at the edge
- Founding member of the SCAiLE consortium

Accelerate your time to market to AI with ReRAM and SCAiLE consortium

CROSSBAR

CROSSBAR

THANK YOU

Sylvain DUBOIS

sylvain.dubois@crossbar-inc.com
Linkedin: linkedin.com/in/link2sylvaindubois
Twitter: @syl20dubois