Hyperconverged Infrastructure Virtualization

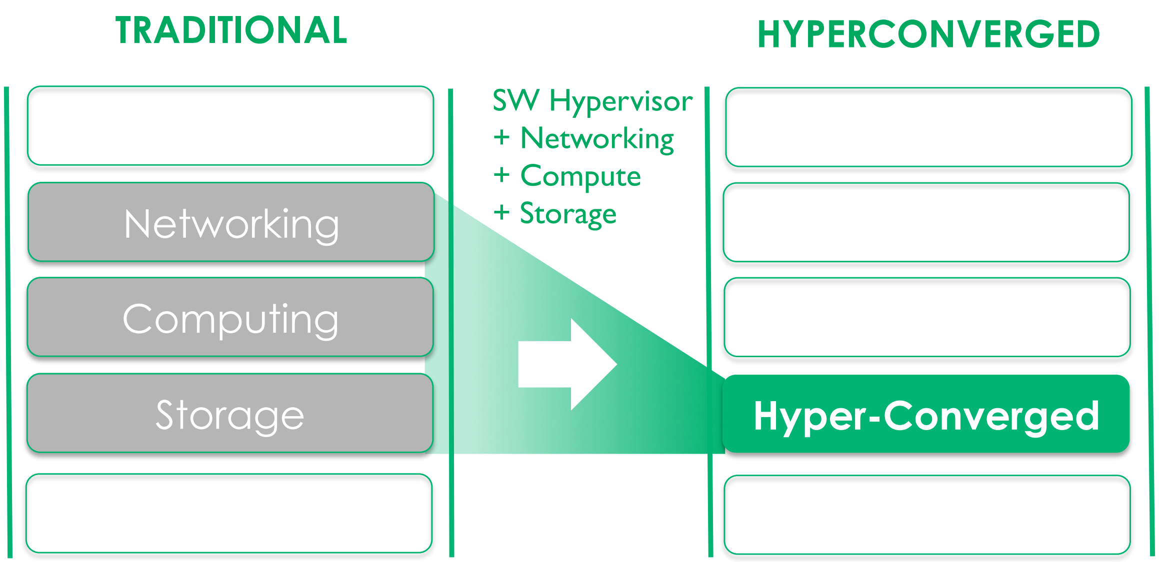
Delivering 1GIOPS Per Server

Sylvain Dubois

Vice President of Strategic Marketing and Business Development

Crossbar

Data centers going towards higher integration



*"**40% of enterprises use hyperconverged** units as a standard building block in the data center, climbing rapidly over the next two years"*

451 Research – Voice of the Enterprise

*"The market for hyperconverged integrated systems (HCIS) **will grow 79%** to reach almost $2 billion in 2016, propelling it toward **mainstream use in the next five years**"*

Gartner, Inc.

Promises and pain points of hyper-convergence infrastructure

Lower TCO
Lower Power

BUT

Compute & Storage can not scale independently

Current NVM challenges in hyper-convergence

COMPUTING

High IOPS
Low latencies

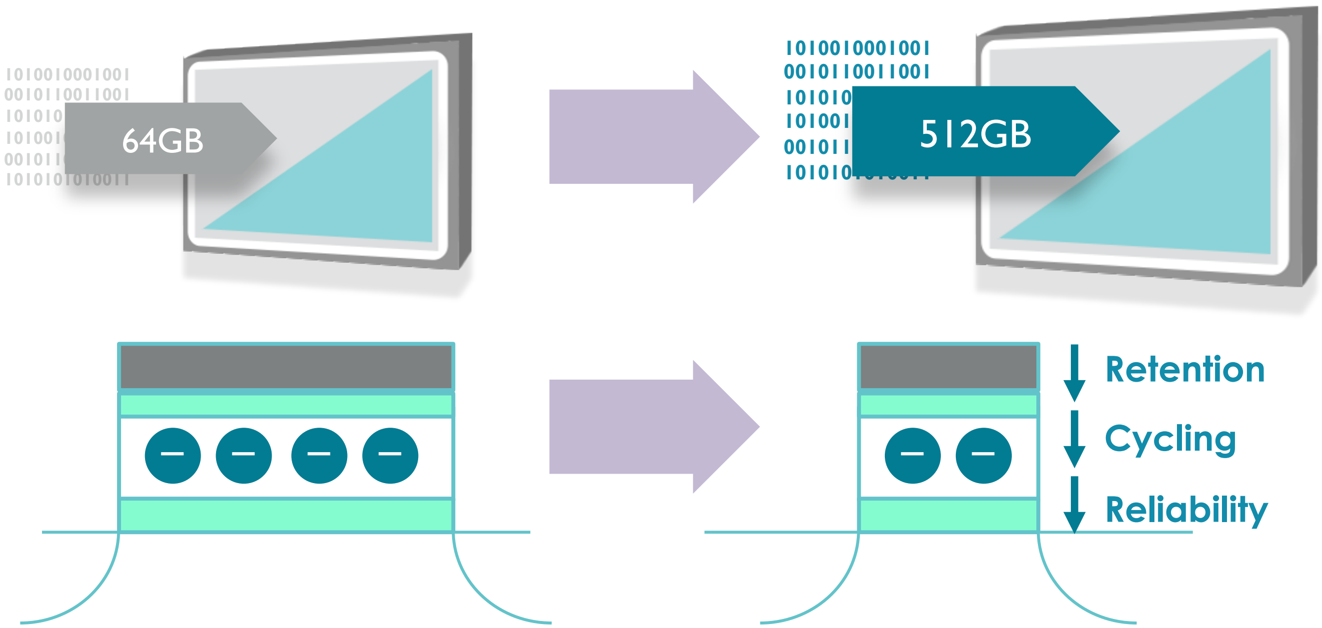
STORAGE

Low IOPS
Long latencies

Need for multi-million IOPS storage solutions

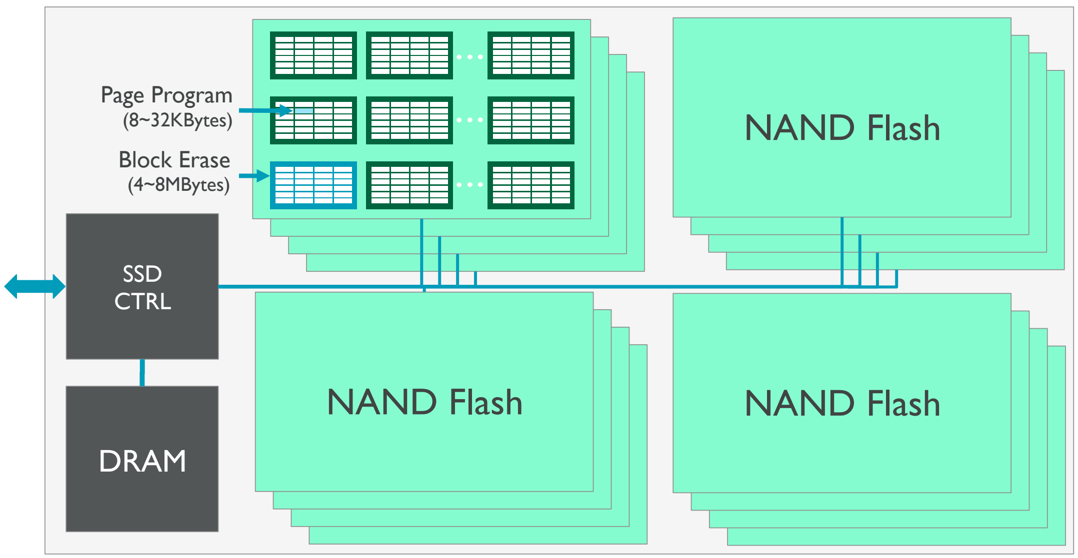
Current performance bottleneck is on storage side

NAND Flash NVM technology running out of steam



Electron-based storage facing scaling challenges with performance degradation

Flash limitations getting harder to hide in SSDs



Limitations that don't get advertised much:

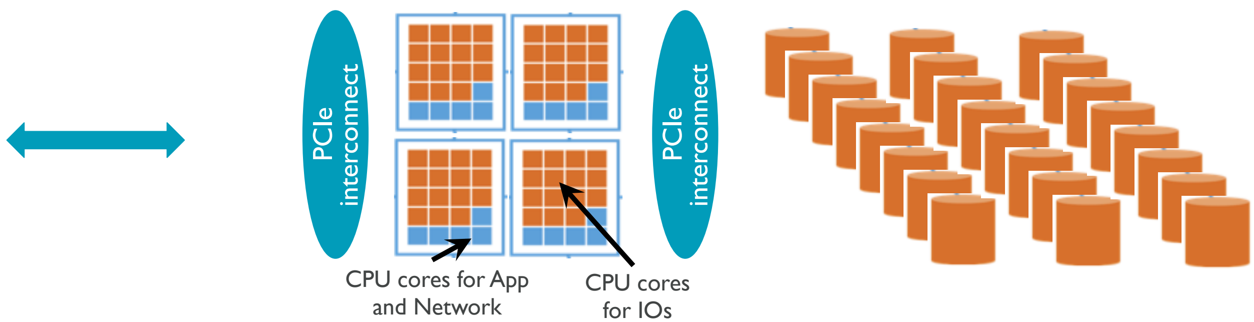
- Write Amplification
- Write Cliff
- Rd/Wr mixed IOPS
- 1k write cycles
- Garbage Collection
- Read Latency at small Queue Depth

Fresh Out-of-Box performance is not representative of performance under constant load

Poor system efficiency in existing infrastructure

CPU time per NVMe IO		100% CPU core saturation	Number of cores to manage 1 MIOPS	Number of cores to manage 18 MIOPS
10,000 CPU cycles	3.7 us	270,000 IOPS	3.3 cores	60 cores

Source: Intel Feb 2015 "Performance Benchmarking for PCIe and NVMe Enterprise Solid-State Drives White Paper



NETWORK
x15 40GE 600Gbit/s

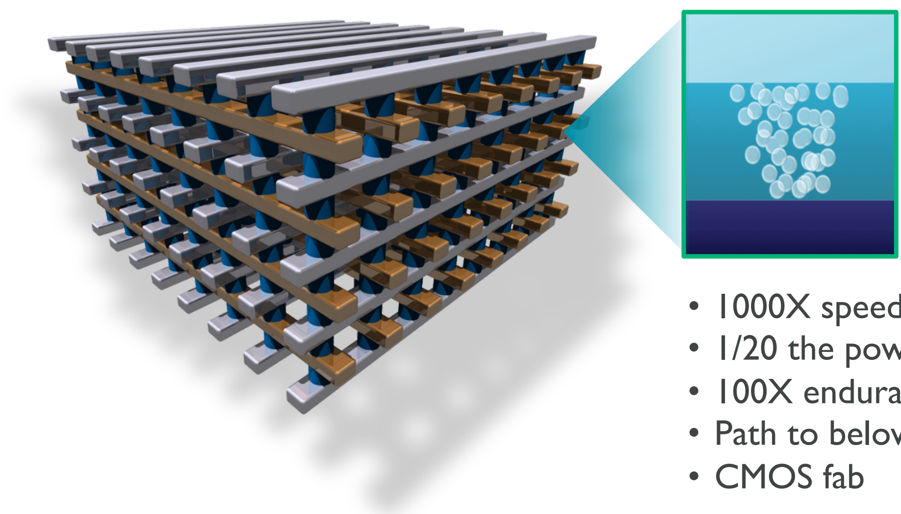
COMPUTE
Quad 20-cores Intel CPU

STORAGE
24 Flash NVMe SSDs @750kIOPS
18 MIOPS

75% of CPU cores used only for IO management !

RRAM revolutionizing NVM technology

NANO-FILAMENTS ARE THE NEW ELECTRONS



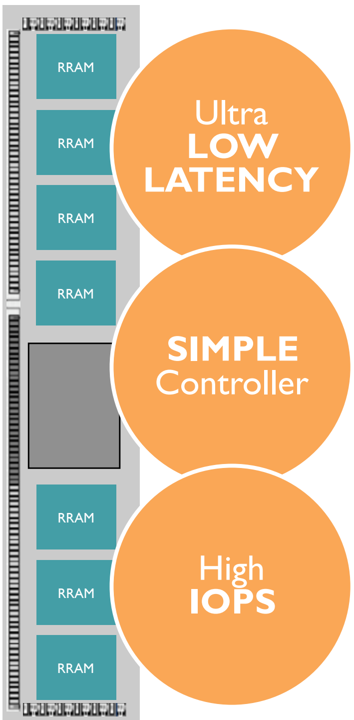
- 1000X speed
- 1/20 the power
- 100X endurance
- Path to below 10nm
- CMOS fab

1ST MANUFACTURING PARTNERSHIP WITH SMIC

READY FOR LICENSING

300mm wafer

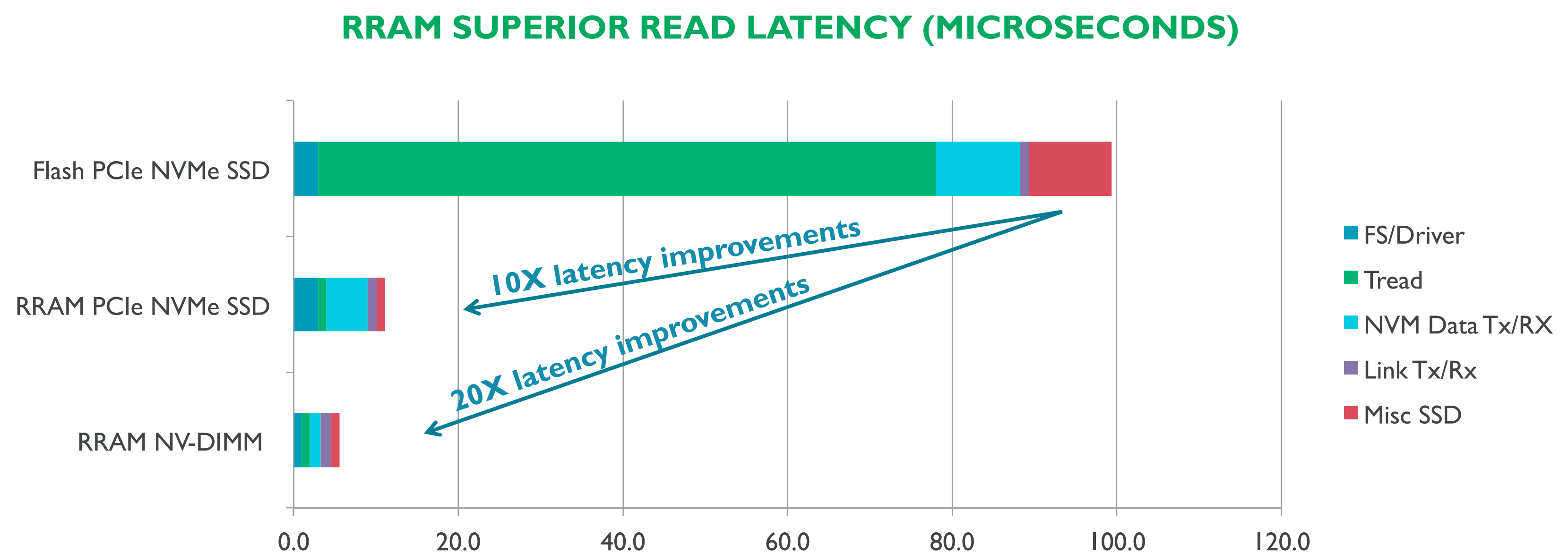
RRAM unleashing storage solutions performances



- At chip level:
  - 100X: **1us** random read vs 100s us for 3D TLC NAND
  - 1000X: **2us** writes vs milliseconds
- At storage level:
  - 4us on NV-DIMM
  - 10us on NVMe SSD
- Overwrite capability:
  - No erase operation required
  - No FTL
  - No Garbage Collection
  - No Over-provisioning
- High endurance
- Small page support. Almost no performance gap between RD and WR
- **24 MIOPS (512B)** on a 4TB NV-DIMM (no DRAM, no super caps) - 3 MIOPS with 4KB IOs
- **6.4 MIOPS (512B)** on a NVMe SSD – 800 KIOPS with 4KB IOs

Unique low latency high density storage solutions

RRAM solving storage latency limitations



Your big data accessible below 5us latencies

Performance bottleneck shift from storage to compute in hyperconverged server

Hyper-converged 

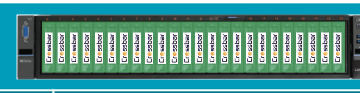
Compute	**Storage & memory**
36 cores • dual 18-core CPU	24 x 2.5" drives: • 4 SSD (400 or 800GB) • 20 HDD (1 or 2 TB) 24 DIMM socket: • 16 DRAM-DIMM (GB)
Max IOPS supported by Compute 12MIOPS **>**	Max IOPS supported by Storage 2.8 MIOPS

Max IOPS: 2.8 MIOPS

limited by Storage

3000 IOPS/VM

Performance bottleneck shift from storage to compute in hyperconverged server

Hyper-converged 		**Hyper-converged** 	
Compute	**Storage & memory**	**Storage & memory**	**Compute**
36 cores • dual 18-core CPU	24 x 2.5" drives: • 4 SSD (400 or 800GB) • 20 HDD (1 or 2 TB) 24 DIMM socket: • 16 DRAM-DIMM (GB)	24 x 2.5" drives • PCIe Crossbar RRAM SSD (up to 8 TB) 24 DIMM socket: • 16 DRAM-DIMM • 8 Crossbar RRAM-DIMM (up to 4 TB)	36 cores • dual 18-core CPU
Max IOPS supported by Compute 12MIOPS >	Max IOPS supported by Storage 2.8 MIOPS	Max IOPS supported by Storage 43 MIOPS >	Max IOPS supported by Compute 36MIOPS
Max IOPS: 2.8 MIOPS **limited by Storage**		Max IOPS: 36 MIOPS **limited by Compute**	

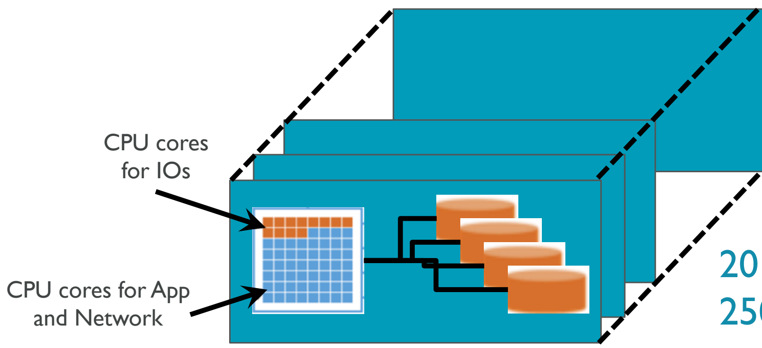
12X →

3000 IOPS/VM

3D RRAM moving performance bottleneck from storage to compute

New era of RRAM-based micro-servers

CPU time per NV-DIMM IO	100% CPU core saturation	Number of cores to manage 1 MIOPS	Number of cores to manage 12 MIOPS
1 us	1,000,000 IOPS	1 core	12 cores



20 microservers per 2U chassis
250 MIOPS per server

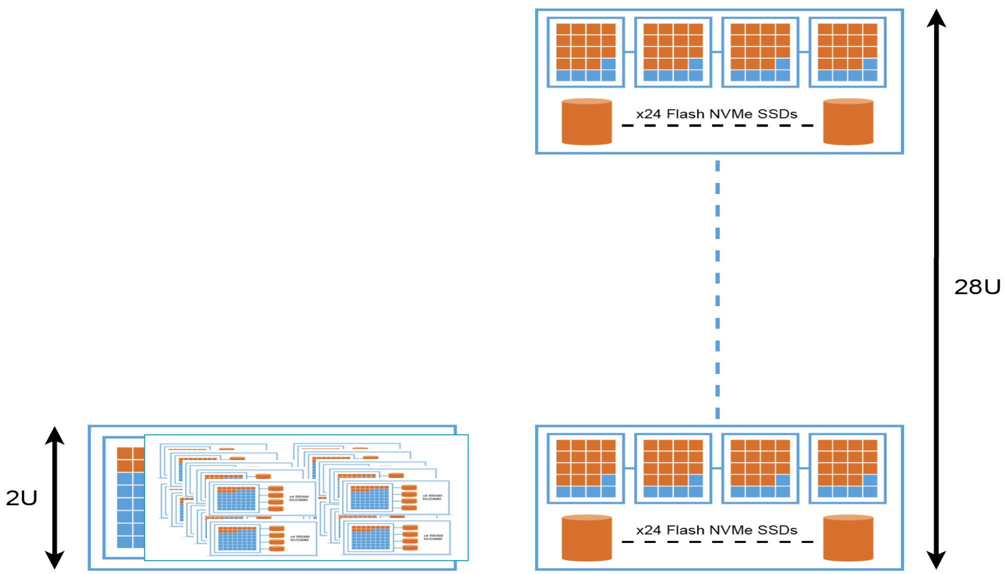
NETWORK
8 Tbit/s

COMPUTE
64-cores ARM CPU

STORAGE
4 RRAM NV-DIMMs @3MIOPS
12 MIOPS per microserver

80% of cores available for computing !

14X more VMs per server with RRAM



Need 14 2U-servers to match same number of VM with Flash-based STORAGE

RRAM-based STORAGE
20 microservers @12MIOPS
250 MIOPS per 2U
83,000 VMs

14X

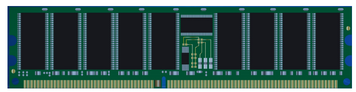
FLASH-based STORAGE
1 server @18MIOPS

6,000 VMs

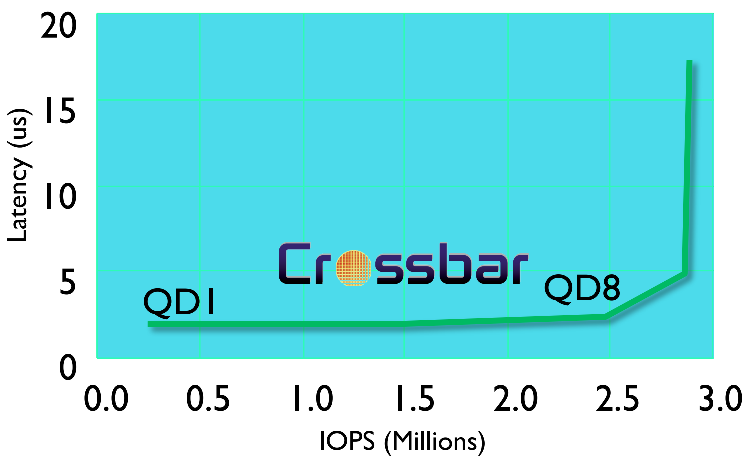
3000 IOPS/VM

Redefining storage solutions with smaller IO size

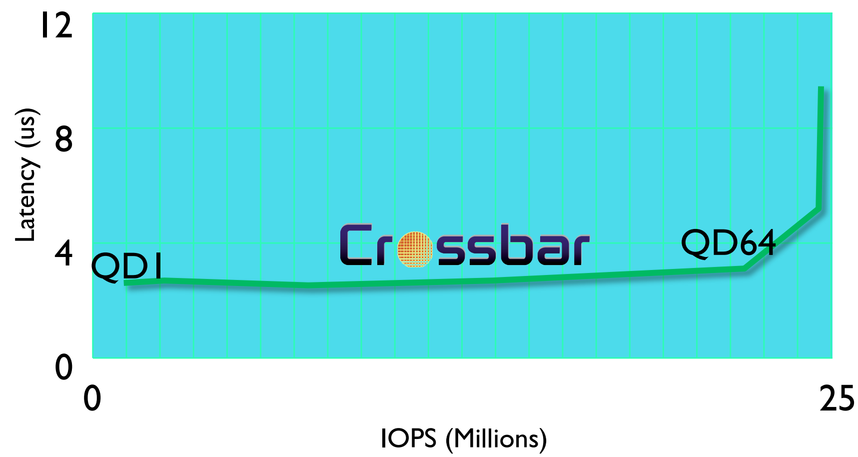
3 MILLION IOPS



24 MILLION IOPS



4K IOPS



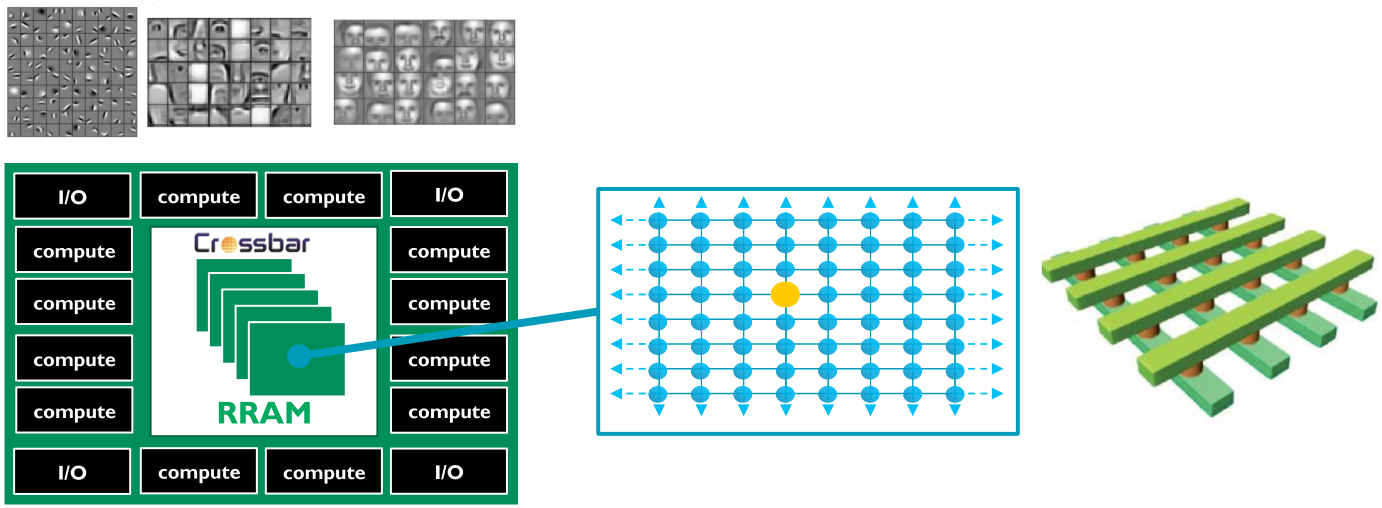
512B IOPS

Delivering GIOPS servers

	4KB IO	512B IO
DIMM IOPS	3 M	24 M
Microserver IOPS (4 DIMMs / microserver)	12 M	96 M
2U server (20 microserver / 2U)	250 MIOPS	2 GIOPS
IOPS/U	125 MIOPS/U	1 GIOPS/U

Performance bottleneck shifted from storage to compute and network

Next - Enabling new era of computing



Unique memory-centric parallel computing platform

Perfect match for deep neural networks hardware acceleration

Thank you