From Edge to Core
Memory-Driven Hardware and Software Co-design for the intelligent enterprise
The Second Coming
— THE SECOND COMING – W. B. Yeats - 1919

Turning and turning in the widening gyre
The falcon cannot hear the falconer;
Things fall apart; the centre cannot hold;
Mere anarchy is loosed upon the world,
The blood-dimmed tide is loosed, and everywhere
The ceremony of innocence is drowned;
The best lack all conviction, while the worst
Are full of passionate intensity.

Surely some revelation is at hand;
Surely the Second Coming is at hand.
The Second Coming! Hardly are those words out
When a vast image out of Spiritus Mundi
Troubles my sight: a waste of desert sand;
A shape with lion body and the head of a man,
A gaze blank and pitiless as the sun,
Is moving its slow thighs, while all about it
Wind shadows of the indignant desert birds.

The darkness drops again but now I know
That twenty centuries of stony sleep
Were vexed to nightmare by a rocking cradle,
And what rough beast, its hour come round at last,
Slouches towards Bethlehem to be born?
### Oh, Inverted World!

<table>
<thead>
<tr>
<th>Compute Abundance</th>
<th>Memory Abundance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moore's Law Scaling</td>
<td>Photonic Scaling</td>
</tr>
<tr>
<td>Imperative</td>
<td>Declarative</td>
</tr>
<tr>
<td>Programming</td>
<td>Training</td>
</tr>
<tr>
<td>Centralized</td>
<td>Distributed</td>
</tr>
<tr>
<td>General Purpose</td>
<td>Precision</td>
</tr>
<tr>
<td>Data as a Cost Center</td>
<td>Data as a Profit Center</td>
</tr>
</tbody>
</table>
The New Normal: Compute is not keeping up

**Microprocessors**

- **Transistors (thousands)**
- **Single-thread Performance (SpecINT)**
- **Frequency (MHz)**
- **Typical Power (Watts)**
- **Number of Cores**

**Data growth**

- Data nearly doubles every two years (2013-2020)
- Data (Zettabytes)

- 2006: 0.3 Zettabytes
- 2008: 0.8 Zettabytes
- 2010: 1.2 Zettabytes
- 2012: 1.8 Zettabytes
- 2014: 4.4 Zettabytes
- 2016: 7.9 Zettabytes
- 2018: 15.8 Zettabytes
- 2020: 44 Zettabytes
The end of scaling at just the wrong or just the right time ...

( 8B people x 20B mobile devices x 100B social infrastructure x 1T apps )
Opportunity - the hyper-competitive digital enterprise

The hyper-competitive digital enterprise:

• Understands data is the new source of competitiveness and economic value creation, of equal or greater value than the underlying commodity or process

• Instruments every physical or digital product, every manufacturing process in the factory, every business process in the enterprise to produce data

• Pushes analytic and machine learning capability as close as possible to the edge for real time insight and action

• Forges a continuum from the enterprise core to the intelligent edge

• Relentlessly and remorselessly turns raw data to economic advantage via process improvement, investment strategy, customer satisfaction, market expansion, warranty reduction, direct monetization
Re-establishing innovation horizons

Business Units
Near future innovation

Hewlett Packard Labs

Applied Research
2 – 5 years

Exploratory Research
5 – 20+ years

Advanced Development
Up to 2 years

EVOLUTIONARY

REVOLUTIONARY
Inversion as Architecture

Today’s architecture
From processor-centric computing

Future architecture
Memory-Driven Computing
HPE introduces the world’s largest single-memory computer
The prototype contains 160 terabytes of memory


- An optimized Linux-based operating system running on ThunderX2, Cavium’s flagship second generation dual socket capable ARMv8-A workload optimized System on a Chip.

- Photonics/Optical communication links, including the new X1 photonics module, are online and operational.

- Software programming tools designed to take advantage of abundant of persistent memory.
HPE’s X1: Fully integrated photonics interconnect chip module
The current prototype

Node

SoC

mem

comp

I/O

Mem

Control

Z-bridge

fabric

P

mem

mem

mem

mem

P

P

P

P

Q

Z-bridge

fabric

P

mem

mem

mem

mem

P

P

P

P

Q

Z-bridge

fabric

P

mem

mem

mem

mem

P

P

P

P

Q

Z-bridge

fabric

P

mem

mem

mem

mem

P

P

P

P
Mature Gen-Z devices

SoC with fabric addresses

Compute accelerator

I/O device (superNIC)

Exascale processor

THINK GLOBAL (ADDRESS SPACE) – ACT LOCAL (TASK SPECIFIC ENDPOINTS)

Memory node
with memory-side accelerator
What are core Memory-Driven Computing components?

**Fast, persistent memory**
Combining memory and storage in a stable environment to increase processing speed and improve energy efficiency.

**Fast memory fabric**
Using photonics where necessary to eliminate cost of distance and create otherwise impossible topologies.

**Task-specific processing**
Optimizing processing from general to specific tasks and embrace novel computational techniques.

**New and Adapted software**
Radically simplifying programming and enabling new applications that we can’t even begin to build today.
How does Memory-Driven Computing enable new applications?

One architecture scales from the dense data center to the intelligent edge

**Memory Abundance**
- Similarity search
- Search space optimization
- Financials futures modeling

**Non-volatility of Memory**
- Scalable transactional key value stores
- Managed data structures
- Energy scalability and retention

**Memory shared with just the right compute**
- Spark in-memory Hadoop
- Deep neural net training
- Network function virtualization

**Dynamic Range**
- Memory-Driven Computing edge
- Consistent node, enclosure, rack, row, data center
Transform performance with Memory-Driven programming

Modify existing frameworks

New algorithms

Completely rethink

In-memory analytics

Similarity search

Large-scale graph inference

Financial models

15x faster

40x faster

100x faster

10,000x faster
Open, open, open

The state of the MFT prototype when we announced at HP Discover, June 2016
Gen-Z: new open interconnect protocol  
Key enabler of the Memory-Driven Computing open architecture

- High Bandwidth
- Low Latency

- Advanced workloads & technologies
- Scalable from IoT to exascale

- Compatible
- Economical
Industry collaboration on interconnect technology

Industry leaders developing a next generation, memory-semantic interconnect

www.genzconsortium.org

Current Membership

- Alpha Data
- AMD
- Amphenol Corporation
- ARM
- Broadcom
- Cavium Inc.
- Cray
- Dell EMC
- Everspin
- FoxxConn Interconnect
- HPE
- Huawei
- IBM
- IDT
- Intelliprop
- Jabil
- Lenovo
- Lotes
- Luxshare ICT
- Mellanox Tech, Ltd
- Mentor Graphics
- Micron
- Microsemi
- Molex
- NetApp
- Nokia
- NumaScale
- PLDA Group
- Red Hat
- Samsung
- Seagate
- Smart Modular
- SK Hynix
- Spin Transfer Tech
- TE
- VMware
- Western Digital
- Xilinx
- YADRO

39 Members (and growing)

12 Board Companies
Memory-Driven Computing Developer Toolkit
Software already available to you

Example Applications
- Image Search
- Large Scale Graph Inference

Data Management & Programming Frameworks
- Sparkle
- Persistent memory toolkit
- Managed data structures
- Fault-tolerant programming
- Fast optimistic engine

Node Operating System
- Linux for Memory-Driven Computing
- Librarian File System (LFS)
- Persistent Memory Library (pmem.io)
- Fabric attached memory atoms library

Management Services
- Librarian

Emulation/Simulation Tools
- Fabric attached memory emulation
- Performance emulation for NVM
- X’86 emulation (Superdome X, MC990x, ProLiant)

Open source components
- Node Operating System
- Persistent Memory
- Library (pmem.io)
- Librarian File System (LFS)
- Fabric attached memory
- Atoms library

- Example Applications
- Programming and analytics tools
- Operating system support
- Emulation/simulation tools

Get access to the toolkit:
https://www.labs.hpe.com/the-machine/developer-toolkit
What can you do in 300ns?
It’s more than just bits per (meter • second • Watt • dollar) between transmit and receive

Budget for the endpoint:
$300\,\text{ns} - 150\,\text{ns} = 150\,\text{ns}$

Budget for the hops:
$150\,\text{ns} - 3 \times 25\,\text{ns} = 75\,\text{ns}$

Give the balance to fiber:
$75\,\text{ns} \times \frac{1\,\text{m}}{5\,\text{ns}} \times \frac{1}{2} = 7.5\,\text{m}$

Describe the sphere:
$\frac{4}{3} \pi (7.5\,\text{m})^3 = 994\,\text{m}^3$

Convert to racks:
$994\,\text{m}^3 \times \frac{1\,\text{rack}}{8\,\text{m}^3} = 124\,\text{racks}$

124 MFT racks = 19PB, 159K cores