High-Dimensional Computing

SDM Memory & Hyperscaling

A session on the Cloud's missing AI component

Gil Russell

WebFeet Research Inc.

In-Memory Computing SUMMIT | NORTH AMERICA 2018

WebFeet Research

Research | Vision | Knowledge

AI - a technology in transition

"Science advances one funeral at a time."

Max Planck

"Randomness is the path of least presumption."

Pentti Kanerva

On the road to Artificial General Intelligence - What we have learned (so far)

Machine Learning ≠ Intelligence

Deep Learning Promise

- *Has entered the trough of disillusionment*

Sequences have become mandatory

- *Requires hierarchical temporal capability*

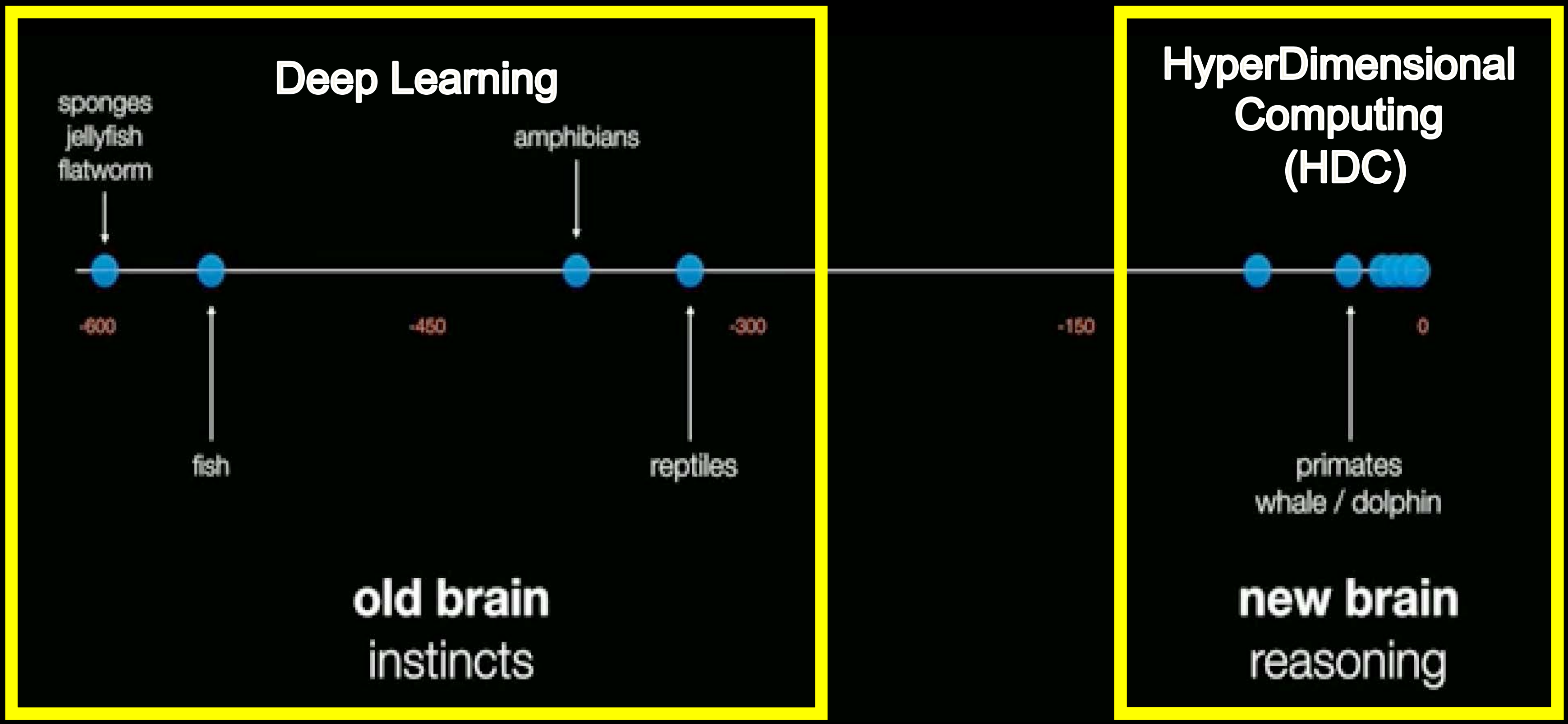
Prediction absolutely required

Fast learning required for anomalous conditions

Broad span solution

- *Embedded Controllers through High-Performance Computing*
- *Cognitive Database compatibilities*

New vs. Old Brain



AI - a technology in transition

Old Brain => Deep Learning (ANNs)

Stimulus, Instinct & Response

Low generalization

Good at high probability inference

Requires large amounts of training data

Limited real time processing and incapable of real time predictive reasoning - brittle failure modes

Autonomous Vehicle developers are rethinking their technology tracks

AI - a technology in transition

New Brain => Hyperdimensional Computing

Relates human intelligence to the properties of abstract mathematical spaces

Associates patterns and analogies – What if and Why?

Single pass training, High Noise Immunity

High range of applicability, from toys to supercomputers

*Can **predict** probable outcomes*

Hyper-Dimensional Computing (HDC)

Google: Visualizing High Dimensional Space

'Old Brain' Memory

Artificial Neural Networks adoption of the Deep Learning algorithm (DL) has limited the development of algorithms that can deal with generality and the mapping of associations

DL designs rapidly adopted Parallel Arrays of Multiply-Accumulate (MAC) solutions using High Bandwidth Memory (HBM)

Memory is internal to the algorithm used and has no external object "association" tether

Dependence on a purely statistical algorithm with little reference to a memory framework is DLs weakest link

New Brain" Memory Fabric

HDC effect is subtle

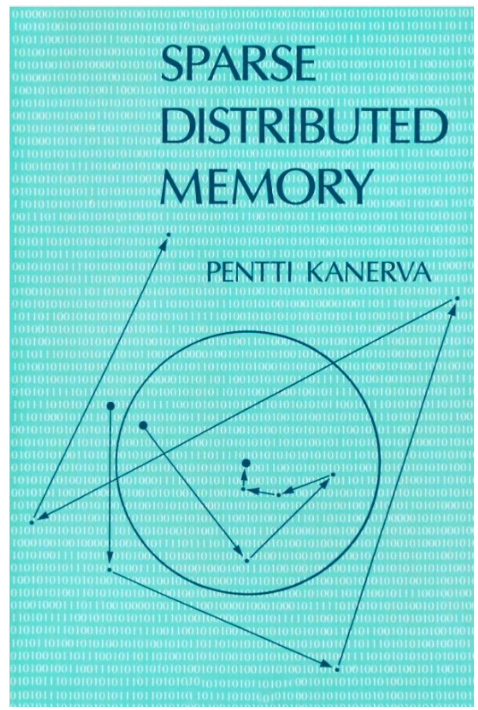
- *Requires overlay of an Cognitive Object Data Base*
- *Node (Object) and Vector Edge represent either Hamming Distance SDM or Cosine Overlap in Hierarchical Temporal SDR to nearest and farthest relations*
- *Vector relationships are being incorporated into Cognitive Object Databases and Probabilistic Programming Languages*

"Sparce, Distributed Memory (SDM)"

Pentti Kanerva

Redwood Center for Theoretical Neuroscience

"Sparse, Distributed Memory", MIT Press, 1988



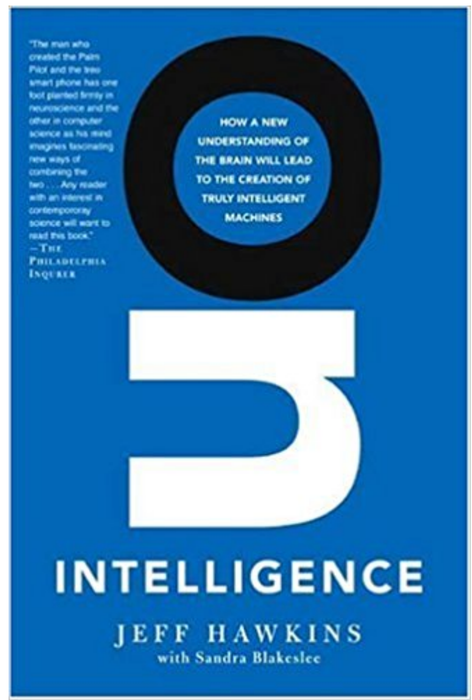
"Hyperdimensional Computing: An Introduction to Computing in Distributed Representation with High-Dimensional Random Vectors", Pentti Kanerva, January 2009

https://redwood.berkeley.edu/

Sparce, Distributed Memory (SDM) + Hierarchical Temporal Memory (HTM)

Jeff Hawkins

Founder: Redwood Center for Theoretical Neuroscience (2002), CEO Numenta (2005)



"On Intelligence, How a New Understanding of the Brain Will Lead to the Creation of Truly Intelligent Machines, 2004"

https://numenta.com/

New Brain - Key People
Redwood NeuroscienceInstitute (RNI)

Pentti Kanerva
Research Affiliate at the Redwood Neuroscience Institute

Dileep George
Cofounder Vicarious AI
Cofounder and CTO Numenta

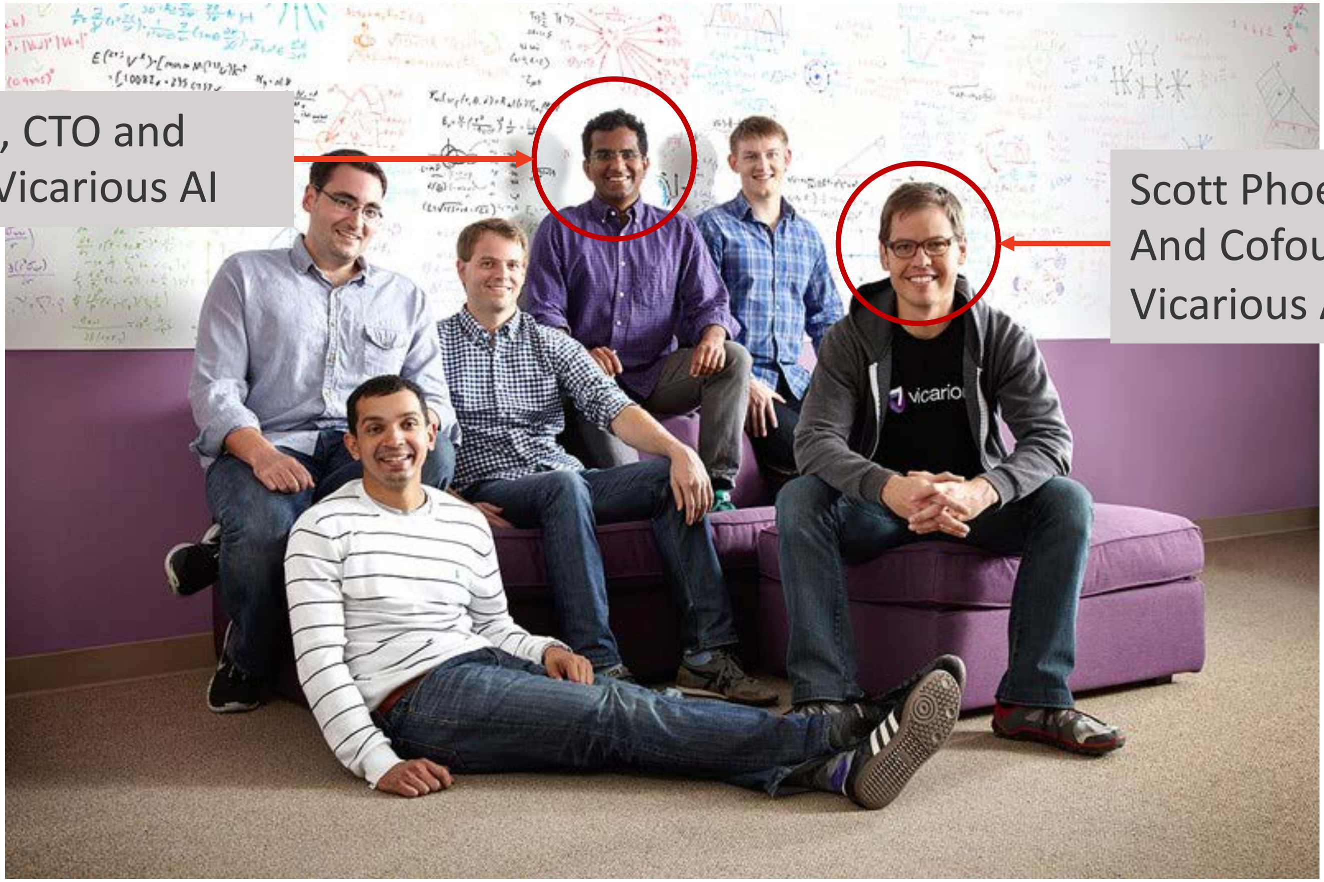
Bruno Olaufson
Professor Helen Wills Neuroscience Institute
Director and Cofounder, Redwood Center for Theoretical Neuroscience and Advisor to Vicarious AI

Jeff Hawkins
Cofounder of Numenta, Redwood Center for Theoretical Neuroscience (2005) and Redwood Neuroscience Institute (2002)

http://www.rctn.org/wiki/History

New Brain – Key People Vicarious AI (2014)

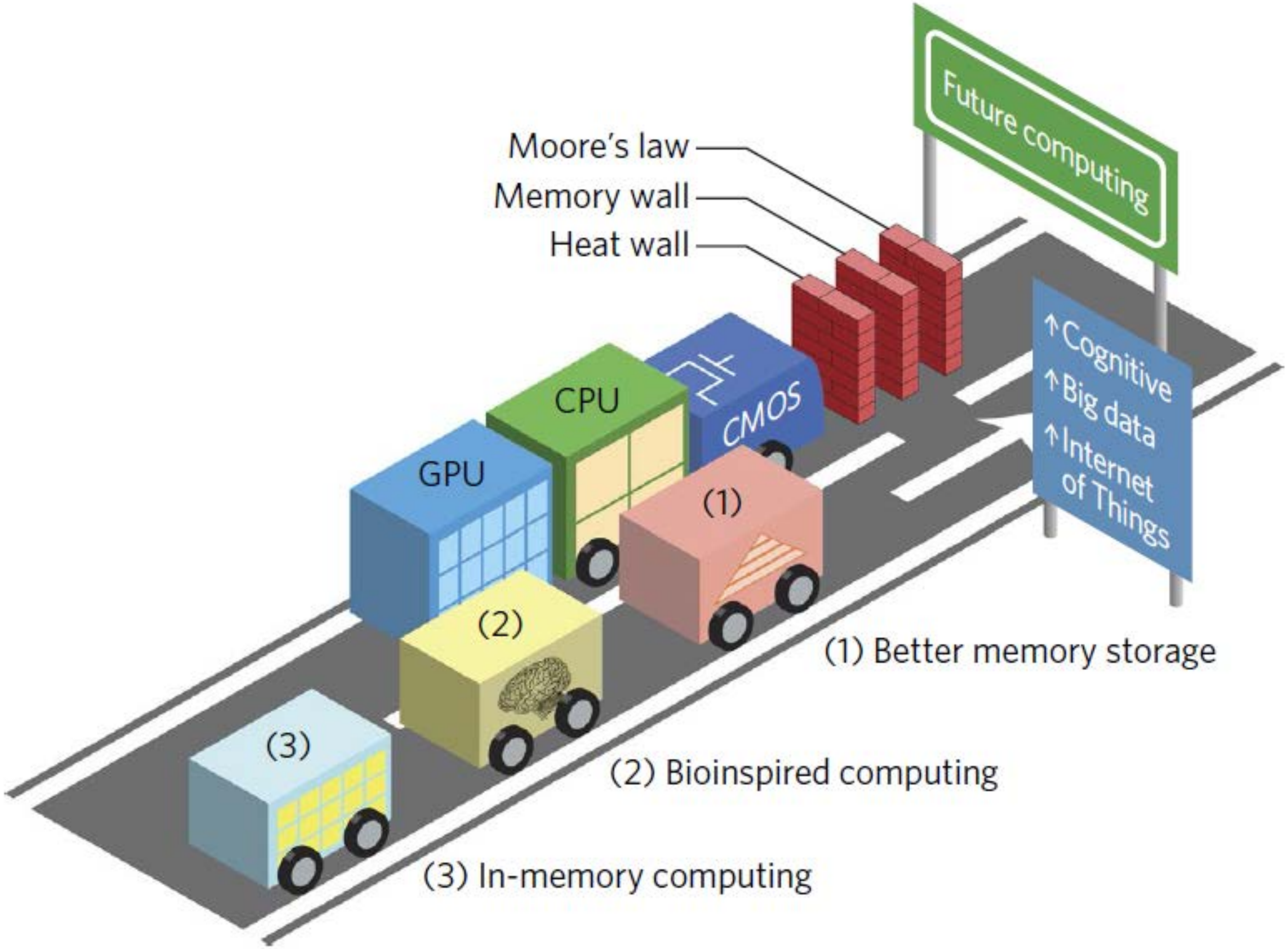
Dileep George, CTO and Cofounder of Vicarious AI



Scott Phoenix, CEO And Cofounder of Vicarious AI

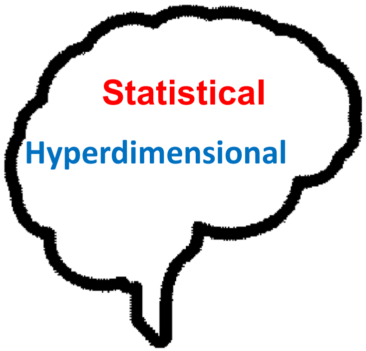
https://www.wired.com/2014/04/vicarious-ai-imagination/

Future Cognitive Computing Highway - Congestion

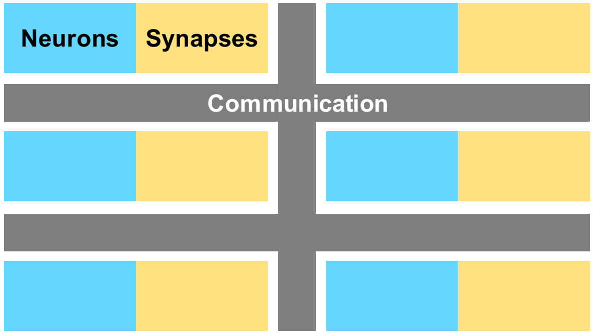


Future Cognitive Computing Highway - Trends

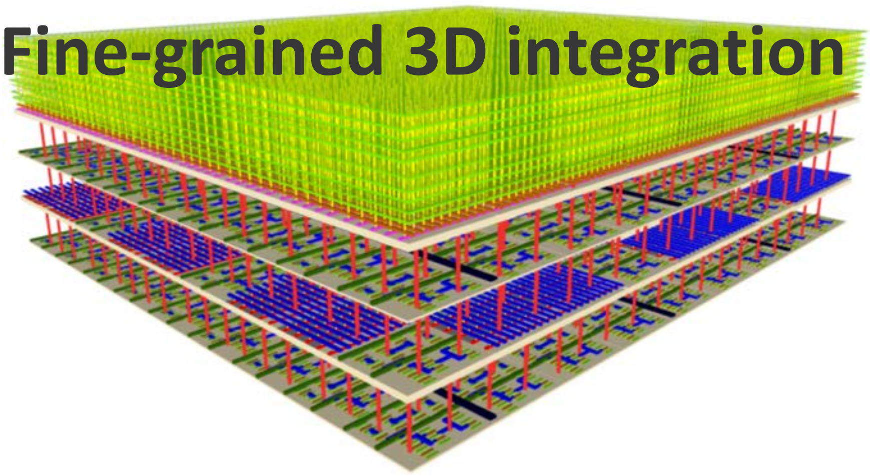
New computation models



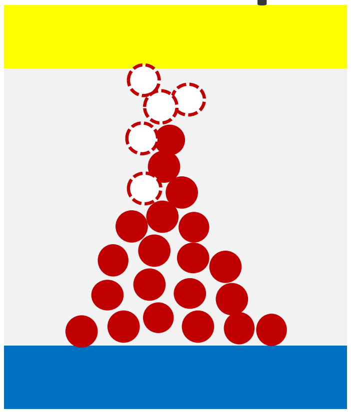
Tailored architectures



Fine-grained 3D integration



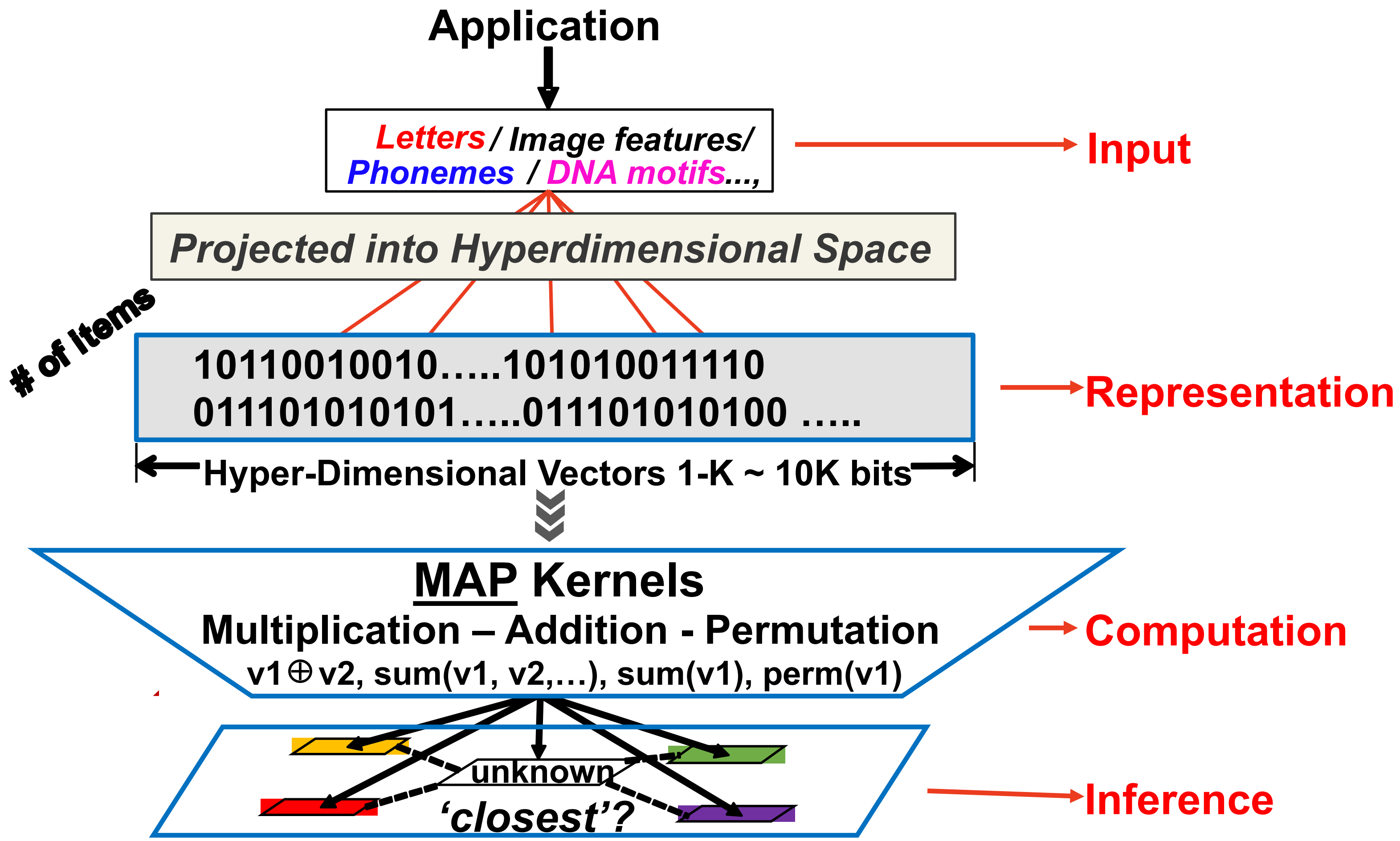
Intrinsic device properties



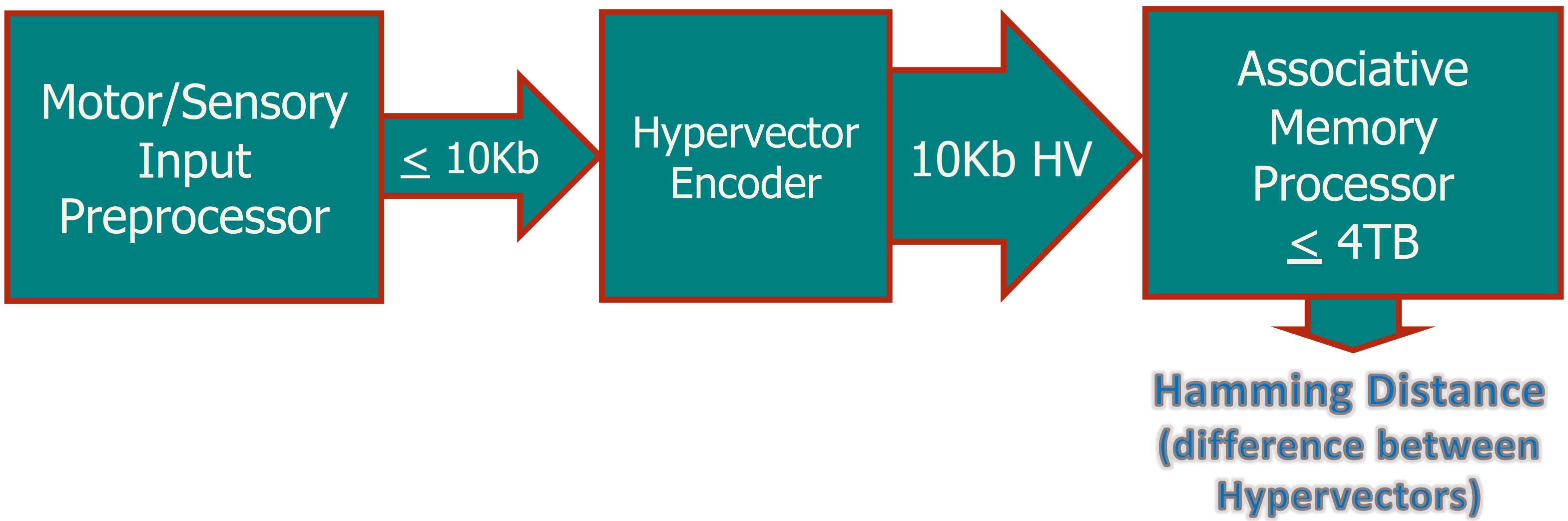
The Grail of Computing

Memory processing unit, can perform different tasks, data, storage, arithmetic, logic and neuromorphic computing using the same physical fabric that is programmable at the finest grain, the individual device level, without the need to move data outside the fabric.

Hyperdimensional Computing (HDC)



Hyperdimensional Computing (Simple Block Diagram)



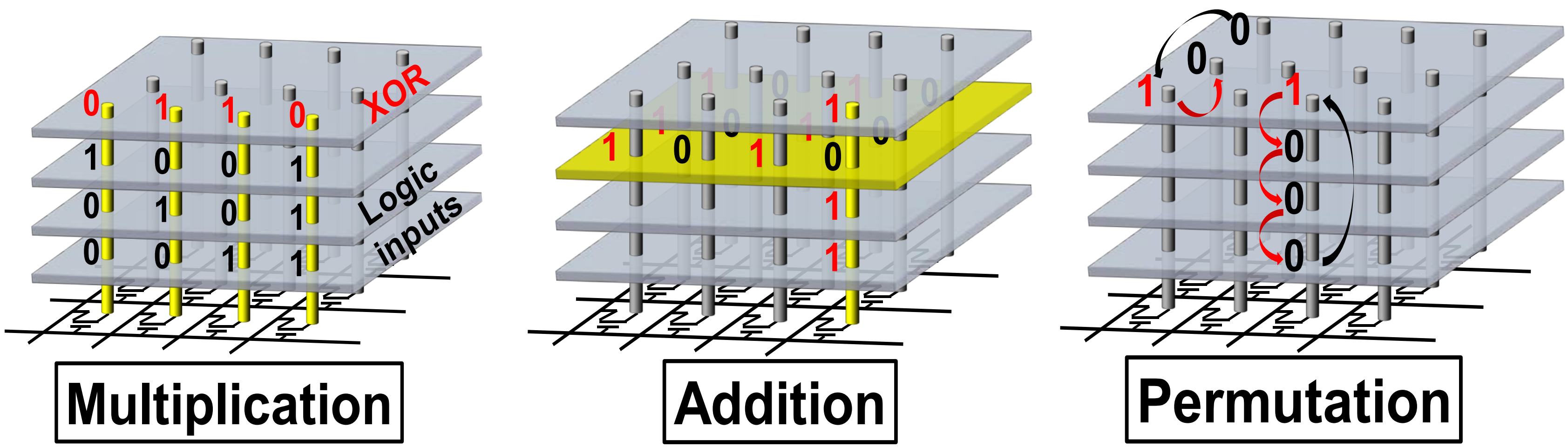
- **HV Address space ($2^{10,000}$ = 1.9950631168807583848837421626 e+3010)**
- **Associative Memory Processor Training + Experiential ≤ 4TB**

In-Memory MAP Kernels using 3D VRRAM

Leverage unique structure & properties of VRRAM

MAP kernels: demonstrated on 4-layer 3D VRRAMs

In-memory HD computing enabled in 3D architecture



In Non-Volatile Memory - NVRAM | NRAM

DDR4 Functionality

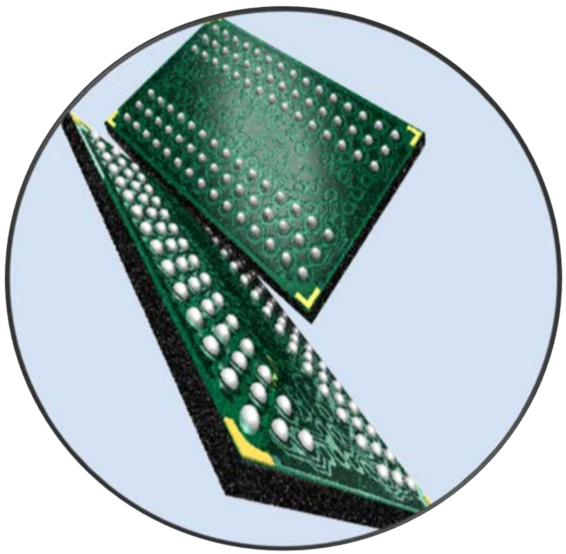
Low latency

CL $\leq$ 13.5 | 5 ns

Higher Density

Roadmaps

8 – 32 | 4 Layers



Persistence

10 years | 300 years

Endurance

10^{12} | 10^{15} PE

Zero Refresh

(or very low overhead) | None

Hyper-Dimensional Properties

General and scalable model of computing with a well-defined set of arithmetic operations

- Fast and one-shot learning
- Auditable (explainability) opens entirely new segment opportunities

Training sets reduced by orders of magnitude while still maintaining classification accuracy

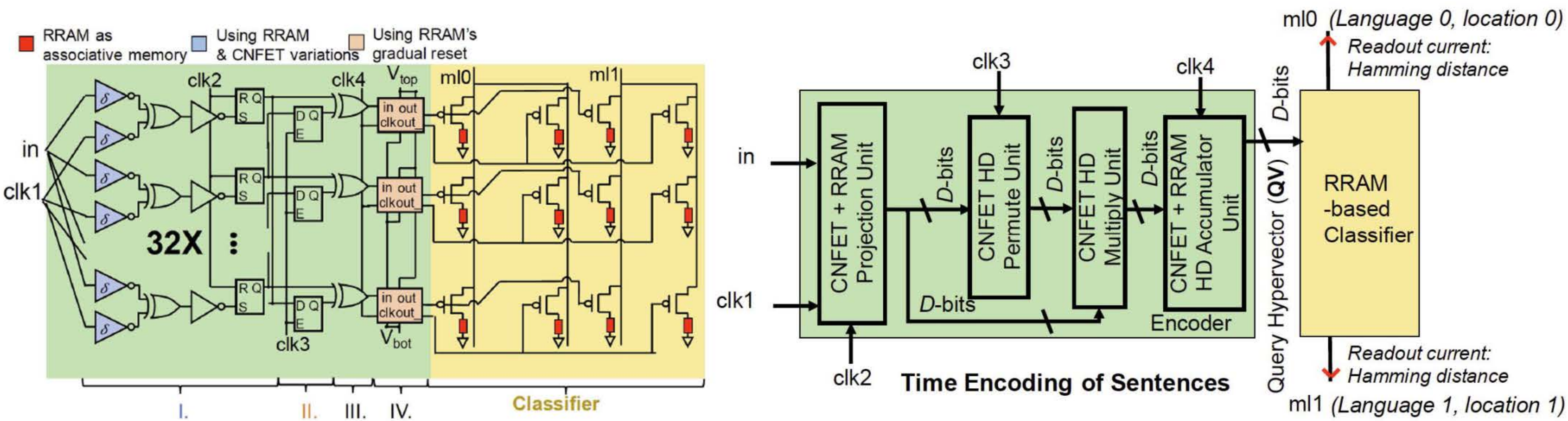
A *memory-centric architecture* with significantly parallelizable operations

Extremely robust against most failure mechanisms and noise

Predictive Analytics with high accuracy easily incorporated

Associative Base technology for the Cognitive Software Pyramid

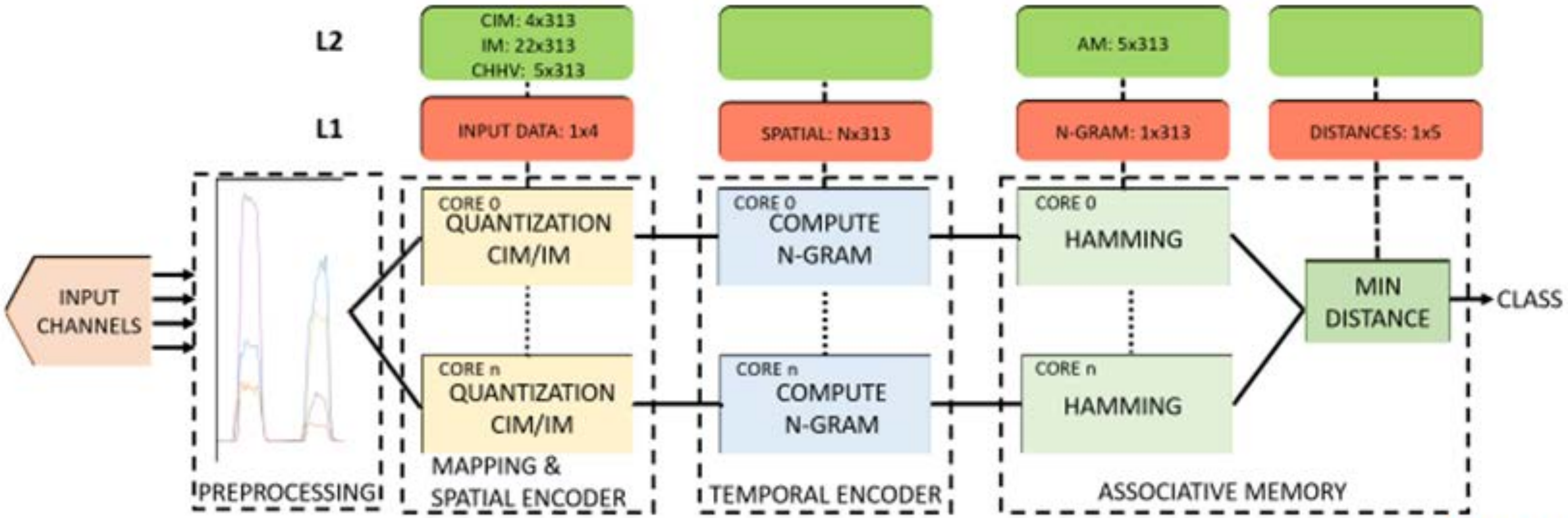
Hyperdimensional Computing Silicon 1



ISSCC 2018 / SESSION 31 / COMPUTATION IN MEMORY FOR MACHINE LEARNING / 31.3

Hyperdimensional Computing Silicon 2

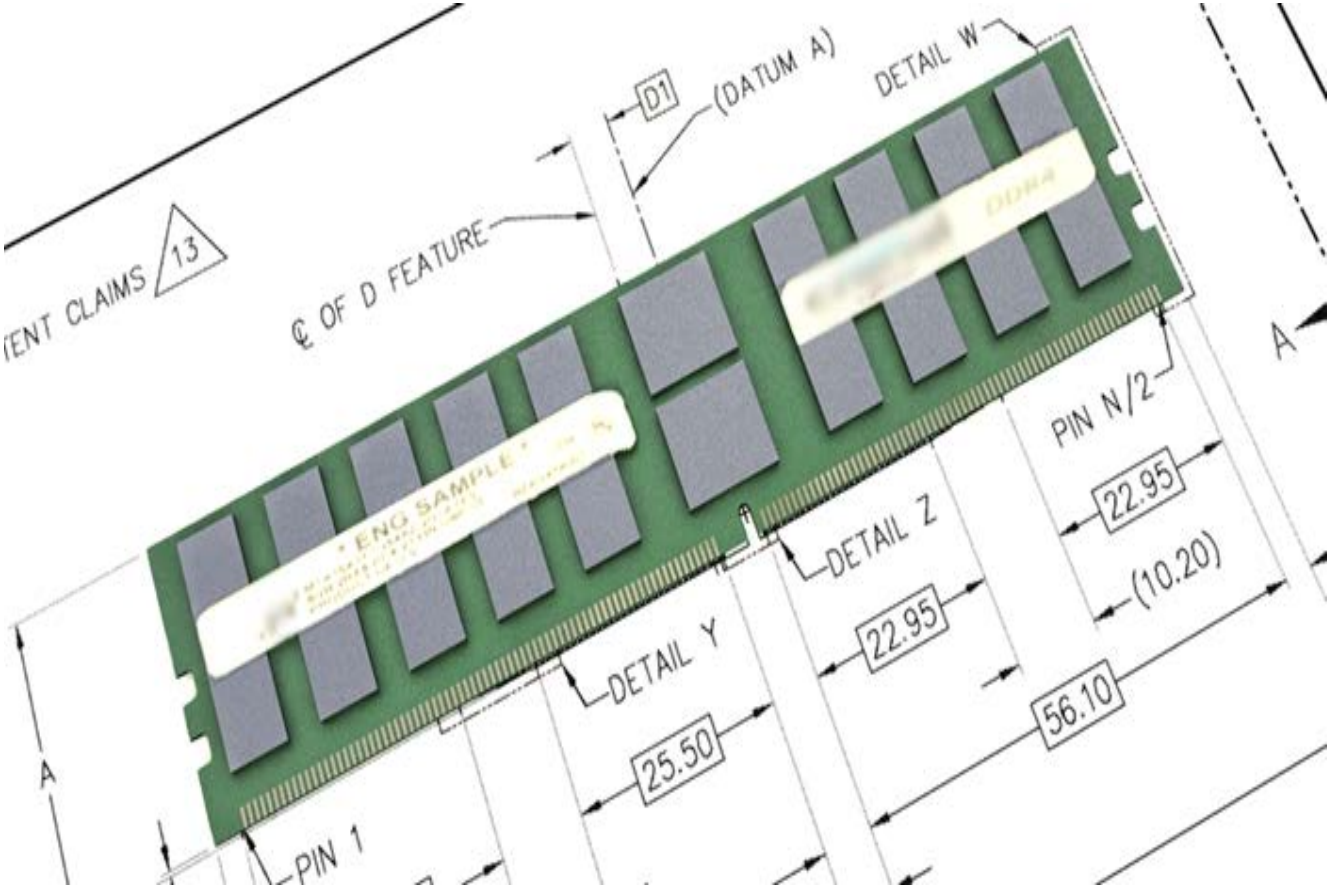
PULP-HD: Accelerating Brain-Inspired High-Dimensional Computing on a Parallel Ultra-Low Power Platform



24 April 2018

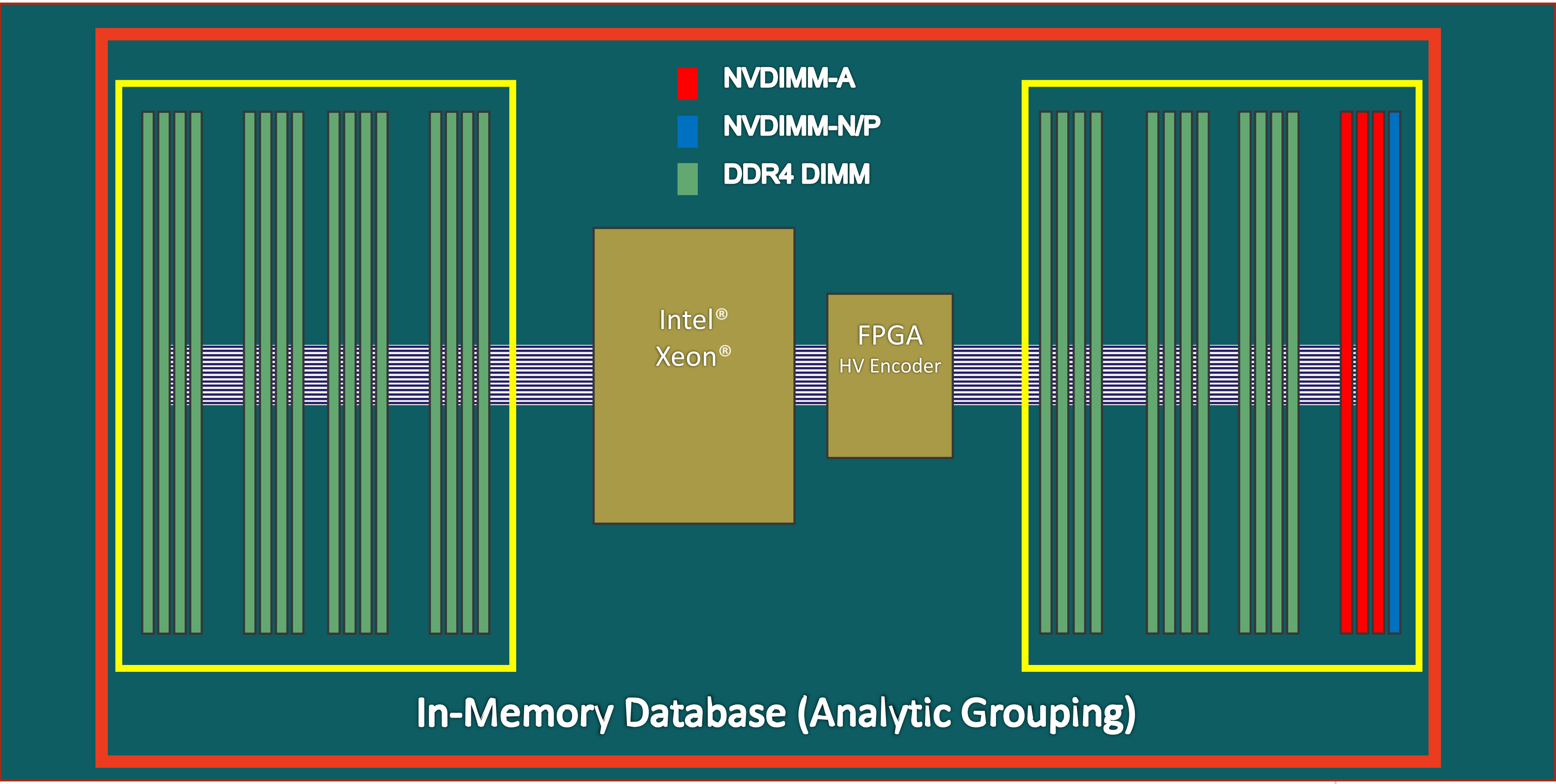
4-core platform (1.5mm2, 2 mW)

Hyperdimensional Computing Associative Processor - NVDIMM-X Form Factor



288 Pin NVDIMM Form Factor

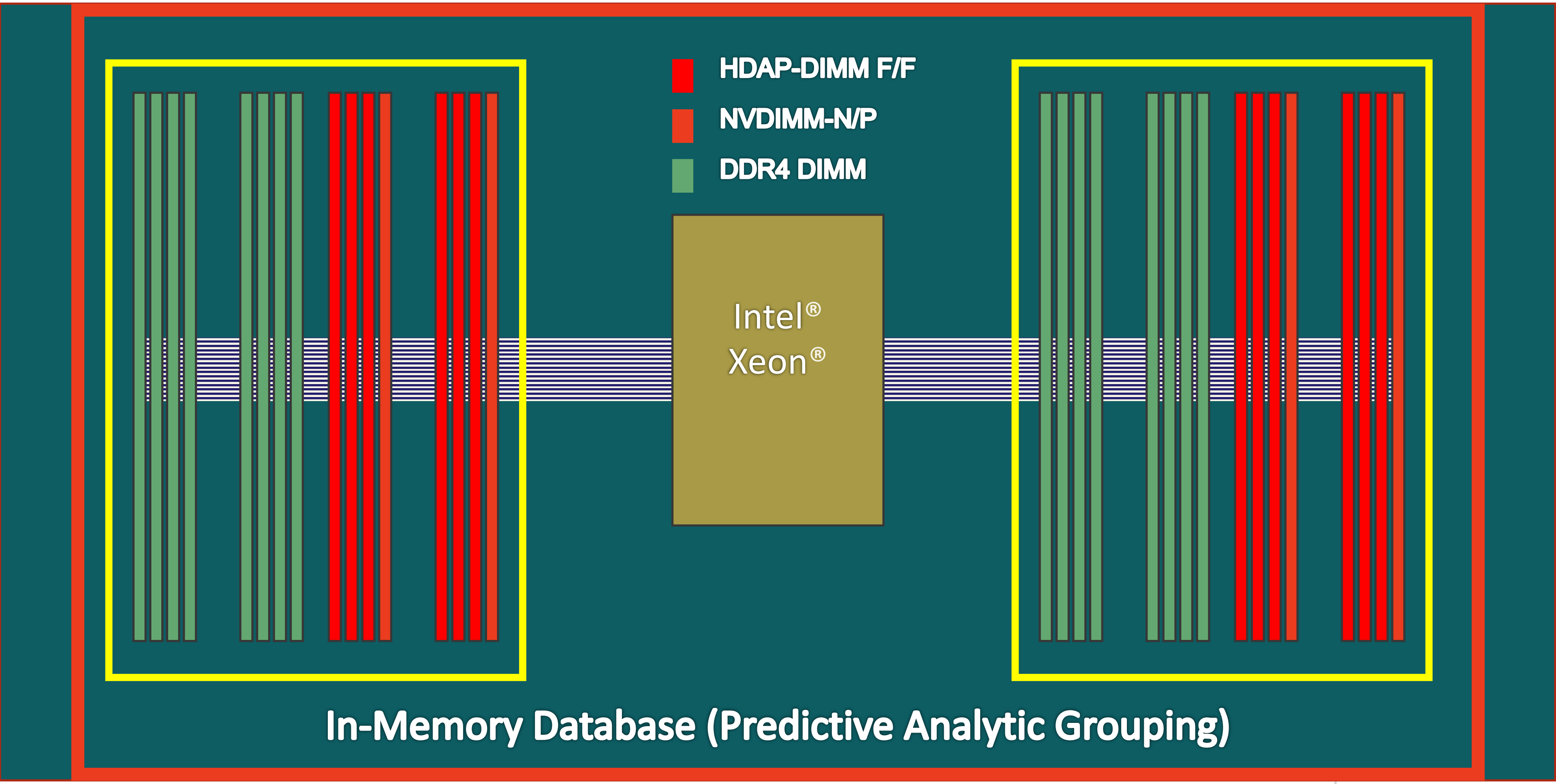
The Kanerva Partition (Database Intersect)
In-Memory Hypothetical System Implementation



The Kanerva Partition

(RT-PA Grouping)

In-Memory Hyperdimensional Associative Processor (Hypothetical)



Next Up:
'Big Problems for Nervous Tissue'

WebFeet Research, Inc.

www.webfeetresearch.com

Alan Niebel, CEO

- O. 831.373.3303
- M.831.402.5754
- Niebel@webfeetresearch.com

Gil Russell, Principal Technology Analyst

- O.510.589.9568
- F.510.489.1701
- Russell@webfeetresearch.com