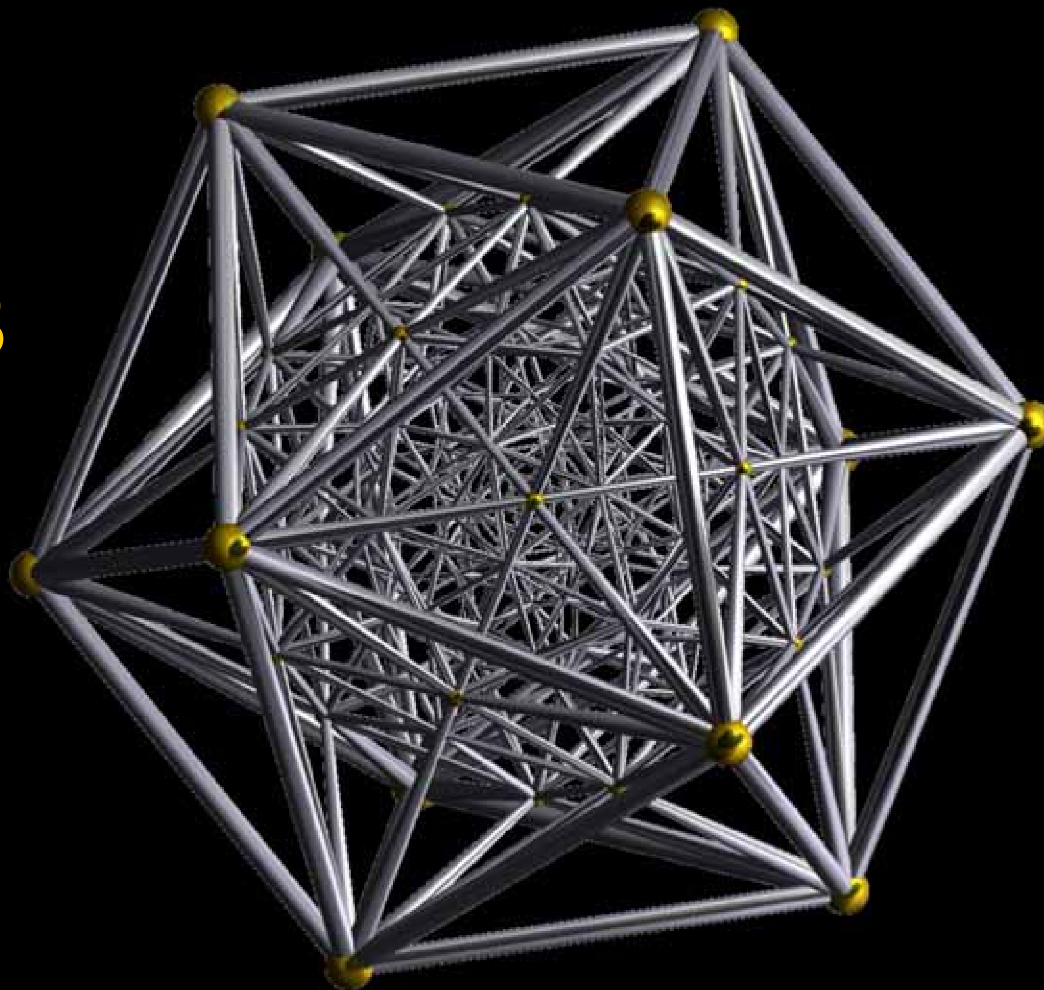
Adventures in High Dimensions



J. Rabaey, University of California @ Berkeley

With Abbas Rahimi, Sohum Dutta, Miles Rusch, Sayeef Salahuddin, Philip Wong, Subhasish Mitra, Penti Kanerva and Bruno Olshausen (and others)

Lund Circuits Workshop, September 25 2018

Rebooting Computing…

The nature of computing is changing

- Programming driven by data and learning, not algorithms
- Truly ubiquitous (smart world, smart humans, …)

While the technologies of old are plateauing

- Traditional computer architecture limited by interconnect
- Variability and leakage constraints limit energy scaling

The Neuroscience Promise

An Amazing Computational Engine

2-3 orders more efficient than today's silicon equivalent (>10^{16} FLOPS with ~20 W)

Robustness in presence of component failure and variations

- Neural response is highly variable (σ/μ≈1) [Faisal]

Amazing performance with mediocre components

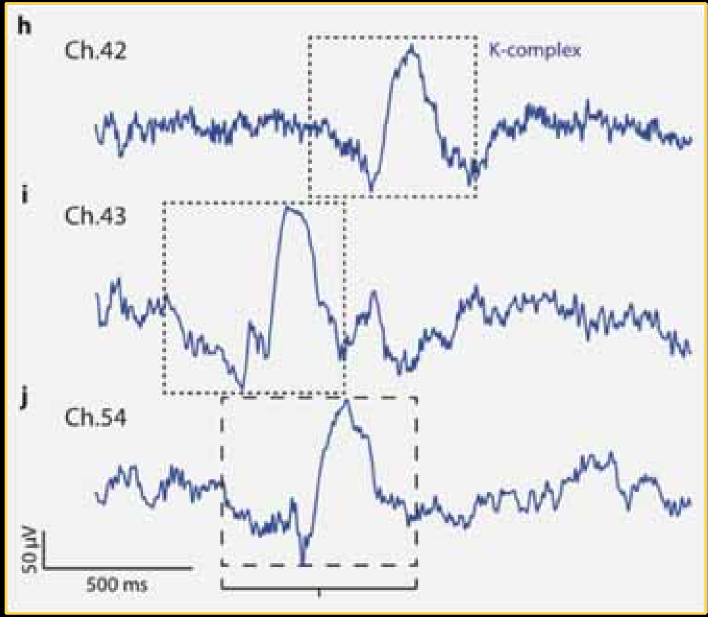
- E.g. sensory pathways– auditory, olfactory, vision, …

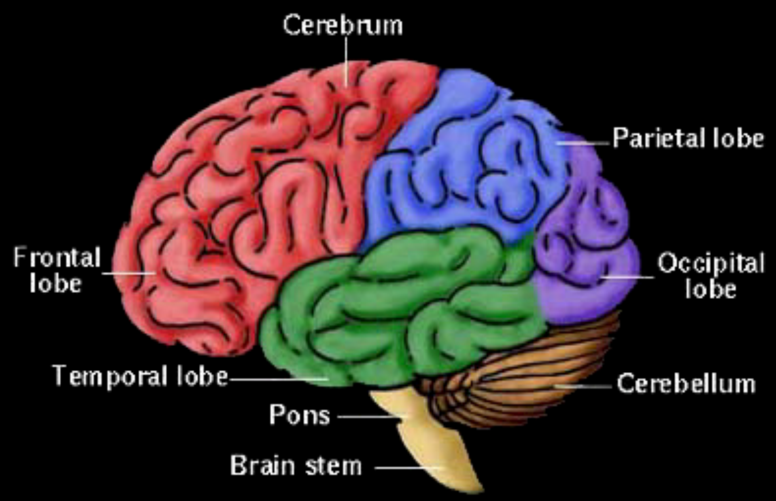
See:
" The return on neuro-inspired computing - Why now?"
Lund SoS workshop, Sept 2014

Still marginally understood, let alone "cloned"

Distinguishing Properties

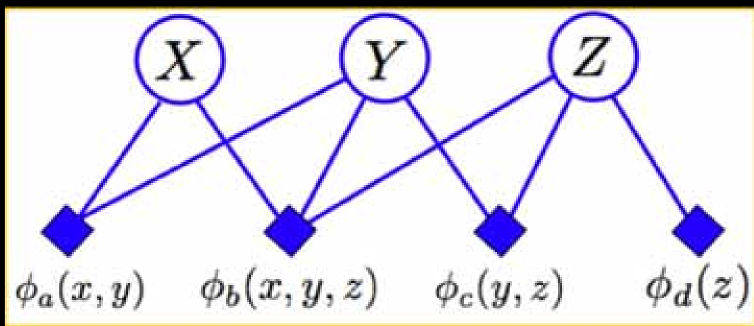
- Learning-based programming paradigm
- Approximate (statistical) & mostly analog
- Overcomplete and redundant
- Data represented in many ways
  - Patterns, phase relations, distributions
- Randomness as a feature





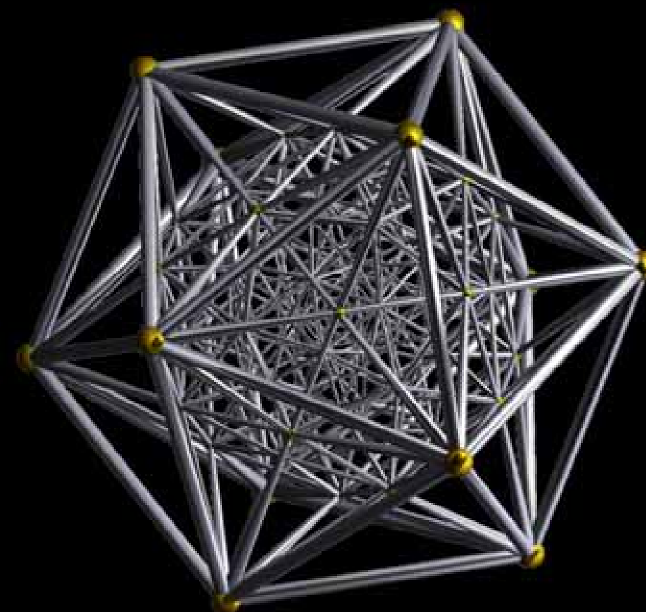
- Function mapped to space
  - no time multiplexing
- Intertwined memory and logic
- Embarrassingly parallel
- Sparse

Learning-Based Computational Models

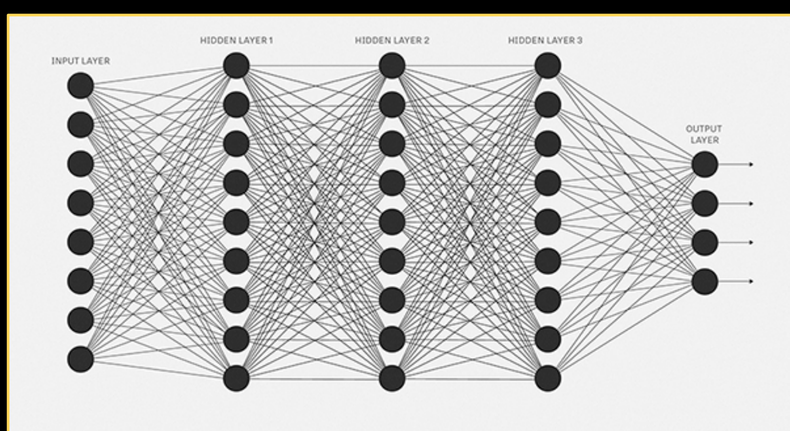


Bayesian Machine learning
(Believe propagation, reinforcement learning, graphical models, support-vector machines)

Model building non-trivial
Executed on standard processors (graph analytics)



High-dimensional computing (SDM, holographic)
Computing with patterns, one-shot learning



Deep Neural Nets
Learning compute and data hungry
Separate from execution
Complex

High-Dimensional Computing* as another approach

Cognitive processing that

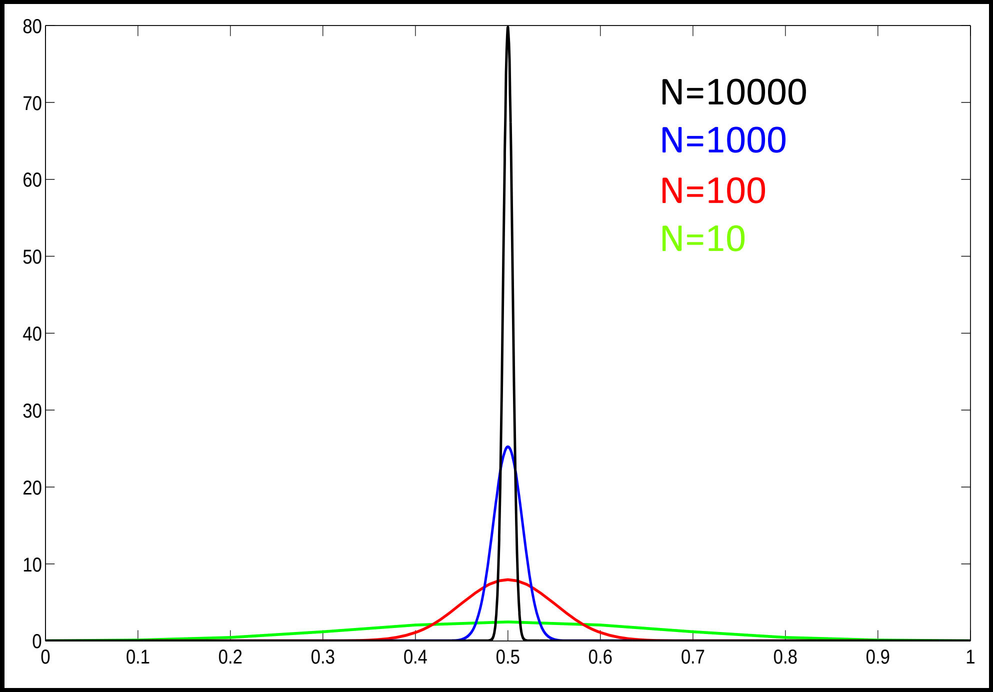
- provides simple and efficient on-line (one-shot) learning
- supports reasoning
- is embarrassingly parallel and memory-centric
- is extremely robust against most failure mechanisms
- offers ultra-low energy potential
- amenable to nanoscale 3D technologies

*: This is just one of many options being explored today

Hyperdimensional Vectors – the Concept

Probability that two N bit vectors, randomly chosen, are different by at least N/2 bits (or normalized Hamming Distance >= 0.5)

n	Norm. Hamming Distance
10	>= 0.400 for 82.8125%
100	>= 0.450 for 86.4373%
1000	>= 0.453 for 99.8671%
10000	>= 0.485 for 99.8694%



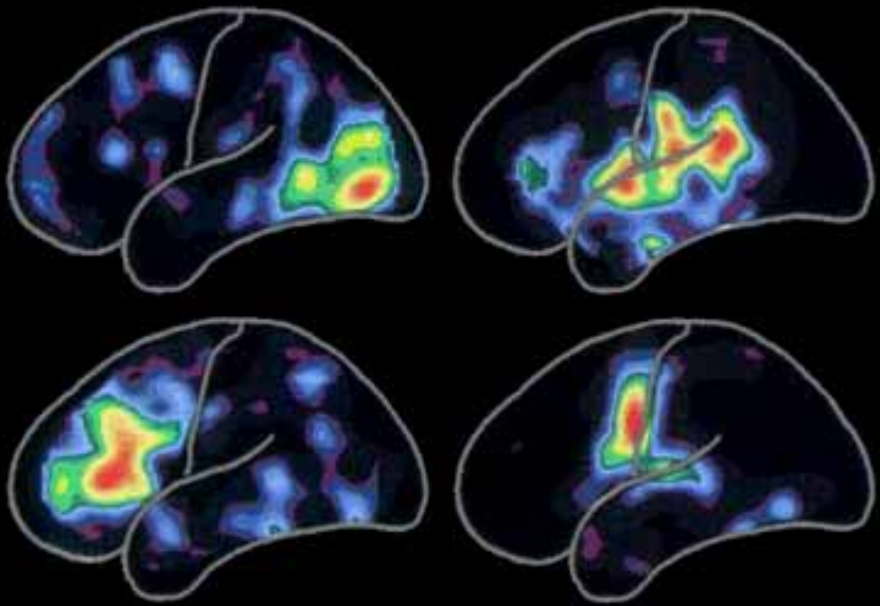
Distance histogram of vectors in N-dimensional space

99.9% probability that normalized Hamming distance between two randomly chosen vectors of length of 10k bits (0's and 1's) > 0.485!

Hyper-Dimensional Computing (HDC)

- Hyperdimensional vectors (N > 10000) as basic computational symbols
  - represent patterns rather then numbers
  - Can be *approximate – that is,* can be compared for similarity

- Mathematical properties of high-dimensional spaces in remarkable agreement with behaviors observed in brain



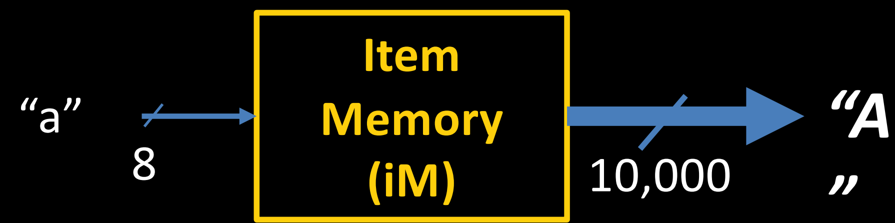
[Ref: P. Kanerva, An Introduction to Computing in Distributed Representation with High-Dimensional Random Vectors, 2009]

Example: Text Processing

(language recognition, text classification)

- Each symbol (letter) is represented by 10,000-D hypervector chosen at *random:*

A = -1 +1 -1 -1 -1 +1 -1 -1 ...
B = +1 -1+1 +1+1 -1 +1 -1 ...
C = -1 -1 -1 +1 +1 -1 +1 -1 ...
D = -1 -1 -1 +1 +1 -1 +1 -1 ...
...
Z = -1 -1 +1 -1 +1 +1 +1 -1 ...



- Every letter hypervector is dissimilar to others, e.g., **⟨*A*, *B*⟩ = 0**
- This assignment is fixed throughout computation

Computing with Patterns (the HD algebra)

- Addition (**+**) is good for representing sets (bundling), since sum vector is similar to its constituent vectors.
  - $\langle A+B, A\rangle = 0.5$
- Multiplication(*****) is good for binding, since product vector is dissimilar to its constituent vectors.
  - $\langle A*B, A\rangle = 0$
- Permutation (**ρ**) makes a dissimilar vector by rotating, and is good for representing sequences.
  - $\langle A, \rho A\rangle = 0$

  NOTE: ***** and **ρ** are invertible and preserve the distance

- Distance Measurement (<>) computes distance between 2 vectors
  - Parallel search for closest match performed in Associative Memory

Computing a Profile Using HD Arithmetic

- Trigram (3-letter sequence) : HD vector computed from its *Letter Vectors* with permutation (cyclic shift) and pointwise multiplication.

example: "EAT" versus "ATE"

E =	**-1**	**-1**	**-1**	**+1**	**+1**	**-1**	**+1**	**-1**	**…**		
A =		-1	+1	-1	-1	-1	+1	-1	-1	…	
T =			-1	-1	+1	-1	+1	+1	+1	-1	…
E A T =			+1	+1	-1	-1	+1	+1	…		

New HD vector representing sequence of characters

A =	**-1**	**+1**	**-1**	**-1**	**-1**	**+1**	**-1**	**-1**	**…**		
T =		-1	-1	+1	-1	+1	+1	+1	-1	…	
E =			-1	-1	-1	+1	+1	-1	+1	-1	…
A T E =			-1	+1	-1	+1	-1	+1	…		

HD vector of complete text:
Sum of trigrams created by sliding window

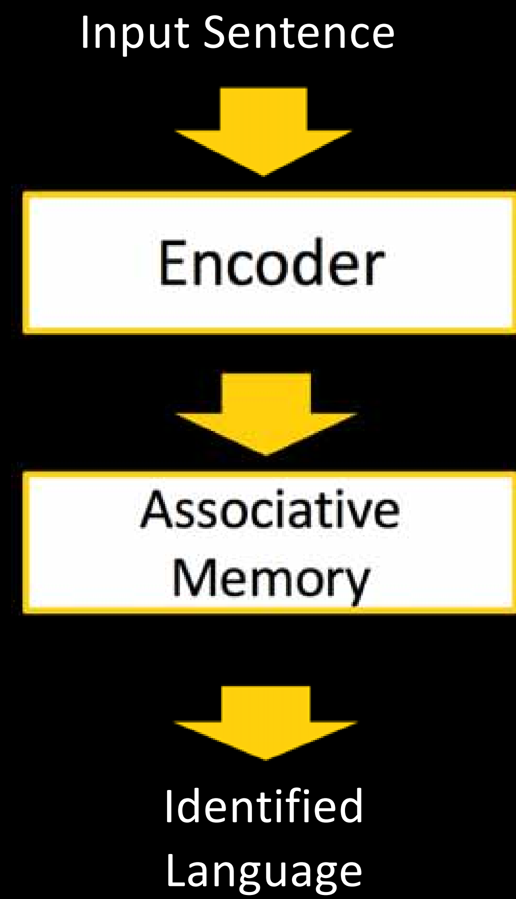
Example: Identifying Languages

LEARNING

21 languages
1000 sentences/language
Letters only

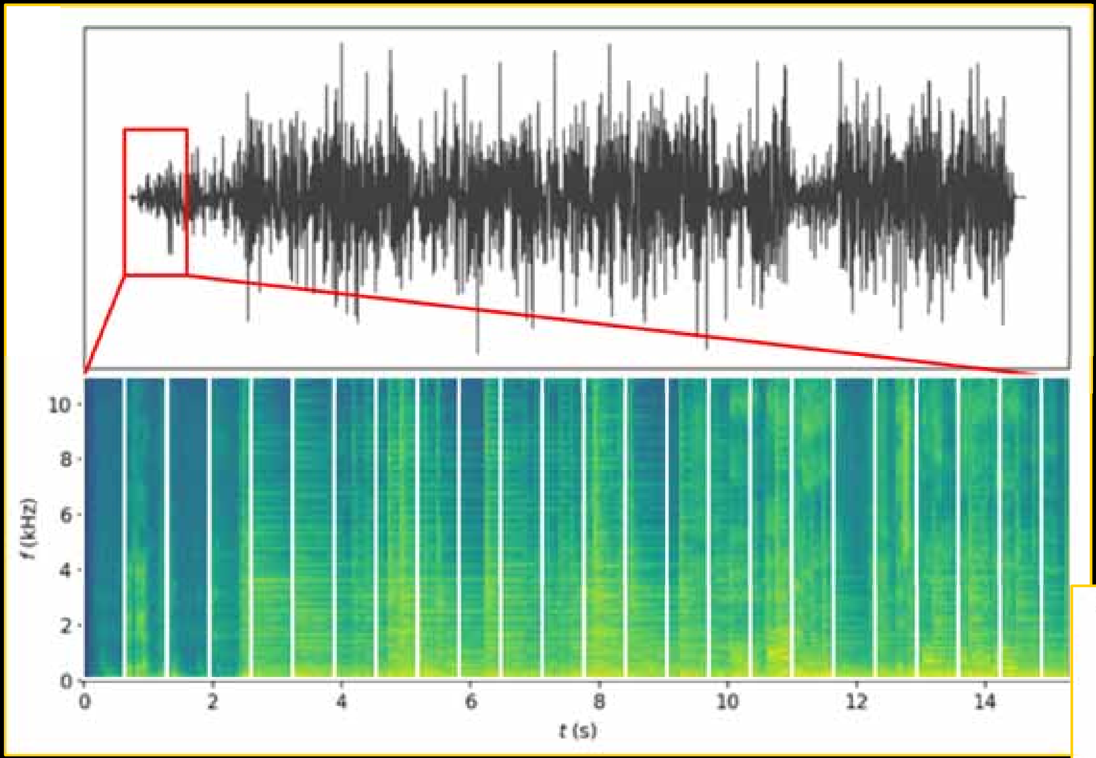
21 10,000D vectors
Stored in Associative Memory

PROCESSING



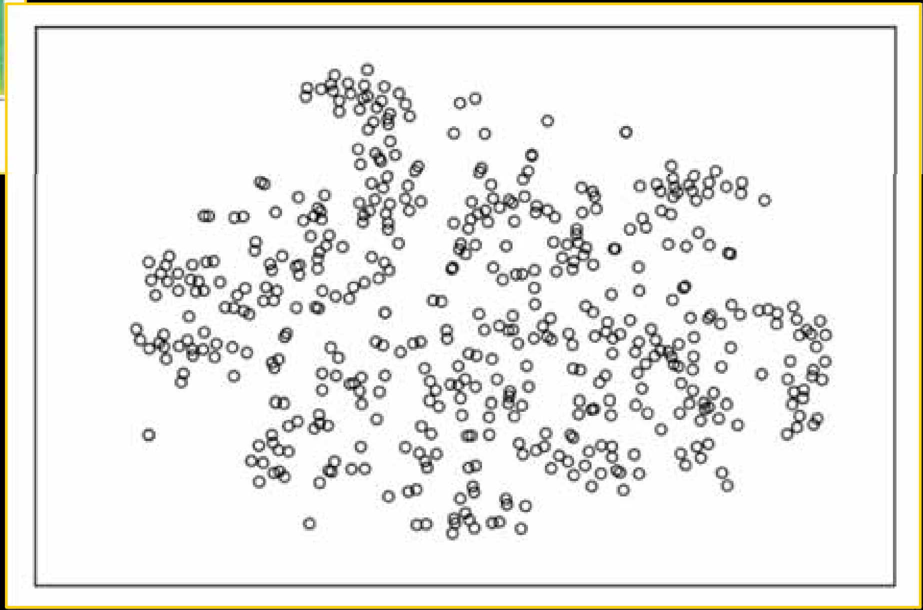
Applications	N-grams	HD	Baseline
Language identification	N=3	96.7	97.9
Text categorization	N=5	94.2	86.4

HD applied to streaming signals



Example: Music signal

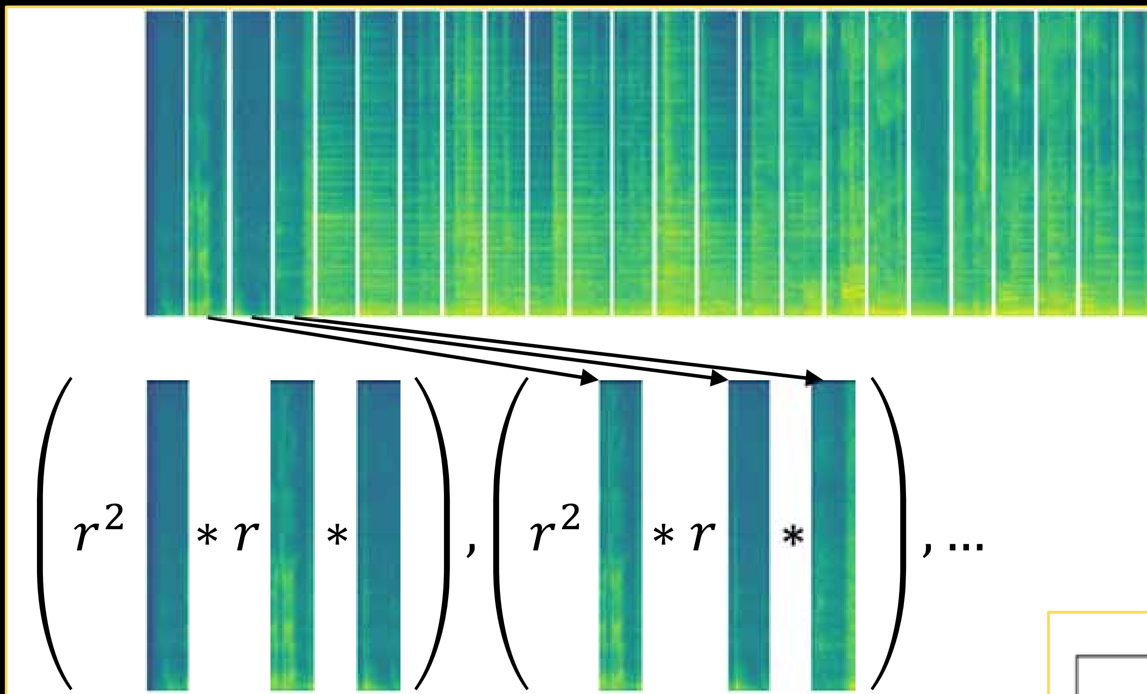
Mapped in HD space:
no visible structure



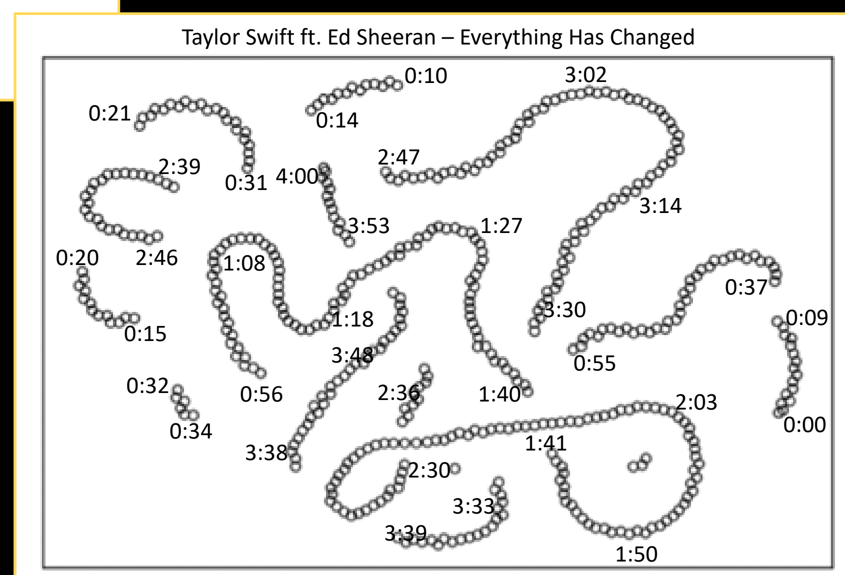
Mapping from high-dimensional to 2-dimensional space
Using t-SNE algorithm

[Courtesy: Justin Wong]

Imposing structure with HD



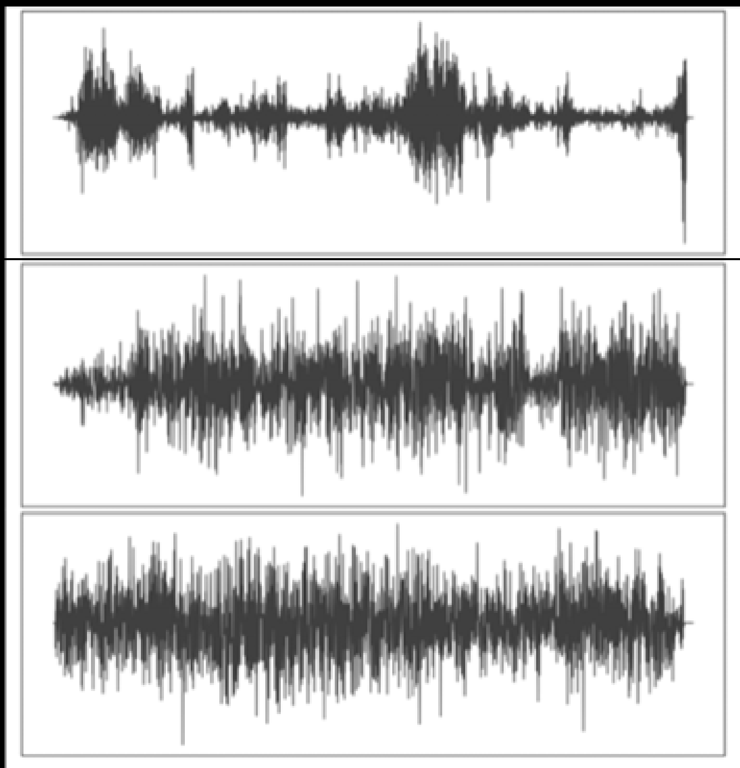
Sliding trigram representation reveals underlying structure



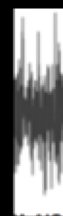
HD song map

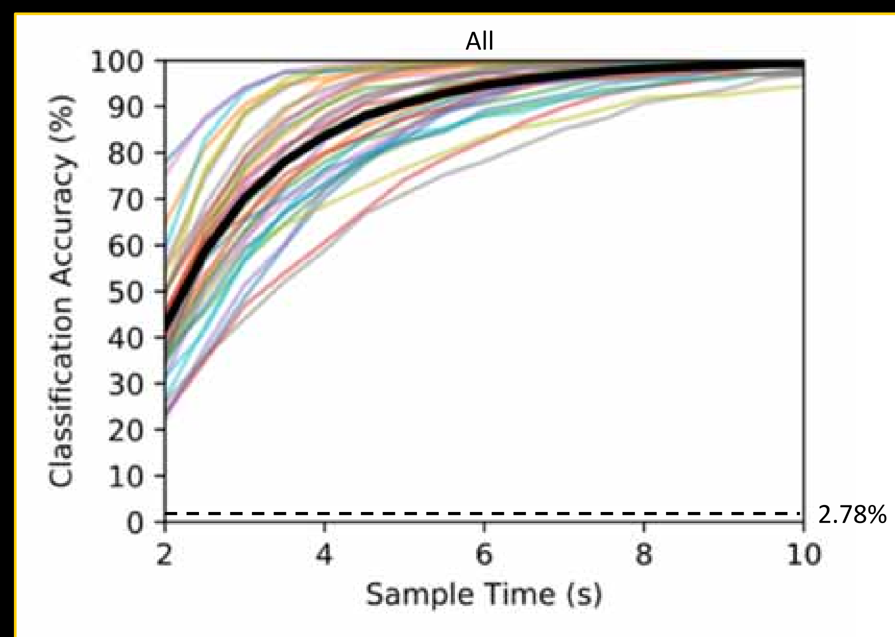
[Courtesy: Justin Wong]

Classifying Music with HD



In which song does segment occur?

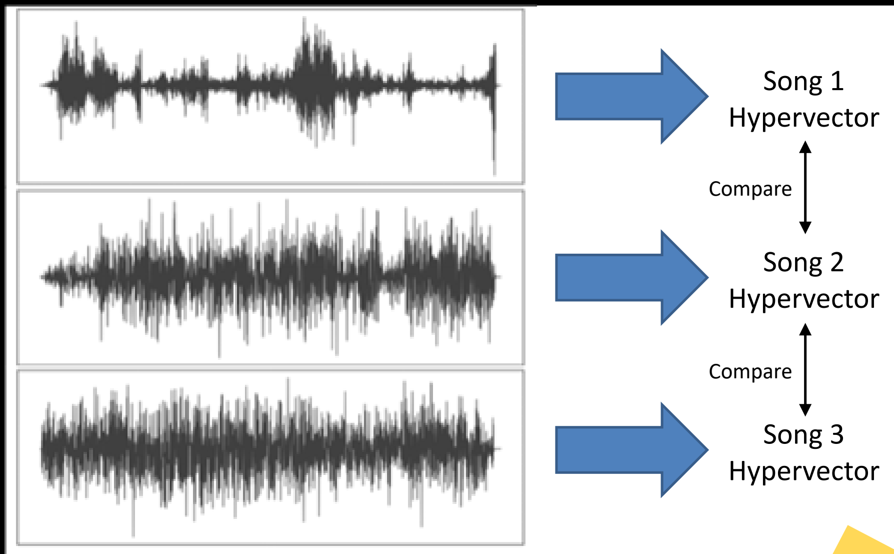




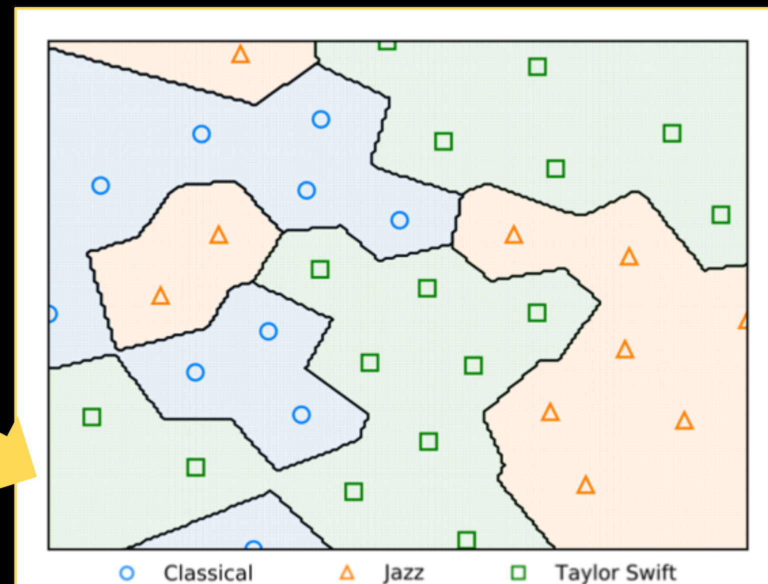
Hypervector size (per song): 86 kB
Compression rate between 65 and 335

[Courtesy: Justin Wong]

HD provides insights



For music classification

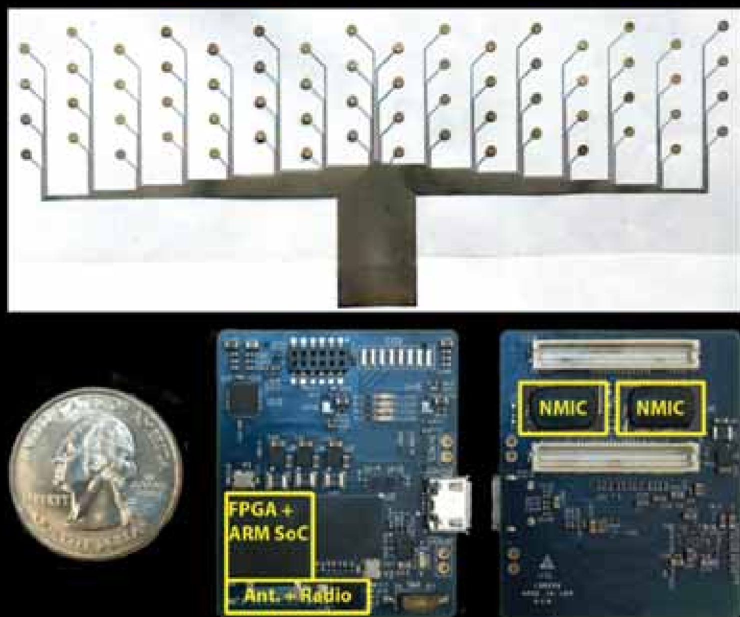


Or languages …

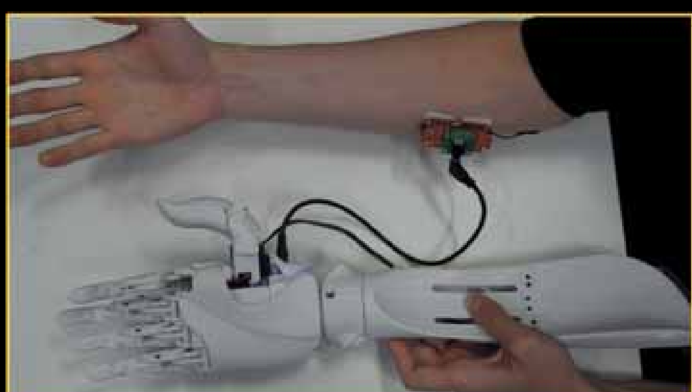
[Courtesy: J. Wong and P. Kanerva]

Example: Space-Time Sequence Classification

Electromyography (EMG) for gesture recognition and prosthesis control



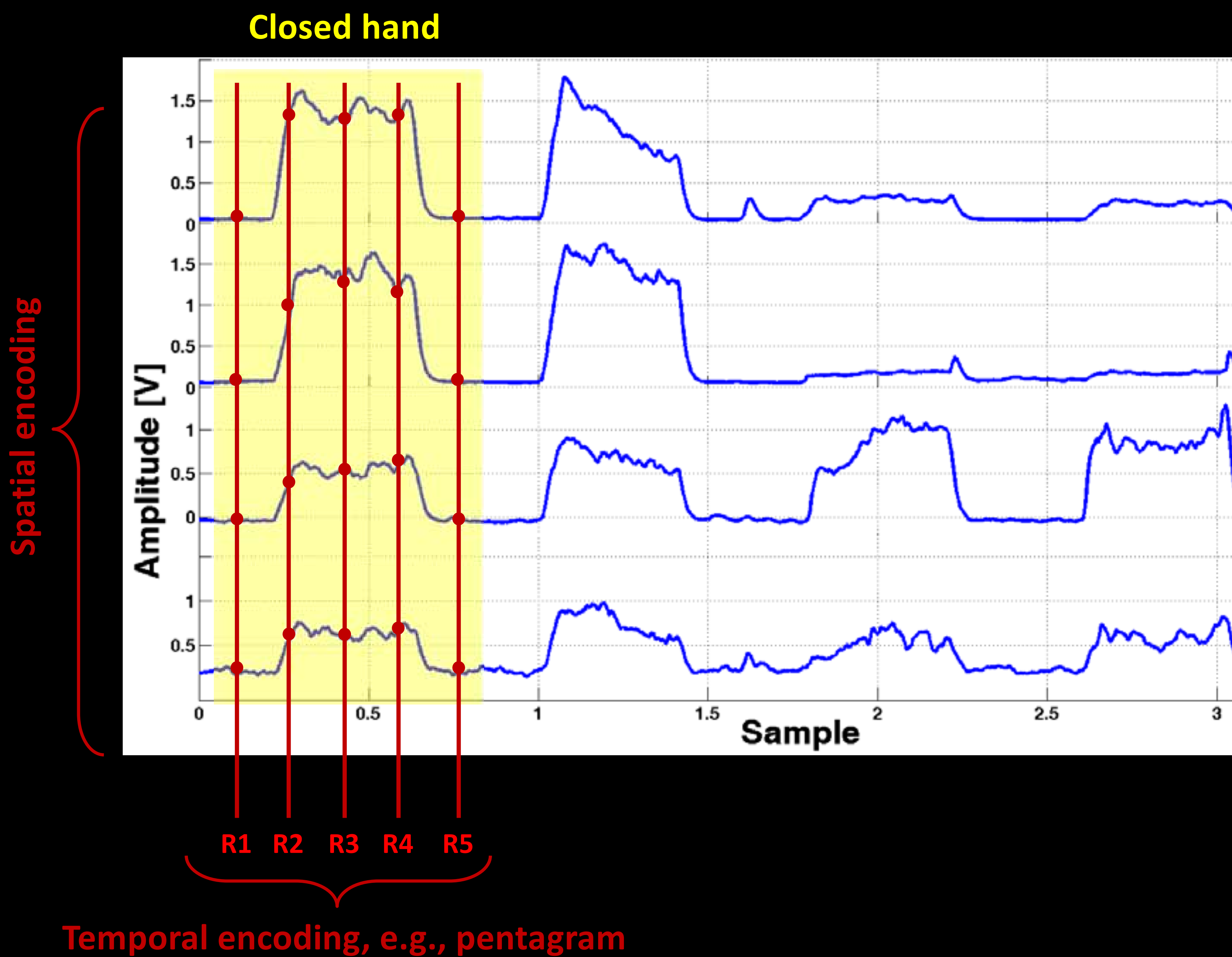
[Berkeley FlexEMG]



[Hackberry Hand]

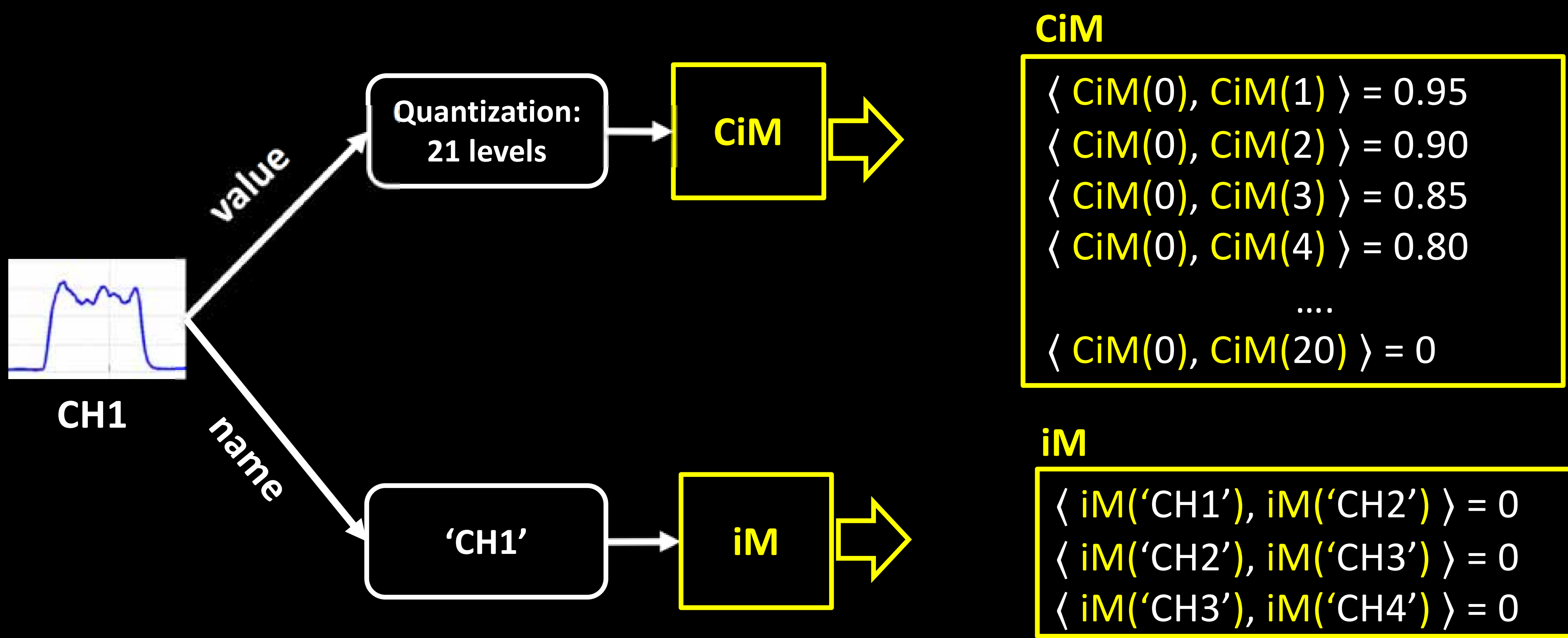
Redundancy in acquisition array (64 electrodes) provides robustness wrt variations (movement, long term wear, day-to-day, …)

Signal Partitioning for Encoding



Mapping to HD Space

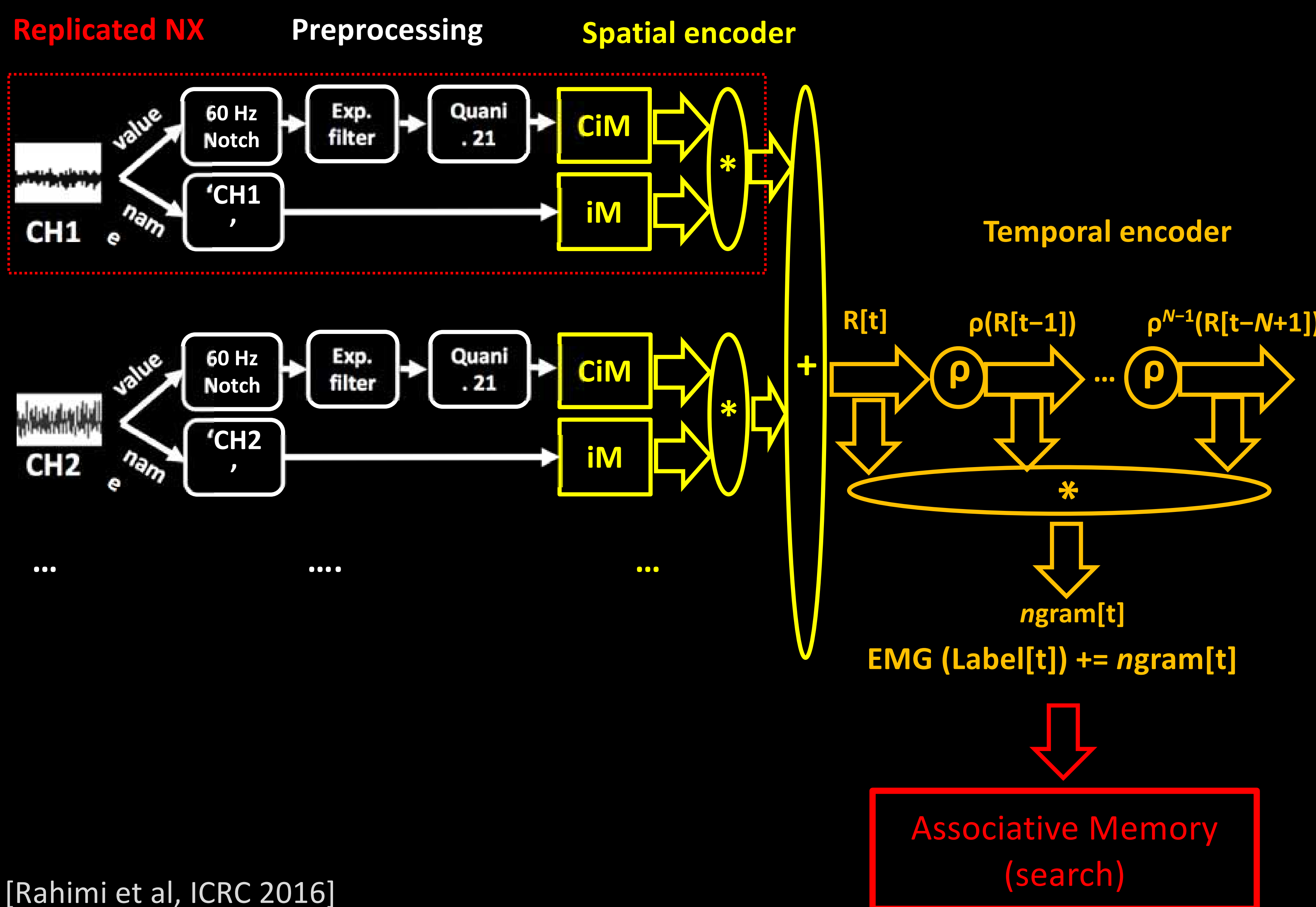
- **Item Memory (iM)** maps channels to orthogonal hypervectors.
- **Continuous iM (CiM)** maps quantities *continuously* to hypervectors.



- **Yellow** color codes components operating with HD distributed representation.

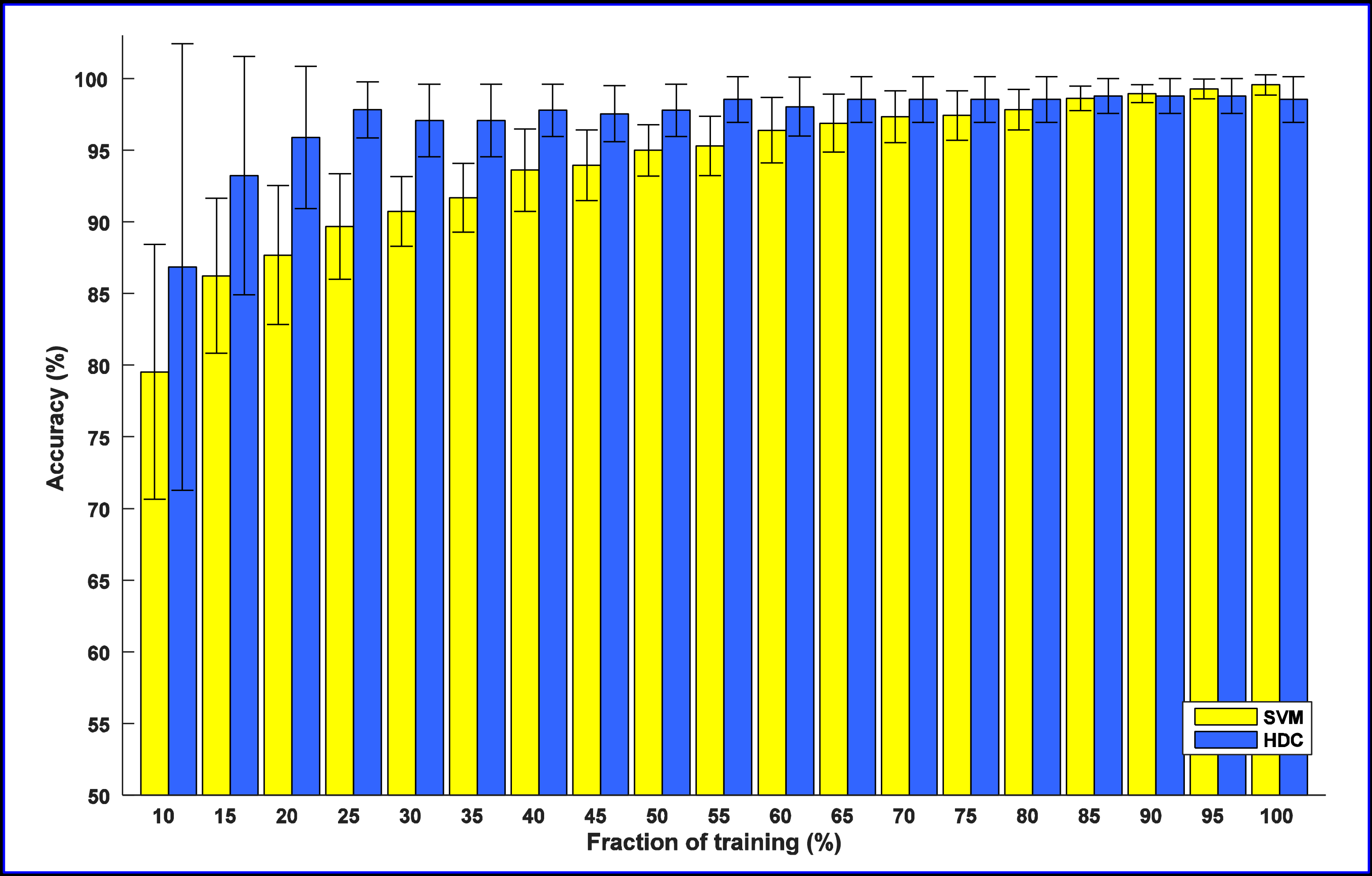
[Rahimi et al, ICRC 2016]

EMG Spatiotemporal Encoder



[Rahimi et al, ICRC 2016]

HDC Learns Fast!



HDC achieves a high level of classification accuracy (**97.8%**) with only **1/3** the training data required by state-of-the-art SVM

EMG Gesture Classification – Learning Speed and Robustness

2 subjects, 5 sessions

	Session 1	Session 2	Session 3	Session 4	Session 5	Overall
Subject 1	95.086%	90.551%	96.756%	94.793%	97.065%	94.850%
Subject 2	95.790%	96.489%	95.283%	99.251%	99.180%	97.199%
Average	95.438%	93.520%	96.020%	97.022%	98.123%	96.024%

Train and test in same session

True One-Shot Learning!

Context Change	Arm position	Wear session and day	Prolonged wear
Subject 1	65.757%	64.170%	87.292%
Subject 2	82.891%	93.126%	99.296%
Average	74.324%	78.648%	93.294%

Train in one session and use in other contexts

Context Change	Arm position		Wear session and day		Prolonged wear	
	New context	Old context	New context	Old context	New context	Old context
Subject 1	91.297%	91.735%	89.147%	87.909%	96.891%	94.852%
Subject 2	93.505%	96.195%	99.122%	98.079%	99.655%	99.381%
Average	92.401%	93.965%	94.135%	92.994%	98.273%	97.117%

Trained data adjusted in new session

[A. Zhou et al, to be published in SRC TechCon Sept. 2018]

From Classification to Reasoning

A simple example

What is the Dollar of Mexico?

Learning (HD database representation)

R1 = Country * USA + MoneyUnit * Dollar + Population * 320M + …
R2 = Country * Mexico + MoneyUnit * Peso + Population * 120M + …
…

Data stored in superposition

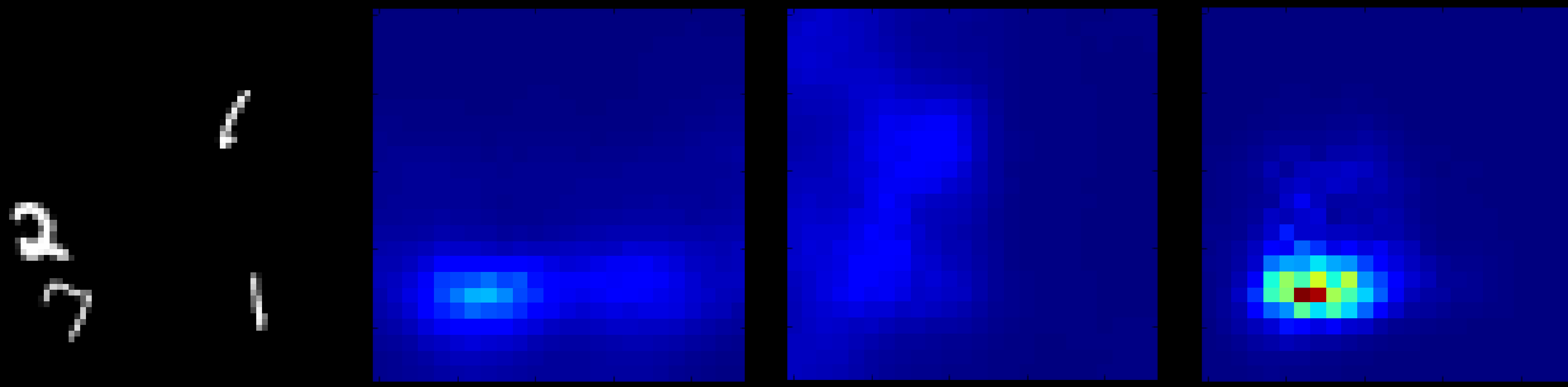
Queries

<R2 * Country> = Mexico
<R2 * <R1 * Dollar> > = Peso
…

- Yellow color codes components operating with HD distributed representation.
- <> stand for associative match

Scene Analysis and Spatial Reasoning

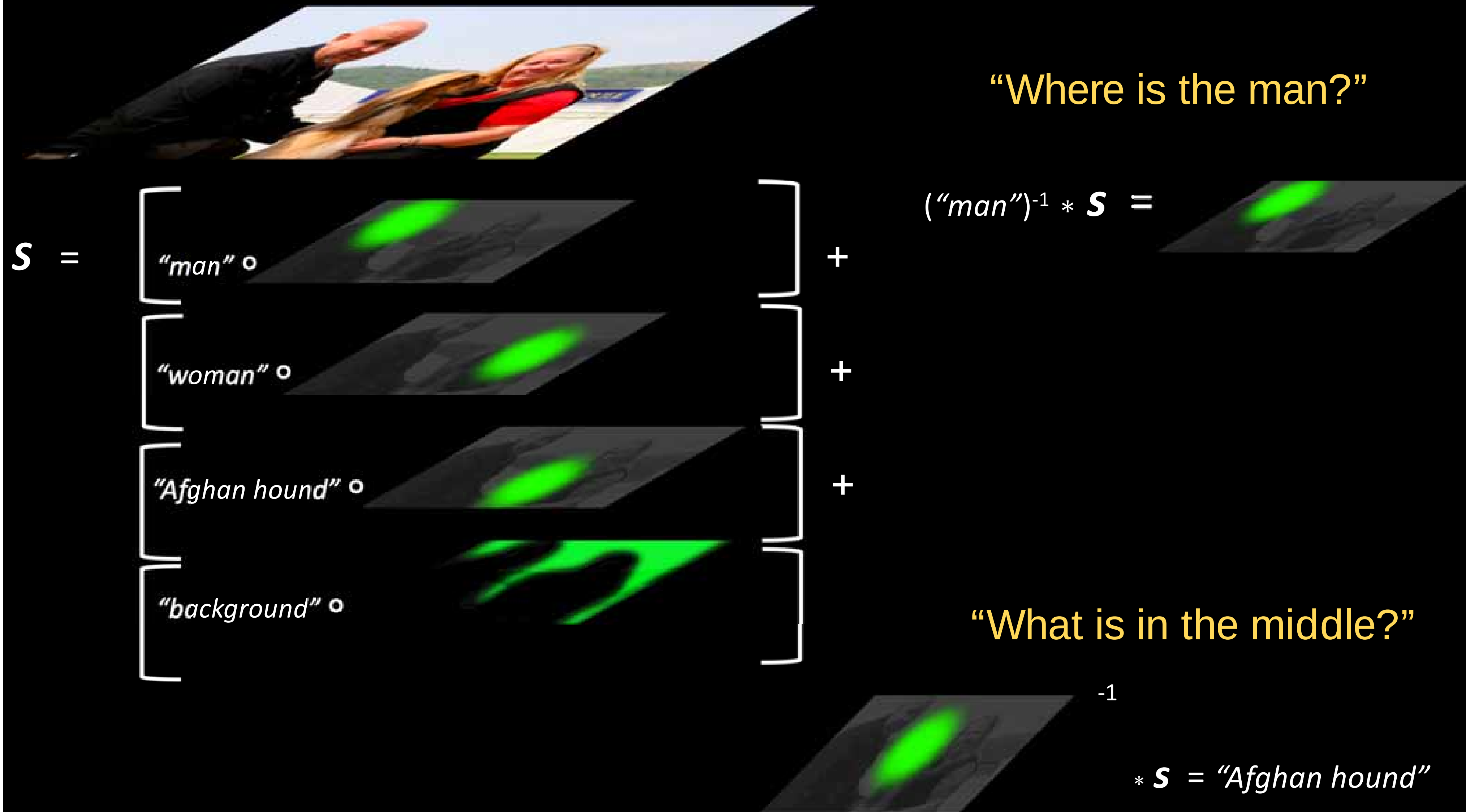
- Each object in a scene (e.g. obtained using DNN): random HD vector
- Object location: random HD vector
- Scene vector: superposition of bindings of object and location vectors.
- Query vector is formed **using same algebra**, and operates on scene vector via multiplication.



Response to the query “What is below a 2 and to the left of a 1?”

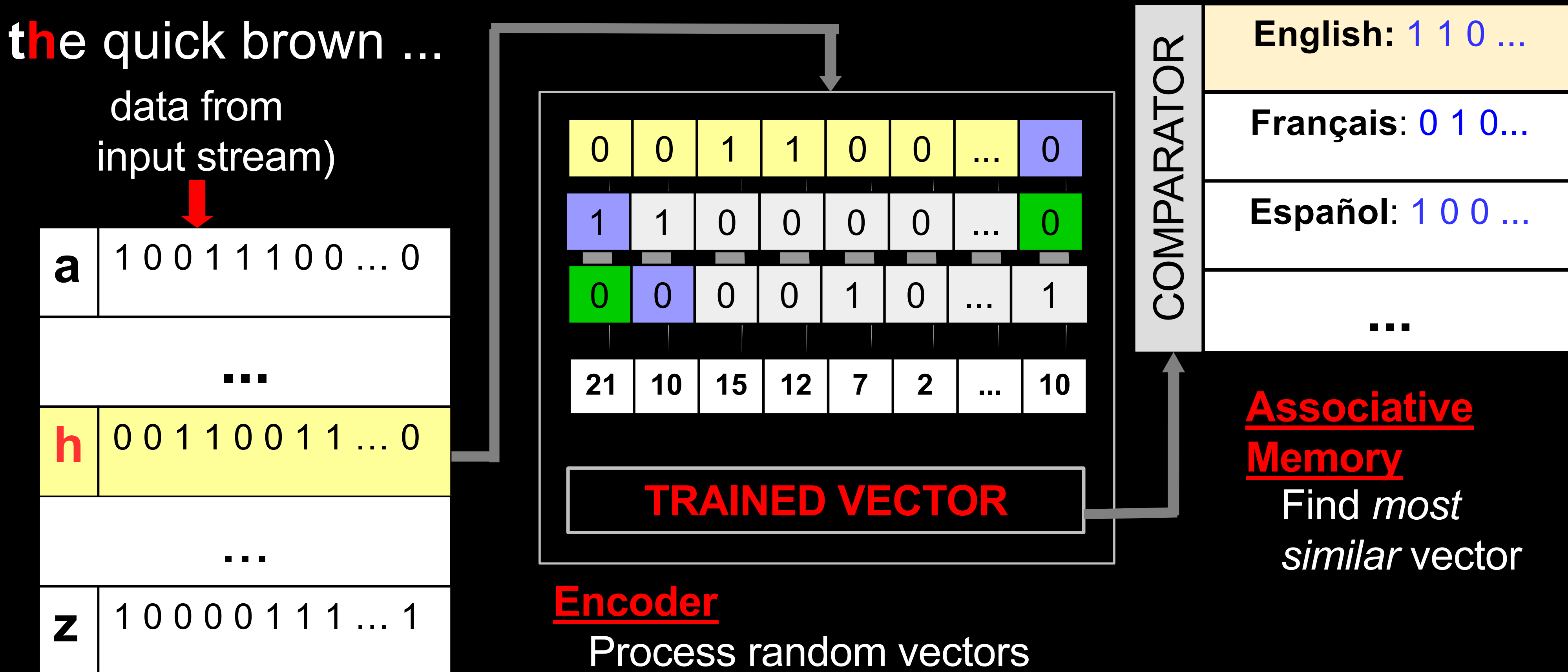
[Courtesy: B. Olshausen and Eric Weiss]

Scene Analysis and Spatial Reasoning



[Courtesy: B. Olshausen and Eric Weiss]

Building a generic HD Processor



Item Memory

Store samples of random HD vectors, the alphabet

Encoder

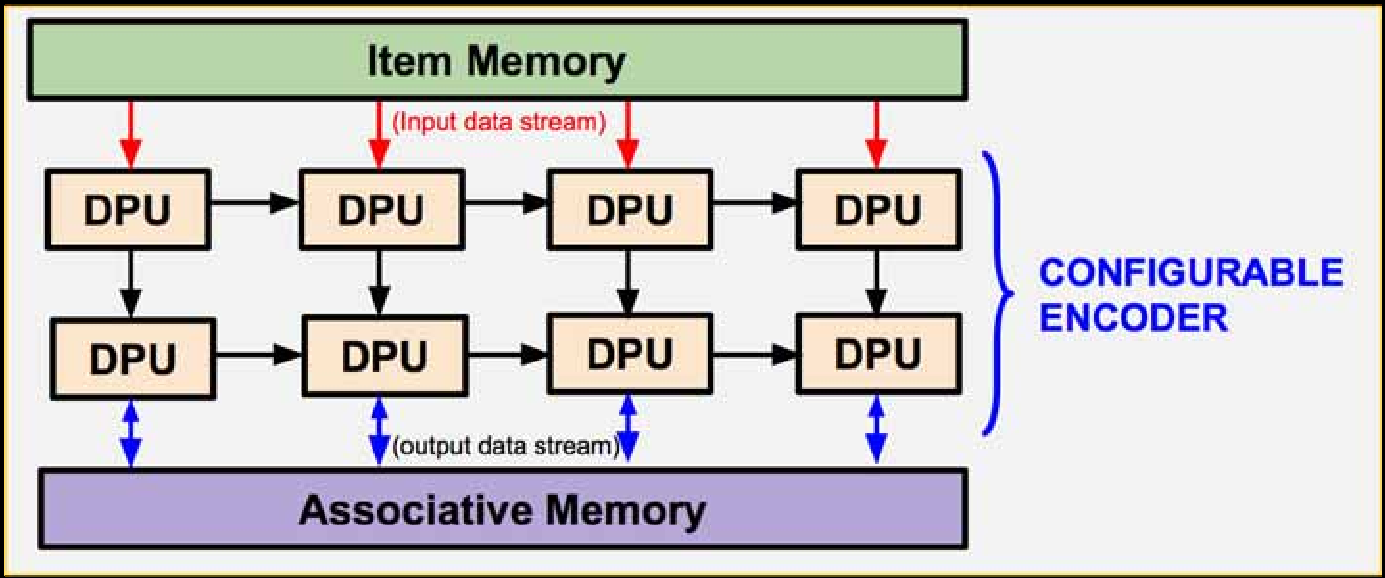
Process random vectors

Configurable to support broad application range

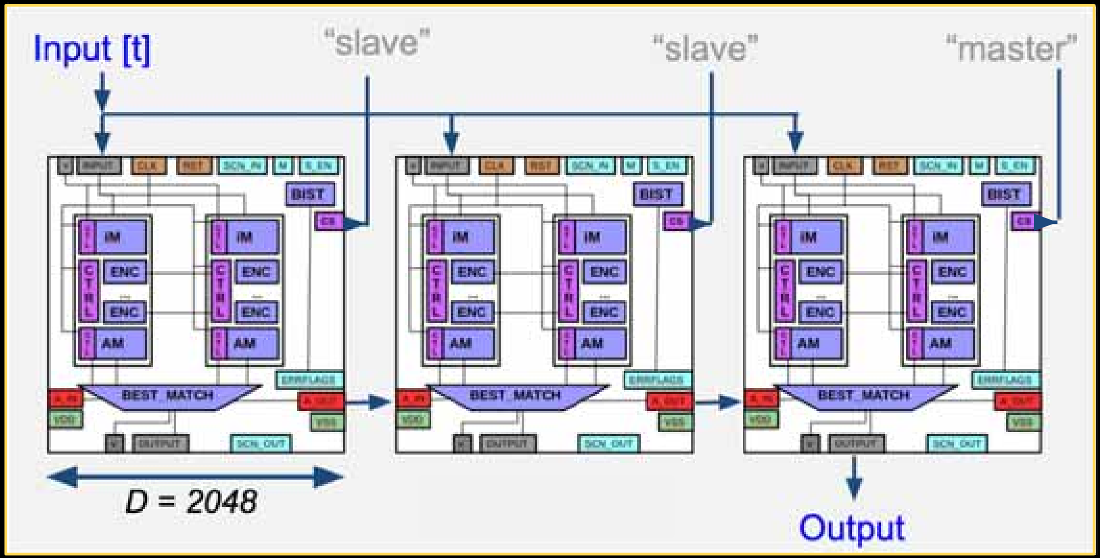
Associative Memory

Find *most similar* vector

A generic HD Processor



- Pipelined array architecture
- Regular, simple DPU network
- No working memory
- Short algorithms



- Scalable in HD dimension

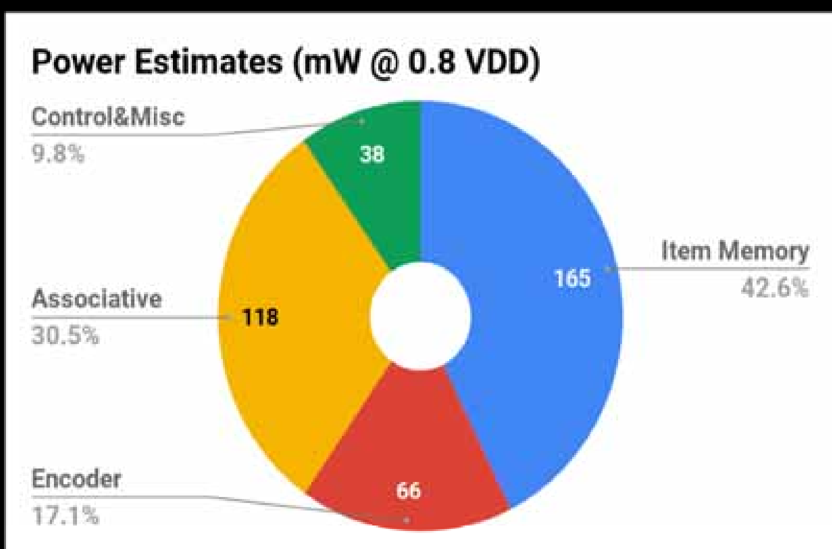
Many ways of using parallelism or to re-use components

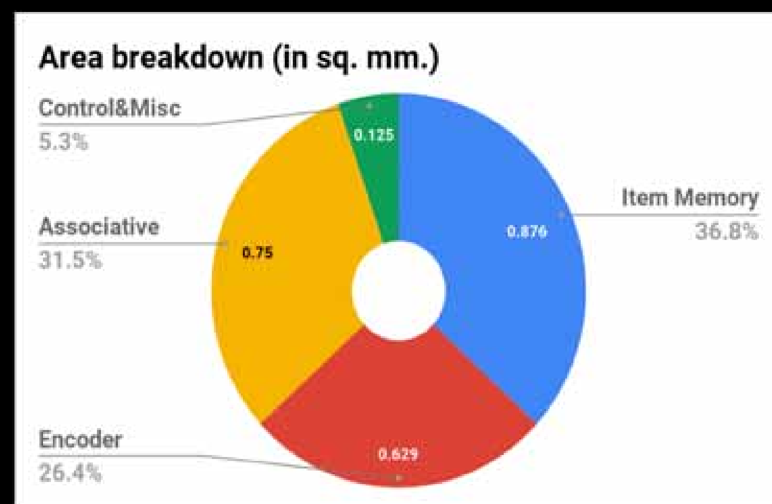
HD Processor Prototype: Applications and results

Applications	Encoding	HD	Known State-of-the-Art ML
Language Recognition	*n*-gram	95.8%	97.1%, n-gram k-Ne
FlexEMG Hand-Gesture Recog.	*n*-gram	**96.8%**	89.7%, Support ector Machine
DNA Sequencing	features	**97.1%**	93.7%, knowledge-based Neural Network
Fetal State Classif. (Cardio)	feature	**92.5%**	90.6%, Support Vector Machine
Page-Block Doc. Classification	features	**91.6%**	85.9%, min-max Hyperplane
UCI Human Activity Recogn.	features	82.5%	89.3%, Support Vector Machine
Spoken Letter Classification	features	80.1%	93.5%, Boosted k-Nearest Neighbors
Human Face Detection	features	84.1%	96.1%, HOG Boosted Decision Trees
MNIST Digit Classification	features	81.4%	99.7%, Convolutional Neural Network

Results based on TensorFlow HDC processor simulator (assume Cosine distance measure)

TSMC 28 HPM Pre-route Estimates:



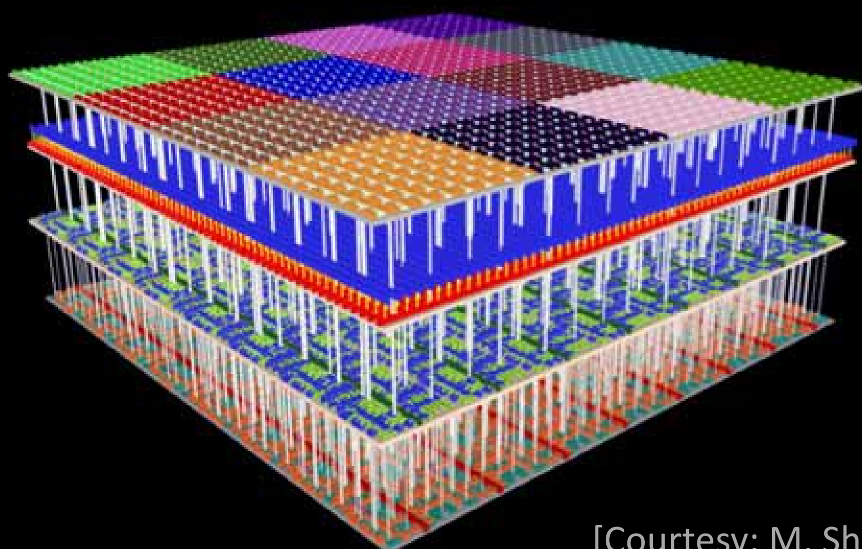


Between 1 and 10 µJ per classification

Note: this implementation does not use any low-energy circuit optimization

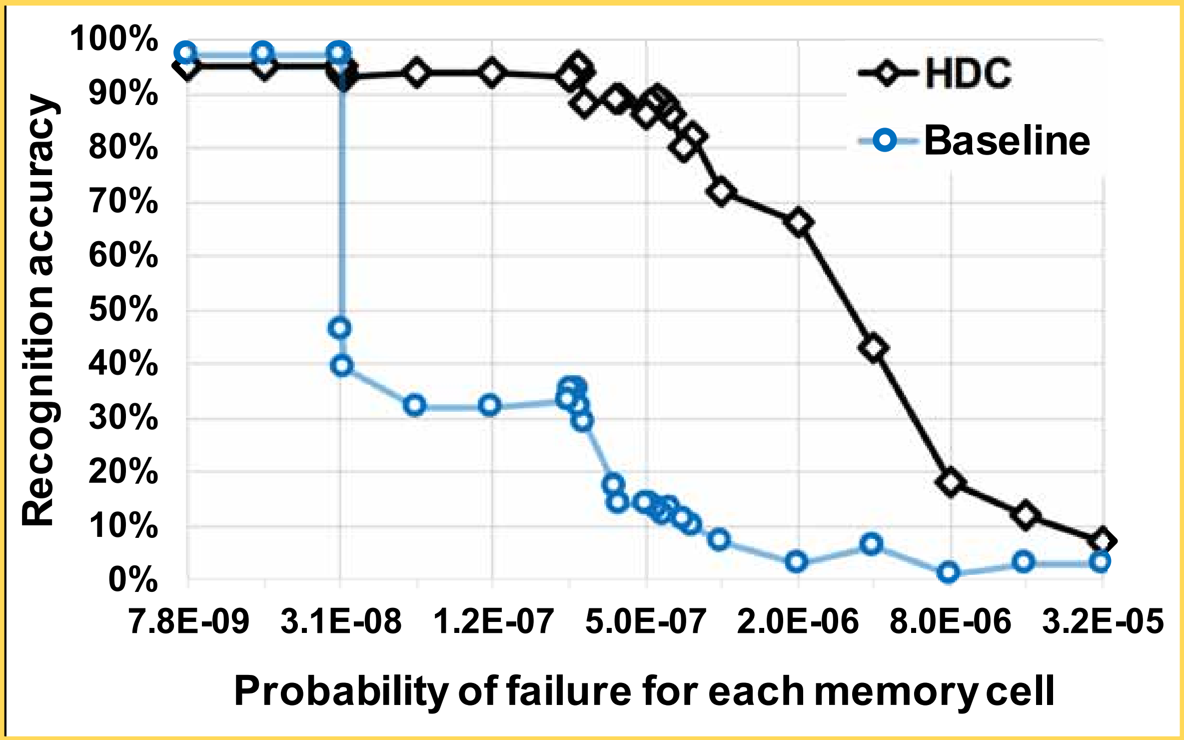
HDC maps well into 3D nanostructures

- Tight interweaving of memory and logic
  - True in-memory computing
- Approximate
  - Extremely robust against failures and errors
  - Allows for low SNR computing
- Scalable



[Courtesy: M. Shulaker]

Graceful degradation



Case study: Language recognition; Baseline: histograms

Near peek accuracy: HDC tolerates **8.8-fold** probability of failure compared to baseline

[Rahimi et al, ISLPED 2016]

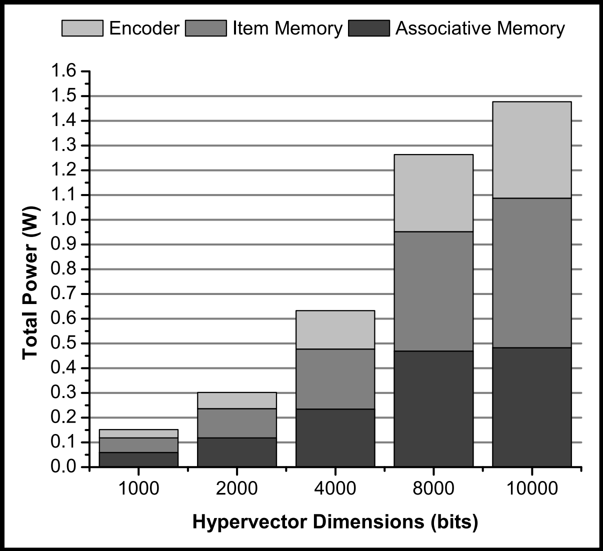
Associative Memory

$$Index = \arg\min_{i=1}^{C}(|s_i, In|)$$

$WL(In, s_i) = D$ (~ 10,000)
C: Number of learned items

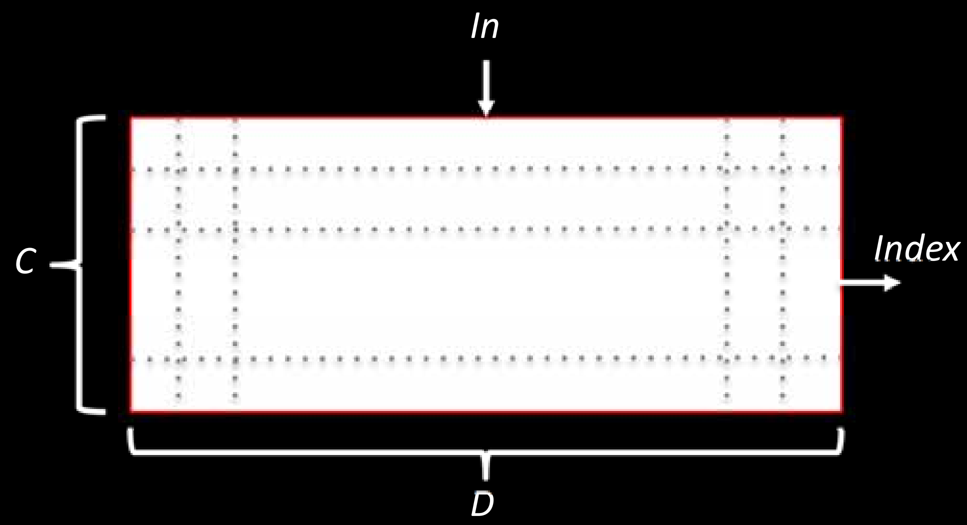
Note: Traditional assoc. memory

$$Index = \arg_{i=1}^{C}(s_i == In)$$



[Rahimi et al, TCAS-I, 2017]

Dominant contributor to power

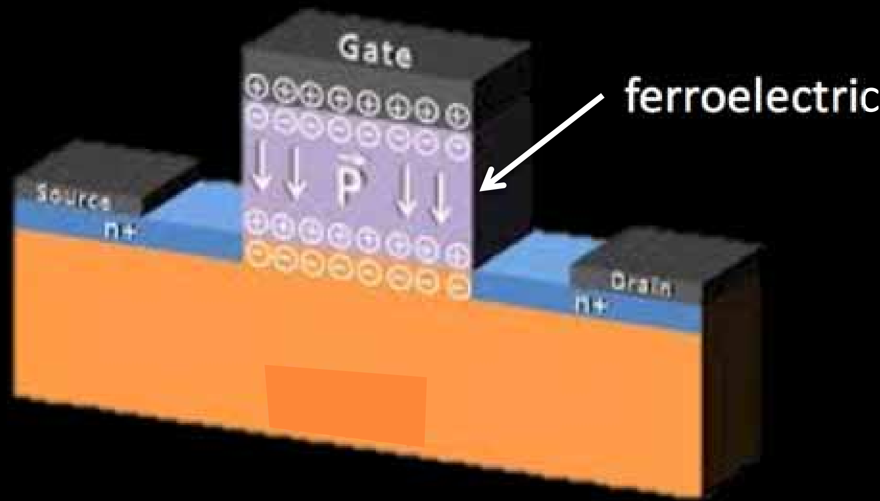


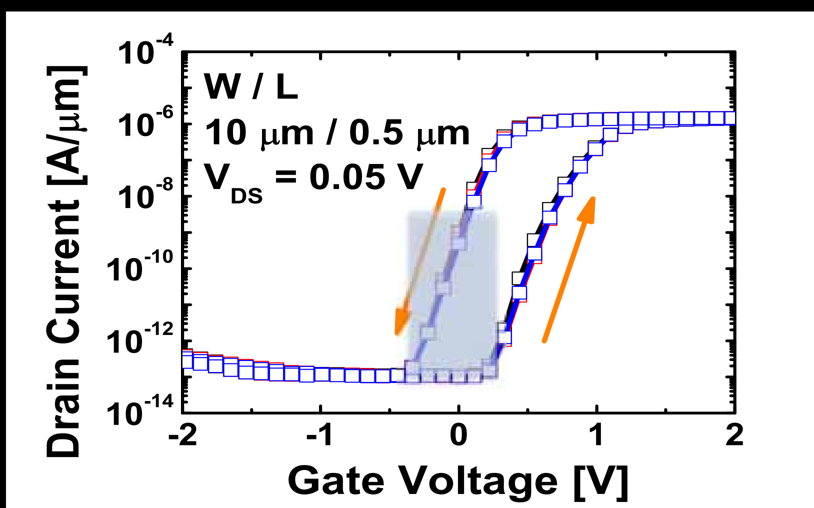
Large range of implementation options:

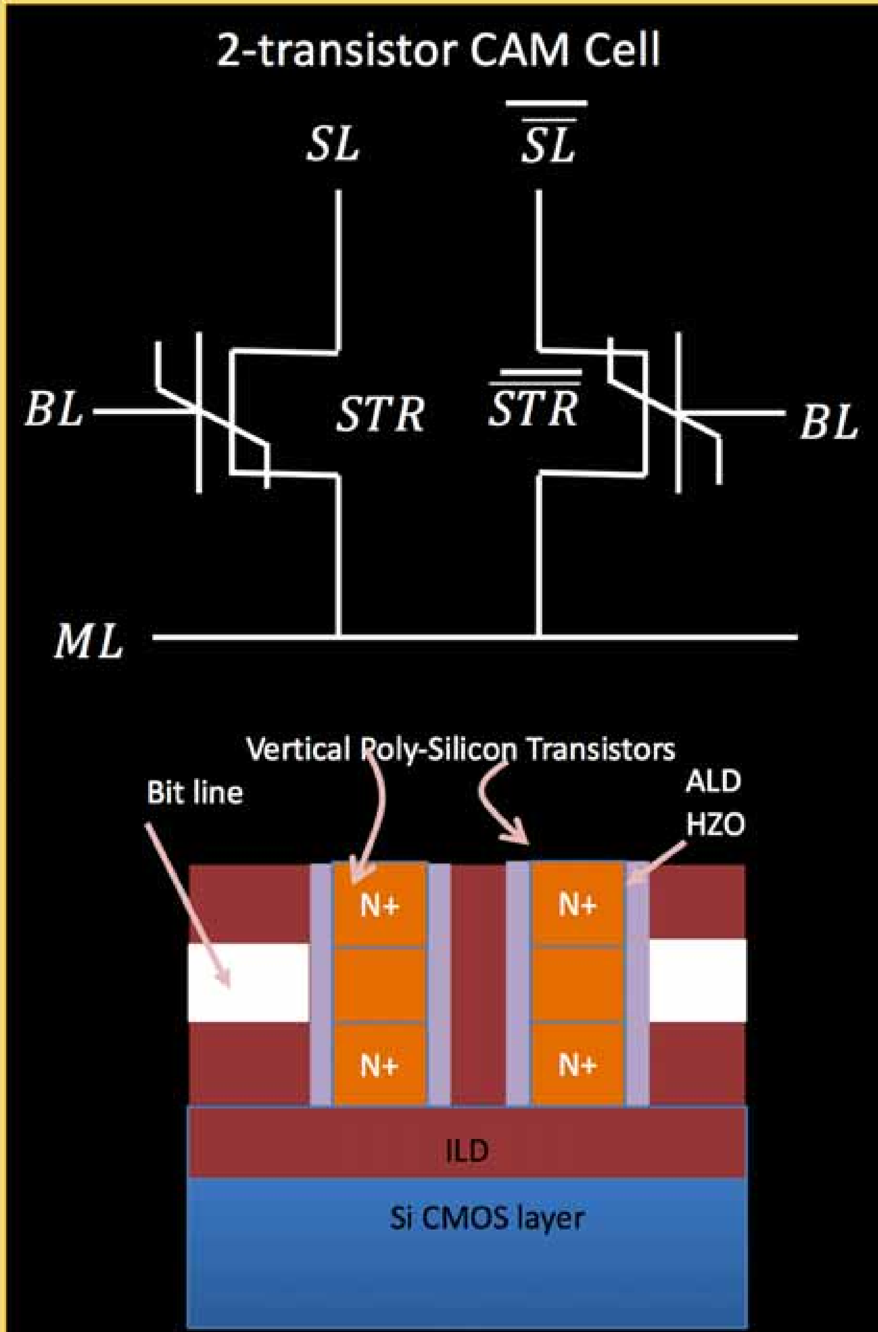
Volatile vs non-volatile
Digital, analog
Accuracy
Memory organization
Data representation

Similar function (but different organization): Sparse distributed memory (SDM)

Device opportunity: Ferroelectric AM







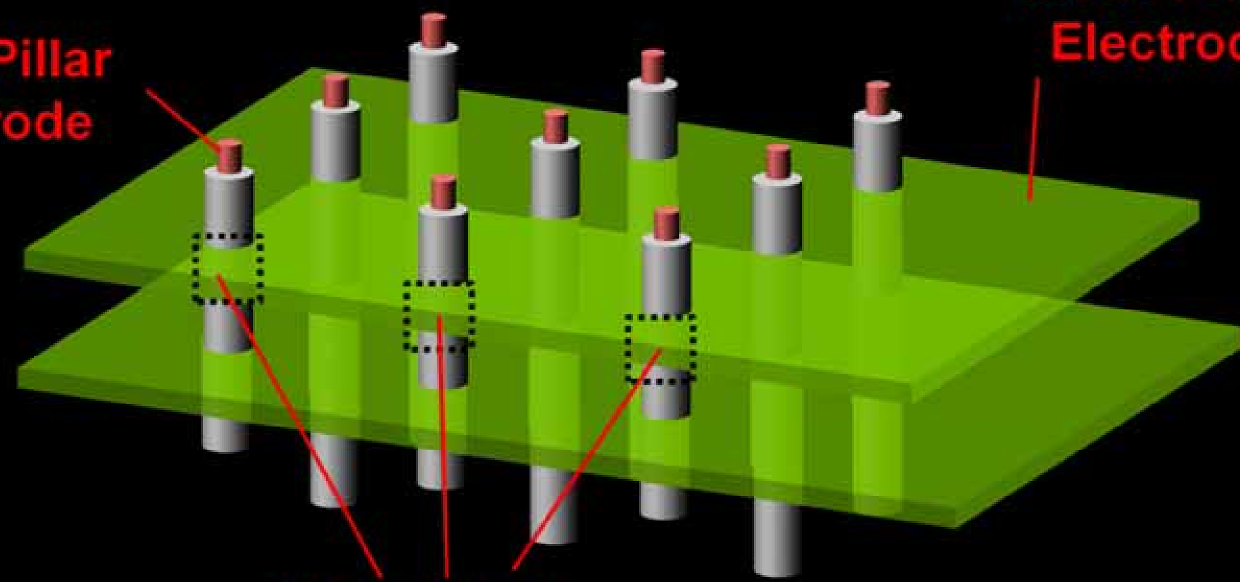
Schematic of a Vertical integrated FECAM cell

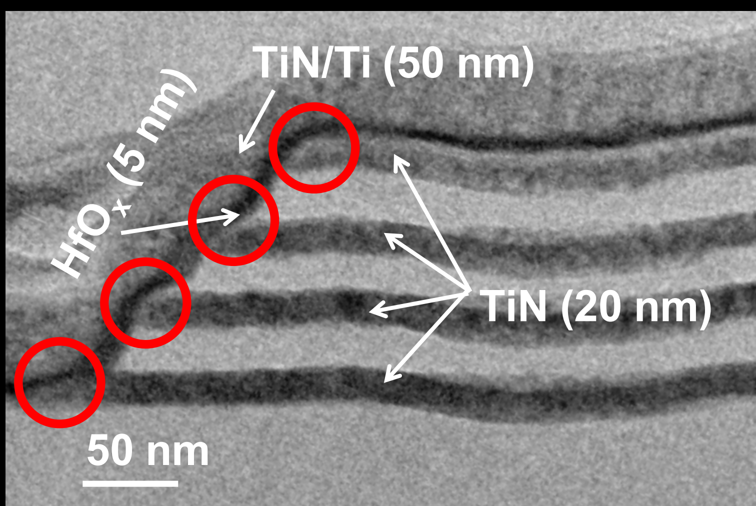
[Courtesy: Sayeef Salahuddin, UCB]

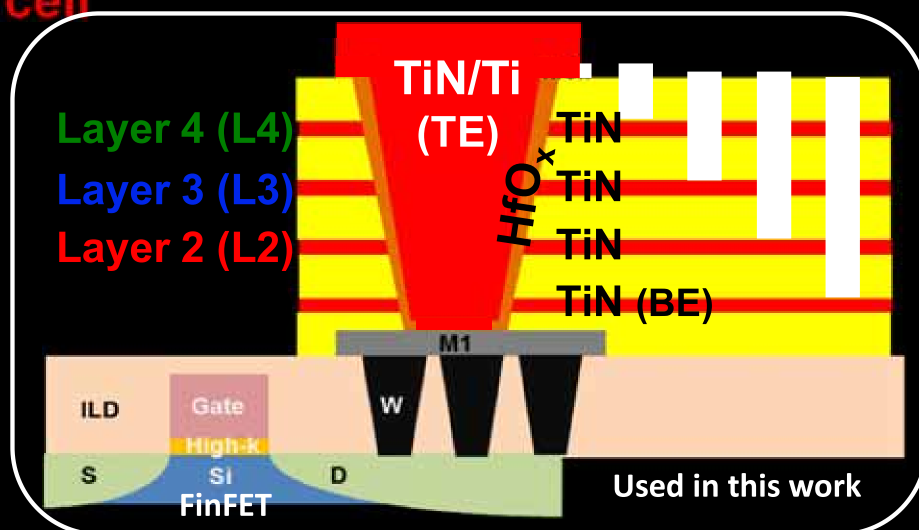
Process Opportunity:

3D integration of non-volatile memory (RRAM) and logic

VRRAM:
vertical
resistive
random
access
memory



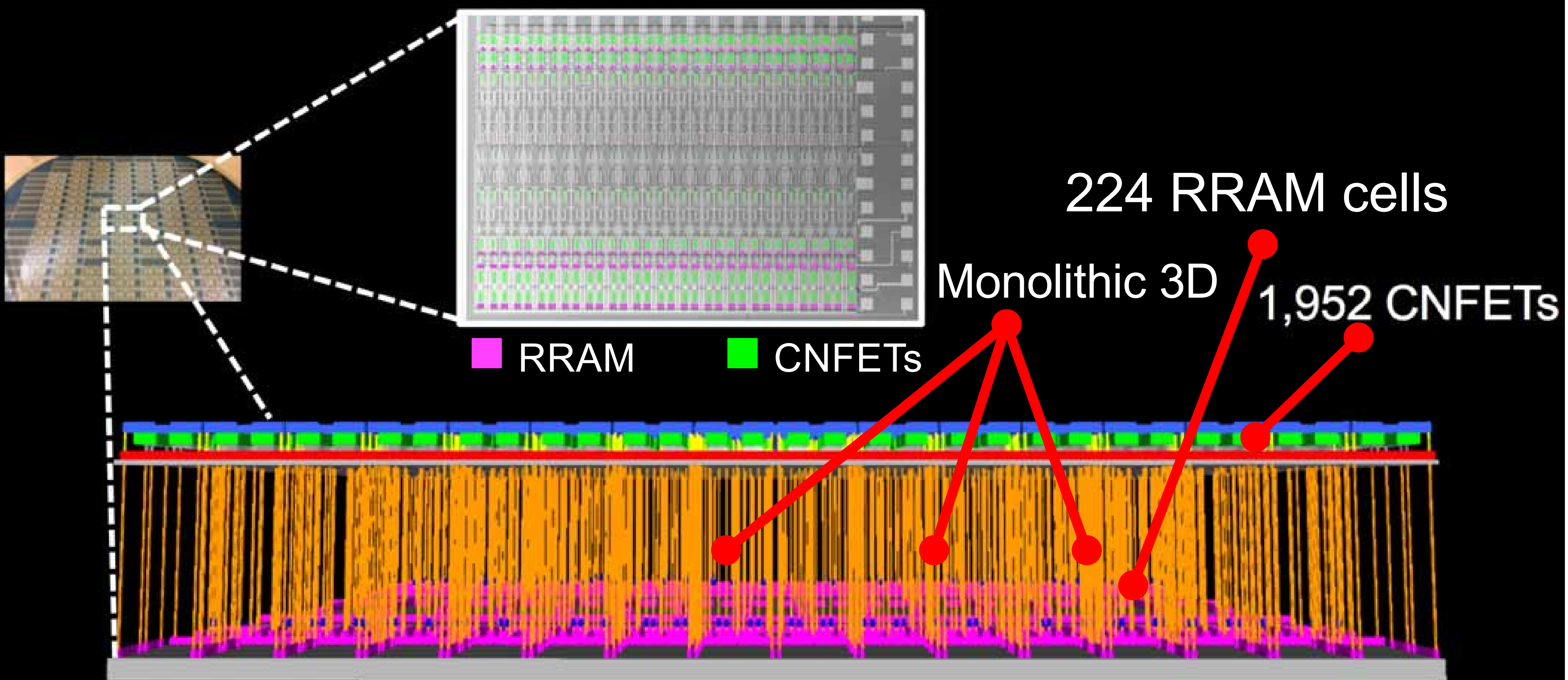




H.-Y. Chen,…, H.-S. P. Wong, ***IEDM***, 2012; H. Li, K.-S. Li,…, H.-S. P. Wong, ***Symp. VLSI Tech.***, 2016

3D Nanosystem for HD Computing

Monolithic 3D integration of logic and non-volatile memory



First 3D integrated HD Processor

[T. Wu et al, ISSCC 2018]

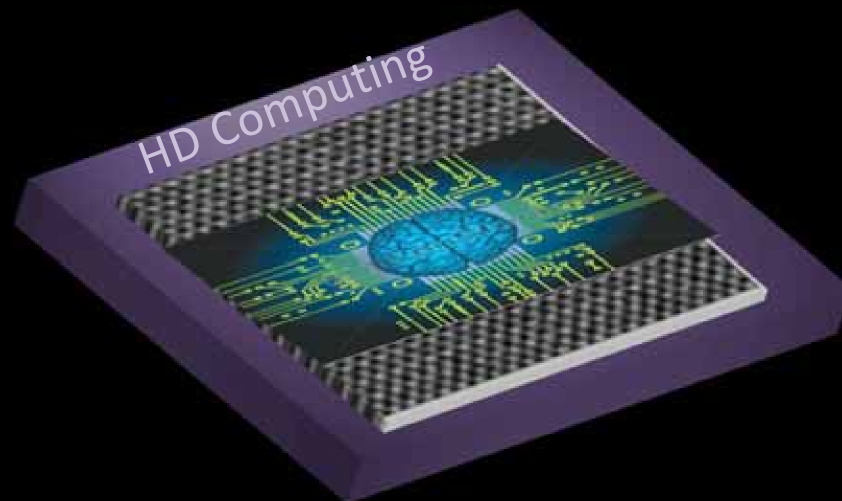
High-Order Bits

- Brain-inspired learning-based computational models as an exciting alternative to traditional algorithmic computing
- Especially for perceptive and cognitive functions
- HD offers exciting opportunity to bring learning-based functionality to low-power, small form-factor devices (smart world, smart human)
- Realizable in todays CMOS, but truly shines in 3D nanoscale technologies, integrating memory and logic

Acknowledgements

The many contributions of my students and many of my colleagues to this presentation are gratefully acknowledged.

The support of the the SRC/NSF Enigma Project, the StarNet SONIC Center, Intel Corp., and the member companies of BWRC and SwarmLab is greatly appreciated.



SWARM Lab
UC BERKELEY

BWRC

Relevant publications

- Abbas Rahimi, Sohum Datta, Denis Kleyko, Edward Paxon Frady, Bruno Olshausen, Pentti Kanerva, Jan M Rabaey, High-Dimensional Computing as a Nanoscalable Paradigm, in IEEE Transactions on Circuits and Systems I , Issue 99, June 2017.
- Abbas Rahimi, Pentti Kanerva, José del R Millán, Jan M Rabaey, Hyperdimensional computing for noninvasive brain–computer interfaces: Blind and one-shot classification of EEG error-related potentials, 10th ACM/EAI International Conference on Bio-inspired Information and Communications Technologies (BICT), 2017. (**Best paper award**)
- Mohsen Imani, Abbas Rahimi, Deqian Kong, Tajana Rosing, Jan M. Rabaey, Exploring hyperdimensional associative memory, in 2017 International Symposium on High Performance Computing Architecture (HPCA), pp. 445-456, Febr. 2017.
- Mohsen Imani, Abbas Rahimi, John Hwang, Tajana Rosing, Jan M. Rabaey, Low-Power Sparse Hyperdimensional Encoder for Language Recognition, in IEEE Design & Test, 2017, *In press*.
- Haitong Li, Tony F Wu, Abbas Rahimi, Kai-Shin Li, Miles Rusch, Chang-Hsien Lin, Juo-Luen Hsu, Mohamed M Sabry, S Burc Eryilmaz, Joon Sohn, Wen-Cheng Chiu, Min-Cheng Chen, Tsung-Ta Wu, Jia-Min Shieh, Wen-Kuan Yeh, Jan M Rabaey, Subhasish Mitra, H-S Philip Wong, Hyperdimensional computing with 3D VRRAM in-memory kernels: Device-architecture co-design for energy-efficient, error-resilient language recognition, 2016 IEEE International Electron Devices Meeting (IEDM), December 2016.
- Abbas Rahimi, Simone Benatti, Pentti Kanerva, Luca Benini, and Jan M. Rabaey, Hyperdimensional Biosignal Processing: A Case Study for EMG-based Hand Gesture Recognition, in IEEE International Conference on Rebooting Computing (ICRC), October 2016.
- Abbas Rahimi, Pentti Kanerva, and Jan M. Rabaey, A Robust and Energy-Efficient Classifier Using Brain Hyperdimensional Computing, in ACM/IEEE International Symposium on Low-Power Electronics and Design (ISLPED), August 2016.
- P. Kanerva. Hyperdimensional computing: An introduction to computing in distributed representation with high-dimensional random vectors. Cognitive Computation, 1(2):139–159, 2009.