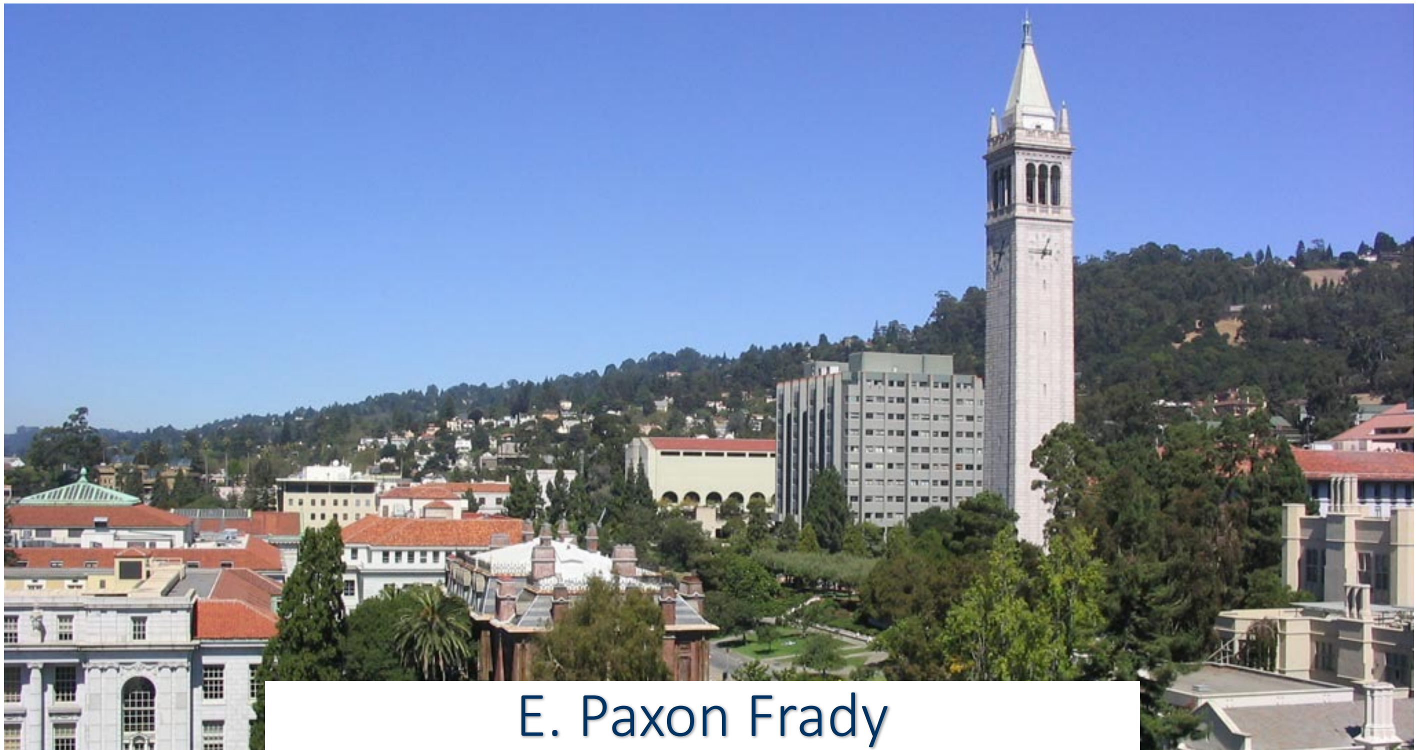
Solving factorization problems with VSAs and resonator networks



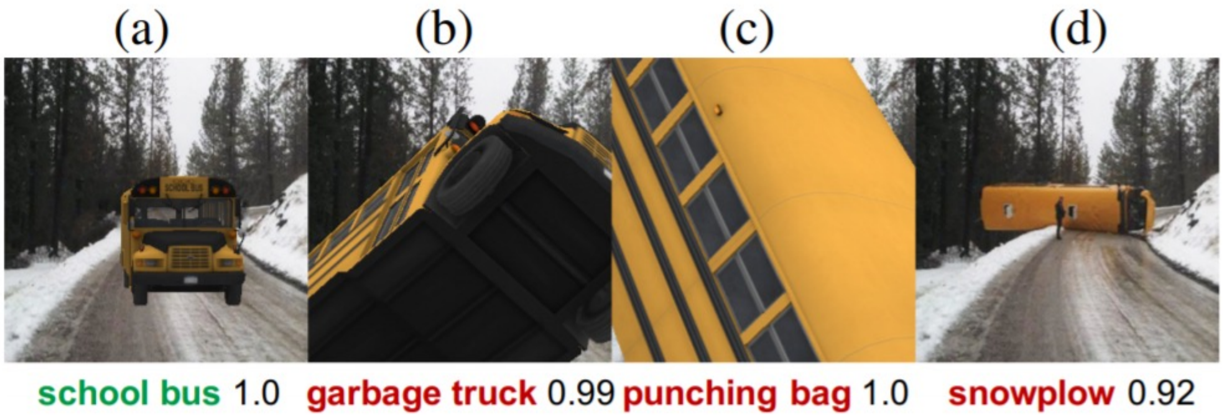
E. Paxon Frady

intel

Berkeley
UNIVERSITY OF CALIFORNIA

REDWOOD CENTER
for Theoretical Neuroscience

Seems easy to me…



Alcorn et al. (2019) Strike (with) a Pose: Neural Networks Are Easily Fooled by Strange Poses of Familiar Objects. ArXiv

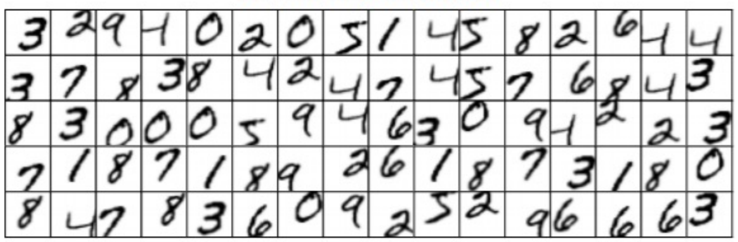
How to incorporate new “machinery” to solve these problems?

Redwood Journal Club!

DISENTANGLING IMAGES WITH LIE GROUP TRANSFORMATIONS AND SPARSE CODING

Ho Yin Chau[1,2] **Frank Qiu**[2,3] **Yubei Chen**[1,2,4] **Bruno Olshausen**[1,2]

2D Translation Dataset

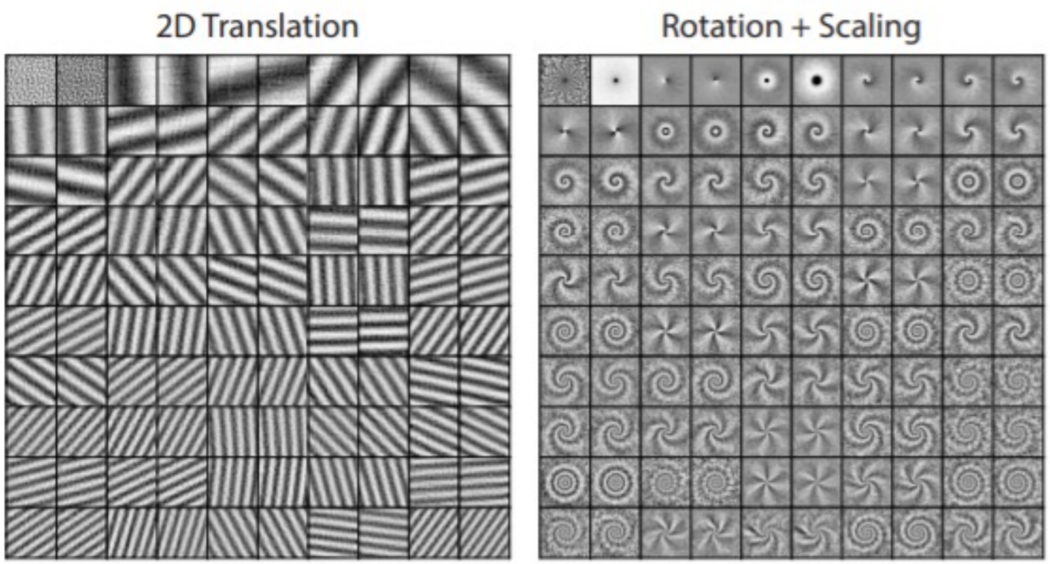


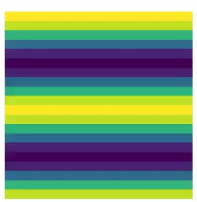
Rotation + Scaling Dataset

$$\mathbf{I} = \boldsymbol{T}(\mathbf{s})\boldsymbol{\Phi}\boldsymbol{\alpha}$$

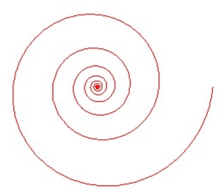
Generative models with transformations
Must infer multiple parameters
VSAs to represent transformation parameters

What did we learn?





Fourier



Log-polar

Why learn when you can engineer?
-- Still have to build-in dimensionality of transform

Vector-Symbolic Architectures (VSA): a framework for computing with distributed representations

$a, b, c \ldots = \{-1, +1\}^N$

$N = 10{,}000$

$a \cdot b \;\; : \; similarity$

(dot product)

$a + b \;\; : set\ operation$

(element-wise sum)

$a \odot b : binding\ operation$

(element-wise multiply)

$\rho(a) : ordering\ operation$

(cyclic shift)

Binary	Bipolar	Real	Complex
Binary spatter code (Kanerva, 1996) Binary sparse distributed code (Rachkovskij, 2001)	Multiply, Add, Permute (Gayler, 1998) Hyperdimensional Computing (Kanerva, 2009)	Holographic Reduced Representation (Plate, 1991) Matrix binding with additive terms (Gallant & Okaywe, 2013)	Fourier Holographic Reduced Representations (Plate, 2003)

VSAs and abstract algebra

- What makes VSA interesting/why did I want to study VSA?
  - VSA includes a binding operation, which is missing in most neural network models
- Abstract algebra defines multiplication as an operation that distributes over addition
  - This immediately leads to the combinatoric nature of multiplication
  - $(a + b + c) * (d + e + f) = ad + ae + af + bd + be + bf + cd + ce + cf$
    - 2 product with 3 sums -> 3^2 terms

HD computing is a field …and more?

- A field has addition/subtraction and multiplication/division
- Multiplication distributes over addition
  - Distributivity is key to multiplication
  - Combinatorics of distributivity $(a + b + c) * (x + y + z)$
- Binding distributes over addition
  - $x \odot (y + z) = x \odot y + x \odot z$
- Permutation also has distributive property
  - $\rho(x + y) = \rho(x) + \rho(y)$
- Permutation also distributes across binding
  - $\rho(x \odot y) = \rho(x) \odot \rho(y)$

Combinatoric representations in connectionism

- The need to represent conjunctions and other types of combinatoric data structures lead connectionists to develop new operations in neural networks.
  - Superposition catastrophe
- Tensor products were natural extension of requirements for representing combinatoric structure in neural networks
  - Dimensionality increases with combinatorics
- VSAs utilize randomness to represent combinatoric structures in vector spaces of fixed dimensionality
  - VSA binding is equivalent to tensor products through lens of compressed sensing.

When are VSAs useful?

- Exploit the combinatoric structure
  - The combinatorics of data structures greatly exceeds the dimensionality of the representation space
    - Trigrams of letters: $26^3 > 17K$, want to reduce dimensions to less than 17K
- Avoid large superpositions
  - Sparsity in the data – only a few of these combinatoric possibilities are represented at a time.

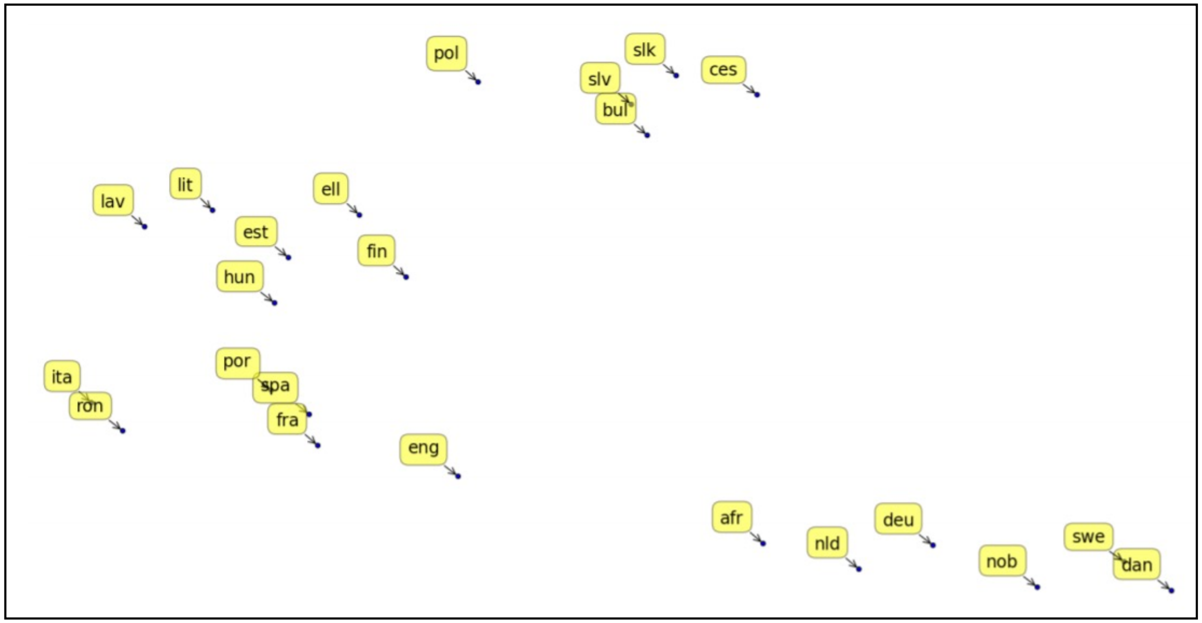
LANGUAGE RECOGNITION USING RANDOM INDEXING

Aditya Joshi Johan T. Halseth Pentti Kanerva

$$V_1 = \rho\rho A * \rho B * C$$
$$V_2 = \rho\rho A * \rho C * B$$

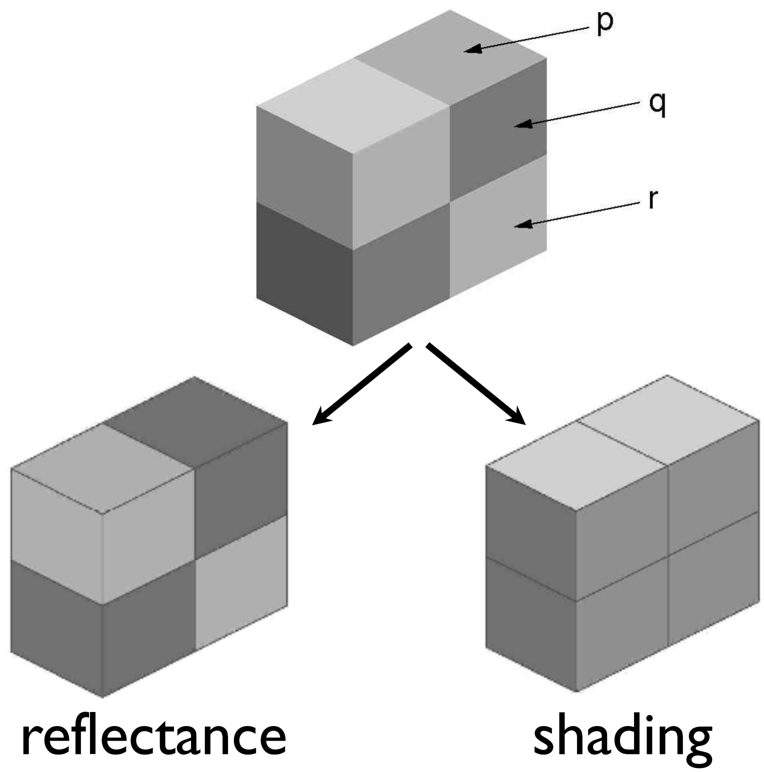
$$10{,}000 < 27^3 = 19{,}683$$

Sparse statistics = less noise



Language Vectors = histogram of letter-trigrams

Factorization in perception



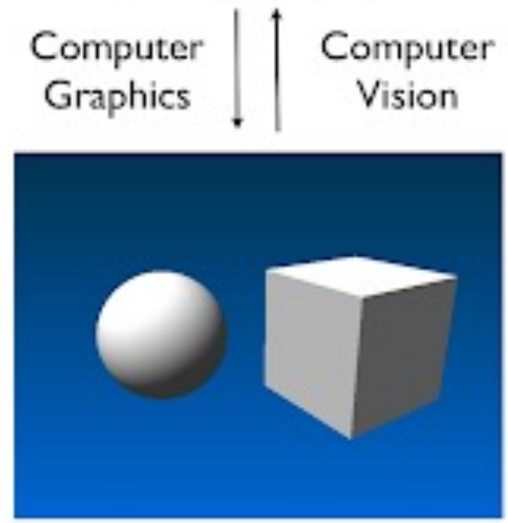
(Adelson, 2000)

Scene understanding and inverse graphics are factoring problems

$$\mathbf{s} = \mathbf{c}_{\mathtt{cyan}} \odot \mathbf{d}_7 \odot \mathbf{v}_{\mathtt{top}} \odot \mathbf{h}_{\mathtt{left}} + \mathbf{c}_{\mathtt{pink}} \odot \mathbf{d}_3 \odot \mathbf{v}_{\mathtt{top}} \odot \mathbf{h}_{\mathtt{right}} + \mathbf{c}_{\mathtt{red}} \odot \mathbf{d}_8 \odot \mathbf{v}_{\mathtt{middle}} \odot \mathbf{h}_{\mathtt{left}}$$

Disentangle `what' and `where'

$$\{cube, size, x_0, y_0, z_0, \theta_{XY}, \theta_{XZ}, \theta_{YZ}, \dots\}$$
$$\{sphere, radius, x_1, y_1, z_1, \dots\}$$



Representing simple scenes

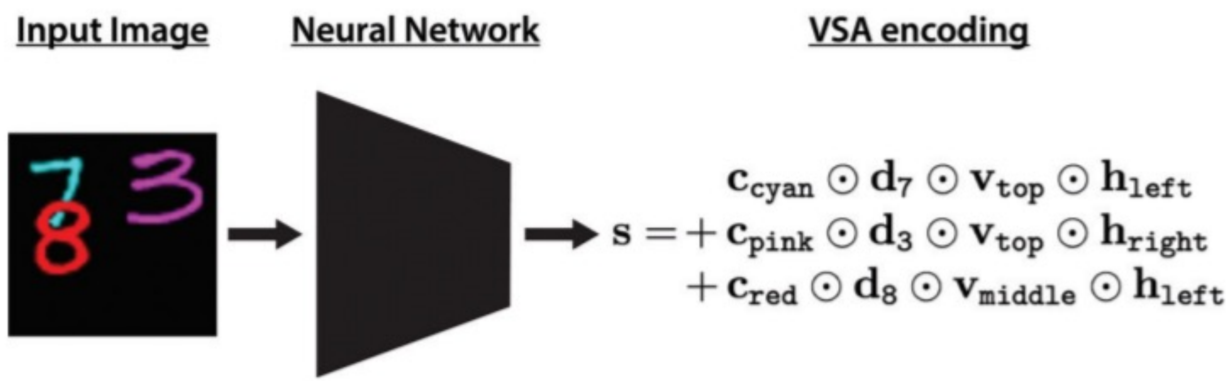


Figure 3: Generating a vector symbolic encoding of a visual scene.

Factorization problems are common in VSAs

Let $$\mathbf{b} = \mathbf{x} \otimes \mathbf{y} \otimes \mathbf{z}$$

$$\mathbf{x} \in \mathbb{X} := \{\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n\}$$
$$\mathbf{y} \in \mathbb{Y} := \{\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_n\}$$
$$\mathbf{z} \in \mathbb{Z} := \{\mathbf{z}_0, \mathbf{z}_1, \ldots, \mathbf{z}_n\}$$

Problem: You are given $\mathbf{b}$, what are $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$?

Consider the following energy function

$$E = -\mathbf{b} \cdot (\mathbf{x} \otimes \mathbf{y} \otimes \mathbf{z})$$

$$\mathbf{x} = \sum_{i=1}^{n} \alpha_i \, \mathbf{x}_i, \quad \mathbf{y} = \sum_{i=1}^{n} \beta_i \, \mathbf{y}_i, \quad \mathbf{z} = \sum_{i=1}^{n} \gamma_i \, \mathbf{z}_i$$

Consider the following energy function

1,000,000 combinations! (n=100)

$$(\alpha_1\beta_1\gamma_1\,\mathbf{x}_1\otimes\mathbf{y}_1\otimes\mathbf{z}_1\;+\;\ldots\;+\;\alpha_i\beta_j\gamma_k\,\mathbf{x}_i\otimes\mathbf{y}_j\otimes\mathbf{z}_k\;+\;\ldots\;+\;\alpha_n\beta_n\gamma_n\,\mathbf{x}_n\otimes\mathbf{y}_n\otimes\mathbf{z}_n)$$

$$E = -\mathbf{b}\cdot(\mathbf{x}\otimes\mathbf{y}\otimes\mathbf{z})$$

$$\mathbf{x} = \sum_{i=1}^{n}\alpha_i\,\mathbf{x}_i, \quad \mathbf{y} = \sum_{i=1}^{n}\beta_i\,\mathbf{y}_i, \quad \mathbf{z} = \sum_{i=1}^{n}\gamma_i\,\mathbf{z}_i$$

Resonator networks for factoring distributed representations of data structures

$$\boldsymbol{s} = \boldsymbol{x} \odot \boldsymbol{y} \odot \boldsymbol{z} + noise$$

$$\boldsymbol{x} \in \boldsymbol{X} : \{\pm 1\}^{N \times D_X}$$

$$\boldsymbol{y} \in \boldsymbol{Y} : \{\pm 1\}^{N \times D_Y}$$

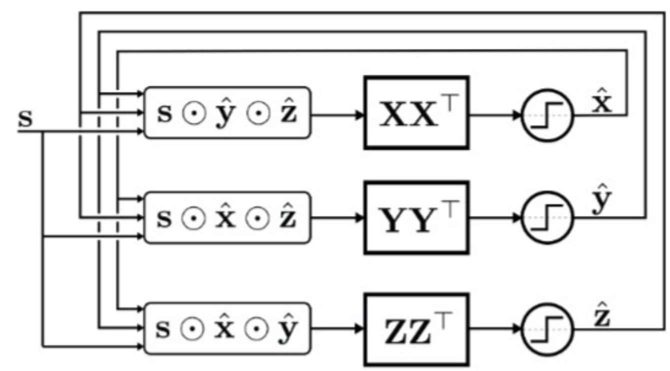
$$\boldsymbol{z} \in \boldsymbol{Z} : \{\pm 1\}^{N \times D_Z}$$

Strategy: Search in *superposition*

$$\hat{\boldsymbol{x}}(t) = f(\boldsymbol{X}\boldsymbol{X}^\top(\boldsymbol{s} \odot \hat{\boldsymbol{y}}(t-1) \odot \hat{\boldsymbol{z}}(t-1)))$$

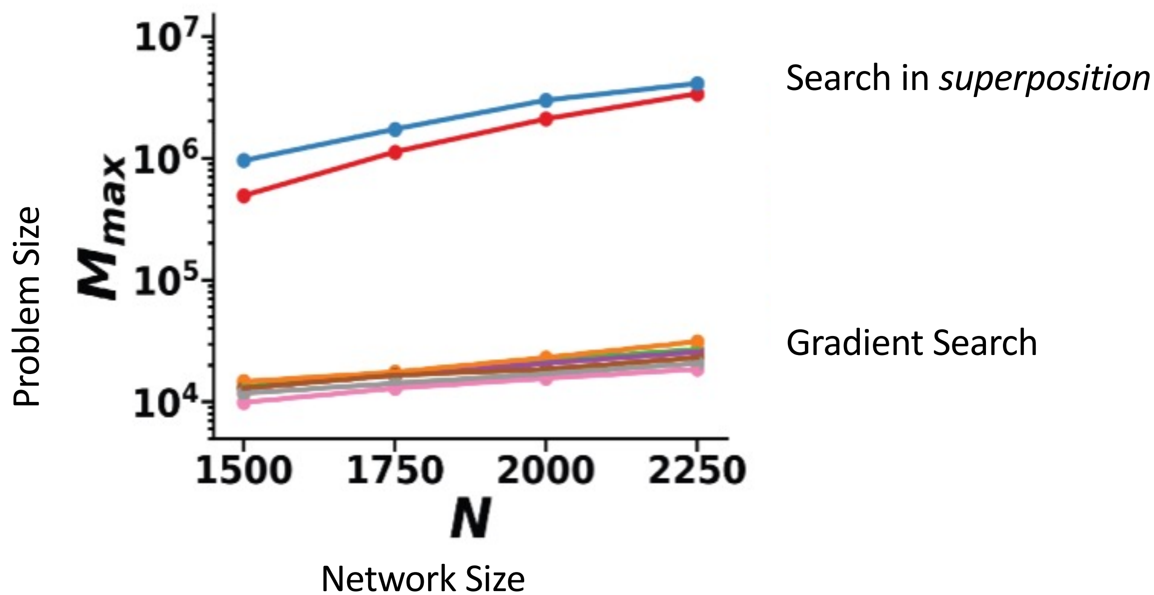
$$\hat{\boldsymbol{y}}(t) = f(\boldsymbol{Y}\boldsymbol{Y}^\top(\boldsymbol{s} \odot \hat{\boldsymbol{x}}(t-1) \odot \hat{\boldsymbol{z}}(t-1)))$$

$$\hat{\boldsymbol{z}}(t) = f(\underbrace{\boldsymbol{Z}\boldsymbol{Z}^\top}_{\text{Constrain } \hat{\boldsymbol{z}} \text{ guesses to codebook } \boldsymbol{Z}}(\underbrace{\boldsymbol{s} \odot \hat{\boldsymbol{x}}(t-1) \odot \hat{\boldsymbol{y}}(t-1)}_{\text{Use guesses in } \hat{\boldsymbol{x}}, \hat{\boldsymbol{y}} \text{ to infer } \hat{\boldsymbol{z}}}))$$



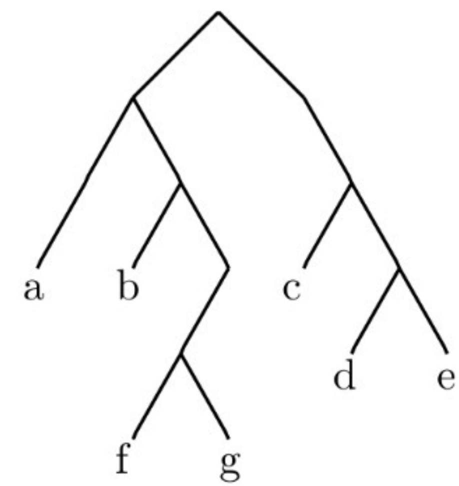
Combination of attractor networks and multiplication/binding

Performance of the resonator network



Tree search as factorization problem

$$
\begin{aligned}
\mathbf{tree} \;=\;& \mathbf{a} \odot \mathbf{left} \odot \rho(\mathbf{left}) \odot \rho^2(\mathbf{left}) \\
&+ \mathbf{b} \odot \mathbf{left} \odot \rho(\mathbf{right}) \odot \rho^2(\mathbf{left}) \\
&+ \mathbf{c} \odot \mathbf{right} \odot \rho(\mathbf{right}) \odot \rho^2(\mathbf{left}) \\
&+ \mathbf{d} \odot \mathbf{right} \odot \rho(\mathbf{right}) \odot \rho^2(\mathbf{right}) \odot \rho^3(\mathbf{left}) \\
&+ \mathbf{e} \odot \mathbf{right} \odot \rho(\mathbf{right}) \odot \rho^2(\mathbf{right}) \odot \rho^3(\mathbf{right}) \\
&+ \mathbf{f} \odot \mathbf{left} \odot \rho(\mathbf{right}) \odot \rho^2(\mathbf{right}) \odot \rho^3(\mathbf{left}) \odot \rho^4(\mathbf{left}) \\
&+ \mathbf{g} \odot \mathbf{left} \odot \rho(\mathbf{right}) \odot \rho^2(\mathbf{right}) \odot \rho^3(\mathbf{left}) \odot \rho^4(\mathbf{right})
\end{aligned}
$$



$$
\mathbf{tree} \odot \big(\mathbf{left} \odot \rho(\mathbf{right}) \odot \rho^2(\mathbf{left})\big) = \mathbf{b} + noise
$$

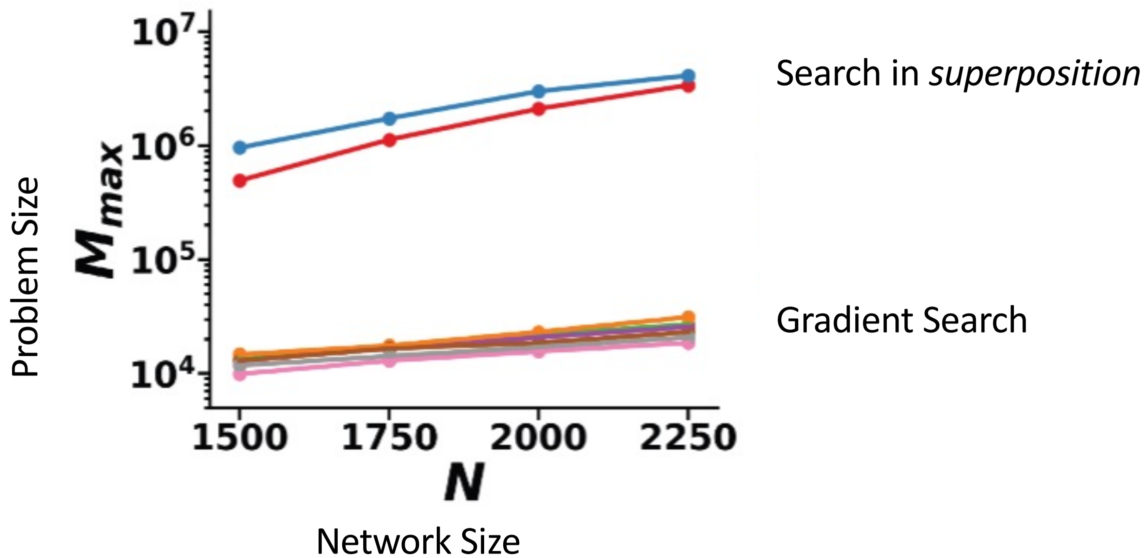
Using the resonator network to solve tree search

$$\mathbf{tree} \odot \mathbf{c} = \mathbf{right} \odot \rho(\mathbf{right}) \odot \rho^2(\mathbf{left}) + \textit{noise}.$$

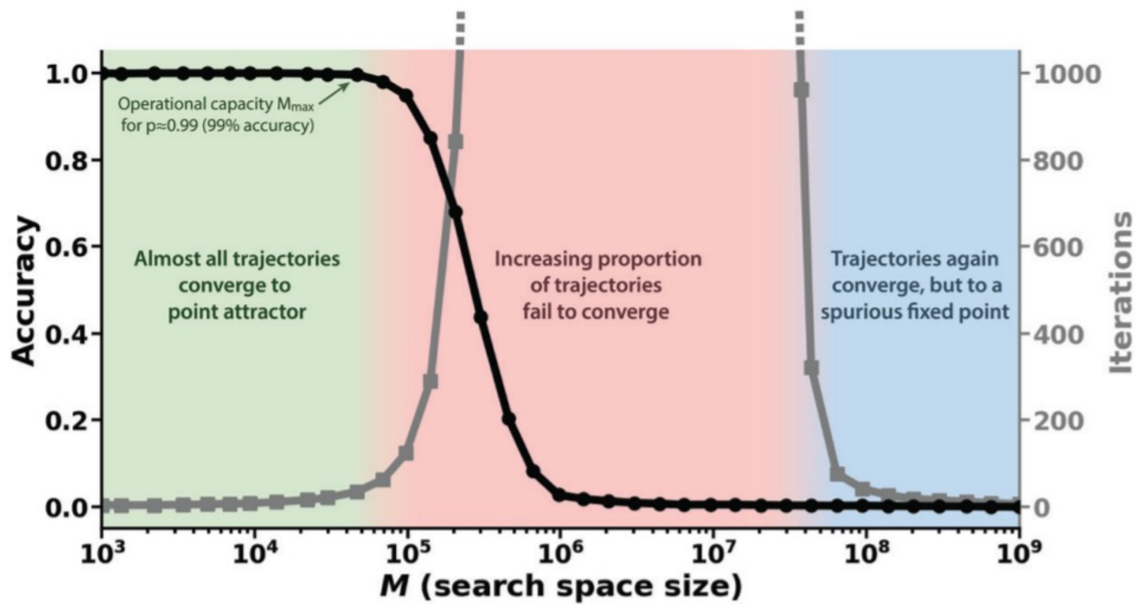
We denote each factor estimate as $\hat{\mathbf{x}}^{(0)}, \hat{\mathbf{x}}^{(1)}, \hat{\mathbf{x}}^{(2)}, \hat{\mathbf{x}}^{(3)}, \hat{\mathbf{x}}^{(4)}$ and the codebook matrices as $\mathbf{X}_0, \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4$. Each codebook matrix contains permuted versions of **left** and **right**, and **1**: $\mathbf{X}_d = [\rho^d(\mathbf{left}), \rho^d(\mathbf{right}), \mathbf{1}]$ where d indicates the depth in the tree. The network is constructed analogous to equation 3.2, but with five factor estimates running in parallel instead of three. For instance, the update equation for the first estimate is

$$\hat{\mathbf{x}}^{(0)}(t+1) = g\big(\mathbf{X}_0\mathbf{X}_0^\top(\mathbf{s} \odot \hat{\mathbf{x}}^{(1)}(t) \odot \hat{\mathbf{x}}^{(2)}(t) \odot \hat{\mathbf{x}}^{(3)}(t) \odot \hat{\mathbf{x}}^{(4)}(t))\big). \tag{4.4}$$

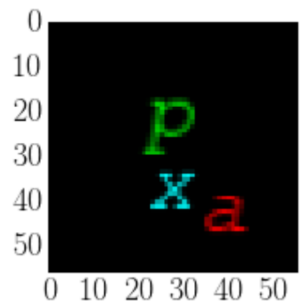
Resonator networks performance comparison

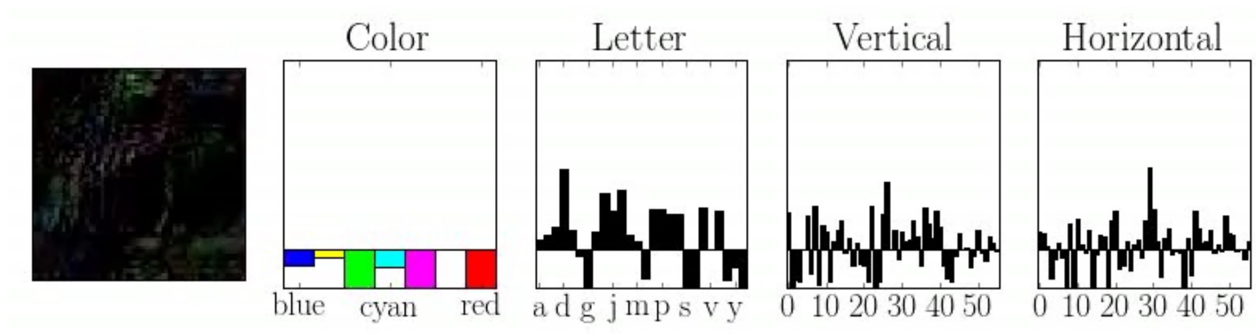


Resonator network convergence properties



Resonator networks for simple scene inference





How to represent continuous manifolds with complex HD vectors

VSA binding in the complex domain can form index patterns of continuous variables.

(1) Distinguishability: Points far apart are encoded by dissimilar “indexing” patterns.

(2) Smoothness: Points close to each other are encoded by similar patterns that smoothly transition.

$$e^{i\phi_1 x(t)}$$

$$\boldsymbol{v} \odot \boldsymbol{v} = \boldsymbol{v}^2$$

$$\boldsymbol{v} \odot \boldsymbol{v} \odot \sqrt{\boldsymbol{v}} = \boldsymbol{v}^{2.5}$$

Spinning Phasors: $\boldsymbol{v}^{x(t)}$

“exponentiation trick”

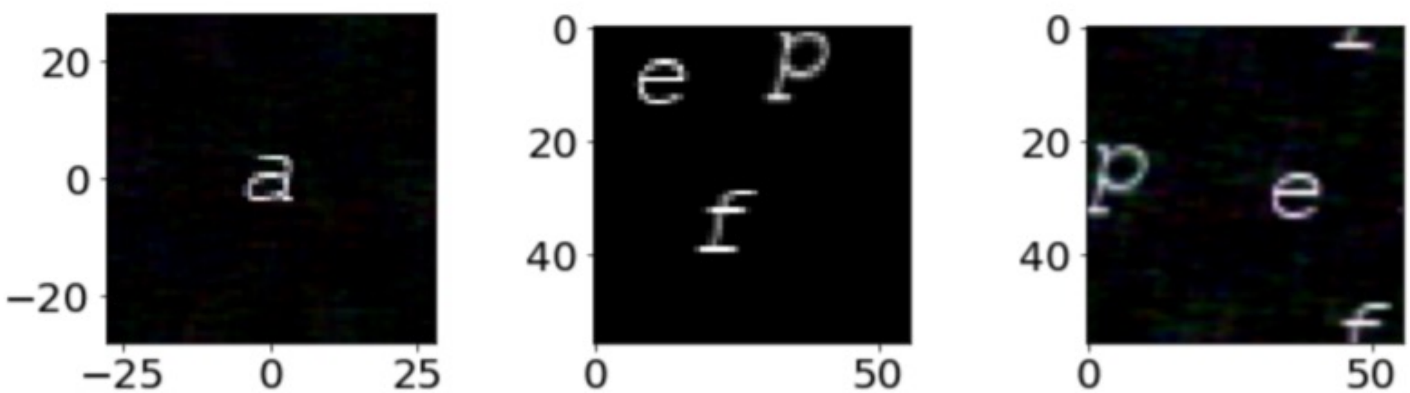
arXiv.org > cs > arXiv:2109.03429

Computer Science > Machine Learning

[Submitted on 8 Sep 2021]

Computing on Functions Using Randomized Vector Representations

E. Paxon Frady, Denis Kleyko, Christopher J. Kymn, Bruno A. Olshausen, Friedrich T. Sommer

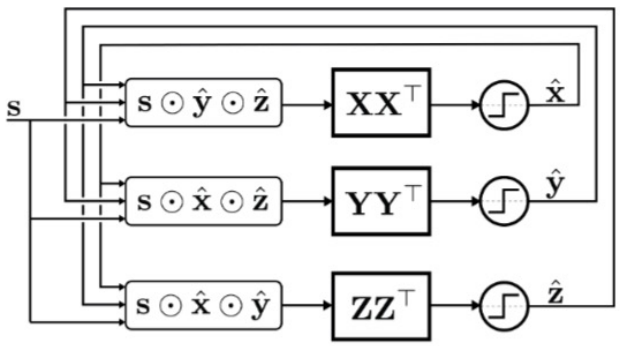


$$\mathbf{v}_a = \sum_{x,y} Im(x,y) \cdot \mathbf{x}^x \odot \mathbf{y}^y$$

$$\begin{aligned}\mathbf{v}_{scene} &= \mathbf{v}_e \odot \mathbf{x}^{12} \odot \mathbf{y}^{10} \\ &+ \mathbf{v}_p \odot \mathbf{x}^{35.86} \odot \mathbf{y}^{7.22} \\ &+ \mathbf{v}_f \odot \mathbf{x}^{22.7} \odot \mathbf{y}^{35}\end{aligned}$$

$$\mathbf{v}^{tr}_{scene} = \mathbf{v}_{scene} \odot \mathbf{x}^{25.8} \odot \mathbf{y}^{20.2}$$

Programming the resonator network



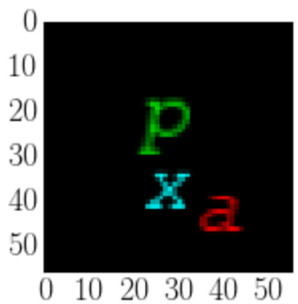
$$\boldsymbol{X} = \left[\boldsymbol{x}^{\boldsymbol{1}}, \boldsymbol{x}^{\boldsymbol{2}}, \dots, \boldsymbol{x}^{\boldsymbol{k}}\right]$$

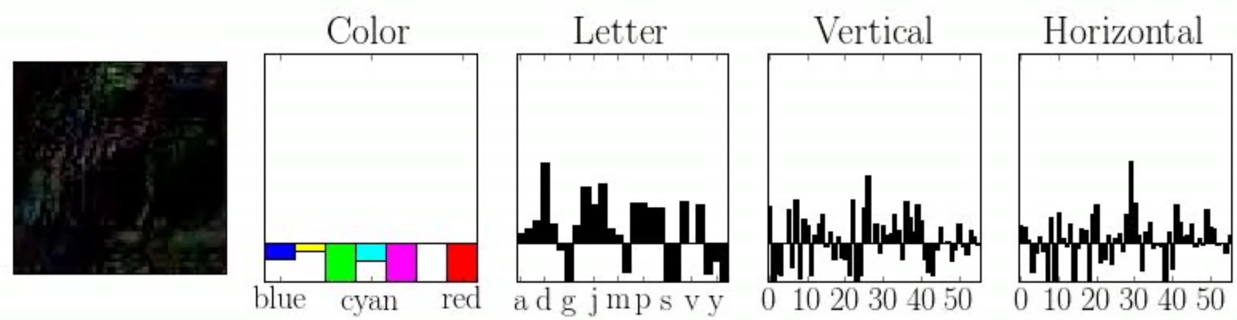
$$\boldsymbol{Y} = \left[\boldsymbol{y}^{\boldsymbol{1}}, \boldsymbol{y}^{\boldsymbol{2}}, \dots, \boldsymbol{y}^{\boldsymbol{l}}\right]$$

$$\boldsymbol{Z} = \left[\boldsymbol{t}^{\boldsymbol{a}}, \boldsymbol{t}^{\boldsymbol{b}}, \dots, \boldsymbol{t}^{\boldsymbol{z}}\right]$$

$$\boldsymbol{t}^{\boldsymbol{a}} = \sum_{x,y} x^{x} \odot y^{y} \cdot Template(x, y)$$

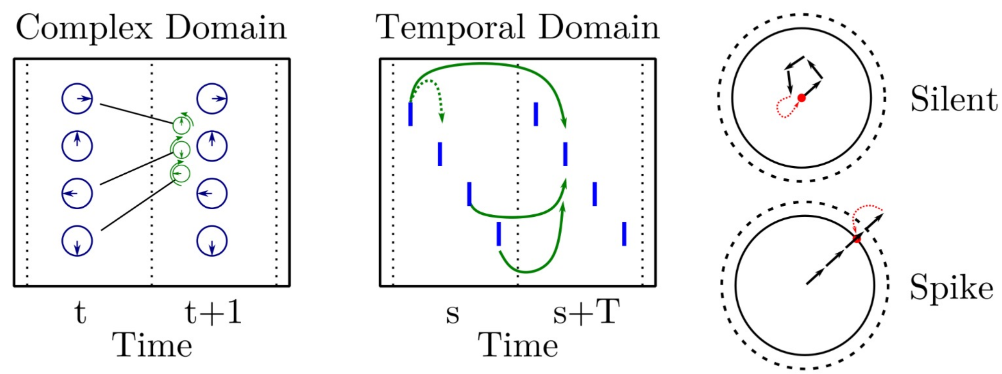
Resonator networks for simple scene inference





How to scale on neuromorphic hardware? How might the brain be doing it?

- Complex-valued matrix operatior
- Patterns of sparse spike-timing activity

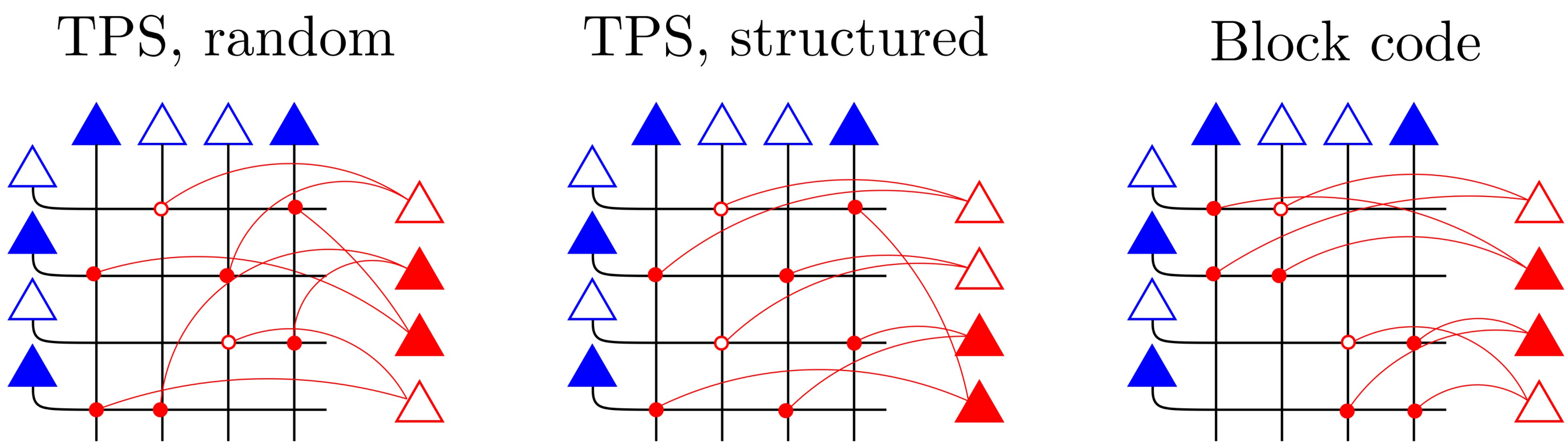


Frady, E.P., Sommer, F.T. (2019) Robust computation with rhythmic spike patterns. PNAS 116(36) 18050-59.

Sparse binding operation?

- Want something like the binding operation, but with sparse vectors
  - Operation produces quasi-orthogonal vector from two inputs
- Want atomic vectors to be sparse, and binding of atomic vectors should stay sparse
- Want minimal scaling of synaptic connections

Sparse VSA strategies



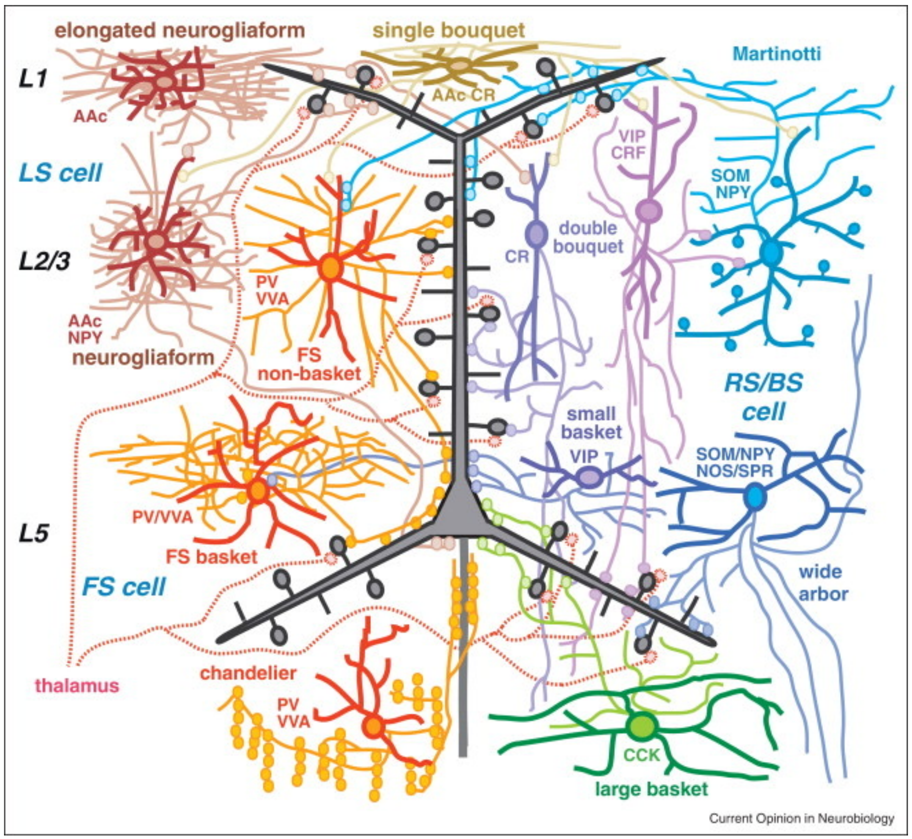
$$\# \, Synapses = \binom{N}{K} N$$

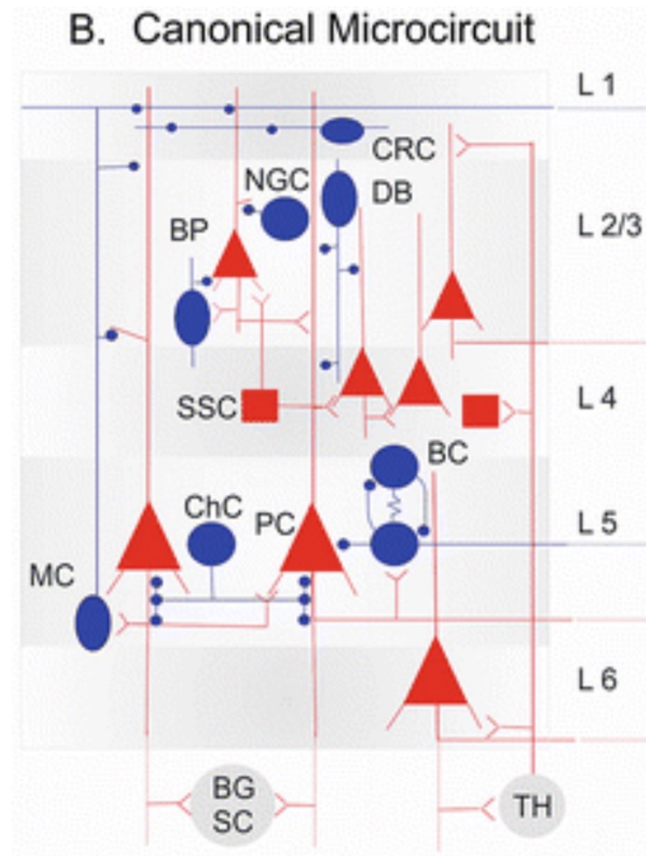
$$\# \, Neurons = N$$

$$\# \, Active = K$$

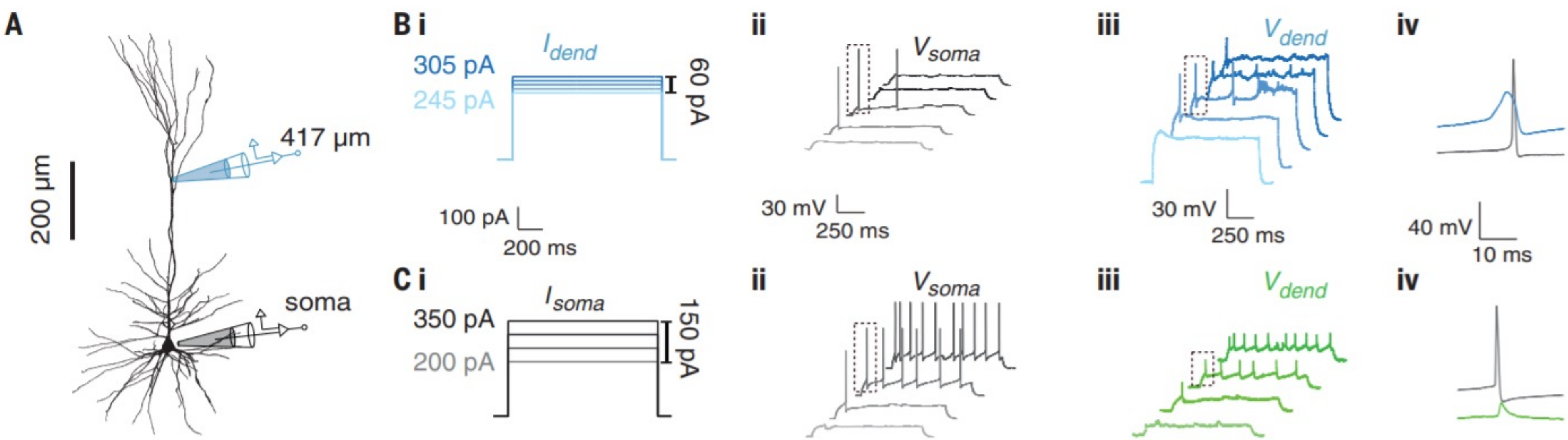
Frady, E. P., Kleyko, D., & Sommer, F. T. (2021). Variable binding for sparse distributed representations: theory and applications. *IEEE Transactions on Neural Networks and Learning Systems*.

Modern neuroscience of cortical circuits





Active dendritic mechanisms in pyramidal neurons

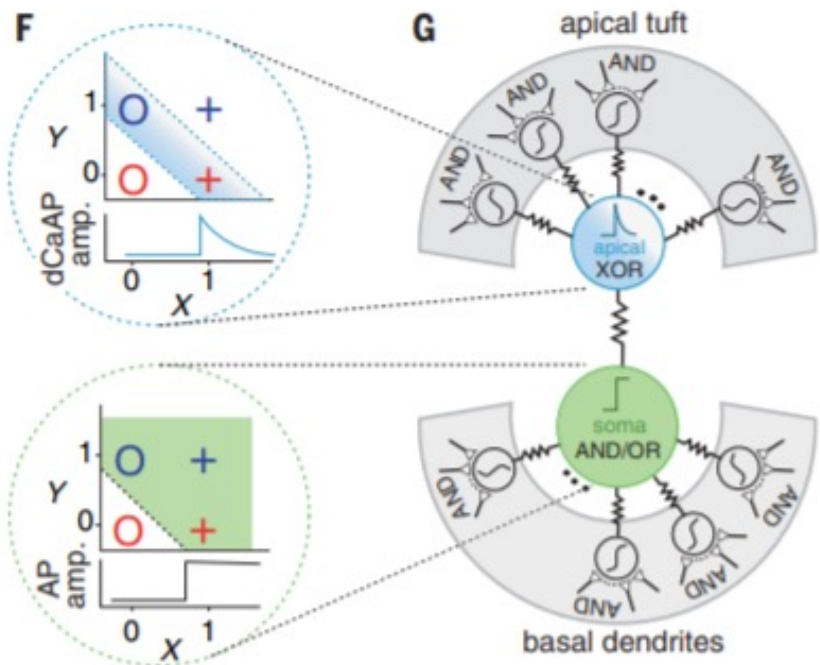


NEUROSCIENCE

Dendritic action potentials and computation in human layer 2/3 cortical neurons

Albert Gidon[1], Timothy Adam Zolnik[1], Pawel Fidzinski[2,3], Felix Bolduan[4], Athanasia Papoutsi[5], Panayiota Poirazi[5], Martin Holtkamp[2], Imre Vida[3,4], Matthew Evan Larkum[1,3]*

Overview of active dendritic processing



was triggered for (X, Y) input pairs of (1, 1), (0, 1), and (1, 0), but not for
presented as logical AND/OR gate with activation function of somatic AP,
ray background, as logical gate AND due to the NMDA spikes (33).

NEUROSCIENCE

Dendritic action potentials and computation in human layer 2/3 cortical neurons

Albert Gidon[1], Timothy Adam Zolnik[1], Pawel Fidzinski[2,3], Felix Bolduan[4], Athanasia Papoutsi[5], Panayiota Poirazi[5], Martin Holtkamp[2], Imre Vida[3,4], Matthew Evan Larkum[1,3]*

Conclusions

- Factorization is an important problem in perception
- Solving factorization problems requires significant computation
- Resonator networks use the strategy of searching in superposition to solve factorization problems efficiently
- VSAs can be extended to represent geometry and binding can be used to perform geometrical transformations
- Sparse and Complex VSAs bring us much closer to linking VSAs with neuroscience
- Binding operations in sparse VSAs resembles active dendritic integration in pyramidal cells
- Cortical microcircuit is a resonator network??

Thanks!

NIH intel NSF

Berkeley DeepDrive

REDWOOD CENTER
for Theoretical Neuroscience