Probabilistic Computing via Sparse Distributed Representations

Lyric Semiconductor, Dec. 14, 2012

Rod Rinkus, President
Neurithmic Systems, LLC

Two Approaches to Non-Binary Computing

Lyric

- Lowest-level signals in system are represented as continuous-valued
- A voltage level on a single wire communicates, in a single transmission cycle, e.g., an 8-bit, quantity, e.g., interpretable as a probability.
- In conventional computing, 8 signals, each a binary voltage level, have to be sent and then combined (decoded) at the destination

Neurithmic

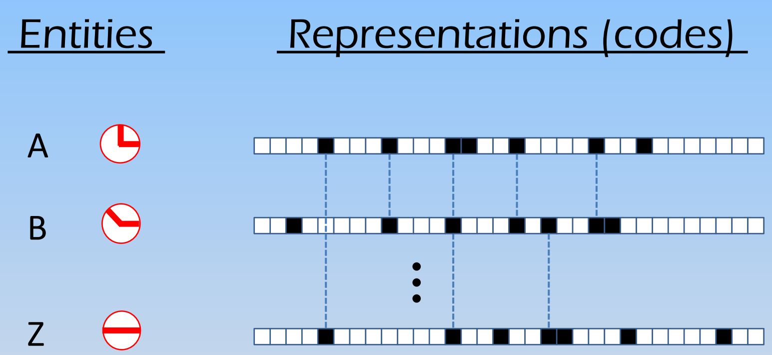
- Lowest-level signals are represented as binary-valued
- All represented values, i.e. of:
    - variables
    - relationships between variables, e.g., conditional probabilities

are represented as *sums* of binary signals

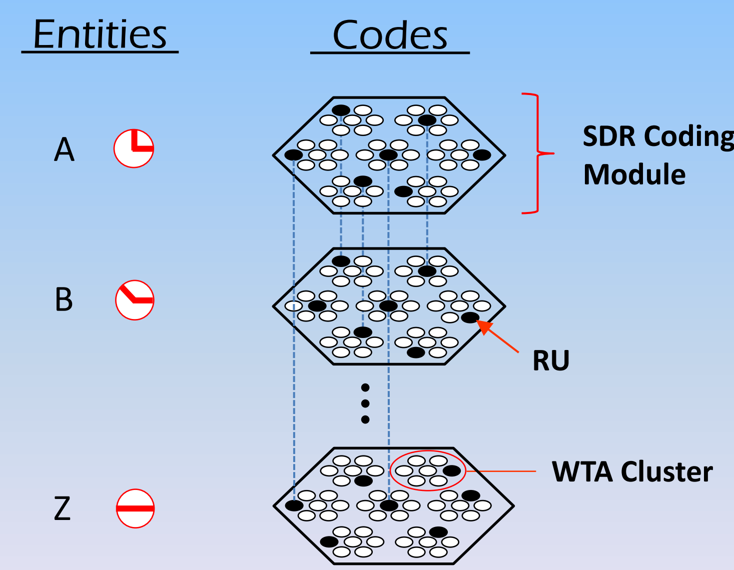
Sparse Distributed Representation

- Every represented entity in the system is represented by a subset of binary representational units (RUs) chosen from a much larger set.
- The subsets can overlap
- It's possible to represent similarity of entities by overlap of their codes

  Similar-inputs-to-similar-codes (SISC) property

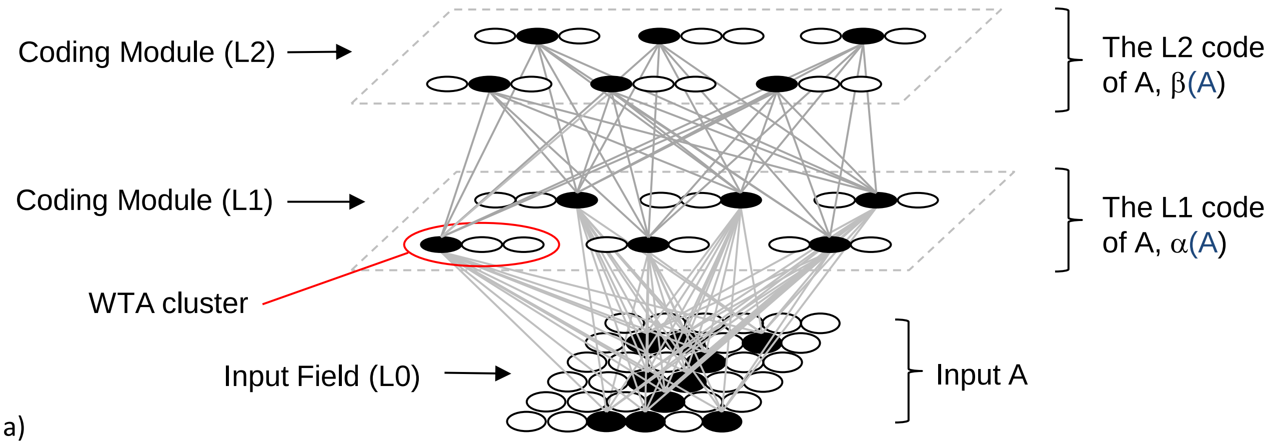


e.g., Kanerva, Willshaw et al, Palm



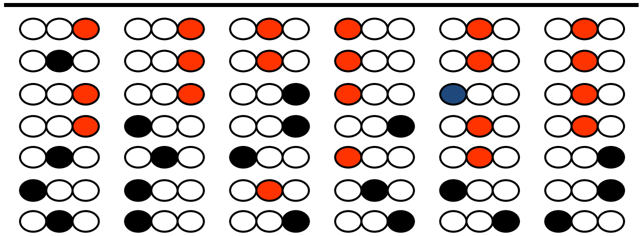
TEMECOR

Realizing SISC via SDR



a)

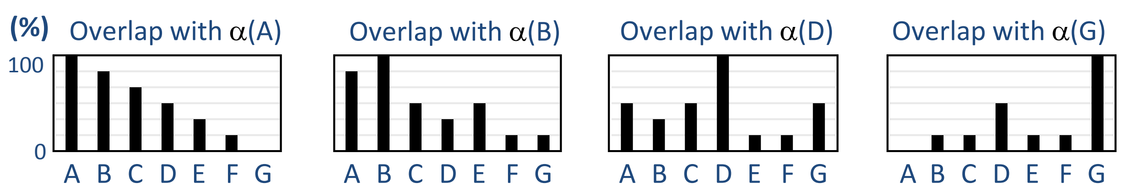
Representation (Code)



Input	Name	Physical Realization	Overlap with α(A)
A	α(A)		6 / 6
B	α(B)		5 / 6
C	α(C)		4 / 6
D	α(D)		3 / 6
E	α(E)		2 / 6
F	α(F)		1 / 6
G	α(G)		0 / 6

b)

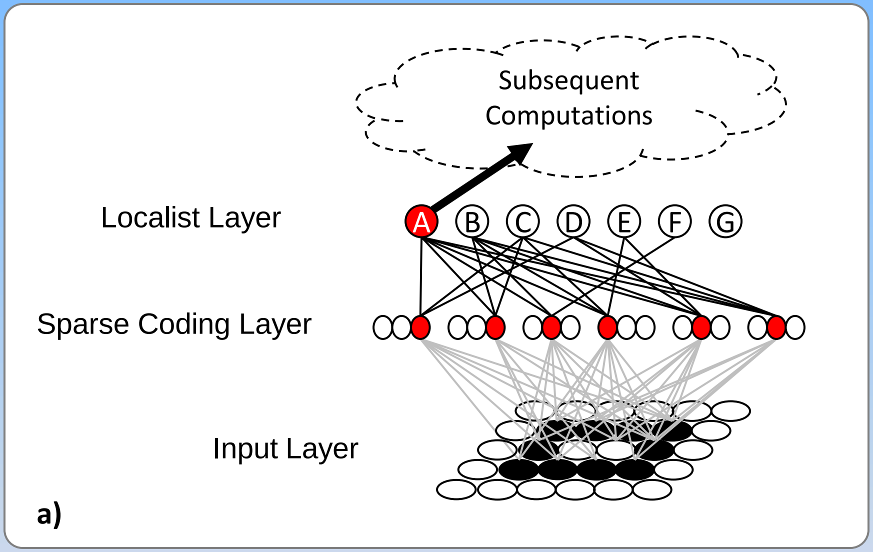
Any single code represents ALL stored codes

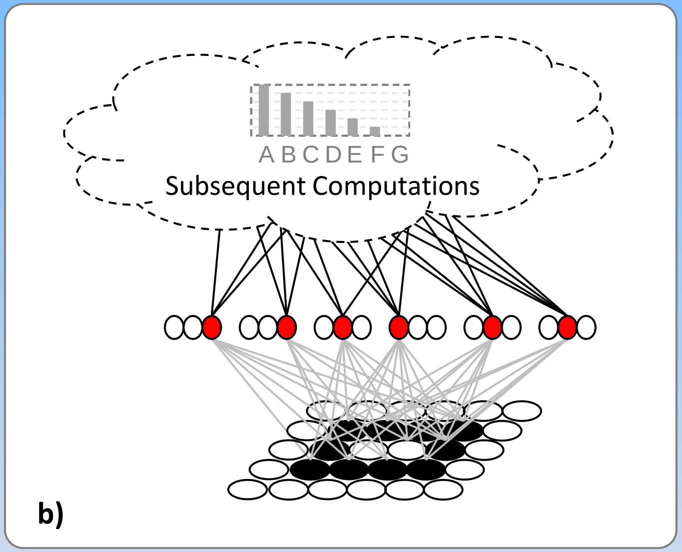


c)

Not Forcing Computation Through a Localist Nexus

- Any single code represents ALL stored codes

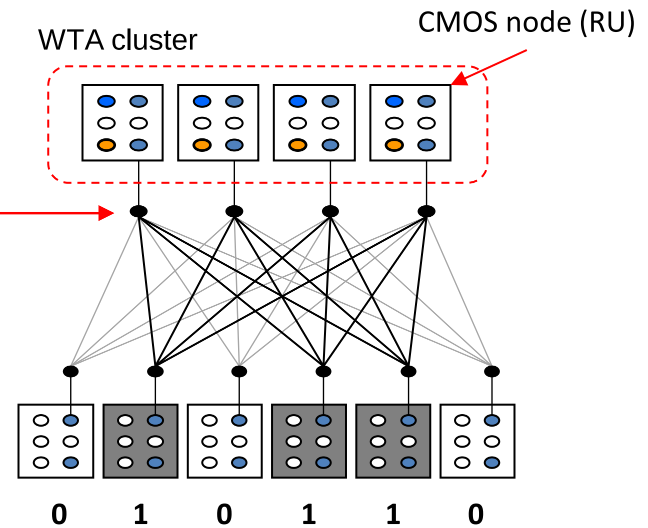




Summing Binary Signals

Physically

- Voltages sum simultaneously on terminal.
- No decoding needed



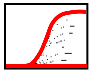
cf. "quantal release" in synaptic transmission

Algorithmically

- Involves iterating over RUs and over each RU's set of incoming weights.
- These are fixed quantities
- → Constant time complexity

- For each RU, i
    - Sum = 0;
    - For all incoming wts, w_{ji}
        - sum_i += w_{ji} * a_j

- For each RU, i
    - $a_i = f(sum_i)$



- Choose winner (soft max)

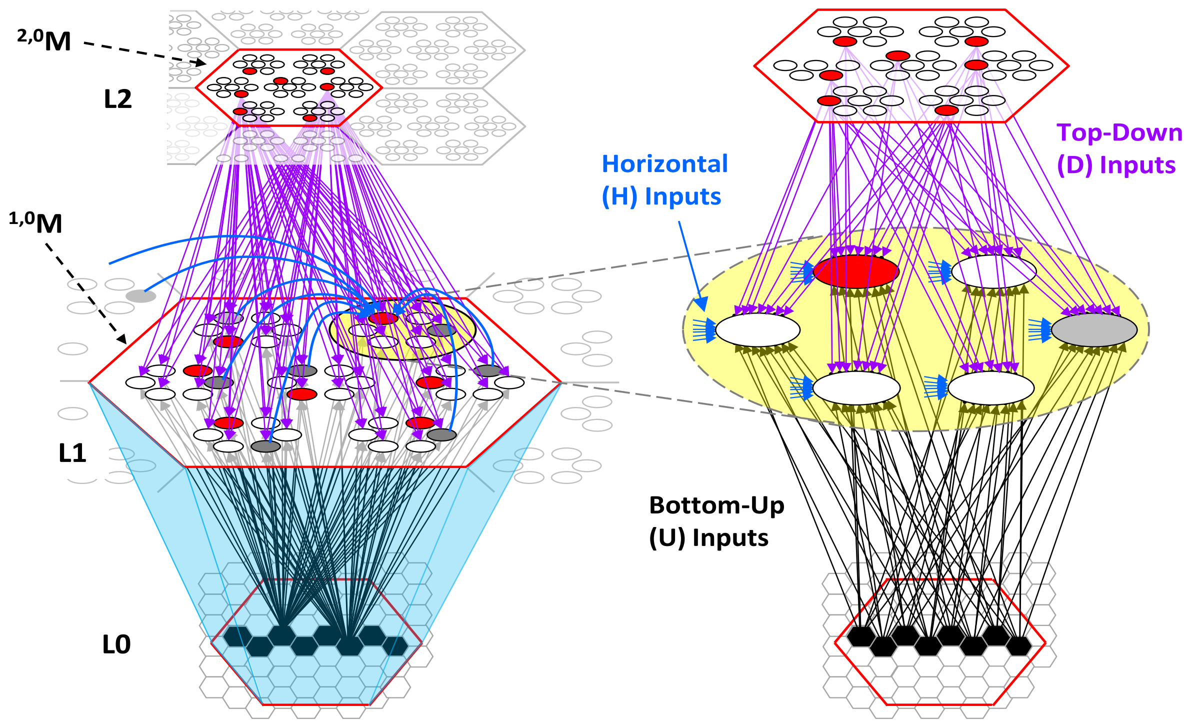
Set up cumulative distribution of a_i

Generate rand. num. in the range →

1
20
1800
1810

Three evidence sources combine to select codes

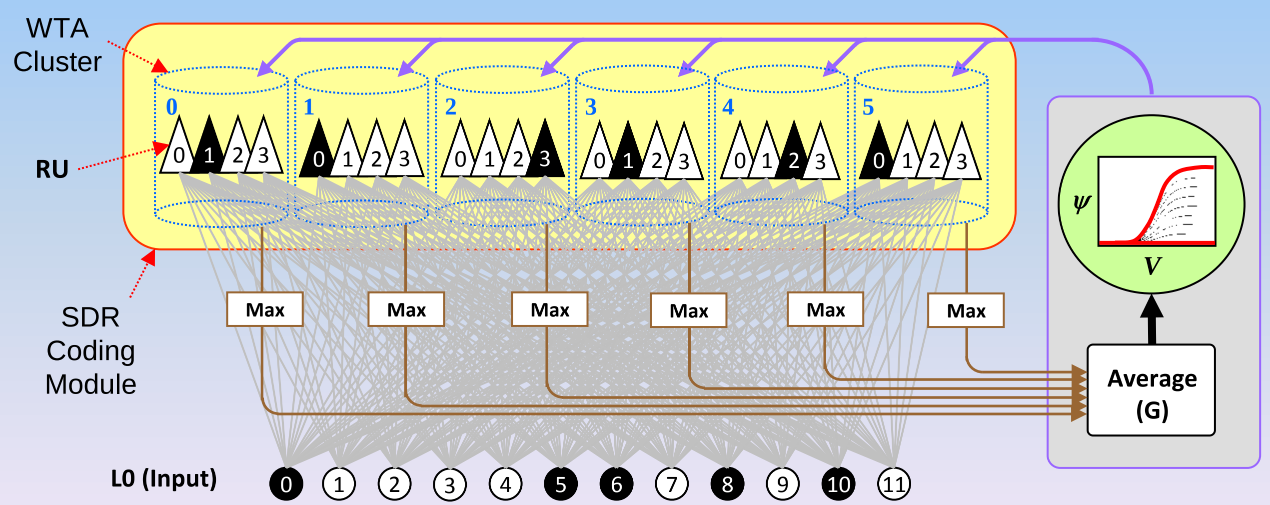
- Bottom-up, Horizontal, and Top-down evidence vectors are combined to choose which code becomes active.



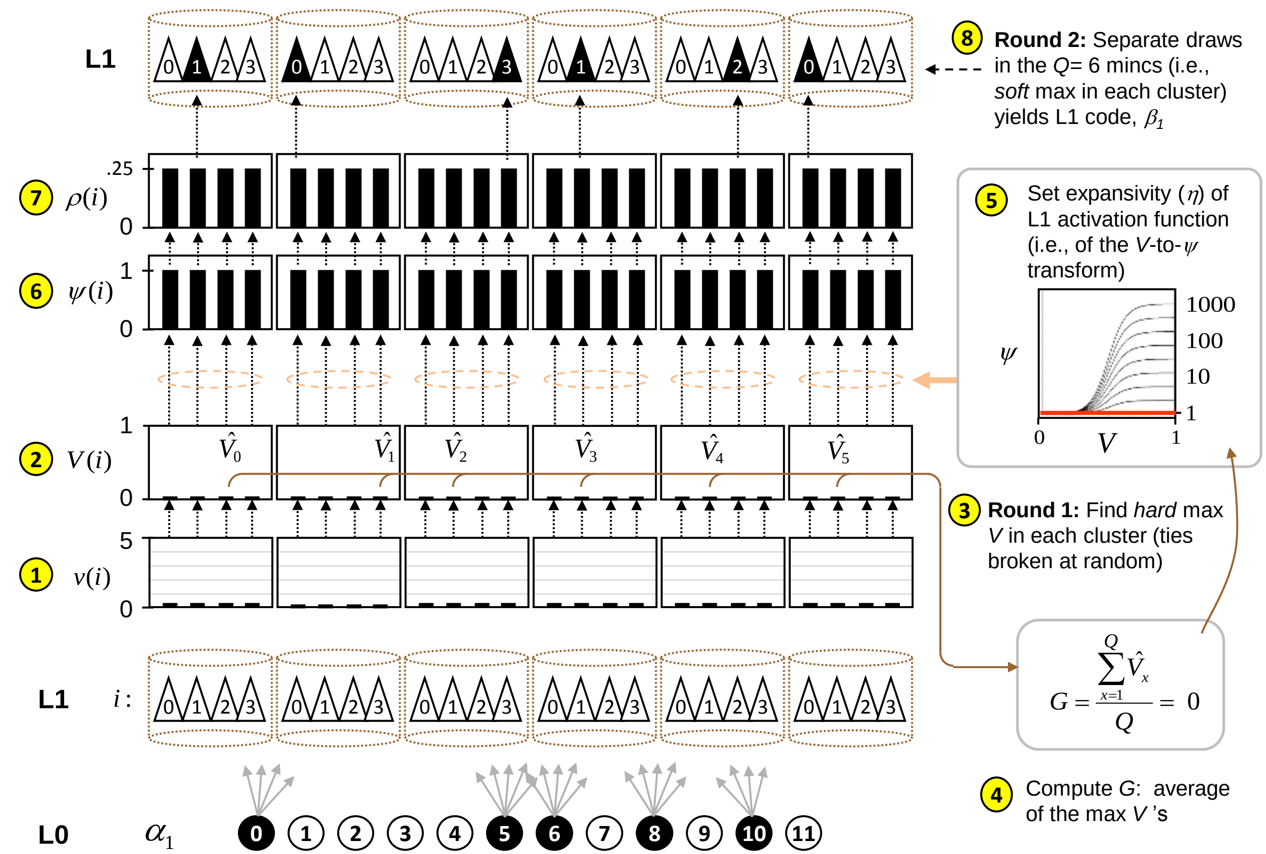
Code Selection Algorithm

1. For each RU, sum its inputs.
2. Normalize each sum to [0..1] range.
3. Multiply the three normalized sums, yielding a *local* degree of support (evidence), *V*.
4. RU with max *V* in each cluster wins (1st round).
5. Compute *G*, the ave. of the max *V*'s over all clusters in a coding module.
6. Modulate the RU activation function, *f(V)*, based on *G*:
   - As $G \rightarrow 0$, make activation function more compressive
   - As $G \rightarrow 1$, make activation function more expansive
7. For each RU, compute $\psi = f(V)$
8. In each cluster, normalize ψ's to probability measure, ρ.
9. In each cluster, choose winner *as draw* from ρ distribution (2nd round)

G is a *global* (to coding module) measure of the familiarity of total input to the module.

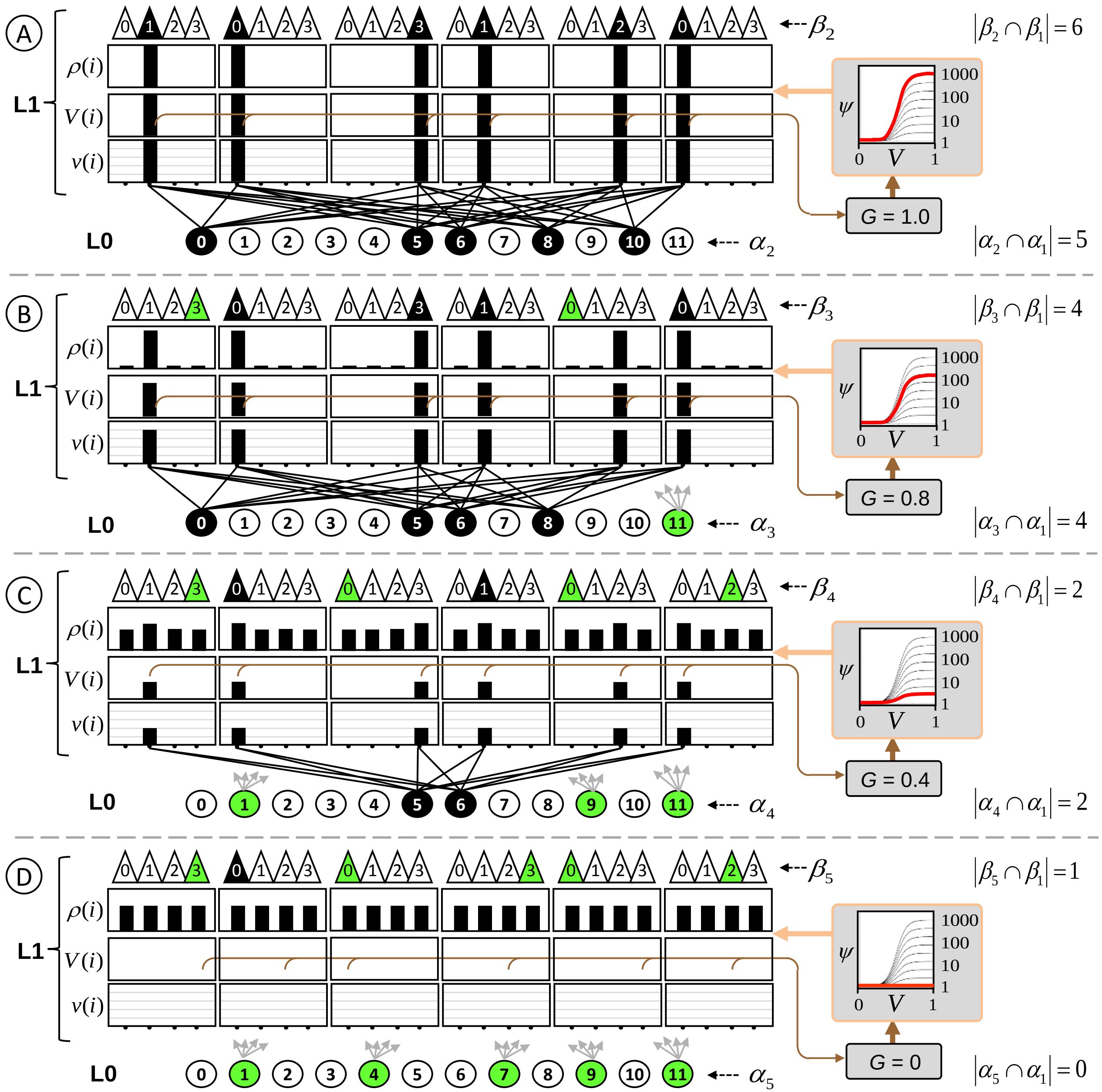


CSA Example



L1
8 Round 2: Separate draws in the Q= 6 mincs (i.e., soft max in each cluster) yields L1 code, β_1
7 $\rho(i)$
.25
0
6 $\psi(i)$
1
0
5 Set expansivity (η) of L1 activation function (i.e., of the V-to-ψ transform)
ψ
1000
100
10
1
0
V
1
2 $V(i)$
$\hat{V}_0$
$\hat{V}_1$
$\hat{V}_2$
$\hat{V}_3$
$\hat{V}_4$
$\hat{V}_5$
3 Round 1: Find hard max V in each cluster (ties broken at random)
1 $v(i)$
5
0
L1
i :
0 1 2 3
$G = \frac{\sum_{x=1}^{Q} \hat{V}_x}{Q} = 0$
4 Compute G: average of the max V 's
L0
α_1
0 1 2 3 4 5 6 7 8 9 10 11

CSA Example (cont'd)



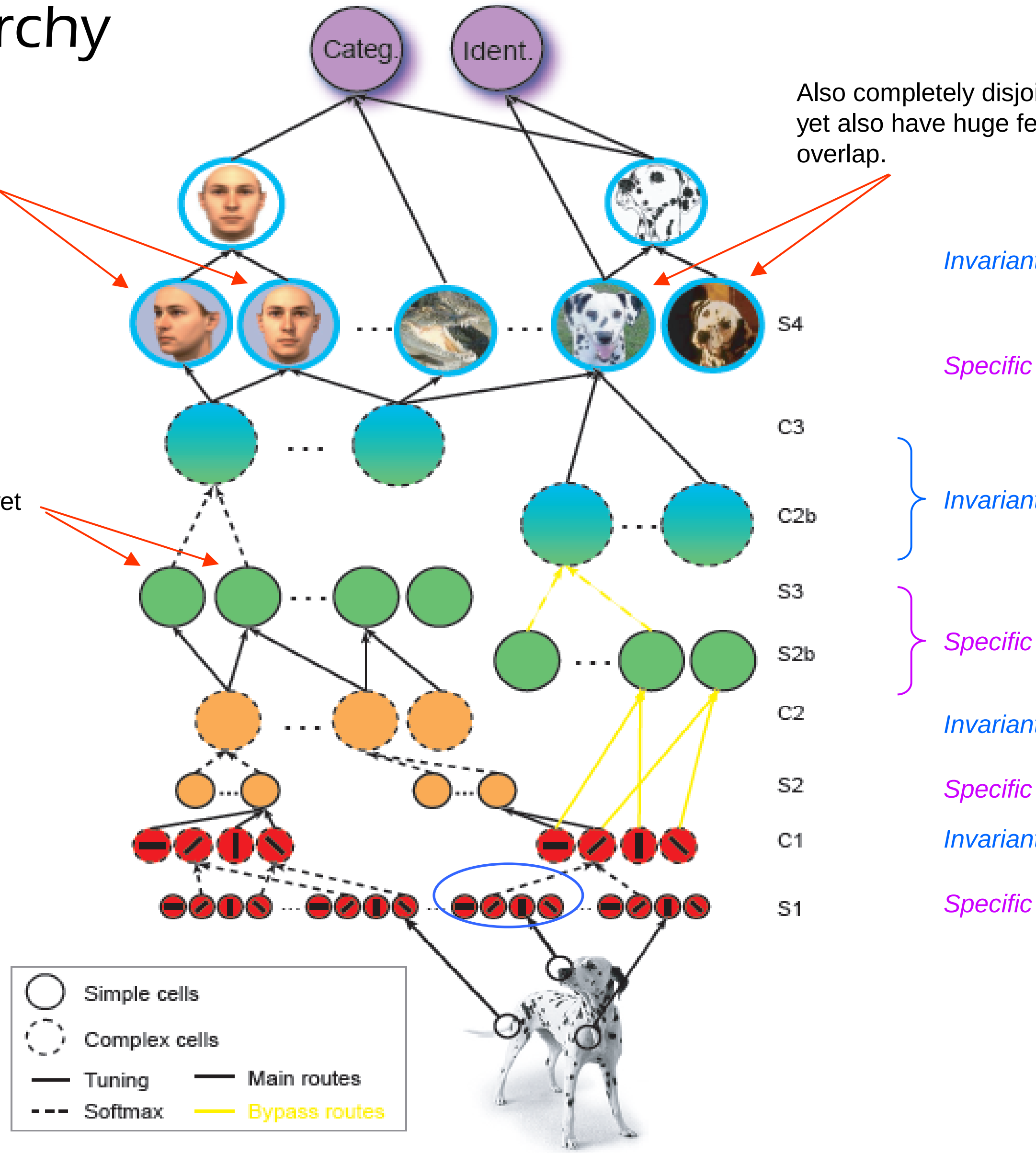
Localist Hierarchy

These two units are completely disjoint, yet they have huge featural overlap.

This is true throughout the entire model; i.e., millions of units representing highly redundant information.

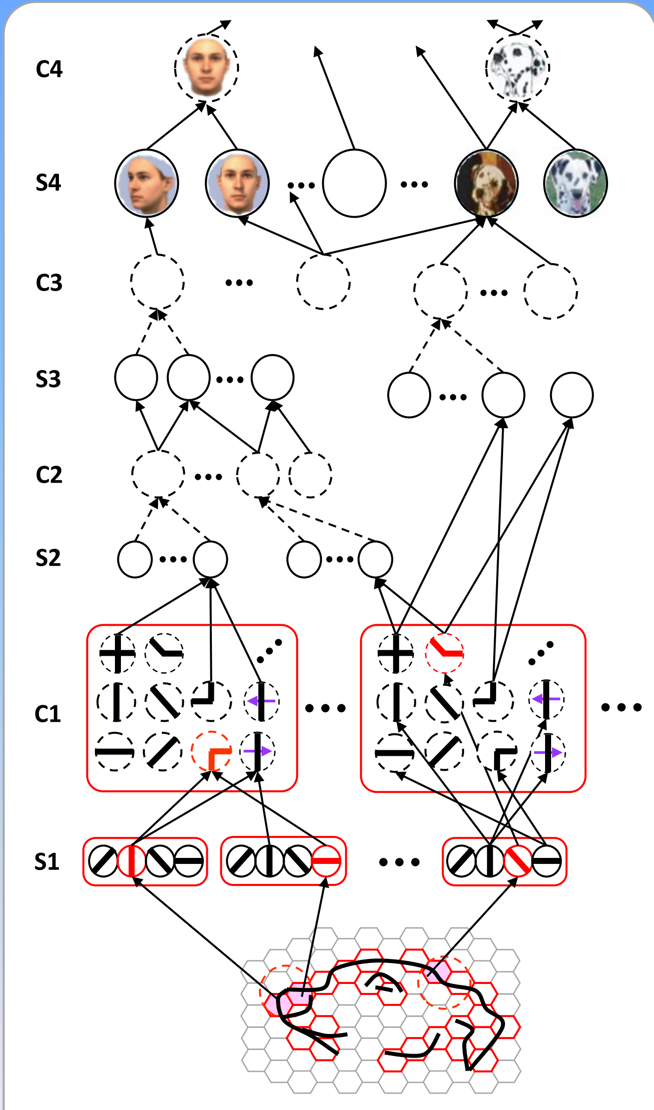
Also completely disjoint, yet also have huge featural overlap.

Also completely disjoint, yet also have huge featural overlap.

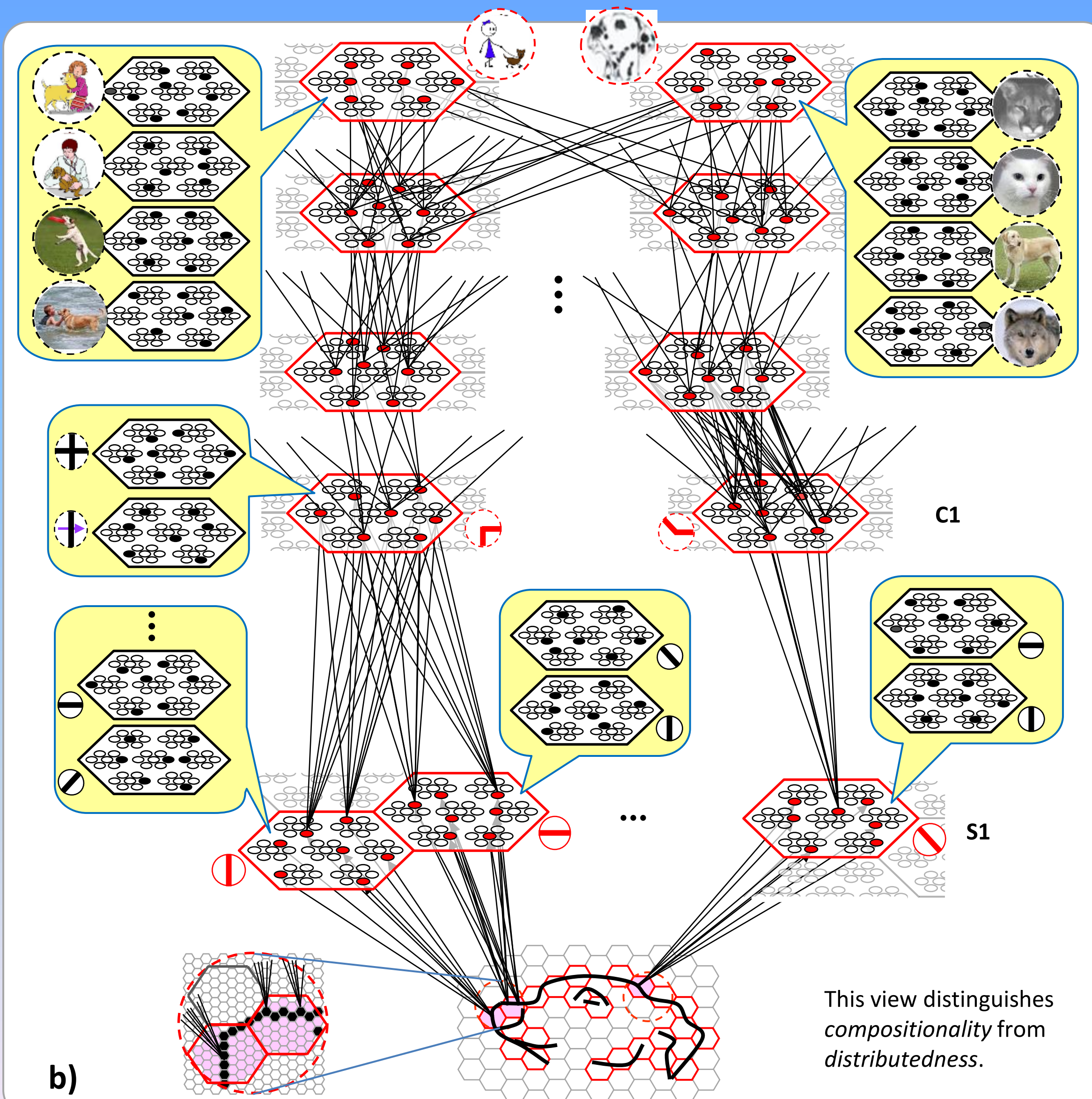


-from: Serre, T., et al., A Theory of Object Recognition: Computations and Circuits in the Feedforward Path of the Ventral Stream in Primate Visual Cortex, in AI Memo 2005-036. 2005, MIT

Localist vs. SDR-based Hierarchical Representations



- Adapted from Serre et al (2005)



This view distinguishes *compositionality* from *distributedness*.