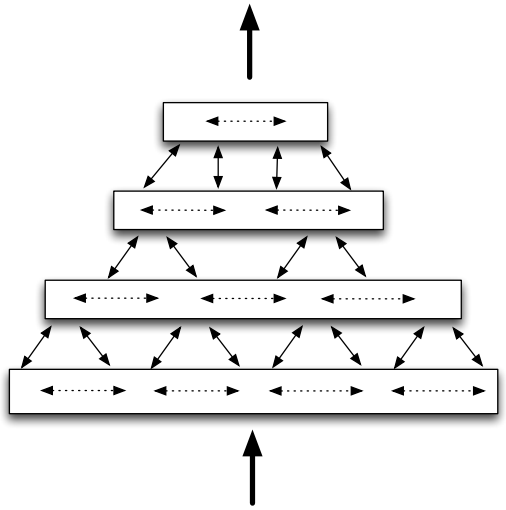
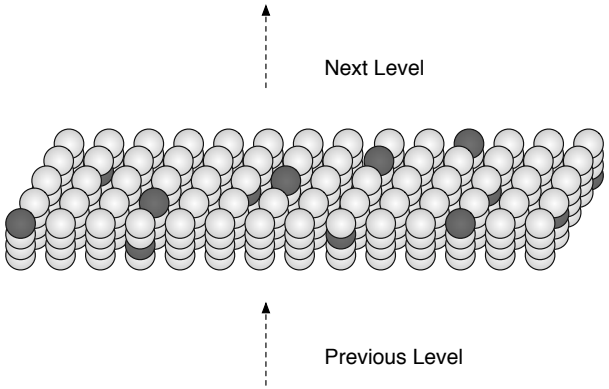
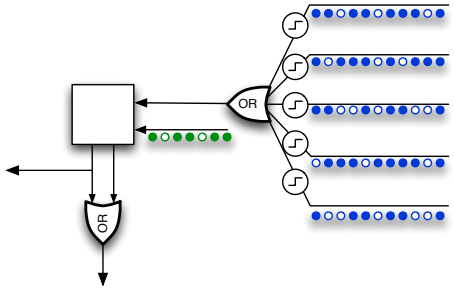
CLA Deep Dive

Subutai Ahmad
subutai@numenta.org

Lots Of Stuff We Could Talk About







But, We Can’t Cover It All

Sample CLA “Quiz” Questions

- Is an untrained CLA spatial pooler just a “random hash”? Why or why not? What happens to the output of the spatial pooler if you randomly change one bit in the input?
- When creating an SDR, what's the difference between picking "5 columns out of 50" vs "50 out of 500"? Both have 10% sparsity.
- How do the SDR representations of input A and input B, compare with input A overlapped with B?
- Suppose a temporal pooler has learned the following two sequences: ABCDE and ABCDF. You now present the sequence ABCD - what will be predicted next? What is the representation of the temporal pooler at this point in time?
- Suppose a temporal pooler has only learned the following two sequences: ABCDE and FGCDH. You now present the sequence FGCD - what will be predicted next? Suppose you present the sequence CD - what will be predicted next? What is the representation of the temporal pooler at this point in time.

Agenda

- One SDR property in depth
  - 15 minutes, slides + whiteboard
  - Key to understanding temporal predictions
  - Key to getting a hierarchy working
  - Key to invariances
  - Will help you understand CLA parameters and numbers better
  - Not covered in depth in talks or whitepaper

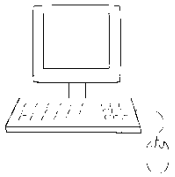
- Open discussion and questions about CLA

Sparse Distributed Representations (SDRs)

- Many bits (thousands)
- Few active
- Example: 2,000 bits, 2% active

  0100000000000000000100000000000000000000000000000000010000.....01000
- Each bit has semantic meaning

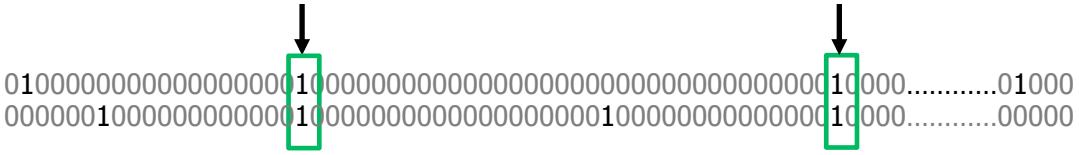
Dense Representations



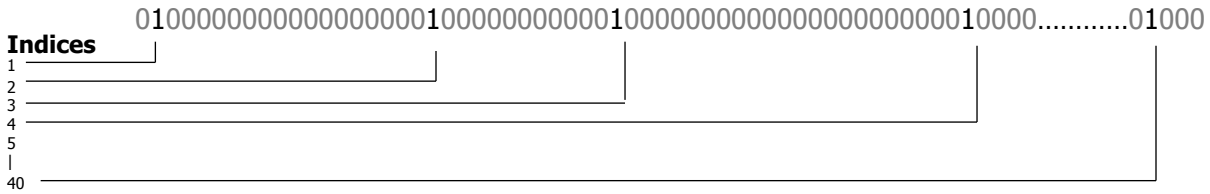
- Few bits (8 to 128)
- All combinations of 1's and 0's
- Example: 8 bit ASCII
- Bits have no inherent meaning

SDR Properties

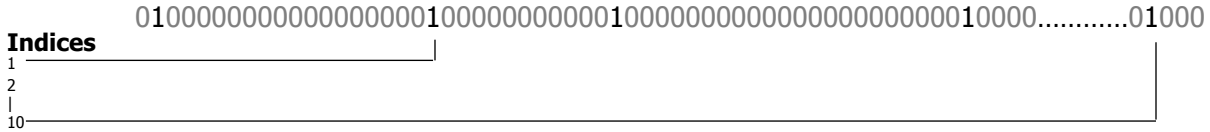
1) **Compare**:
shared bits = semantic similarity



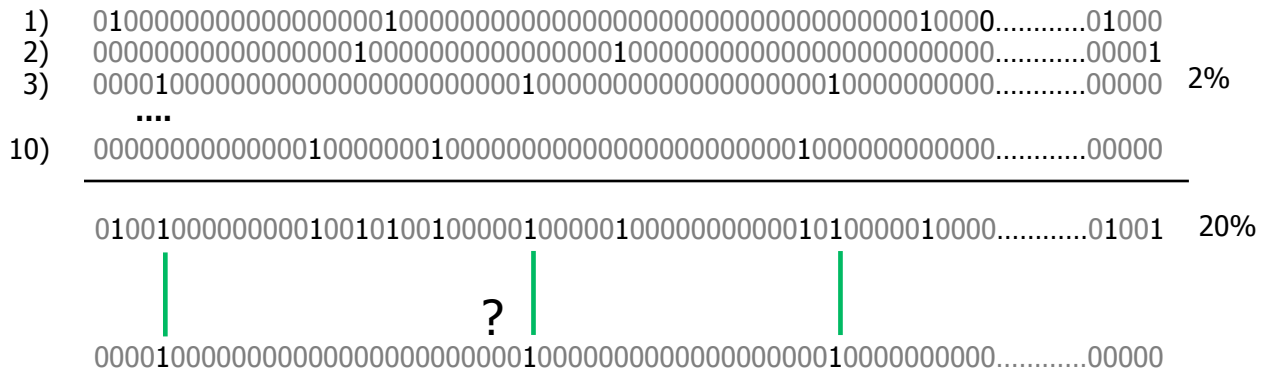
2) **Store**:
store indices of active bits



2b) **Store**:
subsampling is OK

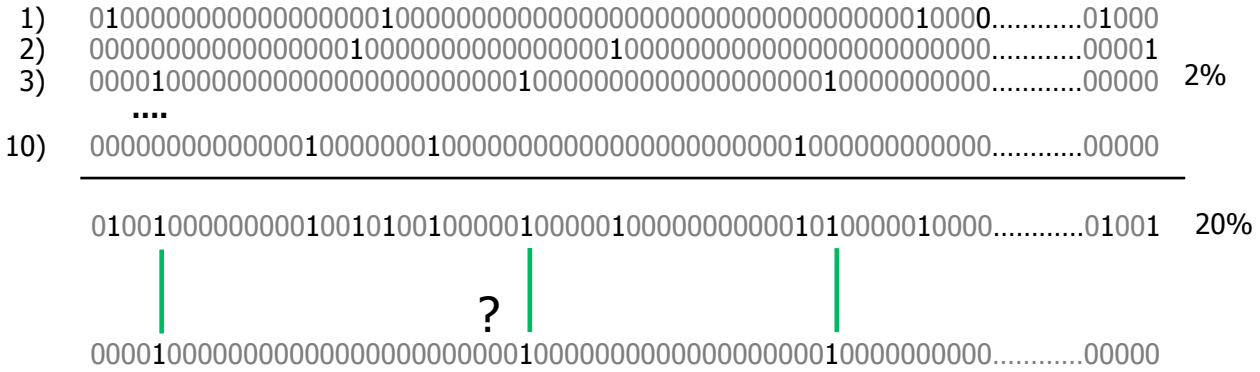


3) **Union**:
can detect membership
in union

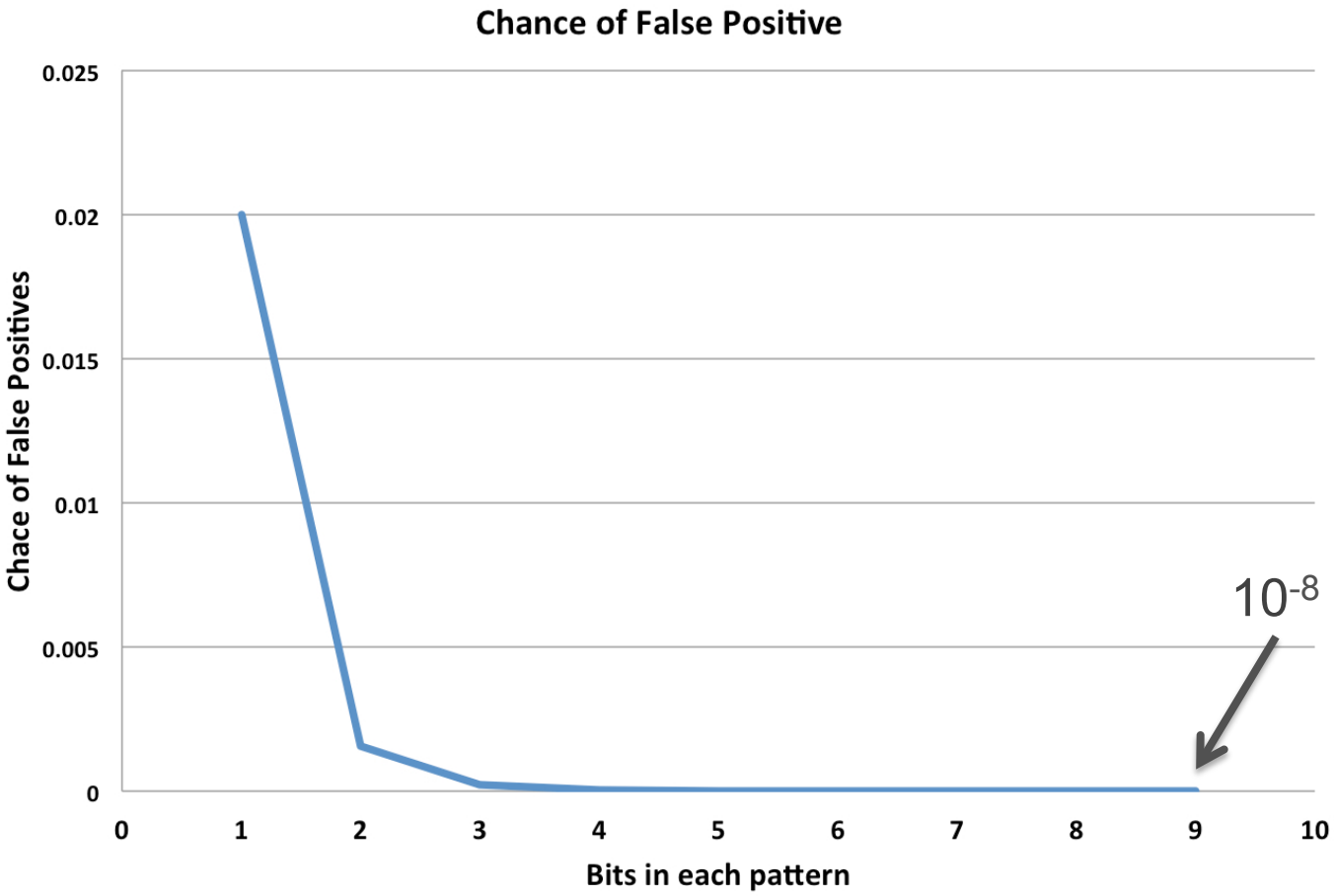


3) **Union**:

can detect membership in union

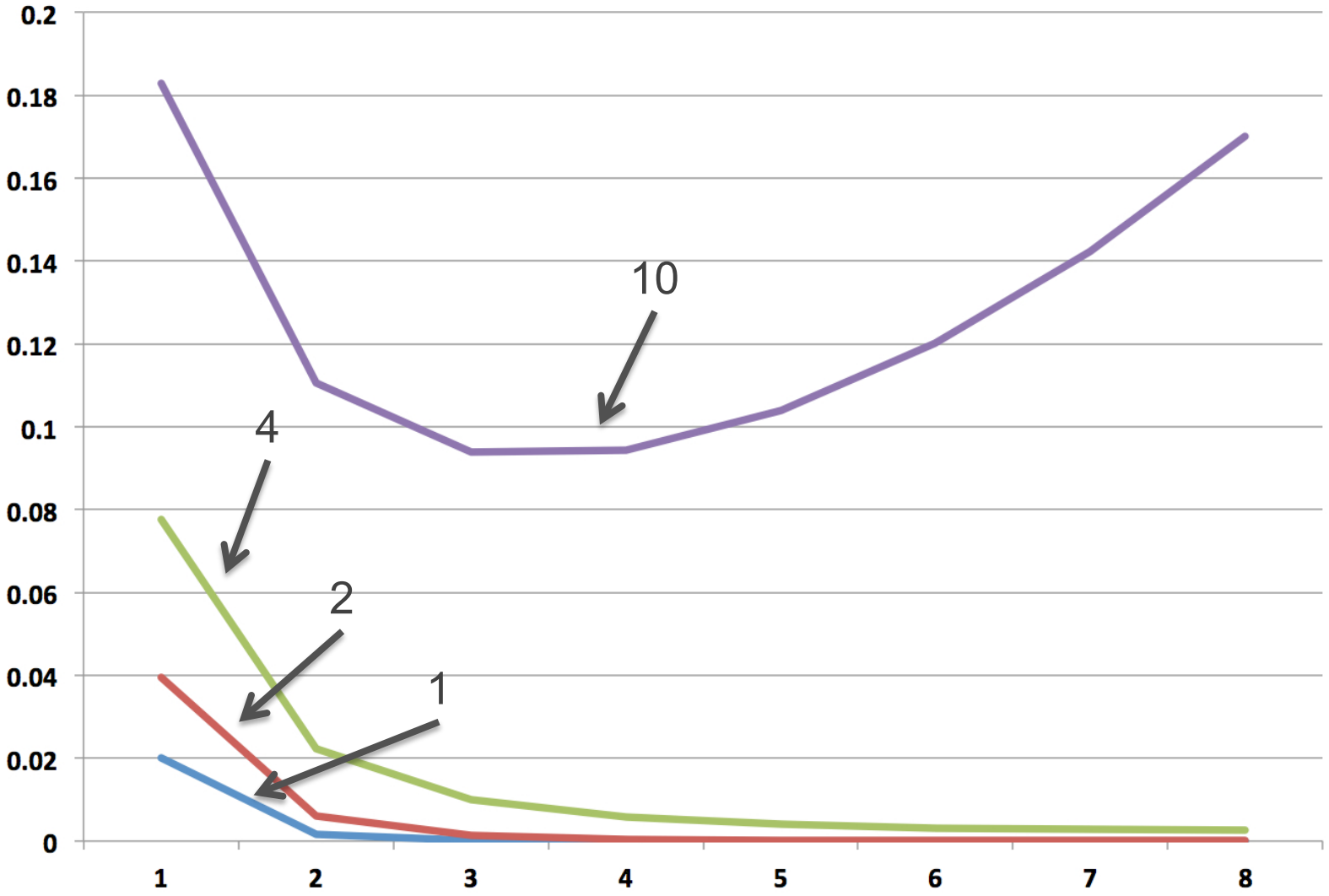


ON bits should be “reasonably high”



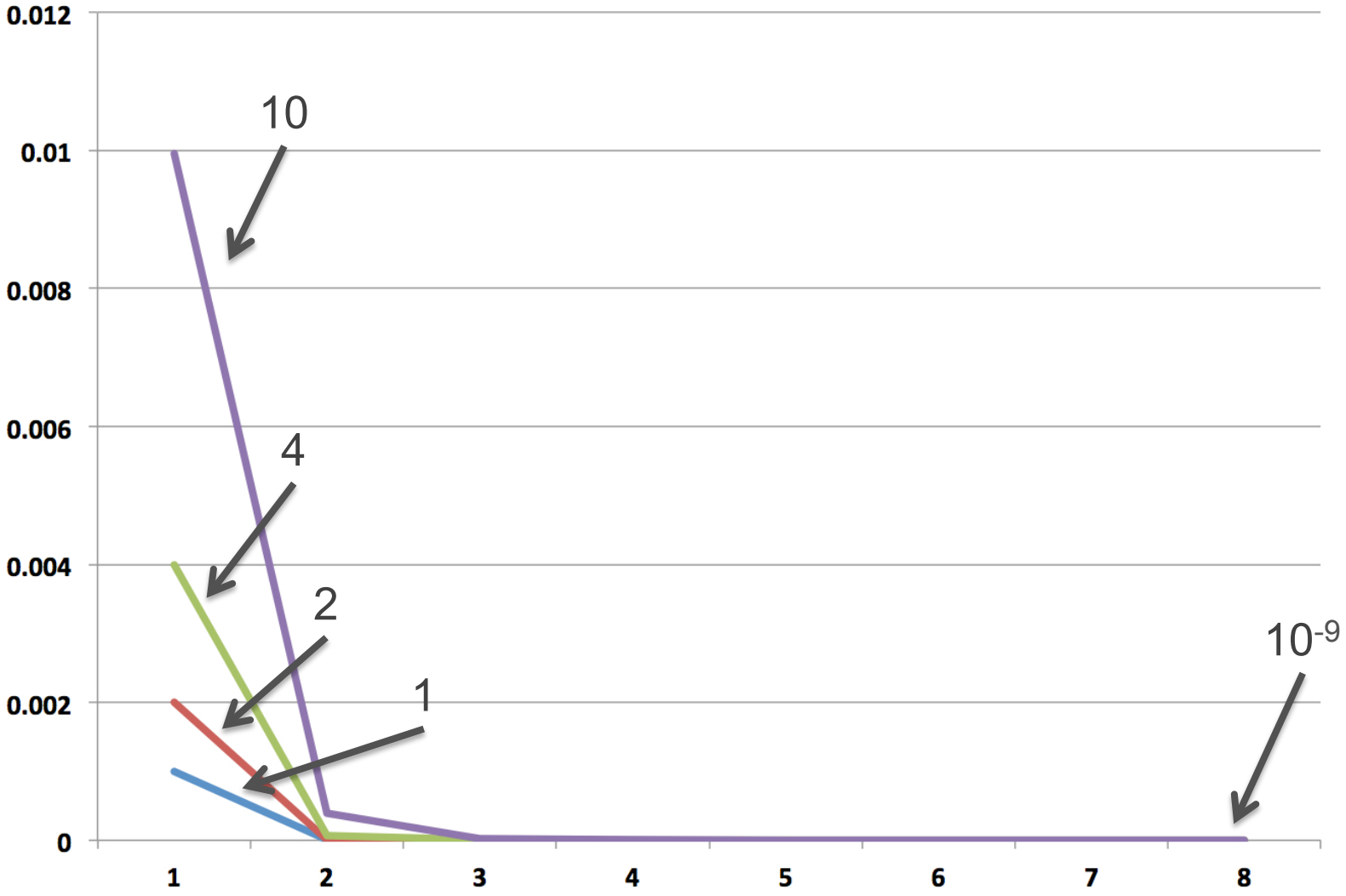
columns = 50

But, # of patterns at a time is low



columns = 50

Answer: just increase number of columns!



columns = 1000

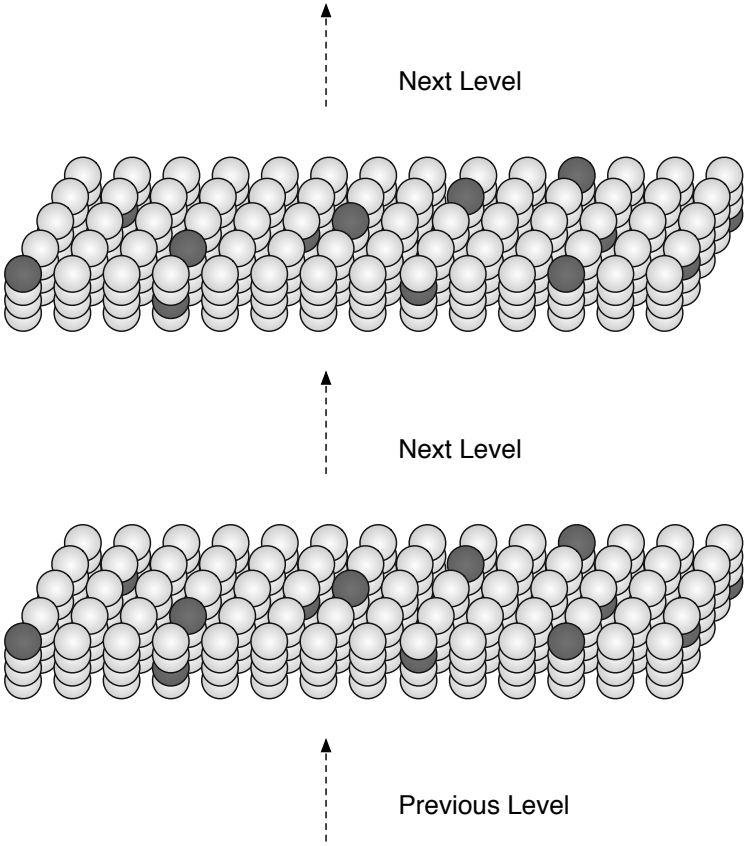
Superposition of patterns

- Through superposition you can store completely separate patterns in the same vector with minimal chance of false positives
- Two variables: # ON bits, # columns
  - ON bits / columns = sparsity
- Sparsity controls how many simultaneous patterns you can store
  - Constraints:
    - Need to have “reasonable” number of ON bits
    - Need a high number of dimensions (columns)
  - Very low sparsity (e.g. 0.01%)
    - Can store tons of patterns but number of columns must be high (e.g. 50,000)

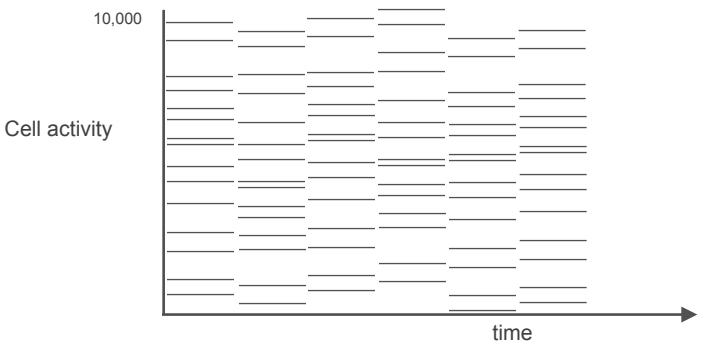
- In NuPIC, we’ve chosen 40 out of 2048
  - Code exercise: figure out how many random patterns we can superimpose

Superposition and the Temporal Pooler

Superposition and hierarchy

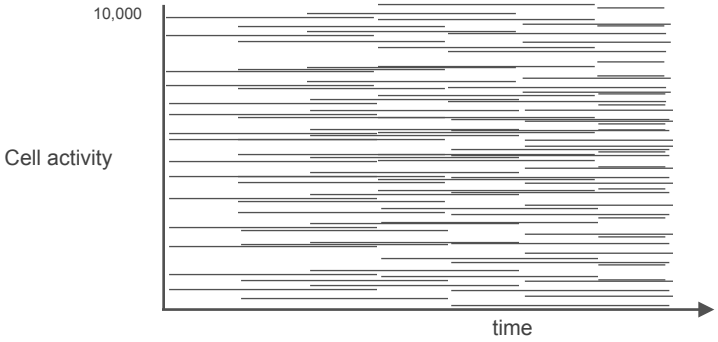


Superposition and hierarchy



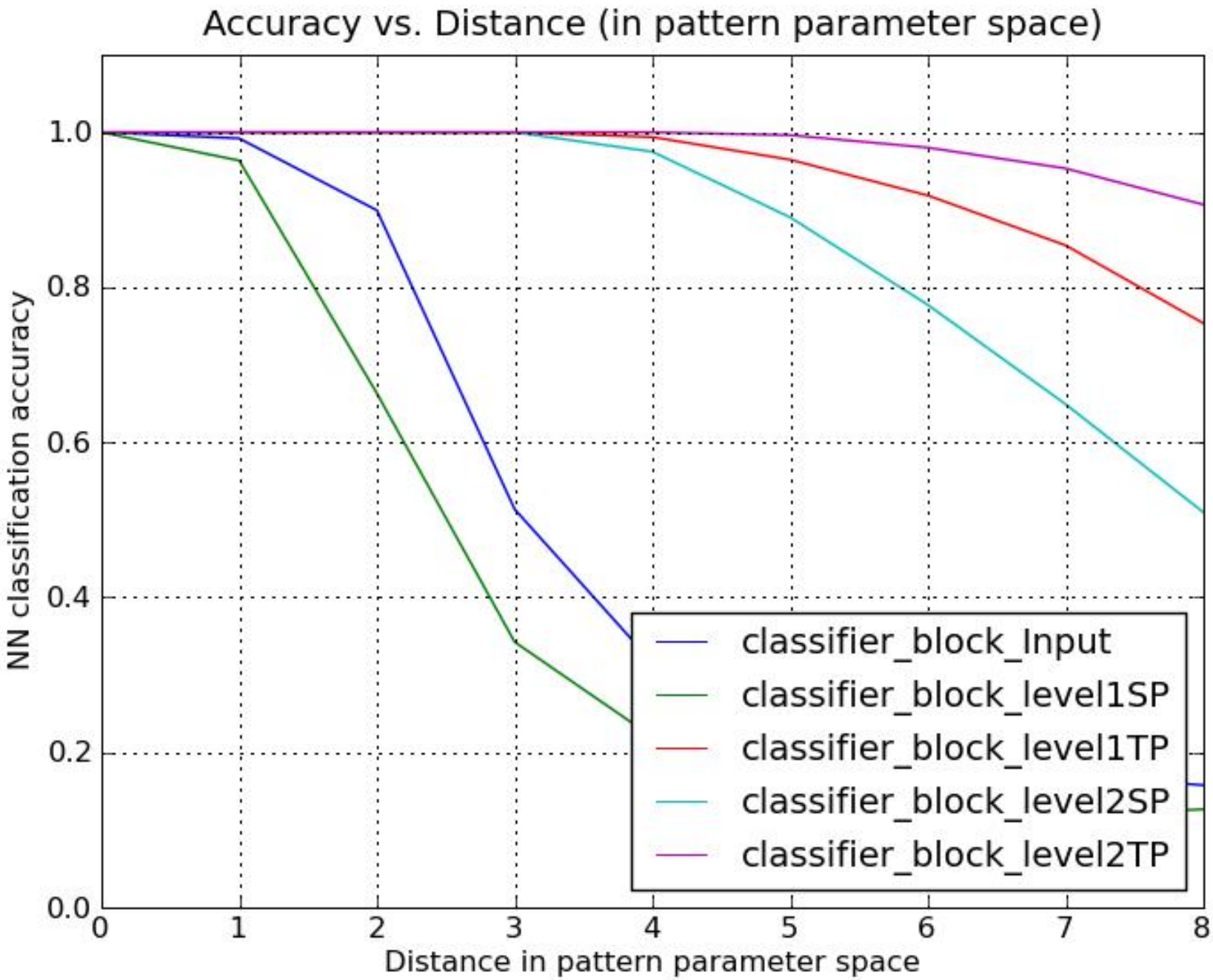
Output before temporal pooling*

- Output of temporal pooler is activity on all 10,000 cells
- These 10,000 cells are input to next level's spatial pooler



Output after learning temporal pooling*

- Temporal pooling causes cells to be active over longer periods of time
- Overlapping activity leads to temporal stability of the next spatial pooler


Accuracy vs. Distance (in pattern parameter space)
NN classification accuracy
Distance in pattern parameter space
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5
6
7
8
classifier_block_Input
classifier_block_level1SP
classifier_block_level1TP
classifier_block_level2SP
classifier_block_level2TP

Agenda

- One SDR property in depth
  - 15 minutes, slides + whiteboard
  - Key to understanding temporal predictions
  - Key to getting a hierarchy working
  - Key to invariances
  - Will help you understand CLA parameters and numbers better
  - Not covered in depth in talks or whitepaper

- Open discussion and questions about CLA