Learning in Boltzmann Machines

Leonardo & Mark

A Learning Algorithm for Boltzmann Machines*

DAVID H. ACKLEY
GEOFFREY E. HINTON

Computer Science Department
Carnegie-Mellon University

TERRENCE J. SEJNOWSKI

Biophysics Department
The Johns Hopkins University

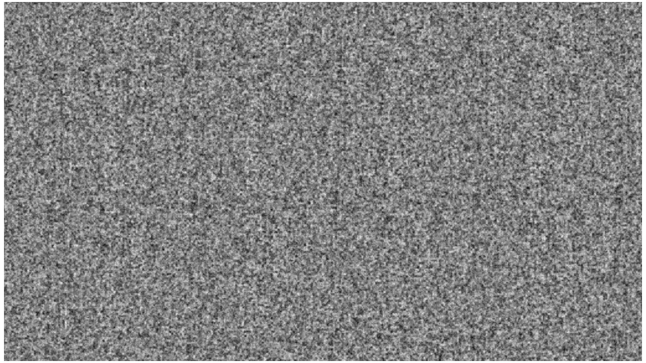
Outline

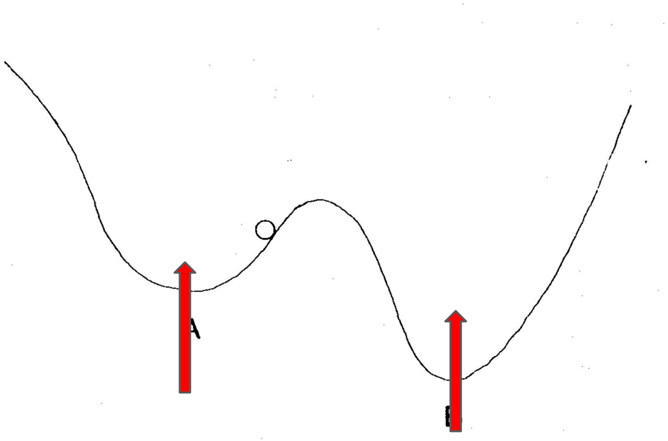
- Part I: The Model
  - The Boltzmann Machine
  - The Encoder Task
- Part II: Reproduction
  - Reproducing the 4-2-4 Encoder
  - Necessary Tricks
- Part III: Investigations
  - Increase the number of hidden units
  - Robustness of Encoding
- Summary and Outlook

Part I: The Model

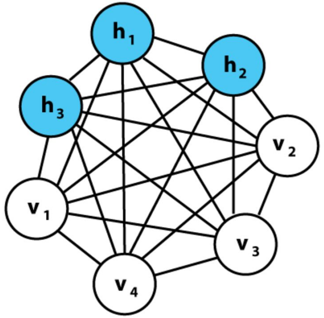
The Boltzmann Machine

- Energy-Based, stochastic Model
- recurrent connections
- add hidden units
- Goal: Learn an Input Distribution
- in other words: Learn to generate Data that was not in the training set





Hinton 1986



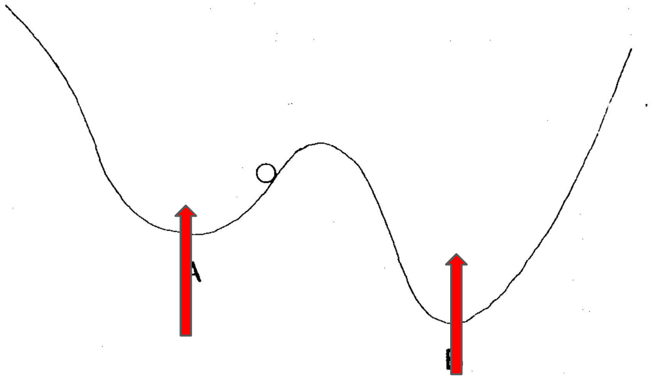
(wikipedia)

Training the BM

- Learn an Input Distribution
- minimize divergence between two distributions
- surprisingly, this leads to two local, Hebbian rules
- Wake phase with the data
- Sleep phase with 'dreams'

$$Q_0 || Q_{BM}$$

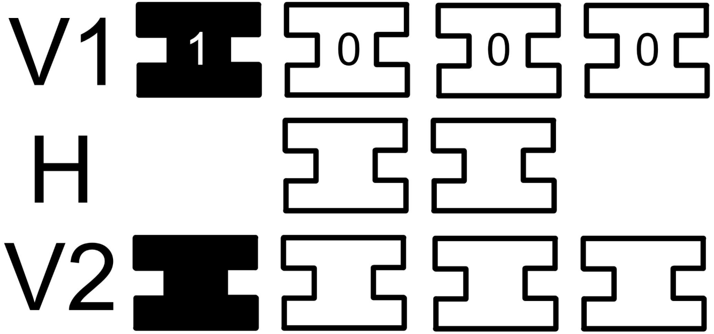
$$\langle s_i s_j \rangle_{data} - \langle s_i s_j \rangle_{model}$$



Hinton 1986

The Encoder Task: Learning Efficient Communication in Boltzmann Machines

- Encoder Task: Two networks have learned concepts separately
- Goal: **V1 = V2**
  - Hidden Layer acts as intermediate channel
- Architecture: 4-2-4 encoder
  - Only one neuron in active (sparse representation)

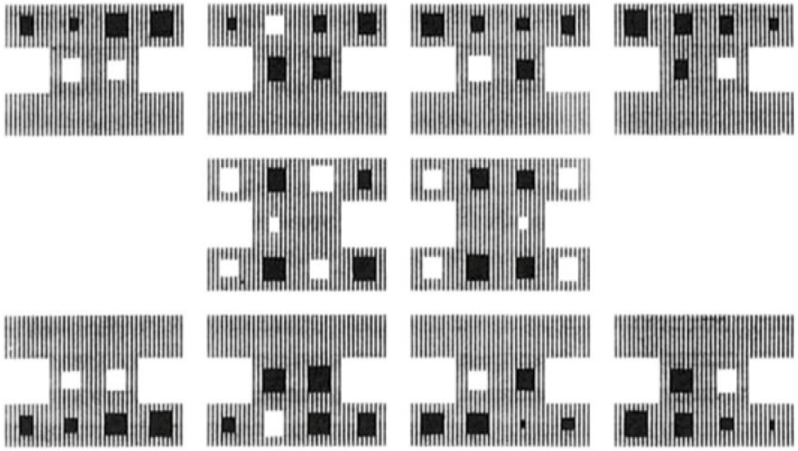


The Encoder Task: Learning Efficient Communication in Boltzmann Machines

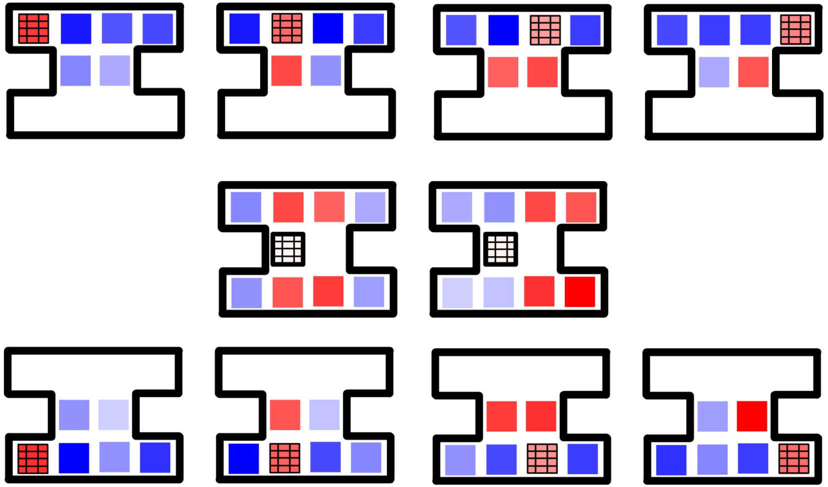
- Communication between two visible Layers
- Hidden Layer as intermediate channel
- Statistics of the ‘Brain Regions’
  - only one active neuron within region
  - correlation between neurons
- Goal: learn to communicate via binary encoding
-

Part II: Reproduction

Reproducing the 4-2-4 Encoder

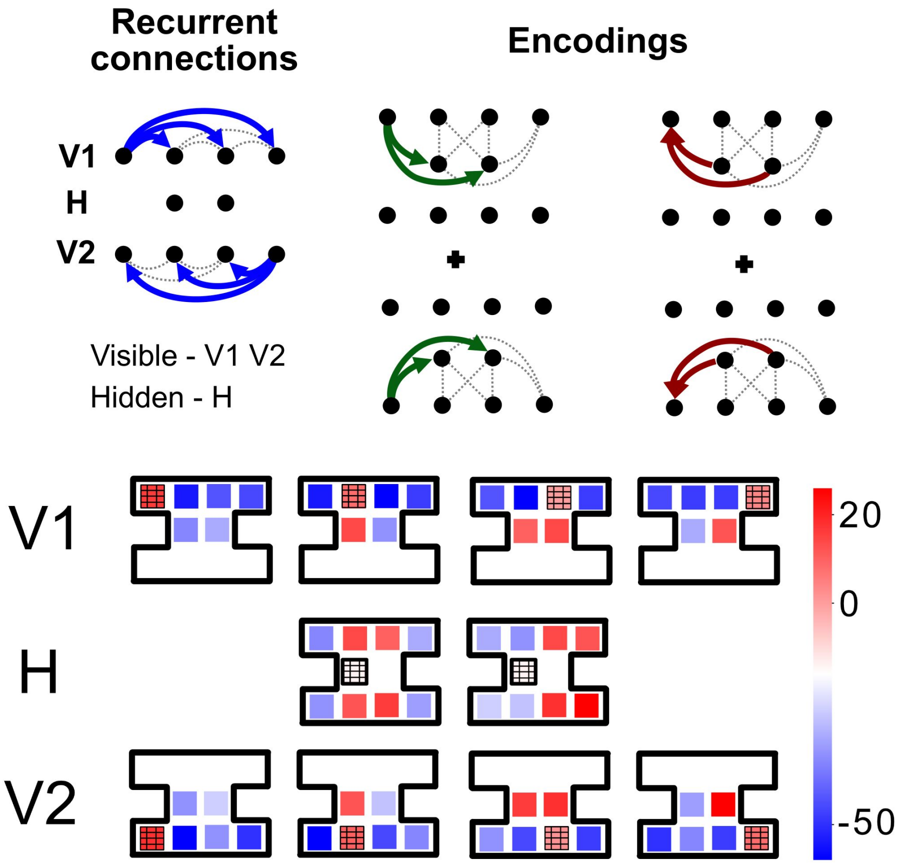


Hinton 1985



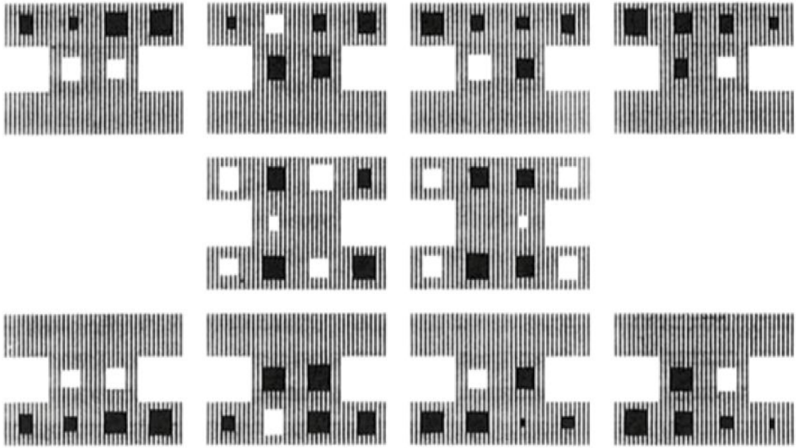
What are we seeing?

- Visible units (V) have recurrent connections within the layer
- Encodings between visible and hidden (H) units
- Recursive plot
  - Small grid recursively represents the connectivity of the neuron to the entire network
- Square Color: Size and sign of the weights
- Biases are shaded



what are we seeing?

- recursive plot (every neuron has the structure of the network)
- boxes are weights
- b&w are +-
- biases are shaded

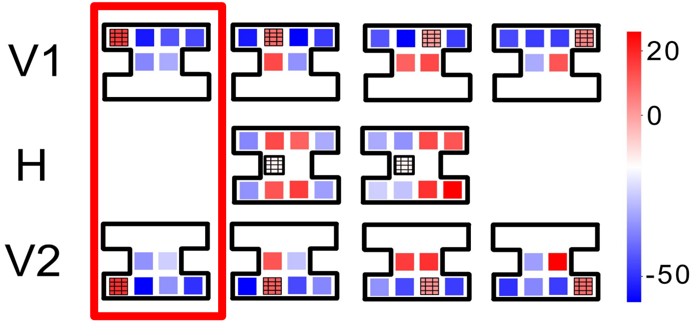


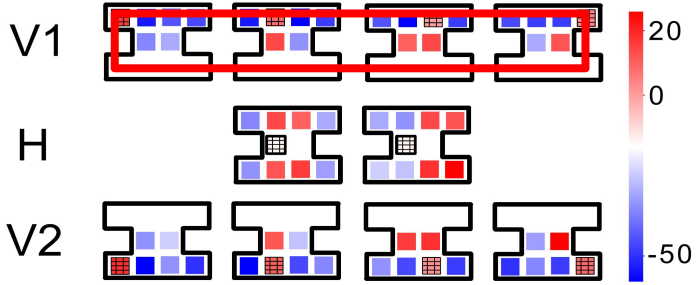
1. the binary encodings (hamming distance)
2. symmetry between V1 and V2

1. receptive fields
2. recurrent connections

Binary Encodings

- V1 = V2
- Weights between V and H correspond to activity of hidden units
- Distinct messages => Distinct encodings
- Observe four distinct encodings
    - (11) (10) (01) (00)





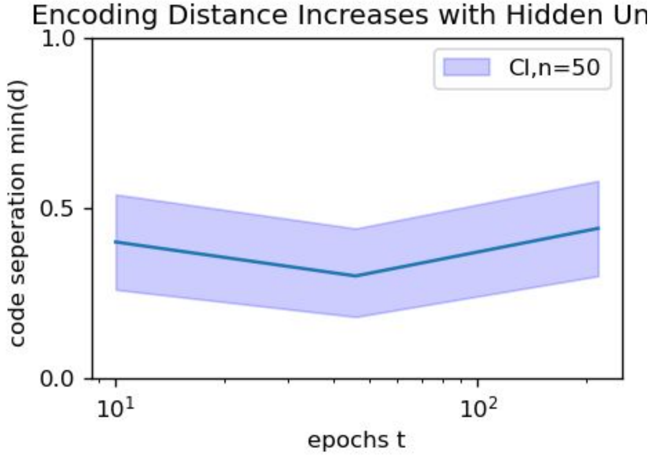
Binary Encodings 4-2-4

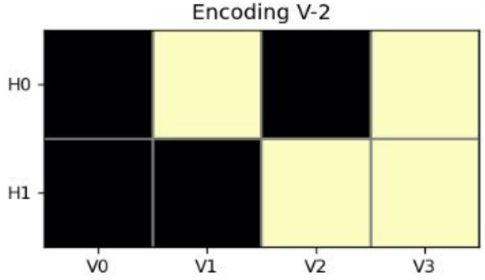
- Use hamming distance to quantify encodings

- Ideally Hamming distance should reach 1
  - Codes entirely separated: min(d) = 1
  - Codes have similarities: min(d) = 0

- Statistics: 95% Confidence intervals & 50 samples

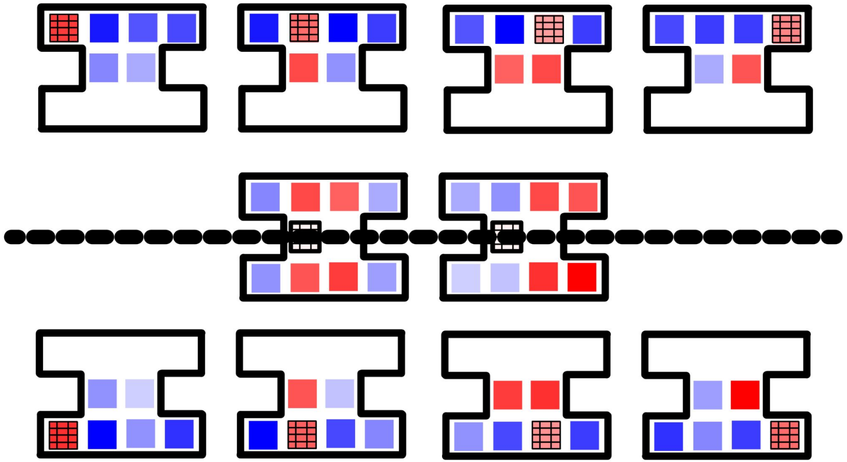
- No perfect encoding:
  - To less epochs
  - Stochastic nature of BM
  - Noise



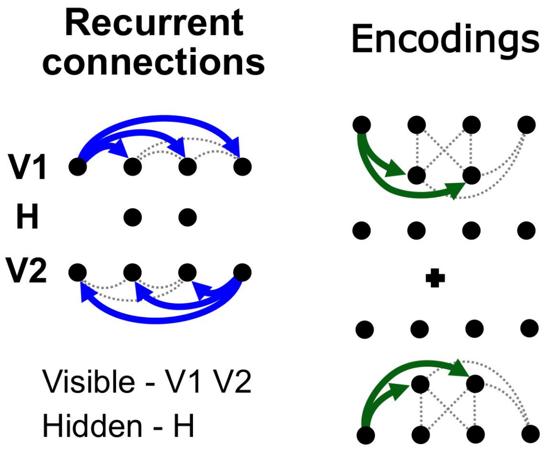


Symmetry in the 4-2-4 Encoder

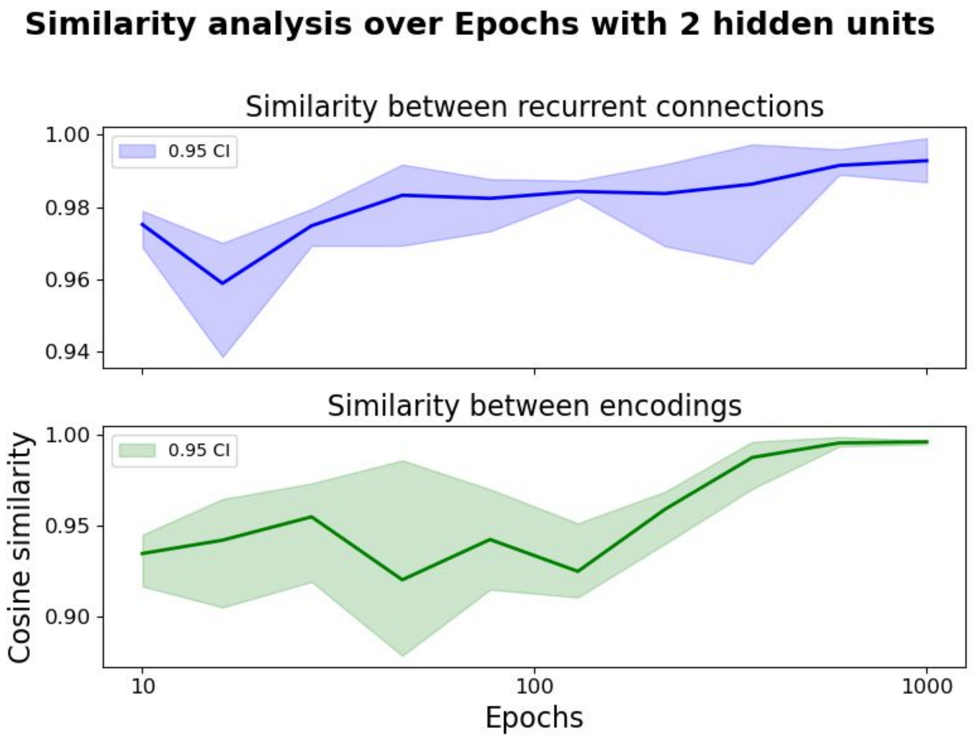
- Consistent communication between V1 and V2
- Mirror symmetry in the weight matrix
- Symmetry ensures that corresponding visible neurons activate the same hidden neurons



Symmetry between V1 and V2



- Measure the symmetry via cosine similarities between
  - Recurrent connections
  - V-H Encodings



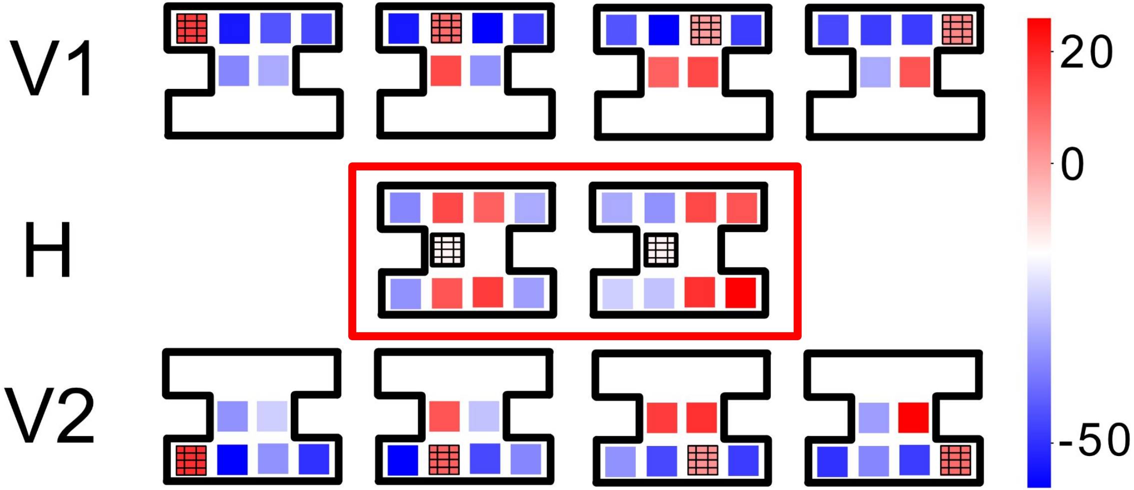
Samples = 3

Non-Obvious Implementation Knowledge

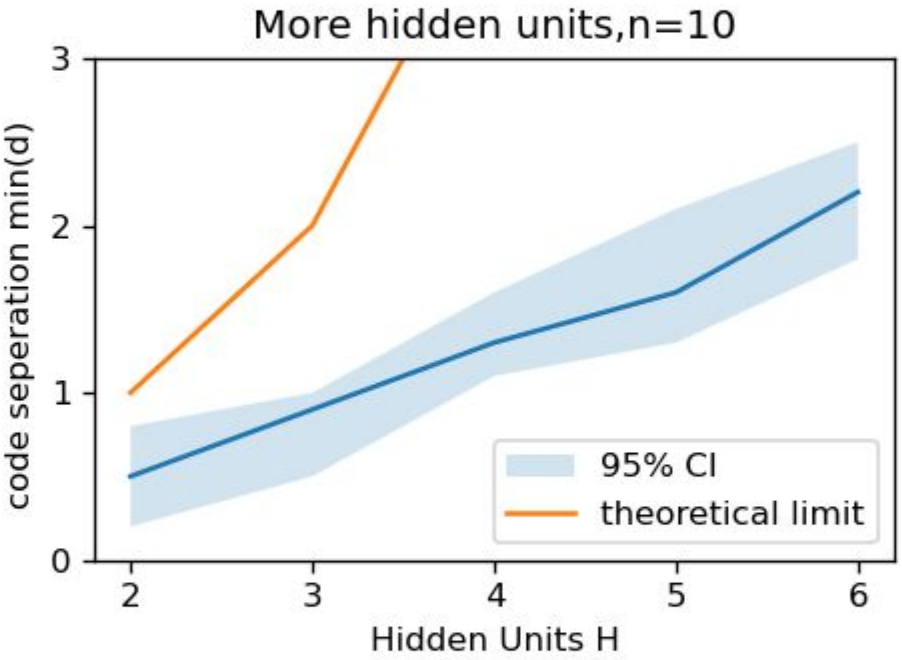
- Network does not work with (+1,-1) neuron states, only (1,0)
- Need to train on noisy data, otherwise weights blow up
- Annealing scheme to get to equilibrium
  - different between train and test
- Perform discrete weight updates instead of doing gradient descent

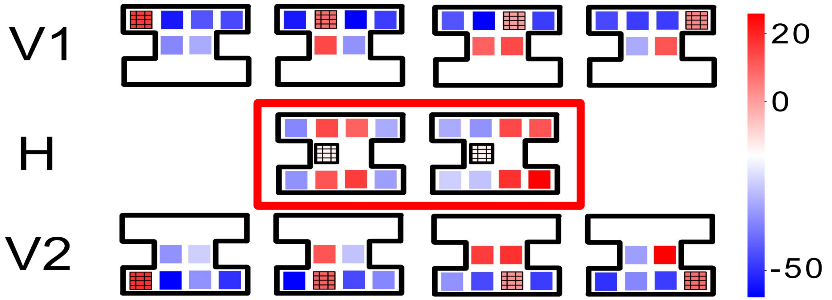
Part III: Investigation

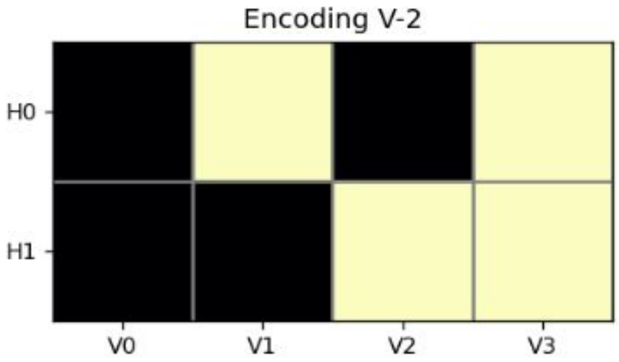
Increase the number of hidden units



Effects of Hidden Units on Separation





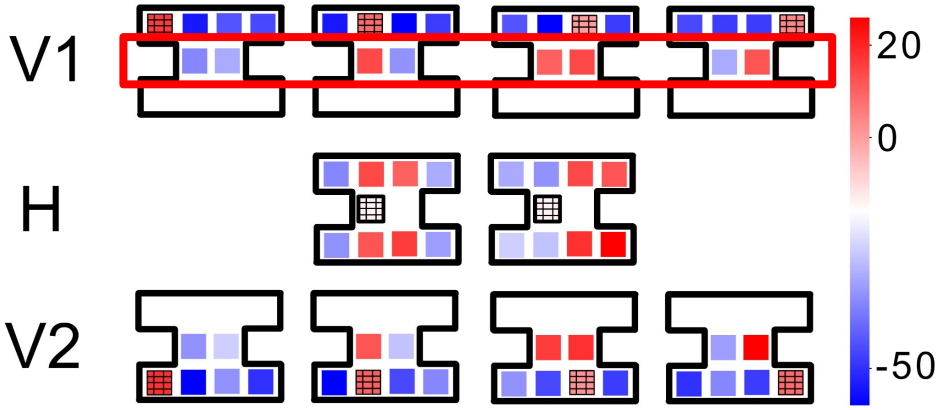


Robustness of Encoding

Why do (+1,-1) activations not work?

Tentative Explanation:

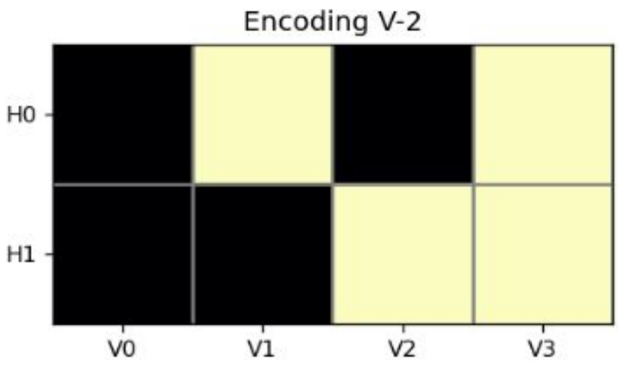
(-1) neurons still ‘send’ their encoding -> interference

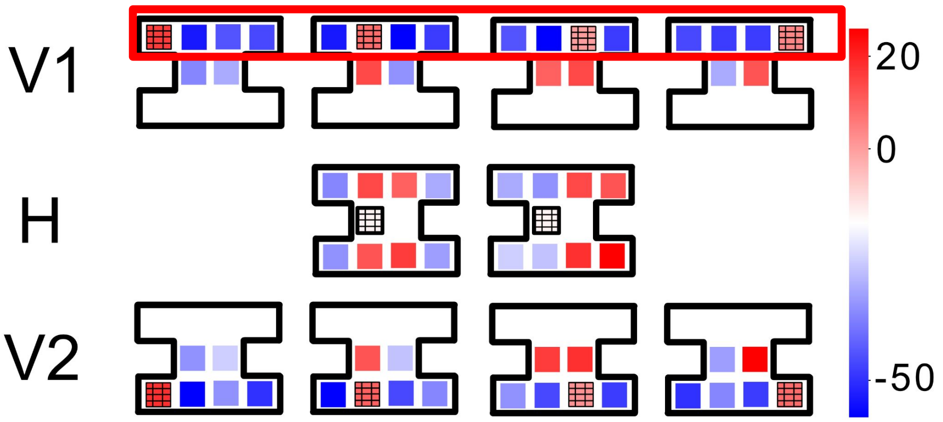


(+1,-1) is standard in Hopfield networks! Cherry picking?

Robustness of Encoding

- Recurrent Connections encode statistics within V1/V2
- could hidden units compensate for this?
- step towards Restricted BM





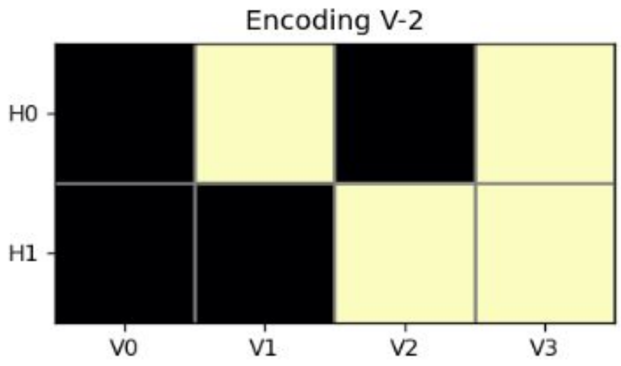
Recurrent Connections

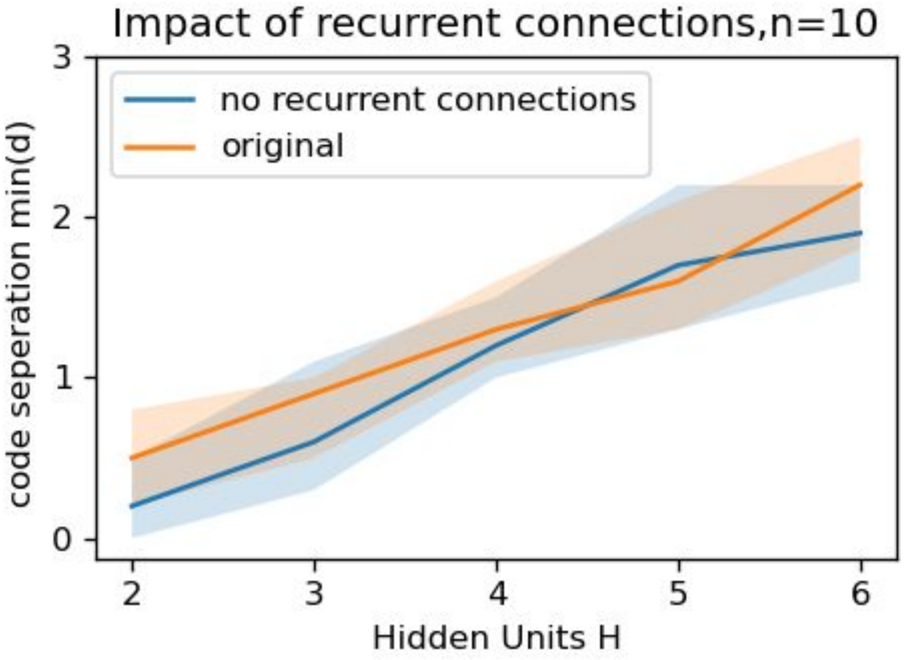
Our Hypothesis:

Removing recurrent will break encodings, a lot of hidden units might be able to compensate

What happened instead:

Encodings still work, maybe even better





Outlook & Conclusions

Outlook:

- Explore more complex datasets, more interesting statistics (e.g. correlations)
- Information Processing, such as Mutual Information (or PID)

Conclusions:

- Hidden units and their representations are important for Boltzmann machines
- sometimes machine learning takes a lot of practical knowledge