A simple and efficient way to store many messages using neural cliques

V. Gripon and C. Berrou

Télécom Bretagne
CNRS Lab-STICC

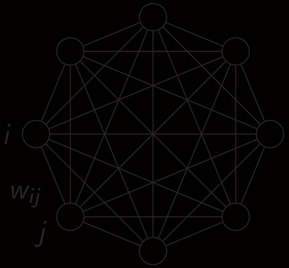
April 15, 2011

Hopfield neural networks

Associative memories: principle

- Learn M messages,
- Retrieve a learnt message in presence of erasures or errors.

The Hopfield network



Example with $n = 8$ neurons.

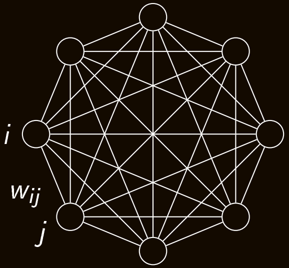
- Learning (static): M binary ($\{-1; 1\}$) messages $\mathbf{d}^m$: $w_{ij} = \sum_{m=1; i,j \leq n}^{M} d_i^m d_j^m$ (w_{ij} is specified on $M+1$ levels),
- Retrieving (iterative): repeat $\forall i, v_i \leftarrow \mathrm{sgn}(\sum_{j \neq i} v_j w_{ij})$.

Hopfield neural networks

Associative memories: principle

- Learn M messages,
- Retrieve a learnt message in presence of erasures or errors.

The Hopfield network



Example with $n = 8$ neurons.

- Learning (static): M binary ($\{-1;1\}$) messages $\mathbf{d}^m$: $w_{ij} = \sum_{m=1; i,j \leq n}^{M} d_i^m d_j^m$ (w_{ij} is specified on $M+1$ levels),
- Retrieving (iterative): repeat $\forall i, v_i \leftarrow \mathrm{sgn}(\sum_{j \neq i} v_j w_{ij})$.

Bounds and limitations

Bounds (n neurons)

- Diversity (number of learnt messages): $\frac{n}{2log(n)}$ sublinear,
- Capacity: $\frac{n^2}{2log(n)}$ subquadratic,
- Binary information used: $\frac{n(n-1)}{2}log_2(M+1)$,
- $\Rightarrow$ Efficiency $\frac{\text{capacity}}{\text{binary information used}} \approx \frac{1}{log(n)log_2(M+1)} \xrightarrow[n\to\infty]{} 0$.

Limitations

- As a plausible model:
  - Sensitive connections, fully interconnected network, retrieving requires a very large number of iterations...
- As an associative memory:

Bounds and limitations

Bounds (n neurons)

- Diversity (number of learnt messages): $\frac{n}{2log(n)}$ sublinear,
- Capacity: $\frac{n^2}{2log(n)}$ subquadratic,
- Binary information used: $\frac{n(n-1)}{2} log_2(M+1)$,
- $\Rightarrow$ Efficiency $\frac{\text{capacity}}{\text{binary information used}} \approx \frac{1}{log(n) log_2(M+1)} \xrightarrow[n\to\infty]{} 0$.

Limitations

- As a plausible model:
  - Sensitive connections, fully interconnected network, retrieving requires a very large number of iterations...
- As an associative memory:

Bounds and limitations

Bounds (n neurons)

- Diversity (number of learnt messages): $\frac{n}{2log(n)}$ sublinear,
- Capacity: $\frac{n^2}{2log(n)}$ subquadratic,
- Binary information used: $\frac{n(n-1)}{2} log_2(M+1)$,
- $\Rightarrow$ Efficiency $\frac{\text{capacity}}{\text{binary information used}} \approx \frac{1}{log(n) log_2(M+1)} \xrightarrow[n \to \infty]{} 0$.

Limitations

- As a plausible model:
  - Sensitive connections, fully interconnected network, retrieving requires a very large number of iterations...
- As an associative memory:

Bounds (n neurons)

- Diversity (number of learnt messages): $\frac{n}{2log(n)}$ sublinear,
- Capacity: $\frac{n^2}{2log(n)}$ subquadratic,
- Binary information used: $\frac{n(n-1)}{2} log_2(M+1)$,
- $\Rightarrow$ Efficiency $\frac{\text{capacity}}{\text{binary information used}} \approx \frac{1}{log(n) log_2(M+1)} \xrightarrow[n \to \infty]{} 0$.

Limitations

- As a plausible model:
  - Sensitive connections, fully interconnected network, retrieving requires a very large number of iterations...
- As an associative memory:

Bounds and limitations

Bounds (n neurons)

- Diversity (number of learnt messages): $\frac{n}{2log(n)}$ sublinear,
- Capacity: $\frac{n^2}{2log(n)}$ subquadratic,
- Binary information used: $\frac{n(n-1)}{2} log_2(M+1)$,
- $\Rightarrow$ Efficiency $\frac{\text{capacity}}{\text{binary information used}} \approx \frac{1}{log(n)log_2(M+1)} \xrightarrow[n\to\infty]{} 0$.

Limitations

- As a plausible model:
  - Sensitive connections, fully interconnected network, retrieving requires a very large number of iterations...
- As an associative memory:
  - Very limited diversity, message length, network size and diversity are bound together, learning both messages and their opposite...

Bounds and limitations

Bounds (n neurons)

- Diversity (number of learnt messages): $\frac{n}{2log(n)}$ sublinear,
- Capacity: $\frac{n^2}{2log(n)}$ subquadratic,
- Binary information used: $\frac{n(n-1)}{2} log_2(M+1)$,
- $\Rightarrow$ Efficiency $\frac{\text{capacity}}{\text{binary information used}} \approx \frac{1}{log(n) log_2(M+1)} \xrightarrow[n \to \infty]{} 0$.

Limitations

- As a plausible model:
  - Sensitive connections, fully interconnected network, retrieving requires a very large number of iterations...
- As an associative memory:
  - Very limited diversity, message length, network size and diversity are bound together, learning both messages and their opposite...

Bounds and limitations

Bounds (n neurons)

- Diversity (number of learnt messages): $\frac{n}{2log(n)}$ sublinear,
- Capacity: $\frac{n^2}{2log(n)}$ subquadratic,
- Binary information used: $\frac{n(n-1)}{2} log_2(M+1)$,
- $\Rightarrow$ Efficiency $\frac{\text{capacity}}{\text{binary information used}} \approx \frac{1}{log(n) log_2(M+1)} \xrightarrow[n \to \infty]{} 0$.

Limitations

- As a plausible model:
  - Sensitive connections, fully interconnected network, retrieving requires a very large number of iterations...
- As an associative memory:
  - Very limited diversity, message length, network size and diversity are bound together, learning both messages and their opposite...

Bounds and limitations

Bounds (n neurons)

- Diversity (number of learnt messages): $\frac{n}{2log(n)}$ sublinear,
- Capacity: $\frac{n^2}{2log(n)}$ subquadratic,
- Binary information used: $\frac{n(n-1)}{2} log_2(M+1)$,
- $\Rightarrow$ Efficiency $\frac{\text{capacity}}{\text{binary information used}} \approx \frac{1}{log(n)log_2(M+1)} \xrightarrow[n\to\infty]{} 0$.

Limitations

- As a plausible model:
  - Sensitive connections, fully interconnected network, retrieving requires a very large number of iterations. . .
- As an associative memory:
  - Very limited diversity, message length, network size and diversity are bound together, learning both messages and their opposite. . .

Bounds and limitations

Bounds (n neurons)

- Diversity (number of learnt messages): $\frac{n}{2log(n)}$ sublinear,
- Capacity: $\frac{n^2}{2log(n)}$ subquadratic,
- Binary information used: $\frac{n(n-1)}{2} log_2(M+1)$,
- $\Rightarrow$ Efficiency $\frac{\text{capacity}}{\text{binary information used}} \approx \frac{1}{log(n)log_2(M+1)} \xrightarrow[n \to \infty]{} 0$.

Limitations

- As a plausible model:
  - Sensitive connections, fully interconnected network, retrieving requires a very large number of iterations...
- As an associative memory:
  - Very limited diversity, message length, network size and diversity are bound together, learning both messages and their opposite...

Bounds and limitations

Bounds (n neurons)

- Diversity (number of learnt messages): $\frac{n}{2log(n)}$ sublinear,
- Capacity: $\frac{n^2}{2log(n)}$ subquadratic,
- Binary information used: $\frac{n(n-1)}{2} log_2(M+1)$,
- $\Rightarrow$ Efficiency $\frac{\text{capacity}}{\text{binary information used}} \approx \frac{1}{log(n)log_2(M+1)} \xrightarrow[n \to \infty]{} 0$.

Limitations

- As a plausible model:
  - Sensitive connections, fully interconnected network, retrieving requires a very large number of iterations. . .
- As an associative memory:
  - Very limited diversity, message length, network size and diversity are bound together, learning both messages and their opposite. . .

Bounds and limitations

Bounds (n neurons)

- Diversity (number of learnt messages): $\frac{n}{2log(n)}$ sublinear,
- Capacity: $\frac{n^2}{2log(n)}$ subquadratic,
- Binary information used: $\frac{n(n-1)}{2} log_2(M+1)$,
- $\Rightarrow$ Efficiency $\frac{\text{capacity}}{\text{binary information used}} \approx \frac{1}{log(n)log_2(M+1)} \xrightarrow[n \to \infty]{} 0$.

Limitations

- As a plausible model:
  - Sensitive connections, fully interconnected network, retrieving requires a very large number of iterations...
- As an associative memory:
  - Very limited diversity, message length, network size and diversity are bound together, learning both messages and their opposite...

Bounds and limitations

Bounds (n neurons)

- Diversity (number of learnt messages): ~~sublinear~~, quadratic
- Capacity: $\frac{n^2}{2log(n)}$ subquadratic,
- $\Rightarrow$ Efficiency $\frac{\text{capacity}}{\text{binary information used}} \approx \frac{1}{log(n)log_2(M+1)} \xrightarrow[n\to\infty]{} 0$.

Limitations

- As a plausible model:
  - Sensitive connections, fully interconnected network, retrieving requires a very large number of iterations...
- As an associative memory:
  - Very limited diversity, message length, network size and diversity are bound together, learning both messages and their opposite...

Bounds and limitations

Bounds (n neurons)

- Diversity (number of learnt messages): ~~sublinear~~, quadratic
- Capacity: ~~subquadratic~~, quadratic
- $\Rightarrow$ Efficiency $\frac{\text{capacity}}{\text{binary information used}} \approx \frac{1}{log(n) log_2(M+1)} \xrightarrow[n\to\infty]{} 0.$

Limitations

- As a plausible model:
  - Sensitive connections, fully interconnected network, retrieving requires a very large number of iterations...
- As an associative memory:
  - Very limited diversity, message length, network size and diversity are bound together, learning both messages and their opposite...

Bounds and limitations

Bounds (n neurons)

- Diversity (number of learnt messages): ~~sublinear~~, quadratic
- Capacity: ~~subquadratic~~, quadratic
- $\Rightarrow$ Efficiency $\frac{\text{capacity}}{\text{binary information used}} \approx 1$.

Limitations

- As a plausible model:
  - Sensitive connections, fully interconnected network, retrieving requires a very large number of iterations...
- As an associative memory:
  - Very limited diversity, message length, network size and diversity are bound together, learning both messages and their opposite...

Bounds and limitations

Bounds (n neurons)

- Diversity (number of learnt messages): ~~sublinear~~, quadratic
- Capacity: ~~subquadratic~~, quadratic
- $\Rightarrow$ Efficiency $\frac{\text{capacity}}{\text{binary information used}} \approx 1$.

~~Limitations~~ Interests

- As a plausible model:
  - Sensitive connections, fully interconnected network, retrieving requires a very large number of iterations...
- As an associative memory:
  - Very limited diversity, message length, network size and diversity are bound together, learning both messages and their opposite...

Bounds and limitations

Bounds (n neurons)

- Diversity (number of learnt messages): ~~sublinear~~, quadratic
- Capacity: ~~subquadratic~~, quadratic
- $\Rightarrow$ Efficiency $\frac{\text{capacity}}{\text{binary information used}} \approx 1$.

~~Limitations~~ Interests

- As a plausible model:
  - ~~Sensitive connections, fully interconnected network, retrieving requires a very large number of iterations...~~ Partition into clusters, use of neural cliques, only winner-take-all and sum, specialized neurons...
- As an associative memory:
  - Very limited diversity, message length, network size and diversity are bound together, learning both messages and their opposite...

Bounds and limitations

Bounds (n neurons)

- Diversity (number of learnt messages): ~~sublinear~~, quadratic
- Capacity: ~~subquadratic~~, quadratic
- $\Rightarrow$ Efficiency $\frac{\text{capacity}}{\text{binary information used}} \approx 1$.

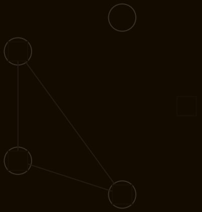
~~Limitations~~ Interests

- As a plausible model:
  - ~~Sensitive connections, fully interconnected network, retrieving requires a very large number of iterations. . .~~ Partition into clusters, use of neural cliques, only winner-take-all and sum, specialized neurons. . .
- As an associative memory:
  - ~~Very limited diversity, message length, network size and diversity are bound together, learning both messages and their opposite. . .~~ Near optimal performance, cumulative non destructive learning, strong resilience. . .

Why cliques?

Strong separability

- Cliques are fully interconnected subsets of nodes,
- Two distinct cliques with c nodes have at least $2(c-1)$ different connections:



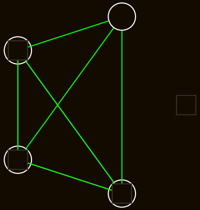
How controlling the size of cliques?

- Partition the network into clusters,
- Force cliques to contain one and only one neuron per cluster.

Why cliques?

Strong separability

- Cliques are fully interconnected subsets of nodes,
- Two distinct cliques with c nodes have at least $2(c-1)$ different connections:



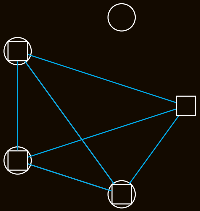
How controlling the size of cliques?

- Partition the network into clusters,
- Force cliques to contain one and only one neuron per cluster.

Why cliques?

Strong separability

- Cliques are fully interconnected subsets of nodes,
- Two distinct cliques with c nodes have at least $2(c-1)$ different connections:



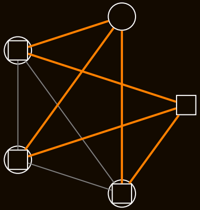
How controlling the size of cliques?

- Partition the network into clusters,
- Force cliques to contain one and only one neuron per cluster.

Why cliques?

Strong separability

- Cliques are fully interconnected subsets of nodes,
- Two distinct cliques with c nodes have at least $2(c-1)$ different connections:



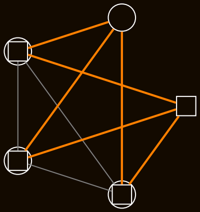
How controlling the size of cliques?

- Partition the network into clusters,
- Force cliques to contain one and only one neuron per cluster.

Why cliques?

Strong separability

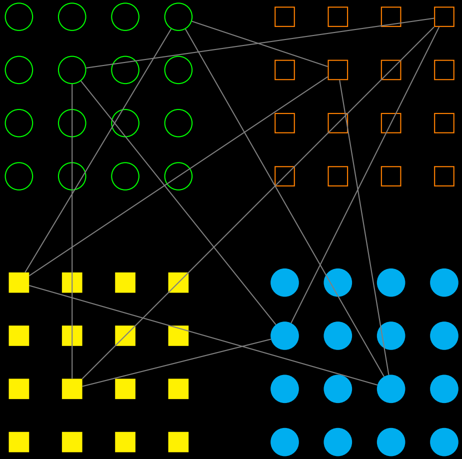
- Cliques are fully interconnected subsets of nodes,
- Two distinct cliques with c nodes have at least $2(c-1)$ different connections:



How controlling the size of cliques?

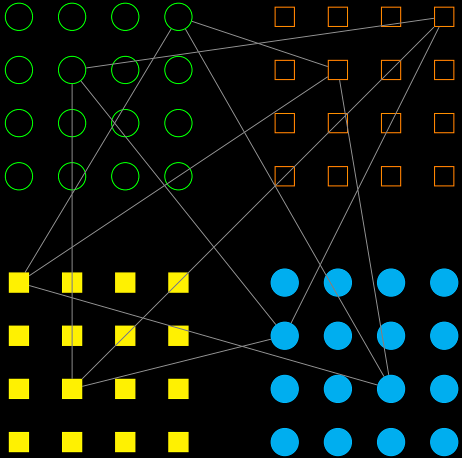
- Partition the network into clusters,
- Force cliques to contain one and only one neuron per cluster.

- Example: $c = 4$ clusters of $l = 16$ neurons each,
- $\underbrace{1000}_{j_1 \text{ in } c_1}\ \underbrace{0011}_{j_2 \text{ in } c_2}\ \underbrace{0010}_{j_3 \text{ in } c_3}\ \underbrace{1001}_{j_4 \text{ in } c_4}$,



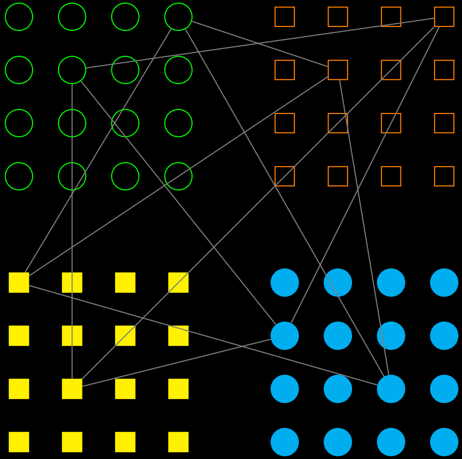
Our model: learning

- Example: $c = 4$ clusters of $l = 16$ neurons each,
- $\underbrace{1000}_{j_1 \text{ in } c_1}\ \underbrace{0011}_{j_2 \text{ in } c_2}\ \underbrace{0010}_{j_3 \text{ in } c_3}\ \underbrace{1001}_{j_4 \text{ in } c_4}$,



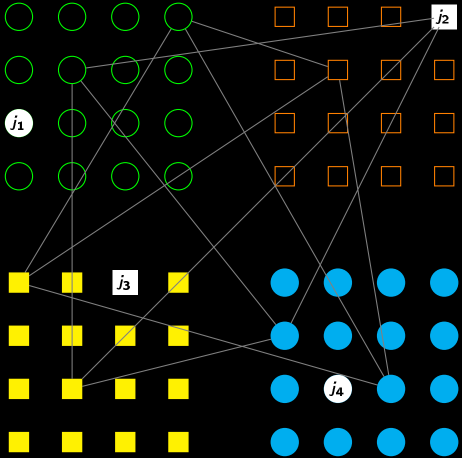
Our model: learning

- Example: $c = 4$ clusters of $l = 16$ neurons each,
- $\underbrace{1000}_{j_1 \text{ in } c_1}\ \underbrace{0011}_{j_2 \text{ in } c_2}\ \underbrace{0010}_{j_3 \text{ in } c_3}\ \underbrace{1001}_{j_4 \text{ in } c_4}$,



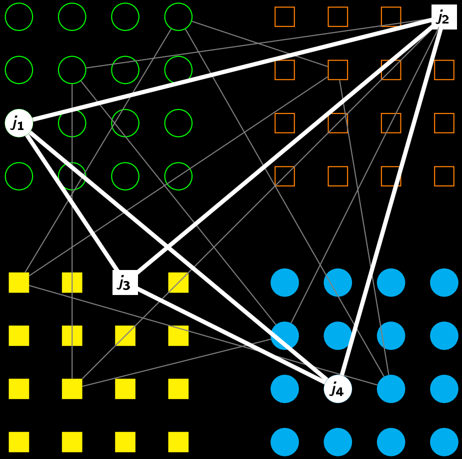
Our model: learning

- Example: $c = 4$ clusters of $l = 16$ neurons each,
- $\underbrace{1000}_{j_1 \text{ in } c_1}\ \underbrace{0011}_{j_2 \text{ in } c_2}\ \underbrace{0010}_{j_3 \text{ in } c_3}\ \underbrace{1001}_{j_4 \text{ in } c_4}$,

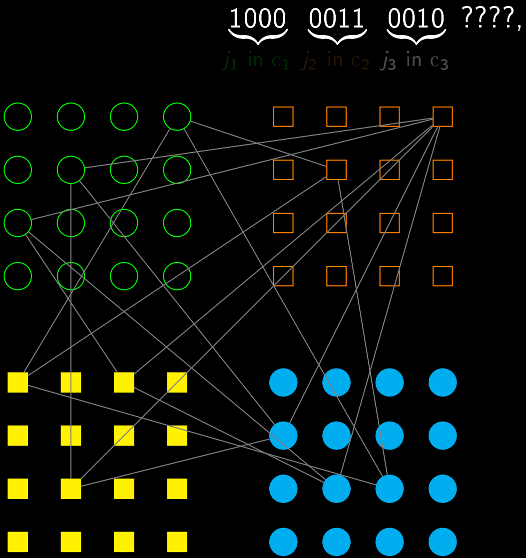


Our model: learning

- Example: $c = 4$ clusters of $l = 16$ neurons each,
- $\underbrace{1000}_{j_1 \text{ in } c_1}\ \underbrace{0011}_{j_2 \text{ in } c_2}\ \underbrace{0010}_{j_3 \text{ in } c_3}\ \underbrace{1001}_{j_4 \text{ in } c_4}$,

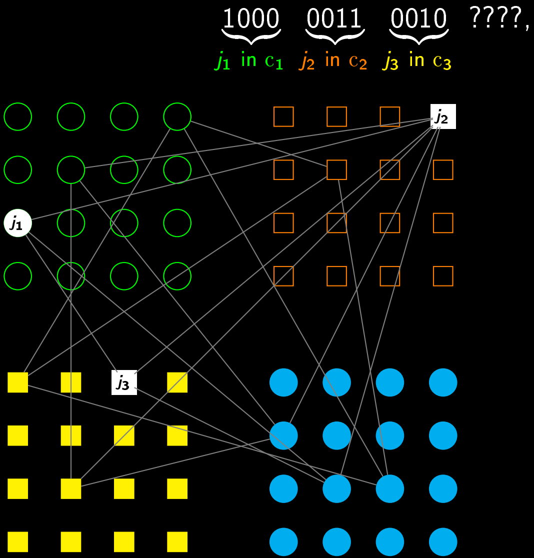


Our model: retrieving



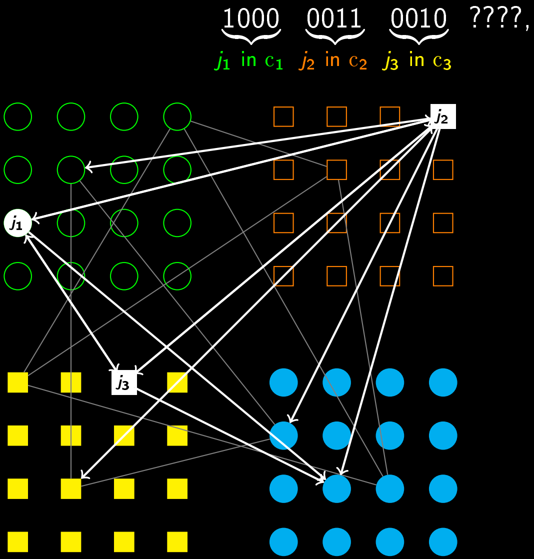
- Local correspondance,
- Global retrieving: sum of inputs,
- Local retrieving: winner-take-all,
- Possibly iterates the two last steps.

Our model: retrieving



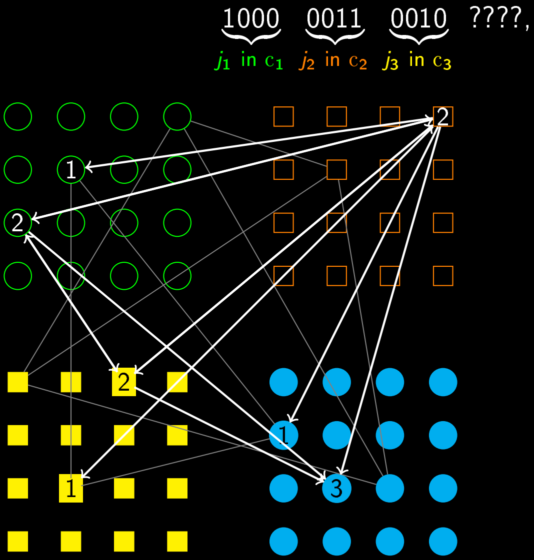
- Local correspondance,
- Global retrieving: sum of inputs,
- Local retrieving: winner-take-all,
- Possibly iterates the two last steps.

Our model: retrieving



- Local correspondance,
- Global retrieving: sum of inputs,
- Local retrieving: winner-take-all,
- Possibly iterates the two last steps.

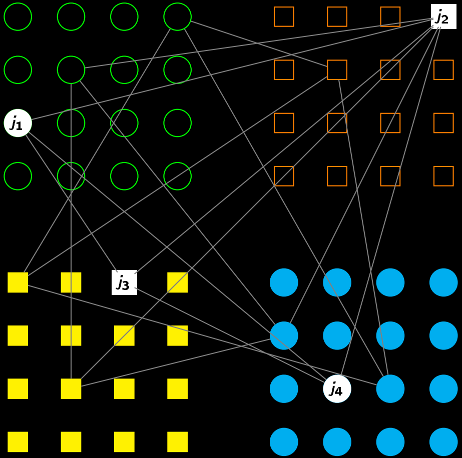
Our model: retrieving



- Local correspondance,
- Global retrieving: sum of inputs,
- Local retrieving: winner-take-all,
- Possibly iterates the two last steps.

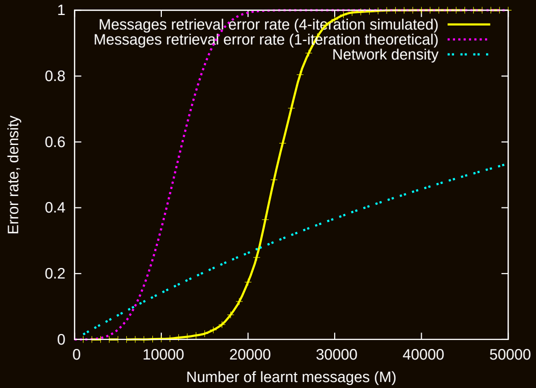
Our model: retrieving

$\underbrace{1000}_{j_1 \text{ in } c_1}\ \underbrace{0011}_{j_2 \text{ in } c_2}\ \underbrace{0010}_{j_3 \text{ in } c_3}\ \underbrace{1001}_{j_4 \text{ in } c_4},$



- Local correspondance,
- Global retrieving: sum of inputs,
- Local retrieving: winner-take-all,
- Possibly iterates the two last steps.

Associative memories

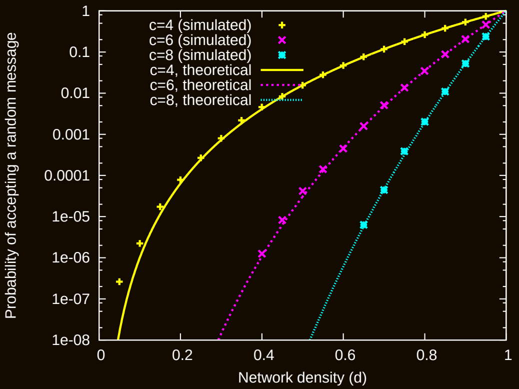


$c = 8$ clusters of $l = 256$ neurons,
Error probability while retrieving learnt messages with 4 out of 8 clusters with no provided information.

- Gain in comparison with the Hopfield network: 250 in diversity, 20 in capacity, and 20 in efficiency ($2.6\% \rightarrow 52\%$).

Performance

Classification

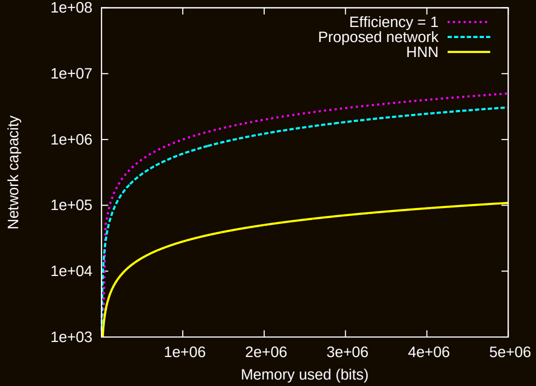


Various c values and $l = 512$ neurons per cluster, Error probability of second kind (probability to accept a non learnt message).

- Dramatic increase in performance compared to Hopfield network: 1071 in diversity, 52 in capacity, 52 in efficiency (2.6% $\rightarrow$ 137%),

Comparison in capacity with Hopfield neural networks

Capacity



Capacity of Hopfield neural networks and of our model (in the case of associative memory and a single iteration) in function of the amount of used information.

Conclusion

Results

- A novel neural architecture has been proposed which improves considerably diversity, capacity and efficiency,
- Biological plausibility has also been increased: binary connections, cliques over clusters, few iterations...

Further details

- See V. Gripon and C. Berrou, "Sparse neural networks with large learning diversity", to appear in IEEE Trans. on Neural Networks.
- Available online at http://arxiv.org/abs/1102.4240 .

Openings

This new kind of recurrent neural networks offer many openings in computational intelligence.

Thank you for your attention!