Information Theory Limits of Neuromorphic Energy Efficiency

Pau Vilimelis Aceituno
Institute of Neuroinformatics (UZH/ETH)

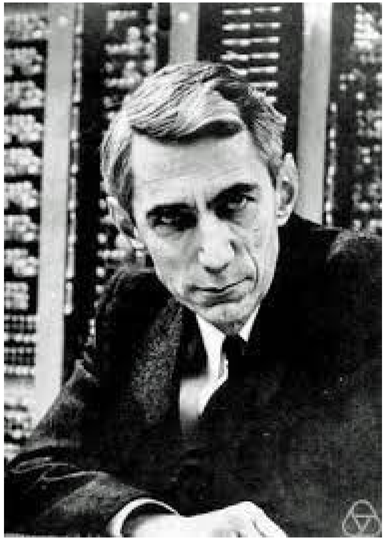
Motivation

- **Neuromorphic systems are energy efficient**
  - How efficient can they be?
  - How do we design them?
- **Neuromorphic systems are (very) heterogeneous**
  - Do we need a new energy analysis for every system?
  - Or for every problem?

- **Warning: this is only theoretical**

Approach: Information theory

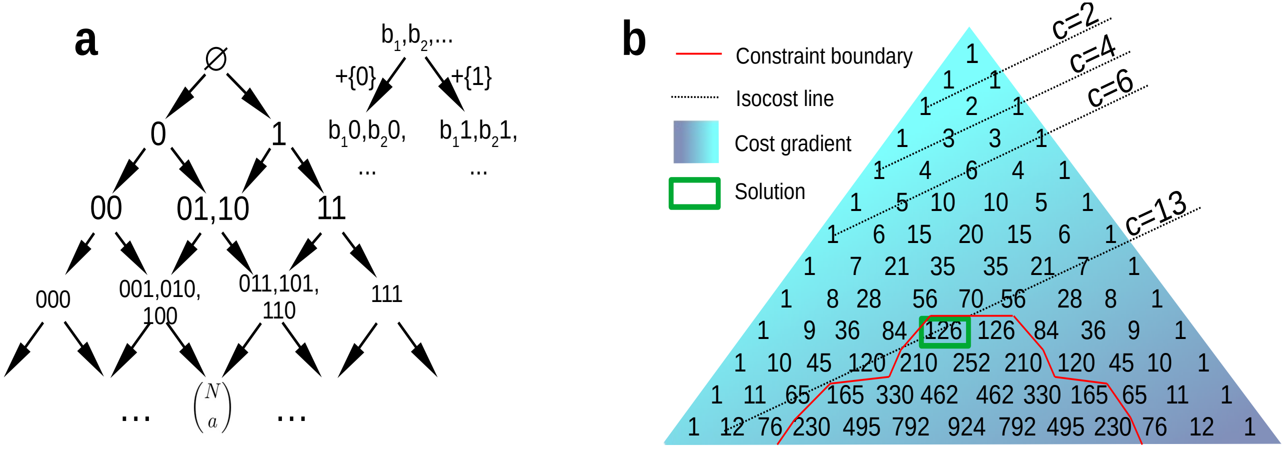
- **History**
  - Right language for information is to use symbols and probabilities
  - Proved that reliable communication on unreliable hardware works



- **Main approach**
  - Minimize cost (code length) → With equal costs for 1/0
  - While respecting code decodability

Source Coding: Combinatorial view

- **Simple example:**
  - Encode 125 symbols, constant activations
  - Each neuron costs 1, each activation costs +1



Similar to Levy & Baxter, Neural Computation 1996

Source Coding: Analytical view

S symbols with prob $\mathbf{p_s}$, **A** maximum number of active neurons, **N** neurons, the cost of each activation is **c(a,N)**

Problem: $$N^*, A^* = \arg\min_{N,A} \sum_{s=1}^{S} p_s c(a(s), N) \quad s.t. \quad S \leq \sum_{a=1}^{A} \binom{N}{a}$$

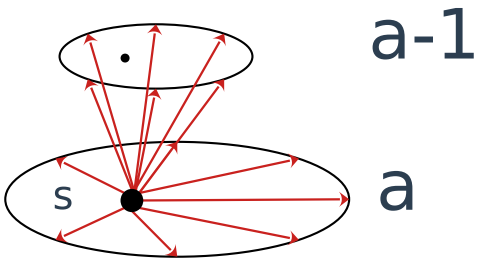
Solution: $$N^* = \arg\min_{N} \sum_{s=1}^{S} p_s c\left(NH^{-1}\left(\frac{\ln(s)}{N}\right), N\right), \quad A^* \approx NH^{-1}\left(\frac{\ln(S)}{N}\right)$$

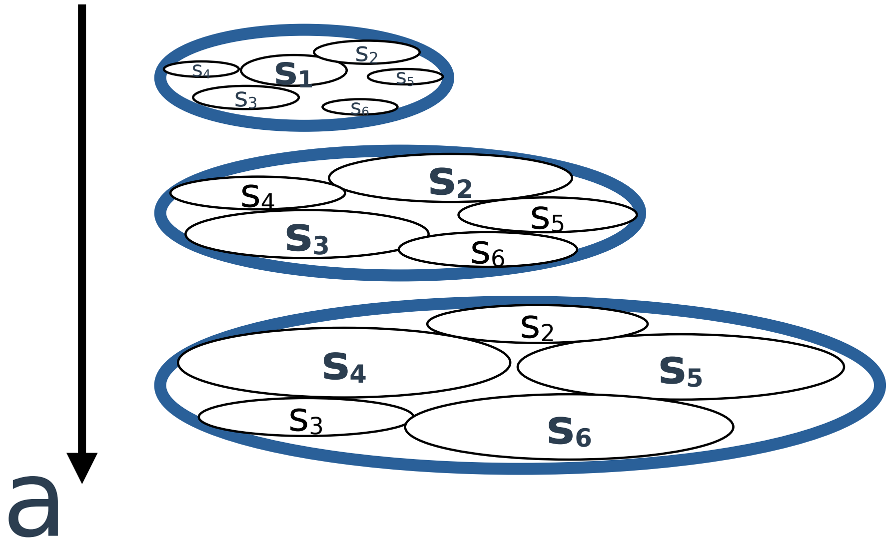
Analytical tricks: Stirling's approx. and Laplace's method

$$\binom{N}{a} \sim exp[NH_2(\frac{a}{N}))] \qquad \lim_{N\to\infty} \int exp[Nf(x)]dx \sim exp[Nf(x)]$$

Noisy Channel Coding: Hand-wavy view

Each Codeword is perturbed by noise and the limit of that noise defines a “ball” of possible codewords





- Each activity level has “volume” *H(a/N)*
- Each Codeword occupies volume in multiple levels
- **How to distribute codeword volume without overlap?**

Limitations and Non-limitations

✅ **Moving from 1-0 to numbers is ok:**

- n spikes is a different symbol → entropy approximation works

❌ **The approximations are on orders of magnitude**

- If you try for 100 neurons it won't work

❌ **Information theory has trouble with structure**

- If encoding values is unintuitive

✅ **Neuroscience has very heterogeneous neurons**

- If there are no heavy-tailed distributions it works

Acknowledgements

- **People who helped**
  - Jean-Pascal Pfister (U. Bern)
  - Hui-An Shen (U. Bern)
  - Stephan Moser (ETH)

- **Non-people who payed/hosted**
  - Institute of Neuroinformatics (UZH/ETH)
  - Grewelab
  - ETH Postdoctoral fellowship