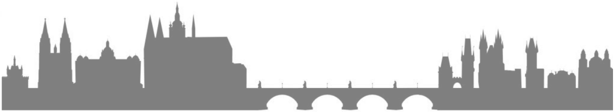
Neuronal information transmission: from finite-size effects to source-channel coding

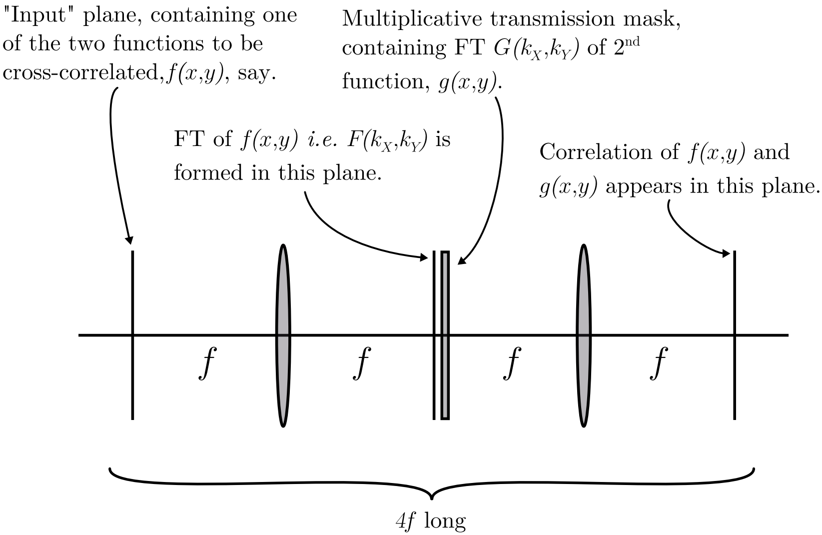
(June 9, 2025, University of Torino, Italy)

Lubomir Kostal

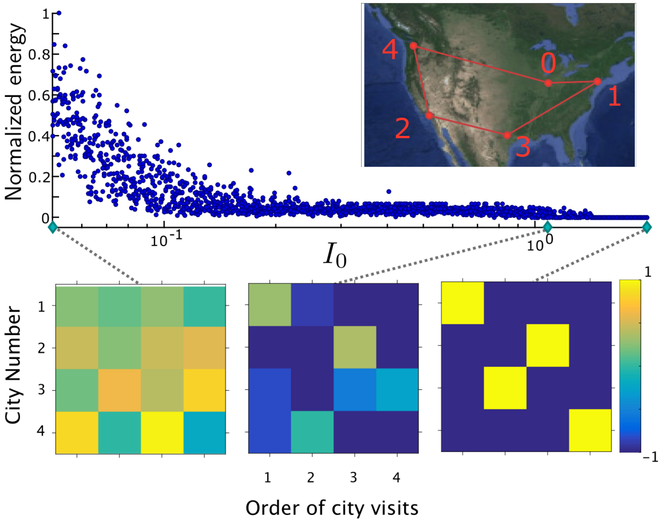
Institute of Physiology CAS, Prague, Czech Republic



Analog computation: 4*f*-correlator (*effortless!*)

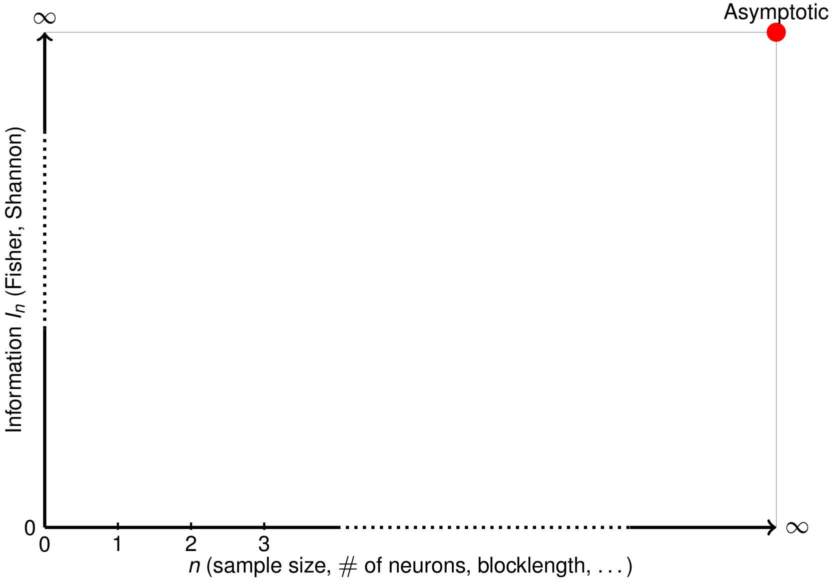


Probabilistic computation (*p-bits*)

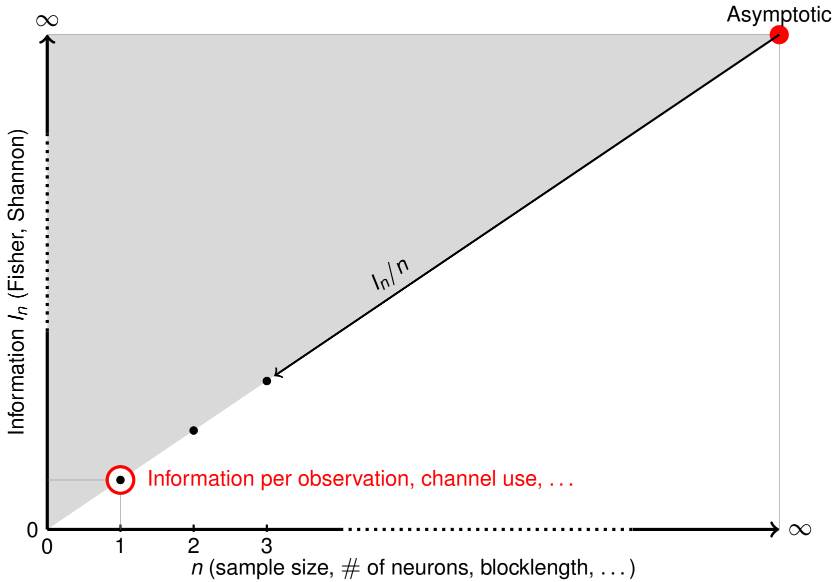


Feynman (1982) ⟶ Camsari et al. (2019)

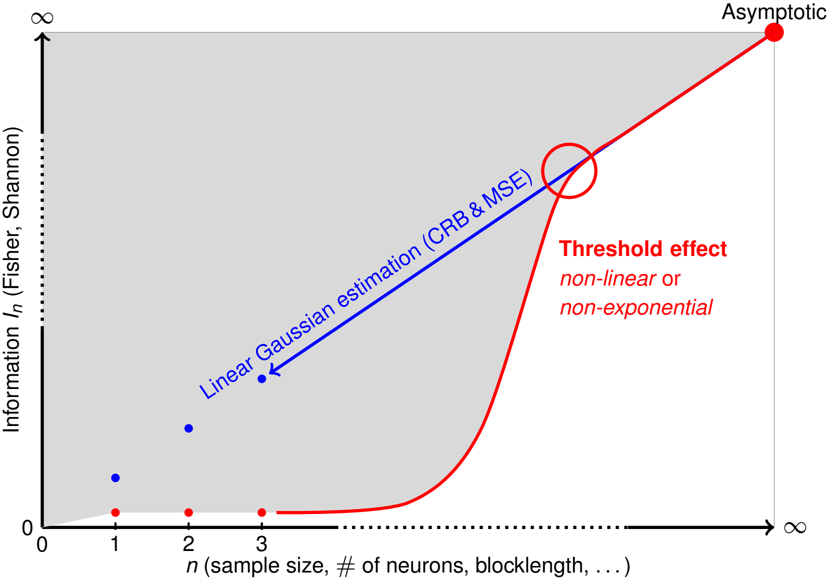
Asymptotic *vs.* non-asymptotic information (*heuristic*)



Asymptotic *vs.* non-asymptotic information (*heuristic*)

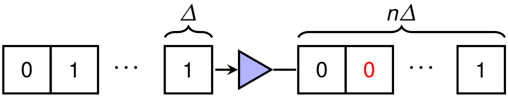


Asymptotic *vs.* non-asymptotic information (*heuristic*)



Shannon's theorem

- 'Reliability' $\Rightarrow$ sequence *vs.* per-symbol decoding, 'errors' *(BSC)*



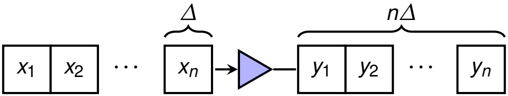
- Information rate (nat/s) assuming m (known) input n-sequences

$$\mathbf{x} = \{x_1, x_2, \ldots, x_n\} \Rightarrow R = \frac{\log m}{n\Delta}$$

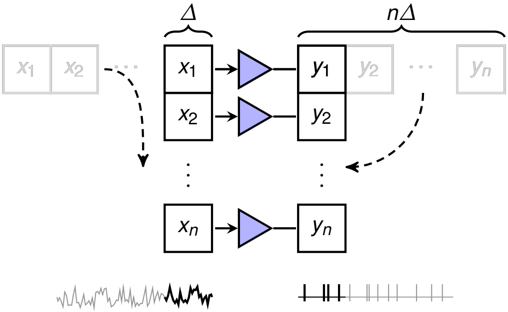
- Shannon's theorem (channel coding) $\doteq$ if $R < C$ then $\exists$ a *set* of $\mathbf{x}$'s such that Pr of $\hat{\mathbf{x}} \neq \mathbf{x}$ is *arbitrarily small* ($\Rightarrow$ n increasing!)

- Signal *estimation* vs. *detection*: up to $m \approx e^{\Delta nC}$ 'patterns' decoded reliably for n *large* enough *(NN classifiers)*

Simple neuronal population model



Simple neuronal population model



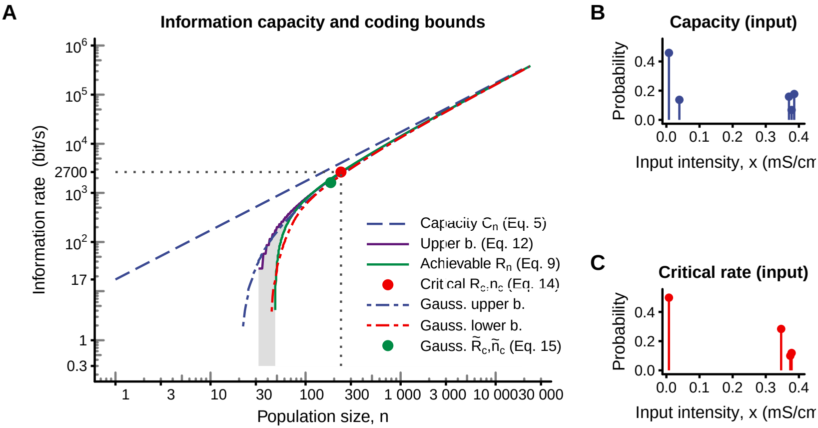
$x = \langle g_e \rangle_\Delta \quad y = \#\mathrm{AP}/\Delta$

- Set of m stimulus 'patterns': $\mathbf{x}^{(j)} = \{x_1^{(j)}, \ldots, x_n^{(j)}\}, j = 1, \ldots, m$
- For $\mathbf{X} = \mathbf{x}$ we observe $\mathbf{Y} = \mathbf{y}$, given by $f(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{n} f(y_i|x_i)$
- ML decoder *(optimality?, cf. ME decoding)*:

$$\mathbf{x}^{(d)} : \quad d = \arg\max_j f(\mathbf{y}|\mathbf{x}^{(j)}),$$

Decoding is difficult (Kostal & Kobayashi, PRE(R) 2019)

HH model + balanced input (Wehr and Zador, 2003; Berg et al., 2007)



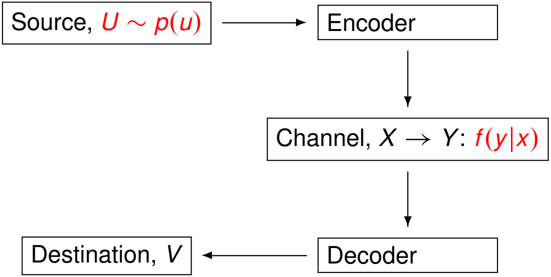
'Critical' rate R_c ($\approx$ bounds eq.) $P_e = 10^{-10}$ (rel. 'noise'), note $I(X_c; Y) \neq R_c$, Reliable information transmission is *not automatic*!

Decoding is a complex task ($\Rightarrow$ delays).

Can coding-decoding be avoided?

Point-to-point communication revisited

- Fundamental limits on the efficacy of: *i) representation* and *ii) reliable communication* of information
- Probabilistic description: *source* and *channel*



- Representation: compression (*source entropy*)
- Reliability: probability of channel decoding error

Mutual information $I(X;Y)$: channel communication

- Given the channel $f(y|x)$ and some input distribution $X \sim p(x)$

$$I(X;Y) = \iint p(x) f(y|x) \log_2 \frac{f(y|x)}{\int f(y|z) p(z)\,\mathrm{d}z}\,\mathrm{d}y\,\mathrm{d}x$$

- Cost of each input $w(x)$ and the average cost

$$P = \int w(x) p(x)\,\mathrm{d}x$$

- Maximal reliable communication at average cost not exceeding W *(capacity-cost function)*

$$C(W) = \max_{p(x):P \leq W} I(X;Y)$$

Mutual information $I(U;V)$: source compression

- Given the source $U \sim p(u)$ and source-destination prob. $g(v|u)$

$$I(U;V) = \iint p(u) g(v|u) \log_2 \frac{g(v|u)}{\int g(v|z) p(z)\, \mathrm{d}z} \,\mathrm{d}u\, \mathrm{d}v$$

- Distortion (fidelity) $d(u,v)$ and the average distortion

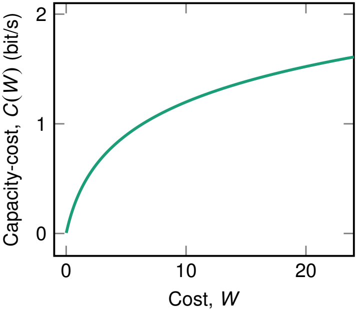
$$\Delta = \iint d(u,v) p(u) g(v|u)\, \mathrm{d}u\, \mathrm{d}v$$

- Maximal compression of source (bit/symbol) at average distortion not exceeding D *(rate-distortion function)*

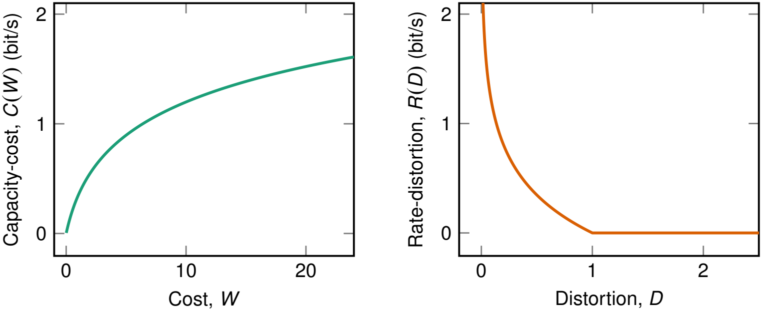
$$R(D) = \min_{g(v|u):\Delta \leq D} I(U;V)$$

N.B.: For simple discrete sources U and a zero-average Hamming distortion (perfect reconstruction) we have $R(D) = H(U)$

Typical shapes of $C(W)$ and $R(D)$

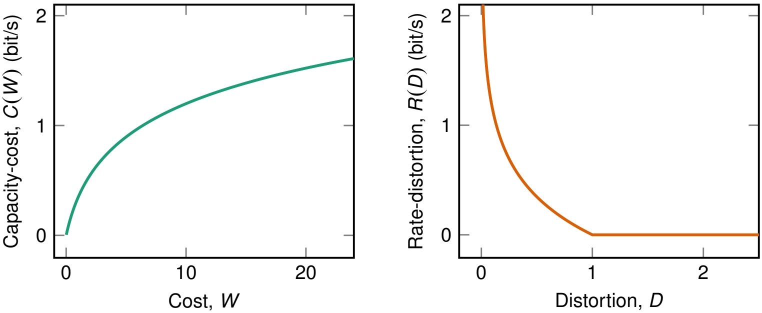


Typical shapes of $C(W)$ and $R(D)$



- Reliable info transfer possible: $C(W) \geq R(D)$

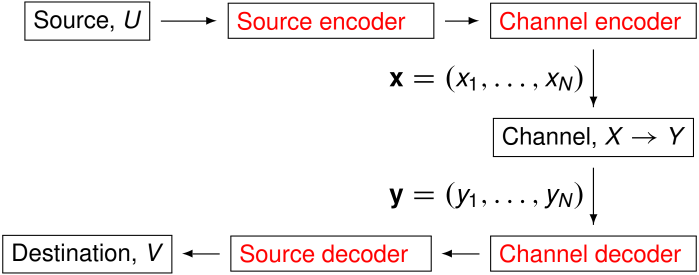
Typical shapes of $C(W)$ and $R(D)$



- Reliable info transfer possible: $C(W) \geq R(D)$

Optimality: $C(W) = R(D)$

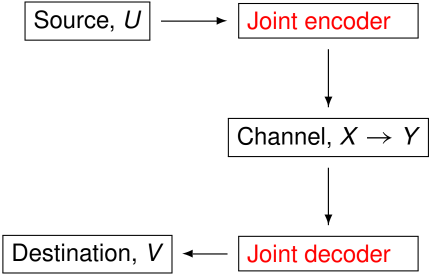
Traditional communication setup: *separation*



- The bounds asymptoticaly achievable by separation cannot be improved by any other approach (*point-to-point*)
- Penalty: delay, computational complexity (*block-coding*)

Joint source-channel coding

- Even though the *fundamental* limits cannot be improved, there are benefits (coding complexity, robustness, . . .)



- No "global" theory, *ad hoc* approaches
- Extreme case: probabilistic source-channel matching

Example: source-channel matching

Source $U \to V$ (1 symb/s): $U \sim N(0, \sigma_U^2)$, $\mathbb{E}(U - V)^2 \leq D$

Example: source-channel matching

Source $U \to V$ (1 symb/s): $U \sim N(0, \sigma_U^2), \quad \mathbb{E}(U - V)^2 \le D$
Channel $X \to Y$ (1 symb/s): $Y = X + Z,\ Z \sim N(0, \sigma_Z^2), \quad \mathbb{E}\, X^2 \le W$

Example: source-channel matching

Source $U \to V$ (1 symb/s): $U \sim N(0, \sigma_U^2), \quad \mathbb{E}(U - V)^2 \leq D$
Channel $X \to Y$ (1 symb/s): $Y = X + Z,\ Z \sim N(0, \sigma_Z^2), \quad \mathbb{E}\, X^2 \leq W$

Separation *(traditional)*

a) lossy *compression* of U with distortion D: min. $R(D)$ bit/s
b) reliable *transfer* of information: max. $C(W)$ bit/s

$$R(D) = \begin{cases} \frac{1}{2}\log_2 \frac{\sigma_U^2}{D}, & D < \sigma_U^2 \\ 0, & D \geq \sigma_U^2 \end{cases}, \quad C(W) = \frac{1}{2}\log_2\left(1 + \frac{W}{\sigma_Z^2}\right).$$

Example: source-channel matching

Source $U \to V$ (1 symb/s): $U \sim N(0, \sigma_U^2), \quad \mathbb{E}(U - V)^2 \leq D$
Channel $X \to Y$ (1 symb/s): $Y = X + Z,\ Z \sim N(0, \sigma_Z^2), \quad \mathbb{E}\, X^2 \leq W$

Separation *(traditional)*

a) lossy *compression* of U with distortion D: min. $R(D)$ bit/s
b) reliable *transfer* of information: max. $C(W)$ bit/s

Optimum: $R(D) = C(W)$ and thus

$$D = \frac{\sigma_U^2 \sigma_Z^2}{W + \sigma_U^2}$$

Block coding, complexity of decoding, ...
Achievability of $R(D)$ and $C(W)$? Only *asymptoticaly* ...
The mapping $U \to V$ is “deterministic”.

Example: source-channel matching

Source $U \to V$ (1 symb/s): $U \sim N(0, \sigma_U^2), \quad \mathbb{E}(U - V)^2 \leq D$
Channel $X \to Y$ (1 symb/s): $Y = X + Z,\ Z \sim N(0, \sigma_Z^2), \quad \mathbb{E}\, X^2 \leq W$

Joint source-channel "coding"

By scaling the inputs and outputs (*symbol-per-symbol*):

$$X = \sqrt{\frac{W}{\sigma_U^2}} U, \quad V = \sqrt{\frac{\sigma_U^2}{W}} \frac{W}{W + \sigma_Z^2} Y, \quad D = \mathbb{E}(U - V)^2 = \frac{\sigma_U^2 \sigma_Z^2}{W + \sigma_U^2}$$

The optimality (separation) is achieved *without coding*!
The mapping $U \to V$ is *stochastic*.

Source-channel match conditions

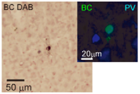
- Shannon(?), Csiszár & Körner (1981), Gastpar *et al.* (2003)
- *Source:* $p(u)$ and distortion $d(u, v)$, $V \sim g(v|u)$
- *Channel:* $f(y|x)$ and cost $w(x)$
- The source-channel matching code is optimal iff:

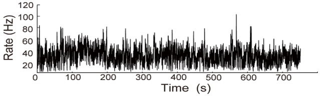
$$
\begin{aligned}
I(U;V) &= I(X;Y), \\
w(x) &= a_1 \int f(y|x) \log \frac{f(y|x)}{p(y)} \, \mathrm{d}y + a_2, \\
d(u,v) &= -b_1 \log g(u|v) + b_2(u)
\end{aligned}
$$

- *What is the source, channel, cost and distortion interpretation?*

Model + comparison with experimental data

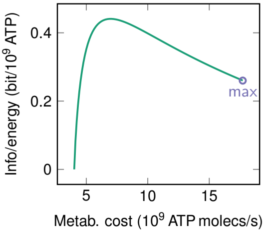
Experimental data:
in vivo (recordings > 30 min.),
layer 2-6 (sensorimotor cortex),
pyramidal and inter-neurons



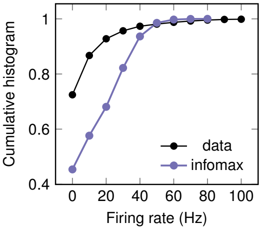


Data: Dr. Tomoki Fukai

Information/energy

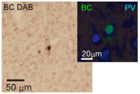


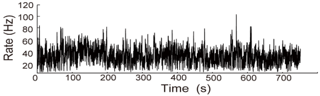
Pyramidal neuron (example)



Model + comparison with experimental data

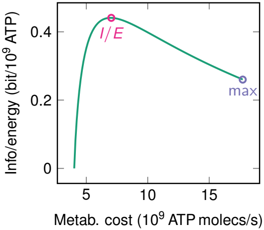
Experimental data:
in vivo (recordings $>$ 30 min.),
layer 2-6 (sensorimotor cortex),
pyramidal and inter-neurons



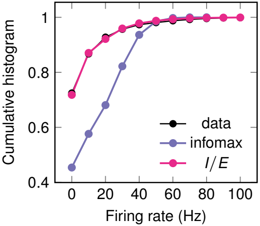


Data: Dr. Tomoki Fukai

Information/energy

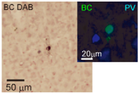


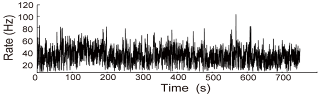
Pyramidal neuron (example)



Model + comparison with experimental data

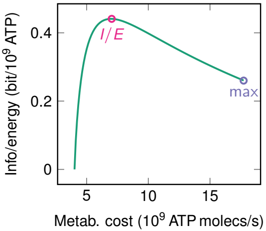
Experimental data:
in vivo (recordings > 30 min.),
layer 2-6 (sensorimotor cortex),
pyramidal and inter-neurons



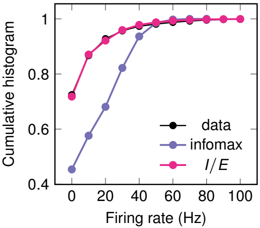


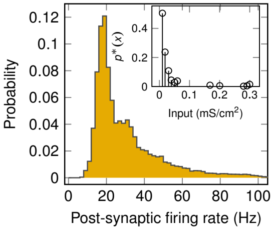
Data: Dr. Tomoki Fukai

Information/energy



Pyramidal neuron (example)

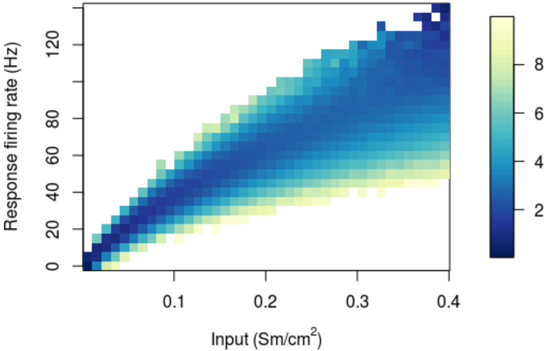




– Predicted PSFR histograms match the data
– Sparse synaptic activity (Brecht *et al.*, *Nature*, 2004)
– Log-normal distribution of synaptic weights (Song *et al.*, *PLoS Biol.*, 2005)

Predicted distortion $d(u, v)$

- *Source-channel matching*: "assume" $U = X$ and $V = Y \Rightarrow$ small loss of efficiency and PSFR difference
- Predicted cost function ($\propto \int f(y|x) \log[f(y|x)/p(y)]\, \mathrm{d}y$) is approx. the metabolic cost!



Interpretation of $d(u, v)$?

Separated vs. *Joint* source-channel coding (SCC)

"Suitable" (✔) vs. "debatable" (✘) in *bio-inspired* systems

Separated SCC

- Optimal and universal in the *point-to-point* scenario ✔
- Source independence (modular character) ✘
- Sensitive to channel perturbations ✘
- Complicated achievability (asymptotic limits) ✘
- Not optimal in networks ✘
- Not optimal in constrained systems ✘

Separated vs. *Joint* source-channel coding (SCC)

"Suitable" (✔) vs. "debatable" (✘) in *bio-inspired* systems

Separated SCC

- Optimal and universal in the *point-to-point* scenario ✔
- Source independence (modular character) ✘
- Sensitive to channel perturbations ✘
- Complicated achievability (asymptotic limits) ✘
- Not optimal in networks ✘
- Not optimal in constrained systems ✘

Joint SCC

- Source-dependent ✔
- Optimal performance in constrained systems ✔
- Not universal (distortion measures, ...) ✘
- Robust against perturbations ✔
- Communication with feedback ✔
- Communication in networks ✔

Summary & Conclusions

- Asymptotic *vs.* achievable: finite-size effects might be important!
- Achievable (*operational*) info is *not* additive even under *iid* setup
- Neural systems: natural ('effortless') computation
- Joint source-channel coding: optimality available for free?
- Sometimes it is better *not to* fight the **randomness.**

Decoding: 'maximum likelihood'

- Set of m stimulus 'patterns': $\mathbf{x}^{(j)} = \{x_1^{(j)}, \ldots, x_n^{(j)}\}, j = 1, \ldots, m$
- For $\mathbf{X} = \mathbf{x}$ we observe $\mathbf{Y} = \mathbf{y}$, given by $f(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{n} f(y_i|x_i)$
- ML decoder *(optimality?, cf. ME decoding)*:

$$\mathbf{x}^{(d)} : \quad d = \arg\max_j f(\mathbf{y}|\mathbf{x}^{(j)}),$$

- Average probability of decoding error

$$P_e = \sum_{j=1}^{m} \Pr(\mathbf{x} = \mathbf{x}^{(j)}) \int_{\mathcal{E}(j)} f(\mathbf{y}|\mathbf{x}^{(j)})\, \mathrm{d}\mathbf{y}$$

$\mathcal{E}(j)$: set of $\mathbf{y}$ such that ML *fails* for $\mathbf{x}^{(j)}$

- How to obtain the inputs $\mathbf{x}$? *(Assume $\Pr(\mathbf{x} = \mathbf{x}^{(j)}) = 1/m$.)*

Lower bound on info rate (*achievable* & general)

- *Ensemble*: generate patterns *i.i.d.* according to some $X \sim \pi(x)$
- Prob. of particular set of m inputs, each of length n:

$$\prod_{j=1}^{m} \pi(\mathbf{x}^{(j)}) = \prod_{j=1}^{m} \pi(x_1^{(j)}) \cdots \pi(x_n^{(j)})$$

- Use the random-coding bound (Gallager, 1968) & invert
- Optimize over $\pi(x)$ to get a tighter result *(convex)*:

$$R_n \geq \frac{n}{\Delta} E_r^{-1}\left(-\frac{\log P_e}{n} \right),$$

$$E_r(\Delta R) = \max_{0 \leq \rho \leq 1} \left[\max_{\pi(x)} E_0\big(\rho, \pi(x)\big) - \rho \Delta R \right],$$

$$E_0\big(\rho, \pi(x)\big) = -\log \int \left(\int f(y|x)^{1/(1+\rho)} \pi(x)\, \mathrm{d}x \right)^{1+\rho} \mathrm{d}y$$

- Note: optimal $\pi^*(X)$: $R \neq I(\pi^*(X), Y)$; 'good' codes? *(not iid)*

Upper bound on info rate

- Technical assumptions, validity ... (*BSC*, Polyanski, 2014)
- *Cf.*: sphere-packing and straight-line bounds (Gallager, 1968)
- Strassen, 1962; Tomamichel, 2013, assume $P_e \leq 1/2$

$$R_n \leq nC - \frac{1}{\Delta}\left[\sqrt{nV(P_e)}Q^{-1}(P_e) + \frac{\log n}{2}\right] + O(1)$$

$$V(P_e) = \min_{\pi \in \mathcal{C}}\left[\mathbb{E}\left(\log \frac{f(Y|X)}{p(Y)}\right)^2 - \Delta^2 C^2\right]$$

$Q(z) = \int_z^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\,\mathrm{d}t$, *n.b.* (Cramér-Esseen, 1937, 1945,)
CLT *vs.* AEP (Feinstein, 1958; Wolfowitz, 1961, $nC + O(\sqrt{n})$)

- Note the const. term (*cf.* Feller, 1972; Tyurin, 2010)

Gaussian approximation

- Gaussian approx. (AWGN: $\mathrm{Var}(X) \leq P$)

$$C_G = \frac{1}{\Delta} \ln\left(1 + \frac{P}{\sigma^2}\right)$$

- Effective SNR: $S = P/\sigma^2$
- Let $S = (e^{2\Delta C} - 1)$ (where C is the single neuron capacity), then *since* $C \propto \log(1 + SNR)$

$$\tilde{R}_c = \frac{1}{2\Delta} \log\left(\frac{1}{2} + \frac{S}{4} + \frac{1}{2}\sqrt{1 + \frac{S^2}{4}}\right),$$

$$\tilde{n}_c = \Big\lceil -4\Big[2 + S - \sqrt{4 + S^2} - 4\log 2 + 2 + \log\left(2 - S + \sqrt{4 + S^2}\right)\Big]^{-1} \log P_e\Big\rceil$$

+ complete closed-form for the error exponents

Asymptotics *vs.* achievable information rates

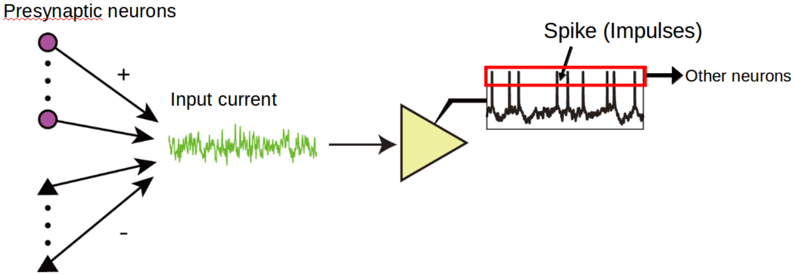
- In fact: the (average) probability of decoding error P_e: phase transition at $R = C$ in the 'thermodynamic' limit $n \to \infty$
(Gallager: $P_e = 0$ for $R < C$, Wolfowitz: $P_e = 1$ for $R > C$)
- Finite-size effects (n): relationship $R \leftrightarrow P_e$?
Perhaps $R > C$ for some 'reasonable' P_e?
- *Maximal* asymptotic *vs.* achievable rates?

$$R_n \approx nC \quad \propto \log m \qquad (n \to \infty)$$
$$R_n = ? \qquad \text{(generaly a function of } P_e, n)$$

Shannon (1959), Gallager (1962–73), . . . , Verdu, Polyanski (2010)

- Non-asymptotics: new relevant parameters
- Price to pay: delays, "complexity": $O(e^N), \ldots$
(Punekar *et al.*, 2013: non-binary LDPC $\sim O(N)$)

Investigated neuronal model (*cortical excitatory*)



- Hodgkin-Huxley + *point-conductance* (*stochasticity*)
  - Excitatory and inhibitory conductances follow Ornstein-Uhlenbeck processes and $\langle g_{E,I} \rangle \propto \lambda_{E,I}$
  - Effective reversal potential: V_r (Miura *et al.*, 2007)

$$\langle g_e \rangle (E_e - V_r) + \langle g_i \rangle (E_i - V_r) = 0$$

  - Balanced input, $\lambda_E \propto \lambda_I$
- Input, $x \equiv \langle g_E \rangle$: mean excitatory conductance (input parameter)
- Output, $y \equiv \#\mathrm{APs}/\Delta$: firing rate

Investigated neuronal model (*cortical excitatory*)

$$\begin{aligned} C_m \frac{\mathrm{d}V}{\mathrm{d}t} &= -I_{\mathrm{Na}}(t) - I_{\mathrm{K}}(t) - g_L(V - E_L) + I_{\mathrm{syn}} \\ I_k &= g_k(V)(V - E_k) \\ I_{\mathrm{syn}} &= -g_e(t)(V - E_e) - g_i(t)(V - E_i) \end{aligned}$$

- Hodgkin-Huxley + *point-conductance* (*stochasticity*)
  - Excitatory and inhibitory conductances follow Ornstein-Uhlenbeck processes and $\langle g_{E,I} \rangle \propto \lambda_{E,I}$
  - Effective reversal potential: V_r (Miura *et al.*, 2007)

$$\langle g_e \rangle (E_e - V_r) + \langle g_i \rangle (E_i - V_r) = 0$$

  - Balanced input, $\lambda_E \propto \lambda_I$
- Input, $x \equiv \langle g_E \rangle$: mean excitatory conductance (input parameter)
- Output, $y \equiv \#\mathrm{APs}/\Delta$: firing rate

Metabolic cost of neuronal activity & efficiency

- *Empirical* metabolic cost given $X = x$ (Attwell & Laughlin, 2001)

$$w(x) = \kappa \times \big(\langle \#\text{APs in } \Delta \rangle \,|\, x\big) + \beta\Delta$$

$[\kappa = 7.1 \times 10^8$ ATPm, $\beta = 4.4 \times 10^8$ ATPm/s$]$

- *Theoretical* (model) cost: only small corrections (RK)✔
- Maximize the information-to-cost ratio (efficiency) E

$$C(W) = \max_{p(x):P \leq W} I(X;Y), \qquad P = \int w(x)p(x)\,\mathrm{d}x,$$

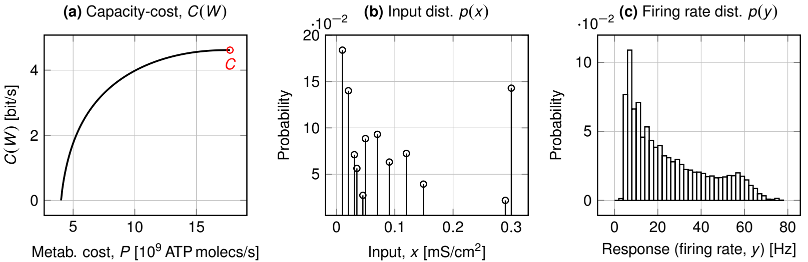
$$E = \max_{W} \frac{C(W)}{W},$$

i.e., $1/E$ is the *minimal cost* of 1 bit

Useful preliminaries

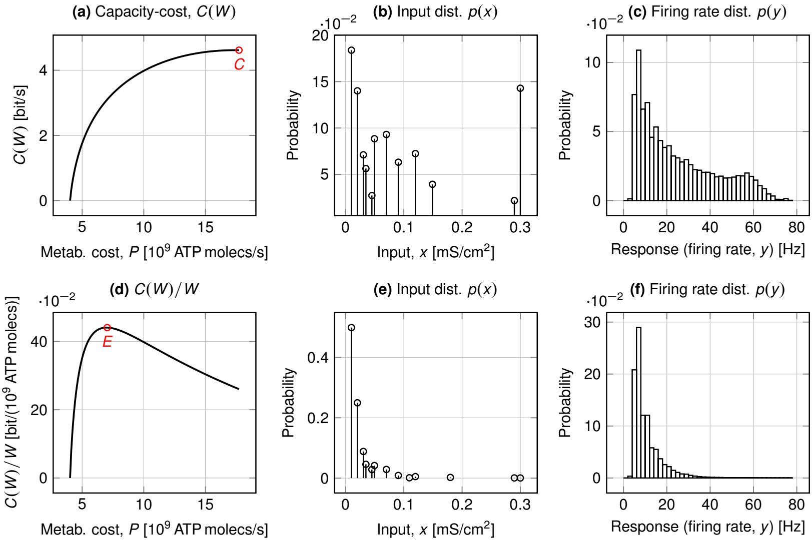
- Y: spike-count in a time window $\Rightarrow \exists$ max. $\Rightarrow$ discrete & finite
- Let $Y \sim f(y|X = x)$: max. K points of support
- Witsenhausen, 1980 ($\Leftarrow$ Dubin's theorem): capacity is achieved by discrete $\pi(X)$ supported at most K points (finite dimensionality, almost no assumptions on X!)
- Extendable to other convex optimization problems:
- 'Model' *vs.* DMC: applicable bounds on R_n
- Numerical methods: cutting-plane (linear programming), ...

Results: predict $p(y) = \int f(y|x)p(x)\,\mathrm{d}x$



($V_r = -60\,\mathrm{mV}$, $g_M = 8\,\mathrm{mS/cm^2}$)

Results: predict $p(y) = \int f(y|x)p(x)\,\mathrm{d}x$



$(V_r = -60\,\mathrm{mV},\ g_M = 8\,\mathrm{mS/cm^2})$