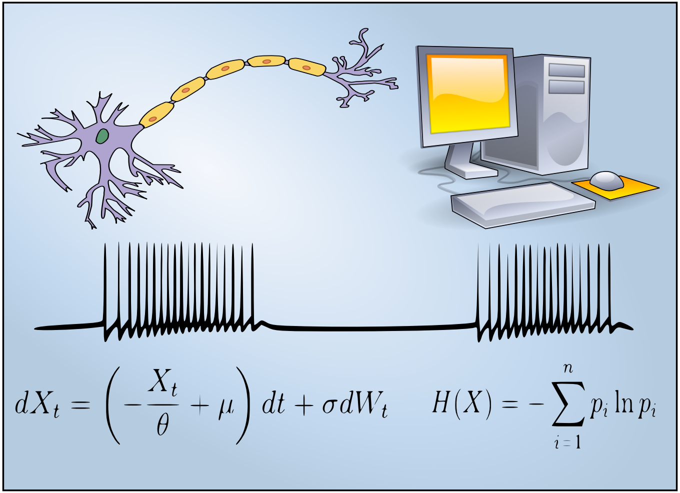
On the efficiency of neuronal information processing

Lubomir Kostal

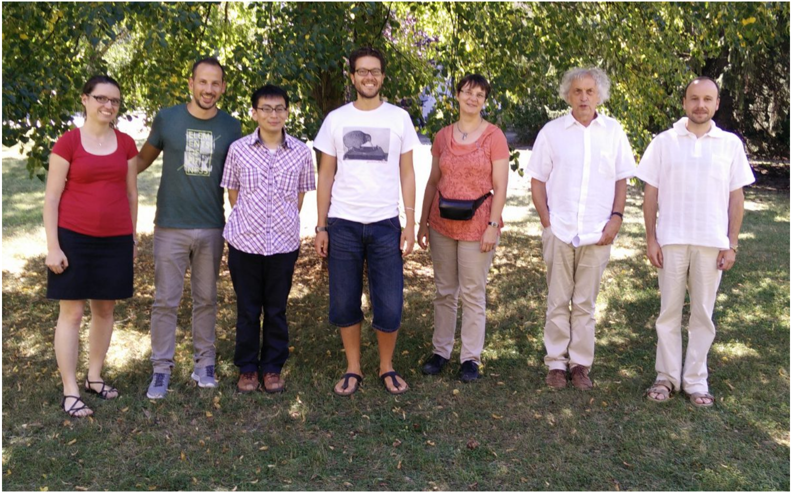
Institute of Physiology CAS, Prague, Czech Republic

Computational Neuroscience Group, IPHYS, Prague



http://comput.biomed.cas.cz

Computational Neuroscience Group, IPHYS, Prague



http://comput.biomed.cas.cz

Outline

1. What is *Computational neuroscience*?
2. Neuronal coding
3. Information-theoretic approach: metabolic efficiency
4. Possible predictions . . .

- Joint work with Ryota Kobayashi

Computational Neuroscience

"The aim of computational neuroscience is to explain how electrical and chemical signals are used in the brain to represent and process information."

T. Sejnowski *et al.*: Computational Neuroscience, *Science*, 1988

Cybernetics: Or Control and Communication in the Animal and the Machine

Norbert Wiener, 1948

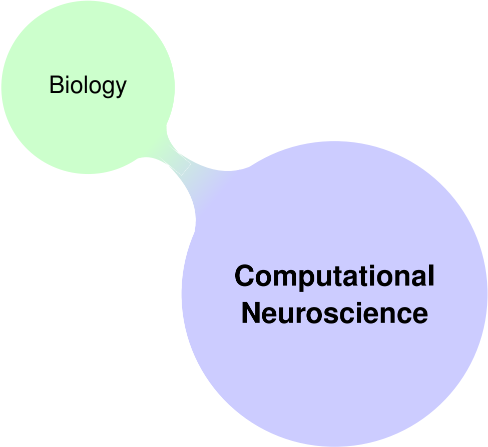
Why? – Progress in *neuroscience* (from molecules to fMRI)
– Progress in *computing power*

But. . . How the nervous system enables us to *see, remember, plan?*

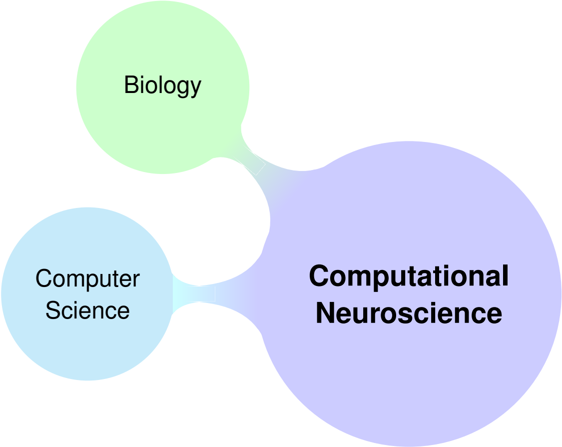
"What are the algorithms used in the brain?"

Computational Neuroscience

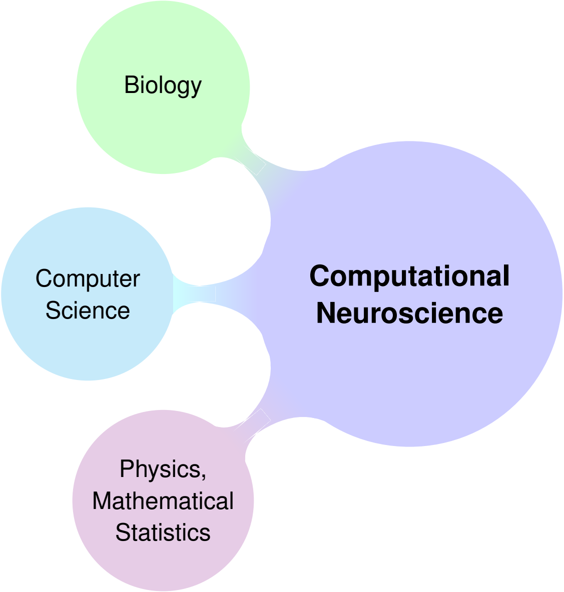
Multidisciplinarity (by definition) ...



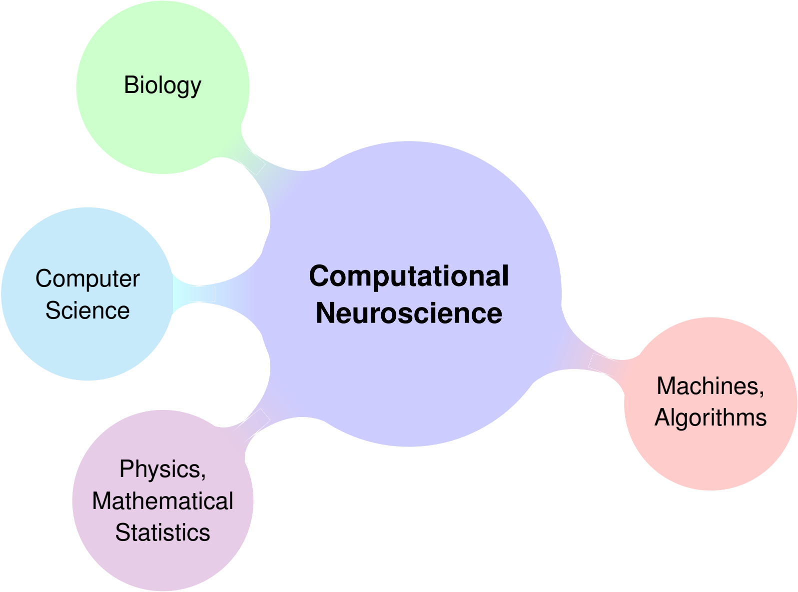
Multidisciplinarity (by definition) ...



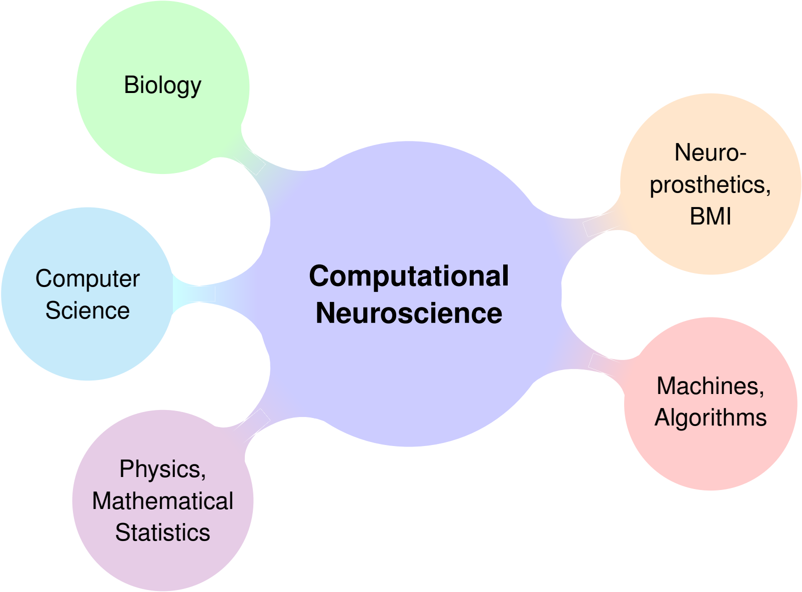
Multidisciplinarity (by definition) . . .



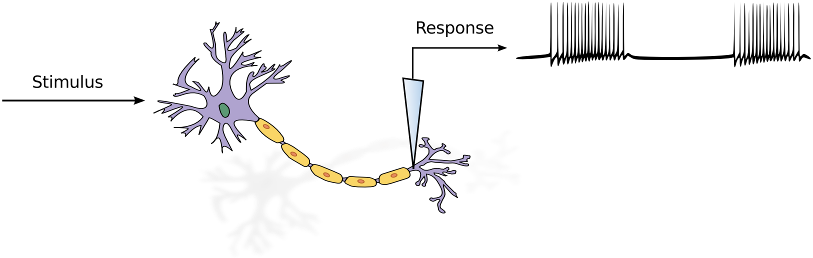
. . . *Fundamental Research* + Applications



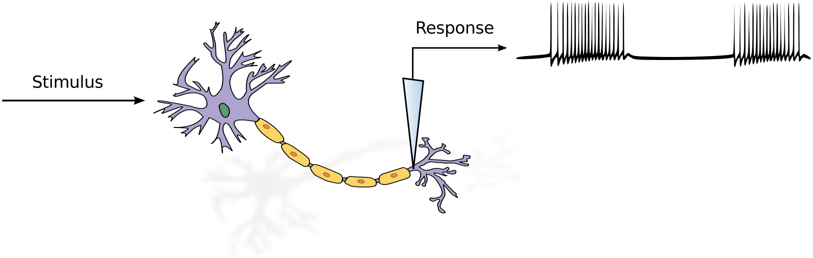
. . . *Fundamental Research* + Applications

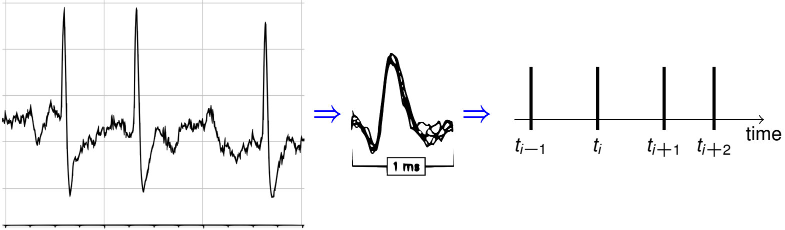


Neuronal code: basic assumptions

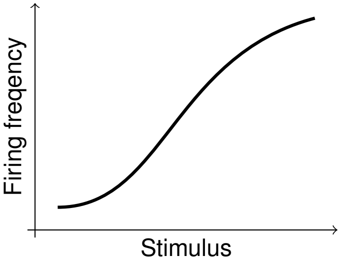


Neuronal code: basic assumptions



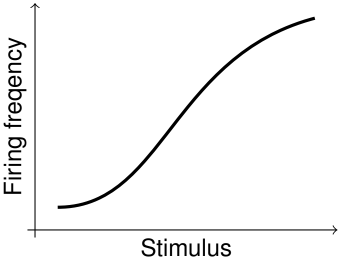


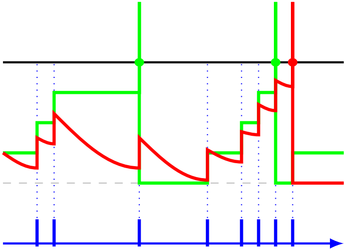
Frequency *vs.* temporal coding



1. Frequency: Adrian (1926), number of pulses (AP) per unit time

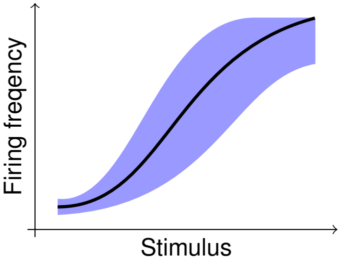
Frequency *vs.* temporal coding

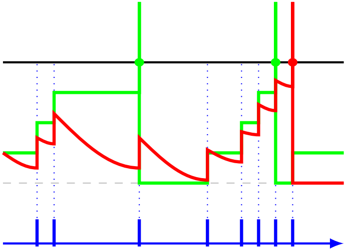




1. Frequency: Adrian (1926), number of pulses (AP) per unit time
2. Temporal: Perkel & Bullock (1968), intervals between APs,

Frequency *vs.* temporal coding



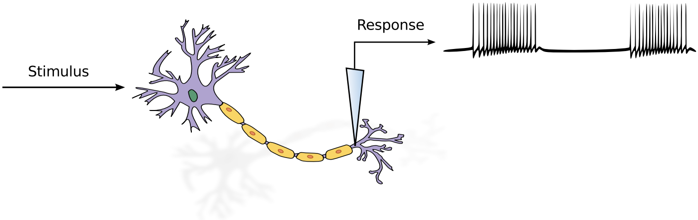


1. Frequency: Adrian (1926), number of pulses (AP) per unit time
2. Temporal: Perkel & Bullock (1968), intervals between APs,

- *Variability, noise*: Stein et al., *Nat. Rev. Neurosci.* 2005

Motivation

- Neural coding: How neurons (populations) encode and process information about their environment?



- *Indirect:* degree to which the response reflects the stimulus

1. "How much information?" (stimulus→response)
   Mutual information (bits)
   MacKay & McCulloch (1952), Stein (1967), Laughlin (1981), Bialek *et al.*, ...
2. Coding precision: the accuracy of stimulus identification
   Fisher information (Cramér-Rao bound):
   Paradiso (1988), Stemmler (1996), Abbott & Dayan (1999), Greenwood *et al.*

Efficient coding hypothesis

Horace B. Barlow, 1961

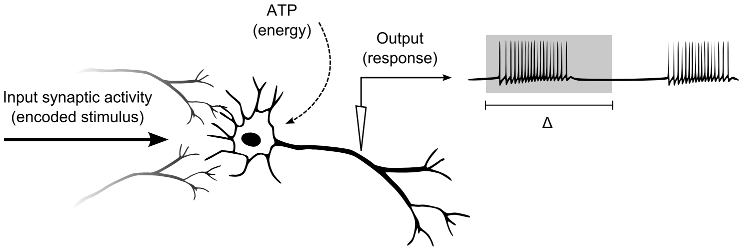
Neurons are adapted, through both evolutionary and developmental processes, to the *statistical characteristics* of their *natural* stimulus.

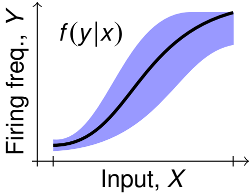
Powerful organizing and *predictive* principle:

- retinal neurons (fly) — Laughlin (1981)
- receptive fields in V1 — Olshausen & Field, *Nature* (1996)
- auditory cortex — Carlson & DeWeese, *PLoS Comput. Biol.* (2012)
- pheromone reception — Kostal *et al*, *PLoS Comput. Biol.* (2008)

Extensions: *metabolic cost*, decoding feasibility, realistic models, . . .

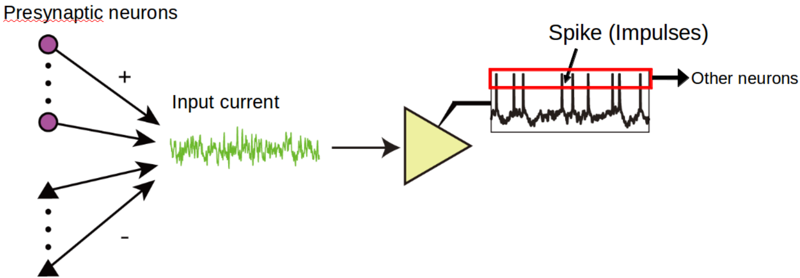
Energy-efficient neural coding?





- Input (exc. conductance, X) – Output (response, Y)
  - Rate coding: $\#$APs in Δ
  - Temporal coding, ...
  - Model parameters, type of stimulation, ...
- Model: $f(y|x)$✔ but $p(x)$**?**
- Efficiency: **Energy** $\times$ **Information** $\times$ **...**

Investigated neuronal model (*cortical excitatory*)



- Extended Hodgkin-Huxley + *point-conductance* (*stochasticity*)
  - Adaptation (I_M)
  - Balanced input, $\lambda_E \propto \lambda_I$
  - Excitatory and inhibitory conductances, $\langle g_{E,I} \rangle \propto \lambda_{E,I}$
  - Effective reversal potential: V_r (Miura *et al.*, 2007)
- Input, $x \equiv \langle g_E \rangle$: mean excitatory conductance (input parameter)
- Output, $y \equiv \#\text{APs}/\Delta$: firing rate

Metabolic cost of neuronal activity & efficiency

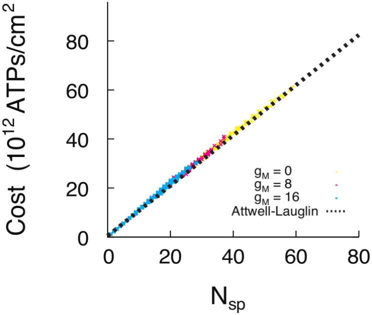
- *Empirical* metabolic cost given $X = x$ (Attwell & Laughlin, 2001)

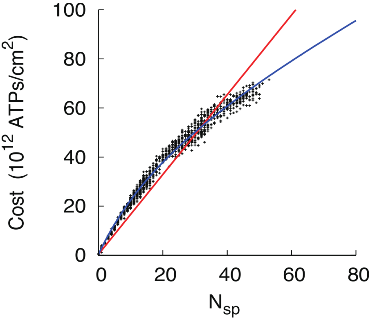
$$w(x) = \kappa \times \big(\langle \#\text{APs in } \Delta \rangle \,|\, x\big) + \beta\,\Delta$$

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

- *Theoretical* (model) cost: only small corrections (RK)✔

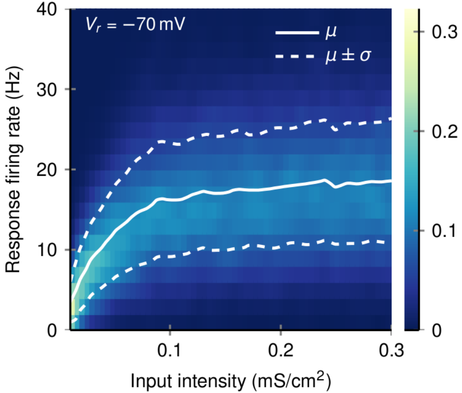
- Excitatory *vs.* inhibitory neurons





Neuronal model: “input-output” relationship

- Procedure: x constant during Δ + repeated trials
- Model $\equiv \Pr(y|x)$ is stochastic (unreliable)



How to *quantify* information transmission?

Methods: mutual information $I(X;Y)$

- Given the “model” $f(y|x)$ and input distribution $X \sim p(x)$:

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

- Interpretation of $I(X;Y)$ – why $I(X;Y)$?

Methods: mutual information $I(X;Y)$

- Given the "model" $f(y|x)$ and input distribution $X \sim p(x)$:

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

- Interpretation of $I(X;Y)$ – why $I(X;Y)$?

1. General: Average statistical dependence (nonlinear) between X and Y
2. "Information transfer" (Shannon): maximum information that can be communicated *reliably* by neuronal 'model' $f(y|x)$ subject to the input statistics $p(x)$
   – conditions are not automatically guaranteed!

Methods continued: capacity-cost function $C(W)$

- *Efficient coding hypothesis*: find the ultimate bounds
- Vary the input distribution $p(x)$ to maximize $I(X;Y)$

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

- *Non-linear* convex optimization problem in $p(x)$ n.p.
- $C(W)$: max. information transfer if the average metabolic cost is less than W
- Classical channel capacity: $C = C(W \to \infty)$

Methods continued: capacity-cost function $C(W)$

- *Efficient coding hypothesis*: find the ultimate bounds
- Vary the input distribution $p(x)$ to maximize $I(X;Y)$

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

- *Non-linear* convex optimization problem in $p(x)$ n.p.
- $C(W)$: max. information transfer if the average metabolic cost is less than W
- Classical channel capacity: $C = C(W \to \infty)$
- Optimal "trade-off" between information and cost (if it exists), the efficiency:

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

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

Finding the optimal input distribution $p(x)$

- Closed-form $p(x)$: only handful of cases (AWGN, ...)
- Structure of optimal $p(x)$: discrete *vs.* continuous *vs.* mixed

- 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 $p(X)$ supported at most K points (finite dimensionality, almost no assumptions on X!)

- **Numerical methods** (converging $C_{\mathrm{low}} \leq C \leq C^{\mathrm{up}}$): Blahut-Arimoto (discrete, 1972), Jimbo-Kunisawa (1979), Davisson-Chang (1988), ...
- Cutting plane (sequence of linear prog.): Huang & Meyn, 2005

Methods: information optimization

- $L[p]$: functional over convex and compact set $\mathcal{F}$ (Smith, 1971)

$$\delta_g L[p_0] = \lim_{\varepsilon \downarrow 0} \frac{L[(1-\varepsilon)p_0 + \varepsilon g] - L[p_0]}{\varepsilon}, \quad \varepsilon \in [0,1]$$

- $L[p]$ diff. and strictly convex $\cap$: unique max. exists at p_0:

$$\delta_g L[p_0] \leq 0 \quad \text{for all } g \in \mathcal{F}$$

Methods: information optimization

- $L[p]$: functional over convex and compact set $\mathcal{F}$ (Smith, 1971)

$$\delta_g L[p_0] = \lim_{\varepsilon \downarrow 0} \frac{L[(1-\varepsilon)p_0 + \varepsilon g] - L[p_0]}{\varepsilon}, \quad \varepsilon \in [0,1]$$

- $L[p]$ diff. and strictly convex $\cap$: unique max. exists at p_0:

$$\delta_g L[p_0] \le 0 \quad \text{for all } g \in \mathcal{F}$$

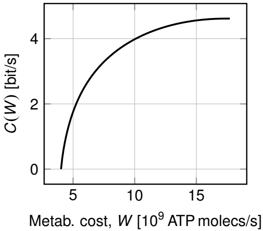
- Finding $C(W)$: strictly convex $\cap$ in $p(x)$, e.g.,

$$\text{''}\, L[p] = I(X;Y) - \lambda \int p(x) w(x)\, \mathrm{d}x\text{''}$$

$$\delta_g L[p] = \int_X g(x) \left[D_{\mathrm{KL}}[f(y|x) \,\|\, p(y)] - 1 - \beta - \lambda w(x)\right] \mathrm{d}x,$$

Results: max. information, efficiency, PS firing rates

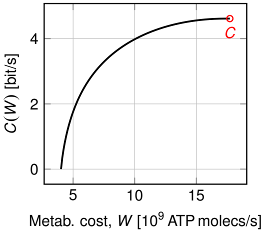
(a) Capacity-cost, $C(W)$



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

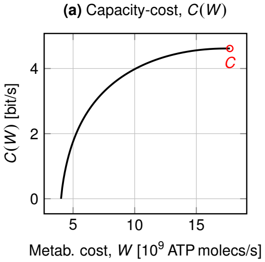
Results: max. information, efficiency, PS firing rates

(a) Capacity-cost, $C(W)$



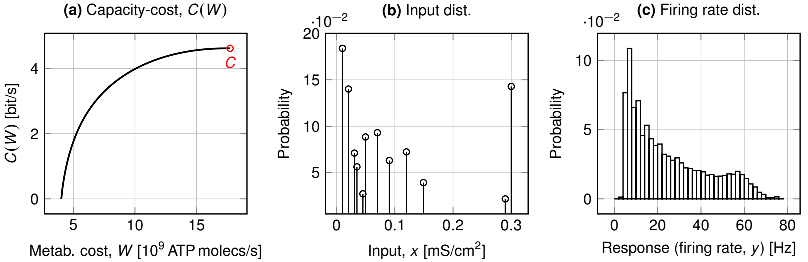
($V_r = -60$ mV, $g_M = 8$ mS/cm^2)

Results: max. information, efficiency, PS firing rates



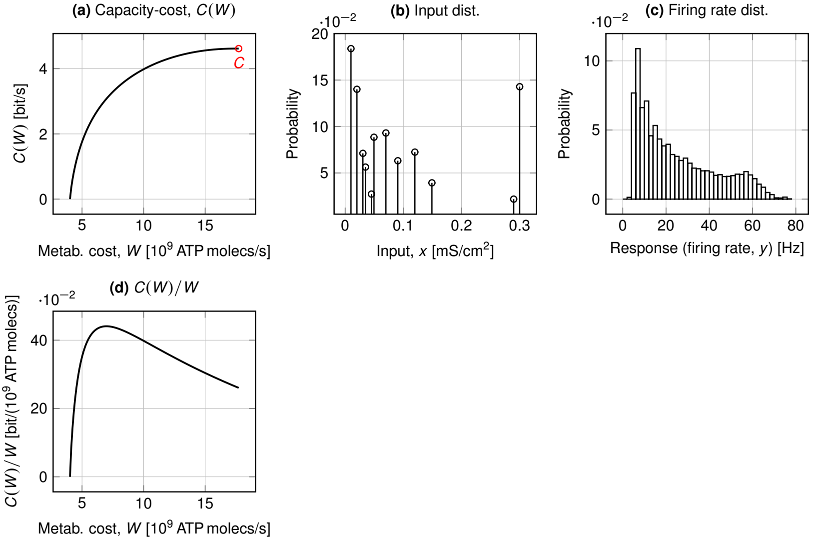
($V_r = -60\,\mathrm{mV}$, $g_M = 8\,\mathrm{mS/cm^2}$)

Results: max. information, efficiency, PS firing rates



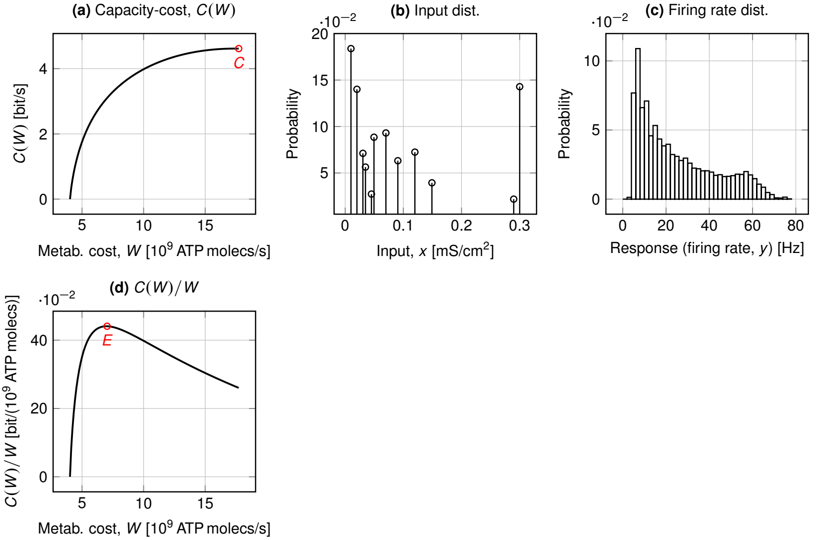
($V_r = -60\,\text{mV}$, $g_M = 8\,\text{mS/cm}^2$)

Results: max. information, efficiency, PS firing rates



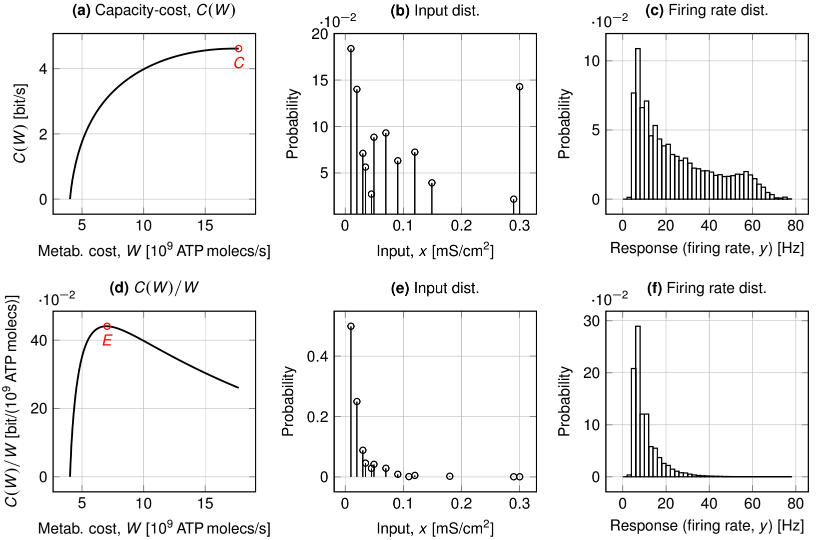
($V_r = -60\,\mathrm{mV}$, $g_M = 8\,\mathrm{mS/cm^2}$)

Results: max. information, efficiency, PS firing rates



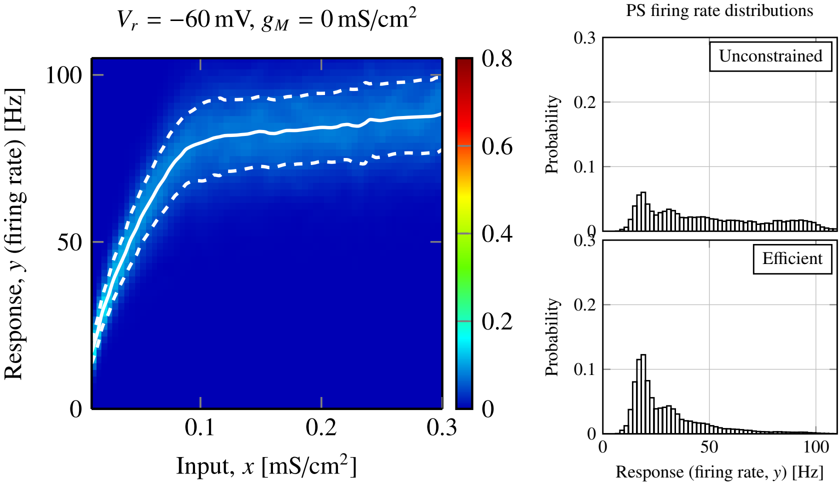
($V_r = -60\,\text{mV}$, $g_M = 8\,\text{mS/cm}^2$)

Results: max. information, efficiency, PS firing rates

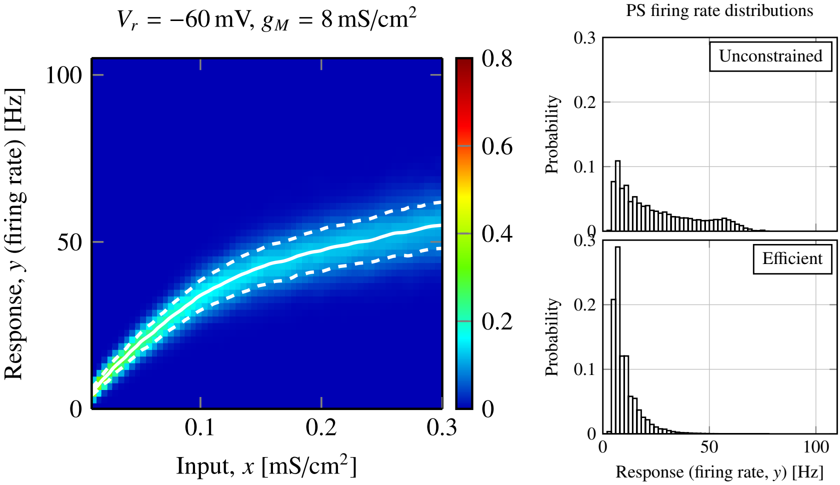


($V_r = -60$ mV, $g_M = 8$ mS/cm^2)

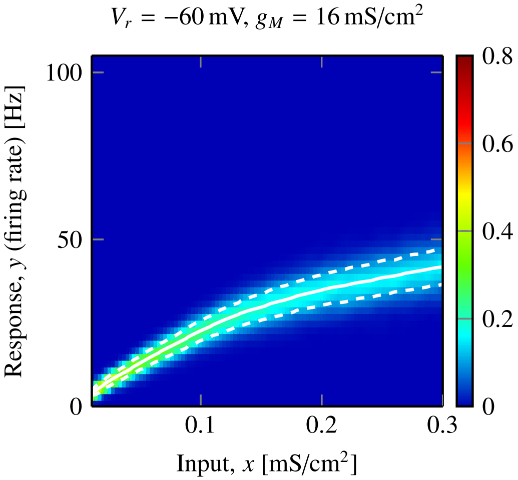
Neuronal model: increasing adaptation, g_M

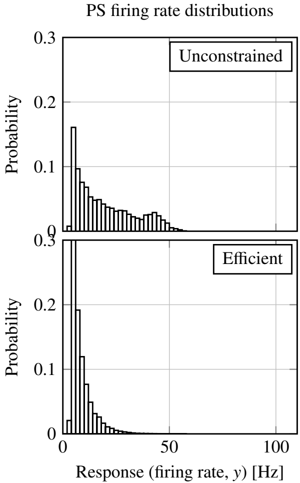


Neuronal model: increasing adaptation, g_M

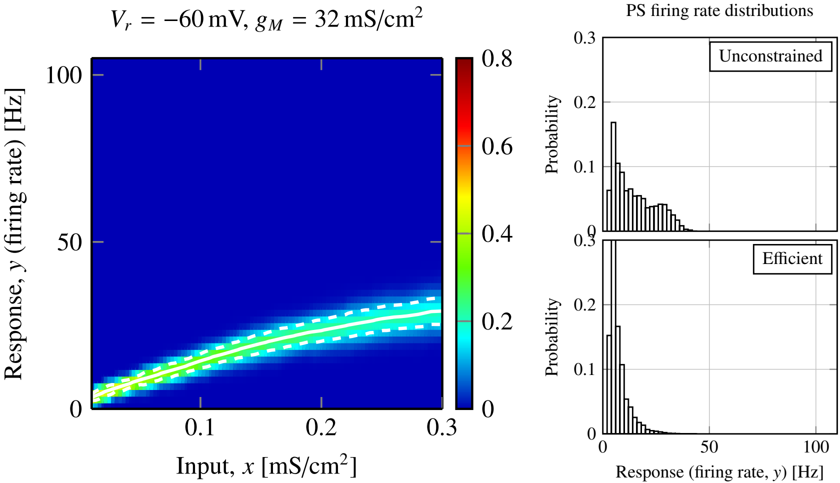


Neuronal model: increasing adaptation, g_M

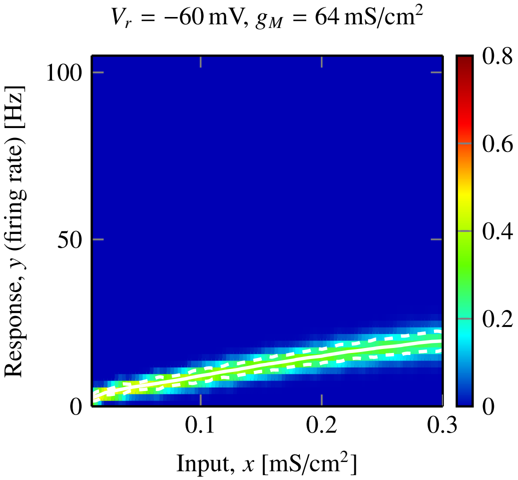


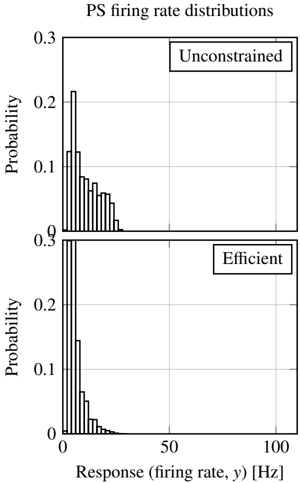


Neuronal model: increasing adaptation, g_M

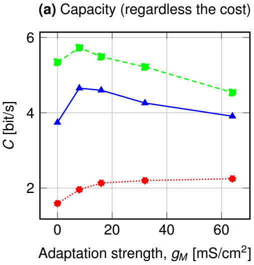


Neuronal model: increasing adaptation, g_M

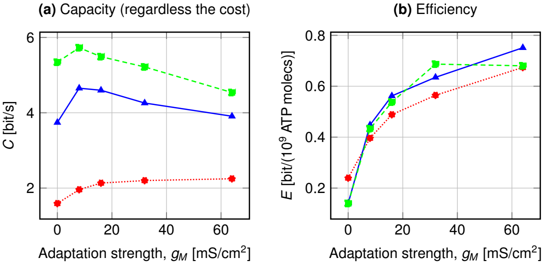




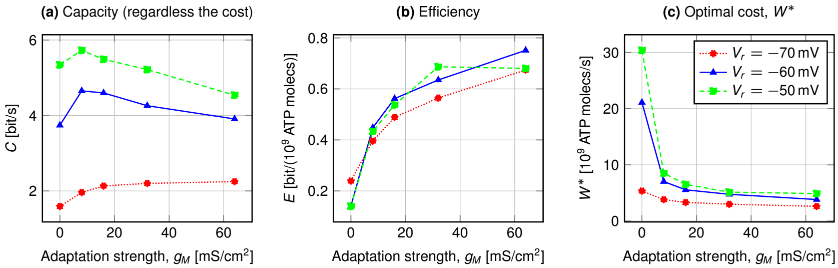
Results: adaptation *vs.* information efficiency



Results: adaptation *vs.* information efficiency



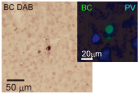
Results: adaptation *vs.* information efficiency

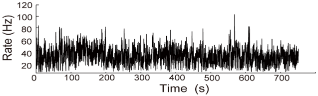


- Preliminary conclusions:
  - Adaptation facilitates the efficiency
  - Adaptation reduces the optimal metabolic workload

Comparison with experimental data

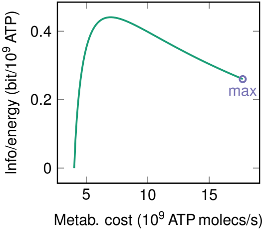
Experimental data:
in vivo (recordings > 30 min.),
layer 2-6 (sensorimotor cortex),
pyramidal and inter-neurons



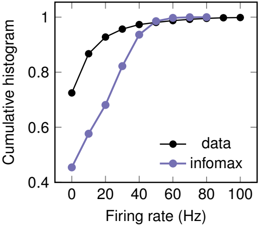


Data: Dr. Tomoki Fukai (RIKEN)

Information/energy

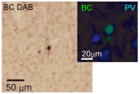


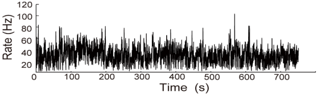
Pyramidal neuron (example)



Comparison with experimental data

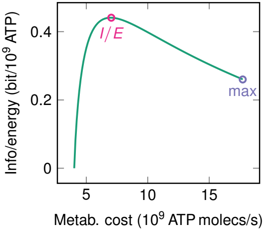
Experimental data:
in vivo (recordings $>$ 30 min.),
layer 2-6 (sensorimotor cortex),
pyramidal and inter-neurons



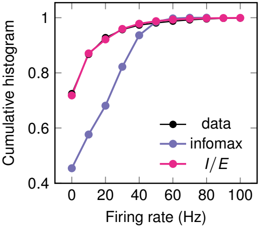


Data: Dr. Tomoki Fukai (RIKEN)

Information/energy

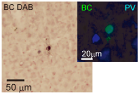


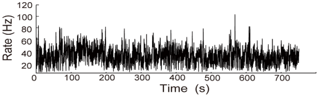
Pyramidal neuron (example)



Comparison with experimental data

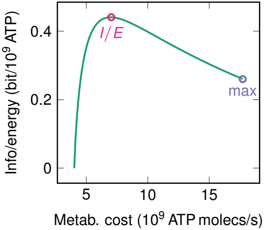
Experimental data:
in vivo (recordings > 30 min.),
layer 2-6 (sensorimotor cortex),
pyramidal and inter-neurons



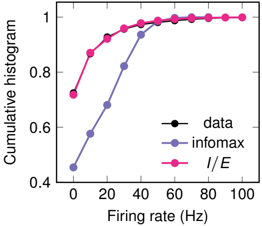


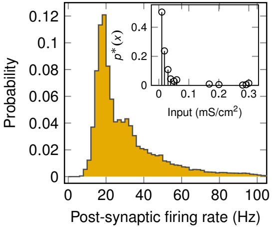
Data: Dr. Tomoki Fukai (RIKEN)

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)
– Low muscarinic-K conductance ($\leq 8\,\mathrm{mS/cm^2}$)

Conclusions

- Metabolic cost considerations seem to have significant impact on the results
- Spike-frequency adaptation improves metabolic efficiency of information transmission,
  e.g., *cost per single bit decreases with increasing g_M*
- For discussion:
  - Interpretation of information-theoretic quantities
  - *Achievability* of information transfer
  - Separated *vs.* joint source-channel coding