

Neural Circuits and Energy Systems for Predictive Coding

Francesco Bullo

Center for Control,
Dynamical Systems & Computation
University of California at Santa Barbara
<https://fbullo.github.io>



**Inaugural Workshop on Active Inference in Italy:
Free-Energy across engineering, complex systems and neurosciences**

23–24 June 2026 — Scuola Superiore Meridionale, Napoli

Acknowledgments



Simone Betteti
Italian Institute of
Artificial Intelligence



Veronica Centorrino
ETH



Anand Gokhale
UC Santa Barbara



William Retnaraj
UC San Diego



Francesca Rossi
Scuola Sup Meridionale
and UCSB



Giacomo Baggio
Univ Padova



Jorge Cortes
UC San Diego



Alexander Davydov
Rice Univ



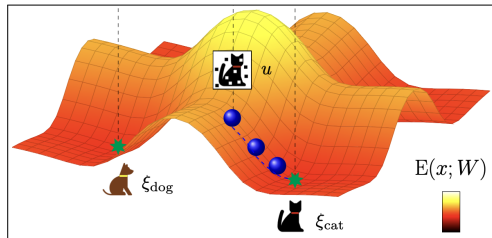
Giovanni Russo
Univ Salerno



Sandro Zampieri
Univ Padova

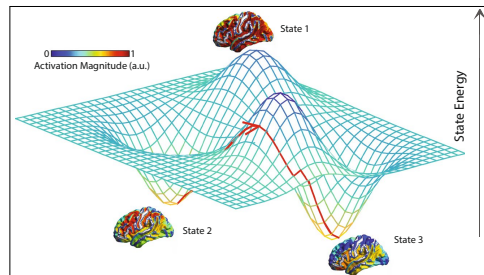
“The idea that the brain functions so as to minimize certain costs pervades theoretical neuroscience.”

S. C. Surace, J.-P. Pfister, W. Gerstner, and J. Brea. On the choice of metric in gradient-based theories of brain function. *PLOS Computational Biology*, 16(4):e1007640, 2020.



Energy landscape for associative memory in Hopfield models

S. Betteti, G. Baggio, F. Bullo, and S. Zampieri. Input-driven dynamics for robust memory retrieval in Hopfield networks. *Science Advances*, 11(17), 2025a.



Energy of neurophysiological activity

S. Gu, M. Cieslak, B. Baird, S. F. Muldoon, S. T. Grafton, F. Pasqualetti, and D. S. Bassett. The energy landscape of neurophysiological activity implicit in brain network structure. *Scientific Reports*, 8(1), 2018.

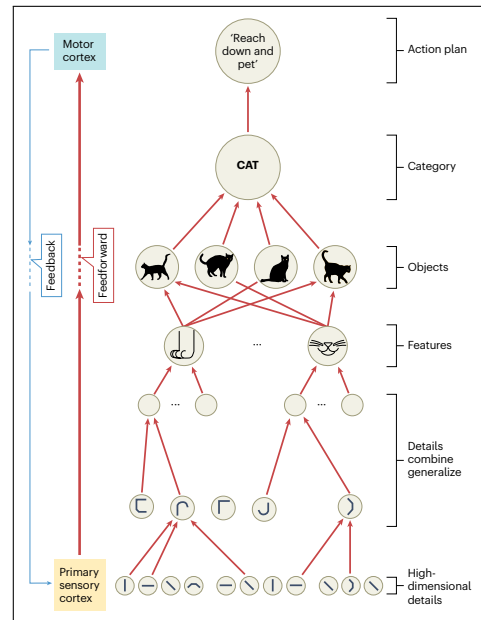
“The idea that the brain actively constructs explanations for its sensory inputs is now generally accepted.”

A. M. Bastos, W. M. Usrey, R. A. Adams, G. R. Mangun, P. Fries, and K. J. Friston. Canonical microcircuits for predictive coding. *Neuron*, 76(4):695–711, 2012.

Predictive coding: hierarchical generative cortex model

R. P. N. Rao and D. H. Ballard. Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1):79–87, 1999.

L. F. Barrett and E. K. Miller. Categorization is ‘baked’ into the brain. *Nature Reviews Neuroscience*, 27(6):435–456, 2026.



Firing-rate network with *stimulus* u :

$$\dot{x} = F_{\text{FR}}(x) := -x + \sigma(Wx + Bu)$$

where σ is *activation function*, W is *synaptic matrix*, and Bu is *stimulus*

- ❶ What *functionality* does F_{FR} implement?
- ❷ What *energy* does F_{FR} minimize?
- ❸ Can such circuits actively *construct explanations* of sensory input?
- ❹ Is there an *optimization-based top-down framework* that derives neural circuits from a mathematical objective?

regularized energy $\mathcal{E}_{\text{regularized}}$ = nominal energy $\mathcal{E}$ + regularizer $\mathcal{R}$



activation function σ , synaptic matrix W , and input matrix B

firing-rate networks are **proximal gradient descent** of a regularized energy

V. Centorrino, A. Gokhale, A. Davydov, G. Russo, and F. Bullo. Positive competitive networks for sparse reconstruction. *Neural Computation*, 36(6):1163–1197, 2024.

proximal operator for the regularizer $\mathcal{R}$

$$\text{prox}_{\mathcal{R}}(x) := \underset{z \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ \mathcal{R}(z) + \frac{1}{2} \|x - z\|_2^2 \right\}$$

- simple regularized problem
- the quadratic term keeps optimal point close to input x
- the proximal operator turns x into a “ $\mathcal{R}$ -better” point

$$\min_x \underbrace{\mathcal{E}(x, u)}_{\text{nominal}} + \underbrace{\mathcal{R}(x)}_{\text{regularizer}}$$

proximal gradient descent:

$$\dot{x} = -x + \text{prox}_{\mathcal{R}}(x - \nabla_x \mathcal{E}(x, u)) =: F_{\text{ProxG}}(x, u)$$

energy system, determined by the energies $\mathcal{E}$ and $\mathcal{R}$

(just like gradient descent $\dot{x} = -\nabla_x \mathcal{E}$ is determined by the energy $\mathcal{E}$)

$$\begin{aligned}\dot{x} &= F_{\text{FR}}(x, u) &:=& -x + \sigma(Wx + Bu) \\ \dot{x} &= F_{\text{ProxG}}(x, u) &:=& -x + \text{prox}_{\mathcal{R}}(x - \nabla_x \mathcal{E}(x, u))\end{aligned}$$

- ❶ $\sigma(x) = \text{prox}_{\mathcal{R}}(x)$
- ❷ $\mathcal{E}$ is quadratic in (x, u)



$$F_{\text{ProxG}} = F_{\text{FR}}$$

Linear-Gaussian generative model

$$y = \Phi x + \text{noise}$$

- *latent variable* x with prior $\exp(-\mathcal{R})$
- *sensory input* y
- *dictionary* matrix Φ
- noise is Gaussian with *precision* matrix Π

Bayesian inference. Compute the maximum a-posteriori estimate of x given y :

$$\begin{aligned} \operatorname{argmax}_x p(x | y) &= \operatorname{argmin}_x \left\{ -\log p(y | x) - \log p(x) \right\} \\ &= \operatorname{argmin}_x \left\{ \frac{1}{2} \|y - \Phi x\|_{\Pi}^2 + \mathcal{R}(x) \right\} =: \operatorname{argmin}_x \mathcal{E}_{\text{MAP}}(x) \end{aligned}$$

Proximal gradient descent of $\mathcal{E}_{\text{MAP}}(x) = \frac{1}{2}\|y - \Phi x\|_{\Pi}^2 + \mathcal{R}(x)$

$$\dot{x} = -x + \text{prox}_{\mathcal{R}}(Wx + \Phi^{\top}\Pi y)$$

activation function $\text{prox}_{\mathcal{R}} :=$ non-linearity determined by the prior $\mathcal{R}$

synaptic matrix $W := I_n - \Phi^{\top}\Pi\Phi$

input matrix $\Phi^{\top}\Pi$

V. Centorrino, A. Gokhale, A. Davydov, G. Russo, and F. Bullo. Positive competitive networks for sparse reconstruction. *Neural Computation*, 36(6):1163–1197, 2024.

Main example: Bayesian inference with generalized measurement models

- *prior on latent variable* x + *exponential-family likelihood on measurement* y

$$x \sim e^{-\mathcal{R}(x)}, \quad p(y|x) = h(y) \exp(y^\top \Phi x - A(\Phi x))$$

- *dictionary* matrix Φ and log-cumulant function A
- *prediction* $g(\Phi x)$, where *prediction function* (canonical response) $g = \nabla A$

Bayesian inference. Compute the maximum a-posteriori estimate of x given y :

$$\begin{aligned} \operatorname{argmax}_x p(x|y) &= \operatorname{argmin}_x \left\{ -\log p(y|x) - \log p(x) \right\} \\ &= \operatorname{argmin}_x \left\{ \underbrace{A(\Phi x) - y^\top \Phi x}_{\text{generalized cross-entropy}} + \mathcal{R}(x) \right\} =: \operatorname{argmin}_x \mathcal{E}_{\text{MAP}}(x) \end{aligned}$$

Main example: Predictive coding network = proximal gradient descent

Proximal gradient descent of $\mathcal{E}_{\text{MAP}}(x) = (A(\Phi x) - y^\top \Phi x) + \mathcal{R}(x)$

$$\dot{x} = -x + \text{prox}_{\mathcal{R}}(x + \Phi^\top \epsilon),$$

$$\epsilon = y - g(\Phi x) \quad (\text{prediction error})$$

which generalizes the linear-Gaussian case

$$\dot{x} = -x + \text{prox}_{\mathcal{R}}(Wx + \Phi^\top \Pi y), \quad W = I_n - \Phi^\top \Pi \Phi$$

Two independent nonlinear knobs:

activation $\text{prox}_{\mathcal{R}} :=$ nonlinearity set by the prior $\mathcal{R}$

prediction $g(\Phi x) :=$ dictionary matrix and nonlinearity set by the likelihood

Catalog #1: prior $\mathcal{R}$ and corresponding activation function

each activation σ is* proximal operator of some regularizer $\mathcal{R}$
different σ used in different cortical areas $\equiv$ different Bayesian priors learned by brain

Prior / regularizer $\mathcal{R}(x)$	Activation $\text{prox}_{\mathcal{R}}(v)$	Biological interpretation
<i>Gaussian</i> : $\frac{\lambda}{2} \ x\ _2^2$	$v/(1 + \lambda)$	linear network
<i>Laplacian</i> : $\lambda \ x\ _1$	$\text{sign}(v) \max(0, v - \lambda)$	sparse coding
<i>non-negativity</i> : $\iota_{\{x \geq 0\}}(x)$	$\text{relu}(v)$	nonneg. $f-I$
<i>box</i> : $\iota_{[0,1]}(x)$	$\text{sat}_{[0,1]}(v)$	bounded nonneg. $f-I$
<i>entropic barrier</i> :	$\text{softmax}(v)$	divisive normalization

Catalog #2: likelihood function and corresponding prediction

each prediction $g(\Phi x)$ is the mean of an exponential-family likelihood
different g used in different cortical areas $\equiv$ different likelihood function learned by brain

Likelihood	Prediction $g(\Phi x)$	Biological interpretation
<i>Gaussian</i>	Φx	graded continuous signals
<i>Bernoulli</i>	$\sigma(\Phi x)$	binary detection (present/absent)
<i>Categorical</i>	$\text{softmax}(\Phi x)$	categorical identity (objects)
<i>Poisson</i>	$e^{\Phi x}$	spike counts (Linear-Nonlinear-Poisson coding)

Catalog #1: the **activation / constraint** = the *prior on latent x*

Catalog #2: the **data / noise model** = the *noise in measurement y*

complementary, not redundant: #1 = “*what prior*”, #2 = “*what’s the measurement*”

§1. Chapter #1: Energy landscapes and predictive coding

§2. Chapter #2: Neural circuits from regularized energies

- Proximal gradient descent
- Case study #1: Sparse signal reconstruction
- Case study #2: Policy composition via free energy
- Case study #3: Hierarchical predictive coding

§3. Conclusion and ongoing research

Case study #1: Sparse signal reconstruction

$$\mathcal{E}_{\text{positive-lasso}}(x) := \frac{1}{2} \|u - \Phi x\|_2^2 + \lambda \|x\|_1 + \iota_{\mathbb{R}_{\geq 0}}(x)$$

where Φ *overcomplete dictionary matrix*, with $\|\Phi_i\| = 1$ and $\Phi_i^\top \Phi_j = \text{similarity between } (i, j)$



resulting proximal gradient dynamics = **positive competitive network**:

$$\dot{x} = -x + \text{relu}\left((I_n - \Phi^\top \Phi)x + \Phi^\top u - \lambda \mathbb{1}_n\right)$$

C. J. Rozell, D. H. Johnson, R. G. Baraniuk, and B. A. Olshausen. Sparse coding via thresholding and local competition in neural circuits. *Neural Computation*, 20(10):2526–2563, 2008.

V. Centorrino, A. Gokhale, A. Davydov, G. Russo, and F. Bullo. Positive competitive networks for sparse reconstruction. *Neural Computation*, 36(6):1163–1197, 2024.

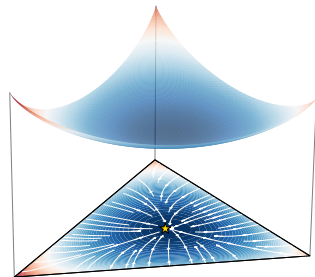
Case study #2: Policy composition via free energy

$$\min_{u \text{ policy probabilities}} \left\{ \underbrace{\text{surprise}(x, u)}_{\text{prior belief vs actual outcomes}} - T \underbrace{\text{entropy}(u)}_{\text{uncertainty}} \right\} \quad (\text{free energy})$$



resulting proximal gradient dynamics = **softmax gradient descent**:

$$\dot{u} = -u + \text{softmax}(-T^{-1} \nabla_u \text{surprise}(x, u))$$



§1. Chapter #1: Energy landscapes and predictive coding

§2. Chapter #2: Neural circuits from regularized energies

- Proximal gradient descent
- Case study #1: Sparse signal reconstruction
- Case study #2: Policy composition via free energy
- Case study #3: Hierarchical predictive coding

§3. Conclusion and ongoing research

From centralized MAP to hierarchical PC: variable splitting

Sensory input is generated by a *deep hierarchy* of linear-Gaussian models:

$$y \approx \Phi^{(1)} \Phi^{(2)} \dots \Phi^{(L)} x$$

Centralized MAP energy:

$$\mathcal{E}_{\text{MAP}}(x) = \frac{1}{2} \|y - \Phi^{(1)} \dots \Phi^{(L)} x\|_{\Pi^{(L)}}^2 + \mathcal{R}(x)$$

- ❶ latent variables $x^{(\ell)}$ at each level (except $x^{(0)} = y$ measurement)
- ❷ $x^{(\ell-1)} \approx \Phi^{(\ell)} x^{(\ell)}$
- ❸ per-level priors $\mathcal{R}^{(\ell)}$

Hierarchical MAP energy:

$$\mathcal{E}_{\text{H-MAP}}(x^{(1:L)}) = \sum_{\ell=1}^L \frac{1}{2} \|x^{(\ell-1)} - \Phi^{(\ell)} x^{(\ell)}\|_{\Pi^{(\ell)}}^2 + \sum_{\ell=1}^L \mathcal{R}^{(\ell)}(x^{(\ell)})$$

R. P. N. Rao and D. H. Ballard. Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1):79–87, 1999.

Hierarchical proximal gradient descent

Prediction estimator at each level $\ell = 1, \dots, L$:

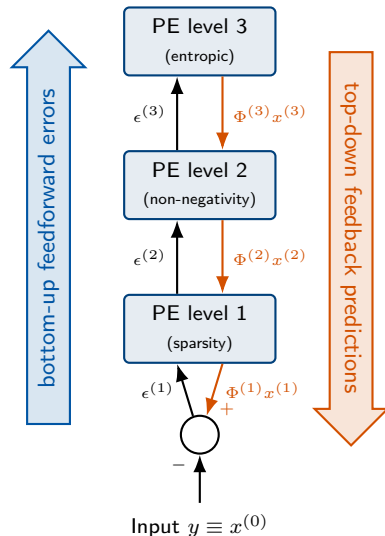
$$\dot{x}^{(\ell)} = -x^{(\ell)} + \text{prox}_{\mathcal{R}^{(\ell)}}(x^{(\ell)} - \nabla_{x^{(\ell)}} \mathcal{E}_{\text{H-MAP}})$$

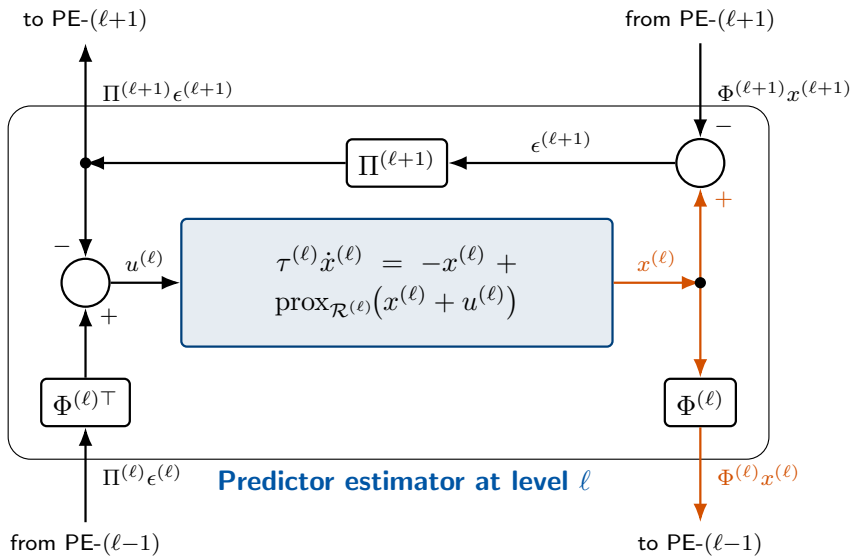
where

$$\nabla_{x^{(\ell)}} \mathcal{E}_{\text{H-MAP}} = \underbrace{\epsilon^{(\ell+1)}}_{\text{top-down}} - \underbrace{\Phi^{(\ell)\top} \epsilon^{(\ell)}}_{\text{bottom-up}}$$

$$\text{error } \epsilon^{(\ell)} = x^{(\ell-1)} - \Phi^{(\ell)} x^{(\ell)}$$

- centralized prox-gradient
= hierarchy of local prox-gradient steps
- each level in closed-loop with neighbors
- each level features: $\text{prox}_{\mathcal{R}^{(\ell)}}$, $\tau^{(\ell)}$, and $\Pi^{(\ell)}$





- bottom-up precision-weighted error enters $+$ port via $\Phi^{(\ell)\top}$
- top-down expectation discrepancy enters $-$ port via $\Pi^{(\ell+1)}$

Probabilistic graphical model: an MRF, not a directed chain

Proposed model is **not directed chain** but **undirected MRF**.

Per-level priors $e^{-\mathcal{R}^{(\ell)}}$ would **over-specify** it.

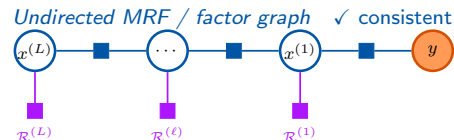
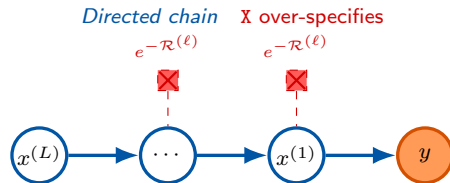
The consistent reading is an **undirected chain MRF**,

$p(y, x^{(1:L)}) = \frac{1}{Z} e^{-\mathcal{E}_{\text{H-MAP}}}$, with

$$\mathcal{E}_{\text{H-MAP}} = \sum_{\ell} \underbrace{\frac{1}{2} \|x^{(\ell-1)} - \Phi^{(\ell)} x^{(\ell)}\|_{\Pi^{(\ell)}}^2}_{\text{edge (pairwise) pot.}} + \sum_{\ell} \underbrace{\mathcal{R}^{(\ell)}(x^{(\ell)})}_{\text{node potential}}$$

Markov property: $x^{(\ell)} \perp \text{rest} \mid x^{(\ell-1)}, x^{(\ell+1)}$.

Locality of $u^{(\ell)} = \text{conditional independence}$



■ edge / pairwise potential ■ node potential $\Rightarrow \text{prox}_{\mathcal{R}^{(\ell)}} \text{ at every level}$

Directed chain = special case $\mathcal{R}^{(\ell)} = 0$ ($\ell < L$),
 $\mathcal{R}^{(L)} = -\log p(x^{(L)})$.

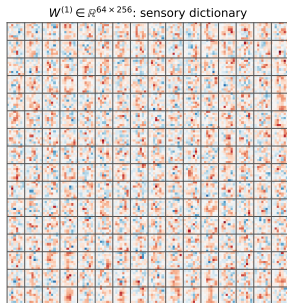
Example: hierarchical classification circuit

Unified normative framework based on proximal gradient:
from continuous feature extraction to discrete classification

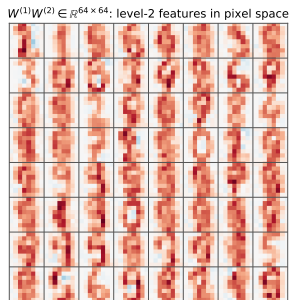
	prior	computational motif
level 1: sensory	Laplace	soft-thresholding , sparse V1 features
level 2: intermediate	non-negativity	relu , non-negative shapes*
level 3: apex	entropy	softmax , categorical decision

Numerical validation

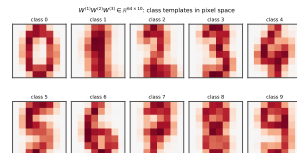
- sklearn digits dataset (1797 grayscale 8×8 images of handwritten digits, 10 classes, pixel values normalised to $[0, 1]$; 80/20 train/test split).
- combined with proximal Hebbian learning



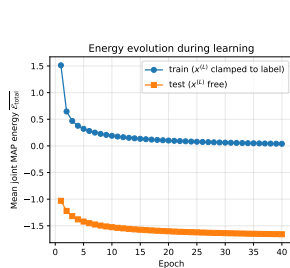
(a) Sensory dictionary $W^{(1)}$ as 8×8 patches.



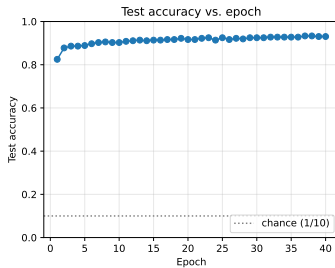
(b) Level-2 features $W^{(1)}W^{(2)}$ shown in pixel space.



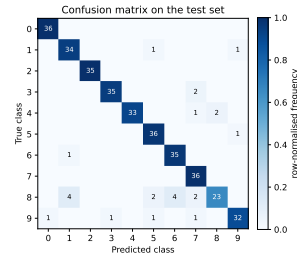
(c) Class templates $W^{(1)}W^{(2)}W^{(3)}$ in pixel space, one per class $k = 0, \dots, 9$.



(d) Mean joint MAP energy on train (with $x^{(L)}$ clamped) and test (with



(e) Top-1 test accuracy vs. epoch.



(f) Test-set confusion matrix at final epoch.

Prior art

- PC as hierarchical Bayesian inference
Rao & Ballard, 1999; Lee & Mumford, 2003; Friston, 2008
- Sparse coding & LCA
Olshausen & Field, 1996; Rozell et al., 2008
- Divisive normalization; canonical microcircuits
Carandini & Heeger, 2012; Bastos et al., 2012
- “Prior → prox” activation dictionary in optimization
Combettes & Pesquet, 2011
- Per-level priors; mixed continuous/discrete models
Lee & Mumford, 2003; Friston et al., 2017
- Biological nonnegative matrix factorization
Lee & Seung 2000, Spratling 2009, Pehlevan & Chklovskii 2014

Our contributions

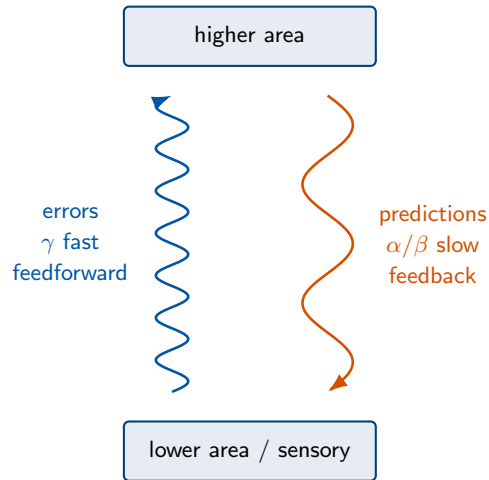
- 1 **Prox gradient descent** as a normative dynamical framework for *flat/hierarchical cortical circuits*
Centorrino et al., 2023; Gokhale et al. 2024; Bullo 2026
- 2 **Positive competitive nets** for sparse reconstruction
Centorrino et al., 2023
- 3 **Entropy-softmax prior** for policy composition and prediction at the categorical end
Rossi et al., 2025
- 4 **Decreasing energy, convergence, contractivity**
Davydov et al., 2023; Gokhale et al., 2024
- 5 **Asymmetric networks and E^k -I motifs**
Betteti et al., 2026 (in progress, arxiv)

Future research: Oscillatory predictive coding

In cortex, the two message streams of predictive coding appear **multiplexed across distinct frequency bands**:

- *bottom-up errors* $\epsilon^{(\ell)}$ ride *fast* feedforward γ rhythms (> 30 Hz)
- *top-down predictions* $\Phi^{(\ell)} x^{(\ell)}$ ride *slow* feedback α/β rhythms ($\sim 8\text{--}30$ Hz)

feedforward γ (errors) and feedback α/β (predictions) form an **oscillatory predictive code**



compelling **phenomenology**, missing **normative “why”**