# A Network of Biologically Inspired Rectified Spectral Units (ReSUs) Learns Hierarchical Features Without Error Backpropagation

**Shanshan Qin[1*], Joshua L. Pughe-Sanford[1], Alexander Genkin[1], Pembe Gizem Ozdil[1,2], Philip Greengard[1], Anirvan M. Sengupta[3,4], Dmitri B. Chklovskii[1,5†]**

[1]Center for Computational Neuroscience, Flatiron Institute, Simons Foundation, New York, NY, USA
[2]EDRS Doctoral Program, Swiss Federal Institute of Technology Lausanne (EPFL), Switzerland
[3]Center for Computational Mathematics, Flatiron Institute, Simons Foundation, New York, NY, USA
[4] Physics Department, Rutgers University, New Brunswick, NJ, USA
[5] Neuroscience Institute, NYU Langone Medical Center, New York, NY, USA
ssqin@sjtu.edu.cn, {jpughesanford,pgreengard,dchklovskii}@flatironinstitute.org,
{alexander.genkin,pgizemozdil,anirvans.physics}@gmail.com

## Abstract

We introduce a biologically inspired, multilayer neural architecture composed of Rectified Spectral Units (ReSUs). Each ReSU projects a recent window of its input history onto a canonical direction obtained via canonical correlation analysis (CCA) of previously observed past–future input pairs, and then rectifies either its positive or negative component. By encoding canonical directions in synaptic weights and temporal filters, ReSUs implement a local, self-supervised algorithm for progressively constructing increasingly complex features.

To evaluate both computational power and biological fidelity, we trained a two-layer ReSU network in a self-supervised regime on translating natural scenes. First-layer units, each driven by a single pixel, developed temporal filters resembling those of Drosophila post-photoreceptor neurons (L1/L2 and L3), including their empirically observed adaptation to signal-to-noise ratio (SNR). Second-layer units, which pooled spatially over the first layer, became direction-selective—analogous to T4 motion-detecting cells—with learned synaptic weight patterns approximating those derived from connectomic reconstructions.

Together, these results suggest that ReSUs offer (i) a principled framework for modeling sensory circuits and (ii) a biologically grounded, backpropagation-free paradigm for constructing deep self-supervised neural networks.

**Code** — https://github.com/ShawnQin/ReSU.git
**Supplemental Material** —
https://github.com/ShawnQin/ReSU.git

## Introduction

Modern deep learning systems outperform human experts on vision and language benchmarks (He et al. 2015; Brown et al. 2020), as well as strategic games (Silver et al. 2016). Yet, they remain conspicuously inferior to biological intelligence in several fundamental aspects. First, contemporary models lack true compositional reasoning and long horizon planning (Keysers et al. 2019), cognitive abilities that humans take for granted. Second, they are vastly less efficient: training state-of-the-art networks requires megawatt-hours of energy and billions of labeled examples (Strubell, Ganesh, and McCallum 2020; Patterson et al. 2021), whereas the human brain draws only $\approx$ 20W (Laughlin 2001) and learns largely from self-supervision. Third, today's models are prone to hallucinations, brittle generalization, and limited motor planning, rarely observed in biological agents (Ji et al. 2023; Recht et al. 2019; Geirhos et al. 2020). If we wish to make AI more powerful, reliable and efficient, understanding the root cause of such gaps is needed.

A potential and often overlooked source of these shortcomings is architectural: nearly all modern AI systems rely on rectified linear units (ReLUs), a mid-20th-century abstraction drawn loosely from early electrophysiology, and are trained by the error backpropagation algorithm (Rumelhart, Hinton, and Williams 1986). ReLUs are inherently static—they sum only simultaneously arriving inputs before rectifying—thereby discarding the rich temporal dynamics that characterize biological neurons. Error backpropagation requires giant labeled training datasets and nonlocal interactions across layers for which no biological substrate has been found.

Decades of experiments in model organisms have now described with cellular precision physiological and anatomical aspects of hierarchical feature emergence. Such growing body of knowledge creates an opportunity to design a more biologically grounded neuronal model, potentially yielding artificial networks that transcend the limitations of current ReLU/backpropagation-based architectures.

## Our Contribution

We introduce a biologically motivated, multi-layer network of Rectified Spectral Units (ReSUs). Each ReSU projects a recent window of its input history onto a canonical direc-

*Present address: Institute of Natural Sciences, Shanghai Jiao Tong University, Shanghai 200240, China.
†Corresponding author.

tion of canonical correlation analysis (CCA) of previously observed past-future input pairs and then rectifies the positive or negative component, implementing a dynamic, potentially local self-supervised algorithm.

To evaluate both computational power and biological fidelity, we applied the ReSU framework to visual motion detection—a well-established, tractable, yet nontrivial benchmark task. The corresponding neural circuit in *Drosophila* is exceptionally well characterized both anatomically and physiologically (Takemura et al. 2013, 2017; Borst and Groschner 2023), making it an ideal guide and testbed. We trained a *Drosophila*-inspired two-layer ReSU network on natural-scene translations, Figure 1, and obtained the following results:

- Layer 1 units, each driven by a single pixel, learn temporal filters similar to those of the L1/L2 and L3 post-photoreceptor neurons of *Drosophila*, including their empirically observed adaptation to SNR.
- Layer 2 units pool outputs of Layer 1 across pixels and develop direction-selective responses analogous to T4 cells; their learned synaptic weight patterns approximate those found in connectomics reconstructions.

These findings demonstrate that ReSU networks: (i) recover salient properties of biological circuits offering a path to principled and interpretable modeling of hierarchical sensory processing, (ii) generate hierarchical features offering a back-prop-free path to constructing more brain-like deep artificial networks. Consequently, a biologically aligned ReSU architecture may inherit the advantages of the human brain over conventional ReLU/backpropagation-based architectures described in the Introduction.

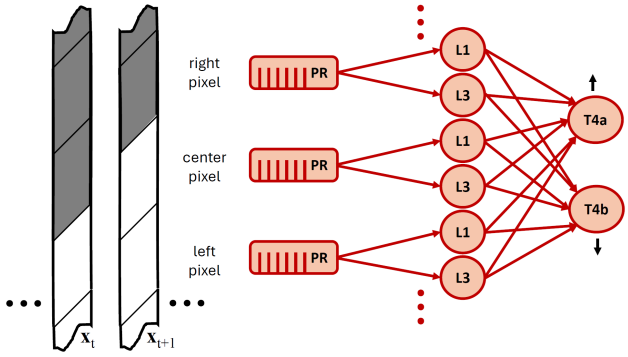


Figure 1: *Drosophila*-inspired two-layer ReSU neural network trained in the self-supervised setting on translating natural images in 1D. Photoreceptors (PR) report the contrast levels of the corresponding pixels. Temporal filters of the first layer neurons (L1, L3) are computed by using CCA on past-future input sequences and act on PR outputs followed by rectification in L1. Second layer neurons (T4) pool information from 3 adjacent pixels. Past-future CCA of the outputs of the first layer learns a directionally selective spatio-temporal filter for T4 (black arrows: preferred directions).

## Related Work

Hierarchical features, like those observed in the cortex, are crucial for solving complex tasks. Yet despite decades of research, the computational principles underlying their emergence in the brain remain unknown.

Unsupervised single-layer networks derived from principled objectives using local (Hebbian and anti-Hebbian) rules and static units can learn principal components and sparse dictionaries (Oja 1982; Olshausen and Field 1997; Hu, Pehlevan, and Chklovskii 2014; Pehlevan and Chklovskii 2019). However, multiple attempts to "stack" such networks into deep architectures failed to generate richer hierarchical representations. Another principled approach is slow feature analysis (SFA) (Wiskott and Sejnowski 2002) which was implemented using local learning rules (Lipshutz et al. 2020). Whereas multi-layer SFA networks exist, obtaining cortex-like features with SFA requires first heuristically guessing hand-made nonlinear features, which is non-trivial.

Single-layer networks using static units and local (non-Hebbian) rules that learn CCA in unsupervised (Lipshutz et al. 2021) and self-supervised (Golkar et al. 2020) settings have been derived from principled objectives. However, because the processing in such networks is linear, unlike in the present work, "stacking" them into multi-layer networks would not produce complex features. Moreover, they do not capture the rich dynamics of biological neurons.

In deep networks built on the predictive coding principle (Rao and Ballard 1999; Rao 2024; Whittington and Bogacz 2017)neurons compute and output prediction errors rather than hierarchical features typically observed in the sensory pathways. Also, neuronal temporal properties in early sensory processing do not seem to align with the predictive coding architecture (Druckmann, Hu, and Chklovskii 2012).

Deep networks constructed using more complex local learning rules integrating self-supervised learning (such as contrastive learning), Hebbian plasticity, and predictive coding can learn hierarchical and invariant object representations (Illing et al. 2021; Dora, Bohte, and Pennartz 2021; Halvagal and Zenke 2023). However, the inherent complexity of these multi-principle frameworks does not facilitate our understanding of the computational primitives operating at the single neuron level, making it difficult to isolate and identify the core mechanisms underlying biologically plausible learning.

Deep networks of static ReLUs trained by backpropagation produce hierarchies matching cortical organization (Yamins et al. 2014) yet rely on error signals or weight symmetries that are difficult to reconcile with neurobiology (Crick 1989; Lillicrap et al. 2020). Efforts to render backpropagation more biologically plausible—random feedback alignment (Lillicrap et al. 2016; Nøkland 2016), dendritic error segregation (Guerguiev, Lillicrap, and Richards 2017), equilibrium propagation(Scellier and Bengio 2017)— result in performance degradation.

Our approach is conceptually related to JEPA (LeCun 2022) and VAMPnets (Mardt et al. 2018), both of which learn representations predictive of future latent states without reconstructing raw inputs. Whereas JEPA and VAMPnets rely on parameterized encoder–predictor networks

trained using backprop end-to-end, ReSU derives analogous predictive primitives analytically via past–future CCA, yielding locally learnable, biologically interpretable units that can be hierarchically composed into deep networks.

A recent application of a backpropagation-trained network of dynamical units to the same *Drosophila* circuit (Lappalainen et al. 2024) produced a mechanism reminiscent of that observed experimentally. However, these networks are trained in a biologically implausible supervised setting using visual motion ground truth and do not consistently reproduce the experimentally observed motion detection computations. This limits their relevance for understanding biologically plausible learning of complex features, particularly in less well-characterized systems.

## Truncated CCA Maximizes Past-Future Mutual Information for OU Processes

We start by postulating that sensory stimuli can be modeled as observations of stochastic dynamical systems. Then the goal of neurons is to learn such dynamics to support prediction and control. As a fully tractable first step towards learning the dynamics of natural stimuli, we consider a discrete-time, linear, time-invariant stochastic system, known as a partially observed multivariate Ornstein–Uhlenbeck (OU) process (Figure 2B) and defined by :

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{v}_t, \tag{1}$$

$$y_t = \mathbf{C}\mathbf{x}_t + w_t, \tag{2}$$

where $\mathbf{x}_t \in \mathbb{R}^n$ is the latent state vector, $y_t \in \mathbb{R}$ is the observed scalar time series, each componet of $\mathbf{v}_t \in \mathbb{R}^n$ and $w_t$ are independent, identically distributed Gaussian noise sources with zero mean and unit covariance. $\mathbf{A}, \mathbf{B}, \mathbf{C}$ are constant but unknown matrices of compatible dimensions.

In this framework, the state vector $\mathbf{x}_t$ acts as an informational interface between past noise inputs, $\mathbf{v}_t$, and future observation, $y_t$. To obtain a reduced-dimensional representation of $\mathbf{x}_t$, that preserves maximal mutual information with the future, we apply balanced truncation (Katayama et al. 2005). Although $\mathbf{x}_t$ is not directly observable, it can be inferred from sequences of partial observations, $y_t$, using past and future lag vectors, as illustrated in Figure 2A:

$$\begin{aligned} \mathbf{p}_t &= [y_t, y_{t-1}, \ldots, y_{t-m+1}]^\top, \\ \mathbf{f}_t &= [y_{t+1}, y_{t+2}, \ldots, y_{t+h}]^\top, \end{aligned} \tag{3}$$

where $m, h \geq n$ denote the lengths of the past (memory) and future (horizon) lag vectors, respectively. Although (centered) $\mathbf{p}_t$ and $\mathbf{f}_t$ are not minimal representations of latent state $\mathbf{x}_t$, they form sufficient statistics for reconstructing its dynamics and estimating its predictive subspace.

The low dimensional latent representation is obtained via a linear projection of the past, Figure 2A:

$$\mathbf{z}_t = \mathbf{\Psi}\mathbf{p}_t, \quad \mathbf{z}_t \in \mathbb{R}^r, \tag{4}$$

where the matrix $\mathbf{\Psi} \in \mathbb{R}^{r \times m}$ compresses the past while retaining maximal information about the future.

For the OU process (Eq.1), the information captured in $\mathbf{z}_t$ about the future can be quantified by the correlation between

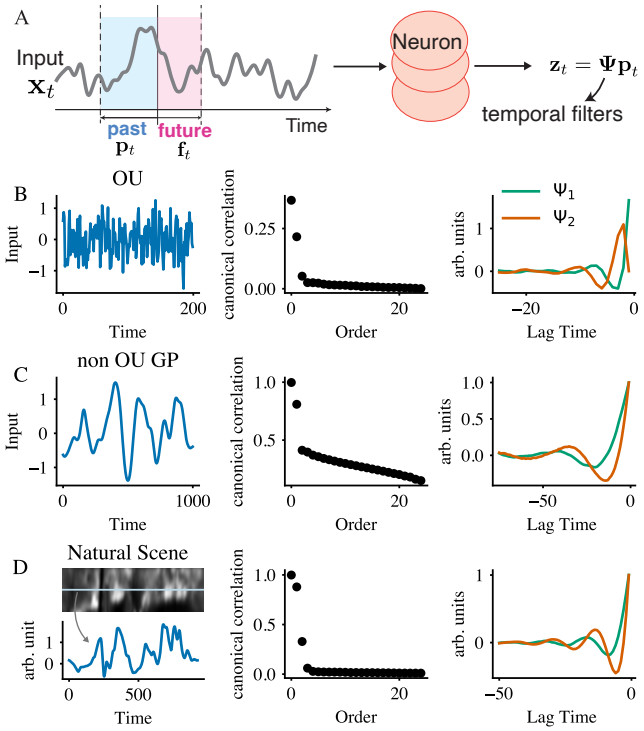


Figure 2: Past–future CCA as a computational primitive applied to different data streams. (A) A sensory stimulus is represented by past and future lag vectors. Each neuron computes its output as a linear projection of the input onto a temporal filter learned from previous observations of the same stimulus. Results of CCA applied to (B) a two-dimensional OU process projected onto one dimension, (C) a non-OU Gaussian process with a rational quadratic kernel, and (D) a contrast profile obtained by scanning a natural image at constant velocity. The first column shows example time series, the second column displays the canonical correlations ($\sigma_i$), and the third column shows the first two normalized temporal filters (canonical directions). In all numerical simulations, a small Gaussian white observation noise was added to the input signal.

$\mathbf{z}_t$ and the whitened future input, $\mathbf{f}_t$, leading to the following constrained optimization problem (Arun and Kung 1990),

$$\max_{\mathbf{\Psi} \in \mathbb{R}^{r \times m}} \left\| \mathbb{E}\left[ \mathbf{C}_{ff}^{-1/2} \mathbf{f}_t (\mathbf{\Psi}\mathbf{p}_t)^\top \right] \right\|_F, \text{ s.t. } \mathbf{\Psi}\mathbf{C}_{pp}\mathbf{\Psi}^\top = \mathbf{I}_r, \tag{5}$$

where the covariance matrices are defined as $\mathbf{C}_{ff} = \mathbb{E}[\mathbf{f}_t\mathbf{f}_t^\top], \mathbf{C}_{fp} = \mathbb{E}[\mathbf{f}_t\mathbf{p}_t^\top], \mathbf{C}_{pp} = \mathbb{E}[\mathbf{p}_t\mathbf{p}_t^\top]$. Substituting $\widetilde{\mathbf{\Psi}} = \mathbf{\Psi}\mathbf{C}_{pp}^{1/2}$ into Eq. 5 yields the equivalent form:

$$\max_{\widetilde{\mathbf{\Psi}} \in \mathbb{R}^{r \times m}} \left\| \mathbf{C}_{ff}^{-1/2} \mathbf{C}_{fp} \mathbf{C}_{pp}^{-1/2} \widetilde{\mathbf{\Psi}}^\top \right\|_F, \text{ s.t. } \widetilde{\mathbf{\Psi}}\widetilde{\mathbf{\Psi}}^\top = \mathbf{I}_r, \tag{6}$$

which is solved via singular value decomposition (SVD) of the whitened cross-covariance matrix $\mathbf{C}_{ff}^{-1/2}\mathbf{C}_{fp}\mathbf{C}_{pp}^{-1/2} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ yielding:

$$\mathbf{\Psi} = \mathbf{V}_r^\top \mathbf{C}_{pp}^{-1/2}, \tag{7}$$

where $\mathbf{V}_r$ consists of the $r$ right singular vectors corresponding to the largest singular values, $\sigma_1 \ldots \sigma_r$, which form the

diagonal of $\mathbf{\Sigma}$, Figure 2B. We also define $\mathbf{\Phi} = \mathbf{U}_r^\top \mathbf{C}_{ff}^{-1/2}$ from the corresponding $r$ left singular vectors. In practice, inverses of covariances may require $l_2$-norm regularization.

This procedure is equivalent to performing CCA between $\mathbf{f}_t$ and $\mathbf{p}_t$, where the singular values $\sigma_i$ represent the canonical correlations (Arun and Kung 1990; Chechik et al. 2003). In other words, each $\sigma_i$ quantifies the correlation between the $i$-th pair of canonical variates, $(\mathbf{\Psi}\mathbf{p}_t)_i$ and $(\mathbf{\Phi}\mathbf{f}_t)_i$.

The mutual information between $\mathbf{z}_t$ and future $\mathbf{f}_t$ is then given by (Arun and Kung 1990; Chechik et al. 2003) (see Supplementary Material):

$$I_r = -\tfrac{1}{2}\sum_{i=1}^{r} \log\bigl(1-\sigma_i^2\bigr). \tag{8}$$

Thus, for an OU process, truncated CCA yields an optimal linear projection of the past lag-vector onto the $r$-dimensional subspace that maximizes the mutual information with the future, providing a principled approach to extracting predictive latent representations from sensory input.

## Truncated CCA Maximizes Past-Future Mutual Information for Gaussian Processes

As the next step towards the dynamics of natural stimuli, we demonstrate that past-future CCA is optimal not just for OU but for a more general class of one-dimensional Gaussian processes (GPs)—specifically, discrete-time GPs with translation-invariant and integrable covariance kernels.

GPs are flexible, non-parametric models widely used in machine learning and statistics, Figure 2C. They are parameterized by covariance kernels that allow practitioners to specify features such as the rate of autocovariance decay in time and spectral domains. This flexibility enables us to generate synthetic data with features that mimic natural stimuli.

Here, we assume that stimuli are generated by GPs defined on a set of equispaced points $t_1, t_2, \ldots$ with stationary and integrable covariance kernel $k(t_i - t_j)$.

Constructing the GP's past/future vectors $\mathbf{p}_t, \mathbf{f}_t$ of (3) to be lag vectors of length $\ell$, the $2\ell \times 2\ell$ covariance matrix $\mathbf{\Sigma}$ of $\mathbf{p}_t, \mathbf{f}_t$ for all $i, j$ is $\mathbf{\Sigma}_{ij} = k(t_i - t_j)$. Similarly, representing $\mathbf{\Sigma}$ as a $2 \times 2$ block matrix, we have

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{C}_{pp} & \mathbf{C}_{pf} \\ \mathbf{C}_{fp} & \mathbf{C}_{ff} \end{bmatrix}. \tag{9}$$

The canonical directions can then be computed as described in the previous section, Figure 2C. For a given lag-vector length $\boldsymbol{p}_t$, CCA maximizes correlation and mutual information with the future. The necessary length of the past lag vector depends on the decay of the covariance kernel, or how far back in time one needs to go until including more past observations is no longer informative about the future.

For most commonly-used GP kernels, the canonical correlations will decay rapidly (Ambikasaran et al. 2016). Intuitively, this means that there's only so much information from the past that can be used to predict the future. For complicated kernels, such as oscillatory ones, the future may depend on many independent features of past observations. For these kernels, the canonical correlations would not decay rapidly (see Supplementary Material for further discussion).

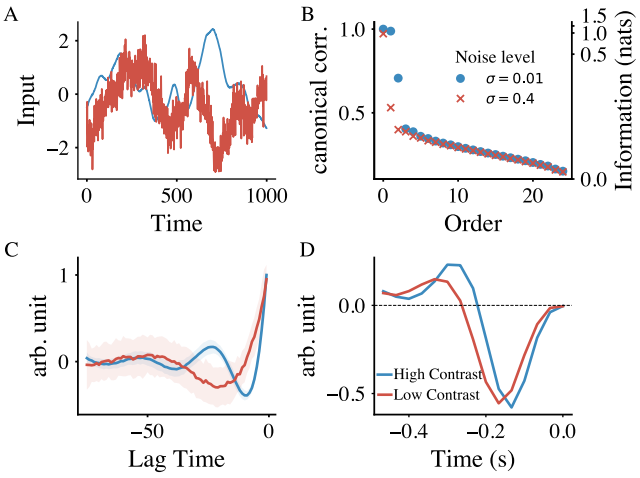


Figure 3: Dependence of past–future CCA results on observation noise in Gaussian processes. (A) Example time series generated from a Gaussian process with a rational quadratic kernel for high (blue) and low (red) SNR, varied through observation noise. (B) Correlation and mutual information between past projections onto canonical directions and the corresponding future signals. (C) Second canonical directions for high (blue) and low (red) SNRs. As observation noise increases, the filter shape transitions from multi-lobed to single-lobed. Shaded regions denote standard deviation across realizations. (D) Experimentally measured temporal filter of a retinal ganglion cell adapts to low contrast, which corresponds to lower SNR (Liu and Gollisch 2015). These filters must pass through the origin (0,0), unlike the model filters (panel C), because physiological filtering is constrained to be both causal and continuous—conditions not enforced in our model.

Next, we demonstrate the truncated CCA framework for GPs through specific examples, Figure 3. We generate a synthetic dataset comprised of time series sampled from a GP with a rational quadratic kernel, i.e., $k(t, t') = \left(1 + \frac{|t-t'|^2}{2\alpha l^2}\right)^{-\alpha}$ for $\alpha, l > 0$. Here, $l = 1$ controls the timescale and $\alpha = 1$ the decay in the time domain of the autocovariance function. We choose this kernel because GPs with this kernel have properties that mimic statistics in natural images, where spatial correlations of contrast decay as a power law (Ruderman and Bialek 1994). Then we derive the past temporal filter by solving the rank-2 CCA optimization problem (5) with $m = 75, h = 50$. The resulting filter $\Psi$ has multiple lobes (blue line in Figure 3C). The latent representation $\boldsymbol{z}_t$ is obtained by projecting the past lag vectors $\mathbf{p}_t$ onto $\Psi$, which is the projection that maximizes mutual information with the future (Figure 3C).

Finally, we investigated how the temporal filter adapts to changes in stimulus statistics, particularly the SNR. Many sensory neurons exhibit such adaptation (Figure 3D) (Srinivasan, Laughlin, and Dubs 1982; van Hateren 1992; Liu and Gollisch 2015), a hallmark of efficient and predictive coding (Srinivasan, Laughlin, and Dubs 1982; van Hateren

1992; Chalk, Marre, and Tkačik 2018). To test whether our model captures this property, we added Gaussian white noise of varying amplitudes to the observed time series and trained the model to derive the corresponding optimal filters. As shown in Figure 3C, under low noise, the filter exhibits multiple lobes, whereas with increasing noise it gradually transitions to a single-lobed shape—mirroring the adaptive temporal filters measured experimentally in sensory neurons (Srinivasan, Laughlin, and Dubs 1982; van Hateren 1992; Liu and Gollisch 2015). Moreover, this adaptation can occur dynamically: when the SNR changes abruptly, the filter adjusts to the new input statistics within roughly ten lag-vector lengths (see Section 3 and Fig. S2 in the Supplementary Material). Thus, the adaptive properties of neuronal temporal filters in our model emerge naturally as a consequence of optimizing predictive information.

## ReSUs Trained on Natural Images Reproduce Physiological Responses

In this Section we use the past-future CCA framework to model neurons by applying it to natural stimuli. As each neuron can only output a scalar as a function of time—a synaptic vesicle release rate or a firing rate—we need to specify how to partition the $r$-dimensional subspace, Eq. 4, among neurons. Gaussian information bottleneck does not specify this as any basis obtained by the orthogonal rotation of the singular vectors should be equally good (Chechik et al. 2003).

We propose that a layer of $r$ neurons compresses the inputs by projecting $\mathbf{p}_t$ onto the temporal filters given by the rows of $\mathbf{\Psi}$ yielding $\mathbf{z}_t$, Eq. 4. As the components of $\mathbf{z}_t$ are whitened, Eq. 5, this has the following advantages. Firstly, it allows straightforward change of the rank in the optimal representation by adding and removing neurons in the order of the corresponding singular values without relearning temporal filters. Secondly, channel whitening is a desirable feature in case of noisy downstream communication and processing (Linsker 1988; Plumbley 1993).

We test this model by comparing its predictions with experimental data from the L1–L3 neurons in the *Drosophila* visual system. These cells receive direct input from photoreceptors that sample the same location in visual space. Assuming that photoreceptor outputs approximate the local contrast of natural scenes, L1–L3 can be viewed as processing a scalar contrast time series for a single pixel, largely independent of neighboring pixels (Rivera-Alba et al. 2014; Borst and Groschner 2023). To model the visual input experienced during ego-motion, we simulate the eye scanning a natural scene at constant velocity by sampling pixel intensities along a straight line of a natural image (Figure 2D). The resulting contrast time series is used to compute neuronal filters via past–future CCA (Eq. 7), which we then compare to the experimentally measured temporal response profiles of L1–L3 (Ketkar et al. 2022).

The first filter is essentially low-pass, while the second filter is a smoothed temporal derivative, predicting a similar output to the activation measured for L3 and L1-L2 respectively (Figure 4, see Supplementary Material for details). To build intuition for our result, consider the so-called "dead leaves model" of natural images. This model assumes that the visual scene is composed of distinct objects, each with a fixed but different contrast from its neighbors. Scanning such a scene along a straight line therefore produces a time series of constant-contrast plateaus separated by transitions. Because in *Drosophila* each photoreceptor samples a relatively large region of visual space ($\approx 5^\circ$), smaller than typical object boundaries, the shape of these contrast transitions reflects the optical properties of the eye and is likely to be stereotyped.

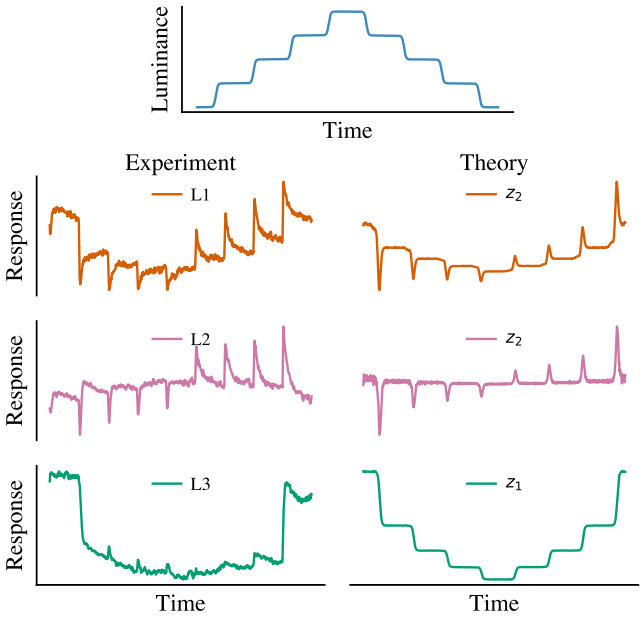


Figure 4: Experimentally measured and theoretically predicted responses to the staircase stimulus. (Top) Luminance as a function of time. (Left column) Experimental measurements of average neuronal activity via calcium imaging in three post-photoreceptor neurons in *Drosophila*: L1, L2 and L3 (Ketkar et al. 2022). (Right column) Output from the first and second temporal filters derived from past-future CCA of natural images. Orange and pink traces differ because of variation in the SNR, i.e., observation noise level: $\sigma = 0.5, 0.05$ respectively.

We can now interpret the roles of L1–L3 neurons in terms of maximizing the mutual information between past and future contrast time series. L3, which outputs a low-pass–filtered version of contrast, captures the strong correlations that persist during contrast plateaus. In contrast, L1 and L2 act as smoothed temporal derivative filters, emphasizing correlations in the rate of change during transitions. Because L1 outputs are rectified by their downstream synapses, they respond selectively to increases in contrast—an ON response. Conversely, L2 outputs are rectified in the opposite direction and respond to decreases in contrast—an OFF response. The plateaus and transitions can thus be viewed as "slow features" that exhibit temporal inertia and are therefore useful for prediction (Wiskott and Sejnowski 2002). These slow features can be further analyzed and combined

downstream, as discussed in the next section.

The linear projection described above would provide an optimal representation under the information bottleneck principle (Arun and Kung 1990; Chechik et al. 2003) if sensory stimuli were generated by linear time-invariant dynamics driven by white Gaussian noise, Eqs. (1),(2). However, this conclusion conflicts with experimental observations showing pronounced output nonlinearities, such as rectification, in most neurons. While such nonlinearities may partly reflect metabolic efficiency (Gjorgjieva, Sompolinsky, and Meister 2014), here we view them as computationally beneficial for extracting predictive latent variables from observations of *nonlinear*—or equivalently, time-varying linear—dynamics. These dynamics can be approximated by switching linear systems, where switching occurs when neural activity crosses a rectification threshold. In this view, each neuron performs a soft clustering of dynamical trajectories and outputs a non-negative membership index (Pehlevan, Genkin, and Chklovskii 2017; Pughe-Sanford et al. 2025). This clustering perspective naturally arises within the Koopman operator framework for nonlinear dynamics which offers a linear representation in lifted feature spaces learnable from data and potentially applicable to hierarchical, deep architectures (Mezić and Wiggins 1999; Williams, Kevrekidis, and Rowley 2015; Kaiser, Kutz, and Brunton 2021; Klus and Conrad 2024; Pughe-Sanford et al. 2025).

Motivated by the above considerations and the non-negativity of neuronal outputs, we propose the Rectified Spectral Unit (ReSU)—a neuron model that projects past input onto a canonical direction and rectifies the projection's positive or negative component:

$$\begin{aligned} \text{ON ReSU}: \quad & z^{+}_{t,i} = \max\left[\mathbf{v}_i^\top \mathbf{C}_{pp}^{-1/2}\mathbf{p}_t,\, 0\right], \\ \text{OFF ReSU}: \quad & z^{-}_{t,i} = \max\left[-\mathbf{v}_i^\top \mathbf{C}_{pp}^{-1/2}\mathbf{p}_t,\, 0\right], \end{aligned} \tag{10}$$

where $\mathbf{v}_i$ is the $i$-th singular vector and $i$-th column of $\mathbf{V}$. The term "Spectral" is used in its linear-algebraic sense and appears in the abbreviation because the projection is derived from the eigendecomposition of the whitened past–future covariance Gramian (6). Following biology, we model L3 using a non-rectified projection (4) with $i = 1$, and L1, L2 using $i = 2$ ON, OFF ReSUs (10), respectively.

Compared to ReLUs, whose synapses are learned by error backpropagation, ReSUs appear more biologically plausible. This is because supervised learning requires labeled ground truth data which is typically not available to neurons and backpropagation relies on nonlocal learning rules for which no biological substrate has been found. At the same time, ReSUs are self-supervised and do not require labeled ground truth. Since synaptic weights in ReSU networks are obtained by past-future CCA of the input, they don't require communication across layers as does backpropagation. Moreover, (Lipshutz et al. 2021; Golkar et al. 2020) have shown that CCA (although not past-future) can be implemented by local learning rules (see Discussion).

## A Two-Layer ReSU Network Trained on Natural Stimuli Detects Motion Similar to *Drosophila*

In this section, we train the second layer of the ReSU network on the outputs of the L1 and L3 analogs from adjacent pixels (Figure 1). We find that, post-training, the ReSU network's neuronal responses and synaptic weights align with the empirical results in *Drosophila* (Figure 5).

To motivate the architecture of the model network (Figure 1), we summarize a few biological facts about *Drosophila* visual processing. ON-motion signals are computed separately from OFF-motion signals by processing the rectified outputs of L1 and the outputs of L3 neurons (Takemura et al. 2013; Borst and Groschner 2023). We simplify the *Drosophila* circuit (Figure 5A) by directly connecting the outputs of L1 and L3 neurons to the T4 neuron (Figure 1) because the intermediate neurons are thought to primarily implement contrast normalization (Matulis et al. 2020). Whereas each L1 and L3 neuron is stimulated by a single "pixel" (Rivera-Alba et al. 2014), the second layer integrates outputs of L1 and L3 corresponding to several adjacent pixels (Takemura et al. 2017). Assuming a one-dimensional retina in the model, we consider three consecutive pixels.

To train the ReSU network, we present the scalar contrast time series, $\{x_t\}$, obtained by scanning the rows of panoramic natural images. By performing CCA on the (centered) past, $\mathbf{p}_t^1$, and future, $\mathbf{f}_t^1$, lag vectors we learn temporal filters and compute the output of L1 and L3 neurons in each pixel, $\mathbf{z}_t : z_1(t) = \Psi_1 \mathbf{p}_t^1, z_2(t) = [\Psi_2 \mathbf{p}_t^1]_+$.

A ReSU in the second layer receives the outputs of L1 and L3 neurons from three adjacent pixels—six channels total—$\mathbf{y}_t = [z_1^L(t), z_2^L(t), z_1^C(t), z_2^C(t), z_1^R(t), z_2^R(t)]^\top \in \mathbb{R}^6$, where the superscripts $L, C, R$ stand for left, center and right pixels, respectively. We learn the second-layer spatial filters by performing CCA on (uncentered) $\mathbf{p}_t^2 = \mathbf{y}_t$ and $\mathbf{f}_t^2 = \mathbf{y}_{t+\text{lag}}$ (Eq. 6) with $\mathbf{y}_t$ elicited by the same contrast time series, $\{x_t\}$, with a motion-induced temporal offset between adjacent pixels. Note that, after rectification in L1, its output is no longer uncorrelated with the output of L3. For details, see Supplementary Material.

We next evaluate the response of the trained two-layer ReSU network to moving grating stimuli (ON/OFF; Figure 5). Filtering the first-layer output with the $i = 1$ ReSU of the second layer produced a motion-direction-independent response, whereas unrectified filtering with the $i = 2$ ReSU yielded a stronger response in the preferred than in the null direction (Figure 5B). The residual null-direction response can be eliminated through appropriate thresholding within the ReSU. In a biological setting, neurons could acquire $i = 2$-like filters if the $i = 1$ projection were suppressed, for example via inhibitory interneurons. The two T4 subtypes (T4a and T4b) can develop opposite direction selectivity when trained on stimuli translating in opposite directions (Figure 5). Given the known gating of synaptic plasticity by postsynaptic activity, we further speculate that when training stimuli contain motion in both directions, such selectivity could emerge through plasticity gated by the rectified output of the corresponding T4 neuron.

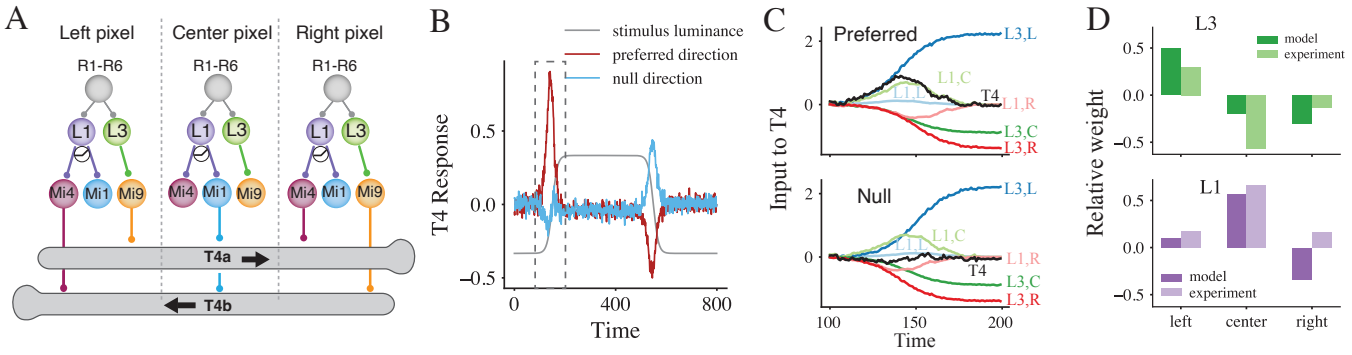


Figure 5: Experimentally observed and learned motion-detection networks exhibit similar response properties and synaptic connectivity. (A) The *Drosophila* ON motion-detection pathway (Takemura et al. 2017; Borst and Groschner 2023). For comparison with the model (Fig. 1), we simplify the circuit by estimating the effective weights of inputs from L1 and L3 to T4 directly, as the intermediate neurons are thought to perform primarily contrast normalization (Matulis et al. 2020). (B) The unrectified response of a T4 analog in the trained ReSU network to a moving grating mirrors experimental results: the strongest, sharply peaked response occurs for an advancing bright ("ON") edge in the preferred direction, whereas weaker responses to "OFF" edges in the null direction can be suppressed by thresholding. (C) Synaptically weighted contributions of each first-layer output channel to the T4 response (black) for ON-edge motion in the preferred and null directions around stimulus onset (dashed box in B). (D) Second-layer synaptic weights in the trained model reproduce the majority of sign patterns and approximate the relative amplitudes of synaptic inputs onto T4a in *Drosophila* (Takemura et al. 2017). Because experimental weights are based on synapse counts without knowledge of neuronal gain factors, we compare L1 and L3 inputs separately.

To better understand the underlying mechanism of motion selectivity, we plot the input to a T4 ReSU, specifically, the L1 and L3 outputs of three adjacent pixels weighted by the spatial filter (Figure 5C). The differential response of T4 to preferred-direction motion is driven by the left L3 input ramping up prior to the drop of the right and center L3 input and, to a lesser degree, center L1 input. After taking into account the simplification of the model architecture compared to the actual connectome, this mechanism aligns with the experiment, Figure 5C,D in (Borst and Groschner 2023).

Finally, we compare the spatial canonical directions of the first-layer outputs with the feedforward synaptic weights onto T4a neurons from connectomics and find qualitative similarity (Figure 5D).

## Discussion

We propose self-supervised ReSU networks as an alternative to the conventional supervised ReLU networks both for modeling biological circuits and for constructing artificial neural networks. A two-layer ReSU network captures salient features of the motion-detection pathway in *Drosophila*. Stacking ReSU layers opens a path to construct deep networks that learn progressively more complex features.

We use 1D motion detection as a well-established, tractable—but nontrivial—testbed to demonstrate that self-supervised learning on natural stimuli can yield bioplausible circuits. Since 2D motion perception builds on a pair of horizontal and vertical 1D detectors, our framework can be readily generated to model 2D motion detection.

While our focus here is on the *Drosophila* visual pathway, ReSU networks can be applied to other sensory modalities and species. Our work was originally motivated by graded-potential neurons whose outputs are rectified by downstream synapses—a property shared by both invertebrate and vertebrate retinas, as well as the *C. elegans* nervous system. Nonetheless, ReSUs can also approximate spiking neurons under a firing-rate representation.

In the current formulation, the temporal memory and prediction horizon are specified *a priori*, rather than learned by the neuron. Future work will aim to develop methods for automatically determining the optimal memory length based on the input statistics.

We demonstrated a self-supervised learning of non-trivial features in a two-layer ReSU network, but whether this approach generalizes to deeper networks remains an open question. We have recently extended this work to a three-layer ReSU network better capturing the fly motion detection pathway (Sharafeldin, Schomburg, and Chklovskii 2026), Figure 5A. While the current two-layer network was trained on images translated in only one direction (left to right), the three-layer ReSU network learns from bidirectionally translated images further establishing a foundation for learning more complex hierarchical features in deeper architectures.

While the theoretically derived neuronal temporal filters qualitatively resemble those observed experimentally using calcium imaging, further validation is needed. Testing the theory would require measurements of neuronal responses and temporal filters at higher temporal resolution, for instance using voltage indicators. Additional promising approaches include examining adaptation to stimulus statistics and probing circuit mechanisms through targeted manipulations such as neuron ablation or silencing.

As early sensory processing is largely feedforward, we expect that ReSU networks trained on natural stimuli offer a principled and interpretable model. Hierarchical fea-

tures learned by self-supervised ReSU networks should support diverse behavioral tasks. As one dives deeper into the brain, the contribution of top-down feedback—and the number of feedback loops—progressively increases. In the future, it would be interesting to explore how to incorporate top-down feedback in our current framework. One potential avenue is to combine the theoretical foundations of the data-driven ReSU framework (Pughe-Sanford et al. 2025) with the data-driven controller neuron model (Moore et al. 2024).

We call our algorithm potentially local because future work will aim to extend the neural network of static units with local learning rules which performs CCA between two concurrent data streams (Lipshutz et al. 2021) to our setting of past-future CCA. We anticipate that incorporating temporal structure will not pose significant challenges as it can be implemented locally within each neuron through the use of distinct ion channels with different time constants. Another difference of our framework from the CCA formulation of (Lipshutz et al. 2021) is that the neuronal output corresponds to a projection of only past inputs onto the canonical direction, with future inputs used exclusively for learning. A similar self-supervised problem has been solved using local learning rules in a static setting (Golkar et al. 2020), and we therefore do not expect it to present major difficulties.

## Acknowledgments

We are grateful to R. Behnia, T. Clandinin, D. Clark, C. Karaneen, I. Nemenman, S.E. Palmer, E. Schneidman, E. Schomburg, D. Schwab, A. Sharafeldin, M. Silies, and N. Srebro for helpful discussions, to R. M. Haret and T. Gollisch for sharing their data. Some of this work was initiated and performed at the Aspen Center for Physics, which is supported by NSF grant PHY-2210452.

## References

Ambikasaran, S.; Foreman-Mackey, D.; Greengard, L.; Hogg, D. W.; and O'Neil, M. 2016. Fast Direct Methods for Gaussian Processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2): 252–265.

Arun, K.; and Kung, S. 1990. Balanced approximation of stochastic systems. *SIAM journal on matrix analysis and applications*, 11(1): 42–68.

Borst, A.; and Groschner, L. N. 2023. How flies see motion. *Annual review of neuroscience*, 46(1): 17–37.

Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J. D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33: 1877–1901.

Chalk, M.; Marre, O.; and Tkačik, G. 2018. Toward a unified theory of efficient, predictive, and sparse coding. *Proceedings of the National Academy of Sciences*, 115(1): 186–191.

Chechik, G.; Globerson, A.; Tishby, N.; and Weiss, Y. 2003. Information bottleneck for Gaussian variables. *Advances in Neural Information Processing Systems*, 16.

Crick, F. 1989. The recent excitement about neural networks. *Nature*, 337(6203): 129–132.

Dora, S.; Bohte, S. M.; and Pennartz, C. M. 2021. Deep gated Hebbian predictive coding accounts for emergence of complex neural response properties along the visual cortical hierarchy. *Frontiers in Computational Neuroscience*, 15: 666131.

Druckmann, S.; Hu, T.; and Chklovskii, D. 2012. A mechanistic model of early sensory processing based on subtracting sparse representations. *Advances in Neural Information Processing Systems*, 25.

Geirhos, R.; Jacobsen, J.-H.; Michaelis, C.; Zemel, R.; Brendel, W.; Bethge, M.; and Wichmann, F. A. 2020. Shortcut learning in deep neural networks. *Nature Machine Intelligence*, 2(11): 665–673.

Gjorgjieva, J.; Sompolinsky, H.; and Meister, M. 2014. Benefits of pathway splitting in sensory coding. *Journal of Neuroscience*, 34(36): 12127–12144.

Golkar, S.; Lipshutz, D.; Bahroun, Y.; Sengupta, A.; and Chklovskii, D. 2020. A simple normative network approximates local non-Hebbian learning in the cortex. *Advances in neural information processing systems*, 33: 7283–7295.

Guerguiev, J.; Lillicrap, T. P.; and Richards, B. A. 2017. Towards deep learning with segregated dendrites. *elife*, 6: e22901.

Halvagal, M. S.; and Zenke, F. 2023. The combination of Hebbian and predictive plasticity learns invariant object representations in deep sensory networks. *Nature neuroscience*, 26(11): 1906–1915.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, 1026–1034.

Hu, T.; Pehlevan, C.; and Chklovskii, D. B. 2014. A hebbian/anti-hebbian network for online sparse dictionary learning derived from symmetric matrix factorization. In *2014 48th Asilomar Conference on Signals, Systems and Computers*, 613–619. IEEE.

Illing, B.; Ventura, J.; Bellec, G.; and Gerstner, W. 2021. Local plasticity rules can learn deep representations using self-supervised contrastive predictions. *Advances in neural information processing systems*, 34: 30365–30379.

Ji, Z.; Lee, N.; Frieske, R.; Yu, T.; Su, D.; Xu, Y.; Ishii, E.; Bang, Y. J.; Madotto, A.; and Fung, P. 2023. Survey of hallucination in natural language generation. *ACM computing surveys*, 55(12): 1–38.

Kaiser, E.; Kutz, J. N.; and Brunton, S. L. 2021. Data-driven discovery of Koopman eigenfunctions for control. *Machine Learning: Science and Technology*, 2(3): 035023.

Katayama, T.; et al. 2005. *Subspace methods for system identification*, volume 1. Springer.

Ketkar, M. D.; Gür, B.; Molina-Obando, S.; Ioannidou, M.; Martelli, C.; and Silies, M. 2022. First-order visual interneurons distribute distinct contrast and luminance information across ON and OFF pathways to achieve stable behavior. *Elife*, 11: e74937.

Keysers, D.; Schärli, N.; Scales, N.; Buisman, H.; Furrer, D.; Kashubin, S.; Momchev, N.; Sinopalnikov, D.; Stafiniak, L.;

Tihon, T.; et al. 2019. Measuring compositional generalization: A comprehensive method on realistic data. *arXiv preprint arXiv:1912.09713*.

Klus, S.; and Conrad, N. D. 2024. Dynamical systems and complex networks: A Koopman operator perspective. *Journal of Physics: Complexity*, 5(4): 041001.

Lappalainen, J. K.; Tschopp, F. D.; Prakhya, S.; McGill, M.; Nern, A.; Shinomiya, K.; Takemura, S.-y.; Gruntman, E.; Macke, J. H.; and Turaga, S. C. 2024. Connectome-constrained networks predict neural activity across the fly visual system. *Nature*, 634(8036): 1132–1140.

Laughlin, S. B. 2001. Energy as a constraint on the coding and processing of sensory information. *Current opinion in neurobiology*, 11(4): 475–480.

LeCun, Y. 2022. A Path Towards Autonomous Machine Intelligence. *OpenReview preprint*. Https://openreview.net/forum?id=BZ5a1r-kVsf.

Lillicrap, T. P.; Cownden, D.; Tweed, D. B.; and Akerman, C. J. 2016. Random synaptic feedback weights support error backpropagation for deep learning. *Nature communications*, 7(1): 13276.

Lillicrap, T. P.; Santoro, A.; Marris, L.; Akerman, C. J.; and Hinton, G. 2020. Backpropagation and the brain. *Nature Reviews Neuroscience*, 21(6): 335–346.

Linsker, R. 1988. Self-organization in a perceptual network. *Computer*, 21(3): 105–117.

Lipshutz, D.; Bahroun, Y.; Golkar, S.; Sengupta, A. M.; and Chklovskii, D. B. 2021. A biologically plausible neural network for multichannel canonical correlation analysis. *Neural Computation*, 33(9): 2309–2352.

Lipshutz, D.; Windolf, C.; Golkar, S.; and Chklovskii, D. 2020. A biologically plausible neural network for slow feature analysis. *Advances in neural information processing systems*, 33: 14986–14996.

Liu, J. K.; and Gollisch, T. 2015. Spike-triggered covariance analysis reveals phenomenological diversity of contrast adaptation in the retina. *PLoS computational biology*, 11(7): e1004425.

Mardt, A.; Pasquali, L.; Wu, H.; and Noé, F. 2018. VAMPnets for deep learning of molecular kinetics. *Nature communications*, 9(1): 5.

Matulis, C. A.; Chen, J.; Gonzalez-Suarez, A. D.; Behnia, R.; and Clark, D. A. 2020. Heterogeneous temporal contrast adaptation in Drosophila direction-selective circuits. *Current Biology*, 30(2): 222–236.

Mezić, I.; and Wiggins, S. 1999. A method for visualization of invariant sets of dynamical systems based on the ergodic partition. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 9(1): 213–218.

Moore, J. J.; Genkin, A.; Tournoy, M.; Pughe-Sanford, J. L.; de Ruyter van Steveninck, R. R.; and Chklovskii, D. B. 2024. The neuron as a direct data-driven controller. *Proceedings of the National Academy of Sciences*, 121(27): e2311893121.

Nøkland, A. 2016. Direct feedback alignment provides learning in deep neural networks. *Advances in neural information processing systems*, 29.

Oja, E. 1982. Simplified neuron model as a principal component analyzer. *Journal of mathematical biology*, 15(3): 267–273.

Olshausen, B. A.; and Field, D. J. 1997. Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision research*, 37(23): 3311–3325.

Patterson, D.; Gonzalez, J.; Le, Q.; Liang, C.; Munguia, L.-M.; Rothchild, D.; So, D.; Texier, M.; and Dean, J. 2021. Carbon emissions and large neural network training. *arXiv preprint arXiv:2104.10350*.

Pehlevan, C.; and Chklovskii, D. B. 2019. Neuroscience-inspired online unsupervised learning algorithms: Artificial neural networks. *IEEE Signal Processing Magazine*, 36(6): 88–96.

Pehlevan, C.; Genkin, A.; and Chklovskii, D. B. 2017. A clustering neural network model of insect olfaction. In *2017 51st Asilomar Conference on Signals, Systems, and Computers*, 593–600. IEEE, IEEE.

Plumbley, M. D. 1993. A Hebbian/anti-Hebbian network which optimizes information capacity by orthonormalizing the principal subspace. In *1993 Third International Conference on Artificial Neural Networks*, 86–90. IET.

Pughe-Sanford, J. L.; Ding, X.; Moore, J. J.; Sengupta, A. M.; Epstein, C.; Greengard, P.; and Chklovskii, D. B. 2025. Neurons as Detectors of Coherent Sets in Sensory Dynamics. *arXiv preprint arXiv:2510.26955*.

Rao, R. P. 2024. A sensory–motor theory of the neocortex. *Nature neuroscience*, 27(7): 1221–1235.

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

Recht, B.; Roelofs, R.; Schmidt, L.; and Shankar, V. 2019. Do imagenet classifiers generalize to imagenet? In *International conference on machine learning*, 5389–5400. PMLR.

Rivera-Alba, M.; Peng, H.; de Polavieja, G. G.; and Chklovskii, D. B. 2014. Wiring economy can account for cell body placement across species and brain areas. *Current Biology*, 24(3): R109–R110.

Ruderman, D. L.; and Bialek, W. 1994. Statistics of natural images: Scaling in the woods. *Physical review letters*, 73(6): 814.

Rumelhart, D. E.; Hinton, G. E.; and Williams, R. J. 1986. Learning representations by back-propagating errors. *nature*, 323(6088): 533–536.

Scellier, B.; and Bengio, Y. 2017. Equilibrium propagation: Bridging the gap between energy-based models and backpropagation. *Frontiers in computational neuroscience*, 11: 24.

Sharafeldin, A.; Schomburg, E.; and Chklovskii, D. B. 2026. Self-supervised learning of motion vision from natural stimuli. In *review*.

Silver, D.; Huang, A.; Maddison, C. J.; Guez, A.; Sifre, L.; Van Den Driessche, G.; Schrittwieser, J.; Antonoglou, I.; Panneershelvam, V.; Lanctot, M.; et al. 2016. Mastering the game of Go with deep neural networks and tree search. *nature*, 529(7587): 484–489.

Srinivasan, M. V.; Laughlin, S. B.; and Dubs, A. 1982. Predictive coding: a fresh view of inhibition in the retina. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 216(1205): 427–459.

Strubell, E.; Ganesh, A.; and McCallum, A. 2020. Energy and policy considerations for modern deep learning research. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, 13693–13696.

Takemura, S.-y.; Bharioke, A.; Lu, Z.; Nern, A.; Vitaladevuni, S.; Rivlin, P. K.; Katz, W. T.; Olbris, D. J.; Plaza, S. M.; Winston, P.; et al. 2013. A visual motion detection circuit suggested by Drosophila connectomics. *Nature*, 500(7461): 175–181.

Takemura, S.-y.; Nern, A.; Chklovskii, D. B.; Scheffer, L. K.; Rubin, G. M.; and Meinertzhagen, I. A. 2017. The comprehensive connectome of a neural substrate for 'ON'motion detection in Drosophila. *Elife*, 6: e24394.

van Hateren, J. H. 1992. Theoretical predictions of spatiotemporal receptive fields of fly LMCs, and experimental validation. *Journal of Comparative Physiology A*, 171(2): 157–170.

Whittington, J. C.; and Bogacz, R. 2017. An approximation of the error backpropagation algorithm in a predictive coding network with local hebbian synaptic plasticity. *Neural computation*, 29(5): 1229–1262.

Williams, M. O.; Kevrekidis, I. G.; and Rowley, C. W. 2015. A data–driven approximation of the koopman operator: Extending dynamic mode decomposition. *Journal of Nonlinear Science*, 25(6): 1307–1346.

Wiskott, L.; and Sejnowski, T. J. 2002. Slow feature analysis: Unsupervised learning of invariances. *Neural computation*, 14(4): 715–770.

Yamins, D. L.; Hong, H.; Cadieu, C. F.; Solomon, E. A.; Seibert, D.; and DiCarlo, J. J. 2014. Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the national academy of sciences*, 111(23): 8619–8624.