

# Neural Circuit Dynamics for Sensory Detection

Sruti Mallik,<sup>1</sup> Srinath Nizampatnam,<sup>1</sup> Anirban Nandi,<sup>2</sup> Debajit Saha,<sup>3</sup>  Baranidharan Raman,<sup>4</sup> and  ShiNung Ching<sup>1,4</sup>

<sup>1</sup>Department of Electrical and Systems Engineering, Washington University in St. Louis, St. Louis, Missouri 63130, <sup>2</sup>Allen Institute of Brain Science, Seattle, Washington 98109, <sup>3</sup>Department of Biomedical Engineering, Michigan State University, East Lansing, Michigan 48824, and <sup>4</sup>Department of Biomedical Engineering, Washington University in St. Louis, St. Louis, Missouri 63130

We consider the question of how sensory networks enable the detection of sensory stimuli in a combinatorial coding space. We are specifically interested in the olfactory system, wherein recent experimental studies have reported the existence of rich, enigmatic response patterns associated with stimulus onset and offset. This study aims to identify the functional relevance of such response patterns (i.e., what benefits does such neural activity provide in the context of detecting stimuli in a natural environment). We study this problem through the lens of normative, optimization-based modeling. Here, we define the notion of a low-dimensional latent representation of stimulus identity, which is generated through action of the sensory network. The objective of our optimization framework is to ensure high-fidelity tracking of a nominal representation in this latent space in an energy-efficient manner. It turns out that the optimal motifs emerging from this framework possess morphologic similarity with prototypical onset and offset responses observed *in vivo* in locusts (*Schistocerca americana*) of either sex. Furthermore, this objective can be exactly achieved by a network with reciprocal excitatory–inhibitory competitive dynamics, similar to interactions between projection neurons and local neurons in the early olfactory system of insects. The derived model also makes several predictions regarding maintenance of robust latent representations in the presence of confounding background information and trade-offs between the energy of sensory activity and resultant behavioral measures such as speed and accuracy of stimulus detection.

## Significance Statement

A key area of study in olfactory coding involves understanding the transformation from high-dimensional sensory stimulus to low-dimensional decoded representation. Here, we examine not only the dimensionality reduction of this mapping but also its temporal dynamics, with specific focus on stimuli that are temporally continuous. Through optimization-based synthesis, we examine how sensory networks can track representations without prior assumption of discrete trial structure. We show that such tracking can be achieved by canonical network architectures and dynamics, and that the resulting responses resemble observations from neurons in the insect olfactory system. Thus, our results provide hypotheses regarding the functional role of olfactory circuit activity at both single neuronal and population scales.

## Introduction

We consider the question of how early sensory networks produce actionable neural representation in response to sensory stimuli arriving in a dynamic fashion. The specific focus here is on the olfactory system, wherein the architecture is schematically

conserved across species (Strausfeld and Hildebrand, 1999). In this system, early networks receive external excitation from periphery and transform it into intermediate representations which are then routed to higher-level brain areas for further processing and behavioral and motor response generation (Kay and Stopfer, 2006; Martin et al., 2011; Aldworth and Stopfer, 2015). Our goal is to identify the functional significance of certain characteristic neural response patterns observed as animals encounter sensory stimuli.

We are specifically motivated by experimental findings illustrating a rich taxonomy of stimulus-evoked, time-varying responses or sensory trajectories associated with both stimulus onset and offset (Stopfer et al., 2003; Kay and Stopfer, 2006; Bathellier et al., 2008; Saha et al., 2013b, 2017). It has been reported that the most noticeable change in activity of sensory neurons occurs immediately following stimulus onset and then

Received Sep. 10, 2019; revised Jan. 19, 2020; accepted Feb. 28, 2020.

Author contributions: S.M., B.R., and S.C. designed research; S.M., S.N., A.N., D.S., and B.R. performed research; S.M., S.N., D.S., B.R., and S.C. analyzed data; S.M. and S.C. wrote the first draft of the paper; S.M., B.R., and S.C. edited the paper; S.M. and S.C. wrote the paper.

The authors declare no competing financial interests.

This work has been supported by Grants EF-1724218 (S.C., B.R.), IIS-1453022 (B.R.), and CMMI-1653589 (S.C.) from the National Science Foundation. S.C. holds a Career Award at the Scientific Interface from the Burroughs-Wellcome Fund.

Correspondence should be addressed to ShiNung Ching at shinung@wustl.edu.

<https://doi.org/10.1523/JNEUROSCI.2185-19.2020>

Copyright © 2020 the authors

following stimulus offset (Stopfer et al., 2003; Mazor and Laurent, 2005; Bathellier et al., 2008; Raman et al., 2010), and these responses are in fact orthogonal to each other at a population level (Saha et al., 2017). In this work, we are not attempting to “pattern match” these responses per se, but rather provide a functional interpretation of what these prototypical spatiotemporal response motifs achieve in terms of their ability to mediate stimulus detection. Through both theory and validation, we address the following: (1) whether there are particular decoding objectives that explain the manifestation of observed sensory response motifs; and (2) whether there are physiologically plausible neural circuit architectures that are capable of realizing motifs generated by optimization of the objective in (1).

The central theoretical premise of our paper is that activity of the early sensory network drives a latent representation of accrued evidence regarding presence or absence of stimulus. The neural responses thus encode not only detection but also “undetection” or withdrawal of the source of excitation. Adapting from approaches used in behavioral neuroscience, we formulate a dynamical decoder that uses high-dimensional, combinatorial input to generate a low-dimensional latent representation. Thereafter, we use a top-down, normative approach wherein we generate the response motifs by minimizing an objective function. We formulate this objective function so that it emphasizes that latent representations should quickly and accurately convey information about peripheral stimulus in an energy-efficient manner. Neural response motifs borne out of such an optimization problem bear morphologic similarity to those observed *in vivo*, leading us to assert our claim of ascribed functional relevance.

Our work builds on a rich theory of olfactory coding focused on the question of how reliable and informative representations propagate through the sensory hierarchy (Bhandawat et al., 2007; Schaefer and Margrie, 2007; Zhang and Sharpee, 2016). In particular, we note recent efforts to ascribe particular functions to the circuit architectures, including enabling stimulus reconstruction (Qin et al., 2019), categorization (Dasgupta et al., 2017), and novelty detection (Dasgupta et al., 2018). However, these findings have mostly been pursued in a static input–output domain (i.e., with tacit assumption of instantaneous, algebraic signal transformation). In contrast, we work in the space of sensory dynamics and time-varying representations, a topic that has received much attention from a descriptive standpoint (Laurent, 1996; Rabinovich et al., 2000; Laurent et al., 2001) but less so from the perspective of normative synthesis. In particular, through formal mathematical arguments, we investigate what advantages the big switch between onset and offset responses achieves in the biological world, where organisms encounter stimuli in a dynamic fashion.

Our results show that our normative model produces emergent phenomena that predict many nuanced features of actual sensory network activity as observed in locust (Saha et al., 2017; Mazor and Laurent, 2005; Nizampatnam et al., 2018). Further, our normative synthesis procedure yields a set of network dynamics that is highly compatible with the known physiology of these circuit *in vivo*. Hereafter, we proceed to formulate our theoretical setup before presenting our key synthesis and validation results.

## Materials and Methods

### Computational model

#### Decoder and latent space

Our problem setup is premised on the canonical architecture of the insect early olfactory system (Kay and Stopfer, 2006; Masse et al., 2009),

wherein chemical cues are transduced to neural signals through olfactory receptor neurons (ORNs), which then propagate activity via glomeruli to projection neurons (PNs) en route to higher brain areas. Our focus is on the dynamical transformations mediated by PNs and their local circuitry.

Our model for sensory tracking hinges on the definition of a latent space,  $\mathbf{v}(t) \equiv [\nu_1(t), \nu_2(t), \dots, \nu_m(t)]$ , which contains information about stimulus identity. Each dimension of the latent space indicates accrued evidence regarding the presence of high-level stimulus features (Raman et al., 2011). The assumption is that such features provide an actionable representation of the stimulus that can then be used to enable higher-level processing and behavior. Specifically, PN activity,  $\mathbf{x} \in \mathbb{R}^n$ , is linearly decoded into the latent space via (1):

$$\dot{\mathbf{v}} = A\mathbf{v} + \mathbf{b}\mathbf{x}, \quad (1)$$

where  $A \in \mathbb{R}^{m \times m}$  such that  $\sigma(A) \in \mathbb{R}^-$ , and  $\sigma(A)$  represents the spectrum of  $A$ . Here,  $A = -a\mathbf{I}$ , where  $a > 0$  and  $\mathbf{I} \in \mathbb{R}^{m \times m}$  is the identity matrix. The matrix  $\mathbf{b} \in \mathbb{R}^{m \times n}$  linearly mixes the contribution of PNs onto each dimension of the latent vector. Equation (1) imposes that the change in accrued evidence at any time  $t$  is proportional to the difference between the information encoded by representative PNs and evidence lost due to intrinsic leaky dynamics of the decoder. Figure 1A illustrates our problem setup, noting that we are focused on activity downstream of the ORNs (i.e., the dashed vertical line). Afferent activity from the receptor neurons is denoted by  $\mathbf{r}(t)$ . It is important to note that our approach does not assume any network structure nor dynamics for the PNs; rather, our goal is to synthesize these dynamics.

### Enabling labile, accurate latent representations via sensory tracking

In particular, the formulation in Equation 1 allows us to introduce our normative premise (i.e., that the PN activity  $\mathbf{x}$  drives accurate representations of accrued evidence, thus conveying information about stimulus presence and identity). Further, the dynamics of  $\mathbf{x}$  should allow for fast transitions in  $\mathbf{v}$ .

This idea is readily captured through the notion of tracking in the latent space. From the initial state  $\mathbf{v} = 0$ , we seek to construct dynamics for the PN  $\mathbf{x}$  such that  $\mathbf{v}$  can quickly and accurately reach an arbitrary point in the latent space,  $\mathbf{z}$ . Mathematically, we formulate a quadratic objective function:

$$\mathbb{J}(\mathbf{x}) = \int_0^\infty \frac{1}{2} [(\mathbf{v}(t)^T - \mathbf{z})\mathbf{Q}(\mathbf{v}(t) - \mathbf{z}) + \mathbf{x}(t)^T\mathbf{S}\mathbf{x}(t) + \dot{\mathbf{x}}(t)^T\mathbf{R}\dot{\mathbf{x}}(t)] dt. \quad (2)$$

The terms in this objective function are highly interpretable from a biological perspective. The first term in the integrand penalizes the deviation of the decoded state from  $\mathbf{z}$ . The second term penalizes the energy of the PN response, whereas the last term penalizes large fluctuations in the PN response. The objective is formulated over an infinite horizon (i.e., there is no bias or expectation regarding the amount of time that the representation is to be tracked).

The objective and normative optimization is schematized in Figure 1B,C. At the stimulus onset (Fig. 1B), the latent state is to track some nominal representation  $\mathbf{z}$ . On termination of the stimulus, the latent state continues tracking  $\mathbf{z}$ , which is now positioned at the origin of the  $m$ -dimensional latent space signifying absence of a stimulus. Here, we considered a pulsatile stimulus structure in accordance with the natural olfactory environment, where an organism perceives olfactory stimulation as pockets of odor plumes (Vickers, 2000). Furthermore, in this work, our focus was on the neural dynamics downstream of the receptor neurons (Fig. 1A); therefore, we assumed that each stimulus is presented at a fixed intensity. The matrix  $\mathbf{Q}$  in our setup enforces that the tracking in the latent space occurs accurately, without transient or asymptotic misrepresentation (see Fig. 1B). The matrix  $\mathbf{S}$

imposes penalty on the energy expenditure, whereas the matrix  $\mathbf{R}$  regularizes rapid fluctuation in firing rate activity. Upon withdrawal of the stimulus (Fig. 1C), the representation returns to neutral. In essence, this model embeds the idea of active stimulus detection and undetection.” The question at hand is how  $\mathbf{x}$  should be specified (according to Equation 2) for robust and accurate detection of odor stimuli.

Table 1 summarizes all mathematical symbols used in this paper as well as their interpretation.

### Reduction of optimization framework to infinite horizon linear quadratic regulator problem

The objective function (as in Eq. 2) subject to the decoder dynamics (Eq. 1) constitutes our optimization problem.

We denote  $\dot{\mathbf{x}} = \mathbf{y}$ . Thus, the problem can be rewritten as follows:

$$\begin{aligned} \min_{\mathbf{y}} \int_0^{\infty} \frac{1}{2} [(\mathbf{v}(t)^T - \mathbf{z})\mathbf{Q}(\mathbf{v}(t) - \mathbf{z}) \\ + \mathbf{x}(t)^T \mathbf{S} \mathbf{x}(t) + \mathbf{y}(t)^T \mathbf{R} \mathbf{y}(t)] dt \\ \text{subject to, } \dot{\mathbf{v}} = \mathbf{A} \mathbf{v} + \mathbf{b} \mathbf{x} \text{ and } \dot{\mathbf{x}} = \mathbf{y}. \end{aligned} \quad (3)$$

Now, we define  $\mathbf{v}_z = [\mathbf{v}^T, \mathbf{x}^T, \mathbf{z}^T]^T$ . We assume that the fixed representation  $\mathbf{z}$  remains constant as long as the same stimulus conditions prevail at the periphery (i.e.,  $\dot{\mathbf{z}} = 0$ ). Therefore, we can write the following:

$$\begin{aligned} \dot{\mathbf{v}}_z = \begin{bmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{0} & \mathbf{0} \\ & \mathbf{0} \end{bmatrix} \mathbf{v}_z + \begin{bmatrix} \mathbf{0} \\ \mathbf{I}_n \\ \mathbf{0} \end{bmatrix} \mathbf{y} = \bar{\mathbf{A}} \mathbf{v}_z \\ + \bar{\mathbf{B}} \mathbf{y} \equiv \mathbf{g}(\mathbf{v}_z, \mathbf{y}) \end{aligned} \quad (4)$$

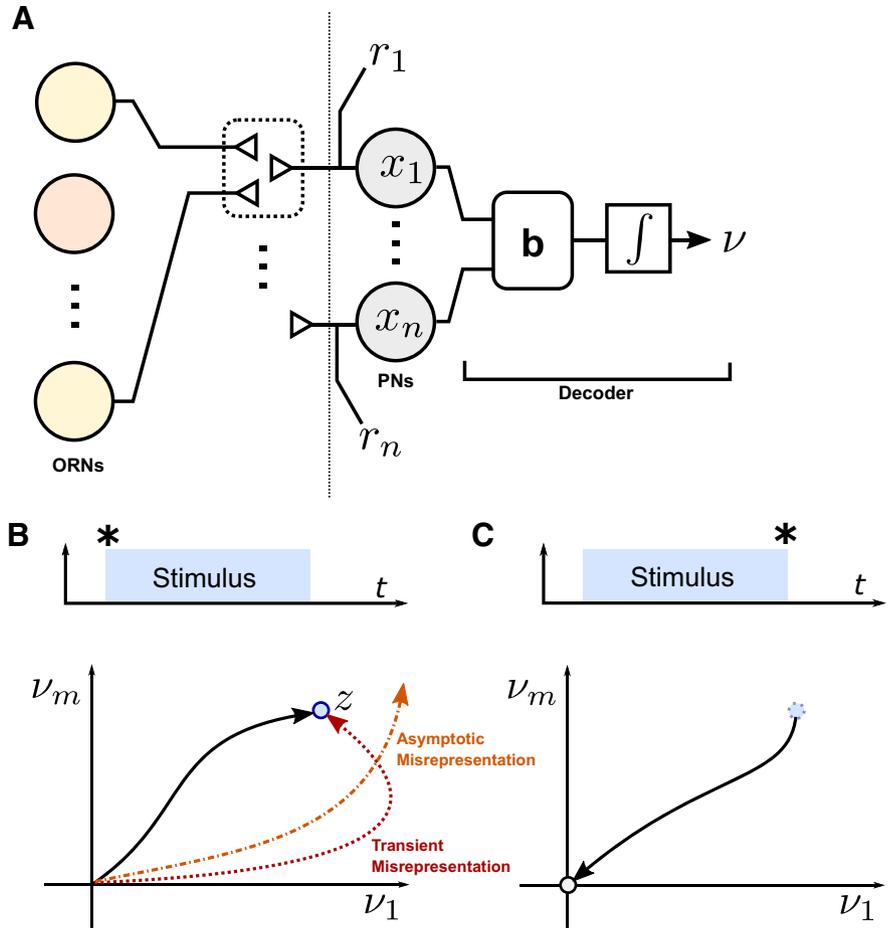
With this, Equation 3 reduces to:

$$\begin{aligned} \min_{\mathbf{y}} \int_0^{\infty} \frac{1}{2} [\mathbf{v}_z(t)^T \bar{\mathbf{Q}}(t) \mathbf{v}_z(t) + \mathbf{y}(t)^T \bar{\mathbf{R}}(t) \mathbf{y}(t)] dt \\ \text{subject to, } \dot{\mathbf{v}}_z = \mathbf{g}(\mathbf{v}_z, \mathbf{y}), \end{aligned} \quad (5)$$

where  $\bar{\mathbf{Q}} = \begin{bmatrix} \mathbf{Q} & \mathbf{0} & -\mathbf{Q} \\ \mathbf{0} & \mathbf{S} & \mathbf{0} \\ -\mathbf{Q} & \mathbf{0} & \mathbf{Q} \end{bmatrix}$  and  $\bar{\mathbf{R}} = \mathbf{R}$ . The transformed optimization problem (OP) (Eq. 5) amounts to an infinite horizon linear quadratic control problem (Boyd and Barratt, 1991; Anderson and Moore, 2007), a classical control theoretic framework for finding the optimal control input to a dynamical system. In our context, the “controls” are the activity of the PNs and the task is to realize these controls by means of a dynamical network.

### Solving the optimization problem

As the matrices  $\mathbf{Q}$ ,  $\mathbf{S}$ , and  $\mathbf{R}$  in Equation 3 are positive definite, it is straightforward to establish that  $\bar{\mathbf{Q}} \geq 0$ ,  $\bar{\mathbf{R}} > 0$ , and hence, problem Equation 5 has a unique solution (Anderson and Moore, 2007). Indeed, the solution to Equation 5 with initial conditions  $\mathbf{v}_z(0) = [\mathbf{v}_0^T, \mathbf{x}_0^T, 0^T]^T$  is given by



**Figure 1.** *A*, Schematic of signal flow through the early olfactory system. We consider response dynamics for the PNs subject to input  $\mathbf{r}$ , which abstracts afferent signals from ORNs. *B*, We posit a decoding objective. At stimulus onset (marked by asterisk), the goal of PNs is to accurately drive a latent representation to a nominal target. *C*, At stimulus offset (marked by asterisk), the latent representation should return to neutral.

$$\mathbf{y}(t) = -\bar{\mathbf{R}}^{-1} \bar{\mathbf{B}}^T \mathbf{K} \mathbf{v}_z = \mathbf{W} \mathbf{v}_z, \quad (6)$$

where  $\mathbf{K}$  is the solution of the algebraic Riccati equation (ARE), given as follows:

$$\bar{\mathbf{A}}^T \mathbf{K} + \mathbf{K} \bar{\mathbf{A}} - \mathbf{K} \bar{\mathbf{B}} \bar{\mathbf{R}}^{-1} \bar{\mathbf{B}}^T \mathbf{K} + \bar{\mathbf{Q}} = 0 \quad (7)$$

Therefore, we can now write

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{y}(t) \\ &= \mathbf{W}_v \mathbf{v}(t) + \mathbf{W}_f \mathbf{x}(t) + \mathbf{W}_z \mathbf{z}(t), \end{aligned} \quad (8)$$

where  $\mathbf{W} = [\mathbf{W}_v : \mathbf{W}_f : \mathbf{W}_z]$  and  $\mathbf{W}_v \in \mathbb{R}^{n \times m}$ ,  $\mathbf{W}_f \in \mathbb{R}^{n \times n}$  and  $\mathbf{W}_z \in \mathbb{R}^{n \times m}$ .

The excitation to the PN layer arrives from first-level ORNs. We denote the afferent input from these receptor neurons as  $\mathbf{r}(t)$ . In the model,

$$\mathbf{r}(t) \equiv \mathbf{W}_z \mathbf{z}(t). \quad (9)$$

It is important to note that the model does not predict ORN dynamics or connectivity per se, only that  $\mathbf{r}(t)$  is provided to the network the PNs. There may be multiple ways to synthesize an ORN network that realizes this transformation, and specifying such a construction is not the primary focus of this paper.

The instantaneous decoded evidence  $\mathbf{v}(t)$  can be written as follows (recall Eq. 1):

**Table 1. Table of math symbols and their interpretation**

Mathematical symbol	Meaning in the model	Significance in the olfactory system
$\mathbf{r}(t)$	Afferent input to PNs	Activity of ORNs
$\mathbf{x}(t), \mathbf{x}_o(t)$	Activity of neural units driving the latent decoder	Firing rate of PNs with baseline activity set to zero.
$\mathbf{x}_i(t)$	Activity of auxiliary population of neurons	Firing rate of LNs with baseline activity set to zero
$\mathbf{v}(t)$	Decoded latent space activity: each dimension indicates accrued evidence regarding presence of high-level stimulus features	Intermediate representation of stimulus information that drives downstream behavioral/motor responses
$A, \mathbf{b}$	Dynamics of the latent space decoder. $A$ represents the internal dynamics of the decoder. Weights of matrix $\mathbf{b}$ represent combinatorial encoding	
$\mathbf{z}$	A fixed representation associated with a particular stimulus	
$\mathbf{Q}$	Penalty on accuracy of latent representation	
$\mathbf{S}$	Penalty on neural resources (i.e., energy used)	
$\mathbf{R}$	Regularization on rapid fluctuation in firing rate activity	
$\mathbf{W}_s$	Connections mapping slow processing of $\mathbf{x}(t)$ onto its dynamics	
$\mathbf{W}_f$	Connections mapping fast processing of $\mathbf{x}(t)$ onto its dynamics	
$\mathbf{W}_o^i$	Excitatory connections from the representative PNs to the auxiliary population	Excitatory synaptic connections from PNs to LNs
$\mathbf{W}_i^o$	Inhibitory connections from the auxiliary population to the putative PNs	Inhibitory synaptic connections from LNs to PNs
$\mathbf{W}_i^i, \mathbf{W}_o^o$	Intrapopulation interactions	Synaptic connections among PNs and LNs
$\widehat{\mathbf{W}}, \widehat{\mathbf{W}}$	Computational parameters (matrix) that must satisfy the following: $\widehat{\mathbf{W}}\widehat{\mathbf{W}} = \mathbf{I}$ , $\text{vec}(\widehat{\mathbf{W}}) \geq 0$ and $\text{vec}(\widehat{\mathbf{W}}, \widehat{\mathbf{W}}) \leq 0$	
$C$	Fraction of neurons responding to both red/blue stimuli	Percentage of projection neurons that respond to more than one sensory stimulus
$\boldsymbol{\eta}(t)$	Unit variance Gaussian noise for the network model developed through Stochastic Linear Quadratic Regulator formulation	Background noise in neural response
$t_s$	Duration of stimulus period	Duration of stimulus period

$$\mathbf{v}(t) = e^{-at}\mathbf{v}(0) + \int_0^t e^{-a(t-\tau)}\mathbf{b}\mathbf{x}(\tau)d\tau. \quad (10)$$

We initialize the decoder to a zero initial state (i.e.,  $\mathbf{v}(0) = \mathbf{0}$ ), leading to

$$\mathbf{v}(t) = \int_0^t e^{-a(t-\tau)}\mathbf{b}\mathbf{x}(\tau)d\tau. \quad (11)$$

Using Equations 11 and 9 in Equation 8, we obtain the optimal neural dynamics:

$$\dot{\mathbf{x}}(t) = \mathbf{W}_v b \int_0^t e^{-a(t-\tau)}\mathbf{x}(\tau)d\tau + \mathbf{W}_f \mathbf{x}(t) + \mathbf{r}(t). \quad (12)$$

When the latent decoder is a noisy one or when the neural entities are modeled as those associated with background noise, we can posit a similar problem. In such a model, the objective function is written as follows:

$$\mathbb{J}(\mathbf{x}) = \mathbb{E} \left[ \int_0^\infty \frac{1}{2} [(\mathbf{v}(t) - \mathbf{z})^T \mathbf{Q} (\mathbf{v}(t) - \mathbf{z}) + \mathbf{x}(t)^T \mathbf{S} \mathbf{x}(t) + \dot{\mathbf{x}}(t)^T \mathbf{R} \dot{\mathbf{x}}(t)] dt \right]. \quad (13)$$

The reduction technique for this stochastic problem is similar to what has been shown above (Boyd and Barratt, 1991) for the deterministic case, and the solution of the ARE for this stochastic linear quadratic regulator is identical to what we have derived above except for the addition of a noise term  $\boldsymbol{\eta}(t)$  (see Eq. 14 below):

$$\dot{\mathbf{x}}(t) = \mathbf{W}_v b \int_0^t e^{-a(t-\tau)}\mathbf{x}(\tau)d\tau + \mathbf{W}_f \mathbf{x}(t) + \mathbf{r}(t) + \sigma_x^2 \boldsymbol{\eta}(t). \quad (14)$$

We have used this stochastic version to generate figures depicting Principal Component (PC) trajectories and correlation of ensemble response.

### Extracting a first-order network to realize the optimal solution

The primary step in extracting a first-order network from the derived optimal motif is to introduce a population of  $n_i$  auxiliary neurons whose activity  $\mathbf{x}_i \in \mathbb{R}^{n_i}$  is given by the following dynamics:

$$\dot{\mathbf{x}}_i = -a\mathbf{x}_i + \widehat{\mathbf{W}}\mathbf{x}_o, \quad (15)$$

where  $\mathbf{x}_o \equiv \mathbf{x}$  is the PN activity,  $n_o$  is the number of PNs, and  $\widehat{\mathbf{W}} \in \mathbb{R}^{n_i \times n_o}$ .

Thus, the function of the matrix  $\widehat{\mathbf{W}}$  is to compute a weighted sum of the PN activity and propagate it to the auxiliary neurons. Using Equations 12 and 15, we can write the following:

$$\dot{\mathbf{x}}_o(t) = \mathbf{W}_s \widehat{\mathbf{W}} \widehat{\mathbf{W}} \int_0^t e^{-a(t-\tau)}\mathbf{x}_o(\tau)d\tau + \mathbf{W}_f \mathbf{x}_o(t) + \mathbf{r}(t). \quad (16)$$

Here,  $\mathbf{W}_s \equiv \mathbf{W}_v b$ ,  $\widehat{\mathbf{W}} \in \mathbb{R}^{n_o \times n_i}$  and  $\widehat{\mathbf{W}}, \widehat{\mathbf{W}}$  must satisfy the condition  $\widehat{\mathbf{W}}\widehat{\mathbf{W}} = \mathbf{I}$  to realize the deduced optimal dynamics. We choose  $n_i < n_o$  (Laurent, 1996). Therefore,  $\widehat{\mathbf{W}}$  is a wide matrix, and the matrix decomposition problem  $\widehat{\mathbf{W}}\widehat{\mathbf{W}} = \mathbf{I}$  is ill posed, which might have no solution or an infinite number of solutions. Instead of solving this underdetermined set of equations for an arbitrary choice of  $\widehat{\mathbf{W}}$ , we reformulate the problem in a tractable format. We introduced a constrained iterative optimization framework (Shlizerman et al., 2014) that minimizes  $\|\widehat{\mathbf{W}}\widehat{\mathbf{W}} - \mathbf{I}\|_2$  (see Algorithm 1) to derive the matrices  $\widehat{\mathbf{W}}, \widehat{\mathbf{W}}$ . The constraints in question are motivated by reports of existence of distinct excitatory/inhibitory neural units in the insect antennal lobe (AL; Wilson and Laurent, 2005; Kay and Stopfer, 2006; Shang et al., 2007). Together, Equations (15) and (16) provide the first-order rate equations of the sensory network (See Figure 2 for schematic of the network).

### Parameter selection

We chose the decoder parameter as  $a = 0.25$  in our simulations (exceptions mentioned in relevant Figure caption). The penalty matrices are chosen as  $\mathbf{Q} = 10\mathbf{I}_m$ ,  $\mathbf{S} = 2\mathbf{I}_n$ ,  $\mathbf{R} = 0.2\mathbf{I}_n$ . Also, unless mentioned

otherwise, we have chosen  $n = 41$  (number of neural units driving the decoder) and  $m = 2$  (dimension of the latent space). For simulations using the stochastic model, we have considered  $\boldsymbol{\eta}(t)$  to be unit variance Gaussian noise with  $\sigma_x = 0.05$ .

The secondary iterative optimization that generates  $\widehat{\mathbf{W}}, \widehat{\mathbf{W}}$  can be outlined as follows.

#### Algorithm 1

```

choose  $\widehat{\mathbf{W}}_{init}$  s.t.  $vec(\widehat{\mathbf{W}}_{init}) \geq 0$ ;
set  $score = \|\widehat{\mathbf{W}}_{init}^\dagger \widehat{\mathbf{W}}_{init} - \mathbf{I}\|_2$ ; set  $flag = 0$ 
 $\widehat{\mathbf{W}}^* = \widehat{\mathbf{W}}_{init}$ 
While  $iter \leq \maxIters$  do
   $\widehat{\mathbf{W}}^* = \underset{\widehat{\mathbf{W}}}{\operatorname{argmin}} \|\widehat{\mathbf{W}}\widehat{\mathbf{W}}^* - \mathbf{I}\|_2 + \lambda_1 \|\widehat{\mathbf{W}}\|_2$ , subject to  $vec(\mathbf{W}_s \widehat{\mathbf{W}}) \leq 0$ 
   $\widehat{\mathbf{W}}^* = \underset{\widehat{\mathbf{W}}}{\operatorname{argmin}} \|\widehat{\mathbf{W}}^* \widehat{\mathbf{W}} - \mathbf{I}\|_2 + \lambda_2 \|\widehat{\mathbf{W}}\|_2$ , subject to  $vec(\widehat{\mathbf{W}}) \geq 0$ 
  If  $\|\widehat{\mathbf{W}}^* \widehat{\mathbf{W}}^* - \mathbf{I}\|_2 < score$  then
    set  $flag = 1$ ;
  else
    continue;
Result: If  $flag == 1$  then
  return  $\widehat{\mathbf{W}}^*, \widehat{\mathbf{W}}^*$ 
Else
  If  $vec(\mathbf{W}_s \widehat{\mathbf{W}}_{init}^\dagger) \leq 0$  then
    return  $\widehat{\mathbf{W}}_{init}^*, \widehat{\mathbf{W}}_{init}^\dagger$ 
  Else
    No solution for this initialization.

```

Here,  $\|\cdot\|_2$  is the Frobenius matrix norm, and regularization coefficients  $\lambda_1, \lambda_2$  are chosen such that  $0 < \lambda_1 < 1, 0 < \lambda_2 < 1$ . We initialize the algorithm with an arbitrary wiring for  $\widehat{\mathbf{W}}$  such that its pseudoinverse  $\widehat{\mathbf{W}}^\dagger$  exists, and we initialize the matrix  $\widehat{\mathbf{W}}$  as  $\widehat{\mathbf{W}}^\dagger$ , as the pseudoinverse is the best fit for a least-squares solution. Thereafter, we proceed using Algorithm 1 to find a pair of matrices  $\widehat{\mathbf{W}}$  and  $\widehat{\mathbf{W}}$  such that the resultant network realizes the optimal motifs as well as the well characterized network constraints are satisfied. The optimization problem is solved using the disciplined convex optimization package CVX implemented in MATLAB (MathWorks; Grant et al., 2008).

#### Dimensionality reduction analysis for simulated PN response

Similar to methods by Saha et al. (2017), we applied principal component analysis (PCA) on our simulated PN response data to visualize the  $n$ -dimensional time-varying activity of PNs. Each exposure to a stimulus of given identity continues for  $\tau$  seconds, followed by a period of no excitation of equal duration. As we synthesized PN response at 10 ms intervals, we generated a time-series data matrix of dimensions  $n \times \Gamma$  (where  $\Gamma = \frac{2\tau}{0.01}$ ) for each stimulus encounter. We concatenated such data matrices obtained for different odorants to compare and contrast the trajectories evoked. The resulting matrix was used to generate an  $n \times n$  response covariance matrix.

To visualize the neural activity during on and off response phases, a time window comprising activity at stimulus onset and withdrawal is selected (4 s on followed by 4 s off). This high-dimensional time-varying data are then projected along the first three eigenvectors of the response covariance matrix. Low-dimensional data points in consecutive time instances were connected to construct trajectories in the three-dimensional space. On the other hand, to compare the activities invoked by red/blue stimulus in on and off phases, a 4 s window comprising either on or off response is used.

#### Definition of metrics to analyze decoder performance

We analyze the quality of the latent representation generated by action of PNs through the following quantities.

- **Accuracy:** Accuracy quantifies the deviation of the latent space representation produced by PN activity from the true representation. This quantification was made at the end of the stimulus period, given mathematically by

$$accuracy = 1 - \frac{d(\mathbf{v}(t_s), \mathbf{z}) - d_{\min}}{d_{\max} - d_{\min}}, \quad (17)$$

where  $d(u, v)$  is a distant metric between the points  $u$  and  $v$ , and  $t_s$  is the time instant when the pulse in withdrawn. We use a Euclidean metric for distance in this case.

- **Latency:** Latency quantifies the time required by the synthesized network model to provide an accurate representation of the stimulus. It is computed as the time necessary to reach  $1 - \epsilon$  of the nominal representation  $\mathbf{z}$ . We used  $\epsilon = 0.2$ .

- **Similarity:** Yet another way of evaluating the quality of the latent representation produced by neural activity is by measuring the cosine of the angle between  $\mathbf{v}(t)$  at time  $t = t_s$  and  $\mathbf{z}$ . This can be interpreted as computing the similarity between the latent representation at the end of stimulus period and the desired nominal representation and is given mathematically as follows:

$$similarity \equiv \cos \theta = \frac{\mathbf{v}(t_s)^T \mathbf{z}}{\|\mathbf{v}(t_s)\| \|\mathbf{z}\|}. \quad (18)$$

However, we find that the trends observed via this quantification closely match that produced by *accuracy*. This is intuitive as the cosine metric can be written in terms of the Euclidean norm. Therefore, in this paper, we provide results using only accuracy and latency measures.

## Experimental design

### Odor stimulation

We followed a similar protocol as in our previous studies for odor stimulation (Saha et al., 2013b, 2017). The odor solutions were diluted in mineral oil (from Sigma-Aldrich) to achieve 1% dilution (v/v). Twenty milliliters of diluted odor solution was placed in a 60 ml sealed glass bottle with inlet and outlet lines. A constant volume ( $0.1 \text{ L min}^{-1}$ ) from the odor bottle headspace was injected into the carrier stream using a pneumatic pico-pump (PV-820; WPI Inc.) during odor presentations. A vacuum funnel was placed right behind the animal preparation to ensure the removal of odor vapors. Odor presentations were 4 s long. Each trial was 40 s in duration with an intertrial interval set to 20 s.

### Extracellular recordings

Post fifth instar locusts (*Schistocerca americana*) of either sex were selected and were immobilized first, and then the brain was exposed, desheathed as reported in previous studies (Laurent and Davidowitz, 1994; Brown et al., 2005; Saha et al., 2013a). Extracellular recordings of PNs were performed using NeuroNexus 16-channel,  $4 \times 4$  silicon probes. Impedances of all electrodes were kept in the 200–300 k $\Omega$  range. Raw extracellular signals were amplified at 10k gain using a custom-made 16-channel amplifier (Biology Electronics Shop; Caltech), filtered between 300 Hz and 6 kHz, and acquired at a 15 kHz sampling rate using a custom-written LabVIEW (National Instruments) software.

### Intracellular recordings

Animal preparation for intracellular recordings was the same as that used for extracellular recordings. Patch-clamp (in vivo) recordings from PNs were performed using the pipettes filled with locust intracellular solution (in mM; Laurent et al., 1993): 155 K aspartate, 1.5 MgCl<sub>2</sub>, 1 CaCl<sub>2</sub>, 10 HEPES, 10 EGTA, 2 ATP disodium salt, 3 D-glucose, 0.1 cAMP. All of these chemicals were purchased from Sigma-Aldrich. The pH of the patch solution was adjusted to 7.0 using 1 M NaOH, and osmolarity was adjusted to the 320–325 mM range using sucrose. The impedances of electrodes ranged between 5 and 15 M $\Omega$ . Raw voltage traces were amplified (Axoclamp 900a, Molecular Devices) and acquired at 16 kHz sampling rate using LabVIEW software.

## Results

### Sensory tracking is achieved by a biologically plausible network architecture

The posited optimization problem is as follows:

$$\min_{\mathbf{x}} \int_0^{\infty} \frac{1}{2} [(\mathbf{v}(t)^T - \mathbf{z})\mathbf{Q}(\mathbf{v}(t) - \mathbf{z}) + \mathbf{x}(t)^T \mathbf{S}\mathbf{x}(t) + \dot{\mathbf{x}}(t)^T \mathbf{R}\dot{\mathbf{x}}(t)] dt$$

subject to,  $\dot{\mathbf{v}} = \mathbf{A}\mathbf{v} + \mathbf{b}\mathbf{x}$ . (19)

The solution of Equation 19 yields the following PN dynamics (see Materials and Methods for details):

$$\dot{\mathbf{x}}(t) = \mathbf{W}_s \int_0^t e^{-a(t-\tau)} \mathbf{x}(\tau) d\tau + \mathbf{W}_f \mathbf{x}(t) + \mathbf{r}(t). \quad (20)$$

Equation (20) reveals the existence of a network that incorporates both slow and fast processing (via  $\mathbf{W}_s$  and  $\mathbf{W}_f$ , respectively) in response to an external stimulus. Also present is dependence on the exogenous stimulus via  $\mathbf{r}(t)$ . However, how the slower time scale is achieved *in vivo* is not explicitly explained by Equation 20. The question here is whether a biologically plausible first-order network architecture can be realized by decomposition of Equation 20 into first-order rate equations. We find that by introducing an auxiliary population of neurons which share reciprocal connections with the second-order PNs, this can be achieved. Equations 21 and 22 below is a representative network architecture arising from Equation 20 that produces the optimal motif as seen in Figure 3:

$$\dot{\mathbf{x}}_o(t) = \mathbf{W}_i^o \mathbf{x}_i(t) + \mathbf{W}_o^o \mathbf{x}_o(t) + \mathbf{r}(t), \quad (21)$$

$$\dot{\mathbf{x}}_i(t) = \mathbf{W}_i^i \mathbf{x}_i(t) + \mathbf{W}_o^i \mathbf{x}_o(t). \quad (22)$$

Here,  $\mathbf{x} \equiv \mathbf{x}_o \in \mathbb{R}^{n_o}$  is the PN activity,  $\mathbf{x}_i \in \mathbb{R}^{n_i}$  is the activity of the auxiliary population, and  $n_o$  and  $n_i$  are the number of neurons belonging to each population subgroup. The connectivity matrices in Equations 21 and 22 are analytically specified as  $\mathbf{W}_i^o = \mathbf{W}_s \widehat{\mathbf{W}}$ ,  $\mathbf{W}_o^o = \mathbf{W}_f$ ,  $\mathbf{W}_i^i = -a\mathbf{I}$ , and  $\mathbf{W}_o^i = \widehat{\mathbf{W}}$ , where  $\widehat{\mathbf{W}} \in \mathbb{R}^{n_i \times n_o}$  and  $\mathbf{W} \in \mathbb{R}^{n_o \times n_i}$  (see Materials and Methods for details). In the insect AL, PNs extend excitatory connections onto an auxiliary population of neurons known as local neurons (LNs), which in turn provide inhibitory control on the PNs (Wilson and Laurent, 2005; Kay and Stopfer, 2006; Shang et al., 2007). In our network architecture, connections between PNs and the auxiliary population are encapsulated in the matrices  $\mathbf{W}_i^o$  and  $\mathbf{W}_o^i$ . The structure of these matrices depend on choice of  $\widehat{\mathbf{W}}$  and  $\mathbf{W}$  (chosen such that  $\widehat{\mathbf{W}}\mathbf{W} = \mathbf{I}$ ). We set the dimensions of these matrices such that  $n_i < n_o$ , as cellular studies have reported that typically the size of the population of LNs is smaller than that of PNs (Laurent, 1996).  $\widehat{\mathbf{W}}$ ,  $\mathbf{W}$  are chosen through a constrained iterative optimization scheme (see Algorithm 1). The constraints for this secondary optimization problem are stated on the basis of the canonical excitatory–inhibitory structure found in the insect antennal lobe: (1)  $\widehat{\mathbf{W}}$  (PN to LN) must be non-negative, and (2)  $\mathbf{W}_s \widehat{\mathbf{W}}$  (LN to PN) must be nonpositive. We initialized our proposed algorithm by selecting  $\widehat{\mathbf{W}}$  to be a positive random block matrix such that its pseudoinverse existed, and we set the initial choice for  $\widehat{\mathbf{W}}$  as  $\widehat{\mathbf{W}}^\dagger$ , the Moore–Penrose pseudoinverse of  $\widehat{\mathbf{W}}$ . The choice of such an initialization was fashioned after our prior knowledge regarding existence of

stimulus-specific tuning of the PNs (Assisi et al., 2011; Saha et al., 2017). It is important to emphasize here that the solution of this iterative scheme is nonunique due to the structure of the problem ( $\widehat{\mathbf{W}}$  is a wide matrix and  $\widehat{\mathbf{W}}^\dagger$  is its left inverse). In other words, in general there will be many network architectures that realize the optimal motif from Equation 19. The common trend observed across various solutions (each a mathematically valid network architecture) is that the connections between PNs and LNs are essentially random (salt and pepper). This is consistent with *in vivo* observations wherein each PN extends its efferent terminals onto only a subset of the auxiliary population and receives inhibitory control only from a subset of the population (Bazhenov et al., 2001; Carey and Carlson, 2011; Saha et al., 2017). It is worthwhile to point out that the response motifs for LNs indeed depend on the resulting  $\widehat{\mathbf{W}}$  matrix, as it encapsulates how each PN projects its synapses on the putative LNs. But the motifs for PNs remain the phasic/tonic ones derived directly from Equation 19 because the cost function of the secondary optimization ensures  $\widehat{\mathbf{W}}\mathbf{W} \simeq \mathbf{I}$ .

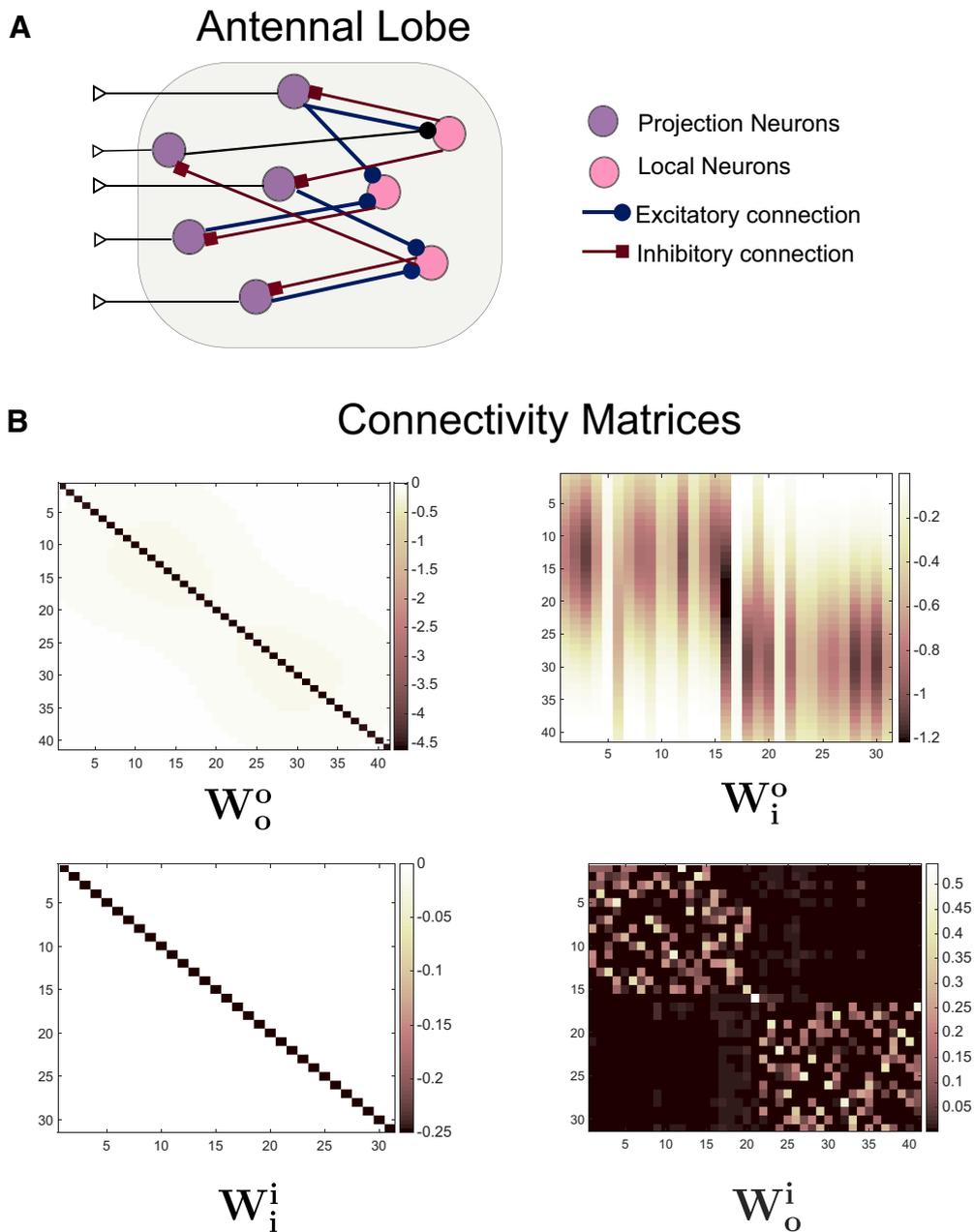
Intrapopulation interactions of PNs and LNs are captured in the matrices  $\mathbf{W}_o^o$  and  $\mathbf{W}_i^i$ . The diagonal terms in these matrices represent the self-decay dynamics associated with each neuron. Our construction also revealed very weak inhibitory connections within the PN population pool. Such connections are not known to exist experimentally, and we surmised that these small negative weights could be proxies of fast inhibitory synapses (Bazhenov et al., 2001; Bazhenov et al., 2005) occurring elsewhere in the antennal lobe (see also Discussion).

### Network dynamics produce stimulus onset and offset responses that are observed *in vivo*

Although the architecture of the normative model is consistent with that of the antennal lobe, it is not yet clear whether its dynamics and activity also match the types of responses observed *in vivo*. To study this issue, we set up an illustrative model for binary discrimination in a combinatorial setting. We emphasize here that the choice of a two-dimensional state space is made without loss of generality (see “Results” for an example of a system with higher-dimensional latent space).

The combinatorial nature of the encoding space is embedded in  $\mathbf{b}$  (see Eq. 1), which determines the tuning of individual neurons. In the example network, we used Gaussian tuning curves as per Figure 3, so that some neurons respond preferentially to one stimulus or the other (here, visualized as red vs blue), whereas some do not exhibit a tuning preference (referred to as untuned). It is observed that the PNs generate two significant phasic bursts of activity: one following odor onset and the other on odor termination (see Fig. 3B). These motifs are comparable to observations made *in vivo* in the locust antennal lobe (see Figure 4). Moreover, the on and off responses have orthogonal orientations in a dimensionality-reduced space and are negatively correlated (see Figure 5A–C and methods for dimensionality reduction details). Although spatiotemporal patterns of stimulus-evoked activity in the AL of insects has been studied previously in the literature (Laurent, 1996; Raman et al., 2010; Saha et al., 2017), our normative formulation provides an insight into the functional relevance of such response patterns (i.e., these motifs allow for robust and accurate stimulus representations for downstream processing in an energy-efficient manner).

Depending on their tuning, PN responses fall into one of the following categories: (1) a rapid increase in firing at the onset of stimulus that subsequently settles to a steady-state response of reduced

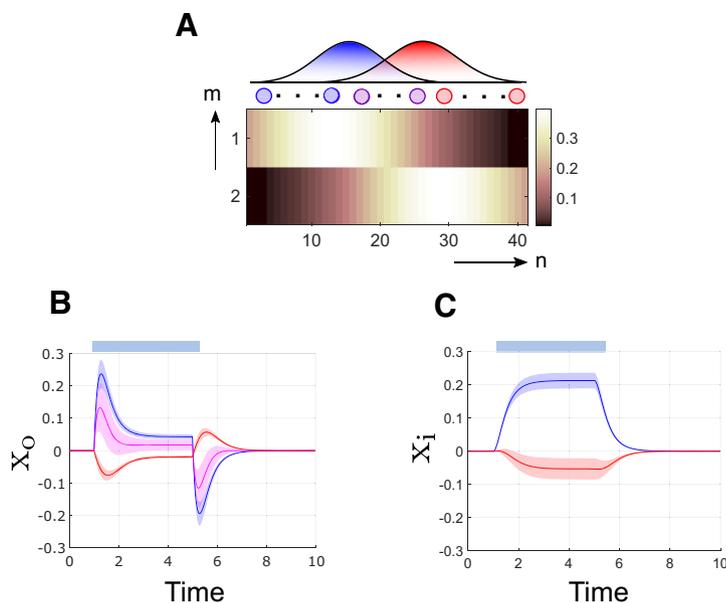


**Figure 2.** *A*, Schematic of synaptic interactions between PNs and LNs in the antennal lobe of insects. The PNs provide excitation to LNs, whereas LNs impose inhibitory control on PNs. *B*, Connectivity matrices arising for a particular choice of  $\bar{W}$ ,  $\hat{W}$  in our mathematical model shows the synaptic interactions between the two populations. It is consistent with the canonical excitatory–inhibitory scheme observed in sensory networks *in vivo*.

amplitude; on removal of stimulus, there exists a brief period of inhibited activity (below baseline) before the system returns to baseline; or (2) a state of suppressed firing activity through the entire duration of stimulus but excitatory (above baseline) firing activity when the stimulus is withdrawn (see Fig. 3*B,C*). On the other hand, the auxiliary neurons (i.e., the putative LNs) display tonic activity in response to stimuli, returning to baseline gradually when the stimulus is no longer active. Although the tuning of PNs has been studied extensively (Stopfer et al., 2003; Wilson et al., 2004), whether similar combinatorial architecture exists in the LNs is yet to be fully characterized. In Figure 3*C*, the blue/red coloration does not imply prior tuning but simply indicates whether the neurons were excited/inhibited by the stimulus (Saha et al., 2017).

Interestingly, these predicted response dynamics have been observed in recordings from the locust olfactory system, wherein

recent findings suggest both phasic and tonic temporal responses associated with stimulus onset and offset (Saha et al., 2017). Specifically, experimental observations indicate that when a pulse of stimulus is provided, the neurons in the antennal lobe exhibit a phasic transient followed by a tonic activity for the duration of the stimulus (referred to as *on response*). On termination of the pulse, there is another short-lived burst of activity before activity slowly returns to the quiescent regime (an *off response*; Saha et al., 2017). Figure 4 shows intracellular voltage traces recorded from four representative PNs in the locust antennal lobe. As can be noted, the stimulus-evoked responses in these PNs occur either when the stimulus is present followed by hyperpolarization (ON type), or following the termination of the stimulus (OFF type; note that membrane potential is hyperpolarized when the odor was presented). This segregation of stimulus-evoked PN



**Figure 3.** *A*, Combinatorial encoding is represented through the weights of matrix **b** in our model. Each row determines the contribution of the  $n = 41$  neural units (marked as colored circles) to evidence accrued corresponding to the high-level feature. *B*, *C*, Response motifs produced by the two populations (PNs and LNs) in presence of a blue pulse. Depending on their tuning to blue/red stimulus, the neurons may be excited or inhibited, respectively. *B*, PN activity displays two phasic transients, one on stimulus onset and the other on stimulus withdrawal; in between these transients the activity is steady but reduced in amplitude. *C*, LN activity is present chiefly as long as the stimulus is in action.

responses into these distinct categories is maintained even when a larger number of neurons probed with a wider panel was considered (Saha et al., 2017). It should be mentioned here that the ensemble response vectors during onset and offset responses have distinct geometric orientations and are negatively correlated with each other as reported by both model simulations and experimental recordings (Fig. 5A–C).

### Orthogonalization of population response achieves robust latent representation

In the previous section, we explored how the PNs achieve stimulus detection in a latent space via spatiotemporal response patterns. We sought to understand how PN response patterns that are optimized for stimulus detection affect other aspects of sensory processing. One prediction that was made by the model pertains to processing stimuli that are encountered in sequence, often without a sufficient interstimulus interval for the system to recover fully. Our model revealed that in such situations, resultant PN responses emphasize, albeit more strongly, on distinct orientations for distinct stimuli (in the reduced dimension space; Fig. 6A–C). This finding has been previously referred to as the contrast enhancement computation (Nizampatnam et al., 2018) that highlights the uniqueness of the current stimulus with respect to prior distractions (Fig. 6E). Here, we used our defined latent space of intermediate representation to analyze the need for such patterning. It turns out that what appeared as contrast enhancement/novelty detection through activity of PNs in fact enabled maintenance of robust latent representation (Fig. 6D) in the face of a variety of background distractions.

Importantly, the objective function in our optimization framework (Equation 2) stressed only on formation of accurate latent representation in an energy-efficient manner for a single stimulus at a time. Regardless, this framework predicted response features that are observed in the biological world where stimuli are often perceived

under a variety of contexts (Nizampatnam et al., 2018). This functional interpretation of what is achieved by neural response modulation when stimuli appear in conjunction with a multitude of distractions is an interesting outcome of this study.

### Active reset enables consecutive fast and accurate representations

As suggested by the above finding, the synthesized network embodies the notion of *active reset* via its off response. We further analyzed the functional advantage of such distinct reversal of firing trends at the conclusion of the active stimulus period by systematic simulations of our model. Here, we considered two model alternatives: (a) the synthesized network and (b) a network whose activity ceases immediately on withdrawal of the stimulus and the decoder relies only on its natural decay to return to neutral (“passive reset”). We simulated each of these models with a sequence of stimuli (Fig. 6A, right). It turned out that the active reset mechanism mediated via the off response prepares the system promptly for the next incoming excitation. In the absence of such swift reversal of firing activity at the conclusion of a stimulus, the system “recovered” at a much slower rate, leading to confounding latent representation of succeeding stimuli (Fig. 7A). Clearly, the extent of misrepresentation depends on the degree of passive reset (i.e., the parameter  $a$ ) as well as the time allowed for the system to recover between pulses (Fig. 7B; the definition of accuracy in this context is included in the Materials and Methods).

### The optimal response motif predicts the existence of a fixed-point attractor

It has been found that PN population response comprising phasic/tonic activity patterns when analyzed in the reduced principal component space traces distinct, odor-specific trajectories. Starting from the origin or a neutral point, each odor-evoked trajectory quickly coalesced to a specific “fixed point” in the neural state space (Galán et al., 2004; Mazor and Laurent, 2005). The model in this study is designed to “track” a nominal representation in the low-dimensional decoded latent space with high confidence. In doing so, the derived optimal neural responses displayed a sharp transient activity at both stimulus onset and offset (Fig. 3B); in the period between these transients, the PNs responded to the stimulus by a steady-state response of low amplitude. Therefore, the model indeed predicts the existence of an odor-specific fixed-point attractor in the neural state space (see Figure 8). Beginning from a neutral state, it took only a few milliseconds to be in the proximity of the attractor determined by stimulus identity; likewise, on withdrawal of the stimulus, it is the transient activity that propelled the system to quickly return to its neutral regime. The interesting observation here is the speed with which the neural response mediated detection and “undetection,” as our optimization framework (Eq. 2) is formulated to produce representations over an infinite time horizon without any bias toward when specific events (i.e., detection) should occur.

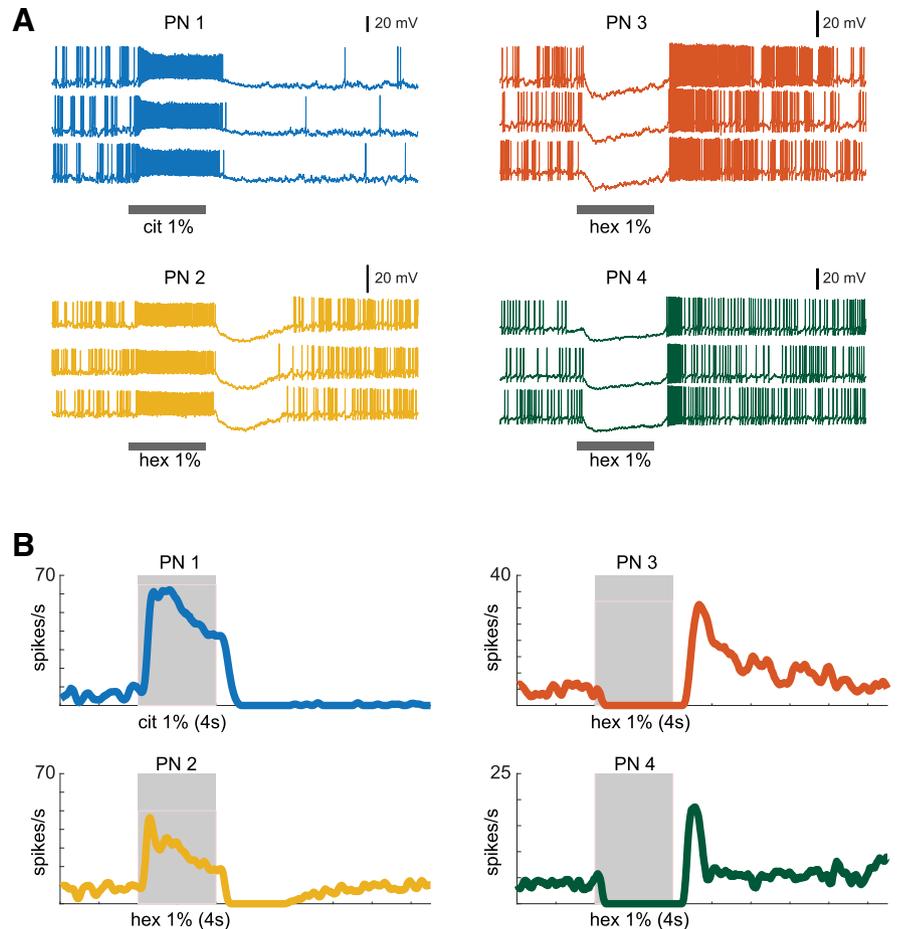
### Encoding performance degrades gracefully with respect to increasing tuning overlap

Combinatorial coding in olfaction is ubiquitously present (Menini, 2009; Kundu et al., 2016), as this scheme provides efficient utilization of limited neural resources for identification and processing of a very large number of odors. The question is, how robust is the performance of the model across different degrees of overlap in the tuning curves? In our mathematical setup,  $C$  is the fraction of neurons that responds to both red/blue stimuli with no overt preference to either stimuli; therefore, it also indicates the extent of overlap between the tuning curves or similarity between the feature space of afferent stimuli. The synthesized model predicted that the network can maintain accurate representations of the stimuli for a large range of values of  $C$  (see Fig. 9A,B). Beyond this range of overlap, however, there is a rapid decline in the quality of representation. It is evident that systems which endured large expenditures in terms of energy and maintained strict demands for error minimization were more tolerant to increased overlap in the combinatorial space (see Fig. 9B) than their counterparts.

In this context the next question is, can this graceful deterioration in performance be attributed to some embedded feature of the optimal motif? On analysis, we found that in fact it is the activity of the competing neurons that alleviated both transient and asymptotic misrepresentation over a considerable range of overlap in the tuning curves. We have previously identified that the neurons preferentially selective to a given stimulus remain in a state of suppressed activity in the presence of a competing stimulus (Figure 3B). When this activity is eliminated by reducing the contribution of the red-tuned neurons to the latent decoder to baseline, we observed that the latent state trajectory tracked  $\mathbf{z}$  poorly. Furthermore, the inherent property of graceful degradation of the synthesized network also disappeared in the absence of competitive inhibition (see Fig. 9C,D).

### Optimal motifs mediate a trade-off between latency, energy, and accuracy

The choice of the error penalty matrix  $\mathbf{Q}$  and energy regularization matrix  $\mathbf{S}$  in Equation 2 shaped the optimal neural responses and subsequently influenced the performance of sensory tracking in the latent space. Intuitively, as  $\mathbf{Q}$  is scaled to higher values, stricter demands are made to ensure high-fidelity tracking, whereas increasing the scale of  $\mathbf{S}$  encourages greater degree of conservation of energy. The notion of a performance trade-off here is embedded, therefore, in the choice of matrices  $\mathbf{Q}$  and  $\mathbf{S}$  (Fig. 10A,B), as they share an antipodal relationship with respect to accuracy as well as latency (time to reach 80% of the nominal evidence level). In the Materials and Methods section of this paper, we have reported how choosing a cosine metric to measure

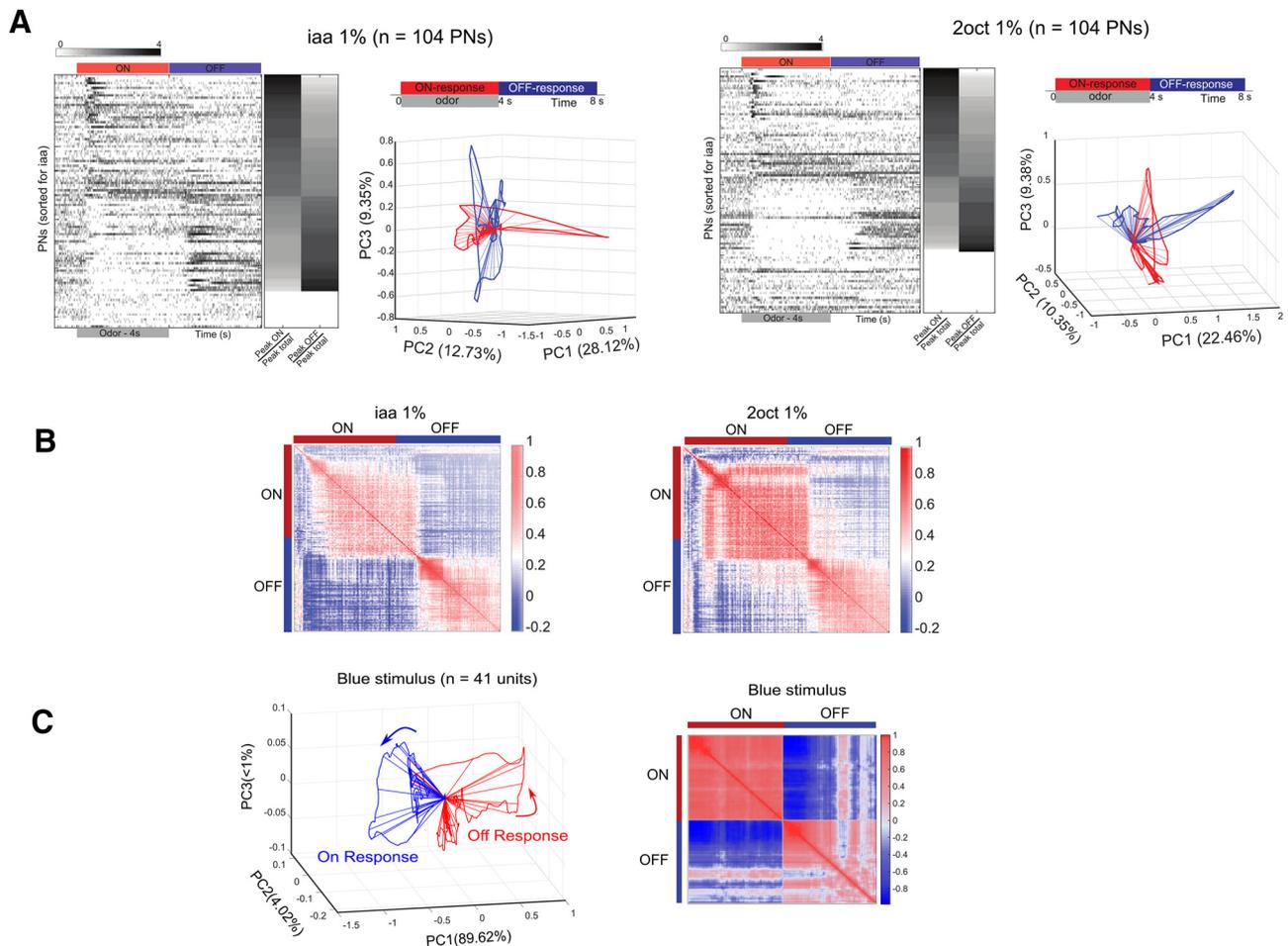


**Figure 4.** *A*, Intracellular voltage traces of four different PNs are shown (different colors). Shaded gray box in each panel shows the 4 s duration of odor exposure (cit, citral; hex, hexanol; both delivered at 1% v/v dilutions), and the black line represents the scale bar (20 mV). Three traces in each panel correspond to three consecutive trials of recording, revealing that the temporal patterns are consistent between trials. PN1 and PN2 responded during the presentation of odor (*on response*), and PN3 and PN4 responded after the termination of odor (*off response*). *B*, PN firing rates in 50 ms time bins were averaged across trials and are shown in *A*.

the quality of latent representation can be used. Figure 10C,D illustrates the response morphology associated with different penalties. This indicated that low penalty on tracking error and high regularization on energy produced smaller overshoots on stimulus onset. Such responses, although still capable of forming latent space representations, did so more slowly and less accurately (Fig. 10A,B, cyan squares on surface plots). Phasic overshoot, conversely, is indicative of a network designed to promote accurate and fast representations (Fig. 10A,B, pink circles on surface plots). Thus, our finding is compatible with the argument that short stimulus exposures are sufficient for olfactory discrimination (Uchida and Mainen, 2003), at the cost of higher energy expenditure within sensory networks.

### Higher-dimensional generalization of the latent space

In this section, we address how the latent space can be generalized to higher dimensions so that a large number of odor inputs can be represented. Here, we consider six odors (blue/magenta/red/olive/green/cyan) that map to a three-dimensional high-level feature space (Fig. 11A,B) through matrix  $\mathbf{b}$ . Similar to the two-dimensional case, neural activity causes accrued evidence  $\mathbf{v}(t)$  toward track  $\mathbf{z}$  in the latent space (Fig. 11C). Through this simulation, we found out that through choice of an appropriate tuning matrix, it



**Figure 5.** *A*, Mean PN firing rates (50 ms time bins and averaged across 10 trials) are shown for each PN (single row) and across the entire PN ensemble (different rows). The 4 s when the odorant was present (i.e., on period) and 4 s after the termination of the stimulus (i.e., off period) are identified using a red and a blue bar, respectively. PNs are ordered based on the difference between the peak firing rates during ON and OFF epochs (Saha et al., 2017). Nonresponsive neurons are shown at the bottom. Normalized peak firing responses during ON and OFF periods for each PN are shown on the right of the panel as a color bar. Neural ensemble response trajectories after PCA dimensionality reduction are shown. Red trajectory corresponds to the 4 s when isoamyl acetate (iaa) was presented, and the blue trajectory corresponds to the 4 s after termination of the stimulus. Similar plots for two different odorants, iaa at 1% v/v (left) and 2-octanol at 1% v/v (right), are shown. *B*, Correlation between ensemble response evoked by an odorant; the ON and OFF responses are negatively correlated. *C*, PCA trajectories and correlation maps generated during 4 s stimulus (blue) onset and 4 s following stimulus withdrawal generated by the noisy model are shown. Arrows indicate the direction of evolution of the trajectories.

is possible to represent a relatively large number of odors in a low-dimensional latent space.

## Discussion

In this work, we have proposed a mechanistic model for sensory detection that identifies the functional significance of spatiotemporal motifs observed ubiquitously across species in the context of olfaction. Previous studies on this topic (Stopfer et al., 2003; Mazor and Laurent, 2005; Bathellier et al., 2008; Raman et al., 2010; Saha et al., 2013b, 2017) have used *post hoc* analysis to surmise the role of the aforementioned responses. In contrast, we present here a normative modeling perspective: we formally optimize PN response dynamics to achieve input separation in an intermediate latent space in an energy-efficient manner.

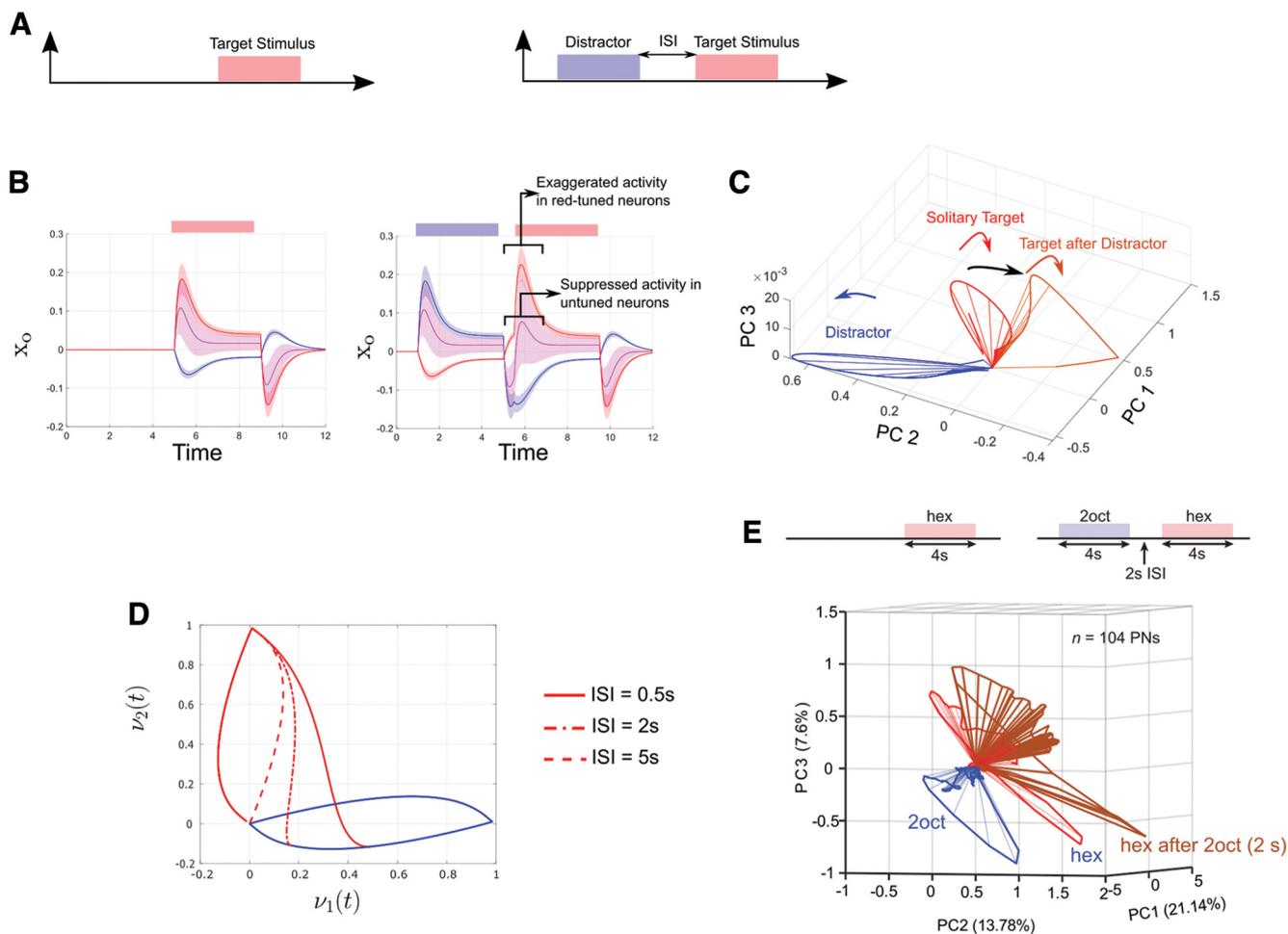
The key contributions of this paper are summarized as follows:

1. We showed that sensory onset and offset responses can be explained by a latent decoding objective function and associated mathematical optimization thereof.

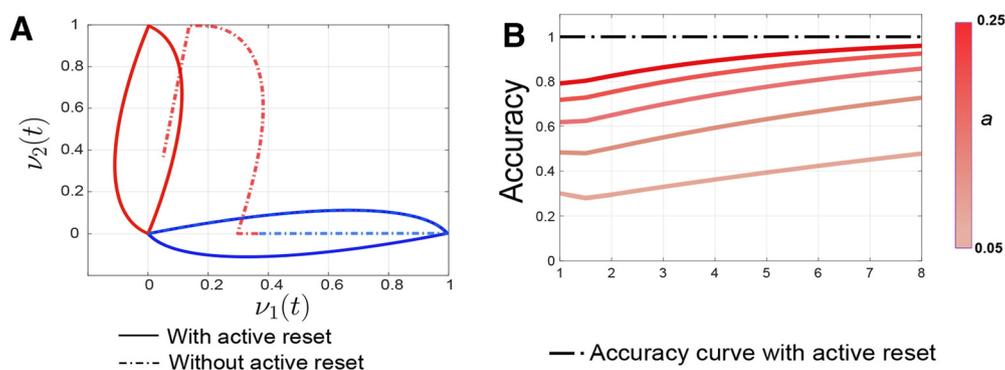
2. We showed that a biologically plausible network architecture is capable of producing the optimal responses elucidated by 1.
3. We showed that this optimization leads to several emergent predictions regarding the mapping from stimulus to latent representation for more complex stimuli.

## Normative interpretation of PN response dynamics

We adopted a top-down approach and formulated a representation tracking problem wherein PNs must fluidly and efficiently encode the stimulus such that it generates a robust and coherent latent representation for downstream processes. It turned out that, to achieve this objective, the optimal PN dynamics exhibited two distinct overshoots above baseline activity, one on onset of stimulus and the other on withdrawal (Fig. 3). The analytical responses also exhibited distinct geometric orientations during stimulus on and off phases, analogous to what has been reported *in vivo* (Fig. 5). It is worth noting here that the optimization problem we have solved amounts to the linear quadratic regulator problem, a classical construct in control theory (Anderson and Moore, 2007).



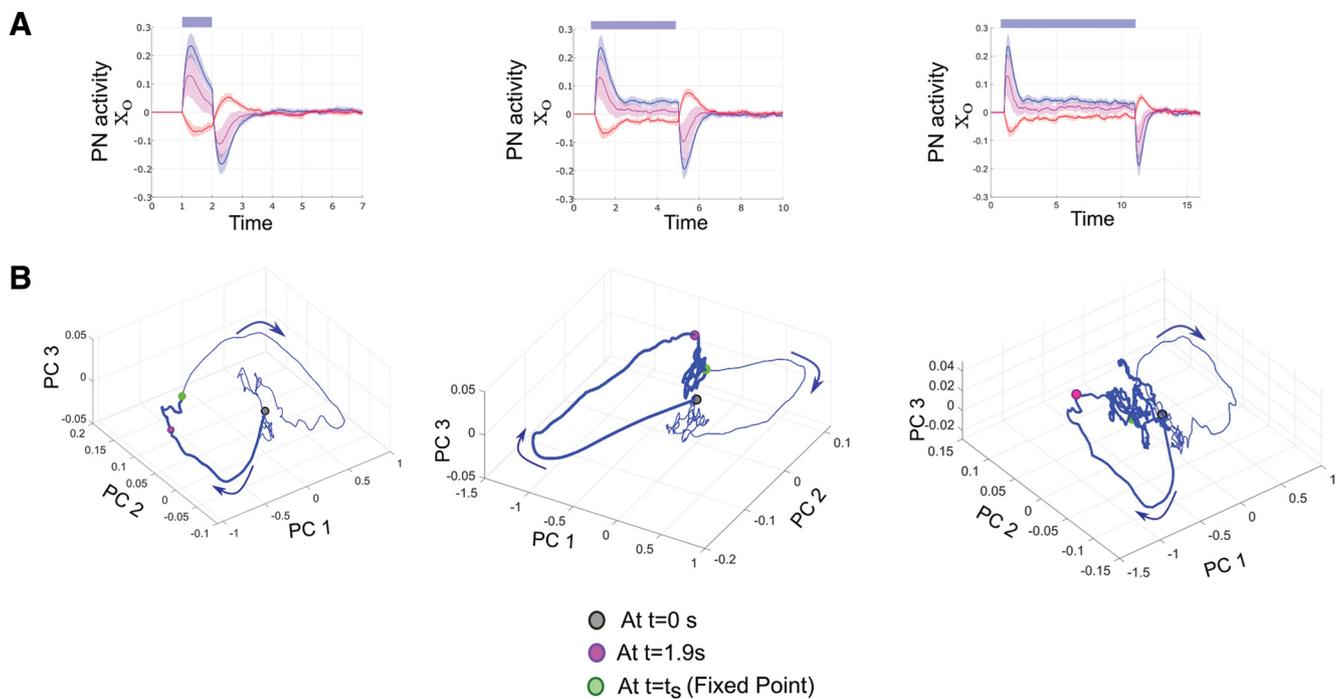
**Figure 6.** *A*, The synthesized model is excited by (a) a solitary red target pulse or (b) a blue distractor pulse followed by a red target pulse. *B*, Change in neural activity due to presence of distractor cue. *C*, PN activity visualized after dimensional reduction: response trajectory in the distractor-target sequence implements contrast enhancement (Nizampatnam et al., 2018). Arrows indicate the direction of evolution of the trajectories. *D*, Latent state representation for a range of interstimulus interval. The model predicts that robust latent realizations are possible even when interstimulus intervals are very small due to contrast enhancement computations. *E*, Similar plots as in Figure 5*A* are shown for a sequential presentation of hexanol (hexanol after 2 s ISI shown as brown trajectory). Neural response trajectories for solitary presentations of 2-octanol (blue trajectory) and hexanol (red trajectory) are also plotted for comparison. Note that, compared with the red trajectory, brown trajectory is more distant from the blue trajectory.



**Figure 7.** *A*, In the absence of active reset, latent state trajectory evolves with a bias introduced by the previous cue. *B*, The amount of bias introduced in the system depends on the intrinsic dynamics of the decoder controlled by  $a$  and the time between two stimuli. A comparison is made with a system implementing active reset (gray line). Here we sweep the parameter  $a$  over a range of 0.05 to 0.25.

We have shown here that an excitatory–inhibitory network architecture with first-order dynamics can implement the motifs generated by the optimization framework. We derived the connectivity weights between the second-order PNs and an auxiliary

population of LNs by means of an iterative optimization scheme (see Fig. 2 and Materials and Methods for details). Our results, however, indicate that there might be several plausible ways to mathematically specify such a network.



**Figure 8.** *A*, PN population response for different lengths of stimulation (i.e.,  $t_s = 1, 4, 10$  s). *B*, PN trajectories as visualized in the PC space. Gray circle indicates the initial point, pink circle indicates the end of transient (phasic) response period, and green circle indicates the fixed point to which the population response converges asymptotically in presence of a stimulus. The thicker part of the trajectory indicates the period for which stimulus was on. The arrows indicate the direction of traversal.

We probed our model further to highlight the relevance of certain response features that appear both in simulations and through experimentation. The tractable construct facilitates a comprehensive understanding through rigorous mathematical simulations for different ranges of parameters and model alternatives. We analyzed the quality of the latent space representations formed using metrics of accuracy and latency to interpret what benefits neural response features, such as competitive inhibition, contrast enhancement, etc., provide in detecting a stimulus. A key result of this study was that the constructed sensory networks were capable of distinguishing between highly similar inputs (see Fig. 9). There is in fact no inherent trade-off between accuracy (i.e., input separation) and speed; rather, both are monotonic with respect to energetic considerations (Fig. 10). Several studies in the literature have discussed these notions through experimental observations and data-driven, bottom-up modeling approaches (Wehr and Laurent, 1999; Rabinovich et al., 2000; Ito et al., 2008; Chittka et al., 2009; Saha et al., 2017). Here, we presented a top-down normative framework that began from the specification of a mathematical form for the decoding objective and systematically navigated the results to provide a functional interpretation of patterns observed *in vivo*.

#### Normative model produces known sensory neural dynamics beyond onset and offset responses

From our high-level detection objective, our normative model predicted several complex neural computations observed in the early olfactory networks beyond onset and offset responses.

For instance, experimental studies (Galán et al., 2004; Mazar and Laurent, 2005) have established that the trajectory in the neural state space quickly converges to a stimulus-specific fixed point: the transient activity being crucial in driving the trajectories toward these fixed-point attractors. Persistence of weak response (tonic activity) after the initial strong onset response (phasic activity) emerged in the synthesized model due to

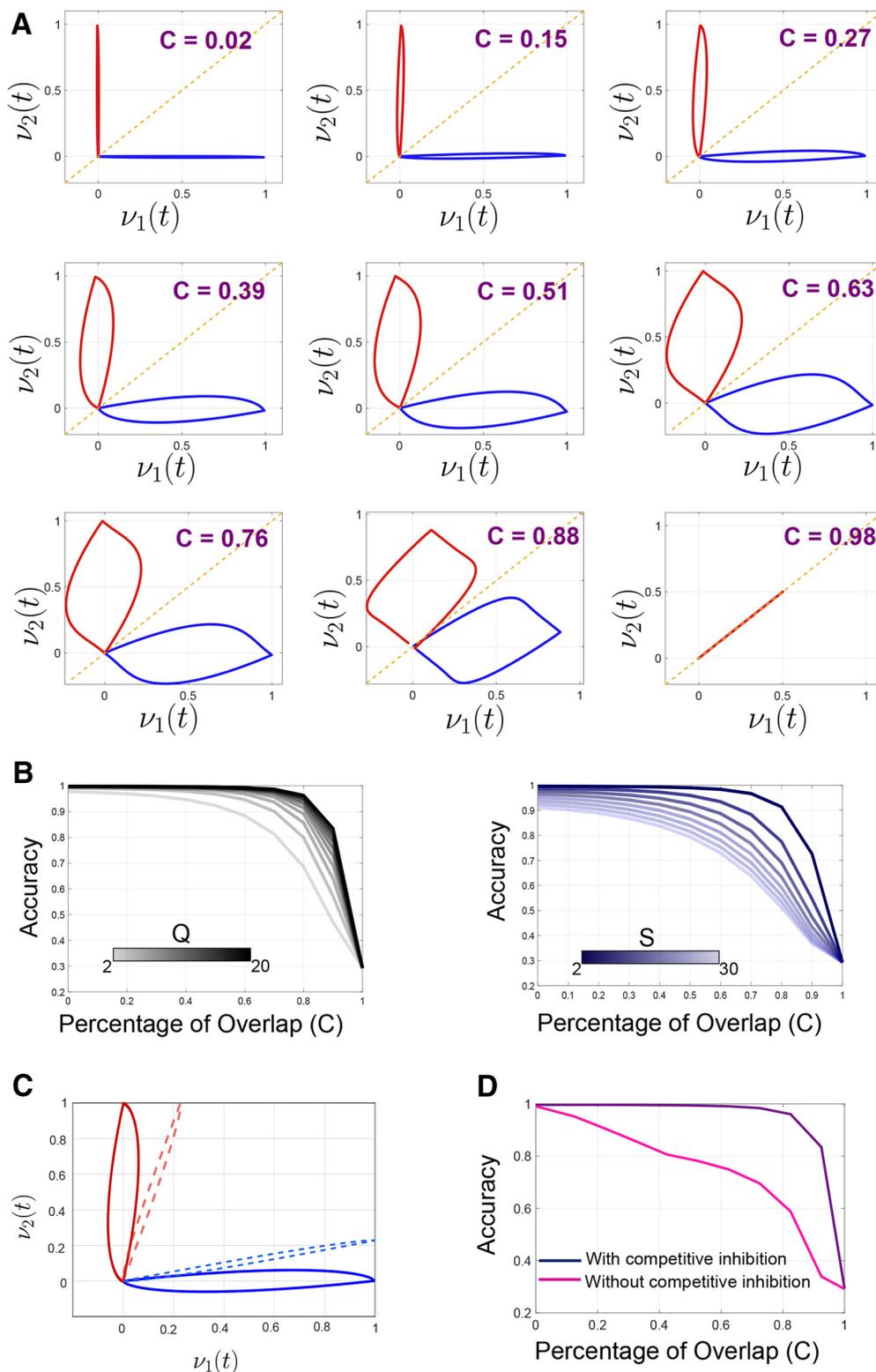
interplay between the first and second terms in the objective function—namely, accuracy of latent representation and energy of sensory activity (Fig. 8). So, the model indeed provides an objective way to think about the need for these fixed-point attractors.

Additionally, our model also provided a novel interpretation of contrast enhancement computation (Nizampatnam et al., 2018) known to exist at a population level in PNs. It turned out that by highlighting the uniqueness of the current stimulus while sensing stimuli in series with very small interstimulus intervals, the PNs ensure robustness of latent representation required for reliable functioning of downstream processes (Fig. 6).

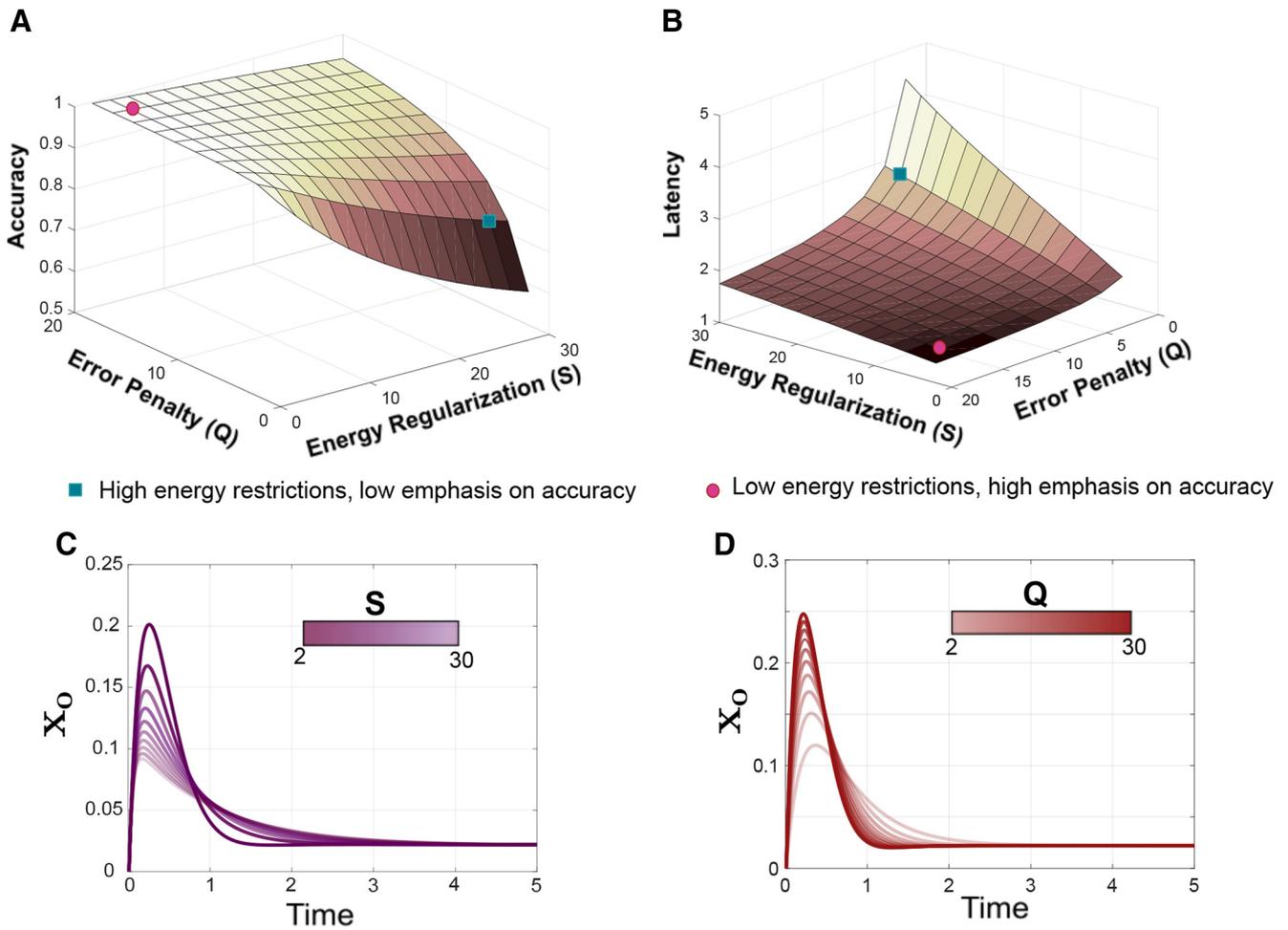
To summarize, we found that our framework, built only to ensure accurate representation tracking in a low-dimensional latent space, exhibits (emergent) observations reported in the activity of olfactory systems, thus providing a unified explanation for these activity patterns.

#### Generality of the decoding model

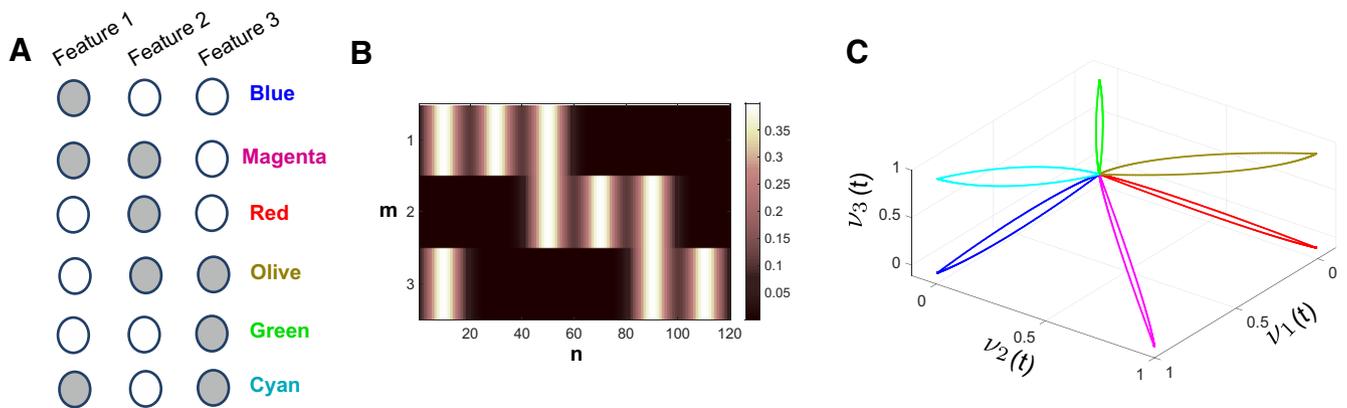
The decoder we have used is a multivariate linear dynamical system driven by PN activity. Mathematically, this is similar to drift-diffusion-type models that have been used to study high-level decision-making [e.g., the two-alternative forced choice task (Bogacz et al., 2006; Colman, 2015)]. In addition to the obvious difference in the level of behavioral abstraction being considered, there are important differences in formulation between our work and these prior results. In particular, our framework caters to the need of representations being fluidly created and then abolished: they must *persist* through the duration of the stimulus and *dissipate* quickly after its withdrawal. This departs from theoretical decision-making paradigms such as “interrogation” or “free response” (Bogacz et al., 2006) where a subject must make a detection within a set amount of time, and where the decoding process essentially terminates at the time of decision. Indeed, there is no



**Figure 9.** *A*, Trajectories in the latent space evolve with reduced spatial separation with increased overlap between Gaussian tuning curves embedded in **b**. However, the degree of misrepresentation (inversely related to spatial separation) is appreciably low across a wide range of overlap. *B*, Accuracy of the normative model as a function of the percentage of overlap. We notice that, after a critical point, the performance degrades drastically. This point along the horizontal axis is a function of penalty matrix  $Q$  (left) and regularization matrix  $S$  (right). *C*, Competitive inhibition is pivotal for ensuring accurate representation. In the absence of competitive inhibition, the trajectories corresponding to red and blue stimulus are more proximal to each other (dashed line). *D*, Comparison between systems with and without competitive inhibition. Fifty percent of the neurons responsive to the red stimulus are switched off while a blue stimulus was provided. The accuracy of the system with impaired competitive inhibition declined rapidly.



**Figure 10.** *A*, Accuracy is plotted as a function of error penalty matrix  $Q$  and energy regularization matrix  $S$ . *B*, Latency of forming a reasonably accurate representation is plotted as a function of error penalty matrix  $Q$  and energy regularization matrix  $S$ . (Note that the axis directions in *A* and *B* are reversed for better visualization of surface plots.) *C*, Response motifs across choice of energy regularization matrix  $S$ . With a higher penalty on energy, the amplitude of the overshoot decreases. *D*, Response motifs across choice of error penalty matrix  $Q$ . With highly scaled error penalty, the systems produce a sharp overshoot followed by a prompt return to baseline.



**Figure 11.** *A*, Odors and their respective mapping to the high-level feature space. *B*,  $b \in \mathbb{R}^{m \times n}$  weighs the contribution of each neuron to each feature. *C*, Latent space trajectory for each of the six representative odors.

“threshold” in our model, only a latent representation that is continuously evolving in its state space.

In the scenario where the number of excitation sources (i.e., odors) is very large (possibly infinite), we can continue to use this normative framework for analysis by choosing an

appropriate  $b$  matrix. In such a case, we could formulate  $b$  so that each row contains weights corresponding to a member of an appropriate basis set. Then, it is possible to translate the infinite set of inputs to their corresponding latent representations by linear combination of the chosen basis functions. The exact weight

distributions for such basis set is yet to be determined and remains a question for future study.

### Features not explained

There are a number of important caveats and limitations that should be pointed out regarding the interpretation of the normative model presented through a set of first-order firing rate dynamics (Dayan and Abbott, 2001). In particular, our synthesized network has linear dynamics wherein neurons can assume both positive and negative firing rates. Such a formulation is acceptable if we envision the rate variables as a change relative to some positive baseline rate of activity. Harder to reconcile in the linear model is a lack of upper and lower bounds of firing rate. Solving the optimization problem with such constraints is a harder problem mathematically (Chachuat, 2007) that we have not yet resolved.

Our construction produces local neurons that are exclusively inhibitory in nature and they interact only via the excitatory neurons. However, there are reports of existence of excitatory local neurons in the antennal lobe of some insects (Shang et al., 2007; Masse et al., 2009; Assisi et al., 2012). The functional role of such excitatory LNs is not accounted for in our model. However, in the mathematical formulation, these excitatory components can easily be realized by relaxing the nonpositivity constraint in the iterative optimization procedure (see Materials and Methods for details). Our construction also produces very weak lateral inhibition directly between PNs ( $\mathbf{W}_0^o$ ; Fig. 2). Although there is no experimental evidence of existence of such PN-PN inhibitory synapses, fast GABAergic synapses (Bazhenov et al., 2001, 2005) are known to exist in the insect antennal lobe (between LN-LN and LN-PN) and thought to be functionally relevant for maintaining the synchrony of PN responses (Bazhenov et al., 2001). We theorize that the inhibitory synaptic weights between PNs produced by solving the reduced linear quadratic regulator problem (Eq. 5) are in fact proxies for these fast inhibitory synapses that have not been explicitly considered in the synthesized network.

In the objective function (Eq. 2), we use a Euclidean distance to evaluate the quality of the latent representation as it tracks the incoming stimulation. However, downstream neurons might achieve input separation before higher-level processing by more involved computations. Experimental studies identify spatial and temporal filtering, decorrelation, etc. (Friedrich, 2013) as operations that could be plausibly executed by sensory neurons. In the present work, we choose to use the well characterized  $\ell_2$  norm as a surrogate for specialized computations that might occur *in vivo*. It is as yet an open question how incorporating such functions would impact our synthesis results.

Additionally, we have not considered any form of adaptation or habituation in our network, choosing to focus on a single exposure regime. The question of how the optimal weights come to be (e.g., via development or learning) is not considered herein.

Finally, our model makes no statements about the translation from latent representations to behavior. Behavior can exhibit nonlinear and sometimes paradoxical characteristics with respect to stimulus intensity, composition, and context. Accounting for these factors within the optimization problem is highly intriguing, but is a more complex formulation that is left for future study.

### References

Aldworth Z, Stopfer M (2015) Trade-off between information format and capacity in the olfactory system. *J Neurosci* 35:1521–1529.

- Anderson B, Moore J (2007) Optimal control: linear quadratic methods. Mineola, NY: Dover Publications.
- Assisi C, Stopfer M, Bazhenov M (2011) Using the structure of inhibitory networks to unravel mechanisms of spatiotemporal patterning. *Neuron* 69:373–386.
- Assisi C, Stopfer M, Bazhenov M (2012) Excitatory local interneurons enhance tuning of sensory information. *PLoS Comput Biol* 8:e1002563.
- Bathellier B, Buhl D, Accolla R, Carleton A (2008) Dynamic ensemble odor coding in the mammalian olfactory bulb: sensory information at different timescales. *Neuron* 57:586–598.
- Bazhenov M, Stopfer M, Rabinovich M, Huerta R, Abarbanel HD, Sejnowski TJ, Laurent G (2001) Model of transient oscillatory synchronization in the locust antennal lobe. *Neuron* 30:553–567.
- Bazhenov M, Stopfer M, Sejnowski T, Laurent G (2005) Fast odor learning improves reliability of odor responses in the locust antennal lobe. *Neuron* 46:483–492.
- Bhandawat V, Olsen SR, Gouwens NW, Schlieff ML, Wilson RI (2007) Sensory processing in the drosophila antennal lobe increases reliability and separability of ensemble odor representations. *Nat Neurosci* 10:1474–1482.
- Bogacz R, Brown E, Moehlis J, Holmes P, Cohen JD (2006) The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced choice tasks. *Psychol Rev* 113:700–765.
- Boyd S, Barratt C (1991) Linear controller design: limits of performance. Englewood Cliffs, NJ: Prentice Hall.
- Brown S, Joseph J, Stopfer M (2005) Encoding a temporally structured stimulus with a temporally structured neural representation. *Nat Neurosci* 8:1568–1576.
- Carey A, Carlson J (2011) Insect olfaction from model systems to disease control. *Proc Natl Acad Sci USA* 108:12987–12995.
- Chachuat B (2007) Nonlinear and dynamic optimization: from theory to practice, Technical report.
- Chittka L, Skorupski P, Raine N (2009) Speed-accuracy tradeoffs in animal decision making. *Trends Ecol Evol (Amst)* 24:400–407.
- Colman A (2015) A dictionary of psychology. Oxford, UK: Oxford UP.
- Dasgupta S, Stevens C, Navlakha S (2018) A neural algorithm for a fundamental computing problem. *Science* 358:793–796.
- Dasgupta S, Sheehan T, Stevens C, Navlakha S (2017) A neural data structure for novelty detection. *Proc Natl Acad Sci USA* 115:13093–13098.
- Dayan P, Abbott L (2001) Theoretical neuroscience: computational and mathematical modeling of neural systems. Cambridge: MIT press.
- Friedrich R (2013) Neuronal computations in the olfactory system of zebrafish. *Annu Rev Neurosci* 36:383–402.
- Galán R, Sachse S, Galizia C, Herz A (2004) Odor-driven attractor dynamics in the antennal lobe allow for simple and rapid olfactory pattern classification. *Neural Comput* 16:999–1012.
- Grant M, Boyd S, Ye Y (2008) CVX: MATLAB software for disciplined convex programming. Available at <http://cvxr.com/cvx/>.
- Ito I, Ong R, Raman B, Stopfer M (2008) Sparse odor representation and olfactory learning. *Nat Neurosci* 11:1177–1184.
- Kay L, Stopfer M (2006) Information processing in the olfactory systems of insects and vertebrates. *Semin Cell Dev Biol* 17:433–442.
- Kundu S, Ganguly A, Chakraborty TS, Kumar A, Siddiqi O (2016) Synergism and combinatorial coding for binary odor mixture perception in drosophila. *eNeuro* 3:ENEURO.0056-14.2016.
- Laurent G (1996) Dynamical representation of odors by oscillating and evolving neural assemblies. *Trends Neurosci* 19:489–496.
- Laurent G, Davidowitz H (1994) Encoding of olfactory information with oscillating neural assemblies. *Science* 265:1872–1875.
- Laurent G, Seymour-Laurent K, Johnson K (1993) Dendritic excitability and a voltage-gated calcium current in locust nonspiking local interneurons. *J Neurophysiol* 69:1484–1498.
- Laurent G, Stopfer M, Friedrich RW, Rabinovich MI, Volkovskii A, Abarbanel HD (2001) Odor encoding as an active, dynamical process: experiments, computation and theory. *Ann Rev Neurosci* 24:263–297.
- Martin JP, Beyerlein A, Dacks AM, Reisenman CE, Riffell JA, Lei H, Hildebrand JG (2011) The neurobiology of insect olfaction: sensory processing in a comparative context. *Prog Neurobiol* 95:427–447.
- Masse N, Turner G, Jefferis G (2009) Olfactory information processing in drosophila. *Curr Biol* 19:R700–R713.
- Mazor O, Laurent G (2005) Transient dynamics versus fixed points in odor representations by locust antennal lobe projection neurons. *Neuron* 48:661–673.

- Menini A (2009) The neurobiology of olfaction. Boca Raton, FL: CRC.
- Nizampatnam S, Saha D, Chandak R, Raman B (2018) Dynamic contrast enhancement and flexible odor codes. *Nat Commun* 9:3062.
- Qin S, Li Q, Tang C, Tu Y (2019) The optimal odor-receptor interaction network is sparse in olfactory systems: compressed sensing by nonlinear neurons with a finite dynamic range. *bioRxiv*. doi:10.1101/464875.
- Rabinovich MI, Huerta R, Volkovskii A, Abarbanel HDI, Stopfer M, Laurent G (2000) Dynamical coding of sensory information with competitive networks. *J Physiology-Paris* 94:465–471.
- Raman B, Joseph J, Tang J, Stopfer M (2010) Temporally diverse firing patterns in olfactory receptor neurons underlie spatiotemporal neural codes for odors. *J Neurosci* 30:1994–2006.
- Raman B, Stopfer M, Semancik S (2011) Mimicking biological design and computing principles in artificial olfaction. *ACS Chem Neurosci* 2:487–499.
- Saha D, Leong K, Katta R, Raman B (2013a) Multi-unit recording methods to characterize neural activity in the locust (*Schistocerca americana*) olfactory circuits. *J Vis Exp*. Advance online publication. Retrieved January 25, 2013. doi: 10.3791/50139.
- Saha D, Leong K, Li C, Peterson S, Siegel G, Raman B (2013b) A spatiotemporal coding mechanism for background-invariant odor recognition. *Nat Neurosci* 16:1830–1839.
- Saha D, Sun W, Li C, Nizampatnam S, Padovano W, Chen Z, Chen A, Altan E, Lo R, Barbour DL, Raman B (2017) Engaging and disengaging recurrent inhibition coincides with sensing and unsensing of a sensory stimulus. *Nat Commun* 8:15413.
- Schaefer A, Margrie T (2007) Spatiotemporal representations in the olfactory system. *Trends Neurosci* 30:92–100.
- Shang Y, Claridge-Chang A, Sjulson L, Pypaert M, Miesenböck G (2007) Excitatory local circuits and their implications for olfactory processing in the fly antennal lobe. *Cell* 128:601–612.
- Shlizerman E, Riffell J, Kutz J (2014) Data-driven inference of network connectivity for modeling the dynamics of neural codes in the insect antennal lobe. *Front Comput Neurosci* 8:70.
- Stopfer M, Jayaraman V, Laurent G (2003) Intensity versus identity coding in an olfactory system. *Neuron* 39:991–1004.
- Strausfeld N, Hildebrand J (1999) Olfactory systems: common design, uncommon origins? *Current opinion in neurobiology* 9:634–639.
- Uchida N, Mainen Z (2003) Speed and accuracy of olfactory discrimination in the rat. *Nat Neurosci* 6:1224–1229.
- Vickers N (2000) Mechanisms of animal navigation in odor plumes. *Biol Bull* 198:203–212.
- Wehr M, Laurent G (1999) Relationship between afferent and central temporal patterns in the locust olfactory system. *J Neurosci* 19:381–390.
- Wilson R, Laurent G (2005) Role of gabaergic inhibition in shaping odor-evoked spatiotemporal patterns in the drosophila antennal lobe. *J Neurosci* 25:9069–9079.
- Wilson R, Turner G, Laurent G (2004) Transformation of olfactory representations in the drosophila antennal lobe. *Science* 303:366–370.
- Zhang Y, Sharpee T (2016) A robust feedforward model of the olfactory system. *PLoS Comput Biol* 12:e1004850.