

RESEARCH

Open Access



Sparse identification of contrast gain control in the fruit fly photoreceptor and amacrine cell layer

Aurel A. Lazar^{1*}, Nikul H. Ukani¹ and Yiyin Zhou¹

*Correspondence:

aurel@ee.columbia.edu

¹Department of Electrical Engineering, Columbia University, New York, USA

Abstract

The fruit fly's natural visual environment is often characterized by light intensities ranging across several orders of magnitude and by rapidly varying contrast across space and time. Fruit fly photoreceptors robustly transduce and, in conjunction with amacrine cells, process visual scenes and provide the resulting signal to downstream targets. Here, we model the first step of visual processing in the photoreceptor-amacrine cell layer. We propose a novel divisive normalization processor (DNP) for modeling the computation taking place in the photoreceptor-amacrine cell layer. The DNP explicitly models the photoreceptor feedforward and temporal feedback processing paths and the spatio-temporal feedback path of the amacrine cells. We then formally characterize the contrast gain control of the DNP and provide sparse identification algorithms that can efficiently identify each the feedforward and feedback DNP components. The algorithms presented here are the first demonstration of tractable and robust identification of the components of a divisive normalization processor. The sparse identification algorithms can be readily employed in experimental settings, and their effectiveness is demonstrated with several examples.

Keywords: Sparse functional identification; Contrast gain control; Fruit fly; Photoreceptor; Divisive normalization

1 Introduction

Sensory processing systems in the brain extract relevant information from stimuli whose amplitude can vary orders of magnitude [1–4]. Consequently, at each layer of processing, starting right from sensory transduction, neurons need to map their output into a range that can be effectively processed by subsequent neural circuits. As an example, photoreceptors [5–8] and olfactory receptor neurons [9, 10] in both vertebrates and invertebrates respectively adapt to a large range of intensity/temporal contrast values of visual and odorant stimuli. Adaptation to mean and variance of the stimuli has been observed in the auditory system [11, 12] as well. Further down the visual pathway, motion sensitive neurons in vertebrates and invertebrates have been shown to be robust at various brightness and contrast levels [13–15].

© The Author(s) 2020. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Early visual circuits, such as the photoreceptor/amacrine cell layer of the fruit fly brain, are believed to perform spatio-temporal intensity and contrast gain control for dynamic adaptation to visual stimuli whose intensity and contrast levels vary orders of magnitude both in space and time. The mechanism underlying temporal and spatio-temporal contrast gain control in the vertebrate retina [16–20] has often been characterized as a change of the receptive fields in response to changing stimulus statistics [21]. However, while linear receptive fields can be estimated at each transition between different gain values, characterizing the full dynamics of contrast gain control received little attention. Current theoretical methods for describing spatio-temporal gain control lack a systematic framework for characterizing its dynamics, and identification algorithms to estimate circuit components are generally not available.

One exception is the use of theory of dynamical systems for modeling nonlinear gain control. Examples include the Volterra series expansion [22, 23] and the nonlinear autoregressive moving average model with exogenous inputs (NARMAX) [24, 25]. Both formalisms exhibit extensive rigorous tools of functional identification. However, the former typically requires high-order Volterra kernels to model the highly nonlinear gain control, and its identification suffers from the curse of dimensionality. In practice, using the second- or third-order Volterra kernels is computationally not tractable with commodity hardware, while at the same time not fully capturing the dynamics of gain control. For the latter, an extension to the spatio-temporal domain is often out of reach.

Divisive normalization provides an alternative nonlinear operator to model gain control. Divisive normalization [26] has been proposed as a canonical circuit model of computation for many sensory processing circuits underlying adaptation and attention [4, 27–31]. Divisive normalization models in sensory systems are often associated with a population of neurons where each receives inputs from the pool [26, 32]. Recent studies have shown that divisive normalization is a suitable candidate for describing the contrast gain control and its dynamics [33].

Existing modeling studies of divisive normalization in sensory systems mostly focus on establishing a connection between gain control and the statistical properties of natural sensory stimuli [28, 34–36]. There is a lack of general mathematical framework for identifying divisive normalization circuits from recorded data.

Early models of divisive normalization [27] and their derivatives [26, 28, 29, 37–40] only consider feedforward divisive normalization circuits. The modeled neural circuits often exhibit, however, extensive feedback circuits [22, 41].

In this paper, we address the above issues by modeling the photoreceptor/amacrine cells layer of the fruit fly as a multi-input multi-output (MIMO) feedforward and feedback temporal and spatio-temporal divisive normalization processor (DNP). The MIMO DNPs are built upon temporal and spatio-temporal feedforward and feedback of divisive normalization operators constructed from low-dimensional Volterra operators. Combining with sparse identification methods [42], we provide efficient algorithms for identifying all the components of the temporal as well as spatio-temporal divisive normalization processors.

This manuscript is organized as follows. In Sect. 2 the overall architecture of the divisive normalization processor (DNP) is introduced and its power of modeling contrast gain control is demonstrated. We first describe in Sect. 2.1 the biological model of photoreceptor-amacrine cell layer. In Sect. 2.2, we introduce a general model for divisive normalization in the time domain. The temporal DNP consists of the ratio of two nonlinear functionals

acting on the input stimulus. In Sect. 2.3 we then extend the model to space-time domain to include models of lateral feedback from amacrine cells and demonstrate its processing power. In Sects. 3 and 4, we provide identification algorithms and show that the temporal and spatio-temporal DNPs can be efficiently identified. We demonstrate the effectiveness of the algorithms with several examples. We conclude the paper with a discussion in Sect. 5.

2 The architecture of divisive normalization processors

In Sect. 2.1 we start by motivating the present work. We then introduce the architecture of divisive normalization processors in the time domain (Sect. 2.2) and space-time domain (Sect. 2.3). Finally, in Appendix 1 we provide examples that characterize the I/O mapping of the class of temporal and spatio-temporal divisive normalization processors previously described.

2.1 Modeling the photoreceptors and amacrine cells layer

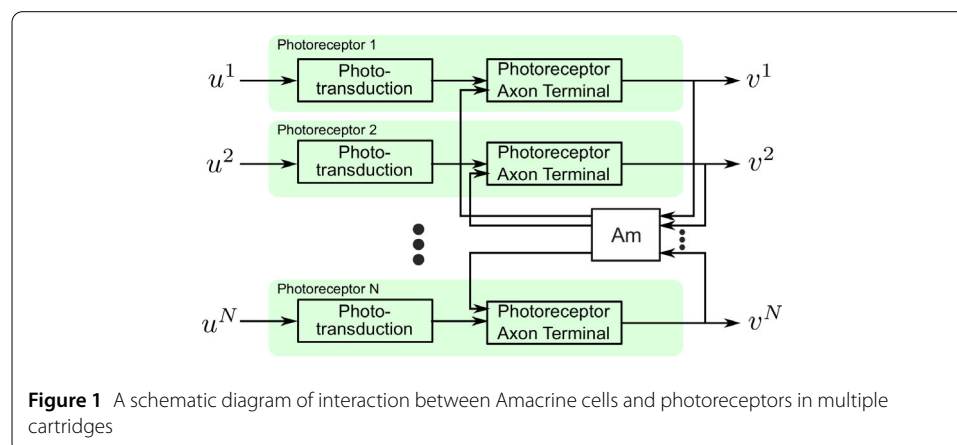
In what follows, we anchor the model description around the photoreceptor-amacrine cell layer of the fruit fly. The fly retina consists of ~ 800 ommatidia, each of which hosts photoreceptors whose axons terminate in a secondary neuropil called lamina. There, they provide inputs to columnar large monopolar cells (LMCs) that project to the third visual neuropil, and to amacrine cells [43]. Amacrine cells are interneurons that innervate axon terminals of multiple photoreceptors. The photoreceptors, in turn, receive lateral feedback from the amacrine cells as well as feedback from LMCs such as L2 neurons [41, 44, 45].

A circuit diagram of the photoreceptor-amacrine cell layer is shown in Fig. 1. For the sake of clarity, we assume here that an ommatidium consists of a single photoreceptor. It has been shown that the outputs of photoreceptors exhibit rapid gain control through both the phototransduction process and the interaction with such feedback loops [25, 46–49].

In what follows we propose a model comprising nonlinear transformations combined with divisive normalization that can model such gain control and can account for diverse dynamics. We will also show that the model we propose can be systematically identified from observed input output pairs.

2.2 Divisive normalization processors in the time domain

In this section we present the modeling of temporal stimuli in 2.2.1 and introduce a class of temporal divisive normalization processors for modeling photoreceptors in Sect. 2.2.2.



2.2.1 Modeling temporal stimuli

We model the temporal varying stimuli $u_1 = u_1(t)$, $t \in \mathbb{D} \subseteq \mathbb{R}$, to be real-valued elements of the space of trigonometric polynomials [50]. The choice of the space of the trigonometric polynomials has, as we will see, substantial computational advantages. A temporal stimulus models the visual field arising at the input of a single photoreceptor.

Definition 1 The space of trigonometric polynomials \mathcal{H}_1 is the Hilbert space of complex-valued functions

$$u(t) = \sum_{l=-L}^L a_l \cdot e_l(t), \quad (1)$$

over the domain $\mathbb{D} = [0, S]$, where

$$e_l(t) = \frac{1}{\sqrt{S}} \exp\left(\frac{j l \Omega}{L} t\right). \quad (2)$$

Here, Ω denotes the bandwidth and L the order of the space. Stimuli $u_1 \in \mathcal{H}_1$ are extended to be periodic over \mathbb{R} with period $S = \frac{2\pi L}{\Omega}$.

\mathcal{H}_1 is a reproducing kernel Hilbert space (RKHS) [51] with reproducing kernel (RK)

$$K_1(t; t') = \sum_{l=-L}^L e_l(t - t'). \quad (3)$$

RKHSs have been previously employed for modeling neural encoding of temporal, auditory, and visual stimuli [52–55].

We denote the dimension of \mathcal{H}_1 by $\dim(\mathcal{H}_1)$ and $\dim(\mathcal{H}_1) = 2L + 1$.

Definition 2 The tensor product space $\mathcal{H}_2 = \mathcal{H}_1 \otimes \mathcal{H}_1$ is an RKHS with reproducing kernel

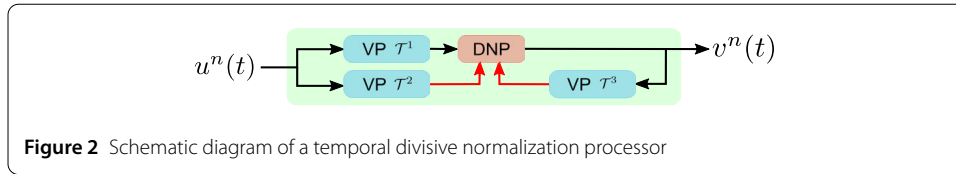
$$K_2(t_1, t_2; t'_1, t'_2) = \sum_{l_1=-L}^L \sum_{l_2=-L}^L e_{l_1}(t_1 - t'_1) \cdot e_{l_2}(t_2 - t'_2). \quad (4)$$

Note that $\dim(\mathcal{H}_2) = \dim(\mathcal{H}_1)^2 = (2L + 1)^2$.

2.2.2 Temporal divisive normalization processors

We first consider single photoreceptors without feedback from the amacrine cells shown in Fig. 1. A schematic of the temporal divisive normalization processor (DNP) modeling the photoreceptor is shown in Fig. 2. For notational simplicity, we consider a single photoreceptor here. The input visual stimulus to the photoreceptor is denoted by $u = u(t)$, $t \in \mathbb{D}$, and the output electric current by $v = v(t)$, $t \in \mathbb{D}$.

Remark 1 Note that, in a single photoreceptor, photons are first absorbed by a large number of microvilli [47] (not shown). Microvilli generate “quantum bumps” in response to photons; the photoreceptor aggregates the bumps and in the process creates the transduction current. Calcium ion influx and calcium diffusion into the photoreceptor cell body



may change the sensitivity of the transduction cascade. A high concentration of calcium (buffer) can result in a photon to be ineffective and may also affect the duration and the magnitude of quantum bumps [56].

The DNP consists of (1) a feedforward Volterra processor (VP) \mathcal{T}^1 , (2) a feedforward normalization VP \mathcal{T}^2 , and (3) a feedback normalization VP \mathcal{T}^3 . The output of the photoreceptor amounts to

$$v^n = \frac{\mathcal{T}^1 u^n}{\mathcal{T}^2 u^n + \mathcal{T}^3 v^n}, \quad n = 1, 2, \dots, N, \tag{5}$$

where

$$\begin{aligned} (\mathcal{T}^l u^n)(t) &= b^l + \int_{\mathbb{D}} h_1^l(s) u^n(t-s) ds \\ &+ \int_{\mathbb{D}^2} h_2^l(s_1, s_2) u^n(t-s_1) u^n(t-s_2) ds_1 ds_2, \quad l = 1, 2, \end{aligned} \tag{6}$$

and

$$(\mathcal{T}^3 v^n)(t) = b^3 + \int_{\mathbb{D}} h_1^3(s) v^n(t-s) ds + \int_{\mathbb{D}^2} h_2^3(s_1, s_2) v^n(t-s_1) v^n(t-s_2) ds_1 ds_2. \tag{7}$$

Here, b^l , $l = 1, 2, 3$, are the zeroth-order Volterra kernels (constants), $h_1^l(t)$, $l = 1, 2, 3$, are first-order Volterra kernels (impulse responses of linear filters), and $h_2^l(t, s)$, $l = 1, 2, 3$, are second-order Volterra kernels. As before, \mathbb{D} denotes the domain of the input space, and $\mathbb{D}^2 = \mathbb{D} \times \mathbb{D}$.

Remark 2 For the sake of tractability, we limit each nonlinear functional in (6) and (7) to be composed of only first- and second-order Volterra kernels. The division in (5) allows us, however, to model nonlinear processing of much higher orders.

We note that v in (5) is invariant under scaling by the same factor of the numerator and denominator. Hence, without loss of generality, we will assume $b^2 + b^3 = 1$. We will also assume that the DNP is bounded-input bounded-output [57].

2.2.3 Modeling temporal DNP feedback filters

Here, we define the filter kernels in equations (6) and (7).

Definition 3 Let $h_p^l \in \mathbb{L}^1(\mathbb{D}^p)$, $l = 1, 2$, $p = 1, 2$, where \mathbb{L}^1 denotes the space of Lebesgue integrable functions. The operator $\mathcal{P}_1 : \mathbb{L}_1(\mathbb{D}) \rightarrow \mathcal{H}_1$ given by

$$(\mathcal{P}_1 h_1^l)(t) = \int_{\mathbb{D}} h_1^l(t') K_1(t; t') dt' \tag{8}$$

is called the projection operator from $\mathbb{L}^1(\mathbb{D})$ to \mathcal{H}_1 . Similarly, the operator $\mathcal{P}_2 : \mathbb{L}_1(\mathbb{D}^2) \rightarrow \mathcal{H}_2$ given by

$$(\mathcal{P}_2 h_2^l)(t_1; t_2) = \int_{\mathbb{D}^2} h_2^l(t'_1; t'_2) K_2(t_1, t_2; t'_1, t'_2) dt'_1 dt'_2 \tag{9}$$

is called the projection operator from $\mathbb{L}^1(\mathbb{D}^2)$ to \mathcal{H}_2 .

Note that, for $u^n \in \mathcal{H}_1, \mathcal{P}_1 u^n = u^n$. Moreover, with $u_2^n(t_1, t_2) = u^n(t_1)u^n(t_2), \mathcal{P}_2 u_2^n = u_2^n$. Thus,

$$\begin{aligned} (\mathcal{T}^l u^n)(t) &= b^l + \int_{\mathbb{D}} (\mathcal{P}_1 h_1^l)(s) u^n(t-s) ds \\ &+ \int_{\mathbb{D}^2} (\mathcal{P}_2 h_2^l)(s_1, s_2) u^n(t-s_1) u^n(t-s_2) ds_1 ds_2, \quad l = 1, 2, \end{aligned} \tag{10}$$

and by assuming that $h_1^l \in \mathcal{H}_1, l = 1, 2$, and $h_2^l \in \mathcal{H}_2$ we recover the simple form of equation (6).

We model the output waveforms $v^o = v^o(t), t \in \mathbb{D} \subseteq \mathbb{R}$, to be real-valued elements of the space of trigonometric polynomials [50].

Definition 4 The space of trigonometric polynomials \mathcal{H}_1^o is the Hilbert space of complex-valued functions

$$v^o(t) = \sum_{l=-L^o}^{L^o} a_l^o \cdot e_l^o(t), \tag{11}$$

over the domain $\mathbb{D} = [0, S]$, where

$$e_l^o(t) = \frac{1}{\sqrt{S^o}} \exp\left(\frac{j l \Omega^o}{L^o} t\right). \tag{12}$$

Here, Ω^o denotes the bandwidth and L^o the order of the space. The output waveforms $v^o \in \mathcal{H}_1^o$ are extended to be periodic over \mathbb{R} with period $S^o = \frac{2\pi L^o}{\Omega^o}$.

\mathcal{H}_1^o is a reproducing kernel Hilbert space (RKHS) [51] with reproducing kernel (RK)

$$K_1^o(t; t') = \sum_{l=-L^o}^{L^o} e_l^o(t-t'). \tag{13}$$

We denote the dimension of \mathcal{H}_1^o by $\dim(\mathcal{H}_1^o)$ and $\dim(\mathcal{H}_1^o) = 2L^o + 1$.

Definition 5 The tensor product space $\mathcal{H}_2^o = \mathcal{H}_1^o \otimes \mathcal{H}_1^o$ is an RKHS with reproducing kernel

$$K_2^o(t_1, t_2; t'_1, t'_2) = \sum_{l_1=-L^o}^{L^o} \sum_{l_2=-L^o}^{L^o} e_{l_1}^o(t_1-t'_1) \cdot e_{l_2}^o(t_2-t'_2). \tag{14}$$

Note that $\dim(\mathcal{H}_2^o) = \dim(\mathcal{H}_1^o)^2 = (2L^o + 1)^2$.

Definition 6 Let $h_p^3 \in \mathbb{L}^1(\mathbb{D}^p)$, $p = 1, 2$, where \mathbb{L}^1 denotes the space of Lebesgue integrable functions. The operator $\mathcal{P}_1^o : \mathbb{L}_1(\mathbb{D}) \rightarrow \mathcal{H}_1^o$ given by

$$(\mathcal{P}_1^o h_1^3)(t) = \int_{\mathbb{D}} h_1^3(t') K_1^o(t; t') dt' \quad (15)$$

is called the projection operator from $\mathbb{L}^1(\mathbb{D})$ to \mathcal{H}_1 . Similarly, the operator $\mathcal{P}_2^o : \mathbb{L}_1(\mathbb{D}^2) \rightarrow \mathcal{H}_2^o$ given by

$$(\mathcal{P}_2^o h_2^3)(t_1; t_2) = \int_{\mathbb{D}^2} h_2^3(t'_1; t'_2) K_2^o(t_1, t_2; t'_1, t'_2) dt'_1 dt'_2 \quad (16)$$

is called the projection operator from $\mathbb{L}^1(\mathbb{D}^2)$ to \mathcal{H}_2^o .

We note that

$$\begin{aligned} (\mathcal{T}^3 v^n)(t) &= b^3 + \int_{\mathbb{D}} (\mathcal{P}_1^o h_1^3)(s) v^n(t-s) ds \\ &\quad + \int_{\mathbb{D}^2} (\mathcal{P}_2^o h_2^3)(s_1, s_2) v^n(t-s_1) v^n(t-s_2) ds_1 ds_2. \end{aligned} \quad (17)$$

If we now assume that $h_1^3 \in \mathcal{H}_1^o$ and $h_2^3 \in \mathcal{H}_2^o$, then $\mathcal{P}_1^o h_1^3 = h_1^3$ and $\mathcal{P}_2^o h_2^3 = h_2^3$, respectively, and the above equation is identical with equation (7) above.

2.3 Divisive normalization processors in the space-time domain

In Sect. 2.2, we described a temporal divisive normalization processor model. The normalization term was the sum of a processed version of the input and the output. However, many biological circuits are thought to exhibit lateral inhibition and gain control [26, 58–61]. An example is provided by the photoreceptor-amacrine cell layer shown in Fig. 1. In Fig. 3, we provide a model of the schematic in Fig. 1. This model is a ‘circuit’ extension of the one shown in Fig. 2 and accounts explicitly for lateral inhibition. We anchor the extension in the visual system where the input domain is spatio-temporal.

In Sect. 2.3.1 we model spatio-temporal stimuli and in Sect. 2.3.2 the spatio-temporal divisive normalization processors.

2.3.1 Modeling spatio-temporal stimuli

In this section we provide a model of the interaction between a group of photoreceptors and an amacrine cell. In each photoreceptor, the phototransduction process converts light into current and excites the membrane of the photoreceptor. The voltage signal is then propagated through its axon to the lamina. While photoreceptors provide inputs to amacrine cells, their axon terminals also receive amacrine cell input. Since an amacrine cell innervates multiple lamina cartridges, it provides spatial feedback to several photoreceptors in a small neighborhood.

We extend the temporal divisive normalization processor depicted in Fig. 2 to process spatio-temporal stimuli as shown in Fig. 3. Each spatially sampled point, or pixel, denoted as $u^i(t)$, $i = 1, 2, \dots, N$, is first processed by a temporal DNP. For simplicity rather than picking different Volterra kernels for each branch, \mathcal{T}^1 , \mathcal{T}^2 , and \mathcal{T}^3 are shared across branches.

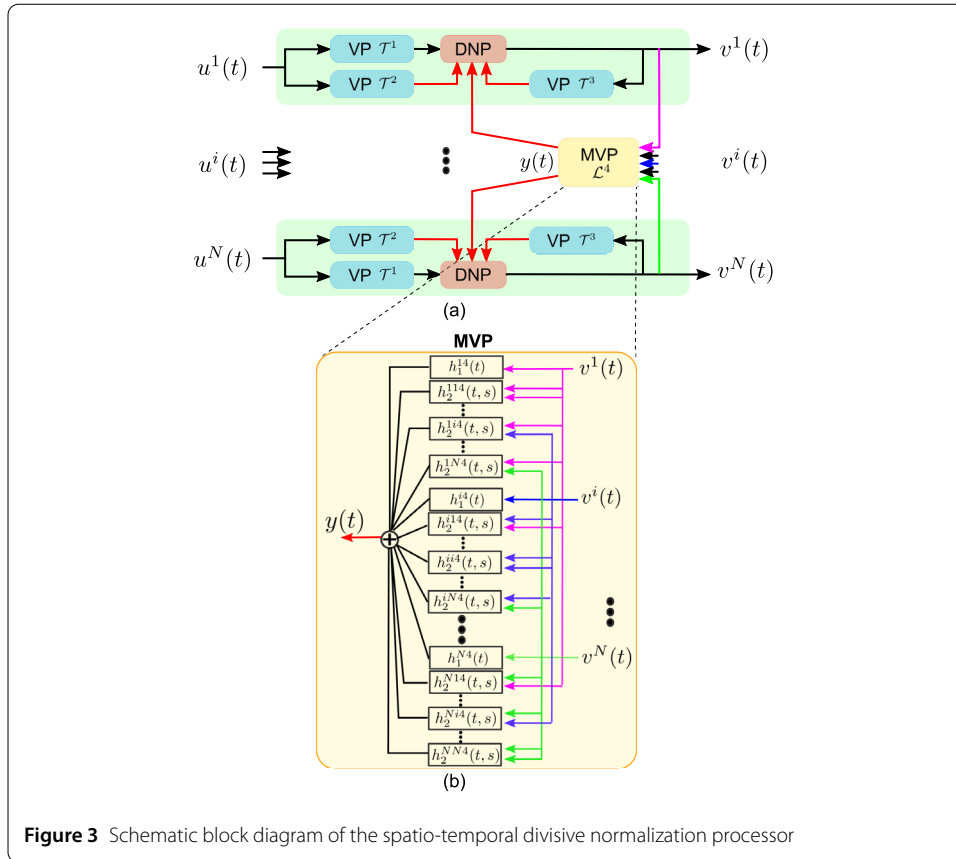


Figure 3 Schematic block diagram of the spatio-temporal divisive normalization processor

In addition, we introduce a multi-input Volterra processor (MVP) to model the spatio-temporal feedback due to the amacrine cell. Each of the branches in Fig. 3, without the input of the MVP block, is equivalent to the model in Fig. 2.

2.3.2 Spatio-temporal divisive normalization processors

As shown in Fig. 3(b), MVPs are comprised of second-order filters acting on DNP output pairs in addition to linear filters independently acting on each DNP output. Thus, the inputs to the MVP are the DNP outputs $v^i(t), i = 1, 2, \dots, N$, and the MVP output amounts to

$$\begin{aligned}
 (\mathcal{L}^4 \mathbf{v})(t) = & b^4 + \sum_{i=1}^N \left(\int_{\mathbb{D}} h_1^{i4}(s) v^i(t-s) ds \right) \\
 & + \sum_{i=1}^N \sum_{j=1}^N \left(\int_{\mathbb{D}^2} h_2^{ij4}(s_1, s_2) v^i(t-s_1) v^j(t-s_2) ds_1 ds_2 \right), \tag{18}
 \end{aligned}$$

where

$$\mathbf{v}(t) = [v^1(t), v^2(t), \dots, v^N(t)]^T, \tag{19}$$

b^4 is the zeros-th-order Volterra kernel (constant). Furthermore, $h_1^{i4} \in \mathcal{H}_1^0, i = 1, 2, \dots, N$, are the first-order Volterra kernels whose inputs are $v^i, i = 1, 2, \dots, N$, respectively, and

$h_2^{ij4} \in \mathcal{H}_2^o$, $i, j = 1, 2, \dots, N$, are the second-order Volterra kernels whose inputs are the pairs (v^i, v^j) , $i, j = 1, 2, \dots, N$, respectively.

The full model in Fig. 3 thus consists of parallel channels as depicted in Fig. 2 with the added cross-channel feedback normalization/gain control provided by the MVP block.

The overall output of the spatio-temporal DNP can be expressed as

$$v^n(t) = \frac{\mathcal{T}^1 u^n}{\mathcal{T}^2 u^n + \mathcal{T}^3 v^n + \mathcal{L}^4 \mathbf{v}}, \quad n = 1, 2, \dots, N. \quad (20)$$

W.L.O.G., we assume that $b^2 + b^3 + b^4 = 1$.

2.3.3 Spatio-temporal DNPs and contrast gain control

The relationship and some intuition behind the modeling power of spatio-temporal DNPs are provided in several examples in Appendix 1. The I/O of three simple spatio-temporal DNPs stimulated with different inputs is briefly mentioned here. In the first example, we evaluated the response of a DNP with four photoreceptors under different background light intensity levels. In Fig. 9 one of the photoreceptors is subjected to an additional flash of different light intensity, while the inputs of the other three are kept at the same background level. The steady state response of the photoreceptor that receives the additional flash is shifted as a function of the background-intensity level. In the second example, contrast gain control exerted by the amacrine cells is demonstrated for the same DNP in Fig. 10. The effect of the MVP block on the RMS contrast can be clearly seen in Fig. 10 and is quantitatively evaluated in Fig. 11. Finally, an example of steady state I/O visualization of a natural image with resolution of 1536×1024 pixels at low, medium, and high luminance values is shown in Fig. 13. The image was divided into 16×16 spatio-temporal DNP blocks with a four-pixel overlap in each direction.

3 Sparse identification of temporal DNPs

In what follows we derive sparse identification algorithms for the components of spatio-temporal DNPs depicted in Fig. 2 and formally defined in equation (5). In what follows, we assume that during experimental trials, the single isolated photoreceptor n is presented with M test stimuli $u^{mm} = u^{mm}(t)$, $m = 1, 2, \dots, M$, and for each trial the outputs $v^{mm} = v^{mm}(t)$, $m = 1, 2, \dots, M$, are recorded. The objective is to identify the model components b_1, h_1^l, h_2^l , $l = 1, 2, 3$, from the knowledge of the inputs and outputs.

3.1 Deriving the sparse identification algorithm for temporal DNPs

Lemma 1 *With M input stimuli presented to a temporal DNP, let the inputs $u^{mm}(t)$ and the outputs $v^{mm}(t)$ be sampled at times (t_k) , $k = 1, 2, \dots, T$. Then*

$$\begin{aligned} b^1 + \langle h_1^1, \phi_{11}^{nmk} \rangle_{\mathcal{H}_1} + \langle h_1^2, \phi_{12}^{nmk} \rangle_{\mathcal{H}_1} + \langle h_1^3, \phi_{13}^{nmk} \rangle_{\mathcal{H}_1^o} \\ + \langle h_2^1, \phi_{21}^{nmk} \rangle_{\mathcal{H}_2} + \langle h_2^2, \phi_{22}^{nmk} \rangle_{\mathcal{H}_2} + \langle h_2^3, \phi_{23}^{nmk} \rangle_{\mathcal{H}_2^o} = q^{nmk}, \end{aligned} \quad (21)$$

where the sampling functions $\phi_{11}^{nmk} \in \mathcal{H}_1, \phi_{12}^{nmk} \in \mathcal{H}_1, \phi_{13}^{nmk} \in \mathcal{H}_1^o, \phi_{21}^{nmk} \in \mathcal{H}_2, \phi_{22}^{nmk} \in \mathcal{H}_2,$ and $\phi_{23}^{nmk} \in \mathcal{H}_2^o$ are given by

$$\begin{aligned} \phi_{11}^{nmk}(t) &= u^{nm}(t_k - t), \\ \phi_{12}^{nmk}(t) &= -q^{nm}(t_k)u^{nm}(t_k - t), \\ \phi_{13}^{nmk}(t) &= -q^{nm}(t_k)(\mathcal{P}_1^o v^{nm})(t_k - t), \\ \phi_{21}^{nmk}(t, s) &= u^{nm}(t_k - t)u^{nm}(t_k - s), \\ \phi_{22}^{nmk}(t, s) &= -q^{nm}(t_k)u^{nm}(t_k - t)u^{nm}(t_k - s), \\ \phi_{23}^{nmk}(t, s) &= -q^{nm}(t_k)(\mathcal{P}_1^o v^{nm})(t_k - t)(\mathcal{P}_1^o v_1^{nm})(t_k - s), \end{aligned} \tag{22}$$

and $q^{nmk} = v^{nm}(t_k)$ for all $m = 1, 2, \dots, M$ and $k = 1, 2, \dots, T$.

Proof See Appendix 2. □

Remark 3 From (21), it can be seen that the identification of the divisive normalization processor has been reformulated as a generalized sampling problem [62] in $\mathbb{R} \oplus \mathcal{H}_1 \oplus \mathcal{H}_1 \oplus \mathcal{H}_1^o \oplus \mathcal{H}_2 \oplus \mathcal{H}_2 \oplus \mathcal{H}_2^o$. Subsequently, the divisive normalization model can be identified by solving a system of linear equations.

In order to solve the system of linear equations in (21), we rewrite them first in matrix form.

Lemma 2 Equation (21) can be expressed in matrix form as follows:

$$\mathbf{c}_1^\top \Phi^{nmk} + \text{Tr}(\mathbf{C}_2^\text{H} \Xi^{nmk}) = q^{nmk} \tag{23}$$

for all $m = 1, 2, \dots, M$ and $k = 1, 2, \dots, T$, where $\text{Tr}(\cdot)$ denotes the trace operator, and

$$\text{Measurements} \leftarrow q^{nmk}, (\text{scalar}) \tag{24}$$

$$\text{Unknowns} \leftarrow \begin{cases} \mathbf{c}_1 = \begin{bmatrix} b^1 & (\mathbf{h}_1^1)^\top & (\mathbf{h}_1^2)^\top & (\mathbf{h}_1^3)^\top \end{bmatrix}_{(4L+2L^o+4) \times 1}^\top, \\ \mathbf{C}_2 = \begin{bmatrix} \mathbf{H}_2^1 & \mathbf{0}_{(2L+1) \times (2L^o+1)} \\ \mathbf{H}_2^2 & \mathbf{0}_{(2L+1) \times (2L^o+1)} \\ \mathbf{0}_{(2L^o+1) \times (2L+1)} & \mathbf{H}_2^3 \end{bmatrix}_{(4L+2L^o+3) \times (2L+2L^o+2)} \end{cases}, \tag{25}$$

Sampling vectors

$$\leftarrow \begin{cases} \Phi^{nmk} = \begin{bmatrix} 1 \\ \mathbf{u}^{nmk} \\ -q^{nmk} \mathbf{u}^{nmk} \\ -q^{nmk} \mathbf{v}^{nmk} \end{bmatrix}_{(4L+2L^o+4) \times 1}, \\ \Xi^{nmk} = \begin{bmatrix} \mathbf{U}_2^{nmk} & \mathbf{0}_{(2L+1) \times (2L^o+1)} \\ -q^{nmk} \mathbf{U}_2^{nmk} & \mathbf{0}_{(2L+1) \times (2L^o+1)} \\ \mathbf{0}_{(2L^o+1) \times (2L+1)} & -q^{nmk} \mathbf{V}_2^{nmk} \end{bmatrix}_{(4L+2L^o+3) \times (2L+2L^o+2)}. \end{cases} \tag{26}$$

Proof See Appendix 3 for more notation and detailed proof. □

A necessary condition on the number of trials and the number of measurements required for identifying the divisive normalization processor for solving the system of equations in Theorem 1 is that the number of trials $M \geq 3 + 2 \cdot \dim(\mathcal{H}_1)$ and the number of total samples $TM \geq 1 + 2 \cdot \dim(\mathcal{H}_1) + 2 \cdot \dim(\mathcal{H}_1)^2$.

It is easy to see that solving the system of equations above suffers from the curse of dimensionality. As the dimension of \mathcal{H}_1 increases, the number of samples needed to identify components increases quadratically. Note that the second-order Volterra kernels h_2^l , $l = 1, 2, 3$, have unique symmetric forms with orthogonal expansions as follows [57]:

$$h_2^l(t_1, t_2) = \sum_{k \in \mathbb{N}} \lambda^{kl} g_1^{kl}(t_1) g_1^{kl}(t_2), \quad \|g_1^{kl}\| = 1, \tag{27}$$

where $g_1^{kl} \in \mathcal{H}_1, k \in \mathbb{N}$, are orthogonal to each other. In what follows, we assume that the second-order Volterra kernels are sparse, *i.e.*, $\lambda^{kl} = 0$ for $k > r_l$, where $r_l \ll \dim(\mathcal{H}_1)$. Sparse kernels often arise in modeling sensory processing, *e.g.*, in complex cells in the primary visual cortex [42]. By exploiting the sparse structure of the second-order kernels, the identification problem can be made tractable.

The sparsity of the kernels can be translated into a low-rank condition on the matrix representation of $h_2^l, l = 1, 2, 3$ (see also Appendix 3). Ideally, the optimization problem would be a rank minimization problem. But rank minimization being NP-hard, we use the surrogate of nuclear norm minimization instead, which is the convex envelope of the rank operator [63].

To perform sparse identification of the divisive normalization processor, we devised Algorithm 1. By optimizing over \mathbf{c}_1 and \mathbf{C}_2 and subsequently assigning the corresponding block entries according to (25), Algorithm 1 identifies $b_1, \mathbf{h}_1^i, i = 1, 2, 3$, and $\mathbf{H}_2^i, i = 1, 2, 3$.

As a surrogate of rank minimization, Algorithm 1 minimizes a linear combination of the nuclear norm of \mathbf{C}_2 and the Euclidean norm of \mathbf{c}_1 . The optimization constraints correspond (i) in (29) to the generalized sampling problem allowing certain amount of error, (ii) in (30) to zero mean slack variables, (iii) in (31) to the zeros in the two blocks in the top-right of \mathbf{C}_2 , (iv) in (32) to the zeros in the block in the lower-left of \mathbf{C}_2 , and (v) in (33), (34), and (35), respectively, $\mathbf{H}_2^1, \mathbf{H}_2^2$, and \mathbf{H}_2^3 are Hermitian.

Algorithm 1 $\hat{\mathbf{c}}_1$ and $\hat{\mathbf{C}}_2$ are the solution to the following optimization problem:

$$\underset{\mathbf{c}_1, \mathbf{C}_2, \boldsymbol{\varepsilon}}{\text{minimize}} \quad \|\mathbf{C}_2\|_* + \lambda_1 \|\mathbf{c}_1\|_2 + \lambda_2 \|\boldsymbol{\varepsilon}\|_2 \tag{28}$$

$$\text{s.t.} \quad \mathbf{c}_1^\top \boldsymbol{\Phi}^{nmk} + \text{Tr}(\mathbf{C}_2^H \boldsymbol{\Xi}^{nmk}) = q^{nmk} + \varepsilon^{(m-1)*T+k}, \tag{29}$$

$$m = 1, \dots, M, k = 1, \dots, T, \tag{29}$$

$$\mathbf{1}^\top \boldsymbol{\varepsilon} = \mathbf{0} \tag{30}$$

$$\begin{bmatrix} \mathbf{I}_{2 \dim(\mathcal{H}_1)} & \mathbf{0} \end{bmatrix} \mathbf{C}_2 \begin{bmatrix} \mathbf{0} \\ \mathbf{I}_{\dim(\mathcal{H}_1^o)} \end{bmatrix} = \mathbf{0}, \tag{31}$$

$$\begin{bmatrix} \mathbf{0} & \mathbf{I}_{\dim(\mathcal{H}_1^o)} \end{bmatrix} \mathbf{C}_2 \begin{bmatrix} \mathbf{I}_{\dim(\mathcal{H}_1)} \\ \mathbf{0} \end{bmatrix} = \mathbf{0}, \tag{32}$$

$$\begin{bmatrix} \mathbf{I}_{\dim(\mathcal{H}_1)} & \mathbf{0} \end{bmatrix} \mathbf{C}_2 \begin{bmatrix} \mathbf{I}_{\dim(\mathcal{H}_1)} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{\dim(\mathcal{H}_1)} \\ \mathbf{0} \end{bmatrix}^\top \mathbf{C}_2^H \begin{bmatrix} \mathbf{I}_{\dim(\mathcal{H}_1)} & \mathbf{0} \end{bmatrix}^\top, \quad (33)$$

$$\begin{aligned} & \begin{bmatrix} \mathbf{0}_{\dim(\mathcal{H}_1)} & \mathbf{I}_{\dim(\mathcal{H}_1)} & \mathbf{0} \end{bmatrix} \mathbf{C}_2 \begin{bmatrix} \mathbf{I}_{\dim(\mathcal{H}_1)} \\ \mathbf{0} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{I}_{\dim(\mathcal{H}_1)} \\ \mathbf{0} \end{bmatrix}^\top \mathbf{C}_2^H \begin{bmatrix} \mathbf{0}_{\dim(\mathcal{H}_1)} & \mathbf{I}_{\dim(\mathcal{H}_1)} & \mathbf{0} \end{bmatrix}^\top, \end{aligned} \quad (34)$$

$$\begin{bmatrix} \mathbf{0} & \mathbf{I}_{\dim(\mathcal{H}_1^o)} \end{bmatrix} \mathbf{C}_2 \begin{bmatrix} \mathbf{0} \\ \mathbf{I}_{\dim(\mathcal{H}_1^o)} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{I}_{\dim(\mathcal{H}_1^o)} \end{bmatrix}^\top \mathbf{C}_2^H \begin{bmatrix} \mathbf{0} & \mathbf{I}_{\dim(\mathcal{H}_1^o)} \end{bmatrix}^\top, \quad (35)$$

where $\|\cdot\|_*$ denotes the nuclear norm defined as $\|\mathbf{C}_2\|_* = \text{Tr}((\mathbf{C}_2^H \mathbf{C}_2)^{\frac{1}{2}})$, λ_1, λ_2 are appropriately chosen hyperparameters, $\varepsilon^i \in \mathbb{R}$, $i = 1, 2, \dots, MT$, represent slack variables, $\mathbf{1}$ represents a vector of all ones, \mathbf{I}_p represents a $p \times p$ identity matrix, $\mathbf{0}_p$ represents a $p \times p$ matrix of all zeros, $\mathbf{0}_{p \times q}$ represents a $p \times q$ matrix of all zeros, and $\mathbf{0}$ represents a matrix of all zeros with dimensions that make the equation consistent.

Theorem 1 *The filters of the DNP are identified as $\hat{b}^1 = \hat{b}^1$ and*

$$\hat{h}_1^1(t) = \sum_{l=-L}^L [\hat{\mathbf{h}}_1^1]_{l+L+1} \cdot e_l(t), \quad (36)$$

$$\hat{h}_2^1(t_1, t_2) = \sum_{l_1=-L}^L \sum_{l_2=-L}^L [\hat{\mathbf{H}}_2^1]_{l_1+L+1, l_2+L+1} \cdot e_{l_1}(t_1) e_{l_2}(t_2), \quad (37)$$

$$\hat{h}_1^2(t) = \sum_{l=-L}^L [\hat{\mathbf{h}}_1^2]_{l+L+1} \cdot e_l(t), \quad (38)$$

$$\hat{h}_2^2(t_1, t_2) = \sum_{l_1=-L}^L \sum_{l_2=-L}^L [\hat{\mathbf{H}}_2^2]_{l_1+L+1, l_2+L+1} \cdot e_{l_1}(t_1) e_{l_2}(t_2), \quad (39)$$

$$\hat{h}_1^3(t) = \sum_{l=-L^o}^{L^o} [\hat{\mathbf{h}}_1^3]_{l+L+1} \cdot e_l^o(t), \quad (40)$$

$$\hat{h}_2^3(t_1, t_2) = \sum_{l_1=-L^o}^{L^o} \sum_{l_2=-L^o}^{L^o} [\hat{\mathbf{H}}_2^3]_{l_1+L^o+1, l_2+L^o+1} \cdot e_{l_1}^o(t_1) e_{l_2}^o(t_2), \quad (41)$$

where

$$\begin{bmatrix} \hat{b}^1 & \hat{\mathbf{h}}_1^1 \top & \hat{\mathbf{h}}_1^2 \top & \hat{\mathbf{h}}_1^3 \top \end{bmatrix}^\top = \hat{\mathbf{c}}_1 \quad \text{and} \quad \begin{bmatrix} \hat{\mathbf{H}}_2^1 & - \\ \hat{\mathbf{H}}_2^2 & - \\ - & \hat{\mathbf{H}}_2^3 \end{bmatrix} = \hat{\mathbf{C}}_2. \quad (42)$$

Remark 4 By exploiting the structure of low-rank second-order Volterra kernels, Algorithm 1 provides a tractable solution to the identification of the components of the divisive normalization processor.

3.2 Examples of sparse identification of temporal DNPs

We provide here identification examples solved using Algorithm 1.

Example 1 Here, we identify a temporal divisive normalization processor in Fig. 2, where

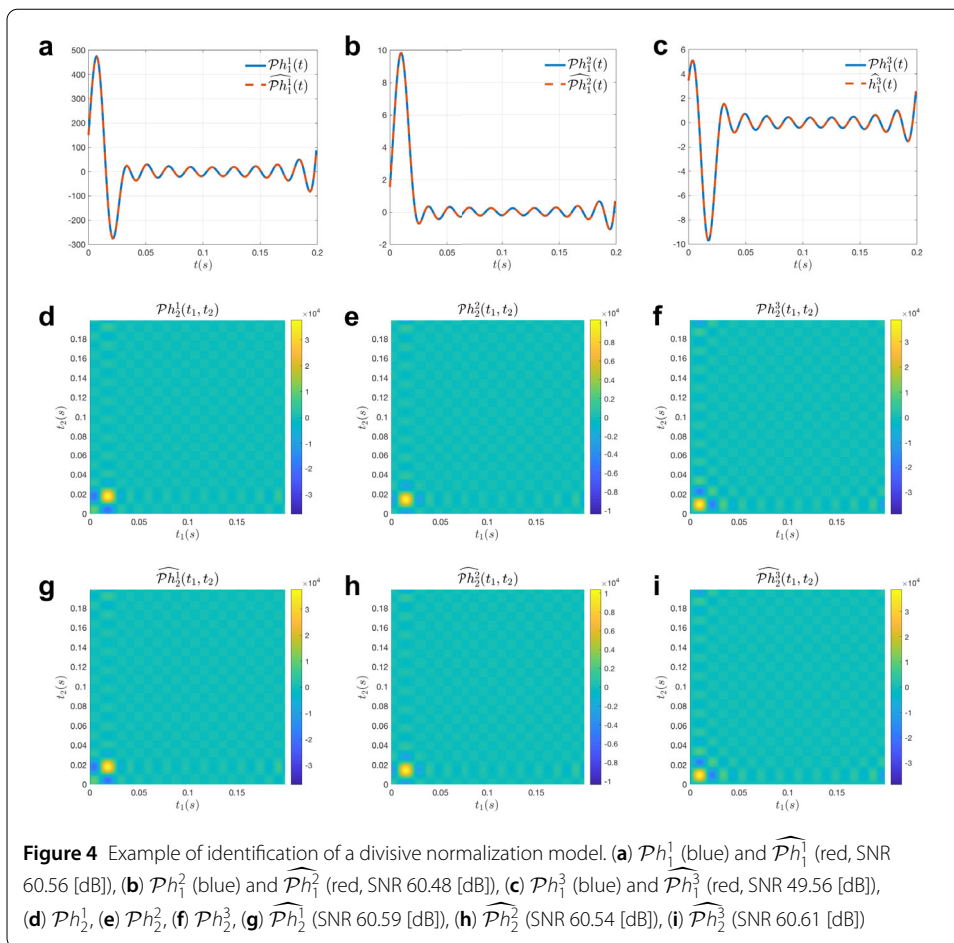
$$\begin{aligned}
 h_1^1(t) &= 2.472 \times 10^{10} t^3 e^{-100\pi t} \cos(36\pi t), \\
 h_1^2(t) &= 3.117 \times 10^8 t^3 e^{-100\pi t} \cos(20\pi t), \\
 h_1^3(t) &= 4.753 \times 10^8 t^3 e^{-100\pi t} \cos(52\pi t), \\
 h_2^1(t, s) &= 9.038 \times 10^{19} t^3 s^3 e^{-100\pi(t+s)} \cos(52\pi t) \cos(52\pi s) \\
 &\quad + 5.3467 \times 10^{14} t^3 s^3 e^{-100\pi(t+s)} \cos(100\pi t) \cos(100\pi s), \\
 h_2^2(t, s) &= 1.533 \times 10^{19} t^3 s^3 e^{-100\pi(t+s)} \cos(68\pi t) \cos(68\pi s) \\
 &\quad + 5.970 \times 10^{14} t^3 s^3 e^{-100\pi(t+s)} \cos(84\pi t) \cos(84\pi s), \\
 h_2^3(t, s) &= 6.771 \times 10^{19} t^3 s^3 e^{-100\pi(t+s)} \cos(100\pi t) \cos(100\pi s) \\
 &\quad + 5.970 \times 10^{16} t^3 s^3 e^{-100\pi(t+s)} \cos(84\pi t) \cos(84\pi s).
 \end{aligned}$$

We choose the input space \mathcal{H}_1 to have $L = 10$, $\Omega = 100\pi$. Thus $S = 0.2s$, $\dim(\mathcal{H}_1) = 21$, and $\dim(\mathcal{H}_2) = 441$. Note that all three second-order Volterra kernels exhibit low-rank structure. We presented the model with 25 stimuli from \mathcal{H}_1 , whose coefficients were chosen to be i.i.d Gaussian variables. Then, a total of 425 measurements were used from the input and the observed output pairs to solve the identification problem using Algorithm 1. The results of the identification are shown in Fig. 4. As can be seen from the figure, Algorithm 1 was able to identify the model with high precision using only 425 measurements, much less than the 1387 measurements that would have been required to solve the generalized sampling problem directly. The factor of reduction in the required measurements is critical when the model needs to be identified in a much larger space, for example, a space of spatio-temporal stimuli as shown in the next example.

Example 2 Here, we identify a detailed biophysical model of *Drosophila* photoreceptors as a DNP with feedforward only (see Fig. 2). The detailed biophysical model consists of 30,000 microvilli [5, 64]. The photon absorption in each microvillus is described by a Poisson process whose rate is proportional to the number of photons per microvillus incident on the ommatidium. Photon absorption leads to a transduction process governed by a cascade of chemical reactions. The entire transduction process is described by 13 successive differential equations and is given in [64]. The total number of equations of the photoreceptor model is 390,000.

For identifying the DNP, we used natural images from the van Hateren database [65] and simulated movements of a fly superimposed on the natural scenes. Each visual scene was captured by a previously developed retina model [64], endowed with realistic optics and the geometry of the fly retina. A 200-second stimulus was generated. We then divided the stimulus into several segments, each of which scaled to a different level of mean light intensity. The resulting stimulus is shown in Fig. 5(a). The bandwidth of the DNP input and output spaces was limited to 50 Hz.

The visual input was presented to the DNP photoreceptor model. We used 10 seconds of the stimulus for identifying the DNP filters. The identified filters are depicted in Fig. 6. The output of the photoreceptor DNP model when stimulated by the other 190 seconds of



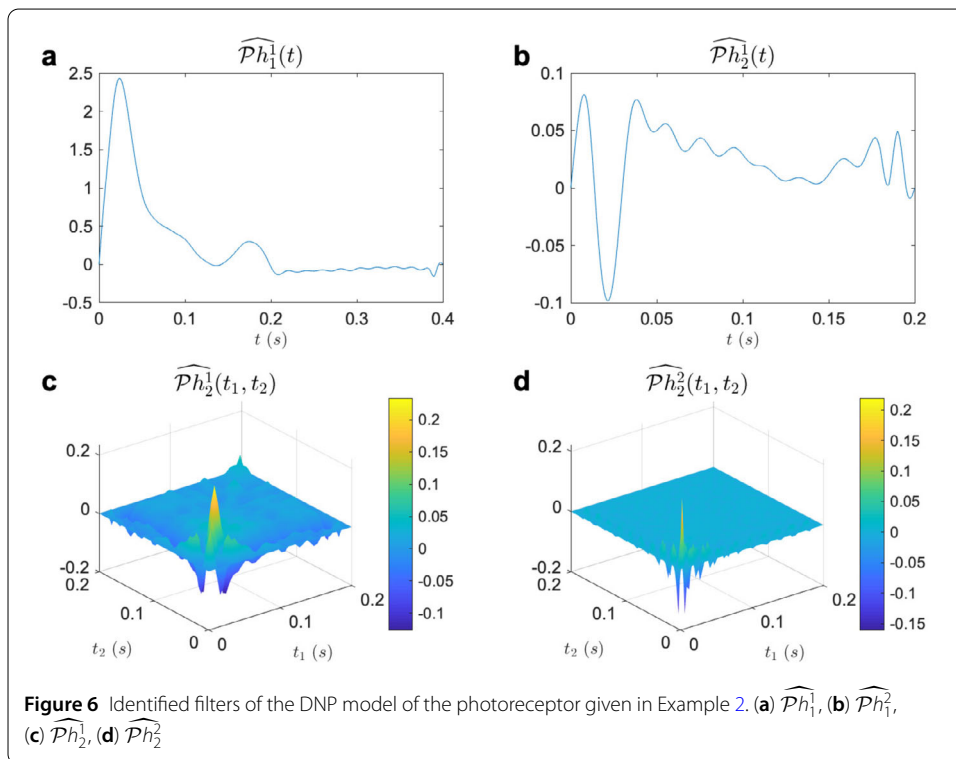
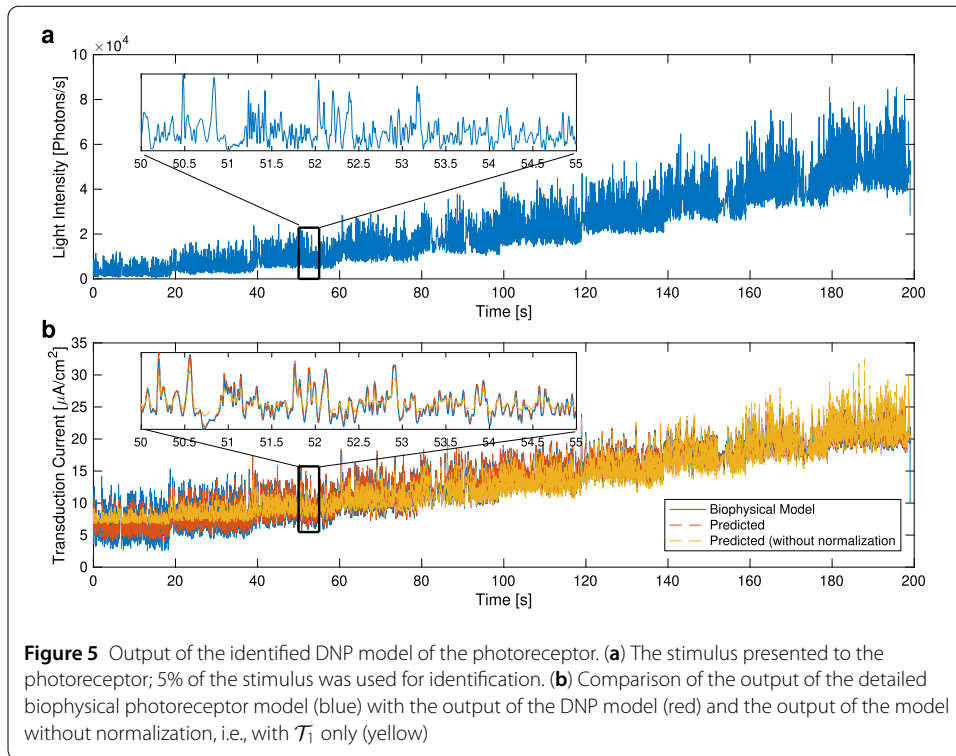
the input stimulus is shown in Fig. 5(b). We evaluated the identified photoreceptor DNP model by computing the SNR between the output of the detailed biophysical model and that of the identified DNP model. The SNR was 26.14 [dB].

We additionally trained a model without normalization (*i.e.*, with only \mathcal{T}^1). As shown in Fig. 5(b) the identified model without normalization does not match the output well across different light intensities. Compared with the output of the DNP model, the SNR was only 14.48 [dB].

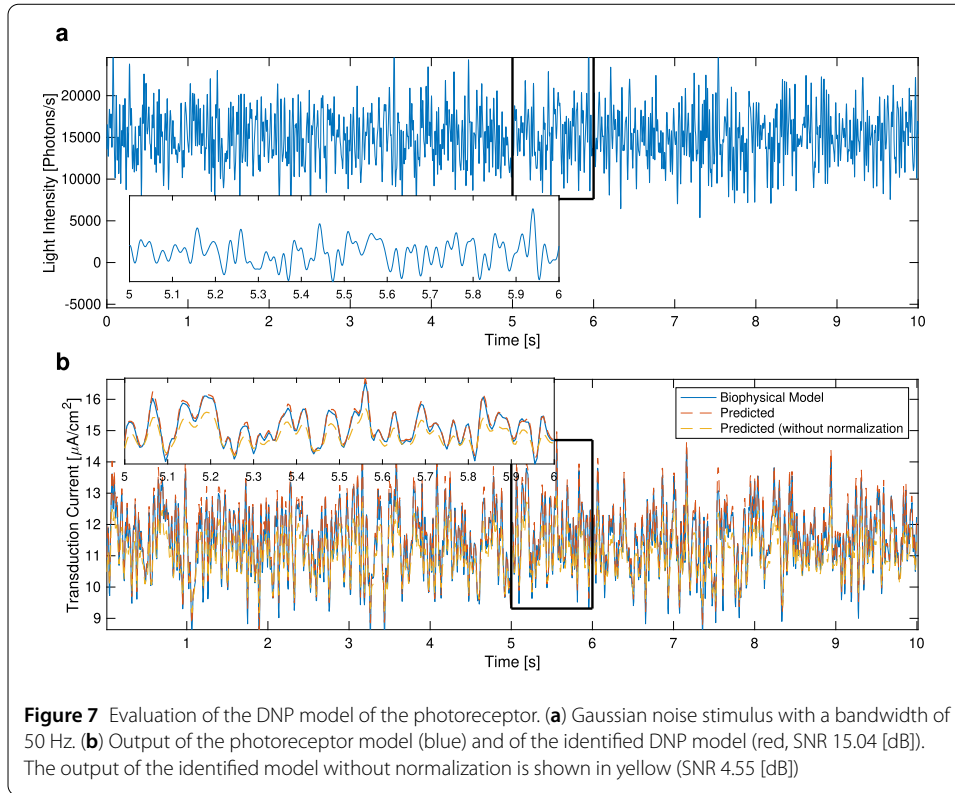
The DNP filters were identified using naturalistic stimuli. To test if the identified filters can also match the output of other types of stimuli, we presented a Gaussian noise stimulus with a bandwidth of 50 Hz to the photoreceptor model. The resulting output was compared with the output of the detailed biophysical model. As shown in Fig. 7, the output of the DNP model closely follows the actual photoreceptor output, and the SNR was 15.04 [dB]. We note that since the input space is defined as an RKHS, the statistics of the input stimuli do not affect the quality of the DNP output. We also evaluated in Fig. 7 the identification model without normalization. The SNR of the output is 4.55 [dB], a significant decrease from the output of the DNP model.

4 Sparse identification of spatio-temporal DNPs

In what follows we derive sparse identification algorithms for the components of spatio-temporal DNPs.



Given the spatio-temporal divisive normalization processor depicted in Fig. 3, we are interested in identifying all the filters from input and output observations. We formulate an optimization problem, which achieves such identification, with high fidelity and with a relatively small number of measurements.



4.1 Deriving the sparse identification algorithm for spatio-temporal DNPs

Here, we make the assumption that the filters $h^{iA}(t)$, $i = 1, 2, \dots, N$, and $h_2^{ijA}(t_1, t_2)$, $i, j = 1, 2, \dots, N$, are followed by the LPF with bandwidth Ω^o and that the output is sampled at a rate of $2f_{\max}$, where $f_{\max} = \frac{\Omega}{2\pi}$ is the maximum of the bandwidth of the filters $h^{iA}(t)$ and $h_2^{ijA}(t_1, t_2)$. By abuse of notation, we will use $v^i(t)$ to denote the low-passed version of the actual output. Note that, based upon the assumptions on the bandlimited nature of the feedback filters acting on the output, the responses of these filters to the low-passed outputs will be the same as the responses to the actual outputs.

We present M trials where input stimuli u^{mm} are chosen to be elements of the space of trigonometric polynomials \mathcal{H}_1 for $n = 1, \dots, N$, $m = 1, \dots, M$. We project the outputs $v^{mm}(t)$ on the Hilbert space of trigonometric polynomials \mathcal{H}_1^o , i.e., $\mathcal{P}_1^o v^{mm}(s)$. Note that $\mathcal{P}_1^o v^{mm}(s)$ is approximately $v^{mm}(s)$ for large values of Ω^o [66]. Further we assume that $v^{mm}, h_1^1, h_1^2, h_1^3, h_1^{iA} \in \mathbb{L}^2(\mathbb{D})$, $i = 1, 2, \dots, N$, the space of square integrable functions over domain \mathbb{D} and $h_2^1, h_2^2, h_2^3, h_2^{ijA} \in \mathbb{L}^2(\mathbb{D}^2)$, $i, j = 1, 2, \dots, N$, the space of square integrable functions over domain \mathbb{D}^2 .

We consider here the identification of the entire DNP circuit at once for two reasons. First, as all channels are connected in the spatial domain through the MVP, the inputs to the MVP are the outputs of the entire DNP circuit. Therefore, all outputs are required to identify the MVP. Second, since h_1^i, h_2^i , $i = 1, 2, 3$, are shared across all channels, fewer trials are needed to identify these filters. We present the following.

Lemma 3 *With M trials presented to the spatio-temporal DNP, let the inputs $u^{mm}(t)$ and the outputs $v^{mm}(t)$ be sampled at times (t_k) , $k = 1, \dots, T$. Then, for $i, j = 1, 2, \dots, N$, we have*

the following equations:

$$\begin{aligned}
 & b^1 + \langle h_1^1, \phi_{11}^{nmk} \rangle_{\mathcal{H}_1} + \langle h_1^2, \phi_{12}^{nmk} \rangle_{\mathcal{H}_1} + \langle h_1^3, \phi_{13}^{nmk} \rangle_{\mathcal{H}_1^o} + \sum_{i=1}^N \langle h_1^{i4}, \phi_{14}^{inmk} \rangle_{\mathcal{H}_1^o} \\
 & + \langle h_2^1, \phi_{21}^{nmk} \rangle_{\mathcal{H}_2} + \langle h_2^2, \phi_{22}^{nmk} \rangle_{\mathcal{H}_2} + \langle h_2^3, \phi_{23}^{nmk} \rangle_{\mathcal{H}_2^o} + \sum_{i=1}^N \sum_{j=1}^N \langle h_2^{ij4}, \phi_{24}^{ijnmk} \rangle_{\mathcal{H}_2^o} \\
 & = q^{nmk},
 \end{aligned} \tag{43}$$

where the sampling functions $\phi_{11}^{nmk} \in \mathcal{H}_1$, $\phi_{12}^{nmk} \in \mathcal{H}_1$, $\phi_{13}^{nmk} \in \mathcal{H}_1^o$, $\phi_{14}^{inmk} \in \mathcal{H}_1^o$, $\phi_{21}^{nmk} \in \mathcal{H}_2$, $\phi_{22}^{nmk} \in \mathcal{H}_2$, $\phi_{23}^{nmk} \in \mathcal{H}_2^o$, and $\phi_{24}^{ijnmk} \in \mathcal{H}_2^o$ are given by

$$\begin{aligned}
 \phi_{11}^{nmk}(t) &= u^{nm}(t_k - t), \\
 \phi_{12}^{nmk}(t) &= -q^{nm}(t_k)u^{nm}(t_k - t), \\
 \phi_{13}^{nmk}(t) &= -q^{nm}(t_k)(\mathcal{P}_1^o v^{nm})(t_k - t), \\
 \phi_{14}^{inmk}(t) &= -q^{nm}(t_k)(\mathcal{P}_1^o v^{im})(t_k - t), \\
 \phi_{21}^{nmk}(t, s) &= u^{nm}(t_k - t)u^{nm}(t_k - s), \\
 \phi_{22}^{nmk}(t, s) &= -q^{nm}(t_k)u^{nm}(t_k - t)u^{nm}(t_k - s), \\
 \phi_{23}^{nmk}(t, s) &= -q^{nm}(t_k)(\mathcal{P}_1^o v^{nm})(t_k - t)(\mathcal{P}_1^o v^{nm})(t_k - s), \\
 \phi_{24}^{ijnmk}(t, s) &= -q^{nm}(t_k)(\mathcal{P}_1^o v^{im})(t_k - t)(\mathcal{P}_1^o v^{im})(t_k - s),
 \end{aligned} \tag{44}$$

and $q_k^{nm} = v^{nm}(t_k)$ for all $m = 1, 2, \dots, M$, $k = 1, 2, \dots, T$, and $i, j, n = 1, 2, \dots, N$.

Proof See Appendix 4. □

Remark 5 Theorem 3 suggests that identifying the lateral divisive normalization model is equivalent to solving a generalized sampling problem with noisy measurements. It also suggests that the output needs to be sampled at a high enough rate, and that the choice of the Hilbert space used to reconstruct the feedback filters is critical since incorrect choices for these parameters can negatively affect the identification by introducing ‘noise’ in the measurements.

We now present the following algorithm to identify the model that exploits the low-rank constraints imposed on the quadratic filters.

Lemma 4 Equation (43) can be expressed in matrix form as follows:

$$\mathbf{c}_1^\top \Phi^{nmk} + \text{Tr}(\mathbf{C}_2^H \Xi^{nmk}) = q^{nmk}, \tag{45}$$

for all $n = 1, 2, \dots, N$, $m = 1, 2, \dots, M$, and $k = 1, 2, \dots, T$, where

$$\text{Measurements} \leftarrow q^{mk}(\text{scalar}) \tag{46}$$

Unknowns

$$\left\{ \begin{aligned} \mathbf{c}_1 &= \begin{bmatrix} b^1 & (\mathbf{h}_1^1)^\top & (\mathbf{h}_1^2)^\top & (\mathbf{h}_1^3)^\top \\ & (\mathbf{h}_1^{14})^\top & (\mathbf{h}_1^{24})^\top & \dots & (\mathbf{h}_1^{N4})^\top \end{bmatrix}_{1 \times (4L+2(N+1)L^0+N+4)}^\top, \\ \mathbf{C}_2 &= \begin{bmatrix} (\mathbf{H}_2^1)^\top & (\mathbf{H}_2^2)^\top & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & (\mathbf{H}_2^3)^\top & (\mathbf{H}_2^{14})^\top \\ \mathbf{0} & \dots & \mathbf{0} & \\ (\mathbf{H}_2^{124})^\top & \dots & (\mathbf{H}_2^{NN4})^\top & \end{bmatrix}_{(2L+2L^0+2) \times (4L+2(N^2+1)L^0+N^2+3)}^\top, \end{aligned} \right. \tag{47}$$

Sampling matrices

$$\left\{ \begin{aligned} \Phi^{nmk} &= \begin{bmatrix} 1 \\ \mathbf{u}^{nmk} \\ -q^{nmk} \mathbf{u}^{nmk} \\ -q^{nmk} \mathbf{v}^{nmk} \\ -q^{nmk} \mathbf{v}^{1mk} \\ -q^{nmk} \mathbf{v}^{2mk} \\ \vdots \\ -q^{nmk} \mathbf{v}^{Nmk} \end{bmatrix}_{(4L+2(N+1)L^0+N+4) \times 1}, \\ \Xi^{nmk} &= \begin{bmatrix} \mathbf{U}_2^{nmk} & \mathbf{0} \\ -q^{nmk} \mathbf{U}_2^{nmk} & \mathbf{0} \\ \mathbf{0} & -q^{nmk} \mathbf{V}_2^{nmk} \\ \mathbf{0} & -q^{nmk} \mathbf{V}_2^{1mk} \\ \mathbf{0} & -q^{nmk} \mathbf{V}_2^{2mk} \\ \vdots & \vdots \\ \mathbf{0} & -q^{nmk} \mathbf{V}_2^{Nmk} \end{bmatrix}_{(4L+2(N^2+1)L^0+N^2+3) \times (2L+2L^0+2)}. \end{aligned} \right. \tag{48}$$

We provide the following algorithm to identify $b_1, \mathbf{h}_1^i, i = 1, 2, 3, \mathbf{h}_1^{i4}, i = 1, 2, \dots, N, \mathbf{H}_2^i, i = 1, 2, 3,$ and $\mathbf{H}_2^{ij4}, i, j = 1, 2, \dots, N.$

Again, we assume that all the second-order filters have sparse structures akin to (27).

Algorithm 2 Let $\hat{\mathbf{c}}_1$ and $\hat{\mathbf{C}}_2$ be the solution to the following optimization problem:

$$\min_{\mathbf{c}_1, \mathbf{C}_2, \boldsymbol{\varepsilon}} \|\mathbf{C}_2\|_* + \lambda_1 \|\mathbf{c}_1\|_2 + \lambda_2 \|\boldsymbol{\varepsilon}\|_2 \tag{49}$$

$$\begin{aligned} \text{s.t.} \quad & \mathbf{c}_1^\top \Phi^{nmk} + \text{Tr}(\mathbf{C}_2^\text{H} \Xi^{nmk}) = q^{nmk} + \varepsilon^{(n-1)TM+(m-1)T+k}, \\ & n = 1, \dots, N, m = 1, \dots, M, k = 1, \dots, T, \end{aligned} \tag{50}$$

$$\mathbf{1}^\top \boldsymbol{\varepsilon} = \mathbf{0} \tag{51}$$

$$\begin{bmatrix} \mathbf{I}_{2 \dim(\mathcal{H}_1)} & \mathbf{0} \end{bmatrix} \mathbf{C}_2 \begin{bmatrix} \mathbf{0} \\ \mathbf{I}_{\dim(\mathcal{H}_1^0)} \end{bmatrix} = \mathbf{0}, \tag{52}$$

$$\begin{bmatrix} \mathbf{0} & \mathbf{I}_{\dim(\mathcal{H}_1^o)} \end{bmatrix} \mathbf{C}_2 \begin{bmatrix} \mathbf{I}_{\dim(\mathcal{H}_1)} \\ \mathbf{0} \end{bmatrix} = \mathbf{0}, \tag{53}$$

$$\begin{bmatrix} \mathbf{I}_{\dim(\mathcal{H}_1)} & \mathbf{0} \end{bmatrix} \mathbf{C}_2 \begin{bmatrix} \mathbf{I}_{\dim(\mathcal{H}_1)} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{\dim(\mathcal{H}_1)} \\ \mathbf{0} \end{bmatrix}^\top \mathbf{C}_2^H \begin{bmatrix} \mathbf{I}_{\dim(\mathcal{H}_1)} & \mathbf{0} \end{bmatrix}^\top, \tag{54}$$

$$\begin{aligned} & \begin{bmatrix} \mathbf{0}_{\dim(\mathcal{H}_1)} & \mathbf{I}_{\dim(\mathcal{H}_1)} & \mathbf{0} \end{bmatrix} \mathbf{C}_2 \begin{bmatrix} \mathbf{I}_{\dim(\mathcal{H}_1)} \\ \mathbf{0} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{I}_{\dim(\mathcal{H}_1)} \\ \mathbf{0} \end{bmatrix}^\top \mathbf{C}_2^H \begin{bmatrix} \mathbf{0}_{\dim(\mathcal{H}_1)} & \mathbf{I}_{\dim(\mathcal{H}_1)} & \mathbf{0} \end{bmatrix}^\top, \end{aligned} \tag{55}$$

$$\begin{bmatrix} \mathbf{0} & \mathbf{I}_{\dim(\mathcal{H}_1^o)} \end{bmatrix} \mathbf{C}_2 \begin{bmatrix} \mathbf{0} \\ \mathbf{I}_{\dim(\mathcal{H}_1^o)} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{I}_{\dim(\mathcal{H}_1^o)} \end{bmatrix}^\top \mathbf{C}_2^H \begin{bmatrix} \mathbf{0} & \mathbf{I}_{\dim(\mathcal{H}_1^o)} \end{bmatrix}^\top, \tag{56}$$

where λ_1, λ_2 are appropriately chosen hyperparameters and $\varepsilon^i \in \mathbb{R}, i = 1, 2, \dots, MTN$, represent slack variables.

The constraints in Algorithm 2 are similar to those in Algorithm 1. Note that H^{ij4} are not constrained to be Hermitian. This follows from the assumption that the MVP block may perform asymmetric processing on any pair of inputs to the block.

Theorem 2 *The identified spatio-temporal divisive normalization is specified as $\hat{b}^1 = \hat{b}^1$ and*

$$\hat{h}_1^1(t) = \sum_{l=-L}^L [\hat{\mathbf{h}}_1^1]_{l+L+1} \cdot e_l(t), \quad \hat{h}_1^2(t) = \sum_{l=-L}^L [\hat{\mathbf{h}}_1^2]_{l+L+1} \cdot e_l(t), \tag{57}$$

$$\hat{h}_1^3(t) = \sum_{l=-L^o}^{L^o} [\hat{\mathbf{h}}_1^3]_{l+L+1} \cdot e_l^o(t), \quad \hat{h}_1^{i4}(t) = \sum_{l=-L^o}^{L^o} [\hat{\mathbf{h}}_1^{i4}]_{l+L+1} \cdot e_l^o(t), \tag{58}$$

$$\hat{h}_2^1(t_1, t_2) = \sum_{l_1=-L}^L \sum_{l_2=-L}^L [\hat{\mathbf{H}}_2^1]_{l_1+L+1, l_2+1-l_2} \cdot e_{l_1}(t_1) e_{l_2}(t_2), \tag{59}$$

$$\hat{h}_2^2(t_1, t_2) = \sum_{l_1=-L}^L \sum_{l_2=-L}^L [\hat{\mathbf{H}}_2^2]_{l_1+L+1, l_2+1-l_2} \cdot e_{l_1}(t_1) e_{l_2}(t_2), \tag{60}$$

$$\hat{h}_2^3(t_1, t_2) = \sum_{l_1=-L^o}^{L^o} \sum_{l_2=-L^o}^{L^o} [\hat{\mathbf{H}}_2^3]_{l_1+L^o+1, l_2^o+1-l_2} \cdot e_{l_1}^o(t_1) e_{l_2}^o(t_2), \tag{61}$$

$$\begin{aligned} \hat{h}_2^{ij4}(t_1, t_2) &= \sum_{l_1=-L^o}^{L^o} \sum_{l_2=-L^o}^{L^o} [\hat{\mathbf{H}}_2^{ij4}]_{l_1+L^o+1, l_2^o+1-l_2} \cdot e_{l_1}^o(t_1) e_{l_2}^o(t_2), \\ i, j &= 1, 2, \dots, N, \end{aligned} \tag{62}$$

where

$$\begin{bmatrix} \hat{b}^1 & \hat{\mathbf{h}}_1^1{}^\top & \hat{\mathbf{h}}_1^2{}^\top & \hat{\mathbf{h}}_1^3{}^\top & \hat{\mathbf{h}}_1^{14}{}^\top & \hat{\mathbf{h}}_1^{24}{}^\top & \dots & \hat{\mathbf{h}}_1^{N4}{}^\top \end{bmatrix}^\top = \hat{\mathbf{c}}_1 \tag{63}$$

and

$$\begin{bmatrix} \widehat{\mathbf{H}}_2^T & \widehat{\mathbf{H}}_2^T & - & - & - & \cdots & - \\ - & - & \widehat{\mathbf{H}}_2^{3T} & \widehat{\mathbf{H}}_2^{114T} & \widehat{\mathbf{H}}_2^{124T} & \cdots & \widehat{\mathbf{H}}_2^{NN4T} \end{bmatrix}^T = \widehat{\mathbf{C}}_2. \tag{64}$$

Remark 6 Assuming that all the second-order kernels in the lateral divisive normalization model have rank r , the expected number of measurements for Algorithm 2 to identify the model is of the order $\mathcal{O}(r \cdot \dim(\mathcal{H}_1) + N^2 r \cdot \dim(\mathcal{H}_1^o))$. When N is large, the N^2 factor may become prohibitive in identifying the model. Additional assumptions on h_2^{ij4} may help mitigate this and maintain tractability of solving the identification problem. For example, with the assumption that $h^{ij4} = h_1^{14}$ if $i = j$ and $h_2^{ij4} = h_1^{24}$ otherwise, the expected number of measurements required will be $\mathcal{O}(r \cdot \dim(\mathcal{H}_1) + Nr \cdot \dim(\mathcal{H}_1^o))$.

4.2 An example of sparse identification of a spatio-temporal DNP

We now present an example of identification obtained by Algorithm 2. We demonstrate here that, in addition to the identification of the Volterra kernels operating within each channel, the MVP in the spatio-temporal DNP can be identified.

Example 3 Here, we choose the DNP in Fig. 3 with $N = 4$, and

$$h_1^1(t) = 25te^{-25t}, \tag{65}$$

$$h_1^2(t) = 25te^{-25t}, \tag{66}$$

$$h_1^{i4}(t) = e^{-\frac{1}{4}(i-2)^2} (25 - 600t)e^{-25t}, \tag{67}$$

$$h_2^{ij4}(t_1, t_2) = 5000e^{-\frac{1}{4}(i-2)^2} e^{-\frac{1}{4}(j-2)^2} \cdot (25t_1 e^{-25t_1})(25t_2 e^{-25t_2}), \tag{68}$$

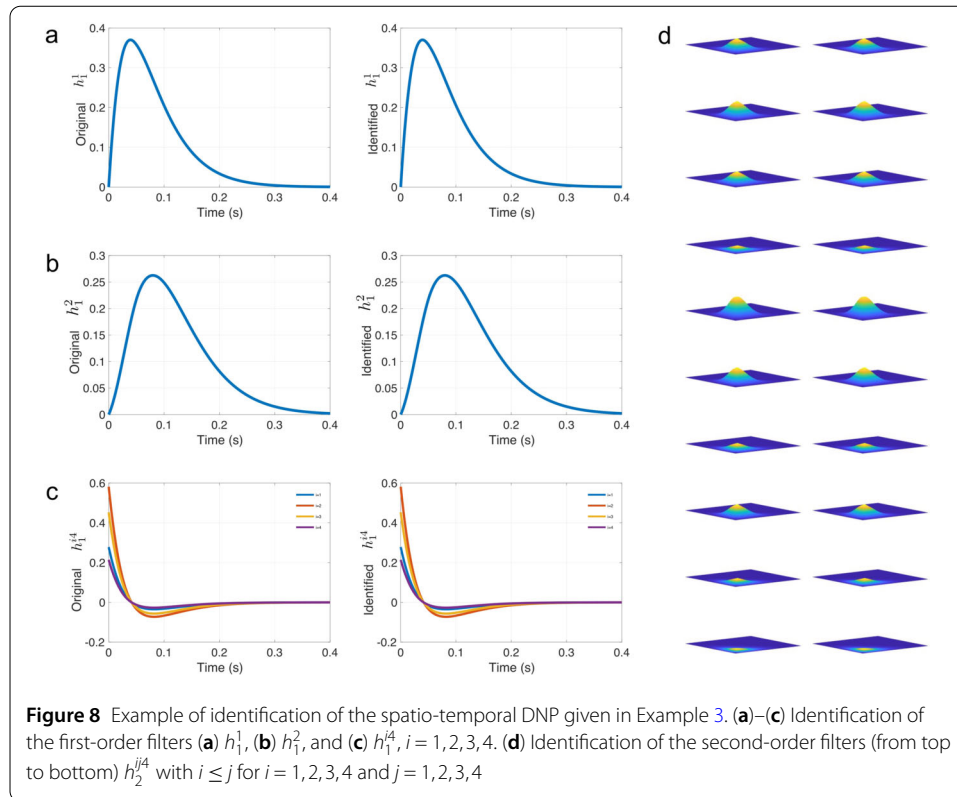
and the other Volterra kernels are set to 0. Note that h_1^1 , and h_1^2 are shared across all channels. In addition, $h^{ij4} = h^{ji4}$ in this case. We assumed knowledge of this symmetric structure of the model and adapted the identification algorithm accordingly and only identified six linear filters and ten quadratic filters.

We performed the identification in the Hilbert space of bandlimited functions with $\Omega = \Omega^o = 40\pi$, and we chose the same space for the input stimuli and for both the feed-forward filters and the feedback ones. We solved for the filters truncated to a period of 0.4 s, and thus there were 17 coefficients to be identified for the linear filters and 17×17 coefficients to be identified for each of the quadratic filters. A total of 1116 measurements (279 measurements from each cell) were used to perform the identification, and the results are depicted in Fig. 8. The average SNR of reconstruction across all filters was more than 150 [dB]. Note that solving the generalized sampling problem directly for the same problem would have required at least 3570 measurements.

5 Discussion

As already mentioned in the [Introduction](#), the photoreceptor/amacrine cell layer of the early vision system of the fruit fly rapidly adapts to visual stimuli whose intensity and contrast vary orders of magnitude both in space and time.

In this paper we presented a spatio-temporal divisive normalization processor that models the transduction and the contrast gain control in the photoreceptor and amacrine cell



layer of the fruit fly. It incorporates processing blocks that explicitly model the feedforward and the temporal feedback path of each photoreceptor and the spatio-temporal feedback from amacrine cells to photoreceptors. We demonstrated that with some simple choice of parameters, the DNP response maintains the contrast of the input visual field across a large range of average spatial luminance values.

We characterized the I/O of the spatio-temporal DNP and highlighted the highly nonlinear behavior of the DNP in contrast gain control. Despite the divisive nonlinearity, we provided an algorithm for the sparse identification of the entire DNP. We showed that the identification of the DNP components can be interpreted as a generalized sampling problem. More importantly, the sparse identification algorithm does not suffer from the curse of dimensionality that would otherwise require a large number of measurements that are quadratically related to the dimension of the input and output spaces.

Neural circuits are often noisy. Although the I/O characterization of MIMO DNPs provided in Sect. 2 is noise free, it can be easily extended to account for noise [52–54]. The identification of DNPs can be interpreted as a generalized sampling problem in an RKHS with noisy measurements. The sparse identification algorithms we provided in this paper include slack variables to account for these inaccurate measurements. Similar algorithms have been shown to be robust to noise [42].

Contrast gain control is an intrinsically nonlinear problem. Early approaches to modeling contrast gain control rely on analyzing the Volterra kernels, *i.e.*, linear or nonlinear receptive fields identified when stimuli of different statistics are presented [19, 20, 23]. The objective of these studies is to observe and characterize the differences between the identified Volterra kernels. To fully capture the nonlinear effect in contrast gain control, these approaches would rely on identifying higher-order Volterra kernels that are too costly to

compute and are simply ignored. The contrast gain control model advanced in this paper, however, results from the divisive normalization operation, albeit having Volterra kernels modeling internal filters in the DNP. The division largely expands upon DNP's ability to model higher-order nonlinearities in contrast gain control.

Biologically, divisive normalization can take place in feedforward or feedback circuits, or both [26]. Most divisive normalization models only consider the feedforward normalization circuit. Although normalization can be achieved by a multiplicative feedback mechanism, it mainly serves as a transformation leading to an equivalent form of feedforward divisive normalization when the input is a constant [27, 67]. The MIMO DNP models described in this paper explicitly include divisive normalization with both feedforward and feedback paths that can be more readily employed in modeling underlying neuronal circuit mechanisms, such as those arising in the photoreceptor and amacrine cell layer.

Predictive coding often takes the form subtractive negative feedback [68]. The MIMO DNP models described in this paper represent an alternative form of predictive coding. Indeed, divisive input modulation has been proposed to implement predictive coding by calculating the residual errors using division rather than subtraction [69]. In flies, it has been suggested that inhibition in the lamina, where retina photoreceptors and amacrine cells interact, leads to predictive coding [70]. Our study advances a spatio-temporal MIMO model for predictive coding and provides an algorithm to identify the components of the MIMO DNP.

The MIMO DNP model opens a new avenue for exploring and quantifying the highly nonlinear nature of sensory processing. The MIMO DNP in Fig. 3 can be further extended to allow the different transformations \mathcal{T}^i , $i = 1, 2, 3$, to incorporate spatio-temporal Volterra kernels, thereby making it more versatile for modeling other types of sensory processing, including (i) interactions between cones and horizontal cells in vertebrate retinas [58], (ii) channels/glomeruli in olfactory circuits and interactions between them through local neurons [9, 71], and (iii) cross-suppression and gain control in the auditory [12, 72], and visual cortices [73–75]. The sparse identification algorithms advanced here can be easily extended to identify the MIMO DNPs of these systems as well.

Appendix 1: Spatio-temporal DNPs and contrast gain control

Here, we characterize the I/O of simple DNPs stimulated with three different inputs. In the first example, we evaluate the response of a 1×4 DNP under different background light intensity levels. In the second example, contrast gain control exerted by the amacrine cells is demonstrated with the same DNP. In the third example, a DNP consisting of 16×16 DNPs tiling a 1536×1024 visual field is stimulated with a natural image taken at low, medium, and high luminance values. The DNP output is evaluated with and without the MVP block.

Example 4 Here, we consider a simple 1×4 DNP consisting of four photoreceptors and a single amacrine cell receiving inputs from and providing feedback to all four photoreceptors. The choices of the component filters of the DNP are as follows:

$$b_2 + b_3 + b_4 = 1, \tag{69}$$

for the transformations \mathcal{T}^1 and \mathcal{T}^2 , the kernels are

$$h_1^1(t) = h_1^2(t) = \frac{t}{0.1} \exp\left(-\frac{t}{0.1}\right), \quad (70)$$

$$h_2^1(t_1, t_2) = h_2^2(t_1, t_2) = 0.0001 h_1^1(t_1) h_1^1(t_2), \quad (71)$$

for the transformation \mathcal{T}^3 , the kernels are

$$h_1^3(t) = 0, \quad (72)$$

$$h_2^3(t_1, t_2) = 0, \quad (73)$$

and for the transformation \mathcal{L}^4 , the kernels are

$$h_1^{i4}(t) = 10,000 \frac{t}{0.8} \exp\left(-\frac{t}{0.2}\right), \quad i = 1, 2, 3, 4, \quad (74)$$

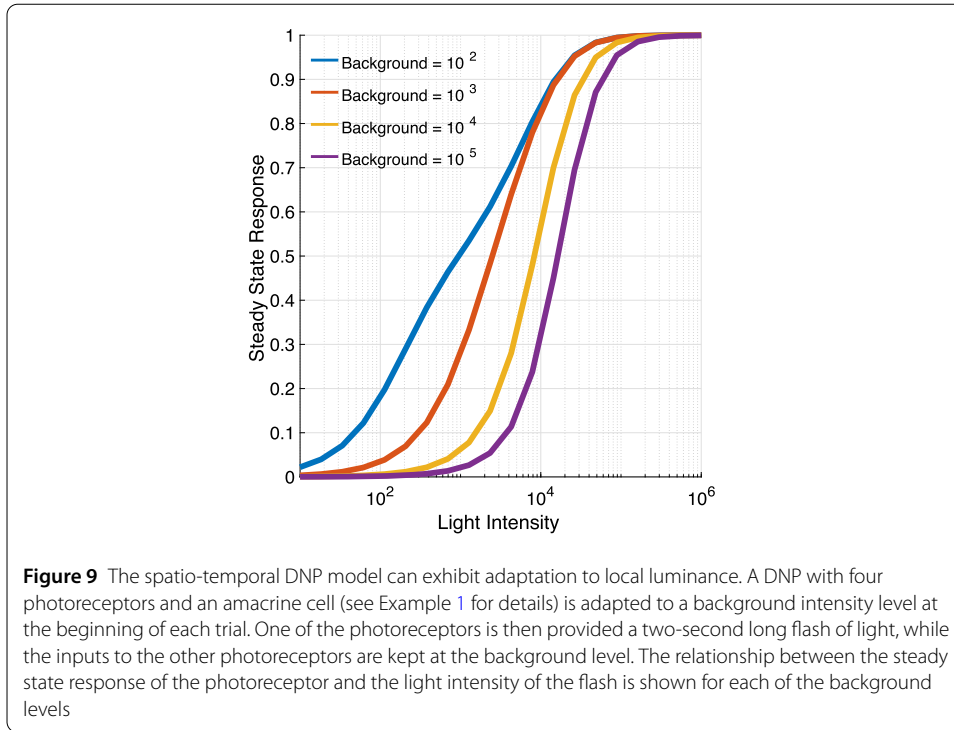
$$h_2^{ij4}(t_1, t_2) = \begin{cases} 12,500 \frac{t_1 t_2}{0.04} \exp\left(-\frac{t_1 + t_2}{0.2}\right), & i = j, \\ 0, & i \neq j, \end{cases} \quad i, j = 1, 2, 3, 4. \quad (75)$$

Here, the bandwidth of \mathcal{H}_1^o was chosen to be $\Omega^o = 10 \cdot 2\pi$ rad/s. The I/O of the fly's individual photoreceptors in steady state is described by a saturating nonlinearity similar in appearance to a sigmoid when plotted on a logarithmic scale [1]. Photoreceptors can deal with the orders of magnitude of the input stimuli while maintaining their output in a suitable range. In addition, the photoreceptors exhibit adaptation so that the sigmoidal curves shift to adjust to the mean local luminance value [26].

We examine here the steady state response of the above DNP under four different background light intensity levels. At the start of each trial, all four photoreceptors are first adapted to the same background light intensity level. One of the photoreceptors is then subjected to an additional flash of different light intensity with a duration of two seconds, while the inputs to the other three are kept at the same background level. We observe the value of the steady state response of the photoreceptor that receives the additional flash. Figure 9 depicts the relationship between the observed steady state response and the light intensity of the flash at the four background intensity levels. It demonstrates that the response of the DNP is background-dependent, and the overall slope and range are similar across different background levels. With the MVP block, the spatio-temporal DNP can reproduce responses of photoreceptors observed in experimental settings [26, 46].

Since all the photoreceptors exhibit a sigmoid like nonlinearity, the output of the retina will be constrained in a suitable current range that can be processed by postsynaptic neurons in the lamina. However, without adaptation to mean local luminance, the saturating nature of the nonlinearities leads to a loss of spatial contrast. The spatial contrast is preserved by spatial gain control or adaptation modeled here with the MVP block.

Example 5 Here, the I/O of DNPs with and without the MVP block is evaluated. Using the same DNP as in the example above, we stimulated the DNP with 25 “images” with a resolution of 1×4 pixels. Each image has a different average luminance and root mean



square (RMS) contrast. The RMS contrast is defined as the standard deviation of pixel intensities normalized by the mean, i.e.,

$$C_{rms} = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (u^i - \bar{u})^2}}{\bar{u}}, \quad (76)$$

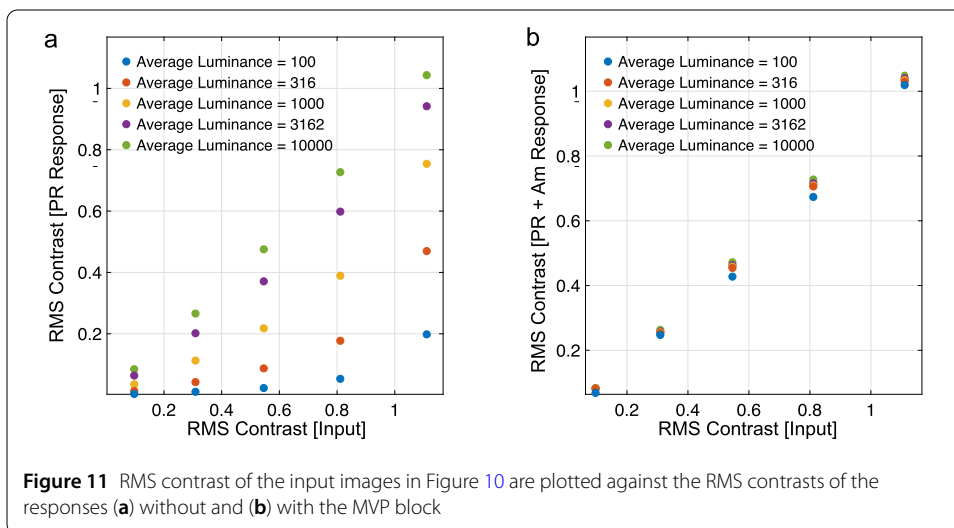
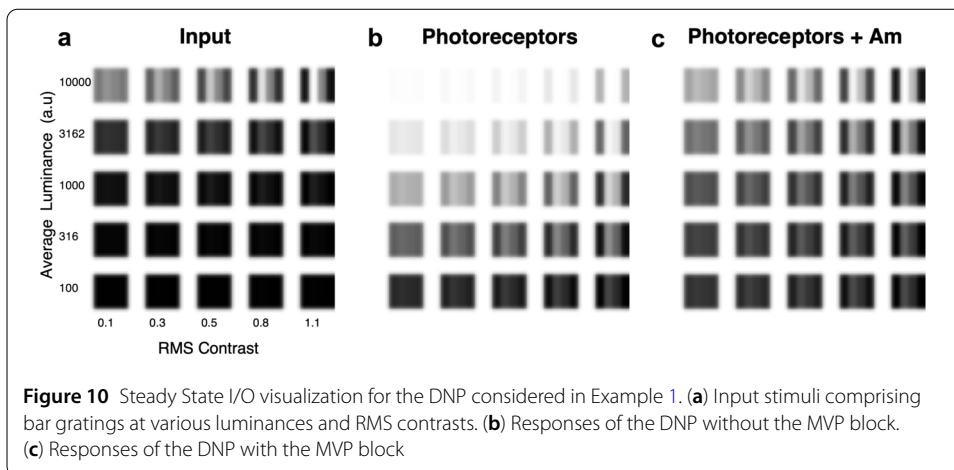
where

$$\bar{u} = \frac{1}{N} \sum_{i=1}^N u^i \quad (77)$$

and $N = 4$. These images are shown in the “input” block in Fig. 10, with each bar representing the input intensity to one photoreceptor. Note that the pixels are extended along the y-axis for a quick visual examination.

In the “Photoreceptors” block in Fig. 10, the steady state responses of the DNP without the MVP block to the respective inputs are shown. Here, pure black and white represent a response of 0 and 1, respectively. This can be interpreted as a circuit in which the reciprocal connections between photoreceptors and amacrine cells are blocked.

In the “Photoreceptors + Am” block in Fig. 10, the steady state responses of the full DNP to their respective inputs are shown. Comparing the “Photoreceptors” and “Photoreceptors + Am” blocks, the responses of the DNP without feedback are washed out, or exhibit low contrast, particularly at three of the four corners of the 5×5 image array, i.e., when either the luminance or contrast is too high or too low. By contrast, the individual bars in the response of the full DNP model are more readily discernible across several orders of magnitude of luminance.



We note that a saturating nonlinearity can maintain its output in a constrained range even when the input varies by orders of magnitude. However, the nonlinearity may lead to a loss of spatial contrast. In contrast, the DNP constrains its output to a suitable range while maintaining the spatial contrast. This is demonstrated in Fig. 11, where the RMS contrast of the input images in Fig. 10 is plotted against the RMS contrasts of the responses for both cases—with and without MVP in the DNP. Spatial contrast, arguably, is an important feature of the image that should be preserved or even enhanced for subsequent stages of extraction of ethologically relevant information from the visual stimulus.

Example 6 Here, we apply a full-scale DNP model to a natural image. The image is taken in raw format so that its pixel values are proportional to the light intensity each pixel is exposed to [65]. The resolution of the image is 1536×1024 . The image is first divided into 16×16 blocks with a four-pixel overlap in each direction. A DNP is assigned to each block, and the filters in the DNP are designed as follows:

$$b_2 + b_3 + b_4 = 1, \tag{78}$$

for the transformations \mathcal{T}^1 and \mathcal{T}^2 , the kernels are

$$h_1^1(t) = h_1^2(t) = \frac{t}{0.04} \exp\left(-\frac{t}{0.04}\right), \tag{79}$$

$$h_2^1(t_1, t_2) = h_2^2(t_1, t_2) = 0.0001h_1^1(t_1)h_1^1(t_2), \tag{80}$$

for the transformation \mathcal{T}^3 , the kernels are

$$h_1^3(t) = 0, \tag{81}$$

$$h_2^3(t_1, t_2) = 0, \tag{82}$$

and for the transformation \mathcal{L}^4 , the kernels are

$$h_1^{i4}(t) = 20 \frac{t}{0.04} \exp\left(-\frac{t}{0.04}\right) \cdot \frac{1}{\sqrt{2\pi} \cdot 16} \exp\left(-\frac{(x^i - x^0)^2 + (y^i - y^0)^2}{32}\right),$$

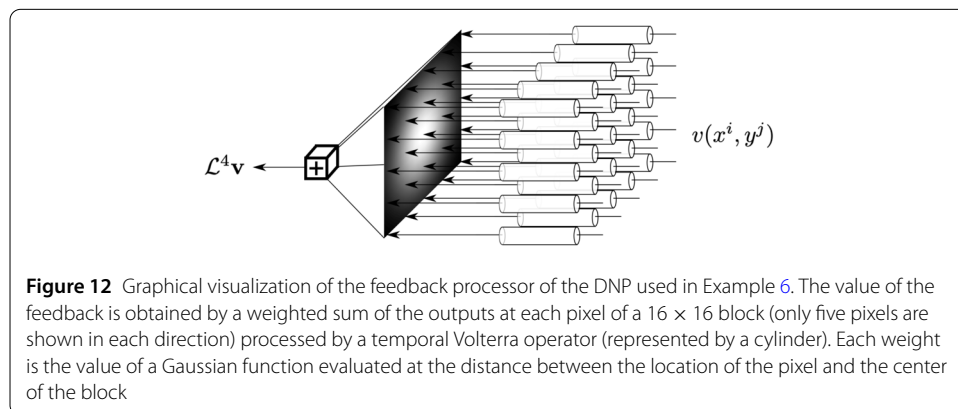
$$i = 1, 2, \dots, 256, \tag{83}$$

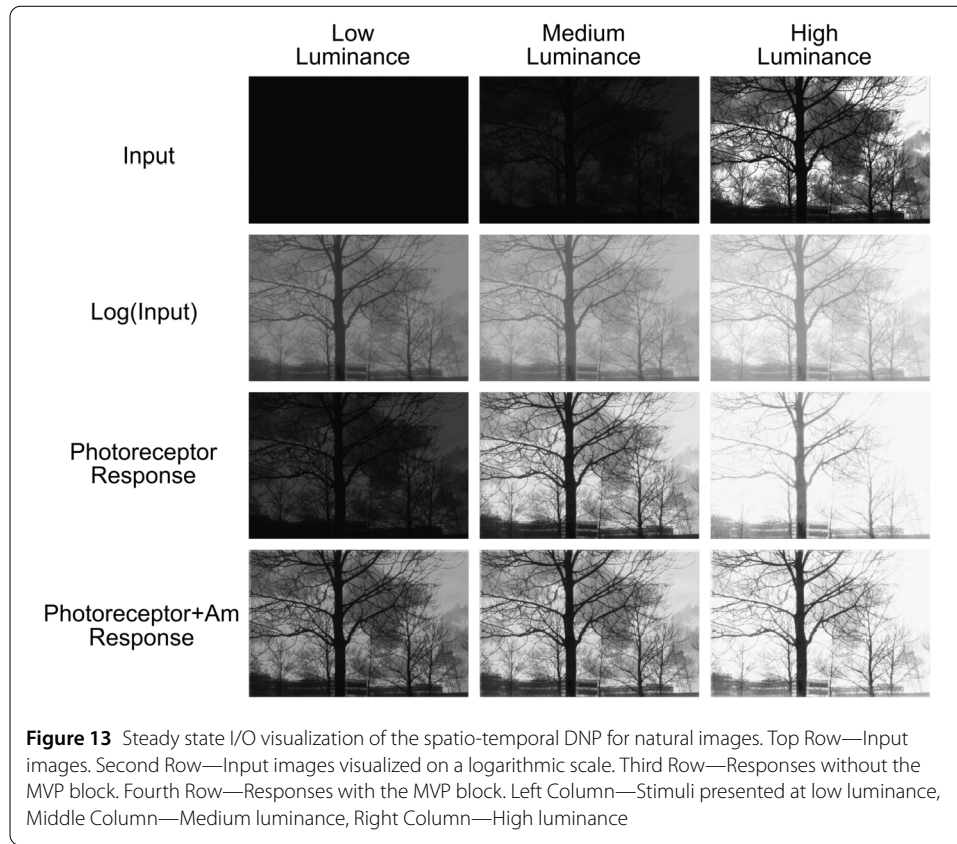
$$h_2^{ij4}(t_1, t_2) = \begin{cases} 100 \frac{t_1 t_2}{0.04^2} \exp\left(-\frac{t_1 + t_2}{0.04}\right) \cdot \sqrt{\frac{1}{2\pi \cdot 16}} \exp\left(-\frac{(x^i - x^0)^2 + (y^i - y^0)^2}{32}\right), & i = j, \\ 0, & i \neq j, \end{cases} \tag{84}$$

$$i, j = 1, 2, \dots, 256,$$

where (x^i, y^i) is the coordinate of pixel i , and (x^0, y^0) is the coordinate of the center pixel in the 16×16 block. Here, the bandwidth of \mathcal{H}_1^o was chosen to be $\Omega^o = 40 \cdot 2\pi$ rad/s.

Figure 12 provides a graphical visualization of the feedback processing of the DNP. We note that the RHS of both equalities (83) and (84) can be separated into a temporal term and a spatial term. The temporal terms correspond to temporal Volterra kernels processing the output at each pixel (see cylinders in Fig. 12). The feedback term is then the weighted sum with a weight according to the spatial term. Each weight is the value of a Gaussian function evaluated at the distance between the location of the pixel and the center of the block. Due to the overlaps between blocks, there are pixels that belong to





multiple blocks. The feedback term for such a pixel is the average of the feedback terms from all the blocks the pixel belongs to.

We tested the DNP using low, medium, and high luminance ($1\times$, $10\times$, and $100\times$) of the original image measured in the number of photons and represented by arbitrary units. They are shown on the top row of Fig. 13. The logarithms of the inputs are shown on the second row. On the third row, we show the response of the DNP without the MVP block, *i.e.*, each photoreceptor processes independently a single pixel. The responses of the full DNP model are shown on the bottom row.

As can be seen from Fig. 13, the response of the DNP with MVP is robust across the different levels of luminance, showing the best quality for all three luminance levels. By contrast, the response of the DNP without MVP is either too dark for low contrast or saturated at high luminance. The contrast within each response is also significantly worse. It is also not ideal to use only a logarithmic nonlinearity to process the images (second row), as the contrast of the images is not sharp enough.

To conclude we note that, with a simple choice of filters, the spatio-temporal DNP proposed in Sect. 2.3 operates in a range of luminance values spanning several orders of magnitude, much like the photoreceptor-amacrine cell layer. The contrast of the output of the DNP is not compromised while its range remains strongly bounded.

Appendix 2: Proof of Lemma 1

For the m th trial and the k th time sample, multiplying out the denominator in (5), we get

$$\begin{aligned}
 v^{nm}(t_k) & \left(1 + \int_{\mathbb{D}} h_1^2(s) u^{nm}(t_k - s) ds + \int_{\mathbb{D}^2} h_2^2(s_1, s_2) u_2^{nm}(t_k - s_1, t_k - s_2) ds_1 ds_2 \right) \\
 & + v^{nm}(t_k) \left(\int_{\mathbb{D}} h_1^3(s) (\mathcal{P}_1^o v^{nm})(t_k - s) ds \right. \\
 & \left. + \int_{\mathbb{D}^2} h_2^3(s_1, s_2) (\mathcal{P}_2^o v_2^{nm})(t_k - s_1, t_k - s_2) ds_1 ds_2 \right) \\
 & = b_1 + \int_{\mathbb{D}} h_1^1(s) u^{nm}(t_k - s) ds + \int_{\mathbb{D}^2} h_2^1(s_1, s_2) u_2^{nm}(t_k - s_1, t_k - s_2) ds_1 ds_2, \tag{85}
 \end{aligned}$$

where $u_2^{nm}(t_1, t_2) \in \mathcal{H}_2$ and $u^{nm}(t_1, t_2) = u^{nm}(t_1)u^{nm}(t_2)$.

With the notation in (22) and after rearranging terms in equation (85) above, we obtain

$$\begin{aligned}
 v^{nm}(t_k) & = b^1 + \int_{\mathbb{D}} h_1^1(s) \phi_{11}^{nmk} ds + \int_{\mathbb{D}} h_1^2(s) \phi_{12}^{nmk} ds + \int_{\mathbb{D}} h_1^3(s) \phi_{13}^{nmk} ds \\
 & + \int_{\mathbb{D}^2} h_2^1(s_1, s_2) \phi_{21}^{nmk} ds_1 ds_2 + \int_{\mathbb{D}^2} h_2^2(s_1, s_2) \phi_{22}^{nmk} ds_1 ds_2 \\
 & + \int_{\mathbb{D}^2} h_2^3(s_1, s_2) \phi_{23}^{nmk} ds_1 ds_2. \tag{86}
 \end{aligned}$$

With $q^{nmk} = v^{nm}(t_k)$, equation (86) becomes

$$\begin{aligned}
 b^1 + \langle h_1^1, \phi_{11}^{nmk} \rangle_{\mathcal{H}_1} + \langle h_1^2, \phi_{12}^{nmk} \rangle_{\mathcal{H}_1} + \langle h_1^3, \phi_{13}^{nmk} \rangle_{\mathcal{H}_1^o} \\
 + \langle h_2^1, \phi_{21}^{nmk} \rangle_{\mathcal{H}_2} + \langle h_2^2, \phi_{22}^{nmk} \rangle_{\mathcal{H}_2} + \langle h_2^3, \phi_{23}^{nmk} \rangle_{\mathcal{H}_2^o} = q^{nmk}
 \end{aligned}$$

for all $m = 1, 2, \dots, M$ and $k = 1, 2, \dots, T$.

Appendix 3: Proof of Lemma 2

Let

$$h_1^1(t) = \sum_{l=-L}^L h_{1l}^1 \cdot e_l(t), \quad h_2^1(t) = \sum_{l_1=-L}^L \sum_{l_2=-L}^L h_{2l_1 l_2}^1 \cdot e_{l_1}(t_1) \cdot e_{l_2}(t_2), \tag{87}$$

$$h_1^2(t) = \sum_{l=-L}^L h_{1l}^2 \cdot e_l(t), \quad h_2^2(t) = \sum_{l_1=-L}^L \sum_{l_2=-L}^L h_{2l_1 l_2}^2 \cdot e_{l_1}(t_1) \cdot e_{l_2}(t_2), \tag{88}$$

$$h_1^3(t) = \sum_{l=-L^o}^{L^o} h_{1l}^3 \cdot e_l^o(t), \quad h_2^3(t) = \sum_{l_1=-L^o}^{L^o} \sum_{l_2=-L^o}^{L^o} h_{2l_1 l_2}^3 \cdot e_{l_1}^o(t_1) \cdot e_{l_2}^o(t_2), \tag{89}$$

and the $(2L + 1) \times 1$ vectors

$$\mathbf{h}_1^1 = \begin{bmatrix} h_{1,-L}^1 \\ h_{1,-L+1}^1 \\ \vdots \\ h_{1,L}^1 \end{bmatrix}, \quad \mathbf{h}_1^2 = \begin{bmatrix} h_{1,-L}^2 \\ h_{1,-L+1}^2 \\ \vdots \\ h_{1,L}^2 \end{bmatrix}, \tag{90}$$

the $(2L^o + 1) \times 1$ vector

$$\mathbf{h}_1^3 = \begin{bmatrix} h_{1,-L^o}^3 \\ h_{1,-L^o+1}^3 \\ \vdots \\ h_{1,L^o}^3 \end{bmatrix}, \tag{91}$$

the $(2L + 1) \times (2L + 1)$ Hermitian matrices

$$\mathbf{H}_2^1 = \begin{bmatrix} h_{2,-L,L}^1 & \cdots & h_{2,L,L}^1 \\ \vdots & \ddots & \vdots \\ h_{2,-L,-L}^1 & \cdots & h_{2,L,-L}^1 \end{bmatrix}, \tag{92}$$

$$\mathbf{H}_2^2 = \begin{bmatrix} h_{2,-L,L}^2 & \cdots & h_{2,L,L}^2 \\ \vdots & \ddots & \vdots \\ h_{2,-L,-L}^2 & \cdots & h_{2,L,-L}^2 \end{bmatrix}, \tag{93}$$

and the $(2L^o + 1) \times (2L^o + 1)$ Hermitian matrix

$$\mathbf{H}_2^3 = \begin{bmatrix} h_{2,-L^o,L^o}^3 & \cdots & h_{2,L^o,L^o}^3 \\ \vdots & \ddots & \vdots \\ h_{2,-L^o,-L^o}^3 & \cdots & h_{2,L^o,-L^o}^3 \end{bmatrix}. \tag{94}$$

Further, let

$$\mathbf{u}^{nmk} = \sqrt{S} \begin{bmatrix} a_{-L}^{nm} \cdot e_{-L}(t_k) \\ a_{-L+1}^{nm} \cdot e_{-L+1}(t_k) \\ \vdots \\ a_L^{nm} \cdot e_L(t_k) \end{bmatrix}, \quad \mathbf{U}_2^{nmk} = \mathbf{u}^{nmk} (\mathbf{u}^{nmk})^H, \tag{95}$$

where a_l^{nm} represents the coefficient of $u^{nm}(t)$ w.r.t. basis element e_l in \mathcal{H}_1 such that $u^{nm} = \sum_{l=-L}^L a_l^{nm} e_l$, and $(\cdot)^H$ represents conjugate transpose, and let

$$\mathbf{v}^{nmk} = \sqrt{S^o} \begin{bmatrix} d_{-L^o}^{nm} \cdot e_{-L^o}^o(t_k) \\ d_{-L^o+1}^{nm} \cdot e_{-L^o+1}^o(t_k) \\ \vdots \\ d_{L^o}^{nm} \cdot e_{L^o}^o(t_k) \end{bmatrix}, \quad \mathbf{V}_2^{nmk} = \mathbf{v}^{nmk} (\mathbf{v}^{nmk})^H, \tag{96}$$

where d_l^{nm} represents the coefficient of $(\mathcal{P}_1^o v^{nm})$ w.r.t. basis element e_l^o in \mathcal{H}_1^o such that $(\mathcal{P}_1^o v^{nm}) = \sum_{l=-L^o}^{L^o} d_l^{nm} e_l^o$.

Finally, we define the $(4L + 2L^o + 4) \times 1$ vector

$$\Phi^{nmk} = \begin{bmatrix} 1 \\ \mathbf{u}^{nmk} \\ -q^{nmk} \mathbf{u}^{nmk} \\ -q^{nmk} \mathbf{v}^{nmk} \end{bmatrix} \tag{97}$$

and $(4L + 2L^o + 3) \times (2L + 2L^o + 2)$ matrix

$$\mathbf{\Xi}^{nmk} = \begin{bmatrix} \mathbf{U}_2^{nmk} & \mathbf{0}_{(2L+1) \times (2L^o+1)} \\ -q^{nmk} \mathbf{U}_2^{nmk} & \mathbf{0}_{(2L+1) \times (2L^o+1)} \\ \mathbf{0}_{(2L^o+1) \times (2L+1)} & -q^{nmk} \mathbf{V}_2^{nmk} \end{bmatrix}. \tag{98}$$

Finally, we define the $(4L + 2L^o + 4) \times 1$ vector

$$\mathbf{c}_1 = \left[b^1 \quad (\mathbf{h}_1^1)^\top \quad (\mathbf{h}_1^2)^\top \quad (\mathbf{h}_1^3)^\top \right]^\top \tag{99}$$

and the $(4L + 2L^o + 3) \times (2L + 2L^o + 2)$ matrix

$$\mathbf{C}_2 = \begin{bmatrix} \mathbf{H}_2^1 & \mathbf{0}_{(2L+1) \times (2L^o+1)} \\ \mathbf{H}_2^2 & \mathbf{0}_{(2L+1) \times (2L^o+1)} \\ \mathbf{0}_{(2L^o+1) \times (2L+1)} & \mathbf{H}_2^3 \end{bmatrix}. \tag{100}$$

The proof is based on two simple observations. First,

$$\begin{aligned} \langle h_1^1, \phi_{11}^{nmk} \rangle_{\mathcal{H}_1} &= \left\langle \sum_{l=-L}^L h_{1l}^1 \cdot e_l(t), \sum_{l=-L}^L a_l^{nm} \cdot e_l(t_k - t) \right\rangle_{\mathcal{H}_1} \\ &= \sum_{l=-L}^L h_{1l}^1 a_l^{nm} \cdot \sqrt{S} e_l(t_k) = (\mathbf{h}_1^1)^\top \mathbf{u}^{nmk}, \end{aligned} \tag{101}$$

and therefore

$$b^1 + \langle h_1^1, \phi_{11}^{nmk} \rangle_{\mathcal{H}_1} + \langle h_1^2, \phi_{12}^{nmk} \rangle_{\mathcal{H}_1} + \langle h_1^3, \phi_{13}^{nmk} \rangle_{\mathcal{H}_1} = \mathbf{c}_1^\top \mathbf{\Phi}^{nmk}.$$

Second,

$$\begin{aligned} \langle h_2^1, \phi_{21}^{nmk} \rangle_{\mathcal{H}_2} &= \left\langle \sum_{l_1=-L}^L \sum_{l_2=-L}^L h_{2l_1 l_2}^1 \cdot e_{l_1}(t) \cdot e_{l_2}(s), \sum_{l_1=-L}^L a_{l_1}^{nm} e_{l_1}(t_k - t) \sum_{l_2=-L}^L a_{l_2}^{nm} e_{l_2}(t_k - s) \right\rangle_{\mathcal{H}_2} \\ &= \sum_{l_1=-L}^L \sum_{l_2=-L}^L h_{2l_1 l_2}^1 a_{l_1}^{nm} a_{l_2}^{nm} \cdot e_{l_1}(t_k) \cdot e_{l_2}(t_k) = \text{Tr}((\mathbf{H}_2^1)^\text{H} \mathbf{U}_2^{nmk}) \end{aligned} \tag{102}$$

and, therefore,

$$\langle h_2^1, \phi_{21}^{nmk} \rangle_{\mathcal{H}_2} + \langle h_2^2, \phi_{22}^{nmk} \rangle_{\mathcal{H}_2} + \langle h_2^3, \phi_{23}^{nmk} \rangle_{\mathcal{H}_2} = \text{Tr}(\mathbf{C}_2^\text{H} \mathbf{\Xi}^{nmk}). \tag{103}$$

Appendix 4: Proof of Lemma 3

For the m th trial, the output of the n th channel, by multiplying by the denominator of (20) with v^{nm} sampled at times t_k , we obtain

$$\begin{aligned} v^{nm}(t_k) &\left(1 + \int_{\mathbb{D}} h_1^2(s) u^{nm}(t_k - s) ds + \int_{\mathbb{D}^2} h_2^2(s_1, s_2) u^{nm}(t_k - s_1) u^{nm}(t_k - s_2) ds_1 ds_2 \right. \\ &\quad \left. + \int_{\mathbb{D}} h_1^3(s) (\mathcal{P}_1^o v^{nm})(t_k - s) ds \right) \end{aligned}$$

$$\begin{aligned}
 & + \int_{\mathbb{D}^2} h_2^3(s_1, s_2) (\mathcal{P}_1^o v^{nm})(t_k - s_1) (\mathcal{P}_1^o v^{nm})(t_k - s_2) ds_1 ds_2 \\
 & + \sum_{i=1}^N \int_{\mathbb{D}} h_1^{i4}(s) (\mathcal{P}_1^o v^{im})(t_k - s) ds \\
 & + \sum_{i=1}^N \sum_{j=1}^N \int_{\mathbb{D}^2} h_2^{ij4}(s_1, s_2) (\mathcal{P}_1^o v^{im})(t_k - s_1) (\mathcal{P}_1^o v^{jm})(t_k - s_2) ds_1 ds_2 \\
 & = b^1 + \int_{\mathbb{D}} h_1^1(s) u^{nm}(t_k - s) ds + \int_{\mathbb{D}^2} h_2^1(s_1, s_2) u^{nm}(t_k - s_1) u^{nm}(t_k - s_2) ds_1 ds_2. \tag{104}
 \end{aligned}$$

Finally, with $v^{nm}(t_k) = q^{nmk}$ and the notation in (44), we get

$$\begin{aligned}
 q^{nmk} & = b^1 + \int_{\mathbb{D}} h_1^1(s) \phi_{11}^{nmk}(s) ds + \int_{\mathbb{D}^2} h_2^1(s_1, s_2) \phi_{21}^{nmk}(s_1, s_2) ds_1 ds_2 \\
 & + \int_{\mathbb{D}} h_1^2(s) \phi_{12}^{nmk}(s) ds + \int_{\mathbb{D}^2} h_2^2(s_1, s_2) \phi_{22}^{nmk}(s_1, s_2) ds_1 ds_2 \\
 & + \int_{\mathbb{D}} h_1^3(s) \phi_{13}^{nmk}(s) ds + \int_{\mathbb{D}^2} h_2^3(s_1, s_2) \phi_{23}^{nmk}(s_1, s_2) ds_1 ds_2 \\
 & + \sum_{i=1}^N \int_{\mathbb{D}} h_1^{i4}(s) \phi_{14}^{inmk}(s) ds + \sum_{i=1}^N \sum_{j=1}^N \int_{\mathbb{D}^2} h_2^{ij4}(s_1, s_2) \phi_{24}^{ijnmk}(s_1, s_2) ds_1 ds_2, \tag{105}
 \end{aligned}$$

and, therefore,

$$\begin{aligned}
 b^1 & + \langle h_1^1, \phi_{11}^{nmk} \rangle_{\mathcal{H}_1} + \langle h_1^2, \phi_{12}^{nmk} \rangle_{\mathcal{H}_1} + \langle h_1^3, \phi_{13}^{nmk} \rangle_{\mathcal{H}_1} + \sum_{i=1}^N \langle h_1^{i4}, \phi_{14}^{inmk} \rangle_{\mathcal{H}_1} \\
 & + \langle h_2^1, \phi_{21}^{nmk} \rangle_{\mathcal{H}_2} + \langle h_2^2, \phi_{22}^{nmk} \rangle_{\mathcal{H}_2} + \langle h_2^3, \phi_{23}^{nmk} \rangle_{\mathcal{H}_2} + \sum_{i=1}^N \sum_{j=1}^N \langle h_2^{ij4}, \phi_{24}^{ijnmk} \rangle_{\mathcal{H}_2} = q_k^{nm} \tag{106}
 \end{aligned}$$

for all $m = 1, 2, \dots, M, k = 1, 2, \dots, T$, and $i, j, n = 1, 2, \dots, N$.

Appendix 5: Proof of Lemma 4

In addition to (87)–(94), let

$$h_1^{i4}(t) = \sum_{l=-L^o}^{L^o} h_{1l}^{i4} e_l^o(t), \quad i = 1, 2, \dots, N, \tag{107}$$

$$h_2^{ij4}(t) = \sum_{l_1=-L^o}^{L^o} \sum_{l_2=-L^o}^{L^o} h_{2l_1 l_2}^{ij4} e_{l_1}^o(t_1) e_{l_2}^o(t_2), \quad i, j = 1, 2, \dots, N, \tag{108}$$

and the $(2L^o + 1) \times 1$ vector

$$\mathbf{h}_1^{i4} = \begin{bmatrix} h_{1,-L^o}^{i4} \\ h_{1,-L^o+1}^{i4} \\ \vdots \\ h_{1,L^o}^{i4} \end{bmatrix}, \quad i = 1, 2, \dots, N, \tag{109}$$

and the $(2L^o + 1) \times (2L^o + 1)$ matrix

$$\mathbf{H}_2^{ij4} = \begin{bmatrix} h_{2,-L^o,L^o}^{ij4} & \cdots & h_{2,L^o,L^o}^{ij4} \\ \vdots & \ddots & \vdots \\ h_{2,-L^o,-L^o}^{ij4} & \cdots & h_{2,L^o,-L^o}^{ij4} \end{bmatrix}, \quad i, j = 1, 2, \dots, N. \tag{110}$$

Note that $\mathbf{H}_2^{ij4}, i, j = 1, 2, \dots, N$, are not necessarily Hermitian matrices. Further, by abuse of notation, let the $(2L + 1) \times 1$ vector

$$\mathbf{u}^{nmk} = \sqrt{S} \begin{bmatrix} a_{-L}^{nm} \cdot e_{-L}(t_k) \\ a_{-L+1}^{nm} \cdot e_{-L+1}(t_k) \\ \vdots \\ a_L^{nm} \cdot e_L(t_k) \end{bmatrix}, \quad \mathbf{U}_2^{nmk} = \mathbf{u}^{nmk} (\mathbf{u}^{nmk})^H, \tag{111}$$

where a_l^{nm} represents the coefficient of $u^{nm}(t)$ w.r.t. basis element e_l in \mathcal{H}_1 such that $u^{nm}(t) = \sum_{l=-L}^L a_l^{nm} e_l$, and let the $(2L^o + 1) \times 1$ vector

$$\mathbf{v}^{nmk} = \sqrt{S^o} \begin{bmatrix} d_{-L^o}^{nm} \cdot e_{-L^o}^o(t_k) \\ d_{-L^o+1}^{nm} \cdot e_{-L^o+1}^o(t_k) \\ \vdots \\ d_{L^o}^{nm} \cdot e_{L^o}^o(t_k) \end{bmatrix}, \quad \mathbf{V}_2^{n_1 n_2 mk} = \mathbf{v}^{n_2 mk} (\mathbf{v}^{n_1 mk})^H, \tag{112}$$

where d_l^{nm} represents the coefficient of $(\mathcal{P}_1^o v^{nm})$ w.r.t. basis element e_l^o in \mathcal{H}_1^o such that $(\mathcal{P}_1^o v^{nm}) = \sum_{l=-L^o}^{L^o} d_l^{nm} e_l^o$.

Additionally, we define the $(1 + 2(2L + 1) + (N + 1)(2L^o + 1)) \times 1$ vector

$$\begin{aligned} \Phi^{nmk} &= [1, (\mathbf{u}^{nmk})^T, -q^{nmk} (\mathbf{u}^{nmk})^T, -q^{nmk} (\mathbf{v}^{nmk})^T, -q^{nmk} (\mathbf{v}^{1mk})^T, \\ &\quad -q^{nmk} (\mathbf{v}^{2mk})^T, \dots, -q^{nmk} (\mathbf{v}^{Nmk})^T]^T, \end{aligned} \tag{113}$$

and the $(2(2L + 1) + (N^2 + 1)(2L^o + 1)) \times (2L + 1 + 2L^o + 1)$ matrix

$$\mathbf{E}^{nmk} = \begin{bmatrix} \mathbf{U}_2^{nmk} & \mathbf{0} \\ -q^{nmk} \mathbf{U}_2^{nmk} & \mathbf{0} \\ \mathbf{0} & -q^{nmk} \mathbf{V}_2^{nmk} \\ \mathbf{0} & -q^{nmk} \mathbf{V}_2^{11mk} \\ \mathbf{0} & -q^{nmk} \mathbf{V}_2^{12mk} \\ \vdots & \vdots \\ \mathbf{0} & -q^{nmk} \mathbf{V}_2^{NNmk} \end{bmatrix}. \tag{114}$$

Finally, we define the $(1 + 2(2L + 1) + (N + 1)(2L^o + 1)) \times 1$ vector

$$\mathbf{c}_1 = [b^1 \quad (\mathbf{h}_1^1)^T \quad (\mathbf{h}_1^2)^T \quad (\mathbf{h}_1^3)^T \quad (\mathbf{h}_1^{14})^T \quad (\mathbf{h}_1^{24})^T \quad \dots \quad (\mathbf{h}_1^{N4})^T]^T, \tag{115}$$

and the $(2(2L + 1) + (N^2 + 1)(2L^o + 1)) \times (2L + 1 + 2L^o + 1)$ matrix

$$\mathbf{C}_2 = \begin{bmatrix} (\mathbf{H}_2^1)^T & (\mathbf{H}_2^2)^T & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & (\mathbf{H}_2^3)^T & (\mathbf{H}_2^{14})^T & (\mathbf{H}_2^{124})^T & \cdots & (\mathbf{H}_2^{MN4})^T \end{bmatrix}^T. \quad (116)$$

With the notation above, the proof follows the same steps as the proof of Lemma 2.

Acknowledgements

Not applicable.

Funding

The research reported here was supported by AFOSR under grant #FA9550-16-1-0410.

Abbreviations

DNP, Divisive normalization processor; LMC, Large monopolar cells; LPF, Low-pass filter; MVP, Multi-input Volterra processor; RKHS, Reproducing kernel Hilbert space; RMS, Root mean square; SNR, Signal-to-noise ratio.

Availability of data and materials

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Ethics approval and consent to participate

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Authors' contributions

AAL, NHU, and YZ conceptualized and designed the study, performed analysis, and wrote the manuscript. NHU and YZ performed the simulations. All authors read and approved the final manuscript.

Authors' information

The authors' names are listed in alphabetical order.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 19 July 2019 Accepted: 28 January 2020 Published online: 12 February 2020

References

- Laughlin SB. Matching coding, circuits, cells, and molecules to signals: general principles of retinal design in the fly's eye. *Prog Retin Eye Res.* 1994;13(1):165–96. Available from <http://www.sciencedirect.com/science/article/pii/S0946294900094>.
- Rodieck RW. *The first steps in seeing.* London: Oxford University Press; 1998.
- de Bruyne M, Clyne PJ, Carlson JR. Odor coding in a model olfactory organ: The *Drosophila* maxillary palp. *J Neurosci.* 1999;19(11):4520–32. Available from <http://www.jneurosci.org/content/19/11/4520>.
- Olsen SR, Bhandawat V, Wilson RI. Divisive normalization in olfactory population codes. *Neuron.* 2010;66(2):287–99. Available from <http://www.sciencedirect.com/science/article/pii/S0896627310002497>.
- Song Z, Postma M, Billings S, Coca D, Hardie RC, Juusola M. Stochastic adaptive sampling of information by microvilli in fly photoreceptors. *Curr Biol.* 2012;22:1371–80.
- Meister M, Berry MJ II. The neural code of the retina. *Neuron.* 1999;22(3):435–50. [https://doi.org/10.1016/S0896-6273\(00\)80700-X](https://doi.org/10.1016/S0896-6273(00)80700-X).
- Scholl B, Latimer KW, Priebe NJ. A retinal source of spatial contrast gain control. *J Neurosci.* 2012;32(29):9824–30. Available from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3432019/>.
- Baccus SA, Meister M. Fast and slow contrast adaptation in retinal circuitry. *Neuron.* 2002;36:909–19.
- Wilson RI. Early olfactory processing in *Drosophila*: mechanisms and principles. *Annu Rev Neurosci.* 2013;36:217–41.
- Firestein S. How the olfactory system makes sense of scents. *Nature.* 2001;413:211. <https://doi.org/10.1038/35093026>.
- Clemens J, Ozeri-Engelhard N, Murthy M. Fast intensity adaptation enhances the encoding of sound in *Drosophila*. *Nat Commun.* 2018;9:134. Available from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5760620/>.
- Rabinowitz NC, Willmore BD, Schnupp JW, King AJ. Contrast gain control in auditory cortex. *Neuron.* 2011;70(6):1178–91. Available from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3133688/>.
- Kohn A, Movshon JA. Neuronal adaptation to visual motion in area MT of the macaque. *Neuron.* 2003;39(4):681–91.
- Rien D, Kern R, Kurtz R. Octopaminergic modulation of contrast gain adaptation in fly visual motion-sensitive neurons. *Eur J Neurosci.* 2012;36(8):3030–9.

15. Heuer HW, Britten KH. Contrast dependence of response normalization in area MT of the rhesus macaque. *J Neurophysiol.* 2002;88(6):3398–408.
16. Shapley RM, Victor JD. The effect of contrast on the transfer properties of cat retinal ganglion cells. *J Physiol.* 1978;285(1):275–98.
17. Shapley RM, Victor JD. How the contrast gain control modifies the frequency responses of cat retinal ganglion cells. *J Physiol.* 1981;318(1):161–79. Available from <https://physoc.onlinelibrary.wiley.com/doi/abs/10.1113/jphysiol.1981.sp013856>.
18. Shapley R, Enroth-Cugell C. Chapter 9 visual adaptation and retinal gain controls. *Prog Retin Eye Res.* 1984;3:263–346. Available from <http://www.sciencedirect.com/science/article/pii/0278432784900117>.
19. Chander D, Chichilnisky EJ. Adaptation to temporal contrast in primate and salamander retina. *J Neurosci.* 2001;21(24):9904–16.
20. Kim KJ, Rieke F. Temporal contrast adaptation in the input and output signals of salamander retinal ganglion cells. *J Neurosci.* 2001;21(1):287–99. Available from <https://www.jneurosci.org/content/21/1/287>.
21. Weber AI, Krishnamurthy K, Fairhall AL. Coding principles in adaptation. *Annu Re Vis Sci.* 2019;5(1):427–49. <https://doi.org/10.1146/annurev-vision-091718-014818>.
22. Shapley R. The contrast gain control and feedback within the retina. *IFAC Proc Vol.* 1982;15(4):489–94. 6th IFAC Symposium on Identification and System Parameter Estimation, Washington USA, 7–11 June. Available from <http://www.sciencedirect.com/science/article/pii/S1474667017630362>.
23. Yu Y, Lee TS. Dynamical mechanisms underlying contrast gain control in single neurons. *Phys Rev E.* 2003;68:011901. Available from <https://link.aps.org/doi/10.1103/PhysRevE.68.011901>.
24. Friederich U, Coca D, Billings S, Juusola M. Data modelling for analysis of adaptive changes in fly photoreceptors. In: *Neural information processing, lecture notes in computer science.* vol. 5863. 2009. p. 34–48.
25. Friederich U, Billings SA, Hardie RC, Juusola M, Coca D. Fly photoreceptors encode phase congruency. *PLoS ONE.* 2016;11(6):1–21. <https://doi.org/10.1371/journal.pone.0157993>.
26. Carandini M, Heeger DJ. Normalization as a canonical neural computation. *Nat Rev Neurosci.* 2012;13:51–62.
27. Heeger DJ. Normalization of cell responses in cat striate cortex. *Vis Neurosci.* 1992;9(2):181–97.
28. Schwartz O, Simoncelli EP. Natural sound statistics and divisive normalization in the auditory system. *Advances in Neural Information Processing Systems.* vol. 13. 2000.
29. Ohshiro T, Angelaki DE, DeAngelis GC. A normalization model of multisensory integration. *Nat Neurosci.* 2011;14(6):775–82. <https://doi.org/10.1038/nn.2815>.
30. Reynolds JH, Heeger DJ. The normalization model of attention. *Neuron.* 2009;61(2):168–85. <https://doi.org/10.1016/j.neuron.2009.01.002>.
31. Montijn J, Klink P, van Wezel RJA. Divisive normalization and neuronal oscillations in a single hierarchical framework of selective visual attention. *Front Neural Circuits.* 2012;6:22. Available from <https://www.frontiersin.org/article/10.3389/fncir.2012.00022>.
32. Sawada T, Petrov AA. The divisive normalization model of V1 neurons: a comprehensive comparison of physiological data and model predictions. *J Neurophysiol.* 2017;118(6):3051–91. <https://doi.org/10.1152/jn.00821.2016>. PMID: 28835531. Available from.
33. Cui Y, Wang YV, Park SJH, Demb JB, Butts DA. Divisive suppression explains high-precision firing and contrast adaptation in retinal ganglion cells. *eLife.* 2016;5:e19460. <https://doi.org/10.7554/eLife.19460>.
34. Wainwright MJ, Schwartz O, Simoncelli EP. In: Rao R, Olshausen B, Lewicki M, editors. *Natural image statistics and divisive normalization: modeling nonlinearity and adaptation in cortical neurons.* Cambridge: MIT Press; 2002. p. 203–22.
35. Lyu S. Divisive normalization: justification and effectiveness as efficient coding transform. In: Lafferty JD, Williams CKI, Shawe-Taylor J, Zemel RS, Culotta A, editors. *Advances in neural information processing systems.* vol. 23. Red Hook: Curran Associates; 2010. p. 1522–30. Available from <http://papers.nips.cc/paper/4065-divisive-normalization-justification-and-effectiveness-as-efficient-coding-transform.pdf>.
36. Coen-Cagli R, Schwartz O. The impact on midlevel vision of statistically optimal divisive normalization in V1. *J Vis.* 2013;13(8):13. <https://doi.org/10.1167/13.8.13>.
37. Lyu S, Simoncelli EP. Nonlinear image representation using divisive normalization. In: *Proceedings IEEE computer society conference on computer vision and pattern recognition.* 2008. p. 1–8. Available from <https://www.ncbi.nlm.nih.gov/pubmed/25346590>.
38. Beck JM, Latham PE, Pouget A. Marginalization in neural circuits with divisive normalization. *J Neurosci.* 2011;31(43):15310–9. Available from <https://www.jneurosci.org/content/31/43/15310>.
39. Rabinowitz NC, Willmore BDB, Schnupp JWH, King AJ. Spectrotemporal contrast kernels for neurons in primary auditory cortex. *J Neurosci.* 2012;32(33):11271–84. Available from <https://www.jneurosci.org/content/32/33/11271>.
40. Malo J, Bertalmio M. Appropriate kernels for divisive normalization explained by Wilson–Cowan equations. 2018. Preprint. [arXiv:1804.05964](https://arxiv.org/abs/1804.05964).
41. Rivera-Alba M, Vitaladevuni SN, Mishchenko Y, Lu Z, Takemura Sy, Scheffer L et al. Wiring economy and volume exclusion determine neuronal placement in the *Drosophila* brain. *Curr Biol.* 2011;21(23):2000–5. <http://www.sciencedirect.com/science/article/pii/S0960982211011468>.
42. Lazar AA, Ukani NH, Zhou Y. Sparse functional identification of complex cells from spike times and the decoding of visual stimuli. *J Math Neurosci.* 2018;8(1):2.
43. Fischbach KF, Ditttrich APM. The optic lobe of *Drosophila melanogaster*. I. A Golgi analysis of wild-type structure. *Cell Tissue Res.* 1989;258:441–75.
44. Nikolaev A, Zheng L, Wardill TJ, O’Kane CJ, de Polavieja GG, Juusola M. Network adaptation improves temporal representation of naturalistic stimuli in drosophila eye: II mechanisms. *PLoS ONE.* 2009;4(1):e4306. <https://doi.org/10.1371/journal.pone.0004306>.
45. Sterling P, Laughlin S. *Principles of neural design.* Cambridge: The MIT Press; 2015.
46. Matić T, Laughlin SB. Changes in the intensity-response function of an insect’s photoreceptors due to light adaptation. *J Comp Physiol.* 1981;145(2):169–77. <https://doi.org/10.1007/BF00605031>.
47. Juusola M, Hardie RC. Light adaptation in *Drosophila* photoreceptors: I. Response dynamics and signaling efficiency at 25°C. *J Gen Physiol.* 2001;117:3–25.

48. Zheng L, de Polavieja GG, Wolfram V, Asyali MH, Hardie RC, Juusola M. Feedback network controls photoreceptor output at the layer of first visual synapses in *Drosophila*. *J Gen Physiol*. 2006;127(5):495–510. Available from <http://jgprupress.org/content/127/5/495>.
49. Laughlin SB. Fly optic lamina as a guide to neural circuit design. 2nd ed. In: Shepard GM, Grillner S, editors. *Handbook of brain microcircuits*. London: Oxford University Press; 2017.
50. Lazar AA, Pnevmatikakis EA, Zhou Y. Encoding natural scenes with neural circuits with random thresholds. *Vis Res*. 2010;50(22):2200–12. Special Issue on Mathematical Models of Visual Coding.
51. Berline A, Thomas-Agnan C. *Reproducing kernel Hilbert spaces in probability and statistics*. Norwell: Kluwer Academic; 2004.
52. Lazar AA, Pnevmatikakis EA. Reconstruction of sensory stimuli encoded with integrate-and-fire neurons with random thresholds. *EURASIP J Adv Signal Process*. 2009;2009(1):682930. <https://doi.org/10.1155/2009/682930>.
53. Lazar AA, Zhou Y. Reconstructing natural visual scenes from spike times. In: *Proceedings of the IEEE*. vol. 102. 2014. p. 1500–19.
54. Lazar AA, Zhou Y. Volterra dendritic stimulus processors and biophysical spike generators with intrinsic noise sources. *Front Comput Neurosci*. 2014;8:95.
55. Lazar AA, Slutskiy YB. Channel identification machines for multidimensional receptive fields. *Front Comput Neurosci*. 2014;8:117.
56. Pumir A, Graves J, Ranganathan R, Shraiman BI. Systems analysis of the single photon response in invertebrate photoreceptors. *Proc Natl Acad Sci*. 2008;105(30):10354–9. Available from <https://www.pnas.org/content/105/30/10354>.
57. Rugh WJ. *Nonlinear system theory: Volterra/Wiener approach*. Baltimore: Johns Hopkins University Press; 1981.
58. VanLeeuwen M, Fahrenfort I, Sjoerdsma T, Numan R, Kamermans M. Lateral gain control in the outer retina leads to potentiation of center responses of retinal neurons. *J Neurosci*. 2009;29(19):6358–66. Available from <http://www.jneurosci.org/content/29/19/6358>.
59. Blakemore C, Carpenter RHS, Georgeson MA. Lateral inhibition between orientation detectors in the human visual system. *Nature*. 1970;228:37–9.
60. Olsen SR, Wilson RI. Lateral presynaptic inhibition mediates gain control in an olfactory circuit. *Nature*. 2008;452(7190):956–60. Available from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2824883/>.
61. Polat U, Sagi D. Lateral interactions between spatial channels: suppression and facilitation revealed by lateral masking experiments. *Vis Res*. 1993;33(7):993–9.
62. Christensen O. *Frames and bases: an introductory course*. Basel: Birkhäuser; 2008.
63. Fazel M, Hindi H, Boyd S. Rank minimization and applications in system theory. In: *Proceedings American control conference*. 2004. p. 3273–8.
64. Lazar AA, Psychas K, Ukani NH, Zhou Y. A parallel processing model of the *Drosophila* retina. *Neurokernel Request for Comments*. Neurokernel RFC #3. 2015 Aug. <https://doi.org/10.5281/zenodo.30036>.
65. van Hateren JH, van der Schaaf A. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc R Soc Lond B, Biol Sci*. 1998;265(1394):359–66.
66. Martin R. Approximation of Ω -bandlimited functions by Ω -bandlimited trigonometric polynomials. *Sampl Theory Signal Image Process*. 2007;6(3):273–96.
67. Ohshiro T, Angelaki DE, DeAngelis GC. A neural signature of divisive normalization at the level of multisensory integration in primate cortex. *Neuron*. 2017;95(2):399–411.e8. Available from <http://www.sciencedirect.com/science/article/pii/S0896627317305950>.
68. Spratling MW. A review of predictive coding algorithms. *Brain Cogn*. 2017;112:92–7. Available from <http://www.sciencedirect.com/science/article/pii/S027826261530035X>.
69. Spratling MW, De Meyer K, Kompass R. Unsupervised learning of overlapping image components using divisive input modulation. *Comput Intell Neurosci*. 2009;2009:381457.
70. Srinivasan MV, Laughlin SB, Dubs A. Predictive coding: a fresh view of inhibition in the retina. *Proc R Soc Lond B, Biol Sci*. 1982;216(1205):427–59.
71. Lazar AA, Yeh CH. A Parallel Processing Model of *Drosophila* Olfactory Sensory Neurons and Its Biological Validation. *Neurokernel Request for Comments* Neurokernel RFC #10. 2017 Dec.
72. Natan RG, Carruthers IM, Mwilambwe-Tshilobo L, Geffen MN. Gain control in the auditory cortex evoked by changing temporal correlation of sounds. *Cereb Cortex*. 2017;27(3):2385–402. <https://doi.org/10.1093/cercor/bhw083>.
73. Allison JD, Smith KR, Bonds AB. Temporal-frequency tuning of cross-orientation suppression in the cat striate cortex. *Vis Neurosci*. 2001;18(6):941–8.
74. Geisler WS, Albrecht DG. Cortical neurons: isolation of contrast gain control. *Vis Res*. 1992;32(8):1409–10. Available from <http://www.sciencedirect.com/science/article/pii/004269899290196P>.
75. Priebe NJ, Ferster D. Mechanisms underlying cross-orientation suppression in cat visual cortex. *Nat Neurosci*. 2006;9(4):552. <https://doi.org/10.1038/nn1660>.