Nonlinear convergence preserves information

Gabrielle J. Gutierrez[1,✉], **Fred Rieke**[2,*], **and Eric Shea-Brown**[1,*]

[1]Department of Applied Mathematics, University of Washington, Seattle, WA
[2]Department of Physiology and Biophysics, University of Washington, Seattle, WA
[*]co-senior authors

Neural circuits are structured with layers of converging and diverging connectivity, and selectivity-inducing nonlinearities at neurons and synapses. These components have the potential to hamper an efficient encoding of the circuit inputs. Past computational studies have optimized the nonlinearities of single neurons, or connection weights of networks, to maximize encoded information, but have not grappled with the simultaneous impact of compressive circuit structure and neural response functions for efficient coding. Our approach is to compare circuits with different combinations of convergence, divergence, and nonlinear responses to discover how interactions between them affect coding efficiency. We find that a combination of divergence, convergence, and nonlinear responses preserves the most information, despite the compressive loss induced by both the convergence and the nonlinearities individually. These results show that the combination of selective nonlinearities and a compressive architecture - both elements that induce lossy compression when acting separately - can promote efficient coding.

Correspondence: *ellag9@uw.edu*

Sensory systems by necessity compress a wealth of information gathered by receptors into the smaller amount of information needed to guide behavior. In many systems, this compression occurs via common circuit motifs - namely convergence of multiple inputs into a single neuron and divergence of inputs to multiple parallel pathways (Jeanne and Wilson, 2015). Here we investigate how these motifs work together to dictate how much and what information is retained in compressive neural circuits.

These issues are highly relevant to signaling in the retina, because the bottleneck provided by the optic nerve means that considerable compression occurs prior to transmission of signals to central targets (Zhaoping, 2006; Nirenberg, et al., 2001). Receptive field subunits are a key feature of the retina?s compressive circuitry. Multiple bipolar cells converge onto a single ganglion cell - forming functional subunits within the receptive field of the ganglion cell (Demb and Singer, 2015; Enroth-Cugell and Robson, 1966). Ganglion cell responses can often be modeled as a linear sum of a population of nonlinear subunits. These subunit models have been used to investigate center-surround interactions (Enroth-Cugell and Freeman, 1987; Hochstein and Shapley, 1976; Barlow, 1953; Turner et al., 2018) and to determine how cells integrate spatial inputs (Enroth-Cugell and Robson, 1966; Turner and Rieke, 2016; Hartline, 1940; Freed and Sterling, 1988).

While it is clear that subunit coding is inherently compressive, it is not known whether this architecture subserves an efficient code. Since the 1950s, information theory has been used to quantify the amount of information that neurons encode (Shannon and Weaver, 1998; Cover and Thomas, 2006; Bell and Sejnowski, 1995). The efficient coding hypothesis proposes that the distribution of neural responses should be one that is maximally informative about the inputs (Attneave, 1954; Barlow, 1961). Take the example of a stimulus variable, such as luminance, where the brightness level is encoded by the number of spikes in the response. An input/output mapping where most of the possible brightness levels are encoded by the same response, i.e. the same number of spikes, ambiguates bright and dim inputs and provides very little information. However, distributing the responses such that they optimally disambiguate inputs leads to the most efficient code.

In this paper, we explore how a combination of divergence of inputs into multiple parallel pathways and convergence in each pathway via nonlinear receptive field subunits impacts coding efficiency. We find that the convergence of nonlinear subunits reduces the loss of information despite the selectivity of the nonlinearities. By reformatting the inputs, nonlinear subunits enable a circuit with divergent pathways to preserve more information than a circuit without any nonlinearities.

DRAFT

Results

We start by quantifying the effect of common circuit motifs, alone and in combination, on coding efficiency. We then explore, geometrically, how nonlinear subunits shape the response distribution to gain intuition as to how they can lead circuits to retain more information. Finally, we explore the implications of nonlinear subunits for which stimulus properties are encoded.

Common circuit components are lossy or inefficient. Our goal is to understand how the divergence of inputs and the convergence of nonlinear subunits impacts the retina's ability to efficiently encode spatial inputs. In particular, we are concerned with the impact of selective nonlinearities on efficient coding. We use Shannon's entropy to describe the maximum amount of information that a distribution of responses could contain about its inputs. We consider deterministic circuits in which the mutual information between the stimulus and response reduces to the entropy of the response. Specifically, we use discrete entropy to compare the information content of continuous distributions of responses generated by different model circuits. The parameters of the discretization were chosen so that distinctions between different distributions could be reasonably interpreted (see Methods).

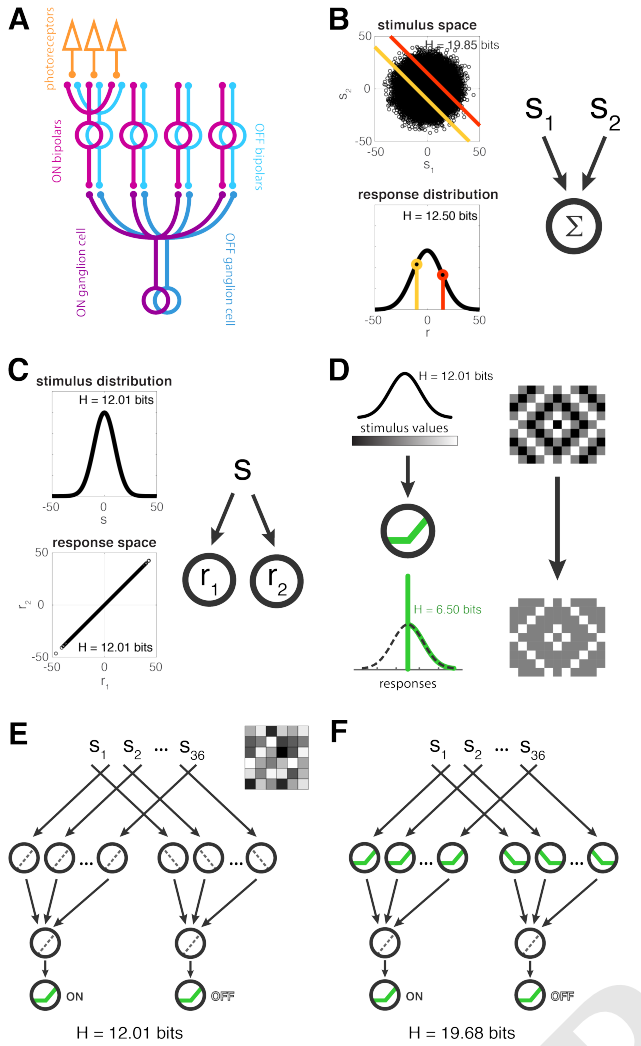

Fig. 1. Neural circuits are composed of inherently lossy components. (A) schematic of retina circuit with its convergent and divergent structure. (B) Converging two inputs results in ambiguities. A 2D stimulus space is reduced to a 1D response space in which one response represents all stimuli along an isoline where s1 + s2 = constant. (C) Diverging to two outputs can produce redundancies. (D) Nonlinear transformation of a gaussian distributed input with a ReLU can distort the distribution, producing a compressed response in which some portion of the stimulus information is discarded. (E-F) Convergent, divergent circuits with (E) nonlinear outputs, or (F) nonlinear subunits. Subunit responses are weighted by $1/\sqrt{36}$. Example stimulus image is shown.

The retina is organized in layers that converge and diverge (Fig. 1A), ultimately leading to the compression and "reformatting" of a high-dimensional visual input into a lower dimensional neural code that can be interpreted by the brain. In addition, nonlinear responses abound in the neurons that compose these layers. These mechanisms may complicate the ability of the circuit to retain information. For example, two converging inputs can result in ambiguities. With linear convergence, the ability to distinguish the stimulus combinations that sum to the same value is lost and hence this is a form of lossy compression (Fig. 1B). The entropy of the full two-dimensional stimulus (Fig. 1B, top) is 19.85 bits - meaning that a given point in the stimulus space provides 19.85 bits of information about the identity of the stimulus (with our choice of bin size, see methods). The entropy of the convergent response is smaller (12.50 bits; Fig. 1B, bottom), thus indicating ambiguity in the stimulus identity.

Diverging motifs are another common neural circuit construction. In the example shown in Figure 1C, the divergent responses are identical and the entropy of the 2-dimensional response space (H = 12.01 bits) is the same as the entropy of the 1-dimensional stimulus distribution shown in the top plot (H = 12.01 bits). This demonstrates that diverging an input into two neurons may produce an inefficient neural architecture by producing redundant (i.e. correlated) signals.

Similarly to convergence, nonlinear transformations can lead to loss of information by introducing ambiguities. Take the example of a rectified-linear transformation that is thresholded at zero and is therefore selective for positive inputs (Fig. 1D). It is a non-invertible nonlinearity where half of the stimulus distribution is encoded by the thresholded response and half is mapped to an output of 0. Therefore, this nonlinearity induces lossy compression: the information that would distinguish these thresholded stimuli has been irretrievably discarded. The entropy of the rectified-linear (ReLU) response (H = 6.50 bits) is nearly half of that for the stimulus distribution (H = 12.01 bits).

Each of the common circuit motifs described above is inefficient or discards information when considered in isolation (Figs. 1A-D). How much information can a neural circuit with all of these components retain? We constructed a model circuit that compresses a high-dimensional spatial input into a low-dimensional output. It has a 36-dimensional input structure that diverges along two pathways, an ON and an OFF pathway, each culminating in a single output neuron. The inputs to each output neuron come from a layer of subunits - the building blocks for the receptive field structure of the output neuron. Each subunit receives input from one of the N stimulus inputs that compose a stimulus image where each pixel is independently drawn from a gaussian distribution. Within each pathway, the normalized subunit responses linearly sum at the output neuron and are then rectified.

The ON and OFF output responses lie in a 2-dimensional space, and form a low-dimensional representation of the N-dimensional input. We compute the entropy of the 2-dimensional output response after showing many stimulus samples to the circuit. The circuit in Figure 1E has linear subunits. Its output has 12.01 bits of entropy. The circuit in Figure 1F has nonlinear subunits and its output response has 19.68 bits of entropy. The greater entropy of the nonlinear subunit circuit is counterintuitive because the nonlinear elements considered in isolation lead to a loss of information (Fig. 1D). This motivated us to consider how each component or structure interacts with the other components to produce this result. Rather than optimizing the circuit weights or input biases, our goal is to explore the extent to which the information retention of a circuit can benefit from selective nonlinearities that operate independently on signals in each subunit. We next investigate how convergence interacts with subunit nonlinearities.

Lossy nonlinear subunits benefit from convergence. To understand the joint impact of nonlinear subunits and con-

DRAFT

vergent connectivity on encoded information, we examined circuit configurations with a single pathway, i.e. without divergence (Fig. 2). Pathways with two subunits are used to visualize the input and response spaces. Stimuli that sum to the same value (example highlighted with dark purple in the top plot of Fig. 2A) are represented by the same response in the circuit pathway with linear subunits because the subunits do not transform the inputs (Fig. 2A, left, 3rd and 4th rows). The nonlinear subunits transform the stimulus space such that all points are compressed into a single quadrant (Fig. 2A right, 2nd row). Summing the nonlinear subunits (Fig. 2A, right, 3rd row) allows the ambiguous stimuli to have a more distributed representation in the output response - meaning that they are represented more distinctly by the nonlinear subunits pathway than the pathway with linear subunits.

For a configuration with a single subunit, the linear and nonlinear subunits circuits would have identical output responses so long as there remained an output nonlinearity. The 2-subunit example in Figure 2A showed improved information transmission in the case of two nonlinear subunits, and prompted us to ask whether there would be a continued improvement with additional nonlinear subunits. We computed the entropy of the output responses for the linear and nonlinear subunit configurations in Figure 2A for a range of subunit dimensions. With increasing subunit dimension, more subunit responses are compressed into the output response. To observe a relative change in entropy, the subunits were normalized (see Methods). The distribution of output responses for the nonlinear subunits pathway qualitatively changes with the number of subunits (Fig. 2B). With few subunits, the output response distribution resembles the truncated gaussian seen for the rectified output response in Figures 1D and 2A. With increasing numbers of subunits, the output response distribution approximates a gaussian (due to the central limit theorem) with a mean that shifts towards more positive values (Fig. 2B). The nonlinear subunits are each individually impacted by their own rectification, but summing the nonlinear subunits broadens and shifts the output distribution away from the nonlinear subunit threshold. In other words, each nonlinear subunit is strongly impacted by its own threshold, but collectively, the nonlinear subunits pull the output response distribution away from the threshold, not only at the output, but also at the level of the subunits. As we will show later, this distribution shift cannot be reproduced by simply adding a bias to the inputs or to the summed linear subunits.

The entropy for the nonlinear subunits pathway increases with increasing subunit dimension (Fig. 2C). It saturates before reaching the entropy of the converged linear subunits (gray dashed line) - which is to say that, although increasing convergence improves the information retention of nonlinear subunits, the entropy of the converged nonlinear subunits is bounded by the entropy of the converged subunits that do not undergo rectification. The converged linear subunits demonstrate what happens when there are no selective nonlinearities: the nonlinear subunits only encode positive inputs whereas the linear subunits encode positive and negative inputs. However, when the summation of the linear subunits is

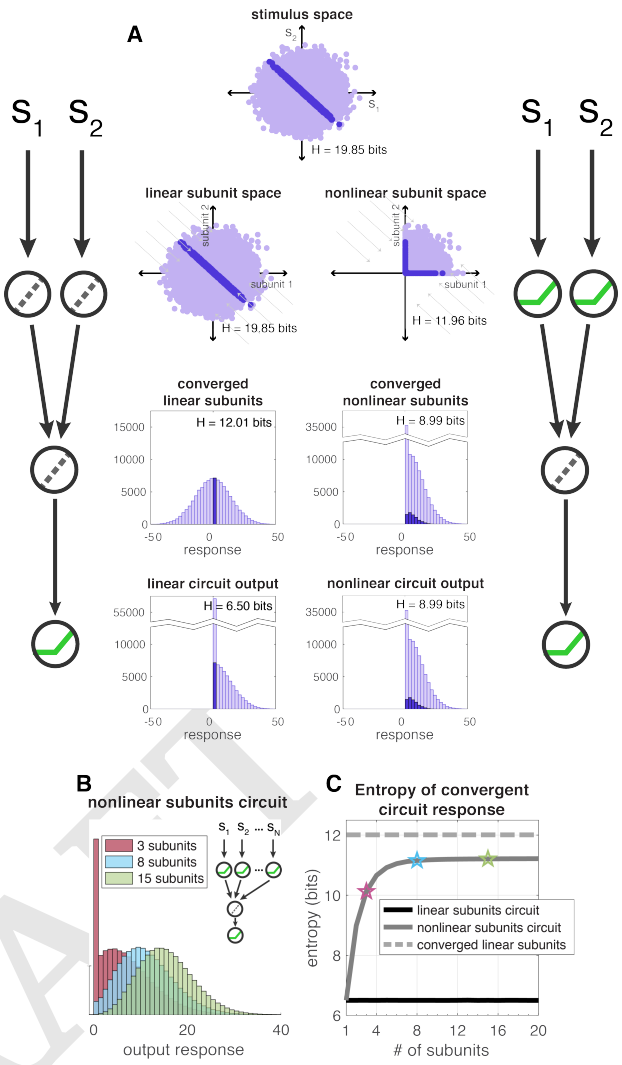

Fig. 2. (A) The encoding of the stimulus space (top) within each circuit layer of a 2-subunit pathway configuration without divergence. Subunits, 2nd row (gray arrows illustrate the effect of summing the subunits); linearly summed response, 3rd row; nonlinear output response, 4th row. Left, linear subunits pathway; right, nonlinear subunits pathway. The output nonlinearity does not have an additional effect on the summed nonlinear subunits without noise. (B) Compression with increasing numbers of converging subunits. The subunit to output weights are normalized so that each subunit has a weight of $1/\sqrt{N}$, thereby preserving the total variance of the summed linear subunits. The inputs [s1,s2,...,sN] are independently drawn from a gaussian distribution. Response histograms for the nonlinear subunits configurations are shown for 3, 10, and 25 subunits. (C) Entropy of output response as a function of number of subunits where subunits are normalized (black solid, entropy of nonlinear output response from linear subunits circuit; gray solid, entropy of nonlinear output response from nonlinear subunits circuit; gray dashed, entropy of the linear summation of linear subunits). Stars correspond to their respective distributions in (B).

then followed by a nonlinear rectification at the output, the response entropy drops to 6.50 bits (Fig. 2C, black line) and does not increase beyond that regardless of the number of convergent subunits.

Figure 2 shows that convergence reduces the impact of the subunit nonlinearity on the entropy of the circuit. The two principal effects of these subunit nonlinearities are to shift responses so that the output nonlinearity will not truncate them, and to differentiate stimuli that are made ambiguous by linear convergence (i.e. that sum to the same value). We wondered whether this effect of nonlinear convergence was sufficient to

DRAFT

explain why the divergent nonlinear subunits circuit in Figure 1F has higher entropy than the divergent linear subunits circuit (Fig. 1E). We next explore the impact of divergence on information coding with nonlinear subunits.

Divergent circuit structure leverages selectivity of nonlinear subunits. To understand the combined impact of divergence, convergence, and nonlinearities, we present a geometrical exploration of the transformations that take place in the different layers of the circuit with either linear or nonlinear subunits. Our demonstration uses circuits with two input dimensions to facilitate the visualization of the stimulus and subunit spaces. Figure 3 shows a 2-dimensional stimulus space that displays each stimulus quadrant in a different color (top, Figs. 3A,D). The points in all subsequent plots are color-coded by the stimulus quadrant from which they originate. The linear ON subunit space (Fig. 3A, 2nd row, left) is identical to the stimulus space because no transformation or compression has taken place through the linear subunits. The OFF subunits receive a negative copy of the same stimulus that the ON subunits receive which rotates the stimuli by 180 degrees (Fig. 3A, 2nd row, right). When the linear subunits are converged within their respective pathways, the ON and OFF responses are compressed onto a diagonal line because they are anti-correlated (Fig. 3B). This emphasizes the fact that the ON and OFF linear subunits do not have stimulus selectivities. When the output nonlinearities are applied, this linear manifold is folded into an L-shape (Fig. 3C).

The entropy for the output response of the linear subunits circuit with diverging pathways (H = 12.01 bits, Fig. 3C,F, black) is higher than it was with just a single pathway (H = 6.50 bits, Fig. 2C, black). However, it is only increased enough to match the entropy of a single pathway response without any nonlinearities in either the subunits or the output (H = 12.01 bits, Fig. 2C, grey dashed). In other words, the OFF pathway in the linear subunits circuit with output nonlinearities (Fig. 3C) encodes the information discarded by the ON pathway, but it does not enable the divergent linear subunits circuit in Figure 3C to do any better than the convergence of only ON linear subunits (Fig. 2C, dashed line). This is because the linear subunits do not select for anything specific and nothing is lost to selectivity; instead the loss of entropy occurs from convergence. Furthermore, when the convergence of the ON linear subunits is followed by a nonlinearity, the positive-summing stimuli are selected. A divergent OFF pathway selects the negative-summing stimuli that the ON pathway discards. Thus, the divergent linear subunits circuit recovers what is lost by the output nonlinearities, but not what is lost by convergence.

To understand how the nonlinear subunits produce an additional advantage in a divergent circuit, observe how the nonlinear subunits transform the inputs (Fig. 3D). Unlike the linear subunits, the stimulus undergoes a transformation within the nonlinear subunits layer, producing a complimentary compression for the ON and OFF pathways. When these subunits converge in their respective pathways (Fig. 3E), the output response has some similarities to that for the linear subunits circuit (Fig. 3C). The L-shaped manifold is still present, but the points representing the stimulus inputs with mixed sign have been projected off it. By virtue of having these points leave the manifold and fill out the response space, entropy is increased. In fact, as more nonlinear subunits converge in a divergent circuit, the entropy continues to increase until saturation (Fig. 3F).

The nonlinear subunits circuit does nothing to save the dually positive (blue quadrant) or dually negative (yellow quadrant) stimuli from information loss by convergence. Those are ultimately encoded in the same way as by the linear subunits circuit. The advantage conferred by the divergent nonlinear subunits is to preserve the variance among the mixed sign stimuli, not only within a single pathway, but also across ON and OFF pathways (hence why adding a bias to the summed linear subunits will not give the same results as the nonlinear subunits circuit). As the stimulus dimension is increased, the mixed sign stimuli make up a larger and larger proportion of all stimuli, resulting in the increasing advantage of the nonlinear subunits circuit.

The output nonlinearities have the effect of decorrelating the ON/OFF output response in the linear subunits circuit, while for the nonlinear subunits circuit, it is the nonlinear subunits themselves that decorrelate the output response (correlation coefficients: linear response = -1, linear subunits, nonlinear output circuit = -0.4670, nonlinear subunits circuit = -0.4669). Indeed, although the output nonlinearity decorrelates the ON/OFF outputs of the linear subunits circuit, this decorrelation does not produce any gains in entropy relative to the linear subunits circuit before output nonlinearities are applied. Furthermore, the ON/OFF responses of the nonlinear subunits circuit are as decorrelated as for the linear subunits circuit, but unlike the linear subunits circuit, it experiences an entropy gain over the converged linear subunits alone. The additional entropy conferred by divergence for the nonlinear subunits circuit is due to how the nonlinear subunits decorrelate the ON and OFF pathways before convergence, and not merely the fact that those outputs have been decorrelated. It is this step that pulls responses off of the linear manifold in the output response space leading to an increase in response entropy.

To determine the optimal nonlinear thresholds, we swept through a range of thresholds for ON and OFF subunits in a divergent, convergent circuit with two inputs. Very low thresholds approximate linear functions while high thresholds are extremely rectifying. We found that the optimal combination of ON/OFF subunit thresholds cross at zero (see supplemental figure). These zero-crossing nonlinearities are used for all other figures.

Increased response entropy could reflect an increased precision in encoding the same stimulus features, or the encoding of new stimulus features. We show next that nonlinear subunits create sensitivity to stimulus features that are not encoded with linear subunits.

Nonlinear subunits circuit encodes both mean and contrast information. To determine whether increases in entropy accompany an encoding of new stimulus features, we visualized the stimulus and response spaces for the two cir-

DRAFT

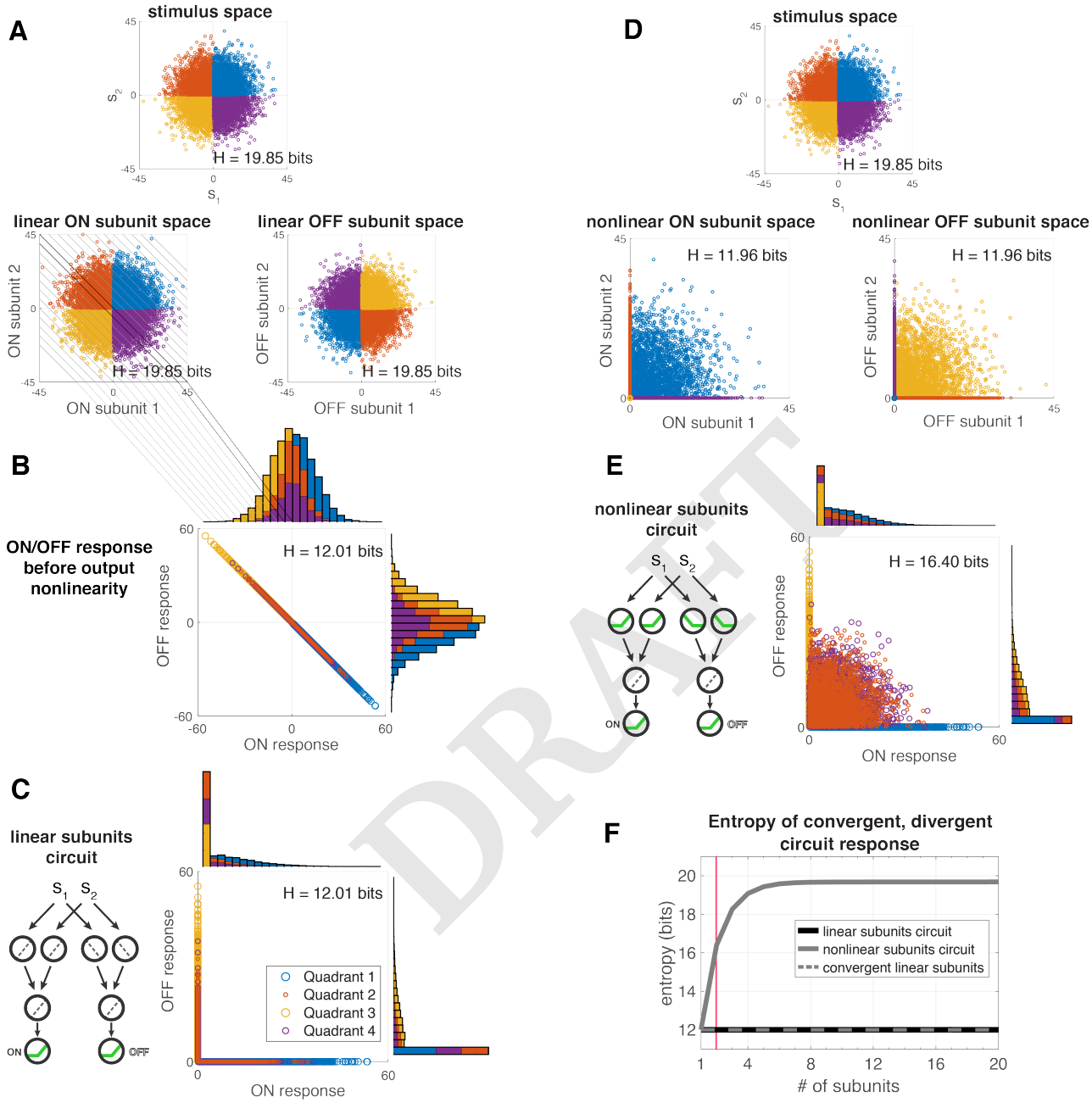

Fig. 3. Visualization of stimulus/response mappings at each level of a convergent, divergent circuit with two inputs, two subunits for each pathway (ON and OFF pathways), and a nonlinear output neuron for each pathway. (A) The spaces are color-coded according to the 4 stimulus quadrants (top). The 2D stimulus space maps onto a 2D linear subunit space for each pathway (second row, left: ON; right: OFF). The subunit spaces are shown before subunit normalization. (B) The response space is shown for the linear sum of subunits before the nonlinearity is applied and (C) after the nonlinear output response. (D) The 2D stimulus space maps onto a 2D **nonlinear** subunit space for each pathway (second row, left: ON; right: OFF). (E) The output response space for the nonlinear subunits circuit. Note that the output response before the output nonlinearity is applied is the same as when the output nonlinearity is applied for the circuit with nonlinear subunits. (F) Entropy of the output response for convergent, divergent circuits with increasing input and subunit dimension (subunits are normalized as before). Black, linear subunits circuit in C; gray solid, nonlinear subunits circuit in E; gray dashed, linear response of linear subunits circuit in B.

DRAFT

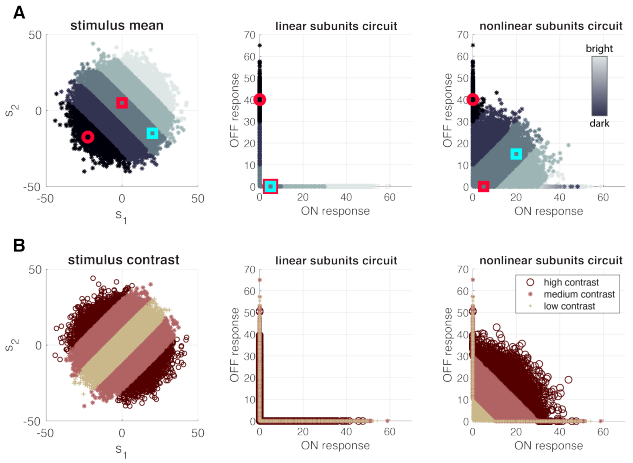

Fig. 4. Mean and contrast encoding of convergent, divergent circuits from Figure 3 (2 subunits, ON and OFF output). (A) Visualization of the stimulus mean representation and response spaces. The red square is an arbitrary reference point. In the stimulus space, the cyan square has the same mean luminance as the red square but a different contrast, while the red circle has the same contrast as the red square but a different mean luminance. (B) Visualization of the stimulus contrast representation and response spaces.

cuit configurations. The stimulus inputs are assumed to represent relative luminance values and the distributions are the same as before. We chose two basic features of visual stimuli to investigate: mean relative luminance and contrast. In Figure 4A, the stimulus space is color-coded by bands of mean luminance levels. In both of the response spaces in Figure 4A, a banded structure is preserved, indicating that there is a separation of the mean luminance levels within the response spaces for the circuits with linear subunits and with nonlinear subunits. This is emphasized by the separation of the red square and red circle in the response spaces (which occupy different bands in the stimulus space). However, the red and cyan squares overlap each other in the output response space for the linear subunits circuit (Fig. 4A, middle). These two symbols represent stimuli with the same mean luminance but different contrasts. Only the nonlinear subunits circuit represents these stimuli with distinct output responses.

The nonlinear subunits circuit encodes stimulus contrast. The bottom row explicitly shows how contrast is encoded by the different circuits (Fig. 4B). The stimulus space is color-coded for three contrast levels (Fig. 4B, left). The highest contrast areas of the space are in the mixed sign quadrants. The representations for low, medium, and high contrast stimuli overlap each other in the output response space of the linear subunits circuits (Fig. 4B, middle). However, there is separation of these contrast levels in the output response space of the nonlinear subunits circuit (Fig. 4B, right). The nonlinear subunits circuit encodes both mean and contrast information whereas the linear subunits circuit only encodes mean luminance.

Discussion

We set out to understand how common neuron and circuit properties impact encoded information. We asked how much stimulus information a compressive circuit could preserve if it also had selective nonlinear subunits. In a circuit like the retina, inputs diverge to distinct cell types while neurons receive converging inputs from many presynaptic neurons. This ultimately leads to a compression and reorganization of the visual inputs via convergence and divergence. To determine the impact of these common circuit properties on information coding, we built a circuit model and compared the entropy of linear and nonlinear subunit configurations. We found that the circuit with nonlinear subunits preserves more information than the circuit with linear subunits despite the fact that the nonlinear subunits, due to their selectivity, are compressive themselves. Divergence, convergence, and non-invertible nonlinear signal transformations each have a negative impact on efficiency or information individually. However, when arranged together they can mitigate the loss of information that is imposed by the reduction in dimension from inputs to outputs.

Selectivity versus efficiency. Nonlinearities can have different functions in a neuron. Nonlinear transformations can induce selectivity in that they can cause a neuron to encode a very particular aspect of the stimulus or its inputs (Gollisch, 2013; Gollisch and Meister, 2010). Nonlinearities can otherwise optimize efficiency by maximizing the entropy of the response distribution (Laughlin, 1981). The ReLU nonlinearity that we used does not maximize the response entropy of the individual neuron that receives gaussian-distributed inputs, but it does enforce a strict selectivity for inputs above threshold. Selectivity, however, is in conflict with efficient coding in that discarding information is a poor way to maximize it. The selective coding of features is often conflated with redundancy reduction, but it is important to make a distinction in the context of efficient coding - where a redundancy reducing code is reversible and is expected to maximize information about the stimulus (Barlow, 2001). Selectivity indicates that some stimulus information will be irreversibly discarded. The existence of selective cell types that compute different aspects of the visual scene appears to confound an efficient coding framework (Pitkow and Meister, 2012). Yet, properties of selectivity are crucial to the functions of a diverse array of cell types, such as object-selective cells in medial temporal lobe (Ison, et al., 2011), face-selective cells in the inferior temporal cortex (Eifuku et al., 2004; Hasselmo, et al., 1989), and direction-selective cells, orientation-selective cells, and edge detector cells in the retina (Sanes and Masland, 2015). Furthermore, many cell types in the retina and other circuits have both an ON and an OFF variant (Gjorgjieva et al., 2014), indicating that this kind of ON/OFF selectivity has an important role to play in sensory processing.

Although it has been shown that a cumulative gaussian function encodes maximal information for gaussian inputs to a single unit, within a population one must carefully consider the weights in order to preserve information (Sharpee and Berkowitz, 2019). Rather than optimizing the subunit weights for information maximization, our approach was to determine whether selective nonlinearities could provide any advantage since the brain, after all, has limited known mechanisms for optimizing information. We chose non-invertible

DRAFT

nonlinearities that exhibit generic selectivity (ON or OFF) and that induce lossy compression of stimuli with no inherent spatial statistical redundancy to exploit. Despite these properties, we found that, within a convergent, divergent circuit architecture, selective subunit nonlinearities preserved more information than lossless linear subunits. This increase comes from a reformatting of the stimulus distribution in a manner that reduces the ambiguities produced by the convergence of multiple inputs (Fig. 2). This reformatting facilitates the encoding of multiple stimulus features in Figure 4. Thus, in the circuits we study here, efficient coding can be achieved with selective nonlinear components. We note that even invertible nonlinearities, when followed by noise, will become difficult to invert and will thus behave like the non-invertible nonlinearity studied here.

Contribution of divergence to information maximization. Past work (Brinkman et al., 2016; Gjorgjieva et al., 2014; Kastner et al., 2015) has explored the optimal nonlinearities and configurations of divergent circuits - though none have explicitly explored convergence or nonlinear compression as we did. Brinkman, et al, found that the optimal nonlinearities are highly dependant on the placement and magnitude of noise in the circuit. Our study did not include noise; however, for low noise conditions, they found that the optimal sigmoidal nonlinearities cross at their "lower bend", similar to the crossing at threshold for the ReLU nonlinearities in our study. In their study, noise had the effect of changing the optimal nonlinearities, but it is unclear whether changing the convergence or divergence in a circuit structure like ours would be more effective for maximizing information than changing the properties of the nonlinearity. Our future studies will investigate how noise impacts the efficiency of the nonlinear subunit circuit.

We find that the efficiency of divergent circuits is enhanced by nonlinearities that decorrelate the outputs, as others have found (Gjorgjieva, et al., 2014; Kastner, et al., 2015). Indeed, our findings show that divergence resulted in efficiency gains for both the linear and nonlinear subunits circuits (compare the entropies in the solid lines for the single pathway configurations in Figure 2C to those for the corresponding divergent circuits in Figure 3F). Our findings also show that nonlinearities facilitate decorrelation among the ON and OFF outputs when comparing across divergent circuits (compare the convergent linear subunits in Figure 3B [corr. coef. = -1] to the circuits in Figure 3C and 3E [corr. coeff. = -0.4669 and -0.4670, respectively]). Despite achieving the same amount of decorrelation, any efficiency gains are dependent on the manner in which this decorrelation is achieved. For the linear subunits circuit, the output nonlinearities induced decorrelations among the outputs but did not result in an increase in entropy beyond that for the fully linear response. The decorrelations due to the nonlinear subunits did, however, lead to an increase in entropy relative to a convergent linear subunit response without any nonlinearities.

Nonlinearities are known to have a special role in decorrelating and separating signals. Pitkow and Meister (2012) showed that nonlinear responses in ganglion cells have more of an effect on decorrelating their responses than their center-surround receptive field properties. However, as they point out, weak correlation is not necessarily weak dependence. In the divergent, convergent circuits in Figure 3, putting nonlinearities in either the outputs or the subunits decorrelates the outputs by the same amount relative to the response without any nonlinearities where the outputs are perfectly anti-correlated. However, only by placing the nonlinearities in the subunits does a gain in entropy result relative to the scenario in which there are no nonlinearities in the circuit. Bell and Sejnowski (1995) showed that nonlinearities have the effect of reducing redundancy between output neurons by separating statistically independent parts of the inputs. Following that, it was shown that the efficient encoding of natural signals is facilitated by a nonlinear decomposition whose implementation is similar to the nonlinear behaviors observed in neural circuits through divisive normalization (Schwartz and Simoncelli, 2001). Our study contributes to this body of work by showing how nonlinear processing contributes to a more informative representation that has lower dimension than the inputs.

Implications for artificial neural networks. Although mean and contrast are elementary features of visual inputs, the striations seen in the response space in Figure 4 (nonlinear subunits circuit, right plots) reinforce the idea that hidden nonlinear neural units can facilitate the categorization of stimulus features (DiCarlo, Zoccolan, Rust, 2012). Indeed, in our study simply inserting nonlinear subunits with uniform weights immediately produced a representation that qualitatively enables linear classification or decoding of the mean and contrast levels of the input.

Feedforward artificial neural networks (ANNs) were inspired by the layered organization of biological neural networks. Neural units have activation functions, or static nonlinearities, that transform inputs. Rectified Linear Units, a.k.a. ReLU activation functions, such as those used in our nonlinear neural units, enforce a strict selectivity for inputs above threshold; whereas smooth nonlinearities implement a less rigid selectivity, if at all. In both cases, selectivity is dependent on the bias and weight parameters, which can be adjusted by learning, to offset the nonlinearity such that it truncates the input distribution to various degrees. The ReLU frequently has the best performance among other nonlinear activation functions (Glorot et al., 2011; LeCun et al., 2015) in tasks ranging from the discrimination of handwritten digits to restricted Boltzmann machines (Nair and Hinton, 2010). The findings presented here of the information preserving capabilities of a selectivity-inducing nonlinear activation within an architecture that is reminiscent of a feedforward ANN complement our knowledge of the ReLU's favorable performance in machine learning and the remarkable classification capabilities of ANNs.

Materials and Methods

We used Shannon's entropy (Shannon and Weaver, 1998) to quantify the information retention of our model circuits be-

DRAFT

cause it quantifies how many distinct neural responses are possible given a particular stimulus distribution, and this relates to the specificity of encoding even though it does not indicate which specific stimulus features are encoded. Since there was no noise anywhere in the circuit, the mutual information between the stimulus and the response reduces to the entropy of the response. Mutual information is defined as $MI = H[r] - \langle H[r|s] \rangle$ (Cover and Thomas, 2006). In our study, there is a deterministic relationship between the response and the stimulus due to the lack of noise. The second term of the MI goes to zero and one is left with the entropy of the response. We were careful to avoid any effects that could distort the interpretation of the entropy. The ReLU was ideal here because it compresses the input signal without necessarily scaling it. In our model, the compression was entirely derived from the non-invertibility of the nonlinearity rather than a linear gain factor.

The convergent structure of the retina reduces the dimension of the high-resolution visual input it receives, placing an upper bound on the amount of information that can possibly be transmitted through the optic nerve. In general, the data compression implemented by the circuit architecture may perform lossless or lossy compression or some combination, depending on the statistics of the inputs. In this study, we focus on lossy compression. By using input images of uncorrelated gaussian random pixels (i.e. no redundant structure), we place the inputs into a regime where lossless compression is impossible or assumed to have already taken place. Therefore, the circuit configuration that experiences less lossy compression has a higher entropy than that which experiences more lossy compression. We thus consider higher entropy to be an indication of better performance.

Model simulations and visualizations. All simulations, visualizations, and entropy computations were done in Matlab. The dimension of the stimulus always matches the dimension of the subunits within a pathway, and a stimulus consists of N stimulus inputs. For example, if there are 5 subunits in each of the ON and OFF pathways, then the stimulus has 5 stimulus inputs (sometimes referred to as pixels). Each stimulus input was independently drawn from a gaussian distribution with arbitrary units and a standard deviation of 10 ($\mu = 0, \sigma^2 = 100$). Each subunit receives input from one stimulus input. For all figures in this paper, linear subunits did not transform stimulus inputs and therefore the ON linear subunit response was equivalent to the stimulus input and the OFF linear subunit response was the negative of the stimulus input.

All weights were uniform with unit weights from stimulus inputs to subunits and normalized weights from subunits to outputs. The subunits were normalized so that the variance of the linear sum of subunits is maintained. With N subunits, each subunit weight is $1/\sqrt{N}$. This normalization facilitated a comparison between circuit configurations with linear and nonlinear subunits. All circuit configurations are subject to the same uniform weighting and subunit normalization here and throughout the paper.

Each nonlinear unit applied a ReLU thresholded at zero with unit slope to the stimulus input - effectively a positive-pass filter for ON subunits, $R_{ON\ subunit\ i}$, and a negative-pass filter for OFF subunits, $R_{OFF\ subunit\ i}$. The output neuron linearly sums the subunit responses in its pathway and then applies the nonlinearity. The output response, $R_{ON\ output}$ or $R_{OFF\ output}$, to a given stimulus image, Υ, was a single value that represents a steady state response, as this model did not have temporal dynamics.

$$R_{ON\ subunit\ i}(s_i) = \begin{cases} s_i, & \text{if } s_i > 0 \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

$$R_{OFF\ subunit\ i}(s_i) = \begin{cases} -s_i, & \text{if } s_i < 0 \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

$$R_{ON\ output}(\vec{R}_{ON\ subunit}) = \tag{3}$$

$$\begin{cases} \sum_i^N \frac{1}{\sqrt{N}} R_{ON\ subunit\ i}, & \text{if } \sum_i^N \frac{1}{\sqrt{N}} R_{ON\ subunit\ i} > 0 \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

$$R_{OFF\ output}(\vec{R}_{OFF\ subunit}) = \tag{5}$$

$$\begin{cases} \sum_i^N \frac{1}{\sqrt{N}} R_{OFF\ subunit\ i}, & \text{if } \sum_i^N \frac{1}{\sqrt{N}} R_{OFF\ subunit\ i} > 0 \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

Visualizations in stimulus, subunit, and response spaces. Each quadrant was color-coded such that: $s_1 > 0, s_2 > 0 : blue, s_1 > 0, s_2 < 0 : purple$, etc. Output response histograms in Figure 3 are also color-coded in this way to show which response bins represent which stimuli. For mean luminance and contrast visualization, spaces were color-coded to indicate bands of mean stimulus luminances, M, and stimulus contrasts, Λ. Each stimulus image, Υ, consists of N stimulus inputs, $\Upsilon = [s_1, s_2, ..., s_N]$. In Figures 3A-E and 4, $N = 2$.

$$M(\Upsilon) = \frac{s_1 + s_2 + ... + s_N}{N} \tag{7}$$

$$\Lambda(\Upsilon) = \left| \sqrt{\frac{1}{2} \sum_i (s_i - \langle s \rangle)^2} \right| \tag{8}$$

Entropy calculations. Information entropy is defined as $H = -\sum P[r] log_2 P[r]$. Discrete entropy was used to quantify continuous stimulus and response distributions. Because stimuli that fall into the same discrete bin are ambiguous, the discretization has a similar effect as noise. A consistent bin width of 0.01 was used for all entropy calculations to facilitate comparison. This bin width was used for all dimensions. For example, in a 2D response space, bins would be boxes that are 0.01 x 0.01. Discretization parameters were chosen carefully to ensure that the bins were sufficiently small to

DRAFT

capture the shape of the distributions, but not so small that the log(N) bound was reached. A Freedman-Diaconis histogram bin approximation was used to determine an appropriate bin width for the stimulus and response distributions (Freedman and Diaconis, 1981).

Entropy computations were done by simulating the circuit responses to batches of 10^5 stimulus samples. Responses were binned and response probabilities where computed from the binned response distributions. These were used to calculate the entropy of the responses with the entropy equation. The entropy quantities presented are the average over five batches.

ACKNOWLEDGEMENTS

We thank Joel Zilberberg, Stephen Baccus, Max Turner, Adree Songco-Aguas, Iris Jianghong Shi, and Matthew Farrell for their helpful feedback and comments on the manuscript and Leenoy Meshulam for their helpful discussions.

Funding sources: GJG is supported by NIH NINDS K22 1K22NS104187-01A1 and WRF UWIN postdoc fellowship, ETSB is supported by the Boeing professorship in Applied Mathematics at UW.

References

1. Attneave, F. Some informational aspects of visual perception. Psychological Review 61, 183-193 (1954).
2. Barlow, H. Redundancy reduction revisited. Network: Computation in Neural Systems 12, 241-253 (2001).
3. Barlow, H. B. Possible principles underlying the transformation of sensory messages. in Sensory communication (ed. Rosenblith, W. A.) 217-234 (MIT Press, 1961).
4. Barlow, H. B. Summation and inhibition in the frog's retina. The Journal of Physiology 119, 69-88 (1953).
5. Bell, A. J. & Sejnowski, T. J. An Information Maximization Approach to Blind Separation and Blind Deconvolution. Neural computation 7, 1129-1159 (1995).
6. Brinkman, B. A. W., Weber, A. I., Rieke, F. & Shea-Brown, E. How Do Efficient Coding Strategies Depend on Origins of Noise in Neural Circuits? PLoS comp bio 12, e1005150 (2016).
7. Cover, T. M. & Thomas, J. A. Elements of information theory. (Wiley-Interscience, 2006).
8. Demb, J. B. & Singer, J. H. Functional Circuitry of the Retina. Annual review of vision science 1, 263-289 (2015).
9. DiCarlo, J. J., Zoccolan, D. & Rust, N. C. How Does the Brain Solve Visual Object Recognition? Neuron 73, 415-434 (2012).
10. Eifuku, S., De Souza, W. C., Tamura, R., Nishijo, H. & Ono, T. Neuronal Correlates of Face Identification in the Monkey Anterior Temporal Cortical Areas. Journal of Neurophysiology 91, 358-371 (2004).
11. Enroth-Cugell, C. & Freeman, A. W. The receptive-field spatial structure of cat retinal Y cells. The Journal of Physiology 384, 49-79 (1987).
12. Enroth-Cugell, C. & Robson, J. G. The contrast sensitivity of retinal ganglion cells of the cat. The Journal of physiology 187, 517-552 (1966).
13. Freed, M. A. & Sterling, P. The ON-alpha ganglion cell of the cat retina and its presynaptic cell types. Journal of Neuroscience 8, 2303-2320 (1988).
14. Freedman, D. & Diaconis, P. On the histogram as a density estimator:L 2 theory. Z. Wahrscheinlichkeitstheorie verw Gebiete 57, 453-476 (1981).
15. Gjorgjieva, J., Sompolinsky, H. & Meister, M. Benefits of pathway splitting in sensory coding. J Neurosci 34, 12127-12144 (2014).
16. Glorot, X., Bordes, A. & Bengio, Y. Deep sparse rectifier neural networks. in Journal of Machine Learning Research 315-323 (Universite de Technologie de Compiegne, Compiegne, France, 2011).
17. Gollisch, T. Features and functions of nonlinear spatial integration by retinal ganglion cells. Journal of physiology, Paris 107, 338-348 (2013).
18. Gollisch, T. & Meister, M. Eye Smarter than Scientists Believed: Neural Computations in Circuits of the Retina. Neuron 65, 150-164 (2010).
19. Hartline, H. K. The effects of spatial summation in the retina on the excitation of the fibers of the optic nerve. American Journal of Physiology-Legacy Content 130, 700-711 (1940).
20. Hasselmo, M. E., Rolls, E. T. & Baylis, G. C. The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. Behavioural Brain Research 32, 203-218 (1989).
21. Hochstein, S. & Shapley, R. M. Linear and Nonlinear Spatial Subunits in Y Cat Retinal Ganglion-Cells. The Journal of physiology 262, 265-284 (1976).
22. Ison, M. J. et al. Selectivity of pyramidal cells and interneurons in the human medial temporal lobe. J Neurophys 106, 1713-1721 (2011).
23. Jeanne, J. M. & Wilson, R. I. Convergence, Divergence, and Reconvergence in a Feedforward Network Improves Neural Speed and Accuracy. Neuron 88, 1014-1026 (2015).
24. Kastner, D. B., Baccus, S. A. & Sharpee, T. O. Critical and maximally informative encoding between neural populations in the retina. Proceedings of the National Academy of Sciences of the United States of America 112, 2533-2538 (2015).
25. Laughlin, S. A Simple Coding Procedure Enhances a Neurons Information Capacity. Zeitschrift Fur Naturforschung C-a Journal of Biosciences 36, 910-912 (1981).
26. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. Nature 521, 436-444 (2015).
27. Meister, M. Multineuronal codes in retinal signaling. PNAS 93, 609-614 (1996).
28. Nair, V. & Hinton, G. E. Rectified linear units improve Restricted Boltzmann machines. in ICML 2010 - Proceedings, 27th International Conference on Machine Learning 807-814 (University of Toronto, Toronto, Canada, 2010).
29. Nirenberg, S., Carcieri, S. M., Jacobs, A. L. & Latham, P. E. Retinal ganglion cells act largely as independent encoders. Nature 411, 698-701 (2001).
30. Pitkow, X. & Meister, M. Decorrelation and efficient coding by retinal ganglion cells. Nature Neuroscience 15, 628-635 (2012).
31. Sanes, J. R. & Masland, R. H. The types of retinal ganglion cells: current status and implications for neuronal classification. Annual review of neuroscience 38, 221-246 (2015).
32. Schwartz, O. & Simoncelli, E. P. Natural signal statistics and sensory gain control. Nat Neurosci 4, 819-825 (2001).
33. Shannon, C. E. & Weaver, W. The mathematical theory of communication. (Univ. of Illinois Press, 1998).
34. Sharpee, T. O. & Berkowitz, J. A. Linking neural responses to behavior with information-preserving population vectors. Current Opinion in Behavioral Sciences 29, 37-44 (2019).
35. Sterling, P. & Laughlin, S. Principles of neural design. (2015).
36. Turner, M. H. & Rieke, F. Synaptic Rectification Controls Nonlinear Spatial Integration of Natural Visual Inputs. Neuron 90, 1257-1271 (2016).
37. Turner, M. H., Schwartz, G. W. & Rieke, F. Receptive field center-surround interactions mediate context-dependent spatial contrast encoding in the retina. eLife 7, e38841 (2018).
38. Zhaoping, L. Theoretical understanding of the early visual processes by data compression and data selection. Network: Computation in Neural Systems 17, 301-334 (2006).

DRAFT