Mind the gap: decoding decreases in tonic firing in populations of spiking neurons

Sylvia C. L. Durian[1], Mark Agrios[2], Gregory W. Schwartz[1,2,3*]

1 Departments of Ophthalmology and Neuroscience, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA.
2 Northwestern Interdepartmental Neuroscience Graduate Program, Northwestern University, Evanston, IL, USA.
3 Department of Neurobiology, Weinberg College of Arts and Sciences, Northwestern University, Evanston, IL, USA.

* Corresponding author: greg.schwartz@northwestern.edu

Abstract

A stimulus can be encoded in a population of spiking neurons through any change in the statistics of their spike patterns. Thus, the baseline spike statistics in the absence of a stimulus can impact the population's encoding capacity. Some neurons maintain a baseline firing pattern and can *decrease* their spike rate in response to a stimulus. Not only do baseline firing rates vary widely among different types of neurons, but so do their higher-order statistics, like the degree to which they tend to group their spikes together into bursts and how those bursts are grouped across the population. We investigated how higher-order statistics of baseline spike patterns impact how much information a neural population can transmit about a stimulus that drives a gap in firing using a novel information-theoretic decoding mechanism which we call an "information train." We discover that there is an optimal level of burstiness for gap detection that is robust to other parameters of the neural population like its size, mean firing rate, and level of correlation. We consider this theoretical result in the context of experimental data from different types of retinal ganglion cells with different baseline spike statistics and determine that the spike statistics of bursty suppressed-by-contrast (bSbC) retinal ganglion cells support nearly optimal detection of both the onset and strength of a contrast step.

Introduction

Most neurons communicate using sequences of action potentials called spike trains. Decoding information from spike trains involves extracting a signal in the face of noise, but which features of the spike train represent signal versus noise is not always clear [1, 2]. Neurons may differ both in the way they respond to a stimulus as well as in their baseline spike patterns in the absence of a stimulus, and spike trains among a population of neurons can vary in their patterns of correlation. Many papers have examined the role of first-order statistics (mean firing rate) and correlations on the efficiency of information transmission by spike trains [3–7]. Higher order statistics beyond the mean and variance of the spike rate, such as temporal coding and burst coding, have also been studied before, especially in the auditory cortex [8, 9], but not in the context of suppression of firing. Rather than focus on the effect of higher-order

statistics as a stimulus *response* feature, we are interested in how suppression of firing codes for a stimulus under different higher-order baseline spike statistic conditions. Here, we study how the burstiness of spike trains affects transmission of information about the timing and duration of a gap in firing rate amongst populations of neurons.

Any change in the statistics of a spike train could theoretically provide information; for sensory neurons we often measure the information with respect to a known sensory stimulus. Many neurons have low or zero spontaneous spike rates [10], so increases in rate in the presence of a stimulus are the most common and well-studied type of information transmission [11–13]. However, the central nervous system (CNS) also contains a wide variety of tonically firing neurons, which are well suited to transmit information by *decreasing* their tonic spike rate. Examples include Purkinje cells [14, 15], lateral geniculate relay neurons [16], spinal cord neurons [17], hippocampal CA1 [18] and CA3 [19] cells, and cortical pyramidal cells [20]. Tonic firing in neurons results from a combination of the influence of synaptic networks and intrinsic electrical properties [21, 22]. Both factors can affect the higher-order statistics of tonic spike trains. Different neurons at the same mean rate can have spike trains which vary along the spectrum from the perfectly periodic clock-like transmission of single spikes to highly irregular bouts of bursting and silence [23, 24].

Retinal ganglion cells (RGCs), the projection neurons of the retina, offer a tractable system for both experimental and theoretical studies of information transmission in spike trains. As the sole retinal outputs, RGCs must carry all the visual information used by an animal to their downstream targets in the brain. RGC classification is more complete, particularly in mice [25], than the classification of virtually any other class of neurons in the CNS. Finally, RGCs of each type form mosaics that evenly sample visual space, creating a clear relationship between the size of a visual stimulus and the number of neurons used to represent it [26].

Inspired by the recent discoveries of how different types of tonically firing RGCs encode contrast [27, 28], here we develop a general theoretical framework to measure the role of burstiness in a neural population's ability to encode the time and duration of a gap in firing. We propose a continuous readout of a spike train into an "information train" which tracks Shannon information [29] over time. Critically, we show that information trains can be summed in a population in a way that captures the raw spiking activity more faithfully and completely and is more robust to population heterogeneity than combining the underlying spike trains by standard methods such as population peri-stimulus-time-histograms. We first develop a model to generate spike trains with varying levels of burstiness and correlation, then establish metrics to decode the time and duration of a firing gap from the spike trains, and finally measure decoding performance when we vary burstiness, population correlation, firing rate, and population size.

Our principal result is a theoretical justification for why burstiness exists in some cell types; there is an optimal level of burstiness for identifying the time of stimulus presentation and stimulus strength. This result could help explain the variability in burstiness between different types of neurons. We compare our theoretical results to spike trains from recorded mouse RGCs to measure how close each type lies to optimality for each decoding task.

Results

A parameterized simulation of populations of tonic and bursty neurons.

To study the role of higher-than-first-order statistics in the encoding of spike gaps by populations of neurons, first we needed a parameterized method to generate spike trains with systematic variation in these statistics which could accurately model experimental data. Two RGC types that our lab has studied extensively represent cases near the edges of the burstiness range. Both types have fairly similar mean tonic firing rates (between 40 and 80 Hz). However, while bursty suppressed-by-contrast (bSbC) RGCs fire in bursts of rapid activity interspersed with periods of silence, OFF sustained alpha (OFFsA) RGCs spike in more regular patterns (**Fig 1A**) [28].

A nested renewal process is a doubly stochastic point process that can simulate both bursty and tonic firing patterns with only four parameters (see **Methods**). This process builds upon one of the most commonly used methods to generate spike trains: the Poisson process. The Poisson model of spike generation assumes that the generation of spikes depends only on the instantaneous firing rate [30]. Using a Poisson process to generate spikes leads to exponentially distributed interspike intervals (ISIs; **Fig 1B** left). One obvious limitation of Poisson spike generation, however, is its inability to model refractoriness – the period of time after a neuron spikes during which it cannot spike again. A renewal process extends the Poisson process to account for the refractory period by allowing the probability of spiking to depend on the time since the last spike as well as the instantaneous firing rate [30, 31]. The resulting spike train has gamma-distributed ISIs (**Fig 1B** middle) since the refractory period does not allow for extremely short intervals between spikes.

Even though a renewal process is more physiologically accurate than a Poisson process, it still fails to model burstiness. Therefore we extended the renewal process again by nesting one renewal process inside another [32], similarly to how others have previously constructed doubly stochastic (although non-renewal) processes to model spikes [33]. The outside renewal process parameterized by κ_1 (a shape parameter) and λ_1 (a rate parameter) determines the number and placement of burst events, while the inside renewal process with parameters κ_2, λ_2 determines the number and placement of spikes within each burst (see **Methods**). Therefore, a nested renewal process can flexibly simulate both regular (1 spike/burst, as in a standard renewal process) and bursty (many spikes per burst) firing patterns. The spike trains generated by a nested renewal process have ISIs which are well fit by a sum of gammas distribution (**Fig 1B** right and **Fig S1**), where the tall, narrow left mode of the distribution corresponds to the intervals between spikes within bursts, and the short, wide right mode of the distribution corresponds to the intervals between bursts. Finally, a nested renewal process models activity of bSbCs and OFFsAs well; tuning its parameters results in good approximations to ISI distributions of experimentally measured bSbC (**Fig 1C**) and OFFsA RGCs (**Fig 1D**), since bSbCs and OFFsAs have ISI distributions which are well fit by sum of gammas as well (**Fig S2**).

In order to quantify burstiness by a single measure, we defined a burstiness factor, β, as the average number of spikes per burst normalized by the firing rate (measured in seconds/burst). **Fig 1E** illustrates the effect of different levels of burstiness on the ISI distribution for a fixed firing rate. Both the number of bursts and the number of spikes are randomly generated within our model, so it is important to note that a burst can contain as little as zero spikes by chance; in other words, a "burst" as defined by the outside renewal process may not always correspond to what we would label "bursts" by eye. This goes against the common view of a burst as needing to contain at least a certain number of spikes within a certain period of time. However, bursts determined by

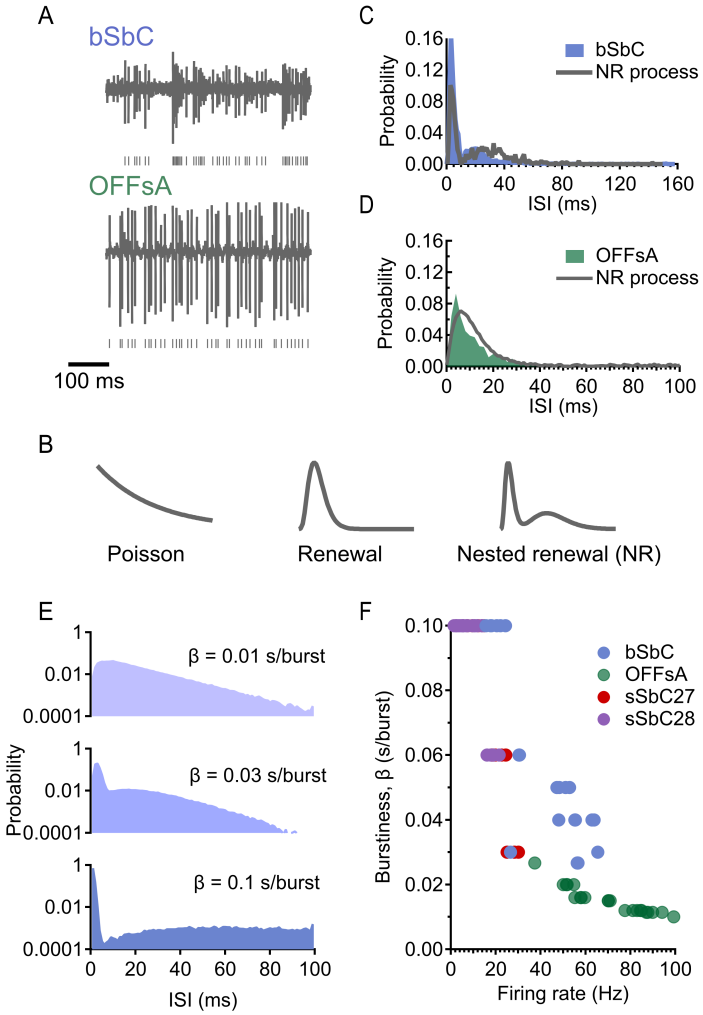


Fig 1. A nested renewal process can simulate bursty and tonic firing patterns with only 4 parameters. (A) Raw traces from bSbC (top) and OFFsA (bottom) in the absence of a stimulus. (B) (Left to right) ISI distributions resulting from a Poisson process (exponential), renewal process (gamma), and nested renewal process (NR; well fit by sum of gammas). (C) ISI distribution of bSbC and closest NR process match with parameters $\kappa_1 = 6$, $\lambda_1 = 180$ bursts/s, $\kappa_2 = 3$, $\lambda_2 = 600$ spikes/s.(D) ISI distribution of OFFsA and closest NR process match with parameters $\kappa_1 = 3$, $\lambda_1 = 150$ bursts/s, $\kappa_2 = 3$, $\lambda_2 = 500$ spikes/s. (E) ISI distributions resulting from NR processes with the same firing rate of 100 Hz, ranging from regular to bursty. (Top: $\kappa = 3$, $\lambda_1 = 300$ bursts/s, $\kappa_2 = 3$, $\lambda_2 = 300$ spikes/s. Middle: $\kappa_1 = 3$, $lambda_1 = 100$ bursts/s, $\kappa_2 = 5$, $\lambda_2 = 1500$ spikes/s. Bottom: $\kappa_1 = 3$, $\lambda_1 = 30$ bursts/s, $\kappa_2 = 3$, $\lambda_2 = 3000$ spikes/s.) Bottom ISI distribution extends out to 500 ms (not shown). Note: ISI distributions are shown on a semilog scale to emphasize differences in their tails. (F) Range of firing rates and burstiness in four RGC types: bSbC and OFFsA as defined in the text, sSbC27 (sustained suppressed-by-contrast type EW27) and sSbC28 (sustained suppressed-by-contrast type EW28). Details of RGC classification can be found in [25]

the outside renewal process will contain, on average, the number of spikes dictated by the inside renewal process, and we prefer our quantification of a burstiness factor since

it eliminates the need to threshold, thus reducing both the number of parameters and arbitrariness. Our quantification of burstiness is dependent upon the parameters of the nested renewal process and cannot be applied to experimental data; therefore, we matched ISI distributions from experimental data with simulated data (**Fig 1C,D**) via a Kolmogorov-Smirov test in order to measure burstiness in spike trains from recorded RGCs (**Fig 1F**). In addition to variable burstiness, the nested renewal process allowed us to independently modulate the firing rate and to simulate neural populations of any desired size with different levels of correlation (**Fig 2A**; see **Methods**).

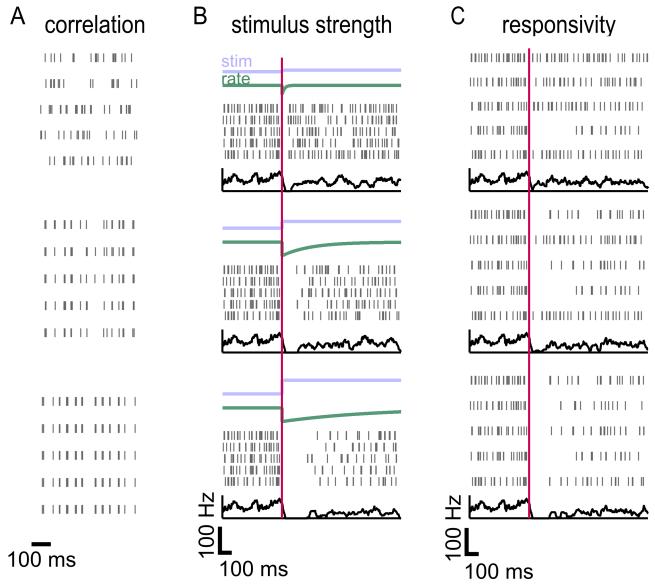


Fig 2. A nested renewal process can simulate populations of varying firing rate, burstiness, size, correlation and responsivity. (A) Populations of varying correlation simulated using a nested renewal (NR) process. (B) Population response to varying levels of stimulus strength, modeled with a drop and subsequent rise in firing rate. Population peri-stimulus-time-histogram (PSTH) is shown in black. (C) Populations of varying responsivity to the stimulus. Population PSTH is shown in black.

We modeled a stimulus-dependent drop in firing as an instantaneous drop to a rate of zero followed by an exponential rise back to the baseline rate, controlled by a variable time constant of recovery (**Fig 2B**). We also introduced heterogeneity into the population of neurons by making a subset of them unresponsive to the stimulus, continuing to spike with baseline statistics after stimulus onset (**Fig 2C**). This stimulus model was chosen because it is consistent with how both bSbC and OFFsA RGCs respond to positive contrast; longer time constants of firing recovery correspond to higher contrast stimuli [28, 34]. However, the stimulus model generalizes to other types of neurons as well. Analogous to models for detection of the onset and duration of an increase in firing, this stimulus allowed us to measure performance in decoding the onset and duration of the firing gap for each simulated population of neurons.

The information train enables aggregation of heterogeneous spike trains.

By aggregating the spike trains of the population of neurons into a one-dimensional signal over time, decoding the onset and duration of a gap in firing can be formulated as a threshold detection task; the choice of threshold determines when the decoder reports that the population is in the baseline state versus the stimulus (gap detected) state (**Fig 3A**). The choice of threshold naturally implies a tradeoff between reaction time, the delay from stimulus onset until the threshold is crossed, and false detection rate, the rate at which the threshold is crossed in the absence of the stimulus. To simplify our analysis, we chose a decoder threshold for each population that achieved a fixed, low error rate of 0.1 false detections per second. Thus, for the detection task, reaction time, δ, was the sole performance metric for the decoder. The performance metric for the duration task was more complex and is considered in the subsequent section.

Next, we considered the choice of the aggregation method used to combine spike trains across cells. The simplest method, which is commonly used in such situations, is to compute the population peri-stimulus-time-histogram (pPSTH) by simply collecting all spikes in the population and smoothing to create a continuous rate signal (**Fig 3F** right). However, there are many properties of population activity that can carry information besides the average firing rate, which is the most commonly used first-order statistic [35, 36]. The pPSTH loses some of the information about the higher-order statistics of each individual spike train that could, in principle, be useful to the decoder. Therefore, we developed a new aggregation method based on the information content of the spike trains. We called the resulting signal the "information train."

The information train method was inspired by neural self-information analysis [37]. In order to build a continuous signal of the information content in a single cell over time, we started by computing a self-information (SI) curve from the ISI distribution using Shannon's definition [29]: $SI = -log_2(p)$, where p is the ISI probability. The SI curve gives the information contained in every observed ISI (**Fig 3B**). SI can be thought of as a measure of surprise: very probable ISIs correspond to small SI, while very improbable ISIs correspond to large SI. A spike train can be equally well described by its ISIs as its spike times; for each ISI observed in the spike train, we may use the SI curve to find the amount of information it contains. Neural self-information analysis then replaces each ISI in a spike train with the corresponding SI value [34], but this creates a discrete signal. Instead, our threshold detection paradigm requires a continuous signal that can be aggregated across cells.

Thus, we developed a new analysis to translate a spike train into a continuous self information train (**Fig 3C** top). We separate ISIs into three cases. In the first case, the current ISI is of the same length as the most probable, or most commonly observed ISI for the cell. Therefore, this ISI contains the lowest possible SI, which we call the "baseline information." The SI train always begins at baseline information, and in this case, remains there for the duration of the ISI. In the second case, the current ISI is longer than the most probable one. However, we do not know that it is longer until the point in time when it has passed the length of the most probable ISI and the cell has not yet spiked. The SI train reflects this by beginning at baseline information and staying constant for the duration of the most probable ISI, then rising according to the SI curve and stopping at the SI value indicated by the total current ISI (blue sections in **Fig 3C** bottom). Finally, in the third case, the current ISI is shorter than the most probable one. In this case, we know that it is shorter as soon as it ends, so the SI train stays at baseline information until the very end of the ISI, then rises instantaneously to the value indicated by the SI curve (green sections in **Fig 3C** bottom). At the start of each ISI, the SI train resets to baseline information, meaning that it has no memory of its history and considers subsequent ISIs independently.

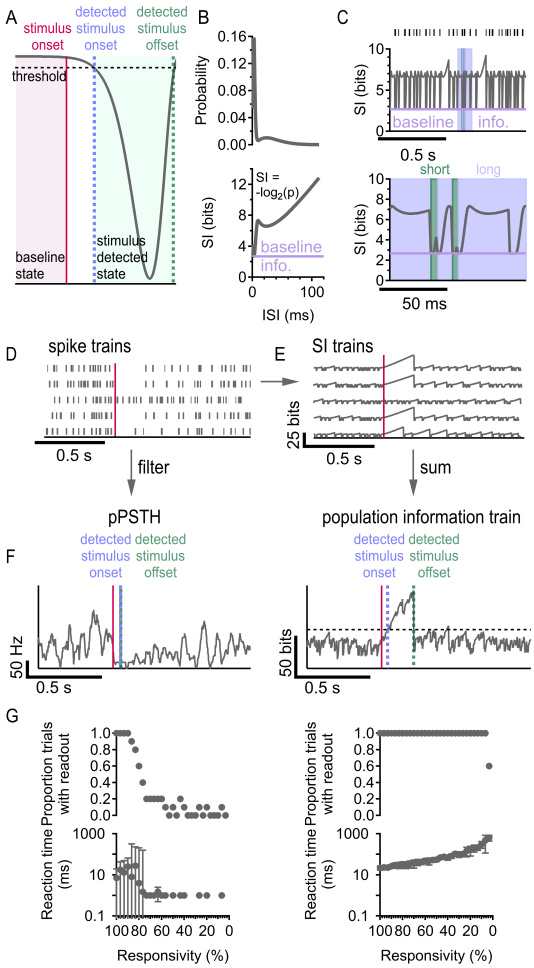


Fig 3. Population firing rate and population information are two plausible readout mechanisms. (A) Detecting onset and duration of a stimulus (pink) from the signal is a threshold detection task: stimulus onset (blue) is detected when the signal crosses the threshold the first time; stimulus offset (green) is detected when the signal crosses the threshold the second time. (B) ISI distribution (top) and corresponding self information (SI) curve (bottom). (C) Top: raster and corresponding single cell information train. Bottom: inset showing long ISIs (blue) and short ISIs (green). (D) Population response to a stimulus (pink) with 80% responsivity. (E) SI trains constructed from spiking activity in (D). (F) Population PSTH (left) and population information train (right) readout from activity in (D). pPSTH detected stimulus onset 50 ms after stimulus presentation; stimulus offset was detected 2 ms after stimulus onset. Information train detected stimulus onset 40 ms after stimulus presentation; stimulus offset was detected 175 ms after stimulus onset. (G) pPSTH (left) and population information train (right) readout on stimulus onset detection task as responsivity decreases. Top: proportion of trials with readout vs. responsivity, where readout does not exist whenever the signal fails to cross the threshold after stimulus onset. Bottom: reaction time vs. responsivity, where reaction time is the time between stimulus onset and stimulus onset detection. Error bars are interquartile range across 10 trials.

For a single cell, the ISI distribution contains all the information in the spike train, so the self information train is a lossless representation. However, for multiple cells,

there are joint ISI distributions to consider. Assuming independence between cells, the information contained in the population is the sum of the information trains of each cell (see **Methods**). Independence is usually not a reasonable assumption, but experimental data from different RGC types show extremely low pairwise noise correlations (**Fig S3**). Furthermore, several studies [38–40], including in the retina [41–43], have shown that only a small amount of information is lost when neural responses are decoded using algorithms that ignore noise correlations – on the order of 10% of the total information. Therefore we have constructed population information trains by summing the SI trains of each cell in the population (**Fig 3F** right) while recognizing that it is a lower bound on the total information present in the full population of spike trains. Correlations can also be induced by a stimulus, and stimulus correlations in populations of neurons are typically larger than noise correlations [42, 44–46]. Our goal, however, was to study the optimal baseline spike statistics for a task in which all responsive neurons were suppressed by the stimulus simultaneously. Thus, non-independence in our population before stimulus onset and after stimulus offset (in other words, when the population was active) was only due to noise correlations.

A critical benefit of the population information train over the pPSTH is that it should automatically amplify the contribution of responsive cells in a population (which will have large SI) relative to unresponsive cells (which will have small SI) without the need to define a cell selection criterion or weighting strategy. To test this intuition, we simulated populations of 30 neurons and varied their responsivity to the stimulus. We decoded the gap onset time for each of these populations using both the pPSTH and the population information train. The pPSTH decoder was extremely sensitive to responsivity, usually failing to reach the error threshold when fewer than 80% of cells were responsive (**Fig 3G left**). The population information train decoder approach was robust to very large fractions of unresponsive cells (**Fig 3G right**). Subsequent analyses used the population information train decoder and 100% responsivity, but our conclusions are robust to lower responsivity (**Fig S4**).

There exists optimal burstiness for minimizing reaction time.

Having developed a readout mechanism – the population information train – we then used it to decode the onset of a gap in firing by measuring reaction time. We were interested in the effects of multiple parameters on reaction time: burstiness, firing rate, population size, and correlation. For three of these parameters, we had an intuition for how they should affect performance. When trying to decode the onset of a gap, temporal precision is key to getting accurate estimates. Increasing the firing rate of a population increases temporal precision, so we expected that increasing the firing rate would improve performance (i.e. decrease reaction time). Another way to gain temporal precision is to increase the size of a population – therefore we also expected population size to have a positive effect on performance. A population which has very low cell correlations is unlikely to have many overlapping spikes from different cells, while a population with high correlation is likely to have major overlap. This implies that temporal precision, and thus performance, should decrease with correlation. In contrast to the other parameters, the intuition for how burstiness affects performance is not simple and was our central question in this part of the study.

We isolated the effect of burstiness on reaction time by holding firing rate, population size, and correlation constant and found a non-monotonic relationship (**Fig 4A**), suggesting that for each combination of parameters there may be a different level of burstiness that minimizes reaction time. We could continue in this way to isolate the effects of each parameter by holding the other parameters constant at different values, but the number of parameters, and their ranges, makes this impractical. A more elegant approach is to try to find a unifying principle which can collapse the

data. Dimensional analysis gives us the tools to identify such a unifying principle.

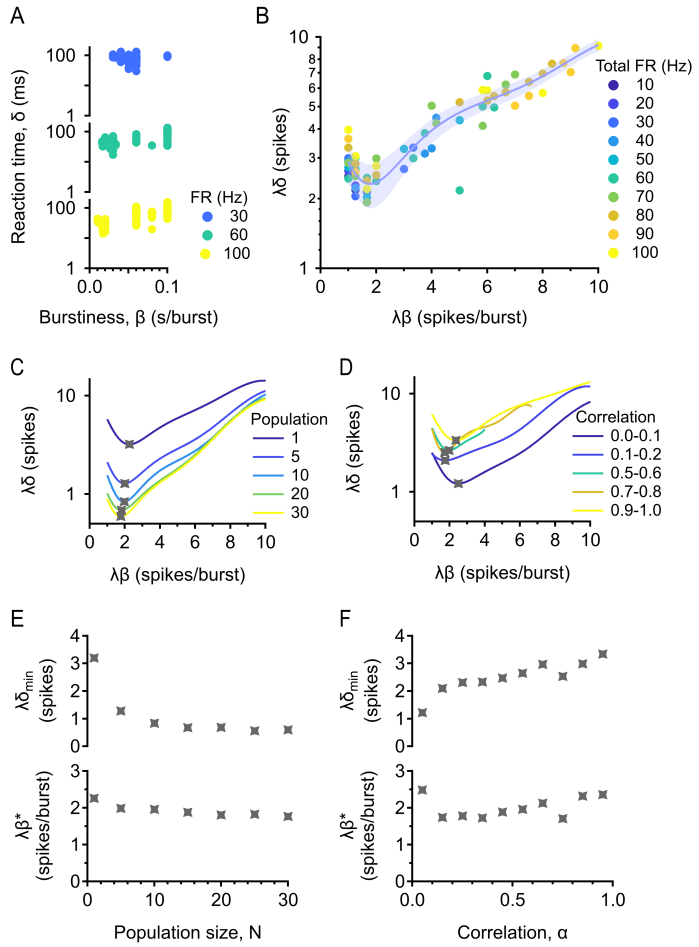


Fig 4. Optimal burstiness for minimizing reaction time. (A) Reaction time δ vs. burstiness β with information train readout. Population is fixed at 5 cells, correlation is fixed between 0-0.2, and total firing rate is fixed at 30, 60, and 100 Hz. (B) Dimensional analysis collapses the data for different firing rates. Dimensionless reaction time is plotted against dimensionless burstiness. Population is fixed at 5 cells and correlation is fixed between 0-0.2. Trial-averaged (median) data is shown. Error bar = standard error estimate of the data around the fit. (C) Existence of a minimum is robust across population sizes. Correlation is fixed between 0-0.2. (D) Existence of a minimum is robust across correlations. Population is fixed at 5 cells. (E) Top: minimum (dimensionless) reaction time against population size. Bottom: optimal (dimensionless) burstiness against population size. (F) Same as in (E) but correlation is varied.

Dimensional analysis uncovers the relationships between variables by tracking their physical dimensions over calculations. Since responsivity is held constant at 100%, there are only five relevant quantities altogether: reaction time δ, burstiness β, firing rate λ, population size N, and correlation α. These are all either dimensionless or some transformation of the physical dimension "time," so there is only one relevant dimension in this problem. By the Buckingham Pi Theorem [47], it is possible to construct $5 - 1 = 4$ independent dimensionless groups which are related by some unknown function. We chose to make reaction time and burstiness dimensionless by multiplying

by firing rate (brackets denote the dimension of the quantity inside), so we obtain

$$[\lambda\delta] = [\lambda\beta] = [N] = [\alpha] = 1 \tag{1}$$

We may write any one of these dimensionless quantities as a function of the rest, but it is difficult to fit functions of three variables, so we fix population size and correlation so that the function no longer depends on them, and then we have

$$\lambda\delta = f(\lambda\beta) \tag{2}$$

This immediately reveals that both reaction time and burstiness are inversely proportional to the firing rate. While burstiness was defined in such a way that it must be inversely proportional to firing rate (see **Methods**), it is illuminating that reaction time is inversely proportional as well. This basic result of dimensional analysis gives us much more information than our intuition, which simply told us that reaction time should decrease with firing rate.

The functional form of f is now possible to find by fitting (**Fig 4B**). Dimensional analysis collapses all the data for different firing rates together, with no pattern, so we can analyze them together. There is a clear minimum, demonstrating that there is some level of burstiness which is optimal for minimizing reaction time across all firing rates. Now we may separately vary population size and correlation (**Fig 4C,D**), illustrating that the existence of optimal burstiness for minimizing reaction time is robust for both parameters. The x and y values of the minima of these curves completely describe how optimal burstiness and minimum reaction time depend on population size and correlation – simply dividing the x and y values of the minimum by the firing rate of the population, we obtain the exact (dimensionful) optimal burstiness and minimum reaction time. Minimum reaction time decreases with population size and increases with correlation (as predicted by our intuition), while optimal burstiness is relatively constant with both parameters, indicating that there is one level of burstiness optimal for detecting stimulus onset at any population size and correlation.

There exists optimal burstiness for discriminating stimulus strength.

Besides "when," another fundamental question to ask about a stimulus is: how strong is it? In our case, the strength of the stimulus corresponds to the duration of the gap in spiking. We varied stimulus strength by changing the recovery time of the burst firing rate (**Fig 2B**) and measured the gap length in the information train (**Fig 3E** bottom right), or the duration of time between stimulus onset and offset detection, in order to make deductions about how the length of the gap in spiking is affected by the suppressed firing rate (**Fig 5A**). The performance metric here – the measure of how well a population can discriminate the stimulus strength – is essentially the accuracy with which the time constant of recovery is captured by the gap length measurement. Therefore the performance metric should actually be an error metric: a measure of how much error there is in the relationship between gap length and the time constant. However, it is not immediately obvious what the relationship between these two variables actually is, much less how much scatter there is around it. Dimensional analysis is again a useful tool here. There are six relevant quantities to this problem: gap length γ, recovery time constant τ, burstiness β, firing rate λ, population size N, and correlation α. Applying the Buckingham Pi Theorem [47] and setting all but two of the dimensionless groups (see **Methods**) to be constant (so that we obtain a function of one variable which relates the gap length and time constant of recovery), we have

$$\lambda\gamma = f(\lambda\tau) \tag{3}$$

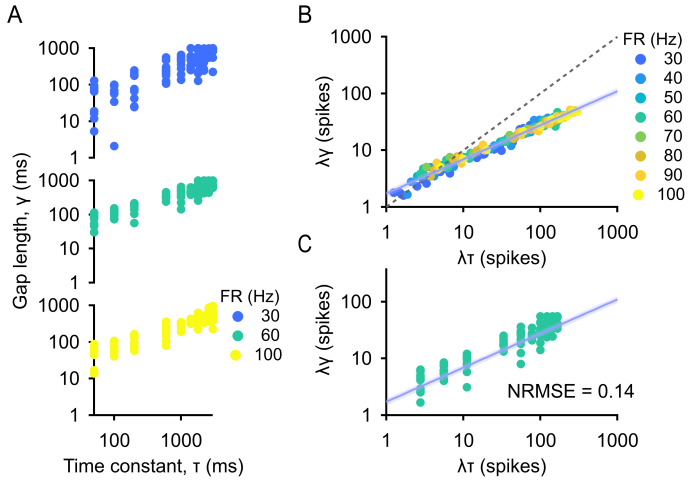


Fig 5. Gap length is exponentially related to the recovery time constant. (A) Gap length γ vs. time constant of recovery τ. Population is fixed at 5 cells, correlation is fixed between 0-0.2, dimensionless burstiness ($\lambda\beta$) is fixed at 3 spikes/burst, and total firing rate is fixed at 30, 60, and 100 Hz. (B) Dimensionless gap length vs dimensionless time constant. Population is 5 cells, correlation is 0-0.2, and dimensionless burstiness is 3 spikes/burst. Trial-averaged (median) data is shown. Dashed line is $y = x$. Error bar = standard error estimate of the data around the fit. (C) Performance metric is the scatter in the data measured with normalized root mean square error (NRMSE). Population is 5 cells, correlation is 0-0.2, and dimensionless burstiness is 3 spikes/burst, and firing rate is 60 Hz. Error bar = standard error estimate of the data around the fit.

By fitting, it is clear that there is an exponential relationship between these variables (**Fig 5B**). Now, for each combination of parameters, for which we have several trials of gap length measurements, we chose the performance metric to be the scatter in the data around the exponential function suggested by dimensional analysis (**Fig 5C**) which we quantified with normalized root mean square error (NRMSE; see **Methods**).

Now that we have a performance metric, we wanted to see how it depended on burstiness in particular, as well as firing rate, population size, and correlation. Once again, we could isolate the effects of each of these parameters by holding all the others constant (**Fig 6A**), but using dimensional analysis simplifies the problem by collapsing data. The relevant quantities are the performance metric NRMSE ϵ, burstiness β, firing rate λ, population size N, and correlation α, so applying the Buckingham Pi Theorem [47] (see **Methods**) and setting population size and correlation to be constant, we have

$$\epsilon = f(\lambda\beta) \tag{4}$$

While burstiness is still inversely proportional to the firing rate, this reveals that NRMSE is constant with firing rate. In other words, the ability to decode the duration of a gap in firing rate does not depend on the spontaneous firing rate. Fitting reveals that f is a function which decays monotonically to a nonzero constant (**Fig 6B**) (see **Methods**). Interestingly, there is a large range of burstiness that optimizes NRMSE, in

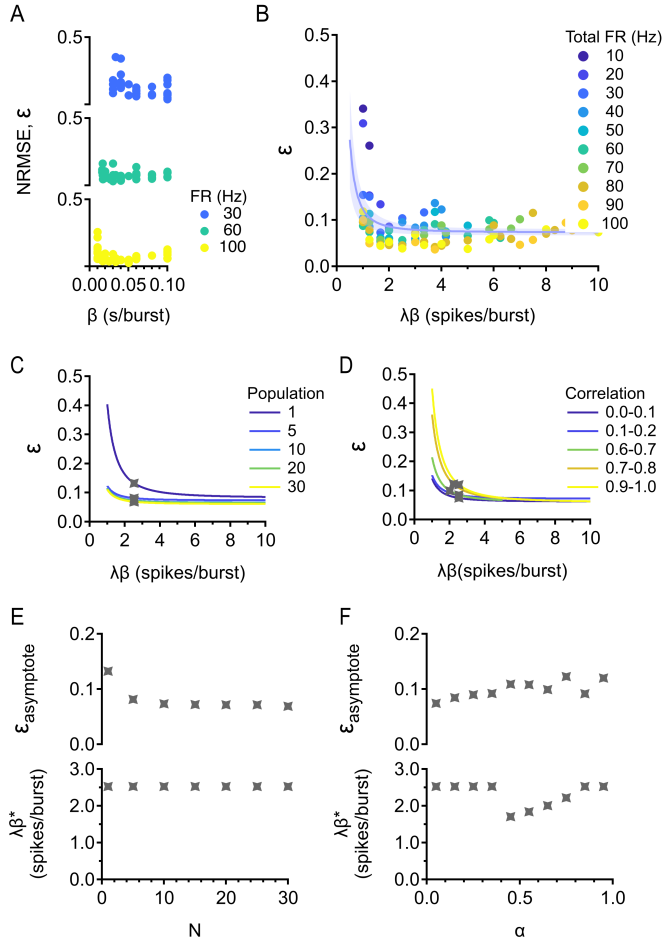


Fig 6. Optimal burstiness for discriminating stimulus strength. (A) NRMSE ϵ vs. burstiness β. Population is fixed at 5 cells, correlation is fixed between 0-0.2, and total firing rate is fixed at 30, 60, and 100 Hz. (B) NRMSE vs. dimensionless burstiness. Population is 5 cells and correlation is 0-0.2. Trial-averaged (median) data is shown. Error bar = standard error estimate of the data around the fit. (C) Existence of an asymptote is robust across population sizes. Correlation is fixed between 0-0.2. (D) Existence of an asymptote is robust across correlations. Population is fixed at 5 cells. (E) Top: asymptotic NRMSE against population size. Bottom: optimal (dimensionless) burstiness against population size. (F) Same as in (E) but correlation is varied.

contrast to how there was a single optimal value of burstiness for minimizing reaction time.

Next we separately varied population size and correlation (**Fig 6C,D**), which showed us that the existence of asymptotic NRMSE and a large range of burstiness resulting in this optimal performance was robust for both parameters. We chose to study the "elbows" of these curves because if we can describe the lowest level of burstiness that optimizes NRMSE, we will know that any burstiness larger than that will also be optimal. Asymptotic NRMSE decreases with population size and increases with correlation, while optimal burstiness is relatively constant with both parameters, implying that the level(s) of burstiness optimal for discriminating stimulus strength is not affected by population size or correlation.

bSbC RGCs have close to optimal burstiness for gap detection.

Having established that optimal burstiness exists for decoding the onset and duration of a gap in firing, our next question was how the baseline spike statistics of the RGC types we studied relate to this optimum. Recall that we fit functions that describe how reaction time (**Fig 4**) and NRMSE (**Fig 6**) depend on dimensionless burstiness. Dimensionless burstiness is simply burstiness multiplied by firing rate, so another way to represent the information in **Figs 4, 6** is as a 3 dimensional graph of reaction time (or NRMSE) vs. burstiness and firing rate (**Fig 7**). We compared performance of different types of experimentally recorded cells by using their firing rate and burstiness to predict how quickly they would detect stimulus onset (**Fig 7A**) and how accurately they would discriminate stimulus strength (**Fig 7B**) according to our model. Thus the experimental data necessarily lie on the surfaces. We set the population size at 5 cells and the correlation between 0-0.2, but as we saw earlier, optimal burstiness is negligibly affected by both the population size and correlation so these results apply to any values of those parameters. Both bSbCs and OFFsAs are quite good at detecting a stimulus quickly, although bSbCs are closer to optimal reaction time, while both types of sustained suppressed-by-contrast RGCs are much worse at this task. However, **Fig 7B** suggests that bSbCs are by far the best at discriminating stimulus strength, since their burstiness puts them right on the lower bound of optimal burstiness; OFFsA RGCs do not perform nearly as well. Therefore, bSbC RGCs have spiking patterns which enable them to both react to a stimulus coming on as quickly as possible and detect the strength of that stimulus with great accuracy.

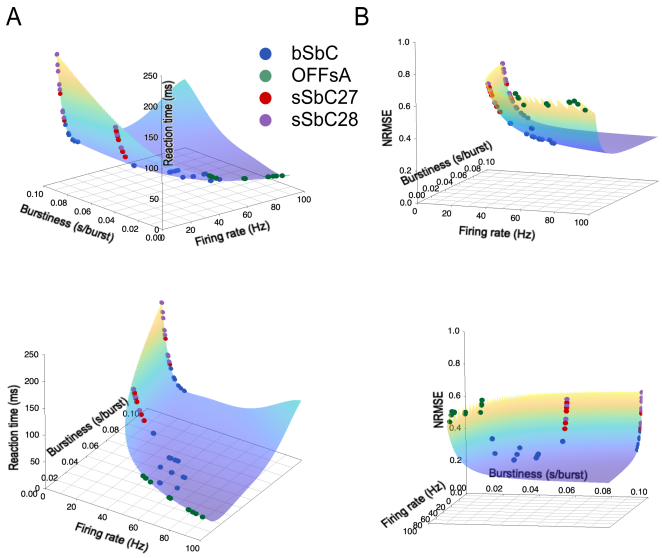


Fig 7. bSbCs have close to optimal burstiness. (A) 3D surface of reaction time predicted from firing rate and burstiness, as described by the function in **Fig 4B**. The surface is colored by elevation for clarity. The position of four RGC types on the surface is shown. Population size is 5 cells, correlation is 0-0.2. Top and bottom are different views. (B) 3D surface of NRMSE predicted from firing rate and burstiness, as described by the function in **Fig 6B**. The position of four RGC types on the surface is shown. Population size is 5 cells, correlation is 0-0.2. Top and bottom are different views.

Discussion

To investigate the role of burstiness in population decoding of firing gaps, we simulated spike trains for populations of neurons using a nested renewal process. This strategy allowed us to capture the statistics of recorded spike trains from RGCs and also to vary burstiness parametrically (**Figs 1, 2**). We then developed a new analysis to combine spike trains across a population into an information train and demonstrated that this method is more robust to unresponsive cells than a population PSTH (**Fig 3**). Using the information trains of different simulated populations, we discovered that there is an optimal burst rate for detecting the onset of a firing gap that is inversely proportional to firing rate and relatively independent of population size and correlation (**Fig 4**). There is also an optimal range of burstiness for detecting gap duration that is relatively independent of all of these other parameters (**Figs 5, 6**). Finally, we considered the baseline spike statistics of four RGC types in the context of these theoretical results and revealed that the burst patterns of bSbC RGCs make them nearly optimal for detecting both the onset and the strength of a contrast step (**Fig 7**).

Baseline spike statistics can be optimized for different decoding tasks.

The fact that some burstiness higher than minimum is optimal for identifying the time of stimulus presentation and stimulus strength leads us to the same overarching principle as [48–50]: different spiking patterns are optimal for encoding different stimulus features. We extended their results by demonstrating two concrete examples of how optimal encoding of different stimulus features depends on spiking patterns. The theoretical implications of this are that different neurons may display different spiking statistics depending on the stimulus features they encode.

Earlier we provided intuition about the effects of firing rate, population size, and correlation on reaction time. The same intuition holds for performance on decoding the duration of a gap in firing, since it is also dependent on temporal precision. Now we will lay out our intuition for why there exists optimal burstiness for both tasks, which will explain why it makes sense that specific spiking patterns are optimal for encoding different stimulus features. We again attribute increased performance on these tasks to increased temporal precision. As explained earlier, a population could theoretically increase its temporal precision by increasing its firing rate or its size, or decreasing its cell correlations; however, there are physiological limits to a population's firing rate and correlation, and the size of the population activated by a stimulus is related to the size of the stimulus [26]. In contrast, rearranging the pattern of spikes is something that a cell can reasonably do, and if it rearranges its spiking pattern so that it has long periods of silence and short periods of rapid activity, or bursts, then it has great temporal precision within those bursts. If that cell is part of a population with low pairwise cell correlations (**Fig S3**), then the bursts from different cells would be staggered in time, allowing the population as a whole to have excellent temporal precision. Therefore it is clear that increasing burstiness can improve performance. Following this line of thinking, it is also easy to see why too much burstiness would be harmful: if the population is trying to detect stimulus onset, the periods of silence between bursts could become so long that even with many cells it is very possible that the population would entirely miss the stimulus simply because no cell was in the middle of a burst when the stimulus was presented. In addition, the fact that firing rate, population size, and correlation had effects on performance that were predictable from this intuition lends further credibility to our result that there is an optimal level of burstiness for identifying the timing and strength of a stimulus.

We found that the baseline firing statistics of bSbC RGCs lie close to the optimum for performance on the gap detection task (**Fig 7**), but what about the other 3

high-baseline-rate RGC types we analyzed and the many lower baseline rate ones we did not analyze? It is important to consider (1) that the gap detection as we defined it is only one out of the many decoding tasks that must be performed simultaneously across the population of RGCs, and (2) that biological constraints, including the energy expenditure of maintaining a high spike rate, also likely drove the evolution of the different RGC types. Like bSbC RGCs, sSbC RGCs also signal exclusively with gaps in firing, but they do so at a substantially longer timescale [27, 51]. Perhaps for these RGCs, the energetic benefit of maintaining a lower baseline firing rate is worth the cost of lower temporal resolution in gap detection because they represent time more coarsely than bSbC RGCs. The OFF alpha RGCs (OFFsA and OFFtA) respond to different stimuli with either increases or decreases in baseline firing [52, 53], and OFFtA RGCs have been implicated in the detection of several specific visual features [54–56], so for these RGC types, performance in representing spike gaps needs to be balanced against metrics for their other functions.

The information train as a way to track population-level changes in spike patterns.

Our work has practical as well as theoretical implications: we proposed the information train readout which tracks the information in a population over time, and which is more comprehensive and robust than a standard pPSTH readout. For example, if you use a pPSTH to analyze a population of direction selective cells without first finding the preferred directions of every single cell in the population, the responses from cells preferring opposite directions will cancel each other out, leading to a pPSTH which fails to reflect change due to the stimulus. The same effect will be observed if you present a light-dark edge to a population of ON or OFF cells. Of course, it is possible to remove this effect by classifying each cell's response to the stimulus first, but that can be time consuming if the population size is large and it requires an (often arbitrary) supervised classification step. The information train will reflect a stimulus change in both of these cases even when the whole population is analyzed together. This is because any ISI length out of the ordinary (i.e. different from the most probable one) causes a positive deflection in the information train, so no matter whether a cell's firing rate increases or decreases in response to a stimulus, the population information rises. Therefore the information train is a convenient readout mechanism to use because it does not require any assumptions to cluster cells, even when the population recording is heterogeneous. It is also easy to implement by simply fitting the observed ISI distributions in the absence of a stimulus to a gamma (many neural circuits have ISI patterns well fit by a gamma distribution [57]) or sum of gammas distribution (depending on whether the cells in question are bursty).

In fact, we believe that the information train nicely bridges a middle ground between the pPSTH and another standard readout mechanism: manifold analysis [58, 59]. The information train is more complete than a pPSTH but less complete than manifold analysis; only slightly more difficult to implement than a pPSTH and far easier to implement than manifold analysis; equally interpretable as a pPSTH and far more interpretable than manifold analysis. We suggest that the information train can be used to determine whether a population-level change has occurred, and then manifold analysis can be used to investigate that change more thoroughly.

Limitations and future directions.

The information train is not a complete description of the full mutual information in the population since it assumes independence. The natural extension of SI to a population is pointwise mutual information (PMI) [29] which measures the association of single

events. Much like entropy is the expected value of SI, mutual information is the expected value of PMI. In the future, a more accurate way to construct the population information train would be to use PMI: the value of the information train at time t is given by measuring the time since the last spike for each cell in the population and then calculating the PMI in the coincidence of those events. This is quite difficult to implement because it requires describing the joint ISI distributions. Not only is it necessary to estimate the joint ISI probabilities during the prestimulus time – when we can generate a lot of data but it is still difficult to estimate the joint probabilities – it is also necessary to be able to accurately extrapolate those joint ISI distributions in order to deduce the stimulus response. Namely, we need to be able to accurately estimate the very tails of the joint ISI distributions. We were able to do this in our study by fitting our observed ISI distributions with a sum of gammas distribution (**Fig S2**), but it could be dangerous to first estimate the joint ISI distributions and then estimate their tail values. Even small errors in the density estimation would lead to drastically different results, since PMI (much like SI) amplifies the significance of extremely improbable events. In other words, small differences in p lead to large differences in $log_2(p)$ for low p. Calculating the true PMI of the population over time is an important future direction, but it has to be done carefully.

Another limitation of this study is that we have not formally seen how sensitive the information train readout is to more complex tasks/stimuli. We posit that the information train should be able to flexibly reflect any change in spiking patterns at the population level since every observed ISI different from the most common one registers a change. However, we only compared the information train to the standard pPSTH for one task; we do not know if the information train gives a more accurate or robust readout than a pPSTH for all stimuli.

Not only is there potential follow up research stemming directly from this study, such as constructing the population information train using PMI and exploring more complicated stimuli, but in addition, the information train will hopefully be used to investigate more complex statistics of spiking and other variables that affect information transmission. Moreover, the information train can be used to study other change detection tasks besides changes in spiking activity. For example, it could be applied to the problem of detecting auditory frequency change, where gaps are similarly important. This theoretical framework is general enough that it can be applied in a variety of research directions.

Methods

Model of spike generation.

In the Poisson model of spike generation, the instantaneous firing rate is the only force which generates spikes. Assuming a constant firing rate λ over time, the number of spikes within an interval of time Δt is a Poisson distributed random variable with parameter λ. Using the Poisson probability density function,

$$P(1\, spike\, during\, \Delta t) = \lambda \Delta t \tag{5}$$

One can therefore generate a sequence of uniformly distributed random numbers $\{x_n\}$ in order to determine when to generate a spike: if $x_i < \lambda \Delta t$, generate a spike in time bin i. Since the number of spikes within any interval of time is a Poisson random variable with parameter λ, the exponential distribution $Exp(\lambda)$ describes the time between spikes, i.e., the interspike interval distribution.

A renewal process extends the Poisson process to depend on both the instantaneous firing rate and the time since the previous spike. One way to model this is to simply generate a Poisson spike train with rate parameter λ and delete all but every κ^{th} spike. A gamma distribution predicts the wait time until the th Poisson process event, so the ISI distribution of a renewal process generated in this way is described by $Gamma(\kappa, \lambda)$. This is a natural extension of the Poisson process since if κ is 1, the ISI distribution is $Gamma(1, \lambda) = Exp(\lambda)$. The average ISI length is $\frac{\kappa}{\lambda}$ and therefore the average firing rate is $\frac{\lambda}{\kappa}$.

We extended the renewal process again, nesting one renewal process inside another. The outside renewal process with parameters κ_1 (a shape), λ_1 (a rate) determines the number and placement of bursts, with an average of $\frac{\lambda_1}{\kappa_1}$ bursts/second. We allowed λ_1 to vary between 30 and 600 bursts/second, and κ_1 to vary between 3 and 6 in order to obtain a range of 10-100 bursts/second. The inside renewal process with parameters κ_2, λ_2 determines the number and placement of spikes within each burst, with an average of $\frac{\lambda_2}{\kappa_2}$ spikes/second within each burst. We let λ_2 range from 300 to 6,000 spikes/second, and κ_2 range from 3 to 6. Thus the average spiking rate within a burst ranged from 100-1000 spikes/second. We inserted spikes generated by the inside process into bursts by setting a burst length τ_b, which we chose to be 10 ms. This results in an average of $\frac{\lambda_2 \tau_b}{\kappa_2}$ spikes/burst, with a range of 1-10 spikes/burst, and therefore an average firing rate

$$\lambda = \frac{\lambda_1 \lambda_2 \tau_b}{\kappa_1 \kappa_2}, \tag{6}$$

with a range of 10-100 Hz.

We varied pairwise correlations between nested renewal process spike trains by using a shared vs. random seed strategy. Just as for a Poisson process, we needed two sequences (for the inside and outside renewal processes) of uniformly distributed random numbers, $\{x_n\}$ and $\{y_n\}$, to determine when to generate spikes (see above). We first generated two shared sequences of uniform random numbers, $\{x_{n_{shared}}\}$ and $\{y_{n_{shared}}\}$, then for each cell in the population generated new independent sequences of uniformly distributed numbers, $\{x_{n_{independent}}\}$ and $\{y_{n_{independent}}\}$. For each cell, $\{x_n\}$ was constructed by drawing from $\{x_{n_{shared}}\}$ with probability α_x and $\{x_{n_{independent}}\}$ with probability $1 - \alpha_x$. The other sequence of uniform numbers, $\{y_n\}$, was constructed similarly with weight α_y. The choices of α_x and α_y control the pairwise correlations of the inside and outside renewal processes. We measured cross correlations between all pairs of spike trains in the population and averaged.

Burstiness.

Definition. *Burstiness* β is the average number of spikes per burst normalized by firing rate: $\beta = \frac{\lambda_2 \tau_b}{\lambda \kappa_2}$.

Remark. Using Eq (6), the formula for burstiness can be simplified to $\beta = \frac{\kappa_1}{\lambda_1}$. The range of burstiness is 0.01-0.1 seconds/burst.

Population readout mechanisms.

Self information is defined as $SI = -log_2(p)$ [29], where p is the ISI probability. We constructed a population information train by summing individual SI trains. This implicitly assumes that cells are independent since if independence holds, then

$$-log_2(p(x, y)) = -log_2(p(x)p(y)) = -log_2(p(x)) - log_2(p(y)). \tag{7}$$

In order to measure reaction time and gap length in the information train, we set a threshold on the information train such that the error rate was 0.1 false crossings per second.

We set the threshold for the pPSTH readout mechanism at 0, and set the filter length such that the pPSTH reached 0 with a rate of 0.1 errors/second during the prestimulus time.

Population information train options B and C considered in **Fig S5** were constructed as follows:

- Option B: The population information train is the SI train constructed from the ensemble spike train, or the spike train obtained by overlaying all the spike trains in the population.
- Option C: The population information train was constructed by summing the SI trains for each cell in the population, but we let ISIs shorter than the most probable one have negative deflections in each SI train, while longer ISIs still had positive deflections.

The analogous threshold (resulting in 0.1 errors/s) was placed on the information trains in options B and C.

Dimensional analysis.

We used dimensional analysis to find the relationship between gap length and the time constant of recovery (**Fig 5A,B**). There are six relevant quantities to this problem: gap length γ, recovery time constant τ, burstiness β, firing rate λ, population size N, and correlation α. These are all either dimensionless or some transformation of "time", so by the Buckingham Pi Theorem [47], we may construct $6 - 1 = 5$ independent dimensionless groups related by an unknown function. We chose to make gap length, the recovery time constant, and burstiness dimensionless by multiplying by firing rate, so we have

$$[\lambda\gamma] = [\lambda\tau] = [\lambda\beta] = [N] = [\alpha] = 1 \tag{8}$$

In order to obtain a function of one variable which relates gap length and the time constant, we set dimensionless burstiness, population size, and correlation to be constant and obtained Eq (3).

Once we defined a performance metric (NRMSE) for the gap length decoding task (**Fig 5C**), we used dimensional analysis again to find its dependence on burstiness and the other model parameters (**Fig 6A,B**). The relevant quantities are the performance metric NRMSE ϵ, burstiness β, firing rate λ, population size N, and correlation α, so by the Buckingham Pi Theorem [47], we may construct $5 - 1 = 4$ independent dimensionless groups related by a function. We again chose to make burstiness dimensionless by multiplying by firing rate, so we have

$$[\epsilon] = [\lambda\beta] = [N] = [\alpha] = 1 \tag{9}$$

Fixing population size and correlation so that the function no longer depends on them, we obtain Eq (4).

Fitting routines.

From dimensional analysis, we obtained Eq (4). Plotting $\lambda\beta$ against ϵ revealed that f is an asymptotically decaying function (**Fig 6B**). We therefore let f be of the form

$$f(x) = a + \frac{b}{x^n} \tag{10}$$

The range of n obtained from this fitting routine was 0.5-4 with no systematic trends, so n was taken as 2, which results in essentially the same good fits and reduces the number of fitting parameters to two. We also considered an exponential fit, but ultimately rejected it because the fit was not as good as the fit given by Eq (10) with fixed n, and it requires three fitting parameters instead of two. Goodness of fit of Eq (10) with variable n as well as fixed n, and the exponential fit, were all assessed with root mean square error (RMSE) returned by the fitting routine.

We fit the exponential relationship between the dimensionless time constant and the dimensionless gap length (**Fig 5B,C**) by fitting a linear relationship of their logarithms. Goodness of fit was assessed with normalized root mean square error (NRMSE):

Definition. The NRMSE of an estimator $\hat{y}$ with respect to observed values of a parameter y is the square root of the mean squared error, normalized by the mean of the measured data, $\bar{y}$: $NRMSE = \frac{1}{\bar{y}}\sqrt{\sum_{i=1}^{n}\frac{1}{n}(y_i - \hat{y}_i)^2}$

We fit ISI distributions of simulated (**Fig S1**) and experimentally recorded data (**Fig S2**) by minimizing the mean square error of the log of the data and log of a sum of gammas distribution, since there was a large dynamic range in the data. This effectively minimizes the percentage difference between the data and the fitting function. Goodness of fit was assessed with RMSE returned by the fitting routine.

RGC spike recordings.

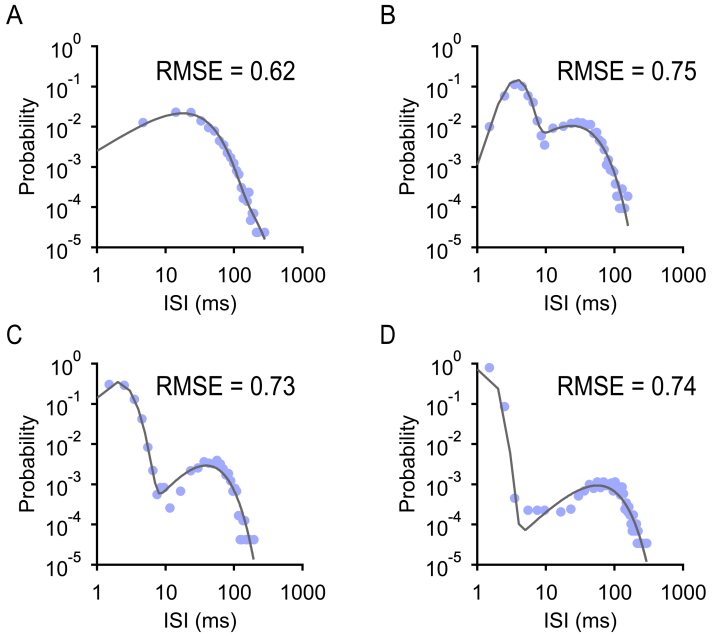
RGCs were recorded in cell-attached configuration in whole-mount preparations of mouse retina as described previously [28, 34]. Cell typology was determined using the methods described in [25]. Baseline spiking was recorded at a mean luminance of 0 and 1000 isomerizations (R*) per rod per second.

Data and code availability.

All simulated data reported in this paper will be available from the GIN database upon manuscript acceptance. Experimental data reported will be shared by the lead contact, Greg Schwartz (greg.schwartz@northwestern.edu), upon request. All code written in support of this publication will be available from the GitHub repository upon manuscript acceptance.

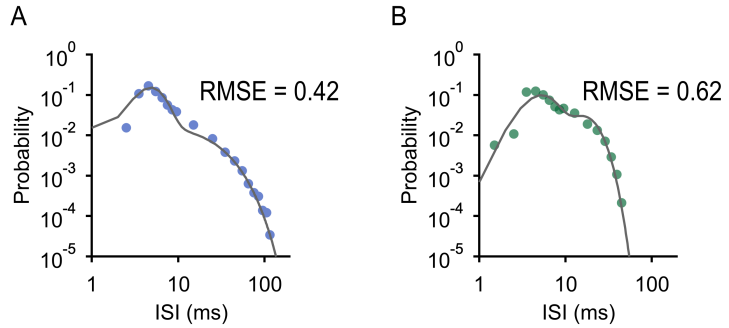
Acknowledgements

We are grateful to all members of the Schwartz Lab for their feedback and support on this project. We would like to thank Stephanie Palmer and Sophia Weinbar for their comments on the manuscript, as well as Sara Solla and Ann Kennedy for their feedback on the project.

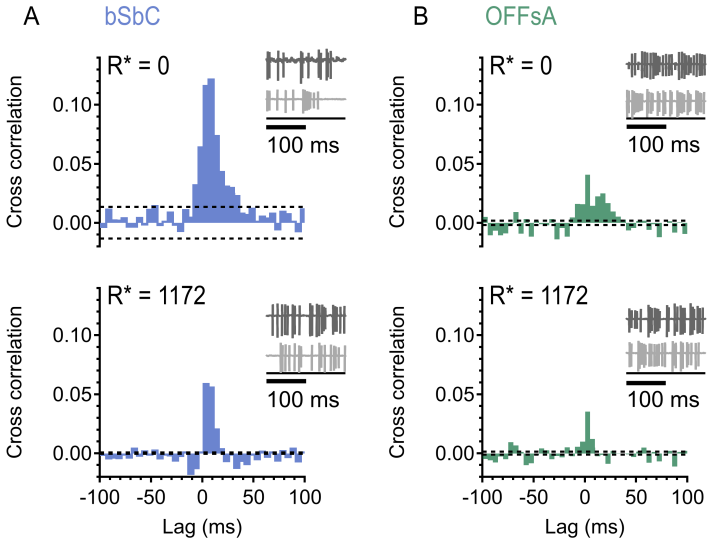


Supporting information

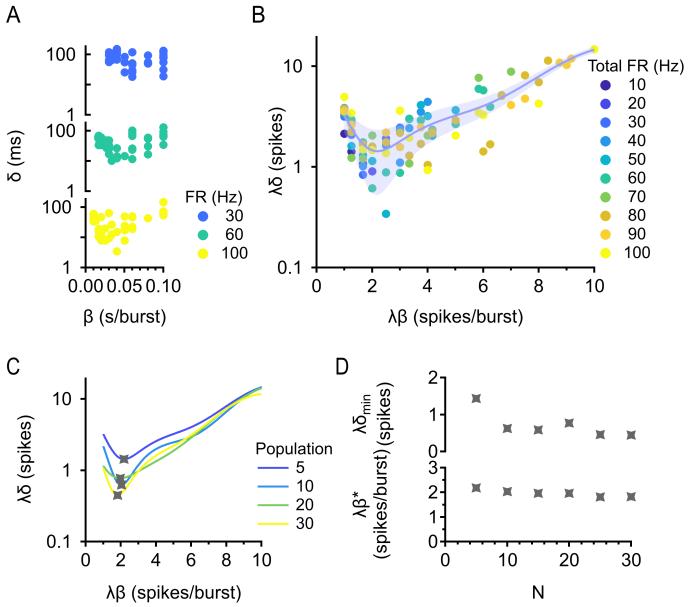
S1 Nested renewal process ISI histograms are well fit by a sum of gammas distribution. (A) ISI distribution from NR process with parameters $\kappa_1 = 4$, $\lambda_1 = 200$ bursts/s, $\kappa_2 = 4$, $\lambda_2 = 400$ spikes/s. Firing rate $\lambda = 50$ Hz, burstiness $\beta = 0.02$ s/burst. Goodness of fit of the sum of gammas distribution is reported as root mean square error (RMSE). (B) ISI distribution from NR process with parameters $\kappa_1 = 4$, $\lambda_1 = 100$ bursts/s, $\kappa_2 = 5$, $\lambda_2 = 1000$ spikes/s. $\lambda = 50$ Hz, $\beta = 0.04$ s/burst. (C) ISI distribution from NR process with parameters $\kappa_1 = 6$, $\lambda_1 = 96$ bursts/s, $\kappa_2 = 3$, $\lambda_2 = 1500$ spikes/s. $\lambda = 80$ Hz, $\beta = 0.06$ s/burst. (D) ISI distribution from NR process with parameters $\kappa_1 = 5$, $\lambda_1 = 50$ bursts/s, $\kappa_2 = 6$, $\lambda_2 = 6000$ spikes/s. $\lambda = 100$ Hz, $\beta = 0.1$ s/burst.



S2 Experimental data is well fit by sum of gamma distributions. (A) ISI distribution of experimentally recorded bSbC in **Fig 1C** with sum of gammas fit. Goodness of fit is reported as RMSE. (B) ISI distribution of experimentally recorded OFFsA in **Fig 1D** with sum of gammas fit.

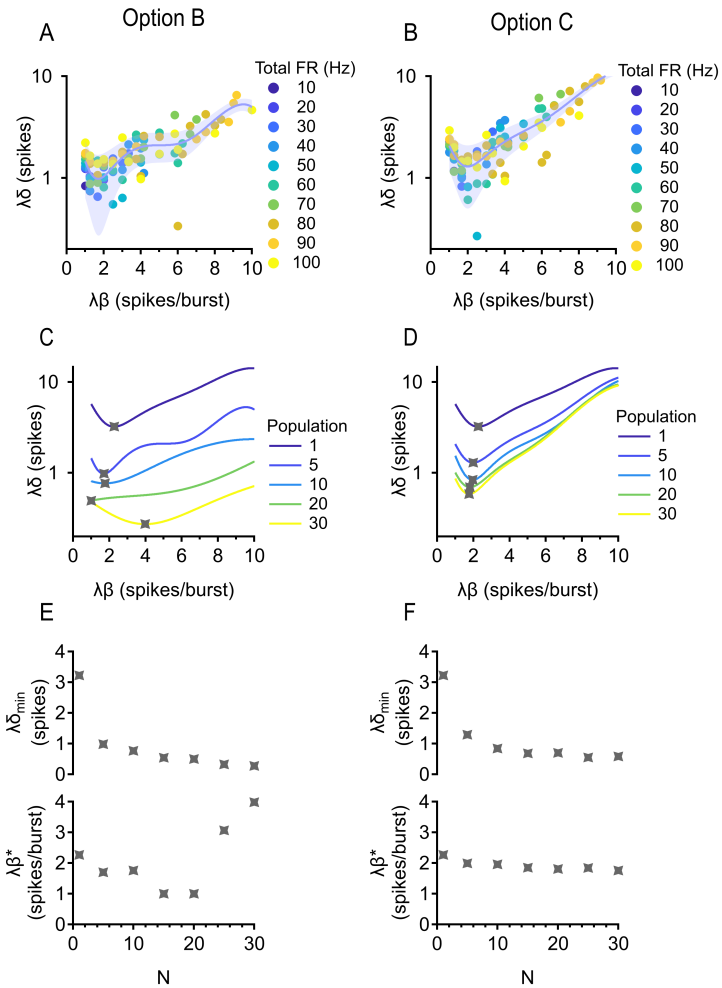


S3 Noise correlations are low. (A) Cross correlation histograms from a pair of bSbCs. Dashed lines are null (epoch-shuffled) correlation. Paired activity was recorded at a mean luminance of 0 and 1172 isomerizations (R*) per rod per second. (B) Cross correlation histograms from a pair of OFFsAs. Dashed lines are null correlation. Paired activity was recorded at a mean luminance of 0 and 1172 isomerizations (R*) per rod per second.



S4 Results are robust to responsivity. (A) Reaction time δ vs. burstiness β with information train readout. Responsivity is fixed at 60%, population is fixed at 5 cells, correlation is fixed between 0-0.2, and total firing rate is fixed at 30, 60, and 100 Hz. (B) Dimensional analysis collapses the data for different firing rates. Dimensionless reaction time is plotted against dimensionless burstiness. Responsivity is fixed at 60%,

population is fixed at 5 cells, and correlation is fixed between 0-0.2. Trial-averaged (median) data is shown. Error bar = standard error estimate of the data around the fit. (C) Existence of a minimum is robust across population sizes. Responsivity is fixed at 60% and correlation is fixed between 0-0.2. (D) Top: minimum (dimensionless) reaction time against population size. Bottom: optimal (dimensionless) burstiness against population size. Responsivity is fixed at 60% and correlation is fixed between 0-0.2.



S5 Results are robust to different ways of constructing population information trains. (A, B) Dimensionless reaction time measured using options B (left) and C (right) vs. dimensionless burstiness. Population is fixed at 5 cells and correlation is fixed between 0-0.2. Trial-averaged (median) data is shown. Error bars = standard error estimate of the data around the fits. (C, D) Existence of a minimum is robust across population sizes. Correlation is fixed between 0-0.2. (E, F) Top: minimum (dimensionless) reaction time against population size. Bottom: optimal (dimensionless) burstiness against population size.

References

1. Rieke F, Warland D, Van Steveninck R de R, Bialek W. Spikes: exploring the neural code. MIT press; 1999.
2. Cessac B, Paugam-Moisy H, Viéville T. Overview of facts and issues about neural coding by spikes. J Physiol Paris. 2010;104: 5–18.
3. da Silveira RA, Rieke F. The Geometry of Information Coding in Correlated Neural Populations. arXiv [q-bio.NC]. 2021. Available: http://arxiv.org/abs/2102.00772
4. Panzeri S, Schultz SR, Treves A, Rolls ET. Correlations and the encoding of information in the nervous system. Proc Biol Sci. 1999;266: 1001–1012.
5. Ainsworth M, Lee S, Cunningham MO, Traub RD, Kopell NJ, Whittington MA. Rates and rhythms: a synergistic view of frequency and temporal coding in neuronal networks. Neuron. 2012;75: 572–583.
6. Ratté S, Hong S, De Schutter E, Prescott SA. Impact of neuronal properties on network coding: roles of spike initiation dynamics and robust synchrony transfer. Neuron. 2013;78: 758–772.
7. de la Rocha J, Doiron B, Shea-Brown E, Josić K, Reyes A. Correlation between neural spike trains increases with firing rate. Nature. 2007;448: 802–806.
8. Bowman DM, Eggermont JJ, Smith GM. Effect of Stimulation on Burst Firing in Cat Primary Auditory Cortex. Journal of Neurophysiology. 1995;74: 1841–1855.
9. King A, Bajo V, Bizley J, Campbell R, Nodal F, Schulz A, Schnupp J. Physiological and Behavioral Studies of Spatial Coding in the Auditory Cortex. Hearing Research. 2007;229: 106–115.
10. Tripathy S, Gerkin R. Spontaneous Firing Rate. 2016 [cited 14 Mar 2022]. In: NeuroElectro [Internet] Available from: `https://neuroelectro.org/ephys_prop/18/`
11. Romo R, Salinas E. Touch and go: decision-making mechanisms in somatosensation. Annu Rev Neurosci. 2001;24: 107–137.
12. Britten KH, Shadlen MN, Newsome WT, Movshon JA. The analysis of visual motion: a comparison of neuronal and psychophysical performance. J Neurosci. 1992;12: 4745–4765.
13. Gold JI, Shadlen MN. The neural basis of decision making. Annu Rev Neurosci. 2007;30: 535–574.
14. Kase M, Miller DC, Noda H. Discharges of Purkinje cells and mossy fibres in the cerebellar vermis of the monkey during saccadic eye movements and fixation. J Physiol. 1980;300: 539–555.
15. Ohtsuka K, Noda H. Discharge properties of Purkinje cells in the oculomotor vermis during visually guided saccades in the macaque monkey. J Neurophysiol. 1995;74: 1828–1840.
16. McCormick DA, Feeser HR. Functional implications of burst firing and single spike activity in lateral geniculate relay neurons. Neuroscience. 1990;39: 103–113.

17. Legendre P, McKenzie JS, Dupouy B, Vincent JD. Evidence for bursting pacemaker neurones in cultured spinal cord cells. Neuroscience. 1985;16: 753–767.

18. Azouz R, Jensen MS, Yaari Y. Ionic basis of spike after-depolarization and burst generation in adult rat hippocampal CA1 pyramidal cells. J Physiol. 1996;492 (Pt 1): 211–223.

19. Raus Balind S, Magó Á, Ahmadi M, Kis N, Varga-Németh Z, Lőrincz A, et al. Diverse synaptic and dendritic mechanisms of complex spike burst generation in hippocampal CA3 pyramidal cells. Nat Commun. 2019;10: 1859.

20. Harris KD, Hirase H, Leinekugel X, Henze DA, Buzsáki G. Temporal interaction between single spikes and complex spike bursts in hippocampal pyramidal cells. Neuron. 2001;32: 141–149.

21. Llinás RR. Intrinsic electrical properties of mammalian neurons and CNS function: a historical perspective. Front Cell Neurosci. 2014;8: 320.

22. Destexhe A, Paré D. Impact of network activity on the integrative properties of neocortical pyramidal neurons in vivo. J Neurophysiol. 1999;81: 1531–1547.

23. Zeldenrust F, Wadman WJ, Englitz B. Neural Coding With Bursts—Current State and Future Perspectives. Front Comput Neurosci. 2018;12. doi:10.3389/fncom.2018.00048

24. De Zeeuw CI, Hoebeek FE, Bosman LWJ, Schonewille M, Witter L, Koekkoek SK. Spatiotemporal firing patterns in the cerebellum. Nat Rev Neurosci. 2011;12: 327–344.

25. Goetz J, Jessen ZF, Jacobi A, Mani A, Cooler S, Greer D, et al. Unified classification of mouse retinal ganglion cells using function, morphology, and gene expression. doi:10.1101/2021.06.10.447922

26. Rodieck RW. The first steps in seeing. Sinauer Associates; 1998. p. 562.

27. Jacoby J, Schwartz GW. Typology and Circuitry of Suppressed-by-Contrast Retinal Ganglion Cells. Front Cell Neurosci. 2018;12: 269.

28. Wienbar S, Schwartz GW. Differences in Spike Generation Instead of Synaptic Inputs Determine the Feature Selectivity of Two Retinal Cell Types. 2021. doi:10.2139/ssrn.3948204

29. Shannon CE. A mathematical theory of communication. SIGMOBILE Mob Comput Commun Rev. 2001;5: 3–55.

30. van Vreeswijk C. Stochastic Models of Spike Trains. In: Grün S, Rotter S, editors. Analysis of Parallel Spike Trains. Boston, MA: Springer US; 2010. pp. 3–20.

31. Heeger D. Poisson model of spike generation. Handout, Stanford University. 2000;5: 76.

32. Yannaros N. On Cox processes and gamma renewal processes. J Appl Probab. 1988;25: 423–427.

33. Gaby S. A doubly stochastic model for the quantification and classification of burstiness and regularity in single spike trains. Front Comput Neurosci. 2011;5. doi:10.3389/conf.fncom.2011.53.00185

34. Jacoby J, Zhu Y, DeVries SH, Schwartz GW. An Amacrine Cell Circuit for Signaling Steady Illumination in the Retina. Cell Rep. 2015;13: 2663–2670.

35. Tiesinga P, Fellous J-M, Sejnowski TJ. Regulation of spike timing in visual cortical circuits. Nat Rev Neurosci. 2008;9: 97–107.

36. Kumbhani RD, Nolt MJ, Palmer LA. Precision, Reliability, and Information-Theoretic Analysis of Visual Thalamocortical Neurons. Journal of Neurophysiology. 2007. pp. 2647–2663. doi:10.1152/jn.00900.2006

37. Li M, Tsien JZ. Neural Code-Neural Self-information Theory on How Cell-Assembly Code Rises from Spike Time and Neuronal Variability. Front Cell Neurosci. 2017;11: 236.

38. Petersen RS, Panzeri S, Diamond ME. Population coding of stimulus location in rat somatosensory cortex. Neuron. 2001;32: 503–514.

39. Averbeck BB, Lee D. Neural noise and movement-related codes in the macaque supplementary motor area. J Neurosci. 2003;23: 7630–7641.

40. Oram MW, Hatsopoulos NG, Richmond BJ, Donoghue JP. Excess synchrony in motor cortical neurons provides redundant direction information with that from coarse temporal measures. J Neurophysiol. 2001;86: 1700–1716.

41. Nirenberg S, Carcieri SM, Jacobs AL, Latham PE. Retinal ganglion cells act largely as independent encoders. Nature. 2001;411: 698–701.

42. Schwartz G, Macke J, Amodei D, Tang H, Berry MJ. Low error discrimination using a correlated population code. J Neurophysiol. 2012;108: 1069–1088.

43. Soo FS, Schwartz GW, Sadeghi K, Berry MJ. Fine Spatial Information Represented in a Population of Retinal Ganglion Cells. J Neurosci. 2011;31: 2145–2155.

44. Cafaro J, Rieke F. Noise correlations improve response fidelity and stimulus encoding. Nature. 2010;468: 964–967.

45. Josić K, Shea-Brown E, Doiron B, de la Rocha J. Stimulus-dependent correlations and population codes. Neural Comput. 2009;21: 2774–2804.

46. Ponce-Alvarez A, Thiele A, Albright TD, Stoner GR, Deco G. Stimulus-dependent variability and noise correlations in cortical MT neurons. Proc Natl Acad Sci U S A. 2013;110: 13162–13167.

47. Buckingham E. On Physically Similar Systems; Illustrations of the Use of Dimensional Equations. Phys Rev. 1914;4: 345–376.

48. Kepecs A, Lisman J. Information encoding and computation with spikes and bursts. Network. 2003;14: 103–118.

49. Oswald A-MM, Chacron MJ, Doiron B, Bastian J, Maler L. Parallel processing of sensory input by bursts and isolated spikes. J Neurosci. 2004;24: 4351–4362.

50. Palmer SE, Wright B, Doupe A, Kao M. Variable but not random: temporal pattern coding in a songbird brain area necessary for song modification. J Neurophysiol. 2021 125:2, 540-555.

51. Schwartz GW. Contrast suppression. Retinal Computation. Academic Press; 2021. p. 103.

52. Homann J, Freed MA. A Mammalian Retinal Ganglion Cell Implements a Neuronal Computation That Maximizes the SNR of Its Postsynaptic Currents. J Neurosci. 2017;37: 1468–1478.

53. Van Wyk M, Wässle H, Taylor WR. Receptive field properties of ON- and OFF-ganglion cells in the mouse retina. Vis Neurosci. 2009;26: 297.

54. Münch TA, Da Silveira RA, Siegert S, Viney TJ, Awatramani GB, Roska B. Approach sensitivity in the retina processed by a multifunctional neural circuit. Nat Neurosci. 2009;12: 1308–1316.

55. Krishnamoorthy V, Weick M, Gollisch T. Sensitivity to image recurrence across eye-movement-like image transitions through local serial inhibition in the retina. Borst A, editor. Elife. 2017;6: e22431–e22431.

56. Wang F, Li E, De L, Wu Q, Zhang Y. OFF-transient alpha RGCs mediate looming triggered innate defensive response. Curr Biol. 2021;31: 2263–2273.e3.

57. Li M, Xie K, Kuang H, Liu J, Wang D, Fox GE, et al. Spike-timing patterns conform to gamma distribution with regional and cell type-specific characteristics. Biorxiv. 2017; 145813.

58. Gallego JA, Perich MG, Miller LE, Solla SA. Neural Manifolds for the Control of Movement. Neuron. 2017;94: 978–984.

59. Wärnberg E, Kumar A. Perturbing low dimensional activity manifolds in spiking neuronal networks. PLoS Comput Biol. 2019;15: e1007074.