LETTER Communicated by Friedrich Sommer

Parallel Neural Multiprocessing with Gamma Frequency Latencies

Ruohan Zhang
zharu@utexas.edu
Dana H. Ballard
danab@utexas.edu
Department of Computer Science, University of Texas at Austin, Austin, TX 78712, U.S.A.

The Poisson variability in cortical neural responses has been typically modeled using spike averaging techniques, such as trial averaging and rate coding, since such methods can produce reliable correlates of behavior. However, mechanisms that rely on counting spikes could be slow and inefficient and thus might not be useful in the brain for computations at timescales in the 10 millisecond range. This issue has motivated a search for alternative spike codes that take advantage of spike timing and has resulted in many studies that use synchronized neural networks for communication. Here we focus on recent studies that suggest that the gamma frequency may provide a reference that allows local spike phase representations that could result in much faster information transmission. We have developed a unified model (gamma spike multiplexing) that takes advantage of a single cycle of a cell's somatic gamma frequency to modulate the generation of its action potentials. An important consequence of this coding mechanism is that it allows multiple independent neural processes to run in parallel, thereby greatly increasing the processing capability of the cortex. System-level simulations and preliminary analysis of mouse cortical cell data are presented as support for the proposed theoretical model.

1 Introduction

The earliest action potential recordings played a crucial role in characterizing the receptive fields in the striate cortex, but with the study of whole circuits (Roelfsema, Lamme, and Spekreijse, 2000; Churchland et al., 2012), the role of oscillations in modulating action potentials has become more important. The possible role of timing-based code was due to Abeles (1991). Subsequent studies showed large areas of synchronization transiting cortical maps (Roelfsema et al., 1994; Singer, 2007; Sejnowski & Paulsen, 2006), which has been refined into a general view of coherent communication (Womelsdorf et al., 2007; Fries, 2015). Now we can study a pool of

Neural Computation **32, 1635–1663** (2020)
https://doi.org/10.1162/neco_a_01301
© 2020 Massachusetts Institute of Technology

synchronized cells representing a specific computation that we will denote as a neural *process*.

With this perspective, we can ask, Is it possible to have multiple cortical processes in parallel? This question surfaced 20 years ago as the binding problem: If, in the cortex, "red" neurons and "blue" neurons are active along with "square" neurons and "circle"' neurons, how does the cortex distinguish the one color-shape pairing from the other? At the time, the great majority recognized the need to solve this problem, but no solutions were proposed (Roskies, 1999). The central problem is partitioning the extensively connected circuitry in a way that avoids possible cross-talk between processes that need to be kept separate.

One easy answer would be to only allow one process to be active at a time. However, this solution is unlikely given the myriad different computations in brain circuitry. Additionally, there is significant evidence for parallel processes. For example, in modeling the multiple spike recordings in a monkey tactile discrimination experiment, Kobak et al. (2016) developed a novel analysis technique that can isolate the spikes that account for the variance in task parameters. Their analysis shows that only 14% of the spikes can account for all the task variance. However, from the perspective of multiprocessing, the remaining 86% of the spikes are presumably doing something other than the controlled task. A second example is Cisek's (2012) monkey experiment in motor choices. A monkey knows that it they will have to choose one of the two movements, but must keep both simultaneously active until the appropriate cue. Neural recordings show separated activities for the possible choices (Cisek, 2012). To distinguish between them, presumably there must be a mechanism to prohibit possible interference between the two.

To allow a cortical network to perform multiprocessing, we propose to use *multiplexing*, wherein the meaning of a spike generated by a neuron can change on short timescales, that is, several active processes may time-share this neuron. We propose a neural coding model, gamma spike multiplexing (GSM), that adopts an efficient phase coding strategy at the level of single cells. We further show that neural networks can be set up in a novel way that allows several processes to run in parallel. Our multiplexing model is unique by using the following conventions in combination:

1. Each process uses a unique gamma frequency.
2. Cells in a process are chosen probabilistically on each gamma cycle.
3. Cells can participate in more than one process at different cycles.
4. The state of a process is represented by different cells on each gamma cycle.

A key element of the GSM model is a phase coding model using gamma oscillations in the somatic membrane potentials as references. The local phase coding model was suggested over 20 years ago (Hopfield, 1995).

The proposed model resembles several existing models (Vinck & Bosman, 2016a; Vinck et al., 2010; Nadasdy, 2009; Ballard & Jehee, 2011) that allow each spike to represent an analog quantity.

To validate the proposed GSM model, we first demonstrate in a simulation that this multiplexing model allows separate processes to be executed in parallel successfully. The mixing process can still result in individual neurons exhibiting classical Poisson statistics. We then analyze mouse cortical cell data recorded with a patch-clamp technique and show that several preconditions of the GSM models are biologically plausible. The patch-clamp technique (Perrenoud, Pennartz, & Gentet, 2016), allows access to the fine structure of a cells membrane potential in awake animals. This capability is necessary for studying gamma oscillations within the membrane potential, which can have very small amplitudes of a few millivolts and can be challenging to separate from large-scale network modulatory oscillation signals (Fries, 2015; Fries, Reynolds, Rorie, & Desimone, 2001; Chalk et al., 2010). Advances in patch-clamp membrane temporal resolution have allowed us to extract the local spike phase automatically from intracellular data (Perrenoud et al., 2016) and provide evidence for our model.

2 Gamma Spike Multiplexing Model

The GSM model has three major components. At the single-cell level, we define a phase coding model that uses minima in gamma frequency oscillations as references. At the network level, we describe how neurons are selected by multiple neural processes in parallel. We then propose a novel multiplexing model that allows a group of neurons to temporarily set up a private network with a designated frequency.

2.1 Gamma Frequency Phase Coding Model. Various population coding models that assume rate coding (Gerstner et al., 1997) have shown enormous usefulness in producing correlates of behavior. Meanwhile, positing the use of population coding as the cortex's generative model is not without debate. Any explicit aggregation of spikes requires counting significant numbers of spikes, which is expensive in both spikes and synapses (Lennie, 2003). A simple theoretical analysis in appendix A shows that with a basic rate coding model, it may take a large number of neurons or spikes to signal a scalar value reliably.

Alternatively, we discuss a gamma frequency phase coding model (Hopfield, 1995; Nadasdy, 2009; Ballard & Jehee, 2011). The advantage of this model is that it allows a single spike to represent an analog quantity instead of a digital quantity (0 or 1), thereby making our model more efficient. The concept is shown in Figure 1a. The spike's delay code Δ is computed as the interval between the spike and the beginning of a gamma cycle. The closer

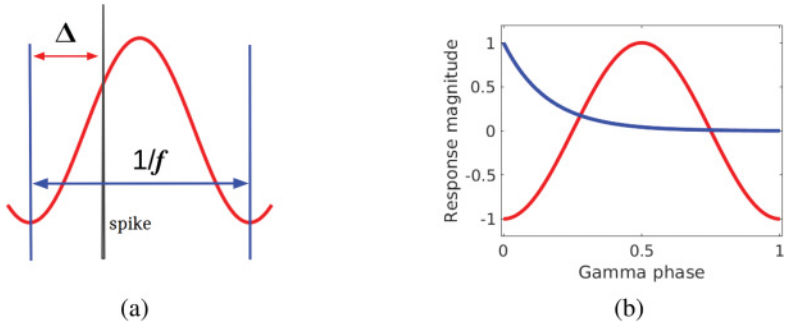

Figure 1: (a) Gamma frequency phase coding model. The red curve indicates the somatic oscillation in the range of gamma frequency. The spike's latency code Δ is then computed as the intervals between the gamma trough and the spike. (b) A spike signals an analog number (blue) coded as a phase delay from the trough. Short delays represent large magnitudes.

the spike is to the start of the cycle, the larger the quantity it represents. The magnitude of response r is computed by

$$r = \exp(-\alpha l), \tag{2.1}$$

where l is the gamma phase and α is a constant, as shown in Figure 1b.

The ability of a phase coding model to represent an analog quantity using a single spike makes it significantly more efficient than the basic rate/population coding. However, an unresolved issue is to select the gamma cycle starting point. Guided by the phase precession model for place cells (O'Keefe & Recce, 1993; Skaggs, McNaughton, Wilson, & Barnes, 1996), we use the gamma trough as shown in Figure 1a. This choice is further resolved by analyzing experimental data. The detailed analysis with mouse cortical cell data that supports this claim is provided in section 4. This model will serve as the basis for the rest of our multiplexing model.

2.2 Probabilistic Parallel Neuron Selection. Given that a spike can represent an analog quantity, we then discuss how the cortex selects neurons to participate in computational processes. The GSM model assumes that the cortical maps are overcomplete and neurons are selected probabilistically in parallel.

2.2.1 Sparse Coding Strategy. The first assumption is the overcompleteness of cortical maps, that is, the cortex contains many times the number of neurons that would be mathematically needed to code an input. Consequently, a sparse coding strategy asserts that the early visual cortex develops neurons with receptive fields to represent the statistics of natural

images, and, when coding a stimulus, relatively few neurons are chosen to be active at any one time (Olshausen & Field, 1996, 1997). The overcompleteness property reduces the probability of competing for the same neuron by multiple processes.

2.2.2 Probabilistic Selection. The neurons are selected to code stimuli probabilistically with odds related to their receptive field projections. Given a target stimulus (represented as a vector), a neuron's response magnitude r_i can be calculated as the projection of the basis function it represents onto the stimulus vector.[1] A standard choice is to use a Boltzmann distribution where a cell is chosen with probability

$$p(r_i) = \frac{\exp(-\eta r_i)}{\sum_j \exp(-\eta r_j)}, \tag{2.2}$$

where η is its "temperature" parameter. The relative magnitudes of the input's projections determine the probabilities of a neuron being selected to be part of the representation. The divisive normalization can be done in one step with lateral connections between neurons. Selecting neurons in this way ensures that all neurons participate in the receptive field learning process appropriately (Jehee, Rothkopf, Beck, & Ballard, 2006). If the cell with the highest response is always selected as in maximum likelihood selection, unselected cells are never able to adjust their receptive fields.

The probabilistic selection has an even more significant consequence. If the stimulus is to be maintained for other gamma cycles, its code must be chosen anew on each such cycle. The state of a process migrates to a new group of cells on each iteration.

2.2.3 Parallel Selection. Given that a single candidate neuron is chosen probabilistically, one still needs to find a set of neurons to represent a stimulus. The standard residual-based method computes this set sequentially. However, in the GSM model, each step of sequential computation would require one gamma cycle, thus ruling it out as practical. To overcome this problem, we propose a parallel selection algorithm. Figures 2a and 2b compare the sequential and parallel approaches. The stimulus vector (black) and basis functions of each neuron (gray) are represented as 2D vectors. Figure 2a shows the sequential method. The contribution of a cell (red) selected for representing the image can be determined by subtraction (green). In contrast, the parallel method, shown in Figure 2b, chooses a fixed number of cells probabilistically using equation 2.2 in parallel. In appendix B, we assess the accuracy of the probabilistic parallel selection algorithm in simulation, but for the moment, an important consequence of the parallel

[1] A neuron's vector of synaptic strengths is referred to as a basis function.

Figure 2: Parallel coding model. (a) A standard algorithm for coding an input (black arrow) picks a neuron (red) and sequentially fit the residual (green), a process that requires multiple iterations. (b) The GSM chooses several representing neurons in a single parallel step.

selection is that such representation is nonstationary, as shown in a related context by Druckmann and Chklovskii (2010).

2.3 Neural Multiplexing Model. Once the phase coding and neuron selection mechanisms are in place, an elegant way of instantaneously connecting large networks implicitly is possible, as shown in Figure 3. If a selected subset of cells were comodulated with the same gamma frequency simultaneously, then all the neurons in that cohort would be behaving as engaged in the same computation. An analogy that helps understand this model is radio communication. A group of neurons can tune to a particular frequency to establish a private communication channel, although gamma oscillations are not used as a signal carrier but as a temporal reference. Collisions are handled by a gating mechanism depicted in Figure 3b. If a neuron is chosen by two or more frequencies, it will not be allowed to spike. Although the exact mechanism is unknown, it is conjectured that such a mechanism is biologically plausible, for instance, through an interneuron's synaptic gating function (Katz, 2003; Herberholz, Antonsen, & Edwards, 2002; Evans, Jing, Rosen, & Cropper, 2003; Ivanov & Calabrese, 2003; Gisiger & Boukadoum, 2011), which might use subtreshold signals (Borg-Graham, Monier, & Fregnac, 1998).

A prerequisite for the model being able to support multiple computations is the ability of a neuron to modulate its oscillation frequency. It has already been observed that the gamma signal can change in a single cycle in vivo (Atallah & Scanziani, 2009; Perrenoud et al., 2016). While there is still no complete understanding of how the neural circuitry could control this functionality, Atallah and Scanziani (2009) have suggested that the needed gamma modulation can be controlled by adjusting excitation and inhibition. Therefore, the membrane potential can be modulated instantaneously. This demonstration is complemented by the circuit model of Brunel and Wang (2003) that illustrates these features and the additional possibility of $\sim$10 ms spike phase delays. A possibility is that some interneurons may

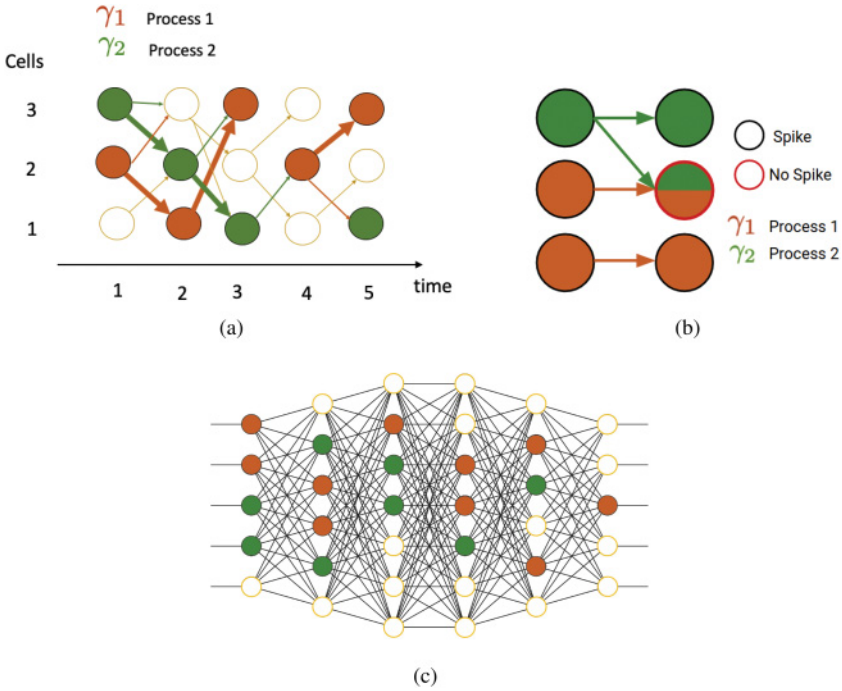

Figure 3: Neural multiplexing model. (a) The figure shows three neurons, each depicted at five successive time instants. The somas are being modulated by two particular gamma frequencies at different times. Consequentially those neurons that are connected at the same frequency can communicate at those instants privately. A neuron can be chosen by different frequencies at different times. (b) Two time steps in neural processing with two processes. If one neuron is chosen by two gamma frequencies, it does not spike. Thus, only neurons that are chosen by a single gamma frequency can spike. If cells are not chosen by a gamma frequency, they do not spike. (c) At any given time, the network is effectively partitioned into multiple separate neural processes; within each, the neurons communicate using their private gamma frequency band.

contribute to columnar synchronization of activity in the gamma frequency range (Llinas, Grace, & Yarom, 1991). In section 4, we provide experimental evidence that frequency modulation is indeed possible.

3 Simulation Results: A Sparse Coding Example

In a general multiplexing situation, one would expect multiple routines to be instantiated at the most abstract forebrain levels and migrate quickly throughout cortical maps, ultimately polling the lateral geniculate nucleus

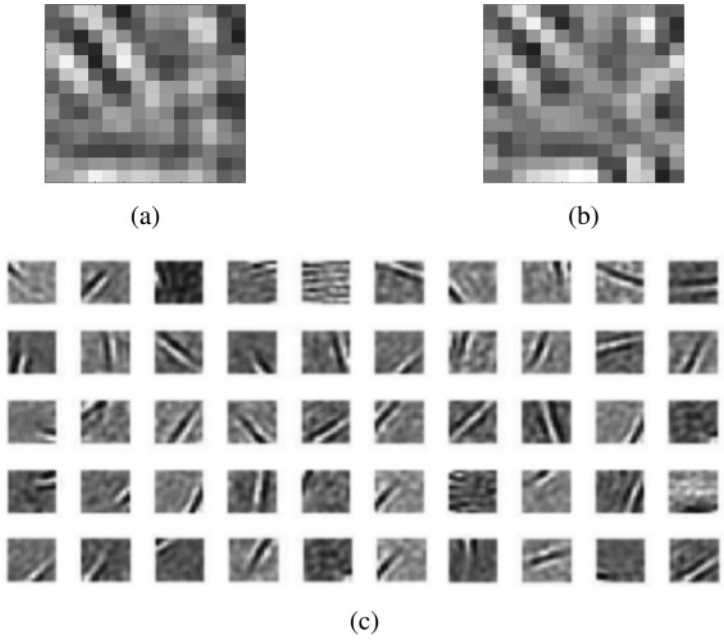

Figure 4: A sparse coding example. (a) In the model, 196 LGN cells appropriately filtered provide input for a small patch sampled from a camera. Their analog values can be depicted abstractly as a vector as shown by the black arrow in Figure 2b. (b) The LGN input is used to choose cortical cells by randomly sampling cells' probability density function of the input projections. (c) Fifty neurons (basis functions) selected by the parallel probabilistic algorithm to code the image patch. The basis functions are learned using the sparse coding algorithm (Lee et al., 2007).

(LGN) for visual input. To sidestep the enormous effort it would take to build a model at this level of completeness, we use a much simpler model–the familiar computation of representing small image patches with receptive fields in the striate cortex using sparse coding (Olshausen & Field, 1996, 1997). Figure 4 provides concrete examples of the neural process we simulated: coding an image patch with 50 neurons. The "LGN" generates the image patches that are sent to the simulated visual area V1 and encoded. Each V1 neural process selects a subset of learned basis functions to encode a single image patch. The V1 basis functions are learned using the algorithm by Lee, Battle, Raina, and Ng (2007). Appendix B includes more examples of learned basis functions and coded image patches.

The basic setting of multiplexing is thus to code multiple image patches in a short time window. At any given discrete time step, several image patches need to be coded simultaneously.

Table 1: Parameter Values Used in the Simulations.

Name	Value	Unit	Description	Function
T	800	Milliseconds	Total length of simulation	simulation
n_b	1000	Scalar	Total number of neurons	simulation
n_s	14×14	Pixels	Dimension of input and basis functions	simulation
n_r	50	Scalar	Number of neurons to code the stimulus	simulation
N_p	16	Scalar	Number of processes	process
f_p	$40 + \mathcal{N}(0, 7)$	Hz	Frequency to modulate a process	process
t_γ	rand(20, 22)	Cycles	Process duration	process
d_γ	5	Milliseconds	Max length of the delay	spike
d_r	4	Milliseconds	Refractory period	spike

Notes: The function rand(a,b) returns a single uniformly distributed random number in the interval [x,y]. $\mathcal{N}(\mu, \sigma)$ is the normal distribution.

First, we need to calculate each neuron's response for an image patch and convert the response to a delay code. We use r_i to denote the ith neuron's response to its input, $\mathbf{x_i}$ to denote the input vector, and $\mathbf{w_i}$ to denote the recipient neuron's basis function. A conventional way to calculate r_i is $r_i = g(\mathbf{x_i} \cdot \mathbf{w_i})$, where g is a nonlinear activation function and only positive responses are maintained. The response is converted to a delay as shown in Figure 1a. Coding neurons for a particular patch are selected using the aforementioned probabilistic parallel selection algorithm.

Table 1 shows the parameters chosen for the simulation. It is important to choose parameters that are biologically plausible for our later analyses:

1. The total number of processes N_p. In primates, the number of simultaneously active processes may be linked to working memory, and in that case, the number would be four (Luck & Vogel, 1997). However much less is known about the use of unconscious, overlearned memory, which may have a significantly higher figure, which we choose to be 16.
2. The number of gamma cycles used by an individual process t_γ. This parameter is reasonably guided by the duration of gaze fixation, which indicates the length of a visual process, typically having a mean of approximately 200 to 600 milliseconds.
3. The time needed for the delay code d_γ. This choice is bounded by the observation that sensitivity to the Pulfrich pendulum illusion (Lit, 1960), which implies a sensitivity to about 5 to 20 milliseconds and the need to keep the phase delay less than half the gamma cycle.

In the simulation, the initial step is to specify the number of processes. For each process, a start time, dedicated gamma frequency, and duration

are chosen. Next, its coding neurons are selected. If a neuron is chosen for a process, during the time for the delay and following a refractory period, it is marked as in use and cannot be selected by another process.

3.1 Validating the Coding Algorithm. We first validate our implementation of the sparse coding algorithm (Lee et al., 2007), and more important, the parallel probabilistic selection algorithm. Appendix B provides a concrete example of the neural process we simulated: coding several image patches in parallel with 50 neurons for each patch. It further shows that the parallel probabilistic selection algorithm results in good coding quality with enough neurons (50 or more). With 50 neurons, the mean coding error per pixel is 4.0×10^{-3} (standard error $= 0.4 \times 10^{-3}$, which is considered small given pixel values are in the range of [0,1]).

3.2 Poisson Statistics and Parameter Sensitivity. We now present the main simulation results. The most basic firing pattern of cortical cells exhibits Poisson distributions (Gerstein & Melbrot, 1964; Shadlen & Newsome, 1998; Berry & Meister, 1998). These distributions comprise the core of various coding models. Therefore, it is crucial for any proposed theoretical model to obey Poisson statistics in simulation, regardless of the parameters chosen. A standard way to test this is to calculate the Fano ratio, that is, the ratio between the variance and the mean of a random process in some time window (Dayan & Abbott, 2001). For a Poisson process, the variance in the count equals the mean count, so its Fano ratio is 1.

The important parameter here is the number of processes. In terms of generating spikes that appear Poisson distributed, the GSM model is surprisingly insensitive to the parameter choice, as shown in Figure 5. It is important to emphasize that many sets of parameters would work to illustrate that the GSM coding strategy produces distributions of spikes that appear Poisson on long timescales.

3.3 Observability of Gamma Oscillation. The proposed model, if true, would drastically change the way we interpret spike data. Now, individual spikes can be identified with the specific image patch that is being coded and can be readily associated with their individual processes. Appendix C presents a visual comparison between the old and the new ways of interpreting spike train data. This new interpretation could potentially resolve a highly debated issue that gamma frequencies are observed in some experiments but not others. This phenomenon has led researchers to doubt the role of gamma oscillation (Ray & Maunsell, 2015) or even raise the question of whether it is in fact neural noise (Burns, Xing, & Shapley, 2011).

In the classical view, data are acquired by trial averaging of a single neuron where the histories from different trials are summed once a common temporal reference is determined. The main assumption is that the overall process is *ergodic*, that is, trial averaging is deemed to be equivalent to

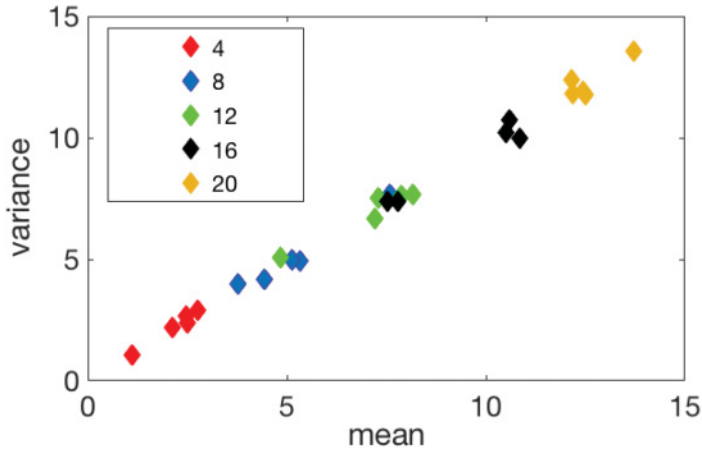

Figure 5: The GSM model generates spikes that appear Poisson distributed in simulation. We calculate the Fano ratio (the ratio between the variance and the mean of a random process) of simulated spike train data while varying the number of processes in simulation. The number of processes varied from 4, the putative value for short-term memory, up to a maximum of 20. Five samples were measured for each setting. The Fano ratios for the simulation samples are very close to an ideal Poisson process, which has a Fano ratio of 1.

averaging many cells in a large network; but according to the GSM model, this equivalence does not hold. From the GSM perspective, trial averaging will have different controlling parameters (those in Table 1) from trial to trial. In contrast, data could also be acquired by ensemble averaging, which uses the same control parameters for all the participating cells from a single trial. The result is that in ensemble averaging, correlations between neurons' oscillations over time can be evident, whereas in trial averaging, they can be weak or nonexistent.

The GSM model allows the evaluation of the extent to which gamma frequencies are observable using simulations. Figure 6 shows the Fourier spectrogram of the trial average case. As the simulation shows, the gamma power is all but absent from the spectrogram. In contrast, Figure 6 further shows the case where the averages are taken in an ensemble, where all coding cells use the same control parameters. The gamma power is very evident. The GSM model obviates the assumption that the neural population in an ensemble obeys an ergodicity where repeatedly sampling one cell is equivalent to sampling a population simultaneously. The bottom line of this spectrogram analysis is to point out that the methodological differences between ensemble averaging and trial averaging may be the reasons why gamma signals are observed in some experiments but not others.

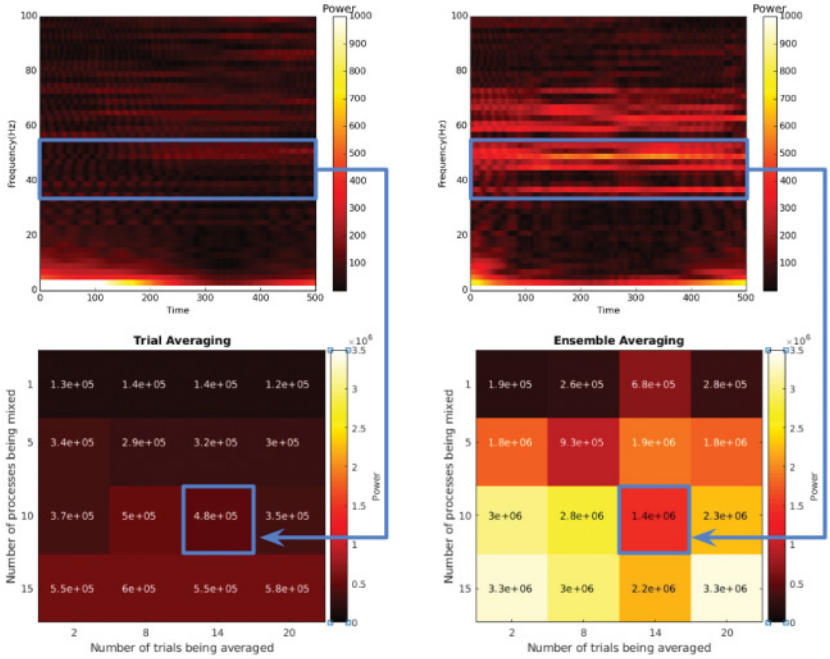

Figure 6: Gamma oscillation is less visible when performing trial averaging. Top row: Fourier spectrograms for trial averaging versus ensemble averaging. Top left: Spectrogram for the trial averaging cases reveals very little power in the gamma band owing to the fact that the controlling parameters are different for each trial. Top right: The ensemble averaging spectrogram reveals marked frequency responses in the gamma band as each process is using a common set of parameters. Bottom row: The effect of the number of processes and number of trials on the aggregated power. The power is aggregated between the 35 Hz to 55 Hz gamma frequency band and over time. The key observation here is that gamma oscillation in the ensemble averaging case is consistently more visible than the trial averaging case regardless of parameters chosen, especially when we mix more than one process. This simulation explains why in experiments, gamma oscillation is sometimes difficult to detect when performing trial averaging.

4 Evidence from Neural Recording Data

We have shown that the GSM model agrees with experimental observations in simulation. The full model cannot be tested until recording technology allows us to perform intracellular recordings of multiple nearby neurons of an awake animal. However, we can test whether some prerequisites of the model are met.

The effects of subthreshold membrane potential can vary (Churchland et al., 2010), but the modulations of spikes by gamma oscillations were observed at least almost 30 years ago (Llinas et al., 1991), where the spike timings were modulated by the gamma oscillations. The recording techniques have greatly improved since then. We used a data set obtained by the Gentet laboratory (Perrenoud et al., 2016) of two-photon targeted patch-clamp recordings in area V1 layers 2/3 of awake mice at 20 kHz. We analyzed recorded somatic membrane potential of 9 pyramidal cells with 536 trials in total. Gamma frequency oscillations (30–80 Hz) are extracted by a Butterworth bandpass filter after spikes are removed. The spike-removal technique was validated experimentally to make sure that it does not produce undesirable artifacts when estimating spike phase. Appendix D shows a representative trial, including the spikes, the somatic potentials, extracted gamma oscillations, and the spike phases.

When we introduced the gamma phase coding model, there was an unsettled issue of picking the gamma cycle starting point. An important observation from the data is that 88.7% of the spikes are between the trough and the peak of a gamma cycle. When the trough is selected as the cycle starting point, the majority of the spikes are in the first half of a gamma cycle, so the delay is in the range of approximately 6 ms to 16 ms (assuming 30 Hz to 80 Hz) as required by the model. The average phase of all 3546 spikes (normalized to 1) is 0.438 (SD = 0.118). As mentioned before, this observation agrees with results from the Pulfrich pendulum illusion study (Lit, 1960), which implies a sensitivity to approximately 5 to 20 milliseconds.

A prerequisite of the multiplexing model that we have noted is that neurons must be able to modulate their somatic gamma oscillation frequencies. Analyses of the patch-clamp data indicate that gamma frequency is modulated by visual stimuli, as shown in Figure 7a. When the visual stimulus is on, this neuron starts to oscillate more in the gamma frequency range. This observation agrees with GSM's prediction that the selected neuron to code the stimulus will be assigned a frequency in that range.

When coding the visual stimulus, our modulation hypothesis requires that different processes use separate frequencies. Thus, the frequency must change over time to allow a neuron to participate in these processes. Besides, it is unlikely that the gamma frequency would be present for just one cycle; it should be stable for several cycles. A standard test for frequency modulation uses the Fourier spectrogram, but this would not have the necessary temporal precision. We locate successive individual zero crossings in the somatic potential and calculate the instantaneous frequencies, as done in Figure 7c. The 20-KHz sampling rate of the data allows for a necessary level of precision.

In Figure 7b, each column shows the frequency histogram from a single trial. The color map emphasizes the maximums in each trial. The result is that certain gamma frequencies are repeated a significant number of times within the recording periods. These preliminary data are encouraging, as

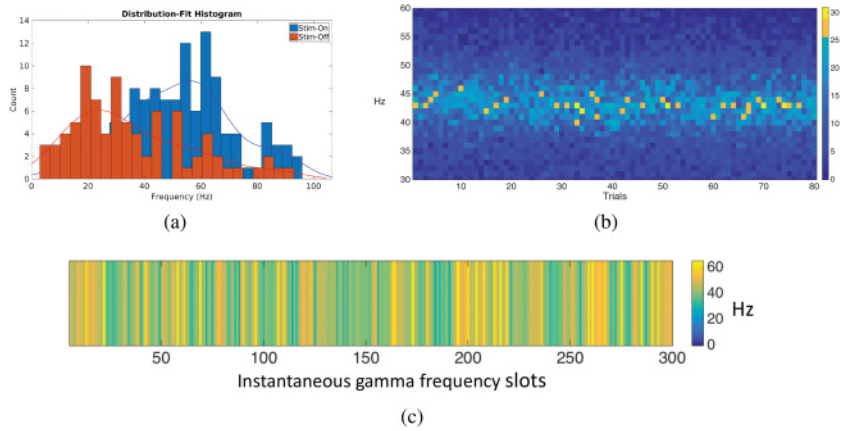

Figure 7: (a) Somatic membrane potential oscillations are modulated by visual stimulus. (b) Instantaneous frequency histograms. Each column represents a histogram (bin size = 1 Hz) of all the gamma frequencies in a single trial. The color map has been adjusted to accentuate the maxima. The hypothesis is that these values may be indicative of particular individual computations. (c) Taking trial 34 as an example, we can ask where the gamma frequencies are in time. This modulation further supports the model as it shows that for a preponderant frequency of 43 Hz, there will be many random "slots" that are available to generate a spike for a computation. The results imply a mixing of gamma frequencies as expected by the model.

the gamma distributions form discrete peaks (instead of the extremes of being focused in a single frequency across trials or being distributed uniformly). Furthermore, the gamma frequency mode is chosen anew on each trial. In summary, these analyses show that somatic gamma oscillations are modulated is a reasonable assumption.

5 Discussion

Conventional spike recording techniques and modulations by oscillations are typically studied at long temporal scales. This fact obscures the details of spike generation that may exploit structure existent at much shorter timescales. This picture is very much echoed by Panzeri, Macke, Gross, and Kayser (2015), in which one question is particularly germane for this study: "How do microscopic-level neural coding mechanisms interact with macroscopic scale aspects of neural activity?"

The focal premise of this letter is that at a fine temporal scale, the cortical neural computation can be factored into independent processes. The model relies crucially on the fine structure of the gamma modulation of a cell's

somatic membrane potential, which allows a spike to send a scalar as a delay and be distinguished with a specific gamma frequency. Our model has the following desirable properties:

1. The phase coding mechanism is significantly more efficient than rate coding models. Since each spike can represent an analog quantity, the circuit is not reduced to counting spikes. The speed of a basic clock cycle of a process is that of a gamma-band frequency.
2. Probabilistic selection ensures that neurons in a large population have a chance of being selected and consequently having their receptive fields being updated regularly.
3. One consequence of this model is that a reconciliation of gamma and Poisson action potential observations is straightforward, as shown in the simulation.
4. A neuron can switch its oscillation frequency rapidly; hence, multiple processes can time-share this neuron, and multiplexing can be achieved.
5. Neurons oscillate in the same frequency can communicate without interference hence multiple processes can coexist.

5.1 Model Assumptions. First, we elaborate the assumptions and issues of the three components of our model in the context of related work.

5.1.1 Phase Coding Model. The gamma frequency phase code is central to the multiplexing model. With rate coding, it is difficult to multiplex and for the downstream neurons to demultiplex the coming spikes. Previous research has shown that cortical circuits can use spike latency codes (Gollisch & Meister, 2008; VanRullen & Thorpe, 2002; Vinck et al., 2010) with reference to a gamma frequency oscillation. The gamma frequency approaches at the network level have demonstrated that long-range communication between distal cells is a reality, in Fries's terms: "communication through coherence" (Fries, 2015).

The proposed model here is similar to the prestigious phase-of-firing model of place cells in the hippocampus with theta frequency oscillation in the local field potentials (LFP) (O'Keefe & Recce, 1993; Skaggs et al., 1996). Note that this is different from our model since the oscillations in the somatic membrane potentials are used as the reference. This type of model was proposed by Hopfield (1995) but limited to the olfactory and auditory systems. The model was further refined by Ballard and Jehee (2011). Some previous work as suggested that this type of phase coding model is indeed plausible and can coexist with rate code (Montemurro, Rasch, Murayama, Logothetis, & Panzeri, 2008; Nadasdy, 2009; Koepsell et al., 2009).

The Pulfrich pendulum illusion (Lit, 1960) strongly suggests that the cortex is sensitive to delays in the range needed by the model. In a model that depends so completely on exact timing, a significant problem to address is

the mechanisms for writing in and reading out its multiplexed messages. From this perspective, the problem of sensory signal acquisition reduces to one of using phase locking to place the input in register with the phase of the process. A natural site for this to happen is the thalamus, and there is evidence that such mechanisms are used in whisking (Yu et al., 2013). Additionally, Koepsell et al. (2009) provide evidence that gamma oscillations at the retina are carried to the cortex through the thalamus.

5.1.2 Probabilistic Parallel Neuron Selection. The selection algorithm utilizes a simple setting of recoding image patches from basic (LGN) representations, but a general capability required is that the new coding cells at the next gamma cycle be created from current coding cells calculated. Such an algorithm has been developed (Druckmann & Chklovskii, 2010) and would be usable by the model with some adjustments to make it parallel.

The ability to code a cell's response in one gamma cycle and also to maintain the code on subsequent cycles is important in its own right, but it can also be compatible with the standard *attractor* model of neural computation introduced by Hopfield and Tank (1986) and Hertz, Krogh, and Palmer (1991), where arbitrary problems can be specified as a fixed point of an appropriate specific network.

5.1.3 Multiplexing Model. At the network level, an issue for larger networks than the one simulated is that the simulation assumes that all neurons that belong to a process are phase-locked. While information traveling between successive neurons can be delayed, there are at least three mitigating circumstances. One is that in the substantially myelinated pyramidal cells, the spike propagation speed is sufficient. The second is that the circuitry between cortical maps may have independent couplings that allow computation to proceed in parallel. The third is that the entire cortex has only 10 levels with connectivity arranged in polytrees, potentially allowing the synchronization procedure to settle fast.

In this study, huge simplifications have been made in the course of the exposition of the main results. A major issue that is sidestepped here is that of the control of a process. Multiplexing implies that there must be a method for initiating a process and determining when it is finished. While a complete account of such a mechanism is very much a research topic, rapidly increasing evidence suggests that lower frequencies–theta, alpha, and beta–may play this role (Michalareas et al., 2016; Van Kerkoerle et al., 2014; van Pelt et al., 2016; Igarashi, Lu, Colgin, Moser, & Moser, 2014; Lisman, 2005; Lisman & Jensen, 2013; McLelland & VanRullen, 2016; Fournier et al., 2019) as they are shown to modulate gamma frequencies.

5.2 Model Implications. Next, we discuss the implications of the GSM model as a whole in the context of previous research.

5.2.1 Implications for Neural Signaling. The GSM model provides several new interpretations of cortical processing. Many similar features of the model have been proposed in other contexts (VanRullen & Thorpe, 2002; Womelsdorf, Fries, Mitra, & Desimone, 2006; Fries, Nikolić, & Singer, 2007; Landau & Fries, 2012; Bastos et al., 2012; Vinck & Bosman, 2016b; Ballard & Jehee, 2011; Ballard & Zhang, 2018), but assembling them into a complete system results in a number of innovative features. The proposed model is naturally sympathetic and complementary to the experimental characterizations of the gamma signal (Fries et al., 2007; Womelsdorf et al., 2007; Landau & Fries, 2012; Bastos et al., 2012; Buzsáki & Wang, 2012), but it has a different interpretation of the role of the gamma frequency band. In these works, gamma synchronization as a medium makes sense of large-scale communication; gamma synchronization in the small addresses the organization of spike codes in the networks. The GSM model posits that its primary message-transmitting role is to allow separate computations to be carried out without crosstalk. In this regard, the model takes a neutral stance as to how the gamma band is sampled, but it may be that this sampling has task-specific components (Brunet et al., 2013).

5.2.2 Membrane Potential Issues. The model has finessed several important issues related to the subthreshold membrane potentials. A variety of studies have shown that the membrane potential is implicated in cell responses outside of its classical receptive field, implying that the cells' functional computations and connectivity are significantly more complex (Grinvald, Lieke, Frostig, & Hildesheim, 1994; Bringuier, Chavane, Glaeser, & Frégnac, 1999; Prechtl, Bullock, & Kleinfeld, 2000). Their illustrations are significant in themselves but may have additional possibilities in different computations that our model and others need. One is in the important suggestion for interarea synchronization by Bastos, Vezoli, and Fries (2015). They propose that the difficult question as to synchronizing feedforward and feedback spikes may be solved with interlaminar delays, which could utilize subthreshold responses that can manipulate the gamma phases (Bringuier et al., 1999). In another issue, our model assumes subthreshold responses could be used for choosing the cells in a network by binding them to a gamma frequency as well as to prohibiting an action potential for a cell when more than one frequency tries to select it (see Figure 3b). Finally, subthreshold potentials could likely be used in setting general attractor computations such as Hopfield and Tank (1986) and Roelfsema, Lamme, and Spekreijse (2000).

5.2.3 Interpreting Receptive Fields. The traditional interpretation of neural spikes, dating from early work by Barlow (1972), is that they belong to a single process and have static interpretations. However, multiplexing changes this interpretation radically and sheds light on an important experiment. In a classical monkey experiment by Moran and Desimone (1985), a V4

receptive field contained two subfields, A and B. By manipulating the attentional state of the animal, either A or B was attended to, resulting in separate spike rates for each of the foci. However, when the animal is attending away from the receptive field, the response rate is the average of the rates from A and B. From the GSM perspective, when attending to subfield A, the spikes for process A dominate the traffic. But when attending away, A and B processes could time-share the receptive field; hence, the spike rate is the average due to multiplexing.

5.2.4 Interpreting Attention. The model also provides an possible interpretation of the effects of attention (Maunsell, 2015; Maunsell & Treue, 2006; Boynton, 2009). This view states that the general effects of attention can be accounted for by a gain factor. The GSM suggests that the gain factor may reflect the use of additional coding neurons by a process. That is, the size of the coding pool is increased directly, resulting in an increased probability of any particular neuron being included in a coding process. This leads to improved coding accuracy analogous to the effects that are observed in our sparse coding example.

While attentional effects can be readily measured experimentally, nonetheless some of the effects appear as small increments on a baseline firing pattern, raising the question of how they can be reliably separated and used (Gilbert & Li, 2013; Roelfsema et al., 2000). The GSM model has a straightforward answer to this issue in that the attentional effects are captured by separate processes that use separate frequencies. This property is particularly evident in Van Kerkoerle, Self, and Roelfsema (2017), where a large neural transient produced by a transient stimulus gap is immediately ignored afterward. The ability of the neuron to immediately filter out such a large transient signal is readily understood if the transient and tracing components can be seen as separate processes that share the recorded neuron but use separate gamma frequencies.

5.2.5 Testing the Model. Eventually the whole model can be tested precisely once we can perform intracellular recordings of multiple nearby neurons of awake animals. One obvious way is to compare trial averaging and ensemble spectrograms, as is done in Figure 6. Power in the gamma range should be much more readily observed in the ensemble recordings. A very exacting test would examine the spectrum for small collections of individual gamma frequencies. This would require estimating the time constants of the coextensive processes, but could be done. The expected result would be that on a timescale of a few hundred milliseconds, the gamma spectra should be discrete, reflecting the number of independent active processes. A final test may come from advancing techniques for identifying neuron assemblies, which could allow the timing relationships between different assemblies to be analyzed in detail (Carrillo-Reid, Han, Taralova, Jebara, & Yuste, 2017).

6 Conclusion

Theoretical computational models are very important tools for understanding the cortex. In this work, we propose the gamma spike multiplexing model that allows neurons to code information efficiently using spike timing and perform multiple computations in parallel. We use system-level simulations as well as in vivo data to characterize the proposed model. The model can be tested thoroughly once the recording technology becomes available. The proposed model, if proven to be true, would change the way we interpret spike data.

Appendix A: Cost of Population Coding

From a purely mathematical perspective, we analyze here the cost of basic rate coding models. We turn to the basic issue of estimating a scalar reliably from spikes that are generated with a probabilistic model. The model chosen is simpler than a Poisson model but is elegant. It illustrates the point that estimating a scalar accurately by counting spikes that are generated probabilistically takes a lot of them.

We use Hoeffding's inequality (Hoeffding, 1994), where m spikes are modeled as $Z_1, \ldots, Z_m$ from a Bernoulli distribution where $P(Z_i = 1) = \phi$ and $P(Z_i = 0) = 1 - \phi$. The estimate of the mean is given by $\hat{\phi} = \frac{1}{m} \sum_i Z_i$. Let any $\mu \geq 0$ be a fixed error criterion. Then applying the inequality,

$$P(|\phi - \hat{\phi}| > \mu) \leq 2e^{-2\mu^2 m}.$$

This formula allows one to estimate the number of cells to be counted to bound the error of the estimate, which is done in Figure 8. This figure shows that the rate/population model may need thousands of neurons or spikes to signal a scalar value reliably (within a reasonable error bound). Meanwhile, the proposed gamma phase coding model requires only a single spike to represent this value.

Appendix B: Sparse Coding and Neuron Selection

Instead of arbitrary computations, we use multiple instances of the familiar computation of representing small image patches with receptive fields in the striate cortex using sparse coding (Olshausen & Field, 1996, 1997). The probabilistic parallel selection algorithm results in good coding quality with 50 to 100 neurons, as shown in Figure 9. Figure 10 shows more coding examples when eight image patches are encoded in parallel, using the proposed neuron selection algorithm. The main result here is that the sparse coding algorithm was implemented correctly.

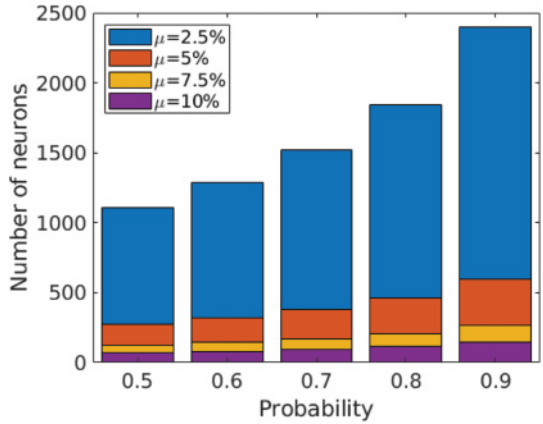

Figure 8: Estimating the cost of population code that assumes rate code. Using a Bernoulli model for spike generation where the mean probability is set to a hypothetical "rate," one can calculate how many spikes are needed in the estimate to bound the error in the estimate. For example, it takes at least 70 coding neurons, such that the coding error will be within $\mu = 10\%$ with probability at least 0.5 (first column, purple). It takes at least 2397 coding neurons, such that coding error will be within $\mu = 2.5\%$ with probability at least 0.9 (last column, blue). Keep in mind that these estimates are for just one scalar.

Appendix C: A New Way of Interpreting Spike Data

The proposed model would greatly change the way we interpret spike data. The spikes generated by the simulation are shown in Figure 11. The simulation uses 1000 coding neurons, but for clarity in the visualization, a representative 50 cells are shown. Figure 11a illustrates the simulation's spikes as might be obtained with a conventional multicell recording. In Figure 11b, all spikes from cells that are coding the same image patch have a common color denoting their instantaneous modulation frequency. Owing to being generated in the context of specific gamma frequencies, individual spikes can be identified with the specific image patch that they are coding. Without the multiprocessing context, the spike trains appear very conventional, but once the generating context is available, the spikes can be readily associated with their processes. Additionally, it can be seen by inspection that a single coding cell's spikes can encode different patches.

Appendix D: Neural Recording Data

One trial from a pyramidal cell is shown in Figure 12 with four spikes extracted from 7 seconds of a cell's membrane potential using patch-clamp

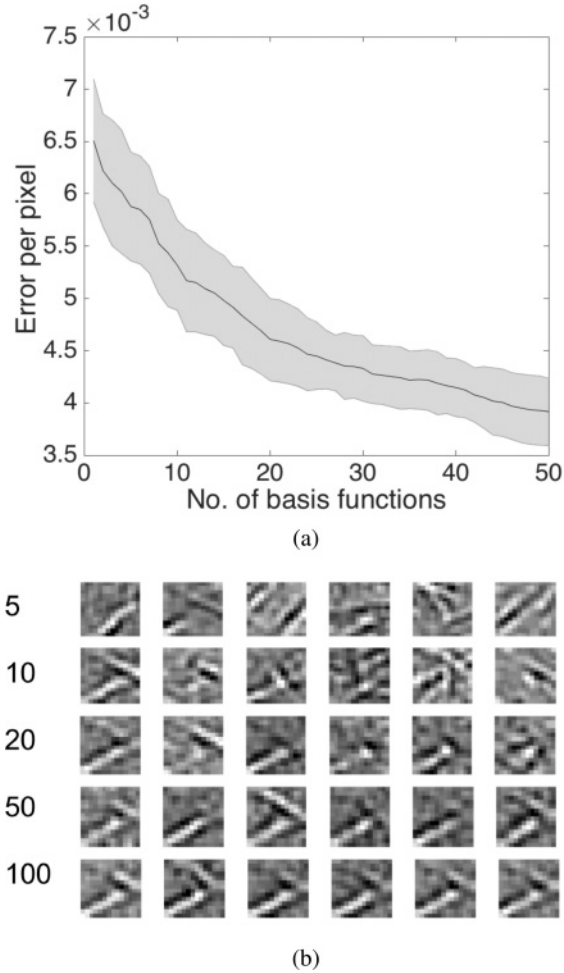

Figure 9: The effect of the number of coding neurons on the accuracy of the parallel probabilistic coding algorithm. (a) Reconstruction mean squared error as a function of the number of coding neurons; the shaded area indicates the standard error of the mean. (b) Sets of six independent reconstructions of the same image patch using increasing numbers of coding cells. We show successive reconstructions using 5, 10, 20, 50, and 100 coding cells using the parallel probabilistic selection algorithm. If 50 to 100 neurons are used, the resultant code becomes accurate.

data. The overlay clearly shows that the spikes are related to the rise in the gamma potential. Each spike codes an analog quantity that is signaled by the delay Δ between the time of the gamma wavelet trough and the spike

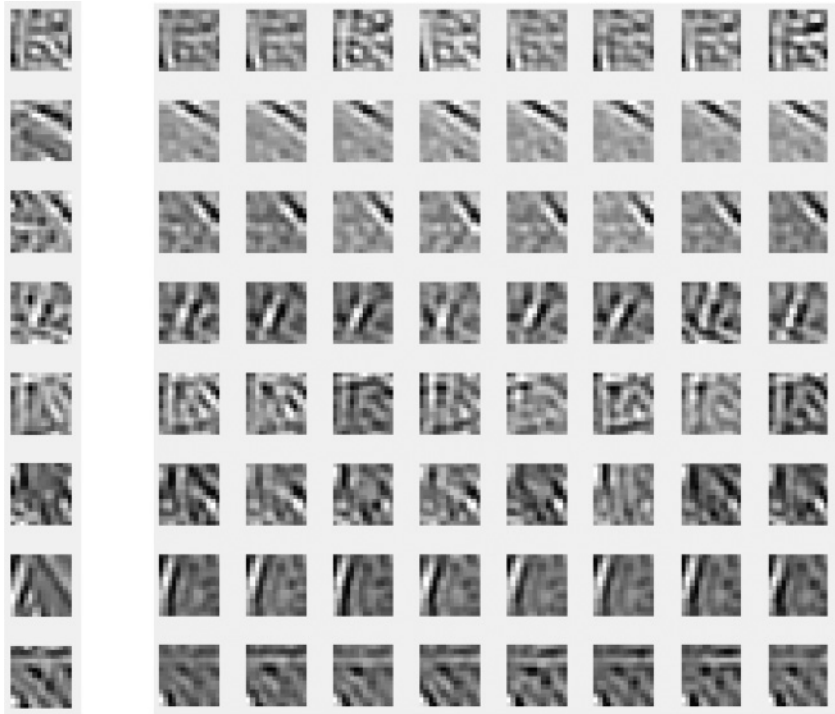

Figure 10: Given an image patch, coding neurons are selected probabilistically. (a) Eight image patches to be coded in the leftmost column simultaneously, representing the eight ongoing neural processes. (b) Each of these is reconstructed eight different times using 50 coding cells. The consequence of overcompleteness and probabilistic selection is that the image patch can be coded using very distinct sets of neurons. The reconstructions are accurate, with a slight tendency to regularize the patterns.

time. For the first three spikes, the delays are 3, 4, and 6 milliseconds. One important observation is that the gamma frequency on a cell's soma can be instantaneously modulated. For identifying the frequency, f is measured from the separation distance of the zero crossings, resulting in frequencies of 43, 46, and 48 Hz.

The putative powerful effect of the small gamma amplitudes can challenge one's intuition, but one needs to keep in mind that the neck of an action potential is exquisitely regulated by the cells in our study to be within

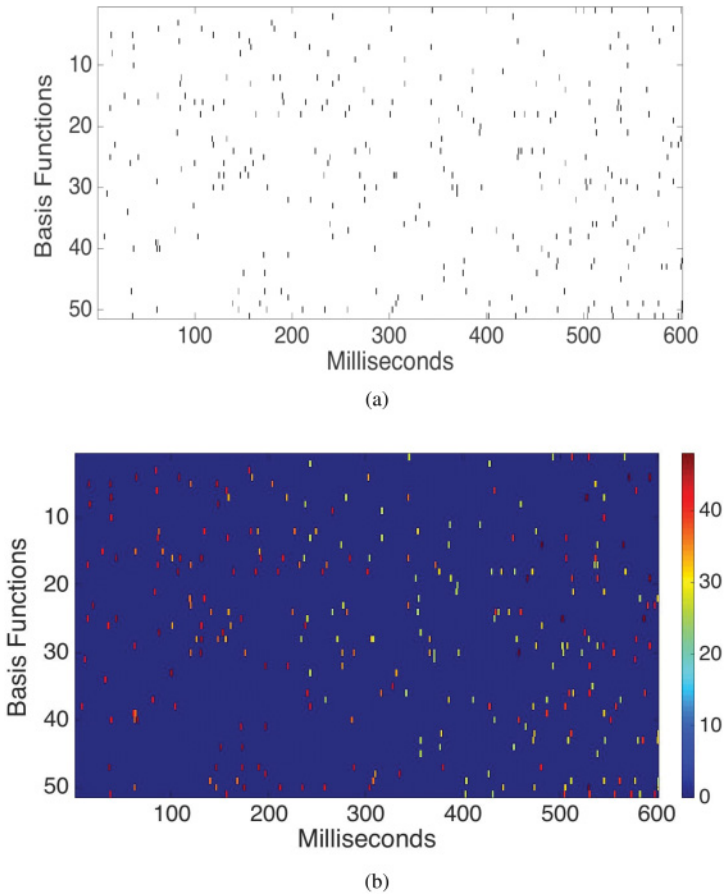

Figure 11: GSM provides a new way of interpreting cell recording data. Spikes are rendered with 2 ms thickness for enhanced visibility. (a) Without the gamma frequency context, spikes can be difficult to interpret. (b) For the same data, we denote each process with a unique color indicating its dedicated gamma frequency. Each spike is then associated with the process that encodes a component of a particular image patch.

a millivolt of negative 37 millivolts, a much smaller excursion compared to the ± 4 millivolts observed in the filtered gamma amplitude signal.

Acknowledgments

The work was supported in part by NIH grant R01 RR09283, and the theoretical concepts benefited with discussions with Ad Aertsen and Martin

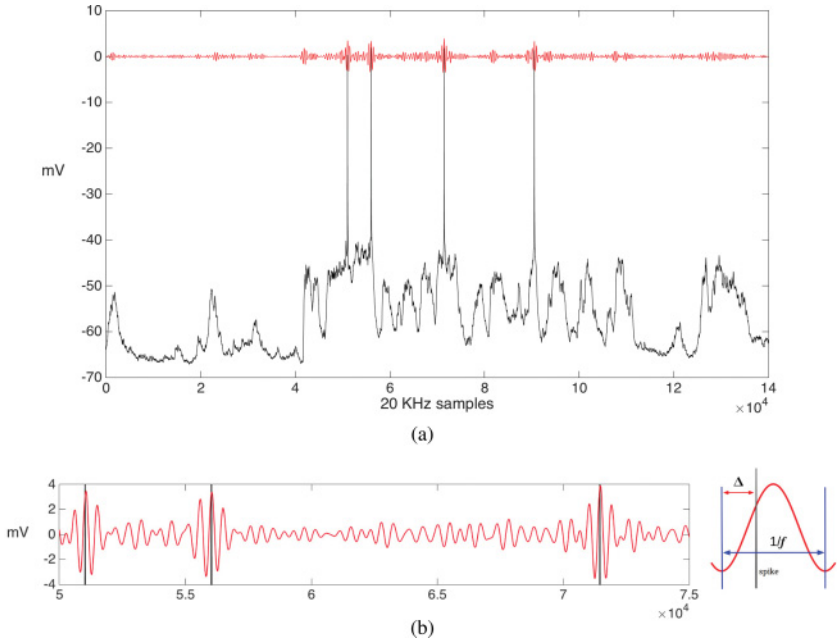

Figure 12: (a) An example patch-clamp data from a single trial. The filtered somatic potential in the gamma frequency range with passband interval [30, 80] Hz shows that spikes are in the first quadrant of the gamma cycle. (b) The first three spikes that are labeled by their associated gamma frequencies. The frequencies and delays are $\{f, \Delta\} = (43 \text{ Hz}, 3 \text{ ms}), (46 \text{ Hz}, 4 \text{ ms}), (48 \text{ Hz}, 6 \text{ ms})$. (Results derived from patch clamp data as a courtesy of the Luc Gentet lab at the Lyon Neuroscience Research Center.)

Vinck. We are also grateful for the membrane potential data provided by Luc Gentet's lab at Lyon Neuroscience Research Center.

References

Abeles, M. (1991). *Corticonics: Neural circuits of the cerebral cortex*. Cambridge: Cambridge University Press.

Atallah, B. V., & Scanziani, M. (2009). Instantaneous modulation of gamma oscillation frequency by balancing excitation with inhibition. *Neuron*, *62*(4), 566–577.

Ballard, D., & Jehee, J. (2011). Dual roles for spike signaling in cortical neural populations. *Frontiers in Computational Neuroscience*, *5*, 22.

Ballard, D., & Zhang, R. (2018). *Cortical spike multiplexing using gamma frequency latencies*. bioRxiv:313320.

Barlow, H. B. (1972). Single units and sensation: A neuron doctrine for perceptual psychology? *Perception*, *1*(4), 371–394.

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron*, *76*(4), 695–711.

Bastos, A. M., Vezoli, J., & Fries, P. (2015). Communication through coherence with inter-areal delays. *Current Opinion in Neurobiology*, *31*, 173–180.

Berry II, M. J., & Meister, M. (1998). Refractoriness and neural precision. In M. I. Jordan, M. J. Kearns, & S. A. Solla (Eds.), *Advances in neural information processing systems*, *10* (pp. 110–116). Cambridge, MA: MIT Press.

Borg-Graham, L. J., Monier, C., & Frégnac, Y. (1998). Visual input evokes transient and strong shunting inhibition in visual cortical neurons. *Nature*, *393*(6683), 369–373.

Boynton, G. M. (2009). A framework for describing the effects of attention on visual responses. *Vision Research*, *49*(10), 1129–1143.

Bringuier, V., Chavane, F., Glaeser, L., & Frégnac, Y. (1999). Horizontal propagation of visual activity in the synaptic integration field of area 17 neurons. *Science*, *283*(5402), 695–699.

Brunel, N., & Wang, X.-J. (2003). What determines the frequency of fast network oscillations with irregular neural discharges? I. Synaptic dynamics and excitation-inhibition balance. *Journal of Neurophysiology*, *90*(1), 415–430.

Brunet, N., Bosman, C. A., Roberts, M., Oostenveld, R., Womelsdorf, T., De Weerd, P., & Fries, P. (2013). Visual cortical gamma-band activity during free viewing of natural images. *Cerebral Cortex*, *25*(4), 918–926.

Burns, S. P., Xing, D., & Shapley, R. M. (2011). Is gamma-band activity in the local field potential of V1 cortex a "clock" or filtered noise? *Journal of Neuroscience*, *31*(26), 9658–9664.

Buzsáki, G., & Wang, X.-J. (2012). Mechanisms of gamma oscillations. *Annual Review of Neuroscience*, *35*, 203–225.

Carrillo-Reid, L., Han, S., Taralova, E., Jebara, T., & Yuste, R. (2017). *Identification and targeting of cortical ensembles*. bioRxiv:226514.

Chalk, M., Herrero, J. L., Gieselmann, M. A., Delicato, L. S., Gotthardt, S., & Thiele, A. (2010). Attention reduces stimulus-driven gamma frequency oscillations and spike field coherence in V1. *Neuron*, *66*(1), 114–125.

Churchland, M. M., Byron, M. Y., Cunningham, J. P., Sugrue, L. P., Cohen, M. R., Corrado, G. S., Shenyoy, K. V. (2010). Stimulus onset quenches neural variability: A widespread cortical phenomenon. *Nature Neuroscience*, *13*(3), 369.

Churchland, M. M., Cunningham, J. P., Kaufman, M. T., Foster, J. D., Nuyujukian, P., Ryu, S. I., & Shenoy, K. V. (2012). Neural population dynamics during reaching. *Nature*, *487*(7405), 51–56.

Cisek, P. (2012). Making decisions through a distributed consensus. *Current Opinion in Neurobiology*, *22*(6) 927–936.

Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience*. Cambridge, MA: MIT Press.

Druckmann, S., & Chklovskii, D. B. (2010). Over-complete representations on recurrent neural networks can support persistent percepts. In J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, & A. Culotta (Eds.), *Advances in neural information processing systems*, *23* (pp. 541–549). Red Hook, NY: Curran.

Evans, C. G., Jing, J., Rosen, S. C., & Cropper, E. C. (2003). Regulation of spike initiation and propagation in anaplysia sensory neuron: Gating-in via central depolarization. *Journal of Neuroscience*, *23*(7), 2920–2931.

Fournier, J., Saleem, A. B., Diamanti, E. M., Wells, M. J., Harris, K. D., & Carandini, M. (2019). *Modulation of visual cortex by hippocampal signals*. bioRxiv:586917.

Fries, P. (2015). Rhythms for cognition: Communication through coherence. *Neuron*, *88*(1), 220–235.

Fries, P., Nikolić, D., & Singer, W. (2007). The gamma cycle. *Trends in Neurosciences*, *30*(7), 309–316.

Fries, P., Reynolds, J. H., Rorie, A. E., & Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. *Science*, *291*(5508), 1560–1563.

Gerstein, G. L., & Mandelbrot, B. (1964). Random walk models for the spike activity of a single neuron. *Biophysical Journal*, *4*(1),41–68.

Gerstner, W., Kreiter, A. K., Markram, H., & Herz, A. V. (1997). Neural codes: Firing rates and beyond. In *Proceedings of the National Academy of Sciences*, *94*(24), 12740–12741.

Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews Neuroscience*, *14*(5), 350.

Gisiger, T., & Boukadoum, M. (2011). Mechanisms gating the flow of information in the cortex: What they might look like and what their uses may be. *Frontiers in Computational Neuroscience*, *5*, 1.

Gollisch, T., & Meister, M. (2008). Rapid neural coding in the retina with relative spike latencies. *Science*, *319*(5866), 1108–1111.

Grinvald, A., Lieke, E. E., Frostig, R. D., & Hildesheim, R. (1994). Cortical point-spread function and long-range lateral interactions revealed by real-time optical imaging of macaque monkey primary visual cortex. *Journal of Neuroscience*, *14*(5), 2545–2568.

Herberholz, J., Antonsen, B. L., & Edwards, D. H. (2002). A lateral excitatory network in the escape circuit of crayfish. *Journal of Neuroscience*, *22*(20), 9078–9085.

Hertz, J. A., Krogh, A. S., & Palmer, R. G. (1991). *Introduction to the theory of neural computation*, vol. *1*. New York: Basic Books.

Hoeffding, W. (1994). Probability inequalities for sums of bounded random variables. In N. I. Fisher, and P. K. Sen (Eds.), *The collected works of Wassily Hoeffding* (pp. 409–426). Berlin: Springer.

Hopfield, J. J. (1995). Pattern recognition computation using action potential timing for stimulus representation. *Nature*, *376*(6535), 33–36.

Hopfield, J. J., & Tank, D. W. (1986). Computing with neural circuits: A model. *Science*, *233*(4764), 625–633.

Igarashi, K. M., Lu, L., Colgin, L. L., Moser, M.-B., & Moser, E. I. (2014). Coordination of entorhinal–hippocampal ensemble activity during associative learning. *Nature*, *510*(7503), 143–147.

Ivanov, A. I., & Calabrese, R. L. (2003). Modulation of spike-mediated synaptic transmission by presynaptic background CA2+ in leech heart interneurons. *Journal of Neuroscience*, *23*(4), 1206–1218.

Jehee, J. F., Rothkopf, C., Beck, J. M., & Ballard, D. H. (2006). Learning receptive fields using predictive feedback. *Journal of Physiology—Paris*, *100*(1–3), 125–132.

Katz, P. S. (2003). Synaptic gating: The potential to open closed doors. *Current Biology*, *13*(14), R554–R556.

Kobak, D., Brendel, W., Constantinidis, C., Feierstein, C. E., Kepecs, A., Mainen, Z. F., & Machens, C. K. (2016). Demixed principal component analysis of neural population data. *Elife*, *5*, e10989.

Koepsell, K., Wang, X., Vaingankar, V., Wei, Y., Wang, Q., Rathbun, D. L., & Sommer, F. T. (2009). Retinal oscillations carry visual information to cortex. *Frontiers in Systems Neuroscience*, *3*, 4.

Landau, A. N., & Fries, P. (2012). Attention samples stimuli rhythmically. *Current Biology*, *22*(11), 1000–1004.

Lee, H., Battle, A., Raina, R., & Ng, A. Y. (2007). Efficient sparse coding algorithms. In B. Schölkopf, J. C. Platt, & T. Hoffman (Eds.), *Advances in neural information processing systems*, *19* (pp. 801–808). Cambridge, MA: MIT Press.

Lennie, P. (2003). The cost of cortical computation. *Current Biology*, *13*(6), 493–497.

Lisman, J. (2005). The theta/gamma discrete phase code occurring during the hippocampal phase precession may be a more general brain coding scheme. *Hippocampus*, *15*(7), 913–922.

Lisman, J. E., & Jensen, O. (2013). The theta-gamma neural code. *Neuron*, *77*(6), 1002–1016.

Lit, A. (1960). The magnitude of the Pulfrich stereophenomenon as a function of target velocity. *Journal of Experimental Psychology*, *59*(3), 165.

Llinas, R. R., Grace, A. A., & Yarom, Y. (1991). In vitro neurons in mammalian cortical layer 4 exhibit intrinsic oscillatory activity in the 10- to 50-Hz frequency range. In *Proceedings of the National Academy of Sciences*, *88*(3), 897–901.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279.

Maunsell, J. H. (2015). Neuronal mechanisms of visual attention. *Annual Review of Vision Science*, *1*, 373–391.

Maunsell, J. H., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in Neurosciences*, *29*(6), 317–322.

McLelland, D., & VanRullen, R. (2016). Theta-gamma coding meets communication-through-coherence: Neuronal oscillatory multiplexing theories reconciled. *PLOS Computational Biology*, *12*(10).

Michalareas, G., Vezoli, J., Van Pelt, S., Schoffelen, J.-M., Kennedy, H., & Fries, P. (2016). Alpha-beta and gamma rhythms subserve feedback and feedforward influences among human visual cortical areas. *Neuron*, *89*(2), 384–397.

Montemurro, M. A., Rasch, M. J., Murayama, Y., Logothetis, N. K., & Panzeri, S. (2008). Phase-of-firing coding of natural visual stimuli in primary visual cortex. *Current Biology*, *18*(5), 375–380.

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science*, *229*(4715), 782–784.

Nadasdy, Z. (2009). Information encoding and reconstruction from the phase of action potentials. *Frontiers in Systems Neuroscience*, *3*, 6.

O'Keefe, J., & Recce, M. L. (1993). Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus*, *3*(3), 317–330.

Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*(6583), 607.

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

Panzeri, S., Macke, J. H., Gross, J., & Kayser, C. (2015). Neural population coding: Combining insights from microscopic and mass signals. *Trends in Cognitive Sciences*, *19*(3), 162–172.

Perrenoud, Q., Pennartz, C. M., & Gentet, L. J. (2016). Membrane potential dynamics of spontaneous and visually evoked gamma activity in V1 of awake mice. *PLOS Biology*, *14*(2), e1002383.

Prechtl, J. C., Bullock, T. H., & Kleinfeld, D. (2000). Direct evidence for local oscillatory current sources and intracortical phase gradients in turtle visual cortex. In *Proceedings of the National Academy of Sciences*, *97*(2), 877–882.

Ray, S., & Maunsell, J. H. (2015). Do gamma oscillations play a role in cerebral cortex? *Trends in Cognitive Sciences*, *19*(2), 78–85.

Roelfsema, P. R., Engel, A. K., König, P., & Singer, W. (1994). Oscillations and synchrony in the visual cortex: Evidence for their functional relevance. In C. Pantev, T. Elbert, & B. Lütkenhöne (Eds.), *Oscillatory event-related brain dynamics* (pp. 99–114). Berlin: Springer.

Roelfsema, P. R., Lamme, V. A., & Spekreijse, H. (2000). The implementation of visual routines. *Vision Research*, *40*(10–12), 1385–1411.

Roskies, A. L. (1999). The binding problem. *Neuron*, *24*(1), 7–9.

Sejnowski, T. J., & Paulsen, O. (2006). Network oscillations: Emerging computational principles. *Journal of Neuroscience*, *26*(6), 1673–1676.

Shadlen, M. N., & Newsome, W. T. (1998). The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding. *Journal of Neuroscience*, *18*(10), 3870–3896.

Singer, W. (2007). Binding by synchrony. *Scholarpedia*, *2*(12), 1657.

Skaggs, W. E., McNaughton, B. L., Wilson, M. A., & Barnes, C. A. (1996). Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus*, *6*(2), 149–172.

Van Kerkoerle, T., Self, M. W., Dagnino, B., Gariel-Mathis, M.-A., Poort, J., Van Der Togt, C., & Roelfsema, P. R. (2014). Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. In *Proceedings of the National Academy of Sciences*, *111*(40), 14332–14341.

Van Kerkoerle, T., Self, M. W., & Roelfsema, P. R. (2017). Layer-specificity in the effects of attention and working memory on activity in primary visual cortex. *Nature Communications*, *8*, 13804.

van Pelt, S., Heil, L., Kwisthout, J., Ondobaka, S., van Rooij, I., & Bekkering, H. (2016). Beta- and gamma-band activity reflect predictive coding in the processing of causal events. *Social Cognitive and Affective Neuroscience*, *11*(6), 973–980.

VanRullen, R., & Thorpe, S. J. (2002). Surfing a spike wave down the ventral stream. *Vision Research*, *42*(23), 2593–2615.

Vinck, M., & Bosman, C. (2016a). More gamma, more predictions: Gamma- synchronization supports the integration of classical receptive field inputs with predictions from the surround. *Frontiers in Systems Neuroscience*, *10*.

Vinck, M., & Bosman, C. A. (2016b). More gamma more predictions: Gamma- synchronization as a key mechanism for efficient integration of classical receptive field inputs with surround predictions. *Frontiers in Systems Neuroscience*, *10*.

Vinck, M., Lima, B., Womelsdorf, T., Oostenveld, R., Singer, W., Neuenschwander, S., & Fries, P. (2010). Gamma-phase shifting in awake monkey visual cortex. *Journal of Neuroscience*, *30*(4), 1250–1257.

Womelsdorf, T., Fries, P., Mitra, P. P., & Desimone, R. (2006). Gamma-band synchronization in visual cortex predicts speed of change detection. *Nature*, *439*(7077), 733.

Womelsdorf, T., Schoffelen, J.-M., Oostenveld, R., Singer, W., Desimone, R., Engel, A. K., & Fries, P. (2007). Modulation of neuronal interactions through neuronal synchronization. *Science*, *316*(5831), 1609–1612.

Yu, C., Horev, G., Rubin, N., Derdikman, D., Haidarliu, S., & Ahissar, E. (2013). Coding of object location in the vibrissal thalamocortical system. *Cerebral Cortex*, *25*(3), 563–577.

Received September 9, 2019; accepted April 19, 2020.