The Bayesian sampling in a canonical recurrent circuit with a diversity of inhibitory interneurons

Eryn Sale[1,2]
eryn.sale@utsouthwestern.edu

Wen-Hao Zhang[1,2*]
wenhao.zhang@utsouthwestern.edu

[1]Lyda Hill Department of Bioinformatics, UT Southwestern Medical Center
[2]O'Donnell Brain Institute, UT Southwestern Medical Center

Abstract

Accumulating evidence suggests stochastic cortical circuits can perform sampling-based Bayesian inference to compute the latent stimulus posterior. Canonical cortical circuits consist of excitatory (E) neurons and types of inhibitory (I) interneurons. Nevertheless, nearly no sampling neural circuit models consider the diversity of interneurons, and thus how interneurons contribute to sampling remains poorly understood. To provide theoretical insight, we build a *nonlinear* canonical circuit model consisting of recurrently connected E neurons and two types of I neurons including Parvalbumin (PV) and Somatostatin (SOM) neurons. The E neurons are modeled as a canonical ring (attractor) model, receiving global inhibition from PV neurons, and locally tuning-dependent inhibition from SOM neurons. We theoretically analyze the *nonlinear* circuit dynamics and *analytically* identify the Bayesian sampling algorithm performed by the circuit dynamics. We found a reduced circuit with only E and PV neurons performs Langevin sampling, and the inclusion of SOM neurons with tuning-dependent inhibition speeds up the sampling via upgrading the Langevin into Hamiltonian sampling. Moreover, the Hamiltonian framework requires SOM neurons to receive no direct feedforward connections, consistent with neuroanatomy. Our work provides overarching connections between nonlinear circuits with various types of interneurons and sampling algorithms, deepening our understanding of circuit implementation of Bayesian inference.

1 Introduction

The brain lives in a world of uncertainty and ambiguity and thus has to infer unobserved world states. The Bayesian inference is a normative framework to implement inference, and extensive studies have suggested that the brain's perception is consistent with the Bayesian inference, forming the concept of Bayesian brain [1], including, e.g., visual processing [2], multisensory integration [3], decision-making [4], sensorimotor learning [5], etc. Studying how neural circuits in the brain realize Bayesian inference has been an active topic in neuroscience [6–8]. Many neural circuit models of Bayesian inference have been developed with distinct representational and algorithmic mechanisms, e.g., parametric-based representation [4, 8–11] and sampling-based representation [12–19].

Despite a large body of neural circuit models of Bayesian inference, there are still gaps between our current Bayesian circuit models and canonical recurrent circuits in the cortex. One obvious distinction is previous Bayesian circuit models haven't considered the rich diversity of neuronal types in the cortex, especially inhibitory interneurons. The canonical cortical microcircuit contains

*Corresponding author.

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

 https://doi.org/10.52202/079017-4301

three major types of inhibitory (I) interneurons [20–24], including Parvalbumin (PV), Somatostatin (SOM), and Vasoactive Intestinal Peptide (VIP) neurons (Fig. 1A). These interneurons have different electrical properties, stimulus-tuning profiles, distinct connectivity, and synaptic modulations with other neurons. For example, PV neurons are weakly tuned to stimulus [20], and *multiplicatively* modulate E neurons in a way called divisive normalization by sending axons to E neurons' cell body [21, 22, 25, 26]. In contrast, SOM neurons are tuned to the stimulus, and module E neurons in an *additive* way via sending axons to the distal dendrites of E neurons [22, 25, 26]. Another gap comes from the *nonlinearity* of cortical circuit dynamics, which impedes the analytical understanding of circuits' Bayesian inference algorithms. Most earlier studies relied on numerical methods to analyze the algorithm in nonlinear circuits (e.g., [15–17, 19]), or considered linear neural dynamics to obtain analytical solutions (e.g., [12, 18, 27, 28]). We still lack a comprehensive understanding about how the nonlinear recurrent circuit dynamics with diversity of interneurons implement Bayesian inference.

To gain insight into Bayesian computation in a nonlinear recurrent circuit with types of interneurons, we build a canonical recurrent circuit model consisting of excitatory (E) and two types of interneurons, including PV and SOM neurons (Fig. 1B), and investigate how the model implements sampling-based Bayesian inference. The E neurons are modeled as a rate-based ring (attractor) model that emerges the tunings over a 1D stimulus such as orientation. The E neurons receive internal Poisson-like variability mimicking stochastic spike generation, which provides a variability source to drive sampling [18]. For simplicity, the PV neurons in the model are not tuned to the stimulus as a limiting case of their weak tuning found in experiments [20], and provides global inhibition to E neurons via divisive normalization to ensure stability [21, 25]. In contrast, the SOM neurons have stimulus tunings and provide locally tuned inhibitory feedback to E neurons in an additive way [20, 29]. The circuit model with the above connectivity successfully reproduces multiplicative and additive modulations on E neurons' tunings from PV and SOM neurons respectively (Fig. 1G-H) [26].

We perform theoretical analysis on the nonlinear recurrent circuit dynamics, and analytically identify the sampling algorithm adopted in the circuit. We find the reduced circuit with only E and PV neurons can implement Langevin sampling in the stimulus feature manifold. The tuning-dependent inhibitory feedback from SOM speeds up the sampling by upgrading the Langevin sampling into Hamiltonian sampling. And the two types of interneurons have different effects on sampling speed. Moreover, we find that Hamiltonian sampling requires SOM neurons not to receive feedforward sensory inputs, consistent with neuroanatomy with few feedforward synapses targeting SOM neurons [22, 24]. The nonlinear circuit model with fixed weights can flexibly sample posteriors with different uncertainties, if located in the linear input-output regime. At last, the circuits can be extended to sample multivariate stimulus posteriors and bimodal posteriors.

2 The recurrent neural circuit with various types of interneurons

The cerebral cortex is a repetition of the canonical neural circuit composed of multiple types of neurons (Fig. 1A), including excitatory (E) neurons and three major types of inhibitory (I) interneurons (PV, SOM, and VIP; classified via biomarkers [30]). To study how sampling-based Bayesian inference is implemented by the canonical neural circuit composed of various types of neurons, we build a recurrent neural circuit model consisting of E neurons and two types of I interneurons of PV and SOM neurons (Fig. 1B). The model doesn't include VIP neurons, which will form our future research (see Discussion). The basic wiring diagram of the proposed circuit model is consistent with the structure of the canonical cortical circuit (Fig. 1A). In the model, the E neurons are selective for a 1D periodic stimulus feature $z \in (-\pi, \pi]$, e.g., the orientation moving direction. Denote θ_j as the preferred stimulus feature of the j-th E neuron, and the preferred stimulus features of all N_E E neurons, $\{\theta_j\}_{j=1}^{N_E}$ uniformly cover the whole range of feature space z (Fig. 1C and F). This setting is the same as the canonical ring network model that has been widely used in modeling cortical circuits (e.g.,[17, 31–36]). Mathematically, in the continuum limit ($\theta_j \to \theta$) corresponding to an infinite number of neurons, the dynamics of the E neurons is [34, 37],

$$\tau \frac{\partial \mathbf{u}_E(\theta,t)}{\partial t} = -\mathbf{u}_E(\theta,t) + \rho \sum_{X=E,F,S} (\mathbf{W}_{EX} * \mathbf{r}_X)(\theta,t) + \sqrt{\tau \mathsf{F}_E[\mathbf{u}_E(\theta,t)]_+}\xi(\theta,t), \quad (1)$$

where $\mathbf{u}_E(\theta,t)$ and $\mathbf{r}_E(\theta,t)$ are the synaptic input and firing rate respectively of the E neuron preferring $z = \theta$. X denotes neuronal types with E, F and S representing E neurons, sensory feedforward inputs, and the SOM neurons respectively. τ is the time constant of synaptic input,

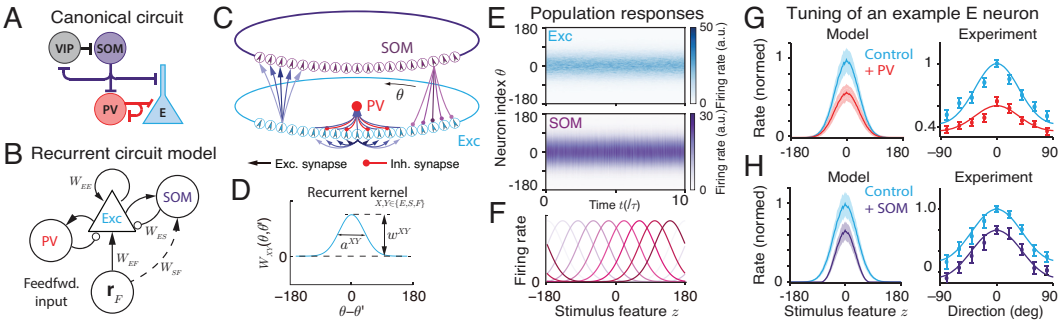

Figure 1: The recurrent circuit model. (A) The canonical cortical circuit consists of E neurons and three types of inhibitory interneurons. (B-C) The recurrent circuit model with two types of interneurons (B) and detailed recurrent circuit structure (C). (D) The Gaussian recurrent connection kernels in the circuit model. (E) An example of population responses of E (top) and SOM (bottom) neurons over time. (F) The tuning curves of E neurons in the model. (G-H) The tuning curve of an example E neuron in control state compared with enhancing PV neurons (G) and SOM neurons (H). The model result is qualitatively similar to the experimental data from [26].

and $\rho = N_E/2\pi$ is the neuronal density covering the space of stimulus feature z. E neurons receive an internal Poisson variability mimicking stochastic spike generation (Eq. 1, last term), and F_E the Fano factor of the injected variability whose value will be adjusted to make the Fano factor of E neurons' activities at the order of 1. $\xi(\theta, t)$ is standard Gaussian white noise, i.e., $\langle \xi(\theta, t)\xi(\theta', t')\rangle = \delta(\theta - \theta')\delta(t - t')$ with $\delta(t - t')$ being the Dirac delta function. The internal Poisson variability provides the variability for the circuit to draw random samples from the posterior [18], which may arise from the E and I balance in the cortex as a complex network effect [38–40].

Recurrent connection kernels. $\mathbf{W}_{YX}(\theta)$ denotes the recurrent connection kernel from neurons with type X to those with type Y, which are modeled as Gaussian functions in the model (Fig. 1D),

$$\mathbf{W}_{YX}(\theta) = w_{YX}\left(\sqrt{2\pi}a_{XY}\right)^{-1}\exp(-\theta^2/2a_{XY}^2), \tag{2}$$

where w_{YX} controls the peak strength of the recurrent weight, and a the connection width across the stimulus feature space. The kernels $\mathbf{W}_{YX}(\theta)$ connecting different neuronal types can have different peak weights w_{YX} ($w_{YX} > 0$ or < 0 regards to E or I synapses respectively). Moreover, different $\mathbf{W}_{YX}(\theta)$ may have different connection widths a_{XY}, although most of them have the same width $a_{XY} = a$ unless noted otherwise (see Supplementary Information (SI.) Sec. 6.1). In Eq. (1), the symbol $*$ denotes the spatial convolution, i.e., $(\mathbf{W} * \mathbf{r})(\theta) = \int \mathbf{W}(\theta - \theta')\mathbf{r}(\theta')d\theta'$, which implies the translation-invariance of the connection weight between neurons in the stimulus feature space.

Sensory feedforward inputs. The recurrent circuit model receives sensory feedforward input $\mathbf{r}_F(\theta, t)$ (Eq. 1) randomly evoked from a stimulus feature z in the world. Given a stimulus feature z, we assume the feedforward inputs $\mathbf{r}_F$ are conditionally independent Poisson spikes with Gaussian tuning (Fig. 2A-B), which has been widely used before (e.g., probabilistic population code [4, 8, 9]).

$$\mathbf{r}_F(\theta|z) \sim \text{Poisson}[\lambda_F(\theta|z)], \quad \lambda_F(\theta|z) = R_F \exp[-(\theta - z)^2/2a^2], \tag{3}$$

where $\lambda_F(\theta|z)$ is the mean firing rate. In simulating our rate-based model, the $\mathbf{r}_F$ is approximated as a continuous Gaussian random variable with multiplicative noise to mimic the Poisson statistics.

2.1 Inhibitory interneurons in the circuit model

PV neurons. The stimulus orientation weakly modulates the PV neurons [20, 26, 41], hence, for simplicity, we consider PV neurons in the model are not tuned for stimulus features and only provide global unstructured inhibition to E neurons to keep stability. Moreover, it was suggested PV neurons provide divisive normalization (DN) to modulate E neurons' responses via shunting inhibition [21, 25, 42]. Hence the proposed circuit model absorbs PV neurons' effects in the divisive normalization of E neurons which has been widely used in circuit models [21, 25, 43, 44],

$$\mathbf{r}_E(\theta, t) = \frac{[\mathbf{u}_E(\theta, t)]_+^2}{1 + \rho w_{EP}\int [\mathbf{u}_E(\theta', t)]_+^2 d\theta'}, \tag{4}$$

https://doi.org/10.52202/079017-4301

where the DN acts as an activation function transferring the instantaneous synaptic input $\mathbf{u}_E(\theta,t)$ of E neurons into their firing rate $\mathbf{r}_E(\theta,t)$. $[x]_+ = \max(x,0)$ denotes negative rectification. The integral $\int[\mathbf{u}_E(\theta,t)]_+^2 d\theta' \equiv \mathbf{r}_{PV}$, reflects PV neurons globally summing all E neurons' activities. w_{EP} is the global inhibition strength characterizing the inhibitory weight from PV to E neurons.

SOM neurons. It is suggested that SOM neurons linearly modulate E neurons' responses, in contrast to multiplicative modulation from PV to E neurons [26]. Therefore the model considers the E neurons receive additive synaptic inputs from SOM neurons ($(\mathbf{W}_{ES} * \mathbf{r}_S)(\theta,t)$, Eq. 1). Furthermore, SOM neurons are tuned to a stimulus feature with a strength comparable to E neurons [20], unlike the weak tunings of PV neurons. Thus, the dynamics of SOM neurons are governed by,

$$\tau\frac{\partial \mathbf{u}_S(x,t)}{\partial t} = -\mathbf{u}_S(x,t) + \rho\sum_{X=E,F}(\mathbf{W}_{SX} * \mathbf{r}_X)(\theta,t); \quad \mathbf{r}_S(\theta,t) = g_S\cdot[\mathbf{u}_S(x,t)]_+, \tag{5}$$

where the g_S (scalar) controls the "gain" of SOM neurons and is set as a fixed value across the study. Two features about SOM neurons' connectivity are worth noting (Eq. 5). First, the model doesn't include recurrent inhibitory connections between SOM neurons, due to few mutual inhibitions between SOM neurons [22, 24]. This simplification will only change the effective gain g_s of SOM neurons without affecting our conclusions of the circuit algorithm. Second, the proposed circuit model allows the existence of the feedforward connections to SOM neurons ($\mathbf{W}_{SF}$ in Eq. 5), even if they are rare in reality [22, 24]. The reason for allowing this rare connection is we want to test whether the Bayesian sampling theory has the power to constrain it.

Overall, the proposed circuit model is consistent with most canonical recurrent circuit models in the field (e.g.,[31–36]). The simplifications considered above reserve the main characteristics of neuronal connectivity and response properties observed in experiments, especially interactions between E neurons and interneurons. For example, it was found enhancing PV neurons will *multiplicatively* modulate E neurons' tuning curves (Fig. 1G, right), whereas SOM neurons modulate E neurons' tuning *additively* (Fig. 1H, right) [26]. Both effects are successfully reproduced in the proposed circuit model (Fig. 1G-H, left; see details in SI. Sec. 6.5).

3 From recurrent circuit dynamics to Bayesian inference

The proposed recurrent circuit dynamics (Eqs. 1 and 5) is supposed to implement Bayesian inference by computing the stimulus posterior based on a received feedforward input,

$$p(z|\mathbf{r}_F) \propto p(\mathbf{r}_F|z)p(z). \tag{6}$$

It regards the stage from external stimulus z to the feedforward input $\mathbf{r}_F$ as the probabilistic generative process (Fig. 2A). Implementing Bayesian inference requires the recurrent circuit to store the generative model. Rather than defining a generative model and deriving its neural circuit implementation as in many previous studies, here we ask the question the other way around: if the proposed recurrent circuit model based on neurophysiology could do Bayesian inference (Eq. 1A), what generative model is stored in the circuit and what stimulus posteriors are computed by the circuit? Furthermore, what is the Bayesian inference algorithm adopted by the circuit dynamics?

Stimulus likelihood. The stochastic feedforward input from the stimulus feature z (Eq. 3) naturally specifies the stimulus likelihood that can be calculated as a Gaussian likelihood (see SI. Sec. 2.1),

$$p(\mathbf{r}_F|z) = \textstyle\prod_\theta \mathrm{Poisson}[\lambda_F(\theta|z)] \propto \mathcal{N}\big(z|\mu_z,\Lambda^{-1}\big), \tag{7}$$

where the Gaussian stimulus likelihood function comes from the the Gaussian profile of feedforward input tuning $\lambda_F(\theta|z)$ (Eq. 3) [9]. The mean μ_z and the precision Λ of the stimulus likelihood can be read out from $\mathbf{r}_F$ via a linear decoder called population vector [45, 46],

$$\mu_z = \textstyle\sum_j \mathbf{r}_F(\theta_j)\theta_j / \sum_j \mathbf{r}_F(\theta_j), \;\; \Lambda = a^{-2}\sum_j \mathbf{r}_F(\theta_j). \tag{8}$$

Geometrically, μ_z is regarded as the location of $\mathbf{r}_F$ in the stimulus feature space, and Λ is proportional to the input spike count (Fig. 2B-C). In this way, a single snapshot of $\mathbf{r}_F$ parametrically conveys the whole stimulus likelihood function $p(\mathbf{r}_F|z)$ [9].

Subjective prior. We suppose the recurrent circuit utilizes its stored *subjective* prior to compute the subjective stimulus posteriors, with a mild assumption that the subjective prior in the circuit matches the *objective* prior in the outside world. Nevertheless, the subjective circuit prior remains unknown at this point. Next, we theoretically analyze the circuit dynamics to identify the stored subjective prior in the circuit and find out the circuit algorithm of Bayesian sampling.

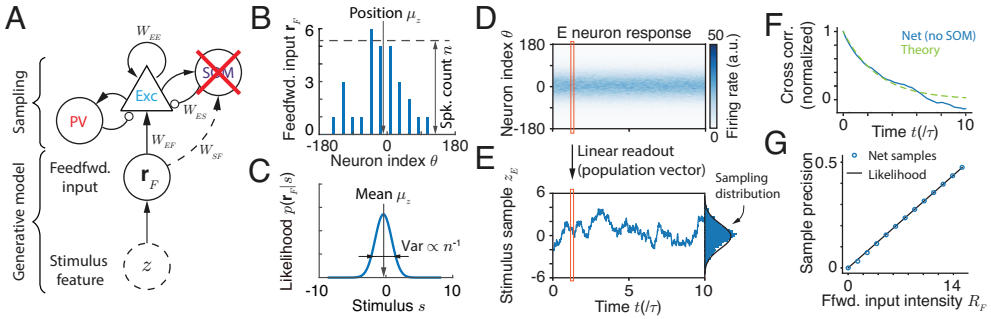

Figure 2: (A) A reduced circuit after blocking SOM neurons. (B-C) A schematic of the feedforward (spiking) input (B) and a linear read out of the stimulus likelihood conveyed by the feedforward input (C). Geometrically, the position and spike count of feedforward input determine the mean and variance of likelihood respectively. (D-E) E neurons' population responses (D) and the location of instantaneous E population response on the y-axis is regarded as stimulus sample z_E generated by the network (E) and can be read out via a linear decoder called population factor. (F) The cross-correlation function of stimulus samples generated by the circuit. (G) The circuit with fixed parameters flexibly samples posteriors with various uncertainties (controlled by feedforward input intensity).

4 The Bayesian sampling in the stochastic circuit dynamics

4.1 Theoretical analysis of the neural dynamics

We theoretically analyze the nonlinear circuit dynamics to investigate how it can implement Bayesian sampling. We perform perturbative analysis of the nonlinear circuit dynamics, identify the low-dimensional stimulus feature manifold (subspace) in the high-dimensional population response space, and eventually study the circuit dynamics on the stimulus feature manifold. Given a feedforward input $\mathbf{r}_F$ (Eq. 3), it can be checked that the synaptic input $\mathbf{u}_X(\theta)$ and the firing rate $\mathbf{r}_X(\theta)$ of neurons X in the equilibrium attractor states are both Gaussian profiles (Fig. 1E-F; see SI. Sec. 3.1) ,

$$\langle \mathbf{u}_X(\theta) \rangle = U_X \exp[-(\theta - \bar{z}_X)^2/4a^2], \ \ \langle \mathbf{r}_X(\theta) \rangle = R_X \exp[-(\theta - \bar{z}_X)^2/2a^2], \ \ (X = E, S) \quad (9)$$

where $\langle \cdot \rangle$ denotes the average over different realizations. U_X and R_X denote the height of the population synaptic input and firing rate respectively, and can be analytically computed (Eq. S37). The position of population activity on the stimulus feature space is $\bar{z}_X = \mu_z$ is the same as the location μ_z of the feedforward inputs (SI. Sec. 3.1). Intuitively, this is because the recurrent circuit is homogeneous along the stimulus feature space, i.e., all neurons are uniformly distributed on the stimulus feature space and the recurrent connections are translational invariance.

With the corruption of sensory noises and the internal Poisson variability, the instantaneous neural responses will deviate from the equilibrium attractor state (Eq. 9). We treat each instantaneous response as a perturbation from its equilibrium state, i.e., $\mathbf{u}_X(\theta, t) = \langle \mathbf{u}_X(\theta) \rangle + \delta \mathbf{u}_X(\theta, t)$, $(X = E, S)$, and the relaxation dynamics of the perturbation $\delta \mathbf{u}_X(\theta, t)$ can be derived [47]. Then performing eigen-analysis of the perturbation dynamics we analytically find out the stimulus feature manifold (subspace), which is specified by its (unnormalized) eigenvector [35, 47, 48],

$$\phi(\theta|z_X) \propto (\theta - z_X) \exp[-(\theta - z_X)^2/4a^2], \ \ (X = E, S). \quad (10)$$

Previous studies have shown the stimulus feature eigenvector has the largest eigenvalue in the perturbation dynamics [47]. Therefore, we project the dynamics of E and SOM neurons (Eqs. 1 and 5) onto their respective stimulus feature eigenvectors, where the projection is computing the inner product between the neuronal responses and the eigenvector, i.e., $\langle \phi(\theta), f(\theta) \rangle = \int \phi(\theta) f(\theta) d\theta$, with $f(\theta)$ representing the left or right handed side of Eqs. (1 and 5). The projection yields the dynamics of the E and SOM neurons on the stimulus feature manifold (see details in SI. Sec. 3.3)

$$\tau_E \dot{z}_E \approx g_{ES}(z_S - z_E) + g_{EF}(\mu_z - z_E) + \sigma_E \sqrt{\tau_E} \xi_t, \quad (11)$$

$$\tau_S \dot{z}_S \approx g_{SE}(z_E - z_S) + g_{SF}(\mu_z - z_S) \quad (12)$$

The approximation comes from ignorance of some negligible nonlinear terms. z_E and z_S denote the instantaneous positions of neural activities at the stimulus feature manifold at time t (Fig. 2E). μ_z is the observed stimulus feature conveyed by the feedforward input (Eq. 8). $\tau_X = \tau U_X$ $(X = E, S)$

https://doi.org/10.52202/079017-4301

is the time constant of the circuit dynamics on the stimulus feature manifold, where U_X is the peak value of population synaptic input (Eq. 9). The coefficients g_{XY} in the above equation denote the coupling strength between neural response positions, where $g_{XY} \propto w_{XY} R_Y$ with R_Y the peak firing rate of the pre-synaptic neural population (Eq. S48). The $\sigma_E^2 = 8a\mathsf{F}_E/(3\sqrt{3\pi})$ is the variance of the internal variability on the stimulus feature manifold coming from the internal Poisson variability (Eq. 1, last term). σ_E is a constant value irrelevant with feedforward inputs and network responses.

4.2 Langevin sampling in the reduced circuit (E and PV neurons)

Since the circuit contains multiple types of neurons, we first analyze a reduced circuit dynamics on the stimulus feature manifold without SOM neuron (Fig. 2A, setting g_{ES} to zero in Eq. 11) and then study how the SOM neurons affect the sampling dynamics. Without SOM neurons (only E and PV neurons), the E neurons' dynamics on the stimulus feature manifold is,

$$\dot{z}_E = \tau_E^{-1} g_{EF}(\mu_z - z_E) + \sigma_E \tau_E^{-1/2} \xi_t, \tag{13}$$

which is a first-order Langevin dynamics. This motivates the possibility that the E and PV neurons can implement Langevin sampling in the stimulus feature manifold. If true, the instantaneous position of E activity, z_E (Fig. 2D), can be regarded as a stimulus feature sample generated by the circuit. In theory, the Langevin sampling of a posterior $p(z|\mathbf{r}_F) \propto p(\mathbf{r}_F|z)p(z)$ corresponds to performing stochastic ascent on the log posterior surface [49, 50],

$$\begin{aligned} \dot{z} &= \tau_z^{-1} \nabla [\ln p(\mathbf{r}_F|z) + \ln p(z)] + (\tau_z/2)^{-1/2} \xi_t, \quad (\nabla \equiv d/dz) \\ &= \tau_z^{-1} [\Lambda(\mu_z - z) + \nabla \ln p(z)] + (\tau_z/2)^{-1/2} \xi_t, \end{aligned} \tag{14}$$

where τ_z is the sampling time constant controlling the sampling speed. The 2nd row is obtained by substituting the Gaussian likelihood (Eq. 7).

Uniform subjective prior. Comparing Eqs. (14) and (13), we can identify a stored *uniform* stimulus prior $p(z)$ in the reduced circuit model stores . This is because the gradient of a uniform prior is $\nabla \ln p(z) = 0$, and then the Langevin sampling dynamics (Eq. 14) reduces to a form similar to the circuit dynamics on the stimulus manifold (Eq. 13). The uniform stimulus prior comes from homogeneous neurons in the circuit, i.e., neurons are uniformly distributed along the stimulus feature space, and the translation-invariant connection profile (Eq. 2). It implies that the circuit needs to break the symmetry of homogeneous neurons to store a non-uniform prior (see Discussion).

Condition for realizing Langevin sampling. Utilizing Langevin dynamics to sample the posterior requires the drift and diffusion coefficients to share the same time constant τ_z (Eq. 14). To satisfy this requirement, the E dynamics on the stimulus feature manifold (Eq. 13) should have $g_{EF}/\sigma_E^2 = \Lambda/2$. With the expressions of g_{EF} and σ_E^2 (Eq. S48), the feedforward connection weight w_{EF} should be,

$$w_{EF} = \frac{\sqrt{\pi}}{a} \sigma_E^2 = \left(\frac{2}{\sqrt{3}}\right)^3 \mathsf{F}_E. \tag{15}$$

Intuitively, larger internal variability F_E increases the sampling variance, and it requires a larger feedforward weigh w_{EF} to compensate for sampling variance increase to match with the posterior variance. To verify our theoretical derivation (Eq. 15), we search whether there is an optimal value of the feedforward weight w_{EF} allowing the circuit without SOM neurons to sample the posterior (likelihood). Indeed, we find once the recurrent circuit model is set with that optimal w_{EF}, the reduced circuit model (consist of E and PV neurons) with all parameters fixed can flexibly sample the likelihood with various uncertainties (Fig. 2G). A characteristic of Langevin sampling is the cross-correlation function of samples (Eq. 14) exponentially decays with time which can be calculated as $\rho(\Delta t) = \exp(-g_{EF}\Delta t/\tau_E)$ (SI. Eqs. S7). To verify whether the reduced circuit performs sampling with the Langevin dynamics as suggested by Eq. (13), we estimate the cross-correlation function of stimulus sample z_E generated by the network, which indeed exhibits an exponential form, and our theoretical calculation $\rho(\Delta t)$ predicts the actual cross-correlation function well (Fig. 2F).

4.3 SOM neurons accelerate Bayesian sampling in E neurons

The SOM neurons augment the dimensionality of the circuit dynamics on the stimulus feature manifold from the first order to the second order (Eqs. 11 and 12). In principle, the second-order

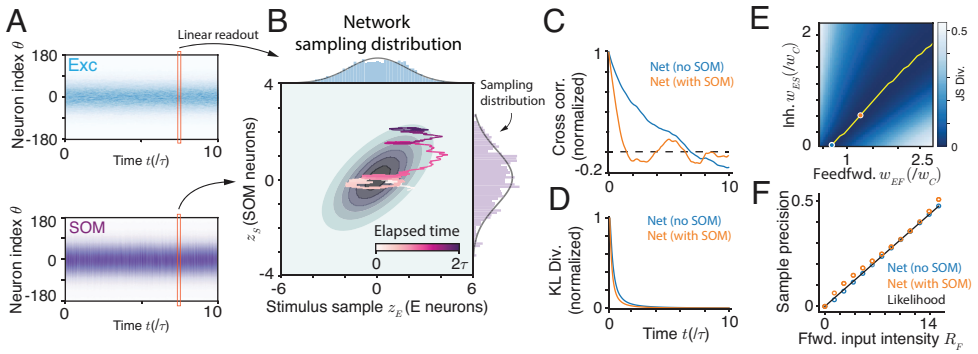

Figure 3: The Bayesian sampling in the full circuit model with PV and SOM neurons. (A) The population responses of E and SOM neurons. (B) The network's sampling distribution read out from E and SOM neurons in a way similar with Fig. 2D-E. The E neuron's position is regarded as stimulus feature sample z_E, while the sample of SOM neurons z_S contribute to the auxiliary variable in Hamiltonian sampling. The distribution of z_E (right marginal) will be used to approximate the posterior. (C-D) The cross-correlation function of stimulus sample z_E (C) and the decaying of KL divergence from posterior and the sampling distribution over time (D). (D) A linear manifold as the combination of feedforward weight w_{EF} and the inhibitory weight from SOM to E neurons w_{ES} allowing the full circuit to sample the posterior correctly. (E) The circuit with fixed weights (Fig. 3D, orange dot) can flexibly sample posteriors with various uncertainties.

Hamiltonian sampling dynamics motivates us to explore whether and how the circuit can implement Hamiltonian-like sampling [50, 51]. The computational benefit from the increased complexity of Hamiltonian sampling compared with Langevin sampling is accelerating the sampling speed by generating less correlated samples over time [50]. To investigate such a possibility, we write one commonly used Hamiltonian sampling dynamics [50, 52, 53],

$$\tau_z \dot{z} = y; \quad \tau_y \dot{y} = -\beta y + \nabla \ln p(z|\mathbf{r}_F) + (2\beta\tau_z)^{1/2}\eta_t. \tag{16}$$

y is the auxiliary variable representing momentum in the Hamiltonian sampling. It can speed up sampling by increasing sampling step size when close to the posterior center, helping the sampling trajectories move toward the other side of the posterior (Fig. 3B; details in SI. Sec. 2.1). β is the friction strength determining how fast the momentum will decay to zero. We next bridge the circuit dynamics (Eqs. 11 and 12) with the Hamiltonian sampling dynamics (Eq. 16).

Intuitively, the sample from SOM neurons z_S (Eq. 12) resemble the auxiliary variable y in the Hamiltonian dynamics (Eq. 16). Nevertheless, there are several gaps between the two: First, in the Hamiltonian sampling, the stimulus sample z is purely driven by the auxiliary variable y (Eq. 16), whereas in the circuit model z_E receives both z_S and μ_z (Eq. 11). Second, Hamiltonian sampling injects variability into the auxiliary dynamics y (Eq. 16), while the variability (from stochastic spike generation) is injected into the dynamics of z_E. To bridge the gap between the circuit model and sampling dynamics, we assume that the circuit is conducting a mixture of Langevin (Eq. 14) and Hamiltonian sampling (Eq. 16), and thus we split the E neurons' sampling dynamics into two parts,

$$\begin{aligned} \tau_E \dot{z}_E &= [g_{ES}(z_S - z_E) + (1-\alpha_L) g_{EF}(\mu_z - z_E)] + [\alpha_L g_{EF}(\mu_z - z_E) + \sigma_E \sqrt{\tau_E}\xi_t], \\ &= \qquad\qquad y_S \qquad\qquad\qquad + [\alpha_L g_{EF}(\mu_z - z_E) + \sigma_E \sqrt{\tau_E}\xi_t], \end{aligned} \tag{17}$$

where $\alpha_L \in [0, 1]$ denotes the proportion of feedforward input contributed by Langevin sampling. We see the y_S resembles the auxiliary variable y in Hamiltonian sampling (Eq. 16), which implies the auxiliary variable in the circuit is a mixture of samples from E neurons z_E and SOM neurons z_S. Given the definition of y_S, the stochastic dynamics of y_S can be derived as a form similar to the one in the Hamiltonian sampling (Eq. 16; details in SI. Sec. 4.1).

$$\dot{y}_S = -\beta_y y_S + \beta_E(\mu_z - z_E) + \beta_S(\mu_z - z_S) + \sigma_y \eta_t. \tag{18}$$

β_y, β_E, β_S and σ_y are functions of the coefficients in Eq. (17) (detailed expressions at Eq. S53). It can be checked that equilibrium distribution of the mixed dynamics is the posterior (details of Fokker-Planck approach in SI. Sec. 2.5).

Conditions for realizing mixed sampling. To utilize the full circuit model with SOM neurons (Eqs. 17-18) to implement Bayesian sampling, the coefficients in the circuit model should satisfy the

https://doi.org/10.52202/079017-4301

relations of coefficients required in the Langevin sampling (Eq. 14) and the Hamiltonian sampling (Eq. 16). First, setting up the Langevin sampling part in the circuit model (Eq. 17, blue) requires $\alpha_L g_{EF}/\sigma_E^2 = \Lambda/2$, which finally leads to,

$$w_{EF} = \sqrt{\pi}\sigma_E^2/(a\alpha_L), \tag{19}$$

whose value is $1/\alpha_L$ times the feedforward weight in the Bayesian sampling circuit without SOM neurons (Eq. 15). Second, realizing the Hamiltonian sampling part in the circuit model (Eq. 17, orange; and Eq. 18) yields three conditions shown below,

$$(a).\ \tau_z = \tau_E; \quad (b).\ \beta_S = 0; \quad (c).\ \frac{\tau_y}{1} = \frac{\beta}{\beta_y} = \frac{\Lambda}{\beta_E} = \frac{(2\beta\tau_z)^{1/2}}{\sigma_y}. \tag{20}$$

The conditions in Eqs. (19) and (20) combined enable the full circuit model with SOM neurons to implement Bayesian sampling. An important insight from Eq. (20b) is that the SOM neurons should not receive feedforward inputs directly ($w_{SF} = 0$ in Eq. 5; Fig. 1A, dashed line removed). This theoretical result is consistent with the anatomy that SOM neurons receive much fewer feedforward synapses than other types of neurons [23, 24]. In addition, substituting the detailed expression of coefficients into Eq. (20), we find there is a low-dimensional manifold in the circuit model's connection weight space for the circuit to sample the posterior $p(z|\mathbf{r}_F)$ (SI. Sec. 4.2),

$$\left(U_E^{-1}R_S\right)\cdot w_{ES} - \left[(1-\alpha_L)U_E^{-1}R_F\right]\cdot w_{EF} = \left[G(\alpha_L)U_S^{-1}R_E\right]\cdot w_{SE}. \tag{21}$$

$G(\alpha_L)$ is a nonlinear function of α_L which specifies the proportion of Langevin sampling in the circuit dynamics, which remains invariant with feedforward input and network activities (SI. Eq. S65). U_X and R_X are the height of the population synaptic input and firing rate of neuronal populations X (Eq. 9). Eq. (21) implies that the sampling in the full circuit model is robust, without the need to fine-tune recurrent weights. To verify the theoretical result (Eq. 21), we fix the weight from E to SOM neurons, w_{SE}, and search whether there is a line manifold in the two-dimensional parameter space of w_{ES} and w_{EF} for the circuit correctly sample the stimulus posterior. Indeed, Fig. 3E numerically confirms the line manifold of weights under which the sampling distribution matches the posterior. Moreover, the introduction of SOM interneruons makes the cross-correlation of sample z_E decay faster (Fig. 3C), suggesting speeding up sampling (Fig. 3D).

Flexible sampling posteriors in the linear regime. Moreover, the circuit model with *fixed weights* should be able to sample the posteriors with different uncertainties. In the circuit model, the posterior uncertainty is determined by the feedforward input rate R_F (Eq. 3), where a larger input rate leading to smaller posterior uncertainty (Eq. 8). We find that when the nonlinear circuit model is located at the *linear regime*, it can flexibly sample posteriors with different uncertainties. To see this effect, we can change the feedforward input rate by multiplying a gain factor g, i.e., $R_F \mapsto gR_F$ that change the likelihood precision $\Lambda \mapsto g\Lambda$ (Eq. 8 and S5). If the circuit is at the linear regime, the peak value of synaptic input, $U_X \mapsto gU_X$, and the population firing rate, $R_X \mapsto gR_X$, are both multiplied with the gain factor g being applied to the feedforward input. And then it can be checked Eq. (21) is still satisfied. This theoretical result is confirmed by numerical simulation (Fig. 3F) which shows the precision of circuit's stimulus samples increases with feedforward rate R_F and aligns well with the likelihood. Here we adjust the recurrent E-to-E weight w_{EE} to set the circuit has an approximately linear response at the range of the feedforward input rate. Fig. S1 shows if the network deviates from the linear regime, the circuit's sampling distribution will deviate from the likelihood.

4.4 The Bayesian sampling performance from interneurons

We further investigate how quantitative measures of sampling, e.g., sampling speed and temporal correlation of samples, will be affected by interneurons such as the inhibitory feedback weight. In principle, both sampling speed and temporal correlation can be revealed by the eigenvalues of the circuit sampling dynamics (Eqs. 11-12). When organizing the circuit sampling dynamics into a matrix form (Eqs. 11-12), $\dot{\mathbf{z}} = -\mathbf{M}\mathbf{z} + \boldsymbol{\mu}$, with $\mathbf{z} = (z_E, z_S)^\top$ and $\boldsymbol{\mu}$ lumping terms exclusive z_E or z_S in Eqs. (11 - 12), the eigenvalues of the circuit sampling dynamics are (SI. Eq. S69),

$$\begin{aligned} &\lambda_\pm = \mathrm{tr}(\mathbf{M}) \pm \sqrt{\mathrm{tr}(\mathbf{M})^2 - 4\det(\mathbf{M})} \triangleq \mathrm{tr}(\mathbf{M}) \pm \sqrt{\Delta}, \\ &\text{where}\ \ \mathrm{tr}(\mathbf{M}) = \tau_E^{-1}(g_{ES} + g_{EF}) + \tau_S^{-1}g_{SE},\ \ \det(\mathbf{M}) = \tau_E^{-1}\tau_S^{-1}g_{EF}g_{SE}. \end{aligned} \tag{22}$$

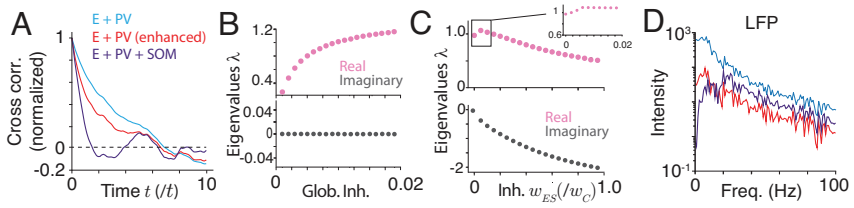

Figure 4: Interneurons' effects on sampling. (A) The cross-correlation of samples for perturbing PV and SOM. (B-C) The smallest eigenvalue of sampling dynamics when changing PV (A) and SOM (B) inhibitions. (A): obtained from the circuit without SOM (Fig. 2A). (D) The local field potential (LFP), defined by the sum of synaptic inputs of E and SOM neurons (color: same as A).

The sampling speed is limited by the smallest real part of eigenvalues, i.e., $\mathrm{Re}(\lambda_-)$ (Eq. 22), in that the KL divergence from the distribution of samples up to time t to the stationary distribution (Fig. 3D) decays exponentially as $\exp[-\mathrm{Re}(\lambda_-)t]$ [54].

PV neurons. PV's inhibition weight w_{EP} modulates the eigenvalues by decreasing the common factor $\tau_E = \tau U_E(w_{EP})$ (Eq. 22), where $U_E(w_{EP})$, the peak value of E population synaptic input (Eq. 9), decreases with w_{EP}. Hence, stronger PV inhibition increases the slowest eigenvalue λ_- (Fig. 4A) and leads to faster sampling, exhibited by the faster decay of the temporal correlation between samples (Fig. 4A, blue and red). Moreover, the multiplicative modulation from PV neurons will not induce the imaginary part of eigenvalues (Fig. 4B, bottom), i.e., no oscillation between samples.

SOM neurons. The SOM inhibitory weight w_{ES} will have non-monotonic effects on sampling speed measured by the real part of the slowest eigenvalue. There is a value of w_{ES} to maximize the sampling speed ($\mathrm{Re}(\lambda_-)$ (Fig. 4C, top). Moreover, SOM's inhibition will induce temporal oscillations between samples, i.e., emerging imaginary part of eigenvalue λ_- (Fig. 4C, bottom). This oscillation is confirmed by the cross-correlation of samples (Fig. 4A, purple). Moreover, to mimic neural experimental data analysis, the oscillation induced by SOM neurons can be revealed by the power spectrum analysis of the local field potential (LFP) (SI. Eq. S90).

5 Sampling complex posteriors in canonical recurrent circuits

High-dimensional stimulus posteriors. As a proof of concept example, we consider sampling bivariate stimulus posteriors by coupled circuits (Fig. 5A) with each circuit the same as Fig. 1B. And only E neurons across circuits are coupled. Each circuit m ($m = 1, 2$) receives a feedforward input generated by a latent stimulus feature z_m, and will sample z_m. Hence, the number of coupled circuits equals to the stimulus feature dimension. We found the coupled circuits store an associative (subjective) prior, i.e., $p(z_1, z_2) \propto \exp[-\Lambda_s(z_1 - z_2)^2/2]$ (Fig. 5C), with the prior precision Λ_s increasing with inter-circuit coupling weights (Fig. 5D; Eq. S85). Math analysis is presented in SI. Sec. 5. Concatenating the samples generated by E neurons in two circuits (with the same readout as Fig. 3A-B), we can obtain the 2D posterior sampled from the coupled circuit (Fig. 5B). Similarly, the SOM interneurons speed up sampling, reflected by the sampling trajectories with SOM transverse over a wider posterior region than the one without SOM in the same period (Fig. 5B).

Bimodal stimulus distributions. We also use the circuit (Fig. 1C) to sample uni-variate bimodal posteriors (Fig. S2), in response to superpositions of two feedforward inputs with each generated by a latent stimulus. The circuit can sample a bimodal distribution, where samples jump between two modes alternatively, due to the bi-stability in the circuit dynamics (Fig. S2C and E). In contrast, without SOM neurons (Fig. 2A), the circuit can only sample a uni-modal distribution, as a uni-modal approximation of the bimodal one (Fig. S2F). Due to the space limit, we haven't comprehensively linked the circuit' bimodal sampling distribution with posteriors, which will form our future work.

6 Conclusion and Discussion

The present study investigates how canonical recurrent circuits with diverse inhibitory interneurons implement Bayesian sampling. The nonlinear circuit model consists of E neurons, and two types of interneurons neurons (PV and SOM neurons). PV and SOM neurons have distinct tuning properties and modulations on E neurons: PV neurons have weak stimulus tuning and multiplicatively modulate E neurons, while SOM neurons have stimulus tuning and send additive inputs to E neurons. Through

https://doi.org/10.52202/079017-4301

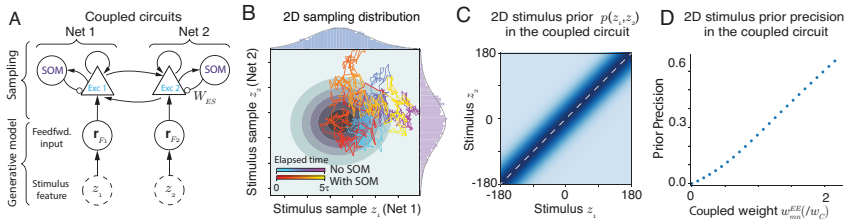

Figure 5: Sampling 2D posteriors in coupled circuits. (A) Each circuit (net) is the same as the one in Fig. 1B (PV neurons are not shown). (B) The 2D sampling distribution generated by the coupled circuit shown in (A). (C-D) The coupled circuits store an associative (subjective) prior of two stimuli (C). The prior correlation increases with inter-circuit coupling weights (D).

theoretical analysis and numerical simulations, we find the reduced circuit with only E and PV neurons implements Langevin sampling in the stimulus manifold (subspace) to compute the stimulus posterior. The SOM neurons accelerate the circuit sampling speed by upgrading the Langevin sampling into Hamiltonian sampling. We further find Hamiltonian sampling requires the SOM neurons to receive no feedforward connections, consistent with neuroanatomy. We also investigate how inhibitory strength from two types of interneurons affects the sampling speed. Our work is one of the earliest studies investigating the Bayesian sampling algorithm in canonical nonlinear recurrent circuits with diverse interneurons, and provides new insight into interneurons in Bayesian computation.

Comparison with other work. The computational mechanism of accelerated sampling by SOM neurons is similar to previous studies considering structured inhibitory feedback to E neurons (e.g., [14, 17, 53, 55]), while some notable differences exist. Earlier studies considered the sampling in neural response space where the posterior dimension is the same as the number of neurons [14, 17], whereas the current circuit samples in the stimulus feature manifold (subspace) in the neural response space. On the other hand, although other previous studies considered sampling in the stimulus feature manifold similar to the current model [53, 55], their structured inhibitory feedback comes from the biophysical mechanism within single neurons, e.g., spike frequency adaptation [53], and potassium channels [55]. From a neurobiology perspective, SOM neurons can be modulated by VIP neurons (Fig. 1A), whereas intracellular mechanisms are hardly modulated [53, 55], suggesting accelerated sampling from SOM neurons might be more flexible than single neuron mechanism in reality. Moreover, the proposed circuit model is similar to a recent one [18] in terms of utilizing internal Possion variability to draw samples. However, the current circuit model is nonlinear while the other study considered a linear circuit model [18]. Lastly, the coupled circuit sampling of bivariate posteriors was also considered in [34], but which didn't figure out the circuit's sampling algorithms.

Limitations and extensions of the model. The proposed circuit model doesn't include VIP neurons that exclusively target SOM neurons (Fig. 1A). Our future work will incorporate them and study their effects on circuit sampling. Based on our current conclusion, it is likely that VIP neurons act as a "knob" to modulate the sampling speed, depending on task needs, by changing the activation level ("gain") of SOM neurons. Moreover, we find the proposed canonical circuit stores a uniform stimulus prior, due to the homogeneity of neurons distributing on the stimulus manifold. The homogeneous neuron simplification has been widely used in neural coding and continuous attractor networks (e.g., [31, 34, 35, 56]), which simplifies the math analysis without altering results substantially. Nevertheless, the circuit has to break the neuronal homogeneity to store a non-uniform prior [57], and certainly, cortical neurons are heterogeneous. The neuronal heterogeneity can be realized by manipulating the translation-invariant recurrent connection matrix (Eq. 2), e.g., introducing randomness (zero mean with certain variance) on recurrent weight which has also been widely used in (chaotic) Excitation and Inhibition (E/I) balanced networks ([38–40, 58]). A potential function of heterogeneity from random recurrent weights is that this puts the spiking networks into the chaotic regime where the network internally generates Poisson variability, which is statistically equivalent to the injected multiplicative variability in our rate-based network (Eq. 1, last term). Moreover, the proposed circuit model only infers a static stimulus, and we will extend to infer a dynamic stimulus described by a hidden Markov model in the future. All of these form our future research.

Acknowledgments

W.H.Z. is supported by the UT Southwestern Endowed Scholars program. The authors thank Chengcheng Huang for fruitful comments.

References

[1] David C Knill and Alexandre Pouget. The bayesian brain: the role of uncertainty in neural coding and computation. *TRENDS in Neurosciences*, 27(12):712–719, 2004.

[2] Alan Yuille and Daniel Kersten. Vision as bayesian inference: analysis by synthesis? *Trends in cognitive sciences*, 10(7):301–308, 2006.

[3] Marc O Ernst and Martin S Banks. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870):429–433, 2002.

[4] Jeffrey M Beck, Wei Ji Ma, Roozbeh Kiani, Tim Hanks, Anne K Churchland, Jamie Roitman, Michael N Shadlen, Peter E Latham, and Alexandre Pouget. Probabilistic population codes for bayesian decision making. *Neuron*, 60(6):1142–1152, 2008.

[5] Konrad P Körding and Daniel M Wolpert. Bayesian integration in sensorimotor learning. *Nature*, 427(6971):244–247, 2004.

[6] József Fiser, Pietro Berkes, Gergő Orbán, and Máté Lengyel. Statistically optimal perception and learning: from behavior to neural representations. *Trends in cognitive sciences*, 14(3):119–130, 2010.

[7] Daniel Kersten, Pascal Mamassian, and Alan Yuille. Object perception as bayesian inference. *Annu. Rev. Psychol.*, 55:271–304, 2004.

[8] Alexandre Pouget, Jeffrey M Beck, Wei Ji Ma, and Peter E Latham. Probabilistic brains: knowns and unknowns. *Nature neuroscience*, 16(9):1170, 2013.

[9] Wei Ji Ma, Jeffrey M Beck, Peter E Latham, and Alexandre Pouget. Bayesian inference with probabilistic population codes. *Nature Neuroscience*, 9(11):1432–1438, 2006.

[10] Rajkumar Vasudeva Raju and Zachary Pitkow. Inference by reparameterization in neural population codes. In *Advances in Neural Information Processing Systems*, pages 2029–2037, 2016.

[11] Ingmar Kanitscheider, Ruben Coen-Cagli, and Alexandre Pouget. Origin of information-limiting noise correlations. *Proceedings of the National Academy of Sciences*, 112(50):E6973–E6982, 2015.

[12] Patrik O Hoyer and Aapo Hyvärinen. Interpreting neural response variability as monte carlo sampling of the posterior. In *Advances in neural information processing systems*, pages 293–300, 2003.

[13] Lars Buesing, Johannes Bill, Bernhard Nessler, and Wolfgang Maass. Neural dynamics as sampling: a model for stochastic computation in recurrent networks of spiking neurons. *PLoS computational biology*, 7(11):e1002211, 2011.

[14] Laurence Aitchison and Máté Lengyel. The hamiltonian brain: efficient probabilistic inference with excitatory-inhibitory neural circuit dynamics. *PLoS computational biology*, 12(12), 2016.

[15] Ralf M Haefner, Pietro Berkes, and József Fiser. Perceptual decision-making as probabilistic inference by neural sampling. *Neuron*, 90(3):649–660, 2016.

[16] Gergő Orbán, Pietro Berkes, József Fiser, and Máté Lengyel. Neural variability and sampling-based probabilistic representations in the visual cortex. *Neuron*, 92(2):530–543, 2016.

[17] Rodrigo Echeveste, Laurence Aitchison, Guillaume Hennequin, and Máté Lengyel. Cortical-like dynamics in recurrent circuits optimized for sampling-based probabilistic inference. *Nature Neuroscience*, 2020.

[18] Wen-Hao Zhang, Si Wu, Krešimir Josić, and Brent Doiron. Sampling-based bayesian inference in recurrent circuits of stochastic spiking neurons. *Nature Communications*, 14(1):7074, 2023.

 https://doi.org/10.52202/079017-4301

[19] Yu Terada and Taro Toyoizumi. Chaotic neural dynamics facilitate probabilistic computations through sampling. *Proceedings of the National Academy of Sciences*, 121(18):e2312992121, 2024.

[20] Hillel Adesnik, William Bruns, Hiroki Taniguchi, Z Josh Huang, and Massimo Scanziani. A neural circuit for spatial summation in visual cortex. *Nature*, 490(7419):226–231, 2012.

[21] Cristopher M Niell. Cell types, circuits, and receptive fields in the mouse visual cortex. *Annual review of neuroscience*, 38:413–431, 2015.

[22] Gord Fishell and Adam Kepecs. Interneuron types as attractors and controllers. *Annual review of neuroscience*, 43:1–30, 2020.

[23] Cristopher M Niell and Massimo Scanziani. How cortical circuits implement cortical computations: mouse visual cortex as a model. *Annual Review of Neuroscience*, 44:517–546, 2021.

[24] Luke Campagnola, Stephanie C Seeman, Thomas Chartrand, Lisa Kim, Alex Hoggarth, Clare Gamlin, Shinya Ito, Jessica Trinh, Pasha Davoudian, Cristina Radaelli, et al. Local connectivity and synaptic dynamics in mouse and human neocortex. *Science*, 375(6585):eabj5861, 2022.

[25] Matteo Carandini and David J Heeger. Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, 13(1):51–62, 2012.

[26] Nathan R. Wilson, Caroline A. Runyan, Forea L. Wang, and Mriganka Sur. Division and subtraction by distinct cortical inhibitory networks in vivo. *Nature*, 488(7411):343–348, August 2012.

[27] Rajesh P Rao. Hierarchical bayesian inference in networks of spiking neurons. *Advances in neural information processing systems*, 17, 2004.

[28] Cristina Savin and Sophie Denève. Spatio-temporal representations of uncertainty in spiking neural networks. *Advances in neural information processing systems*, 27, 2014.

[29] Hiroyuki K Kato, Samuel K Asinof, and Jeffry S Isaacson. Network-level control of frequency tuning in auditory cortex. *Neuron*, 95(2):412–423, 2017.

[30] Adam Kepecs and Gordon Fishell. Interneuron cell types are fit to function. *Nature*, 505(7483):318–326, January 2014.

[31] R Ben-Yishai, R Lev Bar-Or, and H Sompolinsky. Theory of orientation tuning in visual cortex. *Proceedings of the National Academy of Sciences*, 92(9):3844–3848, 1995.

[32] X-J Wang, Jesper Tegnér, C Constantinidis, and Patricia S Goldman-Rakic. Division of labor among distinct subtypes of inhibitory neurons in a cortical microcircuit of working memory. *Proceedings of the National Academy of Sciences*, 101(5):1368–1373, 2004.

[33] Daniel B Rubin, Stephen D Van Hooser, and Kenneth D Miller. The stabilized supralinear network: a unifying circuit motif underlying multi-input integration in sensory cortex. *Neuron*, 85(2):402–417, 2015.

[34] Wen-Hao Zhang, Aihua Chen, Malte J Rasch, and Si Wu. Decentralized multisensory information integration in neural systems. *The Journal of Neuroscience*, 36(2):532–547, 2016.

[35] Si Wu, KY Michael Wong, CC Alan Fung, Yuanyuan Mi, and Wenhao Zhang. Continuous attractor neural networks: candidate of a canonical model for neural information representation. *F1000Research*, 5, 2016.

[36] Guillaume Hennequin, Yashar Ahmadian, Daniel B Rubin, Máté Lengyel, and Kenneth D Miller. The dynamical regime of sensory cortex: stable dynamics around a single stimulus-tuned attractor account for patterns of noise variability. *Neuron*, 98(4):846–860, 2018.

[37] Si Wu, Kosuke Hamaguchi, and Shun-ichi Amari. Dynamics and computation of continuous attractors. *Neural Computation*, 20(4):994–1025, 2008.

[38] Carl Van Vreeswijk and Haim Sompolinsky. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, 274(5293):1724–1726, 1996.

[39] C van Vreeswijk and Haim Sompolinsky. Chaotic balanced state in a model of cortical circuits. *Neural computation*, 10(6):1321–1371, 1998.

[40] Robert Rosenbaum, Matthew A Smith, Adam Kohn, Jonathan E Rubin, and Brent Doiron. The spatial structure of correlated neuronal variability. *Nature neuroscience*, 20(1):107–114, 2017.

[41] Alexandra K. Moore and Michael Wehr. Parvalbumin-Expressing Inhibitory Interneurons in Auditory Cortex Are Well-Tuned for Frequency. *Journal of Neuroscience*, 33(34):13713–13723, August 2013.

[42] James E Cooke, Martin C Kahn, Edward O Mann, Andrew J King, Jan WH Schnupp, and Ben DB Willmore. Contrast gain control occurs independently of both parvalbumin-positive interneuron activity and shunting inhibition in auditory cortex. *Journal of Neurophysiology*, 123(4):1536–1551, 2020.

[43] Tomokazu Ohshiro, Dora E Angelaki, and Gregory C DeAngelis. A normalization model of multisensory integration. *Nature Neuroscience*, 14(6):775–782, 2011.

[44] Sophie Deneve, Peter E Latham, and Alexandre Pouget. Reading population codes: a neural implementation of ideal observers. *Nature Neuroscience*, 2(8):740–745, 1999.

[45] Apostolos P Georgopoulos, Andrew B Schwartz, and Ronald E Kettner. Neuronal population coding of movement direction. *Science*, 233(4771):1416–1419, 1986.

[46] Peter Dayan and Laurence F Abbott. *Theoretical neuroscience*, volume 806. Cambridge, MA: MIT Press, 2001.

[47] C. C Alan Fung, K. Y. Michael Wong, and Si Wu. A moving bump in a continuous manifold: A comprehensive study of the tracking dynamics of continuous attractor neural networks. *Neural Computation*, 22(3):752–792, 2010.

[48] Wen-Hao Zhang and Si Wu. Neural information processing with feedback modulations. *Neural Computation*, 24(7):1695–1721, 2012.

[49] Max Welling and Yee W Teh. Bayesian learning via stochastic gradient langevin dynamics. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pages 681–688, 2011.

[50] Radford M Neal et al. Mcmc using hamiltonian dynamics. *Handbook of markov chain monte carlo*, 2(11):2, 2011.

[51] Aihua Chen, Gregory C DeAngelis, and Dora E Angelaki. Convergence of vestibular and visual self-motion signals in an area of the posterior sylvian fissure. *Journal of Neuroscience*, 31(32):11617–11627, 2011.

[52] Tianqi Chen, Emily Fox, and Carlos Guestrin. Stochastic gradient hamiltonian monte carlo. In *International conference on machine learning*, pages 1683–1691. PMLR, 2014.

[53] Xingsi Dong, Zilong Ji, Tianhao Chu, Tiejun Huang, Wenhao Zhang, and Si Wu. Adaptation accelerating sampling-based bayesian inference in attractor neural networks. *Advances in Neural Information Processing Systems*, 35:21534–21547, 2022.

[54] Sever Silvestru Dragomir. Some gronwall type inequalities and applications. *Science Direct Working Paper*, (S1574-0358):04, 2003.

[55] Yang Qi and Pulin Gong. Fractional neural sampling as a theory of spatiotemporal probabilistic computations in neural circuits. *Nature communications*, 13(1):1–19, 2022.

[56] Mikail Khona and Ila R. Fiete. Attractor and integrator networks in the brain. *Nature Reviews Neuroscience*, 23(12):744–766, December 2022.

 https://doi.org/10.52202/079017-4301

[57] Deep Ganguli and Eero P Simoncelli. Implicit encoding of prior probabilities in optimal neural populations. *Advances in neural information processing systems*, 2010:658, 2010.

[58] Ashok Litwin-Kumar and Brent Doiron. Slow dynamics and high variability in balanced cortical networks with clustered connections. *Nature neuroscience*, 15(11):1498, 2012.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: We perform mathematically rigorous theoretical analysis on the nonlinear circuit dynamics (Eqs. 9 - 21) and verify our theoretical results via numerical simulations (Fig. 2 - 4).

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: It is located at the last part of Conclusion and Discussion.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

https://doi.org/10.52202/079017-4301

Answer: [Yes]

Justification: We clearly write the underlying assumption in constructing the circuit model (Section 2) and its Bayesian inference (Section 3), and present detailed math calculations in the Supplementary Information.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: The main text clearly state the assumptions in constructing the neural circuit model, and the Supplementary Information contains all circuit parameters and details of our numerical simulation.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: Codes are uploaded and shared for reproduction and further experiments.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [NA]

   Justification: The present paper doesn't study the learning problem.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: We define the error bars in Fig. 1G-H, and the define statistics of simulating the proposed circuit model in text corresponding to Fig. 2-4, including the mean, correlation, empirical distribution, and the KL divergence. The Supplementary Information has details of how we obtain these statistics.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

 https://doi.org/10.52202/079017-4301

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: It can be found at the SI. Sec. **??**.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We confirm our research is abide by the Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [No]

    Justification: This is a theoretical study of basic neuroscience research and will not have direct societal impacts.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.

- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: We propose a theoretical model of the neural circuit in the brain and it will not impose any risk.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: We adapt two figures from a published experimental paper (Fig. 1G-H) and we clearly state where the figures are from.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.

 https://doi.org/10.52202/079017-4301

- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: This is a theoretical neuroscience study where the math formulation of the neural circuit model and its theoretical analysis presented in the paper are our main results.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: This is a theoretical neuroscience study.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: This is a theoretical neuroscience study.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.

- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

https://doi.org/10.52202/079017-4301