bioRxiv preprint doi: https://doi.org/10.1101/2025.07.03.663069; this version posted July 4, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

# Natural gradient Bayesian sampling: an innate algorithm emerges in canonical cortical circuits

**Zimei Chen**[1,2], **Yi Ren**[1,2], **Eryn Sale**[1,2], **Ying Nian Wu**[3], **Wen-Hao Zhang**[1,2*]

[1]Lyda Hill Department of Bioinformatics, UT Southwestern Medical Center.
[2]O'Donnell Brain Institute, UT Southwestern Medical Center.
[3]Department of Statistics, University of California, Los Angeles.

## Abstract

Accumulating evidence suggests the canonical cortical circuit, consisting of excitatory (E) and diverse classes of inhibitory (I) interneurons, implement sampling-based Bayesian inference to compute stimulus posteriors. However, most of the identified sampling algorithms in the circuit are still simpler than the nonlinear circuit dynamics. Through comprehensive theoretical analyses of the nonlinear circuit dynamics, we discover the canonical circuit dynamics innately implements natural gradient Bayesian sampling, which is an advanced sampling algorithm that adaptively adjusts the sampling step size based on the local geometry of stimulus posteriors measured by Fisher information. Specifically, the nonlinear circuit dynamics can implement natural gradient Langevin and Hamiltonian sampling of uni- and multi-variate stimulus posteriors, and these algorithms can be switched by interneurons. We also find that the non-equilibrium circuit dynamics when transitioning from the resting to evoked state can further accelerate natural gradient sampling, and analytically identify the neural circuit's annealing strategy. Remarkably, we identify the approximated computational strategies employed in the circuit dynamics, which even resemble the ones widely used in machine learning. Our work provides an overarching connection between canonical circuit dynamics and advanced sampling algorithms, deepening our understanding of the circuit algorithms of Bayesian sampling.

## 1 Introduction

The brain lives in a world of uncertainty and ambiguity, necessitating the inference of unobserved world states. The Bayesian inference provides a normative framework for this process, and extensive studies have suggested that neural computations across domains aligns with Bayesian principles, giving rise to the concept of the "Bayesian brain" [1]. These include visual processing [2], multi-sensory integration [3], decision-making [4], sensorimotor learning [5], reinforcement learning (cite), etc. Recent studies suggested the canonical cortical circuit may naturally implement sampling-based Bayesian inference to compute the posterior [6–15], in that the large cortical response variability is consistent with the stochastic nature of sampling algorithms.

The canonical cortical circuit – the fundamental computational unit repeated throughout the cerebral cortex – consists of excitatory neurons (E) and various classes of inhibitory interneurons (I) (Fig. 1 A) [16–20], including neurons of parvalbumin (PV), somatostatin (SOM), and vasoactive intestinal peptide (VIP). Different interneuron classes have different intrinsic electrical properties and form specific connectivity patterns (Fig. 1B). A very recent study has established connections between

*Corresponding author. wenhao.zhang@utsouthwestern.edu

Preprint. Under review.

bioRxiv preprint doi: https://doi.org/10.1101/2025.07.03.663069; this version posted July 4, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

certain neuronal circuit motifs and sampling-based inference [15]. Specifically, the reduced circuit of only E and PV neurons can implement Langevin sampling in the stimulus feature manifold to compute the posterior. And incorporating SOM into the circuit introduces oscillations that accelerate sampling by upgrading Langevin sampling into more efficient Hamiltonian sampling.

Nevertheless, a significant gap remains between identified Bayesian sampling algorithms and the complex, nonlinear canonical circuit dynamics. A notable distinction is that canonical circuit dynamics is inherently non-linear and substantially more complex than the linear dynamics of Langevin and Hamiltonian samplings identified in previous circuit models and used in machine learning (ML) research. Rather than dismissing the added complexity as incidental to neural dynamics without computational purpose, we explore whether these nonlinear circuit dynamics may serve some advanced function. This raises a compelling question: Can nonlinear circuit dynamics implement more advanced and efficient sampling algorithms? If so, what are advanced circuit algorithms?

To address this question, we perform comprehensive theoretical analyses of the canonical cortical circuit model composed of E neurons and two classes of interneurons (PV and SOM). Our analysis reveals that canonical circuit dynamics not only implements standard Langevin and Hamiltonian sampling as revealed in [15], but innately incorporate the **natural gradient** (NG) to automatically adjust the step size (or the "temperature") in the circuit's Langevin and Hamiltonian sampling based on the local geometry of the posterior distribution measured by the Fisher information (FI).

Specifically, we find the total activity of E neurons monotonically increases with posteriors' FI, dynamically controlling the effective sampling step size in the low-dimensional stimulus feature manifold. Remarkably, the natural gradient sampling in canonical circuit dynamics exhibits computational strategies analogous to established numerical techniques in ML [21–23]. These include, **1**) the recurrent E input acts as a regularization term analogous to adding a small number to the FI before inversion to prevent numerical instabilities (Eq. 8, $\alpha$) – a common practice in NG sampling algorithms; **2**) When coupling multiple canonical circuits interact to sample multivariate stimulus posteriors, the coupled circuit approximates the full FI matrix with its diagonal elements (Sec. 5), similar to the diagonal approximation used in scalable NG samplings. In addition, our analysis reveals that when the circuit transitions from resting state (no feedforward input) to evoked state (with feedforward input), the non-equilibrium canonical circuit dynamics further accelerates sampling beyond the efficiency of standard NG sampling. We analytically identify the **neural annealing** strategy within canonical circuit dynamics (Fig. 2J, Eq. 13b).

**Significance**. The present study provides the first demonstration that canonical cortical circuits with diverse classes of interneurons naturally implement natural gradient Langevin and Hamiltonian sampling. We establish a precise mapping between circuit components and computational elements of advanced sampling algorithms, bridging computational neuroscience and machine learning. And the canonical cortical circuit may inspire more efficient artificial intelligence systems.

## 2 The canonical cortical circuit model

We consider a canonical circuit model widely used in neuroscience research [24, 25, 11, 15], which can be regarded as a hypercolumn in the cortex. The circuit consists of E neurons and two classes of I interneurons (PV and SOM), and is modeled based on continuous attractor networks (CANs, see detailed description in SI. Sec. 3). Briefly, each E neuron has a stimulus tuning and the preference of all E neurons tile the whole stimulus space. E neurons are recurrently connected with a Gaussian kernel in the stimulus space. Both PV and SOM interneurons are driven by E neurons, but differ in function: PV neurons deliver global, divisive normalization to E neurons, whereas SOM neurons provide local, subtractive inhibition. The circuit model reproduces how PV and SOM modulate E neurons' tunings observed in experiments (Fig. S1).

**E neurons' dynamics.** The E neurons in the circuit are selective for a 1D stimulus feature $z$, e.g., spatial location. Denote $\theta_j$ as the preferred stimulus feature of the $j$-th E neuron, and the preferred feature of all $N_E$ E neurons, $\{\theta_j\}_{j=1}^{N_E}$ uniformly cover the whole space of feature space $z$ (Fig. S1), which is widely considered in neuroscience research. Mathematically, in the continuum limit ($\theta_j \to \theta$) corresponding to an infinite number of neurons, the dynamics of E neurons is,

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

bioRxiv preprint doi: https://doi.org/10.1101/2025.07.03.663069; this version posted July 4, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

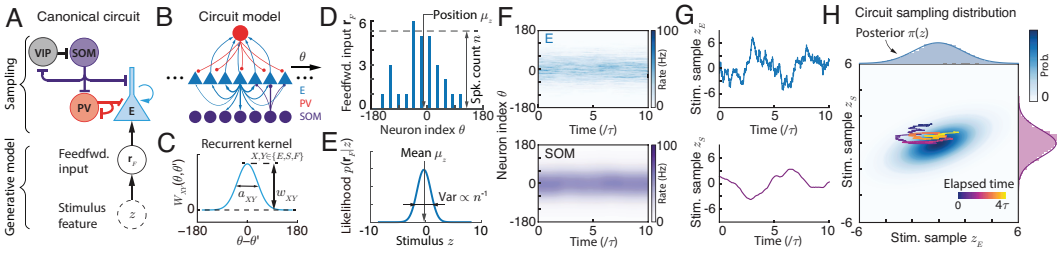



Figure 1: (A) The canonical circuit of E and diverse types of interneurons and sampling-based Bayesian inference. (B) The circuit model in the present study consists of E and two types of interneurons (PV and SOM). (C) The recurrent connection kernel between E neurons. (D) Feedforward input represented as a continuous approximation of Poisson spike trains with Gaussian tuning across the stimulus feature. (E) The feedforward input parametrically encodes the stimulus feature likelihood. (F) The population responses of E (top) and SOM neurons (bottom). (G) Stimulus feature values sampled by the E and SOM neurons respectively. (H) The network's sampling distribution read out from E and SOM neurons. The E neuron's position is regarded as stimulus feature sample $z_E$, while the sample of SOM neurons $z_S$ contributes to the auxiliary momentum variable in Hamiltonian sampling. The distribution of $z_E$ (top marginal) will be used to approximate the posterior.

where $\mathbf{u}_E(\theta,t)$ and $\mathbf{r}_E(\theta,t)$ represent, respectively, the synaptic inputs and firing rates of neurons preferring $z=\theta$. $X$ denotes neuronal types with $E$, $F$ and $S$ stand for E neurons, sensory feedforward inputs, and the SOM neurons respectively. $\tau$ is the time constant, $\rho = N/2\pi$ is the neuronal density covering the stimulus feature space, and$[x]_+ = \max(x,0)$.

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

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

where the scalar $w_{YX}$ is the peak recurrent weight strength and is the only circuit parameter to be adjusted for realizing Bayesian sampling. and $a_{YX}$ the connection width across the stimulus feature space. The symbol $*$ denotes the convolution, i.e., $\mathbf{W}(\theta) * \mathbf{r}(\theta) = \int \mathbf{W}(\theta-\theta')\mathbf{r}(\theta')d\theta'$.

**PV neurons provide divisive normalization.** PV are driven by all E neurons and then provide global, unstructured feedback to E neurons. For simplicity, we consider they are not stimulus tuned, as an extreme case of their weak tuning in reality [16]. It was suggested that PV neurons provide divisive normalization to module E neurons' responses, a canonical operation in the cortex [17, 26, 27]

$$\mathbf{r}_E(\theta,t) = [\mathbf{u}_E(\theta,t)]_+^2/(1+\rho w_{EP} r_P), \quad r_P = \int [\mathbf{u}_E(\theta',t)]_+^2 d\theta', \tag{3}$$

where $r_p$ (scalar) corresponds the mean firing rate of the population of PV neurons. $w_{EP}$ (a positive scalar) characterizes the global inhibition strength from PV neurons to E neurons.

**SOM neurons' dynamics.** Similar to the PV, the SOM neurons are also driven by the E neurons but they have local, *structured* interactions with E neurons. This was supported by experiments that SOM neurons are selective to the stimulus feature and linearly modulate E neurons' responses respectively.

$$\tau\dot{\mathbf{u}}_S(\theta,t) = -\mathbf{u}_S(\theta,t) + \rho(\mathbf{W}_{SE} * \mathbf{r}_E)(\theta,t); \quad \mathbf{r}_S(\theta,t) = g_S \cdot [\mathbf{u}_S(\theta,t)]_+, \tag{4}$$

where $g_S$ (scalar) controls the "gain" of the SOM neurons.

**Stochastic sensory feedforward input.** The feedforward input $\mathbf{r}_F(\theta,t)$ (Eq. 1) is stochastically evoked by an external stimulus $z$, and is modeled as conditionally independent Poisson spikes with Gaussian tuning given a stimulus $z$ (Fig. 1D), consistent with neural coding models (e.g., [28, 12]).

$$\mathbf{r}_F(\theta|z) \sim \text{Poisson}[\lambda_F(\theta|z)], \quad \lambda_F(\theta|z) = R_F\exp[-(\theta-z)^2/2a^2], \tag{5}$$

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

## 3 From circuit dynamics to Bayesian sampling: research philosophy

### 3.1 Subjective posterior sampling

In our framework, the stage from external stimulus $z$ to the feedforward input $\mathbf{r}_F$ is regarded as a generative process (Fig. 1A), and then the circuit dynamics (Eqs. 1 and 4) effectively performs

bioRxiv preprint doi: https://doi.org/10.1101/2025.07.03.663069; this version posted July 4, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Bayesian sampling dynamics to compute the stimulus posterior, $\pi(z) \equiv p(z|\mathbf{r}_F) \propto p(\mathbf{r}_F|z)p(z)$. We hypothesize that the circuit computes **subjective** posterior distributions $\pi(z)$ based on its internally stored generative model [29]. The implicit assumption is the subjective prior in brain's neural circuits matches the *objective* prior (usually not known precisely) of the world. With this hypothesis, we treat the canonical circuit, strongly supported by experiments, as a "ground truth", and aim to identify the circuit's internal generative model and its Bayesian sampling algorithms.

**Subjective prior**. We will leave the subjective prior unspecific for now and will find it through the analysis of the circuit dynamics. This will be shown later at the Eqs. (14 and 17).

**Stimulus likelihood** $L(z)$. The stochastic feedforward input from the stimulus $z$ (Eq. 5) naturally specifies the stimulus likelihood that is calculated as a Gaussian likelihood (see SI. Sec. 3.4),

$$\begin{aligned} &\mathcal{L}(z) \propto p(\mathbf{r}_F|z) = \textstyle\prod_\theta \text{Poisson}[\lambda_F(\theta|z)] \propto \mathcal{N}\big(z|\mu_z, \Lambda^{-1}\big), \\ &\text{where } \ \mu_z = \textstyle\sum_j \mathbf{r}_F(\theta_j)\theta_j / \sum_j \mathbf{r}_F(\theta_j), \ \ \Lambda = a^{-2}\textstyle\sum_j \mathbf{r}_F(\theta_j) = \sqrt{2\pi}\rho a^{-1} R_F. \end{aligned} \tag{6}$$

The Gaussian stimulus likelihood is resulted from the the Gaussian profile of feedforward input tuning $\lambda_F(\theta|z)$ (Eq. 5) [28]. Its mean $\mu_z$ and the precision $\Lambda$ can be read out from $\mathbf{r}_F$ via a linear decoder called population vector [30, 31], Geometrically, $\mu_z$ is regarded as $\mathbf{r}_F$'s location in the stimulus feature space, and $\Lambda$ is proportional to the feedforward input strength $R_F$ (Fig. 1D-E). In this way, a single snapshot of $\mathbf{r}_F$ parametrically conveys the whole stimulus likelihood $p(\mathbf{r}_F|z)$ [28].

### 3.2 Natural gradient Bayesian sampling

A recent study proposed the canonical circuit dynamics can implement both Langevin and Hamiltonian sampling algorithms, depending on the presence of SOM neurons [15]. The present study moves one important step further: we ask whether the circuit can adaptively adjust the sampling step size based on the local geometry (curvature) of the posterior $\pi(z)$, realizing more efficient sampling. Mathematically, Amari's natural gradient is a well-known method to adaptively adjust the sampling step size based on the local geometry characterized by the Fisher information (FI) $G(z)$ [32–34, 22],

$$G(z) = -\mathbb{E}_{\pi(z)}\left[\nabla_z^2 \ln \pi(z)\right]. \tag{7}$$

For a Gaussian distribution $\mathcal{N}(z|\mu, \Lambda^{-1})$, the FI will be its precision $\Lambda$ and doesn't depend on $z$.

**Natural gradient Langevin sampling (NGLS).** For example, the FI will be used to determine the step size of the Langevin sampling dynamics to sample the posterior distribution $\pi(z)$ [22],

$$\dot{z} = \tau_L^{-1}\nabla \ln \pi(z) + (2\tau_L^{-1})^{1/2}\xi_t, \quad \text{where } \ \tau_L = \eta[G(z) + \alpha]. \tag{8}$$

$\alpha$ is a small positive constant acting as a regularization term to improve numerical stability in inverting the FI [21, 23, 35], which is widely used in ML. $\eta$ is a small constant similar to the inverse of "learning rate". In the naive Langevin sampling, $\tau_L$ is fixed rather than proportional to the FI. In the NG Langevin sampling, the $\tau_L$ scales with the FI. If the distribution is widely spread out, the sampling step size will be larger, allowing for faster exploration of the space. Conversely, if the distribution is sharply peaked, the sampling step size will be smaller to explore the local region more thoroughly.

**Natural gradient Hamiltonian sampling (NGHS).** The natural gradient can also be incorporated into Hamiltonian sampling [22] with Hamiltonian function $H(z, p)$ defined as ($p$: auxiliary momentum),

$$H(z, p) = -\ln \pi(z) - \ln \pi(p|z), \quad \pi(p|z) = \mathcal{N}[p|0, G(z)]. \tag{9}$$

$\pi(p|z)$ is the momentum $p$ distribution that is conditioned on $z$, particularly, whose variance (rather than precision) is the FI $G(z)$. The negative logarithm of $\pi(p|z)$ is called kinetic energy. The NGHS with friction is given by [22, 36, 37]

$$\frac{d}{dt}\begin{bmatrix} z \\ p \end{bmatrix} = -\begin{bmatrix} 0 & -\tau_H^{-1} \\ \tau_H^{-1} & \gamma \end{bmatrix}\begin{bmatrix} \nabla_z H \\ \nabla_p H \end{bmatrix} + \sqrt{2}\begin{bmatrix} 0 \\ \gamma^{1/2} \end{bmatrix}\boldsymbol{\xi}_t \tag{10}$$

where $\tau_H$ is the time constant of the Hamiltonian sampling, and $\gamma$ is the friction that dampens momentum. The Hamiltonian dynamics with friction can be interpreted as a Langevin dynamics added into the momentum dynamics [36, 37]. When $\gamma = 0$, the above Hamiltonian dynamics reduces into the naive Hamiltonian dynamics. Our following analysis will show the canonical circuit can automatically implement the natural gradient Langevin sampling and Hamiltonian sampling.

bioRxiv preprint doi: https://doi.org/10.1101/2025.07.03.663069; this version posted July 4, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

# 4 A reduced circuit with E and PV neurons: NG Langevin sampling

## 4.1 Theoretical analysis of the circuit dynamics

To identify the subjective prior in the circuit and investigate how it implements natural gradient sampling, we start by analyzing a reduced circuit model with only E and PV neurons (setting $\mathbf{W}_{ES} = 0$ in Eq. 1). We perform perturbative analysis of the nonlinear circuit dynamics, analytically identify the dominant modes (eigenvectors) of the perturbed dynamics, and then project the dynamics onto these modes. We eventually find the circuit dynamics performs natural gradient sampling in the stimulus feature manifold (subspace) that has the largest eigenvalue in the perturbation dynamics.

**Attractors.** Given a feedforward input $\mathbf{r}_F$ (Eq. 5), it can be checked the synaptic input $\mathbf{u}_E(\theta)$ and firing rate $\mathbf{r}_E(\theta)$ of E neurons both have Gaussian bump attractor states (Fig. S1; SI. Sec. 3.2) ,

$$\bar{\mathbf{u}}_E(\theta) = \bar{U}_E \exp[-(\theta - \bar{z}_E)^2/4a^2], \;\; \bar{\mathbf{r}}_E(\theta) = \bar{R}_E \exp[-(\theta - \bar{z}_E)^2/2a^2]. \tag{11}$$

The Gaussian bump is located at $\bar{z}_X = \mu_z$ on the stimulus feature manifold, with $\mu_z$ is feedforward input's location (Eq. 6). $\bar{U}_E$ and $\bar{R}_E$' are the "height" of the population synaptic input and firing rate respectively, whose analytical solutions are in (Eqs.13a-13b and SI. Sec. 3.1-3.3).

**Perturbative analysis.** The sensory noises and the internal Poisson variability deviate instantaneous neural responses from the attractor state (Eq. 11). We treat each instantaneous response as perturbed from the attractor, i.e., $\mathbf{u}_E(\theta, t) = \bar{\mathbf{u}}_E(\theta) + \delta\mathbf{u}_E(\theta, t)$, and analytically derive the relaxation dynamics of the perturbation $\delta\mathbf{u}_X(\theta, t)$ and its eigenvectors (SI. Sec. 3.3, [38]). The first two dominant eigenvectors correspond to the position $z_x$ and height $U_X$ changes of the population activity [24, 38],

$$\text{Position}: \;\; \phi_1(\theta|z_E) \propto \nabla_z \bar{\mathbf{u}}_E(\theta) \propto (\theta - z_E) \exp[-(\theta - z_E)^2/4a^2], \tag{12a}$$

$$\text{Height}: \;\; \phi_2(\theta|z_E) \propto \;\; \bar{\mathbf{u}}_E(\theta) \propto \exp[-(\theta - z_E)^2/4a^2]. \tag{12b}$$

**Dimensionality reduction for stimulus sampling dynamics.** We further project the E dynamics (Eq. 1) onto the above two dominant 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 hand side of Eq. (1). The projection yields the govening dynamics of the E neurons on the stimulus feature manifold and height manifold (see details in SI. Sec. 3.3)

$$\text{Position}: \;\; \dot{z}_E \approx (\tau U_E)^{-1} U_{EF}(\mu_z - z_E) + \sigma_z (\tau U_E)^{-1/2} \xi_t, \;\; (U_{XY} = \rho w_{XY} R_Y / \sqrt{2}) \tag{13a}$$

$$\text{Height}: \;\; \dot{U}_E \approx \tau^{-1}[-U_E + U_{EE} + U_{EF}] + \sigma_U (\tau^{-1} U_E)^{1/2} \xi_t, \tag{13b}$$

where $U_{XY}$ is the population input magnitude from population $Y$ to $X$. $\sigma_z^2 = 8a\mathsf{F}/(3\sqrt{3\pi})$ and $\sigma_U^2 = \mathsf{F}/(\sqrt{3\pi}a)$ are all constants that don't change with network activities. The approximation comes from omitting negligible nonlinear terms in the circuit dynamics (SI. Sec. 3.3). Our following theoretical analysis on circuit algorithms will be based on the above two equations.

## 4.2 Natural gradient Langevin sampling in the reduced circuit

**E bump position $z_E$ implements Langevin sampling.** We convert the Eq. (13a) into a form consistent with the Langevin sampling (Eq. 8) by rescaling the drift term in Eq. (13a), in that the feedforward input strength is proportional to the likelihod precision, i.e., $U_{EF} \propto \Lambda$ (Eq. 6),

$$\dot{z}_E = (\tau U_E)^{-1} \lambda_z \nabla \ln \mathcal{L}(z_E) + \sigma_z (\tau U_E)^{-1/2} \xi_t, \text{ with } \nabla \ln \mathcal{L}(z) = \Lambda(\mu_z - z), \lambda_z = \frac{w_{EF} a}{2\sqrt{\pi}}. \tag{14}$$

The above equation has several profound implications:

1) The reduced circuit (E and PV neurons) stores a uniform (uninformative) prior and then the circuit dynamics samples the stimulus likelihood $\mathcal{L}(z)$, consistent with the previous study [12, 15].
2) Sampling the likelihood only requires to set the ratio $\sigma_z^2/\lambda_z = 2$ in Eq. (14), which can be realized by setting the feedforward weight as $w_{EF} = \sqrt{\pi}\sigma_z^2/a = (2/\sqrt{3})^3 \mathsf{F}$
3) The E bump height $U_E$ controls the sampling time constant.
4) Eq. (13b)'s drift and diffusion terms both share the E bump height $U_E$, which ensures the circuit with the same weights can sample likelihoods with different uncertainties (Fig. 2B, right).

**E bump height $U_E$ encodes Fisher information.** NGLS requires the circuit's sampling time constant to be proportional to the posterior FI (Eq. 8), i.e., $\tau U_E \propto G(z)$ , and the posterior (likelihood) FI is

bioRxiv preprint doi: https://doi.org/10.1101/2025.07.03.663069; this version posted July 4, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

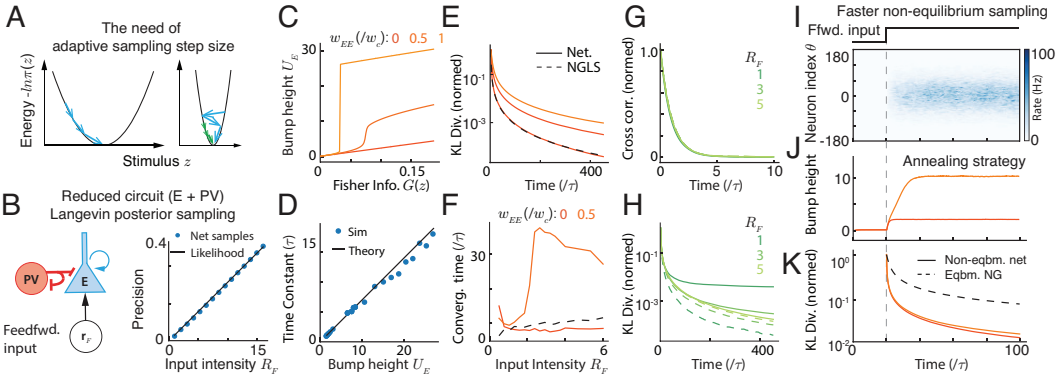



Figure 2: Natural gradient Langevin sampling in the reduced circuit with E and PV neurons. (A) The need for adaptive sampling step size to deal with posteriors with different spreads. (B) The reduced circuit with fixed weights flexibly samples posteriors with different uncertainties in the equilibrium. (C-D) E bump height $U_E$ increases with posterior Fisher information (C) and determines the sampling time constant, or step size, in the stimulus feature manifold. (E-F) The sampling convergence with recurrent weight $w_{EE}$ that acts as the regularization coefficient (Eq. 8) (G) KL divergence; (G) Convergence time. (G-H) The sampling convergence at different posterior uncertainties. The cross correlations of samples (E) and the KL divergence from the sampling distribution to the posterior over time (F). (I-K) Non-equilibrium sampling further accelerates convergence. (I) The non-equilibrium population responses (J) and bump height (L) from the resting state (no feedforward input) to evoked state. (K) The KL divergence from the sampling distribution to the posterior.

$G(z) = \Lambda = \sqrt{2\pi}\rho a^{-1} R_F$ (Eq. 6). We then analyze the relation between the E bump height $U_E$ and $G(z)$. For simplicity, we first focus on the equilibrium mean of $U_E$ by averaging Eq. (13b),

$$\bar{U}_E = U_{EE} + U_{EF}, \quad U_{EF} = \rho w_{EF} R_F / \sqrt{2} = \lambda_z \cdot G(z) \tag{15}$$

where $\bar{R}_E = \bar{U}_E^2/[1 + \sqrt{2\pi} k\rho a \bar{U}_E^2]$ is the average height of the E firing rate bump (Eq. 11), which is realized by substituting the $\bar{U}_E$ into Eq. (3). Combining the above two equations, we find

1) The $U_E$ monotonically increases with the FI, confirming the circuit implements the NGLS.
2) The $U_E$ is augmented by E recurrent input ($w_{EE}\bar{R}_E$ in Eq. 15) that acts as the regularization coefficient $\alpha$ to improve numerical stability in inverting FI (Eq. 8, [21, 23, 35]).
3) The circuit' NGLS coefficients (Eq. 8, Fig. 4E): $\eta = \tau\lambda_z$ and $\alpha = \lambda_z^{-1} U_{EE}$.

The simulations of the circuit model conform above theoretical analyses. Fig. 2C-D shows the bump height $U_E$ indeed increases with the FI, and determines the sampling time constant on the stimulus feature manifold. Without recurrent E weight ($w_{EE} = 0$), the circuit sampling behaves similarly with the NGLS (Fig. 2E). Including the recurrent weight enlarges the sampling time constant, slowing down the sampling (Fig. 2E) as suggested by our theory. Nevertheless, the recurrent E input, acting as the regularizer to stabilize the FI inversion (Eq. 8, $\alpha$), helps the circuit sampling in the case of very low FI, exhibited by higher recurrent weight incurring shorter convergence time at extremely low FI (Fig. 2F). Moreover, a characteristic of the NG Langevin sampling is that the temporal correlation of samples is invariant with the posterior uncertainty controlled by input intensity $R_F$ (Fig. 2F), while the different decay of KL divergence is due to different initial conditions (Fig. 2H)

### 4.3 Non-equilibrium circuit dynamics accelerates natural gradient sampling

The above analysis focuses on the equilibrium mean $\bar{U}_E$ (Eq. 15). Now we extend to its non-equilibrium dynamics (Eq. 13b), especially when the circuit is just evoked by a feedforward input $\mathbf{r}_F$ from the resting state (no feedforward input). After receiving a $\mathbf{r}_F$, the E neurons' population activity will gradually grow up until the equilibirum state (Fig. 2I-J). Since a smaller $U_E$ leads to a larger sampling step size, the sampling step size will be larger at the begining of the circuit dynamics, and then gradually decay to a smaller equilibrium value, similar to the **annealing** in stochastic computation. The larger sampling step size during the non-equibrlium implies the circuit can sample faster than the equilibrium state, confirmed by simulation (Fig. 2K). Furthermore, $U_E$ temporal trajectory (Fig. 2J) describes circuit' sampling **annealing strategy**, governed by Eq. 13b.

bioRxiv preprint doi: https://doi.org/10.1101/2025.07.03.663069; this version posted July 4, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

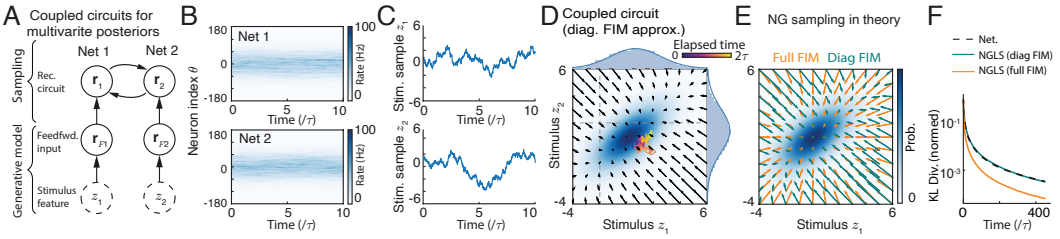



Figure 3: Coupled canonical circuits sample multivariate posteriors, with each circuit $m$ sampling the corresponding marginal posterior of $z_m$. (A) The structure of decentralized circuit with each consisting of E and PV neurons (the same as Fig. 2B). (B-C) The spatiotemporal E neuronal responses in two circuits (B) and the decoded stimulus samples (C). (D) When concatenating the samples from two circuits together, we recover the bivariate sampling distribution. Vector field: the drift term of the sampling dynamics in the circuit. (E) The vector field of natural gradient sampling with full FIM and diagonal FIM approximation. (F) The convergence speed in the decentralized circuit. The diagonal FIM approximation is scalable in high dimensions, while paying the cost of slower sampling speed.

## 5 Coupled circuits: NG Langevin Sampling of multivariate posteriors

The brain needs to sample multivariate stimulus posteriors, which can be implemented by a **decentralized** system consisting of multiple coupled canonical circuit modules (Fig. 3A, [39, 12, 40]). Each circuit module $m$ receives a feedforward input stochastically evoked by a 1D stimulus $z_m$ (Fig. 3), and the cross-talk between circuits enables them to read out the circuit prior, and eventually each circuit $m$ samples the corresponding stimulus $z_m$ distributedly. As a proof of concept, we consider the *smallest* decentralized system of two coupled circuits to sample bivariate posteriors (Fig. 3A). The sampling of higher dim. posteriors can be extended by inserting more circuit modules, with the number of circuit modules determined by the stimulus dimension.

We investigate how the coupled circuits implement bivariate posteriors' NGLS, and what kind of approximation, if there is any, is used in the circuit. The theoretical analysis of the two coupled circuits is similar to a single circuit module, but we perform the analysis on each circuit module individually, yielding the new position and height dynamics (details at SI. Sec. 4),

$$\text{Position}: \ \dot{\mathbf{z}}_E = (\tau \mathbf{D_U})^{-1}\big[-\mathbf{L}\mathbf{z}_E + \mathbf{U}_{EF} \circ (\boldsymbol{\mu}_\mathbf{z} - \mathbf{z}_E)\big] + \sigma_z (\tau \mathbf{D_U})^{-1/2}\boldsymbol{\xi}_t. \tag{16a}$$

$$\text{Height}: \quad \dot{\mathbf{U}}_E = \tau^{-1}\big(-\mathbf{U}_E + \mathbf{U}_{EE} + \mathbf{U}_{EF}\big) + \sigma_U (\tau^{-1}\mathbf{D_U})^{1/2}\boldsymbol{\xi}_t. \tag{16b}$$

$\mathbf{z}_E = (z_1, z_2)^\top$ is two circuits' E bump positions. Similarly for $\boldsymbol{\mu}_\mathbf{z}$ and $\mathbf{R}_F$ (feedfwd. input position and intensity respectively), and $\mathbf{U}_E$ and $\mathbf{R}_E$ (bump height of synaptic input and firing rate respectively). $\mathbf{D}_U = \text{diag}(\mathbf{U}_E)$ is a diagonal matrix of $\mathbf{U}_E$. The $\circ$ denotes the element-wise product.

**The (subjective) bivariate stimulus prior.** The $\mathbf{z}_E$ dynamics (Eq. 16a) has a new term, i.e., $-\mathbf{L}\mathbf{z}_E$, which can be linked to the internal stimulus prior (omitting the subscript $EE$ of $U$ for clarity below).

$$\nabla \ln p(\mathbf{z}) = -\mathbf{L}\mathbf{z}_E, \ \ \mathbf{L} = \begin{pmatrix} U_{12} & -U_{12} \\ -U_{21} & U_{21} \end{pmatrix} \ \Leftrightarrow \ \ p(\mathbf{z}) \propto \exp\left(-\mathbf{z}_E^\top \mathbf{L}\mathbf{z}_E/2\right). \tag{17}$$

Hence the coupling matrix $\mathbf{L}$ is the prior's precision matrix (details at SI. Sec. 4.2). To ease of understanding, we expand the bivariate prior as $p(z_1, z_2) \propto \exp\left[-\Lambda_{12}(z_1 - z_2)^2/2\right]$, with $\Lambda_{12} \propto (U_{12} + U_{12})/2$. The coupled circuit stores an associative (correlational) stimulus prior with each marginal uniform, which was also suggested in previous studies [15, 12]. The identified correlational prior is confirmed by numerical simulation, where the actual sampling distribution of the circuit dynamics matches the subjective posterior predicted via the identified prior (Fig. S2).

**Diagonal Fisher information approximation**. Given the identified circuit's prior, we can calculate Fisher information matrix (FIM) $\mathbf{G}(\mathbf{z})$ of the posterior calculated by the circuit,

$$\text{FI:} \ \mathbf{G}(\mathbf{z}) = \lambda_z^{-1} \begin{pmatrix} U_1 & -U_{12} \\ -U_{21} & U_2 \end{pmatrix} \ \text{ vs. } \ \text{Circuit:} \ \mathbf{D}_U = \begin{pmatrix} U_1 & \\ & U_2 \end{pmatrix}. \tag{18}$$

The time constant matrix $\mathbf{D}_U$ in the circuit dynamics is proportional to the diagonal elements of the FIM, i.e., $\mathbf{D_U} \propto \text{diag}[\mathbf{G}(z)]$, suggesting the circuits utilize the **diagonal approximation** of the FIM

bioRxiv preprint doi: https://doi.org/10.1101/2025.07.03.663069; this version posted July 4, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

to scale the Langevin sampling step size. The diagonal FIM approximation is widely used in ML, as a trade-off of computational efficiency and accuracy [32–35], where the full FIM is hard to estimate.

To illustrate the effect of diagonal FIM approximation on sampling dynamics, we plot the vector field of NGLS with full FIM and diagonal FIM respectively (Fig. 3D). All vector fields under full FIM directly point to the posterior mean, while the ones under diagonal FIM are curled along the long axis of the posterior. Meanwhile, the curled vector fields also exist in coupled circuits' sampling dynamics, confirming diagonal FIM approximation in the circuit (Fig. 3C). The diagonal FIM simplifies the computation, while paying the cost of sampling speed (Fig. 3E).

## 6 The circuit with SOM neurons: natural gradient Hamiltonian sampling

We investigate the Bayeisan sampling in the augmented circuit model with SOM neurons providing structured inhibition to E neurons (Eq. 1). The SOM's structured inhibition can add the Hamiltonian sampling component in the circuit [15]. We further analyze whether the augmented circuit with SOM neurons implements the natural gradient Hamiltonian sampling (NGHS) [15]. For simplicity, we consider a augmented circuit model to sample a univariate stimulus posterior. Similarly, we derive the eigenvectors of the SOM's dynamics and then project the dynamics on dominant eigenvectors (see details in SI. Sec. 5.1). The position dynamics of the E and SOM neurons are,

$$\text{E neurons:}\quad \tau_E \dot{z}_E \approx U_{ES}(z_S - z_E) + U_{EF}(\mu_z - z_E) + \sigma_z\sqrt{\tau_E}\xi_t, \tag{19a}$$

$$\text{SOM:}\quad \tau_S \dot{z}_S \approx U_{SE}(z_E - z_S), \tag{19b}$$

where $\tau_X = \tau U_X$. To understand the circuit's sampling dynamics, we decompose the E neurons' dynamics (Eq. 19a) into two parts, i.e., the Langevin and Hamiltonian part,

$$\tau_E \dot{z}_E = \underbrace{[U_{ES}(z_S - z_E) + (1-\alpha_L)U_{EF}(\mu_z - z_E)]}_{\text{Momentum } p,\ \text{(Hamiltonian part)}} + \underbrace{[\alpha_L U_{EF}(\mu_z - z_E) + \sigma_z\sqrt{\tau_E}\xi_t]}_{\text{Langevin part}}, \tag{20}$$

where the momentum $p$ is a mixture of $z_E$ and $z_S$. $\alpha_L \in [0,1]$ denotes the proportion of Langevin sampling component. Transforming the $(z_E, z_S)$ dynamics (Eqs. 19a - 19b, Fig. 1H) into the $(z_E, p)$ dynamics (Fig. 4B) shows a mixture of Langevin and Hamiltonian sampling in the augmented circuit,

$$\frac{d}{dt}\begin{bmatrix} z_E \\ p \end{bmatrix} = -\begin{bmatrix} \alpha_L\lambda_z\tau_E^{-1} & -\beta_E\Lambda^{-1} \\ \beta_E\Lambda^{-1} & \tau U_E\beta_p\beta_E\Lambda^{-1} \end{bmatrix}\begin{bmatrix} -\nabla_z \ln\pi(z_E) \\ (\tau_E\beta_E)^{-1}\Lambda\cdot p \end{bmatrix} + \begin{bmatrix} \sigma_z\tau_E^{-1/2} \\ \sigma_p \end{bmatrix}\boldsymbol{\xi}_t \tag{21}$$

where $\beta_p$, $\beta_E$ and $\sigma_p$ are functions of the coefficients in Eq. (20) (details at Eq. S52). And the momentum $p$ dynamics has a friction term (Eq. 21), corresponding to a Langevin component.

**Conditions for Hamiltoian sampling.** Comparing the Eq. (21) with the Langevin (Eq. 8) and Hamiltonian sampling (Eq. 10), we find the constraint for the circuit to sample the posterior $\pi(z_E)$

$$(a).\ \alpha_L\lambda_z\tau_E^{-1} = \sigma_z^2\tau_E^{-1}/2 \quad (b).\ \tau_E\beta_p\beta_E\Lambda^{-1} = \sigma_p^2/2$$

Solving the two constraints, we can derive the requirement of posterior sampling on circuit weights,

$$\left(U_E^{-1}R_S\right)\cdot w_{ES} - \left[(1-\alpha_L)U_E^{-1}R_F\right]\cdot w_{EF} = \left[Q(\alpha_L)U_S^{-1}R_E\right]\cdot w_{SE}. \tag{22}$$

$Q(\alpha_L)$ is a nonlinear function of $\alpha_L$ specifying the proportion of Langevin sampling, which remains invariant with feedforward input and network activities (Eq. S57). $U_X$ and $R_X$ are the height of the population synaptic input and firing rate of neuronal populations $X$ (Eq. 11). Eq. (22) suggeest there is a line manifold in the circuit's weight space of $(w_{ES}, w_{EF})$ to ensure the correct posterior sampling, which is confirmed by numerical simulation (Fig. S3). Furthermore, once the weight parameter is set within the line manifold, the circuit with fixed weights can flexibly sample the posterior $\pi(z_E)$ with different uncertainties (Fig. 4C).

**Natural gradient Hamiltonian sampling.** Implementing NGHS requires the precision of the momentum $p$ to be inversely proportional to posterior's FI, $G(z)$ (Eq. 10). To verify this, we calculate the momentum distribution $\pi(p|z)$ in the circuit (comparing Eqs. 21 and 10),

$$-\nabla\ln\pi(p|z) = (\tau_E\beta_E)^{-1}\Lambda\cdot p \ \Rightarrow\ \pi(p|z) = \mathcal{N}(p|0,\Lambda_p^{-1}),\ \text{where}\ \Lambda_p = (\tau_E\beta_E)^{-1}\Lambda \tag{23}$$

We analyze the momentum precision $\Lambda_p$ in the circuit dynamics. Since $\beta_E$ is a complex, quadratic function of neuronal responses, we then use *order* analysis to provide insight (details at SI. Sec. 5.4). In the circuit dynamics, we calculate $\beta_E \sim \mathcal{O}(G(z))$, $U_E \sim \mathcal{O}(G(z))$, and $G(z) = \Lambda$, and then we have $\Lambda_p \propto \mathcal{O}(G(z)^{-1})$, suggesting the momentum precision $\Lambda_p$ decreases with posterior's FI and satisfys the requirement of NGHS. This is confirmed by simulation results (Fig. 4D).

bioRxiv preprint doi: https://doi.org/10.1101/2025.07.03.663069; this version posted July 4, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

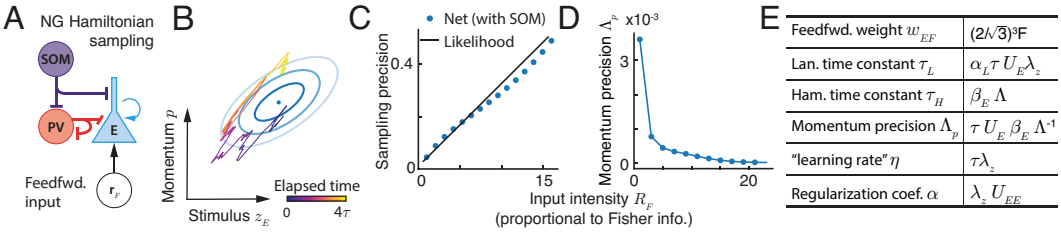



Figure 4: Natural gradient Hamiltonian sampling in the augmented circuit with E, PV, and SOM neurons. (A) The circuit structure. (B) The decoded trajectory of stimulus sample $z_E$ and momentum $p$ exhibits an oscillation pattern, which is a characteristic of Hamiltonian sampling. The momentum $p$ is a weighted average of sample $z_E$ and $z_S$ as shown in Fig. 1H. (C) The circuit with fixed weights can sample posteriors with different uncertainties. (D) The momentum precision decreases with the Fisher information controlled by feedforward input strength, satisfying the requirement of natural gradient Hamiltonian sampling. (E) A table summarizing the circuit' sampling parameters.

## 7 Conclusion and Discussion

The present theoretical study for the first time discovers that the canonical cortical circuit dynamics with E and two classes of interneurons (PV and SOM) innately implement **natural gradient** sampling of stimulus posteriors, deepening our understanding of the computational capability of neural circuits. The circuit dynamics samples stimulus posterior in the stimulus manifold that is geometrically regarded as the E neurons' response bump position. Importantly, we find the E bump height (sum of population responses) encodes the Fisher information of the stimulus posterior, and then utilizing the local geometry of posterior to determines the sampling step size of the bump position. We find the **non-equilibirum** dynamics of the E bump height can further accelerate sampling, and explicitly identify the circuit **annealing** strategy (Fig. 2I-K). Remarkably, we discover the circuit dynamics also utilizes computational approximations widely used in ML algorithms, including the regularization coefficient for inverting FI (Eq. 15) and the diagonal FI matrix approximation in multivariate cases (Eq. 16b), which provides a direct evidence to validate the biological plausibility of artificial ML algorithms. Our work unprecedentedly links the canonical circuit with classes of interneurons to natural gradient sampling and related computational approximations, providing deep insight into the capability of circuit sampling.

**Comparison with previous studies.** To our best knowledge, only one study investigated the natural gradient sampling in recurrent networks [14]. It is difficult to make direct and "fair" comparison since the network being used in two studies are different: The previous study considers a spiking network without explicit defining neuron types, while the present study considers a rate-based network with diverse classes of interneurons. We explicitly link the role of diverse interneuron classes (PV and SOM) to the sampling algorithms (Langevin and Hamiltonian), and even identify the computational approximations in the circuit dynamics. From functional perspective, our proposed circuit with *fixed weights* can flexibly sample posteriors with different mean and uncertainties, which is an essential function for neural circuit. Nevertheless, it remains unknown whether this flexibly can be realized by the previous study. In addition, the proposed circuit model is similar to the one in [15], which also suggested the Langevin and Hamiltonian sampling dynamics in the circuit. However, the previous study didn't link the circuit dynamics to the natural gradient sampling.

**Limitations and future directions.** The proposed circuit model doesn't include VIP neurons (Fig. 1), which forms one of our future directions. Based on our present finding, it is likely that VIP neurons act as a "knob" to modulate the sampling speed and the kinetic energy (SI. Sec. 5.4) by modulating the "gain" of SOM neurons ($g_S$, Eq. 4). Moreover, we find the canonical circuit model widely used in neuroscience studies only stores a uniform univariate stimulus prior or multivariate prior but each marginal uniform, as a result of neurons uniformly distributing on the stimulus manifold and translation-invariant recurrent weights (Eq. 2). The uniform distribution of neurons was proposed long time ago as an ideal case and significantly simplifies dynamical analysis, however, it imposes a limit on the prior stored in the circuit. Our result suggests the circuit has to break the neuronal homogeneity on the stimulus manifold to store a non-uniform (marginal) prior [41]. All of these form our future research.

bioRxiv preprint doi: https://doi.org/10.1101/2025.07.03.663069; this version posted July 4, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

**Acknowledgments**

W.H.Z is supported by the UT Southwestern Endowed Scholars program.

bioRxiv preprint doi: https://doi.org/10.1101/2025.07.03.663069; this version posted July 4, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

## 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] 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.

[7] 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.

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

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

[10] 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.

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

[12] 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.

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

[14] Paul Masset, Jacob Zavatone-Veth, J Patrick Connor, Venkatesh Murthy, and Cengiz Pehlevan. Natural gradient enables fast sampling in spiking neural networks. *Advances in neural information processing systems*, 35:22018–22034, 2022.

[15] Eryn Sale and Wenhao Zhang. The bayesian sampling in a canonical recurrent circuit with a diversity of inhibitory interneurons. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

[16] 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.

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

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

[19] 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.

bioRxiv preprint doi: https://doi.org/10.1101/2025.07.03.663069; this version posted July 4, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

[20] 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.

[21] Dongseong Hwang. Fadam: Adam is a natural gradient optimizer using diagonal empirical fisher information, 2024.

[22] Mark Girolami and Ben Calderhead. Riemann manifold langevin and hamiltonian monte carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(2):123–214, 2011.

[23] Gaétan Marceau-Caron and Yann Ollivier. Natural langevin dynamics for neural networks. In *International Conference on Geometric Science of Information*, pages 451–459. Springer, 2017.

[24] 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.

[25] 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.

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

[27] 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.

[28] 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.

[29] Richard D Lange, Sabyasachi Shivkumar, Ankani Chattoraj, and Ralf M Haefner. Bayesian encoding and decoding as distinct perspectives on neural coding. *Nature Neuroscience*, 26(12):2063–2072, 2023.

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

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

[32] Shun-Ichi Amari. Natural gradient works efficiently in learning. *Neural Computation*, 10(2):251–276, 1998.

[33] Shun-Ichi Amari and Scott C Douglas. Why natural gradient? In *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'98 (Cat. No. 98CH36181)*, volume 2, pages 1213–1216, 1998.

[34] Shun-Ichi Amari. *Information geometry and its applications*, volume 194. Springer, 2016.

[35] Xiaodong Wu, Wenyi Yu, Chao Zhang, and Phil Woodland. An improved empirical fisher approximation for natural gradient descent. *Advances in Neural Information Processing Systems*, 37:134151–134194, 2024.

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

[37] Yi-An Ma, Tianqi Chen, and Emily Fox. A complete recipe for stochastic gradient mcmc. *Advances in neural information processing systems*, 28, 2015.

[38] 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.

bioRxiv preprint doi: https://doi.org/10.1101/2025.07.03.663069; this version posted July 4, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

[39] 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.

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

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