博士学位論文

Nonequilibrium Dynamics in Sparse Neural Activity: Insights from State-Space Kinetic Ising Model

（スパースな神経活動における非平衡ダイナミクス：状態空間キネティック・イジングモデルに基づく知見）

石原 憲

北海道大学大学院生命科学院

2026 年 3 月

Contents

Abstract

Neuronal population activity exhibits nonequilibrium dynamics accompanied by broken time-reversal symmetry, and the resulting irreversibility is linked to entropy dissipation. However, directly evaluating time asymmetry from spike trains is challenging because firing rates and couplings vary over time, i.e., the signals are nonstationary. This dissertation extends the kinetic Ising framework for causal spike dynamics into a state-space model and combines sequential Bayesian estimation with a mean-field approximation to construct a new framework that estimates entropy flow that changes over time. The goal is to clarify how task engagement reorganizes neural couplings and entropy flow, and to demonstrate their relationship to behavioral performance.

The dissertation consists of two chapters. Chapter 1 formalizes a method that integrates an observation model—binary spiking with causal dependence on the immediately preceding bin—with a state model in which time-varying parameters for external drive and couplings evolve smoothly. Using sequential Bayesian filtering/smoothing with a Laplace approximation and the EM algorithm, the method identifies parameter means and uncertainties along the time series, and it derives a mean-field estimator for entropy flow that accounts for temporal variation and scales to large datasets without sampling. Simulations recover ground-truth time-varying parameters accurately in small systems and converge within practical time for datasets with 80 neurons and 550 trials, confirming the validity and scalability of the proposed approach.

Chapter 2 applies the method to the Allen Institute Visual Behavior Neuropixels dataset (37 mice), comparing mouse V1 spike trains for the same natural images under two conditions: active task engagement and passive replay. The analysis used 10 ms bins within a 750 ms post-onset window and focused on the top 80 simultaneously recorded neurons per animal. In the active condition, mean firing rates decreased and the distribution became sparser, while the variance and asymmetry of couplings increased. Trial-shuffled controls confirmed that these changes reflect genuine variations in inter-neuronal couplings rather than estimation noise. Entropy flow exhibited transient increases immediately after stimulus onset and offset, aligning with epochs of declining mean activity. Although the total entropy flow (the sum across all bins) was smaller in the active condition, the

shuffle-difference component—attributable to couplings—was positive in both conditions, suggesting that variations in couplings elevate dissipation. Furthermore, a perturbation analysis that uniformly rescales all model parameters revealed an interaction-driven difference in irreversibility that becomes apparent in a low-gain regime and is stronger during active engagement. With respect to behavior (d'), animals with higher entropy flow per spike (from the shuffle difference) performed better in the active condition; this relationship provided explanatory power independent of sparsification indices, suggesting that efficient computation during task performance is supported by time-asymmetric causal activity.

In summary, the proposed method estimates time-dependent entropy flow from nonstationary, nonequilibrium spike activity and elucidates how task engagement reorganizes coupling dynamics in ways that relate to behavioral outcomes. The scholarly significance of this work lies in presenting a practical tool for testing the thermodynamic underpinnings of neural computation using real data.

Acknowledgements

I am deeply grateful to everyone who supported me during this work.

First and foremost, I thank Professor Hideaki Shimazaki, my principal mentor over the past five years, for sustained guidance, incisive feedback, and unfailing encouragement. His scientific vision and high standards shaped the questions I asked and the way I attempted to answer them.

I also wish to express my sincere appreciation to Professor Shinji Nakaoka for generous advice and remote supervision, and to Ms. Keiko Saito for warm and comprehensive administrative support. I am likewise grateful to Ms. Miwa Nakatsu (Shimazaki Laboratory) for kind and efficient assistance on many occasions. I am indebted to the members of the Shimazaki group and of the Nakaoka Laboratory for stimulating discussions and a collegial atmosphere.

During my master's studies, the Center for Human Nature, Artificial Intelligence, and Neuroscience (CHAIN), Hokkaido University, provided an invaluable cross-disciplinary community. I especially thank Professor Shigeru Taguchi, Director of CHAIN, for leadership and for fostering the environment that made such exchanges possible. Interactions with researchers from diverse graduate schools broadened my perspective and enriched this dissertation.

I am grateful to collaborators and colleagues for constructive conversations and shared efforts, and to friends and family for their patience and encouragement throughout the ups and downs of research.This work would not have been possible without the institutional resources and support I received.

Chapter 1

Nonequilibrium Dynamics in Sparse Neural Activity: Insights from State-Space Kinetic Ising Model

Introduction

Animals—including humans—understand and act upon the external world by representing environmental changes as internal neural states. In this view, the brain encodes sensory inputs into internal representations, and the empirical task of decoding—inferring stimuli or behavior from neural activity—provides a principled way to probe the structure and computational principles of such representations [1]. Classical approaches have established Bayesian state–space point–process models for spike-train decoding with strong performance in tasks such as reconstructing kinematics from motor cortex activity [1]. In parallel, models that explicitly harness population correlations—most notably maximum–entropy/Ising formalisms—have been proposed as decoders capable of millisecond-level stimulus discrimination and codeword extraction in large ensembles [2, 3].

A growing body of empirical work across scales indicates that neural activity generically violates time-reversal symmetry and operates far from thermodynamic equilibrium. Physiologically, synaptic signaling is energetically costly and continuously dissipative, implying sustained ATP consumption and heat production that maintain the brain away from equilibrium [4]. At the mesoscopic circuit level, neural ensembles exhibit robust, time-directed trajectories rather than time-symmetric fluctuations: rotational population dynamics in primate motor cortex during reaching reveal stereotyped phase-advanced components consistent with directed flow in neural state space [5], while hippocampal place-cell sequences in behaving rats anticipate future paths to remembered goals, expressing temporally ordered patterns that underwrite navigation and planning [6].

At macroscopic scales, violations of detailed balance and increases in entropy production covary with cognitive engagement and conscious state. In large-cohort human fMRI, the brain is close to detailed balance at rest but strongly breaks it during physically and cognitively demanding tasks, yielding net entropy flow that scales with task load [7]. Electrocorticography in non-human primates and whole-brain human recordings further show that conscious wakefulness is marked by higher entropy production and stronger probability-flux curls than sleep or anesthesia, i.e., reduced-consciousness states lie closer to equilibrium [8, 9]. Complementarily, model-free irreversibility metrics on EEG and deep-learning–based "arrow-of-time" measures on fMRI demonstrate that temporal asymmetry increases during tasks and decreases in pathology, linking nonequilibrium signatures to functional status [10, 11]. Together, these findings argue that neural systems operate in nonequilibrium steady states, with the magnitude of irreversibility reflecting behavioral demands and level of consciousness.

On the theoretical side, recent work has decomposed the "arrow of time" (entropy production) into contributions from successively higher-order interactions, suggesting that in small neural populations much of the observed irreversibility can be explained by low-order statistics [12]. In addition, nonequilibrium quantification has advanced rapidly: exact or closed-form expressions for entropy production are now available in neural circuit and neural-field models, clarifying how stochastic driving and network structure set irreversibility [13, 14, 15].

A central methodological challenge is that spike trains are inherently nonstationary: not only firing rates but also effective couplings (causal interactions) vary over time. Many existing analyses tacitly rely on equilibrium approximations or stationarity, making it difficult to estimate entropy production—or its key component, entropy flow—in a framework that is simultaneously consistent with nonequilibrium and nonstationarity at the spike level. While Ising-based decoders are powerful for analyzing population codes, classical maximum-entropy formulations often assume equilibrium (symmetric couplings); direct characterization of irreversibility therefore requires extending these models to explicitly accommodate broken detailed balance [2].

In this dissertation, we address these gaps by developing a nonlinear, nonequilibrium time-series framework for spike trains that enables estimation of entropy production (or entropy flow). Specifically, (i) we build on a kinetic Ising model to make causal dependencies explicit; (ii) we capture nonstationarity by representing time-varying firing rates and couplings in a state–space formulation; and (iii) we introduce a mean-field procedure to obtain time-resolved entropy flow estimates. Prior theory and data analyses motivate this direction: balanced networks can exhibit high entropy production via deterministic chaos [13], and large-cohort fMRI shows strengthened temporal asymmetry during task

engagement [11].

Contributions. (1) We introduce an estimation framework for a state–space kinetic Ising model that jointly handles nonstationarity (time-varying rates and couplings) and nonequilibrium (asymmetric couplings). (2) We derive a practically computable estimator of time-dependent entropy flow consistent with the model, and validate it in theory and with experimental spike data. (3) We show that the resulting entropy-flow metrics reveal spike-level irreversibility that complements macroscopic measures, and that they relate to behavioral performance and efficient coding [12, 3].Together, these developments bridge spike-train decoding with thermodynamic indicators of nonequilibrium computation offering a new lens in neural computation.

Results

The state-space kinetic Ising model

In neurophysiological experiments, the experimentalists simultaneously record the activity of multiple neurons while animals are exposed to a stimulus or perform a task, and repeat the recordings multiple times under the same experimental conditions. We analyze the quasi-simultaneous activity of neurons using binarized spike sequences. For this goal, we convert the simultaneous sequences of spike timings of N neurons into sequences of binary patterns by binning them with a bin width of Δ [ms]. We assign a value of 1 if there is one or more spikes in a bin and 0 otherwise. We assume that there are $T+1$ bins for each trial, with an initial bin being the 0-th bin and L trials in total. Below, we treat the bins as discrete time steps and refer to the t-th bin as time t. We let $x^l_{i,t} = \{0, 1\}$ be a binary variable of the i-th neuron at time t in the l-th trial ($i = 1, \ldots, N$, $t = 0, \ldots, T$, $l = 1, \ldots, L$). We collectively denote the binary patterns of simultaneously recorded neurons at time t in the l-th trial using a vector, $\mathbf{x}^l_t = (x^l_{1,t}, \ldots, x^l_{i,t}, ...x^l_{N,t})$. Further, we denote the patterns at time t from all trials by $\mathbf{x}_t = (\mathbf{x}^1_t, \ldots, \mathbf{x}^l_t, \ldots, \mathbf{x}^L_t)$ and denote all the patterns up to time t by $\mathbf{x}_{0:t}$.

We construct the state-space kinetic Ising model to account for the nonequilibrium dynamics of the binary sequences by extending the state-space models developed for equilibrium Ising systems [16, 17]. The state-space model is composed of the observation

model and the state model. The observation model in the t-th bin is

$$p(\mathbf{x}_t|\mathbf{x}_{t-1}, \boldsymbol{\theta}_t) = \prod_{l=1}^{L}\prod_{i=1}^{N} p(x_{i,t}^l|\mathbf{x}_{t-1}^l, \boldsymbol{\theta}_t^i)$$
$$= \prod_{l=1}^{L}\prod_{i=1}^{N} \exp\left[\theta_{i,t}x_{i,t}^l + \sum_{j=1}^{N}\theta_{ij,t}x_{i,t}^l x_{j,t-1}^l - \psi(\boldsymbol{\theta}_t^i, \mathbf{x}_{t-1}^l)\right], \tag{1.1}$$

where $\theta_{i,t}$ is a time-dependent (external) field parameter that determines the bias for inputs to the i-th neuron at time t and $\theta_{ij,t}$ is a time-dependent coupling parameter from the j-th neuron to the i-th neuron. These parameters are collectively denoted as $\boldsymbol{\theta}_t = (\boldsymbol{\theta}_t^1, \ldots, \boldsymbol{\theta}_t^i, \ldots, \boldsymbol{\theta}_t^N)$ and $\boldsymbol{\theta}_t^i = (\theta_{i,t}, \theta_{i1,t}, \ldots \theta_{ij,t}, \ldots \theta_{iN,t})$. $\psi(\boldsymbol{\theta}_t^i, \mathbf{x}_{t-1}^l)$ is the log normalization term defined as

$$\psi(\boldsymbol{\theta}_t^i, \mathbf{x}_{t-1}^l) = \log\left[1 + \exp\left[\theta_{i,t} + \sum_{j=1}^{N}\theta_{ij,t}x_{j,t-1}^l\right]\right]. \tag{1.2}$$

We also specify $p(\mathbf{x}_0)$, a probability mass function of the binary patterns at $t = 0$, which we assume $p(\mathbf{x}_0) = \prod_{i=1}^{N}\prod_{l=1}^{L} p(x_{i,0}^l)$, where $p(x_{i,0}^l) = 0.5$.

Next, we introduce a state model of the time-varying parameters $\boldsymbol{\theta}_t$ for $t = 1, \ldots, T$:

$$p(\boldsymbol{\theta}_{1:T}|\mathbf{w}) = \prod_{i=1}^{N}\left[p(\boldsymbol{\theta}_1^i|\boldsymbol{\mu}^i, \boldsymbol{\Sigma}^i)\prod_{t=2}^{T} p(\boldsymbol{\theta}_t^i|\boldsymbol{\theta}_{t-1}^i, \mathbf{Q}^i)\right], \tag{1.3}$$

where $\mathbf{w}$ denotes the collection of the hyperparameters: $\mathbf{w} = [\boldsymbol{\mu}^1, \ldots, \boldsymbol{\mu}^N, \boldsymbol{\Sigma}^1, \ldots, \boldsymbol{\Sigma}^N, \mathbf{Q}^1, \ldots, \mathbf{Q}^N]$. Namely, we assume independence of the parameters of a neuron from those of the other neurons, which significantly reduces computational costs. The transition of the i-th neuron follows the linear Gaussian models:

$$p(\boldsymbol{\theta}_t^i|\boldsymbol{\theta}_{t-1}^i, \mathbf{Q}^i) \tag{1.4}$$
$$= \frac{1}{\sqrt{|2\pi\mathbf{Q}^i|}}\exp\left[\frac{1}{2}(\boldsymbol{\theta}_t^i - \boldsymbol{\theta}_{t-1}^i)^\top(\mathbf{Q}^i)^{-1}(\boldsymbol{\theta}_t^i - \boldsymbol{\theta}_{t-1}^i)\right], \tag{1.5}$$

while the initial density $p(\boldsymbol{\theta}_1^i|\boldsymbol{\mu}^i, \boldsymbol{\Sigma}^i)$ is given by the Gaussian distribution with mean $\boldsymbol{\mu}^i$ and covariance $\boldsymbol{\Sigma}^i$.

Model fitting and inference

Our goal is to obtain the approximation of the posterior density of the trajectory $\boldsymbol{\theta}_{1:T}$ given the observed neural activity $\mathbf{x}_{0:T}$:

$$p(\boldsymbol{\theta}_{1:T}|\mathbf{x}_{0:T}, \mathbf{w}) = \frac{p(\mathbf{x}_{0:T}|\boldsymbol{\theta}_{1:T})p(\boldsymbol{\theta}_{1:T}|\mathbf{w})}{p(\mathbf{x}_{0:T}|\mathbf{w})}, \tag{1.6}$$

while optimizing the hyperparameters $\mathbf{w}$ under the principle of maximizing marginal likelihood:

$$\begin{aligned} p(\mathbf{x}_{0:T}|\mathbf{w}) &= p(\mathbf{x}_0)\prod_{t=1}^{T} p(\mathbf{x}_t|\mathbf{x}_{0:t-1}, \mathbf{w}) \\ &= p(\mathbf{x}_0)\prod_{t=1}^{T}\prod_{l=1}^{L}\prod_{i=1}^{N}\int p(x_{i,t}^l|\mathbf{x}_{t-1}^l, \boldsymbol{\theta}_t^i)p(\boldsymbol{\theta}_t^i|\mathbf{x}_{0:t-1}^l, \mathbf{w})\, d\boldsymbol{\theta}_t^i. \end{aligned} \tag{1.7}$$

Here, $p(\boldsymbol{\theta}_t^i|\mathbf{x}_{0:t-1}^l, \mathbf{w})$ is the one-step prediction density.

The Expectation-Maximization (EM) algorithm [18] offers a way to construct the approximate posterior with optimized hyperparameters by alternately constructing the approximate posterior density while fixing the hyperparameters (E-step) and optimizing the hyperparameters while fixing the approximate posterior (M-step). The construction of the approximate posterior density at the E-step is performed by sequentially applying Bayes algorithms in a forward and backward manner, where we approximate the posteriors by Gaussian distributions using Laplace's method. Thus, the method yields the mean and variance of the approximated Gaussian posterior at time t, which are denoted as $\boldsymbol{\theta}_{t|T}$ and $\mathbf{W}_{t|T}$, respectively. See Methods and Supplementary Note 1 for the details of the algorithm.

Entropy flow

Using the inferred parameters $\boldsymbol{\theta}_{1:T}$ of the kinetic Ising model from spike data, we estimate entropy flow (also known as bath entropy change) at each time step. The entropy flow at time t is defined as:

$$\sigma_t^{\text{flow}} = \sum_{\mathbf{x}_t, \mathbf{x}_{t-1}} p(\mathbf{x}_t, \mathbf{x}_{t-1}) \log \frac{p(\mathbf{x}_t|\mathbf{x}_{t-1})}{p(\mathbf{x}_{t-1}|\mathbf{x}_t)}, \tag{1.8}$$

where $p(\mathbf{x}_{t-1}|\mathbf{x}_t)$ represents the probability of observing time-reversed processes generated under the forward model. Because we use the natural logarithm, we report entropy flow in units of nats. Eq. 1.8 is related to the entropy production σ_t at time t as follows

[19, 20, 21, 22]:

$$\begin{aligned}\sigma_t &= \sum_{\mathbf{x}_t,\mathbf{x}_{t-1}} p(\mathbf{x}_t,\mathbf{x}_{t-1}) \log \frac{p(\mathbf{x}_t|\mathbf{x}_{t-1})p_{t-1}(\mathbf{x}_{t-1})}{p(\mathbf{x}_{t-1}|\mathbf{x}_t)p_t(\mathbf{x}_t)} \\ &= (S_t - S_{t-1}) + \sigma_t^{\text{flow}}. \end{aligned} \tag{1.9}$$

Here $p_t(\mathbf{x}_t)$ is the marginal probability mass function of the system at time t. S_t is the entropy of the system at time t defined as

$$S_t = -\sum_{\mathbf{x}_t} p_t(\mathbf{x}_t) \log p_t(\mathbf{x}_t). \tag{1.10}$$

The entropy production is non-negative: $\sigma_t \geq 0$. Thus, the positive entropy flow allows a decrease in the system's entropy: namely, the system can be more structured or organized when the entropy flow is positive. Since it is challenging to estimate the system's entropy or its change, here we estimate the entropy flow, which provides the lower bound of the entropy change: $S_t - S_{t-1} \geq -\sigma_t^{\text{flow}}$.

Similarly, since the total entropy production across all time steps $\sigma_{1:T}$ is given as

$$\sigma_{1:T} = \sum_{t=1}^{T} \sigma_t = (S_T - S_0) + \sum_{t=1}^{T} \sigma_t^{\text{flow}}, \tag{1.11}$$

the total entropy flow $\sum_{t=1}^{T} \sigma_t^{\text{flow}}$ provides the lower bound of the system's entropy change from the initial and final time step: $S_T - S_0 \geq -\sum_{t=1}^{T} \sigma_t^{\text{flow}}$. This indicates that the positive total entropy flow enables the systems to be more structured (i.e., lower entropy) at the final time step than at the initial time step.

In this study, we refer to Eq. 1.8 as entropy flow because it is related to heat flow to reservoirs (thermal bath) and the entropy change of the reservoirs in thermodynamics [23]. We note that Eq. 1.8 differs from the entropy flow defined in [24, 25], which was obtained by the decomposition of the dissipation function [22] as an alternative to entropy production. See [22, 26] for their distinct definitions and decompositions for the case of discrete-time systems.

For the case of the kinetic Ising model, the entropy flow is written as

$$\begin{aligned}\sigma_t^{\text{flow}} = &\sum_i \theta_{i,t} \left(E_{\mathbf{x}_t} x_{i,t} - E_{\mathbf{x}_{t-1}} x_{i,t-1} \right) \\ &+ \sum_{i,j} \theta_{ij,t} E_{\mathbf{x}_t,\mathbf{x}_{t-1}} \left(x_{i,t} x_{j,t-1} - x_{i,t-1} x_{j,t} \right) \\ &- \sum_i \left(E_{\mathbf{x}_{t-1}} \psi(\boldsymbol{\theta}_t^i, \mathbf{x}_{t-1}) - E_{\mathbf{x}_t} \psi(\boldsymbol{\theta}_{t-1}^i, \mathbf{x}_t) \right), \end{aligned} \tag{1.12}$$

where $E_{\mathbf{x}_t}$ and $E_{\mathbf{x}_t,\mathbf{x}_{t-1}}$ represents expectation by $p(\mathbf{x}_t)$ and $p(\mathbf{x}_t, \mathbf{x}_{t-1})$, respectively.

Mean-field estimation of entropy flow

Entropy flow (Eq. 1.8) requires the expectation by the joint density $p(\mathbf{x}_t, \mathbf{x}_{t-1})$, which is computationally expensive for large systems. While the mean-field methods for the kinetic Ising model [27, 28, 29, 30, 31, 32] were employed to estimate steady-state entropy flow [32], the mean-field method for estimating time-varying entropy flow remains unexplored. Here, we develop the mean-field method for estimating dynamic entropy flow.

The entropy flow σ_t^{flow} can be decomposed into the forward and reverse components,

$$\sigma_t^{\text{flow}} = -\sigma_t^{\text{forward}} + \sigma_t^{\text{backward}}, \tag{1.13}$$

where $\sigma_t^{\text{forward}}$ and $\sigma_t^{\text{backward}}$ denote the conditional entropies of the forward and time-reversed conditional distributions, respectively. The proposed mean-field method estimates the entropy flow by approximating the forward and time-reversed conditional entropies using the Gaussian integral:

$$\sigma_t^{\text{forward}} \approx \sum_{i=1}^{N} \int \mathcal{D}_z \, \chi(g_{i,t,t-1} + z\sqrt{\Delta_{i,t,t-1}}), \tag{1.14}$$

$$\sigma_t^{\text{backward}} \approx \sum_{i=1}^{N} \int \mathcal{D}_z \, \phi_{i,t}(g_{i,t,t} + z\sqrt{\Delta_{i,t,t}}), \tag{1.15}$$

where $\mathcal{D}_z = \frac{\mathrm{d}z}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right)$. See Methods and Supplementary Note 2 for the derivation of these results. The functions $\chi(h)$ and $\phi_{i,t}(h)$ are given as follows. $\chi(h)$ is entropy of $(0, 1)$ binary random variables with mean $r(h) = 1/(1 + e^{-h})$:

$$\chi(h) = -r(h)h + \psi(h), \tag{1.16}$$

where we redefined the log normalization function ψ as a function of h: $\psi(h) = \log(1+e^h)$. $\phi_{i,t}(h)$ is given by

$$\phi_{i,t}(h) = -m_{i,t-1}h + \psi(h), \tag{1.17}$$

where $m_{i,t-1}$ is the mean-field activation rate of i-th neuron at time $t-1$ (see below for how to obtain it).

Here, the input $h = g_{i,t,s} + z\sqrt{\Delta_{i,t,s}}$ is a Gaussian random variable with mean $g_{i,t,s}$ and variance $\Delta_{i,t,s}$ $(s = t, t-1)$, where z denotes a standardized Gaussian random variable.

$g_{i,t,s}$ and $\Delta_{i,t,s}$ are computed using the mean-field activation rate at time s, $m_{i,s}$, as

$$g_{i,t,s} = \theta_{i,t} + \sum_j \theta_{ij,t} m_{j,s}, \tag{1.18}$$

$$\Delta_{i,t,s} = \sum_j \theta_{ij,t}^2 m_{j,s}(1 - m_{j,s}). \tag{1.19}$$

The mean-field activation rate $m_{i,t}$ can be recursively computed using

$$m_{i,t} \approx \int \mathcal{D}_z \, r(g_{i,t,t-1} + z\sqrt{\Delta_{i,t,t-1}}), \tag{1.20}$$

starting with nominal values of $m_{i,0}$. In this study, we use spiking probability averaged over all time bins and trials for each neuron as $m_{i,0}$.

We also note that under the steady-state assumption, the mean-field approximation can be expressed using the stationary parameters m_i, g_i, and Δ_i as (See Supplementary Note 3):

$$\sigma_t^{\text{flow}} \approx \sum_i \int \mathcal{D}_z \left(r(g_i + z\sqrt{\Delta_i}) - m_i \right) \cdot z\sqrt{\Delta_i}. \tag{1.21}$$

The term $r\left(g_i + z\sqrt{\Delta_i}\right) - m_i$ quantifies how the neuron ' s activity rate deviates from its long-term average, while $z\sqrt{\Delta_i}$ represents the fluctuations of the input it receives. The steady-state mean-field solution thus provides an intuitive view of entropy flow as a measure of a neuron ' s causal responsiveness to input fluctuations – a quantity that captures the correlation underlying Hebbian plasticity in neural systems. However, this equation also clarifies that the approximation depends mainly on the magnitudes of the field and coupling parameters and is thus insensitive to the detailed coupling structure. It should therefore be applied with caution when the degree of coupling asymmetry is the primary determinant of the strength of entropy flow.

Simulation: Estimating the model parameters

We begin by testing the proposed method by estimating the time-dependent parameters of a kinetic Ising model consisting of two simulated neurons (Fig. 1**A**). Figure 1**B** shows the spike data generated using Eq. 1.1 with the number of bins, $T = 400$, and the number of trials, $L = 200$. The time-dependent parameters $\boldsymbol{\theta}_{1:T}$ used to generate binary data were sampled from Gaussian processes (See Methods).

The EM algorithm was applied to this spike data until the log marginal likelihood converged (Fig. 1**C**). Figure 1**D** shows the components of the optimized hyperparameter matrices, $\mathbf{Q}^i (i = 1, 2)$. Figures 1**E** show the MAP estimates of the time-dependent fields

$\theta_{i,t}$ and couplings $\theta_{ij,t}$ under the optimized hyperparameters (solid lines) with 95% credible intervals (shaded areas). The results confirm that the method uncovers the underlying time-dependent parameters (black dashed lines) used to generate the data.

Next, we applied the state-space kinetic Ising model to a network of 12 simulated neurons to estimate the time-varying field and coupling parameters between neurons. Figure 2**A** presents the spike data generated using the observation model with the number of bins set to $T = 75$ and the number of trials $L = 200$. Data generation and model estimation procedures follow the two-neuron case above. Figure 2**B** shows the estimated time-varying coupling parameters $\theta_{ij,t}$ for each neuron. In Fig. 2**C**, we compare the estimated coupling parameters $\boldsymbol{\theta}_{t|T}$ with the true values $\boldsymbol{\theta}_t$ at representative time points ($t = 10, 20, \ldots, 60$). The scatter plot shows agreement between the true and estimated values, with most points aligning closely along the diagonal line, indicating that the model captured the underlying dynamics of the coupling parameters. These results confirm that the proposed state-space kinetic Ising model can reliably estimate time-varying coupling parameters in a network of simulated neurons.

Simulation: Estimation error and computational time

We evaluated the performance of the proposed state-space kinetic Ising model in terms of estimation accuracy and computational time while varying dataset size and population (See Methods for parameter generation).

Estimation error: To assess estimation error, we computed the root mean squared error (RMSE) between the true parameters θ_t and the estimated parameters $\theta_{t|T}$ for both field and coupling parameters. Namely, RMSEs were computed separately for the field parameter $\theta_{i,t}$ and the coupling parameter $\theta_{ij,t}$, then averaged over time bins. The means over 10 independent samplings are shown in Figures 3**A** and **B** with the standard deviations represented by error bars.

For a fixed number of neurons ($N = 80$), RMSEs for both field and coupling parameters decreased as the number of trials L increased (Fig. 3**A**), demonstrating improved estimation accuracy with more data. Conversely, when the number of trials was fixed at $L = 550$, RMSEs exhibited different trends depending on the parameter type. The RMSE for the field parameter increased with N, imposing the challenges of estimating field parameters in larger networks with limited data. The RMSE for coupling parameters remained stable across different neuron numbers in this simulation (Fig. 3**B**).

Computational time: We analyzed the computation time for model fitting. Figure 3**C** illustrates the computation time required to complete the EM algorithm for different numbers of neurons N and trials L. The results indicate that estimation with $N = 80$ and $L = 550$ trials can be completed in approximately one hour, making it feasible for practi-

cal data analysis. Nevertheless, computation time scales with both N and L, highlighting the necessity for further optimization to enable large-scale analysis. The assumption of independent state evolution for individual neurons (Eq. 1.3) significantly reduces computational complexity by enabling independent calculations for filtering, smoothing, and parameter optimization per neuron, which can be further accelerated through parallel updates. Another potential improvement is replacing the current filtering method, which employs exact Newton-Raphson optimization for maximum a posteriori (MAP) estimation, with quasi-Newton or mean-field approximations, as demonstrated in equilibrium state-space Ising models [17].

Simulation: Estimating entropy flow

In this section, we assess the proposed mean-field approximation method for estimating entropy flow. As in the previous section, we generated spike samples from time-dependent parameters $\boldsymbol{\theta}_{1:T}$ sampled from Gaussian processes. All simulations were conducted with $N = 80$, $T = 75$, and $L = 550$ trials. We then estimated the time-dependent field and coupling parameters from the data. Using the posterior mean $\boldsymbol{\theta}_{t|T}$, we obtained the mean-field approximation of the time-dependent entropy flow (Eq. 1.13, using Eqs. 1.14 and 1.15). The solid red line in Fig. 4 represents the entropy flow calculated using the mean-field approximation with the learned parameters.

To verify the consistency of the estimated entropy flow, we calculated the entropy flow using a sampling-based method to compute the expectation over the two-step trajectories (solid black). This approach involves repeatedly running the kinetic Ising model (Eq. 1.1) using the true parameters to sample spike time series data. This process was performed $n_s = 10,000$ times to empirically estimate the joint distribution $p(\mathbf{x}_t, \mathbf{x}_{t-1})$. Using this empirical distribution, we obtained a sample estimate of the entropy flow as follows:

$$\hat{\sigma}_t^{\text{flow}} = \frac{1}{n_s} \sum_{s=1}^{n_s} \log \frac{p\left(\mathbf{x}_t^s | \mathbf{x}_{t-1}^s\right)}{p\left(\mathbf{x}_{t-1}^s | \mathbf{x}_t^s\right)}. \tag{1.22}$$

This sampling estimation using the true parameters serves as the baseline.

The mean-field estimation of the entropy flow (solid red) follows the trajectory of the baseline sampling estimation using the true parameters (solid black). The result confirms that the proposed method is applicable for entropy flow analysis while ensuring computational feasibility. The slight discrepancy between the two lines is due to the errors in estimating the time-dependent parameters and/or the mean-field approximation (in addition to sampling fluctuation inherent to the sampling method). To separate these effects, we estimated the entropy flow by the mean-field approximation using the true parameters

$\boldsymbol{\theta}_{1:T}$ used for the data generation (dashed blue). This estimation deviated from the baseline sampling estimation. In contrast, the sampling method using estimated parameters (dashed green) did not significantly differ from the baseline. Thus, the discrepancy arose from the mean-field approximation, rather than from inaccuracies in parameter estimation. These results suggest that refining the mean-field method could further improve the accuracy of entropy flow estimation.

Chapter 2

Application to mouse V1: Task engagement and entropy flow

Mouse V1 neurons: Experimental design and data description

Having confirmed the applicability of our methods using simulation data, we next applied the state-space kinetic Ising model to empirical data obtained from mice exposed to visual stimuli and estimated its entropy flow.

In this study, we analyzed the Allen Brain Observatory: Visual Behavior Neuropixels dataset provided by the Allen Institute for Brain Science, which contains large-scale recordings of neural spiking activity of mouse brains during the visual change detection task (See [33, 34, 35] for analyses using this data set). The task is designed to analyze the effect of novelty and familiarity of the stimulus on neural responses. One of two image sets (G and H) was presented to animals at the training/habituation and recording sessions with different orders. The G and H image sets contain 8 natural images. We analyzed the recordings of 37 mice available from the Allen dataset, which were exposed to stimulus G in the recording sessions (either day 1 or 2) whereas the same stimulus G was used in the training and habituation sessions prior to the recording sessions (i.e., the case in which G is familiar).

The neural activities were recorded under two distinct conditions, in which the mice were either actively or passively performing the task under the same set of images. The active condition involved the mice performing a go/no-go change detection task, where they earned a water reward upon detection of a change in the visual stimulus, measured by licking behavior. Each of the 8 stimuli was presented for 250 ms, followed by a 500 ms interstimulus interval (gray screen), repeating for one hour while mice actively engaged in the task for reward. In contrast, the passive condition involved replaying the same visual stimuli used in the active condition but without providing any rewards or access to the lick

port. In this study, we analyzed recordings with images labeled im036_r, im012_r, and im115_r, which are used in the training session and classified as Familiar, and compared neural responses under the active and passive conditions. We used all presentations of the images equally and treated one presentation as a trial.

We selected neurons in the V1 area for analysis. For each mouse, we analyzed the simultaneous activities of neurons with a length 750 ms after the onset of an image. While the number of trials varies across mice, the mean trial number was $L = 566$ with 356 and 652 as the minimum and maximum number of trials for the case of an image im036_r.

Mouse V1 neurons: An exemplary result from a single mouse

We constructed binary sequences using a 10 ms bin, which resulted in $T = 75$ time bins. Here, we focused on the analysis of im036_r. Figure 5**A** (Left) shows the probability of a spiking event within a bin averaged over neurons at each time under the active and passive conditions (population-average spike rate) from an exemplary mouse (574078). The overall temporal profiles were similar across the active and passive conditions. In both conditions, the population exhibited higher mean spike rates during the stimulus presentation period (0-250 ms) than the post-stimulus period (250-750 ms). However, their magnitudes significantly differed across the conditions. The passive condition (blue) showed consistently higher spiking probabilities than the active condition (red) throughout the stimulus and post-stimulus periods. In agreement with the population-average spike-rate dynamics, time-averaged spike rates of individual neurons exhibited a sparser distribution during the active condition compared to the passive condition (Fig. 5**A** Right).

We then applied the state-space kinetic Ising model to the binary activities of these neurons. For this goal, we selected the top $N = 80$ neurons with the highest spike rates. The estimated dynamics of the field and coupling parameters exhibited variations in both active and passive conditions (Fig. 5**B**). Notably, the field parameters $\theta_{i,t}$ (the first column) follow the dynamics of the mean spike rate of the population with significant fluctuations. On the contrary, the dynamics of the coupling parameters $\theta_{ij,t}$ exhibited smoother transitions. To clarify the dynamics of the couplings, we show them in the matrix form at specific time points, $t = 5, 25, 35, 50$ (Fig. 5**C**). The neurons are indexed in the ascending order of the average firing rates. The top and bottom rows show the results of the active and passive conditions, respectively. Coupling strength is indicated by graded color, with red and blue representing positive and negative values, respectively. The results show that (i) the couplings exhibit significant variations with positive and negative values; (ii) the variations are stronger in the active condition than in the passive condition; (iii) the diagonal components of the couplings (self-correlations) mostly display

negative correlations.

To corroborate the above observations, we performed the same analysis on the trial-shuffled data (Supplementary Fig. S1). The analysis of trial-shuffled data reveals bias and variance in estimation under the assumption of neuronal independence. The result shows a significant reduction in the magnitude and variability of the couplings, whereas self-couplings remained unchanged (note that the self-coupling remains after trial-shuffling). However, non-zero couplings persisted with stronger variations in the active condition than in the passive condition, reflecting sampling fluctuations due to the lower firing rates in the active condition. These findings indicate that the parameters observed in Fig.5**B,C** include estimation noise, necessitating statistical analyses to confirm their significance.

Mouse V1 neurons: Population analysis across mice

We assessed key features identified in the exemplary mouse (Fig. 5) across all mice by comparing them with trial-shuffled data.

First, the firing rate profiles with reduced activity in the active conditions found in Fig. 5**A** were consistently observed across all mice with a few exceptions (Supplementary Fig. S2 and S3). We compared the average mean activity and sparsity of the distributions under the two conditions for all 37 mice (Supplementary Fig. S4). Sparsity of a non-negative firing rate distribution was quantified by the coefficient of variation (CV) [36]. The V1 neurons exhibited diminished and sparser firing rate distributions in the active condition than in the passive condition, as confirmed by the reduced mean spike rates ($p = 1.556 \times 10^{-8}$, Wilcoxon signed-rank test) and increased CVs ($p = 8.35 \times 10^{-8}$, Wilcoxon signed-rank test).

Next, we assess key statistical features of the estimated parameters of the state-space kinetic Ising model. Figure 6**A-C** illustrates these features for an exemplary mouse (574078). Figure 6**A, B** shows distributions of time-averaged fields $\theta_{i,t}$ and couplings $\theta_{ij,t}$ under the active and passive conditions, while Figure 6**C** shows a scatter plot of time-averaged reciprocal couplings $\theta_{ij,t}$ vs $\theta_{ji,t}$ to evaluate coupling asymmetry. In the active condition, the medians of field parameters decreased, reflecting reduced firing rates, while the medians of couplings remained near zero in both conditions. Field and coupling parameter variances increased, and coupling asymmetry strengthened in the active condition. These trends were consistent across all mice (Fig. 6**D-F**). These characteristics represent key aspects of neural dynamics that are closely related to entropy flow, although they are not entirely independent of each other.

While increased parameter variabilities and coupling asymmetry were observed under the active condition, they may be influenced by the lower neuronal activity. To examine this, we compared results with trial-shuffled data across all mice. Figures 6**G-I** show

field and coupling variances in both conditions, adjusted by subtracting shuffled data values for each mouse. Notably, observed values in both active and passive conditions were significantly higher than shuffled data: $p = 2.91 \times 10^{-11}$ (active), $p = 1.103 \times 10^{-7}$ (passive) for fields, $p = 4.676 \times 10^{-8}$ (active), $p = 1.455 \times 10^{-11}$ (passive) for couplings (Wilcoxon signed-rank test). Note that the observed significant heterogeneity in the field parameters is likely associated with the coupling heterogeneity. These results confirm that the variability observed in active or passive conditions is not explained by noise couplings. The coupling asymmetry was higher than shuffled results only for the active condition ($p = 1.185 \times 10^{-5}$(active) and $p = 0.1287$ (passive) for asymmetry).

Comparisons of these significant changes of the parameter variability (i.e., shuffled results subtracted) between the active and passive conditions showed significantly greater values in the active condition ($p = 8.273 \times 10^{-4}$ for fields, $p = 6.421 \times 10^{-4}$ for couplings, Wilcoxon signed-rank test, Fig.6**G**,**H**), indicating greater variabilities in both field and coupling parameters during active behavior. A similar analysis of the mean couplings across mice revealed slightly but significantly larger values under the active condition (Supplementary Fig. S5). In contrast, coupling asymmetry showed no significant difference ($p = 0.1287$, Wilcoxon signed-rank test, Fig.6**I**). The lack of change in asymmetry in the effective couplings validates the use of the proposed mean-field method for comparing the coupling effect, which primarily arises from variability change. These findings validate enhanced parameter variability in the sparse neuronal activity during active engagement.

Mouse V1 neurons: Entropy flow dynamics

Using the estimated parameters of the state-space kinetic Ising model, we computed entropy flow dynamics. Figure 7**A** shows the time-varying entropy flow of a representative mouse (574078) under the active and passive conditions (red and blue solid lines, respectively). In both cases, transient increases in entropy flow coincided with declines in the mean population spike rate (dashed lines). Similar patterns appeared across all mice analyzed (Supplementary Fig. S6). These increases align with the second law, indicating that greater entropy dissipation is required when the system is transitioning to a lower entropy state, characterized by reduced firing rates.

The entropy flow time courses for this mouse showed no clear differences between the active and passive conditions. To assess population-level effects, we analyzed all 37 mice and computed total entropy flow across time bins for each condition (Fig. 7**B**). The comparison revealed significantly lower total entropy flow in the active condition ($p = 0.01159$, Wilcoxon signed-rank test). Note that neurons exhibited reduced firing rates (Supplementary Fig. S3) and increased parameter variability (Fig. 6**D**,**E**) during the active condition.

To isolate the effect of couplings, we compared the observed total entropy flows with shuffled data results (Fig.7**C**). The estimated entropy flow for shuffled data includes the impact of firing rate dynamics and estimation error on couplings from other neurons; therefore, subtracting shuffling results from observed entropy flow isolates contributions of couplings among different neurons beyond the sampling fluctuation. Positive values of the shuffle-subtracted total entropy flow in both conditions indicate that the couplings caused a significant entropy flow increase ($p = 1.455 \times 10^{-11}$ for active, $p = 1.455 \times 10^{-11}$ for passive, Wilcoxon signed-rank test). These shuffle-subtracted entropy flows behave in agreement with the theoretical prediction by the Sherrington-Kirkpatrick model [37]. In the active condition, the increased coupling variability (and asymmetry) from the shuffle-subtracted values were positively correlated with the shuffle-subtracted entropy flows, while the increased field heterogeneity was negatively correlated (Supplementary Fig. S7**A**-**C**). These effects disappeared in the passive condition, possibly due to small changes in the variabilities and asymmetry introduced by shuffling (Supplementary Fig. S7**D**-**F**).

We analyzed the differences in coupling-related entropy flows between the active and passive conditions for all mice (Fig.7**C**). The result shows no significant difference between the two conditions ($p = 0.1448$, Wilcoxon signed-rank test). However, coupling-related entropy flows of indistinguishable magnitude emerged under distinct neural activity states: sparser, lower activity with increased variability in field and coupling parameters in the active condition; and less sparse, higher activity with reduced variability in the passive condition. Thus, coupled with the previous results, this result indicates that the greater coupling variability in the active condition led to increased total entropy flow, making it comparable to the passive condition despite significantly sparser firing rate distributions. Consistent with this view, a recent study by Aguilera et al. using this dataset complementarily reported that a lower bound on entropy production, derived under steady-state assumptions using a variational framework, was higher in the active condition when normalized per spike [38].

Mouse V1 neurons: Model-based perturbation analysis

To further elucidate the difference in the estimated entropy flow in active and passive conditions, we performed a model-based perturbation analysis by rescaling the fitted model parameters as $\boldsymbol{\theta} \to \beta\boldsymbol{\theta}$ and computing the resulting entropy flow to assess its sensitivity to parameter perturbation. An example result from mouse 574078 (Fig.8**A**) shows that the entropy flows during stimulus presentation and waiting (gray image) periods exhibited distinct behaviors in response to the rescaling. The transient increases in entropy flow caused by firing rate reduction after stimulus onset and offset persisted as the scaling

parameter β increased. In contrast, we observed that the entropy flow peaked at $\beta < 1$ during the waiting period, where the neural activity is relatively stationary.

Both forward and reverse conditional entropies ($\sigma_t^{\text{forward}}$ and $\sigma_t^{\text{backward}}$ in Eq. 1.13) decreased with increasing β during the waiting period (Fig.8**B**,**C**), indicating that both processes became more deterministic. This trend suggests that, as β increases, the system transitions from a disordered phase toward a ferromagnetic phase, rather than into a quasi-chaotic regime [32]. Thus, these results indicate that the subsampled neural population during this period does not operate in a quasi-deterministic chaotic regime, but rather in a subcritical regime.

By subtracting the entropy flow estimated from trial-shuffled data, which preserves only firing rate dynamics, we confirmed that the two bands of increased entropy flow associated with stimulus presentation are attributable to firing rate changes, whereas the increase at $\beta < 1$ arises from interactions, since the former disappeared but the latter persisted after shuffle subtraction (Fig.8**D**). The interaction-driven entropy flow revealed by parameter scaling was stronger during the active condition than the passive condition (Fig.8**D,E**), a result confirmed across all mice (Fig.8**F**). Notably, the previous analysis at $\beta = 1$ showed no difference in shuffle-subtracted (i.e., interaction-driven) entropy flow between the two conditions (Fig.7**C**). Thus, the model-based perturbation analysis uncovered differences in entropy flow between active and passive states that were not apparent at $\beta = 1$.

Mouse V1 neurons: Entropy flow and behavioral performance

Finally, we investigated the relationship between neural dynamics and behavioral performance across individual mice. We quantified task performance by the sensitivity index d' (mean d-prime) defined as the difference between the z-transformed hit and false-alarm rates (Supplementary Note 4). In the following analyses, we extended the analysis to include two additional images (im012_r and im115_r).

First, we examined how sparseness, assessed from individual neurons ' activity rates, relates to behavioral outcomes. As shown in Supplementary Fig. S3, neuronal activities were significantly reduced under active conditions, accompanied by increased sparsity of firing-rate distributions. To further characterize this effect, we examined whether the reduction was uniform across neurons or driven by a subset of neurons by computing the skewness of the firing-rate difference between active and passive conditions (Fig. 9**A**). A uniform reduction results in a skewness of zero, whereas negative skewness indicates that only some neurons decreased their activity, reflecting the sparsification. We found that this sparsification index was significantly correlated with behavioral performance measured by the d-prime, indicating that task engagement is reflected in changes in sparsity quantified

at the level of individual neurons ’ activity rates (Fig. 9**B**).

Having established the link between activity-rate sparsity and behavior, we next turned to entropy flow to ask whether it provides additional explanatory power beyond rate changes alone. The variability of effective couplings was significantly higher during the active condition. To gain insight into the contributions of couplings to entropy flow, we computed the activity rate and mean-field entropy flow of individual neurons as a function of the mean and variability of their inputs (Fig. 9**C,D**). Theoretically, in the low-input and stationary regime, entropy flow increases with both higher mean input and greater variability (Eq. 1.21, background color in Fig. 9**D**). We observed that neurons receiving high mean input tended to have less variable inputs, whereas neurons with low mean input exhibited larger variability (colored circles). These results suggest that total entropy flow is shaped not only by high-input (typically high-firing) neurons but also by low-input neurons with high variability.

These patterns imply two sources of entropy flow: (i) mean-input–driven contributions that track high firing, and (ii) variability-driven contributions that can be substantial even at low firing. To focus more on the latter, we considered entropy flow per activity rate. This normalization reduces the direct dependence on mean rate and makes variability-driven effects, particularly those arising in low-rate neurons, observable on equal footing with high-rate effects. The shift in mean entropy flow per activity rate of individual neurons (active - shuffle) was significantly correlated with behavioral performance (Fig. 9**E**). Moreover, this correlation was weaker and non-significant for trial-shuffled data, indicating contributions from highly variable couplings during active conditions (Fig. 9**F**). This finding suggests that the thermodynamic cost per spiking activity is related to mouse performance, with couplings contributing in addition to activity-rate sparsity.

As an alternative explanation, behavioral performance could be related to entropy-flow changes concentrated in high-firing neurons. We therefore tested whether neurons with higher spike rates tended to increase entropy flow during active engagement in mice with higher task performance (Supplementary Note 5, Supplementary Fig. S8). While this tendency correlated significantly with behavioral performance for one image, it was not significant for the other two images. We therefore infer that high-firing–based changes alone cannot consistently account for performance differences. Instead, the more robust association with entropy flow per activity rate supports a complementary role of variability-driven, coupling-mediated contributions – including those from low-rate neurons – in explaining behavioral performance.

Discussion

This study presents a state-space kinetic Ising model for estimating nonstationary and nonequilibrium neural dynamics, together with a mean-field estimator for time-resolved entropy flow. Within this framework, parameters are inferred sequentially with uncertainty while data-driven smoothness hyperparameters are optimized, allowing heterogeneous temporal variation in both fields and couplings. To our knowledge, time-dependent kinetic Ising inference has not previously been formulated in a sequential Bayesian state-space setting; related Bayesian ideas exist for stationary settings, and most time-varying approaches either keep couplings fixed and vary only fields, or rely on mean-field relations between equal-time and lagged correlations and couplings, which become unstable with limited trial counts typical in animal electrophysiology [39, 29, 31]. Exceptions include point estimates of time-varying couplings and score-driven schemes that rescale all parameters by a single global factor, but these do not capture parameter-wise heterogeneity and often assume equilibrium structure [30, 40, 41, 42]. By contrast, our state-space formulation supports independent temporal variation across parameters and propagates uncertainty, and it yields a practical, time-resolved entropy-flow metric that jointly addresses the intrinsic nonstationarity and nonequilibrium of neural population activity.

Across behavioral conditions in mouse V1, both structure and dynamics of the population activity changed systematically. Structurally, the distributions of fields and couplings differed across conditions; dynamically, the time course and composition of entropy flow shifted. In the active condition the average firing rate was lower, indicating sparser population activity, while the field and coupling distributions were broader, indicating greater heterogeneity. Moreover, entropy flow per spike was positively correlated with behavioral performance. These observations can be situated within several lines of prior work. First, many studies report higher activity during movement or task engagement—for example, enhanced task-related firing in mouse parietal cortex and locomotion-related boosts of visual responses in V1, and even when mean rates drop in higher visual cortex, coding efficiency can improve [43, 44, 45]. The lower mean rate we observe during an active visual discrimination task may appear to diverge from this pattern, but it is consistent with sparse population coding for natural scenes, where only a small subset of neurons responds strongly to a given image and most remain silent [46, 47]. Thus, reduced mean firing alongside intact or improved performance is compatible with efficient, feature-selective coding under naturalistic stimulus statistics.

Second, the broader field and coupling distributions in the active condition indicate increased diversity in neuronal input–output relationships. Theory shows that unobserved (hidden) neurons can skew effective interactions among recorded units, increasing apparent coupling variability, and that internal fluctuations can inflate variability in active

states [48, 49]. Functionally, heterogeneous response patterns—not uniform coactivation —best predict successful detection in mouse V1 [50]. Consistent with these ideas, we find that greater heterogeneity and higher entropy flow per spike covary with better behavioral performance, suggesting that diverse, time-asymmetric causal interactions facilitate information extraction. We did not observe a strong reduction of pairwise correlations during task engagement (Supplementary Fig. S5), a discrepancy that may reflect differences in spatial scale, tasks, or correlation metrics, and that leaves room for additional mechanisms such as inhibitory control or top-down modulation.

Third, neuromodulatory systems offer a plausible physiological substrate for these state changes. Acetylcholine reorganizes neocortical circuits into sparser, more selective, and temporally extended activity modes by reducing the probability of population events and altering temporal recruitment relative to thalamic input [51, 52]. Norepinephrine adjusts cortical gain and promotes sustained active states associated with arousal and attention, contributing to transitions between inattentive and attentive regimes [53, 54, 55]. The sparser activity, increased heterogeneity, and elevated entropy flow we observe during task engagement are qualitatively consistent with cholinergic and noradrenergic modulation, which shift cortex from synchronized, low-information regimes toward desynchronized, information-rich processing.

Fourth, our results connect microcircuit dynamics to macroscopic nonequilibrium observations. At large scales, the resting human brain lies near detailed balance, whereas cognitively demanding tasks break detailed balance and increase entropy production; reduced-consciousness states are correspondingly more reversible [7, 8, 9]. Violations of fluctuation–dissipation relations and arrow-of-time analyses further show greater temporal asymmetry during tasks and reveal cortical hierarchies of nonequilibrium dynamics [56, 57, 58, 59, 60, 61]. By directly estimating time-resolved entropy flow from nonstationary spike trains, we bridge these macroscopic findings with cellular-scale causal interactions and show that entropy flow per spike relates to behavioral performance, suggesting that breaking detailed balance is functionally relevant for computation rather than an epiphenomenon.

Finally, our perspective complements equilibrium energy-landscape approaches based on pairwise maximum-entropy models, which have captured correlation structure, attractor basins, and transition barriers of large-scale brain activity under equilibrium assumptions [62, 63, 64]. While energy landscapes describe the repertoire and stability of accessible states, they do not quantify directional probability fluxes that signal nonequilibrium drive. Our analysis targets this irreversible component by quantifying entropy flow in nonstationary binary spike data. Heuristically, if the energy landscape provides the map of hills and valleys, entropy flow measures the wind that pushes the system along

preferred directions when the brain computes. The stronger directed flux during task engagement indicates greater dissipation and irreversible information processing—features invisible to equilibrium models but central to understanding real cortical computation.

In methodological terms, most time-dependent kinetic Ising schemes relate correlations to couplings via mean-field closures, often allowing only fields to vary in time; point-estimate and score-driven variants relax this to a degree but still limit parameter-wise heterogeneity and robustness with few trials [29, 31, 30, 40]. Our state-space, sequential Bayesian approach overcomes these constraints by allowing independent temporal variation of both fields and couplings, optimizing smoothness hyperparameters from data, and propagating uncertainty through time, yielding a practical estimator of entropy flow that treats nonstationarity and nonequilibrium in a unified fashion [39]. Conceptually, this provides a quantitative link between time-asymmetric causal structure at the spike level and behaviorally relevant computation in naturalistic conditions.

The kinetic Ising–based framework entails theoretical limitations. Although analytically tractable, it assumes pairwise couplings and conditional independence, which simplify neural dynamics but constrain interpretability. Future work should develop principled estimators of entropy flow under non-Markovian conditions while maintaining clarity about the limits of inference.

In summary, by developing a state-space kinetic Ising model that accounts for both nonstationary and nonequilibrium properties, we have demonstrated how task engagement modulates neuronal firing activity and coupling diversity. Our approach incorporates time-varying entropy flow estimation, revealing that time-asymmetric, irreversible activity emerges within sparsely active populations during task engagement—an effect correlated with the mouse ' s behavioral performance. These findings underscore the utility of our approach, offering new insights into the thermodynamic underpinnings of neural computation.

Methods

Estimating time-varying parameters of the kinetic Ising model

We summarize the expectation-maximization algorithm for estimating the state-space kinetic Ising model with optimized hyperparameters. See Supplementary Note 1 for more details.

E-step: Given the hyperparameters $\mathbf{w}$, we obtain the estimate of the state $\boldsymbol{\theta}_t$ given all the data available. When estimating the parameters $\boldsymbol{\theta}_t^i$ ($t = 0, 1, \ldots, T$, $i = 1, \ldots, N$) from the spike data $\mathbf{x}_t$ ($t = 0, 1, \ldots, T$), we first obtain the filter density by the sequentially

applying the Bayes theorem:

$$p(\boldsymbol{\theta}_t|\mathbf{x}_{0:t},\mathbf{w}) = \frac{p(\mathbf{x}_t|\boldsymbol{\theta}_t,\mathbf{x}_{0:t-1},\mathbf{w})p(\boldsymbol{\theta}_t|\mathbf{x}_{0:t-1},\mathbf{w})}{p(\mathbf{x}_t|\mathbf{x}_{0:t-1},\mathbf{w})}. \tag{2.1}$$

Here, the one-step prediction density is computed using the Chapman-Kolmogorov equation:

$$p(\boldsymbol{\theta}_t|\mathbf{x}_{0:t-1},\mathbf{w}) = \prod_{i=1}^{N}\int p(\boldsymbol{\theta}_t^i|\boldsymbol{\theta}_{t-1}^i,\mathbf{Q}^i)p(\boldsymbol{\theta}_{t-1}^i|\mathbf{x}_{t-1})d\boldsymbol{\theta}_{t-1}^i. \tag{2.2}$$

By assuming that the filter density for the i-th neuron at the previous time step $t-1$ is given by the Gaussian distribution with mean $\boldsymbol{\theta}_{t-1|t-1}^i$ and covariance $\mathbf{W}_{t-1|t-1}^i$, the one-step prediction density becomes the Gaussian distribution whose mean $\boldsymbol{\theta}_{t|t-1}^i$ and covariance $\mathbf{W}_{t|t-1}^i$ are given by

$$\boldsymbol{\theta}_{t|t-1}^i = \boldsymbol{\theta}_{t-1|t-1}^i, \tag{2.3}$$

$$\mathbf{W}_{t|t-1}^i = \mathbf{W}_{t-1|t-1}^i + \mathbf{Q}^i, \tag{2.4}$$

with $\boldsymbol{\theta}_{1|0}^i = \boldsymbol{\mu}^i$ and $\mathbf{W}_{1|0}^i = \boldsymbol{\Sigma}^i$ being the hyperparameters of the initial Gaussian distribution, $p(\boldsymbol{\theta}_1^i|\boldsymbol{\mu}^i,\boldsymbol{\Sigma}^i)$. Then, the filter density is given as

$$\begin{aligned} p(\boldsymbol{\theta}_t|\mathbf{x}_{0:t},\mathbf{w}) &= \prod_{i=1}^{N} p(\boldsymbol{\theta}_t^i|\mathbf{x}_{0:t},\mathbf{w}) \\ &\propto \prod_{i=1}^{N}\prod_{l=1}^{L}\exp\left[\theta_{i,t}x_{i,t}^l + \sum_{j=1}^{N}\theta_{ij,t}x_{it}^l x_{j,t-1}^l - \psi(\boldsymbol{\theta}_t^i,\mathbf{x}_{t-1}^l)\right] \\ &\quad\cdot\prod_{i=1}^{N}\exp\left[-\frac{1}{2}(\boldsymbol{\theta}_t^i-\boldsymbol{\theta}_{t|t-1}^i)^\top(\mathbf{W}_{t|t-1}^i)^{-1}(\boldsymbol{\theta}_t^i-\boldsymbol{\theta}_{t|t-1}^i)\right]. \end{aligned} \tag{2.5}$$

Since this filter density is a concave function with respect to $\boldsymbol{\theta}_t^i$ for each neuron, we apply the Laplace approximation independently to the filter densities of individual neurons and obtain the approximate Gaussian distributions, where the mean is approximated by the MAP estimate:

$$\boldsymbol{\theta}_{t|t}^i = \arg\max_{\boldsymbol{\theta}_t^i}\log p(\boldsymbol{\theta}_t^i|\mathbf{x}_{0:t},\mathbf{w}), \tag{2.6}$$

for $i = 1, \ldots, N$, while the covariance is approximated using the Hessian as

$$\begin{aligned}\mathbf{W}^i_{t|t} &= \left[-\frac{\partial^2 \log p(\boldsymbol{\theta}^i_t | \mathbf{x}_{0:t}, \mathbf{w})}{\partial \boldsymbol{\theta}^i_t \partial \left(\boldsymbol{\theta}^i_t\right)^\mathsf{T}} \bigg|_{\boldsymbol{\theta}^i_t = \boldsymbol{\theta}^i_{t|t}} \right]^{-1} \\ &= \left[\mathbf{G}(\boldsymbol{\theta}^i_{t|t}) + \left(\mathbf{W}^i_{t|t-1}\right)^{-1} \right]^{-1}, \end{aligned} \tag{2.7}$$

where $\mathbf{G}(\boldsymbol{\theta}^i_t) \equiv \sum_{l=1}^{L} \frac{\partial^2 \psi(\boldsymbol{\theta}^i_t, \mathbf{x}^l_{t-1})}{\partial \boldsymbol{\theta}^i_t \partial \left(\boldsymbol{\theta}^i_t\right)^\mathsf{T}} \Big|_{\boldsymbol{\theta}^i_t = \boldsymbol{\theta}^i_{t|t}}$ is the Fisher information matrix with respect to $\boldsymbol{\theta}^i_t$ computed for the kinetic Ising model over the trials. We computed the MAP estimate by the Newton-Raphson method utilizing the Hessian evaluated at a search point.

Next, we obtain the smoother density by recursively applying the formula. Because the filter density and state transitions are approximated by normal distributions, we follow the fixed-interval smoothing algorithm developed for the Gaussian distributions [65]. In this method, the smoothed mean and covariance are recursively obtained by the following equations:

$$\boldsymbol{\theta}^i_{t-1|T} = \boldsymbol{\theta}^i_{t-1|t-1} + \mathbf{A}^i_{t-1} \left(\boldsymbol{\theta}^i_{t|T} - \boldsymbol{\theta}^i_{t|t} \right), \tag{2.8}$$

$$\mathbf{W}^i_{t-1|T} = \mathbf{W}^i_{t-1|t-1} + \mathbf{A}^i_{t-1} \left(\mathbf{W}^i_{t|T} - \mathbf{W}^i_{t|t} \right) \mathbf{A}^{i\top}_{t-1}, \tag{2.9}$$

$$\mathbf{A}^i_{t-1} = \mathbf{W}^i_{t-1|t-1} \left(\mathbf{W}^i_{t|t-1} \right)^{-1}, \tag{2.10}$$

for $t = T, T-1, \ldots, 2$.

M-step: We optimize the hyperparameters given the smoothed posteriors. To optimize the hyperparameter $\mathbf{Q}^i$, we used the following update formula that maximizes the lower bound of the log marginal likelihood:

$$\begin{aligned}\mathbf{Q}^i = \frac{1}{T-1} \sum_{t=2}^{T} \big[&(\boldsymbol{\theta}^i_{t|T} - \boldsymbol{\theta}^i_{t-1|T})(\boldsymbol{\theta}^i_{t|T} - \boldsymbol{\theta}^i_{t-1|T})^\top \\ &+ \mathbf{W}^i_{t|T} - \mathbf{W}^i_{t-1,t|T} - \mathbf{W}^i_{t,t-1|T} + \mathbf{W}^i_{t-1|T} \big]. \end{aligned} \tag{2.11}$$

We compute the lag-one smoothing covariance matrix $\mathbf{W}^i_{t,t-1|T}$ following the method of De Jong and Mackinnon [66]: $\mathbf{W}^i_{t,t-1|T} = \mathbf{W}^i_{t|t}(\mathbf{W}^i_{t+1|t})^{-1}\mathbf{W}^i_{t|T}$. We also note that the optimization of a diagonal of the form $\mathbf{Q}^i = \mathrm{diag}[\lambda^i_0, \ldots, \lambda^i_N]$ or $\mathbf{Q}^i = \lambda^i \mathbf{I}$ can be performed by taking diagonal and trace of the r.h.s of the equation above, respectively.

Similarly, we update $\boldsymbol{\Sigma}^i$ according to

$$\boldsymbol{\Sigma}^i = \mathbf{W}^i_{1|T} + (\boldsymbol{\theta}^i_{1|T} - \boldsymbol{\mu})(\boldsymbol{\theta}^i_{1|T} - \boldsymbol{\mu})^\top. \tag{2.12}$$

The convergence of the EM algorithm is assessed by computing the approximate log

marginal likelihood function (Eq.1.7) using the Laplace approximation. Using the mean and covariance of the filter and one-step prediction densities, the approximate log marginal likelihood function for the hyperparameters $\mathbf{w}$ is obtained as

$$\begin{aligned}\log p(\mathbf{x}_{0:T}|\mathbf{w}) &= \log p(\mathbf{x}_0) \\ &+ \sum_{t=1}^{T}\sum_{i=1}^{N}\left[\frac{1}{2}\log|\mathbf{W}^i_{t|t}| - \frac{1}{2}\log|\mathbf{W}^i_{t|t-1}| + q(\boldsymbol{\theta}^i_{t|t})\right].\end{aligned} \tag{2.13}$$

See Supplementary Note 1 for the derivations.

Mean-field approximation of the entropy flow

Here we extend the mean-field approximation method developed for the steady-state kinetic Ising model [32] to make it applicable to nonstationary systems.

First, σ_t^{flow} can be decomposed as follows by introducing the forward and backward conditional entropies:

$$\sigma_t^{\text{flow}} = -\sigma_t^{\text{forward}} + \sigma_t^{\text{backward}}, \tag{2.14}$$

where

$$\sigma_t^{\text{forward}} = -\sum_{\mathbf{x}_t,\mathbf{x}_{t-1}} p(\mathbf{x}_t,\mathbf{x}_{t-1})\log p(\mathbf{x}_t|\mathbf{x}_{t-1}), \tag{2.15}$$

$$\sigma_t^{\text{backward}} = -\sum_{\mathbf{x}_t,\mathbf{x}_{t-1}} p(\mathbf{x}_t,\mathbf{x}_{t-1})\log p(\mathbf{x}_{t-1}|\mathbf{x}_t). \tag{2.16}$$

We calculate these conditional entropies using the Gaussian approximation as follows.

We begin with approximating the forward conditional entropy as

$$\begin{aligned}\sigma_t^{\text{forward}} &= -\sum_{\mathbf{x}_t,\mathbf{x}_{t-1}} p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1})\log p(\mathbf{x}_t|\mathbf{x}_{t-1}) \\ &\simeq -\sum_{\mathbf{x}_{t-1}} Q(\mathbf{x}_{t-1})\sum_{\mathbf{x}_t} p(\mathbf{x}_t|\mathbf{x}_{t-1})\log p(\mathbf{x}_t|\mathbf{x}_{t-1}).\end{aligned} \tag{2.17}$$

Here we replaced $p(\mathbf{x}_{t-1})$ with an independent model $Q(\mathbf{x}_{t-1})$ defined as

$$Q(\mathbf{x}_{t-1}) = \prod_i Q(x_{i,t-1}). \tag{2.18}$$

The conditional probability is written as

$$p(\mathbf{x}_t|\mathbf{x}_{t-1}) = \prod_i p(x_{i,t}|\mathbf{x}_{t-1}), \tag{2.19}$$

where

$$p(x_{i,t}|\mathbf{x}_{t-1}) = e^{x_{i,t}h_{i,t}(\mathbf{x}_{t-1})-\psi(h_{i,t}(\mathbf{x}_t))} \tag{2.20}$$

with

$$h_{i,t}(\mathbf{x}_{t-1}) = \theta_{i,t} + \sum_j \theta_{ij,t}x_{j,t-1}. \tag{2.21}$$

Here, we redefined the log normalization function ψ as a function of $h_{i,t}(\mathbf{x}_t)$: $\psi(h_{i,t}(\mathbf{x}_t)) = \log(1 + e^{h_{i,t}(\mathbf{x}_t)})$.

Note that the expectation of $x_{i,t}$ is given by

$$\begin{aligned} r(h_{i,t}(\mathbf{x}_{t-1})) &= \sum_{x_{i,t}} x_{i,t}p(x_{i,t}|\mathbf{x}_{t-1}) \\ &= \frac{1}{1 + e^{-h_{i,t}(\mathbf{x}_{t-1})}}. \end{aligned} \tag{2.22}$$

Using $r(h_{i,t}(\mathbf{x}_{t-1}))$, we have

$$p(x_{i,t} = 1|\mathbf{x}_{t-1}) = r(h_{i,t}(\mathbf{x}_{t-1})), \tag{2.23}$$

$$p(x_{i,t} = 0|\mathbf{x}_{t-1}) = 1 - r(h_{i,t}(\mathbf{x}_{t-1})). \tag{2.24}$$

Then the forward conditional entropy becomes

$$\begin{aligned} \sigma_t^{\text{forward}} &\simeq -\sum_{\mathbf{x}_{t-1}} Q(\mathbf{x}_{t-1}) \\ &\quad \cdot \sum_i \sum_{x_{i,t}} p(x_{i,t}|\mathbf{x}_{t-1}) \log p(x_{i,t}|\mathbf{x}_{t-1}) \\ &= \sum_i \sum_{\mathbf{x}_{t-1}} Q(\mathbf{x}_{t-1})\chi\left(h_{i,t}(\mathbf{x}_{t-1})\right), \end{aligned} \tag{2.25}$$

where

$$\begin{aligned}\chi\left(h_{i,t}(\mathbf{x}_{t-1})\right) &\equiv -\sum_{x_{i,t}} p(x_{i,t}|\mathbf{x}_{t-1}) \log p(x_{i,t}|\mathbf{x}_{t-1}) \\ &= -r(h_{i,t}(\mathbf{x}_{t-1})) \log r(h_{i,t}(\mathbf{x}_{t-1})) \\ &\quad - (1 - r(h_{i,t}(\mathbf{x}_{t-1}))) \log(1 - r(h_{i,t}(\mathbf{x}_{t-1}))) \\ &= -[r(h_{i,t}(\mathbf{x}_{t-1}))h_{i,t}(\mathbf{x}_{t-1}) - \psi(h_{i,t}(\mathbf{x}_t))] \end{aligned} \tag{2.26}$$

We approximate Eq. 2.25 by a Gaussian distribution based on the central limit theorem for a collection of independent binary signals. Specifically, by using $\mathcal{D}_z = \frac{\mathrm{d}z}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right)$, the forward conditional entropy is approximated as

$$\sigma_t^{\text{forward}} \approx \sum_i \int \mathcal{D}_z \, \chi\left(g_{i,t,t-1} + z\sqrt{\Delta_{i,t,t-1}}\right), \tag{2.27}$$

where $g_{i,t,t-1}$ and $\Delta_{i,t,t-1}$ are the mean and variance of $h_{i,t}(\mathbf{x}_{t-1})$ given by

$$g_{i,t,t-1} = \theta_{i,t} + \sum_j \theta_{ij,t} m_{j,t-1}, \tag{2.28}$$

$$\Delta_{i,t,t-1} = \sum_j \theta_{ij,t}^2 m_{j,t-1}(1 - m_{j,t-1}). \tag{2.29}$$

Here, $m_{i,t}$ is the mean-field approximation of $x_{i,t}$ obtained by the Gaussian approximation method assuming independent activity of neurons at $t-1$:

$$\begin{aligned} m_{i,t} &= \sum_{\mathbf{x}_t, \mathbf{x}_{t-1}} x_{i,t} \, p(\mathbf{x}_t|\mathbf{x}_{t-1}) p(\mathbf{x}_{t-1}) \\ &= \sum_{\mathbf{x}_{t-1}} Q(\mathbf{x}_{t-1}) r(h_{i,t}(\mathbf{x}_{t-1})). \end{aligned} \tag{2.30}$$

Applying the Gaussian approximation to $h_{i,t}(\mathbf{x}_{t-1})$, $m_{i,t}$ is recursively computed as

$$m_{i,t} \approx \int \mathcal{D}_z \, r\left(g_{i,t,t-1} + z\sqrt{\Delta_{i,t,t-1}}\right), \tag{2.31}$$

for $t = 1, \ldots, T$, using Eqs. 2.28 and 2.29, which are functions of $m_{i,t-1}$. Here $m_{i,1}$ was computed using nominal values of $m_{i,0}$ $(i = 1, \ldots, N)$. In the simulation and empirical analyses, we used spiking probability averaged over all time steps and trials for each neuron as $m_{i,0}$.

Next, we approximate $\sigma_t^{\text{backward}}$. It is computed as

$$\begin{aligned}
\sigma_t^{\text{backward}} &= -\sum_{\mathbf{x}_t,\mathbf{x}_{t-1}} p(\mathbf{x}_t,\mathbf{x}_{t-1}) \log p(\mathbf{x}_{t-1}|\mathbf{x}_t) \\
&= -\sum_{\mathbf{x}_t,\mathbf{x}_{t-1}} p(\mathbf{x}_t|\mathbf{x}_{t-1}) \sum_{\mathbf{x}_{t-2}} p(\mathbf{x}_{t-1}|\mathbf{x}_{t-2}) p(\mathbf{x}_{t-2}) \log p(\mathbf{x}_{t-1}|\mathbf{x}_t) \\
&= -\sum_{\mathbf{x}_{t-2}} \sum_{\mathbf{x}_{t-1}} p(\mathbf{x}_{t-1}|\mathbf{x}_{t-2}) p(\mathbf{x}_{t-2}) \\
&\quad \cdot \sum_{\mathbf{x}_t} p(\mathbf{x}_t|\mathbf{x}_{t-1}) \sum_i \left[x_{i,t-1} h_{i,t}(\mathbf{x}_t) - \psi(h_{i,t}(\mathbf{x}_t)) \right].
\end{aligned} \tag{2.32}$$

We approximate the following probabilities by independent distributions:

$$p(\mathbf{x}_{t-2}) = Q(\mathbf{x}_{t-2}), \tag{2.33}$$

$$p(\mathbf{x}_t|\mathbf{x}_{t-1}) = Q(\mathbf{x}_t). \tag{2.34}$$

Using them, $\sigma_t^{\text{backward}}$ can be approximated as

$$\begin{aligned}
\sigma_t^{\text{backward}} &\simeq -\sum_{\mathbf{x}_{t-2}} \sum_{\mathbf{x}_{t-1}} p(\mathbf{x}_{t-1}|\mathbf{x}_{t-2}) Q(\mathbf{x}_{t-2}) \\
&\quad \cdot \sum_{\mathbf{x}_t} Q(\mathbf{x}_t) \sum_i \left[x_{i,t-1} h_{i,t}(\mathbf{x}_t) - \psi(h_{i,t}(\mathbf{x}_t)) \right] \\
&= -\sum_{\mathbf{x}_t} Q(\mathbf{x}_t) \sum_{\mathbf{x}_{t-2}} Q(\mathbf{x}_{t-2}) \\
&\quad \cdot \sum_i \left[r(h_{i,t-1}(\mathbf{x}_{t-2})) h_{i,t}(\mathbf{x}_t) - \psi(h_{i,t}(\mathbf{x}_t)) \right] \\
&= -\sum_i \sum_{\mathbf{x}_t} Q(\mathbf{x}_t) \left[m_{i,t-1} h_{i,t}(\mathbf{x}_t) - \psi(h_{i,t}(\mathbf{x}_t)) \right],
\end{aligned} \tag{2.35}$$

where we used Eq. 2.22 to obtain the second equality and Eq. 2.30 to obtain the last result. By defining

$$\phi_{i,t}(h_{i,t}(\mathbf{x}_t)) = -[m_{i,t-1} h_{i,t}(\mathbf{x}_t) - \psi(h_{i,t}(\mathbf{x}_t))], \tag{2.36}$$

the backward conditional entropy is obtained by the Gaussian integral:

$$\begin{aligned}
\sigma_t^{\text{backward}} &= \sum_i \sum_{\mathbf{x}_t} Q(\mathbf{x}_t) \phi_{i,t}(h_{i,t}(\mathbf{x}_t)) \\
&\approx \sum_i \int \mathcal{D}_z \, \phi_{i,t} \left(g_{i,t,t} + z\sqrt{\Delta_{i,t,t}} \right),
\end{aligned} \tag{2.37}$$

where

$$g_{i,t,t} \equiv \theta_{i,t} + \sum_j \theta_{ij,t} m_{j,t}, \tag{2.38}$$

$$\Delta_{i,t,t} = \sum_j \theta_{ij,t}^2 m_{j,t}(1 - m_{j,t}). \tag{2.39}$$

An alternative approach to obtain the backward conditional entropy is given in Supplementary Note 2.

Thus, the entropy flow is obtained as

$$\begin{aligned} \sigma_t^{\text{flow}} &= -\sigma_t^{\text{forward}} + \sigma_t^{\text{backward}} \\ &\approx \sum_i \int \mathcal{D}_z \Bigg[-\chi\left(g_{i,t,t-1} + z\sqrt{\Delta_{i,t,t-1}}\right) \\ &\quad + \phi_{i,t}\left(g_{i,t,t} + z\sqrt{\Delta_{i,t,t}}\right) \Bigg], \end{aligned} \tag{2.40}$$

which allows us to examine the contributions of each neuron to the total entropy flow.

See also Supplementary Note 3 for the analytical expression of the entropy flow under steady-state conditions or for independent neurons.

Generation of field and coupling parameters for simulation studies

We constructed time-varying field and coupling parameters, from which we generated the binary data. To ensure smooth temporal variations, each coupling parameter $\theta_{ij,t}$ was sampled from a Gaussian process of size T with mean μ and covariance matrix defined by the squared exponential kernel

$$k(t,s) = k_0 \exp\left(-\frac{(t-s)^2}{2\tau^2}\right). \tag{2.41}$$

For the analysis of estimation error and computational time using different system sizes (Fig. 3), we used the scaling mean $\mu = 5/N$ and variance $k_0 = 10/N$, following the convention of the Sherrington-Kirkpatrick model. The characteristic length-scale was specified by $\tau = 30/\sqrt{N}$. Similarly, the external field parameters $\theta_{i,t}$ were independently sampled from the Gaussian process, using $\mu = -3$, $\tau = 50$, and $k_0 = 1$.

To obtain trajectories for the different system sizes, a single set of random values was generated for the maximum system size, and subsets of these values were used to examine the system size N. Specifically, for the coupling parameters, a global three-dimensional array was created with dimensions corresponding to the maximum number

of neurons, time steps, and coupling connections. Similarly, for the field parameters, a two-dimensional array was generated, with dimensions corresponding to the maximum number of time steps and neurons. For a given neuron count N, the relevant subset of values was extracted from these precomputed arrays, ensuring that each N used a subset of the values assigned to larger N. This hierarchical structure ensured that the seed for $N = 80$ encompassed all values used for smaller N, maintaining consistency across different system sizes. We evaluated the model's performance using this data set and repeated the procedure 10 times.

Data availability

We used the publicly available Allen Brain Observatory: Visual Behavior Neuropixels dataset provided by the Allen Institute for Brain Science:
https://portal.brain-map.org/circuits-behavior/visual-behavior-neuropixels.
Large precomputed datasets required to reproduce the figures are archived on Zenodo:
doi:10.5281/zenodo.15220108.

Code availability

The analysis code used in this study is archived on Zenodo and linked to the GitHub repository:
doi:10.5281/zenodo.17504162.
For convenient browsing, see the GitHub mirror:
https://github.com/KenIshihara-17171ken/Non_equ.

Competing Interests

The authors declare no competing interests.

Supplementary Information

2.1 State-space kinetic Ising model

In this Supplementary Note, we provide the filtering and smoothing algorithms for the time-varying kinetic Ising model and an optimization method of its hyperparameters via the Expectation-Maximization algorithm.

2.1.1 Model

Let $x_{i,t} = \{0, 1\}$ be an outcome of a binary random variable of neuron i at time t ($i = 1, \ldots, N$, $t = 0, \ldots, T$). In the kinetic Ising model, the activation of neuron i at time t independently depends on the activities of the neurons in the previous time step $t - 1$. The conditional probability mass function of $x_{i,t}$ is given as

$$p(x_{i,t}|x_{1,t-1}, \ldots, x_{N,t-1}, \boldsymbol{\theta}_t^i) = \frac{\exp\left[\theta_{i,t}x_{i,t} + \sum_{j=1}^N \theta_{ij,t}x_{i,t}x_{j,t-1}\right]}{1 + \exp\left[\theta_{i,t} + \sum_{j=1}^N \theta_{ij,t}x_{j,t-1}\right]}, \tag{S2.1.1}$$

where $\theta_{i,t}$ is a time-dependent field parameter that determines the bias for inputs to the i-th neuron at time t, and $\theta_{ij,t}$ is a time-dependent coupling parameter from the j-th neuron to the i-th neuron at time t. These parameters are collectively denoted as $\boldsymbol{\theta}_t^i = (\theta_{i,t}, \theta_{i1,t}, \ldots \theta_{ij,t}, \ldots \theta_{iN,t})$. Using the log normalization function,

$$\psi(\boldsymbol{\theta}_t^i, \mathbf{x}_{t-1}^l) = \log\left[1 + \exp\left[\theta_{i,t} + \sum_{j=1}^N \theta_{ij,t}x_{j,t-1}^l\right]\right], \tag{S2.1.2}$$

the kinetic Ising model is also written as

$$p(x_{i,t}|x_{1,t-1}, \ldots, x_{N,t-1}, \boldsymbol{\theta}_t^i) = \exp\left[\theta_{i,t}x_{i,t} + \sum_{j=1}^N \theta_{ij,t}x_{i,t}x_{j,t-1} - \psi(\boldsymbol{\theta}_t^i, \mathbf{x}_{t-1})\right]. \tag{S2.1.3}$$

Assuming conditional independence, the joint probability mass function that determines the probabilities of generating patterns of activity across N neurons is given by

$$\prod_{i=1}^N p(x_{i,t}|x_{1,t-1}, \ldots, x_{N,t-1}, \boldsymbol{\theta}_t^i). \tag{S2.1.4}$$

Typical neurophysiological experiments repeat multiple trials of measurement under the same experimental conditions. We let $x_{i,t}^l = \{0, 1\}$ be a binary variable of the i-th neuron at time t in the l-th trial ($i = 1, \ldots, N$, $t = 0, \ldots, T$, $l = 1, \ldots, L$). We collectively denote the binary patterns of simultaneously recorded neurons at time t in the l-th trial using a vector, $\mathbf{x}_t^l = (x_{1,t}^l, \ldots, x_{N,t}^l)$. Further, we denote the patterns at

time t from all trials by $\mathbf{x}_t = (\mathbf{x}_t^1, \ldots, \mathbf{x}_t^l, \ldots, \mathbf{x}_t^L)$ and denote all the patterns up to time t by $\mathbf{x}_{0:t}$. We use the same convention for the time-varying parameters, denoting them as $\boldsymbol{\theta}_t = (\boldsymbol{\theta}_t^1, \ldots, \boldsymbol{\theta}_t^i, \ldots, \boldsymbol{\theta}_t^N)$ and $\boldsymbol{\theta}_{1:t}$ for their trajectories over time.

Given the time-varying parameters $\boldsymbol{\theta}_{1:T}$, the probability mass function observing binary sequences $\mathbf{x}_{0:T}$ is given as

$$p(\mathbf{x}_{0:T}|\boldsymbol{\theta}_{1:T}) = \prod_{l=1}^{L}\prod_{i=1}^{N}\left[p(x_{i,0}^l)\prod_{t=1}^{T}p(x_{i,t}^l|\mathbf{x}_{t-1}^l, \boldsymbol{\theta}_t^i)\right], \tag{S2.1.5}$$

where we use $p(x_{i,0}^l) = 0.5$ for data generation. We assume that the same time-dependent parameters apply across trials.

In the state-space model, the state model defines the discrete-time stochastic processes of the latent variables, which are the time-varying parameters $\boldsymbol{\theta}_{0:T}$ in our model. We use the following Gaussian model by assuming independent processes across neurons:

$$p(\boldsymbol{\theta}_{0:T}) = \prod_{i=1}^{N}\left[p(\boldsymbol{\theta}_0^i|\boldsymbol{\mu}^i, \boldsymbol{\Sigma}^i)\prod_{t=1}^{T}p(\boldsymbol{\theta}_t^i|\boldsymbol{\theta}_{t-1}^i, Q^i)\right], \tag{S2.1.6}$$

where the transition of the i-th neuron is given by

$$p(\boldsymbol{\theta}_t^i|\boldsymbol{\theta}_{t-1}^i, \mathbf{Q}^i) = \frac{1}{\sqrt{|2\pi\mathbf{Q}^i|}}\exp\left[-\frac{1}{2}(\boldsymbol{\theta}_t^i - \boldsymbol{\theta}_{t-1}^i)^\top(\mathbf{Q}^i)^{-1}(\boldsymbol{\theta}_t^i - \boldsymbol{\theta}_{t-1}^i)\right] \tag{S2.1.7}$$

with $\mathbf{Q}^i$ being the noise covariance for the transition of the i-th neuron. The initial density of the i-th neuron $p(\boldsymbol{\theta}_0^i|\boldsymbol{\mu}^i, \boldsymbol{\Sigma}^i)$ is given as a Gaussian distribution with mean $\boldsymbol{\mu}^i$ and covariance $\boldsymbol{\Sigma}^i$. In practice, we used a zero vector and a unit matrix before optimization, respectively. In the followings, we denote a set of hyperparameters $\boldsymbol{\mu}^i, \boldsymbol{\Sigma}^i, \mathbf{Q}^i$ for $i = 1, \ldots, N$ collectively by $\mathbf{w}$.

2.1.2 One-step prediction density

In this section, we derive the one-step prediction density $p(\boldsymbol{\theta}_t|\mathbf{x}_{0:t-1}, \mathbf{w})$, using Chapman–Kolmogorov's equation.

For $t = 1$, we note that the one-step prediction is specified as a prior distribution: $p(\boldsymbol{\theta}_1|\mathbf{x}_0, \mathbf{w}) = p(\boldsymbol{\theta}_1|\mathbf{w}) = \prod_{i=1}^{N}\mathcal{N}\big(\boldsymbol{\theta}_1^i; \boldsymbol{\mu}^i, \boldsymbol{\Sigma}^i\big)$. For $t = 2, \ldots, T$, the one-step prediction density is computed via the Chapman–Kolmogorov equation:

$$\begin{aligned} p(\boldsymbol{\theta}_t|\mathbf{x}_{0:t-1}, \mathbf{w}) &= \int p(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t-1}|\mathbf{x}_{0:t-1}, \mathbf{w})\, d\boldsymbol{\theta}_{t-1} \\ &= \int p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}, \mathbf{x}_{0:t-1}, \mathbf{w})\; p(\boldsymbol{\theta}_{t-1}|\mathbf{x}_{0:t-1}, \mathbf{w})\, d\boldsymbol{\theta}_{t-1}, \end{aligned} \tag{S2.1.8}$$

where $p(\boldsymbol{\theta}_{t-1}|\mathbf{x}_{0:t-1}, \mathbf{w})$ is the filter density at time $t-1$. We assume that the filter density factors into a product of individual neurons. Coupled with the factorized assumption of the state model, this leads to the factorization of the one-step prediction density:

$$p(\boldsymbol{\theta}_t|\mathbf{x}_{0:t-1}, \mathbf{w}) = \prod_{i=1}^{N} \int p(\boldsymbol{\theta}_t^i|\boldsymbol{\theta}_{t-1}^i, \mathbf{w})\, p(\boldsymbol{\theta}_{t-1}^i|\mathbf{x}_{0:t-1}, \mathbf{w})\, d\boldsymbol{\theta}_{t-1}^i. \tag{S2.1.9}$$

We further assume that the filter density at time $t-1$, $p(\boldsymbol{\theta}_{t-1}^i|\mathbf{x}_{0:t-1}, \mathbf{w})$, is approximated by a Gaussian distribution with mean $\boldsymbol{\theta}_{t-1|t-1}^i$ and covariance $\mathbf{W}_{t-1|t-1}^i$ (to be justified at the next filtering step):

$$p(\boldsymbol{\theta}_{t-1}^i|\mathbf{x}_{0:t-1}, \mathbf{w}) = \mathcal{N}\big(\boldsymbol{\theta}_{t-1}^i; \boldsymbol{\theta}_{t-1|t-1}^i, \mathbf{W}_{t-1|t-1}^i\big). \tag{S2.1.10}$$

Here the filter mean is defined as

$$\boldsymbol{\theta}_{t-1|t-1}^i = \int p(\boldsymbol{\theta}_{t-1}^i|\mathbf{x}_{0:t-1})\boldsymbol{\theta}_{t-1}^i d\boldsymbol{\theta}_{t-1}^i = E_{\boldsymbol{\theta}_{t-1}^i|\mathbf{x}_{0:t-1}}\boldsymbol{\theta}_{t-1}^i. \tag{S2.1.11}$$

It represents the expected value of the parameter at time $t-1$ using data up to $t-1$. The filter covariance is

$$\mathbf{W}_{t-1|t-1}^i = E_{\boldsymbol{\theta}_{t-1}^i|\mathbf{x}_{0:t-1}}(\boldsymbol{\theta}_{t-1}^i - E_{\boldsymbol{\theta}_{t-1}^i|\mathbf{x}_{0:t-1}}\boldsymbol{\theta}_{t-1}^i)(\boldsymbol{\theta}_{t-1}^i - E_{\boldsymbol{\theta}_{t-1}^i|\mathbf{x}_{0:t-1}}\boldsymbol{\theta}_{t-1}^i)^\top. \tag{S2.1.12}$$

Given the Gaussian transition model $p(\boldsymbol{\theta}_t^i|\boldsymbol{\theta}_{t-1}^i, \mathbf{w}) = \mathcal{N}(\boldsymbol{\theta}_t^i; \boldsymbol{\theta}_{t-1}^i, \mathbf{Q}^i)$, the one-step prediction density $p(\boldsymbol{\theta}_t|\mathbf{x}_{0:t-1}, \mathbf{w})$ becomes a Gaussian distribution. Namely, by completing the square with resect to $\boldsymbol{\theta}_t^i$ and calculating the integral, we obtain

$$p(\boldsymbol{\theta}_t|\mathbf{x}_{0:t-1}, \mathbf{w}) = \prod_{i=1}^{N} \mathcal{N}\big(\boldsymbol{\theta}_t; \boldsymbol{\theta}_{t|t-1}^i,\, \mathbf{W}_{t|t-1}^i\big), \tag{S2.1.13}$$

where

$$\boldsymbol{\theta}_{t|t-1}^i = \boldsymbol{\theta}_{t-1|t-1}^i, \tag{S2.1.14}$$

$$\mathbf{W}_{t|t-1}^i = \mathbf{W}_{t-1|t-1}^i + \mathbf{Q}^i. \tag{S2.1.15}$$

We also define $\boldsymbol{\theta}_{1|0}^i = \boldsymbol{\mu}^i$ and $\mathbf{W}_{t|0}^i = \boldsymbol{\Sigma}^i$ for the consistent notation of the one-step prediction density for $t = 1, \ldots, T$ in subsequent calculations.

2.1.3 Filtering

Using the observation model and the one-step prediction density $p(\boldsymbol{\theta}_t|\mathbf{x}_{0:t-1}, \mathbf{w})$, the posterior filter density is given as

$$p(\boldsymbol{\theta}_t|\mathbf{x}_{0:t}, \mathbf{w}) \propto \prod_{i=1}^{N}\prod_{l=1}^{L} \exp\left[\theta_{i,t}x_{i,t}^l + \sum_{j=1}^{N}\theta_{ij,t}x_{it}^l x_{j,t-1}^l - \psi(\boldsymbol{\theta}_t^i, \mathbf{x}_{t-1}^l)\right] \cdot \prod_{i=1}^{N} \exp\left[-\frac{1}{2}(\boldsymbol{\theta}_t^i - \boldsymbol{\theta}_{t|t-1}^i)^\top (\mathbf{W}_{t|t-1}^i)^{-1}(\boldsymbol{\theta}_t^i - \boldsymbol{\theta}_{t|t-1}^i)\right]. \tag{S2.1.16}$$

This expression confirms that the filter density at time t is a product of the individual neurons' filter densities, validating the assumption of independent filter densities in constructing the one-step prediction density. The result enables independent filtering for each neuron.

We now approximate the filter density by the Gaussian distribution using Laplace's method. Namely, we obtain the maximum a posteriori (MAP) estimate of the filter density and use the Hessian at around the MAP estimate to obtain the approximate covariance. Using

$$\boldsymbol{\theta}_t^i = [\theta_{i,t}, \theta_{i1,t}, ...\theta_{iN,t}]^\top, \tag{S2.1.17}$$

$$\mathbf{F}(x_{i,t}^l, \mathbf{x}_{t-1}^l) = \left[x_{i,t}^l,\, x_{i,t}^l x_{1,t-1}^l,\, x_{i,t}^l x_{2,t-1}^l,\, \ldots,\, x_{i,t}^l x_{N,t-1}^l\right]^\top, \tag{S2.1.18}$$

we have

$$p(\boldsymbol{\theta}_t|\mathbf{x}_{0:t}, \mathbf{w}) \propto \prod_{i=1}^{N} \exp\left[\sum_{l=1}^{L}(\boldsymbol{\theta}_t^i)^\top \mathbf{F}(x_{i,t}^l, \mathbf{x}_{t-1}^l) - \psi(\boldsymbol{\theta}_t^i, \mathbf{x}_{t-1}^l) - \frac{1}{2}(\boldsymbol{\theta}_t^i - \boldsymbol{\theta}_{t|t-1}^i)^\top (\mathbf{W}_{t|t-1}^i)^{-1}(\boldsymbol{\theta}_t^i - \boldsymbol{\theta}_{t|t-1}^i)\right], \tag{S2.1.19}$$

where $\psi(\boldsymbol{\theta}_t^i, \mathbf{x}_{t-1}^l)$ is now given as

$$\psi(\boldsymbol{\theta}_t^i, \mathbf{x}_{t-1}^l) = \log\left[1 + \exp\left[(\boldsymbol{\theta}_t^i)^\top \mathbf{F}(1, \mathbf{x}_{t-1}^l)\right]\right]. \tag{S2.1.20}$$

First, we obtain the MAP estimate defined as

$$\boldsymbol{\theta}_{\text{MAP}} = \arg\max_{\boldsymbol{\theta}_t} \log p(\boldsymbol{\theta}_t|\mathbf{x}_{0:t}, \mathbf{w}). \tag{S2.1.21}$$

We obtain the MAP estimate through numerical optimization using the Newton–Raphson method. Notably, the MAP estimate for each neurons, $\boldsymbol{\theta}_{\text{MAP}}^i$, can be obtained indepen-

dently of the others. For this goal, we obtain the first and second-order derivatives of the log posterior with respect to $\boldsymbol{\theta}_t^i$. The first-order derivative with respect to $\boldsymbol{\theta}_t^i$ results in

$$\frac{\partial \log p(\boldsymbol{\theta}_t|\mathbf{x}_{0:t}, \mathbf{w})}{\partial \boldsymbol{\theta}_t^i} = \sum_{l=1}^{L} \left[\mathbf{F}(x_{i,t}^l, \mathbf{x}_{t-1}^l) - \frac{\partial \psi(\boldsymbol{\theta}_t^i, \mathbf{x}_{t-1}^l)}{\partial \boldsymbol{\theta}_t^i} \right] - (\mathbf{W}_{t|t-1}^i)^{-1}(\boldsymbol{\theta}_t^i - \boldsymbol{\theta}_{t|t-1}^i). \tag{S2.1.22}$$

Here, the derivative of $\psi(\boldsymbol{\theta}_t^i, \mathbf{x}_{t-1}^l)$ with respect to $\boldsymbol{\theta}_t^i$ is given by:

$$\begin{aligned} \frac{\partial \psi(\boldsymbol{\theta}_t^i, \mathbf{x}_{t-1}^l)}{\partial \boldsymbol{\theta}_t^i} &= \frac{\exp\left[(\boldsymbol{\theta}_t^i)^\top \mathbf{F}(1, \mathbf{x}_{t-1}^l)\right]}{1 + \exp\left[(\boldsymbol{\theta}_t^i)^\top \mathbf{F}(1, \mathbf{x}_{t-1}^l)\right]} \mathbf{F}(1, \mathbf{x}_{t-1}^l) \\ &= \exp\left[(\boldsymbol{\theta}_t^i)^\top \mathbf{F}(1, \mathbf{x}_{t-1}^l) - \psi(\boldsymbol{\theta}_t^i, \mathbf{x}_{t-1}^l)\right] \mathbf{F}(1, \mathbf{x}_{t-1}^l) \\ &= r_{i,t}^l(\mathbf{x}_{t-1}^l) \mathbf{F}(1, \mathbf{x}_{t-1}^l), \end{aligned} \tag{S2.1.23}$$

where we defined the expected rate of i-th neuron at time t given the activity of the previous time step $\mathbf{x}_{t-1}^l$ as

$$\begin{aligned} r_{i,t}^l(\mathbf{x}_{t-1}^l) &\equiv E_{x_{i,t}^l|\mathbf{x}_{t-1}^l} x_{i,t}^l \\ &= \sum_{x_{i,t}^l} p(x_{i,t}^l|\mathbf{x}_{t-1}^l)\, x_{i,t}^l \\ &= \exp\left[(\boldsymbol{\theta}_t^i)^\top \mathbf{F}(1, \mathbf{x}_{t-1}^l) - \psi(\boldsymbol{\theta}_t^i, \mathbf{x}_{t-1}^l)\right]. \end{aligned} \tag{S2.1.24}$$

The second derivative of $\log p(\boldsymbol{\theta}_t|\mathbf{x}_{0:t}, \mathbf{w})$ with respect to $\boldsymbol{\theta}_t^i$ is given by

$$\frac{\partial}{\partial \boldsymbol{\theta}_t^i} \left(\frac{\partial \log p(\boldsymbol{\theta}_t|\mathbf{x}_{0:t}, \mathbf{w})}{\partial (\boldsymbol{\theta}_t^i)^\top} \right) = \sum_{l=1}^{L} \left[-\frac{\partial^2 \psi(\boldsymbol{\theta}_t^i, \mathbf{x}_{t-1}^l)}{\partial \boldsymbol{\theta}_t^i \partial (\boldsymbol{\theta}_t^i)^\top} \right] - (\mathbf{W}_{t|t-1}^i)^{-1}. \tag{S2.1.25}$$

The second derivative of $\psi(\boldsymbol{\theta}_t^i, \mathbf{x}_{t-1}^l)$ with respect to $\boldsymbol{\theta}_t^i$ is given by:

$$\begin{aligned} \frac{\partial^2 \psi(\boldsymbol{\theta}_t^i, \mathbf{x}_{t-1}^l)}{\partial \boldsymbol{\theta}_t^i (\boldsymbol{\theta}_t^i)^\top} &= \frac{\partial}{\partial \boldsymbol{\theta}_t^i} \exp\left[(\boldsymbol{\theta}_t^i)^\top \mathbf{F}(1, \mathbf{x}_{t-1}^l) - \psi(\boldsymbol{\theta}_t^i, \mathbf{x}_{t-1}^l)\right] \mathbf{F}(1, \mathbf{x}_{t-1}^l)^\top \\ &= \exp\left[(\boldsymbol{\theta}_t^i)^\top \mathbf{F}(1, \mathbf{x}_{t-1}^l) - \psi(\boldsymbol{\theta}_t^i, \mathbf{x}_{t-1}^l)\right] \left[\mathbf{F}(1, \mathbf{x}_{t-1}^l) - \frac{\partial \psi(\boldsymbol{\theta}_t^i, \mathbf{x}_{t-1}^l)}{\partial \boldsymbol{\theta}_t^i} \right] \mathbf{F}(1, \mathbf{x}_{t-1}^l)^\top \\ &= r_{i,t}^l(\mathbf{x}_{t-1}^l)\{1 - r_{i,t}^l(\mathbf{x}_{t-1}^l)\} \mathbf{F}(1, \mathbf{x}_{t-1}^l) \mathbf{F}(1, \mathbf{x}_{t-1}^l)^\top. \end{aligned} \tag{S2.1.26}$$

Using the first and second-order derivatives, the MAP estimate $\boldsymbol{\theta}_{\mathrm{MAP}}^i$ for each neurons was found by the Newton-Raphson method.

After finding the MAP estimate, we approximate the filter density by a Gaussian

distribution via the Laplace's method,

$$p(\boldsymbol{\theta}_t^i|\mathbf{x}_{1:t},\mathbf{w}) = \frac{1}{\sqrt{|2\pi\mathbf{W}_{t|t}|}}\exp\left[-\frac{1}{2}(\boldsymbol{\theta}_t^i-\boldsymbol{\theta}_{t|t}^i)^\top\mathbf{W}_{t|t}^{-1}(\boldsymbol{\theta}_t^i-\boldsymbol{\theta}_{t|t}^i)\right] \tag{S2.1.27}$$

with the following mean and variance:

$$\boldsymbol{\theta}_{t|t}^i = \boldsymbol{\theta}_{\mathrm{MAP}}^i, \tag{S2.1.28}$$

and

$$\begin{aligned}\mathbf{W}_{t|t}^i &= \left[-\frac{\partial}{\partial\boldsymbol{\theta}_t^i}\left(\frac{\partial\log p\left(\boldsymbol{\theta}_{t-1}^i|\mathbf{x}_{1-t},\mathbf{w}\right)}{\partial\left(\boldsymbol{\theta}_t^i\right)^\mathsf{T}}\right)\Bigg|_{\boldsymbol{\theta}_t^i=\boldsymbol{\theta}_{t|t}^i}\right]^{-1}\\ &= \left[\mathbf{G}\left(\boldsymbol{\theta}_{t|t}^i\right)+\left(\mathbf{W}_{t|t-1}^i\right)^{-1}\right]^{-1},\end{aligned} \tag{S2.1.29}$$

where $\mathbf{G}\left(\boldsymbol{\theta}_t^i\right)$ is given by

$$\begin{aligned}\mathbf{G}\left(\boldsymbol{\theta}_t^i\right) &= \sum_{l=1}^{L}\frac{\partial^2\psi(\boldsymbol{\theta}_t^i,\mathbf{x}_{t-1}^l)}{\partial\boldsymbol{\theta}_t^i\partial(\boldsymbol{\theta}_t^i)^\top}\\ &= \sum_{l=1}^{L} r_{i,t}^l(\mathbf{x}_{t-1}^l)\{1-r_{i,t}^l(\mathbf{x}_{t-1}^l)\}\mathbf{F}(1,\mathbf{x}_{t-1}^l)\mathbf{F}(1,\mathbf{x}_{t-1}^l)^\top.\end{aligned} \tag{S2.1.30}$$

By sequentially applying the one-step prediction density and the filter density for $t = 1,\ldots,T$, we obtain the filter densities of all time steps.

2.1.4 Smoothing

Given that the filter density is approximated by Gaussian distributions, the smoothing density for the parameters of each neuron can be computed iteratively by using the filter density and the one-step prediction density in a backward manner from the final time step T, following the Rauch-Tung-Striebel smoother [67]:

$$\boldsymbol{\theta}_{t-1|T}^i = \boldsymbol{\theta}_{t-1|t-1}^i + \mathbf{A}_{t-1}^i\left(\boldsymbol{\theta}_{t|T}^i-\boldsymbol{\theta}_{t|t}^i\right), \tag{S2.1.31}$$

$$\mathbf{W}_{t-1|T}^i = \mathbf{W}_{t-1|t-1}^i + \mathbf{A}_{t-1}^i\left(\mathbf{W}_{t|T}^i-\mathbf{W}_{t|t-1}^i\right)\left(\mathbf{A}_{t-1}^i\right)^\top, \tag{S2.1.32}$$

$$\mathbf{A}_{t-1}^i = \mathbf{W}_{t-1|t-1}^i\left(\mathbf{W}_{t|t-1}^i\right)^{-1}, \tag{S2.1.33}$$

for $t = 2,\ldots,T$. For completeness, we provide a compact derivation of these equations below. for $t = 2,\ldots,T$. For completeness, we provide a compact derivation of these equations below.

At the smoothing, we estimate the latent state $\boldsymbol{\theta}_t^i$ given the entire observed data $\mathbf{x}_{0:T}$. The smoother posterior density is given as

$$\begin{aligned} p(\boldsymbol{\theta}_{t-1}^i|\mathbf{x}_{0:T}, \mathbf{w}) &= \int p(\boldsymbol{\theta}_{t-1}^i|\boldsymbol{\theta}_t^i, \mathbf{x}_{0:T}, \mathbf{w}) p(\boldsymbol{\theta}_t^i|\mathbf{x}_{0:T}, \mathbf{w})\, d\boldsymbol{\theta}_t^i \\ &= \int p(\boldsymbol{\theta}_{t-1}^i|\boldsymbol{\theta}_t^i, \mathbf{x}_{0:t-1}, \mathbf{w}) p(\boldsymbol{\theta}_t^i|\mathbf{x}_{0:T}, \mathbf{w})\, d\boldsymbol{\theta}_t^i. \end{aligned} \tag{S2.1.34}$$

Here, we used the Markovian assumption at the second equality. The conditional density $p(\boldsymbol{\theta}_{t-1}^i|\boldsymbol{\theta}_t^i, \mathbf{x}_{0:t-1}, \mathbf{w})$ is obtained as

$$\begin{aligned} p(\boldsymbol{\theta}_{t-1}^i|\boldsymbol{\theta}_t^i, \mathbf{x}_{0:t-1}, \mathbf{w}) &= \frac{p(\boldsymbol{\theta}_{t-1}^i, \boldsymbol{\theta}_t^i|\mathbf{x}_{0:t-1}, \mathbf{w})}{p(\boldsymbol{\theta}_t^i|\mathbf{x}_{0:t-1}, \mathbf{w})} \\ &= \frac{p(\boldsymbol{\theta}_t^i|\boldsymbol{\theta}_{t-1}^i, \mathbf{w}) p(\boldsymbol{\theta}_{t-1}^i|\mathbf{x}_{0:t-1}, \mathbf{w})}{p(\boldsymbol{\theta}_t^i|\mathbf{x}_{0:t-1}, \mathbf{w})}, \end{aligned} \tag{S2.1.35}$$

which is composed of the filter and one-step prediction densities, and the state model. Since we assume that these are Gaussian distributions, given that the smoother density at time t is Gaussian, the linear operations in Eqs. S2.1.34 and S2.1.35 guarantee that the smoother density at time $t-1$ is Gaussian. Therefore, the distribution is specified by the mean and covariance defined as

$$\boldsymbol{\theta}_{t-1|T}^i \equiv E_{\boldsymbol{\theta}_{t-1}^i|\mathbf{x}_{0:T}} \boldsymbol{\theta}_{t-1}^i \tag{S2.1.36}$$

$$\mathbf{W}_{t-1|T}^i \equiv E_{\boldsymbol{\theta}_{t-1}^i|\mathbf{x}_{0:T}} \left(\boldsymbol{\theta}_{t-1}^i - \boldsymbol{\theta}_{t-1|T}^i\right) \left(\boldsymbol{\theta}_{t-1}^i - \boldsymbol{\theta}_{t-1|T}^i\right)^\top. \tag{S2.1.37}$$

To obtain their closed form expressions, first we note that the joint density in Eq. S2.1.35 is written as

$$p(\boldsymbol{\theta}_{t-1}^i, \boldsymbol{\theta}_t^i|\mathbf{x}_{0:t-1}, \mathbf{w}) = \mathcal{N}\left(\begin{pmatrix} \boldsymbol{\theta}_{t-1}^i \\ \boldsymbol{\theta}_t^i \end{pmatrix}; \begin{pmatrix} \boldsymbol{\theta}_{t-1|t-1}^i \\ \boldsymbol{\theta}_{t|t-1}^i \end{pmatrix}, \begin{pmatrix} \mathbf{W}_{t-1|t-1}^i & \mathbf{W}_{t-1,t|t-1}^i \\ \mathbf{W}_{t,t-1|t-1}^i & \mathbf{W}_{t|t-1}^i \end{pmatrix}\right), \tag{S2.1.38}$$

where $\mathbf{W}_{t-1,t|t-1}^i$ is the cross covariance given the data up to time $t-1$. Here, we note that, under the linear Gaussian transition with an identity transition matrix, the one-step prediction mean is

$$\boldsymbol{\theta}_{t|t-1}^i = \boldsymbol{\theta}_{t-1|t-1}^i. \tag{S2.1.39}$$

The cross covariance is obtained as

$$\begin{aligned}\mathbf{W}^i_{t-1,t|t} &\equiv E_{\boldsymbol{\theta}^i_{t-1},\boldsymbol{\theta}^i_t|\mathbf{x}_{0:t}}\left(\boldsymbol{\theta}^i_{t-1}-\boldsymbol{\theta}^i_{t-1|t-1}\right)\left(\boldsymbol{\theta}^i_t-\boldsymbol{\theta}^i_{t|t-1}\right)^\top \\ &= E_{\boldsymbol{\theta}^i_{t-1},\boldsymbol{\xi}_t|\mathbf{x}_{0:t}}\left(\boldsymbol{\theta}^i_{t-1}-\boldsymbol{\theta}^i_{t-1|t-1}\right)\left(\boldsymbol{\theta}^i_{t-1}+\boldsymbol{\xi}_t-\boldsymbol{\theta}^i_{t-1|t-1}\right)^\top \\ &= \mathbf{W}^i_{t-1|t-1}+E_{\boldsymbol{\theta}^i_{t-1},\boldsymbol{\xi}_t|\mathbf{x}_{0:t}}\left[\left(\boldsymbol{\theta}^i_{t-1}-\boldsymbol{\theta}^i_{t-1|t-1}\right)\boldsymbol{\xi}_t^\top\right] \\ &= \mathbf{W}^i_{t-1|t-1}. \end{aligned} \tag{S2.1.40}$$

Here, at the second equality, we inserted the state equation with a state noise $\boldsymbol{\xi}_{t-1}$, and used $\boldsymbol{\theta}^i_{t|t-1}=\boldsymbol{\theta}^i_{t-1|t-1}$. The last equality is obtained due to the orthogonality of the fluctuation of $\boldsymbol{\theta}^i_{t-1}$ and noise $\boldsymbol{\xi}_t$.

Given the joint density, we obtain the conditional density (Eq. S2.1.35). We note that given the multivariate normal distribution,

$$\mathbf{x}=\begin{bmatrix}\mathbf{x}_a\\ \mathbf{x}_b\end{bmatrix}\sim\mathcal{N}\left(\begin{bmatrix}\boldsymbol{\mu}_a\\ \boldsymbol{\mu}_b\end{bmatrix},\begin{bmatrix}\boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab}\\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb}\end{bmatrix}\right). \tag{S2.1.41}$$

The conditional distribution of $\mathbf{x}_a|\mathbf{x}_b$ follows

$$\mathbf{x}_a|\mathbf{x}_b\sim\mathcal{N}(\boldsymbol{\mu}_{a|b},\boldsymbol{\Sigma}_{a|b}) \tag{S2.1.42}$$

with

$$\boldsymbol{\mu}_{a|b}=\boldsymbol{\mu}_a+\boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}(\mathbf{x}_b-\boldsymbol{\mu}_b), \tag{S2.1.43}$$

$$\boldsymbol{\Sigma}_{a|b}=\boldsymbol{\Sigma}_{aa}-\boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba}. \tag{S2.1.44}$$

Applying this formula, we obtain

$$p(\boldsymbol{\theta}^i_{t-1}|\boldsymbol{\theta}^i_t,\mathbf{x}_{0:t-1},\mathbf{w})=\mathcal{N}\left(\boldsymbol{\theta}^i_{t-1};\boldsymbol{\theta}^i_{t-1|t-1}+\mathbf{A}_{t-1}(\boldsymbol{\theta}^i_t-\boldsymbol{\theta}^i_{t-1|t-1}),\mathbf{W}^i_{t-1|t-1}-\mathbf{A}_{t-1}\mathbf{W}^i_{t-1|t-1}\right), \tag{S2.1.45}$$

where $\mathbf{A}_{t-1}=\mathbf{W}^i_{t-1|t-1}(\mathbf{W}^i_{t|t-1})^{-1}$.

Finally, the smoothing density at time t is obtained by multiplying the smoother density at time t and integrating out $\boldsymbol{\theta}^i_t$ according to Eq. S2.1.34. For this, we note that, given the following two normal distributions:

$$p(\mathbf{x}_a|\mathbf{x}_b)=\mathcal{N}(\mathbf{x}_a;\mathbf{A}\mathbf{x}_b+\mathbf{b},\boldsymbol{\Sigma}_{a|b}), \tag{S2.1.46}$$

$$p(\mathbf{x}_b)=\mathcal{N}(\mathbf{x}_b;\boldsymbol{\mu}_b,\boldsymbol{\Sigma}_b), \tag{S2.1.47}$$

the marginal distribution of $\mathbf{x}_a$ is obtained as

$$p(\mathbf{x}_a) = \int p(\mathbf{x}_a|\mathbf{x}_b)\, p(\mathbf{x}_b)\, d\mathbf{x}_b = \mathcal{N}(\mathbf{x}_a; \boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a), \tag{S2.1.48}$$

where

$$\boldsymbol{\mu}_a = \mathbf{A}\boldsymbol{\mu}_b + \mathbf{b}, \tag{S2.1.49}$$

$$\boldsymbol{\Sigma}_a = \mathbf{A}\boldsymbol{\Sigma}_b\mathbf{A}^\top + \boldsymbol{\Sigma}_{a|b}. \tag{S2.1.50}$$

Applying this formula to Eq. S2.1.45 and $p(\boldsymbol{\theta}_t^i|\mathbf{x}_{0:T}, \mathbf{w}) = \mathcal{N}\left(\boldsymbol{\theta}_t^i; \boldsymbol{\theta}_{t|T}^i, \mathbf{W}_{t|T}^i\right)$, we obtain the smoothing density $p(\boldsymbol{\theta}_{t-1}^i|\mathbf{x}_{0:T}, \mathbf{w})$ whose mean and covariance are given by

$$\boldsymbol{\theta}_{t-1|T}^i = \boldsymbol{\theta}_{t-1|t-1}^i + \mathbf{A}_{t-1}(\boldsymbol{\theta}_{t|T}^i - \boldsymbol{\theta}_{t-1|t-1}^i), \tag{S2.1.51}$$

and

$$\begin{aligned}
\mathbf{W}_{t-1|T}^i &= \mathbf{A}_{t-1}\mathbf{W}_{t|T}^i\mathbf{A}_{t-1}^\top + \mathbf{W}_{t-1|t-1}^i - \mathbf{A}_{t-1}\mathbf{W}_{t-1|t-1}^i \\
&= \mathbf{W}_{t-1|t-1}^i + \mathbf{A}_{t-1}\mathbf{W}_{t|T}^i\mathbf{A}_{t-1}^\top - \mathbf{A}_{t-1}\mathbf{W}_{t|t-1}^i(\mathbf{W}_{t|t-1}^i)^{-1}\mathbf{W}_{t-1|t-1}^i \\
&= \mathbf{W}_{t-1|t-1}^i + \mathbf{A}_{t-1}\left(\mathbf{W}_{t|T}^i - \mathbf{W}_{t|t-1}^i\right)\mathbf{A}_{t-1}^\top.
\end{aligned} \tag{S2.1.52}$$

We thus obtained the backward recursion formulae to obtain the smoothing densities.

2.1.5 Optimization of hyperparameters

We consider the problem of optimizing the hyperparameters that maximize the marginal likelihood function. Instead of the marginal likelihood, we optimize its tractable lower bound. In the Expectation-Maximization (EM) algorithm, the posterior density is obtained under given hyperparameters via the algorithm described in the previous section at the E-step. At the M-step, we optimize the hyperparameters that maximize the lower bound, using the given posterior density. Using Jensen's inequality $\log E[X] \geq E[\log X]$,

this lower bound is given by

$$\begin{aligned} l(\mathbf{w}^*) &\equiv \log p(\mathbf{x}_{0:T}|\mathbf{w}^*) \\ &= \log \int p(\boldsymbol{\theta}_{1:T}|\mathbf{x}_{0:T},\mathbf{w}) \frac{p(\mathbf{x}_{0:T},\boldsymbol{\theta}_{1:T}|\mathbf{w}^*)}{p(\boldsymbol{\theta}_{1:T}|\mathbf{x}_{0:T},\mathbf{w})} d\boldsymbol{\theta}_{1:T} \\ &= \log E_{\boldsymbol{\theta}_{1:T}|\mathbf{x}_{0:T},\mathbf{w}} \frac{p(\mathbf{x}_{0:T},\boldsymbol{\theta}_{1:T}|\mathbf{w}^*)}{p(\boldsymbol{\theta}_{1:T}|\mathbf{x}_{0:T},\mathbf{w})} \\ &\geq E_{\boldsymbol{\theta}_{1:T}|\mathbf{x}_{0:T},\mathbf{w}} \log \frac{p(\mathbf{x}_{0:T},\boldsymbol{\theta}_{1:T}|\mathbf{w}^*)}{p(\boldsymbol{\theta}_{1:T}|\mathbf{x}_{0:T},\mathbf{w})} \\ &= E_{\boldsymbol{\theta}_{1:T}|\mathbf{x}_{0:T},\mathbf{w}} \log p(\mathbf{x}_{0:T},\boldsymbol{\theta}_{1:T}|\mathbf{w}^*) - E_{\boldsymbol{\theta}_{1:T}|\mathbf{w}} \log p(\boldsymbol{\theta}_{1:T}|\mathbf{x}_{0:T},\mathbf{w}). \end{aligned} \tag{S2.1.53}$$

The first term is called the Q-function:

$$\begin{aligned} \tilde{Q}(\mathbf{w}) &= E_{\boldsymbol{\theta}_{1:T}|\mathbf{x}_{0:T},\mathbf{w}} \log p(\mathbf{x}_{0:T},\boldsymbol{\theta}_{1:T}|\mathbf{Q}) \\ &= E_{\boldsymbol{\theta}_{1:T}|\mathbf{x}_{0:T},\mathbf{w}} \log p(\mathbf{x}_{0:T}|\boldsymbol{\theta}_{1:T},\mathbf{Q}) + E_{\boldsymbol{\theta}_{1:T}|\mathbf{x}_{0:T},\mathbf{w}} \log p(\boldsymbol{\theta}_{1:T}|\mathbf{Q}). \end{aligned} \tag{S2.1.54}$$

The second term is the entropy of the posterior density, which is fixed at M-step. We thus optimize the hyperparameters that maximize the Q-function. More explicitly, the Q-function can be written as

$$\begin{aligned} \tilde{Q}(\mathbf{w}) = E_{\boldsymbol{\theta}_{1:T}|\mathbf{x}_{0:T},\mathbf{w}} &\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{l=1}^{L}[(\boldsymbol{\theta}_t^i)^T \mathbf{F}(x_{i,t}^l, \mathbf{x}_{t-1}^l) - \psi(\mathbf{x}_{t-1}^l)] \\ + E_{\boldsymbol{\theta}_{1:T}|\mathbf{x}_{0:T},\mathbf{w}} &\sum_{i=1}^{N}\left[-\frac{1}{2}\log|2\pi\boldsymbol{\Sigma}^i| - \frac{1}{2}(\boldsymbol{\theta}_1^i - \boldsymbol{\mu}^i)^\top (\boldsymbol{\Sigma}^i)^{-1}(\boldsymbol{\theta}_1^i - \boldsymbol{\mu}^i)\right] \\ + E_{\boldsymbol{\theta}_{1:T}|\mathbf{x}_{0:T},\mathbf{w}} &\sum_{t=2}^{T}\sum_{i=1}^{N}\left[-\frac{1}{2}\log|2\pi\mathbf{Q}^i| - \frac{1}{2}(\boldsymbol{\theta}_t^i - \boldsymbol{\theta}_{t-1}^i)^\top (\mathbf{Q}^i)^{-1}(\boldsymbol{\theta}_t^i - \boldsymbol{\theta}_{t-1}^i)\right]. \end{aligned} \tag{S2.1.55}$$

Our objective is to choose $\mathbf{Q}$ such that the function $\tilde{Q}(\mathbf{Q})$ attains an extremum. By noting

$$\frac{\partial \log|2\pi\mathbf{Q}^i|}{\partial \mathbf{Q}^i} = \frac{1}{|\mathbf{Q}^i|}\frac{\partial|\mathbf{Q}^i|}{\partial \mathbf{Q}^i} = \frac{1}{|\mathbf{Q}^i|}|\mathbf{Q}^i|(\mathbf{Q}^i)^{-1} = (\mathbf{Q}^i)^{-1}, \tag{S2.1.56}$$

and

$$\begin{aligned} \frac{\partial}{\partial \mathbf{Q}^i}(\boldsymbol{\theta}_t^i - \boldsymbol{\theta}_{t-1}^i)^\top (\mathbf{Q}^i)^{-1}(\boldsymbol{\theta}_t^i - \boldsymbol{\theta}_{t-1}^i) &= \frac{\partial(\mathbf{Q}^i)^{-1}}{\partial \mathbf{Q}^i}\frac{\partial}{\partial(\mathbf{Q}^i)^{-1}}(\boldsymbol{\theta}_t^i - \boldsymbol{\theta}_{t-1}^i)^\top (\mathbf{Q}^i)^{-1}(\boldsymbol{\theta}_t^i - \boldsymbol{\theta}_{t-1}^i) \\ &= -(\mathbf{Q}^i)^{-2}(\boldsymbol{\theta}_t^i - \boldsymbol{\theta}_{t-1}^i)(\boldsymbol{\theta}_t^i - \boldsymbol{\theta}_{t-1}^i)^\top, \end{aligned} \tag{S2.1.57}$$

we obtain

$$\frac{\partial \tilde{Q}(\mathbf{w})}{\partial \mathbf{Q}^i} = E_{\boldsymbol{\theta}_{1:T}|\mathbf{x}_{1:T},\mathbf{w}} \sum_{t=2}^{T} \left[-\frac{1}{2}(\mathbf{Q}^i)^{-1} + \frac{1}{2}(\mathbf{Q}^i)^{-2}(\boldsymbol{\theta}_t^i - \boldsymbol{\theta}_{t-1}^i)(\boldsymbol{\theta}_t^i - \boldsymbol{\theta}_{t-1}^i)^\top \right]. \quad \text{(S2.1.58)}$$

Setting the above derivative equal to zero, it follows that the optimal $\mathbf{Q}^i$ is obtained as

$$\mathbf{Q}^i = \frac{1}{T-1} \sum_{t=2}^{T} E_{\boldsymbol{\theta}_{1:T}|\mathbf{x}_{1:T},\mathbf{w}} (\boldsymbol{\theta}_t^i - \boldsymbol{\theta}_{t-1}^i)(\boldsymbol{\theta}_t^i - \boldsymbol{\theta}_{t-1}^i)^\top. \quad \text{(S2.1.59)}$$

We note that the expectation in the above equation can be decomposed into

$$E_{\boldsymbol{\theta}_{1:T}|\mathbf{x}_{1:T},\mathbf{w}} \left[\boldsymbol{\theta}_t^i (\boldsymbol{\theta}_t^i)^\top - \boldsymbol{\theta}_{t-1}^i (\boldsymbol{\theta}_t^i)^\top - \boldsymbol{\theta}_t^i (\boldsymbol{\theta}_{t-1}^i)^\top + \boldsymbol{\theta}_{t-1}^i (\boldsymbol{\theta}_{t-1}^i)^\top \right]. \quad \text{(S2.1.60)}$$

Hence, using the following definitions of the equal-time covariance matrix:

$$\mathbf{W}_{t|T}^i = E_{\boldsymbol{\theta}_{1:T}|\mathbf{x}_{0:T},\mathbf{w}} \boldsymbol{\theta}_t^i (\boldsymbol{\theta}_t^i)^\top - \boldsymbol{\theta}_{t|T}^i (\boldsymbol{\theta}_{t|T}^i)^\top, \quad \text{(S2.1.61)}$$

and the delayed covariance:

$$\begin{aligned} \mathbf{W}_{t,t-1|T}^i &= E_{\boldsymbol{\theta}_{1:T}|\mathbf{x}_{0:T},\mathbf{w}} (\boldsymbol{\theta}_t^i - \boldsymbol{\theta}_{t|T}^i)(\boldsymbol{\theta}_{t-1}^i - \boldsymbol{\theta}_{t-1|T}^i)^\top \\ &= E_{\boldsymbol{\theta}_{1:T}|\mathbf{x}_{0:T},\mathbf{w}} \boldsymbol{\theta}_t^i (\boldsymbol{\theta}_{t-1}^i)^\top - \boldsymbol{\theta}_{t|T}^i (\boldsymbol{\theta}_{t-1|T}^i)^\top, \end{aligned} \quad \text{(S2.1.62)}$$

the optimal $\mathbf{Q}^i$ is obtained as

$$\mathbf{Q}^i = \frac{1}{T-1} \sum_{t=2}^{T} \left[(\boldsymbol{\theta}_{t|T}^i - \boldsymbol{\theta}_{t-1|T}^i)(\boldsymbol{\theta}_{t|T}^i - \boldsymbol{\theta}_{t-1|T}^i)^\top + \mathbf{W}_{t|T}^i - \mathbf{W}_{t-1,t|T}^i - \mathbf{W}_{t,t-1|T}^i + \mathbf{W}_{t-1|T}^i \right], \quad \text{(S2.1.63)}$$

where $\mathbf{W}_{t-1,t|T}^i = (\mathbf{W}_{t,t-1|T}^i)^\top$. We compute the lag-one smoothed covariance following the method of De Jong and Mackinnon [66]:

$$\mathbf{W}_{t,t-1|T}^i = \mathbf{W}_{t|t}^i (\mathbf{W}_{t+1|t}^i)^{-1} \mathbf{W}_{t|T}^i. \quad \text{(S2.1.64)}$$

Similarly, we update $\boldsymbol{\Sigma}^i$ according to

$$\boldsymbol{\Sigma}^i = \mathbf{W}_{1|T}^i + (\boldsymbol{\theta}_{1|T}^i - \boldsymbol{\mu})(\boldsymbol{\theta}_{1|T}^i - \boldsymbol{\mu})^\top. \quad \text{(S2.1.65)}$$

2.1.6 Approximate log marginal likelihood function

The convergence of the EM algorithm was assessed using the log marginal likelihood. Below, we derive the approximate solution for the log marginal likelihood of the kinetic Ising model.

First, we note that the marginal likelihood function $p(\mathbf{x}_{0:T}|\mathbf{w})$ can be expressed as follows:

$$\begin{aligned} p(\mathbf{x}_{0:T}|\mathbf{w}) &= p(\mathbf{x}_0)\prod_{t=1}^{T} p(\mathbf{x}_t|\mathbf{x}_{0:t-1},\mathbf{w}) \\ &= p(\mathbf{x}_0)\prod_{t=1}^{T}\int d\boldsymbol{\theta}_t p(\mathbf{x}_t|\mathbf{x}_{0:t-1},\boldsymbol{\theta}_t,\mathbf{w})p(\boldsymbol{\theta}_t|\mathbf{x}_{0:t-1},\mathbf{w}) \\ &= p(\mathbf{x}_0)\prod_{t=1}^{T}\int d\boldsymbol{\theta}_t p(\mathbf{x}_t|\mathbf{x}_{t-1},\boldsymbol{\theta}_t)p(\boldsymbol{\theta}_t|\mathbf{x}_{0:t-1},\mathbf{w}) \\ &= p(\mathbf{x}_0)\prod_{t=1}^{T}\prod_{i=1}^{N}\int d\boldsymbol{\theta}_t^i \prod_{l=1}^{L} p(x_{i,t}^l|\mathbf{x}_{t-1},\boldsymbol{\theta}_t^i)p(\boldsymbol{\theta}_t^i|\mathbf{x}_{0:t-1},\mathbf{w}). \end{aligned} \tag{S2.1.66}$$

The observation model and the one-step prediction density in the equation above are written as

$$\begin{aligned} \prod_{l=1}^{L} p(x_{i,t}^l|\mathbf{x}_{t-1}^l,\boldsymbol{\theta}_t^i) &= \prod_{l=1}^{L}\exp\left[\theta_{i,t}x_{i,t}^l + \sum_{j=1}^{N}\theta_{ij,t}x_{it}^l x_{j,t-1}^l - \psi(\boldsymbol{\theta}_t^i,\mathbf{x}_{t-1}^l)\right] \\ &= \exp\left[(\boldsymbol{\theta}_t^i)^T\sum_{l=1}^{L}\mathbf{F}(x_{i,t}^l,\mathbf{x}_{t-1}^l) - \sum_{l=1}^{L}\psi(\boldsymbol{\theta}_t^i,\mathbf{x}_{t-1}^l)\right], \end{aligned} \tag{S2.1.67}$$

and

$$p(\boldsymbol{\theta}_t^i|\mathbf{x}_{0:t-1},\mathbf{w}) = \frac{1}{\sqrt{|2\pi\mathbf{W}_{t|t-1}^i|}}\exp\left[-\frac{1}{2}(\boldsymbol{\theta}_t^i-\boldsymbol{\theta}_{t|t-1}^i)^\top(\mathbf{W}_{t|t-1}^i)^{-1}(\boldsymbol{\theta}_t^i-\boldsymbol{\theta}_{t|t-1}^i)\right]. \tag{S2.1.68}$$

Substituting Eqs.S2.1.67 and S2.1.68 into Eq.S2.1.66, we obtain

$$\begin{aligned} p(\mathbf{x}_{0:T}|\mathbf{w}) = p(\mathbf{x}_0)\prod_{i=1}^{N}\prod_{t=1}^{T}\int d\boldsymbol{\theta}_t^i \frac{1}{\sqrt{|2\pi\mathbf{W}_{t|t-1}^i|}} \\ \cdot\exp\left[(\boldsymbol{\theta}_t^i)^T\sum_{l=1}^{L}\mathbf{F}(x_{i,t}^l,\mathbf{x}_{t-1}^l) - \sum_{l=1}^{L}\psi(\boldsymbol{\theta}_t^i,\mathbf{x}_{t-1}^l) - \frac{1}{2}(\boldsymbol{\theta}_t^i-\boldsymbol{\theta}_{t|t-1}^i)^\top(\mathbf{W}_{t|t-1}^i)^{-1}(\boldsymbol{\theta}_t^i-\boldsymbol{\theta}_{t|t-1}^i)\right]. \end{aligned} \tag{S2.1.69}$$

We now define the function $q(\boldsymbol{\theta}_t^i)$ as follows:

$$q(\boldsymbol{\theta}_t^i) = (\boldsymbol{\theta}_t^i)^T \sum_{l=1}^{L} \mathbf{F}(x_{i,t}^l, \mathbf{x}_{t-1}^l) - \sum_{l=1}^{L} \psi(\boldsymbol{\theta}_t^i, \mathbf{x}_{t-1}^l) - \frac{1}{2}(\boldsymbol{\theta}_t^i - \boldsymbol{\theta}_{t|t-1}^i)^\top (\mathbf{W}_{t|t-1}^i)^{-1}(\boldsymbol{\theta}_t^i - \boldsymbol{\theta}_{t|t-1}^i). \tag{S2.1.70}$$

The Taylor expansion of $q(\boldsymbol{\theta}_t^i)$ around $\boldsymbol{\theta}^*$ up to the second order yields

$$q(\boldsymbol{\theta}_t^i) = q(\boldsymbol{\theta}^*) + \left.\frac{\partial q(\boldsymbol{\theta}_t^i)}{\partial \boldsymbol{\theta}_t^i}\right|_{\boldsymbol{\theta}_t^i=\boldsymbol{\theta}^*} (\boldsymbol{\theta}_t^i - \boldsymbol{\theta}^*) + \frac{1}{2}(\boldsymbol{\theta}_t^i - \boldsymbol{\theta}^*)^\top \left.\frac{\partial^2 q(\boldsymbol{\theta}^i)}{\partial \boldsymbol{\theta}_t^i \partial (\boldsymbol{\theta}_t^i)^\top}\right|_{\boldsymbol{\theta}_t^i=\boldsymbol{\theta}^*} (\boldsymbol{\theta}_t^i - \boldsymbol{\theta}^*). \tag{S2.1.71}$$

The value of $\boldsymbol{\theta}_t^i$ that maximizes the function $q(\boldsymbol{\theta}_t^i)$ is the MAP estimate $\boldsymbol{\theta}_{t|t}^i$ of the filter density. Further, the quadratic term evaluated at the MAP estimate is given by the negative inverse of the filter covariance $\mathbf{W}_{t|t}^i$. Hence, at $\boldsymbol{\theta}^* = \boldsymbol{\theta}_{t|t}^i$, the Taylor expansion becomes

$$q(\boldsymbol{\theta}_t^i) \simeq q(\boldsymbol{\theta}_{t|t}^i) - \frac{1}{2}(\boldsymbol{\theta}_t^i - \boldsymbol{\theta}_{t|t}^i)^\top (\mathbf{W}_{t|t}^i)^{-1}(\boldsymbol{\theta}_t^i - \boldsymbol{\theta}_{t|t}^i). \tag{S2.1.72}$$

With this quadratic approximation, the marginal likelihood is obtained as

$$\begin{aligned} p(\mathbf{x}_{0:T}|\mathbf{w}) &\simeq p(\mathbf{x}_0) \prod_{t=1}^{T} \prod_{i=1}^{N} \int d\boldsymbol{\theta}_t^i \frac{1}{\sqrt{|2\pi \mathbf{W}_{t|t-1}^i|}} \exp\left[q(\boldsymbol{\theta}_{t|t}^i) - (\boldsymbol{\theta}_t^i - \boldsymbol{\theta}_{t|t}^i)^\top \frac{1}{2}[\mathbf{W}_{t|t}^i]^{-1}(\boldsymbol{\theta}_t^i - \boldsymbol{\theta}_{t|t}^i)\right] \\ &= p(\mathbf{x}_0) \prod_{t=1}^{T} \prod_{i=1}^{N} \exp[q(\boldsymbol{\theta}_{t|t}^i)] \frac{\sqrt{|2\pi \mathbf{W}_{t|t}^i|}}{\sqrt{|2\pi \mathbf{W}_{t|t-1}^i|}} \frac{1}{\sqrt{|2\pi \mathbf{W}_{t|t}^i|}} \qquad \text{(S2.1.73)} \\ &\int d\boldsymbol{\theta}_t^i \exp\left[-(\boldsymbol{\theta}_t^i - \boldsymbol{\theta}_{t|t}^i)^\top \frac{1}{2}[\mathbf{W}_{t|t}^i]^{-1}(\boldsymbol{\theta}_t^i - \boldsymbol{\theta}_{t|t}^i)\right] \\ &= p(\mathbf{x}_0) \prod_{t=1}^{T} \prod_{i=1}^{N} \sqrt{\frac{|2\pi \mathbf{W}_{t|t}^i|}{|2\pi \mathbf{W}_{t|t-1}^i|}} \exp[q(\boldsymbol{\theta}_{t|t}^i)]. \qquad \text{(S2.1.74)} \end{aligned}$$

We thus obtain the log marginal likelihood function as follows:

$$\log p(\mathbf{x}_{0:T}|\mathbf{w}) \simeq \log p(\mathbf{x}_0) + \sum_{t=1}^{T} \sum_{i=1}^{N} \left[\frac{1}{2} \log |\mathbf{W}_{t|t}^i| - \frac{1}{2} \log |\mathbf{W}_{t|t-1}^i| + q(\boldsymbol{\theta}_{t|t}^i)\right]. \tag{S2.1.75}$$

2.2 An alternative calculation of the backward conditional entropy

Here, we give an alternative approach to obtaining the backward conditional entropy to the one given in Methods. The result gives an identical approximate solution.

Under the approximation of the following probabilities by independent distributions:

$$p(\mathbf{x}_{t-2}) = Q(\mathbf{x}_{t-2}), \tag{S2.2.1}$$

$$p(\mathbf{x}_t|\mathbf{x}_{t-1}) = Q(\mathbf{x}_t), \tag{S2.2.2}$$

the backward conditional entropy is approximated as

$$\begin{aligned}
\sigma_t^{\text{backward}} &= -\sum_{\mathbf{x}_{t-2}}\sum_{\mathbf{x}_{t-1}} p(\mathbf{x}_{t-1}|\mathbf{x}_{t-2})p(\mathbf{x}_{t-2}) \sum_{\mathbf{x}_t} p(\mathbf{x}_t|\mathbf{x}_{t-1}) \sum_i \left[x_{i,t-1}h_{i,t}(\mathbf{x}_t) - \psi(h_{i,t}(\mathbf{x}_t))\right] \\
&\simeq -\sum_{\mathbf{x}_{t-2}}\sum_{\mathbf{x}_{t-1}} p(\mathbf{x}_{t-1}|\mathbf{x}_{t-2})Q(\mathbf{x}_{t-2}) \sum_{\mathbf{x}_t} Q(\mathbf{x}_t) \sum_i \left[x_{i,t-1}h_{i,t}(\mathbf{x}_t) - \psi(h_{i,t}(\mathbf{x}_t))\right] \\
&= -\sum_i \sum_{x_{i,t-1}} \sum_{\mathbf{x}_{t-2}} p(x_{i,t-1}|\mathbf{x}_{t-2})Q(\mathbf{x}_{t-2}) \sum_{\mathbf{x}_t} Q(\mathbf{x}_t) \left[x_{i,t-1}h_{i,t}(\mathbf{x}_t) - \psi(h_{i,t}(\mathbf{x}_t))\right].
\end{aligned} \tag{S2.2.3}$$

Let us define $\tilde{\phi}_{i,t}(x_{i,t})$ as

$$\tilde{\phi}_{i,t}(x_{i,t-1}) = \sum_{\mathbf{x}_t} Q(\mathbf{x}_t) \left[x_{i,t-1}h_{i,t}(\mathbf{x}_t) - \psi(h_{i,t}(\mathbf{x}_t))\right]. \tag{S2.2.4}$$

Using

$$\gamma(h_{i,t}) = x_{i,t-1}h_{i,t} - \psi(h_{i,t}), \tag{S2.2.5}$$

we approximate $\tilde{\phi}_{i,t}(x_{i,t})$ as

$$\tilde{\phi}_{i,t}(x_{i,t-1}) \approx \int \mathcal{D}_z \, \gamma(g_{i,t} + z\sqrt{\Delta_{i,t}}), \tag{S2.2.6}$$

where $\mathcal{D}_z = \frac{\mathrm{d}z}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right)$.

Then, the backward conditional entropy is written as

$$\begin{aligned}\sigma_t^{\text{backward}} &= -\sum_{\mathbf{x}_{t-2}}\sum_{\mathbf{x}_{t-1}} p(\mathbf{x}_{t-1}|\mathbf{x}_{t-2})Q(\mathbf{x}_{t-2})\tilde{\phi}_{i,t}(x_{i,t-1}) \\ &= -\sum_i \sum_{x_{i,t-1}} \left(\sum_{\mathbf{x}_{t-2}} p(x_{i,t-1}|\mathbf{x}_{t-2})Q(\mathbf{x}_{t-2})\right)\tilde{\phi}_{i,t}(x_{i,t-1}). \end{aligned} \tag{S2.2.7}$$

Note that, from Eq. 2.30, we have

$$\begin{aligned} m_{i,t} &= \sum_{\mathbf{x}_{t-1}} p(\mathbf{x}_t = 1|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}) \simeq \sum_{\mathbf{x}_{t-1}} p(\mathbf{x}_t = 1|\mathbf{x}_{t-1})Q(\mathbf{x}_{t-1}), \\ 1 - m_{i,t} &= \sum_{\mathbf{x}_{t-1}} p(\mathbf{x}_t = 0|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}) \simeq \sum_{\mathbf{x}_{t-1}} p(\mathbf{x}_t = 0|\mathbf{x}_{t-1})Q(\mathbf{x}_{t-1}). \end{aligned} \tag{S2.2.8}$$

Applying these equations for the case of $t-1$, we obtain

$$\sigma_t^{\text{backward}} \simeq -\sum_i \left\{ m_{i,t-1}\tilde{\phi}_{i,t}(x_{i,t-1} = 1) + (1 - m_{i,t-1})\tilde{\phi}_{i,t}(x_{i,t-1} = 0)\right\}. \tag{S2.2.9}$$

Thus, it can be obtained by computing the two Gaussian integral terms.

Since this equation can be further computed as

$$\sigma_t^{\text{backward}} \simeq -\sum_i \left\{ m_{i,t-1}(\tilde{\phi}_{i,t}(x_{i,t-1} = 1) - \tilde{\phi}_{i,t}(x_{i,t-1} = 0)) + \tilde{\phi}_{i,t}(x_{i,t-1} = 0)\right\}. \tag{S2.2.10}$$

and

$$\begin{aligned} \tilde{\phi}_{i,t}(x_{i,t-1} = 1) - \tilde{\phi}_{i,t}(x_{i,t-1} = 0) &= \sum_{\mathbf{x}_t} Q\left(\mathbf{x}_t\right) h_{i,t}(\mathbf{x}_t), \\ \tilde{\phi}_{i,t}(x_{i,t-1} = 0) &= -\sum_{\mathbf{x}_t} Q\left(\mathbf{x}_t\right) \psi(h_{i,t}(\mathbf{x}_t)), \end{aligned} \tag{S2.2.11}$$

it becomes

$$\sigma_t^{\text{backward}} = -\sum_i \sum_{\mathbf{x}_t} Q(\mathbf{x}_t)\left[m_{i,t-1}h_{i,t}(\mathbf{x}_t) - \psi(h_{i,t}(\mathbf{x}_t))\right], \tag{S2.2.12}$$

which is equivalent to Eq. 2.35 in Methods and can be also approximated by the Gaussian integral.

2.3 Mean-field entropy flow under specific conditions

In this section, we derive the mean-field approximation of the entropy flow under the steady-state conditions or for independent neurons.

First, let us summarize the mean-field entropy flow. It is obtained as

$$\begin{aligned}\sigma_t^{\text{flow}} &= -\sigma_t^{\text{forward}} + \sigma_t^{\text{backward}} \\ &\approx \sum_i \int \mathcal{D}_z \left[-\chi\left(g_{i,t,t-1} + z\sqrt{\Delta_{i,t,t-1}}\right) + \phi_{i,t}\left(g_{i,t,t} + z\sqrt{\Delta_{i,t,t}}\right)\right], \end{aligned} \tag{S2.3.1}$$

where $g_{i,t,s}$ and $\Delta_{i,t,s}$ $(s = t, t-1)$ are given as

$$g_{i,t,s} = \theta_{i,t} + \sum_j \theta_{ij,t} m_{j,s}, \tag{S2.3.2}$$

$$\Delta_{i,t,s} = \sum_j \theta_{ij,t}^2 m_{j,s}(1 - m_{j,s}). \tag{S2.3.3}$$

Here $m_{j,s}$ is the mean-field activation rate of the j-th neuron at time s.

Using $r(h) = 1/(1+e^{-h})$ and $\psi(h) = -\log(1 - r(h))$, $\chi(h)$ and $\phi_{i,t}(h)$ are given as

$$\begin{aligned}\chi(h) &= -r(h)\log r(h) - (1 - r(h))\log(1 - r(h)) \\ &= -r(h)\log\frac{r(h)}{1 - r(h)} - \log(1 - r(h)) \\ &= -r(h)h + \psi(h), \end{aligned} \tag{S2.3.4}$$

and

$$\phi_{i,t}(h) = -m_{i,t-1}h + \psi(h). \tag{S2.3.5}$$

2.3.1 Steady-state solution

Under the steady-state assumption ($m_{i,t} = m_{i,t-1} \equiv m_i$), we have $g_{i,t,t-1} = g_{i,t,t} \equiv g_i$ and $\Delta_{i,t,t-1} = \Delta_{i,t,t} \equiv \Delta_i$, making the inputs to χ and $\phi_{i,t}$ common for each neuron. Then, using Eqs. S2.3.4 and S2.3.5 with the common $h = g_i + z\sqrt{\Delta_i}$, we have

$$\begin{aligned}\sigma_t^{\text{flow}} &\approx \sum_i \int \mathcal{D}_z \left(r\left(g_i + z\sqrt{\Delta_i}\right) - m_i\right) \cdot \left(g_i + z\sqrt{\Delta_i}\right) \\ &= \sum_i \int \mathcal{D}_z \left(r\left(g_i + z\sqrt{\Delta_i}\right) - m_i\right) \cdot z\sqrt{\Delta_i}. \end{aligned} \tag{S2.3.6}$$

The term $r\left(g_i + z\sqrt{\Delta_i}\right) - m_i$ represents how the neuron's activity rate deviates from its long-term average, while $z\sqrt{\Delta_i}$ is the fluctuating input to that neuron. Thus, the mean-field solution for the steady state provides an intuitive picture of entropy flow as a measure of the neuron's causal response to fluctuations in its input.

The non-negativity of the mean-field entropy flow can be formally confirmed by Stein's lemma $E(f(X)(X-\mu)) = \sigma^2 E(f'(X))$ for a Gaussian random variable X with expectation μ and variance σ^2. By identifying $f(h) = r(h) - m_i$, $h - g_i = z\sqrt{\Delta_i}$, and $f'(h) = r'(h)$, it can be written as

$$\sigma_t^{\text{flow}} \approx \sum_i \Delta_i \left(\int \mathcal{D}_z \, r'(g_i + z\sqrt{\Delta_i}) \right), \tag{S2.3.7}$$

where $r'(h) = r(h)(1 - r(h))$. Since $\Delta_i \geq 0$ and $r'(h) \geq 0$, the entropy flow is non-negative, which satisfies the requested property of the entropy flow at the steady state. However, while insightful, this form also reveals a key limitation of the approximation: the zero entropy flow is realized only at $\theta_{ij} = 0$ (except for $r = 0, 1$). Consequently, it does not correctly reduce to zero for symmetric couplings, failing to fully incorporate the distinction between symmetric and asymmetric interactions.

2.3.2 Independent neurons

Here we consider independent neurons (i.e., no couplings $\theta_{ij} = 0$) with time-varying field $\theta_{i,t}$. The entropy flow in this system is caused solely by the time-varying fields, or equivalently, the activity rate of individual neurons.

In this case, we have

$$g_{i,t,s} = \theta_{i,t}, \tag{S2.3.8}$$

$$\Delta_{i,t,s} = 0, \tag{S2.3.9}$$

which is independent of s, making the inputs to χ and $\phi_{i,t}$ common once again. Then, we have

$$\begin{aligned} \sigma_t^{\text{flow}} &\approx \sum_i \left(r\left(\theta_{i,t}\right) - m_{i,t-1} \right) \cdot \theta_{i,t} \\ &= \sum_i \left(m_{i,t} - m_{i,t-1} \right) \cdot \theta_{i,t}. \end{aligned} \tag{S2.3.10}$$

For $\theta_{i,t} < 0$, which corresponds to $m_{i,t} < 0.5$, a decrease in the activity rate $m_{i,t} - m_{i,t-1} < 0$ yields positive entropy flow, and an increase in the activity rate induces negative entropy flow.

2.4 The d-prime measure

Here, we provide the definition of the primary behavioral metric, d' (*d-prime*), for clarity. This follows the white paper of "Allen Brain Observatory: Visual Behavior Neuropixels", where further details are available.

To evaluate the sensitivity of the mice to the stimulus, the primary behavioral metric, d', was calculated using data detected only in the active condition with visual changes. The formula for d' is as follows:

$$d' = Z(R_H) - Z(R_F), \tag{S2.4.1}$$

where R_H is the hit rate (the proportion of trials in which the mouse correctly responded to a change in the visual stimulus), and R_F is the false alarm rate (the proportion of trials in which the mouse incorrectly responded to a non-existent change). The function Z represents the inverse of the cumulative distribution function of a standard normal distribution, converting the hit and false alarm rates into z-scores. To prevent extreme values (e.g., 0 or 1) from distorting the results, R_H and R_F were adjusted using the following boundary equations:

$$\frac{1}{2N_H} \leq R_H \leq 1 - \frac{1}{2N_H}, \quad \frac{1}{2N_F} \leq R_F \leq 1 - \frac{1}{2N_F}, \tag{S2.4.2}$$

where N_H and N_F are the total number of trials for the hit and false alarm conditions, respectively. To assess the overall behavioral performance across sessions or experimental conditions, mean d' was used as an aggregated measure, representing the average d' over multiple trials or sessions. For more details, see [68].

2.5 Entropy flow of high-firing neurons and behavioral performance

To elucidate how individual neurons increase total entropy flow in the active condition despite a smaller fraction of neurons exhibiting substantial firing rates (Supplementary Fig.S2, Fig.S3, and Fig.S4), we examined the relationship between the entropy flow and spike rates of individual neurons.

As shown in Eqs. 1.14 and 1.15, the mean-field entropy flow can be decomposed into contributions from individual neurons. We computed the entropy flow of individual neurons under the active and passive conditions and compared them with their firing rates (Supplementary Fig.S8**A**, mouse 574078). The dotted lines connect the values for the active (red) and passive (blue) conditions. We then investigated whether the change in the entropy flow by the behavioral conditions depends on the neuron's firing rate. Supplementary Fig.S8**B** shows the relationship between the geometric mean spike rates of the two conditions (abscissa) and the difference in entropy flow (ordinate) for individual neurons. The difference was computed as 'active' - 'passive,' indicating that the positive value marks a larger entropy flow in the active condition. The positive Spearman rank correlation coefficient ($\rho = 0.22$) for this exemplary mouse suggests that neurons with higher spike rates contributed to increasing total entropy flow in the active condition, despite the summed entropy flow differences across all individual neurons being negative (-3.8331 for this mouse). However, significant variations in the rank correlations were observed across mice.

Assuming that fewer high-firing neurons in the sparsely active populations in the active condition play a critical role in sensory processing (i.e., sparse coding [69, 70, 71]) and that such sensory processing involves time-asymmetric causal patterns, we hypothesized that the above relationship between the spike rates and entropy flow change might be related to mice's cognitive performance. To evaluate the task sensitivity of the mice, we used the primary behavioral metric, d' (*mean d-prime*, see Supplementary Note 2.4 for its definition). The scatter plot in the left panel of Supplementary Fig.S8**C** illustrates the relationship between behavioral measures (mean d-prime) and the rank correlation of entropy flow change with spike rates for all mice for image 'im036_r'. The plot suggests a positive dependency between these two values ($\rho = 0.3578$ measured by the Spearman rank correlation). To confirm this result, we conducted the permutation test that compared the observed rank correlation of the scatter plot with those of the surrogate data constructed by permuting the values of mean d-prime (Supplementary Fig.S8**C** Right). The result confirms the statistical significance of the positive correlation ($p = 0.0304$).

To corroborate that the result does not reflect estimation error in couplings, we ana-

lyzed trial-shuffled data, which showed no clear trend (Supplementary Fig.S8**D**). A permutation test confirmed that the observed correlation yielded a non-significant p-value of 0.5063. This result confirms that the association between higher entropy flow and higher firing neurons in more task-sensitive mice was driven by significant changes in the coupling strengths between the active and passive conditions, rather than firing rate shifts or noise couplings.

However, the additional analyses on the images im012_r and im115_r revealed that these relations were not significantly correlated ('im012_r': $p = 0.574$; 'im115_r': $p = 0.333$, permutation test). Similar analysis replacing the difference of the entropy flow between active and passive conditions with the difference of the entropy flow per activity rate between active and passive conditions yielded non-significant results for these three images.

References

[1] Koyama, S., Eden, U. T., Brown, E. N. & Kass, R. E. Bayesian decoding of neural spike trains. *Annals of the Institute of Statistical Mathematics* **62**, 37–59 (2010).

[2] Schaub, M. T. & Schultz, S. R. The ising decoder: reading out the activity of large neural ensembles. *Journal of Computational Neuroscience* **32**, 101–118 (2012).

[3] Hernández, D. G., Sober, S. J. & Nemenman, I. Unsupervised bayesian ising approximation for decoding neural activity and other biological dictionaries. *eLife* **11**, e68192 (2022).

[4] Harris, J. J., Jolivet, R. & Attwell, D. Synaptic energy use and supply. *Neuron* **75**, 762–777 (2012).

[5] Churchland, M. M. *et al.* Neural population dynamics during reaching. *Nature* **487**, 51–56 (2012).

[6] Pfeiffer, B. E. & Foster, D. J. Hippocampal place-cell sequences depict future paths to remembered goals. *Nature* **497**, 74–79 (2013).

[7] Lynn, C. W., Cornblath, E. J., Papadopoulos, L., Bertolero, M. A. & Bassett, D. S. Broken detailed balance and entropy production in the human brain. *Proceedings of the National Academy of Sciences* **118**, e2109889118 (2021).

[8] Sanz Perl, Y., Kringelbach, M. L., Deco, G., Sitt, J. D. & Tagliazucchi, E. Nonequilibrium brain dynamics as a signature of consciousness. *Physical Review E* **104**, 014411 (2021).

[9] de la Fuente, L. A. *et al.* Temporal irreversibility of neural dynamics as a signature of consciousness. *Cerebral Cortex* **33**, 1856–1865 (2023).

[10] Zanin, M., Güntekin, B., Aktürk, T., Hanoglu, L. & Papo, D. Time irreversibility of resting-state activity in the healthy brain and pathology. *Frontiers in Physiology* **10**, 1619 (2020).

[11] Deco, G. *et al.* The arrow of time of brain signals in cognition: Potential intriguing role of parts of the default mode network. *Network Neuroscience* **7**, 966–998 (2023).

[12] Lynn, C. W., Holmes, C. M., Bialek, W. & Schwab, D. J. Decomposing the local arrow of time in interacting systems. *Physical Review Letters* **129**, 118101 (2022).

[13] Monteforte, M. & Wolf, F. Dynamical entropy production in spiking neuron networks in the balanced state. *Physical Review Letters* **105**, 268104 (2010).

[14] Lucente, D., Gradenigo, G. & Salasnich, L. Entropy production and irreversibility in the linearized stochastic amari neural model. *Entropy* **27**, 1104 (2025).

[15] Nartallo-Kaluarachchi, R., Kringelbach, M. L., Deco, G., Lambiotte, R. & Goriely, A. Nonequilibrium physics of brain dynamics. *arXiv* (2025). Forthcoming in *Physics Reports* (2026), `2504.12144`.

[16] Shimazaki, H., Amari, S.-i., Brown, E. N. & Grün, S. State-space analysis of time-varying higher-order spike correlation for multiple neural spike train data. *PLoS computational biology* **8**, e1002385 (2012).

[17] Donner, C., Obermayer, K. & Shimazaki, H. Approximate inference for time-varying interactions and macroscopic dynamics of neural populations. *PLoS computational biology* **13**, e1005309 (2017).

[18] Dempster, A. P., Laird, N. M. & Rubin, D. B. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society: series B (methodological)* **39**, 1–22 (1977).

[19] Crooks, G. E. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Physical Review E* **60**, 2721 (1999).

[20] Seifert, U. Stochastic thermodynamics, fluctuation theorems and molecular machines. *Reports on progress in physics* **75**, 126001 (2012).

[21] Ito, S., Oizumi, M. & Amari, S.-i. Unified framework for the entropy production and the stochastic interaction based on information geometry. *Physical Review Research* **2**, 033048 (2020).

[22] Yang, Y.-J. & Qian, H. Unified formalism for entropy production and fluctuation relations. *Physical Review E* **101**, 022129 (2020).

[23] Wolpert, D. H. *et al.* Is stochastic thermodynamics the key to understanding the energy costs of computation? *Proceedings of the National Academy of Sciences* **121**, e2321112121 (2024).

[24] Gaspard, P. Time-reversed dynamical entropy and irreversibility in markovian random processes. *Journal of statistical physics* **117**, 599–615 (2004).

[25] Cofré, R., Videla, L. & Rosas, F. An introduction to the non-equilibrium steady states of maximum entropy spike trains. *Entropy* **21**, 884 (2019).

[26] Igarashi, M. Entropy production for discrete-time markov processes. *arXiv:2205.07214* (2022).

[27] Kappen, H. & Spanjers, J. Mean field theory for asymmetric neural networks. *Physical Review E* **61**, 5658 (2000).

[28] Roudi, Y. & Hertz, J. Dynamical tap equations for non-equilibrium ising spin glasses. *Journal of Statistical Mechanics: Theory and Experiment* **2011**, P03031 (2011).

[29] Roudi, Y. & Hertz, J. Mean field theory for nonequilibrium network reconstruction. *Physical review letters* **106**, 048702 (2011).

[30] Mézard, M. & Sakellariou, J. Exact mean-field inference in asymmetric kinetic ising systems. *Journal of Statistical Mechanics: Theory and Experiment* **2011**, L07001 (2011).

[31] Sakellariou, J., Roudi, Y., Mezard, M. & Hertz, J. Effect of coupling asymmetry on mean-field solutions of the direct and inverse sherrington–kirkpatrick model. *Philosophical Magazine* **92**, 272–279 (2012).

[32] Aguilera, M., Moosavi, S. A. & Shimazaki, H. A unifying framework for mean-field theories of asymmetric kinetic ising systems. *Nature communications* **12**, 1197 (2021).

[33] Siegle, J. H. *et al.* Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature* **592**, 86–92 (2021).

[34] Nitzan, N., Bennett, C., Movshon, J. A., Olsen, S. R. & Buzsáki, G. Mixing novel and familiar cues modifies representations of familiar visual images and affects behavior. *Cell reports* **43** (2024).

[35] Ito, S. *et al.* Coordinated changes in a cortical circuit sculpt effects of novelty on neural dynamics. *Cell reports* **43** (2024).

[36] Rolls, E. T. & Tovee, M. J. Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *Journal of neurophysiology* **73**, 713–726 (1995).

[37] Aguilera, M., Igarashi, M. & Shimazaki, H. Nonequilibrium thermodynamics of the asymmetric sherrington-kirkpatrick model. *Nature Communications* **14**, 3685 (2023).

[38] Aguilera, M., Ito, S. & Kolchinsky, A. Inferring entropy production in many-body systems using nonequilibrium maxent. *arXiv preprint arXiv:2505.10444* (2025).

[39] Donner, C. & Opper, M. Inverse ising problem in continuous time: A latent variable approach. *Physical Review E* **96**, 062104 (2017).

[40] Campajola, C., Gangi, D. D., Lillo, F. & Tantari, D. Modelling time-varying interactions in complex systems: the score driven kinetic ising model. *Scientific Reports* **12**, 19339 (2022).

[41] Delamare, G. & Ferrari, U. Time-dependent maximum entropy model for populations of retinal ganglion cells. In *Physical Sciences Forum*, vol. 5, 31 (2022).

[42] Granot-Atedgi, E., Tkačik, G., Segev, R. & Schneidman, E. Stimulus-dependent maximum entropy models of neural population codes. *PLoS computational biology* **9**, e1002922 (2013).

[43] Pho, G. N., Goard, M. J., Woodson, J., Crawford, B. & Sur, M. Task-dependent representations of stimulus and choice in mouse parietal cortex. *Nature communications* **9**, 2596 (2018).

[44] Dadarlat, M. C. & Stryker, M. P. Locomotion enhances neural encoding of visual stimuli in mouse v1. *Journal of Neuroscience* **37**, 3764–3775 (2017).

[45] Christensen, A. J. & Pillow, J. W. Reduced neural activity but improved coding in rodent higher-order visual cortex during locomotion. *Nature communications* **13**, 1676 (2022).

[46] Froudarakis, E. *et al.* Population code in mouse v1 facilitates readout of natural scenes through increased sparseness. *Nature neuroscience* **17**, 851–857 (2014).

[47] Yoshida, T. & Ohki, K. Natural images are reliably represented by sparse and variable populations of neurons in visual cortex. *Nature communications* **11**, 872 (2020).

[48] Brinkman, B. A., Rieke, F., Shea-Brown, E. & Buice, M. A. Predicting how and when hidden neurons skew measured synaptic interactions. *PLoS computational biology* **14**, e1006490 (2018).

[49] Renart, A. & Machens, C. K. Variability in neural activity and behavior. *Current opinion in neurobiology* **25**, 211–220 (2014).

[50] Montijn, J. S., Goltstein, P. M. & Pennartz, C. M. Mouse v1 population correlates of visual detection rely on heterogeneity within neuronal response patterns. *Elife* **4**, e10163 (2015).

[51] Runfeldt, M. J., Sadovsky, A. J. & MacLean, J. N. Acetylcholine functionally reorganizes neocortical microcircuits. *Journal of neurophysiology* **112**, 1205–1216 (2014).

[52] Chen, N., Sugihara, H. & Sur, M. An acetylcholine-activated microcircuit drives temporal dynamics of cortical activity. *Nature neuroscience* **18**, 892–902 (2015).

[53] Reitman, M. E. *et al.* Norepinephrine links astrocytic activity to regulation of cortical state. *Nature Neuroscience* **26**, 579–593 (2023).

[54] Lee, S.-H. & Dan, Y. Neuromodulation of brain states. *Neuron* **76**, 209–222 (2012).

[55] McCormick, D. A., Nestvogel, D. B. & He, B. J. Neuromodulation of brain state and behavior. *Annual review of neuroscience* **43**, 391–415 (2020).

[56] Deco, G., Lynn, C. W., Sanz Perl, Y. & Kringelbach, M. L. Violations of the fluctuation-dissipation theorem reveal distinct nonequilibrium dynamics of brain states. *Physical Review E* **108**, 064410 (2023).

[57] Deco, G. *et al.* The arrow of time of brain signals in cognition: Potential intriguing role of parts of the default mode network. *Network Neuroscience* **7**, 966–998 (2023).

[58] Gilson, M., Tagliazucchi, E. & Cofré, R. Entropy production of multivariate ornstein-uhlenbeck processes correlates with consciousness levels in the human brain. *Physical Review E* **107**, 024121 (2023).

[59] Sekizawa, D., Ito, S. & Oizumi, M. Decomposing thermodynamic dissipation of linear langevin systems via oscillatory modes and its application to neural dynamics. *Physical Review X* **14**, 041003 (2024).

[60] Deco, G., Sanz Perl, Y., Bocaccio, H., Tagliazucchi, E. & Kringelbach, M. L. The insideout framework provides precise signatures of the balance of intrinsic and extrinsic dynamics in brain states. *Communications Biology* **5**, 572 (2022).

[61] Kringelbach, M. L., Perl, Y. S. & Deco, G. The thermodynamics of mind. *Trends in Cognitive Sciences* **28**, 568–581 (2024).

[62] Watanabe, T. *et al.* A pairwise maximum entropy model accurately describes resting-state human brain networks. *Nature communications* **4**, 1370 (2013).

[63] Ezaki, T., Watanabe, T., Ohzeki, M. & Masuda, N. Energy landscape analysis of neuroimaging data. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **375**, 20160287 (2017).

[64] Masuda, N., Islam, S., Thu Aung, S. & Watanabe, T. Energy landscape analysis based on the ising model: Tutorial review. *PLOS Complex Systems* **2**, e0000039 (2025).

[65] Welling, M. The kalman filter. *Lecture Note* 92–117 (2010).

[66] Jong, P. D. & Mackinnon, M. J. Covariances for smoothed estimates in state space models. *Biometrika* **75**, 601–602 (1988).

[67] Rauch, H. E., Tung, F. & Striebel, C. T. Maximum likelihood estimates of linear dynamic systems. *AIAA journal* **3**, 1445–1450 (1965).

[68] Hautus, M. J., Macmillan, N. A. & Creelman, C. D. *Detection theory: A user's guide* (Routledge, 2021).

[69] Olshausen, B. A. & Field, D. J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* **381**, 607–609 (1996).

[70] Olshausen, B. A. & Field, D. J. Sparse coding with an overcomplete basis set: A strategy employed by v1? *Vision research* **37**, 3311–3325 (1997).

[71] Foldiak, P. Sparse coding in the primate cortex. *The handbook of brain theory and neural networks* 895–898 (2003).

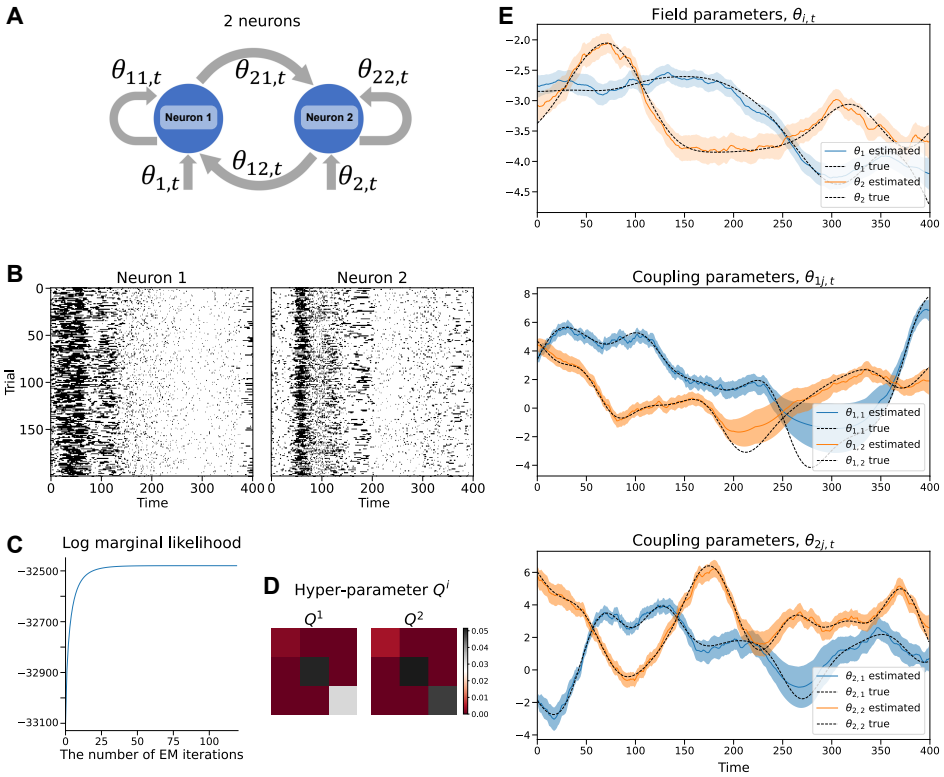

Fig. 1: Application of the state-space kinetic Ising model to two simulated neurons. **A** A schematic of the time-dependent kinetic Ising model for two neurons with field and coupling parameters. The links between the nodes represent the neurons' causal interactions with arrows indicating the time evolution from the past to the present. **B** Raster plots for the two neurons. The vertical axis represents the number of trials, and the horizontal axis shows the number of time bins. **C** The approximate marginal log-likelihood as a function of the iteration steps of the EM algorithm. **D** The optimized hyperparameter $\mathbf{Q}^i$ for neuron 1 (left) and neuron 2 (right). **E** (top) Estimated and true time-dependent field parameters. The solid lines represent the MAP estimates of the field (first-order) parameters obtained from the smoothing posterior, $\boldsymbol{\theta}_{t|T}$. The shaded areas show the 95% credible intervals derived from the diagonal elements of the smoothed covariance matrix, $\mathbf{W}_{t|T}$. The dotted lines are the field parameters from true $\boldsymbol{\theta}_t$ used to generate the data. (middle, bottom) Estimated and true time-dependent coupling (second-order) parameters.

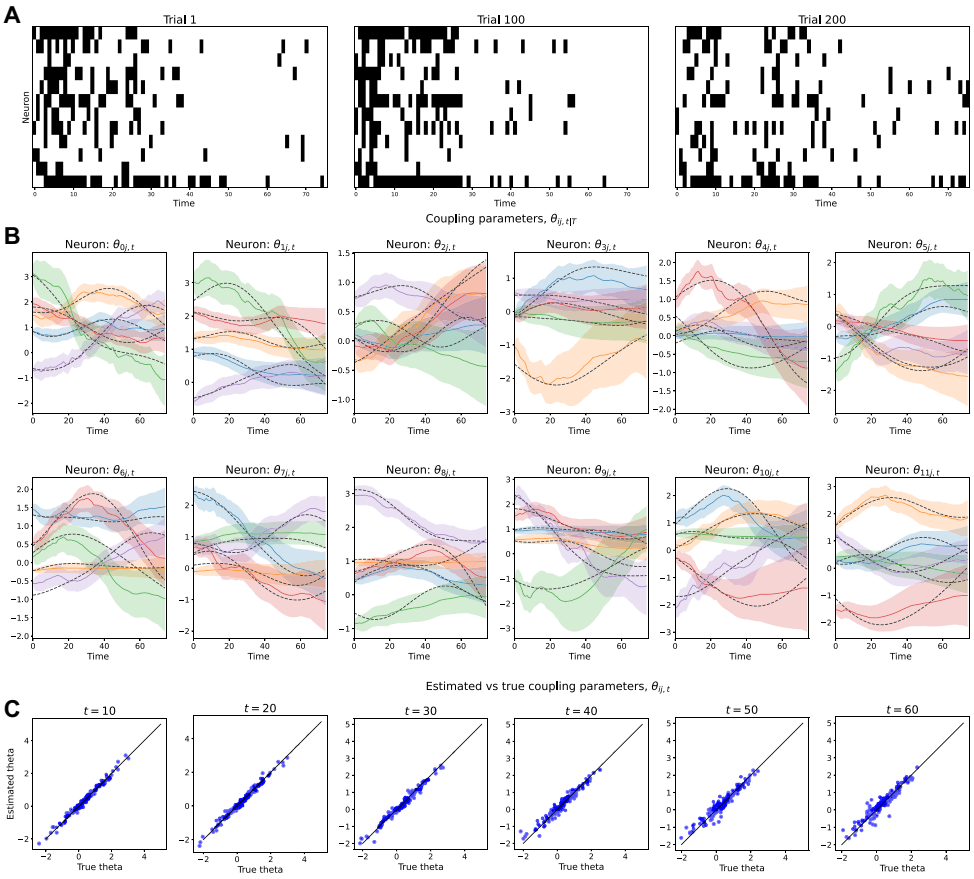

Fig. 2: The application of the state-space kinetic Ising model to 12 simulated neurons. A Simulated spike data for the first, 100th, and last trial out of 200 trials. The vertical axis shows the number of neurons, and the horizontal axis represents the number of bins. **B** Estimated coupling parameters $\boldsymbol{\theta}_{t|T}$ (solid lines), for all neurons and time bins ($i = 1, 2, \ldots, 12$, $t = 1, \ldots, T$). Shaded areas indicate 95% credible intervals, and dashed lines denote the true parameter values used to generate the data. These plots show only the couplings that are significantly deviated from zero: The couplings whose 95% credible interval contains 0 in all bins were excluded. For clarity, only five such significant incoming couplings from other neurons are shown in each panel. **C** Scatter plots comparing the true coupling parameters $\boldsymbol{\theta}_t$ with the estimated values $\boldsymbol{\theta}_{t|T}$ at time $t = 10, 20, \ldots, 60$. The black line is a diagonal line.

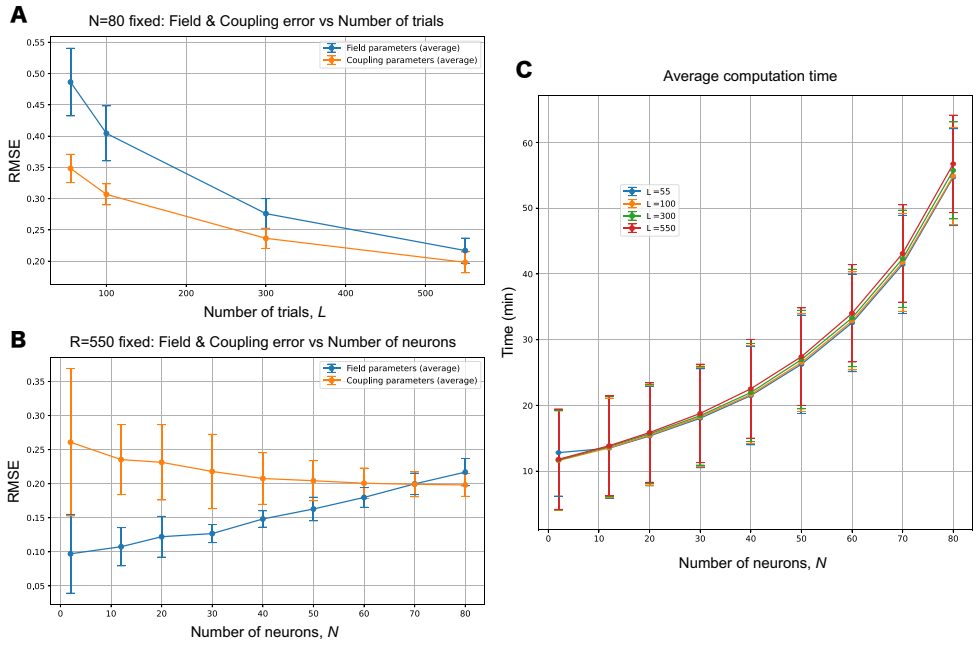

Fig. 3: Estimation error and computational time. **A** Root mean squared error (RMSE) of the field and coupling parameter estimation as a function of trials L, with the number of neurons fixed at $N = 80$. Results are averaged over 10 independent samples, with error bars representing standard deviations. **B** RMSE of the field and coupling parameters as a function of the number of neurons N, with the number of trials fixed at $L = 550$. Averages and standard deviations are computed over 10 independent samples. **C** Average computation time for different numbers of neurons N and trials $L = 55, 100, 300, 550$, with error bars indicating standard deviations. Computation was performed on a Dell PowerEdge R750 server with two Intel Xeon 2.4 GHz CPUs (76 cores / 152 threads).

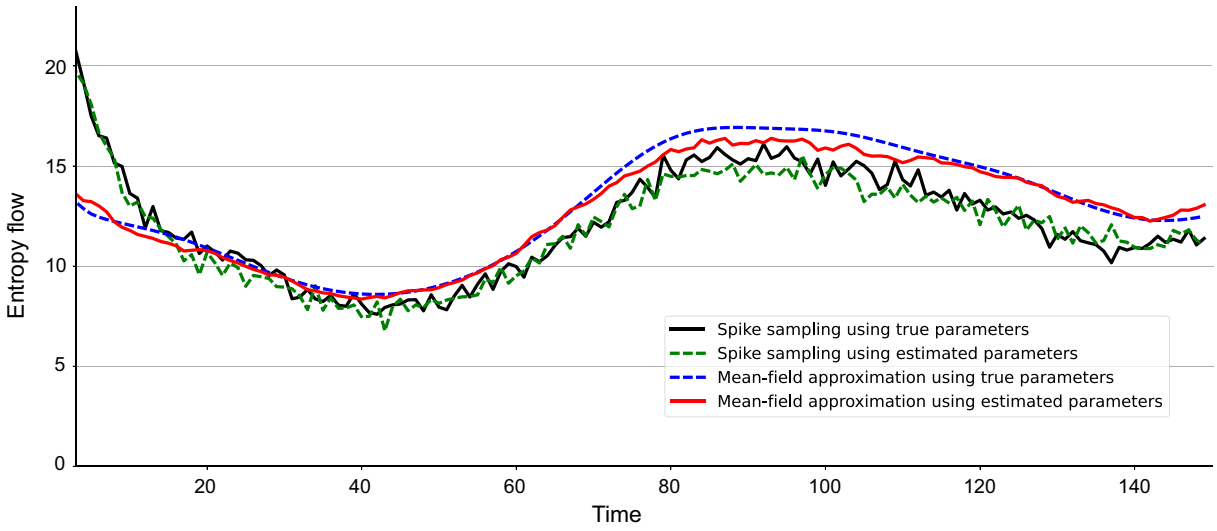

Fig. 4: Comparison of entropy flow estimation methods. Entropy flows estimated using four different approaches: Sampling method with true parameters $\boldsymbol{\theta}_t$ (solid black); sampling method with estimated parameters $\boldsymbol{\theta}_{t|T}$ (dashed green); mean-field method with true parameters $\boldsymbol{\theta}_t$ (dashed blue); and mean-field method with estimated parameters $\boldsymbol{\theta}_{t|T}$ (solid red).

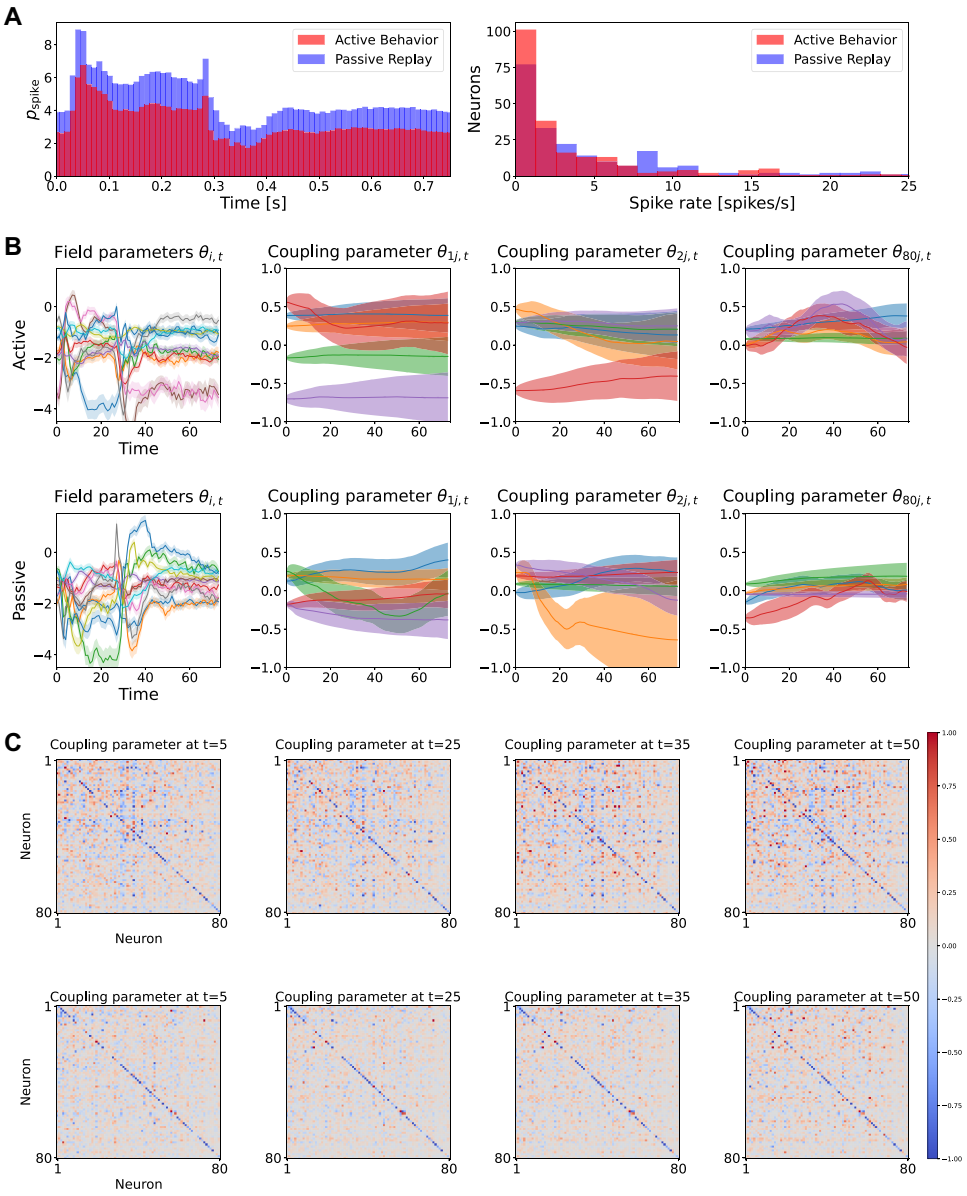

Fig. 5: Estimated neural dynamics under the active and passive conditions for mouse 574078**. A** Spike-rate dynamics and distributions. (Left) Spike-rate averaged across neurons and trials. (Right) Spike-rate distributions of all recorded neurons. **B** Smoothed time–dependent parameters $\boldsymbol{\theta}_{t|T}$ of the kinetic Ising model for the active (top) and passive (bottom) conditions. The first column shows the field $\theta_{i,t}$ (one trace per neuron), and the next three columns show the incoming couplings $\theta_{ij,t}$ for $i = 1, 2, 80$. Solid lines are MAP estimates and shaded areas indicate ± 1 SD (i.e., 68% credible bands) computed from the diagonal of the posterior covariance. For each i, couplings were first screened within the analysis window (bins 21–75) and retained if their credible interval excluded zero at least once in the window (self–couplings excluded). From the retained set, we display the first five couplings per i, ordered by ascending j label for readability. **C** Estimated couplings at $t = 5, 25, 35, 50$ under the active (top) and passive (bottom) conditions.

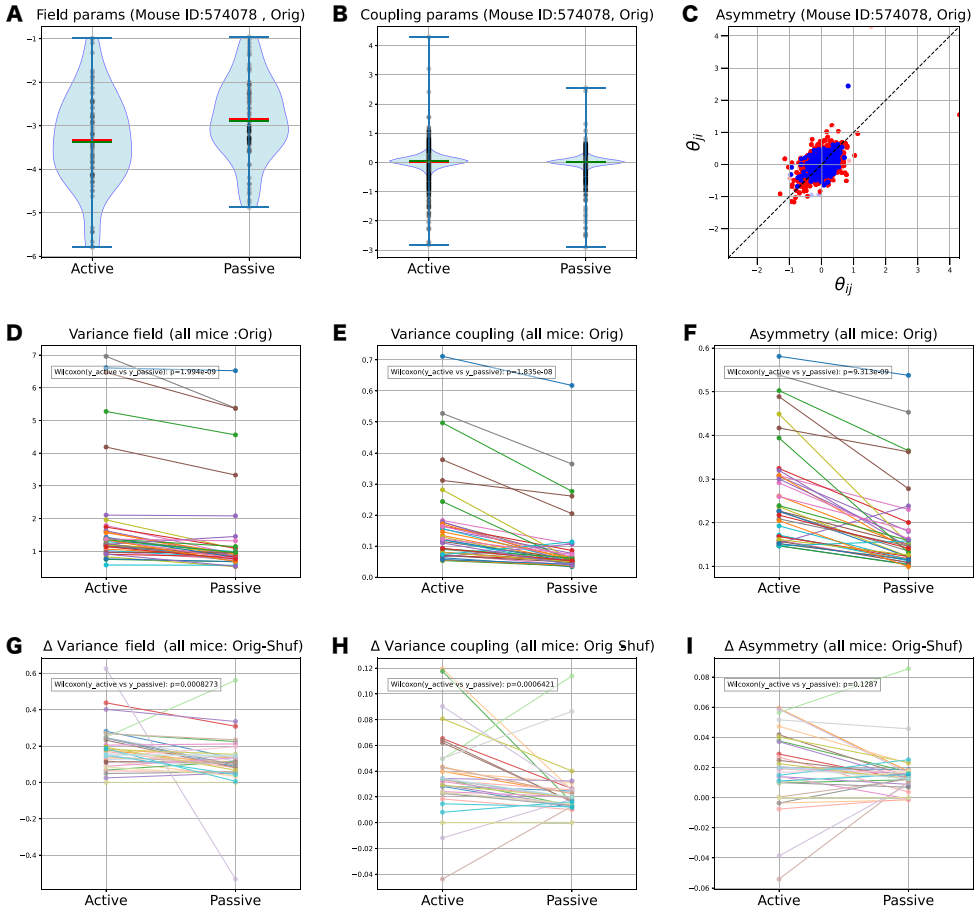

Fig. 6: Variability of estimated model parameters. **A**, **B** Distributions of the time-averaged field values $\bar{\theta}_i$ (A) and the time-averaged couplings $\bar{\theta}_{ij}$ (B) for mouse 574078 under the active and passive conditions. The shaded violin plots depict kernel-density estimates of the empirical distributions; gray dots are the underlying observations (one dot per neuron in A, one per coupling in B). Short horizontal caps at the top and bottom indicate the sample maximum and minimum, respectively. Horizontal red bars mark the mean, while horizontal green bars mark the median. **C**: Scatter plots of coupling strength of reciprocal pairs under the active (red) and passive (blue) conditions for mouse 574078. The coupling asymmetries were 0.147 (active) and 0.105 (passive). The asymmetry was assessed by the average absolute difference of the reciprocal couplings $\langle|\bar{\theta}_{ij} - \bar{\theta}_{ji}|\rangle_{ij}$, where $\bar{\theta}_{ij}$ indicates the time-average of $\theta_{ij,t}$ and $\langle\cdot\rangle_{ij}$ refers to the average over the combinations of i, j. **D-F** Group-level (all mice) comparisons for the original dataset: field variance (D), coupling variance (E), and coupling asymmetry (F). **G-I**: Plots analogous to D-F for shuffle-subtracted parameter variances and coupling asymmetry. Each subplot of D-I contains the p-values of Wilcoxon signed-rank tests for the active vs. passive conditions.

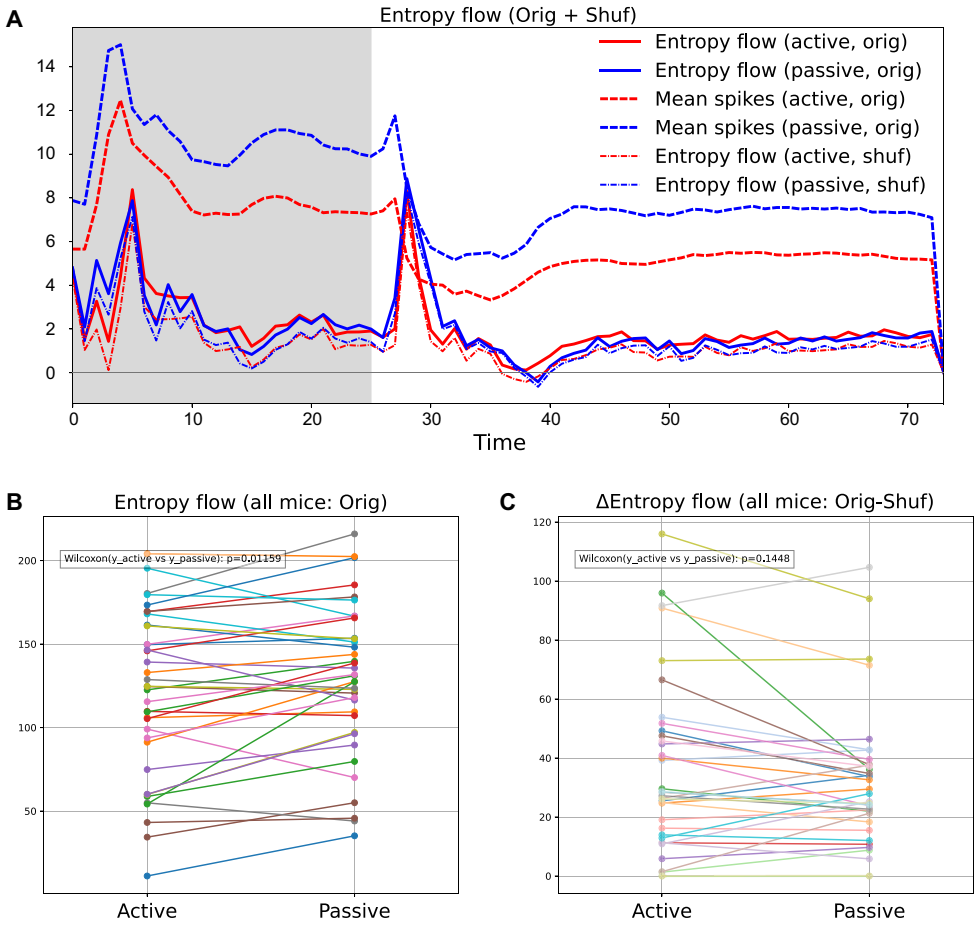

Fig. 7: Estimated entropy flow dynamics. **A** Time courses of entropy flow for the original (solid lines) and shuffled data (dash-dot lines) under the active (red) and passive (blue) conditions. The dashed lines show the corresponding population-averaged spike rates. **B** Total entropy flows summed across all time bins for each mouse under active and passive conditions. Data from the same mouse are connected by a line. **C** Shuffled-subtracted total entropy flow (original - shuffle), shown for each mouse under the active and passive conditions.

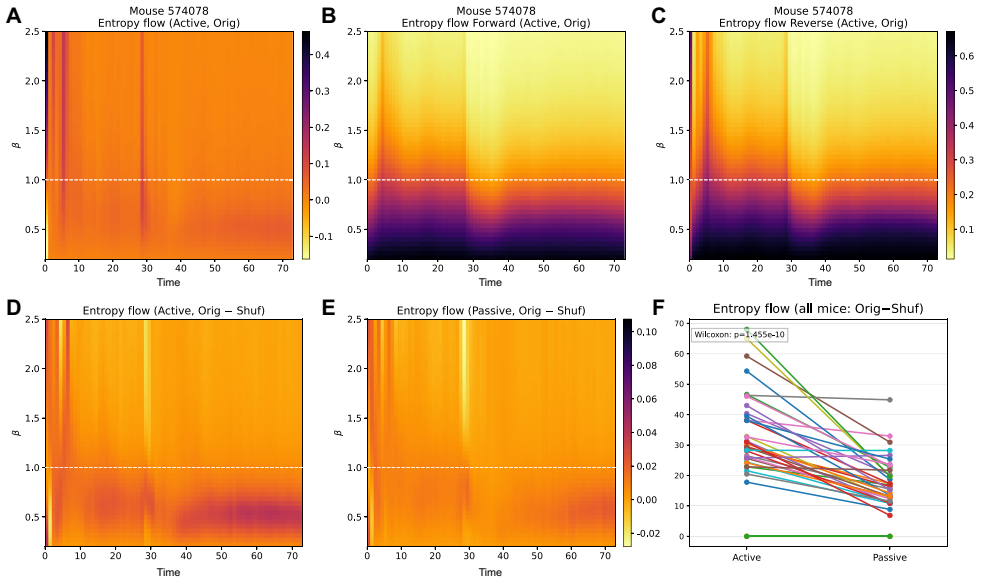

Fig. 8: Model-based perturbation analysis A Entropy flow σ_t of mouse 574078 in the active condition, computed after rescaling the fitted parameters as $\boldsymbol{\theta} \to \beta\boldsymbol{\theta}$. The dashed line indicates $\beta = 1$. **B** Forward entropy flow, $\sigma_t^{\text{forward}}$. **C** Backward entropy flow, $\sigma_t^{\text{backward}}$. **D** Shuffle-subtracted entropy flow in the active condition, isolating the contribution of interactions beyond firing rate dynamics. Entropy flow driven by interactions peaks at $\beta < 1$. **E** Shuffle-subtracted entropy flow in the passive condition. **F** Comparison of shuffle-subtracted entropy flow between active and passive conditions across all mice, showing significantly higher values in the active condition (Wilcoxon test, $p = 1.455 \times 10^{-10}$). Shuffle-subtracted entropy flow is obtained over a low–gain range $\beta \in [0.2, 1.0]$ and across all bins. Lines connect active (left) to passive (right) for each mouse.

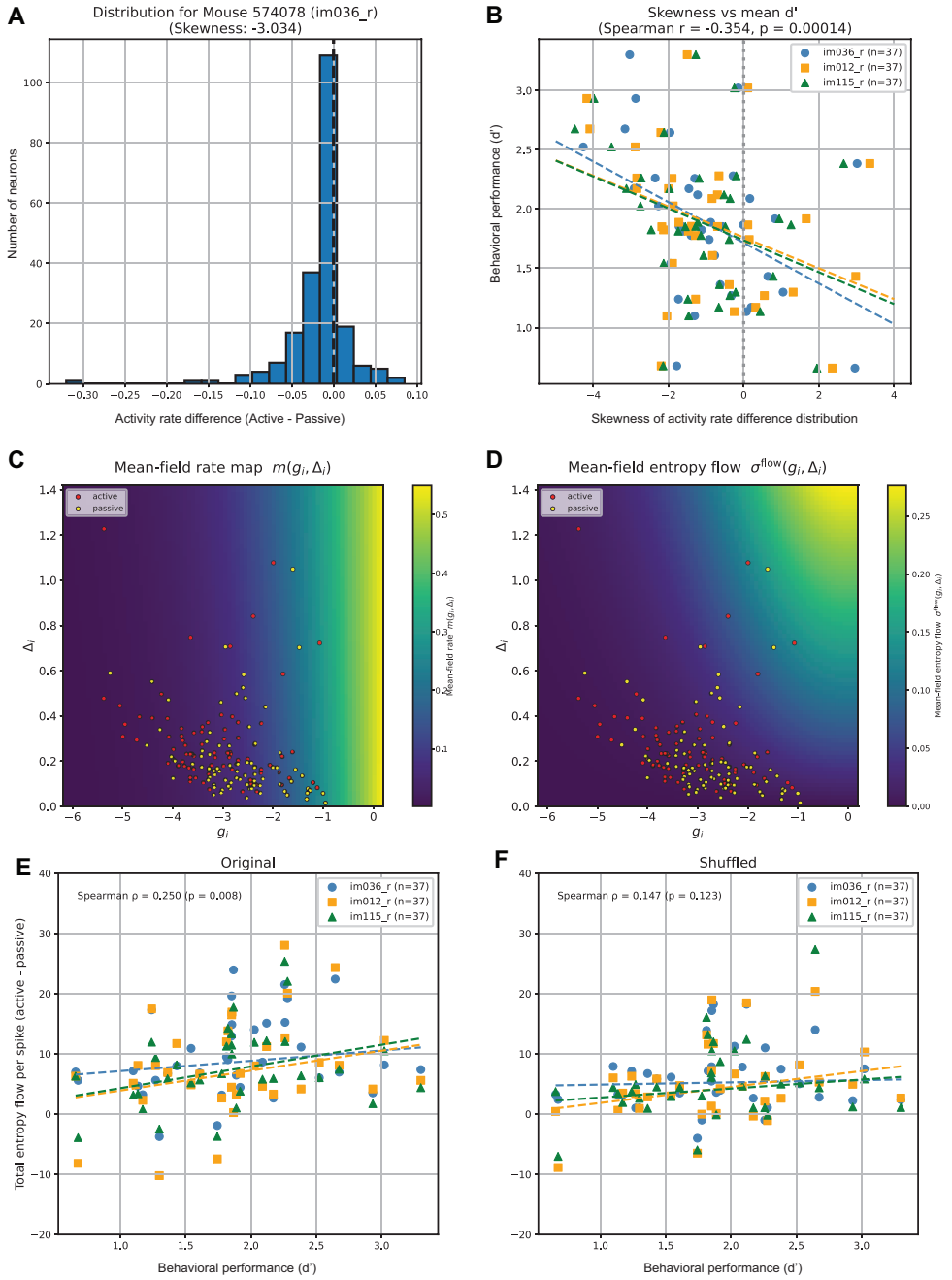

Fig. 9: Correlations between neural dynamics and behavioral performance. **A** A histogram of firing-rate differences (active–passive) across neurons for mouse 574078 with an image im036_r. A skewness was used as a sparsification index. **B** The skewness vs behavioral performance (mean d-prime) for 37 mice with three images. **C** Mean-field rates of individual neurons (colored circles) as a function of time-average mean g_i and variability Δ_i of their inputs for mouse 574078. The background color indicates the theoretical mean-field rate under the steady-state (Eq. 1.20). **D** Mean-field entropy flow of individual neurons (colored circles) as a function of time-average mean and variability of their inputs (mouse 574078). The background color shows theoretical entropy flow under the steady-state (Eq. 1.21). **E** Entropy flow normalized by activity rate vs behavioral performance for 37 mice with three images. **F** Entropy flow normalized by activity rate obtained from trial-shuffled data vs behavioral performance. In E and F, there is one outlier mouse below −20 in the ordinate, which was included in the statistical analysis.

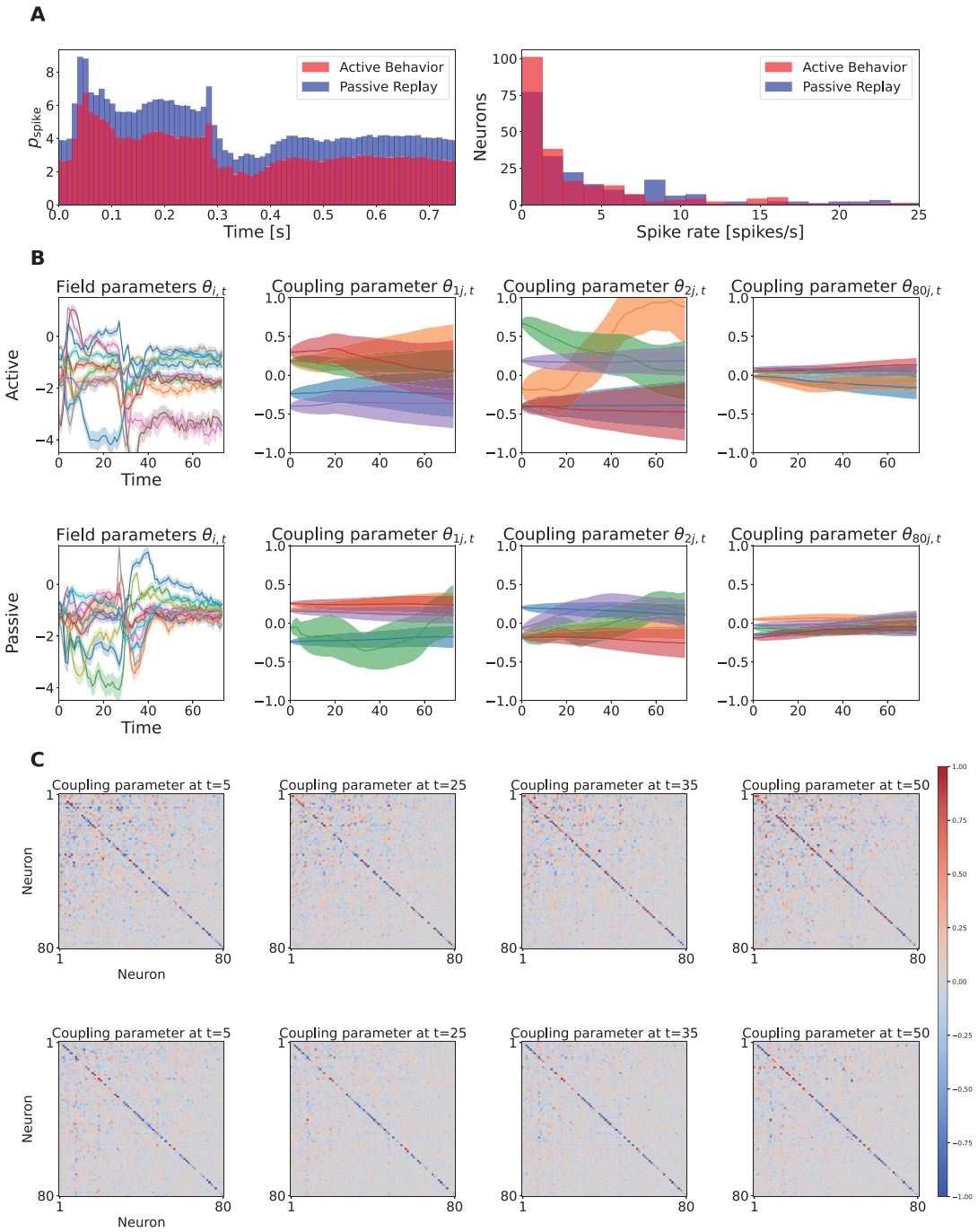

Figure S1. **Estimated neural dynamics under the active and passive conditions for mouse 574078.**

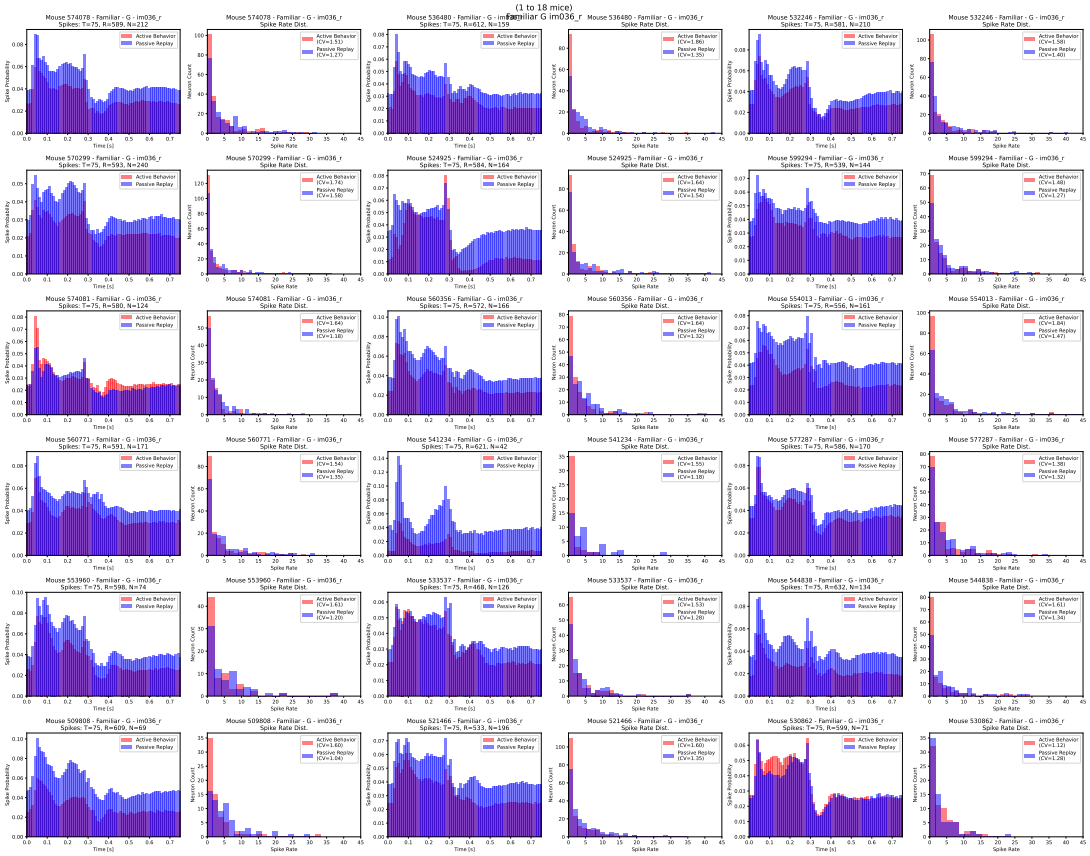

Figure S2 **Spike-rate dynamics and distributions for mice 1-18.** Spike-rate dynamics and distributions under the active (red) and passive (blue) conditions. The presentation styles for each mouse follow Fig. 5**A**. The mice were listed in descending order of behavioral performance measured by d-prime. See Supplementary S3 for the remaining mice.

FIGURES

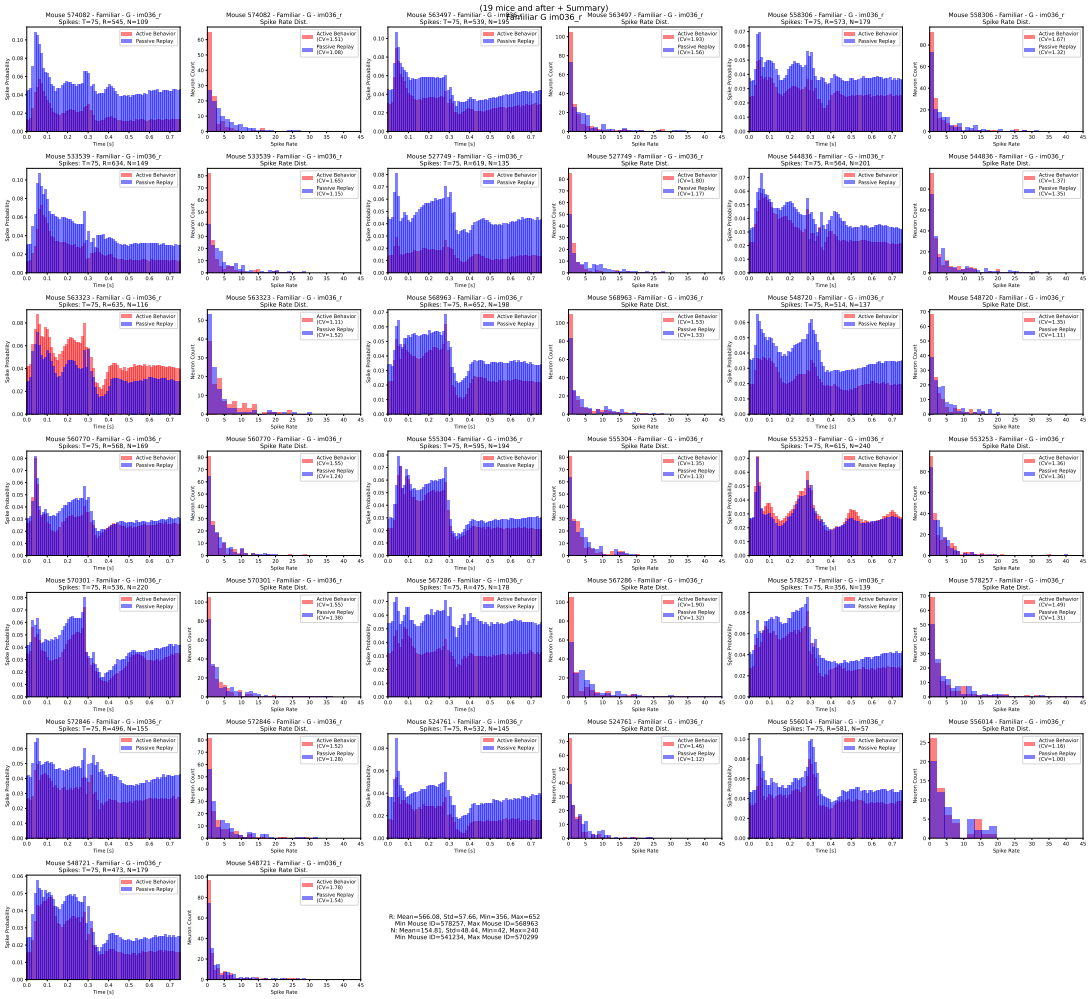

Figure S3 **Spike-rate dynamics and distributions for mice 19-37.** The same as in Supplementary S2 but for the remaining 19 mice.

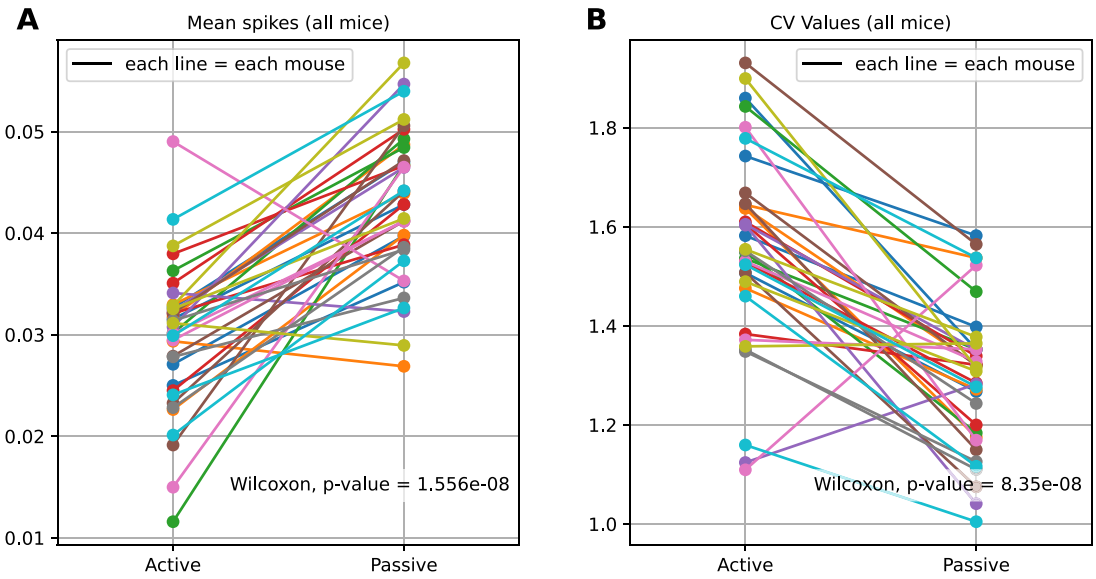

Figure S4 **Comparison of mean spiking probability and coefficient of variation in the active and passive conditions. A** Mean spiking probability across all bins, trials, and neurons in active and passive conditions. Each line represents the same mouse. Neurons showed significantly lower firing rates in the active condition ($p = 1.556 \times 10^{-8}$, Wilcoxon signed-rank test). **B** Coefficient of variations (CVs) of the firing rate distributions, a measure of sparseness, in the active and passive conditions. CV was significantly higher in the active condition ($p = 8.35 \times 10^{-8}$, Wilcoxon signed-rank test).

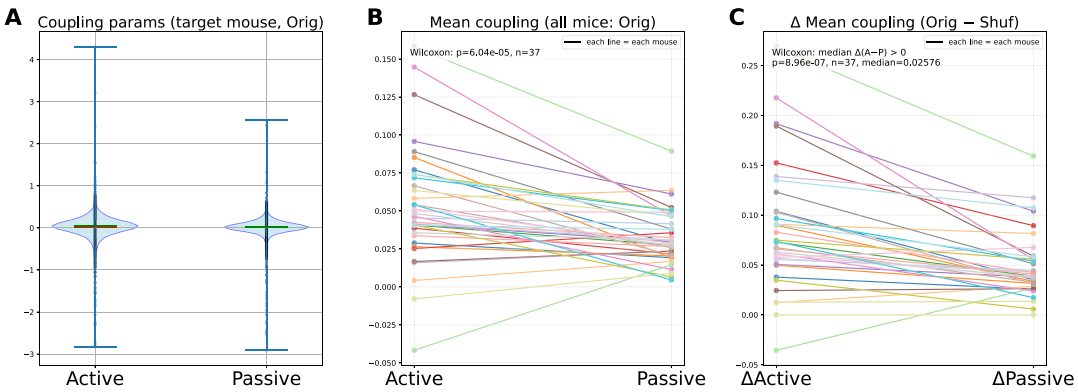

Figure S5 **Mean effective coupling and shuffle control.** **A** Violin plots of time-averaged effective couplings for mouse 574078 under the active and passive conditions. Horizontal bars indicate the mean (red) and median (green); points show individual entries. **B** Population summary of the per-mouse mean coupling in the original data; each line connects the active and passive values from the same mouse. Panel annotations report p-values and sample size (n) from Wilcoxon signed-rank tests across mice (two-sided). **C** Shuffle-adjusted means, where for each mouse the value in each condition is computed as (Original − Shuffle); lines connect paired values. The annotation reports a one-sided Wilcoxon signed-rank test assessing whether the median of {(Original − Shuffle) in Active} minus {(Original − Shuffle) in Passive} is greater than zero. Together, the results indicate that the mean effective coupling is larger in the active condition than in the passive condition, and that this increase persists after shuffle correction.

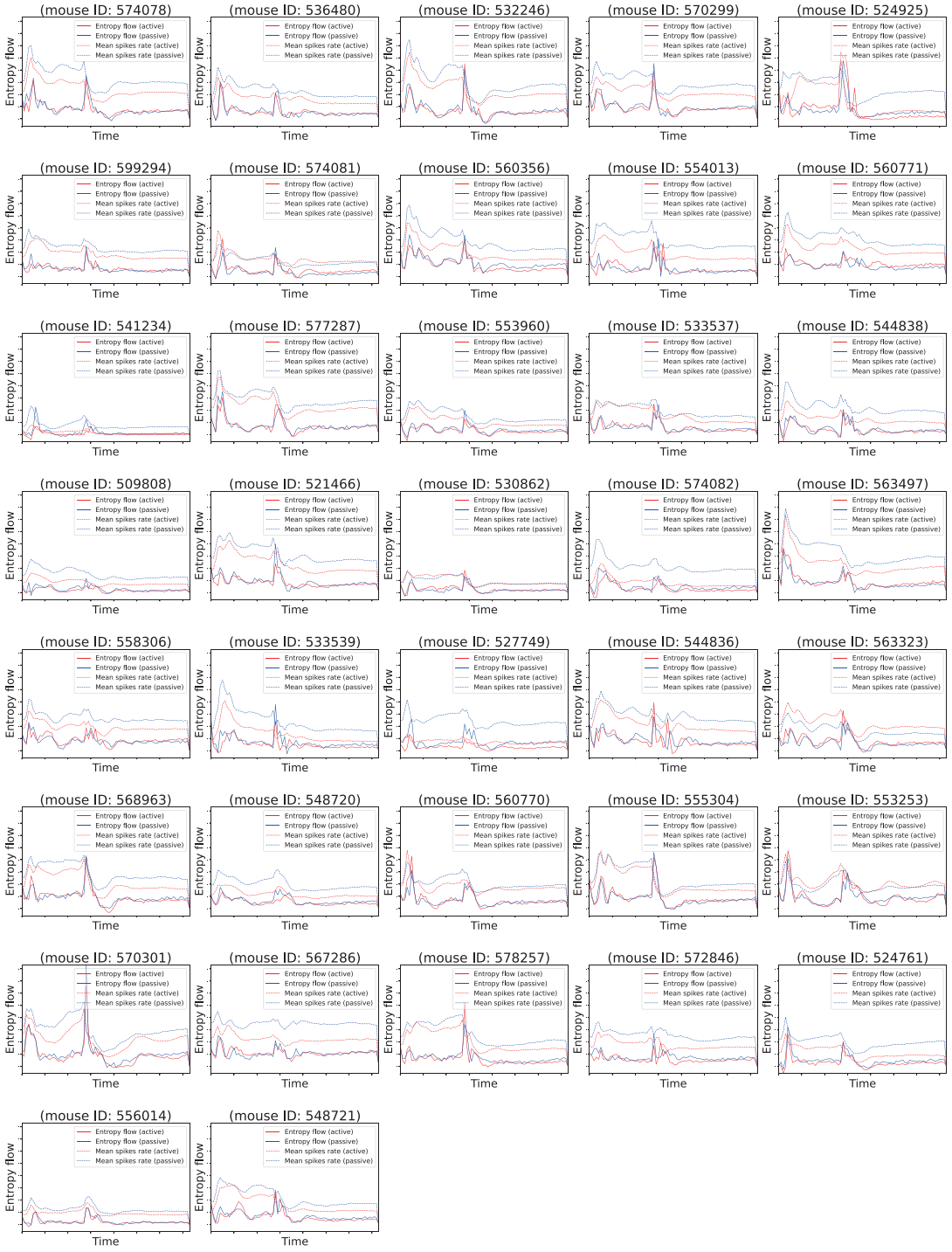

Figure S6 **Time courses of entropy flow and mean spike rates for each mouse under active and passive conditions.** Each subplot represents the dynamics of an individual mouse. Solid lines are entropy flows (red for active, blue for passive) while dashed lines represent the average population spike rate (red for active, blue for passive).

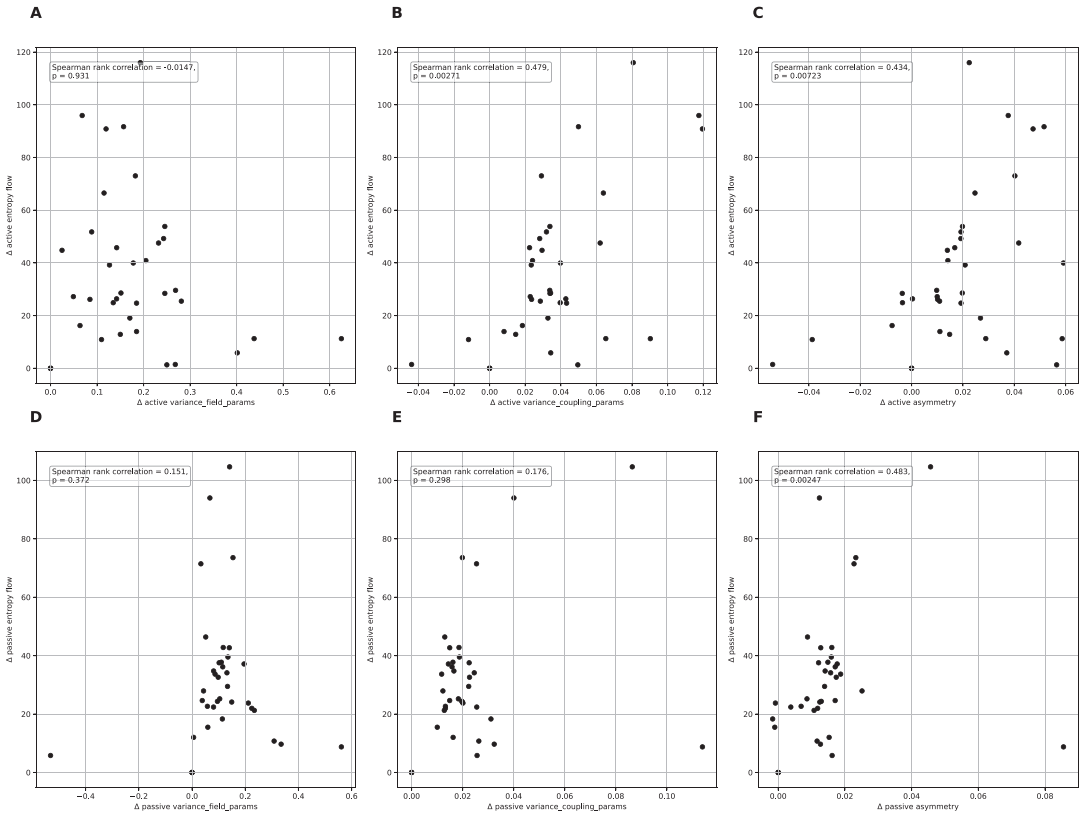

Figure S7 **Comparison of shuffle-subtracted parameter variabilities and coupling asymmetry with entropy flow for all mice.** Each row represents comparisons of parameter variabilities and coupling asymmetry (calculated by subtracting the shuffled-data estimate of the variance from the original-data estimate) and their relationship to the shuffle-subtracted entropy flow. **A, B, C** “ Δactive ” (shuffle-subtracted changes in the field, coupling variabilities, and coupling asymmetry) versus the shuffle-subtracted entropy flow in the active state. **D, E, F** “ Δpassive ” versus the shuffle-subtracted entropy flow in the passive state.

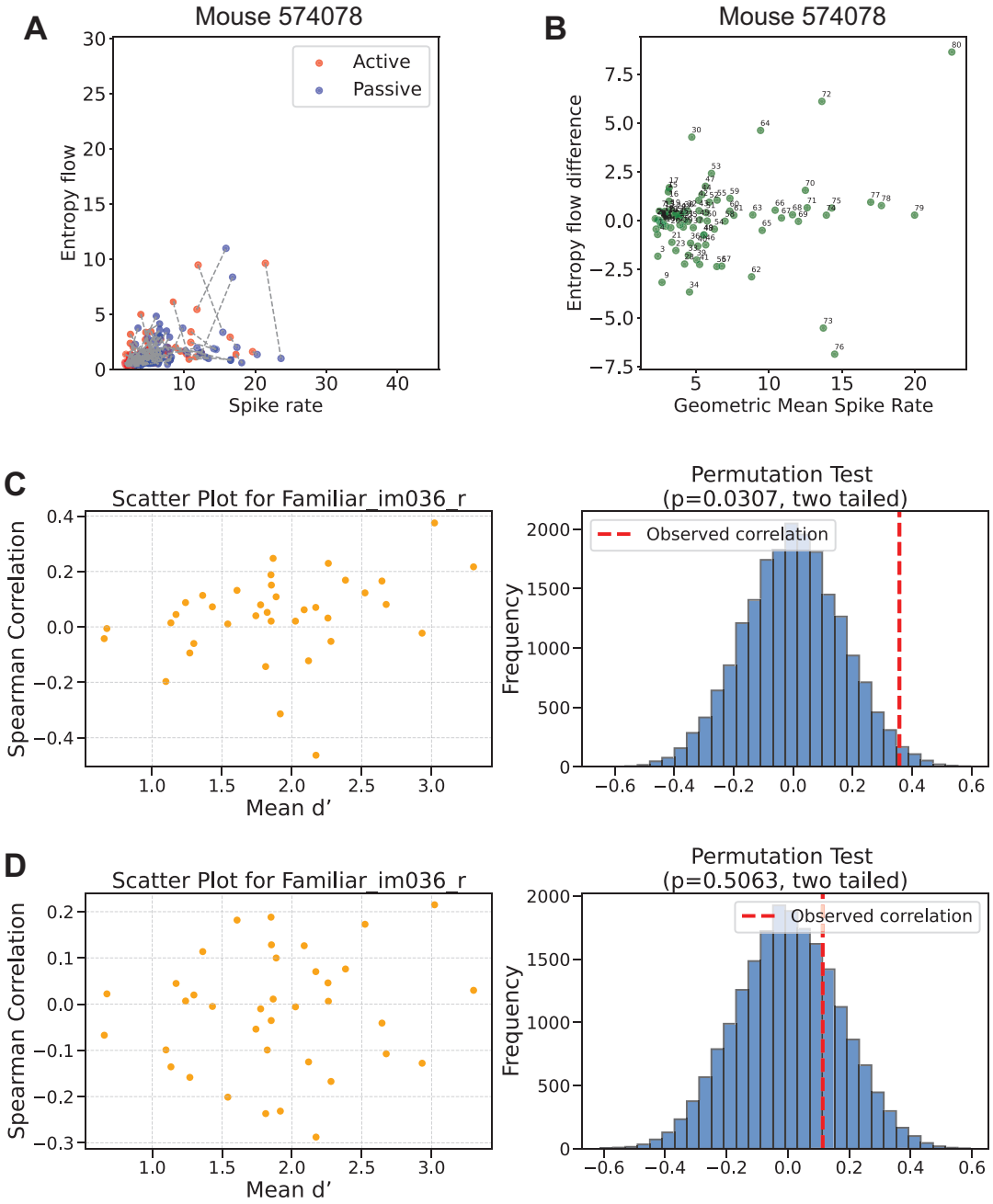

Figure S8 **Relating the dependency of entropy flow change of individual units on firing rates with behavioral performance. A** Mean spike rate vs entropy flow per individual unit under the active and passive conditions (mouse 574078). Dashed lines connect values for the two conditions, highlighting behavioral state-dependent changes. **B** Geometric mean spike rate (abscissa) vs differences in entropy flow (active - passive, ordinate) for individual units. The positive Spearman correlation coefficient ($\rho = 0.22$) suggests that units with higher spike rates increased entropy flow in the active condition. **C** (Left) Scatter plot of behavioral performance (mean d-prime) vs. the Spearman rank correlation between the geometric mean rate and entropy flow change of individual units. Each dot represents a single mouse. The dependency in this scatter plot was assessed again by the Spearman rank correlation coefficient, yielding $\rho = 0.3578$. (Right) A permutation test comparing the observed correlation value ρ with those obtained from the surrogate data. A statistically significant positive relationship was observed ($p = 0.0304$, two-tailed). The surrogate data was constructed by permuting the values of mean d-prime. **D** Results for trial-shuffled data.