Estimating nonstationary input signals from a single neuronal spike train

Hideaki Kim* and Shigeru Shinomoto†

Department of Physics, Graduate School of Science, Kyoto University, Sakyo-ku, Kyoto 606-8502, Japan
(Received 15 July 2012; published 2 November 2012)

Neurons temporally integrate input signals, translating them into timed output spikes. Because neurons nonperiodically emit spikes, examining spike timing can reveal information about input signals, which are determined by activities in the populations of excitatory and inhibitory presynaptic neurons. Although a number of mathematical methods have been developed to estimate such input parameters as the mean and fluctuation of the input current, these techniques are based on the unrealistic assumption that presynaptic activity is constant over time. Here, we propose tracking temporal variations in input parameters with a two-step analysis method. First, nonstationary firing characteristics comprising the firing rate and non-Poisson irregularity are estimated from a spike train using a computationally feasible state-space algorithm. Then, information about the firing characteristics is converted into likely input parameters over time using a transformation formula, which was constructed by inverting the neuronal forward transformation of the input current to output spikes. By analyzing spike trains recorded in vivo, we found that neuronal input parameters are similar in the primary visual cortex V1 and middle temporal area, whereas parameters in the lateral geniculate nucleus of the thalamus were markedly different.

DOI: 10.1103/PhysRevE.86.051903 PACS number(s): 87.19.lo, 05.40.—a, 02.50.—r, 89.70.Cf

I. INTRODUCTION

Multichannel recording techniques have provided a great deal of information about the activities and functional connectivity of neuronal ensembles [1–3]. Yet further analyses are required to obtain a full understanding of computations conducted by mammalian neuronal circuits. One approach is to estimate presynaptic inputs from the spike trains of postsynaptic neurons (Fig. 1). Currently available techniques to examine synaptic inputs, including intracellular or patch recording [4–9], have some disadvantages, such as the need for anesthesia, constraints on animal behavior, damage to recorded neurons, and limited recording time [10–13]. Therefore, analytic techniques are needed to allow analysis of presynaptic inputs based on spike trains recorded from a neuron in vivo for extended periods of time.

Generally, problems of estimating inputs from output signals are ill posed. In the case of neuronal signal transformation, however, a large number of randomly arriving input spikes make it possible to extract some information from an output spike train; a number of irregular synaptic inputs result in uncorrelated fluctuations with means and amplitudes, which can be translated into the activities of presynaptic excitatory and inhibitory neuronal populations. Mathematical methods have been developed, assuming that presynaptic neuronal activities are constant over time [14–17].

Stationary input conditions for standard spiking mechanisms, however, cannot account for spiking statistics of cortical neurons in vivo [18]. To resolve this inconsistency, some studies have suggested introducing nonstationary or correlated fluctuations to the inputs [19,20]. Recently, to analyze situations in which input parameters comprising the mean and the amplitude of uncorrelated fluctuations vary in time, several methods have been proposed using time-dependent stimulus traces [21–23] or averaging over repeated trials [24]. Inputs to individual neurons, however, are not entirely controlled by animal behavior or reproduced exactly under identical behavioral conditions, but instead may fluctuate from trial to trial. Therefore, it is essential to develop a method that is capable of estimating input parameters in a single trial.

In this paper, we constructed a method for estimating nonstationary inputs from a single spike train. We first incorporated an input estimation method, formulated based on the assumption of constant input parameters, into a state-space model, allowing both the mean input and the amplitude of uncorrelated fluctuations to vary over time. This method was computationally complex and the analysis was limited to trains of up to hundreds of spikes. To make the estimation practical for larger data sets, we then developed a method to transform the neuronal firing characteristics into input parameters, from which we can estimate activities in populations of excitatory and inhibitory presynaptic neurons (Fig. 2). The instantaneous firing characteristics consisting of the spike rate and non-Poisson irregularity were estimated using a state-space method that can efficiently process much larger sets of data [25]. We then constructed a formula to transform the firing characteristics into the input parameters; a two-dimensional interpolation formula was obtained by inverting the forward transformation from input to output signals (Fig. 3). This transformation method does not require complex computational analyses.

The method was tested against synthetic data obtained from a simulation of neuronal spiking under fluctuating input parameters. The method was then used to examine temporal variations in input parameters by analyzing in vivo spike trains recorded from the primary visual cortex (V1), middle temporal area (MT), and lateral geniculate nucleus (LGN) of monkeys.

II. MODEL OF A SPIKING NEURON

To estimate inputs based on a train of evoked spikes, we need a spiking neuron model that mimics the neuronal
With the knowledge of EPSPs and IPSPs, the mean drift and uncorrelated fluctuation are related to the rates of excitatory and inhibitory input spikes $r_E$ and $r_I$ \cite{34,36} as follows:

$$R\mu/\tau = a_E r_E - a_I r_I, \quad (R\sigma/\tau)^2 = a_E^2 r_E^2 + a_I^2 r_I^2,$$

(3)

where $a_E$ and $a_I$ are the unitary EPSP and IPSP, respectively. The first equation represents that the mean input is given by the average EPSP subtracted by the average IPSP, and the second equation implies that the input fluctuation is given by the sum of random noisy bombardment of EPSPs and IPSPs.

The linear relation (3) can be inverted so that activities in the populations of excitatory and inhibitory presynaptic neurons can be estimated from the input parameters $\mu$ and $\sigma$:

$$r_E = \frac{(R\sigma/\tau)^2 + a_I (R\mu/\tau)}{a_E (a_E + a_I)},$$

and

$$r_I = \frac{(R\sigma/\tau)^2 - a_E (R\mu/\tau)}{a_I (a_E + a_I)}.$$

(4)

III. A METHOD TO DIRECTLY ESTIMATE INPUT PARAMETERS

The probability that a neuron generates output spikes at times $\{t_j\}_{j=0}^n = \{t_0, t_1, \ldots, t_n\}$ can be obtained as the conditional distribution function given input parameters $\Lambda$. Here, we represented a set of such input parameters as a two-dimensional vector, and modeled them as varying over time: $\Lambda = \Lambda(t)$. For Stein’s model, the input parameters were defined as the mean and the amplitude of uncorrelated fluctuation $\Lambda(t) = [\mu(t), \sigma(t)]$. The input parameters can be estimated from the spike times using the Bayes theorem,

$$P(\Lambda(t)|[t_j]) = \frac{P([t_j]|\Lambda(t))P(\Lambda(t))}{P([t_j])}.$$

(5)

To estimate the probability of an output spike based on variable input parameters, we approximated the input parameters as constant during each ISI as follows:

$$P([t_j]|\Lambda) = \prod_{j=1}^n P(s_j|\Lambda),$$

(6)

where $s_j \equiv t_j - t_{j-1}$ is the jth ISI and $\Lambda_j$ represents the input parameters at the time of the $j$th spike $\Lambda(t_j)$. It should be noted that the constancy of the input parameters does not mean that the input current was constant; the input current fluctuated rapidly with amplitude $\sigma$, reflecting bombardment of a number of input spikes from presynaptic populations of excitatory and inhibitory neurons.

To model the prior distribution of the input parameters represented by the two-dimensional vector $\Lambda(t) \equiv \{\Lambda^1(t), \Lambda^2(t)\}$, we incorporate the tendency to vary slowly by penalizing large gradients:

$$P(\Lambda(t)) = \prod_{j=1}^n P_\gamma(\Lambda_j|\Lambda_{j-1})$$

$$= \prod_{j=1}^n \prod_{k=1}^2 N(\Lambda^k_{j-1}, \gamma^k s_j),$$

(7)
any given set of input parameters to estimate the ISI distribution function. A major cause for this limitation is the need to solve a complex integral equation with a high-performance computer. A major cause for this limitation is the need to solve a complex integral equation with a high-performance computer.

Given a set of spike times, maximization of the marginal likelihood, where \( y \equiv (y^1, y^2) \) is a hyperparameter representing the stationarity of the input parameters, and \( N(x, y) \) is the Gaussian distribution with mean \( x \) and variance \( y \). The initial parameter \( \Lambda_0 \equiv (\Lambda_0^1, \Lambda_0^2) \) was set to the value that has been estimated on the basis of the assumption that input parameters are constant over time. This is identical to input parameters exhibiting a random walk; the variance should therefore be rescaled with the ISI \( s_j \).

Based on the empirical Bayes method, the hyperparameter \( \gamma \equiv (y^1, y^2) \) was determined by maximizing the marginal likelihood,

\[
P_y(t_j) = \prod_{j=1}^{n} \int d\Lambda_j \int d\Lambda_j P(s_j | \Lambda_j) P_y(\Lambda_j | \Lambda_{j-1}). \tag{8}
\]

Given a set of spike times, maximization of the marginal likelihood function can be performed using the expectation maximization (EM) algorithm \([38,39] \). The details are given in Appendix A. When using the state-space method to assess such input parameters as the mean and fluctuation of the input current, we found that the process was computationally complex and only feasible with up to hundreds of spikes even with a high-performance computer. A major cause for this limitation is the need to solve a complex integral equation \([21,40,41] \) to estimate the ISI distribution function \( P(s | \Lambda) \) for any given set of input parameters \( \Lambda = (\mu, \sigma) \).

**IV. TRANSFORMATION METHOD**

To perform this estimation with a larger data set, we developed a computationally feasible state-space method and transformed the information into the mean and fluctuation of the input current. Based on a previous report \([25] \), we approximated the ISI distribution \( P(s | \Lambda) \) with the following gamma distribution:

\[
g(s | \lambda, \kappa) \propto (\kappa \lambda s)^{\kappa-1} \exp(-\kappa \lambda s), \tag{9}
\]

where \( \lambda \) and \( \kappa \) are the scale factor and shape factor representing the mean firing rate and irregularity, respectively. A Poisson random firing is characterized by the exponential distribution of ISIs, which is obtained by setting \( \kappa = 1 \). A deviation of the shape factor from \( \kappa = 1 \) represents non-Poisson irregularity; \( \kappa > 1 \) and \( \kappa < 1 \) indicate non-Poisson regular firing and burst firing, respectively. We used a previously described state-space method to estimate these firing characteristics, which is computationally efficient, and is able to handle larger data sets \([25] \). For each spike train, we can obtain the maximum a posteriori (MAP) estimates of the firing rate \( \hat{\lambda}(t) \) and irregularity \( \hat{k}(t) \).

Then we converted the information into the likely input mean \( \hat{\mu}(t) \) and variation \( \hat{\sigma}(t) \) using a transformation formula given by a set of maps (Fig. 3):

\[
\mu = M(\lambda, \kappa), \quad \sigma = S(\lambda, \kappa). \tag{10}
\]
The first transformation is given by the linear relation (3), and the second one is given by the nonlinear relation (11). According to the latter nonlinear transformation, a circular area on a plane of \((\mu, \sigma)\) is mapped into a deformed area on a plane of \((\lambda, \kappa)\). (b) Reverse transformations from firing characteristics to likely input parameters, and to likely presynaptic neuronal activities. The first transformation is given by the linear relation (4), and the second one is given by the nonlinear relation (10) (see also Table I). A circular area on a plane of \((\lambda, \kappa)\) is mapped into a deformed area on a plane of \((\mu, \sigma)\).

This transformation formula was constructed by inverting the neuronal forward transformation of input signals to output spiking, which was given by

\[
\lambda = L(\mu, \sigma), \quad \kappa = K(\mu, \sigma).
\]

The forward transformation was obtained by fitting the gamma distribution \(g(s|\lambda, \kappa)\) to the ISI distribution of the spiking neuron model \(P(s|\mu, \sigma)\) by minimizing the Kullback-Leibler (KL) divergence. The fitting was obtained by solving the relationship

\[
\int_0^\infty ds P(s|\mu, \sigma) = 1/\lambda,
\]

\[
\int_0^\infty s \log s P(s|\mu, \sigma) - \log \left(\int_0^\infty ds P(s|\mu, \sigma)\right) = \psi(\kappa) - \log(\kappa),
\]

where \(\psi(\kappa)\) is the digamma function. The details are given in Appendix B.

In practice, the transformation formula was obtained in two steps. First, we tested a huge number of possible input parameters \(\mu\) and \(\sigma\) to estimate the firing characteristics \(\lambda\) and \(\kappa\), with an empirical ISI distribution function \(P(s|\mu, \sigma)\) obtained from the OUP simulation [Fig. 3(a)] (for details, see Appendix C). Next, we inverted the relationship from Eq. (11) into Eq. (10) [Fig. 3(b)]. Detailed forms of the polynomial functions approximating Eq. (10) are presented as a usable formula (see Appendix C).

V. RESULTS

We assessed the validity of the estimation methods developed above by examining synthetic data, and using the method to analyze spike trains obtained in vivo.

A. Testing the estimation methods with synthetic data

To compare the direct and transformation methods of input estimation, we generated synthetic data by simulating spiking in the neuron model given nonstationary input parameters comprising the mean and fluctuation of the input current. The model parameters were chosen to match those that were incorporated into the input estimator. We made the following three types of input parameters given by the mean input \(\mu(t)\) and the amplitude of uncorrelated fluctuation \(\sigma(t)\).

(i) The mean current was modulated sinusoidally with a period \(T\), while the amplitude of fluctuation was constant:

\[
\mu(t) = \mu_0 + \delta \mu \sin(2\pi t/T), \quad \sigma(t) = \sigma_0.
\]

(ii) The mean was held constant, while the amplitude of fluctuation was modulated sinusoidally:

\[
\mu(t) = \mu_0, \quad \sigma(t) = \sigma_0 + \delta \sigma \sin(2\pi t/T).
\]

(iii) Both the mean and the amplitude of fluctuation were modulated sinusoidally with the phase shifted by \(\Delta\):

\[
\mu(t) = \mu_0 + \delta \mu \sin(2\pi t/T), \quad \sigma(t) = \sigma_0 + \delta \sigma \sin(2\pi t/T - \Delta).
\]

We then examined spike trains derived from the leaky integrate-and-fire (LIF) model (cf. Sec. II) under the various input parameters with both the direct estimation method and the transformation method. Sample estimations are denoted as \(\hat{\mu}(t)\) and \(\hat{\sigma}(t)\) (Fig. 4). Both the direct and transformation methods have provided an equally accurate estimation. However, the direct method required extensive computation spending several hours even for analyzing a short train of several hundreds of spikes, while the transformation method swiftly completed the computation.

Using the computationally tractable transformation method, we evaluate the goodness of the input estimation in terms of the integrated squared error (ISE) between the intended input parameters and the estimated parameters:

\[
\int_0^\Omega dt \left\{ [\mu(t) - \hat{\mu}(t)]^2 + [\sigma(t) - \hat{\sigma}(t)]^2 \right\},
\]

where \(\Omega\) or the entire observation interval is set to 50 s. The ISEs for the three types of input parameters were plotted...
FIG. 4. (Color online) Three fluctuating input parameters and estimations from sample spike trains. (a) Three input parameters (from left to right): (i) the mean $\mu(t) = \mu_0 + \delta \mu \sin(2\pi t/T)$ and the amplitude of fluctuation $\sigma(t) = \sigma_0$; (ii) $\mu(t) = \mu_0$ and $\sigma(t) = \sigma_0 + \delta \sigma \sin(2\pi t/T - \Delta)$; and (iii) $\mu(t) = \mu_0 + \delta \mu \sin(2\pi t/T)$ and $\sigma(t) = \sigma_0 + \delta \sigma \sin(2\pi t/T - \Delta)$. Numerical parameters: $\mu_0 = 0.5 \text{nA}$, $\delta \mu = 0.15 \text{nA}$, $\sigma_0 = 1 \text{nA ms}^{-1/2}$, $\delta \sigma = 0.6 \text{nA ms}^{-1/2}$, $T = 2.5 \text{s}$, and $\Delta = \pi/2$. (b) Sample spike trains. (c) The estimated firing rate $\hat{\lambda}(t)$ and irregularity $\hat{\kappa}(t)$. (d) Estimated input parameters $\hat{\mu}(t)$ and $\hat{\sigma}(t)$. Red (light gray) and blue (gray) lines represent the estimation results obtained using the direct method and the transformation method, respectively. (e) Orbits in the space of the mean and fluctuation of the input current.

against the period of input modulation $T$ (Fig. 5). Rapid modulation was undetectable, because a small number of output spikes cannot sufficiently represent the temporally fluctuating input parameters. For those cases in which a neuron was firing at a mean rate of approximately 40 spikes/s, the estimation was effective if the time scale of input modulation was longer than 1 s.

**B. Applying the estimation methods to experimental data**

Next, we used the transformation method to examine publicly available spike data that were recorded in vivo from visual cortical areas V1 and MT, and the LGN of anesthetized monkeys (Macaca fascicularis) [42–47]. Recordings from area V1, area MT, and the LGN were obtained while a drifting sinusoidal grating was presented to the monkeys (duration, 6000 or 3000 ms for V1, typically 1280 ms for MT, and 5138 ms for the LGN). We excluded spike sequences with a mean firing rate that was less than 10 spikes/s, because the data was not sufficient for analysis. Consequently, 44, 43, and 52 neurons were selected for area V1, area MT, and the LGN, respectively; neurons in these groups were represented by approximately 15, 25, and 15 trials, each containing approximately 250, 50, and 150 spikes, respectively.

We accounted for the effects of the absolute refractory period, which was estimated from the spike width as $1.74 \pm 0.41 \text{ms}$ [28,31]; we analyzed spike trains that were converted from the original spike trains by subtracting the 2 ms absolute refractory period. ISIs smaller or equal to 2 ms in the original spike train were categorized as misdetected intervals, the second spike was ignored, and consecutive ISIs were summed as a single ISI.

The estimated firing characteristics and input parameters were summarized for individual neurons (Fig. 6). For spike trains obtained from each experimental trial, we estimated the estimated trajectory of the firing characteristics $\hat{\lambda}(t)$.
FIG. 5. (Color online) The ISE between the intended input parameters and estimated parameters. ISEs averaged over 100 samples are plotted against the period of the input modulation $T$. Dashed, dotted, and bold lines represent the three fluctuating input conditions (i)–(iii), respectively, with parameters $\mu_0 = 0.5 \text{nA}$, $\delta \mu = 0.15 \text{nA}$, $\sigma_0 = 1 \text{nA} \text{ms}^{1/2}$, $\delta \sigma = 0.6 \text{nA} \text{ms}^{1/2}$, and $\Delta = \pi/2$.

and $\hat{k}(t)$ with the state-space method [Fig. 6(a)]. The firing characteristics $[\hat{\lambda}(t), \hat{k}(t)]$ were plotted for every spike $t = t_j$ as a scatter diagram in the $(\lambda, \kappa)$ plane [Fig. 6(b)], and the spike trains recorded from a single neuron were collected in the $(\lambda, \kappa)$ plane [Fig. 6(c)]. To obtain another scatter diagram in a plane of input parameters $(\mu, \sigma)$, the firing characteristics were converted into the likely input trajectory $\hat{\mu}(t)$ and $\hat{\sigma}(t)$ using the nonlinear transformation formula [Fig. 6(d)]. Finally, the input parameters of a single neuron were summarized as an ellipse describing a 75% quantile of a two-dimensional Gaussian distribution function fitted to the data.

The inferred input parameters $(\mu, \sigma)$ for groups of neurons in area V1, area MT, and the LGN were summarized by plotting ellipses representing individual neurons [Fig. 7(a)]. The distributions of input parameters for all neurons in the three brain regions were summarized as three ellipses representing 75% quantiles of Gaussian distributions [Fig. 7(b)]. It is noteworthy that the distributions of input parameters are similar for areas V1 and MT, which differ markedly from the results obtained for the LGN; the input fluctuation was not significantly correlated with the mean input in the visual cortical areas, whereas the input fluctuation was negatively correlated with the mean input in the thalamus [Fig. 7(c)].

VI. DISCUSSION

In the present paper, we developed a method to estimate nonstationary population activities of excitatory and inhibitory presynaptic neurons from a single train of spikes (Fig. 2). The model estimates likely input parameters over time from time-varying firing characteristics using a nonlinear transformation (Fig. 3). Instantaneous firing characteristics comprising the firing rate and non-Poisson irregularity of a spike train were
estimated using a state-space model [25]. Our purpose is not simply characterizing an output signal, but is inferring the input parameters that are most likely to have evoked the spike train. For this purpose we constructed a nonlinear transformation formula by inverting the forward neuronal transformation of input signals to output spike times. The forward transformation was modeled as a set of radial basis functions and polynomial functions with coefficients determined using the first-passage time of the OUP, which is a mathematical simplification of neuronal spiking conditions based on the LIF model.

Using synthetic data generated in a simulation, we confirmed that the transformation method matched the accuracy of the state-space method for directly estimating input parameters (Fig. 4). The direct method, however, requires extensive computation analysis, including several hours to examine a spike train obtained during a period of few seconds. By contrast, the transformation method is computationally efficient when estimating the firing characteristics. Thus, we can use the transformation method to analyze experimental data containing as many as 10,000 or 100,000 spikes. We also estimated the goodness of the estimation with the transformation method. The results showed that input variations occurring on the order of 1 s can be tracked (Fig. 5).

We used the method to examine open-access spike train data recorded from visual pathways, revealing similarities between areas V1 and MT, whereas the LGN differed from these cortical locations (Fig. 7). Input fluctuation was not significantly correlated with the mean input in areas V1 and MT, whereas the input fluctuation was negatively correlated with the mean input in the LGN (Fig. 7). If inhibitory activity follows excitatory activity, the input fluctuation could be independent of the mean input, but if inhibitory activity is small or does not follow excitatory activity, the input fluctuation could be negatively correlated with the mean input. Thus, the results suggest the existence of balanced input in areas V1 and MT, and the absence in the LGN. Especially, the covariation between excitatory and inhibitory activities implies that the visual cortex processes information in a recurrent manner [5,48,49]. Although the relative contributions of excitatory and inhibitory activities have been analyzed in the
cerebral cortex by carefully observing temporal modulation of synaptic conductances [5,50,51], our method does not require anesthesia or significant constraints on the animal when estimating inputs. Instead, inputs are estimated solely from spike times obtained via extracellular recordings.

Estimating input parameters from a single neuronal spike train can be helpful even when analyzing multiple spike trains recorded from hundreds of neurons. Because recordable neurons constitute a small portion of local cortical circuitry, which often comprises millions of neurons, most input signals originate from unrecorded neurons (Fig. 1). It is important to capture input conditions of individual neurons when discussing local connections between recorded neurons. Comparing the temporal profiles of estimated input parameters may also allow a better understanding of which neurons are receiving common inputs (Fig. 1).

When estimating inputs from output spike parameters, it is essential to understand the neuronal forward transformation from input signals to output spikes. We adopted the standard LIF model with the input current fluctuating rapidly based on incoming EPSPs and IPSPs. The spiking model was reduced to the OUP by assuming that the amplitudes of the EPSPs and IPSPs are small and that they occur randomly in time. The advantage of this approach was that the OUP contains only two independent parameters, expediting parameter exploration based on the firing characteristics.

Nevertheless, the OUP should be replaced by a more realistic model [52–55] to improve the estimation. One advantage of the transforming method is that any spiking neuron model can replace the LIF model. Because the neuronal transformation is also dependent on individual neurons, it may be possible to adapt the model to individual neurons. The neuron model can be selected according to the accuracy in reproducing spike timing given fluctuating input [54,55]. It is desirable that the model in combination with the estimation method is ultimately validated with experimental data of neurons in vitro applied with controlled input current representing fluctuating input parameters.

A number of methods have been developed to estimate constant input parameters from a spike train [14–17]. Among these approaches, the method proposed by Inoue et al. [14] resembles the basic estimation described in the present study, except that the authors determined input parameters via the mean rate and coefficient of variation (Cv). We have shown in Appendix A that their input estimation corresponds to approximating the ISI distribution with the Gaussian distribution. Because Cv gives unstable estimates of neuronal firing characteristics [56,57] and the Gaussian distribution can provide negative ISIs, we believe that using the gamma distribution family is more appropriate for this analysis. In addition to statistical methods, others have attempted to infer input parameters using the self-consistent analysis of networks of model spiking neurons [49,58]. In all of these methodologies and the present method, estimation of input parameters is based on measuring neuronal firing characteristics, such as the rate and non-Poisson irregularity.

Previous papers showed that the distributions of the firing rate and the firing irregularity differed between functional areas in the brain [25,56,59]. The fact implies that there be specific input conditions to generate output spikes with the firing characteristics specific to each area, although the details about inputs remain to be known. In this paper, we derived a method to estimate input conditions from the resulted neuronal firing characteristics, which provides a way of revealing the input conditions specific to functional areas. Especially in this study, we showed that the difference in the distributions of the neuronal firing characteristics among V1, MT and LGN, which has been reported in Ref. [25], could result from the absence or presence of balanced input.

A key advance in the present study is that our method allows us to track temporally varying input parameters by capturing firing characteristics over time using a state-space method. We characterized the neural activity in terms of the firing rate and irregularity by adjusting the scale and shape factors to fit the gamma distribution function to the ISI sequence [25]. Even if we use a parametric family of functions, there may be other appropriate choices, such as an inverse Gaussian distribution function or log-normal distribution function. In addition, to make the state-space method function, we incorporated a random walk prior to the firing rate and irregularity to ensure that the firing characteristics did not fluctuate rapidly. Hyperparameters controlling the degree of fluctuation were determined using the empirical Bayes method such that the marginal likelihood was maximized. Of note, however, the state-space model is not perfect, and it may not work properly in some cases—for example, when the firing characteristics change too rapidly. The state-space method should therefore be improved to work more robustly.

Our method could be further refined to better represent the experimental data by including a model of an individual neuronal forward transformation. Nevertheless, the present study is an important first step for estimating inputs solely from an output spike train. Because our method does not require any specific information other than the spike times, we can reexamine spike data by reopening large neurophysiology archives. For multiunit data, our method may provide additional information by revealing dynamic activities of unrecorded presynaptic neuronal populations for each neuron, from which we can estimate the independence of individual neuronal processing. This may help to elucidate the dynamics of mesoscopic neuronal circuits.

ACKNOWLEDGMENTS

We thank Keisuke Toyama for valuable comments that helped us to improve the manuscript. This study was supported in part by Grants-in-Aid for Scientific Research to SS from the MEXT Japan (20300083, 20020012) and the Global COE Program “The Next Generation of Physics, Spun from Universality and Emergence.

APPENDIX A: MAXIMIZATION OF THE MARGINAL LIKELIHOOD FUNCTION

We implement a method of searching for a set of hyperparameters \( \gamma \equiv (\gamma^1, \gamma^2) \) that maximizes the marginal likelihood function, using expectation maximization (EM) method [38] in Eq. (8). In the EM method, hyperparameters are determined by iteratively maximizing the expected value of the log likelihood,
the $Q$ function,
\[
Q(\gamma|\gamma_0) = E[\log P_{\gamma}(\{t_j\},\{\Lambda_j\}|\{t_j\},\gamma_0)] \\
= \sum_{j=1}^{n} E[\log P_{\gamma}(\Lambda_j|\Lambda_{j-1})|\{t_j\},\gamma_0] \\
+ \sum_{j=1}^{n} E[\log P(s_j|\Lambda_j)|\{t_j\},\gamma_0],
\]
where $\gamma_0 \equiv (\gamma_0^1, \gamma_0^2)$ and $\gamma_j \equiv (\gamma_j^1, \gamma_j^2)$ are the set of hyperparameters of the $\rho$th iteration, and $E[\{\gamma_j\},\gamma_0]$ represents the expectation with respect to the conditional distribution of $\Lambda$ given $\{t_j\} = \{t_j\}_{j=0}^{\infty}$ under the $\rho$th estimate of hyperparameters. From Eq. (7), the ($p + 1$)st estimate of $\gamma$ is determined by the conditions for $dQ/d\gamma = 0$, leading to a set of the following equations:
\[
\gamma_{k+1} = \frac{1}{n-1} \sum_{j=1}^{n} \frac{1}{s_j} \left( (\Lambda_{j+1}^k - \Lambda_j^k)^2 | \{t_j\}, \gamma_0 \right),
\]
for $k = 1$ and 2.

The expected value appearing in the right-hand side of Eq. (A2) can be evaluated by Kalman filtering and smoothing Gaussian distributed under the assumption that the input parameters $\Lambda$ are obtained using the following algorithm:

(i) Prediction algorithm:
\[
\Lambda_{i-1} = \Lambda_{i-1}, \quad \Sigma_{i-1} = \Sigma_{i-1} + G_{i-1},
\]
where $G_{i}$ is the variance-covariance matrix for the input parameters,
\[
G_{i} = \begin{pmatrix}
\gamma^1 s_i & 0 \\
0 & \gamma^2 s_i
\end{pmatrix}.
\]

(ii) Filtering algorithm: The algorithm recursively computes the filtered distribution [60],
\[
P(\Lambda_i|\{t_j\}_{j=0}^{\infty}) = \frac{P(\Lambda_i|\{t_j\}_{j=0}^{\infty}) P(s_i|\Lambda_i,\{t_j\}_{j=0}^{\infty})}{P(s_i|\{t_j\}_{j=0}^{\infty})} \propto P(\Lambda_i|\{t_j\}_{j=0}^{\infty}) P(s_i|\Lambda_i).
\]

Assuming the filtered distribution to be Gaussian, we can obtain its mean and variance as the mode and negative inverse of the second derivative of its logarithm, respectively [39,61],
\[
\begin{align*}
\frac{d}{d\Lambda_{i}} \log P(\Lambda_i|\{t_j\}_{j=0}^{\infty})|_{\Lambda_i = \Lambda_{i0}} &= 0, \\
\Sigma_{iij}^{-1} &= -\frac{d^2}{d\Lambda_{i}^2} \log P(\Lambda_i|\{t_j\}_{j=0}^{\infty})|_{\Lambda_i = \Lambda_{i0}}.
\end{align*}
\]

(iii) Smoothing algorithm [39,62]:
\[
\Lambda_{i|n} = \Lambda_{ij} + A_i(\Lambda_{i+1|n} - \Lambda_{i+1|j}), \\
\Sigma_{i|n} = \Sigma_{ij} + A_i(\Sigma_{i+1|n} - \Sigma_{i+1|j})A_i^T,
\]
where
\[
A_i = \Sigma_{ij}^{-1} \Sigma_{i+1|j}.
\]

(iv) Covariance algorithm [63]:
\[
\Sigma_{i+1,i|n} = \Lambda_i \Sigma_{i+1|i|n}.
\]

From the equations, we evaluate the conditional variances in Eq. (A2):
\[
E[|\Lambda_{j+1}^k - \Lambda_j^k|^2 | \{t_j\}, \gamma_0] = \Sigma_{j+1,j|n} - 2\Sigma_{j+1,j|n} + (\Lambda_{j+1|n} - \Lambda_{j|n})^2,
\]
for $k = 1$ and 2, where $\Sigma_{(a,b)}$ is the $(a,b)$ component of the matrix $\Sigma$.

APPENDIX B: FITTING A PARAMETRIC DISTRIBUTION FUNCTION TO AN ISI DISTRIBUTION

Here, we derive a general method of fitting a two-parameter exponential family $f(s|\theta^1, \theta^2)$ to an ISI distribution of a spiking neuron model $P(s|\Lambda)$, based on a principle of minimizing the KL divergence, and derive Eq. (12) for a particular gamma distribution family.

A two-parameter exponential family is expressed in a form
\[
f(s|\theta^1, \theta^2) = \exp[\eta^1(\theta^1, \theta^2)A^1(s) + \eta^2(\theta^1, \theta^2)A^2(s) - B(\theta^1, \theta^2) + C(s)],
\]
where $\eta^1$ and $\eta^2$ are called natural parameters, and $A^1(s)$ and $A^2(s)$ are sufficient statistics. The natural parameters $(\eta^1, \eta^2)$ can be connected with input parameters $\Lambda$ by minimizing the KL divergence from $P(s|\Lambda)$ to $f(s|\theta^1, \theta^2)$. Taking extremum in two parameter axes $\theta^1$ and $\theta^2$ is equivalent to taking extremum in two axes of natural parameters $\eta^1$ and $\eta^2$:
\[
\frac{d}{d\eta^k} \int P(s|\Lambda) \log \frac{P(s|\Lambda)}{f(s|\theta^1, \theta^2)} ds = 0,
\]
for $k = 1$ and 2.

By plugging Eq. (B1) into Eq. (B2) we obtain
\[
\int_0^\infty ds A^k(s) P(s|\Lambda) = \int_0^\infty ds A^k(s) f(s|\theta^1, \theta^2),
\]
for $k = 1$ and 2.

This means that sufficient statistics should be congruent between two distribution functions. When considering the
gamma distribution function,
\[
g(s|\lambda, \kappa) = \exp[-\lambda \kappa s + (\kappa - 1) \log s + \kappa \log(\lambda \kappa) - \log \Gamma(\kappa)].
\]  
(B4)
for which two natural statistics are \(s\) and \(\log(s)\), relation (B3) is written down as Eq. (12).

When replacing the normal distribution with the gamma distribution,
\[
n(s|\lambda, \nu) = \exp[(\lambda / \nu) s - (\lambda^2 / 2\nu) s^2 + \log(\lambda / \nu) - 1 / 2\nu - (1 / 2) \log(2\pi)],
\]  
(B5)
we have another set of natural statistics as \(s\) and \(s^2\). In this case, the extremum parameters are given by
\[
\int_0^\infty ds s^2 P(s|\Lambda) = 1 / \nu,
\]  
\[
\int_0^\infty ds s^2 P(s|\Lambda) / \left(\int_0^\infty ds P(s|\Lambda)\right)^2 - 1 = v^2.
\]  
Because \(\lambda\) and \(\nu\) represent the mean firing rate and the coefficient of variation, respectively, fitting to the Gaussian distribution is equivalent to the method proposed by Inoue et al. [14].

**APPENDIX C: TRANSFORMATION FROM FIRING CHARACTERISTICS TO INPUT PARAMETERS**

We constructed a formula representing a backward transformation from the output firing characteristics comprising the firing rate and irregularity, \((\lambda, \kappa)\), to the input parameters comprising the mean and fluctuation of the input current, \((\mu, \sigma)\).

Because any OUP can be linearly transformed into another, all the OUP models are comprehended by investigating a single standard OUP:
\[
\frac{dU(x)}{dx} = -U(x) + \mu + \sigma \xi(x),
\]  
if \(U(x) > 1\), then \(U(x) \to 0\).
(C1)

Thus we obtain a forward transformation for this standard OUP from input mean and fluctuation to output firing rate and irregularity, Eq. (11), and then invert the relation into a backward transformation, Eq. (10).

In considering the LIF model [Eq. (1)] with an assumption of uncorrelated inputs [Eq. (2)], we obtain another OUP:
\[
\tau_m \frac{dV(t)}{dt} = V_l - V(t) + R(\mu + \sigma \xi(t)),
\]  
if \(V(t) > V_{TH}\), then \(V(t) \to V_R\).
(C2)

Once the standard backward transformation Eq. (10) for the standard OUP is established, we may translate it for this particular OUP as
\[
\mu = M(\lambda \tau_m, \kappa) \frac{(V_{TH} - V_R)}{R} + \frac{(V_R - V_l)}{R},
\]  
\[
\sigma = S(\lambda \tau_m, \kappa) \sqrt{\tau_m} \frac{(V_{TH} - V_R)}{R}.
\]  
(C3)

**1. Forward transformation from input to output**

A forward transformation from input parameters to output firing characteristics, Eq. (11), can be constructed by solving Eq. (12) from the distribution of ISIs, \(P(s|\mu, \sigma)\). For the OUP, the firing rate \(\lambda\) can be calculated using a formula for the first moment of the first-passage time distribution [18, 64, 65]:
\[
\lambda^{-1} = \phi(\sqrt{2 / \sigma^2} (1 - \mu)) - \phi(-\sqrt{2 / \sigma^2} \mu),
\]  
(C4)
where
\[
\phi(z) = \left\{ \begin{array}{ll}
100 \left(\frac{\sqrt{2 \pi}}{k!} \Gamma \left(\frac{k}{2}\right) \right) & : z > -5.7, \\
-\left(K_B + \log |z| + \sum_{k=1}^{10} b_k \frac{\exp[(-2)^k]}{k^{2k+1}} \right) & : z \leq -5.7,
\end{array} \right.
\]  
(C5)

The firing irregularity \(\kappa\) cannot be obtained analytically. Thus we should carry out a numerical simulation of the standard OUP, Eq. (C1), to obtain a large number of ISIs \(\{s_1, s_2, \ldots, s_N\}\), and solve Eq. (12) numerically:
\[
\psi(\kappa) - \log(\kappa) = \sum_{i=1}^{N} \log(s_i) - \log \left(\sum_{i=1}^{N} \frac{s_i}{N}\right).
\]  
(C6)

Because \(\psi(\kappa) - \log(\kappa)\) is a monotonically increasing function, we can obtain the root of Eq. (C6) with the bisection method [66]. For each set of input parameters \((\mu, \sigma)\), we generated \(N = 10^5\) ISIs using a fast and accurate simulation algorithm [67].

**2. Backward transformation from output to input**

The backward transformation Eq. (10) can be obtained by inverting the forward transformation Eq. (11). In practice, we first performed a massive numerical simulation over a variety of input parameters \((\mu, \sigma)\) to obtain the firing characteristics \((\lambda, \kappa)\). To approximate the huge number of sample points \(\{\lambda_i, \kappa_i, \mu_i, \sigma_i\}\) generated by the numerical simulation we construct simple and handy polyharmonic splines [68] by selecting \(N_s = 100\) sample points:
\[
M(\lambda, \kappa) = \sum_{i=1}^{N_s} \omega_i^\mu \psi(r_i) + \nu_i^\mu + \nu_i^\kappa \log \lambda + \nu_i^\kappa \log \kappa,
\]  
(C7)
\[
\log[S(\lambda, \kappa)] = \sum_{i=1}^{N_s} \omega_i^\kappa \psi(r_i) + \nu_i^\kappa + \nu_i^\kappa \log \lambda + \nu_i^\kappa \log \kappa,
\]  
where
\[\omega_i^\mu, \nu_i^\mu, \nu_i^\kappa, \omega_i^\kappa, \nu_i^\kappa\] are the weights and knot points, respectively, for the construction of the splines.

\[\psi(r)\] is the polyharmonic spline [68] constructed over the interval \([0, 1]\):
\[
\psi(r) = \sum_{k=0}^{K} \sum_{j=0}^{K} \omega_{kj} r^k (1 - r)^j.
\]  
The coefficients \(\omega_{kj}\) are determined by solving a linear equation system with \(2K + 1\) equations for \(K + 1\) unknowns as follows:
\[
\sum_{k=0}^{K} \sum_{j=0}^{K} \omega_{kj} r^k (1 - r)^j = \frac{1}{2k + 1},
\]  
for \(r = 0.5\).

The formula for the characteristic function of a polyharmonic spline (denoted as \(\psi\)) is:
\[
\psi(r) = \sum_{k=0}^{K} \sum_{j=0}^{K} \omega_{kj} r^k (1 - r)^j.
\]  
We can use the following relation to determine the characteristic function of a polyharmonic spline containing \(2K + 1\) coefficients as follows:
\[
\phi^{(2K)}(z) = \frac{1}{2K + 1},
\]  
where \(\phi^{(2K)}(z)\) is the characteristic function of a polyharmonic spline containing \(2K + 1\) coefficients.
where \( r_i = \sqrt{\log \lambda - \log \lambda_i} + (\log \kappa - \log \kappa_i)^2 \), \( \varphi(r) = r^3 \) is a basic function, \( \omega^\mu_i \) and \( \omega^\rho_i \) are the \( N_i \) weights of the basic functions, and \( v_i^\rho \) and \( v_i^\mu \) are the three weights of the polynomial for input parameters \( \mu \) and \( \rho \) respectively. The weights \( \omega^\rho_i = (\omega^\mu_1, \omega^\mu_2, \ldots, \omega^\mu_{N_i})^T \) and \( v_i^\rho = (v_i^\mu, v_i^\mu, v_i^\mu)^T \) are determined by solving a symmetric, linear system of equations as

\[
\begin{pmatrix}
A & V^T \\
V & 0
\end{pmatrix}
\begin{pmatrix}
\omega^\rho_i \\
v_i^\rho
\end{pmatrix} = \begin{pmatrix}
y \\
0
\end{pmatrix},
\]

where

\[
y = (\mu_1, \mu_2, \ldots, \mu_{N_i})^T,
\]

\[
V = \begin{pmatrix}
\log \lambda_1 & \log \lambda_2 & \cdots & \log \lambda_{N_i} \\
\log \kappa_1 & \log \kappa_2 & \cdots & \log \kappa_{N_i}
\end{pmatrix},
\]

\[
A_{ij} = \varphi((\log \lambda_i - \log \lambda_j)^2 + (\log \kappa_i - \log \kappa_j)^2).
\]

The weights \( \omega^\rho_i = (\omega^\rho_1, \omega^\rho_2, \ldots, \omega^\rho_{N_i})^T \) and \( v_i^\rho = (v_i^\rho, v_i^\rho, v_i^\rho)^T \) are also computed from Eq. (C8) by setting \( y = (\log \sigma_1, \log \sigma_2, \ldots, \log \sigma_{N_i})^T \). The data points and the weights are in Table I.

### Table I. Polyharmonic spline parameters that provide a formula for the reverse transformation representing output to input signals.

<table>
<thead>
<tr>
<th>( \log \lambda_i )</th>
<th>( \log \kappa_i )</th>
<th>( \omega^\mu_i )</th>
<th>( \omega^\rho_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.2040</td>
<td>-0.7840</td>
<td>-0.3640</td>
<td>0.0559</td>
</tr>
<tr>
<td>-1.5020</td>
<td>2.9633</td>
<td>-0.4089</td>
<td>-3.0000</td>
</tr>
<tr>
<td>0.2189</td>
<td>-0.1863</td>
<td>-1.1754</td>
<td>1.0946</td>
</tr>
<tr>
<td>-3.0173</td>
<td>-3.0173</td>
<td>-2.5280</td>
<td>-2.5280</td>
</tr>
<tr>
<td>0.0559</td>
<td>0.4759</td>
<td>-1.2040</td>
<td>-0.7840</td>
</tr>
<tr>
<td>-1.6041</td>
<td>0.2324</td>
<td>-0.5383</td>
<td>3.0197</td>
</tr>
<tr>
<td>-4.3503</td>
<td>1.6684</td>
<td>0.2327</td>
<td>-0.5613</td>
</tr>
<tr>
<td>-2.5280</td>
<td>-2.0387</td>
<td>-2.0387</td>
<td>-2.0387</td>
</tr>
<tr>
<td>-1.0601</td>
<td>-0.0622</td>
<td>-0.1336</td>
<td>0.1936</td>
</tr>
<tr>
<td>-0.3460</td>
<td>0.5594</td>
<td>0.7598</td>
<td>0.9158</td>
</tr>
<tr>
<td>-0.4746</td>
<td>0.2099</td>
<td>-0.1609</td>
<td>-0.0577</td>
</tr>
<tr>
<td>0.1746</td>
<td>0.5201</td>
<td>-0.3004</td>
<td>0.2558</td>
</tr>
<tr>
<td>0.9155</td>
<td>-0.0560</td>
<td>0.0209</td>
<td>0.1333</td>
</tr>
<tr>
<td>0.6860</td>
<td>-0.2214</td>
<td>0.2697</td>
<td>0.1706</td>
</tr>
<tr>
<td>0.2315</td>
<td>-0.3359</td>
<td>0.3718</td>
<td>-0.5005</td>
</tr>
<tr>
<td>0.2255</td>
<td>-0.0816</td>
<td>-0.0816</td>
<td>-0.0816</td>
</tr>
<tr>
<td>0.4777</td>
<td>0.4777</td>
<td>0.4777</td>
<td>0.4777</td>
</tr>
<tr>
<td>0.1253</td>
<td>-0.5092</td>
<td>0.5092</td>
<td>0.5092</td>
</tr>
<tr>
<td>0.3266</td>
<td>-0.0391</td>
<td>-0.2014</td>
<td>0.3718</td>
</tr>
<tr>
<td>0.2255</td>
<td>-0.0816</td>
<td>-0.0816</td>
<td>-0.0816</td>
</tr>
<tr>
<td>0.0078</td>
<td>0.0078</td>
<td>0.0078</td>
<td>0.0078</td>
</tr>
<tr>
<td>0.0078</td>
<td>0.0078</td>
<td>0.0078</td>
<td>0.0078</td>
</tr>
<tr>
<td>0.0078</td>
<td>0.0078</td>
<td>0.0078</td>
<td>0.0078</td>
</tr>
<tr>
<td>0.0078</td>
<td>0.0078</td>
<td>0.0078</td>
<td>0.0078</td>
</tr>
<tr>
<td>0.0078</td>
<td>0.0078</td>
<td>0.0078</td>
<td>0.0078</td>
</tr>
<tr>
<td>0.0078</td>
<td>0.0078</td>
<td>0.0078</td>
<td>0.0078</td>
</tr>
</tbody>
</table>