Statistical structure of the trial-to-trial timing variability in synfire chains

Dina Obeid,[1,*] Jacob A. Zavatone-Veth,[2,†] and Cengiz Pehlevan[1,3,‡]

[1] *John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA*
[2] *Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA*
[3] *Center for Brain Science, Harvard University, Cambridge, Massachusetts 02138, USA*

Timing and its variability are crucial for behavior. Consequently, neural circuits that take part in the control of timing and in the measurement of temporal intervals have been the subject of much research. Here, we provide an analytical and computational account of the temporal variability in what is perhaps the most basic model of a timing circuit, the synfire chain. First, we study the statistical structure of trial-to-trial timing variability in a reduced but analytically tractable model: a chain of single integrate-and-fire neurons. We show that this circuit's variability is well-described by a generative model consisting of local, global, and jitter components. We relate each of these components to distinct neural mechanisms in the model. Next, we establish in simulations that these results carry over to a noisy homogeneous synfire chain. Finally, motivated by the fact that a synfire chain is thought to underlie the circuit that takes part in the control and timing of zebra finch song, we present simulations of a biologically realistic synfire chain model of the zebra finch timekeeping circuit. We find the structure of trial-to-trial timing variability to be consistent with our previous findings, and to agree with experimental observations of the song's temporal variability. Our study therefore provides a possible neuronal account of behavioral variability in zebra finches.

I. INTRODUCTION

Timing is critical for many behaviors, such as speech production, playing a musical instrument, or dancing. However, even the most stereotyped animal behaviors are significantly variable from one iteration to the next. This so-called trial-to-trial variability is ubiquitous and may serve important functions in motor learning and adaptation [1–3]. Its sources have therefore been of great interest to neuroscientists [2, 4].

As behavior is controlled by the nervous system, it is natural to look for the source of some of this variability in the variable activity of neural circuits involved in the production of behavior [2, 4–7]. Indeed, the neural mechanisms underlying behavioral timing have been extensively studied experimentally [2, 8–12], establishing links between temporal variations of behavior and that of neural activity [13–20]. This neural variability could result from multiple sources such as stochastic events at the level of ion channels [21], synapses [22], and neurons [23]; chaotic activity of neural networks [24, 25]; and sensory inference errors [5]. Therefore, understanding the mechanisms and structure of timing variability in neural circuits is necessary for understanding variability in behavior.

In this paper, we study temporal variability in one of the most basic neural network models of timing, the synfire chain [26–28]. The synfire chain is a feedforward network composed of pools of neurons that produce traveling waves of synchronized spiking activity, as illustrated in Figure 1. The synchronization of spikes within a pool, and the sequential propagation of activity across pools (Figure 1b) allow the synfire chain to serve multiple timing functions in a very natural way. First, it can be used to keep time by simply counting the pool which the spiking activity has reached. Second, it can be used to produce precisely timed intervals defined by the time elapsed between when activity arrives at a given pool and when it arrives at a subsequent pool (Figure 1b). The synfire chain can sustain activity indefinitely given sufficiently many pools [27, 29], or by arranging it in a circular topology such that the final pool connects back to the first one [29]. While the synfire chain is theoretically well-studied [27, 29–31], and experiments support its existence in biological systems [32], a theoretical account of its temporal variability is still lacking.

We are interested in the trial-to-trial variability in the timing of neural activity of a synfire chain. Such variability arises from the millisecond-scale tempo differences across multiple propagations of the spiking activity in the chain. We will focus on trial-to-trial variability caused by the inherent noise and fatigue in the neural system.

While it might seem small, millisecond-scale neural variability has been experimentally shown to correlate with behavioral variability at the same timescale in songbirds [16, 18, 20]. This finding is especially relevant since experiments support the existence of a synfire chain architecture in the songbird premotor cortex [32] for millisecond-scale precise time-keeping of the birdsong, with total song durations of few hundreds of milliseconds [13, 33]. Therefore, our findings may have direct implications for behavioral timing. Indeed, we will show that the statistical structure of temporal variability in a synfire chain can possibly explain some of the salient features of temporal variability in birdsong [33–35].

We address these questions first in a simplified and analytically solvable model of trial-to-trial variability in a

* dinaobeid@seas.harvard.edu
† jzavatoneveth@g.harvard.edu
‡ cpehlevan@seas.harvard.edu

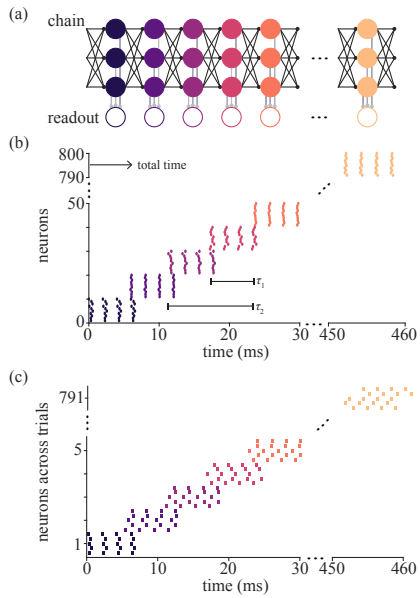

FIG. 1. Timekeeping and trial-to-trial variability in synfire chains. (a) The synfire chain is a feedforward network of pools of neurons. In the figure, circles represent neurons and arrows represent synapses. A simple scheme to measure time intervals composed of K successive pools, is to mark the beginning and end of the interval by the first spike of readout neurons corresponding to first and last pools. (b) Spike trains produced by the synfire chain show a synchronized activity that progresses pool-by-pool. Each point corresponds to the time of a spike produced by a neuron in the chain. Color denotes membership in a layer, each of which consists of a pool of 10 neurons. Various time intervals are shown. (c) Trial-to-trial variability in the synfire chain. Spike times of six neurons from different pools are plotted in different colors. Five different trials for each neuron are shown.

chain of individual neurons (Section II). We derive analytical expressions that describe the magnitude and statistical structure of temporal variability in terms of inherent neural noise and fatigue and verify our results with simulations. We use a generative model introduced by Glaze and Troyer [35] to decompose the variability covariance matrix into three components: independent, global and jitter. Next, we address temporal variability in a noisy homogeneous synfire chain, which includes multiple identical neurons per pool, by numerical simulations (Section III). We show that our results from the analytically tractable model qualitatively carry over to this more complex model. We study the dependence of the various components of variability on the number of neurons per pool of the chain. Further, we relate the distinct neural mechanisms in the model to the different components of variability obtained from the generative model. Finally, we provide an application of our results to birdsong. In zebra finches, experimental studies support the existence of a synfire chain structure in the premotor nucleus HVC [32], which takes part in the production and timing of the birdsong. We simulate a biologically realistic synfire chain model [32] and show that the statistical structure and magnitude of its variability is consistent with that observed in the analytically tractable and homogeneous synfire chain model, and the zebra finch song (Section IV).

II. TRIAL-TO-TRIAL TIMING VARIABILITY IN A CHAIN OF SINGLE NEURONS

In this section, we describe the statistical structure of trial-to-trial timing variability in an analytically tractable model: a chain of N single leaky integrate-and-fire (IF) neurons. In this simple model, each neuron is driven by excitatory synaptic input from the previous neuron in the chain at time t_{ps}, which we model by $I_s\Theta(t-t_{ps})$, where $\Theta(t)$ is the Heaviside step function. Though it would be more biologically realistic to use a input kernel of finite duration, we make this analytically convenient choice as we are only concerned with first-spike times. We model the drive to each neuron from outside the chain by the sum of a constant current I_0, and the noise due to synaptic transmission and other cellular processes by a zero-mean Gaussian process $\sqrt{\tau}\eta(t)$ with autocorrelation $\sigma^2\tau\delta(t-t')$, where τ is the membrane time constant and σ controls the standard deviation of the noise [36, 37].

The sub-threshold dynamics of the membrane potential V of a given neuron in the chain is then governed by the Langevin equation [36, 37]

$$\tau\dot{V}(t) = -V(t) + I_0 + I_s\Theta(t-t_{ps}) + \sqrt{\tau}\eta(t). \qquad (1)$$

When the neuron's membrane potential reaches a firing threshold V_{th}, the neuron produces a spike and resets its membrane potential to V_r.

A. Local variability in a chain of single neurons

We want to study the variability in the first-spike times of successive neurons in the chain. This problem differs from the standard treatment of noisy and leaky IF neurons [36, 37] in that we are interested in trial-to-trial variability of intervals between different neurons' spikes rather than long-time statistics of the intervals between spikes generated by a single neuron. However, we can map this problem to previous results in the literature [36, 37] by dividing it into two threshold-crossing problems. First, we can determine the probability distribution of a given neuron's membrane potential at time t_{ps}, using the fact that it receives no synaptic input before the previous neuron's first spike. Then, given that its membrane potential at time t_{ps} is V_0 with probability $P(V_0)$, we can think of the trial-to-trial variability in that neuron's time to first-spike after t_{ps} as the variability in the inter-spike intervals of a single leaky IF neuron with $V_r = V_0$.

To proceed we assume that t_{ps} is long enough such that $V(t_{ps})$ has equilibrated. This assumption simplifies our calculations and will be validated when our final results are compared to simulations. Then, the solution to the first problem is given by the stationary distribution of the membrane potential, which was found in [36] to be

$$P(V) = \Theta(V_{th} - V)\frac{2\nu\tau}{\sigma}\exp\left(-\frac{(V-I_0)^2}{\sigma^2}\right) \times \int_{(V-I_0)/\sigma}^{(V_{th}-I_0)/\sigma} du\,\Theta\left(u - \frac{V_r - I_0}{\sigma}\right)e^{u^2}, \quad (2)$$

where the firing rate ν is given by

$$\nu = \frac{1}{\tau\sqrt{\pi}}\left[\int_{(V_r-I_0)/\sigma}^{(V_{th}-I_0)/\sigma} du\, e^{u^2}(1+\mathrm{erf}(u))\right]^{-1}. \quad (3)$$

We will be interested in the limit of very low firing rates. This limit is relevant to propagation of spiking activity in synfire chains because a neuron in the chain spikes only when the spiking activity reaches the pool to which the neuron belongs and very rarely otherwise. Because noise is the main driver of firing in the absence of external input, the very low firing rate limit is given by assuming $V_{th} - I_0 \gg \sigma$ [37, 38], which leads to the standard approximations (see Appendix A 1 and [36]) of the firing rate as

$$\nu \approx \frac{V_{th} - I_0}{\sigma\tau\sqrt{\pi}}\exp\left(-\frac{(V_{th}-I_0)^2}{\sigma^2}\right), \quad (4)$$

and the membrane potential distribution as

$$P(V) \approx \frac{1}{\sqrt{\pi}\sigma}\exp\left(-\frac{(V-I_0)^2}{\sigma^2}\right), \quad (5)$$

which is the stationary limit of an Ornstein-Uhlenbeck process with boundaries set at infinity.

We can calculate the mean and variance of the first-spike-interval, T_{fs}, defined as the time elapsed from t_{ps} to the arrival of the first spike, using the mapping between our problem and that of determining the statistics of the inter-spike intervals of a single leaky IF neuron. Conditioned on $V(t_{ps}) = V_0$, these statistics are given by the standard expressions [36, 37]

$$\langle T_{fs}\rangle_{V_0} = \tau\sqrt{\pi}\int_{(V_0-I_0-I_s)/\sigma}^{(V_{th}-I_0-I_s)/\sigma} du\, e^{u^2}(1+\mathrm{erf}(u)) \quad (6)$$

and

$$\langle \delta T_{fs}^2\rangle_{V_0} = 2\pi\tau^2\int_{(V_0-I_0-I_s)/\sigma}^{(V_{th}-I_0-I_s)/\sigma} dx\, e^{x^2} \times \int_{-\infty}^{x} du\, e^{u^2}(1+\mathrm{erf}(u))^2. \quad (7)$$

We can then combine these results to compute the mean and variance of the first-spike-interval across trials.

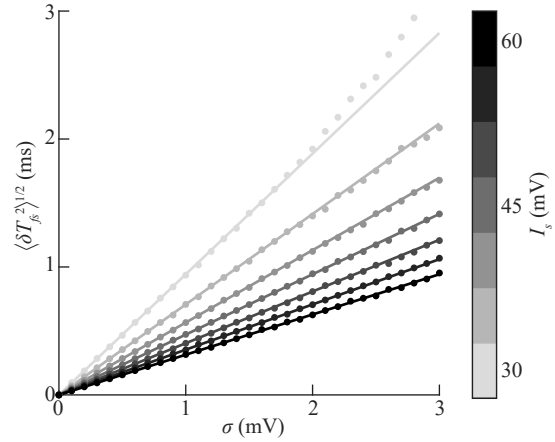

FIG. 2. Scaling of local trial-to-trial timing variability in a chain of single neurons. The results of numerical experiment (Section II D) are shown as dots, and the solid lines show the asymptotic approximations obtained in Section II A. The ordinate shows the standard deviation $\langle \delta T_{fs}^2\rangle^{1/2}$ of the first-spike-interval, and the abscissa shows the standard deviation σ of the noise. Increasing values of I_s are indicated by darker shades of gray.

If we approximate the distribution of initial membrane potentials by the stationary Ornstein-Uhlenbeck process limit (5) in the low firing rate regime $V_{th} - I_0 \gg \sigma$, we obtain the lowest-order asymptotic expansions

$$\frac{\langle T_{fs}\rangle}{\tau} \sim \log\left(\frac{I_s}{I_0 + I_s - V_{th}}\right) - \frac{\sigma^2}{4(I_0+I_s-V_{th})^2} \quad (8)$$

and

$$\frac{\langle \delta T_{fs}^2\rangle}{\tau^2} \sim \frac{\sigma^2}{2(I_0+I_s-V_{th})^2} \quad (9)$$

in the limit of large synaptic input $I_s + I_0 - V_{th} \gg \sigma$ (a detailed derivation of these expressions is given in Appendix A 2). The scaling of this variability with I_s and σ for fixed I_0 and V_{th} is illustrated in Figure 2 (see Section II D). In Appendix A 3, we also derive asymptotics for $\langle T_{fs}\rangle$ and $\langle \delta T_{fs}^2\rangle$ using the alternative approximation $P(V_0) \approx \delta(V_0 - I_0)$.

B. Correlated variability in a chain of single neurons

Thus far, we have only considered sources of variability that are independent across neurons. However, in biological neural networks, there are many possible mechanisms that could introduce correlated variability, such as correlated external input [4, 7, 37, 38]. Here, we consider a simple model for correlated variability due to neural fatigue, i.e. a loss in a neuron's excitability due to effects like synaptic depression [39].

In our model, the spiking threshold in a given trial is increased by some multiple $m \in \{0, 1, \ldots, m_{\max}\}$ of

a small increment δV_{th} relative to the baseline threshold V_{th}. We assume that, across trials, m is drawn from some distribution with mean $\langle m\rangle$ and variance $\langle \delta m^2\rangle$. Working in the regime in which $m_{\max}\,\delta V_{th} \ll I_0+I_s-V_{th}$, relevant to neurons in a synfire chain which receive a barrage of inputs from the synchronous firing of presynaptic pool of neurons, we can use our previously-obtained asymptotic expansions (8, 9) of the mean and variance of the first-spike-interval conditioned on V_{th} to obtain

$$\frac{\langle T_{fs}\rangle}{\tau} \sim \log\left(\frac{I_s}{I_0+I_s-V_{th}}\right) + \frac{\delta V_{th}}{I_0+I_s-V_{th}}\langle m\rangle, \quad (10)$$

and

$$\frac{\langle \delta T_{fs}^2\rangle}{\tau^2} \sim \frac{\sigma^2}{2(I_0+I_s-V_{th})^2}\left[1+2\frac{\delta V_{th}}{I_0+I_s-V_{th}}\langle m\rangle\right] + \frac{\delta V_{th}^2}{(I_0+I_s-V_{th})^2}\langle \delta m^2\rangle, \quad (11)$$

to lowest order in both $\delta V_{th}/(I_0+I_s-V_{th})$ and $\sigma/(I_0+I_s-V_{th})$ (see Appendix A 2 for details). If we now consider two different neurons a and b, the trial-to-trial covariance of their first-spike-intervals T_{fs}^a and T_{fs}^b will be

$$\frac{\langle \delta T_{fs}^a\,\delta T_{fs}^b\rangle}{\tau^a\tau^b} \sim \frac{\delta V_{th}^a \delta V_{th}^b}{(I_0^a+I_s^a-V_{th}^a)(I_0^b+I_s^b-V_{th}^b)}\langle \delta m^2\rangle, \quad (12)$$

to lowest order in $\delta V_{th}^a/(I_0^a+I_s^a-V_{th}^a)$ and $\delta V_{th}^b/(I_0^b+I_s^b-V_{th}^b)$.

Therefore, we obtain a model in which the trial-to-trial covariance matrix of the first-spike-intervals of neurons in the chain is the sum of a diagonal, local-to-a-neuron component and a rank-one global component. Concretely, if we assume for simplicity that all neurons in the chain are identical, the covariance matrix of the first-spike-intervals of the neurons in the chain is given as $\mathbf{\Sigma}_{fs} = \sigma_I^2\mathbf{I} + \sigma_G^2\mathbf{1}\mathbf{1}^\top$, where $\sigma_I^2 = \tau^2\frac{\sigma^2}{2(I_0+I_s-V_{th})^2}\left[1+2\frac{\delta V_{th}}{I_0+I_s-V_{th}}\langle m\rangle\right]$ is the local component of the first-spike-interval variance, $\sigma_G^2 = \tau^2\frac{\delta V_{th}^2}{(I_0+I_s-V_{th})^2}\langle \delta m^2\rangle$ is the global component of the first-spike-interval variance, $\mathbf{I}$ is the identity matrix, and $\mathbf{1}$ is the ones vector. If the neurons were non-identical, σ_I and σ_G would no longer be scalar constants, and the covariance matrix would have the form $\mathbf{\Sigma}_{fs} = \mathbf{\Sigma}_I + \boldsymbol{\sigma}_G\boldsymbol{\sigma}_G^\top$ for a diagonal matrix $\mathbf{\Sigma}_I$ and a vector $\boldsymbol{\sigma}_G$. This decomposition is illustrated in Figure 3a.

In the above calculation, we assumed that one can perfectly read out the first-spike times from the neurons of the chain. However, it is unlikely that noise-free readout is possible in biological timekeeping systems. In the presence of readout noise, the first-spike interval covariance matrix will have an additional component that increases the variance of individual intervals and introduces negative covariance between adjacent intervals, a phenomenon known as timing jitter [35, 40]. Assuming for simplicity that the readout noise is homogeneous, additive, and independent of other forms of variability and has standard deviation σ_J, the overall covariance matrix is given as

$$\mathbf{\Sigma}_{fs} = \sigma_I^2\mathbf{I} + \sigma_G^2\mathbf{1}\mathbf{1}^\top + \sigma_J^2\mathbf{D}\mathbf{D}^\top, \quad (13)$$

where $\mathbf{D}$ is the $N\times(N-1)$ bidiagonal differencing matrix with ones along the diagonal and negative ones along the subdiagonal: $D_{jk} = \delta_{jk} - \delta_{j(k-1)}$.

C. From mechanism to statistical models of timing variability

In the preceding sections, we have characterized the timing variability inherent in a simple neural model. This calculation showed that the covariance matrix of the intervals between the first spikes of successive neurons in the chain could be decomposed into a diagonal local component, a rank-one global component, and a structured component resulting from imperfect readout of spike times. Strikingly, the same covariance structure is present in statistical models of behavioral timing variability [35, 40]. In particular, it matches a Gaussian generative model for behavioral interval durations proposed for zebra finch song by Glaze and Troyer [35]. For a set of P intervals, this model is parameterized by a vector $\mathbf{w}\in\mathbb{R}^P$ and diagonal latent variable covariance matrices $\mathbf{\Psi}\in\mathbb{R}^{P\times P}$ and $\mathbf{\Omega}\in\mathbb{R}^{(P-1)\times(P-1)}$. Then, the vector of interval durations in the μth trial is generated as

$$\mathbf{t}_\mu = \bar{\mathbf{t}} + \sqrt{\mathbf{\Psi}}\boldsymbol{\xi}_\mu + \mathbf{w}z_\mu + \mathbf{D}\sqrt{\mathbf{\Omega}}\mathbf{u}_\mu, \quad (14)$$

where $\bar{\mathbf{t}}$ is the average duration and $z_\mu \sim \mathcal{N}(0,1)$, $\boldsymbol{\xi}_\mu \sim \mathcal{N}(\mathbf{0},\mathbf{I}_P)$, and $\mathbf{u}_\mu \sim \mathcal{N}(\mathbf{0},\mathbf{I}_{P-1})$ are independent latent factors that are independent and identically distributed across trials. The interval duration covariance matrix in this model is thus

$$\mathbf{\Sigma}_{\text{int}} = \mathbf{\Psi} + \mathbf{w}\mathbf{w}^\top + \mathbf{D}\mathbf{\Omega}\mathbf{D}^\top. \quad (15)$$

For homogeneously variable intervals, we can therefore associate each component of the covariance matrix in this behavioral model to one of the components of the first-spike interval covariance matrix (13) in our neural model.

The connection between the statistical structures of neural and behavioral variability also leads to a prediction for how variability should scale with behavioral interval duration under a simple timekeeping model. We assume that the basic unit of time is measured via noisy readout of the first-spike times T_{fs}^a of neurons with covariance (13), and that longer behavioral intervals T_{iK} are formed by summing the durations of sequences of K first-spike intervals:

$$T_{iK} = \sum_{a=(i-1)K+1}^{iK} T_{fs}^a. \quad (16)$$

Then, assuming for convenience that K evenly divides N, the interval-interval covariances are given by sums of

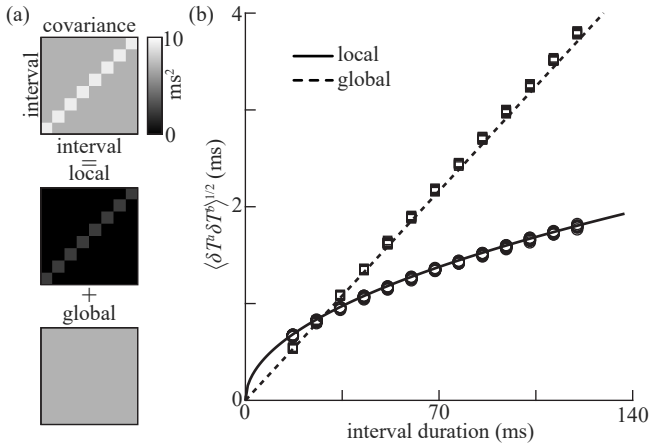

FIG. 3. Correlated trial-to-trial timing variability in a chain of single neurons. (a) Schematic representation of the decomposition of the interval-interval covariance matrix into local and global components as described in Section II B (see expressions for σ_I^2 and σ_G^2). (b) Scaling of the local and global variability with interval duration. The results of numerical experiment (see Section II D) are shown by circles and squares for local and global variability, respectively, while the predictions of asymptotic theory (see Sections II B and II C) are shown by solid and dashed lines. Realizations of the random sampling used to generate intervals are plotted as individual markers.

the corresponding $K \times K$ submatrices of $\mathbf{\Sigma}_{fs}$. This summation yields a $N/K \times N/K$ interval duration covariance matrix

$$\mathbf{\Sigma}_{\text{int}} = K\sigma_I^2 \mathbf{I} + K^2 \sigma_G^2 \mathbf{1}\mathbf{1}^\top + \sigma_J^2 \mathbf{D}\mathbf{D}^\top, \tag{17}$$

where the scaling of the independent and global components is trivial and that of the jitter component follows from the observation that $\mathbf{D}\mathbf{D}^\top$ is a tridiagonal matrix with sub- and super-diagonal elements equal to -1, first and last diagonal elements equal to $+1$, and other diagonal elements equal to $+2$. Therefore, in this simple model for tracking longer intervals of time, the assumption that the first-spike interval covariance matrix has the given structure implies that the interval covariance matrix will have the same structure. Furthermore, the different components of variability have distinct scalings with interval length: the independent component scales linearly, the global component quadratically, and the jitter component is constant. We note that this scaling of local and global components of variability is independent of the details of the single-neuron model, provided that the covariance of the first-spike intervals produced has the form of (13). This variance decomposition and scaling with interval length are illustrated in Figure 3.

D. Numerical simulations

We compare the theoretical asymptotics we obtained in Sections II A, II B, and II C to the results of numerical experiment. To study the scaling of local variability with noise variance and synaptic strength as shown in Figure 2, we perform 10^4 realizations of a single-neuron simulation. In these experiments, we fix $\tau = 20\,\text{ms}$, $I_0 = -70\,\text{mV}$, and $V_{th} = -45\,\text{mV}$ while varying σ and I_s. We integrate the Langevin equation (1) using the Euler-Maruyama method [41] augmented by the reset rule with a timestep of $\Delta t = 10^{-3}\,\text{ms}$. The Euler-Maruyama stochastic integration method is an explicit first-order accurate method in the absence of noise, and accurate to order 1/2 in the presence of noise. We find empirically that increasing or decreasing the timestep by factors of ten does not influence the qualitative results. For all parameter values tested, we observe good agreement between our asymptotic approximation and the experimental results for the mean first-spike-interval. As shown in Figure 2, for the lowest synaptic strengths and largest noise variances, we observe some discrepancy between asymptotic theory and experiment for the first-spike-interval variance. This is unsurprising, since in that regime $I_s + I_0 - V_{th}$ is only around five times greater than the standard deviation of the noise, hence higher-order terms in the expansion are non-negligible (see Appendix A 2).

To study the influence of introducing neural fatigue as described in Section II B, we simulate a chain of 80 identical neurons using the methods described above. In these experiments, we fix $\sigma = 1\,\text{mV}$ and $I_s = 45\,\text{mV}$. Over the 10^4 realizations performed, we draw the parameter m from the discrete uniform distribution on $\{0, \ldots, 249\}$, with the spiking threshold increment set in terms of the baseline threshold V_{th} as $\delta V_{th} = 10^{-3} V_{th}$. We then define intervals of varying lengths by grouping together uniformly randomly sampled sequences of neurons as described in Section II C. We fit the generative factor model described in Section II C to the intervals generated by our network using the expectation-maximization algorithm described in [35]. In Figure 3b, we plot the square root of the local and global components of variability as a function of interval duration to more clearly illustrate their scaling. We observe good qualitative agreement between the theory and the results of the numerical experiments.

III. TRIAL-TO-TRIAL TIMING VARIABILITY IN A NOISY HOMOGENEOUS SYNFIRE CHAIN

In Section II, we considered a chain of single neurons with simplified dynamics and coupling for the sake of analytical tractability. In this section, we study variability in a more realistic neural network, a synfire chain [26, 27, 42–44]. A synfire chain is a feed-forward network of multiple pools of neurons, also termed nodes or layers, which are linked by excitatory synaptic connections. We model the neurons in the synfire chain as bursting neurons, and add a set of readout neurons. Bursts have been known to stabilize synfire chains [30], however, we note that the structure of variability does not change if

we model neurons that emit a single spike rather than bursts.

We construct a chain of N pools, each composed of M identical neurons. Every neuron in a given pool i is connected to all neurons in the next pool $i+1$ with equal weights. As before, we approximate the external input to each neuron $I(t)$ by the sum of an average part $g(t)$ and a zero-mean Gaussian process $\sqrt{\tau_m}\eta(t)$ with autocorrelation $\langle \eta(t)\eta(t')\rangle = \sigma_{\text{neuron}}\delta(t-t')$. Neurons in the first pool receive an extra input $J(t)$, which is modeled as a rectangular pulse of height J_0 and width T_p. In addition to the per-neuron noise, we include another noise term, which we refer to as "shared" noise. This noise $\xi(t)$ is generated by a white Gaussian process but is shared across all neurons in a given pool, with mean $\langle \xi(t)\rangle = 0$ and autocorrelation $\langle \xi(t)\xi(t')\rangle = \sigma_{\text{pool}}\delta(t-t')$. Our motivation for including this additional noise term will become clear later. Then, the sub-threshold dynamics of the membrane potential of neuron j in pool i are given by

$$\tau_m \dot{V}_j^{(i)} = E_l - V_j^{(i)} + J_j^{(i)}(t) + \sqrt{\tau_m}\eta_j^{(i)}(t) + g_j^{(i)}(t) + \sqrt{\tau_m}\xi^{(i)}(t), \tag{18}$$

where $J_j^{(i)}$ is zero for all $i > 1$. The synaptic input $g_j^{(i)}(t)$ is modeled by the low-pass-filtered spike train [37]

$$\tau_s \dot{g}_j^{(i)} = -g_j^{(i)} + \frac{\tau_s I_s}{M}\sum_{k=1}^{M}\sum_{l=0}^{S-1}\delta(t - t_{ps,k}^{(i-1)} - \tau_b l), \tag{19}$$

where $t_{ps,k}^{(i-1)}$ denotes the first-spike time of the kth neuron of the $(i-1)^{\text{th}}$ pool and I_s sets the strength of synaptic coupling. The neurons are modeled to emit a burst of S spikes separated by a fixed interval τ_b, rather than a single spike. When a neuron's membrane potential reaches the firing threshold V_{th}, it is then fixed at that threshold until the specified burst duration has elapsed, at which point it is reset to V_r, and once again evolves according to the sub-threshold dynamics (18). As in Section II B, we model neural fatigue as a small increase in the membrane potential threshold of all the neurons in the chain after each trial. We also considered a different neural mechanism for fatigue, a simplified model of synaptic depression. However, the structure of the resulting trial-to-trial timing variability is independent of which neural mechanism of fatigue we implement.

For each pool in the chain, we have a single readout neuron, which receives synaptic input from all neurons of that pool along with a white Gaussian noise input with mean zero and autocorrelation $\sigma_{\text{readout}}\delta(t-t')$ (Figure 1a). The dynamics of membrane potential and synaptic inputs for the readout neurons are similar to that in (18) and (19) with the appropriate inputs.

We study the statistical structure of timing variability in this synfire chain model using numerical simulations. As for the simple model (II D), we integrate the Langevin dynamics (18), (19) using the Euler-Maruyama method, with a timestep of 10^{-2} ms. The parameters values we use are not unique, and are chosen to be in a biologically plausible range. Unless otherwise noted, we simulate a chain of $N = 81$ pools of $M = 32$ neurons each. We set the reset and resting potentials to -70 mV, the baseline spiking threshold to -45 mV, and the synaptic strength to $I_s = 45$ mV. We fix the membrane constant τ_m to 20 ms, the synaptic time constant τ_s to 5 ms, the number of bursts S to 4, and the spacing of bursts τ_b to 2 ms. Unless otherwise noted, we let the strengths of the per-neuron, per-pool, and readout noise be 0.5 mV, 1 mV, and 3 mV, respectively. As in Section II D, we fix the spiking threshold increment δV_{th} to $10^{-3}V_{th}$, and draw the multiplicative increment factor from the discrete uniform distribution on $\{0,\ldots,249\}$ for each trial. Propagation in a synfire chain is not always successful [27, 44]. We consider a trial to be successful if the total number of spikes in the chain is between $4N$ and $4N + 0.1\,(4N)$, and if all readout neurons fire once. In the experiments on which Figures 4, 5, and 6 were based, two trials out of 1000 were excluded from our simulations that included all sources of noise. In Figure 7c no trials were excluded from the inset.

A. Relating neuronal mechanisms to different components of trial-to-trial timing variability

In Section II B, we showed that introducing correlated variability to a chain of IF neurons through a simple model of neural fatigue yields a spike interval covariance matrix that is the sum of a local component and a rank-one global component. To test whether this structure is present in the trial-to-trial timing variability of the noisy homogeneous synfire chain model, we define intervals by grouping together sequences of ten neurons, yielding intervals with a mean duration of 59.5 ms and the covariance structure shown in Figure 4 a. Then, we take the full model covariance matrix and use the generative model proposed in [35] (see Section II C) to decompose it into three components, a local component, a global component, and a jitter component, Figure 4. We found that the resulting decomposition explained the covariance matrix well, with a standardized root mean squared residual of 0.0067 [35]. Thus, the statistical structure of trial-to-trial timing variability in the noisy homogeneous synfire chain model is consistent with that of the simple model, with the addition of the component corresponding to readout noise.

Next, we delineate the neural mechanisms behind the components of variability by selectively including different sources of noise. First, we include only the chain noise, which comprises of the shared and the neuron-specific noise terms in the input to a neuron shown in (18). In this case, we recover a diagonal covariance matrix, which we identify as local variability Figure 5b. When we include only neural fatigue, we recover the global component of variability producing a nearly rank-

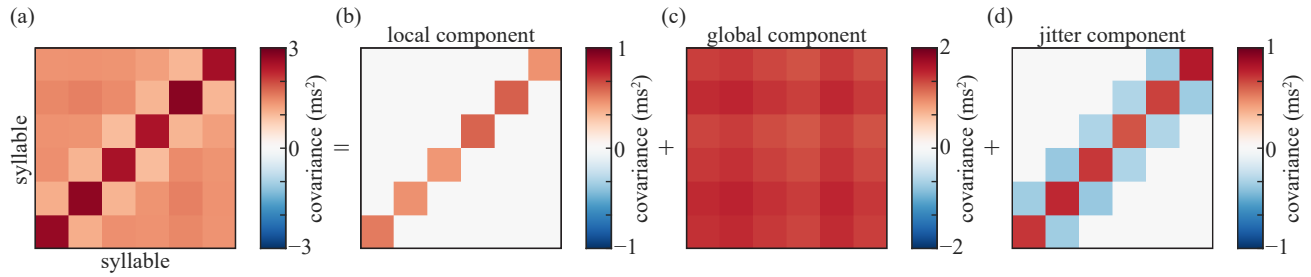

FIG. 4. Factor decomposition of the interval duration covariance matrix of the noisy homogeneous synfire chain model. (a) Full model covariance matrix. (b-d) The covariance matrices of the latent factors resulting from applying the analysis method of [35] to the full model shown in (a).

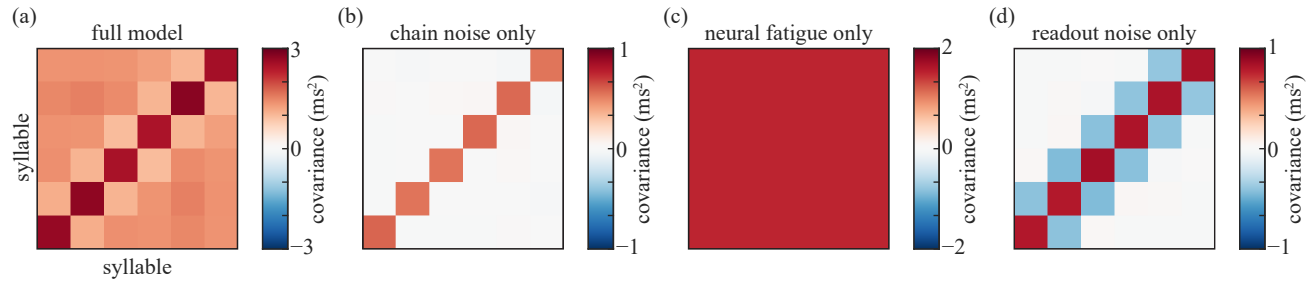

FIG. 5. Interval duration covariance matrices of the noisy homogeneous synfire chain model. (a) Interval duration covariance matrix for the full model. (b-d) Interval duration covariance matrices due to chain noise, neural fatigue, and readout noise alone, respectively. In Figure 7 (a,b) we show the contribution of each of the noise sources in (18) to the interval duration covariance matrix separately.

1 covariance matrix, Figure 5c. If we only include noise in the readout neurons, the duration of neighboring intervals are anti-correlated (Figure 5d), corresponding to jitter. This exercise allows us to identify distinct cellular and synaptic mechanisms that explain distinct components of temporal variability: chain noise contributes to the local component, fatigue to the global component, and readout noise to the jitter component.

B. Scaling of the components of variability with interval duration

In Section II C, we observed that, if one groups multiple neurons together to form intervals, the local component of variability should scale with the square root of the interval length, while the global component should scale linearly with interval length. This scaling is independent of the details of the model, and simply follows from the assumption that the total trial-to-trial variability of the spike interval can be decomposed into a local component and a global component, both of which are uniform in magnitude across neurons. Applying the factor analysis method introduced in [35] to the spike times produced by the noisy homogeneous synfire chain model (Figure 4), we find that the scaling of local and global components of variability with interval length is consistent with this prediction (Figure 6).

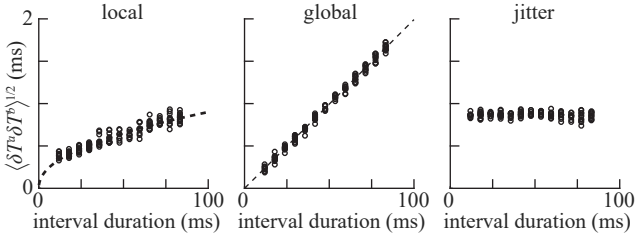

FIG. 6. Scaling of local, global, and jitter components of variability with interval duration in the noisy homogeneous synfire chain model with 32 neurons per pool. Circular markers indicate the results of numerical simulations, and dashed lines show power-law fits to the data, with exponents of 0.46 and 1.00 for the local and global components, respectively. No fit is shown for the jitter component, as there does not exist a statistically significant Spearman correlation between it and the interval duration ($\rho = -0.09$, $p = 0.25$).

C. Scaling of local variability with pool size

In Figure 7a,b, we show the contribution of each of the noise sources in (18), the per-neuron noise and the

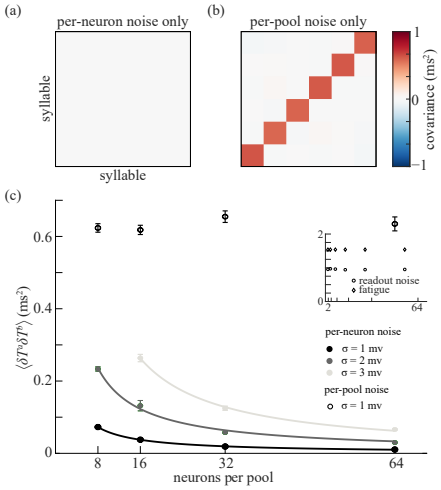

FIG. 7. Noisy homogeneous synfire chain model: (a) and (b) show the contribution of each of the noise sources in (18), the per-neuron noise and the per-pool noise which is shared across neurons in the same pool, to the covariance matrix of interval duration.(c) Scaling of interval variability with the number of neurons per pool due to the two noise sources, the per-neuron noise (filled circles) and shared noise among neurons in the same pool (open circles). The filled black, gray, and light gray circles show the interval variability due to per-neuron noise with $\sigma = 1$ mV, 2 mV, and 3 mV, respectively. Solid lines are power-law fits to the data, with exponents of -0.95 (black line), -0.94 (gray line) and -1.04 (light gray line). Open circles show the same thing but for noise that is shared among neurons in the same pool with $\sigma = 1$ mV. Error bars show the standard error of the mean. The data point for per neuron noise of $\sigma = 3$ mV and $M = 8$ was excluded because the chain propagation failed in more than 10% of trials

.

per-pool noise which is shared across neurons in the same pool, to the covariance matrix of interval duration. If we vary the number of neurons per pool M, we find that the interval duration variance due to per-neuron component of chain noise falls as $\frac{1}{M}$ (Figure 7c). Thus, despite the fact that the system is nonlinear, such noise adds quasi-linearly. Therefore, to have a non-negligible local component of variability in the noisy homogeneous synfire chain model, we must assume that neurons belonging to the same pool receive shared noise. In Figure 7c, we see that the interval variability due to this noise mechanism is roughly independent of the number of neurons per pool. Varying the number of neurons per pool did not have an effect on the readout noise or fatigue. This is shown in the inset of Figure 7c.

IV. THE STRUCTURE OF VARIABILITY IN A BIOLOGICALLY REALISTIC SYNFIRE CHAIN MODEL OF HVC IS CONSISTENT WITH THAT OBSERVED IN ZEBRA FINCH SONG

Zebra finch song is a behavior for which a synfire chain is thought to be involved in the timing of [45]. Zebra finch songs consist of several introductory notes, followed by a few renditions of a motif, sung in a very repetitive manner. Motifs contain 3 to 8 syllables. Syllables range from 50-100 ms and are separated by gaps. The timing of the song is controlled by clock-like bursting in the premotor nucleus HVC, in particular in HVC neurons projecting to Robust Nucleus of the Arcopallium (RA). Many studies suggest that the underlying neural circuit behavior is consistent with a synfire chain model [13, 14, 32, 46, 47]. Further, experimental evidence supports millisecond scale correlations between HVC activity and the song [16, 18, 20]. Thus, we want to test if the trail-to-trail variability in a synfire chain is also consistent with the trial-to-trial variability observed in the song duration. Detailed studies of the trial-to-trial variability in the highly stereotyped zebra finch song were done by Glaze and Troyer [33–35].

We simulate a biologically realistic HVC synfire chain model introduced in [32], which has been shown to agree with measurements of neural variability in HVC. We provide a detailed description of this model in Appendix B. This model consists of a sequentially connected chain of 70 pools, each containing 30 $\mathrm{HVC_{RA}}$ neurons, along with a population of 300 $\mathrm{HVC_I}$ inhibitory interneurons. $\mathrm{HVC_{RA}}$ neurons are modeled as two-compartment bursting neurons incorporating dendritic calcium spikes, while $\mathrm{HVC_I}$ neurons are modeled as single-compartment non-bursting neurons. Inhomogeneity is introduced by randomizing the existence and strength of connections between neurons. Briefly, $\mathrm{HVC_{RA}}$ neurons connect to $\mathrm{HVC_I}$ neurons with probability 0.05, and $\mathrm{HVC_I}$ neurons connect to $\mathrm{HVC_{RA}}$ neurons with probability 0.1. Each $\mathrm{HVC_{RA}}$ connects to an $\mathrm{HVC_{RA}}$ neuron in the next pool with probability $P = 0.5$ and a connection strength drawn from the uniform distribution on $[0, g_{\mathrm{EEmax}}/(30P)]$. Each neuron receives noisy synaptic input in the form of Poisson spike trains which constitutes the chain noise for this model. Synaptic fatigue was modeled by modifying g_{EEmax} as $(1 - m\delta g)g_{\mathrm{EEmax}}$, where $\delta g = 10^{-3}$, with $g_{\mathrm{EEmax}} = 3\,\mathrm{mS/cm^2}$, and m drawn i.i.d. across trials from the uniform distribution on $\{0, \ldots, 249\}$ as before. All remaining model parameters values are set to those used in Long *et al.* [32]. We read out timing information from the chain using the readout model introduced in Section III. The model was integrated using the Euler-Maruyama method with a timestep of 5×10^{-3} ms.

We first observe that the full model covariance matrix of interval duration of this model (Figure 8 a) is similar to that of the homogeneous synfire chain model (Figure 5 a), and has the same structure and magnitude as the song

syllable interval duration covariance matrix reported by Glaze and Troyer [35]. Glaze and Troyer [35] showed that the generative model given in (14) is a good description of the statistical structure of the zebra finch song interval duration variability. We previously showed that the structure of variability of the noisy homogeneous synfire chain model is well-described by this generative model (Section III); we find the same to be true for the HVC synfire chain model (Figure 9). The magnitude of variability in the model for 50 ms interval is of order 1 ms (for each component), which is consistent with experimental findings (Figure 10; data from [35]). As in Section III A, we use the results of Section II C to connect behavioral variability to neural mechanisms. Consistent with our results in Section III A, we find that the chain noise contributes to the local component of the song variability, fatigue contributes to the global component, and readout noise contributes to the jitter component (Figure 8).

We also examine how the different components of variability in syllable duration scale with syllable duration. The scalings of the components of variability of syllable duration (Figure 11) agree with the predictions of Section II C, as do previous models (Figures 3b and 6) and the experimental data (Figure 10).

Finally, we note that unlike the noisy homogeneous synfire chain model (Section III), there is no need to add per-pool noise, which is shared across neurons in the same pool. That per-pool noise was necessary in the noisy homogeneous synfire chain model to obtain a non-negligible local component of variability because the variability in input currents to different neurons in a pool would otherwise be uncorrelated (Section III C). In the HVC synfire chain model, sufficient local variability is created by variable synaptic inputs due to uncorrelated presynaptic Poisson spike trains and correlated noisy inputs from inhibitory neurons connecting to multiple neurons in a pool.

V. DISCUSSION AND CONCLUSION

In this paper, we presented analytical and computational analyses of the trial-to-trial timing variability in synfire chains. We first show how variability scales with input strength and noise level in a simple, analytically tractable chain of single neurons in a low firing rate regime. We also show how trial-to-trial variability scales with interval duration in this simple model. Then, we demonstrated with simulations that our main results carry to noisy homogeneous synfire chains. We found that the statistical structure of timing variability in the chain is well-described by a generative model which consists of local, global and jitter components. Furthermore, we were able to relate each of these components to distinct neural mechanisms in the model. Finally, we show that our main results hold in a biologically realistic synfire chain model of the premotor nucleus HVC in songbirds, and that the structure and magnitude of variability in the model agree with that observed in songbirds song.

Our findings have important implications for the relationship between neural and behavioral variability. Even the most stereotyped of animal behaviors, like the songs of zebra finches, show significant trial-to-trial variability [34]. This variability can be an unavoidable nuisance, or it might be there for an advantageous reason, for example to allow the system to explore more of the behavioral space to help in trial-and-error motor learning [1–3] or help in social contexts [48]. Therefore, understanding the mechanisms which generate and regulate trial-to-trial variability has gained considerable interest [2, 4, 7].

Particularly, it has been argued that some of this variability is rooted in neural activity that controls behavior [2, 4, 7]. Measurements of neural activity show significant trial-to-trial variability [2, 4, 7, 49]. These variations are not independent of behavior; for example, they are known to correlate with behavioral choice in a trial-by-trial basis [50]. Given this background, it is natural to ask whether temporal variability of behavior should reflect the statistics and structure of temporal variability of neural circuits that represent or govern the behavior's timing in a trial-to-trial basis [9, 12]. Indeed, interrelationships between the timings of neural dynamics and behavior have been observed in various experimental studies [13–20]. For example, Srivastava *et al.* [18] show that in Bengalese finches millisecond-scale changes in the timing of a single spike in a burst in respiratory muscle fibers predicts significant differences in breathing dynamics and millisecond-scale variations in precisely-timed electrical stimulation of respiratory muscles strongly modulate breathing output. This millisecond scale spike timing control of behavior extends to other animals and behaviors [20].

As the temporal variability in zebra finch song and in HVC neurons are both on the millisecond scale [13, 15, 32–35], we speculated that they may be linked. We showed that the temporal variability observed in a biologically realistic model of the zebra finch HVC chain [32] is consistent with the magnitude and structure of the timing variability in the zebra finch song. Thus, our findings provide an example of a detailed match between neural and behavioral variability, and suggest a possible neural account of behavioral variability. A direct experimental test of this suggestion would be to look for correlations in a trial-by-trial basis between millisecond scale temporal variations in the song and spike times of RA projecting HVC neurons.

Potential weaknesses of our neural explanation of the zebra finch song variability are the following. First, while the dominant hypothesis for the song-timing circuit is a synfire chain [32], this question is not fully settled. Further, recent work suggests that the song-timing circuit may not even be fully localized within the HVC but is distributed across multiple areas [51]. If the distributed circuit is a synfire chain, which is consistent with the results of Hamaguchi *et al.* [51], our results still remain valid. Second, in zebra finches, variability of song can

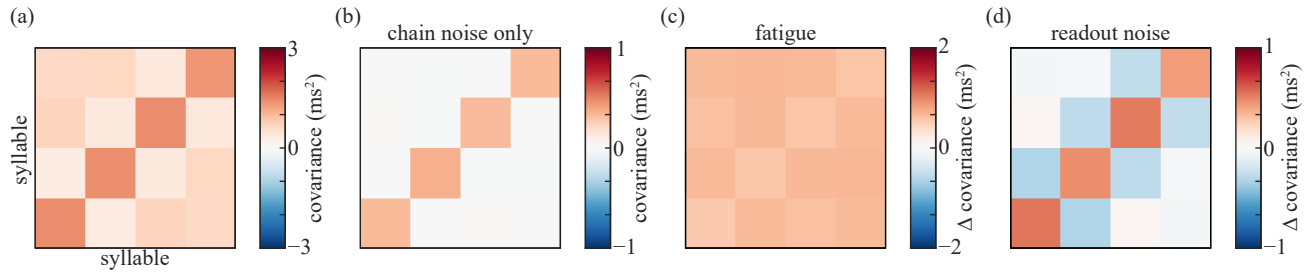

FIG. 8. Contributions of different sources of noise in the HVC synfire chain model to the interval duration covariance matrix (model modified from Long *et al.* [32]). (a) Interval duration covariance matrix for the full model. (b) Interval duration covariance matrix with chain noise only. (c) Difference between covariance matrix with chain noise and fatigue and that with chain noise only. (d) Difference between covariance matrix with readout and chain noise and that with chain noise only.

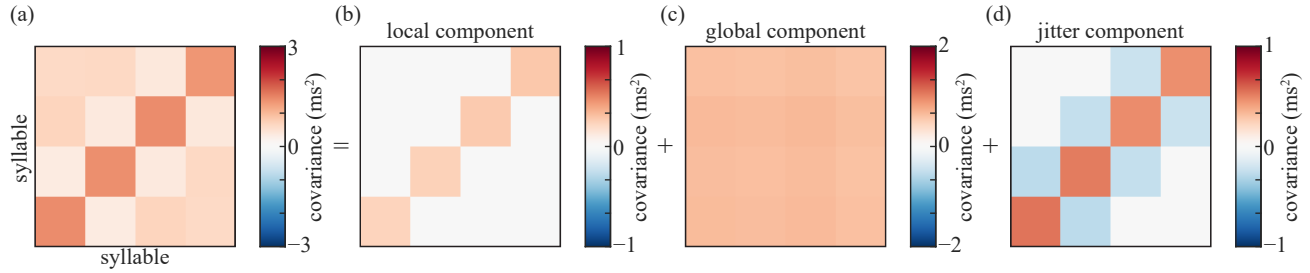

FIG. 9. Factor decomposition of the interval duration covariance matrix of the HVC synfire chain model (model modified from Long *et al.* [32]). (a) Full model covariance matrix. (b-d) The covariance matrices of the latent factors resulting from applying the analysis method of [35] to the full model shown in (a).

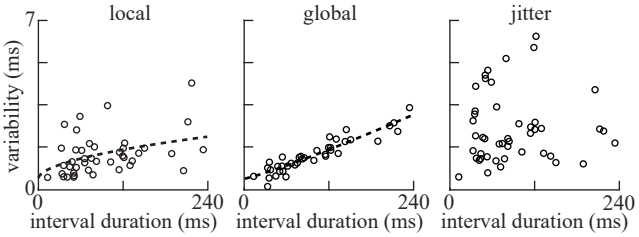

FIG. 10. Scaling of the local, global, and jitter components of syllable timing variability with interval duration in zebra finch song (data from Glaze and Troyer [35]). Glaze and Troyer [35] recorded songs of zebra finches, which are composed of a stereotyped sequence of syllables and gaps, a different sequence for each bird. After identifying syllables and gaps, and their durations in each song, they fitted the generative model described in Section II C to this dataset, separately for each bird. Reported data in the figure is extracted from their Figure 3. Each data point represents a syllable-bird pair. As in Figure 6, we fit the relationships between interval duration and the local and global components of variability with power laws, yielding exponents of 0.53 and 1.14, respectively. No fit is shown for the jitter component, as it is not significantly correlated with interval duration.

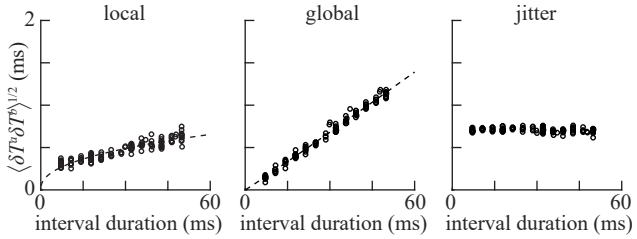

FIG. 11. Scaling of the local, global, and jitter components of interval duration variability in the HVC synfire chain model (model modified from Long *et al.* [32]). As in Figure 6, we fit the relationships between interval duration and the local and global components of variability with power laws, yielding exponents of 0.42 and 1.03, respectively. No fit is shown for the jitter component, as it is not significantly correlated with interval duration.

be actively regulated through involving lateral magnocellular nucleus of the anterior neopallium (LMAN) [52]. Indeed Ali *et al.* [15] observed that LMAN lesions lead to a reduction in the local component of song timing variability, which was speculated to be mediated by indirect LMAN input to HVC [53]. Another possible source of variability is sensory inference errors [5]. HVC receives feedback auditory input through the nucleus interfacialis of the nidopallium [54], and altered auditory feedback can lead to temporal changes in the song [55]. In the HVC synfire chain model [32], noise was introduced as excitatory and inhibitory independent Poisson spike trains, with no specific reference to where such trains may come

from and how could they be regulated. Third, the song production pathway goes from HVC to RA, and then from RA to the tracheosyringeal part of the hypoglossal nerve (nXIIts), which then controls muscular contractions of the syrinx. Neural variability in all these areas as well as variability in muscular contractions contribute to temporal variability of the song. In our model, all of this pathway's contributions to temporal variability are incorporated into the noise injected to a readout neuron, which in turn contributes mostly to temporal jitter. It is very possible that other components of variability are affected by the downstream activity. Finally, a notion of tempo variation that we did not consider arises from structural changes to the chain, such as homeostatic and synaptic plasticity [15, 46, 47], or experimental perturbations [14]. In birdsong, these mechanisms can lead to tempo changes on the order of tens of milliseconds [14, 15, 46, 47], and, when naturally occurring, require thousands of song repetitions to take effect [15].

ACKNOWLEDGMENTS

Part of this work was done when D. Obeid was at the Center for Theoretical Neuroscience at Columbia University. C. Pehlevan acknowledges support from the Intel Corporation. A subset of the computations in this paper were run on the FASRC Cannon cluster supported by the FAS Division of Science Research Computing Group at Harvard University. We thank the anonymous reviewers for a thorough and constructive review.

Appendix A: Asymptotics for the chain of integrate-and-fire neurons

1. Stationary membrane potential distribution in the low-rate limit

Here, we review the approximate firing rate and stationary membrane potential distribution in the low-rate limit $V_{th} - I_0 \gg \sigma$ [36]. Inspecting the equation

$$\nu = \frac{1}{\tau\sqrt{\pi}} \left[\int_{\frac{V_r - I_0}{\sigma}}^{\frac{V_{th}-I_0}{\sigma}} du\, e^{u^2} (1 + \mathrm{erf}(u)) \right]^{-1}, \tag{A1}$$

we can see that the integral is dominated by the upper limit due to the exponential. Making the change of variables $u' \equiv u/h$, $h \equiv (V_{th} - I_0)/\sigma$, $v \equiv (V_0 - I_0)/(V_{th} - I_0)$, we have

$$\begin{aligned} \nu &= \frac{\sigma}{\tau\sqrt{\pi}\,(V_{th} - I_0)} \left[\int_v^1 du\, e^{h^2u^2} (1 + \mathrm{erf}(hu)) \right]^{-1} \\ &\approx \frac{\sigma}{2\tau\sqrt{\pi}\,(V_{th} - I_0)} \left[\int_v^1 du\, e^{h^2u^2} \right]^{-1} \\ &\approx \frac{\sigma}{2\tau\sqrt{\pi}\,(V_{th} - I_0)} \left[\left. \frac{e^{h^2u^2}}{2uh^2} \right|_v^1 \right]^{-1} \\ &\approx \frac{V_{th} - I_0}{\sigma\tau\sqrt{\pi}} \exp\left(-\frac{(V_{th} - I_0)^2}{\sigma^2} \right). \end{aligned} \tag{A2}$$

where we integrated by parts in the third line. By a similar argument, we can approximate the stationary distribution of the membrane potential as

$$\int_{\frac{V - I_0}{\sigma}}^{\frac{V_{th}-I_0}{\sigma}} du\, \Theta\left(u - \frac{V_0 - I_0}{\sigma} \right) e^{u^2} \approx \frac{1}{2\nu\tau\sqrt{\pi}}, \tag{A3}$$

in this limit.

2. Moments of the first-spike-interval in the low-rate stationary limit

To derive the moments of the first-spike-interval in the low-rate stationary approximation, we start with the standard results (given as (6) and (7) in the main text) for the mean and variance conditioned on V_0 [36, 37]:

$$\langle T_{fs} \rangle_{V_0} = \sqrt{\pi}\tau \int_{(V_0 - I_0 - I_s)/\sigma}^{(V_{th} - I_0 - I_s)/\sigma} dy\, e^{y^2} \left[1 + \mathrm{erf}(y) \right] \tag{A4}$$

and

$$\begin{aligned} \langle \delta T_{fs}^2 \rangle_{V_0} &= 2\pi\tau^2 \int_{(V_0 - I_0 - I_s)/\sigma}^{(V_{th} - I_0 - I_s)/\sigma} dx\, e^{x^2} \\ &\quad \times \int_{-\infty}^{0} dy\, e^{(x+y)^2} \left[1 + \mathrm{erf}(x+y) \right]^2. \end{aligned} \tag{A5}$$

Considering the mean first-spike-interval, we use the integral representation of the error function as

$$1 + \mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_{-\infty}^{0} du\, e^{-(u+x)^2} \tag{A6}$$

to write

$$\begin{aligned} \frac{\langle T_{fs} \rangle_{V_0}}{\tau} &= 2 \int_{(V_0 - I_0 - I_s)/\sigma}^{(V_{th} - I_0 - I_s)/\sigma} dy\, e^{y^2} \int_{-\infty}^{0} du\, e^{-(u+y)^2} \\ &= \int_{-\infty}^{0} \frac{du}{u}\, e^{-u^2} \\ &\quad \times \left[e^{2u(I_0 + I_s - V_0)/\sigma} - e^{2u(I_0 + I_s - V_{th})/\sigma} \right], \end{aligned} \tag{A7}$$

where we note the cancellation in the bracketed integrand that ensures that it does not diverge as $u \to 0^-$. We can then easily compute the expectation over the approximate distribution (5) of V_0 to obtain

$$\frac{\langle T_{fs}\rangle}{\tau} = \int_{-\infty}^{0} \frac{du}{u}\, e^{-u^2} \left[e^{u^2+2I_s u/\sigma} - e^{2u(I_0+I_s-V_{th})/\sigma} \right], \tag{A8}$$

which, though it does not have a simple closed-form solution, is the integral of a bounded entire function that decays exponentially fast at infinity (provided that $I_s > 0$), and is therefore well-behaved.

To obtain a similar integral expression for the variance of the first-spike-interval, we recall the law of total variance

$$\langle \delta T_{fs}^2 \rangle = \langle \langle \delta T_{fs}^2 \rangle_{V_0} \rangle + \langle \delta \langle T_{fs} \rangle_{V_0}^2 \rangle, \tag{A9}$$

where the outer angle brackets denote averaging over the distribution of V_0, and follow the same procedure that we used to derive $\langle T_{fs}\rangle$ to obtain

$$\frac{\langle \langle \delta T_{fs}^2 \rangle_{V_0} \rangle}{\tau^2} = 4 \int_{-\infty}^{0} du \int_{-\infty}^{0} dv \int_{-\infty}^{0} dw\, \frac{e^{-(u+v+w)^2+2uv}}{u+v+w} \times \left[e^{(u+v+w)^2+2(u+v+w)I_s/\sigma} - e^{2(u+v+w)(I_0+I_s-V_{th})/\sigma} \right], \tag{A10}$$

and

$$\frac{\langle \delta \langle T_{fs} \rangle_{V_0}^2 \rangle}{\tau^2} = \int_{-\infty}^{0} \frac{du}{u} \int_{-\infty}^{0} \frac{dv}{v}\, e^{-u^2-v^2} \times \left[e^{(u+v)^2+2I_s(u+v)/\sigma} - e^{u^2+v^2+2I_s(u+v)/\sigma} \right]. \tag{A11}$$

With these integral expressions in hand, we can now derive asymptotic expansions for the moments. For brevity, we define the dimensionless scalars $\alpha \equiv (I_0 + I_s - V_{th})/\sigma$ and $\beta \equiv (V_{th} - I_0)/\sigma$; we will work in the limit of low baseline firing rates $\beta \gg 1$ and large synaptic inputs $\alpha \gg 1$. Rescaling u by 2α in (A8), we can write the mean first-spike-interval as

$$\begin{aligned} \frac{\langle T_{fs}\rangle}{\tau} &= \int_{-\infty}^{0} \frac{du}{u}\, e^{u} \left[e^{(\beta/\alpha)u} - e^{-u^2/4\alpha^2} \right] \\ &= \log\left(\frac{\alpha+\beta}{\alpha}\right) - \int_{-\infty}^{0} \frac{du}{u}\, e^{u} \left[e^{-u^2/4\alpha^2} - 1 \right], \end{aligned} \tag{A12}$$

where we have split the integral into two pieces by adding and subtracting one from the integrand and evaluated the first of the remaining integrals. Expanding the remaining integrand other than the overall exponential weight e^u as a power series and integrating term-by-term using the relationship of the integrand to the gamma function [56], we obtain the divergent asymptotic series

$$\begin{aligned} \frac{\langle T_{fs}\rangle}{\tau} &\sim \log\left(\frac{\alpha+\beta}{\alpha}\right) + \sum_{k=1}^{\infty} \frac{(-1)^k (2k-1)!}{4^k k! \alpha^{2k}} \\ &\sim \log\left(\frac{\alpha+\beta}{\alpha}\right) - \frac{1}{4\alpha^2} + \mathcal{O}\left(\alpha^{-4}\right), \end{aligned} \tag{A13}$$

which yields the lowest-order approximation given in the main text.

We now consider $\langle \delta T_{fs}^2 \rangle$. Converting the integral over the negative octant in (A10) to an integral over the positive octant, making the change of variables $x \equiv u$, $y \equiv v + w$, $z \equiv w$, and parameterizing the domain of integration such that we integrate first over $z \in [0, y]$, we have

$$\frac{\langle \langle \delta T_{fs}^2 \rangle_{V_0} \rangle}{\tau^2} = 2 \int_0^{\infty} dx \int_0^{\infty} dy\, \frac{e^{-2\alpha(x+y)}}{x(x+y)} \left[e^{2xy} - 1 \right] \times \left[e^{-(x+y)^2} - e^{-2\beta(x+y)} \right]. \tag{A14}$$

Then, adding the expression for $\langle \delta \langle T_{fs} \rangle_{V_0}^2 \rangle$ given in (A11) to the above expression for $\langle \langle \delta T_{fs}^2 \rangle_{V_0} \rangle$ as prescribed by the law of total variance (A9), we have

$$\frac{\langle \delta T_{fs}^2 \rangle}{\tau^2} = \int_0^{\infty} dx \int_0^{\infty} dy\, e^{-2\alpha(x+y)} \left[e^{2xy} - 1 \right] \times \left[\frac{2e^{-(x+y)^2}}{x(x+y)} + \frac{1}{xy} \left(\frac{x-y}{x+y} \right) e^{-2\beta(x+y)} \right]. \tag{A15}$$

As it is anti-symmetric about the line $y = x$, the second term in the bracketed integrand will vanish under integration over the positive quadrant, leaving

$$\frac{\langle \delta T_{fs}^2 \rangle}{\tau^2} = 2 \int_0^{\infty} dx \int_0^{\infty} dy\, \frac{e^{-(x+y)^2-2\alpha(x+y)}}{x(x+y)} \left[e^{2xy} - 1 \right]. \tag{A16}$$

Rescaling x and y by 2α and making the change of variables $u \equiv x + y$, $v \equiv x$, we have

$$\frac{\langle \delta T_{fs}^2 \rangle}{\tau^2} = 2 \int_0^{\infty} \frac{du}{u} \int_0^{u} \frac{dv}{v}\, e^{-u^2/4\alpha^2 - u} \left[e^{v(u-v)/2\alpha^2} - 1 \right]. \tag{A17}$$

Expanding the bracketed portion of the integrand as a power series and observing that

$$\int_0^{u} dv\, v^k (u-v)^{k+1} = \frac{k!(k+1)!}{(2k+2)!} u^{2k+2}, \tag{A18}$$

we have, integrating over u term-by-term,

$$\frac{\langle \delta T_{fs}^2 \rangle}{\tau^2} = \sum_{k=0}^{\infty} \frac{k!}{2^k (2k+2)! \alpha^{2k+2}} \int_0^{\infty} du\, e^{-u^2/4\alpha^2 - u} u^{2k+1}. \tag{A19}$$

To allow us to apply standard asymptotic results to the remaining integral, we note that it is related to Tricomi's confluent hypergeometric function $U(a,b,z)$ as $\alpha^{2k+2}(2k+1)!\,U(k+1,1/2,\alpha^2)$ [56], hence, shifting indices for convenience, we can write

$$\frac{\langle \delta T_{fs}^2\rangle}{\tau^2} \sim \sum_{k=1}^{\infty} \frac{(k-1)!}{2^k k} U\left(k, \frac{1}{2}, \alpha^2\right). \tag{A20}$$

Using the standard result that

$$U(a,b,z) \sim z^{-a}\Big[1 - \frac{a(a-b+1)}{z} + \frac{a(a+1)(a-b+2)(a-b+1)}{2z^2} + \mathcal{O}(z^{-3})\Big], \tag{A21}$$

for $|z| \gg 1$ [56], we have

$$\frac{\langle \delta T_{fs}^2\rangle}{\tau^2} \sim \frac{1}{2\alpha^2} - \frac{1}{8\alpha^4} + \mathcal{O}(\alpha^{-6}), \tag{A22}$$

which yields the lowest-order approximation given in the main text.

To obtain the asymptotic approximations for the timing variability in the simple model for neural fatigue given in the main text (10, 11, 12), we start from the asymptotic expansions without fatigue (8, 9), and apply the laws of total expectation and total variance given the assumed distribution of the parameter m. We then expand the resulting expressions about the baseline spiking threshold V_{th} to lowest order in $\delta V_{th}/(I_s + I_0 - V_{th})$, assuming that $m_{\max}\,\delta V_{th} \ll I_s + I_0 - V_{th}$, yielding the asymptotic approximations (10) and (11).

3. Moments of the first-spike-interval in a delta-function approximation

In the previous appendix and in the main text, we considered the approximation of the distribution of initial membrane potentials by the stationary Gaussian limit (5). In this appendix, we consider a delta-function approximation $P(V_0) \approx \delta(V_0 - \langle V_0\rangle)$. This approximation maps directly to the standard treatment of leaky IF neurons with the appropriate replacement of V_r by I_0. Here, we review the derivation of the corresponding asymptotic results [37, 38]. In the limit $V_{th} - I_0 \gg \sigma$ of low firing rates, we have $\langle V_0\rangle = I_0$, hence we fix $V_0 = I_0$ in this approximation. Considering the mean first-spike-interval, we again start from the standard expression (6) with V_0 set to I_0, and rescale $\sigma y \mapsto y$, yielding

$$\frac{\langle T_{fs}\rangle}{\tau} = \frac{\sqrt{\pi}}{\sigma}\int_{-I_s}^{V_{th}-I_0-I_s} dy\, e^{y^2/\sigma^2}\left(1 + \operatorname{erf}\left(\frac{y}{\sigma}\right)\right). \tag{A23}$$

In the limit $I_s + I_0 - V_{th} \gg \sigma$ of large synaptic inputs, the quantity y in the above integrand is always negative, and we have $y/\sigma \ll -1$. Using the asymptotic expansion of the error function for $x \ll -1$ [56],

$$\operatorname{erf}(x) \sim -1 + \frac{e^{-x^2}}{\sqrt{\pi}|x|}\left(1 - \frac{1}{2x^2} + \dots\right), \tag{A24}$$

we obtain

$$\frac{\langle T_{fs}\rangle}{\tau} \sim \log\left(\frac{I_s}{I_s + I_0 - V_{th}}\right) - \frac{\sigma^2}{4}\left(\frac{1}{(I_s + I_0 - V_{th})^2} - \frac{1}{I_s^2}\right) \tag{A25}$$

to lowest order. Similarly, for the variance of the first-spike-interval, we start with the standard expression (7) with $V_0 = I_0$. Again rescaling the variables of integration by σ and using the asymptotic form of the error function, we obtain the lowest-order approximation

$$\frac{\langle \delta T_{fs}^2\rangle}{\tau^2} \sim \frac{\sigma^2}{2}\left(\frac{1}{(V_{th} - I_s - I_0)^2} - \frac{1}{I_s^2}\right). \tag{A26}$$

Comparing these expressions to the corresponding results (8, 9) in the approximation of the initial membrane potential distribution by the stationary Gaussian distribution (5), we observe that they are identical up to the presence of the $-I_s^{-2}$ terms in the lowest-order approximations. The presence of these terms in the delta-function approximation means that the variability decreases more rapidly with increasing synaptic strength and increases less rapidly with increasing noise variance σ^2 than in the Gaussian approximation.

Appendix B: Details of the HVC synfire chain model

In this appendix, we provide a detailed description of the HVC synfire chain model from Long *et al.* [32]. This model consists of a chain of 70 sequentially-connected pools of 30 $\mathrm{HVC_{RA}}$ neurons, along with a population of 300 $\mathrm{HVC_I}$ inhibitory interneurons. A given $\mathrm{HVC_{RA}}$ neuron connects to an $\mathrm{HVC_{RA}}$ neuron in the next pool with probability P and an excitatory synaptic conductance drawn from the uniform distribution on $[0, g_{\mathrm{EEmax}}/(30P)]$ mS cm^{-2}, where g_{EEmax} is a dimensionless parameter. An $\mathrm{HVC_{RA}}$ neuron connects to an $\mathrm{HVC_I}$ neuron with probability 0.05 and excitatory synaptic conductance drawn uniformly from $[0, 0.5]$ mS cm^{-2}. Finally, an $\mathrm{HVC_I}$ neuron connects to an $\mathrm{HVC_{RA}}$ neuron with probability 0.1 and an inhibitory synaptic conductance drawn uniformly from $[0, 0.2]$ mS cm^{-2}. Long *et al.* [32] chose these parameter values such that successful spike propagation was possible for many values of g_{EEmax}.

1. $\mathrm{HVC_{RA}}$ dynamics

In the Long *et al.* [32] model, $\mathrm{HVC_{RA}}$ neurons are modeled as two-compartment bursting neurons. The somatic

compartment contains leak, Na+, and delay-rectified K+ conductances, while the dendritic compartment contains leak, high-threshold Ca++, and calcium-activated K+ conductances. This model can generate dendritic calcium spikes, which evoke stereotyped bursts of sodium spikes in the soma. The membrane potentials V_s and V_d of the somatic and dendritic compartments obey the dynamics

$$\begin{aligned} C_\mathrm{m} A_\mathrm{s} \dot{V}_\mathrm{s} &= A_\mathrm{s} \left(I_\mathrm{s,L} + I_\mathrm{s,Na} + I_\mathrm{s,Kdr} + I_\mathrm{s,exc} + I_\mathrm{s,inh} \right) \\ &\quad + I_\mathrm{s,ext} + \frac{V_\mathrm{d} - V_\mathrm{s}}{R_\mathrm{c}} \\ C_\mathrm{m} A_\mathrm{d} \dot{V}_\mathrm{d} &= A_\mathrm{d} \left(I_\mathrm{d,L} + I_\mathrm{d,Ca} + I_\mathrm{d,CaK} + I_\mathrm{d,exc} + I_\mathrm{d,inh} \right) \\ &\quad + I_\mathrm{d,ext} + \frac{V_\mathrm{s} - V_\mathrm{d}}{R_\mathrm{c}}; \end{aligned} \tag{B1}$$

we enumerate the definitions and values of all parameters in Table I. The dynamics of the injected currents $I_\mathrm{s,ext}$ and $I_\mathrm{d,ext}$ are freely chosen, while the remaining currents are given as

$$\begin{aligned} I_\mathrm{s,L} &= -G_\mathrm{L}(V_\mathrm{s} - E_\mathrm{L}) \\ I_\mathrm{s,Na} &= -G_\mathrm{Na} m_\infty^3 h (V_\mathrm{s} - E_\mathrm{Na}) \\ I_\mathrm{s,Kdr} &= -G_\mathrm{Kdr} n^4 (V_\mathrm{s} - E_\mathrm{K}) \\ I_\mathrm{s,exc} &= -g_\mathrm{s,exc}(t) V_\mathrm{s} \\ I_\mathrm{s,inh} &= -g_\mathrm{s,inh}(t)(V_\mathrm{s} - E_\mathrm{I}) \\ I_\mathrm{d,L} &= -G_\mathrm{L}(V_\mathrm{d} - E_\mathrm{L}) \\ I_\mathrm{d,Ca} &= -G_\mathrm{Ca} r^2 (V_\mathrm{d} - E_\mathrm{Ca}) \\ I_\mathrm{d,CaK} &= -G_\mathrm{CaK} \frac{c}{1 + 6[\mathrm{Ca}]^{-1}} (V_\mathrm{d} - E_\mathrm{K}) \\ I_\mathrm{d,exc} &= -g_\mathrm{d,exc}(t) V_\mathrm{d} \\ I_\mathrm{d,inh} &= -g_\mathrm{d,inh}(t)(V_\mathrm{d} - E_\mathrm{I}), \end{aligned} \tag{B2}$$

where $g_\mathrm{s,exc}$, $g_\mathrm{s,inh}$, $g_\mathrm{d,exc}$, and $g_\mathrm{d,inh}$ are the total synaptic conductances of the soma and dendrite. The gating variable $m(t) = m_\infty(V_\mathrm{s})$ is an instantaneous function of V_s, while h, n, r, c all evolve according to the dynamics

$$\tau_x \dot{x} = x_\infty - x \tag{B3}$$

for $x \in \{h, n, r, c\}$, where the activation functions are given as

$$m_\infty(V_\mathrm{s}) = 1/\left(1 + \exp[-(V_\mathrm{s} + 30)/9.5]\right) \tag{B4}$$
$$h_\infty(V_\mathrm{s}) = 1/\left(1 + \exp[(V_\mathrm{s} + 45)/7]\right) \tag{B5}$$
$$n_\infty(V_\mathrm{s}) = 1/\left(1 + \exp[-(V_\mathrm{s} + 35)/10]\right) \tag{B6}$$
$$r_\infty(V_\mathrm{d}) = 1/\left(1 + \exp[-(V_\mathrm{d} + 5)/10]\right) \tag{B7}$$
$$c_\infty(V_\mathrm{d}) = 1/\left(1 + \exp[-(V_\mathrm{d} - 10)/7]\right), \tag{B8}$$

and the time constants are given as

$$\begin{aligned} \tau_h &= 0.1 + 0.75/\left(1 + \exp[(V_\mathrm{s} + 40.5)/6]\right) \\ \tau_n &= 0.1 + 0.5/\left(1 + \exp[(V_\mathrm{s} + 27)/15]\right) \\ \tau_r &= 1 \\ \tau_c &= 10; \end{aligned} \tag{B9}$$

Name	Description	Value
A_s	area of somatic compartment	$5000\,\mu\mathrm{m}^2$
A_d	area of dendritic compartment	$10000\,\mu\mathrm{m}^2$
C_m	membrane capacitance	$1\,\mu\mathrm{F/cm}^2$
R_c	compartment coupling resistance	$55\,\mathrm{M}\Omega$
G_L	leak conductance	$0.1\,\mathrm{mS/cm}^2$
G_Na	Na+ conductance	$60\,\mathrm{mS/cm}^2$
G_Kdr	delay-rectified K+ conductance	$8\,\mathrm{mS/cm}^2$
G_Ca	high-threshold Ca++ conductance	$55\,\mathrm{mS/cm}^2$
G_CaK	Ca-dependent K+ conductance	$150\,\mathrm{mS/cm}^2$
E_L	leak reversal potential	$-90\,\mathrm{mV}$
E_Na	Na+ reversal potential	$55\,\mathrm{mV}$
E_K	K+ reversal potential	$-90\,\mathrm{mV}$
E_Ca	Ca++ reversal potential	$120\,\mathrm{mV}$
E_I	inhibitory reversal potential	$-80\,\mathrm{mV}$

TABLE I. $\mathrm{HVC_{RA}}$ model parameters

the units of all constants are implied. Finally, the calcium concentration [Ca] evolves as

$$[\dot{\mathrm{Ca}}] = 0.1 I_\mathrm{d,Ca} - 0.02[\mathrm{Ca}]. \tag{B10}$$

Synaptic conductances follow "kick-and-decay" kinetics: $g \mapsto g + G$ when a spike arrives at a synapse with conductance G; $\tau \dot{g} = -g$ between spikes, for $g \in \{g_\mathrm{exc}, g_\mathrm{inh}\}$. The synaptic time constants τ_exc and τ_inh are both fixed to 5 ms.

2. $\mathrm{HVC_I}$ dynamics

In the Long *et al.* [32] model, $\mathrm{HVC_I}$ neurons are modeled as single-compartment neurons containing leak, Na+, delay-rectified K+, and high-threshold K+ conductances. The membrane potential V obeys the dynamics

$$C_\mathrm{m} \dot{V} = I_\mathrm{L} + I_\mathrm{Na} + I_\mathrm{Kdr} + I_\mathrm{KHT} + I_\mathrm{exc} + I_\mathrm{inh}; \tag{B11}$$

the values of all parameters are given in Table II. The currents are given as

$$\begin{aligned} I_\mathrm{L} &= -G_\mathrm{L}(V - E_\mathrm{L}) \\ I_\mathrm{Na} &= -G_\mathrm{Na} m^3 h (V - E_\mathrm{Na}) \\ I_\mathrm{Kdr} &= -G_\mathrm{Kdr} n^4 (V - E_\mathrm{K}) \\ I_\mathrm{KHT} &= -G_\mathrm{KHT} w (V - E_\mathrm{K}) \\ I_\mathrm{exc} &= -g_\mathrm{exc}(t) V \\ I_\mathrm{inh} &= -g_\mathrm{inh}(t)(V - E_\mathrm{I}) \end{aligned} \tag{B12}$$

for total excitatory and inhibitory synaptic conductances g_exc and g_inh. The gating variables m, h, and n evolve according to the dynamics

$$\dot{x} = \alpha_x (1 - x) - \beta_x x, \tag{B13}$$

Name	Description	Value
C_{m}	membrane capacitance	$1\,\mu\mathrm{F/cm}^2$
G_{L}	leak conductance	$0.1\,\mathrm{mS/cm}^2$
G_{Na}	Na+ conductance	$100\,\mathrm{mS/cm}^2$
G_{Kdr}	delay-rectified K+ conductance	$20\,\mathrm{mS/cm}^2$
G_{KHT}	high-threshold K+ conductance	$500\,\mathrm{mS/cm}^2$
E_{L}	leak reversal potential	$-65\,\mathrm{mV}$
E_{Na}	Na+ reversal potential	$55\,\mathrm{mV}$
E_{K}	K+ reversal potential	$-80\,\mathrm{mV}$
E_{I}	inhibitory reversal potential	$-75\,\mathrm{mV}$

TABLE II. $\mathrm{HVC_I}$ model parameters

for $x \in \{m, h, n\}$, where

$$\begin{aligned}
\alpha_m &= (V+22)/(1-\exp[-(V+22)/10]) \\
\alpha_h &= 0.7\exp[-(V+34)/20] \\
\alpha_n &= 0.15(V+15)/(1-\exp[-(V+15)/10]) \\
\beta_m &= 40\exp[-(V+47)/18] \\
\beta_h &= 10/(1+\exp[-(V+4)/10]) \\
\beta_n &= 0.2\exp[-(V+25)/80],
\end{aligned} \tag{B14}$$

with implied units throughout. The gating variable obeys

$$\dot{w} = w_\infty - w, \tag{B15}$$

where

$$w_\infty = 1/(1+\exp[-V/5]). \tag{B16}$$

The excitatory and inhibitory conductances obey the same dynamics as for $\mathrm{HVC_{RA}}$ neurons, except for the fact that the excitatory time constant τ_{exc} is set to 2 ms.

3. Noise spike trains

The Long *et al.* [32] model introduces noise into the neurons via independent Poisson spike trains. Each $\mathrm{HVC_{RA}}$ neuron receives excitatory and inhibitory spike trains at both compartments, each generated from a homogeneous Poisson process with a rate of 100 Hz. The conductances of each spike are drawn independently in time from a uniform distribution on $[0, 0.035]\,\mathrm{mS/cm}^2$ for the somatic compartment and $[0, 0.045]\,\mathrm{mS/cm}^2$ for the dendritic compartment. Each $\mathrm{HVC_I}$ neuron also receives excitatory and inhibitory noise spike trains, generated from 250 Hz Poisson processes with conductances drawn uniformly from $[0, 0.45]\,\mathrm{mS/cm}^2$. With this noise model, the RMS fluctuation in the membrane voltage of each compartment of each $\mathrm{HVC_{RA}}$ neuron is about 3 mV, and the $\mathrm{HVC_I}$ neurons spike spontaneously at about 10 Hz.

[1] H. G. Wu, Y. R. Miyamoto, L. N. G. Castro, B. P. Ölveczky, and M. A. Smith, Temporal structure of motor variability is dynamically regulated and predicts motor learning ability, Nature neuroscience **17**, 312 (2014).

[2] A. K. Dhawale, M. A. Smith, and B. P. Ölveczky, The role of variability in motor learning, Annual review of neuroscience **40**, 479 (2017).

[3] A. K. Dhawale, Y. R. Miyamoto, M. A. Smith, and B. P. Ölveczky, Adaptive regulation of motor variability, Current Biology **29**, 3551 (2019).

[4] A. A. Faisal, L. P. Selen, and D. M. Wolpert, Noise in the nervous system, Nature reviews neuroscience **9**, 292 (2008).

[5] L. C. Osborne, S. G. Lisberger, and W. Bialek, A sensory source for motor variation, Nature **437**, 412 (2005).

[6] R. B. Stein, E. R. Gossen, and K. E. Jones, Neuronal variability: noise or part of the signal?, Nature Reviews Neuroscience **6**, 389 (2005).

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

[8] R. B. Ivry and R. M. Spencer, The neural representation of time, Current opinion in neurobiology **14**, 225 (2004).

[9] M. D. Mauk and D. V. Buonomano, The neural basis of temporal processing, Annu. Rev. Neurosci. **27**, 307 (2004).

[10] J. T. Coull, R.-K. Cheng, and W. H. Meck, Neuroanatomical and neurochemical substrates of timing, Neuropsychopharmacology **36**, 3 (2011).

[11] H. Merchant, D. L. Harrington, and W. H. Meck, Neural basis of the perception and estimation of time, Annual review of neuroscience **36**, 313 (2013).

[12] J. J. Paton and D. V. Buonomano, The neural basis of timing: distributed mechanisms for diverse functions, Neuron **98**, 687 (2018).

[13] R. H. Hahnloser, A. A. Kozhevnikov, and M. S. Fee, An ultra-sparse code underlies the generation of neural sequences in a songbird, Nature **419**, 65 (2002).

[14] M. A. Long and M. S. Fee, Using temperature to analyse temporal dynamics in the songbird motor pathway, Nature **456**, 189 (2008).

[15] F. Ali, T. M. Otchy, C. Pehlevan, A. L. Fantana, Y. Burak, and B. P. Ölveczky, The basal ganglia is necessary for learning spectral, but not temporal, features of birdsong, Neuron **80**, 494 (2013).

[16] C. Tang, D. Chehayeb, K. Srivastava, I. Nemenman, and S. J. Sober, Millisecond-scale motor encoding in a cortical vocal area, PLoS biology **12**, e1002018 (2014).

[17] G. B. Mello, S. Soares, and J. J. Paton, A scalable population code for time in the striatum, Current Biology **25**, 1113 (2015).

[18] K. H. Srivastava, C. M. Holmes, M. Vellema, A. R. Pack, C. P. Elemans, I. Nemenman, and S. J. Sober, Motor control by precisely timed spike patterns, Proceedings of the National Academy of Sciences **114**, 1171 (2017).

[19] E. D. Remington, D. Narain, E. A. Hosseini, and M. Jazayeri, Flexible sensorimotor computations through rapid reconfiguration of cortical dynamics, Neuron **98**, 1005 (2018).

[20] S. J. Sober, S. Sponberg, I. Nemenman, and L. H. Ting, Millisecond spike timing codes for motor control, Trends in neurosciences **41**, 644 (2018).

[21] J. A. White, J. T. Rubinstein, and A. R. Kay, Channel noise in neurons, Trends in neurosciences **23**, 131 (2000).

[22] W. H. Calvin and C. F. Stevens, Synaptic noise and other sources of randomness in motoneuron interspike intervals., Journal of neurophysiology **31**, 574 (1968).

[23] Z. F. Mainen and T. J. Sejnowski, Reliability of spike timing in neocortical neurons, Science **268**, 1503 (1995).

[24] C. Van Vreeswijk and H. Sompolinsky, Chaos in neuronal networks with balanced excitatory and inhibitory activity, Science **274**, 1724 (1996).

[25] R. Darshan, W. E. Wood, S. Peters, A. Leblois, and D. Hansel, A canonical neural mechanism for behavioral variability, Nature communications **8**, 15415 (2017).

[26] M. Abeles, Local circuits: an electrophysiological study (1982).

[27] M. Diesmann, M.-O. Gewaltig, and A. Aertsen, Stable propagation of synchronous spiking in cortical neural networks, Nature **402**, 529 (1999).

[28] N. F. Hardy and D. V. Buonomano, Neurocomputational models of interval and pattern timing, Current Opinion in Behavioral Sciences **8**, 250 (2016).

[29] M. Li and H. Greenside, Stable propagation of a burst through a one-dimensional homogeneous excitatory chain model of songbird nucleus hvc, Physical Review E **74**, 011918 (2006).

[30] D. Z. Jin, F. M. Ramazanoğlu, and H. S. Seung, Intrinsic bursting enhances the robustness of a neural network model of sequence generation by avian brain area hvc, Journal of computational neuroscience **23**, 283 (2007).

[31] D. Z. Jin, Generating variable birdsong syllable sequences with branching chain networks in avian premotor nucleus hvc, Physical Review E **80**, 051902 (2009).

[32] M. Long, D. Jin, and M. Fee, Support for a synaptic chain model of neuronal sequence generation, Nature **468**, 394 (2010).

[33] C. M. Glaze and T. W. Troyer, Temporal structure in zebra finch song: implications for motor coding, Journal of Neuroscience **26**, 991 (2006).

[34] C. M. Glaze and T. W. Troyer, Behavioral measurements of a temporally precise motor code for birdsong, Journal of Neuroscience **27**, 7631 (2007).

[35] C. M. Glaze and T. W. Troyer, A generative model for measuring latent timing structure in motor sequences, PLoS One **7**, e37616 (2012).

[36] N. Brunel, Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons, Journal of computational neuroscience **8**, 183 (2000).

[37] H. Tuckwell, *Introduction to theoretical neurobiology: nonlinear and stochastic theories*, Vol. 8 (Cambridge Univ Pr, 2005).

[38] F. Wan and H. Tuckwell, Neuronal firing and input variability, J. theor. Neurobiol **1**, 197 (1982).

[39] R. S. Zucker and W. G. Regehr, Short-term synaptic plasticity, Annual review of physiology **64**, 355 (2002).

[40] A. M. Wing and A. B. Kristofferson, Response delays and the timing of discrete motor responses, Perception & Psychophysics **14**, 5 (1973).

[41] N. A. Gershenfeld, *The nature of mathematical modeling* (Cambridge university press, 1999).

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

[43] J. Hertz, Theoretical aspects of neural computation. in (p. 135-144). wong ky m. king i. and yeung dy (1997).

[44] S. Goedeke and M. Diesmann, The mechanism of synchronization in feed-forward neuronal networks, New Journal of Physics **10**, 015007 (2008).

[45] M. S. Brainard and A. J. Doupe, Translating birdsong: songbirds as a model for basic and applied medical research, Annual review of neuroscience **36**, 489 (2013).

[46] T. M. Otchy, S. B. Wolff, J. Y. Rhee, C. Pehlevan, R. Kawai, A. Kempf, S. M. Gobes, and B. P. Ölveczky, Acute off-target effects of neural circuit manipulations, Nature **528**, 358 (2015).

[47] C. Pehlevan, F. Ali, and B. P. Ölveczky, Flexibility in motor timing constrains the topology and dynamics of pattern generator circuits, Nature communications **9**, 977 (2018).

[48] J. T. Sakata, C. M. Hampton, and M. S. Brainard, Social modulation of sequence and syllable variability in adult birdsong, Journal of neurophysiology **99**, 1700 (2008).

[49] M. M. Churchland, M. Y. Byron, J. P. Cunningham, L. P. Sugrue, M. R. Cohen, G. S. Corrado, W. T. Newsome, A. M. Clark, P. Hosseini, B. B. Scott, *et al.*, Stimulus onset quenches neural variability: a widespread cortical phenomenon, Nature neuroscience **13**, 369 (2010).

[50] K. H. Britten, W. T. Newsome, M. N. Shadlen, S. Celebrini, J. A. Movshon, *et al.*, A relationship between behavioral choice and the visual responses of neurons in macaque mt, Visual neuroscience **13**, 87 (1996).

[51] K. Hamaguchi, M. Tanaka, and R. Mooney, A distributed recurrent network contributes to temporally precise vocalizations, Neuron **91**, 680 (2016).

[52] B. P. Olveczky, A. S. Andalman, and M. S. Fee, Vocal experimentation in the juvenile songbird requires a basal ganglia circuit., PLoS biology **3**, e153 (2005).

[53] K. Hamaguchi and R. Mooney, Recurrent interactions between the input and output of a songbird cortico-basal ganglia pathway are implicated in vocal sequence variability, Journal of Neuroscience **32**, 11671 (2012).

[54] B. Lewandowski, A. Vyssotski, R. H. Hahnloser, and M. Schmidt, At the interface of the auditory and vocal motor systems: Nif and its role in vocal processing, production and learning, Journal of Physiology-Paris **107**, 178 (2013).

[55] J. T. Sakata and M. S. Brainard, Real-time contributions of auditory feedback to avian vocal motor control, Journal of Neuroscience **26**, 9619 (2006).

[56] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, Vol. 55 (US Government printing office, 1948).