Fast and Accurate Langevin Simulations of Stochastic Hodgkin-Huxley Dynamics

Shusen Pu
sxp600@case.edu
Department of Mathematics, Applied Mathematics, and Statistics, Case Western Reserve University, Cleveland, OH 44106, U.S.A.

Peter J. Thomas
pjthomas@case.edu
Department of Mathematics, Applied Mathematics, and Statistics; Biology; Cognitive Science; and Electrical, Computer, and Systems Engineering: Case Western Reserve University, Cleveland, OH 44106, U.S.A.

Fox and Lu introduced a Langevin framework for discrete-time stochastic models of randomly gated ion channels such as the Hodgkin-Huxley (HH) system. They derived a Fokker-Planck equation with state-dependent diffusion tensor D and suggested a Langevin formulation with noise coefficient matrix S such that $SS^\top = D$. Subsequently, several authors introduced a variety of Langevin equations for the HH system. In this article, we present a natural 14-dimensional dynamics for the HH system in which each directed edge in the ion channel state transition graph acts as an independent noise source, leading to a 14×28 noise coefficient matrix S. We show that (1) the corresponding 14D system of ordinary differential equations is consistent with the classical 4D representation of the HH system; (2) the 14D representation leads to a noise coefficient matrix S that can be obtained cheaply on each time step, without requiring a matrix decomposition; (3) sample trajectories of the 14D representation are pathwise equivalent to trajectories of Fox and Lu's system, as well as trajectories of several existing Langevin models; (4) our 14D representation (and those equivalent to it) gives the most accurate interspike interval distribution, not only with respect to moments but under both the L_1 and L_∞ metric-space norms; and (5) the 14D representation gives an approximation to exact Markov chain simulations that are as fast and as efficient as all equivalent models. Our approach goes beyond existing models, in that it supports a stochastic shielding decomposition that dramatically simplifies S with minimal loss of accuracy under both voltage- and current-clamp conditions.

1 Introduction

Many natural phenomena exhibit stochastic fluctuations arising at the molecular scale, the effects of which impact macroscopic quantities. Understanding when and how microscale fluctuations will significantly contribute to macroscale behavior is a fundamental problem spanning the sciences. The impact of random ion channel fluctuations on the timing of action potentials in nerve cells provides an important example. Channel noise can have a significant effect on spike generation (Mainen & Sejnowski, 1995; Schneidman, Freedman, & Segev, 1998), propagation along axons (Faisal & Laughlin, 2007), and spontaneous (ectopic) action potential generation in the absence of stimulation (O'Donnell & van Rossum, 2015). At the network level, channel noise can drive the endogenous variability of vital rhythms such as respiratory activity (Yu, Dhingra, Dick, & Galán, 2017).

Hodgkin and Huxley's quantitative model for active sodium and potassium currents producing action potential generation in the giant axon of *Loligo* suggested an underlying system of gating variables consistent with a multistate Markov process description (Hill & Chen, 1972). The discrete nature of individual ion channel conductances was confirmed experimentally (Neher & Sakmann, 1976). Subsequently, numerical studies of high-dimensional discrete-state, continuous-time Markov chain models produced insights into the effects of fluctuations in discrete ion channel populations on action potentials (Skaugen & Walløe, 1979; Strassberg & De-Felice, 1993), also known as *channel noise* (White, Klink, Alonso, & Kay, 1998; White, Rubinstein, & Kay, 2000).

In the standard molecular-level HH model, which we adopt here, the K^+ channel comprises four identical n gates that open and close independently, giving a five-vertex channel-state diagram with eight directed edges; the channel conducts a current only when in the rightmost state (see Figure 1, top). The Na^+ channel comprises three identical m gates and a single h gate, all independent, giving an eight-vertex diagram with 20 directed edges, of which one is conducting (see Figure 1, bottom).

Discrete-state channel noise models are numerically intensive, whether implemented using discrete-time binomial approximations to the underlying continuous-time Markov process (Skaugen & Walløe, 1979; Schmandt & Galán, 2012) or continuous-time hybrid Markov models with exponentially distributed state transitions and continuously varying membrane potential. The latter were introduced by Clay and DeFelice (1983) and are in principle exact (Anderson, Ermentrout, & Thomas, 2015). Under voltage-clamp conditions, the hybrid conductance-based model reduces to a time-homogeneous Markov chain (Colquhoun & Hawkes, 1981) that can be simulated using standard methods such as Gillespie's exact algorithm (Gillespie, 1977, 2007). Even with this simplification, such Markov chain (MC) algorithms are numerically expensive to simulate with realistic
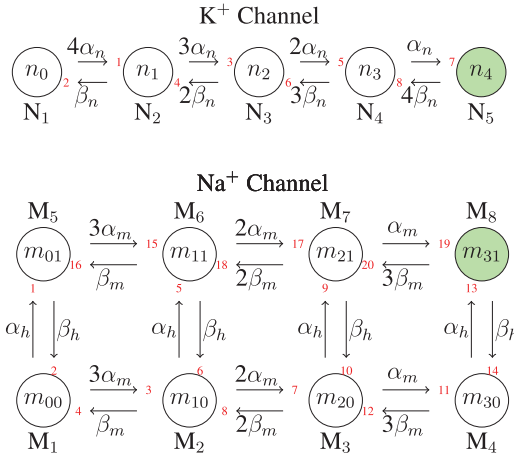
$$K^+ \text{ Channel}$$

$$Na^+ \text{ Channel}$$

Figure 1: Molecular potassium (K^+) and sodium (Na^+) channel states for the Hodgkin-Huxley model. Filled circles mark conducting states n_4 and m_{31}. Per capita transition rates for each directed edge (α_n, β_n, α_m, β_m, α_h and β_h) are voltage dependent (cf. equations B.1 and B.6). Directed edges are numbered 1 to 8 (K^+ channel) and 1 to 20 (Na^+-channel), marked in small red numerals.

population sizes of thousands of channels or greater. Therefore, there is an ongoing need for efficient and accurate approximation methods.

Following Clay and DeFelice's exposition of continuous time Markov chain implementations, Fox and Lu (1994) introduced a Fokker-Planck equation (FPE) framework that captured the first- and second-order statistics of HH ion channel populations in a 14-dimensional representation. Taking into account conservation of probability, one needs four variables to represent the population of K^+ channels, seven for Na^+, and one for voltage, leading to a 12-dimensional state-space description. The resulting high-dimensional partial differential equation is impractical to solve numerically. However, as Fox and Lu observed, "To every Fokker-Planck description, there is associated a Langevin description" (Fox & Lu, 1994). They therefore introduced a Langevin stochastic differential equation:

$$C\frac{dV}{dt} = I_{app}(t) - \bar{g}_{Na}\mathbf{M}_8(V - V_{Na}) - \bar{g}_K\mathbf{N}_5(V - V_K) - g_{leak}(V - V_{leak}),$$

(1.1)

$$\frac{d\mathbf{M}}{dt} = A_{Na}\mathbf{M} + S_1\xi_1,$$

(1.2)

$$\frac{d\mathbf{N}}{dt} = A_K\mathbf{N} + S_2\xi_2,$$

(1.3)

where C is the capacitance, I_{app} is the applied current, maximal conductances are denoted $\bar{g}_{ion}$, with V_{ion} being the associated reversal potential, and ohmic leak current $g_{leak}(V - V_{leak})$. $\mathbf{M} \in \mathbb{R}^8$ and $\mathbf{N} \in \mathbb{R}^5$ are vectors for the fractions of Na$^+$ and K$^+$ channels in each state, with $\mathbf{M}_8$ representing the open channel fraction for Na$^+$, and $\mathbf{N}_5$ the open channel fraction for K$^+$ (see Figure 1). Vectors $\xi_1(t) \in \mathbb{R}^8$ and $\xi_2(t) \in \mathbb{R}^5$ are independent gaussian white noise processes with zero mean and unit variances $\langle \xi_1(t)\xi_1^\top(t') \rangle = I_8\,\delta(t - t')$ and $\langle \xi_2(t)\xi_2^\top(t') \rangle = I_5\,\delta(t - t')$. The state-dependent rate matrices A_{Na} and A_K are given in equations 2.10 and 2.11. In Fox and Lu's formulation, S must satisfy $S = \sqrt{D}$, where D is a symmetric, positive semidefinite $k \times k$ "diffusion matrix" (see appendix D for the D matrices for the standard HH K$^+$ and Na$^+$ channels). We will refer to the 14-dimensional Langevin equations 1.1 to 1.3, with $S = \sqrt{D}$, as the Fox-Lu model.

The original Fox-Lu model, later called the conductance noise model by Goldwyn and Shea-Brown (2011), did not see widespread use until gains in computing speed made the square root calculations more feasible. Seeking a more efficient approximation, Fox and Lu (1994) also introduced a four-dimensional Langevin version of the HH model. This model was systematically studied in Fox (1997), which can be written as

$$C\frac{dV}{dt} = I_{app}(t) - \bar{g}_{Na}m^3h\,(V - V_{Na}) - \bar{g}_K n^4\,(V - V_K) - g_{leak}(V - V_{leak}),$$

(1.4)

$$\frac{dx}{dt} = \alpha_x(1 - x) - \beta_x x + \xi_x(t), \quad \text{where } x = m, h, \text{ or, } n,$$

(1.5)

where $\xi_x(t)$ are gaussian processes with covariance function

$$E[\xi_x(t), \xi_x(t')] = \frac{\alpha_x(1 - x) + \beta_x x}{N}\delta(t - t').$$

(1.6)

Here N represents the total number of Na$^+$channels (respectively, the total number of K$^+$channels), and $\delta(\cdot)$ is the Dirac delta function. This model, referred to as the "subunit noise model" by Goldwyn and Shea-Brown (2011), has been widely used as an approximation to MC ion channel models (see the references in Bruce, 2009; Goldwyn & Shea-Brown, 2011). For example, Schmid et al. (2001) used this approximation to investigate stochastic resonance and coherence resonance in forced and unforced versions of the HH model (e.g. in the excitable regime). However, the numerical accuracy of this method was criticized in several studies (Mino, Rubinstein, & White, 2002; Bruce, 2009), which found that its accuracy does not improve even with increasing numbers of channels.

Although more accurate approximations based on Gillespie's algorithm (using a piecewise constant propensity approximation: Bruce, 2009; Mino et al., 2002) and even based on exact simulations (Clay & DeFelice, 1983; Newby, Bressloff, & Keener, 2013; Anderson et al., 2015) became available, they remained prohibitively expensive for large network simulations. Meanwhile, Goldwyn and Shea-Brown's rediscovery of Fox and Lu's earlier conductance-based model (Goldwyn & Shea-Brown, 2011; Goldwyn, Shea-Brown, & Rubinstein, 2010) launched a flurry of activity seeking the best Langevin-type approximation. Goldwyn and Shea-Brown (2011) introduced a faster decomposition algorithm to simulate equations 1.1 to 1.3 and showed that Fox and Lu's (1994) method accurately captured the fractions of open channels and the interspike interval (ISI) statistics, in comparison with Gillespie-type Monte Carlo (MC) simulations. However, despite the development of efficient singular value decomposition based algorithms for solving $S = \sqrt{D}$, this step still causes a bottleneck in the algorithms based on Fox and Lu (1994), Goldwyn and Shea-Brown (2011), and Goldwyn, Imennov, Famulare, and Shea-Brown (2011).

Many variations on Fox and Lu's (1994) Langevin model have been proposed in recent years (Dangerfield, Kay, & Burrage, 2010, 2012; Linaro, Storace, & Giugliano, 2011; Orio & Soudry, 2012; Güler, 2013b; Huang, Rüdiger, & Shuai, 2013, 2015; Pezo, Soudry, & Orio, 2014; Fox, 2018), including Goldwyn et al.'s work (Goldwyn & Shea-Brown, 2011; Goldwyn et al., 2011), each with its own strengths and weaknesses. One class of methods imposes projected boundary conditions (Dangerfield et al., 2010, 2012); as we will show in section 5, this approach leads to inaccurate interspike interval distribution and is inconsistent with a natural multinomial invariant manifold structure for the ion channels. Several methods implement correlated noise at the subunit level, as in equations 1.5 and 1.6 (Fox, 1997; Linaro et al., 2011; Güler, 2013a, 2013b). However, if one recognizes that at the molecular level, the individual directed edges represent the independent noise sources in ion channel dynamics, then the approach incorporating noise at the subunit level obscures the biophysical origin of ion channel fluctuations. Some methods introduce the noisy dynamics at the level of edges rather than nodes but lump reciprocal edges together into pairs (Orio & Soudry, 2012; Dangerfield et al., 2012; Huang et al., 2013; Pezo et al., 2014). This approach implicitly assumes, in effect, that the ion channel probability distribution satisfies a detailed balance (or microscropic reversibility) condition. However, while detailed balance holds for the HH model under stationary voltage clamp, this condition is violated during active spiking. Finally, the stochastic shielding approximation (Schmandt & Galán, 2012; Schmidt & Thomas, 2014; Schmidt et al., 2018) does not have a natural formulation in the representation associated with an $n \times n$ noise coefficient matrix S; in the cases of rectangular S matrices used in Orio and Soudry (2012) and Dangerfield et al. (2012) stochastic shielding can only be applied to reciprocal pairs of edges. We elaborate on these points in section 6.

In this article, we introduce a new variation of Fox and Lu's conductance-noise model that avoids the limitations we have described. We show that preserving each directed edge in the channel transition graph (see Figure 1) as an independent noise source leads to a natural, biophysically motivated Langevin model that does not require any matrix decomposition step. Our construction lends itself to direct application of stochastic shielding methods, leading to faster simulations that retain the accuracy of Fox and Lu's method.

As an additional benefit, our method answers an open question in the literature, arising from the fact that the decomposition $D = SS^\mathsf{T}$ is not unique. As Fox (2018) recently pointed out, subblock determinants of the D matrices play a major role in the structure of the S matrix elements. Fox conjectured that "a universal form for S may exist." In this article, we obtain the universal form for the noise coefficient matrix S. Moreover, we prove that our model is equivalent to Fox and Lu's 1994 model in the strong sense of pathwise equivalence.

The remainder of the article is organized as follows. In section 2, we review the canonical deterministic 14D version of the HH model. We prove a series of lemmas that show (1) the multinomial submanifold $\mathcal{M}$ is an invariant manifold within the 14D space and (2) the velocity on the 14D space and the pushforward of the velocity on the 4D space are identical. Moreover, we show (numerically) that (3) the submanifold $\mathcal{M}$ is globally attracting, even under current clamp conditions. Figure 2 illustrates the relationship between the 4D and 14D deterministic HH models. Section 3 lays out our 14×28 Langevin HH model. Like Orio and Soudry (2012), Dangerfield et al. (2012), and Pezo et al. (2014), we avoid matrix decomposition by computing S directly. The key difference between our approach and its closest relative (Pezo et al., 2014) is to use a rectangular $n \times k$ matrix S for which each directed edge is treated as an independent noise source rather than lumping reciprocal edges together in pairs. In the new Langevin model, the form of our S matrix reflects the biophysical origins of the underlying channel noise and allows us to apply the stochastic shielding approximation by neglecting the noise on selected individual directed edges. As we prove in section 4, our model (without the stochastic shielding approximation) is pathwise equivalent to all those in a particular class of biophysically derived Langevin models, including those used in Fox and Lu (1994), Goldwyn et al. (2011), Goldwyn and Shea-Brown (2011), Orio and Soudry (2012), Pezo et al. (2014), and Fox (2018). In addition to 4D and 14D deterministic trajectories, Figure 2 shows a stochastic trajectory generated by our Langevin model. Finally, we compare our Langevin model to several alternative stochastic neural models in terms of accuracy (of the full ISI distribution) and numerical efficiency in section 5.

Matlab code to generate the figures throughout the article is available on github from https://github.com/shusenpu/Stochastic_shielding and on ModelDB at accession number 266551.
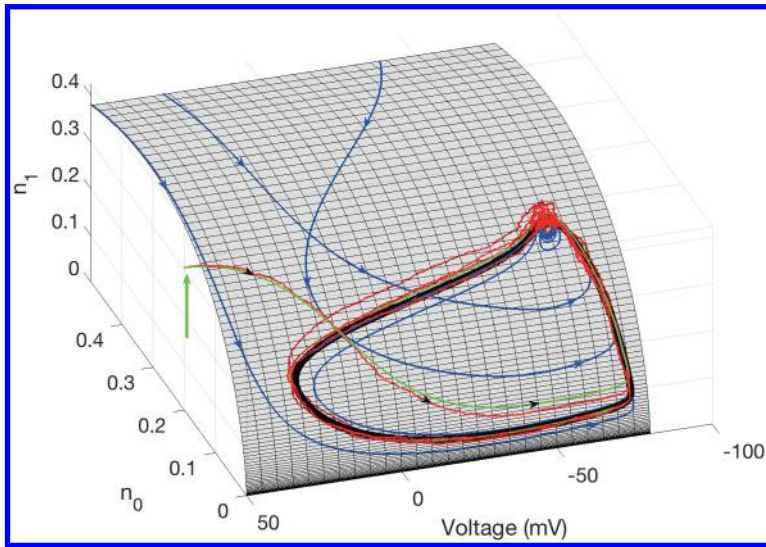
Figure 2: 4D and 14D HH models. The meshed surface is a three-dimensional projection of the 14D state space onto three axes representing the voltage, v, the probability of all four potassium gates being in the closed state, n_0, and the probability of exactly one potassium gate being in the open state, n_1. **Blue curves:** Trajectories of the deterministic 14D HH model with initial conditions located on the 4D multinomial invariant submanifold, $\mathcal{M}$. We prove that $\mathcal{M}$ is an invariant submanifold in section 2.3. **Black curve:** The deterministic limit cycle solution for the 14D HH model, which forms a closed loop within $\mathcal{M}$. **Green curve:** A trajectory of the deterministic 14D HH model with initial conditions (vertical green arrow) off the multinomial submanifold. **Red curve:** A trajectory of the stochastic 14D HH model (see section 3) with the same initial conditions as the green trajectory. The blue and black arrows mark the directions of the trajectories. Note that trajectories starting away from $\mathcal{M}$ converge to $\mathcal{M}$, and all deterministic trajectories converge to the deterministic limit cycle. Parameters of the simulation are given in Table 5.

2 The Deterministic 4D and 14D HH Models

In this section, we review the classical four-dimensional model of Hodgkin and Huxley (1952) HH), as well as its natural 14-dimensional version (Dayan & Abbott, 2001, sec. 5.7), with variables comprising membrane voltage and the occupancies of five potassium channel states and eight sodium channel states. The deterministic 14D model is the mean field of the channel-based Langevin model proposed by Fox and Lu (1994); this article describes both the Langevin and the mean field versions of the 14D Hodgkin-Huxley system. For completeness of exposition, we briefly review

the 4D deterministic HH system and its 14D deterministic counterpart. In section 4, we prove that the sample paths of a class of Langevin stochastic HH models are equivalent; in section 2.3, we review analogous results relating trajectories of the 4D and 14D deterministic ODE systems.

In particular, we show that the deterministic 14D model and the original 4D HH model are dynamically equivalent, in the sense that every flow (solution) of the 4D model corresponds to a flow of the 14D model. The consistency of trajectories between the 14D and 4D models is easy to verify for initial data on a 4D submanifold of the 14D space given by choosing multinomial distributions for the gating variables (Dayan & Abbott, 2001; Goldwyn et al., 2011). Similarly, Keener established results on multinomial distributions as invariant submanifolds of Markov models with ion channel kinetics under several circumstances (Keener, 2009, 2010; Earnshaw & Keener, 2010a, 2010b), but without treating the general current-clamped case. Consistent with these results, we show that the set of all 4D flows maps to an invariant submanifold of the state space of the 14D model. Moreover, we show numerically that solutions of the 14D model with arbitrary initial conditions converge to this submanifold. Thus the original HH model "lives inside" the 14D deterministic model in the sense that the former embeds naturally and consistently within the latter (see Figure 2).

In the stochastic case, the 14D model has a natural interpretation as a hybrid stochastic system with independent noise forcing along each edge of the potassium (8 directed edges) and sodium (20 directed edges) channel state transition graphs. The hybrid model leads naturally to a biophysically grounded Langevin model that we describe in section 3. In contrast to the ODE case, the stochastic versions of the 4D and 14D models are *not* equivalent (Goldwyn & Shea-Brown, 2011).

2.1 The 4D Hodgkin-Huxley Model. The 4D voltage-gated ion channel HH model is a set of four ordinary differential equations:

$$C\frac{dv}{dt} = -\bar{g}_{Na}m^3h(v - V_{Na}) - \bar{g}_K n^4(v - V_K) - g_L(v - V_L) + I_{app}, \quad (2.1)$$

$$\frac{dm}{dt} = \alpha_m(v)(1 - m) - \beta_m(v)m, \quad (2.2)$$

$$\frac{dh}{dt} = \alpha_h(v)(1 - h) - \beta_h(v)h, \quad (2.3)$$

$$\frac{dn}{dt} = \alpha_n(v)(1 - n) - \beta_n(v)n, \quad (2.4)$$

where v is the membrane potential, I_{app} is the applied current, and $0 \leq m, n, h \leq 1$ are dimensionless gating variables associated with Na^+ and K^+ channels. The constant $\bar{g}_{ion}$ is the maximal value of the conductance for the sodium and potassium channel, respectively. Parameters V_{ion} and C

are the ionic reversal potentials and capacitance, respectively. The quantities α_x and β_x, $x \in \{m, n, h\}$ are the voltage-dependent per capita transition rates, defined in appendix B.

This system is a C^∞ vector field on a four-dimensional manifold (with boundary) contained in $\mathbb{R}^4$: $\mathcal{X} = \{-\infty < v < \infty, 0 \leq m, h, n \leq 1\} = \mathbb{R} \times [0, 1]^3$. The manifold is forward and backward invariant in time. If I_{app} is constant then $\mathcal{X}$ has an invariant subset given by $\mathcal{X} \cap \{v_{\min} \leq v \leq v_{\max}\}$, where $v_{\min}$ and $v_{\max}$ are calculated in lemma 1.

As pointed Keener and Sneyd (1998) and Keener (2009) pointed out, for voltage either fixed or given as a prescribed function of time, the equations for m, h, and n can be interpreted as the parameterization of an invariant manifold embedded in a higher-dimensional time-varying Markov system. Several papers developed this idea for a variety of ion channel models and related systems (Keener, 2009; Earnshaw & Keener, 2010b), but the theory developed is restricted to the voltage-clamped case.

Under a fixed voltage clamp, the ion channels form a time-homogeneous Markov process with a unique (voltage-dependent) stationary probability distribution. Under a time-varying current clamp, the ion channels nevertheless form a Markov process, albeit no longer time homogeneous. Under these conditions, the ion channel state converges rapidly to a multinomial distribution indexed by a low-dimensional set of time-varying parameters $(m(t), h(t), n(t))$ (Keener, 2010). In the current-clamped case, the ion channel process, considered alone, is neither stationary nor Markovian, making the analysis of this case significantly more challenging from a mathematical point of view.

2.2 The Deterministic 14D Hodgkin-Huxley Model.

For the HH kinetics given in Figure 1, we define the eight-component state vector $\mathbf{M}$ for the Na^+ gates and the five-component state vector $\mathbf{N}$ for the K^+ gates, respectively, as

$$\mathbf{M} = [m_{00}, m_{10}, m_{20}, m_{30}, m_{01}, m_{11}, m_{21}, m_{31}]^\mathsf{T} \in [0, 1]^8. \tag{2.5}$$

$$\mathbf{N} = [n_0, n_1, n_2, n_3, n_4]^\mathsf{T} \in [0, 1]^5, \tag{2.6}$$

where $\sum_{i=0}^{3} \sum_{j=0}^{1} m_{ij} = 1$ and $\sum_{i=0}^{4} n_i = 1$. The open probability for the Na^+ channel is $\mathbf{M}_8 = m_{31}$ and is $\mathbf{N}_5 = n_4$ for the K^+ channel. The deterministic 14D HH equations may be written (compare equations 2.1 to 2.4)

$$C\frac{dV}{dt} = -\bar{g}_{\mathrm{Na}}\mathbf{M}_8(V - V_{\mathrm{Na}}) - \bar{g}_{\mathrm{K}}\mathbf{N}_5(V - V_{\mathrm{K}}) - g_{\mathrm{L}}(V - V_{\mathrm{L}}) + I_{\mathrm{app}}, \tag{2.7}$$

$$\frac{d\mathbf{M}}{dt} = A_{\mathrm{Na}}(V)\mathbf{M}, \tag{2.8}$$

$$\frac{d\mathbf{N}}{dt} = A_{\mathrm{K}}(V)\mathbf{N}, \tag{2.9}$$

where the voltage-dependent drift matrices A_{Na} and A_K are given by

$$A_{Na}(V) =$$

$$\begin{bmatrix} A_{Na}(1) & \beta_m & 0 & 0 & \beta_h & 0 & 0 & 0 \\ 3\alpha_m & A_{Na}(2) & 2\beta_m & 0 & 0 & \beta_h & 0 & 0 \\ 0 & 2\alpha_m & A_{Na}(3) & 3\beta_m & 0 & 0 & \beta_h & 0 \\ 0 & 0 & \alpha_m & A_{Na}(4) & 0 & 0 & 0 & \beta_h \\ \alpha_h & 0 & 0 & 0 & A_{Na}(5) & \beta_m & 0 & 0 \\ 0 & \alpha_h & 0 & 0 & 3\alpha_m & A_{Na}(6) & 2\beta_m & 0 \\ 0 & 0 & \alpha_h & 0 & 0 & 2\alpha_m & A_{Na}(7) & 3\beta_m \\ 0 & 0 & 0 & \alpha_h & 0 & 0 & \alpha_m & A_{Na}(8) \end{bmatrix}, \quad (2.10)$$

$$A_K(V) = \begin{bmatrix} A_K(1) & \beta_n(V) & 0 & 0 & 0 \\ 4\alpha_n(V) & A_K(2) & 2\beta_n(V) & 0 & 0 \\ 0 & 3\alpha_n(V) & A_K(3) & 3\beta_n(V) & 0 \\ 0 & 0 & 2\alpha_n(V) & A_K(4) & 4\beta_n(V) \\ 0 & 0 & 0 & \alpha_n(V) & A_K(5) \end{bmatrix}, \quad (2.11)$$

and the diagonal elements

$$A_{ion}(i) = -\sum_{j\,:\,j\neq i} A_{ion}(j, i), \text{ for ion } \in \{Na, K\}.$$

2.3 Relation Between the 14D and 4D Deterministic HH Models.
Earnshaw and Keener (2010b) suggest that it is reasonable to expect that the
global flow of the 14D system should converge to the 4D submanifold but
also that it is far from obvious that it must. Existing theory applies to the
voltage-clamped case (Keener, 2009; Earnshaw & Keener, 2010b). Here, we
consider instead the current-clamped case, in which the fluctuations of the
ion channel state influences the voltage evolution, and vice versa.

In the remainder of this section, we (1) define a multinomial submani-
fold $\mathcal{M}$ and show that it is an invariant manifold within the 14D space, and
(2) we show that the velocity on the 14D space and the pushforward of the
velocity on the 4D space are identical. In section 2.4, we (3) provide numer-
ical evidence that $\mathcal{M}$ is globally attracting within the higher-dimensional
space.

In order to compare the trajectories of the 14D HH equations with tra-
jectories of the standard 4D equations, we define lower-dimensional and
higher-dimensional domains $\mathcal{X}$ and $\mathcal{Y}$, respectively, as

$$\mathcal{X} = \{-\infty < v < \infty, 0 \leq m \leq 1, 0 \leq h \leq 1, 0 \leq n \leq 1\} = \mathbb{R} \times [0, 1]^3 \subset \mathbb{R}^4,$$

Table 1: R: Map from the 14D HH Model $(m_{00}, \ldots, m_{31}, n_0, \ldots, n_4)$ to the 4D HH Model (m, h, n).

14D Model	4D Model
$(v, m_{00}, \ldots, m_{31}, n_0, \ldots, n_4)$	(v, m, h, n)
v	v
$\frac{1}{3}(m_{11} + m_{10}) + \frac{2}{3}(m_{21} + m_{20}) + m_{31} + m_{30}$	m
$m_{01} + m_{11} + m_{21} + m_{31}$	h
$n_1/4 + n_2/2 + 3n_3/4 + n_4$	n'

Note: Note that both $\{m_{00}, \ldots, m_{31}\}$ and $\{n_0, \ldots, n_4\}$ follow multinomial distributions.

Table 2: H: Map from the 4D HH Model (m, h, n) and the 14D HH Model $(m_{00}, \ldots, m_{31}, n_0, \ldots, n_4)$.

4D Model	14D Model
(v, m, h, n)	$(v, m_{00}, \ldots, m_{31}, n_0, \ldots, n_4)$
v	v
$(1 - n)^4$	n_0
$4(1 - n)^3 n$	n_1
$6(1 - n)^2 n^2$	n_2
$4(1 - n)n^3$	n_3
n^4	n_4
$(1 - m)^3(1 - h)$	m_{00}
$3(1 - m)^2 m(1 - h)$	m_{10}
$3(1 - m)m^2(1 - h)$	m_{20}
$m^3(1 - h)$	m_{30}
$(1 - m)^3 h$	m_{01}
$3(1 - m)^2 mh$	m_{11}
$3(1 - m)m^2 h$	m_{21}
$m^3 h$	m_{31}

$$\mathcal{Y} = \{-\infty < v < \infty\} \cap \left\{ 0 \le m_{ij}, \sum_{i=0}^{3} \sum_{j=0}^{1} m_{ij} = 1 \right\} \cap \left\{ 0 \le n_i, \sum_{i=0}^{4} n_i = 1 \right\}$$

$$= \mathbb{R} \times \Delta^7 \times \Delta^4 \subset \mathbb{R}^{14}, \tag{2.12}$$

where Δ^k is the k-dimensional simplex in $\mathbb{R}^{k+1}$ given by $y_1 + \ldots + y_{k+1} = 1$, $y_i \ge 0$. The 4D HH model $\frac{dx}{dt} = F(x)$, equations 2.1 to 2.4, is defined for $x \in \mathcal{X}$, and the 14D HH model $\frac{dy}{dt} = G(y)$, equations 2.7 and 2.9, is defined for $y \in \mathcal{Y}$. We introduce a dimension-reducing mapping $R : \mathcal{Y} \to \mathcal{X}$ as in Table 1 and a mapping from lower to higher dimension, $H : \mathcal{X} \to \mathcal{Y}$, as in Table 2. We construct R and H in such a way that $R \circ H$ acts as the identity on $\mathcal{X}$, that is, for all $x \in \mathcal{X}$, $x = R(H(x))$. The maps H and R are consistent with

a multinomial structure for the ion channel state distribution, in the following sense. The space $\mathcal{Y}$ covers all possible probability distributions on the eight sodium channel states and the five potassium channel states. Those distributions, which are products of one multinomial distribution on the K[+]-channel[1] and a second multinomial distribution on the Na[+]-channel,[2] form a submanifold of $\Delta^7 \times \Delta^4$. In this way we define a submanifold, denoted $\mathcal{M} = H(\mathcal{X})$, the image of $\mathcal{X}$ under H.

Before showing that the multinomial submanifold $\mathcal{M}$ is an invariant manifold within the 14D space, we first show that the deterministic 14D HH model is defined on a bounded domain. Having a bounded forward-invariant manifold is a general property of conductance-based models, which may be written in the form

$$\frac{dV}{dt} = f(V, \mathcal{N}_{\text{open}})$$

$$= \frac{1}{C} \left\{ I_{\text{app}} - g_{\text{leak}}(V - V_{\text{leak}}) - \sum_{i \in \mathcal{I}} \left[g_i N_{\text{open}}^i (V - V_i) \right] \right\} \quad (2.13)$$

$$\frac{d\mathcal{N}}{dt} = A(V)\mathcal{N} \text{ and} \quad (2.14)$$

$$\mathcal{N}_{\text{open}} = \mathcal{O}[\mathcal{N}]. \quad (2.15)$$

Here, C is the membrane capacitance, I_{app} is an applied current with upper and lower bounds $I_{\pm}$, respectively, and g_i is the conductance for the ith ion channel. The index i runs over the set of distinct ion channel types in the model, $\mathcal{I}$. The gating vector $\mathcal{N}$ represents the fractions of each ion channel population in various ion channel states, and the operator $\mathcal{O}$ gives the fraction of each ion channel population in the open (or conducting) channel states. The following lemma establishes that any conductance-based model (including the 4D or 14D HH model) is defined on a bounded domain.

Lemma 1. *For a conductance-based model of the form 2.13 to 2.15 and for any bounded applied current $I_- \leq I_{\text{app}} \leq I_+$, there exist upper and lower bounds V_{max} and V_{min} such that trajectories with initial voltage condition $V \in [V_{\text{min}}, V_{\text{max}}]$ remain within this interval for all times $t > 0$, regardless of the initial channel state.*

Proof. Let $V_1 = \min_{i \in \mathcal{I}} \{V_i\} \wedge V_{\text{leak}}$, and $V_2 = \max_{i \in \mathcal{I}} \{V_i\} \vee V_{\text{leak}}$, where the index i runs over $\mathcal{I}$, the set of distinct ion channel types. Note that for all i, $0 \leq$

---

[1] That is, distributions indexed by a single open probability n, with the five states having probabilities $\binom{4}{i} n^i (1-n)^{4-i}$ for $0 \leq i \leq 4$.

[2] That is, distributions indexed by two open probabilities m and h, with the eight states having probabilities $\binom{3}{i} m^i (1-m)^{3-i} h^j (1-h)^{1-j}$, for $0 \leq i \leq 3$, and $0 \leq j \leq 1$.

$N_{\text{open}}^i \leq 1$, and $g_i > 0$, $g_{\text{leak}} > 0$. Therefore, when $V \leq V_1$,

$$\frac{dV}{dt} = \frac{1}{C} \left\{ I_{\text{app}} - g_{\text{leak}}(V - V_{\text{leak}}) - \sum_{i \in \mathcal{I}} \left[g_i N_{\text{open}}^i (V - V_i) \right] \right\} \tag{2.16}$$

$$\geq \frac{1}{C} \left\{ I_{\text{app}} - g_{\text{leak}}(V - V_1) - \sum_{i \in \mathcal{I}} \left[g_i N_{\text{open}}^i (V - V_1) \right] \right\} \tag{2.17}$$

$$\geq \frac{1}{C} \left\{ I_{\text{app}} - g_{\text{leak}}(V - V_1) - \sum_{i \in \mathcal{I}} \left[g_i \times 0 \times (V - V_1) \right] \right\} \tag{2.18}$$

$$= \frac{1}{C} \left\{ I_{\text{app}} - g_{\text{leak}}(V - V_1) \right\}. \tag{2.19}$$

Inequality 2.17 follows because $V_1 = \min_{i \in \mathcal{I}}\{V_i\} \wedge V_{\text{leak}}$, and inequality 2.18 follows because $V - V_1 \leq 0$, $g_i > 0$ and $N_{\text{open}}^i \geq 0$. Let $V_{\text{min}} := \min\left\{ \frac{I_-}{g_{\text{leak}}} + V_1, V_1 \right\}$. When $V < V_{\text{min}}$, $\frac{dV}{dt} > 0$. Therefore, V will not decrease beyond V_{min}. Similarly, when $V \geq V_2$,

$$\frac{dV}{dt} = \frac{1}{C} \left\{ I_{\text{app}} - g_{\text{leak}}(V - V_{\text{leak}}) - \sum_{i \in \mathcal{I}} \left[g_i N_{\text{open}}^i (V - V_i) \right] \right\} \tag{2.20}$$

$$\leq \frac{1}{C} \left\{ I_{\text{app}} - g_{\text{leak}}(V - V_2) - \sum_{i \in \mathcal{I}} \left[g_i N_{\text{open}}^i (V - V_2) \right] \right\} \tag{2.21}$$

$$\leq \frac{1}{C} \left\{ I_{\text{app}} - g_{\text{leak}}(V - V_2) - \sum_{i \in \mathcal{I}} \left[g_i \times 0 \times (V - V_2) \right] \right\} \tag{2.22}$$

$$= \frac{1}{C} \left\{ I_{\text{app}} - g_{\text{leak}}(V - V_2) \right\}. \tag{2.23}$$

Inequality 2.21 holds because $V_2 = \max_{i \in \mathcal{I}}\{V_i\} \vee V_{\text{leak}}$, and inequality 2.22 holds because $V - V_2 \geq 0$, $g_i > 0$ and $N_{\text{open}}^i \geq 0$. Let $V_{\text{max}} = \max\left\{ \frac{I_{\text{app}}}{g_{\text{leak}}} + V_2, V_2 \right\}$. When $V > V_{\text{max}}$, $\frac{dV}{dt} < 0$. Therefore, V will not go beyond V_{max}.

We conclude that if V takes an initial condition in the interval $[V_{\text{min}}, V_{\text{max}}]$, then $V(t)$ remains within this interval for all $t \geq 0$. □

Given that $y \in \mathcal{Y}$ has a bounded domain, lemma 2 follows directly and establishes that the multinomial submanifold $\mathcal{M}$ is a forward-time–invariant manifold within the 14D space. The proof of lemma 2 is in appendix C.

Lemma 2. *Let $\mathcal{X}$ and $\mathcal{Y}$ be the lower-dimensional and higher-dimensional Hodgkin-Huxley manifolds given by equation 2.12, and let F and G be the vector fields on $\mathcal{X}$ and $\mathcal{Y}$ defined by equations 2.1 to 2.4 and 2.7 to 2.9, respectively. Let $H : \mathcal{X} \rightarrow \mathcal{M} \subset \mathcal{Y}$ and $R : \mathcal{Y} \rightarrow \mathcal{X}$ be the mappings given in Tables 2 and 1, respectively, and define the multinomial submanifold $\mathcal{M} = H(\mathcal{X})$. Then $\mathcal{M}$ is forward-time-invariant under the flow generated by G. Moreover, the vector field G, when restricted to $\mathcal{M}$, coincides with the vector field induced by F and the map H. That is, for all $y \in \mathcal{M}$, $G(y) = D_x H(R(y)) \cdot F(R(y))$.*

Lemma 2 establishes that the 14D HH model given by equations 2.7 to 2.9 is dynamically consistent with the original 4D HH model given by equations 2.1 to 2.4.

In section 2.4 we provide numerical evidence that the flow induced by G on $\mathcal{Y}$ converges to $\mathcal{M}$ exponentially fast. Thus, an initial probability distribution over the ion channel states that is not multinomial quickly approaches a multinomial distribution with dynamics induced by the 4D HH equations. Similar results, restricted to the voltage-clamp setting, were established by Keener and Earnshaw (Keener & Sneyd, 1998; Keener, 2009; Earnshaw & Keener, 2010b).

2.4 Local Convergence Rate. Keener and Earnshaw (Keener & Sneyd, 1998; Keener, 2009; Earnshaw & Keener, 2010b) showed that for Markov chains with constant (even time-varying) transition rates, (1) the multinomial probability distributions corresponding to mean-field models (such as the HH sodium or potassium models) form invariant submanifolds within the space of probability distributions over the channel states, and (2) arbitrary initial probability distributions converged exponentially quickly to the invariant manifold. For systems with prescribed time-varying transition rates, such as for an ion channel system under voltage clamp with a prescribed voltage $V(t)$ as a function of time, the distribution of channel states had an invariant submanifold, again corresponding to the multinomial distributions, and the flow on that manifold induced by the evolution equations was consistent with the flow of the full system.

In the preceding section, we established the dynamical consistency of the 14D and 4D models with enough generality to cover both the voltage-clamp and current-clamp systems; the latter is distinguished by not having a prescribed voltage trace, but rather having the voltage coevolve along with the (randomly fluctuating) ion channel states. Here, we give numerical evidence for exponential convergence under current clamp similar to that established under voltage clamp by Keener and Earnshaw.

Rather than providing a rigorous proof, we give numerical evidence for the standard deterministic HH model that $y \rightarrow \mathcal{M}$ under current clamp (spontaneous firing conditions) in the following sense: if $y(t)$ is a solution of $\dot{y} = G(y)$ with arbitrary initial data $y_0 \in \mathcal{Y}$, then $||y(t) - H(R(y(t)))|| \rightarrow 0$ as $t \rightarrow \infty$, exponentially quickly. Moreover, the convergence rate is bounded

by $\lambda = \max(\lambda_v, \lambda_{Na}, \lambda_K)$, where λ_{ion} is the least negative nontrivial eigenvalue of the channel state transition matrix (over the voltage range $V_{min} \leq v \leq V_{max}$) for a given ion, and $-1/\lambda_v$ is the largest value taken by the membrane time constant (for $V_{min} \leq v \leq V_{max}$). In practice, we find that the membrane time constant does not determine the slowest timescale for convergence to $\mathcal{M}$. In fact it appears that the second-least-negative eigenvalues (not the least-negative eigenvalues) of the ion channel matrices set the convergence rate.

Note that $y \in \mathcal{Y}$ can be written as $y = [V; \mathbf{M}; \mathbf{N}]$. As shown in appendix C, the Jacobian matrix $\frac{\partial H}{\partial x}$ consists of three block matrices: one for the voltage terms, $\frac{\partial V}{\partial v}$; one associated with the Na$^+$ gates, given by $\frac{\partial \mathbf{M}}{\partial m}$ and $\frac{\partial \mathbf{M}}{\partial h}$; and one corresponding to the K$^+$ gates, $\frac{\partial \mathbf{N}}{\partial n}$. Fixing a particular voltage v, let λ_i, $i \in \{0, 1, 2, \ldots, 7\}$ be the eight eigenvalues of A_{Na} and v_i be the associated eigenvectors, that is, $A_{Na} v_i = \lambda_i v_i$ for the rate matrix in equations 2.8. Similarly, let η_i, w_i, $i \in \{0, 1, 2, \ldots, 4\}$ be the five eigenvalues and the associated eigenvectors of A_K, that is, $A_K w_i = \eta_i w_i$, for the rate matrix in equations 2.9. If we rank the eigenvalues of either matrix in descending order, the leading eigenvalue is always zero (because the sum of each column for A_{Na} and A_K is zero for every V) and the remainder are real and negative. Let λ_1 and η_1 denote the largest (least negative) nontrivial eigenvalues of A_{Na} and A_K, respectively, and let $v_1 \in \mathbb{R}^8$ and $w_1 \in \mathbb{R}^5$ be the corresponding eigenvectors.

The eigenvectors of the full 14D Jacobian are not simply related to the eigenvectors of the component submatrices, because the first (voltage) row and column contain nonzero off-diagonal elements. However, the eigenvectors associated with the largest nonzero eigenvectors of A_{Na} and A_K (respectively, v_2 and w_2) are parallel to $\partial M/\partial h$ and $\partial N/\partial n$, regardless of voltage. In other words, the slowest decaying directions for each ion channel, v_1 and w_1, transport the flow along the multinominal submanifold of $\mathcal{Y}$. Therefore, it is reasonable to make the hypothesis that if $Y(t)$ is a solution of $\dot{y} = G(y)$ with arbitrary initial data $y \in \mathcal{Y}$, then

$$\frac{\|y(t) - H(R(y(t)))\|}{\|y(0) - H(R(y(0)))\|} \lesssim e^{-\lambda_2 t} \tag{2.24}$$

for λ_2 being the second largest nonzero eigenvalue of A_K and A_{Na} over all v in the range $v_{min} < v < v_{max}$. The convergence behavior is plotted numerically in Figure 3 and is consistent with the Ansatz equation 2.24. We calculate the distance from a point y to $\mathcal{M}$ as

$$y_{max} = \operatorname*{argmax}_{y \in \mathcal{Y}} \left\| y - H(R(y)) \right\|^2. \tag{2.25}$$
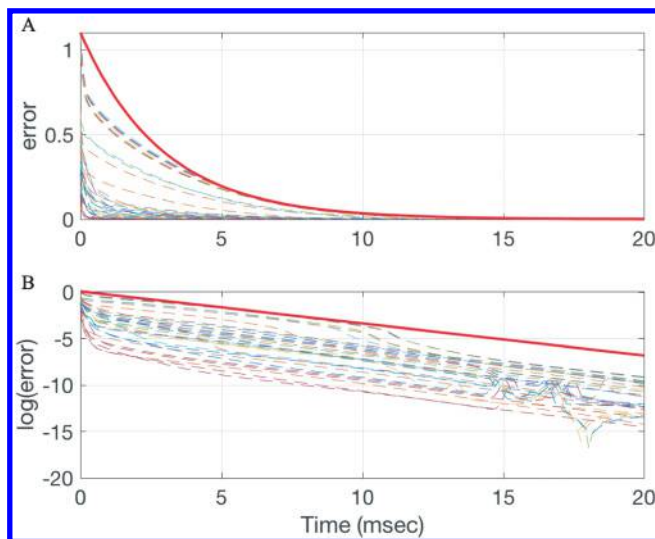
Figure 3: Convergence of trajectories $y(t)$, for arbitrary initial conditions $y_0 \in \mathcal{Y}$, to the multinomial submanifold $\mathcal{M}$, for an ensemble of random initial conditions. (A) Distance (see equation 2.25) between $y(t)$ and $\mathcal{M}$. (B) Logarithm of the distance in panel A. The red solid line shows $d_{\max}e^{-\lambda_2 t}$ in panel A and $\log(d_{\max}) - \lambda_2 t$ in panel B.

In order to obtain an upper bound on the distance as a function of time, we begin with the farthest point in the simplex from $\mathcal{M}$ by numerically finding the solution to the argument 2.25, which is

$$y_{\max} = [v, 0.5, 0, 0, 0.5, 0, 0, 0, 0, 0.5, 0, 0, 0, 0.5].$$

This vector represents the furthest possible departure from the multinomial distribution: all probability equally divided between the extreme states m_{00} and m_{03} for the sodium channel and the extremal states n_0 and n_4 for potassium. The maximum distance from the multinomial submanifold $\mathcal{M}$, $d_{\max}$, is calculated using this point. As shown in Figure 3, the function $d_{\max}\,e^{-\lambda_2 t}$ provides a tight upper bound for the convergence rate from arbitrary initial data $y \in \mathcal{Y}$ to the invariant submanifold $\mathcal{M}$.

3 Stochastic 14D Hodgkin-Huxley Models

Finite populations of ion channels generate stochastic fluctuations ("channel noise") in ionic currents that influence action potential initiation and timing (White et al., 1998; Schneidman et al., 1998). At the molecular level, fluctuations arise because transitions between individual ion channel states

are stochastic (Hill & Chen, 1972; Neher & Sakmann, 1976; Skaugen & Walløe, 1979). Each directed edge in the ion channel state transition diagrams (see Figure 1) introduces an independent noise source. It is of interest to be able to attribute variability of the interspike interval timing distribution to specific molecular noise sources, specifically individual directed edges in each channel state graph. In order to explore these contributions, we develop a system of Langevin equations for the Hodgkin-Huxley equations, set in a 14-dimensional phase space.

Working with a higher-dimensional stochastic model may appear inconvenient, but in fact has several advantages. First, any projection of an underlying 14D model onto a lower (e.g., 4D) stochastic model generally entails loss of the Markov property. Second, the higher-dimensional representation allows us to assess the contribution of individual molecular transitions to the macroscopically observable variability of timing in the interspike interval distribution. Third, by using a rectangular noise coefficient matrix constructed directly from the transitions in the ion channel graphs, we avoid a matrix decomposition step. This approach leads to a fast algorithm that is equivalent to the slower algorithm due to (Fox & Lu, 1994; Goldwyn & Shea-Brown, 2011) in a strong sense (pathwise equivalence) that we detail in section 4.

3.1 Exact Stochastic Simulation of HH Kinetics: The Random–Time-Change Representation. An "exact" representation of the Hodgkin-Huxley system with a population of M_{tot} sodium channels and N_{tot} potassium channels treats each of the 20 directed edges in the sodium channel diagram, and each of the 8 directed edges in the potassium channel diagram, as independent Poisson processes, with voltage-dependent per capita intensities. As in the deterministic case, the sodium and potassium channel population vectors $\mathbf{M}$ and $\mathbf{N}$ satisfy $\sum_{i=0}^{3} \sum_{j=0}^{1} \mathbf{M}_{ij} \equiv 1 \equiv \sum_{i=0}^{4} \mathbf{N}_i$.[3] Thus, they are constrained, respectively, to a 7D simplex embedded in $\mathbb{R}^8$ and a 4D simplex embedded in $\mathbb{R}^5$. In the random-time-change representation (Anderson & Kurtz, 2015) the exact evolution equations are written in terms of sums over the directed edges $\mathcal{E}$ for each ion channel, $\mathcal{E}_{\text{Na}} = \{1, \ldots, 20\}$ and $\mathcal{E}_{\text{K}} = \{1, \ldots, 8\}$, (see Figure 1),

$$\mathbf{M}(t) = \mathbf{M}(0) + \frac{1}{M_{\text{tot}}} \sum_{k \in \mathcal{E}_{\text{Na}}} \zeta_k^{\text{Na}} Y_k^{\text{Na}} \left(M_{\text{tot}} \int_0^t \alpha_k^{\text{Na}}(V(s)) \mathbf{M}_{i(k)}(s) \, ds \right), \quad (3.1)$$

$$\mathbf{N}(t) = \mathbf{N}(0) + \frac{1}{N_{\text{tot}}} \sum_{k \in \mathcal{E}_{\text{K}}} \zeta_k^{\text{K}} Y_k^{\text{K}} \left(N_{\text{tot}} \int_0^t \alpha_k^{\text{K}}(V(s)) \mathbf{N}_{i(k)}(s) \, ds \right). \quad (3.2)$$

---

[3]We annotate the stochastic population vector $\mathbf{M}$ as either $[M_{00}, M_{10}, \ldots, M_{31}]$ or as $[M_1, \ldots, M_8]$, whichever is more convenient. In either notation $M_{31} \equiv M_8$ is the conducting state of the Na$^+$channel. For the K$^+$channel, N_4 denotes the conducting state.

Here ζ_k^{ion} is the stoichiometry vector for the kth directed edge. If we write $i(k)$ for the source node and $j(k)$ for the destination node of edge k, then $\zeta_k^{\text{ion}} = e_{j(k)}^{\text{ion}} - e_{i(k)}^{\text{ion}}.$[4] Each $Y_k^{\text{ion}}(\tau)$ is an independent unit-rate Poisson process, evaluated at "internal time" (or integrated intensity) τ, representing the independent channel noise arising from transitions along the kth edge. The voltage-dependent per capita transition rate along the kth edge is $\alpha_k^{\text{ion}}(v)$, and $\mathbf{M}_{i(k)}(s)$ (resp. $\mathbf{N}_{i(k)}(s)$) is the fractional occupancy of the source node for the kth transition at time s. Thus, for example, the quantity $M_{\text{tot}}\alpha_k^{\text{Na}}(V(s))\mathbf{M}_{i(k)}(s)$ gives the net intensity along the kth directed edge in the Na$^+$ channel graph at time s.

Remark 1. Under voltage-clamp conditions, with the voltage V held fixed, equations 3.1 and 3.2 reduce to a time-invariant first-order transition process on a directed graph (Schmidt & Thomas, 2014; Gadgil, Lee, & Othmer, 2005).

Under current-clamp conditions, the voltage evolves according to a conditionally deterministic current balance equation of the form

$$\frac{dV}{dt} = \frac{1}{C}\{I_{\text{app}}(t) - \bar{g}_{\text{Na}}\mathbf{M}_{31}\,(V - V_{\text{Na}}) - \bar{g}_{\text{K}}\mathbf{N}_4\,(V - V_{\text{K}})$$
$$- g_{\text{leak}}(V - V_{\text{leak}})\}. \tag{3.3}$$

Here, C $(\mu F/cm^2)$ is the capacitance, I_{app} (nA/cm^2) is the applied current, the maximal conductance is $\bar{g}_{\text{chan}}$ (mS/cm^2), V_{chan} (mV) is the associated reversal potential, and the ohmic leak current is $g_{\text{leak}}(V - V_{\text{leak}})$.

The random-time-change representation, equations 3.1 to 3.3, leads to an exact stochastic simulation algorithm, given in Anderson and Kurtz (2015); equivalent simulation algorithms have been used previously (Clay & DeFelice, 1983; Newby et al., 2013). Many authors substitute a simplified Gillespie algorithm that imposes a piecewise-constant propensity approximation, ignoring the voltage dependence of the transition rates α_k^{ion} between channel transition events (Goldwyn et al., 2011; Goldwyn & Shea-Brown, 2011; Orio & Soudry, 2012; Pezo et al., 2014). The two methods give similar moment statistics, provided $N_{\text{tot}}, M_{\text{tot}} \gtrsim 40$ (Anderson & Kurtz, 2015); their similarity regarding path-dependent properties (including interspike interval distributions) has not been studied in detail. Moreover, both Markov chain algorithms are prohibitively slow for modest numbers (e.g., thousands) of channels; the exact algorithm may be even slower than the approximate Gillespie algorithm. For consistency with previous studies, in this article, we use the piecewise-constant propensity Gillespie algorithm with $M_{\text{tot}} = 6000$ Na$^+$ and $N_{\text{tot}} = 1800$ K$^+$ channels as our gold standard Markov chain (MC) model, as in Goldwyn and Shea-Brown (2011).

---

[4]We write e_i^{Na} and e_i^{K} for the ith standard unit vector in $\mathbb{R}^8$ or $\mathbb{R}^5$, respectively.

In section 3.2 we develop a 14D conductance-based Langevin model with 28 independent noise sources – one for each directed edge – derived from the random-time-change representation equations 3.1 to 3.3. In previous work (Schmidt & Thomas, 2014), we established a quantitative measure of edge importance, namely, the contribution of individual transitions (directed edges) to the variance of channel state occupancy under steady-state voltage-clamp conditions. Under voltage clamp, the edge importance was identical for each reciprocal pair of directed edges in the graph, a consequence of detailed balance. Some Langevin models lump the noise contributions of each pair of edges (Dangerfield et al., 2010, 2012; Orio & Soudry, 2012; Pezo et al., 2014). Under conditions of detailed balance, this simplification is well justified. However, as we will show (see Figure 5), under current-clamp conditions (e.g., for an actively spiking neuron) detailed balance is violated, the reciprocal edge symmetry is broken, and each pair of directed edges makes a distinct contribution to ISI variability.

3.2 Langevin Equations of the 14D HH Model. For sufficiently large number of channels, Schmidt and Thomas (2014) and Schmidt et al. (2018) showed that under voltage clamp, equations 3.1 and 3.2 can be approximated by a multidimensional Ornstein-Uhlenbeck (OU) process (or Langevin equation) in the form[5]

$$d\mathbf{M} = \sum_{k=1}^{20} \zeta_k^{\text{Na}} \left\{ \alpha_k^{\text{Na}}(V)\mathbf{M}_{i(k)} dt + \sqrt{\epsilon^{\text{Na}}\alpha_k^{\text{Na}}(V)\mathbf{M}_{i(k)}}\, dW_k^{\text{Na}} \right\}, \qquad (3.4)$$

$$d\mathbf{N} = \sum_{k=1}^{8} \zeta_k^{\text{K}} \left\{ \alpha_k^{\text{K}}(V)\mathbf{N}_{i(k)} dt + \sqrt{\epsilon^{\text{K}}\alpha_k^{\text{K}}(V)\mathbf{N}_{i(k)}}\, dW_k^{\text{K}} \right\}. \qquad (3.5)$$

Here, $\mathbf{M}$, $\mathbf{N}$, ζ_k^{ion}, and α_k^{ion} have the same meaning as in equations 3.1 and 3.2. The channel state increments in a short time interval dt are $d\mathbf{M}$ and $d\mathbf{N}$, respectively. The finite-time increment in the Poisson process Y_k^{ion} is now approximated by a gaussian process, namely, the increment dW_k^{ion} in a Wiener (Brownian motion) process associated with each directed edge. These independent noise terms are scaled by $\epsilon^{\text{Na}} = 1/M_{\text{tot}}$ and $\epsilon^{\text{K}} = 1/N_{\text{tot}}$, respectively.

Equations 3.3 to 3.5 comprise a system of Langevin equations for the HH system (under current clamp) on a 14-dimensional phase space driven by 28 independent white noise sources, one for each directed edge. These equations may be written succinctly in the form

$$d\mathbf{X} = \mathbf{f}(\mathbf{X})\, dt + \sqrt{\epsilon}\mathcal{G}(\mathbf{X})\, d\mathbf{W}(t) \qquad (3.6)$$

---

[5]The convergence of the discrete channel system to a Langevin system under voltage clamp is a special case of Kurtz's theorem (Kurtz, 1981).

where we define the 14-component vector $\mathbf{X} = (V; \mathbf{M}; \mathbf{N})$, and $\mathbf{W}(t)$ is a Wiener process with 28 independent components. The deterministic part of the evolution equation $\mathbf{f}(\mathbf{X}) = \left[\frac{dV}{dt}; \frac{d\mathbf{M}}{dt}; \frac{d\mathbf{N}}{dt}\right]$ is the same as the mean field, equations 2.7 to 2.9. The state-dependent noise coefficient matrix $\mathcal{G}$ is 14×28 and can be written as

$$\sqrt{\epsilon}\mathcal{G} = \left(\begin{array}{c|c} \mathbf{0}_{1\times20} & \mathbf{0}_{1\times8} \\ \hline S_{\text{Na}} & \mathbf{0}_{8\times8} \\ \hline \mathbf{0}_{5\times20} & S_{\text{K}} \end{array}\right).$$

The coefficient matrix S_{K} is

$$S_{\text{K}} = \frac{1}{\sqrt{N_{\text{tot}}}} \begin{bmatrix} -\sqrt{4\alpha_n n_0} & \sqrt{\beta_n n_1} & 0 & 0 \\ \sqrt{4\alpha_n n_0} & -\sqrt{\beta_n n_1} & -\sqrt{3\alpha_n n_1} & \sqrt{2\beta_n n_2} \\ 0 & 0 & \sqrt{3\alpha_n n_1} & -\sqrt{2\beta_n n_2} & \cdots \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} & 0 & 0 & 0 & 0 \\ & 0 & 0 & 0 & 0 \\ \cdots & -\sqrt{2\alpha_n n_2} & \sqrt{3\beta_n n_3} & 0 & 0 \\ & \sqrt{2\alpha_n n_2} & -\sqrt{3\beta_n n_3} & -\sqrt{\alpha_n n_3} & \sqrt{4\beta_n n_4} \\ & 0 & 0 & \sqrt{\alpha_n n_3} & -\sqrt{4\beta_n n_4} \end{bmatrix},$$

and S_{Na} is given in appendix D. Note that each of the 8 columns of S_{K} corresponds to the flux vector along a single directed edge in the K$^+$ channel transition graph. Similarly, each of the 20 columns of S_{Na} corresponds to the flux vector along a directed edge in the Na$^+$ graph (see appendix D).

Remark 2. Although the ion channel state trajectories generated by equation 3.6 are not strictly bounded to remain within the nonnegative simplex, empirically, the voltage nevertheless remains within specified limits with overwhelming probability.

To facilitate comparison of the model, equations 3.3 to 3.5, prior work (Fox & Lu, 1994; Fox, 1997; Goldwyn & Shea-Brown, 2011), we may rewrite the 14×28D Langevin description in the equivalent form:

$$C\frac{dV}{dt} = I_{\text{app}}(t) - \bar{g}_{\text{Na}}\mathbf{M}_8\,(V - V_{\text{Na}}) - \bar{g}_{\text{K}}\mathbf{N}_5\,(V - V_{\text{K}}) - g_{\text{leak}}(V - V_{\text{leak}}),$$

$$(3.7)$$

$$\frac{d\mathbf{M}}{dt} = A_{\mathrm{Na}}\mathbf{M} + S_{\mathrm{Na}}\xi_{\mathrm{Na}}, \qquad (3.8)$$

$$\frac{d\mathbf{N}}{dt} = A_{\mathrm{K}}\mathbf{M} + S_{\mathrm{K}}\xi_{\mathrm{K}}. \qquad (3.9)$$

The drift matrices A_{Na} and A_{K} remain the same as in Fox and Lu (1994), and are the same as in the 14D deterministic model, equations 2.10 and 2.11. S_{Na} and S_{K} are constructed from direct transitions of the underlying kinetics in Figure 1, and $\xi_{\mathrm{Na}} \in \mathbb{R}^{20}$ and $\xi_{\mathrm{K}} \in \mathbb{R}^{8}$ are vectors of independent gaussian white noise processes with zero mean and unit variance.

Fox and Lu's original approach (Fox & Lu, 1994) requires solving a matrix square root equation $SS^{\mathsf{T}} = D$ to obtain a square (8×8 for Na^{+} or 5×5 for K^{+}) noise coefficient matrix consistent with the state-dependent diffusion matrix D. As an advantage, the ion channel representation equations 3.7 to 3.9, uses sparse, nonsquare noise coefficient matrices (8×20 for the Na^{+} channel and 5×8 for the K^{+} channel), which exposes the independent sources of noise for the system.

The new Langevin model in equations 3.7 to 3.9 does not require detailed balance, which gives more insight into the underlying kinetics. Review papers (e.g., Goldwyn & Shea-Brown, 2011; Pezo et al., 2014; Huang et al., 2015) did systematic comparison of various stochastic versions of the HH model. In sections 4 and 5, we quantitatively analyze the connection between the new model and other existing models (Fox & Lu, 1994; Goldwyn et al., 2011; Goldwyn & Shea-Brown, 2011; Dangerfield et al., 2010, 2012; Orio & Soudry, 2012; Huang et al., 2013, 2015; Pezo et al., 2014; Fox, 2018). Problems such as the boundary constraints are beyond the scope of this article; however, we would like to connect the new model to another type of approximation to the MC model; the stochastic shielding approximation.

3.3 Stochastic Shielding for the 14D HH Model. The stochastic shielding (SS) approximation was introduced by Schmandt and Galán (2012), in order to approximate the Markov process using fluctuations from only a subset of the transitions, namely, those corresponding to changes in the observable states. In Schmidt and Thomas (2014), we showed that under voltage clamp, each directed edge makes a distinct contribution to the steady-state variance of the ion channel conductance, with the total variance being a sum of these contributions. We call the variance due to the kth directed edge the *edge importance*; assuming detailed balance, it is given by

$$R_k = J_k \sum_{i=2}^{n} \sum_{j=2}^{n} \left(\frac{-1}{\lambda_i + \lambda_j} \right) (\mathbf{c}^{\mathsf{T}} v_i) \left(w_i^{\mathsf{T}} \zeta_k \right) \left(\zeta_k^{\mathsf{T}} w_j \right) \left(v_j^{\mathsf{T}} \mathbf{c} \right). \qquad (3.10)$$

Here, J_k is the steady-state probability flux along the kth directed edge; $\lambda_n < \lambda_{n-1} \leq \ldots \leq \lambda_2 < 0$ are the eigenvalues of the drift matrix (A_{Na} or A_{K},

(A) HH K channel: R_k as a function of voltage

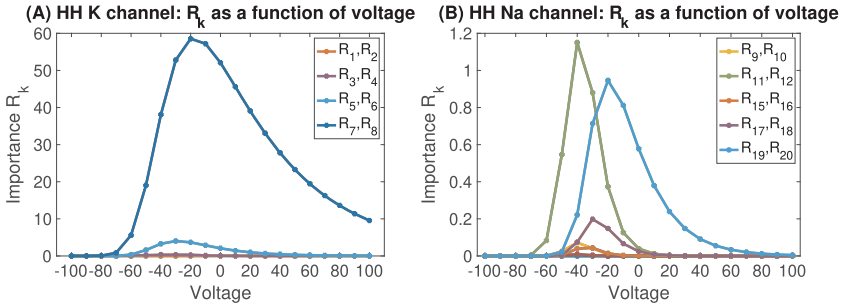**(B) HH Na channel: R_k as a function of voltage**

Figure 4: Stochastic shielding under voltage clamp. Redrawn (with permission) from Figures 10 and 13 of Schmidt and Thomas (2014). Each curve shows the edge importance R_k (see equation 3.10) as a function of voltage in the range $[-100, 100]$ mV for a different edge pair. For the K$^+$ kinetics, $R_7 = R_8$ are the largest R_k value in the voltage range above. For the Na$^+$ kinetics, $R_{11} = R_{12}$ have the largest R_k values in the subthreshold voltage range (c. $[-100, -25]$ mV), and $R_{19} = R_{20}$ have the largest R_k values in the suprathreshold voltage range (c. $[-25, 100]$ mV).

respectively), and v_i (resp. w_i) are the corresponding right (resp. left) eigenvectors of the drift matrix. Each edge's stoichiometry vector ζ_k has components summing to zero; consequently, the columns of A_{Na} and A_K all sum to zero. Thus, each drift matrix has a leading trivial eigenvalue $\lambda_1 \equiv 0$. The vector $\mathbf{c}$ specifies the unitary conductance of each ion channel state; for the HH model, it is proportional to e_8^{Na} or e_5^K, respectively.

Figure 4 shows the edge importance for each pair of edges in the HH Na$^+$ and K$^+$ ion channel graph, as a function of voltage in the range $[-100, 100]$ mV. Note that reciprocal edges have identical R_k due to detailed balance. Under voltage clamp, the largest value of R_k for the HH channels always corresponds to directly observable transitions, that is, edges k such that $|\mathbf{c}^\intercal \zeta_k| > 0$, although this condition need not hold in general (Schmidt, Galán, & Thomas, 2018).

To apply the stochastic shielding method under current clamp, we simulate the model with noise from only a selected subset $\mathcal{E}' \subset \mathcal{E}$ of directed edges, replacing equations 3.8 to 3.9 with

$$\frac{d\mathbf{M}}{dt} = A_{Na}\mathbf{M} + S'_{Na}\xi_{Na}, \tag{3.11}$$

$$\frac{d\mathbf{N}}{dt} = A_K\mathbf{M} + S'_K\xi_K, \tag{3.12}$$

where S'_{Na} (resp. S'_K) is a reduced matrix containing only the noise coefficients from the most important edges $\mathcal{E}'$. That is, $\mathcal{E}'$ contains a subset of edges with the largest edge-importance values R_k.

Schmandt and Galán (2012) assumed that the edges with the largest contribution to current fluctuations under voltage clamp would also make the largest contributions to variability in voltage and timing under current clamp and included edges 7 and 8 of the K^+ channel ($\mathcal{E}'_K = \{7, 8\}$) and edges 11 and 12 and 19 and 20 of the Na^+ channel ($\mathcal{E}'_{Na} = \{11, 12, 19, 20\}$), yielding an 8×4 matrix S'_{Na} and a 5×2 matrix S'_K. They demonstrated numerically that restricting stochastic forcing to these edges gave a significantly faster simulation with little appreciable change in statistical behavior: under voltage clamp, the mean current remained the same, with a small (but noticeable) decrease in the current variance; meanwhile, similar interspike interval (ISI) statistics were observed.

Under current clamp, detailed balance is violated, and it is not clear from mathematical principles whether the edges with the largest R_k under voltage clamp necessarily make the largest contribution under other circumstances. In order to evaluate the contribution of the fluctuations driven by each directed edge on ISI variability, we test the stochastic shielding method by removing all but one column of S_{ion} at a time. That is, we restrict to a single noise source and observe the resulting ISI variance empirically. For example, to calculate the importance of the kth direct edge in the Na^+ channel, we suppress the noise from all other edges by setting $S'_K \xi_K = \mathbf{0}_{5 \times 1}$ and

$$S'_{Na} = [\mathbf{0}_{8 \times 1}, \cdots, S_{Na}(:, k), \cdots, \mathbf{0}_{8 \times 1}],$$

that is, we include only the kth column of S_{Na} and set other columns to be zeros. The ISI variance was calculated from an ensemble of 10^4 voltage traces, each spanning about 500 ISIs.

Figure 5A plots the logarithm of the ISI variance for each edge in $\mathcal{E}_K$. Vertical bars (cyan) show the ensemble mean of the ISI variance, with a 95% confidence interval superimposed (magenta). Several observations are in order. First, the ISI variance driven by the noise in each edge decreases rapidly the farther the edge is from the observable transitions (edges 7,8), reflecting the underlying "stochastic shielding" phenomenon. Second, the symmetry of the edge importance for reciprocal edge pairs—(1,2), (3,4), (5,6), and (7,8)—that is observed under voltage clamp is broken under current clamp. The contribution of individual directed edges to timing variability under current clamp has another important difference compared with the edge importance (current fluctuations) under voltage clamp. A similar breaking of symmetry for reciprocal edges is seen for the Na^+ channel, again reflecting the lack of detailed balance during active spiking.

Figure 5B shows the ISI variance when channel noise is included on individual edges of $\mathcal{E}_{Na}$. Here, the difference between voltage and current clamp is striking. Under voltage clamp, the four most important edges are always those representing observable transitions, in the sense that the transition's stoichiometry vector ζ is not orthogonal to the conductance vector $\mathbf{c}$.
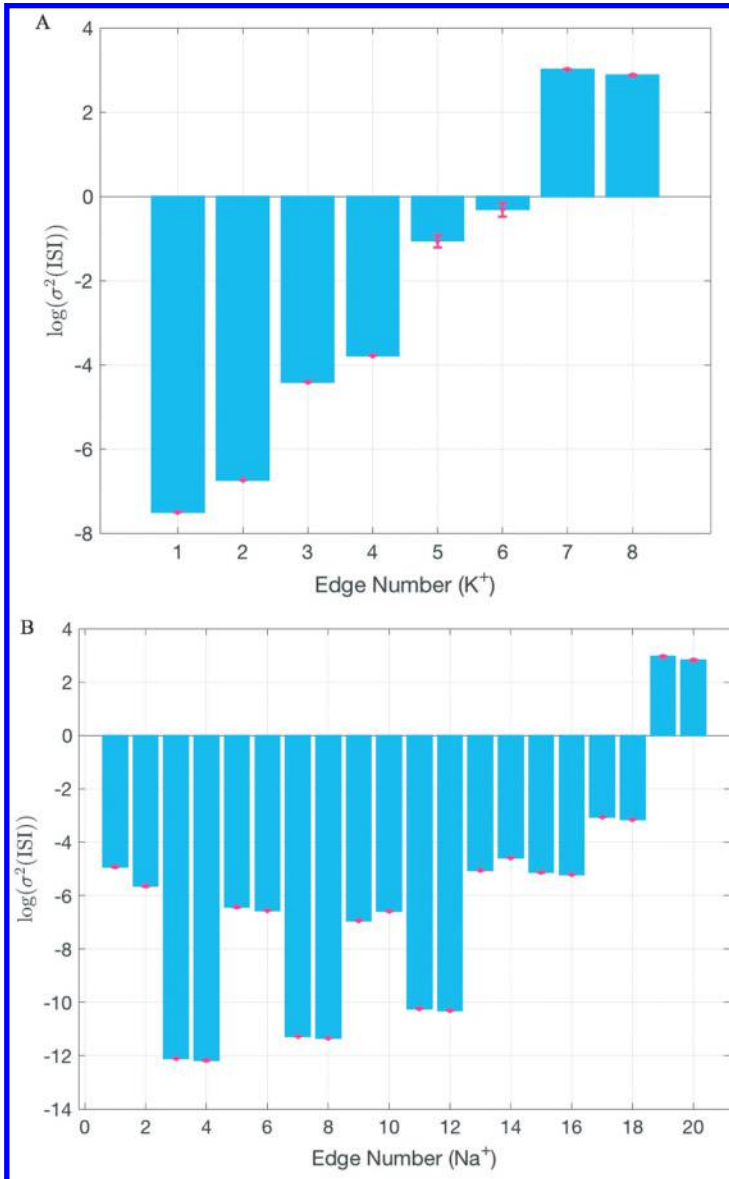
Figure 5: Logarithm of variance of ISI for stochastic shielding under current clamp. The cyan bar is the mean of ISI, and the magenta plots the 95% confidence interval of the mean ISI (see text for details). The applied current is 10 nA with other parameters specified in the appendixes. For the K^+ kinetics, the largest contribution edge is 7, and 8 is slightly smaller, ranking the second largest. For the Na^+ kinetics, the largest contribution pair is 19 and 20, with 20 slightly smaller than 19. Moreover, edge 17 and 18 is the second largest pair.

That is, the four most important pairs are always 11 and 12 and 19 and 20, regardless of voltage (see Figure 4). Under current clamp, the most important edges are 17, 18, 19, and 20. Although edges 11 and 12 are among the four most important sources of channel population fluctuations under voltage clamp, they are not even among the top 10 contributors to ISI variance when taken singly. Even though edges 17 and 18 are hidden, meaning they do not directly change the instantaneous channel conductance, these edges are nevertheless the second most important pair under current clamp. Therefore, when we implement the stochastic-shielding based approximation, we include the pairs 17 and 18 and 19 and 20 in equation 3.11. We refer to the approximate SS model driven by these six most important edges as the 14×6D HH model.

Given the other parameters we use for the HH model (see Table 5 in appendix B), the input current of $I_{\text{app}} = 10$ nA is slightly beyond the region of multistability associated with a subcritical Andronov-Hopf bifurcation. In order to make sure the results are robust against increases in the applied current, we tried current injections ranging from 20 to 100 nA. While injecting larger currents decreased the ISI variance, it did not change the rank order of the contributions from the most important edges.

4 Pathwise Equivalence for a Class of Langevin Models

Fox and Lu's method has been widely used since its appearance (see references in Bruce, 2009; Goldwyn & Shea-Brown, 2011; Huang et al., 2015), and the "best" approximation for the underlying Markov chain (MC) model has been a subject of ongoing discussion for decades. Several studies (Mino et al., 2002; Bruce, 2009; Sengupta, Laughlin, & Niven, 2010) attested to discrepancies between Fox's later approach (Fox, 1997) and the discrete-state MC model, raising the question of whether Langevin approximations could ever accurately represent the underlying fluctuations seen in the gold standard MC models. An influential review paper (Goldwyn & Shea-Brown, 2011) found that these discrepancies were due to the way in which noise is added to the stochastic differential equations, 1.1 to 1.3. Recent studies, including Dangerfield et al. (2010, 2012); Linaro et al. (2011); Goldwyn and Shea-Brown (2011); Goldwyn et al. (2011); Orio and Soudry (2012); Güler (2013b); Huang et al. (2013, 2015); Pezo et al. (2014); and Fox (2018), discussed various ways of incorporating channel noise into HH kinetics based on the original work by Fox and Lu (Fox & Lu, 1994; Fox, 1997), some of which have the same SDEs but with different boundary conditions. Different boundary conditions (BCs) are not expected to have much impact on computational efficiency. Indeed, if BCs are neglected, the main difference between channel-based (or conductance-based) models is the diffusion matrix S in the Langevin euqations 1.2 and 1.3. As the discussion about where and how to incorporate noise into the HH model framework goes on, Fox

(2018) recently asked whether there is a way of relating different models with different S matrices. We give a positive answer to this question below.

In section 4.1 we demonstrate the equivalence (neglecting the boundary conditions) of a broad class of previously proposed channel-based Langevin models (Fox & Lu, 1994; Dangerfield et al., 2010, 2012; Goldwyn & Shea-Brown, 2011; Orio & Soudry, 2012; Huang et al., 2013; Pezo et al., 2014; Fox, 2018) and the 14D Langevin HH model with 28 independent noise sources (one for each directed edge in the channel state transition graph), that is, our 14 × 28D Langevin model.

4.1 When Are Two Langevin Equations Equivalent? Two Langevin models are pathwise equivalent if the sample paths (trajectories) of one model can be made to be identical to the sample paths of the other under an appropriate choice of gaussian white noise samples for each. To make this notion precise, consider two channel-based Langevin models of the form $\mathbf{dX} = \mathbf{f}(\mathbf{X})\,dt + G(\mathbf{X})\,\mathbf{dW}$ with the same mean dynamics $\mathbf{f} \in \mathbb{R}^d$ and two different $d \times n$ matrices (possibly with different values of n_1 and n_2), G_1 and G_2. Denote

$$\mathbf{f} : \mathbb{R}^d \to \mathbb{R}^d, \tag{4.1}$$

$$G_1 : \mathbb{R}^d \to \mathbb{R}^{d \times n_1}, \tag{4.2}$$

$$G_2 : \mathbb{R}^d \to \mathbb{R}^{d \times n_2}. \tag{4.3}$$

Let $\mathbf{X}(t) = [X_1(t), X_2(t), \ldots, X_d(t)]^\mathsf{T}$ and $\mathbf{X}^*(t) = [X_1^*(t), X_2^*(t), \ldots, X_d^*(t)]^\mathsf{T}$ be trajectories produced by the two models, and let $\mathbf{W}(t) = [W_1(t), W_2(t), \ldots, W_{n_1}(t)]^\mathsf{T}$ and $\mathbf{W}^*(t) = [W_1^*(t), W_2^*(t), \ldots, W_{n_2}^*(t)]^\mathsf{T}$ be vectors of Weiner processes. That is, $W_i(t)$, $i = 1, 2, \ldots, n_1$ and $W_j^*(t)$, $j = 1, 2, \ldots, n_2$ are independent Wiener processes with $\langle W_i(s)W_j(t)\rangle = \delta_{ij}\delta(t - s)$ and $\langle W_i^*(s)W_j^*(t)\rangle = \delta_{ij}\delta(t - s)$. Note that n_1 and n_2 need not be equal. As defined in (Allen, Allen, Arciniega, & Greenwood, 2008), the stochastic differential equation (SDE) models

$$\mathbf{dX} = \mathbf{f}(t, \mathbf{X}(t))dt + G_1(t, \mathbf{X}(t))\mathbf{dW}(t) \tag{4.4}$$

and

$$\mathbf{dX}^* = \mathbf{f}(t, \mathbf{X}^*(t))dt + G_2(t, \mathbf{X}^*(t))\mathbf{dW}^*(t) \tag{4.5}$$

are pathwise equivalent if systems 4.4 and 4.5 possess the same probability distribution and, moreover, a sample path solution of one equation is also a sample solution to the other one. Allen et al. (2008) proved a theorem giving general conditions under which the trajectories of two SDEs are equivalent. We follow their construction closely below, adapting it to the case of two

different Langevin equations for the Hodgkin-Huxley system represented in a 14-dimensional state space.

As in section 3, channel-based Langevin models for the stochastic dynamics of HH can be written as

$$d\mathbf{X} = \mathbf{f}(\mathbf{X})\,dt + \mathcal{S}(\mathbf{X})\,d\mathbf{W}(t), \tag{4.6}$$

where the 14-component random vector $\mathbf{X} = (V; \mathbf{M}; \mathbf{N})$ and $\mathbf{f}(\mathbf{x}) = \left[\frac{dV}{dt}; \frac{d\mathbf{M}}{dt}; \frac{d\mathbf{N}}{dt}\right]$ is the same as the mean-field, equations 2.7 to 2.9. Recall that $\mathbf{x} = [v, \mathbf{m}, \mathbf{n}]^\mathsf{T}$. Here we write

$$\mathcal{S}(\mathbf{x}) = \left(\begin{array}{c|c} \mathbf{0}_{1\times m} & \mathbf{0}_{1\times n} \\ \hline S_{\mathrm{Na}}(\mathbf{m}) & \mathbf{0}_{8\times n} \\ \hline \mathbf{0}_{5\times m} & S_{\mathrm{K}}(\mathbf{n}) \end{array}\right), \quad \text{with}$$

$$S_{\mathrm{Na}} : \mathbb{R}^8 \rightarrow \mathbb{R}^{8\times m}, \tag{4.7}$$

for the Na$^+$ channel and

$$S_{\mathrm{K}} : \mathbb{R}^5 \rightarrow \mathbb{R}^{5\times n} \tag{4.8}$$

for the K$^+$ channel. Here, m is the number of independent white noise forcing terms affecting the sodium channel variables, while n is the number of independent noise sources affecting the potassium gating variables. We write

$$\mathbf{W}(t) = [W_1(t), W_2(t), \ldots, W_{m+n}(t)]^\mathsf{T}$$

for a Wiener process incorporating both the sodium and potassium noise forcing. Given two channel-based models with diffusion matrices,

$$S_{\mathrm{Na},1} : \mathbb{R}^8 \rightarrow \mathbb{R}^{8\times m_1}, \tag{4.9}$$

$$S_{\mathrm{Na},2} : \mathbb{R}^8 \rightarrow \mathbb{R}^{8\times m_2}, \tag{4.10}$$

for the Na$^+$ channel, and

$$S_{\mathrm{K},1} : \mathbb{R}^5 \rightarrow \mathbb{R}^{5\times n_1}, \tag{4.11}$$

$$S_{\mathrm{K},2} : \mathbb{R}^5 \rightarrow \mathbb{R}^{5\times n_2}, \tag{4.12}$$

for the K$^+$ channel, we construct the diffusion matrix $\mathcal{D} = \mathcal{S}\mathcal{S}^\mathsf{T}$. In order for the two models to generate equivalent sample paths, it suffices that they have the same diffusion matrix:

$$\mathcal{D} = \mathcal{S}_1\mathcal{S}_1^{\mathsf{T}} = \begin{pmatrix} \mathbf{0}_{1\times 1} & \mathbf{0}_{1\times 8} & \mathbf{0}_{1\times 5} \\ \hline \mathbf{0}_{8\times 1} & D_{\mathrm{Na}} & \mathbf{0}_{8\times 5} \\ \hline \mathbf{0}_{5\times 1} & \mathbf{0}_{5\times 8} & D_{\mathrm{K}} \end{pmatrix} = \mathcal{S}_2\mathcal{S}_2^{\mathsf{T}}.$$

The SDEs corresponding to the two channel-based Langevin models are

$$\mathbf{dX} = \mathbf{f}(t, \mathbf{X}(t))dt + \mathcal{S}_1(t, \mathbf{X}(t))\mathbf{dW}(t), \tag{4.13}$$

$$\mathbf{dX}^* = \mathbf{f}(t, \mathbf{X}^*(t))dt + \mathcal{S}_2(t, \mathbf{X}^*(t))\mathbf{dW}^*(t). \tag{4.14}$$

The probability density function $p(t, \mathbf{x})$ for random variable $\mathbf{X}$ in equation 4.13 satisfies the Fokker-Planck equation:

$$\frac{\partial p(t, \mathbf{x})}{\partial t} = \frac{1}{2} \sum_{i=1}^{8} \sum_{j=1}^{8} \frac{\partial^2}{\partial x_i x_j} \left[p(t, \mathbf{x}) \sum_{l=1}^{m_1+n_1} \mathcal{S}_1^{(i,l)}(t, \mathbf{x})\mathcal{S}_1^{(j,l)}(t, \mathbf{x}) \right]$$

$$- \sum_{i=1}^{8} \frac{\partial}{\partial x_i} \left[\mathbf{f}_i(t, \mathbf{x})p(t, \mathbf{x}) \right]$$

$$= \frac{1}{2} \sum_{i=1}^{8} \sum_{j=1}^{8} \frac{\partial^2}{\partial x_i x_j} \left[\mathcal{D}^{(i,j)}(t, \mathbf{x})p(t, \mathbf{x}) \right] - \sum_{i=1}^{8} \frac{\partial}{\partial x_i} \left[\mathbf{f}_i(t, \mathbf{x})p(t, \mathbf{x}) \right],$$

$$\tag{4.15}$$

where $\mathcal{S}_1^{(i,j)}(t, \mathbf{x})$ is the (i, j)th entry of the diffusion matrix $\mathcal{S}_1(t, \mathbf{x})$. Equation 4.15 holds because

$$\mathcal{D}^{(i,j)}(t, \mathbf{x}) = \sum_{l=1}^{m_1+n_1} \mathcal{S}_1^{(i,l)}(t, \mathbf{x})\mathcal{S}_1^{(j,l)}(t, \mathbf{x}).$$

If $\mathbf{z}_1, \mathbf{z}_2 \in \mathbb{R}^{14}$ and $\mathbf{z}_1 \leqslant \mathbf{z}_2$, then

$$P(\mathbf{z}_1 \leqslant \mathbf{X}(t) \leqslant \mathbf{z}_2) = \int_{z_{1,14}}^{z_{2,14}} \int_{z_{1,13}}^{z_{2,13}} \cdots \int_{z_{1,1}}^{z_{2,1}} p(t, \mathbf{x})dx_1 dx_2 \cdots dx_8.$$

Note that equations 4.13 and 4.14 have the same expression equation 4.15, for the Fokker-Planck equation; therefore, $\mathbf{X}$ and $\mathbf{X}^*$ possess the same probability density function. In other words, the probability density function of $\mathbf{X}$ in equation 4.6 is invariant for different choices of the diffusion matrix $\mathcal{S}$.

4.2 Map Channel-based Langevin Models to Fox and Lu's Model. We now explicitly construct a mapping between Fox and Lu's (1994) 14D model and any channel-based model (given the same boundary conditions). We begin with a channel-based Langevin description

$$\mathbf{dX} = \mathbf{f}(t, \mathbf{X}(t))dt + \mathcal{S}(t, \mathbf{X}(t))\mathbf{dW}(t), \tag{4.16}$$

and Fox and Lu's (1994) model,

$$\mathbf{dX}^* = \mathbf{f}(t, \mathbf{X}^*(t))dt + \mathcal{S}_0(t, \mathbf{X}^*(t))\mathbf{dW}^*(t), \tag{4.17}$$

where S is a d by m matrix satisfying $\mathcal{SS}^\mathsf{T} = \mathcal{D}$ (note that $\mathcal{S}$ is not necessarily a square matrix), and $\mathcal{S}_0 = \sqrt{\mathcal{D}}$.

Let T be the total simulation time of the random process in equations 4.16 and 4.17. For $0 \leqslant t \leqslant T$, denote the singular value decomposition (SVD) of S as

$$\mathcal{S}(t) = P(t)\Lambda(t)Q(t),$$

where $P(t)$ is a $d \times d$ orthogonal matrix (i.e., $P^\mathsf{T}P = PP^\mathsf{T} = I_d$), $Q(t)$ is an $m \times m$ orthogonal matrix, and $\Lambda(t)$ is a $d \times m$ matrix with rank$(\Lambda) = r \leqslant d$ positive diagonal entries and $d - r$ zero diagonal entries.

First, we prove that given a Wiener trajectory, $\mathbf{W}(t)$, $t \in [0, T]$ and the solution to equation 4.16 $\mathbf{X}(t)$, there exists a Wiener trajectory $\mathbf{W}^*(t)$ such that the solution to equation 4.17, $\mathbf{X}^*$, is also a solution to equation 4.16. In other words, for a Wiener process $\mathbf{W}(t)$, we can construct a $\mathbf{W}^*(t)$, such that $\mathbf{X}^*(t) = \mathbf{X}(t)$, for $0 \leq t \leq T$.

Following Allen et al. (2008), we construct the vector $\mathbf{W}^*(t)$ of d independent Wiener processes as follows:

$$\mathbf{W}^*(t) = \int_0^t P(s)\left[\left(\Lambda(s)\Lambda^\mathsf{T}(s)\right)^{\frac{1}{2}}\right]^+ \Lambda(s)Q(s)\mathbf{dW}(s) + \int_0^t P(s)\mathbf{dW}^{**}(s) \tag{4.18}$$

for $0 \leqslant t \leqslant T$, where $\mathbf{W}^{**}(t)$ is a vector of length d with the first r entries equal to 0 and the next $d - r$ entries independent Wiener processes, and $\left[\left(\Lambda(s)\Lambda^\mathsf{T}(s)\right)^{\frac{1}{2}}\right]^+$ is the pseudoinverse of $\left(\Lambda(s)\Lambda^\mathsf{T}(s)\right)^{\frac{1}{2}}$. Consider that

$$\mathcal{D}(t) = \mathcal{S}(t)\mathcal{S}^\mathsf{T}(t) = P(t)\Lambda(t)Q(t)\left[P(t)\Lambda(t)Q(t)\right]^\mathsf{T} \tag{4.19}$$

$$= P(t)\Lambda(t)\Lambda^\mathsf{T}(t)P^\mathsf{T}(t) \tag{4.20}$$

$$= [\mathcal{S}_0(t)]^2, \tag{4.21}$$

where $\mathcal{S}_0(t) = P(t)\left(\Lambda(t)\Lambda^\mathsf{T}(t)\right)^{\frac{1}{2}}P^\mathsf{T}(t)$ is a square root of $\mathcal{D}$, by construction.

The diffusion term on the right side of equation 4.17 with $\mathbf{X}^*(t)$ replaced by $\mathbf{X}(t)$ satisfies

$$
\mathcal{S}_0(t, \mathbf{X}(t))\mathbf{dW}^*(t)
$$

$$
= \mathcal{S}_0(t)\left(P(t)\left[\left(\Lambda(t)\Lambda^\mathsf{T}(t)\right)^{\frac{1}{2}}\right]^+\Lambda(t)Q(t)\mathbf{dW}(t) + P(t)\mathbf{dW}^{**}(t)\right)
$$

$$
= P(t)\left(\Lambda(t)\Lambda^\mathsf{T}(t)\right)^{\frac{1}{2}}P^\mathsf{T}(t)P(t)\left[\left(\Lambda(t)\Lambda^\mathsf{T}(t)\right)^{\frac{1}{2}}\right]^+\Lambda(t)Q(t)\mathbf{dW}(t)
$$

$$
+ P(t)\left(\Lambda(t)\Lambda^\mathsf{T}(t)\right)^{\frac{1}{2}}P^\mathsf{T}(t)P(t)\mathbf{dW}^{**}(t)
$$

$$
= \{P(t)\Lambda(t)Q(t)\}\,\mathbf{dW}(t). \tag{4.22}
$$

From the SVD of $\mathcal{S} = P\Lambda Q$, we conclude that

$$
\mathcal{S}_0(t, \mathbf{X}(t))\mathbf{dW}^*(t) = \mathcal{S}(t, \mathbf{X}(t))\mathbf{dW}(t). \tag{4.23}
$$

Hence, $\mathbf{dX} = \mathbf{f}(t, \mathbf{X}(t))dt + \mathcal{S}_0(t, \mathbf{X}(t))\mathbf{dW}_t^*$, that is, $\mathbf{X}(t)$ is a sample path solution of equation 4.17.

Similarly, given a Wiener trajectory $\mathbf{W}^*(t)$ and the solution to equation 4.17 $\mathbf{X}^*(t)$, we can construct a vector $\mathbf{W}(t)$ of m independent Winner processes as

$$
\mathbf{W}(t) = \int_0^t Q^\mathsf{T}(s)\Lambda^+(s)\left[\Lambda(s)\Lambda^\mathsf{T}(s)\right]^{1/2}P^\mathsf{T}(s)\mathbf{dW}^*(s) + \int_0^t Q^\mathsf{T}(s)\mathbf{dW}^{***}(s) \tag{4.24}
$$

for $0 \leqslant t \leqslant T$, where $\mathbf{W}^{***}(t)$ is a vector of length m with the first r entries equal to 0 and the next $m - r$ entries independent Wiener processes, and $\Lambda^+(s)$ is the pseudoinverse of $\Lambda(s)$. Then, by an argument parallel to equation 4.22, we conclude that

$$
\mathcal{S}(t, \mathbf{X}^*(t))dW(t) = \mathcal{S}_0(t, \mathbf{X}^*(t))\mathbf{dW}^*(t). \tag{4.25}
$$

Hence, $\mathbf{dX}^* = \mathbf{f}(t, \mathbf{X}^*(t))dt + \mathcal{S}(t, \mathbf{X}^*(t))\mathbf{dW}(t)$, that is, $\mathbf{X}^*(t)$ is also a solution to equation 4.16. Therefore, we can conclude that the channel-based Langevin model in equation 4.16 is pathwise equivalent to Fox and Lu's model.

To illustrate pathwise equivalence, Figure 6 plots trajectories of the $14 \times 28D$ stochastic HH model and Fox and Lu's model using noise traces dictated by the preceding construction. In panel A, we generated a sample path for equation 4.16 and plot three variables in $\mathbf{X}$: the voltage V, Na^+ channel open probability M_{31}, and K^+ channel open probability N_4. The corresponding trajectory, $\mathbf{X}^*$, for Fox and Lu's model was generated
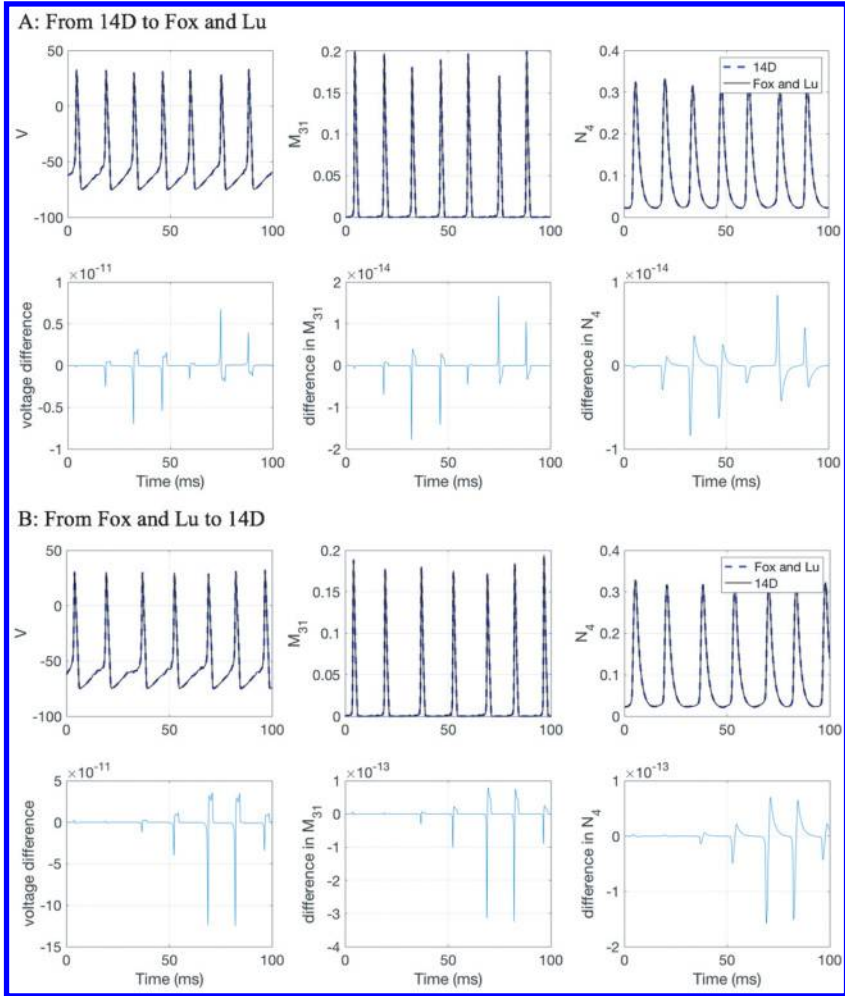
Figure 6: Pathwise equivalency of 14D HH model and Fox and Lu's model. (A) Given a sample path of the 14 × 28D Langevin model in equation 4.16, we construct the noise by equation 4.18 and generate the sample trajectory of Fox and Lu's model using equation 4.17. (B) Given a sample path of Fox and Lu's model in equation 4.17, we construct the noise by equation 4.24 and generate the sample trajectory of the 14 × 28D Langevin model using equation 4.16. For both cases, we plot the voltage V, the open probability of the Na$^+$ channel (M_{31}), and the open probability of the K$^+$ channel (N_4). We obtain excellent agreement in both directions.

from equation 4.17 and the corresponding Wiener trajectory was calculated using equation 4.18. The top three subplots in panel A superposed the voltage V^*, Na$^+$ channel open probability M_{31}^* and K$^+$ channel open probability N_4^* in $\mathbf{X}^*$ against those in $\mathbf{X}$. The bottom three subplots in panel A plot the point-wise differences of each variable. Equations 4.16 and 4.17 are numerically solved in Matlab using the Euler-Maruyama method with a time step $dt = 0.001$ ms. The slight differences observed arise in part due to numerical errors in calculating the singular value decomposition of $\mathcal{S}$ (in equation 4.16); another source of error is the finite accuracy of the Euler-Maruyama method.[6] As shown in Figure 6, most differences occur near the spiking region, where the system is numerically very stiff and the numerical accuracy of the SDE solver accounts for most of the discrepancies (analysis of which is beyond the scope of this article). To illustrate pathwise equivalence, Figure 6 shows superposed voltage trajectories for each simulation, as well as the point-wise voltage differences of each. We can conclude from the comparison in Figure 6 that the 14×28D Langevin model is pathwise equivalent with Fox and Lu's model. Similarly, the same analogy applies for other channel-based Langevin models such that with the same diffusion matrix $\mathcal{D}(\mathbf{X})$.

We have shown that our 14×28D model, with a 14-dimensional state space and 28 independent noise sources (one for each directed edge), is pathwise equivalent to Fox and Lu's original 1994 model, as well as other channel-based models (under corresponding boundary conditions) including Goldwyn et al. (2011); Goldwyn and Shea-Brown (2011); Orio and Soudry (2012); Pezo et al. (2014); Fox (2018). As we shall see in section 5, the pathwise equivalent models give statistically indistinguishable interspike interval distributions under the same BCs. We emphasize the importance of boundary conditions for pathwise equivalence. Two simulation algorithms with the same A_i and S_i matrices will generally have nonequivalent trajectories if different boundary conditions are imposed. For example, Dangerfield et al. (2012) employ the same dynamics as Orio and Soudry (2012) away from the boundary, where ion channel state occupancy approaches zero or one. But where the latter allow trajectories to move freely across this boundary (which leads only to small, short-lived excursions into "nonphysical" values), Dangerfield imposes reflecting boundary conditions through a projection step at the boundary. As we will see ln section 5, this difference in boundary conditions leads to a statistically significant difference in the ISI distribution, as well as a loss of accuracy when compared with the gold standard Markov chain simulation.

---

[6] The forward Euler method is first-order accurate for ordinary differential equations, but the forward Euler-Maruyama method is only $O(\sqrt{dt})$ accurate for stochastic differential equations (Kloeden & Platen, 1999).

5 Model Comparison

In section 3, we studied the contribution of every directed edge to the ISI variability and proposed how stochastic shielding could be applied under current clamp. Moreover, in section 4, we proved that a family of Langevin models is pathwise equivalent.

Here we compare the accuracy and computational efficiency of several models, including the "subunit model" (Fox, 1997; Goldwyn & Shea-Brown, 2011), Langevin models with different S matrices or boundary conditions (Fox & Lu, 1994; Goldwyn & Shea-Brown, 2011; Dangerfield et al., 2012; Orio & Soudry, 2012; Pezo et al., 2014), the 14D HH model (proposed in section 3.2), the 14D stochastic shielding model with six independent noise sources (proposed in section 3.3), and the gold standard Markov chain model (discussed in section 3.1). Where other studies have compared moment statistics such as the mean firing frequency (under current clamp) and stationary channel open probababilies (under voltage clamp), we base our comparison on the entire interspike interval (ISI) distributions, under current clamp with a common fixed driving current. We use two different comparisons of ISI distributions, the first based on the L_1 norm of the difference between two distributions (the Wasserstein distance; Wasserstein, 1969), in section 5.1 and the second based on the L_∞ norm (the Kolmogorov-Smirnov test; Kolmogorov, 1933; Smirnov, 1948), in section 5.2. We find similar results using both measures: as expected, the models that produce pathwise equivalent trajectories (Fox & Lu, 1994; Orio & Soudry; and our $14 \times 28D$ model) have indistinguishable ISI statistics, while the nonequivalent models (Fox, 1997; Dangerfield, 2010, 2012; Goldwyn & Shea-Brown, 2011; our $14 \times 6D$ stochastic-shielding model) have significantly different ISI distributions. Of these, the $14 \times 6D$ SS model is the closest to the models in the $14 \times 28D$ class and as fast as any other model considered.

5.1 L_1 Norm Difference of ISIs. We first evaluate the accuracy of different stochastic simulation algorithms by comparing their ISI distributions under current clamp to that produced by a reference algorithm, namely, the discrete-state Markov chain (MC) algorithm.

Let $X_1, X_2, \ldots, X_n$ be n independent samples of ISIs with a true cumulative distribution function F. Let $F_n(\cdot)$ denote the corresponding empirical cumulative distribution function (ECDF) defined by

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{X_i \leq x\}}, \qquad x \in \mathbb{R}, \tag{5.1}$$

where we write $\mathbf{1}_A$ to denote the indicator function for the set A. Let Q and Q^M be the quantile functions of F and F^M, respectively. The L_1-Wasserstein

Table 3: Summary of the L_1-Wasserstein Distances of ISI Distributions for Langevin Type Hodgkin-Huxley Models Compared to the MC Model.

Model	Variables V+M+N	S Matrix	Noise Dimensions Na+K	L_1 Norm (msec.) (Wasserstein Dist.)	Run Time (sec.)
MC	1+8+5	n/a	20+8	2.27 e-4 ± 7.15 e-5	3790
Fox94	1+7+4	$S = \sqrt{D}$	7+4	4.74 e-2 ± 1.93 e-4	2436
Fox97	1+2+1	n/a	3	8.01 e-1 ± 9.48 e-4	67
Dangerfield	1+8+5	S_{EF}	10+4	2.18 e-1 ± 2.14 e-4	655
Goldwyn	1+8+5	$S = \sqrt{D}$	8+5	1.83 e-1 ± 1.93 e-4	2363
Orio	1+8+5	S_{paired}	10+4	4.52 e-2 ± 2.08 e-4	577
14 × 28D	1+8+5	S_{single}	20+8	4.93 e-2 ± 1.94 e-4	605
SS	1+8+5	S_{ss}	4+2	7.62 e-2 ± 7.57 e-5	73

Notes: Model (see text for details): MC: Markov chain. Fox94; From Fox and Lu (1994). Fox97: Fox (1997). Goldwyn: Goldwyn and Shea-Brown (2011). Dangerfield: Dangerfield et al. (2010, 2012). 14 × 28D: model proposed in section 3.2. SS: stochastic-shielding model (see section 3.3). Variables: number of degrees of freedom in Langevin equation representing voltage, sodium gates, and potassium gates, respectively. S Matrix: Form of the noise coefficient matrix in equations 1.1 to 1.3. Noise dimensions: number of independent gaussian white noise sources represented for sodium and potassium, respectively. L_1 Norm: Empirically estimated L_1-Wasserstein distance between the model's ISI distribution and the MC model's ISI distribution. For MC versus MC, independent trials were compared. $a \pm b$: mean±standard deviation. Run time (in seconds): see text for details.

distance between two CDF's F^M and F can be written as (Shorack & Wellner, 2009)

$$\rho_1(F, F^M) = \int_0^\infty |F(x) - F^M(x)| \, dx = \int_0^1 |Q(x) - Q^M(x)| \, dx. \qquad (5.2)$$

Note that ρ_1 has the same units as dx. Thus, the L_1 distances reported in Table 3 have units of milliseconds.

When two models have the same number of samples, n, equation 5.2 can be estimated by

$$\int_0^1 |Q(x) - Q^M(x)| \, dx \approx \frac{1}{n} \sum_{i=1}^n |X_i - Y_i| := \rho_1(F_n, F_n^M), \qquad (5.3)$$

where $X_1, \cdots, X_n$ and $Y_1, \cdots, Y_n$ are n independent samples sorted in ascending order with CDF F and F^M, respectively.

We numerically calculate $\rho_1(F_n, F_n^M)$ to compare several Langevin models against the MC model. We consider the following models: "Fox94" denotes the original model proposed by Fox and Lu (1994), which requires

a square root decomposition ($S = \sqrt{D}$) for each step in the simulation (see equations 1.1 to 1.3). "Fox97" is the widely used "subunit model" of Fox (1997); see equations 1.4 and 1.5. "Goldwyn" denotes the method taken from Goldwyn and Shea-Brown (2011), where they restrict the 14D system (V, 5 K$^+$ gates and 8 Na$^+$ gates) to the 4D multinomial submanifold (V, m, n, and h; see section 4.2), with gating variables truncated to [0,1]. We write "Orio" for the model proposed by Orio and Soudry (2012), where they constructed a rectangular matrix S such that $SS^\mathsf{T} = D$ (referred to as S_{paired} in Table 3) combining fluctuations driven by pairs of reciprocal edges, thereby avoiding taking matrix square roots at each time step. The model "Dangerfield" represents Dangerfield et al. (2012), which used the same S matrix as in Orio and Soudry (2012) but added a reflecting (no-flux) boundary condition via orthogonal projection (referred to as S_{EF} in Table 3). Finally, we include the 14×28D model we proposed in section 3.2, or "14D" (referred to as S_{single} in Table 3); "SS" is the stochastic shielding model specified in section 3.3.

For each model, we ran 10,000 independent samples of the simulation, holding channel number, injected current ($I_{\text{app}} = 10$ nA), and initial conditions fixed. Throughout the article, we presume a fixed channel density of 60 channels/μm^2 for sodium and 18 channels/μm^2 for potassium in a membrane patch of area 100 μm^2, consistent with prior work such as Goldwyn and Shea-Brown (2011) and Orio and Soudry (2012). The initial condition is taken to be the point on the deterministic limit cycle at which the voltage crosses upward through -60 mV. An initial transient corresponding to 10 to 15 ISIs is discarded, to remove the effects of the initial condition. (See Table 5 in appendix B for a complete specification of simulation parameters.) We compared the efficiency and accuracy of each model through the following steps:

1. For each model, a single run simulates a total time of 84,000 milliseconds (ms) with time step 0.008 ms, recording at least 5000 ISIs.
2. For each model, repeat 10,000 runs in step 1.
3. Create a reference ISI distribution by aggregating all 10,000 runs of the MC model, that is, based on roughly 5×10^7 ISIs.
4. For each of 10^4 individual runs, align all ISI data into a single vector and calculate the ECDF using equation 5.1.
5. Compare the ISI distribution of each model with the reference MC distribution by calculating the L_1-difference of the ECDFs using equation 5.3.
6. To compare the computational efficiency, we take the average execution time of the MC model as the reference. The relative computational efficiency is the ratio of the average execution time of a model with that of the MC model (about 3790 sec).

Table 3 gives the empiricially measured L_1 difference in ISI distribution between several pairs of models.[7] The first row ("MC") gives the average L_1 distance between individual MC simulations and the reference distribution generated by aggregating all MC simulations, in order to give an estimate of the intrinsic variability of the measure. Figure 8 plots the L_1-Wasserstein differences versus the relative computational efficiency of several models against the MC model. These results suggest that the Fox94, Orio, and 14×28D models are statistically indistinguishable when compared with the MC model using the L_1-Wasserstein distance. This result is expected in light of our results (see section 4) showing that these three models are pathwise-equivalent. (We will make pairwise statistical comparisons between the ISI distributions of each model in see section 5.2.) Among these equivalent models, however, the 14×28D and Orio models are significantly faster than the original Fox94 model (and the Goldwyn model) because they avoid the matrix square root computation. The Dangerfield model has speed similar to the 14×28D model, but the use of reflecting boundary conditions introduces significant inaccuracy in the ISI distribution. The imposition of truncating boundary conditions in the Goldwyn model also appears to affect the ISI distribution. Of the models considered, the Fox97 subunit model is the fastest; however, it makes a particularly poor approximation to the ISI distribution of the MC model. Note that the maximum L_1-Wasserstein distance between two distributions is 2. The ISI distribution of the Fox97 subunit model to that of the MC model is more than 0.8, which is 10 times larger than the L_1-Wasserstein distance of the SS model, and almost half of the maximum distance. As shown in Figure 7, the Fox97 subunit model fails to achieve the spike firing threshold and produces longer ISIs. Because of its inaccuracy, we do not include the subunit model in our remaining comparisons. The stochastic shielding model, on the other hand, has nearly the same speed as the Fox97 model, but it is over 100 times more accurate (in the L_1 sense). The SS model is an order of magnitude faster than the 14×28D model and has less than twice the L_1 discrepancy versus the MC model (L_1 norm 76.2 versus 49.3 microseconds). While this difference in accuracy is statistically significant, it may not be practically significant, depending on the application (see section 6 for further discussion of this point).

5.2 Two-Sample Kolmogorov-Smirnov Test. In addition to using the L_1-Wasserstein distances to test the differences between two CDFs, we can also make a pairwise comparison between each model by applying the Dvoretzky-Kiefer-Wolfowitz inequality (Dvoretzky, Kiefer, & Wolfowitz,

---

[7]Run times in Table 3, rounded to the nearest integer number of seconds, were obtained by averaging the run times on a distribution of heterogeneous compute nodes from Case Western Reserve University's high-performance computing cluster.
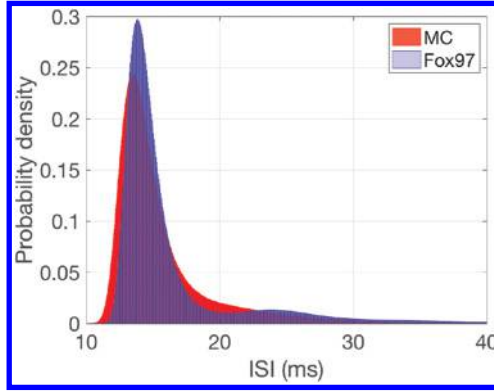
Figure 7: The probability density of interspike intervals (ISIs) for Fox97 (blue) and the MC model (red). The probability densities were calculated over more than 5.4×10^7 ISIs.
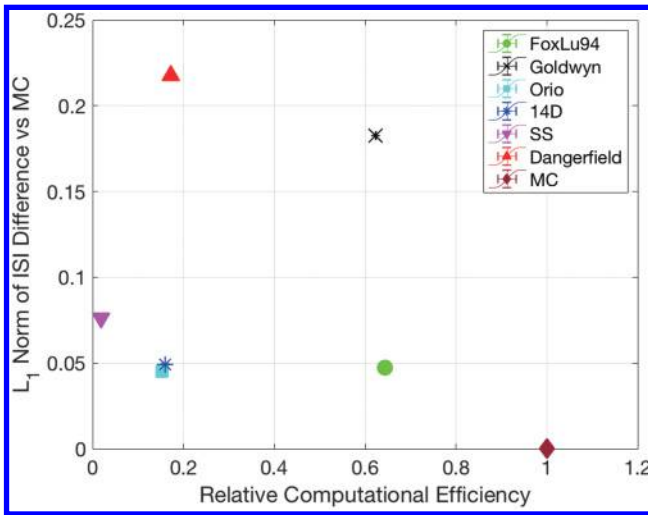

Figure 8: The L_1-Wasserstein distances and relative computational efficiency versus the MC model. Fox94 (green circle), Goldwyn (black cross), Orio (cyan square), 14D (blue star), SS (magenta downward-pointing triangle), Dangerfield (red upward-pointing triangle), and the MC (brown diamond) model. The L_1 error for ISI distribution was computed using the L_1-Wasserstein distance, equation 5.3, with discrete time Gillespie/Monte Carlo simulations as a reference. The relative computational efficiency is the ratio of the recorded run time to the mean recorded time of the MC mode (3790 seconds). The mean and 95% confidence intervals were calculated using 100 repetitions of 10,000 runs each (5×10^9 ISIs total).

1956) and the two-sample Kolmogorov-Smirnov (KS) test (Kolmogorov, 1933; Smirnov, 1948). While the Wasserstein distance is based on the L_1 norm, the KS statistic is based on the L_∞ (or supremum) norm.

The Dvoretzky-Kiefer-Wolfowitz inequality (Dvoretzky et al., 1956) establishes confidence bounds for the CDF. Specifically, the interval that contains the true CDF, $F(\cdot)$, with probability $1 - \alpha$, is given by

$$|F_n(x) - F(x)| \leq \varepsilon \ \text{ where } \varepsilon = \sqrt{\frac{\ln \frac{2}{\alpha}}{2n}}. \tag{5.4}$$

When comparing samples $X_1^M, X_2^M, \ldots, X_n^M$ obtained from an approximate model M against the gold standard, in section 5.1 we computed the L_1 difference of the empirical density functions as an approximation for the L_1 difference of the true distributions. Instead, we work here with the L_∞ norm:

$$\rho_\infty(F_n, F_n^M) = \lim_{p \to \infty} \left(\int_0^\infty \left| F_n^M(x) - F_n(x) \right|^p \, dx \right)^{1/p}$$

$$= \sup_{0 \leq x < \infty} \left(\left| F_n^M(x) - F_n(x) \right| \right). \tag{5.5}$$

For each $x \geq 0$, equation 5.4 bounds the discrepancy between the true and empirical distribution differences as follows. By the triangle inequality and independence of the X_i from the X_i^M, the inequality

$$|F^M - F| = |F^M - F_n^M + F_n - F + F_n^M - F_n|$$

$$\leq |F^M - F_n^M| + |F_n - F| + |F_n^M - F_n|$$

$$\leq 2\varepsilon + |F_n^M - F_n| \tag{5.6}$$

holds with probability $(1 - \alpha)^2$. Similarly,

$$|F_n^M - F_n| = |F_n^M - F^M + F - F_n + F^M - F|$$

$$\leq |F^M - F_n^M| + |F_n - F| + |F^M - F|$$

$$\leq 2\varepsilon + |F^M - F| \tag{5.7}$$

holds with probability $(1 - \alpha)^2$. Together, equations 5.6 and 5.7 indicate that the discrepancy between the difference of empirical distributions and the difference of true distributions is bounded as

$$\left| |F^M - F| - |F_n^M - F_n| \right| \leq 2\varepsilon \tag{5.8}$$

with probability $(1 - \alpha)^2$, for $\varepsilon = \sqrt{\frac{\ln \frac{2}{\alpha}}{2n}}$.

We will use the pointwise difference of the ECDFs for a large sample as an estimate for the pointwise difference between two true CDFs. The two-sample Kolmogorov-Smirnov (KS) test (Kolmogorov, 1933; Smirnov, 1948) offers a statistics to test whether two samples are from the same distribution. The two-sample KS statistic is

$$D_{n,m} = \sup_x |F_{1,n}(x) - F_{2,m}(x)|, \tag{5.9}$$

where $F_{1,n}$ and $F_{2,m}$ are two ECDFs for two samples defined in equation 5.1 and the sup is the supremum function. The reference statistic, $R_{n,m}(\alpha)$, depending on the significance level α, is defined as

$$R_{n,m}(\alpha) = \sqrt{\frac{-\log(\alpha/2)}{2}} \sqrt{\frac{n+m}{nm}}, \tag{5.10}$$

where n and m are the sample sizes. The null hypothesis that "the two samples come from the same distribution" is rejected at the significance level α if

$$D_{n,m} > R_{n,m}(\alpha). \tag{5.11}$$

Figure 9 plots the logarithm of ratio of the two-sample KS statistics, $\frac{D_{n,m}}{R_{n,m}(0.01)}$, for Fox94 (Fox & Lu, 1994), Goldwyn (Goldwyn & Shea-Brown, 2011), Dangerfiled (Dangerfield et al., 2012), Orio (Orio & Soudry, 2012), 14D (the $14 \times 28D$ model we proposed in section 3.2). Data of "self-comparison" (e.g. Fox94 versus Fox 94) was obtained by comparing two ISI ECDFs from independent simulations. As shown in Figure 9, models that we previously proved were pathwise equivalent in section 4, namely, the Fox94, Orio, and 14D models, are not distinguishable at any confidence level justified by our data. Note that those three models use the same boundary conditions (free boundary condition as in Orio & Soudry, 2012) and the ratio $D_{n,m}/R_{n,m}(\alpha)$ of pairwise comparison is on the same magnitude of that for the self-comparisons. The models of Fox and Lu and Orio and Soudry, and our 14D model generate indistinguishable ISI distributions but are distinguishable from Dangerfield's model and Fox's 97 model. Thus, as the bottom panel of Figure 9 shows, Fox94, Orio, and $14 \times 28D$ form a block of statistically indistinguishable samples. However, as pointed out above, these statistically equivalent simulation algorithms have different computational efficiencies (see Figure 8). Among these methods, Orio and Soundry's algorithm (14-dimensional state space with 14 undirected edges as noise sources) and our method (14-dimensional state space with 28 directed edges as noise sources) have similar efficiencies, with Orio's method being about 5% faster than our method. Our $14 \times 28D$ method provides the additional
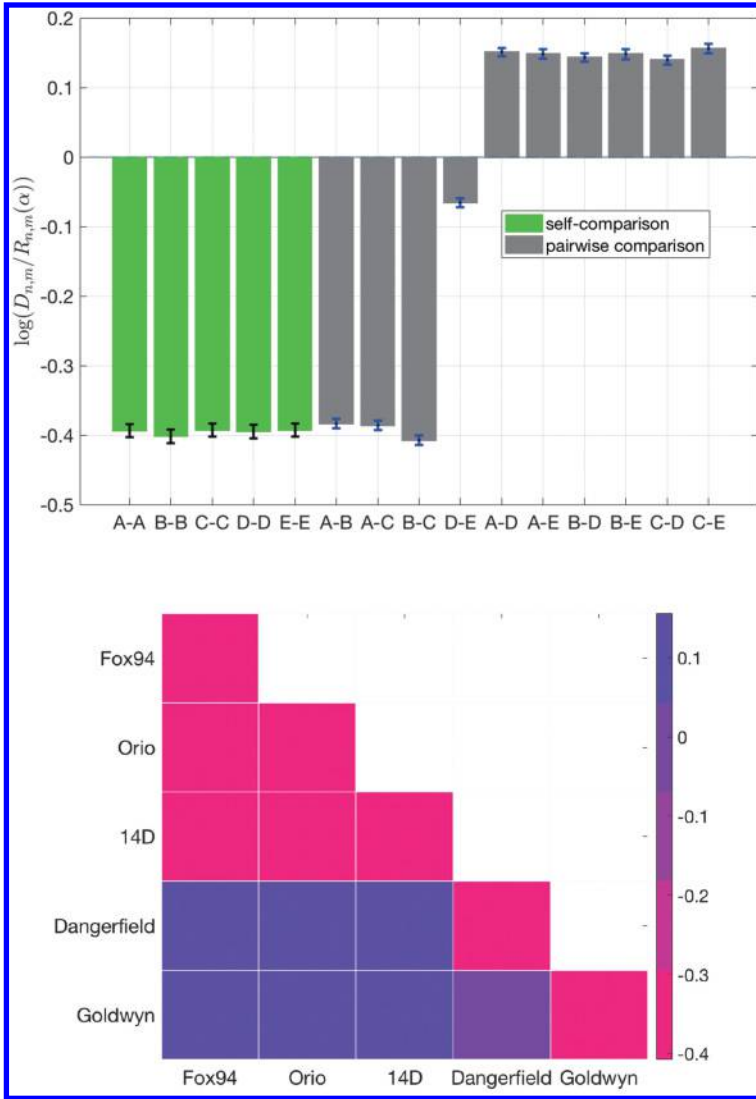
Figure 9: Logarithm of the ratio of Kolmogorov-Smirnov test statistic $D_{n,m}/R_{n,m}(\alpha)$, equations 5.9 and 5.10, for samples from the ISI distribution for each pair of models. (Top) Box and whisker plots showing mean and 95% confidence intervals based on 10,000 pairwise comparisons. The first five plots show self-comparisons (green bars); the remainder compare distinct pairs (gray bars). A: Fox94 (Fox & Lu, 1994), B: Orio (Orio & Soudry, 2012), C: 14D (14×28D model, section 3.2), D: Dangerfield (Dangerfield et al., 2012), E: Goldwyn (Goldwyn & Shea-Brown, 2011). (Bottom) Mean logarithms (as in top panel) for all pairwise comparisons.

advantage that it facilitates further acceleration under the stochastic shielding approximation (see section 6).

In contrast to the statistically equivalent Orio, $14 \times 28D$, and Fox94 models, algorithms using different boundary conditions are not pathwise equivalent, which is again verified in Figure 9. Algorithms with subunit approximation and truncated boundary condition (i.e., "Goldwyn") and the reflecting boundary condition (i.e., "Dangerfield") are significantly different in accuracy (and, in particular, they are less accurate) than models in the $14 \times 28D$ class.

6 Discussion and Conclusions

6.1 Summary. The exact method for Markov chain (MC) simulation for an electrotonically compact (single compartment) conductance-based stochastic model under current clamp is a hybrid discrete (channel state)/continuous (voltage) model of the sort used by Clay and DeFelice (1983), Newby et al. (2013), and Anderson et al. (2015). While MC methods are computationally expensive, simulations based on gaussian/Langevin approximation can capture the effects of stochastic ion channel fluctuations with reasonable accuracy and excellent computational efficiency. Since Goldwyn and Shea Brown's work focusing the attention of the computational neuroscience community on Fox and Lu's Langevin algorithm for the Hodgkin-Huxley system (Fox & Lu, 1994; Goldwyn & Shea-Brown, 2011), several variants of this approach have appeared in the literature.

In this article, we advocate for a class of models combining the best features of conductance-based Langevin models with the recently developed stochastic shielding approximation (Schmandt & Galán, 2012; Schmidt & Thomas, 2014; Schmidt et al., 2018). We propose a Langevin model with a 14-dimensional state space, representing the voltage, five states of the K^+ channel, and eight states of the Na^+-channel; and a 28-dimensional representation of the driving noise: one independent gaussian noise term for each directed edge in the channel-state transition graph. We showed in section 2 that the corresponding mean-field 14D ordinary differential equation model is consistent with the classical HH equations in the sense that the latter correspond to an invariant submanifold of the higher-dimensional model, to which all trajectories converge exponentially quickly. Figure 2 illustrated the relation between the deterministic 4D and 14D Hodgkin-Huxley systems. Building on this framework, we introduced the $14 \times 28D$ model, with independent noise sources corresponding to each ion channel transition (see section 3). We proved in section 4 that given identical boundary conditions, our $14 \times 28D$ model is pathwise equivalent to Fox and Lu's original Langevin model and to a 14-state model with 14 independent noise sources due to Orio and Soudry (2012).

The original 4D HH model, the 14D deterministic HH model, and the family of equivalent 14D Langevin models we consider here form a nested

family, each contained within the next. Specifically, the 14D ODE model is the mean-field version of the 14D Langevin model, and the 4D ODE model forms an attracting invariant submanifold within the 14D ODE model, as we establish in our lemma 2. So in a very specific sense, the original HH equations "live inside" the 14D Langevin equations. Thus, these three models enjoy a special relationship. In contrast, the 4D Langevin equations studied in Fox (1997) do not bear an especially close relationship to the other three.

In addition to rigorous mathematical analysis, we also performed numerical comparisons (see section 5) showing that, as expected, the pathwise equivalent models produced statistically indistinguishable interspike interval (ISI) distributions. Moreover, the ISI distributions for our model (and its equivalents) were closer to the ISI distribution of the gold standard MC model under two different metric space measures. Our method (along with Orio and Soudry's) proved computationally more efficient than Fox and Lu's original method and Dangerfield's model (Dangerfield et al., 2012). In addition, our method lends itself naturally to model reduction (via the stochastic shielding approximation) to a significantly faster 14×6D simulation that preserves a surprisingly high level of accuracy.

6.2 Discrete Gillespie Markov Chain Algorithms. The discrete-state MC algorithm due to Gillespie is often taken to be the gold standard simulation for a single-compartment stochastic conductance-based model. Most former literature on Langevin HH models (Goldwyn & Shea-Brown, 2011; Linaro et al., 2011; Orio & Soudry, 2012; Huang et al., 2013), when establishing a reference MC model, consider a version of the discrete Gillespie algorithm that assumes a piecewise-constant propensity approximation, that is, it does not take into account that the voltage changes between transitions, which changes the transition rates. This approximation can lead to biophysically unrealistic voltage traces for very small system sizes (see Figure 2 of Kispersky & White, 2008; top trace with $N = 1$ ion channel) although the differences appear to be mitigated for $N \gtrsim 40$ channels (Anderson et al., 2015). In this article our MC simulations are based on 6000 Na^+ and 1800 K^+ channels (as in Goldwyn & Shea-Brown, 2011), and we too use the ISI distribution generated by a piecewise-constant propensity MC algorithm as our reference distribution. As shown in Table 3 and Figure 8, the computation time for MC is one order of magnitude larger than efficient methods such as Orio and Soudry (2012) and Dangerfield et al. (2012) and the 14×28D model. The computational cost of the MC model increases dramatically as the number of ion channels grows; therefore, even the approximate MC algorithm is inapplicable for a large number of channels.

6.3 Langevin Models. It is worth pointing out that the accuracy of Fox and Lu's original Langevin equations has not been fully appreciated. In fact, Fox and Lu's model (Fox & Lu, 1994) gives an approximation to the MC

model that is just as accurate as that of Orio and Soudry (2012) both in the gating variable statistics (Goldwyn & Shea-Brown, 2011) and the ISI distribution sense (see section 5), because, as we have established, these models are pathwise equivalent! However, the original implementation requires taking a matrix square root in every time step in the numerical simulation, which significantly reduces its computational efficiency.

Models based on modifications of Fox and Lu's (1994) work can be divided into three classes: the subunit model (Fox, 1997), effective noise models (Linaro et al., 2011; Güler, 2013b), and channel-based Langevin models (e.g., Goldwyn & Shea-Brown, 2011; Dangerfield et al., 2012; Orio & Soudry, 2012; Pezo et al., 2014; Huang et al., 2013).

6.3.1 Subunit Model. The first modification of Fox and Lu's model is the subunit model (Fox, 1997), which keeps the original form of the HH model, and adds noise to the gating variables (m, h, and n) (Fox, 1997; Goldwyn & Shea-Brown, 2011). The subunit approximation model was widely used because of its fast computational speed. However, as Bruce (2009) and others pointed out, the inaccuracy of this model remains significant even for large number of channels. Moreover, Goldwyn and Shea-Brown (2011) and Huang et al. (2015) found that the subunit model fails to capture the statistics of the HH Na^+ and K^+ gates. In this article, we also observed that the subunit model is more efficient than channel-based Langevin models but tends to delay spike generation. As shown in Figure 7, the subunit model generates significantly longer ISIs than the MC model.

6.3.2 Effective Noise Models. Another modification to Fox and Lu's algorithm is to add colored noise to the channel open fractions. Though colored noise models such as those of Linaro et al. (2011) and Güler (2013b) are not included in our model comparison, Huang et al. (2015) found that both of these effective noise models generate shorter ISIs than the MC model with the same parameters. Though the comparison we provided in section 5 only include the Fox and Lu 94, Fox97, Goldwyn, Dangerfield, Orio, SS, and the $14 \times 28D$ models, combining the results from Goldwyn and Shea-Brown (2011) and Huang et al. (2015), the $14 \times 28D$ model could be compared to a variety of models (e.g., Fox & Lu, 1994; Fox, 1997; Goldwyn & Shea-Brown, 2011; Linaro et al., 2011; Orio & Soudry, 2012; Dangerfield et al., 2012; Huang et al., 2013; Güler, 2013b).

6.3.3 Channel-based Langevin Models. The main focus of this article is the modification based on the original Fox and Lu's matrix decomposition method, namely, the channel-based (or conductance-based) Langevin models. We proved in section 4 that under the same boundary conditions, Fox and Lu's original model, Orio's model, and our $14 \times 28D$ model are pathwise equivalent, which was also verified from our numerical simulations in sections 4 and 5. In section 4 we discussed channel-based Langevin

models (Fox & Lu, 1994; Goldwyn & Shea-Brown, 2011; Dangerfield et al., 2012; Orio & Soudry, 2012; Fox, 2018). We excluded Fox's more recent implementation (Fox, 2018) in section 5 for two reasons. First, the algorithm is pathwise equivalent to others considered there. Moreover, the method is vulnerable to numerical instability when the Cholesky decomposition is performed. Specifically, some of the elements in the S matrix from the Cholesky decomposition in Fox (2018) involve square roots of differences of several quantities, with no guarantee that the differences will result in nonnegative terms—even with strictly positive value of the gating variables. Nevertheless, this model would be in the equivalence class and in any case would not be more efficient than the Orio model because of the noise dimension and complicated operations (involving taking multiple square roots) in each step.

6.4 Model Comparisons. If two random variables have similar distributions, then they will have similar moments, but not vice versa. Therefore, comparison of the full interspike-interval distributions produced by different simulation algorithms gives a more rigorous test than comparison of first and second moments of the ISI distribution. Most previous evaluations of competing Langevin approximations were based on the accuracy of low-order moments, for example, the mean and variance of channel state occupancy under voltage clamp, or the mean and variance of the interspike interval distribution under current clamp (Goldwyn et al., 2011; Goldwyn & Shea-Brown, 2011; Schmandt & Galán, 2012; Linaro et al., 2011; Dangerfield et al., 2012; Orio & Soudry, 2012; Huang et al., 2013, 2015). Here, we compare the accuracy of the different algorithms using the full ISI distribution but using the L_1 norm of the difference (Wasserstein metric) and the L_∞ norm (Kolmogorov-Smirnov test). Greenwood, McDonnell, and Ward (2015) previously compared the ISI distributions generated by the Markov chain (Gillespie algorithm) to the distribution generated by different types of Langevin approximations (LA), including the original Langevin models (Fox & Lu, 1994; Goldwyn & Shea-Brown, 2011), the channel-based LA with colored noise (Linaro et al., 2011; Güler, 2013b), and LA with a 14×14 variant of the diffusion coefficient matrix S (Orio & Soudry, 2012). They concluded that Orio and Soudry's method provided the best match to the Markov chain model, specifically, "Fox-Goldwyn, and Orio-Kurtz[8] methods both generate ISI histograms very close to those of "Micro."[9] (Greenwood et al., 2015). We note that the comparison reported in this article simply superimposed plots of the ISI distributions, allowing a qualitative comparison, while our metric-space analysis is fully quantitative. In any case, their conclusions are consistent with our findings; we showed in section 4 that the Fox-Goldwyn and the Orio-Kurtz model

---

[8] We refer to this model as "Orio."
[9] This is the model we refer to as the Markov chain model.

are pathwise equivalent (when implemented with the same boundary conditions), which accounts for the similarity in the ISI histograms they generate. In fact, because of pathwise equivalence, we can conclude that the true distributions for these models are identical, and any differences observed just reflect finite sampling.

6.5 Stochastic Shielding Method. The stochastic shielding (SS) approximation (Schmandt & Galán, 2012) provides an efficient and accurate method for approximating the Markov process with only a subset of observable states. For conductance-based models, rather than aggregating ion channel states, SS affects dimension reduction by selectively eliminating those independent noise sources that have the least impact on current fluctuations. Recent work in (Pezo, Soudry, & Orio, 2014) compared previous methods (Gillespie, 1977; Orio & Soudry, 2012; Dangerfield et al., 2012; Huang et al., 2013; Schmandt & Galán, 2012) in accuracy, applicability, and simplicity, as well as computational efficiency. They concluded that for mesoscopic numbers of channels, stochastic shielding methods combined with diffusion approximation methods can be an optimal choice. Like Orio and Soudry (2012), the stochastic shielding method proposed by Pezo et al. (2014) also assumed a detailed balance of transitions between adjacent states and used edges that are directly connected to the open gates of HH Na^+ and K^+. We calculated the edge importance in section 3.3 and found that the 4 (out of 20) most important directed edges for the Na^+ gates are not the 4 edges directly connected to the conducting state, as assumed in previous application of the SS method (Schmandt & Galán, 2012).

6.6 Which Model to Use? Among all modifications of Fox and Lu's method considered here, Orio and Soudry's approach and our $14 \times 28D$ model provide the best approximation to the gold standard MC model, with the greatest computational efficiency. Several earlier models were studied in the review paper by Goldwyn and Shea-Brown (2011), where they rediscovered that the original Langevin model proposed by Fox and Lu is the best approximation to the MC model among those considered. Later work (Huang et al., 2015) further surveyed a wide range of Langevin approximations for the HH system (Fox & Lu, 1994; Fox, 1997; Goldwyn & Shea-Brown, 2011; Linaro et al., 2011; Güler, 2013b; Orio & Soudry, 2012; Huang et al., 2013) and explored models with different boundary conditions. Huang et al. (2015) concluded that the bounded and truncated-restored Langevin model (Huang et al., 2013) and the unbounded model (Orio & Soudry, 2012) provide the best approximation to the MC model.

As shown in sections 4 and 5, the $14 \times 28D$ Langevin model naturally derived from the channel structure is pathwise equivalent to the Fox94, Fox18, and Orio-Soudry models under the same boundary conditions. The $14 \times 28D$ model is more accurate than the reflecting boundary condition method of Dangerfield et al. (2012) and also better than the approximation

method proposed by Goldwyn and Shea-Brown (2011) when the entire ISI distribution is taken into account. We note that Huang et al. (2015) treated Goldwyn's method (Goldwyn & Shea-Brown, 2011) as the original Fox and Lu model in their comparison; however, the simulation in Goldwyn and Shea-Brown (2011) uses the 4D multinomial submanifold to update gating variables. Our analysis and numerical simulations suggest that the original Fox and Lu model is indeed as accurate as the Orio-Soudry model, although the computational cost remains a major concern.

Though the 14×28D model has similar efficiency and accuracy as that of Orio and Soudry (2012), it has several advantages. First, the rectangular S matrix (in equations 1.2 and 1.3) in Orio's model merges the noise contributions of reciprocal pairs of edges. However, this dimension reduction assumes, in effect, that detailed balance holds along reciprocal edges, which our results show is not the case, under current clamp (see Figure 5). Moreover, the 14×28D model arises naturally from the individual transitions of the exact evolution equations (see equations 3.1 and 3.2) for the underlying Markov chain model, which makes it conceptually easier to understand. In addition, our method for defining the 14×28D Langevin model and finding the best SS model extends to channel-based models with arbitrary channel gating schemes beyond the standard HH model. Given any channel state transition graph, the Langevin equations may be read off from the transitions, and the edge importance under current clamp can be evaluated by applying the stochastic shielding method to investigate the contributions of noise from each individual directed edge. Finally, in exchange for a small reduction in accuracy, the stochastic shielding method affords a significant gain in efficiency. The 14×28D model thus offers a natural way to quantify the contributions of the microscopic transitions to the macroscopic voltage fluctuations in the membrane through the use of stochastic shielding. For general ion channel models, extending a biophysically based Langevin model analogous to our 14×28D HH model, together with the stochastic shielding method, may provide the best available tool for investigating how unobservable microscopic behaviors (such as ion channel fluctuations) affect the macroscopic variability in many biological systems.

6.7 Limitations. All Langevin models, including our proposed 14×28D model, proceed from the assumption that the ion channel population is large enough (and the ion channel state transitions frequent enough) that the gaussian approximations by which the white noise forcing terms are derived are justified. Thus, when the system size is too small, no Langevin system will be appropriate. Fortunately the Langevin approximation appears quite accurate for realistic population sizes.

The 14×28D model uses more noise sources than other approaches. However, stochastic shielding allows us to jettison noise sources that do not significantly affect the system dynamics (the voltage fluctuations and

ISI distribution). Moreover, in order to compare the ISI distribution in detail among several variants of the Fox94′ model versus the Markov chain standard, we have considered a single value of the driving current, while other studies have compared parameterized responses such as the firing rate, ISI variance, or other moments, as a function of applied current. Accurate comparisons require large ensembles of independent trajectories, forcing a trade-off between precision and breadth; we opted here for precise comparisons at a representative level of the driving current.

From a conceptual point of view, a shortcoming of most Langevin models is the tendency for some channel state variables x to collide with the domain boundaries $x \in [0, 1]$ and to cross them during numerical simulations with finite time steps. We adopted the approach advocated by Orio and Soudry (2012) of using "free boundaries" in which gating variables can make excursions into the (unphysical) range $x < 0$ or $x > 1$. Practically, these excursions are always short if the time step is reasonably small, as they tend to be self-correcting.[10] Another approach is to construct reflecting boundary conditions; different implementations of this idea were used in Dangerfield et al. (2010), Fox and Lu (1994), and Schmid, Goychuk, and Hänggi (2001). Dangerfield's method proved both slower and less accurate than the free boundary method. As an alternative method, one uses a biased rejection sampling approach, testing each gating variable of the 14D model on each time step and repeating the noise sample for any time step violating the domain conditions (Fox & Lu, 1994; Schmid et al., 2001). We found that this method had accuracy similar to that of Dangerfield's method (L_1-Wasserstein difference $\approx$ 4.4e-1 msec; see Table 3) and run time similar to that of the Fox94 implementation, about four times slower than our 14D Langevin model.

Yet another approach that in principle can guarantee that the stochastic process remains within proscribed bounds, rederives a "best diffusional approximation" Fokker-Planck equation based on matching a master-equation birth-and -death description of a (binomial) population of two-state ion channels, leading to modified drift and diffusion terms (Goychuk, 2014). This method does not appear to extend readily to the 14D setting, with underlying multinomial structure of the ion channel gates, so we do not dwell on it further.

Table 3 gives the accuracies with which each model reproduces the ISI distribution, compared to a standard reference distribution generated through a large number of samples of the MC method. The mean L_1 difference between a single sample and the reference sample is about 0.227 microseconds. For a nonnegative random variable $T \geq 0$, the difference in the mean under two probability distributions is bounded above by the L_1

---

[10]To avoid complex entries, we use $|x|$ when calculating entries in the noise coefficient matrix.

difference in their cumulative distribution functions.[11] Thus, the L_1 norm gives an idea of the temporal accuracy with which one can approximate a given distribution by another. The mean difference between the ISI distribution generated by a single run of the full $14 \times 28D$ model is about 49 μsec, and the discrepancy produced by the (significantly faster) SS model is about 76 μsec. When would this level of accuracy matter for the function of a neuron within a network? The barn owl *Tyto alba* uses interaural time difference to localize its prey to within 1 to 2 degrees, a feat that requires encoding information in the precise timing of auditory system action potentials at the scale of 5 to 20 microseconds (Moiseff & Konishi, 1981; Gerstner, Kempter, Van Hemmen, & Wagner, 1996). For detailed studies of the effects of channel noise in this system, the superior accuracy of the MC model might be preferred. On the other hand, the timescale of information encoding in the human auditory nerve is thought to be in the millisecond range (Goldwyn, Shea-Brown, & Rubinstein, 2010), with precision in the feline auditory system reported as low as 100 μs (Imennov & Rubinstein, 2009; see also Woo, Miller, & Abbas, 2009). For these and other mammalian systems, the stochastic shielding approximation should provide sufficient accuracy.

6.8 Future Work. In section 3.3 we compared the ISI variance when noise was included one edge at a time and found that the edges making the greatest contribution to population fluctuations under voltage clamp were not identical to the edges having the largest effect on ISI variance when taken one at a time. However, the ISI, considered as a random variable determined through a first-passage time process, depends on the entire trajectory, not just on the occupancy of the conducting states. The HH dynamics are strongly nonlinear, producing a limit cycle in the deterministic case, and it is not immediately clear whether the effects of channel noise on ISI variability should be additive. In future work, we plan to address the question of the additive contribution of individual/molecular noise sources on ISI variability.

A principal motivation for using the stochastic shielding algorithm is to develop fast and accurate algorithms for ensemble simulations of forward models for parameter estimation in a data assimilation framework. We expect that our method may prove useful for such studies based on current-clamp electrophysiological data in the future.

---

[11] For a nonnegative random variable T with cumulative distribution function $F(t) = \mathbb{P}[T \leq t]$, the mean satisfies $E[T] = \int_0^\infty (1 - F(t)) \, dt$ (Grimmett & Stirzaker, 2005). Therefore, the difference in mean under two distributions F_1 and F_2 satisfies $|E_1[T] - E_2[T]| = \left| \int_0^\infty F_1(t) - F_2(t) \, dt \right| \leq \rho_1(F_1, F_2)$.

Appendix A: Table of Common Symbols and Notations

Table 4: Common Symbols and Notations in This Article.

Symbol	Meaning
C	Membrane capacitance (μF/cm^2)
υ	Membrane potential (mV)
V_{Na}, V_K, V_L	Ionic reversal potential for Na$^+$, K$^+$ and leak (mV)
I_{app}	Applied current to the membrane (nA/cm^2)
m, h, n	Dimensionless gating variables for Na$^+$ and K$^+$ channels
$\alpha_x, \beta_x \ x \in \{m, n, h\}$	Voltage-dependent rate constant (1/msec)
x	Vector of state variables
$M = [M_1, M_2, \cdots, M_8]$	Eight-component state vector for the Na$^+$ gates
$[m_{00}, m_{10}, m_{20}, m_{30},$	Components for the Na$^+$ gates
$\ m_{01}, m_{11}, m_{21}, m_{31}]^\mathsf{T}$	
$N = [N_1, N_2, \cdots, N_5]$	Five-component state vector for the K$^+$ gates
$[n_0, n_1, n_2, n_3, n_4]^\mathsf{T}$	Components for the K$^+$ gates
$M_{tot}, \ N_{tot}$	Total number of Na$^+$ and K$^+$ channels
$\mathcal{X}$	4-dimensional manifold domain for 4D HH model
$\mathcal{Y}$	14-dimensional manifold domain for 14D HH model
Δ^k	k-dimensional simplex in $\mathbb{R}^{k+1}$ given by
	$\quad y_1 + \ldots + y_{k+1} = 1, y_i \geq 0$
$\mathcal{M}$	Multinomial submanifold within the 14D space
A_{Na}, A_K	State-dependent rate matrix
D	State diffusion matrix
S, S_1, S_2, S_{Na}, S_K	Noise coefficient matrices
ξ	Vector of independent δ-correlated gaussian white noise with zero mean and unit variance
$\mathbf{X} = [X_1, X_2, \ldots, X_d]$	A d-dimensional random variable for sample path
$\mathbf{W} = [W_1, W_2, \ldots, W_n]$	A Wiener trajectory with n components
$\delta(\cdot)$	The Dirac delta function
δ_{ij}	The Kronecker delta
F_n	Empirical cumulative distribution function with n observations (in section 5, we use $m, \ n$ as sample sizes)

Table 5: Parameters Used for Simulations in This Article.

Symbol	Meaning	Value
C	Membrane capacitance	$1 \ \mu$F/cm^2
$\bar{g}_{Na}$	Maximal sodium conductance	$120 \ \mu$S/cm^2
$\bar{g}_K$	Maximal potassium conductance	$36 \ \mu$S/cm^2
g_{leak}	Leak conductance	$0.3 \ \mu$S/cm^2
V_{Na}	Sodium reversal potential for Na$^+$	50 mV

Table 5: Continued.

Symbol	Meaning	Value
V_K	Potassium reversal potential for K^+	−77 mV
V_{leak}	Leak reversal potential	−54.4 mV
I_{app}	Applied current to the membrane	10 nA/cm^2
$\mathcal{A}$	Membrane area	100 μm^2
M_{tot}	Total number of Na$^+$ channels	6,000
N_{tot}	Total number of K$^+$ channels	18,00

Appendix B: Parameters and Transition Matrices

Subunit kinetics for the Hodgkin and Huxley equations are given by

$$\alpha_m(v) = \frac{0.1(25 - v)}{\exp(2.5 - 0.1v) - 1}, \tag{B.1}$$

$$\beta_m(v) = 4\exp(-v/18), \tag{B.2}$$

$$\alpha_h(v) = 0.07\exp(-v/20), \tag{B.3}$$

$$\beta_h(v) = \frac{1}{\exp(3 - 0.1v) + 1}, \tag{B.4}$$

$$\alpha_n(v) = \frac{0.01(10 - v)}{\exp(1 - 0.1v) - 1}, \tag{B.5}$$

$$\beta_n(v) = 0.125\exp(-v/80) \tag{B.6}$$

$$A_K(v) = \begin{bmatrix} D_K(1) & \beta_n(v) & 0 & 0 & 0 \\ 4\alpha_n(v) & D_K(2) & 2\beta_n(v) & 0 & 0 \\ 0 & 3\alpha_n(v) & D_K(3) & 3\beta_n(v) & 0 \\ 0 & 0 & 2\alpha_n(v) & D_K(4) & 4\beta_n(v) \\ 0 & 0 & 0 & \alpha_n(v) & D_K(5) \end{bmatrix},$$

$$A_{Na} = \begin{bmatrix} D_{Na}(1) & \beta_m & 0 & 0 & \beta_h & 0 & 0 & 0 \\ 3\alpha_m & D_{Na}(2) & 2\beta_m & 0 & 0 & \beta_h & 0 & 0 \\ 0 & 2\alpha_m & D_{Na}(3) & 3\beta_m & 0 & 0 & \beta_h & 0 \\ 0 & 0 & \alpha_m & D_{Na}(4) & 0 & 0 & 0 & \beta_h \\ \alpha_h & 0 & 0 & 0 & D_{Na}(5) & \beta_m & 0 & 0 \\ 0 & \alpha_h & 0 & 0 & 3\alpha_m & D_{Na}(6) & 2\beta_m & 0 \\ 0 & 0 & \alpha_h & 0 & 0 & 2\alpha_m & D_{Na}(7) & 3\beta_m \\ 0 & 0 & 0 & \alpha_h & 0 & 0 & \alpha_m & D_{Na}(8) \end{bmatrix},$$

where the diagonal elements

$$D_{\text{ion}}(i) = -\sum_{j:j\neq i} A_{\text{ion}}(j, i), \qquad \text{ion} \in \{\text{Na, K}\}.$$

Appendix C: Proof of Lemma 2

For the reader's convenience, we restate this lemma.

Lemma 2. *Let $\mathcal{X}$ and $\mathcal{Y}$ be the lower-dimensional and higher-dimensional Hodgkin-Huxley manifolds given by equation 2.12, and let F and G be the vector fields on $\mathcal{X}$ and $\mathcal{Y}$ defined by equations 2.1 to 2.4 and 2.7 to 2.9, respectively. Let $H : \mathcal{X} \to \mathcal{M} \subset \mathcal{Y}$ and $R : \mathcal{Y} \to \mathcal{X}$ be the mappings given in Tables 2 and 1, respectively, and define the multinomial submanifold $\mathcal{M} = H(\mathcal{X})$. Then $\mathcal{M}$ is forward-time-invariant under the flow generated by G. Moreover, the vector field G, when restricted to $\mathcal{M}$, coincides with the vector field induced by F and the map H. That is, for all $y \in \mathcal{M}$, $G(y) = D_x H(R(y)) \cdot F(R(y))$.*

The main idea of the proof is to show that for every $y \in \mathcal{Y}$, $G(y)$ is contained in the span of the four vectors $\left\{ \frac{\partial H}{\partial x_i}(R(y)) \right\}_{i=1}^{4}$.

Proof. The map from the 4D HH model to the 14D HH model is given in Table 2 as $\{H : x \to y \mid x \in \mathcal{X}, y \in \mathcal{Y}\}$, and the map from the 14D HH model to the 4D HH model is given in Table 1 as $\{R : y \to x \mid x \in \mathcal{X}, y \in \mathcal{Y}\}$. The partial derivatives $\frac{\partial H}{\partial x}$ of the map H are given by

$$\frac{dm_{00}}{dm} = -3(1-m)^2(1-h) \qquad \frac{dm_{00}}{dh} = -(1-m)^3$$

$$\frac{dm_{10}}{dm} = 3(1-h)(3m^2 - 4m + 1) \qquad \frac{dm_{10}}{dh} = -3(1-m)^2 m$$

$$\frac{dm_{20}}{dm} = 3(1-h)(2m - 3m^2) \qquad \frac{dm_{20}}{dh} = -3(1-m)m^2$$

$$\frac{dm_{30}}{dm} = 3(1-h)m^2 \qquad \frac{dm_{30}}{dh} = -m^3$$

$$\frac{dm_{01}}{dm} = -3h(1-m)^2 \qquad \frac{dm_{01}}{dh} = (1-m)^3$$

$$\frac{dm_{11}}{dm} = 3h(3m^2 - 4m + 1) \qquad \frac{dm_{11}}{dh} = 3(1 - m)^2 m$$

$$\frac{dm_{21}}{dm} = 3h(2m - 3m^2) \qquad \frac{dm_{21}}{dh} = 3(1 - m)m^2$$

$$\frac{dm_{31}}{dm} = 3hm^2 \qquad \frac{dm_{31}}{dh} = m^3.$$

We can write $\partial H/\partial x$ in matrix form as:

$$\frac{\partial H}{\partial x} = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & -3(1 - m)^2(1 - h) & -(1 - m)^3 & 0 \\
0 & 3(1 - h)(3m^2 - 4m + 1) & -3(1 - m)^2 m & 0 \\
0 & 3(1 - h)(2m - 3m^2) & -3(1 - m)m^2 & 0 \\
0 & 3(1 - h)m^2 & -m^3 & 0 \\
0 & -3h(1 - m)^2 & (1 - m)^3 & 0 \\
0 & 3h(3m^2 - 4m + 1) & 3(1 - m)^2 m & 0 \\
0 & 3h(2m - 3m^2) & 3(1 - m)m^2 & 0 \\
0 & 3hm^2 & m^3 & 0 \\
0 & 0 & 0 & -4(1 - n)^3 \\
0 & 0 & 0 & 4(1 - n)^2(1 - 4n) \\
0 & 0 & 0 & 12n(1 - n)(1 - 2n) \\
0 & 0 & 0 & 4n^2(3 - 4n) \\
0 & 0 & 0 & 4n^3
\end{bmatrix}.$$

We write out the vector fields 2.8 and 2.9 component by component:

$$\frac{dM_1}{dt} = \beta_m M_2 + \beta_h M_5 - (3\alpha_m + \alpha_h)M_1$$

$$= -3(1 - m)^2(1 - h)\left[(1 - m)\alpha_m - m\beta_m\right]$$
$$+ (1 - m)^3 \left[h\beta_h - (1 - h)\alpha_h\right],$$

$$\frac{dM_2}{dt} = 3\alpha_m M_1 + 2\beta_m M_3 + \beta_h M_6 - (2\alpha_m + \beta_m + \alpha_h)M_2$$

$$= 3(1 - h)(3m^2 - 4m + 1)\left[(1 - m)\alpha_m - m\beta_m\right]$$
$$+ 3(1 - m)^2 m \left[h\beta_h - (1 - h)\alpha_h\right],$$

$$\frac{dM_3}{dt} = 2\alpha_m M_2 + 3\beta_m M_4 + \beta_h M_7 - (\alpha_m + 2\beta_m + \alpha_h)M_3,$$

$$= 3(1 - h)(2m - 3m^2)\left[(1 - m)\alpha_m - m\beta_m\right]$$

$$+ 3(1 - m)m^2 \left[h\beta_h - (1 - h)\alpha_h \right],$$

$$\frac{dM_4}{dt} = \alpha_m M_3 + \beta_h M_8 - (3\beta_m + \alpha_h)M_4,$$

$$= 3(1 - h)m^2 \left[(1 - m)\alpha_m - m\beta_m \right] + m^3 \left[h\beta_h - (1 - h)\alpha_h \right],$$

$$\frac{dM_5}{dt} = \beta_m M_6 + \alpha_h M_1 - (3\alpha_m + \beta_h)M_5,$$

$$= -3h(1 - m)^2 \left[(1 - m)\alpha_m - m\beta_m \right]$$
$$+ (1 - m)^3 \left[h\beta_h - (1 - h)\alpha_h \right],$$

$$\frac{dM_6}{dt} = 3\alpha_m M_5 + 2\beta_m M_7 + \alpha_h M_2 - (2\alpha_m + \beta_m + \beta_h)M_6,$$

$$= 3h(3m^2 - 4m + 1) \left[(1 - m)\alpha_m - m\beta_m \right]$$
$$- 3(1 - m)^2 m \left[h\beta_h - (1 - h)\alpha_h \right],$$

$$\frac{dM_7}{dt} = 2\alpha_m M_6 + 3\beta_m M_8 + \alpha_h M_3 - (\alpha_m + 2\beta_m + \beta_h)M_7,$$

$$= 3h(2m - 3m^2) \left[(1 - m)\alpha_m - m\beta_m \right]$$
$$- 3(1 - m)m^2 \left[h\beta_h - (1 - h)\alpha_h \right],$$

$$\frac{dM_8}{dt} = \alpha_m M_7 + \alpha_h M_4 - (3\beta_m + \beta_h)M_8,$$

$$= 3hm^2 \left[(1 - m)\alpha_m - m\beta_m \right] - m^3 \left[h\beta_h - (1 - h)\alpha_h \right],$$

$$\frac{dN_1}{dt} = \beta_n N_2 - 4\alpha_n N_1 = -4(1 - n)^3[\alpha_n(1 - n) - n\beta_n],$$

$$\frac{dN_2}{dt} = 4\alpha_n N_1 + 2\beta_n N_3 - (3\alpha_n + \beta_n)N_2$$

$$= 4(1 - n)^2(1 - 4n)[\alpha_n(1 - n) - n\beta_n],$$

$$\frac{dN_3}{dt} = 3\alpha_n N_2 + 3\beta_n N_4 - (2\alpha_n + 2\beta_n)N_3$$

$$= 12n(1 - n)(1 - 2n)[\alpha_n(1 - n) - n\beta_n],$$

$$\frac{dN_4}{dt} = 2\alpha_n N_3 + 4\beta_n N_5 - (3\alpha_n + 3\beta_n)N_4$$

$$= 4n^2(3 - 4n)[\alpha_n(1 - n) - n\beta_n],$$

$$\frac{dN_5}{dt} = \alpha_n N_4 - 4\beta_n N_5 = 4n^3[\alpha_n(1 - n) - n\beta_n].$$

By extracting common factors from the previous expressions it is clear that $G(y)$ may be written, for all $y \in \mathcal{Y}$, as

$$G(y) = \frac{-\bar{g}_{Na}M_8(V - V_{Na}) - \bar{g}_K N_5(V - V_K) - g_L(V - V_L) + I_{app}}{C} \left\{ \frac{\partial H}{\partial v}(R(y)) \right\}$$

$$+ \left[(1 - m')\alpha_m - m'\beta_m \right] \left\{ \frac{\partial H}{\partial m}(R(y)) \right\}$$

$$- \left[h'\beta_h - (1 - h')\alpha_h \right] \left\{ \frac{\partial H}{\partial h}(R(y)) \right\}$$

$$+ \left[\alpha_n(1 - n') - n'\beta_n \right] \left\{ \frac{\partial H}{\partial n}(R(y)) \right\}, \tag{C.1}$$

where $m' = (M_2 + M_6)/3 + 2(M_3 + M_7)/3 + (M_4 + M_8)$, $h' = M_5 + M_6 + M_7 + M_8$ and $n' = N_2/4 + N_3/2 + 3N_4/4 + N_5$. Thus, $G(y)$ is in the span of the column vectors $\partial H/\partial v$, $\partial H/\partial m$, $\partial H/\partial n$, and $\partial H/\partial h$, as was to be shown.

On the other hand, the vector field for the 4D HH ODE, equations 2.1 to 2.4, is given by

$$F = \begin{bmatrix} \left(-\bar{g}_{Na}m^3h(V - V_{Na}) - \bar{g}_K n^4(V - V_K) - g_L(V - V_L) + I_{app} \right)/C \\ \alpha_m(V)(1 - m) - \beta_m(V)m \\ \alpha_h(V)(1 - h) - \beta_h(V)h \\ \alpha_n(V)(1 - n) - \beta_n(V)n \end{bmatrix}.$$

Referring to equation C.1, we see that $G(y) = D_x H(R(y)) F(R(y))$. Thus we complete the proof of lemma 2. $\qquad\square$

Appendix D: Diffusion Matrix of the 14D Model

As defined in equations 2.5 and 2.6,

$$\mathbf{M} = [m_{00}, m_{10}, m_{20}, m_{30}, m_{01}, m_{11}, m_{21}, m_{31}]^\mathsf{T}, \tag{D.1}$$

$$\mathbf{N} = [n_0, n_1, n_2, n_3, n_4]^\mathsf{T}, \tag{D.2}$$

the diffusion matrices D_{Na} and D_K are given by

$$D_K = \begin{bmatrix} D_K(1,1) & -4\alpha_n n_0 - \beta_n n_1 & 0 \\ -4\alpha_n n_0 - \beta_n n_1 & D_K(2,2) & 3\alpha_n n_1 - 2\beta_n n_2 \\ 0 & -3\alpha_n n_1 - 2\beta_n n_2 & D_K(3,3) & \cdots \\ 0 & 0 & -2\alpha_n n_2 - 3\beta_n n_3 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\begin{array}{cc}
0 & 0 \\
0 & 0 \\
\cdots \quad -2\alpha_n n_2 - 3\beta_n n_3 & 0 \\
D_K(4,4) & -\alpha_n n_3 - 4\beta_n n_4 \\
-\alpha_n n_3 - 4\beta_n n_4 & D_K(5,5)
\end{array} \Bigg],$$

$$D_{Na}^{(1:3)} = \begin{bmatrix}
D_{Na}(1,1) & -3\alpha_m m_{00} - \beta_m m_{10} & 0 \\
-3\alpha_m m_{00} - \beta_m m_{10} & D_{Na}(2,2) & -2\alpha_m m_{10} - 2\beta_m m_{20} \\
0 & -2\alpha_m m_{10} - 2\beta_m m_{20} & D_{Na}(3,3) \\
0 & 0 & -\alpha_m m_{20} - 3\beta_m m_{30} \\
-\alpha_h m_{00} - \beta_h m_{01} & 0 & 0 \\
0 & -\alpha_h m_{10} - \beta_h m_{11} & 0 \\
0 & 0 & -\alpha_h m_{20} - \beta_h m_{21} \\
0 & 0 & 0
\end{bmatrix},$$

$$D_{Na}^{(4:6)} = \begin{bmatrix}
0 & -\alpha_h m_{00} - \beta_h m_{01} & 0 \\
0 & 0 & -\alpha_h m_{10} - \beta_h m_{11} \\
-\alpha_m m_{20} - 3\beta_m m_{30} & 0 & 0 \\
D_{Na}(4,4) & 0 & 0 \\
0 & D_{Na}(5,5) & -3\alpha_m m_{01} - \beta_m m_{11} \\
0 & -3\alpha_m m_{01} - \beta_m m_{11} & D_{Na}(6,6) \\
0 & 0 & -2\alpha_m m_{11} - 2\beta_m m_{21} \\
-\alpha_h m_{30} - \beta_h m_{31} & 0 & 0
\end{bmatrix},$$

$$D_{Na}^{(7:8)} = \begin{bmatrix}
0 & 0 \\
0 & 0 \\
-\alpha_h m_{20} - \beta_h m_{21} & 0 \\
0 & -\alpha_h m_{30} - \beta_h m_{31} \\
0 & 0 \\
-2\alpha_m m_{11} - 2\beta_m m_{21} & 0 \\
D_{Na}(7,7) & -\alpha_m m_{21} - 3\beta_m m_{31} \\
-\alpha_m m_{21} - 3\beta_m m_{31} & D_{Na}(8,8)
\end{bmatrix},$$

where

$$D_{ion}(i,i) = -\sum_{j:\, j\neq i} D_{ion}(j,i), \text{ for ion } \in \{Na, K\}.$$

The matrices S_K and S_{Na} for the 14×28D HH model are given by

$$
S_{Na}^{(1:5)} = \begin{bmatrix}
-\sqrt{\alpha_h m_{00}} & \sqrt{\beta_h m_{01}} & -\sqrt{3\alpha_m m_{00}} & \sqrt{\beta_m m_{10}} & 0 \\
0 & 0 & \sqrt{3\alpha_m m_{00}} & -\sqrt{\beta_m m_{10}} & -\sqrt{\alpha_h m_{10}} \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
\sqrt{\alpha_h m_{00}} & -\sqrt{\beta_h m_{01}} & 0 & 0 & 0 \\
-\sqrt{\beta_h m_{11}} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \sqrt{\alpha_h m_{20}} & -\sqrt{\beta_h m_{21}} \\
0 & 0 & 0 & 0 & 0
\end{bmatrix}
$$

$$
S_{Na}^{(6:10)} = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 \\
\sqrt{\beta_h m_{11}} & -\sqrt{2\alpha_m m_{10}} & \sqrt{2\beta_m m_{20}} & 0 & 0 \\
\sqrt{2\alpha_m m_{10}} & -\sqrt{2\beta_m m_{20}} & -\sqrt{\alpha_h m_{20}} & \sqrt{\beta_h m_{21}} \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
-\sqrt{\beta_h m_{11}} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \sqrt{\alpha_h m_{20}} & -\sqrt{\beta_h m_{21}} \\
0 & 0 & 0 & 0 & 0
\end{bmatrix}
$$

$$
S_{Na}^{(11:15)} = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
-\sqrt{\alpha_m m_{20}} & \sqrt{3\beta_m m_{30}} & 0 & 0 & 0 \\
\sqrt{\alpha_m m_{20}} & -\sqrt{3\beta_m m_{30}} & -\sqrt{\alpha_h m_{30}} & \sqrt{\beta_h m_{31}} & 0 \\
0 & 0 & 0 & 00 & \\
0 & 0 & 0 & 0 & \sqrt{3\alpha_m m_{01}} \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & \sqrt{\alpha_h m_{30}} & -\sqrt{\beta_h m_{31}} & 0
\end{bmatrix}
$$

$$
S_{Na}^{(16:20)} = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
-\sqrt{3\alpha_m m_{01}} & \sqrt{\beta_m m_{11}} & 0 & 0 & 0 \\
-\sqrt{\beta_m m_{11}} & -\sqrt{2\alpha_m m_{11}} & \sqrt{2\beta_m m_{21}} & 0 & 0 \\
0 & \sqrt{2\alpha_m m_{11}} & -\sqrt{2\beta_m m_{21}} & -\sqrt{\alpha_m m_{21}} & \sqrt{3\beta_m m_{31}} \\
0 & 0 & 0 & \sqrt{\alpha_m m_{21}} & -\sqrt{3\beta_m m_{31}}
\end{bmatrix},
$$

where $S_{\text{Na}}^{(i;j)}$ is the ith $-$ jth column of S_{Na}, and

$$S_K = \begin{bmatrix} -\sqrt{4\alpha_n n_0} & \sqrt{\beta_n n_1} & 0 & 0 \\ \sqrt{4\alpha_n n_0} & -\sqrt{\beta_n n_1} & -\sqrt{3\alpha_n n_1} & \sqrt{2\beta_n n_2} \\ 0 & 0 & \sqrt{3\alpha_n n_1} & -\sqrt{2\beta_n n_2} & \cdots \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\cdots \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \sqrt{2\alpha_n n_2} & -\sqrt{3\beta_n n_3} & -\sqrt{\alpha_n n_3} & \sqrt{4\beta_n n_4} \\ 0 & 0 & \sqrt{\alpha_n n_3} & -\sqrt{4\beta_n n_4} \end{bmatrix}.$$

References

Allen, E. J., Allen, L. J., Arciniega, A., & Greenwood, P. E. (2008). Construction of equivalent stochastic differential equation models. *Stochastic Analysis and Applications*, *26*(2), 274–297.

Anderson, D. F., Ermentrout, B., & Thomas, P. J. (2015). Stochastic representations of ion channel kinetics and exact stochastic simulation of neuronal dynamics. *Journal of Computational Neuroscience*, *38*(1), 67–82.

Anderson, D. F., & Kurtz, T. G. (2015). *Stochastic analysis of biochemical systems*, vol. *1*. Berlin: Springer.

Bruce, I. C. (2009). Evaluation of stochastic differential equation approximation of ion channel gating models. *Annals of Biomedical Engineering*, *37*(4), 824–838.

Clay, J. R., & DeFelice, L. J. (1983). Relationship between membrane excitability and single channel open-close kinetics. *Biophys. J.*, *42*(2), 151–157.

Colquhoun, D., & Hawkes, A. (1981). On the stochastic properties of single ion channels. *Proc. R. Soc. Lond. B*, *211*(1183), 205–235.

Dangerfield, C., Kay, D., & Burrage, K. (2010). Stochastic models and simulation of ion channel dynamics. *Procedia Computer Science*, *1*(1), 1587–1596.

Dangerfield, C. E., Kay, D., & Burrage, K. (2012). Modeling ion channel dynamics through reflected stochastic differential equations. *Physical Review E*, *85*(5), 051907.

Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience: Computational and mathematical modeling of neural systems*. Cambridge, MA: MIT Press.

Dvoretzky, A., Kiefer, J., & Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Annals of Mathematical Statistics*, *27*(3), 642–669.

Earnshaw, B., & Keener, J. (2010a). Global asymptotic stability of solutions of nonautonomous master equations. *SIAM Journal on Applied Dynamical Systems*, *9*(1), 220–237.

Earnshaw, B., & Keener, J. (2010b). Invariant manifolds of binomial-like nonautonomous master equations. *SIAM Journal on Applied Dynamical Systems*, *9*(2), 568–588.

Faisal, A. A., & Laughlin, S. B. (2007). Stochastic simulations on the reliability of action potential propagation in thin axons. *PLOS Computational Biology*, *3*, 5.

Fox, R. F. (1997). Stochastic versions of the Hodgkin-Huxley equations. *Biophysical Journal*, *72*(5), 2068–2074.

Fox, R. F. (2018). *Critique of the Fox-Lu model for Hodgkin-Huxley fluctuations in neuron ion channels*. arXiv:1807.07632.

Fox, R., & Lu, Y. (1994). Emergent collective behavior in large numbers of globally coupled independently stochastic ion channels. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics*, *49*(4), 3421–3431.

Gadgil, C., Lee, C. H., & Othmer, H. G. (2005). A stochastic analysis of first-order reaction networks. *Bulletin of Mathematical Biology*, *67*, 901–946.

Gerstner, W., Kempter, R., Van Hemmen, J. L., & Wagner, H. (1996). A neuronal learning rule for sub-millisecond temporal coding. *Nature*, *383*(6595), 76–78.

Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, *81*, 2340–2361.

Gillespie, D. T. (2007). Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.*, *58*, 35–55.

Goldwyn, J. H., Imennov, N. S., Famulare, M., & Shea-Brown, E. (2011). Stochastic differential equation models for ion channel noise in Hodgkin-Huxley neurons. *Physical Review E*, *83*(4), 041908.

Goldwyn, J. H., & Shea-Brown, E. (2011). The what and where of adding channel noise to the Hodgkin-Huxley equations. *PLOS Comput. Biol.*, *7*(11), e1002247.

Goldwyn, J. H., Shea-Brown, E., & Rubinstein, J. T. (2010). Encoding and decoding amplitude-modulated cochlear implant stimuli—a point process analysis. *Journal of Computational Neuroscience*, *28*(3), 405–424.

Goychuk, I. (2014). Stochastic modeling of excitable dynamics, improved Langevin model for mesoscopic channel noise. In *Proceedings of the International Conference on Nonlinear Dynamics of Electronic Systems* (pp. 325–332). Berlin: Springer.

Greenwood, P. E., McDonnell, M. D., & Ward, L. M. (2015). Dynamics of gamma bursts in local field potentials. *Neural Computation*, *27*(1), 74–103.

Grimmett, G., & Stirzaker, D. (2005). *Probability and random processes* (3rd ed.). New York: Oxford University Press.

Güler, M. (2013a). An investigation of the stochastic Hodgkin-Huxley models under noisy rate functions. *Neural Computation*, *25*(9), 2355–2372.

Güler, M. (2013b). Stochastic Hodgkin-Huxley equations with colored noise terms in the conductances. *Neural Computation*, *25*(1), 46–74.

Hill, T. L., & Chen, Y.-D. (1972). On the theory of ion transport across the nerve membrane: IV. Noise from the open-close kinetics of K+ channels. *Biophysical Journal*, *12*(8), 948–959.

Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.*, *117*, 500–544.

Huang, Y., Rüdiger, S., & Shuai, J. (2013). Channel-based Langevin approach for the stochastic Hodgkin-Huxley neuron. *Physical Review E*, *87*(1), 012716.

Huang, Y., Rüdiger, S., & Shuai, J. (2015). Accurate Langevin approaches to simulate Markovian channel dynamics. *Physical Biology*, *12*(6), 061001.

Imennov, N. S., & Rubinstein, J. T. (2009). Stochastic population model for electrical stimulation of the auditory nerve. *IEEE Transactions on Biomedical Engineering*, *56*(10), 2493–2501.

Keener, J. P. (2009). Invariant manifold reductions for Markovian ion channel dynamics. *J. Math. Biol.*, *58*(3), 447–457.

Keener, J. P. (2010). Exact reductions of Markovian dynamics for ion channels with a single permissive state. *J. Math. Biol.*, *60*(4), 473–479.

Keener, J. P., & Sneyd, J. (1998). *Mathematical physiology*, vol. 1. Berlin: Springer.

Kispersky, T., & White, J. A. (2008). Stochastic models of ion channel gating. *Scholarpedia*, *3*(1), 1327.

Kloeden, P. E., & Platen, E. (1999). *Numerical solution of stochastic differential equations*. Berlin: Springer.

Kolmogorov, A. (1933). Sulla determinazione empirica di una lgge di distribuzione. *Inst. Ital. Attuari, Giorn.*, *4*, 83–91.

Kurtz, T. G. (1981). *Approximation of population processes*. Philadelphia: SIAM.

Linaro, D., Storace, M., & Giugliano, M. (2011). Accurate and fast simulation of channel noise in conductance-based model neurons by diffusion approximation. *PLOS Computational Biology*, *7*(3), e1001102.

Mainen, Z. F., & Sejnowski, T. J. (1995). Reliability of spike timing in neocortical neurons. *Science*, *268*(5216), 1503–1506.

Mino, H., Rubinstein, J. T., & White, J. A. (2002). Comparison of algorithms for the simulation of action potentials with stochastic sodium channels. *Ann. Biomed. Eng.*, *30*(4), 578–587.

Moiseff, A., & Konishi, M. (1981). The owl's interaural pathway is not involved in sound localization. *Journal of Comparative Physiology*, *144*(3), 299–304.

Neher, E., & Sakmann, B. (1976). Single-channel currents recorded from membrane of denervated frog muscle fibres. *Nature*, *260*(5554), 799.

Newby, J. M., Bressloff, P. C., & Keener, J. P. (2013). Breakdown of fast-slow analysis in an excitable system with channel noise. *Phys. Rev. Lett.*, *111*(12), 128101.

O'Donnell, C., & van Rossum, M. C. (2015). Spontaneous action potentials and neural coding in unmyelinated axons. *Neural Computation*, *27*(4), 801–818.

Orio, P., & Soudry, D. (2012). Simple, fast and accurate implementation of the diffusion approximation algorithm for stochastic ion channels with multiple states. *PLOS One*, *7*(5), e36670.

Pezo, D., Soudry, D., & Orio, P. (2014). Diffusion approximation-based simulation of stochastic ion channels: Which method to use? *Frontiers in Computational Neuroscience*, *8*, 139.

Schmandt, N. T., & Galán, R. F. (2012). Stochastic-shielding approximation of Markov chains and its application to efficiently simulate random ion-channel gating. *Physical Review Letters*, *109*(11), 118101.

Schmid, G., Goychuk, I., & Hänggi, P. (2001). Stochastic resonance as a collective property of ion channel assemblies. *Europhysics Letters*, *56*(1), 22.

Schmidt, D. R., Galán, R. F., & Thomas, P. J. (2018). Stochastic shielding and edge importance for Markov chains with timescale separation. *PLOS Computational Biology*, *14*(6), e1006206.

Schmidt, D. R., & Thomas, P. J. (2014). Measuring edge importance: A quantitative analysis of the stochastic shielding approximation for random processes on graphs. *Journal of Mathematical Neuroscience*, *4*(1), 6.

Schneidman, E., Freedman, B., & Segev, I. (1998). Ion channel stochasticity may be critical in determining the reliability and precision of spike timing. *Neural Computation*, *10*(7), 1679–1703.

Sengupta, B., Laughlin, S., & Niven, J. (2010). Comparison of Langevin and Markov channel noise models for neuronal signal generation. *Physical Review E*, *81*(1), 011918.

Shorack, G. R., & Wellner, J. A. (2009). *Empirical processes with applications to statistics*. Philadelphia: SIAM.

Skaugen, E., & Walløe, L. (1979). Firing behaviour in a stochastic nerve membrane model based upon the Hodgkin-Huxley equations. *Acta Physiol. Scand.*, *107*(4), 343–363.

Smirnov, N. (1948). Table for estimating the goodness of fit of empirical distributions. *Annals of Mathematical Statistics*, *19*(2), 279–281.

Strassberg, A. F., & DeFelice, L. J. (1993). Limitations of the Hodgkin-Huxley formalism: Effects of single channel kinetics on transmembrane voltage dynamics. *Neural Computation*, *5*(6), 843–855.

Wasserstein, L. N. (1969). Markov processes over denumerable products of spaces describing large systems of automata. *Problems of Information Transmission*, *5*(3), 47–52.

White, J. A., Klink, R., Alonso, A., & Kay, A. R. (1998). Noise from voltage-gated ion channels may influence neuronal dynamics in the entorhinal cortex. *Journal of Neurophysiology*, *80*(1), 262–269.

White, J. A., Rubinstein, J. T., & Kay, A. R. (2000). Channel noise in neurons. *Trends in Neurosciences*, *23*(3), 131–137.

Woo, J., Miller, C. A., & Abbas, P. J. (2009). Simulation of the electrically stimulated cochlear neuron: Modeling adaptation to trains of electric pulses. *IEEE Transactions on Biomedical Engineering*, *56*(5), 1348–1359.

Yu, H., Dhingra, R. R., Dick, T. E., & Galán, R. F. (2017). Effects of ion channel noise on neural circuits: An application to the respiratory pattern generator to investigate breathing variability. *Journal of Neurophysiology*, *117*(1), 230–242.