# Autoregressive Point Processes as Latent State-Space Models: A Moment-Closure Approach to Fluctuations and Autocorrelations

**Michael Rule**
*mrule@alumni.brown.edu*
**Guido Sanguinetti**
*gsanguin@inf.ed.ac.uk*
*Institute for Adaptive and Neural Computation, School of Informatics,*
*University of Edinburgh, Edinburgh EH8 9AB, U.K.*

**Modeling and interpreting spike train data is a task of central importance in computational neuroscience, with significant translational implications. Two popular classes of data-driven models for this task are autoregressive point-process generalized linear models (PPGLM) and latent state-space models (SSM) with point-process observations. In this letter, we derive a mathematical connection between these two classes of models. By introducing an auxiliary history process, we represent exactly a PPGLM in terms of a latent, infinite-dimensional dynamical system, which can then be mapped onto an SSM by basis function projections and moment closure. This representation provides a new perspective on widely used methods for modeling spike data and also suggests novel algorithmic approaches to fitting such models. We illustrate our results on a phasic bursting neuron model, showing that our proposed approach provides an accurate and efficient way to capture neural dynamics.**

## 1 Introduction

Connecting single-neuron spiking to the collective dynamics that emerge in neural populations remains a central challenge in systems neuroscience. As well as representing a major barrier in our understanding of fundamental neural function, this challenge has recently acquired new salience due to the rapid improvements in technologies that can measure neural population activity in vitro and in vivo at unprecedented temporal and spatial resolution (Jun et al., 2017; Maccione et al., 2014). Such technologies hold immense promise in elucidating both normal neural functioning and the etiology of many diseases, yet the high dimensionality and complexity of the resulting data pose formidable statistical challenges. In response to these needs, recent years have seen considerable efforts to develop strategies for extracting and modeling information from large-scale spiking neural recordings. Two of the most successful strategies that emerged in the past decade are

latent state-space models (SSMs) and autoregressive point-process generalized linear models (PPGLMs).

Latent state-space models describe neural spiking as arising from the unobserved latent dynamics of an auxiliary intensity field, which can model both internal and external factors contributing to the dynamics (Macke, Buesing, & Sahani, 2015; Sussillo, Jozefowicz, Abbott, & Pandarinath, 2016; Zhao & Park, 2016; Yu et al., 2009; Smith & Brown, 2003). Mathematically, such models generally take the form of a Cox process (Kingman, 1993), where the intensity field obeys some (discrete or continuous time) evolution equations. This representation therefore recasts the analysis of spike trains within a well-established line of research in statistical signal processing, leveraging both classical tools and more recent developments (Smith & Brown, 2003; Wu, Roy, Keeley, & Pillow, 2017; Gao, Archer, Paninski, & Cunningham, 2016; Pfau, Pnevmatikakis, & Paninski, 2013; Zhao & Park, 2017; Surace, Kutschireiter, & Pfister, 2017). These models have been used in a variety of tasks, such as describing population spiking activity in the motor system (Aghagolzadeh & Truccolo, 2014, 2016; Churchland et al., 2012; Michaels, Dann, Intveld, & Scherberger, 2017). However, while such models can certainly lead to biological insights, latent state-space models remain phenomenological: the recurrent spiking activity itself does not implement the latent state-space dynamics (see Figure 2B).

Autoregressive PPGLMs (see Figure 2A) treat spiking events from neurons as point events arising from a latent inhomogeneous Poisson process (Truccolo, Eden, Fellows, Donoghue, & Brown, 2005; Truccolo, Hochberg, & Donoghue, 2010; Truccolo, 2010, 2016). To fit such models, generalized linear model (GLM) regression is used to map observed spiking events to both extrinsic variables, like stimuli or motor output, and intrinsic spiking history (Truccolo et al., 2005). PPGLMs are especially useful for statistical tests on sources of variability in neural spiking (Rule, Vargas-Irwin, Donoghue, & Truccolo, 2015, 2017) and benefit from a simple fitting procedure that can often be solved by convex optimization. However, they may require careful regularization to avoid instability (Hocker & Park, 2017; Gerhard, Deger, & Truccolo, 2017) and can fail to generalize outside regimes in which they were trained (Weber & Pillow, 2017). Importantly, PPGLMs suffer from confounds if there are unobserved sources of neural variability (Lawhern, Wu, Hatsopoulos, & Paninski, 2010). This is especially apparent when the recorded neural population is a small subsample of the population, and latent state-space models can be more accurate in decoding applications (Aghagolzadeh & Truccolo, 2016).

In this letter, we establish a mathematical connection between autoregressive PPGLMs and SSMs based on low-dimensional, low-order approximation to an exact infinite-dimensional representation of a PPGLM. Unlike previous work, which explored mean-field limits (Gerhard et al., 2017; Chevallier, Duarte, Löcherbach, & Ost, 2017; Galves & Löcherbach, 2015; Delarue, Inglis, Rubenthaler, & Tanré, 2015), we use gaussian

moment closure (Schnoerr, Sanguinetti, & Grima, 2015, 2017) to capture the excitatory effects of fluctuations and process autocorrelations. In doing so, we convert the autohistory effects in spiking into nonlinear dynamics in a low-dimensional latent state space. This converts an autoregressive point process into a latent-variable Cox process, where spikes are then viewed as Poisson events driven by latent states. This connection, as well as being interesting in its own right, also provides a valuable cross-fertilization opportunity between the two approaches. For example, the issue of runaway self-excitation in PPGLMs emerges as divergence in the moment closure ordinary differential equations, leading to practical insights into obtaining a stabilized state-space analog of the autoregressive PPGLM. We illustrate the approach on the case study of the phasic bursting of an Izhikevich (Izhikevich, 2003) neuron model (see Figure 1) considered in Weber and Pillow (2017), showing that our approach achieves both high accuracy in the mean and can capture remarkably well the fluctuations of the process.

## 2 Results

We start by recapitulating some basic notations and definitions from both PPGLMs and SSMs. We then provide a detailed derivation of the mathematical connection between the two frameworks, highlighting all the approximations we make in the process. Finally, we illustrate the performance of the method in an application case study.

**2.1 Point-Process Generalized Linear Models.** A point process (PP) is a subset of dimension zero of a higher-dimensional space (Brillinger, 1988; Truccolo et al., 2005). For our purposes, we consider only PPs over the time domain, so we will equivalently consider a realization of a PP as a series of points in time $y(t)$, where each point (spiking event) is a delta distribution at the event time. We can associate with a PP in time a locally constant counting process $N(t)$ that counts the cumulative number of events up to time $t$. The process $y(t)$ can be thought of as representing the spike train output by a neuron, while the cumulative process $N(t)$ provides a clearer notation for the derivations that follow:

$$N(t) = \# \text{ events} \leq t \,,$$

$$y(t) = \frac{d}{dt} N(t) = \sum_{\tau \in \text{events}} \delta(t = \tau).$$

We restrict our attention to PPs that can be described by an underlying intensity function $\lambda(t)$. In the simplest case, event counts between times $t$ and $t + \Delta$ occur with a Poisson distribution, with a mean rate given by the integral of $\lambda(t)$ over the time window:
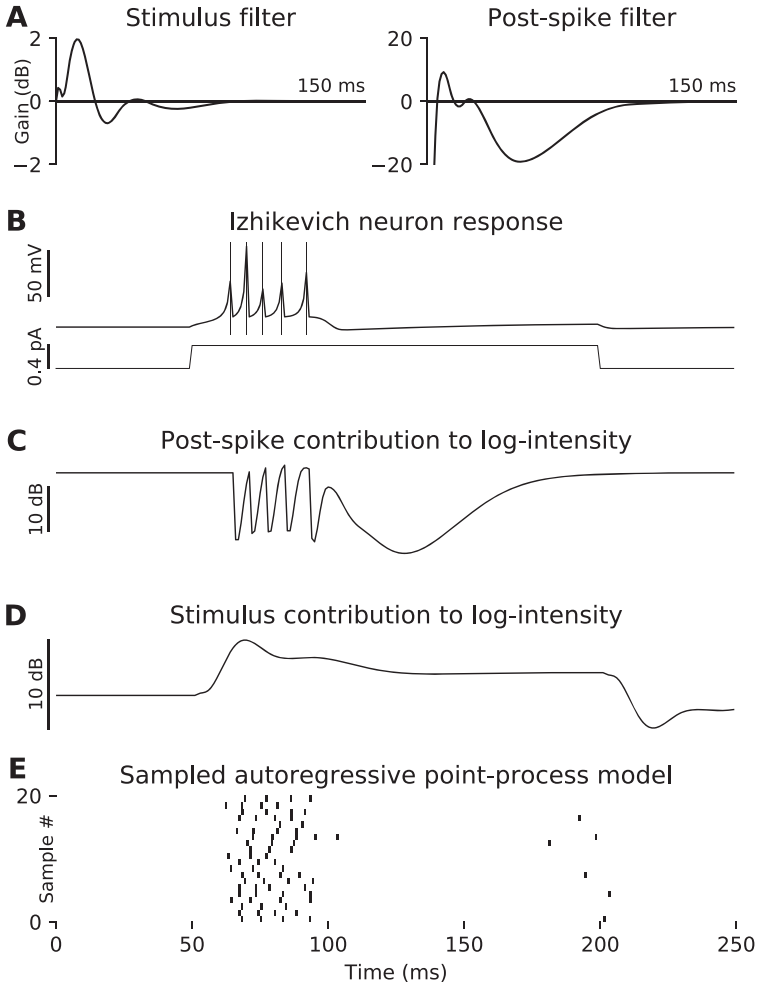
Figure 1: Phasic bursting neuron as emulated by an autoregressive PPGLM model. An autoregressive PPGLM was trained on spiking output from a phasic-bursting Izhikevich neuron model (B) with parameters $a, b, c, d, dt = 0.02, 0.25, -55, 0.05, 1.0$ (Weber & Pillow, 2017). Training stimuli consisted of square current pulses ranging from 0.3 to 0.7 pA and from 10 to 500 ms in duration, with added Ornstein-Uhlenbeck noise with a 200 ms time constant and steady-state variance of 0.01 pA$^2$. (A) Stimulus and postspike filters exhibit ringing associated with the characteristic interspike interval. The postspike filter includes an additional postburst slow-inhibitory component, evident in the postspike contribution to the log intensity (C). This slow postburst inhibition confers additional slow dynamics on top of the stimulus-driven response (D). (E) Samples from the autoregressive PPGLMs reflect the phasic bursting neuronal response.

**A** Autoregressive PPGLM



**B** State-space model



**C** Moment-closure PPGLM
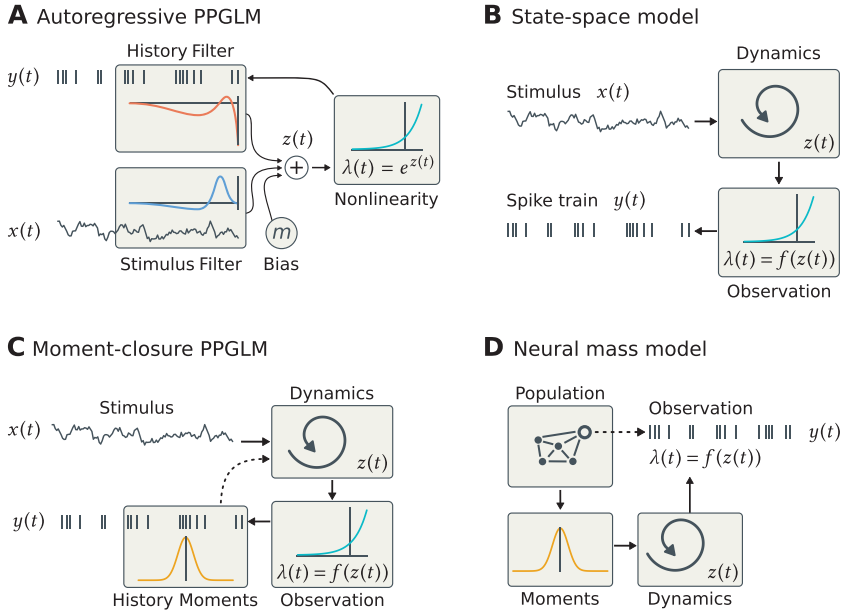


**D** Neural mass model



Figure 2: Moment closure of autoregressive PPGLMs combines aspects of three modeling approaches. (A) Log-linear autoregressive PPGLM framework (e.g., Weber & Pillow, 2017). Dependence on the history of both extrinsic covariates $x(t)$ and the process itself $y(t)$ are mediated by linear filters, which are combined to predict the instantaneous log intensity of the process. (B) Latent state-space models learn a hidden dynamical system, which can be driven by both extrinsic covariates and spiking outputs. Such models are often fit using expectation-maximization, and the learned dynamics are descriptive. (C) Moment closure recasts autoregressive PPGLMs as state-space models. History dependence of the process is subsumed into the state-space dynamics, but the latent states retain a physical interpretation as moments of the process history (dashed arrow). (D) Compare to neural mass and neural field models, which define dynamics on a state space with a physical interpretation as moments of neural population activity.

$$\Pr\left(N(t + \Delta) - N(t) = k\right) \sim \text{Poisson}\left(\int_t^{t+\Delta} \lambda(t)dt\right). \tag{2.1}$$

In the autoregressive PPGLM (see Figure 2A), one models the intensity $\lambda(t)$ conditioned on the past events, as well as extrinsic covariates $x(t)$. Generally,

$$f\left(\lambda(t)\right) = m + F^\top x(t) + \lim_{\epsilon \to 0^+} \int_\epsilon^\infty H(\tau)y(t - \tau)d\tau, \tag{2.2}$$

where $f$ is called the link function, $F$ is a matrix or operator projecting extrinsic covariates down to the dimensionality of the point process, $m$ is a mean or bias parameter, and $H$ is a history filter function. The open integration limit $\epsilon \rightarrow 0^+$ reflects the fact that the spiking at the current time-point $t$ is excluded from the history filter. The inputs and bias are fixed and can be denoted by a single time-dependent input function $I(t) = F^\top x(t) + m$. Here we will take the link function $f$ to be the natural logarithm. Rewriting equation 2.2, making time dependence implicit where it is unambiguous, and denoting the history filter integral as $H^\top y$, we will explore generalized linear models of the form

$$\lambda = \exp\left(m + F^\top x + H^\top y\right). \tag{2.3}$$

Autoregressive PPGLMs can emulate various firing behaviors of real neurons (Weber & Pillow, 2017). For example phasic bursting neurons (see Figure 1) exhibit complex autohistory dependence on both fast and slow timescales. This dependence of the process on intrinsic history confers additional slow dynamics, that is, postburst inhibition on the timescale of tens of milliseconds.

**2.2 Latent State-Space Point-Process Models.** An alternative strategy for capturing slow dynamics in neural spike trains is to postulate a slow, latent dynamical system responsible for bursting and postburst inhibition. This approach is taken by latent state-space models (SSMs), which are often viewed as functionally distinct from autoregressive PPGLMs (see Figure 2B).

In general, a latent state-space model describes how both deterministic and stochastic dynamics of a latent variable $x$ affect the intensity $\lambda$ of a point process:

$$dx(t) = u(x, t)dt + \sigma(x, t)dW,$$
$$\lambda = v(x, t),$$
$$dN \sim \text{Poisson}(\lambda \cdot dt), \tag{2.4}$$

where $dW$ is the derivative of the standard Wiener process, reflecting fluctuations. The functions $u, \sigma$, and $v$ describe, respectively, deterministic evolution, stochastic fluctuations, and the observation model. In the case of, for example, the Poisson linear dynamical system (PLDS; Macke et al., 2011), the latent dynamics are linear with fixed gaussian noise:

$$dx(t) = [Ax + I(t)]dt + \sigma dW,$$
$$\lambda = \exp\left(m + F^\top x + H^\top y\right),$$
$$dN \sim \text{Poisson}(\lambda \cdot dt), \tag{2.5}$$

where $I(t)$ reflects inputs into the latent state-space, and the spiking probability depends on latent states, history, and bias, as in equation 2.3. Latent state-space models of point processes have been investigated in detail (Macke et al., 2011; Smith & Brown, 2003), and mature inference approaches are available to estimate states and parameters from data (Lawhern et al., 2010; Macke et al., 2015; Buesing, Macke, & Sahani, 2012; Rue, Martino, & Chopin, 2009; Cseke, Zammit-Mangion, Heskes, & Sanguinetti, 2016). However, such models are typically phenomenological, lacking a clear physiological interpretation of the latent dynamics. Importantly, the point-process history is typically fit as if it were another extrinsic covariate, and the effects of Poisson fluctuations are either neglected or handled in mean-field limit. This obscures the dynamical role of population spiking history and its fluctuations in the autoregressive PPGLM. In the remainder of this letter, we illustrate that the history dependence of PPGLMs implicitly defines a latent-state space model over moments of the process history.

**2.3 The Auxiliary History Process of a PPGLM.** The (possibly infinite) history dependence makes autoregressive PPGLMs non-Markovian dynamical systems (Truccolo, 2016, equation 6). However, a crucial insight is that since the dependence on history is linear, we can reinterpret the history dependence as a linear filter in time and approximate its effect on the conditional intensity using a low-dimensional linear dynamical system. In some PPGLM formulations, the history basis may already be defined in terms of a linear filter impulse response—for example, the exponential filters considered in Toyoizumi, Rad, and Paninski (2009). In this case, the history process is already Markovian. As the Markovian case has been illustrated elsewhere, we show here how to convert the general non-Markovian point process into a Markovian one.

To formalize this, we introduce an auxiliary history process $h(\tau, t)$ that "stores" the history of the process $y(t)$. One can view $h(\tau, t)$ as a delay line that tracks the signal $y(t)$. The time evolution of $h$ is given by

$$\partial_t h(\tau, t) = \delta_{\tau=0} dN(t) - \partial_\tau h(\tau, t), \tag{2.6}$$

where $\delta_{\tau=0}$ indicates that new events $y(t) = dN(t)$ should be inserted into the history process at $\tau = 0$, and $\partial_\tau$ is the derivative with respect to time lag $\tau$. This converts the autoregressive PPGLM to a stationary Markovian process over an augmented (infinite-dimensional) state space:

$$dN(t) \sim \text{Poisson}(\lambda \cdot dt),$$
$$\lambda(t) = \exp\left(H(\tau)^\top h(\tau, t) + I(t)\right),$$
$$\partial_t h(\tau, t) = \delta_{\tau=0} dN(t) - \partial_\tau h(\tau, t), \tag{2.7}$$

where $H(\tau)$ is the history filter introduced in equation 2.2. In this formulation, the history process $h(\tau, t)$ is still a point process. However, the interaction between history $h(\tau, t)$ and the intensity $\lambda(t)$ is mediated entirely by the projection $H(\tau)^\top h(\tau, t)$, which averages over the process history. To capture the relevant influences of the process history, it then suffices to capture the effects of Poisson variability on this averaged projection.

**2.4 A Continuous Approximation.** In the limit where events are frequent, the Poisson process $dN(t)$ can be approximated as a Wiener process with mean and variance equal to the instantaneous point-process intensity $\lambda(t)$. In the derivations that follow, we omit explicit notation of time dependence (e.g., $\lambda(t)$, $h(\tau, t)$) where unambiguous:

$$dN \approx \lambda dt + \sqrt{\lambda} dW, \tag{2.8}$$

This approximation holds when averaging over a population of weakly coupled neurons (Toyoizumi et al., 2009) or over slow timescales of a single neuron. This approximation is inspired by the chemical Langevin equation (Gillespie, 2000), which remains accurate in the regime of sparse reactions (sparse counts) and is more accurate (in terms of moments) than a linear noise approximation (Schnoerr et al., 2017; see appendix B). We will illustrate (see Figure 3) that this approximation can be surprisingly accurate for even a single neuron. Applying this approximation to the driving noise term in the evolution equation for the auxiliary history process (see equation 2.7), we obtain a continuous (in time and in state) infinite-dimensional approximation of the PPGLM:

$$dh = (\delta_{\tau=0}\lambda - \partial_\tau h)\, dt + \delta_{\tau=0}\sqrt{\lambda} dW,$$
$$\lambda = \exp\left(H^\top h + I(t)\right). \tag{2.9}$$

Because the dimensionality of the history process $h(\tau, t)$ is infinite, this is a stochastic partial differential equation (SPDE). Importantly, this is a system of equations for the history of the process, not the instantaneous rate $\lambda(t)$.

**2.5 Gaussian Moment Closure.** The SPDE in equation 2.9 is analytically intractable due to the exponential inverse link function; however, it is possible to derive an (infinite) set of coupled moment equations for the process. We can then close these equations by setting all cumulants of order greater than two to zero, effectively enforcing the gaussianity of the history process $h(\tau, t)$. The (exact) equation for the process mean $\mu(\tau) = \langle h(\tau) \rangle$ is

$$\partial_t \mu = \partial_t \langle h \rangle$$
$$= \langle \delta_{\tau=0}\lambda - \partial_\tau h \rangle$$
$$= \delta_{\tau=0} \langle \lambda \rangle - \partial_\tau \mu. \tag{2.10}$$

The log-intensity $\ln \lambda = H^\top h + I(t)$ is a linear projection of the history process $h(\tau, t)$, which we approximate as a gaussian process with mean $\mu(\tau)$ and covariance $\Sigma(\tau, \tau')$. Therefore, the log rate is normally distributed with mean $H^\top \mu + I(t)$ and variance $H^\top \Sigma H$, and the firing rate is log normally distributed with mean

$$\langle \lambda \rangle = \exp \left( H^\top \mu + I(t) + \frac{1}{2} H^\top \Sigma H \right). \tag{2.11}$$

Note that this is an approximation, as in general, the higher-order cumulants of the history process will not remain zero due to to the influence of Poisson noise. Nevertheless, we will see that this approximation accurately captures the influence of fluctuations and correlations on the mean. This expectation incorporates second-order effects introduced by fluctuations and correlations mediated through the history filter and therefore couples the time evolution of the first moment to the covariance.

The time derivative of the covariance has both deterministic and stochastic contributions. Overall, the deterministic contribution to the derivative of the covariance can be written as $J\Sigma + \Sigma J^\top$, where $J = \delta_{\tau=0} \langle \lambda \rangle H^\top - \partial_\tau$ (see appendix A). The covariance also has a noise contribution $Q = \delta_{\tau=0} \langle \lambda \rangle \delta_{\tau=0}^\top$ from the spiking noise term entering at time lag 0, with variance proportional to the expected firing rate. In sum, the moment equations for the history process using gaussian moment closure are

$$\partial_t \mu = \delta_{\tau=0} \langle \lambda \rangle - \partial_\tau \mu,$$
$$\langle \lambda \rangle = \exp \left( H^\top \mu + I(t) + H^\top \Sigma H \right),$$
$$\partial_t \Sigma = J\Sigma + \Sigma J^\top + Q,$$
$$J = \delta_{\tau=0} \langle \lambda \rangle H^\top - \partial_\tau,$$
$$Q = \delta_{\tau=0} \langle \lambda \rangle \delta_{\tau=0}^\top. \tag{2.12}$$

This notation resembles a continuous-time Kalman-Bucy filter (Kalman & Bucy, 1961), for which $J(t)$ would be a Jacobian of the mean update and $Q(t)$ would reflect the system noise. Equations 2.12 are also reminiscent of classical neural mass and neural field models (Amari 1975, 1977, 1983; Wilson, Cowan, Baker, Cowan, & van Drongelen, 1972; e.g., Figure 2D). Unlike neural field models, however, the moment equations, 2.12, arise not from population averages but from considering the expected behavior of the stochastic process describing the neural spike train (see Figure 2C).

It is worth reflecting more on this analogy and on the limitations of the moment-closure representation. Spiking events are dramatic all-or-nothing events that cannot be approximated by a continuous stochastic process. Accordingly, one would expect the finite-dimensional moment closure system to fail to capture rapid fluctuations. However, for slow timescales, this

gaussian approximation can be accurate even for a single neuron. In contrast to the neural field interpretation, which averages over a large population at each time instant, one can average over an extended time window and arrive at an approximation for slow timescales (see Figure 3). A pictorial description of the relationship of the proposed moment closure approach to PPGLMs, SSMs, and neural field models is summarized in Figure 2.

**2.6 Case Study: The Izhikevich Neuron Model.** We consider the effectiveness of this approach on the case study of a PPGLM emulation of the Izhikevich neuron model, considered in Weber and Pillow (2017). We compare the accuracy of the gaussian moment closure with a mean-field approach (see appendix B). Figure 3 illustrates moment closure of a phasic-bursting Izhikevich neuron emulated with a PPGLM (see Figure 1). By averaging over the history process, slow timescales in the autoregressive point process are captured in the gaussian moment closure. Unlike a mean-field model, which considers the large-population limit of weakly coupled neurons, moment closure is able to capture the influence of Poisson variability on the dynamics in a single neuron.

Additionally, mean field considers only a single path in the process history, whereas moment closure provides an approximation for a distribution over paths, with fluctuations and autocorrelations taken into account. This has the benefit that the moment-closure system is sensitive to the combined effects of self-excitation and point-process fluctuations and captures, for example, the self-excitation during a burst using the second-order, second-moment terms. This reveals another benefit of the moment-closure approach: runaway self-excitation (Hocker & Park, 2017; Gerhard et al., 2017; Weber & Pillow, 2017) is detected in the moment closure as a divergence of the mean or second-moment terms. This self-excitation, however, introduces some numerical challenges.

The gaussian moment closure captures corrections to the mean evolution due to the second moment (see Figure 3C and Table 1), but exhibits a bias in the second moment owing to unmodeled skewness and higher-order moments. Additionally, the self-excitation mediated by the exponential inverse link function combines with fast-timescale postspike inhibition to make the gaussian moment closure equations stiff and difficult to integrate. Skewness controls the effects of outliers on the expected firing rate and affects both stiffness and the errors in estimating the second moment. We can attenuate these effects by replacing the exponential nonlinearity with a locally quadratic approximation and performing gaussian moment closure on this approximation, thereby stabilizing the effect of outliers on the expected firing rate. This second-order moment-closure approach was first outlined in the context of chemical reaction modeling by Ale, Kirk, and Stumpf (2013). In the second-order moment closure, the mean-field $\bar{\lambda}$ is used for the deterministic evolution of the covariance, and the covariance correction to the mean is approximated at second order as $\exp\left(\frac{1}{2}H^{\top}\Sigma H\right) \approx 1 + \frac{1}{2}H^{\top}\Sigma H$,
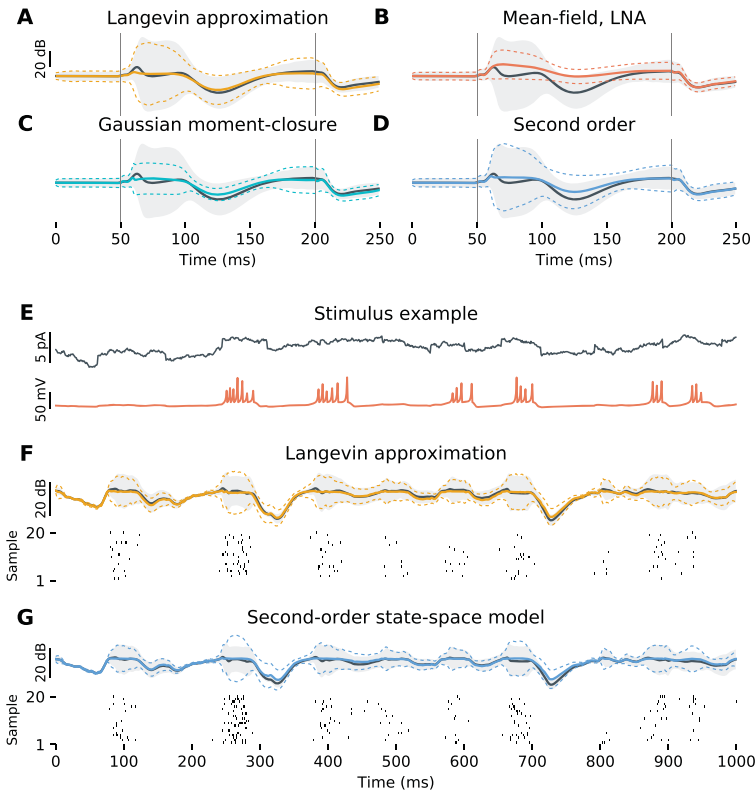
Figure 3: Moment closure captures slow timescales in the mean and fast timescales in the variance. Four approaches (A–D) for approximating the mean (black trace) and variance (shaded, $1\sigma$) of the log intensity of the autoregressive PPGLM phasic bursting model, shown here in response to a 150 ms, 0.3 pA current pulse stimulus (vertical black lines). (A) The Langevin equation (see equation 2.9) retains essential slow-timescale features of point process, but moments must be estimated via computationally intensive Monte Carlo sampling. (B) The mean-field limit with linear noise approximation (LNA; see equation B.11) cannot capture the effects of fluctuations on the mean. (C) Gaussian moment closure (see equations 2.12 and B.9) captures the influence of second-order statistics on the evolution of the mean, but underestimates the variance owing to incorrectly modeled skewness. (D) A second-order approximation (see equations 13 and B.10) better captures the second moment. (E) Example stimulus (black) and Izhikevich voltage response (red). (F) Bursts of spiking are captured by increases in variance in the autoregressive PPGLM (mean: black, $1\sigma$: shaded). Spikes sampled (bottom) from the conditionally OU Langevin approximation (yellow) retain the phasic bursting character. (G) The state-space model derived from moment closure on the Langevin approximation retains essential characteristics of the original system.

Table 1: Moment-Closure Methods Capture Process Moments More Accurately than Mean-Field with LNA (MF/LNA).

|  | Normalized RMSE | | | |
|---|---|---|---|---|
|  | Langevin | MF/LNA | Gaussian MC | Second Order |
| Log mean rate | 0.30 | 0.36 | 0.34 | 0.31 |
| Mean log rate | 0.33 | 0.93 | 0.42 | 0.47 |
| Log rate $\sigma$ | 0.41 | 0.90 | 0.86 | 0.53 |

Notes: Accuracy of various methods compared to Monte Carlo sampling (10,000 samples) of the autoregressive PPGLM model for a phasic bursting neuron, simulated at $dt = 1$ ms. Error is reported as root-mean-squared-error (RMSE) divided by the standard deviation of the relevant signal. The mean rate is captured with similar accuracy for all methods (first row, error computed on the logarithm of the mean rate), with the Langevin and second-order moment equations being the best. The average log rate is captured best by gaussian moment closure (Gaussian MC), whereas the log mean rate and standard deviation of the log rate are handled more accurately by the second-order moment equations. Errors were computed using a stimulus with a baseline inhibitory current of $-0.5$ pA and with of 49 current pulses, amplitudes ranging from 0.5 to 1.0 pA and durations ranging from 50 to 500 ms, with added OU noise with a steady-state variance of $1\,\text{pA}^2$ and time-constant of 100 ms.

yielding the following moment equations:

$$\partial_t \mu = \delta_{\tau=0} \tilde{\lambda} - \partial_\tau \mu,$$

$$\bar{\lambda} = \exp\left(H^\top \mu + I(t)\right),$$

$$\tilde{\lambda} = \bar{\lambda} \cdot (1 + \frac{1}{2} H^\top \Sigma H),$$

$$\partial_t \Sigma_z = J\Sigma + \Sigma J^\top + Q,$$

$$J = \delta_{\tau=0} \bar{\lambda} H^\top - \partial_\tau,$$

$$Q = \delta_{\tau=0} \tilde{\lambda} \delta_{\tau=0}^\top. \tag{2.13}$$

This second-order gaussian moment closure is not only more accurate in the second moment (see Figure 3D and Table 1), it is also less stiff and easier to integrate. This highlights a major benefit of the moment-closure approach: numerical issues that prove difficult or intractable in the original GLM representation can be more readily addressed in a state-space model. Importantly, the basis-projected moment-closure system (see appendix B) is an ordinary differential equation with a form reminiscent of nonlinear (extended) continuous-time Kalman-Bucy filtering (Kalman & Bucy, 1961) and can be viewed as a state-space model (see equations 2.4 and 2.5) explaining

the observed point process. This highlights that the moment-closure state-space equations allow tools for reasoning about the stability of ordinary differential equations to be applied to PPGLMs.

## 3  Discussion

In this letter, we have introduced a mathematical connection between PPGLMs and SSMs that provides an explicit, constructive procedure to fit neural spike train data. Autoregressive point processes and state-space models have been combined before (Zhao & Park, 2016; Smith & Brown, 2003; Lawhern et al., 2010; Eden, Frank, Barbieri, Solo, & Brown, 2004), but so far always in a manner that treats the latent state-space as an extrinsic driver of neural activity. Importantly, the generative, dynamical effects of point-process fluctuations and process autocorrelations are not directly addressed in previous approaches. Additionally, although PPGLMs can condition on population spiking history during training, this conditioning addresses only a single sample path of the process history and does not reflect a recurrent dynamical model in which spiking outputs and their fluctuations lead to the emergence of collective dynamics. The moment-closure approach outlined here can be used to incorporate autohistory effects into a latent state-space model, where conditioning on the process history is replaced by a Bayesian filtering update that updates the moments of the history process at each time step.

Our results highlight the capacity of PPGLMs to implicitly learn hidden causes of neural firing through the autoregressive history filter. For example, a collective network mode at 20 Hz may induce spiking rhythmicity that is detected in the point-process history filter, even if isolated neurons do not exhibit this oscillation. The history dependence of autoregressive PPGLMs defines a latent variable process that captures both process auto-history effects and the influence of unobserved hidden modes or inputs. The moments of the population spiking history can be identified with the latent variables explaining population spiking. This interpretation replaces pairwise coupling in a neural population in the PPGLM formulation with a coupling of single neurons to a shared latent-variable history process.

The identification of latent states with moments of the population history opens up a new interpretation connecting both PPGLMs and latent state-space models to neural field models, a rich area of research in theoretical neuroscience (Amari, 1975, 1977, 1983; Wilson et al., 1972). It suggests that under some conditions, neural field models may be interpreted as latent variable models and fit using modern techniques for latent state-space models. Conversely, this new connection illustrates that some latent-state space models may be viewed not only as modeling hidden causes of spiking, but also as capturing statistical moments of the population that are relevant for neural dynamics in a neural field sense. The precise convergence of a moment closure PPGLM to a neural field model remains to be better

explored mathematically. Convergence of moment closure approximations has been studied extensively in the area of stochastic chemical reactions (Schnoerr, Sanguinetti, & Grima, 2014; Schnoerr et al., 2017). Indeed, our approach was partly inspired by recent work on chemical reaction-diffusion systems (Schnoerr, Grima, & Sanguinetti, 2016), in which pairwise interactions between pointwise agents in space are replaced by a coupling of single agents to a statistical field. In contrast to chemical reaction systems, however, we model self-interactions of a point process over time and also capture the effects of fluctuations on the system.

There are two major benefits of the moment-closure representation of PPGLMs. First, autoregressive time dependencies are converted to a low-dimensional system of ordinary differential equations, reinterpreting the PPGLM as a dynamical latent-state-space model of a similar form as phenomenological latent-dynamics models and population neural field models. Second, moment-closure equations open up new strategies for estimating PPGLMs. A major challenge to fitting PPGLMs to large populations is estimating a large number of pairwise interactions. Our work suggests a different avenue toward estimating such large models: a low-dimensional latent-variable stochastic process with a suitable nonlinearity, and Poisson noise can be interpreted as a process governing the moments of an PPGLM model. This allows the extensive methodological advancements toward identifying low-dimensional state-space models to be applied to autoregressive point processes. For example, efficient estimation methods exist for recurrent linear models (Pachitariu, Petreska, & Sahani, 2013), which can be thought of as a discrete-time, mean-field limit of the system derived here. The interaction of the moment-closure formulation with recent advances for estimating PPGLMs (Sahani, Bohner, & Meyer, 2016; Ramirez & Paninski, 2014) also remains to be explored. In addition to establishing a formal connection to the stochastic process defined by PPGLMs, moment closure provides a second-order model of fluctuations and correlations. This could be especially useful in systems in which the spiking output, and fluctuations therein, influences the population dynamics.

Another challenge in estimating PPGLMs is ensuring that the fitted model accurately captures dynamics (Hocker & Park, 2017; Gerhard et al., 2017). The moment-closure equations outlined here allow process moments

explored mathematically. Convergence of moment closure approximations has been studied extensively in the area of stochastic chemical reactions (Schnoerr, Sanguinetti, & Grima, 2014; Schnoerr et al., 2017). Indeed, our approach was partly inspired by recent work on chemical reaction-diffusion systems (Schnoerr, Grima, & Sanguinetti, 2016), in which pairwise interactions between pointwise agents in space are replaced by a coupling of single agents to a statistical field. In contrast to chemical reaction systems, however, we model self-interactions of a point process over time and also capture the effects of fluctuations on the system.

Here, we employed a gaussian moment closure, but other distributional assumptions may lead to even more accurate state-space formulations. Gaussian moment closure neglects the contribution of higher moments (e.g., skewness) that may arise owing to the influence of Poisson (spiking) noise. More generally, the expected rate (see equation 2.11) is connected to the expected log-likelihood integral for PPGLMs (Park & Pillow, 2011; Ramirez & Paninski, 2014). Other distributional assumptions can be used to compute this integral by matching the first two moments provided from the gaussian moment-closure system. Alternatively, moment equations could also be derived using an alternative parameterization of the history process, for example, log-gaussian.

There are two major benefits of the moment-closure representation of PPGLMs. First, autoregressive time dependencies are converted to a low-dimensional system of ordinary differential equations, reinterpreting the PPGLM as a dynamical latent-state-space model of a similar form as phenomenological latent-dynamics models and population neural field models. Second, moment-closure equations open up new strategies for estimating PPGLMs. A major challenge to fitting PPGLMs to large populations is estimating a large number of pairwise interactions. Our work suggests a different avenue toward estimating such large models: a low-dimensional latent-variable stochastic process with a suitable nonlinearity, and Poisson noise can be interpreted as a process governing the moments of an PPGLM model. This allows the extensive methodological advancements toward identifying low-dimensional state-space models to be applied to autoregressive point processes. For example, efficient estimation methods exist for recurrent linear models (Pachitariu, Petreska, & Sahani, 2013), which can be thought of as a discrete-time, mean-field limit of the system derived here. The interaction of the moment-closure formulation with recent advances for estimating PPGLMs (Sahani, Bohner, & Meyer, 2016; Ramirez & Paninski, 2014) also remains to be explored. In addition to establishing a formal connection to the stochastic process defined by PPGLMs, moment closure provides a second-order model of fluctuations and correlations. This could be especially useful in systems in which the spiking output, and fluctuations therein, influences the population dynamics.

Another challenge in estimating PPGLMs is ensuring that the fitted model accurately captures dynamics (Hocker & Park, 2017; Gerhard et al., 2017). The moment-closure equations outlined here allow process moments

to be estimated, along with model likelihood, using Bayesian filtering. In addition to filtering over a distribution of paths in the process history, filtering can also average over models and thus implicitly capture both fluctuation effects and model uncertainty. However, it remains the subject of future work to apply the moment-closure approach in inference. Other methods, such as particle filtering, may be useful in situations where the latent state-space distribution is highly nongaussian.

**Appendix A: Time Evolution of the Second Moment** ———————

We work with the time evolution of the covariance $\Sigma$, rather than the second moment $\langle hh^\top \rangle$, for improved stability in numerical implementations:

$$\Sigma = \langle hh^\top \rangle - \langle h \rangle \langle h \rangle^\top. \tag{A.1}$$

Differentiating the covariance,

$$
\begin{aligned}
\partial_t \Sigma &= \partial_t (\langle hh^\top \rangle - \langle h \rangle \langle h \rangle^\top) \\
&= \partial_t \langle hh^\top \rangle - \partial_t (\langle h \rangle \langle h \rangle^\top) \\
&= \langle (\partial_t h)h^\top \rangle + \langle h(\partial_t h^\top) \rangle - (\partial_t \langle h \rangle)\langle h \rangle^\top - \langle h \rangle(\partial_t \langle h \rangle^\top).
\end{aligned} \tag{A.2}
$$

This expression consists of two sets of symmetric terms arising from the product rule. Examine one set of terms, and substitute in the delay line evolution, equation 2.6:

$$
\begin{aligned}
\langle (\partial_t h)h^\top \rangle - (\partial_t \langle h \rangle)\langle h \rangle^\top &= \langle [\delta_{\tau=0}\lambda - \partial_\tau h]h^\top \rangle - [\delta_{\tau=0}\langle \lambda \rangle - \partial_\tau \langle h \rangle]\langle h \rangle^\top \\
&= \delta_{\tau=0}[\langle \lambda h^\top \rangle - \langle \lambda \rangle \langle h \rangle^\top] - \partial_\tau [\langle hh^\top \rangle - \langle h \rangle \langle h \rangle^\top].
\end{aligned} \tag{A.3}
$$

This expression is linear in the first two moments, except for the expectation $\langle \lambda h^\top \rangle$. This expectation is taken over the gaussian history process with mean $\langle h \rangle$ and covariance $\Sigma$ and can be computed by completing the square using $m = \langle h \rangle + \Sigma H$ in the gaussian integral:

$$
\begin{aligned}
\langle \lambda h^\top \rangle &= \left\langle h^\top e^{H^\top h + I} \right\rangle \\
&= e^{I(t)} \int_{dh} h e^{H^\top h} \frac{1}{\sqrt{|2\pi \Sigma|}} e^{-\frac{1}{2}(h - \langle h \rangle)^\top \Sigma^{-1}(h - \langle h \rangle)} \\
&= e^{I(t)} e^{\frac{1}{2}(m^\top \Sigma^{-1} m - \langle h \rangle^\top \Sigma^{-1} \langle h \rangle)} \cdot m^\top \\
&= e^{H^\top \langle h \rangle + I(t) + \frac{1}{2}H^\top \Sigma H} \cdot m^\top \\
&= \langle \lambda \rangle (\langle h \rangle + \Sigma H)^\top. 
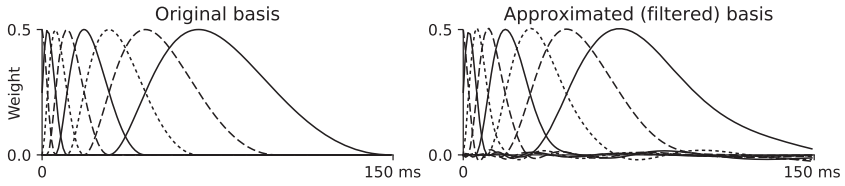\end{aligned} \tag{A.4}
$$

Figure 4: Basis projection of the history process yields a linear filter approximating the original history basis elements. (Left) History dependence in autoregressive point-process models is typically regularized by using a finite history basis. (Right) To convert history basis elements into a linear dynamical system, one projects the infinite-dimensional delay line (see equation 2.6) onto the low-dimensional basis. The resulting linear system has response functions that approximate the history basis. Note, however, the ringing introduced by the approximation.

Substituting the above expression into equation A.3 and simplifying yields the deterministic contribution to the evolution of the covariance:

$$\langle(\partial_t h)h^\top\rangle - (\partial_t \langle h\rangle) \langle h\rangle^\top = \delta_{\tau=0}\left(\langle\lambda\rangle\left(\langle h\rangle + \Sigma H\right)^\top - \langle\lambda\rangle \langle h\rangle^\top\right) - \partial_\tau \Sigma$$

$$= \underbrace{\left(\delta_{\tau=0} \langle\lambda\rangle H^\top - \partial_\tau\right)}_{J} \Sigma. \tag{A.5}$$

**Appendix B: Basis Projection**

The history process $h(\tau, t)$ is infinite dimensional. To make inference and simulation practical, one represents the continuous history filter $H(\tau)$ by a finite collection of basis functions $B(\tau) = \{B_1(\tau), .., B_K(\tau)\}$ (see Figure 4). A common choice is to use a cosine basis, for example, from Weber and Pillow (2017):

$$B_j(t) = \frac{1}{2}\cos(a\log[t + c] - \phi_j) + \frac{1}{2},$$

where parameters $a$ and $c$ select the base and offset of the functions in log time, respectively, and $\phi_j$ are offsets in integer multiples of $\pi/2$. This basis projection moves us from an infinite-dimensional history $h(\tau, t)$ to a finite state space $z(t) = \{z_1(t), .., z_k(t)\}$ defined by the projection $B(\tau) = \{B_0(\tau), .., B_k(\tau)\}$ of $h(\tau, t)$:

$$z_i(t) = \int B_i(\tau)h(\tau, t)d\tau. \tag{B.1}$$

$B$ should be normalized so that volume is preserved at every time $\tau$ (i.e., $\forall \tau, \sum_i B_i(\tau) = 1$) so that the history basis features can be treated as Poisson random variables. In practice, the history will not extend for infinite time, and the final basis functions may be omitted. The continuous history filter $H(\tau)$ is replaced by discrete weights $\beta_i = \int_\tau H(\tau)B_i(\tau)$:

$$H(\tau)^\top h(\tau, t) = \int_0^\infty H(\tau)h(\tau, t)d\tau$$
$$\approx \sum_i \beta_i \int B_i(\tau)h(\tau, t)d\tau. \tag{B.2}$$

The time evolution of $z(t)$ can be written in terms of $h(\tau, t)$:

$$\partial_t z(t) = \partial_t Bh(\tau, t)$$
$$= B\partial_t h(\tau, t)$$
$$= -B\partial_\tau h(\tau, t) + B\delta_{\tau=0}y(t). \tag{B.3}$$

We can approximately recover the state of the delay line $h(\tau, t)$ from the basis projection using the Moore-Penrose pseudoinverse of the basis $B^+$:

$$h(\tau, t) \approx B^+ z(t) = \sum_i z_i(t)B_i(t - \tau). \tag{B.4}$$

This yields a closed approximate dynamical system for computing the convolution of the history basis $B$ with a signal $y(t)$:

$$\partial_t z \approx \partial_t \tilde{z} = -B\partial_\tau B^+ \tilde{z} + B\delta_{\tau=0}y(t). \tag{B.5}$$

This is a finite-dimensional linear system $\tilde{z}$ that approximates the history using basis projection:

$$\partial_t \tilde{z} = Cy(t) - A\tilde{z},$$
$$A = B\partial_\tau B^+,$$
$$C = B\delta_{\tau=0}. \tag{B.6}$$

In silico, the differentiation $\partial_\tau$ and Dirac delta $\delta_{\tau=0}$ operators are implemented as matrices representing the discrete derivative and a point mass over one time step, respectively. The above basis projection then yields low-dimensional linear operators defining a dynamical system. The resulting process is

$$y(t) \sim \text{Poisson}(\lambda),$$
$$\lambda(t) = \exp\left(\beta^\top \tilde{z}(t) + I(t)\right),$$
$$\partial_t \tilde{z}(t) = Cy(t) - A\tilde{z}(t). \tag{B.7}$$

The basis projections integrate over an extended time window. If the intensity $\lambda(t)$ is approximately constant during this time window, the basis-projected history variables $z(t) = (z_1(t), .., z_k(t))$ are Poisson variables with rate and variance $z_i(t)$. These projections, by virtue of integrating over longer timescales, can be approximated as gaussian. Fluctuations that are far from gaussian in the point-process $y(t)$ can be well approximated as gaussian (with mean equal to variance) projections of the history process. In this case, we may approximate the Poisson process as a Wiener process that is continuous in time:

$$dz(t) = [C\lambda(t) - Az(t)]\, dt + C\sqrt{\lambda(t)}dW,$$
$$\lambda(t) = \exp\left(\beta^\top z(t) + I(t)\right). \tag{B.8}$$

Analogous to the moment closure for the infinite-dimensional system, (see equation 2.12 and Figure 3C), one can derive a gaussian moment closure for the low-dimensional basis-projected system. In the equations that follow, denote the deterministic mean rate (without covariance corrections) as $\bar{\lambda} = \exp(\beta^\top \mu_z + I(t))$. Using a gaussian moment closure, the equations for the evolution of the mean and second moment in the finite basis projection are:

$$\partial_t \mu_z = C\langle\lambda\rangle - A\mu_z,$$
$$\langle\lambda\rangle = \bar{\lambda}\exp\left(\frac{1}{2}\beta^\top\Sigma_z\beta\right),$$
$$\partial_t \Sigma_z = J\Sigma_z + \Sigma_z J^\top + Q,$$
$$J = C\langle\lambda\rangle\beta^\top - A,$$
$$Q = C\langle\lambda\rangle C^\top. \tag{B.9}$$

Similarly, the second-order moment closure (see equation 2.13 and Figure 3D) in the basis projection is given by

$$\partial_t \mu_z = C\tilde{\lambda} - A\mu_z,$$
$$\tilde{\lambda} = \bar{\lambda}\cdot\left(1 + \frac{1}{2}\beta^\top\Sigma_z\beta\right),$$
$$\partial_t \Sigma_z = J\Sigma_z + \Sigma_z J^\top + Q,$$
$$J = C\bar{\lambda}\beta^\top - A,$$
$$Q = C\tilde{\lambda}C^\top. \tag{B.10}$$

For comparison, a simpler alternative to moment closure is linear noise approximation (LNA; see Figure 3B), which uses a deterministic mean $\bar{\lambda}$ obtained in the limit of a large, weakly coupled population for which the effect of fluctuations on the mean is negligible (Toyoizumi et al., 2009;

Schnoerr et al., 2017). The LNA describes the second moment as a function of this mean but does not correct for the influence of fluctuations. In the finite basis, the LNA about the deterministic mean-rate $\bar{\lambda}$ is

$$\partial_t \mu_z = C\bar{\lambda} - A\mu_z,$$
$$\partial_t \Sigma_z = J\Sigma_z + \Sigma_z J^\top + Q,$$
$$J = C\bar{\lambda}\beta^\top - A,$$
$$Q = C\bar{\lambda}C^\top. \tag{B.11}$$

## Appendix C:  Numerical Implementation

The basis-projected moment equations (see equations B.9–B.11) can be integrated using forward Euler or other integration methods; code for integrating the moment equations and generating the figures in this letter is published online (Rule, 2018). We can also convert them to a form resembling the discrete-time extended Kalman-Bucy filter (Kalman & Bucy, 1961) by integrating the locally linearized system forward one time step via exponentiation. Consider, for example, a discrete-time version of the gaussian moment-closure system (see equation B.9):

$$\mu_{t+\Delta t} = F\mu_t + C\langle\lambda\rangle\,\Delta t,$$
$$\Sigma_{t+\Delta t} = G\Sigma_t G^\top + Q\Delta t, \tag{C.1}$$

where $F = \exp(-A \cdot \Delta t)$ is the discrete-time mean update, which is constant in time, and $G = \exp(J \cdot \Delta t)$ is the discrete-time covariance update, in which $J$ is the Jacobian as in equations B.9 to B.11 and depends on the current mean.

We assess the accuracy of approximations to the original PP-GLM as a stochastic process by comparing the single-time marginals for the mean log rate, the log mean rate, and the standard deviation of the log rate (see Table 1). Ground-truth estimates of these moments are sampled from the "true" PPGLM using Monte Carlo sampling (10,000 samples) and compared to the predictions from the LNA, Langevin approximation, gaussian moment closure, and second-order moment closure. Of the moment equations, the second-order moment closure best recovers the variance, and the gaussian moment closure best recovers the mean.

Model likelihoods are also a widely used measure of model fit in the point-process literature (see Fernandes, Stevenson, Phillips, Segraves, & Kording, 2013; Ramirez & Paninski, 2014); these are appropriate for the singly stochastic interpretation of the PPGLMs, which treats the latent log intensity as a deterministic, nonstationary parameter. However, the true likelihood of the PPGLM as a doubly stochastic process in principle would result from a computationally intractable marginalization of the latent variables, and it is thus not readily accessible. As a coarse approximation,

Table 2: Moment-Closure Methods Interpret the PPGLM as a Doubly Stochastic Process, Introducing a Penalty for Incorrectly Modeled Dynamics.

| | Log Likelihood | | | |
| --- | --- | --- | --- | --- |
| | GLM | MF/LNA | Gaussian MC | Second Order |
| l.l. (bits/ms) | 0.055 | 0.024 | 0.013 | 0.020 |
| Penalty (bits/ms) | - | 0.031 | 0.042 | 0.034 |
| Relative penalty | - | 1.09 | 1.51 | 1.23 |

Notes: Here, we use Bayesian filtering to estimate the log likelihood of a PPGLM model (reported as normalized log-likelihood "l.l." in bits/ms, first row; Ramirez & Paninski, 2014). The model is originally fit using maximum-likelihood GLM regression (first column), which neglects the dynamical effects of fluctuations. All three moment approximations broadly agree on a likelihood penalty (0.031–0.042 bits/sample), the relative size of which (last row; Fernandes et al., 2013) is on the same order as the distance between the GLM estimated likelihood and the theoretical maximum "saturated" normalized log likelihood of 0.083 bits/ms.

however, moment equations (see equations B.9–B.11) allow us to estimate a likelihood penalty related to slow dynamics (see Table 2). We estimated likelihood penalties via Bayesian filtering on the moment equations as a state-space model, handling the nonconjugate log-gaussian Poisson measurement update via moment matching. All three moment approximations broadly agree on the magnitude of the likelihood penalty. This slow-timescale likelihood estimate could potentially be merged with fast dynamics to build robust, dynamically accurate estimators for autoregressive PPGLM models, and approximation methods for the likelihood of the doubly stochastic interpretation of the PPGLM remain to be explored in future work.

**Acknowledgments**

**References**

Aghagolzadeh, M., & Truccolo, W. (2014). Latent state-space models for neural decoding. In *Proceedings of the Engineering in Medicine and Biology Society 36th Annual International Conference of the IEEE* (pp. 3033–3036). Piscataway, NJ: IEEE.

Aghagolzadeh, M., & Truccolo, W. (2016). Inference and decoding of motor cortex low-dimensional dynamics via latent state-space models. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 24(2), 272–282.

Ale, A., Kirk, P., & Stumpf, M. P. (2013). A general moment expansion method for stochastic kinetic models. *Journal of Chemical Physics*, 138(17), 174101.

Amari, S.-I. (1975). Homogeneous nets of neuron-like elements. *Biol. Cybern.*, 17(4), 211–220.

Amari, S.-I. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biol. Cybern.*, 27(2), 77–87.

Amari, S.-I. (1983). Field theory of self-organizing neural nets. *IEEE Trans. Syst. Man. Cybern.*, SMC-13(5), 741–748.

Brillinger, D. R. (1988). Maximum likelihood analysis of spike trains of interacting nerve cells. *Biological Cybernetics*, 59(3), 189–200.

Buesing, L., Macke, J. H., & Sahani, M. (2012). Spectral learning of linear dynamics from generalised-linear observations with application to neural population data. In F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems, 35* (pp. 1682–1690). Red Hook, NY: Curran.

Chevallier, J., Duarte, A., Löcherbach, E., & Ost, G. (2017). *Mean field limits for nonlinear spatially extended Hawkes processes with exponential memory kernels*. arXiv:1703.05031.

Churchland, M. M., Cunningham, J. P., Kaufman, M. T., Foster, J. D., Nuyujukian, P., Ryu, S. I., & Shenoy, K. V. (2012). Neural population dynamics during reaching. *Nature*, 487(7405), 51–56.

Cseke, B., Zammit-Mangion, A., Heskes, T., & Sanguinetti, G. (2016). Sparse approximate inference for spatio-temporal point process models. *Journal of the American Statistical Association*, 111(516), 1746–1763.

Delarue, F., Inglis, J., Rubenthaler, S., & Tanré, E. (2015). Particle systems with a singular mean-field self-excitation: Application to neuronal networks. *Stochastic Processes and Their Applications*, 125(6), 2451–2492.

Eden, U. T., Frank, L. M., Barbieri, R., Solo, V., & Brown, E. N. (2004). Dynamic analysis of neural encoding by point process adaptive filtering. *Neural Computation*, 16(5), 971–998.

Fernandes, H. L., Stevenson, I. H., Phillips, A. N., Segraves, M. A., & Kording, K. P. (2013). Saliency and saccade encoding in the frontal eye field during natural scene search. *Cerebral Cortex*, 24(12), 3232–3245.

Galves, A., & Löcherbach, E. (2015). *Modeling networks of spiking neurons as interacting processes with memory of variable length*. arXiv:1502.06446.

Gao, Y., Archer, E. W., Paninski, L., & Cunningham, J. P. (2016). Linear dynamical neural population models through nonlinear embeddings. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, & R. Garnett (Eds.), *Advances in neural information processing systems, 29* (pp. 163–171). Red Hook, NY: Curran.

Gerhard, F., Deger, M., & Truccolo, W. (2017). On the stability and dynamics of stochastic spiking neuron models: Nonlinear Hawkes process and point process GLMS. *PLoS Computational Biology*, 13(2), e1005390.

Gillespie, D. T. (2000). The chemical Langevin equation. *Journal of Chemical Physics*, 113(1), 297–306.

Hocker, D., & Park, I. M. (2017). Multistep inference for generalized linear spiking models curbs runaway excitation. In *Proceedings of the 8th International IEEE/EMBS Conference on Neural Engineering* (pp. 613–616). Piscataway, NY: IEEE.

Izhikevich, E. M. (2003). Simple model of spiking neurons. *IEEE Transactions on Neural networks*, *14*(6), 1569–1572.

Jun, J. J., Steinmetz, N. A., Siegle, J. H., Denman, D. J., Bauza, M., Barbarits, B., … Barbic, M. (2017). Fully integrated silicon probes for high-density recording of neural activity. *Nature*, *551*(7679), 232.

Kalman, R. E., & Bucy, R. S. (1961). New results in linear filtering and prediction theory. *Journal of Basic Engineering*, *83*(1), 95–108.

Kingman, J. F. C. (1993). *Poisson processes*. New York: Wiley.

Lawhern, V., Wu, W., Hatsopoulos, N., & Paninski, L. (2010). Population decoding of motor cortical activity using a generalized linear model with hidden states. *Journal of Neuroscience Methods*, *189*(2), 267–280.

Maccione, A., Hennig, M. H., Gandolfo, M., Muthmann, O., Coppenhagen, J., Eglen, S. J., … Sernagor, E. (2014). Following the ontogeny of retinal waves: Pan-retinal recordings of population dynamics in the neonatal mouse. *Journal of Physiology*, *592*(7), 1545–1563.

Macke, J. H., Buesing, L., Cunningham, J. P., Yu, B. M., Shenoy, K. V., & Sahani, M. (2011). Empirical models of spiking in neural populations. In J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems*, *24* (pp. 1350–1358). Red Hook, NY: Curran.

Macke, J. H., Buesing, L., & Sahani, M. (2015). Estimating state and parameters in state space models of spike trains. In Z. Chen (Ed.), *Advanced state space methods for neural and clinical data* (p. 137). Cambridge: Cambridge University Press.

Michaels, J. A., Dann, B., Intveld, R. W., & Scherberger, H. (2017). *Neural dynamics of variable grasp movement preparation in the macaque fronto-parietal network*. bioRxiv, 179143.

Pachitariu, M., Petreska, B., & Sahani, M. (2013). Recurrent linear models of simultaneously-recorded neural populations. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems, 26* (pp. 3138–3146). Red Hook, NY: Curran.

Park, I. M., & Pillow, J. W. (2011). Bayesian spike-triggered covariance analysis. In J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems*, *24* (pp. 1692–1700). Red Hook, NY: Curran.

Pfau, D., Pnevmatikakis, E. A., & Paninski, L. (2013). Robust learning of low-dimensional dynamics from large neural ensembles. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems* (pp. 2391–2399). Red Hook, NY: Curran.

Ramirez, A. D., & Paninski, L. (2014). Fast inference in generalized linear models via expected log-likelihoods. *Journal of Computational Neuroscience*, *36*(2), 215–234.

Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *71*(2), 319–392.

Rule, M. E. (2018). Code to accompany autoregressive point-processes as latent state-space models. github.com/michaelerule/glm_moment_closure/releases/tag/v1.0.1; doi: 10.5281/zenodo.1247109

Rule, M. E., Vargas-Irwin, C., Donoghue, J. P., & Truccolo, W. (2015). Contribution of LFP dynamics to single neuron spiking variability in motor cortex during movement execution. *Frontiers in Systems Neuroscience*, *9*, 89.

Rule, M. E., Vargas-Irwin, C. E., Donoghue, J. P., & Truccolo, W. (2017). Dissociation between sustained single-neuron spiking $\beta$-rhythmicity and transient $\beta$-LFP oscillations in primate motor cortex. *Journal of Neurophysiology*, *11*, 1524–1543.

Sahani, M., Bohner, G., & Meyer, A. (2016). Score-matching estimators for continuous-time point-process regression models. In *Proceedings of the IEEE 26th International Workshop on Machine Learning for Signal Processing* (pp. 1–5). Piscataway, NJ: IEEE.

Schnoerr, D., Grima, R., & Sanguinetti, G. (2016). Cox process representation and inference for stochastic reaction-diffusion processes. *Nat. Commun.*, *7*, 11729.

Schnoerr, D., Sanguinetti, G., & Grima, R. (2014). Validity conditions for moment closure approximations in stochastic chemical kinetics. *Journal of Chemical Physics*, *141*(8). https://aip.scitation.org/doi/abs/10.1063/1.4892836.

Schnoerr, D., Sanguinetti, G., & Grima, R. (2015). Comparison of different moment-closure approximations for stochastic chemical kinetics. *Journal of Chemical Physics*, *143*(18), 185101.

Schnoerr, D., Sanguinetti, G., & Grima, R. (2017). Approximation and inference methods for stochastic biochemical kinetics—a tutorial review. *Journal of Physics A: Mathematical and Theoretical*, *50*(9), 093001.

Smith, A. C., & Brown, E. N. (2003). Estimating a state-space model from point process observations. *Neural Computation*, *15*(5), 965–991.

Surace, S. C., Kutschireiter, A., & Pfister, J.-P. (2017). *How to avoid the curse of dimensionality: Scalability of particle filters with and without importance weights*. arXiv:1703.07879.

Sussillo, D., Jozefowicz, R., Abbott, L., & Pandarinath, C. (2016). *LFADS—latent factor analysis via dynamical systems*. arXiv:1608.06315.

Toyoizumi, T., Rad, K. R., & Paninski, L. (2009). Mean-field approximations for coupled populations of generalized linear model spiking neurons with Markov refractoriness. *Neural Computation*, *21*(5), 1203–1243.

Truccolo, W. (2010). Stochastic models for multivariate neural point processes: Collective dynamics and neural decoding. In S. Grün & S. Rotter (Eds.), *Analysis of parallel spike trains* (pp. 321–341). New York: Springer.

Truccolo, W. (2016). From point process observations to collective neural dynamics: Nonlinear hawkes process GLMS, low-dimensional dynamics and coarse graining. *Journal of Physiology–Paris*, *110*, 336–347.

Truccolo, W., Eden, U. T., Fellows, M. R., Donoghue, J. P., & Brown, E. N. (2005). A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *Journal of Neurophysiology*, *93*(2), 1074–1089.

Truccolo, W., Hochberg, L. R., & Donoghue, J. P. (2010). Collective dynamics in human and monkey sensorimotor cortex: Predicting single neuron spikes. *Nature Neuroscience*, *13*(1), 105–111.

Weber, A. I., & Pillow, J. W. (2017). Capturing the dynamical repertoire of single neurons with generalized linear models. *Neural Computation*, *29*(12), 3260–3289.

Wilson, H. R., Cowan, J. D., Baker, T., Cowan, J., & van Drongelen, W. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys. J.*, *12*(1), 1–24.

Wu, A., Roy, N. G., Keeley, S., & Pillow, J. W. (2017). Gaussian process based nonlinear latent structure discovery in multivariate spike train data. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in neural information processing systems*, *30* (pp. 3498–3507). Red Hook, NY: Curran.

Yu, B. M., Cunningham, J. P., Santhanam, G., Ryu, S. I., Shenoy, K. V., & Sahani, M. (2009). Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. In D. Koller, D. Schuurmans, Y. Bengio, & L. Bottou (Eds.), *Advances in neural information processing systems*, *21* (pp. 1881–1888).

Zhao, Y., & Park, I. M. (2016). Interpretable nonlinear dynamic modeling of neural trajectories. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, & R. Garnett (Eds.), *Advances in neural information processing systems*, *29* (pp. 3333–3341). Red Hook, NY: Curran.

Zhao, Y., & Park, I. M. (2017). Variational latent gaussian process for recovering single-trial dynamics from population spike trains. *Neural Computation*, *29*(5), 1293–1316.