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

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

Michael Rule[1], Guido Sanguinetti[1]

[1]Institute for Adaptive and Neural Computation,
School of Informatics,
University of Edinburgh
10 Crichton St
Edinburgh EH8 9AB

Keywords: Point processes, Moment closure, Mean-field

Abstract

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.

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 saliency due to the rapid improvements in technologies which 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 aetiology 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 last 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 (e.g. Macke et al. 2015; Sussillo et al. 2016; Zhao and Park 2016; Yu et al. 2009; Smith and 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 (e.g. Smith and Brown 2003; Wu et al. 2017; Gao et al. 2016; Pfau et al. 2013; Zhao and Park 2017; Surace et al. 2017). These models have been used in a variety of tasks, such as describing population spiking activity in the motor system (e.g. Aghagolzadeh and Truccolo 2014, 2016; Churchland et al. 2012; Michaels et al. 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 (Fig. 2B).

Autoregressive PPGLMs (Fig. 2A) treat spiking events from neurons as *point events* arising

from a latent inhomogeneous Poisson process (Truccolo et al., 2005, 2010; Truccolo, 2010, 2017). 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 (e.g. Rule et al. 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 and Park, 2017; Gerhard et al., 2017), and can fail to generalize outside of regimes in which they were trained (Weber and Pillow, 2017). Importantly, PPGLMs suffer from confounds if there are unobserved sources of neural variability (Lawhern et al., 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 and 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 et al., 2017; Galves and Löcherbach, 2015; Delarue et al., 2015), we use Gaussian moment-closure (e.g. Schnoerr et al. 2015, 2017) to capture the excitatory effects of fluctuations and process autocorrelations. In doing so, we convert the auto-history 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 analogue of the autoregressive PPGLM. We illustrate the approach on the case study of the phasic bursting of an Izhikevich (Izhikevich, 2003) neuron model (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.

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. We finally illustrate the performance of the method in an application case study.

Point-process Generalized Linear Models (PPGLM)

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 will only consider 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:

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

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.

$$\Pr\left(N(t{+}\Delta) - N(t) = k\right) \sim \text{Poisson}\left(\int_t^{t+\Delta} \lambda(t)dt\right) \tag{1}$$

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

$$f\left(\lambda(t)\right) = m + F^{\top}x(t) + \lim_{\epsilon \to 0^{+}} \int_{\epsilon}^{\infty} H(\tau)y(t-\tau)d\tau \tag{2}$$

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 $\epsilon \to 0^{+}$ 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. Re-writing Eq. 2, making time-dependence implicit where unambiguous, and denoting the history filter integral as $H^{\top}y$, we will explore generalized linear models of the form:

$$\lambda = \exp\left(m + F^{\top}x + H^{\top}y\right) \tag{3}$$

Autoregressive PPGLMs can emulate various firing behaviors of real neurons (Weber and Pillow, 2017). For example phasic bursting neurons (Fig. 1) exhibit complex autohistory dependence on both fast and slow timescales. This dependence of the process on intrinsic history confers additional slow dynamics, i.e. post-burst inhibition on the timescale of tens of milliseconds.

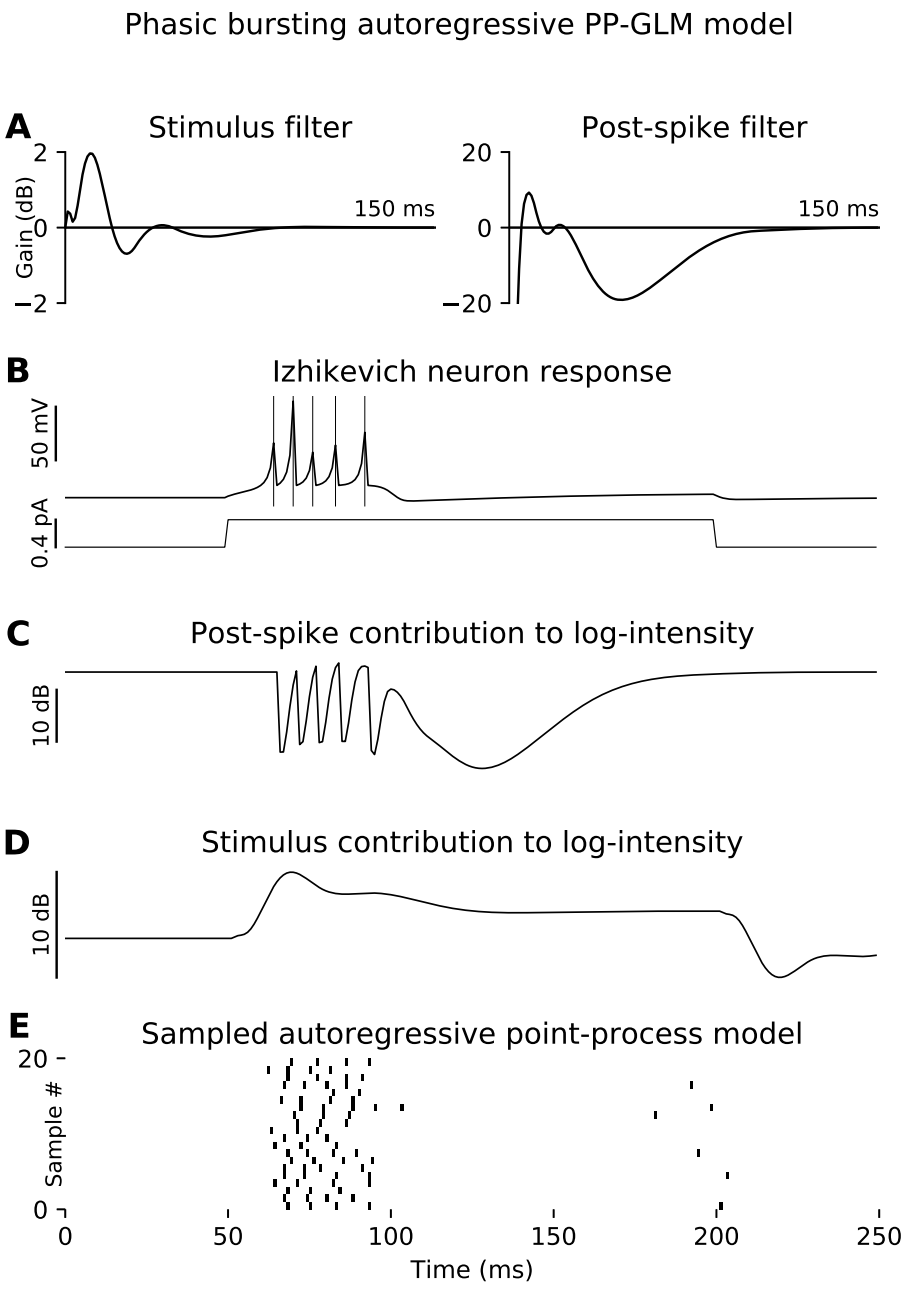

Figure 1: *Phasic bursting neuron as emulated by an autoregressive PPGLM model.* An autoregressive PPGLM was trained on spiking output from a phasic-bursting Izhikevich neuron model (**B**) with parameters $a, b, c, d, dt = 0.02, 0.25, -55, 0.05, 1.0$ (Weber and Pillow, 2017). Training stimuli consisted 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 post-spike filters exhibit ringing associated with the characteristic inter-spike interval. The post-spike filter includes an additional post-burst slow-inhibitory component, evident in the post-spike contribution to the log intensity (**C**). This slow post-burst 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.

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 post-burst inhibition. This approach is taken by latent state-space models (SSMs), which are often viewed as functionally distinct from autoregressive PPGLMs (Fig. 2B).

In general, a latent state-space model (SSM) describes how both deterministic and stochastic dynamics of a latent variable x affect the intensity λ of a point process:

$$\begin{aligned} dx(t) &= u(x,t)dt + \sigma(x,t)dW \\ \lambda &= v(x,t) \\ dN &\sim \text{Poisson}(\lambda \cdot dt), \end{aligned} \tag{4}$$

where dW is the derivative of the standard Wiener process, reflecting fluctuations. The functions u, σ, 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:

$$\begin{aligned} dx(t) &= [Ax + I(t)]dt + \sigma dW \\ \lambda &= \exp\left(m + F^\top x + H^\top y\right) \\ dN &\sim \text{Poisson}(\lambda \cdot dt), \end{aligned} \tag{5}$$

where $I(t)$ reflects inputs into the latent state-space, and the spiking probability depends on latent states, history, and bias, as in Eq. 3. Latent state-space models of point-processes have been investigated in detail (e.g. Macke et al. 2011; Smith and Brown 2003), and mature inference approaches are available to estimate states and parameters from data (Lawhern et al., 2010; Macke et al., 2015; Buesing et al., 2012; Rue et al., 2009; Cseke et al., 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 paper, we illustrate that the history dependence of PPGLMs implicitly defines a latent-state space model over moments of the process history.

The auxiliary history process of a PPGLM

The (possibly infinite) history dependence makes autoregressive PPGLMs non-Markovian dynamical systems (c.f. Truccolo 2017 Eq. 6). However, a crucial insight is that, since the dependence on history is linear, we can re-interpret 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 et al. (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, let us 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:

$$\partial_t h(\tau, t) = \delta_{\tau=0} dN(t) - \partial_\tau h(\tau, t) \tag{6}$$

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

$$\begin{aligned} dN(t) &\sim \text{Poisson}(\lambda \cdot dt) \\ \lambda(t) &= \exp\left(H(\tau)^\top h(\tau, t) + I(t)\right) \\ \partial_t h(\tau, t) &= \delta_{\tau=0} dN(t) - \partial_\tau h(\tau, t) \end{aligned} \tag{7}$$

where $H(\tau)$ is the history filter introduced in equation (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.

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:

$$dN \approx \lambda dt + \sqrt{\lambda} dW, \tag{8}$$

This approximation holds when averaging over a population of weakly-coupled neurons (Toyoizumi et al., 2009), or averaging over slow-timescales of a single neuron. This approximation is inspired by the chemical Langevin equation (Gillespie, 2000), which remains accurate also 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 (Fig. 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 (7), we obtain a continuous (in time, and in state) infinite-dimensional approximation of the PPGLM:

$$\begin{aligned} dh &= (\delta_{\tau=0}\lambda - \partial_\tau h)\, dt + \delta_{\tau=0}\sqrt{\lambda} dW \\ \lambda &= \exp\left(H^\top h + I(t)\right). \end{aligned} \tag{9}$$

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 $\lambda(t)$.

Gaussian moment closure

The SPDE in Eq. 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 Gaussianity of the history process $h(\tau, t)$. The (exact) equation for the process mean $\mu(\tau) = \langle h(\tau) \rangle$ is as follows

$$\begin{aligned} \partial_t \mu &= \partial_t \langle h \rangle \\ &= \langle \delta_{\tau=0} \lambda - \partial_\tau h \rangle \\ &= \delta_{\tau=0} \langle \lambda \rangle - \partial_\tau \mu. \end{aligned} \tag{10}$$

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

$$\langle \lambda \rangle = \exp\left(H^\top \mu + I(t) + \tfrac{1}{2} H^\top \Sigma H \right). \tag{11}$$

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:

$$
\begin{aligned}
\partial_t \mu &= \delta_{\tau=0} \langle \lambda \rangle - \partial_\tau \mu \\
\langle \lambda \rangle &= \exp\left(H^\top \mu + I(t) + H^\top \Sigma H\right) \\
\partial_t \Sigma &= J\Sigma + \Sigma J^\top + Q \\
J &= \delta_{\tau=0} \langle \lambda \rangle H^\top - \partial_\tau \\
Q &= \delta_{\tau=0} \langle \lambda \rangle \delta_{\tau=0}^\top
\end{aligned}
\tag{12}
$$

This notation resembles continuous-time Kalman-Bucy filter (Kalman and Bucy, 1961), for which $J(t)$ would be a Jacobian of the mean update, and $Q(t)$ would reflect the system noise. Equations (12) are also reminiscent of classical neural mass and neural field models (Amari 1975, 1977, 1983; Wilson et al. 1972; e.g. Fig. 2D). Unlike neural field models, however, the moment equations (12) do not arise from population averages, but from considering the expected behavior of the stochastic process describing the neural spike train (Fig. 2C).

It is worth reflecting more on this analogy, and on the limitations of the moment-closure representation. Spiking events are a 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 (e.g. Fig. 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.

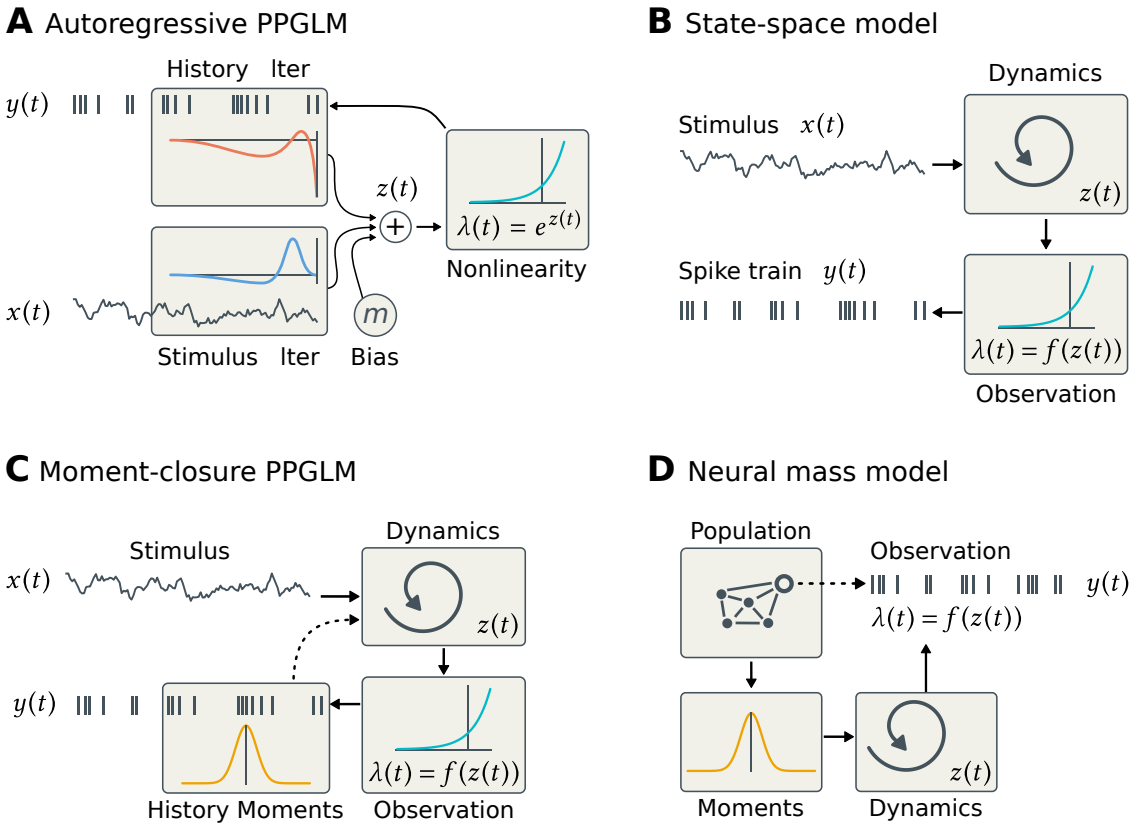

Figure 2: *Moment-closure of autoregressive PPGLMs combines aspects of three modeling approaches.* **A** Log-linear autoregressive PPGLM framework (e.g. Weber and Pillow 2017). Dependence on the history of both extrinsic covariates $x(t)$ and the process itself $y(t)$ are mediated by linear filters, 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.

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 (Appendix B). Figure 3 illustrates moment-closure of a phasic-bursting Izhikevich neuron emulated with a PPGLM (Fig. 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 and Park, 2017; Gerhard et al., 2017; Weber and 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 (Fig. 3C; Table C1), but exhibits a bias in the second moment owing to un-modeled skewness and higher-order moments. Additionally, the self-excitation mediated by the exponential inverse link function combines with fast-timescale post-spike 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 et al. (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 $\exp\left(\frac{1}{2}H^\top \Sigma H\right) \approx 1 + \frac{1}{2}H^\top \Sigma H$, yielding the following moment equations:

$$
\begin{aligned}
\partial_t \mu &= \delta_{\tau=0}\tilde{\lambda} - \partial_\tau \mu \\
\bar{\lambda} &= \exp\left(H^\top \mu + I(t)\right) \\
\tilde{\lambda} &= \bar{\lambda} \cdot (1 + \tfrac{1}{2}H^\top \Sigma H) \\
\partial_t \Sigma_z &= J\Sigma + \Sigma J^\top + Q \\
J &= \delta_{\tau=0}\bar{\lambda} H^\top - \partial_\tau \\
Q &= \delta_{\tau=0}\tilde{\lambda}\delta_{\tau=0}^\top
\end{aligned}
\tag{13}
$$

This second-order Gaussian moment closure is not only more accurate in the second moment (fig. 3D; Table C1), it is also less stiff and easier to integrate. This highlights a major benefit of the moment closure approach: numerical issues which 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 (Appendix B) is an ordinary differential equation with a form reminiscent of nonlinear (extended) continuous-time Kalman-Bucy filtering (Kalman and Bucy, 1961), and can be viewed as a state-space model (Eqs. 4, 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.

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 (e.g. Zhao and Park 2016; Smith and Brown 2003; Lawhern et al. 2010; Eden et al. 2004), but so far always in a manner that treats the latent state-space as an extrinsic driver of neural activity. Importantly, the generative, dynami-

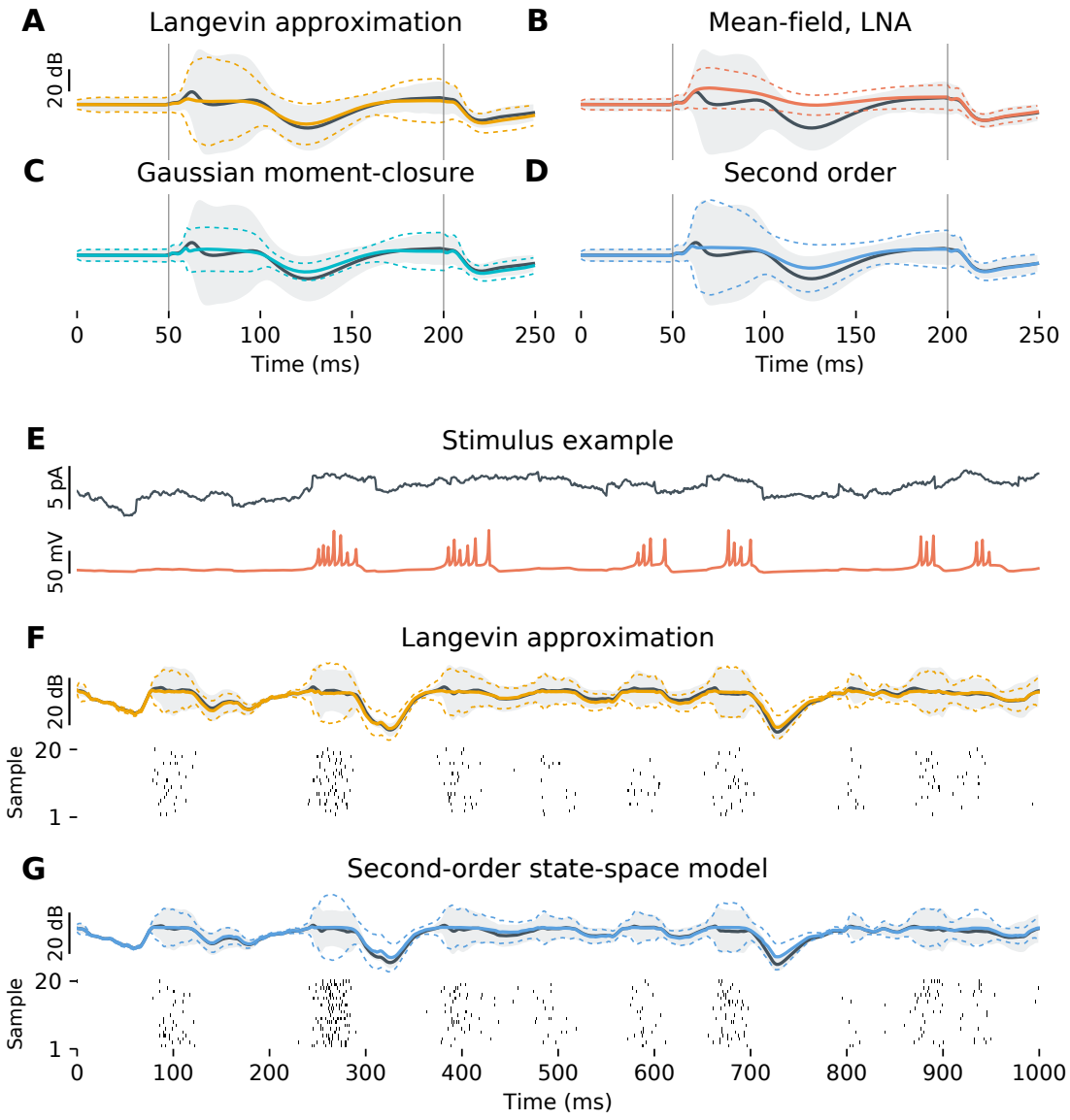

Figure 3: *Moment closure captures slow timescales in the mean, and fast timescales in the variance.* Four approaches (**A-D**) for approximating the mean (black trace) and variance (shaded, 1σ) of the log-intensity of the autoregressive PPGLM phasic bursting model, shown here in response to a 150 ms, 0.3 pA current pulse stimulus (vertical black lines). **A** The Langevin equation (Eq. 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; Eq. B11) cannot capture the effects of fluctuations on the mean. **C** Gaussian moment-closure (Eqs. 12, B9) 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 (Eqs. 13, B10) 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, 1σ: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.

cal 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 auto-history effects into a latent state-space model, where conditioning on the process history is replaced by a Bayesian filtering update which 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 et al., 2014, 2015,

2017). Indeed, our approach was partly inspired by recent work on chemical reaction-diffusion systems (Schnoerr et al., 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 capture also 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 (Eq. 11) is connected to the expected log-likelihood integral for PPGLMs (Park and Pillow, 2011; Ramirez and 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, re-interpreting 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 the challenges in 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 et al., 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 et al., 2016; Ramirez and Paninski, 2014) also

remains to be explored. In addition to establishing a formal connection to the stochastic process defined by PPGLMs, moment-closure a provides 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 and 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 non-Gaussian.

Acknowledgments

The authors acknowledge support from the Engineering and Physical Sciences Research Council of the United Kingdom under grant EP/L027208/1 *Large scale spatio-temporal point processes: novel machine learning methodologies and application to neural multi-electrode arrays*. We are indebted to David Schnoerr, Matthias Hennig, Wilson Truccolo, and the anonymous referees for valuable comments on preliminary versions of this paper.

Appendix

Appendix A: time evolution of the second moment

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

$$\Sigma = \langle hh^\top \rangle - \langle h \rangle \langle h \rangle^\top \tag{A1}$$

Differentiating the covariance:

$$\begin{aligned} \partial_t \Sigma &= \partial_t \left(\langle hh^\top \rangle - \langle h \rangle \langle h \rangle^\top \right) \\ &= \partial_t \langle hh^\top \rangle - \partial_t \left(\langle h \rangle \langle h \rangle^\top \right) \\ &= \langle (\partial_t h) h^\top \rangle + \langle h (\partial_t h^\top) \rangle - \left(\partial_t \langle h \rangle \right) \langle h \rangle^\top - \langle h \rangle \left(\partial_t \langle h \rangle^\top \right) \end{aligned} \tag{A2}$$

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 Eq. 6:

$$\begin{aligned} \langle (\partial_t h) h^\top \rangle - \left(\partial_t \langle h \rangle \right) \langle h \rangle^\top &= \langle [\delta_{\tau=0} \lambda - \partial_\tau h] h^\top \rangle - \left[\delta_{\tau=0} \langle \lambda \rangle - \partial_\tau \langle h \rangle \right] \langle h \rangle^\top \\ &= \delta_{\tau=0} \left[\langle \lambda h^\top \rangle - \langle \lambda \rangle \langle h \rangle^\top \right] - \partial_\tau \left[\langle hh^\top \rangle - \langle h \rangle \langle h \rangle^\top \right] \end{aligned} \tag{A3}$$

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 Σ, and can be computed by completing the square using $m = \langle h \rangle + \Sigma H$ in the Gaussian integral:

$$\begin{aligned} \langle \lambda h^\top \rangle &= \left\langle h^\top e^{H^\top h + I} \right\rangle \\ &= e^{I(t)} \int_{dh} h e^{H^\top h} \frac{1}{\sqrt{|2\pi\Sigma|}} e^{-\frac{1}{2}(h - \langle h \rangle)^\top \Sigma^{-1} (h - \langle h \rangle)} \\ &= e^{I(t)} e^{\frac{1}{2}(m^\top \Sigma^{-1} m - \langle h \rangle^\top \Sigma^{-1} \langle h \rangle)} \cdot m^\top \\ &= e^{H^\top \langle h \rangle + I(t) + \frac{1}{2} H^\top \Sigma H} \cdot m^\top \\ &= \langle \lambda \rangle \left(\langle h \rangle + \Sigma H \right)^\top . \end{aligned} \tag{A4}$$

Substituting the above expression into Eq. A3 and simplifying yields the deterministic contribution to the evolution of the covariance:

$$\begin{aligned}\left\langle (\partial_t h) h^\top \right\rangle - \left(\partial_t \langle h \rangle\right) \langle h \rangle^\top &= \delta_{\tau=0} \left(\langle \lambda \rangle \left(\langle h \rangle + \Sigma H \right)^\top - \langle \lambda \rangle \langle h \rangle^\top \right) - \partial_\tau \Sigma \\ &= \underbrace{\left(\delta_{\tau=0} \langle \lambda \rangle H^\top - \partial_\tau \right)}_{J} \Sigma\end{aligned} \tag{A5}$$

Appendix B: basis projection

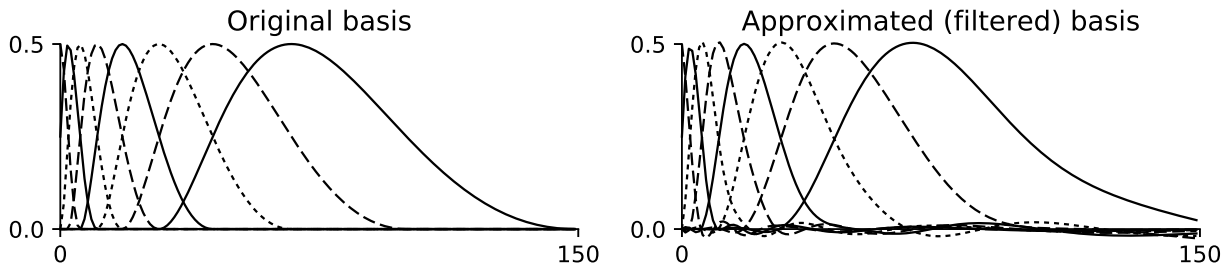

Figure B1: *Basis projection of the history process yields a linear filter approximating the original history basis elements. Left:* History dependence in autoregressive point-process models is typically regularized by using a finite history basis. *Right:* To convert history basis elements into a linear dynamical system, one projects the infinite-dimensional delay-line (Eq. 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.

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 $B(\tau) = \{B_1(\tau), .., B_K(\tau)\}$ (Fig. B1). A common choice is to use a cosine basis, for example from Weber and Pillow (2017):

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

Where parameters a and c select the base and offset of the functions in log-time, respectively, and ϕ_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 $h(\tau, t)$.

$$z_i(t) = \int B_i(\tau)h(\tau, t)d\tau \tag{B1}$$

B should be normalized so that volume is preserved at every time τ, 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 $\beta_i = \int_\tau H(\tau) B_i(\tau)$:

$$\begin{aligned} H(\tau)^\top h(\tau, t) &= \int_0^\infty H(\tau) h(\tau, t) d\tau \\ &\approx \sum_i \beta_i \int B_i(\tau) h(\tau, t) d\tau. \end{aligned} \tag{B2}$$

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

$$\begin{aligned} \partial_t z(t) &= \partial_t B h(\tau, t) \\ &= B \partial_t h(\tau, t) \\ &= -B \partial_\tau h(\tau, t) + B \delta_{\tau=0} y(t) \end{aligned} \tag{B3}$$

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^+:

$$h(\tau, t) \approx B^+ z(t) = \sum_i z_i(t) B_i(t - \tau) \tag{B4}$$

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

$$\partial_t z \approx \partial_t \tilde{z} = -B \partial_\tau B^+ \tilde{z} + B \delta_{\tau=0} y(t) \tag{B5}$$

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

$$\begin{aligned} \partial_t \tilde{z} &= C y(t) - A \tilde{z}, \\ A &= B \partial_\tau B^+ \\ C &= B \delta_{\tau=0} \end{aligned} \tag{B6}$$

In silico, the differentiation ∂_τ 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:

$$\begin{aligned} y(t) &\sim \text{Poisson}(\lambda) \\ \lambda(t) &= \exp\left(\beta^\top \tilde{z}(t) + I(t)\right) \\ \partial_t \tilde{z}(t) &= Cy(t) - A\tilde{z}(t) \end{aligned} \tag{B7}$$

The basis projections integrate over an extended time-window. If the intensity $\lambda(t)$ is approximately constant during this time window, then 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:

$$\begin{aligned} dz(t) &= \left[C\lambda(t) - Az(t)\right] dt + C\sqrt{\lambda(t)}dW \\ \lambda(t) &= \exp\left(\beta^\top z(t) + I(t)\right) \end{aligned} \tag{B8}$$

Analogously to the moment-closure for the infinite-dimensional system (Eq. 12; Fig. 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:

$$\begin{aligned} \partial_t \mu_z &= C\langle\lambda\rangle - A\mu_z \\ \langle\lambda\rangle &= \bar{\lambda}\exp\left(\tfrac{1}{2}\beta^\top \Sigma_z \beta\right) \\ \partial_t \Sigma_z &= J\Sigma_z + \Sigma_z J^\top + Q \\ J &= C\langle\lambda\rangle\beta^\top - A \\ Q &= C\langle\lambda\rangle C^\top. \end{aligned} \tag{B9}$$

Similarly, the second-order moment closure (Eq. 13; Fig. 3D) in the basis projection is given by:

$$\begin{aligned}
\partial_t \mu_z &= C\tilde{\lambda} - A\mu_z \\
\tilde{\lambda} &= \bar{\lambda} \cdot (1 + \tfrac{1}{2}\beta^\top \Sigma_z \beta) \\
\partial_t \Sigma_z &= J\Sigma_z + \Sigma_z J^\top + Q \\
J &= C\bar{\lambda}\beta^\top - A \\
Q &= C\tilde{\lambda}C^\top.
\end{aligned} \tag{B10}$$

For comparison, a simpler alternative to moment-closure is the Linear Noise Approximation (LNA; Fig. 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:

$$\begin{aligned}
\partial_t \mu_z &= C\bar{\lambda} - A\mu_z \\
\partial_t \Sigma_z &= J\Sigma_z + \Sigma_z J^\top + Q \\
J &= C\bar{\lambda}\beta^\top - A \\
Q &= C\bar{\lambda}C^\top
\end{aligned} \tag{B11}$$

Appendix C: numerical implementation

The basis-projected moment equations (Eqs. B9–B11) can be integrated using forward Euler or other integration methods, and code for integrating the moment equations and generating the figures in this paper is published online (Rule, 2018). We can also convert them to a form resembling the discrete-time extended Kalman-Bucy filter (Kalman and 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 (Eq. B9):

$$\begin{aligned} \mu_{t+\Delta t} &= F\mu_t + C\,\langle\lambda\rangle\,\Delta t \\ \Sigma_{t+\Delta t} &= G\Sigma_t G^\top + Q\Delta t, \end{aligned} \tag{C1}$$

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 Eqs. B9–B11, and which 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 (Table C1). Ground truth estimates of these moments are sampled from the "true" PPGLM using Monte Carlo sampling (10K 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 e.g. Fernandes et al., 2013; Ramirez and Paninski, 2014); these are appropriate for the singly-stochastic interpretation of the PPGLMs, which treats the latent log-intensity as a deterministic, non-stationary 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, however, moment equations

	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 σ	0.41	0.90	0.86	0.53

Table C1: *Moment-closure methods capture process moments more accurately than mean-field with LNA (MF/LNA).* Accuracy of various methods compared to Monte-Carlo sampling (10K 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 pA^2 and time-constant of 100 ms.

(B9–B11) allow us to estimate a likelihood penalty related to slow dynamics (Table C2). We estimated likelihood penalties via Bayesian filtering on the moment equations as a state-space model, handling the non-conjugate log-Gaussian Poisson measurement update via moment matching. All three moment approximations broadly agreed 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.

	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

Table C2: *Moment-closure methods interpret the PPGLM as a doubly-stochastic process, introducing a penalty for incorrectly modeled dynamics.* 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; c.f. Ramirez and 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; c.f. 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.

References

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

Aghagolzadeh, M. and 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., and Stumpf, M. P. (2013). A general moment expansion method for stochastic kinetic models. *The 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., and Sahani, M. (2012). Spectral learning of linear dynamics from generalised-linear observations with application to neural population data. In *Advances in neural information processing systems*, pages 1682–1690.

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

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

Cseke, B., Zammit-Mangion, A., Heskes, T., and 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., and 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., and 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., and 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. and Löcherbach, E. (2015). Modeling networks of spiking neurons as interacting processes with memory of variable length. *arXiv preprint arXiv:1502.06446*.

Gao, Y., Archer, E. W., Paninski, L., and Cunningham, J. P. (2016). Linear dynamical neural population models through nonlinear embeddings. In *Advances in Neural Information Processing Systems*, pages 163–171.

Gerhard, F., Deger, M., and 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. *The Journal of Chemical Physics*, 113(1):297–306.

Hocker, D. and Park, I. M. (2017). Multistep inference for generalized linear spiking models curbs runaway excitation. In *Neural Engineering (NER), 2017 8th International IEEE/EMBS Conference on*, pages 613–616. 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., Lee, A. K., Anastassiou, C. A., Andrei, A., Aydın, Ç., et al. (2017). Fully integrated silicon probes for high-density recording of neural activity. *Nature*, 551(7679):232.

Kalman, R. E. and 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*. Wiley Online Library.

Lawhern, V., Wu, W., Hatsopoulos, N., and 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., Berdondini, L., and Sernagor, E. (2014). Following the ontogeny of retinal waves: pan-retinal recordings of population dynamics in the neonatal mouse. *The Journal of physiology*, 592(7):1545–1563.

Macke, J. H., Buesing, L., Cunningham, J. P., Yu, B. M., Shenoy, K. V., and Sahani, M. (2011). Empirical models of spiking in neural populations. In *Advances in neural information processing systems*, pages 1350–1358.

Macke, J. H., Buesing, L., and Sahani, M. (2015). Estimating state and parameters in state space models of spike trains. *Advanced State Space Methods for Neural and Clinical Data*, page 137.

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

Pachitariu, M., Petreska, B., and Sahani, M. (2013). Recurrent linear models of simultaneously-recorded neural populations. In *Advances in neural information processing systems*, pages 3138–3146.

Park, I. M. and Pillow, J. W. (2011). Bayesian spike-triggered covariance analysis. In *Advances in neural information processing systems*, pages 1692–1700.

Pfau, D., Pnevmatikakis, E. A., and Paninski, L. (2013). Robust learning of low-dimensional dynamics from large neural ensembles. In *Advances in neural information processing systems*, pages 2391–2399.

Ramirez, A. D. and 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., and 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*, available: 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., and 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., and Truccolo, W. (2017). Dissociation between sustained single-neuron spiking β-rhythmicity and transient β-lfp oscillations in primate motor cortex. *Journal of Neurophysiology*, pages jn–00651.

Sahani, M., Bohner, G., and Meyer, A. (2016). Score-matching estimators for continuous-time point-process regression models. In *Machine Learning for Signal Processing (MLSP), 2016 IEEE 26th International Workshop on*, pages 1–5. IEEE.

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

Schnoerr, D., Sanguinetti, G., and Grima, R. (2014). Validity conditions for moment closure approximations in stochastic chemical kinetics. *The Journal of chemical physics*, 141(8):08B616_1.

Schnoerr, D., Sanguinetti, G., and Grima, R. (2015). Comparison of different moment-closure approximations for stochastic chemical kinetics. *The Journal of Chemical Physics*, 143(18):11B610_1.

Schnoerr, D., Sanguinetti, G., and 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. and Brown, E. N. (2003). Estimating a state-space model from point process observations. *Neural Computation*, 15(5):965–991.

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

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

Toyoizumi, T., Rad, K. R., and 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 *Analysis of parallel spike trains*, pages 321–341. Springer.

Truccolo, W. (2017). From point process observations to collective neural dynamics: Nonlinear hawkes process glms, low-dimensional dynamics and coarse graining. *Journal of Physiology-Paris*.

Truccolo, W., Eden, U. T., Fellows, M. R., Donoghue, J. P., and 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., and 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. and 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., and 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., and Pillow, J. W. (2017). Gaussian process based nonlinear latent structure discovery in multivariate spike train data. In *Advances in Neural Information Processing Systems*, pages 3498–3507.

Yu, B. M., Cunningham, J. P., Santhanam, G., Ryu, S. I., Shenoy, K. V., and Sahani, M. (2009). Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. In *Advances in neural information processing systems*, pages 1881–1888.

Zhao, Y. and Park, I. M. (2016). Interpretable nonlinear dynamic modeling of neural trajectories. In *Advances in Neural Information Processing Systems*, pages 3333–3341.

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