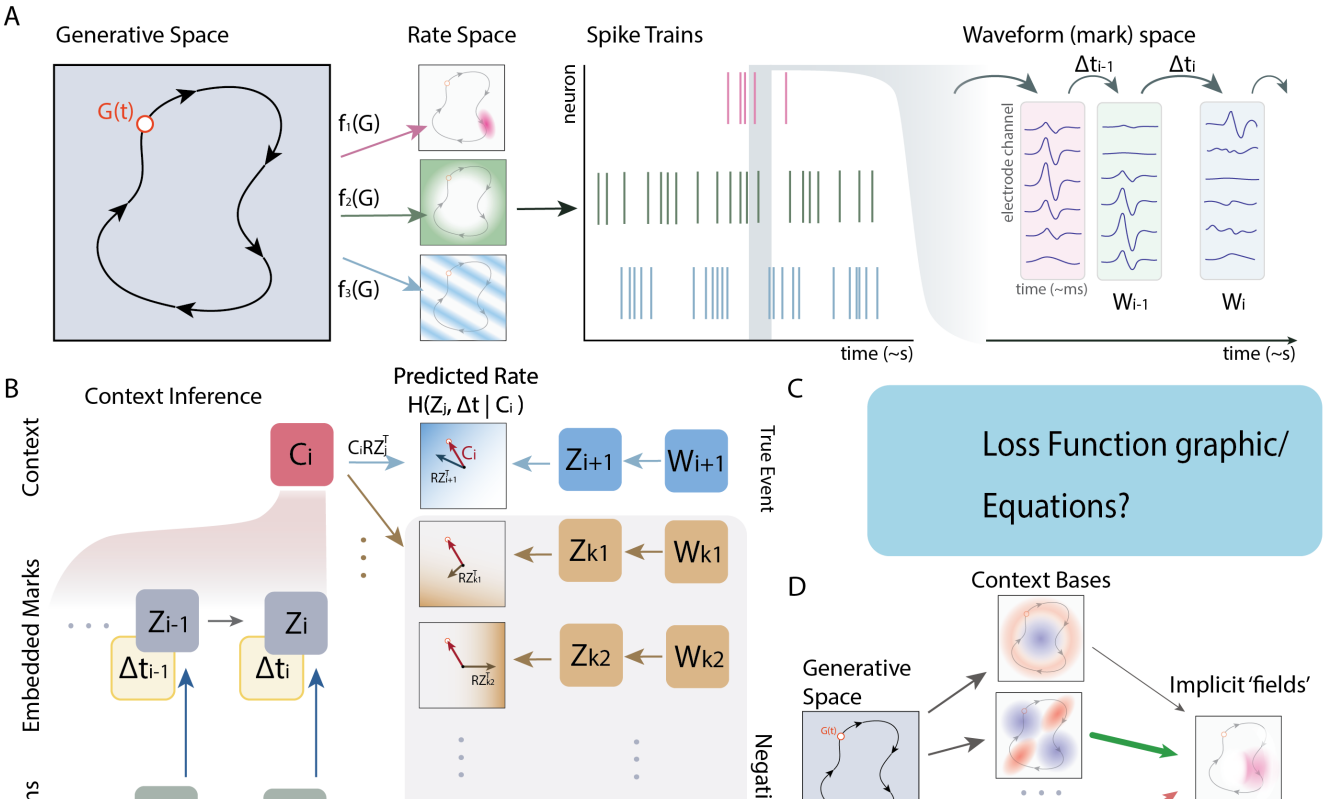
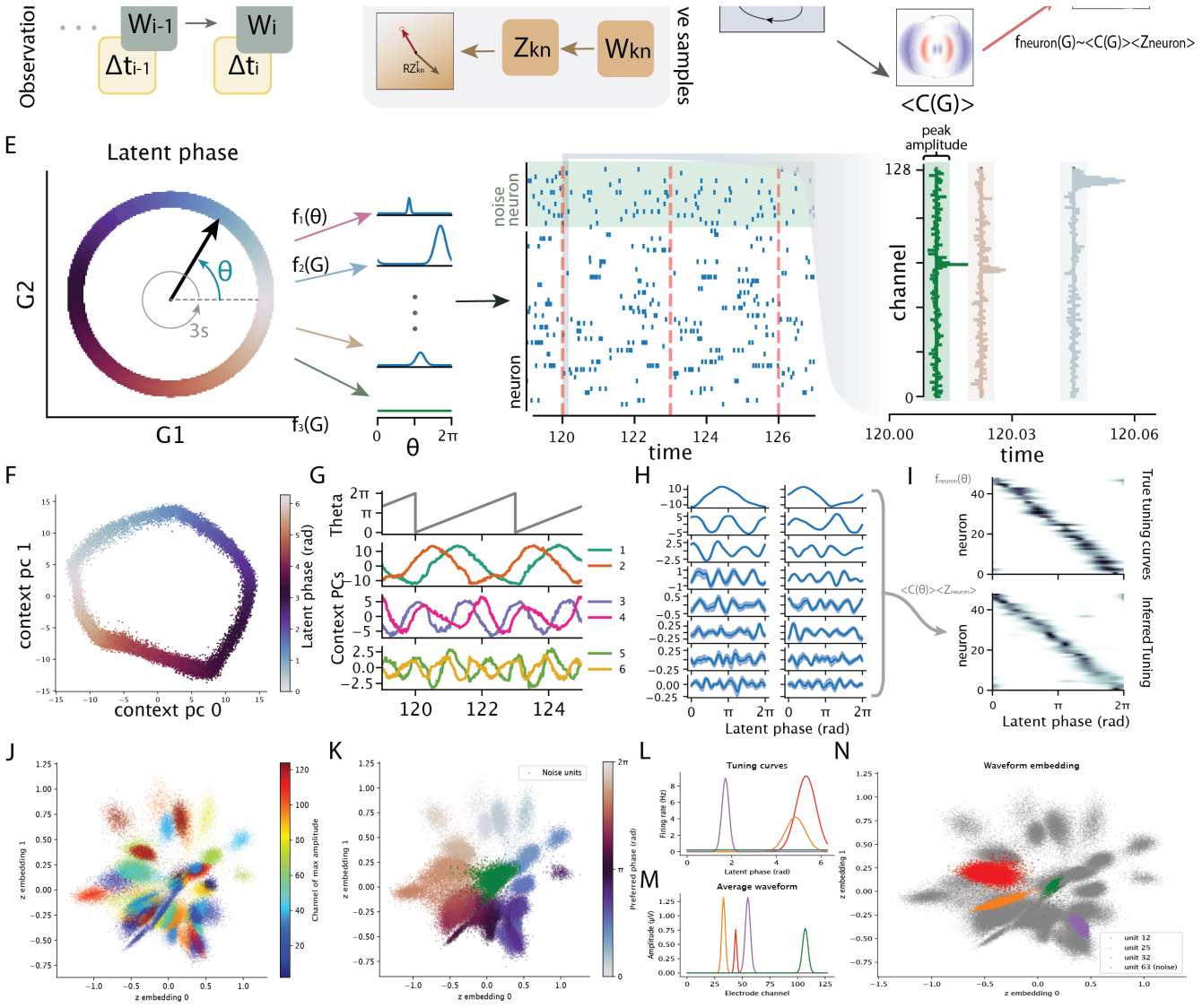
Learning predictive latent structure from unclustered neural spike trains

Abstract

Understanding the low-dimensional generative processes that shape neural population activity is a central goal of systems neuroscience. However, there remains a need for data processing and dimensionality reduction methods that are compatible with the exploding scale of data acquisition. Existing approaches rely on spike sorting (a time intensive process that discards many low-confidence events) and binned firing rates (which imposes temporal smoothing that obscures fast dynamics). Moreover, reconstructive objectives tend to emphasize dominant activity patterns, making it difficult to capture rare but behaviorally informative neural events. Here we introduce Contrastive Prediction of Point-Process Observations (C3PO), a architecture for inferring latent neural generative variables directly from unclustered spike trains. C3PO models neural activity as a marked point process, operating on continuous spike times and waveform features without spike sorting. Using an objective derived from noise-contrastive estimation, the model learns a low-dimensional context representation which captures the predictive information in the waveform data. We demonstrate this method robustly captures known neural-coding dynamics across varied species and brain regions. Furthermore, we leverage the point-process nature of the model to identify higher-frequency fluctuations in the population dynamics and relate these to behavioral variables. Together, these results establish C3PO as a scalable, unsupervised approach for uncovering generative structure in high-dimensional neural data.

Results





A Generative Model of Neural Spike Train Observations (Fig 1A)

We begin by outlining an intuitive generative model for how neural population activity gives rise to the spike train observations analyzed in this work. At any moment in time, the activity of a neural population reflects the accumulated history of internal dynamics, external inputs, and ongoing computation. Rather than treating this history as an unstructured sequence of past events, we assume that it can be summarized by a latent variable G_t, which acts as a sufficient statistic of the system history for predicting future activity. Formally, G_t captures all information from the past that is relevant for determining the distribution of future spiking events. This concept is closely related to the notion of *causal states* in computational mechanics, where histories are grouped according to their predictive equivalence, and to broader results showing that minimal sufficient statistics preserve predictive information between past and future.

Conditional Intensity Function

Within this framework, the probability that a given neuron emits a spike at time t is governed by a conditional intensity (hazard) function:

$$\lambda_n(t) = \lambda_n(G_t),$$

where n indexes neurons. This mapping can be interpreted as a firing rate field defined over the latent state G. In analogy to classical tuning curves, which describe firing rates as functions of experimentally controlled variables, here each neuron exhibits a (generally nonlinear) response function over the latent generative state. These functions may be highly nonlinear and reflect mixed selectivity to multiple underlying variables.

Latent Dynamical Systems Perspective

To build intuition for the role of G, it is helpful to view neural population activity through the lens of low-dimensional dynamical systems. A growing body of work has shown that high-dimensional neural activity often evolves on low-dimensional manifolds governed by latent dynamics. In this perspective, G_t corresponds to a point in this latent state space, whose evolution captures the computation being performed by the circuit.

For example, in hippocampus, the activity of place cells is driven by a combination of sensory inputs, memory, and movement history, yet can be well described as a function of a low-dimensional latent variable such as spatial position. In this case, G_t can be interpreted as the animal's position (or a related internal estimate), and each neuron's firing rate field $\lambda_n(G)$ corresponds to its place field. More generally, in different brain regions, latent variables may encode mixtures of task variables, internal states, or dynamical trajectories, giving rise to distributed and context-dependent firing patterns.

Spike Generation as a Point Process

Given these hazard functions, spikes are generated as stochastic events drawn from a point process. That is, conditioned on the latent state G_t, each neuron emits spikes probabilistically according to its instantaneous firing rate $\lambda_n(G_t)$, consistent with standard point-process formulations of neural activity. This stochastic sampling reflects both intrinsic neural variability and unobserved influences on the system.

Observations as a Marked Point Process

Finally, neural recordings do not directly observe neuron identities or spike events in isolation. Instead, extracellular electrodes measure voltage deflections produced by nearby neurons, resulting in observed waveform snippets whose shapes depend on the spatial and biophysical properties of the underlying sources. Each detected event therefore consists of a time of occurrence and an associated waveform.

We represent the resulting data as a sequence of marked point-process observations:

$$X = (\Delta t_i, W_i),$$

where Δt_i denotes the inter-event interval and W_i denotes the recorded waveform features.

These observations arise from a cascade of transformations: Latent neural dynamics G_t determine firing probabilities, spikes are stochastically generated from these probabilities, and electrode measurements produce high-dimensional, mixed observations of these events

Summary and Challenge

This generative view highlights the central challenge: the latent state G_t that governs neural activity is not directly observed, and must be inferred from indirect, noisy, and high-dimensional measurements. The

following sections develop a framework for learning representations that recover this latent predictive structure directly from the observed spike train.

Model Architecture

Recovering latent structure from marked point-process observations (Fig. 1A)

We now construct a model that inverts the generative process described above. In the generative view, latent neural dynamics G_t give rise to firing rates $\lambda(G_t)$, from which spikes are stochastically generated and subsequently observed as waveform-marked events. Our goal is to learn a representation that recovers the predictive structure of G_t directly from these observations.

Concretely, given a sequence of marked point-process observations

$$X = (\Delta t_i, W_i),$$

we seek to learn a transformation of the form

$$(\Delta t_i, W_i); \rightarrow; Z_i; \rightarrow; C_i,$$

where Z_i represents a low-dimensional embedding of individual events, and C_i represents a latent context variable that approximates the predictive state of the system. In this framework, C_i plays the role of an inferred sufficient statistic of the history, analogous to the latent variable G_t in the generative model.

Mark embedding

The first stage of the model constructs a low-dimensional embedding of the observed waveform features. Each waveform $W_i \in \mathbb{R}^{D_{\text{obs}}}$ is mapped to a latent representation:

$$Z_i = \epsilon(W_i), \quad Z_i \in \mathbb{R}^{D_{\text{embed}}}, \quad D_{\text{embed}} < D_{\text{obs}}.$$

This embedding serves as a representation of the underlying event identity and its functional role within the neural population. Importantly, unlike traditional spike sorting approaches, this representation is not constrained to identify discrete neuron identities nor encode waveform similarity; rather, it is learned to support prediction of future events, and thus organizes observations according to their firing timing

The transformation $\epsilon(\cdot)$ may be implemented using a flexible neural network architecture (e.g., multilayer perceptron or convolutional encoder), provided that gradients can be propagated through the mapping. This flexibility allows the model to accommodate diverse electrode geometries and recording configurations.

From the perspective of the generative model, this stage corresponds to inferring a latent representation of the emission process underlying each observed waveform.

Context Representation of Latent Dynamics

To capture the temporal structure of neural activity, we construct a context variable C_i that integrates information across past events:

$$C_i = \rho(\Delta t_j, Z_{j\,j\leq i}),$$

where ρ is a sequence model that maps the history of observed events to a fixed-dimensional representation.

This formulation generalizes the recursive update

$$C_i = \rho(\Delta t_i, Z_i, C_{i-1}),$$

and allows the context to be implemented using a wide class of architectures. In general, ρ may be any function that takes a sequence of spike events as input and produces a vector-valued representation, including recurrent neural networks, temporal convolutional networks, or transformer-based models.

In this work, we use a WaveNet-style architecture, consisting of stacks of dilated causal convolutions. This architecture is particularly well-suited for neural spiking data, as it efficiently captures long-range temporal dependencies while maintaining causal structure. By exponentially expanding the receptive field with depth, dilated convolutions enable the model to integrate information over extended time horizons without the computational cost associated with recurrent or attention-based models.

The resulting context variable C_i represents a learned summary of population activity that evolves over time. By construction, it is optimized to retain information that is predictive of future events, and can therefore be interpreted as an estimate of the latent state G_t, up to transformations that preserve predictive equivalence.

Conditional hazard and rate modeling

In the generative process, the latent state G_t determines a set of continuously varying firing rates $\lambda(G_t)$, which define the stochastic generation of spike events. In contrast, our observations are discrete events drawn from a point process. Modeling their likelihood therefore requires specifying a conditional hazard function:

$$H(X_i \mid C_{i-1}),$$

which specifies the probability of observing an event X_i given the preceding context.

We parameterize this hazard function through a scoring function $\theta(Z_i, C_{i-1})$, computed using a bilinear interaction between the mark embedding and the context:

$$\theta(Z_i, C_{i-1}) = Z_i^\top R C_{i-1}.$$

This score is then mapped to a valid hazard function according to a chosen point-process model. In general, any parametric form with a closed-form hazard function can be used within this framework. For examples presented in this work, we use a Poisson process model, where the hazard is equal to the conditional firing rate:

$$H(X_i \mid C_{i-1}) = \hat{\lambda}(X_i \mid C_{i-1}) = Z_i^\top R C_{i-1}.$$

For simplicity and interpretability, we further constrain the model such that the embedding and context spaces have equal dimensionality, and set $R = I$, yielding:

$$\hat{\lambda}(X_i \mid C_{i-1}) = Z_i^\top C_{i-1}.$$

This shared-space formulation places both marks and contexts in a common latent space, where their inner product directly determines event likelihood. This structure will later enable a geometric interpretation of the learned representations.

Relationship to the generative model

The model architecture provides a structured approximation to the generative process described in Fig. 1A. The latent state G_t is replaced by the learned context representation C_t, the emission process is captured by the mark embedding Z, and the firing rate field $\lambda(G)$ is replaced by the parameterized hazard function $H(X \mid C)$.

$$G_t ; \leftrightarrow ; C_t, \quad \lambda(Z \mid G) ; \leftrightarrow ; H(X \mid C), \quad \text{emission identity} ; \leftrightarrow ; Z.$$

Rather than explicitly recovering G_t the model learns representations C_t and Z_i that are sufficient for predicting future observations, thereby capturing the predictive structure of the underlying neural dynamics.

Contrastive Learning of Marked Point-Process Likelihoods

Likelihood formulation for marked point-process observations

We begin by deriving the likelihood of a single observed event at time t_i, with observation sequence $X = (\Delta t_i, W_i)$ and context representation C_{i-1} summarizing the event history. The observed waveform is first embedded as $Z_i = \epsilon(W_i)$ as described above

The likelihood of observing an event with embedded mark Z_i at delay Δt_i, with no other events occurring in that interval, is given by the standard marked point-process formulation:

$$p(Z_i, \Delta t_i \mid C_{i-1}) = H_i, \bar{S},$$

where

$$H_i \equiv H(Z_i, \Delta t_i \mid C_{i-1})$$

is the conditional hazard function, and $\bar{S}$ is the cumulative survival probability over all possible marks. Intuitively, this likelihood reflects the probability that an event occurs at Z_i at time ΔT_i, and that no other competing events occur beforehand.

The survival term can be expressed as an integral over the mark space:

$$\bar{S} = \exp\left(\int_Z \log S(Z, \Delta t_i \mid C_{i-1}), dZ \right).$$

and plays a role analogous to a partition function, normalizing over all possible events. Approximating this quantity via sampling yields the following negative log-likelihood:

$$ \mathcal{L}_i = -\log H_i - \sum_{j \in \{i, \text{negatives}\}} \log S_j. $$

Under a Poisson process model, where the hazard is equal to the conditional intensity $H_j = \lambda_j$, this reduces to:

$$ \mathcal{L}*i = -*log *lambda_i + *Delta t_i *sum{j \in {i, \text{negatives}}} \lambda_j. $$

This formulation directly parallels classical maximum likelihood estimation for marked point processes. However, in the present setting, the mark space Z is continuous and high-dimensional, making accurate estimation of the survival term $\bar{S}$ intractable. In particular, sampling-based approximations are both computationally expensive and biased toward regions of high firing rate.

Contrastive reformulation via noise-contrastive estimation

To address this challenge, we reformulate likelihood estimation as a contrastive prediction problem using noise-contrastive estimation (NCE). Instead of directly computing the likelihood, we consider a set of candidate marks:

$$\mathcal{M} = Z_i \cup Z_j{}_{j=1}^{m},$$

consisting of the true next event and a set of negative samples drawn from elsewhere in the data.

We define the probability of each candidate under two models:

- the true model: $p_j = H_j \bar{S}$
- the alternative (noise) model: $q_j = H'_j \bar{S}'$

where H' and $\bar{S}'$ denote the hazard and survival functions of the alternative process.

We then ask: given that exactly one element of $\mathcal{M}$ was generated by the true process and the rest by the noise process, what is the probability that Z_i is the true event?

This yields:

$$p(Z_i = \text{true}) = \frac{p_i \prod_{k \neq i} q_k}{\sum_j p_j \prod_{k \neq j} q_k} = \frac{p_i / q_i}{\sum_j p_j / q_j}.$$

Substituting the point-process likelihood expressions gives:

$$p(Z_i = \text{true}) = \frac{H_i \bar{S} / (H'_i \bar{S}')}{\sum_j H_j \bar{S} / (H'_j \bar{S}')} = \frac{H_i / H'_i}{\sum_j H_j / H'_j}.$$

Critically, the survival terms $\bar{S}$ and $\bar{S}'$ cancel in this ratio. This eliminates the need to approximate the intractable normalization over the continuous mark space, reducing the problem to evaluating relative hazard values.

The resulting loss function is:

$$ \mathcal{L} = -\log p(Z_i = \text{true}) = -\log \frac{H_i}{H'_i}

- \log \sum_j \frac{H_j}{H'_j}. $$

Connection to contrastive predictive coding and mutual information

The ratio $\frac{H(Z|C)}{H'(Z)}$ plays a role directly analogous to the density ratio $\frac{p(x|c)}{p(x)}$ in contrastive predictive coding (CPC). In CPC, the optimal scoring function learned under the InfoNCE objective is proportional to this density ratio:

$$f(x, c) \propto \frac{p(x \mid c)}{p(x)}.$$

Analagously, our formulation learns a function proportional to:

$$\frac{H(Z \mid C)}{H'(Z)},$$

which serves as a density ratio over event likelihoods in mark space.

Importantly, CPC shows that optimizing the InfoNCE objective maximizes a lower bound on the mutual information between context c and future observations x. In our setting, this implies that minimizing the contrastive loss encourages the context representation C to maximize a lower bound on the mutual information between the spike history and future events:

$$I(C_{i-1}; Z_i).$$

This provides a direct connection to the generative framework introduced in Fig. 1A. The latent state G_t was defined as a sufficient statistic of the past for predicting future observations. By maximizing mutual information between C and future events, the model learns a representation C that captures all predictive information in the history, and therefore approximates the latent state G_t up to transformations that preserve predictive equivalence.

Thus, the contrastive objective does not merely provide a tractable approximation to likelihood—it enforces a principled information-theoretic criterion for recovering latent generative structure.

Choice of alternative model

The alternative model H' defines a reference distribution over mark space and influences the geometry of the learned embedding.

Uniform hazard

In the simplest case, the hazard is independent of the mark:

$$H'(Z_j, \Delta t_i \mid C_{i-1}) = H'(\Delta t_i \mid C_{i-1}),$$

which can be set to unity without loss of generality. This yields:

$$\mathcal{L} = -\log H_i + \log \sum_j H_j.$$

Under the additional assumption that the hazard is independent of time, this formulation reduces exactly to the InfoNCE objective used in CPC, establishing a direct correspondence between regularly sampled sequence models and the present continuous-time point-process setting.

Regularization of the mark space

More generally, the alternative model can be used to impose structure on the embedding. We consider a Poisson process independent of context, with intensity decaying with the norm of the embedding:

$$H'_j = \exp\left(-|Z_j|_k^k\right).$$

This choice regularizes the mark space by encouraging embeddings that do not contribute to prediction to collapse toward the origin. Empirically, using the Euclidean norm stabilizes training and improves convergence. In this setting, events that are uninformative for predicting future activity—such as spikes generated independently of network state—are mapped to regions of low norm, reflecting their reduced contribution to the predictive representation.

Inferred context describes an (incomplete) bases over G

As shown above, the training procedure optimizes for C that approximates the predictive information of G and with a linear projection to firing rates. Given that the rates of each neuron are in general an nonlinear function over G ($\lambda_n(|\ G) = f_n(G)$, then the components of C must comprise a set of non-linear basis functions over G. The alignment of a neuron's embedded waveform (Z) with the different C components therefore describe a weighted sum of these bases functions which strives to apprximated the neuron's firing field (Fig. 1D).

Toy model of neural population activity driven by a periodic latent variable (Fig. 1E)

To validate the proposed framework in a setting with known ground truth, we constructed a synthetic dataset in which neural activity is generated by a low-dimensional latent dynamical variable. Specifically, we consider a single periodic latent variable $\theta(t)$ evolving on a circular manifold, which defines the state of the system at each timepoint.

Neural population activity is generated by defining a set of neurons with nonlinear tuning curves over this latent phase. Each neuron n is assigned a preferred phase and emits spikes according to a conditional intensity function $\lambda_n(\theta)$, producing a population of phase-tuned units analogous to place or head-direction cells. In addition, 25% of neurons are designated as non-informative units, with flat tuning curves independent of θ, resulting in spike trains that are uncorrelated with the latent state.

Spiking events are sampled from an inhomogeneous Poisson process with rates determined by these tuning functions. To mimic extracellular recordings, each spike is associated with a high-dimensional waveform feature: template waveforms are generated for each neuron across recording channels, and individual events are constructed by adding noise to these templates. This produces a sequence of marked point-process observations:

$$X = (\Delta ti, Wi)$$

The resulting dataset captures key properties of real neural recordings: (i) low-dimensional latent structure underlying population activity, (ii) nonlinear mappings between latent state and firing rates, (iii) indirect, high-dimensional observations without neuron identity labels, and (iv) the presence of non-informative units that do not encode the latent variable. This controlled setting enables direct evaluation of whether

the model can recover latent structure and distinguish predictive from non-predictive activity in clusterless spike data.