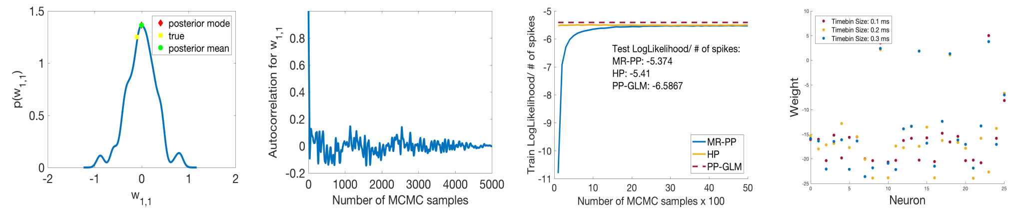

(a) Revised Figure 1.b (b) Revised Figure S1.b (c) Revised Figure 6.b (d) MAUP for Sec 4.3

Figure 1: *Revised experiment results for Mutually Regressive Point Processes.*

We would like to thank the reviewers for their detailed and constructive reviews of our manuscript. Their comments definitely help us clarify many points in the paper. We did our best to address the main areas of concern: i) convergence of the learning algorithm. We ran the learning algorithm for more MCMC iterations, and we revised the relevant Figures 1a, 1b); and ii) limited experiments. We compare against PP-GLMS (discrete-time model capable of capturing general temporal interactions) and Hawkes processes (continuous time, capable of capturing only excitatory interactions) (Figure 1c). We also demonstrate experimentally the sensitivity of the PP-GLM parameters to the boundaries used for the temporal aggregation resulting in a different sign of effect (from excitatory to inhibitory) in Figure 1d.

Convergence of the learning algorithm (Reviewer 2 - comment 4). We reran the learning algorithm for 5000, instead of 1000 MCMC iterations for the synthetic experiment. In Figure 1a, we plot the new posterior obtained. In Figure 1b, we plot the autocorrelation (initially provided in the supplementary material) for that parameter. As was correctly pointed out, the two modes were merged into one at 0 after running the MCMC for more iterations. The oscillations of the autocorrelation around 0 were also reduced after a larger number of samples. The modes in the initial manuscript did not raise a warning flag, since they were insignificant: both of them were close to zero (compared to the very large negative weights for the rest of interactions), indicating no considerable effect from type I coming from the non-linear part of the intensity, as dictated by the prior and due to the self-excitation. The small bumps disappeared for the rest of the parameters. (We will update the camera-ready accordingly in case of acceptance). **Additional experimental results (Reviewer 2 - comment 3, Reviewer 3).** In Figure 1c, we provide the test-logl not only of the PP-GLMs (typically used for spiking data), but also that of a Hawkes process (HP), indicating that loss in the generalization capability of the model stems mostly from the time discretization, although capturing inhibitory effects could further improve its performance. We illustrate one weakness of the discrete-time PP-GLM models briefly discussed in the introduction ("However, the estimated regression coefficients may vary widely depending on the boundaries chosen for aggregation"), known as the Modifiable Areal Unit Problem (MAUP) [1] that is attributed to the temporal aggregation of the spikes. In Figure 1d, we plot the coefficients of the PP-GLM learned for variable size of the time-bin (assuming unit degree of regression). For the effect from neuron 23 on neuron 1, for example, the sign of the interaction changes from positive (for 0.1 msec and 0.3 msec) to negative (for 0.2 msec), indicating a potentially time-varying (in terms of its sign), relationship. Although, this problem could potentially be solved by considering a degree of regression larger than one and finer time-bins, these parameters have to be predetermined (potentially for each neuron/ type separately). Moreover, the change-points may depend dynamically on the temporal history. For that particular experiment, different configuration of time-bin size and degree of regression did not yield improvement in terms of the predictive likelihood. The proposed continuous-time model inherently circumvents these limitations by allowing i) two channels of time-varying interaction one coming from the mutually exciting intensity function, the other from the sigmoidal part which may or may not be mutually exclusive by adjusting the parameters of the prior accordingly ii) the parameters that regulate the excitatory and inhibitory effects to be learned dynamically from the data. We will include a network illustrating the temporal interactions identified by the proposed model and the MAUP for the rest of the neurons in the camera-ready in case of acceptance. **Importance of stable dynamics (Reviewer 2 - comment 5).** One immediate advantage of a model with stable dynamics, is that it allows long-run time predictions. We refer to the paper [2] (Section "Importance of stable point process models for applications"), for a detailed explanation on the importance of obtaining physiological temporal patterns. **Typo L142 - (Reviewer 2).** Thank you for pointing out the typo in L142: the LHS is identical to the product term in the RHS of Equation (14), L138. **Prior Choice - (Reviewer 1).** We thank Reviewer 1 for the positive comments.We will incorporate your feedback (along with suggestions for the prior choice, e.g empirical Bayes) in the discussion section of the camera-ready version in case of acceptance. Based on our current experiments, especially the parameters in the activation functions are critical and hence worth being finely tuned.

[1] Fotheringham, A. S., Wong, D. W. (1991). The modifiable areal unit problem in multivariate statistical analysis. Environment and planning A, 23(7), 1025-1044. **[2]** 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.