Stimulus presentation can enhance spiking irregularity across subcortical and cortical regions

Salah Fayaz^{1*}, Mohammad Amin Fakharian^{1,2*}, Ali Ghazizadeh^{1,2#}

¹Electrical Engineering Department, Sharif University of Technology, Tehran, Iran ²School of Cognitive Sciences, Institute for Research in Fundamental Sciences, Tehran, Iran

* equal contribution

corresponding: alieghazizadeh@gmail.com

Abstract

Stimulus presentation is believed to quench neural response variability as measured by fano-factor (FF). However, the relative contribution of within trial spike irregularity ($n\Psi$) and trial to trial rate variability (nRV) to FF reduction has remained elusive. Here, we introduce a principled approach for accurate estimation of variability components for a doubly stochastic point process which unlike previous methods allows for a time varying $n\Psi$ (aka ϕ). Notably, analysis across multiple subcortical and cortical areas showed across the board reduction in rate variability. However, unlike what was previously thought, spiking irregularity was not constant in time and was even enhanced in some regions abating the quench in the post-stimulus FF. Simulations confirmed plausibility of a time varying $n\Psi$ arising from within and between pool correlations of excitatory and inhibitory neural inputs. By accurate parsing of neural variability, our approach constrains candidate mechanisms that give rise to observed rate variability and spiking irregularity within brain regions.

Keywords: Neural variability, fano-factor, spiking irregularity, rate variability, network models, electrophysiology.

Introduction

While mean firing rate is widely used as a proxy of neural communication code, firing rate variance is also shown to play a significant role in neural coding and to serve as a diagnostic tool for distinguishing underlying neural mechanisms (A. K. Churchland et al., 2011; M. M. Churchland et al., 2006, 2010; M. M. Churchland & Abbott, 2012; Deco & Hugues, 2012; Goris et al., 2014; Shadlen & Newsome, 1998). In particular, neural response variability is believed to change during development and to be correlated with behavioral performance, stimuli conditions and to vary across different neural states (Lombardo et al., 2018; McIntosh et al., 2008; Mišić et al., 2010; Voytek et al., 2015; Waschke et al., 2021). Several studies have shown that spike count variability changes across different conditions in several cortical, subcortical, and cerebellar regions (A. K. Churchland et al., 2011; M. M. Churchland et al., 2010; Goris et al., 2014; Lombardo et al., 2018; Mitchell et al., 2007, 2009; Payne et al., 2019; Poland et al., 2019; Schölvink et al., 2015). Importantly, numerous observations support a drop in neural response variability as measured by fano-factor (FF) in response to sensory or motor stimuli across cortical regions (M. M. Churchland et al., 2006, 2010; Rickert et al., 2009). The variation in neural spike count when viewed as a doubly stochastic process may be due to (1) the variation in generation of the spikes within the trials, (2) the variation between the trials of the same task (law of total variance). The first component is driven by within trial spiking irregularity (SI or Ψ) and gives rise to the expected variance of the count (EVC) while the second component is driven by between trial rate variability (RV) and forms the variance of the expected count (VEC) (A. K. Churchland et al., 2011; Goris et al., 2014; Vinci et al., 2016). Despite this general understanding, a principled approach to estimate these two components of neural variability and to track their possibly concurrent

changes in time is not developed yet. Consequently, it is not known how the widely reported post-stimulus reduction in FF in the brain arise from corresponding changes in RV and Ψ .

A previous attempt to estimate the RV component (A. K. Churchland et al., 2011) (VarCE method) showed that FF reduction is paralleled by the reduction in VEC (VarCE is an estimate of VEC). However, this approach assumed Ψ to be related to firing rate by a coefficient ϕ which was taken to be constant in time (see methods). Other more recent methods also followed the same assumption for spike count variability decomposition (Goris et al., 2014; Vinci et al., 2016). Clearly, such an assumption can result in gross misestimations of Ψ and subsequently VEC if there are temporal changes in ϕ . Even in the absence of such changes, VarCE estimates only the relative VEC up to arbitrary additive shifts. Moreover, the true nature and accurate estimation of ϕ also remains unaddressed.

Here, by invoking minimal assumptions from doubly stochastic renewal theory, we introduce an approach that explicitly relates FF to normalized RV and Ψ (nRV and n Ψ) and reveals the nature of both components in relation to the first and second moments of inter-spike-interval distribution (ISI) (FF asymptote or FFA method). In addition, a second approach, adapted and repurposed from a previously introduced method for estimating components of cross neural spike count correlation (Vinci et al., 2016) is also used. Results from both methods show that unlike what was previously assumed, n Ψ robustly changed along with nRV during the stimulus presentation across several subcortical and cortical regions. While the post-stimulus nRV showed a ubiquitous change across all areas examined, post-stimulus n Ψ change was more

heterogenous: decreasing in some (mostly cortical) areas and increasing in others (mostly subcortical). Moreover, while we found the post-stimulus reduction in FF to be mostly due to nRV decrease, the concurrent change in $n\Psi$ was shown to convey critical information regarding the network state. Our results warn against the constant $n\Psi(\phi)$ assumption when studying post-stimulus response variability and cautions against strong claims to ubiquitous response variability quench across the brain. The heterogeneity in patterns of nRV and $n\Psi$ during an event constrains underlying neural mechanisms differently for different regions.

Results

Parsing neural response variability components: the method

A stochastic process with deterministic parameters would induce variability in the spike count, i.e. point-process variability. Moreover, stochastic selection of such parameters between different trials would impose variability across trials (doubly stochastic process). We assume a doubly stochastic renewal process in which inter-spike interval (ISI) parameters change probabilistically across trials. In this case, sources of spike count variability can be decomposed into at least two components using the law of total variance in the following form:

$$\text{Var}(N_T) = \underbrace{\text{E}[\text{Var}(N_{T_i}|\theta_i)]}_{\text{EVC}} + \underbrace{\text{Var}(\text{E}[N_{T_i}|\theta_i])}_{\text{VEC}} \quad (1)$$

Where N_T denotes the spike count in the time-bin (T) and θ_i represent ISI parameters in trial i. Here the first term denotes the expected variance of spike count (EVC) due to spiking irregularity (Ψ) within a trial given constant parameters, and the second term denotes variance of expected count (VEC) due to rate variability (RV). In the most general sense accurate estimation of these

two components is not feasible as we only can measure one sample of N_{T_i} ; hence, no estimates of the latent variable θ_i is explicitly known. However, progress can be made for renewal processes which has been widely used as models of spiking neurons (Fisch et al., 2012; Koyama & Kostal, 2014; Lansky et al., 2016; Nawrot et al., 2008; Omi & Shinomoto, 2011; Ostojic, 2011; Pipa et al., 2013; Rajdl et al., 2020; Shimokawa et al., 2010; Shinomoto et al., 2005). Fano-Factor estimation shows a functionality of time-bin (T); In particular it can be shown that the fano-factor (FF) of a renewal doubly stochastic process in sufficiently large time-bin follows the formula below (see methods for details):

$$\lim_{T \rightarrow \infty} FF(N_T) = \lim_{T \rightarrow \infty} \frac{\text{Var}(N_T)}{E[N_T]} = n\Psi + T \times nRV \quad (2)$$

$$n\Psi = \lim_{T \rightarrow \infty} \frac{E[\text{Var}(N_{T_i}|\theta_i)]}{E[N_T]} = \frac{E[\lambda_i^3 \sigma_i^2]}{E[\lambda_i]}, \quad nRV = \lim_{T \rightarrow \infty} \frac{\text{Var}(E[N_{T_i}|\theta_i])}{T \times E[N_T]} = \frac{\text{Var}(\lambda_i)}{E[\lambda_i]}$$

Where $\lambda_i = \frac{1}{E[ISI_i]}$ and $\sigma_i^2 = \text{Var}(ISI_i)$ and ISI_i is the random variable representing the inter-spike intervals drawn from a given distribution whose parameters can vary across trials. $n\Psi$ and nRV represent normalized within trial spiking irregularity (Ψ) and between trials rate variability (RV) respectively (see methods). As can be seen in equation 2, $n\Psi$ and nRV depend solely on the first and second moments of ISI distribution for renewal processes. Note that since FF in equation 2 is a linear sum of these two components the changes in FF may reflect change in either or both components.

According to equation 2, FF is asymptotically a linear function of time-bin. It includes an intercept which represents normalized within trial spiking irregularity ($n\Psi$ aka ϕ) and a slope which represents normalized between trials rate variability (nRV). Therefore, a linear regression of FF

against time in the asymptotic part would provide both components of FF. We call this method, the FFA method. For a simple stochastic point process with no trial-to-trial variation in ISI parameters, FF should be constant for sufficiently large time-bins. Figure 1a shows that indeed after a short transient, FF converges to the theoretical prediction based on the parameters of the ISI distribution (i.e. it is 1 for Poisson point process and is $\frac{1}{\kappa}$ for $\text{Gamma}(\kappa, \theta)$, see methods). Consistent with the well-known equality of FF and CV^2 in stationary renewal process (Buckland & Cox, 1964), it can be shown that - even for non-stationary processes - $n\Psi$ can be approximated by $E[\text{CV}_i^2]$ where CV_i is the coefficient of variation in trial i (see methods). This is intuitively encouraging since local CV measures have been widely used as a measure of spiking irregularity (Miura et al., 2007; Nawrot et al., 2008; Payne et al., 2019; Ponce-Alvarez et al., 2010). On the other hand, for a doubly stochastic process with trial-to-trial variation in rate, FF increases in an asymptotically linear fashion as a function of time-bin. Note that in this case the intercept of the line which represents $n\Psi$ remains unchanged as long as the average spiking irregularity (e.g. κ in $\text{Gamma}(\kappa, \theta)$) remains unchanged across trials (Figure 1a). Notably, this result holds for any renewal process regardless of the shape of ISI distribution as illustrated for exponential, gamma and inverse gaussian probability density functions (PDFs). Also note that while in theory FF asymptotic behavior is at infinity limits of time-bin, in practice the linear asymptotic behavior may be observable with much smaller time-bins depending on the firing rate of the neuron (Figure 1a, S1).

Using the FFA method, one would be able to track the temporal fluctuations in $n\Psi$ and $n\text{RV}$ during a trial using a sliding time window. Figure 1b shows a simulated doubly stochastic spike

process with concurrent variations in Ψ and RV across time. Each trial was generated from a renewal point process with ISIs following a two-parameter Gamma(κ, θ) distribution. The process had two sources of variability: Firstly, the rate in each trial was $50 + n_i \times (1 - 5t_+e^{-2t_+})$ where $n_i \sim N(\mu = 0, \sigma = 30)$ and t_+ is time for positive values and zero otherwise. This led to a constant mean firing of 50Hz across trials with a variance that changed during the trial (Fig. 1b top). This variation of rate around the 50Hz mean is responsible for RV. Secondly, in each trial, spikes were generated with the variability inherent to a gamma point process. In this case the parameter κ is inversely proportional to spiking irregularity and was set to change in time (e.g. changes in $\frac{1}{\kappa}$ from 2 to 0.25 to 2 in time). Figure 1c shows spiking realization of this doubly stochastic gamma process across four trials each corresponding to a given rate dynamic. Interestingly, FF intercept and slope in four selected 300ms epochs closely paralleled the expected changes in $n\Psi$ and nRV during the trial. Indeed, the slope showed a reduction concurrent with the reduction of rate variability after time zero while the intercept independently followed changes in $\frac{1}{\kappa}$ that was delayed (onset asynchrony) compared to RV (going from 2 to 0.25 and back to 2).

The FFA method can be used to track the temporal dynamics of $n\Psi$ and nRV for this simulated neuron (Figure 1d-e). As can be seen, the FFA method uncovers concurrent reduction in both spiking irregularity ($n\Psi$) and rate variability (nRV) as well as the asynchrony in this reduction ($n\Psi$ reduction happening after nRV). The overlaid empirical estimate shows the best possible estimate from the data if one had access to a large number of repetitions for each trial with the same ISI parameters (say >200 realizations of the point process with the same θ_i).

While the FFA method has nice theoretical underpinnings and pedagogical value, in practice fitting the asymptotic behavior of FF as a function of time-bin can be noisy especially for the intercept (small change in slope translating to large swings in the intercept). Therefore, we used a more robust estimator of $n\Psi$ and nRV by adapting a previously introduced method (Vinci et al., 2016) and allowing for variations of $n\Psi$ in time (see methods for details). This modified method which we refer to as the Kass method enables us to estimate concurrent changes in $n\Psi$ and nRV with higher fidelity compared to FFA (Figure 1d-e). Thus, along with the FFA slope and intercepts for each region, in practice the Kass method will be used to track sources of variability in time throughout this study. Note that both FFA and Kass methods estimation require sufficiently large number of trials and spikes per time-bin to give accurate estimates of $n\Psi$ and nRV . However as noted, the Kass estimate seems to be more robust compared to FFA across a range of parameters (Figure S1).

Normalized point-process variability ($nPPV$) which represents the inherent randomness of spike generation can be quantified using a measure which we refer to as CV_{local}^2 that estimates the inverse of shape parameter κ . This measure which is based on a local CV measure (known as CV2) is shown to provide robust and unbiased estimation of shape parameter (Holt et al., 1996; Pachitariu et al., 2010) (see methods). In the current simulated example, there are no other sources of $n\Psi$ other than $nPPV$. Hence, as illustrated in figure 1e, CV_{local}^2 also matches the $n\Psi$ estimate both tracking well the changes in $\frac{1}{\kappa}$. In the following sections, we will show that in

general nPPV is the lower bound of $n\Psi$ and that in addition, spiking irregularity is increased by other non-stationarities during the trial.

The fact that the firing rate of real neurons can change during the relatively large time-bins can raise a concern about the inflated estimates in $n\Psi$. However, our analysis and simulation results show that as long as the spiking irregularity remains unchanged during this time period our estimates of $n\Psi$ is largely robust to the underlying rate changes (Figure S2).

Post-stimulus enhancement of $n\Psi$ and reduction of nRV in cerebellum

Cerebellar neurons show sizeable changes in their variability in response to different conditions (Häusser & Clark, 1997; Payne et al., 2019; Shin et al., 2007; ten Brinke, Heiney, Wang, Proietti-Onori, Boele, Bakermans, Medina, Gao, & De Zeeuw, 2017). We used our method to parse rate variability and spiking irregularity across neurons ($n=268$) recorded from Interposed Nucleus (IpN) of mice during the Pavlovian eyeblink conditioning task (ten Brinke, Heiney, Wang, Proietti-Onori, Boele, Bakermans, Medina, Gao, & De Zeeuw, 2017). Figure 2a shows peristimulus time histogram (PSTH) response of an example IpN neuron time locked to the conditioned stimulus (CS) onset (LED light). 250ms after the CS onset the unconditioned stimulus (US: air-puff) was delivered. As can be seen the example neuron showed strong excitation following CS onset and just prior to the US delivery. Figure 2b top shows an initial decrease in the average FF followed by a sustained increase. As can be seen, this temporal pattern was largely due to very similar pattern of change in rate variability (nRV). This pattern of change in nRV and in particular its rise after about 500ms from CS can be easily verified by looking at the spiking raster plot (Figure 2a

bottom). Interestingly, despite the initial decrease in nRV and FF, $n\Psi$ showed a robust post stimulus increase. Notably the increase in $n\Psi$ was only partially due to the increase in nPPV as measured by CV_{local}^2 pointing to extra factors elevating the observed spiking irregularity.

Given the observed changes in $n\Psi$, methods that assume a constant ϕ (i.e. constant $n\Psi$ or spiking irregularity) can misestimate nRV. For instance, the VarCE method which attempts to estimate VEC by finding the largest constant ϕ ($n\Psi$) that keeps VEC positive during the trial, can become error prone:

$$VEC = VarCE = Var(N_T) - \phi E[N_T] \quad (3)$$

The VarCE method is also based on parsing sources of variability using the law of total variance in which VarCE and ϕ here are measures of VEC and $n\Psi$, respectively. Note that even in cases when ϕ is constant, VarCE can only measure VEC up to additive shifts (Figure S3a). More importantly, if the true ϕ changes and in the absence of changes in VEC, VarCE erroneously reports changes in VEC (Figure S3b). In the case of our cerebellar neural example, VarCE estimated the trends of change in VEC accurately but overestimated it particularly during the time when $n\Psi$ was changing (Figure 2c top). More importantly, since estimated ϕ in VarCE was constant, it grossly misestimated the component of variability that was due to spiking irregularity VEC (i.e. $\phi E[N_T]$) (Figure 2c bottom).

Similar to the neural example shown, the pattern of FF reduction and subsequent increase was also observable across the cerebellar IpN population and seemed to be driven primarily by the temporal dynamics in nRV. Once again during the same period $n\Psi$ showed a robust post stimulus

increase which was accompanied by concurrent increase in nPPV which was nevertheless only partially explained the increase observed in $n\Psi$. The post stimulus reduction in nRV and increase in $n\Psi$ can also be verified using the FFA method by looking at the linear asymptotic behavior of FF which showed intercept increase and slope decreased after the stimulus onset.

Simulation results show that in the absence of within trial non-stationarities, $n\Psi$ will be the same as nPPV as measured by CV_{local}^2 regardless of particular ISI distributions used (Figure 3a-b) and the existence of rate variability (Figure 3c). Both FFA and Kass methods track the ideal values of $n\Psi$ and nRV with good accuracies in such cases (Figure 3a-c). So, what can cause the mismatch between $n\Psi$ and nPPV seen in Figure 2b, e? Figure 3d-f shows three possible scenarios in which $n\Psi$ can be inflated beyond nPPV including bursting (Ponce-Alvarez et al., 2010), rate noise (Ermentrout et al., 2008) or rate switching (Deco & Hugues, 2012; Song et al., 2018) within the time-bin. All of these effects create stochastic non-stationarities during the time-bin and can increase spiking irregularity. In the case of rate switching with low transition probability there is also a concurrent increase in between trial rate variability which is well-tracked by both FFA and Kass methods. Note that in all cases CV_{local}^2 proves to be a good estimator of nPPV reflecting its robustness to rate fluctuations within and between the course of the trial. Thus, one may use $n\Psi - CV_{local}^2$ as a measure of spike irregularity induced by sources other than the point process itself which may also interact with nPPV to increase $n\Psi$.

Post-stimulus enhancement of $n\Psi$ canceling reduction of nRV in substantia nigra

Next, we checked response variability in another subcortical area namely substantia nigra reticula (SNr). This region is particularly informative for our analysis since here individual neurons can show large dynamic range of excitatory or inhibitory responses to visual objects based on their past value memory (excitation to low value objects and inhibition to high value objects). Figure 4a-b show responses of example neurons to high value and low value objects, respectively. For the neuron with inhibitory response a post stimulus reduction in FF was observed as expected. Similar to what was observed in cerebellum this decrease was largely due to a drop in post-stimulus rate variability (nRV) rather than spiking irregularity ($n\Psi$) which instead showed a robust post stimulus increase (Figure 4c). Interestingly, the neuron with excitatory response also showed a drop in nRV . However, this drop was more than offset by the increase in the spiking irregularity such that post stimulus FF in this neuron did not show the usually reported quench in spiking variability (Figure 4d). As expected, the radical changes in $n\Psi$ in both neurons meant that the estimates of VEC using VarCE were erroneous (Figure 4e-f). Moreover, there was a gross misestimation of EVC using VarCE especially in the case of the neuron with the inhibitory response. This is because methods with constant ϕ assumption (equivalently constant $n\Psi$), inevitably predict a drop in Ψ for inhibitory responses which turned to be the opposite of what was observed for this neuron (Figure 4e).

Across the population of recorded neurons, stimulus onset was associated with a drop in nRV , on the other hand, $n\Psi$ showed a robust increase in response to the stimulus onset revealing enhanced spiking irregularity (Figure 4g-h). Similar to the cerebellar neuron, population CV_{local}^2 also increased following the stimulus onset suggesting a concurrent increase in $nPPV$ which

nevertheless failed short of explaining the larger increase in $n\Psi$ (Figure 4h). The decrease in nRV and increase in $n\Psi$ can be visualized by the FFA method as well (Figure 4i). Note that unlike the cerebellar IpN neurons due to relatively large size of $n\Psi$, dynamics of the FF and is dissimilar for this region (Figure S4). Indeed, SNr is one brain region which also violates the well accepted reduction of post stimulus FF (M. M. Churchland et al., 2010).

As shown in equation 2, these results were obtained using normalized measures of between and within trial variabilities to control for firing rate change during the stimulus. Nevertheless, and to ensure the robustness of our findings to firing rate changes during the trial (a confounder), we validated our findings using two other approaches: (1) by using the mean-matching method which chooses subsets of neurons at each time point with the constraint of similar rate distributions (M. M. Churchland et al., 2010) and (2) by repeating the analysis separately for neurons in different response types of excitatory, inhibitory and null post-stimulus responses (see methods). Both scenarios resulted in essentially the same behavior for $n\Psi$ and nRV and relatively stable FF during the trial for the SNr population (figure S5).

Stimulus evoked changes in $n\Psi$ and nRV across cortical areas

It was previously shown that stimulus onset tends to reduce variability in cortical neural responses (M. M. Churchland et al., 2010). However, as stated previously due to lack of reliable methods for parsing the sources of spiking variability, it is not known how $n\Psi$ and nRV are affected by the stimulus onset. Here, we applied our methods on multiple cortical regions including areas V1, V2, middle-temporal (MT), premotor cortex (PMd) and ventrolateral

prefrontal cortex (vIPFC) recorded in macaque monkeys by different groups (M. M. Churchland et al., 2010; Ghazizadeh et al., 2018; Zandvakili & Kohn, 2015). As can be seen stimulus presentation or motor action caused a sizeable increase in average firing rate and a decrease in FF of all examined areas (Figure 5a). Moreover, in all cases the reduction in FF was concurrent with stimulus related reduction of nRV (Figure 5b). However almost all areas showed concurrent changes in $n\Psi$ as well. With the exception of vIPFC, the spiking irregularity as measured by $n\Psi$ decreased during stimulus presentation across cortical regions. In vIPFC a transient post-stimulus increase in $n\Psi$ was observed. Note that despite the transient increase in $n\Psi$ in vIPFC, its FF was still quenched during stimulus presentation due to larger size of nRV compared to $n\Psi$ (note the y-scales, Figure S4). Indeed, in all areas one can verify that $n\Psi$ and nRV scaled by time-bin sum up to explain the total FF fluctuations across time (Figure S4). The changes in $n\Psi$ and nRV during stimulus on and off is also evident by looking at the intercept and slope of the FF in the time-bin (Figure 5c). Note that in regions with lower overall firing rate, larger time-bin is required for the asymptotic behavior of FF to emerge (compare 1000ms time-bin in V1 with 200ms time-bin in vIPFC). It is thus possible that for regions with low firing rate, fast temporal dynamics in $n\Psi$ and nRV not be detectable with low number of trials and large time-bins (such as transient increases similar to vIPFC). The relationship between nPPV and $n\Psi$ was more heterogenous between regions. While nPPV hardly changed in V1 and V2, it decreased during stimulus presentation in MT and vIPFC. Once again nPPV changes were relatively minimal and most of the changes in spiking irregularity was not due to nPPV (Figure 5b).

These results were further corroborated using mean matching across the full population as well as separately within clusters of neural responses showing excitatory, inhibitory or null response across all the cortical regions (Figure S6 mean-matching and Figure S7 response clustering). In particular the transient increase in $n\Psi$ in vIPFC was observed in all cases.

Different schemes for coordinated activity in the presynaptic network induce independent changes in $n\Psi$ and nRV

As mentioned previously, accurate estimation of components of neural variability constrains the viable neural mechanisms that underlie the observed neural responses and can serve as a valuable tool for cross region comparisons (A. K. Churchland et al., 2011; Deco & Hugues, 2012; Gómez-Laberge et al., 2016; Rajan et al., 2010; Shadlen & Newsome, 1994; Softky & Koch, 1993). CV of ISIs was previously reported to change as a function of correlations in pre-synaptic inputs both theoretically and using simulations (Salinas & Sejnowski, 2000). However, CV overestimates spiking irregularity in presence of rate fluctuations (Ponce-Alvarez et al., 2010); in contrast, as shown previously $n\Psi$ is robust to rate fluctuations (see Figure S2).

Specifically, we aimed to find some of the characteristics in the presynaptic network which control different sources of neural variability (i.e. $n\Psi$ and nRV). This is critical since the differences in the temporal dynamics of $n\Psi$ in vIPFC compared to other cortical areas or between subcortical and cortical areas, raise questions about underlying pre- and post-synaptic factors that derive spiking in different regions. Presynaptic network characteristics of the neural models are shown to have notable impact on the post-synaptic neuron variability (Cavallari et al., 2014;

Deco & Hugues, 2012; Litwin-Kumar & Doiron, 2012). Rate correlation (Doiron et al., 2016; Salinas & Sejnowski, 2000) as well as spiking synchrony (Denève & Machens, 2016; Gómez-Laberge et al., 2016; Okun & Lampl, 2008; Salinas & Sejnowski, 2001) in presynaptic network are among these characteristics. Moreover, there are a wealth of studies suggesting a tight balance in cortical neurons (Denève & Machens, 2016; Renart et al., 2010) which might also affect the variability of the post-synaptic neurons. Here, we showed how statistical properties of balanced presynaptic network can reproduce some of the pattern observed in $n\Psi$ and nRV for a simple current based leaky integrate and fire (LIF) neuron.

We first show that the degree of trial-to-trial rate correlation between excitatory and inhibitory neurons (E and I Pools) in the presynaptic network can result in variations in nRV with minimal changes in $n\Psi$ in the post-synaptic neuron. Here we assumed two presynaptic pools, one excitatory and the other inhibitory sending efference to one LIF neuron (Figure 6a). The response variability of the LIF neuron was examined as a function of between pool trial-to-trial correlation in firing rate parameter (see methods). Figure 6b shows the firing raster of the LIF neuron for multiple scenarios from fully uncorrelated to fully correlated firing rates between E and I pools. As can be seen the increased rate correlation resulted in lower rate variability (nRV) in the post synaptic LIF neuron. Quantitative analysis showed that indeed nRV was high for zero correlation but decreased steadily for higher correlations. On the other hand, rate correlation had little effect on $n\Psi$ or $nPPV$ (Figure 6d). The firing rate was also largely unaffected by the degree of E/I pool firing rate correlation (Figure 6c).

Next, it is shown that the degree of spiking correlation between and within E/I pools can result in reduction or enhancement of $n\Psi$ without affecting the nRV. Here we assumed two sub-pools (sub-pool #1 and #2) within each excitatory and inhibitory pools. In the first scenario, there was between pool spiking correlation (Figure 7a). In addition, a range of delays between spikes of excitatory and inhibitory neurons was considered (0, 2 and 4ms, aka tight balance (Denève & Machens, 2016)). Figure 7b shows simulated spiking realizations of E/I pools and the resulting spikes in the LIF neuron at 0.3 correlations (Smith & Kohn, 2008) and 0ms (left) or 4ms (right) delays. As can be seen, the spiking irregularity was higher at 0ms compared to 4ms delay in the post-synaptic neuron (Figure 7b top). Quantitative analysis confirmed the reduction in $n\Psi$ as a function of between pool spiking correlation for nonzero delays between E/I pools (Figure 7c). This manipulation did not affect nPPV nor the nRV. In this scenario the firing rate was also increased as the function of between pool spiking correlation. Similar results are found for $n\Psi$, nRV, nPPV and rate by increasing the number of sub-pools (Figure S8).

A different method to impose synchrony is to assume spike correlation within each E and I pools as well while excluding between pool correlation (Figure 7d). This scheme has been previously used to explain reduction of variability in response to partial inactivation of inputs from visual cortex (Gómez-Laberge et al., 2016). In this case increasing the within pool spiking correlation increased spiking irregularity as shown in the example simulated rasters (Figure 7e). Consistent with previous observations high correlations lead to bouts of excitatory activity, in absence of correlated inhibitory activity, which in part leads to burst like activity in the post-synaptic neuron (Gómez-Laberge et al., 2016). In contrast to between pool correlation, this increase in $n\Psi$ showed

opposite polarity with firing rate. Once again, this within pool correlation also had no effect on nRV or nPPV.

Together these results show the plausibility of independent changes in $n\Psi$ and nRV as they were controlled by different presynaptic mechanism (rate correlation in Figure 6 vs spiking correlation in Figure 7). Notably in both cases nPPV as measured by CV_{local}^2 showed little changes similar to observations in the real data (Figure 2, 4, 5).

Discussion

Neurons across the brain often show variable spiking pattern in response to the same stimulus or behavioral event (Carandini, 2004; M. M. Churchland et al., 2010; Shadlen & Newsome, 1994; Tomko & Crapper, 1974). This variability reflects both single cell level characteristics stemming from noisy sensory inputs (Schneeweis & Schnapf, 1999), stochastic synaptic transmission (Ribault et al., 2011), plasticity (Feldman, 2009), and adaptation (Clifford et al., 2007); as well as network level properties including excitatory inhibitory balance (Turrigiano, 2011), attention and arousal levels (Fontanini & Katz, 2008), and neuromodulation effects (Marder, 2012). The total variability from these sources can be formulated into two additive components: (1) variability arising from random generation of spikes given ISI parameter which we referred to as spiking irregularity ($n\Psi$), and (2) the trial-to-trial rate variation in ISI parameters which in the case of the renewal process stems from rate variability (nRV). By introducing new methods (FFA and Kass) for accurate estimation of nRV and $n\Psi$ in time, we were able to reveal concurrent changes in

both components across multiple subcortical and cortical regions during the stimulus presentation in comparison to the spontaneous activity. Importantly, we showed that the previously reported quench in neural variability is solely due to reduction of nRV which was observed across all subcortical and cortical regions tested (A. K. Churchland et al., 2011; M. M. Churchland et al., 2010; Gómez-Laberge et al., 2016; Lombardo et al., 2018; Poland et al., 2019) (Figures 2,4,5). However, unlike what was previously assumed spiking irregularity was not constant but showed robust concurrent changes with nRV. $n\Psi$, unlike nRV, did not necessarily decrease during stimulus presentation. In fact, lPN neurons in the cerebellum and SNr neurons showed robust increases in post-stimulus $n\Psi$ (Figures 2,4). The increase in $n\Psi$ was also observable in the cortical neurons in the vIPFC region (Figure 5). The increase in spiking irregularity in SNr was so prominent that it canceled the reduction in rate variability resulting in the absence of usual quench in overall variability as measured by FF.

Specifically, for lPN neurons $n\Psi$ increased in response to the stimulus onset and also showed a sustained post-stimulus increase in FF after an initial dip (Figure 2). Purkinje cells (PC) which are presynaptic inputs to the lPN neurons are also shown to change their spiking regularity pattern in response to the stimulus onset. They also showed enhanced spiking irregularity due to inputs from molecular layer interneurons (Häusser & Clark, 1997; Payne et al., 2019; Shin et al., 2007). Although optogenetic manipulations in PCs have not shown sizable effects of nPPV on behavior (Payne et al., 2019), the changes in spiking regularity still has information about the presynaptic inputs and network level characteristics. The SNr neurons also showed increased $n\Psi$ in response to the stimulus onset invariant of the rate polarity such that post-stimulus FF remained almost

constant despite the drop in nRV (Figure 4). The robust enhancement of $n\Psi$ in these subcortical regions compared to most cortical regions examined encourage further investigation on potential significance of spiking irregularity for the neural code and behavior and its underlying mechanism in different brain regions in future work.

Furthermore, nPPV, estimated by local measures of CV (Nawrot et al., 2008), has been previously used to explain the total variance caused by spiking irregularity (M. M. Churchland et al., 2010). However, our results showed that nPPV estimated by CV_{local}^2 is not able to explain the total spiking irregularity observed in neural data ($n\Psi$). We proposed several possible mechanisms that could explain this extra portion of spiking irregularity such as bursting, within trial rate noise as well as rate switching which were previously shown to be present in many brain regions and simulations based on neural networks (Deco & Hugues, 2012; Ermentrout et al., 2008; Ponce-Alvarez et al., 2010; Song et al., 2018). All these factors were shown to inflate $n\Psi$ beyond predictions of nPPV (Figure 2). In addition, network modelling results of between and within pool correlations in presynaptic E/I spikes were also shown to change spiking irregularity beyond inherent point process variability (Figure 7). This temporally correlated pre-synaptic activity has also been shown to be a way in which bombarding cortical neurons leads to post-synaptic spike irregularity simulations (Gómez-Laberge et al., 2016; Salinas & Sejnowski, 2000). Thus $n\Psi$ estimates relative to CV_{local}^2 can be used to gauge the influence of sources of spike irregularity in addition to nPPV across brain regions. More research is required to both formulate the relation of these two spiking irregularity components as well as the underlying mechanisms responsible for the inflation of $n\Psi$ beyond nPPV and its possible role in neural processing and behavior.

We note that both methods used for estimation of variability components nRV and $n\Psi$ (i.e. FFA and Kass) require sufficient number spikes per time-bin (Figure S1). Thus, for neurons with low firing rates this would necessitates large time-bin and loss of temporal resolution in estimation of nRV and $n\Psi$ dynamics. Note that the requirement for the sufficient number of spikes per time-bin is due to the nature of doubly stochastic point process and exists even for idealized simulations with access to a large number of trial repetitions with the same parameter (i.e. empirical estimate, Figure S1). The relatively large time-bin used for estimation of $n\Psi$ may raise a concern about the changes in firing rate within time-bin to affect the variability estimates. However, our simulation results show that the spiking irregularity estimates are robust to such firing rate changes during the time-bin (Figure S2). As this figure illustrates, for different patterns of firing rate (Figure S2b-f) in comparison to the spontaneous activity (Figure S2a), the CV_{local}^2 and $n\Psi$ remain largely unchanged. In case of large inhibition or excitation there could be some over and underestimation of our methods. In order to ensure the robustness of our main conclusions on the real neural data, we have implemented two additional strategies to control for variations in firing rate namely mean-matching and separately considering post-stimulus excitation, inhibition and null-responses for the variability analysis. Importantly, these additional controls resulted in virtually the same estimates of variability components and confirmed our main conclusions about relative roles of nRV and $n\Psi$ in total post-stimulus FF.

Different neurons show different spiking irregularity behaviors in vitro and in vivo; some of these characteristics are solely observed in vivo (network level variability) but not in vitro e.g. regularity

of cortical pyramidal cells in vitro in spite of their irregular behavior in vivo. In contrast, in some other cases spiking irregularity is present in both in vivo and in vitro e.g. interneurons of the visual cortex (Englitz et al., 2008; Stiefel et al., 2013). In addition, blockage of tonic inhibitory inputs to the Purkinje cells of the cerebellum also leads to more regular spiking pattern (Häusser & Clark, 1997). In this work, we used LIF current based neural simulations to show effects of network input current properties on different sources of FF which models in vivo portion of variability. However, conductance based models are recently gaining ground as more viable alternatives to explain synaptic effects seen in real neurons (Cavallari et al., 2014; Richardson, 2004). The choice of synaptic model (current or conductance based) affects the spiking irregularity sensitivity of the post-synaptic neuron to the presynaptic firing rate. On the one hand, conductance-based neurons show persistent spiking irregularity roughly invariant of the presynaptic rate, consistent with our observations in V1, V2, MT, PMd cortical regions (Figure 5). On the other hand, the current-based neurons show sizable changes in their spiking irregularity with respect to the presynaptic firing rate which are consistent with our subcortical (Figure 2,4) as well as vIPFC evaluations (Figure 5) (Cavallari et al., 2014). Thus, further evaluation of different sources of variability arising from these network models could provide new insights and diagnostic tools for proposing underlying neuronal and neural circuit mechanisms of the observed variability component across various brain regions.

Our results revealed a ubiquitous quench in post-stimulus nRV across many cortical and subcortical regions. Importantly we also find significant concurrent changes in $n\Psi$ which represents spiking irregularity in all areas examined. However, unlike nRV the pattern of post-

stimulus changes in $n\Psi$ was heterogenous across regions. Previous work (M. M. Churchland et al., 2010) has indicated a robust reduction in network noise and a constant private noise during the stimulus. Our results do not address shared and private variability components across a population. However, the fact that both $n\Psi$ and nRV components showed concurrent post stimulus changes across neurons and had plausible network level implementations (Figure 6-7) suggest that they contribute to network level noise. On the other hand, $nPPV$ seemed to be resistant to change in real data and thus may represent the private noise component for each neuron (Nawrot et al., 2008; Ponce-Alvarez et al., 2010). Interestingly, previous simulations and in vivo injections (Miura et al., 2006) have revealed that the balanced network of excitatory and inhibitory networks are plausible models explaining the constant $nPPV$ phenomena as used in our simulations. Analysis of simultaneously recorded neurons using the Kass method (Vinci et al., 2016) which is readily available to address network level organization of $n\Psi$ and nRV across the population can further address factors underlying event related changes in neurons response variability components.

Materials and Methods

Subjects and datasets:

Neural Data:

MT and PMd data (n = 7031, 4264 sessions respectively) were recorded in macaques (M. M. Churchland et al., 2010) and are available online at <https://churchland.zuckermaninstitute.columbia.edu/content/code>.

The IpN data is from deep cerebellar nucleus of mice (ten Brinke, Heiney, Wang, Proietti-Onori, Boele, Bakermans, Medina, Gao, & De Zeeuw, 2017) and is available at (ten Brinke, Heiney, Wang, Proietti-Onori, Boele, Bakermans, Medina, Gao, & Zeeuw, 2017). We only used the first 270 sessions which were recorded during conditioning. V1 and V2 datasets (n = 1272, 296 respectively) are from anesthetized macaque monkeys in 5 different sessions (Zandvakili & Kohn, 2019). We used the session with largest number of neurons for each of these two regions and merged all different grating conditions. vIPFC and SNr data is recorded in macaques (Ghazizadeh et al., 2018; Ghazizadeh & Hikosaka, 2021). For all datasets, we removed low-rate neurons (sessions) with less than 3 spike/bin average for the given time-bin (number of used sessions are indicated in Figure 5).

Simulated spike trains:

Spike trains in each trial were viewed as renewal point processes and ISIs were drawn from various distributions including Poisson, Gamma, and Inverse Gaussian processes (Figure 1a). For Figure 1b and Figure 3 Gamma distribution was used. In Figure 3d bursts were added to the spike trains by adding 5 consecutive spikes with ISI = 3ms on 5% of spikes. For rate noise (Figure 3e) we

generated a subsampled white noise with $\Delta t = 50\text{ms}$ which was added to the rate of each trial randomly. The rate switching was simulated using a Markov Process with two fast and slow rate states (Figure 3f). The transition rates from slow to fast and fast to slow state are determined by k_{sf} and k_{fs} respectively (Song et al., 2018). The process on average is in the fast state with probability p_f :

$$p_f = \frac{k_{sf}}{k_{sf} + k_{fs}} \quad (1)$$

where we set $k_{fs} = 0.005\text{ms}^{-1}$ and $p_f = 0.4$, hence the $k_{sf} \approx 0.003\text{ms}^{-1}$. In order to constrain the average firing rate of the total process we set the slow rate based on the desired fast and average rate. The fast rate was set to $r_f = 25\text{Hz}$ and the slow rate would be determined using the following formulation:

$$\bar{r} = p_f r_f + (1 - p_f) r_s \quad (2)$$

For average rate of $\bar{r} = 10\text{Hz}$ the slow rate equals $r_s \approx 7.14\text{Hz}$.

Between trial rate variability (RV) was imposed by drawing rates of different trials from a gaussian distribution with different standard deviations specified for each figure (Figure 1b, 3c).

Network simulations:

We used LIF neuron with $N_E = N_I = 50$ neurons with the following formulation (Softky & Koch, 1993). We further set the parameters as in (Gómez-Laberge et al., 2016; Van Vreeswijk & Sompolinsky, 1996):

$$v(t) = v_0 + \tau \frac{dv}{dt} + rI, \quad (3)$$

where $v_{th} = -55mv$, $v_0 = -65mv$, $v_{min} = -70mv$, time constant $\tau = 20ms$ with synaptic resistance of $r = 80M\Omega$ and refractory period $t_{ref} = 1ms$. The current I was:

$$I = \left(\sum_i^{N_E} J_i^E(t) + \sum_j^{N_I} J_j^I(t) \right) \quad (4)$$

where J_i^E and J_j^I are presynaptic currents by each E/I pool neurons. We modeled the synaptic current response function by $te^{-\frac{t}{t_p}}$, where $t_p = 0.5ms$ for both excitatory and inhibitory pools.

The presynaptic E/I pools (50 neurons/pool) were generated using correlated Poisson processes (Macke et al., 2009). We added a Laplacian distributed noise to the ISIs of the generated correlated neurons with mean zero and standard deviation of 2ms. For the correlation tightness we changed the mean to 2-4ms for inhibitory neurons in order to shift inhibitory neurons after excitatory ones on average. The rates of presynaptic pools in Figure 6 at each trial are generated using the bivariate normal distribution as illustrated in Figure 6b.

Spike count variability decomposition:

Fano Factor Asymptote (FFA) Method:

Let $I_1, I_2, \dots, I_n$ denote the ISIs of a neural spike train. If we consider the distribution of each I_j from an i.i.d. distribution the resulting point process is a renewal process. We leverage law of total variance and law of total expectation to decompose the spike count variability given the ISI distribution parameters in each trial θ_i .

$$\text{Var}(N_T) = \underbrace{\mathbb{E}[\text{Var}(N_{T_i}|\theta_i)]}_{\text{EVC}} + \underbrace{\text{Var}(\mathbb{E}[N_{T_i}|\theta_i])}_{\text{VEC}}, \quad (5)$$

$$E[N_T] = E[E[N_T|\lambda_i]] \quad (6)$$

The first term in equation (5) represent the expected variance of count (EVC) while the second term represents variance of expected count (VEC) which was termed VarCE previously (A. K. Churchland et al., 2011).

We first derive equation (5) for doubly stochastic Poisson Process (exponentially distributed ISIs). For the ISIs from exponential distribution we have:

$$\text{Var}(N_T) = E[N_T] = \lambda T \quad (7)$$

Substituting in (5) we have:

$$\text{Var}(N_T) = E[\lambda_i T] + \text{Var}(\lambda_i T) = T \times E[\lambda_i] + T^2 \times \text{Var}(\lambda_i) \quad (8)$$

From (6) we have:

$$E[N_T] = T \times E[\lambda_i] \quad (9)$$

Moreover, from (7-9) we have:

$$\text{FF}(N_T) = \frac{\text{Var}(N_T)}{E[N_T]} = 1 + T \times \frac{\text{Var}(\lambda_i)}{E[\lambda_i]} \quad (10)$$

Hence, FF for the doubly stochastic Poisson process is a linear function of time bin T . As expected for stationary Poisson Process $\text{FF} = 1$.

In contrast to Poisson process, the aforementioned formulations do not have a closed form for general point processes. However, for renewal processes estimates of mean spike count and its variance can be calculated in the limit of large time bins and is shown to depend only on the mean and variance of the ISI distribution (Buckland & Cox, 1964). If $E[\text{ISI}_i] = \frac{1}{\lambda_i}$ and $\text{Var}(\text{ISI}_i) = \sigma_i^2$, as such:

$$\lim_{T \rightarrow \infty} E[N_T | \lambda] = \lambda T, \quad \lim_{T \rightarrow \infty} \text{Var}(N_T | \lambda, \sigma^2) = \lambda^3 \sigma^2 T \quad (11)$$

By substituting (11) in (5) we have:

$$\lim_{T \rightarrow \infty} \text{Var}(N_T) = T \times \Psi + T^2 \times \text{RV} \quad (12)$$

where

$$\Psi = E[\lambda_i^3 \sigma_i^2], \quad \text{RV} = \text{Var}(\lambda_i) \quad (13)$$

In addition, by substituting (11) in (6) the same equation (9) is derived for general renewal process. Fano factor (FF), the ratio of neural variability to its mean, is in effect a measure of neural count variability normalized by its mean. The decomposition presented in equations 12 allows one to parse FF into two components. Using equation (12) and (9), one can derive the asymptotic FF for a general renewal process in relatively large time bins:

$$\lim_{T \rightarrow \infty} \text{FF}(N_T) = \lim_{T \rightarrow \infty} \frac{\text{Var}(N_T)}{E[N_T]} = n\Psi + T \times n\text{RV} \quad (14)$$

where

$$n\Psi = \frac{E[\lambda_i^3 \sigma_i^2]}{E[\lambda_i]}, \quad n\text{RV} = \frac{\text{Var}(\lambda_i)}{E[\lambda_i]} \quad (15)$$

$n\Psi$ and $n\text{RV}$ are normalized measures of Ψ and RV and also intercept and slope of FF in equation (14) respectively. Note that FF and $E[N_T]$ can be estimated empirically for specified time-bin (T). Note that the slope term in (10) and (14) representing the $n\text{RV}$ are identical and invariant of the distribution type. Rather, they are proportional to changes of rate parameter across trials.

18

The FFA method works by first estimating the FF as a function of time-bin and then by fitting its linear asymptotic ([0.8:1] range of the estimation time-bin by default) behavior for which the slope represents $n\text{RV}$ and the intercept measures $n\Psi$ (Figures 1a, c bottom, 2f, 4i, 5c).

Note that for a general doubly stochastic renewal process $CV_i^2 = \lambda_i^2 \sigma_i^2$, hence we have:

$$n\Psi = \frac{E[\lambda_i^3 \sigma_i^2]}{E[\lambda_i]} = E[CV_i^2] + \frac{\text{Cov}(\lambda_i, CV_i^2)}{E[\lambda_i]} \approx E[CV_i^2] \quad (16)$$

where the last approximation is by assuming that the CV^2 as a measure of nPPV is not related to neuron's firing rate; Note that while observations in cortical regions support this assumption (Ponce-Alvarez et al., 2010), it might not be true for all regions (Payne et al., 2019).

However, for the Gamma process in Figure 1b-e ($ISI \sim \text{Gamma}(\kappa, \theta)$, $\lambda_i = \frac{1}{\kappa \theta_i}$, $\sigma_i^2 = \kappa \theta_i^2$), where θ_i but not κ is randomly varying across trials, the equivalence of $n\Psi$ and CV^2 is exact:

$$n\Psi = \lim_{T \rightarrow \infty} \frac{E[\lambda_i^3 \sigma_i^2]}{E[\lambda_i]} = \frac{1}{\kappa} \frac{E\left[\frac{1}{\theta_i}\right]}{E\left[\frac{1}{\theta_i}\right]} = \frac{1}{\kappa} \quad (17)$$

$$CV^2 = \frac{\text{Var}(ISI)}{(E[ISI])^2} = \frac{\kappa \theta_i^2}{(\kappa \theta_i)^2} = \frac{1}{\kappa} \quad (18)$$

The same equality, with mild assumptions, is shown for the inverse-gaussian process in the last section (equations 25-27).

Modified Kass Method:

The Kass method was originally introduced for decomposition of spike count correlation (SCC) between neurons into two components firing rate correlation (FRC) and within-trial point process noise (Vinci et al., 2016). This method was shown to be able to reliably extract the latent FRC from the observed SCC by a non-parametric approach. Obviously, the problem of spike count

variance of a neuron can be considered as a special case of SCC between two neurons by substituting autocorrelations instead of cross correlations. Notably, the Kass method can be easily adapted to allow for a time-varying $n\Psi$ (aka ϕ), unlike previous methods which either took $n\Psi$ to be equal to one (Poisson process assumption) (Goris et al., 2014) or as a tuning variable to find relative RV respectively (A. K. Churchland et al., 2011).

Kass method requires a set of hyperparameters: (1) K which indicates the duration in which cross-correlation between two neurons is significant (duration of significant auto-correlation in our case) (2) bin-count (m) which indicates the number of bins that the time-bin is divided to (see (Vinci et al., 2016)). These parameters were set manually in the original implementation. In our implementation we used $m = 100$. For simulated data, the parameter K was set automatically for each neuron using its estimated shape parameter values (κ). We used equation (21) (see Local CV^2 method) to find the shape parameter of the gamma process. Using the shape parameter and rate we first simulated a gamma process and empirically estimated the cross-correlation. K is estimated by the number of correlated lags (minimum correlation $\alpha \geq 0.01$). This method can be helpful in cases where there are low number of spikes per bin or limited number of trials. We further manually tuned the K parameter for real datasets to get more stable results.

VarCE Method:

VarCE finds the largest constant ϕ such that VEC stays positive during the trial:

$$VEC = \text{VarCE} = \text{Var}(N_T) - \phi E[N_T] \quad (19)$$

We used the code provided by (A. K. Churchland et al., 2011) to compute the VarCE and ϕ .

1 Local CV² Method:

2 Several local measures of CV, as a proxy of spiking irregularity, have been introduced in order to
3 reduce sensitive to the rate variations in time (Davies et al., 2006; Holt et al., 1996; Miura et al.,
4 2007; Ponce-Alvarez et al., 2010; Shinomoto et al., 2003). Theoretically for gamma process CV is
5 only related to the shape parameter. Robust and unbiased estimation of the shape parameter
6 could be achieved using the CV2 measure (Holt et al., 1996; Pachitariu et al., 2010). Normalized
7 point-process variability (nPPV) which represents the inherent randomness of spike generation
8 can be quantified using inverse of the shape parameter estimate which we refer to as CV_{local}^2 .
9 Although the gamma distribution of ISI has previously been examined by different evaluation on
10 fitted distribution (Li et al., 2017; Maimon & Assad, 2009), even in non-renewal case, like
11 occurrence of bursting, CV_{local}^2 would persistently estimate the hidden gamma shape parameter
12 (Pachitariu et al., 2010).

13
14 We used gamma shape estimator for non-stationary gamma processes to estimate the
15 normalized point-process variability (nPPV) based on equation (17-18). CV2 is a measure of
16 spiking irregularity which is robust to within trial non-stationarities (Holt et al., 1996). $CV2(i)$ is
17 computed by the following formulation:

$$18 \quad CV2(i) = \frac{|ISI_{i+1} - ISI_i|}{(ISI_{i+1} + ISI_i)/2}, \quad (20)$$

19 The shape parameter of the gamma process can be estimated using CV2 as follows (Pachitariu et
20 al., 2010):

$$21 \quad \hat{\kappa} = \frac{2}{\langle CV2^2 \rangle} - \frac{1}{2}, \quad (21)$$

where, $\langle . \rangle$ denotes average. We then used this unbiased estimator to calculate the nPPV of gamma process which equals $\frac{1}{\kappa}$ (equation 17-18) and which we refer to as CV_{local}^2 :

$$CV_{\text{local}}^2 = \frac{2\langle CV^2 \rangle}{4 - \langle CV^2 \rangle} = \frac{1}{\hat{\kappa}} \quad (22)$$

Mean Matching:

Mean matching selects sub-population of neurons/conditions for each time-bin in trial period so as to the rate distribution remains constant for all time-bins. We used the code provided by the (M. M. Churchland et al., 2010) (Figure S5a-6).

Response polarity clustering:

We used pre- and post-stimulus periods of the tasks to compute the average spike count for the two periods of each neuron. We then calculated the p-values for differences between pre- and post-stimulus rate of each neuron to cluster them into (1) excited (positive and $p < 0.05$), (2) null ($p > 0.05$), and (3) inhibited (negative and $p < 0.05$) groups. We then used these clusters to calculate the average n Ψ and nRV to test if the observed changes are due to rate fluctuations (Figure S5b-d, 7).

Double stochasticity in the presented LIF network models:

Approximation of network model in Figures 6 & 7 by simplifying assumption of integrate and fire (IF) model can be performed using Wiener process (no synaptic leakage). Hence, ISIs come from

1 inverse gaussian distribution with following first and second statistical moments (for $a_e \lambda_k^e >$
2 $a_i \lambda_k^i$).

$$3 \quad \mu_k = E[ISI_k] = \frac{\beta}{a_e \lambda_k^e - a_i \lambda_k^i}, \quad (23)$$

$$4 \quad \sigma_k^2 = \text{Var}(ISI_k) = \frac{\beta(a_e^2 \lambda_k^e + a_i^2 \lambda_k^i)}{(a_e \lambda_k^e - a_i \lambda_k^i)^3}, \quad (24)$$

5 where, β represents threshold of post-synaptic IF neuron for spiking, λ_k^e and λ_k^i are excitatory
6 and inhibitory firing rate of presynaptic neurons in trial k , and a_e and a_i are excitatory and
7 inhibitory synaptic weights which are considered constant and equal to each other across trials
8 (Tuckwell, 1988).

9 Hence, for large enough time bins, $n\Psi$ (or ϕ) can be computed by substituting (23) and (24) in
10 (15) as follows:

$$11 \quad n\Psi_{\text{stoch}} = \frac{E\left[\frac{\sigma_k^2}{\mu_k^3}\right]}{E\left[\frac{1}{\mu_k}\right]} = \frac{1}{\beta} \frac{E[a_e^2 \lambda_k^e + a_i^2 \lambda_k^i]}{E[a_e \lambda_k^e - a_i \lambda_k^i]} = \frac{1}{\beta} \frac{a_e^2 E[\lambda_k^e] + a_i^2 E[\lambda_k^i]}{a_e E[\lambda_k^e] - a_i E[\lambda_k^i]}, \quad (25)$$

$$12 \quad n\Psi_{\text{non-stoch}} = \frac{E\left[\frac{\sigma^2}{\mu^3}\right]}{E\left[\frac{1}{\mu}\right]} = \frac{1}{\beta} \frac{E[a_e^2 \lambda^e + a_i^2 \lambda^i]}{E[a_e \lambda^e - a_i \lambda^i]} = \frac{1}{\beta} \frac{a_e^2 \lambda^e + a_i^2 \lambda^i}{a_e \lambda^e - a_i \lambda^i}, \quad (26)$$

$$13 \quad CV_{\text{non-stoch}}^2 = \frac{\text{Var}(ISI)}{E[ISI]^2} = \frac{1}{\beta} \frac{a_e^2 \lambda^e + a_i^2 \lambda^i}{a_e \lambda^e - a_i \lambda^i}, \quad (27)$$

14 Equations (25-27) show that $n\Psi_{\text{stoch}} = n\Psi_{\text{non-stoch}} = CV_{\text{non-stoch}}^2$. Given simplifying
15 assumption of $a_e = a_i = a$, we have:

$$16 \quad n\Psi = \frac{a \nu + 1}{\beta \nu - 1}, \quad (28)$$

17 where, $\nu = \frac{\lambda^e}{\lambda^i}$ is the excitatory to inhibitory ratio.

Acknowledgements

This research was supported by the Research Program at the Sharif University of Technology (Grant #G980703) and intramural grant at School of Cognitive Science, IPM.

Author Contributions

S.F. devised the original ideas under A.G. supervision. S.F. and M.A.F did all the analysis and plots. All authors contributed to writing the manuscript.

Competing Interests: The authors declare that they have no competing interests.

Data and Code Availability: All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. All code and data are available upon request.

List of supplementary materials: Figures S1-S8

Figure Captions

Figure 1. Parsing components of variability in doubly stochastic renewal processes in time. (a)

Fano factor (FF) as function of time bin ($[0,1]$ ratios of time-bin) for simple (solid lines) and doubly stochastic (dashed lines) renewal processes with different ISI distributions (exponential, gamma and inverse gaussian). The asymptotic part of FF for large time bins (gray shaded area) is used for independent estimates of normalized spiking irregularity ($n\Psi$) as the intercept and the normalized rate variability (nRV) as the slope. Note the non-zero identical slopes only for doubly stochastic processes ($\lambda_i \sim N(\mu = 50\text{Hz}, \sigma = 10\text{Hz})$). (b) Rate parameter ($\lambda = \frac{\kappa}{\theta}$) and inverse of

shape parameter ($\frac{1}{\kappa}$) for a doubly stochastic gamma process as a function of time. Rate parameter was variable across trials according to $50 + n_i \times (1 - 5t_+e^{-2t_+})$ where $n_i \sim N(0,30)$ and t_+ is time for positive values and zero otherwise. (c) top, raster plot realization of color-coded trials from the stochastic gamma process (rate of each trial is indicated in panel b), bottom, FF as a

function of time bin within the specified time-bin (300ms), blue curve represents FF estimates as a function of time. Dashed line shows extrapolation of fitted line in linear part; red dot represents intercept. Slopes of each line are indicated by bar plots. (d) nRV estimates made empirically by assuming access to a large number of realizations given constant parameter (black) along with estimates using smoothed FFA (green) and modified Kass (red) methods. (e) same format as d but for $n\Psi$ estimates. Blue trace represents CV_{local}^2 estimates. All error bars indicating the sem.

Figure 2. Post-stimulus changes in rate variability and enhancement of spiking irregularity in cerebellum. (a) Example IpN neuron PSTH with standard deviation as the shaded area (top) and raster plot of 20 example trials (bottom) aligned to the conditioned stimulus (CS) onset (LED light) followed by unconditioned stimulus (US, air-puff) in 250ms. (b) Post-stimulus FF and nRV showing an initial decrease followed by sustained increase (top) and post-stimulus $n\Psi$ showing a robust increase. CV_{local}^2 also shows a moderate increase following stimulus onset. (f) VEC and EVC estimations using the Kass vs VarCE method (top and bottom, respectively). VarCE overestimates the rise in VEC (related to rate variability) and underestimates EVC (related to spiking irregularity). (d) Average PSTH of IpN neural population (n=268) neurons. Black and gray traces show PSTHs with 10ms and 300ms bin-widths, respectively. Longer bin-width is equal to time-bin used for variability estimation in the Kass and FFA methods. (e) Average FF and nRV (top) and $n\Psi$ and CV_{local}^2 (bottom) as a function of time within the sliding time-bin (300ms, 50ms sliding). Dashed line indicates CS onset, horizontal line indicates CS on period and arrow indicates US onset. (f) FF as a function of time bin within time-bin (300ms) centered on pre- and post-stimulus onset periods (-200 and 250ms respectively). Gray area indicates the linear zone at which

asymptotic FF is estimated. The increase in intercept and the decrease in slope is consistent with changes in $n\Psi$ and nRV respectively. Error bars are sem.

Figure 3. Within trial non-stationarities as well as normalized point process variability (nPPV) contribute to spiking irregularity ($n\Psi$). Empirical estimations of $n\Psi$ and nRV along with estimations made by FFA, Kass methods. Theoretical nPPV and its estimate using CV_{local}^2 is also shown. (a) Poisson process with rate 10Hz with no rate variability. (b) Gamma process ($\kappa = 2$) with no rate variability. (c) same as b but doubly stochastic with rate variability. (d) same as b but with bursting imposed with probability 0.05 on a given spike (5 spikes with 3ms ISI per burst). (e) same as b but with within trial white noise added to the constant rate ($\sim N(0,3)$ and $\Delta t = 50ms$, see methods). (f) same as b but with within trial state switching between high and low-rates imposed (see methods). Manipulations shown in d-f can increase $n\Psi$ beyond irregularity imposed by nPPV (red portion in empirical). In all cases CV_{local}^2 is high fidelity estimated of nPPV. State switching can also result in increased nRV . FFA and Kass methods provide good estimate of $n\Psi$ and nRV in all scenerios.

Figure 4. Post-stimulus drop in rate variability is compensated by concurrent rise of spiking irregularity in SNr. (a-f) Same format as Figure 2a-c but for one example SNr neuron showing inhibition (a, c, e) and another example SNr neuron showing excitation (b, d, f) to object onset. Raster plots show 15 sample trials for each neuron. (g-i) Same format as Figure 2d-f showing population average of 302 SNr neurons.

Figure 5. Concurrent changes in rate variability and spiking irregularity during stimulus presentation or movement execution across the cortex. (a) Average population PSTH across 5 different cortical regions responding to the stimulus presentation (areas V1, V2, MT, vIPFC) or movement execution (PMd). Black and gray traces show PSTHs with 10ms and EW bin-widths, respectively (time-bin = 1000, 1000, 400, 400, 200ms sliding by 100ms for V1/2 and 50ms for the rest respectively). Longer bin-width is equal to time-bin used for variability estimation in the Kass and FFA method (top), and corresponding change of FF during the trial (bottom). Horizontal line indicates stimulus on period and arrow indicates stimulus onsets. (b) Concurrent temporal changes in nRV (top) and in $n\Psi$ (bottom) across cortical regions. Temporal dynamics of CV_{local}^2 is plotted alongside $n\Psi$. (c) Same format as Figure 2f for each cortical area pre- and post- sensory or movement event.

Figure 6. Firing rate correlation between balanced presynaptic excitatory and inhibitory pools controls post-synaptic nRV. (a) Schematic of the network model including 50 Excitatory (E pool) and 50 Inhibitory (I pool) Poisson neurons. The synaptic current is modeled using the same kernel for E and I connections. (b) E/I rate distributions of different trials ($\sim N(40,4)$) with different correlation values (top) along with simulated sample rasters in three correlation conditions (correlations equal to 0, 0.6 and 1). (c) The effect of E/I rate correlation on the firing rate of post synaptic neuron. (d) The effect of E/I rate correlation on the nRV (left) and $n\Psi$ (right) of the post synaptic neuron. The effect on CV_{local}^2 is shown alongside $n\Psi$. Higher rate correlation leads to lower nRV in the post-synaptic neuron but had minimal effect on $n\Psi$ or CV_{local}^2 . Firing rate, nRV, $n\Psi$ and CV_{local}^2 for nonrandom presynaptic firing rate (deter) is shown alongside for comparison

in c-d. All error bars are SEM.

Figure 7. Spike correlation between and within balanced presynaptic excitatory and inhibitory pools controls post-synaptic $n\Psi$. (a) Schematic of the network model consisting of 50 Excitatory (E pool) and 50 Inhibitory (I pool) Poisson neurons. Each pool is further divided to two sub-pools. Between pool spike correlation is imposed among sub-pools (between E and I sub-pool#1 and between E and I sub-pool#2). (b) Simulated spiking pattern in the E/I sub-pools 1 and 2 and the resulting spiking in the postsynaptic neuron when there is a 0.3 spiking correlation between E/I sub-pools with zero delay (left) and 4ms delay (right) between E/I neurons (see methods). (c) Stronger between pool spike correlation decreases $n\Psi$ for nonzero delays (2-4ms). nRV or CV_{local}^2 are not affected by between pool spike correlations. A concurrent increase in firing rate is also observed for stronger between pool spike correlation for nonzero delays (2-4ms). (d) The same network as (a) but with only within E and I pool spike correlation. (e) same format as b but for within pool correlations of zero (left) and 0.3 (right). (f) Stronger within pool spike correlation increases $n\Psi$. nRV or CV_{local}^2 are not affected by within pool spike correlations. A concurrent increase in firing rate is also observed for stronger within pool spike correlation.

Supplemental Figure Captions

Figure S1. Dependence of $n\Psi$ and nRV estimates using Kass and FFA methods on the number of trials and the average number of spikes within the time-bin. (a) Each heat map shows estimated $n\Psi$ values for a gamma processe (10Hz) for different nRV (columns) and $n\Psi$ ($\frac{1}{\kappa}$, rows).

X-axis shows average number of spikes within time-bin and y-axis shows number of trials. Color-code is adjusted such that green mean correct estimation, while red and blue mean over- and under-estimation of the true $n\Psi$ values. The heat bar over each square heatmap indicates the empirical estimates (best possible estimate given the repetition of each trial data). (b-d) same format as a but for estimated nRV values using the Kass method, for estimated $n\Psi$ values using FFA and for estimated nRV using FFA, respectively.

Figure S2. $n\Psi$ estimates are robust to changes in firing rate during the time-bin. (a-f, top) firing rate pattern within an time-bin of 500ms for a gamma process with $\kappa = 2$. The shading shows firing rate variation across trials which should only effect nRV not $n\Psi$, as in Figure 3c. (a-f, bottom) Empirical estimates of $n\Psi$ (red bar) and the CV_{local}^2 estimates (green bar) along with $n\Psi$ estimates by Kass and FFA method.

Figure S3. VarCE can fail when estimating VEC. (a) Similar to Figure 1b example with a time varying nRV component but a constant $n\Psi$ component (constant κ) (top two rows). Empirical estimates of VEC and $n\Psi$ (same as ϕ) along with estimates made by Kass and VarCE methods. (b) same format as (a) but for a case when there is no rate variation ($VEC = T^2 \times RV = 0$) but there is time varying spiking irregularity (changing κ). VarCE method assigns fluctuations caused by $n\Psi$ to VEC while Kass method correctly disentangles the two different sources.

Figure S4. $n\Psi$ and $nRV \times T$ sum up to equal FF. Temporal dynamics of normalized spiking

irregularity $n\Psi$ and normalized rate variability nRV estimates and the fact that they sum up to almost fully explain the FF dynamics. Results are shown for all subcortical and cortical regions analyzed in the main paper. The relative size and contribution of $n\Psi$ and nRV in driving the fluctuations in FF can also be examined.

Figure S5. Robustness of $n\Psi$ and nRV pattern in SNr to variations in firing rate. (a) Mean-matching as well as clustering the SNr neurons to three groups with (b) excitatory response, (c) inhibitory response, (d) and null response yield similar patterns in $n\Psi$ and nRV estimates as shown in Figure 5. The gray curve in (a, top) plot shows the average and sem mean matched firing using a sub-selection of neurons for each time-bin. The gray curve in (b-d, top) is the PSTH evaluated with the same time-bin (200ms) as used for variability estimates $n\Psi$ and nRV .

Figure S6. Robustness of $n\Psi$ and nRV pattern across cortical regions by mean-matching. Mean matching across cortical regions yield similar patterns in $n\Psi$ and nRV estimates as shown in Figure 4. The gray curved in the PSTH plot shows the average mean-matched firing rate.

Figure S7. Robustness of $n\Psi$ and nRV pattern across cortical regions within response type clusters. Clustering cortical neurons to three groups with (a) excitatory response, (b) null response, (c) and inhibitory response yield similar patterns in $n\Psi$ and nRV estimates as shown in Figure 4.

Figure S8. Effect of number of sub-pools on post-synaptic $n\Psi$ changes as a function of between

1 **pool correlation.** $n\Psi$ and rate sensitivity to between pool correlation decreases as we increase
2 the number of sub-pools (for constant number of neurons). nRV and CV_{local}^2 remain unchanged
3 in all scenarios.

4

5

6

7

References

- Buckland, W. R., & Cox, D. R. (1964). Renewal Theory. *Biometrika*, 51(1/2), 290.
<https://doi.org/10.2307/2334228>
- Carandini, M. (2004). Amplification of trial-to-trial response variability by neurons in visual cortex. *PLoS Biol*, 2(9), e264.
- Cavallari, S., Panzeri, S., & Mazzoni, A. (2014). Comparison of the dynamics of neural interactions between current-based and conductance-based integrate-and-fire recurrent networks. *Frontiers in Neural Circuits*, 8(MAR), 1–23.
<https://doi.org/10.3389/fncir.2014.00012>
- Churchland, A. K., Kiani, R., Chaudhuri, R., Wang, X. J., Pouget, A., & Shadlen, M. N. (2011). Variance as a Signature of Neural Computations during Decision Making. *Neuron*, 69(4), 818–831. <https://doi.org/10.1016/j.neuron.2010.12.037>
- Churchland, M. M., & Abbott, L. F. (2012). Two layers of neural variability. *Nature Neuroscience*, 15(11), 1472–1474. <https://doi.org/10.1038/nn.3247>
- Churchland, M. M., Yu, B. M., Cunningham, J. P., Sugrue, L. P., Cohen, M. R., Corrado, G. S., Newsome, W. T., Clark, A. M., Hosseini, P., Scott, B. B., Bradley, D. C., Smith, M. A., Kohn, A., Movshon, J. A., Armstrong, K. M., Moore, T., Chang, S. W., Snyder, L. H., Lisberger, S. G., ... Shenoy, K. V. (2010). Stimulus onset quenches neural variability: A widespread cortical phenomenon. *Nature Neuroscience*, 13(3), 369–378. <https://doi.org/10.1038/nn.2501>
- Churchland, M. M., Yu, B. M., Ryu, S. I., Santhanam, G., & Shenoy, K. V. (2006). Neural variability in premotor cortex provides a signature of motor preparation. *Journal of Neuroscience*, 26(14), 3697–3712. <https://doi.org/10.1523/JNEUROSCI.3762-05.2006>

1 Clifford, C. W. G., Webster, M. A., Stanley, G. B., Stocker, A. A., Kohn, A., Sharpee, T. O., &
2 Schwartz, O. (2007). Visual adaptation: Neural, psychological and computational aspects.
3 *Vision Research*, 47(25), 3125–3131.

4 Davies, R. M., Gerstein, G. L., & Baker, S. N. (2006). Measurement of time-dependent changes in
5 the irregularity of neural spiking. *Journal of Neurophysiology*, 96(2), 906–918.

6 Deco, G., & Hugues, E. (2012). Neural network mechanisms underlying stimulus driven
7 variability reduction. *PLoS Computational Biology*, 8(3).
8 <https://doi.org/10.1371/journal.pcbi.1002395>

9 Denève, S., & Machens, C. K. (2016). Efficient codes and balanced networks. *Nature*
10 *Neuroscience*, 19(3), 375–382. <https://doi.org/10.1038/nn.4243>

11 Doiron, B., Litwin-Kumar, A., Rosenbaum, R., Ocker, G. K., & Josić, K. (2016). The mechanics of
12 state-dependent neural correlations. *Nature Neuroscience*, 19(3), 383–393.

13 Englitz, B., Stiefel, K. M., & Sejnowski, T. J. (2008). Irregular firing of isolated cortical
14 interneurons in vitro driven by intrinsic stochastic mechanisms. *Neural Computation*, 20(1),
15 44–64.

16 Ermentrout, G. B., Galán, R. F., & Urban, N. N. (2008). Reliability, synchrony and noise. *Trends in*
17 *Neurosciences*, 31(8), 428–434.

18 Feldman, D. E. (2009). Synaptic mechanisms for plasticity in neocortex. *Annual Review of*
19 *Neuroscience*, 32, 33–55.

20 Fisch, K., Schwalger, T., Lindner, B., Herz, A. V. M., & Benda, J. (2012). Channel noise from both
21 slow adaptation currents and fast currents is required to explain spike-response variability
22 in a sensory neuron. *Journal of Neuroscience*, 32(48), 17332–17344.

<https://doi.org/10.1523/JNEUROSCI.6231-11.2012>

Fontanini, A., & Katz, D. B. (2008). Behavioral states, network states, and sensory response variability. *Journal of Neurophysiology*, 100(3), 1160–1168.

Ghazizadeh, A., & Hikosaka, O. (2021). Common coding of expected value and value uncertainty memories in the prefrontal cortex and basal ganglia output. *Science Advances (in Press)*.

Ghazizadeh, A., Hong, S., & Hikosaka, O. (2018). Prefrontal Cortex Represents Long-Term Memory of Object Values for Months. *Current Biology*, 28(14), 2206-2217.e5.

<https://doi.org/10.1016/j.cub.2018.05.017>

Gómez-Laberge, C., Smolyanskaya, A., Nassi, J. J., Kreiman, G., & Born, R. T. (2016). Bottom-Up and Top-Down Input Augment the Variability of Cortical Neurons. *Neuron*, 91(3), 540–547.

<https://doi.org/10.1016/j.neuron.2016.06.028>

Goris, R. L. T., Movshon, J. A., & Simoncelli, E. P. (2014). Partitioning neuronal variability. *Nature Neuroscience*, 17(6), 858–865. <https://doi.org/10.1038/nn.3711>

Häusser, M., & Clark, B. A. (1997). Tonic synaptic inhibition modulates neuronal output pattern and spatiotemporal synaptic integration. *Neuron*, 19(3), 665–678.

[https://doi.org/10.1016/S0896-6273\(00\)80379-7](https://doi.org/10.1016/S0896-6273(00)80379-7)

Holt, G. R., Softky, W. R., Koch, C., & Douglas, R. J. (1996). Comparison of discharge variability in vitro and in vivo in cat visual cortex neurons. *Journal of Neurophysiology*, 75(5), 1806–

1814. <https://doi.org/10.1152/jn.1996.75.5.1806>

Koyama, S., & Kostal, L. (2014). The effect of interspike interval statistics on the information gain under the rate coding hypothesis. *Mathematical Biosciences and Engineering*, 11(1),

63–80. <https://doi.org/10.3934/mbe.2014.11.63>

1 Lansky, P., Sacerdote, L., & Zucca, C. (2016). The Gamma renewal process as an output of the
2 diffusion leaky integrate-and-fire neuronal model. *Biological Cybernetics*, 110(2–3), 193–
3 200. <https://doi.org/10.1007/s00422-016-0690-x>

4 Li, M., Xie, K., Kuang, H., Liu, J., Wang, D., Fox, G. E., Wei, W., Li, X., Li, Y., Zhao, F., Chen, L., Shi,
5 Z., Cui, H., Mao, Y., & Tsien, J. Z. (2017). Spike-timing pattern operates as gamma-
6 distribution across cell types, regions and animal species and is essential for naturally-
7 occurring cognitive states. *BioRxiv*, 145813. <https://doi.org/10.1101/145813>

8 Litwin-Kumar, A., & Doiron, B. (2012). Slow dynamics and high variability in balanced cortical
9 networks with clustered connections. *Nature Neuroscience*, 15(11), 1498–1505.
10 <https://doi.org/10.1038/nn.3220>

11 Lombardo, J. A., Macellaio, M. V., Liu, B., Palmer, S. E., & Osborne, L. C. (2018). State
12 dependence of stimulus-induced variability tuning in macaque MT. *PLoS Computational*
13 *Biology*, 14(10), 1–28. <https://doi.org/10.1371/journal.pcbi.1006527>

14 Macke, J. H., Berens, P., Ecker, A. S., Tolias, A. S., & Bethge, M. (2009). Generating spike trains
15 with specified correlation coefficients. *Neural Computation*, 21(2), 397–423.
16 <https://doi.org/10.1162/neco.2008.02-08-713>

17 Maimon, G., & Assad, J. A. (2009). Beyond Poisson: Increased Spike-Time Regularity across
18 Primate Parietal Cortex. *Neuron*, 62(3), 426–440.
19 <https://doi.org/10.1016/j.neuron.2009.03.021>

20 Marder, E. (2012). Neuromodulation of neuronal circuits: back to the future. *Neuron*, 76(1), 1–
21 11.

22 McIntosh, A. R., Kovacevic, N., & Itier, R. J. (2008). Increased brain signal variability

1 accompanies lower behavioral variability in development. *PLoS Computational Biology*,
2 4(7). <https://doi.org/10.1371/journal.pcbi.1000106>

3 Mišić, B., Mills, T., Taylor, M. J., & McIntosh, A. R. (2010). Brain noise is task dependent and
4 region specific. *Journal of Neurophysiology*, 104(5), 2667–2676.
5 <https://doi.org/10.1152/jn.00648.2010>

6 Mitchell, J. F., Sundberg, K. A., & Reynolds, J. H. (2007). Differential Attention-Dependent
7 Response Modulation across Cell Classes in Macaque Visual Area V4. *Neuron*, 55(1), 131–
8 141. <https://doi.org/10.1016/j.neuron.2007.06.018>

9 Mitchell, J. F., Sundberg, K. A., & Reynolds, J. H. (2009). Spatial Attention Decorrelates Intrinsic
10 Activity Fluctuations in Macaque Area V4. *Neuron*, 63(6), 879–888.
11 <https://doi.org/10.1016/j.neuron.2009.09.013>

12 Miura, K., Okada, M., & Amari, S. I. (2006). Estimating spiking irregularities under changing
13 environments. *Neural Computation*, 18(10), 2359–2386.
14 <https://doi.org/10.1162/neco.2006.18.10.2359>

15 Miura, K., Tsubo, Y., Okada, M., & Fukai, T. (2007). Balanced excitatory and inhibitory inputs to
16 cortical neurons decouple firing irregularity from rate modulations. *Journal of*
17 *Neuroscience*, 27(50), 13802–13812.

18 Nawrot, M. P., Boucsein, C., Rodriguez Molina, V., Riehle, A., Aertsen, A., & Rotter, S. (2008).
19 Measurement of variability dynamics in cortical spike trains. *Journal of Neuroscience*
20 *Methods*, 169(2), 374–390. <https://doi.org/10.1016/j.jneumeth.2007.10.013>

21 Okun, M., & Lampl, I. (2008). Instantaneous correlation of excitation and inhibition during
22 ongoing and sensory-evoked activities. *Nature Neuroscience*, 11(5), 535–537.

1 Omi, T., & Shinomoto, S. (2011). Optimizing time histograms for non-Poissonian spike trains.
2 *Neural Computation*, 23(12), 3125–3144. https://doi.org/10.1162/NECO_a_00213

3 Ostojic, S. (2011). Interspike interval distributions of spiking neurons driven by fluctuating
4 inputs. *Journal of Neurophysiology*, 106(1), 361–373.
5 <https://doi.org/10.1152/jn.00830.2010>

6 Pachitariu, M., Brody, C., Jun, P., & Holmes, P. (2010). Probabilistic models for spike trains of
7 single neurons [Citeseer]. In *Gatsby.Ucl.Ac.Uk*.
8 <http://www.gatsby.ucl.ac.uk/~marius/papers/SpikTrainStats.pdf>

9 Payne, H. L., French, R. L., Guo, C. C., Nguyen-Vu, T. D. B., Manninen, T., & Raymond, J. L.
10 (2019). Cerebellar purkinje cells control eye movements with a rapid rate code that is
11 invariant to spike irregularity. *ELife*, 8. <https://doi.org/10.7554/eLife.37102>

12 Pipa, G., Grün, S., & van Vreeswijk, C. (2013). Impact of spike train autostructure on probability
13 distribution of joint spike events. *Neural Computation*, 25(5), 1123–1163.
14 https://doi.org/10.1162/NECO_a_00432

15 Poland, E., Donner, T. H., Müller, K.-M., Leopold, D. A., & Wilke, M. (2019). Thalamus exhibits
16 less sensory variability quenching than cortex. *Scientific Reports*, 9(1), 1–12.

17 Ponce-Alvarez, A., Kilavik, B. E., & Riehle, A. (2010). Comparison of local measures of spike time
18 irregularity and relating variability to firing rate in motor cortical neurons. *Journal of*
19 *Computational Neuroscience*, 29(1–2), 351–365. [https://doi.org/10.1007/s10827-009-](https://doi.org/10.1007/s10827-009-0158-2)
20 0158-2

21 Poulet, J. F. A., & Petersen, C. C. H. (2008). Internal brain state regulates membrane potential
22 synchrony in barrel cortex of behaving mice. *Nature*, 454(7206), 881–885.

1 Rajan, K., Abbott, L. F., & Sompolinsky, H. (2010). Stimulus-dependent suppression of chaos in
2 recurrent neural networks. *Physical Review E*, 82(1), 11903.

3 Rajdl, K., Lansky, P., & Kostal, L. (2020). Fano Factor: A Potentially Useful Information. *Frontiers*
4 *in Computational Neuroscience*, 14. <https://doi.org/10.3389/fncom.2020.569049>

5 Renart, A., De La Rocha, J., Bartho, P., Hollender, L., Parga, N., Reyes, A., & Harris, K. D. (2010).
6 The asynchronous state in cortical circuits. *Science*, 327(5965), 587–590.

7 Ribault, C., Sekimoto, K., & Triller, A. (2011). From the stochasticity of molecular processes to
8 the variability of synaptic transmission. *Nature Reviews Neuroscience*, 12(7), 375–387.

9 Richardson, M. J. E. (2004). Effects of synaptic conductance on the voltage distribution and
10 firing rate of spiking neurons. *Physical Review E - Statistical Physics, Plasmas, Fluids, and*
11 *Related Interdisciplinary Topics*, 69(5), 8. <https://doi.org/10.1103/PhysRevE.69.051918>

12 Rickert, J., Riehle, A., Aertsen, A., Rotter, S., & Nawrot, M. P. (2009). Dynamic encoding of
13 movement direction in motor cortical neurons. *Journal of Neuroscience*, 29(44), 13870–
14 13882. <https://doi.org/10.1523/JNEUROSCI.5441-08.2009>

15 Salinas, E., & Sejnowski, T. J. (2000). Impact of correlated synaptic input on output firing rate
16 and variability in simple neuronal models. *Journal of Neuroscience*, 20(16), 6193–6209.

17 Salinas, E., & Sejnowski, T. J. (2001). Correlated neuronal activity and the flow of neural
18 information. *Nature Reviews Neuroscience*, 2(8), 539–550.

19 Schneeweis, D. M., & Schnapf, J. L. (1999). The photovoltage of macaque cone photoreceptors:
20 adaptation, noise, and kinetics. *Journal of Neuroscience*, 19(4), 1203–1216.

21 Schölvinck, M. L., Saleem, A. B., Benucci, A., Harris, K. D., & Carandini, M. (2015). Cortical state
22 determines global variability and correlations in visual cortex. *Journal of Neuroscience*,

1 35(1), 170–178. <https://doi.org/10.1523/JNEUROSCI.4994-13.2015>

2 Shadlen, M. N., & Newsome, W. T. (1994). Noise, neural codes and cortical organization.

3 *Current Opinion in Neurobiology*, 4(4), 569–579. <https://doi.org/10.1016/0959->

4 4388(94)90059-0

5 Shadlen, M. N., & Newsome, W. T. (1998). The variable discharge of cortical neurons:

6 Implications for connectivity, computation, and information coding. *Journal of*

7 *Neuroscience*, 18(10), 3870–3896. <https://doi.org/10.1523/jneurosci.18-10-03870.1998>

8 Shimokawa, T., Koyama, S., & Shinomoto, S. (2010). A characterization of the time-rescaled

9 gamma process as a model for spike trains. *Journal of Computational Neuroscience*, 29(1–

10 2), 183–191. <https://doi.org/10.1007/s10827-009-0194-y>

11 Shin, S. L., Hoebeek, F. E., Schonewille, M., De Zeeuw, C. I., Aertsen, A., & De Schutter, E. (2007).

12 Regular patterns in cerebellar Purkinje cell simple spike trains. *PLoS ONE*, 2(5), 1–9.

13 <https://doi.org/10.1371/journal.pone.0000485>

14 Shinomoto, S., Miura, K., & Koyama, S. (2005). A measure of local variation of inter-spike

15 intervals. *BioSystems*, 79(1-3 SPEC. ISS.), 67–72.

16 <https://doi.org/10.1016/j.biosystems.2004.09.023>

17 Shinomoto, S., Shima, K., & Tanji, J. (2003). Differences in Spiking Patterns among Cortical

18 Neurons. *Neural Computation*, 15(12), 2823–2842.

19 <https://doi.org/10.1162/089976603322518759>

20 Smith, M. A., & Kohn, A. (2008). Spatial and temporal scales of neuronal correlation in primary

21 visual cortex. *Journal of Neuroscience*, 28(48), 12591–12603.

22 Softky, W. R., & Koch, C. (1993). The highly irregular firing of cortical cells is inconsistent with

temporal integration of random EPSPs. *Journal of Neuroscience*, 13(1), 334–350.
<https://doi.org/10.1523/jneurosci.13-01-00334.1993>

Song, S., Lee, J. A., Kiselev, I., Iyengar, V., Trapani, J. G., & Tania, N. (2018). Mathematical Modeling and Analyses of Interspike-Intervals of Spontaneous Activity in Afferent Neurons of the Zebrafish Lateral Line. *Scientific Reports*, 8(1), 11–13.
<https://doi.org/10.1038/s41598-018-33064-z>

Stiefel, K. M., Englitz, B., & Sejnowski, T. J. (2013). Origin of intrinsic irregular firing in cortical interneurons. *Proceedings of the National Academy of Sciences*, 110(19), 7886–7891.

Tan, A. Y. Y., Chen, Y., Scholl, B., Seidemann, E., & Priebe, N. J. (2014). Sensory stimulation shifts visual cortex from synchronous to asynchronous states. *Nature*, 509(7499), 226–229.

ten Brinke, M. M., Heiney, S. A., Wang, X., Proietti-Onori, M., Boele, H.-J., Bakermans, J., Medina, J. F., Gao, Z., & Zeeuw, C. I. De. (2017). *Eyelid behavior and spike activity of cerebellar interpositus nucleus neurons during eyeblink conditioning in awake behaving mice*. <https://doi.org/http://dx.doi.org/10.6080/K0B85697>

ten Brinke, M. M., Heiney, S. A., Wang, X., Proietti-Onori, M., Boele, H. J., Bakermans, J., Medina, J. F., Gao, Z., & De Zeeuw, C. I. (2017). Dynamic modulation of activity in cerebellar nuclei neurons during pavlovian eyeblink conditioning in mice. *eLife*, 6, 1–27.
<https://doi.org/10.7554/eLife.28132>

Tomko, G. J., & Crapper, D. R. (1974). Neuronal variability: non-stationary responses to identical visual stimuli. *Brain Research*, 79(3), 405–418.

Tuckwell, H. C. (1988). Introduction to Theoretical Neurobiology. In *Introduction to Theoretical Neurobiology*. <https://doi.org/10.1017/cbo9780511623271>

Turrigiano, G. (2011). Too many cooks? Intrinsic and synaptic homeostatic mechanisms in cortical circuit refinement. *Annual Review of Neuroscience*, 34, 89–103.

Van Vreeswijk, C., & Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, 274(5293), 1724–1726.
<https://doi.org/10.1126/science.274.5293.1724>

Vinci, G., Ventura, V., Smith, M. A., & Kass, R. E. (2016). Separating spike count correlation from firing rate correlation. *Neural Computation*, 28, 2709–2733.
https://doi.org/10.1162/NECO_a_00831

Voytek, B., Kramer, M. A., Case, J., Lepage, K. Q., Tempesta, Z. R., Knight, R. T., & Gazzaley, A. (2015). Age-related changes in 1/f neural electrophysiological noise. *Journal of Neuroscience*, 35(38), 13257–13265. <https://doi.org/10.1523/JNEUROSCI.2332-14.2015>

Waschke, L., Kloosterman, N. A., Obleser, J., & Garrett, D. D. (2021). II Perspective Behavior needs neural variability. *Neuron*, 1–16. <https://doi.org/10.1016/j.neuron.2021.01.023>

Zandvakili, A., & Kohn, A. (2015). Coordinated Neuronal Activity Enhances Corticocortical Communication. *Neuron*, 87(4), 827–839. <https://doi.org/10.1016/j.neuron.2015.07.026>

Zandvakili, A., & Kohn, A. (2019). *Simultaneous V1-V2 neuronal population recordings in anesthetized macaque monkeys*. <https://doi.org/http://dx.doi.org/10.6080/K0B27SHN>

Figure S3