Cell assembly dynamics of sparsely-connected inhibitory networks: a simple model for the collective activity of striatal projection neurons

David Angulo-Garcia*
CNR - Consiglio Nazionale delle Ricerche - Istituto dei Sistemi Complessi, via Madonna del Piano 10, I-50019 Sesto Fiorentino, Italy

Joshua D. Berke†
Department of Psychology, University of Michigan, Ann Arbor, 530 Church St., Ann Arbor, MI 48104, USA

Alessandro Torcini‡
CNR - Consiglio Nazionale delle Ricerche - Istituto dei Sistemi Complessi, via Madonna del Piano 10, I-50019 Sesto Fiorentino, Italy and INFN Sez. Firenze, via Sansone, 1 - I-50019 Sesto Fiorentino, Italy

Striatal projection neurons form a sparsely-connected inhibitory network, and this arrangement may be essential for the appropriate temporal organization of behavior. Here we show that a sparse inhibitory network of artificial Leaky-Integrate-and-Fire neurons can reproduce key features of striatal population activity, as observed in brain slices [*Carrillo-Reid et al., J. Neurophysiology* **99** *(2008) 1435–1450*]. In particular we develop a new metric to determine the conditions under which sparse inhibitory networks form anti-correlated cell assemblies with variable firing rates of individual cells. We find that in this parameter range, the network displays an input-specific sequence of cell assembly switching, and can optimally discriminate between similar inputs. Our results support the proposal [*Ponzi and Wickens, PLoS Comp Biol* **9** *(2013) e1002954*] that striatal network topology is set up to allow stimulus-selective, temporally-extended sequential activation of cell assemblies.

Author Summary

The striatum is a brain structure involved in many functional roles, from action selection to learning and motor control. Some noticeable experiments demonstrated that the striatum encodes information received from sensory areas by the alternating activation of small assemblies of Medium Spiny Neurons firing synchronously. We propose a simple mathematical model of inhibitory neurons which is able to capture some of the most relevant features of the striatal dynamics. The simplicity of the model allows to identify the minimal ingredients required by a neural network to display the rich behavior observed in the striatum. The derivation of an easily tractable model can allow a deeper understanding of the fundamental mechanisms at the basis of the striatal dynamics and eventually to elucidate in a near future how neural diseases, like Parkinson and Huntigthon disease, alter the normal behaviour of Medium Spiny Neurons.

* david.angulo@fi.isc.cnr.it
† jdberke@umich.edu
‡ alessandro.torcini@cnr.it

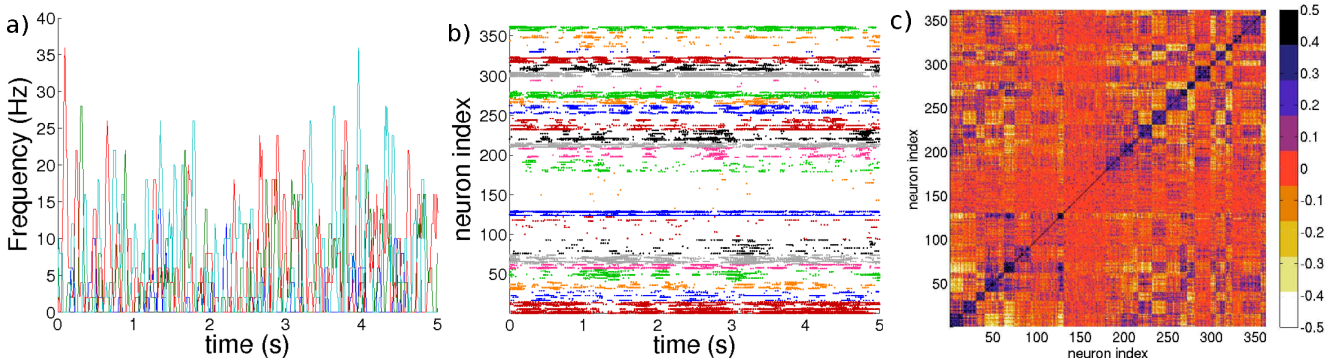

FIG. 1. **Cell activity characterization.** a) Firing rates ν_i of 5 selected neurons; b) raster plot activity, the firing times are colored according to the assembly the neurons belong to; c) cross-correlation matrix $C(\nu_i, \nu_j)$ of the firing rates. The neurons in panel b) and c) are clustered according to the correlation of their firing rates, by employing the *k-means* algorithm (for more details see Methods). The firing rates are calculated over overlapping time windows of duration 1 s, the origins of successive windows are shifted by 50 ms. The system is left to evolve during 10^6 spikes, after discarding an initial transient of 10^5 spike events. Other parameters used in the simulation: $g = 8$, $K = 20$, $N = 400$, $k_{mean} = N_{act}/15$, $\Delta V = 5$ mV and $\tau_\alpha = 20$ ms.

I. INTRODUCTION

The basal ganglia are critical brain structures for behavioral control, whose organization has been highly conserved during vertebrate evolution [44]. Disruptions of the basal ganglia underlie a wide range of human neurological and psychiatric disorders, but the specific computations normally performed by these circuits remain elusive. The largest component of the basal ganglia is the striatum, which appears to have a key role in adaptive decision-making based on reinforcement history [8], and in timing behavior on the scale of tenths of seconds to tens of seconds [22].

The great majority ($> 90\%$) of striatal neurons are GABAergic medium spiny neurons (MSNs), which project to other basal ganglia structures but also make local collateral connections within striatum [30, 49]. These local connections were proposed in early theories to achieve action selection through strong winner-take-all lateral inhibition [2, 11], but this idea fell out of favor once it became clear that MSN connections are actually sparse (nearby connection probabilities $\simeq 10 - 25\%$ [45, 47]), unidirectional and relatively weak [15, 46]. Nonetheless, striatal networks are intrinsically capable of generating patterns of cell activation that evolve over time, even in brain slice preparations with uniform excitatory input [6]. An important recent set of modeling studies proposed that the sparse connections between MSNs, though individually weak, can collectively mediate sequential switching between cell assemblies [35, 36], and may even be optimally configured for this purpose [37]. This proposal is of high potential significance, since sequential dynamics may be central to the striatums functional role in the organization and timing of behavioral output [4].

In their work [35–37], Ponzi and Wickens used conductance-based model neurons (with persistent Na^+ and K^+ currents [14]), in proximity to a bifurcation from a stable fixed point to a tonic firing regime. We show here that networks based on simpler leaky integrate-and-fire (LIF) neurons can also exhibit sequences of cell assembly activation. The simpler model allows us to more clearly identify the minimal ingredients needed to observe dynamics resembling that of the striatal MSN network, and we use a novel measure of structured bursting to identify the best parameters for observing such dynamics. Among other results, we show that the duration of GABAergic post-synaptic currents is crucial for the networks ability to discriminate different input patterns. Finally, we qualitatively replicate the observed response of striatal networks in brain slices to altered excitatory drive [6].

II. RESULTS

Measuring cell assembly dynamics

The model is composed of N leaky integrate-and-fire (LIF) inhibitory neurons, with each neuron receiving inputs from a randomly selected 5% of the other neurons (i.e. a directed Erdös-Renyi graph with constant in-degree $K = pN$, where $p = 5\%$). In addition, each neuron i is subject to an excitatory input current mimicking the cortical and thalamic inputs received by the striatal network (for more details see Eq. (2) in Methods).

We are interested in model parameters by which uniformly distributed inputs $I = \{I_i\}$, where $I_i \in [V_{th}, V_{th} + \Delta V]$, produce cell assembly-like sequential patterns in the network. The main aspects of the MSN activity can

be summarized as follows: (i) single neurons should exhibit large variability in the firing rates ($CV > 1$); (ii) the dynamical evolution of neuronal assemblies should reveal strong correlation within an assembly and anti-correlation with neurons out of the assembly. As suggested by many authors [33, 46] the dynamics of MSNs cannot be explained in terms of a *winners take all* (WTA) mechanism, which would imply a limited number of highly firing neurons, while the remaining would be silenced. Therefore we will search, in addition to the requirements (i) and (ii), for a regime where the number of neurons actively involved in the network dynamics is significant. Fig. 1 shows an example of such dynamics in the LIF model, with three pertinent similarities to previously observed dynamics of real MSN networks [6]. Firstly, cells are organized into correlated groups, and these groups are mutually anticorrelated (as evident from the cross-correlation matrix of the firing rates reported in Fig. 1 (c)). Secondly, individual cells within a group show irregular firing, that would be reflected as a coefficient of variation (CV) of inter-spike-intervals (ISIs) definitely greater than one (see Fig. 1 (a) and the black curve in Fig. 3 (b)) as observed experimentally for the dynamics of rat striatum *in-vitro* [46, 47]. Thirdly, a substantial fraction of neurons show activity, rather than a small number of active winners with the remainder silenced, as shown in the raster plot reported in Fig. 1 (b).

A novel metric for the structured cell assembly activity

The properties characterizing MSN activity can be quantified in terms of a single scalar metric Q_0, as follows:

$$Q_0 \equiv \langle CV \rangle_N \times \sigma(C(\nu_i, \nu_j)) \times n^* \quad ; \tag{1}$$

where $\langle \cdot \rangle_N$ denotes average over the ensemble of N neurons, $n^* = N_{act}/N$ is the fraction of active neurons N_{act} out of the total number, $C(\nu_i, \nu_j)$ is the $N \times N$ zero-lag cross-correlation matrix between all the pairs of single neuron firing rates (ν_i, ν_j), and $\sigma(\cdot)$ is the standard deviation of such matrix (for details see Methods). Our metric has been inspired by the metric introduced in [13] to identify some structure in the dynamics of cell groups for a detailed computational model of the striatal microcircuit. However, at variance with the metric proposed in [13], Q_0 takes into account also the large variability in the firing rates. Furthermore, the values of the parameters controlling the lateral inhibition (namely, the synaptic strength g and the duration of the post-synaptic potentials $\tau_\alpha \equiv 1/\alpha$) have been chosen within biologically meaningful ranges in order to maximize Q_0. As we show below, these choices allow us not only to obtain MSN-like network dynamics but also to optimize the networks computational capabilities, in the sense of producing a consistent, temporally-structured response to a given pattern of input while also discriminating between distinct, similar inputs.

The role of lateral inhibition

In this sub-section we examine how network dynamics is affected by the strength of inhibitory connections (Fig. 2). When these lateral connections are very weak (parameter g close to zero), the dominant input to each neuron is the constant excitation. As a result, most individual neurons are active (fraction of active neurons, n^*, is close to 1) and firing regularly (CV close to zero). As lateral inhibition is made stronger, single neuron firing becomes irregular, and n^* initially declines towards a minimum fraction of $\simeq 50\%$ (at $g = g_{min}$). As noted by Ponzi and Wickens [37], this is a winner-take-all (WTA)-like mechanism: faster-firing neurons depress or even silence the neurons to which they are connected. The *winners* (*losers*) are characterized by an effective input W_i which is on average supra-threshold (below-threshold), and we can definitely affirm that their firing activity is mean-driven (fluctuation-driven) [41] (for more details see Figs. S2 (c) and (f)). This reflects in the corresponding distribution $P(\overline{ISI})$ of the average interspike intervals ($\overline{ISI}$) (Fig. 2(b), red curve). *Winner* neurons produce peak of very short $\overline{ISI}$ (i.e. high firing rate) while the remaining neurons (*losers*) produce the long tail of the distribution extending up to $\overline{ISI} \simeq 10^3$ s. In the distribution of coefficients of variation (CV; Fig. 2(b) inset, red curve) the winner neurons produce a peak of very low CV (i.e. highly-regular firing), suggesting that they are not strongly influenced by the other neurons in the network and therefore have an effective input on average supra-threshold. Instead the *losers* are associated with a smaller peak at $CV \simeq 1$, confirming that their firing is fluctuation-driven.

Counterintuitively however, further increases in lateral inhibition strength result in increased neuronal participation, with n^* progressively returning towards $\simeq 1$. The same effect was previously reported by Ponzi and Wickens [37] in a different, more complex, model. When the number of active neurons returns almost to 100%, i.e. for sufficiently large coupling $g > g_{min}$, most of the neurons appear to be below threshold, as revealed by the distribution of the effective inputs W_i reported in Figs. S2 (c) and (f). Therefore in this case the network firing is essentially fluctuation-driven, as a matter of fact the $P(\overline{ISI})$ distribution is now characterized by a broader distribution and by the absence of the peak at short $\overline{ISI}$ (as shown in Fig. 2 (b), blue line; see also Figs. S2(a) and (d)). Furthermore the single neuron

dynamics is definitely irregular, as testified by the fact that the CV distribution is now centered around $CV \simeq 2$ (inset of Fig. 2 (b), blue line; see also Figs. S2(b) and (e)).

The transition among the two dynamical regimes, occurring at $g = g_{min}$, is due to a passage from a state where a part of the neurons identified as the *winners* were mean-driven and were able to depress all the other neurons, to a state at $g >> g_{min}$ where almost all neurons are fluctuation-driven and all the neurons contribute to the network activity. The transition occurs because at $g < g_{min}$ the fluctuations of the effective input currents W_i are small and insufficient to drive the losers towards the firing threshold (as shown in the insets of Fig. S2 (c) and (f)). At $g \simeq g_{min}$ the amplitude of fluctuations begin to be sufficient to lead some losers to rise again above threshold and to contribute to the number of active neurons. This will also reduce the winners' activity. For $g >> g_{min}$ the fluctuations of W_i are sufficient to restore all losers to the firing activity and at the same time no clear distinction among losers and winners can be done. The transition is due to the fact that not only the inhibitory action is proportional to the synaptic strength, but also the fluctuation amplitude increases linearly with g, at least for $g > g_{min}$ (as shown in Figs. S3(a) and (b)).

The reported results explain why the variability $\sigma(C)$ of the cross-correlation matrix has a non monotonic behaviour with g (as shown in the middle panel in Fig. 2(a)). At low coupling $\sigma(C)$ is almost zero, since the single neuron dynamics are essentially independent one from another, while the increase of the coupling leads to an abrupt rise of $\sigma(C)$. This growth is clearly associated to the fact that the neuronal population splits into winners and losers, which are definitely anti-correlated. The variability of the cross-correlation matrix achieves a maximum value for coupling slightly larger than g_{min}, where fluctuations in the effective currents begin to play a relevant role in the network dynamics. At larger coupling $\sigma(C)$ begins to decay towards a finite non zero value. These results confirm that the most interesting region to examine for the MSNs behaviour is the one with coupling $g > g_{min}$, as already suggested in [37].

The observed behaviour of CV, n^* and $\sigma(C)$ suggest that we should expect a maximum in Q_0 at some intermediate coupling $g > g_{min}$, as indeed we have found for both studied cases, as shown in Fig. 2 (c) and (d). The initial increase in Q_0 is due to the increase in CV and n^*, while the final decrease, following the occurrence of the maximum, is essentially driven by the decrease of $\sigma(C)$. For larger ΔV the neurons tend to fire regularly in a wider range of coupling at small g (see Fig. 2 (d)), thus indicating that due to their higher firing rates a larger synaptic inhibition is required to influence their dynamics. On the other hand, their bursting activity observable at large g is more irregular (see the upper panel in Fig. 2 (a); dashed line and empty symbols).

In order to understand if our metric Q_0 is capable of determining the range of parameters where the network can distinguish among different inputs, we perform the following analysis. The network is stimulated with two different stimuli: each stimulus is presented for a period T_{sw}, the input is then switched to the second stimuli for another time window of duration T_{sw}, and this procedure is repeated several times. During this stimulation protocol, the state transition matrix (STM) and the associated quantity ΔM_d^0 are evaluated. The metric ΔM_d^0 has been previously introduced in [37] to analyse the ability of the network to distinguish among two inputs. In particular, ΔM_d^0 is simply the difference between the average values of the STM measured during the presentation of each of the two stimuli (a detailed description of the procedure is reported in the sub-section *Discriminative and computation capability* and in Methods).

Furthermore, we consider a new metric ΔM_d, obtained by substituting in Eq. (1) the standard deviation of the correlation matrix with ΔM_d^0. As it can be appreciated from Figs. 2(c) and 2(d) the metrics ΔM_d and Q_0 behave in a quite similar manner, indicating that indeed Q_0 becomes maximal for parameter values where the network operates optimally in order to distinguish different stimuli. In particular, the maximal value of Q_0 is attained for $g = 4$ ($g = 8$) for $\Delta V = 1$ mV ($\Delta V = 5$ mV) corresponding to $A_{PSP} = 0.368$ mV ($A_{PSP} = 0.736$ mV), which can be considered as realistic values accordingly to the experimental results reported in the literature [33, 46, 47] (for more details see the model description reported in Methods).

The role of the post-synaptic time scale

The results shown so far have been obtained for slow synapses, namely by considering $\tau_\alpha = 20$ ms. In this section we will examine the effect of decreasing the post-synaptic decay time τ_α. The results are shown in Fig. 3(a), where it is clear that by decreasing τ_α the values of both metrics Q_0 and ΔM_d steadily decrease. To gain more insights on the role of the PSP in shaping the structured dynamical regime, we show for the same network the distribution of the CVs of the single cell, for three different τ_α (Fig. 3(b)). Narrow pulses ($\tau_\alpha \simeq 2$ ms) are associated with a distribution of CV values ranging from 0.5 to 1, with a predominant peak at one. By increasing τ_α one observes that the CV distributions shift to larger and larger CV values. Therefore, one can conclude that at small τ_α the activity is mainly Poissonian, while by increasing the decay time of the PSPs leads to more irregular behaviours, similar to the one observed in the MSN activity [23]. To confirm this analysis we have estimated also the distribution of the

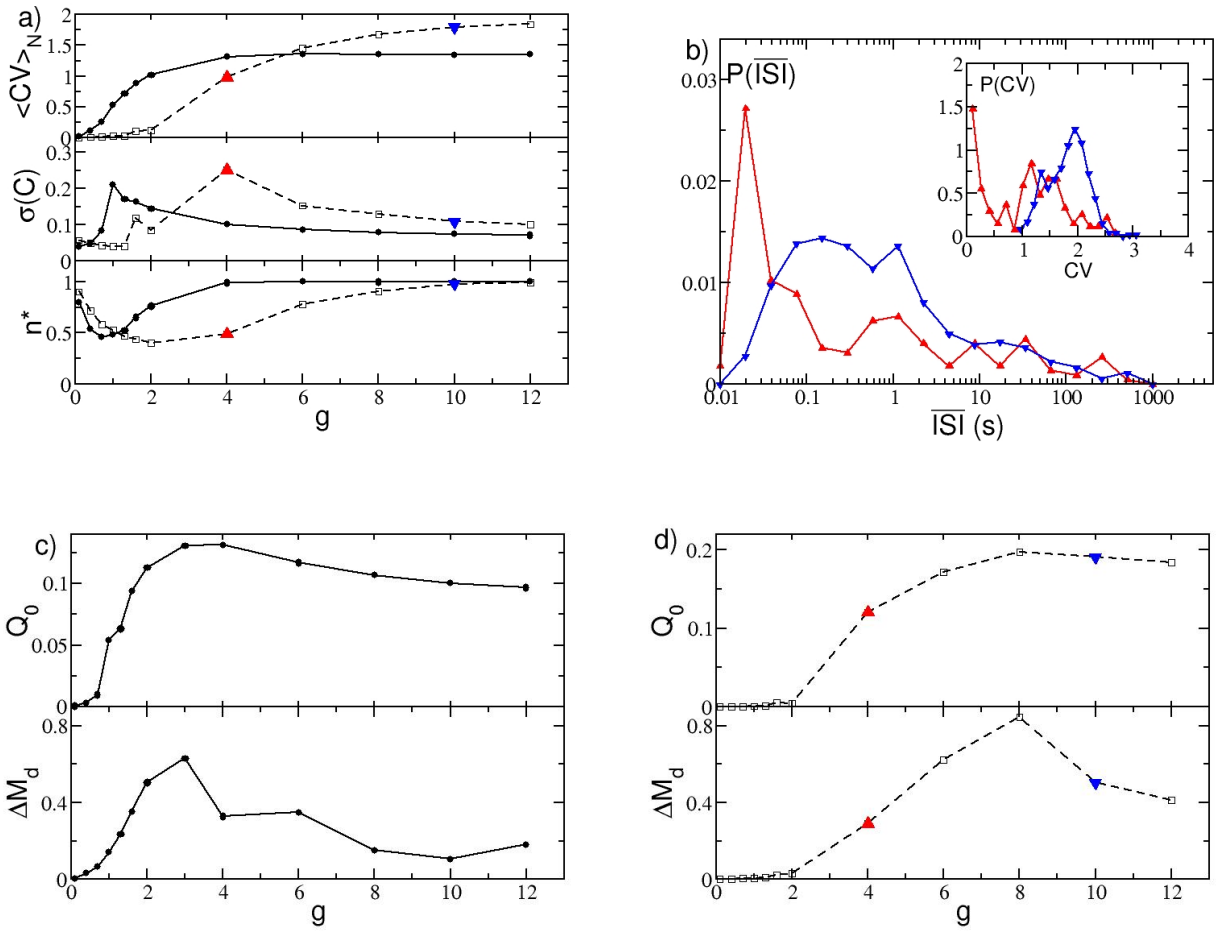

FIG. 2. **Metrics of structured activity vs lateral inhibition strength.** a) Parameters defining Q_0 and their dependence by the coupling g. From top to bottom: Averaged coefficient of variation CV, standard deviation of the cross-correlation matrix $\sigma(C)$, and the fraction of active neurons n^*. The solid (dashed) line refers to the case $\Delta V = 1$ mV ($\Delta V = 5$ mV). The minimum number of active neurons is achieved at $g = g_{min}$, this corresponds to a peak amplitude of the PSP $A_{PSP} = 0.064$ mV ($A_{PSP} = 0.184$ mV) for $\Delta V = 1$ mV ($\Delta V = 5$ mV) (for more details see Methods). b) Distributions $P(\overline{ISI})$ of the average ISI for a fixed $\Delta V = 5$ mV and for two different coupling strengths, $g = 4$ (red triangle-up symbol) and $g = 10$ (blue triangle-down). Inset, the distribution $P(CV)$ of the CV of the single neurons for the same two cases. c) Q_0 and ΔM_d, as defined in Eqs. (1) and (15), versus g for $\Delta V = 1$ mV. d) Same as c) for $\Delta V = 5$ mV.

CV_2: A CV_2 distribution with a peak around zero denotes a very regular firing, while a peak around 1 indicates the presence of long silent periods followed interrupted by rapid firing. Finally a flat distribution denotes a Poisson spike emission. It is clear from Fig. 3(c) that by increasing τ_α from 2 to 20 ms this leads the system from an almost Poissonian behaviour to a bursting dynamics, where an almost regular firing inside the burst (intra-burst) is followed by a long quiescent period (inter-burst) before starting again.

The distinct natures of the distributions of CV for short and long-tailed pulses raises the question of what mechanism underlies such differences. To answer this question we analyzed the distribution of the ISI of a single cell in the network for two cases: in a cell assembly bursting regime (corresponding to $\tau_\alpha = 20$ ms) and for Poissonian unstructured behavior (corresponding to $\tau_\alpha = 2$ ms). The dynamics of the two neurons are quite different, as seen by the results presented in Fig. 4(a) and (b). In both cases one observes a long tailed exponential decay of $P(ISI)$ corresponding to a Poissonian like behaviour. However the decay rate ν_D is slower for $\tau_\alpha = 20$ ms with respect to $\tau_\alpha = 2$ ms, namely $\nu_D \simeq 2.74$ Hz versus $\nu_D \simeq 20.67$ Hz. Interestingly, the main macroscopic difference between the two distributions arises at short times. For $\tau = 2$ ms, (see Fig. 4(b)) an isolated and extremely narrow peak appears at low $\mathrm{ISI}_0 \simeq 12$ ms, followed by a gap, and finally by the exponential decay. The origin of the gap resides in the fact $\mathrm{ISI}_0 >> \tau_\alpha$, because if the neuron is firing tonically with its period ISI_0 and receives a single PSP, this has time to decay almost to zero

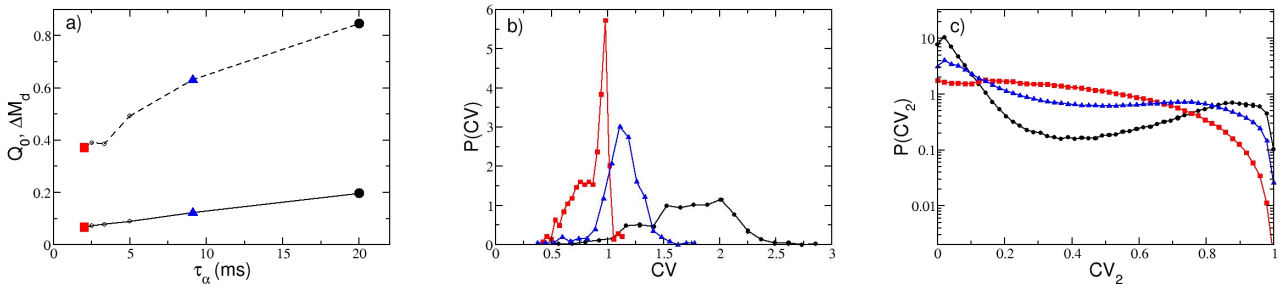

FIG. 3. **Metrics of structured activity vs post-synaptic time duration.** a) Metrics Q_0 (in solid line) and ΔM_d (dashed) as a function of the pulse width for the pair of parameter values $\{\Delta V, g\} = \{5\ \text{mV}, 8\}$ corresponding to the maximum in Q_0 in Fig. 2(d). Probability distribution functions $P(CV)$ ($P(CV_2)$) for the coefficient of variation CV (local coefficient of variation CV_2) are shown in b) (in c)) for three representative $\tau_\alpha = \{2, 9, 20\}$ ms, displayed by employing the same symbols and colors as indicated in a). Other parameters as in Fig. 1

before the next spike emission. Thus a single PSP will delay the next firing event of a fixed amount corresponding to the gap in Fig. 4(b). The arrival of several PSPs will then give rise to the exponential tail. In this case the contribution to the CV comes essentially from the exponential tail, while the isolated peak at ISI_0 has a negligible contribution.

On the other hand, if $\tau_\alpha > \text{ISI}_0$, as in the case reported in Fig. 4(a), $P(ISI)$ does not show anymore a gap, but instead a continuous distribution of values. This because now the inhibitory effects of the received PSPs sum up leading to a continuous range of delayed firing times of the neuron. The presence of this peak of finite width at short ISI in the $P(ISI)$ plus the exponential delay is at the origin of the observed $CV > 1$.

We would like to understand whether it is possible to reproduce similar distributions of the ISIs by considering an isolated cell receiving Poissonian distributed inhibitory inputs. In order to verify this, we simulate a single cell receiving K uncorrelated spike trains at a rate $\langle\nu\rangle_N$, or equivalently, a single Poissonian spike train with rate $K\langle\nu\rangle_N$. Here, $\langle\nu\rangle_N$ is the average firing rate of a single neuron in the original network. The corresponding $P(ISI)$ are plotted in Fig. 4 (c) and 4 (d), for $\tau_\alpha = 20$ ms and 2 ms, respectively. There is a remarkable similarity between the reconstructed ISI distributions and the real ones (shown in Fig. 4(a) and (b)) , in particular at short ISIs. These results demonstrate that the bursting activity of inhibitory coupled cells is not a consequence of complex correlations among the incoming spike trains, but rather a property related to intrinsic properties of the single neuron, like its tonic firing, the synaptic strength, and the post-synaptic time decay. Obviously this analysis cannot explain other collective effects, like the non trivial dependence of the number of active cells on the synaptic strength, discussed in the previous sub-section, or the emergence of the correlations and anti-correlations among the neural firing activities, subject of the next sub-section.

Discriminative and computational capability

In this sub-section we examine the ability of the network to perform different tasks: namely, to respond in a reproducible manner to equal stimuli and to discriminate between similar inputs by encoding the differences in distinct dynamical evolution. In particular, for this analysis we have always compared the responses of the network obtained for a set of parameters corresponding to the maximum Q_0 value shown in Fig. 2(d), where $\tau_\alpha = 20$ ms, and for the same parameters but with a shorter PSP decay time, namely $\tau_\alpha = 2$ ms.

To check for the capability of the network to respond to cortical inputs with a reproducible sequences of states of the network, we perform a simple experiment where two different inputs $I^{(1)}$ and $I^{(2)}$ are presented sequentially to the system. Each input stimulates for a time duration T_{sw} the system and then the stimulus is switched to the other one and this process is repeated for the whole simulation time. The raster plot measured during such experiment is shown in Fig. 5 (a) for $\tau_\alpha = 20$ ms. Whenever one of the stimuli is presented, a specific sequence of pattern activations can be observed. Furthermore, the sequence of emerging activity patterns is reproducible when the same stimulus is again presented to the system, as it can be appreciated by observing the patterns encircled with black lines in Fig. 5 (a). Recall that the clustering algorithm here employed to identify the different groups is applied only during the presentation of the first stimulus, therefore the sequential dynamics is more evident for that particular stimuli.

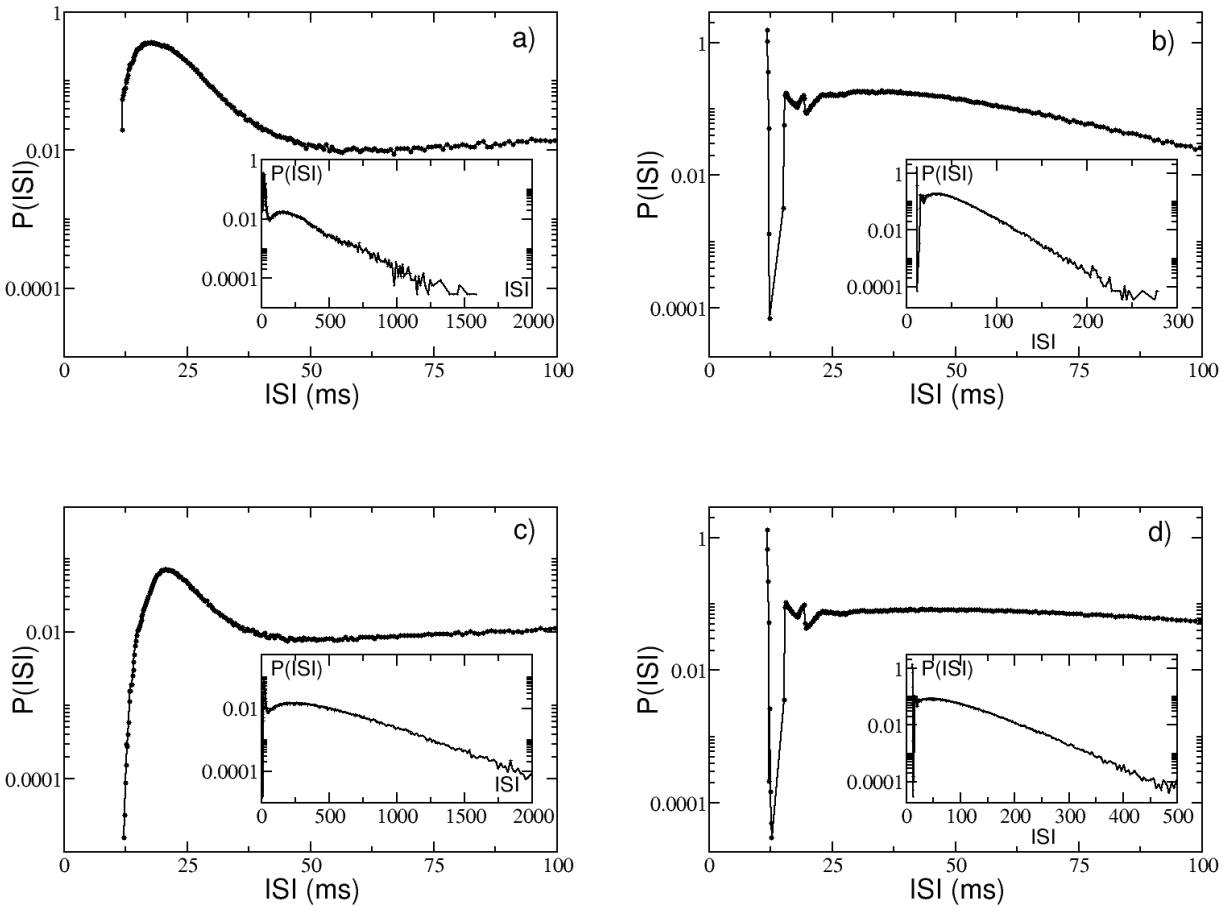

FIG. 4. **Single neuron statistics.** Distributions $P(ISI)$ for one representative cell in the network for a) $\tau_\alpha = 20$ ms and b) $\tau_\alpha = 2$ ms. The corresponding Poissonian reconstruction of the $P(ISI)$ are reported in c) and d). The network parameters are $\Delta V = 5$ mV and $g = 8$, and the others as in Fig. 1, both the examined neurons have $I_i = -45.64$ mV. For the Poissonian reconstruction the frequencies of the incoming uncorrelated spike trains are set to $\langle\nu\rangle_N \approx 75$ Hz ($\langle\nu\rangle_N \approx 88.5$ Hz) for $\tau_\alpha = 20$ms ($\tau_\alpha = 2$ms), as measured from the corresponding network dynamics. The distributions are obtained by considering a sequence of 10^9 spikes in the original network, and 10^7 events for the Poissonian reconstruction. In all plots the main figure displays the short-time ISI, while the inset is a zoom out of the distribution.

Furthermore, we can quantitatively calculate how similar is the firing activity in the network at different times during the experiment, this is done by estimating the STM. The similarity is quantified by computing the normalized scalar product of the instantaneous firing rates of the N neurons measured at time t_i and t_j. We observe that the similarity of the activity at a given time t_0 and at a successive time $t_0 + 2mT_{sw}$ is high, thus suggesting that the response to the same stimulus is similar, while it appears as uncorrelated with the response at a time corresponding to the presentation of a different stimulus, i.e. at $t_0 + (2n-1)T_{sw}$ (here, $n, m = 1, 2, 3...$). This results in a STM with a periodic structure of period T_{sw} with alternating high correlated blocks followed by low correlated blocks (see Fig. S5(b)). An averaged version of the STM is shown in Fig. 5 (b) (for details of the calculation see Methods). This matrix shows not only the reproducible nature of the system response but also the capability of the network to distinguish between the stimuli.

To verify that the network response was indeed related to the type of stimulus and not to the order of presentation, we repeated the same experiment by exchanging $I^{(1)}$ and $I^{(2)}$. The results confirm that the response is related to the presented stimulus and not to their order of presentation. Furthermore, to test if there was some memory effect that can modify the network response we performed a further test were the system dynamics was completely reset after the stimulation of each stimulus and before the presentation of the next one. As a matter of fact, we do not observe any relevant change in the network response. We repeated the experiment for another random realization of the inputs,

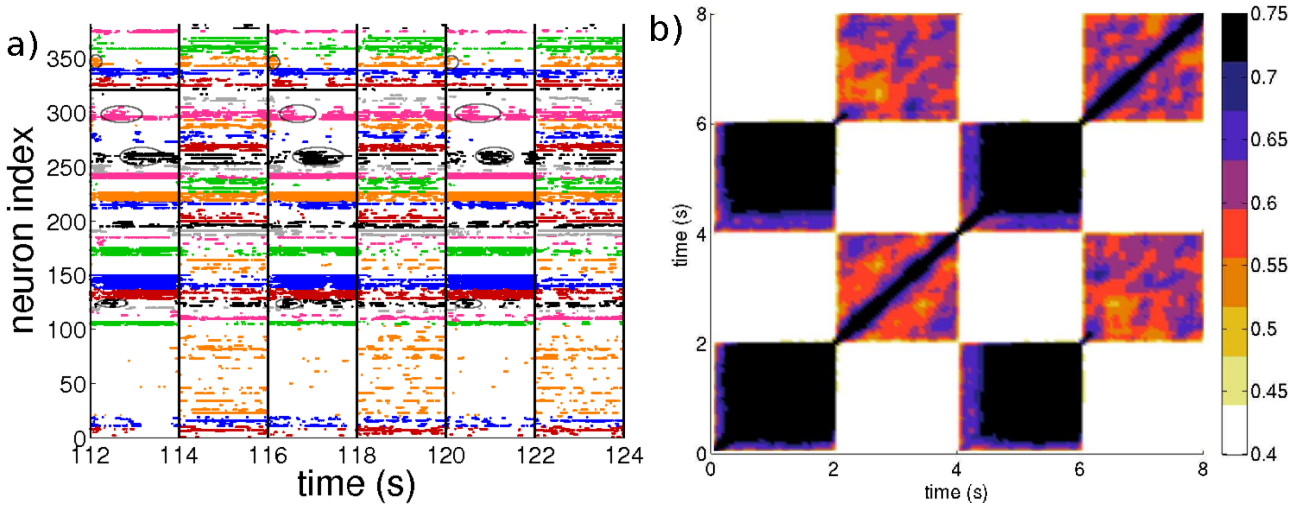

FIG. 5. **Sequential switching.** a) Raster plot associated to the two input protocols $I^{(1)}$ and $I^{(2)}$. The circles denote the clusters of active neurons appearing repetitively after the presentation of the stimulus $I^{(1)}$. Vertical lines denote the switching times between stimuli. b) Averaged State Transition Matrix $\overline{D}$, created by considering a $4T_{sw} \times 4T_{sw}$ sub-matrix (see methods section for details). The inputs $I^{(1)}$ and $I^{(2)}$ are different realization of the same random process, they are obtained by selecting N current values I_i from the flat interval $[V_{th},\, V_{th} + \Delta V]$. The input stimuli are switched every $T_{sw} = 2$ s. Number of clusters $k = 35$ in a). Other parameters as in Fig. 1.

noticing essentially the same features previously reported (as shown in Fig. S5(a-c)). Thus we can conclude that our results are robust.

We have finally examined the influence of the PSP time scale on the observed results, as shown in Fig. S5(d-f) for $\tau_\alpha = 2$ ms. For this value of τ_α the network does not reveal a large variability in the response showing a reduced number of patterns of activity. In particular, it responds in a quite uniform manner during the presentation of each stimulus, thus the corresponding STM shows just highly correlated blocks alternating to low correlated ones, but these blocks do not reveal any structure useful to distinguish among different inputs.

We proceeded to check the ability of the network to discriminate among similar inputs and how this ability depends on the temporal scale of the synaptic response. In particular, we tried to answer to the following question: if we present two inputs that differ only for a fraction f of the stimulation currents, which is the minimal difference between the inputs that the network can discriminate ? In particular, we considered a control stimulation $I^{(c)} = I_i \in [V_{th},\, V_{th}+\Delta V]$ and a perturbed stimulation $I^{(p)}$, where the stimulation currents differ only over a fraction f of currents I_i (which are randomly chosen from the same distribution as the control stimuli). We measure the differences of the responses to the control and to the perturbed stimulations by measuring, over an observation window T_E, the dissimilarity metric $d^f(t)$, defined in Methods. The time averaged dissimilarity metric $\bar{d}^f$ is reported as a function of f in Fig. 6 for two different values τ_α. It is clear that for any f-value the network with longer synaptic response always discriminates better between the two different stimuli than the one with shorter PSP decay. We have also verified that the metric is robust to the modification of the observation times T_E, this is verified because the dissimilarity $d^f(t)$ rapidly reaches a steady value (as shown in Fig. S6).

In order to better characterize the computational capability of the network and the influence due to the different duration of the PSPs, we have examined the response of the network to a sequence of three consecutive stimuli. In particular, we considered an analogous test to the two input protocol, where now three different stimuli are consecutively presented to the network. The system is stimulated with each input for a time period T_{sw}, and the stimulation sequence is repeated for the whole experiment duration T_E.

The output of the network can be represented by the instantaneous firing rates of the N neurons measured over a time window $\Delta T = 100$ ms, this is a high dimensional signal, where each dimension is represented by the activity of a single neuron. The similarity of the signals can be estimated by measuring how many dimensions are indeed explored in the phase space, more similar are the output signals among the neurons less directions are visited.

A principal component analysis (PCA) performed over $T_E/\Delta T$ observations of the N firing rates reveals that for $\tau_\alpha = 2$ ms the 80% of the variance is recovered already with a projection over a two dimensional sub-space (red bars in Fig. 7 (a)). On the other hand, for $\tau_\alpha = 20$ ms a higher number of principal components is required to reconstruct the dynamical evolution (black bars in Fig. 7 (a)), thus suggesting the higher computational capability of the system with longer PSPs [31]. In particular, for $\tau_\alpha = 20$ ms the network respond to a quenched stimulation, constant in time,

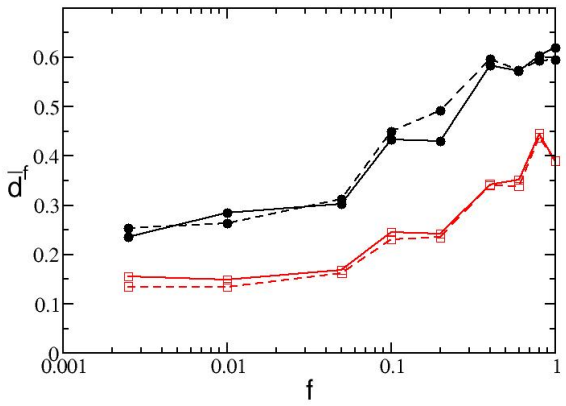

FIG. 6. **Pattern separation.** Average dissimilarity as a function of the fraction of shuffled inputs f for the values of $\tau_\alpha = 20$ms (black circles) and $\tau_\alpha = 2$ms (red squares) with two different observation windows $T_E = 2$s (continuous line) and $T_E = 10$s (dashed line). Other parameters used: $\Delta T = 50$ms, $\Delta V = 5$ mV. Remaining parameters as in Fig. 1.

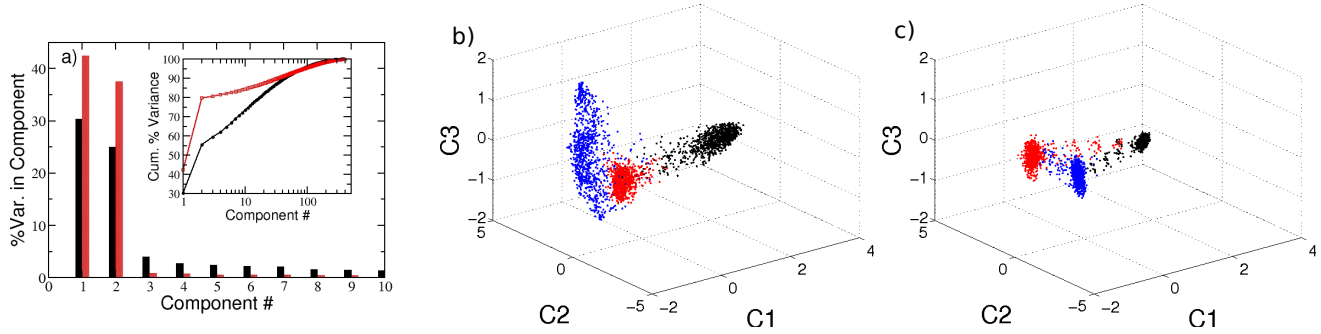

FIG. 7. **Computational capability of the network.** Characterization of the firing activity of the network, obtained as response to three consecutive inputs presented in succession. a) Percentage of the variance of the neuronal firing activity reproduced by each of the first 10 principal components. The inset displays the corresponding cumulative percentage as a function of the considered component. Filled black and shaded red (bar or symbols) correspond to $\tau_\alpha = 20$ ms and $\tau_\alpha = 2$ ms, respectively. Projection of the neuronal response along the first three principal components. for b) $\tau_\alpha = 20$ ms and c) $\tau_\alpha = 2$ ms. The three colors denote the response to the three different inputs, which are quenched stimulation currents randomly taken as $I^{(j)} \in [V_{th},\, V_{th} + \Delta V]$ for $j = 1, 2, 3$, the experiment is then performed as explained in the text.

with a temporal evolution in a high-dimensional space, thus indicating that in this case complex computations are indeed possible.

These results are confirmed by analyzing the projections of the firing rates in the subspace spanned by the first three principal components $(C1, C2, C3)$ shown in Fig. 7 (b) and (c) for $\tau_\alpha = 20$ ms and $\tau_\alpha = 2$ ms, respectively. The responses to the three different stimuli can be effectively discriminate by both networks, since they lie in different parts of the phase space. However, the response to the three stimuli correspond essentially to three fixed points for $\tau_\alpha = 2$ ms, while trajectories evolving in a higher dimension are associated to each stimulus for $\tau_\alpha = 20$ ms.

These analyses confirm that the network parameters selected by employing the maximal Q_0 criterion, are also suitable to maximize the system capability to respond in a reproducible way to different stimuli, as well as to discriminate among different inputs.

Response of the network to an increased level of excitability

The analysis here reported has been inspired by the experiment performed by Carrillo *et al.* [6]. In that experiment the authors considered a striatal network *in vitro*, which displays sporadic and asynchronous activity under control conditions. Therefore in order to induce spatio-temporal patterned activity they perfuse the slice preparation with

N-methyl-D-aspartate (NMDA). Since it is known that, similarly to what occurs in other microcircuits [10], NMDA administration brings about an excitatory tonic drive, which favours the occurence of up-states as well as of recurrent bursting activity [48]. The crucial role of the synaptic inhibition in shaping the patterned activity in striatal dynamics is also demonstrated in [6], by applying the $GABA_A$ receptor antagonist bicuculline to effectively decrease the inhibitory synaptic effect.

In our simple model, ionic channels and NMDA-receptors are not modeled; nevertheless it is possible to recreate the effect of NMDA administration by increasing the excitability of the cells in the network, and the effect of the bicuculline as an effective decrease in the synaptic strength. We then verify *at posteriori* whether these assumptions lead to results similar to those reported in [6].

In our model the single cell excitability is controlled by the parameter I_i. In this sense, the computational experiment consists in setting the system in a low firing regime corresponding to the control conditions with $I^{(c)} = \{I_i^{(c)}\} \in [-53, -49.5]$ mV and in enhancing, after 20 seconds, the system excitability to the range of values $I^{(e)} = \{I_i^{(e)}\} \in [-60, -45]$ mv, for another 20 seconds. This latter stage of the numerical experiment corresponds to the NMDA bath in the real experiment. The process is repeated several times and the resulting raster plot is coarse grained as explained in Methods (subsection *Synchronized Event Transition Matrix*).

From the coarse grained version of the raster plot, we calculate the *Network Bursting Rate* (NBR) as the fraction of neurons participating in a burst event in a certain time window. Whenever the instantaneous NBR is larger than the average NBR plus two standard deviations, this is identified as a synchronized bursting event (see Fig. 8(a)). In Fig. 8(b) we plot all the neurons participating in a series of $S_s = 20$ synchronized events. Here the switching times between control conditions and the regimes of increased excitability are marked by vertical dashed lines. Once identified a series of synchronized events, we measure their similarity in terms of the *Synchronized Event Transition Matrix* (SETM) (for more details see Methods), which is shown in Fig. 8(c). Furthermore, using the SETM we divide the synchronized events in clusters according to an optimal clustering algorithm [25] (see Methods). In the present case we have identified 3 distinct states (clusters), if we project the vectors $W_s(t)$ characterizing each single synchronized event (see Methods for a definition) on the two dimensional space spanned by the first two principal components $(C1, C2)$ we observe a clear division among the 3 states (see Fig. 8(d)). It is now important to understand whether the cells firing during the events classified as a state are the same or not. We observe that the groups of neurons recruited for each synchronized event, corresponding to a certain state, largely overlap, while the number of neurons participating to different states is limited. As shown in Fig. 8(e), the number of neurons participating to the events associated to a certain state is of the order of 40-50, while the coactive neurons (those participating in more than one state) ranges from 12 to 25. Furthermore, we have a core of 9 neurons which are firing in all states. Thus we can safely identify a distinct assembly of neurons active for each state.

As shown in Fig. 8(c), we observe, in analogy to what found in [6], that the system alternates his activity among the previously identified cell assemblies. In particular, the probability that two successive events belong to the same state is quite high $\simeq 25 - 40\%$; while the probability that the next synchronization event belongs to a different state is uniformly distributed among all the possible states with a probability of $\simeq 20\%$. The only exception is the strong correlation observed for the sequence of states $3 \to 2$ which has a probability of 50 % of occurrences.

If we repeat the same experiment after a long simulation interval $t \simeq 200$ s we find that the the dynamics can be always described in terms of few states (3-4), however the cells contributing to these states are different from the ones presently identified. This is due to the fact that the dynamics is in our case chaotic, as we have verified in the Supporting Information Text S1 (*Linear Stability Analysis*).

To check for the effect of bicuculline, the same experiment is performed again with a much smaller synaptic coupling, namely $g = 1$, the results are shown in Fig. 8(f-j). The first important difference can be identified in higher NBR values with respect to the previous analyzed case ($g = 8$) Fig. 8(f). This is due to the decreased inhibitory effect, which allows most of the neurons to fire almost tonically, and therefore a higher number of neurons participate in the bursting events. In Fig. 8(g) it is clearly visible a highly repetitive pattern of synchronized activity (identified as state 2, blue symbols), this state emerges immediately after that the excitability is enhanced. After this event we observe a series of bursting events, involving a large number of neurons (namely, 149), which have been identified as an unique cluster (state 1, red symbols). The system, analogously to what shown in [6], is now locked in an unique state which is recurrently visited until the return to control conditions. Interestingly, both synchronized events corresponding to state 1 and state 2 are highly correlated among them when compared with the $g = 8$ case, as seen by the SETM in Fig. 8(h). Despite this, it is still possible to identify both states when projected on the two dimensional space spanned by the first two principal components (see Fig. 8(i)). This high correlation can be easily explained by the fact that the neurons participating in state 2 are a subset of the neurons participating in state 1, as shown in Fig. 8(j).

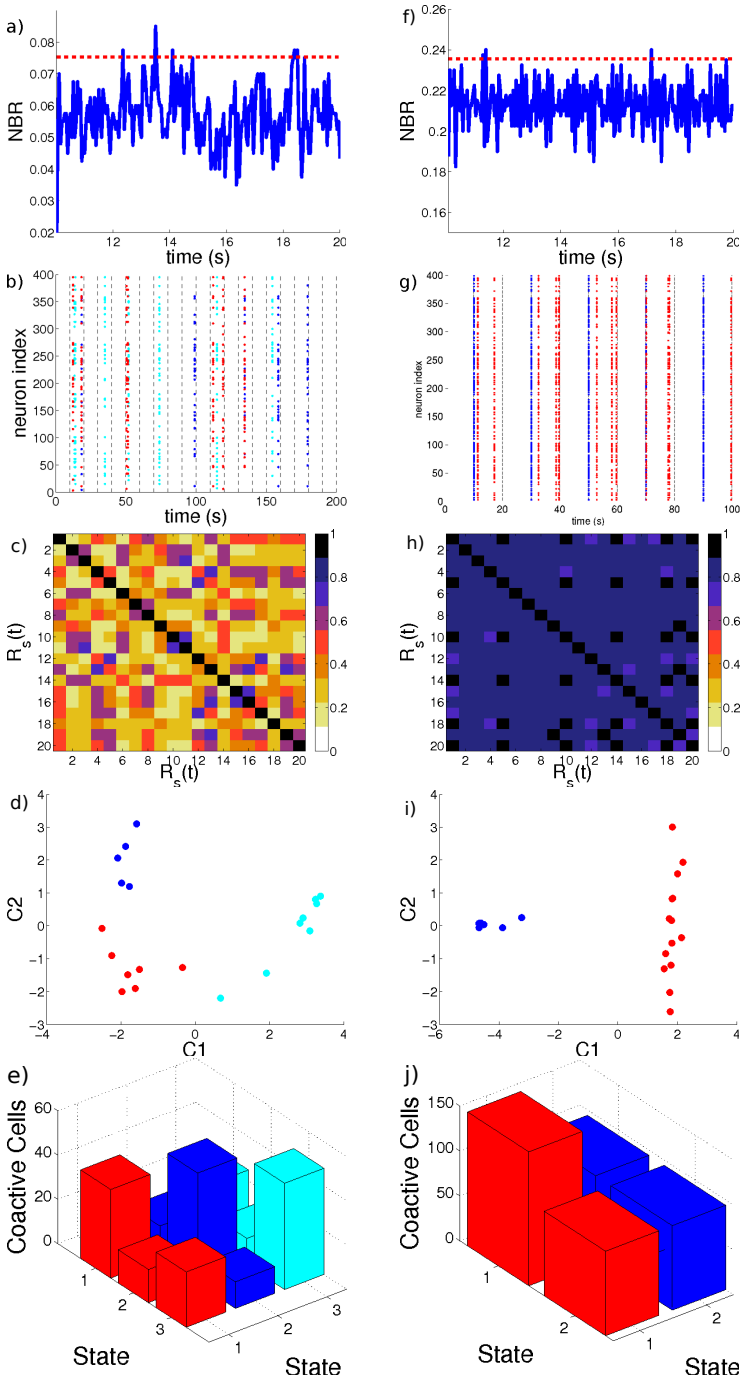

FIG. 8. **Response of the network to an increase in the excitability.** a,f) Network Bursting Rate, and the threshold defined for considering a synchronized event. b,g) Neurons involved in the synchronized events, vertical lines denoted the switching times between the excited $I^{(e)}$ and control $I^{(c)}$ inputs. Synchronized event only occur when the stimuli $I^{(e)}$ is presented to the system. Colors in the raster indicates the group assigned to the synchronous event using an optimal community partition algorithm. c,h) Synchronized Event Transition Matrix, calculated with a window $T_W = 50$ ms and number of events $S_s = 20$. d,i) Projection of the synchronized events in the 2D principal component space. e,j) Number of coactive cells in each state. The diagonal elements of the bar graph represents the total number of neurons contributing to one state. Panels (a-e) correspond to $g = 8$, while panels (f-j) depict the case $g = 1$. In both cases the system is recorded during the time span required to identify $S_s = 20$. Remaining parameters as in Fig. 1.

III. DISCUSSION

In summary, we have shown that the increase in the lateral inhibition is fundamental to drive the network dynamics from a situation where a few neurons, tonically firing at a high rate, depress a large part of the network, to a situation where all neurons are active and fire with similar slow rates. In particular, for low inhibition the *winner takes all* mechanism is at work and the activity of the network is mainly mean-driven. On the other hand for sufficiently large inhibitory strength almost all the neurons are on average sub-threshold and the dynamical activity is fluctuation-driven. These two regimes reflect in the distribution of the coefficient of variation which passes from a peaked distribution centered around one, to a wider distribution centered around two, as reported in experimental measurements of MSNs [17].

In order to observe structured bursting activity, another fundamental ingredient is the ratio between the single neuron firing rate and the duration of the post-synaptic potentials, bursting similar to MSN activity can be observed only for neurons with firing periods of the order (or smaller) than the post-synaptic decay times.

Therefore we can affirm that the structured activity observed for Medium Spiny Neurons (MSNs) is the consequence of the lateral inhibition in sparsely-connected networks with slow synapses, the mechanism at work is akin to the *Winerless competition* reported to explain the functioning of olfactory networks in order to discriminate different odors [21]. Winnerless competition refers to a dynamical mechanism, initially revealed in asymmetrically coupled inhibitory rate models [38], displaying a transient slow switching evolution along a series of metastable saddles (for a recente review on the subject see [39]). In our case, the sequence of metastable states can be represented by the firing activity of the cells assembly, switching over time. In particular, as in our analysis, slow synapses have been recognized as a fundamental ingredient, besides asymmetric inhibitory connections, to observe the emergence of winnerless competition mechanism in realistic neuronal models [18, 28].

We have introduced a new effective metric capable to encompass in a single indicator key aspects of the MSNs firing activity, with the help of this metric we have identified the values of the parameters entering in the model to optimally reproduce MSN dynamics. Furthermore, for the same choice of parameters the network is able to respond in a reproducible manner to the same stimulus, presented at different times, while presenting complex computational capability by responding to a quenched stimulus with an evolution in a high dimensional space [31].

Finally, we have been able to reproduce general experimental features reported in [6]. In particular, we have observed, as in the experiment, a structured activity alternating among a small number of distinct cell assemblies. Furthermore, we have reproduced the dynamical effects induced by decreasing the inhibitory coupling: the drastic reduction of the inhibition lead essentially to the emergence of an unique highly correlated neuronal assembly. Thus confirming the essential role of inhibition in the MSN dynamics.

In a recent work Ponzi and Wickens [37] have noticed that in their model the striatally relevant regimes correspond to marginally stable dynamical evolution. In the Supporting Information Text S1 we devote a sub-section *Linear stability analysis* to the investigation of this specific point, our conclusion is that for our model the striatally relevant regimes are definitely associated to chaotic regimes, but located in proximity of a transition to linearly stable dynamics. However for inhibitory networks it is known that even linearly stable networks can display erratic dynamics (resembling chaos) due to finite amplitude perturbations [1, 16, 24, 50]. This suggests that the usual linear stability analysis, corresponding to the estimation of the maximal Lyapunov exponent [3], is unable to distinguish between regular and irregular evolution, at least for the studied inhibitory networks [1].

IV. METHODS

The model

We considered a network of N LIF inhibitory neurons coupled via α pulses, which can be represented via the following set of $3N$ equations.

$$\dot{v}_i(t) = a_i - v(t) - gE_i(t) \tag{2a}$$

$$\dot{E}_i(t) = P_i(t) - \alpha E_i(t) \qquad i = 1, \ldots, N \tag{2b}$$

$$\dot{P}_i(t) = -\alpha P_i(t) + \frac{\alpha^2}{K} \sum_{n|t_n<t} \mathcal{C}_{i,j}\delta(t - t_n)\,. \tag{2c}$$

In this model v_i is the membrane potential of neuron i, K denotes the number of pre-synaptic connections, $g > 0$ is the strength of the synapses, the variable E_i accounts for the past history of all the previous recurrent post-synaptic potentials (PSP) that arrived to the neuron i at times t_n, and P_i is an auxiliary variable. a_i is the external excitation

and α is inversely proportional to the decaying time of the PSP. The inhibition is introduced via the minus sign in front of the synaptic strength in Eq. (2a). The matrix $\mathcal{C}_{i,j}$ is the connectivity matrix where the entry i,j is equal to 1 if there exists a synaptic connection from neuron j to neuron i. When the membrane potential of the q-th neuron arrives to the threshold $v_{th} = 1$, it is reset to the value $v_r = 0$ and the cell emits a spike which is instantaneously transmitted to all its post-synaptic neurons.

The equations (2a) to (2c) can be exactly integrated from the time $t = t_n$, just after the deliver of the n-th pulse, to time $t = t_{n+1}$ corresponding to the emission of the $(n+1)$-th spike, thus obtaining an *event driven map* [29, 51] which reads as

$$E_i(n+1) = E_i(n)\mathrm{e}^{-\alpha\tau(n)} + P_i(n)\tau(n)\mathrm{e}^{-\alpha\tau(n)} \tag{3a}$$

$$P_i(n+1) = P_i(n)\mathrm{e}^{-\alpha\tau(n)} + \mathcal{C}_{iq}\frac{\alpha^2}{K} \tag{3b}$$

$$v_i(n+1) = v_i(n)\mathrm{e}^{-\tau(n)} + a(1-\mathrm{e}^{-\tau(n)}) - gH_i(n)\,, \tag{3c}$$

where $\tau(n) = t_{n+1} - t_n$ is the inter-spike interval associated with two successive neuronal firing in the network, which can be determined by solving the transcendental equation

$$\tau(n) = \ln\left[\frac{a - v_q(n)}{a - gH_q(n) - 1}\right]\,, \tag{4}$$

here q identifies the neuron which will fire at time t_{n+1} by reaching the threshold value $v_q(n+1) = 1$.

The explicit expression for $H_i(n)$ appearing in equations (3c) and (4) is

$$H_i(n) = \frac{\mathrm{e}^{-\tau(n)} - \mathrm{e}^{-\alpha\tau(n)}}{\alpha - 1}\left(E_i(n) + \frac{P_i(n)}{\alpha - 1}\right) - \frac{\tau(n)\mathrm{e}^{-\alpha\tau(n)}}{\alpha - 1}P_i(n)\,. \tag{5}$$

The model is now rewritten as a discrete-time map with $3N - 1$ degrees of freedom, since one degree of freedom $v_q(n+1) = 1$, is lost due to the event driven procedure, which corresponds to perform a Poincaré section at any time a neuron fires.

The model so far introduced contains only adimensional units, however, the evolution equation for the membrane potential (2) can be easily re-expressed in terms of dimensional variables as follows

$$\tau_m\dot{V}_i(\tilde{t}) = I_i - V_j(\tilde{t}) - \tau_m G\tilde{E}_i(\tilde{t}) \quad i = 1, \cdots, N \quad ; \tag{6}$$

where we have chosen $\tau_m = 10$ ms as the membrane time constant in agreement with the values reported in literature for MSNs in the up state in mice [17, 32, 34], I_i represents the neural excitability and the external stimulations, which takes in account the cortico-thalamic inputs received by the striatal network. Furthermore, $\tilde{t} = t \cdot \tau_m$, the field $\tilde{E}_i = E_i/\tau_m$ has the dimensionality of a frequency and G of a voltage. The currents $\{I_i\}$ have also the dimensionality of a voltage, since they include the membrane resistance.

For the other parameters/variables the transformation to physical units is simply given by

$$V_i = V_r + (V_{th} - V_r)V_i \tag{7}$$

$$I_i = V_r + (V_{th} - V_r)a_i \tag{8}$$

$$G = (V_{th} - V_r)g \tag{9}$$

where $V_r = -60$ mV, $V_{th} = -50$ mV. The isolated i-th LIF neuron is supra-threshold whenever $I_i > V_{th}$, however due to the inhibitory coupling the effective input is $W_i = I_i - \tau_m G\tilde{E}_i$. Therefore, the neuron is able to deliver a spike if $\overline{W}_i > V_{th}$, in this case the firing of the neuron can be considered as mean-driven. However, even if $\overline{W}_i < V_{th}$, the neuron can be lead to fire from fluctuations in the effective input and the firing is in this case fluctuation-driven. It is clear that the fluctuations $\sigma(W_i)$ are directly proportional to the strength of the inhibitory coupling for constant external currents I_i.

For what concerns the PSPs the associated time constant is $\tau_\alpha = \tau_m/\alpha$, and the peak amplitude is given by

$$A_{PSP} = \frac{\alpha}{K}Ge^{-1} = g \times 92 \ \mu V \quad ; \tag{10}$$

where the last equality allows for a direct transformation from adimensional units to dimensional ones, for the connectivity considered in this paper, namely $K = 20$, and for $\alpha = 0.5$, which is the value mainly employed in our

analysis. The experimentally measured peak amplitudes of the inhibitory PSPs for spiny projection neurons ranges from $\simeq 0.16 - 0.32$ mV [47] to $\simeq 1 - 2$ mV [33]. These values depend strongly on the measurements conditions, a renormalization of all the reported measurements nearby the firing threshold gives for the PSP peak $\simeq 0.17 - 0.34$ mV [46]. Therefore from Eq. (10) one can see that realistic values for A_{PSP} can be obtained in the range $g \in [2:10]$. For $\alpha = 0.5$ one gets $\tau_\alpha = 20$ ms, which is consistent with the PSPs duration and decay values reported in literature for inhibitory transmission among spiny neurons [19, 47]

Our model does not take in account the depolarizing effects of inhibitory PSPs for $V \leq E_{cl}$ [33]. The GABA neurotrasmitter acts as depolarizing neurotrasmitter in mature projection spiny neurons, however this depolarization does not lead to a direct excitation of the spiny neurons. Therefore our model can be considered as an effective model encompassing only the polarizing effects of the PSPs for $V > E_{cl}$, while the depolarizing effects of the lateral inhibition are somehow included in the excitatory current terms I_i. This is the reason why we have assumed that the membrane potential varies in the range $[-60:-50]$ mV, since $E_{cl} \simeq -60$ mV and the threshold is $\simeq -50$ mV [33].

In the paper we have always employed dimensional variables (for simplicity we neglect the tilde on the time variable), apart for the amplitude of the synaptic coupling, where we have found more convenient to use the adimensional quantity g.

Characterization of the firing activity

We define active neurons, as opposite to silent neurons, as cells that deliver a number of spikes larger than a certain threshold S_Θ during the considered numerical experiments. In particular, we show in Fig. S1 of the Supporting Information that the value of the chosen threshold does not affect the reported results for $0 \leq S_\Theta \leq 100$.

Furthermore, the characterization of the dynamics of the active neurons is performed via the coefficient of variation CV, the local coefficient of variation CV_2 and the zero lag cross-correlation matrix of the firing rates $C(\nu_i, \nu_j)$. The coefficient of variation associated to the i-th neuron is then defined as the ratio:

$$CV^{(i)} = \frac{\sigma(ISI^{(i)})}{\overline{ISI^{(i)}}} \; ;$$

where $\sigma(A)$ and $\overline{A}$ denotes the standard deviation and mean value of the quantity A. The distribution of the coefficient of variation $P(CV)$ reported in the article refer to the values of the CV associated to all the active cells in the network.

Another useful measure of the spike statistics is the local coefficient of variation. For each neuron i and each spike emitted at time $t_n^{(i)}$ from the considered cell the local coefficient of variation is measured as

$$CV_2^{(i)}(n) = \frac{|ISI_n^{(i)} - ISI_{n-1}^{(i)}|}{ISI_n^{(i)} + ISI_{n-1}^{(i)}}$$

where the n-th inter-spike interval is defined as $ISI_n^{(i)} = t_n^{(i)} - t_{n-1}^{(i)}$. The above quantity clearly ranges between zero and one: a zero value corresponds to a perfectly periodic firing, while the value one is attained in the limit $ISI_n^{(i)}/ISI_{n-1}^{(i)} \to 0$ (or $ISI_n^{(i)}/ISI_{n-1}^{(i)} \to \infty$). The probability distribution function $P(CV_2)$ is then computed by employing the values of the $CV_2^{(i)}(n)$ for all the active cells of the network estimated at each firing event, examples of $P(CV_2)$ are shown in Fig. 3(c).

The level of correlated activity between firing rates is measured via the cross-correlation matrix $C(\nu_i, \nu_j)$. The firing rate $\nu_i(t)$ of each neuron i is calculated at regular intervals $\Delta T = 50$ ms by counting the number of spikes emitted in a time window of $10\Delta T$, starting from the considered time t. For each couple of neuron i and j the corresponding element of the $N \times N$ symmetric cross-correlation matrix $C(i,j)$ is simply the Pearson correlation coefficient measured as follows

$$C(i,j) = \frac{cov(\nu_i, \nu_j)}{\sigma(\nu_i)\sigma(\nu_j)} \; ,$$

where $cov(\nu_i, \nu_j)$ is the covariance between signals $\nu_i(t)$ and $\nu_j(t)$.

State Transition Matrix (STM) and measure of dissimilarity

The STM is constructed by calculating the firing rates $\nu_i(t)$ of the N neurons at regular time intervals $\Delta T = 50$ ms. At each time t the rates are measured by counting the the number of spikes emitted in a window $2\Delta T$, starting

at the considered time. Notice that the time resolution here used is higher with respect to that employed for the cross-correlation matrix, since we are interested in the response of the network to a stimulus presentation evaluated on a limited time window. The firing rates can be represented as a state vector $R(t) = \{\nu_i(t)\}$ with $i = 1, \ldots, N$. For an experiment of duration T_E, we have $S = \lfloor T_E/\Delta T \rfloor$ state vectors $R(t)$ representing the network evolution ($\lfloor \cdot \rfloor$ denotes the integer part). In order to measure the similarity of two states at different times $t_m = m \times \Delta T$ and $t_n = n \times \Delta T$, we have calculated the normalized scalar product

$$D(m,n) = \frac{R(t_m) \cdot R(t_n)}{|R(t_m)||R(t_n)|} \tag{11}$$

for all possible pairs $m, n = 1, \ldots, S$ during the time experiment T_E. This gives rise to a $S \times S$ matrix called the state transition matrix [43].

In the case of the two input experiment reported in the results section, the obtained STM has a periodic structure of period T_{sw} with high correlated blocks followed by low correlated ones (see Figs. S5(b) and (e) for the complete STM). In Fig 5 b) is presented a coarse grained version of the entire STM obtained by performing an average of the complete STM over sub-matrices $4T_{sw} \times 4T_{sw}$. This has been obtained as follows:

$$\overline{D(m,n)} = \frac{1}{(\lfloor S/4 \rfloor)^2} \sum_{i,j=1}^{\lfloor S/4 \rfloor} D(4i+m, 4j+n) \tag{12a}$$

$$\forall\, m, n \leq \lfloor T_{sw}/\Delta T \rfloor \quad .$$

Similarly to what done above we can define a dissimilarity metric to distinguish between the response of the network to two different inputs. We define a control input $I^c = \{I_i^c\} \in [V_{th}, V_{th} + \Delta V]$, and we register the network state vectors $R^c(t)$ at S regular time intervals for a time span T_E. Now we repeat the numerical experiment with a new input I^f, where a fraction f of the initial control stimuli $\{I_i^c\}$ is randomly modified by reassigning their values from the same distribution employed for the control input. For the modified input we register another sequence $R^f(t)$ of state vectors on the same time interval, with the same resolution. The istantaneous dissimilarity $d^f(t)$ between the response to the control and perturbed stimuli is defined as:

$$d^f(t_m) = 1 - \frac{R^c(t_m) \cdot R^f(t_m)}{|R^c(t_m)||R^f(t_m)|} \tag{13}$$

its average in time is simply given by $\bar{d}^f = \frac{1}{S} \sum_{i=1}^{S} d^f(t_i)$.

Distinguishability metric ΔM_d:

Following [37] a metric of the ability of the network to distinguish between two different inputs ΔM_d^0 can be defined in terms of the STM. In particular, let us consider the STM obtained for two inputs $I^{(1)}$ to $I^{(2)}$, each presented for a time lag T_{sw}. In order to define ΔM_d^0 the authors in [37] have considered the correlations of the state vector R taken at a generic time t_{m^*} with all the other configurations, this amounts to examine the elements $D(m^*, n)$ of the STM $\forall n$. By defining M_1 (M_2) the average of $D(m^*, n)$ taken in correspondence of the input $I^{(1)}$ ($I^{(2)}$), the distinguishablity metric between the two inputs can be finally defined as

$$\Delta M_d^0 = |M_1 - M_2| \,. \tag{14}$$

In order to take in account the single neuron variability and the number of active neurons involved in the network dynamics we have modified ΔM_d^0 by multiplying this quantity by the fraction of active neurons and the average coefficient of variation, as follows

$$\Delta M_d = \Delta M_d^0 n^* \langle CV \rangle_N \,. \tag{15}$$

The above metric is reported in Figs. 2(c),(d) and Fig. 3 (a).

Synchronized Event Transition Matrix (SETM)

In order to obtain a Synchronized Event Transition Matrix (SETM), we first coarse grain the raster plot of the network. This is done by considering a series of windows of duration $T_W = 50$ ms covering the entire raster plot.

A bursting event is identified whenever a neuron i fires 3 or more spikes within the considered window. To signal the burst occurrence a dot is drawn at the beginning of the window. From this coarse grained raster plot we obtain the Network Bursting Rate (NBR) by measuring the fraction of neurons that are bursting within the considered window. When this fraction is larger or equal to the average NBR plus two standard deviation, a synchronized event is identified. Each synchronized event is encoded in the the synchronous event vector $W_s(t)$, a N dimensional binary vector where the i-th entry is 1 if the i-th neuron participated in the synchronized event and zero otherwise. To measure the similarity between two synchronous events, we make use of the normalized scalar product between all the pairs of vectors W_s obtained at the different times t_i and t_j in which a synchronized event occurred. This represents the element i, j of the SETM.

Principal Components Analysis (PCA):

In subsection *Discriminative and computational capability*, a Principal Component Analysis (PCA) is performed by collecting S state vectors $R(t)$, measured at regular intervals ΔT for a time interval T_E, then by estimating the covariance matrix $cov(\nu_i, \nu_j)$ associated to these state vectors. Similarly, in subsection *Response of the network to an increased level of excitability* the PCA is computed collecting the S_s synchronous event vectors W_s, and the covariance matrix calculated from this set of vectors.

The principal components are the eigenvectors of theses matrices, ordered from the largest to the smallest eigenvalue. Each eigenvalue represents the variance of the original data along the corresponding eigendirection. A reconstruction of the original data set can be obtained by projecting the state vectors along a limited number of principal eigenvectors, obviously by employing the first eigenvectors will allow to have a more faithful reconstruction.

Clustering algorithms.

The *k-means* algorithm is a widespread mining technique in which N data points of dimension M are organized in clusters as follows. As a first step a number k of clusters is defined a-priori, then from a sub-set of the data k samples are chosen randomly. From each sub-set a centroid is defined in the M-dimensional space. at a second step, the remaining data are assigned to the closest centroid according to a distance measure. After the process is completed, a new set of k centroids can be defined by employing the data assigned to each cluster. The procedure is repeated until the position of the centroids converge to their asymptotic value.

An unbiased way to define a partition of the data can be obtained by finding the optimal cluster division [25]. The optimal number of clusters can be found by maximizing the following cost function, termed *modularity*:

$$\mathcal{M} = \frac{1}{A_{tot}} \sum_{ij} \left(A_{ij} - \mathcal{N}_{ij}\right) \delta(c_i, c_j), \tag{16}$$

where, $A \equiv \{A_{i,j}\}$ is the matrix to be clusterized, the normalization factor is $A_{tot} = \sum_{ij} A_{ij}$; $\mathcal{N}_{ij}$ accounts for the matrix element associated to the *null model*; c_i denotes the cluster to which the i-th element of the matrix belongs to, and $\delta(i, j)$ is the Kronecker delta. In other terms, the sum appearing in Eq. (16) is restricted to elements belonging to the same cluster. In our specific study, A is the *similarity matrix* corresponding to the SETM previously introduced. Furthermore, the elements of the matrix $\mathcal{N}$ are given by $\mathcal{N}_{ij} = \eta_i \eta_j / A_{tot}$, where $\eta_i = \sum_j A_{ij}$, these correspond to the expected value of the similarity for two randomly chosen elements [12, 27]. If two elements are similar than expected by chance, this implies that $A_{ij} > \mathcal{N}_{ij}$, and more similar they are larger is their contribution to the modularity $\mathcal{M}$. Hence they are likely to belong to the same cluster. The problem of modularity optimization is NP-hard [9], nevertheless some heuristic algorithms have been developed for finding local solutions based on greedy algorithms [5, 20, 40, 42]. In particular, we make use of the algorithm introduced for connectivity matrices in [7, 27], which can be straightforwardly extended to similarity matrices by considering the similarity between two elements, as the weight of the link between them [26]. The optimal partition technique is used in subsection *Response of the network to an increased level of excitability*, where it is applied to the similarity matrix $\mathcal{S}_{ij} = 1 - \mathcal{E}_{ij}$ where the distance matrix $\mathcal{E}_{ij} = \frac{\|x_i^p - x_j^p\|_2}{\max(\mathcal{E})}$. Here x_i^p is the vector defining the i^{th} synchronized event projected in the first p principal components, which accounts for the 80% of the variance.

ACKNOWLEDGMENTS

The authors had useful interactions with Robert Schmidt at an early stage of the project. D.A.-G. and A.T. also acknowledge helpful discussions with Diego Garlaschelli, Stefano Luccioli, and Mel MachMahon. This work has been supported by the European Commission under the program "Marie Curie Network for Initial Training", through the project N. 289146, "Neural Engineering Transformative Technologies (NETT)". D. A.-G. would like also to acknowledge the partial support provided by "Departamento Adminsitrativo de Ciencia Tecnologia e Innovacion - Colciencias" through the program "Doctorados en el exterior - 2013".

[1] Angulo-Garcia, D. and Torcini, A. (2014). Stable chaos in fluctuation driven neural circuits. *Chaos, Solitons & Fractals*, 69(0):233 – 245.

[2] Beiser, D. G. and Houk, J. C. (1998). Model of cortical-basal ganglionic processing: encoding the serial order of sensory events. *Journal of Neurophysiology*, 79(6):3168–3188.

[3] Benettin, G., Galgani, L., Giorgilli, A., and Strelcyn, J.-M. (1980). Lyapunov characteristic exponents for smooth dynamical systems and for hamiltonian systems; a method for computing all of them. part 1: Theory. *Meccanica*, 15(1):9–20.

[4] Berke, J. D., Breck, J. T., and Eichenbaum, H. (2009). Striatal versus hippocampal representations during win-stay maze performance. *Journal of neurophysiology*, 101(3):1575–1587.

[5] Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008.

[6] Carrillo-Reid, L., Tecuapetla, F., Tapia, D., Hernández-Cruz, A., Galarraga, E., Drucker-Colin, R., and Bargas, J. (2008). Encoding network states by striatal cell assemblies. *Journal of neurophysiology*, 99(3):1435–1450.

[7] Clauset, A., Newman, M. E., and Moore, C. (2004). Finding community structure in very large networks. *Physical review E*, 70(6):066111.

[8] Daw, N. D. and Doya, K. (2006). The computational neurobiology of learning and reward. *Current opinion in neurobiology*, 16(2):199–204.

[9] Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, 486(3):75–174.

[10] Grillner, S., Hellgren, J., Menard, A., Saitoh, K., and Wikström, M. A. (2005). Mechanisms for selection of basic motor programs–roles for the striatum and pallidum. *Trends in neurosciences*, 28(7):364–370.

[11] Groves, P. M. (1983). A theory of the functional organization of the neostriatum and the neostriatal control of voluntary movement. *Brain Research Reviews*, 5(2):109–132.

[12] Humphries, M. D. (2011). Spike-train communities: finding groups of similar spike trains. *The Journal of Neuroscience*, 31(6):2321–2336.

[13] Humphries, M. D., Wood, R., and Gurney, K. (2009). Dopamine-modulated dynamic cell assemblies generated by the gabaergic striatal microcircuit. *Neural Networks*, 22(8):1174–1188.

[14] Izhikevich, E. M. and Moehlis, J. (2008). Dynamical systems in neuroscience: The geometry of excitability and bursting. *SIAM review*, 50(2):397.

[15] Jaeger, D., Kita, H., and Wilson, C. J. (1994). Surround inhibition among projection neurons is weak or nonexistent in the rat neostriatum. *Journal of neurophysiology*, 72(5):2555–2558.

[16] Jahnke, S., Memmesheimer, R.-M., and Timme, M. (2009). How chaotic is the balanced state? *Front. Comp. Neurosci.*, 3(13).

[17] Klapstein, G. J., Fisher, R. S., Zanjani, H., Cepeda, C., Jokel, E. S., Chesselet, M.-F., and Levine, M. S. (2001). Electrophysiological and morphological changes in striatal spiny neurons in r6/2 huntington's disease transgenic mice. *Journal of neurophysiology*, 86(6):2667–2677.

[18] Komarov, M., Osipov, G., and Zhou, C. (2013). Heteroclinic contours in oscillatory ensembles. *Physical Review E*, 87(2):022909.

[19] Koos, T., Tepper, J. M., and Wilson, C. J. (2004). Comparison of ipscs evoked by spiny and fast-spiking neurons in the neostriatum. *The Journal of neuroscience*, 24(36):7916–7922.

[20] Lancichinetti, A. and Fortunato, S. (2009). Community detection algorithms: a comparative analysis. *Physical review E*, 80(5):056117.

[21] Laurent, G. (2002). Olfactory network dynamics and the coding of multidimensional signals. *Nature Reviews Neuroscience*, 3(11):884–895.

[22] Merchant, H., Harrington, D. L., and Meck, W. H. (2013). Neural basis of the perception and estimation of time. *Annual review of neuroscience*, 36:313–336.

[23] Miller, B. R., Walker, A. G., Shah, A. S., Barton, S. J., and Rebec, G. V. (2008). Dysregulated information processing by medium spiny neurons in striatum of freely behaving mouse models of huntington's disease. *Journal of neurophysiology*, 100(4):2205–2216.

[24] Monteforte, M. and Wolf, F. (2012). Dynamic flux tubes form reservoirs of stability in neuronal circuits. *Phys. Rev. X*, 2:041007.

[25] Newman, M. (2010). *Networks: an introduction*. Oxford University Press.

[26] Newman, M. E. (2004a). Analysis of weighted networks. *Physical Review E*, 70(5):056131.
[27] Newman, M. E. (2004b). Fast algorithm for detecting community structure in networks. *Physical review E*, 69(6):066133.
[28] Nowotny, T. and Rabinovich, M. I. (2007). Dynamical origin of independent spiking and bursting activity in neural microcircuits. *Physical review letters*, 98(12):128106.
[29] Olmi, S., Livi, R., Politi, A., and Torcini, A. (2010). Collective oscillations in disordered neural networks. *Phys. Rev. E*, 81(4 Pt 2).
[30] Oorschot, D. E. (1996). Total number of neurons in the neostriatal, pallidal, subthalamic, and substantia nigral nuclei of the rat basal ganglia: a stereological study using the cavalieri and optical disector methods. *Journal of Comparative Neurology*, 366(4):580–599.
[31] Ostojic, S. (2014). Two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons. *Nature neuroscience*, 17(4):594–600.
[32] Planert, H., Berger, T. K., and Silberberg, G. (2013). Membrane properties of striatal direct and indirect pathway neurons in mouse and rat slices and their modulation by dopamine. *PloS one*, 8(3):e57054.
[33] Plenz, D. (2003). When inhibition goes incognito: feedback interaction between spiny projection neurons in striatal function. *Trends in neurosciences*, 26(8):436–443.
[34] Plenz, D. and Kitai, S. T. (1998). Up and down states in striatal medium spiny neurons simultaneously recorded with spontaneous activity in fast-spiking interneurons studied in cortex–striatum–substantia nigra organotypic cultures. *The Journal of Neuroscience*, 18(1):266–283.
[35] Ponzi, A. and Wickens, J. (2010). Sequentially switching cell assemblies in random inhibitory networks of spiking neurons in the striatum. *The Journal of Neuroscience*, 30(17):5894–5911.
[36] Ponzi, A. and Wickens, J. (2012). Input dependent cell assembly dynamics in a model of the striatal medium spiny neuron network. *Frontiers in systems neuroscience*, 6.
[37] Ponzi, A. and Wickens, J. R. (2013). Optimal balance of the striatal medium spiny neuron network. *PLoS computational biology*, 9(4):e1002954.
[38] Rabinovich, M., Huerta, R., Volkovskii, A., Abarbanel, H., Stopfer, M., and Laurent, G. (2000). Dynamical coding of sensory information with competitive networks. *Journal of Physiology-Paris*, 94(5):465–471.
[39] Rabinovich, M. I. and Varona, P. (2011). Robust transient dynamics and brain functions. *Frontiers in computational neuroscience*, 5.
[40] Reichardt, J. and Bornholdt, S. (2006). Statistical mechanics of community detection. *Physical Review E*, 74(1):016110.
[41] Renart, A., Moreno-Bote, R., Wang, X.-J., and Parga, N. (2007). Mean-driven and fluctuation-driven persistent activity in recurrent networks. *Neural. Comput.*, 19(1):1–46.
[42] Ronhovde, P. and Nussinov, Z. (2010). Local resolution-limit-free potts model for community detection. *Physical Review E*, 81(4):046114.
[43] Schreiber, S., Fellous, J., Whitmer, D., Tiesinga, P., and Sejnowski, T. J. (2003). A new correlation-based measure of spike timing reliability. *Neurocomputing*, 52:925–931.
[44] Stephenson-Jones, M., Samuelsson, E., Ericsson, J., Robertson, B., and Grillner, S. (2011). Evolutionary conservation of the basal ganglia as a common vertebrate mechanism for action selection. *Current Biology*, 21(13):1081–1091.
[45] Taverna, S., Van Dongen, Y. C., Groenewegen, H. J., and Pennartz, C. M. (2004). Direct physiological evidence for synaptic connectivity between medium-sized spiny neurons in rat nucleus accumbens in situ. *Journal of neurophysiology*, 91(3):1111–1121.
[46] Tepper, J. M., Koós, T., and Wilson, C. J. (2004). Gabaergic microcircuits in the neostriatum. *Trends in neurosciences*, 27(11):662–669.
[47] Tunstall, M. J., Oorschot, D. E., Kean, A., and Wickens, J. R. (2002). Inhibitory interactions between spiny projection neurons in the rat striatum. *Journal of Neurophysiology*, 88(3):1263–1269.
[48] Vergara, R., Rick, C., Hernández-López, S., Laville, J., Guzman, J., Galarraga, E., Surmeier, D., and Bargas, J. (2003). Spontaneous voltage oscillations in striatal projection neurons in a rat corticostriatal slice. *The Journal of physiology*, 553(1):169–182.
[49] West, M. J., Østergaard, K., Andreassen, O. A., and Finsen, B. (1996). Estimation of the number of somatostatin neurons in the striatum: an in situ hybridization study using the optical fractionator method. *Journal of Comparative Neurology*, 370(1):11–22.
[50] Zillmer, R., Livi, R., Politi, A., and Torcini, A. (2006). Desynchronization in diluted neural networks. *Phys. Rev. E*, 74(3):036203.
[51] Zillmer, R., Livi, R., Politi, A., and Torcini, A. (2007). Stability of the splay state in pulse-coupled networks. *Phys. Rev. E*, 76:046102.

Supporting Information:

Text S1: Silent Neurons, Mechanisms for the resurgence of silent neurons, Linear stability analysis, State Transition Matrices for different regimes.

Fig. S1: Dependence of the value $n*$ on the chosen threshold S_Θ Fraction of silent neurons n^* vs the coupling strength, for several threshold definitions. A neuron is considered silent whenever it does not spike at least

S_Θ-times during the observation time. Panel a) for $\Delta V = 1$ mV and b) for $\Delta V = 5$ mV. The system is left to evolve during 10^7 spikes, after discarding 10^5 spike events of transient. Other parameters used in the simulation: $K = 20$, $N = 400$ and $\tau_\alpha = 20$ms.

Fig. S2 Neuronal statistics. Neuronal distributions of the average $\overline{ISI}$ (a,d), of the coefficient of variation CV (b,e) and of the average effective synaptic input $\overline{W}_i$ (c,f). The data in the first row are for $\Delta V = 1$ mV and in the second one for $\Delta V = 5$ mV. For $\Delta V = 1$ mV ($\Delta V = 5$ mV) black solid line with filled circles correspond to $g = 1$ ($g = 4$) and red dashed lines with open squares to $g = 4$ ($g = 10$). The system is left to evolve during 10^7 spikes, after discarding 10^5 spike events. Other parameters used in the simulation: $K = 20$, $N = 400$ and $\tau_\alpha = 20$ms.

Fig. S3: Effective input currents and fluctuations. Effective input currents $\langle\overline{W}\rangle$ (red filled circles) and the corresponding standard deviation $\langle\sigma(W)\rangle$ (black filled squares) averaged over all the neurons in the network for $\Delta V = 1$ mV (a) and $\Delta V = 5$ mV (b). The effective currents are shifted by V_{th}: negative (positive) values correspond to neurons on average below (above) threshold. The dashed vertical lines indicated the position of g_{min}. The system is left to evolve for 10^7 spikes, after discarding 10^5 spike events and the units for W and $\sigma(W)$ are in mV. Other parameters used in the simulation: $K = 20$, $N = 400$ and $\tau_\alpha = 20$ms.

Fig. S4: Linear stability analysis. a) Maximal Lyapunov exponent λ at a fixed $\tau_\alpha = 20$, as a function of connectivity strength for $\Delta V = 1$ mV (continuous line, filled circles) and $\Delta V = 5$ mV (dashed line, empty squares). b) Maximal Lyapunov exponent λ as a function of the pulse duration τ_α for the parameters $\{\Delta V, g\} = \{1\ \text{mV}, 4\}$ (continuous line with filled circles) and $\{5\ \text{mV}, 8\}$ (dashed line with empty squares). In both panels, the blue filled square indicates the triad $\{\Delta V, g, \tau_\alpha\} = \{5\ \text{mV}, 8, 20\ \text{ms}\}$, and the red filled circle to $\{\Delta V, g, \tau_\alpha\} = \{1\ \text{mV}, 4, 20\ \text{ms}\}$; these values are associated to the maximum values of Q_0 obtained for excitability distributions with fixed width ΔV. The tangent space Eq. (S12) is evolved during a period corresponding to 10^6 spikes, after discarding a transient of 10^5 spikes. Other parameters used in the simulation: $K = 20$, $N = 400$.

Fig. S5: Two stimuli presentation. Upper panel, another realization of the network with the same parameters as chosen in the main text and $\tau_\alpha = 20$ms. Lower panel, network with $\tau_\alpha = 2$ms. From left to right it is depicted the raster plot colored according to the *k-means* algorithm with k=25, vertical lines indicates the change of the presented stimulus. In the middle column is reported the state transition matrix calculated over a time span of 20 seconds (the stimulation protocol is repeated 5 times). Rightmost column reports the state transition matrix averaged over blocks of duration 4 s $\times$ 4s.

Fig. S6: Pattern Separation. Dissimilarity measure in time for an observation window of length a) $T_E = 2$s and b) $T_E = 10$s, for two values of τ_α=20ms (black circles) and τ_α=2ms (red squares) at a fixed value of $f = 0.2$. It is clearly observed that $\tau_\alpha = 20$ms more effectively differentiates the similar inputs in both observation windows, as seen by the larger values of dissimilarity respect to the $\tau_\alpha = 2$ms. Other parameters used: $\Delta T = 50$ ms, $g = 8$, $N = 400$, $K = 20$.