Detecting single-cell stimulation in a large network of integrate-and-fire neurons

Davide Bernardi* and Benjamin Lindner
Bernstein Center for Computational Neuroscience Berlin, Philippstraße 13, Haus 2, 10115 Berlin, Germany and Physics Department of Humboldt University Berlin, Newtonstraße 15, 12489 Berlin, Germany

(Received 30 October 2018; published 11 March 2019)

Several experiments have shown that the stimulation of a single neuron in the cortex can influence the local network activity and even the behavior of an animal. From the theoretical point of view, it is not clear how stimulating a single cell in a cortical network can evoke a statistically significant change in the activity of a large population. Our previous study considered a random network of integrate-and-fire neurons and proposed a way of detecting the stimulation of a single neuron in the activity of a local network: a threshold detector biased toward a specific subset of neurons. Here, we revisit this model and extend it by introducing a second network acting as a readout. In the simplest scenario, the readout consists of a collection of integrate-and-fire neurons with no recurrent connections. In this case, the ability to detect the stimulus does not improve. However, a readout network with both feed-forward and local recurrent inhibition permits detection with a very small bias, if compared to the readout scheme introduced previously. The crucial role of inhibition is to reduce global input cross correlations, the main factor limiting detectability. Finally, we show that this result is robust if recurrent excitatory connections are included or if a different kind of readout bias (in the synaptic amplitudes instead of connection probability) is used.

DOI: 10.1103/PhysRevE.99.032304

How can we understand the ways in which the brain processes sensory input? One classical approach used both in experiments and theoretical investigations is to present a stimulus (e.g., a movie, a sound sequence, or a movement of a tactile sensor) and to analyze how it is represented at different stages of sensory processing. While this kind of analysis has provided tremendous insights in particular at the sensory periphery [1], it can become more difficult to interpret the results in higher brain areas. Here, neurons are strongly and recurrently connected to each other and receive huge amounts of feedback from other regions. Their complex autonomous quasistochastic dynamics can also be captured with recurrent spiking network models [2–8] but even in these models it is often hard to predict how this activity is affected by the feed-forward sensory input.

One alternative way to explore how information about sensory stimuli is encoded in networks of the brain is the approach of reverse physiology [9]. Instead of provoking a response by presenting a sensory stimulus, neurons in the sensory cortex are directly stimulated to elicit a behavioral response. Surprisingly, it was found on several occasions that single-cell stimulation can have a "macroscopic" impact on the brain state [10–12] or behavior [13–16]. These observations are in marked contrast to common textbook wisdom that only the activity of large neuronal populations can signal something meaningful. It is a challenging theoretical endeavor to understand how networks can display a spontaneous activity compatible with brain recordings and at the same time exhibit such sensitivity to the stimulation of a single neuron [17].

The behavioral response is probably the result of a long chain of events possibly involving many brain areas. A reasonable simplification of this problem focuses on an important first step that has to take place if the single-cell stimulation is to be perceivable by the animal: the stimulation should cause a statistically significant change in a macroscopic part of the network. Put differently, the (strong) single-cell signal has to be transferred into a (possibly weak) population signal. A further constraint for the model is that the network's spontaneous activity should be asynchronous and irregular, as typically observed in cortical networks when the animal is attentive and performing a task [5,18].

In our previous study [19], we explored this problem in a random network of integrate-and-fire neurons and found that single-cell stimulation can be detected in the activity of a subpopulation of the network if this readout is biased toward specific neurons. Here, we build on those results and show that a second network can be used to detect the perturbation. Remarkably, in the simplest implementation of the second network we cannot achieve any improvement in detecting the single-cell stimulus, whereas an incorporation of more biological details enables a significant enhancement of the detection performance compared to the case without an explicit readout network.

Our paper is organized as follows. In Sec. I we first present the core structure of the model and then describe its various parts in more detail. In Sec. II we show how the detectability of the single-cell stimulation in the theoretical model introduced here compares to our previous model [19]. In the final Sec. III, we discuss the relevance of our results and we mention some limitations of this study and possible future directions. Three Appendices provide more mathematical details on our analytical approach and on the statistics of the detection process.

*Corresponding author: davide.bernardi@bccn-berlin.de

I. MODEL

As in Ref. [19], we mimic the single-cell stimulation experiment in the rat's barrel cortex [14] in a large (10^5 neurons) random recurrent network of integrate-and-fire neurons. In this paper, we compare how well the stimulus can be detected in two theoretical setups (Fig. 1). Both setups have in common the network representing a portion of the barrel cortex, in which a randomly selected cell is stimulated. Although this recurrent network model does not possess features specific to a particular cortical area, for concreteness we refer to it as the "barrel cortex network" (BCN). The two setups differ in how the activity of the BCN is read out and forwarded to the detector.

In setup A [Fig. 1(a)], a subset of neurons ($\mathcal{S}^A$) is chosen at random but with a bias toward the set of neurons (indicated as $\mathcal{B}_1$) receiving direct input from the stimulated cell. The filtered activity of $\mathcal{S}^A$ is fed to a detector, whose sensitivity is parametrized by the position of a decision threshold. By varying the threshold, the receiver-operating-characteristic (ROC) curve of the detector can be obtained. In this standard way of characterizing the performance of a detector, the correct detection rate (also known as hit rate or true positive rate) is plotted versus the false positive rate. The distance from the diagonal (the chance level) for a suitably chosen value of the threshold defines the effect size $\bar{\mathcal{Y}}$, the final output of the model. Apart from some minor differences discussed in the following, this is the setup considered in Ref. [19].

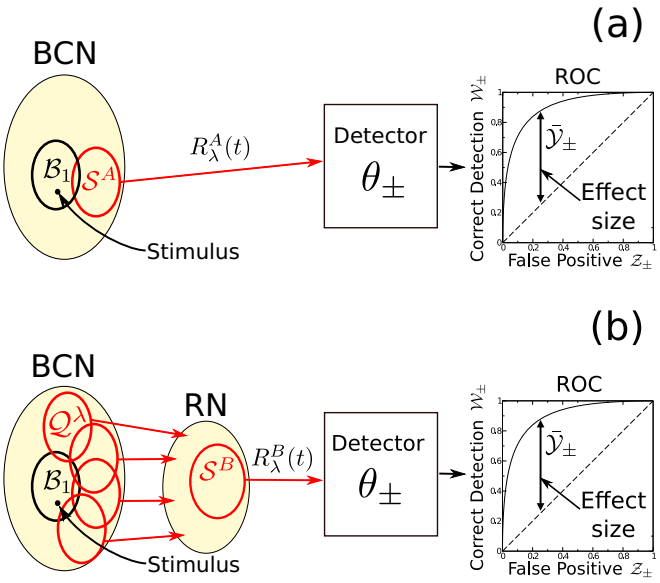

FIG. 1. Illustration of the two detection schemes compared in this paper. One neuron embedded in a large recurrent network representing the barrel cortex (BCN) is stimulated. We compare the readout scheme A [panel (a)], introduced in Ref. [19], with the readout scheme B [panel (b)]. In the readout scheme A, the activity of a subpopulation $\mathcal{S}^A$ is fed to the detector. Varying the detector sensitivity ($\theta_\pm$) yields the receiver operating characteristic (ROC) curve, which describes the performance of the detector as a function of its sensitivity. The distance of the ROC curve from the diagonal for a false positive rate of 25% defines the effect size $\bar{\mathcal{Y}}$. In the readout scheme B, the readout network (RN) receives feed-forward input from the barrel cortex network. The activity of the excitatory neurons within the RN, $\mathcal{S}^B$, is then fed to the detector. In both cases, the parameter λ quantifies the bias of the readout toward the set of neurons $\mathcal{B}_1$, the neurons one synapse away from the stimulated neuron.

In setup B [Fig. 1(b)], a readout network (RN) receiving feed-forward input from the BCN is introduced. Each neuron in the RN receives input from a different subset of the BCN. These sets of input neurons are chosen again at random with a bias toward $\mathcal{B}_1$. The activity of all excitatory neurons in the RN, $\mathcal{S}^B$, is then fed to the same detector as in setup A. For both setups, the bias is parametrized by $0 \leqslant \lambda \leqslant 1$, where $\lambda = \lambda_0$ means no bias, $\lambda = 1$ is the maximum bias toward $\mathcal{B}_1$, and $\lambda = 0$ is the maximum bias against $\mathcal{B}_1$. In the following subsections the components of the two setups are described in more detail. Furthermore, in Table I all parameters used in this paper are reported with their numerical value.

A. Barrel cortex network

The BCN is modeled as a variant of the Amit-Brunel network [20,21]. All neurons are leaky integrate-and-fire (LIF) point neurons [22]: the membrane potential of the kth neuron, v_k, evolves according to

$$\tau_m \dot{v}_k(t) = -v_k(t) + R_m[I_{\text{ext},k}(t) + I_{\text{rec},k}(t)], \tag{1}$$

where $\tau_m = 20\,\text{ms}$ is the membrane time constant, R_m is the membrane resistance, and $I_{\text{ext},k}(t)$ and $I_{\text{rec},k}(t)$ are the external background input and the recurrent input from other neurons in the network, respectively. Action potentials are mimicked by the fire-and-reset rule: whenever $v_k(t)$ crosses the firing threshold $v_T = 20\,\text{mV}$ (voltages are measured with respect to the resting potential) a spike is emitted and $v_k(t)$ is set and clamped to $v_R = 10\,\text{mV}$ for the duration of the absolute refractory period $\tau_{\text{ref}} = 2\,\text{ms}$. The times of threshold crossings of the kth neuron, $t_{i,k}$, define $x_k(t)$, the spike train emitted by neuron k:

$$x_k(t) = \sum_i \delta(t - t_{i,k}). \tag{2}$$

The BCN network consists of $N_E = 8 \times 10^4$ excitatory and $N_I = \gamma N_E = 2 \times 10^4$ inhibitory neurons, which corresponds to the size of about five cortical barrels [23]. Neurons are coupled by current-based synapses, so that the input current from the BCN to neuron k is

$$I_{\text{rec},k}(t) = \frac{\tau_m}{R_m}\left[\sum_{j\in\mathcal{P}_e(k)} J_{kj}x_j(t-D_{kj}) - g\sum_{\ell\in\mathcal{P}_i(k)} J_{k\ell}x_\ell(t-D_{k\ell})\right], \tag{3}$$

where $\mathcal{P}_e(k)$ is a random set of $C_E = 4000$ excitatory neurons, $\mathcal{P}_i(k)$ is a random set of $C_I = \gamma C_E = 1000$ inhibitory neurons, J_{kj} and $J_{k\ell}$ are independent exponentially distributed random numbers with mean $J = 0.1\,\text{mV}$, and D_{kj} and D_{kl} are transmission delays, randomly drawn in the interval 0.5 ms to 2.0 ms. Autapses (self-coupling) were excluded; i.e., $J_{ii} = 0$ for any i. By construction, the connection probability between two neurons is approximately independent of the neuron type and is $p_c \approx C/N = 0.05$, where $N = N_E + N_I = 10^5$ is the total network size and $C = C_E + C_I = 5 \times 10^3$ is the total number of inputs per neuron. Hence, the connectivity of the network is sparse.

TABLE I. List of parameters used in this paper with respective numerical values.

Symbol	Value	Description
τ_m	20 ms	membrane time constant
τ_{ref}	2 ms	refractory period
v_T	20 mV	threshold voltage
v_R	10 mV	reset voltage
$R_m I_0$	5.2 mV	constant external input ($R_m I_0 = -18$ mV for the RN in Sec. II A)
C_{ext}	700	number of excitatory external Poisson inputs per neuron
r_{ext}	12 Hz	rate of excitatory external Poisson inputs
N_E	80 000	number of excitatory neurons in the BCN
γ	0.25	ratio of inhibitory to excitatory neurons
N_I	γN_E	number of inhibitory neurons
C_E	4000	number of excitatory inputs per neuron
C_I	γC_E	number of inhibitory inputs per neuron
J	0.1 mV	average synaptic coupling strength
g	7	strength of inhibitory relative to excitatory coupling
D_{min}	0.5 mV	minimum transmission delay
D_{max}	2.0 ms	maximum transmission delay
$\hat{C}$	4000	feed-forward (from the BCN to the RN) inputs per neuron ($\hat{C} = 1000$ in Sec. II C)
N_B	10 000	number of excitatory neurons in the RN, i.e., in $\mathcal{S}^B$
$N_{\mathcal{I}}$	γN_B	number of inhibitory neurons in the RN
$C_{\mathcal{E}}$	$C_E - \hat{C}$	number of recurrent excitatory inputs per neuron in the RN
$C_{\mathcal{I}}$	C_I	number of recurrent inhibitory inputs per neuron in the RN
T_s	400 ms	stimulus duration
$R_m \Delta I_0$	23 mV	stimulus intensity
T_w	1200 ms	time window for single-cell detection
τ_f	100 ms	width of time filter for detection
N_A	$\hat{C}$	number of neurons in the readout set $\mathcal{S}^A$
T_{ic}	500 ms	initial simulation time to forget initial conditions
T	3000 ms	simulation time (data acquisition)
Δt	0.1 ms	simulation time step
N_{trials}	900	number of trials for each network simulation

The strength of inhibition relative to excitation $g = 7$ is chosen such that the spontaneous firing of the BCN is asynchronous and irregular (AI) with low mean firing rate $r_{sp} \approx 2$ Hz. This choice is motivated by the fact that the AI firing regime is generally associated with the attentive state [5] and spontaneous firing rates in the barrel cortex are typically quite low [24,25]. The external input is the sum of a constant part I_0 and of Poissonian (temporally uncorrelated) shot noise

$$I_{\text{ext},k}(t) = I_0 + \frac{\tau_m}{R_m}\left[\sum_{j=1}^{C_{\text{ext}}}\sum_{l} J_{k,j,l}\,\delta(t - t_{k,j,l})\right], \qquad (4)$$

where $t_{k,j,l}$ are independent spiking times with mean rate $r_{\text{ext}} = 12$ Hz, $C_{\text{ext}} = 700$ is the number of external inputs per neuron, and $J_{k,j,l}$ are i.i.d. samples from an exponential distribution with mean value $J = 0.1$ mV. The constant input $R_m I_0 = 5.2$ mV is chosen such that the total mean external input is slightly above the firing threshold $R_m I_0 + \tau_m C_{\text{ext}} J r_{\text{ext}} = 22$ mV. Note that the mean recurrent input is inhibitory. Therefore, the *total* mean input is below the firing threshold and neurons fire driven by input fluctuations. The mean voltage can be roughly estimated by

$$\langle v \rangle = R_m I_0 + \tau_m J[C_{\text{ext}} r_{\text{ext}} + C_E r_{\text{sp}}(1 - g\gamma)] \approx 10\,\text{mV}, \quad (5)$$

which is well below the firing threshold (the actual $\langle v \rangle$ is slightly lower because of the effect of spikes). A crude estimate of the magnitude of the voltage fluctuations can be obtained by assuming (spatially and temporally) uncorrelated Poisson input:

$$\sigma_v \approx \left(\int_{-\infty}^{+\infty} df\, \frac{\tau_m^2 J^2}{1 + (2\pi\tau_m f)^2}[C_{\text{ext}} r_{\text{ext}} + C_E r_{\text{sp}}(1 + g^2\gamma)]\right)^{\frac{1}{2}}$$
$$= \left(\frac{\tau_m J^2}{2}[C_{\text{ext}} r_{\text{ext}} + C_E r_{\text{sp}}(1 + g^2\gamma)]\right)^{\frac{1}{2}} \approx 3.5\,\text{mV}. \qquad (6)$$

The measured σ_v is actually ≈20% larger than predicted by Eq. (6) but still rather small compared to the distance from $\langle v \rangle$ to the threshold, which is consistent with irregular fluctuation-driven firing. Previous theoretical work has shown that combining total mean external input above threshold and net negative recurrent feedback is necessary to obtain AI firing at low rate for this kind of network model [4].

B. Stimulus and firing rate response

In the experiment [14], each cell was stimulated several times (15 times on average). However, the final result was averaged over many cells (51) from many animals. For simplicity, we chose here to simulate a comparable total number

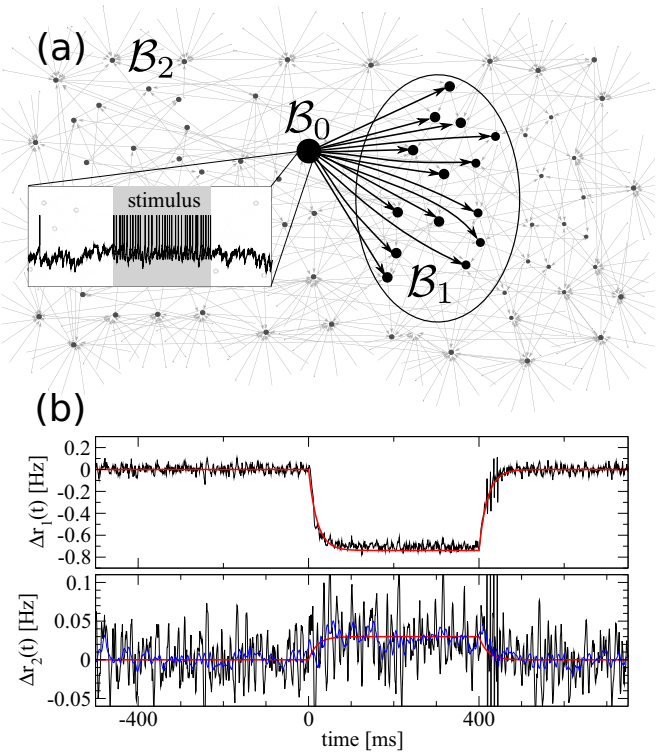

FIG. 2. (a) Illustration of the network representing the barrel cortex (BCN). $\mathcal{B}_0$ is the stimulated cell; $\mathcal{B}_1$ is the set of cells receiving direct input from $\mathcal{B}_0$; $\mathcal{B}_2$ is the set of cells receiving no direct input from $\mathcal{B}_0$. Note that the network is completely random and unstructured, and cells are grouped together only for visualization purposes. (b) Firing rate deviation of $\mathcal{B}_1$ and $\mathcal{B}_2$ in response to the stimulus, filtered with two different time steps (black curve for 1 ms, blue for 20 ms) to reduce fluctuations due to the finite number of trials (1000). The thicker red line is the theory resulting from Eqs. (9) and (11).

of trials (900) and to change the network on every trial. As a consequence, trial average implies also averaging over the network connectivity (including weights and delays), over the random initial conditions, and over realizations of the input noise. With this premise, we can now describe how we mimicked the single-cell stimulation experiment.

On each trial, the network was first run for $T_{\rm ic} = 500$ ms to forget initial conditions. Afterward, the network is simulated for a total time of $T = 3000$ ms. In the middle of the simulation time, at $t = 0$, one randomly selected cell is stimulated for $T_s = 400$ ms. The stimulus consists of a current step of intensity $R_m \Delta I_0 = 23$ mV which brings the firing rate of the stimulated cell to about 80 Hz, comparable to the evoked rate increase in the experiment [14]. The network was implemented in C++ and integrated with an Euler scheme with time step 0.1 ms.

The stimulation protocol and the notation are illustrated in Fig. 2(a). The stimulated cell is labeled as $\mathcal{B}_0$ and can be either excitatory or inhibitory. Cells receiving direct input from the stimulated neuron, i.e., neurons that are—postsynaptically—one connection away from $\mathcal{B}_0$, are labeled as $\mathcal{B}_1$. Neurons that receive no direct input from $\mathcal{B}_0$ form the set $\mathcal{B}_2$, which are almost surely exactly two links away from $\mathcal{B}_0$ (the probability of observing a neuron three synapses away is extremely small). Note that the network has no spatial structure, and neurons in Fig. 2 are grouped only for illustration convenience.

During the stimulus, the network quickly reaches a new steady state, in which the average firing rates of $\mathcal{B}_0$, $\mathcal{B}_1$, and $\mathcal{B}_2$, are different from each other and from the spontaneous value $r_{\rm sp}$. For the time-dependent firing rate of population B_ℓ we employ the common definition [22]

$$r_\ell(t) = \left\langle \frac{1}{N_\ell} \sum_{j \in \mathcal{B}_\ell} x_j(t) \right\rangle, \tag{7}$$

where $\ell = 0, 1, 2$ and N_ℓ is the size of the respective population. Angular brackets indicate here (as well as in the following) averaging over different realizations of the network connectivity, of the input noise, and of the random initial conditions.

To avoid an overladen notation, we will indicate with r_ℓ with no time argument the *steady-state* firing rate the population B_ℓ reaches during the stimulation. These rates can be approximately predicted by solving a system of equations numerically, in which each output firing rate is a function of its excitatory ($\nu_{e,\ell}$, with $\ell = 0, 1, 2$) and inhibitory ($\nu_{i,\ell}$) input firing rates. These equations describe the state of the system during the stimulation in the static picture (i.e., after all transients have passed) and for an *average* network realization. The input to the ℓth subpopulation is determined by the sum of the input firing rates from all other subpopulations multiplied by the respective average number of connections to $\mathcal{B}_\ell$. Excitatory input rates also have an additional term due to the external Poissonian drive. For the case of excitatory $\mathcal{B}_0$ the system is

$$\begin{aligned}
r_0 &= \phi_{sn}(\nu_{e,0}, \nu_{i,0}, I_0 + \Delta I_0),\\
r_1 &= \phi_{sn}(\nu_{e,1}, \nu_{i,1}, I_0),\\
r_2 &= \phi_{sn}(\nu_{e,2}, \nu_{i,2}, I_0),\\
\nu_{e,0} &= p_c C_E r_1 + (1 - p_c) C_E r_2 + C_{\rm ext} r_{\rm ext},\\
\nu_{e,1} &= r_0 + p_c (C_E - 1) r_1 + (1 - p_c)(C_E - 1) r_2 + C_{\rm ext} r_{\rm ext},\\
\nu_{e,2} &= p_c C_E r_1 + (1 - p_c) C_E r_2 + C_{\rm ext} r_{\rm ext},\\
\nu_{i,0} &= p_c \gamma C_E r_1 + (1 - p_c) \gamma C_E r_2,\\
\nu_{i,1} &= p_c \gamma C_E r_1 + (1 - p_c) \gamma C_E r_2,\\
\nu_{i,2} &= p_c \gamma C_E r_1 + (1 - p_c) \gamma C_E r_2.
\end{aligned} \tag{8}$$

In these equations we assume $(C_E - 1)/(N_E - 1) \approx C_E/(N_E - 1) \approx p_c$. Within this approximation, $\mathcal{B}_1$ contains on average $N_E p_c = C_E$ excitatory neurons. Consider the excitatory inputs to $\mathcal{B}_0$ and to $\mathcal{B}_2$ (fourth and sixth lines in the above equations): the probability of receiving input from $\mathcal{B}_0$ is zero, either because autapses are forbidden or by definition of $\mathcal{B}_2$; the probability of receiving input from $\mathcal{B}_1$ is $C_E/N_E = p_c$; the probability of receiving input from $\mathcal{B}_2$ is $1 - p_c$. Therefore, the prefactors multiplying r_0, r_1, and r_2 are zero, $p_c C_E$, and $(1 - p_c) C_E$, respectively. If we consider a neuron belonging to $\mathcal{B}_1$ (fifth line of the above set of equations), the probability of receiving input from $\mathcal{B}_0$ is unity by definition, which imposes the prefactor one to r_0 and leaves only $C_E - 1$ connections to assign to r_1 and

r_2. These remaining $C_E - 1$ connections will originate from $\mathcal{B}_1$ with probability p_c and from $\mathcal{B}_2$ with probability $1 - p_c$, which explains the factors $p_c(C_E - 1)$ and $(1 - p_c)(C_E - 1)$ multiplying r_1 and r_2, respectively. Similar considerations motivate the prefactors of the inhibitory input terms (again neglecting corrections of order $1/N$).

For inhibitory $\mathcal{B}_0$ only two equations change compared to Eq. (8):

$$\begin{aligned} &\dots, \\ &\nu_{e,1} = p_c C_E r_1 + (1 - p_c) C_E r_2 + C_{\text{ext}} r_{\text{ext}}, \\ &\dots, \\ &\nu_{i,1} = r_0 + p_c(\gamma C_E - 1) r_1 + (1 - p_c)(\gamma C_E - 1) r_2, \\ &\dots. \end{aligned} \tag{9}$$

In both sets of equations, the input-output relation is a straightforward adaptation of Eq. (10) from Ref. [26], the firing rate of a LIF neuron driven by excitatory and inhibitory shot noise:

$$\phi_{sn}(\nu_e, \nu_i) = \left[\tau_{\text{ref}} + \tau_m \int_0^{1/J} \frac{ds}{s} Z_0^{-1}(s, \nu_e, \nu_i) \times \left(\frac{e^{s\hat{v}_T}}{1 - Js} - e^{s\hat{v}_R} \right) \right]^{-1}, \tag{10}$$

where $Z_0^{-1}(s, \nu_e, \nu_i) = (1 - Js)^{\tau_m \nu_e}(1 + gJs)^{\tau_m \nu_i}$, $\hat{v}_R = v_R - R_m I_0$, and $\hat{v}_T = v_T - R_m I_0$.

Note that Eq. (10) is exact for Poissonian shot noise with exponentially drawn weights. Neurons in our network receive an external input noise that satisfies these conditions; however, inputs from recurrent connections do have a temporal correlation [8,27], and weights are drawn only once when the network is constructed, but then remain unchanged. Still, given the low firing rate and the high number of connections per neuron, both conditions are approximately satisfied also by the recurrent input. Indeed, the solution of Eq. (8) predicts the network spontaneous and evoked firing rates rather well also in the absence of external shot-noise input (see Ref. [19]). The prediction is more accurate than the more commonly used diffusion approximation.

The time-dependent firing rates of the populations $\mathcal{B}_1$ and $\mathcal{B}_2$ in response to the single-cell stimulation can be described by an exponential relaxation from the spontaneous value to the new steady-state value. If we include also the part of the time course after we have switched off the stimulus the response of $\mathcal{B}_\ell$ can be well approximated by

$$\begin{aligned} \Delta r_\ell(t) &= r_\ell(t) - r_{\text{sp}} \approx (r_\ell - r_{\text{sp}}) \Delta a(t) \\ &\approx \Delta r_\ell \left[H(t)\left(1 - e^{-\frac{t}{\tau_1}}\right) - H(t - T_s)\left(1 - e^{-\frac{t - T_s}{\tau_2}}\right) \right], \end{aligned} \tag{11}$$

where $\Delta r_\ell = r_\ell - r_{\text{sp}}$ is the steady-state deviation of the firing rate of population $\mathcal{B}_\ell$ and $H(t)$ is the Heaviside step function. In the asynchronous state and for a weak signal, a "quasistationary" approximation can be attempted (see [22], chapter 15), in which both time constants are taken equal to the membrane time constant. Indeed, setting $\tau_{1,2} = \tau_m$ provides a reasonable agreement with the rate changes observed in the simulations, as seen in Fig. 2(b). Here, the firing rate response for the case of inhibitory $\mathcal{B}_0$ is plotted. Intuitively, the firing rate response is stronger for $\mathcal{B}_1$ than $\mathcal{B}_2$. Furthermore, the two effects are of opposite sign. To understand why stimulating an inhibitory cell has an excitatory effect on $\mathcal{B}_2$, we recall that the net recurrent input is inhibitory. Therefore, the inhibiting effect of $\mathcal{B}_0$ on $\mathcal{B}_1$ reduces the amount of recurrent inhibition that $\mathcal{B}_2$ receives from $\mathcal{B}_1$, causing a weak increase in the firing rate of $\mathcal{B}_2$. In the case of excitatory $\mathcal{B}_0$, the firing rate response of $\mathcal{B}_1$ is positive, while that of $\mathcal{B}_2$ is negative (not shown; see also [19], Fig. 1).

Of course, the actual network response is more complex than Eq. (11). For instance, a transient, high-frequency oscillation is excited just after the stimulus is switched off [Fig. 2(b), just after $t = 400$ ms]. However, one must consider that this brief oscillation is rather weak in the first place, if compared to the noise level of a *single* trial [curves in Fig. 2(b) are averaged over 1000 trials]; then, it is low-pass filtered by the detector (see next section). Hence, it is unlikely that the synchronization seen in Fig. 2(b) or other high-frequency transients can play a significant role in the detection of the stimulus. Even the most prominent feature of the time-dependent rate dynamics—i.e., the exponential relaxation seen in Eq. (11)—happens on a timescale that is much smaller than the stimulus duration and has no large impact on the detectability of the stimulus. Therefore, when discussing the firing-rate response in the Results section, we will neglect any time dependence of $r_\ell(t)$ and assume that firing rates jump almost instantaneously to the new constant value r_ℓ as predicted by Eqs. (8) and (9).

C. Readout

The detector receives input from the set of neurons $\mathcal{S}^X$. We recall that $\lambda \in [0, 1]$ quantifies the bias when selecting input neurons from $\mathcal{B}_1$ (the subpopulation that receives input directly from the stimulated cell $\mathcal{B}_0$). The value of λ_0 (no bias) is the size of $\mathcal{B}_1$ divided by the total size of the BCN.

In the first setup [Fig. 1(a)], $X = A$, and $\mathcal{S}^A$ is a subset of $\hat{C}$ excitatory neurons excluding the stimulated cell. In line with the meaning of λ explained above, $\mathcal{S}^A$ is constructed by randomly selecting $\lambda\hat{C}$ neurons from $\mathcal{B}_1$ and $(1 - \lambda)\hat{C}$ from $\mathcal{B}_2$. In the following we use $\hat{C} = 4000$ except for Sec. II C, where $\hat{C} = 1000$.

In the second setup [Fig. 1(b)], $\mathcal{S}^B$ is the set of all excitatory neurons of the RN. The size of $\mathcal{S}^B$ is $N_B = 1 \times 10^4$, which is the order of magnitude of a cortical column [23]. Each neuron in $\mathcal{S}^B$ receives feedforward input from $\hat{C}$ neurons in the BCN in addition to the local recurrent input and external shot noise. This feedforward input is purely excitatory, in line with the view that long-range projections are mainly excitatory. Hence, the kth neuron in $\mathcal{S}^B$ evolves according to

$$\tau_m \dot{v}_k(t) = -v_k(t) + R_m[I_{\text{ext},k}(t) + I_{\text{rec},k}(t) + I_{\text{FF},k}(t)], \tag{12}$$

where $I_{\text{ext},k}(t)$ are $I_{\text{rec},k}(t)$ analogous to Eqs. (3) and (4), respectively. The third input term represents the input from the BCN to the RN and reads

$$I_{\text{FF},k}(t) = \frac{\tau_m}{R_m} \sum_{j \in \mathcal{Q}_\lambda(k)} J_{kj}^{\text{FF}} x_j\left(t - D_{kj}^{\text{FF}}\right), \tag{13}$$

where the weights J_{kj}^{FF} are distributed exponentially with mean $J = 0.1$ mV and the delays D_{kj}^{FF} are uniformly distributed in

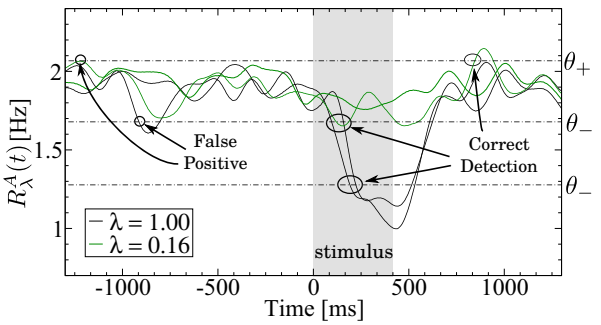

FIG. 3. Illustration of the detector. Two trials of the readout activity R^A_λ are plotted for two values of the bias parameter λ. Crossings of the threshold θ_+ or of the threshold θ_- define false positive events if they occur before the stimulus onset, i.e., for $-T_w < t < 0$, or correct detection events for $0 < t < T_w$.

the range 0.5 ms to 2.0 ms, as for the BCN. The presynaptic population of the kth neuron, $\mathcal{Q}_\lambda(k)$, consists of $\hat{C}$ excitatory neurons drawn from $\mathcal{B}_1$ with probability λ, otherwise taken from $\mathcal{B}_2$ (to be consistent with the first setup, $\mathcal{B}_0$ is excluded). Therefore, λ represents here the *average* overlap between $\mathcal{B}_1$ and $\mathcal{Q}_\lambda(k)$; the overlap for each $\mathcal{Q}_\lambda(k)$ is binomially distributed with a narrow relative standard deviation as the number of feed-forward inputs per neuron is large.

While Eq. (12) describes the general case considered in this paper, we investigate and compare various configurations for the RN, so that some details differ for each case and are therefore discussed in the respective subsection of the Results.

D. Detection and effect size

The detection is common to the two setups and works as follows. First, the input spike trains are summed up and filtered to obtain the readout activity $R^X_\lambda(t)$:

$$R^X_\lambda(t) = \frac{1}{N_X} \sum_{j \in \mathcal{S}^X} x_j(t) \star F_{\tau_f}(t), \tag{14}$$

where N_X is the number of neurons in $\mathcal{S}^X$ and $\star$ means convolution with the filter, a truncated Gaussian

$$F_{\tau_f}(t) = \frac{H(t)H(3\tau_f - t)}{\sqrt{\pi \tau_f^2/2}} \exp\left[\frac{-(t - 3\tau_f/2)^2}{\tau_f^2/2}\right]. \tag{15}$$

The filter is causal and its width is controlled by the parameter $\tau_f = 100$ ms. The chosen value for this timescale is well in the range of slow NMDA synapses [28] and reflects a typical speed for the integration of information.

As an example, two trials for the activity $R^A_\lambda(t)$ are plotted in Fig. 3 for inhibitory $\mathcal{B}_0$ and for two particular values of λ. Before the stimulus onset (occurring at $t = 0$), $R^A_\lambda(t)$ fluctuates around the network spontaneous firing rate. Afterward, it reaches transiently a new plateau $r_\lambda = r_1\lambda + r_2(1 - \lambda)$, where r_1 and r_2 can be calculated by either Eq. (8) or Eq. (9) [19].

Because the stimulated cell can be either excitatory or inhibitory, we consider in fact two distinct detectors (differently from Ref. [19]), and consider their output separately. One detector reacts to deviations in the positive direction while the other one responds to deflections in the negative direction. More precisely, we set both an upper threshold θ_+ and a lower threshold θ_- together with a detection window $T_w = 1200$ ms. The θ_+ detector will react whenever a trajectory exceeds θ_+ (we use the same symbol to indicate the detector and the corresponding threshold) while the θ_- detector whenever a trajectory falls below θ_- in the relevant time interval: If the crossing event happens before the stimulus onset, i.e., in the interval $(-T_w, 0)$, it is considered as a *false positive* and indicated with either $\mathcal{Z}^+(\theta_+)$ or $\mathcal{Z}^-(\theta_-)$ depending on the detector type. Otherwise, if it happens for $0 < t < T_w$ it is classified as a *correct detection* and indicated as $\mathcal{W}^+(\theta_+)$ or $\mathcal{W}^-(\theta_-)$. The fraction of trials in which an event is detected defines the respective detection rates as a function of the threshold. For the θ_+ detector

$$\mathcal{Z}^+_\lambda(\theta_+) = \left\langle \max_{t \in (-T_w, 0)} \left\{ H\left(R^X_\lambda(t) - \theta_+\right) \right\} \right\rangle, \tag{16}$$

where $H(t)$ is again the Heaviside function. For downward crossings

$$\mathcal{Z}^-_\lambda(\theta_-) = \left\langle \max_{t \in (-T_w, 0)} \left\{ H\left(\theta_- - R^X_\lambda(t)\right) \right\} \right\rangle, \tag{17}$$

and correct detection rates $\mathcal{W}^\pm_\lambda(\theta)$ are analogously defined but for $t \in (0, T_w)$. Plotting the correct detection rate as a function of the false positive rate upon variation of the detector sensitivity (a high θ_+ or a low θ_- corresponds to low sensitivity, i.e., low detection rates) yields the receiver-operating-characteristic (ROC) curve, the standard way of visualizing the performance of a detector (see Fig. 1). Following [14], we define the effect size as the difference between correct detection and false positive rate,

$$\mathcal{Y}^\pm_\lambda(\theta_\pm) = \mathcal{W}^\pm_\lambda(\theta_\pm) - \mathcal{Z}^\pm_\lambda(\theta_\pm), \tag{18}$$

which depends both on λ and $\theta_\pm$. In experiments, the sensitivity, i.e., the guessing or false positive rate, can be indirectly influenced by rewards during training but cannot be directly controlled or varied, being a property of each individual experimental subject. Here, we choose a value for the FP rate of 25% which roughly corresponds to the average false positive rate measured experimentally [14]. Hence, our measure for the overall performance of the detector is

$$\bar{\mathcal{Y}}^\pm(\lambda) = \mathcal{Y}^\pm_\lambda(\bar{\theta}_\pm), \tag{19}$$

with the threshold $\bar{\theta}_\pm$ obeying

$$\mathcal{Z}^\pm_\lambda(\bar{\theta}_\pm) = 0.25. \tag{20}$$

With this choice, the maximum effect size that can be achieved is $1 - \mathcal{Z}^\pm_\lambda(\bar{\theta}_\pm) = 0.75$, and the minimum is $-\mathcal{Z}^\pm_\lambda(\bar{\theta}_\pm) = -0.25$, i.e., the effect size can be negative, which means that the detector reacts less frequently to a stimulus than chance level. For each case considered here, the effect size was obtained from 900 trials and p values were calculated with Fisher's exact test to assess the statistical significance at the standard level of 5%.

We note that the detector considered in Ref. [19] was similar but showed two differences: First, instead of two detectors with one threshold, a single detector with two symmetric thresholds was considered there, which relies on a rather precise *a priori* knowledge of the spontaneous firing rate.

Second, because in the experiments the detection task was preceded by a training phase, it was assumed that the detection threshold is optimized during learning. Hence, the effect size was defined as the maximum of Eq. (18), which corresponds to the maximum distance of the ROC curve from the diagonal. The problem of this choice lies in the determination of the p value: Briefly, applying the standard significance test to the same data set used to find the optimal threshold does not give the correct p value.

One proper procedure would be to use separate data sets to select the optimal threshold and to calculate the effect size, but this is computationally expensive. Therefore, we choose here not to optimize the threshold but to select a fixed value, as explained above. In Appendix B we show that results obtained by the detector as defined here and as in Ref. [19] are rather similar. Furthermore, in Appendix C we give more details on significance tests for the detector with fixed false positive rate and with optimized threshold.

E. Effect size and signal-to-noise ratio

For a given stimulation time and under simplifying circumstances, there is a monotonic relationship between the effect size and the (signed) signal-to-noise ratio (SNR) of the readout activity, defined as

$$\delta_X(\lambda) = \frac{\Delta r_\lambda^X}{\sigma_X} = \frac{r_\lambda^X - r_{\mathrm{sp}}}{\sigma_X}, \tag{21}$$

where only r_λ^X, the steady-state deviation of the firing rate of the population $\mathcal{S}^X$, is considered as signal, and σ_X is the stationary standard deviation of the population activity

$$\sigma_X = \sqrt{\langle R_\lambda^{X2}\rangle - \langle R_\lambda^X\rangle^2}. \tag{22}$$

Here, angular brackets indicate both averaging over trials and over the time interval that can be considered stationary, that is, before the stimulus onset. The dependence of σ_X on λ will be mostly ignored in the following and not explicitly indicated. Note that our SNR (needed for the calculation of the effect size) differs from the standard definition [29,30] in that we take into account by the sign whether the signal consists of a positive or negative deviation from the spontaneous value.

As the SNR is a simpler quantity than the effect size, we will use it to interpret the qualitative behavior of the effect size for the two setups. To this end, it suffices to know that the effect size is an increasing function solely of the SNR, as long as the activity is approximately Gaussian and the signal is not too short. The relation between SNR and the effect size is further described in Appendix A.

In the following, we will make use of the fact that the variance of the population activity can be decomposed into two parts, one of which is proportional to the low-frequency limit of the spike-train power spectrum of a single neuron $S_{xx}(0)$, and the other one proportional to the low-frequency limit of the cross spectrum between spike trains of different neurons $S_{x_1x_2}(0)$:

$$\sigma_X^2 \approx \frac{S_{xx}(0)}{\sqrt{\pi}\tau_f N_X} + \frac{S_{x_1x_2}(0)}{\sqrt{\pi}\tau_f}. \tag{23}$$

Here, we exploited the fact that the size of the population N_X is large and that spectral measures depend weakly on frequencies up to $1/\tau_f$, above which the filter is essentially zero. We note that the approximation in Eq. (23) is rather precise, unlike many approximations we make in the following, but valid only in the stationary situation. More details on the definition of spectral measures and on the relation in Eq. (23) are given in Appendix A.

II. RESULTS

We now study and compare the detection performance of the setups A and B, depicted in Fig. 1. As explained above, setup B differs from setup A in that the detector does not read in input directly from the BCN, where the single-cell stimulation occurs, but from an intermediate processing stage, the RN. We consider three configurations for the RN, in order of increasing complexity: (i) a population of neurons receiving only feed-forward input and no recurrent connections, (ii) a population of neurons receiving both feed-forward excitation and local recurrent inhibition, and (iii) a fully recurrent E-I network receiving feed-forward input from the BCN.

A. Purely feed-forward readout

The first configuration for the RN we consider is a population of LIF neurons receiving feed-forward input from the BCN, additional shot noise, and no further recurrent input (Fig. 4). Accordingly, each neuron in $\mathcal{S}^B$ obeys

$$\tau_m \dot{v}_k(t) = -v_k(t) + R_m[I_{\mathrm{ext},k}(t) + I_{\mathrm{FF},k}(t)], \tag{24}$$

where the external and feed-forward input currents are given by Eqs. (4) and (13), respectively, and the usual fire-and-reset rule is applied. Neurons in $\mathcal{S}^B$ receive as much excitatory input as neurons in the BCN but no inhibition. To replace the missing recurrent inhibition and keep the average firing rate low, the constant input current [included in $I_{\mathrm{ext},k}(t)$; see Eq. (4)] is set here to $R_m I_0 = -18.0\,\mathrm{mV}$. With this choice, r_{sp}^B, the spontaneous firing rate of the RN, is very close to the spontaneous firing rate of the BCN, that is, $r_{\mathrm{sp}}^B \approx 2\,\mathrm{Hz}$. Therefore, the filtered readout activities of the two setups, $R_\lambda^A(t)$ and $R_\lambda^B(t)$, fluctuate around a similar mean value. However, the magnitude of the fluctuations for the two setups is very different: fluctuations are small for $R_\lambda^A(t)$ ($\sigma_A \approx 0.09\,\mathrm{Hz}$) and much larger for $R_\lambda^B(t)$ ($\sigma_B \approx 1.6\,\mathrm{Hz}$). The strong fluctuations in setup B are due to the amplification of the common input noise in the driven population of uncoupled cells (cf. [31,32] for a similar setup and for analytical approaches in this situation).

As discussed above, the input is not perfectly Poissonian. However, Eq. (10) can still be exploited to obtain a rough estimate of the output fluctuations. We start from the decomposition of the readout variance Eq. (23) (see also Appendix A):

$$\sigma_B^2 \approx \frac{S_{xx}^{\mathcal{E}}(0)}{\sqrt{\pi}\tau_f N_B} + \frac{S_{x_1x_2}^{\mathcal{E}\mathcal{E}}(0)}{\sqrt{\pi}\tau_f}. \tag{25}$$

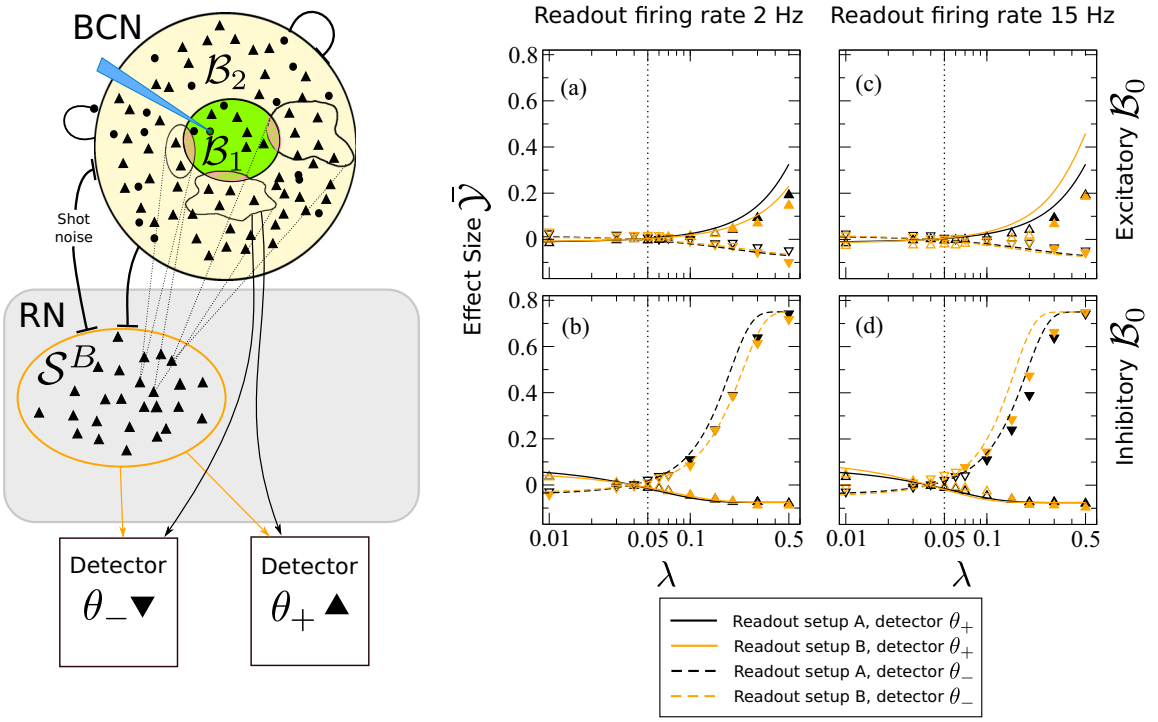

FIG. 4. Detectability of single-cell stimulation does not improve if the readout network (RN) has no recurrent connections but only integrates feed-forward input from the barrel cortex network (BCN), in which a randomly selected cell ($\mathcal{B}_0$) is stimulated. The four panels on the right side show the effect size as a function of the bias λ for setup A (black symbols and lines) and setup B (orange symbols and lines). Detection via the θ_+ detector is represented by upward pointing triangles (simulations) and continuous lines (theoretical approximation discussed in Appendix A). Detection via the θ_- detector is depicted with downward pointing triangles (simulation) and dashed lines (theory). In all panels, the vertical dotted line marks λ_0, which corresponds to no bias in the input to the detector. Closed symbols indicate statistically significant points (Fisher's exact test, p value <0.05). Panels (a) and (c) show results for excitatory $\mathcal{B}_0$, while (b) and (d) show results for inhibitory $\mathcal{B}_0$. In panels (a) and (b) the mean input is set such that the spontaneous firing rate of $\mathcal{S}^B$ is equal to that of the BCN ($\approx$2 Hz). In panels (c) and (d) the firing rate of the readout is higher ($\approx$15 Hz) to make the readout activity more Gaussian.

The magnitude of the cross correlations at low frequencies is $S^{\mathcal{E}\mathcal{E}}_{x_1x_2}(0) \sim 0.5\,\mathrm{Hz}$, which has to be compared with $S^{\mathcal{E}}_{xx}(0)/N_B \approx r^B_{\mathrm{sp}}/N_B \approx 2\times 10^{-4}\,\mathrm{Hz}$, so that the first term in Eq. (25) can be safely neglected. In the linear approximation, output correlations are proportional to the cross correlation between the input to two neurons (here arbitrarily labeled with 1 and 2) [33,34],

$$S^{\mathcal{E}\mathcal{E}}_{x_1x_2}(0) \approx \left|\frac{dr^B_{\mathrm{sp}}}{d\mu}\right|^2 S^{\mathcal{E}}_{\eta_1\eta_2}(0). \tag{26}$$

In the above equation the derivative of r^B_{sp} is taken with respect to the mean input $\mu = R_m I_0$ and $S^{\mathcal{E}}_{\eta_1\eta_2} = \langle\tilde{\eta}_1\tilde{\eta}_2^*\rangle/T$, where the tilde indicates Fourier transformation, asterisk the complex conjugate, and η_k is the input to the kth neuron in $\mathcal{S}^B$. This input term can be written explicitly as

$$\tilde{\eta}_k = \tau_m \sum_{j\in\mathcal{Q}^\lambda(k)} J_{kj} e^{2\pi i D_{kj} f}\tilde{x}_j. \tag{27}$$

The delay term $e^{2\pi i D_{kj} f}$ becomes irrelevant in the limit $f \to 0$. Equation (27) can be multiplied by its complex conjugate and averaged term by term to find an expression for $S^{\mathcal{E}}_{\eta_1\eta_2}(0)$. Inserting the result into Eq. (26) yields

$$\begin{aligned} S^{\mathcal{E}\mathcal{E}}_{x_1x_2}(0) &\approx \left|\frac{dr^B_{\mathrm{sp}}}{d\mu}\right|^2 \tau_m^2 J^2\left[\lambda_c\hat{C}S^E_{xx}(0) + (\hat{C}^2 - \lambda_c\hat{C})S^{EE}_{x_1x_2}(0)\right] \\ &\approx \left|\frac{dr^B_{\mathrm{sp}}}{d\mu}\right|^2 \tau_m^2 J^2\hat{C}^2\left[\frac{\lambda_c S^E_{xx}(0)}{\hat{C}} + S^{EE}_{x_1x_2}(0)\right] \\ &= \alpha^2\left[\frac{\lambda_c S^E_{xx}(0)}{\hat{C}} + S^{EE}_{x_1x_2}(0)\right], \end{aligned} \tag{28}$$

where we recall that $\hat{C}$ is the number of feed-forward inputs per neuron. Approximately $\lambda_c\hat{C} = [\lambda^2 + (1-\lambda)\lambda_0]\hat{C}$ inputs are common to the two neurons. These inputs yield terms proportional to the single-neuron spike-train spectrum $S^E_{xx}(0)$, while all remaining pairs yield a term proportional to cross correlations between excitatory neurons in the BCN $S^{EE}_{x_1x_2}(0)$. We defined $\alpha = \tau_m J\hat{C}dr^B_{\mathrm{sp}}/d\mu$ as shorthand for the linearization of the input-output relation.

For our choice of parameters, the term proportional to the power spectrum in Eq. (28) can be neglected except for the largest values of λ. Exploiting this fact, we can insert Eq. (28) without the first term into Eq. (25), which yields

$$\sigma_B^2 \approx \frac{S^{\mathcal{E}\mathcal{E}}_{x_1x_2}(0)}{\sqrt{\pi}\tau_f} \approx \frac{\alpha^2 S^{EE}_{x_1x_2}(0)}{\sqrt{\pi}\tau_f} \approx \alpha^2\left[\sigma_A^2 - \frac{S^E_{xx}(0)}{N_A\sqrt{\pi}\tau_f}\right], \tag{29}$$

where in the last step we have used Eq. (23) with $X = A$. Neglecting also the second term related to the power spectrum in the last equation gives a simple proportionality between the variances of $R^A_\lambda(t)$ and $R^B_\lambda(t)$:

$$\sigma_B^2 \approx \alpha^2 \sigma_A^2. \tag{30}$$

In the linear response approximation the numerators of the SNR are also proportional to each other:

$$\Delta r^B_\lambda \approx \frac{dr^B_{\rm sp}}{d\mu} \tau_m J \hat{C} \Delta r^A_\lambda = \alpha \Delta r^A_\lambda. \tag{31}$$

Therefore, by combining the above equations one finds that the SNRs for the two setups should be approximately the same:

$$\delta_B(\lambda) = \frac{\Delta r^B_\lambda}{\sigma_B} \approx \frac{\alpha \Delta r^A_\lambda}{\alpha \sigma_A} = \delta_A(\lambda). \tag{32}$$

Although not precise [35], these approximations suggest that the effect sizes for setups A and B should be similar. In Fig. 4 we plot the effect size for the two setups as a function of the bias λ. The vertical dotted line indicates $\lambda = \lambda_0$, i.e., no readout bias. Black symbols and lines correspond to setup A, while orange symbols and lines correspond to setup B. Upward pointing triangles (simulation results) and continuous lines (theoretical approximation discussed in Appendix A) stand for detection by the upper-boundary detector $\bar{\theta}_+$, while downward pointing triangles (simulations) and dashed lines (theory) represent the effect size obtained from the lower-boundary detector $\bar{\theta}_-$. Closed symbols are statistically significant points. In Fig. 4(a) we consider the case of excitatory $\mathcal{B}_0$: a significant detection is possible only for rather large values of the readout bias and setup A slightly outperforms setup B. Only for $\lambda = 0.5$ is the lower barrier significantly different from zero, but the effect size is negative (i.e., the detector reacts less frequently to a stimulus than chance). Otherwise, only the upper-barrier detector can detect the stimulus. If the stimulated cell is inhibitory [Fig. 4(b)], the roles of the two detectors are reversed, and the effect size is much larger and rather similar for both setups, although setup A is again slightly better than setup B. The minimum bias needed for significant detection here is $\lambda \approx 0.1$. The better detectability of inhibitory neurons is consistent with the experimental results [14,15] and in our model is due to the stronger weights of inhibitory neurons. We emphasize that strong inhibition is needed to achieve a stable spontaneous AI firing at low rates and was not chosen to increase the detectability of inhibitory neurons.

Because the LIF model performs a nonlinear transformation of the input, the parameters of the nonlinearity determine to which degree Eq. (32) approximately holds. For instance, we can choose not to match the spontaneous firing rates of the two populations and increase the baseline input to $\mathcal{S}^B$ and thus its spontaneous firing rate to $r^B_{\rm sp} \approx 15$ Hz. In this case, the SNR and the effect size change slightly and the performance of setup B is increased (nothing changes for setup A), as can be seen in Fig. 4(c) for excitatory and in Fig. 4(d) for inhibitory $\mathcal{B}_0$, but the change is moderate.

The results of this subsection show that a population of cells purely integrating input from the BCN does not give a clear advantage to setup B over setup A because both signal and noise are similarly amplified by the input-output relation of the LIF model. For a linear system, the SNR would not change under any circumstance. However, as any integrate-and-fire neuron model is certainly nonlinear, playing with parameters can improve the detection performance in principle. As shown above, increasing the output firing rate of the readout can enhance the SNR for setup B but yields only a modest improvement. Another simple way to influence the SNR would be to change the size of $\mathcal{S}_B$, thus reducing the first term in Eq. (25) and the total readout variance σ_B^2. However, cortical populations cannot be arbitrarily large and firing rates in the cortex *are* typically low, which constrains the possible improvement with the readout considered in this subsection.

B. Readout population with local inhibition

In cortical circuits, excitatory neurons are interconnected with inhibitory neurons; therefore it is natural to extend the RN and include a second population of inhibitory neurons. As for the BCN we keep the ratio of excitatory to inhibitory neurons at four to one. Therefore, the size of this local population of inhibitory neurons $\mathcal{I}$ is $N_\mathcal{I} = \gamma N_B = 2500$. Feed-forward connections do not specifically target excitatory neurons [36]. Therefore, all excitatory and inhibitory neurons in the RN receive the same number of feed-forward excitatory inputs from the BCN and the same local recurrent inhibition (see Fig. 5). Each neuron in the RN evolves according to

$$\tau_m \dot{v}_k(t) = -v_k(t) + R_m[I_{{\rm ext},k}(t) + I_{{\rm rec},k}(t) + I_{{\rm FF},k}(t)]. \tag{33}$$

The external input term is as in the previous subsection, while the recurrent input is purely inhibitory,

$$I_{{\rm rec},k}(t) = \frac{\tau_m}{R_m}\left[-g \sum_{\ell \in \mathcal{L}_i(k)} J^r_{k\ell} x_\ell\left(t - D^r_{k\ell}\right)\right], \tag{34}$$

for both excitatory and inhibitory neurons in the RN. Here, $\mathcal{L}_i(k)$ are sets of $C_\mathcal{I} = 1000$ neurons selected at random within $\mathcal{I}$. The feed-forward input is the same as in Eq. (13),

$$I_{{\rm FF},k}(t) = \frac{\tau_m}{R_m} \sum_{j \in \mathcal{Q}^\lambda_e(k)} J^{\rm FF}_{kj} x_j\left(t - D^{\rm FF}_{kj}\right), \tag{35}$$

where $\lambda = \lambda_e$ if the considered neuron is in $\mathcal{S}^B$ and $\lambda = \lambda_i$ if it is in $\mathcal{I}$. Put differently, the readout bias to excitatory and inhibitory readout neurons is regulated by two separate parameters. To avoid confusion, from now on we indicate with λ_A the readout bias for setup A.

Because each neuron in $\mathcal{S}^B$ and $\mathcal{I}$ receives the same number of excitatory and inhibitory inputs, the spontaneous firing rate of the two populations is similar to the spontaneous rate of the BCN, $r^B_{\rm sp} \approx r^\mathcal{I}_{\rm sp} \approx r_{\rm sp} \approx 2$ Hz.

We first consider the scenario in which learning involves primarily excitatory-excitatory feed-forward connections. To this end, we vary λ_e and fix $\lambda_i = \lambda_0$. We also set $\lambda_A = \lambda_e$ to compare the two setups. In Figs. 5(a) and 5(b) we show the effect size as a function of λ_e for the case of excitatory and inhibitory $\mathcal{B}_0$, respectively. Colors and symbols are as in Fig. 4. The difference between the two setups is striking: For excitatory $\mathcal{B}_0$ the detection is always significant except for values of λ_e very close to the unbiased case $\lambda_e = \lambda_0$ [Fig. 5(a)].

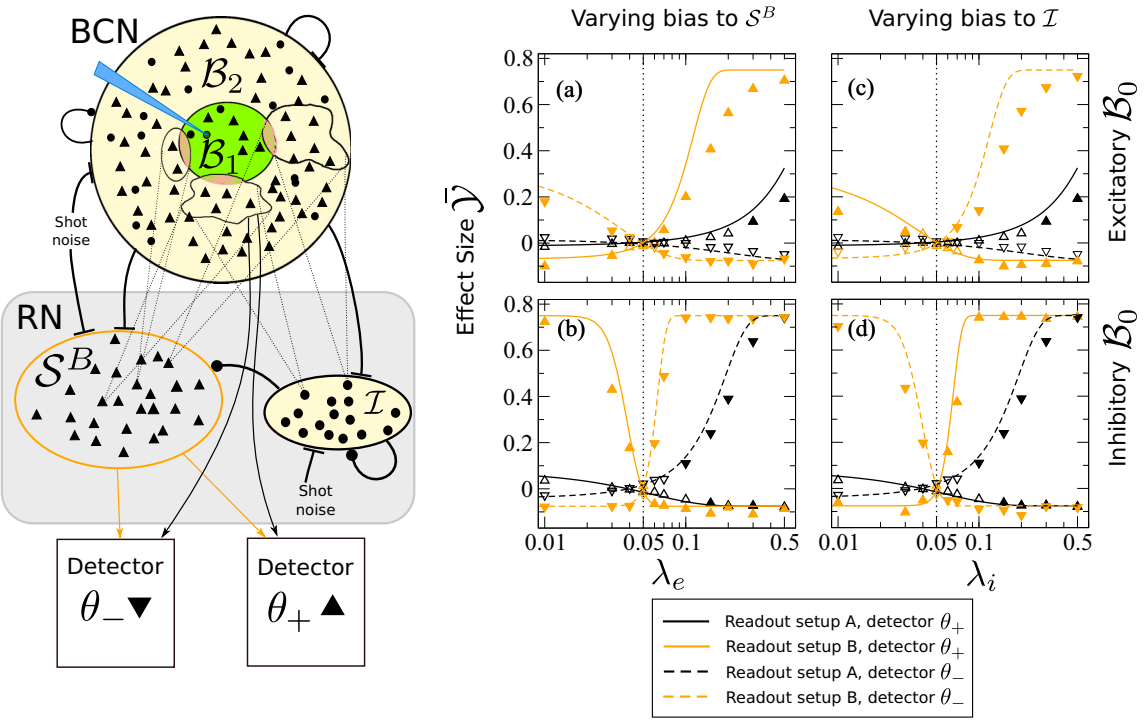

FIG. 5. Large enhancement in the detectability of single-cell stimulation owing to local inhibition. Here, the readout network (RN) consists of the population $\mathcal{S}^B$ and the local inhibitory population $\mathcal{I}$. Both $\mathcal{S}^B$ and $\mathcal{I}$ receive feed-forward input from the barrel cortex network (BCN), in which a random cell is stimulated, and inhibitory input from $\mathcal{I}$. Panels on the right side show the effect size as a function of the bias for setup A, setup B, and both detector types $\theta_\pm$. Color code and symbols are as in Fig. 4. Panels (a) and (c) show results for excitatory $\mathcal{B}_0$, while panels (b) and (d) show results for inhibitory $\mathcal{B}_0$. In panels (a) and (b) λ_e (the connection bias from the BCN to $\mathcal{S}^B$) is varied while $\lambda_A = \lambda_e$ and $\lambda_i = \lambda_0$. In panels (c) and (d) λ_i (the connection bias from the BCN to $\mathcal{I}$) is varied while $\lambda_A = \lambda_i$ and $\lambda_e = \lambda_0$.

For inhibitory $\mathcal{B}_0$, the only point for which the detection is not significant is $\lambda_e = \lambda_0$ [Fig. 5(b)]. For excitatory $\mathcal{B}_0$ and $\lambda_e = 0.5$ [Fig. 5(a)], and for inhibitory $\mathcal{B}_0$ and many values of λ_e [Fig. 5(b)], the maximum effect size of 0.75 is achieved. For excitatory $\mathcal{B}_0$ and $\lambda_e > \lambda_0$ the upper threshold detects the signal [upward pointing triangles and continuous lines in Fig. 5(a)]. For $\lambda_e < \lambda_0$ the lower threshold detects the signal [downward pointing triangles and dashed lines in Fig. 5(a)], because here the detector is biased *against* $\mathcal{B}_1$ and toward $\mathcal{B}_2$, which responds to the stimulus in the opposite direction (see also Sec. I B). The role of the two barriers is swapped in the case of inhibitory $\mathcal{B}_0$ [Fig. 5(b)].

We can interpret this striking difference between the two setups in terms of the SNR, similarly to the previous subsection. To this end, we first consider the linear response of the inhibitory interneurons of the RN population,

$$\Delta r^{\mathcal{I}} \approx \frac{dr^B_{\rm sp}}{d\mu}\left[\tau_m J\hat{C}\Delta r^A_{\lambda_i} - g\tau_m J C_{\mathcal{I}}\Delta r^{\mathcal{I}}\right], \tag{36}$$

and solve it for $\Delta r^{\mathcal{I}}$,

$$\Delta r^{\mathcal{I}} = \alpha\frac{\Delta r^A_{\lambda_i}}{1+g\gamma\alpha}, \tag{37}$$

where we used the fact that $C_{\mathcal{I}} = \gamma C_E = \gamma\hat{C}$ and $\Delta r^A_\lambda = \lambda\Delta r_1 + (1-\lambda)\Delta r_2$. Note that $dr^B_{\rm sp}/d\mu$ (and thus α) here is smaller than in the previous subsection, presumably because of the much larger input noise due to the 1000 inhibitory inputs absent in Sec. II A. We can use Eq. (37) to calculate the linear response of the readout neurons $\mathcal{S}^B$:

$$\Delta r^B \approx \alpha\left(\Delta r^A_{\lambda_e} - g\gamma\Delta r^{\mathcal{I}}\right) = \alpha\left(\Delta r^A_{\lambda_e} - \frac{g\gamma\alpha\Delta r^A_{\lambda_i}}{1+g\gamma\alpha}\right). \tag{38}$$

If we assume that $\gamma g\alpha \gg 1$ and $\Delta r^A_{\lambda_i}/\Delta r^A_{\lambda_e} \approx \lambda_i/\lambda_e$ (this second approximation is not very precise if λ_e, λ_i are very small), we obtain

$$\Delta r^B \approx \alpha\Delta r^A_{\lambda_e}\left(1 - \frac{\lambda_i}{\lambda_e}\right). \tag{39}$$

Turning to the variance of the readout activity σ_B, we first consider input cross correlations that affect σ_B via Eqs. (25) and (26). Input cross correlations can be computed in the same way as in the previous subsection. However, here neuron pairs must be distinguished depending on the neuron type, as average cross correlations are different for different neuron types [37]. Carrying out the averaging yields

$$\begin{aligned} S^{\mathcal{E}}_{\eta_1\eta_2}(0) &\approx \tau_m^2 J^2\hat{C}\big[\lambda_c S^E_{xx}(0) + \hat{\lambda}\gamma g^2 S^{\mathcal{I}}_{xx}(0) + (\hat{C}-\lambda_c)S^{EE}_{x_1x_2}(0) \\ &\quad - 2g\gamma\hat{C}S^{E\mathcal{I}}_{x_1x_2}(0) + g^2\gamma(\gamma\hat{C}-\hat{\lambda})S^{\mathcal{I}\mathcal{I}}_{x_1x_2}(0)\big] \\ &\approx \tau_m^2 J^2\hat{C}^2\left[\frac{\lambda_c S^E_{xx}(0) + \hat{\lambda}\gamma g^2 S^{\mathcal{I}}_{xx}(0)}{\hat{C}} + S^{EE}_{x_1x_2}(0)\right. \\ &\quad \left. - 2g\gamma S^{E\mathcal{I}}_{x_1x_2}(0) + g^2\gamma^2 S^{\mathcal{I}\mathcal{I}}_{x_1x_2}(0)\right], \end{aligned} \tag{40}$$

where $S^E_{xx}(0)$ and $S^{\mathcal{I}}_{xx}(0)$ are low-frequency limits of the single spike-train power spectra of excitatory neurons in the BCN

and of inhibitory neurons in $\mathcal{I}$, respectively. Furthermore, $\hat{\lambda} = C_E/N_B$, $\lambda_c = \lambda_e^2 + (1-\lambda_e)^2\lambda_0$, the term $S^{E\mathcal{I}}_{x_1x_2}(0)$ represents cross correlations between excitatory neurons in the BCN and neurons in $\mathcal{I}$, and the term $S^{\mathcal{I}\mathcal{I}}_{x_1x_2}(0)$ represents cross correlations between pairs of neurons within $\mathcal{I}$. The cross spectra $S^{\mathcal{I}\mathcal{I}}_{x_1x_2}(0)$ and $S^{E\mathcal{I}}_{x_1x_2}(0)$ both depend nontrivially on λ_i, which we do not indicate for simplicity.

Because of the dense connectivity in $\mathcal{I}$, the term $S^{\mathcal{I}\mathcal{I}}_{x_1x_2}(0)$ is negative, while the term $S^{E\mathcal{I}}_{x_1x_2}(0)$ is positive, because excitatory neurons in the BCN drive neurons in $\mathcal{I}$. Taking into account the prefactors, we see that both terms contribute negatively to the sum in Eq. (40) and thus reduce the total input cross correlation for low frequencies. Again, the linear response ansatz Eq. (26) largely overestimates the output cross correlation. However, numerical measurements reveal that $S^{\mathcal{E}\mathcal{E}}_{x_1x_2}(0) \approx S^{EE}_{x_1x_2}(0)$ and that in the end (except for large values of $\lambda_e \gtrsim 0.2$, for which the first term proportional to the power spectrum becomes large),

$$\sigma_B^2 \approx \sigma_A^2. \tag{41}$$

We can now combine this last observation with Eq. (38), which yields for the SNRs δ_A, δ_B of the two setups

$$\frac{\delta_B}{\delta_A} \approx \frac{\alpha}{\Delta r^A_{\lambda_A}}\left(\Delta r^A_{\lambda_e} - \frac{g\gamma\alpha\Delta r^A_{\lambda_i}}{1+g\gamma\alpha}\right) \approx \frac{\alpha}{\Delta r^A_{\lambda_A}}\left(\Delta r^A_{\lambda_e} - \Delta r^A_{\lambda_i}\right). \tag{42}$$

With the help of Eq. (42) we can interpret the results of Figs. 5(a) and 5(b) discussed above. There, $\lambda_A = \lambda_e$ and $\lambda_i = \lambda_0$, so that Eq. (42) is further reduced to (again under the assumption $\gamma g\alpha \gg 1$ and $\Delta r^A_{\lambda_0}/\Delta r^A_{\lambda_e} \approx \lambda_0/\lambda_e$)

$$\frac{\delta_B}{\delta_A} \approx \alpha\left(1 - \frac{g\gamma\alpha\Delta r^A_{\lambda_0}}{(1+g\gamma\alpha)\Delta r^A_{\lambda_e}}\right) \approx \alpha\left(1 - \frac{\lambda_0}{\lambda_e}\right). \tag{43}$$

The input-output linearization is rather steep, $\alpha \approx 8$. Hence, for $\lambda_e \neq \lambda_0$ we expect a sizable improvement in the SNR and, thus, in the effect size. In Figs. 5(a) and 5(b) we see indeed that the effect size is much larger for setup B whenever λ_e is sufficiently larger or smaller than λ_0, in line with Eq. (43). When λ_e is smaller than λ_0, the approximation Eq. (43) predicts a sign flip and a divergence. In Fig. 5(b) we can see a small range (between $0.04 \leqslant \lambda_e \leqslant 0.05$) where the sign is indeed swapped: there, the role of the two barriers for setups B and A is interchanged. However, for smaller values of λ_e, the ratio of the two SNRs is actually positive, which means that Eq. (43) is no longer valid. Going back to Eq. (42) (again with $\lambda_A = \lambda_e$, $\lambda_i = \lambda_0$) and letting $\lambda_e \to 0$ we obtain that the ratio of the SNRs should saturate to

$$\frac{\delta_B}{\delta_A} \xrightarrow[\lambda_e \to 0]{} \alpha\left(1 - \frac{\Delta r_1}{\Delta r_2}\lambda_0\right). \tag{44}$$

We recall that Δr_1 and Δr_2 have opposite signs (see Sec. I B). Hence, the term in parentheses is larger than one, which explains why setup B yields a better effect size also for $\lambda_e \ll \lambda_0$, as seen in Figs. 5(a) and 5(b).

If we consider the scenario of learning taking place at the excitatory-to-inhibitory (BCN to RN) synapses, λ_i is varied while $\lambda_e = \lambda_0$, i.e., the bias for feed-forward E-E connections is left at its "natural" value. To compare with setup A, we now set $\lambda_A = \lambda_i$. Performing these substitutions in Eq. (42) and making the same approximation as in Eq. (43) yields

$$\frac{\delta_B}{\delta_A} \approx -\alpha\left(1 - \frac{\lambda_0}{\lambda_i}\right). \tag{45}$$

The last equation suggests that the increase in the effect size should be very similar in magnitude to the previous case but reversed in sign. Indeed, simulation results for setup B in Figs. 5(c) and 5(d) are almost a copy of Figs. 5(a) and 5(b) with the role of the barriers exchanged.

A further prediction of Eq. (42) is that for $\lambda_e = \lambda_i = \lambda_A$ setup B should bring no improvement compared to setup A. Simulations confirm that in this case setup B performs always worse than setup A (not shown). In addition, if λ_e and λ_i are both changed in opposite directions, for instance by setting $\lambda_e = \lambda_0 + \Delta\lambda$, $\lambda_i = \lambda_0 - \Delta\lambda$, setup B can significantly detect the stimulus for $\Delta\lambda$ as small as ≈ 0.01 for excitatory $\mathcal{B}_0$ and $\Delta\lambda \approx 0.005$ for inhibitory $\mathcal{B}_0$ (not shown).

To sum up the results of this subsection, we conclude that feed-forward inhibition removes to a great extent input cross correlations, thus suppressing the main source of noise in the detection. If the bias of the excitatory readout neurons is equal to the bias of the inhibitory neurons (i.e., $\lambda_e = \lambda_i$), there is also a similar cancellation effect for the signal and, consequentially, the detection performance is similar or worse to that for pure feed-forward excitation. If, however, the fraction of inputs from $\mathcal{B}_1$ to the excitatory and inhibitory neurons in RN is significantly different ($\lambda_e \neq \lambda_i$) we find a strong enhancement of the detection performance.

C. Fully recurrent readout population

Although including a local population of inhibitory interneurons in the RN adds to the realism of the setup, it is still a somewhat artificial assumption to exclude recurrent excitatory connections among neurons in the RN. In this subsection, we allow for recurrent connections between excitatory neurons in the RN. Specifically, we set $\hat{C} = 1000$; otherwise, the feed-forward input term is as in Eq. (13). The number of recurrent excitatory inputs within the RN is set to $C_{\mathcal{E}} = C_E - \hat{C} = 3000$ so that the total number of excitatory inputs per neuron remains the same for all neurons as in the previous cases. Each neuron in RN (both $\mathcal{S}^B$ and $\mathcal{I}$) evolves according to Eq. (33), where now the recurrent input term is

$$I_{\mathrm{rec},k}(t) = \frac{\tau_m}{R_m}\left[\sum_{j\in\mathcal{L}_e(k)} J^r_{kj}x_j\left(t - D^r_{kj}\right) - g\sum_{\ell\in\mathcal{L}_i(k)} J^r_{k\ell}x_\ell\left(t - D^r_{k\ell}\right)\right], \tag{46}$$

where $\mathcal{L}_e(k)$ are randomly selected subsets of $C_{\mathcal{E}}$ neurons in $\mathcal{S}^B$.

We consider again the special case $\lambda_i = \lambda_0$ and set $\lambda_A = \lambda_e$ to compare the two setups [Figs. 6(a) and 6(b)]. We can see that the qualitative picture is very similar to that of the previous subsection: the effect size for setup B is everywhere larger than for setup A, and detection is significant for many values of λ_e both for excitatory [Fig. 6(a)] and inhibitory [Fig. 6(b)]. Again, the detection drops in the vicinity of $\lambda_e = \lambda_0$.

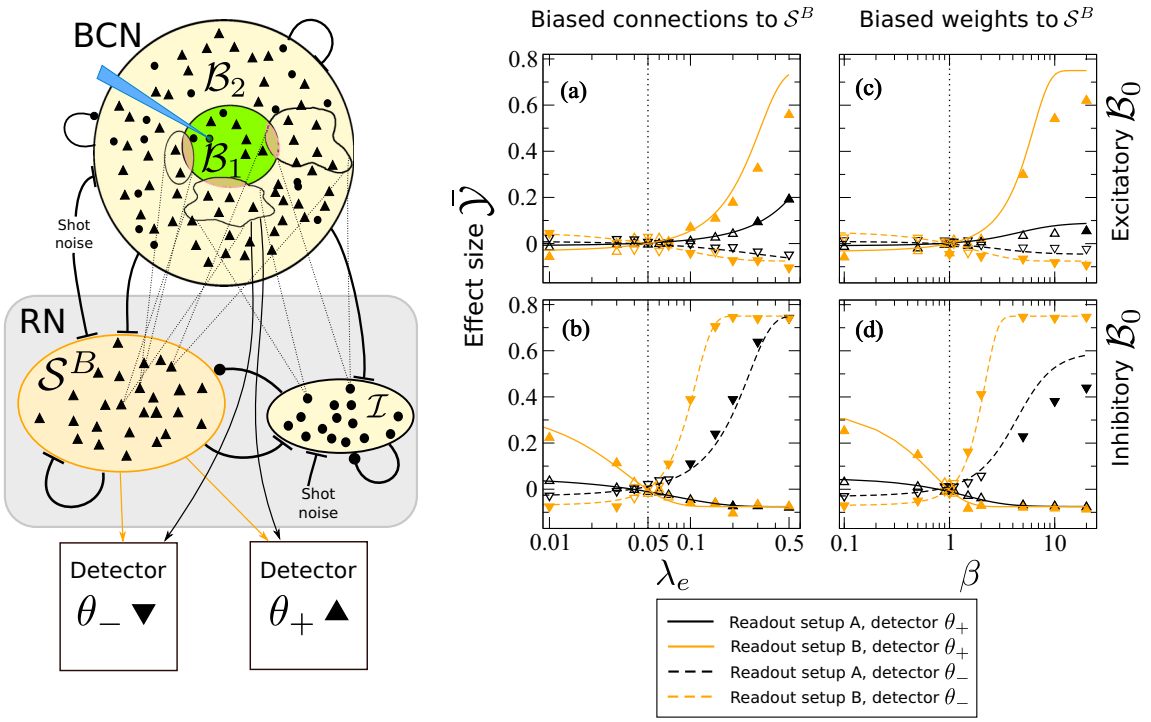

FIG. 6. Fully recurrent E-I network also enhances the detectability of the single-cell stimulation. Here, the readout network (RN) consists of an excitatory population $\mathcal{S}^B$ and an inhibitory population $\mathcal{I}$, both recurrently connected. The RN receives recurrent excitatory input from $\mathcal{S}^B$ (75% of the total recurrent input) and feed-forward excitatory input from the barrel cortex network (BCN) (25% of total input). Panels (a) and (c) show results for excitatory $\mathcal{B}_0$, while panels (b) and (d) show results for inhibitory $\mathcal{B}_0$. In panels (a) and (b) the effect size as a function of the connection bias for excitatory feed-forward projections λ_e (as in the previous cases) is plotted ($\lambda_i = \lambda_0$ and $\lambda_A = \lambda_e$), while in panels (c) and (d) the bias β represents the relative strength of connections from $\mathcal{B}_1$ (see text). Color code and symbols are as in Figs. 4 and 5.

Although the difference between the two setups is quantitatively smaller than in Sec. II B, the difference is still very pronounced and can be explained similarly to the previous subsection (justifying the lengthy discussion therein). In analogy with Eq. (43), we find

$$\frac{\delta_B}{\delta_A} \approx \hat{\alpha}\left(1 - \frac{\lambda_0}{\lambda_e}\right), \tag{47}$$

where

$$\hat{\alpha} = \alpha\left[1 - \frac{\alpha C_{\mathcal{E}}}{\hat{C} + g\gamma\alpha(\hat{C} + C_{\mathcal{E}})}\right]^{-1}. \tag{48}$$

The effective amplification $\hat{\alpha}$ is here the linear response as in Sec. II B but corrected by the factor in square brackets due to recurrent excitation. We note that α is here smaller by a factor four compared to Sec. II B because of its proportionality to $\hat{C}$. This reduction is partially compensated by the recurrent excitation (the factor in square brackets is $\approx$1.7). As a consequence, $\hat{\alpha}$ is still significantly larger than one. The most crucial implication of this in Eq. (47) is that the SNR for setup B is significantly larger than for setup A if λ_e is sufficiently different from λ_i.

Although formation and elimination of synapses is indeed observed in the adult brain as a consequence of learning [38], an interesting question is whether the detection of the stimulus requires (in our model implicit) rewiring of the graph or can be realized by changing synaptic weights. In the following, we address this question by considering this last setup, depicted in Fig. 6 and modifying the definition of readout bias: Instead of biasing the connection probability, we now bias the connection strength. More precisely, we draw connections with probability λ_0 from neurons in $\mathcal{B}_1$ to neurons in $\mathcal{S}^B$ from an exponential distribution with mean βJ and with probability $1-\lambda_0$ from neurons in $\mathcal{B}_2$ to $\mathcal{S}^B$ from an exponential distribution with mean $\hat{J} = J(1-\beta\lambda_0)/(1-\lambda_0)$. By this construction, the average coupling amplitude remains J and is thus independent of β. From the definition of $\hat{J}$ it also follows that β can be at most $1/\lambda_0$. The feed-forward input to each neuron in the RN reads

$$I_{\mathrm{FF},k}(t) = \frac{\tau_m}{R_m}\left[\sum_{j\in\mathcal{Q}_e(k)\cap\mathcal{B}_1} \beta J^{\mathrm{FF}}_{kj} x_j\left(t - D^{\mathrm{FF}}_{kj}\right) + \sum_{i\in\mathcal{Q}_e(k)\cap\mathcal{B}_2} \hat{J}^{\mathrm{FF}}_{ki} x_i\left(t - D^{\mathrm{FF}}_{ki}\right)\right]. \tag{49}$$

For setup A, the readout bias is implemented by weighting differently neurons in $\mathcal{B}_1$ and $\mathcal{B}_2$. In other words, the readout activity is now

$$R^A_\beta(t) = \frac{1}{N_A}\left[\beta \sum_{j\in A\cap\mathcal{B}_1} x_j(t) + \frac{1-\beta\lambda_0}{1-\lambda_0}\sum_{i\in A\cap\mathcal{B}_2} x_i(t)\right] \star F_{\tau_f}(t), \tag{50}$$

where the prefactors guarantee again that the mean of $R^A_\beta(t)$ does not depend on β. In Figs. 6(c) and 6(d) we plot the

effect size as a function of β for excitatory and inhibitory $\mathcal{B}_0$, respectively. Here, the vertical dotted line marks $\beta = 1$, i.e., no bias. Otherwise, the color coding and symbols are as in the previous cases. The results are qualitatively very similar to those of Figs. 6(a) and 6(b). In more detail, we see that the effect size is rather small for setup A and significant detection is possible only for very large values of β (only for the maximum β in the case of excitatory $\mathcal{B}_0$). On the contrary, a significant effect is observed for many values of β for setup B. For inhibitory $\mathcal{B}_0$, the effect is significant for all $\beta \neq 1$.

To sum up the results of this section, we have added recurrent excitatory connections to the RN and considered an alternative definition of the readout bias based on different connection strength rather than on different connection probability. In both cases we found a very similar qualitative picture: the single-cell stimulation can be much better detected in the activity of a recurrent readout network than in the activity of the directly stimulated network.

III. DISCUSSION

In this paper we considered the problem of detecting the strong but brief activation of a single neuron in a large network of integrate-and-fire neurons. We extended our previous model of Ref. [19] in ways that are both more biologically plausible and effective.

The detection scheme introduced in Ref. [19] consisted of a detector reacting to crossings of an upper or lower threshold by the filtered activity of a subpopulation of the stimulated network. This represented a cartoon of a readout population integrating input from the barrel cortex and triggering a downstream decision. If the detector represents an excitatory population, the downstream reaction would be reasonably triggered by crossing a certain activity level, i.e., an upper threshold. If the detector stands for an inhibitory readout population, the downstream event would be triggered by disinhibition, i.e., by reducing the inhibition to its target. This mechanism requires a dip in the instantaneous firing, i.e., the crossing of a lower threshold. Because neuronal populations are either excitatory or inhibitory, a detector playing both roles at the same time disagrees with this picture. Furthermore, using two symmetric thresholds requires high precision in centering the detector on the mean firing rate: if the detector is misaligned, one threshold becomes useless causing the detection performance to decrease. Requiring such fine-tuning is another weakness of the detector introduced previously. Here, by using two separate detectors we solve both issues at the same time: θ_+ may represent a caricature for the effect triggered by an excitatory neuronal population, while θ_- could be the threshold for a disinhibitory pathway.

The second difference in the detector used here is that we no longer optimized the detection threshold, but chose a level corresponding to a fixed false positive rate. Although it is conceivable that the animal optimizes the threshold during training, using the optimal threshold poses a technical problem (discussed in Appendix C): it requires generating separate data sets for the extraction of the optimal threshold and the calculation of significance levels, which is computationally expensive. In addition, learning the optimal threshold requires training during the single-cell stimulation. In reality, most of the training is done by extracellular microstimulation, which activates a group of cells in the local network [39,40]. On the contrary, choosing a threshold corresponding to a fixed false positive rate, as done here, only requires the spontaneous activity and is therefore more consistent with the experiments.

These considerations suggest that the detector we use here is a better, although very simplistic, representation of how the activity of the readout population can trigger the behavioral effect. However, feeding the detector with the activity of a subpopulation of the stimulated network (as done in setup A, akin to the procedure in Ref. [19]) is equivalent to assuming that the decision about the presence of the stimulus is taken within the stimulated network itself. Functionally, this is highly unlikely and thus including a second population is an important extension of the model.

The first possibility we considered for the readout network (Sec. II A) can be regarded as an implementation of the single-barrier detector with LIF neurons: the neuronal dynamics play the role of the filtering and the firing that of the decision threshold (the reset mechanism has no parallel in the detector). Therefore, it may appear in the hindsight not too surprising that the readout performance is similar. Still, a population of threshold units with independent noise can in principle decode a signal better than the single unit, an effect known as suprathreshold stochastic resonance [41–44]. In our case, however, the input noise is highly correlated, which makes adding more readout neurons effectively redundant.

In the second configuration (Sec. II B), we added inhibition, thus enhancing the plausibility of the readout model. Local inhibitory neurons track global fluctuations in the input and, by inhibiting themselves and the excitatory readout neurons, actively cancel input cross correlations [5]. Reducing cross correlations decreases the noise; however, the signal itself can be seen a source of cross correlation that feed-forward inhibition can track and remove [45], which does happen, for instance, in the case of no learning, i.e., of no readout bias.

The presence of inhibitory neurons in the readout also implies that two different sets of feed-forward connections can be biased: those from the stimulated network to the excitatory readout neurons (corresponding to the bias parameter λ_e) and those from the stimulated network to the inhibitory readout neurons (bias parameter λ_i). We showed that increasing λ_e is tantamount to increasing λ_i, provided that the upper threshold detector θ_+ is exchanged with the lower threshold detector θ_-. Furthermore, decreasing λ_e or λ_i below the natural value λ_0 also reverses the role of the barrier, which makes possible to detect both excitatory or inhibitory $\mathcal{B}_0$ by using only the θ_+ or only the θ_- detection scheme. Although *one type* of detector is enough to detect both kinds of stimuli, two separate detectors are still required. Using the very same detector, for instance θ_+, would require increasing λ_e to detect an excitatory $\mathcal{B}_0$ and increasing λ_i to detect an inhibitory $\mathcal{B}_0$. However, if both bias parameters are changed in parallel, the two signal pathways (the direct one and the inhibitory one) are still balanced and the signal is suppressed. Consequently, two different detectors are needed to detect excitatory and inhibitory $\mathcal{B}_0$.

We showed plots for simulations in which λ_e and λ_i were varied separately but another way of breaking the balance is to change both at the same time. As discussed at the end of

Sec. II B, the two parameters must be changed in opposite directions to improve detectability. If we imagine the bias as the product of a Hebbian-like learning rule, it is intuitive that maximizing correlation between the firing rate of $\mathcal{B}_1$ and of $\mathcal{S}_B$ can be done both by potentiating direct connections between the two and by weakening feed-forward inhibition. Put differently, changing λ_e and λ_i in opposite directions is two ways of obtaining the same effect. Conversely, it seems unlikely that a learning rule would increase or decrease λ_e and λ_i by the same amount, as the two effects cancel each other.

It must be noted that the bias increases the correlations due to common inputs. However, the larger component of input cross correlations consists of global oscillations and not of shared input. These global cross correlations are removed by inhibition regardless of the bias, which hence enables the readout inhibition to cancel a large portion of noise without eliminating the signal.

In the third part of the results (Sec. II C) we checked the robustness of the results by adding recurrent excitation to the readout network and decreasing the number of feed-forward inputs. The recurrent excitatory connections increase cross correlations within the readout network, thus increasing the readout noise, while the smaller number of feed-forward inputs reduces the signal. Nevertheless, the single-cell stimulation remained detectable even if the feed-forward input from the stimulated network was only one fourth of the total recurrent input (and an even smaller fraction of the total input, considering the external shot noise). Still, a sufficiently large number of feed-forward inputs is needed to have amplification of the signal-to-noise ratio (to increase what we called α in this study), or possibly fewer inputs with stronger synaptic efficacy.

Finally, we considered a change of synaptic amplitudes as a different possibility to bias the readout: we strengthened (or weakened) direct connections from $\mathcal{B}_1$ to $\mathcal{S}_B$ while keeping the mean input to each neuron constant, which in turn ensures that the mean firing rate does not change too much. An explicit learning model is beyond the scope of this paper, but it is conceivable that such a synaptic change could be achieved by means of a Hebbian learning rule with local homeostasis.

In all configurations considered here, the intermediate readout network of setup B considerably lowered the necessary bias for detection, which means that a smaller rewiring or synaptic change from the naive state is required. Importantly, *some* learning is still necessary, which is consistent with the experiments: if the stimulus were detectable in our model with no bias at all, it would be at odds with the fact that untrained animals cannot report single-cell stimulation [14]. One potential inconsistency of our model is that the bias refers to a specific $\mathcal{B}_1$, and therefore to a $\mathcal{B}_0$, while the training is not specific to a particular cell (it is done by using microstimulation, which affects a larger area). Still, microstimulation is repeated in between trials to keep the rat attentive [14]. It is possible that microstimulation redirects the bias in between trials toward the area around $\mathcal{B}_0$, thus effectively adjusting λ. For this picture to make sense, a network with spatial structure is needed, which is missing in our model. Although the kind of dynamics and cross correlations in networks with spatial structure can be very different from those of a homogeneous random network, a global hidden variable can be a major source of correlations, i.e., affecting the whole population [46]. In such a case, feed-forward inhibition could still effectively remove cross correlations while the biased input would permit detection. It is an interesting open problem to consider the detectability in spatially structured networks.

Besides neglecting a spatial profile in the connectivity, we made further simplifying assumptions regarding the synaptic interactions in our model. To ease the theoretical analysis, we modeled the input from other cells as delta-current pulses. A more realistic choice would be to use conductance-based synapses with an additional filter dynamics (for a comparison of the network dynamics with conductance-based and current-based synapses, see [47]). Changing the synaptic dynamics in this way will cause quantitative differences in the firing statistics and response behavior of the network. However, as long as the network still resides in the asynchronous state with low firing rates, we do not expect that fast synaptic dynamics would impact our results on the detection problem significantly. Likewise, we do not believe that the exact choice of the transmission delays (whether they are randomly distributed, as in our study, or set to a sharp value, as in many other studies) is crucial for the detectability of the single-cell stimulation.

This paper focused on the role of feed-forward configurations, although feedback projections from other cortical areas to the barrel cortex exist [36]. How could feedback projections from the readout network to the stimulated network change our results? For concreteness, let us focus on the case of excitatory $\mathcal{B}_0$, θ_+ detector, and positive λ_e as in Fig. 5(a), solid line. We can imagine that a beneficial effect for detectability could be provided by excitatory feedback projections from the readout $\mathcal{S}^B$ to *excitatory* neurons in the stimulated BCN. Such connections could form as the result of learning and would not need to be biased toward particular excitatory neurons to build a positive feedback loop: the excitatory $\mathcal{B}_0$ with positive bias λ_e raises the firing rate of $\mathcal{S}^B$; the excitatory feedback would then raise the firing rate of neurons in the BCN thus increasing Δr^A and, remembering Eq. (39), Δr^B. For the case of inhibitory $\mathcal{B}_0$, a second dedicated readout is required, for instance a θ_+ detector and positive λ_i bias, as in Fig. 5(d), solid line. In this case, the feedback projections should target *inhibitory* neurons of the stimulated network, thus amplifying the negative firing-rate response of both the BCN and the readout network. These considerations suggest that a suitable feedback may potentially improve detectability by providing additional drive to the BCN that goes in the same direction as the stimulus. However, a positive feedback loop will also amplify fluctuations in the readout activity, thus exerting a negative effect on the detectability. It is hard to predict which effect would prevail and whether the feedback would provide a net gain in the detectability.

In this study we showed how stimulating a single cell can transiently trigger a detectable change in the activity of a second network. Is the activation of this second network enough to trigger the behavioral response, or do we need to consider a chain of networks? Anatomical studies show that there are direct connections from the somatosensory area S1 (to which the barrel cortex belongs) to the motor area M1 [36]; it has also been shown that stimulating a single neuron in M1 can elicit whisker movements [13]. One could speculate that the readout population considered here may be embedded in M1 and thus directly cause the response.

Still, compared to the brief wiggle of a whisker, the licking response is a more complex movement that may require a "conscious" decision and the concerted involvement of other cortical areas. It is therefore an interesting open problem to consider more elaborate readout configurations involving multiple populations.

ACKNOWLEDGMENT

This work has been funded by BMBF (FKZ:01GQ1001A) and DFG (GRK 1589/2).

APPENDIX A: SPECTRAL MEASURES, THEORETICAL APPROXIMATION OF THE EFFECT SIZE, AND ITS CONNECTION TO THE SIGNAL-TO-NOISE RATIO

We define the spike-train power spectrum of a single neuron as

$$S_{xx}(f) = \lim_{T\to\infty} \frac{\langle \tilde{x}(f)\tilde{x}^*(f)\rangle}{T}, \tag{A1}$$

where the tilde indicates Fourier transform

$$\tilde{x}(f) = \int_0^T dt\, e^{2\pi i f t} x(t), \tag{A2}$$

the asterisk the complex conjugate, and T is the duration of the spike train. Angular brackets indicate averaging over trials and over different neurons in a homogeneous population. The cross spectrum between two neurons is defined analogously:

$$S_{x_1x_2}(f) = \lim_{T\to\infty} \frac{\langle \tilde{x}_1(f)\tilde{x}_2^*(f)\rangle}{T}, \tag{A3}$$

where the averaging is here meant over trials and neuron pairs. All spectral measures defined here as well as the following derivation Eq. (A4) hold for stationary processes and refer to the spontaneous activity of the network, that is, before the stimulus onset.

First, we would like to derive Eq. (23), i.e., the subdivision of σ_X^2, the variance of $R_\lambda^X(t)$, and the filtered readout activity Eq. (14). We start by expressing σ_X^2 as the integral over its power spectrum [48], isolate terms with the same index, and use the above definitions. We then make use of the fact that all spectral measures we consider are rather flat for frequencies up to the inverse of the filter time constant $1/\tau_f = 10\,\mathrm{Hz}$ (not shown), and that the size of the readout population N_X is large for both $X = A, B$. Furthermore, when integrating the filter we neglect the Heaviside function in Eq. (15):

$$\begin{aligned}
\sigma_X^2 &= \int_{-\infty}^{+\infty} df\, S_{RR}(f) = \frac{1}{T}\int_{-\infty}^{+\infty} df\, \langle \tilde{R}(f)\tilde{R}^*(f)\rangle \\
&= \int_{-\infty}^{+\infty} df\, \frac{|\tilde{\mathcal{F}}_{\tau_f}(f)|^2}{TN_X^2} \sum_{x_i,x_j\in\mathcal{S}^X} \langle \tilde{x}_i(f)\tilde{x}_j^*(f)\rangle \\
&= \int_{-\infty}^{+\infty} df |\tilde{\mathcal{F}}_{\tau_f}(f)|^2 \left(\frac{S_{xx}(f)}{N_X} + \frac{N_X(N_X-1)S_{x_1x_2}(f)}{N_X^2}\right) \\
&\approx \left(\frac{S_{xx}(0)}{N_X} + \frac{N_X(N_X-1)S_{x_1x_2}(0)}{N_X^2}\right)\int_{-\infty}^{+\infty} df\, |\tilde{\mathcal{F}}_{\tau_f}(f)|^2 \\
&\approx \frac{S_{xx}(0)}{N_X\sqrt{\pi}\tau_f} + \frac{S_{x_1x_2}(0)}{\sqrt{\pi}\tau_f}.
\end{aligned} \tag{A4}$$

We now give more details on the relation between effect size and SNR. The population activity $R_\lambda^X(t)$ is the sum of many weakly correlated low-pass filtered spike trains; therefore it can be approximated as a Gaussian process. The effect size is defined as the difference between correct detection and false positive rate for a given threshold $\theta_\pm$. Both rates are related to an integral over the first passage time (FPT) density of the stochastic process $R_\lambda^X(t)$ either in the absence or presence of the stimulus. Finding the exact FPT density is a difficult problem even within the Gaussian approximation.

As already discussed in [19], a crude "sampling" approximation yields fair results. We imagine that the probability for the Gaussian *process* $R_\lambda^X(t)$ exceeding a certain threshold within the time window T_w is equivalent to asking how likely it is that at least one out of n independent draws of a Gaussian *variable* exceeds this threshold. The number of draws can be estimated as the ratio of the time window and the correlation time of the stochastic process, i.e., $n = T_w/\tau_c$. Because the filter time constant τ_f is longer than the correlation time of the unfiltered network activity, one can approximate $\tau_c \approx \tau_f$. Within this picture, the correct detection rate is

$$\mathcal{W}(\theta_\pm) \approx 1 - \prod_{k=0}^{T_w/\tau_c} p_R(\theta_\pm, k\tau_c), \tag{A5}$$

and $p_R(\theta_\pm, k\tau_c)$ is the probability of $R_\lambda^X(t)$ being below θ_+ or above θ_- at time $t = k\tau_c$, respectively. For a Gaussian process, this probability is simply the cumulative density for θ_+ or the complementary density for the case of θ_-. The presence of the stimulus renders both the mean and variance time-dependent. However, to simplify the theory, we make the following approximations: (i) we treat the variance as time-independent and equal to the stationary value σ_X^2; (ii) we approximate the time dependence of the mean by a box-shaped function of duration T_s and height Δr_λ^X, added on top of the stationary mean. Under these assumptions, the correct detection rate becomes

$$\mathcal{W}(\theta_\pm) \approx 1 - p_R(\theta_\pm, \delta_X)^{T_s/\tau_f} p_R(\theta_\pm, 0)^{\frac{T_w - T_s}{\tau_f}}, \tag{A6}$$

where

$$\begin{aligned}
p_R(\theta_+, \delta_X) &= \int_{-\infty}^{\theta_+} da\, \mathcal{N}\left(a, \Delta r_\lambda^X, \sigma_X\right) \\
&= \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{\theta_+ - \Delta r_\lambda^X}{\sqrt{2}\sigma_X}\right)\right] \\
&= \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{\theta_+}{\sqrt{2}\sigma_X} - \frac{\delta_X}{\sqrt{2}}\right)\right],
\end{aligned} \tag{A7}$$

where erf is the error function. The false positive rate can be approximated in the same way. Here, the probability of exceeding the threshold for a single sample does not depend on time so that the result is even simpler:

$$\mathcal{Z}(\theta_\pm) \approx 1 - p_R(\theta_\pm, 0)^{T_w/\tau_f}. \tag{A8}$$

Equation (A8) can be solved for $\theta_\pm$ and inserted into Eq. (A6) to obtain an approximation for the ROC curve.

To obtain the effect size, we assume a false positive rate of 0.25 and solve Eq. (A8) for the corresponding value of the

threshold-to-noise ratio:

$$\frac{\bar{\theta}_+}{\sqrt{2}\sigma_X} = \mathrm{erf}^{-1}\left[2\left(\frac{3}{4}\right)^{\tau_f/T_w} - 1\right]. \quad \text{(A9)}$$

The final expression for the effect size reads

$$\bar{\mathcal{Y}}^+ = p_R(\bar{\theta}_+, 0)^{\frac{T_w - T_s}{\tau_f}}\left[p_R(\bar{\theta}_+, 0)^{\frac{T_s}{\tau_f}} - p_R(\bar{\theta}_+, \delta_X)^{\frac{T_s}{\tau_f}}\right]. \quad \text{(A10)}$$

For given values of T_w, T_s, τ_f, the effect size is a monotonically increasing function of the SNR, a property we exploit in the Results section. Theoretical lines are obtained by evaluating Eqs. (A7), (A9), and (A10) (and the analogous expressions for the lower boundary detector). The numerator of δ_A is obtained from the mean-field equations (8) and (9) and the numerator of δ_B is calculated via linear response theory, as explained in Secs. II A and II B. For the calculation of the cross correlations, linear response theory is inadequate. Therefore, for the denominator of the SNR we use σ_A and σ_B as measured from network simulations in the case of $\lambda = \lambda_0$, thus ignoring the dependence of cross correlations on λ. The only exception is the case of setup A in Figs. 6(c) and 6(d), where the dependence of the variance σ_A^2 on the bias parameter β is rather strong and follows a simple quadratic equation

$$\sigma_A^2 = \sigma_A^2(\beta = 0) + (\beta^2 - 2\beta)\frac{S_{xx}^E \lambda_0}{N_A(1-\lambda_0)\sqrt{\pi}\tau_f}, \quad \text{(A11)}$$

which can be calculated by using Eq. (50) and Eq. (A4).

APPENDIX B: SINGLE-BARRIER VERSUS DOUBLE-BARRIER DETECTOR

Here, we compare the results of Sec. II B to those obtained with the detector introduced in Ref. [19], which employs two symmetric thresholds instead of a single one and picks the optimal threshold instead of a fixing a false positive rate.

We first study the effect of using two thresholds placed at $w \pm \theta_d$, in the following referred to as the θ_d detector. In Fig. 7(a) we reconsider the case of Fig. 5(a) and plot the effect size obtained with $w = 2$ Hz with orange circles together with results for the θ_+ and θ_- detectors for comparison. If the detector is centered on the correct spontaneous firing rate, i.e., $w = r_{\mathrm{sp}}$, it can detect both a positive and a negative deflection. In Fig. 7(a) we see that the θ_d detector can indeed detect both positive ($\lambda > \lambda_0$) and negative ($\lambda < \lambda_0$) deviations from the mean and always lies between the other two. The effect size for the θ_d detector is smaller than for the best of the other two detectors because of the interference of the second threshold. If the detector is not perfectly centered on the correct mean r_{sp}, one of the two thresholds becomes less relevant: if the baseline level w is chosen too low, the upper threshold becomes less relevant; if it is chosen too high, the opposite happens. For weak signal, as for $\lambda < \lambda_0$ in Fig. 7(a), even a small imprecision in the choice of w becomes apparent with the curve for the θ_d detector clinging to that of the θ_- detector.

In Fig. 7(b) we plot the effect size obtained by optimizing the threshold instead of fixing a false positive rate. In other words, for each ROC curve we select the threshold that maximizes the distance to the diagonal. For reasons explained in Appendix C, the statistical significance for this case requires particular care and was omitted here. While all points are slightly higher compared to the case of fixed threshold of Fig. 7(a), the qualitative picture is very similar, and even quantitatively differences are not large.

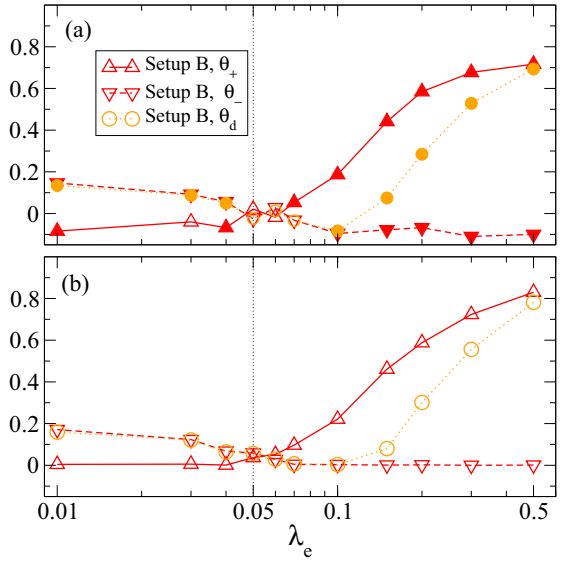

FIG. 7. Qualitative effects do not depend on the choice of the detector. Effect size obtained from the detector used in this study and in Ref. [19], applied to the case considered in Sec. II B. The detector with two symmetric thresholds (orange circles) detects both above and below λ_0 but the effect size is smaller in magnitude. In panel (a) are plotted results for a fixed false positive rate, as done in the main text; in panel (b) we plot results for an optimized threshold, as done in Ref. [19]. With this definition the effect size cannot be negative. Otherwise, for positive effect size it can be seen that optimizing the threshold has a very small effect as comparison to panel (a) reveals. In (b) we omit indicating the statistical significance (see Appendix C).

APPENDIX C: STATISTICAL SIGNIFICANCE OF EFFECT SIZE

Here, we assess the applicability of the statistical significance test used here and in Ref. [19]. Specifically, we illustrate how optimizing the threshold from a data set and then calculating the significance level from the same data set leads to potentially wrong p values. To this end, we use the simplified description of the detection process described in Appendix A. For concreteness, we restrict ourselves to the case of the single upper threshold and use the readout activity $R^B(t)$ with parameters as in Sec. II B. However, the following considerations apply to all detectors considered and do not depend on the particular choice of parameters.

For different scenarios we mimic a so-called catch trial, in which false positive and correct detection rates are both calculated in the absence of the stimulus. We first measure the histogram of the stationary activity $R^B(t)$ without stimulus. We then draw independently two sets of $n = 10$ independent samples from the histogram. Again, the particular choice of n does not influence the following considerations. If at least one sample out of n in the first (second) set is above the detection

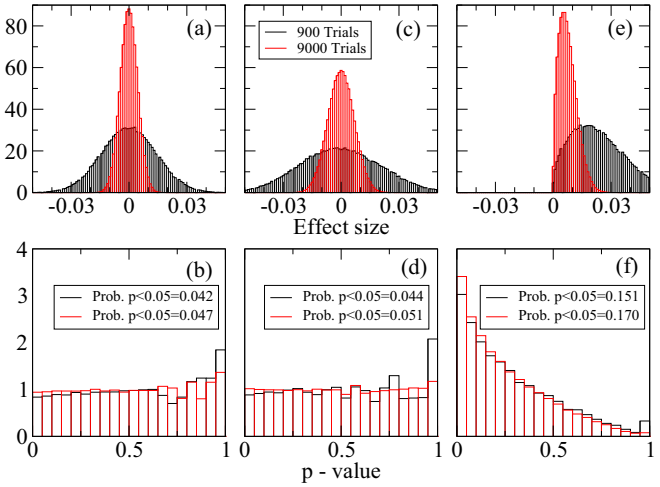

FIG. 8. Simplified detection experiment without signal in order to mimic catch trials. Distribution of effect size (here only due to finite-size fluctuations) and p value calculated with Fisher's test for three ways of choosing the threshold: fixed threshold [(a), (b)], threshold corresponding to fixed false positive rate [(c), (d)], and optimal threshold [(e), (f)].

threshold θ_+, a false positive (correct detection) event is registered. Because the two sets are drawn from the same distribution, this represents a "catch trial" in the detection experiment. By repeating this random sampling $N_{\text{trials}} = 900$ times—the number of trials used in all simulations—and averaging, we obtain a false positive and a correct detection rate and, hence, an effect size. A p value can be calculated for this effect size by using Fisher's exact test. Because of the finite size of the sample, the effect size will not be exactly zero, but randomly distributed. Depending on the procedure we use to choose the threshold and, therefore, the effect size, the resulting distribution will be different. We consider three possibilities: (i) we fix a threshold *a priori* and use it to determine both false positive and correct detection rate; (ii) we fix a false positive rate, find the corresponding threshold, then use this threshold to calculate the correct detection rate, which is the procedure we use in the main text; (iii) we select the threshold that maximizes the effect size and use it to determine both false positive and correct detection rate, which is the procedure used in Ref. [19]. For each case, we construct a histogram for effect size and p value obtained from 200 000 repetitions of the "catch trial" procedure described above.

If the threshold is fixed beforehand (case i) the effect size has the distribution of the difference of two binomial variables. For a large number of trials it is approximately Gaussian and symmetric [Fig. 8(a)]. For this case, p values are roughly uniformly distributed as expected [49], although not perfectly because of the finite number of trials [Fig. 8(b)]. Increasing the number of trials (red histogram) renders the histogram flatter.

If a false positive level is fixed and used to determine a threshold (case ii) the histogram is still symmetric [Fig. 8(c)] but has an increased width, due to the spread in the threshold. Still, p values are uniformly distributed as in the case of fixed threshold [Fig. 8(d)].

Finally, if the threshold is optimized with respect to the effect size (case iii) we obtain a histogram that is not symmetric [Fig. 8(e)]. The mean value of the effect size is here not zero, and the histogram of p values is not flat [Fig. 8(f)]. In particular, the probability of $p < 0.05$ is three times larger than 5%, and therefore its value does not express the intended statistical significance. One conceptually simple but computationally very expensive way to solve this problem would be to generate two distinct data sets, and use the first one to determine the optimal threshold and the second one to compute effect sizes.

[1] E. R. Kandel, J. H. Schwartz, and T. M. Jessell, *Principles of Neural Science* (McGraw-Hill, New York, 2000).
[2] A. Treves, Network **4**, 259 (1993).
[3] C. van Vreeswijk and H. Sompolinsky, Science **274**, 1724 (1996).
[4] N. Brunel, J. Comput. Neurosci. **8**, 183 (2000).
[5] A. Renart, J. de la Rocha, P. Bartho, L. Hollender, N. Parga, A. Reyes, and K. D. Harris, Science **327**, 587 (2010).
[6] A. Litwin-Kumar and B. Doiron, Nat. Neurosci. **15**, 1498 (2012).
[7] S. Ostojic, Nat. Neurosci. **17**, 594 (2014).
[8] S. Wieland, D. Bernardi, T. Schwalger, and B. Lindner, Phys. Rev. E **92**, 040901 (2015).
[9] M. Brecht, M. S. Fee, O. Garaschuk, F. Helmchen, T. W. Margrie, K. Svoboda, and P. Osten, J. Neurosci. **24**, 9223 (2004).
[10] C.-y. T. Li, M.-m. Poo, and Y. Dan, Science **324**, 643 (2009).
[11] P. Bonifazi, M. Goldin, M. A. Picardo, I. Jorquera, A. Cattani, G. Bianconi, A. Represa, Y. Ben-Ari, and R. Cossart, Science **326**, 1419 (2009).
[12] A. C. Kwan and Y. Dan, Curr. Biol. **22**, 1459 (2012).
[13] M. Brecht, M. Schneider, B. Sakmann, and T. W. Margrie, Nature (London) **427**, 704 (2004).
[14] A. R. Houweling and M. Brecht, Nature (London) **451**, 65 (2008).
[15] G. Doron, M. von Heimendahl, P. Schlattmann, A. R. Houweling, and M. Brecht, Neuron **81**, 653 (2014).
[16] N. Tanke, J. G. G. Borst, and A. R. Houweling, J. Neurosci. **38**, 2057 (2018).
[17] J. Wolfe, A. R. Houweling, and M. Brecht, Curr. Opin. Neurobiol. **20**, 306 (2010).
[18] K. D. Harris and A. Thiele, Nat. Rev. Neurosci. **12**, 509 (2011).
[19] D. Bernardi and B. Lindner, Phys. Rev. Lett. **118**, 268301 (2017).
[20] D. J. Amit and N. Brunel, Cereb. Cortex **7**, 237 (1997).
[21] D. J. Amit and N. Brunel, Network **8**, 373 (1997).
[22] W. Gerstner, W. M. Kistler, R. Naud, and L. Paninski, *Neuronal Dynamics: From Single Neurons to Networks and Models of Cognition* (Cambridge University Press, Cambridge, UK, 2014).
[23] H. S. Meyer, V. C. Wimmer, M. Oberlaender, C. P. J. de Kock, B. Sakmann, and M. Helmstaedter, Cereb. Cortex **20**, 2277 (2010).

[24] M. Brecht and B. Sakmann, J. Physiol. **543**, 49 (2002).
[25] C. P. J. de Kock, R. M. Bruno, H. Spors, and B. Sakmann, J. Physiol. **581**, 139 (2007).
[26] M. J. E. Richardson and R. Swarbrick, Phys. Rev. Lett. **105**, 178102 (2010).
[27] A. Lerchner, C. Ursta, J. Hertz, M. Ahmadi, P. Ruffiot, and S. Enemark, Neural Comput. **18**, 634 (2006).
[28] A. C. Flint, U. S. Maisch, J. H. Weishaupt, A. R. Kriegstein, and H. Monyer, J. Neurosci. **17**, 2469 (1997).
[29] B. McNamara and K. Wiesenfeld, Phys. Rev. A **39**, 4854 (1989).
[30] L. Gammaitoni, P. Hänggi, P. Jung, and F. Marchesoni, Rev. Mod. Phys. **70**, 223 (1998).
[31] A. Kruscha and B. Lindner, Phys. Rev. E **92**, 052817 (2015).
[32] A. Kruscha and B. Lindner, Phys. Rev. E **94**, 022422 (2016).
[33] S. Ostojic, N. Brunel, and V. Hakim, J. Neurosci. **29**, 10234 (2009).
[34] R. D. Vilela and B. Lindner, Phys. Rev. E **80**, 031909 (2009).
[35] In Eq. (29) the contribution of the second term (proportional to the single-neuron power spectrum) accounts for about 30% of the total variance, so neglecting it is not a good approximation. On the other hand, the linear approximation of the input-output relation, Eq. (28), underestimates the output cross spectrum, so that in the end the two errors partially compensate and the measured ratio of the two standard deviations ($\approx$18) is not too far from the predicted value of $\alpha \approx 16$.
[36] D. Feldmeyer, M. Brecht, F. Helmchen, C. C. H. Petersen, J. F. A. Poulet, J. F. Staiger, H. J. Luhmann, and C. Schwarz, Prog. Neurobiol. **103**, 3 (2013).
[37] M. Helias, T. Tetzlaff, and M. Diesmann, New J. Phys. **15**, 023002 (2013).
[38] D. B. Chklovskii, B. Mel, and K. Svoboda, Nature (London) **431**, 782 (2004).
[39] S. Butovas and C. Schwarz, J. Neurophysiol. **90**, 3024 (2003).
[40] M. H. Histed, V. Bonin, and R. C. Reid, Neuron **63**, 508 (2009).
[41] N. G. Stocks, Phys. Rev. Lett. **84**, 2310 (2000).
[42] T. Hoch, G. Wenning, and K. Obermayer, Phys. Rev. E **68**, 011911 (2003).
[43] S. Durrant, Y. Kang, N. Stocks, and J. Feng, Phys. Rev. E **84**, 011923 (2011).
[44] M. Beiran, A. Kruscha, J. Benda, and B. Lindner, J. Comput. Neurosci. **44**, 189 (2018).
[45] Y. Hu, J. Zylberberg, and E. Shea-Brown, PLoS Comput. Biol. **10**, e1003469 (2014).
[46] R. Rosenbaum, M. A. Smith, A. Kohn, J. E. Rubin, and B. Doiron, Nat. Neurosci. **20**, 107 (2017).
[47] A. Kumar, S. Schrader, A. Aertsen, and S. Rotter, Neural Comput. **20**, 1 (2008).
[48] R. Stratonovich, *Topics in the Theory of Random Noise* (Gordon & Breach, New York, 1963), Vol. 2.
[49] D. R. Cox and D. V. Hinkley, *Theoretical Statistics* (Chapman & Hall, London, 1974).