J Comput Neurosci (2013) 34:125–136
DOI 10.1007/s10827-012-0410-z

Inhibition enhances memory capacity: optimal feedback, transient replay and oscillations

Axel Kammerer · Álvaro Tejero-Cantero · Christian Leibold

Received: 17 February 2012 / Revised: 18 May 2012 / Accepted: 21 June 2012 / Published online: 11 July 2012

Abstract Recurring sequences of neuronal activation in the hippocampus are a candidate for a neurophysiological correlate of episodic memory. Here, we discuss a mean-field theory for such spike sequences in phase space and show how they become unstable when the neuronal network operates at maximum memory capacity. We find that inhibitory feedback rescues replay of the sequences, giving rise to oscillations and thereby enhancing the network's capacity. We further argue that transient sequences in an overloaded network with feedback inhibition may provide a mechanistic picture of memory-related neuronal activity during hippocampal sharp-wave ripple complexes.

Keywords Associative memory · Sequence memory · Memory capacity · Sequence replay · Inhibition · Neuronal network dynamics · Learning and plasticity · Sharp-wave ripples · Oscillations

Action Editor: David Golomb

Axel Kammerer and Álvaro Tejero-Cantero contributed equally to this paper.

A. Kammerer (✉) · Á. Tejero-Cantero · C. Leibold
Department Biologie II, Ludwig-Maximilians University Munich, Großhadernerstrasse 2, 82152 Planegg, Germany
e-mail: kammerer@bio.lmu.de

A. Kammerer · Á. Tejero-Cantero · C. Leibold
Graduate School for Systemic Neurosciences, Ludwig-Maximilians-University Munich, Großhaderner Str. 2, 82152 Planegg, Germany

Á. Tejero-Cantero · C. Leibold
Bernstein Center for Computational Neuroscience Munich, Großhaderner Str. 2, 82152 Planegg, Germany

1 Introduction

The hippocampus is a brain structure crucially involved in the formation of autobiographic, episodic memories. Electrophysiological recordings of hippocampal neurons in behaving rodents have revealed the existence of *place cells*, which are active at only a few particular locations in a known environment, and silent elsewhere (O'Keefe and Dostrovsky 1971). The response of a population of place cells hence encodes the position of an animal in an environment and a spatial trajectory of this animal is represented by a sequence of active place cells (e.g. Dragoi and Buzsáki 2006; Davidson et al. 2009).

During sleep and resting states, spontaneous hippocampal activity bursts have been observed in which neurons were activated in an order similar to exploratory phases (Wilson and McNaughton 1994; Nádasdy et al. 1999; Lee et al. 2002; Diba and Buzsáki 2007; Gupta et al. 2010; Dragoi and Tonegawa 2011). The activity sequences observed in slow-wave sleep mostly correspond to previously experienced trajectories, whereas under awake rest, they have been found to even code for future trajectories. Therefore, these network bursts have been hypothesized to reflect memories and imaginations of spatial episodes (for review, see Buhry et al. 2011).

The activity sequences are correlated with the occurrence of sharp-wave ripple complexes (Lee and Wilson 2002; Diba and Buzsáki 2007), a brief (∼100 ms) field potential deflection that is superimposed with a high-frequency oscillation in the hippocampal CA1 pyramidal cell layer (∼200 Hz; Buzsáki et al. 1992). Although it is still debated how these spontaneous sequences are related to memory in the psychological sense, they

constitute a fascinating example of a biophysical memory phenomenon realized by local network mechanisms (Csicsvari et al. 2000; Sullivan et al. 2011; Maier et al. 2011).

In this paper, we extend on a model of a sequence memory network using a dynamical systems approach. The model network is operated in a regime at which it can robustly reproduce all stored sequences. The summed length of all stored sequences is called the capacity of the network. It is known that neuronal inhibition can improve this capacity and the robustness of sequence retrieval (e.g. Tsodyks and Feigel'man 1988; Treves 1990; Hirase and Recce 1996). Here we show that, as the network is operated close to its capacity limit, inhibition gives rise to oscillations. Beyond maximum capacity, inhibition enables transient replay, i.e. the partial reproduction of stored sequences—just as observed during sharp-wave-ripple-associated replay in the hippocampus.

2 Model

We model neuronal sequence generation in a network of binary neurons with binary synapses (Willshaw et al. 1969; Golomb et al. 1990; Nadal 1991; Hirase and Recce 1996; Leibold and Kempter 2006). The network consists of a randomly connected network of N excitatory neurons $i = 1, \ldots, N$. The neurons are simplified as binary with state $x_i = 1$ if neuron i fires and state $x_i = 0$ if it is silent. In the hippocampus, neuronal firing during the spontaneous sequences is phase-locked to ripple oscillations (Maier et al. 2011) at a frequency around 200 Hz. We therefore formulate the dynamics in discrete time t indicative of the oscillation cycle. A model neuron i receives an input from neuron j if there is a connection by an active "synapse". Following Gibson and Robinson (1992), the synapses are described by two independent binary stochastic processes. One stochastic variable indicates the presence ($w_{ij} = 1$) or absence ($w_{ij} = 0$) of a morphological connection, with probability $\mathrm{prob}(w_{ij} = 1) = c_m$. The constant c_m thereby denotes morphological connectivity. The other stochastic variable s_{ij} describes the synaptic state, which will be used to store memories. In the potentiated state ($s_{ij} = 1$) a synapse translates a presynaptic spike into a postsynaptic potential whereas in the silent state ($s_{ij} = 0$) it does not influence the postsynaptic neuron. The model neuron i fires a spike at cycle $t + 1$ if the sum of its inputs $h_i(t)$ in the previous cycle t exceeds a threshold θ. In summary, the network dynamics is described by the equation $x_i(t+1) = \Theta\big[\sum_{j=1}^{N} w_{ij} s_{ij} x_j(t) - \theta\big]$, with Θ denoting the Heaviside step function.

The memories stored in the network are sequences of activity patterns $\boldsymbol{\xi}$ described by binary vectors of dimension N, $\boldsymbol{\xi} \in \{0, 1\}^N$. A memory sequence of length Q is an ordered occurrence of activity patterns $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_Q$ (Fig. 1). The number M of active neurons in each pattern is called pattern size, and is the same for all patterns. Memory sequences are stored in the synapses using the learning rule by Willshaw et al. (1969): A synapse is potentiated *only* if it connects two neurons that are activated in sequence at least once. Then the number P of stored associations between a cue and a target pattern is related to the fraction c/c_m of activated synapses by

$$P = \frac{\ln(1 - c/c_m)}{\ln(1 - f^2)} \,. \tag{1}$$

The coding ratio $f = M/N$ is the fraction of active neurons and fixes the firing rate. If all P stored associations can be replayed, the number $\alpha \equiv P/(N\, c_m)$ is called the capacity of the network, and counts the stored associations per number of synapses at a neuron. Note that the sequence length Q is generally much smaller than the total number P of associations; the number of retrievable sequences of length Q is given by $\lfloor P/(Q-1) \rfloor$. Whereas P reflects the combinatorics of the synaptic matrix (determined by N, M, c_m, c), the requirement of a minimum Q implies a stability constraint for the network as a dynamical system.

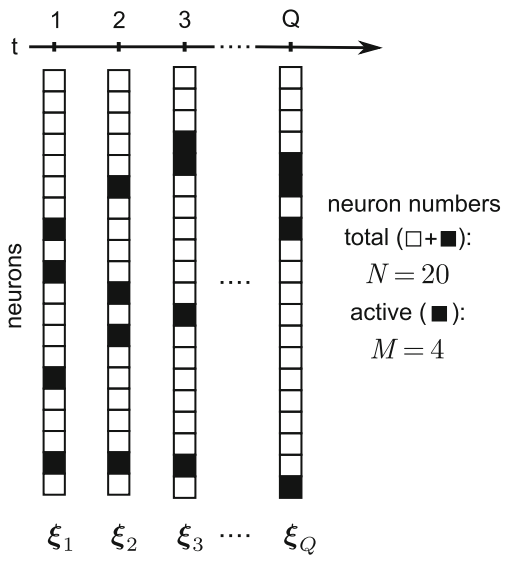

Fig. 1 Sequence of activity patterns. A network of N neurons (*boxes*) is considered in discrete time t. At each time step M neurons are active (*filled boxes*). A sequence that lasts for Q time steps is described by the binary vectors $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_Q$, where $(\xi_k)_i = 1$ if neuron i fires in pattern k and $(\xi_k)_i = 0$ if not. If P transitions between activity patterns are stored in the network, the number of stored sequences of length Q is $\lfloor P/(Q-1) \rfloor$

In the biologically relevant scenario of low coding ratio f, capacity grows like $\alpha \propto M^{-2}$ for fixed c, c_m. Maximum capacity thus corresponds to the minimum pattern size M_{opt} at which the last (Q-th) element of the sequence can still be recalled. The minimum pattern size has been shown to scale like $\ln N$ (Willshaw et al. 1969; Gardner 1987), and hence $\alpha \propto N/(\ln N)^2$.

Sequence retrieval is described by two macroscopic dynamical variables: the number $m_t \in [0, M]$ of correctly activated neurons (*hits*) and the number $n_t \in [0, N-M]$ of incorrectly activated neurons (*false alarms*). For large network sizes N and large pattern sizes M, we can assume Gaussian distributions for the number of inputs $h(t)$, and reinterpret the variables m and n in a mean-field sense as their respective expectation values over realizations of the connectivity matrix. The distributions of inputs are thus characterized by the means $\mu \equiv \langle h(t) \rangle$ and variances $\sigma^2 \equiv \langle h(t)^2 \rangle - \langle h(t) \rangle^2$; for "hit" neurons

$$\begin{aligned} \mu_{\text{On}} &= c_m\, m + c\, n\,, \\ \sigma^2_{\text{On}}(m,n) &= c_m\,(1-c_m)\, m \\ &\quad + c\left[(1-c) + c\,\text{CV}^2_{\text{q}}(n-1)\right] n, \end{aligned} \tag{2}$$

and for "false-alarm" neurons,

$$\begin{aligned} \mu_{\text{Off}} &= c\,(m+n)\,, \\ \sigma^2_{\text{Off}}(m,n) &= c\left[(1-c) + c\,\text{CV}^2_q\,(m+n-1)\right] \\ &\quad \times (m+n)\,. \end{aligned} \tag{3}$$

The terms proportional to

$$\text{CV}^2_q = (1-f^2)^P\, \frac{\left(1 - \frac{f^2}{1+f}\right)^P - (1-f^2)^P}{[1-(1-f^2)^P]^2} \tag{4}$$

originate from correlations in the synaptic states that are induced by Willshaw's learning rule and are computed in the Appendix following Gibson and Robinson (1992).

The network dynamics is then implemented as an iterated map:

$$(m_{t+1}, n_{t+1}) = [T_{\text{On}}(m_t, n_t),\, T_{\text{Off}}(m_t, n_t)]\,. \tag{5}$$

Since only those neurons fire whose input $h(t)$ exceeds the threshold θ, we estimate the expectation values of hits m_{t+1} and false alarms n_{t+1} from the cumulative distribution function (cdf) of the normal distribution, $\Phi(z) \equiv [1 + \text{erf}\,(z/\sqrt{2})]/2$,

$$\begin{aligned} T_{\text{On}}(m,n) &= M\,\Phi[(\mu_{\text{On}} - \theta)/\sigma_{\text{On}}] \\ T_{\text{Off}}(m,n) &= (N-M)\,\Phi[(\mu_{\text{Off}} - \theta)/\sigma_{\text{Off}}]\,. \end{aligned} \tag{6}$$

The nullclines (e.g. $n - T_{\text{On}}(n,m) = 0$) of this dynamical system are shown in Fig. 2(A) for a case of stable retrieval, i.e., there exists an asymptotically stable fixed point $(m_\infty, n_\infty) = [T_{\text{On}}(m_\infty, n_\infty),\, T_{\text{Off}}(m_\infty, n_\infty)]$ with many hits $m_\infty \simeq M$ and few false alarms $n_\infty \ll N - M$.

If the firing threshold is too low or the pattern size is too large (Fig. 2(C)), the nullclines do not cross in a retrieval regime: After initialization at the condition of perfect retrieval $(m_0, n_0) = (M, 0)$, all neurons immediately start to fire and the network falls into an all-active state, $(m,n) \simeq (M, N-M)$. If the firing threshold is too high or the pattern size is too low (Fig. 2(B)) only an unstable fixed point exists in the retrieval region. After initialization at perfect retrieval, the network immediately falls into an all-silent state $(m,n) \simeq (0,0)$.

The phase diagram reveals the three phases of our model sequence memory network (Fig. 2(D)): all silent, all active, and retrieval. The region in which retrieval is possible is wedge-shaped with a thin tip at low pattern sizes M. It turns out that the dynamics usually converges to the fixed points in only a few iterations, meaning that if sequence retrieval is stable for some finite length $Q \gtrapprox 10$, it is likely to be stable for $Q \to \infty$ (see thin grey area in Fig. 2(D) which indicates transient replay between 4 and 99 iterations). Note that, technically, $Q \leq P$, and thus the limit $Q \to \infty$ should be interpreted as having stored a cyclic sequence with $\boldsymbol{\xi}_{Q+1} = \boldsymbol{\xi}_1$.

The region of retrieval obtained from the mean-field equations can be validated with computer simulations of the corresponding networks of binary neurons. As expected, owing to the finite size of the simulated network, the region of retrieval is overestimated by mean-field theory, yet the deviations are relatively small (white discs in Fig. 2(D)). According to Eq. (1), the number P of stored associations increases with decreasing coding ratio $f = M/N$, and thus the network obtains the highest memory capacity at the wedge tip $M = M_{\text{opt}}$. There the stability of the fixed point is particularly sensitive to noise and thus the high capacity is not accessible unless the dynamics can be stabilized. A natural way to stabilize replay is to include feedback inhibition (see Section 3).

2.1 Optimal firing threshold

A different view on the task of the excitatory neurons during sequence replay is that of an optimal detector:

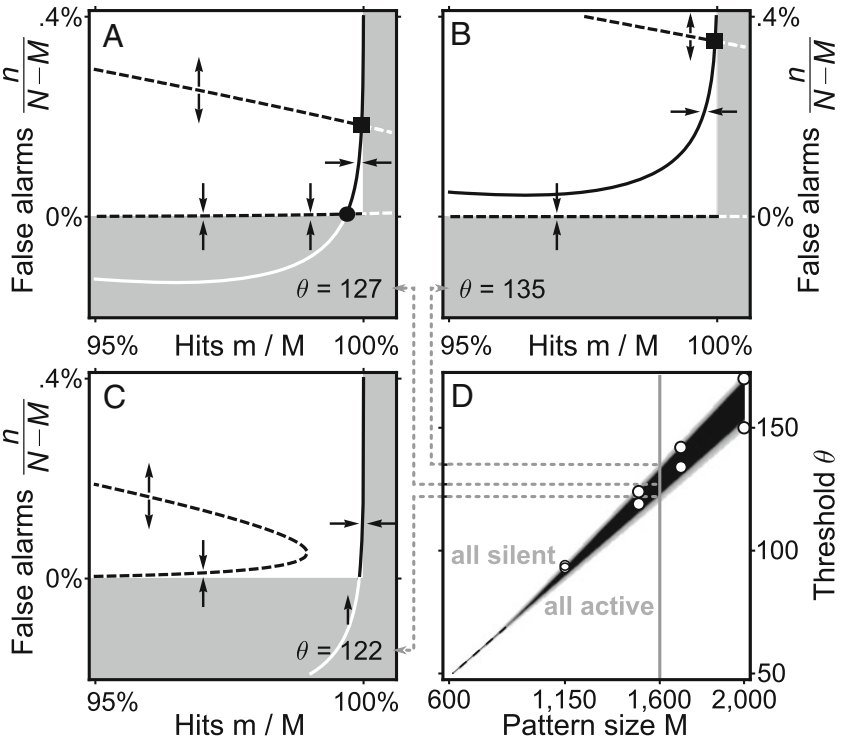

Fig. 2 Phase space and phase diagram. (**A**) Phase space is spanned by the numbers m of hits and n of false alarms. The m nullcline (*solid line*) intersects twice with the n nullcline (*dashed*), producing stable (*disc*) and unstable (*square*) fixed points. *Arrows* indicate attractive or repulsive character of the nullcline; *gray* areas correspond to unphysical values $n < 0$ or $m > M$. (**B**) Same as A for higher threshold θ. Only the unstable fixed point remains. (**C**) Same as A for lower θ. Both fixed points disappear. (**D**) Phase diagram. M-θ space consists of three areas: All silent; the sequence dies out. All active; all neurons fire at maximum rate. Sequence retrieval (*black*); the fraction of hits is much larger than the fraction of false alarms for infinitely many time steps (here tested as $m_t/M > 0.9, n_t/(N-M) < 0.1$ for $t \leq 100$). The *dashed line* separates the all-silent and all-active phases for M-values at which no retrieval phase exists. Areas in *light gray* correspond to transient retrieval of at least 4 time steps. *White discs* mark the boundary of the retrieval region as obtained from simulations of N binary neurons for exemplary values of M. Parameters here and elsewhere are $N = 10^5$, $M = 1{,}600$, $c_m = 0.1$, and $c = 0.05$, unless specified otherwise

the detector neuron is supposed to fire if it belongs to the hit population $\mathcal{A}_{\text{On}}$ of the current time step, or not, in which case it belongs to the false-alarm-population $\mathcal{A}_{\text{Off}}$. The prior probabilities for a neuron to belong to either of these populations, $\Pr(\mathcal{A}_{\text{On}}) = f$ and $\Pr(\mathcal{A}_{\text{Off}}) = 1 - f$, are given by the coding ratio f, which stipulates how many active neurons code for a pattern at any one time step. The basis for the decision whether to fire or not is a one-shot sample from the distributions of synaptic input levels. We again approximate these distributions as Gaussians whose mean and variance depend on whether the detector neuron is target in a pattern or not.

A Bayesian strategy to solve this problem ideally seeks to maximize the *probability of success* S, i.e. the probability of taking the right decision:

$$S = \Pr\left(\text{spike}|\mathcal{A}_{\text{On}}\right)\Pr(\mathcal{A}_{\text{On}}) + \Pr(\text{silence}|\mathcal{A}_{\text{Off}})\Pr(\mathcal{A}_{\text{Off}})\,.$$

Given the spike generation model (spike $\equiv h \geqslant \theta$), the conditional probabilities of spike or silence correspond to integrals of the respective probability densities over regions of synaptic input separated by the threshold θ:

$$S(\theta) = \Pr(h \geqslant \theta|\mathcal{A}_{\text{On}})\Pr(\mathcal{A}_{\text{On}}) + \Pr(h < \theta|\mathcal{A}_{\text{Off}})\Pr(\mathcal{A}_{\text{Off}})\,.$$

The mean-field considerations leading to Eq. (6) allow to rewrite the success probability in terms of Gaussian cdfs:

$$S(\theta) = \Phi(z_{\text{On}})\ f + (1 - \Phi(z_{\text{Off}}))\ (1 - f)\ ,$$

with

$$z_{\text{On/Off}} = \frac{\mu_{\text{On/Off}} - \theta}{\sigma_{\text{On/Off}}}\,.$$

The threshold θ_{opt} that maximizes the success probability can be readily obtained by demanding

$\mathrm{d}S(\theta)/\mathrm{d}\theta = 0$. Since $\partial\Phi(z)/\partial z = \mathrm{e}^{-z^2/2}/\left(\sqrt{2\pi}\sigma\right)$ we have

$$\frac{\mathrm{d}S}{\mathrm{d}\theta} = \frac{1}{\sqrt{2\pi}}\left(-f\frac{\mathrm{e}^{-z_{\mathrm{On}}^2/2}}{\sigma_{\mathrm{On}}} + (1-f)\frac{\mathrm{e}^{-z_{\mathrm{Off}}^2/2}}{\sigma_{\mathrm{Off}}}\right),$$

i.e. the optimal threshold is at the crossing point of the weighted Gaussians. The resulting equation is quadratic in θ,

$$z_{\mathrm{Off}}^2 - z_{\mathrm{On}}^2 = 2\log\left(\frac{1-f}{f}\frac{\sigma_{\mathrm{On}}}{\sigma_{\mathrm{Off}}}\right),$$

and has roots

$$\theta_{\pm} = \left(\sigma_{\mathrm{On}}^2 - \sigma_{\mathrm{Off}}^2\right)^{-1}\left[c(m+n)\sigma_{\mathrm{On}}^2 - (c_m m + cn)\sigma_{\mathrm{Off}}^2 \pm \sigma_{\mathrm{On}}\sigma_{\mathrm{Off}}\sqrt{(c-c_m)^2 m^2 + \left(\sigma_{\mathrm{Off}}^2 - \sigma_{\mathrm{On}}^2\right)\log\left(\frac{M^2}{F^2}\frac{\sigma_{\mathrm{Off}}^2}{\sigma_{\mathrm{On}}^2}\right)}\right]. \tag{7}$$

Generally, one of the thresholds is positive and the other negative when $m \simeq M, n \ll N - M$, which enables heuristic identification of the sign leading to maximization of $S(\theta)$.

2.2 Optimal threshold adaptation

Foreshadowing our interest in adaptive regulation of the threshold, we ask how the threshold should change with the excitatory activities m and n. Figure 3 displays the optimal threshold $\theta_{\mathrm{opt}}(m,n)$ from Eq. (7): In the phase-space region of retrieval (large m, small n), the level curves of $\theta_{\mathrm{opt}}(m,n)$ can be very well approximated by a linear function, which, using Taylor expansion, is

$$\theta_{\mathrm{opt}}(m,n) = \theta_{\mathrm{opt}}(M,0) + \partial_m\theta_{\mathrm{opt}}(M,0)(m-M) + \partial_n\theta_{\mathrm{opt}}(M,0)\,n\,.$$

Figure 3(A) thus demonstrates that the optimal threshold derived from analytic optimality considerations linearly increases with activity levels m and n as has been found numerically by Hirase and Recce (1996).

The coupling coefficients $\partial_m\theta_{\mathrm{opt}}$ and $\partial_n\theta_{\mathrm{opt}}$ are plotted in Figure 3(B). Both partial derivatives have positive values, which, as intuitively expected, corresponds to an increase in threshold for growing activity. Two things should be noted further: First, the two coefficients depend on M only little (at least for $M >$ 880 at which replay is stable; see Figure 2). Second, they are of similar value (between c and $2c$ in our example network). Together this indicates that the optimal adapting threshold may be approximately realized by a single multiplicative coupling constant $b \approx \partial_m\theta_{\mathrm{opt}} \approx \partial_n\theta_{\mathrm{opt}}$ that is the same for both hits m_t and false alarms n_t (see Section 3.1).

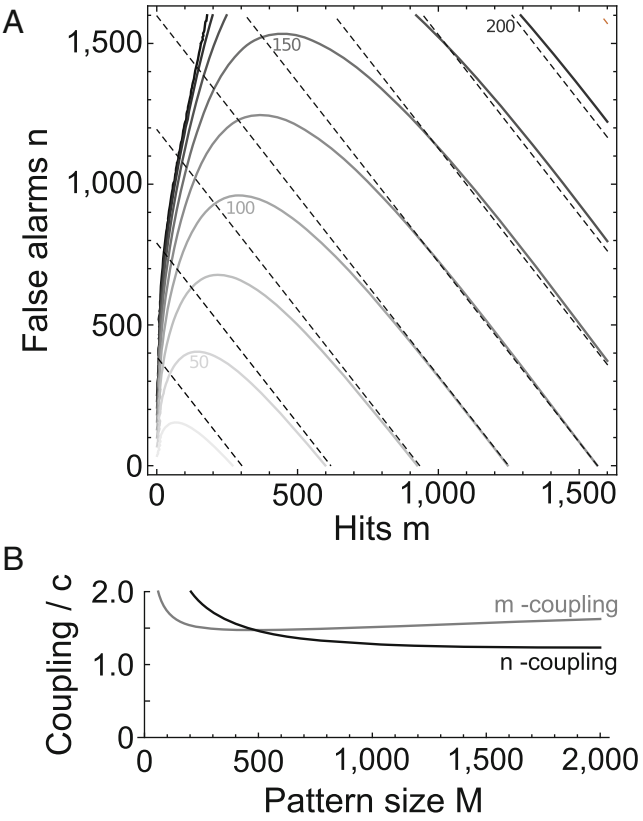

Fig. 3 Optimal threshold is linear in activity. (**A**) Level curves (*grey*) of the optimal threshold as a function of the numbers of hits m and false-alarms n. In this exemplary network with $c = 0.05$, $c_m = 0.1$, $N = 10^5$ and $M = 1{,}600$, the linear approximation (*dashed*; $\theta(m,n) \simeq 1.118 + 0.079\,m + 0.062\,n$) is very good in the retrieval region ($m \approx M, n/N \approx 0$) of the phase plane. (**B**) Coefficients $\partial_{m/n}\theta_{\mathrm{opt}}$ that couple the optimal detector threshold θ_{opt} and excitatory population activity of On (*grey*) and Off cells (*black*)

3 Role of inhibition

It has been shown previously that adaptive thresholds can be interpreted as instantaneous global feedback inhibition and can improve the capacity of associative memory networks (Golomb et al. 1990; Treves 1990; Hirase and Recce 1996). We therefore have investigated the effect of inhibition with respect to its phase-space behavior in our model. First, we consider inhibition to provide an instantaneous negative feedback. Second, and unlike previous approaches, we treat global inhibition as an additional dynamical variable.

3.1 Instantaneous global inhibition

Motivated by previous results on optimal thresholds (Section 2.2 and Hirase and Recce 1996), we introduce an instantaneous negative feedback proportional to the total number $m+n$ of active neurons. The dynamics is derived from Eq. (6) by substituting $\theta \to \theta + b\ (m_t + n_t)$, where the positive b acts as a feedback gain.

The main effect of inhibition is as follows. When the threshold θ is too low (as in Fig. 2(C)), inhibition moves the n-nullcline rightward; when θ is too high (as in Fig. 2(B)), inhibition moves the m-nullcline downward. Finally, in cases for which M is below the optimal pattern size M_{opt} of the purely excitatory model, and no threshold exists for which replay is stable (as in Fig. 4(A)), inhibition moves both nullclines at the same time. Thus, inhibition restores the stable fixed point and therefore effectively enlarges the retrieval phase (Fig. 4(B)). In particular, inhibition lowers the optimal pattern size M_{opt}, thereby enhancing memory capacity $\alpha \propto M_{\text{opt}}^{-2}$ (by a factor of about 2 in the example of Fig. 4(C)). Interestingly, the optimal range for the feedback gain ($b \gtrsim c$) fits well to that for the Bayes-optimal threshold in Fig. 3(B). For such optimal values of b the lower border of the wedge becomes roughly horizontal and the threshold θ is close to zero (not shown). Physiologically, the feedback gain b may be adjusted into this range by plasticity of inhibitory synapses.

To investigate the scaling behavior of the memory capacity, we determined the minimum pattern size M_{opt} for different network sizes N and found the well-known logarithmic dependence $M_{\text{opt}} \propto \ln N$ regardless of the inhibitory gain b (Fig. 4(D)). The capacity thus still grows with network size as $\alpha \propto N/(\ln N)^2$.

An alternative view on instantaneous global feedback inhibition can be derived from the mean values in Eqs. (2) and (3), viz., the substitution $\theta \to \theta + b\ (m_t + n_t)$ effectively reduces the connectivities c_m, and c to $c_m - b$ and $c - b$. The ratio $r = (c_m - c)/c$ between non-activated and activated synapses can be interpreted as the plasticity resources of the network and was shown in Leibold and Kempter (2006) to critically define the signal-to-noise ratio at the post-synaptic neurons (with maximal capacity at $r \approx 10$ for fixed N, M, c_m). By substituting $c_m \to c_m - b$ and $c \to c - b$, instantaneous global inhibition with $b < c$ can be formally interpreted to increase the signal-to-noise ratio like $r \to r/(1 - b/c)$. Since, in the present paper, we initially assume $r = 1$, a feedback gain of $b > 0$ thus generally increases the signal-to-noise ratio for fixed c and thereby enhances capacity (Fig. 4(C)).

3.2 Dynamic global inhibition

We next asked whether the effects observed with instantaneous global inhibition can be reproduced in a more physiological scenario in which inhibitory activity

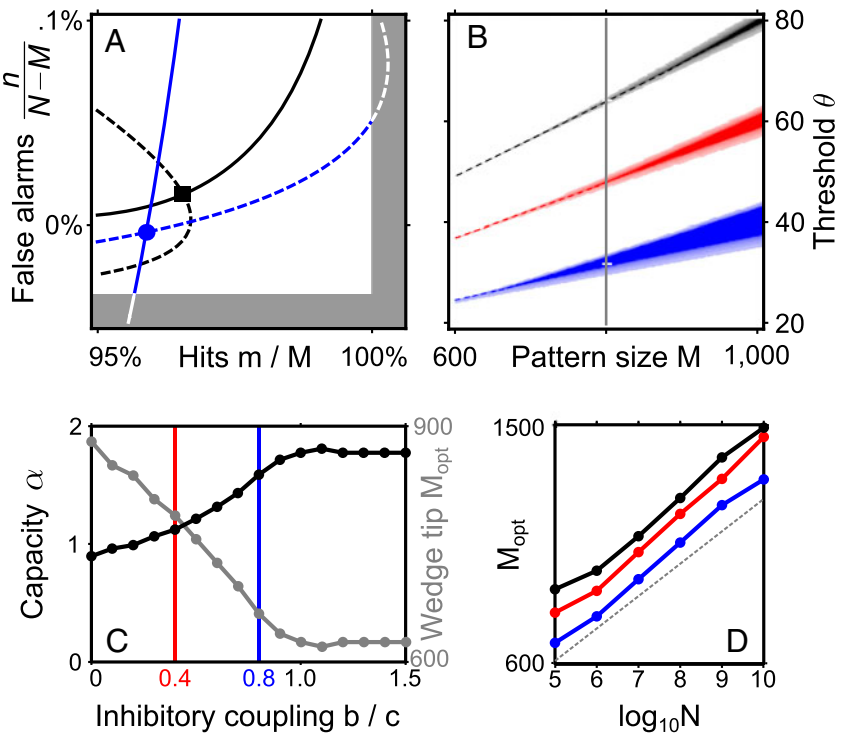

Fig. 4 Retrieval with instantaneous global inhibition $b = 0$ (*black*), $b = 0.4\,c$ (*red*), and $b = 0.8\,c$ (*blue*). (**A**) Nullclines (*red* omitted). Firing threshold θ for the *blue* nullclines is offset by $-b$ M to account for lower mean input h. (**B**) Phase diagram. *Grey vertical line* and *white dashes* indicates M, θ values used in A. *Light colors* show transient retrieval of at least 4 time steps. (**C**) Capacity (*black*) and minimum pattern size M_{opt} (*gray*) as a function of b. (**D**) The minimum pattern size grows sublinearly with N so that the capacity shows an overall increase. The *dashed grey line* indicates a logarithmic dependence $M \propto \ln N$

has its own dynamics, and what additional features such inhibitory dynamics would give rise to. To this end, we extended the model from Eqs. (5) and (6) by including a third dynamical variable k_t that accounts for the number of spikes in an inhibitory pool of K neurons. Each neuron in this pool is assumed to project to and to receive input from excitatory neurons with probabilities c_{IE} and c_{EI}, respectively. Analogous to Eq. (5), the dynamics of k_t is implemented as the map $k_{t+1} = T_{\mathrm{Inh}}(m_t, n_t)$ with

$$T_{\mathrm{Inh}}(m,n) = K\,\Phi[(\mu_{\mathrm{Inh}} - \eta)/\sigma_{\mathrm{Inh}}]. \tag{8}$$

The mean synaptic input and its variance are

$$\mu_{\mathrm{Inh}}(m,n) = c_{EI}\,w_{EI}\,(m+n) \tag{9}$$

$$\sigma^2_{\mathrm{Inh}}(m,n) = w^2_{EI}\,c_{EI}\,(1-c_{EI})\,(m+n). \tag{10}$$

The parameter w_{EI} denotes the synaptic weight of the connections from excitatory to inhibitory neurons. The inhibitory action on the sequence-related variables m and n is implemented by replacing the thresholds in Eq. (6) by $\theta \to \theta + w_{IE}\,c_{IE}\,k_t$, and the variances by $\sigma^2_{\mathrm{On/Off}}(m,n) \to \sigma^2_{\mathrm{On/Off}}(m,n) + k\,w^2_{IE}\,c_{IE}\,(1-c_{IE})$. Again, w_{IE} is the corresponding synaptic weight.

To test for sequence retrieval, the map is initialized with a perfect pattern and matching inhibition, $(m_0, n_0, k_0) = [M, 0, T_{\mathrm{Inh}}(M,0)]$. The resulting phase diagram reveals again regions of stable and transient retrieval (Fig. 5). In agreement with the linear instantaneous inhibition model, the retrieval region in the phase diagram extends to lower pattern sizes M (higher capacities). However, the non-linearity of the sigmoidal Gaussian cdf in Eq. (8) introduces a shearing of this region that can be explained as follows: The Gaussian cdf is roughly linear in the vicinity of the inhibitory threshold η and virtually flat elsewhere. Hence, as an approximation, inhibition has no effect at low total activities $m+n$, it adds a constant $w_{IE}\,c_{IE}\,K$ to the threshold θ at high total activities and establishes a nearly linear coupling for intermediate regimes, similar to the instantaneous-inhibition model from Figure 4. During sequence retrieval, total activity is approximately constant, $m+n \simeq M$, and therefore the retrieval region of the dynamic-inhibition model can be understood as a combination of three retrieval regions of the instantaneous-inhibition model for different feedback gains b and thresholds θ.

The broadening in θ of the region of retrieval (for constant M) with both instantaneous and dynamic inhibition suggests that sequence memory becomes more robust. Mean field theories, however, generally overestimate the regions of stable sequence retrieval (Latham and Nirenberg 2004). To assess the predictive power

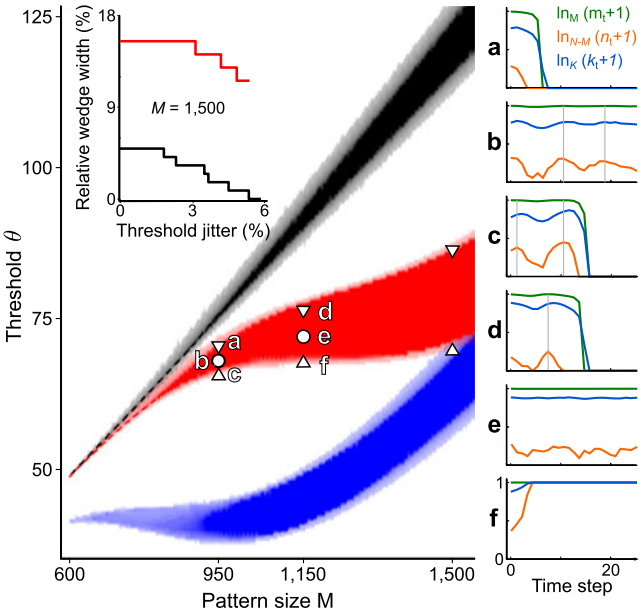

Fig. 5 Comparison of the retrieval regions in M-θ space for the 2-dimensional model without inhibition (*black*) vs. the 3-dimensional model with increasing dynamical feedback inhibition (*red, blue*). *Triangles* on the *red* region mark the first values of θ (in integer steps) for which sequence retrieval is unstable in simulations. *Light colors* show transient retrieval of at least 4 time steps. (**a**–**f**): Example trajectories from network simulations for M-θ pairs as indicated by white markers a–f (m_t: *green*; n_t: *orange*, k_t: *blue*). The *grey vertical lines* indicate Hilbert phase zero for the false-alarm neuron activity n_t. Parameters are $K = 5{,}000$, $w_{IE} = 0.012$, $c_{IE} = 1$, $w_{EI} = 1$, $c_{EI} = 0.01$, and $\eta = 13$ for the *red* region; $\eta = 8.8$ for the *blue* region. *Inset*: Robustness of sequence retrieval against threshold jitter with (*red*, $K = 5{,}000$) and without inhibition (*black*) at $M = 1{,}500$ for simulated networks with threshold noise

of our mean field results, we ran further simulations where neuronal thresholds θ were jittered according to a Gaussian process. The results show that the increase of the relative range of thresholds by inhibition indeed withstands threshold noise (Fig. 5 Inset). At high capacities, the demand of robustness against threshold noise implies that the area of retrieval should be broadest at minimum $M = M_{\mathrm{opt}}$.

We suggest two heuristic criteria for the parameters of dynamic inhibition. First, to achieve maximum sensitivity of the inhibitory feedback, the linear region of $T_{\mathrm{Inh}}(m,n)$ should be centered at the average total input $m+n \simeq M$ during retrieval. This requirement is granted by setting the inhibitory threshold to $\eta = \mu_{\mathrm{Inh}}(M,0)$. Second, the slope at this average total input should yield maximum capacity according to the instantaneous inhibition model (Figs. 3(B) and 4(C)), i.e., it should take a value of at least c. This requirement can be met by appropriately choosing the coupling factor $w_{IE}\,K$. The blue region in Fig. 5 illustrates the outcome

of such an optimization at $M = 880$ with an effective slope of $1.6\,c$ (the red region is obtained at $M = 1{,}300$ and slope $1.3\,c$). The region of stable retrieval is almost flat in M-θ space, suggesting that replay is robust against variability in pattern size M. To the left of the region of stable retrieval, we observe in lighter color a substantial region of transient sequences. Such large regions of transient retrieval only occur for slopes larger than c (not shown), which corresponds to the optimal gain factors for the threshold adaptation from Fig. 3(B). The minimum pattern size M_{opt} of stable retrieval, however, does not decrease further for slopes above c (as in Fig. 4(C)).

Simulations confirm the shape of the fundamental regimes all active, all silent, and retrieval predicted by the three-dimensional mean-field model. Figure 5(a–f) displays simulated trajectories (m_t, n_t, k_t) for typical situations. Interestingly, all-silent states can also sometimes be observed for low threshold values, where inhibition overcompensates the false alarms and transiently allows for sequence retrieval before the network falls back into silence (Fig. 5(c)).

In the retrieval phase the network typically exhibits oscillatory behavior (Fig. 5(b–d)) arising from the interplay between excitatory neurons m_t, n_t and inhibition k_t that manifests itself in oscillations of the two with the phase of inhibition slightly delayed (by about one timestep). The periods of these oscillations are about 5 to 10 time steps corresponding to gamma-range frequencies of 20 to 40 Hz, under our initial assumption that one time step corresponds to a ripple cycle of 5 ms. The oscillatory activity components are present during both transient (Fig. 5(c), (d)) and ongoing replay (Fig. 5(b)). We further analyzed the oscillations based on the inhibitory activities k_t during ongoing replay from cellular simulations (Fig. 6). As a measure for the oscillation amplitude we computed the standard deviation over time std(k/K), and found that it increases towards the edges of the region of replay. As a consequence, the oscillations are particularly strong at the low-M tip of the replay wedge, where the network realizes its maximum capacity. From this, we conclude that gamma oscillations herald the onset of dynamical instability as it is the case at the capacity limit.

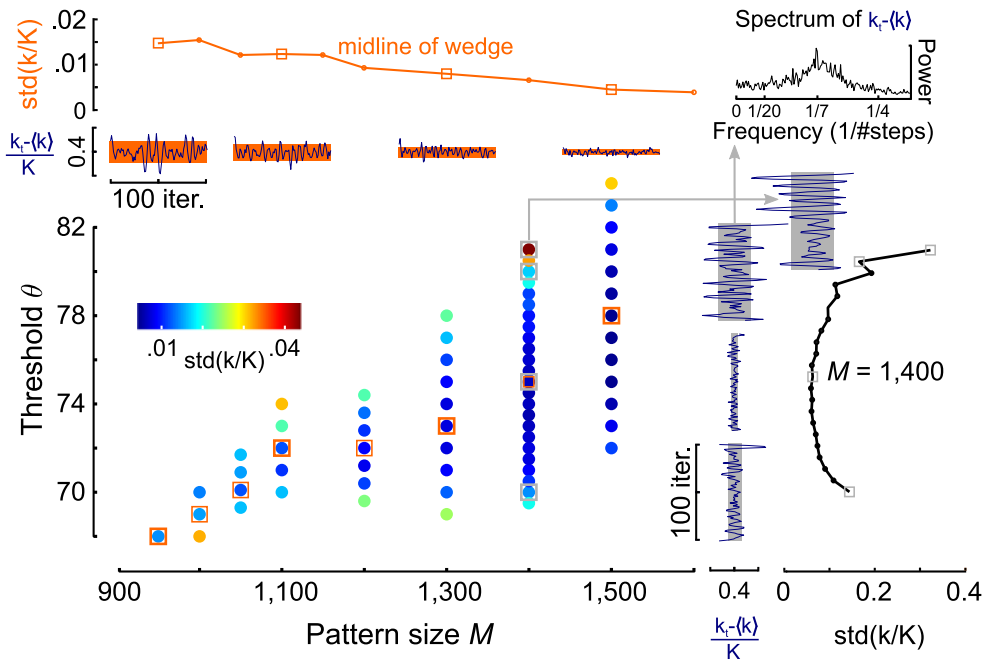

Fig. 6 Transition to instability is marked by increase in amplitude of gamma oscillations. At *lower left*, *colored discs* mark combinations of M and θ for which numerical simulations revealed stable retrieval. The standard deviation over time of the inhibitory activity k_t (normalized to K) is represented by the color code as indicated. At *top left*, oscillation amplitudes (measured as std(k/K) are shown for networks with M, θ along the midline of the wedge. Examples of k_t/K are given with std(k/K) as *orange bars*. At right, oscillation amplitudes std(k/K) are shown for networks of fixed $M = 1{,}400$ and different thresholds; the corresponding k_t/K are given with std(k/K) as *grey bars*. The *top right* panel shows an exemplary power spectrum of the inhibitory activation $k - \langle k \rangle$ for a simulation with M-θ values as indicated. The peak of the spectrum at around 1/7 per time step corresponds to 30 Hz for a time step of 5 ms

4 Discussion

This paper presents a dynamical-systems extension of time-discrete sequence memory models with inhibitory feedback. The effect of instantaneous global feedback inhibition in memory networks has been well studied (e.g. Tsodyks and Feigel'man 1988; Golomb et al. 1990; Treves 1990; Hirase and Recce 1996; Amit and Huang 2010). Our model shows that also dynamical feedback inhibition can stabilize the retrieval of memory sequences and thereby increase both memory capacity and robustness. The optimal instantaneous global inhibitory feedback is a roughly linear function of the total network activity as numerically found by Hirase and Recce (1996) and semi-analytically confirmed by a probabilistic approach in the present paper. Extending the model to dynamic global inhibition, we find that, at the edges of stable replay, inhibition induces strong oscillations, which can be interpreted as gamma oscillations.

Gamma oscillations are ubiquitous in the brain and their origin is generally attributed to local inhibitory networks (Wang 2010). Several cognitive functions have been related to increased gamma power and coherence, such as sensory integration, attention, and memory (Jutras and Buffalo 2010). Specifically, gamma coherence between subregions in the hippocampal formation and prefrontal cortex has been shown to correlate with involvement in a short-term memory task (Sigurdsson et al. 2010). This finding fits well into the general view of gamma rhythms as a mechanism that facilitates communication between brain areas (e.g. Colgin 2011). In our model, gamma occurs as a side effect of feedback stabilization during replay. In combination with these findings, our model suggests that memory networks may have to be critically loaded to be able to transfer information to other brain areas.

Our model also reveals parameter regions in which transient retrieval occurs that lasts for only a few time steps. These regions of transient retrieval (light color in Fig. 5) extend far into low pattern sizes M for strong inhibitory feedback, and thus correspond to the regimes of largest memory capacity. Neuronal networks exhibiting activity sequences hence operate with optimal memory performance if they are in a regime of hyperexcitability that is stabilized by delayed inhibition. This transient retrieval regime is consistent with the dynamic features of sequence replay during sharp wave ripple complexes in the hippocampus, which typically extends over 5 to 10 cycles of an approximately 200 Hz oscillation, and that are accompanied by delayed inhibitory feedback (Maier et al. 2011).

In large environments, sequence replay *in vivo* can span several ripple episodes (Davidson et al. 2009), showing that long sequences can be constructed by concatenating multiple transient replay episodes. Our model argues that such fragmentation solves the dilemma between stability and capacity by not having to trade capacity for stability. Instead it uses dynamic feedback inhibition to break sequence replay into short stable fragments of transient replay. It remains open though, how information transfer between these fragments is realized and whether it occurs intra- or extrahippocampally.

Throughout the paper, we consider the connectivity parameters c and c_m as constants, based on the assumption that the morphological connectivity c_m is mainly determined by geometrical constraints such as the size of the cell surface, or the volume requirement of wiring. The functional connectivity c is assumed to result from the specific learning rules that ensure that the network always remains plastic: In order to store new memories a large fraction of synapses has to be able to change its state. In the parameter regime used for our analysis, this requirement is fulfilled by fixing $c/c_m = 0.5$. Moreover, the connectivities employed are small, since experiments indicate that hippocampal networks are sparsely connected (Miles and Wong 1986).

Limitations of our model arise from specific assumptions underlying our analysis. One of them is that of a constant pattern size M. In reality pattern sizes may be variable (as discussed for a different learning rule in Amit and Huang 2010), leading to a decreased capacity. Another significant simplification of our model is the discreteness in time. Dynamical interactions of synaptic currents and membrane processes during sharp-wave ripples may also reduce capacities. In this sense the capacity values derived in this paper can only be considered as upper bounds and for determining scaling behavior.

Extending the model to more realistic dynamics is necessary to investigate how close to the upper bound of capacity a real spiking network can get. Such a translation to biophysically more realistic neuron models, however, raises difficult problems. The massive bombardment by synaptic inputs (specifically inhibition) sets the cells into a high-conductance regime, in which the effective time constants become short and spike generation very sensitively depends on the timing of inputs. Further, the interplay between excitation and inhibition not only has to keep the cell in a balanced state in which spiking is sparse, but also has to ensure synchrony of spiking in the time slots of roughly 5 ms.

Other models of hippocampal sharp-wave ripples focused on dynamic features of sharp-wave ripples as a network phenomenon mostly disregarding functional aspects of sequence replay. In the model of

Memmesheimer (2010) sharp waves arise from spontaneous bursts of excitatory activity that are shaped by non-linear dendritic integration. Such a scenario requires a relatively high level of background activity (high n) and it is not yet clear how well this can work together with sequence replay at high memory capacities, where false alarms n are not desired. In another model by Vladimirov et al. (2012) synaptic integration plays no role in evoking action potentials. Spiking is propagated across axons by axo-axonal gap junctions (Schmitz et al. 2001). Also in this model the relation of these axonal spike patterns to memory-related processes has not been evaluated. Moreover, it's unclear how inhibition could physiologically be realized in such a scenario. We thus conclude that, despite these considerable efforts, we still lack a model of sharp wave ripples that combines realistic physiological phenomenology with functional hypotheses of the hippocampal memory networks.

Acknowledgements The authors are grateful to Alexander Mathis for comments on a previous version of the manuscript. This work was funded by the German Ministry for Education and Research (BMBF) under grant number 01GQ0440 (BCCN Munich) and the German Science Foundation (DFG) under grant numbers Le-2250/2 and GRK 1190.

Appendix: First and second moments

The dynamics underlying neuronal activity sequences is formulated as a two-dimensional iterated map in Eqs. (2)–(6). This time-discrete dynamics is simplified using Gaussian approximations for the distributions of the number h of synaptic inputs to a specified neuron. The Gaussian approximation therefore requires expressions for the means and variances of the input sums h.

Inputs can be of two kinds, hits m and false alarms n. The input sum $h = \sum_{j=1}^{m+n} w_j s_j$ thus runs over all $m+n \leq N$ active (firing) neurons in the network and depends on two binary random variables for each potential input: $w \in \{0,1\}$ indicating the presence of a synaptic connection, and $s \in \{0,1\}$ indicating its state (Gibson and Robinson 1992). The stochasticity of s is inherited from the randomness of the activity patterns underlying the memory sequences via Willshaw's learning rule.

The distribution of w is given by the morphological connectivity such that $\mathrm{prob}(w=1) = c_m$. The probability $\mathrm{prob}(s=1)$ of a synapse having been potentiated depends on whether it connects or not neurons that should fire in sequence at the particular point in time.

The Willshaw rule ensures that synapses that connect sequentially firing neurons are in the potentiated state, i.e. $\mathrm{prob}(s=1)=1$, and thus for this subset of synapses the input sum depends on a binomial process with probability $\mathrm{prob}(w=1) = c_m$.

For the other synapses, the probability $\mathrm{prob}(s=1) = q_x$ depends on the the number $x \leq P$ of associations the specific postsynaptic neuron is involved in. Note that if the postsynaptic neuron is never supposed to fire, the Willshaw rule will activate none of its synapses and thus $q_0 = 0$. In general, the probability that a neuron is not a target in one specific step of the sequence (association) is $1-f$, and thus the probability that it is not a target in any one of x associations is $(1-f)^x$. Conversely, the probability of such a synapse being potentiated is $q_x = 1-(1-f)^x$. Hence, assuming independence of the two binomial processes, the input sum h for this subset of synapses is binomial with probability

$$\mathrm{prob}(w_i s_i = 1) = c_m\, q_x \,. \tag{11}$$

The probability distribution of the input h can then be determined as

$$p(h) = \sum_{x=0}^{P} p(h|x)\; p(x) \,, \tag{12}$$

in which the conditional probability $p(h|x) = \binom{m+n}{h} (c_m\; q_x)^h\, (1 - c_m\; q_x)^{m+n-h}$ is derived from Eq. (11), and the probability $p(x)$ that a neuron is involved in x associations is also binomial, viz. $p(x) = \binom{P}{x} f^x\, (1-f)^{P-x}$.

To compute expected values of h, we have to discern between neurons that should be active at time step $t+1$ (and are supposed to generate the hits) and those that should be silent (and potentially give rise to false alarms). For the potential false alarms, we obtain

$$\begin{aligned}
\langle h\rangle_{\mathrm{Off}} &= \sum_{h=0}^{m+n} h \sum_{x=0}^{P} p(h|x)\; p(x) = \sum_{x=0}^{P} p(x) \sum_{h=0}^{m+n} h\; p(h|x) \\
&= \sum_{x=0}^{P} p(x)\; (m+n)\; (c_m\, q_x) \\
&= (m+n)\, c_m \sum_{x=0}^{P} [1-(1-f)^x] \binom{P}{x} f^x (1-f)^{P-x} \\
&= (m+n)\, c_m \left[1 - (1-f)^P \sum_{x=0}^{P} \binom{P}{x} f^x \right] \\
&= (m+n)\, c_m \left[1 - (1-f)^P\, (1+f)^P \right] \\
&= (m+n)\, c \,.
\end{aligned}$$

Note that the last step makes use of the capacity of the Willshaw rule, Eq. (1). Similarly, for the potential hits, we obtain

$$\langle h\rangle_{\mathrm{On}} = \sum_{h'=0}^{n} h' \sum_{x=0}^{P} p(h'|x)\, p(x) + \sum_{h''=0}^{m} h''\, p(h'') = n\,c + m\,c_m\ .$$

Here the expected value sums over two independent subsets of neurons, the first one (h') representing the false alarms, and the second (h'') representing the hits during the previous time step.

The corresponding variances can be obtained analogously employing the formula of the geometric series several times, and introducing the abbreviation $\mathrm{CV}_q^2 = \mathrm{var}_x\, q/\langle q\rangle_x^2$ with expected values according to the distribution $p(x)$:

$$\sigma^2_{\mathrm{On}}(m,n) = c_m\, m\,(1-c_m) + n\,c\left[(1-c) + c\,\mathrm{CV}_q^2\,(n-1)\right]$$

$$\sigma^2_{\mathrm{Off}}(m,n) = (m+n)\,c \times\left[(1-c) + c\,\mathrm{CV}_q^2\,(m+n-1)\right].$$

Note that $\mathrm{CV}_q \to 0$ for $f \to 0$, and, in this limit, the variance formulas $\sigma^2_{\mathrm{On}} \to m\,c_m\,(1-c_m) + n\,c\,(1-c)$, $\sigma^2_{\mathrm{Off}} \to (m+n)\,c\,(1-c)$ from the present theory approximate those in Leibold and Kempter (2006).

References

Amit, Y., & Huang, Y. (2010). Precise capacity analysis in binary networks with multiple coding level inputs. *Neural Computation, 22,* 660–688.

Buhry, L., Azizi, A. H., & Cheng, S. (2011). Reactivation, replay, and preplay: How it might all fit together. *Neural Plasticity, 2011*, 203462.

Buzsáki, G., Horvath, Z., Urioste, R., Hetke, J., & Wise, K. (1992). High-frequency network oscillation in the hippocampus. *Science, 256*, 1025–1027.

Colgin, L. L. (2011). Oscillations and hippocampal-prefrontal synchrony. *Current Opinion in Neurobiology, 21*, 467–474.

Csicsvari, J., Hirase, H., Mamiya, A., & Buzsáki, G. (2000) Ensemble patterns of hippocampal CA3-CA1 neurons during sharp wave-associated population events. *Neuron, 28*, 585–594.

Davidson, T. J., Kloosterman, F., & Wilson, M. A. (2009). Hippocampal replay of extended experience. *Neuron, 63*, 497–507.

Dragoi, G., & Tonegawa, S. (2011). Preplay of future place cell sequences by hippocampal cellular assemblies. *Nature, 469*, 397–401.

Diba, K., & Buzsáki, G (2007). Forward and reverse hippocampal place-cell sequences during ripples. *Nature Neuroscience, 10*, 1241–1242.

Dragoi, G., & Buzsáki, G. (2006). Temporal encoding of place sequences by hippocampal cell assemblies. *Neuron, 50*, 145–157.

Gardner, E. (1987). Maximum storage capacity in neural networks. *Europhysics letters, 4*, 481–485.

Gibson, W. G., & Robinson, J. (1992). Statistical analysis of the dynamics of a sparse associative memory. *Neural Networks, 5*, 645–661.

Golomb, D., Rubin, N., & Sompolinsky, H. (1990). Willshaw model: Associative memory with sparse coding and low firing rates. *Physical Review. A, 41*, 1843–1854.

Gupta, A. S., van der Meer, M. A., Touretzky, D. S., & Redish, A. D. (2010). Hippocampal replay is not a simple function of experience. *Neuron, 65*, 695–705.

Hirase, H., & Recce, M. (1996). A search for the optimal thresholding sequence in an associative memory. *Network, 7*, 741–756.

Jutras, M. J., & Buffalo, E. A. (2010). Synchronous neural activity and memory formation. *Current Opinion in Neurobiology, 20*, 150–155.

Latham, P. E., & Nirenberg, S. (2004). Computing and stability in cortical networks. *Neural Computation, 16*, 1385–1412.

Lee, A. K., & Wilson, M. A. (2002). Memory of sequential experience in the hippocampus during slow wave sleep. *Neuron, 36*, 1183–1194.

Leibold, C., & Kempter, R. (2006) Memory capacity for sequences in a recurrent network with biological constraints. *Neural Computation, 18*, 904–941.

Maier, N., Tejero-Cantero, Á., Dorrn, A., Winterer, J., Beed, P., Morris, G., et al. (2011). Coherent phasic excitation during hippocampal ripples. *Neuron, 72*, 137–152.

Memmesheimer, R. (2010). Quantitative prediction of intermittent high-frequency oscillations in neural networks with supralinear dendritic interactions. *Proceedings of the National Academy of Sciences of the United States of America, 107*, 11092–11097.

Miles, R., & Wong, R. K. (1986). Excitatory synaptic interactions between CA3 neurones in the guinea-pig hippocampus. *Journal of Physiology, 373*, 397–418.

Nadal, J.-P. (1991). Associative memory: On the (puzzling) sparse coding limit. *Journal of Physics. A. Mathematical and General, 24*, 1093–1101.

Nádasdy, Z., Hirase, H., Czurkó, A., Csicsvari, J., & Buzsáki, G. (1999). Replay and time compression of recurring spike sequences in the hippocampus. *Journal of Neuroscience, 21*, 9497–9507.

O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Research, 34*, 171–175.

Schmitz, D., Schuchmann, S., Fisahn, A., Draguhn, A., Buhl, E., Petrasch-Parwez, E., et al. (2001). Axo-axonal coupling: A novel mechanism for ultrafast neuronal communication. *Neuron, 31*, 831–840.

Sigurdsson, T., Stark, K. L., Karayiorgou, M., Gogos, J. A., & Gordon, J. A. (2010). Impaired hippocampal-prefrontal synchrony in a genetic mouse model of schizophrenia. *Nature, 464*, 763–767.

Sullivan, S., Csicsvari, J., Mizuseki, K., Montgomery, S., Diba, K., & Buzsáki, G. (2011). Relationships between hippocampal sharp waves, ripples, and fast gamma oscillation: Influence of dentate and entorhinal cortical activity. *Journal of Neuroscience, 31*, 8605–8616.

Vladimirov, N., Tu, Y., & Traub, R. D. (2012). Shortest loops are pacemakers in random networks of electrically coupled axons. *Frontiers in Computational Neuroscience, 6*, 17.

Treves, A. (1990). Graded-response neurons and information encodings in autoassociative memories. *Physical Review A, 42*, 2418–2430.

Tsodyks, M. V., & Feigel'man, M. V. (1988). Enhanced storage capacity in neural networks with low level of activity. *Europhysics Letters, 6*, 101–105.

Wang, X.-J. (2010). Neurophysiological and computational principles of cortical rhythms in cognition. *Physiological Reviews, 90*, 1195–1268.

Willshaw, D. J., Buneman, O. P., & Longuet-Higgins, H. C. (1969). Non-holographic associative memory. *Nature, 222*, 960–962.

Wilson, M. A., & McNaughton, B. L. (1994) Reactivation of hippocampal ensemble memories during sleep. *Science, 265*, 676–679.