



Technical Report No. 183

**The effect of pairwise neural
correlations on global population
statistics**

Jakob H. Macke¹, Manfred Opper²,
Matthias Bethge¹

March 2009

This Technical Report has been approved by:

Director at MPIK

Postdoc at MPIK



Technical Report No. 183

**The effect of pairwise neural
correlations on global population
statistics**

Jakob H. Macke¹, Manfred Opper²,
Matthias Bethge¹

March 2009

¹ MPI for Biological Cybernetics,
Computational Vision and Neuroscience Group
Spemannstraße 41
72076 Tübingen

² Technical University Berlin,
Artificial Intelligence Group
Franklinstraße 28/29
10587 Berlin,

The effect of pairwise neural correlations on global population statistics

Jakob H. Macke¹, Manfred Opper², Matthias Bethge¹

Abstract. Simultaneously recorded neurons often exhibit correlations in their spiking activity. These correlations shape the statistical structure of the population activity, and can lead to substantial redundancy across neurons. Here, we study the effect of pairwise correlations on the population spike count statistics and redundancy in populations of threshold-neurons in which response-correlations arise from correlated Gaussian inputs. We investigate the scaling of the redundancy as the population size is increased, and compare the asymptotic redundancy in our models to the corresponding maximum- and minimum entropy models.

1 Introduction

Finding suitable models for capturing the statistical structure of firing patterns distributed across multiple neurons is a major challenge in sensory neuroscience, and a prerequisite for understanding population codes. Recently, the Ising model [1], originally introduced by Ising in 1925 to understand ferromagnetism, has become popular for studying the statistical structure of neural activity recorded simultaneously from large populations of neurons [2, 3]. The use of the Ising model for neural data analysis originates from the fact that it constitutes the optimum with respect to the maximum entropy (MaxEnt) rationale [4]. These studies have raised the question of the importance of higher-order correlations in binarized spike trains, and how the joint entropy of neural populations scales with population size [5].

The Dichotomized Gaussian distribution (DG) constitutes an alternative model for correlated binary patterns [6]. In this model, binary patterns are thought of being generated by thresholding a multivariate Gaussian random variable, and correlations between neurons arise from the correlations in the underlying Gaussian. The DG can be interpreted as a phenomenological model of how correlations in the inputs to threshold-neurons shape the distribution of spike-counts in the population [7]. In contrast to maximum entropy models [8], the DG is easily fit even to large population sizes, and it is possible to generate exact samples via sampling from the latent Gaussian.

Bethge and Berens [9] used the DG to model the statistics of symmetrically thresholded natural images, and empirically found very small entropy-differences between the DG and Ising model on small images. However, the properties of the DG for large populations have not been characterized analytically yet. Here, we investigate whether this close match between the two models persist for large populations of neurons, and across all possible firing rates μ and pairwise correlation coefficients ρ . For simplicity, we focus on models with exchangeable correlation structure, i.e. with constant firing rates and pairwise correlations for all neurons.

We present a derivation of the spike-count distribution of the DG, from which one can calculate its entropy rate and characterize its phase-transitions. We generalize these results to models consisting of a finite number of homogeneous pools, with correlations both within and across pools. In addition, we show how the model can be generalized to common input models with arbitrary input distributions, and to neurons which are not necessarily simple threshold devices. Finally, we derive the minimum entropy distribution (MinEnt) for given first and second moments. The entropy difference between the MaxEnt and MinEnt models determine to what extent the entropy is constrained by pairwise correlations.

In this technical report, we aimed to present a very detailed and rigorous derivation of all relevant results, at the cost of having to sacrifice readability and smoothness of exposition. Motivations, a more stringent exposition, explanation and interpretation of results, and conclusions will be given elsewhere.

2 The Dichotomized Gaussian: A common input model for binary neurons

2.1 The Dichotomized Gaussian

We want to model a multivariate, n -dimensional binary random variable $X \in \{0, 1\}^n$ with mean μ and covariance matrix Σ . In the DG, samples from X are constructed by generating samples from a Normal random variable Z with mean γ and covariance Λ , and binarizing it at 0, i. e. setting $X_i = 1$ iff $Z_i > 0$. For X to have the desired moments μ and Σ , we choose γ and Λ such that

$$\Lambda_{ii} = 1, \mu_i = \Phi(\gamma_i) \quad (1)$$

$$\Sigma_{ij} = \Phi_2(\gamma_i, \gamma_j, \Lambda_{ij}) - \Phi(\gamma_i)\Phi(\gamma_j). \quad (2)$$

Here, $\Phi(\cdot)$ is the cumulative distribution function (cdf) of a univariate Gaussian, and $\Phi_2(\cdot, \cdot, \lambda)$ the cdf of a bivariate Gaussian with unit variances and correlation coefficient λ . Equation (2) has a unique solution for any admissible moments μ and Σ , and can be solved numerically. In the special case of $\mu_i = \mu_j = 1/2$, $\Lambda_{ij} = \sin(2\pi\Sigma_{ij})$.

Fitting the parameters via equations (2) is easy, and possible even for large population sizes n . As samples from multivariate Gaussians can also be obtained in a straightforward manner, the Dichotomized Gaussian can be used as an efficient tool for generating binary random variables with specified second order correlation structure [10, 11]. For alternative methods, see [12, 13, 14, 15]. We here do not consider temporal correlations or point process models of spiking in continuous time. Other studies [16, 17] have used doubly-stochastic processes with Gaussian process rate functions to generate spike trains with controlled correlation structure in continuous time.

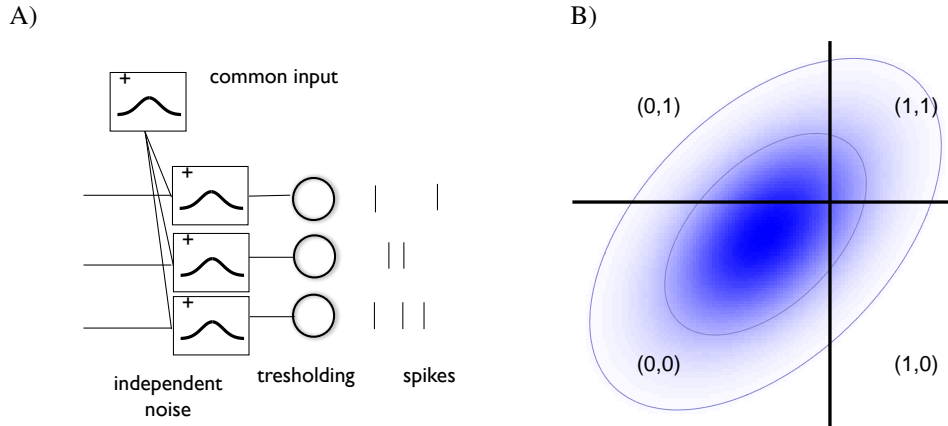


Figure 1: Dichotomized Gaussian: **A)** In this model, correlations are modelled as arising from common, Gaussian input into threshold neurons. **B)** The parameters of this model can be fitted by moment matching: Given the thresholds γ and the correlation of the common input, $\lambda = \Lambda_{12}$, the probability mass in the top left corners determine the probability of two simultaneous spikes, and hence the correlation in the spiking activity.

2.2 Relationship between input- and output correlations

Equation (2) implies a characteristic relationship between the correlations of the Gaussian inputs and the correlations of the binary threshold neurons: Given the same amount of input correlation, the correlation of the binary firing rates will grow monotonically with their firing rates. The qualitatively similar relationship has been observed in integrate-and-fire neurons [18], continuous time threshold crossing processes [19, 20], intra-cellular recordings of slices receiving common input [18], as well as, e.g., in in-vivo recordings of neural population activity in the somatosensory cortex of anesthetized rats [21]. In general, the absolute value of the output correlation is always less than the absolute magnitude of the input correlation: The binary thresholding operation 'destroys' correlation. This effect is most prominent for very low firing rates, and smallest in the symmetric case, i.e. $\mu = 0.5$.

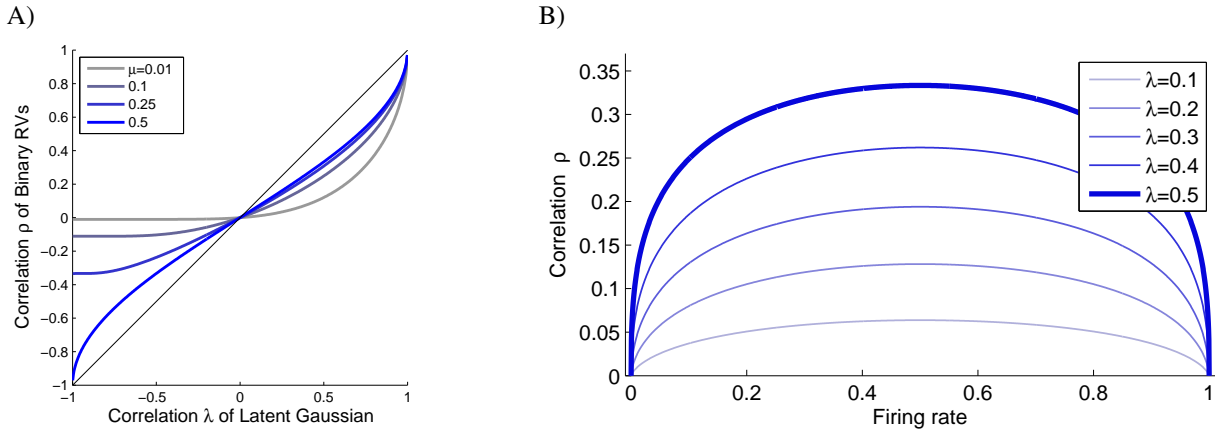


Figure 2: Relationship between input and output correlations: **A)** The transfer-function which maps input-correlation $\lambda = \Lambda_{12}$ to output correlation changes with firing rate. **B)** Similarly, we plot the output correlation as a function of firing rate, for fixed input correlations. Given the same level of input correlation, the output correlation increases monotonically with firing rate up to $\mu = 0.5$.

2.3 Homogeneous population model

In this report, we want to model a homogeneous population with n neurons, each having firing rate μ , constant pairwise covariance c and correlation coefficient $\rho = c/(\mu(1 - \mu))$ [22]. Binary samples from the model are obtained by dichotomizing a latent Gaussian Z with mean γ , unit variances and pairwise covariance (and correlation coefficient) $\lambda > 0$.

We set

$$Z \sim \mathcal{N}(\gamma, \Lambda), \text{ where } \Lambda = \mathbb{I}(1 - \lambda) + \lambda e_n e_n^\top \quad (3)$$

$$Z = \gamma + \sqrt{1 - \lambda}T + \sqrt{\lambda}e_n S \quad (4)$$

$$\text{where } T \sim \mathcal{N}(0, \mathbb{I}_n) \quad (5)$$

$$S \sim \mathcal{N}(0, 1), \quad (6)$$

where e_n is a column vector consisting of all ones.

The binary patterns X are generated by $X = 1$. In this model, $P(X = x) = P(X = y)$ whenever $\|x\| = \|y\|$, where $\|x\| = \sum_i x_i$. Such a homogeneous population model is fully characterized by the number of neurons that spike synchronously, i.e. the population spike count distribution: We set $K = \|X\|$, the population spike count, and $R = K/n$, the proportion of neurons spiking at any time. The probability of seeing a given pattern with k spikes is $Q_{DG}(k) = P_{DG}(K = k)/\binom{n}{k}$. For finite population sizes n , the population spike count distribution is given by

$$P_{DG}(X = x) = \int_z P(X = x|z)p(z)dz \quad (7)$$

$$= \int_{-\infty}^{\infty} \phi_{\lambda}(s)P(X = x|s)ds \quad (8)$$

$$= \int_{-\infty}^{\infty} \phi_{\lambda}(s) \prod_{i=1}^n P(X_i = x_i|s)ds \quad (9)$$

$$= \int_{-\infty}^{\infty} \phi_{\lambda}(s) \prod_{i=1}^n (1 - L(s))^{1-x_i} L(s)^{x_i} ds, \quad (10)$$

$$\text{where } L(s) = P(Z_i > 0|s) \quad (11)$$

$$= P\left(T_i > \frac{-s - \gamma}{\sqrt{1 - \lambda}}\right) \quad (12)$$

$$= \Phi\left(\frac{s + \gamma}{\sqrt{1 - \lambda}}\right). \quad (13)$$

Here, $\phi_{\lambda}(s)$ is the probability density function of a one-dimensional Gaussian with mean 0 and variance λ .

2.4 Asymptotic spike count distribution

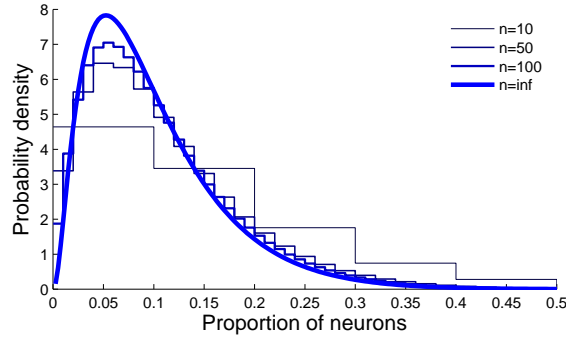


Figure 3: Convergence of population spike count: For increasing population size but fixed input correlation and firing rate, the population spike count distribution converges to its asymptotic distribution.

For large population sizes n , the integral in (10) can be approximated using the saddle-point approximation: We set $U(s) = r \log L(s) + (1 - r) \log L(-s)$ to get

$$Q_{DG}(k) = \int_{-\infty}^{\infty} \phi_{\lambda}(s) L(s)^{rn} (1 - L(s))^{n-rn} ds \quad (14)$$

$$= \int_{-\infty}^{\infty} \phi_{\lambda}(s) e^{nU(s)} ds \quad (15)$$

$$\text{for large } n \approx \sqrt{\frac{2\pi}{-nU''(s_o)}} \phi_{\lambda}(s_o) e^{nU(s_o)} \quad (16)$$

$U(s)$ has a unique maximum u_o , as it is just a binomial likelihood function in L , and L is monotonic in s . Therefore, we get that

$$L(s_o) = r \quad (17)$$

$$s_o = \sqrt{1-\lambda}\Phi^{-1}(r) - \gamma \quad (18)$$

$$e^{nU(s_o)} = (r^r(1-r)^{1-r})^n \quad (19)$$

$$U''(s_o) = -\frac{L'(s_o)^2}{r(1-r)} \quad (20)$$

$$L'(s_o) = \phi\left(\frac{s+\gamma}{\sqrt{1-\lambda}}\right) \frac{1}{\sqrt{1-\lambda}} \quad (21)$$

$$\phi_\lambda(s) = \frac{\phi(s/\sqrt{\lambda})}{\sqrt{\lambda}} \quad (22)$$

Putting this together, we get

$$Q_{DG}(k) = \frac{\phi(s_o/\sqrt{\lambda})}{\phi\left(\frac{s_o+\gamma}{\sqrt{1-\lambda}}\right)} (r^r(1-r)^{1-r})^n \sqrt{\frac{2\pi(1-\lambda)r(1-r)}{\lambda n}} \quad (23)$$

$$(24)$$

and the discrete 'large n ' population spike count distribution

$$P_{DG}(K = k) = Q_{DG}(k) \binom{n}{k} \quad (25)$$

$$= \frac{\phi(s_o/\sqrt{\lambda})}{\phi(\Phi^{-1}(r))} \sqrt{\frac{2\pi(1-\lambda)r(1-r)}{\lambda}} \left(\frac{(r^r(1-r)^{1-r})^n}{\sqrt{n}} \binom{n}{nr} \right) \quad (26)$$

$$\approx \frac{\phi(s_o/\sqrt{\lambda})}{\phi(\Phi^{-1}(r))} \sqrt{\frac{2\pi(1-\lambda)r(1-r)}{\lambda}} \left(\frac{1}{n\sqrt{2\pi r(1-r)}} \right). \quad (27)$$

Renormalizing, this yields the asymptotic, continuous valued spike count distribution:

$$f(r) = \frac{\phi(s_o/\sqrt{\lambda})}{\phi(\Phi^{-1}(r))} \sqrt{\frac{1-\lambda}{\lambda}} \quad (28)$$

$$= \frac{\phi\left(\sqrt{\frac{1-\lambda}{\lambda}}\eta - \frac{\gamma}{\sqrt{\lambda}}\right)}{\phi(\eta)} \sqrt{\frac{1-\lambda}{\lambda}} \quad (29)$$

$$= \exp\left(-\frac{1}{2\lambda} \left((1-2\lambda)\eta^2 - 2\gamma\eta\sqrt{1-\lambda} + \gamma^2\right)\right) \sqrt{\frac{1-\lambda}{\lambda}} \quad (30)$$

$$= \exp\left(-\frac{1}{2} \frac{\left(\eta - \frac{\gamma\sqrt{1-\lambda}}{(1-2\lambda)}\right)^2}{\lambda/(1-2\lambda)}\right) \exp\left(\frac{\gamma^2}{2-4\lambda}\right) \sqrt{\frac{1-\lambda}{\lambda}} \quad (31)$$

where $\eta = \Phi^{-1}(r)$.

In the symmetric case, $\gamma = 0$, we have that

$$f(r) = C \exp\left(-\frac{(1-2\lambda)\eta^2}{2\lambda}\right), \quad (32)$$

which is symmetric around the mode $r = 1/2$, as expected. Also, we want to recover the uniform distribution if $c = \Phi_\lambda(0, 0) - 1/4 = 1/12$. This is the case for $\lambda = \sin\left(\frac{2c}{\pi}\right)$, i.e. $\lambda = 1/2$, in which case $f(r)$ is a constant as desired.

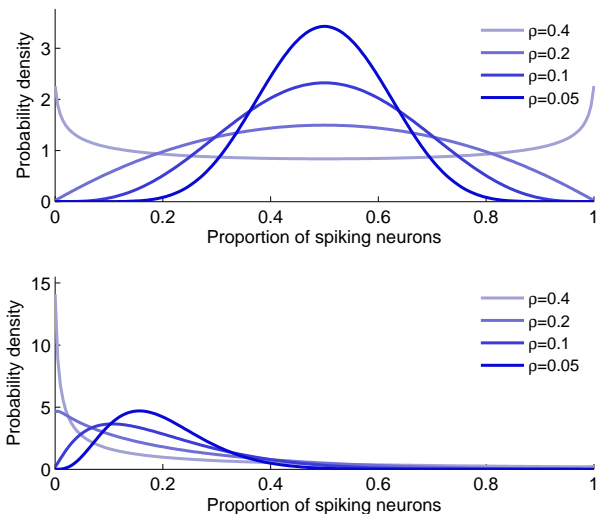


Figure 4: Asymptotic spike count distribution of the DG for mean firing rate $\mu = 0.5$ (top panel) and $\mu = 0.2$, for four different correlation strengths.

2.5 Phase transitions and modes of the distribution

We are interested in how the spike-count distribution changes as the pairwise correlations ρ are increased. The partition-function has a discontinuity at $\lambda = 1/2$. Thus, the transition point is at $\lambda = 1/2$, independently of the mean γ of the latent Gaussian. However, the resulting correlation coefficient ρ of the binary vectors X strongly depends on γ , as can be seen from (2). Translating back from λ to the correlation coefficient ρ , one can see that the transition point is at

$$\rho_c = \frac{\Phi_2(\gamma, \gamma, 1/2) - \Phi(\gamma)^2}{\Phi(\gamma)\Phi(-\gamma)}. \quad (33)$$

For low firing rates μ , even very weak binary correlations lead to a latent covariance of $1/2$, and are sufficient for reaching the critical point. In this model, the ‘strength’ of correlation between binary random variables depends strongly on the firing rates μ [7]. Thus, in this model, the parameter λ measures correlation-strength independently of firing rates, whereas the correlation-coefficient ρ confounds the two [23].

From equation (31), we can see that, for $\lambda \neq 1/2$, the distribution $f(r)$ has a local extremum at

$$r_m(\lambda, \gamma) = \Phi\left(\frac{\gamma\sqrt{1-\lambda}}{1-2\lambda}\right) \quad (34)$$

$$\frac{\partial r_m}{\partial \lambda} = \phi\left(\frac{\gamma\sqrt{1-\lambda}}{1-2\lambda}\right) \frac{\gamma(3-2\lambda)}{2(1-2\lambda)^2\sqrt{1-\lambda}} \quad (35)$$

For $\lambda < 1/2$, r_m is the mode of the distribution, whereas for $\lambda > 1/2$, it constitutes a local minimum.

At the transition point $\lambda = 1/2$, the quadratic term in (31) vanishes, and the spike-count distribution is given by

$$f_c(r) = C \exp\left(\frac{1}{\sqrt{2}}\Phi^{-1}(r)\gamma\right). \quad (36)$$

2.6 Entropy rate

We note that in this model $H(X) = H(X|K) + H(K)$. The entropy of the the spike count $H(K)$, is bounded above by $\log(n)$, which is dominated by $H(X|k)$, which grows linearly. Hence, for large populations, we can

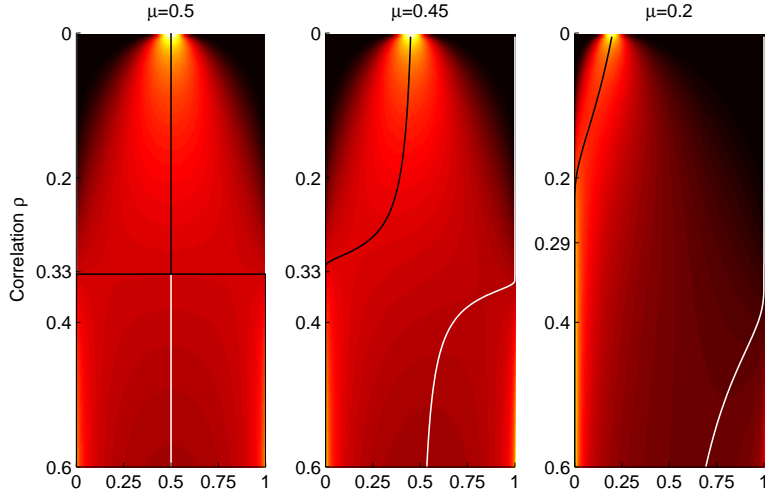


Figure 5: Changes in spike count statistics for varying correlation strength ρ and constant mean $\mu = 0.5$, $\mu = 0.45$ and $\mu = 0.2$. Each line in each panel represents the spike count distribution for a given correlation and mean. The black and gray lines mark the mode and minimum of the distribution. The transition points are at $\rho = 1/3, 0.332, 0.29$.

calculate the entropy by considering only $H(X|k)$:

$$h_{DG} = \frac{1}{n} H(X) \quad (37)$$

$$= \frac{1}{n} \sum_k P(K = k) \log \binom{n}{k} \quad (38)$$

$$\approx \sum_k P(K = k) \left(-\frac{k}{n} \log \frac{k}{n} - \left(1 - \frac{k}{n}\right) \log \left(1 - \frac{k}{n}\right) \right) \quad (39)$$

$$\approx \int_0^1 f(r) [-r \log r - (1-r) \log(1-r)] dr \quad (40)$$

Alternatively, the entropy can be computed noting that it can be decomposed into the entropy of the model given the common input S , and the mutual information between common input and X , $MI(S : X)$:

$$H_{DG}(n) = - \sum_x p(x) \log(p(x)) \quad (41)$$

$$= - \int_{-\infty}^{\infty} \sum_x p(x, s) \log(p(x)) ds \quad (42)$$

$$= - \int_{-\infty}^{\infty} \sum_x p(x, s) \left(\log p(x|s) - \frac{\log p(x|s)}{p(x)} \right) ds \quad (43)$$

$$= - \int_{-\infty}^{\infty} p(s) \sum_x p(x|s) \log p(x|s) ds \quad (44)$$

$$+ \int_{-\infty}^{\infty} p(s) \sum_x p(x|s) \log \frac{p(x|s)}{p(x)} ds \quad (45)$$

$$=: H(X|S) + MI(S : X) \quad (46)$$

The conditional entropy $H(X|S)$ is given by

$$H(X|S) = n \int_{-\infty}^{\infty} p(s) [L(s) \log L(s) + L(-s) \log L(-s)] ds \quad (47)$$

The second term, $MI(S : X)$ is negligibly small for large n . This follows (e. g.) from the data-processing inequality:

$$MI(S : X) \leq MI(S : Z) \quad (48)$$

$$= H(Z) - H(Z|S) \quad (49)$$

$$= \frac{1}{2} (\log(\det(\Lambda)) - \log(\det((1 - \lambda)\mathbb{I}_n))) \quad (50)$$

$$= \frac{1}{2} \log \left(1 + \frac{\lambda n}{1 - \lambda} \right) \quad (51)$$

Therefore, for large n :

$$h_{DG} = \int_{-\infty}^{\infty} \phi_{\lambda}(s) [L(s) \log L(s) + L(-s) \log L(-s)] ds \quad (52)$$

By changing variables $L(s) = r$, one can see that the two derivations lead to the same result.

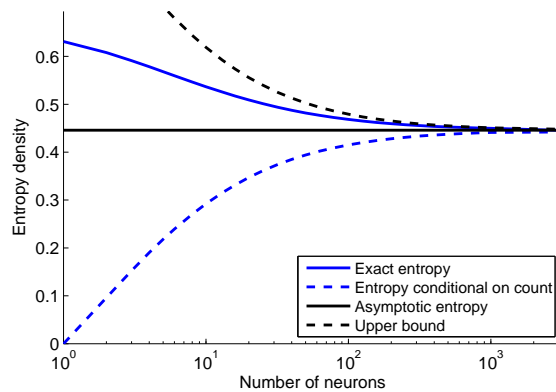


Figure 6: Behaviour of the entropy rate as the population size is increased: The entropy can be decomposed in the entropy of the spike count, and the entropy of the model conditional on the spike count. For large population sizes, the conditional entropy (blue dashed) dominates. Therefore, the asymptotic entropy can be computed by just considering this conditional entropy. Alternatively, the entropy can be computed using the fact that the mutual information between X and the common input S is bounded by the mutual information between S and Z . For large n , this bound becomes tight (black dashed).

3 The Ising Model

3.1 Spike count distribution for finite population sizes n

For the Ising model [1, 24], the probability of observing a given pattern x with k spikes ($\|x\| = k$) is given by

$$Q_{isi}(\|X\| = k) = \frac{1}{Z_{isi}} \exp\left(hk + \frac{Jk^2}{2n}\right), \quad (53)$$

and the probability of observing k spikes is

$$P_{isi}(K = k) = \binom{n}{k} \frac{1}{Z_{isi}} \exp\left(hk + \frac{Jk^2}{2n}\right). \quad (54)$$

The partition function Z_{isi} is given by

$$Z_{isi} = \sum_x \exp\left(h\|x\| + \frac{J\|x\|^2}{2n}\right) \quad (55)$$

$$= \sum_x \exp(h\|x\|) \sqrt{\frac{Jn}{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{Jnm^2}{2} + Jm\|x\|\right) dm \quad (56)$$

$$= \sqrt{\frac{Jn}{2\pi}} \int_{-\infty}^{\infty} \exp\left(\frac{-Jnm^2}{2}\right) (1 + \exp(Jm + h))^n dm \quad (57)$$

$$= \sqrt{\frac{Jn}{2\pi}} \int_{-\infty}^{\infty} \exp\left(\frac{-Jnm^2}{2} + n \log(1 + \exp(Jm + h))\right) dm \quad (58)$$

For finite n , we fit the parameters numerically [25] and calculate the entropy via $H(X) = H(K) + H(X|K)$ to avoid having to sum across all 2^n patterns.

For large n , the integral can be approximated [26] by

$$Z_{isi} \approx \sqrt{\frac{Jn}{2\pi}} \sqrt{\frac{2\pi}{-nU''(m_o)}} \exp(nU(m_o)) \quad (59)$$

where

$$U(m) = -\frac{Jm^2}{2} + \log(1 + \exp(Jm + h)) \quad (60)$$

and m_o is the solution to

$$m = \frac{1}{1 + \exp(-Jm - h)}. \quad (61)$$

This gives

$$U''(m_o) = J^2 \exp(-Jm - h) m^2 - J \quad (62)$$

$$Z \approx \frac{\exp(nU(m_o))}{1 - \sqrt{J \exp(-Jm - h) m^2}} \quad (63)$$

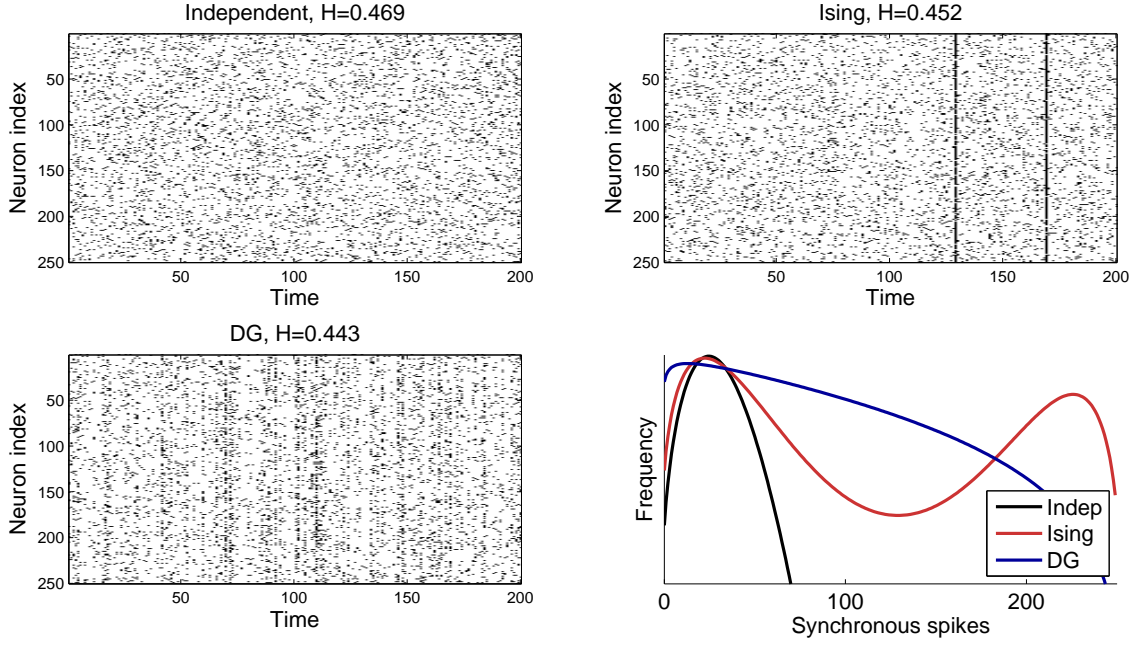


Figure 7: Comparison between the Dichotomized Gaussian and the Ising model for large population sizes: For comparison, we show simulated spike rasters from a model with no correlation but same mean firing rate (top left). The spike count distribution of the Ising model is bimodal (bottom right). Thus, it behaves like a mixture of two independent models: For most of the time, neurons firing independently with low firing rate, but this firing mode is punctuated by events in which very neurons fire synchronously. Interestingly, these synchronous firing events do not require higher order interactions. In contrast, the spike-count distribution of the DG has a much wider range, with firing patterns of different orders occurring (bottom left).

3.2 Asymptotic entropy

For large population sizes n , $H(K)$ is negligibly small compared to $H(X|K)$, and we find the MaxEnt distribution directly by finding the spike-count distribution P_{isi} which maximizes

$$H(X|K) = \frac{1}{n} \sum_k P_{isi}(K = k) \log \binom{n}{k}. \quad (64)$$

This is really the linear program

$$\max_P H(X|K) = \frac{1}{n} \sum_k P_k \log \binom{n}{k} \quad (65)$$

$$\text{subject to } P_k \geq 0 \forall k \quad (66)$$

$$\sum_k P_k = 1 \quad (67)$$

$$\frac{1}{n} \sum_k P_k k = \mu \quad (68)$$

$$\frac{1}{n^2} \sum_k P_k k^2 = \mu^2 n^2 + cn(n-1) + \mu(1-\mu)n. \quad (69)$$

$$(70)$$

In the limit of $n \rightarrow \infty$, we can write it as

$$\max_f H(X|K) = - \int_0^1 f(r)r \log(r) + (1-r) \log(1-r) dr \quad (71)$$

$$\text{subject to } f(r) \geq 0 \quad \forall r \in [0, 1] \quad (72)$$

$$\int f(r) dr = 1 \quad (73)$$

$$\int f(r)r dr = \mu \quad (74)$$

$$\int f(r)r^2 dr = \mu^2 + c. \quad (75)$$

$$(76)$$

By finding the function f that maximizes it, we can find the asymptotic entropy of the maximum-entropy distribution without having to use the exponential form of the MaxEnt distribution, and in particular, without identifying the parameters h and J . We make the informed guess that the functional form of P which maximizes is a mixture of two delta-distributions, $f_{isi}(r) = p_- \delta(r-r_-) + p_+ \delta(r-r_+)$, where $r_+ = 1-r_-$ and the locations r_{\pm} and probabilities p_{\pm} are chosen such that the mean and variance are as desired, i.e. by

$$\mu = p_- r_- + p_+ r_+ \quad (77)$$

$$c + \mu^2 = p_- r_-^2 + p_+ r_+^2 \quad (78)$$

$$r_{\pm} = \frac{1}{2} \pm \sqrt{1/4 - \mu + \mu^2 + c} \quad (79)$$

$$p_+ = \frac{\mu + r_+ - 1}{2r_+ - 1} \quad (80)$$

$$p_- = 1 - p_+. \quad (81)$$

The entropy rate is given by

$$h_{isi} = \lim_{n \rightarrow \infty} \left(\frac{p_-}{n} \log \binom{n}{nr_-} + \frac{p_+}{n} \log \binom{n}{nr_+} \right) \quad (82)$$

$$= p_- H(r_-) + p_+ H(r_+) \quad (83)$$

$$\text{where } H(r) = -r \log(r) - (1-r) \log(1-r) \quad (84)$$

Summing up, the entropy is given by:

$$r = \frac{1}{2} \pm \sqrt{1/4 - \mu + \mu^2 + c} \quad (85)$$

$$= \frac{1}{2} \pm \sqrt{1/4 + \mu(1-\mu)(1-\rho)} \quad (86)$$

$$H = -r \log(r) - (1-r) \log(1-r) \quad (87)$$

Thus, we can calculate the asymptotic entropy rate of the Ising model without having to identify the parameters h and J explicitly. Also, we see that an Ising model with rate μ_0 and no correlations has the same entropy as one rate μ_1 with pairwise covariance c provided that $\mu_0^2 - \mu_0 = \mu_1^2 - \mu_1 + c$.

We are yet to show that the form of f_{isi} that we guessed above is indeed the solution to the linear program. In general, one can prove the optimality [27] of a feasible solution x^* for the linear program

$$\min v^T x \quad (88)$$

$$\text{subject to } x \geq 0 \quad (89)$$

$$Ax = b \quad (90)$$

by finding a solution λ^* to the dual program

$$\max \lambda^\top b \quad (91)$$

$$A^\top \lambda \leq v \quad (92)$$

which has the same result, i.e. $v^\top x^* = \lambda^{*\top} b$. Rewriting the MaxEnt problem as minimizing the negative entropy, we need to find a vector λ which satisfies

$$\text{a) } \lambda_1 + \lambda_2 \mu + \lambda_3(\mu^2 + c) = -h_{isi} \quad (93)$$

$$\text{b) } \lambda_1 + \lambda_2 s + \lambda_3 s^2 \leq s \log(s) + (1-s) \log(1-s). \forall s \in [0, 1], \quad (94)$$

We take

$$\lambda_1 = -h_{isi} + \lambda_2(r^2 - r) \quad (95)$$

$$\lambda_2 = \log\left(\frac{r}{1-r}\right)/(1-2r) \quad (96)$$

$$\lambda_3 = -\lambda_2. \quad (97)$$

(The values of λ are obtained by requiring that the equality a) holds, and that inequality b) holds with equality at the point $s = \mu - \mu^2 - c$, and that the derivatives of the lhs and rhs of equation b) also match at this point).

Noting that $\mu - \mu^2 - c = r - r^2$, condition a) holds:

$$\lambda_1 + \lambda_2 \mu + \lambda_3(\mu^2 + c) = \quad (98)$$

$$= -h_{isi} + \lambda_2(r^2 - r) + \lambda_2(\mu - \mu^2 - c) \quad (99)$$

$$= -h_{isi} \quad (100)$$

For condition b) to hold, we check that r is a local minimum of the function $\Delta(s) = r h s(s) - l h s(s)$. As $\Delta(r) = 0$, and that it does not have any other stationary points for $s \leq 1/2$, we get that $\Delta(s) \geq 0 \forall s \in (0, 1)$.

3.3 Minimum entropy distribution

For the minimum entropy model (assuming exchangeability), we can use the same type of argument, by guessing the solution to a linear program, and verifying its correct-ness by constructing the solution to the dual problem. In that case, the solution turns out to be a sum of three delta-peaks at locations 0, r_m and 1. The mean and correlation determine both the location of the middle-peak r_m and the relative weights p , $f_{min}(r) = p_0 \delta(r-0) + p_m \delta(r-r_m) + p_1 \delta(r-1)$. Without loss of generality, we only consider the case $\mu \leq 1/2$. The functional form of solution is different for small and big correlations, and changes at a transition point given by $c_o = 1/2\mu - \mu^2$. For $c \leq c_o$

$$r_m = \mu + (1-\mu)\rho \quad (101)$$

$$p_m = \mu/r_m \quad (102)$$

$$p_0 = 1 - p_m, p_1 = 0 \quad (103)$$

$$h_{min} = \lim_{n \rightarrow \infty} p_0 \frac{1}{n} \log \binom{n}{0} + p_m \frac{1}{n} \log \binom{n}{nr_m} \quad (104)$$

$$= 0 - p_m (r_m \log r_m + (1-r_m) \log(1-r_m)) \quad (105)$$

$$= -\mu (\log r_m + (1/r_m - 1) \log(1-r_m)) \quad (106)$$

For $c > c_o$

$$p_m = 4\mu(1-\mu)(1-\rho) \quad (107)$$

$$r_m = 1/2 \quad (108)$$

$$p_0 = p_1 \quad (109)$$

$$h_{min} = \lim_{n \rightarrow \infty} p_0 \frac{1}{n} \log \binom{n}{0} + p_m \frac{1}{n} \log \binom{n}{n/2} + p_1 \frac{1}{n} \log \binom{n}{n} \quad (110)$$

$$= 0 + p_m + 0 = p_m \quad (111)$$

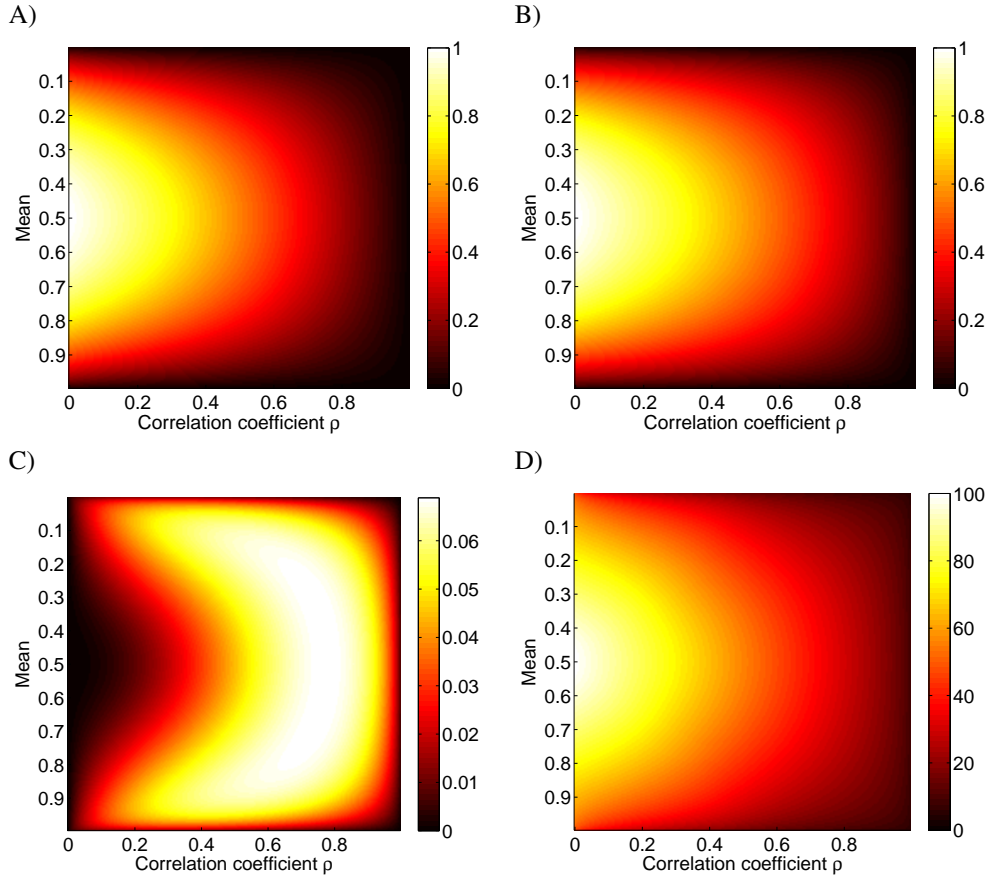


Figure 8: Comparison of the entropy rate for the DG and Ising model. **A** Entropy rate of the Dichotomized Gaussian for different values of the mean μ and correlation coefficient ρ . **C** Absolute entropy difference between the Ising model and DG, $h_{isi} - h_{DG}$. **D** Scaled entropy of the DG-model, $(h_{DG} - h_{min}) / (h_{isi} - h_{min})$ in percent.

4 More general covariance structures

4.1 A finite number of mutually uncorrelated pools

We first consider a population consisting of l pools of size n_1, n_2, \dots, n_l , $\sum_i n_i = n$, and relative sizes $b_i = \frac{n_i}{n}$, $\sum_i b_i = 1$. We define $B = \sum_i b_i^2$. Each neuron has mean firing rate μ and pairwise correlation ρ (covariance c), but only with neurons within its pool. In this case, the average covariance is given by

$$c_{av} = c \frac{\sum_i n_i(n_i - 1)}{n(n - 1)} \quad (112)$$

$$\approx c \sum_i b_i^2 \quad (113)$$

$$= cB \quad (114)$$

and the entropy density by

$$h_{\text{pools}} = \frac{1}{n} \sum_i H(\text{pool } i) \quad (115)$$

$$= \frac{1}{n} \sum_i n_i h_{DG} \quad (116)$$

$$= h_{DG} \quad (117)$$

The entropy density of the pool-model is the same as that of the infinite-range model. However, the average correlation between any pair of neurons is only cB . By making B smaller (e.g. by using more pools), we can make the average correlations arbitrarily small without changing the (asymptotic) entropy h .

The population spike-count distribution of this model is the convolution of the single-pool spike-count distribution with itself n times.

More generally, if the l pools have mean firing rates and with-in pool correlations μ_i and ρ_i , then, in the asymptotic limit, we get

$$h_{\text{pools}} = \frac{1}{n} \sum_i H(\text{pool}_i) \quad (118)$$

$$= \frac{1}{n} \sum_i n_i h_{DG}(\mu_i, \rho_i) \quad (119)$$

4.2 Scaling up general covariance matrices

Suppose we have recorded a population of l neurons, with some covariance matrix C_l . How can we make bigger populations that have 'the same second order structure'? One approach is to make bigger matrices C' by random sampling elements from C , or by randomly sampling elements from the underlying correlation-coefficient matrix, and thereby generating bigger, synthetic covariance matrices. The problem of that approach is that we have no guarantee as to whether the synthetically generated matrix C' will indeed be 'valid', i.e. whether it will be the covariance matrix of any multivariate binary distribution. Even if C' is positive definite, and even if each entry C'_{ij} satisfies the constraints that the means of two binary random variables impose on their covariance, validity of C' is not guaranteed [28, 11].

Alternatively, one can envisage a model in which correlations are modelled as arising from a common input. Then, we extrapolate to larger population sizes under the assumption that the common input is fixed. In other words, we assume that the extrapolation corresponds to a denser and denser sampling of the population.

4.3 Multivariate common input model

In this model, we assume that the latent Gaussian can be decomposed into a common input L and independent noise T :

$$Z = \gamma + DT + BU, \quad (120)$$

where D is a diagonal matrix of size $l \times l$ which determines the variance of the independent noise, and B is an $l \times b$ dimensional, low rank matrix of 'loadings', which determines how much each of the neurons gets as a common input. T and U are l dimensional standard Gaussians, and γ is an l dimensional offset. The dimensionality of U , b , is the rank of the common input. This model has covariance matrix

$$\hat{\Lambda} = \text{Cov}(Z) = D^2 + BB^\top \quad (121)$$

We first find the covariance matrix of the Gaussian inputs Z , Λ , via equation (2) Then, perform a factor analysis [29] on Λ to identify B and D from Λ . If $b < l$, Λ and $\hat{\Lambda}$ will not be identical, but should be similar. It should be noted that both of these steps could be combined into a single maximum likelihood estimation of B and D from the binary observations X —algorithms for this task are variants of 'binary factor analysis'.

4.4 Extrapolations: Scaling up the model

We extrapolate the model to larger population sizes by keeping the common input fixed. For each neuron which receives a common input and adds independent noise, we create m 'copies' which receive, on each trial, the same common input, and have independent noise with the same statistical properties. In other words, we consider a sequence of populations $Z^{(1)}, Z^{(2)}, \dots, Z^{(m)}$ constructed by the following procedure: $Z^{(1)} = \gamma + DT^{(1)} + BU$. In the general case, the population

$$Z^{(m)} = \gamma^{(m)} + D^{(m)}T^{(m)} + B^{(m)}U \quad (122)$$

is defined by: $\gamma^{(m)} = \mathbf{e}_m \otimes \gamma$, $D^{(m)} = \mathbf{I}_s \otimes D$, $T^{(m)}$ is an $n = lm \times 1$ dimensional standard Gaussian, and $B^{(m)} = \mathbf{e}_l \otimes B$. We also define the random variable $S^{(m)} = B^{(m)}U$, i.e. the common input, with associated covariance matrix $\Lambda_S = BB^\top$.

Thus, $Z^{(m)}$ has covariance matrix

$$\text{Cov}(Z^{(m)}) = \mathbf{I}_m \otimes D^2 + \mathbf{e}_m \mathbf{e}_m^\top \otimes BB^\top. \quad (123)$$

It will sometimes be convenient to rewrite $Z^{(m)}$ as a matrix of size $l \times m$, where $Z_{ij}^{(m)}$ denotes the input to the j -th element of the i th pool—whenever we use two indices, we mean the matrix versions of Z or X . As before, binary random variables X are obtained by cutting of Z at 0.

4.5 Asymptotic entropy

We calculate the entropy via the formula

$$H_{DG}(n) = H(X|S) + MI(S : X) \quad (124)$$

$$\approx H(X|S) = \int_s H(X|S = s)p(s)ds \quad (125)$$

$$= \int_s p(s) \sum_{i=1}^n H(X_i|S = s)ds \quad (126)$$

$$= \sum_{j=1}^m \sum_{i=1}^l \int_s p(s)H(X_{ij}|S_i = s)ds \quad (127)$$

$$= m \sum_{i=1}^l \int_{s_i} p(s_i)H(X_{i1}|S_i = s_i)ds_i \quad (128)$$

$$= m \sum_{i=1}^l \int_{s_i} p(s_i) [L_i(s_i) \log(L_i(s_i)) + (1 - L_i(s_i)) \log(1 - L_i(s_i))] ds_i \quad (129)$$

where $L_i(s_i) = P(Z_{i1} > 0|s_i) = \Phi\left(\frac{s_i + \gamma}{D_{ii}}\right)$ and $p(s_i) = \sqrt{\Lambda_{S,ii}}\phi\left(s_i/\sqrt{\Lambda_{S,ii}}\right)$. We can see that the formula only depends on the diagonal of the correlation input Λ_{ii} : In other words, the cross-pool correlations do not contribute to the asymptotic entropy at all, only the within-pool correlations matter.

4.6 Asymptotic spike count distribution

We also want to write down the 'generalized spike count distribution' of $P(R = r)$, where we define K to be the l dimensional vector with $K_i = \sum_{k=1}^m X_{ik}$, and $R_i = \frac{1}{m}k_i$.

$$P(X = x) = \int P(X = x|S = s)P(S = s)ds \quad (130)$$

$$= \int \phi_\Lambda(s) \prod_{i=1}^l \prod_{j=1}^m P(X_{ij}|s_i)ds \quad (131)$$

$$= \int \phi_\Lambda(s) \prod_{i=1}^l \prod_{j=1}^m L_i(s_i)^{x_{ij}} L_i(-s_i)^{1-x_{ij}} ds \quad (132)$$

$$= \int \phi_\Lambda(s) \prod_{i=1}^l L_i(s_i)^{k_i} L_i(-s_i)^{m-k_i} ds \quad (133)$$

$$= \int \phi_\Lambda(s) \exp(mU(s)) ds \quad (134)$$

with $U(s) = \sum_{i=1}^l r_i \log L_i(s_i) + (1 - r_i)L_i(-s_i)$ Now, using argument entirely analogously to the case of a single pool, we get that, in the limit of large m , this expression becomes:

$$P(X = x) \approx \frac{(2\pi)^{l/2}}{\sqrt{-m \det \left(\frac{\partial^2 U}{\partial s^2}(s) \right)}} \phi_{\Lambda_S}(s) e^{mU(s)} \quad (135)$$

where

$$L_i(s_i) = r_i \quad (136)$$

$$s_i = D_{ii} \Phi^{-1}(r_i) - \gamma \quad (137)$$

$$e^{mU(s)} = \left(\prod_{i=1}^l r_i^{r_i} (1 - r_i)^{1-r_i} \right)^m \quad (138)$$

$$\det \left(\frac{\partial^2 U}{\partial s^2}(s) \right) = \prod_{i=1}^l \frac{-L'_i(s_i)^2}{r_i(1-r_i)} \quad (139)$$

$$L'_i(s_i) = \phi \left(\frac{s_i + \gamma}{D_{ii}} \right) \frac{1}{D_{ii}} \quad (140)$$

Hence, for large n , we have that:

$$P(X = x) \approx \frac{(2\pi)^{l/2}}{m^{l/2} \sqrt{\prod_{i=1}^l \frac{\left(\phi \left(\frac{s_i + \gamma}{D_{ii}} \right) \frac{1}{D_{ii}} \right)^2}{r_i(1-r_i)}}} \left(\prod_{i=1}^l r_i^{r_i} (1 - r_i)^{1-r_i} \right)^m \phi_{\Lambda_S}(s) \quad (141)$$

$$(142)$$

From this, we can derive the asymptotic generalized spike count distribution $f(r)$:

$$P(R = r) = P(X = x) \prod_{i=1}^l \binom{m}{mr_i} \quad (143)$$

$$= \frac{(2\pi)^{l/2}}{\sqrt{\prod_{i=1}^l \frac{\left(\phi \left(\frac{s_i + \gamma}{D_{ii}} \right) \frac{1}{D_{ii}} \right)^2}{r_i(1-r_i)}}} \left(\prod_{i=1}^l \frac{1}{\sqrt{m}} (r_i^{r_i} (1 - r_i)^{1-r_i})^m \binom{m}{mr_i} \right) \phi_{\Lambda_S}(s) \quad (144)$$

$$\approx \frac{(2\pi)^{l/2}}{\sqrt{\prod_{i=1}^l \frac{\left(\phi \left(\frac{s_i + \gamma}{D_{ii}} \right) \frac{1}{D_{ii}} \right)^2}{r_i(1-r_i)}}} \left(\prod_{i=1}^l \frac{1}{m \sqrt{2\pi r_i(1-r_i)}} \right) \phi_{\Lambda_S}(s) \quad (145)$$

$$= \frac{1}{m^l} \left(\prod_{i=1}^l \frac{\left(\phi \left(\frac{s_i + \gamma}{D_{ii}} \right) \frac{1}{D_{ii}} \right)^2}{r_i(1-r_i)} \prod_{i=1}^l r_i(1-r_i) \right)^{-1/2} \phi_{\Lambda_S}(s) \quad (146)$$

$$(147)$$

Hence, finally, the generalized spike count density is given by

$$f(r) = \frac{\phi_{\Lambda_S}(D\Phi^{-1}(r) - \gamma)}{\prod_{i=1}^l \frac{\phi(\Phi^{-1}(r_i))}{D_{ii}}} \quad (148)$$

$$= \det D \frac{\phi_{\Lambda_S}(D\Phi^{-1}(r) - \gamma)}{\phi(\Phi^{-1}(r))} \quad (149)$$

$$= \frac{\det(D)}{\det(\Lambda_S)^{1/2}} \exp\left(-\frac{1}{2}((D\eta - \gamma)^\top \Lambda_S^{-1}(D\eta - \gamma) - \eta^\top \eta)\right) \quad (150)$$

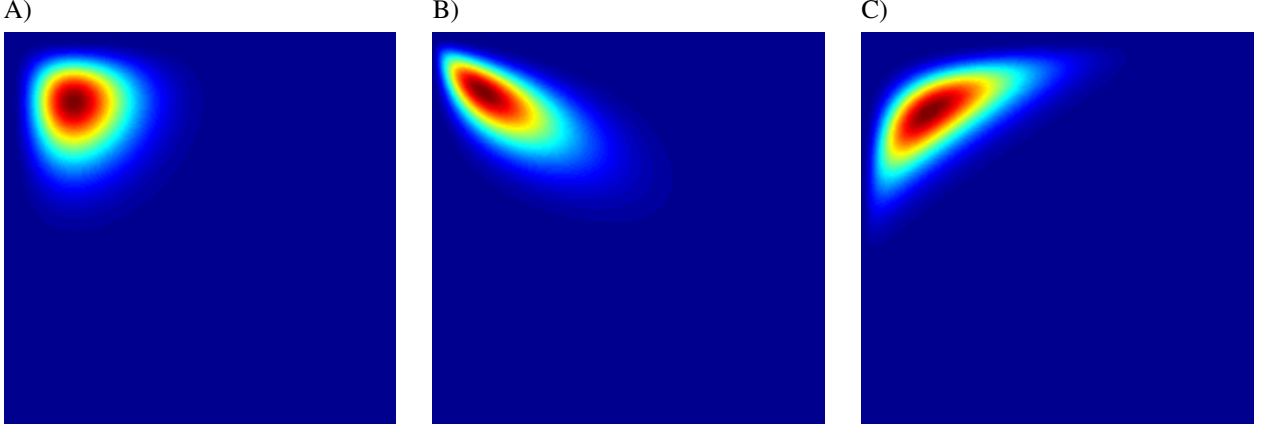


Figure 9: Generalized spike count distribution for a model with two pools and within-pool correlations. In A), the pools are not correlated, in B), they have a positive correlations, whereas in C), the correlation is negative.

4.7 General input distributions

Up to now, we assume that the correlated input into the cells were Gaussian, and that each individual cell had additive, Gaussian noise. We can easily generalize the spike-count distribution $f(r)$ to an arbitrary common input with distribution $f_s(s)$, and a noise distribution $P(X_i = 1|S_i = s_i) = L_i(s)$. In this case,

$$f(r) = \frac{f_s(L^{-1}(r))}{\prod_{i=1}^l L'_i(L_i^{-1}(r_i))} \quad (151)$$

In summary, the generalized spike count distribution can be obtained by starting with the distribution of the common input, and performing a change of variables based on the noise models $L_i(s)$ on each dimension. This makes sense intuitively: The probability of observing a proportion of r of the neurons spiking is the same as the probability of observing a common input which, in each individual neuron, has the corresponding probability of eliciting a spike.

Acknowledgments

JHM and MB were supported by the Bernstein Prize to MB (BMBF; FKZ: 01GQ0601) and the Max Planck Society. We thank S. Gerwinn for useful discussions, and Philipp Berens for comments on the manuscript.

References

- [1] G. Parisi. *Statistical Field Theory*. Perseus Books Group, 1998.
- [2] Elad Schneidman, Michael J 2nd Berry, Ronen Segev, and William Bialek. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440(7087):1007–12, 2006.
- [3] J. Shlens, G. D. Field, J. L. Gauthier, M. I. Grivich, D. Petrusca, A. Sher, A. M. Litke, and E. J. Chichilnisky. The structure of multi-neuron firing patterns in primate retina. *J Neurosci*, 26(32):8254–8266, 2006.
- [4] E.T. Jaynes. Information Theory and Statistical Mechanics. *Physical Review*, 106(4):620–630, 1957.
- [5] Y. Roudi, S. Nirenberg, and P Latham. Pairwise maximum entropy models for studying large biological systems: when they can and when they can’t work. *Plos Computational Biology*, 2009.
- [6] D. R. Cox and N. Wermuth. On some models for multivariate binary variables parallel in complexity with the multivariate gaussian distribution. *Biometrika*, 89:462–469, June 2002.
- [7] de La, Brent Doiron, Eric Shea-Brown, Kresimir Josic, and Alex Reyes. Correlation between neural spike trains increases with firing rate. *Nature*, 448(7155):802–806, 2007.
- [8] Tamara Broderick, Miroslav Dudik, Gasper Tkacik, Robert E Schapire, and William Bialek. Faster solutions of the inverse pairwise ising problem. *arXiv*, q-bio.QM, Dec 2007.
- [9] M. Bethge and P. Berens. Near-maximum entropy models for binary neural representations of natural images. In *Proceedings of the Twenty-First Annual Conference on Neural Information Processing Systems*, volume 21, 2007.
- [10] L. J. Emrich and M. R. Piedmonte. A method for generating high-dimensional multivariate binary variates. *The American Statistician*, 45(4):302–304, 1991.
- [11] Jakob H Macke, Philipp Berens, Alexander S Ecker, Andreas S Tolias, and Matthias Bethge. Generating spike trains with specified correlation coefficients. *Neural Computation*, 2009.
- [12] A. J. Lee. Generating Random Binary Deviates Having Fixed Marginal Distributions and Specified Degrees of Association. *The American Statistician*, 47(3), 1993.
- [13] A. D. Lunn and S. J. Davies. A note on generating correlated binary variables. *Biometrika*, 85(2):487, 1998.
- [14] Bahjat Qaqish. A family of multivariate binary distributions for simulating correlated binary variables with specified marginal means and correlations. *Biometrika*, 90(2):455–463, 2003.
- [15] E. Niebur. Generation of synthetic spike trains with defined pairwise correlations. *Neural Comp.*, 19:1720–1738, July 2007.
- [16] R. Brette. Generation of correlated spike trains. *Neural Computation*, 21(1):188–215, 2009.
- [17] M. Krumin and S. Shoham. Generation of Spike Trains with Controlled Auto-and Cross-Correlation Functions. *Neural Computation*, pages 1–23, 2009.
- [18] Jaime de la Rocha, Brent Doiron, Eric Shea-Brown, Kresimir Josic, and Alex Reyes. Correlation between neural spike trains increases with firing rate. *Nature*, 448(7155):802–806, 2007 Aug 16.
- [19] Tatjana Tchumatchenko, Aleksey Malyshev, Theo Geisel, Maxim Volgushev, and Fred Wolf. Correlations and synchrony in threshold neuron models, 2008.
- [20] Y. Burak, S. Lewallen, and H. Sompolinsky. Stimulus-dependent correlations in threshold-crossing spiking neurons. *Arxiv preprint arXiv:0902.2037*, 2009.

- [21] David S Greenberg, Arthur R Houweling, and Jason N D Kerr. Population imaging of ongoing neuronal activity in the visual cortex of awake rats. *Nat Neurosci*, 11(7):749–751, 2008 Jul.
- [22] Shun-Ichi Amari, Hiroyuki Nakahara, Si Wu, and Yutaka Sakai. Synchronous firing and higher-order interactions in neuron pool. *Neural Comput*, 15(1):127–142, 2003 Jan.
- [23] SI Amari. Measure of Correlation Orthogonal to Change in Firing Rate. *Neural computation*, 2008.
- [24] S. M. Bohte, H. Spekreijse, and P. R. Roelfsema. The effects of pair-wise and higher-order correlations on the firing rate of a postsynaptic neuron. *Neural Comp.*, 12:153–179, 2000.
- [25] J.N. Darroch and D. Ratcliff. Generalized iterative scaling for log-linear models. *The Annals of Mathematical Statistics*, 43(5):1470–1480, 1972.
- [26] H. Nishimori. *Statistical physics of spin glasses and information processing*. Oxford University Press New York, 2001.
- [27] W. Karush. Minima of functions of several variables with inequalities as side constraints. Master’s thesis, University of Chicago, 1939.
- [28] N. R. Chaganty and H. Joe. Range of correlation matrices for dependent bernoulli random variables. *Biometrika*, 93(10):197–206, 2006.
- [29] B. Everitt. *An introduction to latent variable models*. Chapman and Hall New York, 1984.