Phase transitions in optimal neural coding: Complexity via optimisation

N G Stocks
[1] School of Engineering, University of Warwick
E-mail: n.g.stocks@warwick.ac.uk

1. Description of the mini-project

It is, of course, well known that systems consisting of many interacting elements can undergo phase transitions as a control parameter (e.g. temperature) is varied. However, it is perhaps less obvious that systems of non-interacting elements can also be governed by systems of equations that display phase transitions. In this context, it has recently been discovered that the *parameters* that govern the optimal coding of an *uncoupled* neural population (network) display phase transitions Ref. [1]. The coupling that ultimately leads to the phase transition arises from the optimization process itself. This occurs because the set of parameters that optimise the performance of the network are inter-dependent and hence in some sense are coupled. For example, changing and fixing one parameter will require all other parameters to be adjusted to keep the system operating optimally. In addition, the objective function to be optimised (in our case the mutual information) depends on these parameters in a highly nonlinear way. Consequently, one can view the optimisation process as that of navigating a 'potential landscape' that exhibits multiple attracting states - these states can display bifurcations (similar to saddle node) as parameters are changed. All of these considerations illustrate that the optimisation of an uncoupled network of neurons can result in a complex solution that exhibits multiple transitions/bifurcations.

At present, there is no theory that predicts the critical parameters of the transition or their type (e.g.some look like 1st order transitions and others like 2nd order). However, we have been able to map the theory into a minimization problem that looks very much like the minimisation of a free energy in statistical physics (details below). Interestingly, the partition function associated with the free energy can be identified to be the Fisher information of the system. In the project you will investigate these transition, initially using numerical techniques but hopefully it will be able to develop a theortical understanding. In particular we will identify the type of transition and characterise the critical control paramaters at which they occur.

Although these results are of immediate interest to the coding of information by neural systems we also believe that some of our main results are generic in the sense that they are applicable to a wide range of other optimized systems. For example, there are strong analogies between to phase transitions observed in the context of deterministic annealing applied to fuzzy clustering Ref [2]. We also suspect that the results may be relevant to optimised voting systems (and other Boolean type networks) but this has yet to be established. Part of this project will be to look at the applicability of the results/methods to other types of networks that commonly occur in connection with studies of complex systems.

2. Background Information - Optimization of neural systems

We consider the network/population of McCulloch-Pitts type neurons shown in Fig. 1. This network is used to model the quantization of analogue signals by analogue-to-digital converters as well as to study the coding of information by neural systems. It is of sufficient simplicity that theoretical progress can be made but at the same time captures the primary feature of neurons i.e. the presence of a neural threshold.

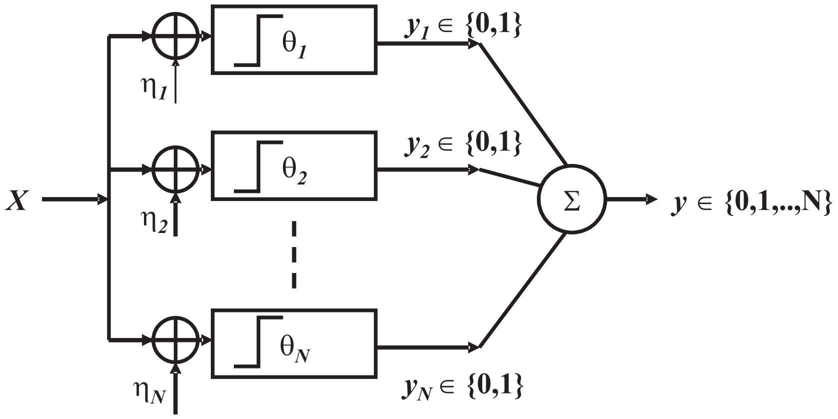

Figure 1. Array (population) of McCulloch-Pitts neuorns. Each of the K neurons, with threshold values $\theta_n,\ n = 1, .., K$, is subject to independent additive noise, $\eta_n,\ n = 1, .., K$. The common input to each neuron is a random signal, X, which in the results presented here is assumed to be Gaussian. Each neuron thresholds its input to produce a binary signal, y_n. The overall output is the sum of each y_n.

Each neuron is connected to a common random signal, x, but has an independent noise source η_n. The response of each neuron is governed by

$$y_n = \Theta(x + \eta_n), \quad n = 1, .., K, \tag{1}$$

where $\Theta(.)$ is the Heaviside function and the response of the population, Y is

$$Y = \sum_{n=1}^{K} y_n. \tag{2}$$

If the input signal has probability density function $P_X(x)$, then the mutual information between the input and output is given by [5]

$$I(X;Y) = -\sum_{i=0}^{K} P_Y(i) \log_2 P_Y(i) - \left(-\int_{-\infty}^{\infty} P_X(x) \sum_{i=0}^{K} P_{Y|X}(i|x) \log_2 P_{Y|X}(i|x) dx \right), \tag{3}$$

where

$$P_Y(n) = \int_{-\infty}^{\infty} P_{Y|X}(n|x) P_X(x) dx.$$

Therefore for a given signal distribution, the mutual information depends entirely on the conditional probabilities, $P_{Y|X}(n|x)$. If the noise distribution is specified, the only free variables are the population size, K, and the threshold values. Let $\hat{P}_n$ be the probability of neuron n 'spiking' in response to signal value x. Then

$$\hat{P}_n = \int_{\theta_n - x}^{\infty} f_\eta(\eta) d\eta = 1 - F_\eta(\theta_n - x), \tag{4}$$

where F_η is the cumulative distribution function of the noise and $n = 1, .., K$.

Given a noise density and arbitrary threshold values, $\hat{P}_n$ can be calculated exactly for any value of x from (4). These can then by used in an algorithm specified in [3] for calculating the conditional probabilities, $P_{Y|X}(n|x)$.

We now aim to find the threshold settings that maximize the mutual information as the input SNR varies. This goal can be formulated as the following nonlinear optimization problem:

$$\begin{aligned} \text{Find:} \quad & \max_{\{\theta_n\}} \quad I(X;Y) \\ \text{subject to:} \quad & \{\theta_n\} \in \mathbb{R}^K. \end{aligned} \tag{5}$$

It is necessary to solve (5) numerically using standard unconstrained non-linear optimization methods such as the conjugate gradient method [4]. However, note that the objective function is not convex, and there exist a number of local optima. This problem can be overcome by employing random search techniques such as simulated annealing. We present here in Figs. 2-5 results for the optimal thresholds obtained by solving Problem (5) for Gaussian signal and Gaussian noise, as a function of $\sigma := \frac{\sigma_\eta}{\sigma_x}$, where σ_η and σ_x are the noise and signal standard deviation respectively; we have set $\sigma_x = 1$ for all results. In Figs. 2, 3, 4, 5 we have $K = 3, 4, 5, 15$ respectively [1].

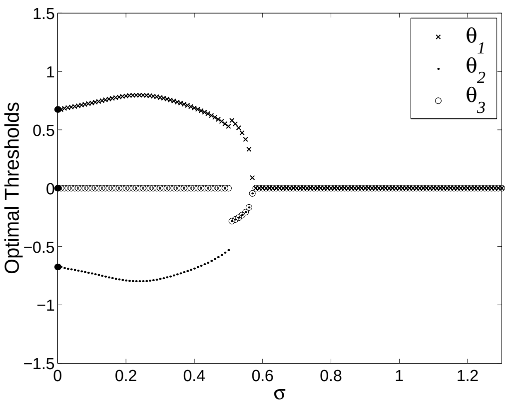

Figure 2. Optimal thresholds against noise intensity, σ, for $K = 3$. Reproduced from Ref. [1]

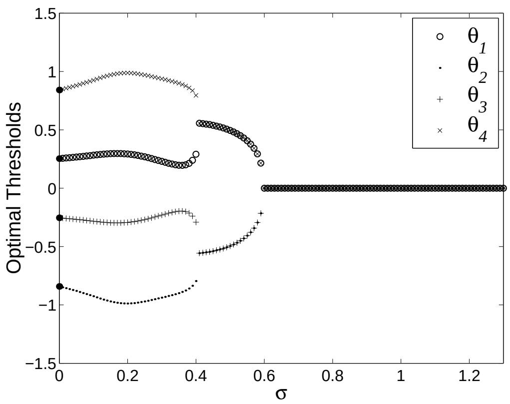

Figure 3. Optimal thresholds against noise intensity, σ, for $K = 4$. Reproduced from Ref. [1]

It can be observed that the dependence of the optimal threshold values on the noise level is similar for all results. For large σ it is optimal to set all the thresholds equal to the signal mean which is zero in this case. As the noise is reduced this optimal solution bifurcates, at a value σ_1, in favour of grouping all the thresholds at two distinct values. Further reduction in the noise level results in three distinct values becoming optimal. This pattern is repeated as the noise is reduced further. In general it is observed that the number of distinct threshold values increases as $1 \to 2 \to 3 \to \ldots \to K$ as the noise level is reduced. It should be noted that the exact values of the noise at which the transitions occur (e.g. σ_1) depends on the size of the population size

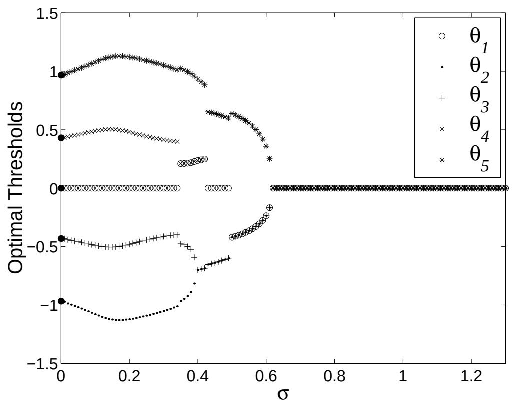

Figure 4. Optimal thresholds against noise intensity, σ, for $K = 5$. Reproduced from Ref. [1]

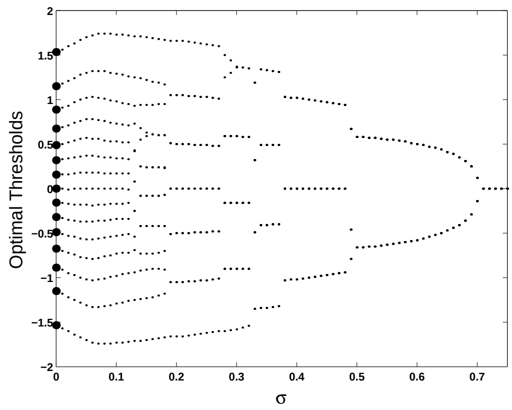

Figure 5. Optimal thresholds against noise intensity, σ, for $K = 15$. Reproduced from Ref. [1]

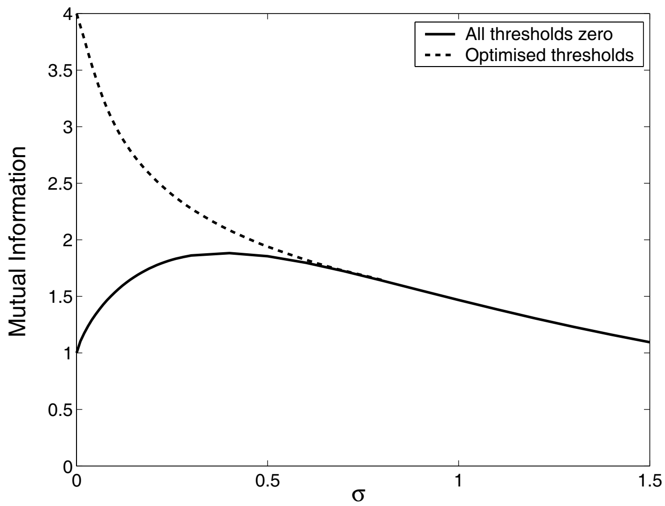

Figure 6. Plot of $I(X, Y)$ against noise intensity, σ, for $K = 15$, the optimal thresholds (dashed line) and $\theta_i = 0 \;\forall\; i$ (solid line). The solid line shows the suprathreshold stochastic resonance (SSR) effect [5, 3]. Reproduced from Ref. [1]

K. This problem is discussed in Chapter 8 of Ref. [1]. For the limit $K \to \infty$, it is observed that an asymptotic limit of σ_1 exists, but that its value is dependent on the distributions of the signal and noise. For example, for Gaussian signal and noise the limiting σ_1 is about 0.8, while for logistic signal and noise it is about 1.2. However, qualitatively, the pattern is always the same.

These results raise a number of interesting points. First, the structure of the optimal threshold distribution is discrete. Numerical experiments indicate that this bifurcational structure is not changed even in the limit $K \to \infty$. Consequently, the clustering of thresholds into distinct groups is not a consequence of the finite number of neurons but is rather a property of the system (channel) itself. Second, the existence of these distinct groups indicate an optimal coding strategy is to group neurons with similar thresholds together into *subpopulations*; the number of distinct subpopulations depends on the level of noise. Third, the optimal design solution is strongly dependent on the level of the noise. Indeed, this point is graphically highlighted in Fig. 6.

In Fig. 6 the MI is plotted as a function of the noise level for two different cases i) for the optimal threshold values (dashed line) shown in Fig. 5 and ii) for the optimal threshold solution for $\sigma \geqslant \sigma_1$ (solid line); i.e. the thresholds are held constant at $\theta_i = 0 \; \forall \; i$ as the noise level is varied. The two curves coalesce at σ_1. It can be observed that while the solution $\theta_i = 0 \; \forall \; i$ is optimal for $\sigma \geqslant \sigma_1$ this solution performs very badly in the limit $\sigma \to 0$. Indeed, it achieves a MI of only 1-bit compared to the optimal solution at $\sigma = 0$ which achieves 4-bits. Consequently, the correct choice of neural parameters (and hence the resulting neural code) is seen to strongly depend on the level of noise.

An interesting consequence of the sub-optimal nature of the solution $\theta_i = 0 \; \forall \; i$ at small σ is that it gives rise to a suprathreshold stochastic resonance [5, 3] i.e. the MI is maximised at a non-zero level of noise. Such counter-intuitive 'noise-benefits' rely on some aspect of a system being 'sub-optimal' [5, 1].

Some theoretical progress in calculating the optimal information can be made by utilising a connection between mutual information and Fisher information. Following [6, 7] we can write,

$$Y(x) = T(x) + \sqrt{V(x)}\xi \tag{6}$$

where $T(x)$ is the mean output for a give input x and $V(x)$ is the variance of the output conditioned on the input, and ξ is a random variable with zero mean and unit variance. For not too large $V(x)/N^2$ we can write

$$I(X;Y) \approx H(x) - \frac{1}{2}\int_x f_x(x) \log_2 \left(\frac{2\pi e}{J(x)} \right) dx, \tag{7}$$

where the Fisher information $J(x)$ is given by

$$J(x) \approx \frac{\left(\frac{dT(x)}{dx} \right)^2}{V(x)}. \tag{8}$$

Given that the input entropy $H(x)$ only depends on the statistics of the signal and not on the properties of the channel, maximisation of (7) can be seen to be equivalent to a minimisation of the function

$$F = -\langle \log_2 Z \rangle_x, \tag{9}$$

where $Z = J(x)$ and the average, $\langle . \rangle_x$ is taken over the input signal distribution. It is interesting to note that this looks similar to the minimisation of a Free Energy with a partition function Z.

These expressions can be simplified by noting that the responses of the neurons, $y_n(x)$, are (conditionally) independent. Consequently, the mean response and variance can be obtained by summing the individual means, $T_n(x)$, and variances,$\mathrm{Var}(y_n|x)$ [1]. Hence,

$$T(x) = \mathrm{E}(Y|x) = \sum_{n=1}^{N} T_n(x), \tag{10}$$

and

$$V(x) = \mathrm{Var}(Y|x) = \sum_{n=1}^{N} \mathrm{Var}(y_n|x) = \sum_{n=1}^{N} T_n(1 - T_n). \tag{11}$$

The final expression for the variance comes from the fact that $\hat{P}_n = T_n$ and the variance of each neuron is governed by Bernoulli statistics, which gives $\mathrm{Var}(y_n|x) = \hat{P}_n(1-\hat{P}_n)$. Consequently, substituting these expressions into Eq. (8) yields

$$Z = J(x) = \frac{\left(\sum_{n=1}^{N} \frac{dT_n(x)}{dx}\right)^2}{\sum_{n=1}^{N} T_n(1-T_n)}. \tag{12}$$

It is interesting to note that Eq. (12) only requires specification of the mean response $T_n(x)$, i.e. the tuning curve of each neuron. Furthermore, $T_n(x)$ is, itself, simply related to the cumulative noise distribution of each neuron as $T_n(x) = 1 - F_\eta(\theta_n - x)$. Consequently, it is largely the cumulative noise distribution that ultimately specifies the information flow in these systems.

A particularly simple expression for the mutual information can be derived under the assumption that the noise has a logistic distribution. In this case the cumulative distribution function of the noise is $F_\eta(\theta_n - x) = [1 + \exp(-\beta(\theta_n - x))]^{-1}$ where β is the inverse noise level. With this assumption it can be shown that T_n, via F_η, satisfies the differential equation associated with the logistic function, that is

$$\frac{dT_n}{dx} = \beta T_n(1-T_n). \tag{13}$$

Noting that the expression for the Fisher information simplifies to $J(x) = \beta^2 \sum_{n=1}^{N} T_n(1-T_n)$, Eq. (9) can be written

$$F(\{\theta_n\},\beta) = -\left\langle \log_2 \sum \frac{\beta^2}{4}\left[\cosh^2\left(\frac{\beta}{2}(\theta_n - x)\right)\right]^{-1} \right\rangle_x . \tag{14}$$

To obtain the optimal information it is now necessary to find the set $\{\theta_n\}$ that minimises $F(\{\theta_n\},\beta)$ for a given β. Although we have used a similar approach to find the location of the first bifurcation point for large N we have not been successful in describing the entire sequence of bifurcations. We speculate that further progress might be made by using a *replica* approach, but we have been unable to verify this.

References

[1] McDonnell M D, Stocks N G, Pearce C E M and Abbott D 2008 *Stochastic Resonance: From Suprathreshold Stochastic Resonance to Stochastic Signal Quantisation* (Cambridge: Cambridge University Press)

[2] Rose K, 1998,Deterministic Annealing for Clustering, Compression, Classification, Regression and Related Optimization Problems. *Proc. IEEE* **86**, 2210

[3] McDonnell M D, Abbott D and Pearce C E M 2002 *Microelectronics Journal* **33** 1079

[4] Nocedal J and Wright S J 1999 *Numerical Optimization* (New York: Springer-Verlag)

[5] Stocks N G 2000 *Phys. Rev. Lett.* **84** 2310

[6] Brunel N and Nadal J 1998 *Neural Computation* **10** 1731

[7] McDonnell M D and Stocks N G 2008 *Phys. Rev. Lett.* **101** 058103