

Optimal information transmission in nonlinear arrays through suprathreshold stochastic resonance

Mark D. McDonnell^{a,*}, Nigel G. Stocks^b, Charles E.M. Pearce^c, Derek Abbott^a

^a School of Electrical & Electronic Engineering and Centre for Biomedical Engineering (CBME), The University of Adelaide, Adelaide, SA 5005, Australia

^b School of Engineering, The University of Warwick, Coventry CV4 7AL, United Kingdom

^c School of Mathematical Sciences, The University of Adelaide, Adelaide, SA 5005, Australia

Received 10 June 2005; received in revised form 15 November 2005; accepted 16 November 2005

Available online 1 December 2005

Communicated by C.R. Doering

Abstract

We examine the optimal threshold distribution in populations of noisy threshold devices. When the noise on each threshold is independent, and sufficiently large, the optimal thresholds are realized by the suprathreshold stochastic resonance effect, in which case all threshold devices are identical. This result has relevance for neural population coding, as such noisy threshold devices model the key dynamics of nerve fibres. It is also relevant to quantization and lossy source coding theory, since the model provides a form of stochastic signal quantization. Furthermore, it is shown that a bifurcation pattern appears in the optimal threshold distribution as the noise intensity increases. Fisher information is used to demonstrate that the optimal threshold distribution remains in the suprathreshold stochastic resonance configuration as the population size approaches infinity. © 2005 Elsevier B.V. All rights reserved.

PACS: 05.40.Ca; 87.19.La; 89.70.+c; 02.40.Xx

Keywords: Information theory; Neural coding; Suprathreshold stochastic resonance; Quantization; Optimal quantization; Population coding; Bifurcations; Point density function

1. Introduction

A fascinating aspect of the behavior of populations of neurons is their capability of reliability in the presence of very low signal-to-noise ratios (SNRs) [1]. It has been established by many studies that improved performance in individual neurons can be achieved in the presence of large ambient noise, by a mechanism known as stochastic resonance (SR) [2–4]. However, the vast majority of studies on SR in static nonlinearities and neurons have been restricted to the case of subthreshold signals since, for a single device and suprathreshold stimuli, noise enhanced signal transmission disappears. For such systems, it has been pointed out that SR is a sub-optimal means

for improving system performance, since optimal performance can be gained by adjusting the threshold value [1,5]. Given that a central idea of current theoretical neural coding is that neural systems optimize information transmission [6], it is not clear how this is reconciled with SR as a neural coding mechanism.

Here, however, we examine the performance of SR in a noisy multi-threshold system, and show for the first time that a form of SR can be considered to be an optimal coding scheme. Such a study has relevance to both neural population coding [6,7], as well as signal processing and lossy source coding theory [8,9], and the growing overlap between the two fields [10]. The approach we take is to analyze the optimal encoding of a random input signal by a population of simple threshold devices, all of which are subject to additive *iid* (independent and identically distributed) input noise. Although this approach greatly simplifies the dynamics of realistic neural models, it does encapsulate the main nonlinearity: that of a threshold that generates an output spike when crossed. A natural measure to use for measuring information transmission in noisy neural systems, and one

* Corresponding author. Tel.: +61 8 8303 6296; fax: +61 8 8303 4360.

E-mail addresses: mmcdonne@eleceng.adelaide.edu.au (M.D. McDonnell), ngs@eng.warwick.ac.uk (N.G. Stocks), cpearce@maths.adelaide.edu.au (C.E.M. Pearce), dabbott@eleceng.adelaide.edu.au (D. Abbott).

which has been used extensively in computational neuroscience [1], is that of mutual information.

1.1. Suprathreshold stochastic resonance

Previously, it has been shown for suprathreshold signal levels in such a system that when all threshold values are equal to the signal mean, the mutual information between the input and output signals has a maximum value for a nonzero noise intensity. This phenomenon was termed suprathreshold stochastic resonance (SSR) to illustrate the fact that it is a form of stochastic resonance that is not restricted to subthreshold signals [11]. Subsequently, the effect was also shown to occur in FitzHugh–Nagumo model neurons [12] and applied to cochlear implant encoding [13].

1.2. Summary of new results

In this current Letter, we discuss the optimality of SSR by examining whether the mutual information can be increased by adjusting the set of thresholds, which we denote as $\{\theta_i\}$, $i = 1, \dots, N$, while keeping the noise constant. For a Gaussian signal and *iid* Gaussian noise, we show numerically that above a certain noise intensity the optimal threshold settings occur when all thresholds are equal to the signal mean. Hence, we show the SSR effect is a case of SR where, given such large ambient noise conditions, it is not possible to improve performance by adjusting the threshold values.

Furthermore, we show that as noise increases from zero—where the optimal thresholds are widely distributed across the dynamic range of the signal—the values of the optimal threshold settings go through a number of transitions where more and more threshold values accumulate to fewer and fewer points, in a series of bifurcations. Such a clustering of optimal thresholds to a small number of *point singularities* appears to persist if we let N approach infinity and hence, in this case there must be a transition from continuous to singular solutions of the optimal threshold values.

The paper is organized as follows. Section 2 discusses the model in which SSR occurs, extends it to arbitrary threshold values, and discusses the method we use to calculate the mutual information between its input and output signals. Section 3 then expresses our goal of optimizing the threshold values as a non-linear optimization problem. Section 4 presents the results of solving this problem, and discusses their main features. Finally, Section 5 describes a method for approximately solving our optimization problem in the event of a large number of threshold devices.

2. Array of threshold devices

The model we consider is shown in Fig. 1. This system consists of N static threshold devices, which all receive the same random input signal. This random signal is assumed to consist of a sequence of independent samples, x , drawn from a distribution with probability density function (pdf), $P(x)$.

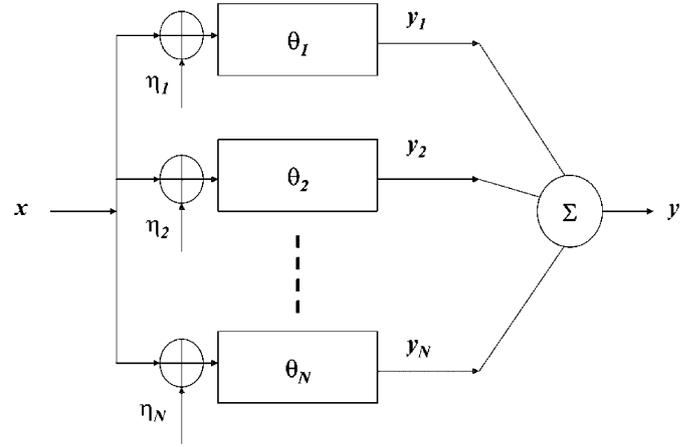


Fig. 1. Array of N noisy threshold devices. Each device receives the same input signal sample, and is subject to independent additive noise. The output from each device, y_i , is unity if the sum of the signal, x , and noise at its input, η_i , is greater than the corresponding threshold, θ_i , and zero otherwise. The overall output, y , is the sum of the individual outputs, $y = \sum_{i=1}^N y_i$, and is therefore a discrete encoding of the continuously valued input signal, x .

The i th device in the model is subject to continuously valued *iid* additive noise, η_i ($i = 1, \dots, N$), drawn from a probability distribution with pdf $R(\eta)$. Each noise signal is required to also be independent of the signal, x . For each individual threshold device, the output signal, y_i , is unity if the input signal, x , plus the noise on that threshold, η_i , is greater than the threshold value, θ_i . The output signal is zero otherwise. The outputs from each threshold device, y_i , are summed to give the overall system output signal, y . Hence, y is a discrete signal that can take integer values, $n \in \{0, \dots, N\}$, and is a lossy and nondeterministic quantization of the input signal, x . Thus, the overall output of the array of threshold devices is $y = \sum_{i=1}^N y_i$. This signal can also be expressed as a random variable dependent on x , in terms of the signum (sign) function, as

$$y(x) = \frac{1}{2} \sum_{i=1}^N \text{sign}[x + \eta_i - \theta_i] + \frac{N}{2}. \quad (1)$$

Now, mutual information is defined as

$$I(x, y) = H(y) - H(y|x),$$

where $H(y)$ is the entropy of the output signal, and $H(y|x)$ is the average conditional entropy of the output signal given the input signal [14]. If we define $P(n|x)$ as the conditional probability that the output is in state $y = n$, given the input signal is x , then

$$H(y|x) = - \int_{-\infty}^{\infty} P(x) \sum_{n=0}^N P(n|x) \log_2 P(n|x) dx,$$

where $P(x)$ is assumed known. For our model, since the output is discretely valued, we have the output entropy as

$$H(y) = - \sum_{n=0}^N P_y(n) \log_2 P_y(n),$$

where the output probability mass function is given by

$$P_y(n) = \int_{-\infty}^{\infty} P(n|x)P(x) dx, \quad n \in \{0, \dots, N\}. \quad (2)$$

Thus, the mutual information depends only on the conditional probability distribution of the output given the input, which we will refer to as the *transition probabilities*, and is given by

$$I(x, y) = - \sum_{n=0}^N P_y(n) \log_2 P_y(n) - \left(- \int_{-\infty}^{\infty} P(x) \sum_{n=0}^N P(n|x) \log_2 P(n|x) dx \right). \quad (3)$$

If we impose the constraint that all thresholds are set to the same value, the SSR effect can occur, in which case the mutual information has a maximum for nonzero noise intensity [11]. Such behavior has been shown to occur for noise with various pdfs such as Gaussian [11], uniform [15], Rayleigh, exponential [16] and Laplacian [17], as well as a variety of signal types [18,19]. The effect is maximized when the thresholds are all set to the signal mean [20]. SSR has also been shown to exist when measures other than mutual information are used [18,21,22].

Our objective here however, is to relax the constraint that all thresholds are identical, and to find the optimal threshold values, $\{\theta_i^*\}$. By optimal, we mean in the sense of maximum mutual information between the input signal, x , and output signal, y . In order to find the mutual information for arbitrary thresholds, we need a method for numerically calculating the transition probabilities.

2.1. Calculating the transition probabilities

Let $\hat{P}_i(x)$ be the probability of device i being “on” (that is, signal plus noise exceeding the threshold θ_i), for a given value of input signal, x . Then

$$\hat{P}_i(x) = \int_{\theta_i - x}^{\infty} R(\eta) d\eta = 1 - F_R(\theta_i - x), \quad (4)$$

where $F_R(\cdot)$ is the cumulative distribution function (cdf) of the noise and $i = 1, \dots, N$. For the particular case when the thresholds all have the same value, then each $\hat{P}_i(x)$ has the same value for all i and we have $P(n|x)$ given by the binomial distribution [11]. However, in general it is difficult to find analytical expressions for $P(n|x)$ and we will rely on numerics. Given any arbitrary N , $R(\eta)$, and $\{\theta_i\}$, the set of probabilities, $\{\hat{P}_i(x)\}$, can be calculated exactly for any value of x from Eq. (4), from which $P(n|x)$ can be found using an efficient recursive formula [16], and hence the mutual information calculated by numerical integration of Eqs. (2) and (3).

3. Problem formulation

Our problem of finding the threshold settings that maximize the mutual information can now be expressed as a nonlinear

optimization problem, where the cost function to maximize is the mutual information, and there are structural constraints on how the transition probabilities are obtained,

$$\begin{aligned} \text{Find:} & \quad \max_{\{P(n|x)\}} I(x, y) \\ \text{Subject to:} & \quad \{P(n|x)\} \text{ is a function of } \{\hat{P}_i(x)\}, \\ & \quad \sum_{n=0}^N P(n|x) = 1 \quad \forall x, \text{ and } \{\theta_i\} \in \mathbf{R}^N. \end{aligned} \quad (5)$$

This formulation is similar to previous work on clustering and neural coding problems solved using a method known as deterministic annealing [8,10,23]. In particular, the formulation reached in [10] can be expressed in a fashion identical to problem (5) with the structural constraints removed. Due to this difference though, the solution method used in that work to find the optimal conditional distribution, $\{P(n|x)\}$, cannot be used here, and instead we concentrate on optimizing the only free variable, the set $\{\theta_i\}$. This can be achieved with standard unconstrained optimization techniques.

However, the objective function is not convex in $\{\theta_i\}$, and there exist a number of local optima. This problem can be overcome by judicious selection of initial conditions or by employing random search techniques such as simulated annealing. The results presented below were obtained by solving problem (5) using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm, combined with many different initial conditions intended to track all local optima. The BFGS algorithm falls into the sub-category of optimization algorithms known as *quasi-Newton methods* [24].

4. Numerical results

We present results for the case of *iid* zero mean Gaussian signal and noise distributions. If the noise has variance σ_η^2 then we have

$$\hat{P}_i(x) = \frac{1}{2} + \frac{1}{2} \operatorname{erf}\left(\frac{x - \theta_i}{\sqrt{2}\sigma_\eta}\right),$$

where erf is the error function. Let $\sigma = \sigma_\eta/\sigma_x$, where σ_x^2 is the variance of the Gaussian signal. Thus, σ is the inverse of the square root of the input SNR. It is shown in [20] that when the signal and noise both have the same distribution, the mutual information is a function of the ratio, σ . Note for the case of $\sigma_\eta = 0$ that it is possible to analytically determine the optimal thresholds [16], and that in this case, each threshold has a unique value. Figs. 2, 3 and 4 show our results for the mutual information and optimal threshold settings, $\{\theta_i^*\}$, for $N = 15$ plotted against increasing σ . We have arbitrarily set $\sigma_x = 1$. For other values of σ_x , the actual values of $\{\theta_i^*\}$ differ proportionally to σ_x . Several interesting features are present in these results.

Firstly, Fig. 2 shows that the mutual information obtained with the optimal thresholds is strictly decreasing with increasing σ . This means that no SR effect is seen for optimized thresholds. The mutual information for SSR is also shown for comparison. Note that for $\sigma = 0$, SSR provides only one bit

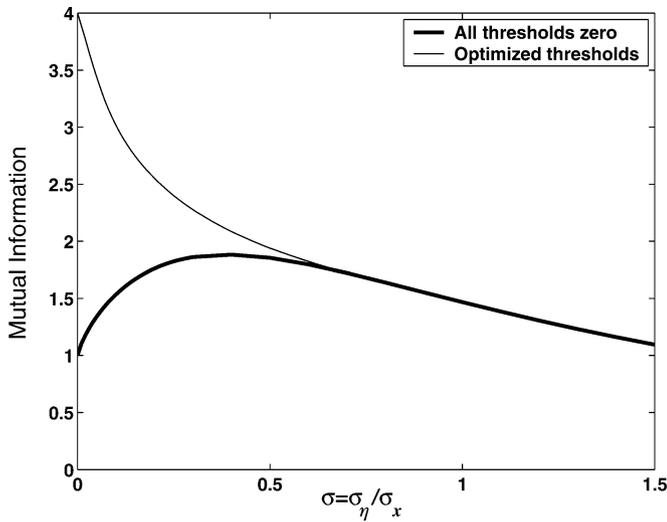


Fig. 2. Plot of mutual information, $I(x, y)$, against σ for $N = 15$ and Gaussian signal and noise. The thick line shows the mutual information obtained for the SSR situation of all thresholds equal to the signal mean. The thin line shows the mutual information that occurs for optimized thresholds. For small σ , optimizing the thresholds provides a far higher mutual information than SSR. However, as σ increases, the difference in mutual information between the two cases decreases, until for sufficiently large σ , SSR becomes optimal.

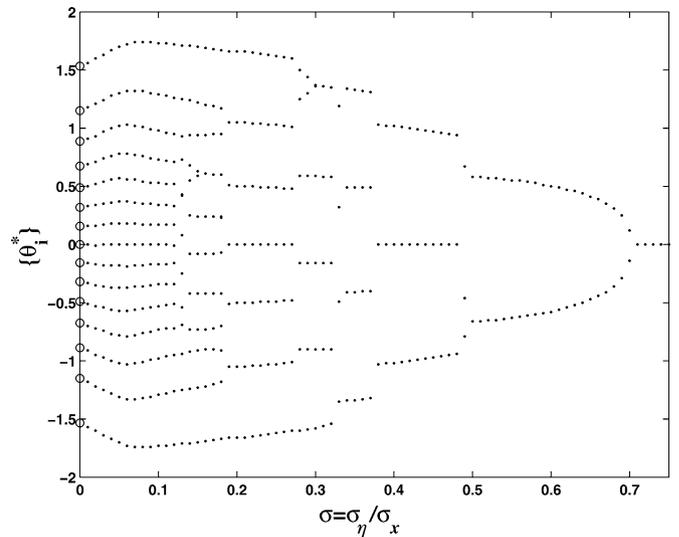


Fig. 3. Plot of optimal thresholds, $\{\theta_i^*\}$ against σ for $N = 15$ and Gaussian signal and noise, with $\sigma_x = 1$. The optimal noiseless thresholds ($\sigma = 0$) are shown with circles. Otherwise, the optimal thresholds are shown with dots. As with $\sigma = 0$, for small σ all optimal threshold values are unique. However, as the noise intensity, σ , increases, bifurcations occur so that more and more thresholds cluster to fewer unique values. For sufficiently large σ , there is a single cluster, and all optimal thresholds have the same value as the signal mean. The bifurcations are due to the presence of many local optima. The local optimum which provides the global optimum changes with σ , and therefore most bifurcations are discontinuous. Thus, it does not make sense to join up the paths of individual thresholds with increasing σ , as individual thresholds within clusters cannot be differentiated.

per sample, while if the thresholds are optimized the mutual information is 4 bits per sample, since the output entropy is $\log_2(N + 1) = 4$. However, as σ increases, the difference in mutual information between the two cases decreases, until for $\sigma \simeq 0.7$, SSR becomes optimal. While the fact that the mutual information decreases with increasing σ means that there is no advantage to be gained by increasing the noise level in such an optimized system, the fact that SSR is optimal for sufficiently large noise is still highly significant. Unlike a single threshold device, where the optimal threshold value is always at the signal mean regardless of the noise level and therefore SR can never occur, this result shows that when the noise is large, changing the thresholds from a situation where SR can occur gains no advantage.

The actual optimal thresholds are shown in Figs. 3 and 4. From Fig. 3 we observe, firstly, that for very small noise, the optimal thresholds are consistent with the optimal noiseless values. There does not appear to be a discontinuity in the optimal thresholds as the noise intensity increases from zero to some small nonzero value. However, the most striking feature is the fact that *bifurcations* are present. For σ between zero and some critical value greater than zero, the optimal placement of the 15 thresholds are at unique values. However, for larger σ , a bifurcation occurs, and some of the optimal thresholds coincide to the same value. As σ increases, more bifurcations occur, and increasing fractions of the total number of thresholds tend to cluster to fewer identical values. We will refer to the number of thresholds with the same value as the *size* of a cluster, and the actual threshold value of those thresholds as the *value* of a cluster. Fig. 4 indicates with stem plots the fraction of the total number of thresholds at each cluster value for various values of σ .

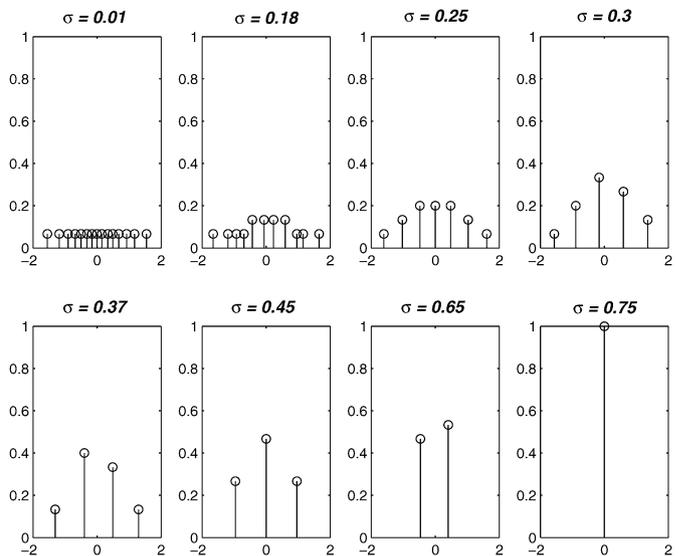


Fig. 4. Threshold point density function, $\lambda(x, \sigma)$, obtained for $N = 15$ and various values of σ . The y-axes give the fraction of thresholds in each cluster, $v(j, \sigma)$ and the x-axis gives the cluster values, $x = \theta_j$. It is clear that as σ increases, the number of clusters decreases.

This tendency of the optimal thresholds to form clusters at identical values leads to regions of asymmetry about the x -axis since, where for example there are two clusters of size seven and eight, then the value of the cluster of size seven is larger in magnitude than the value of the cluster of size eight. Note that in such regions of asymmetry, there are two globally opti-

mal threshold vectors. The second global solution is simply the negative of the set of thresholds in the first global solution, that is, the mutual information evaluated at $\{\theta^*\}$ is equal to the mutual information evaluated at $-\{\theta^*\}$. This result stems from the fact that both the signal and noise pdfs are even functions.

For most of the bifurcations, the number of clusters decreases with increasing σ . Sometimes a continuous bifurcation occurs as more than one cluster converges to the same value, as σ increases, to form a larger, merged cluster. On other occasions a discontinuous bifurcation occurs, and two clusters with completely different values merge to form a larger cluster with a value somewhere between the two values of the two merging clusters. It does not appear possible for the number of clusters to increase with increasing σ . However, further bifurcations can occur within a region of σ with k clusters, when the sizes and values of the k clusters change. For example, at $\sigma = 0.49$, there are three clusters of size $\{3, 5, 7\}$, while for $\sigma \in [0.38, 0.48]$, there are three clusters of size $\{4, 7, 4\}$.

The bifurcation structure is quite surprising, but appears to be fundamental to the problem type, as we have found similar bifurcation structures in the optimal threshold settings for measures other than mutual information, including correlation coefficient and mean square error distortion (error variance), and other signal and noise distributions. Finally, it is evident that above a certain value of σ the SSR situation is optimal. That is, the optimal quantization for large noise is to set all thresholds to the signal mean.

4.1. Point density function

We now mathematically describe the observations made above. For the purposes of optimization, the ordering of the optimal threshold vector, $\{\theta_i^*\}$, is not important. However, to simplify the mathematical description, we now introduce an ordered sequence notation for the optimal thresholds. Specifically, we label the i th optimal threshold value as θ_i^* , so that the sequence $(\theta_i^*)_{i=1}^N$ is nondecreasing. In the absence of noise, it is straightforward to show that each optimal threshold is given by $\theta_i^* = F_x^{-1}\left(\frac{i}{N+1}\right)$, where $F_x^{-1}(\cdot)$ is the inverse cdf of the signal.

We now introduce a concept used in the theoretical analysis of high resolution quantizers in information theory—that of a quantizer *point density function*, $\lambda(x)$, defined over the same support variable as the source pdf, $P(x)$ [9]. The point density function has the property that $\int_x \lambda(x) dx = 1$, and usually is only used in the context where the number of thresholds is very large. In this situation, the point density function gives the density of thresholds across the support of the signal pdf.

We observe from Fig. 4 that our empirically optimal threshold sequence, $(\theta_i^*)_{i=1}^N$, can have at most $k(\sigma)$ unique values, where $1 \leq k \leq N$. When bifurcations occur as σ increases, $k(\sigma)$ may either decrease or remain constant.

We now denote $v(j, \sigma)$ as the fraction of the total thresholds in the j th cluster at noise intensity σ , where $j \in \{1, \dots, k(\sigma)\}$, so that $\sum_{j=1}^{k(\sigma)} v(j, \sigma) = 1$. Thus, $v(j, \sigma)$ is the size of the j th cluster divided by N . Denote the value of the j th cluster as Θ_j , so that the size of the cluster at $x = \Theta_j$ is $Nv(j, \sigma)$. As with the

ordered optimal threshold sequence, we can define an ordered sequence of cluster values as $(\Theta_j)_{j=1}^{k(\sigma)}$. Unlike the optimal threshold sequence, this sequence is strictly increasing with j .

We are now able to write a point density function as a function of σ to describe our empirically optimal threshold configuration. This is

$$\lambda(x, \sigma) = \sum_{j=1}^{k(\sigma)} v(j, \sigma) \delta(x - \Theta_j), \quad (6)$$

where $\delta(\cdot)$ is the delta function. We also note that $\int_{x=-\infty}^a \lambda(x, \sigma) dx$ is the fraction of thresholds with values less than or equal to a , and that $\int_{x=-\infty}^{\infty} \lambda(x, \sigma) dx = 1$. For the special case of $\sigma = 0$ we can write the analytically optimal point density function as

$$\begin{aligned} \lambda(x, 0) &= \sum_{j=1}^N v(j, 0) \delta(x - \Theta_j) \\ &= \sum_{i=1}^N \frac{1}{N} \delta\left(x - F_x^{-1}\left(\frac{i}{N+1}\right)\right). \end{aligned} \quad (7)$$

4.2. Explanation of the bifurcations

The discontinuous bifurcations in the optimal threshold diagram are due to the presence of many locally optimal threshold configurations. In fact, numerical experiments find that for every value of σ , there is at least one locally optimal solution—that is a set of threshold values giving a gradient vector of the mutual information with respect to $\{\theta_i\}$ of zero—corresponding to every possible *integer partition* of N . For each partition, there are as many locally optimal solutions as there are unique orderings of that partition. For small σ , all of these local optima are unique. As σ increases, more and more of these local optima bifurcate continuously to become coincidental with other local optima. For example, a local optimum corresponding to $k = 3$ clusters, with $\{Nv(j, \sigma)\} = \{4, 7, 4\}$ might have Θ_2 and Θ_3 converge to the same value with increasing σ . At the point of this convergence, a bifurcation occurs, and the local optimum becomes one consisting of $k = 2$ clusters, with $\{Nv(j, \sigma)\} = \{7, 8\}$.

5. Behavior for large N

We can make some analytical progress on the solution to problem (5) by allowing the population size N to become very large, a case that is biologically relevant [25]. Hence, if we let $N \rightarrow \infty$, and divide the optimal threshold sequence by N , then the result approximates a strictly nondecreasing, function, $\Theta(z)$, defined on the continuous interval $z \in [0, 1]$. For the noiseless case $\Theta(z) = F_x^{-1}(z)$, so that

$$\lim_{N \rightarrow \infty} \lambda(x, 0) = \int_{z=0}^1 \delta(x - F_x^{-1}(z)) dz = P(x),$$

that is, the point density function is the pdf of the signal.

However, for nonzero noise, our numerical solutions of problem (5) indicate that even for very large N the clusters and bifurcation structure persists. Hence, if we assume that this is the case also for infinite N , there must be a transition at some σ from a continuously valued to a discretely valued optimal $\Theta(z)$. This is the reason that we claim the optimal threshold distribution contains point singularities for $\sigma > 0$.

Furthermore, our numerical results indicate that the locations of each bifurcation tend to converge to the same value of σ as N increases. Under the assumption that this holds for infinite N , we are able to make use of an approximation to the mutual information to find the location of the final bifurcation, that is the smallest noise intensity for which SSR is the optimal coding strategy.

This approximation relies on an expression for a lower bound on the mutual information involving the Fisher information, $J(x)$, and the entropy of an efficient estimator for x . Fisher information has previously been used to study SSR [18,26] and is a measure of how well the input signal, x , can be estimated from a set of N observations. In the limit of large N , the entropy of an efficient estimator approaches the entropy of the input signal, and if the distribution of this estimator is Gaussian then the lower bound becomes asymptotically equal to the actual mutual information [6,26]. This means that the mutual information can be written as

$$I(x, y) = H(x) - 0.5 \int_x P(x) \log_2 \frac{2\pi e}{J(x)} dx. \quad (8)$$

For a zero mean Gaussian input signal Eq. (8) becomes

$$I(x, y) = 0.5 \int_x P(x) \log_2(\sigma_x^2 J(x)) dx. \quad (9)$$

The Fisher information for a given input sample, x , for the system in Fig. 1 is

$$J(x) \approx \frac{(\sum_{i=1}^N \frac{d\hat{P}_i(x)}{dx})^2}{\sum_{i=1}^N \hat{P}_i(x)(1 - \hat{P}_i(x))} = \frac{(\sum_{i=1}^N R(\theta_i - x))^2}{\sum_{i=1}^N \hat{P}_i(x)(1 - \hat{P}_i(x))}.$$

Thus, $\Theta(z)$ can be obtained by maximizing Eq. (9), given assumptions on values of $k(\sigma)$ and $v(j, \sigma)$. Our empirical results for small N show that for σ just smaller than the first bifurcation, $v(1, \sigma) = v(2, \sigma) = 0.5$, and that $\lambda(x, \sigma) = 0.5\delta(x - t) + 0.5\delta(x + t)$, where $t \geq 0$. Under the assumption that this holds for very large N , it is straightforward to numerically find the value of t that maximizes Eq. (9) for any given σ . This maximization finds that the asymptotic location of the first bifurcation is at $\sigma \simeq 0.865$. This has been verified without recourse to the Fisher information approximation for N up to 20000.

Note that this value of σ corresponds to an input SNR in the order of 0 dB. This is particularly interesting given that SNRs of about this magnitude are typical for sensory neural coding [1], and indicates that perhaps sensory neural populations have evolved to make use of this noise. If the noise signals in a number of otherwise identical neurons are independent, then as demonstrated in [12], the SSR effect can be exploited to overcome the noise, while simultaneously encoding the signal in a quantized manner. Furthermore as demonstrated in principle

here, for sufficiently large noise, it is optimal in this situation for all neurons to be identical, rather than to have different threshold values.

6. Conclusion

To summarize, we have shown that the optimally quantized encoding of a Gaussian input signal by an array of independently noisy threshold devices contains point singularities in its threshold distribution, the number of which decreases in a series of bifurcations as the noise intensity increases. We have also found that for large enough noise, the optimal encoding is for all thresholds to be equal to the signal mean. This shows that SSR is a form of threshold-based SR that can be optimal. Finally, a Fisher information approach has shown that for very large population sizes, and Gaussian signal and noise, the noise intensity at which SSR becomes optimal converges to $\sigma \simeq 0.865$, which means an input SNR in the order of 0 dB. This result corresponds well with measurements of the SNR in real sensory neurons.

Acknowledgements

The authors wish to acknowledge funding support from the Australian Research Council, The Australian Academy of Science, the Cooperative Research Centre for Sensor Signal and Information Processing (Australia), and the Leverhulme Trust, grant number F/00 215/J. We would also like to thank the anonymous referees for their valuable comments, which have greatly improved the clarity of this Letter.

References

- [1] M. DeWeese, W. Bialek, *Nuovo Cimento* 17 (7–8) (1995) 733.
- [2] S. Fauve, F. Heslot, *Phys. Lett. A* 97 (1–2) (1983) 5.
- [3] L. Gammaitoni, P. Hänggi, P. Jung, F. Marchesoni, *Rev. Mod. Phys.* 70 (1) (1998) 223.
- [4] G.P. Harmer, B.R. Davis, D. Abbott, *IEEE Trans. Instrum. Meas.* 51 (2) (2002) 299.
- [5] D. Abbott, *Chaos* 11 (3) (2001) 526.
- [6] N. Brunel, J. Nadal, *Neural Comput.* 10 (1998) 1731.
- [7] A. Pouget, P. Dayan, R.S. Zemel, *Annu. Rev. Neurosci.* 26 (2003) 381.
- [8] K. Rose, *Proc. IEEE* 86 (11) (1998) 2210.
- [9] R.M. Gray, D.L. Neuhoff, *IEEE Trans. Inf. Theory* 44 (6) (1998) 2325.
- [10] A.G. Dimitrov, J.P. Miller, *Network* 12 (4) (2001) 441.
- [11] N.G. Stocks, *Phys. Rev. Lett.* 84 (11) (2000) 2310.
- [12] N.G. Stocks, R. Mannella, *Phys. Rev. E* 64 (2001) 030902(R).
- [13] N.G. Stocks, D. Allingham, R.P. Morse, *Fluct. Noise Lett.* 2 (3) (2002) L169.
- [14] T.M. Cover, J.A. Thomas, *Elements of Information Theory*, John Wiley and Sons, New York, 1991.
- [15] N.G. Stocks, *Phys. Lett. A* 279 (5–6) (2001) 308.
- [16] M.D. McDonnell, D. Abbott, C.E.M. Pearce, *Microelectron. J.* 33 (12) (2002) 1079.
- [17] M.D. McDonnell, D. Abbott, Signal reconstruction via noise through a system of parallel threshold nonlinearities, in: *Proc. 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Montreal, Canada, 17–21 May, 2004, vol. 2, 2004, pp. 809–812.
- [18] D. Rousseau, F. Duan, F. Chapeau-Blondeau, *Phys. Rev. E* 68 (2003) 031107.
- [19] F. Martorell, M.D. McDonnell, D. Abbott, A. Rubio, Generalized noise resonance: Using noise for signal enhancement, in: D. Abbott, S.M.

- Bezrukov, A. Der, A. Sánchez (Eds.), Proc. SPIE Fluctuations and Noise in Biological, Biophysical, and Biomedical Systems II, vol. 5467, 2004, pp. 163–174.
- [20] N.G. Stocks, Phys. Rev. E 63 (4) (2001) 041114.
- [21] M.D. McDonnell, D. Abbott, C.E.M. Pearce, Fluct. Noise Lett. 2 (3) (2002) L205.
- [22] D. Rousseau, F. Chapeau-Blondeau, Phys. Lett. A 321 (2004) 280.
- [23] K. Rose, E. Gurewitz, G.C. Fox, Phys. Rev. Lett. 65 (8) (1990) 945.
- [24] J. Nocedal, S.J. Wright, Numerical Optimization, Springer-Verlag, New York, 1999.
- [25] S. Wu, S. Amari, H. Nakahara, Neurocomputing 44–46 (2002) 697.
- [26] T. Hoch, G. Wenning, K. Obermayer, Phys. Rev. E 68 (3) (2003) 011911.