

Optimal coding of a random stimulus by a population of parallel neuron models

Mark D. McDonnell^{*a}, Nigel G. Stocks^b and Derek Abbott^a,

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

^bSchool of Engineering,
The University of Warwick, Coventry CV4 7AL, UK

ABSTRACT

We examine the question of how a population of independently noisy sensory neurons should be configured to optimize the encoding of a random stimulus into sequences of neural action potentials. For the case where firing rates are the same in all neurons, we consider the problem of optimizing the noise distribution for a known stimulus distribution, and the converse problem of optimizing the stimulus for a given noise distribution. This work is related to suprathreshold stochastic resonance (SSR). It is shown that, for a large number of neurons, the SSR model is equivalent to a single rate-coding neuron with multiplicative output noise.

Keywords: neural coding, population coding, suprathreshold stochastic resonance, information theory, noisy neurons

1. INTRODUCTION

Biological sensory neurons communicate information about stimuli to the brain along nerve fibres via short duration electrical pulses known as ‘spikes’, or, more formally, *action potentials*. Different methods are used in different sensory system to encode information, such as the average rate of action potentials, and the timing between consecutive action potentials. Another important method is ‘place coding,’ where the physical position of the neuron producing action-potentials indicates important information about a stimulus.

Here, we are interested in the neural communication method known as *rate coding*. We relate recent results¹ on a general model known as the Suprathreshold Stochastic Resonance (SSR) model^{2–13} to rate coding in sensory neural populations. Information theory is used to demonstrate that the SSR model is equivalent to a large population of sensory neurons that can overcome the effects of noise to produce an accurate rate-code representation of an analog random stimulus.

Suprathreshold Stochastic Resonance (SSR) is a form of noise or randomness enhanced signal transmission—i.e., Stochastic Resonance (SR)^{14, 15}—that occurs in a parallel array of identical, but independently noisy, non-linear threshold devices, including neuron models.^{5, 8, 13} We refer to this parallel array as the *SSR model*. Like other forms of SR, the performance of the SSR model is improved by an optimal level of noise. However, unlike conventional SR, this occurs not just for weak input signals, or subthreshold input signals, but for input signals with arbitrary magnitude.

The SSR model was originally intended to model neural populations, however it has not previously been related to optimal neural tuning curves. Discussion of optimal stimulus distributions has been restricted to the specific case of Gaussian noise.⁸ Here we demonstrate that if the SSR model is viewed as a population of parallel rate-coding neurons, information theory can be used to find optimal stimulus distributions, or optimal tuning curves for the model.

This paper is structured as follows. In Section 2 we provide details of the general SSR model. Section 3 defines what we mean by optimal neural tuning curves, in the context of rate coding. It also provides a brief

*mmdonne@eleceng.adelaide.edu.au; phone +61 8 8303 3165; fax +61 8 8303 4360

literature review of this topic and its relationship to information theory. Next, Section 4 explains how the SSR model may be viewed as a population of parallel neurons, with an overall tuning curve. This perspective is based on known expressions for the conditional probability distribution of the output of the model, for a given input value. Section 5 summarizes our previous optimization results for the input or noise distribution in the SSR model, before outlining how these results can be viewed in the neural coding context. The possible implications of our results are discussed in Section 6.

2. THE SSR MODEL

The system in which SSR occurs is a specific case of the block diagram shown in Fig. 1. For SSR, each of N parallel *identical*—i.e. $\theta_i = \theta \forall i$ —threshold device operates on nonidentical—due to independent additive noise—but correlated, signals. It has been shown for suprathreshold—meaning signals that can have values both above and below the threshold value, as opposed to entirely subthreshold—signal levels in such systems, that the mutual information between the input and output signals has a maximum value for nonzero noise intensity.²⁻⁶ This phenomenon was termed *suprathreshold* stochastic resonance to illustrate the fact that it is a form of stochastic resonance that is not restricted to subthreshold signals.

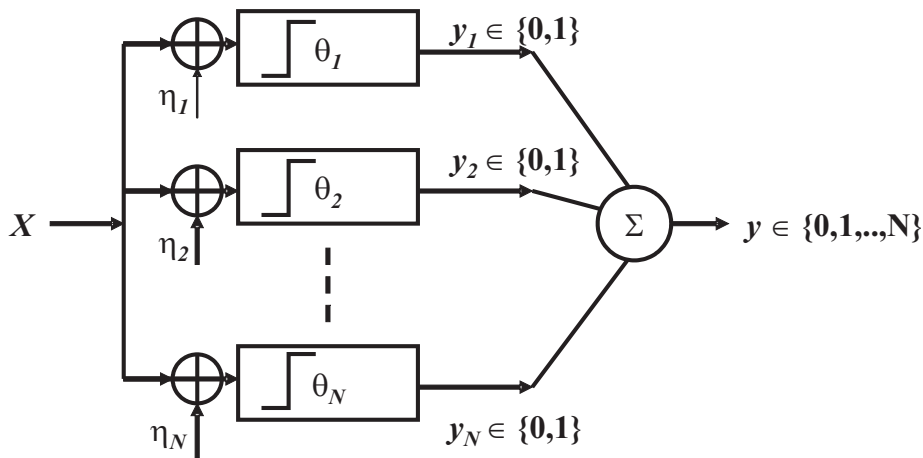


Figure 1. Generic Stochastic Quantization Model. Each of N threshold devices, with values θ_i , $i = 1, \dots, N$, is subject to independent additive noise, η_i , $i = 1, \dots, N$. The common input to each device is the random signal, x . A specific case of this model is the Suprathreshold Stochastic Resonance (SSR) model, which results when $\theta_i = \theta \forall i$.

Initial work on SSR considered the input to the array of thresholds to be a random signal, with either a Gaussian or uniform distribution, and the *iid* noise on each threshold to likewise be samples from a Gaussian or uniform distribution.^{2,3} The array was considered to be a communications channel, and the measure used was the mutual information between the input signal and the discretely valued output. It has since been shown that the SSR effect is very general¹² and can also be described as *stochastic quantization*—i.e. non-deterministic quantization—of a signal.^{10,11}

3. NEURONS, INFORMATION THEORY AND OPTIMAL TUNING CURVES

Empirical recordings of neural activity are often characterized by ‘firing rates,’ that is, the *average* number of ‘spikes’ observed while a given stimulus x , is held fixed. If this is repeated for a range of x , an average *tuning curve*, $T(x)$, can be found.¹⁶ Since $T(x)$ is an average, there will be some variance about $T(x)$, which would normally be classified as noise. If this noise is dependent on the actual value of x , it is called multiplicative noise. We note that if there is an upper limit, N , to the number of action potentials during duration, t , the tuning curve may be normalized to a firing rate between 0 and 1.

There have been many previous studies examining the rate of information transmission through a neuron or population of neurons.¹⁷⁻¹⁹ This approach requires an assumption that the input stimulus to a neuron is a random signal. The focus is usually on finding the maximum mutual information stimulus distribution. By definition, such a distribution is one that achieves *channel capacity*.

Less frequently, results have appeared discussing the optimal *tuning curve* for a neuron or neural population. This is equivalent to maximizing information transfer by optimizing the channel to suit a given stimulus distribution. To achieve this, one usually has to assume a form of channel model. For example, Brunel and Nadal,²⁰ suppose that the output, y , of a single neuron, in response to stimulus x , can be written as

$$y(x) = a(x) + \sqrt{b(x)}\xi, \quad (1)$$

where $a(x)$ and $b(x)$ are deterministic and invertible functions, and ξ is a zero mean, unit variance random variable. Note that this means that $y(x)$ is a random variable. With a channel of this form, the conditional mean of y given x is $E[y|x] = a(x)$, and the conditional variance is $\text{var}[y|x] = b(x)$.

The case where $b(x) = 1$ has been studied by Nadal and Parga,²¹ who show that the mutual information between x and y is maximized when the PDF of the stimulus, $f_x(x)$ is proportional to the derivative of $a(x)$ with respect to x . This result can be viewed two ways. Firstly, given a channel of the form of Eqn. (1), with $b(x) = 1$ and a specified $a(x)$, channel capacity is achieved with a stimulus distribution proportional to $\frac{da(x)}{dx}$. Alternatively, given a stimulus PDF, $f_x(x)$, optimality is achieved when

$$a(x) = A \int_{-\infty}^x f_x(\xi)d\xi + B, \quad (2)$$

where B is a constant.²⁰

The case where $a(x) = b(x)$ —i.e. a large-rate stimulus-dependant Poisson neuron—has been studied by Brunel and Nadal.²⁰ It is shown that the optimal transfer function is achieved when $a(x)$ —the Poisson rate, for a given stimulus value—satisfies

$$\sqrt{a(x)} = A \int_{-\infty}^x f_x(\xi)d\xi + B. \quad (3)$$

As an example, supposing the stimulus is uniformly distributed between 0 and 1, Eqn. (3) is satisfied if $a(x) = x^2$, $A = 2$ and $B = 0$, since $a'(x) = 2x$.²⁰

In Section 4 we discuss how the SSR model can be viewed as a neuron with a transfer function of the form of Eqn. 1.

4. THE TUNING CURVE FOR THE SSR MODEL

We assume that the input stimulus to the SSR model is a time varying random aperiodic signal with a stationary distribution. The analysis below is simplified by considering only a single value of the input stimulus, and then averaging over its probability distribution. When we consider the SSR model as a population of neurons, we represent the production of an action-potential in the i -th neuron, in response to input stimulus, x , by $y_i = 1$. When no action-potential is produced in response to x , we have $y_i = 0$. The overall population response, y , is the total number of action-potentials produced in response to x .

The conditional probability distribution, $P_{y|x}$ of the total number of action-potentials, given stimulus, x , is given by the binomial distribution,

$$P_{y|x}(n|x) = \binom{N}{n} P_{1|x}^n (1 - P_{1|x})^{(N-n)}, \quad (4)$$

where

$$P_{1|x} = 1 - F_\eta(\theta - x), \quad (5)$$

and $F_\eta(\cdot)$ is the cumulative distribution function (CDF) of the noise.

Therefore, using the well known expression for the mean of a binomial distribution, the mean conditional number of action potentials is

$$E[y|x] = NP_{1|x} = N - NF_{\eta}(\theta - x). \quad (6)$$

We also have the conditional variance as

$$\text{var}[y|X] = NP_{1|x}(1 - P_{1|x}) = N(1 - F_{\eta}(\theta - x))F_{\eta}(\theta - x). \quad (7)$$

From Eqns. (6) and (7), since F_{η} is a CDF, when the stimulus is at its mean value, the neural population has a firing rate of $N/2$, with a maximum variance of $0.25N$. When the stimulus is well below its mean, the firing rate approaches zero, with little variance. When the signal is well above its mean, the firing rate approaches saturation, i.e. $y = N$, with little variance.

The conditional mean and variance can be normalized, and made invariant to N , by dividing by the total number of neurons, N . Fig. 2 shows (a) the normalized mean and (b) the variance of the firing rate, as a function of stimulus intensity, x , for the SSR model, for the cases of uniform, cosine and Gaussian noise. These plots illustrate the sigmoidal nature of the tuning curve, and the fact that the variance is maximum at the stimulus mean.

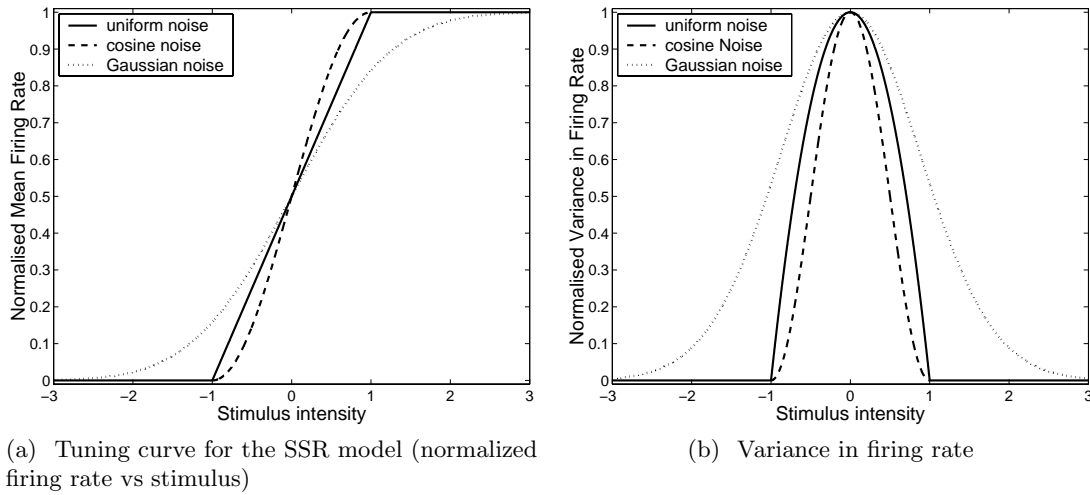


Figure 2. Normalised mean firing rate of the SSR model (a), and variance in firing rate (b), as a function of stimulus intensity, x , for the cases of uniform, cosine and Gaussian noise.

If the threshold value, θ , is equal to the mean of the stimulus, then when the stimulus is equal to its mean, this could be considered as *spontaneous firing* due to noise alone. Furthermore, each individual neuron in our model has the same characteristics as the population, and therefore exhibit spontaneous firing at the signal mean, increased firing when the signal is above its mean, and decreased firing when the signal is below its mean. This type of neuron is quite different to those discussed in previous work relating mutual information to Fisher information and optimal tuning curves, which instead focus on neurons with Poisson firing statistics.^{20, 22} Even though this means that tuning curves are sigmoidal, or at least increasing functions of the stimulus, the stimulus itself is predominantly subthreshold, meaning that spontaneous firing rates are quite low.

Recalling Eqn. (1), we now use Eqns. (6) and (7) to write an expression for the output of the SSR model,

$$y(x) = NP_{1|x} + \sqrt{NP_{1|x}(1 - P_{1|x})}\xi(x). \quad (8)$$

Note that here we have written the noise random variable, ξ , as a function of x , since we cannot assume that the output noise has the same distribution for every value of x . However, for large N , since the binomial distribution

approaches a Gaussian distribution, we are able to write

$$y(x) \simeq NP_{1|x} + \sqrt{NP_{1|x}(1 - P_{1|x})}\xi_G, \quad (9)$$

where ξ_G is a zero mean unity variance Gaussian random variable.^{1, 12} Regardless of whether the distribution of ξ is constant or not, we have highlighted that the output noise of our model is multiplicative with respect to the input signal, since the random part of the output is signal dependent.

4.1. Optimal Tuning Curves for the SSR model

Previously¹—see Section 5— we have shown that the mutual information in the SSR model for a large number of threshold devices, N , depends solely on N and an auxiliary PDF, $f_Q(\cdot)$. This PDF is that of the random variable describing the conditional *average* output of the SSR model, given that the input signal is $X = x$.

In this current paper, the general model¹ is translated into a neural population consisting of N identical *parallel* neurons. In this context $f_Q(\cdot)$ describes the PDF of the mean spike count distribution—or normalized tuning curve—of the entire population, for a given stimulus intensity.

Since the maximum value of the SSR model’s output is N , upon letting $\tau = \frac{\bar{y}}{N} \in [0, 1]$, then τ represents the *normalized* average rate of ‘firing’. If the mutual information through the SSR model is maximized, assuming large N , the optimal $f_Q(\cdot)$ is¹

$$f_Q^o(\tau) = \frac{1}{\pi\sqrt{\tau(1 - \tau)}}, \quad \tau \in [0, 1], \quad (10)$$

which is the PDF of the arcsine distribution, i.e. the beta distribution with parameters 0.5 and 0.5. This result holds *regardless* of whether the stimulus PDF is optimized for a given noise PDF or *vice versa*. This Beta distribution is bimodal, with the most probable values of the stimulus those near zero and unity. Similar results for an optimal input distribution in an information theoretic optimization of a neural system have previously been found numerically.²³

5. ANALYTICAL OPTIMIZATION OF STIMULUS AND NOISE DISTRIBUTIONS

Hoch *et. al.*²⁴ have shown that the Fisher information^{25, 26} for the SSR model is given by

$$J(x) = \left(\frac{dP_{1|x}}{dx} \right)^2 \frac{N}{P_{1|x}(1 - P_{1|x})}. \quad (11)$$

For large N , it is possible to write the mutual information for the SSR model in terms of the Fisher information as

$$I(x, y) = H(x) - 0.5 \int_{x=-\infty}^{x=\infty} f_x(x) \log_2 \left(\frac{2\pi e}{J(x)} \right) dx. \quad (12)$$

Eqn. (12) was first discussed in the context of the SSR model by Hoch *et. al.*²⁴ It has previously been utilized for neural modeling by Brunel and Nadal,²⁰ who showed that in the limit of large N , the mutual information in a system becomes equal to the mutual information between the input signal and an *efficient* Gaussian estimator for that signal.^{8, 20, 27–29}

We have subsequently shown¹ that Eqns. (11) and (12) lead to a simple, but general, sufficient condition for the maximum mutual information in the SSR model for large N . This condition exists due to the fact that the function

$$f_S(x) = \frac{\sqrt{J(x)}}{\int_{\phi} \sqrt{J(\phi)} d\phi} = \frac{\sqrt{J(x)}}{\pi\sqrt{N}} \quad (13)$$

can be proven to be a probability density function (PDF). Such a PDF that is a normalization of the square root of the Fisher information is known as *Jeffrey’s Prior*.^{30, 31} We have not found any previous work giving examples of Jeffrey’s prior for which $\int_{\phi} \sqrt{J(\phi)} d\phi$ can be integrated analytically, as is the case for the SSR model.

Using Eqn. (13), we can rewrite Eqn. (12) as

$$I(x, y) = 0.5 \log_2 \left(\frac{N\pi}{2e} \right) - D(f_x || f_S), \quad (14)$$

where $D(\cdot || \cdot)$ is relative entropy.

Since relative entropy is always non-negative, from Eqn. (14) a sufficient condition for achieving the large N channel capacity is that

$$f_x(x) = f_S(x) \quad \forall x, \quad (15)$$

with the resultant capacity as¹

$$C(x, y) = 0.5 \log_2 \left(\frac{N\pi}{2e} \right) \simeq 0.5 \log_2 N - 0.3956. \quad (16)$$

The derived sufficient condition of Eqn. (15) leads to two ways in which maximum mutual information can be achieved. In a neural coding context, these are (i) an optimal stimulus PDF for a given noise PDF, and (ii) an optimal noise PDF for a given stimulus PDF. Optimizing the noise PDF can be considered as optimizing the neural channel, i.e. the tuning curve.

5.1. Optimizing the stimulus

With $\theta = 0$, from Eqns. (15), (13) and (11), the optimal stimulus PDF—i.e. that which achieves maximum mutual information—can be expressed in terms of the noise PDF and CDF as¹

$$f_x^o(x) = \frac{f_\eta(-x)}{\pi \sqrt{F_\eta(-x)(1 - F_\eta(-x))}}. \quad (17)$$

5.1.1. Examples

Suppose the noise in the SSR model is uniformly distributed on the interval $[-\sigma_\eta/2, \sigma_\eta/2]$, i.e.,

$$f_\eta(\eta) = \frac{1}{\sigma_\eta}, \quad \eta \in [-\sigma_\eta/2, \sigma_\eta/2]. \quad (18)$$

Substituting Eqn. (18) and its associated CDF into Eqn. (17), we find the optimal stimulus PDF is

$$f_x^o(x) = \frac{1}{\pi \sqrt{\frac{\sigma_\eta^2}{4} - x^2}}, \quad x \in [-\sigma_\eta/2, \sigma_\eta/2]. \quad (19)$$

This is the PDF of a sine-wave with uniformly random phase and amplitude $\sigma_\eta/2$, and is illustrated in Fig. 3. Similar *numerical* results for an optimal stimulus distribution in an information theoretic optimization of a neural system have been found previously.²³ This means that our analytical results are potentially highly significant, given their generality.

Suppose the noise has a Gaussian distribution with mean zero and variance σ_η^2 ,

$$f_\eta(\eta) = \frac{1}{\sqrt{2\pi\sigma_\eta^2}} \exp\left(-\frac{\eta^2}{2\sigma_\eta^2}\right). \quad (20)$$

Substituting Eqn. (20) and its associated CDF into Eqn. (17), gives the optimal signal PDF, which cannot be simplified very much. However, we have previously verified¹ that the resultant PDF has the correct shape via previous numerical results of Hoch *et. al.*⁸ The analytically optimal signal PDF, $f_x^o(x)$, from Eqs. (20) and (17), with $\theta = 0$, is plotted in Fig. 4, as well as a Gaussian PDF with variance $0.25\pi^2$. The optimal signal PDF clearly has almost the same shape as the Gaussian PDF. This is quite different to the uniform noise case, where the optimal stimulus is the bimodal arcsine distribution.

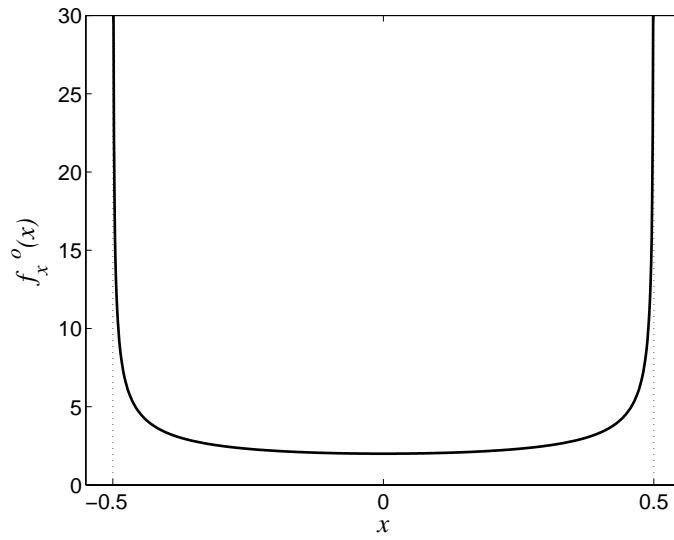


Figure 3. The optimal stimulus PDF, $f_x^o(x)$, for uniform noise with $\sigma_\eta = 1$, and threshold value $\theta = 0$. This is the PDF of the arcsine distribution, and is clearly bimodal.

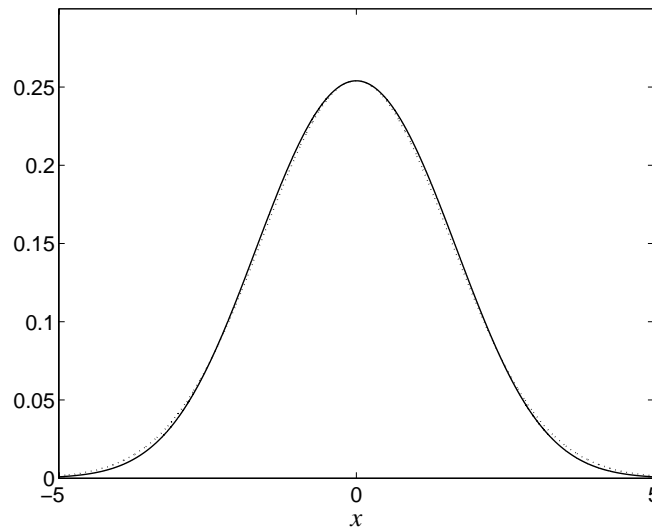


Figure 4. The optimal stimulus PDF, $f_x^o(x)$, for zero mean, unity variance Gaussian noise, and threshold value $\theta = 0$, is shown with a solid line. This was obtained from Eqn. (17). Superimposed with a dotted line is a Gaussian PDF with the same peak value as $f_x^o(x)$. Clearly the optimal stimulus PDF is close to being Gaussian, for Gaussian noise.

5.2. Optimizing the tuning curve via the noise distribution

We now assume knowledge of the stimulus distribution. We wish to maximize the mutual information by finding the optimal noise PDF. This is equivalent to optimizing the neural channel. For $\theta = 0$, the CDF corresponding to $f_S(\cdot)$, can be written in terms of the CDF of the noise distribution as¹

$$F_S(x) = 1 - \frac{2}{\pi} \arcsin \left(\sqrt{F_\eta(-x)} \right). \quad (21)$$

Applying the sufficient condition for optimality that $f_x(x) = f_S(x)$, we have $F_x(s) = F_S(x)$, and from Eqn. (21) the optimal noise CDF can be expressed in terms of the signal CDF as

$$F_\eta^o(x) = 0.5 + 0.5 \cos(\pi F_x(-x)). \quad (22)$$

Recalling Eqn. (6), where the noise CDF is related to a neural tuning curve, Eqn. (22) is equivalent to the optimal tuning curve for the SSR model.

This is a very interesting result in the context of neural coding and tuning curves. There is much evidence in the neuroscience literature to support the belief that cosine tuning curves can be optimal.^{16,22} For the special case of a stimulus uniformly distributed between $\phi \in [0, 2\pi]$, the optimal noise CDF can be written as

$$F_\eta^o(\phi) = 0.5 - 0.5 \cos\left(\frac{\phi}{2}\right), \quad \phi \in [0, 2\pi]. \quad (23)$$

This compares with the peaked tuning curve centered at π given by Bethge *et. al.*,²² which can be written as

$$T(\phi) = 0.5 + 0.5 \cos(\phi - \pi), \quad \phi \in [0, 2\pi], \quad (24)$$

if it is assumed that the maximum firing rate is unity and the minimum rate is zero. Clearly Eq. (23) has the same shape on $[0, 2\pi]$ as Eq. (24) does on $[0, \pi]$, but the former is a sigmoidal curve, while the latter has a peak.

6. DISCUSSION

Tuning curves nearly always imply that a large amount of spike rate variability exists for a fixed stimulus. Our analysis indicates that (i) sigmoidal neural tuning mechanisms might *require* the presence of such random fluctuations—without the randomness, tuning curves would not provide a near linear increase in rate with stimulus; and (ii) this randomness could be modeled as additive noise at the input to each neuron in a population. Note that it is not suggested that every neural system has a tuning curve that depends on randomness. However one possible sensory system where our modelling could apply is for hair cells in the inner ear. Hair cells are subject to substantial Brownian motion at the input,³² and synapse with groups of parallel identical afferent neurons.

Another interesting consequence of our results is that it turns out that whether we optimize the stimulus for a given noise distribution, or optimize the noise for a given stimulus, the optimal Fisher information can be written as a function of the stimulus PDF,¹

$$J^o(x) = N\pi^2(f_x(x))^2. \quad (25)$$

Therefore, when the mutual information is maximized, and N is large the Fisher information is constant iff the stimulus is uniformly distributed. This is another very interesting result in the context of neural coding, since the optimality of constant Fisher information in such a context has been previously studied by Bethge *et. al.*³³

7. CONCLUSIONS AND FUTURE WORK

We have shown that the basic SSR model, even though consisting of very primitive neuron models, has a number of properties that can be related to previous computation neuroscience research. Future work will expand our results to the case of non-identical threshold values. We will also expand on our result showing that the Fisher information for a uniformly distributed stimulus is constant, and relate this to results from estimation²⁵ and rate-distortion theory.²⁶

ACKNOWLEDGMENTS

The work was supported by the CASS Foundation Limited, the Australian Research Council, and EPSRC grant EP/C523334/1, and we thank them for their support.

REFERENCES

1. M. D. McDonnell, N. G. Stocks, and D. Abbott, "Optimal stimulus and noise distributions for information transmission via suprathreshold stochastic resonance," *Physical Review E* **75**, In Press, 2007.
2. N. G. Stocks, "Suprathreshold stochastic resonance in multilevel threshold systems," *Physical Review Letters* **84**, pp. 2310–2313, 2000.
3. N. G. Stocks, "Suprathreshold stochastic resonance: An exact result for uniformly distributed signal and noise," *Physics Letters A* **279**, pp. 308–312, 2001.
4. N. G. Stocks, "Information transmission in parallel threshold arrays: Suprathreshold stochastic resonance," *Physical Review E* **63**, Art. No. 041114, 2001.
5. N. G. Stocks and R. Mannella, "Generic noise enhanced coding in neuronal arrays," *Physical Review E* **64**, Art. No. 030902(R), 2001.
6. M. D. McDonnell, D. Abbott, and C. E. M. Pearce, "An analysis of noise enhanced information transmission in an array of comparators," *Microelectronics Journal* **33**, pp. 1079–1089, 2002.
7. M. D. McDonnell, D. Abbott, and C. E. M. Pearce, "A characterization of suprathreshold stochastic resonance in an array of comparators by correlation coefficient," *Fluctuation and Noise Letters* **2**, pp. L205–L220, 2002.
8. T. Hoch, G. Wenning, and K. Obermayer, "Optimal noise-aided signal transmission through populations of neurons," *Physical Review E* **68**, Art. No. 011911, 2003.
9. D. Rousseau, F. Duan, and F. Chapeau-Blondeau, "Suprathreshold stochastic resonance and noise-enhanced Fisher information in arrays of threshold devices," *Physical Review E* **68**, Art. No. 031107, 2003.
10. M. D. McDonnell, N. G. Stocks, C. E. M. Pearce, and D. Abbott, "Quantization in the presence of large amplitude threshold noise," *Fluctuation and Noise Letters* **5**, pp. L457–L468, 2005.
11. M. D. McDonnell, N. G. Stocks, C. E. M. Pearce, and D. Abbott, "Optimal information transmission in nonlinear arrays through suprathreshold stochastic resonance," *Physics Letters A* **352**, pp. 183–189, 2006.
12. M. D. McDonnell, N. G. Stocks, C. E. M. Pearce, and D. Abbott, *Stochastic Resonance: From Suprathreshold Stochastic Resonance to Stochastic Signal Quantisation*, Cambridge University Press, 2007 (In Press).
13. S. Durrant and J. Feng, "Suprathreshold stochastic resonance in neural processing tuned by correlation." *To be submitted*, 2007.
14. L. Gammaitoni, P. Hänggi, P. Jung, and F. Marchesoni, "Stochastic resonance," *Reviews of Modern Physics* **70**, pp. 223–287, 1998.
15. F. Moss, L. M. Ward, and W. G. Sannita, "Stochastic resonance and sensory information processing: a tutorial and review of application," *Clinical Neurophysiology* **115**, pp. 267–281, 2004.
16. P. Dayan and L. F. Abbott, *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*, The MIT Press, 2001.
17. A. Rapoport and W. J. Horvath, "The theoretical channel capacity of a single neuron as determined by various coding systems," *Information and Control* **3**, pp. 335–350, 1960.
18. R. Stein, "The information capacity of nerve cells using a frequency code," *Biophysics Journal* **7**, pp. 70–86, 1967.
19. F. Rieke, D. Warland, R. de Ruyter van Steveninck, and W. Bialek, *Spikes: Exploring the Neural Code*, MIT Press, Cambridge, MA., 1997.
20. N. Brunel and J. Nadal, "Mutual information, Fisher information and population coding," *Neural Computation* **10**, pp. 1731–1757, 1998.
21. J. Nadal and N. Parga, "Nonlinear neurons in the low noise limit: A factorial code maximizes information-transfer," *Network: Computation in Neural Systems* **5**, pp. 565–581, 1994.
22. M. Bethge, D. Rotermund, and K. Pawelzik, "Optimal neural rate coding leads to bimodal firing rate distributions," *Network: Computation in Neural Systems* **14**, pp. 303–319, 2003.

23. S. Schreiber, C. K. Machens, A. V. Herz, and S. B. Laughlin, "Energy-efficient coding with discrete stochastic events," *Neural Computation* **14**, pp. 1323–1346, 2002.
24. T. Hoch, G. Wenning, and K. Obermayer, "Adaptation using local information for maximizing the global cost," *Neurocomputing* **52-54**, pp. 541–546, 2003.
25. E. L. Lehmann and G. Casella, *Theory of Point Estimation*, Springer, New York, 1998.
26. T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley and Sons, New York, 1991.
27. B. S. Clarke and A. R. Barron, "Information-theoretic asymptotics of Bayes methods," *IEEE Transactions on Information Theory* **36**, pp. 453–471, 1990.
28. M. Stemmler, "A single spike suffices: The simplest form of stochastic resonance in model neurons," *Network: Computation in Neural Systems* **7**, pp. 687–716, 1996.
29. K. Kang and H. Sompolinsky, "Mutual information of population codes and distance measures in probability space," *Physical Review Letters* **86**, pp. 4958–4961, 2001.
30. J. J. Rissanen, "Fisher information and stochastic complexity," *IEEE Transactions on Information Theory* **42**, pp. 40–47, 1996.
31. E. T. Jaynes, *Probability Theory: The Logic of Science*, Cambridge University Press, Cambridge, UK, 2003.
32. J. F. Lindner, M. Bennett, and K. Wiesenfeld, "Stochastic resonance in the mechano-electrical transduction of hair cells," *Physical Review E* **72**, Art. No. 051911, 2005.
33. M. Bethge, D. Rotermund, and K. Pawelzik, "Optimal short-term population coding: When Fisher information fails," *Neural Computation* **14**, pp. 2317–2351, 2002.