Invited Paper

High Resolution Optimal Quantization for Stochastic Pooling Networks

Mark D. McDonnell*[a], Pierre-Olivier Amblard[b], Nigel G. Stocks[c], Steeve Zozor[b] and Derek Abbott[a],

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

[b]Laboratoire des Images et des Signaux (UMR CNRS 5083)
BP 46, 38 402 Saint Martin d'Hères cedex, France

[c]School of Engineering,
The University of Warwick, Coventry CV4 7AL, UK

ABSTRACT

Pooling networks of noisy threshold devices are good models for natural networks (*e.g.* neural networks in some parts of sensory pathways in vertebrates, networks of mossy fibers in the hippothalamus, ...) as well as for artificial networks (*e.g.* digital beamformers for sonar arrays, flash analog-to-digital converters, rate-constrained distributed sensor networks, ...). Such pooling networks exhibit the curious effect of suprathreshold stochastic resonance, which means that an optimal *stochastic* control of the network exists.

Recently, some progress has been made in understanding pooling networks of *identical*, but independently noisy, threshold devices. One aspect concerns the behavior of information processing in the asymptotic limit of large networks, which is a limit of high relevance for neuroscience applications. The mutual information between the input and the output of the network has been evaluated, and its extremization has been performed. The aim of the present work is to extend these asymptotic results to study the more general case when the threshold values are no longer identical. In this situation, the values of thresholds can be described by a density, rather than by exact locations. We present a derivation of Shannon's mutual information between the input and output of these networks. The result is an approximation that relies a weak version of the law of large numbers, and a version of the central limit theorem. Optimization of the mutual information is then discussed.

Keywords: pooling networks, suprathreshold stochastic resonance, stochastic signal quantization, channel capacity, optimal quantization, high resolution quantization, lossy compression, neurons, noise

1. INTRODUCTION

Two challenges are ubiquitous for many forms of biological and artificial signal and information processing tasks. These are (i) robustness to the effects of random noise, and (ii) extraction of only the 'information' which is relevant for the task. The latter challenge can requires the use of lossy compression, by which we mean the situation where 'information' is intentionally discarded—because it is considered to be either redundant or irrelevant—and can never be recovered. Robustness to noise is often achieved using a network or array of sensors. It is less obvious that a network approach could lead to lossy compression. In this paper we consider a network model in which noise reduction and lossy compression can be simultaneously achieved. This model is a form of *stochastic pooling network*.

Fig. 1 shows a schematic diagram of our stochastic pooling network model. The network consists of N parallel 'sensors'—such as antennas, hydrophones or sensory 'hair' cells—each of which receive the same sample of a random input signal, x. This random signal is assumed to consist of a sequence of independent samples drawn

*mmcdonne@eleceng.adelaide.edu.au; phone +61 8 8303 6296; fax +61 8 8303 4360

Complexity and Nonlinear Dynamics, edited by Axel Bender, Proc. of SPIE
Vol. 6417, 641706, (2006) · 0277-786X/06/$15 · doi: 10.1117/12.695984

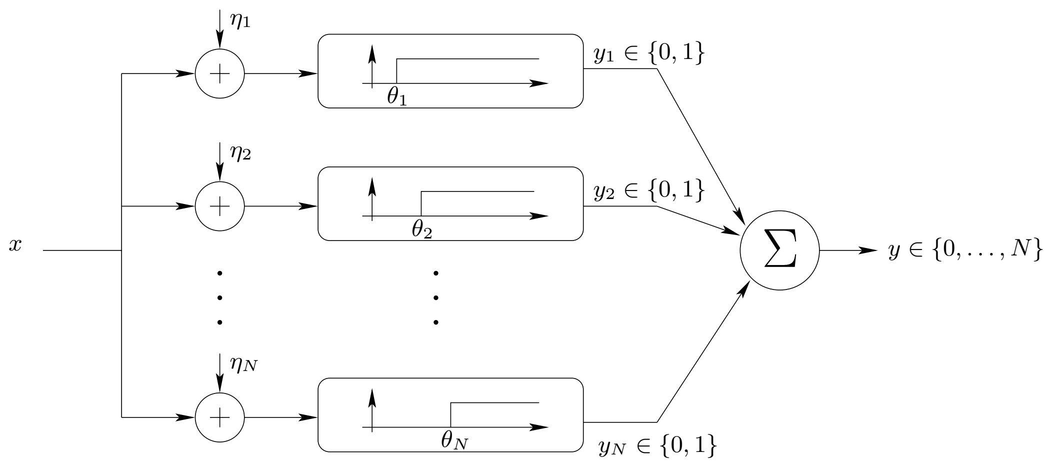

Figure 1. Abstract model of a stochastic pooling network. An analog sample, x, from the source distribution, $P(x)$, is received by N sensors, all of which are subject to *iid* additive input noise, η_i. The output from the i-th sensor, y_i, is unity if the sum of x and η_i is greater than the sensor's threshold value, θ_i, and zero otherwise. The overall network output, y, is the sum of the y_i's, and provides a stochastic quantization of x.

from a distribution with stationary probability density function (PDF), $P(x)$. The i–th device in the model is subject to continuously valued *iid*—independent and identically distributed—additive noise, η_i ($i = 1,..,N$), drawn from a probability distribution with PDF $R(\eta)$. Each noise signal is required to also be independent of the signal, x. Each sensor processes its input with a static threshold device to give an output signal, y_i, which is unity if the input signal, x, plus the noise on that device's threshold, η_i, is greater than the threshold value, θ_i. The output signal is zero otherwise. The outputs from each device, y_i, are summed to give an overall output signal, $y = \sum_{i=1}^{N} y_i$.

There are several possible ways of analyzing the performance of a pooling network. Information theory is used in this paper, since we consider several different applications for the model, and Shannon's mutual information provides a measure that is largely problem independent. The aim is to analyze the theoretical limits of performance of large stochastic pooling networks, i.e. we allow N to approach infinity. There are several free variables which can be modified in order to obtain optimal performance. These include (i) the signal distribution; (ii) the noise distribution; and (iii) the threshold values.

Before proceeding, the remainder of the introduction discusses a range of situations for which a stochastic pooling network provides a good model.

1.1. Various Forms of Pooling Networks

1.1.1. Parallel Sensory Cells in the Mammalian Auditory System

Our main motivation for studying stochastic pooling networks comes from our previous studies of the *suprathreshold stochastic resonance* (SSR) effect, which occurs in the network shown in Fig. 1 when all threshold values are identical.[1,2] This means that optimal performance—for transmission or estimation of a signal—is achieved for nonzero noise, rather than the absence of noise. Unlike most forms of stochastic resonance,[3–5] noise enhanced performance occurs for both sub and suprathreshold signals, and for very large input SNRs.

In the original SSR work[1] the parallel network of 'sensors' was intended to model parallel neurons, such as sensory neurons in the inner ear. Although most of the work in this area has studied neuron models of the form shown in Fig. 1, there have been several studies which replace the simple threshold nonlinearities with FitzHugh-Nagumo neuron models[6] and Hodgkin-Huxley neuron models,[7] and found the same qualitative results as for the simple threshold case. Further discussion on modeling sensory neural coding using Fig. 1 can be found elsewhere.[8–10]

1.1.2. Analog-to-Digital Converter Circuits

The input of our stochastic pooling network is a continuously valued random signal. The output is a discretely valued signal. Therefore, a stochastic pooling network is a quantizer.[11] The most familiar form of signal quantization occurs in analog-to-digital converter (ADC) circuits. In the absence of input noise, the stochastic pooling network shown in Fig. 1 is identical to a flash ADC,[12] where our 'sensors' are in fact simply comparator components, each with their own voltage threshold value. If an implementation of a flash ADC was such that independent additive noise existed at the input to each threshold device, possibly due to thermal noise, then a flash ADC could be considered to be a stochastic pooling network.[13] Note that this is different to dithering in an ADC, since we assume that each thresholding operation is *independently* noisy.

1.1.3. DIMUS Beamforming Sonar Arrays

The architecture of pooling networks has been used in signal detection applications, dating back to early digital beamforming sonar arrays in the 1960s. The technique used was known as Digital Multibeam Steering (DIMUS).[14, 15] Due to hardware constraints, the acoustic input to each of a number of spatially distributed hydrophones in an array were 'clipped' to a single bit, before being delayed and beamformed by summation. Although the DIMUS array is quite old, other more modern sonar systems, such as the Barra sonobuoy[16] have incorporated similar single-bit clipping of correlated signals. The use of quantization via pooling networks for detection applications has also received more recent attention.[17]

1.1.4. Bandwidth-Constrained Sensor Networks

Recent advances in wireless technology has led to a need for new distributed signal processing techniques that take into account severe rate or bandwidth constraints in distributed communication or sensor networks.[18–22] In particular, optimization of such networks is likely to depend on new forms of distributed compression or source coding.[23, 24] A model of this problem that has been studied in the information theory literature is known as the *CEO problem*.[25] In the CEO problem, each of N spatially separated sensors measure independently noisy observations of the same sequence of *iid* source samples. These observations are vector quantized at some arbitrary rate before transmission to a fusion center. Performance is usually analyzed via rate-distortion theory—that is, the problem is that of minimizing the sum-rate of communication to the central hub, subject to a specified mean square error distortion constraint. This is a form of stochastic pooling network; if the bandwidth constraint is such that each sensor can only single-bit scalar quantize its observations, then we have the model discussed in this paper—see Fig. 1.[26]

2. MATHEMATICAL DESCRIPTION

The output signal of the stochastic pooling network shown in Fig. 1 is an integer valued signal between 0 and N that can be expressed as a function of x in terms of the signum (sign) function as

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

We will describe the conditional distribution of the output given the input, denoted by $P(y = n|x)$, as the *transition probabilities*, since $P(y = n|x)$ gives the probability that input value, x, is encoded to output state n. From here on we will abbreviate the notation $P(y = n|x)$ to $P(n|x)$. The transition probabilities can be used to obtain $P_y(n)$ as

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

We always assume that the signal PDF, $P(x)$, is known. The mutual information between the input signal, x, and the output, y, of the network is that of a semi-continuous channel, and can be written as[1]

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

where we recall that $P(x)$ is the PDF of the input x. To progress further requires a method for calculating the transition probabilities.

2.1. All Threshold Identical

Suppose the threshold value of all devices in the network is identical and equal to θ. Let $P_{1|x}$ be the probability of any threshold device being 'on', *given* that the input signal value, x, is known. If the noise has PDF $R(\eta)$ and cumulative distribution function (CDF), $F_R(\cdot)$, then

$$P_{1|x} = 1 - F_R(\theta - x). \tag{4}$$

The transition probabilities as a function of x are given by the binomial distribution as[1]

$$P(n|x) = \binom{N}{n} P_{1|x}^n (1 - P_{1|x})^{N-n} \quad n \in 0, \ldots, N. \tag{5}$$

For this special case, Eq. (3) reduces to[1]

$$I(x,y) = -\sum_{n=0}^{N} P_y(n) \log_2 P^*(n) + N \int P(x) P_{1|x} \log_2 P_{1|x} dx + N \int P(x)(1 - P_{1|x}) \log_2 (1 - P_{1|x}) dx, \tag{6}$$

where $P^*(n) = P_y(n)/\binom{N}{n}$ so that $P^*(n) = \int P(x) P_{1|x}^n (1 - P_{1|x})^{N-n} dx$. Introducing a mathematically convenient change of probability measure via $\tau = P_{1|x}$ we obtain

$$I(x,y) = -\sum_{n=0}^{N} P_y(n) \log_2 P^*(n) + N \int_0^1 Q(\tau) \tau \log_2 \tau d\tau + N \int_0^1 Q(\tau)(1 - \tau) \log_2 (1 - \tau) d\tau, \tag{7}$$

where

$$Q(\tau) = \frac{P(x)}{R(\theta - x)}\Big|_{x = \theta - F_R^{-1}(1-\tau)}. \tag{8}$$

Note that throughout this paper, $Q(\cdot)$ is also a function of σ. The above result of course supposes that the support of $P(x)$ is contained in the support of $R(\theta - x)$, since otherwise division by zero occurs. It can be shown that $Q(\tau)$ is a PDF defined on support $[0, 1]$. Note that $P^*(n) = P_y(n)/\binom{N}{n}$ now reads

$$P^*(n) = \int_0^1 Q(\tau) \tau^n (1 - \tau)^{N-n} d\tau. \tag{9}$$

Provided both the signal and noise PDFs are even functions about their means, and θ is equal to the signal mean, then Eq. (7) can be further simplified to

$$I(x,y) = -\sum_{n=0}^{N} P_y(n) \log_2 P^*(n) + 2N \int_0^1 Q(\tau) \tau \log_2 \tau d\tau. \tag{10}$$

In the special case where $P(x) = R(\theta - x) \; \forall \; x$, it can be shown that[27]

$$I(x,y) = \log_2 (N + 1) - \frac{N}{2 \ln 2} - \frac{1}{N+1} \sum_{n=2}^{N} (N + 1 - 2n) \log_2 n, \tag{11}$$

regardless of the actual signal or noise distribution.

2.2. Noise Parameter, σ

We assume that the variance of the noise is always a function of the single parameter, σ_η, and the variance of the signal is a function of the single parameter, σ_x. In the case where the signal and noise both have the same distribution but different variances, it has previously proven convenient[1,28] to parameterize mutual information results in terms of the ratio of noise standard deviation to signal standard deviation, $\sigma = \sigma_\eta / \sigma_x$. The reason for this is that the mutual information is a function of the ratio, σ, and is therefore invariant to a change in σ_x provided σ_η changes by the same proportion. This result can be proved for certain families of distributions by writing the signal and noise PDFs in a normalized format. This is illustrated by deriving[29] the function $Q(\tau)$ for some specific cases, as shown in Table 1. For signal and noise distribution for which this property does not hold, one would instead select a fixed value of σ_x, and analyse the mutual information as a function of the noise standard deviation, or equivalently, the input signal-to-noise ratio.

Table 1. This table shows the PDF, $Q(\tau)$ for five different 'matched' signal an noise distributions, as well as $H(Q)$, the entropy of $Q(\tau)$. The label 'NAS' indicates that there is no analytical solution for the entropy.

Distribution	$Q(\tau)$	$H(Q)$
Gaussian	$\sigma \exp\left(\left(1-\sigma^2\right)\left(\text{erf}^{-1}(2\tau-1)\right)^2\right)$	$-\log_2(\sigma) - \frac{1}{2\ln 2}\left(\frac{1}{\sigma^2}-1\right)$
Uniform, $\sigma \geq 1$	$\begin{cases} \sigma, & -\frac{1}{2\sigma}+0.5 \leq \tau \leq \frac{1}{2\sigma}+0.5, \\ 0, & \text{otherwise.} \end{cases}$	$\log_2 \sigma$
Laplacian	$\begin{cases} \sigma(2\tau)^{(\sigma-1)} & \text{for } 0 \leq \tau \leq 0.5, \\ \sigma(2(1-\tau))^{(\sigma-1)} & \text{for } 0.5 \leq \tau \leq 1. \end{cases}$	NAS
Logistic	$\sigma\frac{(\tau(1-\tau))^{(\sigma-1)}}{(\tau^\sigma+(1-\tau)^\sigma)^2}$	NAS
Cauchy	$\sigma_\lambda \frac{1+\tan^2(\pi(\tau-0.5))}{(1+\sigma_\lambda^2\tan^2(\pi(\tau-0.5)))}$	NAS

3. MUTUAL INFORMATION IN A LARGE NETWORK FOR IDENTICAL THRESHOLD VALUES

This section reviews previous results[29–31] on the large N asymptotic performance of a stochastic pooling network for the case of all thresholds identical. We aim here to be slightly more rigorous in our proofs.

3.1. Output Distribution

It has been shown[29–31] that the output distribution for large N can be approximated by

$$P_y(n) = \frac{Q\left(\frac{n}{N}\right)}{N+1} \qquad n = 1, \ldots, N-1. \tag{12}$$

This result was derived by approximating the binomial distribution by a Gaussian. Note that we cannot always obtain the value of $P_y(n)$ for $n = 0$ or $n = N$, since for small values of σ, $Q(0)$ and $Q(1)$ may be infinite. We will come back to this problem in Section 3.1.2.

3.1.1. Beta Distribution Approach

A different approach follows from noting the output distribution of our model can be expressed in terms of the PDF of a Beta distribution,[32] with parameters $n+1$ and $N-n+1$, i.e. from Eq. (9),

$$\begin{aligned} P_y(n) &= \binom{N}{n}\int_0^1 \tau^n(1-\tau)^{(N-n)}Q(\tau)d\tau = \frac{1}{N+1}\int_0^1 \left(\frac{\tau^n(1-\tau)^{(N-n)}}{\int \phi^n(1-\phi)^{(N-n)}d\phi}\right)Q(\tau)d\tau \\ &= \frac{1}{N+1}\int_0^1 T(\tau)Q(\tau)d\tau, \end{aligned} \tag{13}$$

where $T(\tau)$ is the PDF of the Beta distribution with parameters $n+1$, $N-n+1$, and we have used the fact that the Beta function, $\beta(n+1, N-n+1) = \int \phi^n(1-\phi)^{(N-n)}d\phi = \frac{(N+1)!}{(n)!(N-n)!}$ cancels out most of the combinatorial term. For the case where $P(x) = R(\theta - x)$, $Q(\tau)$=1, and since $T(\tau)$ is a PDF, $P_y(n) = 1/(N+1)$, which is a result derived previously.[33] The PDF of a Beta distribution is monomodal, and can be shown to behave as a delta function located at $\tau = n/N$, when N is large, i.e. $T(\tau) \rightarrow \delta\left(\frac{n}{N}\right)$.[31] Substituting this into Eq. (13) gives the result of Eq. (12).

3.1.2. Laplace's Method

In order to derive this result more rigorously, we normalise the output of the network to the interval $[0, 1]$, by dividing y by N, so that the output probability mass function asymptotically approaches a probability density function on $[0, 1]$ as N approaches infinity. Let $\lambda \in [0, 1]$. Then, for a given N, there exists a unique k such that $\lambda \in [\frac{2k-1}{2N}, \frac{2k+1}{2N}]$. Hence, the density we search for is given by

$$P\Big(y \in (\lambda, \lambda + d\lambda)\Big) = f_y(\lambda)d\lambda \approx P(y = k/N)/N.$$

We have seen previously that this probability can be written as

$$P_y(\lambda) = \binom{N}{N\lambda} \int_0^1 Q(\tau)\tau^{N\lambda}(1-\tau)^{N(1-\lambda)} d\tau.$$

<u>**Case $\lambda \neq 0$ and $\lambda \neq 1$**</u>. We now make use of Laplace's method[34] (also known as the steepest descent method or saddle-point approximation method). This method allows approximation of the following form, which holds for large N,

$$\int Q(\tau) \exp\left(Nf(\tau)\right) d\tau \simeq Q(\tau_0) \sqrt{\frac{2\pi}{N|f''(\tau_0)|}} \exp\left(Nf(\tau_0)\right), \tag{14}$$

where τ_0 is the global maximum of $f(\tau)$, and is not an endpoint. This result can be derived by Taylor expanding $f(\tau)$ about τ_0, discarding all terms above τ^2, noting that $f'(\tau_0) = 0$, and that the resultant integral is a constant multiplied by a Gaussian PDF. For our purpose, we have to calculate

$$I_N(\lambda) \;=\; \int_0^1 Q(\tau)\tau^{N\lambda}(1-\tau)^{N(1-\lambda)} d\tau = \int_0^1 Q(\tau) e^{Nf(\tau)} d\tau, \text{ where } f(\tau) = \log\left(\tau^{\lambda}(1-\tau)^{1-\lambda}\right).$$

We have

$$f'(\tau) = \frac{\lambda - \tau}{\tau(1-\tau)} \text{ and } f''(\tau) = \frac{-\lambda + 2\lambda\tau - \tau^2}{\tau^2(1-\tau)^2}.$$

Therefore, f is maximum at $\tau_0 = \lambda$ where $f''(\lambda) = -1/(\lambda(1-\lambda))$ and we can apply Laplace's method. Before doing so, we might be aware that when λ is small, convergence may be very slow. Indeed, if $\sigma < 1$ and for monomodal densities, $Q(\tau)$ is diverging at 0 and 1. If $\sigma < 1$ and λ is close to 0 or 1, N has to be very large to ensure the validity of the approximation. This can be studied in the particular case when the noise and the input are Gaussian. For this case we have previously noted that

$$Q(\tau) = \sigma \exp\left((1-\sigma^2)\left(\operatorname{erf}^{-1}(2\tau - 1)\right)^2\right).$$

It can be shown[35] that $\left(\sqrt{2}\operatorname{erf}^{-1}(2\tau-1)\right)^2 \sim W\left(1/(2\pi\tau^2)\right)$ where $W(.)$ is the so-called Lambert function. Further, an asymptotic for the Lambert function is known, and we have $\lim_{x\to+\infty} W(x) = \log x - \log\log x$. Thus we obtain

$$Q(\tau)\tau^{N\lambda}(1-\tau)^{N(1-\lambda)} \;\sim\; \sigma(2\pi)^{\frac{\sigma^2-1}{2}} \tau^{N\lambda+\sigma^2-1}\left(-\log(2\pi\tau^2)\right)^{\frac{\sigma^2-1}{2}} |_{\tau\to 0}$$

which goes to zero as soon as $\lambda > (1-\sigma^2)/N$. Hence we see that N must at least greater than $1/\lambda$ in order to apply Laplace's method. In practice, N must be much greater.

Applying Laplace's method to I_N we obtain

$$I_N(\lambda) \sim Q(\lambda)\lambda^{N\lambda}(1-\lambda)^{N(1-\lambda)} \frac{\sqrt{2\pi\lambda(1-\lambda)}}{\sqrt{N}}.$$

Since $P_y(\lambda) = \binom{N}{N\lambda} I_N(\lambda)$, using Stirling's approximation $\Gamma(N+1) \sim \sqrt{2\pi} N^{N+1/2} e^{-N}$ leads to

$$P_y(\lambda) \;=\; \frac{\Gamma(N+1)}{\Gamma(N\lambda+1)\Gamma(N(1-\lambda)} Q(\lambda) \frac{(2\pi)^{1/2}\lambda^{N\lambda+1/2}(1-\lambda)^{N(1-\lambda)+1/2}}{N^{1/2}} \sim \frac{1}{N} Q(\lambda)$$

which of course is the desired result.

<u>**Cases $\lambda = 0$ and $\lambda = 1$**</u>. For $\lambda = 0$ and 1 there is a discontinuity in the analysis for $\sigma = 0$. In the absence of noise, the output of the network takes only two values and the density reads $P_y(\lambda) = p\delta(\lambda) + (1-p)\delta(\lambda - 1)$ for some p ($p = 1/2$ for symmetrical densities). Now, if $\sigma > 0$, we have $P_y(\lambda = 0) \to 0$ as $N \to +\infty$ since there

is a non zero probability for $x + \eta$ to be overthreshold, whatever x with unbounded support. The case $\lambda = 1$ is symmetrical and is not treated here.

Therefore, the case $\lambda = 0$ is tedious, since $Q(0) = +\infty$ if $\sigma < 1$ and if the densities are monomodal. In fact, we have

$$P_y(0) \quad = \quad \int_0^1 Q(\tau)(1-\tau)^N d\tau,$$

and the main contribution to the integral comes from what is happening at $\tau = 0$. We can obtain an asymptotic in the Gaussian case as follows. If we insert in this integral the approximation for Q at 0 we see that we have to evaluate

$$\sigma(2\pi)^{\frac{\sigma^2-1}{2}} \int_0 \tau^{\sigma^2-1} \left(-\log(2\pi\tau^2)\right)^{\frac{\sigma^2-1}{2}} (1-\tau)^N d\tau.$$

For small τ, we can further use the fact that $(1-\tau)^N \sim e^{-N\tau}$, then change $\sqrt{2\pi}\tau = u$ to obtain

$$\sigma(2\pi)^{\frac{\sigma^2-1}{2}} 2^{\frac{\sigma^2-1}{2}} \frac{1}{\sqrt{2\pi}^{\sigma^2-1}} \int_0 u^{\sigma^2-1} \left(-\log(u)\right)^{\frac{\sigma^2-1}{2}} e^{-Nu/\sqrt{2\pi}} d\tau.$$

This integral has a logarithmic singularity and can be approximated.[34] We finally obtain

$$P(0) \sim \sigma(2\pi)^{\frac{\sigma^2-1}{2}} \frac{1}{N^{\sigma^2}} \left(\log \frac{N^2}{2\pi}\right)^{\frac{\sigma^2-1}{2}},$$

showing the extremely slow convergence of $P(0)$ to zero as N grows when $\sigma < 1$.

3.2. Output Entropy

Consider the entropy of the discrete random variable y. Making use of Eq. (12), we have

$$H(y) = -\sum_{n=0}^{N} P_y(n) \log_2\left(P_y(n)\right) = -\frac{1}{N} \sum_{n=0}^{N} Q\left(\frac{n}{N}\right) \log_2\left(Q\left(\frac{n}{N}\right)\right) + \frac{\log_2(N)}{N} \sum_{n=0}^{N} Q\left(\frac{n}{N}\right). \tag{15}$$

The sums on the r.h.s are Riemann sums and converge to integrals as N goes to infinity. Therefore we obtain

$$\begin{aligned} H(y) &\simeq -\int_0^1 Q(\lambda)\log_2\left(Q(\lambda)\right)d\lambda + \log_2 N \int_0^1 Q(\lambda)d\lambda \\ &= \log_2(N) - \int_0^1 Q(\lambda)\log_2\left(Q(\lambda)\right)d\lambda = \log_2(N) + H_d(Q), \end{aligned} \tag{16}$$

where $H_d(Q)$ is the differential entropy a random variable which PDF is Q. This analysis agrees with a well known result,[36] which states that the entropy of an M bit quantization of a continuous random variable Z is approximately the sum of M and the entropy of Z. Here, we have λ as the continuous random variable that approximates the $N+1$ state discrete output distribution. The discrete output distribution is a $\log_2(N+1)$ bit quantization of λ, and has entropy approximately equal to the differential entropy of λ plus $\log_2(N+1)$, which for large N agrees with Eq. (16). Performing a change of variable in Eq. (16) of $\tau = F_R(x)$ gives

$$\begin{aligned} H(y) &\simeq \log_2 N - \int_{-\infty}^{\infty} P(x)\log_2\left(\frac{P(x)}{R(x)}\right)dx \\ &= \log_2(N) - D_{KL}(P||R), \end{aligned} \tag{17}$$

where $D_{KL}(P||R)$ is the relative entropy between the PDFs, $P(\cdot)$ and $R(\cdot)$, also known as Kullback-Leibler divergence.[36] This shows that $H(y)$ for large N is approximately the sum of the number of output bits and the negative of the relative entropy between $P(x)$ and $R(x)$. Since relative entropy is always non-negative, the approximation to $H(y)$ given by Eq. (17) is always less than or equal to $\log_2(N)$. This agrees with the known expression for $H(y)$ in the specific case of $\sigma = 1$ of $\log_2(N+1)$, which is the maximum output entropy, and holds for any N.

3.3. Conditional Output Entropy

An approximation to the conditional output entropy, $H(y|x)$ can be derived by noting that for large N, the binomial distribution can be approximated by a Gaussian with the same mean and variance that is, $NP_{1|x}$ and $NP_{1|x}(1-P_{1|x})$ respectively. Provided $0 \ll NP_{1|x} \ll N$ we have

$$P(n|x) \simeq \frac{1}{\sqrt{2\pi NP_{1|x}(1-P_{1|x})}} \exp\left(-\frac{(n-NP_{1|x})^2}{2NP_{1|x}(1-P_{1|x})}\right). \tag{18}$$

This approximation breaks down when $P_{1|x}$ is close to zero or unity, in which case, $P(n|x)$ can be approximated by the Poisson distribution, or the Edgeworth series approximation.[37] Note that more precise asymptotic expansions exist (*e.g.* the work of Knessl[38]) and that we could include higher-order terms. This study has not been done yet. Here we will use Eq. (18), since we have found that it is valid for our needs.[29] The *average* conditional output entropy is $H(y|x) = \int_x P(x)\hat{H}(y|x)dx$, where $\hat{H}(y|x) = -\sum_{n=0}^{N} P(n|x)\log_2(P(n|x))$. Using the well known result for the entropy of a Gaussian random variable, we can write

$$\hat{H}(y|x) \simeq 0.5\log_2(2\pi eNP_{1|x}(1-P_{1|x})). \tag{19}$$

Multiplying both sides of Eq. (19) by $P(x)$ and integrating over all x gives

$$\begin{aligned} H(y|x) &\simeq 0.5\log_2(2\pi eN) + 0.5\int_{-\infty}^{\infty} P(x)\log_2\left(P_{1|x}(1-P_{1|x})\right)dx \\ &= 0.5\log_2(2\pi eN) + \int_0^1 Q(\tau)\log_2\tau d\tau, \end{aligned} \tag{20}$$

since $P(x)$ and $R(\eta)$ are even functions about means of zero.

3.4. Mutual Information

The above equations for $H(y)$ and $H(y|x)$ can be combined to give a large N approximation to the mutual information as $I(x,y) = H(y) - H(y|x)$ and we obtain

$$I(x,y) \simeq 0.5\log_2\left(\frac{N}{2\pi e}\right) - \int_0^1 Q(\tau)\log_2(\tau Q(\tau))d\tau \tag{21}$$

$$= 0.5\log_2\left(\frac{N}{2\pi e}\right) - \int_0^1 Q(\tau)\log_2(\tau)d\tau + H(Q). \tag{22}$$

The integral on the RHS of Eq. (21) is independent of N and therefore for large N, the mutual information scales with $0.5\log_2(N)$. As noted previously,[27] this is half the maximum channel capacity for a network of size N. The integral on the RHS of Eq. (21) is insignificant when compared to $\log_2(N)$, but its importance is that it determines how the mutual information varies from $0.5\log_2\left(\frac{N}{2\pi e}\right)$ as σ varies.

For the specific case of $Q(\tau) = 1$, the mutual information reads $I(x,y) = 0.5\log_2\left(\frac{Ne}{2\pi}\right)$, which for large N agrees precisely with a known result, derived[29,30] for the case where $P(x) = R(\theta - x)$, validating the new formula in this specific case.

3.5. Maximising the Mutual Information

3.5.1. Matched Signal and Noise

Maximisation of the mutual information for 'matched' signal and noise distributions means finding the optimal value of the parameter, σ, which we label as σ_o. This can be expressed as[30]

$$\lim_{N\to\infty} \sigma_o = \min_\sigma \int_0^1 Q(\tau)\ln(\tau Q(\tau))d\tau, \tag{23}$$

or as

$$\lim_{N\to\infty} \sigma_o = \min_\sigma D_{KL}(P||R) + \int_{-\infty}^{+\infty} P(x)\log_2(P_{1|x})d\tau. \tag{24}$$

3.5.2. Arbitrary Signal and Noise

Maximising the mutual information means maximising the function

$$f(Q) = H_d(Q) - 0.5 \int_\tau Q(\tau) \log_2(\tau) d\tau - 0.5 \int_\tau Q(\tau) \log_2(1-\tau) d\tau. \qquad (25)$$

Maximising $f(Q)$ is equivalent to maximising the Lagrangian function that results from maximising $H_d(Q)$ subject to constraints on $\mathrm{E}[\log(\tau)]$ and $\mathrm{E}[\log(1-\tau)]$ and $\tau \in [0,1]$. This means we have a constrained maximum entropy problem.

The maximum entropy PDF for these constraints is given by the Beta(1/2,1/2) distribution,[39] and the maximum entropy PDF, Q, is given by

$$Q(\tau) = \frac{1}{\pi\sqrt{\tau(1-\tau)}}, \qquad \tau \in [0,1].$$

This can be verified by noting that the entropy of this $Q(\tau)$ is $H_d(\lambda) = \log_2(\pi) - 2$, and $\int_0^1 Q(\tau)\log_2(\tau)d\tau = -2$, so that the resultant mutual information is $I(x,y) = 0.5\log_2\left(\frac{N\pi}{2e}\right)$, which agrees with a result derived previously in a different way.[30]

Making a change of variable from τ to x via $\tau = F_R(x)$ leads to the optimality condition that

$$P(x) = \frac{R(x)}{\pi\sqrt{P_{1|x}(1-P_{1|x})}}. \qquad (26)$$

Note that it can be verified by direct integration that the RHS of Eq. (26) is a PDF, since the derivative of $P_{1|x}$ is $R(x)$.

An expression for the Fisher information, $J(x)$, in the network of threshold devices[40] is

$$J(x) = \frac{NR(x)^2}{P_{1|x}(1-P_{1|x})}, \qquad (27)$$

and as N grows, the mutual information can be approximated in terms of the Fisher information as[40]

$$I(x,y) = H(x) - 0.5\int_{-\infty}^{+\infty} P(x)\log_2\left(\frac{2\pi e}{J(x)}\right)dx. \qquad (28)$$

Equations of this form can be traced to older works.[41–43] Furthermore,[30] the integral of the square root of $J(x)$ is equal to $\pi\sqrt{N}$, regardless of the noise distribution. This means that the necessary condition for optimally can be written as

$$P(x) = \frac{\sqrt{J(x)}}{\pi\sqrt{N}} = \frac{\sqrt{J(x)}}{\int_x \sqrt{J(\phi)}d\phi}.$$

Any PDF which is a normalization of the square root of the Fisher information is known as Jeffrey's prior.[44, 45] Such equations have appeared previously in a neural coding context[42, 43] but were not referred to by this name. Furthermore, we see here that maximising with respect to P or R is equivalent.

4. MUTUAL INFORMATION IN A LARGE NETWORK FOR ARBITRARY THRESHOLD VALUES

We now turn to the presentation of our study of stochastic pooling networks with arbitrary threshold values. We adopt the asymptotic point of view, which is of high relevance in neuroscience applications. We begin by developing the asymptotic approximation of the mutual information and then turn to a discussion on the Fisher information. Note that such networks have been discussed for small N and the threshold values optimized as a function of σ.[26, 46, 47] This work showed that the SSR situation, where all thresholds are identical, is optimal for sufficiently large σ. For smaller σ, the optimal solution is for more than one threshold value to be identical, but for the total array to consist of 'clusters.' That is, each threshold device within a cluster has an identical threshold value, but different clusters have different threshold values.

To get the asymptotic situation we use the high-rate quantization[11] framework in which the thresholds are supposed to have a density that we denote as $p(\theta)$. In the limit as the number of levels, N, goes to infinity, p is the number of levels contained in an infinitesimal interval divided by N. In other words, around a level θ_0, the characteristic size of the cells in the asymptotic limit is given by $d\theta = 1/(Np(\theta_0))$. Note that a regular scalar quantizer can be seen as the cascade of "a nonlinearity (compressor)–uniform quantizer– the inverse nonlinearity (expander)" denoted as a compander. In the high-rate limit, the density of thresholds is the derivative of the compressor.

In the asymptotic regime, the general network can be seen as a pooling network of identical thresholds set to zero by combining the density of thresholds with the density of the noise. In this view however, the noises $\eta_i + \theta$ added on the input are no longer independent, and the theory recalled previously cannot be applied by replacing R by $R*p$, where $*$ denotes the convolution product. To be precise here, the result is identical if we only consider the law of the output (see the result below for the entropy), but once we study conditional quantity, the result differs.

4.1. Mutual Information in the asymptotic case

To get the entropy of the output of the network, we first show that, conditional on the input, the network is equivalent in law with a deterministic quantizer. In this section we use y to denote the output of the network divided by N. From Eq. (1), y can be written as a function of x as

$$y(x) = \frac{1}{N}\sum_{i=1}^{N} U(x+\eta_i-\theta_i),$$

where $U(x) = \frac{1}{2}\,\mathrm{sign}(x) + \frac{1}{2}$. We are interested in the large N behavior of y. We use the Lindeberg theorem,[48] which states that N independent random variables, X_i, with means m_i and variances σ_i^2 satisfies the law of large numbers and the central limit theorem whenever they are uniformly bounded. This is the case for us when we consider the variables $U(x+\eta_i-\theta_i)$ conditional on x (indeed U takes only two values).

First, we use the law of large numbers. Its conclusion is that for all $\varepsilon > 0$

$$P\left(\frac{1}{N}\Big|\sum_{i=1}^{N} U(x+\eta_i-\theta_i) - \sum_{i=1}^{N} E\left[U(x+\eta_i-\theta_i)\right]\Big| > \varepsilon \Big| x\right) \longrightarrow 0 \text{ as } N \to +\infty.$$

But we have

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N} E\left[U(x+\eta_i-\theta_i)\big|x\right] &= \frac{1}{N}\sum_{i=1}^{N}\int U(x+z)R(z+\theta_i)dz \\
&= \int U(x+z)\left(\frac{1}{N}\sum_{i=1}^{N} R(z+\theta_i)\right)dz \\
&\approx \int U(x+z)\left(\int p(\theta)R(z+\theta)d\theta\right)dz \\
&= \int U(x+z)f(z)dz = \int_{-x}^{+\infty} f(z)dz = 1 - F(-x),
\end{aligned}$$

where we have used $f(z) = \int p(\theta)R(z+\theta)d\theta$. Asymptotically, this means that the network is equivalent to the uniform quantization $1 - F(-x)$, that is the uniform quantization of the nonlinear transform $1 - F(-x)$, also called the compressor (F is the cumulative density function of f). Recall that in quantization theory, the threshold density is precisely the derivative of the compressor.[11] The conclusion of this is that asymptotically, the output of the network is equivalent in probability to the output of a deterministic quantizer with threshold density $g(x) = \frac{d(1-F(-x))}{dx} = f(-x) = \int p(\theta)R(\theta - x)d\theta$. Note that $g(x) = p(x) * R(-x)$ is the convolution between the threshold density and the noise PDF, $R(-x)$. To get the entropy of the output we use Eq. (16) to obtain

$$H(y) = \log_2 N - D_{KL}(P||g),$$

which generalizes the result given in Eq. (17).

We now evaluate the conditional output entropy $H(y|x)$. As before, y is the sum of conditionally independent random variables that are uniformly bounded. Therefore, Lindeberg's theorem[48] applies, and we conclude that as N goes to infinity, $y|x$ tends to be normally distributed with mean $\sum_i E[U(X + \eta_i - \theta_i)]$ and variance $\sum_i \mathrm{Var}[U(X+\eta_i-\theta_i)]$. Now, $U(X+\eta_i-\theta_i)$ is equal to 1 with probability $P_{1|x}(\theta_i) = 1-F_R(\theta_i-x)$ and to 0 with probability $1-P_{1|x}(\theta_i)$. Hence the mean of the random variable is $P_{1|x}(\theta_i)$ and its variance $P_{1|x}(\theta_i)\big(1-P_{1|x}(\theta_i)\big)$.

Since the thresholds have a density $p(\theta)$, we can approximate the discrete sums by integrals *via*

$$\begin{aligned} \sum_i P_{1|x}(\theta_i) &= N \int p(\theta) P_{1|x}(\theta) d\theta, \\ \sum_i P_{1|x}(\theta_i)\big(1 - P_{1|x}(\theta_i)\big) &= N \int p(\theta) P_{1|x}(\theta)\big(1 - P_{1|x}(\theta)\big) d\theta. \end{aligned}$$

Since given x the output of the network is asymptotically Gaussian, we obtain for the conditional entropy

$$H(y|x) = E_x\Big[\frac{1}{2}\log\Big(2\pi e N \int p(\theta)P_{1|x}(\theta)\big(1 - P_{1|x}(\theta)\big)d\theta\Big)\Big].$$

Finally, we have the following approximation for the mutual information, retaining $O(1)$ terms,

$$\begin{aligned} I(x, y) &= H(y) - H(y|x) \\ &= \frac{1}{2}\log\frac{N}{2\pi e} - D_{KL}(P(x)||p(x) * R(-x)) - \frac{1}{2}\int P(x)\log\Big[\int p(\theta)P_{1|x}(\theta)\big(1 - P_{1|x}(\theta)\big)d\theta\Big]dx \end{aligned}$$

which of course gives back the result of Eq. (22) for identical threshold values, in which case the density p is a δ function.

4.2. Maximizing the mutual Information for arbitrary thresholds

Optimization of the mutual information can be performed with respect to the input distribution $P(x)$, or with respect to the noise distribution $R(x)$, or with respect to the threshold density $p(\theta)$. Maximizing with respect to P or R is linked to the notion of capacity of a channel, and Shannon's theorem states that the capacity is the maximum amount of information that can be transmitted through the channel that can be recovered without error at the output. Maximizing with respect to $p(\theta)$ places the channel in its infomax regime. This last maximization is not developed here, since our results are too preliminary in this respect. To optimize with respect to P or R, we write the mutual information as

$$I(x, y) = \frac{1}{2}\log\frac{N}{2\pi e} + H(x) + \int P(x)\log\left[\frac{\int p(\theta)R(\theta - x)d\theta}{\sqrt{\int p(\theta)F_R(\theta - x)\big(1 - F_R(\theta - x)\big)d\theta}}\right]dx,$$

since $P_{1|x}(\theta) = 1 - F_R(\theta - x)$. We have now to solve the following problems

$$\begin{aligned} P_o(x) &= \operatorname{Arg}\max_P \left\{ H(x) + \int P(x)\log\left[\frac{\int p(\theta)R(\theta - x)d\theta}{\sqrt{\int p(\theta)F_R(\theta - x)\big(1 - F_R(\theta - x)\big)d\theta}}\right]dx - \gamma\int P(x)dx \right\}, \\ R_o(x) &= \operatorname{Arg}\max_R \left\{ \int P(x)\log\left[\frac{\int p(\theta)R(\theta - x)d\theta}{\sqrt{\int p(\theta)F_R(\theta - x)\big(1 - F_R(\theta - x)\big)d\theta}}\right]dx - \gamma\int R(x)dx \right\}, \end{aligned}$$

where the term $\gamma \int P(x)dx$ (or R) reflects the constraint that P or R must be PDFs. In the second problem, instead of looking for R, we look for its cumulative density F_R such that $R = F_R'$. In each problem, the maximum is attained if the Euler equation is fulfilled. For the first problem, it is immediate to obtain

$$P_o(x) \propto \frac{\int p(\theta)R(\theta - x)d\theta}{\sqrt{\int p(\theta)F_R(\theta - x)\big(1 - F_R(\theta - x)\big)d\theta}}, \tag{29}$$

which gives back the result for identical threshold by setting $p(\theta) = \delta(\theta)$.

For the second problem, the functional to maximize writes

$$\mathcal{L}(F_R(x)) = \int P(x) \log \left[\frac{\int p(\theta)F_R'(\theta - x)d\theta}{\sqrt{\int p(\theta)F_R(\theta - x)\big(1 - F_R(\theta - x)\big)d\theta}} \right] dx - \gamma \int F_R'(x)dx.$$

In this form, we cannot apply directly the Euler equation since $\mathcal{L}$ depends on F and F' through convolution. We therefore apply the functional derivative and demand that it is equal to zero

$$\frac{\delta \mathcal{L}(F_R)}{\delta F_R(y)} \quad = \quad \frac{d}{d\varepsilon}\mathcal{F}(F(x) + \varepsilon\delta(x - y))\Big|_{\varepsilon=0} = 0.$$

Some easy algebra leads to the following equivalent condition

$$\int \frac{P(x)p'(x + y)}{\int p(\theta)F_R'(\theta - x)d\theta}dx = -\frac{1}{2}\int \frac{P(x)p(x + y)(1 - 2F_R(y))}{\int p(\theta)F_R(\theta - x)\big(1 - F_R(\theta - x)\big)d\theta}dx. \tag{30}$$

In the case $p(\theta) = \delta(\theta)$, this equation reduces to

$$\frac{P'(-y)F_R'(y) + P(-y)F"_R(y)}{(F_R'(y))^2} = -\frac{(1 - 2F_R(y))P(-y)}{2F_R(y)(1 - F_R(y))},$$

which can be integrated to lead to $F_R'(y) \propto P(-y)\sqrt{F_R(y)(1 - F_R(y)}$ which is the result obtained for identical threshold values. In the general case however, we did not succeed in solving (30). Even allowing equality of the integrands (after an integration by parts in the LHS) in Eq. (30) does not lead to a solution.

4.3. Fisher Information Approximation for Arbitrary Thresholds

A simple formula like Eq. (27) for the Fisher information does not exist for arbitrary thresholds. This may be because y is no longer a sufficient statistic for x. Nevertheless, the Cramér-Rao information bound[45] says

$$\mathrm{var}[y|x] \geq \frac{\left(\frac{d}{dx}\mathrm{E}[y|x]\right)^2}{J(x)}, \tag{31}$$

where $J(x)$ is the Fisher information, which for arbitrary thresholds is given by

$$J(x) = \sum_{n=0}^{N} \frac{\left(\frac{dP(n|x)}{dx}\right)^2}{P(n|x)}. \tag{32}$$

Since $P_{1|x,i} = 1 - F_R(\theta_i - x)$, we have for the arbitrary threshold model that

$$\begin{aligned} \frac{d}{dx}\mathrm{E}[y|x] &= \sum_{i=1}^{N} \frac{dP_{1|x,i}}{dx} \\ &= \sum_{i=1}^{N} R(\theta_i - x) \\ &\approx N \int p(\theta)R(\theta - x)d\theta. \end{aligned} \tag{33}$$

In the same way we have $\mathrm{var}[y|x] \approx N \int p(\theta) F_R(\theta - x)\big(1 - F_R(\theta - x)\big) d\theta$ and therefore, an inequality for the Fisher information is

$$J(x) \geq \frac{N \left(\int p(\theta) R(\theta - x) d\theta\right)^2}{\int p(\theta) F_R(\theta - x)\big(1 - F_R(\theta - x)\big) d\theta}. \tag{34}$$

The Fisher information for arbitrary thresholds and signal and noise distributions can be calculated numerically from Eqn. (32), and compared with numerical calculations of the RHS of Inequality (34). Experiments with such calculations indicate that the bound does not hold exactly, apart from the SSR situation, but that, even for small N, the bound has a maximum error when compared to the exact Fisher information, in the order of one percent. It is possible this error is attributable to numerical errors, and that the bound does hold with equality. In any case, we are able to state

$$J(x) \simeq \frac{\left(\sum_{i=1}^{N} R(x - \theta_i)\right)^2}{\sum_{1=1}^{N} P_{1|x,i}(1 - P_{1|x,i})} \approx \frac{N \left(\int p(\theta) R(\theta - x) d\theta\right)^2}{\int p(\theta) F_R(\theta - x)\big(1 - F_R(\theta - x)\big) d\theta}. \tag{35}$$

Combining this last result with Eq. (29), we conclude that the capacity of the channel is achieved if $P(x)$ is chosen as Jeffrey's prior

$$P_o(x) = \frac{\sqrt{J(x)}}{\int_x \sqrt{J(\phi)} d\phi},$$

where $J(x)$ is given by Eq. (35).

5. CONCLUSIONS

In this paper we have discussed the limits of information transmission performance in a stochastic pooling network. This is potentially an important question for rate-constrained wireless networks that can be modeled in this way, as well as for understanding sensory neural coding.

For the situation where all nodes in the network are threshold devices with identical threshold values, the mutual information scales with $0.5 \log_2(N)$, and is maximized when the signal and noise distributions are such that Eq. (26) is satisfied. In the more general case of arbitrary thresholds, we have used the concept of a *point density function*, such as used in high resolution quantization theory, and found the more general result of Eq. (29).

Further work on this problem will address the problem of finding the optimal threshold point density function. If a solution can be found, features found—i.e. clustering and bifurcations—in the numerical optimization of the threshold values may be explained.[47]

ACKNOWLEDGMENTS

This work was funded by the Australian Research Council, The Australian Academy of Science and the Leverhulme Trust, grant number F/00 215/J and we thank them for their support.

REFERENCES

1. N. G. Stocks, "Suprathreshold stochastic resonance in multilevel threshold systems," *Physical Review Letters* **84**(11), pp. 2310–2313, 2000.
2. 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**(3), pp. L205–L220, 2002.
3. L. Gammaitoni, P. Hänggi, P. Jung, and F. Marchesoni, "Stochastic resonance," *Reviews of Modern Physics* **70**(1), pp. 223–287, 1998.

4. 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.
5. S. Zozor and P.-O. Amblard, "Noise-aided processing: revisiting dithering in a sigma-delta quantizer," *IEEE Transactions on Signal Processing* **53**, pp. 3202–3210, 2005.
6. N. G. Stocks and R. Mannella, "Generic noise enhanced coding in neuronal arrays," *Physical Review E* **64**, Art. No. 030902(R), 2001.
7. T. Hoch, G. Wenning, and K. Obermayer, "Optimal noise-aided signal transmission through populations of neurons," *Physical Review E* **68**(3), Art. No. 011911, 2003.
8. N. G. Stocks, D. Allingham, and R. P. Morse, "The application of suprathreshold stochastic resonance to cochlear implant coding," *Fluctuation and Noise Letters* **2**(3), pp. L169–L181, 2002.
9. D. Allingham, N. G. Stocks, R. P. Morse, and G. F. Meyer, "Noise enhanced information transmission in a model of multichannel cochlear implantation," in *Proc. SPIE Fluctuations and Noise in Biological, Biophysical, and Biomedical Systems II*, D. Abbott, S. M. Bezrukov, A. Der, and A. Sánchez, eds., **5467**, pp. 139–148, 2004.
10. R. P. Morse and N. G. Stocks, "Enhanced cochlear implant coding using multiplicative noise," in *Proc. SPIE Fluctuations and Noise in Biological, Biophysical and Biomedical Systems III*, N. G. Stocks, D. Abbott, and R. P. Morse, eds., **5841**, pp. 23–30, 2005.
11. A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, 1992.
12. V. K. Madisetti and D. B. Williams, *The Digital Signal Processing Handbook*, CRC Press, 1999.
13. M. D. McDonnell, D. Abbott, and C. Pearce, "Neural mechanisms for analog to digital conversion," in *Proc. SPIE BioMEMS and Nanoengineering*, D. V. Nicolau, ed., **5275**, pp. 278–286, 2003.
14. P. Rudnick, "Small signal detection in the DIMUS array," *The Journal of the Acoustical Society of America* **32**(7), pp. 871–877, 1960.
15. W. R. Remley, "Some effects of clipping in array processing," *The Journal of the Acoustical Society of America* **39**(4), pp. 702–707, 1966.
16. D. G. Gray, "Private communication." 2006.
17. S. Zozor, P. O. Amblard, and C. Duchêne, "Random quantizers as suboptimal detectors." Submitted to *Fluctuation and Noise Letters*, 2006.
18. M. Gastpar and M. Vetterli, "Power, spatio-temporal bandwidth, and distortion in large sensor networks," *IEEE Journal on Selected Areas in Communications* **23**(4), pp. 745–754, 2005.
19. A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks—Part I: Gaussian case," *IEEE Transactions on Signal Processing* **54**(3), pp. 1131–1143, 2006.
20. M. Gastpar, M. Vetterli, and P. L. Dragotti, "Sensing reality and communicating bits: a dangerous liaison," *IEEE Signal Processing Magazine* **14**, pp. 70–83, 2006.
21. J.-J. Xiao, A. Ribiero, Z.-Q. Luo, and G. B. Giannakis, "Distributed compression-estimation using wireless sensor networks," *IEEE Signal Processing Magazine* **14**, pp. 27–41, 2006.
22. Q. Zhao, A. Swami, and L. Tong, "The interplay between signal processing and networking in sensor networks," *IEEE Signal Processing Magazine* **14**, pp. 84–93, 2006.
23. Z. Xiong, A. D. Liveris, and S. Cheng, "Distributed source coding for sensor networks," *IEEE Signal Processing Magazine* **21**, pp. 80–94, 2004.
24. Z.-Q. Luo, M. Gastpar, J. Liu, and A. Swami, "Distributed signal processing in sensor networks," *IEEE Signal Processing Magazine* **14**, pp. 14–15, 2006.
25. T. Berger, Z. Zhang, and H. Viswanathan, "The CEO problem," *IEEE Transactions on Information Theory* **42**(3), pp. 887–902, 1996.
26. M. D. McDonnell, N. G. Stocks, and D. Abbott, "Complexity and distortion reduction in quantizers with independently random threshold levels." Submitted to *IEEE Transactions on Signal Processing*, 2006.
27. N. G. Stocks, "Information transmission in parallel threshold arrays: Suprathreshold stochastic resonance," *Physical Review E* **63**(4), Art. No. 041114, 2001.
28. 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**(12), pp. 1079–1089, 2002.

29. M. D. McDonnell, *Theoretical Aspects of Stochastic Signal Quantisation and Suprathreshold Stochastic Resonance*. PhD thesis, School of Electrical and Electronic Engineering, The University of Adelaide, Australia, 2006.
30. M. D. McDonnell, N. G. Stocks, and D. Abbott, "High resolution approximations for information transfer and channel capacity in the suprathreshold stochastic resonance model." Submitted to *Physical Review E*, 2006.
31. P.-O. Amblard, S. Zozor, and O. J. J. Michel, "Networks of the pooling type and optimal quantization," in *Proc. 2006 IEEE International Conference on Acoustics, Speech, and Signal Processing*, **3**, pp. 716–719, 2006.
32. R. D. Yates and D. J. Goodman, *Probability and Stochastic Processes: A Friendly Introduction for Electrical and Computer Engineers*, John Wiley and Sons, Inc., second ed., 2005.
33. N. G. Stocks, "Suprathreshold stochastic resonance: An exact result for uniformly distributed signal and noise," *Physics Letters A* **279**(5-6), pp. 308–312, 2001.
34. R. Wong, *Asymptotic Approximation of Integrals*, Academic Press, 1989.
35. D. E. Dominici, "The inverse of the cumulative standard normal probability function," *Integrals, Transforms and Special Functions* **14**(4), pp. 281–292, 2003.
36. T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley and Sons, New York, 1991.
37. J. V. Harrington, "An analysis of the detection of repeated signals in noise by binary integration," *IRE Transactions on Information Theory* **1**(1), pp. 1–9, 1955.
38. C. Knessl, "Integral representations and asymptotic expansions for Shannon and Rényi entropies," *Applied Mathematics Letters* **11**(2), pp. 69–74, 1998.
39. J. N. Kapur and H. K. Kesavan, *Entropy Optimization Principles with Applications*, Academic Press, 1992.
40. T. Hoch, G. Wenning, and K. Obermayer, "Adaptation using local information for maximizing the global cost," *Neurocomputing* **52-54**, pp. 541–546, 2003.
41. R. Stein, "The information capacity of nerve cells using a frequency code," *Biophysics Journal* **7**, pp. 70–86, 1967.
42. 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.
43. N. Brunel and J. Nadal, "Mutual information, Fisher information and population coding," *Neural Computation* **10**, pp. 1731–1757, 1998.
44. J. J. Rissanen, "Fisher information and stochastic complexity," *IEEE Transactions on Information Theory* **42**(1), pp. 40–47, 1996.
45. E. L. Lehmann and G. Casella, *Theory of Point Estimation*, Springer, New York, 1998.
46. 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**(3), pp. L457–L468, 2005.
47. 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.
48. W. Feller, *An Introduction to Probability and its Application*, Wiley, 1968.