Optimal stimulus and noise distributions for information transmission via suprathreshold stochastic resonance

Mark D. McDonnell,[1,*] Nigel G. Stocks,[2,†] and Derek Abbott[1,‡]
[1]*School of Electrical and Electronic Engineering & Centre for Biomedical Engineering, The University of Adelaide, South Australia 5005, Australia*
[2]*School of Engineering, The University of Warwick, Coventry CV4 7AL, United Kingdom*
(Received 29 March 2006; revised manuscript received 24 January 2007; published 6 June 2007)

Suprathreshold stochastic resonance (SSR) is a form of noise-enhanced signal transmission that occurs in a parallel array of independently noisy identical threshold nonlinearities, including model neurons. Unlike most forms of stochastic resonance, the output response to suprathreshold random input signals of arbitrary magnitude is improved by the presence of even small amounts of noise. In this paper, the information transmission performance of SSR in the limit of a large array size is considered. Using a relationship between Shannon's mutual information and Fisher information, a sufficient condition for optimality, i.e., channel capacity, is derived. It is shown that capacity is achieved when the signal distribution is Jeffrey's prior, as formed from the noise distribution, or when the noise distribution depends on the signal distribution via a cosine relationship. These results provide theoretical verification and justification for previous work in both computational neuroscience and electronics.

DOI: 10.1103/PhysRevE.75.061105 PACS number(s): 05.40.Ca, 02.50.−r, 89.70.+c

I. INTRODUCTION

The term "stochastic resonance" describes the situation where a system's response to some signal is optimized by the presence of random noise, rather than its absence. It occurs in a wide variety of nonlinear physical [1] and biological [2] systems.

In many of the systems and models in which stochastic resonance (SR) has been observed, the essential nonlinearity is a single static threshold, e.g., [3–6]. It is generally thought that SR cannot occur in such systems for suprathreshold signals, meaning that the amplitude of the input signal needs to be restricted to values smaller than the amplitude of the threshold for SR to occur [7].

However, the 2000 discovery of a novel form of SR—known as *suprathreshold stochastic resonance* (SSR)—showed that this is not always true [8]. SSR occurs in an array of identical threshold nonlinearities, each of which is subject to independently random additive noise. We refer to this array as the *SSR model*—see Fig. 1. In this model SR occurs regardless of whether the input signal is entirely subthreshold or not. Furthermore, SSR occurs even for very large input signal-to-noise ratios (SNRs). This is a further difference from conventional SR, for which the signal is required to be weak compared to the noise.

SSR is a form of aperiodic stochastic resonance [4,9,10] that was first shown to occur by calculating Shannon's average mutual information for the SSR model [8]. It was subsequently found that the performance achievable via SSR is maximized when all threshold values are set to the signal mean [11], and that, for sufficiently small input SNRs, modifying the thresholds in the model cannot improve information transfer [12].

The SSR model was originally motivated as a model for parallel sensory neurons, such as those synapsing with hair cells in the inner ear [13]. Although the basic SSR model is nondynamical, and does not model the many complexities of real neurons, each threshold nonlinearity is equivalent to a Pitts-McCulloch neuron model, and encapsulates the neural coding properties we are interested in—i.e., the generation of action potentials in response to a noisy aperiodic random stimulus. The small input SNRs we focus on are biologically relevant [14], particularly so for hair cells, which are subject to substantial Brownian motion [15]. This leads to much randomness in the release of neurotransmitters at synapses with afferent neurons leading to the cochlear nucleus.

Further justification of the SSR model's relevance to neural coding is discussed in [16–18], and by extensions of the model to include more biologically realistic neural features. For example, the parallel array has been modified to consist of parallel FitzHugh-Nagumo neuron models [19], leaky integrate-and-fire neuron models [16,17], and Hodgkin-Huxley models [16], and for the case of signal-dependent

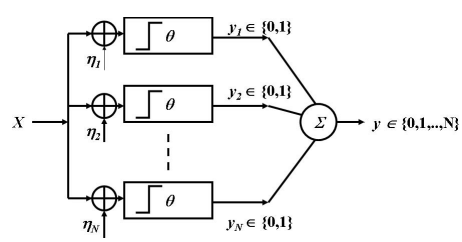

FIG. 1. SSR model: There are N parallel threshold devices, each with the same threshold value θ. The common input signal is a continuously valued random signal X, consisting of a sequence of discrete time-uncorrelated samples. Each device receives independently noisy versions of X. The noise signals η_i are i.i.d. additive random signals that are independent of X. The output from the ith device, y_i, is unity if $X+\eta_i > \theta$ and zero otherwise. The overall output y is the sum of the individual outputs y_i.

*Electronic address: mmcdonne@eleceng.adelaide.edu.au
†Electronic address: ngs@eng.warwick.ac.uk
‡Electronic address: dabbott@eleceng.adelaide.edu.au

 ©2007 The American Physical Society

(multiplicative) noise [18]. In all cases the same qualitative results as for the simple threshold model were obtained. The SSR effect has also led to a proposal for improving the performance of cochlear implants for suprathreshold stimuli [13], based on the idea that the natural randomness present in functioning cochlear hair cells is missing in patients requiring implants [20].

The purpose of this paper is to analyze, in a *general* manner, the information theoretic upper limits of performance of the SSR model. This requires allowing the array size N to approach infinity. Previous work has discussed the scaling of the mutual information through the SSR model with N for specific cases, and found conditions for which the maximum mutual information—i.e., channel capacity—occurs [11,16,21]. In a neural coding context, the question "What is the optimal stimulus distribution?" for a given noise distribution is discussed numerically for the SSR model in [16].

In Sec. II, we significantly extend the results in [11,16,21], by showing that the mutual information and output entropy can both be written in terms of simple relative entropy expressions—see Eqs. (21) and (22). This leads to a very general sufficient condition, Eq. (25), for achieving capacity in the large-N regime that can be achieved by either optimizing the signal distribution for a given noise distribution, or optimizing the noise for a given signal. Given the neuroscience motivation for studying the SSR model, this result is potentially highly significant in computational neuroscience, where both optimal stimulus distributions and optimal tuning curves are often considered [16,22].

Furthermore, the optimal signal for the special case of uniform noise is shown to be the arcsine distribution (a special case of the Beta distribution), which has a relatively large variance and is bimodal. This result provides theoretical justification for a proposed heuristic method for analog-to-digital conversion based on the SSR model [23]. In this method, the input signal is transformed so that it has a large variance and is bimodal.

As a means of verification of our theory, in Sec. III our general results are compared to the specific capacity results contained in [11,16,21]. This leads us to find and justify improvements to these previous results.

Before we proceed, however, the remainder of this section outlines our notation, describes the SSR model, and derives some important results that we utilize.

A. Information theoretic definitions

Recent work using the SSR model has described performance using measures other than mutual information [24–29]. However, in line with much theoretical neuroscience research [14], here we use the information theoretic viewpoint where the SSR model can be considered to be a communications channel [8].

Throughout, we denote the probability mass function (PMF) of a discrete random variable α as $P_\alpha(\cdot)$, the probability density function (PDF) of a continuous random variable β as $f_\beta(\cdot)$, and the cumulative distribution function (CDF) of β as $F_\beta(\cdot)$.

All signals are discrete-time memoryless sequences of samples drawn from the same stationary probability distribution. This differs from the detection scenario often considered in SR research, in which the input signal is periodic. Such a signal does not convey new information with an increasing number of samples, and cannot be considered from an information theoretic viewpoint [7].

Consider two continuous random variables X and Y, with PDFs $f_X(x)$ and $f_Y(x)$, with the same support S. The *relative entropy*—or Kullback-Liebler divergence—between the two distributions is defined as [30]

$$D(f_X \| f_Y) = \int_{\eta \in S} f_X(\eta) \log_2\left(\frac{f_X(\eta)}{f_Y(\eta)}\right) d\eta. \tag{1}$$

Suppose X and Y have joint PDF $f_{XY}(x,y)$. Shannon's mutual information between X and Y is defined as the relative entropy between the joint PDF and the product of the marginal PDFs [30],

$$\begin{aligned} I(X,Y) &= \int_x \int_y f_{XY}(x,y) \log_2\left(\frac{f_{XY}(x,y)}{f_X(x) f_Y(y)}\right) dx\, dy \\ &= H(Y) - H(Y|X) \quad \text{bits per sample}, \end{aligned} \tag{2}$$

where $H(Y)$ is the entropy of Y and $H(Y|X)$ is the average conditional entropy of Y given X.

The definition of mutual information also holds for discrete random variables, and for one variable discrete and one continuous. The entropy of a discrete random variable Y is given by

$$H(Y) = -\sum_{n=0}^{N} P_Y(n) \log_2 P_Y(n), \tag{3}$$

while a continuous random variable X has *differential* entropy

$$H(X) = -\int_{\eta \in S} f_X(\eta) \log_2[f_X(\eta)] d\eta. \tag{4}$$

In this paper we are interested in the case of X continuous with support S and Y discrete, with N states, in which case the average conditional entropy of Y given X is

$$H(Y|X) = -\int_{x \in S} f_X(x) \sum_{n=0}^{N} P_{Y|X}(n|x) \log_2[P_{Y|X}(n|x)] dx. \tag{5}$$

In information theory, the term *channel capacity* is defined as being the maximum achievable mutual information of a given channel [30]. Suppose X is the source random variable, and Y is the random variable at the output of the channel. Usually, the channel is assumed to be fixed and the maximization performed over all possible source PDFs $f_X(x)$. The channel capacity C can be expressed as the optimization problem

$$\text{Find} \quad C = \max_{\{f_X(x)\}} I(X,Y). \tag{6}$$

Usually there are prescribed constraints on the source distribution such as a fixed average power, or a finite alphabet [30]. In Sec. III we will also consider the more stringent

constraint that the PDF of the source is known other than its variance. In this situation, channel capacity is determined by finding the optimal source variance, or, as is often carried out in SR research, the optimal noise variance.

B. SSR model

Figure 1 shows a schematic diagram of the SSR model. The array consists of N parallel threshold nonlinearities—or "devices"—each of which receive the same random input signal X with PDF $f_X(\cdot)$. The ith device in the model is subject to continuously valued independent and identically distributed (i.i.d.) additive random noise η_i $(i=1,\ldots,N)$, with PDF $f_\eta(\cdot)$. Each noise signal is required to also be independent of the signal X. The output of each device, y_i, is unity if the input signal X plus the noise on that device's threshold, η_i, is greater than the threshold value θ. The output signal is zero otherwise. The outputs from each device, y_i, are summed to give the overall output signal, $y=\Sigma_{i=1}^N y_i$. This output is integer valued, $y\in[0,\ldots,N]$, and is therefore a quantization (digitization) of X [28].

The conditional PMF of the output given the input is $P_{y|X}(y=n|X=x)$, $n\in[0,\ldots,N]$. We abbreviate this to $P_{y|X}(n|x)$. The output distribution is

$$P_y(n)=\int_x P_{y|X}(n|x)f_X(x)dx, \quad n\in 0,\ldots,N. \tag{7}$$

The mutual information between X and y is that of a semi-continuous channel [8], and can be written as

$$\begin{aligned} I(X,y) &= H(y)-H(y|X) \\ &= -\sum_{n=0}^{N} P_y(n)\log_2 P_y(n) \\ &\quad -\left(-\int_{-\infty}^{\infty} f_X(x)\sum_{n=0}^{N} P_{y|X}(n|x)\log_2 P_{y|X}(n|x)dx\right). \end{aligned} \tag{8}$$

To progress further we use the notation introduced in [8]. Let $P_{1|x}$ be the probability of the ith threshold device giving output $y_i=1$ in response to input signal value $X=x$. If the noise CDF is $F_\eta(\cdot)$, then

$$P_{1|x}=1-F_\eta(\theta-x). \tag{9}$$

As noted in [8], $P_{y|X}(n|x)$ is given by the binomial distribution as

$$P_{y|X}(n|x)=\binom{N}{n}P_{1|x}^n(1-P_{1|x})^{N-n}, \quad n\in 0,\ldots,N, \tag{10}$$

and Eq. (8) reduces to

$$\begin{aligned} I(X,y) &= -\sum_{n=0}^{N} P_y(n)\log_2\left(\frac{P_y(n)}{\binom{N}{n}}\right)+N\int_x f_X(x)P_{1|x}\log_2 P_{1|x}dx \\ &\quad +N\int_x f_X(x)(1-P_{1|x})\log_2(1-P_{1|x})dx. \end{aligned} \tag{11}$$

Numerically evaluating Eq. (11) as a function of input SNR for given signal and noise distributions shows that the mutual information has a unimodal stochastic resonance curve for $N>1$, even when the signal and noise are both suprathreshold—i.e., the threshold value θ is set to the signal mean [11,24].

Further analytical simplification of Eq. (8) is possible in the case where the signal and noise PDFs are identical, with the same variance, i.e., $f_X(x)=f_\eta(\theta-x)\ \forall\ x$ [11]. The result is

$$I(X,y)=\log_2(N+1)-\frac{N}{2\ln 2}-\frac{1}{N+1}\sum_{n=2}^{N}(N+1-2n)\log_2 n. \tag{12}$$

What is quite remarkable about this result is that the mutual information is independent of the shape of the PDFs of the signal and noise, other than that $f_X(x)=f_\eta(\theta-x)\ \forall\ x$. This means that both PDFs have the same shape, but may possibly have different means, and be mutually reversed along the x axis about their means. In Sec. II D we compare the mutual information of Eq. (12) with our calculations of the general channel capacity.

C. Describing SSR using a single PDF $f_Q(\tau)$

We now show that the mutual information in the SSR model depends solely on N and an auxiliary PDF $f_Q(\cdot)$. This PDF is shown to be that of the random variable describing the conditional *average* output of the SSR model, given that the input signal is $X=x$.

1. $f_Q(\tau)$ as the PDF of the average transfer function

Although the output of the SSR model, y, is a discrete random variable, the conditional *expected* value of y, given the input is $X=x$, is a continuous random variable, since X is. We label this random variable as $\bar{Y}$. Since the PMF of y given $X=x$ is the binomial PMF as in Eq. (10), we know that $\bar{Y}$ is the random variable that results from $\bar{y}=\mathrm{E}[y|X=x]=NP_{1|x}$. Inverting this gives $x=\theta-F_\eta^{-1}(1-\bar{y}/N)$.

The PDF of $\bar{Y}$ can be derived from $f_X(\cdot)$, since $\bar{y}=NP_{1|x}$ provides an invertible transformation of X, with PDF $f_X(x)$, to $\bar{Y}$, with PDF $f_{\bar{Y}}(\bar{y})$. Using the well-known expression for the resultant PDF, and provided the support of $f_X(x)$ is contained in the support of $f_\eta(\theta-x)$—since otherwise $dx/d\bar{y}$ does not necessarily exist—we have

TABLE I. The auxiliary PDF $f_Q(\tau)$ for five different matched signal and noise distributions (i.e., the same distribution but with different variances), as well as $H(Q)$, the entropy of $f_Q(\tau)$. The threshold value θ and the signal and noise means are assumed to be zero, so that these results are independent of θ. The noise intensity $\sigma=\sigma_\eta/\sigma_x$ is the ratio of the noise standard deviation to the signal standard deviation. For the Cauchy case, σ_λ is the ratio of the full-width-at-half-maximum parameters. The label NAS indicates that there is no analytical solution for the entropy.

Distribution	$f_Q(\tau)$	$H(Q)$
Gaussian	$\sigma \exp\{(1-\sigma^2)[\mathrm{erf}^{-1}(2\tau-1)]^2\}$	$-\log_2(\sigma)-\frac{1}{2\ln 2}\left(\frac{1}{\sigma^2}-1\right)$
Uniform, $\sigma \geqslant 1$	$\begin{cases}\sigma, & -\frac{1}{2\sigma}+0.5 \leqslant \tau \leqslant \frac{1}{2\sigma}+0.5,\\ 0 & \text{otherwise}\end{cases}$	$\log_2 \sigma$
Laplacian	$\begin{cases}\sigma(2\tau)^{(\sigma-1)} & \text{for } 0 \leqslant \tau \leqslant 0.5\\ \sigma[2(1-\tau)]^{(\sigma-1)} & \text{for } 0.5 \leqslant \tau \leqslant 1\end{cases}$	$-\log_2(\sigma)-\frac{1}{2\ln 2}\left(\frac{1}{\sigma}-1\right)$
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

$$f_{\bar{Y}}(\bar{y}) = f_X(x)\left|\frac{dx}{d\bar{y}}\right|\Bigg|_{x=\theta-F_\eta^{-1}(1-\bar{y}/N)} = \frac{f_X(x)}{Nf_\eta(\theta-x)}\Bigg|_{x=\theta-F_\eta^{-1}(1-\bar{y}/N)}, \quad \bar{y}\in[0,N]. \tag{13}$$

Our condition regarding the supports of the signal and noise ensures that $f_\eta(\cdot)\neq 0$. If we make a further change to a new random variable, Q, via $\tau=\bar{y}/N$, the PDF of Q is

$$f_Q(\tau) = \frac{f_X(x)}{f_\eta(\theta-x)}\Bigg|_{x=\theta-F_\eta^{-1}(1-\tau)}, \quad \tau\in[0,1], \tag{14}$$

and the PDF of $\bar{Y}$ can be written as

$$f_{\bar{Y}}(\bar{y}) = \frac{f_Q(\bar{y}/N)}{N}, \tag{15}$$

which illustrates the physical significance of the auxiliary PDF $f_Q(\cdot)$ as the PDF of $\bar{y}/N$.

2. Mutual information in terms of $f_Q(\tau)$

Making a change of variable in Eq. (11) from x to τ via $\tau=P_{1|x}=1-F_\eta(\theta-x)$ gives

$$I(X,y) = -\sum_{n=0}^{N} P_y(n)\log_2\left(\frac{P_y(n)}{\binom{N}{n}}\right) + N\int_{\tau=0}^{\tau=1} f_Q(\tau)\tau\log_2\tau\, d\tau + N\int_{\tau=0}^{\tau=1} f_Q(\tau)(1-\tau)\log_2(1-\tau)d\tau, \tag{16}$$

where

$$P_y(n) = \binom{N}{n}\int_{\tau=0}^{\tau=1} f_Q(\tau)\tau^n(1-\tau)^{N-n}d\tau. \tag{17}$$

Equations (16) and (17) show that the PDF $f_Q(\tau)$ encapsulates the behavior of the mutual information in the SSR model.

3. Entropy of the random variable Q

If we make a change of variable from τ to x, and note that $f_X(x)dx=f_Q(\tau)d\tau$, the entropy of Q can be written as

$$H(Q) = -\int_0^1 f_Q(\tau)\log_2[f_Q(\tau)]d\tau = -\int_x f_X(x)\log_2\left(\frac{f_X(x)}{f_\eta(\theta-x)}\right)dx = -D(f_X(x)\|f_\eta(\theta-x)), \tag{18}$$

which is the negative of the relative entropy between the signal PDF and the noise PDF reversed about $x=0$ and shifted by θ. In the event that the noise PDF is an even function about its mean, and θ is equal to the signal mean, then the entropy of Q is simply the negative of the relative entropy between the signal and noise PDFs.

4. Examples of the PDF $f_Q(\tau)$

The PDF $f_Q(\tau)$ can be derived for specific signal and noise distributions. Table I lists $f_Q(\tau)$ for several cases where the signal and noise share the same distribution and a mean of zero, but with not necessarily equal variances. The threshold value θ is also set to zero.

For each case considered, the standard deviation of the noise can be written as $a\sigma_\eta$, where a is a positive constant,

and the standard deviation of the signal can be written $a\sigma_x$. We find that $f_Q(\tau)$ in each case is a function of a single parameter that we call the *noise intensity*, $\sigma=\sigma_\eta/\sigma_x$. Given this, from Eq. (16), it is clear that the mutual information must be a function only of the ratio σ, so that it is invariant to a change in σ_x provided σ_η changes by the same proportion. This fact is noted to be true for the Gaussian case in [8], and the uniform case in [11], but here we have illustrated why.

We note however, that if θ is not equal to the signal mean, then $f_Q(\tau)$ will depend on the ratio θ/σ_x, as well as θ and σ, and therefore so will the mutual information.

Table I also lists the entropy of Q for three cases where an analytical expression could be found.

D. Large-*N* SSR: Literature review and outline of this paper

In the absence of noise, the maximum mutual information is the maximum entropy of the output signal $\log_2(N+1)$. It has been shown for very specific signal and noise distributions that the mutual information in the SSR model scales as $0.5\log_2(N)$ for large N [11,21]. This means that the channel capacity for large N under the specified conditions is about half the maximum noiseless channel capacity. This situation is discussed in Sec. III.

The only other work to consider SSR in the large-N regime finds that the optimal noise intensity for Gaussian signal and noise occurs for $\sigma\simeq 0.6$ [16]. Unlike [21]—which uses the exact expression of Eq. (12), and derives a large-N expression by approximating the summation with an integral—[16] begins by using a *Fisher-information*-based approximation to the mutual information.

In Appendix A 1 we rederive the formula of [16] in a different manner, which results in new large-N approximations for the *output entropy*, as well as the mutual information. These approximations provide the basis for the central result of this paper, which is a general sufficient condition for achieving channel capacity in the SSR model, for any arbitrary specified signal or noise distribution. This is discussed in Sec. II. These new general results are compared with the specific results of [11,16,21] in Sec. III.

II. A GENERAL EXPRESSION FOR THE SSR CHANNEL CAPACITY FOR LARGE *N*

Fisher information [30,31] has previously been discussed in numerous papers on both neural coding [32] and stochastic resonance [33], and both [34,35]. However, most SR studies using Fisher information consider only the case where the signal itself is not a random variable. When it is a random variable, it is possible to connect Fisher information and Shannon mutual information under special conditions, as discussed in [16,22,34,36].

It is demonstrated in [16] that the Fisher information at the output of the SSR model as a function of input signal value $X=x$, is given by

$$J(x)=\left(\frac{dP_{1|x}}{dx}\right)^2\frac{N}{P_{1|x}(1-P_{1|x})}. \tag{19}$$

In [16], Eq. (19) is used to approximate the large-N mutual information in the SSR model via the formula

$$I(X,y)=H(X)-0.5\int_{x=-\infty}^{x=\infty}f_X(x)\log_2\left(\frac{2\pi e}{J(x)}\right)dx. \tag{20}$$

This expression—which is derived under much more general circumstances in [22,37]—relies on an assumption that an efficient Gaussian estimator for x can be found from the output of the channel, in the limit of large N.

In Appendix A 1 we outline an alternative derivation to Eq. (20)—from which Eq. (19) can be inferred—that is specific to the SSR model, and provides additional justification for its large-N asymptotic validity. This alternative derivation allows us to find individual expressions for both the output entropy and conditional output entropy. This derivation makes use of the auxiliary PDF $f_Q(\tau)$ derived in Sec. I C. The significance of this approach is that it leads to our demonstration of the new results that the output entropy can be written for large N as

$$H(y)\simeq\log_2(N)-D(f_X(x)\|f_\eta(\theta-x)), \tag{21}$$

while the mutual information can be written as

$$I(X,y)\simeq 0.5\log_2\left(\frac{N\pi}{2e}\right)-D(f_X\|f_S), \tag{22}$$

where $f_S(\cdot)$ is a PDF known as Jeffrey's prior,

$$f_S(x)=\frac{\sqrt{J(x)}}{\pi\sqrt{N}}. \tag{23}$$

It is proven in Appendix A 2 that for the SSR model Eq. (23) is indeed a PDF. This is a remarkable result, as in general Jeffrey's prior has no such simple form. Substitution of Eq. (23) into Eq. (22) and simplifying leads to Eq. (20), which verifies this result.

By inspection of Eq. (19), $f_S(x)$ can be derived from knowledge of the noise PDF $f_\eta(\eta)$, since

$$f_S(x)=\frac{f_\eta(\theta-x)}{\pi\sqrt{F_\eta(\theta-x)[1-F_\eta(\theta-x)]}}. \tag{24}$$

A. A sufficient condition for optimality

Since relative entropy is always non-negative, from Eq. (22) a sufficient condition for achieving the large-N channel capacity is that

$$f_X(x)=f_S(x)\quad\forall\ x, \tag{25}$$

with the resultant capacity as

$$C(X,y)=0.5\log_2\left(\frac{N\pi}{2e}\right)\simeq 0.5\log_2 N-0.3956. \tag{26}$$

Equation (26) holds, provided the conditions for the approximation given by Eq. (20) hold. Otherwise, the right-hand sides of Eqs. (21) and (22) give lower bounds. This means that for the situations considered previously in [16,21], where the signal and noise both have the same distribution (but different variances), we can expect to find channel capacity that is less than or equal to that of Eq. (26). This is discussed in Sec. III.

The derived sufficient condition of Eq. (25) leads to two ways in which capacity can be achieved: (i) an optimal signal PDF for a given noise PDF, and (ii) an optimal noise PDF for a given signal PDF.

B. Optimizing the signal distribution

Assuming Eq. (20) holds, the channel capacity achieving input PDF $f_X^o(x)$ can be found for any given noise PDF from Eqs. (24) and (25) as

$$f_X^o(x) = \frac{f_\eta(\theta - x)}{\pi\sqrt{F_\eta(\theta - x)[1 - F_\eta(\theta - x)]}}. \tag{27}$$

1. Example: Uniform noise

Suppose the i.i.d. noise at the input to each threshold device in the SSR model is uniformly distributed on the interval $[-\sigma_\eta/2, \sigma_\eta/2]$ so that it has PDF

$$f_\eta(\xi) = \frac{1}{\sigma_\eta}, \quad \xi \in [-\sigma_\eta/2, \sigma_\eta/2]. \tag{28}$$

Substituting Eq. (28) and its associated CDF into Eq. (27), we find that the optimal signal PDF is

$$f_X^o(x) = \frac{1}{\pi\sqrt{\sigma_\eta^2/4 - (x-\theta)^2}}, \quad x \in [\theta - \sigma_\eta/2, \theta + \sigma_\eta/2]. \tag{29}$$

This PDF is in fact the PDF of a sine wave with uniformly random phase, amplitude $\sigma_\eta/2$, and mean θ. A change of variable to the interval $\tau \in [0,1]$ via the substitution $\tau = (x - \theta)/\sigma_\eta + 0.5$ results in the PDF of the Beta distribution with parameters 0.5 and 0.5, also known as the arcsine distribution. As mentioned in Sec. I, this result provides some theoretical justification for the analog-to-digital conversion method proposed in [23].

This Beta distribution is bimodal, with the most probable values of the signal those near zero and unity. Similar results for an optimal input distribution in an information theoretic optimization of a neural system have been found in [38]. These results were achieved numerically using the Blahut-Arimoto algorithm often used in information theory to find channel capacity achieving source distributions, or rate-distortion functions [30].

2. Gaussian noise

Suppose the i.i.d. noise at the input to each threshold device has a zero mean Gaussian distribution with variance σ_η^2, with PDF

$$f_\eta(\xi) = \frac{1}{\sqrt{2\pi\sigma_\eta^2}} \exp\left(-\frac{\xi^2}{2\sigma_\eta^2}\right). \tag{30}$$

Substituting Eq. (30) and its associated CDF into Eq. (27) gives the optimal signal PDF. The resultant expression for $f_X^o(x)$ does not simplify much, and contains the standard error function erf(·) [39].

We are able to verify that the resultant PDF has the correct shape via Fig. 8 in [16], which presents the result of *numerically* optimizing the signal PDF $f_X(x)$ for unity variance zero mean Gaussian noise, $\theta = 0$, and $N = 10\,000$. As with the work in [38], the numerical optimization is achieved using the Blahut-Arimoto algorithm. It is remarked in [16] that the optimal $f_X(x)$ is close to being Gaussian. This is illustrated by plotting both $f_X(x)$ and a Gaussian PDF with nearly the same peak value as $f_X(x)$. It is straightforward to show that a Gaussian with the same peak value as our analytical $f_X^o(x)$ has variance $0.25\pi^2$. If the signal was indeed Gaussian, then we would have $\sigma = 2/\pi \simeq 0.6366$, which is very close to the value calculated for actual Gaussian signal and noise in Sec. III.

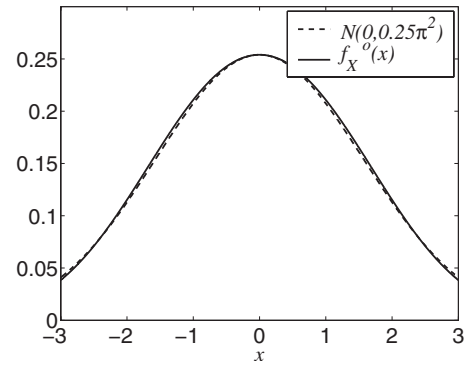

FIG. 2. Optimal signal PDF $f_X^o(x)$ for zero mean, unity variance Gaussian noise, and threshold value $\theta = 0$, as obtained from Eq. (27). Superimposed is a Gaussian PDF with the same peak value as $f_X^o(x)$, so that it has variance $0.25\pi^2$. This figure uses our new theoretical results to analytically replicate Fig. 8 in [16], which was calculated numerically.

Our analytical $f_X^o(x)$ from Eqs. (30) and (27), with $\theta = 0$, is plotted on the interval $x \in [-3, 3]$ in Fig. 2, along with a Gaussian PDF with variance $0.25\pi^2$. Clearly the optimal signal PDF is very close to the Gaussian PDF. Our Fig. 2 is virtually identical to Fig. 8 in [16]. It is emphasized that the results in [16] were obtained using an entirely different method that involves numerical iterations, and therefore provides excellent validation of our theoretical results.

C. Optimizing the noise distribution

We now assume that the signal distribution is known and fixed. We wish to achieve channel capacity by finding the optimal noise distribution. It is easy to show by integrating Eq. (24) that the CDF corresponding to the PDF $f_S(\cdot)$, evaluated at x, can be written in terms of the CDF of the noise distribution as

$$F_S(x) = 1 - \frac{2}{\pi} \arcsin[\sqrt{F_\eta(\theta - x)}]. \tag{31}$$

If we now let $f_X(x) = f_S(x)$, then $F_X(s) = F_S(x)$, and rearranging Eq. (31) gives the optimal noise CDF in terms of the signal CDF as

$$F_\eta^o(x) = \sin^2\left(\frac{\pi}{2}[1 - F_X(\theta - x)]\right) = 0.5 + 0.5\cos[\pi F_X(\theta - x)]. \tag{32}$$

Differentiating $F_\eta^o(x)$ gives the optimal noise PDF as a function of the signal PDF and CDF,

$$f_\eta^o(x) = \frac{\pi}{2}\sin\{\pi[1 - F_X(\theta - x)]\}f_X(\theta - x). \tag{33}$$

Unlike optimizing the signal distribution, which is the standard way for achieving channel capacity in information theory [30], we have assumed a signal distribution, and found the "best" noise distribution, which is equivalent to optimizing the channel, rather than the signal.

1. Example: Uniform signal

Suppose the signal is uniformly distributed on the interval $x \in [-\sigma_x/2, \sigma_x/2]$. From Eqs. (32) and (33), the capacity achieving noise distribution has CDF

$$F_\eta^o(x) = 0.5 + 0.5\sin\left(\frac{\pi(x - \theta)}{\sigma_x}\right), \quad x \in [\theta - \sigma_x/2, \theta + \sigma_x/2] \tag{34}$$

and PDF

$$f_\eta^o(x) = \frac{\pi}{2\sigma_x}\cos\left(\frac{\pi(x - \theta)}{\sigma_x}\right), \quad x \in [\theta - \sigma_x/2, \theta + \sigma_x/2]. \tag{35}$$

Substitution of $F_\eta^o(x)$ and $f_\eta^o(x)$ into Eq. (19) finds the interesting result that the Fisher information is constant for all x,

$$J(x) = N\frac{\pi^2}{\sigma_x^2}. \tag{36}$$

This is verified in Eq. (37) below.

D. Consequences of optimizing the large-N channel capacity

1. Optimal Fisher information

Regardless of whether we optimize the signal for given noise, or optimize the noise for a given signal, it is straightforward to show that the Fisher information can be written as a function of the signal PDF,

$$J(x) = N\pi^2[f_X(x)]^2. \tag{37}$$

Therefore, the Fisher information at large-N channel capacity is constant for the support of the signal if and only if the signal is uniformly distributed. The optimality of constant Fisher information in a neural coding context is studied in [32].

2. The optimal PDF $f_Q(\tau)$

A further consequence that holds in both cases is that the ratio of the signal PDF to the noise PDF is

$$\frac{f_X(x)}{f_\eta(\theta - x)} = \frac{2}{\pi\sin\{\pi[1 - F_X(x)]\}}. \tag{38}$$

This is not a PDF. However, if we make a change of variable via $\tau = 1 - F_\eta(\theta - x)$ we get the PDF $f_Q(\tau)$ discussed in Sec. I C, which for channel capacity is

$$f_Q^o(\tau) = \frac{1}{\pi\sqrt{\tau(1 - \tau)}}, \quad \tau \in [0,1]. \tag{39}$$

This optimal $f_Q(\tau)$ is in fact the PDF of the Beta distribution with parameters 0.5 and 0.5, i.e., the arcsine distribution. It is emphasized that this result holds regardless of whether the signal PDF is optimized for a given noise PDF or vice versa.

3. Output entropy at channel capacity

From Eq. (18), the entropy of Q is equal to the negative of the relative entropy between $f_X(x)$ and $f_\eta(\theta - x)$. The entropy of Q when capacity is achieved can be calculated from Eq. (39) using direct integration as

$$H(Q) = \log_2(\pi) - 2. \tag{40}$$

From Eqs. (21) and (18), the large-N output entropy at channel capacity in the SSR model is

$$H(y) = \log_2\left(\frac{N\pi}{4}\right). \tag{41}$$

4. The optimal output PMF is Beta binomial

Suppose we have signal and noise such that $f_Q(\tau) = f_Q^o(\tau)$—i.e., the signal and noise satisfy the sufficient condition Eq. (25)—but that N is not necessarily large. We can derive the output PMF for this situation, by substituting Eq. (39) into Eq. (17), to get

$$P_y(n) = \binom{N}{n}\frac{1}{\pi}\int_0^1 \tau^{(n-0.5)}(1 - \tau)^{(N-n-0.5)}d\tau = \binom{N}{n}\frac{\beta(n + 0.5, N - n + 0.5)}{\beta(0.5, 0.5)}, \tag{42}$$

where $\beta(a,b)$ is a Beta function. This PMF can be recognized as that of the Beta binomial—or negative hypergeometric—distribution with parameters N, 0.5, 0.5 [40]. It is emphasized that Eq. (42) holds as an exact analytical result for any N.

5. Analytical expression for the mutual information

The exact expression for the output PMF of Eq. (42) allows exact calculation of both the output entropy and the mutual information without need for numerical integration, using Eq. (16). This is because, when $f_Q(\tau) = f_Q^o(\tau)$, the integrals in Eq. (16) can be evaluated exactly to get

$$I_o(X,y) = -\sum_{n=0}^{N} P_y(n)\log_2\left(\frac{P_y(n)}{\binom{N}{n}}\right) + N\log_2\left(\frac{e}{4}\right). \tag{43}$$

The exact values of $I_o(X,y)$ and the corresponding output entropy $H_o(y)$ are plotted in Fig. 3(a) for $N = 1, \ldots, 1000$. For

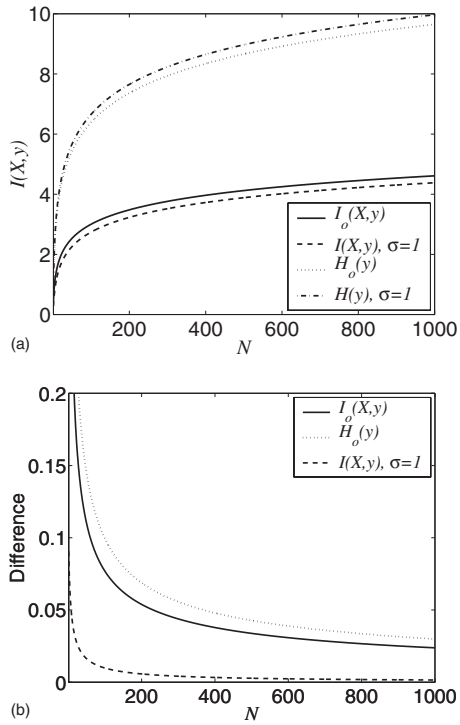

FIG. 3. (a) Exact expressions obtained using $f_Q^o(\tau)$ for $I_o(X,y)$ and $H_o(y)$, as well as the exact mutual information and output entropy when $f_X(x)=f_\eta(\theta-x)$ (denoted as $\sigma=1$), as a function of N. (b) Difference between the exact expressions for $I_o(X,y)$, $H_o(y)$, and $I(X,y)$ for $f_X(x)=f_\eta(\theta-x)$, and the corresponding large-N expressions given by Eqs. (22), (41), and (49).

comparison, the exact $I(X,y)$ of Eq. (12), which holds for $f_X(x)=f_\eta(\theta-x)$, is also plotted, as well as the corresponding entropy $H(y)=\log_2(N+1)$. It is clear that $I_o(X,y)$ is always larger than the mutual information of the $f_X(x)=f_\eta(\theta-x)$ case, and that $H_o(y)$ is always less than its entropy, which is the maximum output entropy.

To illustrate that the large-N expressions derived are lower bounds to the exact formula plotted in Fig. 3(a), and that the error between them decreases with N, Fig. 3(b) shows the difference between the exact and the large-N mutual information and output entropy. This difference clearly decreases with increasing N.

E. A note on the output entropy

The SSR model has been described in terms of signal quantization theory in [28], and compared with the related process of companding in [41]. In this context *quantization* means the conversion of a continuously valued signal to a discretely valued signal that has only a finite number of possible values. Quantization in this sense occurs in analog-to-digital converter circuits, lossy compression algorithms, and histogram formation [42]. For a deterministic scalar quantizer with $N+1$ output states, N threshold values are required. In quantization theory, there is a concept of high-resolution quantizers, in which the distribution of $N\to\infty$ threshold values can be described by a point density function $\lambda(x)$. For such quantizers, it can be shown that the quantizer output y in response to a random variable X has entropy $H(y)\simeq\log_2 N-D(f_X\|\lambda)$ [42]. This is strikingly similar to our Eq. (21) for the large-N output entropy of the SSR model. In fact, since the noise that perturbs the fixed threshold value, θ, is additive, each threshold acts as an i.i.d. random variable with PDF $f_\eta(\theta-x)$, and therefore, for large N, $f_\eta(\theta-x)$ acts as a density function describing the relative frequency of threshold values as a function of x, just as $\lambda(x)$ does for a high-resolution deterministic quantizer.

For deterministic quantizers, the point density function can be used to approximate the high-resolution distortion incurred by the quantization process. For the SSR model, however, since the quantization has a random aspect, the distortion has a component due to randomness as well as lossy compression, and cannot be simply calculated from $f_\eta(\cdot)$. Instead, one can use the Fisher information to calculate the asymptotic mean square error distortion, which is not possible for deterministic high-resolution quantizers.

III. CHANNEL CAPACITY FOR LARGE N AND "MATCHED" SIGNAL AND NOISE

Unlike the previous section, we now consider channel capacity under the constraint of matched signal and noise distributions—i.e., where both the signal and noise, while still independent, have the same distribution, other than their variances. The mean of both signal and noise is zero and the threshold value is also $\theta=0$. In this situation the mutual information depends solely on the ratio $\sigma=\sigma_\eta/\sigma_x$, which is the only free variable. Finding channel capacity is therefore equivalent to finding the optimal value of noise intensity σ. Such an analysis provides verification of the more general capacity expression of Eq. (26), which cannot be exceeded.

Furthermore, inspection of Eq. (A10) shows that the large-N approximation to the mutual information consists of a term that depends on N and a term that depends only on σ. This shows that for large N the channel capacity occurs for the same value of σ—which we denote as σ_o—for all N.

This fact is recognized in both [21] for uniform signal and noise—where $\sigma_o\to 1$—and [16] for Gaussian signal and noise. Here, we investigate the value of σ_o and the mutual information at σ_o for other signal and noise distributions, and compare the channel capacity obtained with the case where $f_X(x)=f_S(x)$. This comparison finds that the results of [16] overstate the true capacity, and that large-N results in [11,21] need to be improved to be consistent with the central results of this paper.

From Eq. (22), channel capacity for large N occurs for the value of σ that minimizes the relative entropy between f_X and f_S. If we let

$$f(\sigma)=\int_{x=-\infty}^{x=\infty} f_X(x)\ln\left(\frac{1}{J(x)}\right)dx, \tag{44}$$

then from Eq. (20) it is also clear that this minimization is equivalent to solving the following problem:

TABLE II. Large-N channel capacity and optimal σ for matched signal and noise.

Distribution	$C(X,y)-0.5\log_2(N)$	σ_o	$C(X,y)-I(X,y)\|_{\sigma=1}$
Gaussian	−0.3964	0.6563	0.208
Logistic	−0.3996	0.5943	0.205
Laplacian	−0.3990	0.5384	0.205

$$\sigma_o = \min_{\sigma} f(\sigma). \tag{45}$$

This is exactly the formulation stated in [16]. Problem (45) can be equivalently expressed as

$$\sigma_o = \min_{\sigma}\left(f(\sigma) = D(f_X \| f_\eta) + \int_{x=-\infty}^{x=\infty} f_X(x)\log_2(P_{1|x})dx \right), \tag{46}$$

where we have assumed that both the signal and noise PDFs are even functions. The function $f(\sigma)$ can be found for any specified signal and noise distribution by numerical integration, and problem (46) easily solved numerically. If an exact expression for the relative entropy term is known, then only $g(\sigma)=\int_{x=-\infty}^{x=\infty} f_X(x)\log_2(P_{1|x})dx$ needs to be numerically calculated.

Table II gives the result of numerically calculating the value of σ_o, and the corresponding large-N channel capacity, $C(X,y)$, for a number of distributions. In each case, $C(X,y)-0.5\log_2(N)<-0.3956$, as required by Eq. (26). The difference between capacity and $0.5\log_2(N)$ is about 0.4 bits per sample. In the limit of large N, this shows that capacity is almost identical, regardless of the distribution. However, the value of σ_o at which this capacity occurs is different in each case.

As discussed in Sec. I B, the mutual information is identical whenever the signal and noise PDFs are identical, i.e., $\sigma=1$. It is shown below in Eq. (48) that for large N the mutual information at $\sigma=1$ is $I(X,y)=0.5\log_2(N)-0.6444$. Given that the channel capacity is slightly larger than this, as indicated by Table II, for each case there is a constant difference between the channel capacity and the mutual information at $\sigma=1$. This value is also listed in Table II.

A. Improvements to previous large-N approximations

We now use the results of Sec. II to show that previous large-N expressions for the mutual information in the literature for the $\sigma=1$, Gaussian, and uniform cases can be improved.

1. SSR for large N and $\sigma=1$

We now consider the situation where $f_X(x)=f_\eta(x)$, so that $\sigma=1$. It is shown in [11] that in this case as N approaches infinity, Eq. (12) reduces to

$$I(X,y) \simeq 0.5\log_2\left(\frac{N+1}{e}\right) \simeq 0.5\log_2(N+1) - 0.7213. \tag{47}$$

To show that this expression can be improved, we begin with the version of Eq. (20) given by Eq. (A10). When $\sigma=1$ we have $f_Q(\tau)=1$ and $H(Q)=0$. The integrals in Eq. (A10) can be solved to give the large-N mutual information at $\sigma=1$ as

$$I(X,y) \simeq 0.5\log_2\left(\frac{Ne}{2\pi}\right) \simeq 0.5\log_2 N - 0.6044. \tag{48}$$

Although Eqs. (47) and (48) agree as $N\to\infty$, the constant terms do not agree. It is shown in Appendix A 3 that the discrepancy can be resolved by improving on the approximation to the average conditional entropy, $H(y|X)$, made in [11]. The output entropy at $\sigma=1$ can be shown to be simply $H(y)=\log_2(N+1)$ [11]. Subtracting Eq. (A18) from $H(y)$ and letting N approach infinity gives

$$I(X,y) \simeq 0.5\log_2\left(\frac{(N+2)e}{2\pi}\right), \tag{49}$$

which does have a constant term which agrees with Eq. (48). The explanation of the discrepancy is that [11] uses the Euler-Maclaurin summation formula to implicitly calculate $\log_2(N!)$ in the large-N approximation to $H(y|X)$. Using Stirling's approximation for $N!$, as done here, gives a more accurate approximation.

The increased accuracy of Eq. (48) can be verified by numerically comparing both Eq. (48) and Eq. (47) with the exact expression for $I(X,y)$ of Eq. (12), as N increases. The error between the exact expression and Eq. (48) approaches zero as N increases, whereas the error between Eq. (12) and Eq. (47) approaches a nonzero constant for large N of $0.5\log_2(e^2/2\pi)\simeq 0.117$ bits per sample.

2. Uniform signal and noise

A derivation is given in [21] of an exact expression for $I(X,y)$ for uniform signal and noise and $\sigma\leqslant 1$. In addition, [21] finds a large-N approximation to the mutual information. Using the same arguments as for the $\sigma=1$ case, this approximation can be improved to

$$I(X,y) \simeq \frac{\sigma}{2}\log_2\left(\frac{(N+2)e}{2\pi}\right) + (1-\sigma)[1-\log_2(1-\sigma)] - \sigma\log_2(\sigma). \tag{50}$$

The accuracy of Eq. (50) can be verified by numerical comparison with the exact formula in [21], as N increases. If one replicates Fig. 3 of [21] in this manner, it is clear that Eq. (50) is the more accurate approximation.

Differentiating Eq. (50) with respect to σ and setting to zero obtains the optimal value of σ as

$$\sigma_o = \frac{\sqrt{(N+2)}}{\sqrt{(N+2)}+\sqrt{(8\pi/e)}}. \tag{51}$$

The channel capacity at σ_o is

$$C(X,y) = 1 - \log_2(1-\sigma_o) = \log_2\left(2 + \sqrt{\frac{(N+2)e}{2\pi}}\right). \tag{52}$$

Clearly, $\lim_{N\to\infty}\sigma_o = 1$, and the capacity approaches $0.5\log_2[(N+2)e/(2\pi)]$, which agrees with Eq. (49). Expressions for σ_o and the corresponding capacity for large N are also given in [21]. Again, these are slightly different to Eqs. (51) and (52), due to the slightly inaccurate terms in the large-N approximation to $H(y|X)$. However, the important qualitative result remains the same, which is that the channel capacity scales with $0.5\log_2(N)$ and the value of σ which achieves this asymptotically approaches unity.

3. Gaussian signal and noise

In [16], an analytical approximation for σ_o for the specific case of Gaussian signal and noise is derived using a Taylor expansion of the Fisher information inside the integral in Eq. (20). We give a slightly different derivation of this approach that uses the PDF $f_Q(\tau)$.

We begin with problem (46). Solving this problem requires differentiating $f(\sigma)$ with respect to σ and solving for zero. From Table I, the derivative of the relative entropy between f_X and f_η is

$$\frac{d}{d\sigma}D(f_X\|f_\eta) = \frac{1}{\ln 2}(\sigma^{-1} - \sigma^{-3}). \tag{53}$$

For the second term $g(\sigma)$, we take the lead from [16] and approximate $\ln(P_{1|x})$ by its second-order Taylor series expansion [39]. The result is that

$$g(\sigma) = -\int_{x=-\infty}^{x=\infty} f_X(x)\log_2(P_{1|x})dx \simeq 1 + \frac{1}{\pi\sigma^2\ln 2}. \tag{54}$$

Numerical testing finds that the approximation of Eq. (54) appears to be quite accurate for all σ, as the relative error is no more than about 10% for $\sigma > 0.2$. However, as we will see, this is inaccurate enough to cause the end result for the approximate channel capacity to significantly overstate the true channel capacity.

Taking the derivative of Eq. (54) with respect to σ, subtracting it from Eq. (53), setting the result to zero and solving for σ gives the optimal value of σ found in [16], $\sigma_o \simeq \sqrt{1-2/\pi} \simeq 0.6028$.

An expression for the mutual information at σ_o can be found by back substitution. Carrying this out gives the large-N channel capacity for Gaussian signal and noise as

$$C(X,y) \simeq 0.5\log_2\left(\frac{2N}{e(\pi-2)}\right), \tag{55}$$

which can be written as $C(X,y) \simeq 0.5\log_2 N - 0.3169$.

Although Eq. (55) is close to correct, recall from Sec. II that capacity must be less than $0.5\log_2 N - 0.3956$ and hence Eq. (55) significantly overstates the true capacity.

ACKNOWLEDGMENTS

This work was funded by the Australian Research Council, the Australian Academy of Science, as well as EPSRC Grant No. EP/C523334/1, and we gratefully acknowledge this support. The authors would also like to thank Priscilla Greenwood of Arizona State University, Pierre-Olivier Amblard of Laboratoire des Images et des Signaux, France, Thinh Nguyen of Oregon State University, and Simon Durrant of the University of Plymouth, U.K., for valuable discussions and generous provision of manuscripts prior to publication, and the anonymous referees, whose comments have led to significant improvements in this paper.

APPENDIX A: DERIVATIONS

1. Mutual information for large N and arbitrary σ

This appendix contains derivations of the large-N approximations to the output entropy and mutual information discussed in Sec. II.

a. 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 distribution with the same mean and variance—i.e., $NP_{1|x}$ and $NP_{1|x}(1-P_{1|x})$, respectively. Provided $0 \ll NP_{1|x} \ll N$ we have

$$P_{y|X}(n|x) \simeq \frac{1}{\sqrt{2\pi NP_{1|x}(1-P_{1|x})}}\exp\left(-\frac{(n-NP_{1|x})^2}{2NP_{1|x}(1-P_{1|x})}\right). \tag{A1}$$

The average conditional output entropy is $H(y|X) = \int_x f_X(x)\hat{H}(y|x)dx$, where

$$\hat{H}(y|x) = -\sum_{n=0}^{N} P_{y|X}(n|x)\log_2[P_{y|X}(n|x)]. \tag{A2}$$

Using the well-known result for the entropy of a Gaussian random variable [30], we can write

$$\hat{H}(y|x) \simeq 0.5\log_2[2\pi eNP_{1|x}(1-P_{1|x})]. \tag{A3}$$

Multiplying both sides of Eq. (A3) by $f_X(x)$ and integrating over all x gives

$$\begin{aligned} H(y|X) &\simeq 0.5\log_2(2\pi eN) \\ &\quad + 0.5\int_{x=-\infty}^{\infty} f_X(x)\log_2[P_{1|x}(1-P_{1|x})]dx \\ &= 0.5\log_2(2\pi eN) + 0.5\int_{\tau=0}^{\tau=1} f_Q(\tau)\log_2(\tau)d\tau \\ &\quad + 0.5\int_{\tau=0}^{\tau=1} f_Q(\tau)\log_2(1-\tau)d\tau. \end{aligned} \tag{A4}$$

Equation (A4) can be verified for the case where $f_X(x)=f_\eta(\theta-x)$, since this means $f_Q(\tau)=1$ and $\int_{\tau=0}^{\tau=1} f_Q(\tau)\log_2(\tau)d\tau=-\log_2(e)$. Consequently Eq. (A4) reduces to $H(y|X)\simeq 0.5\log_2(2\pi N/e)$ which agrees precisely with Eq. (A19). This approximation breaks down when $P_{1|x}$ is close to zero or unity. Furthermore, Eq. (A4) holds exactly only for values of x for which $P_{y|X}(n|x)$ is exactly Gaussian. Otherwise, $H(y|X)$ is strictly less than the approximation given.

b. Output distribution and entropy

For large N, since $P_{y|X}(n|x)$ is Gaussian, y/N approaches a δ function located at $P_{1|x}=n/N$. From Eqs. (7) and (17), this means that $P_y(n)$ can be written in terms of the PDF of the average transfer function, $f_Q(\cdot)$, as

$$P_y(n)\simeq\frac{f_Q(n/N)}{N}. \tag{A5}$$

This result can be derived more rigorously using saddle-point methods [22].

Consider the case where the signal and noise both have the same distribution but different variances. When the noise intensity $\sigma>1$, then $f_Q(0)=f_Q(1)=0$, whereas for $\sigma<1$, we have $f_Q(0)=f_Q(1)=\infty$. From Eq. (A5), this means $P_y(0)$ and $P_y(N)$ are either zero or infinite. However, for finite N, there is some finite nonzero probability that all output states are on or off. Indeed, at $\sigma=1$, we know that $P_y(n)=1/(N+1)\ \forall\ n$, and at $\sigma=0$, $P_y(0)=P_y(N)=0.5$. Furthermore, for finite N, Eq. (A5) does not guarantee that $\Sigma_{n=0}^{N}P_y(n)=1$. To increase the accuracy of our approximation by ensuring that $P_y(0)$ and $P_y(N)$ are always finite, and that $P_y(n)$ forms a valid PMF, we define a new approximation as

$$P'_y(n)=\begin{cases}\dfrac{f_Q(n/N)}{N} & \text{for } n=1,\ldots,N-1,\\ 0.5\left(1-\displaystyle\sum_{m=1}^{N-1}\dfrac{f_Q(m/N)}{N}\right) & \text{for } n=0,\ n=N.\end{cases} \tag{A6}$$

Figure 4 shows that the approximation given by $P'_y(n)$ is highly accurate for N as small as 63, for σ both smaller and larger than unity.

Consider the entropy of the discrete random variable y. Making use of Eq. (A5), we have

$$\begin{aligned}H(y)&=-\sum_{n=0}^{N}P_y(n)\log_2[P_y(n)]\\&=-\frac{1}{N}\sum_{n=0}^{N}f_Q\left(\frac{n}{N}\right)\log_2\left[f_Q\left(\frac{n}{N}\right)\right]+\frac{\log_2(N)}{N}\sum_{n=0}^{N}f_Q\left(\frac{n}{N}\right).\end{aligned} \tag{A7}$$

Suppose that the summations above can be approximated by integrals, without any remainder terms. Carrying this out and then making the change of variable $\tau=n/N$ gives

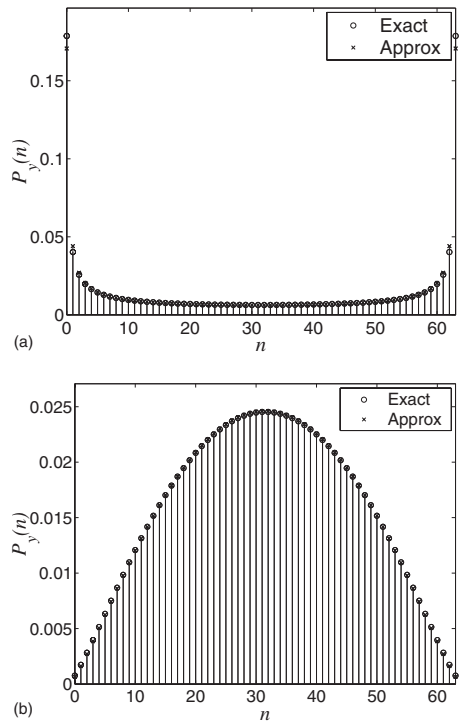

FIG. 4. Approximation to the output PMF $P_y(n)$, given by Eq. (A6), for $N=63$. Circles indicate the exact $P_y(n)$ obtained by numerical integration and the crosses show approximations. $\sigma=$(a) 0.4 and (b) 1.6, Gaussian signal and noise.

$$H(y)\simeq\log_2 N-\int_{\tau=0}^{\tau=1}f_Q(\tau)\log_2[f_Q(\tau)]d\tau=\log_2 N+H(Q), \tag{A8}$$

where $H(Q)$ is the differential entropy of the random variable Q. Performing a change of variable in Eq. (A8) of $\tau=1-F_\eta(\theta-x)$ gives

$$H(y)\simeq\log_2(N)-D(f_X(x)\|f_\eta(\theta-x)). \tag{A9}$$

This result 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 f_X and f_η. Therefore, since relative entropy is always non-negative, the approximation to $H(y)$ given by Eq. (A9) 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 holds for any N.

c. Mutual information

Subtracting Eq. (A4) from Eq. (A8) gives a large-N approximation to the mutual information as

$$\begin{aligned}I(X,y)\simeq 0.5\log_2\left(\frac{N}{2\pi e}\right)+H(Q)-0.5\\ \times\int_{\tau=0}^{\tau=1}f_Q(\tau)\log_2[\tau(1-\tau)]d\tau.\end{aligned} \tag{A10}$$

As discussed in the main text, the mutual information scales with $0.5\log_2(N)$. The importance of the N-independent terms

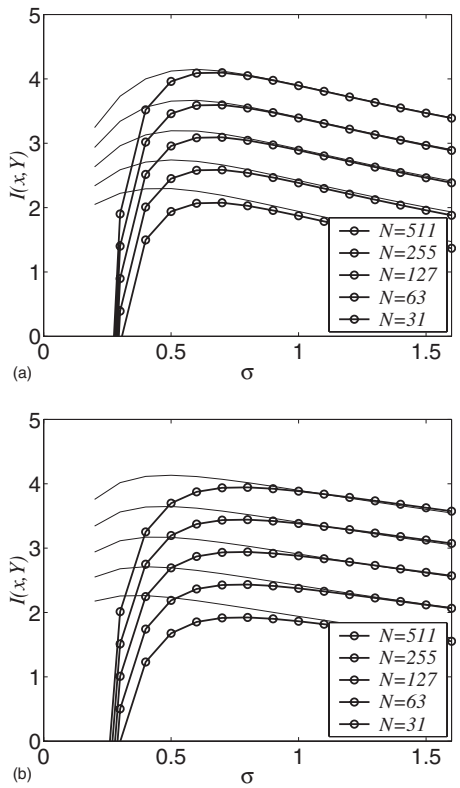

FIG. 5. Large-N approximation to mutual information given by Eq. (A10) and exact mutual information calculated numerically. The exact expression is shown by thin solid lines, and the approximation by circles, with a thicker solid line interpolating between values of σ as an aid to the eye. The approximation can be seen to always be a lower bound on the exact mutual information. (a) Gaussian and (b) Laplacian.

in Eq. (A10) is that they determine how the mutual information varies from $0.5\log_2(N/2\pi e)$ for different PDFs $f_Q(\tau)$.

Figure 5 shows, as examples, the approximation of Eq. (A10), as well as the exact mutual information—calculated by numerical integration—for the Gaussian and Laplacian cases, for a range of σ and increasing N. As with the output entropy, the mutual information approximation is quite good for $\sigma > 0.7$, but worsens for smaller σ. However, as N increases the approximation improves.

Equation (A10) can be rewritten via the change of variable, $x=\theta-F_\eta^{-1}(1-\tau)$, as

$$I(X,y) = 0.5\log_2\left(\frac{N}{2\pi e}\right) - \int_{x=-\infty}^{x=\infty} f_X(x)\log_2[P_{1|x}(1-P_{1|x})]dx - D(f_X(x)\|f_\eta(\theta-x)). \tag{A11}$$

Rearranging Eq. (A11) gives Eq. (20)—with the Fisher information $J(x)$ given by Eq. (19)—which is precisely the same as that derived in [16] as an asymptotic large-N expression for the mutual information. Our analysis extends [16] by finding large-N approximations to both $H(y)$ and $H(y|X)$, as well as the output distribution $P_y(n)$. We have also illustrated the role of the PDF $f_Q(\tau)$ in these approximations, and justified the use of Eq. (20) for the SSR model.

2. Proof that $f_S(x)$ is a PDF

As shown in [16], the Fisher information for the SSR model is given by Eq. (19). Consider $f_S(x)$ as in Eq. (24). Since $f_\eta(x)$ is a PDF and $F_\eta(x)$ is the CDF of η evaluated at x, we have $f_S(x) \geqslant 0\ \forall\ x$. Letting $h(x)=F_\eta(\theta-x)$, Eq. (24) can be written as

$$f_S(x) = \frac{-h'(x)}{\pi\sqrt{h(x)-h(x)^2}}. \tag{A12}$$

Suppose $f_\eta(x)$ has support $x\in[-a,a]$. Integrating $f_S(x)$ over all x gives

$$\begin{aligned}\int_{x=-a}^{x=a} f_S(x)dx &= \int_{x=-a}^{x=a} \frac{-h'(x)}{\pi\sqrt{h(x)-h(x)^2}}dx \\ &= -\frac{1}{\pi}\{2\arcsin[\sqrt{h(x)}]|_{x=-a}^{x=a}\} \\ &= -\frac{2}{\pi}[\arcsin(0)-\arcsin(1)] = 1,\end{aligned} \tag{A13}$$

which means $f_S(x)$ is a PDF.

3. $H(y|X)$ for large N and $\sigma=1$

Here we derive a large-N approximation to $H(y|X)$ used in Sec. III A 1. For $\sigma=1$ the output PMF is $P_y(n)=1/(N+1)\ \forall\ n$ [11]. Using this, it can be shown that

$$-\sum_{n=0}^{N} P_y(n)\log_2\binom{N}{n} = \log_2(N!) - \frac{2}{N+1}\sum_{n=1}^{N} n\log_2 n. \tag{A14}$$

We will now see that both terms of Eq. (A14) can be simplified by approximations that hold for large N. First, for the $\log_2(N!)$ term, we can make use of Stirling's formula [39], which is valid for large N,

$$N! \sim \sqrt{(2\pi N)}N^N\exp(-N). \tag{A15}$$

This approximation is particularly accurate if the logarithm is taken of both sides, which we require. Second, the sum in the second term of Eq. (A14) can be approximated by an integral and simplified by way of the Euler-Maclaurin summation formula [39]. The result is

$$\frac{2}{N+1}\sum_{n=1}^{N} n\log_2 n \simeq N\log_2(N+1) - \frac{N(N+2)}{2\ln 2(N+1)} + O\left(\frac{\log N}{N}\right). \tag{A16}$$

Subtracting Eq. (A16) from the logarithm of Eq. (A15) gives

$$-\sum_{n=0}^{N} P_y(n)\log_2\binom{N}{n} \simeq 0.5\log_2\left(\frac{N}{e^2}\right) - \frac{N}{2\ln 2}\left(2-\frac{N+2}{N+1}\right) + 0.5\log_2(2\pi) - O\left(\frac{\log N}{N}\right), \quad \text{(A17)}$$

where we have used $N\log_2(1+1/N)=1/\ln 2+O(1/N)$. When Eq. (A17) is substituted into an exact expression for $H(y|X)$ given in [11], we get

$$H(y|X) = \frac{N}{2\ln 2} - \sum_{n=0}^{N} P_y(n)\log_2\binom{N}{n} \simeq 0.5\log_2 N + 0.5\left(\frac{N}{N+1}-2\right)\log_2(e) + 0.5\log_2(2\pi) - O\left(\frac{\log N}{N}\right). \quad \text{(A18)}$$

The final result is that, for large N,

$$H(y|X) \simeq 0.5\log_2\left(\frac{2\pi N}{e}\right). \quad \text{(A19)}$$

[1] L. Gammaitoni, P. Hänggi, P. Jung, and F. Marchesoni, Rev. Mod. Phys. **70**, 223 (1998).
[2] F. Moss, L. M. Ward, and W. G. Sannita, Clin. Neurophysiol. **115**, 267 (2004).
[3] K. Wiesenfeld, D. Pierson, E. Pantazelou, C. Dames, and F. Moss, Phys. Rev. Lett. **72**, 2125 (1994).
[4] A. R. Bulsara and A. Zador, Phys. Rev. E **54**, R2185 (1996).
[5] F. Chapeau-Blondeau, Phys. Rev. E **53**, 5469 (1996).
[6] P. E. Greenwood, L. M. Ward, and W. Wefelmeyer, Phys. Rev. E **60**, 4687 (1999).
[7] M. DeWeese and W. Bialek, Nuovo Cimento Soc. Ital. Fis., D **17**, 733 (1995).
[8] N. G. Stocks, Phys. Rev. Lett. **84**, 2310 (2000).
[9] J. J. Collins, C. C. Chow, and T. T. Imhoff, Phys. Rev. E **52**, R3321 (1995).
[10] R. P. Morse and P. Roper, Phys. Rev. E **61**, 5683 (2000).
[11] N. G. Stocks, Phys. Rev. E **63**, 041114 (2001).
[12] M. D. McDonnell, N. G. Stocks, C. E. M. Pearce, and D. Abbott, Phys. Lett. A **352**, 183 (2006).
[13] N. G. Stocks, D. Allingham, and R. P. Morse, Fluct. Noise Lett. **2**, L169 (2002).
[14] W. Bialek, M. DeWeese, F. Rieke, and D. Warland, Physica A **200**, 581 (1993).
[15] J. F. Lindner, M. Bennett, and K. Wiesenfeld, Phys. Rev. E **72**, 051911 (2005).
[16] T. Hoch, G. Wenning, and K. Obermayer, Phys. Rev. E **68**, 011911 (2003).
[17] S. Durrant and J. Feng (unpublished).
[18] A. Nikitin, N. G. Stocks, and R. P. Morse, Phys. Rev. E **75**, 021121 (2007).
[19] N. G. Stocks and R. Mannella, Phys. Rev. E **64**, 030902(R) (2001).
[20] R. P. Morse, P. F. Morse, T. B. Nunn, K. A. M. Archer, and P. Boyle, J. Assoc. Res. Otolaryngol. **8**, 42 (2007).
[21] N. G. Stocks, Phys. Lett. A **279**, 308 (2001).
[22] N. Brunel and J. Nadal, Neural Comput. **10**, 1731 (1998).
[23] T. Nguyen, IEEE Trans. Signal Process. **55**, 2735 (2007).
[24] M. D. McDonnell, D. Abbott, and C. E. M. Pearce, Microelectron. J. **33**, 1079 (2002).
[25] M. D. McDonnell, D. Abbott, and C. E. M. Pearce, Fluct. Noise Lett. **2**, L205 (2002).
[26] D. Rousseau, F. Duan, and F. Chapeau-Blondeau, Phys. Rev. E **68**, 031107 (2003).
[27] D. Rousseau and F. Chapeau-Blondeau, Phys. Lett. A **321**, 280 (2004).
[28] M. D. McDonnell, N. G. Stocks, C. E. M. Pearce, and D. Abbott, Fluct. Noise Lett. **5**, L457 (2005).
[29] Y. Wang and L. Wu, Fluct. Noise Lett. **5**, L435 (2005).
[30] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (John Wiley and Sons, New York, 1991).
[31] E. L. Lehmann and G. Casella, *Theory of Point Estimation* (Springer, New York, 1998).
[32] M. Bethge, D. Rotermund, and K. Pawelzik, Neural Comput. **14**, 2317 (2002).
[33] P. E. Greenwood, U. U. Müller, and L. M. Ward, Phys. Rev. E **70**, 051110 (2004).
[34] M. Stemmler, Network Comput. Neural Syst. **7**, 687 (1996).
[35] P. E. Greenwood, L. M. Ward, D. F. Russell, A. Neiman, and F. Moss, Phys. Rev. Lett. **84**, 4773 (2000).
[36] K. Kang and H. Sompolinsky, Phys. Rev. Lett. **86**, 4958 (2001).
[37] B. S. Clarke and A. R. Barron, IEEE Trans. Inf. Theory **36**, 453 (1990).
[38] S. Schreiber, C. K. Machens, A. V. Herz, and S. B. Laughlin, Neural Comput. **14**, 1323 (2002).
[39] M. R. Spiegel and J. Liu, *Mathematical Handbook of Formulas and Tables* (McGraw-Hill, New York, 1999).
[40] M. Evans, N. Hastings, and B. Peacock, *Statistical Distributions*, 3rd ed. (John Wiley and Sons, New York, 2000).
[41] P.-O. Amblard, S. Zozor, and O. J. J. Michel, in *Proceedings of the 2006 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2006* (IEEE, Toulouse, France, 2006), Vol. 3, pp. 716–719.
[42] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression* (Kluwer Academic, Dordrecht, 1992).