Information capacity of stochastic pooling networks is achieved by discrete inputs

Mark D. McDonnell*

Institute for Telecommunications Research, University of South Australia, SA 5095, Australia

(Received 18 December 2008; published 6 April 2009)

Stochastic pooling networks (SPN) are sensor networks where multiple sensors make independently noisy and compressed measurements of the same information source, which are combined via *pooling*. Examples of SPNs range from nanoelectronics to biological sensory neurons. Here it is shown that optimal information transmission in SPNs with nodes that quantize to a finite number of states requires the input signal distribution to be discrete. This is illustrated numerically for a simple SPN consisting of N binary-quantizing sensors. The resultant information capacity is shown to be independent of the noise distribution when the signal distribution can be freely chosen, but to imply an optimal noise distribution if the signal distribution is fixed. While larger than the best performance of previously studied continuously valued input signals, the capacity does not scale faster than the previous best result of $\log_2(\sqrt{N})$ bits per channel use. It is also shown that a plot of the optimal input distribution contains bifurcations as N increases, and that suprathreshold stochastic resonance occurs when the mutual information is determined for a suboptimal noise distribution.

DOI: 10.1103/PhysRevE.79.041107 PACS number(s): 05.40.Ca, 02.50.−r, 87.19.lo, 89.70.−a

I. INTRODUCTION AND BACKGROUND

The goal of both artificial sensors and biological senses is to accurately transduce and represent an information source. Often, accuracy in representation is achieved in the presence of perturbing noise via the redundancy inherent in multiple independent measurements of the same source, for example noise reduction by beamforming or coherent averaging. If the measurements need to be stored or communicated prior to use, there is also a need for efficient representation via *compression*.

Stochastic pooling networks (SPNs) are an abstract framework for modeling some sensor networks where both signal compression and noise reduction via averaging occur simultaneously. Both can be viewed as emergent properties of the physical manner in which such networks combine individual measurements and the nonlinear interaction between redundancy, lossy compression and random noise [1,2]. SPNs are intended to model either (i) engineered sensors where energy or physical constraints do not permit use of optimal signal processing techniques, e.g., noisy analog-to-digital converters [1,3,4] or nanoscale electronics [5]; (ii) biological sensory processing, such as parallel neurons, receptor cells, or synapses [6–10]. The suboptimality of SPNs compared to ideal engineered sensor networks lead to some surprising emergent phenomena—see [2].

Part of the current paper is focused on suprathreshold stochastic resonance (SSR), a counter-intuitive emergent property of some SPNs that occurs when random noise is required for good performance [11,12]. However, the main result is determination of the previously unknown upper limit to the information transmission performance of finite size SPNs, that is, its *information capacity*, or *channel capacity*. For a memoryless input signal (which is assumed here), this is defined as the maximum achievable *mutual information* between two random variables [13], which in this case are the SPN input and output. Finding it generally means assuming a fixed channel and searching for the optimal input distribution [13].

Only the simple N-node SPN often studied in the context of SSR [3,8,11,14–16] or noise-enhanced detection [1,17] will be considered here. This SPN has N identical binary-quantizing nodes, and an output that *pools* its N measurements by summation. However, the results reported here can easily be generalized to any SPN for which the output signal is a discrete random variable.

Previous work on information transmission in this SPN has always assumed the input is a continuous random variable, e.g., [3,8,11,14,16,18], and capacity results have been found only in the large-N limit [8,12]. In all cases, mutual information is known to scale with $\log_2(\sqrt{N})$. What has not been known is whether there exists some input distribution for which the capacity for *small* N is significantly larger than reported previously.

Here it is shown that capacity also scales with $\log_2(\sqrt{N})$ for small N. This result is achieved by numerical calculation of the signal distribution that achieves maximum mutual information, via an algorithm that relies on some fundamental properties of all memoryless information channels that have a finite number of input or output states. In particular, it has not been previously realized that capacity for the SPN considered here (and any SPN for which the output signal is a discrete random variable) is achieved by an input signal with a *discrete* distribution since the output is discrete and finite. This claim is presented as the following restatement of [19] (Corollaries 2 and 3, pp. 96), for the context of SPNs:

Corollary 1. Channel capacity in an N-node SPN, where each node's output is a discrete random variable, is achieved by input signals that have a discrete distribution and a finite number of mass points. If the smallest number of points that achieves capacity is M^o, then for this M^o a unique capacity achieving input distribution exists, and M^o must be no larger than the number of output states.

Note that for the SPN considered in this paper, with identical binary nodes and pooling by summation, Corollary 1 says that $M^o \leq N+1$.

*mark.mcdonnell@unisa.edu.au

The focus of the current paper is similar to other recent work aimed at establishing the performance limits of systems that exhibit stochastic resonance (SR) [20], or more precisely, aperiodic stochastic resonance (ASR) [21–23] (SSR is a form of ASR). For example, calculations of the noise distribution that optimizes performance for suboptimal randomized detectors are given in [24]. A second example are theorems that provide simple conditions for predicting when SR will occur for random binary signals in static threshold systems [25] and spiking neuron models [26]. These are collectively known as the *forbidden interval theorem* [25], and provide proof that SR effects should be expected in nearly all threshold systems.

When the consequences of the forbidden interval theorem are combined with the observation that SR occurs for impulsive (infinite variance) noise [27], it is clear that SR is very robust. This fact highlights that although published SR research usually includes an assumption of finite variance noise—very often Gaussian—this is not a necessary condition for SR to occur. The optimal signal distribution, i.e., the one that best utilizes the fact that SR may occur, may be very different to Gaussian, as illustrated in this paper. Indeed, Corollary 1 also applies to previous work demonstrating subthreshold ASR in static threshold systems via calculations of mutual information. For example, the system of [23,25] can be viewed as a trivial SPN where $N=1$, and $M^o=2$, and equivalent to a binary asymmetric channel [28]. Therefore the binary input signal used in [23,25] is optimal in terms of the number of mass points (although not in the mass locations, given the subthreshold assumption used).

The remainder of this paper provides verification of Corollary 1 via comparison of the numerical capacity results of Sec. III with previous analytical results (listed in Sec. II) for mutual information in the binary-node SPN. Results providing an upper bound for the channel capacity and a lower bound on M^o are also presented. It is shown that suboptimal discrete input distributions give rise to SSR, and that the capacity results can be interpreted in terms of finding an optimal *noise distribution* for a fixed discrete signal distribution. The paper finishes in Sec. IV with discussion of the significance of the optimality of discrete distributions for SPNs in general, as well as for aperiodic stochastic resonance, SSR, and biological neurons. First however some notation is defined, and the capacity problem is described mathematically.

A. Problem formulation

Consider an N-node SPN where each sensor quantizes its input z_i using a threshold value θ to produce the binary output

$$y_i(z_i) = \begin{cases} 1 & z_i \geq \theta \\ 0 & z_i < \theta \end{cases} \qquad i = 1, \ldots, N. \tag{1}$$

Let the output of this SPN be the sum of the binary measurements, $y=\Sigma_i y_i$. This is a discrete random variable defined on the non-negative integers $n=0,\ldots,N$, with probability mass function (PMF) $P_y(n)$.

In general, all nodes of an SPN operate on independently noisy versions of the same signal sample, x [2]. For *iid* additive noise, η_i, the input to each sensor is $z_i=x+\eta_i$. For the binary-node case of Eq. (1), if the cumulative distribution function (CDF) of the noise is $F_\eta(\cdot)$, define the probability that a sensor produces a 1 in response to x as

$$P_{1|x} = 1 - F_\eta(\theta - x). \tag{2}$$

An SPN can be thought of as a communications channel where measurements of a random signal x are corrupted via noisy transmission. The channel is entirely described by the *transition probabilities*, i.e., the conditional distribution of the output given the input. For a binary-quantizing SPN with N nodes, this is given by the binomial distribution [11],

$$P_{y|x}(y = n|x) = \binom{N}{n}(P_{1|x})^n(1 - P_{1|x})^{N-n}, \tag{3}$$

where $n=0,\ldots,N$. If the SPN's input has CDF $F_x(x)$, the output PMF is $P_y(n)=\int_x dF_x(x)P_{y|x}(n|x)$, while the mutual information is the difference between the entropy of the output, $H(y)=-\Sigma_{n=0}^N P_y(n)\log_2 P_y(n)$, and the average conditional entropy, $H(y|x)=\int_x dF_x(x)\Sigma_{n=0}^N P_{y|x}(n|x)\log_2 P_{y|x}(n|x)$, i.e.,

$$I(x;y) = H(y) - H(y|x). \tag{4}$$

Channel capacity can be expressed as

$$C(x;y) = \max_{\{F_x(x)\}} I(x;y). \tag{5}$$

In some practical cases finding capacity requires additional constraints on the source distribution such as maximum amplitude or average power [13]. Here no such constraint need be considered, although it is certainly possible to do so.

Note that $I(x;y)=I[x;(y_1,\ldots,y_N)]$, which means that summation of the N nodes' outputs does not reduce information in any way. This is due to the summation being a sufficient statistic for the vector of individual node outputs [2]. This fact captures what is meant by *pooling* of information in an SPN. If the nodes were not identical, this would no longer be true, and by the *data processing inequality* [13], the capacity for the summed SPN nodes may only be smaller than that without summation. However, as discussed in [2], the decrease in capacity may be very small.

Since the capacity achieving distribution must be discrete, it is assumed in the following, except where otherwise stated, that x is a discrete random variable with M points of support and PMF $P_x(m)>0$, $m=1,\ldots,M$. The mass points are denoted as x_m, $m=1,\ldots,M$.

It is important to note that since the channel transition probabilities depend on the mass points via Eqs. (2) and (3), altering any x_m actually alters the channel. However, instead of finding optimal mass points, note that the channel capacity can be equivalently written as

$$C(x;y) = \max_{\{M,P_x(1),\ldots,P_x(M),P_{1|x_1},\ldots,P_{1|x_M}\}} I(x;y), \tag{6}$$

which means searching for M optimal values of $P_{1|x}$ instead of x.

For use in Sec. III, by [19] (Theorem 4.5.1) necessary and sufficient conditions for $P_x(\cdot)$ to achieve capacity for a given channel are that for all m with $P_x(m)>0$,

$$i(x_m) \coloneqq \sum_{n=0}^{N} P_{y|x}(n|x_m)\log_2\left(\frac{P_{y|x}(n|x_m)}{P_y(n)}\right) = C(x;y), \quad (7)$$

while $i(x_m) \leq C(x;y)\,\forall x$ s.t. $P_x(m)=0$. Consequently, a necessary condition for capacity for the SPN is that $C(x;y) = \log_2[P_y^o(0)] = \log_2[P_y^o(N)]$, where $P_y^o(n)$ is the output distribution induced by the capacity achieving input distribution. This follows by noting that nonzero mass should always occur at $P_{1|x}=0$ or $P_{1|x}=1$ (since the channel will be deterministic in these cases) and substituting these values into Eq. (3).

Corollary 1 provides a guide to an algorithm for numerically calculating the optimal discrete distribution for the SPN. Before outlining such an algorithm and results in Sec. III, some theoretical discussion and useful results are provided.

II. THEORETICAL RESULTS

A. Capacity for the binary-node SPN is independent of θ and the noise distribution

When there are no constraints on the signal distribution, channel capacity depends only on N, and is invariant to changes in the noise distribution or θ. This can be seen by noting from Eq. (6) that capacity can be found by optimizing M values of $P_{1|x}$. If the SPN is defined in terms of a known noise distribution with CDF $F_\eta(\eta)$, then the optimal mass points of the signal can be determined via inversion of Eq. (2) since from Eq. (2),

$$x_m = \theta - F_\eta^{-1}(1 - P_{1|x_m}), \quad \forall\, m = 1, \ldots, M, \quad (8)$$

where $F_\eta^{-1}(\cdot)$ is the inverse CDF of the noise.

Consequently, channel capacity is a property only of N, i.e., the size of the SPN.

B. Lower bounds on capacity

Three previous exact theoretical expressions for the mutual information in the binary-quantizing SPN can be used to verify the numerical results presented in Sec. III. This is because these expressions were derived assuming *continuously valued* signal distributions [with PDF $f_x(\cdot)$] and specified noise distributions [with PDF $f_\eta(\cdot)$], and therefore by Corollary 1 must provide lower bounds to the channel capacity.

First, when the signal is a continuously valued random variable with the same distribution as the noise, i.e., $f_x(x) = f_\eta(\theta - x)$, the mutual information is [29]

$$I_1 \coloneqq \log_2(N+1) - \frac{N}{2\ln 2} - \frac{1}{N+1}\sum_{n=2}^{N}(N+1-2n)\log_2 n. \quad (9)$$

Second, when

$$f_x(x) = \frac{f_\eta(\theta - x)}{\pi\sqrt{F_\eta(\theta-x)[1 - F_\eta(\theta - x)]}}, \quad (10)$$

the mutual information is [12]

$$I_2 \coloneqq -\sum_{n=0}^{N} P_y(n)\log_2\left[\frac{P_y(n)}{\binom{N}{n}}\right] + N\log_2\left(\frac{e}{4}\right), \quad (11)$$

where $P_y(n)$ is beta-binomially distributed,

$$P_y(n) = \binom{N}{n}\frac{\beta(n+0.5, N-n+0.5)}{\beta(0.5,0.5)}, \quad (12)$$

and $\beta(a,b)$ is a beta function [30]. It is shown numerically in [12] that $I_2 > I_1$.

A third case of an exact expression derived in the context of SSR is that of continuously uniform signal and noise with equal mean, but where the noise' support is smaller than the signal's support, and has ratio $\sigma \leq 1$ ([14], Eq. (7)). This is equivalent to an input signal distribution that is a mixture of discrete mass at the extremes of the noise' support, each with probability $0.5(1-\sigma)$, and uniform density between. From Ref. [16], Eqs. (4.59) and (4.60), there is an optimal value of σ,

$$\sigma^o = \frac{1}{\frac{N-1}{N+1} + 2^{1-F}},$$

where F is a function of N and I_1,

$$F = \frac{1}{N-1}[(N+1)I_1 - 2\log_2(N+1)].$$

The resultant mutual information can be expressed as

$$I_3 \coloneqq \log_2\left[2 + \left(\frac{N-1}{N+1}\right)2^F\right]. \quad (13)$$

In this case for any σ,

$$P_y(0) = P_y(N) = \frac{\sigma}{N+1} + 0.5(1-\sigma).$$

It is straightforward to show that when $\sigma = \sigma^o$, $I_3 = -\log_2[P_y(N)]$. Consequently, the necessary condition for capacity mentioned in Sec. I A is satisfied for $n=0$ and $n=N$. This does not mean I_3 gives capacity, as the necessary condition is not met for other values of n.

Using an approximation based on Fisher information [8], a fourth lower bound to the channel capacity can be derived as [12]

$$C(x;y) \geq \lim_{N\to\infty} I_2 = 0.5\log_2\left(\frac{N\pi}{2e}\right) \coloneqq I_L. \quad (14)$$

C. Upper bounds on capacity

The entropy of any discrete probability distribution is upper bounded by the log of its cardinality, while mutual information can be no larger than the lesser of its input or output entropy [13]. Hence, if the entropy of the capacity achieving input with M^o points is $H^o(x)$, then

$$C(x;y) \leq H^o(x) \leq \log_2(M^o) \leq \log_2(N+1). \quad (15)$$

Numerical determination of M^o and $C(x;y)$ as in Sec. III finds that these inequalities are far from tight, except that $H^o(x)$ is close to $\log_2(M^o)$. A much tighter (although analytically unproven) bound for $N>2$ that is in agreement with the numerics for at least $N=2,\ldots,100$ can be derived by taking the large N limit of I_3 [16],

$$C(x;y) < \lim_{N\to\infty} I_3 = \log_2\left(2 + \sqrt{\frac{(N+2)e}{2\pi}}\right) := I_U. \quad (16)$$

Consequently, the channel capacity for small N also scales with $O[\log_2(N)]$, as was shown either for *suboptimal* input distributions, and/or for large N in [8,11,12,14,29]. Notice also that

$$\lim_{N\to\infty} I_U = \lim_{N\to\infty} I_L = \log_2(\sqrt{N}),$$

and that therefore the upper and lower bounds coincide at $N=\infty$.

D. Lower bound on M^o

In Sec. III, it is shown numerically that I_3 is the lower bound that is closest to capacity for small N, and $I_3 > I_L$. Consequently,

$$I_L < I_3 \leq C(x;y) \leq \log_2(M^o).$$

Therefore, the number of points of mass of the capacity achieving input distribution must satisfy

$$N+1 \geq M^o \geq \lceil 2^{I_3} \rceil > \left\lceil \sqrt{\frac{N\pi}{2e}} \right\rceil. \quad (17)$$

This bound states that the number of points of mass in the input distribution must increase at a rate faster than $\sqrt{N}$ but slower than N in order to achieve capacity with a minimal number of input mass points.

III. NUMERICAL RESULTS

A. Procedure for finding capacity

When a communication channel is defined by conditional distributions that are independent of the values of M source symbols, finding channel capacity means finding only the optimal probabilities, $P_x(m)$. This can be achieved using a convergent iterative algorithm known as the Blahut-Arimoto (BA) algorithm [13].

However since the binary-node SPN depends on $P_{1|x_m}$, the BA algorithm cannot be used in isolation, except via an approximation where M is assumed to be very large, and the $P_{1|x_m}$ uniformly spaced. The BA algorithm will theoretically converge, possibly with the result that most $P_x(x_m)=0$. However convergence in such a situation is extremely slow, as there are likely to be many near optimal solutions for P_x on this assumed support.

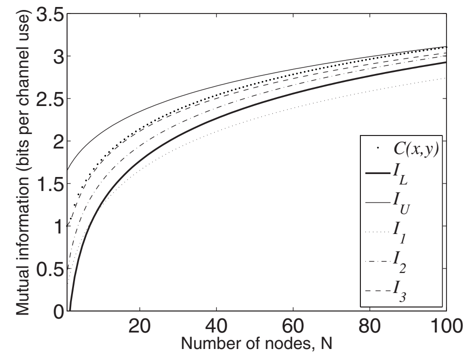

FIG. 1. Numerically calculated channel capacity, $C(x;y)$ as N increases. Also shown are the upper and lower bounds to capacity given by Eqs. (9), (11), (13), (14), and (16).

Nevertheless, this approach was successfully applied for binary-node SPNs [8], but only for large N. The resultant capacity achieving PMF was assumed to estimate a PDF, and is in agreement with an analytical formula derived using the large N assumption [12], in which it is also assumed that the input signal must have a continuous distribution. However, by Corollary 1 the optimal signal distributions of [8,12] can only be optimal in the sense that they are the continuous PDF to which a discrete PMF will converge, as $N\to\infty$.

The necessity to preselect values of $P_{1|x_m}$ in this way can be removed for small fixed M by combining the BA algorithm with a gradient descent algorithm. Local optima of $I(x;y)$ will be found by allowing $P_{1|x_m}$ to be free variables in the gradient descent algorithm, and the BA algorithm used to find the corresponding optimal $P_x(x_m)$ for each trial solution of $P_{1|x_m}$.

The value of M^o for some N cannot be known in advance, and needs to be determined. A way to achieve this follows from the previously mentioned necessary and sufficient condition for optimality of a discrete memoryless channel [19] (Theorem 4.5.1)—see Eq. (7)—and use of an algorithm given in [31], where it is noted that it suffices to determine the optimal $P_x(\cdot)$ for increasing values of M, starting with $M=2$, and stopping once the conditions are met. Furthermore, if a local optima satisfies [19] (Theorem 4.5.1), then it is actually a global optima.

An alternative approach is to first explicitly calculate the channel capacity and the optimal *output distribution*, $P_y^o(n)$ by applying standard minimax optimization algorithms to Eq. (7), as outlined in [32]. The optimal input distribution can then be found by numerically finding the maxima of $i(x_m)$ that results from $P_y^o(n)$.

B. Unconstrained capacity results

The results of applying the above method to find channel capacity for the binary-node SPN, for N between 1 and 100, are shown as $C(x;y)$ in Fig. 1. Also shown are the lower bounds I_1, I_2, I_3, and I_L, evaluated exactly from Eqs. (9), (11), (13), and (14), and the conjectured upper bound, I_U

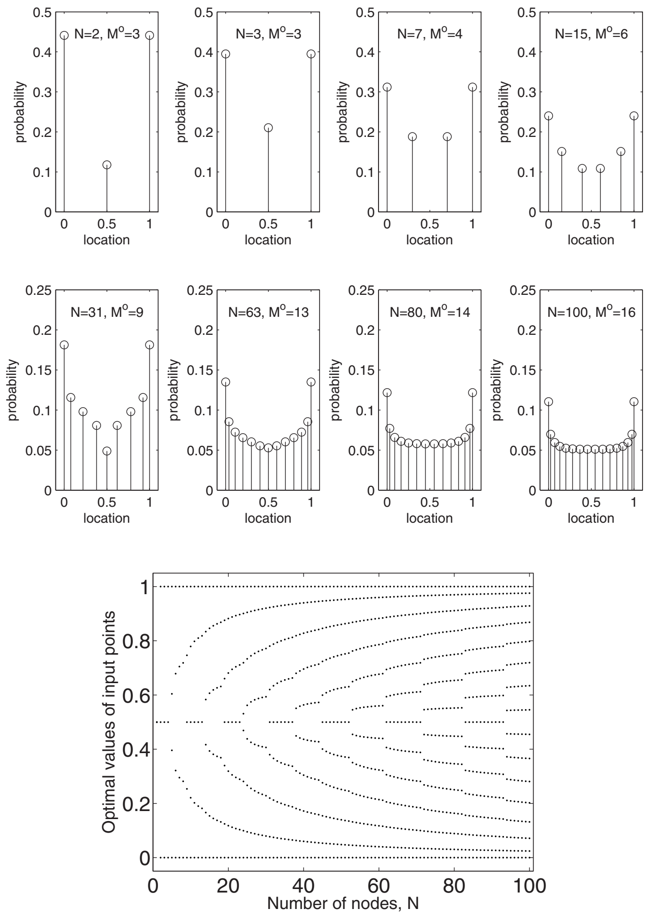

FIG. 2. Capacity achieving input distributions as the network size, N, varies. The upper panel shows the optimal input probability mass function for selected N, where M^o denotes the optimal number of points for each N. The ordinate is the optimal probability mass of each point, $P_x(m)$, for $m=1,\ldots,M^o$ and the abscissa the optimal locations, $P_{1|x_m}$. Note that the optimal $P_{1|x_m}$ can be converted into the optimal points of support for $P_x(m)$, i.e., x_m, for any absolutely continuous noise distribution, using Eq. (8). The lower panel shows the locations of optimal mass points for $N=1,\ldots,100$.

from Eq. (16). It is clear that all the stated inequalities are verified.

The input distribution that achieved $C(x;y)$ is shown in Fig. 2. The upper panel shows the optimal $P_x(m)$ and the corresponding points, $P_{1|x_m}$ for selected values of N, while the lower panel shows only the optimal $P_{1|x_m}$ for $N=1,\ldots,100$.

The output distribution at capacity is shown for selected values of N in Fig. 3, while Fig. 4 shows the optimal number of points, M_o and also the lower bounds of Eq. (17).

C. Varying the noise level: Suprathreshold stochastic resonance

As discussed in Sec. II A, capacity for the binary-node SPN is independent of the noise distribution. However, this does not alter the fact that capacity is achieved by the presence of noise, and therefore that SSR occurs. This can be demonstrated by assuming that the capacity achieving input distribution is used and remains unaltered, and then allowing

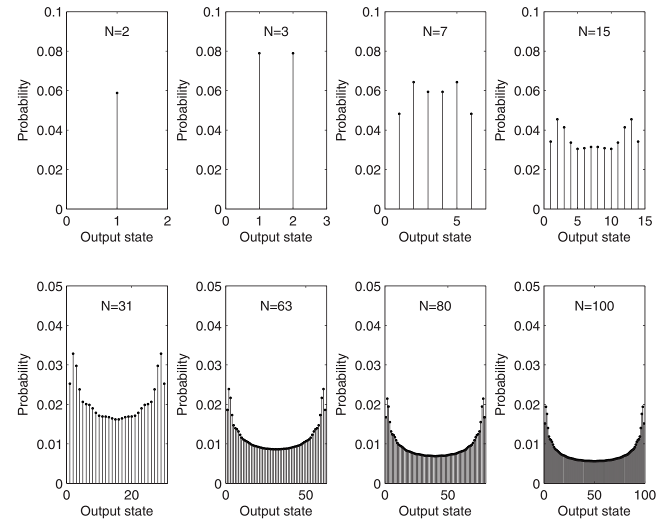

FIG. 3. Optimal output distribution (i.e., that induced by the input distribution that achieves capacity), $P_y^o(n)$, $n=1,\ldots,N-1$ for selected values of the network size, N. $P_y^o(0)$ and $P_y^o(N)$ are not shown, in order to aid visibility, however, as noted in Sec. I A, at capacity it is necessary that $P_y^o(0)=P_y^o(N)=2^{-C(x;y)}$.

the noise intensity to change, as is now described.

Suppose the optimal solution is given by $\{M^o, P_x^o(1),\ldots,P_x^o(M), P_{1|x_1}^o,\ldots,P_{1|x_M}^o\}$. From Eq. (8), for any given continuously valued noise distribution with inverse CDF $F_\eta^{-1}(\eta)$, the optimal input mass points are

$$x_m^o = \theta - F_\eta^{-1}(1-P_{1|x_m}^o), \qquad \forall\, m=1,\ldots,M. \qquad (18)$$

Clearly the actual mass points of the optimal signal distribution (and consequently its statistics, e.g., mean and variance) depend on the noise distribution, even though the capacity does not.

As an example, suppose for a given N, the input distribution is optimal for zero mean, unity variance Gaussian noise, and the threshold is $\theta=0$. This means that

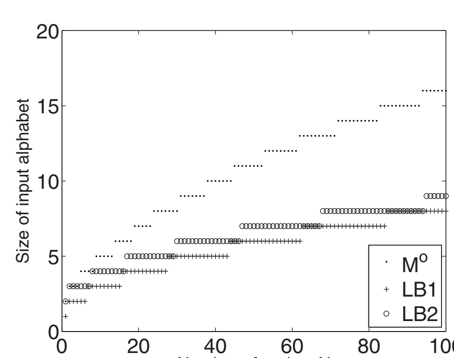

FIG. 4. Optimal number of points in the capacity achieving input distribution, M^o, and the two lower bounds to M^o given by Eq. (17), i.e., $LB_1=\lceil 2^{I_3}\rceil$ and $LB_2=\lceil\sqrt{\frac{N\pi}{2e}}\rceil$. As N increases, the lower bounds become less accurate since the input entropy becomes smaller than $\log_2(M^o)$.

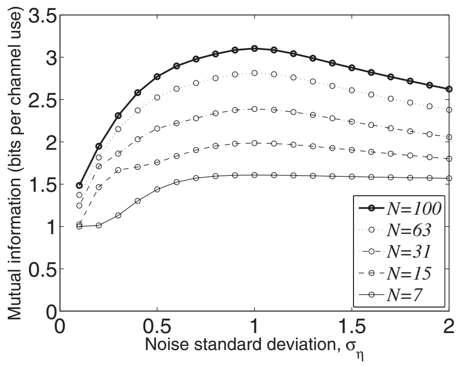

FIG. 5. Mutual information when the noise distribution is allowed to vary from that used to calculate the capacity achieving signal mass points for various N. The data is for an example of Gaussian noise with zero mean and standard deviation σ_η, when the mass points are capacity achieving for $\sigma_\eta=1$. SSR can be seen to occur since noise with larger and smaller standard deviations than optimal reduce the mutual information from channel capacity.

$$x_m^o = -\sqrt{2}\text{erf}^{-1}(1-2P^o_{1|x_m}), \quad \forall\, m=1,\ldots,M, \tag{19}$$

where $\text{erf}^{-1}(\cdot)$ is the inverse error function. If the Gaussian noise is now allowed to have a variance of σ_η^2, this has the affect of changing the $P_{1|x}$ from the optimal values to

$$P_{1|x_m} = 0.5 + 0.5\text{erf}\left(\frac{x_m^o}{\sqrt{2}\sigma_\eta}\right). \tag{20}$$

Clearly $P_{1|x_m}=P^o_{1|x_m}$ only if $\sigma_\eta=1$. The resultant combination of the optimal signal distribution and suboptimal noise distribution are no longer capacity achieving.

The reduction in mutual information can be calculated numerically, as is shown in Fig. 5 for N=7, 15, 31, 63, and 100 for a range of Gaussian noise variances.

D. Equivalent optimal discrete noise distribution

A corollary of the observations of Sec. III C is that given complete control over the noise distribution, capacity can be achieved for any specified signal distribution by optimizing the noise. Since calculation of mutual information depends on $P_{1|x_m}$ rather than x_m, then *given* a signal distribution with the mass points of Eq. (18), there exists an *optimal noise distribution*.

While noise with CDF $F_\eta(\eta)$ would achieve capacity in this case, this distribution is not unique when x is discrete since any noise distribution with CDF $F_\zeta(\cdot)$ such that $P^o_{1|x_m}=1-F_\zeta(\theta-x_m^o)\,\forall m$ will also be optimal. This includes the possibility of a discrete noise distribution, which can be understood with the following example.

Suppose the noise distribution is discrete on [0,1]. Let its mass points be the values

$$\zeta_m = 0.5(P^o_{1|x_{m-1}} + P^o_{1|x_m}), \quad m=2,\ldots,M$$

and $\zeta_1=P^o_{1|x_1}$. Let its probability mass values be

$$P_\zeta(m) = P^o_{1|x_m} - \sum_{i=1}^{m-1} P^o_{1|x_i}, \quad m=1,\ldots,M.$$

This gives a CDF that has points of increase at ζ_m, i.e.,

$$\lim_{\delta\to 0} F_\zeta(\zeta_m+\delta) = P^o_{1|x_m}.$$

Consequently

$$F_\zeta(x_m) = P^o_{1|x_m}.$$

IV. DISCUSSION AND CONCLUSIONS

A. Interpretation of the numerical results

It has been shown that the channel capacity for small N is significantly larger than that achieved with continuously valued signal distributions, e.g., I_1 and I_2. This is due to such cases not exploiting the fact that the channel is deterministic when $P_{1|x}$ is zero or unity. As shown in Fig. 2, it is always optimal to place the largest P_x at these points. This is also the reason why I_3 is very close to capacity since it includes probability mass at these points.

The optimal input distribution can be seen to be close to uniform other than the points at $M=0$ and $m=M^o$, which is likely the reason why the lower bound I_3 is close to capacity. The output distribution can be seen to be multimodal, with the smallest probabilities close to $N/2$ and at $n=1$ and $n=N-1$. Intuitively this can be expected to be the case since (i) $E[y|P_{1|x}=0.5]=N/2$ and the conditional variance of y is largest when $P_{1|x}=0.5$ since $\text{var}[y|x]=NP_{1|x}(1-P_{1|x})$; (ii) the smallest conditional variance of y occurs when $P_{1|x}=0$ and $P_{1|x}=1$, and consequently $P_y(0)$ and $P_y(N)$ should be large, while other values of y near 0 and N should have small probability.

The lower panel of Fig. 2 shows what appear to be discontinuous bifurcations as N increases. This behavior is qualitatively similar to related but different optimizations of the binary-quantizing SPN, where instead of optimizing the input distribution, the optimization is of variable threshold values [33].

Here this result is more easily understood, in that bifurcations occur when it is necessary for M^o to increase to achieve capacity, similar to the results of [31]. For example, when N=1, one bit per channel use can be achieved by M^o=2, while for N=2, more than one bit per channel use can be achieved with M^o=3, which requires an additional mass point to appear at $P_{1|x}=0.5$.

The increasing gap between M^o and the lower bounds shown in Fig. 4 is due to the fact that the entropy of x is significantly larger than the mutual information as N increases. This is because the average conditional entropy of x given y increases with N, as it becomes harder to discriminate between points in the input distribution as the number of points grows.

B. Implications for ASR and SSR

For the demonstration of SSR in Fig. 5, the value at $\sigma_\eta=1$ is the channel capacity shown in Fig. 1. In this context,

$\sigma_\eta=1$ is the optimal noise intensity and leads to the SSR peak in mutual information. For values of σ_η larger and smaller than optimal, the mutual information is reduced from capacity, in a qualitatively identical manner to all previous work on SSR.

Although the greater focus here is on SSR, as discussed in Sec. I trivial cases of the SPN model are equivalent to the single threshold model previously used to demonstrate subthreshold ASR using mutual information. The current focus on optimizing the signal distribution, rather than on finding the optimal noise intensity for assumed signal and noise distributions, makes it clear that it is the thresholding operation, and the number of measurements, N, that limits the information capacity. The capacity is also independent of the noise distribution. The fact that SSR occurs is due to the facts that (i) $N>1$ and (ii) all nodes in the SPN are identical; while subthreshold ASR occurs in the $N=1$ case only because of the subthreshold constraint.

C. Implications for information transfer in biological neurons

One of the reasons for studying the binary-quantizing SPN is its similarity to populations of parallel sensory neurons, and the fact that its behavior, e.g., SSR, has been shown to be qualitatively the same in more complex neural population models [6,8,9]. As mentioned in Sec. I, SPNs exhibit information pooling. The consequences of such an effect has been of recent interest in neuroscience [7], where it has also been called *aggregation* [10].

Provided biologically relevant models of pooling by networks of neurons can be mapped to an SPN framework where the output signal is discrete with a finite number of states, Corollary 1 states that the information capacity of that network when the input signal is memoryless will be achieved by a discrete input distribution.

Similar results on optimal discreteness was suggested for a Poisson neuron model and an assumption of rate coding in [34]. One difference between the Poisson case and the SPN studied here is that while the output distribution of a Poisson neuron is discrete; it is also infinite in cardinality. This means that Corollary 1 does not apply, and a proof that the optimal input is discrete has only been provided recently, as part of a more general formulation [35] (although similar results outside the neural context for Poisson channels have long been known [10,36,37]).

One implication of the results for the Poisson channel that has been overlooked is that the discreteness of the capacity achieving input distribution can be converted to mean that the optimal neural *tuning curve* is also discrete [38]. Section III C contained an analogous result for the binary-quantizing SPN, except that instead of a discrete tuning curve, it is the noise distribution that should be discrete to optimize information transfer for an arbitrary continuous signal distribution.

The case where the input signal is continuous, and information capacity is achieved by ensuring the tuning curve or noise distribution is optimally discrete can be viewed as achieving optimal *source coding*, i.e., the maximally informative representation of a signal obtained through noisy and quantized observations. On the other hand, when the noise or tuning curve is fixed, ensuring information capacity via a discrete input signal distribution can be viewed as optimal *symbol coding* for transmission in a noisy channel.

This "optimal discrete noise" result means that coding of a continuous random variable for information transmission through an SPN is achieved when the noise is added in discrete packets or quanta. One biological scenario where it may be worth exploring whether this actually occurs is the synaptic junction between sensory receptors cells and afferent nerve fibers, e.g., in the auditory system. While the potential induced in the sensory cell by an external stimulus is effectively continuously valued, transmission of that potential to each of a number of nerve fibers involves (i) a discretized representation prior to the nerve fiber since synaptic transmission is both *quantal* and random [39] and (ii) discrete outputs in the form of action potentials.

ACKNOWLEDGMENTS

Mark D. McDonnell is funded by the Australian Research Council, Grant No. DP0770747. The author would also like to thank Nigel G. Stocks, Shiro Ikeda, Alex G. Grant, and Terrence H. Chan for valuable discussions.

[1] S. Zozor, P. Amblard, and C. Duchêne, Fluct. Noise Lett. **7**, L39 (2007).
[2] M. D. McDonnell, N. G. Stocks, and P. O. Amblard, J. Stat. Mech.: Theory Exp. (2009), P01012.
[3] M. D. McDonnell, D. Abbott, and C. E. M. Pearce, Microelectron. J. **33**, 1079 (2002).
[4] T. Nguyen, IEEE Trans. Signal Process. **55**, 2735 (2007).
[5] F. Martorell and A. Rubio, Microelectron. J. **39**, 1041 (2008).
[6] N. G. Stocks and R. Mannella, Phys. Rev. E **64**, 030902(R) (2001).
[7] S. Panzeri, F. Petroni, R. S. Petersen, and M. E. Diamond, Cereb. Cortex **13**, 45 (2003).
[8] T. Hoch, G. Wenning, and K. Obermayer, Phys. Rev. E **68**, 011911 (2003).
[9] A. Nikitin, N. G. Stocks, and R. P. Morse, Phys. Rev. E **75**, 021121 (2007).
[10] D. H. Johnson and I. N. Goodman, Network Comput. Neural Syst. **19**, 13 (2008).
[11] N. G. Stocks, Phys. Rev. Lett. **84**, 2310 (2000).
[12] M. D. McDonnell, N. G. Stocks, and D. Abbott, Phys. Rev. E **75**, 061105 (2007).
[13] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. (Wiley, New York, 2006).
[14] N. G. Stocks, Phys. Lett. A **279**, 308 (2001).
[15] D. Rousseau, F. Duan, and F. Chapeau-Blondeau, Phys. Rev. E **68**, 031107 (2003).

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

[17] D. Rousseau and F. Chapeau-Blondeau, Signal Process. **85**, 571 (2005).

[18] D. Rousseau, J. Rojas Varela, and F. Chapeau-Blondeau, Phys. Rev. E **67**, 021102 (2003).

[19] R. G. Gallager, *Information Theory and Reliable Communication*, 2nd ed. (Wiley, New York, 1968).

[20] L. Gammaitoni, P. Hänggi, P. Jung, and F. Marchesoni, Rev. Mod. Phys. **70**, 223 (1998).

[21] J. J. Collins, C. C. Chow, A. C. Capela, and T. T. Imhoff, Phys. Rev. E **54**, 5575 (1996).

[22] C. Heneghan, C. C. Chow, J. J. Collins, T. T. Imhoff, S. B. Lowen, and M. C. Teich, Phys. Rev. E **54**, R2228 (1996).

[23] A. R. Bulsara and A. Zador, Phys. Rev. E **54**, R2185 (1996).

[24] H. Chen, P. K. Varshney, S. M. Kay, and J. H. Michels, IEEE Trans. Signal Process. **55**, 3172 (2007).

[25] B. Kosko and S. Mitaim, Phys. Rev. E **70**, 031911 (2004).

[26] A. Patel and B. Kosko, Neural Networks **18**, 467 (2005).

[27] B. Kosko and S. Mitaim, Phys. Rev. E **64**, 051110 (2001).

[28] P. O. Amblard, O. J. J. Michel, and S. Morfu, in Proceedings of SPIE: Noise in Complex Systems and Stochastic Dynamics III, edited by L. B. Kish, K. Lindenberg, and Z. Gingl (2005), Vol. 5845, pp. 50–60.

[29] N. G. Stocks, Phys. Rev. E **63**, 041114 (2001).

[30] M. R. Spiegel and J. Liu, *Mathematical Handbook of Formulas and Tables* (McGraw-Hill, New York, 1999).

[31] T. H. Chan, S. Hranilovic, and F. R. Kschischang, IEEE Trans. Inf. Theory **51**, 2073 (2005).

[32] M. Chiang and S. Boyd, IEEE Trans. Inf. Theory **50**, 245 (2004).

[33] M. D. McDonnell, N. G. Stocks, C. E. M. Pearce, and D. Abbott, Phys. Lett. A **352**, 183 (2006).

[34] R. B. Stein, Biophys. J. **7**, 797 (1967).

[35] S. Ikeda and J. H. Manton, Neural Computation (to be published).

[36] Y. M. Kabanov, Theor. Probab. Appl. **23**, 143 (1978).

[37] S. Shamai, IEE Proc., Part I (Communications, Speech, and Vision) **137**, 424 (1990).

[38] A. P. Nikitin, N. G. Stocks, R. P. Morse, and M. D. McDonnell, e-print arXiv:0809.1549.

[39] C. Koch, *Biophysics of Computation: Information Processing in Single Neurons* (Oxford University Press, New York, 1999).