The data processing inequality and stochastic resonance

Mark D. McDonnell[a], Nigel G. Stocks[b], Charles E. M. Pearce[c], Derek Abbott[a]
[a]Centre for Biomedical Engineering (CBME) and School of Electrical & Electronic Engineering, The University of Adelaide, SA 5005, Australia;
[b]School of Engineering, The University of Warwick, Coventry CV4 7AL, UK;
[c]School of Applied Mathematics, The University of Adelaide, SA 5005, Australia

ABSTRACT

The data processing inequality of information theory states that given random variables X, Y and Z that form a Markov chain in the order $X \longrightarrow Y \longrightarrow Z$, then the mutual information between X and Y is greater than or equal to the mutual information between X and Z. That is $I(X;Y) \geq I(X;Z)$. In practice, this means that no more information can be obtained out of a set of data then was there to begin with, or in other words, there is a bound on how much can be accomplished with signal processing. However, in the field of stochastic resonance, it has been reported that a signal to noise ratio gain can occur in some nonlinear systems due to the addition of noise. Such an observation appears to contradict the data processing inequality. In this paper, we investigate this question by using an example model system.

Keywords: stochastic resonance, information theory, data processing inequality, SNR gain, channel capacity

1. INTRODUCTION

The Data Processing Inequality (DPI) states that given random variables X, Y and Z that form a Markov chain in the order $X \longrightarrow Y \longrightarrow Z$, then the mutual information between X and Y is greater than or equal to the mutual information between X and Z.[1] That is

$$I(X;Y) \geq I(X;Z).$$

In practice, this means that you cannot get more information out of a set of data then was there to begin with, or in other words, no signal processing on Y can increase the information that Y contains about X.

However, it has been noted in some papers that it is possible to get a signal to noise ratio (SNR) gain in some nonlinear systems by the addition of noise.[2–12] Such an occurrence appears on the surface to contradict the DPI. We also note that general investigations of such SNR gains due to noise have previously been published,[13] which criticize the use of SNR in such systems and indicate more appropriate measures to use, at least for signal detection or estimation problems.[14, 15] In this paper we discuss the literature relating to this issue and investigate the apparent contradiction by using an example model system.

It should be noted that the terminology *Markov chain* used in connection with the DPI is somewhat more inclusive than that prevalent in applied probability. DPI usage requires only the basic Markov property that Z and X are conditionally independent given Y. By contrast the usage in applied probability requires also that X, Y, Z range over the same set of values and that the distribution of Z given Y be the same as that of Y given X.

Further author information: (Send correspondence to Mark D McDonnell.)
Mark D. McDonnell.: E-mail: mmcdonne@eleceng.adelaide.edu.au, Telephone: +61 8 83036296 Fax: +61 8 8303 4360
Nigel G. Stocks.: E-mail: es2003@eng.warwick.ac.uk
Charles E. M. Pearce.: E-mail: cpearce@maths.adelaide.edu.au
Derek Abbott.: E-mail: dabbott@eleceng.adelaide.edu.au, Telephone: +61 8 83035748

2. HISTORY OF SNR GAINS IN STOCHASTIC RESONANCE

A SNR gain is not in itself a remarkable thing; SNR gains are routinely obtained by filtering. Inchiosa *et al* give an excellent description of the issue: "... *a nonlinear signal processor may output a signal that has infinite SNR but is useless because it has no correlation with the input signal. Such a system would be one which simply generates a sine wave at the signal frequency, total ignoring its input.*"[14] They point out the example of a bandpass filter, for which the output SNR may exceed the input SNR.

However in the stochastic resonance literature (stochastic resonance is the name given to a broad range of phenomena where an optimal nonzero value of noise intensity at the input can induce an optimal output from a nonlinear system[16]), the reported SNR gains are due to the *addition of noise* to an already noisy signal, rather than a deliberately designed filter, which is why the SNR gains are taken to be quite remarkable. In this section we briefly summarize some of the literature describing the possibility of SNR gains due to stochastic resonance.

Stochastic resonance was at first thought to occur only in bistable dynamical systems, generally driven by a periodic input signal, $A\sin(\omega_0 t + \phi)$ and broadband noise. Since the input to such systems was a simple sinusoid, the SNR at the input was a natural measure to use, with the following definition most commonly used,

$$\mathrm{SNR} = \frac{P(\omega_0)}{S_N(\omega_0)},$$

where $P(\omega_0)$ is the input signal power and $S_N(\omega_0)$ is the power spectral density of the noise at frequency ω_0. The work on such systems showed that the ratio of the output power at frequency ω_0 to the background noise spectral density at ω_0 was maximized by a nonzero value of noise intensity. This was the definition of stochastic resonance.

We note in passing, that it is well known in electronic engineering that nonlinear devices cause output frequency distortion, that is, for a single frequency input, the output will consist of various harmonics of the input.[17] Hence, basic circuit design requires the use of filters that remove unwanted output frequencies. For example, this *harmonic distortion* in audio amplifiers is very undesirable. However on the other hand high frequency oscillators make use of this effect, by starting with a very stable low frequency oscillator, and putting the generated signal through a chain of frequency multipliers. The final frequency is harmonically related to the low frequency source.

For more than one input frequency, the output will contain the input frequencies plus integer multiples of the sum and difference between all frequencies.[17] This effect of creating new frequencies is known as *intermodulation distortion*. In the field of optics, this phenomenon can be used to generate lower frequency signals, (for example, teraHertz rays) from different optical frequencies by a method known as *optical rectification*.[18,19]

We note that a study of such higher harmonics generated by a nonlinear system exhibiting stochastic resonance has been published,[20] and the phenomenon also discussed in subsequent work.[21,22] However, in general we are only interested in the output frequency component that corresponds to the input periodic driving signal. Hence, the output SNR ignores all other output harmonics. Therefore, if we define the signal to noise ratio gain in a nonlinear system as $G = \mathrm{SNR}_o/\mathrm{SNR}_i$ then

$$G = \frac{P_o(\omega_0)}{P_i(\omega_0)}\frac{S_{N,i}(\omega_0)}{S_{N,o}(\omega_0)}.$$

It can be shown theoretically[23,24] for the case of stationary Gaussian noise, and a signal that is small compared to the noise, that for nonlinear systems G must be less than or equal to unity, and that hence no SNR gain can be induced by utilizing the stochastic resonance effect. This proof is based on the use of linear response theory, where, since the signal is small compared to the noise, both the signal and noise are transferred linearly to the output, and as in a linear system, no SNR gain is possible. Much attention was given to this fact, since most studies on stochastic resonance remained in the linear response limit to ensure that the output was not subject to the above mentioned harmonic distortion.[10]

Once this fact was established, researchers still hoping to be able to find systems in which SNR gains due to noise could occur, turned their attention to situations not covered by the proof, that is, the case of a signal that is not small compared to the noise, or broadband signals or non-Gaussian noise. For example, work by Kiss[2] used a broadband input signal and hence required a different SNR definition to the conventional one, and Chapeau-Blondeau *et al*[5,6] used the conventional SNR definition, but the large signal regime to show the existence of SNR gains. Furthermore, Chapeau-Blondeau has also considered the case of non-Gaussian noise.[7]

When the input signal is broadband (or aperiodic), use of the ratio of the output fundamental frequency spike to the background noise is no longer relevant. Consider how much input signal power can be discarded in the case of an input square wave. This signal is periodic, but broadband, since all frequencies are present. In this case, if the output signal is taken to only be the first harmonic, the only information known at the output is the period of the input and the SNR at that frequency, but nothing is known about the shape of the input. Hence, when authors first turned their attention to the concept of stochastic resonance and aperiodic input signals, they used other methods (such as cross-correlation[25] and cross-spectra[2]), which do indicate how well the shape of the output signal corresponds to the shape of the input. Mutual information has also been used for aperiodic input signals, and this measure indirectly describes the shape. DeWeese *et al* address the benefits of using the mutual information measure.[23]

In order to compare the detectability at the output to the input for broadband input signals, a measure analogous to the input-output SNR ratio for periodic input signals is required. As mentioned, Kiss[2] defined an SNR measure (based on cross-spectra) as a function of frequency to obtain an input-output SNR measure. An alternative measure used by Chapeau-Blondeau[7] was channel capacity. Here, Chapeau-Blondeau stated that comparing the channel capacity at the input and output to a system for an aperiodic input signal was analogous to a comparison of the input and output SNRs for periodic input signals. In this paper we concentrate on giving an investigation into the use of channel capacity (maximum possible mutual information[1]) in simple threshold based systems where stochastic resonance can occur.

3. THE MODEL

Consider a system where a signal, s, is subject to independent additive random noise, n, to form another random signal, $y = s + n$. The signal y is then subjected to a nonlinear function, g to give a final random signal, $z = g(y)$. A block diagram of such a system is shown in Figure 1.

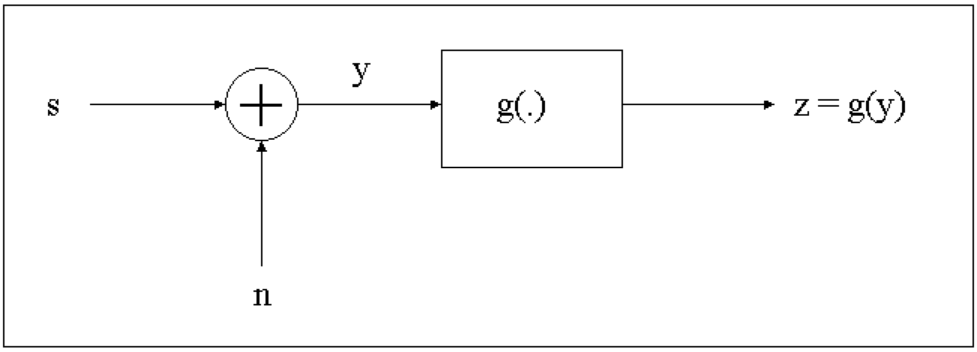

Figure 1. System of input signal, s, noise, n, output, z and transfer function, g.

Such a system is similar to those reported to show SNR gains for aperiodic input signals, in which the input is a random pulse train. As mentioned, Chapeau-Blondeau calculated the maximum mutual information between the output and input (C_{out}), and appeared to calculate the maximum mutual information between the signal by itself and the signal plus noise (C_{in}), and showed that it was possible for $C_{\text{out}} > C_{\text{in}}$.[7] This appears on the surface to be a clear contradiction of the DPI. We investigate this contradiction in the following sections.

In this paper we consider only random input signals. We note that in the case of a random pulse train consisting of only two levels, such as 0 and 1 or -1 and $+1$ that the model is equivalent to the binary memoryless channel often considered in information theory textbooks. The channel consists of the noise and the threshold, and the addition of noise to the signal causes the input signal to be transmitted through the channel with errors. For the particular case of a binary *symmetric* memoryless channel, the analysis is performed using the probability p that the input is inverted, so that $p(z = 1|s = 0) = p(z = 0|s = 1) = p$. If we had $p(z = 1|s = 0) \neq p(z = 0|s = 1)$ then the channel would not be symmetric.

3.1. Mutual information

The mutual information (also known as the transmitted information, transinformation, or Average Mutual Information (AMI)) between two random variables X and Y is defined as

$$I(X,Y) = H(Y) - H(Y|X),$$

where H is the entropy of the random quantity. For a discrete random variable, X_d,with n possible states that occur with probabilities $p_i, i = 1, .., n$ where $0 \leq p_i \leq 1$ and $\sum_{i=1}^{n} p_i = 1$, the entropy of X_d is defined as

$$H(X_d) = -\sum_{i=1}^{n} p_i \log_2 p_i.$$

For a continuous random variable, X_c, with probability density function $p_{X_c}(x)$ where, $q \leq x \leq r$, the entropy of X_c is defined as

$$H(X_c) = -\int_q^r p_{X_c}(x) \log_2 p_{X_c}(x) dx.$$

We also have the conditional entropy of two discrete random variables, X (n states) and Y (m states) given by

$$H(Y|X) = -\sum_{i=1}^{n}\sum_{j=1}^{m} P(X_i, Y_j) \log_2 P(Y_j|X_i),$$

where $P(X_i, Y_j)$ is the joint probability function of X and Y and $P(Y_j|X_i)$ is the conditional probability that Y is Y_j given X is X_i. If Y is a continuous random variable with probability density $p_Y(y)$ and X (n states) is a discrete random variable, the conditional entropy of Y given X is

$$H(Y|X) = \sum_{i=1}^{n} \int_y P(X_i, Y) \log_2 P(Y|X_i) dy,$$

where $p(X_i, Y)$ is the joint probability density of X_i and Y and $P(Y|X_i)$ is the conditional probability density of Y given X_i.

3.2. Channel capacity

The channel capacity of a discrete memoryless channel is defined as

$$C = \max_{p(x)} I(X,Y)$$

where the maximum is taken over all possible input distributions, $p(x)$. Therefore, for a binary channel, the channel capacity is

$$C = \max_{(p(0),p(1))} I(X,Y),$$

where $p(1) = 1 - p(0)$.

3.3. Is the model a Markov chain?

The DPI states that $I(s,y) \geq I(s,z)$ provided s, y and z form a Markov chain $s \longrightarrow y \longrightarrow z$. This means that the conditional distribution of z depends only on y and is conditionally independent of s.[1] For our model, since we have $z = g(y)$, z is a function of y and thus z is conditionally independent of s. Hence our model forms a Markov chain $s \longrightarrow y \longrightarrow z$.

4. EXAMPLE 1: ASYMMETRIC BINARY CHANNEL

Let $g(y)$ be given by

$$z = g(y) = \begin{cases} 1 & \text{if} \quad y = s + n \geq \theta, \\ 0 & \text{otherwise.} \end{cases}$$

Let the input signal, s, be a random binary signal such that both values of s are equally likely. Hence, s is a discrete random variable with probability distribution $P(s) = 0.5$, $s \in \{0, 1\}$. We assume that the noise is continuously valued.

4.1. Noiseless case

In the noiseless case (that is, $P(n = 0) = 1$), if $0 < \theta \leq 1$, then the output z will be identical to the input and $P(z = s) = 1$, and we have a symmetric binary channel with $p = 0$. But if $\theta > 1$ then the output will be always zero (that is, $P(z = 0) = 1$). Hence, for a threshold greater than the maximum input signal (or less than the minimum input signal) then all information is lost, since it is impossible to correlate the output with the input.

4.2. Finite, nonzero noise and subthreshold signal

Let the threshold, $\theta = t > 1$, so that the maximum value of the signal is below the threshold by $t - 1$. Let the noise be uniformly distributed between 0 and b. Hence, the density function of the noise is

$$p_n(n) = \begin{cases} \frac{1}{b} & 0 \leq n \leq b, \\ 0 & \text{otherwise.} \end{cases}$$

In this section we only consider the case of $t - 1 \leq b < t$, so that the output, z, will sometimes be equal to 1 when $s = 1$, but never equal to 1 when $s = 0$. If $b < t - 1$, then the output, z will always be zero, since the threshold is never crossed. If $b \geq t$ then sometimes the threshold will be crossed when $s = 0$. Thus we have an asymmetric binary channel. Hence, by the addition of noise to the input signal, the output becomes correlated with the the input signal (some information is conveyed; the output is only ever 1 when the input is 1) – an improvement over the case of zero noise, where the output conveys no information about the input.

Such a channel is a specific case of those already analyzed by Chapeau-Blondeau[26] in the stochastic resonance context, who calculated the mutual information for the general case of an arbitrary noise distribution and arbitrary $P(s)$. He then went on to calculate the channel capacity and showed a noise induced maximum. Here we consider the specific case simply as a means of illustrating the validity of the DPI.

4.2.1. Joint and conditional probability functions of s and z

We can express the above discussion in terms of the conditional probability function of z given s as

$$p(z|s) = (1 - 2z)s\left(\frac{t-1}{b} - 1\right) + (1 - z).$$

We can also state the probability distribution of z by observing that $z = 0$ when $s = 0$ *or* when $s = 1$ *and* $n < t - 1$, and that $z = 1$ when $s = 1$ *and* $n \geq t - 1$. Thus

$$P(z) = 0.5 + 0.5(1 - 2z)\left(\frac{t-1}{b}\right).$$

Now the joint distribution of z and s is $p(s, z) = p(s)p(z|s)$. Therefore we can write

$$p(s, z) = 0.5s\left((1 - 2z)\left(\frac{t-1}{b} - 1\right)\right) + 0.5(1 - z).$$

From this joint distribution, it is possible to calculate the probability of an error at the output as $P_e = p(s = 0, z = 1) + p(s = 1, z = 0) = \frac{t-1}{2b}$.

4.2.2. Joint and conditional probability functions of s and y

When $b < 1$, y can never take on values between b and 1. When $b \geq 1$, y can take on any value between 0 and $1 + b$. We assume for this section that $b < 1$. Note that we have defined b such that $t - 1 \leq b < t$, and that $t > 1$. Hence $b > 0$ always, and when $b < 1$ we must have $t < 2$. Therefore $0 < t - 1 < 1$.

Since $y = s + n$, then $p(y|s) = p(s + n|s) = p_n(y - s)$. Hence

$$p(y|s) = \begin{cases} \frac{1}{b} & s \leq y \leq b + s, \\ 0 & \text{otherwise.} \end{cases}$$

The joint probability density of s and y is given by $p(s, y) = p(s)p(y|s)$. Hence

$$p(s, y) = \begin{cases} \frac{1}{2b} & s \leq y \leq b + s, \\ 0 & \text{otherwise.} \end{cases}$$

The probability density function of y is given by

$$p_y(y) = \sum_s p(y|s)p(s) = \begin{cases} \frac{1}{2b} & 0 \leq y \leq b \text{ and } 1 \leq y \leq b + 1, \\ 0 & \text{otherwise.} \end{cases}$$

4.2.3. Mutual information between s and y

From the previous sections we can write an expression for $I(s, y)$. Recall that we assume $b < 1$. The entropy of y is then

$$H(y) = -\int p_y(y) \log_2 p_y(y) dy = -\int_0^b \frac{1}{2b} \log_2 \frac{1}{2b} dy - \int_1^{b+1} \frac{1}{2b} \log_2 \frac{1}{2b} dy = 1 + \log_2 b.$$

This makes sense, as the entropy of a uniform distribution of width c is $\log_2 c$ and the entropy of s is 1. The entropy of y given s is

$$H(y|s) = -\sum_s \int P(s, y) \log_2 P(y|s) dy = -\sum_s \int_s^{b+s} \frac{1}{2b} \log_2 \frac{1}{b} dy$$
$$= -\int_0^b \frac{1}{2b} \log_2 \frac{1}{b} dy + \int_1^{b+1} \frac{1}{2b} \log_2 \frac{1}{b} dy = \log_2 b.$$

This indicates that $H(y|s)$ is simply the entropy of the noise, which should be expected, since s is known, and both values of s are equally probable. Thus, the mutual information between s and y is $I(s, y) = H(y) - H(y|s) = 1$ bit per sample. This is consistent with the fact that the noise is small enough that $b < 1$ and therefore the input signal can be determined from the noisy signal without error, if a threshold were placed between b and 1.

4.2.4. Mutual information between s and z

The entropy of z is

$$H(z) = -\sum_z p_z(z) \log_2 p_z(z) = -\sum_z \left(0.5 + 0.5(1 - 2z) \left(\frac{t-1}{b}\right)\right) \log_2 \left(0.5 + 0.5(1 - 2z) \left(\frac{t-1}{b}\right)\right).$$

The entropy of z given s is

$$H(z|s) = -\sum_s \sum_z P(s, z) \log_2 P(z|s) = -\left(\frac{t-1}{2b}\right) \log_2 \left(\frac{t-1}{b}\right) - \left(\frac{b-t+1}{2b}\right) \log_2 \left(\frac{b-t+1}{b}\right).$$

Recall that the probability of error is $P_e = \frac{t-1}{2b}$. Then

$$H(z) = -(0.5 + P_e) \log_2 (0.5 + P_e) - (0.5 - P_e) \log_2 (0.5 - P_e) \quad (1)$$

and

$$
\begin{aligned}
H(z|s) &= -P_e \log_2 2P_e - (0.5 - P_e)\log_2(1 - 2P_e) \\
&= -0.5 - P_e \log_2 P_e - (0.5 - P_e)\log_2(0.5 - P_e).
\end{aligned} \tag{2}
$$

Equations (1) and (2) are in exact agreement with Equations (7) and (8) of Chapeau-Blondeau[26] for this specific case. Therefore

$$
\begin{aligned}
I(s,z) &= H(z) - H(z|s) \\
&= 0.5 + P_e \log_2 P_e - (0.5 + P_e)\log_2(0.5 + P_e) \quad \text{bits per sample.}
\end{aligned} \tag{3}
$$

4.2.5. Comparing $I(s,y)$ and $I(s,z)$

Recall that the probability of an error is $P_e = \frac{t-1}{2b}$, but $b \geq t-1$, $0 < b < 1$ and $t - 1 > 0$, therefore $0 < P_e \leq 0.5$. Hence $0.5 < 0.5 + P_e \leq 1$. Therefore the second term in Equation (3) is always negative and the maximum value of the third term is 0.5. Therefore $I(s,z) < 1$ always. Since $I(s,y) = 1$ we therefore have $I(s,y) > I(s,z)$ always and since from Section 3.3, s, y and z form a Markov chain $s \longrightarrow y \longrightarrow z$, the DPI holds for this example. We note that since $t > 1$ that $P_e > 0$. As $t \to 1$, $P_e \to 0$ and $I(s,z) \to 1$. When $t - 1 = b$, $P_e = 0.5$ and $I(s,z) = 0$. It might seem strange that even though the probability of error is 0.5 that the mutual information is zero, however this corresponds to the case where if the output is always zero, and therefore nothing can be said about the input.

5. EXAMPLE 2: BINARY ERASURE CHANNEL

Another system considered by Chapeau-Blondeau is the binary erasure channel.[7] In this case, the output has three values: 1, -1 and 0, and there are two thresholds, $\pm\theta$. Specifically, we have $g(y)$ given by

$$
z = g(y) = \begin{cases} 1 & \text{if} \quad y = s + n \geq \theta, \\ -1 & \text{if} \quad y = s + n \leq -\theta, \\ 0 & \text{otherwise,} \end{cases}
$$

where an output value of zero indicates the complete erasure of an input value, rather than its corruption.

However, in the textbooks,[1] the binary erasure channel is a channel in which there is no possibility of error, only a possibility of erasure, and hence there is no possibility of an inverted output. Chapeau-Blondeau's channel does allow for this possibility, so it is not equivalent to the classical binary erasure channel. In this paper, however, we will still refer to the channel considered by Chapeau-Blondeau as the binary erasure channel.

Chapeau-Blondeau showed, for a system where $\theta > 1$ and therefore, the signal is subthreshold for even noise probability densities, that the channel capacity occurred when $p(s = -1) = p(s = 1) = 0.5$ and as expected showed that this capacity has a maximum for nonzero noise, as the addition of noise allows threshold crossings to occur. He then aimed to show that the capacity at the output (that is, the maximum mutual information about s contained in z) could be greater than the capacity at the input (that is, the maximum mutual information about s contained in $s + n$). We note that this possibility appears to be ruled out by the DPI. Here we indicate the source of this apparent contradiction.

5.1. Binary symmetric channel

Let the input signal, s, be a random binary signal. For a binary memoryless channel, if p is the probability of error given the input is known and $p = p(z = 1|s = 0) = p(z = 0|s = 1)$, then this is known as a symmetric binary channel. For such a channel, the mutual information[1] is

$$
I(s,z) = H(z) - H(p),
$$

and the channel capacity, or maximum mutual information, between the input and the output occurs when $p(s = -1) = p(s = 1) = 0.5$ and is given by

$$
C = 1 + p\log_2(p) + (1-p)\log_2(1-p). \tag{4}
$$

5.2. Capacity at the input to binary erasure channel

Chapeau-Blondeau stated that when the noise has an even probability density, the input capacity to the binary erasure channel was given by the capacity of the binary symmetric channel, with the probability of error given the input given by $p = 1 - F_n(1) = 1 - P(n \leq 1) = P(n > 1)$, where F_n is the cumulative distribution function of the noise. This is correct, since if the noise is always between ± 1, a threshold at zero would correctly give the input signal. For noise values outside ± 1, the probability of error equals the probability that the noise is outside the thresholds.

Hence, Chapeau-Blondeau stated that the channel capacity at the input, C_{in}, was given by Equation (4).

5.3. Capacity at the output to binary erasure channel

For the classical binary erasure channel, where there is no possibility of error, only of erasure, if we denote the probability of erasure as α, then the capacity is given by $C = 1 - \alpha$, where capacity is achieved when $p(s = -1) = p(s = 1) = 0.5$.[1]

The channel considered by Chapeau-Blondeau, in which there is a possibility of error as well as erasure, is, as he noted, still a symmetric memoryless discrete channel, provided the noise has an even probability density function. Hence, for this case, capacity is achieved when $p(s = -1) = p(s = 1) = 0.5$. Chapeau-Blondeau derived formulae for the mutual information for such a channel, from which channel capacity can be found be taking the input probabilities as 0.5. These formulae depend on the cumulative distribution function of the noise.

For example, let the thresholds be $\theta = \pm 1.1$ (subthreshold signal) and the noise be uniform with a variance of $\sigma_n^2/12$. Hence, the noise has probability density

$$p(n) = \begin{cases} \frac{1}{\sigma_n} & -\sigma_n/2 \leq n \leq \sigma_n/2, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Therefore, the cumulative distribution function of the noise is

$$p(n < u) = F_n(u) = \begin{cases} 1 & u \geq \sigma_n/2, \\ \frac{u}{\sigma_n} + \frac{1}{2} & -\sigma_n/2 \leq u \leq \sigma_n/2, \\ 0 & u \leq -\sigma_n/2. \end{cases} \tag{6}$$

Using Equation (6), we can use Chapeau-Blondeau's Equations (10)-(17)[7] to calculate numerically the channel capacity for the binary erasure channel. Figure 2 shows a plot of the capacity of the binary erasure channel against the rms value of a uniform noise distribution when $\theta = \pm 1.1$. Note that the rms value of the noise is given by $\sigma_n/\sqrt{12}$. It also shows the capacity of the binary symmetric channel (calculated using Equation (4) with, as Chapeau-Blondeau stated, $p = F_n(1)$, since the noise density is symmetric). Note that the maximum value of the capacity for the binary erasure channel corresponds to a nonzero value of noise. This shows that stochastic resonance can occur in the channel capacity measure, verifying the results of Chapeau-Blondeau.[26]

The capacity of the binary symmetric channel is a decreasing function of the noise rms amplitude, except for small values, where the capacity is one bit per sample. This makes sense, since for small values of noise, the probability of error is zero (for example, by using a threshold at zero), and for larger values of noise, the probability of error increases as the rms noise amplitude increases.

5.4. Source of the apparent contradiction

Chapeau-Blondeau took C_{in} to be the maximum mutual information at the *output* of a binary symmetric channel, and C_{out} to be the maximum mutual information at the output of a binary erasure channel. Using the results indicated by Figure 2, we can obtain a graph for the ratio of C_{out} to C_{in}, as shown in Figure 3 for a uniform noise distribution and for $\theta = 1.1$. We note that Figure 3 is identical to the plot produced by Chapeau-Blondeau for uniform noise and $\theta = 1.1$.[7]

Chapeau-Blondeau interpreted this result to mean that for noise values where the ratio was greater than one, the capacity at the output of the binary erasure channel was greater than the capacity at the input.

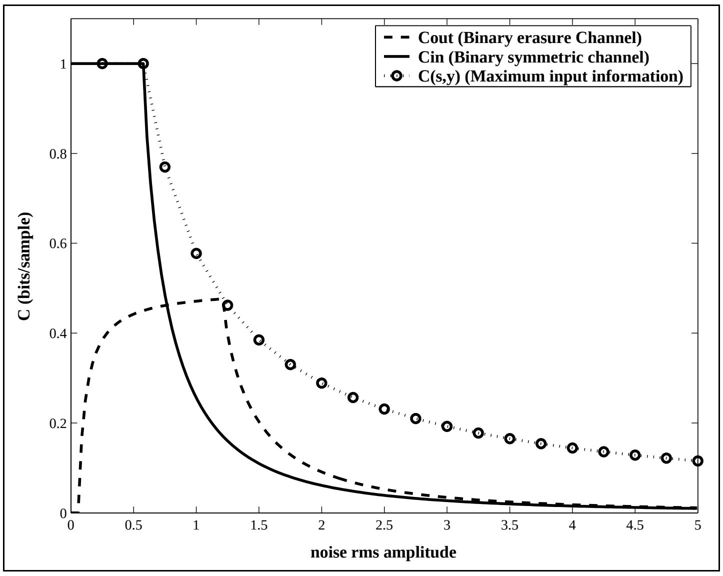

Figure 2. Channel capacity for the binary symmetric channel (Cin), the binary erasure channel ($Cout$), and the maximum input capacity ($C(s, y)$).

The DPI on the other hand states that the mutual information at the output of the binary erasure channel is less than or equal to the maximum mutual information at the input of that channel, which is also the *input* to the binary symmetric channel. Thus, Chapeau-Blondeau has shown only that for the values of noise rms where his ratio $C_{\text{out}}/C_{\text{in}} > 1$, more information can be obtained about s by the binary erasure channel than for the binary symmetric channel. This does not show that the output of the binary erasure channel is more detectable than the input, as claimed.

In actual fact, the capacity at the input should be taken as the maximum of the mutual information between the input signal, s, and the input signal plus noise, $y = s + n$. As we did in the example of Section 4, we can derive theoretically the mutual information between y and s, this time using Equation (5) for the noise density. Such a derivation must take into account that the probability density function of y depends on the value of σ_n. When $\sigma_n < 2$, s is retrievable without error, using a threshold at 0. In this case, the mutual information between s and y is the entropy of the input signal, s. However, when $\sigma_n \geq 2$, there exists a range of values $1 - \sigma_n/2 \leq y \leq -1 + \sigma_n/2$ for which it is impossible to tell whether s was -1 or 1. We assume that $p(s = 1) = p(s = -1) = 0.5$. Thus for $\sigma < 2$ we have

$$p_y(y) = \sum_s p(y|s)p(s) = \begin{cases} \frac{1}{2\sigma_b} & -1 - \frac{\sigma_n}{2} \leq y \leq -1 + \frac{\sigma_n}{2} \text{ and } 1 - \frac{\sigma_n}{2} \leq y \leq 1 + \frac{\sigma_n}{2}, \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

For $\sigma_n \geq 2$,

$$p_y(y) = \begin{cases} \frac{1}{2\sigma_b} & -1 - \frac{\sigma_n}{2} \leq y < 1 - \frac{\sigma_n}{2} \text{ and } -1 + \frac{\sigma_n}{2} < y \leq 1 + \frac{\sigma_n}{2}, \\ \frac{1}{\sigma_b} & 1 - \frac{\sigma_n}{2} \leq y \leq -1 + \frac{\sigma_n}{2}, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

Using Equations (7) and (8) we can derive the entropy of y as

$$H(y) = -\int p_y(y) \log_2 P_y(y) dy = \begin{cases} 1 + \log_2 \sigma_n & \sigma_n < 2, \\ \frac{2}{\sigma_n} + \log_2 \sigma_n & \sigma_n \geq 2. \end{cases}$$

We note that as σ_n becomes large, $H(y) \to \log_2 \sigma_n$, which is the entropy of a uniform distribution. As expected, this indicates that for very large noise y is very unlike the original binary signal, which has an entropy of 1 bit per sample.

As in Section 4.2.3, the entropy of y given s is simply the entropy of the uniform noise. Thus $H(y|s) = \log_2 \sigma_n$.

Hence, the mutual information between s and y is

$$I(s, y) = H(y) - H(y|s) = \begin{cases} 1 & \sigma_n < 2, \\ \frac{2}{\sigma_n} & \sigma_n \geq 2. \end{cases} \tag{9}$$

We assume without proof that due to the symmetry involved, capacity does occur when $p(s = 1) = p(s = -1) = 0.5$. We denote the capacity between s and y as $C(s, y)$. This is given by the mutual information formula of Equation (9) and is plotted in Figure 2. It can be seen that the capacity at the output of both the binary symmetric channel and binary erasure channel is less than or equal to the capacity at the input, $C(s, y)$, and that hence the DPI holds. We note that although we can obtain theoretically a formula for channel capacity at the input to a channel, $C(s, y)$, the theory does not tell us how to achieve that capacity. Figure 2 indicates that capacity is reached by the binary symmetric channel when the rms noise value is less than $2/\sqrt{12}$, and capacity is reached by the binary erasure channel when the rms noise value is about 1.215, however for other values we don't know how to reach capacity.

For the binary symmetric channel, where there is a noise induced maximum in the capacity, this maximum can be interpreted to mean that a certain nonzero value of noise can minimize the information lost in the channel.

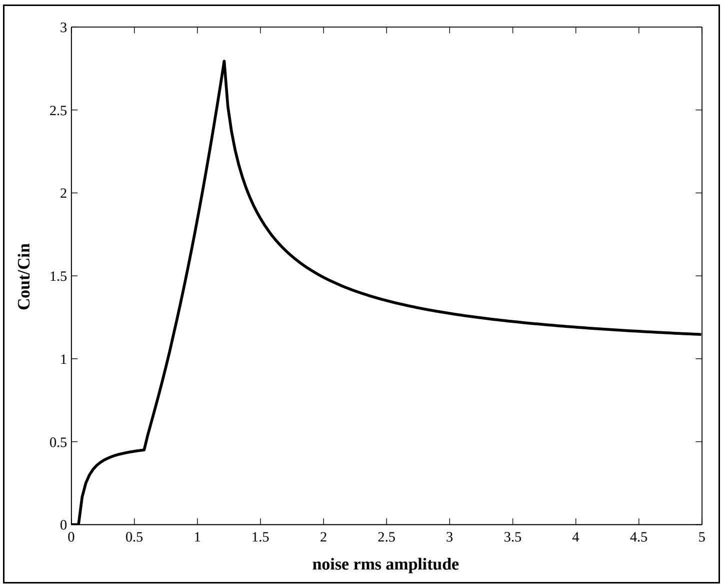

Figure 3. Ratio of Cout (capacity of binary erasure channel, with $\theta = 1.1$) to Cin (capacity of binary symmetric channel) against the rms value of uniformly distributed noise. For values of noise rms amplitude where the ratio is greater than unity, the binary erasure channel has greater capacity than the binary symmetric channel.

6. CONCLUSIONS

From this investigation, it is clear that although SNR gains may exist for periodic input signals, for random noisy aperiodic signals the DPI holds and hence no more information about the original signal can be gained by operating on that noisy signal, whether by signal processing, filtering, or by stochastic resonance. Specifically, this means that the addition of more noise to a noisy signal cannot be of benefit as far as obtaining an input-output mutual information gain is concerned. However, it certainly does not rule out the fact that the addition of noise at the input to a channel can maximize the mutual information at the output, in other words, the effect of stochastic resonance in aperiodic signals is perfectly valid. A new perspective is that for certain systems, it is possible to maximize the output mutual information by an optimal value of input noise, which means minimizing the information lost in the channel.

By analogy, this analysis lends further weight to arguments that the SNR metric for periodic signals is inadequate, and that other means of quantifying a system's performance are required. A number of authors address this problem, and, specifically, suggest more complete methods of quantifying stochastic resonance in threshold systems, for example Galdi *et al*,[15] Hänggi *et al*,[13] and Robinson *et al*.[27, 28]

ACKNOWLEDGMENTS

This work was funded by the Australian Research Council. We also warmly thank François Chapeau-Blondeau of the Université d'Angers, France, for useful discussions.

REFERENCES

1. T.M. Cover and J.A. Thomas, *Elements of Information Theory*, John Wiley and Sons, New York, 1991.
2. L.B. Kiss, "Possible breakthrough: significant improvement of signal to noise ratio by stochastic resonance," in *Chaotic, Fractal and Nonlinear Signal Processing*, R. Katz, ed., *American Institute of Physics Conference Proceedings* **375**, pp. 382–396, 1996.
3. K. Loerincz, Z. Gingl, and L.B. Kiss, "A stochastic resonator is able to greatly improve signal-to-noise ratio," *Physics Letters A* **224**, pp. 63–67, 1996.
4. J.M.G. Vilar and J.M. Rubí, "Divergent signal-to-noise ratio and stochastic resonance in monostable systems," *Physical Review Letters* **77**, pp. 2863–2866, 1996.
5. F. Chapeau-Blondeau, "Input-output gains for signal in noise in stochastic resonance," *Physics Letters A* **232**, pp. 41–48, 1997.
6. F. Chapeau-Blondeau and X. Godivier, "Theory of stochastic resonance in signal transmission by static nonlinear systems," *Physical Review E* **55**, pp. 1478–1495, 1997.
7. F. Chapeau-Blondeau, "Periodic and aperiodic stochastic resonance with output signal-to-noise ratio exceeding that at the input," *International Journal of Bifurcation and Chaos* **9**, pp. 267–272, 1999.
8. F. Liu, Y. Yu, and W. Wang, "Signal-to-noise ratio gain in neuronal systems," *Physical Review E* **63**, 2001.
9. Z. Gingl, P. Makra, and R. Vajtai, "High signal-to-noise ratio gain by stochastic resonance in a double well," *Fluctuation and Noise Letters* **1**, pp. L181–L188, 2002.
10. Z. Gingl, R. Vajtai, and L.B. Kiss, "Signal-to-noise ratio gain by stochastic resonance in a bistable system," *Chaos, Solitons and Fractals* **11**, pp. 1929–1932, 2000.
11. P. Makra, Z. Gingl, and L.B. Kish, "Signal-to-noise ratio gain in non-dynamical and dynamical bistable stochastic resonators," *Fluctuation and Noise Letters* **2**, pp. L147–L155, 2002.
12. J. Casado-Pascual, C. Denk, J. Gomez-Ordonez, M. Marillo, and P. Hänggi, "Gain in stochastic resonance: Precise numerics versus linear response theory beyond the two-mode approximation." To appear in Physical Review E, 2003.
13. P. Hänggi, M.E. Inchiosa, D. Fogliatti, and A.R. Bulsara, "Nonlinear stochastic resonance: The saga of anomalous output-input gain," *Physical Review E* **62**, pp. 6155–6163, 2000.
14. M.E. Inchiosa and A.R. Bulsara, "Nonlinear dynamic elements with noisy sinusoidal forcing: Enhancing response via nonlinear coupling," *Physical Review E* **52**, pp. 327–339, 1995.
15. V. Galdi, V. Pierro, and I.M. Pinto, "Evaluation of stochastic-resonance-based detectors of weak harmonic signals in additive white Gaussian noise," *Physical Review E* **57**, pp. 6470–6479, 1998.
16. L. Gammaitoni, P. Hänggi, and P. Jung, "Stochastic resonance," *Reviews of Modern Physics* **70**, pp. 223–287, 1998.
17. J.R. Cogdell, *Foundations of Electrical Engineering*, Prentice Hall International, London, 1996.
18. S. Mickan, D. Abbott, J. Munch, X.-C. Zhang, and T. van Doorn, "Analysis of system trade-offs for terahertz imaging," *Microelectronics Journal* **31**, pp. 503–514, 2000.
19. S. Mickan and X.-C. Zhang, "Chapter 8: T-ray sensing and imaging," in *Terahertz Sensing Technology: Electronic Devices Advanced Technology*, D. Woolard, M. S. Shur, and W. Leorop, eds., Word Scientific Publishing Company, 2003 (Accepted).
20. R. Bartussek, P. Hänggi, and P. Jung, "Stochastic resonance in optical bistable systems," *Physical Review E* **49**, pp. 3930–3939, 1994.
21. A.R. Bulsara and M.E. Inchiosa, "Noise-controlled resonance behavior in nonlinear dynamical systems with broken symmetry," *Physical Review Letters* **77**, pp. 2162–2165, 1996.

22. M.E. Inchiosa and A.R. Bulsara, "DC signal detection via dynamical asymmetry in a nonlinear device," *Physical Review E* **58**, pp. 115–127, 1998.
23. M. DeWeese and W. Bialek, "Information flow in sensory neurons," *Il Nuovo Cimento* **17**, pp. 733–741, 1995.
24. M.I. Dykman, D.G. Luchinsky, R. Mannella, P.V.E. McClintock, N.D. Stein, and N.G. Stocks, "Stochastic resonance in perspective," *Il Nuovo Cimento* **17**, pp. 661–683, 1995.
25. J.J. Collins, C.C. Chow, and T.T. Imhoff, "Aperiodic stochastic resonance in excitable systems," *Physical Review E* **52**, pp. R3321–R3324, 1995.
26. F. Chapeau-Blondeau, "Noise-enhanced capacity via stochastic resonance in an asymmetric binary channel," *Physical Review E* **55**, pp. 2016–2019, 1997.
27. J.W.C. Robinson, D.E. Asraf, A.R. Bulsara, and M.E. Inchiosa, "Information-theoretic distance measures and a generalization of stochastic resonance," *Physical Review Letters* **81**, pp. 2850–2853, 1998.
28. J.W.C. Robinson, J. Rung, A.R. Bulsara, and M.E. Inchiosa, "General measures for signal-noise separation in nonlinear dynamical systems," *Physical Review E* **63**, p. 011107, 2001.