SCIENTIFIC REP♦RTS

OPEN

Capacity of very noisy communication channels based on Fisher information

Fabing Duan[1], François Chapeau-Blondeau[2] & Derek Abbott[3]

We generalize the asymptotic capacity expression for very noisy communication channels to now include coloured noise. For the practical scenario of a non-optimal receiver, we consider the common case of a correlation receiver. Due to the central limit theorem and the cumulative characteristic of a correlation receiver, we model this channel noise as additive Gaussian noise. Then, the channel capacity proves to be directly related to the Fisher information of the noise distribution and the weak signal energy. The conditions for occurrence of a noise-enhanced capacity effect are discussed, and the capacity difference between this noisy communication channel and other nonlinear channels is clarified.

It is well known that, for an additive Gaussian noise channel and an energy constrained input signal, the channel capacity can be explicitly calculated[1–4]. In practical applications, however, communication systems frequently encounter non-Gaussian noise environments, for instance, underwater acoustic noise and low-frequency atmospheric noise[5–7]. Of all channels with power-constrained noise, the capacity of a Gaussian channel is the smallest[1–4]. Thus, the capacities of non-Gaussian channels are of great interest[5–15]. Moreover, from theoretical and practical viewpoints, a very interesting topic is the investigation of the channel capacity with very weak input signals, e.g. deep space communication channels[2,9] and qubit depolarizing channels[15]. A very noisy channel was introduced by Reiffen[8], and extended by Gallager[2] and Majani[9] to model many physical communication channels operating at very low signal-to-noise ratio (SNR). "Very noisy" channels with very low capacity are of significant interest to communications, since Shannon's theorem guarantees reliable communication as long as the capacity is nonzero[1–4,9,16]. Following the approaches developed in[2,8,9] and using a power series of characteristic functions, Nirenberg[5] derived a simple formula of the capacity for the coherent threshold channel with an optimum receiver. For memoryless channels with very weak inputs, Kullback[10], Verdú[11] and Prelov[12] explicitly expressed the asymptotic expressions of the channel capacity closely related to the Fisher information matrix. Recently, Kostal and Lansky[14] presented an approximate expression for the information capacity in a broad class of discrete-time channels under the constraint of vanishing input amplitude or power, which allows us to analyse the capacity of channels with memory in a convenient way[13,14].

In this paper, under the assumption of low SNR, we will further derive the capacity of a very noisy communication channel, wherein the optimum receiver may be unavailable and noise is not restricted to be white. Based on the central limit theorem, we argue that, for sufficiently large observation times and with the constraint of weak signal energy, the receiver output tends to be Gaussian distributed, and the channel capacity is then computed by a simple formula being directly related to the Fisher information of the noise distribution. We demonstrate the enhancement of capacity via stochastic resonance will not occur in very noisy communication channel with an optimum receiver, but it can occur with generalized correlation receivers suited for practical implementation. Finally, we compare the asymptotic capacity expressions of this noisy communication channel with other capacity formulas in refs 10–14.

Results

Channel capacity for coloured noise. For the M-ary communication channel shown in Fig. 1, the observation data vector **X** contains the additive noise vector **Z** and the signal vector $\mathbf{S}_m$, $m = 1, 2, \cdots, M$. With the assumptions of white noise and very low SNR, Nirenberg[5] derived the capacity for the coherent threshold channel with an optimum receiver. We briefly present the conclusions of ref. 5 for reference (see **Methods**). However, the

[1]Institute of Complexity Science, Qingdao University, 266071 Qingdao, China. [2]Laboratoire Angevin de Recherche en Ingénierie des Systèmes (LARIS), Université d'Angers, 49000 Angers, France. [3]Centre for Biomedical Engineering (CBME) and School of Electrical & Electronic Engineering, The University of Adelaide, Adelaide, SA 5005, Australia. Correspondence and requests for materials should be addressed to F.D. (email: fabing.duan@gmail.com)
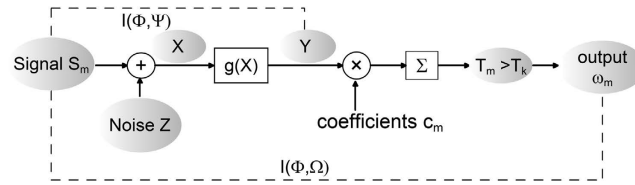
Figure 1. Mutual information $I(\Phi, \Omega)$ of the communication channel and $I(\Phi, \Psi)$ of the nonlinear channel.

idealized assumption of white noise is unpractical, and the coloured noise has practical significance[2–4]. We here further derive a general asymptotic expression of the channel capacity for coloured noise, which applies to not only the optimum receiver but also an arbitrary correlation receiver.

In the case of coloured noise and for very low SNR, the conditional probability function can be expanded to the first order

$$f(\mathbf{x}|\mathbf{s}_m) = f_{\mathbf{Z}}(\mathbf{x} - \mathbf{s}_m) \approx f_{\mathbf{Z}}(\mathbf{x}) - (\nabla f_{\mathbf{Z}}(\mathbf{x}))^T \mathbf{s}_m = f_{\mathbf{Z}}(\mathbf{x})[1 + \Gamma(\mathbf{s}_m, \mathbf{x})],\tag{1}$$

where the operator $\nabla \triangleq \left[\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \cdots, \frac{\partial}{\partial x_N}\right]^T$ and the statistic $\Gamma(\mathbf{s}_m, \mathbf{x}) = -(\nabla \ln f_{\mathbf{Z}}(\mathbf{x}))^T \mathbf{s}_m$. Here, from an information theory point of view by Reiffen[8] and Gallager[2], the module $|\Gamma(\mathbf{s}_m, \mathbf{x})| \ll 1$ indicates the channel is very noisy in the sense that the channel output is almost independent of the input. For M equiprobable signals $\mathbf{S}_m$, the receiver takes the maximum likelihood rule

$$f(\mathbf{x}|\mathbf{s}_m) > f(\mathbf{x}|\mathbf{s}_k), \quad (k \neq m)\tag{2}$$

to optimally choose mth signal[5,17]. Substituting equation (1) into equation (2), the optimum receiver

$$\Gamma(\mathbf{s}_m, \mathbf{x}) > \Gamma(\mathbf{s}_k, \mathbf{x}), \quad (k \neq m)\tag{3}$$

enables us to decide if the mth signal was transmitted. For clarity, we state that the statistic $\Gamma(\mathbf{s}_m, \mathbf{x})$ and the maximum likelihood decoding rule of equation (2) compose an optimum correlation receiver. The channel output is the decoding signal ω_m of the receiver, as shown in Fig. 1.

Then, supposing the zero-mean $E_{\mathbf{S}}(\mathbf{s}) = \mathbf{0}$ and extending the very noisy vector channel ($|\Gamma(\mathbf{s}_m, \mathbf{x})| \ll 1$)[2,5,8,9], the mutual information between the input signal space $\boldsymbol{\Phi}$ and the channel output space $\boldsymbol{\Omega}$ is given by

$$
\begin{aligned}
I(\boldsymbol{\Phi}, \boldsymbol{\Omega}) &= \frac{1}{2} \iint f_{\mathbf{X}}(\mathbf{x}) f_{\mathbf{S}}(\mathbf{s}) \Gamma(\mathbf{s}, \mathbf{x})^T \Gamma(\mathbf{s}, \mathbf{x}) d\mathbf{s} d\mathbf{x} \\
&\quad - \frac{1}{2} \int f_{\mathbf{X}}(\mathbf{x}) \left| \int f_{\mathbf{S}}(\mathbf{s}) \Gamma(\mathbf{s}, \mathbf{x}) d\mathbf{s} \right|^2 d\mathbf{x} \\
&= \frac{1}{2} E_{\mathbf{S}}[\mathbf{s}^T \mathbf{J}(f_{\mathbf{Z}}) \mathbf{s}],
\end{aligned}\tag{4}
$$

where the Fisher information matrix of the noise distribution is defined as[3,7]

$$\mathbf{J}(f_{\mathbf{Z}}) = E_{\mathbf{Z}}[(\nabla \ln f_{\mathbf{Z}}(\mathbf{z}))(\nabla \ln f_{\mathbf{Z}}(\mathbf{z}))^T].\tag{5}$$

It is noted that $\mathbf{J}(f_{\mathbf{Z}})$ is also called the Fisher information of a location parameter or the shift-invariant Fisher information[3,6,7,18], which can be viewed as a special case of the Fisher information measuring the statistical information contained in data about an unknown parameter. Therefore, with the energy constraint of $E_{\mathbf{S}}(\mathbf{s}^T\mathbf{s}) \leq \varepsilon$ and for the standardized vector $\mathbf{u} = \mathbf{s}/\sqrt{E_{\mathbf{S}}(\mathbf{s}^T\mathbf{s})}$, the channel capacity can be expressed as

$$C = \max_{f_{\mathbf{S}}} I(\boldsymbol{\Phi}, \boldsymbol{\Omega}) = \max_{f_{\mathbf{S}}} \frac{1}{2} E_{\mathbf{S}}(\mathbf{s}^T\mathbf{s}) \, E_{\mathbf{S}}[\mathbf{u}^T \mathbf{J}(f_{\mathbf{Z}})\mathbf{u}] = \frac{1}{2}\varepsilon\Lambda,\tag{6}$$

where Λ is the largest eigenvalue of the matrix $\mathbf{J}(f_{\mathbf{Z}})$ and $\mathbf{u}$ takes the corresponding eigenvector.

For positive definite matrixes $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{N \times N}$ and an arbitrary column vector $\mathbf{X} \in \mathbb{R}^N$, the inequality $\mathbf{X}^T(\mathbf{A} - \mathbf{B})\mathbf{X} \geq 0$ is abbreviated as $\mathbf{A} \succeq \mathbf{B}$. Then, for the positive semidefinite matrix

$$E_{\mathbf{Z}}\left[\left(\nabla \ln f_{\mathbf{Z}}(\mathbf{z}) - \sum_{\mathbf{Z}}^{-1}\mathbf{z}\right)\left(\nabla \ln f_{\mathbf{Z}}(\mathbf{z}) - \sum_{\mathbf{Z}}^{-1}\mathbf{z}\right)^T\right] \succeq \mathbf{0}$$

and the noise covariance matrix $\sum_{\mathbf{Z}} = E_{\mathbf{Z}}(\mathbf{z}\mathbf{z}^T)$, we have

$$\mathbf{J}(f_{\mathbf{Z}}) \succeq E_{\mathbf{Z}}[(\nabla \ln f_{\mathbf{Z}}(\mathbf{z}))\mathbf{z}^T] \sum_{\mathbf{Z}}^{-1} E_{\mathbf{Z}}[(\nabla \ln f_{\mathbf{Z}}(\mathbf{z}))^T \mathbf{z}] = \sum_{\mathbf{Z}}^{-1},\tag{7}$$

where the equality occurs for N-dimensional Gaussian distribution $f_{\mathbf{Z}}(\mathbf{z}) = \exp(-\mathbf{z}^T\sum_{\mathbf{Z}}^{-1}\mathbf{z}/2)/\sqrt{(2\pi)^N \det(\sum_{\mathbf{Z}})}$ with its Fisher information matrix $\mathbf{J}(f_{\mathbf{Z}}) = \sum_{\mathbf{Z}}^{-1}$. Thus, equation (7) indicates the maximum eigenvalue of $\sum_{\mathbf{Z}}^{-1}$ is

less than that of Fisher information matrix of non-Gaussian noise. This result extends the conclusion of equation (36) by Nirenberg[5], and also confirms that, in terms of the channel capacity, zero-mean Gaussian noise is the worst case given that the noise vector has a fixed covariance matrix[3,4].

However, we note the channel capacity of equation (6) is achieved by the optimum receiver of equation (3). In many practical cases, the optimum receiver may be not implementable for the unknown noise distribution or the non-closed form of distributions (e.g. α-stable noise[19]). Thus, we further consider the generalized correlation receiver

$$T_m = \mathbf{c}_m^T g(\mathbf{x}) > T_k = \mathbf{c}_k^T g(\mathbf{x}), \quad (k \neq m) \tag{8}$$

where the coefficient vector $\mathbf{c}_m = [c_{m1}, c_{m2}, \cdots, c_{mN}]^T$ and the function $g(\mathbf{x})$ is not restricted to be memoryless. For the zero-mean vector of $E_{\mathbf{Z}}[g(\mathbf{z})] = \mathbf{0}$ (for a shift in mean)[6] under $f_{\mathbf{Z}}$ and for very low SNR, $g(\mathbf{x})$ can be expanded to the first-order

$$g(\mathbf{x}) = g(\mathbf{z} + \mathbf{s}_m) \approx g(\mathbf{z}) + (\nabla g(\mathbf{z}))^T \mathbf{s}_m. \tag{9}$$

Then, for a large observation size N, the statistic T_m has the mean $E_{\mathbf{Z}}(T_m) \approx \mathbf{c}_m^T E_{\mathbf{Z}}[(\nabla g(\mathbf{z}))^T] \mathbf{s}_m$ and the variance $\mathrm{var}(T_m) \approx \mathbf{c}_m^T E_{\mathbf{Z}}[g(\mathbf{z})g(\mathbf{z})^T] \mathbf{c}_m$. Using the Cholesky decomposition of the symmetrical matrix $\mathbf{V} = E_{\mathbf{Z}}[g(\mathbf{z})g(\mathbf{z})^T] = \mathbf{L}\mathbf{L}^T$, the output SNR of the receiver can be calculated as

$$
\begin{aligned}
\frac{E_{\mathbf{Z}}^2(T_m)}{\mathrm{var}(T_m)} &= \max_{\mathbf{c}_m} \frac{|\mathbf{c}_m^T E_{\mathbf{Z}}[(\nabla g(\mathbf{z}))^T]\mathbf{s}|^2}{\mathbf{c}_m^T E_{\mathbf{Z}}[g(\mathbf{z})g(\mathbf{z})^T]\mathbf{c}_m} \\
&= \max_{\mathbf{c}_m} \frac{|(\mathbf{L}^T\mathbf{c}_m)^T (\mathbf{L}^{-1}E_{\mathbf{Z}}[(\nabla g(\mathbf{z}))^T]\mathbf{s})|^2}{(\mathbf{L}^T\mathbf{c}_m)^T(\mathbf{L}^T\mathbf{c}_m)} \\
&= \mathbf{s}^T E_{\mathbf{Z}}[\nabla g(\mathbf{z})]\mathbf{V}^{-1}E_{\mathbf{Z}}[(\nabla g(\mathbf{z}))^T]\mathbf{s},
\end{aligned}
\tag{10}
$$

by optimally choosing $\mathbf{c}_m = \mathbf{V}^{-1}E_{\mathbf{Z}}[(\nabla g(\mathbf{z}))^T]\mathbf{s}$. Then, we argue that, for sufficiently large observation times and with the constraint of weak signal energy, the receiver output tends to be Gaussian distributed, and the capacity can be approximately calculated as

$$C_g = \frac{1}{2} \ln\left(1 + \frac{E_{\mathbf{Z}}^2(T_m)}{\mathrm{var}(T_m)}\right) \approx \max_{f_{\mathbf{s}}} \frac{1}{2} E_{\mathbf{S}}\{\mathbf{s}^T E_{\mathbf{Z}}[\nabla g(\mathbf{z})]\mathbf{V}^{-1}E_{\mathbf{Z}}[(\nabla g(\mathbf{z}))^T]\mathbf{s}\} = \frac{1}{2}\varepsilon\Lambda_g, \tag{11}$$

where Λ_g is the largest eigenvalue of the matrix $E_{\mathbf{Z}}[\nabla g(\mathbf{z})]\mathbf{V}^{-1}E_{\mathbf{Z}}[(\nabla g(\mathbf{z}))^T]$. Observing

$$E_{\mathbf{Z}}[\nabla g(\mathbf{z})] = E_{\mathbf{Z}}[(-\nabla \ln f_{\mathbf{Z}}(\mathbf{z}))g(\mathbf{z})^T], \tag{12}$$

and for the positive semidefinite matrix

$$E_{\mathbf{Z}}[(-\nabla \ln f_{\mathbf{Z}}(\mathbf{z}) - \mathbf{A})(-\nabla \ln f_{\mathbf{Z}}(\mathbf{z}) - \mathbf{A})^T] \succeq \mathbf{0},$$

we have

$$J(f_{\mathbf{Z}}) \succeq E_{\mathbf{Z}}[\nabla g(\mathbf{z})]\mathbf{V}^{-1}E_{\mathbf{Z}}[(\nabla g(\mathbf{z}))^T], \tag{13}$$

with $\mathbf{A} = E_{\mathbf{Z}}[(-\nabla \ln f_{\mathbf{Z}}(\mathbf{z}))g(\mathbf{z})^T]\mathbf{V}^{-1}g(\mathbf{z})$ and the equality occurring for $g(\mathbf{x}) = -\nabla \ln f_{\mathbf{Z}}(\mathbf{x})$. This inequality (13) indicates that the eigenvalue Λ of $J(f_{\mathbf{Z}})$ is not less than the eigenvalue Λ_g of the matrix $E_{\mathbf{Z}}[\nabla g(\mathbf{z})]\mathbf{V}^{-1}E_{\mathbf{Z}}[(\nabla g(\mathbf{z}))^T]$. Therefore, based on equations (6), (11) and (13), we find

$$C_g \leq C, \tag{14}$$

which extends the conclusion of ref. 5 to the case of coloured noise. In addition, the equality in equation (13) also demonstrates the receiver of equation (8) is optimal when $g(\mathbf{x}) = -\nabla \ln f_{\mathbf{Z}}(\mathbf{x})$, i.e. the optimum receiver of equation (3).

We argue that the asymptotic capacity expression of equation (11) has a broader applicability for an arbitrary correlation receiver operated in coloured or white noise environments. As a simple check for the consistency of the results from equation (11) to equation (14), we consider the case of white noise. Immediately, due to the statistical independence of $g(\mathbf{z})$, the expectation matrices $E_{\mathbf{Z}}[\nabla g(\mathbf{z})] = E_z[g'(z)]\mathbf{I}$ and $\mathbf{V} = E_z[g^2(z)]\mathbf{I}$. Here, the derivative $g'(z) = dg(z)/dz$ and $\mathbf{I}$ is the unit matrix. Therefore, the matrix $E_{\mathbf{Z}}[\nabla g(\mathbf{z})]\mathbf{V}^{-1}E_{\mathbf{Z}}[(\nabla g(\mathbf{z}))^T]$ in equation (11) has N identical eigenvalues $\Lambda_g = E_z^2[g'(z)]/E_z[g^2(z)]$, and the channel capacity becomes

$$C_g \approx \frac{1}{2}\varepsilon\Lambda_g = \frac{1}{2}\varepsilon\frac{E_z^2[g'(z)]}{E_z[g^2(z)]} \leq C = \frac{1}{2}\varepsilon\Lambda = \frac{1}{2}\varepsilon J(f_z), \tag{15}$$

where the eigenvalue $\Lambda = J(f_z)$ corresponds to the Fisher information matrix $J(f_{\mathbf{Z}})$ in equation (6). Using the Cauchy-Schwarz inequality and integration by parts $E_z^2[g'(z)] = E_z^2[g(z)(-f_z'(z)/f_z(z))] \leq E_z[g^2(z)]J(f_z)$,

and the equality in equation (15) occurs when $g(z) = -f_z'(z)/f_z(z)$ that specifies the optimum receiver in the presence of white noise[5].

Conditions for noise-enhanced capacity.

Since the emergence of the concept of stochastic resonance[20], the employment of noise in enhancing the performance of nonlinear systems has become an interesting option[13,14,21–36]. Initially, the mechanism of stochastic resonance manifests itself as a time-scale matching condition for the noise-induced characteristic time of systems and the signal period[20,27]. Later, the notion of stochastic resonance has been widened to a number of different mechanisms, e.g. aperiodic stochastic resonance[22] and suprathreshold stochastic resonance[31]. For such stochastic resonance effects[22,31], there is no matching time-scale that corresponds to the input aperiodic or information-carrying random signal, but the system performance still reaches a maximum at an optimal non-zero noise level. Therefore, the noise-enhanced effect, instead of stochastic resonance, becomes a more appropriate term for describing the enhancement effect of system responses via the addition of noise. Here, if the channel capacity reaches a maximum at an optimal non-zero noise level, then the noise-enhanced capacity effect occurs. Otherwise, upon increasing the noise level, the channel capacity monotonically decreases, this is to say, the noise-enhanced capacity effect does not exist.

There are two approaches for varying the noise in stochastic resonance. One is tuning the noise level but not changing the noise type, and the other is adding extra noise to a given noisy signal, while the extra noise type may be different form the original one. Next, we will demonstrate the occurrence or nonoccurrence condition of the noise-enhanced capacity effect by the above mentioned methods.

First, we will prove that no noise-enhanced capacity effect exists for tuning the scaled noise level in an optimum receiver. For the scaled noise vector $\mathbf{Z} = \mathbf{D}\mathbf{Z}_n$, the covariance matrix $\sum_{\mathbf{Z}}$ can be factored as $\sum_{\mathbf{Z}} = \mathbf{D}\mathbf{D}^T$ and the standardized noise vector $\mathbf{Z}_n$ has a covariance matrix being the unit matrix $\mathrm{E}(\mathbf{z}_n\mathbf{z}_n^T) = \mathbf{I}$[7]. A well-known scaling property of the Fisher information matrix is[7,18,37–40]

$$\mathbf{J}(f_{\mathbf{Z}}) = \mathbf{D}^{-1}\mathbf{J}(f_{\mathbf{Z}_n})(\mathbf{D}^T)^{-1}, \tag{16}$$

which implies the largest eigenvalue Λ of $\mathbf{J}(f_{\mathbf{Z}})$ is a monotonically decreasing function of $\Lambda_n/\det(\sum_{\mathbf{Z}})$ for the determinants $\det^2(\mathbf{D}) = \det^2(\mathbf{D}^T) = \det(\sum_{\mathbf{Z}})$. Here, the largest eigenvalue of $\mathbf{J}(f_{\mathbf{Z}_n})$ is Λ_n that is a fixed quantity for $\mathbf{Z}_n$. For such a channel with its optimum receiver, equation (11) indicates the channel capacity $C \approx \frac{1}{2}\varepsilon\Lambda_n/\det(\sum_{\mathbf{Z}})$ monotonically decreases as the noise intensity increases. Thus, no noise-enhanced capacity phenomenon will occur by tuning the noise level.

For instance, we consider a threshold receiver based on the function $g(x) = \mathrm{sign}(x)$ and the Laplacian white noise with its distribution $f_z(x) = \exp(-\sqrt{2}|x|/\sigma)/(\sqrt{2}\sigma)$. We note that the threshold receiver is optimum for the Laplacian noise, and $g(x) = \mathrm{sign}(x) = -f_z'(z)/f_z(z)$ satisfies the equality condition in equation (15). In this case, the channel capacity in equation (15) can be calculated as $C \approx \frac{1}{2}\varepsilon J(f_z) = \varepsilon/\sigma^2$, which monotonically decreases as the noise level σ increases. Thus, there is no noise-enhanced capacity effect.

Secondly, we usually have a given signal corrupted by noise, and the initial noise level is unadjustable. We will prove that the addition of extra noise cannot further improve the channel capacity achieved by the optimum receiver. Under this circumstance, we add an extra noise vector $\mathbf{W}$, independent of $\mathbf{Z}$ and $\mathbf{S}_m$, to the observation $\mathbf{X}$, and the updated data vector is

$$\widetilde{\mathbf{X}} = \mathbf{Z} + \mathbf{W} + \mathbf{S}_m = \mathbf{U} + \mathbf{S}_m, \tag{17}$$

where the composite noise vector $\mathbf{U} = \mathbf{Z} + \mathbf{W}$ with its distribution $f_{\mathbf{U}}$. In this case, we should employ the statistic $\Gamma(\mathbf{s}_m, \widetilde{\mathbf{x}}) = (-\nabla \ln f_{\mathbf{U}}(\widetilde{\mathbf{x}}))^T \mathbf{s}_m$ to specify the optimum receiver, and the corresponding capacity is then given by

$$\widetilde{C} = \frac{1}{2}\varepsilon\widetilde{\Lambda} \tag{18}$$

with the largest eigenvalue $\widetilde{\Lambda}$ of the Fisher information matrix $\mathbf{J}(f_{\mathbf{U}})$. For any nonsingular matrix $\mathbf{A} \in \mathbb{R}^{N\times N}$, the Fisher information matrix inequality[3,37–40] holds for

$$\mathbf{A}\mathbf{J}(f_{\mathbf{Z}})\mathbf{A}^T + (\mathbf{I} - \mathbf{A})\mathbf{J}(f_{\mathbf{W}})(\mathbf{I} - \mathbf{A})^T \succeq \mathbf{J}(f_{\mathbf{U}}), \tag{19}$$

we then find the largest eigenvalue Λ of $\mathbf{J}(f_{\mathbf{Z}})$ is not less than the largest eigenvalue $\widetilde{\Lambda}$ of $\mathbf{J}(f_{\mathbf{U}})$ and

$$\widetilde{C} \leq C. \tag{20}$$

This result of equation (20) clearly shows that stochastic resonance cannot further improve the channel capacity achieved by the optimum receiver, regardless of adding white or coloured noise vector $\mathbf{W}$.

Thirdly, we note that the above two negative conditions of the noise-enhanced capacity effect arise with the optimum receiver matched to the distribution of the background noise. By Contrast, if the generalized correlation receivers of equation (8) are not optimal for the background noise, stochastic resonance may play an important role in the enhancement of capacity. For example, we consider non-scaled Gaussian mixture noise vector $\mathbf{W}$ with its distribution[6,21,28,33]
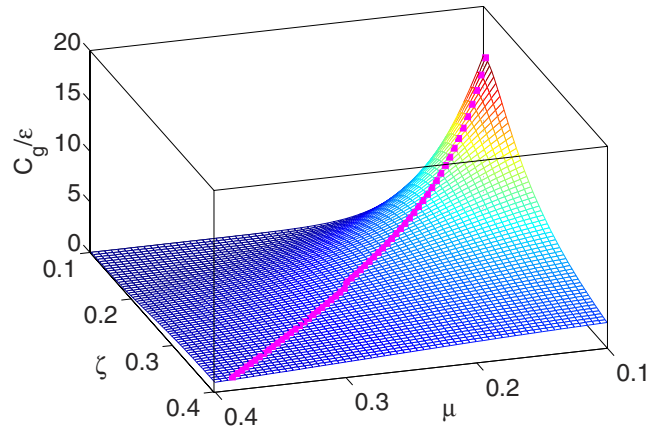
Figure 2. Stochastic resonance effect of the capacity per signal energy $C_g/\varepsilon = \Lambda_g/2$ in equation (11) versus the noise parameters μ and ζ in equation (21). Here, the correlation coefficient $\rho_1 = 0.2$ and $\rho_2 = 0$ in the coloured noise model of equation (22). The corresponding maxima of C_g/ε versus optimal values of ζ are also marked by squares.

$$f_w(x) = \frac{1}{2\sqrt{2\pi\zeta^2}}\left[\exp\left(\frac{-(x-\mu)^2}{2\zeta^2}\right) + \exp\left(\frac{-(x+\mu)^2}{2\zeta^2}\right)\right],$$

(21)

where the variance $\sigma_w^2 = \mu^2 + \zeta^2$ and parameters $\mu, \zeta \geq 0$. A useful coloured noise model of the first-order moving-average[41] as

$$Z_n = \rho_1 W_{n-1} + W_n + \rho_2 W_{n+1},$$

(22)

where the correlation coefficients are $\rho_{1,2}$ and $\mathbf{W} = [W_1, W_2, \cdots, W_N]^T$ is an independent identically distributed (i.i.d.) random vector. For small values of $\rho_{1,2}$ ($|\rho_{1,2}| \ll 1$), the dependence among noise samples Z_n will be weak[41]. The signum function $g(x) = \text{sign}(x)$ is adopted to construct the generalized correlation receiver of equation (8), which is not optimal for the coloured noise $\mathbf{Z}$. The optimum receiver indicated in equation (3) for the coloured noise $\mathbf{Z}$ is rather complicated, since the distribution $f_{\mathbf{Z}}$ does not have a tractable analytic expression[41]. Using the approach developed in ref. 41, we have the expectation matrix

$$E_{\mathbf{Z}}[\nabla g(\mathbf{z})] \approx E_W[g'(w)]\mathbf{I},$$

(23)

with the unit matrix $\mathbf{I}$ and $E_W(\cdot) = \int \cdot f_W(x)\,dx$, and the matrix $\mathbf{V}$ becomes tridiagonal with elements

$$\mathbf{V}_{i,i} = E_W[g^2(w)],$$

(24)

$$\mathbf{V}_{i,i+1} = \mathbf{V}_{i+1,i} = (\rho_1 + \rho_2)E_W[wg(w)]E_W[g'(w)],$$

(25)

for $i = 1, 2, \cdots, N$, and other elements are higher-order infinitesimal of $\rho_1 + \rho_2$ ($|\rho_{1,2}| \ll 1$). Then, we calculate the largest eigenvalue of the matrix $E_{\mathbf{Z}}[\nabla g(\mathbf{z})]\mathbf{V}^{-1}E_{\mathbf{Z}}[(\nabla g(\mathbf{z}))^T]$ as

$$\Lambda_g = \frac{E_W^2[g'(w)]}{E_W[g^2(w)] - 2|\rho_1 + \rho_2|E_W[wg(w)]E_W[g'(w)]}$$

$$= \frac{\frac{2}{\pi\zeta^2}\exp\left(-\frac{\mu^2}{\zeta^2}\right)}{1 - 2|\rho_1 + \rho_2|\left[\frac{2}{\pi}\exp\left(-\frac{\mu^2}{\zeta^2}\right) + \sqrt{\frac{2}{\pi}}\frac{\mu}{\zeta}\exp\left(-\frac{\mu^2}{2\zeta^2}\right)\text{erf}\left(\frac{\mu}{\sqrt{2}\zeta}\right)\right]},$$

(26)

where the error function $\text{erf}(x) = 2/\sqrt{\pi}\int_0^x \exp(-t^2)\,dt$. In Fig. 2, we show the capacity per signal energy $C_g/\varepsilon = \Lambda_g/2$ in equation (11) versus the noise parameters μ and ζ in equation (21). Here, the correlation coefficient $\rho_1 = 0.2$ and $\rho_2 = 0$ in the coloured noise model of equation (22). We regard the parameters $\pm\mu$ as the peak locations of the Gaussian mixture distribution in equation (21), while the parameter ζ as the noise level. It is then clearly shown in that Fig. 2, upon increasing ζ for a fixed value of μ (the noise variance σ_w^2 also increases), the noise-enhanced capacity effects exist. The corresponding maxima of C_g/ε versus optimal values of ζ are also marked by squares in Fig. 2.

We emphasize that the above noise-enhanced capacity effect is an illustrative case of stochastic resonance that exists for a suboptimal receiver not matching the background noise. However, this mismatch condition is not the decision criteria for the occurrence of the noise-enhanced effect, since the example illustration is under

the assumptions of a small signal and a correlation receiver with a large observation size. Beyond these restrictive assumptions, the noise-enhanced effect has been frequently observed[21,24,25,28–31]. For instance, the noise-enhanced effect has been demonstrated for non-weak signals in threshold neurons[25,29,31], where an optimal matching condition is inapplicable to the neuronal model immersed in complex noisy environments. It is sufficiently recognized that a well-established criterion for the noise-enhanced effect is to observe an optimal noise level at which the system response can be optimized.

Discussion

In this paper, we analyse the capacity of a very noisy communication channel with correlation receivers. With the weak signal energy constraint and for very low SNR, we generalize an asymptotic expression of capacity achieved by the optimum receivers in a coloured noisy environment. Moreover, for the case when the optimum receiver is unavailable in practice, a capacity formula is presented for the communication channel with a generalized correlation receiver. We further discuss the occurrence condition of the noise-enhanced capacity effect in the considered communication channel.

A similar asymptotic expression of capacity is also obtained in memoryless[10,11] or memory additive-noise channels[12–14]. We emphasize the asymptotic capacity expressions of equations (6) and (11) are different from that in previous literature[10–14]. In Fig. 1, for the channel output $\mathbf{Y} = g(\mathbf{X})$, these studies assume the conditional probability density as $f_{\mathbf{Y}|\mathbf{S}}(\mathbf{y}|\mathbf{s})$. Then, the Fisher information matrix is defined as[10–14]

$$\mathbf{J}(f_{\mathbf{Y}|\mathbf{S}}) = \mathrm{E}_{\mathbf{Y}|\mathbf{S}}[(\nabla_{\mathbf{S}} \ \ln f_{\mathbf{Y}|\mathbf{S}} (\mathbf{y}|\mathbf{s}))(\nabla_{\mathbf{S}} \ \ln f_{\mathbf{Y}|\mathbf{S}} (\mathbf{y}|\mathbf{s}))^T], \tag{27}$$

with the operator $\nabla_{\mathbf{S}} \triangleq \left[\frac{\partial}{\partial s_1}, \frac{\partial}{\partial s_2}, \cdots, \frac{\partial}{\partial s_N}\right]^T$. Then, for the zero-mean signal vector $\mathrm{E}_{\mathbf{S}}(\mathbf{s}) = \mathbf{0}$ and the weak signal energy ε, the mutual information between the input space $\mathbf{\Phi}$ and the output space $\mathbf{\Psi}$ is approximated as[10–14]

$$I(\mathbf{\Phi}, \mathbf{\Psi}) \approx \frac{1}{2}\mathrm{E}_{\mathbf{S}}[\mathbf{s}^T\mathbf{J}(f_{\mathbf{Y}|\mathbf{S}})|_{\mathbf{s}=\mathbf{0}}\mathbf{s}], \tag{28}$$

which is different from the mutual information $I(\mathbf{\Phi}, \mathbf{\Omega})$ of equation (4) based on the Fisher information matrix $\mathbf{J}(f_{\mathbf{Z}})$ of the noise distribution $f_{\mathbf{Z}}$. It is shown in Fig. 1 that the receiver multiplies nonlinear transformation $g(\mathbf{x})$ with optimized coefficients, and obtains a cumulative statistic T_m that decides whether the mth signal $\mathbf{S}_m$ is sent or not. Then, the considered communication channel chooses an optimal signal $\mathbf{S}_m$ from the signal space to maximize the average mutual information. Since the receiver collecting the weighted nonlinear outputs as the statistic $T_m = \mathbf{c}_m^T g(\mathbf{x})$, and for any nonlinear function g, the distribution of T_m tends to be Gaussian. This leads to the asymptotic expressions of capacity of equations (6) and (11). We recognize the asymptotic capacity expressions in equations (6) and (11) have application in the context of a very noisy communication channel with a correlation receiver. As a new analytical result of the channel capacity, it has theoretical significance and deserves some exposition.

We also note that, for the linear transfer function of $\mathbf{Y} = \mathbf{Z} + \mathbf{S}$, the conditional probability density $f_{\mathbf{Y}|\mathbf{S}}(\mathbf{y}|\mathbf{s}) = f_{\mathbf{Z}}(\mathbf{y} - \mathbf{s})$, the Fisher information matrix of equation (27) becomes

$$\mathbf{J}(f_{\mathbf{Y}|\mathbf{S}}) = \mathrm{E}_{\mathbf{Y}|\mathbf{S}}[(\nabla_{\mathbf{S}} \ \ln f_{\mathbf{Z}} (\mathbf{y} - \mathbf{s}))(\nabla_{\mathbf{S}} \ \ln f_{\mathbf{Z}} (\mathbf{y} - \mathbf{s}))^T] = \mathrm{E}_{\mathbf{Z}}[(\nabla \ \ln f_{\mathbf{Z}} (\mathbf{z}))(\nabla \ \ln f_{\mathbf{Z}} (\mathbf{z}))^T] = \mathbf{J}(f_{\mathbf{Z}}), \tag{29}$$

where the differentiation operator ∇ with respect to $\mathbf{S}$ is equivalent to differentiation with respect to $\mathbf{Z}$[3]. Therefore, for the linear additive-noise channel, the considered communication channel has the same capacity as that denoted in refs 10–14.

Besides a linear channel capacity defined and calculated by Shannon[1], only a few analytical results exist for a variety of different nonlinear channel models. We argue that our asymptotic capacity expression for a nonlinear channel may be valuable for practical channels and coding techniques developed for communication applications in order to approach the established linear Shannon limit, and deserves further extensive study. We here only consider a single correlation receiver for detecting the weak signal, however recent studies in general provide evidence that, besides an optimal noise intensity, an optimal network configuration exists, at which the best system response can be obtained[22,31,42–46]. Thus, an interesting extension for future work is to investigate the capacity of a very noisy communication channel with receivers connected in various network configurations.

Methods

Very noisy communication channel model. Consider a coherent M-ary communication channel transmitting M possible signals $\mathbf{S}_m$ for $m = 1, 2, \cdots, M$, as shown in Fig. 1. In an interval, the observation vector

$$\mathbf{X} = \mathbf{S}_m + \mathbf{Z}, \tag{30}$$

where $\mathbf{X} = [X_1, X_2, \cdots, X_N]^T$ contains the noise vector $\mathbf{Z} = [Z_1, Z_2, \cdots, Z_N]^T$ and the signal vector $\mathbf{S}_m = [S_{m1}, S_{m2}, \cdots, S_{mN}]^T$. Then, a receiver multiplies the transformation $g(\mathbf{X})$ with optimized coefficients, resulting in a cumulative statistic $T_m(\mathbf{X})$ for deciding whether the mth signal $\mathbf{S}_m$ is sent or not. The capacity C of a communication channel is given by the maximum of the mutual information $I(\mathbf{\Phi}, \mathbf{\Omega})$ between the input signal space $\mathbf{\Phi}$ and the channel output space $\mathbf{\Omega}$

$$C = \max_{f_{\mathbf{S}}} I(\mathbf{\Phi}, \mathbf{\Omega}), \tag{31}$$

where the maximization is with respect to the input distribution $f_{\mathbf{S}}$ over the signal space $\mathbf{\Phi}$[1–5].

Nirenberg's approach for white noise. The white noise $\mathbf{Z}$ has the multivariate distribution $f_{\mathbf{Z}}(\mathbf{z}) = \prod_{n=1}^{N} f_z(z_n)$ with zero-mean and variance σ_z^2. Let the statistically independent signal components be constrained to satisfy $s_{mn}^2 \leq \varepsilon_n$, and the total signal energy has a constraint $\sum_{n=1}^{N} s_{mn}^2 \leq \sum_{n=1}^{N} \varepsilon_n = \varepsilon$. Then, for very low SNR of $\varepsilon/\sigma_z^2 \ll 1$, the conditional probability density can be approximated as

$$f(\mathbf{x}|\mathbf{s}) = f_{\mathbf{Z}}(\mathbf{x} - \mathbf{s}_m) = \prod_{n=1}^{N} f_z(x_n - s_{mn}) = \prod_{n=1}^{N} [f_z(x_n) - s_{mn}f_z'(x_n)] = f_{\mathbf{Z}}(\mathbf{x})[1 + \gamma(\mathbf{s}_m, \mathbf{x})]$$

(32)

with the first two terms of Taylor series. Here, $f_z'(z) = df_z(z)/dz$ and the statistic $\gamma(\mathbf{s}_m, \mathbf{x}) = -\sum_{n=1}^{N} s_{mn}f_z'(x_n)/f_z(x_n)$[5]. Using the maximum likelihood rule[17], the conditional probability density on the knowledge that the mth signal satisfies

$$f(\mathbf{x}|\mathbf{s}_m) > f(\mathbf{x}|\mathbf{s}_k), \quad (k \neq m)$$

(33)

which leads to the optimum receiver

$$\gamma(\mathbf{s}_m, \mathbf{x}) > \gamma(\mathbf{s}_k, \mathbf{x}), \quad (k \neq m)$$

(34)

to decide if the mth signal was sent.

To simplify the mathematical manipulations, Nirenberg[5] assumes the even noise distribution function $f_z(z) = f_z(-z)$ and a very noisy channel ($|\gamma(\mathbf{s}_m, \mathbf{x})| \ll 1$)[2,8,9] yielding the mutual information between the output space $\mathbf{\Omega}$ and the input signal space $\mathbf{\Phi}$ as

$$
\begin{aligned}
I(\mathbf{\Phi}, \mathbf{\Omega}) &= \frac{1}{2}\iint f_{\mathbf{X}}(\mathbf{x})f_{\mathbf{S}}(\mathbf{s})\gamma^2(\mathbf{s}, \mathbf{x})d\mathbf{s}d\mathbf{x} - \frac{1}{2}\int f_{\mathbf{X}}(\mathbf{x})\left[\int f_{\mathbf{S}}(\mathbf{s})\gamma(\mathbf{s}, \mathbf{x})d\mathbf{s}\right]^2 d\mathbf{x} \\
&= \frac{1}{2}J(f_z)\sum_{n=1}^{N}[\mathrm{E}_s(s_{mn}^2) - \mathrm{E}_s^2(s_{mn})],
\end{aligned}
$$

(35)

where the Fisher information $J(f_z) = \mathrm{E}_z[f_z'^2(z)/f_z^2(z)] = \int f_z'^2(z)/f_z(z)dz$ of the noise density f_z and the expectation $\mathrm{E}_s(\cdot) = \int \cdot f_s(s)ds$. Since the same bias $\sum_{n=1}^{N} \mathrm{E}_s^2(s_{mn}) = \sum_{n=1}^{N} \mathrm{E}_s^2(s_{kn})$ does not affect the decision of inequality of equation (34), it may be conveniently assumed to be zero[5]. Then, over the class of signal distributions $f_{\mathbf{S}}$, the channel capacity is computed as[5]

$$C = \max_{f_{\mathbf{S}}} I(\mathbf{\Phi}, \mathbf{\Omega}) = \frac{1}{2}J(f_z)\sum_{n=1}^{N}\varepsilon_n = \frac{1}{2}J(f_z)\varepsilon,$$

(36)

which is applicable to various white noise types.

Furthermore, for a fixed noise variance $\mathrm{E}_z(z^2) = \sigma_z^2$ and an arbitrary noise density function f_z, $J(f_z) \geq \sigma_z^{-2}$[17], where the equality occurs for Gaussian distribution $f_z(z) = \exp(-z^2/\sigma_z^2)/\sqrt{2\pi\sigma_z^2}$ with its Fisher information $J(f_z) = 1/\sigma_z^2$[27,18,37]. Accordingly, the additive Gaussian noise channel is the worst one, and has the minimum capacity, as indicated in equation (36). It is well known that, for very low SNR of $\varepsilon/\sigma_z^2 \ll 1$, the capacity of Gaussian vector channel is approximately calculated as[2–4]

$$C_G = \sum_{n=1}^{N} \frac{1}{2}\ln\left(1 + \frac{\varepsilon_n}{\sigma_z^2}\right) \approx \frac{1}{2}\frac{\varepsilon}{\sigma_z^2},$$

(37)

which accords well with equation (36)[5].

References

1. Shannon, C. E. A mathematical theory of communication. *Bell System Technical Journal* **27,** 379–423 (1948).
2. Gallager, R. G. *Information Theory and Reliable Communication* (Wiley, New York, 1968).
3. Cover, T. M. & Thomas, J. A. *Elements of Information Theory* (Wiley, New York, 1991).
4. Yeung, R. W. *Information Theory and Network Coding* (Springer, New York, 2008).
5. Nirenberg, L. M. Low SNR digital communication over certain additive non-Gaussian channels. *IEEE Transactions on Communications* **23,** 332–341 (1975).
6. Kassam, S. A. *Signal Detection in Non-Gaussian Noise* (Springer-Verlag, New York, 1988).
7. Kay, S. *Fundamentals of Statistical Signal Processing* (Prentice-Hall, Englewood Cliffs, New Jersey, 1998).
8. Reiffen, B. A note on 'very noisy' channel. *Information and Control* **6,** 126–130 (1963).
9. Majani, E. E. *A Model for the Study of Very Noisy Channels and Applications* (PhD. Thesis, California Institute of Technology, California, 1988).
10. Kullback, S. *Information Theory and Statistics* (Dover, New York, 1968).
11. Verdú, S. On channel capacity per unit cost. *IEEE Transactions on Information Theory* **36,** 1019–1030 (1990).
12. Prelov, V. V. & van der Meulen, E. C. An asymptotic expression for the information and capacity of a multidimensional channel with weak input signals. *IEEE Transactions on Information Theory* **39,** 1728–1735 (1993).
13. Kostal, L. Information capacity in the weak-signal approximation. *Physical Review E* **82,** 026115 (2010).
14. Kostal, L. & Lansky, P. Information transfer for small-amplitude signals. *Physical Review E* **81,** 050901(R) (2010).
15. DiVincenzo, D. P., Shor, P. W. & Smolin, J. A. Quantum-channel capacity of very noisy channels. *Physical Review A* **57,** 830–838 (1998).
16. Abdel-Ghaffar, K. & McEliece, R. J. The ultimate limits of information density. *Proceeding of the NATO Advanced Study Institute on Performance Limits in Communication Theory and Practice*, Il Ciocco, Italy, **142,** 267–279 (1986).
17. Huber, P. J. *Robust Statistics* (Wiley, New York, 1981).

18. Stam, A. J. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Information and Control* **2,** 101–112 (1959).
19. Patel, A. & Kosko, B. Noise benefits in quantizer-array correlation detection and watermark decoding. *IEEE Transactions on Signal Processing* **59,** 488–505 (2011).
20. Benzi, R., Sutera, A. & Vulpiani, A. The mechanism of stochastic resonance. *Journal of Physics A: Mathematical and General* **14,** L453–L457 (1981).
21. Chapeau-Blondeau, F. & Godivier, X. Theory of stochastic resonance in signal transimission by static nonlinear systems. *Physical Review E* **55,** 1478–1495 (1997).
22. Collins, J. J., Chow, C. C. & Imhoff, T. T. Stochastic resonance without tuning. *Nature* **376,** 236–238 (1995).
23. Floriani, E., Mannella, R. & Grigolini, P. Noise-induced transition from anomalous to ordinary diffusion: The crossover time as a function of noise intensity. *Physical Review E* **52,** 5910–5917 (1995).
24. Bulsara, A. R. & Zador, A. Threshold detection of wideband signals: A noise-induced maximum in the mutual inforamtion. *Physical Review E* **54,** R2185–R2188 (1996).
25. Greenwood, P. E., Ward, L. M. & Wefelmeyer, W. Statistical analysis of stochastic resonance in a simple setting. *Physical Review E* **60,** 4687–4695 (1999).
26. Duan, F., Chapeau-Blondeau, F. & Abbott, D. Fisher information as a metric of locally optimal processing and stochastic resonance. *PLoS One* **7,** e34282 (2012).
27. Gammaitoni, L., Hänggi, P., Jung, P. & Marchesoni, F. Stochastic resonance. *Reviews of Modern Physics* **70,** 233–287 (1998).
28. Kay, S. Can detectabilty be improved by adding noise? *IEEE Signal Processing Letters* **7,** 8–10 (2000).
29. McDonnell, M. D., Stocks, N. G., Pearce, C. E. M. & Abbott, D. *Stochastic Resonance: From Suprathreshold Stochastic Resonance to Stochastic Signal Quantization* (Cambridge University Press, Cambridge, 2008).
30. Moss, F., Ward, L. M. & Sannita, W. G. Stochastic resonance and sensory information processing: A tutorial and review of application. *Clinical NeuroPhysiology* **115,** 267–281 (2004).
31. Stocks, N. G. Suprathreshold stochastic resonance in multilevel threshold systems. *Physical Review Letters* **84,** 2310–2313 (2000).
32. Martignoli, S., Gomez, F. & Stoop, R. Pitch sensation involves stochastic resonance. *Scientific Reports* **3,** 2676 (2013).
33. Zozor, S. & Amblard, P. O. Stochastic resonance in locally optimal detectors. *IEEE Transactions on Signal Processing* **51,** 3177–3181 (2003).
34. Duan, F., Chapeau-Blondeau, F. & Abbott, D. Stochastic resonance with coloured noise for neural signal detection. *PLoS One* **9,** e91345 (2014).
35. Czaplicka, A., Holyst, J. A. & Sloot, P. M. A. Noise enhances information transfer in hierarchical networks. *Scientific Reports* **3,** 1223 (2013).
36. Han, J., Liu, H., Sun, Q. & Huang, N. Reconstruction of pulse noisy images via stochastic resonance. *Scientific Reports* **5,** 10616 (2015).
37. Blachman, N. M. The convolution inequality for entropy power. *IEEE Transactions on Information Theory* **IT-11,** 267–271 (1965).
38. Dembo, A. Simple proof of the concavity of the entropy power with respect to added Gaussian noise. *IEEE Transactions on Information Theory* **35,** 887–888 (1989).
39. Dembo, A., Cover, T. M. & Thomas, J. A. Information theoretic inequalities. *IEEE Transactions on Information Theory* **37,** 1501–1518 (1991).
40. Zamir, R. A proof of the Fisher information inequality via a data processing argument. *IEEE Transactions on Information Theory* **44,** 1246–1250 (1998).
41. Poor, H. V. Signal detection in the presence of weakly dependent noise—Part I: Optimum detection. *IEEE Transactions on Information Theory* **28,** 735–744 (1982).
42. Perc, M. Stochastic resonance on excitable small-world networks via a pacemaker. *Physical Review E* **76,** no. 066203 (2007).
43. Benzi, R. Stochastic resonance in complex systems. *Journal of Statistical Mechanics: Theory and Experiment* **1,** P01052 (2009).
44. Perc, M. Stochastic resonance on weakly paced scale-free networks. *Physical Review E* **78,** 036105 (2008).
45. Wang, Q., Perc, M., Duan, Z. & Chen, G. Delay-induced multiple stochastic resonances on scale-free neuronal networks. *Chaos* **19,** 023112 (2009).
46. Gan, C., Perc, M. & Wang, Q. Delay-aided stochastic multiresonances on scale-free FitzHugh-Nagumo neuronal networks. *Chinese Physical B* **19,** no. 040508 (2010).

Acknowledgements

Author Contributions

All authors performed the theoretical analyses and participated in the writing of the manuscript. F.D. performed the experiments.

Additional Information

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Duan, F. *et al.* Capacity of very noisy communication channels based on Fisher information. *Sci. Rep.* **6,** 27946; doi: 10.1038/srep27946 (2016).