Capacity of very noisy communication channels based on Fisher information

Fabing Duan1, François Chapeau-Blondeau2 & Derek Abbott3

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 calculated1–4. In practical applications, however, communication systems frequently encounter non-Gaussian noise environments, for instance, underwater acoustic noise and low-frequency atmospheric noise5–7. Of all channels with power-constrained noise, the capacity of a Gaussian channel is the smallest1–4. Thus, the capacities of non-Gaussian channels are of great interest5–15. 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 channels2,9 and qubit depolarizing channels15. A very noisy channel was introduced by Reiffen8, and extended by Gallager2 and Majani9 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 nonzero1–4,9,16. Following the approaches developed in2,8,9 and using a power series of characteristic functions, Nirenberg5 derived a simple formula of the capacity for the coherent threshold channel with an optimum receiver. For memoryless channels with very weak inputs, Kullback10, Verdú11 and Prelov12 explicitly expressed the asymptotic expressions of the channel capacity closely related to the Fisher information matrix. Recently, Kostal and Lansky14 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 way13,14.

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 \( S_m, m = 1, 2, \ldots, M \). With the assumptions of white noise and very low SNR, Nirenberg5 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
idealized assumption of white noise is unpractical, and the coloured noise has practical significance. 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

\[ f(x|s_m) = f_Z(x - s_m) \approx f_Z(x) - (\nabla f_Z(x))^T s_m = f_Z(x)[1 + \Gamma(s_m, x)], \]  

(1)

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

\[ f(x|s_m) > f(x|s_k), \quad (k \neq m) \]  

(2)

to optimally choose \( m \)th signal. Substituting equation (1) into equation (2), the optimum receiver

\[ \Gamma(s_m, x) > \Gamma(s_k, x), \quad (k \neq m) \]  

(3)

enables us to decide if the \( m \)th signal was transmitted. For clarity, we state that the statistic \( \Gamma(s_m, x) \) and the maximum likelihood decoding rule of equation (2) compose an optimum correlation receiver. The channel output is the decoding signal \( \omega_m \) of the receiver, as shown in Fig. 1.

Figure 1. Mutual information \( I(\Phi, \Omega) \) of the communication channel and \( I(\Phi, \Psi) \) of the nonlinear channel.

Then, supposing the zero-mean \( E(s) = 0 \) and extending the very noisy vector channel \( |\Gamma(s_m, x)| \ll 1 \), the mutual information between the input signal space \( \Phi \) and the channel output space \( \Omega \) is given by

\[
I(\Phi, \Omega) = \frac{1}{2} \int \int f_X(x)f_S(s)\Gamma(s, x)^T\Gamma(s, x)dsdx \approx \frac{1}{2} \int f_X(x) \left[ \int f_S(s)\Gamma(s, x)ds \right]^2 dx \approx \frac{1}{2}E_s[s^T J(f_Z) s], \]

(4)

where the Fisher information matrix of the noise distribution is defined as

\[
J(f_Z) = E_s[(\nabla \ln f_Z(z)) / (\nabla \ln f_Z(z))^T]. \]

(5)

It is noted that \( J(f_Z) \) is also called the Fisher information of a location parameter or the shift-invariant Fisher information, 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 \( E_s(s^T s) \leq \varepsilon \) and for the standardized vector \( u = s/\sqrt{E_s(s^T s)} \), the channel capacity can be expressed as

\[
C = \max_{f_Z} I(\Phi, \Omega) = \max_{f_Z} \frac{1}{2}E_s(s^T s)E_s[u^T J(f_Z) u] = \frac{1}{2} \varepsilon \Lambda, \]

(6)

where \( \Lambda \) is the largest eigenvalue of the matrix \( J(f_Z) \) and \( u \) takes the corresponding eigenvector.

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

\[
E_Z \left[ (\nabla \ln f_Z(z) - \sum_z \frac{1}{Z}z)^T (\nabla \ln f_Z(z) - \sum_z \frac{1}{Z}z) \right] \succeq 0
\]

and the noise covariance matrix \( \sum_Z = E_Z(ZZ^T) \), we have

\[
J(f_Z) \geq E_Z[(\nabla \ln f_Z(z))^T Z] \sum_z \frac{1}{Z} E_Z[(\nabla \ln f_Z(z))^T Z] \succeq \sum_z \frac{1}{Z},
\]

(7)

where the equality occurs for \( N \)-dimensional Gaussian distribution \( f_Z(z) = \exp(-x^T \sum_z \frac{1}{Z}z^2/2)/(2\pi)^{N/2} \text{det}(\sum_Z)^{1/2}) \) with its Fisher information matrix \( J(f_Z) = \sum_z \frac{1}{Z} \). Thus, equation (7) indicates the maximum eigenvalue of \( \sum_z \frac{1}{Z} \) is
less than that of Fisher information matrix of non-Gaussian noise. This result extends the conclusion of equation (36) by Nirenberg\textsuperscript{2}, 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\textsuperscript{1,4}. 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. \( \alpha \)-stable noise\textsuperscript{19}). Thus, we further consider the generalized correlation receiver
\[
T_m = c_m^T g(x) > T_k = c_k^T g(x), \quad (k \neq m)
\]
where the coefficient vector \( c_m = [c_{m1}, c_{m2}, \ldots, c_{mN}]^T \) and the function \( g(x) \) is not restricted to be memoryless. For the zero-mean vector of \( E_Z[g(z)] = 0 \) (for a shift in mean\textsuperscript{6} under \( f_Z \) and for very low SNR, \( g(x) \) can be expanded to the first-order
\[
g(x) = g(z + s_m) \approx g(z) + (\nabla g(z))^T s_m.
\]
Then, for a large observation size \( N \), the statistic \( T_m \) has the mean \( E_{Z_m}(T_m) \approx c_m^T E_Z[ (\nabla g(z))^T ] s_m \) and the variance \( \text{var}(T_m) \approx c_m^T E_Z[ (\nabla g(z))^T ] c_m \). Using the Cholesky decomposition of the symmetrical matrix \( V = E_Z[g(z)g(z)^T] = LL^T \), the output SNR of the receiver can be calculated as
\[
\frac{E_Z^2(T_m)}{\text{var}(T_m)} = \max_{s_m} \frac{c_m^T E_Z[ (\nabla g(z))^T ] s_m^T}{c_m^T E_Z[ g(z)g(z)^T ] c_m} = \max_{s_m} \frac{[(Lc)^T (L^{-1} E_Z[ (\nabla g(z))^T ] s_m)]^T}{(Lc)^T (L^{-1} E_Z[ (\nabla g(z))^T ] c_m)} = s_m^T E_Z[ (\nabla g(z))^T ] L^{-1} E_Z[ (\nabla g(z))^T ] s_m,
\]
by optimally choosing \( c_m = V^{-1} E_Z[ (\nabla g(z))^T ] s \). 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
\[
C_x = \frac{1}{2} \ln \left( 1 + \frac{E_Z^2(T_m)}{\text{var}(T_m)} \right) \approx \frac{1}{2} \ln \left( \frac{1}{f_s} \right) = \frac{1}{2} \ln \left( 1 + \frac{1}{f_s} \right),
\]
where \( \Lambda_x \) is the largest eigenvalue of the matrix \( E_Z[ (\nabla g(z))^T ] V^{-1} E_Z[ (\nabla g(z))^T ] \). Observing
\[
E_Z[ (\nabla g(z)) ] = E_Z[ ( - \nabla \ln f_Z(z) ) g(z)^T ],
\]
and for the positive semidefinite matrix
\[
E_Z[ ( - \nabla \ln f_Z(z) ) - A \nabla \ln f_Z(z) ] \geq 0,
\]
we have
\[
J(f_Z) \geq E_Z[ (\nabla g(z))^T ] V^{-1} E_Z[ (\nabla g(z))^T ],
\]
with \( A = E_Z[ ( - \nabla \ln f_Z(z) ) g(z)^T ] V^{-1} g(z) \) and the equality occurring for \( g(x) = - \nabla \ln f_Z(x) \). This inequality (13) indicates that the eigenvalue \( \Lambda_x \) of \( J(f_Z) \) is not less than the eigenvalue \( \Lambda_x \) of the matrix \( E_Z[ (\nabla g(z))^T ] V^{-1} E_Z[ (\nabla g(z))^T ] \). Therefore, based on equations (6), (11) and (13), we find
\[
C_x \leq C,
\]
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 \( g(x) = - \nabla \ln f_Z(x) \), 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 \( g(x) \), the expectation matrices \( E_Z[ (\nabla g(z))^T ] = E_Z[ g(z)^T ] I \) and \( V = E_Z[ g(z)^2 ] I \). Here, the derivative \( g'(z) = dg(z)/dz \) and I is the unit matrix. Therefore, the matrix \( E_Z[ (\nabla g(z))^T ] V^{-1} E_Z[ (\nabla g(z))^T ] \) in equation (11) has \( N \) identical eigenvalues \( \Lambda_x \). In equation (11) is not less than the eigenvalue \( \Lambda_x \) of the matrix
\[
C_x \approx \frac{1}{2} \epsilon \Lambda_x \leq \frac{1}{2} \epsilon \frac{E_Z^2[ g'(z)^2 ]}{E_Z^2[ g(z)^2 ]} \leq C = \frac{1}{2} \epsilon \Lambda = \frac{1}{2} \epsilon J(f_Z),
\]
where the eigenvalue \( \Lambda = J(f_Z) \) corresponds to the Fisher information matrix \( J(f_Z) \) in equation (6). Using the Cauchy-Schwarz inequality and integration by parts \( E_Z^2[ g'(z)^2 ] = E_Z^2[ g(z) (-f'_Z(z)/f_Z(z)) ] \leq E_Z[ g(z)^2 ] J(f_Z) \).
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.

**Conditions for noise-enhanced capacity.** Since the emergence of the concept of stochastic resonance\(^8\), the employment of noise in enhancing the performance of nonlinear systems has become an interesting option\(^3,13,14,21–36\). 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\(^26,37\). Later, the notion of stochastic resonance has been widened to a number of different mechanisms, e.g. aperiodic stochastic resonance\(^32\) 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 \( Z = DZ_m \), the covariance matrix \( \Sigma_Z \) can be factored as \( \Sigma_Z = DD^T \) and the standardized noise vector \( Z_m \) has a covariance matrix being the unit matrix \( E(Z_m Z_m^T) = \Gamma \). A well-known scaling property of the Fisher information matrix is\(^13,18,37–40\)

\[
J(f_{z_m}) = D^{-1}J(f_{z_m})(D^T)^{-1},
\]

which implies the largest eigenvalue \( \Lambda \) of \( J(f_{z_m}) \) is a monotonically decreasing function of \( \Lambda = \det(\Sigma_Z) \) for the determinants \( \det(D) = \det(D^T) = \det(\Sigma_Z) \). Here, the largest eigenvalue of \( J(f_{z_m}) \) is \( \Lambda = \) a fixed quantity for \( Z_m \). For such a channel with its optimum receiver, equation (11) indicates the channel capacity \( C \approx = \frac{1}{2} \epsilon \Lambda \) 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 \( g(x) = \text{sign}(x) \) 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 \( g(x) = \text{sign}(x) = -f_z'(z)/f_z(z) \) satisfies the equality condition in equation (15). In this case, the channel capacity in equation (15) can be calculated as \( C \approx = \frac{1}{2} \epsilon f_z(x) = \epsilon/\sigma^2 \), which monotonically decreases as the noise level \( \epsilon \) 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 \( W \) independent of \( Z \) and \( S_m \) to the observation \( X \), and the updated data vector is

\[
\tilde{X} = Z + W + S_m = U + S_m,
\]

where the composite noise vector \( U = Z + W \) with its distribution \( f_U \). In this case, we should employ the statistic \( \Gamma \langle s_m, x \rangle = (\nabla \ln f_U(\tilde{x}))^T s_m \) to specify the optimum receiver, and the corresponding capacity is then given by

\[
\tilde{C} = \frac{1}{2} \epsilon \Lambda
\]

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

\[
AJ(f_{z_m})A^T + (I - A)J(f_w)(I - A)^T \geq J(f_U),
\]

we then find the largest eigenvalue \( \Lambda \) of \( J(f_{z_m}) \) is not less than the largest eigenvalue \( \Lambda \) of \( J(f_U) \) and

\[
\tilde{C} \leq C.
\]

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 \( W \).

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 \( W \) with its distribution\(^8,21,28,33\)
as an independent identically distributed \( \mathbf{w} \), the dependence among noise samples is then calculated \( \mathbf{g} \), and the matrix \( \mathbf{g}_{12} \). In Fig. 2, we show the capacity per signal energy \( \frac{C}{\varepsilon} \), and other elements are higher-order infinitesimal of \( \mu \), 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 \( f(x) \) does not have a tractable analytic expression. Using the approach developed in ref. 41, we have the expectation matrix

\[ E_W[\nabla g(z)] \approx E_W[g'(w)] \mathbf{I} \quad (23) \]

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

\[ V_{i,i} = E_W[g^2(w)], \quad V_{i,i+1} = V_{i+1,i} = \left( \rho_1 + \rho_2 \right) E_W[wg(w)] E_W[g'(w)], \quad (24) \]

for \( i = 1, 2, \ldots, N \), and other elements are higher-order infinitesimals of \( \rho_1 + \rho_2 (|\rho_1| < 1) \). Then, we calculate the largest eigenvalue of the matrix \( \mathbf{E}_W[\nabla g(z)] \mathbf{V}^{-1} \mathbf{E}_W[\nabla g(z)]^T \) as

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

\[ = \frac{2}{\pi \xi} \exp \left(-\frac{\mu^2}{\xi}\right) \left[ 1 - 2[\rho_1 + \rho_2]^{2} \exp \left(-\frac{\mu^2}{\xi}\right) + \sqrt{\frac{2}{\pi \xi}} \exp \left(-\frac{\mu^2}{2 \xi}\right) \text{erf} \left(\frac{\mu}{\sqrt{2 \xi}}\right) \right], \quad (26) \]

where the error function \( \text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt \). In Fig. 2, we show the capacity per signal energy \( \frac{C}{\varepsilon} = \Lambda_g/2 \) in equation (11) versus the noise parameters \( \mu \) and \( \zeta \) 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 \( \zeta \) as the noise level. It is then clearly shown in that Fig. 2, upon increasing \( \zeta \) for a fixed value of \( \mu \) (the noise variance \( \sigma^2 \) also increases), the noise-enhanced capacity effects exist. The corresponding maxima of \( \frac{C}{\varepsilon} \) versus optimal values of \( \zeta \) 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

\[ f_w(x) = \frac{1}{2\sqrt{2\pi\zeta}} \left[ \exp \left(-\frac{(x - \mu)^2}{2\zeta^2}\right) + \exp \left(-\frac{(x + \mu)^2}{2\zeta^2}\right) \right], \quad (21) \]

where the variance \( \sigma^2 = \mu^2 + \zeta^2 \) and parameters \( \mu, \zeta \geq 0 \). A useful coloured noise model of the first-order moving-average as

\[ Z_n = \rho_1 W_{n-1} + W_n + \rho_2 W_{n+1}, \quad (22) \]

where the correlation coefficients are \( \rho_{1,2} \) and \( \mathbf{W} = [W_1, W_2, \ldots, W_N]^T \) 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. The signum function \( sgn(x) \) is adopted to construct the generalized correlation receiver of equation (8), which is not for the coloured noise \( \mathbf{Z} \).
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\(^\text{21,24,25,28-31}\). For instance, the noise-enhanced effect has been demonstrated for non-weak signals in threshold neurons\(^\text{25,29,31}\), where an optimal matching condition is incapable 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\(^\text{10,11}\) or memory additive-noise channels\(^\text{12-14}\). We emphasize the asymptotic capacity expressions of equations (6) and (11) are different from those in previous literature\(^\text{10-14}\). In Fig. 1, for the channel output \(Y = g(X)\), these studies assume the conditional probability density as \(f_{YS}(y|s)\). Then, the Fisher information matrix is defined as\(^\text{10-14}\)

\[
J(f_{YS}) = E_{YS}[(\nabla_s \ln f_{YS}(y|s))(\nabla_s \ln f_{YS}(y|s))^T],
\]

with the operator \(\nabla_s \triangleq \left[ \frac{\partial}{\partial s_1}, \frac{\partial}{\partial s_2}, \ldots, \frac{\partial}{\partial s_N} \right]^T\). Then, for the zero-mean signal vector \(E_s = 0\) and the weak signal energy \(\varepsilon\), the mutual information between the input space \(\Phi\) and the output space \(\Psi\) is approximated as\(^\text{10-14}\)

\[
I(\Phi, \Psi) \approx \frac{1}{2}E_{s'}[J(f_{YS})]_{s \rightarrow 0},
\]

which is different from the mutual information \(I(\Phi, \Omega)\) of equation (4) based on the Fisher information matrix \(J(f_z)\) of the noise distribution \(f_z\). It is shown in Fig. 1 that the receiver multiplies nonlinear transformation \(g(x)\) with optimized coefficients, and obtains a cumulative statistic \(T_m\) that decides whether the \(m\)th signal \(S_m\) is sent or not. Then, the considered communication channel chooses an optimal signal \(S_m\) from the signal space to maximize the average mutual information. Since the receiver collecting the weighted nonlinear outputs as the statistic \(T_m = c_m g(x)\), 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).

**Methods**

**Very noisy communication channel model.** Consider a coherent \(M\)-ary communication channel transmitting \(M\) possible signals \(S_m\) for \(m = 1, 2, \ldots, M\), as shown in Fig. 1. In an interval, the observation vector

\[
X = S_m + Z,
\]

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

\[
C = \max_I(\Phi, \Omega),
\]

where the maximization is with respect to the input distribution \(f_S\) over the signal space \(\Phi\).
Nirenberg’s approach for white noise. The white noise Z has the multivariate distribution
\[ f_Z(z) = \prod_{n=1}^{N} f_{Z_n}(z_n) \] with zero-mean and variance \( \sigma_n^2 \). Let the statistically independent signal components be
constrained to satisfy \( s_{mn} \leq \varepsilon_n \) and the total signal energy has a constraint \( \sum_{n=1}^{N} s_{mn}^2 \leq \varepsilon \). Then, for
very low SNR of \( \varepsilon / \sigma_n^2 \ll 1 \), the conditional probability density can be approximated as
\[
f(x|s_m) = f_Z(x - s_m) = \prod_{n=1}^{N} f_{Z_n}(x_n - s_{mn}) = \prod_{n=1}^{N} [f_{Z_n}(x_n) - s_{mn} f'_{Z_n}(x_n)] = f_Z(x)[1 + \gamma(s_m, x)]
\]
with the first two terms of Taylor series. Here, \( f'_z(z) = df_z(z)/dz \) and the statistic \( \gamma(s_m, x) = -\sum_{n=1}^{N} s_{mn} f'_{Z_n}(x_n)/f_z(x_n)^2 \). Using the maximum likelihood rule\(^7\), the conditional probability density on the knowl-
edge that the mth signal satisfies
\[
f(x|s_m) > f(x|s_k), \quad (k = m)
\]
which leads to the optimum receiver
\[
\gamma(s_m, x) > \gamma(s_k, x), \quad (k = m)
\]
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 \( (|\gamma(s_m, x)| \ll 1)^{2,9} \) yielding the mutual information between the output
space \( \Phi \) and the input signal space \( \Omega \) as
\[
I(\Phi, \Omega) = \frac{1}{2} \int f_Z(x)f_z(s) \gamma^2(s, x) ds dx - \frac{1}{2} \int f_Z(x) \int f_z(s) \gamma(s, x) ds^2 dx
\]
\[
= \frac{1}{2} J(f_z) \sum_{n=1}^{N} [E_z(s_{mn}) - E_z^2(s_{mn})],
\]
where the Fisher information \( J(f_z) = E_z[f_z^2(z)/f_z^2(z)] = \int [f_z^2(z)/f_z(z)] dz \) of the noise density \( f_z \) and the expectation \( E_z(\cdot) = \int f_z(z) ds \). Since the same bias \( \sum_{n=1}^{N} E_z(s_{mn}) = \sum_{n=1}^{N} E_z^2(s_{mn}) \) does not affect the decision of ine-
quality of equation (34), it may be conveniently assumed to be zero\(^6\). Then, over the class of signal distributions \( f_z \),
the channel capacity is computed as\(^5\)
\[
C = \max_{f_z} I(\Phi, \Omega) = \frac{1}{2} J(f_z) \sum_{n=1}^{N} \varepsilon_n = \frac{1}{2} J(f_z) \varepsilon,
\]
which is applicable to various white noise types.

Furthermore, for a fixed noise variance \( E_z(\varepsilon^2) = \sigma_z^2 \) and an arbitrary noise density function \( f_z \), \( J(f_z) \geq \sigma_z^{-2}\)\(^17\),
where the equality occurs for Gaussian distribution \( f_z(z) = \exp(-z^2/2\sigma_z^2)/\sqrt{2\pi\sigma_z^2} \) with its Fisher information
\( J(f_z) = 1/\sigma_z^{-2}\)\(^27,3,37\). 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\(^2,4\)
\[
C_G = \sum_{n=1}^{N} \frac{1}{2} \ln \left( 1 + \frac{\varepsilon_n}{\sigma_z} \right) \approx \frac{1}{2} \frac{\varepsilon}{\sigma_z^2},
\]
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, 322–341 (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. Reifen, B. A note on ‘very noisy’ channel. Information and Control 6, 126–130 (1963).
9. Majani, E. A. 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, 061105 (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, II Ciocco, Italy, 142, 267–279 (1986).
17. Huber, P. J. Robust Statistics (Wiley, New York, 1981).
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).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/