Robust stochastic resonance for simple threshold neurons

Bart Kosko[1] and Sanya Mitaim[2]
[1]*Department of Electrical Engineering, Signal and Image Processing Institute, University of Southern California, Los Angeles, California 90089-2564, USA*
[2]*Department of Electrical Engineering, Faculty of Engineering, Thammasat University, Pathumthani 12120, Thailand*
(Received 27 January 2003; revised manuscript received 6 May 2004; published 27 September 2004)

Simulation and theoretical results show that memoryless threshold neurons benefit from small amounts of almost all types of additive noise and so produce the stochastic-resonance or SR effect. Input-output mutual information measures the performance of such threshold systems that use subthreshold signals. The SR result holds for all possible noise probability density functions with finite variance. The only constraint is that the noise mean must fall outside a "forbidden" threshold-related interval that the user can control—a new theorem shows that this condition is also necessary. A corollary and simulations show that the SR effect occurs for *right*-sided beta and Weibull noise as well. These SR results further hold for the entire uncountably infinite class of alpha-stable probability density functions. Alpha-stable noise densities have infinite variance and infinite higher-order moments and often model impulsive noise environments. The stable noise densities include the special case of symmetric bell-curve densities with thick tails such as the Cauchy probability density. The SR result for alpha-stable noise densities shows that the SR effect in threshold and thresholdlike systems is robust against occasional or even frequent violent fluctuations in noise. Regression analysis reveals both an exponential relationship for the optimal noise dispersion as a function of the alpha bell-curve tail thickness and an approximate linear relationship for the SR-maximal mutual information as a function of the alpha bell-curve tail thickness.

DOI: 10.1103/PhysRevE.70.031911 PACS number(s): 87.10.+e

I. ALMOST ALL THRESHOLD SYSTEMS EXHIBIT STOCHASTIC RESONANCE

Several researchers have found that threshold neurons and other threshold systems exhibit stochastic resonance [1–8]: Small amounts of noise improve the threshold neuron's input-output correlation measure [9,10] or mutual information [1,8,11]. All of these simulations and analyses assume a noise probability density function that has finite variance. Most further assume that the noise is simply Gaussian or uniform. Yet the statistics of real-world noise can differ substantially from these simple and finite-variance probability descriptions. The noise can be impulsive and irregular and have infinite variance and infinite higher-order moments. Computer simulations alone cannot decide whether this uncountable class of noise densities produces the SR effect in threshold systems. Theoretical techniques can decide the issue and we show that the answer is positive: Almost all threshold systems exhibit the SR effect in terms of mutual information or a bit-based measure of system performance.

The two theorems in [12] show the SR effect in simple (memoryless) threshold neurons as often found in the literature of neural networks [13–15]. We state these two theorems below (Theorems 1.1 and 2.1) without proof and derive a corollary and two new related theorems. The first theorem (Theorem 1.1) shows that threshold neurons exhibit the SR effect for all finite-variance noise densities if the system performance measure is Shannon's mutual information and if the mean or location parameter falls outside a "forbidden" interval that one can often pick in advance. A corollary shows that this SR effect still occurs for right-sided beta and Weibull noise. Traditional SR research has focused almost exclusively on two-sided noise. The second theorem (Theorem 2.1) shows that this also holds for all infinite-variance densities that belong to the large class of stable distributions. Both theorems assume that all signals are subthreshold signals. The two new theorems (Theorems 1.2 and 2.2) show that there is no SR effect if the mean or location parameters fall within the forbidden threshold interval. Figure 4 shows a simulation instance of this predicted forbidden-interval effect for Gaussian and Cauchy noise.

The paper then presents several regression analyses of simulation experiments that confirm and extend the exponential relationship between the optimal noise dispersion and alpha bell-curve tail thickness [16]. This exponential relationship corresponds to a similar one for infinite-variance SR systems that use a signal-to-noise ratio or a cross correlation for the system performance measure [16]. Regression also shows that the SR-maximal mutual information in noisy threshold neurons depends approximately linearly on the bell-curve tail thickness for symmetric alpha-stable noise.

Figure 1 shows the system-flow diagram of a noisy threshold neuron system that processes subthreshold signals. Figure 2 shows the first use of (right-sided) beta noise for

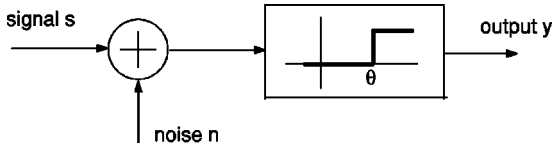

FIG. 1. System-flow diagram of a noisy threshold neuron. The neuron's signal function has the form (1) with threshold parameter θ, where s is the input signal and n is the input additive noise. We assume subthreshold signals: $A < \theta$, where A is the amplitude of the Bernoulli input s.

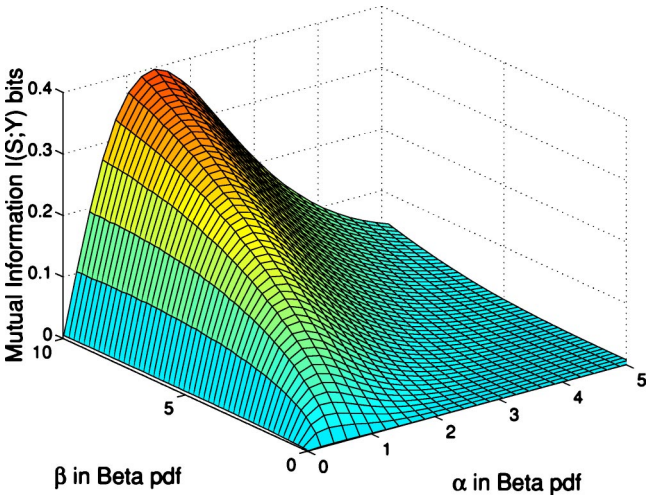

FIG. 2. (Color online) Stochastic resonance with (right-sided) beta noise. The noisy signal-forced threshold neuron has the form (1). The beta noise n_t adds to the bipolar input Bernoulli signal s_t. The parametrized interval $[a, b]$ of the beta density (14) has $a=0$ and $b=10$. The neuron has threshold $\theta=1$. The input Bernoulli signal has amplitude $A=0.8$ with success probability $p=\frac{1}{2}$. Each trial produced 10 000 input-output samples $\{s_t,y_t\}$ that estimated the probability densities to obtain the mutual information. The graph shows the smoothed input-output mutual information of a threshold neuron as a function of the parameters α and β of additive white beta noise n_t. The neuron's mutual information has a nonzero noise optimum $\sigma_{\text{opt}}>0$ where the variance has the form $\sigma_n^2=[(b-a)^2\alpha\beta]/[(\alpha+\beta)^2(\alpha+\beta+1)]$.

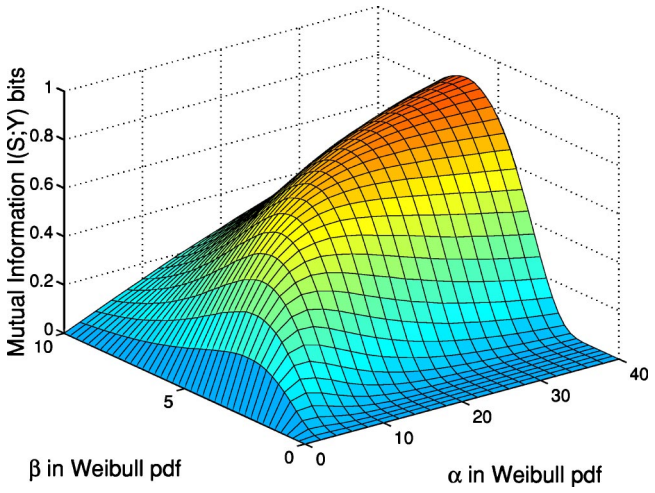

FIG. 3. (Color online) Stochastic resonance with (right-sided) Weibull noise. The noisy signal-forced threshold neuron has the form (1). The Weibull noise n_t adds to the bipolar input Bernoulli signal s_t. The neuron has threshold $\theta=0.5$. The input Bernoulli signal has amplitude $A=0.2$ with success probability $p=\frac{1}{2}$. Each trial produced 10 000 input-output samples $\{s_t,y_t\}$ that estimated the probability densities to obtain the mutual information. The graph shows the smoothed input-output mutual information of a threshold neuron as a function of the parameters α and β of additive white Weibull noise n_t. The neuron's mutual information has a nonzero noise optimum $\sigma_{\text{opt}}>0$ where the variance has the form $\sigma_n^2=(\beta/\alpha)^{2/\beta}[\Gamma(1+2/\beta)-\{\Gamma(1+1/\beta)\}^2]$.

SR. The beta density generalizes the uniform density and is popular in Bayesian statistics [17] because it allows analysts to control the shape of the density with two parameters and scale or translate the finite-length domain. Figure 3 shows the first use of (right-sided) Weibull noise for SR. The Weibull density generalizes the exponential and Rayleigh densities and has an infinite-length domain. Figure 4 shows several symmetrical alpha-stable noise densities whose bell curves have thick tails that produce infinite variance and often highly impulsive noise spikes. Figure 5 shows a simulation instance of both Theorem 2.1 and the empirical trends in Figs. 7 and 8. Infinite-variance Cauchy noise produces the SR effect when plotted against the Shannon mutual information of the threshold system. The linear regression results in Table I and Fig. 7 reveal the exponential relationship between the optimal noise dispersion and the alpha bell-curve tail thickness. The linear dependence of the log-transformed optimal noise dispersion on the bell-curve thickness becomes quadratic when the signal amplitude is too small or too close to the neuron's threshold. The regression results in Table II and Fig. 8 show a similar pattern. The linear dependence of the SR-maximal mutual information on the bell-curve thickness also becomes quadratic when the signal amplitude is too small or too close to the neuron's threshold.

II. THRESHOLD NEURONS AND SHANNON'S MUTUAL INFORMATION

We use the standard discrete-time threshold neuron model [1,2,6,15,16,18]

$$y_t = \text{sgn}(s_t + n_t - \theta) = \begin{cases} 1 & \text{if } s_t + n_t \geq \theta \\ 0 & \text{if } s_t + n_t < \theta \end{cases}, \tag{1}$$

where $\theta>0$ is the neuron's threshold, s_t is the bipolar input Bernoulli signal (with arbitrary success probability p such that $0<p<1$) with amplitude $A>0$, and n_t is the additive white noise with probability density $p(n)$. Figure 1 shows the system flow of the threshold system.

The threshold system uses subthreshold binary signals. The symbol "0" denotes the input signal $s=-A$ and output signal $y=0$. The symbol "1" denotes the input signal $s=A$ and output signal $y=1$. We assume subthreshold input signals: $A<\theta$. Then the conditional probabilities $P_{Y|S}(y|s)$ are

$$P_{Y|S}(0|0) = \Pr\{s + n < \theta\}_{s=-A} \tag{2}$$

$$= \Pr\{n < \theta + A\}$$

$$= \int_{-\infty}^{\theta+A} p(n)dn \tag{3}$$

$$P_{Y|S}(1|0) = 1 - P_{Y|S}(0|0) \tag{4}$$

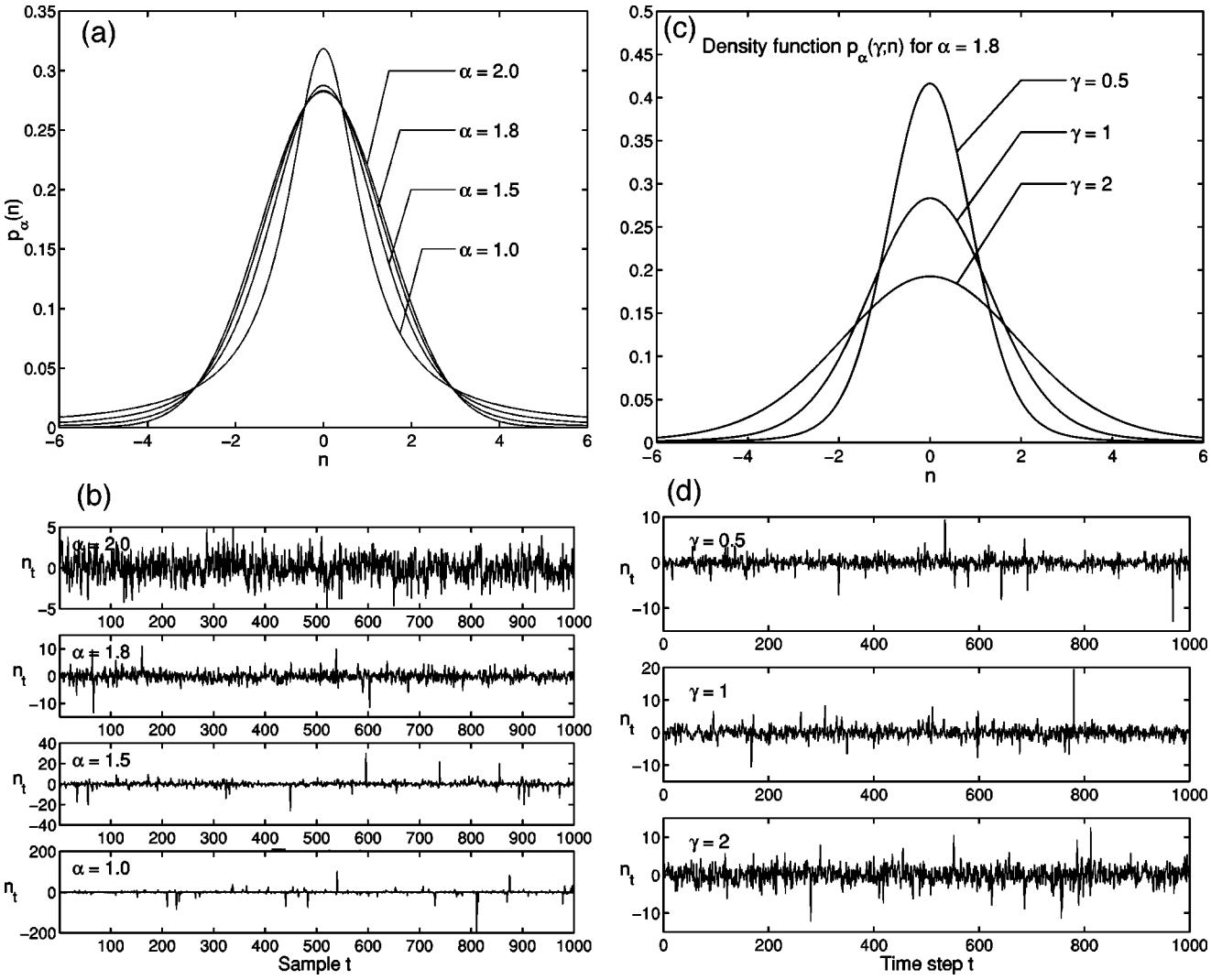

FIG. 4. Samples of standard symmetric alpha-stable probability densities and their realizations. (a) Density functions with zero location ($a=0$) and unit dispersion ($\gamma=1$) for $\alpha=2, 1.8, 1.5$, and 1. The densities are bell curves that have thicker tails as α decreases and thus that model increasingly impulsive noise as α decreases. The case $\alpha=2$ gives a Gaussian density with variance 2 (or unit dispersion). The parameter $\alpha=1$ gives the Cauchy density with infinite variance. (b) Samples of alpha-stable random variables with zero location and unit dispersion. The plots show realizations when $\alpha=2$, 1.8, 1.5, and 1. Note the scale differences on the y axes. The alpha-stable variable n becomes more impulsive as the parameter α falls. The algorithm in [39,40] generated these realizations. (c) Density functions for $\alpha=1.8$ with dispersions $\gamma=0.5$, 1, and 2. (d) Samples of alpha-stable noise n for $\alpha=1.8$ with dispersions $\gamma=0.5$, 1, and 2.

$$P_{Y|S}(0|1) = \Pr\{s + n < \theta\}_{s=A} \tag{5}$$

$$= \Pr\{n < \theta - A\}$$

$$= \int_{-\infty}^{\theta - A} p(n)dn, \tag{6}$$

$$P_{Y|S}(1|1) = 1 - P_{Y|S}(0|1) \tag{7}$$

and the marginal density is

$$P_Y(y) = \sum_s P_{Y|S}(y|s)P_S(s). \tag{8}$$

Other researchers have derived the conditional probabilities $P_{Y|S}(y|s)$ of the threshold system with Gaussian noise with bipolar inputs [1] and Gaussian inputs [8]. We neither restrict the noise density to be Gaussian nor require that the density have finite variance even if the density has a bell-curve shape.

We use Shannon mutual information [19] to measure the noise enhancement or "stochastic resonance" (SR) effect [1,3,8,20,21]. The discrete Shannon mutual information of the input S and output Y is the difference between the output

unconditional entropy $H(Y)$ and the output conditional entropy $H(Y|X)$:

$$I(S,Y) = H(Y) - H(Y|S) \tag{9}$$

$$= -\sum_y P_Y(y)\log_2 P_Y(y) + \sum_s \sum_y P_{SY}(s,y)\log_2 P_{Y|S}(y|s) \tag{10}$$

$$= -\sum_y P_Y(y)\log_2 P_Y(y) + \sum_s P(s)\sum_y P(y|s)\log_2 P(y|s) \tag{11}$$

$$= \sum_{s,y} P_{SY}(s,y)\log_2 \frac{P_{SY}(s,y)}{P_S(s)P_Y(y)}. \tag{12}$$

So the mutual information is the expectation of the random variable $\log_2\{[P_{SY}(s,y)]/[P_S(s)P_Y(y)]\}$:

$$I(S,Y) = E\left[\log_2 \frac{P_{SY}(s,y)}{P_S(s)P_Y(y)}\right]. \tag{13}$$

Here $P_S(s)$ is the probability density of the input S, $P_Y(y)$ is the probability density of the output Y, $P_{Y|S}(y|s)$ is the conditional density of the output Y given the input S, and $P_{SY}(s,y)$ is the joint density of the input S and the output Y. Simple bipolar histograms of samples can estimate these densities in practice.

Mutual information also measures the pseudodistance between the joint probability density $P_{SY}(s,y)$ and the product density $P_S(s)P_Y(y)$. This holds for the Kullback [19] pseudodistance measure

$$I(S,Y) = \sum_s \sum_y P_{SY}(s,y)\log_2 \frac{P_{SY}(s,y)}{P_S(s)P_Y(y)}.$$

Then Jensen's inequality implies that $I(S,Y) \geqslant 0$. Random variables S and Y are statistically independent if and only if $I(S,Y)=0$. Hence $I(S,Y)>0$ implies some degree of dependence. The proofs in [12] and the Appendix use this fact.

III. SR FOR THRESHOLD SYSTEMS WITH FINITE-VARIANCE NOISE

Almost all finite-variance noise densities produce the SR effect in threshold neurons with subthreshold signals. This holds for all probability density functions defined on a two-symbol alphabet. The proof of Theorem 1.1 in [12] shows that if $I(S,Y)>0$ then eventually the mutual information $I(S,Y)$ tends toward zero as the noise variance tends toward zero. So the mutual information $I(S,Y)$ must *increase* as the noise variance increases from zero. The only limiting assumption is that the noise mean $E[n]$ does not lie in the "forbidden" signal-threshold interval $(\theta-A, \theta+A)$.

Theorem 1.1. Suppose that the threshold signal system (1) has noise probability density function $p(n)$ and that the input signal S is subthreshold $(A<\theta)$. Suppose that there is some statistical dependence between input random variable S and output random variable Y [so that $I(S,Y)>0$]. Suppose that the noise mean $E[n]$ does not lie in the signal-threshold interval $(\theta-A, \theta+A)$ if $p(n)$ has finite variance. Then the threshold system (1) exhibits the nonmonotone SR effect in the sense that $I(S,Y)\to 0$ as $\sigma \to 0$.

Corollary 1.1. The threshold neuron (1) exhibits stochastic resonance for the additive beta and Weibull noise densities under the hypotheses of Theorem 1.1.

The generalized beta probability density function has the form

$$p(n) = \begin{cases} \dfrac{1}{b-a}\dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\left(\dfrac{n-a}{b-a}\right)^{\alpha-1}\left(\dfrac{b-n}{b-a}\right)^{\beta-1} & \text{if } a \leqslant n \leqslant b \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

Parameters α and β are positive shape constants, parameters a and b are constants $-\infty < a < b < \infty$, and Γ is the gamma function

$$\Gamma(x) = \int_0^\infty y^{x-1}e^y dy, \quad x>0. \tag{15}$$

The mean and variance of the beta density are

$$m_n = a + (b-a)\frac{\alpha}{\alpha+\beta}, \tag{16}$$

$$\sigma_n^2 = \frac{(b-a)^2\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}. \tag{17}$$

So the beta density is right-sided for $a \geqslant 0$. We used $a=0$ and $b=10$ and so defined the beta density in the interval $[0,10]$ for the SR simulation instance in Fig. 2. The algorithm in [22] generated the beta noise. Bayesian statisticians often use a beta density to encode prior information about a parameter (such as a binomial success parameter p) over a fixed-length interval [23]. The beta density can also model the semblance or the ratio of stacked energy to total energy across a signal array [24], fluctuations of the radar-scattering cross sections

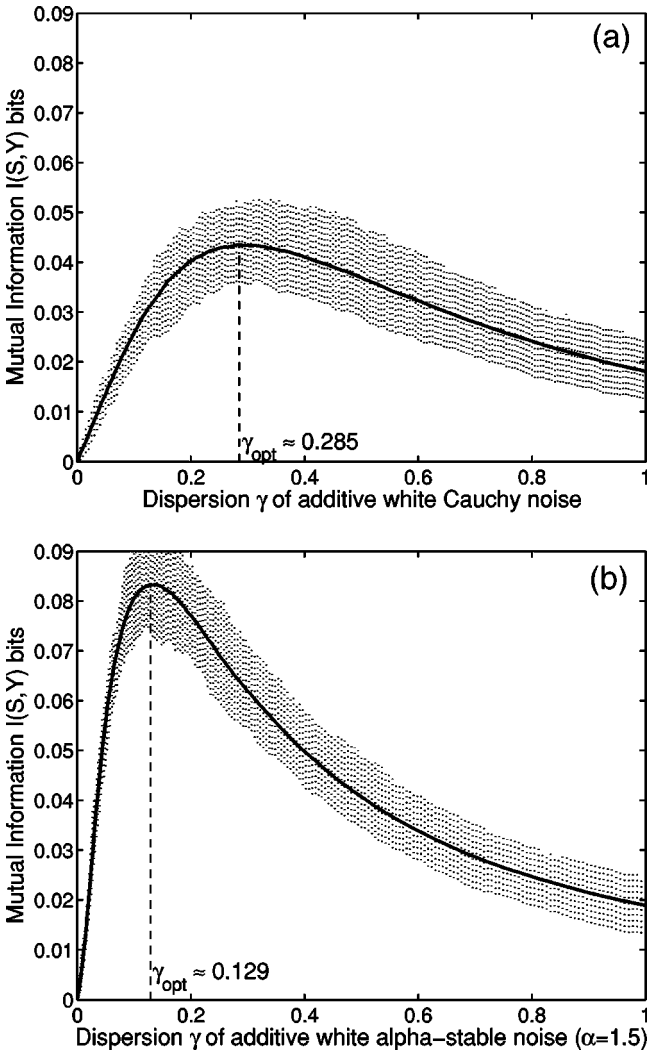

FIG. 5. Stochastic resonance with highly impulsive (infinite-variance) alpha-stable noise. The graphs show the smoothed input-output mutual information of a threshold system as a function of the dispersion of additive white alpha-stable noise n_t with $\alpha=1$ (Cauchy noise) in (a) and $\alpha=1.5$ in (b). The vertical dashed lines show the absolute deviation between the smallest and largest outliers in each sample average of 100 outcomes. The system has a nonzero noise optimum at $\gamma_{\text{opt}} \approx 0.285$ for $\alpha=1$ and $\gamma_{\text{opt}} \approx 0.129$ for $\alpha=1.5$ and thus shows the SR effect. The noisy signal-forced threshold system has the form (1). The alpha-stable noise n_t adds to the bipolar input Bernoulli signal s_t. The system has threshold $\theta=0.5$. The input Bernoulli signal has amplitude $A=0.3$ with success probability $p=\frac{1}{2}$. Each trial produced 10 000 input-output samples $\{s_t, y_t\}$ that estimated the probability densities to obtain the mutual information. Note that decreasing the tail-thickness parameter α increases the optimal noise dispersion γ_{opt} as in Fig. 7 and decreases the SR-maximal mutual information $I_{\max}(S,Y)$ as in Fig. 8.

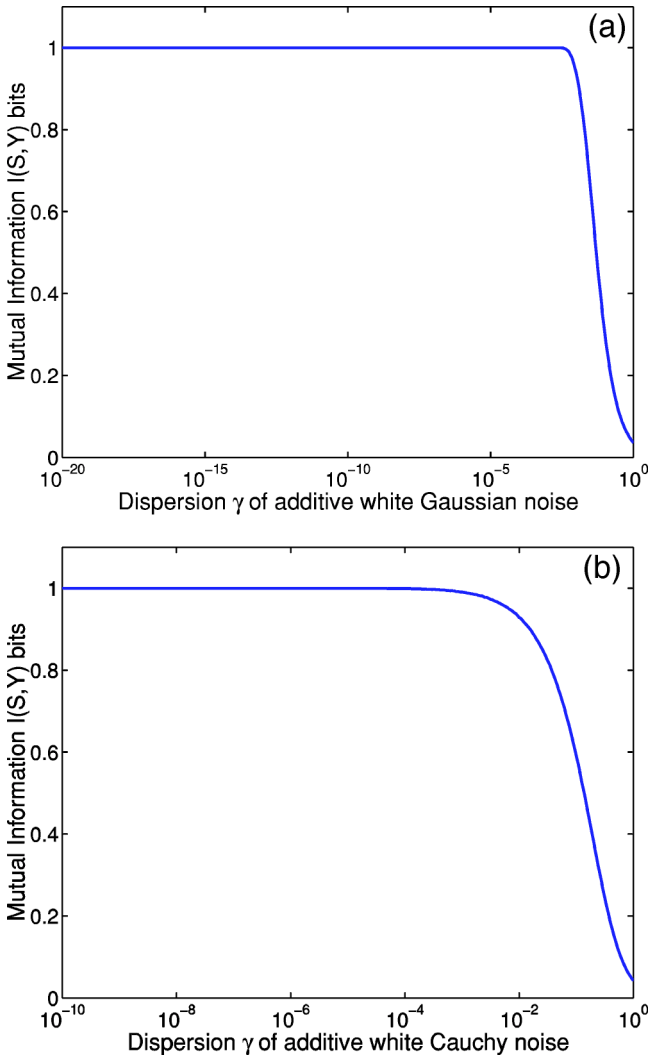

FIG. 6. No SR in the "forbidden" interval (per Theorems 1.2 and 2.2)—mutual information versus alpha-stable noise dispersion when the noise mean (location) lies in the "forbidden" signal-threshold interval: $a \in (\theta-A, \theta+A)$. The graphs show the smoothed input-output mutual information of 100 trials of a threshold system as a function of the dispersion of additive white alpha-stable noise n_t with $\alpha=2$ (Gaussian) in (a) and $\alpha=1$ (Cauchy noise) in (b). The system is optimal when $\gamma \to 0$ and thus *does not* show the SR effect: The mutual information $I(S,Y)$ is maximum when it equals the input entropy $H(S)$. The noisy signal-forced threshold system has the form (1). The alpha-stable noise n_t has location $a=0.4$ and adds to the bipolar input Bernoulli signal s_t. The system has threshold $\theta=0.5$. The input Bernoulli signal has amplitude $A=0.4$ with success probability $p=\frac{1}{2}$. Each trial produced 10 000 input-output samples $\{s_t, y_t\}$ that estimated the probability densities to obtain the mutual information.

of targets [25], the self-similar process of video traffic [26], and the variation of the narrowband vector channels or spatial signature variations due to movement [27].

The Weibull probability density function has the form

$$p(n) = \begin{cases} \alpha n^{\beta-1} e^{-\alpha n^{\beta}/\beta} & \text{if } n \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

for positive shape parameters α and β. The mean and variance of the Weibull density are

$$m_n = \left(\frac{\beta}{\alpha}\right)^{1/\beta}\Gamma\left(1+\frac{1}{\beta}\right), \tag{19}$$

$$\sigma_n^2 = \left(\frac{\beta}{\alpha}\right)^{2/\beta}\left[\Gamma\left(1+\frac{2}{\beta}\right) - \left\{\Gamma\left(1+\frac{1}{\beta}\right)\right\}^2\right]. \tag{20}$$

Figure 3 shows simulation realizations of this corollary for the Weibull noise density. MATLAB 6.5 [28] generated the Weibull noise. Weibull [29] first proposed this parametric probability density function to model the fracture of materials under repetitive stress. This density has become a standard model of multipart system reliability [30]. It can also effectively model signals and noise in many data-rich systems such as radar clutter [31] and confocal laser scanning microscopy [32].

We next state a result that shows that we cannot in general omit the threshold-interval condition in the hypothesis of Theorem 1.1. Noise does not help a threshold θ that already lies between $\theta - A$ and $\theta + A$.

Theorem 1.2. Suppose that the threshold signal system (1) has noise probability density function $p(n)$ and that the input signal S is subthreshold ($A < \theta$). Suppose that the noise mean $E[n]$ lies in the signal-threshold interval $(\theta - A, \theta + A)$ if $p(n)$ has finite variance. Then the threshold system (1) does not exhibit the nonmonotone SR effect in the sense that $I(S,Y)$ is maximum as $\sigma \to 0$:

$$I(S,Y) = H(Y) = H(S) \quad \text{as} \quad \sigma \to 0. \tag{21}$$

The Appendix gives the proof.

IV. SR FOR THRESHOLD SYSTEMS WITH INFINITE-VARIANCE NOISE

We now proceed to the more general (and more realistic) case where infinite-variance noise interferes with the threshold system. The SR effect also occurs in other systems with impulsive infinite-variance noise [16,33]. We can model many types of impulsive noise with *symmetric* alpha-stable bell-curve probability density functions with parameter α in the characteristic function $\varphi(\omega) = \exp\{-\gamma|\omega|^\alpha\}$. Here γ is the *dispersion* parameter [34–37].

The parameter α controls tail thickness and lies in $0 < \alpha \leqslant 2$. Noise grows more impulsive as α falls and the bell-curve tails grow thicker. The (thin-tailed) Gaussian density results when $\alpha = 2$ or when $\varphi(\omega) = \exp\{-\gamma\omega^2\}$. So the standard Gaussian random variable has zero mean and variance $\sigma^2 = 2$ (when $\gamma = 1$). The parameter α gives the thicker-tailed Cauchy bell curve when $\alpha = 1$ or $\varphi(\omega) = \exp\{-|\omega|\}$ for a zero *location* ($a = 0$) and unit dispersion ($\gamma = 1$) Cauchy random variable. The moments of stable distributions with $\alpha < 2$ are finite only up to order k for $k < \alpha$. The Gaussian density alone has finite variance and higher moments. Alpha-stable random variables characterize the class of normalized sums of independent random variables that converge in distribution to a random variable [34] as in the famous Gaussian special case called the "central limit theorem."

Alpha-stable models tend to work well when the noise or signal data contain "outliers"— and all do to some degree. Models with $\alpha < 2$ can accurately describe impulsive noise in telephone lines, underwater acoustics, low-frequency atmospheric signals, fluctuations in gravitational fields and financial prices, and many other processes [37,38]. Note that the best choice of α is an *empirical* question for bell-curve phenomena. Bell-curve behavior alone does not justify the (extreme) assumption of the Gaussian bell curve. Figure 4 shows realizations of four symmetric alpha-stable noise random variables.

Theorem 2.1 applies to *any* alpha-stable noise model. The density need not be symmetric. A general alpha-stable probability density function f has characteristic function φ [36,37,41,42]

$$\varphi(\omega) = \exp\left\{ia\omega - \gamma|\omega|^\alpha\left(1 + i\beta\,\mathrm{sgn}(\omega)\tan\frac{\alpha\pi}{2}\right)\right\} \quad \text{for } \alpha \neq 1 \tag{22}$$

and

$$\varphi(\omega) = \exp\{ia\omega - \gamma|\omega|[1 - 2i\beta\,\ln|\omega|\mathrm{sgn}(\omega)/\pi]\} \quad \text{for } \alpha = 1, \tag{23}$$

where

$$\mathrm{sgn}(\omega) = \begin{cases} 1 & \text{if } \omega > 0 \\ 0 & \text{if } \omega = 0 \\ -1 & \text{if } \omega < 0 \end{cases} \tag{24}$$

and $i = \sqrt{-1}$, $0 < \alpha \leqslant 2$, $-1 \leqslant \beta \leqslant 1$, and $\gamma > 0$. The parameter α is the characteristic exponent. Again the variance of an alpha-stable density does not exist if $\alpha < 2$. The location parameter a is the "mean" of the density when $\alpha > 1$. β is a skewness parameter. The density is symmetric about a when $\beta = 0$. The theorem below still holds even when $\beta \neq 0$. The dispersion parameter γ acts like a variance because it controls the width of a symmetric alpha-stable bell curve. There are no known closed forms of the α-stable densities for most α's. Numerical integration of φ produced the simulation results in Fig. 4.

The proof of Theorem 2.1 in [12] is simpler than the proof in the finite-variance case because all stable noise densities have a characteristic function with the exponential form in Eqs. (22) and (23). So zero noise dispersion gives φ as a simple complex exponential and hence gives the corresponding density as a delta spike that can fall outside the interval $(\theta - A, \theta + A)$.

Theorem 2.1. Suppose $I(S,Y) > 0$ and the threshold system (1) uses alpha-stable noise with location parameter $a \notin (\theta - A, \theta + A)$. Then the system (1) exhibits the nonmonotone SR effect if the input signal is subthreshold.

Figure 5 gives a typical example of the SR effect for highly impulsive noise with infinite variance. The alpha-stable noises have $\alpha = 1$ (Cauchy) and $\alpha = 1.5$. So frequent and violent noise spikes interfere with the signal. Figure 5 also illustrates the empirical trends in Figs. 7 and 8: A falling tail-thickness parameter α produces an increasing optimal noise dispersion γ_{opt} but a decreasing SR-maximal mutual information $I_{\max}(S,Y)$. We next state a new sufficient condition for SR not to occur in an impulsive threshold system.

TABLE I. Linear regression estimates of the SR-optimal log dispersion γ_{opt} as a function of the bell-curve tail-thickness parameter α from a symmetric alpha-stable noise density. The parameters β_0 and β_1 relate $\log_{10}\gamma_{opt}$ and α through a linear relationship: $\log_{10}\gamma_{opt}(\alpha)=\beta_0+\beta_1\alpha$. The coefficient of determination r_l^2 shows how well the linear model fits the log-transformed data. The last column shows the coefficient of determination r_q^2 for the quadratic model $\log_{10}\gamma_{opt}(\alpha)=\beta_0+\beta_1\alpha+\beta_2\alpha^2$. All observed significance levels or p-values were less than 10^{-4}.

Signal amplitude	Linear model Regression coefficients			Quadratic model
A	$\hat{\beta}_0$	$\hat{\beta}_1$	r_l^2	r_q^2
0.025	0.0701	−0.5944	0.9003	0.9444
0.050	0.1002	−0.6087	0.9321	0.9723
0.075	0.1124	−0.6192	0.9490	0.9842
0.100	0.1180	−0.6261	0.9558	0.9888
0.125	0.1090	−0.6228	0.9594	0.9910
0.150	0.1078	−0.6251	0.9679	0.9921
0.175	0.1026	−0.6273	0.9672	0.9933
0.200	0.0915	−0.6214	0.9699	0.9942
0.225	0.0810	−0.6161	0.9737	0.9950
0.250	0.0694	−0.6172	0.9781	0.9959
0.275	0.0595	−0.6149	0.9826	0.9964
0.300	0.0439	−0.6148	0.9869	0.9961
0.325	0.0290	−0.6184	0.9903	0.9962
0.350	0.0116	−0.6211	0.9935	0.9961
0.375	−0.0134	−0.6215	0.9957	0.9960
0.400	−0.0313	−0.6367	0.9947	0.9951
0.425	−0.0705	−0.6432	0.9903	0.9950
0.450	−0.1107	−0.6688	0.9757	0.9944
0.475	−0.1837	−0.7217	0.9408	0.9911
0.490	−0.2805	−0.8053	0.8987	0.9863

Theorem 2.2. Suppose that the threshold signal system (1) has subthreshold input signal and use alpha-stable noise with location parameter $a \in (\theta-A, \theta+A)$. Then the threshold system (1) does not exhibit the nonmonotone SR effect: $I(S,Y)$ is maximum as $\gamma \to 0$:

$$I(S,Y) = H(Y) = H(S) \quad \text{as } \gamma \to 0. \tag{25}$$

The Appendix gives the proof. Figure 6 shows the noise-mutual information profile of the subthreshold signal system with noise location (mean) in the "forbidden" signal-threshold interval.

Statistical regression confirmed an exponential relationship between the optimal noise dispersion γ_{opt} and the bell-curve tail-thickness parameter α: $\gamma_{opt}(\alpha)=10^{\beta_0+\beta_1\alpha}$ for parameters β_0 and β_1 that depend on the signal amplitude A. Then the log-transformation of the optimal dispersion gives the linear model $\log_{10}\gamma_{opt}(\alpha)=\beta_0+\beta_1\alpha$. Table I shows the estimated parameters $\hat{\beta}_0$ and $\hat{\beta}_1$ and the coefficient of determination r_l^2 for 20 signal amplitudes in the threshold neuron using SPSS software. All observed significance levels or p-values were less than 10^{-4}. The p-values measure the credibility of the null hypothesis that the regression lines have zero slope or other coefficients. The exponential trend's exponent is linear for most amplitudes but becomes quadratic for very small amplitudes and for amplitudes close to the threshold $\theta=\frac{1}{2}$ [or $\gamma_{opt}(\alpha)=10^{\beta_0+\beta_1\alpha+\beta_2\alpha^2}$ for a quadratic fit to the data]. Figure 7 shows 6 of the 20 log-linear plots.

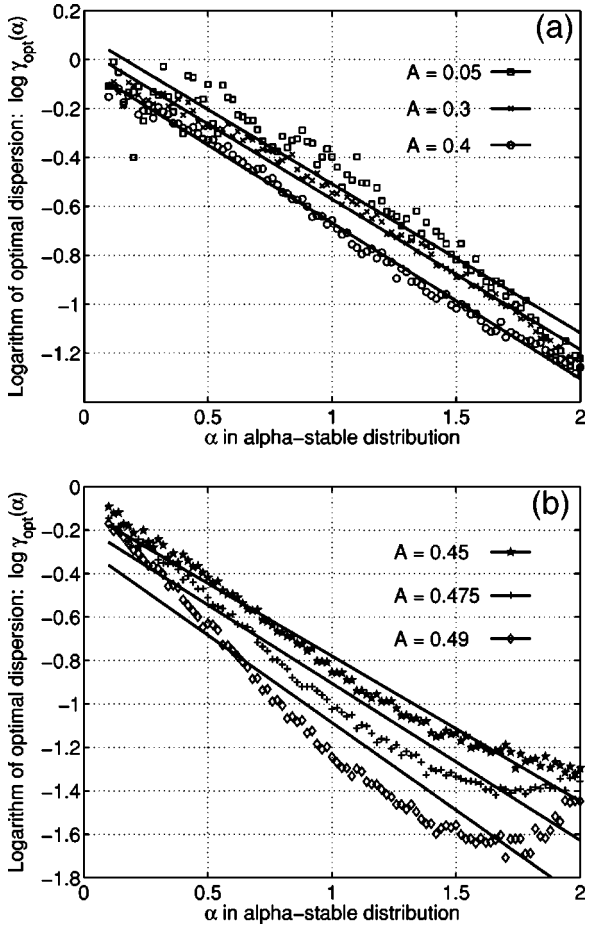

FIG. 7. Exponential law for optimal noise dispersion γ_{opt} as a function of bell-curve thickness parameter α for the mutual-information performance measure and for different signal amplitudes A. The optimal noise dispersion γ_{opt} depends on the parameter α through the exponential relation $\gamma_{opt}(\alpha)=10^{\beta_0+\beta_1\alpha}$ for parameters β_0 and β_1 [or $\gamma_{opt}(\alpha)=10^{\beta_0+\beta_1\alpha+\beta_2\alpha^2}$ for a quadratic fit to the data]. Table I shows the estimated parameters $\hat{\beta}_0$ and $\hat{\beta}_1$ for 20 input Bernoulli signal amplitudes A. The exponential trend's exponent is linear for most amplitudes but becomes quadratic for very small amplitudes and for amplitudes close to the threshold $\theta=\frac{1}{2}$. All observed significance levels or p-values were less than 10^{-4}.

We also found an approximate linear relationship $I_{max}(S,Y;\alpha)=\beta_0+\beta_1\alpha$ for the SR-maximal mutual information $I_{max}(S,Y)$ as a function of the tail-thickness parameter α. Table II shows the estimated parameters $\hat{\beta}_0$ and $\hat{\beta}_1$ and the coefficient of determination r_l^2 for 20 signal amplitudes in the threshold neuron. All observed significance levels or p-values were less than 10^{-4}. There is a clear linear trend for most amplitudes A. The trend becomes quadratic for very

TABLE II. Linear regression of the SR-maximal mutual information $I_{\max}(S,Y)$ as a function of the bell-curve tail-thickness parameter α from a symmetric alpha-stable noise density. The parameters β_0 and β_1 relate $I_{\max}(S,Y)$ and α through a linear relationship: $I_{\max}(S,Y;\alpha)=\beta_0+\beta_1\alpha$. The coefficient of determination r_l^2 shows how well the linear model fits the data. The last column shows the coefficient of determination r_q^2 for the quadratic model $I_{\max}(S,Y;\alpha)=\beta_0+\beta_1\alpha+\beta_2\alpha^2$. All observed significance levels or p-values were less than 10^{-4}.

Signal amplitude	Linear model Regression coefficients			Quadratic model
A	$\hat{\beta}_0$	$\hat{\beta}_1$	r_l^2	r_q^2
0.025	−0.0001	0.0006	0.9312	0.9907
0.050	−0.0008	0.0022	0.9370	0.9972
0.075	−0.0018	0.0049	0.9401	0.9985
0.100	−0.0031	0.0086	0.9440	0.9990
0.125	−0.0048	0.0134	0.9477	0.9993
0.150	−0.0068	0.0190	0.9521	0.9995
0.175	−0.0090	0.0256	0.9558	0.9997
0.200	−0.0113	0.0329	0.9612	0.9998
0.225	−0.0138	0.0411	0.9658	0.9998
0.250	−0.0161	0.0500	0.9715	0.9997
0.275	−0.0185	0.0596	0.9764	0.9995
0.300	−0.0207	0.0698	0.9816	0.9993
0.325	−0.0224	0.0807	0.9866	0.9990
0.350	−0.0236	0.0920	0.9913	0.9987
0.375	−0.0240	0.1039	0.9951	0.9984
0.400	−0.0229	0.1161	0.9976	0.9981
0.425	−0.0196	0.1286	0.9972	0.9977
0.450	−0.0120	0.1408	0.9905	0.9975
0.475	0.0058	0.1513	0.9655	0.9973
0.490	0.0336	0.1527	0.9145	0.9959

small amplitudes and for amplitudes close to the threshold $\theta=\frac{1}{2}$. Figure 8 shows 6 of the 20 linear plots.

V. CONCLUSIONS

Both theory and detailed simulations show that almost all noise types produce stochastic resonance in threshold systems that use subthreshold signals. These results help explain the widespread occurrence of the SR effect in mechanical and biological threshold systems [43–49]. The broad generality of the results suggests that SR should occur in any nonlinear system whose input-output structure approximates a threshold system as in the many models of continuous neurons [50–52]. The infinite-variance theoretical and simulation results further imply that such widespread SR effects should be robust against violent noise impulses.

ACKNOWLEDGMENTS

National Science Foundation Grant No. ECS-0070284 and Thailand Research Fund Grant No. RSA4680001 supported this research.

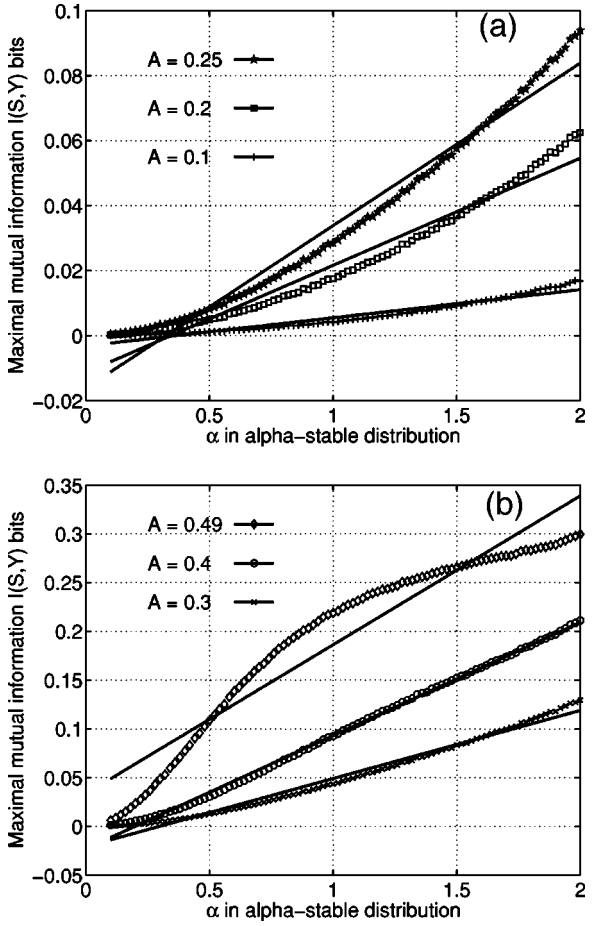

FIG. 8. Linear regression for maximal mutual information $I_{\max}(S,Y)$ as a function of bell-curve thickness parameter α for different signal amplitudes A. The maximal mutual information $I_{\max}(S,Y)$ depends on the parameter α through the linear relationship $I_{\max}(S,Y;\alpha)=\beta_0+\beta_1\alpha$ for parameters β_0 and β_1 [or $I_{\max}(\alpha)=\beta_0+\beta_1\alpha+\beta_2\alpha^2$ for a quadratic fit to the data]. Table II shows the estimated parameters $\hat{\beta}_0$ and $\hat{\beta}_1$ for 20 input Bernoulli signal amplitudes A. The linear trend is strong for most amplitudes A. The trend becomes quadratic for very small amplitudes and for amplitudes close to the threshold $\theta=\frac{1}{2}$. All observed significance levels or p-values were less than 10^{-4}.

APPENDIX: PROOFS OF THEOREMS 1.2 AND 2.2

The two proofs below use the same idea as do the proofs for Theorems 1.1 and 2.1 [12]. Assume $0<P_S(s)<1$ to avoid triviality when $P_S(s)=0$ or 1. We show that $H(Y)\to H(S)$ and $H(Y|S)\to 0$ as $\sigma\to 0$ or $\gamma\to 0$. So $I(S,Y)\to H(S)$ as $\sigma\to 0$ or $\gamma\to 0$ and is maximum since $I(S,Y)=H(Y)-H(Y|S)$ and $I(S,Y)\leqslant H(S)$ by the data processing inequality: $I(S,S)\geqslant I(S,g(S))=I(S,Y)$ for a Markov chain $S\to S\to Y$ [19]. The boundary case $I(S,S)=H(S)$ implies $I(S,Y)\leqslant H(S)$.

Finite-variance noise case (Theorem 1.2)

Now we show that $P_{Y|S}(y|s)$ is either 1 or 0 as $\sigma\to 0$ or $\gamma\to 0$. Let the mean of the noise be $m=E[n]$ and the variance be $\sigma^2=E[(n-m)^2]$. Then $m\in(\theta-A,\theta+A)$ by hypothesis.

Consider first $P_{Y|S}(0|0)$. Pick $\varepsilon = \frac{1}{2}d(\theta+A,m) = \frac{1}{2}(\theta+A-m) > 0$. So $\theta+A-\varepsilon = \theta+A-\varepsilon+m-m = m+(\theta+A-m)-\varepsilon = m+2\varepsilon-\varepsilon = m+\varepsilon$. Then

$$P_{Y|S}(0|0) = \int_{-\infty}^{\theta+A} p(n)dn \tag{A1}$$

$$\geqslant \int_{-\infty}^{\theta+A-\varepsilon} p(n)dn \tag{A2}$$

$$= \int_{-\infty}^{m+\varepsilon} p(n)dn \tag{A3}$$

$$= 1 - \int_{m+\varepsilon}^{\infty} p(n)dn \tag{A4}$$

$$= 1 - \Pr\{n \geqslant m+\varepsilon\} = 1 - \Pr\{n-m \geqslant \varepsilon\} \tag{A5}$$

$$\geqslant 1 - \Pr\{|n-m| \geqslant \varepsilon\} \tag{A6}$$

$$\geqslant 1 - \frac{\sigma^2}{\varepsilon^2} \quad \text{by Chebyshev's inequality} \tag{A7}$$

$$\to 1 \quad \text{as } \sigma \to 0. \tag{A8}$$

So $P_{Y|S}(0|0) = 1$.

Similarly for $P_{Y|S}(1|1)$: Pick $\varepsilon = \frac{1}{2}d(\theta-A,m) = \frac{1}{2}(m-\theta+A) > 0$. So $\theta-A+\varepsilon = \theta-A+\varepsilon+m-m = m+(\theta-A-m)+\varepsilon = m-2\varepsilon+\varepsilon = m-\varepsilon$. Then

$$P_{Y|S}(1|1) = \int_{\theta-A}^{\infty} p(n)dn \tag{A9}$$

$$\geqslant \int_{\theta-A+\varepsilon}^{\infty} p(n)dn \tag{A10}$$

$$= \int_{m-\varepsilon}^{\infty} p(n)dn \tag{A11}$$

$$= 1 - \int_{-\infty}^{m-\varepsilon} p(n)dn \tag{A12}$$

$$= 1 - \Pr\{n \leqslant m-\varepsilon\} = 1 - \Pr\{n-m \leqslant -\varepsilon\} \tag{A13}$$

$$\geqslant 1 - \Pr\{|n-m| \geqslant \varepsilon\} \tag{A14}$$

$$\geqslant 1 - \frac{\sigma^2}{\varepsilon^2} \quad \text{by Chebyshev's inequality} \tag{A15}$$

$$\to 1 \quad \text{as} \quad \sigma \to 0. \tag{A16}$$

So $P_{Y|S}(1|1) = 1$.

Alpha-stable noise case (Theorem 2.2)

The characteristic function $\varphi(\omega)$ of alpha-stable noise density $p(n)$ has the exponential form (22) and (23). This reduces to a simple complex exponential in the zero-dispersion limit:

$$\lim_{\gamma \to 0} \varphi(\omega) = \exp\{ia\omega\} \tag{A17}$$

for *all* characteristic exponents α, skewnesses β, and locations a. So Fourier transformation gives the corresponding density function in the limiting case ($\gamma \to 0$) as a translated delta function δ:

$$\lim_{\gamma \to 0} p(n) = \delta(n-a). \tag{A18}$$

Then $a \in (\theta-A, \theta+A)$ gives

$$P_{Y|S}(0|0) = \int_{-\infty}^{\theta+A} p(n)dn \tag{A19}$$

$$\to \int_{-\infty}^{\theta+A} \delta(n-a)dn = 1 \quad \text{as } \gamma \to 0. \tag{A20}$$

Similarly

$$P_{Y|S}(1|1) = \int_{\theta-A}^{\infty} p(n)dn \tag{A21}$$

$$\to \int_{\theta-A}^{\infty} \delta(n-a)dn = 1 \quad \text{as} \quad \gamma \to 0. \tag{A22}$$

The two conditional probabilities for both the finite-variance and infinite-variance cases likewise imply that $P_{Y|S}(0|1) = P_{Y|S}(1|0) = 0$ as $\sigma \to 0$ or $\gamma \to 0$. These four probabilities further imply that

$$H(Y|S) = -\sum_s \sum_y P_{SY}(s,y) \log_2 P_{Y|S}(y|s) \tag{A23}$$

$$= \sum_s P_S(s) \sum_y P_{Y|S}(y|s) \log_2 P_{Y|S}(y|s) \tag{A24}$$

$$= 0, \tag{A25}$$

where we use the fact (L'Hôpital) that $0 \log_2 0 = 0$. The unconditional entropy $H(Y)$ becomes

$$H(Y) = -\sum_y P_Y(y) \log_2 P_Y(y) \tag{A26}$$

$$= -\sum_s P_S(s) \log_2 P_S(s) \tag{A27}$$

$$= H(S) \tag{A28}$$

because

$$P_Y(y) = \sum_s P_{Y|S}(y|s)P_S(s) \tag{A29}$$

$$= P_{Y|S}(y|0)P_S(0) + P_{Y|S}(y|1)P_S(1) \tag{A30}$$

$$= P_{Y|S}(y|0)P_S(0) + P_{Y|S}(y|1)[1 - P_S(0)] \tag{A31}$$

$$= [P_{Y|S}(y|0) - P_{Y|S}(y|1)]P_S(0) + P_{Y|S}(y|1) \tag{A32}$$

$$= [P_{Y|S}(y|1) - P_{Y|S}(y|0)]P_S(1) + P_{Y|S}(y|0) \tag{A33}$$

$$= \begin{cases} P_S(1) & \text{if } y = 1 \\ P_S(0) & \text{if } y = 0. \end{cases} \tag{A34}$$

[1] A. R. Bulsara and A. Zador, Phys. Rev. E **54**, R2185 (1996).
[2] L. Gammaitoni, Phys. Lett. A **208**, 315 (1995).
[3] X. Godivier and F. Chapeau-Blondeau, Int. J. Bifurcation Chaos Appl. Sci. Eng. **8**, 581 (1998).
[4] S. M. Hess and A. M. Albano, Int. J. Bifurcation Chaos Appl. Sci. Eng. **8**, 395 (1998).
[5] N. Hohn and A. N. Burkitt, in *Proceedings of the 2001 IEEE International Joint Conference on Neural Networks (IJCNN '01)* (IEEE, Washington, D.C., 2001), pp. 644–647.
[6] P. Jung, Phys. Lett. A **207**, 93 (1995).
[7] P. Jung and G. Mayer-Kress, Nuovo Cimento D **17**, 827 (1995).
[8] N. G. Stocks, Phys. Rev. E **63** 041114 (2001).
[9] J. J. Collins, C. C. Chow, A. C. Capela, and T. T. Imhoff, Phys. Rev. E **54**, 5575 (1996).
[10] J. J. Collins, C. C. Chow, and T. T. Imhoff, Nature (London) **376**, 236 (1995).
[11] S. Mitaim and B. Kosko, in *Proceedings of the 2002 IEEE International Joint Conference on Neural Networks (IJCNN '02), Honolulu* (IEEE, Honolulu, 2002), pp. 1980–1985.
[12] B. Kosko and S. Mitaim, Neural Networks **16**, 755 (2003).
[13] S. Grossberg, *Studies of Mind and Brain: Neural Principles of Learning, Perception, Development, Cognition, and Motor Control* (Kluwer Academic Publishers, Dordrecht, 1982).
[14] S. Haykin, *Neural Networks: A Comprehensive Foundation* (McMillan, 1994).
[15] B. Kosko, *Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence* (Prentice Hall, Englewood Cliffs, NJ, 1992).
[16] B. Kosko and S. Mitaim, Phys. Rev. E **64**, 051110 (2001).
[17] I. Miller and M. Miller, *Mathematical Statistics*, 6th ed. (Prentice Hall, Englewood Cliffs, NJ, 1999).
[18] J. J. Hopfield, Proc. Natl. Acad. Sci. U.S.A. **79**, 2554 (1982).
[19] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (John Wiley & Sons, 1991).
[20] G. Deco and B. Schürmann, Physica D **117**, 276 (1998).
[21] M. E. Inchiosa, J. W. C. Robinson, and A. R. Bulsara, Phys. Rev. Lett. **85**, 3369 (2000).
[22] R. C. H. Cheng, Commun. ACM **21**, 317 (1978).
[23] E. A. Tanis and R. V. Hogg, *Probability and Statistical Inference*, 6th ed. (Prentice Hall, Englewood Cliffs, NJ, 2001).
[24] C. V. Kimball and D. J. Scheibner, Geophysics **63**, 345 (1998).
[25] R. L. Kulp, Electromagnetics **4**, 165 (1984).
[26] N. Ansari, H. Liu, Y. Q. Shi, and H. Zhao, IEEE Trans. Broadcasting **48**, 337 (2002).
[27] A. Kavak, M. Torlak, W. J. Vogel, and G. Xu, IEEE Trans. Microwave Theory Tech. **46**, 930 (2000).
[28] L. Devroye, *Non-Uniform Random Variate Generation* (Springer-Verlag, Berlin, 1986).
[29] W. Weibull, J. Appl. Mech. **18**, 293 (1951).
[30] W. Mendenhall and T. Sincich, *Statistics for Engineering and the Sciences*, 4th ed. (Prentice Hall, Englewood Cliffs, NJ, 1995).
[31] M. A. Sletten, IEEE Trans. Antennas Propag. **46**, 45 (1998).
[32] J. Hu, A. Razden, G. M. Nielson, G. E. Farin, D. P. Baluch, and D. G. Capco, IEEE Trans. Vis. Comput. Graph. **9**, 320 (2003).
[33] S. Mitaim and B. Kosko, Proc. IEEE **86**, 2152 (1998), special issue on intelligent signal processing.
[34] L. Breiman, *Probability* (Addison-Wesley, Reading, MA, 1968).
[35] W. Feller, *An Introduction to Probability Theory and Its Applications* (John Wiley & Sons, Reading, MA,1966) Vol. II.
[36] M. Grigoriu, *Applied Non-Gaussian Processes* (Prentice Hall, Englewood Cliffs, NJ, 1995).
[37] C. L. Nikias and M. Shao, *Signal Processing with Alpha-Stable Distributions and Applications* (John Wiley & Sons, 1995).
[38] B. Kosko, *Fuzzy Engineering* (Prentice Hall, Englewood Cliffs, NJ, 1996).
[39] J. M. Chambers, C. L. Mallows, and B. W. Stuck, J. Am. Stat. Assoc. **71**, 340 (1976).
[40] P. Tsakalides and C. L. Nikias, IEEE Trans. Signal Process. **44**, 1623 (1996).
[41] V. Akgiray and C. G. Lamoureux, J. Bus. Econ. Stat. **7**, 85 (1989).
[42] H. Bergstrom, Arkiv Feur Matematik **2**, 375 (1952).
[43] H. A. Braun, H. Wissing, K. Schäfer, and M. C. Hirsch, Nature (London) **367**, 270 (1994).
[44] J. K. Douglass, L. Wilkens, E. Pantazelou, and F. Moss, Nature (London) **365**, 337 (1993).
[45] S. Fauve and F. Heslot, Phys. Lett. **97A**, 5 (1983).
[46] L. Gammaitoni, P. Hänggi, P. Jung, and F. Marchesoni, Rev. Mod. Phys. **70**, 223 (1998).
[47] J. E. Levin and J. P. Miller, Nature (London) **380**, 165 (1996).
[48] V. I. Melnikov, Phys. Rev. E **48**, 2481 (1993).
[49] D. F. Russell, L. A. Willkens, and F. Moss, Nature (London) **402**, 291 (1999).
[50] A. R. Bulsara, A. J. Maren, and G. Schmera, Biol. Cybern. **70**, 145 (1993).
[51] A. Longtin, J. Stat. Phys. **70**, 309 (1993).
[52] H. E. Plesser and S. Tanaka, Phys. Lett. A **225**, 228 (1997).