Adaptive Stochastic Resonance in Noisy Neurons Based on Mutual Information

Sanya Mitaim and Bart Kosko

Abstract—**Noise can improve how memoryless neurons process signals and maximize their throughput information. Such favorable use of noise is the so-called "stochastic resonance" or SR effect at the level of threshold neurons and continuous neurons. This paper presents theoretical and simulation evidence that 1) lone noisy threshold and continuous neurons exhibit the SR effect in terms of the mutual information between random input and output sequences, 2) a new statistically robust learning law can find this entropy-optimal noise level, and 3) the adaptive SR effect is robust against highly impulsive noise with infinite variance. Histograms estimate the relevant probability density functions at each learning iteration. A theorem shows that almost all noise probability density functions produce some SR effect in threshold neurons even if the noise is impulsive and has infinite variance. The optimal noise level in threshold neurons also behaves nonlinearly as the input signal amplitude increases. Simulations further show that the SR effect persists for several sigmoidal neurons and for Gaussian radial-basis-function neurons.**

Index Terms—**Alpha-stable noise, impulsive noise, infinite-variance statistics, mutual information, noise processing, sigmoidal neurons and radial basis functions, stochastic gradient learning, stochastic resonance (SR), threshold neurons.**

I. NOISE AND ADAPTIVE STOCHASTIC RESONANCE

NOISE is an unwanted signal or source of energy. Scientists and engineers have largely tried to filter noise or cancel it or mask it out of existence. The Noise Pollution Clearinghouse condemns noise outright: "Noise is unwanted sound. It is derived from the Latin word 'nausea' meaning seasickness. Noise is among the most pervasive pollutants today. Noise from road traffic, jet planes, jet skis, garbage trucks, construction equipment, manufacturing processes, lawn mowers, leaf blowers, and boom boxes, to name a few, are among the unwanted sounds that are routinely broadcast into the air."

The new field of *stochastic resonance* or SR [3], [4], [9], [26], [32], [52], [53], [61], [71] rests on an exception to this undeclared war on noise. SR occurs when noise enhances a faint signal in a nonlinear system. It occurs when the addition of a small amount of noise increases a nonlinear system's performance measure such as its signal-to-noise ratio (SNR), cross-correlation, or mutual information. The nonlinearity is often as simple as a memoryless threshold. So a great deal of SR research has focused on how dither-like noise can help spiking neurons process data streams [12], [33], [38]. SR occurs in physical systems such as ring lasers [56], threshold hysteretic Schmitt triggers [27], superconducting quantum interference devices (SQUIDs) [36], Josephson junctions [7], chemical systems [25], and quantum-mechanical systems [34]. SR also occurs in biological systems such as the rat [18], crayfish [23], cricket [48], river paddlefish [66], and in many types of model neurons [8], [10], [16], [17], [63].

Fig. 1 shows how uniform pixel noise can improve our subjective perception of an image. The system quantizes the original gray-scale "Lena" image into a binary image of black and white pixels. It emits a white pixel as output if the input gray-scale pixel equals or exceeds a threshold. It emits a black pixel as output if the input gray-scale pixel falls below the threshold. This quantizer is biased because it does not set the threshold at the midpoint of the gray scale. So the quantized version of the original image contains almost no information. A small level of noise sharpens the image contours and helps fill in features when it adds to the original image before the system applies the threshold. Too much noise swamps the image and degrades its contours. Gammaitoni [29] and others [70] have proposed a dithering argument for this SR effect and still others [55] have applied this argument to still images. The argument involves adding dither noise to a signal before quantization. Consider gray-scale pixel $x \in [0, 1]$ and binary output pixel $y \in \{0, 1\}$ with threshold $\theta = 1/2$. Then the dithered quantizer gives $E[Y|x] = 1 - \Pr\{n < \theta - x\} = x$ if and only if the noise is uniform on $(-1/2, 1/2)$. But the subjective SR result in Fig. 1 holds for nonuniform *infinite*-variance Cauchy noise and for many other types of nonuniform noise. So the dithering argument only partially explains this subjective SR effect.

We first show that noise added to a memoryless threshold neuron produces the SR effect in terms of the Shannon mutual information $I(S, Y)$ between realizations of a random (Bernoulli) bipolar input signal S and realizations of the thresholded output random variable Y. Fig. 2 shows a typical simulation confirmation of this SR result for additive Gaussian noise. The theorem holds for more general bell curves that have thicker tails and thus that have infinite variance and can produce impulsive noise. Extensive simulations reproduce these SR effects for several standard continuous sigmoidal neurons and for Gaussian radial basis functions (see Fig. 13).

We next show that a new robust learning law can find the optimal noise variance and dispersion for both threshold and continuous neurons and for both finite-variance and infinite-variance noise. We introduced adaptive stochastic resonance in [57]
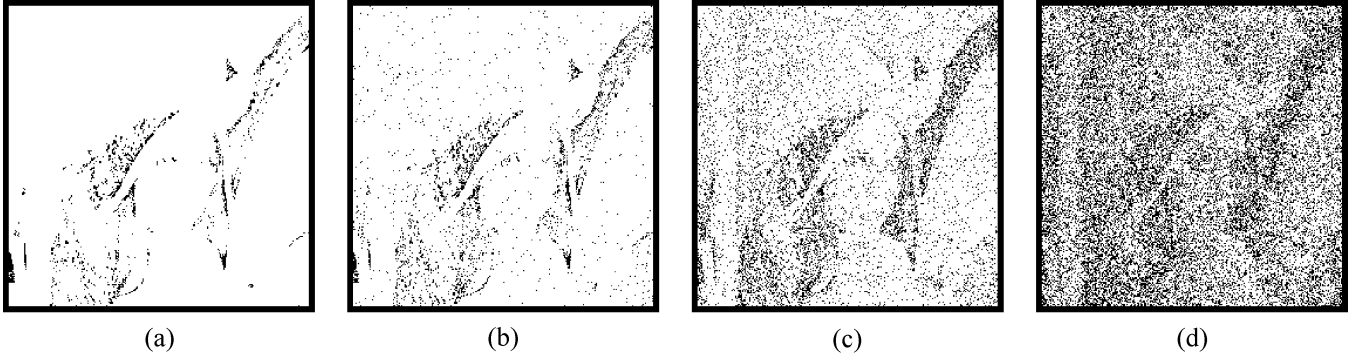
Fig. 1. A "dithering" Cauchy pixel noise can improve subjective image quality. The noise produces a nonmonotonic response: A small level of noise sharpens the image features while too much noise degrades them. These noisy images result when we apply a pixel threshold to the popular "Lena" image used in signal processing [60]: $y = g((x + n) - \theta)$ where $g(x) = 1$ if $x \geq 0$ and $g(x) = 0$ if $x < 0$ for an input pixel value $x \in [0, 1]$ and output pixel value $y \in \{0, 1\}$. The input image's gray-scale pixels vary from 0 (black) to 1 (white). The threshold is $\theta = 0.06$. Thresholding the original "Lena" image gives the faint image in (a). The Cauchy noise n has zero location and its dispersion γ_n grows from (b)–(d): $\gamma_n = 0.01$ in (b), $\gamma_n = 0.08$ in (c), and $\gamma_n = 0.50$ in (d).
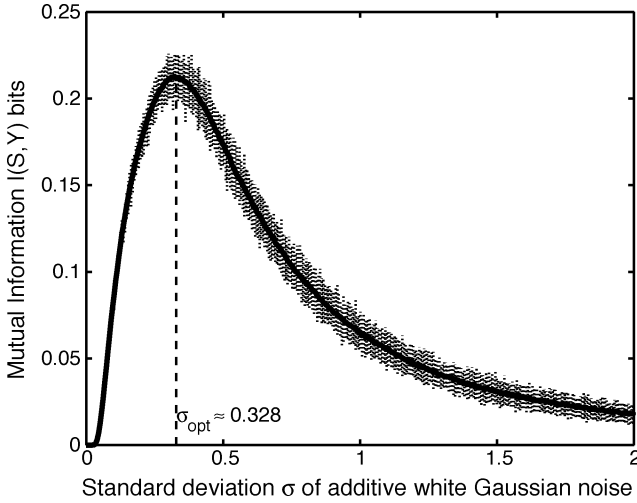

Fig. 2. The nonmonotonic signature of stochastic resonance. The graph shows the smoothed input-output mutual information of a threshold system as a function of the standard deviation of additive white Gaussian noise n_t. The vertical dashed lines show the absolute deviation between the smallest and largest outliers in each sample average of 100 outcomes. The system has a nonzero noise optimum at $\sigma_{\mathrm{opt}} \approx 0.328$ and thus shows the SR effect. The noisy signal-forced threshold system has the form (6). The Gaussian noise n_t adds to the external forcing bipolar signal s_t.

and [47] as a robustified stochastic gradient ascent algorithm that slowly finds the optimal noise variance or dispersion given thousands of joint samples of the noise input and the nonlinear system's spectral SNR or its cross correlation. This paper extends adaptive SR to the mutual-information performance measure. The last section derives and tests a new robustified learning law that finds the entropically optimal noise level given histogram estimates of the underlying marginal and conditional probability density functions. This statistically robust algorithm uses only the sign of the noise gradient rather than the gradient itself.

The results show that model neurons can exploit low levels of crosstalk or other forms of noise in their local environment. Even highly impulsive noise can help neurons maximize their throughput information. Such noise-based information maximization is consistent with Linsker's principle of information maximization in neural networks [49], [50]. These findings support the implicit SR conjecture that biological neurons have

evolved to computationally exploit their noisy environments [11], [18], [19], [23], [48], [58], [64], [69]. Further support is that these adaptive SR effects still hold for other sigmoidal and nonsigmoidal (Gaussian) neurons as Fig. 13 shows. These results suggest that biological neurons should experience less mutual information if they do not use their local noise.

II. MUTUAL INFORMATION AND SR IN NEURON MODELS

This section reviews Shannon's measure of mutual information between two random variables. Then it reviews the simple nonlinear threshold model of a neuron and the continuous neuron model that show the SR effect for bipolar signals.

A. Mutual Information Measure

Mutual information [20] can measure the SR effect [12], [22], [33], [43], [67]. The discrete Shannon mutual information of the input S and output Y has the form

$$I(S, Y) = H(Y) - H(Y|S) \tag{1}$$

$$= -\sum_y P_Y(y) \log P_Y(y)$$

$$+ \sum_s \sum_y P_{SY}(s, y) \log P_{Y|S}(y|s) \tag{2}$$

$$= -\sum_y P_Y(y) \log P_Y(y)$$

$$+ \sum_s P(s) \sum_y P(y|s) \log P(y|s) \tag{3}$$

$$= \sum_{s,y} P_{SY}(s, y) \log \frac{P_{SY}(s, y)}{P_S(s) P_Y(y)}. \tag{4}$$

We can view the mutual information in the form of expectation of a random variable $\log(P_{SY}(s, y)/P_S(s)P_Y(y))$:

$$I(S, Y) = E\left[\log \frac{P_{SY}(s, y)}{P_S(s) P_Y(y)}\right]. \tag{5}$$

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 joint density of the input S and the output Y.

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 [20] pseudodistance measure $I(S, Y) = \sum_s \sum_y P_{SY}(s, y)\log(P_{SY}(s, y)/P_S(s)P_Y(y))$. Then Jensen's inequality implies that $I(S, Y) \geq 0$. 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.

B. Noisy Threshold Neuron

We use the discrete-time threshold neuron model [12], [29], [39], [44], [45]

$$y_t = \text{sgn}(s_t + n_t - \theta) = \begin{cases} 1 & \text{if } s_t + n_t \geq \theta \\ -1 & \text{if } s_t + n_t < \theta \end{cases} \quad (6)$$

where $\theta > 0$ is the neuron's threshold, s_t is the bipolar input Bernoulli signal (with success probability 1/2) with amplitude $A > 0$, and n_t is the additive white noise with probability density $p(n)$. Experiments with other success probabilities near 1/2 did not produce substantially different simulation results.

C. Noisy Continuous Neurons

We use the additive continuous neuron model with a neuronal signal function $S(x)$ [45]

$$\dot{x} = -x + S(x) + s(t) + n(t) \quad (7)$$
$$y(t) = \text{sgn}(x(t)). \quad (8)$$

Here $s(t)$ and $n(t)$ are the input and additive noise of the neuron and $y(t)$ is the binary output. The neuron feeds its output signal $S(x)$ back to itself and emits the threshold bipolar signal $y(t)$ as output.

- **Hyperbolic Tangent** This signal function gives an additive neuron model that is bistable [2], [10], [15], [39], [40], [45]

$$S(x) = 2\tanh x. \quad (9)$$

- **Linear-Threshold** This simple linear-threshold signal function [45] also gives the SR effect in the neuron

$$S(x) = \begin{cases} cx & |cx| < 1 \\ 1 & cx > 1 \\ -1 & cx < -1 \end{cases} \quad (10)$$

for a constant $c > 0$. We use $c = 2$.
- **Exponential** This signal function is asymmetric with the form [45]

$$S(x) = \begin{cases} 1 - \exp\{-cx\} & x > 0 \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

for a constant $c > 0$. We use $c = 10$.
- **Gaussian**. The Gaussian or "radial basis" signal function [45] differs from the other signal functions above because it is nonmonotonic

$$S(x) = \exp\{-cx^2\} \quad (12)$$

for a constant $c > 0$. We use $c = 100$.

III. MUTUAL INFORMATION OF THE THRESHOLD NEURON WITH BIPOLAR INPUT SIGNALS

A. SR in Memoryless Threshold Neurons

This section derives analytical SR results for the noisy threshold neuron based on the marginal probability density function of the output $P_Y(y)$ and the conditional density $P_{Y|S}(y|s)$. The system is the binary neuron with a fixed threshold θ. The bipolar (Bernoulli with success probability p) input signal s_t has amplitude A: $s_t \in \{-A, A\}$ with probability density $P_S(s)$. The noise n_t adds to the signal s_t before it enters the neuron. So the neuron's output y_t has the form (6). Fig. 5 plots the mutual information $I(S, Y)$ for four standard closed-form noise probability density functions 18, 24, 29, and 38. The central result is a theorem that holds for almost all noise probability densities so long as the mean noise falls outside a user-controlled interval that depends on the threshold θ.

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

$$P_{Y|S}(0|0) = \Pr\{s + n < \theta\}|_{s=-A}$$
$$= \Pr\{n < \theta + A\} = \int_{-\infty}^{\theta+A} p(n)dn \quad (13)$$

$$P_{Y|S}(1|0) = 1 - P_{Y|S}(0|0) \quad (14)$$
$$P_{Y|S}(0|1) = \Pr\{s + n < \theta\}|_{s=A}$$
$$= \Pr\{n < \theta - A\} = \int_{-\infty}^{\theta-A} p(n)dn \quad (15)$$

$$P_{Y|S}(1|1) = 1 - P_{Y|S}(0|1) \quad (16)$$

and the marginal density is

$$P_Y(y) = \sum_s P_{Y|S}(y|s)P_S(s). \quad (17)$$

Researchers have derived the conditional probabilities $P_{Y|S}(y|s)$ of the threshold system with *Gaussian* noise with bipolar inputs [12] and Gaussian inputs [67]. We next derive $P_{Y|S}(y|s)$ for uniform, Laplace, and (infinite-variance) Cauchy noise as well. Fig. 3 shows four examples of the unimodal noise densities and their realizations. Then we introduce stable distributions to model a spectrum of impulsive noise types.

- **Gaussian Noise** The Gaussian density with zero mean and variance $\sigma_n^2 = \sigma^2$ has the form

$$p(n) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{n^2}{2\sigma^2}\right\}. \quad (18)$$

Then the conditional probabilities $P_{Y|S}(y|s)$ are

$$P_{Y|S}(0|0) = \int_{-\infty}^{\theta+A} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{n^2}{2\sigma^2}\right\} dn$$
$$= \frac{1}{2} + \frac{1}{2}\text{erf}\left(\frac{\theta + A}{\sigma\sqrt{2}}\right) \quad (19)$$

$$P_{Y|S}(1|0) = \frac{1}{2} - \frac{1}{2}\text{erf}\left(\frac{\theta + A}{\sigma\sqrt{2}}\right) \quad (20)$$
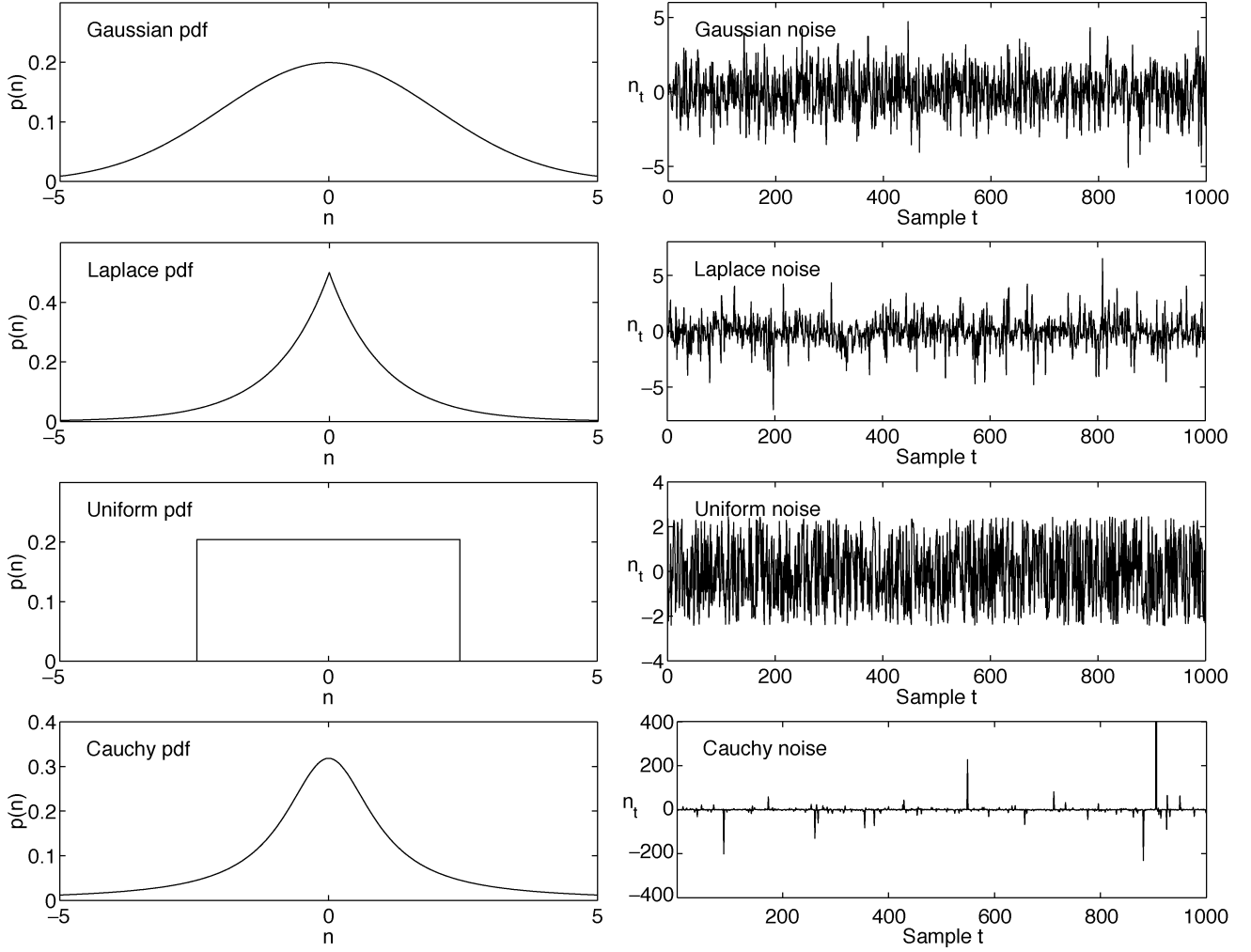
Fig. 3. Probability density functions and sample realizations. The figure shows Gaussian, Laplace, and uniform random variables n with zero mean and variance of two: $E[n] = 0$ and $E[n^2] = \sigma^2 = 2$. The Cauchy density function has zero location and unit dispersion but infinite variance. The pseudorandom number generators in [65] act as noise sources for these probability densities.

$$P_{Y|S}(0|1) = \frac{1}{2} + \frac{1}{2}\mathrm{erf}\left(\frac{\theta - A}{\sigma\sqrt{2}}\right) \qquad (21)$$

$$P_{Y|S}(1|1) = \frac{1}{2} - \frac{1}{2}\mathrm{erf}\left(\frac{\theta - A}{\sigma\sqrt{2}}\right). \qquad (22)$$

The error function erf is

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp\{-t^2\}dt. \qquad (23)$$

- **Uniform Noise** The uniform density with zero mean and variance $\sigma_n^2 = a^2/12$ has the form

$$p(n) = \begin{cases} \frac{1}{a} & \text{if } -\frac{a}{2} < n < \frac{a}{2} \\ 0 & \text{otherwise.} \end{cases} \qquad (24)$$

Then the conditional probabilities $P_{Y|S}(y|s)$ are

$$P_{Y|S}(0|0) = \begin{cases} 1 & \text{if } \frac{a}{2} < \theta + A \\ \frac{1}{2} + \frac{A+\theta}{a} & \text{otherwise} \end{cases}$$

$$= \min\left\{1, \frac{1}{2} + \frac{\theta + A}{a}\right\} \qquad (25)$$

$$P_{Y|S}(1|0) = \max\left\{0, \frac{1}{2} - \frac{\theta + A}{a}\right\} \qquad (26)$$

$$P_{Y|S}(0|1) = \min\left\{1, \frac{1}{2} + \frac{\theta - A}{a}\right\} \qquad (27)$$

$$P_{Y|S}(1|1) = \max\left\{0, \frac{1}{2} - \frac{\theta - A}{a}\right\}. \qquad (28)$$

- **Laplace Noise** The Laplace density with zero mean and variance $\sigma_n^2 = 2\beta^2$ has the form

$$p(n) = \frac{1}{2\beta} \exp\left\{-\left|\frac{n}{\beta}\right|\right\}. \qquad (29)$$

Then the conditional probabilities $P_{Y|S}(y|s)$ are

$$P_{Y|S}(0|0) = 1 - \frac{1}{2} \exp\left\{-\frac{\theta + A}{\beta}\right\} \qquad (30)$$

$$P_{Y|S}(1|0) = \frac{1}{2} \exp\left\{-\frac{\theta + A}{\beta}\right\} \qquad (31)$$

$$P_{Y|S}(0|1) = 1 - \frac{1}{2} \exp\left\{-\frac{\theta - A}{\beta}\right\} \qquad (32)$$

$$P_{Y|S}(1|1) = \frac{1}{2} \exp\left\{-\frac{\theta - A}{\beta}\right\}. \qquad (33)$$

- **Cauchy Noise** The Cauchy density with zero location and finite dispersion γ (but infinite variance) has the form

$$p(n) = \frac{1}{\pi} \frac{\gamma}{n^2 + \gamma^2}. \tag{34}$$

Then the conditional probabilities $P_{Y|S}(y|s)$ are

$$P_{Y|S}(0|0) = \frac{1}{2} + \frac{1}{\pi} \tan^{-1} \frac{\theta + A}{\gamma} \tag{35}$$

$$P_{Y|S}(1|0) = \frac{1}{2} - \frac{1}{\pi} \tan^{-1} \frac{\theta + A}{\gamma} \tag{36}$$

$$P_{Y|S}(0|1) = \frac{1}{2} + \frac{1}{\pi} \tan^{-1} \frac{\theta - A}{\gamma} \tag{37}$$

$$P_{Y|S}(1|1) = \frac{1}{2} - \frac{1}{\pi} \tan^{-1} \frac{\theta - A}{\gamma}. \tag{38}$$

- **Symmetric Alpha-Stable Noise: Thick-Tailed Bell Curves**

We 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 [6], [28], [35], [62]. The parameter α controls tail thickness and lies in $0 < \alpha \leq 2$. 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 [6] 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 contains "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 [46], [62]. Note that the best choice of α is an *empirical* question for bell-curve phenomena. Bell-curve behavior alone does not justify the assumption of the Gaussian bell curve.

Fig. 4 shows realizations of four symmetric alpha-stable random variables. A general alpha-stable probability density function f has characteristic function φ [1], [5], [35], [62]

$$\varphi(\omega) = \exp\left\{ia\omega - \gamma|\omega|^\alpha \left(1 + i\beta \text{sign}(\omega) \tan \frac{\alpha\pi}{2}\right)\right\} \quad \text{for } \alpha \neq 1 \tag{39}$$

and

$$\varphi(\omega) = \exp\left\{ia\omega - \gamma|\omega| \left(1 - \frac{2i\beta \ln|\omega| \text{sign}(\omega)}{\pi}\right)\right\} \quad \text{for } \alpha = 1 \tag{40}$$

where

$$\text{sign}(\omega) = \begin{cases} 1 & \text{if } \omega > 0 \\ 0 & \text{if } \omega = 0 \\ -1 & \text{if } \omega < 0 \end{cases} \tag{41}$$

and $i = \sqrt{-1}$, $0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$, 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 shown 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 alpha-stable densities for most α's. Numerical integration of φ gives the probability densities in Fig. 4.

The following theorem shows that noisy threshold neurons produce some SR effect for almost all noise probability descriptions. The proof shows that if $I(S, Y) > 0$ then eventually the mutual information $I(S, Y)$ tends toward zero as the noise variance or dispersion tends toward zero. So the mutual information $I(S, Y)$ must increase as the noise variance increases from zero. The crucial assumption is that the noise mean $E[n]$ (or location parameter) not lie in the signal-threshold interval $(\theta - A, \theta + A)$.

Theorem: Suppose that the threshold signal system (6) 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. Suppose that $a \notin (\theta - A, \theta + A)$ for the location parameter a of an alpha-stable noise density with characteristic function (39), (40). Then the threshold system (6) exhibits the nonmonotone SR effect in the sense that $I(S, Y) \to 0$ as $\sigma \to 0$ or $\gamma \to 0$.

Proof: Assume $0 < P_S(s) < 1$ to avoid triviality when $P_S(s) = 0$ or 1. We show that S and Y are asymptotically independent: $I(S, Y) \to 0$ as $\sigma \to 0$ (or as $\gamma \to 0$). Recall that $I(S, Y) = 0$ if and only if S and Y are statistically independent [20]. So we need to show only that $P_{SY}(s, y) = P_S(s)P_Y(y)$ or $P_{Y|S}(y|s) = P_Y(y)$ as $\sigma \to 0$ (or as $\gamma \to 0$) for all signal symbols $s \in S$ and $y \in Y$. The two-symbol alphabet set S gives

$$P_Y(y) = \sum_s P_{Y|S}(y|s)P_S(s) \tag{42}$$

$$= P_{Y|S}(y|0)P_S(0) + P_{Y|S}(y|1)P_S(1) \tag{43}$$

$$= P_{Y|S}(y|0)P_S(0) + P_{Y|S}(y|1)(1 - P_S(0)) \tag{44}$$

$$= \left(P_{Y|S}(y|0) - P_{Y|S}(y|1)\right) P_S(0) + P_{Y|S}(y|1). \tag{45}$$

So we need to show only that $P_{Y|S}(y|0) - P_{Y|S}(y|1) = 0$ as $\sigma \to 0$ (or as $\gamma \to 0$). This condition implies that $P_Y(y) = P_{Y|S}(y|1)$ and $P_Y(y) = P_{Y|S}(y|0)$. We assume for simplicity that the noise density $p(n)$ is integrable. The argument below still holds if $p(n)$ is discrete and if we replace integrals with appropriate sums.

Consider $y = $ '0'. Then (13) and (15) imply that

$$P_{Y|S}(0|0) - P_{Y|S}(0|1) = \int_{-\infty}^{\theta+A} p(n)dn - \int_{-\infty}^{\theta-A} p(n)dn \tag{46}$$
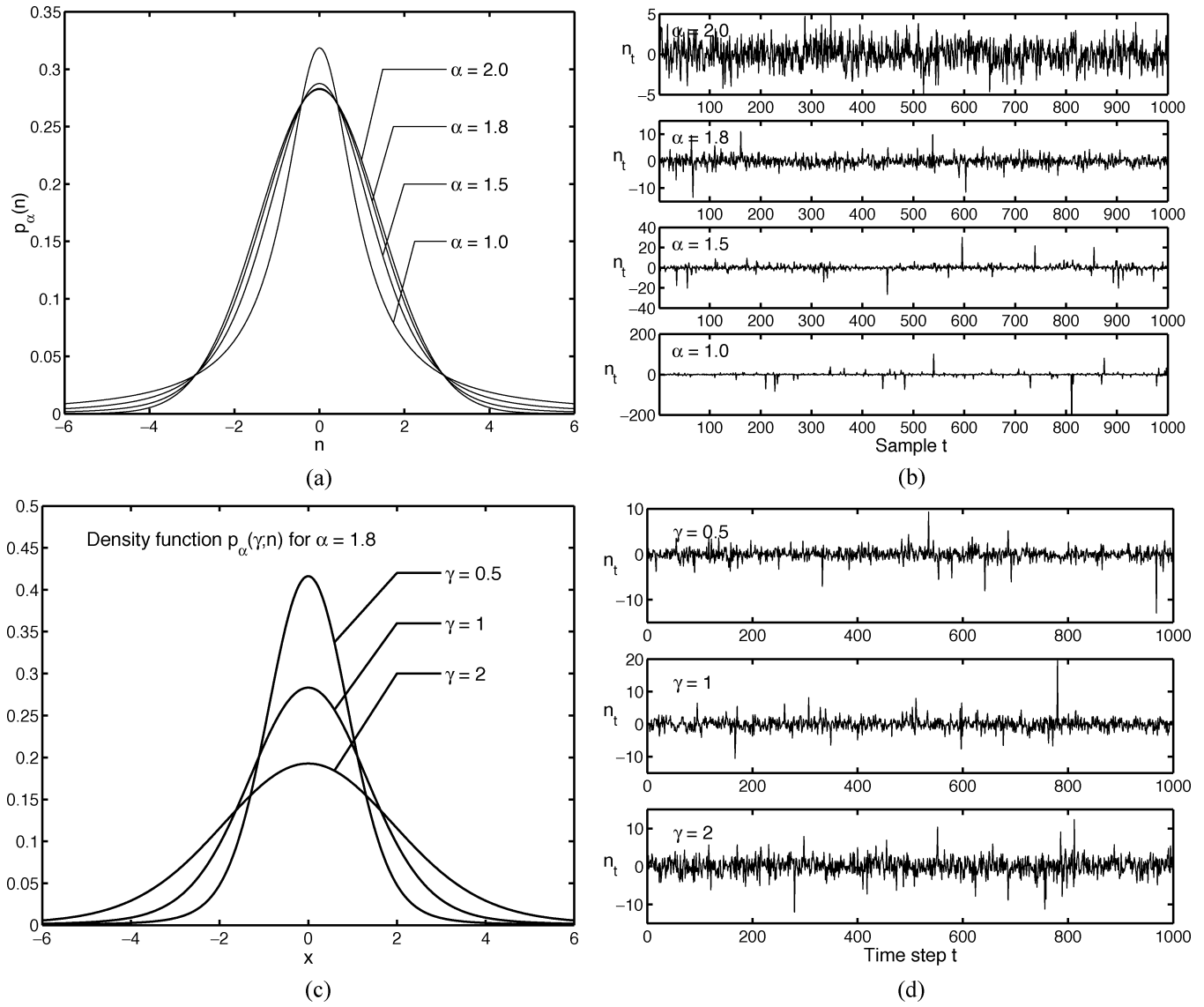
Fig. 4. Samples of standard symmetric alpha-stable probability densities and their realizations. (a) Density functions with zero location ($a = 0$) and unit dispersion ($\gamma = 1$) for $\alpha = 2$, 1.8, 1.5, and 1. The densities are bell curves that have thicker tails as α decreases and thus that model increasingly impulsive noise as α decreases. The case $\alpha = 2$ gives a Gaussian density with variance two (or unit dispersion). The parameter $\alpha = 1$ gives the Cauchy density. (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 noise n becomes more impulsive as the parameter α falls. The algorithm in [13], [68] generates 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.

$$= \int_{\theta-A}^{\theta+A} p(n)dn. \qquad (47)$$

Similarly for $y = $ '1'

$$P_{Y|S}(1|0) = \int_{\theta+A}^{\infty} p(n)dn \qquad (48)$$

$$P_{Y|S}(1|1) = \int_{\theta-A}^{\infty} p(n)dn. \qquad (49)$$

Then

$$P_{Y|S}(1|0) - P_{Y|S}(1|1) = -\int_{\theta-A}^{\theta+A} p(n)dn. \qquad (50)$$

The result now follows if we can show that

$$\int_{\theta-A}^{\theta+A} p(n)dn \to 0 \qquad \text{as } \sigma \to 0 \text{ or } \gamma \to 0. \qquad (51)$$

Case 1) Finite-variance noise. Let the mean of the noise be $m = E[n]$ and the variance be $\sigma^2 = E[(n-m)^2]$. Then $m \notin (\theta - A, \theta + A)$ by hypothesis.

Now suppose that $m < \theta - A$. Pick $\varepsilon = (1/2)d(\theta - A, m) = (1/2)(\theta - A - m) > 0$. So $\theta - A - \varepsilon = \theta - A - \varepsilon + m - m = m + (\theta - A - m) - \varepsilon = m + 2\varepsilon - \varepsilon = m + \varepsilon$. Then

$$P_{Y|S}(0|0) - P_{Y|S}(0|1)$$

$$= \int_{\theta-A}^{\theta+A} p(n)dn \qquad (52)$$

$$\leq \int_{\theta - A}^{\infty} p(n)dn \tag{53}$$

$$\leq \int_{\theta - A - \varepsilon}^{\infty} p(n)dn \tag{54}$$

$$= \int_{m+\varepsilon}^{\infty} p(n)dn \tag{55}$$

$$= \Pr\{n \geq m + \varepsilon\} = \Pr\{n - m \geq \varepsilon\} \tag{56}$$

$$\leq \Pr\{|n - m| \geq \varepsilon\} \tag{57}$$

$$\leq \frac{\sigma^2}{\varepsilon^2} \quad \text{by Chebyshev's inequality} \tag{58}$$

$$\to 0 \quad \text{as } \sigma \to 0. \tag{59}$$

Suppose next that $m > \theta + A$. Then pick $\varepsilon = (1/2)d(\theta + A, m) = (1/2)(m - \theta - A) > 0$ and so $\theta + A + \varepsilon = \theta + A + \varepsilon + m - m = m - (m - \theta - A) + \varepsilon = m - 2\varepsilon + \varepsilon = m - \varepsilon$. Then

$$P_{Y|S}(0|0) - P_{Y|S}(0|1)$$

$$= \int_{\theta - A}^{\theta + A} p(n)dn \tag{60}$$

$$\leq \int_{-\infty}^{\theta + A} p(n)dn \tag{61}$$

$$\leq \int_{-\infty}^{\theta + A + \varepsilon} p(n)dn \tag{62}$$

$$= \int_{-\infty}^{m - \varepsilon} p(n)dn \tag{63}$$

$$= \Pr\{n \leq m - \varepsilon\} = \Pr\{n - m \leq -\varepsilon\} \tag{64}$$

$$\leq \Pr\{|n - m| \geq \varepsilon\} \tag{65}$$

$$\leq \frac{\sigma^2}{\varepsilon^2} \quad \text{by Chebyshev's inequality} \tag{66}$$

$$\to 0 \quad \text{as } \sigma \to 0. \tag{67}$$

Case 2) Impulsive noise: Alpha-stable noise. The characteristic function $\varphi(\omega)$ of alpha-stable density $p(n)$ has the exponential form (39), (40). This reduces to a simple complex exponential in the zero-dispersion limit

$$\lim_{\gamma \to 0} \varphi(\omega) = \exp\{ia\omega\} \tag{68}$$

for *all* α, skewness β, and location a. So Fourier transformation gives the corresponding density function in the limiting case ($\gamma \to 0$) as a translated delta function

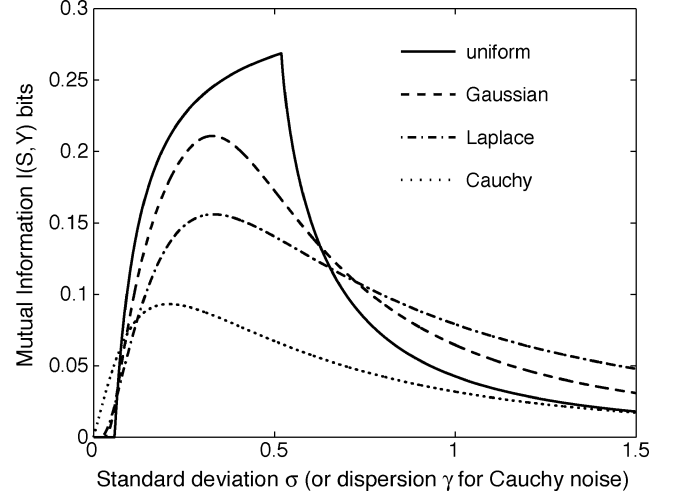$$\lim_{\gamma \to 0} p(n) = \delta(n - a). \tag{69}$$

Fig. 5. Mutual information I profiles of a threshold system with bipolar input for four kinds of noise. The system has threshold $\theta = 0.5$. The input Bernoulli signal is bipolar with amplitude $A = 0.4$.

Then

$$P_{Y|S}(0|0) - P_{Y|S}(0|1) = \int_{\theta - A}^{\theta + A} p(n)dn \tag{70}$$

$$= \int_{\theta - A}^{\theta + A} \delta(n - a)dn \tag{71}$$

$$= 0 \tag{72}$$

because $a \notin (\theta - A, \theta + A)$.

Then $P_Y(y) = P_{Y|S}(y|s)$ as $\gamma \to 0$. So Cases 1 and 2 imply that $I(S, Y) \to 0$ as $\sigma \to 0$ for finite-variance noise or as $\gamma \to 0$ for alpha-stable noise.   Q.E.D.

B. Theoretical Results for Closed-Form Noise Densities

Inserting Gaussian or other specific closed-form conditional probability densities $P_{Y|S}(y|s)$ from (19)–(38) into (1)–(4) gives exact solutions of the mutual information $I(S, Y)$ as a function of the noise parameter σ. Fig. 5 shows I-versus-σ profiles of a threshold system with four kinds of noise: Gaussian, uniform, Laplace, and Cauchy. The I profile of the uniform noise has the highest peak among the four noise densities for the same system (same threshold θ and same input amplitude A). And the I profile has a distinct shape: it drops sharply after it reaches its peak as σ grows. Gaussian noise gives the second highest I while Cauchy gives the lowest. The threshold system requires different optimal standard deviations (or dispersions) for different kinds of noise.

The closed form of the I versus σ profiles in Fig. 5 also allows a direct analysis of how the optimal noise depends on the signal amplitude A for Gaussian, uniform, Laplace, and Cauchy noise. Suppose the signal amplitude A is a subthreshold input in a noisy threshold neuron with threshold θ: $A < \theta$. Then will the optimal noise σ_{opt} (or γ_{opt}) decrease as the signal amplitude A moves closer to the threshold θ?

Intuition might suggest that the threshold system should need less noise to produce the entropic SR effect as the amplitude
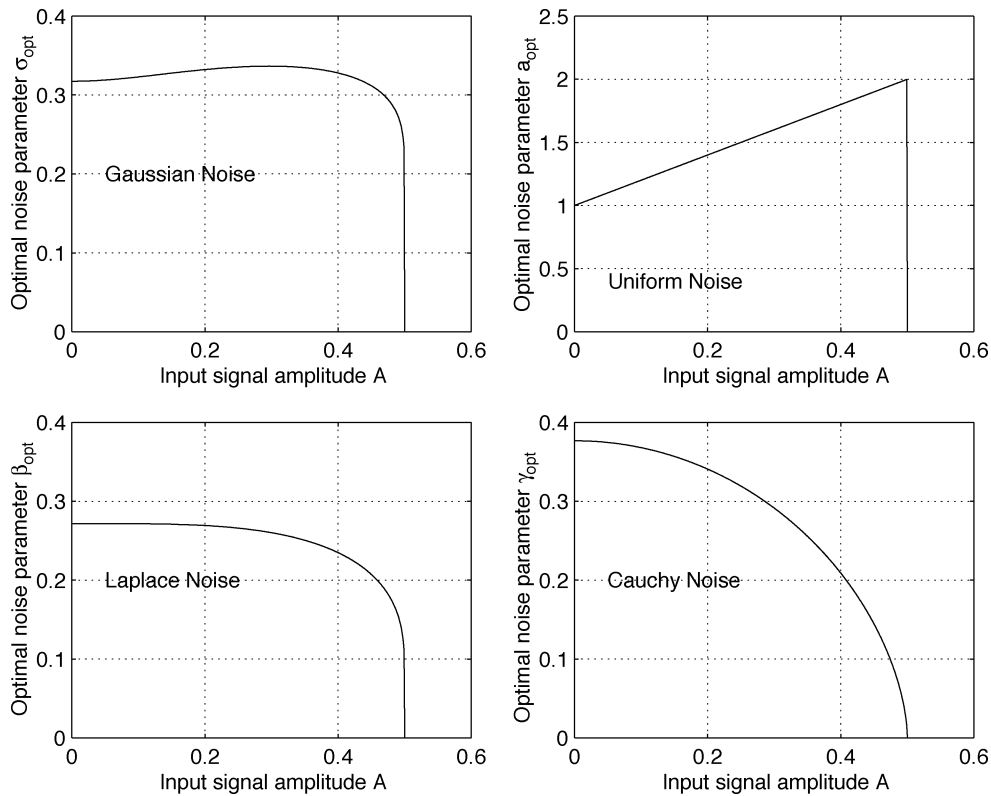
Fig. 6. Optimal SR noise schedules for a noisy threshold neuron with threshold $\theta = 0.5$. The schedules show how optimal noise variance or dispersion depends on signal amplitude A for the four closed-form noise results in Fig. 5.

moves closer to the threshold θ. But the results in Fig. 6 show that the compound nonlinearities involved produce no such simple relationship. The different noise types produce different SR optimality schedules. Fig. 6 shows four optimal noise schedules for the threshold value $\theta = 0.5$. Other threshold values produced similar results. Only optimal Laplace and Cauchy noise produce the more intuitive monotone decrease in the optimal noise level with rising signal amplitude A. Optimal uniform noise grows linearly with signal amplitude while optimal Gaussian noise defines a nonmonotonic schedule.

IV. STOCHASTIC RESONANCE IN COMPUTER SIMULATIONS

Discrete simulations can model continuous-time nonlinear dynamical systems if a stochastic numerical scheme approximates the system dynamics and its signal and noise response. We used a simple stochastic version of the Euler scheme to model a nonlinear system with input forcing signal and noise. We measured how the system performed based on only the system's input-output samples.

Consider the forced dynamical system with additive forcing input signal s and "white" noise n

$$\dot{x} = f(x) + s(t) + n(t) \qquad (73)$$
$$y(t) = g\left(x(t)\right). \qquad (74)$$

These models simply add a noise term to a differential equation rather than use formal Ito or Stratonovich stochastic differentials [14], [24], [31]. "Whiteness" of a random variable n here means that n is white only over some large but finite frequency bandwidth interval $[-B, B]$ for some large $B > 0$. Random num-

bers from the algorithms in [13], [65], [68] act as noise from various probability densities in our simulations. The next sections show how discretized continuous-time systems produced the discrete-time systems we used for computer simulations.

A. Nonlinear Systems With White Gaussian Noise

Consider the dynamical system (73) with initial condition $x(t_0) = x_0$. Here the white Gaussian noise w has zero mean and unit variance so that $n = \sigma w$ has zero mean and variance σ^2. This system corresponds to the stochastic initial value problem [31]

$$dX = \tilde{f}(t, X) + \sigma(t, X)dW \qquad (75)$$

for initial condition $X(t_0) = X_0$. Here $\tilde{f}(t, X) = f(X) + s(t)$, $\sigma(t, X) = \sigma$, and W is the standard Wiener process [31]. We used Euler's method (the Euler-Maruyama scheme) [21], [31], [42] to obtain the discrete form for computer simulation

$$x_{t+1} = x_t + \Delta T \left(f(x_t) + s_t\right) + \sigma\sqrt{\Delta T}w_t \qquad (76)$$
$$y_t = g(x_t) \qquad (77)$$

for $t = 0, 1, 2, \dots$ and initial condition x_0. The input sample s_t has the value of the signal $s(t\Delta T)$ at time step t. The zero-mean white Gaussian noise sequence $\{w_t\}$ has unit variance $\sigma_w^2 = 1$. The term $\sqrt{\Delta T}$ scales w_t so that $\sqrt{\Delta T}w_t$ conforms with the Wiener increment [31], [42], [59]. The output sample y_t is some transformation g of the system's state x_t.

This simple algorithm gives fairly accurate results for moderate nonlinear systems [31], [42], [51], [59]. Other algorithms may give more accurate numerical solutions of the stochastic

differential equations for more complicated system dynamics [31], [54]. All of our simulations used the Euler's scheme in (76), (77).

The numerical algorithm in [65] generates a sequence of pseudo-random numbers from a Gaussian density with zero mean and unit variance for $\{w_t\}$ in (76). Fig. 3 shows the Gaussian and other densities that have zero mean and a variance of two.

B. Nonlinear Systems With Other Finite-Variance Noise

We next consider a system (73) with finite-variance noise n. Suppose the noise n has variance σ^2 and again apply Euler's method

$$x_{t+1} = x_t + \Delta T \left(f(x_t) + s_t \right) + \sigma \sqrt{\Delta T} w_t \qquad (78)$$

$$y_{t+1} = g(x_{t+1}). \qquad (79)$$

Here the random sequence $\{w_t\}$ has density function $p(w)$ with zero mean and unit variance. The numerical algorithms in [65] generate sequences of random variables for Laplace and uniform density functions. Fig. 3 plots these probability density functions and their realizations with mean zero and variance of two: $E[n] = 0$ and $E[n^2] = 2$.

C. Nonlinear Systems With Alpha-Stable Noise

Figs. 3 and 4 show realizations of the symmetric alpha-stable random variable for several characteristic exponents α. Again we assume that the Euler's method above applies to this class of random variables with infinite variance. Let w be a standard alpha-stable random variable with parameter α and zero location and unit dispersion: $a = 0$ and $\gamma = 1$. Let $\kappa = \gamma^{1/\alpha}$ denote a "scale" factor of a random variable. Then $n = \kappa w$ has zero location and dispersion $\gamma = \kappa^\alpha$. This leads to the Euler's numerical solution

$$x_{t+1} = x_t + \Delta T \left(f(x_t) + s_t \right) + \kappa \sqrt{\Delta T} w_t \qquad (80)$$

$$y_t = g(x_t). \qquad (81)$$

The algorithm in [13], [68] generates a standard alpha-stable random variable w.

V. DERIVATION OF SR LEARNING LAW

We show that a memoryless neuron can use stochastic gradient ascent to learn the SR effect [47], [57]

$$\sigma_{k+1} = \sigma_k + \mu_k \frac{\partial I}{\partial \sigma}. \qquad (82)$$

We assume that $P(s)$ does not depend on σ and we use the natural logarithm. Then the learning term $\partial I / \partial \sigma$ has the form

$$\frac{\partial I}{\partial \sigma} = \frac{\partial}{\partial \sigma} \left(-\sum_y P(y) \log P(y) \right.$$

$$\left. + \sum_s P(s) \sum_y P(y|s) \log P(y|s) \right) \qquad (83)$$

$$= -\sum_y \left(P(y) \frac{1}{P(y)} \frac{\partial P(y)}{\partial \sigma} + \log P(y) \frac{\partial P(y)}{\partial \sigma} \right)$$

$$+ \sum_s \sum_y \left(P(s) P(y|s) \frac{1}{P(y|s)} \frac{\partial P(y|s)}{\partial \sigma} \right.$$

$$\left. + P(s) \log P(y|s) \frac{\partial P(y|s)}{\partial \sigma} \right) \qquad (84)$$

$$= -\sum_y \left(\frac{\partial P(y)}{\partial \sigma} + \log P(y) \frac{\partial P(y)}{\partial \sigma} \right)$$

$$+ \sum_s \sum_y \left(P(s) \frac{\partial P(y|s)}{\partial \sigma} \right.$$

$$\left. + P(s) \log P(y|s) \frac{\partial P(y|s)}{\partial \sigma} \right). \qquad (85)$$

The sum $\sum_y P(y) = 1$ implies $\sum_y (\partial P(y)/\partial \sigma) = (\partial/\partial \sigma) \sum_y P(y) = 0$. And $\sum_s \sum_y (\partial P(y|s)/\partial \sigma) = 0$ because $\sum_y P(y|s) = 1$. So

$$\frac{\partial I}{\partial \sigma} = -\sum_y \log P(y) \frac{\partial P(y)}{\partial \sigma} + \sum_s \sum_y P(s) \log P(y|s) \frac{\partial P(y|s)}{\partial \sigma}. \qquad (86)$$

We estimate the partial derivative with a ratio of time differences and replace the denominator with the signum function to avoid numerical instability

$$\frac{\partial P(y)}{\partial \sigma} \approx \frac{P_k(y) - P_{k-1}(y)}{\sigma_k - \sigma_{k-1}}$$

$$\approx \text{sgn}(\sigma_k - \sigma_{k-1}) \left[P_k(y) - P_{k-1}(y) \right] \qquad (87)$$

$$\frac{\partial P(y|s)}{\partial \sigma} \approx \frac{P_k(y|s) - P_{k-1}(y|s)}{\sigma_k - \sigma_{k-1}}$$

$$\approx \text{sgn}(\sigma_k - \sigma_{k-1}) \left[P_k(y|s) - P_{k-1}(y|s) \right] \qquad (88)$$

where $P_k(y)$ is the marginal density function of the output Y at time k and $P_k(y|s)$ is the conditional density function at time k. Then the learning term becomes

$$\frac{\partial I}{\partial \sigma} \approx \text{sgn}(\sigma_k - \sigma_{k-1}) \left(-\sum_y \left[P_k(y) - P_{k-1}(y) \right] \log P_k(y) \right.$$

$$\left. + \sum_s \sum_y P_k(s) \left[P_k(y|s) - P_{k-1}(y|s) \right] \log P_k(y|s) \right). \qquad (89)$$

Our previous work [47], [57] on adaptive SR found through statistical tests that the random learning term $\partial P / \partial \sigma$ had an approximately Cauchy distribution for the spectral signal-to-noise and cross-correlation performance measures P. These frequent and energetic Cauchy impulse spikes destabilized the stochastic learning process. So we "robustified" the learning term with the standard Cauchy error suppressor $\phi(z_k) = 2z_k/(1 + z_k^2)$ [37], [41]. This included the threshold neuron given a periodic input sequence.

But detailed simulations revealed a special pattern in the case of mutual information: The density $P_k(y)$ tends to stay close to the past density $P_{k-1}(y)$ if the values of σ_k and σ_{k-1} are close. This causes the learning paths σ_k to converge quickly near the
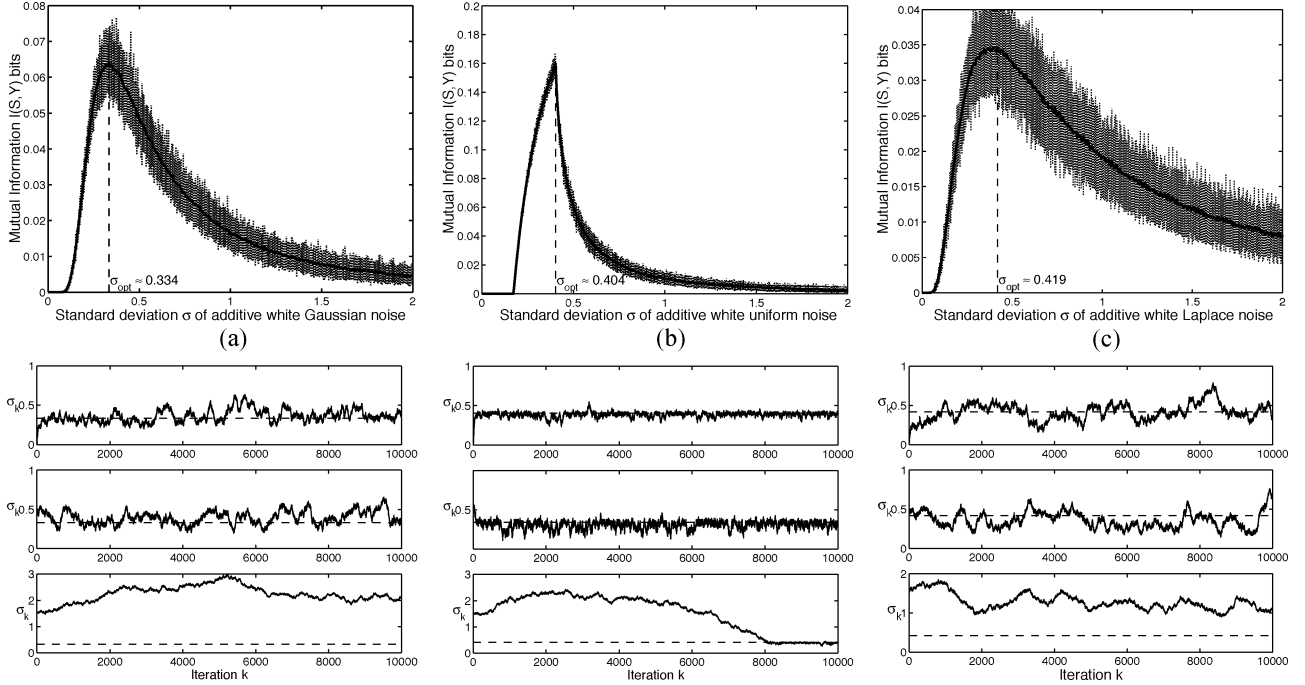
Fig. 7. Finite-variance noise cases: adaptive stochastic resonance for the noisy threshold neuron (6) with bipolar input signal s_t, amplitude $A = 0.2$, and threshold $\theta = 0.5$. The additive noise are (a) Gaussian, (b) uniform, and (c) Laplace. The graphs at the top show the nonmonotonic signatures of SR. The sample paths at the bottom plots show the convergence if the initial condition σ_0 is close to the optimal noise level σ_{opt}. Distant initial conditions may lead to divergence as the third learning path in (a) shows. The constant learning rates are $\mu_k = 0.01$ for Gaussian and uniform noise and $\mu_k = 0.02$ for Laplace noise.

initial conditions. So we can replace the learning term $\partial I/\partial \sigma$ with its sign $\text{sgn}(\partial I/\partial \sigma)$ and the learning law simplifies to

$$\sigma_{k+1} = \sigma_k + \mu_k \text{sgn}\left(\frac{\partial I}{\partial \sigma}\right). \tag{90}$$

The signum is a simple robustifier and formally consistent with a two-sided Laplacian distribution [37].

VI. SIMULATION RESULTS

We tested the robust learning law in (90) with the approximation of the learning term in (89). We needed to estimate the marginal and conditional probability densities $P_k(y), P_k(s)$, and $P_k(y|s)$ at each iteration k. So at each k we collected 1000 input-output samples $\{s_t, y_t\}$ and used them to estimate the densities with histograms for the threshold system. We used 500 of the input-output symbols to estimate the probability densities for the continuous neuron model. We chose the neurons' and signals' parameters below to demonstrate the algorithm. Other parameters gave similar results.

A. Noisy Threshold Neuron

The threshold neuron had a fixed threshold $\theta = 0.5$. The bipolar input Bernoulli signal has probability $P_S(-A) = P_S(A) = 1/2$ where the amplitude A varied from $A = 0.1$ to $A = 0.4$ (subthreshold inputs). We tried several noise densities that included the Gaussian, uniform, Laplace, and the impulsive alpha-stable densities that include the Cauchy density. All noise densities had zero mean (zero location for Cauchy). We tried to learn the optimal standard deviation σ_{opt} (or optimal dispersion γ_{opt} for alpha-stable noise). We used constant learning rates

$\mu_k = 0.01$ for Gaussian and uniform noise, $\mu_k = 0.02$ for Laplace and Cauchy noise, and $\mu_k = 0.02$ for alpha-stable noise with $\alpha = 1.9$ and $\alpha = 1.5$. We started the learning from several initial conditions with different noise seeds.

Figs. 7–9 show the adapted SR profiles and the σ_{opt} learning paths for different noise types. The learning paths converged to the optimal standard deviation σ_{opt} (or dispersion γ_{opt}) if the initial value was near σ_{opt}. The learning paths tended to stay nearer the optimal values for larger input amplitudes.

B. Noisy Continuous Neuron

We used the discrete model in Section IV for simulations. We used $dt = 0.01$ s and let each input symbol stay for 50 s. So for each input symbol we presented the corresponding "spikes" (plus noise) 5000 times to the neuron. And we collected 5000 discrete-time output "spikes" and averaged them to get the output symbol. This procedure applied to all types of signal functions and for all types of noise.

1) Continuous Neurons with Hyperbolic Tangent Signal Function: We tested the continuous neuron model with hyperbolic tangent signal function with several noise densities such as the Gaussian, uniform, Laplace, and alpha-stable (which included the Cauchy density). All noise densities had zero mean (zero location for Cauchy). The bipolar input Bernoulli signal had success probability $P_S(-A) = P_S(A) = 1/2$ where the amplitude A varied from $A = 0.1$ to $A = 0.4$ (subthreshold inputs). We used constant learning rates $\mu_k = 0.03$ for Gaussian, uniform, and Laplace noise. We used the smaller learning rates $\mu_k = 0.02$ for alpha-stable noise with $\alpha = 1.9$ and $\alpha = 1.5$ and used the still smaller learning rate $\mu_k = 0.005$ for Cauchy noise. We started the learning from several initial
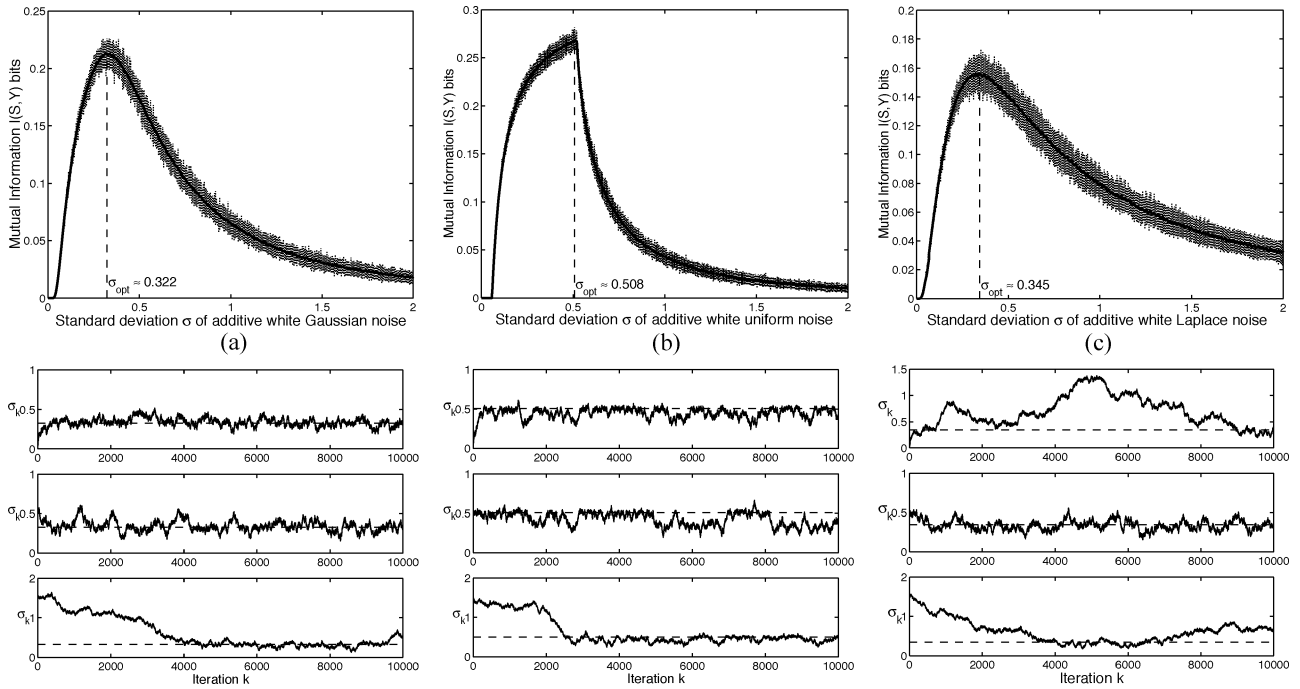
Fig. 8. Finite-variance noise cases: adaptive stochastic resonance for the noisy threshold neuron (6) with bipolar input signal s_t, amplitude $A = 0.4$, and threshold $\theta = 0.5$. The additive noise are (a) Gaussian, (b) uniform, and (c) Laplace. The graphs at the top show the nonmonotonic signatures of SR. The sample paths at the bottom plots show the convergence of the noise standard deviation σ_k to the noise optimum σ_{opt} for each noise density. The constant learning rates are $\mu_k = 0.01$ for Gaussian and uniform noise and $\mu_k = 0.02$ for Laplace noise.
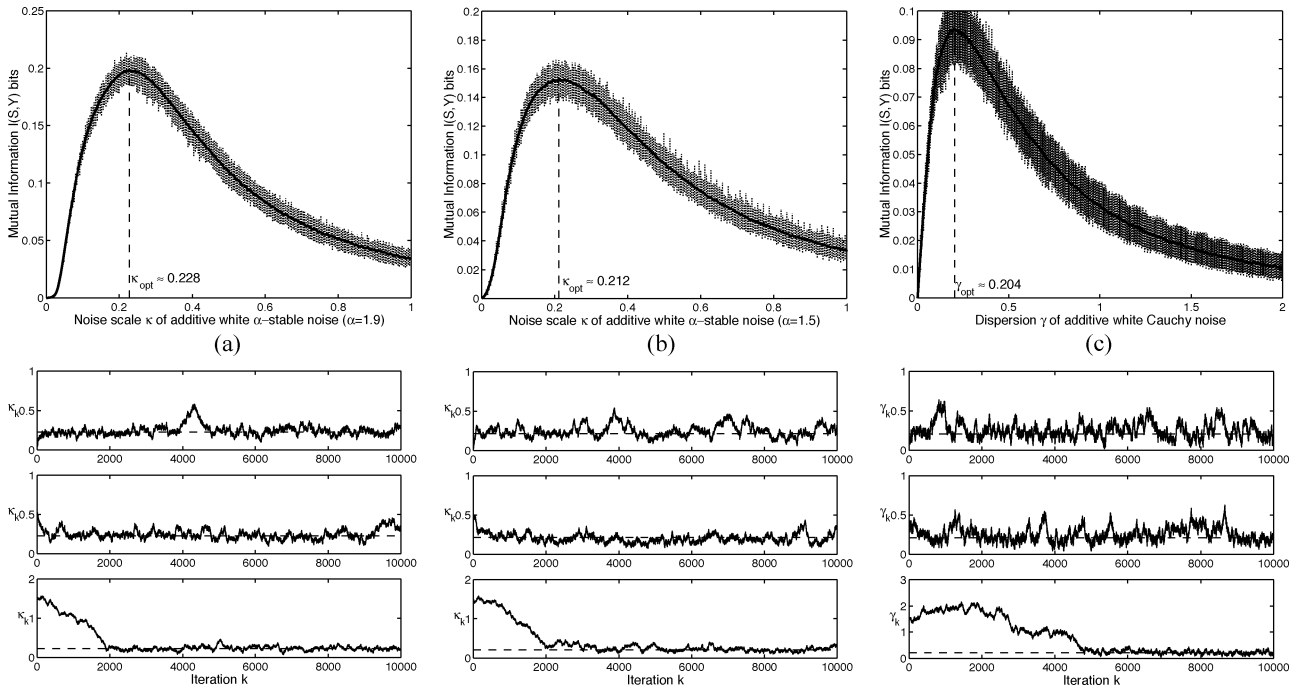

Fig. 9. Impulsive noise cases: Adaptive stochastic resonance for the noisy threshold neuron (6) with bipolar input signal s_t, amplitude $A = 0.4$, and threshold $\theta = 0.5$. The additive noise are α-stable distributed with the parameter (a) $\alpha = 1.9$, (b) $\alpha = 1.5$, and (c) $\alpha = 1$ or Cauchy density. The graphs at the top show the nonmonotonic signatures of SR. The sample paths at the bottom plots show the convergence of the noise scale κ_k to the noise optimum κ_{opt} for each noise density. The corresponding dispersions are $\gamma = \kappa^\alpha$ for each α-stable noise. The constant learning rates are $\mu_k = 0.01$ for $\alpha = 1.9$ and $\alpha = 1.5$ noise and $\mu_k = 0.02$ for Cauchy noise $\alpha = 1$.

conditions with different noise seeds. Figs. 10–12 show the adapted SR profiles and the σ_{opt} learning paths for different noise types. The learning paths converged near the optimal standard deviation σ_{opt} (or dispersion γ_{opt}) if the initial value was near σ_{opt}.

2) Continuous Neurons with Linear-Threshold, Exponential, and Gaussian (Radial Basis) Signal Functions: We further tested the continuous neuron model with linear-threshold, exponential, and Gaussian (radial basis) signal functions in Gaussian noise to show the generality of the SR effect. We
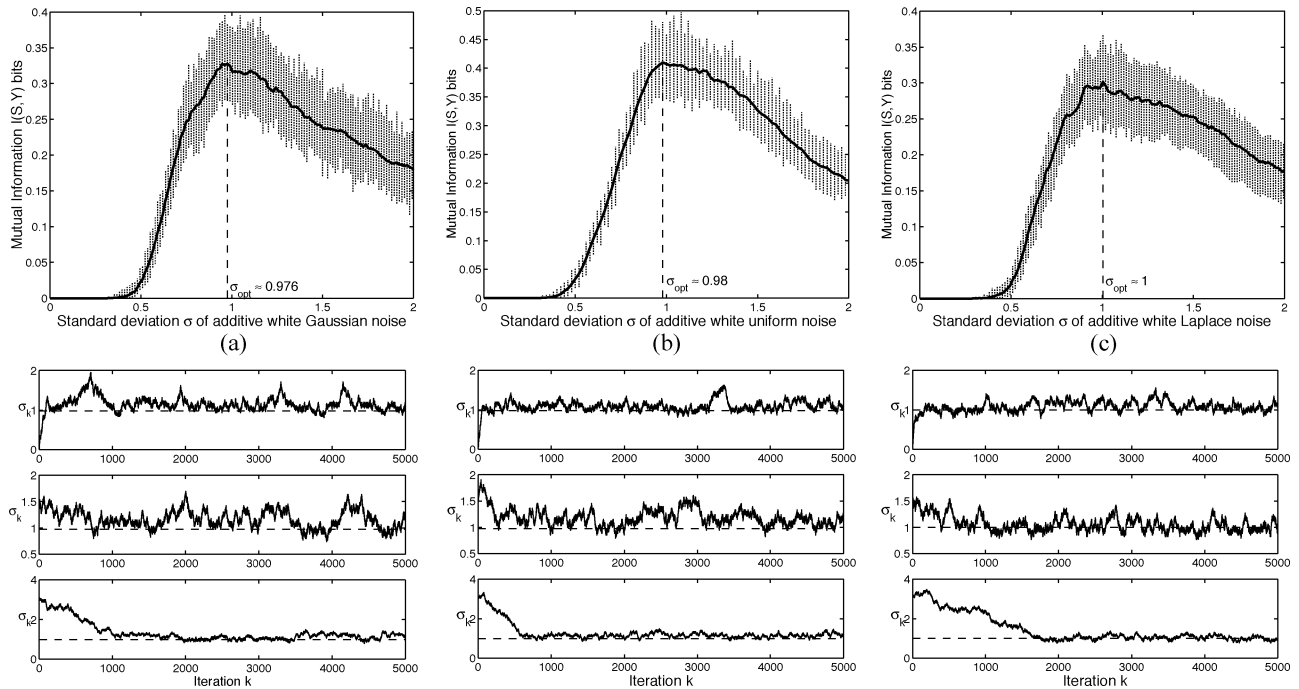
Fig. 10. Finite-variance noise cases: Adaptive stochastic resonance for the noisy continuous neuron (7) with hyperbolic signal function (9) and bipolar input signal s_t with amplitude $A = 0.2$. The additive noise are (a) Gaussian, (b) uniform, and (c) Laplace. The graphs at the top show the nonmonotonic signatures of SR. The sample paths at the bottom plots show the convergence of the noise standard deviation σ_k to the noise optimum σ_{opt} for each noise density. The constant learning rates are $\mu_k = 0.03$ for all cases.
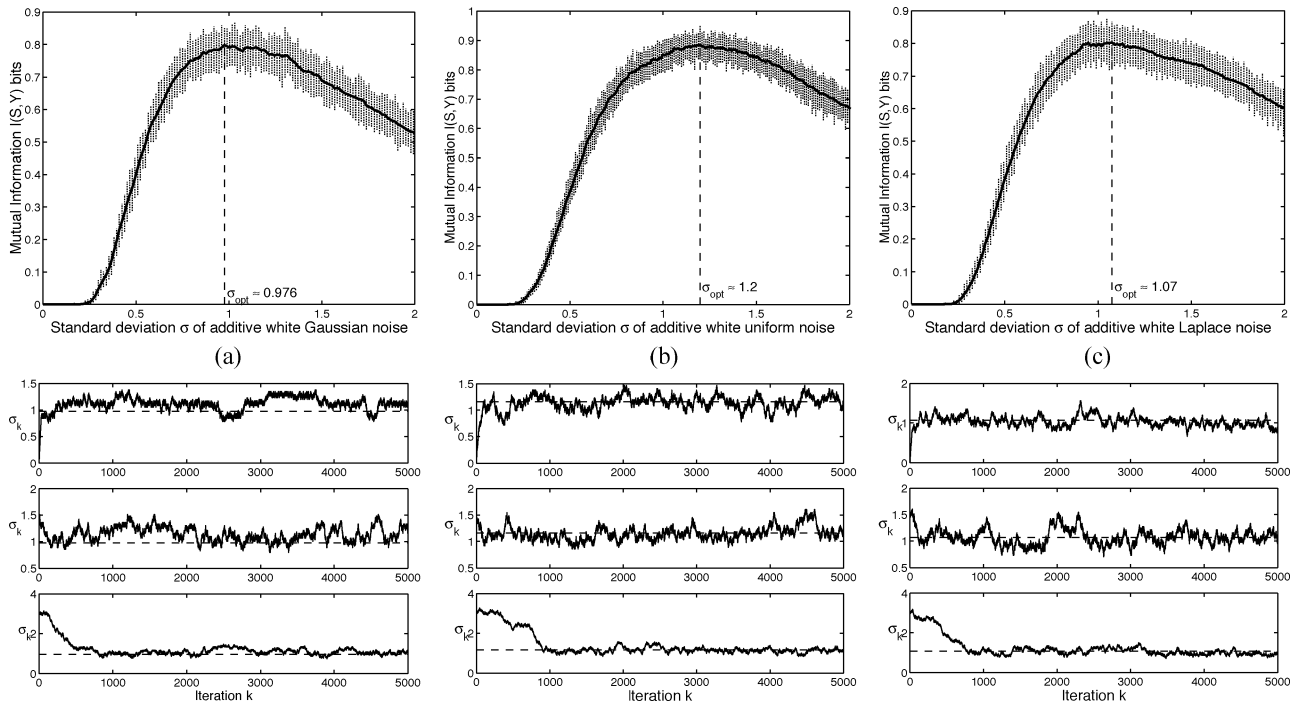

Fig. 11. Finite-variance noise cases: Adaptive stochastic resonance for the noisy continuous neuron (7) with hyperbolic signal function (9) and bipolar input signal s_t with amplitude $A = 0.4$. The additive noise are (a) Gaussian, (b) uniform, and (c) Laplace. The graphs at the top show the nonmonotonic signatures of SR. The sample paths at the bottom plots show the convergence of the noise standard deviation σ_k to the noise optimum σ_{opt} for each noise density. The constant learning rates are $\mu_k = 0.03$ for all cases.

used the same bipolar input Bernoulli signal with success probability $P_S(-A) = P_S(A) = (1/2)$ where the amplitude is $A = 0.4$ for the linear-threshold and Gaussian signal functions and $A = 0.6$ for the exponential signal function. The input amplitudes were "subthreshold" for the neuron models

with these signal functions. We used constant learning rates $\mu_k = 0.02$ for the exponential and Gaussian signal functions and $\mu_k = 0.05$ for the linear-threshold signal functions. We started the learning from several initial conditions with different noise seeds. Fig. 13 shows the adapted SR profiles and the σ_{opt}
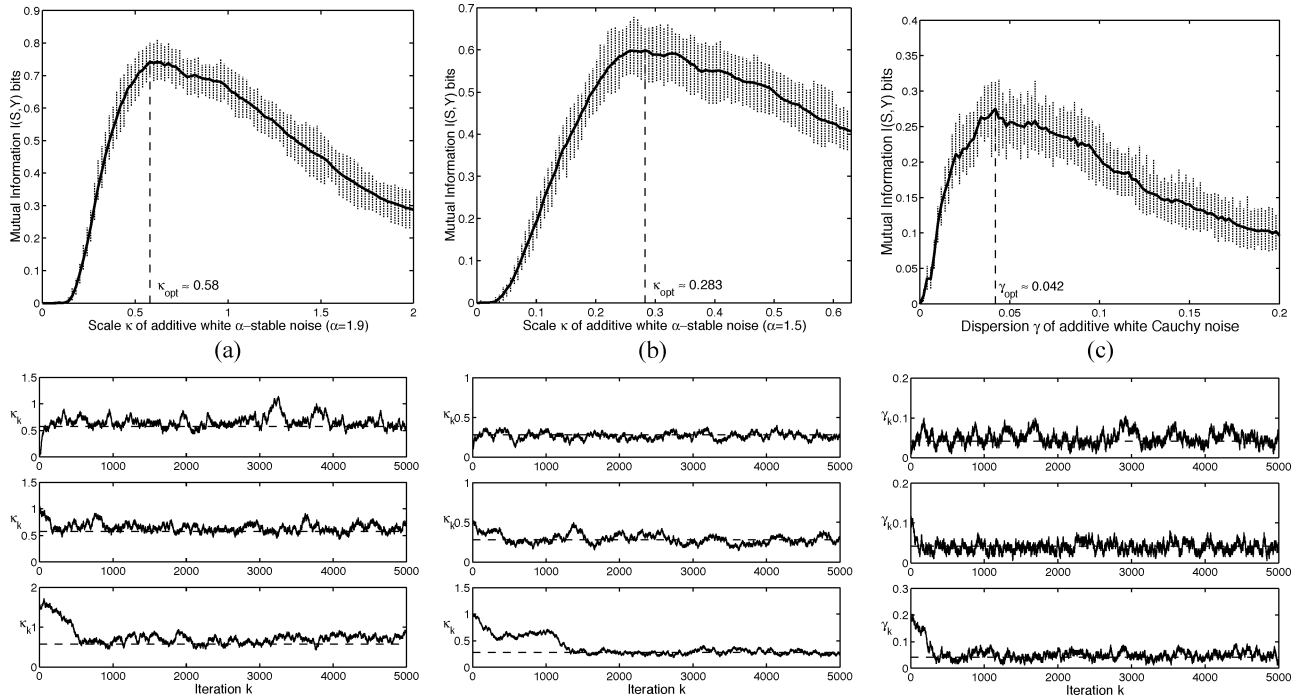
Fig. 12.   Impulsive noise cases: Adaptive stochastic resonance for the noisy continuous neuron (7) with hyperbolic signal function (9) and bipolar input signal s_t with amplitude $A = 0.4$. The additive noise are α-stable distributed with the parameter (a) $\alpha = 1.9$, (b) $\alpha = 1.5$, and (c) $\alpha = 1$ or Cauchy density. The graphs at the top show the nonmonotonic signatures of SR. The sample paths at the bottom show the convergence of the noise standard deviation σ_k to the noise optimum σ_{opt} for each noise density. The constant learning rates are $\mu_k = 0.02$ for $\alpha = 1.9$, $\mu_k = 0.01$ for $\alpha = 1.5$, and $\mu_k = 0.005$ for $\alpha = 1$.
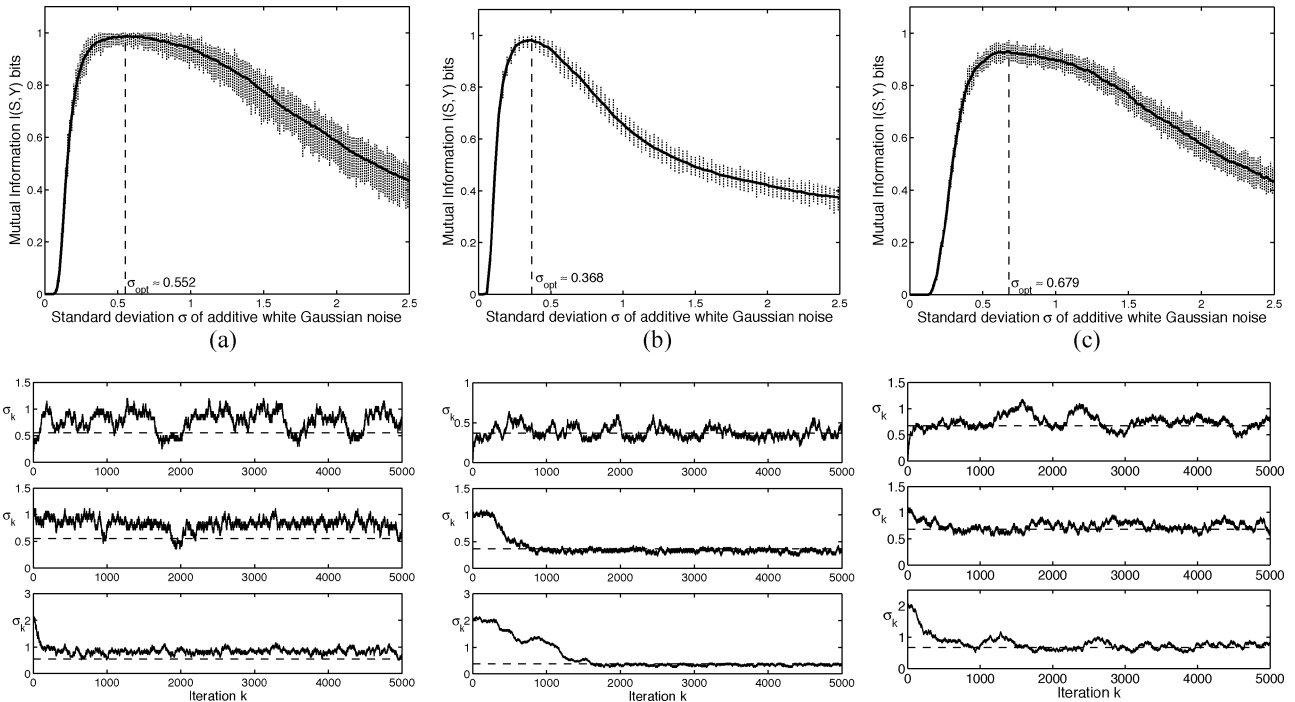

Fig. 13.   Adaptive stochastic resonance for continuous neurons with linear-threshold, exponential, and Gaussian (radial basis) signal functions. The bipolar input signal s_t has amplitude $A = 0.4$ for the linear-threshold and Gaussian signal functions and $A = 0.6$ for the exponential signal function. The additive noise n_t is Gaussian. The graphs at the top show the nonmonotonic signatures of SR. The sample paths at the bottom show the convergence of the noise standard deviation σ_k to the noise optimum σ_{opt} for each case of signal functions: (a) linear-threshold, (b) exponential, and (c) Gaussian. The constant learning rates are $\mu_k = 0.05$ for the linear-threshold signal function and $\mu_k = 0.02$ for the exponential and Gaussian signal functions.

learning paths for the three other signal functions. The learning paths converged near the optimal standard deviation σ_{opt} if the initial value was near σ_{opt}.

VII. CONCLUSION

Threshold neurons exhibit stochastic resonance—they increase their throughput mutual information when faint input

noise increases in intensity. A theorem shows that this holds for almost all noise densities. Such noise-based information maximization is consistent with Linsker's principle of information maximization in neural networks [49], [50]. Closed-form noise densities allow us to derive the exact dependence of mutual information on noise dispersion and to observe the nonlinear relationships between the optimal noise level and the magnitude of the input signal amplitude. Extensive simulations confirmed this entropic SR effect for noisy threshold (memoryless) neurons and for simple continuous neurons.

A simple robust stochastic learning law can find the entropically optimal noise level for both threshold and continuous neurons that process noisy bipolar input signals. This result holds for many types of finite-variance and infinite-variance (impulsive) noise. These noise types can model energetic disturbances that range from thermal jitter to unmodeled environmental effects to the random crosstalk of neurons in large neural networks. This robust finding supports the implicit SR conjecture that biological neurons [11], [18], [19], [23], [48], [58], [64], [69] have evolved over genetic eons to exploit the noise energy freely available in their local environment.

REFERENCES

[1] V. Akgiray and C. G. Lamoureux, "Estimation of stable-law parameters: a comparative study," *J. Bus. Econ. Statist.*, vol. 7, pp. 85–93, Jan. 1989.

[2] S. Amari, "Neural theory of association and concept formation," *Biol. Cybern.*, vol. 26, pp. 175–185, 1977.

[3] R. Benzi, G. Parisi, A. Sutera, and A. Vulpiani, "Stochastic resonance in climatic change," *Tellus*, vol. 34, pp. 10–16, 1982.

[4] R. Benzi, A. Sutera, and A. Vulpiani, "The mechanism of stochastic resonance," *J. Phys. A, Math. Gen.*, vol. 14, pp. L453–L457, 1981.

[5] H. Bergstrom, "On some expansions of stable distribution functions," *Arkiv för Matematik*, vol. 2, pp. 375–378, 1952.

[6] L. Breiman, *Probability*. Reading, MA: Addison-Wesley, 1968.

[7] P. Bryant, K. Wiesenfeld, and B. McNamara, "The nonlinear effects of noise on parametric amplification: an analysis of noise rise in Josephson junctions and other systems," *J. Appl. Phys.*, vol. 62, no. 7, pp. 2898–2913, Oct. 1987.

[8] A. R. Bulsara, T. C. Elston, C. R. Doering, S. B. Lowen, and K. Lindenberg, "Cooperative behavior in periodically driven noisy integrate-fire models of neuronal dynamics," *Phys. Rev. E*, vol. 53, no. 4, pp. 3958–3969, Apr. 1996.

[9] A. R. Bulsara and L. Gammaitoni, "Tuning in to noise," *Phys. Today*, pp. 39–45, Mar. 1996.

[10] A. R. Bulsara, E. W. Jacobs, T. Zhou, F. Moss, and L. Kiss, "Stochastic resonance in a single neuron model: theory and analog simulation," *J. Theoret. Biol.*, vol. 152, pp. 531–555, 1991.

[11] A. R. Bulsara, A. J. Maren, and G. Schmera, "Single effective neuron: dendritic coupling effects and stochastic resonance," *Biolog. Cybern.*, vol. 70, pp. 145–156, 1993.

[12] A. R. Bulsara and A. Zador, "Threshold detection of wideband signals: a noise-induced maximum in the mutual information," *Phys. Rev. E*, vol. 54, no. 3, pp. R2185–R2188, Sept. 1996.

[13] J. M. Chambers, C. L. Mallows, and B. W. Stuck, "A method for simulating stable random variables," *J. Amer. Statist. Assoc.*, vol. 71, no. 354, pp. 340–344, 1976.

[14] K. L. Chung and R. J. Williams, *Introduction to Stochastic Integration*, 2nd ed, Germany: Birkhäuser, 1990.

[15] M. A. Cohen and S. Grossberg, "Absolute stability of global pattern formation and parallel memory storage by competitive neural networks," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-13, pp. 815–826, Sept. 1983.

[16] J. J. Collins, C. C. Chow, A. C. Capela, and T. T. Imhoff, "A periodic stochastic resonance," *Phys. Rev. E*, vol. 54, no. 5, pp. 5575–5584, Nov. 1996.

[17] J. J. Collins, C. C. Chow, and T. T. Imhoff, "Stochastic resonance without tuning," *Nature*, vol. 376, pp. 236–238, July 1995.

[18] J. J. Collins, T. T. Imhoff, and P. Grigg, "Noise-enhanced information transmission in Rat SA1 cutaneous mechanoreceptors via aperiodic stochastic resonance," *J. Neurophysiol.*, vol. 76, no. 1, pp. 642–645, July 1996.

[19] ——, "Noise-enhanced tactile sensation," *Nature*, vol. 383, p. 770, Oct. 1996.

[20] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[21] G. Dahlquist and Å. Björck, *Numerical Methods*. Englewood Cliffs, NJ: Prentice-Hall, 1974.

[22] G. Deco and B. Schürmann, "Stochastic resonance in the mutual information between input and output spike trains of noisy central neurons," *Phys. D*, vol. 117, pp. 276–282, 1998.

[23] J. K. Douglass, L. Wilkens, E. Pantazelou, and F. Moss, "Noise enhancement of information transfer in crayfish mechanoreceptors by stochastic resonance," *Nature*, vol. 365, pp. 337–340, Sept. 1993.

[24] R. Durrett, *Stochastic Calculus: A Practical Introduction*: CRC Press, 1996.

[25] M. I. Dykman, T. Horita, and J. Ross, "Statistical distribution and stochastic resonance in a periodically driven chemical system," *J. Chem. Phys.*, vol. 103, no. 3, pp. 966–972, July 1995.

[26] J.-P. Eckmann and L. E. Thomas, "Remarks on stochastic resonances," *J. Phys. A: Math. General*, vol. 15, pp. L261–L266, 1982.

[27] S. Fauve and F. Heslot, "Stochastic resonance in a bistable system," *Phys. Lett. A*, vol. 97, no. 1, 2, pp. 5–7, Aug. 1983.

[28] W. Feller, *An Introduction to Probability Theory and Its Applications*. New York: Wiley, 1966, vol. II.

[29] L. Gammaitoni, "Stochastic resonance and the dithering effect in threshold physical systems," *Phys. Rev. E*, vol. 52, no. 5, pp. 4691–4698, Nov. 1995.

[30] ——, "Stochastic resonance in multi-threshold systems," *Phys. Lett. A*, vol. 208, pp. 315–322, Dec. 1995.

[31] T. C. Gard, *Introduction to Stochastic Differential Equations*. New York: Marcel Dekker, 1988.

[32] J. Glanz, "Mastering the nonlinear brain," *Science*, vol. 277, pp. 1758–1760, Sept. 1997.

[33] X. Godivier and F. Chapeau-Blondeau, "Stochastic resonance in the information capacity of a nonlinear dynamic system," *Int. J. Bifurc. Chaos*, vol. 8, no. 3, pp. 581–589, 1998.

[34] M. Grifoni, M. Sassetti, P. Hänggi, and U. Weiss, "Cooperative effects in the nonlinearly driven spin-boson system," *Phys. Rev. E*, vol. 52, no. 4, pp. 3596–3607, Oct. 1995.

[35] M. Grigoriu, *Applied Non-Gaussian Processes*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[36] A. Hibbs, E. W. Jacobs, J. Bekkedahl, A. R. Bulsara, and F. Moss, "Signal enhancement in a r.f. SQUID using stochastic resonance," *Il Nuovo Cimento*, vol. 17 D, no. 7–8, pp. 811–817, Luglio–Agosto 1995.

[37] R. V. Hogg and A. T. Craig, *Introduction to Mathematical Statistics*, 5th ed. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[38] N. Hohn and A. N. Burkitt, "Enhanced stochastic resonance in threshold detectors," in *Proc. 2001 IEEE Int. Joint Conf. on Neural Networks (IJCNN '01)*, 2001, pp. 644–647.

[39] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," in *Proc. National Acad. Sci.*, vol. 79, 1982, pp. 2554–2558.

[40] ——, "Neural networks with graded response have collective computational properties like those of two-state neurons," in *Proc. National Acad. Science*, vol. 81, 1984, pp. 3088–3092.

[41] P. J. Huber, *Robust Statistics*. New York: Wiley, 1981.

[42] M. E. Inchiosa and A. R. Bulsara, "Nonlinear dynamic elements with noisy sinusoidal forcing: Enhancing response via nonlinear coupling," *Phys. Rev. E*, vol. 52, no. 1, pp. 327–339, July 1995.

[43] M. E. Inchiosa, J. W. C. Robinson, and A. R. Bulsara, "Information-theoretic stochastic resonance in noise-floor limited systems: The case for adding noise," *Phys. Rev. Lett.*, vol. 85, pp. 3369–3372, Oct. 2000.

[44] P. Jung, "Stochastic resonance and optimal design of threshold detectors," *Phys. Lett. A*, vol. 207, pp. 93–104, Oct. 1995.

[45] B. Kosko, *Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence*. Englewood Cliffs, NJ: Prentice-Hall, 1992.

[46] ——, *Fuzzy Engineering*. Englewood Cliffs, NJ: Prentice-Hall, 1996.

[47] B. Kosko and S. Mitaim, "Robust stochastic resonance: Signal detection and adaptation in impulsive noise," *Phys. Rev. E*, vol. 64, no. 051 110, October 2001.

[48] J. E. Levin and J. P. Miller, "Broadband neural encoding in the cricket cercal sensory system enhanced by stochastic resonance," *Nature*, vol. 380, pp. 165–168, Mar. 1996.

[49] R. Linsker, "Self-organization in a perceptual network," *Computer*, vol. 21, no. 3, pp. 105–117, Mar. 1988.

[50] ——, "A local learning rule that enables information maximization for arbitrary input distributions," *Neural Comput.*, vol. 9, no. 8, pp. 1661–1665, Nov. 1997.

[51] A. Longtin, "Autonomous stochastic resonance in bursting neurons," *Phys. Rev. E*, vol. 55, no. 1, pp. 868–876, Jan. 1997.

[52] J. Maddox, "Toward the brain-computer's code?," *Nature*, vol. 352, p. 469, Aug. 1991.

[53] ——, "Bringing more order out of noisiness," *Nature*, vol. 369, p. 271, May 1994.

[54] R. Mannella and V. Palleschi, "Fast and precise algorithm for computer simulation of stochastic differential equations," *Phys. Rev. A*, vol. 40, no. 6, pp. 3381–3386, Sept. 1989.

[55] R. J. Marks II, B. Thompson, M. A. El-Sharkawi, W. L. J. Fox, and R. T. Miyamoto, "Stochastic resonance of a threshold detector: image visualization and explanation," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS 2002)*, vol. 4, 2002, pp. IV-521–IV-523.

[56] B. McNamara, K. Wiesenfeld, and R. Roy, "Observation of stochastic resonance in a ring laser," *Phys. Rev. Lett.*, vol. 60, no. 25, pp. 2626–2629, June 1988.

[57] S. Mitaim and B. Kosko, "Adaptive stochastic resonance," *Proc. IEEE: Special Issue on Intelligent Signal Processing*, vol. 86, pp. 2152–2183, Nov. 1998.

[58] R. P. Morse and E. F. Evans, "Enhancement of vowel coding for cochlear implants by addition of noise," *Nature Medicine*, vol. 2, no. 8, pp. 928–932, Aug. 1996.

[59] F. Moss and P. V. E. McClintock, Eds., *Noise in Nonlinear Dynamical Systems*. Cambridge, UK: Cambridge Univ. Press, 1989, vol. I–III.

[60] D. C. Munson, "A note on Lena," *IEEE Trans. Image Processing*, vol. 5, p. 3, Jan. 1996.

[61] C. Nicolis, "Stochastic aspects of climatic transitions-response to a periodic forcing," *Tellus*, vol. 34, pp. 1–9, 1982.

[62] C. L. Nikias and M. Shao, *Signal Processing With Alpha-Stable Distributions and Applications*. New York: Wiley, 1995.

[63] X. Pei, K. Bachmann, and F. Moss, "The detection threshold, noise and stochastic resonance in the Fitzhugh-Nagumo neuron model," *Phys. Lett. A*, vol. 206, pp. 61–65, Oct. 1995.

[64] X. Pei, L. Wilkens, and F. Moss, "Light enhances hydrodynamic signaling in the multimodal caudal photoreceptor interneurons of the crayfish," *J. Neurophysiol.*, vol. 76, no. 5, pp. 3002–3011, Nov. 1996.

[65] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge, UK: Cambridge Univ. Press, 1993.

[66] D. F. Russell, L. A. Willkens, and F. Moss, "Use of behavioral stochastic resonance by paddle fish for feeding," *Nature*, vol. 402, pp. 291–294, Nov. 1999.

[67] N. G. Stocks, "Information transmission in parallel threshold arrays," *Phys. Rev. E*, vol. 63, no. 041 114, 2001.

[68] P. Tsakalides and C. L. Nikias, "The Robust Covariation-Based MUSIC (ROC-MUSIC) algorithm for bearing estimation in impulsive noise environments," *IEEE Trans. Signal Processing*, vol. 44, pp. 1623–1633, July 1996.

[69] M. Usher and M. Feingold, "Stochastic resonance in the speed of memory retrieval," *Biolog. Cybern.*, vol. 83, pp. L11–L16, 2000.

[70] R. A. Wannamaker, S. P. Lipshitz, and J. Vanderkooy, "Stochastic resonance as dithering," *Phys. Rev. E*, vol. 61, no. 1, pp. 233–236, Jan. 2000.

[71] K. Wiesenfeld and F. Moss, "Stochastic resonance and the benefits of noise: From ice ages to crayfish and SQUIDs," *Nature*, vol. 373, pp. 33–36, Jan. 1995.

Sanya Mitaim received the B.Eng. degree in control engineering from King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand. He received the M.S. and Ph.D. degrees in electrical engineering from the University of Southern California, Los Angeles.

He is an Assistant Professor with the Department of Electrical Engineering, Faculty of Engineering, Thammasat University, Pathumthani, Thailand. His research interests include neural and fuzzy techniques in nonlinear signal processing and noise processing.

Bart Kosko received the degrees in philosophy, economics, applied mathematics, and electrical engineering.

He is a Professor of electrical engineering with the University of Southern California, Los Angeles. He is the author of the textbooks *Neural Networks and Fuzzy Systems* (Prentice-Hall: Englewood Cliffs, NJ, 1992), and *Fuzzy Engineering* (Prentice-Hall: Englewood Cliffs, NJ, 1997); the novel *Nanotime* (Avon: New York, 1997); the trade books *Fuzzy Thinking* (Hyperion: New York, 1993), and *Heaven in a Chip* (Random House: New York, 2000). He edited *Neural Networks for Signal Processing* (Prentice-Halll: Englewood Cliffs, NJ, 1992) and coedited (with Simon Haykin) *Intelligent Signal Processing* (IEEE/Wiley: New York, 2001).

Dr. Kosko is an elected governor of the International Neural Network Society and has chaired and co-chaired many neural and fuzzy system conferences.