Robust stochastic resonance: Signal detection and adaptation in impulsive noise

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*
[2]*Department of Electrical Engineering, Faculty of Engineering, Thammasat University Rangsit Campus, Klong Luang, Pathumthani 12121, Thailand*
(Received 23 October 2000; revised manuscript received 7 May 2001; published 22 October 2001)

Stochastic resonance (SR) occurs when noise improves a system performance measure such as a spectral signal-to-noise ratio or a cross-correlation measure. All SR studies have assumed that the forcing noise has finite variance. Most have further assumed that the noise is Gaussian. We show that SR still occurs for the more general case of impulsive or infinite-variance noise. The SR effect fades as the noise grows more impulsive. We study this fading effect on the family of symmetric α-stable bell curves that includes the Gaussian bell curve as a special case. These bell curves have thicker tails as the parameter α falls from 2 (the Gaussian case) to 1 (the Cauchy case) to even lower values. Thicker tails create more frequent and more violent noise impulses. The main feedback and feedforward models in the SR literature show this fading SR effect for periodic forcing signals when we plot either the signal-to-noise ratio or a signal correlation measure against the dispersion of the α-stable noise. Linear regression shows that an exponential law $\gamma_{opt}(\alpha)=cA^{\alpha}$ describes this relation between the impulsive index α and the SR-optimal noise dispersion γ_{opt}. The results show that SR is robust against noise ''outliers.'' So SR may be more widespread in nature than previously believed. Such robustness also favors the use of SR in engineering systems. We further show that an adaptive system can learn the optimal noise dispersion for two standard SR models (the quartic bistable model and the FitzHugh-Nagumo neuron model) for the signal-to-noise ratio performance measure. This also favors practical applications of SR and suggests that evolution may have tuned the noise-sensitive parameters of biological systems.

DOI: 10.1103/PhysRevE.64.051110 PACS number(s): 05.40.−a

I. IMPULSIVE NOISE AND STOCHASTIC RESONANCE

Most noise processes have infinite variance. This mathematical fact is almost trivial. Even most bell-curve probability densities do not have finite variance or any finite higher-order moments. Yet this fact finds scant expression in over a century of published research in science and engineering. A review of the published statistical research in any field shows a common practice. Most random models assume that the dispersion of a random variable equals its squared-error measure of variance. But other measures of dispersion may be finite while the variance measure is infinite. The popularity of the finite-variance assumption may attest to its usefulness in many cases. But that does not lessen its severity. The assumption persists even though such a squared-error term seldom exists in any formal generality and even though such a squared-error term is not robust against data ''outliers'' when it does exist. Celebrated examples of the finite-variance hypothesis range from the Heisenberg uncertainty principle in quantum mechanics to the least-squares regression framework that underlies statistical curve fitting and forecasting in fields as disparate as astronomy and sociology.

The presence of infinite variance in a random model does not itself nullify the model or count as some sort of stochastic *reductio ad absurdum*. Infinite variance does not imply that we lack all statistical knowledge about the position or momentum of a random particle or about the value of any random variable if we assume only that the random variable has a probability density function in the shape of a bell curve. Many infinite-variance bell curves are locally indistinguishable from the thinner-tailed Gaussian bell curve. Infinite-variance noise itself produces impulses of only finite magnitude. Nor does infinite variance imply that a real system must have infinite energy. This holds for the same reason that the use of a Gaussian bell curve in a model does not imply that axes extend to infinity in the real world. Other events can explain the presence of infinite variance. We may have measured the random dispersion involved with the wrong measure. We may have applied a good but approximate measure to extreme cases that lie outside the measure's particular structure. Or we may simply have used or encountered a bell curve that has thicker tails than a Gaussian bell curve has.

Stochastic resonance [1–13] offers a recent and stark example of the finite-variance assumption. A dynamical system stochastically resonates or shows the stochastic resonance (SR) effect when noise increases its signal-to-noise ratio or other system performance measure. Almost all SR research has assumed that the noise process is Gaussian and hence has finite variance. A few SR studies have explored uniform and other non-Gaussian but finite-variance noise-types [14–17]. The SR signature of a nonmonotonic signal-to-noise graph gives perhaps the best evidence of the universality of the finite-variance assumption in SR research. All SR studies plot the dynamical system's signal-to-noise ratio against either the variance or the standard deviation of the driving noise process. So the very notation excludes the presence of infinite variance. This practice rules out a vast set of possible SR scenarios and suggests that SR is not robust against noise outliers. The simulation results below show that the SR effect can indeed occur when infinite-variance noise drives nonlinear feedback and feedforward systems.

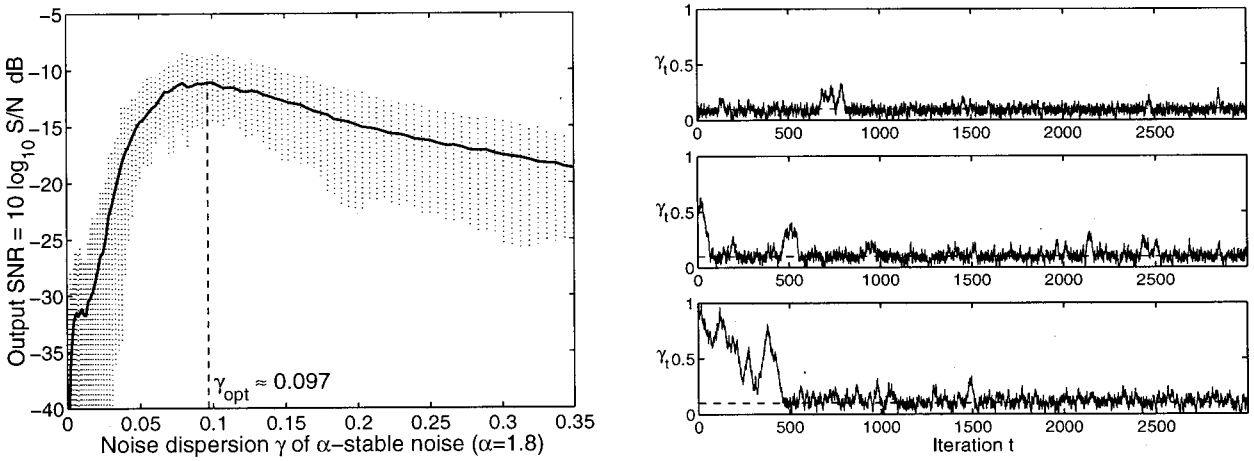

FIG. 1. Impulsive stochastic resonance SNR measure and convergence of learned dispersion to the SR effect. (a) The graph shows the smoothed output SNR as a function of the dispersion of additive α-stable (infinite-variance) noise with $\alpha=1.8$. The vertical dotted lines show the absolute deviation between the smallest and largest SNR outliers in each sample average of 100 outcomes. The vertical dashed line shows the optimal noise level at the dispersion $\gamma\approx 0.097$. The noisy signal-forced quartic bistable dynamical system has the form $\dot{x}=x-x^3+s+n$ with binary output $y(t)=\mathrm{sgn}(x(t))$. The α-stable noise $n(t)$ (with $\alpha=1.8$) adds to the external forcing narrow band signal $s(t)=0.1\sin 2\pi(0.01)t$. (b) Learning paths of γ_t with the Cauchy impulse suppressor $\phi(z)=2z/(1+z^2)$ for the quartic bistable system with sinusoidal input. The Cauchy impulse suppressor $\phi(\partial \mathrm{SNR}_t/\partial\sigma)$ replaces $\partial \mathrm{SNR}_t/\partial\sigma$ in the SR learning law [16] as in Eq. (43) below. The learning paths converged to and wander about the optimal noise dispersion $\gamma_{opt}\approx 0.097$.

Stochastic resonance occurs in a signal-forced dynamical system when noise improves its performance by increasing its signal-to-noise ratio (SNR) [18–22] or some other performance measure such as a signal cross correlation [23–27] or mutual entropy [25–27]. Then the noise process $n(t)$ and signal process $s(t)$ force a feedback dynamical system of the form $\dot{x}=f(x)$ to give $\dot{x}=f(x)+s(t)+n(t)$. The forced system's signal-to-noise ratio has the form $\mathrm{SNR}=S/N$ where S measures the spectral content of the forcing signal $s(t)$ in the forced system and N measures the spectral content of the noise $n(t)$ [as entangled with each other and with the system state dynamics $\dot{x}=f(x)$]. Most SR systems in the literature have assumed that the forcing signal has the simple periodic form of a sinusoid. Aperiodic SR [23,24] is an important exception that we do not consider here.

The figures show the main results of this research. Figure 1 shows an SR profile when the additive forcing impulsive noise has infinite variance. The noise has alpha value $\alpha=1.8$ and so the noise is only mildly impulsive compared to the noise that arises from bell curves with thicker tails. Figure 1 also shows the more complex result that a stochastic learning algorithm can learn to locate the SR-optimal dispersion value in this impulsive environment and do so based not on the functional form of the dynamical stable (the quartic bistable system in this case) but based on only input-output training samples of dispersion and SNR values. Each SNR value depends on the noise-corrupted system dynamics. This allows the learning process to in effect slowly estimate the system dynamics. The presence of system dynamics means that the same dispersion value or the same noise impulse will at different times produce different SNR values. Learning based on a correlation measure requires direct use of the state dynamics.

Figure 2 shows four α-stable bell curves and the noise samples they produce [28,29]. It also shows three infinite-variance curves for $\alpha=1.8$ based on three dispersion values and the resulting samples of impulsive noise. The three impulsive SR profiles for the SNR measure in Fig. 3 show that the SR mode occurs for even smaller dispersion values as the impulsiveness grows (as α falls). Figure 4 shows that the pattern in Fig. 3 generalizes. Impulsiveness decreases stochastic resonance because the exponential law $\gamma_{opt}(\alpha)=cA^{\alpha}$ tends to hold for all the dynamical systems we studied. Figure 5 confirms this pattern for the cross-correlation performance measure for a quartic bistable system. Figure 6 shows that any SNR-based learning scheme faces Cauchy-like impulsiveness as it approaches the first-order condition for an SR optimum. This impulsiveness occurs for all noise-types including the Gaussian. This in turn implies that both biological and engineering systems must find some way to suppress this second level of impulsiveness if they try to learn the SR optimum or otherwise search for it based on noisy training data.

II. SYMMETRIC α-STABLE NOISE: THICK-TAILED BELL CURVES

We use a class of symmetric α-stable bell-curve probability density functions with parameter α in the characteristic function $\phi(\omega)=\exp[-\gamma|\omega|^{\alpha}]$ where γ is the *dispersion* parameter [30–33]. The parameter α lies in $0<\alpha\leqslant 2$ and gives the Gaussian random variable 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 disper-

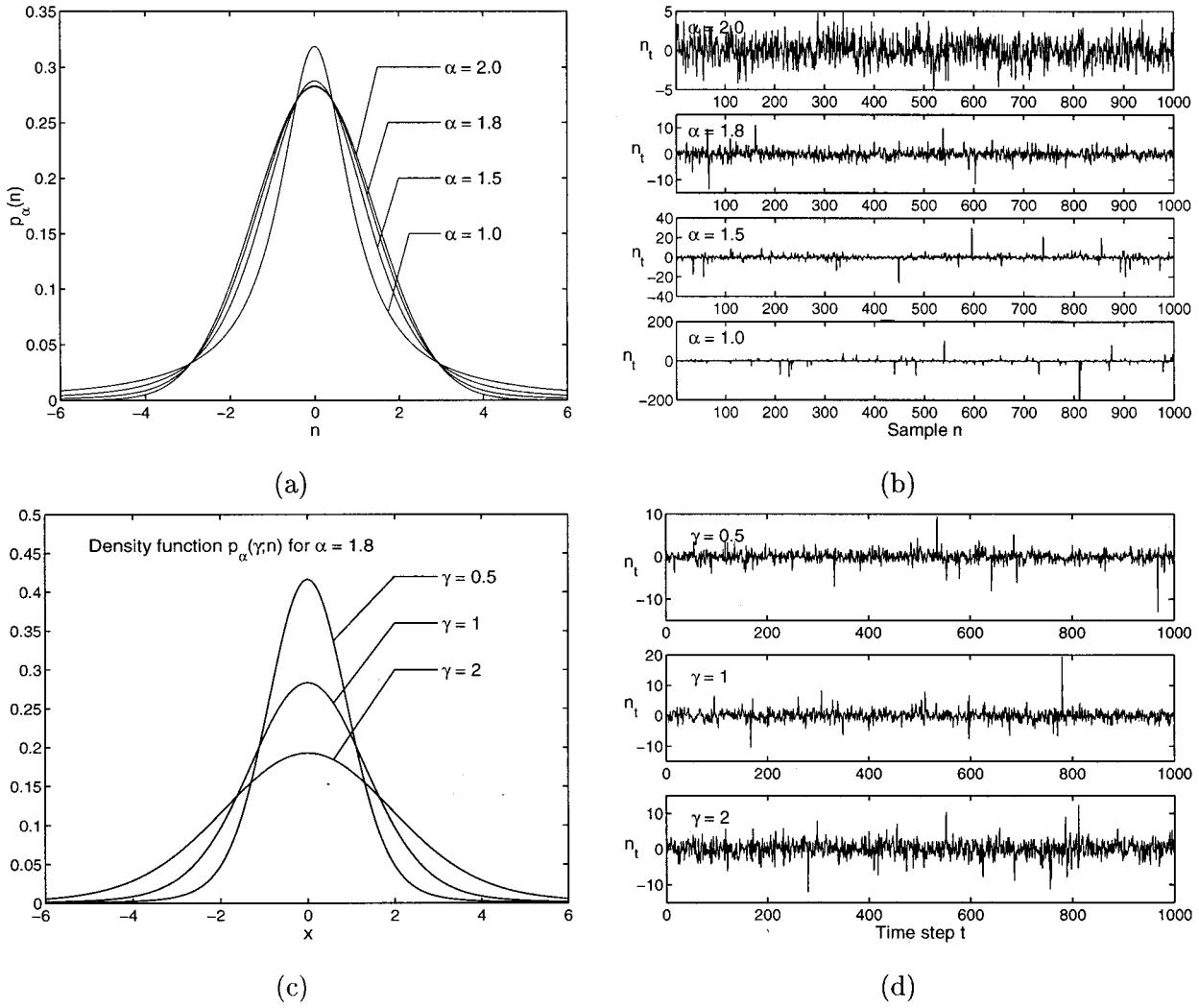

FIG. 2. Samples of standard symmetric α-stable probability densities and their realizations. (a) Density functions with zero location ($a=0$) and unit dispersion ($\gamma=1$) for $\alpha=2$, 1.8, 1.5, and 1. The densities are bell curves that have thicker tails as α decreases. The case $\alpha=2$ gives a Gaussian density with variance two (or unit dispersion). The parameter $\alpha=1$ gives the Cauchy density. (b) Samples of α-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 α-stable variable x becomes more impulsive as the parameter α falls. The algorithm in [28,29] generates these realizations. (c) Density function for $\alpha=1.8$ with dispersion $\gamma=0.5$, 1, and 2. (d) Samples of α-stable noise n for $\alpha=1.8$ with dispersions $\gamma=0.5$, 1, and 2.

sion ($\gamma=1$) Cauchy random variable. The moments of stable distributions with $\alpha<2$ are finite only up to the order k for $k<\alpha$. The Gaussian density alone has finite variance and higher moments. α-stable random variables characterize the class of normalized sums of independent random variables that converge in distribution to a random variable [30] as in the famous Gaussian special case called the ''central limit theorem.'' α-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 [33,34]. The best choice of α is always an *empirical* question for bell-curve phenomena.

Figure 2 shows realizations of four symmetric α-stable random variables. An α-stable probability density f has the characteristic function [32,33,35,36] φ:

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

and

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

where

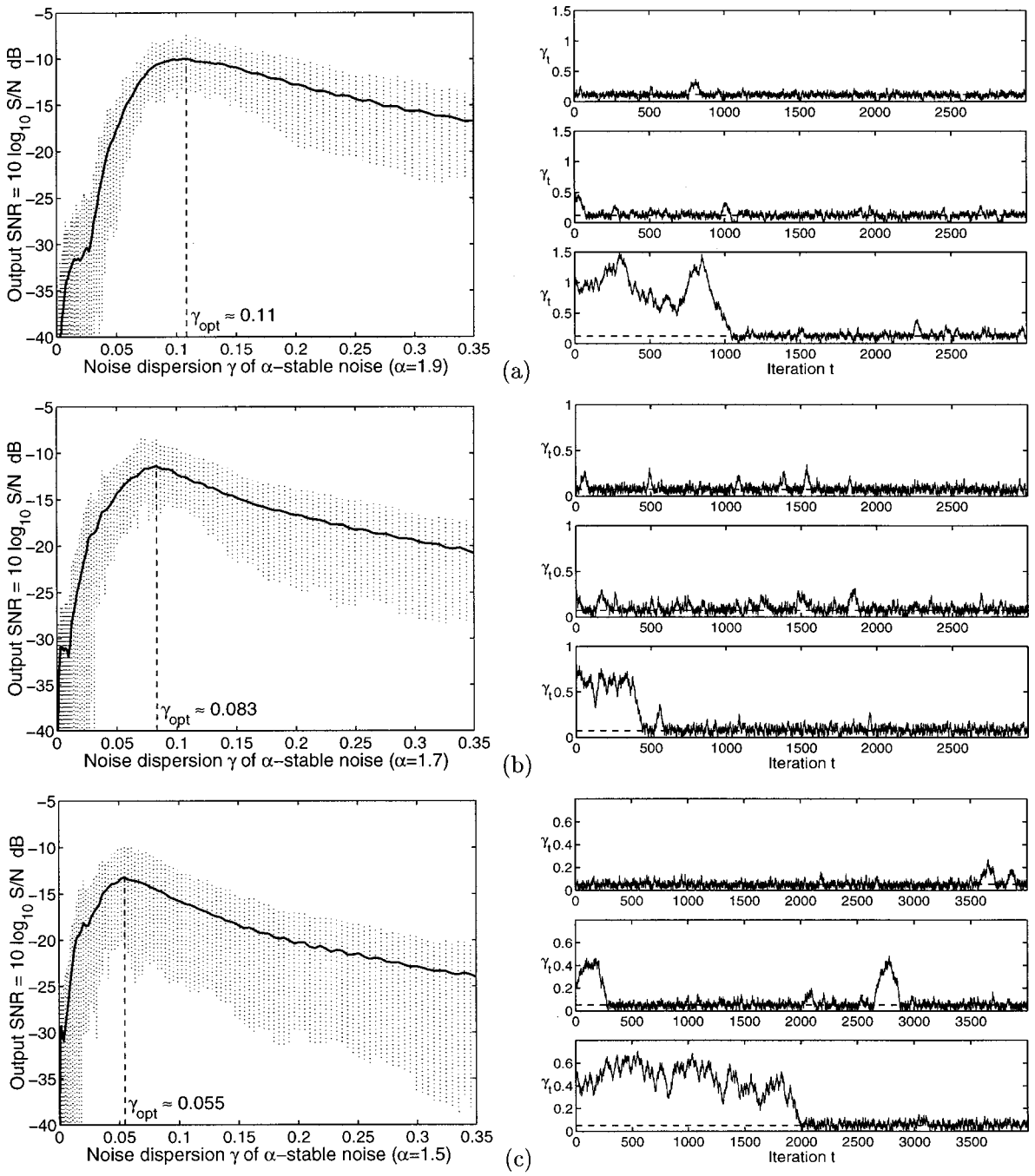

FIG. 3. The optimal dispersion γ_{opt} becomes smaller as the tails on the noise bell curves become thicker and thus as the infinite-variance noise becomes more impulsive. The three SR profiles show that γ_{opt} shifts to the left as α falls. Figure 4 shows that this trend generalizes to an exponential relationship between α and γ_{opt}. The dynamical system is the quartic bistable $\dot{x}=x-x^3+s+n$ modified for saturations effects and where the signal s is the sinusoid $s(t)=0.1\sin 2\pi(0.01)t$. The plots on the left side show the SNR-dispersion profiles for (a) $\alpha=1.9$, (b) $\alpha=1.7$, and (c) $\alpha=1.5$. The dotted lines show the absolute deviation between the smallest and largest SNR outliers in each sample average of 100 outcomes. The vertical dashed lines show the SR effect or mode at the optimal noise dispersion γ_{opt}. The plots on the right side of (a)–(c) show the learning paths of γ as it slowly and noisily converges to γ_{opt} per the robustified learning law in Eq. (43).

$$\mathrm{sgn}(\omega)=\begin{cases}1 & \text{if } \omega>0\\ 0 & \text{if } \omega=0\\ -1 & \text{if } \omega<0,\end{cases} \tag{3}$$

and $i=\sqrt{-1}$, $0<\alpha\leqslant 2$, $-1\leqslant\beta\leqslant 1$, and $\gamma>0$. The α is the characteristic exponent parameter. An α-stable density with $\alpha<2$ has finite moments only of order less than α. Again the variance of an α-stable density distribution does not exist if $\alpha<2$. The location parameter a is the ''mean'' of the density when $\alpha>1$ and β is a skewness parameter. The density is symmetric about a when $\beta=0$. The dispersion parameter γ acts like a variance because it controls the width of a symmetric α-stable bell curve.

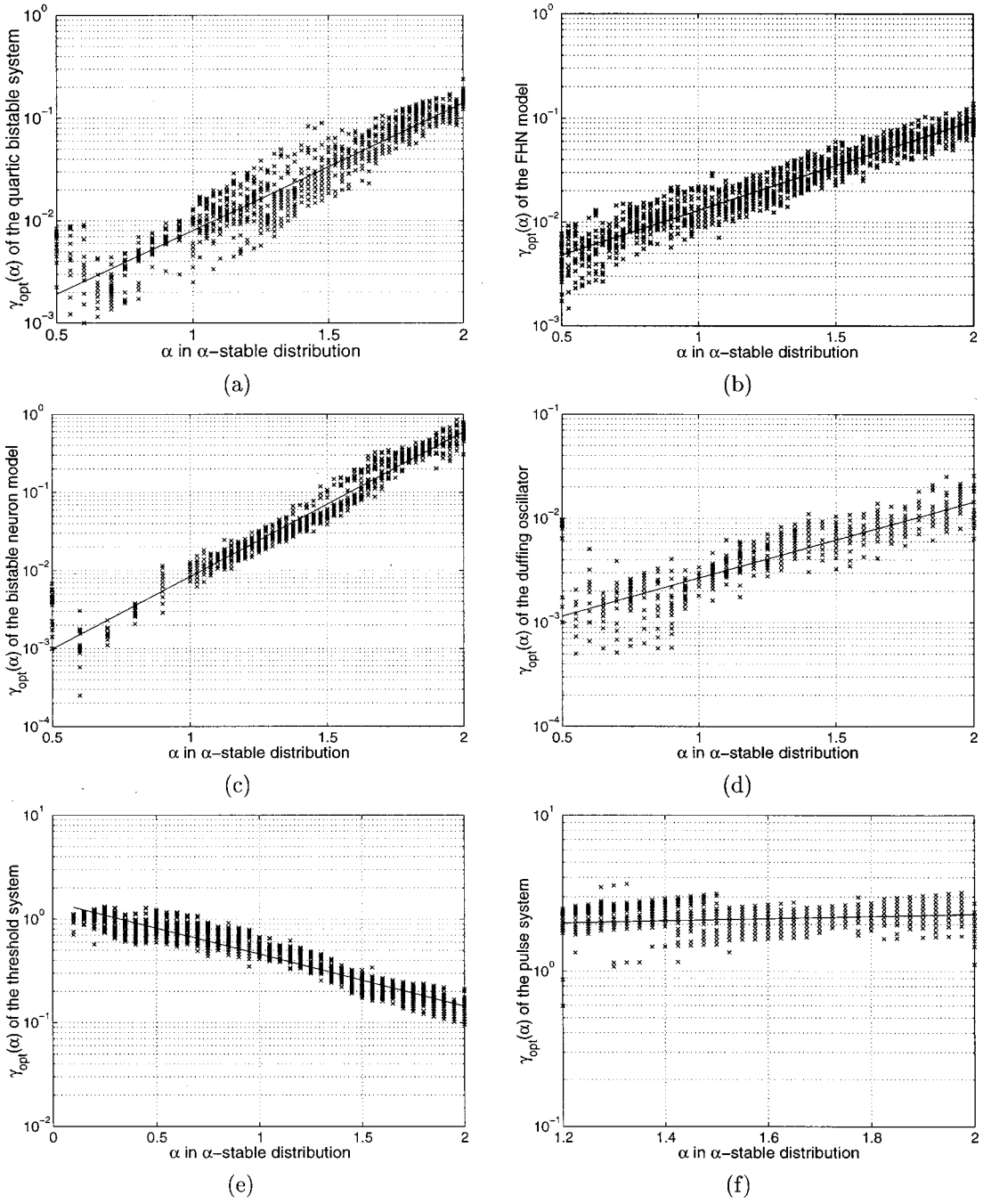

FIG. 4. Exponential laws for optimal noise dispersion γ and parameter α for the SNR performance measure. The optimal noise dispersion γ depends on the parameter α through the exponential relation $\gamma_{opt}(\alpha)=cA^{\alpha}$ for some constants c and A. Table I shows the constants c and A for the dynamical systems we tested: (a) the Quartic bistable system (modified), (b) the FHN model (modified), (c) the bistable neuron model (Hopfield), (d) the duffing oscillator, (e) the feedforward threshold system, and (f) the random pulse system. The slope of the pulse-system in (f) is so close to zero as to undermine the log-linear (exponential) relationship. The small correlation coefficient for the pulse system in Table I reflects this nearly flat log-linear relationship.

III. AN EXPONENTIAL LEARNING LAW: IMPULSIVENESS DECREASES RESONANCE

This section lists the SR performance measures and state models that we used in the simulations. Four of the six state models are feedback or dynamical systems. The neuron and pulse models are feedforward models. All give rise to the exponential law $\gamma_{opt}(\alpha)=cA^{\alpha}$ but the pulse model does so with only a small correlation coefficient of linear regression because its log-plot is almost flat.

A. SR performance measures

This section reviews the two most popular measures of SR. These performance measures depend on the forcing sig-

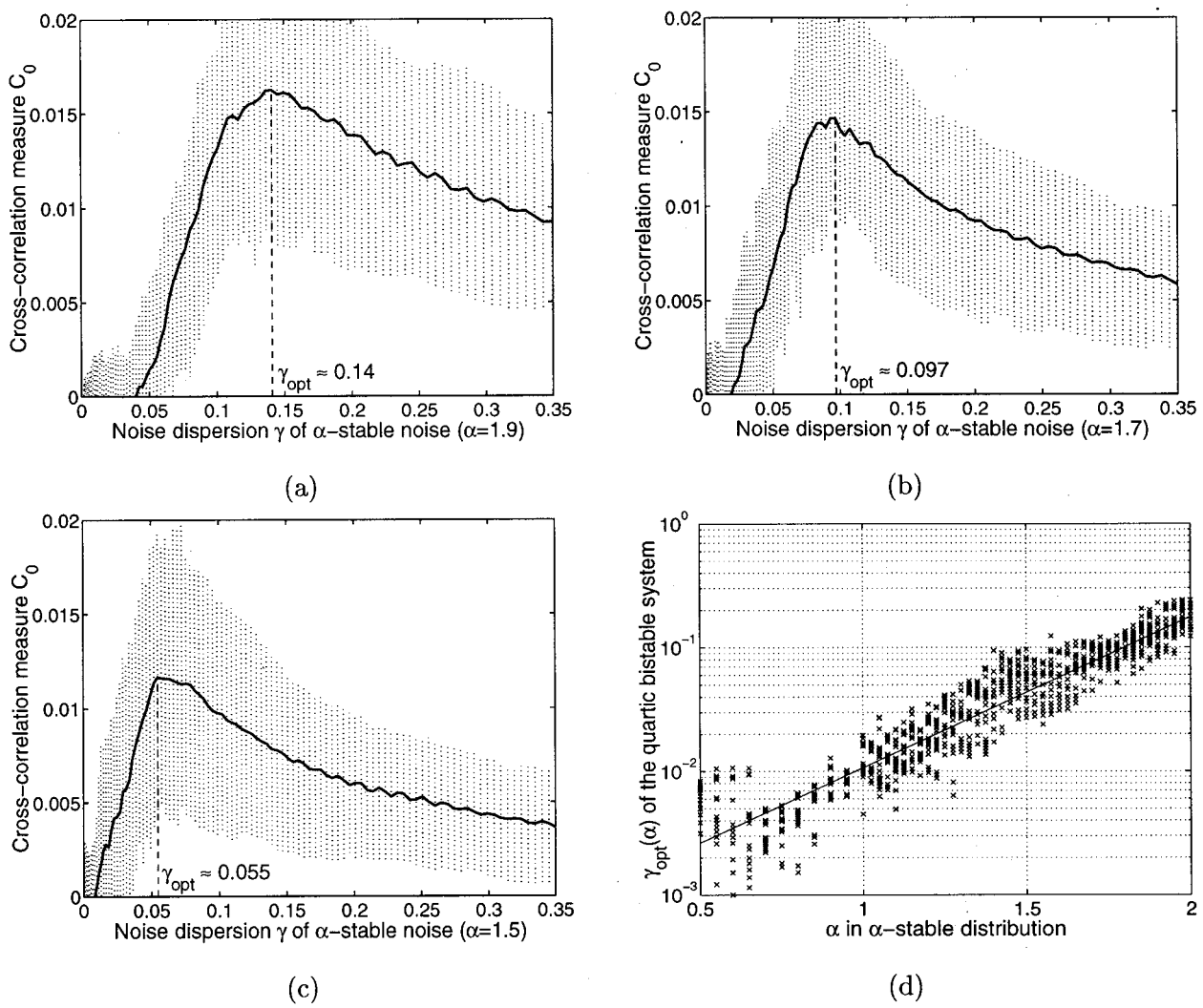

FIG. 5. The optimal dispersion γ_{opt} still becomes smaller as the infinite-variance noise becomes more impulsive for a cross-correlation performance measure. The dynamical system is the quartic bistable $\dot{x}=x-x^3+s+n$ modified for the saturation effects. The signal s is the sinusoid $s(t)=0.1\sin 2\pi(0.01)t$ as in Fig. 3 but with cross-correlation measure C_0. The plots (a)–(c) show the C_0-dispersion profiles for (a) $\alpha=1.9$, (b) $\alpha=1.7$, and (c) $\alpha=1.5$. The dotted lines show the absolute deviation between the smallest and largest cross-correlation outliers in each sample average of 100 outcomes. The vertical dashed lines show the SR effect or mode at the optimal noise dispersion γ_{opt}. The plot (d) shows the exponential law for optimal noise dispersion γ and parameter α.

nal and noise and can vary from system to system. There is no consensus in the SR literature on how to measure the SR effect.

a. Signal-to-noise ratio. The most common SR measure is some form of a signal-to-noise ratio (SNR) [18–22,37]. This seems the most intuitive measure even though there are many ways to define a SNR.

Suppose the input signal is the sinewave $s(t)=\varepsilon\sin\omega_0 t$. Then the SNR measures how much the system output $y=g(x)$ contains the input signal frequency ω_0:

$$\mathrm{SNR}=10\log_{10}\frac{S}{N} \tag{4}$$

$$=10\log_{10}\frac{S(\omega_0)}{N(\omega_0)}\ \mathrm{dB}. \tag{5}$$

The signal power $S=|Y(\omega_0)|^2$ is the magnitude of the output power spectrum $Y(\omega)$ at the input frequency ω_0. The background noise spectrum $N(\omega_0)$ at input frequency ω_0 is some average of $|Y(\omega)|^2$ at nearby frequencies [21,26,38]. The discrete Fourier transform (DFT) $Y[k]$ for $k=0,\ldots,L-1$ is an exponentially weighted sum of elements of a discrete-time sequence $\{y_0,y_1,\ldots,y_{L-1}\}$ of output signal samples

$$Y[k]=\sum_{t=0}^{L-1} y_t e^{-i(2\pi kt/L)}. \tag{6}$$

The signal frequency ω_0 corresponds to bin k_0 in the DFT for integer $k_0=L\Delta T f_0$ and for $\omega_0=2\pi f_0$. This gives the output signal in terms of a DFT as $S=|Y[k_0]|^2$. The noise

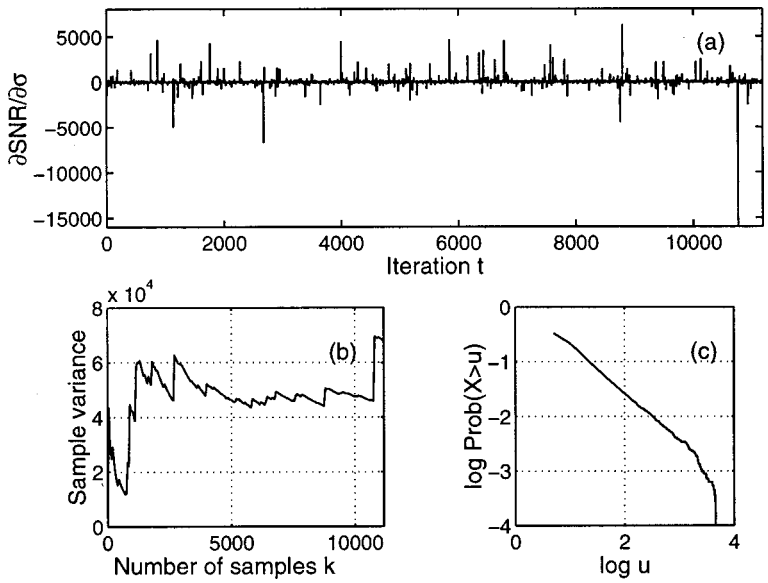

FIG. 6. Visual display of sample statistics of $\partial \mathrm{SNR}_t/\partial\sigma$ for the saturation-modified quartic bistable system $\dot{x}=x-x^3+s+n$ with sinusoidal input $s(t)=0.1\sin 2\pi(0.01)t$ and α-stable noise $n(t)$ with $\alpha=1.8$. The system has binary output $y(t)=\mathrm{sgn}(x(t))$. (a) Cauchy-like samples of $\partial \mathrm{SNR}_t/\partial\sigma$ at each iteration t at the noise dispersion $\gamma=0.1$ (which is the optimal dispersion for this signal system). The plot shows impulsiveness of the random variable $\partial \mathrm{SNR}_t/\partial\sigma$. (b) Test of infinite variance. The sequence of sample variances converges to a finite value if the underlying probability density has finite variance. Else it has infinite variance. (c) Log-tail test of the parameter α in for an α-stable bell curve. The test plots $\log \mathrm{Prob}(X>u)$ versus $\log_{10} u$ for large u. If the underlying density is α-stable with $\alpha<2$ then the slope of this plot is approximately $-\alpha$. This test found that $\alpha\approx 1$ and so the density was approximately Cauchy. The result is that we need to apply the Cauchy impulse suppressor [53] $\phi(x)=2x/(1+x^2)$ to the approximate SR gradient $\partial \mathrm{SNR}_t/\partial\sigma$.

power $N=N[k_0]$ is the average power in the adjacent bins $k_0-M,\ldots,\ k_0-1,\ k_0+1,\ldots,\ k_0+M$ for some integer M [22,39]

$$N=\frac{1}{2M}\sum_{j=1}^{M}\left(|Y[k_0-j]|^2+|Y[k_0+j]|^2\right). \tag{7}$$

There is no standard definition of system-level signal and noise in nonlinear systems. We work with a SNR that is easy to compute and that depends on standard spectral power measures in signal processing. We start with a sinewave input and view the output state $y(t)=g(x(t))$ of the dynamical system as a mixture of signal and noise. We arrange the DFT computation so that the energy of the sine term lies in frequency bin k_0. The squared magnitude of this energy spectrum $Y[k_0]$ acts as the system-level signal: $S=2|Y[k_0]|^2$. We view all else in the spectrum as noise: $N=P-S=P-2|Y[k_0]|^2$ where the total energy is $P=\Sigma_{k=0}^{L-1}|Y[k]|^2$. We ignore the factor L that scales S and P since the ratio S/N cancels its effect.

b. Cross-correlation measures. These ''shape matchers'' can measure SR when inputs are not periodic signals. Researchers coined the term ''aperiodic stochastic resonance'' [23,40–42] for such cases. They defined cross-correlation measures for the input signal s and the system response in terms of the mean transition rate r in the FHN model in Eqs. (16)–(18):

$$C_0=\max\{\overline{s(t)r(t+\tau)}\}, \tag{8}$$

$$C_1=\frac{C_0}{\overline{[s^2(t)]}^{1/2}\{\overline{[(t)-r(t)]^2}\}^{1/2}}. \tag{9}$$

where $\overline{x}$ is the time average $\overline{x}=1/T\int_0^T x(t)dt$.

B. SR systems and simulation models

The computer simulation uses a discrete version

$$x_{t+1}=x_t+\Delta T[f(x_t)+s_t]+\sqrt{\Delta T}\kappa w_t, \tag{10}$$

$$y_{t+1}=g(x_{t+1}), \tag{11}$$

with initial condition x_0 and output y_t. We assume that this discrete model applies to systems with α-stable noise. The zero location white α-stable random sequence $\{w_t\}$ has unit dispersion $\gamma_w=1$. So $n_t=\kappa w_t$ has dispersion $\gamma=\kappa^\alpha$. Note that a unit dispersion for Gaussian density (when $\alpha=2$) equals a variance of two. We tested the following six models:

(a) Quartic bistable system. The forced quartic bistable system has the form

$$\dot{x}=x-x^3+s(t)+n(t), \tag{12}$$

$$y(t)=\mathrm{sgn}(x(t)), \tag{13}$$

for binary output $y(t)$. We tested the quartic bistable system model with the sinusoid input $s(t)=\varepsilon\sin 2\pi f_0 t$ for $\varepsilon=0.1$

and $f_0=0.01$. The discrete version of the quartic bistable follows from Eqs. (10)–(11) as

$$x_{t+1}=x_t+\Delta T(x_t-x_t^3+s_t)+\sqrt{\Delta T}\kappa w_t, \tag{14}$$

$$y_{t+1}=\mathrm{sgn}(x_{t+1}). \tag{15}$$

We limit the magnitude of the system state x_t to 10 in the simulation model (14) to account for physical and computer saturation effects. We put $x_{t+1}=10$ when $x_{t+1}>10$ and put $x_{t+1}=-10$ when $x_{t+1}<-10$ in the discrete dynamic system (14). This gives a modified version of the quartic bistable system. The optimal dispersion γ_{opt} has the form $\gamma_{opt}(\alpha)=\kappa^{\alpha}$ for the noise scale κ in Eq. (14).

(b) FHN model. The forced FHN model has the form

$$\epsilon\dot{x}=-x\left(x^2-\frac{1}{4}\right)-z+A'+s'(t)+n'(t), \tag{16}$$

$$\dot{z}=x-z, \tag{17}$$

$$y(t)=x(t), \tag{18}$$

for $\epsilon=0.005$ and $A'=-(5/12\sqrt{3}+0.07)=-0.310\,56$ as in [42] and linear output $y(t)$. We use a sinusoidal input $s'(t)=\varepsilon\sin 2\pi f_0 t$ where $\varepsilon=0.01$ and $f_0=0.5$. We can rewrite Eqs. (16)–(18) as

$$\begin{aligned}\dot{x}&=-\frac{1}{\epsilon}x\left(x^2-\frac{1}{4}\right)-\frac{1}{\epsilon}z+\frac{1}{\epsilon}A+\frac{1}{\epsilon}s'(t)+\frac{1}{\epsilon}n'(t)\\&=-\frac{1}{\epsilon}x\left(x^2-\frac{1}{4}\right)-\frac{1}{\epsilon}z+A+s(t)+n(t),\end{aligned} \tag{19}$$

$$\dot{z}=x-z, \tag{20}$$

$$y(t)=x(t), \tag{21}$$

for $A=A'/\epsilon$. Then Eqs. (10) and (11) give the discrete version to simulate the FHN model as

$$x_{t+1}=x_t+\Delta T\left[-\frac{1}{\epsilon\varepsilon}x\left(x^2-\frac{1}{4}\right)-\frac{1}{\epsilon}z+A+s_t\right]+\sqrt{\Delta T}\kappa w_t, \tag{22}$$

$$z_{t+1}=z_t+\Delta T(x_t-z_t), \tag{23}$$

$$y_{t+1}=x_{t+1}. \tag{24}$$

We also modify the recursive relation (22) to allow for saturation effects by requiring the magnitude of x_{t+1} not to exceed 2. The optimal dispersion γ_{opt} has the form $\gamma_{opt}(\alpha)=\kappa^{\alpha}$ for the noise scale κ in Eq. (22).

(c) Bistable potential neuron model [43]. The bistable potential neuron model [44] with stable white noise has the form

$$\dot{x}=-x+2\tanh x+s(t)+n(t), \tag{25}$$

$$y(t)=\mathrm{sgn}(x(t)). \tag{26}$$

The sinusoid input is $s(t)=\varepsilon\sin 2\pi f_0 t$ for $\varepsilon=0.1$ and $f_0=0.01$. The discrete version has the form

$$x_{t+1}=x_t+\Delta T(-x_t+2\tanh x_t+s_t)+\sqrt{\Delta T}\kappa w_t, \tag{27}$$

$$y_{t+1}=\mathrm{sgn}(x_{t+1}). \tag{28}$$

We test this neuron model with sinusoidal input $s(t)=\varepsilon\sin 2\pi f_0 t$ where $\varepsilon=0.1$ and $f_0=0.01$.

(d) Duffing oscillator [45]. The forced duffing oscillator has the form

$$\ddot{x}=-0.15\dot{x}+x-x^3+\epsilon\sin(\omega_0 t)+n(t), \tag{29}$$

$$y(t)=x(t). \tag{30}$$

We test the duffing oscillator with sinusoidal input $s(t)=\varepsilon\sin 2\pi f_0 t$ for $\varepsilon=0.3$ and $f_0=0.01$. The discrete version of the duffing oscillator has the form

$$x_{t+1}=x_t+\Delta T z_t, \tag{31}$$

$$z_{t+1}=z_t+\Delta T(-\delta z_t+x_t-x_t^3+s_t)+\sqrt{\Delta T}\kappa w_t, \tag{32}$$

$$y_{t+1}=\mathrm{sgn}(x_{t+1}). \tag{33}$$

(e) Threshold system [15,46–50]. The output y of a simple feedforward threshold system has the form

$$y_t=\mathrm{sgn}(s_t+n_t-\Theta)=\mathrm{sgn}(s_t+\kappa w_t-\Theta). \tag{34}$$

The optimal dispersion γ_{opt} has the form $\gamma_{opt}(\alpha)=\kappa^{\alpha}$ for κ in Eq. (34).

(f) Pulse system [51]. This doubly Poisson process generates a pulse train with probability r that depends on the input $V(t)=s(t)+n(t)$

$$r(V(t))=r(0)\exp(V(t)). \tag{35}$$

Here we let $r(0)=1$. The sinusoid input is $s(t)=\varepsilon\sin 2\pi f_0 t$ for $\varepsilon=0.5$ and $f_0=0.05$. The system generates an output $y(t)$ as a unit pulse with a rate $r(t)$.

1. Exponential law with linear least-squares fit of log data

The optimal dispersion $\gamma_{opt}(\alpha)$ of the system obeys the exponential law

$$\gamma_{opt}(\alpha)=cA^{\alpha} \tag{36}$$

for real constants c and A. Then

$$\log_{10}\gamma_{opt}(\alpha)=\log_{10}c+\alpha\log_{10}A=a\alpha+c' \tag{37}$$

for $a=\log_{10}A$ and $c'=\log_{10}c$. The least-squares method gives the a and c' values as

$$a=\frac{\sum_{i=1}^{N}(\alpha_i-\overline{\alpha})w_i}{\sum_{i=1}^{N}\alpha_i^2-N(\overline{\alpha})^2}\quad\text{and}\quad c'=\overline{w}-a\overline{\alpha}, \tag{38}$$

for N data pairs (α_i,w_i) where $w_i=\log_{10}\gamma_{opt}(\alpha_i)$ at the experiment i with the parameter α_i. This method is the same

TABLE I. Linear least-squares fit of the log of optimal dispersion γ and the parameter α in an α-stable density. The parameters a and c' relate $\log_{10}\gamma$ and α through a straight line: $\log_{10}\gamma(\alpha)=a\alpha+c'$.

	SNR		Cross correlation	
	Parameters	r^2	Parameters	r^2
Quartic bistable	$a=1.2444$, $c'=-3.3411$	0.8923	$a=1.2177$, $c'=-3.1889$	0.8463
FHN	$a=0.8622$, $c'=-2.7496$	0.9098	$a=0.6518$, $c'=-2.4869$	0.7510
Bistable neuron	$a=1.8552$, $c'=-3.9344$	0.9593	$a=1.9581$, $c'=-4.0252$	0.9641
Duffing oscillator	$a=0.7320$, $c'=-3.3057$	0.7444	$a=0.8912$, $c'=-3.3204$	0.8175
Threshold	$a=-0.5020$, $c'=0.1638$	0.9215	$a=-0.5036$, $c'=0.1658$	0.9196
Pulse	$a=0.0692$, $c'=0.2267$	0.0406	$a=0.2478$, $c'=0.2516$	0.3361

as the minimum variance method for arbitrary random variables and the maximum likelihood method for normal random variables [52].

The correlation coefficient r^2 indicates how good the linear model fits the data

$$r^2=\frac{\Sigma(\hat{w}_i-\overline{w})^2}{\Sigma(w_i-\overline{w})^2}=\frac{[\Sigma(\alpha_i-\overline{\alpha})(w_i-\overline{w})]^2}{\Sigma(ga_i-\overline{\alpha})^2\Sigma(w_i-\overline{w})^2}, \tag{39}$$

where $0\leqslant|r|\leqslant 1$ and $|r|=1$ iff $w_i=\hat{w}_i=a\alpha_i+c'$ for every i. The positive and negative signs reflect the positive and negative slopes.

2. Test results

Table I shows the parameters a and c' of the linear least-squares fit of logarithm of the optimal dispersion γ_{opt} and the parameter α. The correlation coefficients r^2 measure how well the regression $a\alpha+b$ fits the data and how much $\log_{10}\gamma_{opt}$ linearly depends on α. Figure 4 shows the SR-optimal dispersion $\gamma_{opt}(\alpha)$ versus the parameter α. The plots in Figs. 4(a)–4(d) for feedback systems agree with the exponential law. Figures 4(e) and 4(f) show the plots for the threshold and feedforward pulse systems. The correlation coefficients r^2 for the pulse system for both the SNR and cross-correlation measures are small due to the small slopes a and the large spread of the data $\log_{10}(\gamma_{opt})$. But their trends still show a linear relationship.

Note also that the slopes of the plots can be positive or negative or zero depending on the time scale factor of the dynamical system and on the noise when we consider the noise scale κ that gives the dispersion $\gamma=\kappa^\alpha$. Consider, for example, the two FHN models (16)–(18) and (19)–(21) are the same system. But the noise $n'(t)=\kappa' w(t)$ in Eq. (16) differs from the optimal noise $n(t)=\kappa w(t)$ in Eq. (19) by the scale ϵ. So at SR the two optimal noise scales obey the relation $\kappa'_{opt}=\epsilon\kappa_{opt}$. Then $\gamma'_{opt}(\alpha)=\kappa'_{opt}(\alpha)^\alpha=c(AB)^\alpha$ if $\gamma_{opt}(\alpha)=\kappa_{opt}(\alpha)^\alpha=cA^\alpha$. So the factor ϵ can change the slope of the plot from positive to negative for this FHN model.

IV. LEARNING THE OPTIMAL NOISE DISPERSION IN IMPULSIVE ENVIRONMENTS

We applied the stochastic SR gradient-ascent learning law of [15] to the problem of finding the optimal noise dispersion γ_{opt} for infinite-variance noise. This learning law has the form

$$\gamma_{t+1}=\gamma_t+\mu_t\frac{\partial\mathrm{SNR}}{\partial\gamma}, \tag{40}$$

where μ_t is a decreasing sequence of learning coefficients. A like learning law holds for the correlation measure in Eq. (9). The spectral relation $\mathrm{SNR}=S/N$ and the chain rule of calculus show that

$$\frac{\partial\mathrm{SNR}}{\partial\gamma}=\frac{\partial\mathrm{SNR}}{\partial S}\frac{\partial S}{\partial\gamma}+\frac{\partial\mathrm{SNR}}{\partial N}\frac{\partial N}{\partial\gamma} \tag{41}$$

$$=\frac{1}{N}\frac{\partial S}{\partial\gamma}-\frac{\mathrm{SNR}}{N}\frac{\partial N}{\partial\gamma}. \tag{42}$$

The first-order condition for an SR maximum is $\partial\mathrm{SNR}/\partial\gamma=0$. This leads to the optimality condition $S/N=S'/N'$ where $S'=\partial S/\partial\gamma$. But the optimality error process $\mathcal{E}=S/N-S'/N'$ itself is impulsive. Indeed a converging-variance test and log-tail test confirm that this random process obeys the highly impulsive Cauchy probability density (with $\alpha\approx 1$). Figure 6 shows samples of this Cauchy-like error process. These impulses destabilized all attempts to learn γ_{opt} with Eq. (42). This Cauchy impulsiveness holds for forcing noise with finite as well as infinite variance and for all the SR models and performance measures. It is systemic to the gradient-learning process. But its Cauchy nature suggests an immediate remedy. We can apply the well-known Cauchy impulse suppressor $\phi(z_t)=2z_t/(1+z_t^2)$ from the theory of robust statistics [53]. This gives the final robustified form of the learning law:

$$\gamma_{t+1}=\gamma_t+\mu_t\phi\left(\frac{\partial\mathrm{SNR}}{\partial\gamma}\right). \tag{43}$$

The robustified learning law (43) learned the optimal dispersions γ_{opt} in Figs. 1 and 3. It successfully found γ_{opt} for α values in the range [1.4, 2) for both the quartic bistable and Fitzhugh-Nagumo models but only for the SNR performance measure. The learning law often converged to γ_{opt} for α values in [1, 1.4) but with decreasing frequency and accuracy for the lower α values. The learning scheme often did not converge when the forcing noise was Cauchy ($\alpha=1$).

Learning with the SNR measure did not require knowledge of the system dynamics while learning with the correlation measure did require some knowledge of the system Jacobian. Learning is slow in any case because the system must in effect estimate at least part of the system dynamics based on the sampled SNR inputs to the learning process. The robustified gradient scheme (43) can use other performance measures or can include more information from the system dynamics to help the system more accurately estimate the stochastic term $\partial \mathrm{SNR}/\partial \gamma$.

V. CONCLUSION

We have shown that stochastic resonance is robust against noise outliers. Sufficiently large and sufficiently frequent noise impulses can overwhelm any SR system. But an SR effect still emerges even for the wide range of infinite-variance noise-types that lie between the extremes of the wildly impulsive Cauchy bell curve and the nonimpulsive Gaussian. The approximate exponential relationship $\gamma_{opt}(\alpha) = cA^{\alpha}$ shows this. This result is encouraging because all real noise is impulsive to some degree—the best-fit α is seldom the Gaussian case of $\alpha = 2$. This robustness favors engineering designs that may not conform to the ideal standards of Gaussian noise. It also suggests that SR may occur more widely in nature than many had believed.

The success of the dispersion-learning simulations further suggests that evolution could have tuned biological parameters to exploit the SR affect for signal detection in noisy environments. No living organism can control the noise structure of the environment. But gene selection over thousands of generations might act as if a gene pool slowly and noisily tuned its own noise parameters. Each act of reproductive fitness would count as only a lone noisy spike in evolution's learning process. The battle of genetic countermeasures between predator and prey suggest that if the predator or prey evolved SR-sensitive signal detection (as Moss [11,54] has shown for crayfish that use SR to detect a largemouth bass's periodic fin pattern or paddlefish [55] that use SR to detect plankton) then they would have to evolve new SR parameter settings as their opponents evolved new countermeasures.

The problem with such an SR evolutionary hypothesis is the Cauchy impulsiveness of gradient-ascent learning (40) for either a signal-to-noise or correlation performance measure. Biological systems would have to further evolve a robustifier of some sort to suppress extremely large learning outliers as Eq. (43) does with the Cauchy impulse suppresser. A meta-level threshold system might suffice for that task.

ACKNOWLEDGMENTS

National Science Foundation Grant Nos. ECS-9906251 and ECS-0070284 and the Thailand Research Fund Grant No. PDF/29/2543 partly supported this research.

[1] R. Benzi, G. Parisi, A. Sutera, and A. Vulpiani, Tellus **34**, 10 (1982).
[2] R. Benzi, G. Parisi, A. Sutera, and A. Vulpiani, SIAM (Soc. Ind. Appl. Math.) J. Appl. Math. **43**, 565 (1983).
[3] R. Benzi, A. Sutera, and A. Vulpiani, J. Phys. A **14**, L453 (1981).
[4] K. S. Brown, New Sci. **150**, 28 (1996).
[5] A. R. Bulsara and L. Gammaitoni, Phys. Today **49** (3), 39 (1996).
[6] J.-P. Eckmann and L. E. Thomas, J. Phys. A **15**, L261 (1982).
[7] L. Gammaitoni, P. Hänggi, P. Jung, and F. Marchesoni, Rev. Mod. Phys. **70**, 223 (1998).
[8] J. Glanz, Science **277**, 1758 (1997).
[9] P. Jung and K. Wiesenfeld, Nature (London) **385**, 291 (1997).
[10] F. Moss, D. Pierson, and D. O'Gorman, Int. J. Bifurcation Chaos Appl. Sci. Eng. **4**, 1383 (1994).
[11] F. Moss and K. Wiesenfeld, Sci. Am. **273**, 66 (1995).
[12] C. Nicolis, Tellus **34**, 1 (1982).
[13] K. Wiesenfeld and F. Moss, Nature (London) **373**, 33 (1995).
[14] F. Chapeau-Blondeau and X. Godivier, Phys. Rev. E **55**, 1478 (1997).
[15] L. Gammaitoni, Phys. Lett. A **208**, 315 (1995).
[16] S. Mitaim and B. Kosko, Proc. IEEE **86**, 2152 (1998).
[17] B. R. Parnas, IEEE Trans. Biomed. Eng. **43**, 313 (1996).
[18] S. Fauve and F. Heslot, Phys. Lett. **97A**, 5 (1983).
[19] R. F. Fox, Phys. Rev. A **39**, 4148 (1989).
[20] G. Hu, G. Nicolis, and N. Nicolis, Phys. Rev. A **42**, 2030 (1990).
[21] B. McNamara and K. Wiesenfeld, Phys. Rev. A **39**, 4854 (1989).
[22] T. Zhou and F. Moss, Phys. Rev. A **41**, 4255 (1990).
[23] J. J. Collins, C. C. Chow, A. C. Capela, and T. T. Imhoff, Phys. Rev. E **54**, 5575 (1996).
[24] J. J. Collins, C. C. Chow, and T. T. Imhoff, Nature (London) **376**, 236 (1995).
[25] A. R. Bulsara and A. Zador, Phys. Rev. E **54**, R2185 (1996).
[26] A. Neiman, B. Shulgin, V. Anishchenko, W. Ebeling, L. Schimansky-Geier, and J. Freund, Phys. Rev. Lett. **76**, 4299 (1996).
[27] M. Stemmler, Network Comput. Neural Syst. **7**, 687 (1996).
[28] J. M. Chambers, C. L. Mallows, and B. W. Stuck, J. Am. Stat. Assoc. **71**, 340 (1976).
[29] P. Tsakalides and C. L. Nikias, IEEE Trans. Signal Process. **44**, 1623 (1996).
[30] L. Breiman, *Probability* (Addison-Wesley, Reading, MA, 1968).
[31] W. Feller, *An Introduction to Probability Theory and Its Applications*, (Wiley, New York, 1966), Vol. II.
[32] M. Grigoriu, *Applied Non-Gaussian Processes* (Prentice Hall, Englewood Cliff, NJ, 1995).
[33] C. L. Nikias and M. Shao, *Signal Processing with Alpha-Stable Distributions and Applications* (Wiley, New York, 1995).
[34] B. Kosko, *Fuzzy Engineering* (Prentice Hall, Englewood Cliffs, NJ, 1996).

[35] V. Akgiray and C. G. Lamoureux, J. Bus. Econ. Stat. **7**, 85 (1989).
[36] H. Bergstrom, Ark. Math. **2**, 375 (1952).
[37] L. Gammaitoni, E. Menichella-Saetta, S. Santucci, F. Marchesoni, and C. Pressilla, Phys. Rev. A **40**, 2114 (1989).
[38] M. E. Inchiosa and A. R. Bulsara, Phys. Rev. E **53**, R2021 (1996).
[39] A. S. Asdi and A. H. Tewfik, in Proceedings of the 1995 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-95), Detroit, MI, 1995, Vol. 2, pp. 1332–1335.
[40] D. R. Chialvo, A. Longtin, and J. Müller-Gerking, Phys. Rev. E **55**, 1798 (1997).
[41] J. J. Collins, C. C. Chow, and T. T. Imhoff, Phys. Rev. E **52**, R3321 (1995).
[42] C. Heneghan, C. C. Chow, J. J. Collins, T. T. Imhoff, S. B. Lowen, and M. C. Teich, Phys. Rev. E **54**, R2228 (1996).
[43] A. R. Bulsara, E. W. Jacobs, T. Zhou, F. Moss, and L. Kiss, J. Theor. Biol. **152**, 531 (1991).
[44] M. A. Cohen and S. Grossberg, IEEE Trans. Neural Netw. **SMC-13**, 815 (1983).
[45] G. Nicolis, C. Nicolis, and D. McKernan, J. Stat. Phys. **70**, 125 (1993).
[46] Z. Gingl, L. B. Kiss, and F. Moss, Europhys. Lett. **29**, 191 (1995).
[47] X. Godivier and F. Chapeau-Blondeau, Signal Process. **56**, 293 (1997).
[48] P. Jung, Phys. Rev. E **50**, 2513 (1994).
[49] P. Jung, Phys. Lett. A **207**, 93 (1995).
[50] A. Restrepo (Palacios), L. F. Zuluaga, and L. E. Pino, in Proceedings of the 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-97), Munich, Germany, 1997, Vol. III, pp. 2365–2368.
[51] S. M. Bezrukov and I. Vodyanoy, Nature (London) **385**, 319 (1997).
[52] A. Papoulis, *Probability and Statistics* (Prentice Hall, Englewood Cliffs, NJ, 1990).
[53] P. J. Huber, *Robust Statistics* (Wiley, New York, 1981).
[54] J. K. Douglass, L. Wilkens, E. Pantazelou, and F. Moss, Nature (London) **365**, 337 (1993).
[55] D. F. Russell, L. A. Willkens, and F. Moss, Nature (London) **402**, 291 (1999).