10-1-2014

# Noise-Enhanced Information Systems

Hao Chen
*Boise State University*

Lav R. Varshney
*University of Illinois at Urbana-Champaign*

Pramod K. Varshney
*Syracuse University*

# Noise-Enhanced Information Systems

Hao Chen, *Member, IEEE,* Lav R. Varshney, *Member, IEEE*, and Pramod K. Varshney, *Fellow, IEEE*

*Invited Paper*

*Abstract*—**Noise, traditionally defined as an unwanted signal or disturbance, has been shown to play an important constructive role in many information processing systems and algorithms. This noise enhancement has been observed and employed in many physical, biological, and engineered systems. Indeed stochastic facilitation (SF) has been found critical for certain biological information functions like detection of weak, subthreshold stimuli or suprathreshold signals through both experimental verification and analytical model simulations.**

**In this paper, we present a systematic noise-enhanced information processing framework to analyze and optimize the performance of engineered systems. System performance is evaluated not only in terms of signal-to-noise ratio but also in terms of other more relevant metrics such as probability of error for signal detection or mean square error for parameter estimation. As an important new instance of SF, we also discuss the constructive effect of noise in associative memory recall. Potential enhancement of image processing systems via the addition of noise is discussed with important applications in biomedical image enhancement, image denoising and classification.**

*Index Terms*—**Stochastic facilitation, noise-enhanced signal processing, stochastic resonance**

## I. INTRODUCTION

**L**OOSELY defined as an unwanted signal or disturbance to a system, understanding and handling noise is an important research problem in modern science and engineering [1], [2]. Whether considering problems of communication [3], detection [4], estimation [5], or learning [6], designing systems to deal with noise has been the centerpiece of information processing research for decades.

Generally more system noise leads to less channel capacity, worse detection performance, degraded estimation accuracy, and reduced ability to learn in a generalizable way. Hence noise is often removed or mitigated by a variety of filters and signal processing algorithms. Despite its generally disruptive nature, noise may surprisingly play an important *constructive* role in many nonlinear information processing systems. Noise enhancement has been employed in many areas such as dithering in quantization, stochastic optimization techniques such as genetic algorithms or simulated annealing, and in learning [7]. Noise enhancement has also been observed in physical systems as stochastic resonance (or more generally as stochastic facilitation). Moreover, understanding the functional

H. Chen is with the Department of Electrical and Computer Engineering, Boise State University, Boise, ID 83725 USA (e-mail: haochen@boisestate.edu).

L. R. Varshney is with the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, 61801 USA (e-mail: varshney@illinois.edu).

P. K. Varshney is with the Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY, 13244 USA (e-mail: varshney@syr.edu).

role of noise in information processing has shed light into the way biological systems operate [8]–[11].

Building on inspiration from biology, electronic information processing circuits are now being built to take advantage of stochastic facilitation [12] for several applications, including some not discussed here. Stochastic electronics designs are especially useful at nanoscale, where randomness is abundant.

The primary purpose of this paper is to present an overview of mathematical theories of information processing where noise enhancement arises, to indicate how these effects can be optimally utilized, and to demonstrate their utility in a variety of engineering applications.

As a starting vignette to provide intuition, we demonstrate the value of noise in quantization through the technique called *dithering*. The sequel delves into other information processing tasks for which a little bit of noise helps rather than hurts.

### A. A Vignette: Dithering

Consider the simplest of image quantization tasks: take a grayscale image with pixel values between 1 and 256 and convert to a black and white image with only one bit per pixel. Figure 1 shows an example. Pane (a) is the original image; pane (b) is the original image quantized to one bit with a uniform quantizer; and pane (c) is the same original image quantized with the same quantizer, however, independent Gaussian noise with mean zero and standard deviation 45 was added to the image before quantization, relative to pixel values in the range 1 to 256. As can be observed both contours and textures are better preserved.

The same example can be extended to a setting for detecting subthreshold signals. If a one-bit quantizer, such that all pixel values are below the threshold, is used then no image will be seen, pane (d). The addition of noise for the same quantizer can recover some visual information. Pane (e) shows the result when independent Gaussian noise of mean zero and standard deviation 100 is added before quantization.

In this elementary setting, we can viscerally observe that noise has a benefit in sensory processing. In the remainder of the paper, we go through a variety of information processing settings where this is true. Before closing this vignette, though, some details are in order.

Adding some noise to the signal before quantization— the process of dithering—has been shown to improve signal quality and mitigate the artifact effect introduced by quantization [13]–[18]. Let the signal to be quantized be $\mathbf{x} = [x_1, \ldots, x_N] \in R^N$, the dithering noise be $\mathbf{n} = [n_1, \ldots, n_N] \in R^N$ and the multi-bit quantizer be $q(\cdot)$. The quantized signal $\mathbf{z}$ is $\mathbf{z} = q(\mathbf{x} + \mathbf{n})$. Depending on whether the receiver knows the
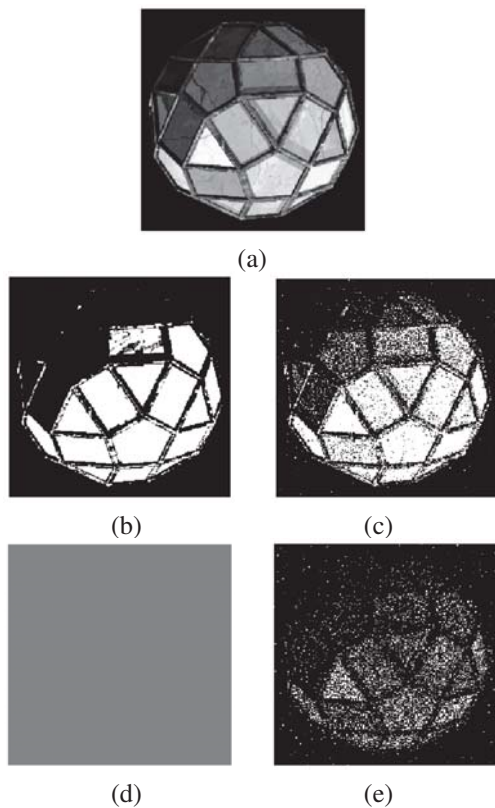
Fig. 1: One-bit image quantization. (a) the original 8-bit image. (b) 1-bit uniformly quantized image. (c) 1-bit uniformly quantized image with additive noise. (d) 1-bit subthreshold quantized image. (e) 1-bit subthreshold quantized image with additive noise.

dithering noise $\mathbf{n}$ or not, there are two approaches to dithering: *subtractive* dithering and *nonsubtractive* dithering where the estimated signal $\hat{\mathbf{x}}$ is given by $\hat{\mathbf{x}} = \mathbf{z} - \mathbf{n} = q(\mathbf{x} + \mathbf{n}) - \mathbf{n}$, or $\hat{\mathbf{x}} = \mathbf{z} = q(\mathbf{x} + \mathbf{n})$ respectively.

When applying subtractive dithering methods under certain conditions, the error residual $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}$ is independent of $\mathbf{x}$ and thus reduces artifacts [14]–[16]. As it requires the exact value of the dithering noise $\mathbf{n}$, subtractive dithering provides a better statistical result but is impractical in many instances. Nonsubtractive dithering is of more interest in quantizer design [19]. For nonsubtractive dithering, the residual is no longer independent of the signal but moments of the total error can be made independent of $\mathbf{x}$ by choosing a suitable probability density function (pdf) of $\mathbf{n}$ [17], [18]. Dithering may also improve signal quality. For example, for a $\Sigma\Delta$ quantizer with a weak sinusoid signal input, the output signal-to-noise ratio can be improved by choosing the right amount of dithering noise [20]. There are precise connections between quantization, dithering, and stochastic resonance [21].

### B. Preview and Organization of Paper

In the above vignette, we demonstrated that introducing some noise improves visualization of quantized images. Introduction of noise or other randomization has been observed in many biological systems and provides motivation to explore

this bio-inspired concept while designing or enhancing system performance. Although we discuss this for imaging systems which is the theme of this special issue, we also provide a much broader presentation encompassing several areas under the umbrella of information systems.

The remainder of the paper is organized as follows. In Sec. II, we describe the phenomenon of stochastic resonance, the term used by many investigators in the field, and present two quantitative metrics to characterize system performance. In Sec. III, we give several illustrative examples of biological systems where SF has been observed. In Secs. IV and V, we discuss noise-enhanced detection and estimation as well as image processing systems. In Secs. VI and VII, we point out the role randomization plays in stochastic search algorithms for optimization and associative memory recall.

## II. STOCHASTIC RESONANCE

The focus of this paper is on noise-enhanced systems for which extensive research has also been carried out under the name *stochastic resonance* (SR). Since proposed by Benzi et al. in 1981 [22] to explain the periodicity of the earth's ice ages, the SR effect, or more generally *stochastic facilitation* (SF), has been observed and applied in many systems— physical, biological, and engineered [8], [9], [23], [24]. The classic signature of SR is the output signal-to-noise ratio (SNR) being greater than the input SNR when an appropriate amount of noise (usually Gaussian) is added to weak periodic input signals [25]–[39].

Later, for a set of parallel networks consisting of multiple SR elements and a sum unit, suprathreshold SR (SSR) effects were demonstrated and investigated in terms of mutual information improvement via additive noise [40]–[42] or overall detection performance [43]. Unlike conventional SR, the input signals to the parallel arrays are predominantly suprathreshold. For a given noise form, the optimal noise variance was determined for hypotheses testing, estimation, and watermark decoding [44]–[46], though some results are valid under the rather strong assumption of asymptotic normality of the output signal, which is not likely to be true in practice due to statistical dependence of the SR elements arising from the common input signal.

Due to its richness and close relationship with the topic of this paper, we provide a brief background on SF theory and applications in the remainder of this section as well as in the next sections. This is by no means a thorough treatment of the subject; interested readers are referred to [8], [9], [21], [23] and references therein for further development.

Although noise manifests additively in many information systems with stochastic resonance, this is not the only way. When engineering systems to have noise enhancement [47], one may also introduce randomness into signals via other random transformations. For example, one could consider multiplicative noise. Also, unlike most current approaches, one may consider signal-dependent noise. Next, we discuss some fundamental metrics to characterize system performance enhancement due to SF.

## A. Signal-to-noise ratio Gain by Stochastic Resonance

In many information processing systems, performance is quantified in terms of SNR. Some approaches have been proposed to tune the SR system by maximizing SNR [23]. For some SR systems, robustness enhancement using non-Gaussian noises was reported in [32]. For a fixed type of noise, Mitaim and Kosko [48] proposed an adaptive stochastic learning scheme performing a stochastic gradient ascent search on the SNR to determine the optimal noise level based on the samples from the process. Rather than adjusting the input noise level, Xu et al. [49] proposed a numerical method for realizing SR by tuning system parameters to maximize SNR gain. As an example, stochastic resonance in a 3-level quantizer is illustrated as follows [50].

Consider a symmetric 3-level quantizer with thresholds $-\zeta$ and $\zeta$ driven by a sequence $x[n] = s[n] + w[n]$ which is the sum of a subthreshold sinusoid signal $s[n] = A\cos(2\pi n/N - \theta_0)$ with amplitude $A$, frequency $N$, and phase $\theta_0$ and $w[n]$, an independent and identically distributed (i.i.d.) zero-mean noise signal with variance $\sigma^2$. The output $y[n]$ is given by

$$y[n] = \begin{cases} -1, & x[n] \leq -\zeta \\ 0, & -\zeta < x[n] \leq \zeta \\ 1, & x[n] > \zeta. \end{cases} \quad (1)$$

Further, consider the case where $w[n]$ belongs to the generalized Gaussian family such that its pdf satisfies

$$f(w) = \frac{a}{\sigma} e^{(-b|w/\sigma|^p)}$$

where

$$a = \frac{p}{2} \frac{\Gamma^{1/2}(3/p)}{\Gamma^{3/2}(2/p)}$$

and

$$b = \left[\frac{\Gamma(3/p)}{\Gamma(1/p)}\right]^{p/2}.$$

The Laplacian, Gaussian, and uniform pdfs belong to the family defined by $p = 1, 2,$ and $\infty$, respectively.

The input SNR is $\text{SNR}_i = A^2/4\sigma^2$. The output SNR is given by [29] as $\text{SNR}_o = |Y_1|^2/\bar{\sigma}_y^2$, where

$$Y_1 = \frac{1}{N} \sum_{n=0}^{N-1} E(y[n]) e^{j2\pi n/N}$$

and

$$\bar{\sigma}_y^2 = \frac{1}{N} \sum_{n=0}^{N-1} \sigma_y^2[n].$$

Thus, the SNR gain is $G = \text{SNR}_o/\text{SNR}_i$.

When the signal amplitude $\frac{A}{\sigma} \ll 1$ is very small, the Taylor series expansion implies at the output

$$|Y_1|^2 = \frac{A^2}{\sigma^2} f^2\left(\frac{\zeta}{\sigma}\right) + O\left(\left(\frac{A}{\sigma}\right)^4\right)$$

and

$$\bar{\sigma}_y^2 \approx 2\left[1 - F\left(\frac{\zeta}{\sigma}\right)\right] + O\left(\left(\frac{A}{\sigma}\right)^2\right).$$
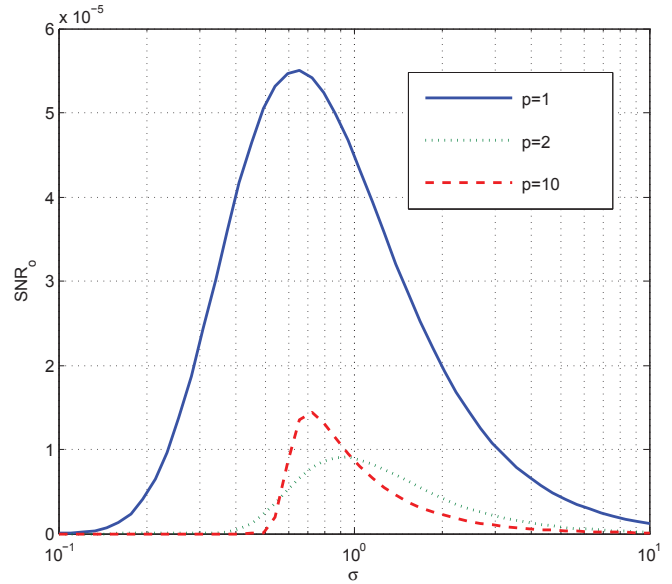


Fig. 2: Output $\text{SNR}_o$ as a function of the standard deviation of input noise, with threshold $\zeta = 1$ and $A = 0.01$. Input noise pdf is generalized Gaussian, with $p = 1, 2, 10$.

Thus, the output SNR and the SNR gain $G$ are given by

$$\text{SNR}_o \cong \frac{A^2 f^2\left(\frac{\zeta}{\sigma}\right)}{\sigma^2 \left[2 - 2F\left(\frac{\zeta}{\sigma}\right)\right]},$$

$$G = \frac{2f^2\left(\frac{\zeta}{\sigma}\right)}{1 - F\left(\frac{\zeta}{\sigma}\right)},$$

respectively, where $F$ is the cumulative distribution function of the normalized Generalized Gaussian.

$\text{SNR}_o$ as a function of $\sigma$ is shown in Fig. 2 when the quantization threshold $\zeta = 1$. Clearly, the maximum output SNR occurs when $\sigma = \sigma_{opt} > 0$, i.e., the output SNR is maximum when there exists an appropriate amount of noise. When $\sigma < \sigma_{opt}$, increasing the input noise $\sigma$ will increase $\text{SNR}_o$. Instead of adding noise to the input signal, the authors adjusted the quantizer threshold to maximize the output SNR and achieve positive SNR gains [50].

## B. Mutual Information Gain by Stochastic Resonance

As a well-known metric, mutual information (MI) measures the mutual dependence between two signals. SR has also been found to enhance the MI between input and output signals [51]–[56], thereby helping information flow through the nonlinear system [57]. While McDonnell, et al. point out that the capacity of an SR channel cannot exceed the actual capacity at the input due to the data processing theorem in information theory [58], Mitaim and Kosko showed that almost all noise pdfs produce some SR effect resulting in increased mutual information in threshold neurons [56]. A new statistically robust learning law was proposed to find the optimal noise level.

As an illustrative example, let us consider a single threshold image processing system, similar to the ones evaluated in [59] and also shown in Fig. 1. In such systems, the input gray level images are passed through a binary quantization system to convert with a relatively low threshold before presenting to the human subjects for visual inspection. Due to the extreme threshold, the resulting images are not clearly seen.

The output image $y$ is equal to $I((x + n) - T))$ where $x$ is the 'Lena' image, $n$ is the additive noise, $T$ is a predefined threshold and $I(x) = 1$ if $x \geq 0$ and 0 otherwise. Panes (a)-(e) of Fig. 3 show the effect of different noise levels. Clearly, the subjective performance is best when an intermediate level of noise is added. The mutual information between $X$ and $Y$ as a function of standard deviation $\sigma$ is shown in Fig. 4, the maximum MI value is obtained when $\sigma \neq 0$, i.e., when some SR noise is applied. Here, the input image has been normalized to $[0, 1]$ and $T = 0.9$. The joint distributions between the input and output images for mutual information calculations are estimated using the 2-D histogram.
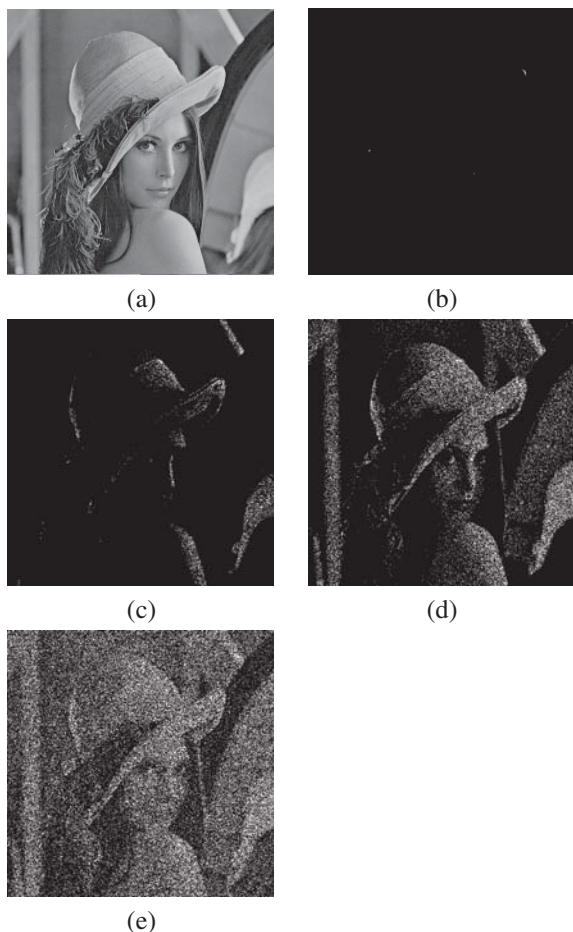


Fig. 4: The non-monotonic signature of stochastic resonance in mutual information



Fig. 3: Uniform noise can improve subjective image quality. (a) Original 'Lena' image; (b) No noise; (c) Little noise; (d) 'Just right' amount of noise; (e) Too much noise.

## III. STOCHASTIC RESONANCE IN BIOLOGICAL SYSTEMS

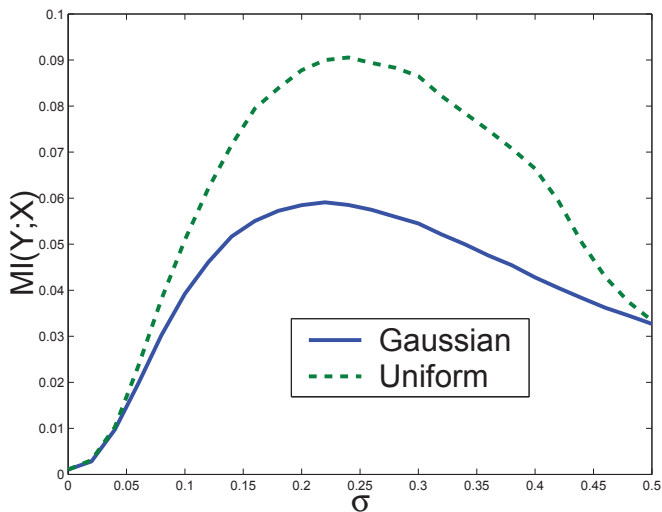The stochastic resonance phenomenon has been observed not only in a variety of theoretical models of information processing as illustrated in Sec. II and in natural systems in physics [23], but also in many experimental neuroscience studies on information processing either in single neurons or in neuronal networks [9], [60]–[62]. Further, it has been seen in biological information processing systems outside of neuroscience, such as biochemical reaction networks [63]. In this section, we briefly review some neurobiological examples of stochastic resonance, especially focusing on sensory processing. The goal is not to be comprehensive, but rather illustrative.

An early experiment demonstrating stochastic resonance in sensory processing was in the cricket cercal system. The cercal system detects the presence of either predators or other crickets by changes in air currents. In the experiment, cercal receptors were stimulated with naturalistic air currents modulated either at a single frequency or at multiple frequencies in the range due to predator attack, together with noisy broadband air currents as would occur in the natural environment. Spike trains from the cercal receptors to connected interneurons were recorded. Operating in the frequency domain, the SNR was extracted from the spike train, and further the mutual information between the air current stimulus and the spike train was computed. Both SNR and MI had maximum values at intermediate, non-zero levels of broadband noise, demonstrating the stochastic facilitation phenomenon [64].

An example of sensory stochastic resonance with wide-scale clinical applications is in tactile sensing and motor control [65]. Somatosensory function declines in people as we age, and further such changes are associated with diminished motor performance. Indeed diminished somatosensation in adults 65 years or older has been associated with increased likelihood of falling, since somatosensory feedback is crucial for balance. It has been demonstrated that input noise can enhance sensory [66] and motor [67] function, in measures such as SNR or accuracy. Thus, noise-based devices such as randomly vibrating shoe insoles are an effective way to enhance performance

on dynamic balance activities such as walking and enable older adults to overcome age-related instability. The same basic noise-enhancement principles are also useful for neural prosthetics for patients with loss of neural function [68].

SR has also been observed in human visual processing, and suggests methods for improving visual systems involving both people and machines. We will see this in Sec. V. Two interesting experiments to characterize primary visual processing were carried out at the level of individual neurons in the visual cortex of the human brain. First, it was found that when the signal is a contrast-reversing square-wave grating and flicker noise uniform in spatial and temporal frequency is added, then the evoked spike train in visual processing is enhanced [69]. Next, it was found that when the visual stimulus signal is a flickering 5 Hz wave and the noise is 10 Hz to 70 Hz roughly uniform flicker noise, this enhanced the 10 Hz entrainment response of the visual system [70].

Stochastic resonance has even been found in experiments on visual attention control, a higher human brain function. When subjects were asked to detect a weak gray-level target inside a marker box either in the left or the right visual field, signal detection performance was optimized by presenting randomly flickering noise between and outside the two possible target locations, because noise increased eye movement rates. Thus, noise at the cognitive level is associated with enhanced switching behavior between multi-stable attention states [71].

Stochastic resonance appears to be a prevalent design principle for neural information processing, but strong computational hypotheses are needed to understand it fully [9].

## IV. NOISE-ENHANCED SIGNAL DETECTION AND ESTIMATION

In this section, we consider some fundamental theories of noise-enhanced signal processing, especially for signal detection and estimation.

### A. Noise-Enhanced Binary Hypothesis Testing

In signal detection theory, noise may play a very important role in improving signal detectability. Some studies investigated the potential detection performance gain for certain suboptimal detection schemes with a few particular types of noise. In [38], [72], improvement of detection performance of a weak sinusoid signal via addition of noise is reported. To detect a constant signal in a Gaussian mixture noise background, Kay showed that under certain conditions, performance of the sign detector can be enhanced by adding some white Gaussian noise [73]. For another suboptimal detector, the locally optimal detector (LOD), Zozor and Amblard pointed out that detection performance is optimum when the noise parameters and detector parameters are matched [39]. A study of noise enhancement in quantizers by Saha and Anand showed that better detection performance is achieved by proper choice of quantizer thresholds [50] and for a fixed quantizer, by adding a suitable amount of noise.

In binary hypothesis testing, a likelihood ratio test (LRT) detector is optimal in both the Bayesian and Neyman-Pearson

frameworks. However, implementing the LRT detector requires complete knowledge of the pdfs $p_0(\cdot)$ and $p_1(\cdot)$ under the respective hypothesis $H_0$ and $H_1$, which may not be available in practice. Also, the input data statistics may vary with time or may change from one application to another. To make matters worse, there are many detection problems where the exact form of the LRT is too complicated to implement. Therefore, simpler and more robust suboptimal detectors are often used [74].

In [75], [76], improving performance of any given detection system through additive noise is considered. Some of the findings are summarized as follows. Consider a general problem with observations $\mathbf{x} \in \mathcal{X}$ and known probability distribution under both hypotheses.

- Binary hypotheses.

$$H_0 : p_{\mathbf{X}}(\mathbf{x}; H_0) = p_0(\mathbf{x})$$
$$H_1 : p_{\mathbf{X}}(\mathbf{x}; H_1) = p_1(\mathbf{x}),$$

  where $p_0(\mathbf{x})$ and $p_1(\mathbf{x})$ are the pdfs of $\mathbf{x}$ under $H_0$ and $H_1$, respectively.
- Decision function $\phi(\mathbf{x}) \in [0, 1]$. The detector for this problem can be completely characterized by a decision function such that the detector output $U = H_1$ with probability $\phi(\mathbf{x})$.
- Additive noise $\mathbf{n} \in \mathcal{N}$. As an attempt to improve system performance, an independent additive noise $\mathbf{n}$ with pdf $p_{\mathbf{n}}(n) \in \mathcal{P}_{\mathcal{N}}$ is added to the observed data $\mathbf{x}$ to obtain a "new" observation $\mathbf{y} = \mathbf{x} + \mathbf{n} \in \mathcal{Y}$ as the input to the detector. $\mathcal{X}, \mathcal{Y}, \mathcal{N}, \mathcal{P}_{\mathcal{N}}$ are the domains of $\mathbf{x}, \mathbf{y}, \mathbf{n}$, and $p_{\mathbf{n}}(n)$, respectively.
- Performance measures. The detection performance is evaluated via three key metrics: probability of detection $P_d = P(U = H_1; H_1)$, probability of false alarm $P_f = P(U = H_1; H_0)$ under the Neyman-Pearson framework, and probability of error $P_e = \pi_1(1 - P_d) + \pi_0 P_f$ where $\pi_i = P(H_i)$ is the prior probability of $H_i$ under the Bayesian setting, $i = 0, 1$.

For this binary detection problem, with the noise $\mathbf{n}$ added, the probability of detection is given by

$$\begin{aligned} P_d &= E_1(\mathbf{y}) \qquad\qquad (2)\\ &= E_1 E_{\mathbf{n}}\left(\phi(\mathbf{x} + \mathbf{n})\right)\\ &= E_{\mathbf{n}}\left\{E_1\left(\phi(\mathbf{x} + \mathbf{n})\right)\right\}\\ &= E_{\mathbf{n}} F_1(\mathbf{n}), \end{aligned}$$

where $E_i(\cdot)$ is the expected value under $H_i$, $E_{\mathbf{n}}$ is the expectation with noise distribution $p_{\mathbf{n}}(n)$ and $F_i(\mathbf{n}) \equiv E_i\left(\phi(\mathbf{x} + \mathbf{n})\right)$, $i = 0, 1$.

Similarly, the noise-modified probability of false alarm and probability of error become

$$P_f = E_0(\mathbf{y}) = E_{\mathbf{n}} F_0(\mathbf{n}), \qquad (3)$$

and

$$P_e = \pi_1 + E_{\mathbf{n}}\left(\pi_0 F_0(\mathbf{n}) - \pi_1 F_1(\mathbf{n})\right) = \pi_1 + E_{\mathbf{n}}\left(F_e(\mathbf{n})\right), \quad (4)$$

where $F_e(\mathbf{n}) \equiv \pi_0 F_0(\mathbf{n}) - \pi_1 F_1(\mathbf{n})$.

*1) Optimal Noise Distributions:* In this subsection, we discuss the effect of the noise distribution in detection performance and optimal noise distributions that achieve the best possible performance, i.e. the best $p_{\mathbf{n}} \in \mathcal{P}_{\mathcal{N}}$. Noise-enhanced detectors can be treated as equivalent to randomization (due to the noise distribution) between detectors corresponding to a particular noise level. Due to the complexity of the problems, next we focus on finding *simple* noise distribution forms that attain the maximum achievable detection performance and the conditions on whether a detector can be improved or not.

Consider the case where $\mathcal{P}_{\mathcal{N}}$ is the set of all possible probability distributions, though discussion can be easily extended to cases where the distribution functions are limited to certain sets. Let $\mathcal{F}_i$ be the domain of $F_i(\mathbf{n})$, $i = 0, 1, e$. Clearly, under the Bayesian setting, when the minimum exists in $\mathcal{F}_e$, i.e., when $\mathcal{F}_e$ is a closed set, $P_e$ can be minimized by letting $\mathbf{n} = \mathbf{n}_0 \equiv \arg\min_{\mathbf{n} \in \mathcal{N}} F_e(\mathbf{n})$ with probability 1. That is, the optimal "noise" is a constant signal $\mathbf{n_0}$ under the Bayesian setting. However, when the domain of $\mathcal{F}_e$ is not a closed set, the minimum may not exist. When it happens, certainly there does not exist a noise distribution which can obtain the so-called "minimum". Nevertheless, the "noise" with a single constant level can obtain the detection performance that any other optimal noise distribution can achieve. Overall, we conclude that the optimal noise form to minimize the probability of error is a single constant noise such that

$$p_{\mathbf{n,e}}^{\mathbf{o}} = \delta(\mathbf{n} - \mathbf{n}_0), \tag{5}$$

where $\delta(\cdot)$ is the Dirac delta function.

The optimization problem under the Neyman-Pearson framework is similar. Due to the extra constraint on the probability of false alarm $P_f \leq \alpha$, from Carathéodory's theorem in convex optimization theory [77], under this linear constraint on noise distributions, we can conclude that the optimal noise form is a randomization of at most two constant levels such that

$$P_{\mathbf{n}}^{\mathbf{o}} = (1 - \lambda)\delta(\mathbf{n} - \mathbf{n}_1) + \lambda\delta(\mathbf{n} - \mathbf{n}_2), \tag{6}$$

where $\lambda \in [0, 1]$, $\mathbf{n}_1, \mathbf{n}_2$ are appropriate noise parameters to be tuned to satisfy the constraint on probability of false alarm while maximizing the probability of detection. In other words, it is sufficient to only consider the "2-peak" noises of the form given in (6) when designing and optimizing noise-enhanced detectors under the Neyman-Pearson criterion.

Improvability of the given detector when the noise is added can be determined by computing and comparing $P_{f,opt}^y$ and $P_D^x$ where the latter is the special case of $\mathbf{n} = 0$. When $P_{d,opt}^y > P_d^x$, the given detector is improvable by adding additional noise. However, it requires the complete knowledge of $F_1(\cdot)$ and $F_0(\cdot)$ and significant computation. To determine the improvability of the detector under this noise-enhanced framework, let us now consider a function $J(t)$ such that $J(t) = \max(F_1 : F_0 = t)$ is the maximum value of $F_1$ given $F_0(\mathbf{n}) = t$. Clearly, $J(P_f^x) \geq F_1(0) = P_d^x$. It follows that for any noise $p_{\mathbf{n}}$, we have $P_D^y(p_{\mathbf{n}}) \leq E_{\mathbf{n}}J(F_0(\mathbf{n}))$. Therefore, the optimum $P_d^y$ is attained when $F_1(F_0(\mathbf{n})) \leq J(f_0)$ and $P_{D,opt}^y = E_{\mathbf{n}}(J)$. For a large class of detectors, defined by the local properties of $J$, we may determine sufficient conditions

for improvability and non-improvability more easily. Proofs of the following theorems can be found in [75].

**Theorem 1** (Improvability of detection via noise addition)**.** *If $J(P_f^x) > P_d^x$ or $J''(P_f^x) > 0$ when $J(t)$ is second-order continuously differentiable around $P_f^x$ (local convexity), then there exists at least one noise process $\mathbf{n}$ that can improve the detection performance.*

**Theorem 2** (Non-improvability of detection via noise addition)**.** *If there exists a non-decreasing concave function $\Psi(t)$ where $\Psi(P_{FA}^x) = J(P_{FA}^x) = F_1(0)$ and $\Psi(t) \geq J(t)$ for every $t$, then $P_D^y \leq P_D^x$ for any independent noise, i.e., the detection performance cannot be improved by adding noise.*

When the domain of $\mathcal{F}_0$ or $\mathcal{F}_1$ is not a closed set, the maximum/minimum may not exist. However, the optimal noise forms proposed below can above the detection performance that any other noise distribution can. This observation as well as optimal noise form determination under certain linear constraints were also reported in [78].

In the above discussion, the detector was assumed fixed and we could only add noise to the input signals. The above results hold true even when the detector $\phi(\cdot)$ can be changed/adjusted along with the choice of noise distributions [76]. In fact, adding/changing the input noise distribution can be considered as a particular way of adjusting the detector.

When the additive noise is constrained to a particular distribution family but with tunable parameters (like standard deviation), a set of forbidden interval theorems to determine the improvability of SR effects were proven. These theorems are used in a range of applications such as signal detection in carbon nanotubes and quantum computing, with a variety of noise distributions [21], [79]–[83].

Besides improving detection performance, adding noise can also improve the security performance of a network. It has been shown in [84], [85] that for a distributed inference network consisting of malicious sensors [86], an appropriate addition of noise makes the system more robust to attacks and increases the minimum number of attacked sensors required to deteriorate the network's performance.

*2) A Detection Example:* Consider detecting a constant signal in a mixture Gaussian noise background, a problem first considered in [73] and later revisited in [75], [76]. The two hypotheses $H_0$ and $H_1$ are given as

$$H_0 : x[i] = w[i]$$
$$H_1 : x[i] = A + w[i]$$

for $i = 0, 1, \cdots, N - 1$, $A > 0$ is fixed and known, and $w[i]$ are i.i.d. noise samples with a symmetric Gaussian mixture noise pdf $p_w(w) = \frac{1}{2}\gamma(w; -\mu, \sigma_0^2) + \frac{1}{2}\gamma(w; \mu, \sigma_0^2)$ where

$$\gamma(w; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(w - \mu)^2}{2\sigma^2}\right]$$

is the Gaussian pdf with mean $\mu$ and variance $\sigma^2$. A simple but suboptimal sign detector is employed such that when null hypothesis $H_0$ is rejected when the number of positive
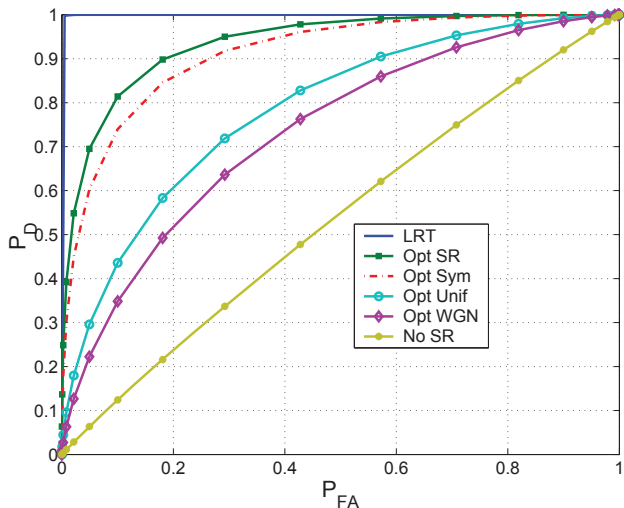
Fig. 5: ROC curves for different SR noise enhanced sign detectors, $N = 30$ [75, Fig. 9]: LRT is optimal likelihood ratio test; Opt SR is noise-enhanced detector with optimal i.i.d. noise; Opt Sym is noise-enhanced detector with optimal symmetric i.i.d. noise; Opt Unif is noise-enhanced detector with optimal i.i.d. uniform noise; Opt WGN is noise-enhanced detector with optimal symmetric i.i.d. Gaussian noise; and No SR is the original detector with no noise added. LRT performance is nearly perfect ($P_D \approx 1$ for all $P_f$ values).

observation samples is greater than a given threshold. When there is only one observation ($N = 1$), the detector test statistic becomes $x$ with threshold $\eta = 0$.

Depending on the parameter settings, the detection performance of the sign detector can be very poor. For example, when $A = 3, \mu = 3, \sigma = 1$, the probability of false alarm is $P_f^x = 0.5$ and the probability of detection is $P_d^x = 0.5114$. The problem of determining the optimal "noise" to be injected to the system is to determine the optimal $p(n)$ where for the new observation $y = x + n$, the probability of detection $P_D^y = p(y > 0; H_1)$ is maximum without increasing the probability of false alarm. From before, the simplest optimal noise form is no more than randomization between two constant values. When $A = 3, \mu = 3, \sigma = 1$ the optimal noise pdf is $p_n^{opt}(n) = 0.3085\delta(n + 3.5) + 0.6915\delta(n - 2.5)$ [75]. By injecting this noise, the resulting $P_d^y = 0.6915$, far greater than the original $P_d^x = 0.5114$.

Determining the optimal noise for $N > 1$ is more difficult as it involves an optimization problem with $2N$ parameters. However, due to the particular structure of this detector, we can inject i.i.d. noise to each data sample $x[i]$. Fig. 5 shows the receiver operating characteristic (ROC) curves when $N = 30$ and different types of i.i.d. noise added; different degrees of improvement are observed for different noises. The optimal SR detector and the optimal symmetric SR detector performance levels are superior to those of the optimal uniform and optimal Gaussian SR detectors and more closely approximate the LRT curve.

*3) Sequential Detection:* Aiming to reduce the expected sample size (ESS) to make a decision with a given desired

accuracy, the potential of stochastic facilitation in sequential detection was investigated in [87], [88]. Sequential detection techniques have been widely applied in a broad range of applications where observations are collected sequentially [89]. For a binary detection problem with completely known distribution functions under both hypotheses, the sequential probability ratio test (SPRT) achieves the same probability of decision errors with minimum expected sample size (ESS) among all possible detectors [90]. In practice, data at the input of a sequential detector may be degraded since measurements may be subject to impairments like quantization or transmission over noise channels [91]–[93]. Such impairments may also arise for some nonparametric sequential detectors [94], [95], resulting in significant performance loss. Development of methods to combat performance loss without changing the system structure is highly desirable.

Since sample size is often not a constant in sequential detection, optimal noise distributions in (5) or (6) are no longer applicable. In [87], the sequential detection problem was investigated under a framework where data samples are collected sequentially and transformed one sample at a time through a nonlinear, memoryless transformation system, and the SPRT is applied at the output.

If the additive noise introduced at the input is i.i.d., the ESS under the hypothesis $H_0$ ($H_1$) is minimized by maximizing the corresponding Kullback-Leibler divergence (KLD) between output sample distributions under $H_0$ and $H_1$ ($H_1$ and $H_0$). Using the concavity of the KLD [96], it can be shown the optimal noise is a constant vector per data sample, under either hypothesis. However, the optimal noise form to minimize the overall average sample size is still unknown as it involves optimization of the geometrical average of KLDs. The stochastic facilitation effect in suboptimal sequential detectors (not SPRT-based) for shift-in-mean binary hypothesis testing problems with detectors is investigated in [88]. It is found that certain sequential detection procedures can be made more efficient by randomly adding or subtracting a suitable constant value to the data at the input of the (suboptimal) detector, similar to the conclusions in [87].

### B. Noise-Enhanced Parameter Estimation

Noise enhancement can also be performed for estimation problems when the estimator is a nonlinear system. It has been shown that the estimation performance may be improved by adding a suitable noise at the input of such systems. For example, in [97], for a distributed estimation problem, a better Cramer-Rao lower bound has been obtained by adding suitable noise to the observed data at local sensors before quantization. In [43], the problem of estimating the frequency of a periodic wave employing the optimum Bayesian estimator based on the output of the one bit quantization system is evaluated and a non-monotonic relationship between the estimation performance and noise power is demonstrated. In [98], noise-enhanced systems for general parameter estimation problems have been considered, not only for additive noise but also for other non-additive noise scenarios. The problem formulations and main findings are summarized as follows.

Similar to the binary hypothesis detection setting, we consider an estimator that consists of an input signal $\mathbf{x}$, a parameter to be estimated $\theta$, and an estimator $\hat{\theta} = T(\mathbf{x})$ which infers $\theta$ from $\mathbf{x}$. The estimator performance is evaluated by a risk/cost function $r_i(\theta, \hat{\theta})$, $i = 1, \ldots, I$ where $I$ is the total number of metrics. The risk/cost can also be evaluated in the average sense such that $R_i = E[r_i(\theta, \hat{\theta})]$. Many widely employed estimation metrics can be characterized under this framework. For example the bias, covariance matrix, and mean-square-error (MSE) of the estimator are, respectively:

$$B(\theta) = E_{\mathbf{x}}(\hat{\theta} - \theta),$$
$$V(\theta) = E_{\mathbf{x}}(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T,$$
$$M = E_{\mathbf{x}} E_{\theta}(\hat{\theta} - \theta)^T(\hat{\theta} - \theta).$$

Instead of merely focusing on additive noise, we consider a general form of stochastic facilitation. Here, the relationship between the noise modified input signal $\mathbf{y}$, the input signal $\mathbf{x}$ and the introduced $\mathbf{n}$ is determined by $p(\mathbf{y}|\mathbf{x}, \mathbf{n}) = \zeta(\mathbf{x}, \mathbf{n})$ where $\zeta$ is a prespecified stochastic transform function. The modified input signal $\mathbf{y}$ is then used for the estimation process. Similar to the detection problem, the fundamental optimization questions become: Given the estimator $T$ and the transform $\zeta$, determine the simplest noise distribution $p(\mathbf{n})$ for any achievable $R_I$ subject to the risk/cost constraints

$$R_i \leq \alpha_i, i = 1, \cdots, I - 1. \tag{7}$$

For this general noise-modified estimator, notice the cost functions can be rewritten as

$$\begin{aligned} R_i &= Er_i\left(\theta, \hat{\theta}(\mathbf{y})\right) \\ &= E E_{\mathbf{n}} r_i\left(\theta, \hat{\theta}(\mathbf{y}(\mathbf{x}, \mathbf{n}))\right) \\ &= E_{\mathbf{n}}\left\{ Er_i\left(\theta, \hat{\theta}(\mathbf{y}(\mathbf{x}, \mathbf{n}))\right)\right\} \\ &= E_{\mathbf{n}} R_{\mathbf{n}, i}(\mathbf{n}). \end{aligned}$$

For optimization, it can be shown it is sufficient to limit consideration to "I-peak" noises with distributions

$$p_{\mathbf{n}}(\mathbf{n}) = \sum_{i=1}^{I} \lambda_i \delta(\mathbf{n} - \mathbf{n}_i), \tag{8}$$

where $\lambda_i \geq 0, \sum \lambda_i = 1$, and $\mathbf{n_1}, \cdots, \mathbf{n_I}$ are suitable constant vectors.

For example, if the overall goal is to minimize the MSE, then the optimal noise is a single constant, which is equivalent to choosing a single noise parameter $\mathbf{n}_o$. Similarly, for a scalar estimation problem where the goal is to reduce the estimation variance while keeping the unbiasedness of the estimator, the optimal noise form is $P_{\mathbf{n}}^{\mathbf{o}} = (1 - \lambda)\delta(\mathbf{n} - \mathbf{n}_1) + \lambda\delta(\mathbf{n} - \mathbf{n}_2)$.

With some transformations and reformulations, the basic framework established here can be extended to cover other scenarios, e.g., the cases with further constraints on noise distributions, such as noise power constraints.

**Theorem 3.** *In addition to the constraints on the risk/cost functions, if there are other "linear" constraints on the noise distributions such as*

$$C_j = E_{\mathbf{n}}(c_j(\mathbf{n})) \leq \epsilon_j,$$

*for* $j = 1, \cdots, J$, *then it is sufficient to consider the "$I + J$" peak noises such that* $p_{\mathbf{n}}(\mathbf{n}) = \sum_{i=1}^{I+J} \lambda_i \delta(\mathbf{n} - \mathbf{n}_i)$.

In some applications, especially while studying biological systems, sources of noise are often inherent in the system [47], [99] and are not easily adjustable except possibly tuning a few parameters. In such cases, $\mathcal{P}_{\mathcal{N}}$ is no longer the set of all possible noise distributions, but can be, say, a mixture of certain types of distributions, like a Gaussian mixture resulting from randomization between different Gaussian noise sources. The same analysis procedure still applies in these cases and the optimal noise forms remain the same except replacing the $\delta(\cdot)$ function with the suitable noise distribution kernels. By relating the risk/cost function $R_i$ with $P_d$, $P_f$ and $P_e$, it can be shown that the same noise forms are also optimal for noise-enhanced detectors.

We point out here that the optimal noise forms obtained above are mainly for the optimization problems with linear constraints on noise distribution. There are many applications where the constraints are nonlinear, for example, when the performance is measured by the geometrical average of KLDs or when the input noises are i.i.d. and the system performance depends on convolutions of noise distributions. Other than determining the optimal parameters for certain distribution families [21], little is known for the optimal noise distributions in the nonlinear cases.

## V. NOISE-ENHANCED IMAGE PROCESSING

The structure of and algorithms in image processing systems vary significantly from one application to another; performance evaluation is also carried out via different metrics [100]. Image modeling plays an important role in image processing theory. Whether explicitly or implicitly, image compression, image restoration, image visualization, and other image processing applications can benefit from a suitable statistical image model [59]: performance is maximized when the underlying model fits the actual image [101]. Due to the nature of images, a perfect model is usually very difficult if not impossible to find. That is, no matter how one models the image, some degree of mismatch always exists. As a result, no matter how sophisticated it is, an image processing algorithm may not achieve best visual performance even if it is the optimum based on the assumed model.

Noise enhancement has been observed and applied to some image processing systems in the past. For a subthreshold image in a threshold detector, it has been shown that adding some noise to the image before thresholding can improve the visual perception of the thresholded images [59]. The constructive roles of noise in human vision are further examined in [10], [11], [102]–[104].

For a prespecified type of noise, the optimal level of additive noise to maximize mutual information between input and output images was determined in [56]. For a binary quantizer, the relationship among the expected value of the output image, the noise distribution, and the input image is discussed in [105]. Stochastic facilitation has also been applied to the Radon transform to extract weak lines from

a noisy image and some other image enhancement applications in [106]. Some applications of stochastic resonance for image denoising, enhancement, and edge detection have been reported in [104], [107]–[113]. A suitable amount of noise has also been shown to improve image segmentation quality [114], [115], watermark/logo recognition, and image resizing detection [116]–[118]. Utilizing noisy quantizer arrays, the SSR effect was also explored in watermark detection [46].

At first glance, one may cast noise-enhanced image processing problems into the estimation framework, and determine that the optimal noisy image to be "injected" to the image processing system is a suitable randomization of several different images. This approach, although theoretically sound, underestimates the huge complexity of image processing problems due to image size, image modeling and performance evaluations. Determination of the optimal image is nearly impossible in practice. Therefore, in many applications, the noises to be added are restricted to certain types of noises such as images generated with i.i.d. Gaussian noises.

Under this i.i.d. noise assumption, except for a few notable exceptions, the constraint on the noise distribution is no longer "linear", and as a result, the results obtained in the previous section cannot be directly applied. To improve the quality of noise-enhanced image processing systems, one can use more than one noise-modified system in parallel, each with the same or different noise distribution. The final image is obtained via data fusion of the output images.

Next, let us give a rather counter-intuitive example, a noise-enhanced image denoising system which improves denoising performance by adding more noise to the input image.

**Example 1.** Let us examine the possibility of improving a median filter via the addition of noise. For a noisy image, the median filter $\mathrm{Med}(L)$ estimates the original image value by replacing each pixel value with the median value of its $(2L+1) \times (2L+1)$ local neighborhood. Let the original noise-free image be $\mathbf{x}_0$, the cost function the MSE between the estimated image and the noise-free image.

Next, let us examine the denoising performance when the original image is contaminated by a symmetric Gaussian mixture noise with mean $\pm\mu$ and Gaussian variance $\sigma^2$. In this experiment, the frequently used 'Lena' image of size $512\times512$ is tested. Here, due to the large dimensionality of the image, the optimal constant image to be added is very difficult to find. Alternatively, a random Gaussian noise image with zero mean and $\sigma_n^2$ variance is applied in the single noise framework. We consider five different noise-enhanced systems: two single-noise-modified image processing systems (SNMIPSs) are employed here with $\sigma_n = 10$ and $30$ respectively, one multiple-noise-modified image processing system (MNMIPS) with 9 subsystems with each noisy image being a random Gaussian image with zero mean and standard deviation $\sigma_i = 5(i-1)$ for $i = 1, 2, \cdots, 9$; and another two INMIPS (identical-noise-modified image processing systems) with $\sigma_n = 10$ and $\sigma_n = 30$, respectively. The experiment is repeated 30 times and the system performance is evaluated in terms of their means and variances as shown in panes (a) and (b) of Fig. 6.

It can be seen that adding some noise to the observed image can improve the denoising performance significantly.

For example, when $\sigma = 10$, for the single noise system, the mean MSE is reduced from $361.8$ for the original system to $284.1$ and $227.8$ for SNMIPS with $\sigma_n = 10$ and $30$ respectively. The denoising performance is further improved when an MNMIPS is employed. The MNMIPS using different noises achieved a MSE of $200.6$ and the mean MSE is $260.8$ for an INMIPS with $\sigma_n = 10$ and is $153.6$ for the INMIPS with $\sigma_n = 30$. The uncertainty of denoising performance is also shown to decrease when an MNMIPS is applied. Compared to the SNMIPSs, the MSE variance of the MNMIPSs are also much smaller which indicates a more stable and predictable performance.
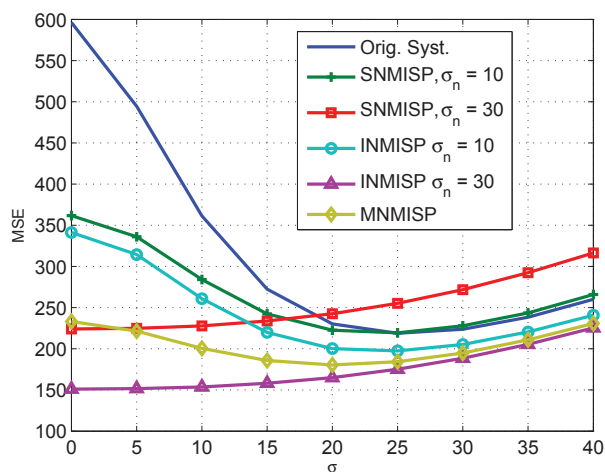


(a) Mean MSE of the noise modified median filters as a function of the original noisiness $\sigma$.
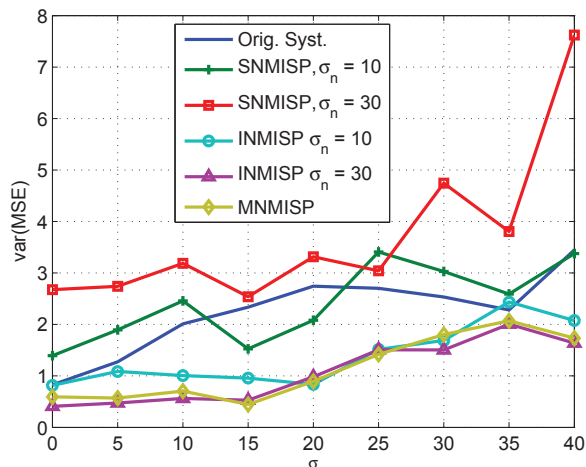


(b) Variance of the MSE of the noise modified median filters as a function of the original noisiness $\sigma$.

Fig. 6: Denoising performance of the noise modified median filters for a mixture Gaussian noise contaminated 'Lena' image with $\mu = 30$.
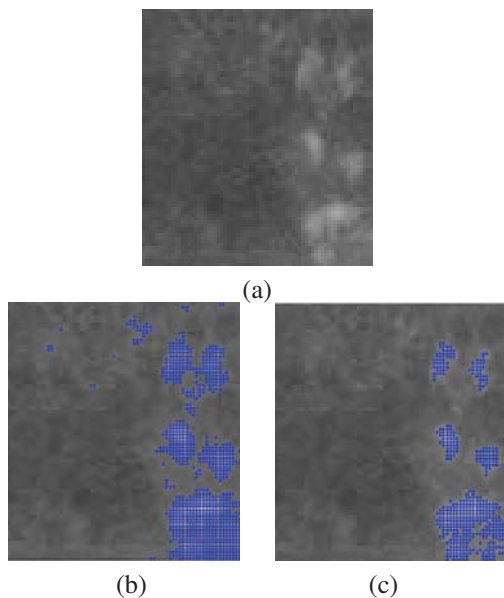
Fig. 7: Comparison of Lesion Detection Performance (a) the original mammogram image. (b) Detection Performance of the iterative mode separation (IMS) method [121]. (c) Noise Enhanced IMS Detection Result [120, Fig. 2].

### A. Noise-Enhanced Detection of Micro-Calcifications in Digital Mammograms

There is clear evidence showing early diagnosis and treatment of breast cancer can significantly increase the chance of survival for patients. One of the important early symptoms of breast cancer in mammograms is the appearance of micro-calcification clusters and so accurate detection of micro-calcifications is highly desirable to ensure early diagnosis [119]. Many existing detectors treat lesion detection as an anomaly detection problem using some particular known background models. These models are often based on Gaussian assumptions. However, in practical datasets, the Gaussian assumptions are often not true. As a result, the detection performance of those lesion detectors is degraded due to model mismatch.

In [120], a noise enhancement approach to improve some micro-calcification detectors has been proposed wherein suitable noise is added to the digital mammograms before applying the detectors. Instead of relying on a prespecified model and parameters, the additive noise distribution in this case is learnt adaptively. The framework and algorithms were tested on a set of 75 representative abnormal mammograms. They yield superior performance when compared with several existing classification and detection approaches available in the literature. One detection result of an image in the testing dataset is presented in Fig. 7. For more detailed information and results, please refer to [120].

### VI. NOISE-ENHANCED SEARCH ALGORITHMS

Another area where introduction of randomness enhances performance is optimization. This is especially true when search for an optimum is likely to get trapped in local minima.

In these cases, randomization assists in the search for optimal or near-optimal solutions. As an example of a search technique for the optimum solution, a genetic algorithm reformulates the parameters of a solution to a chromosome often written as a sequences of binary bits. The optimum solution is searched by creating and reproducing sets of chromosomes to find the optimum solution under predefined conditions. In GA, the new chromosomes are created by both *crossover* where a fraction of the chromosomes are swapped between two chromosomes and *mutation* where some bits of the chromosomes are changed randomly with certain probabilities [122]. With the randomness introduced using *crossover* and *mutation*, GA avoids local minima by preventing the population of chromosomes from becoming too similar to each other. Note that here, the role of mutation is actually similar to injecting noise into the system. That is, if the mutation operation changes a bit $x_i$ with probability $\alpha$, the same result can be obtained by adding noise $n_i$ to it such that the noisy version of the signal $x_i \oplus n_i$ instead of the noise free $x_i$ is used for future optimization. Here, '$\oplus$' is the XOR operator with and $n_i$ is an independent Bernoulli random variable with $p(n_i = 1) = \alpha$. Also similar to dithering, choice of a suitable mutation rate is very important to the system.

In simulated annealing (SA), to search for an optimal solution, a random sequence of solutions is generated that eventually converges to a final solution. Unlike many other approaches, SA involves a temperature parameter $T$ to control the randomness of the search procedure. Higher the $T$, better the chance that SA will accept a solution which is worse than the current one. This probability goes down as $T$ becomes smaller. In the limit, $T$ eventually reaches zero and a final decision is made. The randomness in SA can also be modeled as noise. Compared to dithering and GA, the noise variance goes down as $T$ becomes smaller and equals 0 when $T = 0$ where the randomness disappears.

In learning with discrete finite-state Markov chains, addition of a suitable amount of noise helps the Markov chain explore improbable regions of the state space, improving convergence speed to equilibrium [123]. Findings for noise-enhanced hidden Markov models and speech recognition were reported in [124]. The positive effects of noisy was also shown to speed up the average convergence of the Expectation-Maximization (EM) algorithm [125], and centroid-based clustering algorithms such as K-means algorithm [126], under certain conditions.

One interesting application of randomization of inputs is in the complexity analysis of algorithms. Conventional analysis of algorithms is carried out in two ways, either by worst-case or average-case analysis. Neither of these approaches provide realistic and convincing complexity estimates. In [127], a smoothed analysis approach is presented where slight random perturbations of arbitrary inputs are employed to evaluate algorithm complexity. This approach has been used to show, for example, that the simplex method for linear programming has polynomial smoothed complexity.

## VII. Noise-Enhanced Associative Memory Recall

In the last section, we saw how noise can facilitate search-based optimization procedures, essentially by smoothing out rough energy surfaces with many local minima. Here we consider the recall phase of associative memory, which is essentially a form of nearest-neighbor search. The associative memory model is how many brain regions such as olfactory cortex and hippocampus are thought to work, and stochastic facilitation has been noted in hippocampal memory [128]. The basic idea is that a set of patterns are memorized using a learning rule in the training phase.[1] Then noisy or incomplete versions of these patterns are presented during recall: the memory produces the closest match.

Besides aiding in understanding neurobiology, content-addressable associative memories are of growing technological importance in the era of big data. Storage systems need not only store more and more information (as location-addressed memory systems also do), but help in determining whether there is data relevant to the task at hand, and then retrieving it. The goal is to store a desired set of states—the memories—as fixed points of a network such that errors in an input representation of a memory are corrected and the memory retrieved. One approach for building content-addressable associative memory is to use ideas from modern coding theory, where the matrix of synaptic weights is like a code matrix [129].

Traditional models of associative memory recall have assumed that the algorithm and circuit implementation are noiseless, however it has recently been shown that adding noise in algorithm steps can improve the final error probability achieved [129]. Here we review this coding-theoretic associative memory architecture and recall algorithm.

The set of patterns stored via the learning algorithm span a subspace of the larger space of possible patterns, and may be represented using a bipartite graph with variable nodes and check nodes that enforce subspace constraints among variables. The bipartite graph is organized in a clustered fashion similar to the cortical column structure of the mammalian brain, where nodes within a cluster are well-connected and there is a small level of connection between clusters. In recall, a noisy version of a pattern is presented and a two-level iterative message-passing algorithm is used to propagate information so as to perform error correction by local computations to enforce subspace constraints. The first level of the algorithm operates within a cluster like belief propagation, whereas the second level spreads information between clusters like sequential peeling algorithms in decoding [130]. The algorithms operate iteratively.

As part of noise enhancement, any message passed in the algorithm is perturbed with additive noise. Messages that go from variable-to-check nodes have noise level $\upsilon$ whereas messages that go from check-to-variable nodes have noise level $\nu$. Let the fraction of external errors (as part of the query) corrected by a noiseless recall algorithm after $T$ iterations be $\Lambda(T)$ and that of a recall algorithm with internal noise be $\Lambda_{\upsilon,\nu}(T)$, for the same set of patterns memorized with the same storage capacity. Further let the $T \to \infty$ values be $\Lambda^*$ and $\Lambda_{\upsilon,\nu}^*$.

**Theorem 4.** *For an appropriately chosen design (such that a noiseless query is successful), for the same realizations of external errors, $\Lambda_{\upsilon,\nu}^* \geq \Lambda^*$.*

The high-level idea why a noisy network outperforms a noiseless one comes from understanding stopping sets—realizations of external errors where sequential peeling cannot make progress towards error correction. Stopping sets shrink as internal noise is added, and so in the $T \to \infty$ limit, the noisy network can correct any error pattern that can be corrected by the noiseless version and it can also get out of stopping sets that cause the noiseless network to fail. Although the theorem does not say whether the noisy neural network may need more iterations to achieve the same error correction performance, empirical experiments demonstrate many settings where even the running time improves when using a noisy network.

Noise facilitates recall in associative memory, but since the basic approach is the same as in other iterative message-passing algorithms, whether for decoding low-density parity-check codes or other applications [131], we conjecture the same stochastic facilitation property will hold in finitary regimes for these other problems (cf. [132] and references therein).

## VIII. Conclusion

In this paper, we presented a state-of-the-art review of several seemingly unrelated research areas with the underlying common theme of system performance enhancement due to introduction of some noise or random transformations in the system. Such phenomenon had been observed in many biological systems over the years but had not been explored in the context of performance enhancement of information systems. Most of our presentation was illustrative in nature in that many examples were provided. In some instances such as detection and estimation, the noise-enhanced system framework has been formulated mathematically and optimum noise to be introduced has been determined. It has been observed that noise has a tendency to convexify problems which leads to improved performance. This is the case in detection, where stochastic facilitation is intimately related to randomized decision rules [133]. However, mathematical formulation of this bio-inspired phenomenon in most cases is still in its infancy and much research still needs to be done. Mathematical models to characterize the phenomenon need to be developed and achievability results in terms of performance need to be derived. We presented a number of examples where promising and encouraging results have been obtained. Finding new areas where this phenomenon could be exploited is expected to be a fruitful endeavor.

## Acknowledgment

---

[1] Interestingly, adding noise to training data enhances certain neural network learning algorithms [124].

### REFERENCES

[1] L. Cohen, "The history of noise," *IEEE Signal Process. Mag.*, vol. 22, no. 6, pp. 20–45, Nov. 2005.

[2] B. Kosko, *Noise*. Viking Adult, 2006.

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

[4] E. L. Lehmann, *Testing Statistical Hypotheses*, 2nd ed. John Wiley & Sons, 1986.

[5] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ: Prentice Hall PTR, 1993.

[6] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of Machine Learning*. Cambridge, MA: MIT Press, 2012.

[7] W. Maass, "Noise as a resource for computation and learning in networks of spiking neurons," *Proc. IEEE*, vol. 102, no. 5, pp. 860–880, May 2014.

[8] F. Moss, L. M. Ward, and W. G. Sannita, "Stochastic resonance and sensory information processing: a tutorial and review of application," *Clin. Neurophysiol.*, vol. 115, no. 2, pp. 267–281, Feb. 2004.

[9] M. D. McDonnell and L. M. Ward, "The benefits of noise in neural systems: bridging theory and experiment," *Nat. Rev. Neurosci.*, vol. 12, no. 7, pp. 415–426, Jul. 2011.

[10] E. Sejdic and L. A. Lipsitz, "Necessity of noise in physiology and medicine," *Comput. Methods Programs Biomed.*, vol. 111, no. 2, pp. 459–470, Aug. 2013.

[11] M. Wang, R. Sun, W. Huang, and Y. Tu, "Internal noise induced pattern formation and spatial coherence resonance for calcium signals of diffusively coupled cells," *Physica A*.

[12] T. J. Hamilton, S. Afshar, A. van Schaik, and J. Tapson, "Stochastic electronics: A neuro-inspired design paradigm for integrated circuits," *Proc. IEEE*, vol. 102, no. 5, pp. 843–859, May 2014.

[13] L. G. Roberts, "Picture coding using pseudo-random noise," *IRE Trans. Inf. Theory*, vol. IT-8, no. 2, pp. 145–154, Feb. 1962.

[14] L. Schuchman, "Dither signals and their effect on quantization noise," *IEEE Trans. Commun. Technol.*, vol. 12, no. 4, pp. 162–165, Dec. 1964.

[15] J. O. Limb, "Design of dither waveforms for quantized visual signals," *Bell Syst. Tech. J.*, vol. 48, no. 7, pp. 2555–2582, Sep. 1969.

[16] N. S. Jayant and L. R. Rabiner, "The application of dither to the quantization of speech signals," *Bell Syst. Tech. J.*, vol. 51, no. 6, pp. 1293–1304, July-Aug. 1972.

[17] R. M. Gray and T. G. Stockham, Jr., "Dithered quantizers," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 805–812, May 1993.

[18] R. A. Wannamaker, S. P. Lipshitz, J. Vanderkooy, and J. N. Wright, "A theory of nonsubtractive dither," *IEEE Trans. Signal Process.*, vol. 48, no. 2, pp. 499–516, Feb. 2000.

[19] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2325–2383, Oct. 1998.

[20] S. Zozor and P.-O. Amblard, "Noise-aided processing: Revisiting dithering in a $\Sigma \Delta$ quantizer," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3202–3210, Aug. 2005.

[21] M. D. McDonnell, N. G. Stocks, C. E. M. Pearce, and D. Abbott, *Stochastic Resonance: From Suprathreshold Stochastic Resonance to Stochastic Signal Quantization*. Cambridge: Cambridge University Press, 2008.

[22] R. Benzi, A. Sutera, and A. Vulpiani, "The mechanism of stochastic resonance," *J. Phys. A: Math. Gen.*, vol. 14, no. 11, pp. L453–L457, Nov. 1981.

[23] L. Gammaitoni, P. Hänggi, P. Jung, and F. Marchesoni, "Stochastic resonance," *Rev. Mod. Phys.*, vol. 70, no. 1, pp. 223–287, Jan. 1998.

[24] V. S. Anishchenko, A. B. Neiman, F. Moss, and L. Schimansky-Geier, "Stochastic resonance: noise-enhanced order," *Physics-Uspekhi*, vol. 42, no. 1, pp. 7–36, 1999.

[25] J. J. Collins, C. C. Chow, and T. T. Imhoff, "Stochastic resonance without tuning," *Nature*, vol. 376, no. 6537, pp. 236–238, Jul. 1995.

[26] K. Lorincz, Z. Gingl, and L. B. Kiss, "A stochastic resonator is able to greatly improve signal-to-noise ratio," *Phys. Lett. A*, vol. 224, no. 1–2, pp. 63–67, Dec. 1996.

[27] S. M. Bezrukov and I. Vodyanoy, "Stochastic resonance in non-dynamical systems without response thresholds," *Nature*, vol. 385, no. 6614, pp. 319–321, Jan. 1997.

[28] F. Chapeau-Blondeau, "Input-output gains for signal in noise in stochastic resonance," *Phys. Lett. A*, vol. 232, no. 1–2, pp. 41–48, Jul. 1997.

[29] F. Chapeau-Blondeau and X. Godivier, "Theory of stochastic resonance in signal transmission by static nonlinear systems," *Phys. Rev. E*, vol. 55, no. 2, pp. 1478–1495, Feb. 1997.

[30] F. Chapeau-Blondeau, "Periodic and aperiodic stochastic resonance with output signal-to-noise ratio exceeding that at the input," *Int. J. Bifurc. Chaos*, vol. 9, no. 1, pp. 267–272, Jan. 1999.

[31] P. Hänggi, M. E. Inchiosa, D. Fogliatti, and A. R. Bulsara, "Nonlinear stochastic resonance: The saga of anomalous output-input gain," *Phys. Rev. E*, vol. 62, no. 5, pp. 6155–6163, Nov. 2000.

[32] F. J. Castro, M. N. Kuperman, M. Fuentes, and H. S. Wio, "Experimental evidence of stochastic resonance without tuning due to non-gaussian noises," *Phys. Rev. E*, vol. 64, no. 5, p. 051105, Oct. 2001.

[33] Z. Gingl, P. Makra, and R. Vajtai, "High signal-to-noise ratio gain by stochastic resonance in a double well," *Fluct. Noise Lett.*, vol. 1, no. 3, pp. L181–L188, Sep. 2001.

[34] G. P. Harmer, B. R. Davis, and D. Abbott, "A review of stochastic resonance: Circuits and measurement," *IEEE Trans. Instrum. Meas.*, vol. 51, no. 2, pp. 299–309, Apr. 2002.

[35] P. Makra, Z. Gingl, and L. B. Kish, "Signal-to-noise ratio gain in non-dynamical and dynamical bistable stochastic resonators," *Fluct. Noise Lett.*, vol. 2, no. 3, pp. L145–L153, Sep. 2002.

[36] P. Makra, Z. Gingl, and T. Fülei, "Signal-to-noise ratio gain in stochastic resonators driven by coloured noises," *Phys. Lett. A*, vol. 317, no. 3–4, pp. 228–232, Oct. 2003.

[37] S. Zozor and P.-O. Amblard, "Stochastic resonance in discrete time nonlinear AR(1) models," *IEEE Trans. Signal Process.*, vol. 47, no. 1, pp. 108–122, Jan. 1999.

[38] ——, "On the use of stochastic resonance in sine detection," *Signal Process.*, vol. 7, no. 3, pp. 353–367, Mar. 2002.

[39] ——, "Stochastic resonance in locally optimal detectors," *IEEE Trans. Signal Process.*, vol. 51, no. 12, pp. 3177–3181, Dec. 2003.

[40] N. G. Stocks, "Suprathreshold stochastic resonance: an exact result for uniformly distributed signal and noise," *Phys. Lett. A*, vol. 279, no. 5-6, pp. 308–312, Feb. 2001.

[41] ——, "Information transmission in parallel threshold arrays: Suprathreshold stochastic resonance," *Phys. Rev. E*, vol. 63, no. 4, p. 041114, Mar. 2001.

[42] N. G. Stocks, D. Allingham, and R. P. Morse, "The application of suprathreshold stochastic resonance to cochlear iimplant coding," *Fluct. Noise Lett.*, vol. 2, no. 3, pp. L169–L181, Sep. 2002.

[43] F. Chapeau-Blondeau and D. Rousseau, "Noise-enhanced performance for an optimal Bayesian estimator," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1327–1334, May 2004.

[44] A. Patel and B. Kosko, "Error-probability noise benefits in threshold neural signal detection," *Neural Netw.*, vol. 22, no. 5-6, pp. 697–706, July-Aug. 2009.

[45] ——, "Optimal mean-square noise benefits in quantizer-array linear estimation," *IEEE Signal Process. Lett.*, vol. 17, no. 12, pp. 1005–1009, Dec. 2010.

[46] ——, "Noise benefits in quantizer-array correlation detection and watermark decoding," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 488–505, Feb. 2011.

[47] M. D. McDonnell, "Is electrical noise useful?" *Proc. IEEE*, vol. 99, no. 2, pp. 242–246, Feb. 2011.

[48] S. Mitaim and B. Kosko, "Adaptive stochastic resonance," *Proc. IEEE*, vol. 86, no. 11, pp. 2152–2183, Nov. 1998.

[49] B. Xu, J. Li, and J. Zheng, "How to tune the system parameters to realize stochastic resonance," *J. Phys. A: Math. Gen.*, vol. 36, no. 48, pp. 11 969–11 980, Dec. 2003.

[50] A. A. Saha and G. V. Anand, "Design of detectors based on stochastic resonance," *Signal Process.*, vol. 83, no. 6, pp. 1193–1212, Jun. 2003.

[51] 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, Mar. 1998.

[52] I. Goychuk and P. Hänggi, "Stochastic resonance in ion channels characterized by information theory," *Phys. Rev. E*, vol. 61, no. 4, pp. 4272–4280, Apr. 2000.

[53] N. G. Stocks, "Suprathreshold stochastic resonance in multilevel threshold systems," *Phys. Rev. Lett.*, vol. 84, no. 11, pp. 2310–2313, Mar. 2000.

[54] B. Kosko and S. Mitaim, "Stochastic resonance in noisy threshold neurons," *Neural Netw.*, vol. 16, no. 5–6, pp. 755–761, June-July 2003.

[55] ——, "Robust stochastic resonance for simple threshold neurons," *Phys. Rev. E*, vol. 70, no. 3, p. 031911, Sep. 2004.

[56] S. Mitaim and B. Kosko, "Adaptive stochastic resonance in noisy neurons based on mutual information," *IEEE Trans. Neural Netw.*, vol. 15, no. 6, pp. 1526–1540, Nov. 2004.

[57] V. Anantharam and V. S. Borkar, "An information theoretic view of stochastic resonance," in *Proc. 2007 IEEE Int. Symp. Inf. Theory*, Jun. 2007, pp. 966–970.

[58] M. D. McDonnell, N. G. Stocks, C. E. M. Pearce, and D. Abbott, "Stochastic resonance and data processing inequality," *IEE Electron. Lett.*, vol. 39, no. 17, pp. 1287–1288, Aug. 2003.

[59] E. Simonotto, M. Riani, C. Seife, M. Roberts, J. Twitty, and F. Moss, "Visual perception of stochastic resonance," *Phys. Rev. Lett.*, vol. 78, no. 6, pp. 1186–1189, Feb. 1997.

[60] J. Tougaard, "Stochastic resonance and signal detection in an energy detector – implications for biological receptor systems," *Biol. Cybern.*, vol. 83, no. 5, pp. 471–480, Oct. 2000.

[61] ——, "Signal detection theory, detectability and stochastic resonance effects," *Biol. Cybern.*, vol. 87, no. 2, pp. 79–90, Aug. 2002.

[62] L. M. Ward, A. Neiman, and F. Moss, "Stochastic resonance in psychophysics and in animal behavior," *Biol. Cybern.*, vol. 87, no. 2, pp. 91–101, Aug. 2002.

[63] Y. Hirano, Y. Segawa, T. Kawai, and T. Matsumoto, "Stochastic resonance in a molecular redox circuit," *J. Phys. Chem. C*, vol. 117, no. 1, pp. 140–145, Jan. 2013.

[64] J. E. Levin and J. P. Miller, "Broadband neural encoding in the cricket cercal sensory system enhanced by stochastic resonance," *Nature*, vol. 380, no. 6570, pp. 165–168, Mar. 1996.

[65] I. Mendez-Balbuena, E. Manjarrez, J. Schulte-Mönting, F. Huethe, J. A. Tapia, M.-C. Hepp-Reymond, and R. Kristeva, "Improved sensorimotor performance via stochastic resonance," *J. Neurosci.*, vol. 32, no. 36, pp. 12 612–12 618, Sep. 2012.

[66] W. Liu, L. A. Lipsitz, M. Montero-Odasso, J. Bean, D. C. Kerrigan, and J. J. Collins, "Noise-enhanced vibrotactile sensitivity in older adults, patients with stroke, and patients with diabetic neuropathy," *Arch. Phys. Med. Rehabil.*, vol. 83, no. 2, pp. 171–176, Feb. 2002.

[67] A. Priplata, J. Niemi, M. Salen, J. Harry, L. A. Lipsitz, and J. J. Collins, "Noise-enhanced human balance control," *Phys. Rev. Lett.*, vol. 89, no. 23, p. 238101, Nov. 2002.

[68] L. E. Medina, M. A. Lebedev, J. E. O'Doherty, and M. A. L. Nicolelis, "Stochastic facilitation of artificial tactile sensation in primates," *J. Neurosci.*, vol. 32, no. 41, pp. 14 271–14 275, Oct. 2012.

[69] R. Srebro and P. Malladi, "Stochastic resonance of the visually evoked potential," *Phys. Rev. E*, vol. 59, no. 3, pp. 2566–2570, Mar. 1999.

[70] T. Mori and S. Kai, "Noise-induced entrainment and stochastic resonance in human brain waves," *Phys. Rev. Lett.*, vol. 88, no. 21, p. 218101, May 2002.

[71] K. Kitajo, K. Yamanaka, L. M. Ward, and Y. Yamamoto, "Stochastic resonance in attention control," *Europhys. Lett.*, vol. 76, no. 6, pp. 1029–1036, Dec. 2006.

[72] A. S. Asdi and A. H. Tewfik, "Detection of weak signals using adaptive stochastic resonance," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP 1995)*, vol. 2, May 1995, pp. 1332–1335.

[73] S. Kay, "Can detectability be improved by adding noise?" *IEEE Signal Process. Lett.*, vol. 7, no. 1, pp. 8–10, Jan. 2000.

[74] J. B. Thomas, "Nonparametric detection," *Proc. IEEE*, vol. 58, no. 5, pp. 623–631, May 1970.

[75] H. Chen, P. K. Varshney, S. M. Kay, and J. H. Michels, "Theory of the stochastic resonance effect in signal detection: Part I–fixed detectors," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3172–3184, Jul. 2007.

[76] H. Chen and P. K. Varshney, "Theory of the stochastic resonance effect in signal detection: Part II–variable detectors," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5031–5041, Oct. 2008.

[77] A. W. Roberts and D. E. Varberg, *Convex Functions*. New York and London: Academic Press, 1973.

[78] A. Patel and B. Kosko, "Optimal noise benefits in Neyman-Pearson and inequality-constrained statistical signal detection," *IEEE Trans. Signal Process.*, vol. 57, no. 5, pp. 1655–1669, May 2009.

[79] I. Lee, X. Liu, C. Zhou, and B. Kosko, "Noise-enhanced detection of subthreshold signals with carbon nanotubes," vol. 5, no. 6, pp. 613–627, Nov. 2006.

[80] A. Patel and B. Kosko, "Stochastic resonance in continuous and spiking neuron models with Levy noise," *IEEE Trans. Neural Netw.*, vol. 19, no. 12, pp. 1993–2008, Dec. 2008.

[81] B. Kosko, I. Lee, S. Mitaim, A. Patel, and M. M. Wilde, "Applications of forbidden interval theorems in stochastic resonance," in *Applications of Nonlinear Dynamics*, V. In, P. Longhini, and A. Palacios, Eds. Springer Berlin Heidelberg, 2009, pp. 71–89.

[82] M. M. Wilde and B. Kosko, "Quantum forbidden-interval theorems for stochastic resonance," *J. Phys. A: Math. Gen.*, vol. 42, no. 46, p. 465309, Nov. 2009.

[83] D. Applebaum, "Extending stochastic resonance for neuron models to general Lévy noise," *IEEE Trans. Neural Netw.*, vol. 20, no. 12, pp. 1993–1995, Dec. 2009.

[84] M. Gagrani, P. Sharma, S. Iyengar, V. S. S. Nadendla, A. Vempaty, H. Chen, and P. K. Varshney, "On noise-enhanced distributed inference in the presence of Byzantines," in *Proc. 49th Annu. Allerton Conf. Commun. Control Comput.*, Sep. 2011, pp. 1222–1229.

[85] A. Vempaty, V. S. S. Nadendla, and P. K. Varshney, "Further results on noise-enhanced distributed inference in the presence of Byzantines," in *Proc. 16th Int. Symp. Wirel. Pers. Multimed. Commun. (WPMC)*, Jun. 2013.

[86] A. Vempaty, L. Tong, and P. K. Varshney, "Distributed inference with Byzantine data: State-of-the-art review on data falsification attacks," *IEEE Signal Process. Mag.*, vol. 30, no. 5, pp. 65–75, Sep. 2013.

[87] H. Chen, P. K. Varshney, and J. H. Michels, "Improving sequential detection performance via stochastic resonance," *IEEE Signal Process. Lett.*, vol. 15, pp. 685–688, 2008.

[88] M. Guerriero, S. Marano, V. Matta, and P. Willett, "Stochastic resonance in sequential detectors," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 2–15, Jan. 2009.

[89] N. Mukhopadhyay, S. Datta, and S. Chattopadhyay, Eds., *Applied Sequential Methodologies: Real-World Examples with Data Analysis*. CRC, 2004.

[90] A. Wald, *Sequential Analysis*. New York: John Wiley and Sons, 1947.

[91] P. K. Varshney, *Distributed Detection and Data Fusion*. Springer, 1997.

[92] V. V. Veeravalli, "Sequential decision fusion: Theory and applications," *J. Franklin Inst.*, vol. 336, no. 2, pp. 301–322, Mar. 1999.

[93] B. Chen, L. Tong, and P. K. Varshney, "Channel-aware distributed detection in wireless sensor networks," *IEEE Signal Process. Mag.*, vol. 23, no. 4, pp. 16–26, Jul. 2006.

[94] S. Tantaratana and J. B. Thomas, "On sequential sign detection of a constant signal," *IEEE Trans. Inf. Theory*, vol. IT-23, no. 11, pp. 304–315, Nov. 1977.

[95] R. Chandramouli and N. Ranganathan, "A generalized sequential sign detector for binary hypothesis testing," *IEEE Signal Process. Lett.*, vol. 5, no. 11, pp. 295–297, Nov. 1998.

[96] S. Kullback, *Information Theory and Statistics*. Mineola, NY: Dover Publications Inc., 1997, reprint of second edition (1968).

[97] H. C. Papadopoulos, G. W. Wornell, and A. V. Oppenheim, "Sequential signal encoding from noisy measurements using quantizers with dynamic bias control," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 978–1002, Mar. 2001.

[98] H. Chen, P. K. Varshney, and J. H. Michels, "Noise enhanced parameter estimation," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5074–5081, Oct. 2008.

[99] M. D. McDonnell and D. Abbott, "What is stochastic resonance? Definitions, misconceptions, debates, and its relevance to biology," *PLoS Comput. Biol.*, vol. 5, no. 5, p. e1000348, May 2009.

[100] K. R. Castleman, *Digital Image Processing*, 2nd ed. Prentice Hall, 1995.

[101] E. P. Simoncelli, "Statistical modeling of photographic images," in *Handbook of Image and Video Processing*, 2nd ed., A. Bovik, Ed. Academic Press, 2005.

[102] M. Piana, M. Canfora, and M. Riani, "Role of noise in image processing by the human perceptive system," *Phys. Rev. E*, vol. 62, no. 1, pp. 1104–1109, Jul. 2000.

[103] T. Aihara, K. Kitajo, D. Nozaki, and Y. Yamamoto, "Internal noise determines external stochastic resonance in visual perception," *Vis. Res.*, vol. 48, no. 14, pp. 1569–1573, Jun. 2008.

[104] M.-O. Hongler, Y. L. de Meneses, A. Beyeler, and J. Jacot, "The resonant retina: Exploiting vibration noise to optimally detect edges in an image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1051–1062, Sep. 2003.

[105] 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, May 2002, pp. 521–523.

[106] Q. Ye, H. Huang, X. He, and C. Zhang, "A SR-based Radon transform to extract weak lines from noise images," in *Proc. IEEE Int. Conf. Image Process. (ICIP 2003)*, vol. 1, Sep. 2003, pp. 849–852.

[107] Q. Ye, H. Huang, and C. Zhang, "Image enhancement using stochastic resonance," in *Proc. IEEE Int. Conf. Image Process. (ICIP 2004)*, vol. 1, Oct. 2004, pp. 263–266.

[108] A. Histace and D. Rousseau, "Constructive action of noise for impulsive noise removal in scalar images," *IEE Electron. Lett.*, vol. 42, no. 7, pp. 393–395, Mar. 2006.

[109] R. Peng, H. Chen, P. K. Varshney, and J. H. Michels, "Stochastic resonance: An approach for enhanced medical image processing," in *Proc. IEEE/NIH Life Science Systems and Applications Workshop (LISA 2007)*, Nov. 2007, pp. 253–256.

[110] V. P. S. Rallabandi, "Enhancement of ultrasound images using stochastic resonance-based wavelet transform," *Comput. Med. Imaging Graph.*, vol. 32, no. 4, pp. 316–320, Jun. 2008.

[111] V. P. S. Rallabandi and P. K. Roy, "Magnetic resonance image enhancement using stochastic resonance in Fourier domain," *Magn. Reson. Imaging*, vol. 28, no. 9, pp. 1361–1373, Nov. 2010.

[112] C. Ryu, S. G. Kong, and H. Kim, "Enhancement of feature extraction for low-quality fingerprint images using stochastic resonance," *Pattern Recognit. Lett.*, vol. 32, no. 2, pp. 107–113, Jan. 2011.

[113] R. K. Jha and R. Chouhan, "Noise-induced contrast enhancement using stochastic resonance on singular values," *Signal Image Video Process.*, vol. 8, no. 2, pp. 339–347, Feb. 2014.

[114] S. Janpaiboon and S. Mitaim, "Adaptive stochastic resonance in color object segmentation," in *Proc. 2006 Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2006, pp. 2508–2515.

[115] O. Krishna, R. K. Jha, A. K. Tiwari, and B. Soni, "Noise induced segmentation of noisy color image," in *Proc. 2013 Nat. Conf. Commun. (NCC)*, Feb. 2013.

[116] S. Sun and B. Lei, "On an aperiodic stochastic resonance signal processor and its application in digital watermarking," *Signal Process.*, vol. 88, no. 8, pp. 2085–2094, Aug. 2008.

[117] R. K. Jha, P. K. Biswas, and S. Shrivastava, "Logo extraction using dynamic stochastic resonance," *Signal Image Video Process.*, vol. 7, no. 1, pp. 119–128, Jan. 2013.

[118] L. Nataraj, A. Sarkar, and B. S. Manjunath, "Adding Gaussian noise to denoise JPEG for detecting image resizing," in *Proc. IEEE Int. Conf. Image Process. (ICIP 2009)*, vol. 1, Nov. 2009, pp. 1493–1496.

[119] T. C. Wang and N. B. Karayiannis, "Detection of microcalcifications in digital mammograms using wavelets," *IEEE Trans. Med. Imag.*, vol. 17, no. 4, pp. 498–509, Aug. 1998.

[120] R. Peng, H. Chen, and P. K. Varshney, "Noise-enhanced detection of micro-calcifications in digital mammograms," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 1, pp. 62–73, Feb. 2009.

[121] G. McLachlan and D. Peel, *Finite Mixture Models*. New York: John Wiley & Sons, 2000.

[122] K. S. Tang, K. F. Man, S. Kwong, and Q. He, "Genetic algorithms and their applications," *IEEE Signal Process. Mag.*, vol. 13, no. 6, pp. 22–37, Nov. 1996.

[123] B. Franzke and B. Kosko, "Noise can speed convergence in Markov chains," *Phys. Rev. E*, vol. 84, no. 4, p. 041112, Oct. 2011.

[124] K. Audhkhasi, O. Osoba, and B. Kosko, "Noise benefits in backpropagation and deep bidirectional pre-training," in *Proc. 2013 Int. Joint Conf. Neural Netw. (IJCNN)*, Aug. 2013.

[125] O. Osoba, S. Mitaim, and j. . F. v. . . n. . . m. . s. y. . . p. . . d. . S. Bart Kosko", title = "The Noisy Expectation-Maximization Algorithm".

[126] O. Osoba and B. Kosko, "Noise-enhanced clustering and competitive learning algorithms," *Neural Netw.*, vol. 37, pp. 132–140, Jan. 2013.

[127] D. A. Spielman and S.-H. Teng, "Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time," *J. ACM*, vol. 51, no. 3, pp. 385–463, May 2004.

[128] M. Yoshida, H. Hayashi, K. Tateno, and S. Ishizuka, "Stochastic resonance in the hippocampal CA3-CA1 model: a possible memory recall mechanism," *Neural Netw.*, vol. 15, no. 10, pp. 1171–1183, Dec. 2002.

[129] A. Karbasi, A. H. Salavati, A. Shokrollahi, and L. R. Varshney, "Noise-enhanced associative memories," in *Advances in Neural Information Processing Systems 26*, C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, Eds. Cambridge, MA: MIT Press, 2013, pp. 1682–1690.

[130] T. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge: Cambridge University Press, 2008.

[131] S. M. Aji and R. J. McEliece, "The generalized distributive law," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 325–343, Mar. 2000.

[132] F. Leduc-Primeau, S. Hemati, S. Mannor, and W. J. Gross, "Dithered belief propagation decoding," *IEEE Trans. Commun.*, vol. 60, no. 8, Aug. 2012.

[133] S. Kay, "Noise enhanced detection as a special case of randomization," *IEEE Signal Process. Lett.*, vol. 15, pp. 709–712, 2008.

**Hao Chen** (S'06M'08) received the Ph.D. degree in electrical engineering from Syracuse University, Syracuse, NY, USA, in 2007. He was a Post-Doctoral Research Associate from 2007 to 2010 and then a Research Assistant Professor with Syracuse University. Since August 2010, he has been an Assistant Professor with the Department of Electrical and Computer Engineering and Computer Science, Boise State University, Boise, ID, USA. His current research interests include statistical signal and image processing, and communications.

**Lav R. Varshney** (S'00M'10) received the B. S. degree with honors in electrical and computer engineering (magna cum laude) from Cornell University, Ithaca, NY, in 2004. He received the S. M., E. E., and Ph. D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, in 2006, 2008, and 2010, respectively. His S. M. thesis earned the E. A. Guillemin Thesis Award and his doctoral thesis earned the J.-A. Kong Award Honorable Mention.

He is an assistant professor in the Department of Electrical and Computer Engineering, a research assistant professor in the Coordinated Science Laboratory, and a research affiliate in the Beckman Institute for Advanced Science and Technology, all at the University of Illinois at Urbana-Champaign. He was a research staff member at the IBM Thomas J. Watson Research Center, Yorktown Heights, NY from 2010 until 2013. He was previously a postdoctoral associate, instructor, research assistant, and National Science Foundation graduate research fellow at MIT from 2004 until 2010, a visitor at the École Polytechnique Fédérale de Lausanne, Switzerland in 2006, and a visiting scientist at Cold Spring Harbor Laboratory, Cold Spring Harbor, New York in 2005. His current research interests include information and coding theory, statistical signal processing, and neuroscience.

Dr. Varshney is a member of Tau Beta Pi, Eta Kappa Nu, and Sigma Xi. He has received several best paper awards.

**Pramod K. Varshney** (S'72–M'77–SM'82–F'97) was born in Allahabad, India, on July 1, 1952. He received the B. S. degree in electrical engineering and computer science (with highest honors), and the M. S. and Ph. D. degrees in electrical engineering from the University of Illinois at Urbana-Champaign in 1972, 1974, and 1976 respectively.

During 1972-76, he held teaching and research assistantships at the University of Illinois. Since 1976 he has been with Syracuse University, Syracuse, NY where he is currently a Distinguished Professor of Electrical Engineering and Computer Science and the Director of CASE: Center for Advanced Systems and Engineering. He served as the Associate Chair of the department during 1993-96. He is also an Adjunct Professor of Radiology at Upstate Medical University in Syracuse, NY. His current research interests are in distributed sensor networks and data fusion, detection and estimation theory, wireless communications, physical layer security and image processing, radar. He has published extensively. He is the author of *Distributed Detection and Data Fusion*, published by Springer-Verlag in 1997. He has served as a consultant to several major companies.

While at the University of Illinois, Dr. Varshney was a James Scholar, a Bronze Tablet Senior, and a Fellow. He is a member of Tau Beta Pi and is the recipient of the 1981 ASEE Dow Outstanding Young Faculty Award. He was elected to the grade of Fellow of the IEEE in 1997 for his contributions in the area of distributed detection and data fusion. He was the guest editor of the special issue on data fusion of the *Proceedings of the IEEE*, January 1997. In 2000, he received the Third Millennium Medal from the IEEE and Chancellor's Citation for exceptional academic achievement at Syracuse University. He is the recipient of the IEEE 2012 Judith A. Resnik Award. He received an honorary Doctor of Engineering degree from Drexel University in 2014. He is on the editorial boards of the *Journal on Advances in Information Fusion* and *IEEE Signal Processing Magazine*. He was President of the International Society of Information Fusion during 2001.