STOCHASTIC RESONANCE IN CONTINUOUS AND SPIKING NEURON MODELS WITH LEVY NOISE

Ashok Patel and Bart Kosko

Signal and Image Processing Institute
Department of Electrical Engineering
University of Southern California
Los Angeles, CA 90089-2564, USA

Keywords: Levy noise, Signal detection, Stochastic resonance, Neuron models, Mutual information

ABSTRACT

Levy noise can help neurons detect faint or subthreshold signals. Levy noise extends standard Brownian noise to many types of impulsive jump-noise processes found in real and model neurons as well as in models of finance and other random phenomena. Two new theorems and the Itô calculus show that white Levy noise will benefit subthreshold neuronal signal detection if the noise process’s scaled drift velocity falls inside an interval that depends on the threshold values. These results generalize earlier ‘forbidden interval’ theorems of neuronal ‘stochastic resonance’ or noise-injection benefits. Global and local Lipschitz conditions imply that additive white Levy noise can increase the mutual information or bit count of several feedback neuron models that obey a general stochastic differential equation. Simulation results show that the same noise benefits still occur for some infinite-variance stable Levy noise processes even though the theorems themselves apply only to finite-variance Levy noise. The Appendix proves the two Itô-theoretic lemmas that underlie the new Levy noise-benefit theorems.
I. STOCHASTIC RESONANCE IN NEURAL SIGNAL DETECTION

Stochastic resonance (SR) occurs when noise benefits a system rather than harms it. Small amounts of noise can often enhance some forms of nonlinear signal processing while too much noise degrades it [12, 13, 22, 27, 44, 48, 57, 59, 60, 68, 70, 71, 82]. SR has many useful applications in physics, biology, and medicine [5, 6, 7, 11, 14, 17, 18, 21, 23, 32, 40, 41, 42, 51, 52, 54, 55, 61, 69, 74, 81, 83, 87, 89]. SR in neural networks is itself part of the important and growing area of stochastic neural networks [9, 10, 38, 84, 85, 86, 88]. We show that a wide range of general feedback continuous neurons and spiking neurons benefit from a broad class of additive white Levy noise. This appears to be the first demonstration of the SR effect for neuron models subject to Levy noise perturbations.

Figure 1: Stochastic resonance in the Kanisza square illusion with symmetric \( \alpha \)-stable noise (\( \alpha = 1.9 \)) in a thick-tailed bell curve with infinite-variance but finite intensity or dispersion \( \gamma \) [30]. The Kanisza square illusion improves as the noise dispersion \( \gamma \) increases from 0.047 to 0.3789 and then it degrades as the dispersion increases further. Each pixel represents the output of the noisy bistable potential neuron model (1)-(2) and (5) that uses the pixel values of the original Kanisza square image as subthreshold input signals. The additive \( \alpha \)-stable noise dispersions are (a) \( \gamma = 0.047 \), (b) \( \gamma = 0.1015 \), (c) \( \gamma = 0.3789 \), (d) \( \gamma = 1 \), and (e) \( \gamma = 3.7321 \).

Figure 1 shows how impulsive Levy noise can enhance the Kanisza-square visual illusion in which four dark-corner figures give rise to an illusory bright interior square. Each pixel is the thresholded output of a noisy bistable neuron whose input signals are subthreshold and quantized pixel values of the original noise-free Kanisza image. The outputs of the bistable neurons do not depend on the input signals if there is no
additive noise because the input signals are subthreshold. Figure 1(a) shows that adding infinite-variance
Levy noise induces a slight correlation between the pixel input and output signals. More intense Levy
noise increases this correlation in Figures 1(b) - (c). Still more intense Levy noise degrades the image
and undermines the visual illusion in Figures 1(d) - (e). Figure 2 shows typical sample paths from different
types of Levy noise. Figure 3 shows the characteristic inverted-U or non-monotonic signature of SR for
white Levy noise that perturbs a continuous bistable neuron.

Figure 2: Sample paths from one-dimensional Levy processes: (a) Brownian motion with drift \( \mu = 0.1 \) and variance
\( \sigma = 0.15 \), (b) jump diffusion with \( \mu = 0.1, \sigma = 0.225 \), Poisson jump rate \( \lambda = 3 \), and uniformly distributed jump
magnitudes in the interval \([-0.2, 0.2]\) (and so with Levy measure \( \nu(dy) = (3/0.4)dy \) for \( y \in [-0.2, 0.2] \) and zero else),
(c) normal inverse Gaussian (NIG) process with parameters \( \alpha = 20, \beta = 0, \delta = 0.1, \) and \( \mu = 0 \), (d) infinite-variance
\( \alpha \)-stable process with \( \alpha = 1.9 \) and dispersion \( \kappa = 0.0272 \) (\( \mu = 0, \sigma = 0, \) and \( \nu(dy) \) is of the form \( \frac{k}{|y|^{1+\alpha}} dy \)).

We generalize the recent ‘forbidden interval’ theorems [49, 50, 60, 64, 65] for continuous and spiking
neuron models to a broad class of finite-second-moment Levy noise that may depend on the neuron’s mem-
Figure 3: Mutual information Levy noise benefits in the continuous bistable neuron (1)-(2) and (5). Additive white Levy noise $dL_t$ increases the mutual information of the bistable potential neuron for the subthreshold input signals $s_1 = -0.3$ and $s_2 = 0.4$. The types of Levy noise $dL_t$ are (a) Gaussian with uniformly distributed jumps, (b) pure-jump normal inverse Gaussian (NIG), and (c) symmetric $\alpha$-stable noise with $\alpha = 1.9$ (thick-tailed bell curve with infinite variance [62]). The dashed vertical lines show the total min-max deviations of the mutual information in 100 simulation trials.

brane potential. The original forbidden interval theorem [49, 50] states that simple threshold neurons will have an SR noise benefit in the sense that noise increases the neuron’s mutual information or bit count if and only if the noise mean or location parameter $\mu$ does not fall in a threshold-related interval: SR occurs if and only if $\mu \notin (T - A, T + A)$ for threshold $T$ where $-A < A < T$ for bipolar subthreshold signal $\pm A$.

The theorems below show that such an SR noise benefit will occur if the additive white Levy noise process has a bounded scaled drift velocity that does not fall within a threshold-based interval. This holds for general feedback continuous neuron models that include common signal functions such as logistic sigmoids or Gaussians. It also holds for spiking neurons such as the FitzHugh-Nagumo, leaky integrate-and-fire, and reduced Type-I neuron models. We used the Itô stochastic calculus to prove our results under the assumption that the Levy noise has a finite second moment. But Figure 1 and the (c) sub-figures of Figures 3-8 all show that the SR noise benefit still occurs in the more general infinite-variance case of some types of $\alpha$-stable Levy noise. So the SR effect is not limited to finite-second-moment Levy noise. We were not able to prove that these stable infinite-variance SR effects must occur as we did prove with simpler neuron
Levy noise has advantages over standard Gaussian noise in neuron models despite its increased mathematical complexity. A Levy noise model more accurately describes how the neuron’s membrane potential evolves than does a simpler diffusion model because the more general Levy model includes not only pure-diffusion and pure-jump models but jump-diffusion models as well [35, 73]. Neuron models with additive Gaussian noise are pure-diffusion models. These neuron models rely on the classical central limit theorem for their Gaussian structure and thus they rely on special limiting-case assumptions of incoming Poisson spikes from other neurons. These assumptions require at least that the number of impinging synapses is large and that the synapses have small membrane effects due to the small coupling coefficient or the synaptic weights [28, 46]. The Gaussian noise assumption may be more appropriate for signal inputs from dendritic trees because of the sheer number of dendrites. But often fewer inputs come from synapses near the post-synaptic neuron’s trigger zone and these inputs produce impulses in noise amplitudes because of the higher concentration of voltage-sensitive sodium channels in the trigger zone [29, 45, 63]. Engineering applications also favor the more general Levy model because physical devices may be limited in their number of model-neuron connections [58] and because real signals and noise can often be impulsive [30, 62, 79].

II. NOISY FEEDBACK NEURON MODELS

We study Levy SR noise benefits in the noisy feedback neuron models of the general form

\[
\dot{x} = -x(t) + f(x(t)) + s(t) + n(t) \\
y(t) = g(x(t))
\]

with initial condition \(x(t_0) = x_0\). Here \(s(t)\) is the additive net excitatory or inhibitory input forcing signal—either \(s_1\) or \(s_2\). The additive noise term \(n(t)\) is Levy noise with mean or location \(\mu\) and intensity scale \(\kappa\) (or dispersion \(\gamma\) for symmetric \(\alpha\)-stable noise where \(\gamma = \kappa^\alpha\) with characteristic function \(\phi(u) = e^{i\mu u - \gamma|u|^\alpha}\)). The neuron feeds its activation or membrane potential signal \(x(t)\) back to itself through \(-x(t) + f(x(t))\) and emits the (observable) thresholded or spike signal \(y(t)\) as output. Here \(g\) is a static transformation...
function. We use the threshold

\[
g(x) = \begin{cases} 
1 & \text{if } x > 0 \\
0 & \text{else} 
\end{cases}
\]  

(3)

for continuous neuron models. We use a related threshold \( g \) in spiking neuron models where \( g \) determines the spike occurrence. The neuronal signal function \( f(x) \) of (1) can be of quite general form for continuous neuron models [65]:

- **Logistic.** The logistic signal function [47] is sigmoidal and strictly increasing

\[
f(x) = \frac{1}{1 + e^{-cx}}
\]  

(4)

for scaling constant \( c > 0 \). We use \( c = 8 \). This signal function gives a bistable additive neuron model.

- **Hyperbolic Tangent.** This signal function is also sigmoidal and gives a bistable additive neuron model [2, 15, 37, 47]:

\[
f(x) = 2 \tanh x
\]  

(5)

- **Linear Threshold.** This linear-threshold signal has the form [47]:

\[
f(x) = \begin{cases} 
\text{cx} & |\text{cx}| < 1 \\
1 & \text{cx} > 1 \\
-1 & \text{cx} < -1 
\end{cases}
\]  

(6)

for constant \( c > 0 \). We use \( c = 2 \).

- **Exponential.** This signal function is asymmetric and has the form [47]

\[
f(x) = \begin{cases} 
1 - \exp\{-cx\} & \text{if } x > 0 \\
0 & \text{else} 
\end{cases}
\]  

(7)

for constant \( c > 0 \). We use \( c = 8 \). 

6
• Gaussian. The Gaussian or ‘radial basis’ signal function [47] differs in form from the signal functions above because it is nonmonotonic:

\[ f(x) = \exp\{-cx^2\} \quad (8) \]

for constant \( c > 0 \). We use \( c = 20 \).

The above neuron models can have up to three fixed points depending on the input signal and the model parameters. The input signal is subthreshold in the sense that switching it from \( s_1 \) to \( s_2 \) or vice versa does not change the output \( Y_t \) of (22). There exist \( \theta_1 \) and \( \theta_2 \) such that the input \( S \) is subthreshold when \( \theta_1 \leq s_1 < s_2 \leq \theta_2 \). The values of \( \theta_1 \) and \( \theta_2 \) depend on the model parameters. Consider the linear threshold neuron model (1)-(2) and (6) with \( c = 2 \). A simple calculation shows that if the input signal \( S_t \in \{s_1, s_2\} \) satisfies \(-0.5 < s_1 < s_2 < 0.5\) then the linear threshold neuron has two stable fixed points (one positive and the other negative) and has one unstable fixed point between them. The Gaussian neuron model (1)-(2) and (8) has only one fixed point if \( 0 < s_1 < s_2 \). So the input is subthreshold because switching it from \( s_1 \) to \( s_2 \) or vice versa does not change the output \( Y_t \). Figure 3 shows the mutual information noise benefits in the bistable neuron model (1)-(2) and (5) for three different additive white Levy noise cases when the input signals are subthreshold. Note the signature nonmonotonic shape of all three SR noise-benefit curves in Figure 3.

The membrane potential dynamics (1) are one-dimensional for all our neuron models except for the two-dimensional FHN spiking neuron model below. So next we briefly describe multi-dimensional Levy processes and set up a general multi-dimensional Levy stochastic differential equation framework for our feedback continuous and spiking neuron models.

III. LEVY PROCESSES AND STOCHASTIC DIFFERENTIAL EQUATIONS

Levy processes [67, 75] form a wide class of random processes that include Brownian motion, \( \alpha \)-stable processes, compound Poisson processes, generalized inverse Gaussian processes, and generalized hyperbolic processes. Figure 2 shows some typical scalar Levy sample paths. Levy processes can account for the im-
pulsiveness or discreteness of both signals and noise. Researchers have used Levy processes to model diverse phenomena in economics [4, 78], physics [77], electrical engineering [1, 4, 62, 66], biology [76], and seismology [80]. A Levy process $L_t = (L^1_t, ..., L^m_t)'$ for $t \geq 0$ in a given probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \leq t \leq \infty}, P)$ is a stochastic process taking values in $\mathbb{R}^m$ with stationary and independent increments (we assume that $L_0 = 0$ with probability 1). The Levy process $L_t$ obeys three properties:

1. $L_t - L_s$ is independent of sigma-algebra $\mathcal{F}_s$ for $0 \leq s < t \leq \infty$

2. $L_t - L_s$ has the same distribution as $L_{t-s}$

3. $L_s \to L_t$ in probability if $s \to t$.

The Levy-Khintchine formula gives the characteristic function $\phi$ of $L_t$ as [3]

$$
\phi(u) = E(e^{i\langle u, L_t \rangle}) = e^{\eta(u)} \text{ for } t \geq 0 \text{ and } u \in \mathbb{R}^m
$$

where $\langle \cdot, \cdot \rangle$ is the Euclidean inner product (so $|u| = \langle u, u \rangle^{1/2}$). The characteristic exponent or the so-called Levy exponent is

$$
\eta(u) = i\langle \mu, u \rangle - \frac{1}{2} \langle u, Ku \rangle + \int_{\mathbb{R}^m - \{0\}} \left[ e^{i\langle u, y \rangle} - 1 - i\langle u, y \rangle \chi_{|y|<1}(y) \right] \nu(dy)
$$

for some $\mu \in \mathbb{R}^m$, a positive-definite symmetric $m \times m$ matrix $K$, and measure $\nu$ on Borel subsets of $\mathbb{R}^m_0 = \mathbb{R}^m \setminus \{0\}$ (or $\nu(\{0\}) = 0$). Then $\nu$ is a Levy measure such that

$$
\int_{\mathbb{R}^m_0} \min\{1, |y|^2\} \nu(dy) < \infty.
$$

A Levy process $L_t$ combines a drift component, a Brownian motion (Gaussian) component, and a jump component. The Levy-Khintchine triplet $(\mu, K, \nu)$ completely determines these components. The Levy measure $\nu$ determines both the average number of jumps per unit time and the distribution of jump magnitudes in the jump component of $L_t$. Jumps of any size in a Borel set $B$ form a compound Poisson process with rate $\int_B \nu(dy)$ and jump density $\nu(dy)/\int_B \nu(dy)$ if the closure $\bar{B}$ does not contain 0. $\mu$ gives the velocity of the drift component. $K$ is the covariance matrix of the Gaussian component. If $K = 0$ and $\nu = 0$ then (9) becomes $E(e^{i\langle u, L_t \rangle}) = e^{it\langle \mu, u \rangle}$. Then $L_t = \mu t$ is a simple $m$-dimensional deterministic
motion (drift) with velocity vector $\mu$. If $K \neq 0$ and $\nu = 0$ then $L_t$ is a $m$-dimensional Brownian motion with drift because (9) takes the form $E(e^{i(u,L_t)}) = e^{t[i(\mu,u)-\frac{1}{2}(u,Ku)]}$ and because this exponential is the characteristic function of a Gaussian random vector with mean vector $\mu t$ and covariance matrix $tK$. If $K \neq 0$ and $\nu(R^m) < \infty$ then $L_t$ is a jump-diffusion process while $K = 0$ and $\nu(R^m) = \infty$ give a compound Poisson process. If $K = 0$ and $\nu(R^m) < \infty$ then $L_t$ is a purely discontinuous jump process and has an infinite number of small jumps in any time interval of positive length. We consider only the Levy processes whose components $L^k_t$ have finite second moments: $E[(L^k_t)^2] < \infty$. This excludes the important family of infinite-variance $\alpha$-stable processes (including the $\alpha = 0.5$ Levy stable case) where $\alpha \in (0,2]$ measures the tail thickness and where symmetric $\alpha$-stable distributions have characteristic functions $\phi(u) = e^{iu\mu-\gamma|u|^\alpha}$ for each $\gamma > 0$. Normal inverse Gaussian NIG($\alpha, \beta, \delta, \mu$) distributions are examples of semi-thick-tailed pure-jump Levy processes that have infinite Levy measure and yet have finite moments of all order [33, 72]. They have characteristic functions of the form $\phi(u) = e^{iu\mu+\delta(\sqrt{\alpha^2-\beta^2}-\sqrt{\alpha^2-(\beta+iu)^2})}$ where $0 \leq |\beta| < \alpha$ and $\delta > 0$.

Let $L_t(\mu, K, \nu) = (L^1_t, ..., L^m_t)$ be a Levy process that takes values in $R^m$ where $L^j_t(\mu^j, \sigma^j, \nu^j)$ are real-valued independent Levy processes for $j = 1, ..., m$. We denote the Levy-Itô decomposition [3] of $L^j_t$ for each $j = 1, ..., m$ and $t \geq 0$ as

$$L^j_t = \mu^j t + \sigma^j B^j_t + \int_{|y|<1} \int_0^t y^j N^j(t, dy^j) + \int_{|y| \geq 1} y^j N^j(t, dy^j)$$

$$= \mu^j t + \sigma^j B^j_t + \int_0^t \int_{|y|<1} y^j N^j(ds, dy^j) + \int_0^t \int_{|y| \geq 1} y^j N^j(ds, dy^j).$$

Here $\mu$ determines the velocity of the deterministic drift process $\mu^j t$ while the $B^j_t$ are real-valued independent standard Brownian motions. Then $\mu = (\mu^1, ..., \mu^m)'$ and $K = \text{diag}[(\sigma^1)^2, ..., (\sigma^m)^2]$. The $N^j$ are independent Poisson random measures on $R^+ \times R_0$ with compensated (mean-subtracted) Poisson processes $\tilde{N}^j$ and intensity/Levy measures $\nu^j$.

Define the Poisson random measure as

$$N^j(t, B) = \# \{ \Delta L_s \in B \text{ for } 0 \leq s \leq t \}$$

for each Borel set $B$ in $R_0$. The Poisson random measure gives the random number of jumps of $L_t$ in the
time interval \([0, t]\) with jump size \(\Delta L_t\) in the set \(B\). \(N^j(t, B)\) is a Poisson random variable with intensity \(\nu^j(B)\) if \(\nu^j(B) < \infty\) and if we fix \(t\) and \(B\). But \(N^j(t, \cdot)\) is a measure if we fix \(\omega \in \Omega\) and \(t \geq 0\). This measure is not a martingale. But the compensated Poisson random measure

\[
\tilde{N}^j(t, B) = N^j(t, B) - t \nu^j(B)
\]

is a martingale and gives the compensated Poisson integral (12) (the second term on the right-hand side of (12)) as

\[
\int_{|y^j| < 1} y^j \tilde{N}^j(t, dy^j) = \int_{|y^j| < 1} y^j N^j(t, dy^j) - t \int_{|y^j| < 1} y^j \nu^j(dy^j) \quad \text{for} \quad j = 1, \ldots, m. \tag{16}
\]

We assume again that each \(L^j_t\) has a finite second moment \((E|L^j_t|^2 < \infty)\). But if \(L^j_t\) is a Levy process with triplet \((\mu^j, \sigma^j, \nu^j)\) then \(L^j_t\) has a finite \(p\)th moment for \(p \in \mathbb{R}^+\) if and only if \(\int_{|y^j| > 1} |y^j|^p \nu^j(dy^j) < \infty\) [75]. The drift velocity \(\mu^j\) relates to the expected value of a Levy Process \(L^j_t\) by

\[
E(L^j_1) = \mu^j + \int_{|y^j| > 1} y^j \nu^j(dy^j) \quad \text{and} \quad E(L^j_t) = t E(L^j_1) \quad \text{for} \quad j = 1, \ldots, m.
\]

The variance of the Levy process in (12) is

\[
\Var(L^j_t) = \Var(\sigma^j B^j_t) + \Var(\int_{|y^j| > 1} y^j \tilde{N}^j(t, dy^j)) + \Var(\int_{|y^j| < 1} y^j N^j(t, dy^j)) \tag{17}
\]

because the underlying processes are independent. The variance terms on the right-hand side of (17) have the following form [3]:

\[
\Var(L^j_t) = t(\sigma^j)^2 \tag{18}
\]

\[
\Var\left(\int_{|y^j| > 1} y^j \tilde{N}^j(t, dy^j)\right) = t \int_{|y^j| > 1} |y^j|^2 \nu^j(dy^j) \tag{19}
\]

\[
\Var\left(\int_{|y^j| < 1} y^j \tilde{N}^j(t, dy^j)\right) \leq E\left(\int_{|y^j| < 1} y^j \tilde{N}^j(t, dy^j)\right)^2 = t \int_{|y^j| < 1} |y^j|^2 \nu^j(dy^j). \tag{20}
\]

The last equality follows from the Itô isometry identity (Proposition 8.8 in [19]). Then (17) and (18)-(20) imply that the \(\Var(L^j_t) \to 0\) if and only if \(\sigma^j \to 0\) and \(\nu^j \to 0\).

We can rewrite (1)-(2) as a more general Itô stochastic differential equation (SDE) [3]

\[
dX_t = b(X_{t-})dt + c(X_{t-})dL_t \tag{21}
\]

\[
Y_t = g(X_t) \tag{22}
\]
with initial condition $X_0$. Here $b(X_{t^-}) = -X_{t^-} + f(X_{t^-}) + S_t$ is a Lipschitz continuous drift term, $c(X_{t^-})$ is a bounded Levy diffusion term, and $dL_t$ is a white Levy noise with noise scale $\kappa$.

Our continuous neuron models are again one-dimensional but the spiking FHN neuron model is two-dimensional. So consider the general $d$-dimensional SDE in the matrix form with $m$-dimensional Levy noise $L_t = (L_t^1, ..., L_t^m)'$

$$dX_t = b(X_{t^-})dt + c(X_{t^-})dL_t \quad (23)$$

which is shorthand for the system of SDEs

$$dX_t^i = b^i(X_{t^-})dt + \sum_{j=1}^m c^j_i(X_{t^-})dL_t^j \quad \text{for } i = 1, ..., d \quad (24)$$

with initial condition $X_0^i$. Here $X_t = (X_t^1, ..., X_t^d)'$, $b(X_t) = (b^1(X_t), ..., b^d(X_t))'$, and $c$ is a $d \times m$ matrix with rows $c^i(X_t) = (c^1_i(X_t), ..., c^m_i(X_t))$. The functions $b^i: \mathbb{R}^d \to \mathbb{R}$ are locally or globally Lipschitz measurable functions. The functions $c^j_i: \mathbb{R}^d \to \mathbb{R}$ are bounded globally Lipschitz measurable functions such that $|c^j_i|^2 \leq H_j^2 \in \mathbb{R}^+$. The $L_t^j$ terms are independent Levy processes as in (13) with $\mu_j^i = 0$ for $j = 1, ..., m$. Then

$$dX_t^i = b^i(X_{t^-})dt + \sum_{j=1}^m [c^j_i(X_{t^-})\mu_j^i]dt + \sum_{j=1}^m [c^j_i(X_{t^-})\sigma_j^i]dB_t^j + \sum_{j=1}^m \int_{|y_j^i| < 1} [c^j_i(X_{t^-})y_j^i]L^j(dt, dy_j^i) + \sum_{j=1}^m \int_{|y_j^i| \geq 1} [c^j_i(X_{t^-})y_j^i]N^j(dt, dy_j^i)

= b^i(X_{t^-})dt + \sum_{j=1}^m [\mu_j^i(X_{t^-})]dt + \sum_{j=1}^m [\sigma_j^i(X_{t^-})]dB_t^j + \sum_{j=1}^m \int_{|y_j^i| < 1} F_j^i(X_{t^-}, y_j^i)L^j(dt, dy_j^i) + \sum_{j=1}^m \int_{|y_j^i| \geq 1} G_j^i(X_{t^-}, y_j^i)N^j(dt, dy_j^i). \quad (25)$$

where $\mu_j^i(X_{t^-}) = c^j_i(X_{t^-})\mu^i = 0$, $\sigma_j^i(X_{t^-}) = c^j_i(X_{t^-})\sigma^i$, $F_j^i(X_{t^-}, y^i) = c^j_i(X_{t^-})y^i$, and $G_j^i(X_{t^-}, y^i) = c^j_i(X_{t^-})y^i$ are all globally Lipschitz functions. This equation has the integral form with initial condition
IV. LEVY NOISE BENEFITS IN CONTINUOUS NEURON MODELS

We now prove that Levy noise can benefit the noisy continuous neurons (21)-(22) with signal functions (4)-(8) and subthreshold input signals. We assume that the neuron receives a constant subthreshold input signal \( S_t \in \{s_1, s_2\} \) for time \( T \). Let \( S \) denote the input signal and let \( Y \) denote the output signal \( Y_t \) for a sufficiently large randomly chosen time \( t \leq T \).

Noise researchers have used various system performance measures to detect SR noise benefits \([8, 17, 44, 51, 59, 60, 64, 71]\). These include the output signal-to-noise ratio, cross-correlation, error probability, and Shannon mutual information between input and output signals. We use Shannon mutual information to measure the Levy noise benefits. Mutual information measures the information that the neuron’s output conveys about the input signal. It is a common detection performance measure when the input signal is random \([8, 39, 60, 82]\).

Define the Shannon mutual information \( I(S,Y) \) of the discrete input random variable \( S \) and the output random variable \( Y \) as the difference between the output’s unconditional and conditional entropy \([20]\):

\[
I(S,Y) = H(Y) - H(Y|S) 
\]

\[
= -\sum_y P_Y(y) \log P_Y(y) 
+ \sum_s \sum_y P_{SY}(s,y) \log P_{Y|S}(y|s) 
= -\sum_y P_Y(y) \log P_Y(y) 
+ \sum_s P_S(s) \sum_y P_{Y|S}(y|s) \log P_{Y|S}(y|s) 
= \sum_{s,y} P_{SY}(s,y) \log \frac{P_{SY}(s,y)}{P_S(s)P_Y(y)}. 
\]
So the mutual information is the expectation of the random variable \( \log \frac{P_{SY}(s,y)}{P_{S}(s)P_{Y}(y)} \):

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

(31)

Here \( P_{S}(s) \) is the probability density of the input \( S \), \( P_{Y}(y) \) is the probability density of the output \( Y \), \( P_{Y|S}(y|s) \) is the conditional density of the output \( Y \) given the input \( S \), and \( P_{SY}(s,y) \) is the joint density of the input \( S \) and the output \( Y \). An SR noise benefit occurs in a system if and only if an increase in the input noise variance or dispersion increases the system’s mutual information (31).

We need the following lemma to prove that noise improves the continuous neuron’s mutual information or bit count. The Appendix gives the proof of Lemma 1.

**Lemma 1:** Let \( b^{i} : \mathbb{R}^{d} \rightarrow \mathbb{R} \) and \( c_{j}^{i} : \mathbb{R}^{d} \rightarrow \mathbb{R} \) in (23)-(24) be measurable functions that satisfy the global Lipschitz conditions

\[
||b^{i}(x_{1}) - b^{i}(x_{2})|| \leq K_{1}||x_{1} - x_{2}||
\]  

(32)

\[
||c_{j}^{i}(x_{1}) - c_{j}^{i}(x_{2})|| \leq K_{2}||x_{1} - x_{2}||
\]  

(33)

and

\[
|c_{j}^{i}|^{2} \leq H_{j}^{i} \quad \text{for all } x_{1}, x_{2} \in \mathbb{R}^{d} \text{ and for } i = 1, \ldots, d \text{ and } j = 1, \ldots, m.
\]  

(34)

Suppose

\[ dX_{t} = b(X_{t-})dt + c(X_{t-})dL_{t} \]  

(35)

\[ d\hat{X}_{t} = b(\hat{X}_{t})dt \]  

(36)

where \( dL_{t} \) is a Levy noise with \( \mu = 0 \) and finite second moments. Then for every \( T \in \mathbb{R}^{+} \) and for every \( \varepsilon > 0 \):

\[
E \left[ \sup_{0 \leq t \leq T} ||X_{t} - \hat{X}_{t}||^{2} > \varepsilon \right] \rightarrow 0 \quad \text{as } \sigma^{j} \rightarrow 0 \text{ and } \nu^{j} \rightarrow 0 \text{ for all } j = 1, \ldots, m,
\]  

(37)

and hence

\[
P \left[ \sup_{0 \leq t \leq T} ||X_{t} - \hat{X}_{t}||^{2} > \varepsilon \right] \rightarrow 0 \quad \text{as } \sigma^{j} \rightarrow 0 \text{ and } \nu^{j} \rightarrow 0 \text{ for all } j = 1, \ldots, m
\]  

(38)

since mean-square convergence implies convergence in probability.

We prove the Levy SR theorem with the stochastic calculus and a special limiting argument. This avoids trying to solve for the process \( X_{t} \) in (21). The proof strategy follows that of the ‘forbidden interval’
Theorems [49, 50, 64]: what goes down must go up. Jensen’s inequality implies that $I(S,Y) \geq 0$ [20]. Random variables $S$ and $Y$ are statistically independent if and only if $I(S,Y) = 0$. Hence $I(S,Y) > 0$ implies some degree of statistical dependence. So the system must exhibit the SR noise benefit if $I(S,Y) > 0$ and if $I(S,Y) \to 0$ when noise parameters $\sigma \to 0$ and $\nu \to 0$. Theorem 1 uses Lemma 1 to show that $I(S,Y) \to 0$ when noise parameters $\sigma \to 0$ and $\nu \to 0$. So some increase in the noise parameters must increase the mutual information.

**Theorem 1:** Suppose that the continuous neuron models (21)-(22) and (4)-(8) have a bounded globally Lipschitz Levy diffusion term $c(X_t^-) \leq H$ and that the additive Levy noise has drift velocity $\mu$. Suppose also that the input signal $S(t) \in \{s_1, s_2\}$ is subthreshold: $\theta_1 \leq s_1 < s_2 \leq \theta_2$ and that there is some statistical dependence between the input random variable $S$ and the output spike-rate random variable $R$ so that $I(S,R) > 0$. Then the neuron models (21)-(22) with signal functions including (4)-(8) exhibit the nonmonotone SR effect in the sense that $I(S,Y) \to 0$ as the Levy noise parameters $\sigma \to 0$ and $\nu \to 0$ if $\theta_1 - s_1 \leq H\mu \leq \theta_2 - s_2$.

**Proof:** Let $\{\sigma_k, \nu_k\}_{k=1}^\infty$ be any decreasing sequence of Levy noise parameters such that $\sigma_k \to 0$ and $\nu_k \to 0$ as $k \to \infty$. Define $X(t)_k$ and $Y(t)_k$ as solution processes of the continuous neuron models with Levy noise parameters $\sigma_k$ and $\nu_k$ instead of $\sigma$ and $\nu$.

Suppose that $\mu \neq 0$. We can absorb the drift $c(X_t^-)\mu$ into the input signal $S$ because the Levy noise $L_t$ is additive in the neuron models. Then the new input signal $S' = S + c(X_t^-)\mu$ and it does not affect the Lipschitz continuity of $b(X_t^-)$ in equation (21). Note that $S'$ is subthreshold ($\theta_1 \leq S' \leq \theta_2$) if $\theta_1 - s_1 \leq H\mu \leq \theta_2 - s_2$. So we lose no generality if we consider the noise $dL_t$ with $\mu = 0$ and let $S \in \{s_1, s_2\}$ be subthreshold in the continuous neuron models (21). This allows us to use Lemma 1.

Let the symbol ‘0’ denote the input signal $S = s_1$ and the output signal $Y = 0$. Let the symbol ‘1’ denote the input signal $S = s_2$ and the output signal $Y = 1$. Assume that $0 < P_S(s) < 1$ to avoid triviality when $P_S(s) = 0$ or 1. We show that $S$ and $Y$ are asymptotically independent by using the fact that $I(S,Y) = 0$ if and only if $S$ and $Y$ are statistically independent [20]. So we need to show only that
$P_{SY}(s,y) = P_{S}(s)P_{Y}(y)$ or $P_{Y|S}(y|s) = P_{Y}(y)$ as $\sigma_k \to 0$ and $\nu_k \to 0$ as $k \to \infty$ for signal symbols $s \in S$ and $y \in Y$. The theorem of total probability and the two-symbol alphabet set $S$ give

$$P_{Y}(y) = \sum_{s} P_{Y|S}(y|s)P_{S}(s)$$

$$= P_{Y|S}(y|0)P_{S}(0) + P_{Y|S}(y|1)P_{S}(1)$$

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

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

So we need to show only that $P_{Y_k|S}(y|0) - P_{Y_k|S}(y|1) = 0$ as $\sigma_k \to 0$ and $\nu_k \to 0$ for $y \in \{0, 1\}$. We prove the case for $y = 0$ only: $\lim_{k \to \infty} \{P_{Y_k|S}(0|0) - P_{Y_k|S}(0|1)\} = 0$ since the proof for $y = 1$ is similar. Then the desired limit goes to zero because

$$\lim_{k \to \infty} \{P_{Y_k|S}(0|0) - P_{Y_k|S}(0|1)\} = \lim_{k \to \infty} P_{Y_k|S}(0|0) - \lim_{k \to \infty} P_{Y_k|S}(0|1)$$

$$= \lim_{k \to \infty} P[Y_k = 0|S = 0] - \lim_{k \to \infty} P[Y_k = 0|S = 1]$$

$$= \lim_{k \to \infty} P[X(t)_k < 0|S = 0] - \lim_{k \to \infty} P[X(t)_k < 0|S = 1] \text{ for large } t$$

$$= \lim_{k \to \infty} \left\{P[X(t)_k < 0, \hat{X}(t) < 0|S = 0] + P[X(t)_k < 0, \hat{X}_t > 0|S = 0]\right\}$$

$$- \lim_{k \to \infty} \left\{P[X(t)_k > 0, \hat{X}(t) < 0|S = 1] + P[X(t)_k > 0, \hat{X}_t > 0|S = 1]\right\} \text{ for large } t$$

$$= \lim_{k \to \infty} \left\{P[X(t)_k < 0, \hat{X}(t) < 0, S = 0]P[\hat{X}(t) < 0|S = 0] \right.$$  

$$+ P[X(t)_k < 0, \hat{X}_t > 0, S = 0]P[\hat{X}(t) > 0|S = 0]\right\}$$

$$- \lim_{k \to \infty} \left\{P[X(t)_k > 0, \hat{X}(t) < 0, S = 1]P[\hat{X}(t) < 0|S = 1]\right.$$  

$$+ P[X(t)_k > 0, \hat{X}_t > 0, S = 1]P[\hat{X}(t) > 0|S = 1]\right\} \text{ for large } t$$

$$= \{1 \cdot \frac{1}{2} + 0 \cdot \frac{1}{2}\} - \{0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2}\} \text{ by Lemma 1 and the assumption that }$$

$$P[\hat{X}(t) < 0|S = s_i] = P[\hat{X}(t) > 0|S = s_i] = 1/2 \text{ for } i = 1, 2.$$

$$= 0 \quad \text{Q.E.D.}$$

Sub-figures (a) and (b) of Figures 4-6 show simulation instances of Theorem 1 for finite-variance jump-diffusion and pure-jump additive white Levy noise in logistic, linear-threshold, and Gaussian neuron models.
Small amounts of additive Levy noise in continuous neuron models produce the SR effect by increasing the Shannon mutual information $I(S,Y)$ between realizations of a random (Bernoulli) subthreshold input signal $S$ and the neuron’s thresholded output random variable $Y$. The SR effect in Figures 2 (c) and 4-6 (c) lie outside the scope of the theorem because it occurs for infinite-variance $\alpha$-stable noise. Thus the SR effect in continuous neurons is not limited to finite-second-moment Levy noise.

V. LEVY NOISE BENEFITS IN SPIKING NEURON MODELS

We next demonstrate Levy SR noise benefits in three popular spiking neuron models: the leaky integrate-and-fire model [17, 28], the reduced Type I neuron model [53], and the FitzHugh-Nagumo (FHN) model [26, 16]. This requires the use of Lemma 2 as we discuss below. These neuron models have a one- or two-dimensional form of (1). A spike occurs when the membrane potential $x(t)$ crosses a threshold value from below. We measure the mutual information $I(S,R)$ between the input signal $s(t)$ and the output spike-rate response $R$ of these spiking neuron models. We define the average output spike-rate response
Figure 5: Mutual information Levy noise benefits in the linear-threshold continuous neuron (1)-(2) and (6). Additive white Levy noise $dL_t$ increases the mutual information of the linear-threshold neuron for the subthreshold input signal $s_1 = -0.4$, $s_2 = 0.4$, and $c = 2$. The types of Levy noise $dL_t$ are (a) Gaussian with uniformly distributed jumps, (b) pure-jump normal inverse Gaussian (NIG), and (c) symmetric $\alpha$-stable noise with $\alpha = 1.95$ (thick-tailed bell curve with infinite variance [62]). The dashed vertical lines show the total min-max deviations of the mutual information in 100 simulation trials.

$R$ in the time interval $(t_1, t_2]$ as

$$R = \frac{N(t_1, t_2]}{t_2 - t_1}$$

where $N(t_1, t_2]$ is the number of spikes in the time interval $(t_1, t_2]$.

**The Leaky Integrate-and-Fire Neuron Model**

The leaky integrate-and-fire neuron model has the form [17]

$$\dot{v} = -av + a - \delta + S + n$$

where $v$ is the membrane voltage, $a$ and $\delta$ are constants, $\delta/a$ is the barrier height of the potential, $S$ is an input signal, and $n$ is independent Gaussian white noise in the neural literature but here is Levy white noise. The input signal $S$ is subthreshold when $S < \delta$. The neuron emits a spike when the membrane voltage $v$ crosses the threshold value of 1 from below to above. The membrane voltage $v$ resets to $1 - \delta/a$ just after the neuron emits a spike.
Figure 6: Mutual information Levy noise benefits in the Gaussian or ‘radial basis’ continuous neuron (1)-(2) and (8). Additive white Levy noise $dL_t$ increases the mutual information of the Gaussian neuron for the subthreshold input signal $s_1 = 0.2$, $s_2 = 0.4$, and $c = 20$. The types of Levy noise $dL_t$ are (a) Gaussian with uniformly distributed jumps, (b) pure-jump normal inverse Gaussian (NIG), and (c) symmetric $\alpha$-stable noise with $\alpha = 1.95$ (thick-tailed bell curve with infinite variance [62]). The dashed vertical lines show the total min-max deviations of the mutual information in 100 simulation trials.

The Reduced Type I Neuron Model

The reduction procedure in [31, 36] gives a simple one-dimensional normal form [53] of the multi-dimensional dynamics of Type I neuron models:

\[
\dot{v} = \beta + v^2 + \sigma n
\]

where $v$ is the membrane potential, $\beta$ is the value of input signal, and $\sigma$ is the standard deviation of Gaussian white noise $n$ in the neural literature but here is Levy white noise. This reduced model (12) operates in a subthreshold or excitable regime when the input $\beta < 0$.

The FitzHugh-Nagumo (FHN) Neuron Model

The FHN neuron model [16, 26, 28] is a two-dimensional simplification of the Hodgkin and Huxley neuron model [34]. It describes the response of a so-called Type II excitable system [28, 54] that undergoes a Hopf bifurcation. The system first resides in the stable rest state for subthreshold inputs as do multistable systems. Then the system leaves the stable state in response to a strong input but returns to it after passing
through firing and refractory states in a manner that differs from the behavior of multistable systems. The
FHN neuron model is a limit-cycle oscillator of the form

\begin{align}
\epsilon \dot{v} &= -v(v^2 - \frac{1}{4}) - w + A + s(t) + n(t) \\
\dot{w} &= v - w
\end{align}

where \( v(t) \) is a fast (voltage) variable, \( w(t) \) is slow (recovery) variable, \( A \) is a constant (tonic) activation
signal, and \( \epsilon = 0.005 \). \( n(t) \) is a white Levy noise and \( s(t) \) is a subthreshold input signal—either \( s_1 \) or \( s_2 \).
We measure the neuron’s response to the input signal \( s(t) \) in terms of the transition (firing) rate \( r(t) \).

We can rewrite (42)-(43) as

\begin{align}
\epsilon \dot{v} &= -v(v^2 - \frac{1}{4}) - w + A_T - (B - s(t)) + n(t) \\
\dot{w} &= v - w
\end{align}

where \( B \) is a positive constant parameter that corresponds to the distance that the input signal \( s(t) \) must
overcome to cross the threshold. Then \( B - s(t) \) is the signal-to-threshold distance and so \( s(t) \) is subthreshold
when \( B - s(t) > 0 \). Our simulations used \( B = 0.007 \) and hence \( A = -(5/(12\sqrt{3} + 0.007)) \).

The deterministic FHN model \((n(t) \equiv 0 \text{ in (44)})\) performs relaxation oscillations and has an action
potential \( x(t) \) that lies between -0.6 and 0.6. The system emits a spike when \( x(t) \) crosses the threshold
value \( \theta = 0 \). We use a lowpass-filtered version of \( x(t) \) to avoid false spike detections due to the additive
noise. The lowpass filter is a 100-point moving-average smoother with a 0.001 second time-step.

We rewrite equations (42)-(43) as

\begin{align}
\dot{x}^1 &= -\frac{x^1}{\epsilon}((x^1)^2 - \frac{1}{4}) - \frac{x^2}{\epsilon} + \frac{A}{\epsilon} + \frac{s(t)}{\epsilon} + \frac{n(t)}{\epsilon} \\
\dot{x}^2 &= x^1 - x^2
\end{align}

Here \( x^1 = v \) and \( x^2 = w \). The corresponding matrix Itô stochastic differential equation is

\[ dX_t = b(X_{t^-})dt + c(X_{t^-})dL_t \]

where \( X_t = (X^1_t, X^2_t)^T \), \( L_t = (L^1_t, L^2_t)^T \).
\[ b(X_{t^-}) = \begin{bmatrix} b^1(X^1_{t^-}, X^2_{t^-}) \\ b^2(X^1_{t^-}, X^2_{t^-}) \end{bmatrix} = \begin{bmatrix} -\frac{X^1_{t^-}}{\epsilon}((X^1_{t^-})^2 - \frac{1}{4}) - \frac{X^2_{t^-}}{\epsilon} + \frac{A}{\epsilon} + \frac{3}{4} \\ X^1_{t^-} - X^2_{t^-} \end{bmatrix}, \]

and \[ c(X_{t^-}) = \begin{bmatrix} \frac{\sigma}{\epsilon} \\ 0 \end{bmatrix}. \]

Thus all of the above spiking neuron models have the SDE form (23). Note that the drift term of the leaky integrate-and-fire neuron model is globally Lipschitz while the drift term of the reduced Type I neuron model is locally Lipschitz. The Lipschitz condition is not easy to verify in the FHN model.

We now show that the drift term \( b^1(X_t) \) in the preceding equation does not satisfy the global Lipschitz condition. Note that \( b^1(X_t) \) is differentiable on \( \mathbb{R}^2 \) because the partial derivatives of \( b^1(X^1_{t^-}, X^2_{t^-}) \) exist and are continuous on \( \mathbb{R}^2 \). Suppose that \( b^1(X_t) \) satisfies the following global Lipschitz condition: There exists a constant \( K > 0 \) such that

\[ |b^1(Z_t) - b^1(Y_t)| \leq K||Z_t - Y_t|| \quad \text{for all } Z_t \text{ and } Y_t \in \mathbb{R}^2 \text{ and } t \in [0, T]. \]

Then the mean-value theorem gives

\[
\begin{align*}
b^1(Z_t) - b^1(Y_t) &= \left[ \frac{\partial b^1(\zeta)}{\partial X^1_t} \right] \cdot [Z^1_t - Y^1_t] + \left[ \frac{\partial b^1(\zeta)}{\partial X^2_t} \right] \cdot [Z^2_t - Y^2_t] \quad \text{for some } \zeta \text{ between } Z_t \text{ and } Y_t \text{ in } \mathbb{R}^2. \tag{49}
\end{align*}
\]

Then

\[
\begin{align*}
|b^1(Z_t) - b^1(Y_t)| &\geq \left| \frac{\partial b^1(\zeta)}{\partial X^1_t} \right| ||Z^1_t - Y^1_t|| - \left| \frac{\partial b^1(\zeta)}{\partial X^2_t} \right| ||Z^2_t - Y^2_t|| \\
&= \left| \frac{\partial b^1(\zeta)}{\partial X^1_t} \right| ||Z^1_t - Y^1_t|| - \frac{1}{\epsilon} ||Z^2_t - Y^2_t|| \quad \text{because } \frac{\partial b^1}{\partial X^2_t} = -\frac{1}{\epsilon} \\
&= \left| \frac{\partial b^1(\zeta)}{\partial X^1_t} \right| ||Z^1_t - Y^1_t|| \quad \text{choosing } Z_t \text{ and } Y_t \text{ such that } Z^2_t = Y^2_t \\
&= \left| \frac{\partial b^1(\zeta)}{\partial X^1_t} \right| ||Z_t - Y_t|| \\
&> K||Z_t - Y_t|| \quad \text{for some } Z_t \in \mathbb{R}^2 \text{ and } Y_t \in \mathbb{R}^2
\end{align*}
\]
because \( |\frac{\partial b^1}{\partial X_1} | = \frac{-3(X_1^2)}{\varepsilon} + \frac{1}{16} \) is unbounded and continuous on \( R^2 \) and so there is a domain \( D \subset R^2 \) such that \( |\frac{\partial b^1(\zeta)}{\partial X_1^2} | > K \) for all \( \zeta \in D \). Thus \( b^1(X_t) \) is not globally Lipschitz. So we cannot use Lemma 1 to prove the sufficient condition for the SR effect in the FHN neuron model (44)-(45).

But \( b^1(X_t) \) is locally Lipschitz. The partial derivatives of \( b^1(X_{1t}, X_{2t}) \) exist and are continuous on \( R^2 \). So \( \frac{\partial b^1}{\partial X_1} \) and \( \frac{\partial b^1}{\partial X_2} \) achieve their respective maxima on the compact set \( \{ \zeta \in R^2 : ||\zeta|| \leq n \} \). Then (49) gives the required local Lipschitz condition:

\[
|b^1(Z_t) - b^1(Y_t)| \leq \max \{ \sup_{\zeta} |\frac{\partial b^1(\zeta)}{\partial X_1^1}|, \sup_{\zeta} |\frac{\partial b^1(\zeta)}{\partial X_2^1}| \} ||Z_t - Y_t||
\]

for all \( Z_t \) and \( Y_t \) such that \( ||Z_t|| \leq n, ||Y_t|| \leq n, \) and \( ||\zeta|| \leq n \). Lemma 2 extends the conclusion of Lemma 1 to the locally Lipschitz drift terms \( b^i(X_t) \).

Theorem 2 below gives a ‘forbidden-interval’ sufficient condition for a Levy SR noise benefits in spiking neuron models such as the leaky integrate-and-fire model [17, 28], the reduced Type I neuron model [53], and the FitzHugh-Nagumo (FHN) model [26, 16]. It shows that these neuron models enjoy SR noise benefits if the noise mean \( \mu \) falls to the left of a bound. Theorem 2 requires Lemma 2 to extend the conclusion of Lemma 1 to the locally Lipschitz drift terms \( b^i(X_t) \). The Appendix gives the proof of Lemma 2.

**Lemma 2:** Let \( b^i : R^d \rightarrow R \) and \( c^i_j : R^d \rightarrow R \) in (23)-(24) be measurable functions that satisfy the respective local and global Lipschitz conditions

\[
||b^i(z) - b^i(y)|| \leq C_n||z - y|| \quad \text{when } ||z|| \leq n \text{ and } ||y|| \leq n \\
||c^i_j(z) - c^i_j(y)|| \leq K_1||z - y|| \quad \text{for all } z \text{ and } y \in R^d
\]

and

\[
|c^i_j|^2 \leq H^i_j \quad \text{for } i = 1, ..., d \text{ and } j = 1, ..., m.
\]

Suppose

\[
dX_t = b(X_t)dt + c(X_t-)dL_t
\]

\[
d\hat{X}_t = b(\hat{X}_t)dt
\]

where \( dL_t \) is a Levy noise with \( \mu = 0 \) and finite second moments. Then for every \( T \in R^+ \) and for every \( \varepsilon > 0 \):
\[
E[ \sup_{0 \leq t \leq T} ||X_t - \hat{X}_t||^2 > \varepsilon ] \to 0 \quad \text{as} \quad \sigma^j \to 0 \quad \text{and} \quad \nu^j \to 0 \quad \text{for all} \quad j = 1, \ldots, m,
\]  
(55)

and hence
\[
P[ \sup_{0 \leq t \leq T} ||X_t - \hat{X}_t||^2 > \varepsilon ] \to 0 \quad \text{as} \quad \sigma^j \to 0 \quad \text{and} \quad \nu^j \to 0 \quad \text{for all} \quad j = 1, \ldots, m
\]  
(56)

since mean-square convergence implies convergence in probability.

We can now state and prove Theorem 2.

**Theorem 2:** Suppose that the spiking neuron models (40)-(41) and (42)-(43) have the form of the Levy SDE (23) with a bounded globally Lipschitz Levy diffusion term \( c(X_{t^-}) \leq H \) and that the additive Levy noise has drift velocity \( \mu \). Suppose that the input signal \( S(t) \in \{s_1, s_2\} \) is subthreshold: \( S(t) < B \). Suppose there is some statistical dependence between the input random variable \( S \) and the output spike-rate random variable \( R \) so that \( I(S, R) > 0 \). Then the spiking neuron models (40)-(41) and (42)-(43) exhibit the SR effect in the sense that \( I(S, R) \to 0 \) as the Levy noise parameters \( \sigma \to 0 \) and \( \nu \to 0 \) if \( H\mu < B - s_2 \).

**Proof:** Let \( \{\sigma_k, \nu_k\}_{k=1}^{\infty} \) be any decreasing sequence of Levy noise parameters such that \( \sigma_k \to 0 \) and \( \nu_k \to 0 \) as \( k \to \infty \). Define \( X(t)_k \) and \( R_k \) as the respective solution process and spiking rate process of the FHN spiking neuron model (48) with Levy noise parameters \( \sigma_k \) and \( \nu_k \) instead of \( \sigma \) and \( \nu \).

Suppose that \( \mu \neq 0 \). We can absorb the drift \( c(X_{t^-})\mu \) into the input signal \( S \) because the Levy noise \( L_t \) is additive in all the neuron models. Then the new input signal \( S' = S + c(X_{t^-})\mu \) and this does not affect the Lipschitz continuity of \( b(X_{t^-}) \) in (21). \( S' \) is subthreshold \( \langle S' < B \rangle \) because \( c(X_{t^-})\mu < H\mu < B - s_2 \) where \( s_2 = \max\{s_1, s_2\} \). So we lose no generality if we consider the noise \( dL_t \) with \( \mu = 0 \) and let \( S \in \{s_1, s_2\} \) be subthreshold in the continuous neuron models (21). This allows us to use Lemma 2.

Recall that \( I(S, R) = 0 \) if and only if \( S \) and \( R \) are statistically independent [20]. So we need to show only that \( f_{SR}(s, r) = f_S(s)f_{R(r)} \) or \( f_{R|r}(r|s) = f_R(r) \) as \( \sigma \to 0 \) and \( \nu \to 0 \) for signal symbols \( s \in \{s_1, s_2\} \).
and for all \( r \geq 0 \). Here \( f_{SR} \) is the joint probability density function and \( f_{S|R} \) is the conditional density function. This is logically equivalent to \( F_{R|S} = F_R \) as \( \sigma_k \to 0 \) and \( \nu_k \to 0 \) as \( k \to 0 \) where \( F_{R|S} \) is the conditional distribution function \([25]\). Again the theorem of total probability and the two-symbol alphabet set \( \{s_1, s_2\} \) give

\[
F_R(r) = \sum_s F_{R|S}(r|s)P_S(s)
\]

\[
= F_{R|S}(r|s_1)P_S(s_1) + F_{R|S}(r|s_2)P_S(s_2)
\]

\[
= F_{R|S}(r|s_1)P_S(s_1) + F_{R|S}(r|s_2)(1 - P_S(s_1))
\]

\[
= (F_{R|S}(r|s_1) - F_{R|S}(r|s_2))P_S(s_1) + F_{R|S}(r|s_2)
\]

So we need to show that \( \lim_{k \to \infty} F_{R_k|S}(r|s_1) - F_{R_k|S}(r|s_2) = 0 \) for all \( r \geq 0 \). This holds if and only if

\[
\lim_{k \to \infty} P[R_k > r|S = s_1] - P[R_k > r|S = s_2] = 0
\]

We prove that \( \lim_{k \to \infty} P[R_k > r|S = s_i] = 0 \) for \( i = 1 \) and \( i = 2 \). Note that if \( r > 0 \) for (48) then \( X_1(t)_k \) must cross the firing or spike threshold \( \theta \). Then

\[ P[R_k > r|S = s_i] \leq P[\sup_{t_1 \leq t \leq t_2} X_1(t)_k > \theta|S = s_i]. \]

Then Lemma 2 shows that the required limit goes to zero:

\[
\lim_{k \to \infty} P[R_k > r|S = s_2] \leq \lim_{k \to \infty} P[\sup_{t_1 \leq t \leq t_2} X_1(t)_k > \theta|S = s_i]
\]

\[
= \lim_{n \to \infty} P[\sup_{t_1 \leq t \leq t_2} X_1(t)_k > \theta, \hat{X}_1(t) < \theta|S = s_i]
\]

because \( \hat{X}_1(t) \) converges to the FHN fixed-point \( Z_{F_1} < \theta \) for large \( t \)

\[
= 0 \quad \text{by Lemma 2.}
\]

Q.E.D.

Sub-figures (a) and (b) of Figures 7-8 show simulation instances of Theorem 2 for finite-variance diffusion and jump-diffusion white Levy noise in the leaky integrate-and-fire and the FHN neuron models. Small
Figure 7: Mutual information Levy noise benefits in the leaky integrate-and-fire spiking neuron model (40). Additive white Levy noise $dL_t$ increases the mutual information of the IF neuron with parameters $a = 0.5$ and $\delta = 0.02$ for the subthreshold input signal $s_1 = 0.005$ and $s_2 = 0.012$. The types of Levy noise $dL_t$ are (a) Gaussian, (b) Gaussian with uniformly distributed jumps, and (c) symmetric $\alpha$-stable noise with $\alpha = 1.95$ (thick-tailed bell curve with infinite variance [62]). The dashed vertical lines show the total min-max deviations of the mutual information in 100 simulation trials.

Amounts of additive Levy noise in these spiking neuron models produce the SR effect in terms of the noise-enhanced Shannon mutual information $I(S,Y)$ between realizations of a random (Bernoulli) subthreshold input signal $S$ and the neuron’s thresholded output random variable $Y$. The SR effects in Figures 7-8(c) again lie outside the scope of Theorem 2 because they occur for infinite-variance $\alpha$-stable noise and because Theorem 2 requires noise with finite second moments. Thus the SR effect in spiking neurons is not limited to finite second moment Levy noise.

VI. CONCLUSIONS

Levy noise processes can benefit several continuous and spiking neuron models because general forms of the SR ‘forbidden interval’ theorem hold for several types of Levy noise. The generality of Levy noise extends simple Brownian models of noise to more complex and realistic Poisson jump models of noise that can affect biological and model neurons. But both Levy SR theorems require the finite second-moment restrictions of the two lemmas. This rules out the important class of stable noise distributions in all but
Figure 8: Mutual information Levy noise benefits in the FHN spiking neuron (42)-(43). Additive white Levy noise $dL_t$ increases the mutual information of the FHN neuron for the subthreshold input signal $s_1 = -0.0045$ and $s_2 = 0.0045$. The types of Levy noise $dL_t$ are (a) Gaussian, (b) Gaussian with uniformly distributed jumps, and (c) symmetric $\alpha$-stable noise with $\alpha = 1.9$ (thick-tailed bell curve with infinite variance [62]). The dashed vertical lines show the total min-max deviations of the mutual information in 100 simulation trials.

the Gaussian or pure-diffusion case.

Relaxing the second-moment assumption may produce stochastic differential equations that are not mathematically tractable. Yet the simulation evidence of Figure 1 and sub-figure (c) of Figures 3-8 shows that the SR noise benefit continues to hold for several stable models where the noise has infinite variance and infinite higher-order moments. It is an open research question whether a more general Levy SR result can include these and other observed noise benefits in continuous and spiking neuron models.

APPENDIX: PROOF OF LEMMAS

The proof of Lemma 2 relies on the proof technique of Lemma 1 in which we bound a mean-squared term by four additive terms and then show that each of the four terms goes to zero in the limit.

**Lemma 1:** Let $b^i : \mathbb{R}^d \rightarrow \mathbb{R}$ and $c^i_j : \mathbb{R}^d \rightarrow \mathbb{R}$ in (23)-(24) be measurable functions that satisfy the global Lipschitz conditions
\[ ||b^i(x_1) - b^i(x_2)|| \leq K_1||x_1 - x_2|| \]  
\[ ||c^i_j(x_1) - c^i_j(x_2)|| \leq K_2||x_1 - x_2|| \]

and \( |c^i_j|^2 \leq H^i_j \) for all \( x_1 \) and \( x_2 \in \mathbb{R}^d \) and for \( i = 1, ..., d \) and \( j = 1, ..., m. \)

Suppose \( dX_t = b(X_t)dt + c(X_t)dL_t \)
\[ d\hat{X}_t = b(\hat{X}_t)dt \]
where \( dL_t \) is a Levy noise with \( \mu = 0 \) and finite second moments. Then for every \( T \in \mathbb{R}^+ \) and for every \( \varepsilon > 0: \)
\[ E[\sup_{0 \leq t \leq T} ||X_t - \hat{X}_t||^2 > \varepsilon] \rightarrow 0 \] as \( \sigma^j \rightarrow 0 \) and \( \nu^j \rightarrow 0 \) for all \( j = 1, ..., m, \)

and hence
\[ P[\sup_{0 \leq t \leq T} ||X_t - \hat{X}_t||^2 > \varepsilon] \rightarrow 0 \] as \( \sigma^j \rightarrow 0 \) and \( \nu^j \rightarrow 0 \) for all \( j = 1, ..., m \)

since mean-square convergence implies convergence in probability.

**Proof:**

The Lipschitz conditions (60) and (61) ensure that the process \( X_t \) exists [3] for \( t \geq 0 \) in (63). Then the proof commences with the inequality
\[ \sup_{0 \leq t \leq T} ||X_t - \hat{X}_t||^2 \leq \sum_{i=1}^{d} \sup_{0 \leq t \leq T} (X^i_t - \hat{X}^i_t)^2 \]

which implies that
\[ E[\sup_{0 \leq t \leq T} ||X_t - \hat{X}_t||^2] \leq \sum_{i=1}^{d} E[\sup_{0 \leq t \leq T} (X^i_t - \hat{X}^i_t)^2]. \]

Thus the result follows if \( E[\sup_{0 \leq t \leq T}(X^i_t - \hat{X}^i_t)^2] \rightarrow 0 \) as \( \gamma \rightarrow 0 \) for \( i = 1, ..., d. \) Equations (26) and (63)-(64) imply
\[ X^i_t - \hat{X}^i_t = \int_0^t [b^i(X_s) - b^i(\hat{X}_s)]ds + \sum_{j=1}^{m} \int_0^t \sigma^i_j(X_s)dB^j_s + \sum_{j=1}^{m} \int_{|y^j| < 1} F^i_j(X_s, y^j)\tilde{N}^j(ds, dy^j) \]
\[ + \sum_{j=1}^{m} \int_{|y^j| \geq 1} G^i_j(X_s, y^j)N^j(ds, dy^j). \]
This gives an upper bound on the squared difference as

\[(X_i^t - \hat{X}_i^t)^2 \leq (3m + 1) \left( \left\{ \int_0^t [b^i(X_s^t) - b^i(\hat{X}_s^t)]ds \right\}^2 + \sum_{j=1}^m \left\{ \int_0^t \sigma^i_j(X_s^t)dB^i_s \right\}^2 \right) + \sum_{j=1}^m \left\{ \int_0^t \int_{|y^j|<1} F^j_i(X_s^t, y^j)\tilde{N}^j(ds, dy^j) \right\}^2 + \sum_{j=1}^m \left\{ \int_0^t \int_{|y^j|\geq 1} G^j_i(X_s^t, y^j)N^j(ds, dy^j) \right\}^2 \]  

(70)

because \((u_1 + ... + u_n)^2 \leq n(u_1^2 + ... + u_n^2)\). The Cauchy-Schwartz inequality gives

\[\left\{ \int_0^t [b^i(X_s^t) - b^i(\hat{X}_s^t)]ds \right\}^2 \leq \left( \int_0^t ds \right) \left( \int_0^t [b^i(X_s^t) - b^i(\hat{X}_s^t)]^2 ds \right). \]  

(71)

Now put (71) in the first term of (70) and then take expectations of the supremum on both sides to get four additive terms as an upper bound:

\[E \left[ \sup_{0 \leq t \leq T} (X_i^t - \hat{X}_i^t)^2 \right] \leq (3m + 1) \left( E \left[ \sup_{0 \leq t \leq T} t \int_0^t [b^i(X_s^t) - b^i(\hat{X}_s^t)]^2 ds \right] + \sum_{j=1}^m E \left[ \sup_{0 \leq t \leq T} \left\{ \int_0^t \sigma^i_j(X_s^t)dB^i_s \right\}^2 \right] \right. \]

\[\left. + \sum_{j=1}^m E \left[ \sup_{0 \leq t \leq T} \left\{ \int_0^t \int_{|y^j|<1} F^j_i(X_s^t, y^j)\tilde{N}^j(ds, dy^j) \right\}^2 \right] \right. \]

\[\left. + \sum_{j=1}^m E \left[ \sup_{0 \leq t \leq T} \left\{ \int_0^t \int_{|y^j|\geq 1} G^j_i(X_s^t, y^j)N^j(ds, dy^j) \right\}^2 \right] \right). \]  

(72)

We next show that each of the four terms goes to zero. Consider the first term on the right-hand side of (72):

\[E \left[ \sup_{0 \leq t \leq T} t \int_0^t [b^i(X_s^t) - b^i(\hat{X}_s^t)]^2 ds \right] \leq TE \left[ \sup_{0 \leq t \leq T} \int_0^t [b^i(X_s^t) - b^i(\hat{X}_s^t)]^2 ds \right] \]

\[\leq TK_i^2E \left[ \sup_{0 \leq t \leq T} \int_0^t [X_s^t - \hat{X}_s^t]^2 ds \right] \]  

by the Lipschitz condition (60)

\[\leq TK_i^2 \int_0^T E \left[ \sup_{0 \leq u \leq s} [X_u^t - \hat{X}_u^t]^2 \right] ds. \]  

(73)

The second term

\[E \left[ \sup_{0 \leq t \leq T} \left\{ \int_0^t \sigma^i_j(X_s^t)dB^i_s \right\}^2 \right] \leq 4E \left[ \int_0^T \left\{ \sigma^i_j(X_s^t)dB^i_s \right\}^2 \right] \]
because \( \int_0^t \sigma_j^i (X_{s-}) dB^i_s \) is a martingale and so we can apply Doob's \( L^p \) inequality \([56]\): 

\[
E \left[ \sup_{a \leq t \leq b} |U_t|^p \right] \leq \left( \frac{p}{p-1} \right)^p E|U_b|^p
\]

if \( \{U_t\}_{t \geq 0} \) is a real-valued martingale, \([a, b] \) is a bounded interval of \( \mathbb{R}^+ \), \( U_t \in L^p(\Omega, \mathbb{R}) \), and if \( p > 1 \) (\( p = 2 \) in our case). But

\[
4E \left[ \left\{ \int_0^T \sigma_j^i (X_{s-}) dB^i_s \right\}^2 \right] = 4 \int_0^T E \left[ \left\{ \sigma_j^i (X_{s-}) \right\}^2 \right] ds
\]

by Itô isometry \([3]\): 

\[
E \left( \left\{ \int_0^T f(t, y) dB_t \right\}^2 \right) = \int_0^T E((f(t, y))^2) ds
\]

if \( f \in \mathcal{H}_2([0, T], \mathbb{R}) \) where \( \mathcal{H}_2([0, T], \mathbb{R}) \) is the space of all real-valued measurable \( \{\mathcal{F}_t\} \)-adapted processes such that \( E \left( \int_0^T |f(t, y)|^2 ds \right) < \infty \). Then

\[
4 \int_0^T E \left[ \left\{ \sigma_j^i (X_{s-}) \right\}^2 \right] ds \leq 4(\sigma_j^i)^2 \int_0^T E \left[ \left\{ \sigma_j^i (X_{s-}) \right\}^2 \right] ds \quad \text{by definition of} \quad \sigma_j^i (X_{s-})
\]

\[
\leq 4(\sigma_j^i)^2 TH_j^i \quad \text{because} \quad |c_j^i|^2 \leq H_j^i.
\]

Note that

\[
E \left[ \sup_{0 \leq t \leq T} \left\{ \int_0^t \int_{|y| < 1} F_j^i (X_{s-}, y^j) \tilde{N}^j (ds, dy^j) \right\}^2 \right] \leq 4E \left[ \left\{ \int_0^T \int_{|y| < 1} F_j^i (X_{s-}, y^j) \tilde{N}^j (ds, dy^j) \right\}^2 \right] \quad \text{by Doob's} \quad L^p \quad \text{inequality}
\]

\[
= 4E \left[ \left\{ \int_0^T \int_{|y| < 1} c_j^i (X_{s-}) y^j \tilde{N}^j (ds, dy^j) \right\}^2 \right] \quad \text{by definition of} \quad F_j^i (X_{s-}, y^j)
\]

\[
\leq 4H_j^i E \left[ \left\{ \int_0^T \int_{|y| < 1} y^j \tilde{N}^j (ds, dy^j) \right\}^2 \right] \quad \text{because} \quad |c_j^i|^2 \leq H_j^i
\]

\[
\leq 4H_j^i E \left[ \left\{ \int_{|y| < 1} y^j \tilde{N}^j (T, dy^j) \right\}^2 \right] \quad \text{by Itô isometry and (20)}.
\]

Similar arguments and (19) give

\[
E \left[ \sup_{0 \leq t \leq T} \left\{ \int_0^t \int_{|y| \geq 1} G_j^i (X_{s-}, y^j) N^j (ds, dy^j) \right\}^2 \right] \leq 4H_j^i T \int_{|y| \geq 1} |y|^2 \nu^j (dy^j).
\]

Substituting the above estimates (73), (74), (75), and (76) in inequality (72) gives

\[
E \left[ \sup_{0 \leq t \leq T} (X_t^i - \tilde{X}_t^i)^2 \right] \leq (3m + 1) \left( TK_1^2 \cdot \int_0^T E \left[ \sup_{0 \leq u \leq s} [X_{u-} - \tilde{X}_{u-}]^2 \right] ds \right.
\]

\[
\left. + 4mTH_j^i \cdot \max_j \left[ (\sigma_j^i)^2 + \int_{\mathbb{R}} |y|^2 \nu^j (dy^j) \right] \right).
\]
Rewrite this inequality as

\[ z(T) \leq A + Q \int_0^T z(s) d(s) \]  

(78)

where \( z(T) = E \left[ \sup_{0 \leq t \leq T} (X^i_t - \hat{X}^i_t)^2 \right] \), \( A = (3m + 1)4mTH_i^j \cdot \max_j \left[ (\sigma^j)^2 + \int_{\mathbb{R}} |y^j|^2 \nu^j(dy^j) \right] \), and \( Q = (3m + 1)TK_i^2 \). Then we get \( z(T) \leq Ae^{QT} \) by Gronwall’s inequality [24]: \( \phi(t) \leq \alpha e^{\beta t} \) for all \( t \in [0, T] \) and for real continuous \( \phi(t) \) in \( [0, T] \) such that \( \phi(t) \leq \alpha + \beta \int_0^t \phi(\tau) d\tau \) where \( t \in [0, T] \) and \( \beta > 0 \). Note that \( A \to 0 \) as \( \sigma^j \to 0 \) and \( \nu^j \to 0 \). Hence

\[ E[ \sup_{0 \leq t \leq T} (X^i_t - \hat{X}^i_t)^2 ] \to 0 \]  

as \( \sigma^j \to 0 \) and \( \nu^j \to 0 \).  

(79)

for each \( j = 1, \ldots, m \). Applying (79) to inequality (68) implies (65) and hence implies (66). Q.E.D.

**Lemma 2:** Let \( b^i : \mathbb{R}^d \to \mathbb{R} \) and \( c^i_j : \mathbb{R}^d \to \mathbb{R} \) in (23)-(24) be measurable functions that satisfy the respective local and global Lipschitz conditions

\[ ||b^i(z) - b^i(y)|| \leq C_n||z - y|| \text{ when } ||z||, ||y|| \leq n \]  

(80)

\[ ||c^i_j(z) - c^i_j(y)|| \leq K_1||z - y|| \text{ for all } z, y \in \mathbb{R}^d \]  

(81)

and \( |c^i_j|^2 \leq H_i^j \) for \( i = 1, \ldots, d \) and \( j = 1, \ldots, m \).  

(82)

Suppose \( dX_t = b(X_t) dt + c(X_t) dL_t \)

(83)

\[ d\hat{X}_t = b(\hat{X}_t) dt. \]

(84)

where \( dL_t \) is a Levy noise with \( \mu = 0 \) and finite second moments. Then for every \( T \in \mathbb{R}^+ \) and for every \( \varepsilon > 0 \):

\[ E[ \sup_{0 \leq t \leq T} ||X_t - \hat{X}_t||^2 > \varepsilon ] \to 0 \]  

as \( \sigma^j \to 0 \) and \( \nu^j \to 0 \) for all \( i = 1, \ldots, m \),

(85)

and hence

\[ P[ \sup_{0 \leq t \leq T} ||X_t - \hat{X}_t||^2 > \varepsilon ] \to 0 \]  

as \( \sigma^j \to 0 \) and \( \nu^j \to 0 \) for all \( i = 1, \ldots, m \).  

(86)
since mean-square convergence implies convergence in probability.

Proof:
First define the function \( \tilde{b}_i^r \) such that

(i) \( \tilde{b}_i^r(x) = b_i(x) \) for \( ||x|| \leq r \)
(ii) \( \tilde{b}_i^r(x) = 0 \) for \( ||x|| \geq 2r \)
(iii) \( \tilde{b}_i^r(x) = ((2r - ||x||)/r)b_i(rx/||x||) \) for \( r \leq ||x|| \leq 2r \).

We then show that the function \( \tilde{b}_i^r \) is globally Lipschitz:
\[
||\tilde{b}_i^r(x) - \tilde{b}_i^r(y)|| \leq C'_r ||x - y|| \quad \text{for } x, y \in \mathbb{R}^n.
\]

Consider the function \( \tilde{b}_i^r(x) \). Write
\[
\tilde{b}_i^r(x) = \begin{cases} 
  b_i(x) & \text{if } ||x|| \leq r \\
  f(x)g_i^r(x) & \text{if } r \leq ||x|| \leq 2r 
\end{cases}
\]
where
\[
f(x) = ((2r - ||x||)/r) \quad \text{and} \quad g_i^r(x) = b_i(rx/||x||).
\]

The definition of \( \tilde{b}_i^r \) implies that it is Lipschitz continuous on the region \( D_1 = \{ ||x|| \leq r \} \):
\[
||\tilde{b}_i^r(x) - \tilde{b}_i^r(y)|| \leq C_r ||x - y|| \quad \text{for all } x, y \in \mathbb{R}^d.
\]

We first show that \( \tilde{b}_i^r(x) \) is Lipschitz continuous on the region \( D_2 = \{ r \leq ||x|| \leq 2r \} \). For \( x, y \in D_2 = \{ r \leq ||x|| \leq 2r \} \):
\[
||f(x) - f(y)|| = \frac{||y|| - ||x||}{r} \quad \text{by definition of } f
\]
\[
\leq \frac{||x - y||}{r}
\]
and

\[ \|g^i(x) - g^i(y)\| = \|b^i(rx/\|x\|) - b^i(ry/\|y\|)\| \quad \text{by definition of } g^i \]  
\[ \leq C_r \frac{r x}{\|x\|} - \frac{r y}{\|y\|} \]  
\[ \text{because } \frac{r x}{\|x\|} \in D_1 \text{ for all } s \in \mathbb{R}^d \text{ and } b^i \text{ is Lipschitz continuous on } D_1 \]  
\[ \leq \frac{C_r}{2} \|x - y\| \quad \text{because } r \geq \|x\|, \|y\| \geq 2r. \]  

Hence

\[ \|\tilde{b}_r^i(x) - \tilde{b}_r^i(y)\| = \|f(x)g^i(x) - f(y)g^i(y)\| \]  
\[ \leq \|f(x)g^i(x) - f(z)g^i(z)\| + \|f(z)g^i(z) - f(y)g^i(y)\| \]  
\[ = \|f(x)g^i(x) - f(x)g^i(y)\| + \|f(x)g^i(y) - f(y)g^i(y)\| \]  

by choosing \( z \) on the line segment between 0 and \( y \) such that \( \|z\| = \|x\| \)

\[ = \|f(x)\|\|g^i(x) - g^i(y)\| + \|g^i(y)\|\|f(x) - f(y)\| \]  
\[ \leq \|f\|_{\infty,2}\|g^i(x) - g^i(y)\| + \|g\|_{\infty,2}\|f(x) - f(y)\| \]  

where we define \( \|v\|_{\infty,i} = \sup\{\|v(s)\| : s \in D_i\} \)

\[ \leq \|f\|_{\infty,2} C_r \frac{r}{2} \|x - y\| + \|g\|_{\infty,2} \frac{r}{r} \]  
\[ \leq C'_r \|x - y\| \quad \text{where } C'_r = \|f\|_{\infty,2} \frac{C_r}{2} + \|g\|_{\infty,2} \frac{r}{r}. \]  

So \( \tilde{b}^r(x) \) is Lipschitz continuous on \( D_2 \).

We next show that \( \tilde{b}^r(x) \) is Lipschitz continuous on \( D_1 \) and \( D_2 \). Choose \( x \in D_1, y \in D_2 \), and a point \( z \) of \( \partial D_1 \) on the line segment between \( x \) and \( y \). Then

\[ \|\tilde{b}_r^i(x) - \tilde{b}_r^i(y)\| \leq \|\tilde{b}_r^i(x) - \tilde{b}_r^i(z)\| + \|\tilde{b}_r^i(z) - \tilde{b}_r^i(y)\| \]  
\[ \leq C_r \|x - z\| + C'_r \|z - y\| \]  
\[ \leq C'_r \|x - y\| \quad \text{because } C'_r \geq C_r \text{ and } \|x - z\| + \|z - y\| = \|x - y\|. \]  

So \( \tilde{b}^r(x) \) is Lipschitz continuous with coefficient \( C'_r \) on \( \|x\| \leq 2r \). We now choose \( x \in (D_1 \cup D_2), y \in \)
\((D_1 \cup D_2)^c\), and a point \(z\) of \(\partial(D_1 \cup D_2)^c\) on the line segment between \(x\) and \(y\). Then

\[
\|\tilde{b}_r(x) - \tilde{b}_r(y)\| \leq \|\tilde{b}_r(x) - \tilde{b}_r(z)\| + \|\tilde{b}_r(z) - \tilde{b}_r(y)\|
\]

(104)

\[
\leq C_r\|x - z\| + 0
\]

(105)

\[
\leq C'_r\|x - z\| + C'_r\|z - y\|
\]

(106)

\[
= C'_r\|x - y\|
\]

(107)

Then (89), (100), (103), and (107) show that \(\tilde{b}_r(x)\) is Lipschitz continuous with coefficient \(C'_r\) on \(\mathbb{R}^d\).

Consider next the SDE

\[
d\tilde{X}_t = \tilde{b}_r(\tilde{X}_t)dt + c(\tilde{X}_t - dL_t.
\]

(108)

Lemma 1 holds for (108) and so we can write

\[
P[\sup_{0 \leq t \leq T} \|\tilde{X}_t - \tilde{X}_t\|^2 > \varepsilon] \to 0 \quad \text{as} \quad \sigma^i \to 0 \quad \text{and} \quad \nu^i \to 0 \quad \text{for all} \quad j = 1, ..., m.
\]

(109)

Now define \(\tilde{T}_r = \inf\{t \geq 0 : \|\tilde{X}_t\| > r\}, T_r = \inf\{t \geq 0 : \|X_t\| > r\}, \text{and} \tau_r = \inf\{t : \|X_t\| > r \text{ or } \|	ilde{X}_t\| > r\}

= \min\{\tilde{T}_r, T_r\}. \) Then \(X_t\) and \(\tilde{X}_t\) satisfy (83) on \([0, \tau_r]\). Note that \(\tilde{T}_r\) and \(T_r\) are stopping times and thus \(\tau_r\) is also a stopping time. So arguments similar to those of the proof of Lemma 1 ((68)-(76) with appropriate modifications) give

\[
E\left[\sup_{0 \leq u \leq \min\{t, \tau_r\}} (X_u^i - \tilde{X}_u^i)^2\right] \leq Q' \int_0^t E\left[\sup_{0 \leq u \leq \min\{s, \tau_r\}} (X_u^i - \tilde{X}_u^i)^2\right] ds
\]

(110)

Then

\[
E\left[\sup_{0 \leq u \leq \min\{t, \tau_r\}} (X_u^i - \tilde{X}_u^i)^2\right] \leq 0 \quad \text{by Gronwall’s inequality.}
\]

(111)

Hence \(X_t^i = \tilde{X}_t^i\) holds almost surely on \([0, \tau_r]\) for \(i = 1, ..., d\). So \(X_t = \tilde{X}_t\) holds almost surely on \([0, \tau_r]\).

This result and (109) give

\[
P[\sup_{0 \leq t \leq \tau_r} \|X_t - \tilde{X}_t\|^2 > \varepsilon] \to 0 \quad \text{as} \quad \sigma^i \to 0 \quad \text{and} \quad \nu^i \to 0 \quad \text{for all} \quad j = 1, ..., m.
\]

(112)

We need now show only that \(\tau_r \to \infty\) almost surely as \(r \to \infty\) to prove (86). So we need to prove that \(\tilde{T}_r \to \infty\) and \(T_r \to \infty\) almost surely as \(r \to \infty\). Chebychev’s inequality implies that

\[
P(\tilde{T}_r \leq T) = P\left(\sup_{0 \leq t \leq T} \tilde{X}_t^i \geq r\right) \leq E\left[\sup_{0 \leq u \leq \tau_r} (\tilde{X}_u^i)^2\right] / r^2.
\]

(113)
Again arguments similar to those of the proof of Lemma 1 ((68)-(76) with appropriate modifications) give

\[
E \left[ \sup_{0 \leq t \leq T} (\tilde{X}_t)^2 \right] \leq A'' + Q'' \int_0^T E \left[ \sup_{0 \leq u \leq s} (\tilde{X}_u)^2 \right] ds. \tag{114}
\]

Thus

\[
E \left[ \sup_{0 \leq t \leq T} (\tilde{X}_t)^2 \right] \leq A'' e^{Q'' T} \quad \text{by Gronwall's inequality} \tag{115}
\]

where \(A''\) and \(Q''\) do not depend on \(r\) because we do not use the Lipschitz condition in the derivation of (114). Then (113) and (115) imply that

\[
P(\tilde{T}_r \leq T) \leq \frac{E \left[ \sup_{0 \leq t \leq T} (\tilde{X}_t)^2 \right]}{r^2} \rightarrow 0 \quad \text{as} \quad r \rightarrow \infty. \tag{116}
\]

Thus \(\tilde{T}_r \rightarrow \infty\) almost surely as \(r \rightarrow \infty\). We can similarly show that \(T_r \rightarrow \infty\) almost surely as \(r \rightarrow \infty\). This implies the claim (86). Q.E.D.

References


