IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 28 February 2024, accepted 12 March 2024, date of publication 19 March 2024, date of current version 26 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3379324

RESEARCH ARTICLE

Does Adding of Neurons to the Network Layer Lead to Increased Transmission Efficiency?

BARTOSZ PAPROCKI[1], AGNIESZKA PREGOWSKA[2], AND JANUSZ SZCZEPANSKI[2]
[1]Institute of Mechanics and Applied Computer Science, Kazimierz Wielki University, 85-074 Bydgoszcz, Poland
[2]Institute of Fundamental Technological Research, Polish Academy of Sciences, 02-106 Warsaw, Poland

Corresponding author: Janusz Szczepanski (jszczepa@ippt.pan.pl)

This work was supported in part by the National Centre for Research and Development under Grant Infostrateg I/0042/2021-00.

ABSTRACT The aim of this study is to contribute to the important question in Neuroscience of whether the number of neurons in a given layer of a network affects transmission efficiency. Mutual Information, as defined by Shannon, between the input and output signals for certain classes of networks is analyzed theoretically and numerically. A Levy-Baxter probabilistic neural model is applied. This model includes all important qualitative mechanisms involved in the transmission process in the brain. We derived analytical formulas for the Mutual Information of input signals coming from Information Sources as Bernoulli processes. These formulas depend on the parameters of the Information Source, neurons and network. Numerical simulations were performed using these equations. It turned out, that the Mutual Information starting from a certain value increased very slowly with the number of neurons being added. The increase is of the rate m^{-c} where m is the number of neurons in the transmission layer, and c is very small. The calculations also show that for a practical number (up to 15000) of neurons, the Mutual Information reaches only approximately half of the information that is carried out by the input signal. The influence of noise on the transmission efficiency depending on the number of neurons was also analyzed. It turned out that the noise level at which transmission is optimal increases significantly with this number. Our results indicate that a large number of neurons in the network does not mean an essential improvement in transmission efficiency, but can contribute to reliability.

INDEX TERMS Shannon communication theory, neural network, network layer, transmission efficiency, mutual information, model of neuron, spike trains, information source, entropy.

I. INTRODUCTION

The human brain contains billions of neurons, linked to one another via hundreds of trillions of tiny contacts called synapses [1]. It is known that more than 80% of neurons are l small branch cells located in the cerebellum and have received only a few electrical impulses (spikes) from to 4-7 synapses, while the rest of the neurons have up to 200 000 connections [3]. Over the last two decades, significant progress has been made in explaining the evolution and role of brain size [2]. In this context, it is important to understand how the size of the network, and more specifically the number of neurons, affects the efficiency of information transmission.

Research on the impact and effects of scales on the functioning of neural networks has been conducted in several aspects. Neural activity at the microscopic level was modeled using phenomenological equations [4]. Schwalger and co-authors proposed a system of equations for several interacting populations at a mesoscopic scale, starting from a microscopic model of randomly connected generalized integrate-and-fire neuron models for networks varying between 50-2000 neurons [5]. In turn, in [6], it was shown that structural networks are a crucial component of the stochastic brain model on the mesoscopic scale. A metric called multiscale relevance (*MSR*) was proposed in [7] to capture the dynamic variability of the activity of single neurons across different scales. It was shown that neurons with a low *MSR* tend to have low Mutual Information, whereas neurons with a high *MSR* contain significant information on spatial

The associate editor coordinating the review of this manuscript and approving it for publication was Luca Barletta.

navigation and allow the decoding of the spatial position. In [8], it was suggested by exploiting the Hindmarsh-Rose neuron model that in small-scale networks information, is exchanged using temporal codes, while on a macroscopic scale, there would typically be pairs of neurons not directly connected because of the brain's sparsity, firing rate, and interspike-interval codes would be the most efficient codes.

The scale problems also constitute a challenge for the efficient implementation and performance of advanced networks. In [9], linear models were considered to analyze the memory consumption of the constituent components of neuronal simulators as a function of the network size and number of cores used. It was found out that as the network model sizes approach the regime of meso- and macroscale simulations, memory consumption on individual compute nodes became a critical bottleneck. Large-scale models of neuronal activity describing the activity of whole neural populations were considered in [10]. The authors combined simulations of brain network models with microscopically detailed spiking neuron network models. This approach enables the integration of different information sources and analysis of the biophysiological mechanisms in the network.

However, an important question regarding the role of the brain is its effectiveness in information processing. Two criteria are important to quantitatively measure this effectiveness. The first criterion simply maximizes the rate of information transmission, and the second criterion maximize information in relation to the energy used to transmit it. This study focuses on the first criterion. For the second criterion, it is important to estimate the energy costs. Attwell and Laughlin [11] analyzed the metabolic cost of different components of excitatory signaling and suggested that signaling-related energy consumption increases linearly with spiking frequency. More recently, detailed models and experimental results [12] have extended these calculations to the nonlinear regime. In [13], it was shown by direct mathematical analysis for the bi-stable neuron model that there exists an optimal number of neurons in the network where the average energy cost compared to the Mutual Information achieved per neuron passes through a global minimum.

The quantitative measurement of information requires the application of adequate mathematical tools. In general, there are two approaches to the quantitative analysis of information transfer processes, the Wiener and Shannon approaches [14]. The Shannon formulation differs from the Wiener approach in the nature of the transmitted signal and in the type of decision made at the receiver. In the Shannon model, a randomly generated message produced by a source of information is encoded, that is, each possible message that the source can produce is associated with a signal belonging to a specified set. On the other hand, in the Wiener model, a random signal communicated directly through the channel; the encoding step is absent. Furthermore, the channel model is essentially fixed. The channel is generally taken to be a device that adds to the input signal a randomly generated "noise". In Shannon Information Theory [15], [16], neural networks are treated as communication channels and the information transmitted is measured as the Mutual Information between stimuli and response signals [16], [19], [20], [21]. When studying information transmission processing, it is important to select both neural [23] and network architecture models [22]. In previous studies, we directly investigated the transmission for simple neuronal ring architectures composed of a few Levy-Baxter neurons [24] paying particular attention to the role of inhibitory neurons, long-range connections, and adaptation of neuronal networks to the presence of noise [18], [25]. This neuron model has a probabilistic character and exploits the binary representation of neuronal signals. Moreover, it contains all essential qualitative mechanisms participating in the transmission process, and provides results consistent with physiologically observed values [24].

In this study, we focus on the problem of the influence of the number of neurons in the network on transmission efficiency. We analyze both theoretically and numerically the Mutual Information between the input and output signals in the case of a simple class of neural networks with an increasing number of neurons. This type of analysis provides insight and intuition regarding complex situations. We present the results characterizing *MI* dependence on the size of the network as well as on the adopted parameters of the neurons. It is worth emphasizing that finding the maximum *MI* actually means finding the Shannon capacity of the transmission channel, which directly characterizes optimal decoding opportunities.

The remainder of this paper is organized as follows. In Section II, we briefly recall the following subsections: the basic concepts and notations of Shannon's Communication Theory, standard digitization of the spike trains, the neuron model used, and the assumed network architecture. Section III presents theoretical and numerical results. A discussion and concluding remarks are preseented in Section IV.

II. MATHEMATICAL BACKGROUNDS AND MODELS

In this Section, we present basic information regarding the fundamental concepts of Shannon's theory, neuronal signals digitization, neuron model used and assumed network architecture. We introduce notations based on mathematical formalism, but also refer to certain intuitions and a more accessible understanding of the process of information transfer in neural networks.

A. SHANNON COMMUNICATION THEORY

In this Section, we provide a brief overview of the basic concepts of Shannon's Communication Theory. The two fundamental concepts of this theory are entropy and Mutual Information (*MI*) between two random variables X and Z [16]. These concepts have been extensively used in many problems related to the application of learning methods that use neural networks to data classification problems [26].

Mutual Information can be expressed in terms of entropy as

$$MI(X;Z) := H(X) - H(X|Z) = H(X) + H(Z) - H(X,Z), \quad (1)$$

where $H(X|Z)$ is the entropy of X conditional on Z and $H(X,Z)$ is the joint entropy of X and Z [27], [28]. Clearly, $0 \leq MI(X;Z) \leq H(X)$. There is a lot of estimators of entropy and consequently also Mutual Information developed in the literature [29]. These estimators require large empirical data to be effective; in the case of a neural network and a neuron model, they would require the generation of very long strings of bits (output signals), and taking into account that in order to find the maximum MI, this generation operation would have to be repeated, it would be very computationally expensive. However, in practice, it is difficult to obtain reliable results. Therefore, in the following sections we present first theoretical and then numerical results. We derived analytical formulas for Mutual Information between input signals coming from an Information Source, such as Bernoulli processes and output signals. Then, we performed numerical simulations based on these formulae, as described in the next subsection.

The basic idea of Mutual Information (expressed bits) is to determine the reduction of uncertainty (measured by entropy) of random variable X provided that we know the values of discrete random variables Z. Maximal MI is linked with the channel capacity for a given communication channel through the Shannon Fundamental Theorem, which characterizes the optimal decoding schemes.

B. SPIKE TRAINS CODING

It is commonly known [17], [18], [30], [31], [32] that the carriers of information between neurons are electrical signals, specifically sequences of potential actions called spike-trains. Considering the physiological issues associated with the spike train appearance, each spike was detected with a limited time resolution. This led to the idea of representing spike trains by using a sequence of symbols. The binary digitalization of spike trains is the most natural and commonly used representation [30]. Because a spike train is observed with a limited time resolution δ, a spike is either present in each time bin (denoted by "1") or absent (assigned by "0"). Then, if we look at a time interval of length T, each spike train is represented by a binary sequence (additionally with some probability of occurrence). Mathematically, such sequences can be treated as a part of a trajectory of a stochastic process that can be analyzed from the point of view of the information they carry using Shannon's Information Theory [16].

C. NEURON MODEL APPLIED

In this study, we assume a probabilistic Levy-Baxter neuron model [24], (see Fig. 1). In general, it considers all essential qualitative mechanisms involved in the information transmission process and provides results that are consistent with the physiologically observed values. In this model, the synaptic noise s is a success rate parameter ($0 < s < 1$) that a spike will be transmitted through the synapse (this means that when the synaptic noise parameter s is equal to 1 there is no noise, while the noise increases when s is smaller), amplitude modulation Q_i is a random variable with uniform distribution in the interval [0,1] and activation threshold height g ($g > 0$), are the neuron model's parameters. The input to each neuron at a given moment in time, is a sequence of bits $X = [X^{(1)}, \ldots, X^{(n)}]$ where n denotes the number of synapses. The neuron output at each moment is a bit ("0" or "1"). To simplify the notation used in the original study by Levy and Baxter, we denote inputs E and I by X with an appropriate index.

Neuron acts in the following manner. Each binary input (block of n bits) to a given neuron is subject in synapses to quantal failures ϕ being a Bernoulli distributed random variable (with parameter s) and quantal amplitude modulation Q_i and which are summed to σ. This is the input to the spike generator $g(\sigma)$. A spike is generated if the magnitude of its excitation σ exceeds the assumed threshold g.

Because the type of stimuli coming from inhibitory neurons is an internal mechanism in the brain, while the aim of this study was to estimate the transmission of information from external sources (information provided by external stimuli), we assumed only external Sources of Information.

D. NETWORK ARCHITECTURE ASSUMED

To analyze how an increasing number of neurons in the network can affect the transmission efficiency, we directly considered a network with a simple architecture (see Figure 2) enabling the addition of subsequent neurons in the transmission layer (i.e., the Second Layer N) and having the option, after connecting a new neuron, to transmit the same input information. Thus, we assume that, at every moment in time, each neuron in the Second Layer N receives the same information represented by a block of bits. Therefore, the length of the input block (Input Layer X) is equal to the assumed number of synapses n for a single neuron. Each bit (component) in the input block appears at each subsequent moment with probability p when it is equal to "1" and with probability 1-p when it is equal to "0". Therefore, according to Shannon's terminology, the source of the information is an n-dimensional stochastic process. To introduce this notation, we now move on to a more formal description in mathematical language using Shannon's formalism. Let n denote the number of synapses for each neuron and m be the number of neurons in the network in the transmission layer. The Information Source is assumed to be an n-dimensional stochastic process, $X(t_i) = [X^{(1)}(t_i), \ldots, X^{(n)}(t_i)], i = 1, 2, \ldots,$ and $t_{i+1} - t_i = \Delta$, where Δ is the assumed time resolution. For example, one can assume that this information come from n neurons and is modeled by process X. The components of process $X^{(k)}(t_i), k = 1, 2, \ldots, n$ are considered independent Bernoulli processes with parameter p. This parameter can be

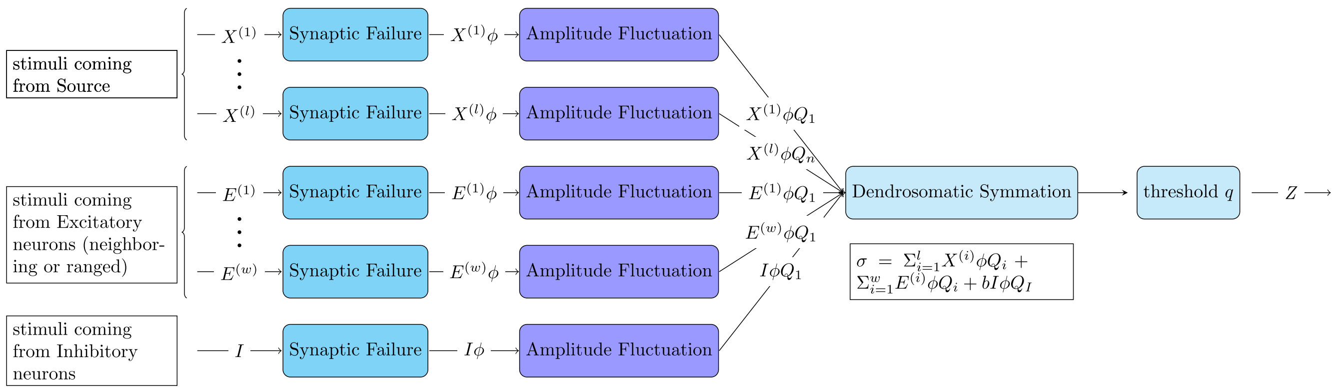

FIGURE 1. The scheme of the neuron model proposed by Levy and Baxter [24]. Additionally, similarly to [25] the type of inputs was also emphasized. Encoded stimuli is modeled by a discrete, binary stationary stochastic process with firing-rate (f_r) being the probability of a spike occuring and $1 - fr$ the probability of no spike. Parameters $E^{(i)}$ describe excitatory strength, b addresses inhibition strength, quantal failures ϕ is a random variable taking 0 for the input 0 and 1 with probability s for the input 1, amplitude fluctuations Q are implemented as random variables $U[0; 1]$ with uniform distribution function. The activation threshold is denoted by g. When σ is greater than g a "1" bit is generated, otherwise a "0" is output.

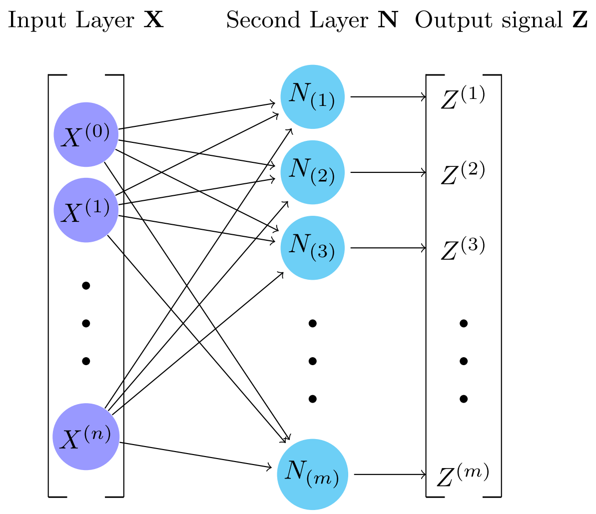

FIGURE 2. Architecture of the neural network under consideration. Each neurons $X^{(i)}$, $i = 1, 2, \ldots, n$ in the input layer X is supported by signals from other sources, for example, from earlier neurons. Each neuron $N^{(j)}$, $j = 1, 2, \ldots, m$ in the second layer N is supported by inputs from Information Source X (consisting of n neurons). Thus, we assumed that the neurons in the second layer had n synapses. In this study, the Mutual Information $MI(X; Z)$ between the information delivered by neurons from layer X and the information carried out by the output layer Z is evaluated and analyzed.

understood as the firing rate f_r of the input spike trains to a single synapse of a neuron located in the Second Layer N. The joint neuronal output from layer N of m neurons (which is a m-dimensional random variable) is denoted by $Z = [Z^{(1)}, \ldots, Z^{(m)}]$, where $Z^{(i)}$ is a binary random variable (see Fig. 2). Note that because $MI(X; Z) \leq H(X)$ and $H(X) \leq n$, consequently $MI(X; Z)$ must be less than the number n of synapses in a single neuron.

III. RESULTS

In this Section, we first present the theoretical and then numerical results. We derived analytical formulas for Mutual Information between input signals coming from an Information Sources, such as Bernoulli processes and output signals. Then, we performed numerical simulations based on these formulae, as described in the next subsection.

A. THEORETICAL RESULTS

Now, let us assume that x^k is the event that, at a given moment of time, k of specific components $X^{(i)}$ in X being inputs to neurons in layer N are equal to 1, and the other $n - k$ is equal to 0. Similarly, let z^j be the event that at a given moment in time, j of specific components $Z^{(i)}$ in Z being output from layer Z is equal to 1 and the other $m - j$ is equal to 0.

Because the random variables $X^{(i)}$, are independent, the probability of event x^k is

$$P(\mathbf{x^k}) = P(\mathbf{X} = \mathbf{x^k}) = f_r^k (1 - f_r)^{n-k}. \tag{2}$$

For each neuron from the output layer, the input spike can pass through the synapse with a success rate s. Next, the amplitudes of the transmitted signals are modulated by a random function Q with the uniform distributions on the interval $[0; 1]$. Thus, the conditional probability of the activation of a single neuron, provided that event $\mathbf{x^k}$ occurs, is

$$P(\mathbf{Z} = \mathbf{z^1} | \mathbf{X} = \mathbf{x^k}) = \sum_{i=0}^{k} \binom{k}{i} s^i (1 - s)^{k-i} P(iQ \geq g) \tag{3}$$

$$P(\mathbf{Z} = \mathbf{z^0} | \mathbf{X} = \mathbf{x^k}) = 1 - P(\mathbf{Z} = \mathbf{z^1} | \mathbf{x^k}) \tag{4}$$

where $P(iQ \geq g)$ denotes the probability that the sum iQ of i random variables of type Q reaches the activation threshold g and z^1 is the probability of activation of a single neuron. Because each Q is uniformly distributed and independent, the random variable iQ has an Irwin-Hall distribution [33]. Then, the probability $P(iQ \geq g)$ can be expressed as:

$$P(iQ \geq g) = 1 - P(iQ < g), \tag{5}$$

where the cumulative distribution function (CDF) is of the form [33]

$$P(iQ < g) = \frac{1}{i!}\sum_{h=0}^{|g|}(-1)^h \binom{i}{h}(g-h)^i. \tag{6}$$

Note that the calculation of probability (5) is directly obtained from the CDF of iQ at point g. By substituting (5) and (6) into (3) we obtain

$$P(\mathbf{Z} = \mathbf{z^1}|\mathbf{X} = \mathbf{x^k}) = \sum_{i=0}^{k}\binom{k}{i}s^i(1-s)^{k-i}(1 - \frac{1}{i!}\sum_{h=0}^{|g|}(-1)^h\binom{i}{h}(g-h)^i). \tag{7}$$

Since the components in the output $\mathbf{Z} = [Z^{(1)}, \ldots, Z^{(m)}]$ are independent, thus we have

$$P(\mathbf{Z} = \mathbf{z^j}|\mathbf{X} = \mathbf{x^k}) = (P(\mathbf{Z} = \mathbf{z^1}|\mathbf{X} = \mathbf{x^k}))^j(P(\mathbf{Z} = \mathbf{z^0}|\mathbf{X} = \mathbf{x^k}))^{m-j}, \tag{8}$$

$$P(\mathbf{Z} = \mathbf{z^j}, \mathbf{X} = \mathbf{x^k}) = P(\mathbf{x^k})P(\mathbf{Z} = \mathbf{z^j}|\mathbf{X} = \mathbf{x^k}) \tag{9}$$

$$P(\mathbf{Z} = \mathbf{z^j}) = \sum_{x\in X}P(\mathbf{x}, \mathbf{z^j}) = \sum_{k=0}^{n}\binom{n}{k}P(\mathbf{X} = \mathbf{x^k}, \mathbf{Z} = \mathbf{z^j}). \tag{10}$$

Thus, we can calculate all the components (i.e. corresponding entropies) that are needed to determine the Mutual Information $MI(\mathbf{X}; \mathbf{Z})$ in (1). The entropies are expressed as follows:

$$\begin{aligned} H(\mathbf{X}) &= H(\mathbf{X}^{(1)}) + H(\mathbf{X}^{(2)}) + \ldots + H(\mathbf{X}^{(n)}) \\ &= -n[f_r \log f_r + (1-f_r)\log(1-f_r)], \end{aligned} \tag{11}$$

$$\begin{aligned} H(\mathbf{Z}) &= -\sum_{z\in Z}P(\mathbf{z})\log P(\mathbf{z}) \\ &= -\sum_{j=0}^{m}\binom{m}{j}P(\mathbf{z^j})\log P(\mathbf{z^j}), \end{aligned} \tag{12}$$

$$\begin{aligned} H(\mathbf{Z}|\mathbf{X}) &= -\sum_{x\in X}\sum_{\mathbf{z}\in\mathbf{Z}}P(\mathbf{z}, \mathbf{x})\log P(\mathbf{z}|\mathbf{x}) \\ &= -\sum_{k=0}^{n}\binom{n}{k}P(\mathbf{x^k})\sum_{j=0}^{m}\binom{m}{j}P(\mathbf{z^j}|\mathbf{x^k})\log P(\mathbf{z^j}|\mathbf{x^k}), \end{aligned} \tag{13}$$

$$\begin{aligned} H(\mathbf{X}, \mathbf{Z}) &= -\sum_{x\in X}\sum_{\mathbf{z}\in\mathbf{Z}}P(\mathbf{x}, \mathbf{z})\log P(\mathbf{x}, \mathbf{z}) \\ &= -\sum_{k=0}^{n}\sum_{j=0}^{m}\binom{n}{k}\binom{m}{j}P(\mathbf{x^k}, \mathbf{z^j})\log P(\mathbf{x^k}, \mathbf{z^j}). \end{aligned} \tag{14}$$

Thus, the Mutual Information $MI(\mathbf{X}; \mathbf{Z})$ is of the form

$$\begin{aligned} MI(\mathbf{X}; \mathbf{Z}) &= -n[f_r \log f_r + (1-f_r)\log(1-f_r)] \\ &\quad + -\sum_{j=0}^{m}\binom{m}{j}P(\mathbf{z^j})\log P(\mathbf{z^j}) \\ &\quad + \sum_{k=0}^{n}\sum_{j=0}^{m}\binom{n}{k}\binom{m}{j}P(\mathbf{x^k}, \mathbf{z^j})\log P(\mathbf{x^k}, \mathbf{z^j}) \end{aligned} \tag{15}$$

and can be calculated by substituting equations (8), (9), and (10) into equation (15).

To summarize, in this subsection we have derived the formulas that allow the calculation of Mutual Information $MI(X; Z)$ between the input signal $\mathbf{X}$ and the output signal $\mathbf{Z}$ expressed in terms of the parameters of the information source (f_r), neuron (s, g) and the number of neurons m in the second layer N of the considered neural network.

In the next section, we apply (15) to find the Mutual Information between the input signals $\mathbf{X}$ and output signals $\mathbf{Z}$ for the considered networks with an increasing number of neurons m, and perform calculations for the full range of parameters characterizing a Levy-Baxter neuron and Information Source.

B. NUMERICAL RESULTS

We performed a numerical simulation to evaluate $MI(\mathbf{X}; \mathbf{Z})$ by exploiting formulas (8), (9), (10), and (15) developed in the previous section. The application of these formulas allows us to reduce the computational cost essentially and consequently allow us to evaluate $MI(X; Z)$ for a larger number of neurons, even up to $m = 60$. It is worth emphasizing that directly applying formula (1) to a reliable estimate of $MI(X; Z)$ would require estimating the probabilities needed to calculate the entropy $H(Z)$ and $H(X, Z)$, which, given the length of the Z strings of even $m = 20$ bits, would require generating very long strings and would be computationally very expensive. In addition, such calculations would have to be repeated for different parameters s, and f_r to calculate the maximum $MI(X; Z)$, which constitutes a significant additional computational cost. To find the maximal MI with satisfactory accuracy, we needed to go through the neuron parameter space of synaptic failure s as $0 < s < 1$ and through firing frequency f_r ($0 < f_r < 1$) with a relatively small step equal to 0,01. Because the Levy-Baxter model of a neuron has a probabilistic nature, simulating the input-output process for a single neuron requires the use of randomizing generators (working according to a given probability distribution). This implies that MI calculations can be performed successfully for a network containing up to several dozen of neurons. To perform the computations with a possibly large number of neurons m and taking into account the information concerning the number of synapses in [3]. we performed an analysis for neurons with five synapses. The threshold parameter g was assumed to be 5% of the maximal possible value that could be reached by a neuron

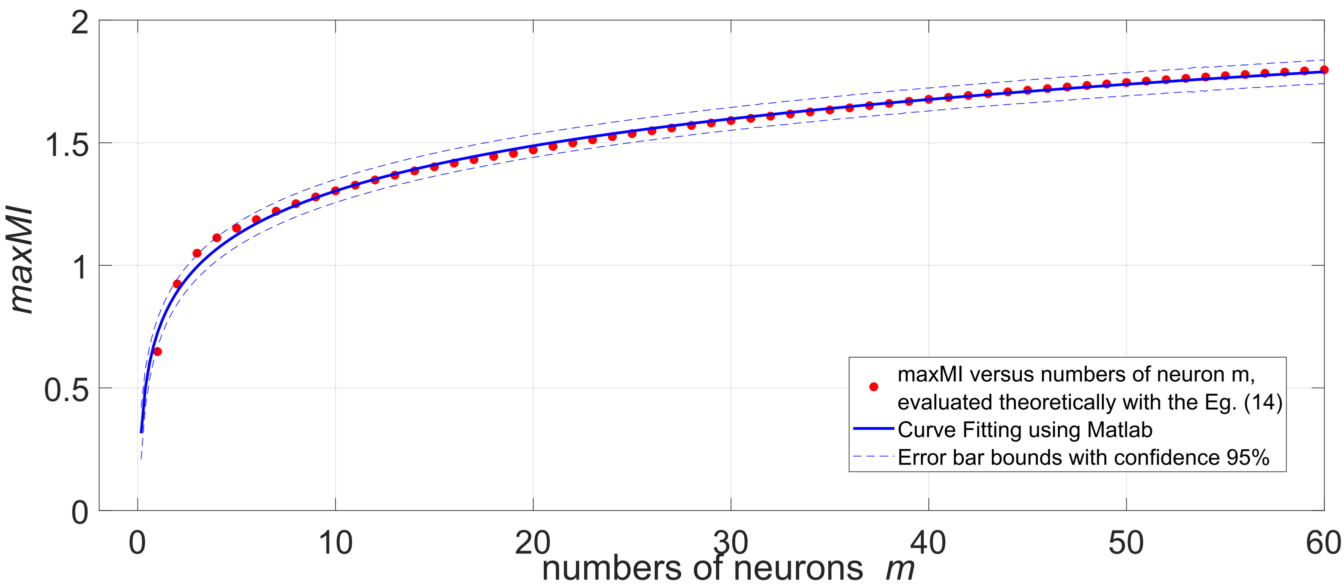

FIGURE 3. The influence of an increasing number (m) of neurons in the layer N on maximal Mutual Information $MI(\mathbf{X};\mathbf{Z})$. For each number m the maximum was taken over the source (i.e. layer X) parameter f_r and neuron parameter s being the success rate (i.e. probability s that spike will be successfully transmitted over a given synapse) in layer N. Activation threshold g is assumed to be 0.25. The calculations are presented for n=5, what means that the number of synapses of neurons in layer N is already 5. It turned out that the best fitting curve is of the form $f(m) = a - \frac{b}{m^c}$ ($a = 12.05, b = 11.37, c = 0,025$), where m is the number of neurons in the Layer N. The goodness of fitting according to MATLAB Curve Fitting Toolbox is: SSE : 0.02096, $RMSE$: 0.01917. The confidence bounds with 95% of confidence level are also depicted ($\pm 2SD$).

with five synapses. Despite these limitations, our results showed quantitative and qualitative behavior of maximal $MI(\mathbf{X};\mathbf{Z})$ as a function of the number of neurons (m). The results are presented in Fig. 3, 4 and Table 1.

In Fig. 3, the influence of increasing the number of neurons m on the maximal Mutual Information $MI(X;Z)$ for neural networks with the architecture presented in Fig. 2 is shown. We applied the MATLAB Curve Fitting Toolbox and found that the best fitting curve is the function $maxMI(m) = 12.05 - \frac{11.37}{m^{0.025}}$ with a goodness of fit $RMSE = 0.019$ and $SSE = 0.02$. Here, $RMSE$ and SSE are the root mean square error ($RMSE$) and summation of the square error (SSE), respectively. The smaller the SSE, the better the fitting. $MaxMI$ is the maximum value of $MI(X;Z)$ for all $0 < s < 1$ and $0 < f_r < 1$. This shows that $maxMI$ is asymptotically limited, and another important observation is that it increases very slowly starting from the number of m about 50 – 60. Moreover, because we assumed in our simulations that the number of synapses for each neuron is equal to five and that the input signals to each synapse come from the Bernoulli process, taking into account the classical inequality $MI(X;Z) \leq H(X)$, we have that $MI(X;Z)$ can be up to five. Therefore, we see that for a practical number of neurons m up to 15000 the value of $MI(X;Z) = f(m) \sim 2.57$ is only approximately half of the maximum possible value. This means that the uncertainty of correctly decoding the input signal given the output signal is, in this case, reduced in average by half.

In Fig. 4, Mutual Information $MI(\mathbf{X};\mathbf{Z}) = MI(f_r, s, g)$ as a function of the source parameter firing rate f_r and neuron parameter synaptic noise s for increasing

TABLE 1. Maximal Mutual Information (expressed in bits) for a selected number of neurons m (5, 10, 30, 60). The parameters s (the probability that an action potential will pass through the synapse) and f_r (probability of "1"/action potential occurrence in the input sequence) for which these maxima are achieved are also given. The threshold $g = 0,25$ is assumed 5% of the maximal possible σ which is equal to 5 (since the number of synapses to a given neuron is $n = 5$).

network size m	threshold g	$maxMI$	s	f_r
5	0.250	1.203	0.980	0.172
10	0.250	1.304	0.900	0.199
30	0.250	1.597	0.821	0.262
60	0.250	1.797	0.695	0.317

numbers m (5, 10, 30, 60) of neurons in layer **N** is presented. It can be observed that, with an increase in the network size m the values of $maxMI$ (i.e., the capacity of the transmission channel) are reached for smaller s and larger f_r values (see also Table 1). Because s is a parameter responsible for the level of noise in synapses, it is the probability of successful transmission of an action potential through the synapse; therefore, the larger s, the higher the success and the lower the noise. The results in Table 1 show that as the number of neurons increases, achieving maximum transmission efficiency is accompanied by a slight increase in noise (column "s"). We see that in the case of a network with a size of $m = 5$ neurons, the maximum Mutual Information is achieved for $s = 0.980$, while for $m = 10$ neurons the value of s for which this maximum is achieved is 0.9. Then, at $m = 30$ neurons, it increases the value of the most favorable noise in synapses reaching $s = 0.821$ and for $m = 60$ this value is $s = 0.695$. In turn, the observation that for a larger

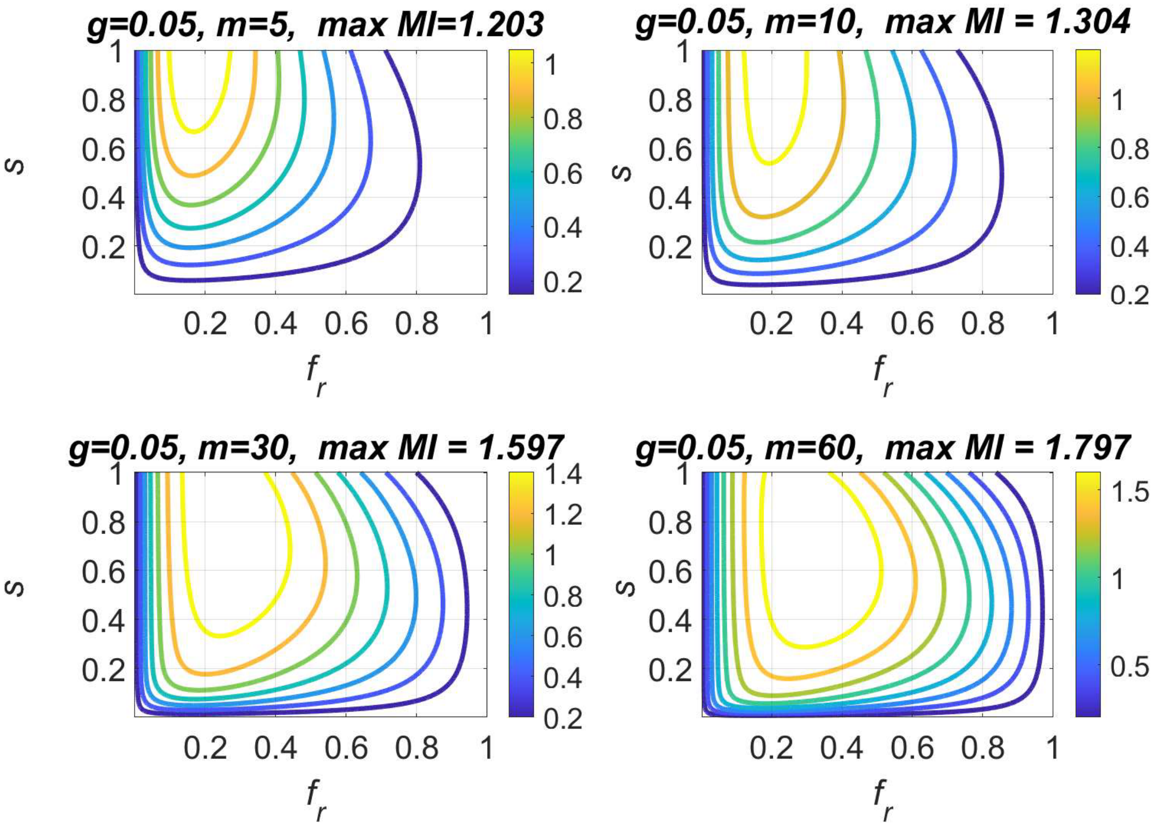

FIGURE 4. Mutual Information $MI(\mathbb{X};\mathbb{Z}) = MI(f_r, s, g)$ for neural networks shown in Figure 2 stimulated by a Bernoulli information source with firing rate f_r (horizontal axe) and synaptic noise s (vertical axe). The number m of neurons in the second layer N increases successively, $m = 5, 10, 30, 60$. The isolines with step 0.05 are depicted. One can see that the maximum MI is achieved for the synaptic noise s significantly less than 1. This is especially visible when the number of neurons m is larger.

network the maximum MI is achieved for larger f_r confirms that more energy must be used [11] to obtain a more efficient transmission with increasing size and, as compensation, this maximum is achieved for a more noisy channel (i.e. for a lower success rate s). This implies that a network with a larger number of neurons is more resistant to noise and consequently is more reliable.

To summarize the main numerical results, it was found that, for these neural networks, the maximum Mutual Information increased very slowly at the rate m^{-c}, with a small $c = 0.02473$, where m is the number of neurons. This indicates, among other things, that a further increase in the number of neurons does not mean a significant increase in transmission efficiency. Moreover, it was also noticed that the maximum transmission efficiency is achieved for a certain noise s in synapses, which shows that, in a sense, neural mechanisms can cope with this natural phenomenon in biological systems.

IV. DISCUSSION AND CONCLUSION

Mechanisms have been developed over the course of evolution to enable more efficient and reliable information processing. However, the key question is the impact of the size of the brain and thus the role of the number of neurons on the efficiency of these processes. Can such a performance be significantly improved by a simple increasing the size of the neural network. To provide insight into this problem, it seems natural to analyze these issues by examining the relevant models of the networks and neurons themselves [34], [35]. However, realistic models of real neural networks are analytically and computationally intractable. One of the major difficulties is the selection of the appropriate size and topology of these networks. Hunter et co-authors [36] discussed hot issues, including different learning algorithms, efficiency of different network topologies, and importance of choosing the proper size of neural networks. In [37] a new formalism that borrows from many-body statistical physics methods to analyze finite-size effects in spiking neural networks was introduced. Therefore, a very important question arises: does increasing the size of the network always lead to greater efficiency of information transfer? Another important question is whether increasing network noise always leads to a decrease in the effectiveness of information transfer.

Traditional mathematical approaches to analytically study the dynamics of neural networks rely on mean-field approximation, which is rigorously applicable only to infinite-sized networks [38]. However, all existing biological networks consist of a finite number of neurons, often consisting of only a few dozen neurons, such as the microscopic circuits in invertebrates. Therefore, it is important to extend our ability to analytically study neural dynamics in small networks. At present, systematic analytical solutions for the dynamics of finite-size neural networks require further analysis. In this study, to provide insight into the impact of network size on the efficiency of information transmission and noise influence, we considered the case of fully connected networks consisting of Levy-Baxter neurons with a single hidden layer with an increasing number of neurons. We treated this problem quantitatively using the Shannon approach, and analyzed the Mutual Information between the input and output signals. The L-B neuron, which exhibits the basic properties of a biological neuron, is described in probabilistic language, which allows us to derive analytical formulas for *MI* expressed in terms of the size of the network and the parameters of the neuron. Numerical results obtained using these formulas have shown that for a practical number of neurons for which the input signal can reach a given moment (up to 15000), the Mutual Information between the input and output signals is approximately 50% of the maximum possible information that can be achieved. It is also important to note that the increase in *MI* is very slow as the number of neurons increases. The results of our study concerning the influence of network size are consistent with those obtained in [13]. In this study, the effectiveness of information transmission with an increasing number of neurons for a network architecture similar to that used in our study was examined. The authors used a bistable neuron model and consider the information carried by the signal in such a network in relation to its metabolic cost. More precisely, they consider the $\frac{E}{MI}$ quotient. It was demonstrated that there exists an optimal number of neurons in this model for which the energy cost per Mutual Information passes through a global minimum. Our work shows that Mutual Information, starting from a certain level of the number of neurons, grows very slowly, and it is known that the energy consumption increases with the number of neurons practically linearly [11], [39], or even exponentially [12], in which the $\frac{MI}{E}$ indicator passes through a global maximum (because we consider the inverse quotient) for an increasing number of neurons, which is consistent with the results in [13].

In turn, in the case of the influence of noise on the transmission efficiency, our results are consistent with the results obtained in the works of [40], [41], [42], and [43], which also found that a certain level of noise in the considered system (neural networks, signal recovery through an array of saturating sensors, estimator design) may be beneficial in the context of information processing. This means that since the maximum transmission rate is reached as the number of neurons increases, for higher and higher noise values s, it shows that larger networks are more reliable.

To summarize, the results of our paper show that large number of neurons in actual biological networks (brain) is related mostly to the fact that individual areas in the brain are dedicated to different types of stimuli and it is rather due to the tendency to achieve reliability and noise immunity rather than a significant increase in the information performance.

Studies related to the presented results regarding the saturation of the information transmission rate as a function of the number of neurons and the impact of noise on the transmission efficiency are the subject of further research. It is particularly important to clarify these issues in brain-inspired networks (such as spiking neural networks or large-scale brain networks) with more advanced architectures and with the application of a variety of neuron models starting from physiological neuron such as integrate-and fire models based on the insight into pulsating electrical activity or from the biophysical Hodgkin–Huxley like models that describe ion channels.

REFERENCES

[1] J. L. van Hemmen and T. Sejnowski, "How have brains evolved?" "How is the cerebral cortex organized?" in *Problems in Systems Neurosciences*. London, U.K.: Oxford Univ. Press, 2006, ch. 1, pp. 1–132.
[2] G. A. Ascoli, B.-X. Huo, and P. P. Mitra, "Sizing up whole-brain neuronal tracing," *Sci. Bull.*, vol. 67, no. 9, pp. 883–884, May 2022.
[3] J. W. Shepherd, J. B. Deregowski, and H. D. Ellis, "A cross-cultural study of recognition memory for faces," *Int. J. Psychol.*, vol. 9, no. 3, pp. 205–212, Jan. 1974.
[4] W. Gerstner, H. Sprekeler, and G. Deco, "Theory and simulation in neuroscience," *Science*, vol. 338, no. 6103, pp. 60–65, Oct. 2012.
[5] T. Schwalger, M. Deger, and W. Gerstner, "Towards a theory of cortical columns: From spiking neurons to interacting neural populations of finite size," *PLOS Comput. Biol.*, vol. 13, no. 4, Apr. 2017, Art. no. e1005507.
[6] G. Barzon, G. Nicoletti, B. Mariani, M. Formentin, and S. Suweis, "Criticality and network structure drive emergent oscillations in a stochastic whole-brain model," *J. Phys., Complex.*, vol. 3, no. 2, Jun. 2022, Art. no. 025010.
[7] R. J. Cubero, M. Marsili, and Y. Roudi, "Multiscale relevance and informative encoding in neuronal spike trains," *J. Comput. Neurosci.*, vol. 48, no. 1, pp. 85–102, Feb. 2020.
[8] C. G. Antonopoulos, E. Bianco-Martinez, and M. S. Baptista, "Evaluating performance of neural codes in model neural communication networks," *Neural Netw.*, vol. 109, pp. 90–102, Jan. 2019.
[9] S. Kunkel, T. C. Potjans, J. M. Eppler, H. E. Plesser, A. Morrison, and M. Diesmann, "Meeting the memory challenges of brain-scale network simulation," *Frontiers Neuroinform.*, vol. 5, no. 35, pp. 1325–1331, Jan. 2012.
[10] P. Ritter, "Multi-scale personalized brain modeling," *Biophysical J.*, vol. 121, no. 3, p. 28a, Feb. 2022.
[11] D. Attwell and S. B. Laughlin, "An energy budget for signaling in the grey matter of the brain," *J. Cerebral Blood Flow Metabolism*, vol. 21, no. 10, pp. 1133–1145, Oct. 2001.
[12] J. J. Harris, R. Jolivet, E. Engl, and D. Attwell, "Energy-efficient information transfer by visual pathway synapses," *Current Biol.*, vol. 25, no. 24, pp. 3151–3160, Dec. 2015.
[13] L. Yu, C. Zhang, L. Liu, and Y. Yu, "Energy-efficient population coding constrains network size of a neuronal array system," *Sci. Rep.*, vol. 6, no. 1, pp. 1–8, Jan. 2016.
[14] A. Ephremides, "How information theory changed the world—A brief review of the history of the information theory society," in *Proc. IEEE Conf. Hist. Tech. Societies*, Aug. 2009, pp. 1–7.
[15] T. M. Cover and J. A. Thomas, "Entropy, relative entropy, and mutual information, 'Channel capacity,"' in *Elements of Information Theory*. New York, NY, USA: Wiley, 1991, ch. 2, pp. 13–241.
[16] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Jul. 1948.
[17] M. Diesmann, M.-O. Gewaltig, and A. Aertsen, "Stable propagation of synchronous spiking in cortical neural networks," *Nature*, vol. 402, no. 6761, pp. 529–533, Dec. 1999.

[18] A. Pregowska, E. Kaplan, and J. Szczepanski, ''How far can neural correlations reduce uncertainty? Comparison of information transmission rates for Markov and Bernoulli processes,'' *Int. J. Neural Syst.*, vol. 29, no. 08, Oct. 2019, Art. no. 1950003.

[19] H. Awan, R. S. Adve, N. Wallbridge, C. Plummer, and A. W. Eckford, ''Information theoretic based comparative analysis of different communication signals in plants,'' *IEEE Access*, vol. 7, pp. 117075–117087, 2019.

[20] D. Sengupta, P. Gupta, and A. Biswas, ''A survey on mutual information based medical image registration algorithms,'' *Neurocomputing*, vol. 486, pp. 174–188, May 2022.

[21] D. Wen, R. Li, M. Jiang, J. Li, Y. Liu, X. Dong, M. I. Saripan, H. Song, W. Han, and Y. Zhou, ''Multi-dimensional conditional mutual information with application on the EEG signal analysis for spatial cognitive ability evaluation,'' *Neural Netw.*, vol. 148, pp. 23–36, Apr. 2022.

[22] H.-J. Park and K. Friston, ''Structural and functional brain networks: From connections to cognition,'' *Science*, vol. 342, no. 6158, Nov. 2013, Art. no. 1238411.

[23] W. Gerstner and R. Naud, ''How good are neuron models?'' *Science*, vol. 326, no. 5951, pp. 379–380, Oct. 2009.

[24] W. B. Levy and R. A. Baxter, ''Energy-efficient neuronal computation via quantal synaptic failures,'' *J. Neurosci.*, vol. 22, no. 11, pp. 4746–4755, Jun. 2002.

[25] B. Paprocki and J. Szczepanski, ''Transmission efficiency in ring, brain inspired neuronal networks. Information and energetic aspects,'' *Brain Res.*, vol. 1536, pp. 135–143, Nov. 2013.

[26] D. U. Jo, S. Yun, and J. Y. Choi, ''How much a model be trained by passive learning before active learning?'' *IEEE Access*, vol. 10, pp. 34677–34689, 2022.

[27] B. Salafian, E. F. Ben-Knaan, N. Shlezinger, S. De Ribaupierre, and N. Farsad, ''MICAL: Mutual information-based CNN-aided learned factor graphs for seizure detection from EEG signals,'' *IEEE Access*, vol. 11, pp. 23085–23096, 2023.

[28] M. Bayram and M. A. Arserim, ''Analysis of epileptic iEEG data by applying convolutional neural networks to low-frequency scalograms,'' *IEEE Access*, vol. 9, pp. 162520–162529, 2021.

[29] S. Verdú, ''Empirical estimation of information measures: A literature guide,'' *Entropy*, vol. 21, no. 8, p. 720, Jul. 2019.

[30] F. Rieke, D. D. Warland, R. R. de Ruyter van Steveninck, and W. Bialek, ''Introduction,'' ''Foundations,''' in *Spikes: Exploring the Neural Code*. Cambridge, MA, USA: MIT Press, 1997, ch. 1 and 2, pp. 1–20.

[31] D. M. MacKay and W. S. McCulloch, ''The limiting information capacity of a neuronal link,'' *Bull. Math. Biophys.*, vol. 14, no. 2, pp. 127–135, Jun. 1952.

[32] A. Pregowska, J. Szczepanski, and E. Wajnryb, ''Temporal code versus rate code for binary information sources,'' *Neurocomputing*, vol. 216, pp. 756–762, Dec. 2016.

[33] P. Hall, ''The distribution of means for samples of size N drawn from a population in which the variate takes values between 0 and 1, all such values being equally probable,'' *Biometrika*, vol. 19, nos. 3–4, pp. 240–245, Dec. 1927.

[34] P. H. E. Tiesinga, J.-M. Fellous, J. V. José, and T. J. Sejnowski, ''Optimal information transfer in synchronized neocortical neurons,'' *Neurocomputing*, vols. 38–40, pp. 397–402, Jun. 2001.

[35] J. Kitazono, R. Kanai, and M. Oizumi, ''Efficient search for informational cores in complex systems: Application to brain networks,'' *Neural Netw.*, vol. 132, pp. 232–244, Dec. 2020.

[36] D. Hunter, H. Yu, M. S. PukishIII, J. Kolbusz, and B. M. Wilamowski, ''Selection of proper neural network sizes and architectures—A comparative study,'' *IEEE Trans. Ind. Informat.*, vol. 8, no. 2, pp. 228–240, May 2012.

[37] M. A. Buice and C. C. Chow, ''Dynamic finite size effects in spiking neural networks,'' *PLoS Comput. Biol.*, vol. 9, no. 1, Jan. 2013, Art. no. e1002872.

[38] D. Fasoli, A. Cattani, and S. Panzeri, ''Transitions between asynchronous and synchronous states: A theory of correlations in small neural circuits,'' *J. Comput. Neurosci.*, vol. 44, no. 1, pp. 25–43, Feb. 2018.

[39] A. Hasenstaub, S. Otte, E. Callaway, and T. J. Sejnowski, ''Metabolic cost as a unifying principle governing neuronal biophysics,'' *Proc. Nat. Acad. Sci. USA*, vol. 107, no. 27, pp. 12329–12334, Jul. 2010.

[40] L. Huan and W. You-Guo, ''Noise-enhanced information transmission of a non-linear multilevel threshold neural networks system,'' *Acta Phys. Sinica*, vol. 63, no. 12, 2014, Art. no. 120506.

[41] L. Xu, T. Vladusich, F. Duan, L. J. Gunn, D. Abbott, and M. D. McDonnell, ''Decoding suprathreshold stochastic resonance with optimal weight,'' *Phys. Lett. A*, vol. 379, no. 38, pp. 2277–2283, Oct. 2015.

[42] F. Duan, Y. Pan, F. Chapeau-Blondeau, and D. Abbott, ''Noise benefits in combined nonlinear Bayesian estimators,'' *IEEE Trans. Signal Process.*, vol. 67, no. 17, pp. 4611–4623, Sep. 2019.

[43] T. T. Xie, Y. D. Ji, Z. S. Yang, F. B. Duan, and D. Abbott, ''Decoding suprathreshold stochastic resonance with optimal weight,'' *Chaos, Solitons Fractals*, vol. 166, Jan. 2023, Art. no. 112887.

BARTOSZ PAPROCKI received the M.Sc. degree in computer science from the Faculty of Mathematics and Computer Science, Nicolaus Copernicus University, Toruń, Poland, in 2008, and the Ph.D. degree from the Institute of Fundamental Technological Research, Polish Academy of Sciences, Warsaw, in 2015. Currently, he is with the Institute of Mechanics and Applied Computer Science, Kazimierz Wielki University, Bydgoszcz, Poland. His research interests include neural modeling, neural computation, and information theory.

AGNIESZKA PREGOWSKA received the M.Sc. (Eng.) degree in the application of dynamical systems methods in mechanics and machine design from the Faculty of Mechatronics, Warsaw University of Technology, in 2008, the Ph.D. degree in developing control and optimization methods to the study of mechanical systems dynamics from the Polish Academy of Sciences, Warsaw, in 2013, and the D.Sc. (Habilitation) degree in information and communication technology (in the field of neuroinformatics), in 2023. Currently, she is an Assistant Professor with the Institute of Fundamental Technological Research, Polish Academy of Sciences. Her scientific interests focus on the application of artificial intelligence, information theory and computational methods in neuroinformatics, and biomedical signals processing and classification. She is also involved in the applications of extended reality in medicine.

JANUSZ SZCZEPANSKI received the M.Sc. degree in mathematics from Warsaw University and the Ph.D. degree in applied mathematics and the Habilitation degree (Hons.) in computer sciences from the Polish Academy of Sciences (PAS). In 2014, he received the title of Full Professor from the President of the Republic of Poland. In 2004, he was a Visiting Scholar with the University of California at San Diego. From 2000 to 2003, he was a Visiting Scientist with Miguel Hernandez University, Elche, Spain. He is currently a Professor with the Institute of Fundamental Technological Research (IFTR), PAS. His research interests include information theory, neuroscience, the application of dynamical systems, stochastic processes to biological systems, and cryptography. He is a member of the Scientific Council and the Deputy Director for Research with IFTR.