Dynamic Image Quantization Using Leaky Integrate-and-Fire Neurons

Effrosyni Doutsi, *Member, IEEE*, Lionel Fillatre, Marc Antonini, *Member, IEEE*, and Panagiotis Tsakalides, *Member, IEEE*

***Abstract*— This paper introduces a novel coding/decoding mechanism that mimics one of the most important properties of the human visual system: its ability to enhance the visual perception quality in time. In other words, the brain takes advantage of time to process and clarify the details of the visual scene. This characteristic is yet to be considered by the state-of-the-art quantization mechanisms that process the visual information regardless the duration of time it appears in the visual scene. We propose a compression architecture built of neuroscience models; it first uses the leaky integrate-and-fire (LIF) model to transform the visual stimulus into a spike train and then it combines two different kinds of spike interpretation mechanisms (SIM), the time-SIM and the rate-SIM for the encoding of the spike train. The time-SIM allows a high quality interpretation of the neural code and the rate-SIM allows a simple decoding mechanism by counting the spikes. For that reason, the proposed mechanisms is called Dual-SIM quantizer (Dual-SIMQ). We show that (i) the time-dependency of Dual-SIMQ automatically controls the reconstruction accuracy of the visual stimulus, (ii) the numerical comparison of Dual-SIMQ to the state-of-the-art shows that the performance of the proposed algorithm is similar to the uniform quantization schema while it approximates the optimal behavior of the non-uniform quantization schema and (iii) from the perceptual point of view the reconstruction quality using the Dual-SIMQ is higher than the state-of-the-art.**

***Index Terms*— Uniform quantization, non-uniform quantization, leaky integrate-and-fire model, spikes, rate coding, time coding.**

I. Introduction

COMPRESSION is undoubtedly considered as one of the most important and necessary processing steps in image communication. Images are highly correlated signals as they consist of a lot of redundancy. A lot of effort has been deployed to justify how to efficiently eliminate this redundancy while ensuring high reconstruction quality (lossy compression). The definition of redundancy is often associated with the sensitivity of the human visual system (HVS) to specific spatiotemporal frequencies. Thus, understanding and modeling the visual perception seems to be very beneficial to the progress of compression algorithms [1], [2].

The core of the state-of-the-art lossy compression algorithms is *quantization*. The quantization objective is to figure out the best possible way to map a range of values from the input signal into a single quantum value. There are two different kinds of quantization; the uniform and the non-uniform. Both methods are able to process the input signal at once and achieve only one reconstruction quality depending on their parameters. However, this single-step approach is in contrast to the HVS where *time* has a key role in the quality refinement of the reconstructed images. In other words, if we neglect the saccadic eye movements, the HVS takes advantage of the duration the input signal is available to enhance the quality of the perceived image.

In this work, we are interested in adopting into a compression architecture, the aforementioned time-dependent capacity of the HVS in order to improve the perception quality of the reconstructed signal in time. We propose a novel coding/decoding architecture inspired by the neurons which are the main processing units of the HVS. In the literature, there are plenty of *spike generation mechanisms (SGM)* which approximate the way the neurons transform a constant positive input stimulus I into a sequence of $N \in \mathbb{N}^+$ discrete events called a spike train. Each discrete event, namely a spike, is generated if the input intensity is stronger than a threshold θ, otherwise the neuron remains silent. The spikes are treated as identical stereotype events, because their shape does not seem to carry any information. Rather, it is the number of spikes and/or the spike arrival times which matter [3].

During the last decade, the neural spiking mechanisms have attracted the interest of the signal processing society. The most challenging part of using these mechanisms is to find out the best *spike interpretation mechanism (SIM)* which allows us to use the code of spikes and reconstruct the highest quality input signal (Fig. 1), since the brain uses the code of spikes to learn, analyze, and take decisions instead of reproducing the input stimulus. However, there are several architectures that use neural models in order to encode signals using spikes such as the rank order coders (ROC) [4], [5], time encoding machines [6] and asynchronous pulse sigma-delta modulators

Manuscript received May 8, 2020; revised December 1, 2020; accepted March 19, 2021. Date of publication April 9, 2021; date of current version April 16, 2021. This work was supported in part by the Hellenic Foundation for Research and Innovation (HFRI) and the General Secretariat for Research and Technology (GSRT) through the HFRI Faculty under Grant 1725 and in part by the Stavros Niarchos Foundation within the framework of the Project Advancing Young Researchers' Human Capital in Cutting Edge Technologies in the Preservation of Cultural Heritage and the Tackling of Societal Challenges (ARCHERS). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Yun He. *(Corresponding author: Effrosyni Doutsi.)*

Effrosyni Doutsi is with the Foundation for Research and Technology-Hellas (FORTH), 711 10 Heraklion, Greece, and also with the Laboratoire d'Informatique, Signaux et Systèmes de Sophia Antipolis (I3S Laboratory), CNRS, Université Côte d'Azur, 06108 Nice, France (e-mail: edoutsi@ics.forth.gr).

Lionel Fillatre and Marc Antonini are with the Laboratoire d'Informatique, Signaux et Systèmes de Sophia Antipolis (I3S Laboratory), CNRS, Université Côte d'Azur, 06108 Nice, France (e-mail: lionel.fillatre@i3s.unice.fr; am@i3s.unice.fr).

Panagiotis Tsakalides is with the Department of Computer Science, University of Crete, 700 13 Heraklion, Greece, and also with Foundation for Research and Technology-Hellas (FORTH), 711 10 Heraklion, Greece (e-mail: tsakalid@ics.forth.gr).

Digital Object Identifier 10.1109/TIP.2021.3070193

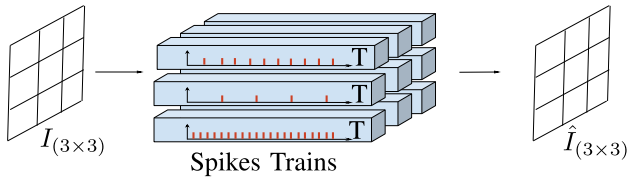


Fig. 1. General framework of the proposed architecture. A 3×3 input image I is fed to a group of 9 neurons (in blue). Each neuron is associated and excited by an individual pixel. After a given observation window T the spike trains are used to reconstruct an approximation of the input image $\hat{I}$.

(APSDM) [7]. The importance of the time parameter was mentioned for the first time in [4] where the authors show that when the observation window is long, the number of spikes that participate in the reconstruction process is higher, resulting in a better quality.

The main contributions of this paper are the following. First, the paper introduces a novel quantizer, namely the *Dual-SIM Quantizer (Dual-SIMQ)*, which is based on two complementary aspects of SIM: i) the input value is converted into a sequence of spikes by using a time-encoding and ii) it is reconstructed by using a rate-decoding which counts the spikes. It is shown that the combination of time-encoding and rate-decoding leads to a natural quantization of the input value. Second, the maximum number of spikes is controlled by a given observation duration $T > 0$. The duration T is interpreted as the maximum time period which is allowed to encode and decode the spike train. The behavior of the quantizer depends of the parameter T. Hence, this time constraint generates a dynamic quantizer whose behavior evolves in time. The dynamic properties of the Dual-SIMQ give rise to a ground-breaking compression system that permits a time-dependent quality refinement of the reconstructed signal. This is a great breakthrough compared to the conventional quantizers which process the input stimulus in a single-step without taking advantage of the observation duration. It is also mathematically proven that this novel neuro-inspired mechanism performs as a uniform or non-uniform quantizer by choosing adequately some of its parameters. Last but not least, this paper shows that the Dual-SIMQ outperforms the capacity of a uniform scalar quantizer (USQ) without deadzone, it coincides with the performance of a USQ with deadzone and it approximates the optimal Lloyd-Max quantizer (LQ).

Section II describes the principle of the neuro-inspired quantization based on spike trains. Section III is an overview of the spike generation and interpretation mechanisms, focusing especially on the well-known leaky integrate-and-fire (LIF) model. Section IV presents our main contribution introducing how to combine the rate-SIM and time-SIM to derive the Dual-SIMQ. Numerical results on simulated data and real data are presented in Section V. A concise discussion and the conclusion of this work are drawn in section VI.

II. Principle of Neuro-Inspired Quantization

This section briefly recalls the main concepts in quantization and rate-distortion theory and how they are related to the neuro-inspired quantization.

A. Basics of Quantization

Let I be a real random variable with the probability density function (pdf) $p(I)$ and let the representation of I be denoted as $\hat{I}$. If we are given r bits to represent I, the value $\hat{I}$ can take on 2^r values. The general problem of quantization is to find the optimum set of values for $\hat{I}$, called the code points $\hat{I}_1$, $\hat{I}_2$, …and the regions $S_1, S_2, \ldots$, that are associated with each code point.

According to quantization theory [8], a 2^r-rate distortion code consists of an encoding function,

$$f : \mathbb{R} \to \mathcal{C}, \tag{1}$$

where $\mathcal{C}$ is a subset with 2^r elements of the set of all integers $\mathbb{Z}$ and a decoding function

$$g : \mathcal{C} \to \mathbb{R}. \tag{2}$$

The encoding function defines a partition $\{S_1, \ldots, S_{2^r}\}$ of $\mathbb{R}$ such that $S_i \cap S_j = \emptyset$ for all $i \neq j$ and $\cup_{m=1}^{2^r} S_m = \mathbb{R}$. The interval S_m is called the m-th quantization region and it is defined such that $f(I)$ is constant for all $I \in S_m$ and $g(f(I)) = \hat{I}_m$ for all $I \in S_m$.

The quantization can be uniform or non-uniform [9], [10]. A uniform quantizer is recommended when the input source is either uniformly or non-uniformly distributed. In the latter case, it is mandatory that the quantizer is followed by an entropy coder where the statistics of the input source are taken into consideration. Otherwise, it is better to calculate some non-uniform quantization regions such that finer regions are associated to more likely values. This paper does not make any assumption on the probability distribution of the input source.

Most uniform quantizers for signed input value can be classified as being of one of two types: mid-rise and mid-tread. A typical mid-rise uniform scalar quantizer with a quantization step size $q > 0$ can be expressed as

$$Q_q(x) = q \times \left(\left\lfloor \frac{x}{q} \right\rfloor + \frac{1}{2} \right), \tag{3}$$

where $\lfloor x \rfloor$ corresponds to the greatest integer less than or equal to x. For simplicity, it is assumed that $\mathcal{C} = \mathbb{Z}$ is not finite.

The definition of the mid-tread uniform scalar quantizer with deadzone $\lambda > 0$ is given by:

$$Q_{q,\lambda}(x) = \text{sgn}(x) \max \left(0, \left\lfloor \frac{|x| - \lambda/2}{q} + 1 \right\rfloor \right) \times q, \tag{4}$$

where $\text{sgn}(I)$ denotes the sign of I: $\text{sgn}(I) = 1$ if $I > 0$, $\text{sgn}(I) = -1$ if $I < 0$ and $\text{sgn}(0) = 0$. The zero output of the quantizer is the interval $\left[-\frac{\lambda}{2}, \frac{\lambda}{2}\right]$ called the deadzone. The standard mid-tread quantizer corresponds to $\lambda = q$.

It has been proven that given a certain statistical distribution of the signal it is possible to compute the best partition that minimizes the power of noise using the non-uniform Lloyd-Max quantizer which is explicitly described in [11], [12].

B. Neuro-Inspired Quantization

The neuro-inspired quantizer proposed in this paper encodes the input value as a spike train and it exploits this spike train

to estimate $\hat{I}$. More formally, it is assumed that the input value I takes the form of a constant signal

$$I(t) = I\,\mathbf{1}_{[0\le t\le T]}(t), \tag{5}$$

for a given duration $T > 0$ where $\mathbf{1}$ is the indicator function which equals 1 if $0 \le t \le T$, and 0 otherwise. Without any loss of generality, it is assumed that $I \ge 0$. It will be mathematically defined later on (see section IV-B), that if I is real, the sign of I will be coded separately with a dedicated single bit. Here, the encoding function is a function $f(I)$ which transforms the signal $I(t)$ into a spike train

$$f : I \mapsto \left\{t^1(I), t^2(I), \ldots, t^{N(I)}(I)\right\} = \left\{t^1, t^2, \ldots, t^N\right\}$$

of $N = N(I)$ increasing positive time values $t^j = t^j(I)$ depending on I. The decoding function $g(\cdot)$ transforms the spike train $f(I)$ in an estimated real value $\hat{I} = g(f(I))$. In the rest of the paper, the symbol I is omitted in the spike times t^j and N in order to simplify the notations. The duration T acts as a parameter to control the number N of spikes.

C. Basics of Rate Distortion Theory

The distortion of a quantizer can be measured by the Mean Squared Error (MSE):

$$D = \mathrm{MSE}(I, \hat{I}) = \sum_{m=1}^{L} \int_{S_m} (I - \hat{I}_m)^2 p(I)\,\mathrm{d}I, \tag{6}$$

where L is the number of the quantization layers. In the case of a high resolution uniform scalar quantizer, when q is small (or equivalently L sufficiently large), assuming that the pdf $p(I)$ is smooth enough, it is shown in [9] that

$$D \approx \frac{q^2}{12}. \tag{7}$$

The rate r is given in bits/symbol (here, a symbol corresponds to a codeword) by the entropy of the codewords:

$$r = -\sum_{m=1}^{L} p_m \log_2 p_m, \tag{8}$$

where $p_m = \int_{S_m} p(I)\,\mathrm{d}I$ is the probability of the quantization interval S_m. We also get from [13] that:

$$r \approx H(I) - \log_2 q, \tag{9}$$

where $H(I) = \int_{-\infty}^{+\infty} p(I) \log_2 p(I)\,\mathrm{d}I$ is the Shannon entropy. Finally, we obtain the famous rate-distortion approximation

$$D = D(r) \approx \frac{1}{12} 2^{2H(I)} 2^{-2r}, \tag{10}$$

which gives the optimum value of D for a given rate r in the case of a high resolution uniform quantizer.

III. Spike Generation and Interpretation

This section describes usual mechanisms to transform a signal into a spike train and to recover it from the spike train.

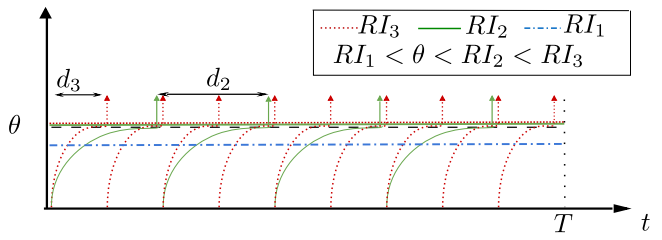


Fig. 2. LIF model with observation window T and threshold θ. If the intensity I satisfies $RI > \theta$, the neuron spikes (case $I \in \{I_2, I_3\}$), otherwise it remains silent (case $I = I_1$).

A. Spike Generation Mechanism (SGM)

In the literature, there are several models which approximate the neural activation. Hodgkin and Huxley [14] reproduced the neural activity with high accuracy deriving a set of four nonlinear differential equations which approximate the neural behavior with a lot of details at the level of ion channels. However, these equations are difficult to manipulate. A possible reduction of these equations leads to either a system of two-dimensions [15]–[18] or the Spike Response Model (SRM) [3], [19], [20]. On the one hand, the advantage of the two-dimension simplification is the plane analysis of the neural behavior. On the other hand, based on the SRM, it is proven in [19] that the Hodgkin-Huxley equations can be approximated by the simpler Leaky Integrate-and-Fire (LIF) model [3].

The well-known LIF model is simple [3]. It approximates the neuronal encoding process by a first order differential equation derived from a resistor-capacitor circuit:

$$I(t) = \frac{u(t)}{R} + C\frac{du}{dt}(t), \tag{11}$$

where $I(t)$ is the input signal, C is the capacitance, R is the resistance and $u(t)$ is the voltage across the resistor. The voltage $u(t)$ models the membrane potential of a neuron. It is assumed that $u(t = t^{(k)}) = 0$ mV after the emission of a spike at time $t^{(k)}$, $k \ge 1$, with the convention that $t^{(0)} = 0$ ms.

The solution $u_k(t)$ of the differential equation (11) for the constant signal $I(t)$ in (5) after the emission of the k-th spike at time $t^{(k)}$ is given by:

$$u_k(t) = RI\left[1 - \exp\left(-\frac{t - t^{(k)}}{\tau}\right)\right], \quad \forall t \ge t^{(k)}, \tag{12}$$

where $\tau = RC$ is the time constant. The neuron spikes when $u_k(t)$ crosses the threshold $\theta > 0$. The moment $t^{(k+1)}$ the neuron spikes is called the $(k+1)$-th firing time and it satisfies

$$u_k(t^{(k+1)}) = \theta. \tag{13}$$

It follows that

$$t^{(k+1)} = \begin{cases} +\infty, & \text{if } RI \le \theta, \\ t^{(k)} - \tau \ln\left[1 - \dfrac{\theta}{RI}\right], & \text{if } RI > \theta. \end{cases} \tag{14}$$

Just after the emission of the $(k+1)$-th spike at time $t^{(k+1)}$, the potential is reset to zero, i.e., $u_{k+1}(t^{(k+1)}) = 0$, and the integration of the potential starts all over again for $t > t^{(k+1)}$ until the next spike emission. The asymptotic value RI determines the generation of the spikes: if $RI \le \theta$, there is no spike, otherwise, a spike is emitted (see Fig. 2). This paper

does not consider any absolute refractory period [3] after the spike emission.

B. Spike Interpretation Mechanism (SIM)

Due to the fact that spikes are characterized as stereotype events, the information which is carried on a spike train is either the number of spikes (rate) or the exact time each spike arrives. This subsection is dedicated to the analysis and comparison of the most widely used Spike Interpretation Mechanisms (SIMs), the *rate-SIM* and the *time-SIM*.

1) Rate-SIM: The spiking activity of a neuron over time is usually represented by a graph called the raster plot. Under the assumption that the neurons are independent, it has been proven that for a given input $I(t)$ the firing rate is a stochastic process which causes irregular interspike intervals reflecting a random process [21], [22]. Then, the instantaneous spike rate (mean firing rate) can be obtained either by averaging the spikes of an individual neuron (*spike count*), or by averaging the firing rate over multiple repetitions of the same experiment (*spike density*) [3]. The performance of Rate-SIM methods is poor if (i) it is impossible to repeat the experiment, and (ii) when the observation window is too short such that the neurons are able to emit only a small number of spikes.

2) Time-SIM: An alternative strategy is to interpret a code of spikes by exploiting the time a neuron emits its spikes. Generally, the time-to-first-spike is a time-SIM which assumes that the neuron which fires shortly after the onset of the stimulus is more sensitive to the input comparing to other neurons which are activated somewhat later [23]–[27]. Another famous time-SIM code is the Rank-Order-Coder (ROC) which identifies the spike train of a neuron by ranking the arrival of the first spike [4], [23], [28]. Finally, the LIF can also be considered as a time-SIM as discussed hereafter.

According to Subsection III-A, the LIF encodes the input stimulus into the spike train $\left\{t^{(1)}, \ldots, t^{(N)}\right\}$. From the definition of the arrival times $t^{(k+1)}$ in (14), it follows that the delay $d = d(I)$ between two spikes arrivals is constant because I is assumed constant in the observation window T, i.e., $d = t^{(k+1)} - t^{(k)}$ for any k, and satisfies

$$d(I) = \begin{cases} +\infty, & \text{if } RI < \theta, \\ h(I) = -\tau \ln\left[1 - \dfrac{\theta}{RI}\right], & \text{if } RI > \theta. \end{cases} \tag{15}$$

The stronger the input signal is, the smaller the delay between spikes. On the contrary, a weak input signal corresponds to a larger delay. Figure 2 illustrates the LIF model for three different temporally constant inputs $I_1 < I_2 < I_3$ and a threshold θ. Based on (15), the intensities I_2 and I_3 are able to spike with delays $3 < 2$. The third intensity I_1 remains silent because $RI_1 < \theta$ so its spiking delay turns to infinity.

Let us denote $h^{-1}(d)$ the inverse function of $h(I)$ given by

$$h^{-1}() = \frac{\theta}{R\left(1 - \exp\left(-\frac{}{\tau}\right)\right)}, \quad \text{for } d \neq 0. \tag{16}$$

If the delay , finite or infinite, was perfectly known, the reconstructed value would be $\hat{I}$:

$$\hat{I} = \begin{cases} 0, & \text{if } d > T, \\ h^{-1}(), & \text{if } d \leq T. \end{cases} \tag{17}$$

When d is larger than the observation duration T, the receiver does not receive any spike. Hence, any arbitrary value of $\hat{I}$ is acceptable; the zero value is a reasonable choice. In addition, there is no error of reconstruction when the delay is smaller than T. Based on the analysis above, the substitution of the delay d with the observation window T in (16) results in a new threshold λ associated to the reconstruction error

$$\lambda = R\,h^{-1}(T) = \theta\left(1 - \exp\left(-\frac{T}{\tau}\right)\right)^{-1}. \tag{18}$$

Therefore, according to the aforementioned example where the delay is known, if $RI > \lambda$ there will be no reconstruction error. It can be noted that $\lambda > \theta$ (since $T > 0$ and $\tau > 0$) and λ converges to θ as T becomes arbitrarily large.

The characteristic function of such a "perfect" coding/decoding system, called the *Perfect-LIF*, is a thresholding function [29]. The temporal constraint T implies that all the input value I such that $\theta < RI \leq \lambda$ cannot be recovered by the time-constrained perfect-LIF. As discussed in [30], the transmission of the exact value of the delay d is very expensive regarding the number of coding bits. To decrease the binary rate of the perfect-LIF, this paper proposes to combine the spike counter rate-SIM and the delay coder time-SIM mechanisms resulting in the Dual-SIMQ. The proposed quantizer ensures (i) time-dependency, similar to the HVS behavior, (ii) low memory cost, (iii) high reconstruction quality, (iv) simplicity in terms of the overall processing and (v) extension feasibility to higher dimensional signals.

IV. Dual-SIM Quantizer

This section is dedicated to the analysis of our novel Dual-SIMQ which was first briefly introduced in [30].

A. Dual-SIMQ Coder/Decoder

The first step of the Dual-SIMQ encoder consists in encoding the input value I as a spike train by using the LIF encoder (14). When the input signal is constant, since the interspike delay (15) is constant, we propose, as a second step, to count the spikes instead of coding the interspike delay. The theoretical number of spikes over the time interval $[0, T]$ is:

$$N = N(I) = \begin{cases} 0, & \text{if } RI \leq \lambda, \\ \left\lfloor \dfrac{T}{d(I)} \right\rfloor, & \text{if } RI > \lambda. \end{cases} \tag{19}$$

Once the number of spikes, N, has been transmitted to the decoder, one can easily estimate the interspike delay by

$$\hat{d} = \begin{cases} \infty, & \text{if } N = 0, \\ \dfrac{T}{N} & \text{if } N > 0. \end{cases} \tag{20}$$

Fig. 3 illustrates how the Dual-SIMQ counts the number of spikes with respect to the input intensity I. For any input intensity I such that $RI < \lambda$, there will be no spikes emitted ($N = 0$) because $> T$. Consequently, all the input values which belong to interval

$$S_0 = \{I > 0 : RI < \lambda\} = \left[0, h^{-1}(T)\right) \tag{21}$$

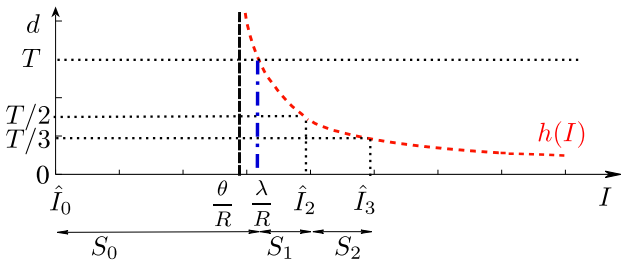


Fig. 3. The input values I is arranged in quantization regions S_k depending on the number k of emitted spikes.

will be recovered by the single output intensity $\hat{I}_0 = 0$. Based on the above equation, it is obvious that the length ℓ_0 of the interval S_0 is $\ell_0 = \lambda/R$. Let us now suppose that only one spike arrives for the input signal I, i.e., $N = 1$. According to (19), all the input intensities I which have caused the generation of a single spike belong to

$$S_1 = \left\{ I > 0 : \frac{T}{2} < d(I) \leq T \right\}. \tag{22}$$

According to the quantization theory [9], assuming that the pdf, $p(I)$, is uniform over S_1, it is well known that the MSE error is minimized when the quantization interval is represented by its center. Hence, we choose to reconstruct any value I associated to S_1 as its centroid value

$$\hat{I}_1 = \frac{1}{2}\left(h^{-1}\left(\frac{T}{2}\right) + h^{-1}(T) \right). \tag{23}$$

With the same reasoning, let us define S_k as the quantization region associated to the input value I which has generated exactly k spikes for any $k \geq 1$, i.e.,

$$S_k = \left\{ I > 0 : \frac{T}{k+1} < d(I) \leq \frac{T}{k} \right\}. \tag{24}$$

The length ℓ_k of an interval S_k for $k \geq 1$ is given by

$$\ell_k = h^{-1}\left(\frac{T}{k+1}\right) - h^{-1}\left(\frac{T}{k}\right). \tag{25}$$

A value $I \in S_k$ is reconstructed by the interval's centroid

$$\hat{I}_k = \frac{1}{2}\left(h^{-1}\left(\frac{T}{k+1}\right) + h^{-1}\left(\frac{T}{k}\right)\right). \tag{26}$$

B. Dealing With Real Values

In signal processing, it is very common that an input source has to be first transformed before the quantization. The transformation enables to concentrate most of the signal information in few low frequency components. However, after the transformation, most of the times, occur negative values, so here we describe how the proposed Dual-SIMQ deals with negative inputs.

Suppose that the input value I corresponds to one of the pixel values of an image. We have decided to assign 1-bit per pixel to encode the sign of each input intensity as following

$$\text{sgn}(I) = \begin{cases} 1, & \text{if} \quad I \geq 0, \\ -1, & \text{otherwise.} \end{cases} \tag{27}$$

Thus, the Dual-SIMQ coder receives as an input the absolute value of each input intensity $|I|$ and computes the number of the emitted spikes k within the observation window T. Then, the decoder receives the sign information, $\text{sgn}(I)$, and the number of spikes k which are associated to the quantization interval S_k represented by the centroid value $|\hat{I}|$. Finally, the output of the decoder is given by $\tilde{I} = \text{sgn}(I)|\hat{I}|$ and the reconstructed values belong to the set

$$\tilde{I} \in \left\{ \tilde{I}_0, \tilde{I}_1, \ldots, \tilde{I}_k, \ldots \right\}. \tag{28}$$

C. Dynamic Properties of the Dual-SIM Quantization

As explained in the previous sections, the performance of the Dual-SIMQ is mainly driven by the threshold parameter θ. Figure 4(a) shows for a zero-mean normal distribution input with $\sigma = 2$ that when the threshold value increases, the distortion generated by the Dual-SIMQ increases. This is also obvious by the characteristic function of the Dual-SIMQ (see Fig. 4(b)) where the length ℓ of the quantization steps are wider as theta increases. However besides θ, there are other parameters that also influence the Dual-SIMQ response such as the observation window T and the resistance R.

1) Time-Dependent Dual-SIMQ: The "dynamic" behavior of the Dual-SIMQ is one of its most important properties associated with the fact that the number of spikes depends on the length of the observation window T. According to (19), the longer the input signal is "flashed" in front of a LIF neuron, the more the number of spikes that correspond to this input intensity. On the other hand, if the observation window is too small, the number of spikes will fail to precisely describe the input signal.

As depicted in Fig. 4(c), for a normal distribution input, while increasing the observation window T, the quality of the reconstructed signal substantially improves. It is remarkable that when time is too short ($T < 20$ ms) the Dual-SIMQ is not able to perceive any information regarding the input signal. This is a natural coincidence due to the neuroscience models and parameters embedded in the Dual-SIMQ. This time could be intuitively explained as the propagation time of the visual stimulus to the spiking neurons. In addition, it is obvious that at a given time ($T \approx 100$ ms) the reconstruction quality vanishes into an asymptotic value.

2) Resistance-Dependent Dual-SIMQ: It has been shown in [30] that the Dual-SIMQ can be approximated by a USQ for very large values of the resistance R. In this work, we extend this proof and we show that R determines the Dual-SIMQ response that varies from *uniform* to *non-uniform*.

Proposition 1: Given a value of R arbitrarily large. Let us assume that the input value I has generated exactly $N = k$ spikes for any $k \geq 1$ during T milliseconds. Then, the Dual-SIMQ behaves as a uniform quantizer where the length ℓ_k of each quantization interval is constant for all k and given by:

$$\ell_k = \ell = \frac{\theta C}{T} + o\left(\frac{1}{R}\right), \quad \forall k, \tag{29}$$

where the notation $o(\cdot)$ is the little-o notation, which is used to express the asymptotic behavior of a function.

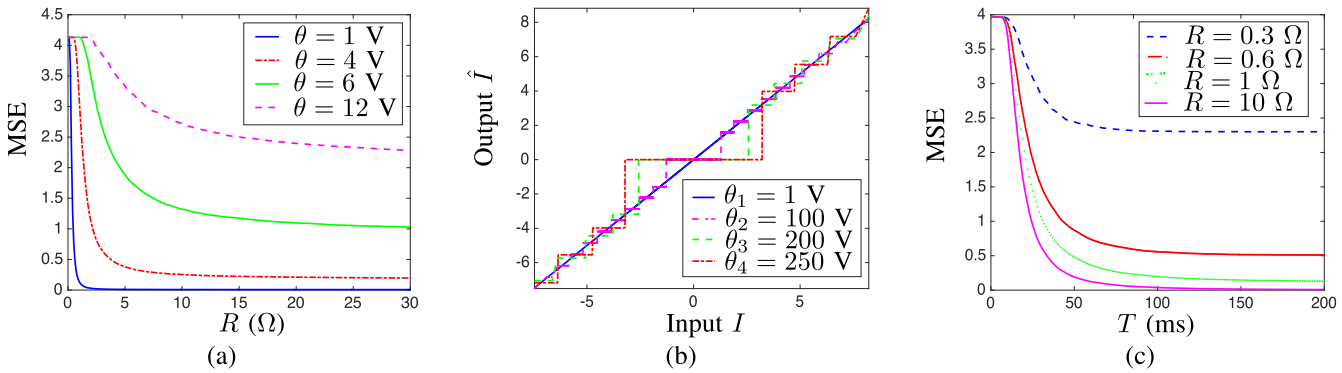


Fig. 4. (a) Impact of the threshold θ on the performance of the Dual-SIMQ for a zero-mean normal distribution with $\sigma = 2$; the distortion increases with theta (set of parameters: $C = 50$ F, $T = 200$ ms). (b) Dual-SIMQ characteristic function for different θ values (parameters: $T = 150$ ms, $R = 100$ Ω, $C = 1$ F). (c) Impact of the size of the observation window T on the performance of the Dual-SIMQ. The reconstruction quality improves when the size of the observation window increases (set of parameters: $C = 50$ F, $\theta = 1$ V).

Then, the number of the generated spikes is

$$N = N(I) = \left\lfloor \frac{I}{\ell} \right\rfloor = \left\lfloor \frac{T}{\theta C} I \right\rfloor. \tag{30}$$

Proof: Using the Taylor series it follows that, for $k \geq 1$,

$$h^{-1}\left(\frac{T}{k}\right) = \frac{\theta C}{T}k + \frac{\theta}{2R} + \frac{\theta T}{12\,R^2 C k} + o\left(\frac{1}{R^2}\right). \tag{31}$$

Combining (25) and (31) yields (29). Furthermore, we get

$$\frac{\lambda}{R} = h^{-1}(T) = \frac{\theta C}{T} + \frac{\theta}{2\,R} + \frac{\theta T}{12\,R^2\,C} + o\left(\frac{1}{R^2}\right).$$

Finally, a short calculation based on the Taylor series of the logarithm shows that

$$d(I) = h(I) = \frac{\theta C}{I} + \frac{C\theta^2}{2RI^2} + o\left(\frac{1}{R}\right). \tag{32}$$

Incorporating (32) in (19) yields (30). □

Proposition 1 shows that the Dual-SIMQ coincides with a USQ, $Q_{q=\ell_\infty,\lambda=2\ell_\infty}(x)$, as R becomes arbitrarily large, with a quantization step $q = \ell_\infty$. This confirms that a large T yields an accurate quantization. On the opposite, a large value of θ or C decreases the accuracy of the quantizer.

Proposition 2: Given a value of R relatively small, then the Dual-SIMQ behaves as a non-uniform quantizer. The length of each quantization interval depends on the number of spikes k. When k increases, the length ℓ_k converges to an asymptotic value

$$\lim_{k\to\infty} \ell_k = \frac{\theta C}{T}. \tag{33}$$

Proof: See the proof in the Appendix A. □

Figure 5 illustrates how the value of R affects the length ℓ of the quantization intervals in function of the number of spikes. When R is small and $k \geq 1$, the length of the quantization interval is a strictly increasing function which is upper-bounded by ℓ_0. However, when R is large, the Dual-SIMQ becomes completely uniform. As we show later on, when the Dual-SIMQ is applied to a normal distribution signal, it is expected to better encode the low than the high intensities when R is small. On the contrary, whatever the intensity is, if R is high it will behave towards a uniform manner. The interpretation of the above behavior will be more evident in section V.

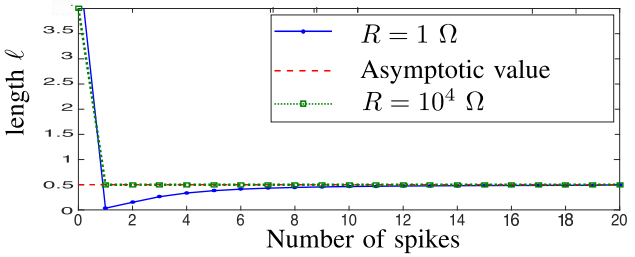


Fig. 5. Performance of the Dual-SIMQ as (i) non-uniform quantizer (small R) and (ii) uniform quantizer (R arbitrarily large). The number of spikes varies in function of the intensity I (set of parameters: $\theta = 5$ V, $C = 10$ F and $T = 100$ ms.

V. Numerical Results

The proposed neuro-inspired Dual-SIMQ is compared to the state-of-the-art and evaluated in terms of the rate- distortion trade-off on both simulated and real data against.

A. Experiments With Simulated Data

In this section, we aim to study the validity of the rate-distortion theory which is determined by the comparison of the rate-distortion approximation and the performance of the Dual-SIMQ when the distribution of the input signal is normal. It has been proven in Section IV-C, that the Dual-SIMQ is a dynamic quantizer that performs either as a uniform or as a non-uniform quantizer. The distortion approximation (7) is only related to an asymptotic behavior thus, the comparison is considered against the uniform Dual-SIMQ. If the constant length of the quantization intervals ℓ_k (33) takes the place of the quantization step q in (7), it is trivial to visualize that the distortion approximation perfectly overfits the behavior of the Dual-SIMQ leading to the following formula:

$$D = \frac{(\theta C)^2}{12T^2} \sim \frac{\alpha}{T^2}. \tag{34}$$

Interestingly, we can see that the distortion of the Dual-SIMQ quantizer is inversely proportional to the observation time T for given θ and C.

Fig. 6 compares the performance of the Dual-SIMQ and the USQ with deadzone for input samples following zero mean Laplacian and Gaussian *i.i.d.* distributions with $\sigma = 2$. According to [13] for a USQ with deadzone, there is an

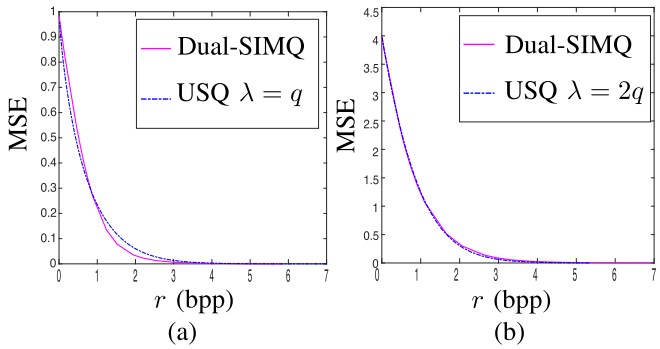


Fig. 6. Comparison of the asymptotic performance of the Dual-SIMQ with the USQ with deadzone for inputs following (a) the Laplacian distribution and (b) the Gaussian distribution (set of parameters: C = 1 F, $R = 10^3$ Ω, $\theta \in$[5,800] V and $T = 150$ ms).

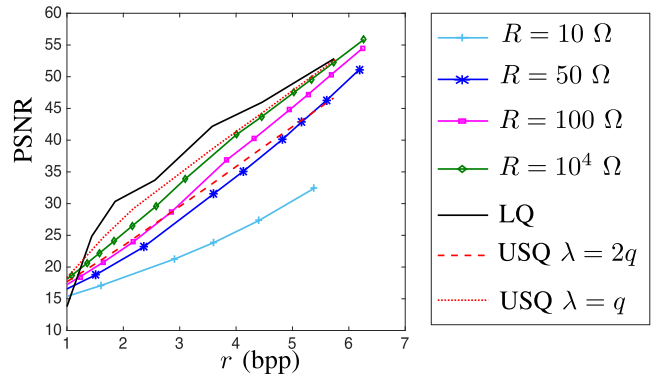


Fig. 7. Comparison between (i) the Dual-SIMQ, (ii) the USQ with deazone $\lambda = q$, (iii) the USQ without deadzone ($\lambda = 2q$) and (iv) the non-uniform Lloyd quantizer for 100 synthetic normal zero mean distribution signals (set of parameters: $C = 1$ F, $T = 150$ ms, $\theta \in \{1, R, \ldots, 10R\}$ V, $\lambda = q$, $q = \{1, 8, 10, 15, 20, 40, 60, 80, 100\}$).

optimal relationship between the quantization step q and the deadzone λ; for Laplacian distributions the deadzone equals $\lambda = 2q$ and for Gaussian distributions $\lambda = q$. Proposition 1 has shown that the performance of the Dual-SIMQ is asymptotically equivalent to the optimal USQ with deadzone.

It is proven in Section IV, that by tuning some of the Dual-SIMQ parameters its behavior might be uniform or non-uniform. For this reason, we have decided to contrast the performance of the Dual-SIMQ with the uniform USQ and the non-uniform LQ. Figure 7 illustrates the average behavior of (i) the proposed Dual-SIMQ, (ii) the USQ with deadzone, (iii) the USQ without deadzone and (iv) the LQ when trained with 100 synthetic zero mean normal distribution images with $\sigma = 2$. As expected, for the same rate values the performance of Dual-SIMQ, outperforms the USQ without deadzone, coincides with the optimal USQ with deadzone $\lambda = q$, while it approximates the capacity of the optimal LQ quantizer. The quality evaluation of the results was measured by the PSNR metric (35) while the rate was computed according to (9). Throughout this paper the entropy is given in bits per pixel (bpp).

$$\mathrm{PSNR}(I, \hat{I}) = 10 * \log_{10}\left(\frac{255^2}{\mathrm{MSE}(I, \hat{I})}\right), \tag{35}$$

where $\mathrm{MSE}(I, \hat{I})$ is defined by (6).

B. Dual-SIMQ on Real Data

This section is dedicated to the comparison of the proposed neuro-inspired quantizer to the state-of-the-art. Figure 8 illustrates the schema of every quantization architecture that participates to this comparison. The first quantization is when the input value I is quantized by a USQ or the LQ (see Fig. 8 (a) and (b) respectively). The second type of quantization is composed of three steps: (i) the input value I is transformed in a spike train with a constant interspike delay $d(I)$, (ii) the interspike delay $d(I)$ is quantized with a USQ or the LQ (see Fig. 8 (c) and (d), respectively) (iii) the reconstructed value $\hat{I}$ is given by:

$$\hat{I} = \begin{cases} 0, & \text{if} \quad \hat{d} > T, \\ h^{-1}(\hat{d}), & \text{otherwise,} \end{cases} \tag{36}$$

where $h^{-1}(\cdot)$ is defined in (17). The third quantization is given by the Dual-SIMQ (see Fig. 8 (e)).

The first goal of this section is to show that, in terms of compression, counting the number of spikes is more efficient than quantizing the delays (see Fig. 8 (e) versus (c)-(d)). The second mission is to compare the proposed neuro-inspired quantizer to the state-of-the-art (i) USQ with and/or without deadzone and (ii) LQ when applied directly to the pixel intensities (see Fig. 8 (e) versus (a)-(b)). Let the input intensities $I_1, \ldots, I_n$ correspond to the pixel values of each input image $I = (I_1, \ldots, I_n)$.

Figure 9 visually compares the performance of all the aforementioned quantization architectures. As expected, the Dual-SIMQ (see cases (e.1) and (e.1)) outperforms the state-of-the-art USQ (see cases (c.1) and (c.2)) and LQ (see cases (d.1) and (d.2)) applied to the delays. This is evident both numerically, by the fact that for similar rates the quality assessment using the PSNR metric (35) is higher, and visually especially for lower rates (see cases (c.2),(d.2) and (e.2)), where not only the details but also the intensities range of the input image are better approximated by the neuro-inspired method. For instance, paying attention to the brim of the hat or the woman's hand we observe that the Dual-SIMQ has better approximated the original intensities. Figure 9 verifies that for the same rate r (bpp) counting the number of spikes (see cases (e.1) and (e.2)) is as efficient as quantizing the pixel intensities uniformly (see cases (a.1) and (a.2)) but less efficient than the non-uniform manner (see cases (b.1) and (b.2)).

Similar to Fig. 7, the experiment in Fig. 10 compares the performance of Dual-SIMQ to the state-of-the-art for real data. In particular, it exploits 100 images with $n = 256 \times 256$ pixels taken from the USC-SIPI database [31]. The best performance of the proposed neuro-inspired quantizer is the one of the uniform Dual-SIMQ (blue diamond curve). Its rate-distortion behavior is similar to the USQ with deadzone (red dotted curve) but much better than the USQ without deadzone (red dashed curve). As expected, the optimal quantizer still remains the LQ. From the computational complexity point of view, the LQ requires a training process which is time demanding whereas Dual-SIMQ and USQ require only 0.0137 sec. and 0.0112 sec., respectively to quantize/dequantize a still image

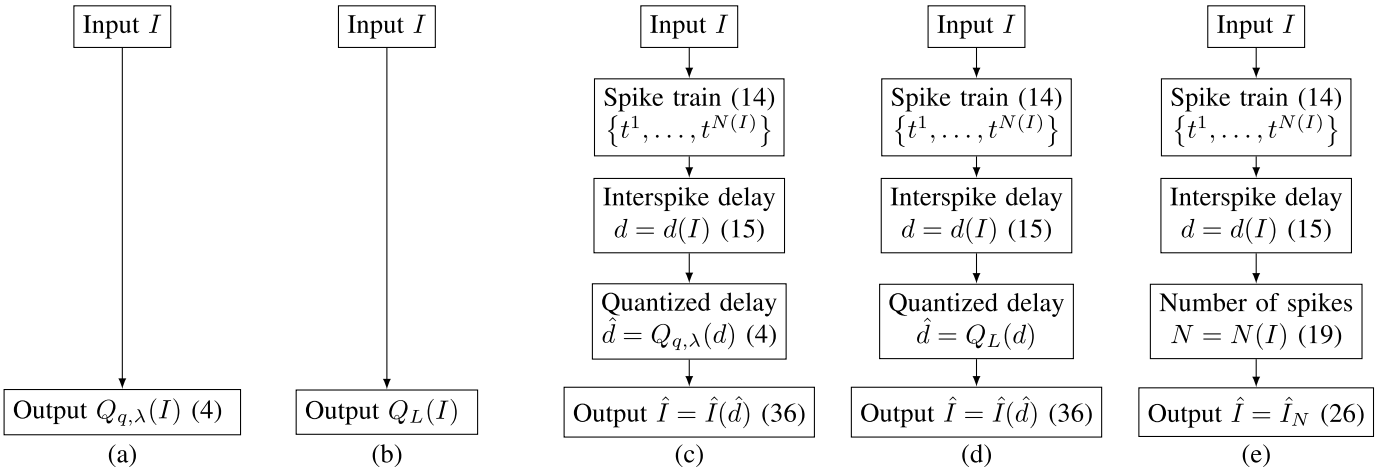


Fig. 8. Quantization schemes used to evaluate the performance of Dual-SIMQ: (a) USQ $Q_{q,\lambda}$ applied to I, (b) Lloyd Quantizer Q_L applied to I, (c) USQ $Q_{q,\lambda}$ applied to delays $d(I)$, (d) Lloyd Quantizer Q_L applied to delays, and (e) Dual-SIMQ.

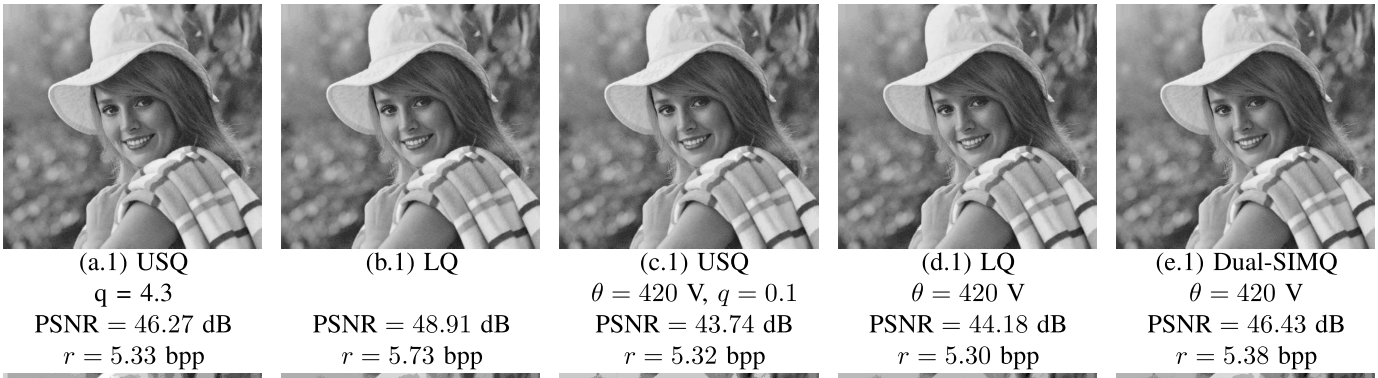


Fig. 9. Visual comparison between Dual-SIMQ (e.1) and (e.2)), USQ applied to the delays (c.1)-(c.2), Lloyd applied to the delays (d.1)-(d.2), USQ applied to the pixel intensities (a.1)-(a.2) and Lloyd applied to the pixel intensities (b.1)-(b.2) for similar rates (set of parameters: $R = 10^4\ \Omega$, $C = 1$ F, $T = 100$ ms).

with 230×230 pixels on a MacBook Pro with a 2.6GHz Intel Core i7 processor.

C. Progressive Reconstruction

The Dual-SIMQ is a time-dependent quantizer as discussed in section IV-C. It is also evident according to Proposition 2 and Proposition 1 that the lengths of the quantization regions depend on T, especially the quantization step ℓ_∞ vanishes as T becomes arbitrarily large. Fig. 11 illustrates the dynamic behavior of the Dual-SIMQ comparing its reconstruction performance for different observation windows T.

As expected, when the available observation time of the Dual-SIMQ is short, the number of spikes corresponding to high intensities is limited. As a result, the quality of the reconstruction is poor because most of the small intensities will be represented by one or none spikes. On the other hand, when the observation window is large almost all the pixel intensities will generate some spikes improving in that sense the reconstruction quality. As a consequence, the progressive enhancement of the reconstructed signal is definitely among the most important and ground-breaking benefits of the Dual-SIMQ taking under consideration that none of the

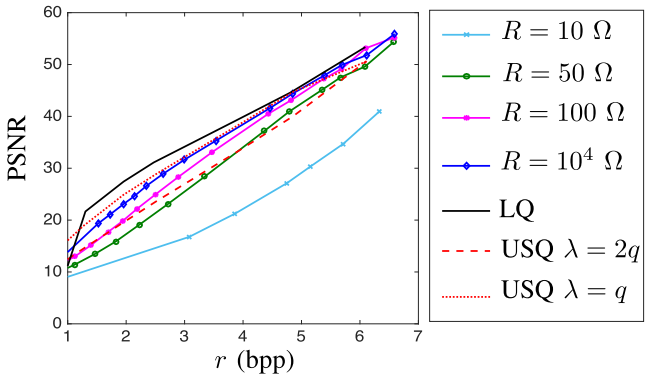
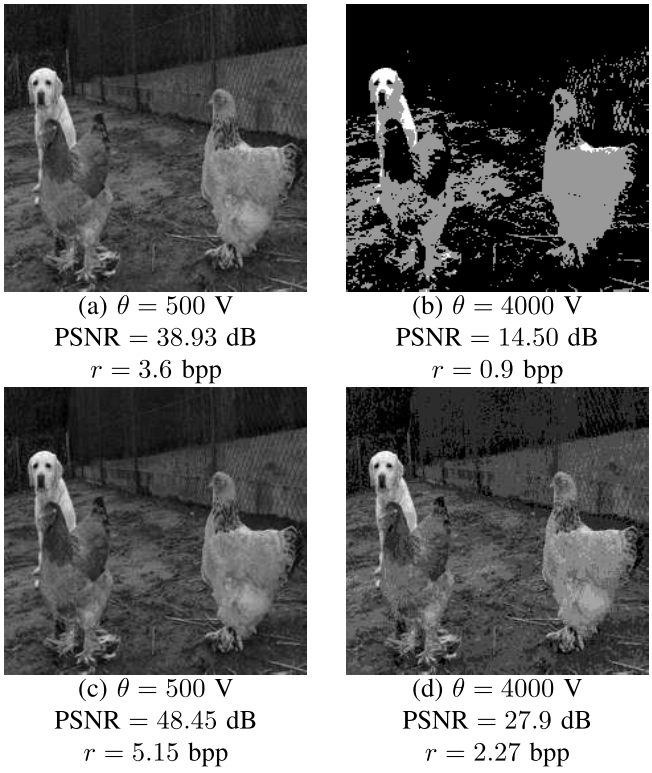


Fig. 10. Comparison between (i) the Dual-SIMQ, (ii) the USQ with deazone $\lambda = q$, (iii) the USQ without deadzone ($\lambda = 2q$) and (iv) the non-uniform Lloyd quantizer for 100 real images taken from [31] (set of parameters: $C = 1$ F, $T = 150$ ms, $\theta \in \{1, R, \ldots, 10R\}$ V, $\lambda = q$, $q = \{1, 8, 10, 15, 20, 40, 60, 80, 100\}$).



(a) $\theta = 500$ V, PSNR $= 38.93$ dB, $r = 3.6$ bpp

(b) $\theta = 4000$ V, PSNR $= 14.50$ dB, $r = 0.9$ bpp

(c) $\theta = 500$ V, PSNR $= 48.45$ dB, $r = 5.15$ bpp

(d) $\theta = 4000$ V, PSNR $= 27.9$ dB, $r = 2.27$ bpp

Fig. 11. Dual-SIMQ progressive reconstruction for $T = 50$ ms (a)-(b)and $T = 150$ ms (c)-(d) (set of parameters: $R = 10^3$ Ω and $C = 1$ F).

state-of-the-art quantization methods is able to improve the quality of the signal along time.

VI. Conclusion

This paper has introduced a novel, bio-inspired encoder/decoder of natural images called the Dual-SIMQ. The Dual-SIMQ encoder is based on the LIF model, a very efficient spike generation mechanism which approximates the neural spiking process. The Dual-SIMQ decoder is a combination of two spike interpretation mechanisms which approximates the spike arrival delay by counting the number of spikes within a given observation window.

The Dual-SIMQ framework can play a pivotal role in the signal, image and video processing fields because it allows to encode the input values in a simple and dynamic manner, mimicking the neural behavior. At the same time, it enables to progressively reconstruct the input value.

We aim at extending this work and apply the Dual-SIMQ to videos where time is an important parameter which is directly linked to the frame rate of the video stream. Last but not least, the "bigger picture" of this work is the development of a compression system that understands the visual word according to the human visual perception. Within this framework, merging different neuro-inspired processing tools, such as the retina-inspired filter [32] and the proposed Dual-SIMQ, could establish an alternative signal reconstruction methodology depending on neurons capabilities.

Appendix A
Proof of Proposition 2

Let $f : [1, +\infty) \mapsto \mathbb{R}$ be the differentiable function:

$$f(x) = \frac{1}{1 - \exp\left(-\frac{\alpha}{x}\right)} \tag{37}$$

where $\alpha = T/\tau > 0$ and let $g(x) = f(x+1) - f(x)$. It is straightforward to verify that

$$\ell_k = \frac{\theta}{R} g(k), \quad \forall k \geq 1.$$

Let us show that g is an increasing function. The first derivative of g is

$$g'(x) = f'(x+1) - f'(x) \tag{38}$$

where

$$f'(x) = \frac{\alpha \exp\left(-\frac{\alpha}{x}\right)}{x^2\left(1 - \exp\left(-\frac{\alpha}{x}\right)\right)^2}. \tag{39}$$

A short calculation shows that

$$f'(x) = \frac{\alpha}{4x^2 \sinh^2\left(\frac{\alpha}{2x}\right)} \tag{40}$$

where $\sinh(\cdot)$ is the hyperbolic sine. Let $u : [1, +\infty) \mapsto \mathbb{R}$ be the function defined by

$$u(x) = x \sinh\left(\frac{\alpha}{2x}\right).$$

The function $u(x)$ is strictly decreasing over $[1, +\infty)$ since its first derivative is strictly negative. Indeed, we get

$$u'(x) = \sinh\left(\frac{\alpha}{2x}\right) - \frac{\alpha}{2x}\cosh\left(\frac{\alpha}{2x}\right)$$

where $\cosh(\cdot)$ is the hyperbolic cosine. So, $u'(x) < 0$ is equivalent to

$$\tanh\left(\frac{\alpha}{2x}\right) < \left(\frac{\alpha}{2x}\right).$$

A short calculation shows that

$$\tanh(y) < y, \ \forall y > 0,$$

where $\tanh(\cdot)$ is the hyperbolic tangent, which proves that $u'(x) < 0$ for all $x \geq 1$. Since $u(x)$ is strictly positive and strictly decreasing, $u^2(x)$ is also strictly decreasing. It follows that $f'(x)$ is strictly increasing. From (38), it follows that

$g'(x) > 0$ for all $x \geq 1$. This shows that g is strictly increasing and, hence, ℓ_k is a strictly increasing sequence of reals.

Let us calculate the limit of the sequence ℓ_k. The Taylor series of $h^{-1}(x)$ at $x = 0$ is given by

$$h^{-1}(x) = \frac{\theta C}{x} + \frac{\theta}{2R} + \frac{\theta x}{12\,R^2 C} + \frac{\theta}{R} o\left(\frac{x}{RC}\right), \qquad (41)$$

where $o(\cdot)$ is the little-o notation such that $f = o(g)$ means that there exists a function $\varepsilon(x)$ satisfying $f = g\varepsilon$ and $\varepsilon(x) \to 0$ as $x \to 0$. Indeed, a short calculation shows that:

$$\frac{1}{1 - \exp(-x)} = \frac{1}{x} + \frac{1}{2} + \frac{1}{12}x + o(x), \qquad (42)$$

and the derivation of (41) is straightforward. Assuming that k is large and applying (41) to each term of (25) yields

$$\ell_k = \frac{\theta C}{T} + o\left(\frac{1}{k}\right). \qquad (43)$$

The limit ℓ_∞ is immediate.

Finally, let us show that $\ell_0 > \ell_\infty$. It is well known that $\exp(-x) > 1 - x$ for all $x \neq 0$. Hence, it follows that

$$\ell_0 = \frac{\lambda}{R} = \frac{\theta}{R}\left(1 - \exp\left(-\frac{T}{RC}\right)\right)^{-1} > \frac{\theta}{R}\frac{RC}{T} = \ell_\infty. \qquad (44)$$

REFERENCES

[1] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. Image Process.*, vol. 1, no. 2, pp. 205–220, Apr. 1992.

[2] Z. Liu, L. J. Karam, and A. B. Watson, "JPEG2000 encoding with perceptual distortion control," *IEEE Trans. Image Process.*, vol. 15, no. 7, pp. 1763–1778, Jul. 2006.

[3] W. Gerstner and W. Kistler, *Spiking Neuron Models: Single Neurons, Populations, Plasticity*. Cambridge, U.K.: Cambridge Univ. Press, 2002.

[4] R. V. Rullen and S. J. Thorpe, "Rate coding versus temporal order coding: What the retinal ganglion cells tell the visual cortex," *Neural Comput.*, vol. 13, no. 6, pp. 1255–1283, Jun. 2001.

[5] K. Masmoudi, M. Antonini, and P. Kornprobst, "Frames for exact inversion of the rank order coder," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 2, pp. 353–359, Feb. 2012.

[6] A. A. Lazar and L. T. Tóth, "Perfect recovery and sensitivity analysis of time encoded bandlimited signals," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 51, no. 10, pp. 2060–2073, Oct. 2004.

[7] Y. C. Yoon, "LIF and simplified SRM neurons encode signals into spikes via a form of asynchronous pulse sigma–delta modulation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 5, pp. 1192–1205, May 2017.

[8] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Toronto, ON, Canada: Wiley, 2006.

[9] P. C. Cosman, R. M. Gray, and M. Vetterli, "Vector quantization of image subbands: A survey," *IEEE Trans. Image Process.*, vol. 5, no. 2, pp. 202–225, Feb. 1996.

[10] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2325–2384, Oct. 1998.

[11] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 2, pp. 129–137, Mar. 1982.

[12] J. Max, "Quantizing for minimum distortion," *IEEE Trans. Inf. Theory*, vol. IT-6, no. 1, pp. 7–12, Mar. 1960.

[13] C. Parisot, "Allocations basées modéles et transformée en ondelettes au fil de l'eaupour le codage d'images et de vidéeos," Ph.D. dissertation, Dept. Discipline Autom. Singal Image Process., Université de Nice-Sophia Antipolis, Nice, France, 2003.

[14] A. L. Hodgkin and A. F. Huxley, "Currents carried by sodium and potassium ions through the membrane of the giant axon of *Loligo*," *J. Physiol.*, vol. 116, no. 4, pp. 449–472, Apr. 1952.

[15] R. FitzHugh, "Impulses and physiological states in theoretical models of nerve membrane," *Biophys. J.*, vol. 1, no. 6, pp. 445–466, Jul. 1961.

[16] J. Nagumo, S. Arimoto, and S. Yoshizawa, "An active pulse transmission line simulating nerve axon," *Proc. IRE*, vol. 50, no. 10, pp. 2061–2070, Oct. 1962.

[17] J. Rinzel, "Excitation dynamics: Insights from simplified membrane models," *Theor. Trends Neurosci., Fed. Proc.*, vol. 44, no. 15, pp. 2944–2946, 1985.

[18] J. Rinzel and G. B. Ermentrout, "Analysis of neural excitability and oscillations," in *Methods in Neuronal Modeling: From Synapses to Networks*, C. Koch and I. Segev, Eds., 2nd ed. Cambridge, MA, USA: MIT Press, 1998, pp. 251–291.

[19] W. Gerstner, *A Framework for Spiking Neuron Models: The Spike Response Model*, vol. 4. Amsterdam, The Netherlands: North Holland, 2001.

[20] L. F. Abbott and T. B. Kepler, "Model neurons: From Hodgkin-Huxley to hopfield," in *Statistical Mechanics of Neural Networks*, L. Garrido, Ed. Berlin, Germany: Springer, 1990.

[21] E. D. Adrian, "The impulses produced by sensory nerve endings," *J. Physiol.*, vol. 61, no. 1, pp. 47–72, 1926.

[22] D. Heeger, "Poisson model of spike generation," *Handout*, vol. 5, pp. 1–13, Sep. 2000.

[23] S. Thorpe, D. Fize, and C. Marlot, "Speed of processing in the human visual system," *Nature*, vol. 381, no. 6582, pp. 520–522, Jun. 1996.

[24] J. J. Hopfield, "Pattern recognition computation using action potential timing for stimulus representation," *Nature*, vol. 376, pp. 33–36, Jul. 1995.

[25] O. Jensen and J. E. Lisman, "Hippocampal CA3 region predicts memory sequences: Accounting for the phase precession of place cells," *Learn. Memory*, vol. 3, nos. 2–3, pp. 279–287, Jan. 1996.

[26] W. Maass, "Lower bounds for the computational power of networks of spiking neurons," *Neural Comput.*, vol. 8, no. 1, pp. 1–40, Jan. 1996.

[27] J. O'Keefe, "Hippocampus, theta, and spatial memory," *Current Opinion Neurobiol.*, vol. 3, no. 6, pp. 917–924, Dec. 1993.

[28] S. J. Thorpe and J. Gautrais, "Rank order coding: A new coding scheme for rapid processing in neural network," in *Computational Neuroscience: Trends in Research*. 1998, pp. 113–118.

[29] E. Doutsi, L. Fillatre, M. Antonini, and J. Gaulmin, "Bio-inspired sparse representation of images," in *Proc. Groupe d'Etudes du Traitement du Signal et des Images (Gretsi)*, 2017, pp. 1–4.

[30] E. Doutsi, L. Fillatre, M. Antonini, and J. Gaulmin, "Neuro-inspired quantization," in *Proc. IEEE Int. Conf. Image Process.*, Athens, Greece, Oct. 2018, pp. 689–693.

[31] University of Southern California Signal and Image Processing Institute. *USC-SIPI Image Database*. [Online]. Available: http://sipi.usc.edu/database/

[32] E. Doutsi, L. Fillatre, M. Antonini, and J. Gaulmin, "Retina-inspired filter," *IEEE Trans. Image Process.*, vol. 27, no. 7, pp. 3484–3499, Jul. 2018.

Effrosyni Doutsi (Member, IEEE) received the bachelor's degree in computer science and biomedical informatics from the University of Thessaly, Greece, in 2012, the master's degree in computational biology and biomedicine from the University of Nice Sophia Antipolis, France, in 2013, and the Ph.D. degree in computer science from Laboratoire d'Informatique, Signaux et Systèmes de Sophia Antipolis (I3S Laboratory), CNRS, Université Côte d'Azur, in 2017. Since 2018, she has been an Associate Postdoctoral Researcher with the Signal Processing Laboratory, FORTH-ICS. She is also a Visiting Professor with the Department of Electrical and Computer Engineering, Technical University of Crete. She has coauthored more than 20 journal and conference papers. Her research interests include signal processing, brain-like computing, applied neuroscience, machine learning, and computational biology. She has received several scholarships and research grants. She is also a Principal Investigator of the BRIEFING Project funded by the French Government within the framework of the Make Our Planet Great Again (MOPGA) Call.

Lionel Fillatre received the M.Sc. degree in decision and information engineering and the Ph.D. degree in systems optimization from the University of Technology of Troyes (UTT), France, in 2001 and 2004, respectively. From 2005 to 2007, he worked with Tèlècom Bretagne, Brest, France. From 2007 to 2012, he was an Associate Professor with the Systems Modeling and Dependability Laboratory, UTT. Since 2012, he has been a Full Professor with the Laboratoire d'Informatique, Signaux et Systèmes de Sophia Antipolis (I3S Laboratory), Université Côte d'Azur. His current research interests include statistical decision theory, machine learning, deep learning, signal and image processing, and bio-inspired processing.

Marc Antonini (Member, IEEE) received the Ph.D. degree in electrical engineering and the Habilitation à Diriger des Recherches degree from the University of Nice Sophia Antipolis, France, in 1991 and 2003, respectively. He was a Post-Doctoral Fellow with the Centre National d'Etudes Spatiales (CNES), Toulouse, France, in 1991 and 1992. In 1993, he joined the Laboratoire d'Informatique, Signaux et Systèmes de Sophia Antipolis (I3S Laboratory), CNRS, Université Côte d'Azur, where he has been the Directeur de Recherche CNRS, since 2004. From 1995 to 2001, he was involved in the Earth Observation Program "Pléiades" at CNES, for the definition of the on-board image coder and the image analysis solution he developed with his research group has been implemented inside the satellite (first launch in 2011). He is currently leading the MediaCoding Research Team, I3S Laboratory. He is also the Co-Founder of Cintoo, a Start-Up founded in July 2013, spin-off from Université Côte d'Azur and CNRS. Cintoo develops technologies and solutions for managing and leveraging the 3D data coming from reality capture devices in the cloud. More recently, he started an activity on the storage of digital images onto synthetic DNA. His works on wavelet transform have been included in the JPEG2000 image coding standard. He is the author of more than 250 articles and seven book chapters and holds 13 patents. His research activities cover in particular image and video coding as well as geometric processing and compression of static surface meshes and animations. He is also interested in the analysis of the information contained by the neural code in the visual system, with bio-inspired applications in image and video compression. He is a member of the IEEE MMSP Technical Committee and an Expert Member of the JPEG-DNA Ad-hoc Group. He was the Prize-Winner in 2013 for the French National Innovation Prize for the creation of start-up (BPI France " Emergence "). He was awarded the Medal of the University of Nice Sophia Antipolis in 2013. He is also an Associate Editor of the journal IEEE TRANSACTIONS ON IMAGE PROCESSING and the *Journal on Image and Video Processing* (Elsevier).

Panagiotis Tsakalides (Member, IEEE) received the Diploma degree in electrical engineering from the Aristotle University of Thessaloniki, Greece, in 1990, and the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, CA, USA, in 1995. He is currently a Professor in computer science with the University of Crete, Greece, and the Head of the Signal Processing Laboratory, FORTH-ICS. He has an extended experience of transferring research and interacting with the industry. During the last ten years, he has been the project coordinator in seven European Commission and 12 national research and innovation projects totaling more than €5.5 M in actual funding for FORTH-ICS and the University of Crete. His research interests include the fields of statistical signal processing and machine learning with emphasis in non-Gaussian estimation and detection theory, sparse representations, and applications in sensor networks, audio, imaging, and multimedia systems. He has coauthored more than 200 technical publications in these areas.