Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 6647135, 12 pages
https://doi.org/10.1155/2021/6647135

Research Article

Symmetric Quantile Quantizer Parameterization for the Laplacian Source: Qualification for Contemporary Quantization Solutions

Zoran Perić,[1] **Jelena Nikolić**,[1] **Danijela Aleksić**,[2] **and Anastasija Perić**[1]

[1]*Faculty of Electronic Engineering, University of Niš, Niš 18000, Serbia*
[2]*Department of Mobile Network Niš, Telekom Srbija, Voždova 11, Niš, Serbia*

Correspondence should be addressed to Jelena Nikolić; jelena.nikolic@elfak.ni.ac.rs

Received 23 November 2020; Revised 11 January 2021; Accepted 24 January 2021; Published 8 February 2021

Academic Editor: A. M. Bastos Pereira

In this paper, we consider the opportunities and constraints, which rest on quantization as a guiding principle for data representation and compression. In particular, we propose a novel model of Symmetric Quantile Quantizer (SQQ) and we describe in detail its parameterization. We suggest a simple method for offline precalculation of its parameters and we examine the inevitable loss of information introduced by SQQ, as an important part of bit optimization task at the traditional network level, which can be globally mapped out in many contemporary solutions. Our anticipation is that such precalculated values can be leveraged in deterministic quantization process. We highlight that this notice heavily relies on the fact that the values of interest are distributed according to the Laplacian distribution, which we consider in the paper. The basic difference of our SQQ and the previously established asymptotically optimal quantizer model, that is, Scalar Companding Quantizer (SCQ), is reflected in the fact that, in SCQ model, both decision thresholds and representation levels are determined in accordance with the specified compressor function, whereas in our SQQ model, a precedence of SCQ model for the straightforward decision thresholds calculation is used, while the representation levels are optimally determined for the specified decision thresholds and assumed Laplacian distribution. As a result, our SQQ outperforms SCQ in terms of signal-to-quantization noise ratio (SQNR). As stated in this paper, there are numerous indications to make us believe that appropriate quantizer parameterization will move us closer to an optimization in the amount of the transferred data in bits, which is strongly dependent on the amount of SQNR.

1. Introduction

Humans are unique in their ability to weigh perceptions and make decisions based on them. Apparently, human observation is inherently very subjective and thus varies across individuals. Additionally, emotional, cognitive, and social processes can override basic instincts and deductive or inductive conclusion and guide behavioral choices. In this respect, the human decision to infer is a consequence of an interaction of all these factors. In practice, it is not realistic to identify and adequately quantify all the factors that play a role in a certain phenomenon, even if the phenomenon works simply. As the amount of data we need to generalize grows, the number of features or dimensions also grows. Assuming that the explanatory part of many problems is characterized as highly dimensional, in order to fix the data in a linear space or on some space of a significantly smaller dimension, it is desirable to involve a prediction model or model ensembles [1–3].

When analyzing the effectiveness of predictions, it is preferable to make the prediction models close to the actual data. Although it is possible to achieve high accuracy on the actual training set, the real challenge is to develop a model that generalizes well to data not known in advance. The outlined criterion to the predictive model logically leads to a solution, which lets some amounts of data to dictate algorithms and solutions that penalize underfitting and overfitting. More generally, the presence of overfitting is a common problem of many neural network (NN) models [4]. Additionally, it is not possible to partially prevent from occurring possible extrapolation errors, commonly appearing in the uncovered regions, with no data. The

fundamental question is imposed about how to build robust systems that perform reasonably and safely in the real world. Achieving robustness is certainly a goal not only for the full-precision NN, but also for quantized neural network (QNN) [5]. Inspired jointly by bit optimization and robustness, the exploration and interpretation of the different quantization principles behind the QNN remain as intriguing directions for future research.

A growing interest in deep neural network is directed towards the efficient inferencing and training using quantization [6]. In [7], it has been recently proved that with appropriately trained NNs using binary weights and activations, quite well performance could be achieved even by extremely low-precision networks, such as binarized neural network. It turns out that quantization is a desirable mechanism that can dictate the entire NN performance. A promising deployment of QNN in many contemporary approaches and artificial intelligence applications reduces the overall computational and memory cost [5]. As shown in [8], this is of great importance since the amount of data to be processed constantly grows so that more powerful hardware (CPU/GPU) is required. Moreover, the QNN approach addressed in [5] is particularly beneficial for implementing in models with the extreme memory requirements, such as mobile devices and edge devices.

Whereas many contemporary applications are inherently decentralized or heterogeneous, data are distributed throughout a network without an aggregation, due to various communication constraints [9]. It is implied that compressed or quantized, rather than raw, data are transferred [10–13]. In other words, quantization is seen as a preferred mechanism for typically indispensable bit reduction. In brief, quantization plays a prominent role in optimizing various data transmission models in the existing network solutions. There are numerous indications to make us believe that appropriate quantizer parameterization will move us closer to an optimization in the amount of the transferred data in bits. Often, the same quantizer is applied to a variety of signals with distinct statistical features. For this reason, it is desirable to design the quantizer to be robust to the mentioned features of the input signal. In numerous literature studies, it has been assumed that the underlying distributions are known and with a specified parametric form [10, 14–26]. However, in many situations, typically, there is no prior information about the distribution of the signal to be processed, so that one can accept widespread opinion that the Laplacian distribution fits many natural, economical, and social phenomena [27]. Broadly speaking, as a consequence of the partial or total lack of information about the distribution, accepting some of distributions, such as the Laplacian distribution, offers the opportunity for a more complete view of the input attributes and the relations among their representatives. This explains our choice of considering the Laplacian distribution in the analysis presented in the paper.

Numerous papers have given a general framework to stimulate discussion on many different possible interpretations of the impact of the number of quantization levels and the support region width on both the granular and the overload distortion of the quantizer designed for the assumed Laplacian source [16–18, 21, 22, 25]. Having in mind the definition of the support region, as the area where distortion is small or at least appropriately bounded [11, 22], the main trade-off in quantizer design is done between an accommodation to the signal's amplitude dynamic and minimization of the overall distortion. In this regard, as shown in [22], the support region threshold is probably the most striking feature of the quantizer. To this end, we address the problem of determining this striking feature of our SQQ, as well as how to perform its parameterization in order to qualify our novel SQQ model for sources having Laplacian probability density function (pdf) with improved overall quantization efficiency.

One of the main goals of a quantizer designing or parameterization is to provide minimal possible distortion, that is, maximal possible SQNR, for a given bit rate, or equally, for a given number of quantization levels N. Designing an optimal N-level scalar quantizer consists of determining the set of $N+1$ decision thresholds and the set of N representation levels minimizing distortion that measures quantizer performance [11, 12]. In general, as we highlighted in [20], it is extremely hard to determine the global minimum of distortion function simultaneously with respect to all the decision thresholds and representation levels. However, Lloyd and Max found that it is possible to solve two partial problems [11]. The first problem is to determine the optimal representation levels for a given set of the decision thresholds. The second problem is to determine the optimal decision thresholds given the set of representation levels. In doing so, Lloyd and Max developed an iterative algorithm for designing an optimal quantizer for a source with known pdf. As stated in [11], the necessary and sufficient conditions for optimal quantizers are that each decision threshold is the midpoint of the adjacent representation levels and that each representation level is the centroid of its respective cell with respect to the given pdf. Due to the mutual dependence of these key quantizer parameters, the optimal quantizer has to be designed iteratively by applying the Lloyd–Max algorithm [11]. However, this algorithm is too complex and time-consuming and its application is due to the prominent problem with the algorithm initialization mainly limited to very low bit rates. Accordingly, novel quantizers with lower design complexities are very beneficial. Taking into account the fact that SCQ is significantly simpler to design than Lloyd–Max's quantizer [11] and that SCQ has worse performance, especially for small and medium bit rates, in this paper we came up with the idea to determine the set of decision thresholds according to the SCQ model and to determine the optimal set of N representation levels for the given set of decision thresholds and the assumed Laplacian pdf, in a similar manner as specified by Lloyd–Max's quantizer. In brief, the outcome of this paper is one completely novel model of scalar quantizer, called SQQ, which is not only simple to design, since it does not require iterative parameterization, but also outperforms asymptotically optimal SCQ in terms of SQNR. Unlike with the SCQ model, where representation levels are determined from the compressor

function, in our SQQ each representation level is determined as the centroid of the corresponding quantization cell with respect to the assumed Laplacian pdf. An important aspect of our interest in SQQ design is that it dictates the assumptions one can make on the statistics of the input signal, emerging from maximal SQNR. In particular, recasting the bit optimization problem, as the mean squared error (MSE) prediction problem, we offer the manner for direct computation of the support region threshold and SQNR from the formulas derived in the paper with the goal of characterizing the performance of our SQQ and providing a fair comparison with other well-known quantizer models. Moreover, as we have recently highlighted in [16] that, in general case, the clipping process formulation can be directly related to the support region determination, our anticipation in this paper is that such an offline precalculated value of support region can be stored in a lookup table for a fast retrieval, which can be indeed beneficial in numerous applications.

The paper is organized as follows. Section 2 provides an insight into the Laplacian pdf and briefly summarizes known facts from the literature about it. The main results of the paper are given in Sections 3 and 4. In particular, parameterization of our SQQ is presented in Section 3, whereas in Section 4, the discussion on our results and the comparison with the up-to-date related results are provided. Finally, in Section 5, we summarize the paper goals and we conclude with our research results.

In order to improve the readability of the rest of the paper, we opt to summarize here the list of abbreviations and notations we utilize: Symmetric Quantile Quantizer (SQQ); Scalar Companding Quantizer (SCQ); signal-to-quantization noise ratio (SQNR); neural network (NN); quantized neural network (QNN); probability density function (pdf); mean squared error (MSE); $p(x)$: Laplacian pdf of zero mean and unit variance; N: the number of representation levels; $\lambda(x)$: quantization level density function; $c^{\mathrm{SQQ}}(x)$: compressor function of SQQ; x_i^{SQQ}: decision thresholds of SQQ; y_i^{SQQ}: representation levels of SQQ; Δ_i^{SQQ}: distance between neighbouring decision thresholds of SQQ; D_g^{SQQ}: granular distortion of SQQ; $D_{\mathrm{over}}^{\mathrm{SQQ}}$: overload distortion of SQQ; $D_{\mathrm{Asym}}^{\mathrm{SQQ}}$: asymptotic distortion of SQQ; $D_{\mathrm{Exact}}^{\mathrm{SQQ}}$: exact distortion of SQQ; x_i^{SCQ}: decision thresholds of SCQ; y_i^{SCQ}: representation levels of SCQ; D_g^{SCQ}: granular distortion of SCQ; $D_{\mathrm{over}}^{\mathrm{SCQ}}$: overload distortion of SCQ; D^{SCQ}: total asymptotic distortion of SCQ; and $D_{\mathrm{over}}^{\mathrm{B.I.}}$: overload distortion of SQQ specified by Bennett's integral application.

2. Why Laplacian pdf?

Legacy voice is mature, robust, and high performing. Voice service continues to be globally forecasted as the challenge, which traditionally has a relatively predictable and small use of bandwidth but requires real-time transmission. It is well known that the distinctive attributes of speech signals are well modelled by the Laplacian pdf or the Gaussian one [10, 11, 27]. In addition, the most common assumption is that the pixel difference values can be described by either the Laplacian or the generalized Gaussian pdf [12]. However, the Laplacian pdf is more tailed than a Gaussian one. Moreover, the Laplacian pdf provides typically tractable closed-form formulas necessary for the design and performance assessment of various transmission systems. This tractability is often impossible in the case where one or more Gaussian pdfs are assumed [28], so that approximations of the special functions, such as erfc function and Q-function, or approximation of the compressor function, are required [29, 30]. For that reason, Laplacian pdf is a more preferable pdf, which has been recently used to model not only weights and activations in NNs, but also speech and audio signals [15–20, 25, 31]. Namely, according to [10], the Laplacian pdf (see Figure 1) is arguably the most suitable form for modelling statistical properties of both audio and speech signals. Since audio signals may exhibit varying degrees of stationarity, it is often advantageous to allow for flexible coders backed by the flexible transmission rates, able to adjust in real time to continuously time-varying signal attributes. As known, Laplacian and Gaussian pdfs belong to the family of the log-concave functions. The concept of log-concavity was widely studied in the literature for many symmetric distributions such as the Laplacian one, for instance, in [14]. The log-concavity can facilitate the accurate anticipation of the quantizer design from the standpoint of its symmetry. According to [14], as the Laplacian pdf satisfies the log-concavity condition, the optimal quantizer is itself symmetric. More generally, the Laplacian pdf can be used to model symmetric shaped data, such as weights used in NN [15]. Rather than striving for a full precision, it has been shown in [15] that, by quantizing the NN into low-bit representation and with an analytical clipping assessment, negligible accuracy degradation can be incurred. As we have already mentioned that, in general case, the clipping process formulation can be directly related to the support region determination, we believe that such an offline precalculated value of the support region threshold can be stored in a lookup table for a fast retrieval, which can be indeed beneficial. Given the acceptable quantization error, the question arises whether the approximated quantized network still has the same ability to extract features from the input data. The outlined clipping assessment task is clearly possible by minimizing the MSE. The main reason why we deal with the MSE is because the error function for the observed Laplacian pdf is a convex function [11]. To our knowledge, the convexity of the MSE distortion has recently received a lot of attention in [17]. Essentially, one of the preferences of convexity is that the error of averaging two guesses is always lower than the average error from the same two guesses. The lack of convexity has been seen and proved for odd numbers of quantizer representation level, while for any symmetric density, such as the Laplacian, the MSE is convex for even numbers of representation levels [17].

Before moving on to the quantizer modelling, we should highlight here that some of the bases in our further analysis are straight forwarded by the rigorously stated partial distortion theorem in the case of the Laplacian pdf [18]. We should also highlight that some novel insights in the results of the abovementioned approach applied in our SQQ design are provided in this paper, which could be of great

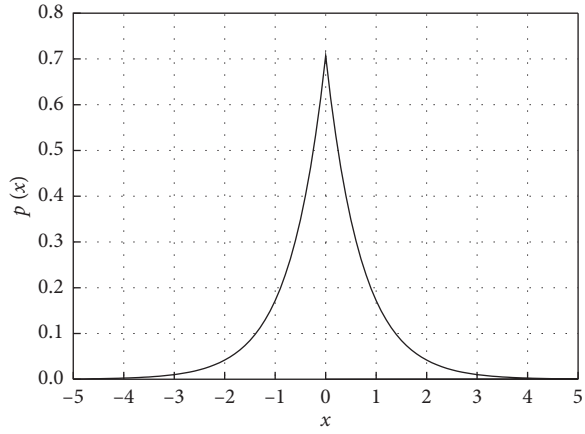

Figure 1: Laplacian pdf of zero mean and unit variance.

significance in the further analysis of quantization and in qualification for contemporary quantization solutions.

3. Symmetric Quantile Quantizer Designed for the Laplacian pdf

Signal coding generally refers to the process of extracting and transforming of the raw analog information from the signal, into digitally coded representations, to facilitate the access to descriptive and invariant attributes of this signal [11]. Signal decoding, as the reciprocal process of signal coding, on the receiver end, has the main task to obtain the closest possible signal to an original signal, using the extracted attributes. Well-designed coder or quantizer can reduce the potential influence of unpredictable statistical characteristics of the signals to the quality, thereby contributing to increased robustness. Following the main aspect of data coding and compression and a constant intention to decrease the number of bits necessary to deliver original analog signal, companding quantizers, based on some compression laws [11, 19–21, 32], have been receiving much research interest, especially due to the fact that they are well suited in terms of complexity of designing and implementation, as shown in [33].

In general, an input signal is divided by the quantization procedure into a granular region and an overload region, or alternatively, in an inner and an outer region, which are for the symmetric quantizer separated by the support region thresholds denoted by $-x_{\max}$ and $x_{\max}$, respectively [22] (see Figure 2). Both analyses for the simplest quantizer model, that is for uniform quantizer, emerged recently, are authored by Sangsin Na et al. [18, 25]. Paper [25] deals with the monotonicity-by-two for the symmetric uniform quantizers in the case of the four most common pdfs, including Laplacian pdf, we observe here. Another interesting fact herein stated and proved for the Laplacian pdf and sufficiently large number of quantization levels N is that the optimal step size for the uniform quantizer decreases, as bit rate increases by one. Initially derived from "partial distortion theorem," an analysis given in [18] uncovers several insights about the relationship between the MSE distortion, number and position of quantization levels in inner and outer region, or granular and overload region.

One of the conclusions from [18] is that microscopical contribution to the total distortion is the same from each cell, whereas microscopically is more from outer cells. In what follows, we will ascertain some novel insights in the results of the abovementioned approach applied in our SQQ design.

Let us consider the variance-matched case of our interest in which the Laplacian pdf of zero mean and unit variance is assumed:

$$p(x) = \frac{1}{\sqrt{2}} \exp\{-\sqrt{2}|x|\}. \tag{1}$$

Referring to the aforementioned log-concavity and the MSE convexity proved for the Laplacian pdf [14], of particular interest, in this paper we propose one simple manner for the SQQ parameterization. Our N-level SQQ is defined by mapping Q: $\mathbb{R} \longrightarrow Y$ [22], where $Q(\bullet)$ is a symmetric characteristic of our quantizer, $\mathbb{R}$ is a set of real numbers, and $Y \equiv \{-y_{N/2}, \ldots - y_1, y_1, \ldots, y_{N/2}\} \subset \mathbb{R}$ is a set of representation levels (see Figure 2), which makes the code book of size $|Y| = N = 2K$. Subsequently, one can accept that, for every symmetric quantizer with an even number of nonoverlapping cells, quantization levels are symmetrically placed about the mean. Without losing of generality, we are focusing only on the K positive counterparts, as shown in Figure 3, where some of the positive counterparts for the odd compressor function $c^{\mathrm{SQQ}}(x)$ are presented.

The necessary parameters of our N-level SQQ are given as follows: {quantization level density function; compressor function; decision thresholds; representation levels; support region threshold; distortion}.

Quantization level density function of our novel model of quantizer named SQQ is given by

$$\lambda(x) = p^{1/3}(x) \left[2 \int_0^{x_K^{\mathrm{SQQ}}} [p(t)]^{1/3} \mathrm{d}t \right]^{-1}. \tag{2}$$

Foremost, the substitution of equation (1) in (2) yields

$$\lambda(x) = \left[3\sqrt{2} \exp\left\{ \frac{\sqrt{2}\,x}{3} \right\} \left(1 - \exp\left\{ -\frac{\sqrt{2}\,x_K^{\mathrm{SQQ}}}{3} \right\} \right) \right]^{-1}. \tag{3}$$

Our odd compressor function is $c^{\mathrm{SQQ}}(x)$: $[0,+\infty) \longrightarrow [0, 1)$

$$c^{SQQ}(x) = 2 \int_0^x \lambda(t) \mathrm{d}t = \left(1 - \exp\left\{ -\frac{\sqrt{2}\,x}{3} \right\} \right) \left(1 - \exp\left\{ -\frac{\sqrt{2}\,x_K^{\mathrm{SQQ}}}{3} \right\} \right)^{-1}, \quad x \geq 0. \tag{4}$$

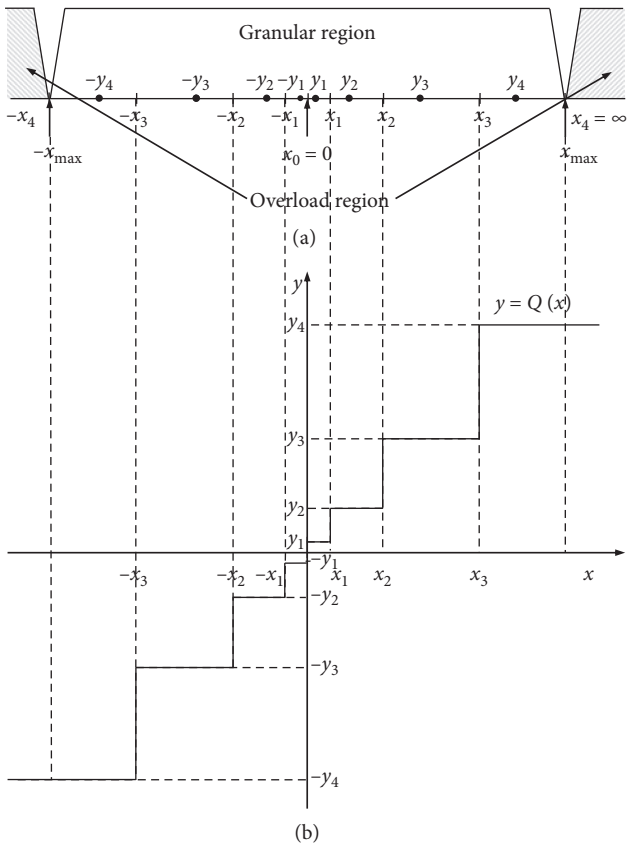

FIGURE 2: (a) Granular and overload regions of the symmetric nonuniform quantizer with $N = 8$ quantization levels. (b) Characteristic of the nonuniform quantizer $Q(x)$ for $N = 8$.

Nonnegative decision thresholds of our SQQ are determined from the compressor function as

$$\begin{aligned} x_i^{\mathrm{SQQ}} &= c^{\mathrm{SQQ}-1}\left(\frac{i}{K}\right), \quad i = 0, 1, \ldots, K-1, \\ x_0^{\mathrm{SQQ}} &= 0, \\ x_K^{\mathrm{SQQ}} &= +\infty, \\ x_{-i}^{\mathrm{SQQ}} &= -x_i^{\mathrm{SQQ}}, \quad i = 1, 2, \ldots, K, \end{aligned} \tag{5}$$

whereas the negative counterparts explicitly follow from the symmetry. Unlike with the SCQ model, where representation levels are also determined from the compressor function, in our SQQ, each representation level is determined as the centroid of the corresponding quantization cell $[x_{i-1}^{\mathrm{SQQ}}, x_i^{\mathrm{SQQ}})$ with respect to $p(x)$:

$$\begin{aligned} y_i^{\mathrm{SQQ}} &= E_p\left\{X | x_{i-1}^{\mathrm{SQQ}} \le X < x_i^{\mathrm{SQQ}}\right\}, \\ y_{-i}^{\mathrm{SQQ}} &= -y_i^{\mathrm{SQQ}}, \quad i = 1, 2, \ldots, K, \end{aligned} \tag{6}$$

where X and $E_p\{\cdot\}$ denote the continuous random variable with pdf $p(x)$ and expectation with respect to $p(x)$:

$$y_i^{\mathrm{SQQ}} = \int_{x_{i-1}^{\mathrm{SQQ}}}^{x_i^{\mathrm{SQQ}}} x p(x)\mathrm{d}x \left[\int_{x_{i-1}^{\mathrm{SQQ}}}^{x_i^{\mathrm{SQQ}}} p(x)\mathrm{d}x\right]^{-1}, \quad i = 1, 2, \ldots, K, \tag{7}$$

$$y_K^{\mathrm{SQQ}} = x_{K-1}^{\mathrm{SQQ}} + \frac{1}{\sqrt{2}}. \tag{8}$$

For the assumed pdf and for $i = 1,2, \ldots, K-1$, we derive

$$y_i^{\mathrm{SQQ}} = x_{i-1}^{\mathrm{SQQ}} + \frac{1}{\sqrt{2}} - \frac{\Delta_i^{\mathrm{SQQ}} \exp\left\{-\sqrt{2}\,\Delta_i^{\mathrm{SQQ}}\right\}}{1 - \exp\left\{-\sqrt{2}\,\Delta_i^{\mathrm{SQQ}}\right\}}, \tag{9}$$

$$\Delta_i^{\mathrm{SQQ}} = x_i^{\mathrm{SQQ}} - x_{i-1}^{\mathrm{SQQ}} = \frac{3}{\sqrt{2}} \ln\left(\frac{K-i+1}{K-i}\right), \tag{10}$$

where Δ_i^{SQQ} is a distance between neighbouring decision thresholds of our SQQ.

Let us substitute $x_K^{\mathrm{SQQ}} = +\infty$ in equations (3) and (4) so that the corresponding quantization level density function and compressor function of our SQQ are derived as follows:

$$\begin{aligned} \lambda(x) &= \frac{1}{3\sqrt{2}} \exp\left\{-\frac{\sqrt{2}\,x}{3}\right\}, \\ c^{\mathrm{SQQ}}(x) &= \left(1 - \exp\left\{-\frac{\sqrt{2}\,x}{3}\right\}\right), \quad x \ge 0. \end{aligned} \tag{11}$$

Accepting such a defined $c^{\mathrm{SQQ}}(x)$, for the nonnegative decision thresholds in our SQQ defined by equation (5), we obtain

$$x_i^{\mathrm{SQQ}} = \frac{3}{\sqrt{2}} \ln\left(\frac{K}{K-i}\right), \quad i = 0, 1, \ldots, K-1, \tag{12}$$

and for $i = K-1$, from equations (5) and (12), we directly derive the closed-form formula for clipping:

$$x_{K-1}^{\mathrm{SQQ}} = \frac{3}{\sqrt{2}} \ln(K). \tag{13}$$

By applying Bennett's integral [23] to approximate the granular distortion of our novel SQQ

$$D_g^{\mathrm{SQQ}} = \frac{2\Delta^2}{12} \int_0^{x_{K-1}^{\mathrm{SQQ}}} \left[\frac{\partial c^{\mathrm{SQQ}}(x)}{\partial x}\right]^{-2} p(x)\mathrm{d}x, \tag{14}$$

where in our model it holds $\Delta = 2/N = 1/K$, we derive

$$D_g^{\mathrm{SQQ}} = \frac{9}{2N^2}\left(1 - \exp\left\{-\frac{\sqrt{2}\,x_{K-1}^{\mathrm{SQQ}}}{3}\right\}\right). \tag{15}$$

Further substituting equation (13) in (15) gives us the closed-form formula for the granular distortion:

$$D_g^{\mathrm{SQQ}} = \frac{9}{2N^2}\left(1 - \frac{2}{N}\right). \tag{16}$$

To calculate the overload distortion, we use y_K^{SQQ} obtained from equation (8):

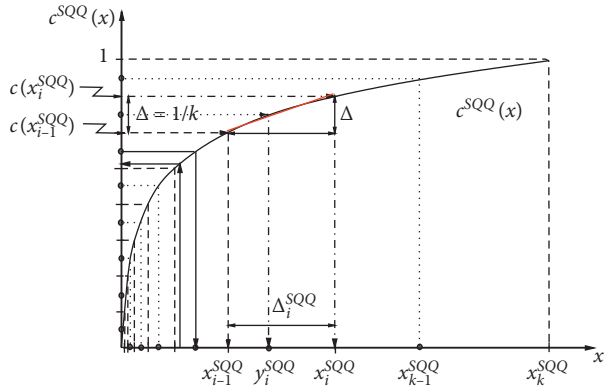

Figure 3: Positive counterparts for the odd compressor function $(c)^{SQQ}(x)$.

$$D_{\text{over}}^{\text{SQQ}} = 2\int_{x_{K-1}^{\text{SQQ}}}^{x_K^{\text{SQQ}}} \left(x - y_K^{\text{SQQ}}\right)^2 p(x)\mathrm{d}x, \tag{17}$$

$$D_{\text{over}}^{\text{SQQ}} = \frac{\left(x_{K-1}^{\text{SQQ}} + (1/\sqrt{2})\right)^2 + 0.5 - y_K^{\text{SQQ}}\left(2x_{K-1}^{\text{SQQ}} + \sqrt{2}\right) + \left(y_K^{\text{SQQ}}\right)^2}{\exp\left\{\sqrt{2}\, x_{K-1}^{\text{SQQ}}\right\}} = \frac{1}{2}\exp\left\{-\sqrt{2}\, x_{K-1}^{\text{SQQ}}\right\}. \tag{18}$$

Direct substitution of equation (13) in (18) then yields

$$D_{\text{over}}^{\text{SQQ}} = \frac{4}{N^3}. \tag{19}$$

Eventually, by summing equations (16) and (19) for the total distortion, composed of the granular and the overload distortion, we obtain

$$D_{\text{Asym}}^{\text{SQQ}} = \frac{9}{2N^2}\left(1 - \frac{10}{9N}\right), \tag{20}$$

where subscript Asym refers to the application of Bennett's integral in our asymptotic analysis. In order to ascertain the accuracy of the derived asymptotic formula for the overall distortion in what follows, we derive one beneficial exact formula for the overall distortion $D_{\text{Exact}}^{\text{SQQ}}$, for the Laplacian pdf of zero mean and unit variance.

Let us start our derivation with the general formula for the exact distortion for any even N value:

$$D_{\text{Exact}}^{\text{SQQ}} = 2\sum_{i=1}^{K}\int_{x_{i-1}^{\text{SQQ}}}^{x_i^{\text{SQQ}}} \left(x - y_i^{\text{SQQ}}\right)^2 p(x)\mathrm{d}x, \tag{21}$$

which can be rewritten as

$$D_{\text{Exact}}^{\text{SQQ}} = 2\sum_{i=1}^{K}\int_{x_{i-1}^{\text{SQQ}}}^{x_i^{\text{SQQ}}} x^2 p(x)\mathrm{d}x - 4\sum_{i=1}^{K}\int_{x_{i-1}^{\text{SQQ}}}^{x_i^{\text{SQQ}}} x y_i^{\text{SQQ}} p(x)\mathrm{d}x + 2\sum_{i=1}^{K}\int_{x_{i-1}^{\text{SQQ}}}^{x_i^{\text{SQQ}}} \left(y_i^{\text{SQQ}}\right)^2 p(x)\mathrm{d}x, \tag{22}$$

where

$$\int_{-\infty}^{+\infty} x^2 p(x)\mathrm{d}x = 1, \tag{23}$$

so that we derive

$$D_{\text{Exact}}^{\text{SQQ}} = 1 - 2\sum_{i=1}^{K-1}\left(y_i^{\text{SQQ}}\right)^2 P_i^{\text{SQQ}},$$

$$P_i^{\text{SQQ}} = \int_{x_{i-1}^{\text{SQQ}}}^{x_i^{\text{SQQ}}} p(x)\mathrm{d}x = \frac{\exp\left\{-\sqrt{2}\, x_{i-1}^{\text{SQQ}}\right\} - \exp\left\{-\sqrt{2}\, x_i^{\text{SQQ}}\right\}}{2}. \tag{24}$$

By using equation (12), we finally derive

$$P_i^{\text{SQQ}} = \frac{1}{2K^3}\left(3(K-i)^2 + 3(K-i) + 1\right),$$

$$D_{\text{Exact}}^{\text{SQQ}} = 1 - 2\sum_{i=1}^{K-1}\left(y_i^{\text{SQQ}}\right)^2 \frac{1}{2K^3}\left(3(K-i)^2 + 3(K-i) + 1\right). \tag{25}$$

To conclude, the necessary parameters of our N-level SQQ are given as follows:

$$\left\{\lambda(x)=\frac{1}{3\sqrt{2}}\exp\left\{-\frac{\sqrt{2}\,x}{3}\right\};c^{\mathrm{SQQ}}(x)=\left(1-\exp\left\{-\frac{\sqrt{2}\,x}{3}\right\}\right),x\geq 0;x_i^{\mathrm{SQQ}}=\frac{3}{\sqrt{2}}\ln\left(\frac{K}{K-i}\right),i=0,1,\ldots,K-1;\right. \tag{26}$$
$$\left.y_i^{\mathrm{SQQ}}=E_p\left\{X|x_{i-1}^{\mathrm{SQQ}}\leq X<x_i^{\mathrm{SQQ}}\right\},i=1,2,\ldots,K;x_{K-1}^{\mathrm{SQQ}}=\frac{3}{\sqrt{2}}\ln(K);D_{\mathrm{Asym}}^{\mathrm{SQQ}}=\frac{9}{2N^2}\left(1-\frac{10}{9N}\right)\right\}.$$

Quantizer parameterization dictates its entire performance. If decision thresholds and representation levels are chosen more appropriately, the overall distortion is smaller; that is, SQNR is higher, which translates to a reduction in the number of bits required from the quantizer for achieving a certain SQNR. Let us highlight here that SQQ is a modified form of nonuniform quantizer named as compandor. Namely, compandor or SCQ is conceptually realized as cascade of three intrinsic functional blocks: compressor, uniform quantizer, and expandor [11]. In particular, in the considered SQQ model, a precedence of compandor model for the decision thresholds calculation is used, while the representation levels are optimally defined, so that it makes the basic difference with SCQ model where both decision thresholds and representation levels are determined in accordance with the specified compressor function. Therefore, we highlight here that our SQQ and SCQ models are characterized with the similar design and implementation complexity level so that both models can be used for the efficient Lloyd–Max algorithm initialization [24]. Moreover, SQQ and SCQ are characterized by the elegant determining of asymptotic quantizer's performances for the Laplacian pdf.

Towards determining the cases of interest where the proposed SQQ outperforms SCQ for the same support region threshold value $x_{K-1}^{\mathrm{SQQ}}=x_{K-1}^{\mathrm{SCQ}}$, we define the necessary parameters of N-level SCQ as

$$\left\{\lambda(x)=\frac{1}{3\sqrt{2}}\exp\left\{-\frac{\sqrt{2}\,x}{3}\right\};c^{\mathrm{SCQ}}(x)=\left(1-\exp\left\{-\frac{\sqrt{2}\,x}{3}\right\}\right),x\geq 0;x_i^{\mathrm{SCQ}}=c^{\mathrm{SCQ}^{-1}}\left(\frac{i}{K}\right),i=0,1,\ldots,K-1\right. \tag{27}$$
$$\left.y_i^{\mathrm{SCQ}}=c^{\mathrm{SCQ}^{-1}}\left(\frac{2i-1}{2K}\right),i=1,2,\ldots,K;x_{K-1}^{\mathrm{SCQ}}=\frac{3}{\sqrt{2}}\ln(K);D^{\mathrm{SCQ}}=\frac{9}{2N^2}-\frac{1-8\ln 8\cdot(\ln\sqrt{8}-1)}{N^3}\right\}.$$

We can highlight here that the above described novel SQQ model and the appropriate formulas derived for this model are main results of this paper. In what follows, we show that the overload distortion of the SCQ is not equal to the one of the SQQ. Specifically, for the support region threshold x_{K-1}^{SCQ} of SCQ, designed for the Laplacian pdf of zero mean and unit variance, in the following, we determine the overall distortion and we show that it differs from the distortion of SQQ given by equation (20).

Briefly, we omit the complete proof and only highlight its main steps:

(i) Recalling equations (14), (16), and (17) and discussions about the granular distortion of the SQQ, from the condition $x_{K-1}^{\mathrm{SQQ}}=x_{K-1}^{\mathrm{SCQ}}$, it follows $D_g^{\mathrm{SQQ}}=D_g^{\mathrm{SCQ}}$. Paying in-depth attention to these equations, we can intuitively propose how to calculate the granular distortion $D_g^{\mathrm{SCQ}}=(9/2N^2)(1-(2/N))$.

(ii) Repeat the calculation of y_K^{SCQ}, where it holds $c(y_K^{\mathrm{SCQ}})=(N-1)/N$, so that we derive $y_K^{\mathrm{SCQ}}=(3/\sqrt{2})\ln(N)$.

(iii) Note that y_K^{SCQ} and y_K^{SQQ} have direct effect on the overload distortion.

(iv) Applying y_K^{SCQ} in equation (17) gives us $D_{\mathrm{over}}^{\mathrm{SCQ}}$ in the following form:

$$D_{\mathrm{over}}^{\mathrm{SCQ}}=2\int_{x_{K-1}^{\mathrm{SQQ}}}^{x_K^{\mathrm{SQQ}}}\left(x-y_K^{\mathrm{SCQ}}\right)^2p(x)\mathrm{d}x=\frac{8[1+\ln 8\cdot(\ln\sqrt{8}-1)]}{N^3}. \tag{28}$$

Eventually, the total distortion and SQNR of SCQ are

$$D^{\mathrm{SCQ}} = D_g^{\mathrm{SCQ}} + D_{\mathrm{over}}^{\mathrm{SCQ}} = \frac{9}{2N^2} - \frac{1 - 8\ln 8 \cdot (\ln\sqrt{8} - 1)}{N^3},$$
$$\mathrm{SQNR}^{\mathrm{SCQ}} = -10\log_{10}\left(D^{\mathrm{SCQ}}\right). \tag{29}$$

4. Numerical Results and Analysis

As earlier outlined in Section 3, we analyze the Laplacian SQQ with an even number of quantization levels, where each level is the centroid of the respective cell. Considering both the overload and the granular region, we can adopt that the large, compressed signal values occur very rarely compared to the small values. For our SQQ analysis refinement, it seems reasonable to accept the monotonicity of the neighbouring cells, which stems from the cell size determination, as illustrated in Figure 3. Note that, from equation (10), we can conclude that Δ_i^{SQQ} decreases strictly monotonically to Δ_0^{SQQ} as i decreases, since the source pdf has a finite support x_{K-1}^{SQQ} and $\Delta_{K-1}^{\mathrm{SQQ}}$ is the cell of a finite length; i.e., this is the cell that lies the most outwardly in the granular region.

With the increase of N ($N \longrightarrow \infty$) and with the neglection of the overload distortion due to narrowing the overload region, one can notice that equation (20) approaches the PD high-resolution formula $D^{\mathrm{PD}} = 9/(2N^2)$ [26]:

$$D_N^{\mathrm{SQQ}} | N \longrightarrow \infty = \lim_{N\longrightarrow\infty}\left(\frac{9}{2N^2}\left(1 - \frac{10}{9N}\right)\right) = \frac{9}{2N^2}. \tag{30}$$

We can explicitly compute the overload distortion by applying Bennett's integral [20] and adopting x_{K-1}^{SQQ} from equation (13):

$$D_{\mathrm{over}}^{\mathrm{B.I.}} = \frac{2\Delta^2}{12}\int_{x_{K-1}^{\mathrm{SQQ}}}^{x_K^{\mathrm{SQQ}}}\left[\frac{\partial c^{\mathrm{SQQ}}(x)}{\partial x}\right]^{-2} p(x)\mathrm{d}x, \tag{31}$$

$$D_{\mathrm{over}}^{\mathrm{B.I.}} = \frac{9}{N^3}. \tag{32}$$

From equations (30) and (32), one can conclude that it is not cumbersome or demanding to derive the following formula for the granular distortion of our SQQ:

$$D_g^{\mathrm{SQQ}} = D_N^{\mathrm{SQQ}} | N \longrightarrow \infty - D_{\mathrm{over}}^{\mathrm{B.I.}} = \frac{9}{2N^2}\left(1 - \frac{2}{N}\right), \tag{33}$$

which is expectedly equal to equation (16).

For the sake of the overload distortion analysis, we can compare the formulas given in equations (19) and (32), with the one in equation (28) that we derived for the companding quantizer. Obviously, the following inequality holds: $D_{\mathrm{over}}^{\mathrm{SQQ}} < D_{\mathrm{over}}^{\mathrm{SCQ}} < D_{\mathrm{over}}^{\mathrm{B.I.}}$. This is indeed one of the beneficial features of the proposed SQQ model since it introduces smaller distortion compared to the asymptotically optimal quantizer model, that is, compared to SCQ.

Having in mind that relative competitive relations defined as

$$\delta_{\mathrm{over}}^{\mathrm{SQQ/B.I.}} = \left|\frac{D_{\mathrm{over}}^{\mathrm{SQQ}} - D_{\mathrm{over}}^{\mathrm{B.I.}}}{D_{over}^{\mathrm{SQQ}}}\right| \cdot 100 = 125\,[\%],$$
$$\delta_{\mathrm{over}}^{\mathrm{SCQ/B.I.}} = \left|\frac{D_{\mathrm{over}}^{\mathrm{SCQ}} - D_{\mathrm{over}}^{\mathrm{B.I.}}}{D_{\mathrm{over}}^{\mathrm{SCQ}}}\right| \cdot 100 = 3,93\,[\%] \tag{34}$$

do not depend on N, we can highlight here that this interesting notice deserves further attention, especially when the importance of the overload distortion is more pronounced. We can also highlight here that, as long as the overload distortion is not predominantly negligible to the granular distortion, $D_{\mathrm{over}}^{\mathrm{SQQ}}$ provides significant advantage over $D_{\mathrm{over}}^{\mathrm{SCQ}}$ and $D_{\mathrm{over}}^{\mathrm{B.I.}}$.

An important aspect of our interest in SQQ design, as a means of data conversion or compression is, that it dictates the assumptions one can make on the statistics of the input signal, emerging from maximal SQNR. In particular, by recasting the bit optimization problem, as the MSE prediction problem, we can derive SQNR directly from the affordable formulas characterizing its performance, given in equations (20) and (21):

$$\mathrm{SQNR}_{\mathrm{Asym}}^{\mathrm{SQQ}} = -10\log_{10}\left(D_{\mathrm{Asym}}^{\mathrm{SQQ}}\right), \tag{35}$$

$$\mathrm{SQNR}_{\mathrm{Exact}}^{\mathrm{SQQ}} = -10\log_{10}\left(D_{\mathrm{Exact}}^{\mathrm{SQQ}}\right). \tag{36}$$

In order to provide fare performance comparison, we further define relative errors:

$$\delta^A = \left|\frac{\mathrm{SQNR}_{\mathrm{Asym}}^{\mathrm{SQQ}} - \mathrm{SQNR}_{\mathrm{Exact}}^{\mathrm{SQQ}}}{\mathrm{SQNR}_{\mathrm{Exact}}^{\mathrm{SQQ}}}\right|,$$
$$\delta^{\mathrm{PD}} = \left|\frac{\mathrm{SQNR}^{\mathrm{PD}} - \mathrm{SQNR}_{\mathrm{Exact}}^{\mathrm{SQQ}}}{\mathrm{SQNR}_{\mathrm{Exact}}^{\mathrm{SQQ}}}\right|. \tag{37}$$

The values of SQNR presented in Table 1 show that $\mathrm{SQNR}_{\mathrm{Asym}}^{\mathrm{SQQ}}$ provides significant performance gain when compared to $\mathrm{SQNR}^{\mathrm{PD}}$. It is worthy to notice that, for $N \geq 128$, $\mathrm{SQNR}_{\mathrm{Asym}}^{\mathrm{SQQ}}$ achieves roughly the same level of accuracy as that of $\mathrm{SQNR}_{\mathrm{Exact}}^{\mathrm{SQQ}}$ or $\mathrm{SQNR}^{\mathrm{PD}}$, for the same number of levels N, which we have intuitively expected. Our approach of determining distortion, or directly related SQNR, $\mathrm{SQNR} = -10\log_{10}(D)$) [11] provides a more complete insight into the impact of both the granular and the overload distortion to the total distortion.

In what follows, we provide the following observations about some of the results of our analysis and the performances of the proposed SQQ model:

Table 1: The comparison of SQNR values for the Asym, Exact, and PD approach and relative error overview for the Asym and PD approach.

N	$\mathrm{SQNR}_{\mathrm{Asym}}^{\mathrm{SQQ}}[dB]$	$\mathrm{SQNR}_{\mathrm{Exact}}^{\mathrm{SQQ}}[dB]$	$\mathrm{SQNR}^{\mathrm{PD}}[dB]$	δ^{A}	δ^{PD}
4	6.9224	7.1865	5.5091	0.0368	0.2334
6	9.9203	10.1329	9.0309	0.0210	0.1087
8	12.1791	12.3501	11.5297	0.0138	0.0664
10	13.9794	14.1212	13.4679	0.0100	0.0463
12	15.4734	15.5943	15.0515	0.0077	0.0348
14	16.7496	16.8547	16.3904	0.0062	0.0275
16	17.8629	17.9558	17.5503	0.0052	0.0226
18	18.8500	18.9333	18.5733	0.0044	0.0190
20	19.7367	19.8121	19.4885	0.0038	0.0163
28	22.5869	22.6416	22.4110	0.0024	0.0102
30	23.1742	23.2254	23.0103	0.0022	0.0093
32	23.7243	23.7725	23.5709	0.0020	0.0085
64	29.6675	29.6919	29.5915	0.0008	0.0034
128	35.6499	35.6622	35.6121	0.0003	0.0014
256	41.6516	41.6577	41.6327	0.0001	0.0006

(1) SQQ model introduces smaller distortion compared to the observed asymptotically optimal quantizer model, that is, compared to SCQ.

(2) SQQ is very simple to design since it does not require iterative parameterization as it is the case with the optimal Lloyd–Max quantizer.

(3) Although outputting approximate values, the derived asymptotic formula for $\mathrm{SQNR}_{\mathrm{Asym}}^{\mathrm{SQQ}}$ is very accurate, even for a small N value, which make it very beneficial.

(4) Since δ^{A} and δ^{PD} amount approximately less than or equal to 1% for $N = 10$ and $N = 30$, respectively, the correctness of the analysis performed for our SQQ model for small and medium bit rates is confirmed. This is of great importance since the accuracy of the asymptotic theory decreases noticeably for the classes of small or medium numbers of the quantization cells N.

(5) The values of N at which relative error δ^{A} increases steeply are in the very narrow range $N \in [4, 8]$.

(6) The largest relative error $\delta^{\mathrm{A}} = 3.68\%$ occurs for the smallest observed number of levels $N = 4$.

To perceive the model of SQQ and SCQ, we perform an analysis for a wide-ranged set of the quantization levels, where for SQQ we specify a relative competitive benefit over the high-resolution theory (asymptotic theory), that is, the performance gain:

$$G^{\mathrm{SQQ-PD}} = 10\log_{10}\left(\frac{D^{\mathrm{PD}}}{D^{\mathrm{SQQ}}}\right) = 10\log_{10}\left(\frac{9N}{9N-10}\right), \quad (38)$$

whereas to provide the comparison with SCQ we specify the following performance gain:

$$G^{\mathrm{SQQ-SCQ}} = 10\log_{10}\left(\frac{D^{\mathrm{SCQ}}}{D^{\mathrm{SQQ}}}\right) = 10\log_{10}\left(\frac{9N - 2\left(1 - 8\ln 8 \cdot \left(\ln\sqrt{8} - 1\right)\right)}{9N-10}\right). \quad (39)$$

By observing Table 2, we can come to the conclusion that for every bit rate R ($R = \log_2 N$); it holds $G^{\mathrm{SQQ-PD}}[dB] > G^{\mathrm{SQQ-SCQ}}[dB]$ and that $G^{\mathrm{SQQ-PD}}[dB]$ ranges up to about 1.4 dB.

In what follows, we compare SQNR values and we analyze the SQNR curves overlapping along the finite part of their length. Note that curve overlapping is an illustrative indicator and hence, more specifically, our goal is to emphasize strong overlapping for $\mathrm{SQNR}_{\mathrm{Asym}}^{\mathrm{SQQ}}$ and $\mathrm{SQNR}_{\mathrm{Exact}}^{\mathrm{SQQ}}$, as well as $\mathrm{SQNR}^{\mathrm{PD}}$and $\mathrm{SQNR}^{\mathrm{SCQ}}$ when R ranges from 2 bits/sample to 5 bits/sample (Figure 4(a)). We can also observe that, as shown in Figure 2(b), for R ranging from 3 bits/sample almost to 5 bits/sample, the derived $\mathrm{SQNR}_{\mathrm{Asym}}^{\mathrm{SQQ}}$ overperforms other asymptotic solutions as $\mathrm{SQNR}^{\mathrm{PD}}$ and

Table 2: SQQ performance gain over the high-resolution theory and SCQ model.

N	$D^{\mathrm{SQQ}}_{\mathrm{Asym}}$	D^{PD}	D^{SCQ}	$G^{\mathrm{SQQ-PD}}[dB]$	$G^{\mathrm{SQQ-SCQ}}[dB]$
4	0.20313	0.28125	0.27595	1.4133	1.3307
6	0.10185	0.12500	0.12342	0.8894	0.8345
8	0.06055	0.07031	0.06965	0.6494	0.6083
10	0.04000	0.04500	0.04466	0.5115	0.4787
12	0.02836	0.03125	0.03105	0.4220	0.3946
14	0.02114	0.02296	0.02284	0.3591	0.3357
16	0.01636	0.01758	0.01750	0.3125	0.2921
18	0.01303	0.01389	0.01383	0.2767	0.2585
20	0.01063	0.01125	0.01121	0.2482	0.2318
32	0.00424	0.00440	0.00438	0.1535	0.1432
64	0.00108	0.00110	0.00110	0.0760	0.0710
128	0.000272	0.000275	0.000274	0.0379	0.0353
256	0.0000684	0.0000687	0.0000686	0.0189	0.0176

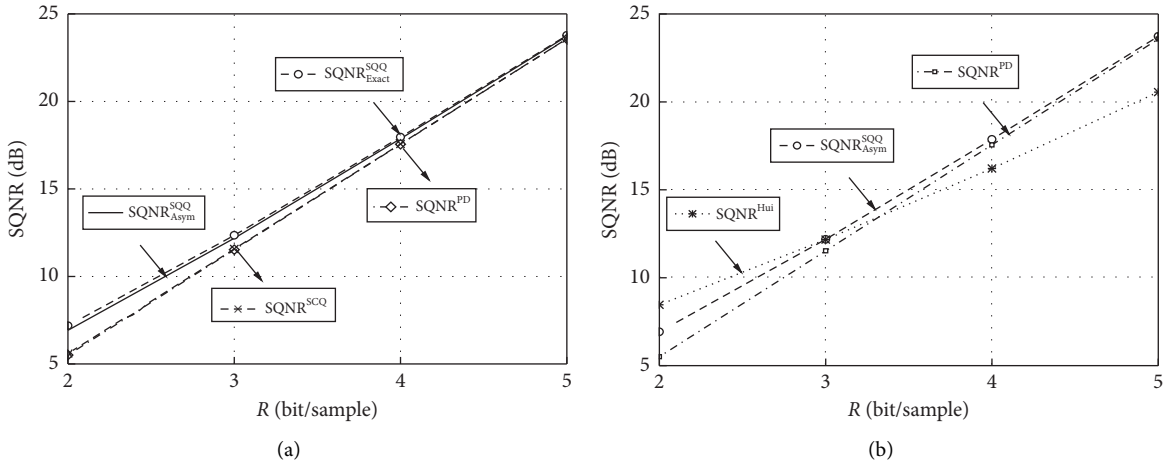

Figure 4: SQNR dependence on R for small and medium bit rates. (a) Application of equations (29), (35), (36), and (40). (b) Application of equations (35), (40), and (41).

$\mathrm{SQNR}^{\mathrm{Hui}}$ calculated by using the following equations from [26, 34], respectively.

$$\mathrm{SQNR}^{\mathrm{PD}} = -10\log_{10}\left(D^{\mathrm{PD}}\right) = 10\log_{10}\left(\frac{2N^2}{9}\right), \tag{40}$$

$$\mathrm{SQNR}^{\mathrm{Hui}} = 10\log_{10}\left(\frac{3N^2}{2\ln^2(N)+3}\right), \tag{41}$$

while for $R = 2$ bit/sample, $\mathrm{SQNR}^{\mathrm{Hui}}$ outperforms SQNR of our SQQ. This can be explained by the fact that, for such low bit rate, the value of the support region threshold, optimized in the asymptotic analysis for $N >> 1$, is more inaccurate in the case of a nonuniform quantizer than in the case of a uniform one, which is consequently reflected in the SQNR values of these two quantizer models.

From the analysis presented in the paper, one can conclude that the granular distortion can be applicable as the total distortion, as long as the overload distortion does not occur, or it is predominantly negligible to the granular distortion. While in the most cases where quantization is traditionally used, for instance, in [16, 19–21, 24], the high-resolution theory ($N \longrightarrow \infty$) is well justified, it totally breaks down for the classes of small or medium numbers of the quantization cells N. Large values of N lead to more fine-leveled quantization, but they result in a less efficient computation and storage consumption. Many quantizers often operate at a very modest bit rate, i.e., at medium bit rates and, for that reason, they indeed make sense of imposing various constraints on the number of used bits. Accordingly, we believe that our analysis derived not only for higher bit rates ($R \geq 5$ bit/sample) but also for small and medium bit rates could be of great significance.

In brief, the model of SQQ that we have proposed is fully specified with the necessary tunable parameter set given by equation (26) and it has very beneficial properties: in pursuit of the reduced total distortion, the number of quantization levels can be gradually incremented; i.e., some compression gain can be provided. Having recast the total distortion asymptotic formula given by equation (20) and by specifying apparently simple solution for the SQQ support region (see equation (13)), in this paper we intended to deal with an opportunity in considering a wide usage of SQQ in NN

training process. However, this has been left for the future research. Eventually, we can anticipate that our results for the successful deployment of clipping in neural networks [15] are very encouraging.

5. Summary and Conclusions

As a main contribution in this paper, we have tackled SQQ parameterization by suggesting one simple method for offline calculation of fully qualified SQQ parameter set, which can be stored for fast retrieval. Our anticipation has been based on the fact that such precalculated values can be leveraged in the deterministic quantization process. We have highlighted that this notice heavily relies on the fact that the values of interest are distributed according to the Laplacian pdf. Namely, we have shown that the practical constraint that rests on quantization, as a guiding principle for data compression, imposes an inevitable information loss, which with the proposed SQQ model we have made smaller compared to the asymptotically optimal SCQ. We have not only analyzed the particular use cases where formidable convenience of SQQ could be applied, but we have also addressed explicit SQQ parameterization, emphasizing the wide area of quantizers usage, as well as the possibility for its qualification for contemporary quantization solutions. Although we intended to consider the application of SQQ in NN training process, this has been left for our future research since in this paper we opted to describe one general quantizer solution applicable for processing any data modelled by the Laplacian pdf. However, from the analysis presented in this paper, we can anticipate that our results for the successful deployment of clipping in NNs are very encouraging and, for that reason, our future work will be directed toward the research on quantization benefits in compression of NN parameters.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was supported by the Science Fund of the Republic of Serbia, 6527104, AI-Com-in-AI.

References

[1] J. P. Garcia-Laencina, "Improving predictions using linear combination of multiple extreme learning machines," *Information Technology and Control*, vol. 42, no. 1, pp. 86–93, 2013.

[2] W. Tian, F. Zhao, Z. Sun et al., "A novel performance prediction model for the machining process based on the interval type-2 fuzzy neural network," *Mathematical Problems in Engineering*, vol. 2020, Article ID 5740362, 10 pages, 2020.

[3] N. Jiang and T. Liu, "An improved speech segmentation and clustering algorithm based on SOM and *k*-means," *Mathematical Problems in Engineering*, vol. 2020, Article ID 3608286, 19 pages, 2020.

[4] N. Strivastava, G. Hinton, A. Krizhevsky et al., "Dropout: a simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[5] Y. Guo, "A survey on methods and theories of quantized neural networks," 2018, https://arxiv.org/abs/1808.04752.

[6] X. Long, X. R. Zeng, Z. Ben et al., "A novel low-bit quantization strategy for compressing deep neural networks," *Comutational Intelligence and Neuroscience*, vol. 2020, Article ID 7839064, 2020.

[7] I. Hubara, M. Courbariaux, D. Soudry et al., "Quantized neural networks: training neural networks with low precision weights and activations," *Journal of Machine Learning Research*, vol. 18, no. 187, pp. 1–30, 2018.

[8] N. Naz, A. H. Malik, A. B. Khurshid et al., "Efficient processing of image processing applications on CPU/GPU," *Mathematical Problems in Engineering*, vol. 2020, Article ID 4839876, 14 pages, 2020.

[9] X. Cui, X. Li, and B. Wang, "Communication optimization technology based on network dynamic performance model," *Mathematical Problems in Engineering*, vol. 2020, Article ID 8890721, 13 pages, 2020.

[10] S. Gazor and W. Wei Zhang, "Speech probability distribution," *IEEE Signal Processing Letters*, vol. 10, no. 7, pp. 204–207, 2003.

[11] S. Jayant and P. Noll, *Digital Coding of Waveforms*, Prentice Hall, Upper Saddle River, NJ, USA, 1984.

[12] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Pearson, London, UK, 4th edition, 2018.

[13] K. Sawada and S. Shin, "Numerical optimization design of dynamic quantizer via matrix uncertainty approach," *Mathematical Problems in Engineering*, vol. 2013, Article ID 250683, 12 pages, 2013.

[14] P. Kabal, "Quantizers for the Gamma distribution and other symmetrical distributions," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, no. 4, pp. 836–841, 1984.

[15] R. Banner, Y. Nahshan, and D. Soudry, "Postraining 4-bit quantization of convolutional networks for rapid-deployment," in *Proceedings of the 33rd Conference On Neural Information Processing Systems (NeurIPS 2019)*, pp. 7948–7956, Vancouver, Canada, December 2019.

[16] D. Aleksić and Z. Perić, "Analysis and design of robust quasilogarithmic quantizer for the purpose of traffic optimisation," *Information Technology and Control*, vol. 47, no. 4, pp. 615–622, 2018.

[17] S. Na and D. L. Neuhoff, "On the convexity of the MSE distortion of symmetric uniform scalar quantization," *IEEE Transactions on Information Theory*, vol. 64, no. 4, pp. 2626–2638, 2018.

[18] J. Lee and S. Na, "A rigorous revisit to the partial distortion theorem in the case of a Laplacian source," *IEEE Communications Letters*, vol. 21, no. 12, pp. 2554–2557, 2017.

[19] S. Tomić, Z. Perić, M. Tančić, and J. Nikolić, "Backward adaptive and quasi-logarithmic quantizer for sub-band coding of audio," *Information Technology and Control*, vol. 47, no. 1, pp. 131–139, 2018.

[20] Z. Perić, M. Petković, and J. Nikolić, "Optimization of multiple region quantizer for Laplacian source," *Digital Signal Processing*, vol. 27, no. 15, pp. 150–158, 2014.

[21] Z. Perić and J. Nikolić, "High-quality Laplacian source quantisation using a combination of restricted and

unrestricted logarithmic quantisers," *IET Signal Processing*, vol. 6, no. 7, pp. 633–640, 2012.

[22] S. Na and D. Neuhoff, "On the Support of MSE-optimal, fixed-rate, scalar quantizers," *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 2972–2982, 2001.

[23] W. R. Bennett, "Spectra of quantized signals," *Bell System Technical Journal*, vol. 27, no. 3, pp. 446–472, 1948.

[24] Z. Perić and J. Nikolić, "An effective method for initialization of Lloyd-Max's algorithm of optimal scalar quantization for Laplacian source," *Informatica*, vol. 18, no. 2, pp. 279–288, 2007.

[25] S. Na and D. L. Neuhoff, "Monotonicity of step sizes of MSE-optimal symmetric uniform scalar quantizers," *IEEE Transactions on Information Theory*, vol. 65, no. 3, pp. 1782–1792, 2019.

[26] P. F. Panter and W. Dite, "Quantization distortion in pulse-count modulation with nonuniform spacing of levels," *Proceedings of the IRE*, vol. 39, no. 1, pp. 44–48, 1951.

[27] S. Kotz, T. Kozubowski, and K. Podgorski, *The Laplace Distribution and Generalizations*, Birkhäuser, Boston, MA, USA, 2001.

[28] J. Huang and Y. Ma, "Bat algorithm based on an integration strategy and Gaussian distribution," *Mathematical Problems in Engineering*, vol. 2020, Article ID 9495281, 22 pages, 2020.

[29] J. Nikolić, Z. Perić, and A. Marković, "Proposal of simple and accurate two-parametric approximation for the *Q*-function," *Mathematical Problems in Engineering*, vol. 2017, Article ID 8140487, 10 pages, 2017.

[30] Z. Perić, N. Simić, and J. Nikolić, "Design of single and dual-mode companding scalar quantizers based on piecewise linear approximation of the Gaussian PDF," *Journal of the Franklin Institute*, vol. 357, no. 9, pp. 5663–5679, 2020.

[31] S. M. Naik, R. P. K. Jagannath, and V. Kuppili, "Bat algorithm-based weighted Laplacian probabilistic neural network," *Neural Computing and Applications*, vol. 32, no. 4, pp. 1157–1171, 2020.

[32] N. Nicodemo, G. Naithani, K. Drossos, T. Virtanen, and R. Saletti, "Memory requirement reduction of deep neural networks for field programmable gate arrays using low-bit quantization of parameters," in *Proceedings of the 28th European Signal Processing Conference EUSIPCO*, pp. 466–470, Amsterdam, Netherlands, January 2021.

[33] J. Nikolić, Z. Perić, and D. Pokrajac, "Average complexity analysis of scalar quantizer design," in *Proceedings of the 6th WSEAS International Conference On Telecommunications and Informatics*, pp. 22–27, Dallas, Texas, USA, March 2007.

[34] D. Hui and D. L. Neuhoff, "Asymptotic analysis of optimal fixed-rate uniform scalar quantization," *IEEE Transactions on Information Theory*, vol. 47, no. 3, pp. 957–977, 2001.