Algorithm based on 2-bit adaptive delta modulation and fractional linear prediction for Gaussian source coding

Zoran Peric | Bojan Denic | Vladimir Despotovic

1Department of Telecommunications, University of Nis, Faculty of Electronic Engineering, Nis, Serbia
2Department of Computer Science, University of Luxembourg, Esch-sur-Alzette, Luxembourg

Abstract
A novel 2-bit adaptive delta modulation (ADM) algorithm is presented based on uniform scalar quantization and fractional linear prediction (FLP) for encoding the signals modelled by a Gaussian probability density function. The study focuses on two major areas: realization of a 2-bit adaptive quantizer based on Q-function approximation that significantly facilitates quantizer design; and implementation of a recently introduced FLP approach with the memory of two samples, which replaces the first-order linear prediction used in standard ADM algorithms and enables improved performance without increasing transmission costs. It furthermore represents the first implementation of FLP in signal encoding, therefore confirming its applicability in a real signal-processing scenario. Based on the performance analysis conducted on a real speech signal, the proposed ADM algorithm with FLP is demonstrated to outperform other 2-bit ADM baselines by a large margin for the gain in signal-to-noise ratio achieved over a wide dynamic range of input signals. The results of this research indicate that ADM with adaptive quantization based on Q-function approximation and adaptive FLP represents a promising solution for encoding/compression of correlated time-varying signals following the Gaussian distribution.

1 | INTRODUCTION AND MOTIVATION

Delta modulation (DM) is a low-complexity technique widely used in quantization and compression of correlated signals such as speech or images [1, 2]. The main idea behind this technique is encoding of the prediction error signal, that is, the difference between the original and predicted signal values using one-bit code words [1–4]. Owing to its attractiveness expressed by a relatively simple encoder/decoder structure and high level of compression, DM has also been applied successfully in many other research fields, such as compression of the hologram [5, 6], electroencephalogram waveforms [7], photoplethysmogram signals [8], and data conversion for portable applications [9].

Two commonly exploited versions of the standard DM—also known as linear delta modulation (LDM)—are adaptive delta modulation (ADM) [1, 2, 10–17] and sigma delta modulation [1, 2, 18, 19]. ADM has been introduced to resolve the main drawbacks of LDM, i.e. the slope overload and the granular noise, caused by the inadequate choice of quantization step size. A slope overload is caused when a too-small step size is available for large variations in the input signal, whereas granular noise is present when a too-large step size is used for small variations in the input signal. Therefore, ADM attempts to fit the quantization step size according to variations in the input signal.

In general, ADM algorithms can be classified into instantaneously based [10–14] and frame-based [15–17] algorithms. In instantaneous ADM [10–14], the adaptation is performed at the sample level, and accordingly, the step size is updated at each sampling instant using the predefined rule. The most prominent members of ADM algorithms are constant factor delta modulation (CFDM) and continuously variable slope delta modulation (CVSDM). Typically, during adaptation process, CFDM uses the memory of one or two samples, whereas in CVSDM, the memory of three or four samples is used. As indicated in [1, 2, 13, 16, 17], these two algorithms can provide a certain level of robustness over a wide range of input signal variances.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. IET Signal Processing published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology.
In frame-based ADM algorithms [15–17], the input signal is processed frame-by-frame (frame denotes a group of consecutive samples of a certain length), where adaptation is performed at the frame level and not the sample level as in the former ADM case. Various statistical properties of the signal can be used to adapt the parameters of the employed quantizer; however, the estimated variance of the signal frame seems to be the most reasonable choice. This is opposed to the case of instantaneous ADM, which is mainly intended for signals with unknown variance. The advantage of the frame-based ADM algorithms over the instantaneous ones is reflected in higher signal-to-noise ratio (SNR) values and a significantly wider dynamic range [16, 17].

Various modifications of basic ADM with 1-bit quantization have been proposed in the literature that utilize 2-bit quantizers [13, 14, 17] while keeping the property of DM that the quantized output signal is determined by adding or subtracting the step size from the predicted signal [14]. The algorithms discussed in [13, 14] are characterized by an exponentially variable rate in step size changes, where the produced output code words contain information about both the sign and the absolute value of the prediction error. The adaptation logic implemented in [13] is relatively simple and consists of scaling the previous step size by an appropriately chosen factor to obtain the step size for the current sample. On the other hand, more complex adaptation logic has been observed in the algorithm presented in [14], where the scaling factor depends on the sign of the current and previous samples of prediction error and the magnitude of the prediction error. Two algorithms with frame-wise adaptation have been proposed in [17], named 2-bit hybrid and 2-bit optimal ADM, with the goal to improve the performance of the algorithm in [14]. In a 2-bit hybrid ADM, a few additional steps have been introduced compared with the algorithm in [14], where the adaptation has been performed at both the frame and the sample level. In particular, the frame variance has been estimated and used for determination of the initial step size for each frame, while to process each sample of the frame, the same adaptive procedure for step size as in [14] has been exploited. In this manner, embedded step size initialization has been ensured, as opposed to [14], which requires an external LDM for the initial step size choice that may not be optimal. The 2-bit optimal ADM has been introduced to improve the performance of the ADM in [14] in terms of the employed quantizer.

We want to enhance the performance of the existing 2-bit ADM algorithms in a manner as simple as possible, that is, without introducing additional complexity. This can be achieved by improving the performance of the main building components—quantizer and/or predictor—and we address both of these components. Non-uniform scalar quantization and first-order linear prediction (LP) have been commonly used thus far in ADM. Our study is focused on two major modifications of basic ADM: realization of a 2-bit adaptive quantizer based on Q-function approximation that significantly facilitates quantizer design, and implementation of a recently introduced fractional linear prediction (FLP) approach with the memory of two samples, which replaces the first-order LP used in standard ADM algorithms and enables improved performance without increasing transmission costs. The resulting 2-bit ADM is intended to encode Gaussian sources that can be used for modelling of various real signals, such as speech [20] or Orthogonal Frequency-Division Multiplexing-modulated signals [21].

Utilization of the 2-bit uniform scalar quantization instead of the non-uniform (Lloyd–Max) counterpart can be explained by the fact that both provide an approximately equal signal-to-quantization noise ratio (SQNR) value for a Gaussian probability density function (PDF) [1, 2], whereas the former is simpler for design. We propose an iterative design method based on the Q-function by which performance comparable to Lloyd–Max quantization can be achieved but with notably reduced computational complexity (given as the number of arithmetical operations). The main reason for employing the Q-function approximations is that the Gaussian Q-function cannot be represented via only elementary functions and therefore requires numerical calculations. Some useful Q-function approximations used in the context of facilitating the performance analysis of the communications systems can be found in [22–27], while those used in scalar quantization can be found in [28, 29]. A recently proposed parametric Q-function approximation from [28] is used with its parameter optimized for the problem under consideration. Additionally, it has already been indicated in [30, 31] that 2-bit uniform quantization can be efficiently used for the compression of weights used in neural networks (from the 32-bit floating point to the low-bit fixed point), thus resulting in a significant reduction in storage requirements and computational complexity. Because weight coefficients can be well described by Gaussian PDF [30], one can expect suitability of the proposed quantizer in neural network applications.

FLP [32–34] is the approach based on the application of the fractional calculus technique [35] in predictor design. Additionally, the technique of fractional calculus has been employed in various digital signal processing applications, such as the detection of ECG signals [36] and the design of 1-D [37] and 2-D [38] digital filters. The FLP model we use was recently proposed in [33, 34] and employs the memory of two samples, as in the second-order LP, but only requires one coefficient, as in the first-order LP. Hence, the employed FLP falls between the first-order and the second-order LP. It has already been shown in [32, 33] that FLP can outperform the first-order LP in terms of prediction gain and offers comparable gain values to those of the second-order LP, but there have been no implementations as of yet in real signal-processing scenarios. Therefore, we would like to investigate the applicability of FLP in ADM.

The rest of this paper is organized as follows. In Section 2, a design procedure of the quantizer (with and without Q-function approximation) is provided in detail along with its application in a forward adaptive algorithm. In Section 3, the FLP model is described. In Section 4, the proposed 2-bit ADM algorithm is introduced. Section 5 summarizes and discusses the experimental results, and Section 6 concludes the paper.
2.2 | Design of optimal 2-bit uniform scalar quantizer

The design of an optimal scalar quantizer assumes specification of the parameters of the quantizer (decision thresholds and representative levels) for the so-called reference variance \( \sigma_{ref}^2 \), that is, using \( p(x, \sigma_{ref}) \) such that minimal MSE distortion is inserted. Usually, the unit variance is used as the reference \( \langle \sigma_{ref}^2 \rangle = 1 \) [1]. In that case, MSE distortion in Equation (2) takes the form

\[
D = 1 - \sqrt{\frac{2}{\pi}} x_1 \left( 1 + 2 \exp \left( -\frac{x_1^2}{2} \right) \right) + x_1^2 \left( \frac{1}{4} + 4Q(x_1) \right).
\]  

(4)

Equation (4) suggests that parameter \( x_1 \) is of extreme importance in the design of an MSE-optimal quantizer.

Lemma 1 The step size \( x_1 \) of the symmetric 2-bit uniform scalar quantizer of a Gaussian source optimized in terms of MSE distortion can be obtained by solving the following iterative equation:

\[
x_1^{(i+1)} = \frac{1}{\sqrt{2\pi}} \left( 1 + 2 \exp \left( -\frac{(x_1^{(i)})^2}{2} \right) \right) \left( \frac{1}{4} + 4Q(x_1^{(i)}) \right).
\]

(6)

Proof By finding the first derivative of MSE distortion with respect to \( x_1 \) and equating it to zero, \( \partial D / \partial x_1 = 0 \), one can derive the following integral equation:

\[
\frac{1}{\sqrt{2\pi}} \left( 1 + 2 \exp \left( -\frac{x_1^2}{2} \right) \right) = x_1 \left( \frac{1}{4} + 4Q(x_1) \right).
\]

(7)

From the last equation, \( x_1 \) can be expressed as

\[
x_1 = \frac{1}{\sqrt{2\pi}} \left( 1 + 2 \exp \left( -\frac{x_1^2}{2} \right) \right) \left( \frac{1}{4} + 4Q(x_1) \right),
\]

(8)

showing that the required step size value can be obtained iteratively, which concludes the proof.

As can be seen from Equation (6), to obtain the exact value of the step size (and accordingly all other quantizer parameters), numerical calculation of the \( Q \)-function must be
performed because its closed-form expression is not available. Typically, in numerical calculations, the Q-function is given as the sum of exponential terms. If \( s \) denotes the number of terms used in summation and \( I \) denotes the number of iterations, the computational complexity \( C \) of this design approach is given by \( C = I \cdot s + 4 \).

### 2.3 Design of optimal 2-bit uniform scalar quantizer based on Q-function approximation

The proposed design approach is based on the replacement of the Gaussian Q-function with the corresponding Q-function approximation with the goal of reducing the computational complexity of the quantizer presented in the previous section while preserving the performance as much as possible. Although there are many Q-function approximations available in the literature for the problem, for example \([22–29]\), we will use the Q-function approximation recently proposed in \([28]\).

The proof of its suitability will be given later in this section.

The parametric Q-function approximation \([28]\) is

\[
F(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{x^2 + a}} \exp\left( -\frac{x^2}{2} \right), \quad 1 \leq a < 2, \quad (9)
\]

and defines the class of approximations that are the upper bounds for \( x > \sqrt{(4a + 1 + 2a^2 - 2a - 1)/(4 - 2a)} \).

Obviously, the aim is to find the best value of parameter \( a \) related to the problem we are interested in. We provide a method for optimization of Equation (9) with respect to \( a \) within the interval (range) of arguments \( x \) that is highly important for our analysis. The bounds of the range are determined from the upper \([23, \text{Equation (18)}]\) and lower \([25, \text{Equation (2)}]\) bounds of the Q-function, respectively given by Equations (10) and (11):

\[
F^{[23]}(x) = \frac{1}{\sqrt{2\pi}} \frac{x}{x^2 + 1} \exp\left( -\frac{x^2}{2} \right), \quad (10)
\]

\[
F^{[25]}(x) = \frac{1}{\sqrt{2\pi}} \frac{x^2 + b - 1}{x(x^2 + b)} \exp\left( -\frac{x^2}{2} \right), \quad (11)
\]

The lower bound of the range \( x = x^{\text{low}}_1 \) is obtained by replacing \( Q(x) \) in Equation (6) with its approximation in Equation (10). Similarly, the upper bound of the range \( x = x^{\text{up}}_1 \) is obtained by replacing \( Q(x) \) with its approximation in Equation (11). The initial decision threshold value in both cases is set to \( x_1^{[0]} = 2\sqrt{\log N}, N = 4 \) \([39]\).

In the optimization method we propose, for each argument value \( x = x_i(i) \), \( i = 1, \ldots, N_x \), from the interval \([x^{\text{low}}_1, x^{\text{up}}_1]\), the parameter value denoted as \( a^{\text{opt}}(x_i(i)) \) that minimizes the relative error in the approximating Q-function is determined as

\[
a^{\text{opt}}(x_i(i)) = \arg\min_a \left\{ \frac{F(x_i(i), a) - Q(x_i(i))}{Q(x_i(i))} \right\}, \quad (12)
\]

where \( i = 1, \ldots, N_x \), \( x^{\text{low}}_1 = x_1(1) \) and \( x^{\text{up}}_1 = x_1(N_x) \). Afterwards, the linear approximation \( a(x) = k \cdot x + w \) on the established interval is provided such that the minimal MSE is introduced.

Figure 2 presents the results obtained by applying the described optimization method, and as a result, the following Q-function approximation is obtained:

\[
F^p(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{x^2 + 0.48x + 0.84}} \exp\left( -\frac{x^2}{2} \right). \quad (13)
\]

Substituting Equation (13) into Equation (6), one can estimate the approximate value of the step size denoted as \( x^{\text{opt}}_1 \):

\[
(x^{\text{opt}}_i)^{(i+1)} = \frac{1}{\sqrt{2\pi}} \left( 1 + 2\exp\left( -\frac{(x^{[i]}_i)^{[0]}}{2} \right) \right). \quad (14)
\]

It is clear that Equation (14) requires a certain number of iterations (denoted as \( l_1 \)) to specify the step size but avoids numerical calculation of the Q-function. The computational complexity is equal to \( C = l_1 + 4 \); therefore, the computational complexity is reduced \( l_1 \cdot (s+4)/l_1 + 4 \) times with respect to the former iterative design method.

Furthermore, the following approximate closed-form expressions for distortion and SQNR can be derived:
$D^* = 1 - \sqrt[4]{2} x^*\left(1 + 2\exp\left(-\left(\frac{x^*}{2}\right)^2\right)\right) + \left(x^*\right)^2\left(\frac{1}{4} + 4F^*\left(x^*\right)\right), \quad (15)$

$SQNR' = 10\log_{10}\left(\frac{1}{1 - \sqrt[4]{2} x^*\left(1 + 2\exp\left(-\left(\frac{x^*}{2}\right)^2\right)\right) + \left(x^*\right)^2\left(\frac{1}{4} + 4F^*\left(x^*\right)\right)}\right), \quad (16)$

### 2.4 Accuracy of derived formulas

To evaluate the accuracy of the derived approximate Equations (14)–(16), they are compared with the corresponding ones obtained without the Q-function approximation (Section 2.2). In addition, in the comparison analysis, we include approximations with analytical forms similar to the one derived in our study ([22, 24, 27]):

$F^{[22]}(x) = \frac{\left(1 - \exp\left(-\frac{ax}{\sqrt{2}b}\right)\right)\exp\left(-\frac{x^2}{2}\right)}{\sqrt{2\pi bx}}, \quad (17)$

$F^{[24]}(x) = \frac{\left(1 - \exp(-Cx)\right)\exp\left(-\frac{x^2}{2}\right)}{\sqrt{2\pi x}}, \quad C = \sqrt{\frac{\pi}{2}}, \quad (18)$

$F^{[27]}(x) = \frac{1}{\sqrt{2\pi b}} \frac{\pi}{\sqrt{2\pi + bx}} \exp\left(-\frac{x^2}{2}\right). \quad (19)$

Note that values $A = 1.98$ and $B = 1.135$ in Equation (17) are computed to minimize the integral of the absolute error for a wide range of arguments $x \in [0, 20]$ [22].

Table 1 summarizes the calculated values of the step size, distortion, and resulting SQNR for various Q-function approximations as well as the relative errors for the corresponding parameters, defined as $|\delta x| = |x^* - x_j| / x_j^*$, $\delta D = |D^* - D_j| / D^*$, and $\delta_{SQNR} = |SQNR^* - SQNR_j| / SQNR^*$, $\delta_{SQNR} = |SQNR^* - SQNR_j| / SQNR^*$.

---

### Table 1 Approximated decision threshold, distortion, and bit rate for various Q-function approximations

| Approximation | $x_1$ | $\delta x$ (%) | $D$ | $\delta D$ (%) | $SQNR$ (dB) | $\delta_{SQNR}$ (%) |
|---------------|-------|----------------|-----|----------------|-------------|---------------------|
| $F^{[22]}(x)$ | 0.9759 | 1.9886         | 0.1270 | 6.9024         | 8.9620      | 3.1135              |
| $F^{[24]}(x)$ | 0.8742 | 12.2025        | 0.1753 | 47.5589        | 7.5622      | 18.2465             |
| $F^{[27]}(x)$ | 1.2704 | 27.5886        | 0.0409 | 65.5724        | 13.8828     | 50.0843             |
| $F(x)$        | 0.9936 | 0.2109         | 0.1197 | 0.7576         | 9.2191      | 0.3341              |

The best achieved performance is emphasised in bold. Abbreviation: SQNR, signal-to-quantization noise ratio.
\[ \sigma_j^2 = 1/M \sum_{n=1}^{M} x_j^2(n), \quad j = 1, \ldots, F. \]  

(21)

Note that in Equation (21), we are assuming that input frame has a zero-mean value. The quantization of this parameter is performed using the log-uniform quantizer (Q_{LU}) because it attains better performance than other available quantization models—for example, uniform quantization \([40]\). Hence, using Q_{LU}, the logarithmic variance \( V_j [\text{dB}] = 20 \cdot \log_{10}(\sigma_j/\sigma_{\text{ref}}) \) is quantized to one of the \( L \) possible values, defined as \([15–17]\):

\[ V_j \in V_k \implies V_k = V_{\text{min}} + \frac{2k - 1}{L} B \quad k = 1, \ldots, L, \]  

(22)

where \( B[\text{dB}] = V_{\text{max}}[\text{dB}] - V_{\text{min}}[\text{dB}] \) defines the dynamic range of the signal, and \( V_{\text{max}} \) and \( V_{\text{min}} \) denote the maximal and minimal estimated value of the frame variance, respectively.

The employed level of Q_{LU} is encoded using the fixed-length code words composed of \( R_{LU} = \log_2 L \) bits (index \( J \)) and transmitted as the side information.

Step 3 Adaptive 2-bit uniform quantization.

The output level of Q_{LU}, that is, the quantized variance, is further used to scale the parameters of the fixed 2-bit uniform quantizer (designed for \( \sigma^2 = \sigma_{\text{ref}}^2 = 1 \)), and as a result, the adaptive quantizer is obtained. The key parameter is specified as follows:

\[ x^a_1(g) = g x^a_1(\sigma_{\text{ref}}), \]  

(23)

and accordingly, the adaptive decision thresholds and the representative levels are determined as

\[ t_3(g) = 0, t_4(g) = x^a_1(g), t_5(g) = 2x^a_1(g), \]  

(24)

\[ y_3(g) = x^a_1(g)/2, y_4(g) = 3x^a_1(g)/2, \]  

(25)

where \( g \) is the scaling factor defined as

\[ g = 10^{\frac{V_k}{20}}. \]  

(26)

After the quantizer codebook, designed for variance \( g^2 \), is well matched for the \( j \)-th frame of variance \( \sigma_j^2 \), each of \( M \) samples contained in the \( j \)-th frame is passed through the quantizer and encoded using 2-bit code words (index \( I \)).

Step 4 Repeat the previous steps until all frames are processed.

To estimate the performance of the described forward adaptive algorithm the following equations for distortion, SQNR and bit rate are valid:

\[
D(\sigma, g) = \sigma^2 - \sqrt{\frac{2}{\pi}} x^a_1(g) \left( 1 + \exp\left( -\frac{(x^a_1(g))^2}{2} \right) \right) + \\
+ \left( x^a_1(g) \right)^2 \left( \frac{1}{4} + 4F(\sigma^a_1(g)) \right),
\]  

(27)
\[ SQNR(\sigma, g) = 10\log_{10}\left(\frac{\sigma^2}{D(\sigma, g)}\right), \tag{28} \]

\[ R = 2 + \frac{\log_2 L}{M}. \tag{29} \]

where the second term in Equation (29) refers to the side information.

### 2.6 Performance of adaptive 2-bit uniform scalar quantizer

Figure 4 plots the theoretical SQNR for the non-adaptive and adaptive 2-bit uniform quantizer (Section 2.2) over the broad variance range (-20 dB, 20 dB) in terms of the reference (designed) variance \( \sigma_{\text{ref}}^2 = 1 \), when \( L = 32 \)-level \( Q_{LU} \) is used. The advantage of adaptation is reflected in an increased dynamic range in comparison with the non-adaptive quantizer, whereas the maximal SQNR is same in both cases. The inferior performance of the non-adaptive quantizer in processing the Gaussian signals having different variance than the designed one \( \sigma^2 \neq \sigma_{\text{ref}}^2 \) are caused by the limit cycle effects, which leads to unfavourable SQNR degradation. The implementation of quantizer in forward adaptive algorithm can significantly suppress this effect, as expressed by almost constant and periodical SQNR in Figure 4, where the number of periods corresponds to the number of levels of the \( Q_{LU} \).

In Figure 5, in the same variance range as above, we depict the theoretical SQNR for the forward adaptive 2-bit uniform quantizer, designed without (Section 2.2) and with the Q-function approximation (Section 2.3). Observe the closeness of the SQNR curves in the whole range of variances. The graph of relative error illustrated in Figure 6, shows the periodicity and small range of errors. The maximum relative error amounts to 0.49\%, whereas the average relative error (see the dashed line) equals 0.31\%, indicating high accuracy.

For comparison purposes, in Figure 5 we also show the SQNR of the forward adaptive 2-bit Lloyd-Max's quantizer designed for Gaussian source [1,2], which serves as the upper bound. Observe that the proposed uniform quantizer with Q-function approximation is almost as effective as the Lloyd-Max's quantizer, providing only slightly lower SQNR values, but with much simpler design. In particular, for the problem of 2-bit quantization, the computational complexity of the Lloyd-Max's algorithm is equal to \( C = I_2 \cdot (6m + 5) \), where \( I_2 \) is the number of iterations. Comparing it with the one obtained in Section 2.3 leads to the conclusion that the computational complexity of the proposed design method is reduced with respect to the Lloyd-Max's algorithm by \( I_2(6s + 5)/I_1 + 4 \) times, where \( I_1 < I_2 \), showing the benefit of our approach.

**FIGURE 4** SQNR of non-adaptive \( (\sigma_{\text{ref}}^2 = 1) \) and adaptive 2-bit uniform quantizer \( (L = 32\text{-level } Q_{LU}) \) designed using Q-function approximation in wide dynamic range of input signal variances. SQNR, signal-to-quantization noise ratio.

**FIGURE 5** Performance (SQNR) comparison of different 2-bit scalar quantizers applied in forward adaptive algorithm, \( L = 32\text{-level } Q_{LU} \). SQNR, signal-to-quantization noise ratio.

**FIGURE 6** Relative error in SQNR attained in considered variance range of width 40 dB. SQNR, signal-to-quantization noise ratio.
3 | MODEL OF FRACTIONAL LINEAR PREDICTOR

FLP [32–34] is a novel approach in signal prediction, based on application of the fractional calculus technique [35]. Meanwhile, the classical LP model estimates the current sample as the linear combination of previous samples:

\[ x_p(n) = \sum_{i=1}^{p} a_i x(n-i), \quad (30) \]

where \( x_p(n) \) is the estimated sample, \( p \) is the prediction order and \( a_i \) are the LP coefficients, the predicted sample in FLP is obtained as a linear combination of \( q \) ‘fractional terms’:

\[ x_p(n) = \sum_{i=1}^{q} b_i D^\lambda x(n-1), \quad (31) \]

where \( \lambda_i \) is the real number, \( b_i \) stands for FLP coefficients and \( D^\lambda x(n-1) \) represents the fractional derivative operator of order \( \lambda_i \) of time delayed signal. Therefore, FLP can be observed as the generalized form of LP.

FLP model that we will use is based on the one recently elaborated in [33, 34] using only one fractional term to predict the current sample:

\[ x_p(n) = b D^\lambda x(n-1), \quad (32) \]

where the fractional derivative operator can be interpreted using the Grünwald–Letnikov definition [35]:

\[ D^\lambda x(n-1) = 1/b^\lambda \sum_{j=0}^{m} (-1)^j \binom{\lambda}{j} x(n-j), \quad (33) \]

where \( b \) is the sampling period and \( m \) is the small positive number representing the upper bound of the summation. Using the recurrent relationship for estimation of binomial coefficients, \( \omega_{j}^{(\lambda)} = (-1)^j \binom{\lambda}{j}, \) as

\[ \omega_{0}^{(\lambda)} = 1, \quad \omega_{j}^{(\lambda)} = (-1)^j \binom{\lambda+1}{j} \omega_{j-1}^{(\lambda)}, \quad j = 1, 2, \ldots, \quad (34) \]

and further adopting \( m = 1 \), one can arrive to [33, 34]:

\[ D^\lambda x(n-1) = 1/b^\lambda (x(n-1) - \lambda x(n-2)). \quad (35) \]

Substituting Equation (35) in Equation (32) results in determination of the target FLP model:

\[ x_p(n) = b/b^\lambda (x(n-1) - \lambda x(n-2)). \quad (36) \]

The defined FLP model uses the memory of two samples, same as in the second-order LP (\( p = 2 \) in Equation (30)). On the other hand, if \( \lambda \) is fixed then the FLP model in Equation (36) is specified by only one coefficient, same as in the first-order LP (\( p = 1 \) in Equation (30)). This implies that the considered FLP model is between the first- and second-order LPs. It is also interesting to note that for \( \lambda = 0 \) the FLP model is equivalent to the first-order LP. Therefore, \( \lambda \) should be selected so that performance as high as possible is attained for the case \( \lambda = 0 \). The authors in [32] have proposed an approximate expression for \( \lambda \), estimated as the inverse of predictor memory (\( \lambda \approx 1/2 \), for the memory of two samples).

The validity of the approximated formula will also be checked in case of ADM.

In view of implementation in adaptive systems performing frame-by-frame signal processing (e.g. ADM), both first-order LP and FLP require the same transmission costs since only one coefficient should be updated per frame, whereas transmission costs are reduced with respect to the second-order LP where two coefficients have to be updated per frame. This effect is especially visible when the signal frames are short, that is, the case where side information has to be sent more often.

The optimal FLP coefficient \( b \) in Equation (36) can be found by minimizing the MSE of the prediction error signal. The prediction error signal \( e(n) \) can be determined according to

\[ e(n) = x(n) - x_p(n) = x(n) - b/b^\lambda (x(n-1) - \lambda x(n-2)), \quad (37) \]

that is as the difference between the current sample value \( x(n) \) and its predicted value \( x_p(n) \), while the mean-squared prediction error, \( \sigma_e^2 \), is specified as

\[ \sigma_e^2 = E[(x(n) - x_p(n))^2] = \sigma_e^2 \left( 1 - \frac{2b}{b^\lambda} (\rho_1 - \lambda \rho_2) + \left( \frac{b^\lambda}{b^\lambda} \right)^2 \left( 1 - 2\lambda \rho_1 + \lambda^2 \right) \right), \quad (38) \]

where \( E \) is the mathematical expectation, and \( \rho_1 \) and \( \rho_2 \) are the correlation coefficients at lags 1 and 2, respectively. By differentiating \( \sigma_e^2 \) over \( b \) and equating it to zero, that is, \( \partial \sigma_e^2 / \partial b = 0 \), one can obtain the optimal FLP coefficient [32]:

\[ b^{opt} = b^\lambda \frac{\rho_1 - \lambda \rho_2}{1 - 2\lambda \rho_1 + \lambda^2}. \quad (39) \]

Substituting Equation (39) in Equations (37) and (38) we obtain

\[ e(n) = x(n) - \left( \frac{\rho_1 - \lambda \rho_2}{1 - 2\lambda \rho_1 + \lambda^2} \right)(x(n-1) - \lambda x(n-2)), \quad (40) \]
Using Equation (41), we can define the prediction gain in case of considered FLP model:

\[ G = 10 \log_{10} \left( \frac{\sigma_e^2}{\sigma_x^2} \right) = 10 \log_{10} \left( 1 - \frac{(\rho_1 - \lambda \rho_2)^2}{(1 - 2\rho_1 + \lambda^2)} \right). \]  

\[ (42) \]

4  |  A 2-BIT ADAPTIVE DELTA MODULATION ALGORITHM BASED ON FRACTIONAL LINEAR PREDICTION

The components described above, a 2-bit uniform quantizer (Section 2) and an FLP predictor with the memory of two samples (Section 3), are further implemented in ADM algorithm intended for processing of non-stationary Gaussian sources, as shown in Figure 7. We introduce the following notations: \( x(n) \) is the original signal, \( e(n) \) is the prediction error signal, \( e_q(n) \) is the quantized prediction error signal, \( y(n) \) is the reconstructed signal and \( x_p(n) \) is the predicted signal.

Similar to the algorithm described in Section 2 (Figure 3), the present one performs frame-by-frame analysis of the input signal, but here we quantize (encode) the difference between the original samples \( x(n) \) and their predicted values \( x_p(n) \) (it actually represents the weighted sum of the delayed reconstructed samples \( y(n) \)), that is, the prediction error signal \( e(n) \) and not the original sample values. Hence, the quantizer should be adapted to the variance of the prediction error, and this information needs to be quantized and transmitted. On the other hand, the FLP coefficient (see Equation (39)) is estimated from the original frame (as \( \lambda \) is known at both encoder–decoder sides), which is further quantized and transferred to the receiver, increasing the side information in comparison with Section 2. Additionally, the error-free channel is assumed for transmission of the quantized prediction error signal \( e_q(n) \) and side information.

The 2-bit ADM algorithm displayed in Figure 7 operates in one of two modes, depending on whether zero frame or non-zero frame is available. In the first mode (non-zero frame, i.e. \( \sigma_e^2 \neq 0 \)), the principle of work can be described by the following steps:

Step 1 Frame buffering. Same as in Step 1 of the algorithm in Section 2.5.

Step 2 Estimation of the frame variance. Same as in Step 2 of the algorithm in Section 2.5.

Step 3 Estimation of the correlation coefficients. The correlation coefficients at lags 1 and 2, denoted as \( \rho_1 \) and \( \rho_2 \), for the current \( j \)-th frame can be calculated as [1, 2]:

\[ \rho_{kj} = \frac{\sum_{n=1}^{M-k} x_j(n) x_j(n + k)}{\sqrt{\sum_{n=1}^{M} x_j^2(n)} \sqrt{\sum_{n=1}^{M} x_j^2(n + k)}}, \quad j = 1, \ldots, F, \quad k = 1, 2. \]

\[ (43) \]

Step 4 Estimation of the variance of the prediction error and quantization. For the \( j \)-th frame, the variance of the prediction error can be estimated according to Equation (41) based on calculated frame variance, correlation...
coefficients $\rho_{1j}$ and $\rho_{2j}$, and parameter $\lambda$, which is a priori available in the encoder/decoder part. Variance is further quantized using the log-uniform quantizer ($Q_{LU}$) with $L$-1 representative levels, and index of the employed level using $\log_2 L$ bits (index $j$) is transmitted to the receiver end. Recall that the dynamic range of $Q_{LU}$ is smaller with respect to the former case (Section 2.5), as indicated by Equation (41).

Step 5 Estimation of FLP coefficient and quantization. The FLP coefficient $b$ for the $j$-th frame can be estimated using the correlation coefficients $\rho_{1j}$ and $\rho_{2j}$ and parameter $\lambda$, according to Equation (39). It is uniformly quantized ($Q_U$) to the one of $T$ representative levels:

$$b_t \in b_t | b_t = b_{\min} + \frac{2t - 1}{2} \Delta_t, \text{ } t = 1, \ldots, T \quad (44)$$

where $\Delta_t = (b_{\max} - b_{\min})/T$ is the step size, and $b_{\min}$ and $b_{\max}$ denote the minimal and maximal estimated values of the FLP coefficient, respectively.

Based on the $Q_U$ output, the FLP coefficient is specified and the information about this is conveyed to the receiver as side information using $R_U = \log_2 T$ bits once per frame (index $K$).

Step 6 Determination of prediction error. The prediction error signal for the $j$-th frame is determined as

$$e_j(n) = x_j(n) - x_{pj}(n) \quad (45)$$

where $x_{pj}(n) = b_j^{opt}/b_j^y(y_j(n - 1) - \lambda y_j(n - 2)), n = 1, \ldots, M_j$, is the predicted sample value, and $y_j(n)$ is the reconstructed value:

$$y_j(n) = x_{pj}(n) + e_{qj}(n) \quad (46)$$

Step 7 Adaptive quantization. The adaptive 2-bit uniform quantizer is designed as in Step 3 of the algorithm in Section 2.5, that is, by multiplying the codebook of the fixed 2-bit uniform quantizer with the scaling factor determined by Equation (26). Using it, the prediction error of the $j$-th frame, $e_j(n)$, is quantized and signal $e_{qj}(n)$ is obtained. The encoded samples are transmitted using index $L$.

Step 8 Repeat all previous steps until all frames are processed.

If the zero frame is available (e.g. silence in active speech), the variance of the input signal frame and of the prediction error is zero, and hence, all quantizer levels are set to zero. Furthermore, the prediction error signal is zero, and the quantized prediction error is zero as well. This case is equivalent to the ideal reconstruction scenario, as the input and output of the algorithm are identical. For this characteristic case, encoding should not be performed ($M \times 2$-bit code words are not sent in this case), only the information about the zero variance should be transferred to the receiver to inform it about the zero input. Therefore, one reserved code word of the log-uniform quantizer is used (e.g., 00000), while the remaining 5-bit code words are used for the non-zero frames. More specifically, when the 00000 code word is received, the decoder should generate a frame containing all zeros.

It should be emphasized that because of the adaptation of both the quantizer and the predictor as well as processing only small portions of signal within one frame, the possibility for large error accumulation and propagation is rather small.

The bit rate and SNR are used to specify the performance of the proposed 2-bit ADM algorithm. Particularly, bit rate is defined by

$$R^{FA} = 2 + \frac{\log_2 L + \log_2 T}{M} \quad (47)$$

which is larger than the one defined in Equation (29) for $\log_2 T$ bits per frame (information about FLP coefficient). On the other hand, in predictive coding algorithms such as DM, it holds [1, 2]

$$\text{SNR} = \text{SQNR} + G \quad (48)$$

that is, the overall SNR can be decomposed into the contribution of the adaptive quantizer (SQNR, Equation (28)) and adaptive predictor ($G$, Equation (42)).

5 | EXPERIMENTAL RESULTS AND DISCUSSION

The efficiency of the proposed 2-bit ADM algorithm is analyzed on speech signal because short-term statistics of speech are well described by the Gaussian PDF [1, 20]. In our analysis, we use $L = 32$-level $Q_{LU}$ for quantization of the variance of the prediction error ($R_{LU} = 5$ bits) and $T = 32$-level $Q_U$ ($R_U = 5$ bits) for quantization of the FLP coefficient, as it was recently shown in [17] that such quantizers represent the compromise rate-quality solution. As performance measure the segmental SNR (SNR$_{seg}$) is used [1, 2, 16, 17], that is, the SNR is calculated over each speech frame and then averaged. The frame length is set to 5 ms.

The training sequence of approximately 30 min of speech extracted from the TIMIT dataset (8 kHz sampling rate) [41] was used to estimate the parameter $\lambda$ of the FLP predictor. The estimated parameter $\lambda$ is then applied to speech that was not included in the training sequence.

Figure 8 illustrates the impact of different values of $\lambda$ ranging from 0 to 0.6 on the algorithm performance measured by SNR. Recall that for $\lambda = 0$, the FLP model becomes the first-order LP. It can be perceived that as $\lambda$ increases ($\lambda > 0$), performance gets better, attaining its maximum for $\lambda = 0.48$. The difference in maximal SNR and SNR obtained when $\lambda = 0$ is actually the maximal gain in performance over the case when the first-order LP is implemented in 2-bit ADM algorithm. Therefore, the parameter $\lambda$ is fixed to 0.48. This is also in
accord with the result provided in [32] where $\lambda$ is approximated as the inverse of the predictor memory ($\lambda \approx 1/s$, for the memory of two samples); hence, correctness of the estimation formula is confirmed for DM. Further increasing of $\lambda$ ($\lambda > 0.48$) results in performance drop, meaning that values $\lambda > 0.48$ are out of interest.

Experiments are performed on a speech signal extracted from the TIMIT dataset [41] sampled at a frequency of 8 kHz and having a length of 24,000 samples. The available speech includes both voiced and unvoiced segments.

Table 2 lists the performance (SNR$_{seg}$ and bit rate) of the 2-bit ADM with FLP model ($\lambda = 0.48$) and the cases when 2-bit ADM (with quantizer defined in Section 2.3) is implemented using the optimal LP of the first and the second order. For comparison purposes, we also present the results for 2-bit optimal ADM algorithm (32-level $Q_{LU}$ for frame variance quantization and 32-level $Q_{LU}$ for the first-order LP coefficient) reported in [17] (see 2-bit ADM [17]), as well as the non-predictive algorithm described in Section 2.5.

For the optimal first-order LP, the coefficient is determined as [1,2]

$$d^F = \rho_1,$$  (49)

while the coefficients of the optimal second-order LP are determined as [1,2]

$$d'^F = \frac{\rho_1(1 - \rho_2)}{1 - \rho_1^2}, \quad d'^S = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2},$$  (50)

The coefficients in Equations (49) and (50) are uniformly quantized using 5 bits (i.e. 32-level $Q_{LU}$ is used). It can be seen in Table 2 that the implementation of FLP provides nearly 1.5 dB higher SNR$_{seg}$ in comparison with the optimal first-order LP, where the same bit rate is observed (2.25 bit/sample). Furthermore, observe that implementation with FLP model provides SNR$_{seg}$ value close to the one achieved with the ADM with the second-order LP, but at the lower bit rate. Moreover, since both SQNR and $G$ contribute to the overall SNR in the predictive schemes (see Equation (48)), instead of presenting the results for SNR only, we also assessed what are the individual contributions of different employed predictors. Thus, in Figure 9 we shown the prediction gain achieved across different speech frames and we also provided the average gain values ($G_{av}$). Note the high predicting capability of FLP in comparison with the first-order LP, indicated by higher prediction gain $G$ for almost all speech frames and consequently, higher average prediction gain $G_{av}$. Note also the competitive performance with respect to the second-order LP. Owing to that fact and the lower transmission costs, the considered FLP model is a better choice for real-time applications where bit rate is a critical parameter. Additionally, Table 2 and Figure 9 suggest that the quantizer contribution is approximately 9 dB, which is consistent with the theoretical results presented in Figure 5.

Figure 10a allows us to take a better insight into the SNR performance of the respective algorithms at the frame level. It can be seen that the proposed 2-bit ADM algorithm is able to ensure higher SNR values for both voiced and unvoiced frames in comparison with the baselines. By comparing their average values given in Table 2, an approximately 1.7 dB higher SNR$_{seg}$ is achieved with respect to the algorithm in [17], both having the same bit rate (2.25 bit/sample). The superiority of the proposed predictive coding method vs. the non-predictive one is clearly visible as well, where the gain of more than 4 dB is observed. Additionally, the results of non-predictive algorithm are in accord with the theoretical ones given in Figure 5. In Figure 10b, the waveforms of the reconstructed speech and error signal are plotted, showing nearly perfect reconstruction of the original speech signal.

Finally, we want to indicate the superiority of the proposed 2-bit ADM algorithm over the instantaneously based 2-bit ADM reported in [13,14]. For the baseline algorithm in [13], we use $\alpha = 2$, whereas for the one in [14] we adopt $\alpha = 1.1$, $\beta = 1.9$ and $\gamma = 1.5$. The initial step size is the one that

![Figure 8](image_url)  
**FIGURE 8**  SNR versus order of fractional derivative operator $\lambda$. SNR, signal-to-noise ratio.

| Algorithm | 2-bit ADM (First-order LP) | 2-bit ADM (FLP) | 2-bit ADM (Second-order LP) | 2-bit ADM [17] | Non-predictive algorithm (Section 2.5) |
|-----------|--------------------------|----------------|-----------------------------|----------------|-------------------------------------|
| SNR (dB)  | 12.67                    | 14.13          | 14.80                       | 12.45          | 9.70                                |
| $R$ (b/s) | 2.25                     | 2.25           | 2.375                       | 2.25           | 2.125                               |

Abbreviations: ADM, adaptive delta modulation; FLP, fractional linear prediction; LP, linear prediction; SNR, signal-to-noise ratio.
maximizes the SNR of LDM, while the variable step size is limited to the upper $\Delta_{\text{max}}$ and lower $\Delta_{\text{min}}$ value, satisfying $\Delta_{\text{max}}/\Delta_{\text{min}} = 1000$ (60 dB dynamic range). The results are depicted in Figure 11, where SNR is given as the function of the input speech variance and $\sigma_{\text{ref}}^2$ denotes the variance of the original speech signal. One can notice that the proposed algorithm significantly improves the dynamic range of the baselines and offers higher SNR values (obtained by averaging the SNR over different speech frames, where the total number of speech frames is 600 (see Figures 9 and 10)). In particular, the average SNR converges to constant value (see the SNR value achieved at $\sigma^2 = \sigma_{\text{ref}}^2$). If the adaptation is well performed, then for each speech signal having variance $\sigma^2$ that is different from the reference one ($\sigma^2 \neq \sigma_{\text{ref}}^2$), the same average SNR values may be obtained:

However, the enhanced performance is achieved at the expense of slightly increasing the bit rate by 0.25 bit/sample (information about the variance of the prediction error and

FIGURE 9 Prediction gain over different speech frames for FLP and LP of orders one and two. FLP, fractional linear prediction; LP, linear prediction.

FIGURE 10 (a) SNR over speech frames for different algorithms; (b) Speech waveforms of the original, reconstructed and error signal in case of the proposed 2-bit ADM algorithm ($L = 32$-level QL, $T = 32$-level QL, $\lambda = 0.48$, frame length 5 ms). ADM, adaptive delta modulation; SNR, signal-to-noise ratio.
the FLP coefficient). Note that for Gaussian quantizers, the expected decrease in signal quality when the bit rate drops by 0.25 bit/sample is approximately 1.32 dB [1, 2]. Thus, assuming the equivalent bit rate of 2 bits/sample, the achieved gains in maximal and average SNR within the considered range (−40 dB, 40 dB) are 5.9 and 7.8 dB, respectively, with respect to the baselines in [13]. For the baseline in [14], the gains in maximal and average SNR are 5 and 8.5 dB, respectively.

6 | CONCLUSION

The 2-bit ADM based on uniform scalar quantization and FLP for frame-by-frame processing of Gaussian sources has been presented. The iterative design method using the Q-function approximations has been proposed for the quantizer implementation, which provides smaller computational complexity. On the other hand, the FLP model with the memory of two samples provides higher prediction gain with respect to the optimal first-order LP without increasing the transmission costs. The established algorithm has shown increased performance in terms of SNR versus its existing counterparts, either frame-based or instantaneously based, using speech as a test signal. In particular, the performance analysis has revealed that it achieves approximately 1.7 dB higher SNR than the frame-based optimal 2-bit ADM baseline while working in same dynamic range and at the same bit rate. On the other hand, in comparison with the two instantaneously based baselines, significantly higher SNR values have been observed—approximately 5–5.9 dB in maximal SNR and approximately 7.8–8.5 dB in the average SNR. Furthermore, a much wider dynamic range has been provided by the proposed algorithm.

A comparison of theoretical and experimental results has been provided in which good matching is perceived.

The proposed algorithm can be successfully applied in quantization/compression of different highly correlated signals following the Gaussian distribution, such as speech signal. Other potential areas of application include analog-to-digital (A/D) conversion, in scenarios where the application of other A/D systems is limited because of lower dynamic range, or as the approximations for high-complexity systems. Future work will include its implementation in biomedical signal processing for signals that do not follow the Gaussian distribution.

ACKNOWLEDGEMENTS

This work has been supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia, Science Fund of the Republic Serbia (Grant No. 6527104, AI- Com-in-AI) as well as bilateral project (337-00-1072019-0911) jointly funded with the Slovak Research and Development Agency.

REFERENCES

1. Jayant, N.S., Noll, P.: Digital Coding of Waveforms: Principles and Applications to Speech and Video. Prentice Hall, New Jersey (1984)
2. Chu, W.C.: Speech Coding Algorithms: Foundation and Evolution of Standardized Coders. John Wiley & Sons, Hoboken (2003)
3. Zrilic, D.G.: Circuits and Systems Based on Delta Modulation. Springer, Berlin (2005)
4. Despotovic, V., et al.: Switched adaptive quantiser for speech compression based on optimal companding and correlation. IET Signal Process. 5(7), 701–707 (2011)
5. Tsang, P., et al.: Low-complexity compression of holograms based on delta modulation. Opt. Commun. 284, 2113–2117 (2011)
6. Tsang, P.W.M., Cheung, W.K., Poon, T.C.: Near computation-free compression of Fresnel holograms based on adaptive delta modulation. Opt. Eng. 50(8), 085802 (2011)
7. Liu, Y., Tang, W.: A delta sigma based finite impulse response filter for EEG signal processing. In: Proceedings of the IEEE International Midwest Symposium on Circuits and Systems (MWSCAS), Fort Collins (2015)
8. Chonga, K.S., et al.: Evaluation of the effect of step size on delta modulation for photoplethysmogram compression. In: Proceedings of the International Conference on Electrical Engineering and Informatics (ICEEI 2013), pp. 815–822.Selangor, Malaysia (2013)
9. Naderi, A., Sawan, M., Savaria, Y.: A low-power 2 GHz data conversion using delta modulation for portable application. Integration. 42, 68–76 (2009)
10. Dandach, S.H., Dasgupta, S., Anderson, B.D.O.: Stability of adaptive delta modulators with forgetting factor and constant inputs. Int. J. Adapt. Contr. Signal Process. 25, 723–739 (2011)
11. Liang-Chien, C., Yi-Chang, C.: A wideband adaptive delta modulator for digitalized video signal. In: Proceedings of the Cross Strait Quad-Regional Radio Science and Wireless Technology Conference, Banqiao District (2012)
12. Gómez-Estern, F., Canudas-de-Wit, C., Rubio, F.R.: Adaptive delta modulation in networked controlled systems with bounded disturbances. IEEE Trans. Automat. Contr. 56(1), 129–134 (2011)
13. Aldajani, M.A., Sayed, A.H.: A stable adaptive structure for delta modulation with improved performance. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal, Salt Lake City (2001)
14. Prosalentis, E.A., Tombras, G.S.: 2-bit adaptive delta modulation system with improved performance. EURASIP J. Adv. Signal Process. 16286 (2006)
