

Quantization Errors in Overlapped Block Digital Filtering Methods

Mustafa Daloglu
 Dept. of Electrical and Electronics Eng.
 Bilkent University
 Bilkent, Ankara, Turkey
 email: mudaloglu@gmail.com

Erchin Serpedin
 Dept. of Electrical and Computer Eng.
 Texas A&M University
 College Station, Texas, USA
 email: serpedin@ece.tamu.edu

Abstract—In digital signal processing applications involving filtering long sequences, block filtering methods like overlap-save and overlap-add are widely used. Like all finite-precision applications, overlap-save and overlap-add methods are also affected by quantization errors. The goal of this paper is to calculate and make a quantitative comparison of the overall quantization noise resulting from the two methods in terms of power (variance) of the quantization noise. Multiple quantization noise sources are taken into consideration in the computation of the variances. The calculations reveal that the overlap-add approach is more prone to quantization noise compared to the overlap-save approach due to the addition of overlapping sections between overlap-add output blocks.

Keywords—block filter; quantization noise; overlap-add; overlap-save;

I. INTRODUCTION

Overlap-save and overlap-add block filtering algorithms are frequently used to implement block finite impulse response (FIR) filters [1], particularly when the input sequences are long. In these widely used approaches, segments from the long input signal are processed using the discrete Fourier transform (DFT), which is generally implemented via the fast Fourier transform (FFT) [2]. In all the computational stages of the block filtering approaches, quantization errors occur due to the necessity of expressing numerical quantities with finite precision in digital signal processors. The goal of this study is to estimate the overall quantization noise in the output blocks resulting from the overlap-save (OLS) and overlap-add (OLA) methods, and compare these errors in terms of quantization noise power. In order to compute the overall quantization noise in the output blocks, multiple error sources related to the different stages of the computation must be considered, as for example, it is the case with the errors resulting from the analog-to-digital (A/D) conversion and the DFT computation stage.

Individual quantization error sources have been amply investigated in previous studies. These include the quantization errors from the A/D conversion process [3,4], errors generated from a standard DFT computation [3,4] and quantization errors from an FFT process [3,4,5]. However, there is no previous attempt to investigate the overall quantization error present in the output blocks of the OLA approach and OLS scheme, to the best of our knowledge. In the following

sections, we will try to estimate and compare the power of the overall quantization error in the output blocks of the OLA and OLS processes computed via the standard DFT.

The rest of this paper is organized as follows. In Section II, the probabilistic properties of the quantization noise generated in the different stages of the computational process are described. In Section III, the variance of the quantization error present in the output of the OLS blocks is calculated. In Section IV, the same calculation is conducted for the output blocks of the OLA. A brief numerical study is conducted to illustrate the difference in terms of quantization noise power between OLA and OLS methods. In Section V, the error variances from the two different OLA and OLS block filters are compared and several concluding remarks are presented.

Throughout the paper, the DFT and the inverse DFT (IDFT) of any signal block are computed via multiplication with the DFT and IDFT matrix, respectively. The $N \times N$ DFT matrix for an N -point DFT computation is defined as [6]:

$$W_N = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & e^{-j\frac{2\pi}{N}} & \dots & e^{-j\frac{2\pi(N-1)}{N}} \\ \vdots & \vdots & & \vdots \\ 1 & e^{-j\frac{2\pi(N-1)}{N}} & \dots & e^{-j\frac{2\pi(N-1)^2}{N}} \end{bmatrix}. \quad (1)$$

The $N \times N$ IDFT matrix for an N -point IDFT computation is given by [6]:

$$W_N^{-1} = \frac{1}{N} W_N^H, \quad (2)$$

where H stands for the complex conjugate transpose (Hermitian). In addition, x denotes the time-domain input sequence and \hat{x} represents the DFT of x , and it is considered to belong to the frequency domain. This relationship can be expressed compactly as:

$$\hat{x} = W_N x. \quad (3)$$

Similarly, h denotes the time-domain impulse response of the filter and \hat{h} represents the DFT of h .

II. PROBABILISTIC PROPERTIES OF THE QUANTIZATION NOISE

In this section, we describe from a statistical perspective the quantization noises generated in the different stages of

the OLA and OLS methods. The quantization noise in all of these stages is modeled as an additive noise (or error) signal [4, p.119]. A statistical model is used for this additive noise with the assumptions that: “the error sequence e is a sample sequence of a stationary random process, e is uncorrelated with the original signal, the random variables (RV) of the error process are uncorrelated and the probability distribution functions (PDF) of the error process are uniform over the range of the quantization error [4, p.120].” For quantizers using rounding in quantization, the amplitude of e is in the range:

$$\frac{-\Delta}{2} < e < \frac{\Delta}{2}. \quad (4)$$

With the assumptions above, the mean value of e is zero and its variance (power) is [4, p.120]:

$$\sigma_e^2 = \frac{\Delta^2}{12}. \quad (5)$$

A. Quantization Noise in the A/D Conversion

Quantization noise occurs at the A/D conversion stage, where the analog-to-digital converter (ADC) is the quantizer. Referring to the noise model above, the step size Δ_{adc} of a $B + 1$ -bit two’s complement ADC is in general given by [4, p.118]:

$$\Delta_{adc} = \frac{2X_m}{2^{B+1}} = \frac{X_m}{2^B}, \quad (6)$$

where X_m is called the full-scale level of the ADC [4, p.118]. Consequently, the variance of the ADC quantization noise σ_{adc}^2 is expressed as:

$$\sigma_{adc}^2 = \frac{\Delta_{adc}^2}{12} = \frac{X_m^2}{3} 2^{-(2B+2)}. \quad (7)$$

B. Quantization of the Filter Coefficients

The filter coefficients, or the impulse response h , are stored in a finite-precision digital environment in digital filtering applications which also require quantization. If the impulse response coefficients are rounded to $B + 1$ bits, then the step size of this quantization Δ_h will be [4,p.347], [3,p.580]:

$$\Delta_h = 2^{-B}. \quad (8)$$

Then, the variance of this noise σ_h^2 will be:

$$\sigma_h^2 = \frac{\Delta_h^2}{12} = \frac{2^{-(2B+2)}}{3}. \quad (9)$$

C. Quantization Noise in an N -Point DFT

The DFT operation is also implemented in a finite-precision digital environment, for this reason quantization noise is also present in the computation of a DFT. The main source of the error is represented by the round-off errors due to the multiplications performed in the DFT with fixed point arithmetic; therefore, we assume that the DFT coefficients themselves and the addition operations are not quantized. Assuming that the input signal is also complex

valued, there are N complex-valued multiplications (thus $4N$ real-valued multiplications) in the direct computation of the DFT. We also assume that these $4N$ quantization errors are mutually uncorrelated and also uncorrelated with the input sequence [3, p.487]. If the numbers are represented with $B + 1$ bits, the variance of the error resulting from a single real multiplication would be $2^{-2B}/12$ [4, p.630]. Since we have $4N$ mutually uncorrelated errors, the variance of the overall error would be expressed as [4, p.631], [3, p.487]:

$$\sigma_{dft}^2 = \frac{N}{3} 2^{-2B}. \quad (10)$$

D. Quantization Noise in the Multiplication $\hat{x} \odot \hat{h}$

After computing the DFTs of the input signal and the impulse response, the next step is to multiply the two N -point frequency domain sequences term by term using the Hadamard product “ \odot ”. This operation corresponds to a circular convolution in the time-domain and it is actually representing the filtering process. For each term, there will be a single complex-valued multiplication or 4 real-valued multiplications. Using the same approach as above, we can directly state that the error corresponding to this operation (supposing that the representation assumes $B + 1$ bits) will have a variance of:

$$\sigma_{mult}^2 = \frac{4}{12} 2^{-2B} = \frac{1}{3} 2^{-2B}. \quad (11)$$

E. Quantization Noise in an N -Point IDFT

The IDFT operation is symmetric to the DFT operation. Therefore, the previous analysis conducted for the DFT is directly extendable to the IDFT operation. The variance of the error generated by the IDFT operation will be given by:

$$\sigma_{idft}^2 = \frac{N}{3} 2^{-2B}. \quad (12)$$

III. QUANTIZATION NOISE IN THE OLS METHOD

One of the most widely used approaches for block digital filtering is the overlap and save method [7]. Considering an N -point input segment x and a P -point impulse response h , the OLS method corresponds to implementing an N -point circular convolution and identifying the part of this convolution that corresponds to a linear convolution. Then, patching these segments together to form the output block. This can be obtained by dividing x into sections of length N such that each input section overlaps the preceding section by $P - 1$ points, and by discarding the first $P - 1$ samples of the output block that result from the N -point circular convolution. The remaining $N - (P - 1)$ samples correspond to the linear convolution and can be patched block by block to form the output [4, p.558].

In order to implement the N -point circular convolution, we will first pad the impulse response with $N - P$ zeros so that it will have the same length as the input block (assuming that $N > P$). Then both of the time-domain

sequences will be transformed with the DFT operation and multiplied in the frequency-domain. The resulting sequence will be converted back to the time-domain via the IDFT operation. The quantization errors will be introduced step by step during the process and will be carried to the output. Any quantization error is denoted as an additive random vector with the same length as that of the original signal, i.e., it is modeled as a vector of uncorrelated complex random variables with a length N .

A. Computation of the Error Term

The first source of quantization error is the quantization of the input signal by the ADC and the quantization of the filter coefficients:

$$\begin{aligned} \mathbf{x} &= \mathbf{x}_0 + \Delta\mathbf{x} \\ \mathbf{h} &= \mathbf{h}_0 + \Delta\mathbf{h}, \end{aligned} \quad (13)$$

where \mathbf{x}_0 , \mathbf{h}_0 are the original vectors and $\Delta\mathbf{x}$, $\Delta\mathbf{h}$ are the random vectors, all of length N . The next step is to transform both of these vectors to the frequency-domain via the multiplication with the N -point DFT matrix \mathbf{W} , which will also generate quantization errors:

$$\begin{aligned} \widehat{\mathbf{x}} &= \mathbf{W}\mathbf{x}_0 + \mathbf{W}\Delta\mathbf{x} + \widehat{\delta\mathbf{x}} \\ &= \widehat{\mathbf{x}}_0 + \widehat{\Delta\mathbf{x}} + \widehat{\delta\mathbf{x}}, \end{aligned} \quad (14)$$

where $\widehat{\delta\mathbf{x}}$ is the quantization noise generated from the DFT operation. In a similar fashion:

$$\begin{aligned} \widehat{\mathbf{h}} &= \mathbf{W}\mathbf{h}_0 + \mathbf{W}\Delta\mathbf{h} + \widehat{\delta\mathbf{h}} \\ &= \widehat{\mathbf{h}}_0 + \widehat{\Delta\mathbf{h}} + \widehat{\delta\mathbf{h}}. \end{aligned} \quad (15)$$

At this stage we multiply both vectors using the Hadamard product " \odot " in the frequency-domain to form the output:

$$\widehat{\mathbf{y}} = \widehat{\mathbf{h}} \odot \widehat{\mathbf{x}} + \widehat{\mu\mathbf{y}}, \quad (16)$$

where $\widehat{\mu\mathbf{y}}$ is the quantization error vector generated from the finite-precision multiplications at this stage. After taking the Hadamard product of $\widehat{\mathbf{h}}$ and $\widehat{\mathbf{x}}$ and discarding the terms that result from the multiplication of two error terms, it follows that

$$\begin{aligned} \widehat{\mathbf{y}} &= \widehat{\mathbf{h}}_0 \odot \widehat{\mathbf{x}}_0 + \widehat{\mathbf{h}}_0 \odot \widehat{\Delta\mathbf{x}} + \widehat{\mathbf{h}}_0 \odot \widehat{\delta\mathbf{x}} \\ &\quad + \widehat{\Delta\mathbf{h}} \odot \widehat{\mathbf{x}}_0 + \widehat{\delta\mathbf{h}} \odot \widehat{\mathbf{x}}_0 + \widehat{\mu\mathbf{y}}. \end{aligned} \quad (17)$$

The last quantization error source will be the one induced by the IDFT operation. We will model the noise induced by the IDFT in a similar way to that generated by the DFT process:

$$\mathbf{y} = \frac{1}{N} \mathbf{W}^H \widehat{\mathbf{y}} + \delta\mathbf{y}, \quad (18)$$

where $\delta\mathbf{y}$ stands for the quantization error vector generated by the IDFT operation.

In order to compute the power of the overall error, we express \mathbf{y} in terms of the original signal and an error term:

$$\mathbf{y} = \mathbf{y}_0 + \Delta\mathbf{y} \rightarrow \mathbf{y}_0 = \frac{1}{N} \mathbf{W}^H (\widehat{\mathbf{h}}_0 \odot \widehat{\mathbf{x}}_0). \quad (19)$$

Then we can express the final error in open form as:

$$\begin{aligned} \Delta\mathbf{y} &= \frac{1}{N} \mathbf{W}^H (\widehat{\mathbf{h}}_0 \odot \widehat{\Delta\mathbf{x}}) + \frac{1}{N} \mathbf{W}^H (\widehat{\Delta\mathbf{h}} \odot \widehat{\mathbf{x}}_0) \\ &\quad + \frac{1}{N} \mathbf{W}^H (\widehat{\mathbf{h}}_0 \odot \widehat{\delta\mathbf{x}}) + \frac{1}{N} \mathbf{W}^H (\widehat{\delta\mathbf{h}} \odot \widehat{\mathbf{x}}_0) \\ &\quad + \frac{1}{N} \mathbf{W}^H (\widehat{\mu\mathbf{y}}) + \delta\mathbf{y}. \end{aligned} \quad (20)$$

As expressed above, the overall error is composed of 6 terms, which will be denoted respectively as \mathbf{z}_i , where $1 \leq i \leq 6$, for easy reference in the further stages:

$$\Delta\mathbf{y} = \sum_{i=1}^6 \mathbf{z}_i. \quad (21)$$

The error vector \mathbf{z}_i is an $N \times 1$ vector of complex random variables generated from the i^{th} quantization error source.

B. Variance of the Error Terms

The calculation of the output signal \mathbf{y} , showed that there are multiple quantization noise sources in the process, which are assumed to generate mutually uncorrelated and zero-mean noise vectors. For the computation of the variance, it is important to note that the noise vectors are mutually uncorrelated and their entries are also uncorrelated one with respect to the other. In other words, the noise terms in an individual vector are all uncorrelated with each other. With these assumptions, we will determine our strategy to calculate the variance of the overall noise term using $\mathbf{z}_{N \times 1}$ as a generic vector with a variance of σ_z^2 . Note that \mathbf{z} is a zero-mean noise vector:

$$E(\mathbf{z}) = \begin{bmatrix} E(z_1) \\ \vdots \\ E(z_N) \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0}. \quad (22)$$

Using this property and the fact that the noise terms are uncorrelated with each other, it follows that:

$$E(\mathbf{z}\mathbf{z}^H) = \begin{bmatrix} E(|z_1|^2) & \dots & 0 \\ & \ddots & \\ 0 & \dots & E(|z_N|^2) \end{bmatrix} = \sigma_z^2 \mathbf{I}. \quad (23)$$

Note that σ_z^2 is the variance of a single term in the vector \mathbf{z} . In order to calculate the variance of the vector, we need to add up all the variances corresponding to all the entries in the vector. Mathematically, this operation amounts to

$$\sigma_{z\text{-vector}}^2 = E(\mathbf{z}^H \mathbf{z}) = N\sigma_z^2. \quad (24)$$

Once the variances of the individual noise sources are calculated, it is possible to find the overall variance just by adding the individual ones due to the uncorrelated nature

of the noise sources [8, p.220]. We will denote the variance of an individual noise vector \mathbf{z}_i as σ_i^2 , where $1 \leq i \leq 6$ denotes the respective order.

1) σ_1^2 and σ_2^2 : Let's start with \mathbf{z}_1 and proceed using the strategy developed for the generic vector.

$$\begin{aligned} \mathbf{z}_1 &= \frac{1}{N} \mathbf{W}^H (\widehat{\mathbf{h}}_0 \odot \widehat{\Delta \mathbf{x}}) \rightarrow E(\mathbf{z}_1^H \mathbf{z}_1) = \\ &E\left(\frac{1}{N^2} (\widehat{\mathbf{h}}_0 \odot \widehat{\Delta \mathbf{x}})^H \mathbf{W} \mathbf{W}^H (\widehat{\mathbf{h}}_0 \odot \widehat{\Delta \mathbf{x}})\right) \quad (25) \\ &= E\left(\frac{1}{N} (\widehat{\mathbf{h}}_0 \odot \widehat{\Delta \mathbf{x}})^H (\widehat{\mathbf{h}}_0 \odot \widehat{\Delta \mathbf{x}})\right), \end{aligned}$$

since $\mathbf{W} \mathbf{W}^H = N \mathbf{I}$. Writing in open form:

$$\begin{aligned} E(\mathbf{z}_1^H \mathbf{z}_1) &= E\left(\frac{1}{N} \sum_{i=1}^N |\widehat{h}_0[i]|^2 |\widehat{\Delta x}[i]|^2\right) \quad (26) \\ &= \frac{1}{N} \sum_{i=1}^N |\widehat{h}_0[i]|^2 E(|\widehat{\Delta x}[i]|^2), \end{aligned}$$

where $\widehat{\Delta x}$'s are the probabilistic terms. Note that $\widehat{\Delta \mathbf{x}} = \mathbf{W} \Delta \mathbf{x}$, and $\Delta \mathbf{x}$ is the zero-mean noise vector generated at the ADC composed of terms with a variance σ_{adc}^2 as defined previously. Therefore, the following relations hold:

$$\begin{aligned} \widehat{\Delta \mathbf{x}} &= \mathbf{W} \Delta \mathbf{x}, \quad E(\widehat{\Delta \mathbf{x}}) = 0 \\ E(\widehat{\Delta \mathbf{x}} \widehat{\Delta \mathbf{x}}^H) &= E(\mathbf{W} \Delta \mathbf{x} (\mathbf{W} \Delta \mathbf{x})^H) \quad (27) \\ &= \mathbf{W} E(\Delta \mathbf{x} \Delta \mathbf{x}^H) \mathbf{W}^H \\ &= \mathbf{W} \sigma_{adc}^2 \mathbf{I} \mathbf{W}^H = \sigma_{adc}^2 N \mathbf{I}. \end{aligned}$$

This means that each term in the vector $\widehat{\Delta \mathbf{x}}$ has a variance equal to $\sigma_{adc}^2 N$. Thus, we can replace any $E(|\widehat{\Delta x}[i]|^2)$ term with $\sigma_{adc}^2 N$. Therefore, it follows further that

$$\begin{aligned} E(\mathbf{z}_1^H \mathbf{z}_1) &= \frac{1}{N} \sum_{i=1}^N |\widehat{h}_0[i]|^2 \sigma_{adc}^2 N \quad (28) \\ &= \sigma_{adc}^2 \sum_{i=1}^N |\widehat{h}_0[i]|^2. \end{aligned}$$

Note also that $\widehat{\mathbf{h}}_0 = \mathbf{W} \mathbf{h}_0$. Thus:

$$\begin{aligned} \sum_{i=1}^N |\widehat{h}_0[i]|^2 &= \widehat{\mathbf{h}}_0^H \widehat{\mathbf{h}}_0 = (\mathbf{W} \mathbf{h}_0)^H (\mathbf{W} \mathbf{h}_0) \\ &= \mathbf{h}_0^H \mathbf{W}^H \mathbf{W} \mathbf{h}_0 = N \mathbf{h}_0^H \mathbf{h}_0 \quad (29) \\ &= N \sum_{i=1}^N h_0^2[i]. \end{aligned}$$

To sum up, the variance (σ_1^2) of the quantization noise from the first source (\mathbf{z}_1) is found to be:

$$\sigma_1^2 = N \sigma_{adc}^2 \sum_{i=1}^N h_0^2[i]. \quad (30)$$

This also means that any term in the vector \mathbf{z}_1 has a variance equal to $\sigma_{adc}^2 \sum_{i=1}^N h_0^2[i]$.

We notice that the variance (σ_2^2) calculation of the quantization noise from the second source (\mathbf{z}_2) is similar with the first calculation. For this reason, we can directly state that:

$$\sigma_2^2 = N \sigma_h^2 \sum_{i=1}^N x_0^2[i], \quad (31)$$

where σ_h^2 is the noise caused by the quantization of the impulse response.

2) σ_3^2 and σ_4^2 : The strategy for the calculation of these variances is the same as the one used for the first variance. The vector variance σ_3^2 of the noise generated from the third source \mathbf{z}_3 could be directly expressed as:

$$\sigma_3^2 = \frac{1}{N} \sum_{i=1}^N (|\widehat{h}_0[i]|^2 E(|\widehat{\delta x}[i]|^2)). \quad (32)$$

As stated previously, $\widehat{\delta \mathbf{x}}$ is the zero-mean noise vector generated by the N -point DFT process with a variance σ_{dft}^2 . Then, we can replace any $E(|\widehat{\delta x}[i]|^2)$ term with σ_{dft}^2 . We have previously found out that $\sum_{i=1}^N |\widehat{h}_0[i]|^2 = N \sum_{i=1}^N h_0^2[i]$. Thus:

$$\sigma_3^2 = \sigma_{dft}^2 \sum_{i=1}^N h_0^2[i]. \quad (33)$$

As it was the case for the first set of variances, calculation of σ_3^2 and σ_4^2 is quite similar. Therefore, one can directly express σ_4^2 as follows:

$$\sigma_4^2 = \sigma_{dft}^2 \sum_{i=1}^N x_0^2[i]. \quad (34)$$

3) σ_5^2 and σ_6^2 : The computation of these variances is less complex with respect to the previous variance calculations because \mathbf{z}_5 and \mathbf{z}_6 are generated at the final stages of the OLS process. For this reason, these noise vectors are carried through fewer computational stages in the OLS process with respect to the previous noise vectors. Calculation of \mathbf{z}_5 relies on the following relation:

$$\begin{aligned} \sigma_5^2 &= E\left(\left[\frac{1}{N} \mathbf{W}^H \widehat{\boldsymbol{\mu}} \mathbf{y}\right]^H \left[\frac{1}{N} \mathbf{W}^H \widehat{\boldsymbol{\mu}} \mathbf{y}\right]\right) \quad (35) \\ &= \frac{1}{N} E(\widehat{\boldsymbol{\mu}} \mathbf{y}^H \widehat{\boldsymbol{\mu}} \mathbf{y}) = \frac{1}{N} \sum_{i=1}^N E(|\widehat{\boldsymbol{\mu}} \mathbf{y}[i]|^2). \end{aligned}$$

From our previous definitions, $\widehat{\boldsymbol{\mu}} \mathbf{y}$ was the zero-mean noise vector generated by the Hadamard product of two complex N -point vectors in a finite precision environment with a variance σ_{mult}^2 . Replacing $E(|\widehat{\boldsymbol{\mu}} \mathbf{y}[i]|^2)$ with σ_{mult}^2 , we obtain:

$$\sigma_5^2 = \sigma_{mult}^2. \quad (36)$$

The last quantization noise source is the IDFT process, generating a noise vector $\mathbf{z}_6 = \delta\mathbf{y}$ with terms having a variance σ_{idft}^2 :

$$\sigma_6^2 = E(\mathbf{z}_6^H \mathbf{z}_6) = \sum_{i=1}^N E(|\delta y[i]|^2). \quad (37)$$

The resulting value is:

$$\sigma_6^2 = N\sigma_{idft}^2. \quad (38)$$

Exploiting the assumption that all these 6 noise sources are mutually uncorrelated, we express the overall vector variance in an N -point output block computed via the OLS method as:

$$\begin{aligned} \sigma_{OLS}^2 &= (N\sigma_{adc}^2 + \sigma_{dft}^2) \sum_{i=1}^N h_0^2[i] \\ &+ (N\sigma_h^2 + \sigma_{dft}^2) \sum_{i=1}^N x_0^2[i] \\ &+ \sigma_{mult}^2 + N\sigma_{idft}^2. \end{aligned} \quad (39)$$

All these variances were numerically determined in the second section. Considering a $(B+1)$ -bit digital system that assumes a $(B+1)$ -bit ADC with a full-scale level of X_m , we find the variance in an N -point OLS output block to be:

$$\begin{aligned} \sigma_{OLS}^2 &= \frac{N}{3} 2^{-2B} \left(\left(\frac{X_m^2 + 4}{4} \right) \sum_{i=1}^N h_0^2[i] \right. \\ &\left. + \left(\frac{5}{4} \right) \sum_{i=1}^N x_0^2[i] + \frac{N^2 + 1}{N} \right). \end{aligned} \quad (40)$$

IV. QUANTIZATION NOISE IN THE OLA METHOD

The other widely used block digital filtering method is the overlap and add method [1]. Considering an L -point input signal block \mathbf{x} and a P -point filter impulse response \mathbf{h} , the OLA method convolves \mathbf{x} and \mathbf{h} using a linear convolution operation. Thus, it results in an output block of length $N = L + P - 1$. In order to perform this operation in the frequency-domain via DFT-IDFT, we must pad both \mathbf{x} and \mathbf{h} with zeros such that both of them become vectors of the same length N , which corresponds to performing an N -point circular convolution. When patching the N -point output blocks together, the first $P - 1$ terms of each block are overlapped and added with the last $P - 1$ terms of the previous block [4, p.558].

In terms of variance calculation, an important difference between the OLS scheme and OLA method is that the OLA method requires the addition of the first $P - 1$ terms of the current output block with the overlapping $P - 1$ terms of the previous output block. This means that in these overlapping sections, the quantization errors present in the two individual blocks are also added. Other than this

difference, the computation stages for both the OLS scheme and the OLA method are the same. Thus, we can use our previous variance calculations for the OLS and modify them appropriately to include the additional noise coming from overlapping $P - 1$ terms.

Assume that the OLA method results in an output block of length N , and consider an N -point slice from the output sequence of the OLS. Just before overlapping and adding the first $P - 1$ terms of the current OLA block with the last $P - 1$ terms of the previous block, both blocks from the OLA and the OLS process present the same variance. Assuming that the overlapping blocks of the OLA are uncorrelated, the $P - 1$ noise terms being added to each other are also uncorrelated. It would be correct to state that the variance of the terms in the overlapping section of the OLA block will be doubled. With this reasoning, we can express the variance of the quantization noise in an N -point OLA output block (σ_{OLA}^2) in terms of the variance of the quantization noise in an N -point OLS output block (σ_{OLS}^2) as follows:

$$\sigma_{OLA}^2 = \frac{N + P - 1}{N} \sigma_{OLS}^2. \quad (41)$$

Adopting a $(B+1)$ -bit digital system that assumes a $(B+1)$ -bit ADC with a full-scale level of X_m , we can express the numerical value of σ_{OLA}^2 as:

$$\begin{aligned} \sigma_{OLA}^2 &= \frac{N + P - 1}{3} 2^{-2B} \\ &\left(\left(\frac{X_m^2 + 4}{4} \right) \sum_{i=1}^N h_0^2[i] + \left(\frac{5}{4} \right) \sum_{i=1}^N x_0^2[i] + \frac{N^2 + 1}{N} \right). \end{aligned} \quad (42)$$

In practice, the difference in terms of quantization noise variance between OLS and OLA is relatively small. For example, when $N = 256$ and $P = 32$, according to (41), the difference is $10 \log_{10}((N + P - 1)/N) = 0.49dB$, while for $N = 256$ and $P = 128$, the difference amounts to $1.7dB$.

V. CONCLUSION

The quantization noise is a side effect that is inherently present in any digital system due to the necessity of representing the signal samples in finite-precision. Having an important role in digital signal processing applications that involve the filtering of long sequences, OLA and OLS methods are also prone to quantization noise once that they get implemented in a digital system. Both methods were implemented in multiple stages like the ADC, DFT or the IDFT stage and multiple quantization noise sources from these stages were taken into consideration. It was observed that the power of the quantization noise in an OLA output block tended to be higher than the quantization noise in an OLS output block, mainly because of the addition of the overlapping sections in the OLA process. However, it was also observed that the difference in the quantization noise power was highly dependent on the length of the input sequences.

The computation of the variances would have been more realistic if an FFT algorithm was used instead of a direct DFT computation because most digital signal processing applications take advantage of FFT algorithms. This study can be expanded to include an FFT algorithm instead of a direct DFT implementation. However, it is expected that the quantitative comparison of the quantization noises in OLA and OLS will not change.

ACKNOWLEDGMENT

This work was made possible by the support offered by QNRF-NPRP grants 09-341-2128 and 4-1293-2-513.

REFERENCES

- [1] M. J. Narasimha, "Linear convolution using skew-cyclic convolutions," *IEEE Signal Processing Letters*, vol. 14, 2007, pp. 173–176.
- [2] A. Daher, E. H. Baghious, G. Burel, and E. Radoi, "Overlap-save and overlap-add filters: optimal design and comparison," *IEEE Transactions on Signal Processing*, vol. 58, 2010, pp. 3066–3075.
- [3] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing: Principles, Algorithms and Applications*. Upper Saddle River: Prentice Hall, 1996.
- [4] A. V. Oppenheim and R. W. Schaffer, *Discrete Time Signal Processing*. Englewood Cliffs: Prentice Hall, 1989.
- [5] A. V. Oppenheim and C. J. Weinstein, "Effects of finite register length in digital filtering and the fast Fourier transform," *Proceedings of IEEE*, vol. 60, 1972, pp. 957–976.
- [6] A. E. Cetin, *Lecture Notes on Discrete-Time Signal Processing: EE424 Course at Bilkent University, Ankara, Turkey, 2012*.
- [7] G. Burel, "Optimal design of transform-based block digital filters using a quadratic criterion," *IEEE Transactions on Signal Processing*, vol. 52, 2004, pp. 1964–1974.
- [8] D. P. Bertsekas and J. N. Tsitsiklis, *Introduction to Probability*. Belmont: Athena Scientific, 2008.