IMTC 2006 - Instrumentation and Measurement
Technology Conference
Sorrento, Italy, April 24–27, 2006

Digital Non-Subtractive Dither: Necessary and Sufficient Condition for Unbiasedness, with Implementation Issues

I. Kollár

Department of Measurement and Information Systems,
Budapest University of Technology and Economics,
Budapest, Hungary, H-1521, P.O.Box 91.
Phone: +36 1 463-1774, Fax: +36 1 463-4112, Email: kollar@mit.bme.hu.

Abstract – *Dither theory gives the condition of unbiasedness for non-subtractive dither, which can only be fulfilled with continuous-amplitude dithers. However, within a computer or a DSP, only digital dither can be used. A condition was given also for digital dither, however, the condition and the proof were only approximate. This paper makes his statement precise, and gives a full proof for unbiasedness. Implementation issues are also discussed.*

Keywords – *Dither, roundoff error, quantization, bias, digital signal processing, necessary and sufficient conditions.*

I. INTRODUCTION

In quantization theory, the main concern is unbiased reconstruction of moments of the input random variable from quantized data. This is assured by the so-called Sheppard corrections [1], [2], [3], [4], if the quantization conditions for the input variable hold: the probability density function (PDF) is band-limited, that is, its characteristic function (CF) is zero (or approximately zero) outside the band $(2\pi/q)$, where q is the quantum size.

In digital signal processors, quantization occurs not only as input A/D conversion, but also as roundoff after each arithmetic operation. Therefore, the conditions of unbiasedness after re-quantization of already quantized data are of high importance. However, in these cases, the already discrete inputs to the roundoff operation usually do not fulfill the required conditions. In such cases, quantization theory offers the use of a dither. This paper discusses the use of dither in re-quantization.

II. PRELIMINARIES

Dither theory was first studied in the sixties [5], [6]. Their theorems deal with the so-called subtractive dither. This means that an auxiliary signal, called dither, is added to the signal before quantization, and the error of the output signal is studied after subtraction of the dither. This is can be the case after coding (transmission and/or storage) and reconstruction of the quantized noise with reconstructed dither.

[1] with support of OTKA TS49743

Later it was recognized that in many cases, the dither cannot be removed from the output (non-subtractive dither, NSD), either because it cannot be reconstructed even when it is pseudo-random, or the subtraction cannot be executed because of the limited bit length. This is most often the case in DSP's. For this case, Wannamaker [7] proved a general theorem[2] as

In an NSD quantizing system, $\mathrm{E}[\varepsilon^l]$ *is independent of the distribution of the system input,* x*, for* $l = 1, 2, \dots, M$ *if and only if*

$$\Phi_{\mathrm{d}}^{(i)}\left(\frac{k}{\Delta}\right) \tag{1}$$

for $\forall k \in Z_0$ *and* $i = 0, 1, 2, \dots M-1$*, with* $\varepsilon = d + \nu$ *being the total quantization noise, in which* d *denotes the dither, and* ν *denotes the error caused by quantization,* Φ *being the characteristic function, and the superscript* $^{(i)}$ *denoting the* i^{th} *derivative.* Z_0 *is the set of nonzero integers.*

Examples for such dithers are the uniform dither in $(-q/2, q/2)$ ($M = 1$, zero-order dither), or triangular dither in $(-q, q)$ ($M = 2$, first-order dither).

Already Wannamaker recognized that a digital dither (a dither represented by finite bit numbers) can never exactly fulfill this theorem, since a discrete PDF has a periodic CF which cannot have all the required zeros in the derivatives. Therefore, he formulated another theorem for digital dither.

Wannamaker's theorem for digital dither:
For a digital NSD quantizing system in which requantization is used to remove the L *least significant bits of binary data,* $\mathrm{E}[\varepsilon^l]$ *is independent of the distribution of the system input,* x*, for* $l = 1, 2, \dots, M$ *if a non-subtractive digital dither (with the same precision as the input data) is applied for which*

$$\widetilde{\Phi}_{\mathrm{d}}^{(i)}\left(\frac{k}{\Delta}\right) . \tag{2}$$

for $\forall k \in Z_0$ *and* $i = 0, 1, 2, \dots M-1$*, with* $\varepsilon = d + \nu$ *being the total quantization noise, in which* d *denotes the dither, and* ν *denotes the error caused by quantization,* $\widetilde{\Phi}$ *being the characteristic function, and the superscript* $^{(i)}$ *denoting the* i^{th} *derivative.* Z_0 *is the set of nonzero integers.*

[2] His notations were slightly changed for this paper.

However, this theorem, while it seems to comply with the requirements of digital processing, cannot be precisely fulfilled, since this is an equivalent of the theorem above, with the requirement that the CF of the discrete dither has zeros at all the required places, which is impossible to fulfill. Wannamaker circumvented the problem by requiring *approximate* fulfillment of the conditions.

Figure 1 illustrates distribution of a digital dither in the amplitude domain, and Fig. 2 its characteristic function.

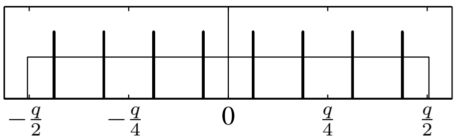

Fig. 1. Illustration of the distribution of a digital dither in the amplitude domain

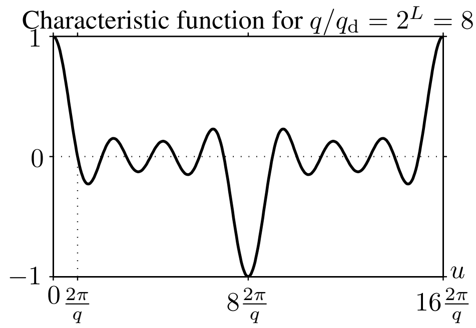

Fig. 2. Characteristic function of a digital dither illustrated in Fig. 1

The first observation one can make is that the characteristic function has the required zeros, *except* at integer multiples of $2\pi/q_\mathrm{d}$, where q_d is the quantum size (amplitude resolution) of the dither. The is a general rule: since the digital dither has discrete distribution, that is, its PDF is sampled, at these places the CF cannot be zero.

It can be checked if these nonzero values disturb the moments described in Wannamaker's theorem or not. A little experimentation shows that they do not disturb it. Even a modified theorem can be formulated for digital nonsubtractive dither:

Quantizing Theorem with Digital Dither (QTDD)

For a digital system in which re-quantization is used to remove the L least significant bits of binary data, $\mathrm{E}\{\varepsilon^m|x\}$ *has the same value for all representable values of* x *for* $m = 1, 2, \ldots, r$*, if a non-subtractive digital dither (with the same precision as the input data) is applied for which*

$$\left.\frac{\mathrm{d}^t\,\Phi_\mathrm{d}(u)}{\mathrm{d}u^t}\right|_{u=l\Psi} = 0\,, \qquad (3)$$

for $t = 0, 1, \ldots, r-1$, at $l = 1, 2, \ldots, 2^L - 1$ ($(r-1)$-order digital dither).

The proof follows from examination of the conditional CF of ε, given in Eq. (4).

$$\begin{aligned}\Phi_{\varepsilon|x}(u_\varepsilon) &= \sum_{l=-\infty}^{\infty} \Phi_\mathrm{d}\left(u_\varepsilon + l\Psi\right) e^{jl\Psi x}\,\mathrm{sinc}\left(\frac{q(u_\varepsilon + l\Psi)}{2}\right)\\ &= \Phi_\mathrm{d}\left(u_\varepsilon\right)\mathrm{sinc}\left(\frac{qu_\varepsilon}{2}\right)\\ &\quad + \sum_{\substack{l=-\infty\\ l\neq 0}}^{\infty} \Phi_\mathrm{d}\left(u_\varepsilon + l\Psi\right) e^{jl\Psi x}\,\mathrm{sinc}\left(\frac{q(u_\varepsilon + l\Psi)}{2}\right). \qquad (4)\end{aligned}$$

The required moments are not influenced by the value of x in the infinite sum of the last part, because

- For the first moment, we need to examine the sum (see Eq. (4)), for the values $x = kq_\mathrm{d} = k2^{-L}q$. We will look at the terms for which $l \neq 2^L\lambda$ (λ is an integer), and at the terms for which $l = 2^L\lambda$, separately.

$$\begin{aligned}\mathrm{E}\{\varepsilon|x\} &= \frac{1}{j}\frac{\mathrm{d}}{\mathrm{d}u_\varepsilon}\left(\sum_{l=-\infty}^{\infty}\Phi_\mathrm{d}\left(u_\varepsilon + l\Psi\right)e^{jl\Psi x}\right.\\ &\qquad\left.\left.\times\ \mathrm{sinc}\left(\frac{q(u_\varepsilon + l\Psi)}{2}\right)\right)\right|_{u_\varepsilon=0}\\ &= \mu_\mathrm{d}\\ &\quad + \frac{1}{j}\sum_{\substack{l=-\infty\\ l\neq 2^L\lambda}}^{\infty}\Phi_\mathrm{d}\left(l\Psi\right)e^{jl\frac{2\pi}{q}kq_\mathrm{d}}\\ &\qquad\times\left.\frac{\mathrm{d}}{\mathrm{d}u}\,\mathrm{sinc}\left(\frac{q(u_\varepsilon + l\Psi)}{2}\right)\right|_{u_\varepsilon=0}\\ &\quad + \frac{1}{j}\sum_{\substack{\lambda=-\infty\\ \lambda\neq 0}}^{\infty}\Phi_\mathrm{d}\left(2^L\lambda\Psi\right)e^{j2^L\lambda\frac{2\pi}{q}kq_\mathrm{d}}\\ &\qquad\times\left.\frac{\mathrm{d}}{\mathrm{d}u}\,\mathrm{sinc}\left(\frac{q(u_\varepsilon + 2^L\lambda\Psi)}{2}\right)\right|_{u_\varepsilon=0}\\ &= \mu_\mathrm{d}\\ &\quad + \frac{1}{j}\sum_{\substack{l=-\infty\\ l\neq 2^L\lambda}}^{\infty}\Phi_\mathrm{d}\left(l\frac{2\pi}{q}\right)e^{jlk\frac{2\pi}{2^L}}\frac{q}{\pi}\frac{1}{l}\\ &\quad + \frac{1}{j}\sum_{\substack{\lambda=-\infty\\ \lambda\neq 0}}^{\infty}\Phi_\mathrm{d}\left(\lambda\frac{2\pi}{q_\mathrm{d}}\right)e^{j\lambda k2\pi}\frac{q_\mathrm{d}}{\pi}\frac{1}{\lambda}\,. \qquad (5)\end{aligned}$$

  The first sum equals zero because of the theorem's condition (3) for $m = 1$, and the second sum does not depend on k (and thus, does not depend on x).[3] This proves the theorem for $m = 1$.

[3] In addition, if the dither has a distribution symmetric to zero, or a distribution symmetric to any of $nq_\mathrm{d}/2$, $n = \pm1, \pm2, \ldots$, the second sum also equals zero since $e^{-jn\frac{q_\mathrm{d}}{2}\frac{2\pi}{q_\mathrm{d}}}\,\Phi_\mathrm{d}\left(\lambda\frac{2\pi}{q_\mathrm{d}}\right) = e^{jn\frac{q_\mathrm{d}}{2}\frac{2\pi}{q_\mathrm{d}}}\,\Phi_\mathrm{d}\left(-\lambda\frac{2\pi}{q_\mathrm{d}}\right)$, that is, $\Phi_\mathrm{d}\left(\lambda\frac{2\pi}{q_\mathrm{d}}\right) = \Phi_\mathrm{d}\left(-\lambda\frac{2\pi}{q_\mathrm{d}}\right)$, therefore the terms in the sum are negatively symmetric for $\pm\lambda$, so these terms cancel out pairwise.

- Similarly, for $m > 1$, independence of x (or of k) will be provided, since the terms for $l \neq 2^L\lambda$ disappear because of (3), and x disappears from each term corresponding to $l = 2^L\lambda$:

$$e^{jl\Psi x} = e^{j2^L\lambda\frac{2\pi}{q}kd_{\mathrm{d}}} = e^{j\lambda k2\pi} \equiv 1\,. \tag{6}$$

While functional independence of x is provided, the sum usually does not equal zero. For $m = 1$, an extra condition was given to assure zero value (footnote 3). For $m = 2$, the nonzero value is even more common. If (3) is fulfilled for $r = 2$, it is enough to examine the possibly nonzero elements:

$$\begin{aligned}
\mathrm{E}\{\varepsilon^2|x\} &= \frac{1}{j^2}\frac{\mathrm{d}^2}{\mathrm{d}u_\varepsilon^2}\left(\sum_{l=-\infty}^{\infty}\Phi_{\mathrm{d}}\left(u_\varepsilon + l\Psi\right)e^{jl\Psi x}\right.\\
&\qquad\left.\left.\times\ \mathrm{sinc}\left(\frac{q(u_\varepsilon+l\Psi)}{2}\right)\right)\right|_{u_\varepsilon=0}\\
&= \mathrm{E}\{d^2\} + \mathrm{E}\{n^2\}\\
&\quad - \sum_{\substack{\lambda=-\infty\\ \lambda\neq 0}}^{\infty}\dot{\Phi}_{\mathrm{d}}\left(2^L\lambda\Psi\right)e^{j2^L\lambda\frac{2\pi}{q}kd_{\mathrm{d}}}\frac{\mathrm{d}}{\mathrm{d}u_\varepsilon}\\
&\qquad\left.\times\ \mathrm{sinc}\left(\frac{q(u_\varepsilon+2^L\lambda\Psi)}{2}\right)\right|_{u_\varepsilon=0}\\
&\quad - \sum_{\substack{\lambda=-\infty\\ \lambda\neq 0}}^{\infty}\Phi_{\mathrm{d}}\left(2^L\lambda\Psi\right)e^{j2^L\lambda\frac{2\pi}{q}kd_{\mathrm{d}}}\frac{\mathrm{d}^2}{\mathrm{d}u_\varepsilon^2}\\
&\qquad\left.\times\ \mathrm{sinc}\left(\frac{q(u_\varepsilon+2^L\lambda\Psi)}{2}\right)\right|_{u_\varepsilon=0}\\
&= \mathrm{E}\{d^2\} + \mathrm{E}\{n^2\}\\
&\quad - \sum_{\substack{\lambda=-\infty\\ \lambda\neq 0}}^{\infty}\dot{\Phi}_{\mathrm{d}}\left(\lambda\frac{2\pi}{q_{\mathrm{d}}}\right)\frac{q_{\mathrm{d}}}{\pi}\frac{1}{\lambda}\\
&\quad + \sum_{\substack{\lambda=-\infty\\ \lambda\neq 0}}^{\infty}\Phi_{\mathrm{d}}\left(\lambda\frac{2\pi}{q_{\mathrm{d}}}\right)\frac{q_{\mathrm{d}}^2}{\pi^2}\frac{1}{\lambda^2}\,.
\end{aligned} \tag{7}$$

This is often somewhat larger than $\mathrm{E}\{d^2\} + \mathrm{E}\{n^2\}$. However, the deviation is relatively small, since the first sum is zero for a dither which has a distribution symmetric to zero and has values only on the grid kq_{d} (or at least only on the grid $kq_{\mathrm{d}}/2$, $k = 0, \pm1, \pm2, \ldots$), while the second sum can be upper bounded since $|\,\Phi(u)| \leq 1$:

$$\mathrm{E}\{\varepsilon^2|x\} \leq \frac{q_{\mathrm{d}}^2}{3}\,, \tag{8}$$

which is usually negligible when compared to $(\mathrm{E}\{d\})^2 + (\mathrm{E}\{n\})^2 = (\mathrm{E}\{d\})^2 + q^2/12$. Since this deviation does not depend on x, it can be corrected for.

From QTDD, the important consequence is that the resolution (LSB) of the digital dither should be the same as of the data to be quantized: $q_{\mathrm{d}} = 2^{-L}q$. The theorem provides that this is sufficient. Finer resolution of the dither would be superfluous, coarser resolution would not be sufficient. This theorem is useful for dithering in digital computers and digital signal processors.

A. *Dirac Delta Functions at $q/2 + kq$*

While the above proof is correct, the theorem has an important application limitation. Quantization theory is considered as area sampling of a smooth PDF. When at the edge of such an area there is a Dirac delta function, it is tacitly assumed in the derivation that half of the integral of the Dirac delta belongs to this area, and half of it to the next area. This is a property of Fourier transform pairs, and the proofs are based upon Fourier transform. This corresponds to random-direction quantization of input values equal to $(\text{integer} + 0.5)q$: half of the values at the comparison levels are rounded downwards, half of them are rounded upwards.

The existence of such Dirac delta functions is a common case in re-quantization. In practical processors, however, as well as in simulations in MATLAB, a deterministic algorithm is implemented: such values are either rounded always upwards, or always downwards, or towards zero, or towards $\pm\infty$ (like in MATLAB's `round(·)` function), or convergent rounding is implemented (rounding towards the closest *even* number when the input is exactly at 0.5 LSB distance from two representable numbers). Quantization theory *does not deal with these cases*. Therefore, we have to content ourselves by

- either accepting that convergent rounding averages out the bias for the given input signal,
- or assuming that the probability of the values just at 0.5 LSB from two representable values is very small,
- or saying that if $q_{\mathrm{d}} \ll q$, deviation between theory and practice is negligible, or
- implementing in the processor (in the simulation program) a modification of rounding to correspond to theory: when having a number which equals $(\text{integer}+0.5)$ LSB, either additional dither should determine if rounding is done in the upwards or downwards direction, or the program takes care to do upwards/downwards rounding alternately for the same level.

III. DIGITAL DITHER WITH APPROXIMATELY NORMAL DISTRIBUTION

In a computer, it is easy to generate normally distributed numbers. Either a pseudo-random normal number generator can be used, or several independent, identically distributed random numbers can be added. These normally distributed numbers are then quantized to q_{d} to have a digital dither.

The characteristic function of this dither is

$$\begin{aligned}
\Phi_{\mathrm{d}}(u) = \sum_{\lambda=-\infty}^{\infty} & e^{j(u+\lambda\Psi_{\mathrm{d}})\mu}e^{-\frac{\sigma^2(u+\lambda\Psi_{\mathrm{d}})^2}{2}}\\
&\times\ \mathrm{sinc}\left(\frac{q_{\mathrm{d}}(u+\lambda\Psi_{\mathrm{d}})}{2}\right)\,,
\end{aligned} \tag{9}$$

with $\Psi_\mathrm{d} = 2\pi/q_\mathrm{d}$.

If the common rule $\sigma > q$ is followed for the normally distributed dither, and its mean value is equal to zero, the moments of the dither can be well reconstructed from samples of the digital dither, using Sheppard's corrections. Some similar corrections can be used between moments of $(x+d)'$ and of $(x+d)$, or between moments of $(x+d)' - d$, and of x.

For Gaussian dither with $\sigma > q$, the condition (3) of QTDD is fulfilled with good approximation, thus the moments are independent of x, if x is representable on the grid kq_d. This does not mean however that for *any* value of x, the moments would be unbiased. It is heuristically clear that the digital dither has "roughness" q_d, therefore, if x is arbitrary, the error in Sheppard's first correction may reach $\pm q_\mathrm{d}/2$, and in Sheppard's second correction it may be in the order of magnitude of $q_\mathrm{d}^2/6$. We cannot go into these details here, the errors of this kind can be studied in detail by investigation of the corresponding CF's.

IV. GENERATION OF DIGITAL DITHER

Let us turn now to the generation of digital dither. Random number generators can be realized based on different principles [8]. One of the most popular methods is based on feedback shift registers. These generate $2^N - 1$ pseudorandom bits, where N is the register length, and these can be used to generate pseudorandom numbers. If 2^N is not very large, we notice that the generated dither has a periodic nature, and that using a full period, it goes through every individual step. This latter fact may be used for increased efficiency in averaging: in this special case the sequence to be averaged contains all possible values just once. Therefore, the result of averaging is exact, with no randomness due to dithering.

The first thing we have to decide is the distribution of the dither. We can approximate any distribution by digital means; however, uniform and triangular dithers are by far the most popular ones. We will deal here with these. Gaussian and sinusoidal dithers are also usual and reasonable choices. Their properties can be determined with similar analysis.

As for the number of bits, according to QTDD, it is not reasonable to use a dither which has finer resolution than the variable to be quantized. Thus, the difference of the bit numbers of the accumulator and of the memory (the storage bit number) determine the reasonable bit number of the dither.

When the number of bits is known, the digital representation of the distribution is to be selected. We can consider digital dither as a *finely quantized* version of the continuous one. Therefore, using at least a few bits, we can apply the approximation that the variance is $\mathrm{var}\{d\} \approx \mathrm{var}\{d_\mathrm{c}\} + q_\mathrm{d}^2/12 \approx \mathrm{var}\{d_\mathrm{c}\}$, with d_c being the continuous-time dither, and q_d denoting the dither LSB.

A. Uniformly Distributed Digital Dither

For uniform dither, we have a few, almost equivalent, solutions (Fig. 3).

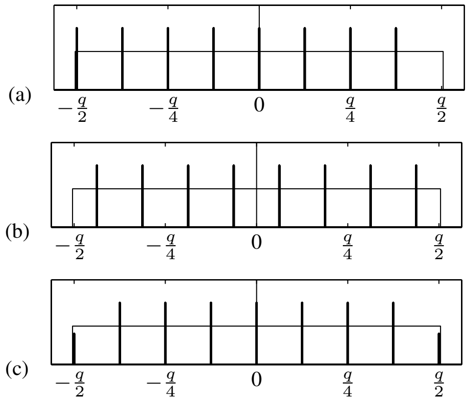

Fig. 3. Discrete uniform dithers with $L = 3$ ($q_\mathrm{d} = q/2^3$): (a) simple (two's complement) binary representation which has mean value $-q_\mathrm{d}/2$; (b) unbiased (shifted) binary representation; (c) unbiased binary representation with half-probability boundary samples (needs $L+1$ bits for coding all the possible values).

In Fig. 3a, the digital dither clearly has a bias of $\mathrm{E}\{d\} = -q_\mathrm{d}/2$. The representation is simple and straightforward. The number of different values is 2^L, with $L = \log_2(q/q_\mathrm{d})$. The variance is $\mathrm{var}\{d\} = \mathrm{var}\{d_\mathrm{c}\} - q_\mathrm{d}^2/12 = q^2/12 - q_\mathrm{d}^2/12$. In $1/2^L$ part of the cases $x+d$ will be equal to $(\mathrm{integer}+0.5)$ LSB (see the remark above section III).

In Fig. 3b, we have removed the bias. The dither can be represented with L bits, keeping in mind that each dither sample has an additional 1 at the bit position $0.5\,\mathrm{LSB}_\mathrm{dither}$.[4]

In Fig. 3c, the digital dither needs $L+1$ bits for representation, since it can have $2^L + 1$ different values. The 0.5 LSB problem arises also here similarly to the case of Fig. 3a.

All three cases behave similarly.

In Fig 4 we have illustrated the behavior of the most important characteristics of the quantization noise of the dither of Fig. 3b. We can observe that even for a few-bit dither, some of the characteristics of the noise are good enough, but the variance still can have large variations: it changes between $[0, q^2/4]$. The cause of the anomaly is that the dither is only zero-order. The CF of the dither in Fig. 4b is:

$$\begin{aligned}\Phi_\mathrm{d}(u) &= \int_{-\infty}^{\infty} \sum_{n=-2^{L-1}+1}^{2^{L-1}} \frac{q_\mathrm{d}}{q}\, \delta\Big(x - (i-0.5)q_\mathrm{d}\Big) e^{jux}\, \mathrm{d}x \\ &= \frac{q_\mathrm{d}}{q} \frac{e^{j\frac{uq}{2}} - e^{-j\frac{uq}{2}}}{e^{j\frac{uq_\mathrm{d}}{2}} - e^{-j\frac{uq_\mathrm{d}}{2}}} \frac{q_\mathrm{d}}{q} \frac{\sin\left(\dfrac{qu}{2}\right)}{\sin\left(\dfrac{q_\mathrm{d}u}{2}\right)} \\ &= \frac{\sin\left(\dfrac{qu}{2}\right)}{2^L \sin\left(2^{-L}\dfrac{qu}{2}\right)}\,. \end{aligned} \tag{10}$$

[4] In this case, for proper quantization we need to round values $x + d = (k + 0.5)q$ upwards. This is even simpler to implement than up or down rounding with probability 0.5-0.5.

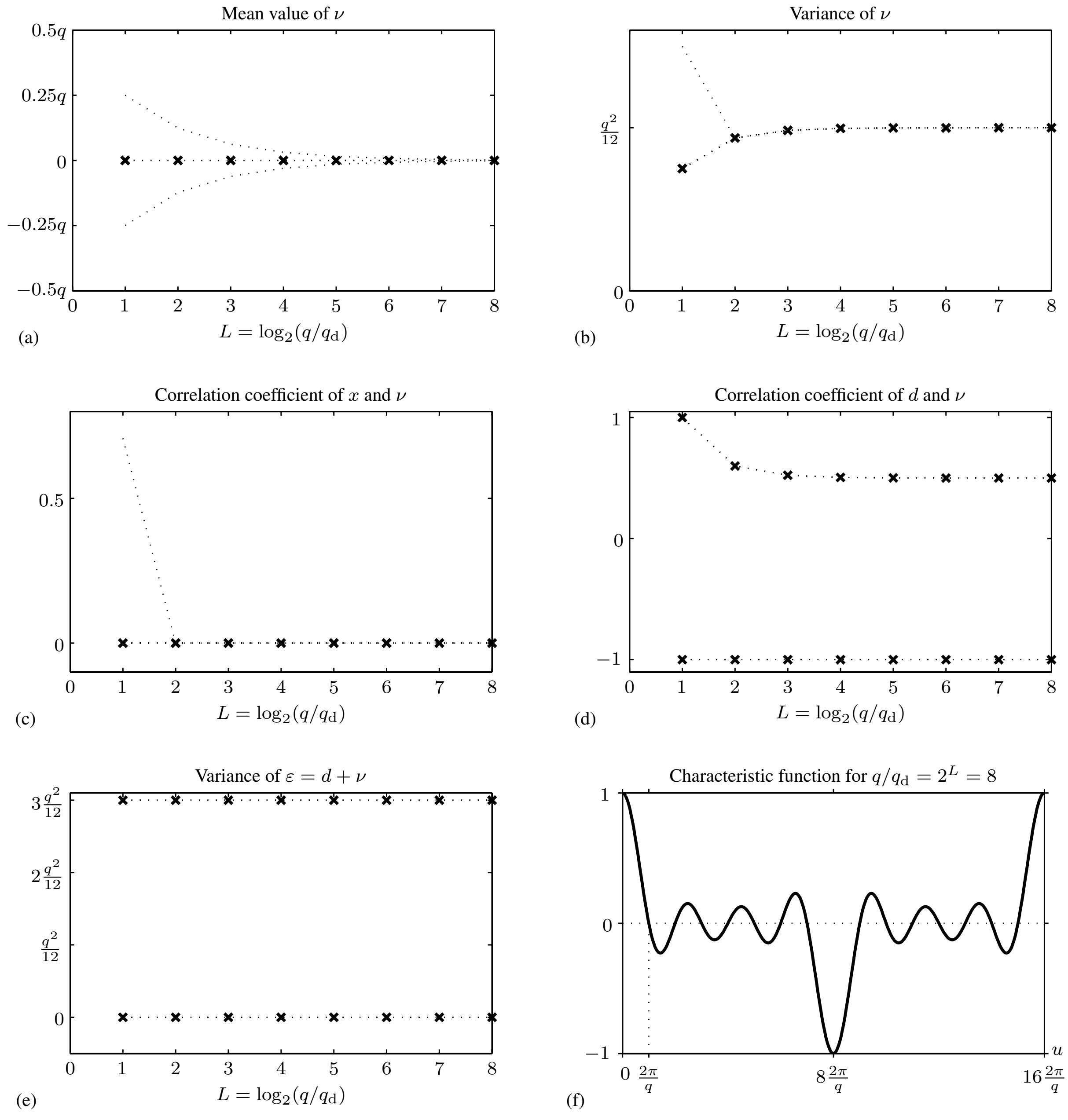

Fig. 4. Quantization noise characteristics for digital uniform dither as in Fig. 3b. Minimum-maximum values for constant x in [0,q] (the **x** marks were calculated from the cases when x is discrete: $k2^{-L}q$, $k = 0, 1, \ldots$; the dotted upper/lower bounds were calculated for continuous-amplitude x values): (a) mean value; (b) variance; (c) correlation coefficient with input x; (d) correlation coefficient with input d; (e) variance of $\varepsilon = d + \nu$; (f) the CF of the dither for $q/q_\mathrm{d} = 2^L = 8$.

The values of the characteristic function of the dither are zero at $l \cdot 2\pi/q$, $l = \pm 1, \pm 2, \cdots$ except when $l = k \cdot 2^L$, $k = \pm 1, \pm 2, \ldots$ Therefore, this dither is only *digitally zero-order* dither. The exceptional peaks (see Fig. 4f) have no influence on the first moment, see Eq. (5). However, the derivatives are not zero, allowing for significant correlation values between d and ν.

The characteristic functions of the other two dithers are similar.

B. Triangularly Distributed Digital Dither

For triangular dither, we have again three almost equivalent forms.

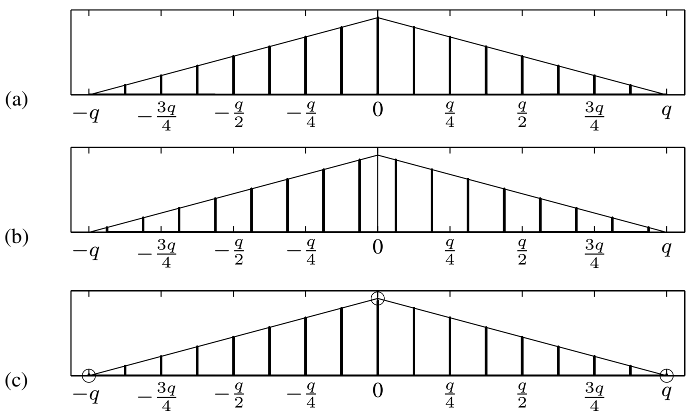

Fig. 5. Discrete triangular dithers with $L = 3$ ($q_\mathrm{d} = q/2^3$): (a) combination of two dithers of Fig. 3b, or mean-corrected combination of two dithers of Fig. 3a; (b) continuous-amplitude triangular dither, quantized with a mid-riser quantizer to resolution q_d; (c) continuous-amplitude triangular dither, quantized with a mid-tread quantizer to resolution q_d.

In Fig. 5a, the digital triangular dither can be obtained by simply adding two dithers of Fig. 3b, or by adding two dithers of Fig. 3a and subtracting the bias $-q_\mathrm{d}$. The representation is simple and straightforward. The number of values is $2 \cdot 2^L - 1$, so the necessary number of bits is $L + 1$. The variance is $\mathrm{var}\{d\} = \mathrm{var}\{d_\mathrm{c}\} - 2q_\mathrm{d}^2/12 = 2q^2/12 - 2q_\mathrm{d}^2/12$ (double of variance of the first digital dither).

This digital dither cannot be obtained by direct quantization of the continuous-time triangular one. A possibility to have this, is illustrated in Fig. 5b. This dither can still be represented with $L+1$ bits ($2 \cdot 2^L$ different values), keeping in mind that each dither sample has an additional 1 at the bit position 0.5 LSB, like in Fig. 3a.

The dither form of Fig. 5c is the result of mid-tread uniform quantization of the continuous-amplitude dither. Mathematically, the distribution can be obtained by correcting the dither shown in Fig. 5a by subtracting a probability $P_1 = (q_\mathrm{d}/2)/q \cdot 1/q \cdot q_\mathrm{d}/2 = q_\mathrm{d}^2/4q^2$ at the center, and executing similar corrections at the edges. The number of amplitude levels is $2 \cdot 2^L + 1$.

Condition (3) is fulfilled for $r = 1$, therefore these dithers are zero-order digital dithers.

In Fig 6 we have illustrated the behavior of the CF of the noise of Fig. 5a. Its expression is:

$$\Phi_\mathrm{d}(u) = \left(\frac{\sin\left(\dfrac{qu}{2}\right)}{2^L \sin\left(2^{-L}\dfrac{qu}{2}\right)} \right)^2 . \tag{11}$$

The characteristic functions of the other two dithers are similar but slightly different.

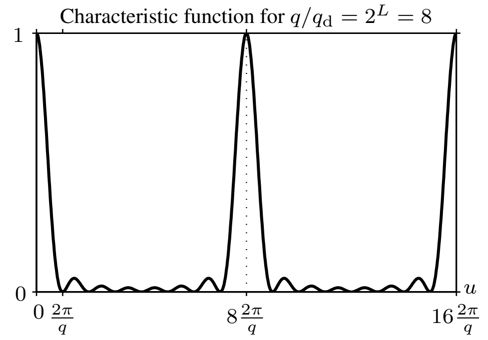

Fig. 6. Characteristic function of digital triangular dither for $2^L = q/q_\mathrm{d} = 8$

The value and the first derivative of the characteristic function of the dither are equal zero at the places required in Eq. (3), so this dither is a *digitally first-order* dither. The peaks shown in the plots have no x-dependent effect on the second moments when the input is digital with $\mathrm{LSB} = q_\mathrm{d}$, as provided by theorem QTDD, therefore they can be corrected for, having knowledge of the dither. This is the dither which can be recommended for digital systems.

V. CONCLUSIONS

Precise condition of unbiasedness with digital dither was given. No approximation is necessary. Eliminating L bits from intermediate results, there is no need at all for dithers with resolution below $2^{-L}q$. Such dithers are exactly realizable in DSP's.

VI. REFERENCES

[1] B. Widrow, "Statistical analysis of amplitude-quantized sampled-data systems," *Trans. AIEE, Part II.: Applications and Industry*, vol. 79, no. 52, pp. 555–68, Jan'61 Section 1961.

[2] I. Kollár, "Quantization noise," Doct. Sci. Thesis, Hungarian Academy of Sciences, Budapest, 1997.

[3] B. Widrow and I. Kollár, *Quantization Noise*, in preparation (2006), `http://www.mit.bme.hu/books/quantization/`.

[4] B. Widrow, I. Kollár, and M.-C. Liu, "Statistical theory of quantization," *IEEE Trans. on Instrumentation and Measurement*, vol. 45, no. 6, pp. 353–61, 1995.

[5] G. L. Roberts, "Picture coding using pseudo-random noise," *IRE Trans. on Information Theory*, vol. IT-8, no. 2, pp. 145–54, Feb 1962.

[6] L. Schuchman, "Dither signals and their effect on quantization noise," *IEEE Trans. on Communication Theory*, vol. 12, pp. 162–65, Dec 1964.

[7] R. A. Wannamaker, S. P. Lipshitz, J. Vanderkooy, and J. Nelson Wright, "A mathematical theory of non-subtractive dither," *IEEE Trans. on Signal Processing*, vol. 48, no. 2, pp. 499–516, Feb. 2000.

[8] K. Godfrey, Ed., *Perturbation Signals for System Identification*, Prentice-Hall, Englewood Cliffs, NJ, 1993.