TECHNICAL RESEARCH REPORT

Algorithm-Based Low-Power Transform Coding Architectures - Part II: Logarithmic Complexity, Unified Architecture, and Finite-Precision Analysis

by A-Y. Wu and K.J.R. Liu

T.R. 95-33

INSTITUTE FOR SYSTEMS RESEARCH

Sponsored by
the National Science Foundation
Engineering Research Center Program,
the University of Maryland,
Harvard University,
and Industry

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **1995**	2. REPORT TYPE	3. DATES COVERED **00-00-1995 to 00-00-1995**
4. TITLE AND SUBTITLE **Algorithm-Based Low-Power Transform Coding Architectures- Part II: Logarithmic Complexity, Unified Architecture, and Finite-Precision Analysis**		5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Department of Electrical Engineering,Institute for Systems Research,University of Maryland,College Park,MD,20742**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**		
13. SUPPLEMENTARY NOTES		
14. ABSTRACT **see report**		
15. SUBJECT TERMS		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**33**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

Algorithm-Based Low-Power Transform Coding Architectures-Part II: Logarithmic Complexity, Unified Architecture, and Finite-Precision Analysis

An-Yeu Wu and K. J. Ray Liu
Electrical Engineering Department and Institute for Systems Research
University of Maryland
College Park, MD 20742
Phone: (301) 405-6619, Fax: (301) 405-6707

ABSTRACT

In the companion paper, we addressed the low-power DCT/IDCT VLSI architectures of linear complexity increase based on the multirate approach. In this paper, we will discuss other aspects of the low-power design. Firstly, we consider the design of low-power architectures that can lower the power consumption at only $O(\log M)$ increase in hardware complexity. Next, we will extend the low-power DCT design to other orthogonal transforms such as Modulated Lapped Transform (MLT) and Extended Lapped Transform (ELT). A unified programmable IIR low-power transform module, which can perform most of the existing discrete sinusoidal transforms, is also proposed. Finally, we perform the finite-precision analysis of the DCT architecture under the normal and multirate operations. In VLSI design, the assignment of the system wordlength will directly affect the total switching events and routing capacities, hence the power consumption. Using the analytical results, we can choose the optimal wordlength for each DCT channel under required signal-to-noise ratio (SNR) constraint. The material presented in this paper, together with the multirate architectures in the companion paper, provides a framework for the algorithm-based low-power transform coding kernel design.

This work was supported in part by the ONR grant N00014-93-10566 and the NSF NYI Award MIP9457397.

1 Introduction

In the companion paper [1], we introduced the algorithm-based low-power design based on the multirate approach. Specifically, we showed that the power consumption can be reduced provided that we can perform the DCT/IDCT from the decimated-by-M input sequences at $O(M)$ increase in hardware complexity. In practice, the $O(M)$ overhead may not be desirable when M is large and total chip area is limited. Therefore, the search for compensation scheme with less hardware overhead is desired. In this paper, we will show a scheme to perform the polyphase decomposition in such a way that only $O(\log M)$ overhead is required to compensate the speed penalty. The resulting structure reduces the operating frequency on a stage-by-stage base: In each stage, the operating frequency is reduced by half. After reaching to the $(\log M)^{th}$ stage, we can operate at M-times slower clock rate of the original data rate. We shall refer to this as *logarithmic low-power architecture.* This multiple operation frequency environment allows us to perform different speed compensation at each stage; *i.e.*, different low supply voltages can be used to lower the power consumption. In general, the power savings of the logarithmic architecture is between the normal IIR architecture [2] and the full multirate architecture presented in the companion paper [1].

Next we extend the low-power design presented in the companion paper to a larger class of orthogonal transforms. We start with the low-power design of the Modulated Lapped Transform (MLT) and Extended Lapped Transform (ELT). The MLT and ELT, which belong to the family of Lapped Orthogonal Transforms (LOT), are very attractive in the applications of transform coding since they can diminish the blocking effect encountered in low bit-rate block transforms [3][4][5]. Recently, Frantzeskakis *et al.* [6][7] proposed the time-recursive MLT and ELT architectures that are suitable for VLSI implementation due to their modularity and regularity. However, since the updating of the MLT and ELT coefficients should be as fast as the input data rate, those architectures cannot compensate the speed penalty under low supply voltage. In this paper, we will derive the low-power time-recursive MLT and ELT structures. By applying the polyphase decomposition to their IIR transfer functions, the MLT/ELT coefficients can be updated at M-times slower rate with linear hardware overhead; hence, the low-power operation is allowed. Later, based on the derivations of the MLT and ELT, we propose a unified low-power IIR structure which can be implemented as a programmable DSP co-processor to perform most of the existing sinusoidal transforms.

In the last part of this paper, we will consider the finite-wordlength effect of the proposed low-

power DCT architectures. The effect of wordlength on the power consumption was discussed in [8]. In summary, shorter wordlengths will result in fewer switching events, lower capacitance, and shorter average routing length in the system. As a result, low power consumption of the chip can be achieved. On the other hand, if the wordlengths are too short, the rounding error caused by finite-precision operations can be severe enough to hazard the signal-to-noise ratio (SNR). Thus, choosing minimum wordlengths without degrading the SNR requirement is an important issue in the low-power VLSI design. Motivated by this, we perform the finite-precision analysis for the IIR DCT structure and its low-power design. Our study can precisely predict the finite-precision behavior under different block sizes and decimation factors. Using these analytical results, we can assign the optimal wordlength for each DCT channel given the SNR constraint. Moreover, our analyses show that the average SNR's of the proposed low-power architectures are better than that of the normal design given the same wordlength assignment. This indicates the multirate design has better numerical properties under fix-point arithmetic.

The organization of the this paper is as follows: Section 2 presents the low-power DCT architecture of logarithmic complexity. In section 3, we derive the multirate MLT and ELT algorithms and architectures. Then, a unified low-power IIR structure for most sinusoidal transforms are described. The fixed-point analysis is presented in Section 4 followed by a conclusion.

2 Low-Power Architecture of Logarithmic Complexity

In the companion paper, we have shown how to perform the DCT/IDCT from the decimated-by-M input sequences so that the speed penalty under low-power operation can be compensated at the algorithmic/architectural level. The advantage of this design is obtained by applying the polyphase decomposition to the IIR transfer function until the resulting transfer function is fully expanded with all exponents being multiples of M. However, such manipulation requires $O(M)$ overhead in hardware, which may not be acceptable when M is large and the chip area is limited. In this section, we will show how to achieve low-power consumption with only logarithmic complexity overhead. The basic priciple is to repeat the polyphase decomposition in a certain way instead of fully expanding them. By doing so, the lower-rate operations can be obtained while the complexity will grow slower. The price paid is that the resulting architecture will be operated at multiple low frquencies rather than at the uniform low frequency as discussed in [1]. Nevertheless, the multiple frequency environment

enables us to perform different speed compensations at different stages of the design. Therefore, different low supply voltages can be applied according to the given speed constraint, and the total power consumption can be still reduced. In what follows, we will derive the logarithmic low-power DCT architecture. The results can be extended to other low-power transformation designs to be discussed in Section 3.

2.1 Low-Power DCT Architecture of Logarithmic Complexity

The multirate IIR DCT transfer function with $M = 2$ can be written as [1]

$$H_{DCT,k}(z) = \frac{(-1)^k C(k)}{D(z^2)} \left[H_0(z^2) + z^{-1} H_1(z^2) \right] \tag{1}$$

where $C(k)$ is the scaling factor of the DCT and

$$\begin{aligned} D(z^2) &= 1 - 2\cos 4\omega_k z^{-2} + z^{-4}, \\ H_0(z^2) &= (\cos\omega_k - \cos 3\omega_k z^{-2}), \\ H_1(z^2) &= (\cos 3\omega_k - \cos\omega_k z^{-2}). \end{aligned} \tag{2}$$

Substituting the polyphase decomposition

$$\frac{1}{D(z^2)} = \frac{H_0'(z^4) + z^{-2} H_1'(z^4)}{D'(z^4)} \tag{3}$$

with

$$\begin{aligned} D'(z^4) &= 1 - 2\cos 8\omega_k z^{-4} + z^{-8}, \\ H_0'(z^4) &= 1 + z^{-4}, \\ H_1'(z^4) &= 2\cos 4\omega_k, \end{aligned} \tag{4}$$

into (1) and rearranging, we can rewrite $H_{DCT,k}(z)$ so that the DCT can be computed at four times slower clock rate [1]. Nevertheless, this multirate design requires $O(M)$ hardware overhead to directly lower the input clock rate by four. In order to save the hardware complexity, we may rewirte (1) in

a cascade form after the substitution is made, *i.e.*,

$$H_{DCT,k}(z) = (-1)^k C(k) \left[H_0(z^2) + z^{-1}H_1(z^2)\right] \left[H_0'(z^4) + z^{-2}H_1'(z^4)\right] \cdot \frac{1}{D'(z^4)}. \tag{5}$$

Fig.1(a) shows the polyphase implementation of (5), which leads to the cascade multirate DCT architecture depicted in Fig.1(b). There are two major blocks. One operates at 50% sample rate and the other at 25% sample rate. Due to the special form of the denominator of the transfer function, we can repeatedly perform the polyphase decomposition on the denominator and retain the cascade form. We then have

$$H_{DCT,k}(z) = (-1)^k C(k) \left[H_0(z^2) + H_1(z^2)\right] \frac{\prod_{i=1}^{\log M - 1} \left[(1 + z^{-2^{i+1}}) + 2z^{-2^i}\cos(2^{i+1}\omega_k)\right]}{1 - 2\cos(2M\omega_k)z^{-M} + z^{-2M}} \tag{6}$$

for any M, $M \in 2^{Z^+}$. The resulting architecture decimates the operating frequency on a stage-by-stage base: In each stage, the operating frequency is reduced by half. After reaching the $(\log M)^{th}$ stage, we will have M times slower clock rate of the original data rate.

2.2 Power Consumption

When low-power implementation is taken into consideration, the feature of multiple operating frequencies in the above architecture implies that different supply voltages will be used according to the slowest allowable operating speed. That is, the operators to realize $H_0(z^2)$ and $H_1(z^2)$ in (5) can be operated at $3.1V$ due to the two times slower clock rate, while all other operators to realize $H_0'(z^4)$ and $H_1'(z^4)$ can be operated at $2.1V$ due to the four times slower clock rate [9][1]. As a consequence, the power consumption of the 16-point low-power DCT architecture in Fig.1(b) can be estimated as

$$(\frac{N_2}{N_0}C_{eff})(\frac{3.1V}{5V})^2(\frac{1}{2}f) + (\frac{N_4}{N_0}C_{eff})(\frac{2.1V}{5V})^2(\frac{1}{4}f) \approx 0.24P_0, \tag{7}$$

where $N_0 = 30$ is the total multipliers used in the normal DCT ($M = 0$); $N_2 = 30$ and $N_4 = 30$ are the number of multipliers in the $M = 2$ stage and $M = 4$ stage, respectively. From (7), we can see that the overall power consumption of the logarithmic low-power design will be in between $M = 2$ and $M = 4$ of the full multirate DCT systems discussed in [1].

On the other hand, by examing (6), we can see that in order to have M-times slower operating

frequency at the final stage, we need a total of $(\log M + 2)$ multipliers to realize the multirate transfer function. The comparison of the logarithmic low-power architecture with other approaches is listed in Table 1. Although the total power savings of the logarithmic structure is less than that of the full multirate structure given the same decimation factor M, the $O(\log M)$ hardware overhead is preferable when we want to achieve low-power consumption without trading too much chip area.

The multiple-frequency feature of the cascade low-power architecture also allows us to achieve more power and area savings at the arithmetic level. For example, we can use look-ahead adders in the $M = 2$ region to match the data throughput rate, whereas we can employ low-speed carry-ripple adders in the $M = 4$ region due to the much relaxed speed constraint.

3 Unified Low-Power Module Design

3.1 The IIR MLT Algorithm

The MLT operates on segments of data of length $2N$, $x(t+n-2N+1), n = 0, 1, \cdots, 2N-1$, and produces N output coefficients, $X_{MLT,k}(t)$, $k = 0, 1, \cdots, N-1$, as follows [4]:

$$X_{MLT,k}(t) = S(k)\sqrt{\frac{2}{N}}\sum_{n=0}^{2N-1} \sin\frac{\pi}{2N}(n+\frac{1}{2})\cos[\frac{\pi}{N}(k+\frac{1}{2})(n+\frac{1}{2}+\frac{N}{2})]\,x(t+n-2N+1) \tag{8}$$

where $S(k) = (-1)^{(k+2)/2}$ if k is even, and $S(k) = (-1)^{(k-1)/2}$ if k is odd. After some algebraic manipulations, the MLT can be decomposed into [7]

$$X_{MLT,k}(t) = -S(k)[\,X_{C,k+1}(t) + X_{S,k}(t)\,], \tag{9}$$

where

$$X_{C,k}(t) \triangleq \beta_1 \sum_{n=0}^{L-1} \cos[(2n+1)\omega_k + \theta_k]x(t+n-2N+1), \tag{10}$$

$$X_{S,k}(t) \triangleq \beta_1 \sum_{n=0}^{L-1} \sin[(2n+1)\omega_k + \theta_k]x(t+n-2N+1), \tag{11}$$

with block size $L = 2N$ and

$$\beta_1 \triangleq \frac{1}{\sqrt{2N}}, \quad \omega_k \triangleq \frac{\pi k}{2N}, \quad \text{and} \quad \theta_k \triangleq \frac{\pi}{2}(k+\frac{1}{2}). \tag{12}$$

The IIR transfer functions for (10) and (11) can be computed as

$$H_{C,k}(z) = \beta_1(1-z^{-L})\frac{\cos((2L-1)\omega_k+\theta_k)-\cos((2L+1)\omega_k+\theta_k)z^{-1}}{1-2\cos 2\omega_k z^{-1}+z^{-2}}, \tag{13}$$

$$H_{S,k}(z) = \beta_1(1-z^{-L})\frac{\sin((2L-1)\omega_k+\theta_k)-\sin((2L+1)\omega_k+\theta_k)z^{-1}}{1-2\cos 2\omega_k z^{-1}+z^{-2}}. \tag{14}$$

The corresponding IIR module for the dual generation of $X_{C,k}(t)$ and $X_{S,k}(t)$ is depicted in Fig.2, where

$$\Gamma_1 \triangleq \beta_1\cos((2L-1)\omega_k+\theta_k), \qquad \Gamma_2 \triangleq -\beta_1\cos((2L+1)\omega_k+\theta_k),$$
$$\Gamma_3 \triangleq \beta_1\sin((2L-1)\omega_k+\theta_k), \qquad \Gamma_4 \triangleq -\beta_1\sin((2L+1)\omega_k+\theta_k). \tag{15}$$

This IIR module can be used as a basic building block to implement MLT according to (9). Fig.3 illustrates the overall time-recursive MLT architecture for the case $N = 8$. It consists of two parts: One is the *IIR module array* which computes $X_{C,k}(t)$ and $X_{S,k}(t)$ with different index k in parallel. The other is the *combination circuit* which selects and combines the outputs of the IIR array to generate the MLT coefficients.

3.2 Low-Power Design of the MLT

As with the low-power DCT, we can have a low-power MLT architecture if each MLT module can compute $X_{C,k}(t)$ and $X_{S,k}(t)$ using the decimated input sequences. After performing the polyphase decomposition on (13) and (14), we can compute the multirate IIR transfer functions for $H_{C,k}(z)$ and $H_{S,k}(z)$ as

$$\begin{aligned} H_{C,k}(z) &= \frac{\beta_1(1-z^{-L/2})}{1-2\cos(4\omega_k)z^{-1}+z^{-2}}\Big([\cos((2L-3)\omega_k+\theta_k)-\cos((2L+1)\omega_k+\theta_k)z^{-1}]X_e(z) \\ &+ [\cos((2L-1)\omega_k+\theta_k)-\cos((2L+3)\omega_k+\theta_k)z^{-1}]X_o(z)\Big), \end{aligned} \tag{16}$$

and

$$\begin{aligned} H_{S,k}(z) &= \frac{\beta_1(1-z^{-L/2})}{1-2\cos(4\omega_k)z^{-1}+z^{-2}}\Big([\sin((2L-3)\omega_k+\theta_k)-\sin((2L+1)\omega_k+\theta_k)z^{-1}]X_e(z) \\ &+ [\sin((2L-1)\omega_k+\theta_k)-\sin((2L+3)\omega_k+\theta_k)z^{-1}]X_o(z)\Big). \end{aligned} \tag{17}$$

The parallel architecture for (16) and (17) is shown in Fig.4, where

$$\begin{aligned}
\Gamma_{1,e} &= \beta_1 \cos((2L-3)\omega_k+\theta_k), & \Gamma_{2,e} &= -\beta_1 \cos((2L+1)\omega_k+\theta_k),\\
\Gamma_{3,e} &= \beta_1 \sin((2L-3)\omega_k+\theta_k), & \Gamma_{4,e} &= -\beta_1 \sin((2L+1)\omega_k+\theta_k),\\
\Gamma_{1,o} &= \beta_1 \cos((2L-1)\omega_k+\theta_k), & \Gamma_{2,o} &= -\beta_1 \cos((2L+3)\omega_k+\theta_k),\\
\Gamma_{3,o} &= \beta_1 \sin((2L-1)\omega_k+\theta_k), & \Gamma_{4,o} &= -\beta_1 \sin((2L+3)\omega_k+\theta_k).
\end{aligned} \tag{18}$$

It consists of two MLT modules in Fig.2. The upper module computes part of the $X_{C,k}(t)$ and $X_{S,k}(t)$ from the even sequence, while the lower one computes the remaining part from the odd sequence. The two adders at the right end are used to combine the even and odd outputs. Through such manipulation, only decimated sequences are processed inside the module. Hence, the MLT module can operate at the half of the original frequency by doubling the hardware complexity. The comparison of hardware cost is shown in Table 2. Suppose that P_0 denotes the power consumption of the MLT module in Fig.2. From the CMOS power model, it can be shown that the power consumption for the low-power MLT modules are $0.38P_0$ and $0.17P_0$ for the case $M = 2$ and $M = 4$, respectively. Basically, this savings is obtained at the expense of linear increase in hardware.

3.3 Low-Power Design of the ELT

The ELT with basis length equal to $4N$ operates on data segment of length $4N$, $x(t+n-4N+1), n = 0, 1, \cdots, 4N-1$, and produces N output coefficients, $X_{ELT,k}(t), k = 0, 1, \cdots, N-1$. One good choice for the ELT is as follows [10][5]:

$$X_{ELT,k}(t) = \sqrt{\frac{2}{N}} \sum_{n=0}^{4N-1} [\frac{1}{2\sqrt{2}} - \frac{1}{2} \cos \frac{\pi}{N}(n+\frac{1}{2})] \cos[\frac{\pi}{N}(k+\frac{1}{2})(n+\frac{1}{2}+\frac{N}{2})]\, x(t+n-4N+1) \tag{19}$$

By the use of some trigonometric identities, we can rewrite (19) as

$$X_{ELT,k}(t) = -\tilde{X}_{S,k+1}(t) + \sqrt{2}\tilde{X}_{C,k}(t) + \tilde{X}_{S,k-1}(t), \tag{20}$$

where

$$\tilde{X}_{C,k}(t) \triangleq \beta_2 \sum_{n=0}^{L-1} \cos[(2n+1)\omega'_k + \theta'_k]\, x(t+n-4N+1), \tag{21}$$

$$\tilde{X}_{S,k}(t) \triangleq \beta_2 \sum_{n=0}^{L-1} \sin[(2n+1)\omega'_k + \theta'_k]\, x(t+n-4N+1), \tag{22}$$

with

$$L = 4N, \quad \beta_2 \triangleq \frac{1}{2\sqrt{2N}}, \quad \omega'_k \triangleq \frac{\pi}{2N}(k+\frac{1}{2}), \text{ and } \theta'_k \triangleq \frac{\pi}{2}(k+\frac{1}{2}). \tag{23}$$

Define the relationship in (9) and (20) as the *combination functions.* After comparing (9)-(12) with (20)-(23), we see that the MLT and ELT have identical mathematical structures except for the definitions of parameters and the combination functions. Therefore, the IIR MLT module in Fig.2, as well as the low-power MLT module in Fig.4, can be readily applied to ELT by simply modifying those multiplier coefficients. Also, the overall ELT architecture is similar to the MLT architecture in Fig.3 except that the combination circuit performs according to (20).

Moreover, it can be verified that $X_{S,-1}(t) = -X_{S,0}(t)$ and $X_{S,N}(t) = X_{S,N-1}(t)$. Hence, we can compute the 0th and $(N-1)^{th}$ ELT coefficients from

$$\begin{aligned} X_{ELT,0}(t) &= -X_{S,1}(t) + \sqrt{2}X_{C,0}(t) - X_{S,-1}(t), \\ X_{ELT,N-1}(t) &= -X_{S,N-1}(t) + \sqrt{2}X_{C,N-1}(t) + X_{S,N-2}(t), \end{aligned} \tag{24}$$

instead of implementing two extra ELT modules for $X_{S,-1}(t)$ and $X_{S,N}(t)$. The hardware cost for the ELT can be found in Table 2. Since the number of multipliers of the ELT is about the same as that of the MLT, the power savings for both transforms are similar.

3.4 Unified Low-Power IIR Transform Module Design

From the transform functions described in (9)-(12) and (20)-(23), we observe that the low-power MLT module in Fig.4 can be used to realize most existing discrete sinusoidal transforms by suitably setting the parameters and defining the combination functions. For example, $X_{C,k}(t)$ in (10) is equivalent to the DCT by setting

$$L = N, \quad \beta_1 = C(k), \quad \omega_k = \frac{k\pi}{2N}, \text{ and } \theta_k = 0. \tag{25}$$

As a result, the multirate MLT module in Fig.4 can compute the DCT with different index k in parallel.

The other example is the discrete Fourier transform (DFT) with real-valued inputs. With the following parameter setting

$$L = N, \quad \beta_1 = \frac{1}{\sqrt{N}}, \quad \omega_k = \frac{-k\pi}{N}, \text{ and } \theta_k = -\omega_k, \tag{26}$$

(10) and (11) become

$$X_{C,k}(t) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \cos(\frac{-2\pi}{N} kn)\, x(t+n-N+1), \tag{27}$$

$$X_{S,k}(t) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \sin(\frac{-2\pi}{N} kn)\, x(t+n-N+1), \tag{28}$$

which are the real part and the imaginary part of the DFT, respectively. The discrete Hartley transform (DHT) can be computed using the same parameter setting as the DFT except that the combination circuit in Fig.3 performs as

$$X_{DHT,k}(t) = X_{C,k}(t) + X_{S,k}(t). \tag{29}$$

The parameter settings as well as the corresponding combination functions for other orthogonal transforms are summarized in Table 3.

The programmable feature of the unified low-power module design makes it very attractive in transform coding applications. Firstly, the unified structure can be implemented as a high-performance programmable co-processor which performs various transforms for the host processor by loading the suitable parameters. Secondly, by hard-wiring the multiplier coefficients of the modules to preset values according to the transformation type, we can perform any one of the discrete sinusoidal transforms using the same architecture. This can significantly reduce the design cycle as well as the manufacturing cost.

3.5 Extension to Low-Power 2-D Transforms

Extension of our low-power transform algorithms to low-power two-dimensional (2-D) transforms can be achieved by employing the time-recursive 2-D DCT architecture proposed by Chiu and Liu [11]. In general, the architecture in [11] can be applied to all transformations with SIPO property. Therefore, we can apply it to implement the low-power 2-D transforms with some minor modifications.

4 Finite-Precision Analysis of The IIR DCT Architecture

In low-power VLSI implementation, the choice of wordlength is an important issue since it will directly affect the total switching activities inside the opeartors as well as the total effective capacitance. Besides, an underestimated wordlength will degrade the system performance due to the increased rounding errors. Therefore, we should carefully determine the minimum allowable system wordlength that meets the accuracy criteria for cost-effective implemention. In this section, we will consider the finite-precision effects of the proposed low-power DCT architectures. The results can be easily extended to other transform architectures. We will start with the DCT architecture under the normal operation, then the analysis is extended to the low-power design with $M = 2$. The general results for arbitrary M is also presented. Throughout the derivations, the "statistical error model" for fixed-point analysis is used [12, chap.6]:

1. The rounding error is treated as wide-stationary additive white noise with magnitude uniformly distributed over one quantization level.

2. Rounding error occurs only in multiplication.

3. All errors are uncorrelated with the input signal, and are independent of each other.

4.1 Basic Considerations in Finite-Precision Analysis

There are two basic considerations in the fixed-point analysis. One is the *rounding error* behavior. It occurs when we multiply two $(B+1)$-bit numbers together while only $(B+1)$-bit product is kept. The mean and variance of the rounding error are given by [12, chap.6]

$$m_R = 0, \qquad \sigma_R^2 = \frac{2^{-2B}}{12}, \tag{30}$$

respectively. Understanding the roudning error behavior will allow us to minimize the wordlength to achieve a desired output signal-to-noise ratio.

The other is the *dynamic range* issue. In fixed-point implementation, each number in the system is treated as a fraction. The magnitude of each node in the circuit cannot exceed one, otherwise overflow occurs and will result in great distortion in the final output. Therefore, to prevent overflow, a suitable scaling of the input signal is usually employed according to the dynamic range of the

system. In practice, the signal-to-noise ratio of the scaled system, SNR', will be degraded by the scaling process and is given by [12, chap.6]

$$SNR' = s^2 SNR_0, \tag{31}$$

where s is the scaling factor and SNR_0 is the signal-to-noise ratio of the original system. This implies that knowing the dynamic range will enable us to perform minimally necessary scaling to prevent further degradation in SNR.

4.2 IIR DCT Using Direct Form I Structure

4.2.1 Rounding Errors

Using the statistical error model, the rounding error of the IIR DCT structure can be modeled as (see Fig.5)

$$e(t) = e_1(t) + e_2(t) \tag{32}$$

where $e_i(t)$, $i = 1, 2$ is the rounding error caused by the ith multiplier in the circuit [1]. Then the actual output of the DCT circuit after N iterations can be represented as

$$\hat{X}_{DCT,k}(t) = X_{DCT,k}(t) + f(t) \tag{33}$$

where $f(t)$ is the output error due to the noise error $e(t)$.

Let $H_{ef}(z)$ denote the transfer function of the system from the node at which $e(t)$ is injected to the output, and $h_{ef}(n)$ be the corresponding unit-sample response. From Fig.5, $H_{ef}(z)$ is given by

$$H_{ef}(z) = \frac{1}{1 - 2\cos 2\omega_k z^{-1} + z^{-2}}, \tag{34}$$

and $h_{ef}(n)$ can be derived as

$$h_{ef}(n) = \frac{1}{\sin(2\omega_k)} \sin[(n+1)2\omega_k]u(n) \tag{35}$$

[1]Since the 0th channel of the DCT is computed by a simple add-and-accumulate operation, we will not consider the finite-wordlength effect of this channel.

where $u(n)$ denotes the step function. Since only N iterations are performed in the IIR circuit, the mean and variance of $f(t)$ of the kth DCT channel can be computed as

$$m_f = m_e \sum_{n=0}^{N-1} h_{ef}(n) = m_e \sum_{n=0}^{N-1} \frac{1}{\sin(2\omega_k)} \sin[(n+1)2\omega_k], \tag{36}$$

$$\sigma_f^2 = \sigma_e^2 \sum_{n=0}^{N-1} |h_{ef}(n)|^2 = \frac{\sigma_e^2}{\sin^2(2\omega_k)} \sum_{n=0}^{N-1} \sin^2[(n+1)2\omega_k] = \frac{\sigma_e^2}{\sin^2(2\omega_k)} \left(\frac{N}{2}\right), \tag{37}$$

where

$$m_e = E\{e(t)\} = 0, \tag{38}$$

$$\sigma_e^2 = E\{e^2(t)\} = E\{(e_1(t))^2\} + E\{(e_2(t))^2\} = (1 + N_s(k)) \cdot \sigma_R^2, \tag{39}$$

and $N_s(k)$ is the number of the noise sources contributed by the multiplier $M_2 = 2\cos(2\omega_k)$ in the IIR loop:

$$N_s(k) = \begin{cases} 4, & \text{if } |2\cos(2\omega_k)| > 1, \\ 1, & \text{if } |2\cos(2\omega_k)| < 1, \\ 0, & \text{if } |2\cos(2\omega_k)| = 1. \end{cases} \tag{40}$$

When $|2\cos(2\omega_k)| < 1$, a normal multiplication is performed and $E\{(e_2(t))^2\} = \sigma_R^2$. In the case of $|2\cos(2\omega_k)| > 1$, since a left-shift is performed after the multiplication with $\cos(2\omega_k)$, the rounding error is amplified by 2 and its power becomes $E\{(2e_2(t))^2\} = 4 \cdot \sigma_R^2$. In the case of $|2\cos(2\omega_k)| = 1$, no mulplication is performed, hence $E\{(e_2(t))^2\} = 0$. Now using (36)-(40), we can represent the total noise power at the kth DCT channel as

$$P_f = m_f^2 + \sigma_f^2 = \frac{N(N_s(k)+1)}{2\sin^2(2\omega_k)} \left(\frac{2^{-2B}}{12}\right). \tag{41}$$

As we can see, given the system wordlength B, the rounding error grows linearly with the block size N. This indicates that we will have 3 dB degradation in the SNR as N doubles; however, such degradation can be compensated by adding 1/2 (in average) bit in the wordlength. On the other hand, the noise power is inversely proportional to $\sin^2(2\omega_k)$. That is, the effect of the rounding error in each channel of the IIR DCT greatly depends on the pole locations of the IIR transfer function. The closer $2\omega_k$ is to 0 or π, the larger the rounding error is. As a consequence, the 1st and $(N-1)^{th}$ DCT channels suffer most from the finite-wordlength effect, while the middle channels have good

SNR in terms of rounding error. This phenomenon is quite different from what we have seen in other DCT algorithms (*cf*, Fig.7 in [13]).

4.2.2 Dynamic Range

In fixed-point arithmetic, the input sequence $x(t)$ is represented as a fraction and is bounded by $|x(t)| \leq 1$. Hence, the dynamic range of the circled nodes in Fig.6 can be computed as

$$D_1 = 2, \tag{42}$$

$$\begin{aligned} D_2 &= \max\{X_{DCT,k}(t)\} = \max\{C(k)\sum_{n=0}^{N-1}\cos[(2n+1)\omega_k]x(n)\} \\ &= C(k)\sum_{n=0}^{N-1}|\cos[(2n+1)\omega_k]|\cdot\max\{x(n)\} = C(k)\sum_{n=0}^{N-1}|\cos[(2n+1)\omega_k]|, \end{aligned} \tag{43}$$

and the dynamic range of the overall architecture is given by

$$D = \max\{D_1, D_2\}. \tag{44}$$

Suppose that a one-time scaling scheme is provided at the input end to avoid overflow, and it is done by shifting the data to the right by K bits. We have

$$K = \lceil \log_2 D \rceil, \tag{45}$$

and the scaling factor s is given by

$$s = \frac{1}{2^K}. \tag{46}$$

4.2.3 Optimal Wordlength Assignment

Assume that the input sequence $x(t)$ is uniformly distributed over $(-1, 1)$ with zero mean. From (31), (41), and (46), we have

$$SNR' = s^2\frac{E\{(X_{DCT,k}(t))^2\}}{P_f} = \frac{8\sin^2(2\omega_k)}{N(N_s(k)+1)}\cdot 2^{2B-2K} \tag{47}$$

where the fact that [13]

$$E\{(X_{DCT,k}(t))^2\} = E\{x^2(t)\} = 1/3, \quad k = 1, 2, \ldots, N-1, \tag{48}$$

is used. If we want to achieve a performance of 40 dB in SNR for the kth DCT component, the optimal wordlength B_k for that channel can be computed from (47) as

$$B_k = \left\lceil \frac{4 - \log_{10}[\sin^2(2\omega_k) \cdot \frac{8}{N(N_s(k)+1)}]}{2 \cdot \log_{10} 2} + K \right\rceil. \tag{49}$$

As an example, the B_k's for the case $N = 8$ and 16 under the constraint SNR = 40 dB are listed in Table 4(a), where B_A denotes the average system wordlength. As we can see, $B_A = 12$ bit is sufficient to meet the accuracy criteria. Compared with the DCT implemenation in [14], in which B_A was chosen to be 16 bit based on the experimental simulation results, our system wordlength is much shorter. Suppose that the silicon area of the multiplier is dominant in the chip and the size of the multipliers is proportional to $(B_A)^2$. Using the optimal wordlengths in Table 4, we can reduce the total chip area to 56% of the original design without degrading the SNR performance. This shows that our analysis approach provides more insights to determine the architectural specifications than the experimental approach. Moreover, in the applications of transform coding, we can shorten the wordlengths for the high-frequency channels since the human vision system is less sensitive to these components. Thus, the total wordlength can be further reduced.

4.3 IIR DCT Using Direct Form II Structure

Given the IIR DCT transfer function, we can also implement it using the direct form II structure as shown in Fig.7. Following the above derivations for the direct form I structure, the fixed-point analytical results can be derived as:

1. Rounding error:

$$P_f = (N_s(k) + 1)\frac{2^{-2B}}{12}. \tag{50}$$

2. The dynamic range:

$$D_1 = \frac{1}{|\sin(2\omega_k)|} \sum_{n=0}^{N-1} |\sin[(2n+1)\omega_k]|,$$

$$\begin{aligned} D_2 &= 2, \\ D &= \max\{D_1, D_2\}. \end{aligned} \tag{51}$$

In contrast to the direct form I structure, the dynamic range of the direct form II structure is affected by the factor $\frac{1}{\sin(2\omega_k)}$ in D_1; that is, we will have non-uniform dynamic ranges for different DCT channels. This feature is not desirable in real implementations even though the SNR results of both structures are comparative to each other (see simulation results in Section 4.5)–It not only requires different scaling scheme in each DCT channel, but also makes the data interface between VLSI modules complicated (*e.g.* 2-D DCT in which two DCT modules are connected.). Therefore, the direct form I is a better choice for the VLSI implementation of the IIR DCT structures.

4.4 Analysis for the Low-Power IIR DCT with $M = 2$

In the low-power IIR DCT architecture with $M = 2$, the injected rounding error can be modeled as (see Fig.8)

$$e(t) = e_1(t) + e_2(t) + e_3(t) \tag{52}$$

and its power is given by

$$\sigma_e^2 = E\{e^2(t)\} = (2 + N_s(k))\sigma_R^2. \tag{53}$$

Note that

$$H_{ef}(z) = \frac{1}{1 - 2\cos 4\omega_k z^{-1} + z^{-2}}, \tag{54}$$

and the total iteration is reduced to $N/2$. Thus, the total power of the rounding error at the output becomes

$$\sigma_f^2 = \sigma_e^2 \sum_{n=0}^{N/2-1} |h_{ef}(n)|^2 = \frac{\sigma_e^2}{\sin^2(4\omega_k)} \left(\frac{N}{4}\right) = \frac{(2 + N_s(k))N\sigma_R^2}{4\sin^2(4\omega_k)}. \tag{55}$$

From (55), we observe that

1. Although the total number of noise sources increases, the total noise power is compensated by the halved number of iterations.

2. Compared with the factor $\frac{1}{\sin(2\omega_k)^2}$ in (41), the factor $\frac{1}{\sin(4\omega_k)^2}$ in (55) will have similar effect on the SNR of each DCT channel but with halved period.

Now let us consider the dynamic range of the low-power DCT structure with $M = 2$. Given the assumption that the input sequence $x(t)$ is an i.i.d. sequence, the decimated inputs $x_e(t)$ and $x_o(t)$ are also i.i.d. sequences and are uncorrelated with each other. Thus, we can apply the technique of "superposition" to analyze the dynamic range of the system: We first set $x_o(t)$ to zero while analyzing the dynamic range contributed by $x_e(t)$; then we perform the same analysis for $x_o(t)$ by setting $x_e(t)$ to zero. The overall D can be found from the summation of the two dynamic ranges, which is given by (see Appendix)

$$\begin{aligned} D_1 &= 2C(k)\left(|\cos\omega_k| + |\cos 3\omega_k|\right), \\ D_2 &= C(k)\sum_{n=0}^{N/2-1}\left(|\cos[(4n+1)\omega_k]| + |\cos[(4n+3)\omega_k]|\right), \\ D &= \max\{D_1, D_2\}. \end{aligned} \tag{56}$$

Using the analytical results in (55) and (56), we can also find the optimal wordlengths for $N = 8$ and 16 under the 40dB SNR constraint. The results are listed in Table 4(b). It is interesting to note that the average wordlengths of the multirate DCT architectures are even less than those of the normal DCT architectures. This is due to the fact that the number of the iterations in the IIR loop will be reduced to N/M. As M increases, the accumulation of the rounding errors becomes smaller and thus less wordlength can be allocated. This indicates that the multirate DCT architecture can not only reduce low-power consumption, its numerical properties also become better as M increases.

The above analyses can be extended to the low-power DCT design with decimation factor equal to M ($M \geq 2, M \in 2^{+Z}$). The results are given by

$$P_f = (M + N_s(k))\left(\frac{N}{2^{m+1}}\right)\frac{\sigma_R^2}{\sin^2(2^{m+1}\omega_k)}, \tag{57}$$

and

$$\begin{aligned} D_1 &= M \cdot C(k)\sum_{n=0}^{M-1}|\cos[(2n+1)\omega_k]|, \\ D_2 &= C(k)\sum_{n=0}^{\frac{N}{M}-1}\sum_{i=0}^{M-1}\left|\cos[(2^{m+1}n+2i+1)\omega_k]\right|, \\ D &= \max\{D_1, D_2\}. \end{aligned} \tag{58}$$

with $m = \log_2 M$.

4.5 Simulation Results

To verify our analytical results, computer simulations are carried out by using the aforementioned DCT architectures. The input sequence is a random sequence with uniform probability distribution over the interval (-1,1). All the results are based on the average of 1000 independent DCT computations. Fig.10 shows the average SNR as a function of the DCT channel number k. As we can see, there is a close agreement between the theoretical and experimental results. Basically, the SNR distribution is affected by the factor $\sin^2(2^{m+1}\omega_k)$ in (57) so that its period varies with the decimation factor M. It should be noted that although Fig.10 (a) and (b) yield similar SNR results, the uniform dynamic range of the direct form I structure makes it a better choice for VLSI implementations.

Fig.11 shows the relationship between the average SNR and the wordlength for $N = 16$. Compared to the simulation results in [13], the three IIR DCT architectures give comparative SNR performance to the DCT architectures by Hou [15] and Lee [16] under fixed-point arithmetic. It is worth noting that the multirate DCT architectures have better SNR results than the normal IIR DCT architectures; *i.e.*, the multirate DCT has better numerical properties under fixed-point arithmetic, which is consistent with what we have seen in Table 4.

In summary, the analytical results presented in this section can be used as a good index for future applications as N and/or M changes. Furthermore, we can assign the optimal wordlength for each individual DCT channel given the SNR criteria, while this is not the case in the fast-algorithm based PIPO DCT structures [15][16]. Due to the characteristics of global interconnections in the PIPO DCT structure, each operator at each stage will affect part or all of the outputs. Therefore, it is not easy to find optimal wordlength for each channel in the PIPO structure.

5 Conclusions

In this paper, we presented some new aspects of the multirate low-power design discussed in the companion paper [1]; namely, the logarithmic-complexity low-power architecture, unified low-power IIR module design, and optimal wordlength assignment. We have shown that logarithmic architecture is a good choice for VLSI implementation when both low-power dissipation and chip area are taken into consideration. The unified IIR module presented in Section 3 allows us to perform various

sinusoidal transforms using the same dedicated VLSI architecture. The real-time operations as well as the programmability of this design makes it a promising candidate to be incorporated into the design of video co-processor. Finally, the finite-wordlength analysis gives us a tool to achieve a desired SNR by choosing minimal wordlength. It not only reduces the total switching events (hence the power dissipation), but also provides a good control over the total chip area under the SNR constraint. The materials presented in this paper, together with the multirate approach in the companion paper, constitute a framework of the algorithm-based low-power design with application to transform coding kernel design.

Appendix

Derivation of (56)

Setting $x_o(t)$ to zero, Fig.8 is reduced to the IIR structure depicted in Fig.9, where $w_i(t)$, $i = 1, 2$, are the nodes that may have overflow. It is easy to see that

$$D_{1,e} = \max\{w_1(t)\} = C(k)\left(|\cos\omega_k| + |\cos 3\omega_k|\right). \tag{59}$$

From the transfer function of $w_2(t)$

$$H_2(z) = \frac{W_2(z)}{X_e(z)} = C(k)\frac{\cos 3\omega_k - \cos\omega_k z^{-1}}{1 - 2\cos 4\omega_k z^{-1} + z^{-2}}, \tag{60}$$

we can derive the unit-sample response as

$$h_2(n) = C(k)\cos[(4n+1)\omega_k]u(n). \tag{61}$$

Thus,

$$D_{2,e} = \max\{w_2(t)\} = C(k)\sum_{n=0}^{N/2-1}|\cos[(4n+1)\omega_k]|\cdot\max\{x(n)\} = C(k)\sum_{n=0}^{N/2-1}|\cos[(4n+1)\omega_k]|. \tag{62}$$

Similarly, by setting $x_e(t) = 0$, we can derive the dynamic ranges of the two circled nodes, $D_{1,o}$ and $D_{2,o}$, as

$$\begin{aligned} D_{1,o} &= C(k)\left(|\cos\omega_k| + |\cos 3\omega_k|\right), \\ D_{2,o} &= C(k)\sum_{n=0}^{N/2-1} |\cos[(4n+3)\omega_k]|. \end{aligned} \tag{63}$$

Combining (62) and (63) together, we can write the overall dynamic range of the multirate DCT as

$$\begin{aligned} D_1 &= D_{1,e} + D_{1,o} = 2C(k)\left(|\cos\omega_k| + |\cos 3\omega_k|\right), \\ D_2 &= D_{2,e} + D_{2,o} = C(k)\sum_{n=0}^{N/2-1} \left(|\cos[(4n+1)\omega_k]| + |\cos[(4n+3)\omega_k]|\right), \\ D &= \max\{D_1, D_2\}. \end{aligned} \tag{64}$$

References

[1] A.-Y. Wu and K. J. R. Liu, "Algorithm-based low-power transform coding architectures- Part I: The multirate approach," *submit to IEEE Trans. Circuits Syst. II: Analog and Digital Signal Processing*, 1995.

[2] K. J. R. Liu, C. T. Chiu, R. K. Kolagotla, and J. F. J. Ja', "Optimal unified architectures for the real-time computation of time-recursive discrete sinusoidal transforms," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 4, pp. 168–180, April 1994.

[3] H. S. Malvar and D. H. Staelin, "The LOT: Tansform coding without blocking effects," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. 37, pp. 553–559, April 1989.

[4] H. S. Malvar, "Lapped transforms for efficient transform/subband coding," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. 38, pp. 969–978, June 1990.

[5] H. S. Malvar, "Extended Lapped Transforms: Properties, applications, and fast algorithms," *IEEE Trans. Signal Processing*, vol. 40, pp. 2703–2714, Nov. 1992.

[6] E. Frantzeskakis, J. S. Baras, and K. J. R. Liu, "Time-recursive computation and real-time parallel architectures, Part I: Framework," Tech. Rep. TR 93-17r1, Institute for Systems Research, University of Maryland, 1993.

[7] E. Frantzeskakis, J. S. Baras, and K. J. R. Liu, "Time-recursive computation, Part II: Methodology, and application on QMF banks and ELT," Tech. Rep. TR 93-18r1, Institute for Systems Research, University of Maryland, 1993.

[8] A. P. Chandrakasan, M. Potkonjak, J. Rabaey, and R. W. Brodersen, "An approach for power minimization using transformations," in *VLSI signal processing V* (K. Yao, R. Jain, W. Przytula, and J. Rabaey, eds.), pp. 41–50, IEEE Press, 1992.

[9] A. P. Chandrakasan, S. Sheng, and R. W. Brodersen, "Low-power CMOS digital design," *IEEE J. Solid-State Circuits*, vol. 27, pp. 473–484, April 1992.

[10] H. S. Malvar, "Fast algorithm for modulated lapped transform," *Electron. Lett.*, vol. 27, pp. 775–776, Apr. 1991.

[11] C.-T. Chiu and K. J. R. Liu, "Real-time parallel and fully pipelined two-dimensional DCT lattice structures with application to HDTV systems," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 2, pp. 25–37, March 1992.

[12] A. V. Oppenheim and R. W. Schafer, *Discrete-time Signal Processing.* Prentice Hall, 1989.

[13] I. D. Yun and S. U. Lee, "On the fixed-point-error analysis of several fast DCT algorithms," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 3, pp. 27–41, Feb. 1993.

[14] V. Srinivasan and K. J. R. Liu, "Full custom VLSI inplementation of high-speed 2-D DCT/IDCT chip," in *Proc. IEEE Int. Conf. Image Processing*, (Austin, Texas), pp. III.606–610, 1994.

[15] H. S. Hou, "A fast recursive algorithm for computing the discrete cosine transform," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. 35, pp. 1455–1461, Oct. 1987.

[16] B. G. Lee, "A new algorithm to compute the discrete cosine transform," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. 32, pp. 1243–1245, Dec. 1984.

[17] Z. Wang, "Fast algorithms for the discrete W transform and for the discrete Fourier transform," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. 32, pp. 803–816, Aug. 1984.

	Normal DCT architecture in [2]	Logarithmic low-power DCT architecture	Full low-power DCT architecture in [1]
Multipliers	$2N-2$	$(\log M+2)N$ (in order)	$(M+1)N$ (in order)
Adders	$2N$	$(2\log M+1)N$ (in order)	$(M+1)N$ (in order)
Power consumption for 16-point DCT	P_0	$0.24P_0$ $(M=4)$	$0.11P_0$ $(M=4)$

Table 1: Comparison of hardware cost and power consumption of the logarithmic low-power DCT architecture with other approaches.

	Normal Operation		Downsampling by 2		Downsampling by 4	
	Multiplier	Adder	Multiplier	Adder	Multiplier	Adder
IIR MLT	$5N$	$5N$	$10N$	$11N$	$20N$	$23N$
IIR ELT	$6N$	$6N$	$11N$	$12N$	$21N$	$24N$

Table 2: Comparison of hardware cost for the MLT and ELT with their low-power designs in terms of 2-input multipliers and 2-input adders.

	L	β_1	ω_k	θ_k	Combination Function
DCT	N	$C(k)$	$\frac{k\pi}{2N}$	0	$X_{DCT,k}(t) = X_{C,k}(t)$
IDCT	N	$C(1)$	$\frac{\pi}{2N}(k+\frac{1}{2})$	$-\omega_k$	$X_{IDCT,k}(t) = X_{C,k}(t) + (C(0) - C(1))x(n-N+1)$
DST-IV in [17]	N	$C(1)$	$\frac{\pi}{2N}(k+\frac{1}{2})$	0	$X_{DST,k}(t) = X_{S,k}(t)$
IDST-IV in [17]	N	$C(1)$	$\frac{\pi}{2N}(k+\frac{1}{2})$	0	$X_{IDST,k}(t) = X_{S,k}(t)$
MLT	$2N$	$\frac{1}{\sqrt{2N}}$	$\frac{k\pi}{2N}$	$\frac{\pi}{2}(k+\frac{1}{2})$	$X_{MLT,k}(t) = -S(k)[\, X_{C,k+1}(t) + X_{S,k}(t)\,]$
ELT	$4N$	$\frac{1}{2\sqrt{2N}}$	$\frac{\pi}{2N}(k+\frac{1}{2})$	$\frac{\pi}{2}(k+\frac{1}{2})$	$X_{ELT,k}(t) = -X_{S,k+1}(t) + \sqrt{2}X_{C,k}(t) + X_{S,k-1}(t)$
DFT	N	$\frac{1}{\sqrt{N}}$	$\frac{-k\pi}{N}$	$-\omega_k$	$Re\{X_{DFT,k}(t)\} = X_{C,k}(t),\ Im\{X_{DFT,k}(t)\} = X_{S,k}(t).$
DHT	N	$\frac{1}{\sqrt{N}}$	$\frac{-k\pi}{N}$	$-\omega_k$	$X_{DHT,k}(t) = X_{C,k}(t) + X_{S,k}(t).$

Table 3: Parameter settings for the unified low-power IIR transformation architecture, where $Re\{X_{DFT,k}(t)\}$ and $Im\{X_{DFT,k}(t)\}$ denote the the real part and the imaginary part of the DFT, respectively.

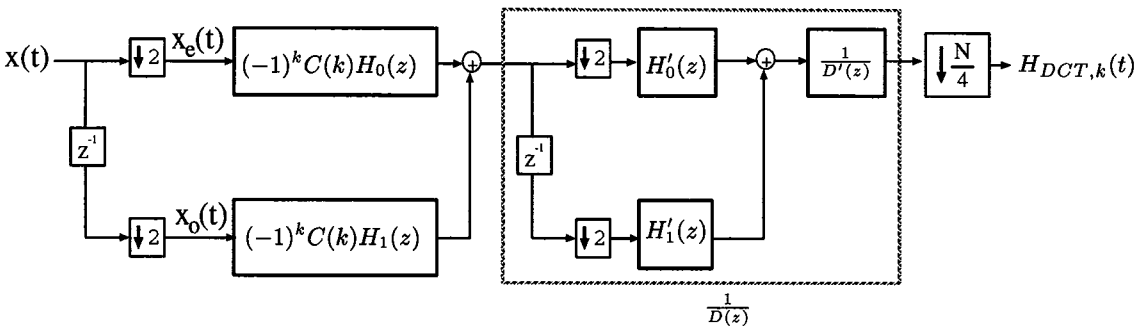
DCT channel k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	B_A
B_k ($N = 8$)	12	11	10	9	10	11	12	N/A								10.7
B_k ($N = 16$)	13	12	12	11	11	10	10	10	10	10	11	11	12	12	13	11.2

(a)

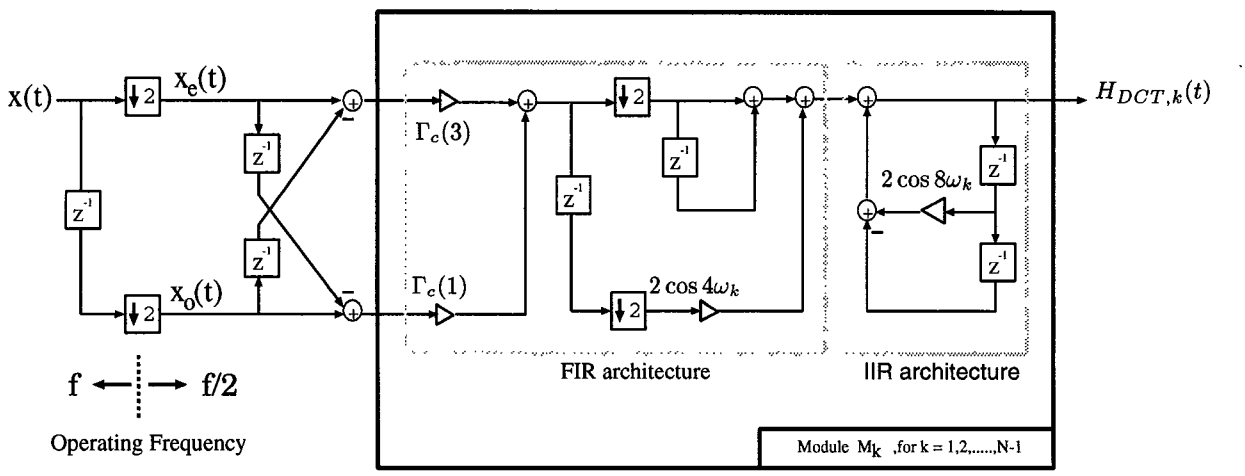
DCT channel k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	B_A
B_k ($N = 8, M = 2$)	10	9	10	11	10	9	10	N/A								9.9
B_k ($N = 16, M = 2$)	12	11	10	10	10	11	12	12	12	11	10	10	10	11	12	10.9

(b)

Table 4: Optimal wordlength assignment under the constraint SNR = 40dB, where B_A is the average wordlength. (a) Normal IIR DCT. (b) Low-power DCT with $M = 2$.



(a)



(b)

Figure 1: (a) Polyphase representation of $H_{DCT,k}(z)$ in cascade form. (b) Multirate DCT architecture with logarithmic complexity, where ω_k, $\Gamma_c(m)$, $m = 1, 3$, are defined in [1].

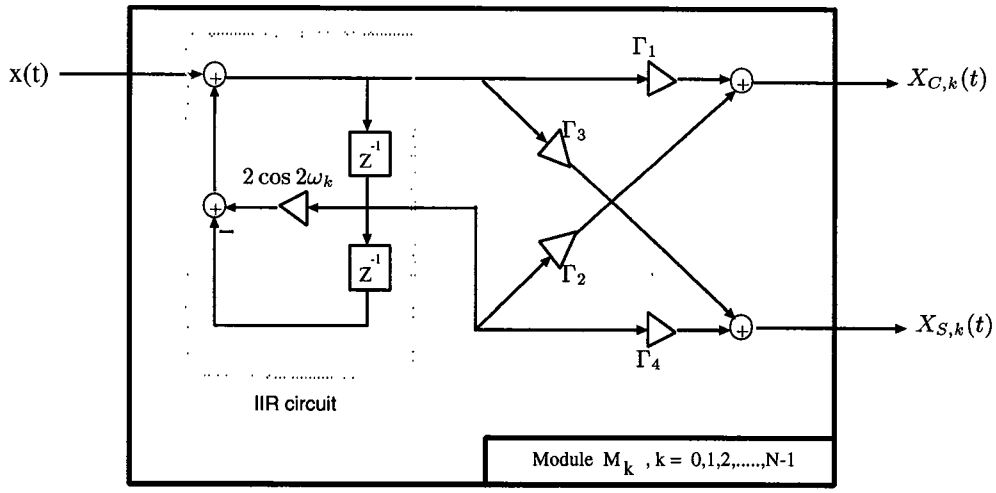


Figure 2: IIR MLT module design.

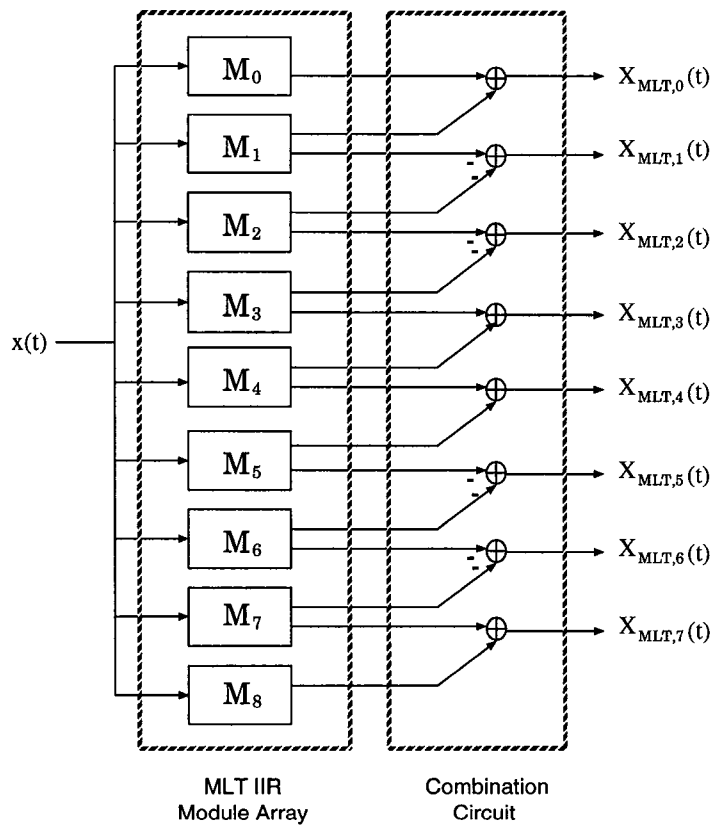


Figure 3: The time-recursive MLT architecture.

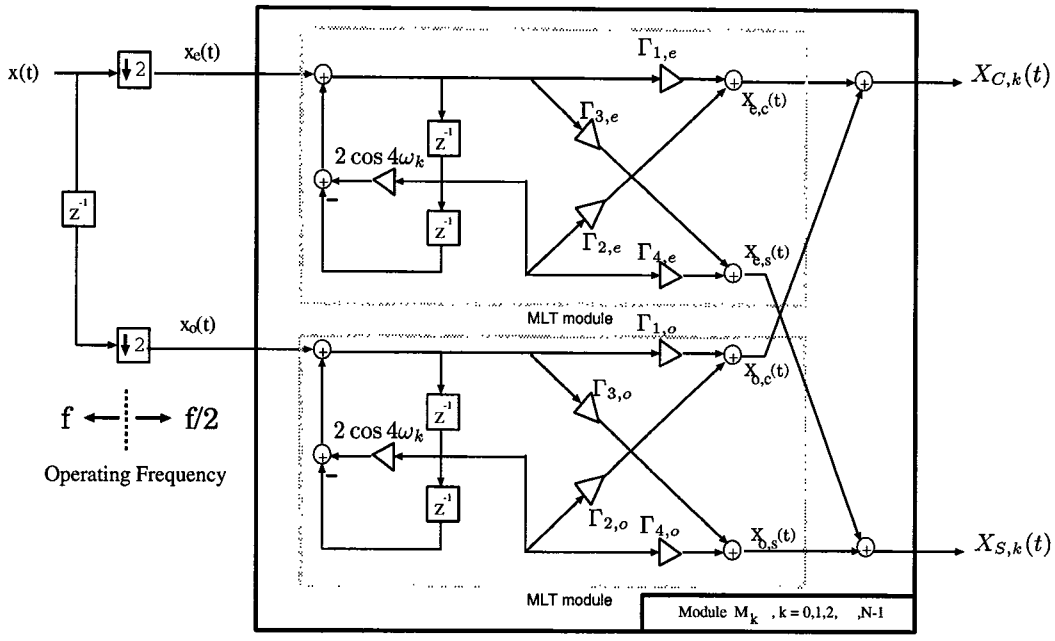


Figure 4: Low-power IIR MLT module design.

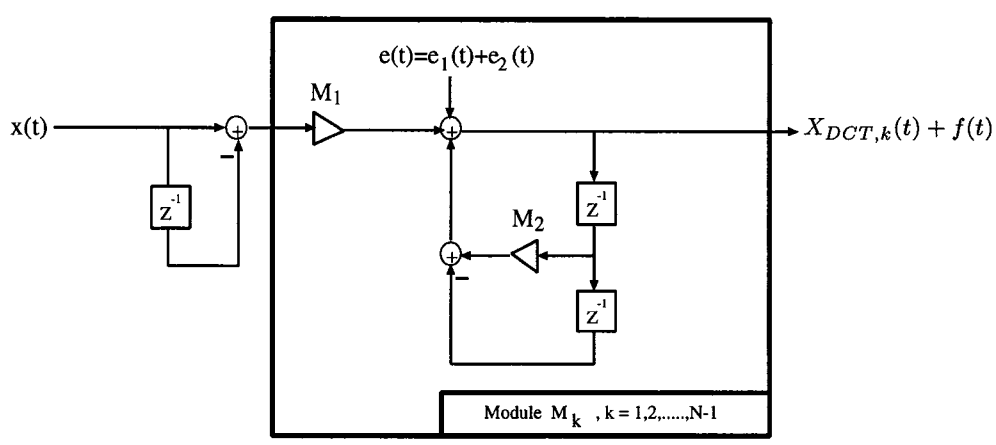


Figure 5: Rounding error in the IIR DCT architecture.

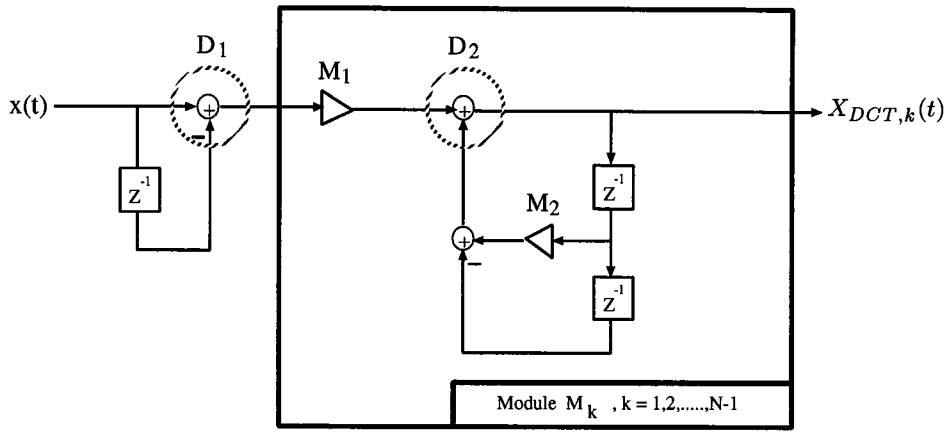


Figure 6: Dynamic range of the IIR DCT architecture.

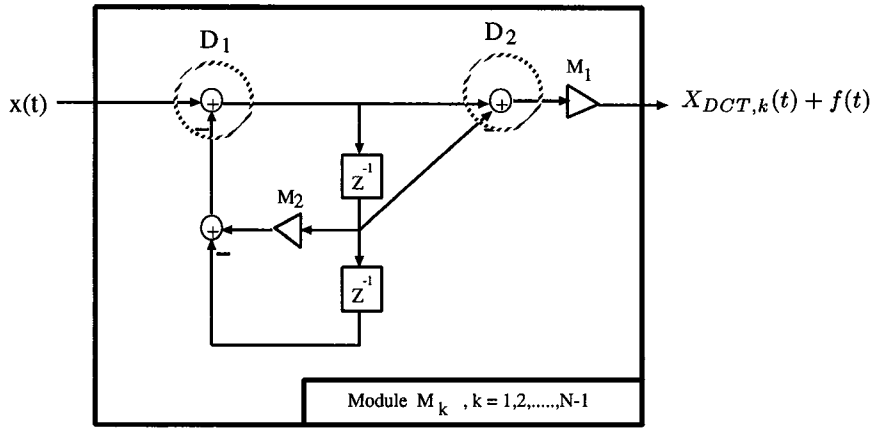


Figure 7: IIR DCT using the direct form II structure.

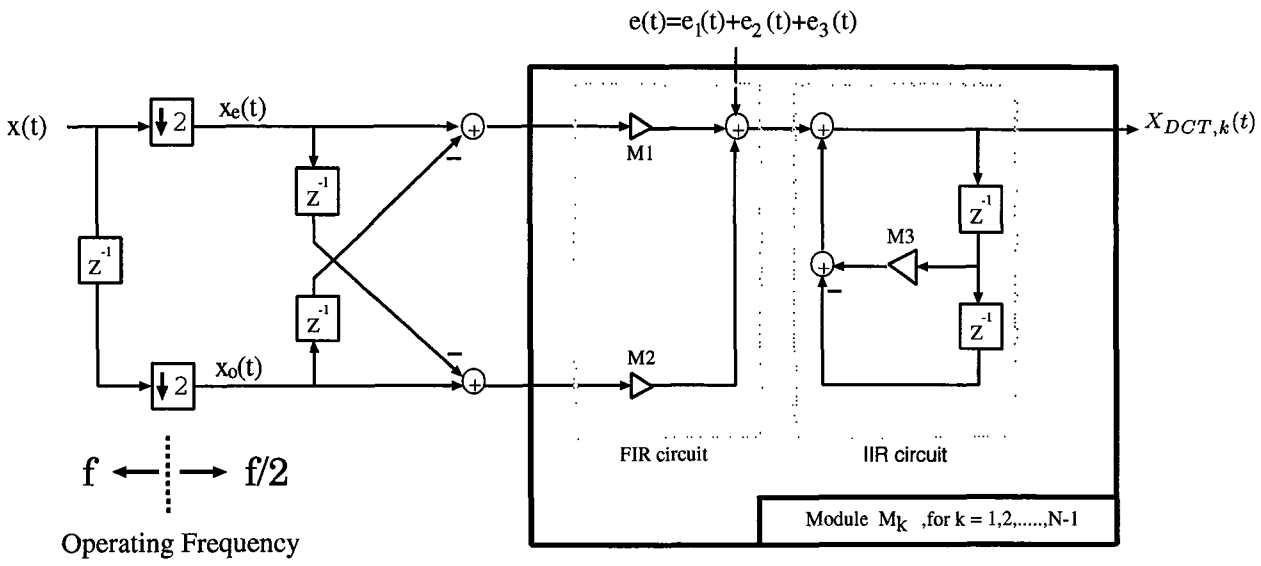


Figure 8: Rounding noise in the low-power IIR DCT architecture with $M = 2$.

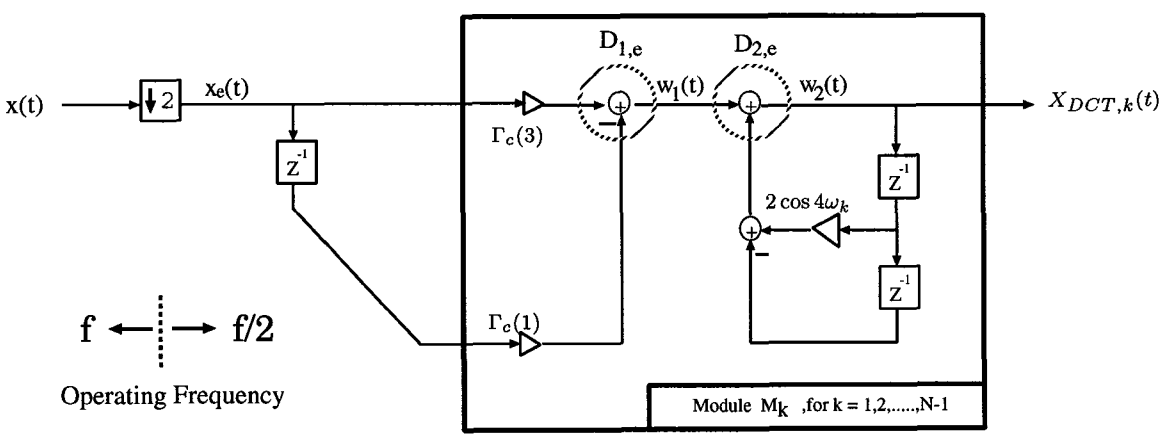


Figure 9: Reduced IIR DCT architecture with $M = 2$.

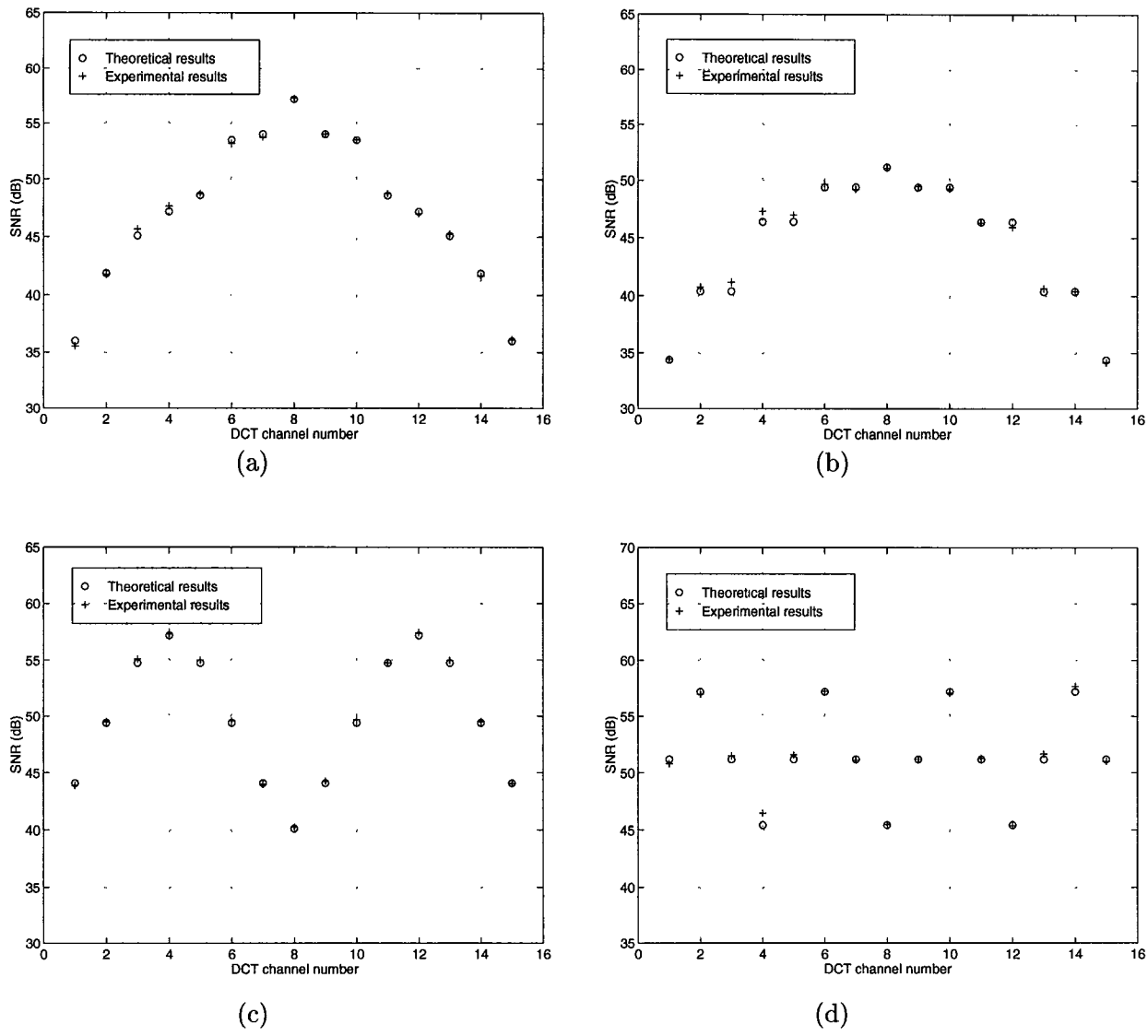


Figure 10: Average SNR as a function of DCT channel number under fixed-point arithmetic ($N = 16$, $B = 12$). (a) Normal IIR DCT using direct form I structure. (b) Normal IIR DCT using direct form II structure. (c) Low-power DCT with $M = 2$. (d) Low-power DCT with $M = 4$.

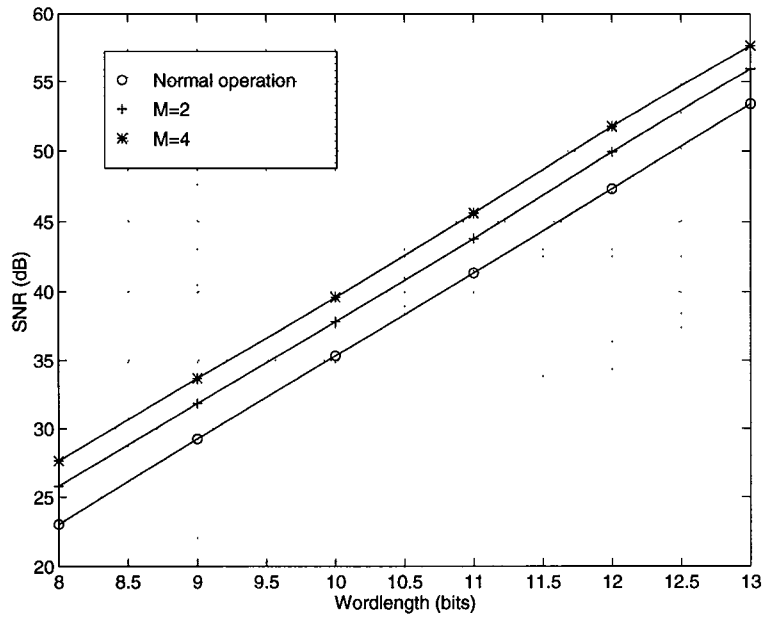


Figure 11: Average SNR as a function of wordlength under fixed-point arithmetic (N=16). The multirate low-power architectures have better SNR as M increases.