**MDCT/IMDCT- properties and applications**

**Introduction:**
MDCT is a lapped transform which is a type of linear discrete block transformation where the basis functions of the transformation overlap the block boundaries. However, the number of coefficients which results from a series of overlapping block transforms remains the same as any other non-overlapping block transformation.

The MDCT is based on the type-IV discrete cosine transform (DCT-IV) [11], which is given by

$$X_k = \sum_{n=0}^{N-1} x_n \cos \left[ \frac{\pi}{N} \left( n + \frac{1}{2} \right) \left( k + \frac{1}{2} \right) \right]$$

where, $k=0, 1, ..., N-1$

$x_n = 0, 1, ..., 2N-1$.

It is designed to be applied on consecutive blocks of a larger dataset, in cases where subsequent blocks overlap such that the last half of one block coincides with the first half of the next block.

MDCT is unusual compared to other Fourier-related transforms in that it has half as many outputs as inputs (instead of the same number). It is linear function i.e. $F: \mathbb{R}^{2n} \rightarrow \mathbb{R}^N$ where $\mathbb{R}$ denotes the set of real numbers.

**Definition**

**MDCT:**

$$X_k = \sum_{n=0}^{2N-1} x_n \cos \left[ \frac{\pi}{N} \left( n + \frac{1}{2} + \frac{N}{2} \right) \left( k + \frac{1}{2} \right) \right]$$

$k=0, 1, ..., N-1$

$x_n = h_n a_n$ is the windowed input signal; $a_n$ is the input signal with $2N$ samples and $h_n$ is the window function.

This window is an identical analysis-synthesis time window, given by-

$h_0 = h_{2N-1-k}$

$h_n^2 + h_{n+N}^2 = 1$

A sine window is widely used in audio coding as it offers good stop-band attenuation. It also provides attenuation of the block edge effect and allows perfect reconstruction. However, other optimized windows can also be applied. The sine window mentioned here can be defined as:

$h_n = \sin[h_n(n + \frac{1}{2})/ 2N]$,
**IMDCT:**

\[ y_n = \frac{1}{N} \sum_{k=0}^{N-1} X_k \cos \left[ \frac{\pi}{N} \left( n + \frac{1}{2} \right) \left( k + \frac{1}{2} \right) \right] \]

\[ n=0,1,\ldots,2N-1 \]

where, \( X_k \) the output signal of MDCT.

The \( y_n \) in the equation contain time domain aliasing [14]:

\[
y_n = \begin{cases} 
-x \frac{1}{2} a_n - \frac{1}{2} a_{N-1-n}, & n = 0,1,\ldots,N-1 \\
-x \frac{1}{2} a_n + \frac{1}{2} a_{3N-1-n}, & n = 0,1,\ldots,2N-1 
\end{cases}
\]

Though the IMDCT obtains \( y_n \), the original \( x_n \) can be perfectly obtained by adding the overlapped IMDCTs of subsequent overlapping blocks as shown in fig 1. This leads to cancellation of errors and hence the original data can be retrieved.

**Properties of MDCT [2]:**

1) MDCT is not an orthogonal transform. Perfect signal reconstruction can be achieved in the overlap-add (OA) process. For the overlap-add window of 2N samples, the first N and last N samples of the signal will remain modified. One can easily see this from the fact that performing MDCT and IMDCT of an arbitrary signal \( x_k \) reconstructs signal \( X_k \).

**Overlap-add process:**

![Fig 1. simple overlap and add algorithm.](image)

The Fig 1 describes the simple overlap and add algorithm. Here, first the signal \( x[n] \) is partitioned into non-overlapping sequences. After this the discrete Fourier transforms of the sequences \( y_k[n] \) are calculated [2]. This is done by multiplying the FFT of \( x[n] \) with the FFT of \( h[n] \). Then we recover the \( y_k[n] \) using inverse FFT [7], the resulting
output signal can be reconstructed using overlapping and adding the $y_k[n]$ as described in the Fig 1. The overlap is based on the idea that a linear convolution is always longer than the original sequences [2].

2) NONORTHOGONAL PROPERTY OF MDCT
   a) If a signal exhibits local symmetry such that
   \[ x_k = x_{N-k+1}, \quad k=0,1, \ldots, N-1 \]
   \[ x_k = -x_{3N-k-1}, \quad k=N,N+1, \ldots, 2N-1 \]
   Then its MDCT degenerates to zero; i.e. \( X_k = 0 \) for \( k = 0, \ldots, N-1 \).
   This is an example that MDCT does not fulfill Parseval’s theorem, i.e. the time domain energy is not equal to the frequency domain energy.

   b) If a signal exhibits local symmetry such that
   \[ x_k = -x_{N-k+1}, \quad k=0,1, \ldots, N-1 \]
   \[ x_k = x_{3N-k-1}, \quad k=N,N+1, \ldots, 2N-1 \]
   Then MDCT and IMDCT will perfectly reconstruct the original time domain samples.

3) MDCT becomes an orthogonal transform if the signal length is infinite. This is different from the traditional definition of orthogonality, which require a square transform matrix.

4) MDCT is similar to the orthogonal transforms such as DFT, DCT, DST and it also possesses energy compaction capability.

5) Invertibility: Since the number of inputs and outputs are not equal it may seem that the MDCT is not invertible, however MDCT is perfectly invertible which is achieved by adding the overlapped IMDCTs of subsequent overlapping blocks, this leads the errors to cancel and the original data will be retrieved. This technique is known as time-domain aliasing cancellation (TDAC) [10].

6) Performing MDCT and then IMDCT with one single frame of time domain samples, the original time samples cannot be perfectly reconstructed, instead the reconstructed samples are normally an alias-embedded version. MDCT by itself is a lossy process.
7) 90% energy is concentrated within 10% of the normalized frequency scale for most of the test signals for all transforms concerned [2]. The energy compaction property of different transforms becomes more unified with increasing window length.

8) Window shape has an impact on MDCT energy compaction property. In the case of rectangular window, DCT has always the best energy compaction property since DCT corresponds to an even extension of the signal.

9) MDCT is very useful with its time domain alias cancellation (TDAC) characteristics [10]. However, its mismatch with the DFT domain based psychoacoustic model must be kept in mind when developing a MDCT based audio codec with its full potential in terms of coding performance.

10) The distinct advantage of MDCT lies in its critical sampling property, reduction of block effect and the possibility of adaptive window switching.

11) In comparison to the orthogonal transforms, the MDCT has a special property i.e. the input signal cannot be perfectly reconstructed from a single block of MDCT coefficients of the sliding discrete Fourier transform- SDFT_{(N-1)/2,1/1} [12] are lost in the MDCT.
In terms of energy compaction property, MDCT does not have any advantage in comparison to DFT and DCT as indicated in Fig 2.

Fig 2. Cumulative power spectra of DCT (diamond), SDFT \((n+1)/2, -1/2\) (circle), DFT (star) and MDCT (triangle) with rectangular windows. The window size is 256. The length of the test sequence is 5292000 PCM samples [2].

**Variants of MDCT**

**Low Delay MDCT (LD-MDCT) [3]:** Non-overlapping transforms, such as DCT-IV leads to aliasing in lossy compression coding. MDCT on the other hand eliminate this aliasing effect. However to do this the MDCT requires a 50% overlap-add (OLA), and as a result it leads to a delay. The LD_MDCT (Low Delay MDCT) is developed to reduce this delay needed for the lookahead.
The 2N-point MDCT and N-point IMDCT are given as

\[ X(k) = \sqrt{\frac{2}{N}} \sum_{n=0}^{2N-1} x(n) w(k) \cos \left( \frac{(2n + 1 + D)(2k + 1)\pi}{4N} \right) \]

\[ \tilde{x}(n) = w(n) \sqrt{\frac{2}{N}} \sum_{k=0}^{N-1} X(k) \cos \left( \frac{(2n + 1 + D)(2k + 1)\pi}{4N} \right) \]

where, \( N \) is the frame length, \( x(n) \) is the original signal, \( \tilde{x}(n) \) is the reconstruction signal, \( w(n) \) is the reduced overlap window, and \( D \) is the phase for aliasing cancellation.

\( D \) in the case of conventional MDCT is fixed to \( N \). on the other hand, \( D \) in the case of Low delay MDCT changes according to the desired delay. \( D \) is equal to the lookahead length and is defined as:

\[ D = \frac{N}{2^d - 1}, \quad 2^d - 1 < N, \]

where \( d \) is a positive number, the result of the MDCT or IMDCT is changed by the value of \( D \).
Special characteristics of MDCT:

![Image of signal analysis/synthesis with MDCT, overlap-add procedure and perfect reconstruction of time domain samples.](image)

Fig3. Illustration of signal analysis/synthesis with MDCT, overlap-add procedure and perfect reconstruction of time domain samples. (a) a phase/frequency-modulated time signal; (b)(d)(f) MDCT spectra in different time slots, indicated as frames 1,2,3 in (a); (c)(e)(g) reconstructed time domain samples (with IMDCT) of frames 1,2,3 respectively; (h) the reconstructed time samples after the overlap-add procedure [1].

The Fig 3(a) illustrates the 50% window overlap. However, MDCT spectra of different time slots in Fig 4(b) Fig 4(d) Fig 4(f) are calculated with rectangular windows. The IMDCT time domain samples of frame 1,2,3 are shown in Fig 4(c) Fig 4(e) Fig 4(g) respectively. The reconstructed time domain samples after overlap-add (OA) procedure is shown in Fig 4(h).
Comparison of DFT, SDFT \((N+1)/2,1/2\) and MDCT:

The Fig 4 shows the fluctuation of MDCT spectrum in comparison with DFT and SDFT \((N+1)/2,1/2\) spectra. With a frequency-modulated time signal in Fig 2(a), the DFT power spectrum is very stable despite a moving window. Conversely, the MDCT spectrum is very unstable. The SDFT
The spectrum is in between. This shows that the SDFT can be used as a bridge to connect MDCT and DFT in audio coding applications [2].

**Application of MDCT:**

MDCT is employed in most modern lossy audio formats such as MP3, AX-3, Vorbis, Windows Media Audio, Cook and AAC which are different lossy audio compression codecs [13].

In MP3, the MDCT is not applied to the audio signal directly, but instead it is applied to the output of a 32-band polyphase quadrature filter (PQF) bank. The output of this MDCT is post processed by an alias reduction formula to reduce the typical aliasing of the PQF filter bank. Such as combination of a filter bank with an MDCT is called hybrid bank or a subband MDCT. AAC on the other hand normally uses as pure MDCT only the MPEG-4 AAC-SSR (AAC Scalable Sample Rate) variant by Sony electronics uses a four-band PQF bank followed by an MDCT. Similar to MP3 the ATRAC uses stacked quadrature mirror filters (QMF) followed by an MDCT.

The direct application of MDCT formula require $O(N^2)$ operations, however the number of operations can be reduced to only $O(N \log_2 N)$ complexity. This reduction can be done by recursively factorizing the computation as in the cases of an FFT. It is also possible to compute MDCT using other transforms such as DFT (FFT) or a DCT combined with $O(N)$ pre-and post-processing steps.

MDCT limits the sources of output distortion at the quantization stage. It is also used as an analysis filter. MDCT performs a series of inner products between the input data $x(n)$ and the analysis filter $h_k(n)$. This eliminates the blocking artifacts that would cause a problem during the reconstruction of the sample. The inverse MDCT reconstructs the samples without the blocking artifacts.

The signal representation in the MDCT is a dominant tool in high-quality audio coding because of its special properties. In addition to an energy compaction capability similar to DCT, the MDCT simultaneously achieves critical sampling, a reduction of the block effect and flexible window switching.

MDCT makes use the concept of time-domain alias cancellation (TDAC) whereas the quadrature mirror filter bank (QMF) uses the concept of the frequency-domain alias cancellation [10]. This can be viewed as a duality of MDCT and QMF. However, it is to be noted that MDCT also cancels frequency-domain aliasing, whereas QMF does not cancel time-domain aliasing this means that the MDCT is designed to achieve perfect reconstruction, QMF on the other hand does not produce perfect reconstruction.
Overlapped windows allow for better frequency response functions but carry the penalty of additional values in the frequency domain, thus these transforms are not critically sampled. MDCT thus has solved the paradox satisfactorily and is currently the best solution.

**Fig5. MDCT-based perceptual audio encoder [1].**

**Fig6. Modified Structure of MDCT-based perceptual audio encoder with less computation [1].**
Bit Stream in

\[\text{Frame unpacking (Decoding bitstream)}\]

\[\text{Reconstruction (Inverse Quantization)}\]

\[\text{Inverse Mapping (Synthesis filter bank, IMDCT)}\]

Audio out

\[\text{Fig7. Block diagram of an Audio Decoder.}\]

\[\text{MDCT-Domain Error Concealment:}\]

\[\text{Fig8. Illustration of a special problem with repetition scheme in MDCT domain. Shaded rectangles-corrupted data units; blank rectangle-error free ones; heavily shaded rectangles-uncalled alias window shape. Arrows indicate packet repetition operations. } n \text{ is an integer number representing data unit index [1].}\]

\[\text{Fig9. Example of window-type mismatch problem in case of simple packet repetition [1].}\]
Fig 10. Possible quadruple-drumbeat problem in case of beat replacement when using a long MDCT block. Original (beat 1) and inserted (beat 2) beats are not aligned in time, and their aliases (alias 1 and alias 2) do not cancel each other [1].

The MDCT spectrum of a signal is the Fourier spectrum of the signal mixed with its alias. This compromises the performance of MDCT as a Fourier spectrum analyzer and leads to possible mismatch problems between MDCT- and DFT-based perceptual models. However, MDCT has been applied successfully to perceptual audio compression without major problems if a proper window, such as a sine window, is employed [13].

The TDAC of an MDCT filter bank can only be achieved with the overlap-add process in the time domain [10]. Although MDCT coefficients are quantized in an individual data block MDCT is usually analyzed in the context of a continuous stream. In the case of discontinuity, such as editing or error concealment, the aliases of the two neighboring blocks in the overlapped area are not able to cancel each other out.

MDCT can achieve perfect reconstruction only without quantization, which is never the case in coding applications. On modelling the quantization as a superposition of quantization noise to the MDCT coefficients, the time-domain alias of the input signal will still be canceled, but the noise components will be extended as additional “noise alias”. In order to have 50% window overlap and critical sampling simultaneously, the MDCT time-domain window is twice as long as that of ordinary transforms such as DCT. Because of the increased time-domain window length, the quantization noise is spread to the whole window, thus making pre-echo more likely to be audible. Well-known solutions this problem is the window switching and temporal noise shaping. In very low bit-rate coding the high-frequency components are often removed. This corresponds to a very steep low-pass filter. Due to the increased window size, the ringing effect caused by high-frequency cutting is longer.


References:


