4.1

A Low-Power IDCT Macrocell for MPEG2 MP@ML Exploiting Data Distribution Properties for Minimal Activity

Thucydides Xanthopoulos† and Anantha Chandrakasan
Massachusetts Institute of Technology, Cambridge, MA 02139

I. Introduction and Background

This work describes the implementation of a low power IDCT chip targetted to medium and low bitrate applications. Our strategy for reducing the chip power was two-fold: First, we selected an IDCT algorithm that minimizes activity by exploiting the relative occurence of zero-valued DCT coefficients in compressed video. Previous IDCT implementations (e.g. [1],[2]) have relied on conventional fast IDCT algorithms that perform a constant number of operations per block independent of the data distribution. Our approach performs a variable number of operations that depends on the statistical properties of the input data. Second, we minimized the energy through aggressive voltage scaling using deep pipelining and appropriate circuit techniques so that the chip could produce 14 Msamples/sec (640x480, 30 fps, 4:2:0) at 1.3V in a standard 3.3V process ($VTP = -0.9V, VTN = 0.7V$) and meet the requirement for MPEG2 MP@ML.

II. Algorithm, Architecture and Circuits

The 8-point 1-D Inverse DCT of a coefficient vector X is given by the following equation:

$$x[n] = \sum_{k=0}^{7} \frac{c[k]}{2} X[k] cos(\frac{(2n+1)k\pi}{16}), c[k] = \frac{1}{\sqrt{2}} \; if \; k = 0,1 \; o.w. \quad (1)$$

The 2-D IDCT of an 8×8 coefficient block can be computed by applying eq.1 on each block row, transposing the intermediate result and reapplying eq.1 on each column. Although there have been numerous fast IDCT algorithms that minimize the number of multiplications and additions implied by eq.1, we have decided to adopt a different approach based on the DCT coefficient characteristics of compressed images and video. MPEG-compressed video exhibits a large percentage of zero-valued DCT coefficients. In this case, we observe that direct application of eq.1 will result in a small total number of operations since multiplication and accumulation with a zero-valued $X[k]$ coefficient may constitute a NOP. Appropriate implementation of eq.1 can result in an algorithm with a variable number of operations per block as opposed to conventional architectures that perform a constant number of operations and do not exploit data properties. Conventional butterfly-style operations absorb the zero-valued coefficients early in the signal path without affecting average switched capacitance.

The IDCT chip consists of two 1-D IDCT units implementing eq.1 and a transposition memory structure in between (Fig.1). Every 8 cycles, the 1-D IDCT of a block row has converged within the 8 accumulators of the first stage. The intermediate result is transposed on-the fly by the structure (TRAM) of Fig.2. The second 1-D stage performs the same operations on the columns. The multiply-accumulate operations must be pipelined by a factor of 4 (Fig.3) to meet the bandwidth requirement at $V_{DD} = 1.3V$. The 24-T hybrid full adder of Fig.4 has been used because it eliminates long PMOS pullup chains from the signal path and is 37% faster than a 24-T static CMOS adder at $V_{DD} = 1.3V$. The bit width of the arithmetic units has been carefully selected to meet IEEE IDCT precision standard 1180-1990.

The presence of many zero-valued coefficients must be exploited in order to reduce the switching activity and reap the low power benefits of the selected algorithm. Clock gating in a pipeline can be implemented if each stage uses a separate clock net gated by an appropriate qualifying pulse. The qualifying pulse propagates from stage to stage along with the non-zero coefficient that requires processing. If a zero-valued coefficient enters the pipeline, only the stage that corresponds to the zero is powered down. The other upstream and downstream stages remain unaffected. The IDCT chip features ten separate clock nets in addition to the master clock for fully qualifying all steps of the entire pipeline formed by both 1-D stages. A clock-gated pipeline is susceptible to race conditions since the clock nets are not nominally equipotential: If $t_1 - t_0 > t_{CLK \rightarrow Q} + t_{pd} - t_{hold}$ (Fig.4a), the wrong data will be sampled at the second stage. This problem becomes acute when $t_{pd} \rightarrow 0$ (i.e. back-to-back shift registers). In such cases a negative level-sensitive latch was inserted (Fig.4b) to ensure functional correctness with a minimal penalty (<2%) in power and no effect on the critical path.

III. Test Results

The chip (Fig.6) is functional over a wide range of frequencies and power supplies (Fig.7). Over 2.8 million separate power measurements on an 8×8 block basis have been taken during several weeks while the chip was stimulated from 6 different MPEG2 sequences. The results of these measurements (average and 95% confidence interval) are plotted in Fig.8 vs. the number of non-zero coefficients whithin each DCT block. The block non-zero content histogram is also shown. Power dissipation shows strong correlation vs. the non-zero coefficients because of the data-dependent processing algorithm. Conventional architectures on the other hand exhibit a virtually flat power dissipation profile vs. non-zero block content. This property enables the IDCT chip to tradeoff image quality (quantization level) and power. The average power dissipation for this data set is 4.65 mWatts at 1.3 V, 14 MHz (5.68 NZ coefficients on average per block). The IDCT chip exhibits lower switched-capacitance per sample (factor of 2) from past IDCT chips ([1],[2]) when normalized for equal process feature size. The power performance of the chip improves significantly at lower bitrates (coarser quantization). This makes our approach ideal for emerging quality-on-demand compression protocols that tradeoff power dissipation and video quality.

† T. Xanthopoulos is partially supported by a National Semiconductor Fellowship. This work is supported by SHARP Corporation.

References

[1] M. Matsui et al., "200 MHz video compression macrocells using low-swing differential logic," in *IEEE International Solid-State Circuits Conference*, Feb. 1994, pp. 76–77.

[2] T. Kuroda et al., "A 0.9V, 150-MHz, 10-mW, $4mm^2$, 2-D discrete cosine transform core processor with variable-threshold-voltage (VT) scheme," *IEEE Journal of Solid State Circuits*, vol. 31, no. 11, pp. 1770–1777, Nov. 1996.

 0-7803-4766-8/98/$10.00 © 1998 IEEE

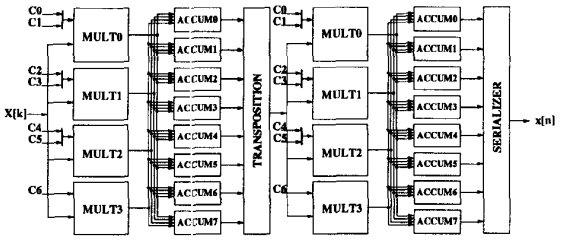

Fig. 1. IDCT Chip Block Diagram

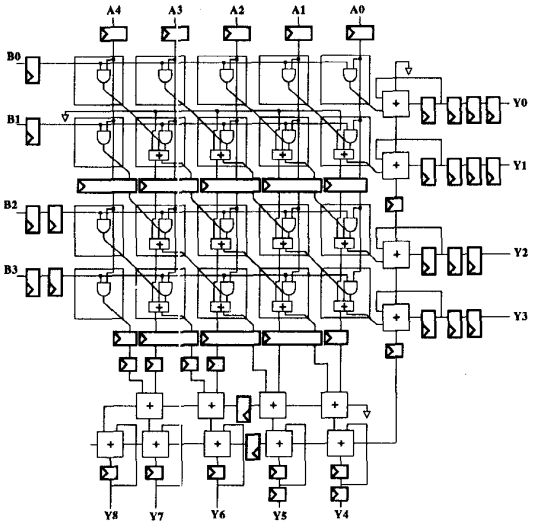

Fig. 3. Pipelined MAC Structure. A 14x14 MAC is actually used

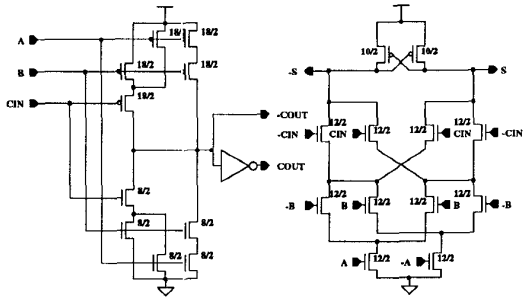

Fig. 5. 24-T Hybrid Adder (Static CMOS, CVSL)

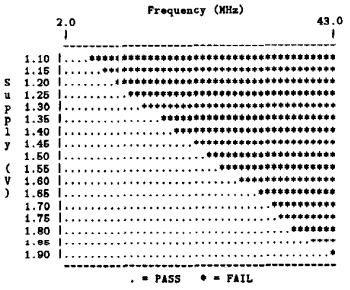

Fig. 7. IDCT Chip Schmoo Plot

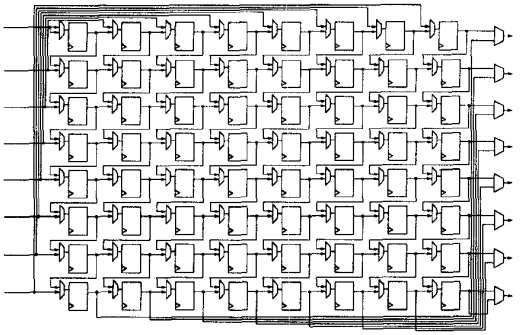

Fig. 2. Transposition Structure (TRAM). Data is Transposed On-The-Fly by Changing the Shifting Direction (T → B or L → R.)

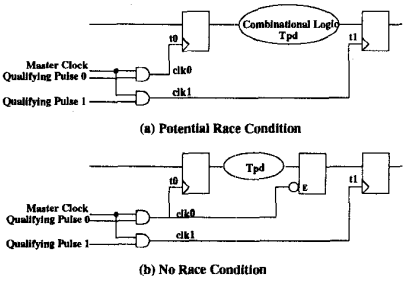

Fig. 4. Potential Race Conditions in Clock-Gated Pipelines

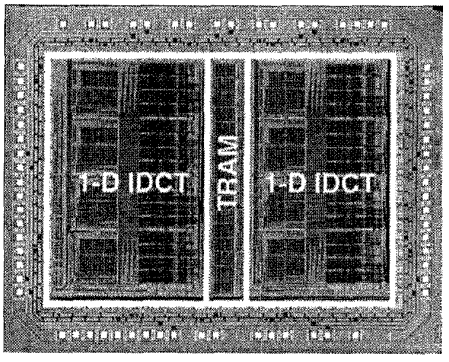

Fig. 6. IDCT Chip Microphotograph

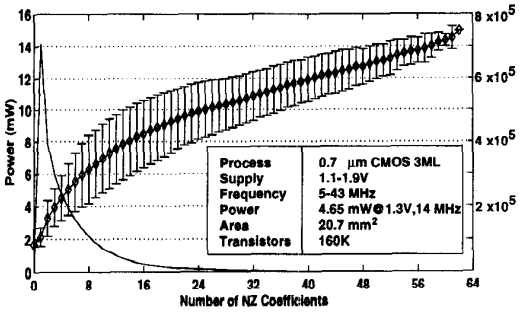

Fig. 8. IDCT Chip Measured Power Results at 1.32V, 14 MHz. This plot represents over 2.8 million separate power measurements.