

ARQUITETURA DE HARDWARE DA MTS PARA O CODIFICADOR VERSATILE VIDEO CODING

BIANCA SILVEIRA¹; DANIEL PALOMINO²; CLÁUDIO DINIZ³;
GUILHERME CORREA⁴

¹Universidade Federal de Pelotas, ViTech – bscsilveira@inf.ufpel.edu.br

²Universidade Federal de Pelotas, ViTech - dpalomino@inf.ufpel.edu.br

³Universidade Federal do Rio Grande do Sul, PGMICRO – cmdiniz@inf.ufrgs.br

⁴Universidade Federal de Pelotas, ViTech – gcorrea@inf.ufpel.edu.br

1. INTRODUÇÃO

Devido ao aumento na demanda de utilização de vídeos digitais, o processo de compreensão de vídeo tem se tornado de extrema importância para acelerar a transmissão. Considerando os dados da CISCO, 2018, estima-se que até o ano de 2022 somente o processo de transmissão de vídeo ocupará o 82% do tráfego de IP da internet.

Um dos novos padrões de compressão de vídeo que tem sido desenvolvido nos últimos anos, é o Versatile Video Coding (VVC). Este padrão apresenta novas ferramentas de codificação que permitem um ganho de codificação quando comparado com outros padrões (até 40% em comparação com o HEVC). Somada à essas vantagens de codificação, surge uma elevada complexidade computacional das operações, acarretando a necessidade do uso de técnicas de otimização algorítmica e implementação de hardware de baixo consumo.

Uma das inovações no VVC é o tipo de particionamento dos blocos. O tamanho dos blocos podem variar de 4x4 até 128x128, inclusive, podendo ser de tamanhos retangulares do tipo $N \times N/2$, $N/2 \times N$, $N \times N/4$ e $N/4 \times N$. Uma outra característica deste novo padrão é a introdução da Multiple Transform Selection (MTS) nas etapas de predição Intra e Inter. A MTS é uma ferramenta composta por três tipos de transformadas: transformadas discretas do cosseno tipo II e tipo VIII (DCT-II e DCT-VIII) e a transformada discreta do seno tipo VII (DST-VII). Estas transformadas possuem tamanhos de blocos que podem chegar a 64×64 para DCT-II e 32×32 para DCT-VIII/DST-VII. A utilização de técnicas que permitam acelerar o processo de codificação torna-se necessária para aproveitar estes novos recursos, e desta maneira reduzir a complexidade das operações. Assim mesmo, como citado por KAMMOUN, 2019, é crucial fornecer um projeto de alto desempenho para as restrições de hardware, desde que o suporte as múltiplas transformadas da ferramenta MTS apresenta consequências relacionadas aos recursos de memória e lógica de alocação.

2. METODOLOGIA

A introdução da ferramenta MTS no codificador VVC permitiu a utilização de mais de um tipo de transformada na codificação de um mesmo bloco de resíduo ao contrário do antecessor HEVC que apenas era utilizada a DCT-II. No VVC, duas novas transformadas foram incluídas: DCT-VIII e DST-VII. Outra diferença importante entre o VVC e HEVC é que neste novo codificador os tamanhos de blocos transformados podem ser quadrados e retangulares de tamanhos 4x4 até

64x64 para a DCT-II e 4x4 até 32x32 para as DCT-VIII e DST-VII. Para os blocos retangulares, os tamanhos podem ser de $N/2 \times N$, $N \times N/2$, $N/4 \times N$ e $N \times N/4$.

Para cada bloco residual é aplicada uma transformada horizontal em uma dimensão (1D), resultando em um bloco intermediário $N \times M$. Após, uma transformada vertical 1D é aplicada neste bloco intermediário resultando no bloco final transformado. O que caracteriza a MTS é que os tipos de transformadas utilizadas nessas duas dimensões podem ser diferentes. A escolha de qual transformada a ser utilizada em cada dimensão é realizada durante o tempo de codificação do vídeo. As possíveis combinações entre os tipos de transformadas são realizados por *flags* encontradas no software de referência do padrão.

A arquitetura proposta neste trabalho é projetada para realizar os três tipos de transformadas 1D permitidas na ferramenta MTS e as suas combinações em transformadas 2D. A arquitetura suporta tamanhos de blocos quadrados e retangulares: 4x4, 4x8, 4x16, 8x4, 8x8, 8x16, 16x4, 16x8 e 16x16.

Nos software de referência do VVC, pode-se notar que parte da matriz de 16 pontos é composta pela matriz de 8 pontos, e parte desta, por sua vez, é composta pela matriz de 4 pontos. Assim, é possível identificar operações a serem reutilizadas das matrizes menores para as maiores.

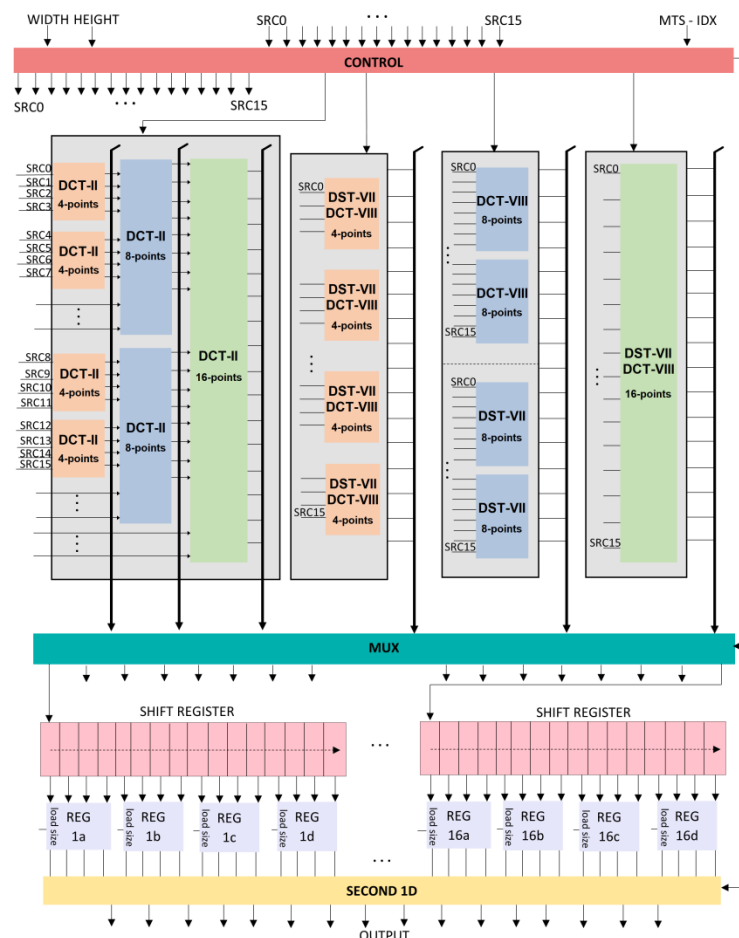


Figura 1: Arquitetura proposta da MTS

Para as transformadas DCT-VIII e DST-VII, não foi possível identificar semelhança entre os coeficientes para diferentes tamanhos de matriz, de modo que não podem ser reutilizados como em DCT-II. Por outro lado, ao comparar os

coeficientes DCT-VIII e DST-VII nota-se uma semelhança entre eles: seus valores são espelhados, muitas vezes com sinais invertidos.

A Figura 1 apresenta a arquitetura proposta da MTS unificada. A arquitetura é projetada para processar um bloco residual do tamanho de 16x16 que pode ser dividido em todos os diferentes sub-blocos permitidos no VVC.

Para concluir o cálculo 1D de todo o bloco 16x16 residual, a arquitetura leva 20 ciclos de clock (16 ciclos + 4 ciclos de latência inicial para preencher o pipeline). No quarto ciclo de clock, os valores das quatro primeiras linhas do bloco 16x16 intermediário estão disponíveis, então essas saídas já podem ser utilizadas como entrada para realizar os cálculos da segunda transformada 1D. A arquitetura tem uma latência inicial para executar as duas dimensões de todo o bloco 16x16 de 36 ciclos de clock. Depois disso, são necessários 16 ciclos para processar a transformação 2D.

3. RESULTADOS E DISCUSSÃO

O projeto proposto foi sintetizado usando a ferramenta *Cadence RTL Compiler 14.02* com a tecnologia de células padrão TSMC 40nm. A tabela 1 mostra os resultados em termos de área do circuito (*Kgates*) e potência total (mW) para diferentes frequências alvo (MHz). As frequências alvo foram calculadas com base na taxa mínima de processamento para diferentes resoluções de vídeo a serem codificadas em tempo real. As resoluções de vídeo consideradas são: HD 720p a 30 quadros por segundo (fps), FHD 1080p a 30 fps e 60 fps e UHD 4K a 30 fps e 60 fps.

TABELA 1. Resultados de Síntese

Resolução	Frequência (MHz)	Área (kgates)	Potência (mW)
HD@30fps	15,55	838,48	54,37
FHD@30fps	34,99		86,81
FHD@60fps	69,98		183,89
4K@30fps	139,97		362,68
4K@60fps	274,94		583,9

A tabela 1 mostra os resultados de potência total extraídos da síntese realizada para cada resolução e sua respectiva frequência alvo. Os resultados mostram que, como esperado, a potência total aumenta à medida que a frequência aumenta, portanto, é natural que a potência total mais alta seja para o throughput de 4K@60fps. Ressalta-se também que, à medida que a frequência dobra de valor, a potência total também apresenta comportamento semelhante, exceto entre HD@30fps e FHD@30fps, onde a potência não aumenta proporcionalmente à frequência.

A maioria das arquiteturas de hardware encontradas na literatura suportam apenas tamanhos de blocos quadrados e são sintetizadas apenas para FPGA. É o caso dos trabalhos em MERT, 2017, KAMMOUN, 2019 e GARRIDO, 2019. O trabalho em ZENG, 2021 suporta blocos retangulares mas implementa apenas as transformadas DCT-VIII e DST-VII. Os dois trabalhos que apresentam resultados de síntese para ASIC são FAN, 2019 e FARHAT, 2020. A solução em FAN, 2019 inclui apenas as transformadas DCT-VIII e DST-VII e apresenta resultados para a tecnologia de 65 nm. A arquitetura de hardware em FARHAT, 2020 suporta todos os tamanhos de bloco permitidos no módulo MTS, mas tem como alvo o módulo das transformadas inversas. Além disso, os resultados de área e frequência são

apresentados apenas para as transformadas 1D. A arquitetura proposta neste trabalho tem como objetivo o codificador, suporta todas as combinações de transformadas permitidas na ferramenta MTS e suporta todos os tamanhos de blocos de 4x4 a 16x16, incluindo formas retangulares.

4. CONCLUSÕES

Este trabalho propôs uma arquitetura de hardware unificada para o módulo MTS do codificador de vídeo VVC. A arquitetura suporta tamanhos de bloco $M \times N$ ($M, N = 4, 8, 16$) e todas as combinações de transformadas da MTS. Os resultados sugerem que a ideia de desenvolver e implementar uma arquitetura de hardware eficiente utilizando a reutilização de operações entre diferentes tipos e tamanhos de transformadas permite otimizar os processos de codificação do VVC para processamento em tempo real. A arquitetura proposta é capaz de processar vídeos de até 4K@60fps em tempo real, com dissipação de potência de 583 mW. Vale ressaltar que a arquitetura pode ser estendida para tamanhos maiores (32x32) aplicando as mesmas técnicas de reutilização de hardware.

5. REFERÊNCIAS BIBLIOGRÁFICAS

CISCO, “Cisco visual networking index: Forecast and trends, 2017–2022,” Tech. Rep., Nov. 2018. [Online]. Available: www.cisco.com

MERT, A. C.; KALALI E.; AND HAMZAOGLU, I., “High performance 2D transform hardware for future video coding,” IEEE Transactions on Consumer Electronics, vol. 63, no. 2, pp. 117–125, 2017.

KAMMOUN A.; HAMIDOUCHE, W.; PHILIPP, P.; BELGHITH, F.; MASSMOUDI, N.; AND NEZAN, J.F., “Hardware acceleration of approximate transform module for the versatile video coding standard,” in 27th European Signal Processing Conference (EUSIPCO). IEEE, 2019, pp. 1–5.

GARRIDO, M. J.; PESCADOR, F.; CHAVARRIAS, M.; LOBO, P. J.; SANZ, C., “A 2-D multiple transform processor for the versatile video coding standard,” IEEE Transactions on Consumer Electronics, vol. 65, no. 3, pp. 274–283, 2019

ZENG, Y.; SUN, H.; KATTO, J.; AND FAN, Y., “Approximated reconfigurable transform architecture for VVC,” in IEEE International Symposium on Circuits and Systems (ISCAS). IEEE, 2021, pp. 1–5.

FAN, Y.; ZENG, Y.; SUN, H.; KATTO, J.; AND ZENG, X., “A pipelined 2d transform architecture supporting mixed block sizes for the vvc standard,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 30, no. 9, pp. 3289–3295, 2019.

FARHAT, I.; HAMIDOUCHE, W.; GRILL, A.; MENARD, D.; AND DÉFORGES, O., “Lightweight hardware implementation of VVC transform block for ASIC decoder,” in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020, pp. 1663–1667.