

ARQUITETURA DE HARDWARE COM DUPLO MODO DE OPERAÇÃO PARA A PREDIÇÃO INTRA-QUADRO DIRECIONAL DO CODIFICADOR AV1

LUIZ NETO¹; MARCEL CORREA^{1,2}; DANIEL PALOMINO¹; LUCIANO AGOSTINI¹; GUILHERME CORREA¹

¹Video Technology Research Group – Universidade Federal de Pelotas

²Instituto Federal de Educação, Ciência e Tecnologia Sul-rio-grandense
{lcmdcneto, mmc Correa, dpalomino, agostini, gcorrea}@inf.ufpel.edu.br

1. INTRODUÇÃO

Devido a popularização significativa de serviços baseados em vídeos digitais nas últimas décadas, o que é percebido pelo crescimento das plataformas de streaming para entretenimento, TV digital, principalmente transmissões em tempo real e vários outros serviços de mídia. Tais serviços acabam criando diversos desafios na área de codificação de vídeo. Um vídeo *Full HD* (1920x1080 pixels) com uma taxa de quadros de 30 quadros por segundo e composto por apenas 30 segundos requer aproximadamente 1,8 GB para ser representado quando descompactado, o que dificulta bastante sua transmissão e armazenamento. Por conta destes fatos, a codificação de vídeo se torna indispensável.

Muitos formatos para codificação de vídeo foram desenvolvidos nos últimos anos, como *Advanced Video Coding* (H.264/AVC) (WIEGAND et al, 2003), *High Efficiency Video Coding* (HEVC) (SULLIVAN et al, 2012) e *Versatile Video Coding* (VVC) (ISO/IEC, 2020), lançados em 2003, 2013 e 2020, respectivamente. No entanto, esses formatos estão cobertos por royalties e este fato dificulta sua distribuição. Por causa disso, os formatos livres de royalties para codificação de vídeo têm sido desenvolvidos, como VP9 (MUKHERJEE, 2013). O mais novo formato de codificação de vídeo livre de royalties é chamado de *AOMedia Video 1* (AV1) (CHEN et al, 2018), e foi desenvolvido por a *Alliance for Open Media* (AOM). Desde então, o software de referência do AV1, o *libaom* (RIVAZ; HAUGHTON, 2018), atinge bons resultados em termos de compressão e qualidade de imagem, porém, devido a complexidade de algumas ferramentas inclusas no mesmo, o seu custo computacional acaba se tornando elevado, fazendo com que a codificação em tempo real não seja praticável, por conta da lentidão do processo. Portanto, soluções em hardware de algumas destas ferramentas se tornam uma boa maneira de fazer com que o processo de codificação seja mais rápido, possibilitando a codificação em tempo real.

Este trabalho propõe uma arquitetura de hardware dedicada com dois modos de operação para a predição intra-quadro utilizando os modos direcionais do AV1, onde um modo visa maior qualidade de imagem, o *High Quality* (HQ), ativando a etapa de pré-processamento e todos os modos direcionais. O outro modo visa uma menor dissipação de potência da arquitetura, o *Low Power* (LP), desabilitando a etapa de pré-processamento e utilizando apenas os ângulos nominais. A arquitetura desenvolvido visa processar resoluções até ultra-alta definição (UHD) em tempo real.

2. REFERENCIAL TEÓRICO

Antes que o processo de predição intra-quadro aconteça, o AV1 passa por um processo de particionamento de blocos, ou seja, cada quadro do vídeo que

deve ser codificado é dividido em blocos de píxels chamados de super blocos (SBs) e pode assumir o tamanho de 64×64 ou 128×128. A partir desta primeira divisão, cada SB pode ser sub-dividida de nove maneiras diferentes, sendo que sub-divisões resultantes em blocos quadrados podem ser sub-divididas novamente, até o tamanho de 4x4. O processo de predição intra-quadro do AV1 suporta, para cada tamanho de bloco, modos direcionais e não direcionais. Na predição com os modos direcionais, a última linha do bloco reconstruído acima do bloco atual é utilizada como referência (*AboveRow*), assim com a última coluna do bloco à esquerda do atual (*LeftCol*). Além disso, a predição direcional é dividida em duas etapas, uma etapa de pré-processamento, que inclui a filtragem das amostras de referência e um processo de *upsampling* e a outra etapa é onde ocorre a predição em si.

Visando suavizar as amostras de referência, a etapa de filtragem acontece com o intuito de adicionar maior precisão ao processo de predição. Existem três intensidades de filtros diferentes disponíveis e são chamados de filtro de força um, dois e três, onde os filtros de força um e dois utilizam três amostras de referência para filtrar uma única amostra, já o filtro de força três utiliza cinco amostras para filtrar uma, além desses filtros, existe a possibilidade de não aplicar os filtros. Após a etapa de filtragem, acontece o processo de *upsampling*, que serve para dobrar a quantidade amostras de referência utilizadas no processo de predição, as amostras são criadas entre as amostras originais, através de interpolação de quatro amostras espacialmente vizinhas.

O AV1 tem suporte para 56 modos direcionais na predição intra quadro, tendo como base oito ângulos nominais (45, 67, 90, 113, 135, 157, 180 e 203), o restante dos ângulos são derivados dos oito nominais aplicando um deslocamento de -9, -6, -3, +3, +6, +9 graus. Ângulos menores que 90 utilizam as amostras do *AboveRow*, ângulos maiores que 180 utilizam as amostras do *LeftCol* e ângulos entre 90 e 180 utilizam ambos. Para realizar a predição de uma única amostra, é realizada a projeção dela no *array* de referência, essa projeção é representada pelas equações (1-3), onde *RefArray* é o *array* que armazena as amostras de referência, *base* é o índice da amostra de referência resultante da projeção da amostra a ser predita e *shift* é um coeficiente usado para realizar a multiplicação pelas amostras do *array* de referência e *Pred_{x,y}* é o valor da amostra predita localizada na posição determinada por x e y.

$$mult1 = RefArray_{base} \times (32 - shift) \quad (1)$$

$$mult2 = RefArray_{base+1} \times shift \quad (2)$$

$$Pred_{x,y} = (mult1 + mult2 + 16) \gg 5 \quad (3)$$

3. ARQUITETURA DESENVOLVIDA

A arquitetura proposta opera a partir de um SB de tamanho 64x64 e funciona como uma busca em profundidade até o tamanho 4x4, passando por todas as possibilidades de particionamento de blocos, isso é feito pela unidade *Control*. A saída da arquitetura é uma linha ou coluna de cada modo angular por ciclo. A Figura 1 mostra o diagrama de blocos da arquitetura desenvolvida.

A *Filtering Unit* serve para implementar os três filtros disponíveis em paralelo, recebendo as amostras que estão armazenadas nos buffers *AboveRow* e *LeftCol*, os valores das amostras filtradas são utilizadas como entrada na *Upsampling Unit*, que faz o processo de duplicar a quantidade de amostras utilizadas como referência. Tanto a saída da *Filtering Unit* quanto a da *Upsampling Unit* são utilizadas como entrada para a *Filter Selector Unit*, que faz a escolha de qual filtro

será utilizado para cada ângulo, a escolha é feita por uma expressão booleana através de alguns parâmetros, como: tamanho do bloco, ângulo do modo e *filter type*. A *Mode Prediction Unit* (MPU) organiza todos os modos angulares conforme os cálculos necessários para realizar a predição de uma amostra. Para realizar a predição de uma única amostra, as equações (1-3) foram implementadas na arquitetura *Sample Prediction Unit* (SPU), 16 diferentes SPU's foram implementadas e instanciadas de acordo com a necessidade de cada MPU. Um total de 101 MPUs são implementadas na arquitetura, 54 para os modos normais e 47 para os modos que usam o *upsampling*. O número de SPU's em uma MPU varia de acordo com o ângulo do modo. Além disso, o ângulo do modo também define se uma linha ou coluna do bloco será predita.

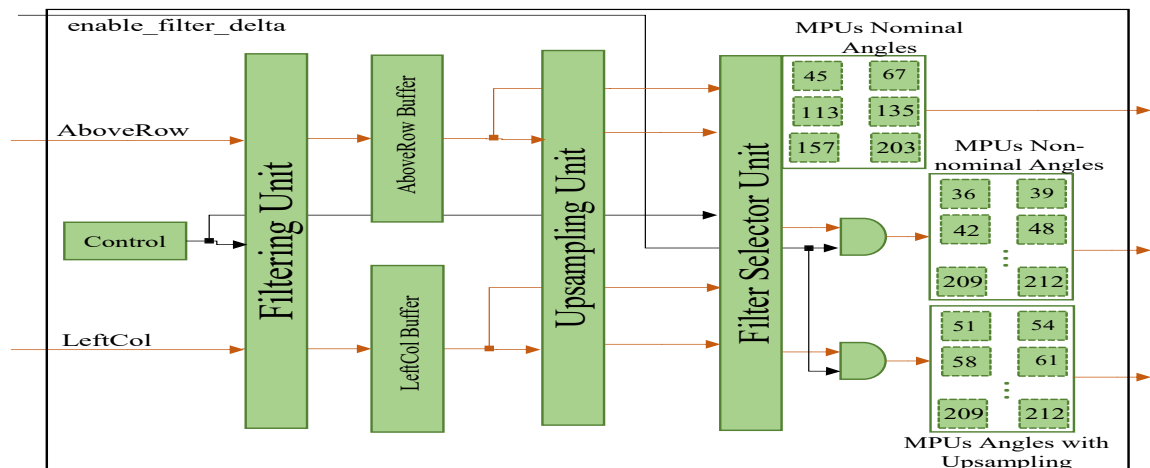


Figura 1. Diagrama de blocos da arquitetura proposta.

Para controlar a operação quando os modos HQ e LP são usados, um sinal externo denominado *enable_filter_delta* foi definido. Com base neste sinal, a estratégia do *operand isolation* foi utilizada. O *operand isolation* é aplicado na entrada dos módulos de predição para ângulos não nominais e modos que usam matrizes pós-*upsampling*, forçando-os a operar em um valor de entrada zerado quando o modo LP está habilitado, evitando chaveamento em unidades que não são necessárias, reduzindo assim a dissipação de energia da arquitetura.

4. RESULTADOS E DISCUSSÕES

A arquitetura proposta descrita na seção anterior foi sintetizada para a tecnologia TSMC 40nm usando a ferramenta Cadence Encounter RTL Compiler 14.02. A Tabela 1 mostra os resultados obtidos em termos de dissipação de potência (mW) e a área do circuito para cada modo de operação, considerando cada taxa de transferência alvo (fps) e a frequência necessária (MHz) para operação em tempo real.

Os resultados da Tabela 1 mostram que, como esperado, a dissipação de potência no modo HQ é maior que no modo LP para todas as frequências e em ambos os modos a potência aumenta conforme a frequência aumenta. É interessante notar que os valores de área não diferem tanto para as três primeiras frequências, além disso, a potência entre HD@60 e FHD@30 é muito baixa.

Em relação aos trabalhos que implementam a predição intra-quadro do AV1 em hardware, o trabalho (NETO et al, 2020) apresenta uma arquitetura que implementa apenas os oito modos nominais e a etapa de pré-processamento, nesse caso, nosso projeto obteve uma dissipação de potência 75% menor. Por

outro lado, a configuração HQ proposta neste trabalho requer uma dissipação de potência 35%, mas isso é esperado devido à inclusão dos 48 modos não nominais. O trabalho (CORREA et al, 2020) implementa todos os modos direcionais, mas sem a etapa de pré-processamento, isso permitiu aos autores realizar otimizações no processo de particionamento de blocos, reduzindo significativamente as frequências alvo e consequentemente reduzindo a potência dissipada e a área do circuito.

Tabela 1. Resultados de Síntese

Taxa de transferência	Freq. (MHz)	Área (Kgates)	Potência Total (mW)	
			Modo HQ	Modo LP
HD@30fps	96	1624,3	174,94	37,77
HD@60fps	191	1628,9	210,11	35,78
FHD@30fps	215	1639,3	212,11	36,26
FHD@60fps	476	1976,2	263,03	43,74
UHD@30fps	951	2890,5	469,87	70,39
UHD@60fps	1902	2794,0	631,76	82,76

5. CONCLUSÕES

Este trabalho apresentou uma solução de hardware para a predição intra-quadro do AV1 usando modos direcionais. A arquitetura é capaz de processar todas as possibilidades de particionamento de blocos para um SB de 64x64 pixels em tempo real, permitindo a utilização de dois modos de operação distintos, um priorizando a qualidade visual (*High Quality*) e outro priorizando o baixo consumo de energia (*Low Power*).

Os resultados da síntese mostram que o modo *Low Power* dissipa 84,4% menos potência do que o modo *High Quality*. Quando comparado com o trabalho de última geração encontrado na literatura, este projeto atinge uma redução de dissipação de energia de 75% no modo *Low Power*.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- WIEGAND, T.; SULLIVAN, G. J.; BJØNTEGAARD, G.; LUTHRA, A. Overview of the H.264/AVC video coding standard. **IEEE Transactions on Circuits and Systems for Video Technology**, [S.l.], v.13, n.7, p.560–576, 2003.
- SULLIVAN, G. J.; OHM, J.; HAN, W.; WIEGAND, T. Overview of the High Efficiency Video Coding (HEVC) Standard. **IEEE Transactions on Circuits and Systems for Video Technology**, [S.l.], v.22, n.12, p.1649–1668, 2012.
- CHEN, Y. et al. An Overview of Core Coding Tools in the AV1 Video Codec. In: PICTURE CODING SYMPOSIUM (PCS), 2018., 2018. **Anais**. . . , 2018. p.41–45.
- RIVAZ, P.; HAUGHTON, J. **AV1 Bitstream Decoding Process Specification**. Disponível em: <<https://aomedia.org/av1/specification/>>.
- MUKHERJEE, D. et al. The latest open-source video codec VP9 - An overview and preliminary results. In: PICTURE CODING SYMPOSIUM (PCS), 2013., 2013. **Anais** [...].
- NETO, L. et al. Directional Intra Frame Prediction Architecture with Edge Filter and Upsampling for AV1 Video Coding. In: SYMPOSIUM ON INTEGRATED CIRCUITS AND SYSTEMS DESIGN (SBCCI). 2020, Evento Online. **Anais** [...].
- CORREA, M. et al. ASIC Solution for the Directional Intra Prediction of the AV1 Encoder Targeting UHD 4K Videos. In: IEEE INTERNATIONAL SYMPOSIUM ON CIRCUITS AND SYSTEMS (ISCAS). 2020, Evento Online. **Anais** [...].
- ISO/IEC. **Information technology -- Coded representation of immersive media -- Part 3: Versatile video coding**. ISO/IEC DIS 23090-3. 2020.