

ACELERADOR DE HARDWARE MULTIVERSÃO DE BAIXO CONSUMO PARA OS FILTROS DE INTERPOLAÇÃO DO CODIFICADOR AV1

DAIANE FONSECA FREITAS¹; CLÁUDIO DINIZ²; MATEUS GRELLERT³;
GUILHERME CORRÊA⁴

¹Universidade Federal de Pelotas – dffreitas@inf.ufpel.edu.br

²Universidade Federal do Rio Grande do Sul – cmdiniz@inf.ufrgs.br

³Universidade Federal do Rio Grande do Sul – mateusgrellert@gmail.com

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

1. INTRODUÇÃO

Com a pandemia da Covid-19, as pessoas tornaram-se ainda mais dependentes de serviços conectados à internet. Um estudo aponta que em janeiro de 2022 92% dos usuários da internet acessaram conteúdos *online* através de dispositivos móveis (DATAREPORTAL, 2022). Outra tendência, é o aumento no consumo de mídias em plataformas de *streaming* de vídeo, bem como o uso de redes sociais dedicadas ao compartilhamento rápido de vídeo. Com base nestas tendências, especialistas estão desenvolvendo soluções eficientes para a compressão destes vídeos.

Entretanto, a maior eficiência de compressão esperada nos codificadores de vídeos emergentes é alcançada devido à adoção de novas ferramentas de codificação, que contribuem para o seu custo computacional adicional. No codec livre de *royalties* AV1, a etapa de predição inter-quadros é responsável por mais de 70% do tempo total de execução (BENDER et al., 2023). Além disso, a complexidade é aumentada com a adoção do esquema de filtragem adaptativa. Esse recurso é aplicado à parte fracionária da estimação de movimento (ME), a estimação de movimento fracionária (FME), bem como à compensação de movimento (MC). Nesse esquema, o AV1 permite o uso de 75 filtros do tipo FIR (Resposta Finita ao Impulso) diferentes para a geração de amostras de pixels de posição fracionária (HAN et al., 2021). Esses filtros estão agrupados em três famílias com 15 filtros cada, chamadas Regular, *Sharp* e *Smooth*, além de versões aproximadas das famílias Regular e *Smooth*. Logo, dado o elevado número de filtros no processo de interpolação do AV1 e a compreensão geral de que os filtros de interpolação desempenham um papel crítico no desempenho de um codificador de vídeo (MUKHERJEE et al., 2013), torna-se imperativo adotar estratégias para mitigar essa complexidade, como a concepção de aceleradores de hardware, e ainda requer cuidados em relação a eficiência energética.

Com base no exposto, esse trabalho propõe um acelerador de *hardware* multiversão para o módulo de interpolação de amostras do AV1, que incorpora técnicas de redução de potência e computação aproximada.

2. METODOLOGIA

A arquitetura do acelerador de *hardware* proposto neste trabalho é mostrada na Figura 1, e foi fragmentada em duas partes principais nomeadas *Multiversion Interpolation Core* (MVIC) e *Multiversion Interpolation Processor* (MVIP).

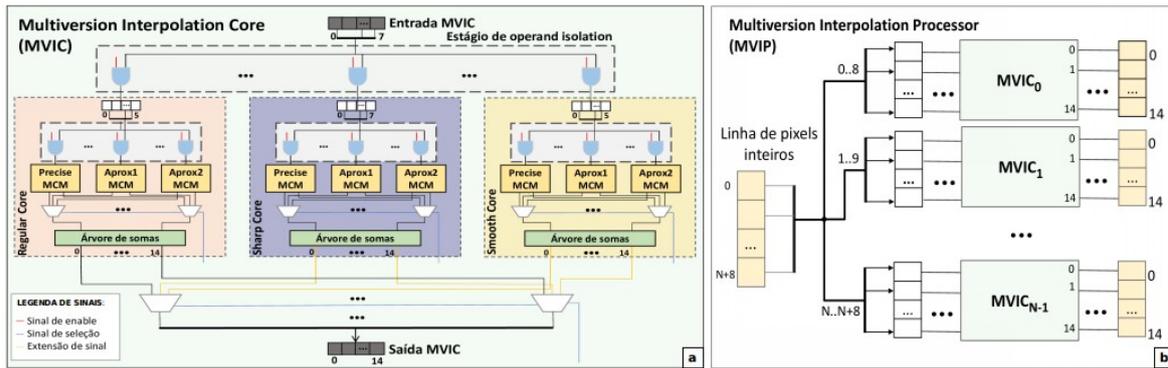


Figura 1: (a) Arquitetura MVIC. Para que cada *core* e as suas versões de precisão operem de forma independente, foram adicionados estágios de *operand isolation*. Os blocos em cinza representam registradores; e (b) MVIP, que recebe como entrada uma linha do bloco mais uma zona de preenchimento (N+8).

A MVIC apresentada na Figura 1(a) consiste em três *cores* que implementam os principais filtros de interpolação usados no ME/MC do AV1, chamados de *Regular Core*, *Sharp Core* e *Smooth Core*. Esses filtros foram descritos sem utilizar somadores a partir do algoritmo Hcub MCM (VORONENKO; PÜSCHEL, 2007). Na MVIC, além das versões exatas dos filtros também foram implementadas versões aproximadas para redução do custo de codificação na ME, com diferentes níveis de precisão. O módulo *Regular Core* consiste em três versões diferentes de precisão para cada conjunto de coeficientes da família Regular. A primeira versão, denominada *Precise MCM*, implementa os coeficientes precisos dos filtros AV1 Regular, que têm 6 taps, sendo necessários 6 MCMs. A segunda versão, é a primeira proposta aproximada denominada *Approx1 MCM*, projetada com base na versão *Precise MCM*. Nela, novos coeficientes múltiplos de 2 e mais próximos possíveis dos coeficientes originais foram gerados, com no máximo um somador no caminho crítico do MCM. Por fim, a terceira versão de MCMs, denominada *Approx2 MCM*, é uma aproximação sem somadores, que usa apenas deslocamentos para derivar coeficientes aproximados. No módulo, os MCMs descrevem os coeficientes dos filtros no modo transposto. As árvores de somas e deslocamentos adicionam as operações aritméticas para uma descrição completa das equações do filtro de interpolação. As três versões dos MCMs de cada família de filtros compartilham a mesma árvore de somas e deslocamentos, onde nenhuma aproximação foi aplicada. A arquitetura MVIC tem 8 entradas, para suportar o maior número de taps apresentados nos filtros de interpolação do AV1. A seleção de qual família de filtros será escolhida e a versão do filtro usada depende apenas das entradas primárias. Para reduzir a dissipação de potência, a técnica *operand isolation* foi empregada para diminuir a dissipação de potência dinâmica, sendo representado na Figura 1(a) pelos operadores booleanos AND incluídos antes de cada *core* da família e cada versão de filtro. Por fim, a saída da MVIC corresponde a 15 *pixels* em posições de *subpixel* obtidas no processo de interpolação. Ao final, os sinais de saída dos 15 filtros são obtidos usando combinações de árvores de somadores e deslocamentos. Os módulos *Sharp Core* e *Smooth Core* seguem a mesma metodologia para sua implementação.

A Figura 1(b) exibe a estrutura *top-level*, denominada MVIP, que foi projetada para processar blocos 4×4 (logo N=4), o menor tamanho de bloco definido no AV1. Assim, a entrada da arquitetura é capaz de suportar um paralelismo de 12 amostras em posições de *pixel* inteiro, considerando a zona de

preenchimento (*padding-left* e *padding-right*), isto é, $N+8$. Na saída do MVIP, são obtidas as amostras em posições de *subpixel* necessárias para a interpolação de um bloco de 4×4 . Para possibilitar esse processamento, a MVIP é formada por instâncias do MVIC. A interpolação de um bloco completo é realizada em 51 ciclos de *clock* para ME, e 16 ciclos de *clock*. Ambos os cálculos consideram que o MVIP é capaz de interpolar uma linha do bloco acrescida dos *pixels* da zona de preenchimento em cada ciclo de *clock*.

3. RESULTADOS E DISCUSSÃO

O acelerador de *hardware* apresentado nesse trabalho foi descrito em VHDL (*Very High Speed Integrated Circuits Hardware Description Language*), sintetizado para tecnologia de célula padrão TSMC de 40nm, com 1,1V e 25°C usando a ferramenta *Cadence RTL Compilation tool*, e contagem de portas com base em NANDs 2. A potência foi estimada a partir de entradas reais. Os resultados de síntese apresentados na 6ª coluna da Tabela 1, mostram que a uma frequência máxima de 833,3MHz, a MVIP atinge um *throughput* de 2.656,14 amostras/segundo em um cenário ME. Ou seja, a arquitetura MVIP pode interpolar vídeos com resolução 8K a 50 fps considerando o cenário MC.

Também comparamos os resultados obtidos com os trabalhos relacionados. A Tabela 1 mostra que a MVIP requer uma área maior do que todos os trabalhos relacionados, o que é esperado, pois a proposta implementa várias versões de MCMs para os filtros de interpolação do AV1, bem como estágios de *operand isolation*. Quando comparado com Domanski et al., (2019), a MVIP opera em uma frequência máxima 2,98× maior, dissipa e 49,21% menos potência. Em comparação com o hardware aproximado em Domanski et al., (2021), a MVIP é capaz de processar vídeo de resolução mais alta, mas dissipa 1,54× mais potência. Esse aumento é justificado, pois a MVIP implementa filtros precisos e aproximados. Em comparação com Kolodziejski et al. (2022), a MVIP dissipa menos potência, e alcança um *throughput* mais alto. Por fim, quando comparado aos filtros precisos em Freitas et al., (2022), a MVIP dissipa 19,2% menos potência, mas uma frequência máxima menor, o que leva a um *throughput* e taxa de quadros menor.

Tabela 1: Resultados de síntese e comparação com trabalhos relacionados.

Trabalhos Correlatos	Domanski (2019)	Domanski (2021)	Kolodziejski (2022)	Freitas (2022)	MVIP
Famílias de Filtros	Todos+Bil. Preciso	Todos+Bil. Aprox.	Todos+Bil. Aprox.	3 famílias Preciso	3 famílias Prec./Aprox.
Tec. (nm)	40	40	40	40	40
Freq. (MHz)	279,9	686,0	650,0	1000 833,3	833,3
Portas (K)	141,10	72,64	102,62	324,79	473,01
Potência(mW)	81,31	26,79	82,91	51,15	41,30
Resol. MC	4320p @30qps	4320p @30qps	4320p @30qps	4320p @60qps	4320p @50qps
Through. (Mam.s)	-	-	-	3.187,5	2.656,14

4. CONCLUSÕES

Este trabalho apresentou um acelerador de hardware multiversão para os filtros de interpolação com foco na ME/MC do AV1 chamada MVIP, que explora diferentes níveis de aproximação para a descrição dos filtros. A fim de reduzir a potência dinâmica dissipada, foi utilizada a técnica *operand isolation*. Assim, arquitetura MIP é capaz de processar vídeos com resolução de 8K a 50 fps, com a menor dissipação de potência quando comparada com trabalhos relacionados.

Logo, em trabalhos futuros propomos uma análise do impacto na eficiência de codificação ao utilizar versões aproximadas dos filtros de interpolação AV1, bem como a exploração de um modelo preditivo baseado em aprendizado de máquina para estimar a melhor combinação de filtros de interpolação para cada cenário.

5. REFERÊNCIAS BIBLIOGRÁFICAS

BENDER, I. et al. Complexity and compression efficiency analysis of libaom AV1 vídeo codec. **Journal of Real-Time Image Processing**, [S.l.], v.20, n.3, p.50, 2023.

DATAREPORTAL. "**Digital 2022: time Spent Using Connected Tech Continues To Rise**". Disponível em: <<https://datareportal.com/reports/digital-2022-time-spent-with-connected-tech>>.

DOMANSKI, R. et al. High-throughput multifilter interpolation architecture for AV1 motion compensation. **IEEE Transactions on Circuits and Systems II: Express Briefs**, v. 66, n. 5, p. 883-887, 2019.

DOMANSKI, R. et al. Low-power and high-throughput approximated architecture for AV1 FME interpolation. In: **2021 IEEE International Symposium on Circuits and Systems (ISCAS)**. IEEE, 2021. p. 1-5.

FREITAS, D. et al. High-Throughput Multifilter VLSI Design for the AV1 Fractional Motion Estimation. In: **SBC/SBMICRO/IEEE/ACM SYMPOSIUM ON INTEGRATED CIRCUITS AND SYSTEMS DESIGN (SBCCI)**, 2022., 2022. **Anais**. . . [S.l.: s.n.], 2022. p.1–6.

HAN, J. et al. A technical overview of AV1. **Proceedings of the IEEE**, v. 109, n. 9, p. 1435-1462, 2021.

KOŁODZIEJSKI, W. et al. Ultra-High Definition AV1 FME Interpolation Architectures Exploring Approximate Computing. **Journal of Integrated Circuits and Systems**, [S.l.], v.17, n.2, p.1–12, 2022.

MUKHERJEE, D. et al. A technical overview of VP9—the latest open-source video codec. In: **SMPTE 2013 Annual Technical Conference & Exhibition**. SMPTE, 2013. p. 1-17.

VORONENKO, Y.; PÜSCHEL, M.. Multiplierless multiple constant multiplication. **ACM Transactions on Algorithms (TALG)**, v. 3, n. 2, p. 11-es, 2007.