



SOMADOR DE PREFIXO PARALELO APROXIMADO

LEONARDO ANTONIETTI FERREIRA¹; MORGANA M. A. DA ROSA¹; EDUARDO ANTONIO CESAR DA COSTA²; RAFAEL IANKOWSKI SOARES¹

¹ Universidade Federal de Pelotas – laferreira@inf.ufpel.edu.br; mmarosa@inf.ufpel.edu.br, rafael.soares@inf.ufpel.edu.br

² Universidade Católica de Pelotas – eduardo.costa@ucpel.edu.br

1. INTRODUÇÃO

O declínio da escala clássica ou escala de Dennard (BOHR, 2007) levou a uma desaceleração no crescimento da frequência do clock e a uma redução na voltagem a cada geração de tecnologia, impulsionando a transição para a era de múltiplos núcleos. No entanto, até mesmo esse novo paradigma de escala está ameaçado pela incapacidade de encaixar mais núcleos em um orçamento de energia limitado. Como resultado, os desenvolvedores são forçados a buscar novas fontes de eficiência computacional, como a computação aproximada (AxC).

AxC é estabelecida como uma alternativa de design para aumentar a eficiência de sistemas de computação, aproveitando a resiliência intrínseca a erros de várias aplicações, como segurança de hardware, processamento digital de sinais, imagem e vídeo, eletrônicos de consumo, e aprendizado de máquina (ML). AxC introduz a precisão/qualidade como uma nova dimensão para o *trade-off* em termos de otimização de design, o que pode levar a reduções significativas na área de hardware VLSI, atraso e dissipação de potência, reduzindo o consumo de energia. Impulsionado pelo alto potencial de economia de energia, o projeto de arquiteturas de aritmética aproximada tem despertado interesse significativo de pesquisa nos dias de hoje (ROSA, 2023).

No nível de hardware personalizado, a AxC tem se concentrado principalmente em somadores (ROSA, 2023) pois eles são os componentes principais de todos os cálculos e várias aplicações tolerantes a erros (ZERVAKIS, 2021). Pode-se destacar processamento de sinal, imagem e vídeo, visão computacional e aprendizado de máquina supervisionado entre as aplicações tolerantes a erros e intensivas em computação. Essas aplicações podem usar diferentes estratégias de AxC, promovendo otimizações durante sua execução de acordo com a métrica de erro. Portanto, a otimização de unidades de somadores é um tópico importante de pesquisa para aumentar a eficiência energética especialmente para circuitos integrados de aplicação específica (ASICs).

Muitos trabalhos têm proposto arquiteturas para somadores aproximados (AxAs) (LEE, 2021; ROSA, 2023). Os AxAs eficientes em energia realizam a aproximação dos bits menos significativos (LSBs) para os bits mais significativos (MSB). Os somadores de prefixo paralelo (PPAs) estão entre as unidades de soma mais rápidas e eficientes em área, pois usam a redução logarítmica para calcular o caminho de propagação de *carry*, assim otimizando o caminho crítico da soma. Este trabalho apresenta somadores paralelos de prefixo aproximados (AxPPA) que combinam aproximações do LSB para os MSB e rápida propagação de *carry*, fornecendo uma solução híbrida de somadores rápidos e eficientes em energia. Nossa ideia de combinar velocidade e eficiência energética por meio de aproximações nos PPAs mostra uma nova Frente de Pareto com redução de área e energia para o mesmo nível de aproximação. A ideia-chave por trás de nossa proposta é aproximar a lógica de propagação (P) e geração (G) de *carry* dos



operadores de prefixo (POs). Para demonstrar a proposta de operadores de prefixo aproximados (AxPOs), implementamos e avaliamos quatro AxPPAs usando as arquiteturas de Brent Kung (AxPPA-BK), Kogge Stone (AxPPA-KS), Ladner Fischer (AxPPA-LF) e Sklansky (AxPPA-SK).

2. METODOLOGIA

Para realizar uma avaliação orientada à precisão do AxPPA, utilizou-se como métrica de erro a Distância de Erro Relativa Média (MRED). Realizaram-se experimentos comparativos de precisão com outros somadores aproximados (AxAs) apresentados na literatura, como COPY, ETA-I, LOA, Trunc, e as versões de VLSPPAs (ESPOSITO, 2015; ESPOSITO, 2016) do inglês *Variable Latency Speculative Parallel Prefix Adders*. O VLSPPA são versões da literatura de PPAs com computação de prefixo aproximada, como a computação de prefixo dá a natureza do PPA, foram propostas seis topologias de VLSPPA a base de Carry-Increment (VLSPPA-CI), Han Carlson (VLSPPA-HC), Brent Kung (VLSPPA-BK), Kogge Stone (VLSPPA-KS), Ladner Fischer (VLSPPA-LF) e Sklansky (VLSPPA-SK).

Obteu-se os resultados de precisão no processo de co-simulação entre os softwares MATLAB e Modelsim. Calculamos os valores de MRED para 1,000,000 de entradas pseudoaleatórias de 16 bits. Para os resultados baseados em ASIC as arquiteturas foram descritas em VHDL e sintetizadas usando a ferramenta de síntese Genus da Cadence na frequência de 200 MHz. As sínteses utilizaram a biblioteca de células padrão comercial ST 65nm de baixa potência com tensão de alimentação de 1,25V. Utilizou-se a ferramenta Cadence Incisive para simular todas as listas de redes considerando o arquivo SDF para atrasos precisos na propagação do sinal e falhas temporais. A simulação gera um arquivo *Toggle Count Format* (TCF), carregado na ferramenta de síntese para extração de energia realista. A metodologia de estimativa de potência utiliza a ferramenta de síntese Genus no modo PLE para gerar a netlist em nível de porta Verilog e o formato SDF.

3. RESULTADOS E DISCUSSÃO

A Fig. 1-(a) mostra o *trade-off* entre economia de energia e qualidade, e a Fig. 1-(b) mostra o *trade-off* entre economia de área e qualidade para a estrutura proposta do AxPPA e outros AxAs da literatura, com variações no passo de aproximação, com $K=1$ à 16 bits. A Frente de Pareto na Fig.1 resume o *trade-off* entre aproximação e qualidade MRED em perspectiva com os resultados de síntese de energia e área.

O AxPPA resulta em uma economia de área de até 60% em comparação com um somador exato. Como a parte aproximada do AxAs usa menos ou nenhuma porta lógica, requer menos área de circuito do que o somador exato. Além disso, o AxPPA alcança economia de energia de até 80% em comparação com o somador exato. Existem variações significativas na área de circuito dos AxAs, e isso se deve às diferenças em sua composição lógica. Como resultado, os AxAs dissipam menos energia do que o somador exato para todas as propostas de AxAs e AxPPAs, como visto na Fig. 1. Por outro lado, os AxAs implicam em um erro no resultado da soma. É importante observar que, à medida que a aproximação aumenta, o consumo de energia diminui. O VLSPPA tem o maior consumo de energia entre os AxAs: AxPPA, COPY, ETA-I e LOA.

O AxPPA-LF com $K=16$ alcança as seguintes reduções nas métricas de design em comparação com o somador exato: (i) economia de energia de 85,43%, (ii) economia de área de 60,05%; com $K=1$: (a) economia de energia de 72,63%,

(b) economia de área de 40,41% c) pequeno erro de $6,5 \times 10^{-4}$ considerando a métrica MRED. O AxPPA-LF com K=12 alcança as seguintes reduções nas métricas de design em comparação ao COPY e Trunc: (i) economia de energia de 11,34%, (ii) economia de área de 2,29%, (iii) redução de erro com MRED=36. Por outro lado, como a parte aproximada do COPY e Trunc não possui portas lógicas, de K=13 a 16 indica pequenas economias de área em comparação com o AxPPA. Note que, para K=16, o Trunc tem economia de área de 7,76% em comparação com o AxPPA também com K=16. É importante destacar que o AxPPA-LF oferece os melhor *trade-off* entre resiliência a erros e economia de área e energia.

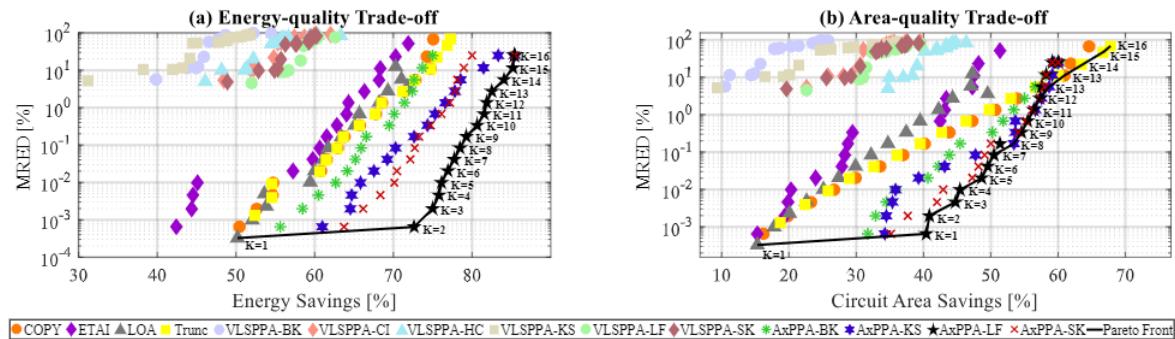


Fig. 1: MRED para o AxPPA em relação aos AxAs da literatura: a) Economia de energia versus compensação de qualidade, b) Economia de área de circuito versus compensação de qualidade. O *baseline* empregado para o cálculo da economia é o somador exato selecionado automaticamente pelo a ferramenta de síntese.

A Fig. 2 mostra a taxa de erro média para a previsão de *carry*. Essa taxa é cerca de 1,67 vezes maior para os somadores da literatutura com 8, 14 e 16 bits em comparação com o AxPPA. A falta de previsão de *carry* nas versões COPY, ETA-I, LOA, Trunc e VLSPPA é evidenciada por uma taxa de erro de previsão de *carry* superior a 60% para K = 16. Observe que, a entrada de *carry* do somador Trunc é definida como '0', então, para um somador com entrada de 16 bits e K = 16, a taxa de erro é igual a 100% (todos os bits em nível lógico '0' e sem propagação de *carry*). Os AxAs ETA-I e LOA incluem a especulação de *carry* baseada em AND, o que reduz a taxa de erro para cerca de 25% em comparação com a Trunc. Observe que, o VLSPPA-SK se comporta de forma semelhante ao Trunc em termos de taxa de erro média para a previsão de *carry*. O AxPPA é extremamente resistente a erros. Observe que, para K = 8, a técnica do AxPPA de previsão de *carry* tem uma taxa de erro de 0,27%, ou seja, para 1,000,000 de entradas apenas 27 destas irão propagar *carry* com valor errado. Para K = 14, a taxa de erro para o AxPPA é igual a 14,13%, enquanto para K = 16, a taxa de erro é de 23,91%.

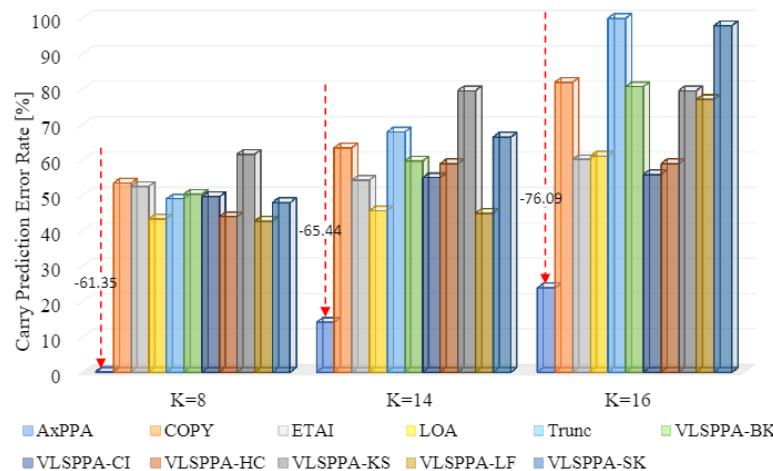


Fig. 2: Comparação das taxas de erro de previsão de transporte de AxAs com K = 8, 14 e 16.

4. CONCLUSÕES

O AxPPA é uma nova arquitetura de somador de prefixo paralelo aproximado. A técnica de aproximação desenvolvida calcula as saídas dos somadores prefixados. Avalia-se as propostas do AxPPA de forma independente de aplicação com K variando de 0 à 16 bits de aproximação. O AxPPA supera os AxAs energeticamente eficientes na literatura considerando a métrica MRED bem conhecida na literatura. A proposta do AxPPA mostra as melhores economias para resultados de síntese em comparação com os AxAs consolidados na literatura. O AxPPA alcança excelentes padrões de qualidade e suporta níveis de aproximação mais altos. Em particular, o AxPPA-LF oferece a melhor troca entre aproximação e qualidade, resultando em economias significativas de energia e área.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- BOHR, M. A 30 year retrospective on Dennard's MOSFET scaling paper. IEEE SolidState Circuits Society Newsletter, v.12, n.1, p.11–13, 2007.
- ROSA, M. M. A. d. et al. AxPPA: Approximate Parallel Prefix Adders. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, v.31, n.1, p.17–28, 2023.
- ZERVAKIS, G. et al. Approximate Computing for ML: State-of-the-art, Challenges and Visions. In: ASIA AND SOUTH PACIFIC DESIGN AUTOMATION CONFERENCE (ASP-DAC), 2021. p.189–196.
- J. Lee, H. Seo, H. Seok, and Y. Kim, “A Novel Approximate Adder Design Using Error Reduced Carry Prediction and Constant Truncation,” IEEE Access, vol. 9, pp. 119 939–119 953, 2021.
- ESPOSITO, D.; CARO, D. D.; STROLLO, A. G. M. Variable Latency Speculative Parallel Prefix Adders for Unsigned and Signed Operands. IEEE Transactions on Circuits and Systems I: Regular Papers, v.63, n.8, p.1200–1209, Aug 2016.
- ESPOSITO, D. et al. Variable Latency Speculative Han-Carlson Adder. IEEE Transactions on Circuits and Systems I: Regular Papers, v.62, n.5, p.1353–1361, 2015.