

GEM5X: Uma Extensão do GEM5 para Computação Aproximada em Nível de Instrução

Hytalo Bassi¹, Pedro e Silva¹, Jailton Nunes Filho¹, Eduardo Paes¹,
Jhullya Martins¹, Daniela Catelan¹

¹ Faculdade de Computação – Universidade Federal de Mato Grosso do Sul (FACOM - UFMS)
Av. Costa e Silva, s/n – Campo Grande – MS – Brasil

{hytalo_bassi, pedro.candia, jailton_nunes_filho}@ufms.br

{eduardo_paes, jhullya.e.s.martins, daniela.catelan}@ufms.br

Abstract. *This work implements approximate arithmetic instructions in RISC-V on GEM5, enabling cycle-accurate microarchitectural evaluation of performance, power and energy. A Design Space Exploration (DSE) combining NSGA-II with Pearson correlation found configurations with 28% energy and 27% time reductions in neural network inference, preserving behavioral fidelity and showing application-dependent effectiveness.*

Resumo. *Este trabalho implementa instruções aritméticas aproximadas em RISC-V no GEM5, viabilizando avaliação microarquitetural com métricas de ciclo preciso. A DSE que combina NSGA-II e Correlação de Pearson identificou configurações com reduções de 28% em energia e 27% em tempo na inferência de redes neurais, preservando a fidelidade comportamental e demonstrando efetividade dependente da aplicação.*

1. Introdução

A miniaturização dos transistores trouxe limitações como o dark silicon, exigindo equilíbrio entre desempenho e eficiência. A Computação Aproximada (CA) admite erros controlados em aplicações tolerantes, reduzindo energia, área e tempo; instruções aproximadas operam com precisão reduzida e apresentam resultados promissores [Catelan 2025].

A arquitetura RISC-V [RISC-V 2022], aberta e extensível, é ambiente ideal para experimentação. Este trabalho implementa instruções aproximadas em RISC-V no GEM5 [Binkert 2011], viabilizando avaliação microarquitetural com métricas de ciclo preciso.

Catelan propôs extensões aproximadas para RISC-V no simulador SPIKE [SPIKE 2019]. Este trabalho avança avaliando microarquiteturalmente no GEM5, combinando métricas de ciclo preciso com DSE baseada em NSGA-II [Verma 2021] estendido por Correlação de Pearson.

2. Metodologia

As instruções aproximadas [Catelan 2025] foram portadas do SPIKE para o GEM5, aqui denominado GEM5X, com correção funcional validada comparando saídas ao SPIKE. Os experimentos usaram O3CPU (out-of-order, 64 bits, 1GHz) em modo SE, 512MiB de RAM e caches L1/L2; as estimativas de energia empregaram o modelo de Catelan derivado do McPAT. O GEM5X em modo SE é determinístico, executando cada configuração uma vez.

Foram avaliadas quatro aplicações: **EMA** [Livera 2011], **Série de Leibniz** [Burden 2016], **Inferência de Redes Neurais** [Goodfellow et al. 2016] e

FFT [Cooley and Tukey 1965]. A qualidade funcional foi medida pelo erro histórico médio (EHM), erro relativo (ER) e erro absoluto (EA); apresentam-se apenas os resultados da Inferência de Redes Neurais.

A DSE¹ empregou **Sweep**, **NSGA-II** (população 24, 20 gerações) e **NSGA-II com Correlação de Pearson**, conforme sintetiza a Figura 1.

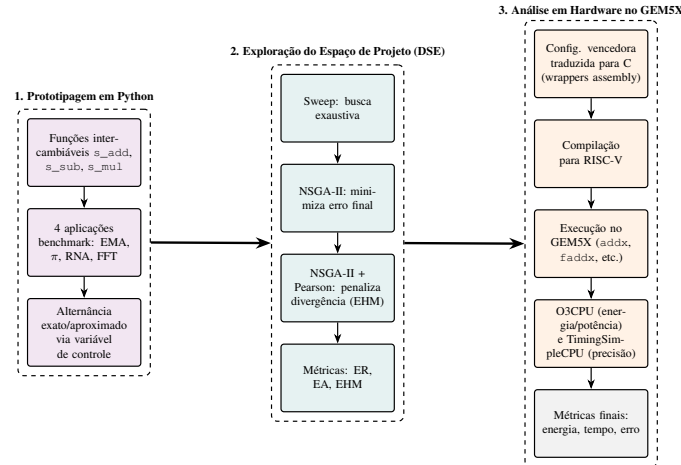


Figura 1. Fluxo de trabalho da metodologia: da prototipagem em Python à análise em hardware no GEM5X.

3. Resultados e Discussões

A Tabela 1 mostra o Candidato 5 reduzindo energia em 28% e tempo em 27% em relação à implementação exata. Embora seu ER final seja de 88%, o EHM foi próximo de zero: pela amplitude da amostra, esse ER torna-se desprezível, resultando em $\text{EHM} \approx 0$. A Figura 2 mostra trajetória sobreposta à execução exata, evidenciando que Correlação de Pearson e EHM são essenciais para identificar configurações que preservam a dinâmica em aplicações iterativas.

Os identificadores dos candidatos seguem a numeração original das configurações geradas na exploração do espaço de projeto. A Tabela 1 traz apenas as cinco configurações selecionadas para comparação detalhada; as demais não atenderam simultaneamente aos critérios de interesse definidos para a análise.

Tabela 1. Comparação entre a implementação exata e as configurações aproximadas selecionadas pelo NSGA-II.

Implementação	Ops. Aprox.	Energia (mJ)	Tempo (ms)	Erro Relativo (%)	Erro Histórico Médio
Exata	–	2,138	0,423	–	–
Aprox. Candidato 2	8 (h1_*, h2_*, out_*)	2,155	0,427	43,3%	0,77
Aprox. Candidato 3	8 (h1_*, h2_*, out_*, accum_add)	1,536	0,306	105%	1,1
Aprox. Candidato 4	7 (h1_*, h2_*, out_add)	2,169	0,429	24,7%	1,7
Aprox. Candidato 5	1 (accum_add)	1,533	0,305	88%	0
Aprox. Candidato 9	3 (h1_mul_x1, h1_add, h2_mul_x2)	2,225	0,440	3,21%	0,49

¹Código-fonte e artefatos experimentais disponíveis em: <https://github.com/hytalo-bassi/Approximate-Instructions>

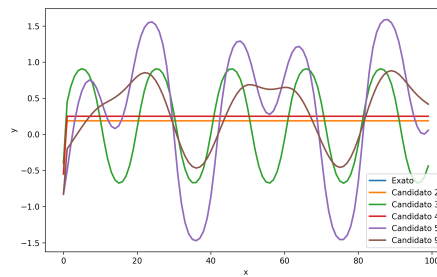


Figura 2. Inferência de Redes Neurais

4. Conclusão

Este trabalho demonstrou a avaliação microarquitetural de instruções aproximadas em RISC-V no GEM5X, alcançando reduções de 28% em energia e 27% em tempo na inferência de redes neurais. A Correlação de Pearson na DSE foi determinante para selecionar configurações que preservam a dinâmica temporal.

Trabalhos futuros incluem avaliação em visão computacional, caracterização de perfis algorítmicos, técnicas de compensação adaptativa e extensões de compiladores.

O presente trabalho foi realizado com apoio da Universidade Federal de Mato Grosso do Sul - UFMS/MEC - Brasil.

Referências

- Binkert, N. e. a. (2011). The gem5 simulator. *SIGARCH Comput. Archit. News*, 39(2):1–7.
- Burden, A. e. a. (2016). *Numerical Analysis, 10th ed.* Cengage Learning.
- Catelan, D. L. (2025). Approximate computing: Contributions to the design of arithmetic circuits and instruction-set architectures. *Repositório Institucional da UFMS*.
- Cooley, J. W. and Tukey, J. W. (1965). An algorithm for the machine calculation of complex fourier series. *Mathematics of Computation*, 19(90):297–301.
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press. <http://www.deeplearningbook.org>.
- Livera, A. M. e. a. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association*, 106(496):1513–1527.
- RISC-V (2022). Risc-v toolchain. Available at: <https://github.com/riscv-collab/riscv-gnu-toolchain>. Accessed: June/2025.
- SPIKE (2019). Spike risc-v isa simulator. Available at: <https://github.com/riscv-software-src/riscv-isa-sim>. Accessed: June/2025.
- Verma, S. e. a. (2021). A comprehensive review on nsga-ii for multi-objective combinatorial optimization problems. *IEEE Access*, 9:57757–57791.