

Programação Multicore

Contexto e Motivação

Demetrios A. M. Coutinho - NADIC/IFRN

Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte

12-13 de maio de 2023



https://github.com/nadic-ifrn/multicore_programming/

Agenda

Part 1 - **Contexto e Motivação**

Part 2 - **Paralelismo de Hardware e Software**

Part 3 - **Computação Paralela com OpenMP**

Part 4 - **Computação Paralela em CUDA**

Next

1 Sobre mim

2 Contexto

3 Motivação

4 Aplicações

5 Resumo



IFRN



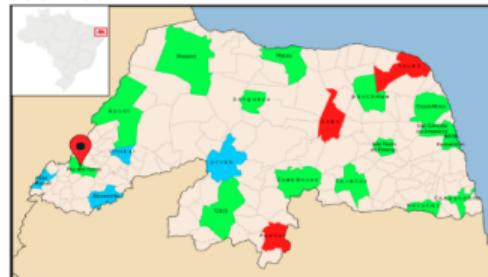
70 cursos
técnicos de nível
médio

36 cursos de
graduação



22 campi divididos
totalizando 19
municípios

Cerca de
41.000 discentes



Sobre mim

📄 Pequeno Resumo:

- 👤 Doutor em Engenharia de Elétrica e Computação pela UFRN.
- 🎓 Desde 2012, professor do IFRN e Coordenador do curso superior TADS por 4 anos.
- 🏛️ Doutorado Sanduíche na universidade de Bristol na área de software energeticamente eficiente no Reino Unido em 2021.

📘 Meus interesses:

- 📊 Análise e visualização de dados.
- 🔧 Machine Learning.
- 🖥️ Programação paralela e suas aplicações.
- ⚡ Computação de alto desempenho e eficiência energética.

🌐 Onde me encontrar?

- 💻 <https://gitlab.com/demetriosamc>
- /github <https://github.com/DemetriosCoutinho>.
- LinkedIn: <https://www.linkedin.com/in/demetrios-coutinho/>.
- 📄 Lattes: <http://lattes.cnpq.br/6749444529357582>.



Next

1 Sobre mim

2 Contexto

3 Motivação

4 Aplicações

5 Resumo



Paralelismo, Um conceito Familiar



Paralelismo, Um conceito Familiar



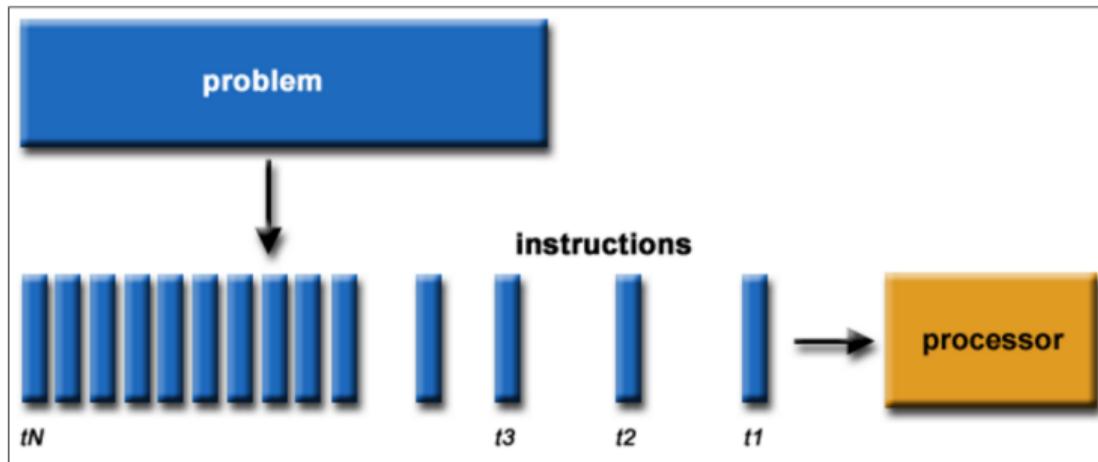
Embora familiares, essas formas de paralelismo são diferentes.

Paralelismo na computação



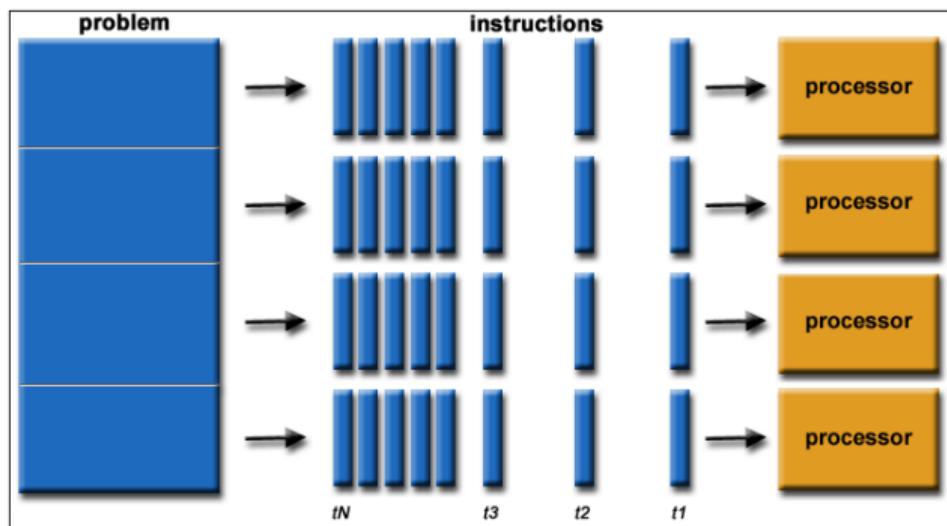
Paralelismo na computação

Figure: Problema Serial.



Paralelismo na computação

Figure: Problema Paralelo.



Next

1 Sobre mim

2 Contexto

3 Motivação

4 Aplicações

5 Resumo



O que está acontecendo na computação

Operação de software é um conceito bastante recente

Por décadas, o ideal era único núcleo, máxima frequência.

- ⌚ Sistemas de único núcleo eram o padrão.
 - ≡ Sistemas paralelos negados pela Lei de Amdahl.
 - ↗ Crescimento exponencial de desempenho:
 - ▶ Lei de Moore → frequências de operação mais altas.
 - ▶ Melhor ILP, caches, out-of-order, hw threads.



O que está acontecendo na computação

 Energia foi mantida em níveis aceitáveis e controlados

 Densidade de energia aumentada por

- ▶ Frequências de operação mais altas;
- ▶ Lei de Moore: transistores menores.

 Dissipação de energia (e calor) controlada por

- ▶ Avanços em sistemas de refrigeração → aumento na potência geral do sistema.
- ▶ Lei de Moore: transistores menores → menos dissipação de energia por transistor.

O que está acontecendo na computação?

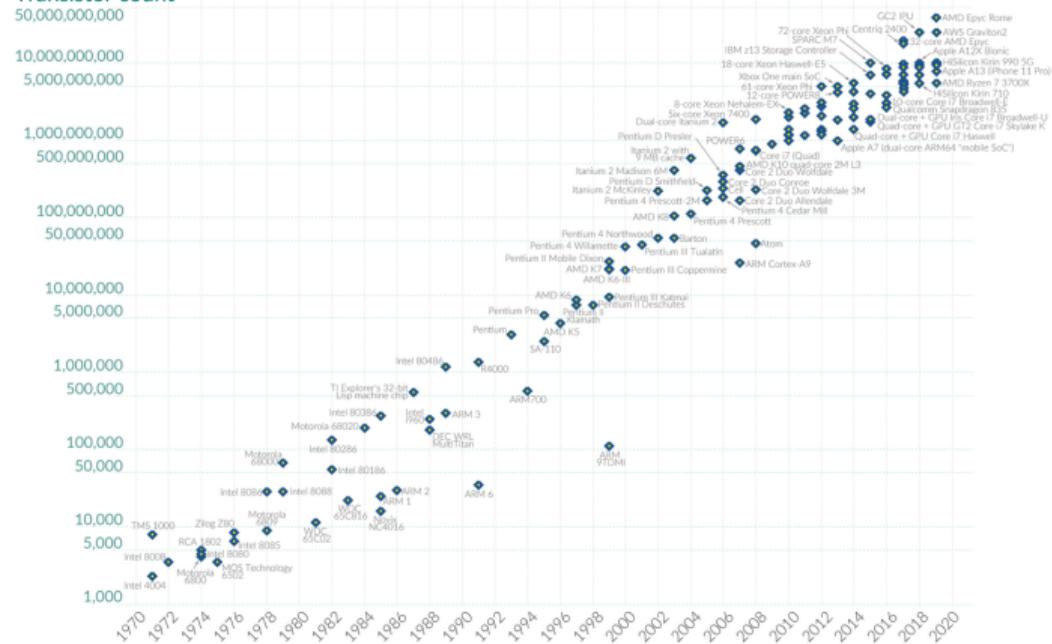
Figure: Lei de Moore criada em 1965.

Moore's Law: The number of transistors on microchips doubles every two years

Moore's law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years. This advancement is important for other aspects of technological progress in computing – such as processing speed or the price of computers.

Our World
in Data

Transistor count



Lei de Moore

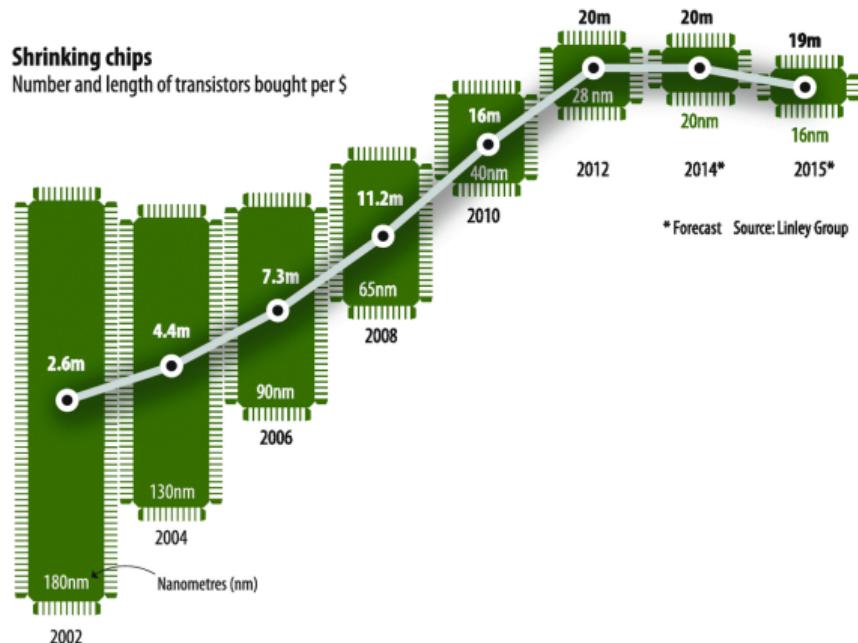
diz que a cada 1 ano e meio a quantidade de transistores dobraria.

O que está acontecendo na computação?

Shrink Chips

Assumia-se que quanto mais transistores em um chip maior desempenho era adquirido.

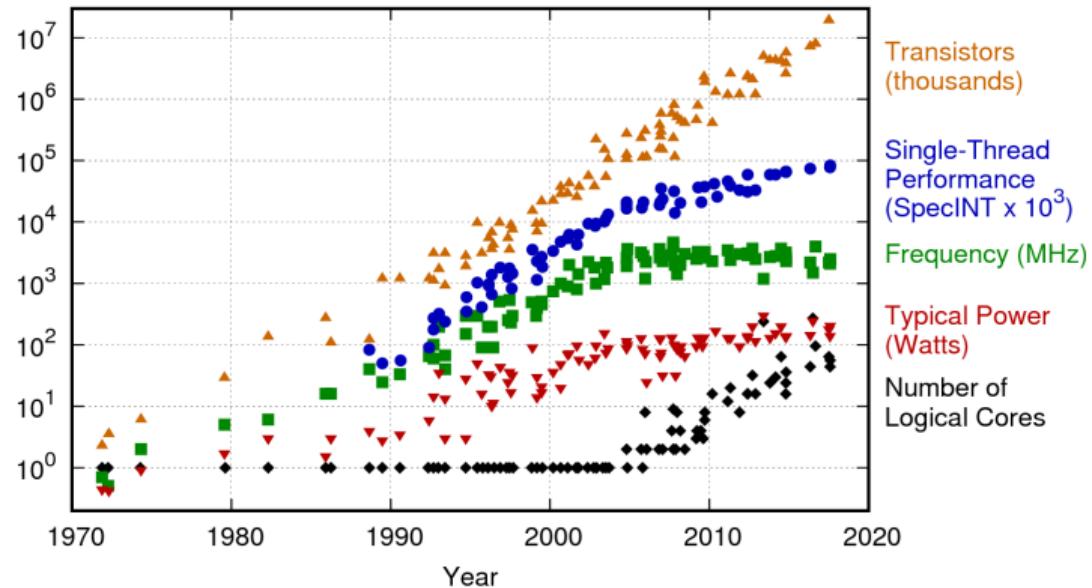
Figure: Diminuição do tamanho dos chips.



O que está acontecendo na computação?

Figure: Historical trends in CPU performance.

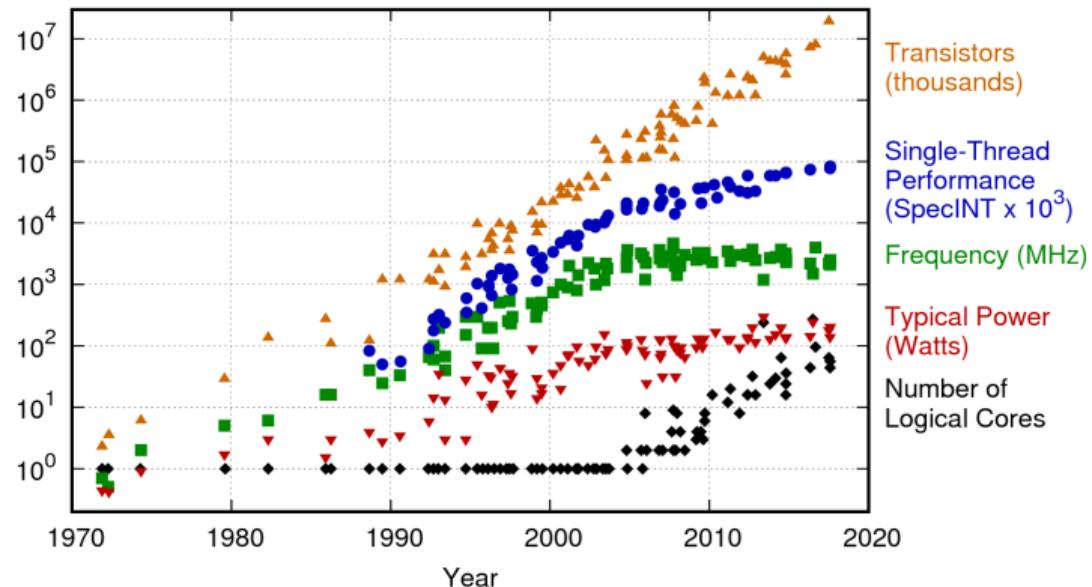
42 Years of Microprocessor Trend Data



O que está acontecendo na computação?

Figure: Historical trends in CPU performance.

42 Years of Microprocessor Trend Data



Em meados dos anos 2000, a indústria percebeu que **single-core era impraticável**

💡 Em 2030, 1 cm² de silício equivaleria a 1 cm² da superfície do sol.

O que está acontecendo com a computação

☰ A Era Multicore nasceu

- ▶ Crescimento **sustentado** do desempenho nominal do hardware.
 - ▶ Mais núcleos de processamento → mais **desempenho nominal**.

Desempenho Nominal X Efetivo

- 💡 **Desempenho nominal** é o desempenho máximo teórico de um sistema.
- ✳️ **Desempenho efetivo** é o desempenho real observado em condições reais de uso.
- ▶ **Nem sempre** o desempenho nominal se traduz em desempenho efetivo.

O **software** desempenha um papel **crucial** por **transformar** o desempenho nominal em desempenho real.

Como ceifar o máximo o poder computacional?

Programação paralela se torna uma necessidade ao invés de uma alternativa.



It is not a piece of cake! 🎂

Desafios da computação paralela

- ❗ Programas de computador paralelos são mais difíceis de programar que sequenciais!
- ❗ Introduz diversas novas classes de armadilhas potenciais que podem ocorrer.
- ❗ **"syncronização x Computação"**

Paralelismo e Economia de Energia na Computação

Tecnologia da Informação e Comunicação

Consume mundialmente cerca de 10% dos recursos energéticos.

- ✈ Toda a indústria da aviação consome apenas cerca de 3% .
- 🌿 **Energia:** importante em diferentes escalas por diferentes razões:
 - ⚡ **Sistemas grandes:** custos financeiros e ambientais
por exemplo, Datacenters, Supercomputadores, infraestrutura de TI corporativa, etc.
 - 🔋 **Sistemas pequenos:** autonomia de bateria/gerador, tamanho da bateria
por exemplo, Dispositivos vestíveis, sistemas embarcados e móveis, sem fio, etc.

Necessidade presente e futura:

Sistemas mais eficientes em termos de energia, grandes e pequenos.



Formas de reduzir o consumo de energia em sistemas de computação:

Especialização de hardware

- ▶ ASICs, FPGA, Aceleração, etc.
- ▶ **Menor flexibilidade**

Otimização de hardware (processadores programáveis)

- ▶ ILP, multiple issue, out-of-order issue, hw threading, etc.
- ▶ **Retornos decrescentes**

Desafios de Energia no Software

- ◎ Dois níveis de otimização de software:
 - ▶ **projeto de software,**
 - ▶ **operação de software**

Problema 1: Projeto de software inadequado

⚠ pode desperdiçar grande parte dos esforços de otimização de hardware.

Problema 2: Operação de software inadequada

⚠ desperdício de esforços de otimização de hardware e projeto de software.

Em última análise, o hardware é **controlado por** software.

Explorando paralelismo para economia de energia

- ⚡ Paralelismo permite executar tarefas em menor frequência, mantendo a performance e reduzindo o consumo de energia.
- ➡ Redução do consumo de energia em CPUs e GPUs através do paralelismo e arquiteturas heterogêneas.
- 💡 Melhoria na eficiência energética através do uso de técnicas de computação paralela e otimizações específicas de software.

Next

1 Sobre mim

2 Contexto

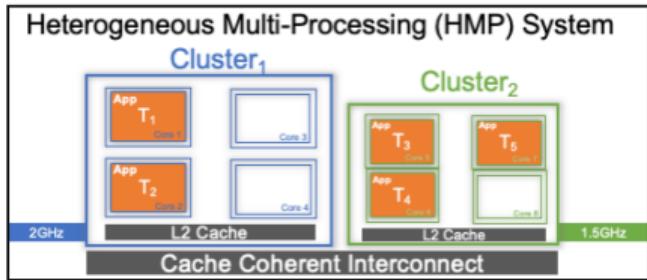
3 Motivação

4 Aplicações

5 Resumo



IPerfEcT - Identificando Ponto Ótimo de Operação de Software



Potential of energy efficiency scales with the degree of parallelism.

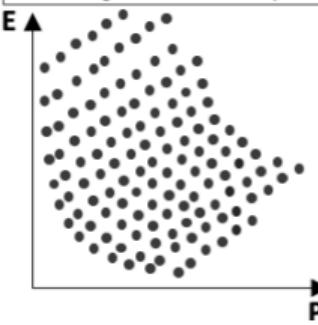
$$\begin{array}{c} \text{Frequency} \\ \text{#Cores} \end{array} \oplus \begin{array}{c} \text{Frequency} \\ \text{#Cores} \end{array} \Rightarrow \text{Vast Performance-Energy (P-E) Trade-off configuration search space.}$$

How to find optimal P-E trade-offs tailored to a parallel application for a given HMP?

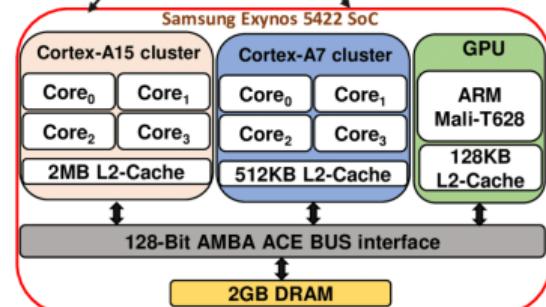
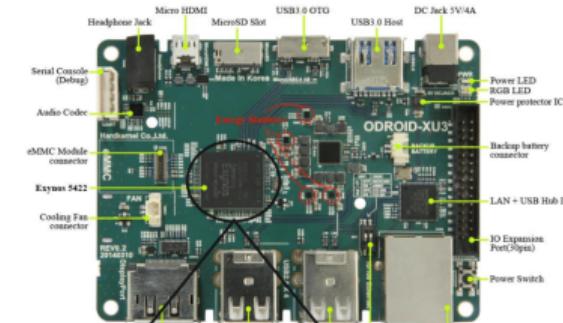
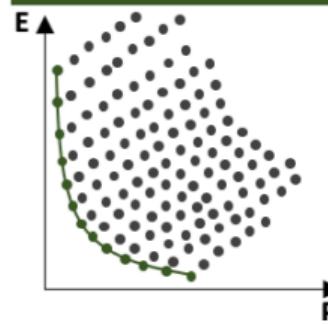
① Analytical Models

$$\begin{aligned} \text{Performance Model} \\ + \\ \text{Power Model} \\ = \\ \text{Energy Model} \end{aligned}$$

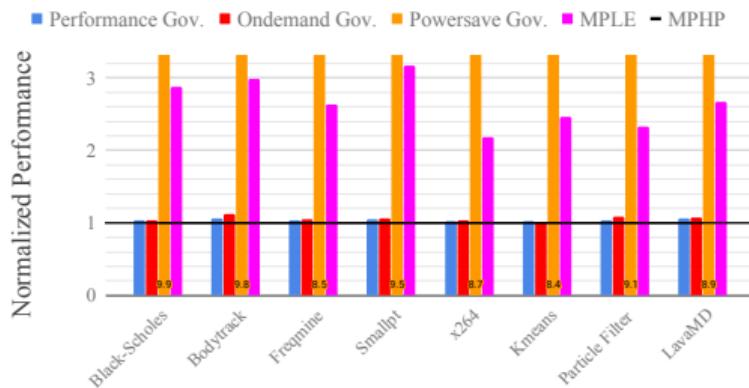
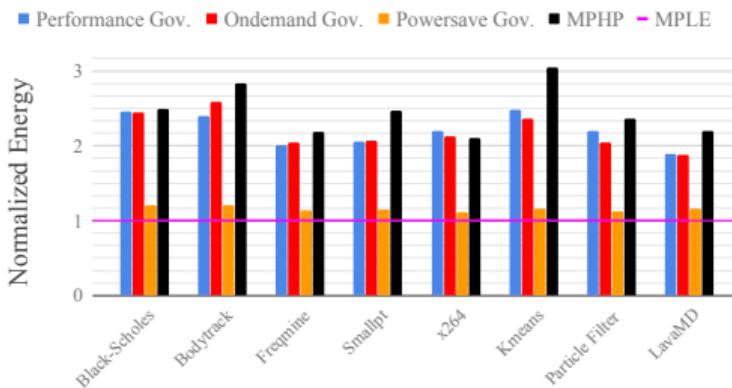
② Performance and Energy Configuration Search Space



③ Optimal Configurations on Pareto Frontier



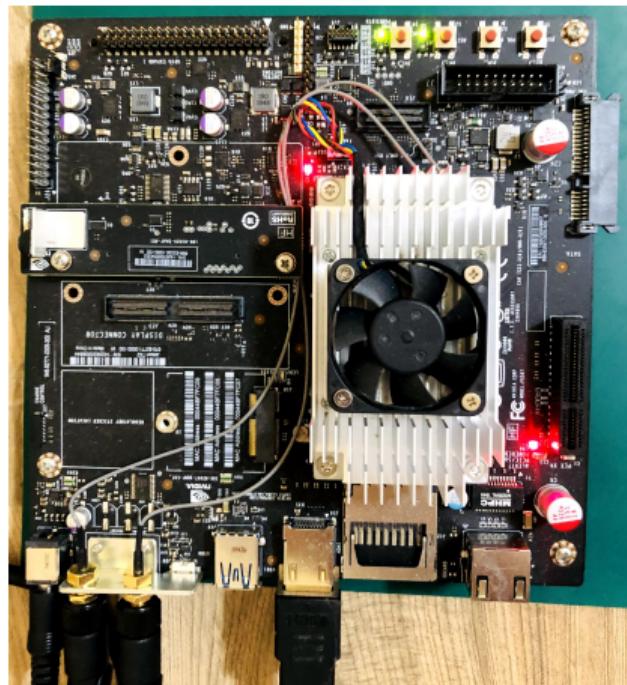
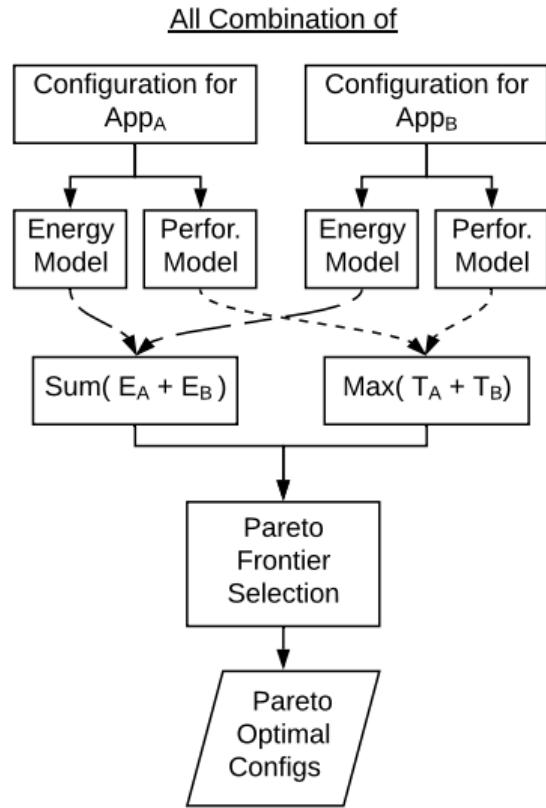
IPerFECT- Identificando Ponto Ótimo de Operação de Software



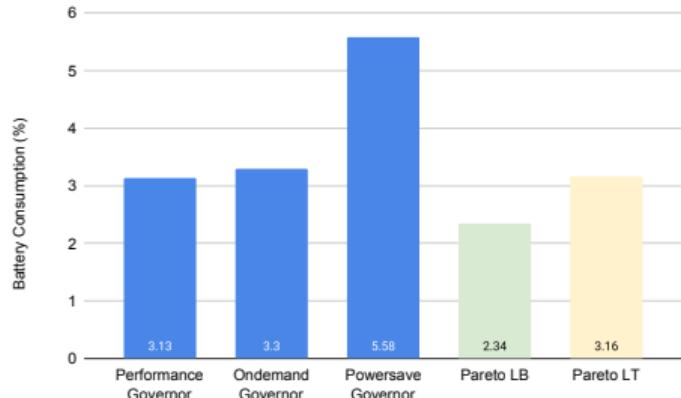
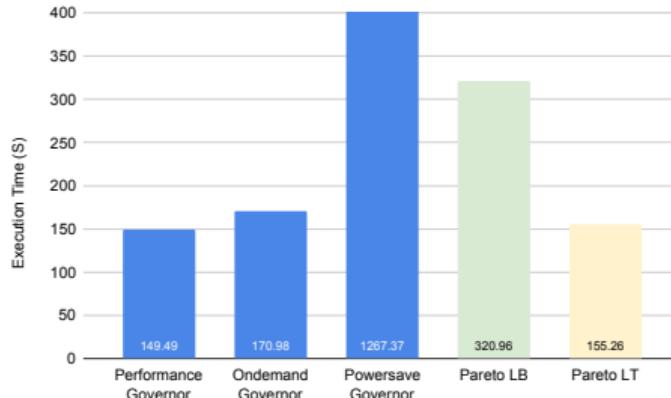
Tutorial ministrado no HiPEAC 2021:
<https://gitlab.com/lappsufrn/iperfect-tutorial>



IPerfEcT - Uso de Caso do uso de Drone

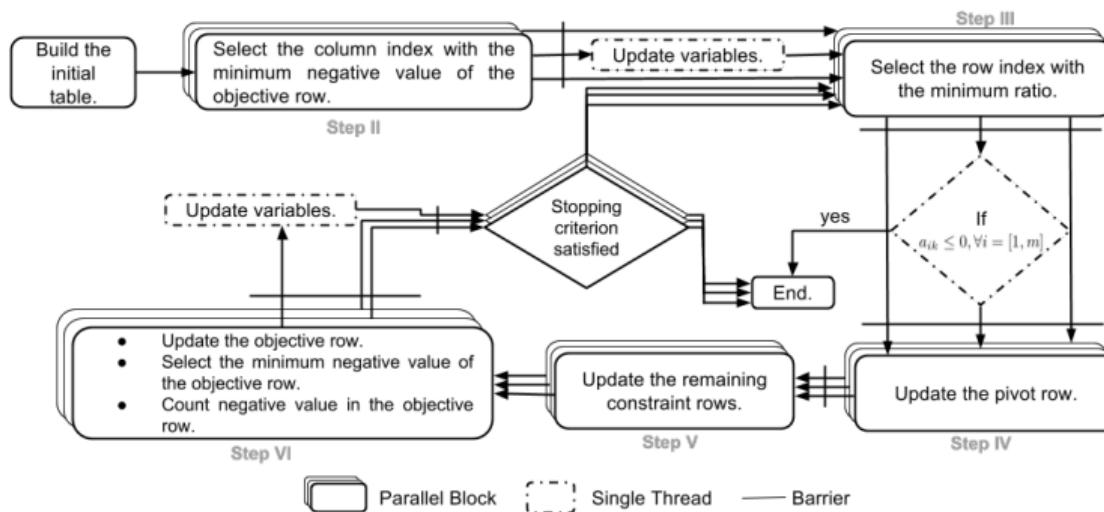


IperfEcT - Uso de Caso do uso de Drone



Tutorial ministrado no HiPEAC 2021:
<https://gitlab.com/lappsufrn/iperfect-tutorial>

Simplex Paralelo de Memória Compartilhada

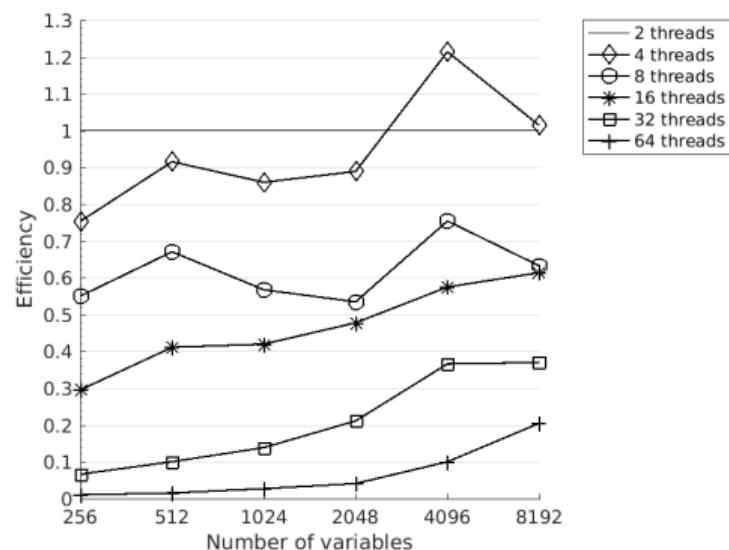


Servidor

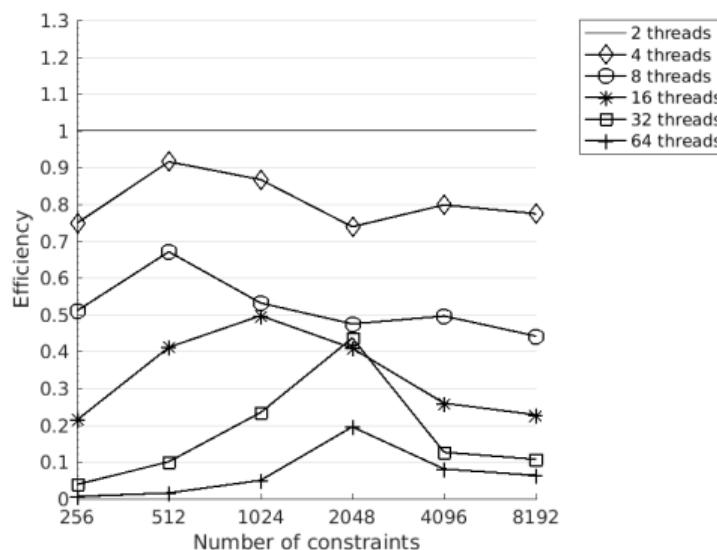
- AMD Opteron 6376 com 64 núcleos.
- 256 GB de RAM DDR3.
- 16MB de Cache L3

Paper: <https://arxiv.org/abs/1804.04737>

Simplex Paralelo de Memória Compartilhada



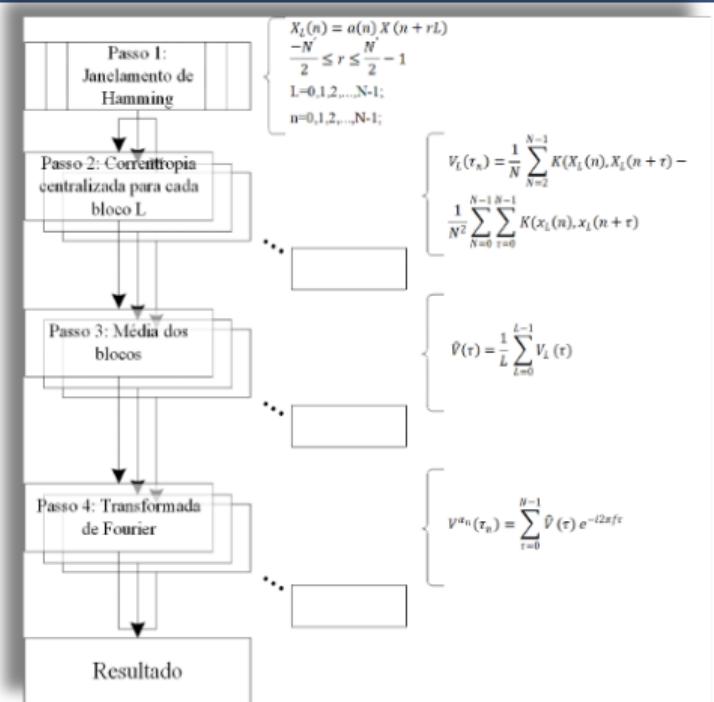
(i) Problems with 512 constraints.



(j) Problems with 512 variables.

Figure: Parallel efficiency of the proposed implementation for fixing one of the problem dimensions in 512 for primal (a) and dual (b) formulations.

Correntropia Cíclica Multicore



Servidor

- AMD Opteron 6376 com 64 núcleos.
- 256 GB de RAM DDR3.
- 16MB de Cache L3

Paper: <https://latamt.ieeer9.org/>

Correntropia Cíclica Multicore

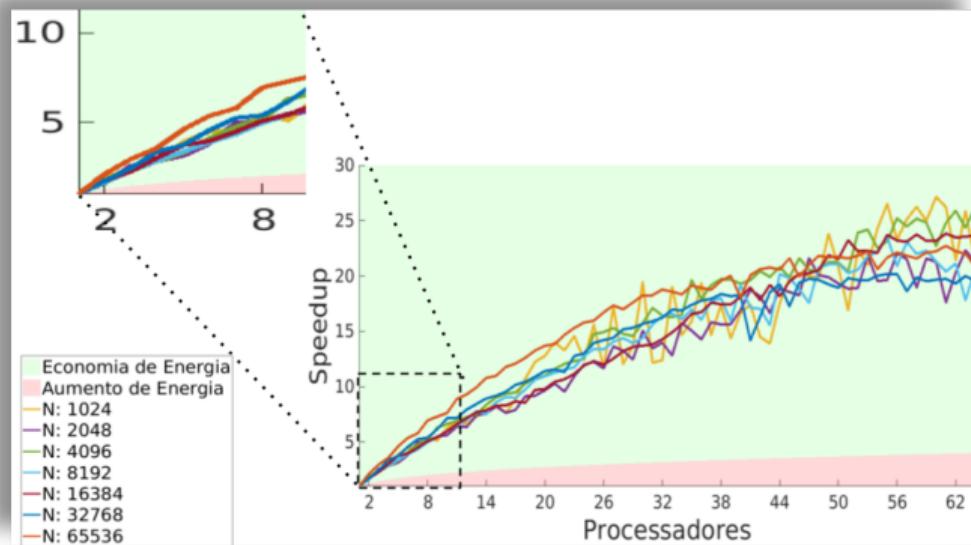
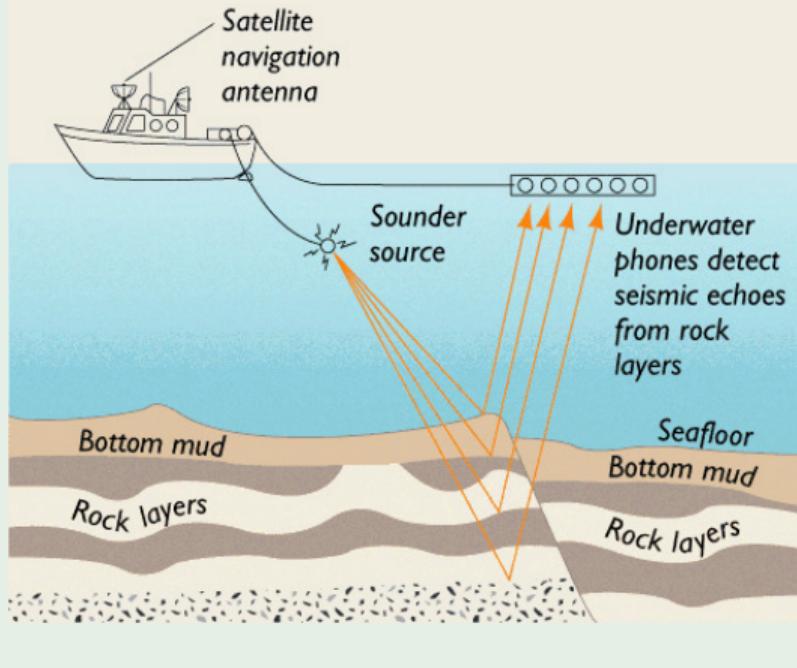


Figure: Speedups para diferentes entradas N aumentando o número de processadores.

Processamento de dados sísmicos

Aquisição sísmica para Exploração de petróleo e gás



Desafios em Sismologia

■ Processamento de dados sísmicos:

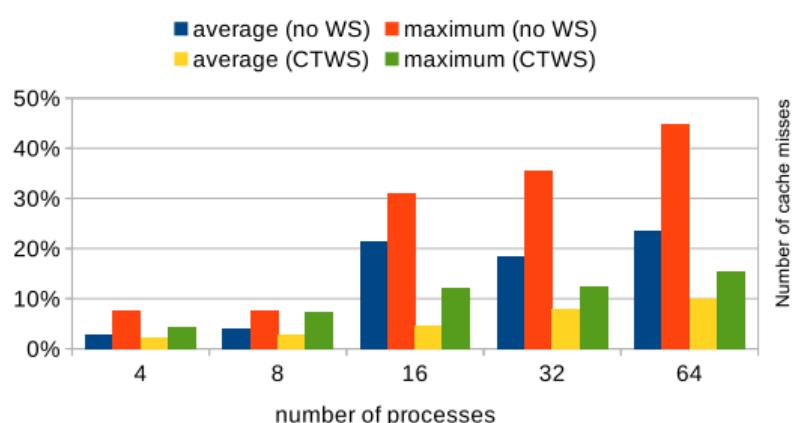
- ❑ Extração de características da subsuperfície:

- ❑ Imagem 3D
- ↔ Velocidades de propagação da onda
- etc

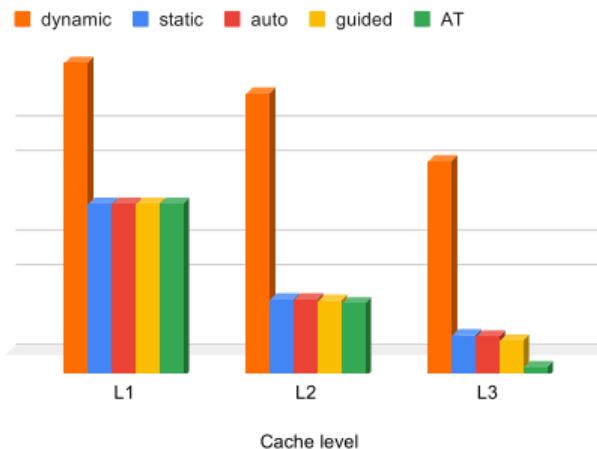
■ Desafios computacionais:

- ⚠ Alto custo computacional
- 叟 Necessidade de computação paralela em supercomputadores

Desafios em Sismologia



(a) Redução do tempo ocioso.



(b) Aumento do reúso de cache.

Figure: Resultados obtidos usando técnicas de paralelismo no método processamento sismico, migração Reversa no Tempo (RTM) 3D.

Next

1 Sobre mim

2 Contexto

3 Motivação

4 Aplicações

5 Resumo



O que vimos até aqui

Computação Paralela

Fundamental para desafios computacionais atuais e futuros.

Desenvolvimento

Requer habilidades e conhecimentos específicos.

Aplicações

Importância e impacto na solução de problemas complexos.

Agenda

Obrigado pela atenção 🙋

Próximos tópicos

Part 1 - **Contexto e Motivação**

Part 2 - **Parallelismo de Hardware e Software**

Part 3 - **Computação Paralela com OpenMP**

Part 4 - **Computação Paralela em CUDA**