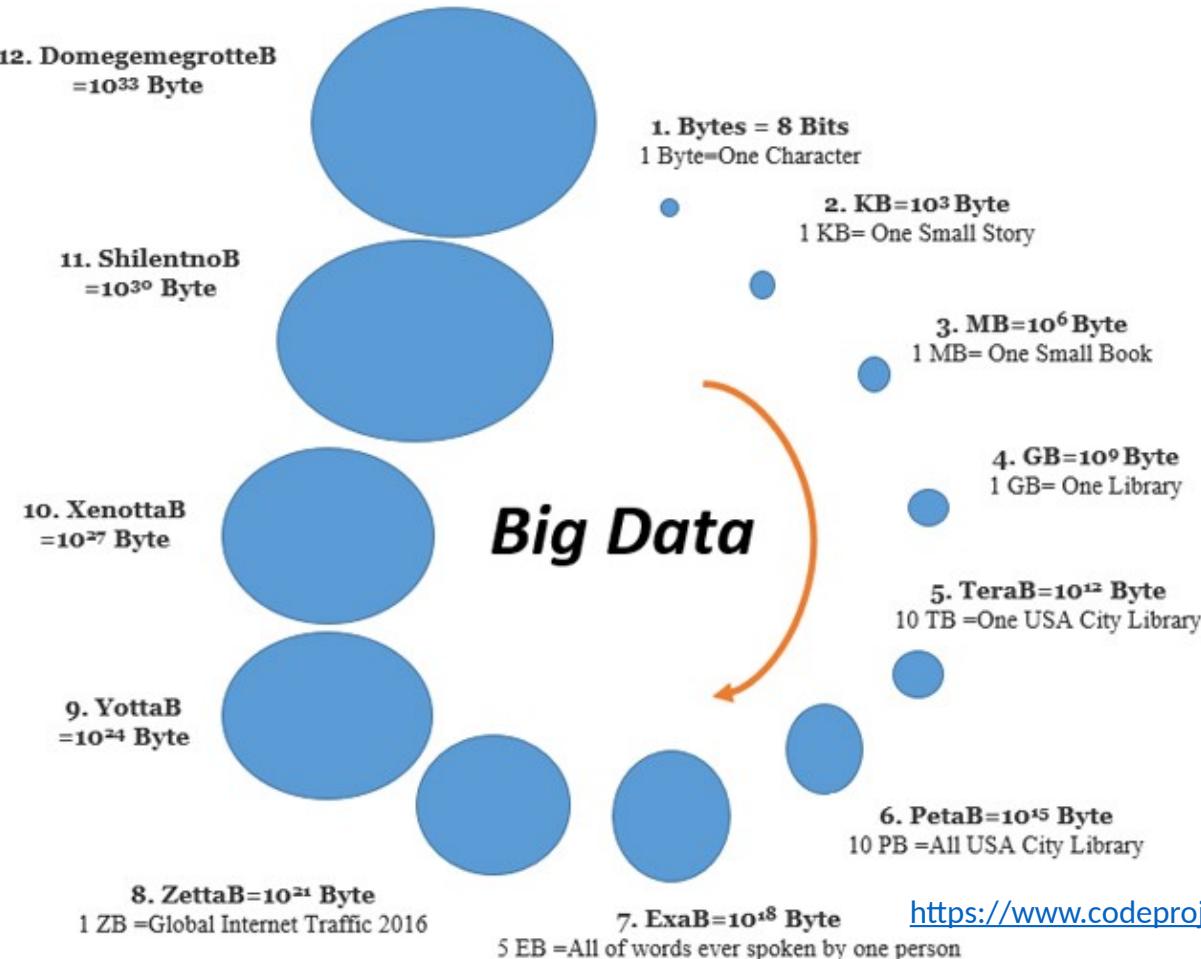


# Arquitetura de Computadores III

Arquiteturas para Big Data

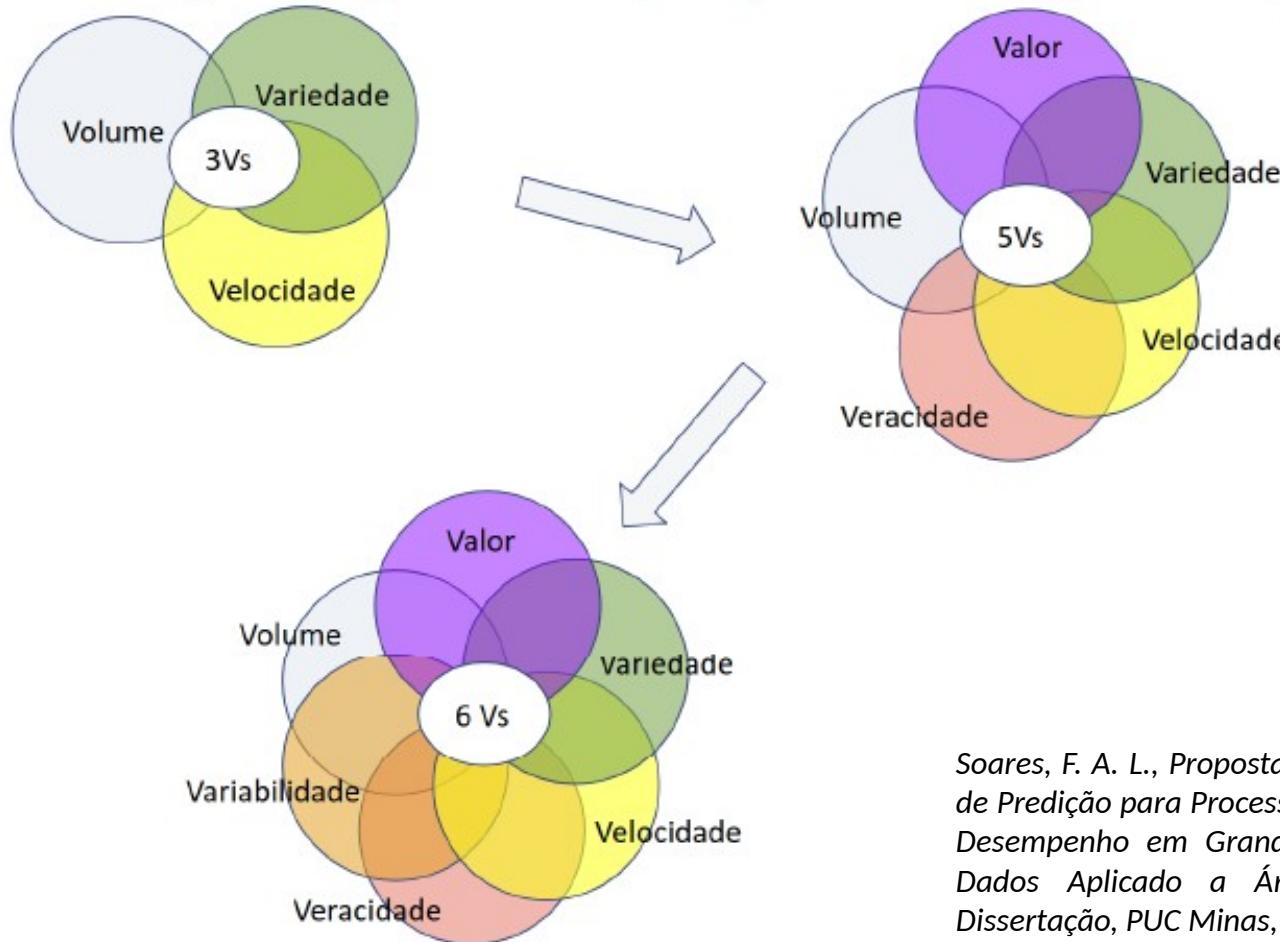
# Big Data



Mahsa Hassankashi

<https://www.codeproject.com/Articles/1239650/Big-Data-2>

# Big Data



Soares, F. A. L., *Proposta de um Sistema de Predição para Processamento de Alto Desempenho em Grandes Volumes de Dados Aplicado a Área da Saúde, Dissertação, PUC Minas, 2020*

# Movimentação de dados

Pequena e rápida

Maior custo

Menor custo

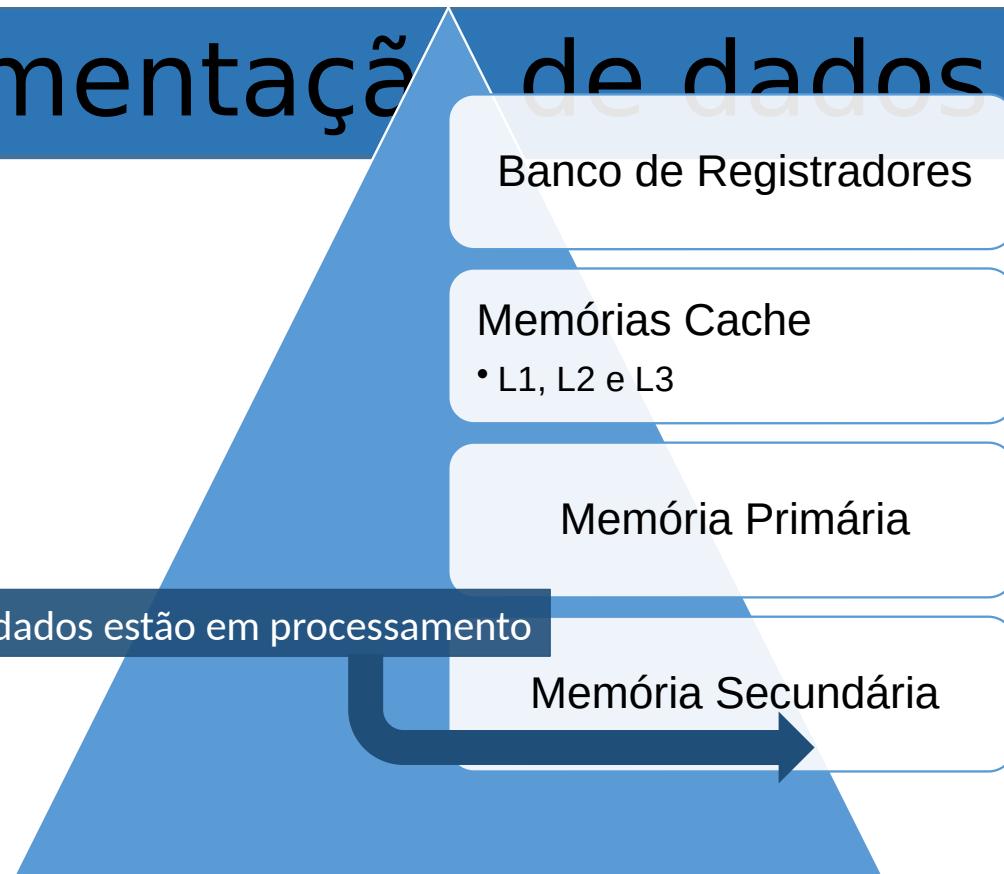
Grande e lenta

Chip do processador

RAM, física, principal, real

Disco, SSD

# Movimentação de dados

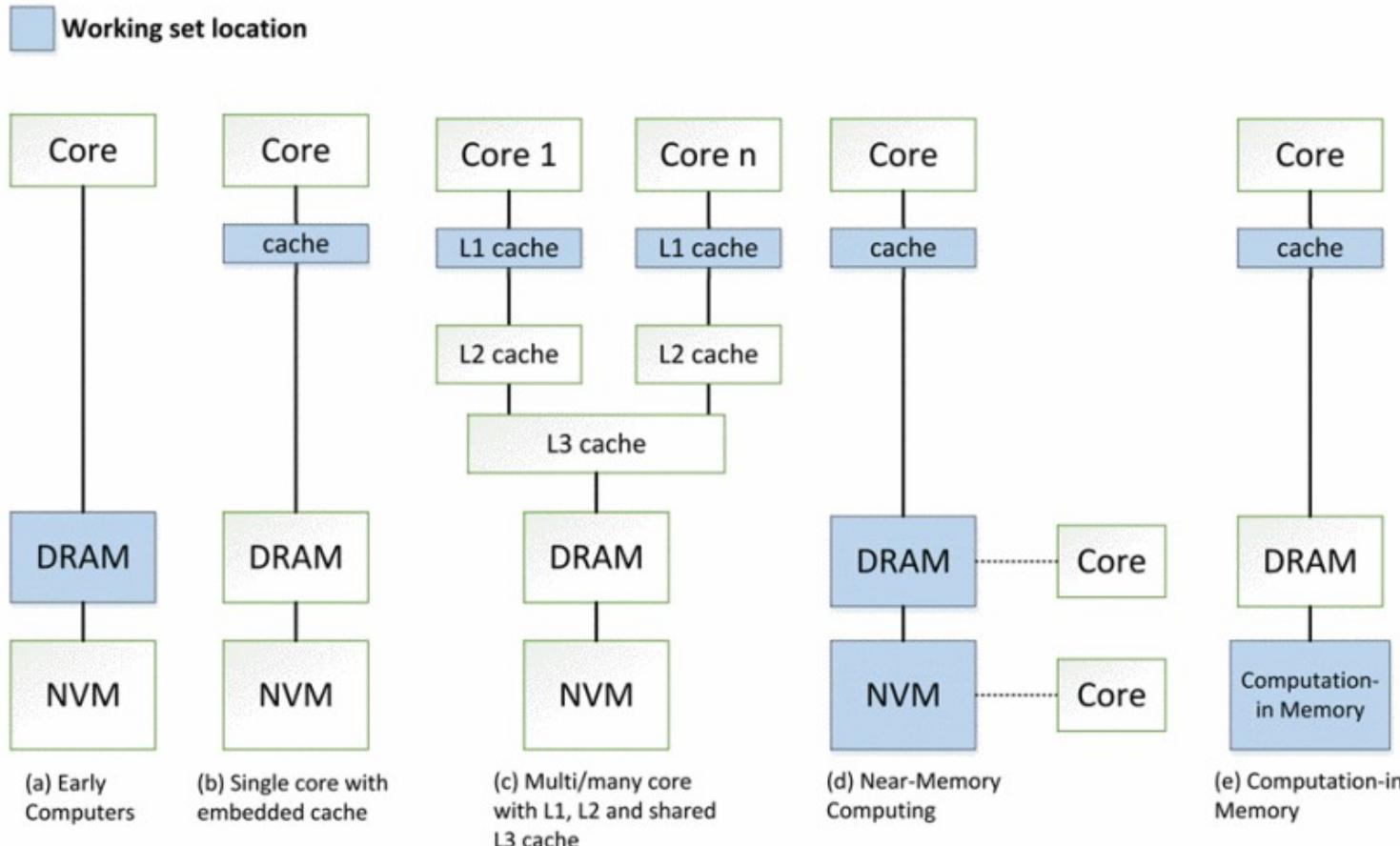


# Processor in Memory

- De 1980 a 2000, o desempenho dos processadores cresceu cerca de 60%, e das memórias DRAM, cerca de 10%.
- O barramento de memória é um gargalo conhecido nas arquiteturas von Neumann.
- Processor in Memory (PIM) é uma estratégia para mitigar o gargalo das arquiteturas von Neumann.
- Computações em memória podem reduzir o número de instruções de load e store.

<https://doi.org/10.1109/HPEC.2019.8916496>

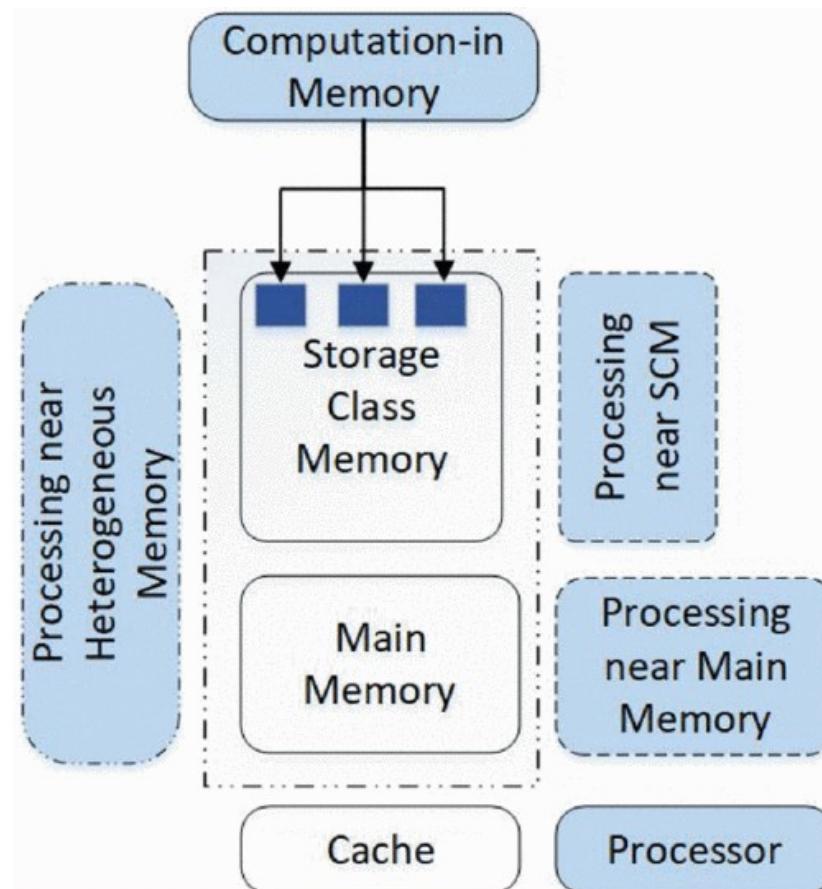
# Classificação baseada no working set location



NVM: Non-volatile Memory

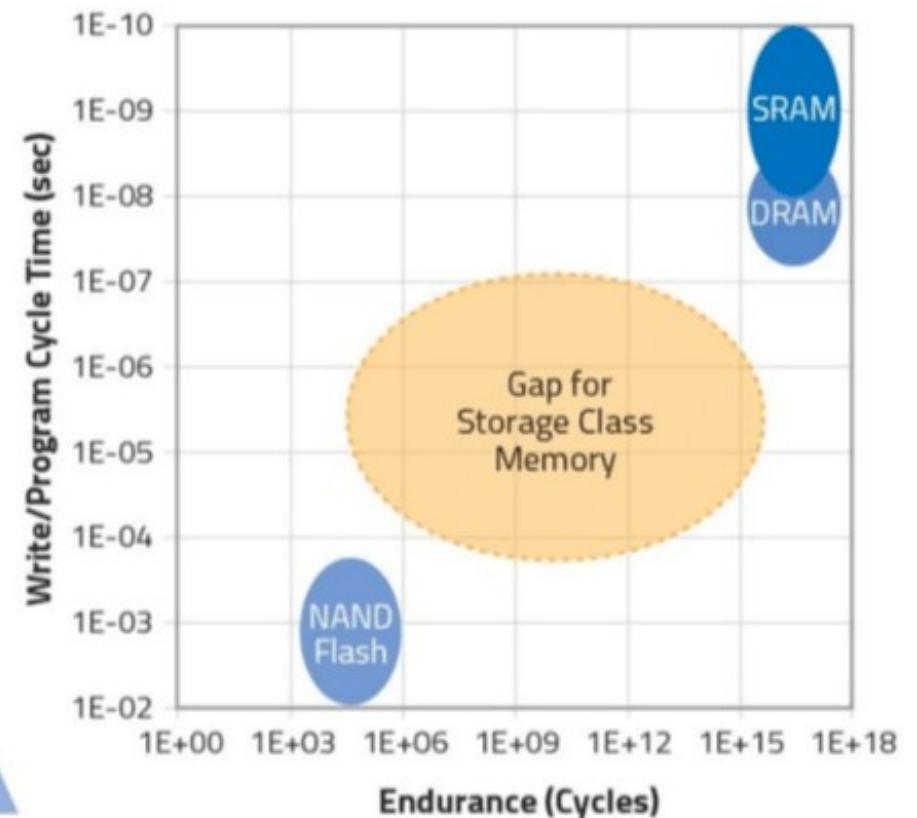
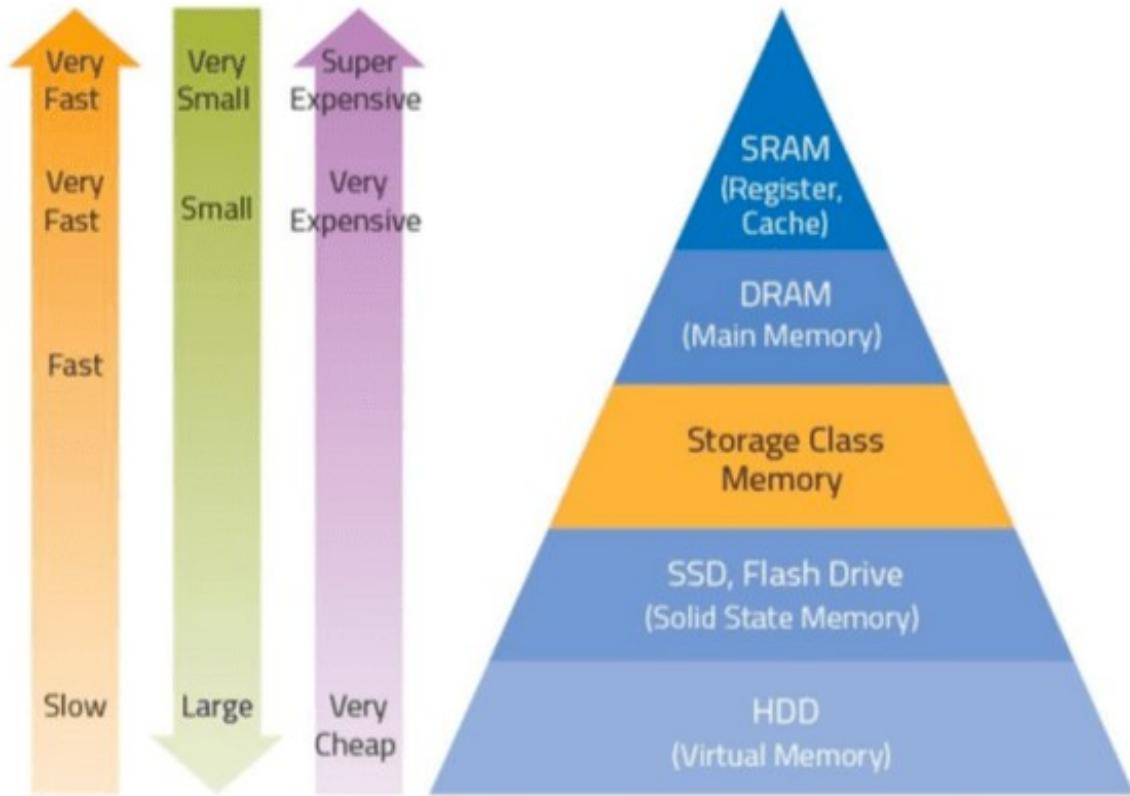
<https://doi.org/10.1109/DSD.2018.00106>

# Opções de computação



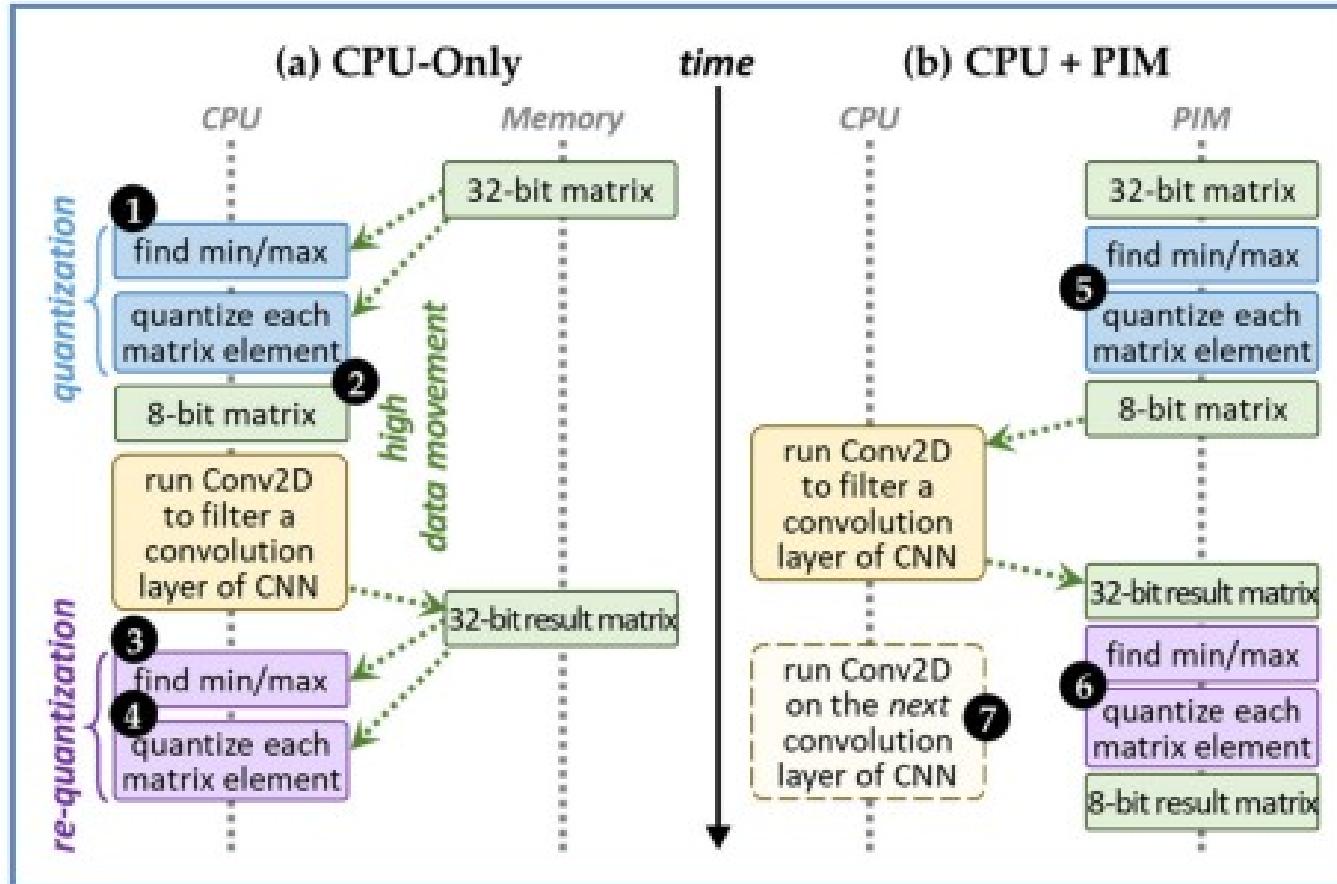
<https://doi.org/10.1109/DSD.2018.00106>

# Memory/storage hierarchy



<https://semiengineering.com/in-memory-vs-near-memory-computing/> (Source: Lam Research)

# Perspectiva de execução de uma carga de trabalho

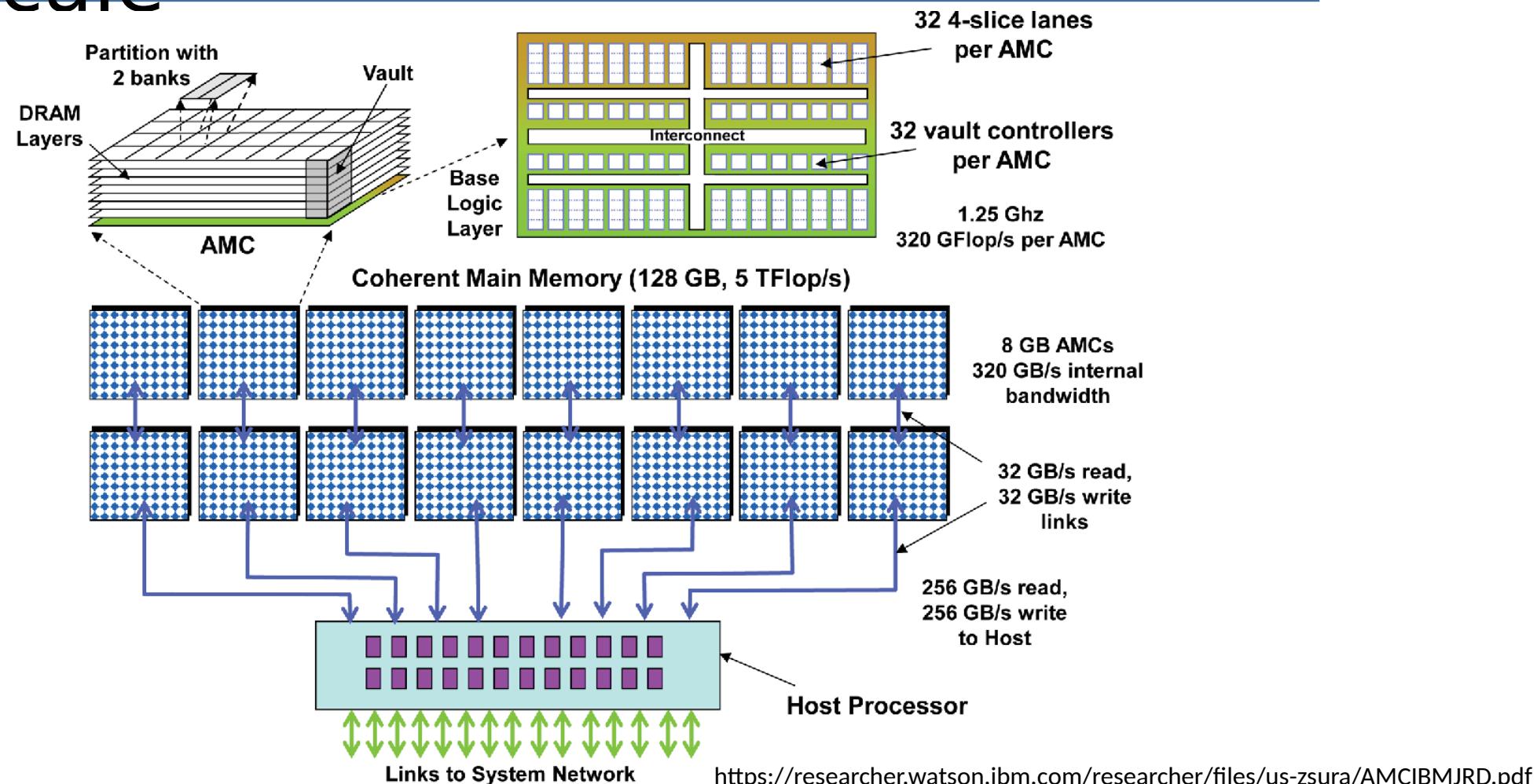


5 inicia nova sequencia.

- Em (b), há redução na movimentação de dados.

[https://www.pdl.cmu.edu/PDL-FTP/associated/19ibmrd\\_pim.pdf](https://www.pdl.cmu.edu/PDL-FTP/associated/19ibmrd_pim.pdf)

# Uma arquitetura PIM para sistemas Exascale



# Arquitetura de Computadores III

Arquiteturas de aprendizado de máquina

# AI, Machine Learning, Deep Learning

## Artificial Intelligence

Sistema de computador que desempenha tarefas típicas do ser humano: reconhecimento de fala; tradução entre linguagens

## Machine learning

Aprendizado supervisionado ou não supervisionado

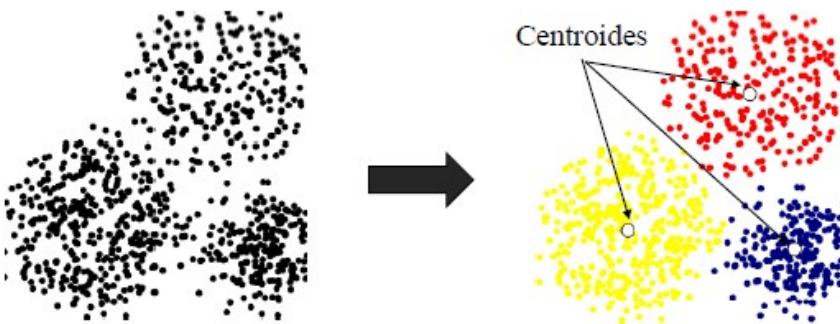
## Deep Learning

Mais acurácia e mais computação  
Deep Artificial Neural Network

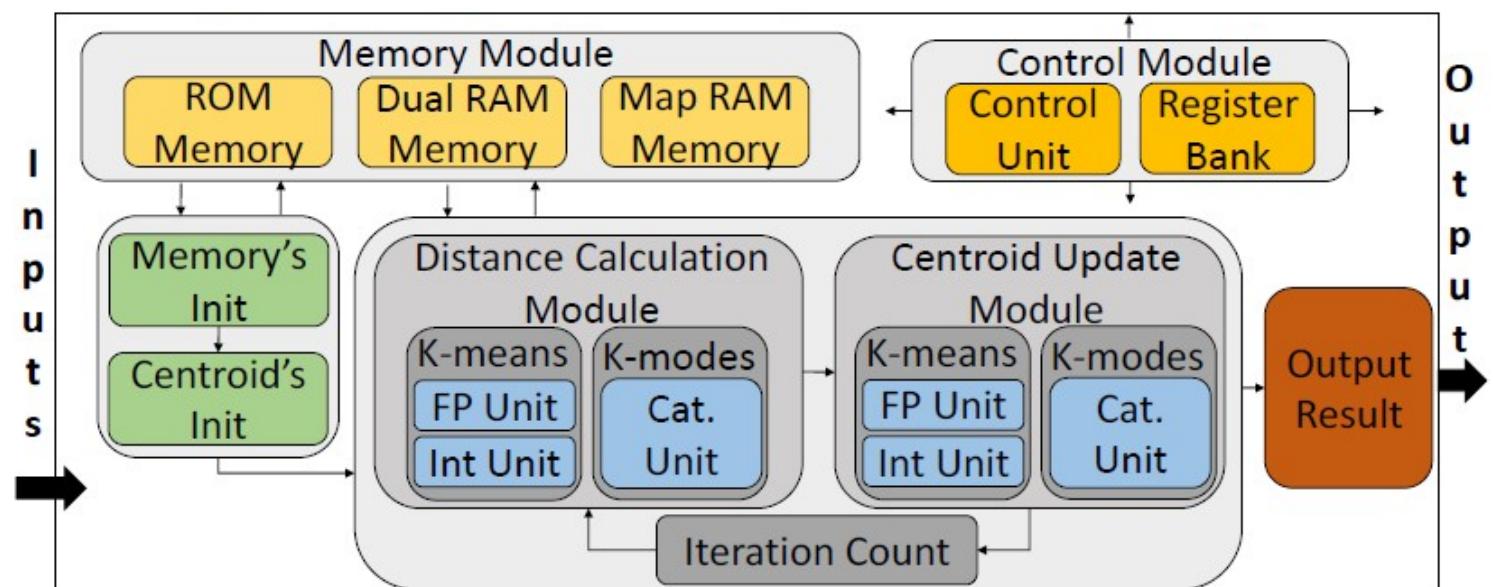
# Machine Learning

- Aprendizado supervisionado: o usuário sabe o que está procurando e especifica uma variável de destino, e a máquina aprende com os dados identificando padrões.
  - Exemplos: Árvores de Decisão, Support Vector Machines (SVM), Redes Neurais
- Aprendizado não supervisionado: o usuário não sabe o que está procurando e a máquina é responsável por definir de acordo com o que os dados fornecidos têm em comum.
  - Exemplos: Apriori, K-means, K-modes, Redes Neurais
- Reinforcement Learning não é classificado como supervisionado ou não supervisionado. Método baseado em recompensa. Em alguns casos pode ser classificado como supervisionado em ambientes de feedbacks esparsos.

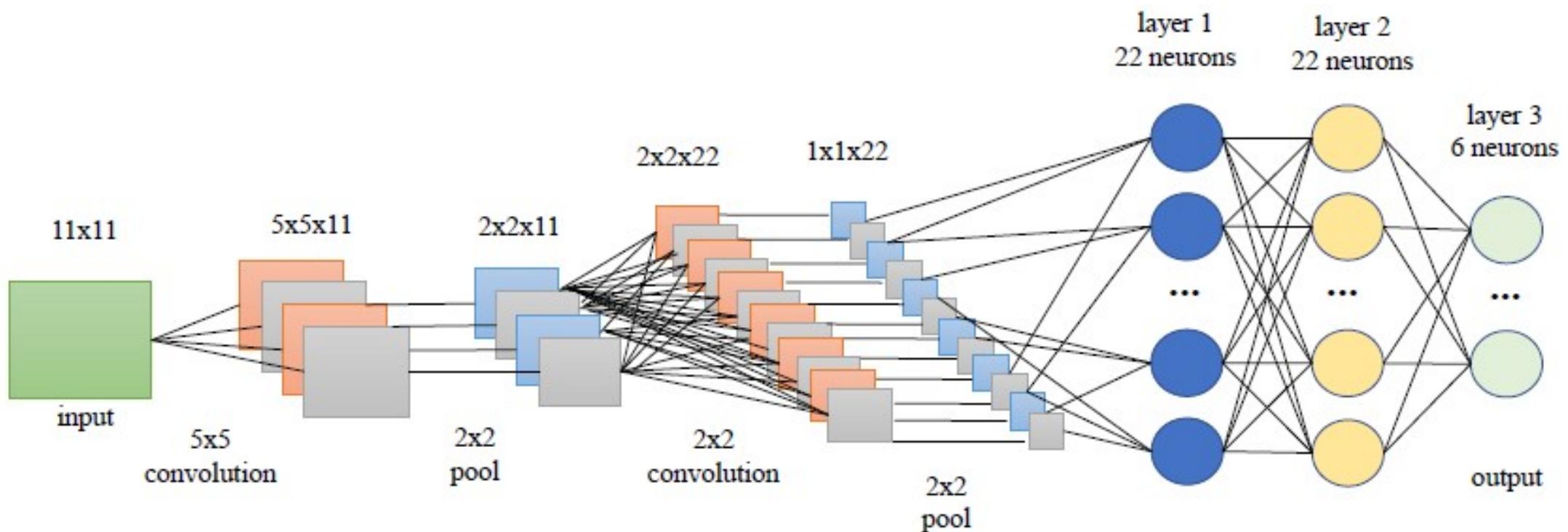
# Arquitetura K-means/K-modes em FPGA



MACIEL, LUCAS ANDRADE ; SOUZA, MATHEUS ALCANTARA ; DE FREITAS, HENRIQUE COTA . Reconfigurable FPGA-based K-means/K-modes Architecture for Network Intrusion Detection. IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II-EXPRESS BRIEFS, v. 67, p. 1459-1463, 2020  
<https://doi.org/10.1109/TCSII.2019.2939826>

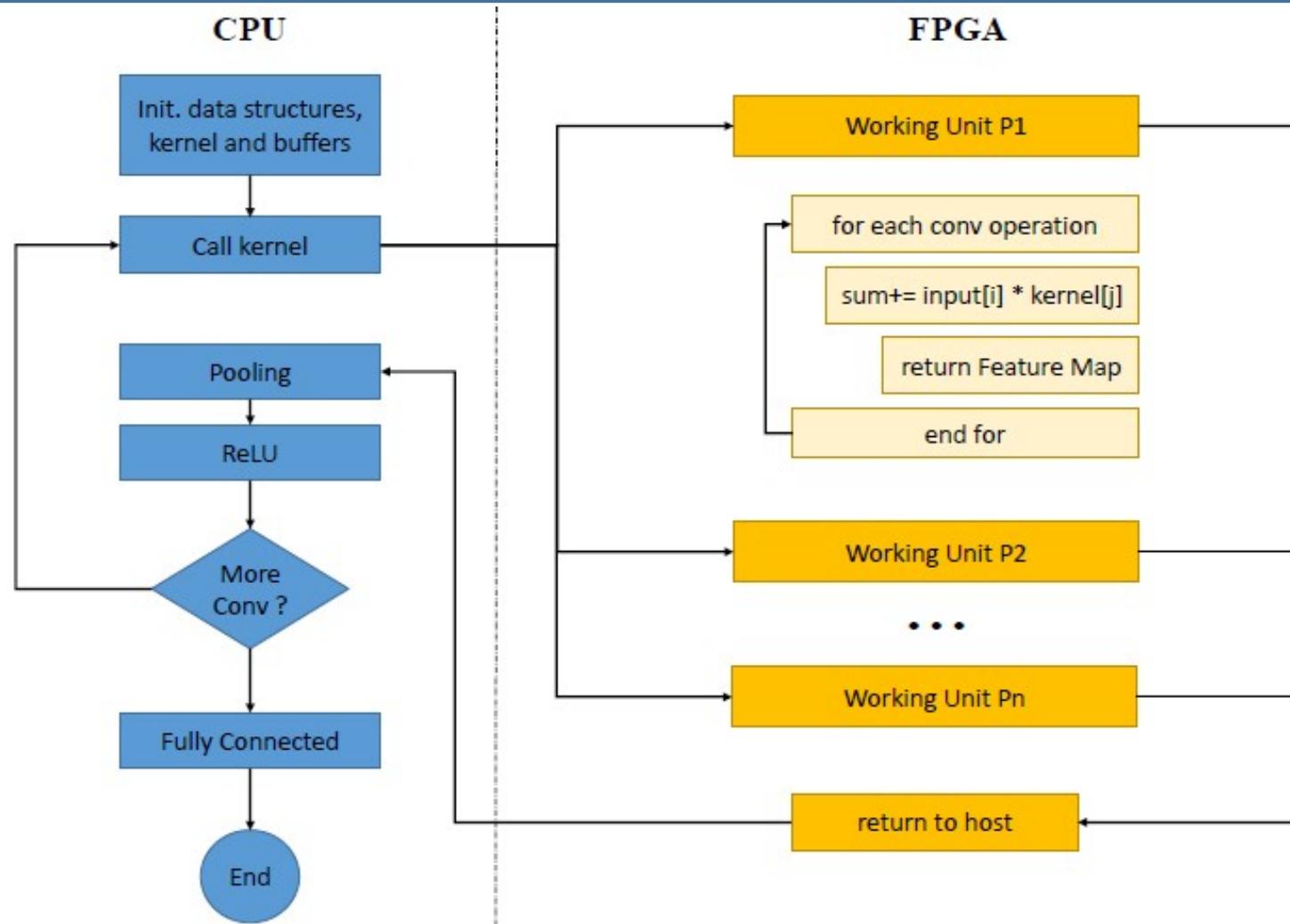


# Rede Neural Convolucional em FPGA



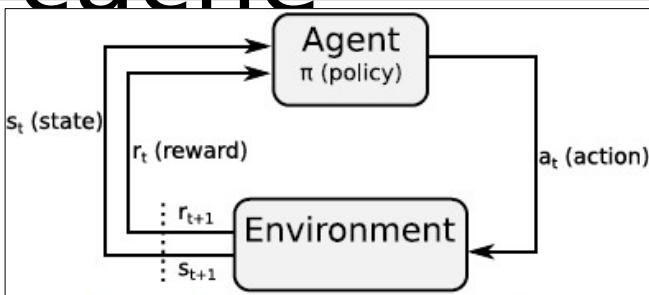
Maciel, L. A., Arquiteturas Reconfiguráveis para Aprendizado de Máquina com Eficiência Energética, Dissertação, PUC Minas, 2020

# Rede Neural Convolucional em FPGA



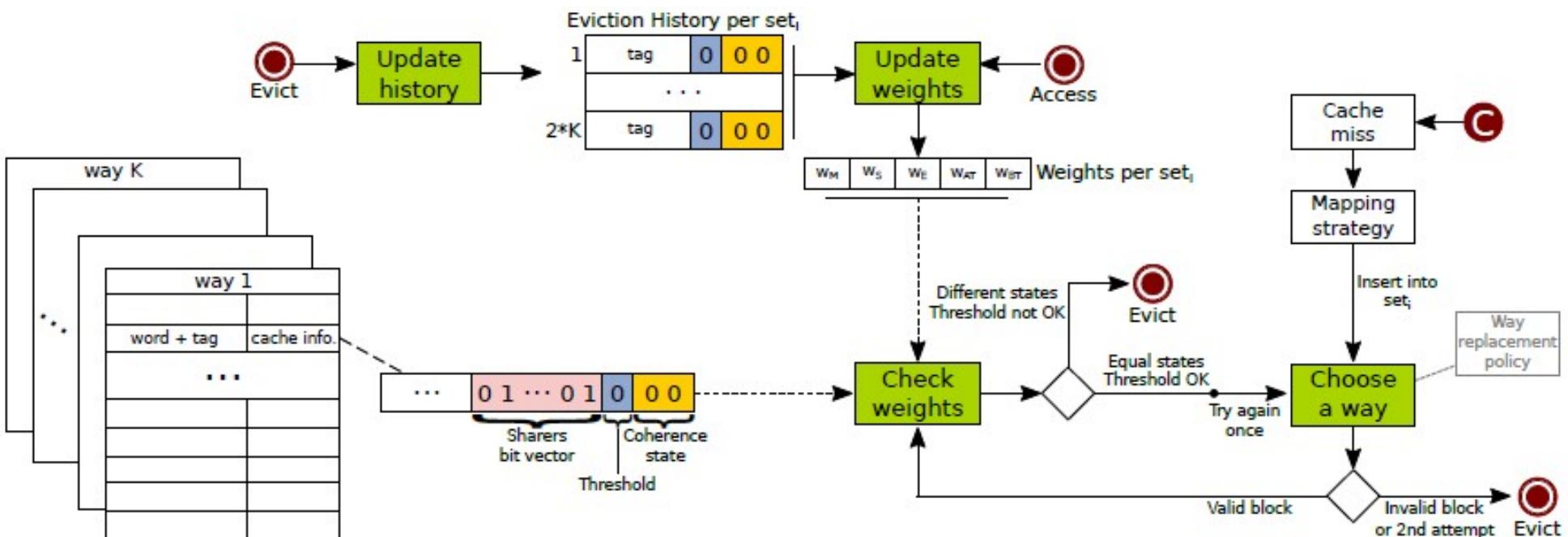
Maciel, L. A., Arquiteturas Reconfiguráveis para Aprendizado de Máquina com Eficiência Energética, Dissertação, PUC Minas, 2020

# RL para substituição de blocos em cache



Souza, M. A., Way-Replacement Algorithms for Multicore Processors Based on Coherence and Sharing States with Reinforcement Learning, Tese, PUC Minas, 2021

## Estratégia de despejo baseada em reinforcement learning

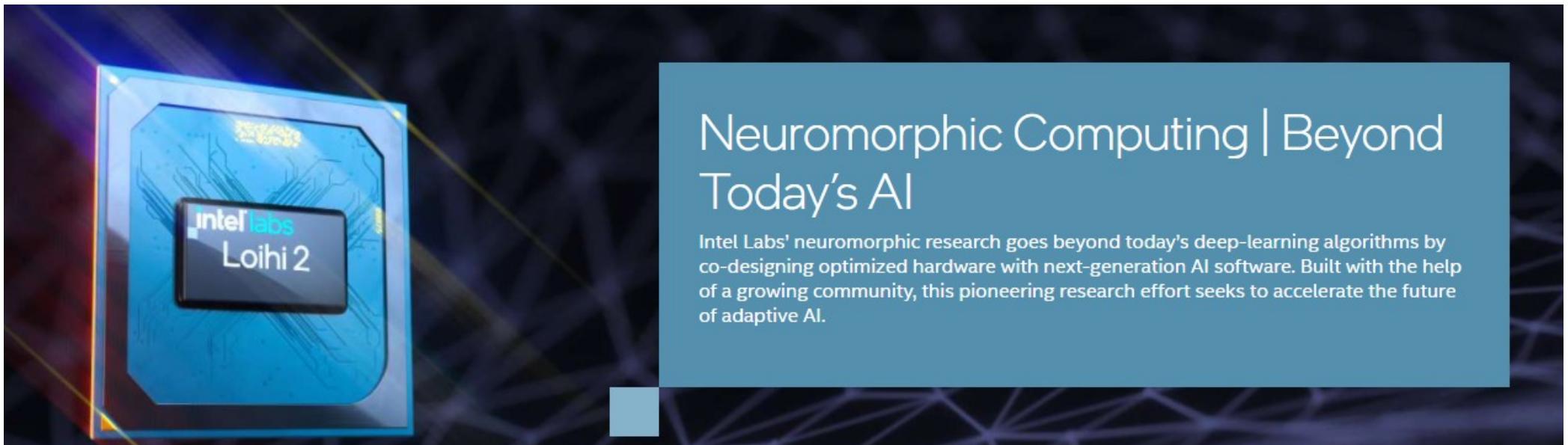


# Arquitetura de Computadores III

Arquiteturas Neuromórficas

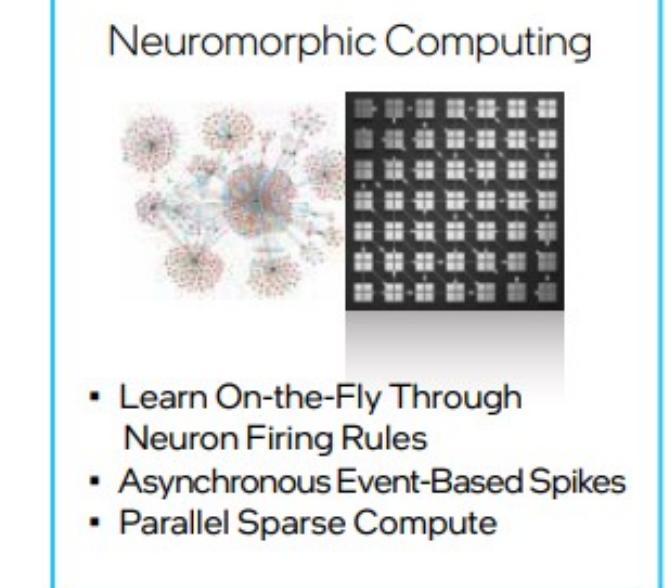
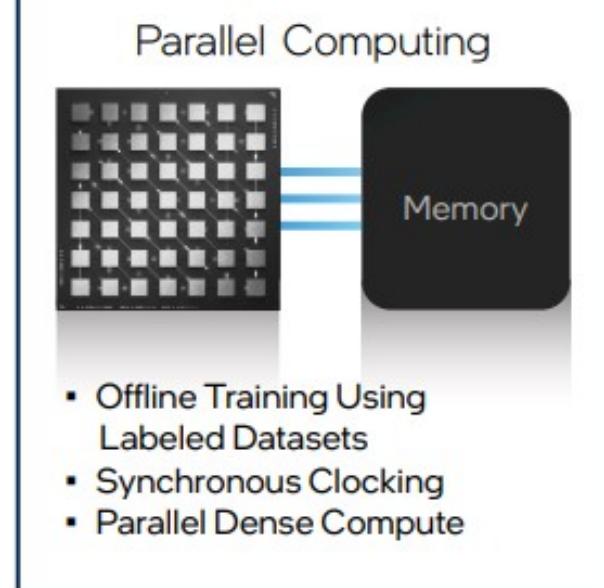
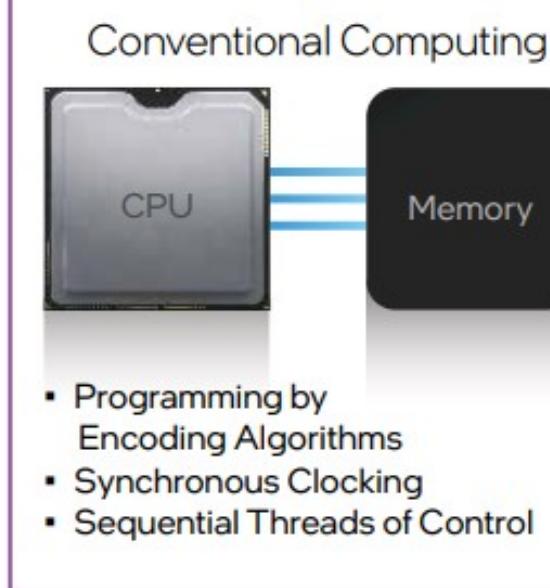
# Computação Neuromórfica

- É o uso de sistemas VLSI (*Very Large Scale Integration*) e eletrônica analógica para simular arquiteturas neurobiológicas presentes no sistema nervoso e cérebro humano. <https://github.com/mikeroyal/Neuromorphic-Computing-Guide>



<https://www.intel.com/content/www/us/en/research/neuromorphic-computing.html>

# Computação Neuromórfica



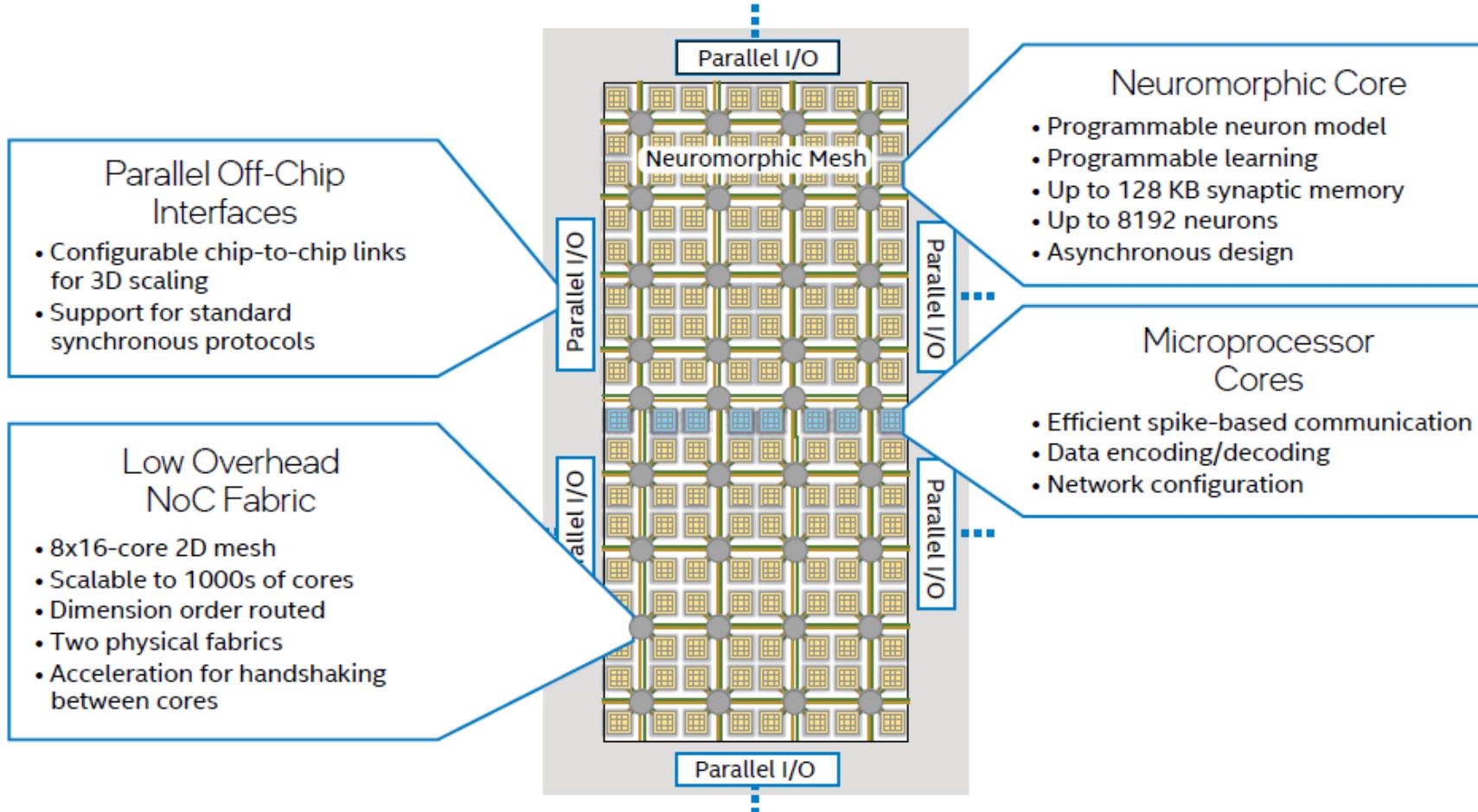
<https://www.intel.com/content/www/us/en/research/neuromorphic-computing-loihi-2-technology-brief.html>

# Conjunto de Instruções do Intel Loihi 2

OP CODES	DESCRIPTION
<b>RMW, RDC</b> <i>read-modify-write, read-and-clear</i>	Access neural state variables in the neuron's local memory space
<b>MOV, SEL</b> <i>move, move if 'c' flag</i>	Copy neuron variables and parameters between registers and the neuron's local memory space
<b>AND, OR, SHL</b> <i>and, or, shift left</i>	Bitwise operations
<b>ADD, NEG, MIN</b> <i>add, negate, minimum</i>	Basic arithmetic operations
<b>MUL_SHR</b> <i>multiply shift right</i>	Fixed precision multiplication
<b>LT, GE, EQ</b> <i>less than, not equal, equals</i>	Compare and write result to 'c' flag
<b>SKP_C, JMP_C</b> <i>skip ops, jump to program address based on 'c' flag</i>	Branching to navigate program
<b>SPIKE, PROBE</b> <i>spike, send probe data</i>	Generate spike or send probe data to processor

<https://www.intel.com/content/www/us/en/research/neuromorphic-computing-loihi-2-technology-brief.html>

# Arquitetura do Intel Loihi 2

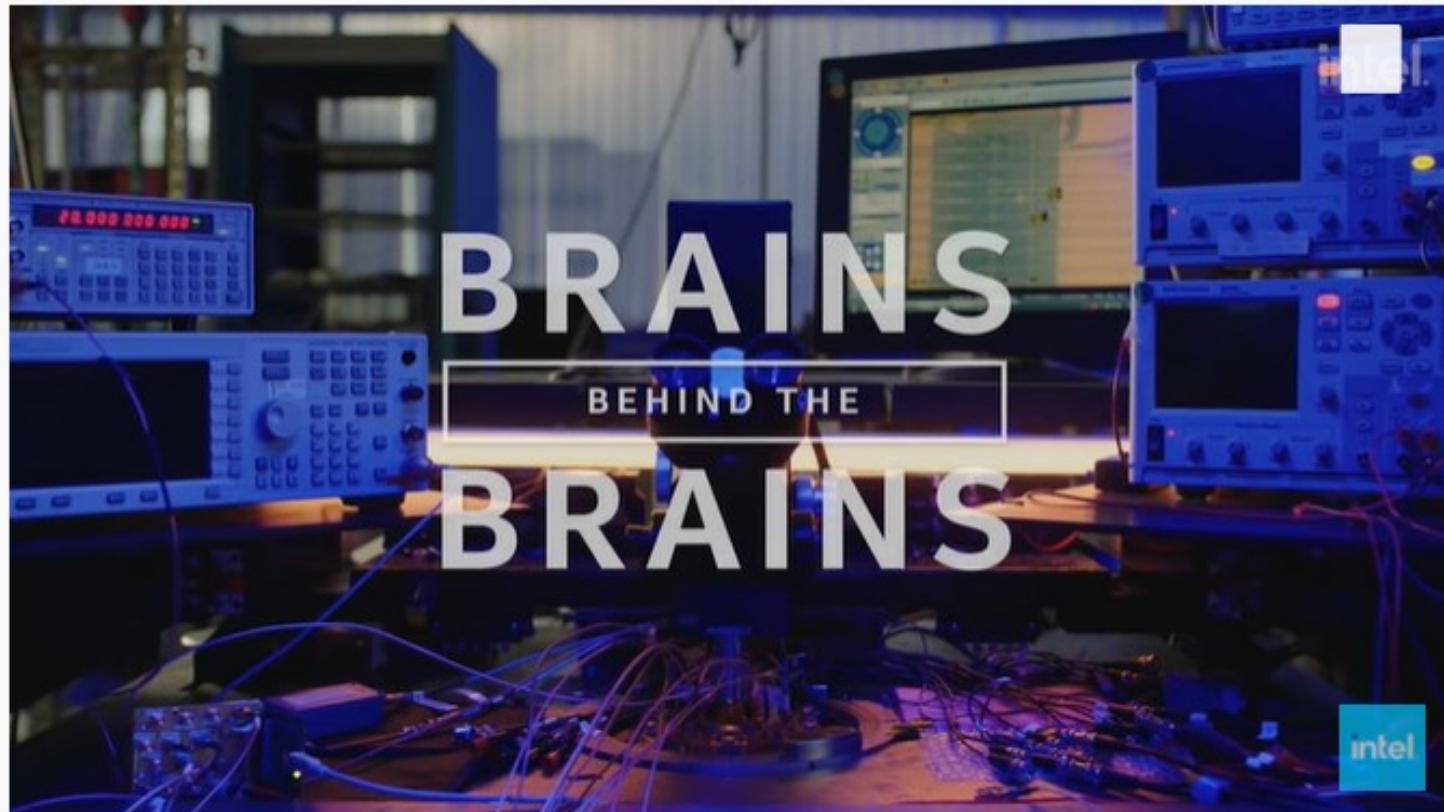


# Loihi vs. Loihi 2

<https://www.intel.com/content/www/us/en/research/neuromorphic-computing-loihi-2-technology-brief.html>

Resources/Features	Loihi	Loihi 2
Process	Intel 14nm	Intel 4
Die Area	60 mm <sup>2</sup>	31 mm <sup>2</sup>
Core Area	0.41 mm <sup>2</sup>	0.21 mm <sup>2</sup>
Transistors	2.1 billion	2.3 billion
Max # Neuron Cores/Chip	128	128
Max # Processors/Chip	3	6
Max # Neurons/Chip	128,000	1 million
Max # Synapses/Chip	128 million	120 million
Memory/Neuron Core	208 KB, fixed allocation	192 KB, flexible allocation
Neuron Models	Generalized LIF	Fully programmable
Neuron State Allocation	Fixed at 24 bytes per neuron	Variable from 0 to 4096 per neuron depending on neuron model requirements
Connectivity Features	Basic compression features: <ul style="list-style-type: none"><li>• Variety of sparse and dense synaptic compression formats</li><li>• Weight sharing of source neuron fanout lists</li></ul>	In addition to the Loihi 1 features: <ul style="list-style-type: none"><li>• Shared synapses for convolution</li><li>• Synapses generated from seed</li><li>• Presynaptic weight-scaling factors</li><li>• Core fan-out list compression and sharing</li><li>• Broadcast of spikes at destination chip</li></ul>
Information Coding	Binary spike events	Graded spike events (up to 32-bit payload)
Neuron State Monitoring (for development/debug)	Requires remote pause and query of neuron memory	Neurons can transmit their state on-the-fly
Learning Architecture	Programmable rules applied to pre-, post-, and reward traces	Programmable rules applied to pre-, post-, and generalized "third-factor" traces
Spike Input	Handled by embedded processors	Hardware acceleration for spike encoding and synchronization of Loihi with external data stream
Spike Output	1,000 hardware-accelerated spike receivers per embedded processor	In addition to the Loihi 1 feature, hardware accelerated spike output per chip for reporting graded payload, timing, and source neuron
External Loihi Interfaces	Proprietary asynchronous Interface	Support for standard synchronous (SPI) and asynchronous (AER) protocols, GPIO, and 1000BASE-KX, 2500BASE-KX, and 10GBase-KR Ethernet
Multi-Chip Scaling	2D tile-able chip array Single inter-chip asynchronous protocol with fixed pin-count	3D tile-able chip array Range of inter-chip asynchronous protocols with variable pipelining and pin-counts optimized for different system configurations
Timestep Synchronization	Handled by cores	Accelerated by NoC routers

# Brains behind the Brains



**Brains Behind the Brains: Mike Davies and Neuromorphic Computing at Intel  
Labs | Intel** <https://www.youtube.com/watch?v=GN3eSMoJcM8>

# Arquitetura de Computadores III

Arquiteturas Quânticas

# Computação Quântica

- Bibliografia sugerida:
  - National Academies of Sciences, Engineering, and Medicine. 2019. **Quantum Computing: Progress and Prospects**. Washington, DC: The National Academies Press. <https://doi.org/10.17226/25196>
  - João Bosco M. Sobral, Renato Bobsin Machado, **Computação Quântica: Aspectos Físicos e Matemáticos - Uma Abordagem Algébrica**, Série Pensamento Matemático @ Ciência da Computação, UFSC/UNIOESTE, 1<sup>a</sup> Edição, 2019. <https://repositorio.ufsc.br/handle/123456789/197576>
    - “A computação quântica é mais uma teoria sobre o formalismo mecânico quântico abstrato, construída em conceitos como qubits e portas quânticas...”

# Computação Quântica

- Na computação clássica, o computador é baseado na arquitetura de John von Neumann.
- A miniaturização, de acordo com a Lei de Moore, é um fator fundamental para a evolução da computação, no entanto, a indústria de fabricação de circuitos integrados tem enfrentado problemas devido ao chamado **tunelamento quântico**.
- Solução: **computação baseada em mecânica quântica**.

# Computação Quântica

- O computador Quântico é baseado em leis da física quântica:
  - sobreposição quântica
  - entrelaçamento quântico
- Sobreposição quântica: é um princípio fundamental da mecânica quântica que em que dois ou mais estados quânticos podem ser combinados e o resultado será um novo estado, uma sobreposição dos anteriores.
- Entrelaçamento quântico: é uma propriedade da mecânica quântica em que os estados quânticos de duas ou mais partículas podem ser descritos em referência umas as outras, independente da distância entre elas no universo.

# O *qubit* (bit quântico)

- O bit na computação clássica, vale 0 ou 1. No entanto, na computação quântica, o *qubit* pode valor 0 e 1 ao mesmo tempo.
- Analogia com uma moeda (cara ou coroa): uma moeda é um objeto macroscópico, governado pela física clássica. Por isso, ao jogarmos uma moeda, o resultado vai ser cara (0) ou coroa (1). No entanto, em um mundo microscópico, governado pela mecânica quântica, ao jogarmos a moeda, os dois lados poderiam ser vistos, devido a propriedade da sobreposição.  
$$|\varphi\rangle = \alpha|0\rangle + \beta|1\rangle$$

Significa sobreposição

Está simultaneamente em 0 e 1
- O *qubit* é representado pela seguinte expressão:  
onde  $\alpha$  e  $\beta$  são números complexos, associados à probabilidade de se encontrar, no lançamento de uma moeda, 0 e 1, respectivamente.

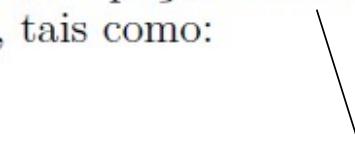
# O *qubit* (bit quântico)

$$|\varphi\rangle = |a\rangle + |b\rangle \text{ (o sinal de '+' significa sobreposição)}$$

onde os estados  $|a\rangle$  e  $|b\rangle$  existem simultaneamente. Contudo, quando observarmos o sistema, ou seja, quando efetuarmos uma medição, um dos estados **colapsa** (deixa de existir) e o outro (o de maior probabilidade) é então detectado. Nesse sentido, o exemplo da moeda faz alusão a essa curiosa propriedade quântica de superposição.

Agora, se pensarmos nos bits 0 e 1, teremos, para a computação quântica, os estados quânticos correspondentes para representar os bits clássicos 0 e 1 e, portanto, o chamaremos *qubit*. Os valores 0 e 1 que um bit clássico pode assumir serão substituídos pelos vetores  $|0\rangle$  e  $|1\rangle$  representados por dois vetores do espaço de Hilbert 2-dimensional de números complexos  $\mathbb{C}$ , com produto interno, tais como:

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ e } |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$



Os elementos de espaço de Hilbert abstrato são chamados vetores. Em aplicações, eles são tipicamente sequências de números complexos ou funções. Na *mecânica quântica*, por exemplo, um sistema físico é descrito por um espaço de Hilbert complexo que contém os vetores de estado, que contém todas as informações do sistema com suas complexidades.

João Bosco M. Sobral, Renato Bobsin Machado, **Computação Quântica: Aspectos Físicos e Matemáticos – Uma Abordagem Algébrica**, Série Pensamento Matemático @ Ciência da Computação, UFSC/UNIOESTE, 1ª Edição, 2019. <https://repositorio.ufsc.br/handle/123456789/197576>

# Computação Clássica vs. Quântica

## A diferença entre bit e qubit

### BIT

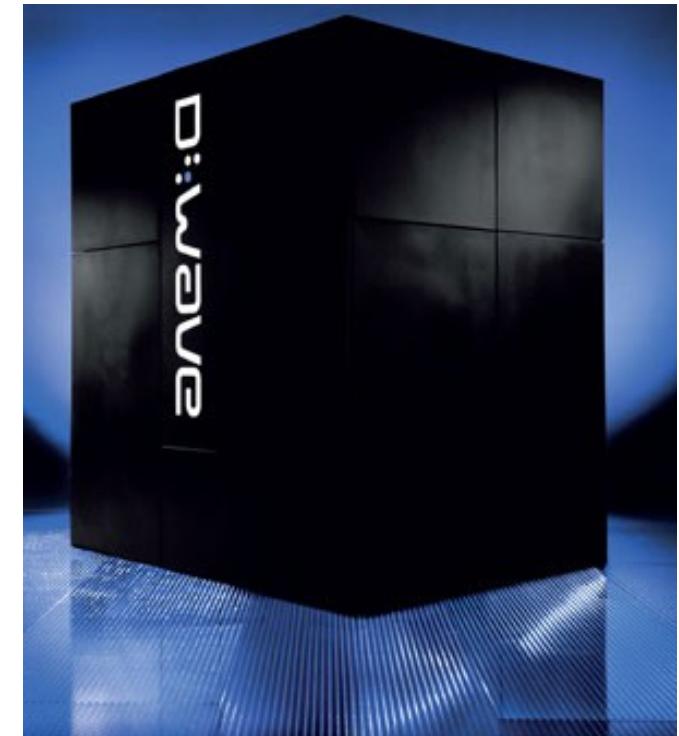
- 0
  - 
  - 
  - 1
- É a menor unidade de informação dos computadores atuais. Representa apenas um de dois valores possíveis, 0 ou 1

### QUBIT

- 0
  - 
  - 
  - 1
- Análogo quântico do bit clássico. Pode assumir dois valores simultaneamente, 0 e 1. Essa característica amplia a capacidade de fazer cálculos em paralelo

Investimento em aceleradores e computadores quânticos

<https://revistapesquisa.fapesp.br/a-nova-onda-dos-qubits/>



computador quântico D-Wave One: máquina tem o dobro da altura de um homem, ocupa área de 10 metros quadrados e funciona perto da temperatura do zero absoluto

# Aceleradores Quânticos

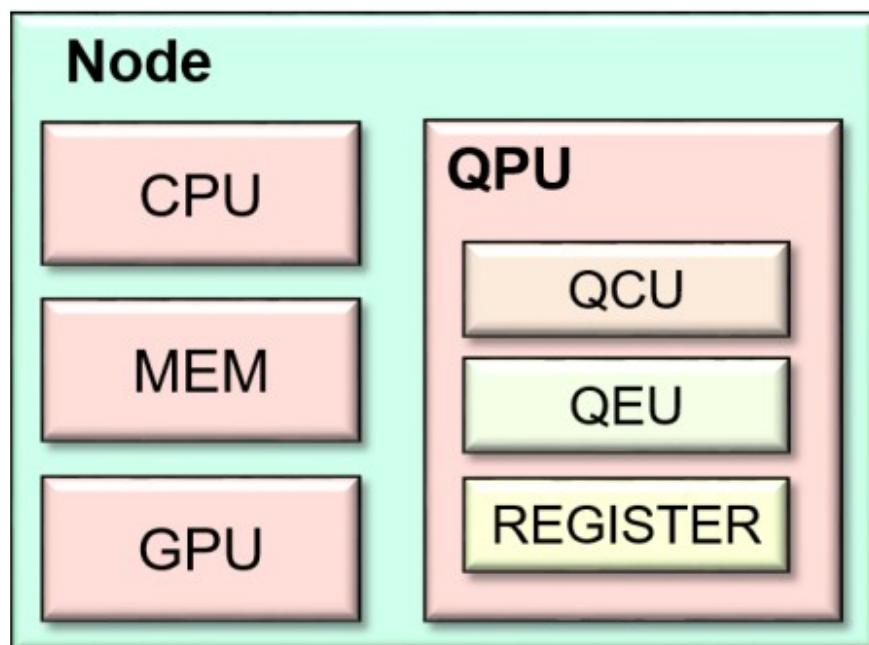


Figure 1. A typical node within the QPU-accelerated HPC system

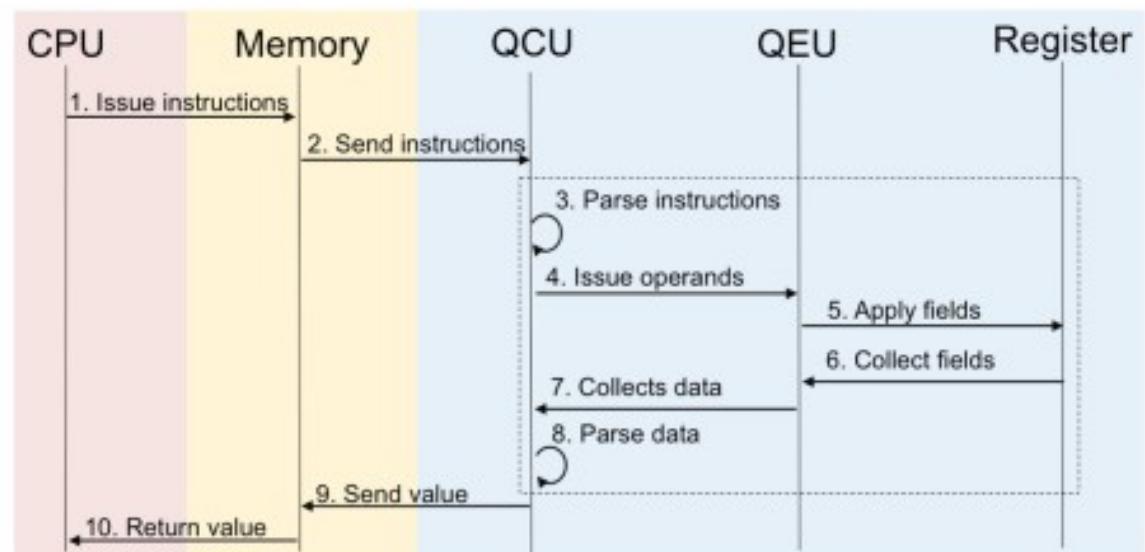
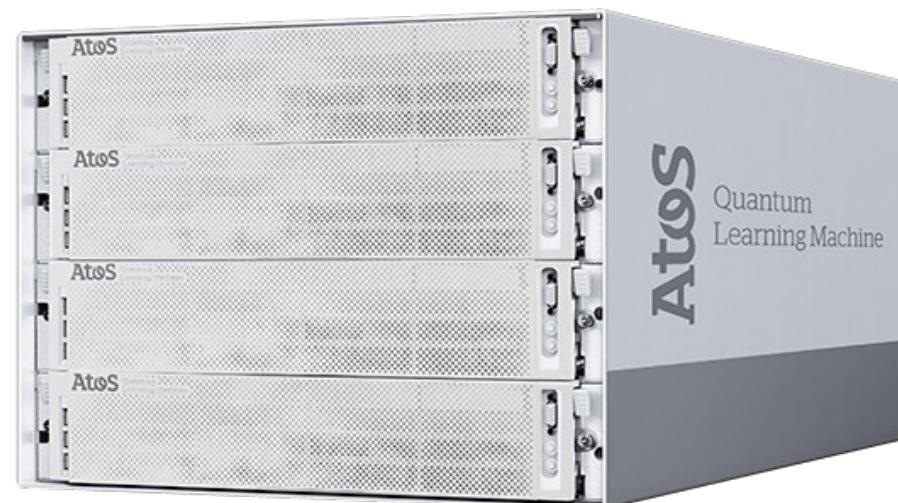


Figure 4. The execution model for a QPU-accelerated host node passing instruction from a CPU to memory to a QPU.

Quantum Accelerators for High-Performance Computing  
Systems, <https://arxiv.org/pdf/1712.01423.pdf>

# Simulador Quântico



<https://atos.net/en/solutions/quantum-learning-machine>

## A global strategy for QC

- 1 | Quantum Programming Platform** ➤ Complete programming and simulation environment for quantum software/hardware developers and for education/training
- 2 | Quantum Expert Consulting Services** ➤ Assisting our customers in discovering Quantum Computing, detecting relevant use cases, assessing quantum implementation benefits on the QLM simulator
- 3 | New Generation Architectures** ➤ Designing the new quantum-powered accelerators for supercomputers or hybrid systems
- 4 | Quantum Algorithms** ➤ Atos' own research, focused on Quantum Machine Learning, one of the most promising application areas of QC
- 5 | Quantum Safe Cryptography** ➤ Preparing the cryptographies and hardware security modules, resistant to quantum computer attacks

A Atos e o SENAI CIMATEC criaram o Latin America Quantum Computing Center (LAQCC). O Atos QLM aproveitará o hardware clássico de supercomputação para simular sistemas quânticos e algoritmos de até 35 qubits. A Atos pretende lançar um acelerador quântico até 2023.

[https://atos.net/pt-br/2021/comunicados-de-imprensa\\_2021\\_05\\_05/atos-e-senai-cimatec-lancam-centro-de-excelencia-em-computacao-quantica-no-brasil](https://atos.net/pt-br/2021/comunicados-de-imprensa_2021_05_05/atos-e-senai-cimatec-lancam-centro-de-excelencia-em-computacao-quantica-no-brasil)

# Algoritmos quânticos

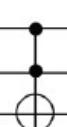
- Algoritmo de Fatoração de Shor:
  - *fatorar exponencialmente mais rápido que o melhor algoritmo conhecido atualmente rodando em um computador clássico.*
- Algoritmo Quântico de Lov Grover:
  - *algoritmo de busca em listas desordenadas com complexidade O(1).*
- O problema de Deutsch-Jozsa:
  - “*Consiste em saber, se uma determinada função  $f(x)$  é constante ou balanceada. Uma função constante é quando  $f(x)$  é igual para qualquer  $x$ , e uma função balanceada é quando  $f(x)$  é diferente*”.

$x$	$f_1(x)$	$f_2(x)$	$f_3(x)$	$f_4(x)$
0	1	0	0	1
1	1	0	1	0

Computador Clássico faz operação duas vezes.

Computador Quântico apenas uma vez.

# Portas/Operações Quânticas

Gate name	# Qubits	Circuit Symbol	Unitary Matrix	Description
Hadamard	1		$\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$	Transforms a basis state into an even superposition of the two basis states.
T	1		$\begin{bmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{bmatrix}$	Adds a relative phase shift of $\pi/4$ between contributing basis states. Sometimes called a $\pi/8$ gate, because diagonal elements can be written as $e^{-i\pi/8}$ and $e^{i\pi/8}$ .
CNOT	2		$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$	Controlled-not; reversible analogue to classical XOR gate. The input connected to the solid dot is passed through to make the operation reversible.
Toffoli (CCNOT)	3		$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$	Controlled-controlled-not; a three-qubit gate that switches the third bit for states where the first two bits are 1 (that is, switches $ 110\rangle$ to $ 111\rangle$ and vice versa).
Pauli-Z	1		$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$	Adds a relative phase shift of $\pi$ between contributing basis states. Maps $ 0\rangle$ to itself and $ 1\rangle$ to $- 1\rangle$ . Sometimes called a "phase flip."
Z-Rotation	1		$\begin{bmatrix} e^{-i\theta/2} & 0 \\ 0 & e^{i\theta/2} \end{bmatrix}$	Adds a relative phase shift of (or rotates state vector about z-axis by) $\theta$ .
NOT	1		$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$	Analogous to classical NOT gate; switches $ 0\rangle$ to $ 1\rangle$ and vice versa.

<https://www.nap.edu/read/25196/chapter/4>