

**PSPD**

**ATIVIDADE  
EXTRACLASSE**

**ARQUITETURA DE  
CPUS E GPUS**

# **INTEGRANTES**

André Emanuel Bispo da Silva - 221007813  
Vinicius de Oliveira - 202017263

# **VISÃO GERAL SOBRE CPUS**

# Visão Geral sobre CPUs

- Executa operações aritméticas, controle e I/O.
- Evolução: de single-core a multi-core.
- Etapas do ciclo: busca, decodificação, execução.

# Visão Geral sobre CPUs

CPU	Ano	Tecnologia	Arquitetura / Palavras	Frequência (MHz)	Transistores / Componentes	Núcleos
UNIVAC 1103	1953	Válvulas eletrônicas	36 bits	~0.1 MHz	~18.000 válvulas	1
IBM 7090	1959	Transistores	36 bits	~0.5 MHz	~50.000 transistores	1
Intel 4004	1971	LSI (circuito integrado)	4 bits	0.74 MHz	2.300 transistores	1
Intel 8080	1974	NMOS	8 bits	2 MHz	6.000 transistores	1
Intel 8086	1978	HMOS	16 bits	5-10 MHz	29.000 transistores	1
Motorola 68000	1979	HMOS	16/32 bits	8 MHz	68.000 transistores	1

# Visão Geral sobre CPUs

Intel 80386	1985	CMOS	32 bits	16-33 MHz	275.000 transistores	1
Intel Pentium	1993	BiCMOS	32 bits (x86)	60-300 MHz	3,1 milhões de transistores	1
AMD Athlon XP	2001	CMOS	32 bits	1.3-2.3 GHz	~37,5 milhões de transistores	1
Intel Core 2 Duo	2006	65 nm CMOS	64 bits (x86-64)	~2.0 GHz	~291 milhões de transistores	2
Intel i7 (Nehalem)	2008	45 nm CMOS	64 bits, multi-core	2.66-3.33 GHz	~731 milhões de transistores	4-8
Apple M1	2020	5 nm FinFET	64 bits (ARM)	3.2 GHz	16 bilhões de transistores	8 (4P+4E)
AMD Ryzen 9 7950X	2022	5/6 nm	64 bits (x86-64)	Até 5.7 GHz	~13,1 bilhões de transistores	16
Intel Core Ultra 9 (Meteor Lake)	2023	Intel 4 (7nm equiv.)	64 bits, heterogênea	Até 5.5 GHz	~15 bilhões (estimado)	16 (6P+8E +2LPE)

# Visão Geral sobre CPUs

- Multi-core para execução simultânea.
- Frequência de clock limitada por fatores térmicos.

# **ARQUITETURA DAS GPUS**

# Arquitetura das GPUs

- Inicialmente para renderização gráfica.
- Hoje: uso em IA, simulações, GPGPU.
- Altamente paralelas → milhares de threads.
- Foco em throughput vs. baixa latência (CPUs).

# Arquitetura das GPUs

Característica	CPU	GPU
Núcleos	Poucos, potentes	Muitos, simples
Latência	Baixa	Alta
Paralelismo	Limitado	Massivo
Aplicação	Tarefas gerais	Cálculos vetoriais, IA, ML

# Arquitetura das GPUs

- Hierarquia: núcleos → SMs → clusters.
- Unidade de execução: SIMD.
- Tensor Cores (deep learning).
- Exemplo: arquitetura Hopper da NVIDIA.

# Arquitetura das GPUs

- G80: 1<sup>a</sup> com CUDA.
- Fermi: suporte a C++, funções complexas.
- Kepler: SMX, Hyper-Q.
- Volta: introduz Tensor Cores.
- Turing/Ampere: RT Cores (Ray Tracing).
- Grace Hopper: integração CPU+GPU.

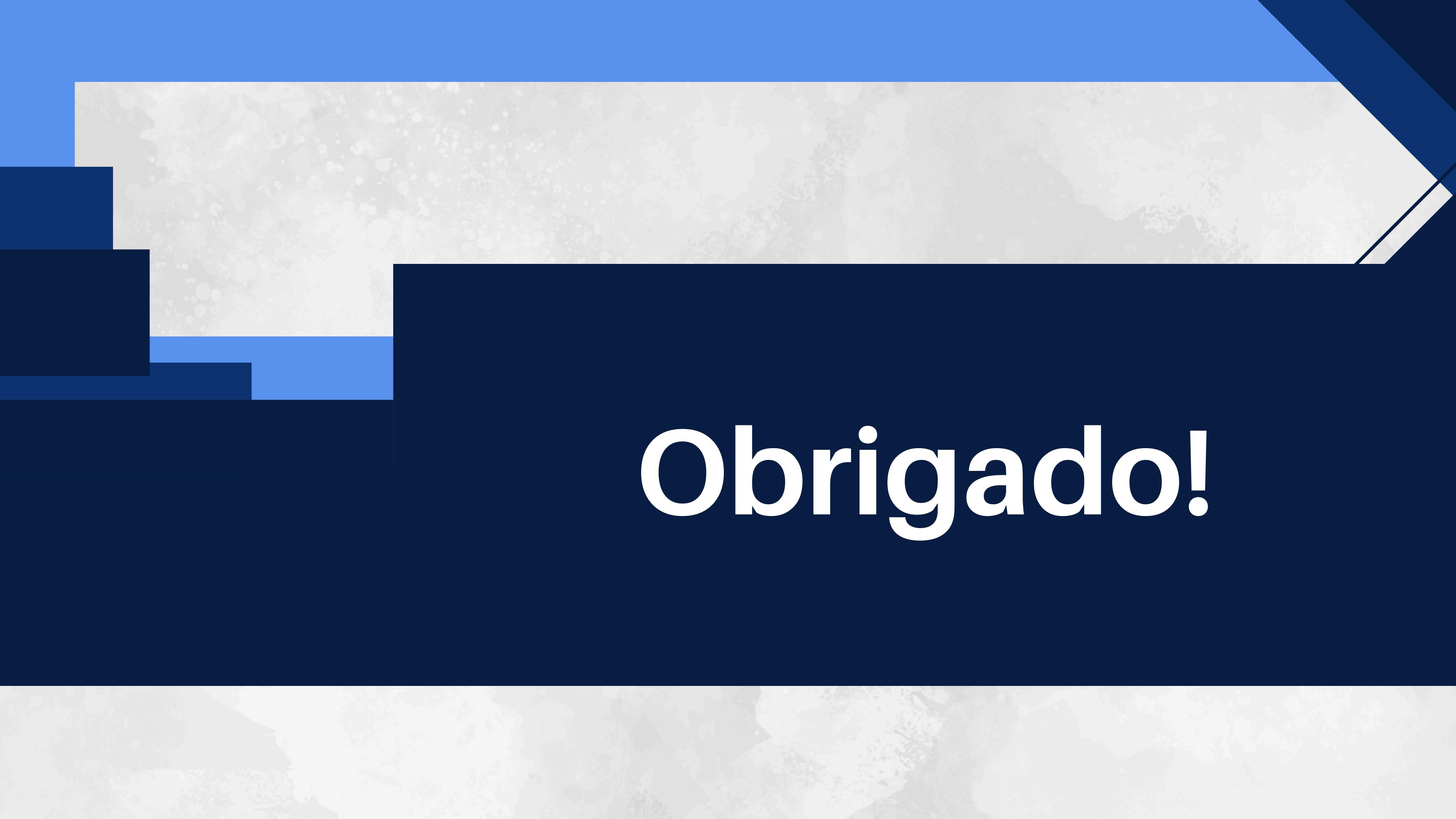
# **MODELO DE PROGRAMAÇÃO**

# CUDA

- Desenvolvido pela NVIDIA.
- Controle direto de threads e blocos.
- Sintaxe: <<<N, M>>>.
- Suporte a linguagens como C++, Python.
- Exemplo visual: soma vetorial.

# OpenMP

- API de programação paralela para CPU e GPU.
- Modelo fork-join.
- Suporte a instruções SIMD (Single Instruction, Multiple Data),
- Diretivas como `#pragma omp parallel, simd, target`.
- Suporte a offloading para GPU (v4.0+).



# Obrigado!