

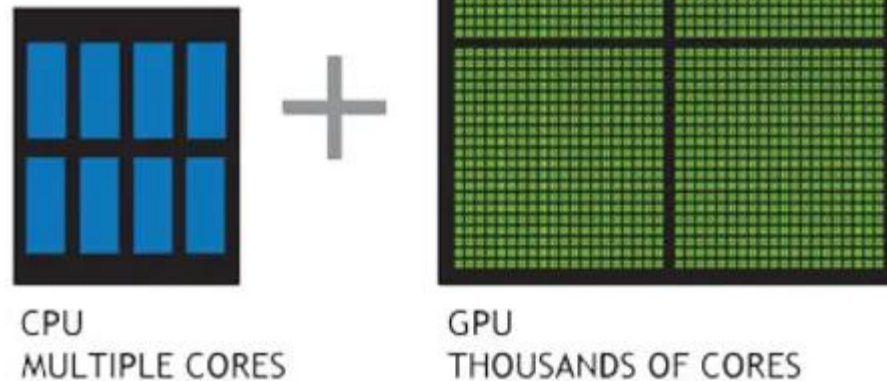
ACELERACIÓN GPU

PARA DEEP LEARNING EN AMD

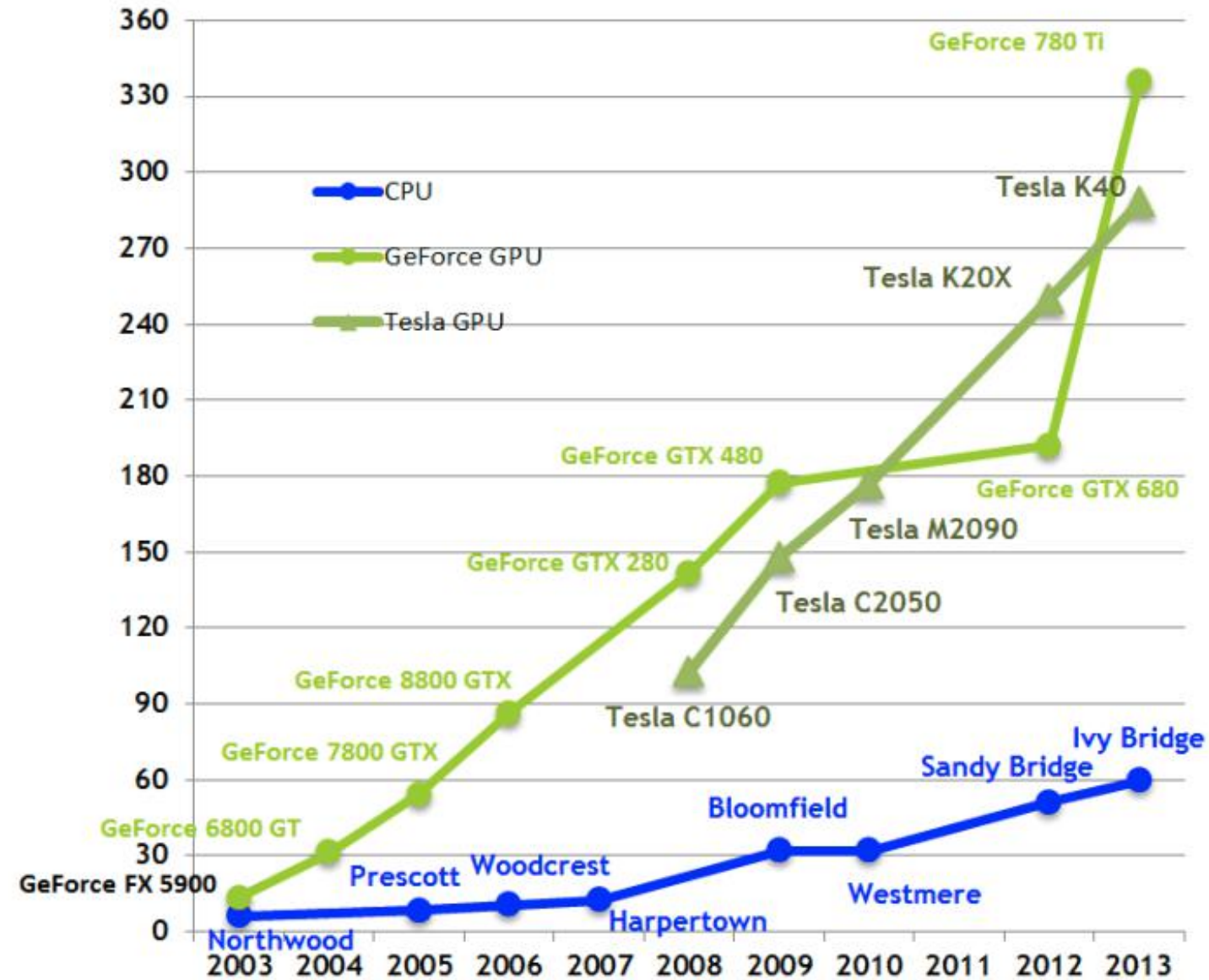
¿QUÉ ES LA ACCELERACIÓN GPU?

CPU VS GPU

- Las CPUs para pocas tareas secuenciales y complejas.
- Las GPUs para muchas tareas paralelas simples y específicas.
- Ferrari VS. Camión



Theoretical GB/s



[Comparison of bandwidth for CPUs and GPUs over time.]

SUPREMACÍA DE NVIDIA

- Dominancia del mercado (IBM usa NVIDIA)
- Desarrollo de hardware específico
- Drivers y frameworks

Entonces... ¿qué pasa si tengo AMD?

ALTERNATIVAS

- ROCm
- plaidML
- Cloud

ROCm

- Plataforma universal de aceleración GPU
- Open Source
- Fork de tensorflow (importante para deep learning)
- Independiente del lenguaje de programación
- Varios sistemas operativos soportados:
 - Ubuntu
 - CentOS
 - SLES 15 SP1
 - RHEL (Red Hat Enterprise)
- Algunas características:
 - Multi-GPU
 - Concurrencia de procesos
 - Grandes asignaciones de memoria

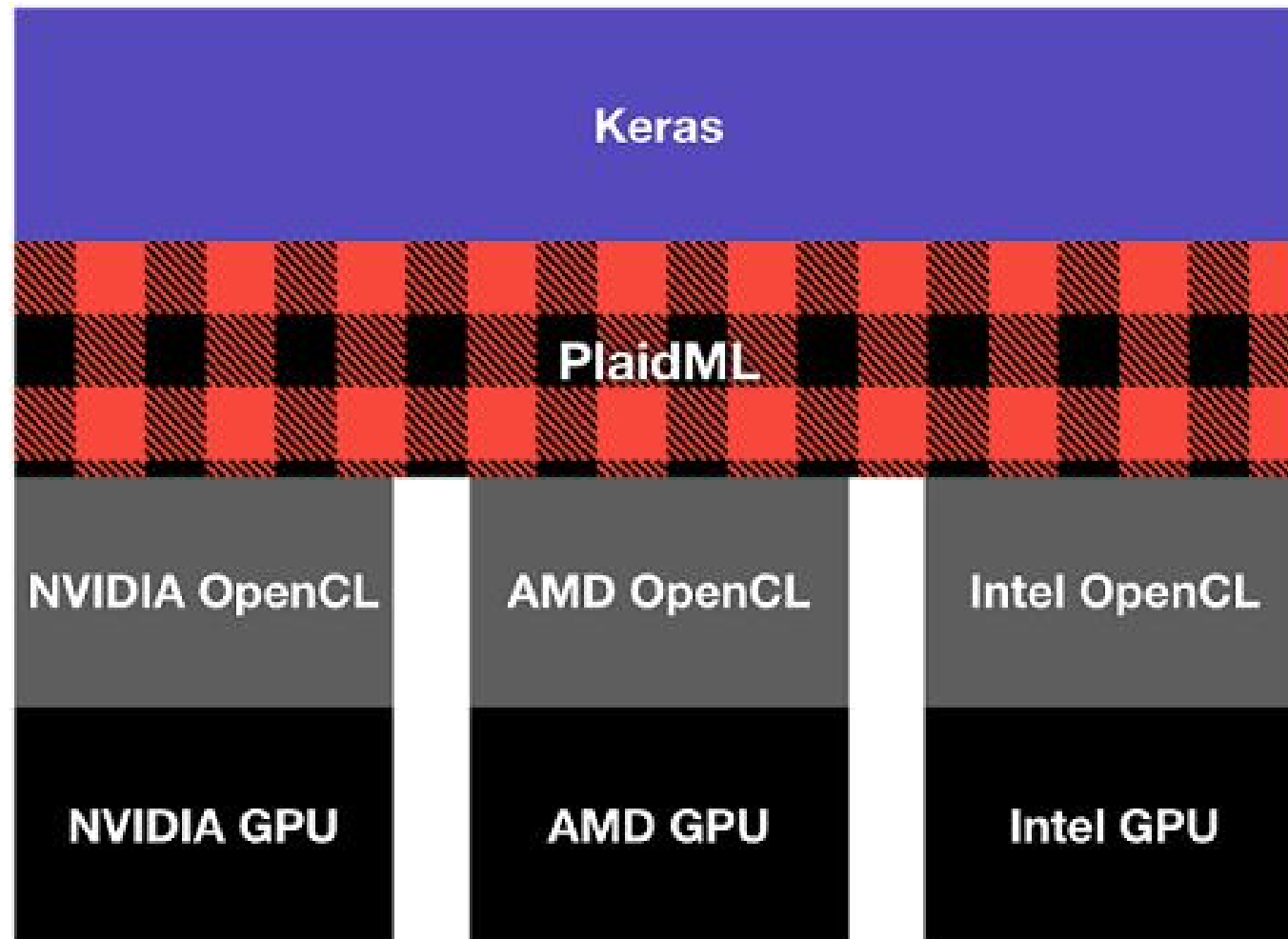
GPUs SOPORTADAS

- GFX8 GPUs
 - "Fiji" chips, por ejemplo: AMD Radeon R9 Fury X y Radeon Instinct MI8
 - "Polaris 10" chips, por ejemplo: AMD Radeon RX 580 y Radeon Instinct MI6
- GFX9 GPUs
 - "Vega 10" chips, por ejemplo: AMD Radeon RX Vega 64 y Radeon Instinct MI25
 - "Vega 7nm" chips, por ejemplo: Radeon Instinct MI50, Radeon Instinct MI60 o AMD Radeon VII
- GPUs "experimentales":
 - GFX8 GPUs
 - "Polaris 11" chips, por ejemplo: AMD Radeon RX 570 y Radeon Pro WX 4100
 - "Polaris 12" chips, por ejemplo: AMD Radeon RX 550 y Radeon RX 540
 - GFX7 GPUs
 - "Hawaii" chips, por ejemplo: AMD Radeon R9 390X y FirePro W9100

PlaidML

- Compilador de tensores
- Software libre
- Usable como backend para Keras
- Buen funcionamiento sobre GPUs
- Soportado por la mayoría de sistemas operativos
- Uso del driver OpenCL

ESTRUCTURA



CLOUD: LA FORMA FÁCIL

- Google Cloud
- Google Colab
- Azure
- AWS

Para terminar...

EJEMPLO USANDO AMD

DEMOSTRACIÓN USANDO NVIDIA