



UNIVERSITAT POLITÈCNICA DE CATALUNYA  
BARCELONATECH

Departament d'Arquitectura de Computadors

# Tarjetas Gráficas y Aceleradores

## Ejemplos Comerciales

### Agustín Fernández

Departament d'Arquitectura de Computadors

Facultat d'Informàtica de Barcelona

Universitat Politècnica de Catalunya

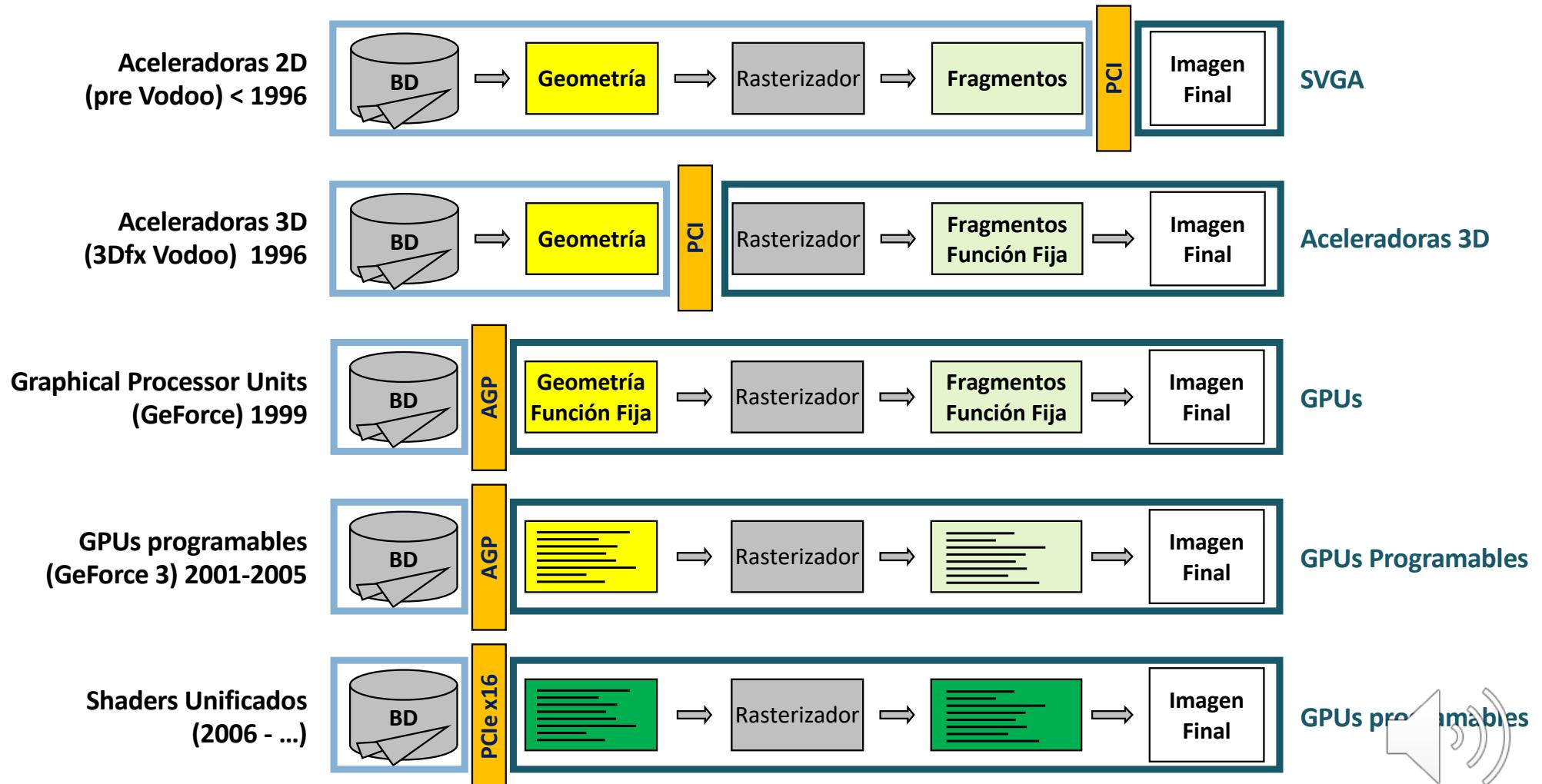


# Evolución de las Tarjetas Gráficas

- Es muy complejo seguir la historia de los dispositivos hardware.
- Es muy confuso distinguir entre:
  - Fecha en que una tecnología es anunciada por 1<sup>a</sup> vez.
  - Fecha en que un chip basado en esa tecnología es anunciado por 1<sup>a</sup> vez.
  - Fecha en que el chip empieza a venderse.
  - Fecha en que una tarjeta con ese chip es anunciada por 1<sup>a</sup> vez.
  - Fecha en que los primeros prototipos de esa tarjeta están disponibles.
  - Fecha en que la tarjeta empieza a venderse.
- ¡Esto no sólo se aplica a las tarjetas gráficas!

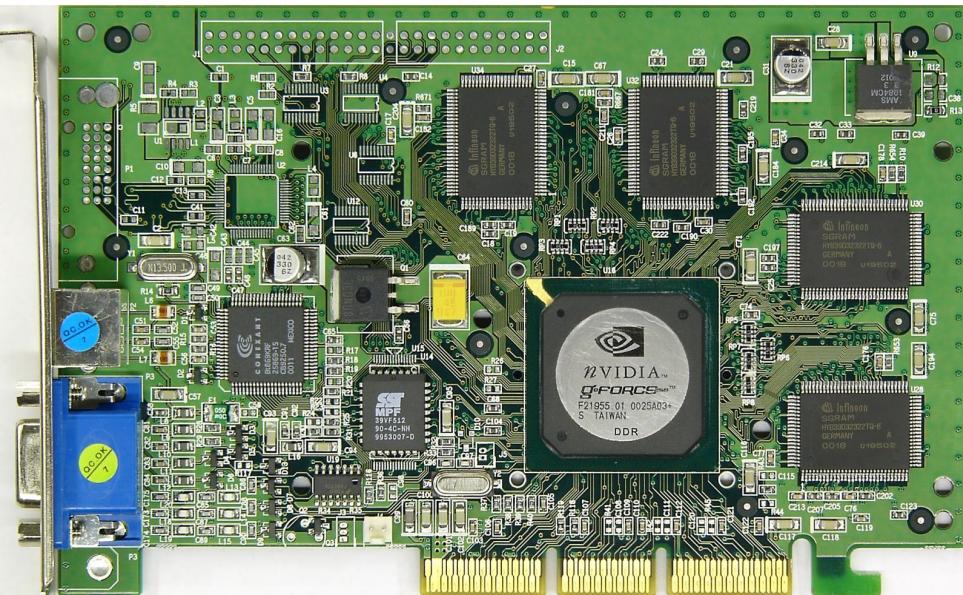


# Evolución de la Aceleración 3D



# NVIDIA GeForce 256

- **Año de Lanzamiento:** 1999
- **Conexión:** AGP 4x y PCI
- **Versión DirectX:** 7
- **Versión OpenGL:** 1.3
- **Shader Model:** T&L (hardware)
- **Memoria:** 32-64 MB DDR, 128 bits
- GPU (NV10) de 256 bits
- **Transistores:** 23 millones (220nm)
- **Frecuencia:** 120 MHz
- **Frecuencia Memoria:** 300 MHz
- **Ancho de Banda:** 4,8 GB/s
- **Configuración:** 4 Pixel Shaders, 4 Texture Units y 4 ROPs
- 480 MOPs



1<sup>a</sup> GPU de la historia

Acabó con la competencia.  
Sólo aguantó ATI.

Pentium III a 450 MHz (Feb 1999)

# The 3D Graphics Pipeline [1999]

**Application Tasks.** The application controls the movements of objects (including the camera) and their interaction in the 3D world. Problems such as realistic physics calculations (objects have momentum) and collision detection (did my race car bump the wall or another car?) affect how an object moves in the scene.

**Scene-level Tasks (culling, LOD, display list).** Scene-level tasks include object level culling (car #3 is completely behind the camera, so it doesn't need to be drawn to the next stage), selecting the appropriate detail level (car #1 is far away, so a low-detail model is best), and creating the list of all objects in the current camera view.

**Transform.** The transform engine converts 3D objects from one frame of reference to a new frame of reference. The system must transform them to the current view before performing the following steps (lighting, triangle setup and rendering). Every object that is displayed and some objects that are not displayed must be transformed every time the scene is redrawn.

**Lighting.** Lighting is the next step in the 3D pipeline and provides high visual impact. Lighting effects are essential for enhancing the realism of a scene and bringing rendered images one more step closer to our perception of the real world.

**Triangle Setup and Clipper.** The triangle setup engine is a floating-point math processor that receives vertex data and calculates all of the parameters that the rendering engine will require. This unit 'sets up' the triangle for the rendering engine.

**Rendering.** Rendering is calculating the correct color for each pixel on the screen, given all of the information delivered by the setup engine. The rendering engine must consider the color of the object, the color and direction of light hitting the object, whether the object is translucent, and what textures apply to the object.

Propaganda de la NVIDIA GeForce 256

3D Application  
and API  
  
3D Graphics  
Pipeline



# The 3D Graphics Pipeline [1999]

**Application Tasks.** La aplicación controla los movimientos de los objetos (incluyendo la cámara) y su interacción en la escena 3D. Problemas tales como los cálculos físicos realistas (los objetos tienen inercia) y la detección de colisiones (mi coche de carreras ¿ha golpeado la pared o a otro coche?), afectan a cómo se mueve un objeto en la escena.

**Scene-level Tasks (culling, LOD, display list).** Las tareas a nivel de escena incluyen el nivel de culling (el coche #3 está detrás de la cámara, por lo que no necesita ser enviado a la siguiente etapa), seleccionar el nivel de detalle apropiado (el coche #1 está lejos, por lo que un bajo nivel de detalle es suficiente), y la creación de la lista con todos los objetos visibles por la cámara.

**Transform.** El motor de transformación convierte los datos 3D de un marco de referencia a otro. El sistema debe transformar los datos a la vista actual antes de realizar los siguientes pasos (iluminación, triangle setup y rendering). Cada objeto que se muestra y algunos objetos que no se muestran deben transformarse cada vez que la escena se vuelve a dibujar.

**Lighting.** La iluminación es el siguiente paso en el pipeline 3D y proporciona un alto impacto visual. Los efectos de iluminación son esenciales para mejorar elrealismo de una escena y dejan las imágenes renderizadas un paso más cerca de nuestra percepción del mundo real.

**Triangle Setup and Clipping.** El triangle setup es un procesador en coma flotante que recibe los datos de cada vértice y calcula todos los parámetros necesarios para el renderizado. Esta unidad prepara el triángulo para el motor de renderizado.

**Rendering.** Renderizar es calcular el color correcto para cada píxel de la pantalla, a partir de toda la información generada previamente. El motor de renderizado debe tener en cuenta el color del objeto, el color y la dirección de la luz que incide en el objeto, si el objeto es translúcido, y qué texturas se aplican al objeto.

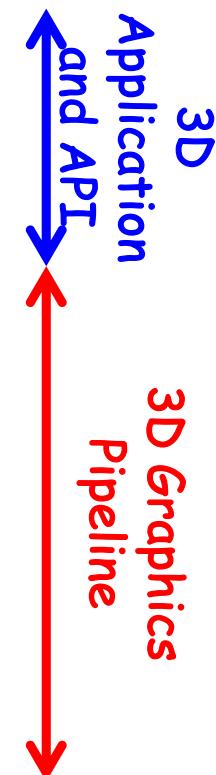
3D Application and API  
3D Graphics Pipeline



# The 3D Graphics Pipeline [1999]

Application Tasks (move objects according to application, move/aim camera).	CPU	CPU	CPU	CPU
Scene-level calculations (object level culling, select detail level, create object mesh).	CPU	CPU	CPU	CPU
Transform.	CPU	CPU	CPU	GPU
Lighting.	CPU	CPU	CPU	GPU
Triangle Setup and Clipping.	CPU	Graphics Processor	Graphics Processor	GPU
Rendering.	Graphics Processor	Graphics Processor	Graphics Processor	GPU

1996            1997            1998            1999  
ATI 3D        NVIDIA        ATI        NVIDIA  
Rage II      Riva128      Rage 128    GeForce 256



# NVIDIA GeForce 6800

- **Año de Lanzamiento:** 2004
- **Conexión:** PCIe ×16 nativa
- **Versión DirectX:** 9.0c
- **Versión OpenGL:** 2.0
- **Salidas de vídeo:** 2
- **RAMDAC:** 2 × 400 MHz
- **Shader Model:** 3.0
- **Consumo:** 70-80 W
- **MultiGPU:** Tecnología SLI
- Filtrado Anisotrópico (×16)
- Antialiasing (×8)
- Decodificación hardware de vídeo



PassMark - G3D Mark: 879



# NVIDIA GeForce 6800

## Utiliza la GPU NV40

- **Tecnología de integración:** 0,13 micras
- **Número de transistores:** 222 millones
- **Frecuencia:** 400 MHz
- **Vertex shaders:** 6, puede procesar 600 millones de vértices/seg.
- **Fragment shaders:** 16, puede procesar 16 fragmentos por ciclo.
- **Potencia cálculo:** 120 GFLOPS

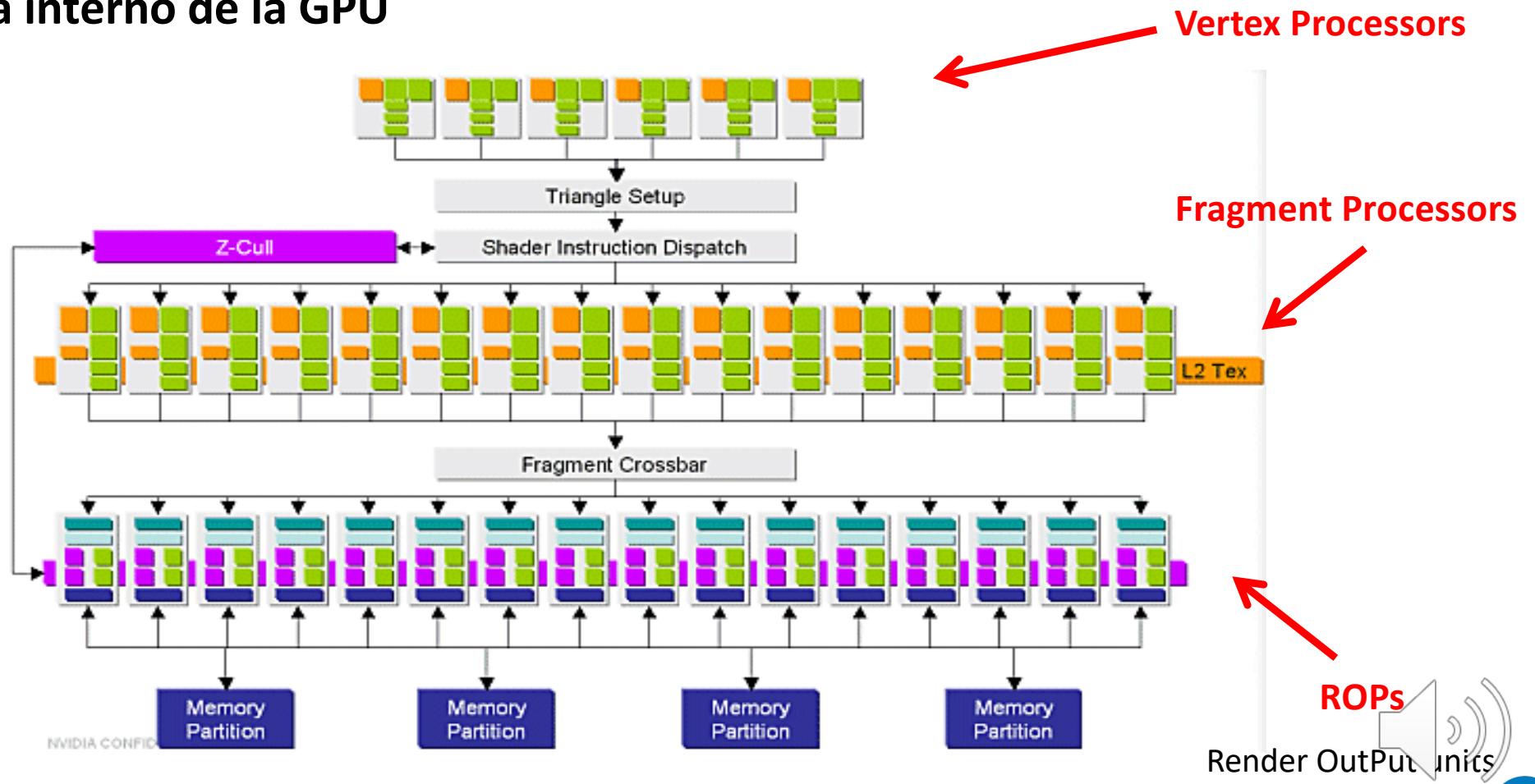
## Utiliza Memoria GDDR3

- **Tamaño:** 128 / 256 / 512 MB
- **Frecuencia:** 1100 MHz (550 x 2)
- **Anchura del bus de memoria:** 256 bits (en 4 canales)
- **Ancho de banda:** 35,2 GB/s

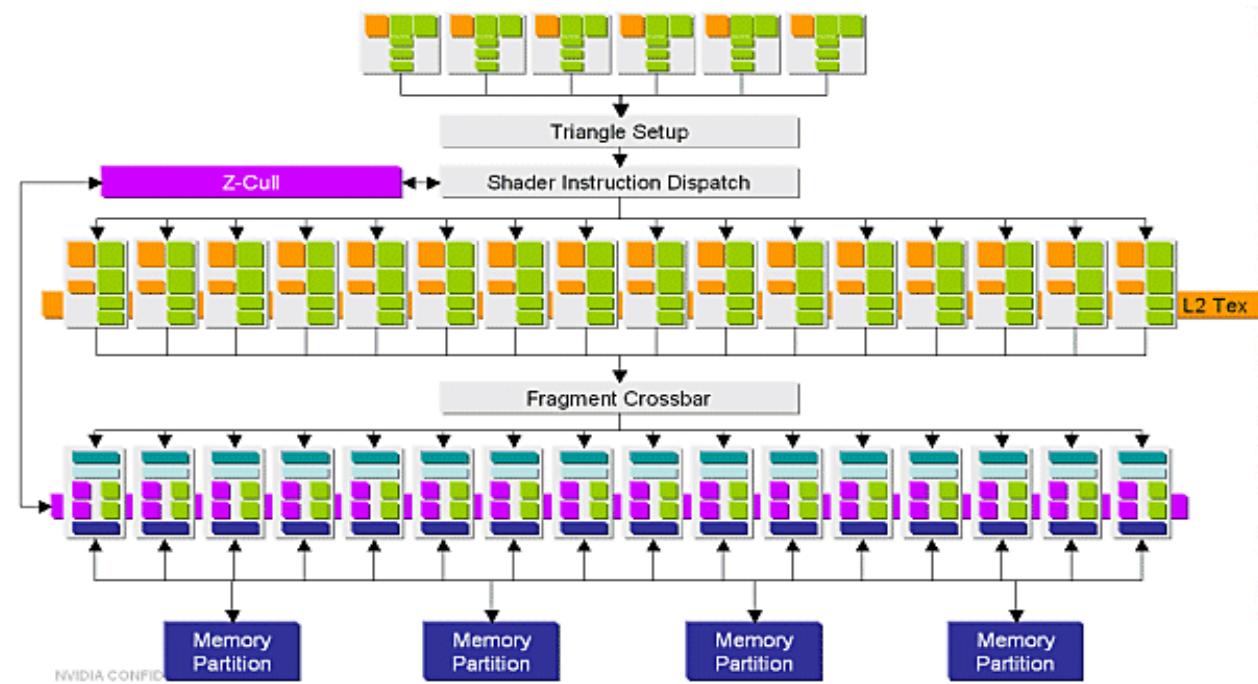
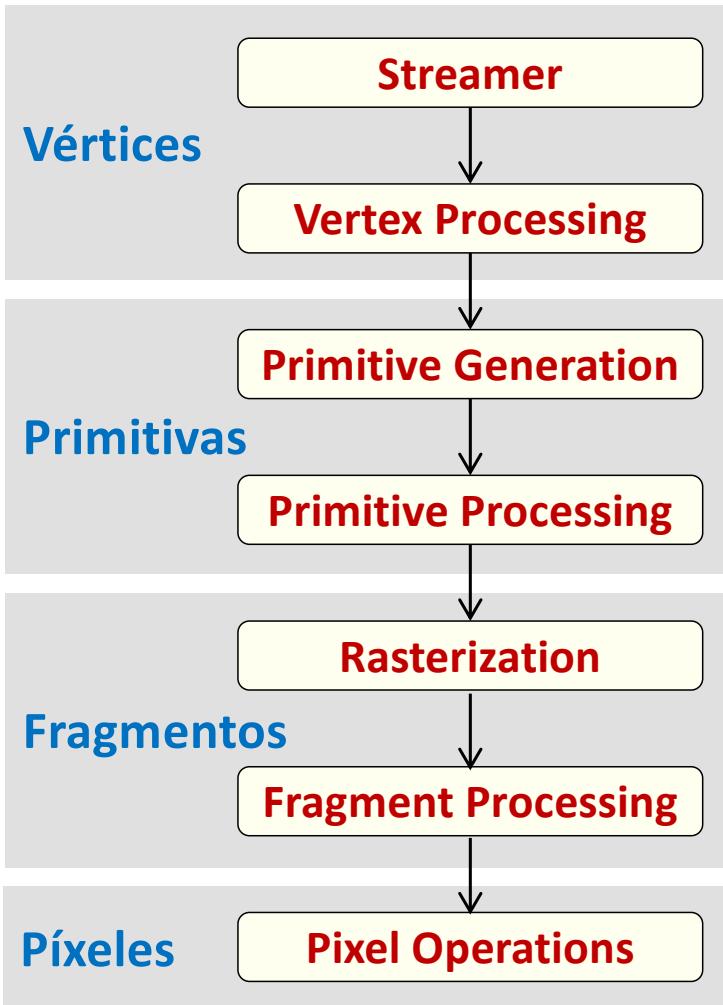


# NVIDIA GeForce 6800

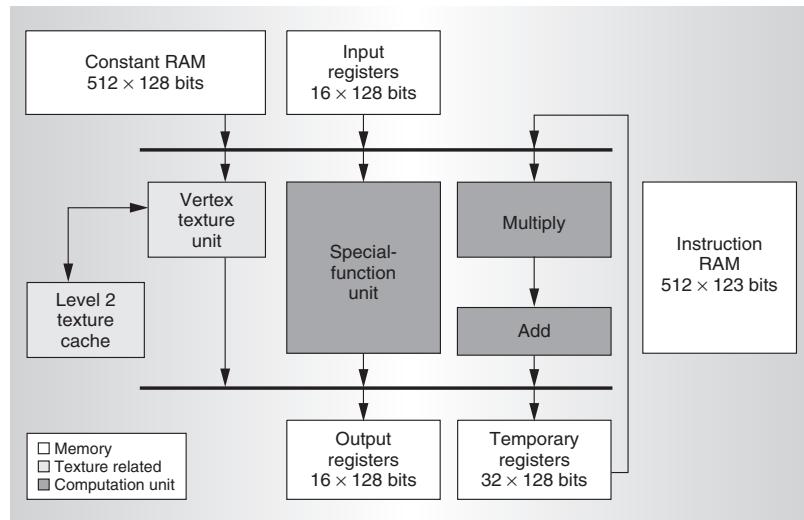
## Esquema interno de la GPU



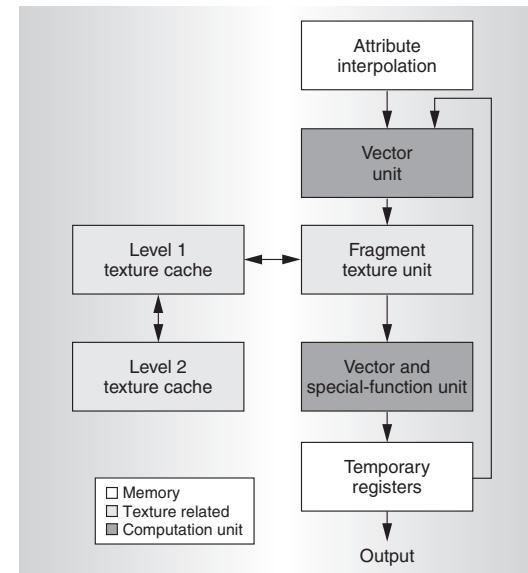
# El Pipeline Gráfico / NVIDIA GeForce 6800



# NVIDIA GeForce 6800



Vertex Processor



Fragment Processor

# NVIDIA GeForce 6800

## Tecnología SLI

- Permite la interconexión de 2 tarjetas gráficas **IDÉNTICAS** para que trabajen en paralelo.
- **Funcionamiento:** Cada tarjeta renderiza una parte de la imagen. Se comunican entre ellas para repartir la carga de trabajo.
- Requisitos necesarios:
  - Placa base con 2 slots PCIe 16x
  - Conector SLI
  - 2 tarjetas gráficas NVIDIA GeForce 6 idénticas y compatibles con SLI
  - Fuente de alimentación: **500W mínimo.**



# ATI Radeon X800 XT

- **Año de Lanzamiento:** 2004
- **Conexión:** AGP
- **Versión DirectX:** 9.0b
- **Versión OpenGL:** 2.0
- **Shader Model:** No soporta 3.0
- **Salidas de vídeo:** 1
- **MultiGPU:** ATI Crossfire (2005)
- **Consumo:** 70 W
- Filtrado Anisotrópico ( $\times 16$ )
- Antialiasing ( $\times 8$ )
- Decodificación hardware de vídeo



PassMark - G3D Mark: 600



# ATI Radeon X800 XT

## Utiliza la GPU R420

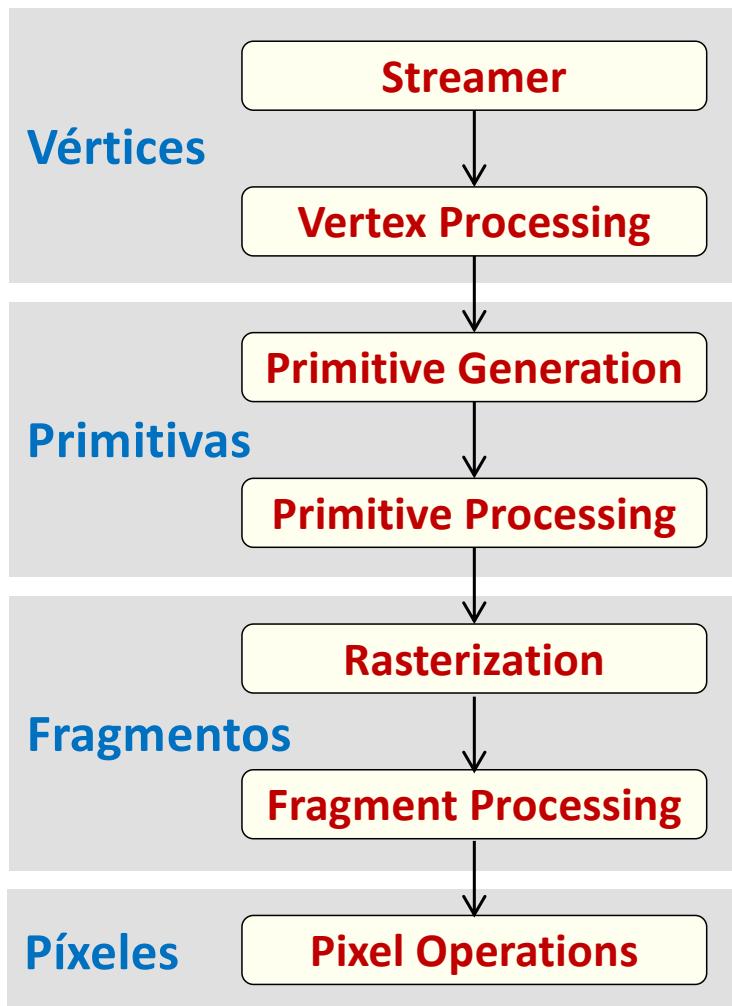
- **Tecnología de integración:** 0,13 micras
- **Número de transistores:** 160 millones
- **Frecuencia:** 520 MHz
- **Vertex shaders:** 6, puede procesar 780 millones de vértices/seg.
- **Fragment shaders:** 16
- La GPU dispone de instrucciones de control de flujo: saltos, bucles y subrutinas

## Utiliza Memoria GDDR3

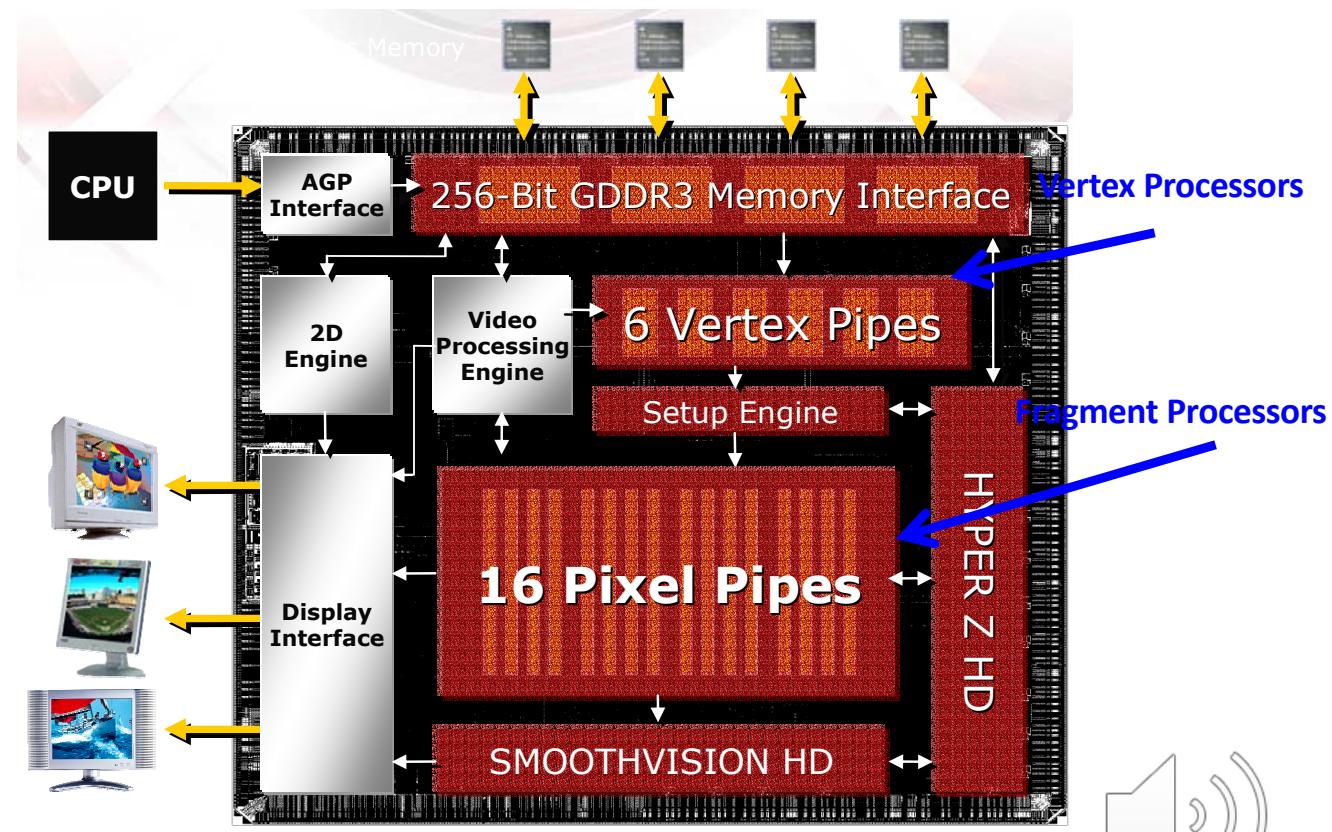
- **Tamaño:** 256 MB
- **Frecuencia:** 1120 MHz (560 x 2)
- **Anchura del bus de memoria:** 256 bits (en 4 canales)
- **Ancho de banda:** 35,8 GB/s



# ATI Radeon X800 XT

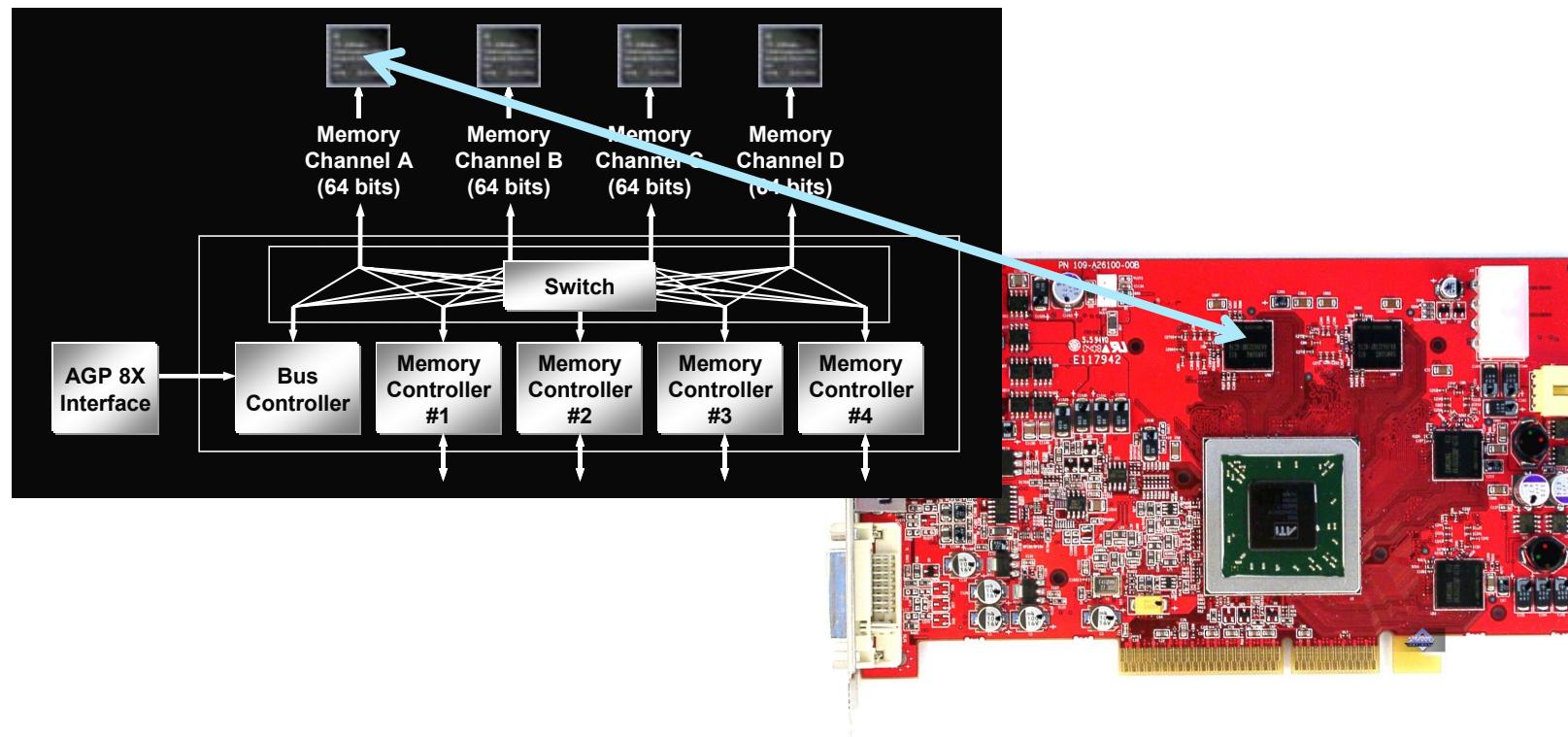


Esquema interno de la GPU



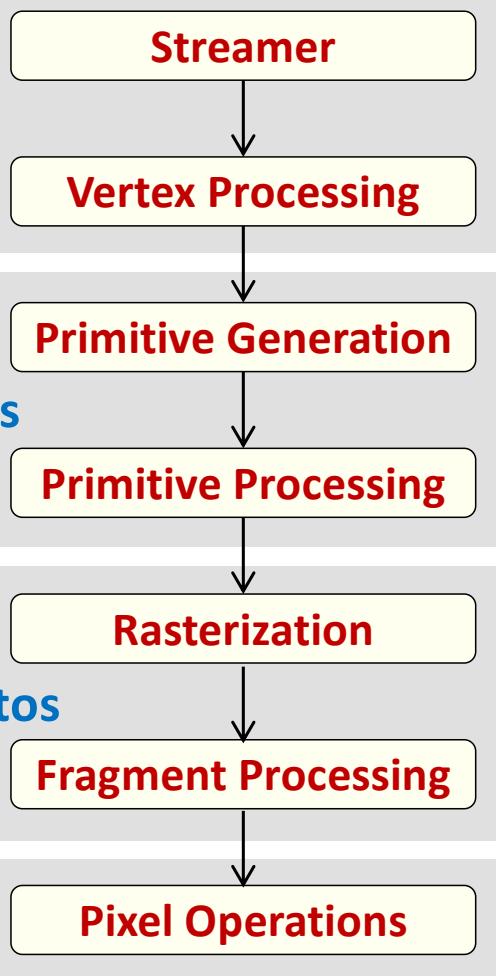
# ATI Radeon X800 XT

## Memoria de la GPU



# ATI Radeon X800 XT

Vértices

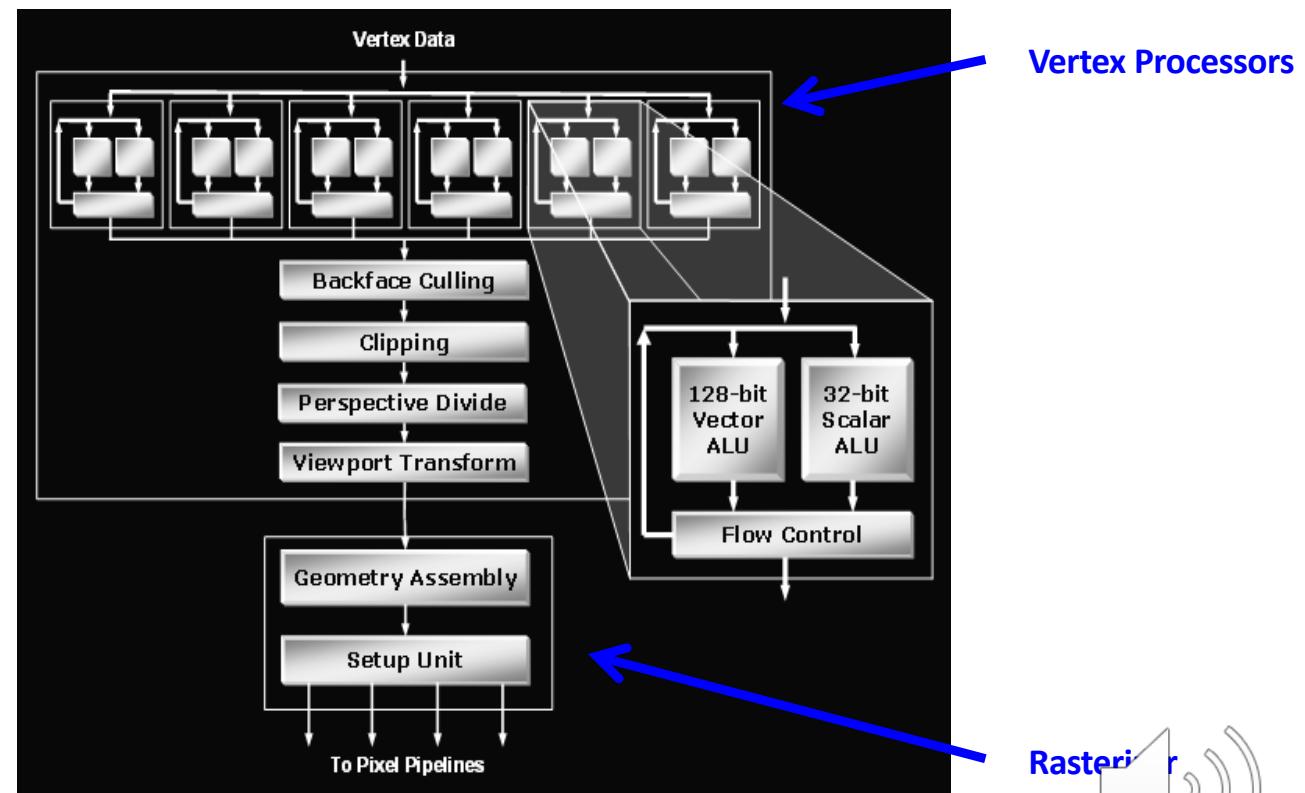


Primitivas

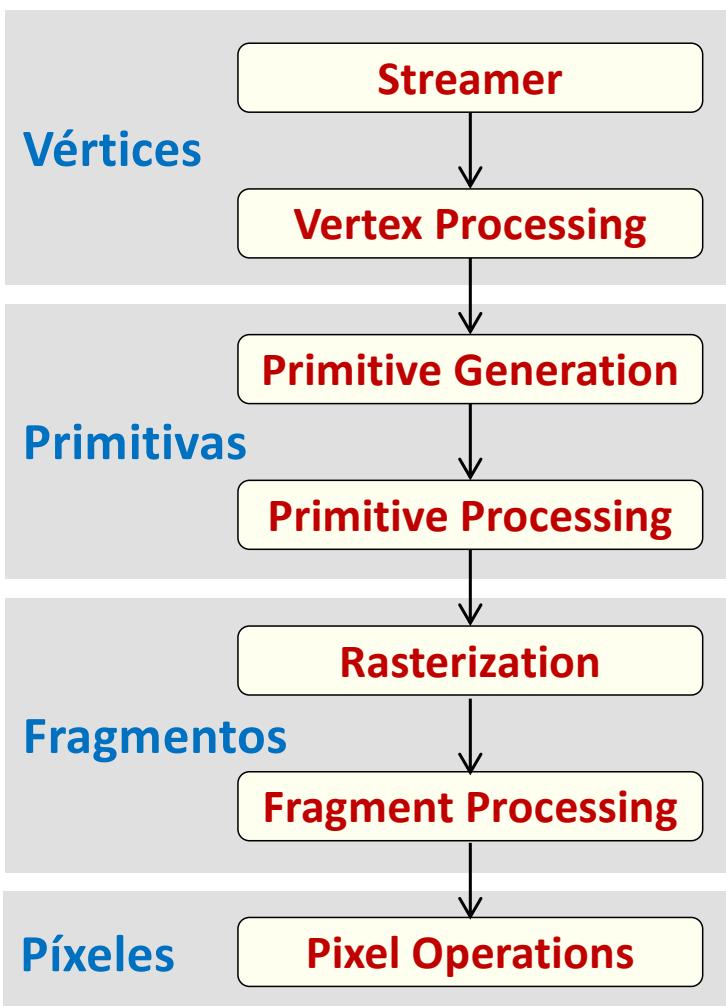
Fragmnetos

Píxeles

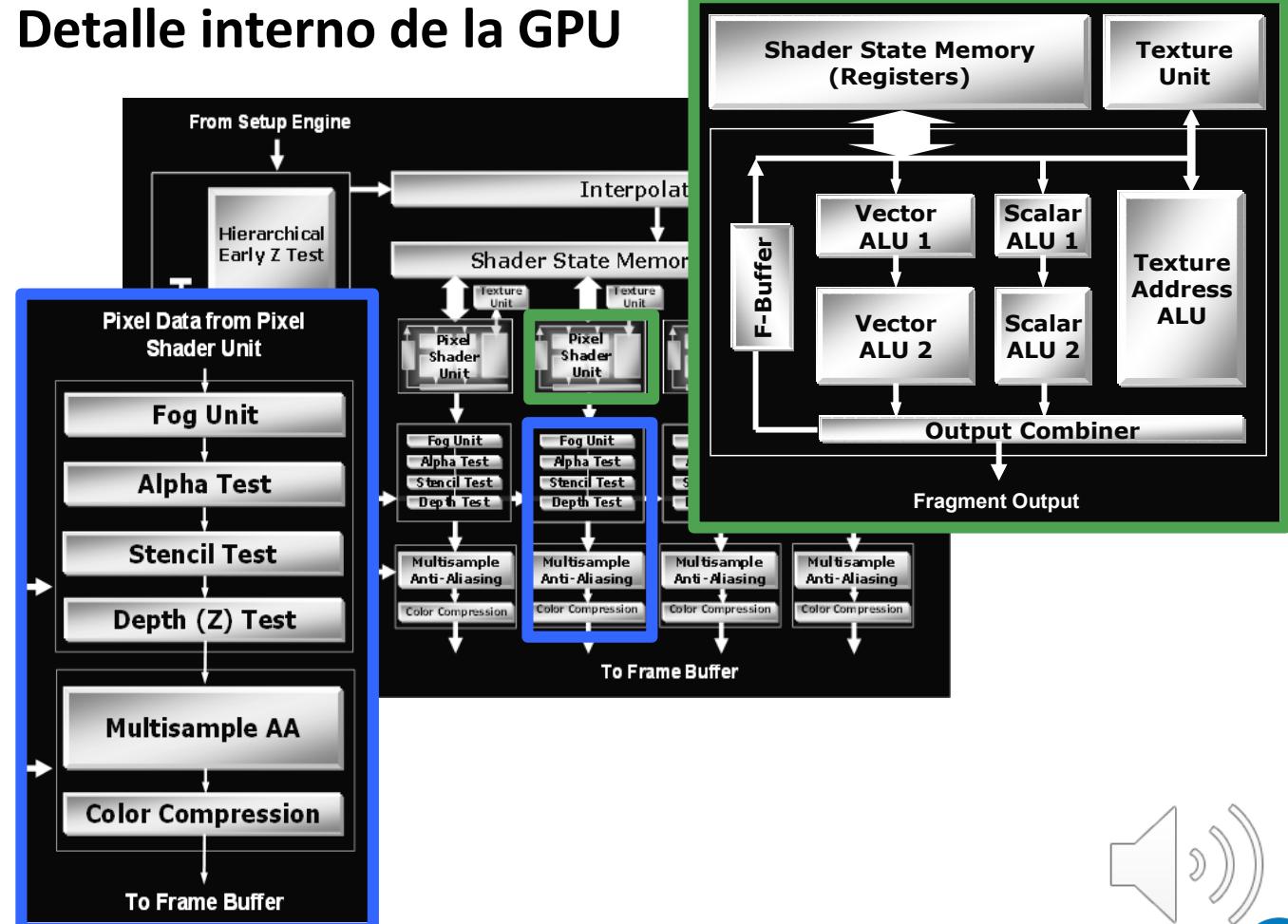
## Detalle interno de la GPU



# ATI Radeon X800 XT



## Detalle interno de la GPU



# ATI Radeon X800 XT

## Tecnología ATI Crossfire

- Similar al SLI de nVIDIA
- Permite la interconexión de 2 tarjetas gráficas **similares** para que trabajen en paralelo.
- Inicialmente la conexión se hacía a través del conector del vídeo
- Requisitos necesarios:
  - Placa base con 2 slots PCIe 16x
  - Placa base compatible con Crossfire
  - Fuente de alimentación adecuada.
- ¿Antecedentes?** Alternate Frame Rendering en la ATI Rage Fury MAXX (2 chips Rage 128)



# NVIDIA 8800

- **Año de Lanzamiento:** 2006-2007
- **Conexión:** PCIe ×16 2.0
- **Versión DirectX:** 10
- **Versión OpenGL:** 3.3
- **Shader Model:** 4.0
- **MultiGPU:** SLI
- **Resolución máxima:** 2560×1600



GeForce 8800 GTS 512

Familia: Tesla Architecture

PassMark - G3D Mark: 858



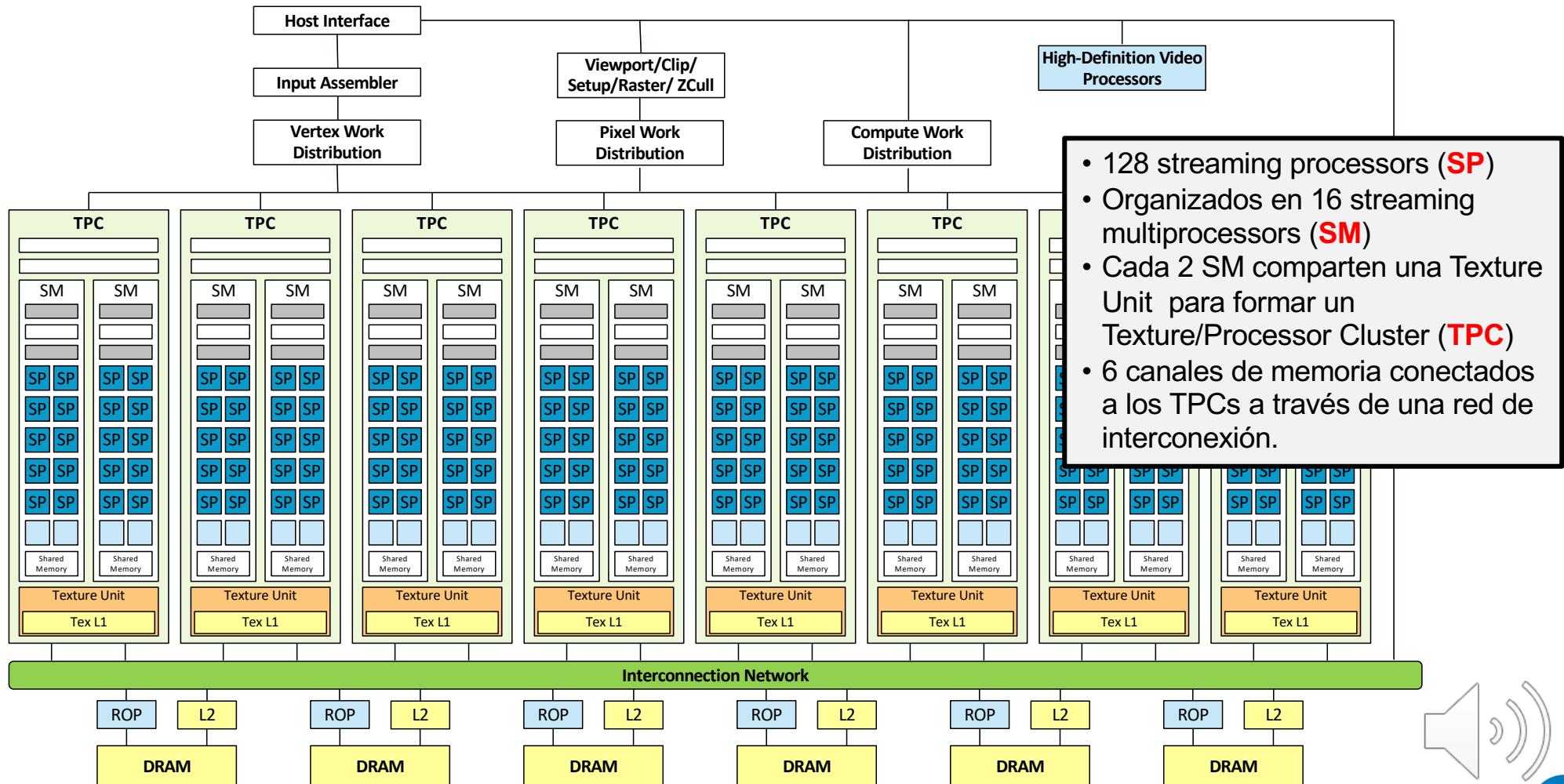
# NVIDIA 8800

## Especificaciones Técnicas nVIDIA 8800 GTS

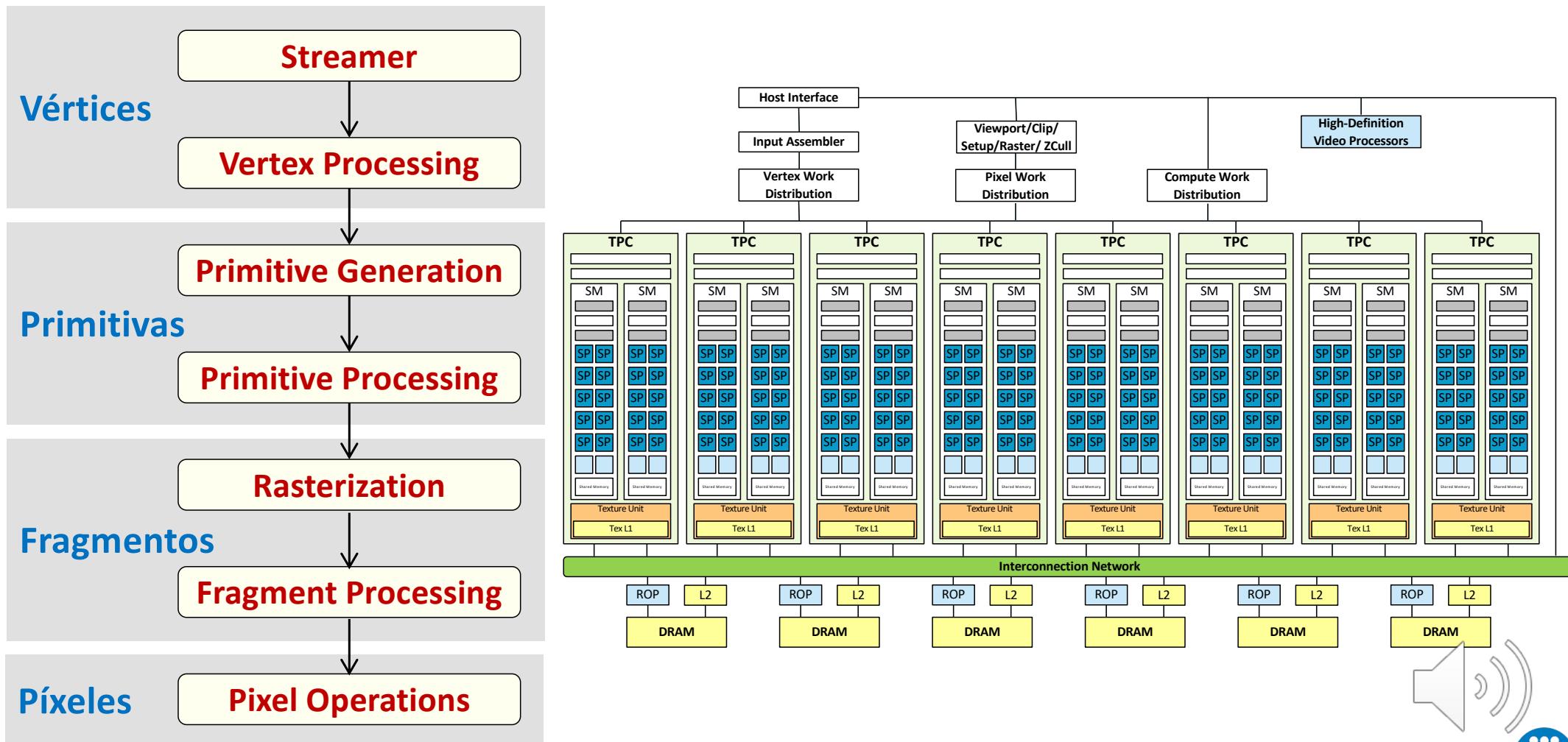
- **Tecnología de integración:** 65nm
- **Número de transistores:** 690 millones
- **Frecuencia GPU:** 650 MHz
- **Frecuencia shader:** 1,625 GHz
- **Stream Processors:** 128
- **Potencia de cálculo:** 624 GFLOPS
- **Texture Units:** 64
- **Raster Operators:** 16
- **Memoria:** 512MB GDDR3, en 6 canales
- **Frecuencia Memoria:** 970 MHz
- **Ancho de banda:** 62,1 GB/s
- **Consumo:** 135 W



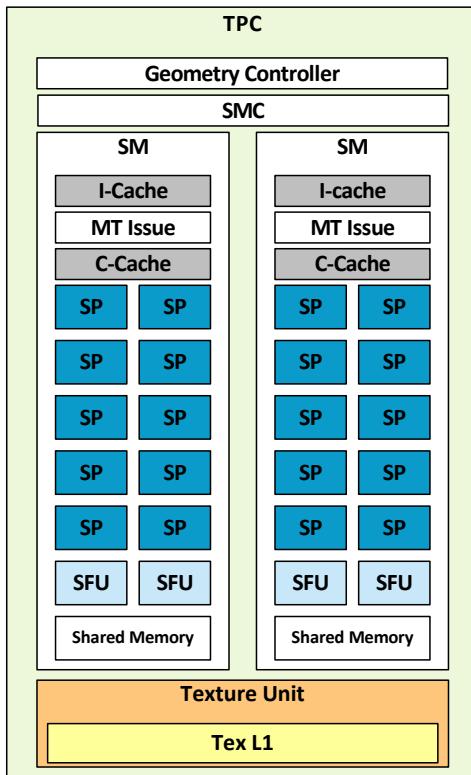
# NVIDIA 8800



# NVIDIA 8800



# NVIDIA 8800



- SP y SFUs funcionan a 1.625 GHz, dando 39 GFLOPS por SM
- Las SFUs realizan funciones especiales (sin, cos, ...)
- Cada SM puede soportar 768 threads a coste zero.
- Instrucciones de acceso a memoria load / store como en cualquier CPU.  
Pero acceden explícitamente a memoria local, compartida y global.
- Unidades en coma flotante de 32 bits.
- Se aparta del concepto clásico de GPU (memoria local, memoria compartida). **Pensadas para GPGPU con CUDA.**
- Las siguientes GPUs de NVIDIA siguen diseños similares con pequeñas variantes:
  - ↑ elementos de cálculo
  - ↑ threads
  - Coma flotante de 64 bits



# GeForce GTX 680

- **Año de Lanzamiento:** 2012
- **Conexión:** PCIe ×16 3.0
- **Versión DirectX:** 11
- **Versión OpenGL:** 4.3
- **Versión OpenCL:** 1.2
- **MultiGPU:** SLI



Familia: Kepler Architecture

PassMark - G3D Mark: **5561**



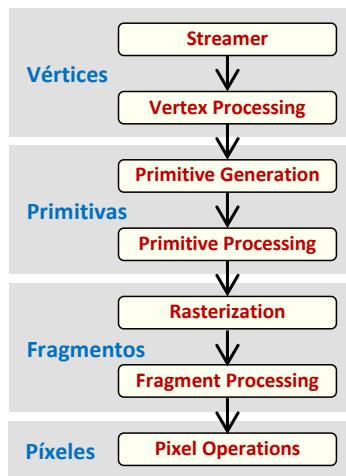
# GeForce GTX 680

## Especificaciones Técnicas

- **Tecnología de integración:** 28nm
- **Número de transistores:** 3540 millones
- **Frecuencia:** 1006 MHz
- **Stream Processors:** 1.536
- **Potencia de cálculo:** 3.090 GFLOPS
- **Texture Units:** 128
- **ROPs:** 32
- **Memoria:** 2-4GB DDR5, en 8 canales, 256 bits
- **Frecuencia Memoria:** 6000 MHz
- **Ancho de banda:** 192,3 GB/s
- **Consumo:** 195 W



# GeForce GTX 680

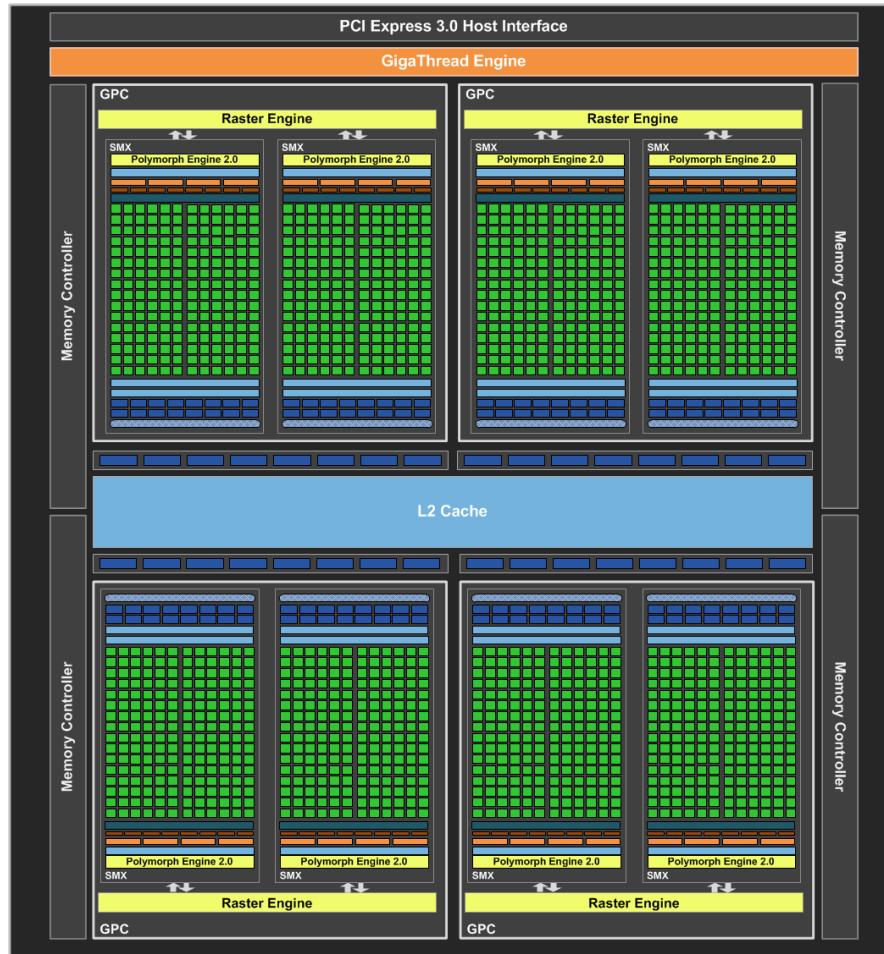


La estructura es "similar" a una GeForce 8800.

192 CUDA Cores



# ¿Cuál es la tarjeta gráfica más potente HOY?



GeForce GTX 680

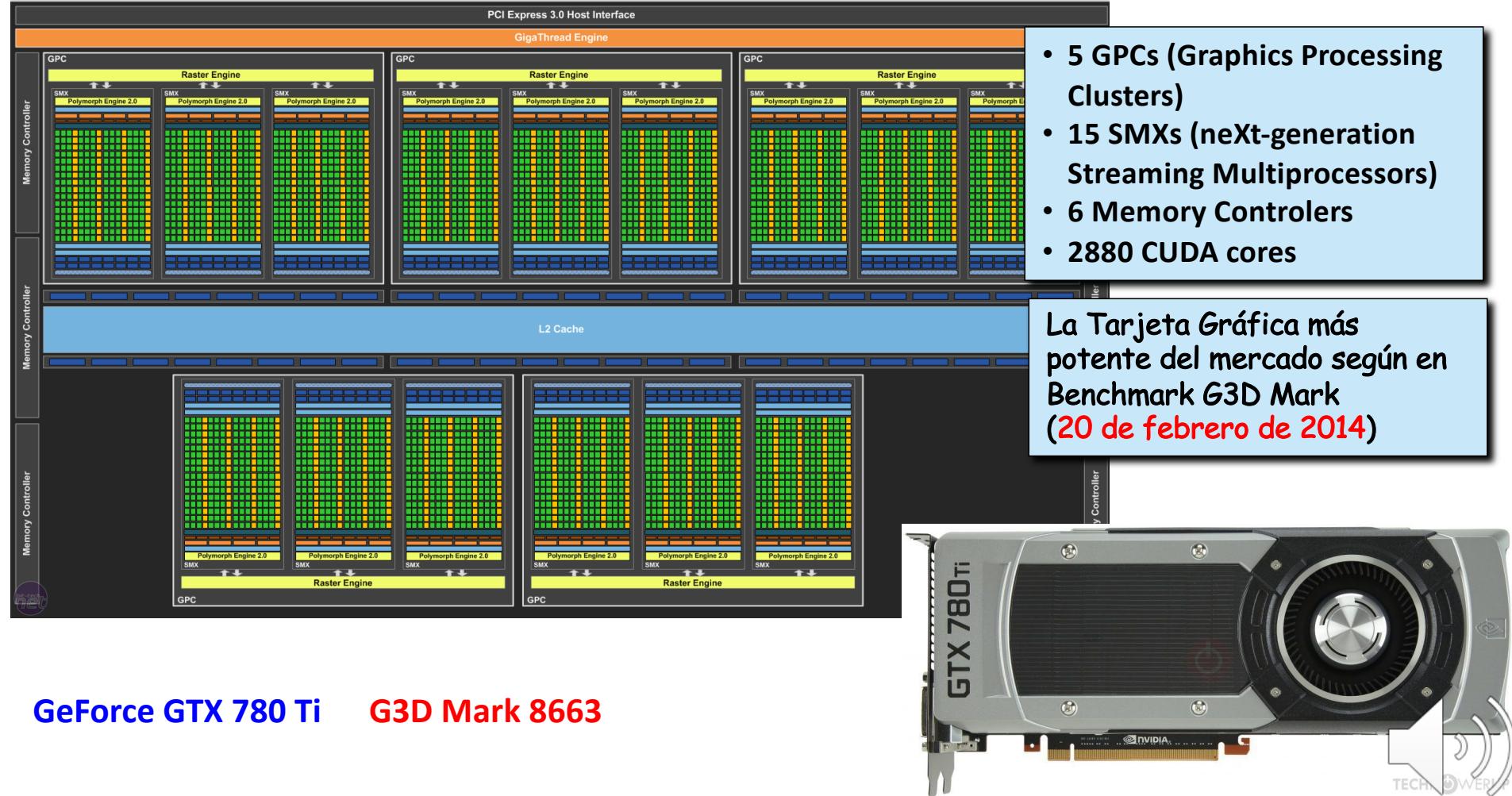
G3D Mark 5708

La Tarjeta Gráfica más potente del mercado según en Benchmark G3D Mark  
(7 de febrero de 2013)

- 4 GPCs (Graphics Processing Clusters)
- 8 SMXs (neXt-generation Streaming Multiprocessors)
- 4 Memory Controllers
- 1536 CUDA cores



# ¿Cuál es la tarjeta gráfica más potente HOY?



# ¿Cuál es la tarjeta gráfica más potente HOY?



¿Dónde está el pipeline gráfico?

GeForce GTX 980

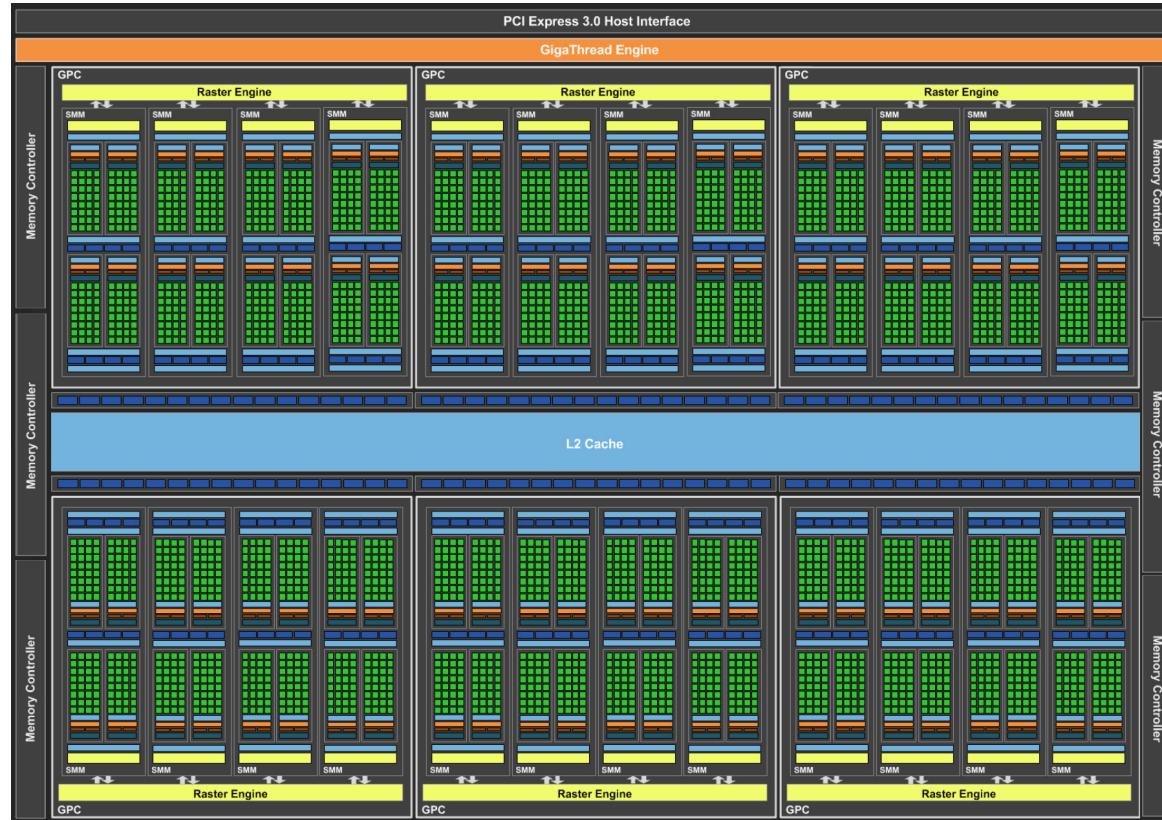
G3D Mark 9674

- 4 GPCs (Graphics Processing Clusters)
- 16 Maxwell SMs (SMM)
- 4 Memory Controllers
- 2048 CUDA cores

La Tarjeta Gráfica más potente del mercado  
según en Benchmark G3D Mark  
(25 de febrero de 2015)



# ¿Cuál es la tarjeta gráfica más potente HOY?



GeForce GTX 980 Ti    G3D Mark 11605

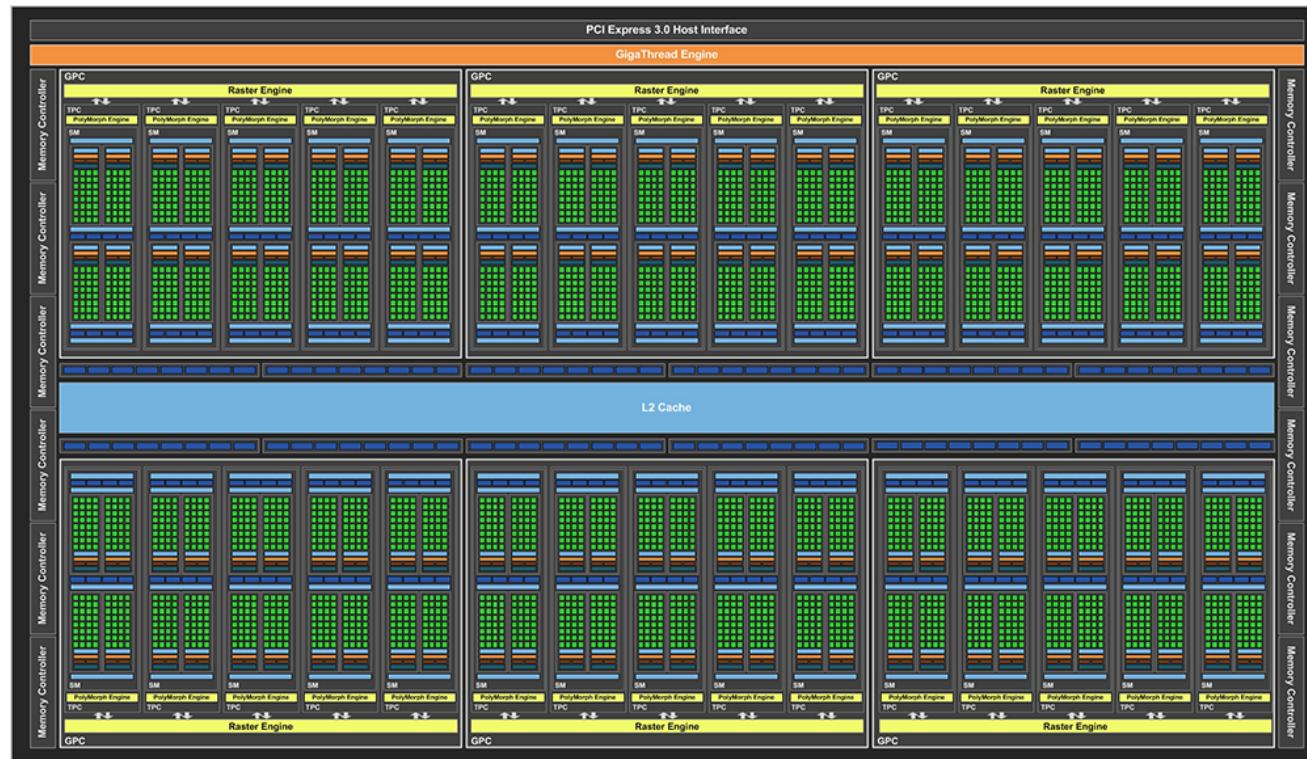
- 6 GPCs (Graphics Processing Clusters)
- 24 Maxwell SMs (SMM)
- 6 Memory Controllers
- 3072 CUDA cores

La Tarjeta Gráfica más potente del mercado según en Benchmark G3D Mark  
(24 de febrero de 2016)



¿Dónde está el pipeline gráfico?

# ¿Cuál es la tarjeta gráfica más potente HOY?



NVIDIA Titan X Pascal **G3D Mark 13060**

- 6 GPCs (Graphics Processing Clusters)
- 28 Pascal SMs
- 6 Memory Controlers
- 3584 CUDA cores
- 1.200 \$

La Tarjeta Gráfica más potente del mercado según en Benchmark G3D Mark (14 de febrero de 2017)



¿Dónde está el pipeline gráfico?

# ¿Cuál es la tarjeta gráfica más potente HOY?



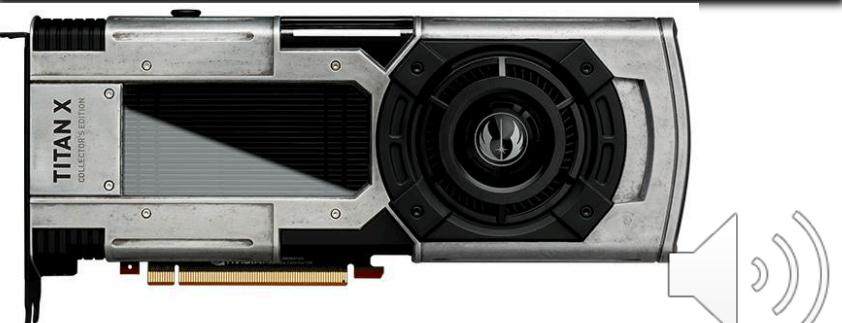
Pascal GP100 Streaming Multiprocessor

¿Dónde está el pipeline gráfico?

NVIDIA Titan Xp COLLECTORS EDITION  
G3D Mark 14.836

- 3.840 CUDA cores (PASCAL)
- 1582 MHz
- 12 GB (GDDR5X)
- 547,7 GB/s
- Resolución 7680x4320 60Hz
- 1.699 \$ (amazon.com)

La Tarjeta Gráfica más potente del mercado  
según en Benchmark G3D Mark  
(26 de febrero de 2018)



# ¿Cuál es la tarjeta gráfica más potente HOY?



Turing SM

NVIDIA Titan RTX  
G3D Mark 16.356

- 4.608 CUDA cores (TURING)
- 1.350 MHz
- 24 GB (GDDR6)
- 672 GB/s
- Resolución 7680x4320 60Hz
- 2.499 \$ (amazon.com)

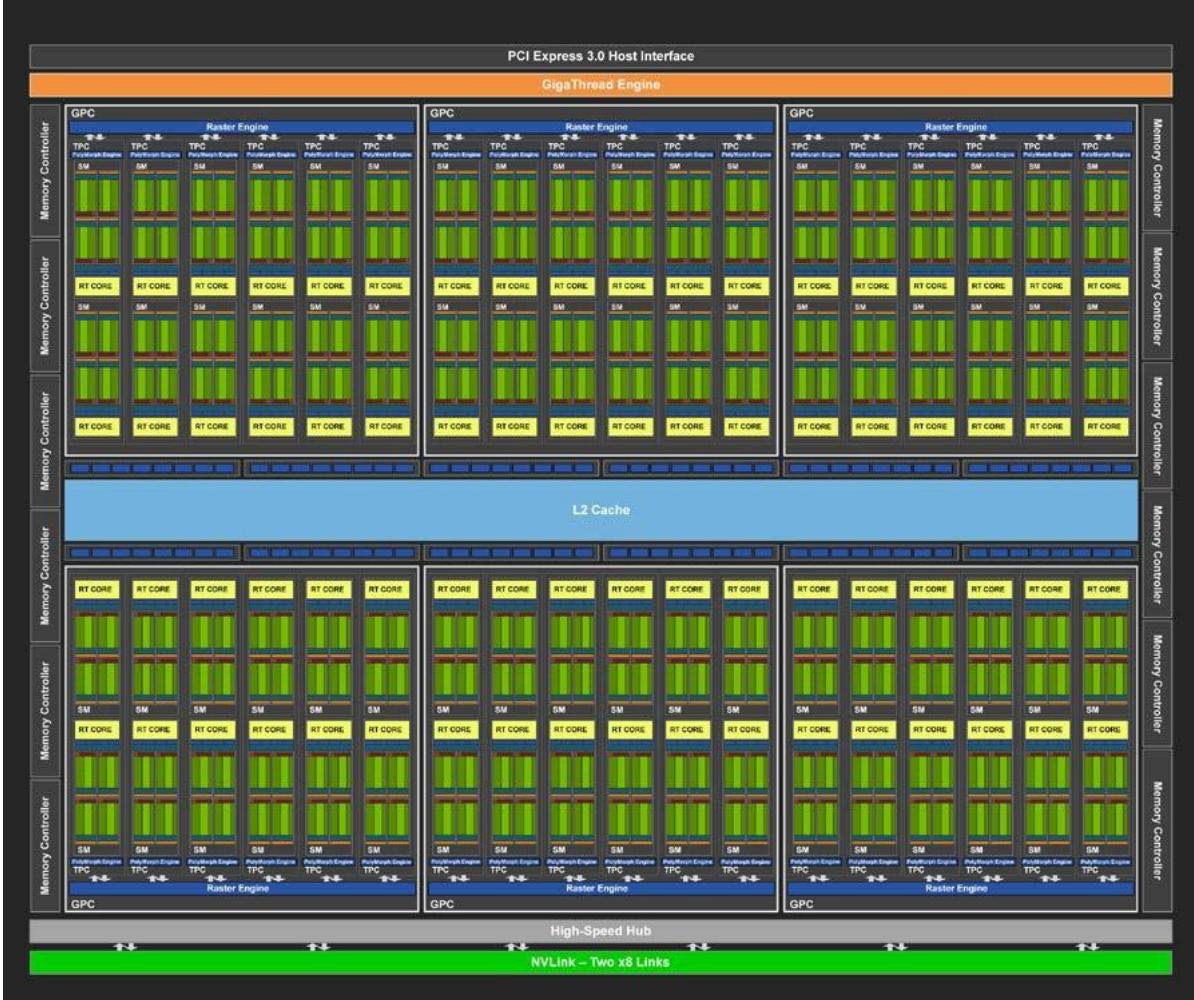
La Tarjeta Gráfica más potente del mercado  
según en Benchmark G3D Mark  
(18 de febrero de 2019)

¿Dónde está el pipeline gráfico?

x72



# ¿Cuál es la tarjeta gráfica más potente HOY?



NVIDIA Titan RTX    G3D Mark 16.356

- 4.608 CUDA cores (TURING)
- 1.350 MHz
- 24 GB (GDDR6)
- 672 GB/s
- Resolución 7680x4320 60Hz
- 2.499 \$ (amazon.com)

La Tarjeta Gráfica más potente del mercado según en Benchmark G3D Mark (18 de febrero de 2019)



# ¿Cuál es la tarjeta gráfica más potente HOY?



¿Dónde está el pipeline gráfico?

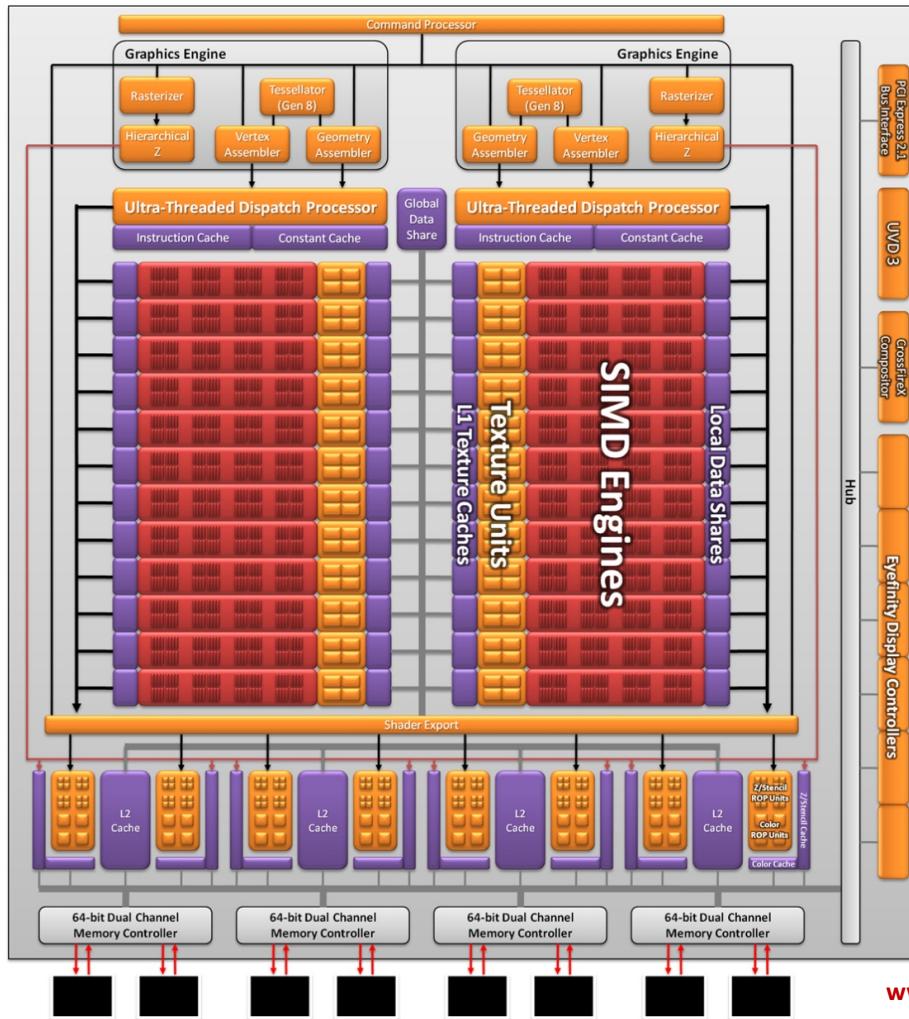
TITAN V CEO Edition G3D Mark 16.998

- 5.120 CUDA cores (VOLTA)
- 640 Tensor cores
- 1.200 MHz
- 32 GB (GDDR6)
- 868 GB/s
- ≈10.000 \$ (no está a la venta)

La Tarjeta Gráfica más potente del mercado según en Benchmark G3D Mark  
(29 de enero de 2020)



# AMD Radeon HD 6970



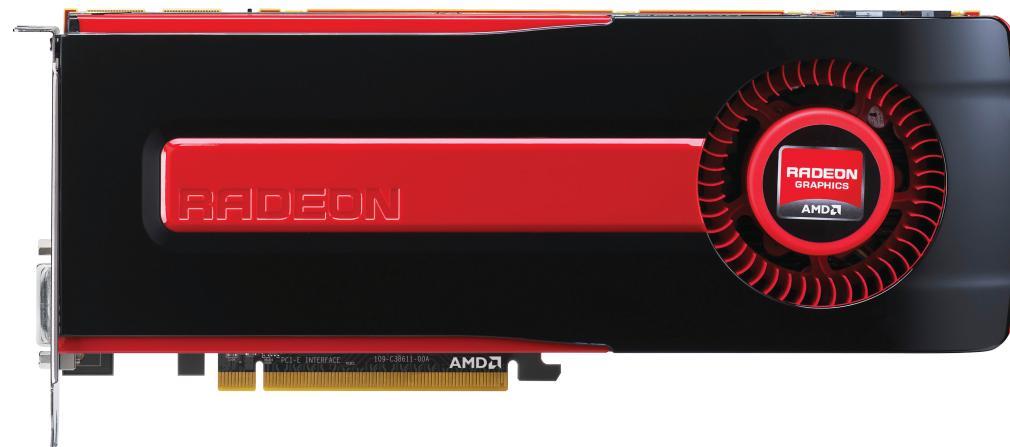
[www.amd.com](http://www.amd.com)

- Lanzamiento: Diciembre 2010
- Multi-core (24 cores)
- Multi-thread
- VLIW
- SIMD
- Frecuencia de la GPU: 880 MHz
- 2 GB de memoria GDDR5
- Frecuencia de memoria 1375 MHz
- 8 canales con memoria
- 176 GB/s de acho de banda de memoria
- 2,7 TFLOPs en simple precisión
- 683 GFLOPs en doble precisión
- 1536 elementos de cálculo
- G3dmark: 3487



# AMD Radeon HD 7970

- **Año de Lanzamiento:** 2012
- **Conexión:** PCIe ×16 3.0
- **Versión DirectX:** 11
- **Versión OpenGL:** 4.2
- **Versión OpenCL:** 1.2
- **MultiGPU:** AMD Crossfire
- Filtrado Anisotrópico ( $\times 16$ )
- Antialiasing ( $\times 24$ )
- Tecnología AMD Eyefinity (máx. 6 pantallas)
- Controlador HD audio 7:1
- Soporte 3D estereoscópico
- **IEEE 754, simple y doble precisión**
- **Soporte interrupciones: debugging**



**Familia: Southern Islands Architecture**

**PassMark - G3D Mark: 4946**



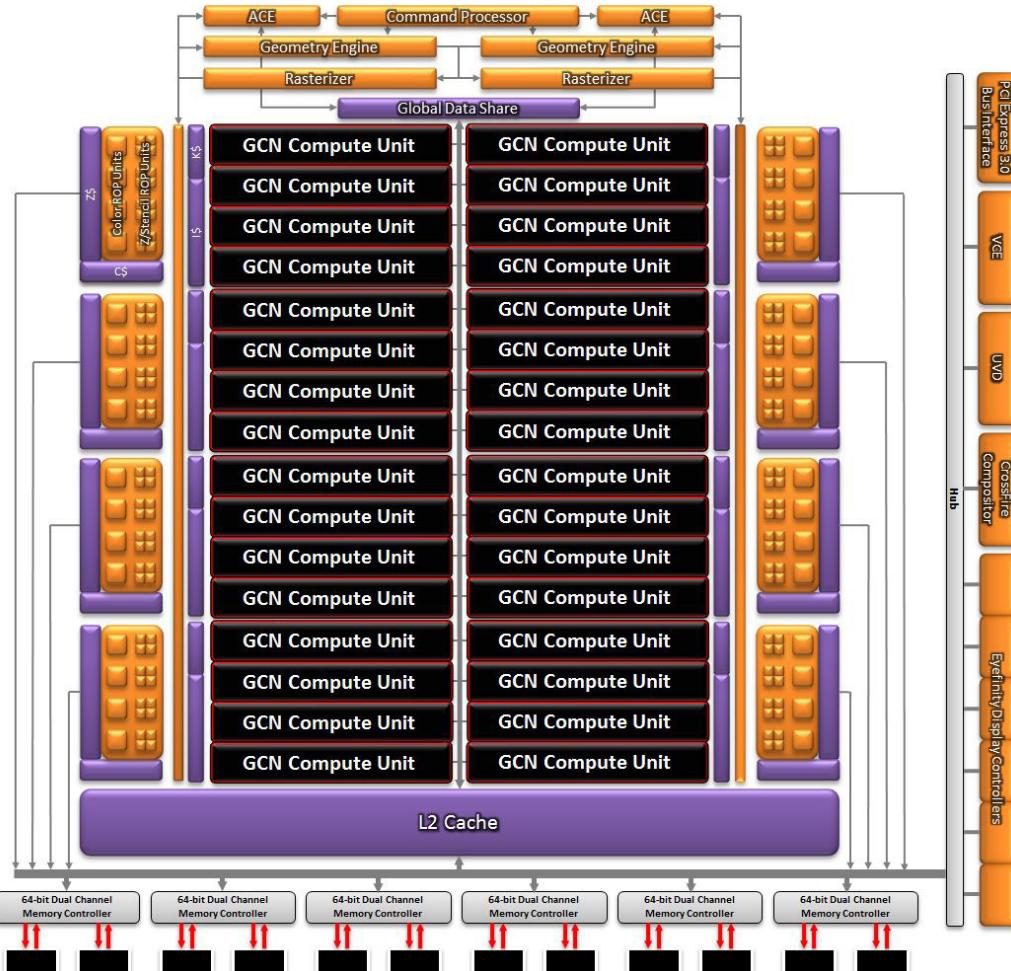
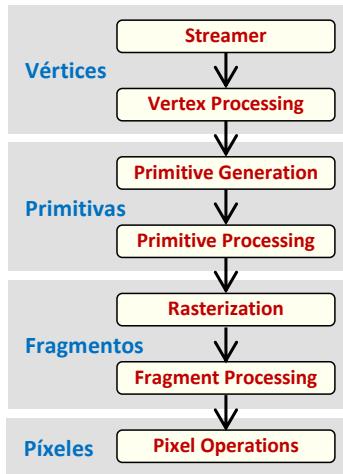
# AMD Radeon HD 7970

## Especificaciones Técnicas

- **Tecnología de integración:** 28nm
- **Número de transistores:** 4310 millones
- **Frecuencia:** 925 MHz
- **Stream Processors:** 2.048
- **Potencia de cálculo:** 3.790 GFLOPS (**3.790.000.000.000 ops en FP por segundo**)
- **Texture Units:** 128
- **ROPs:** 32
- **Z/Stencil:** 128
- **Memoria:** 3GB GDDR5, en 6 canales
- **Frecuencia Memoria:** 1375 MHz
- **Ancho de banda:** 264 GB/s
- **Consumo:** 250 W



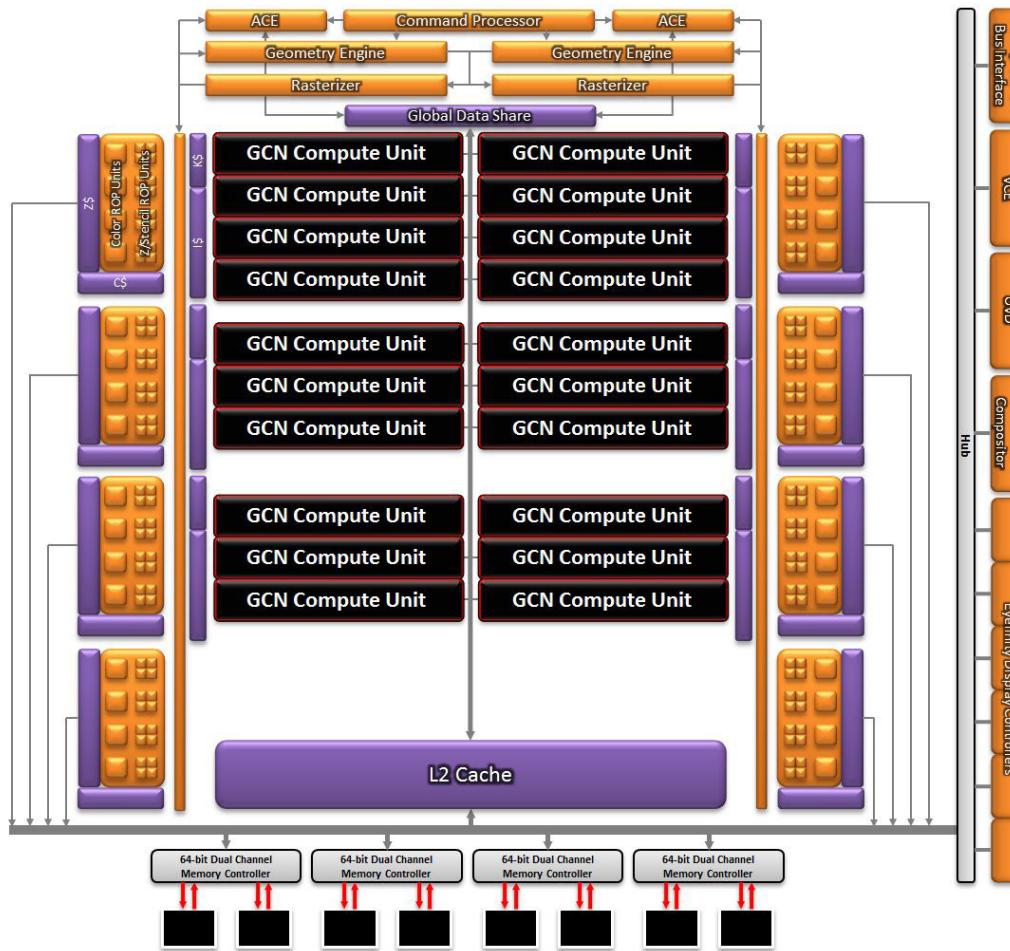
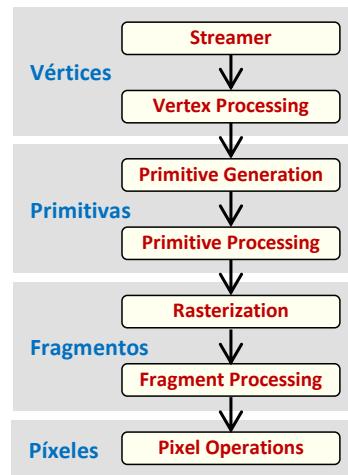
# AMD Radeon HD 7970



**"Tahiti"**  
32 GCN CUs  
81.920 items en vuelo  
6 canales memoria  
384 bits



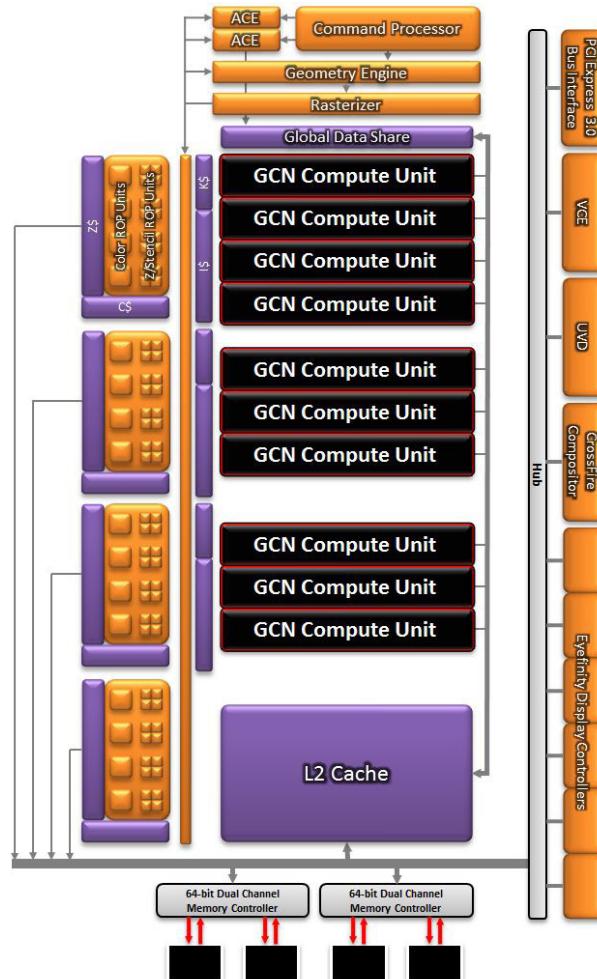
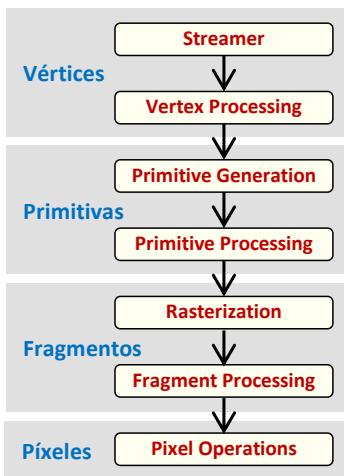
# AMD Radeon HD 7870



**"Pitcairn"**  
20 GCN Cus  
51.200 items en vuelo  
4 canales memoria  
256 bits



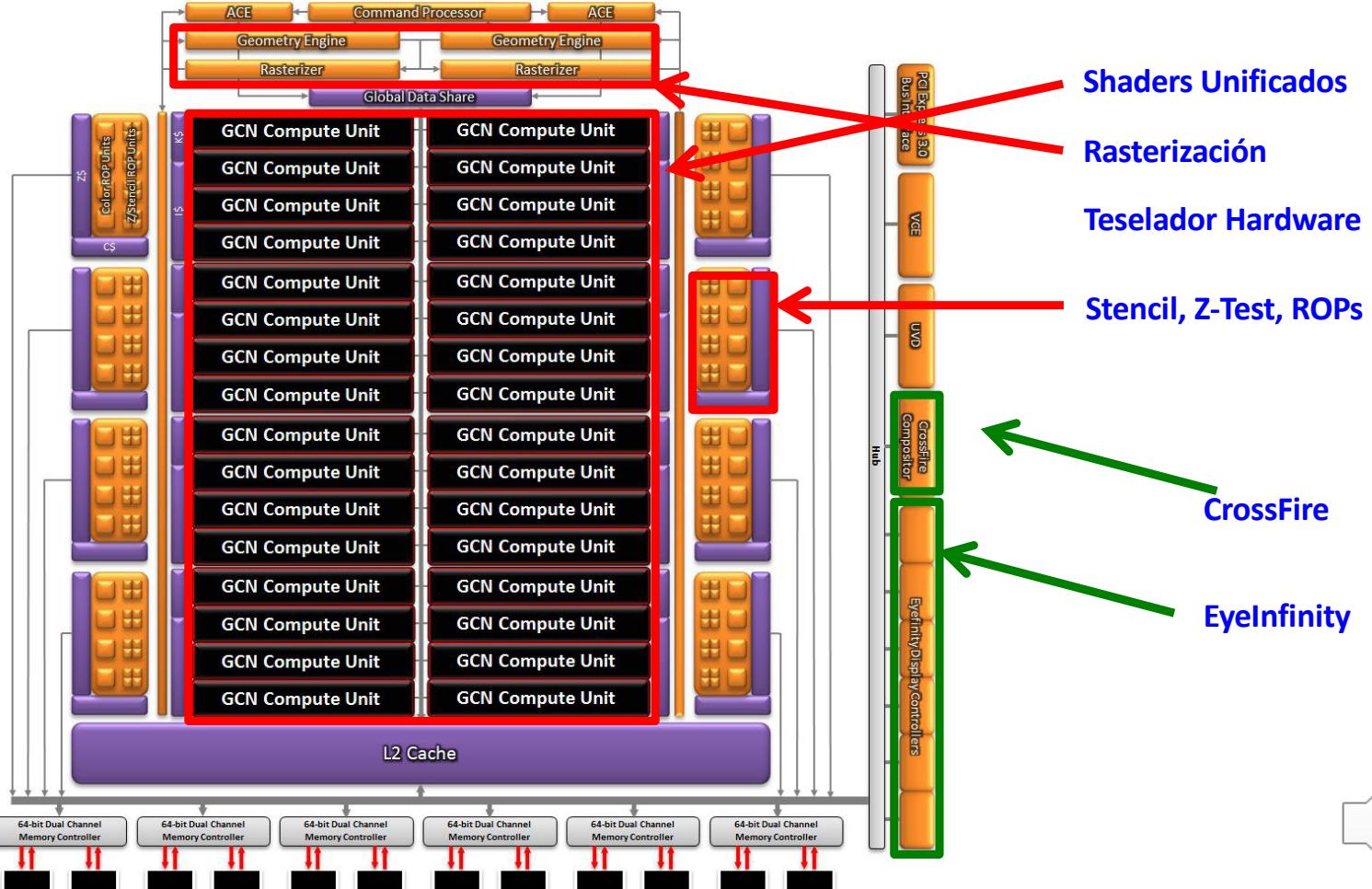
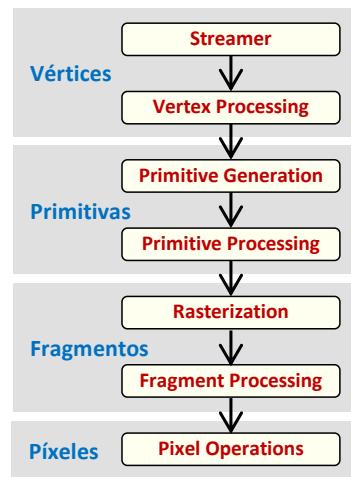
# AMD Radeon HD 7770



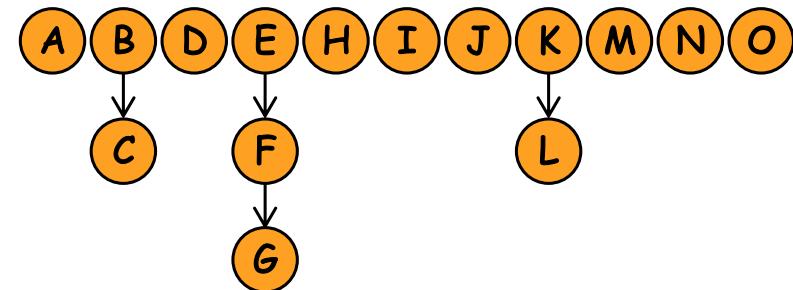
**"Cape Verde"**  
10 GCN Cus  
25.600 items en vuelo  
2 canales memoria  
128 bits



# AMD Radeon HD 7970: Arquitectura Gráfica



# Graphics Core Next: The Southern Islands Architecture



## Dependencias entre wavefronts.

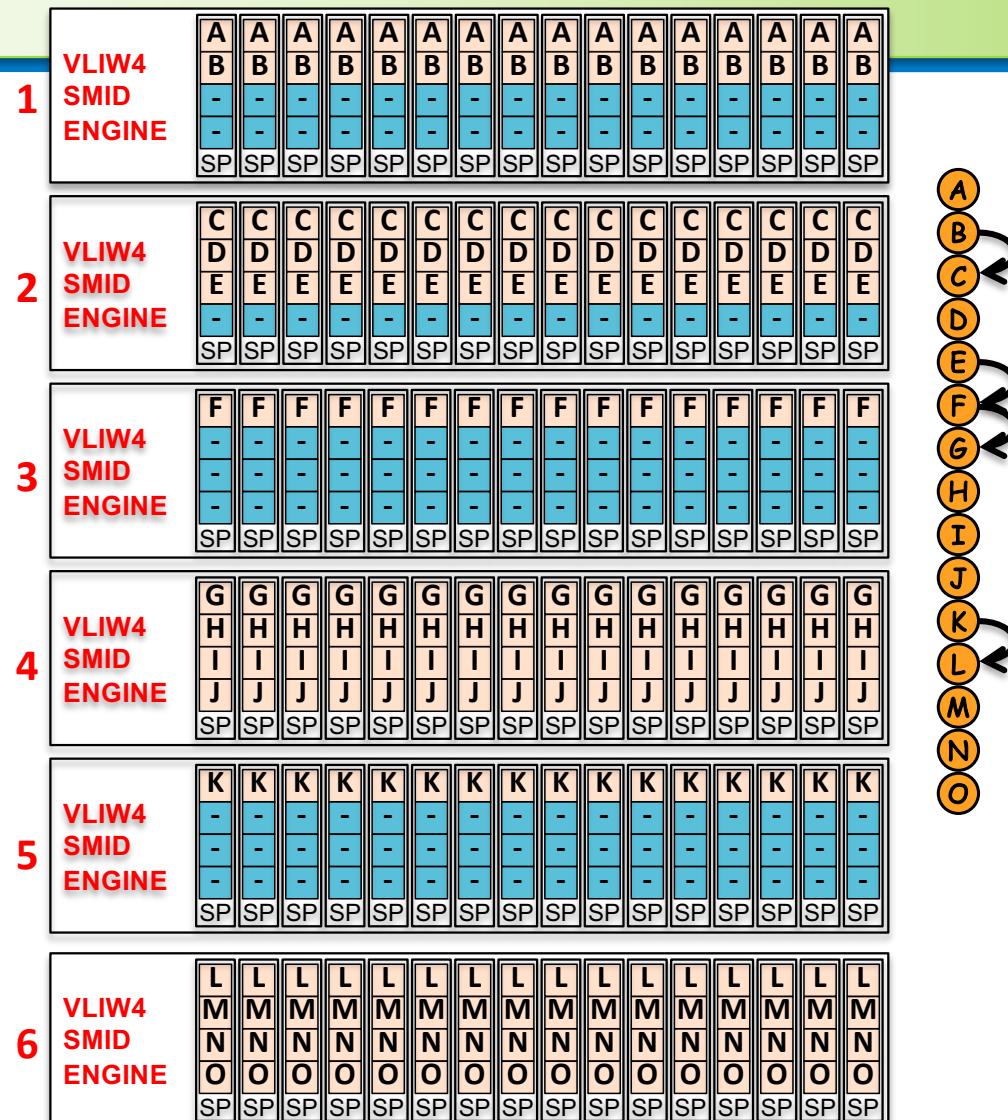
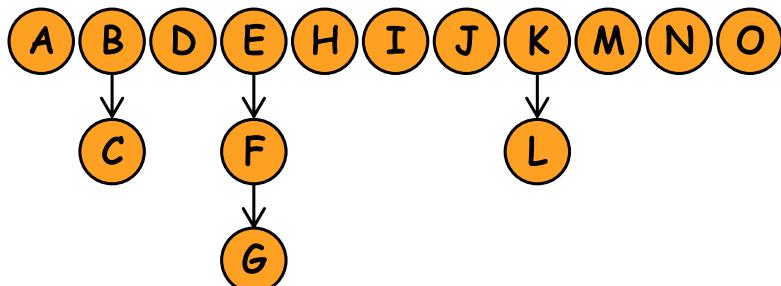
Planificación de wavefronts (instrucciones) realizada por el compilador.



# Ejecución VLIW4

## Ejecución INEFICIENTE

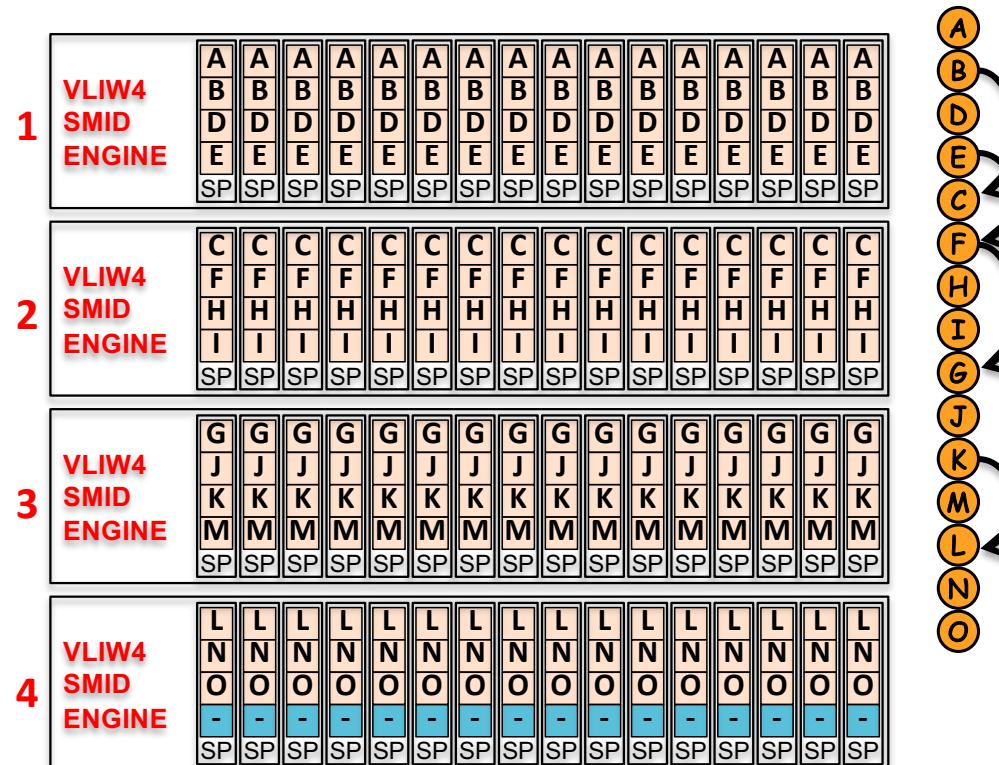
- El compilador es el encargado de definir la planificación de instrucciones.
- El análisis de dependencias no siempre es efectivo en tiempo de compilación.



# Ejecución VLIW4

## Ejecución EFICIENTE

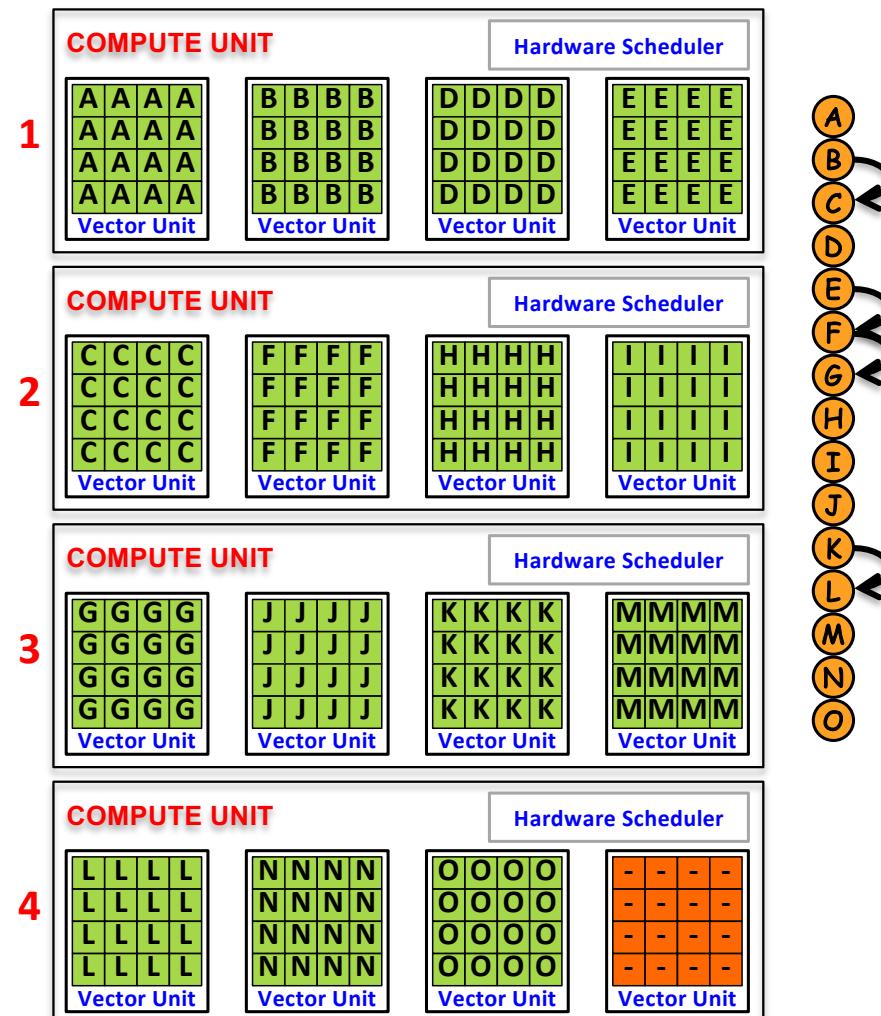
- El compilador es el encargado de definir la planificación de instrucciones.
- El análisis de dependencias no siempre es efectivo en tiempo de compilación.
- Las diferencias de rendimiento pueden ser muy importantes.
- Los conflictos con los recursos (memoria, registros) son difíciles de tratar en tiempo de compilación.  
**Hay que ser conservador.**



# Ejecución SIMD OoO

## Ejecución Out of Order

- El compilador define una planificación de instrucciones inicial.
- El hardware se encarga de ejecutar las instrucciones que estén preparadas para su ejecución.
- Las dependencias entre instrucciones se identifican fácilmente en tiempo de ejecución.
- Los conflictos con los recursos (memoria, registros) son fáciles de identificar.
- Los rendimientos son buenos



# VLIW SIMD vs GCN Quad SIMD

Elementos cálculo anteriores de AMD

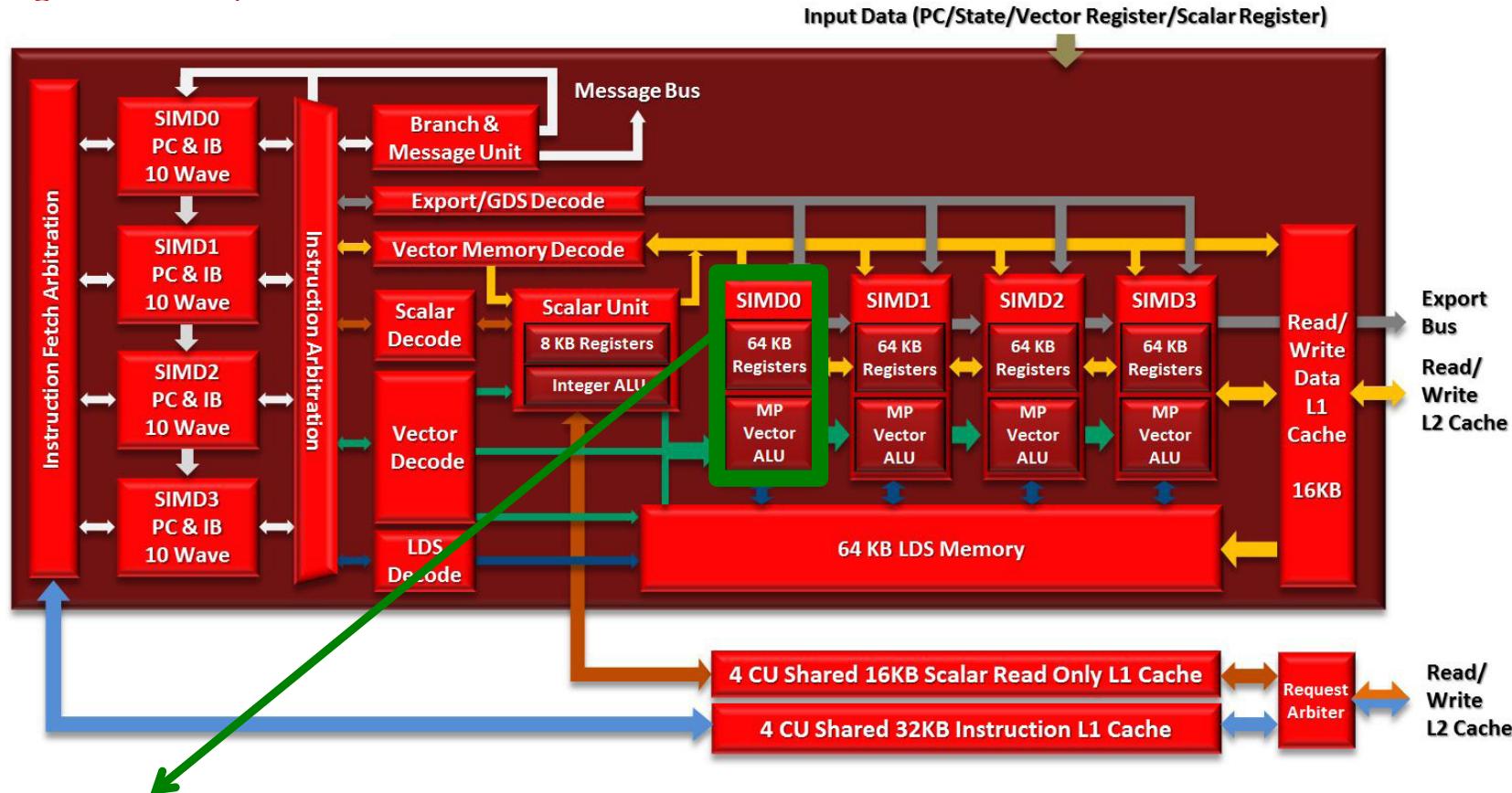
- 1 VLIW  $\times 4$  ALU ops:  
Limitado por las dependencias.
- El compilador gestiona los conflictos en el banco de registros. Necesita conocimiento profundo del hardware.
- Scheduling en manos del compilador: muy especializado y complejo.
- Útil para aplicaciones gráficas, menos flexible para computación de propósito general.
- Optimización cuidadosa para obtener buenos rendimientos.

Nuevos elementos de cálculo: GCN

- 4 SIMD  $\times 1$  ALU ops:  
Limitado por la ocupación de los recursos
- Sin conflicto en el banco de registros
- Compilador estándar. El scheduling se realiza online vía hardware
- Desarrollo y soporte software normal.
- Rendimiento estable y predecible.



# Graphics Core Next Architecture (1st gen)

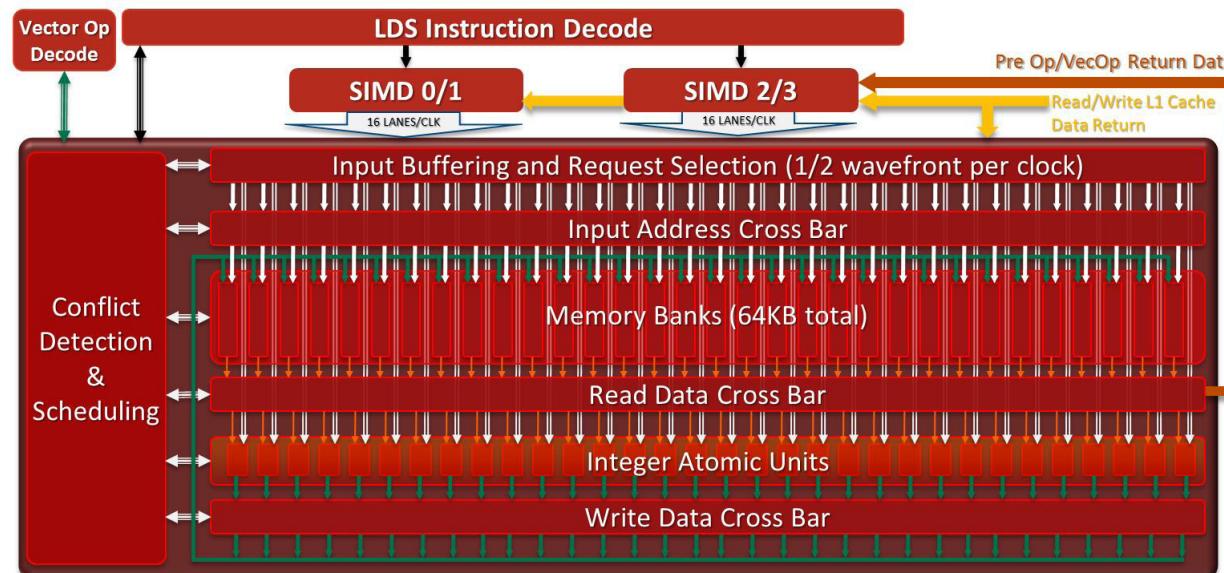


1 operación sobre 16 items por ciclo, 64 shaders en total por GCN  
10 wavefronts simultáneos, 2560 items en vuelo por GCN (quads)



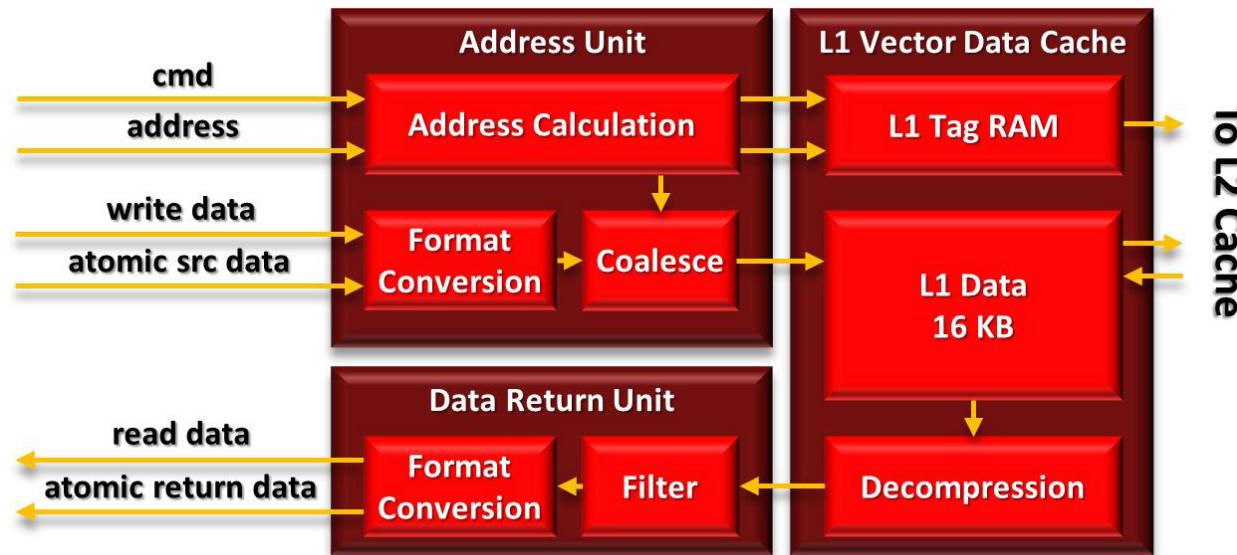
# Graphics Core Next Architecture (1st gen)

- Un GCN puede tener 2560 items en vuelo.
- Alimentar los SIMD para funcionar a esa velocidad no es una tarea trivial.
- Cada GCN Dispone de una zona de memoria local de 64KB, gestionada de forma explícita por el programador. Junto con la vector cache unit consiguen el Ancho de Banda necesario. [Este elemento dificulta la programación de propósito general]



# Graphics Core Next Architecture (1st gen)

## L1 Vector Data Cache

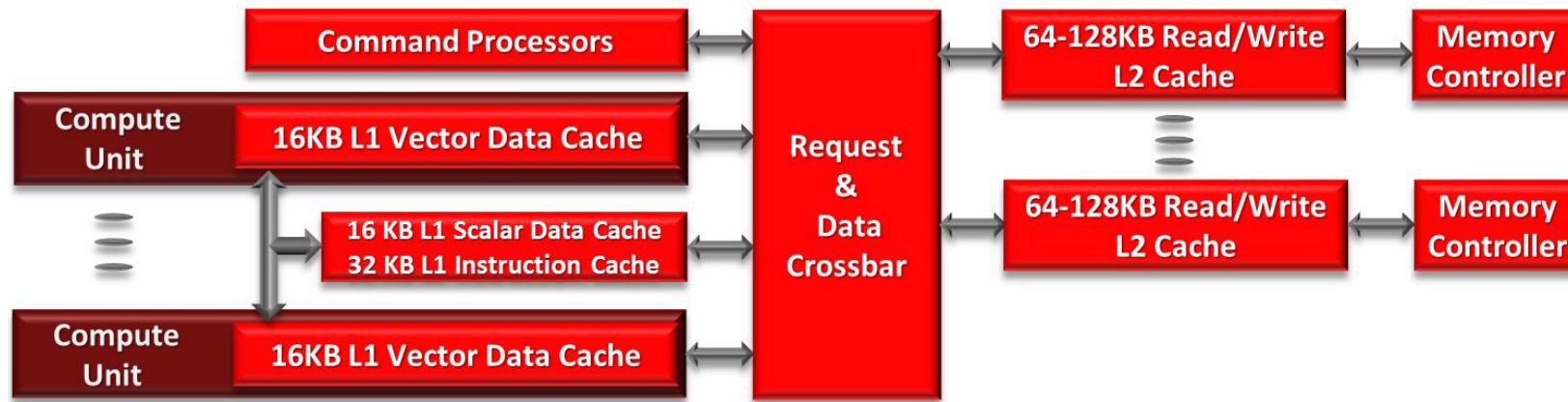


- Elemento muy innovador. No necesario en el entorno de gráficos. Pero muy interesante para las aplicaciones de propósito general.
- 16KB, 4-way, 64B por línea, write through (dirty bit mask) + write allocate (protocolo coherencia)
- En “modo gráfico” la jerarquía se altera para optimizar el acceso a texturas.



# Graphics Core Next Architecture (1st gen)

## Jerarquía de Memoria muy avanzada



- **Cache Coherente**, L1 coherente a nivel de cada grupo de 4 CU, L2 mantiene la coherencia global
- Cache instrucciones y datos compartida cada 4 CUs.
- La comunicación entre CU se realiza con instrucciones explícitas de sincronización.
- L2: 64-128KB (16-way 64B por línea) por “slice” de L2, asociada a 1 canal de MP. **Escalable**.
- Soporta **Memoria Virtual**. Especialmente diseñado para ser compatible con la Memoria Virtual de los procesadores x86. [GPU integrada chips AMD].
- **MULTITASKING**



# AMD Radeon R9 Fury X

- **Año de Lanzamiento:** 2015
- **Conexión:** PCIe ×16 3.0
- **Versión DirectX:** 12
- **Versión OpenGL:** 4.5
- **Versión OpenCL:** 2.2
- Vulkan & Mantle
- **MultiGPU:** AMD Crossfire
- Tecnología AMD Eyefinity (máx. 6 pantallas)
- Resolución 4K
- **Memoria HBM**

**Familia: Volcanic Islands Architecture**

**PassMark - G3D Mark: 8117**



# AMD Radeon R9 Fury X

## Especificaciones Técnicas

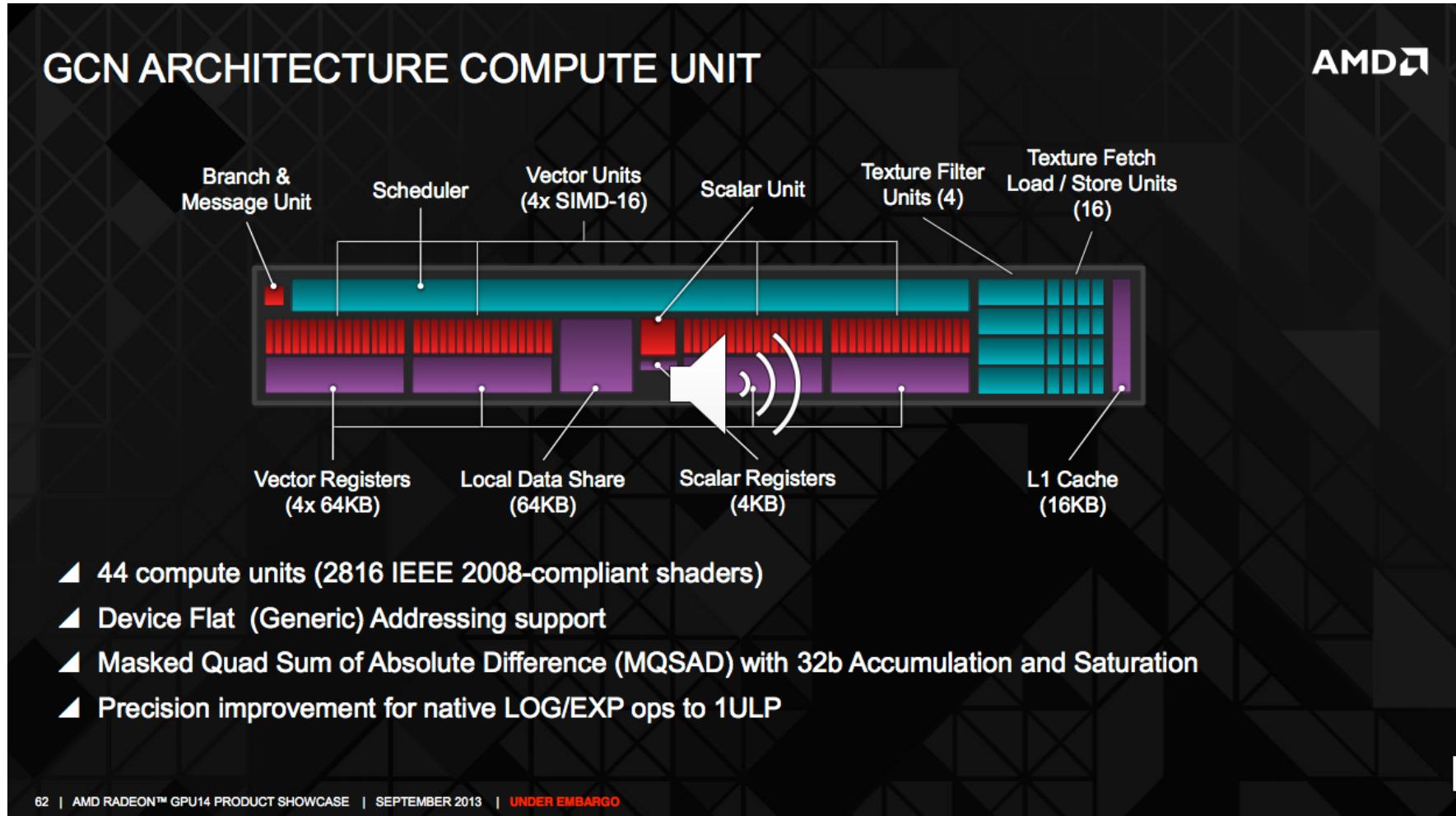
- **Tecnología de integración:** 28nm
- **Número de transistores:** 8900 millones
- **Frecuencia:** 1050 MHz
- **Compute Units:** 64 (GCN 3<sup>a</sup> gen.)
- **Stream Processors:** 4.096
- **Potencia de cálculo:** 8.602 GFLOPS fp32 [538 GFLOPS fp64]
- **Texture Units:** 256
- **ROPs:** 64
- **Z/Stencil:** 256
- **Memoria:** 4GB HBM, bus 4096 bits, 8 canales
- **Frecuencia Memoria:** 500 MHz
- **Ancho de banda:** 512 GB/s
- **Consumo:** 275 W



# AMD Radeon R9 Fury X



# AMD Radeon R9 Fury X



AMD



[www.amd.com](http://www.amd.com)



# Evolución del Pipeline Gráfico: AMD

AMD	Año	VS	FS	TU	ROP	Mpix/s	Mtex/s
Rage 128 Pro	1999	0	2	2	2	250	250
Radeon 7200	2000	0	2	6	2	333	966
Radeon 7500	2001	0	2	6	3	580	1740
Radeon 9000 Pro	2002	1	4	4	4	1.100	1.100
Radeon 9800 XT	2003	4	8	8	8	3.296	3.296
Radeon X850 XT	2004	6	16	16	16	8.320	8.320
Radeon X1800 XT	2005	8	16	16	16	10.000	10.000
Radeon X1950 XT	2006	8	48	16	16	10.000	10.000
Radeon HD 2900 XT	2007	320		16	16	11.900	11.900
Radeon HD 4870	2008	800		40	16	12.000	30.000
Radeon HD 4890	2009	800		40	16	13.600	34.000
Radeon HD 5870	2009	1.600		80	32	27.200	68.000
Radeon HD 6970	2010	1.536		96	32	28.200	84.500
Radeon HD 6930	2011	1.280		80	32	24.000	60.000
Radeon HD 7970	2012	2.048		128	32	29.600	118.400
Radeon HD 8970	2013	2.048		128	32	33.600	134.400
Radeon R9 290X	2013	2.816		176	64	64.000	176.000
Radeon R9 280	2014	1.792		112	32	26.500	92.600
Radeon R9 Fury X	2015	4.096		256	64	67.200	268.800
Radeon RX 480	2016	2.304		144	32	35.800	161.300
Radeon RX Vega 64	2017	4.096		256	64	98.900	395.800
Radeon RX 590	2018	2.304		144	32	47.000	211.500
Radeon VII	2018	3.840		240	64	115.000	432.000
Radeon RX 5700 XT	2019	2.560		160	64	102.700	256.800
Radeon RX 6900 XT	2020	5.120		320	128	288.000	720.000

Geometría en la CPU o función fija

Primeras GPUs programables

Shaders diferenciados

Shaders unificados

VS: Vertex Shaders

FS: Fragment Shaders

TU: Texture Units

ROP: Render Output Units  
(Pixel Operations)



# Evolución del Pipeline Gráfico: NVIDIA

NVIDIA	Año	VS	FS	TU	ROP	Mpix/s	Mtex/s
Riva TNT2	1999	0	2	2	2	250	250
GeForce 256 DDR	2000	0	4	4	4	480	480
Geforce2 Pro	2000	0	4	8	4	800	1.600
GeForce3	2001	1	4	8	4	800	1.600
GeForce4 Ti 4600	2002	2	4	8	4	1.200	2.400
GeForce FX 5950 Ultra	2003	3	4	8	4	1.900	3.800
GeForce 6800 GT	2004	6	16	16	16	5.600	5.600
GeForce 6800 Ultra	2005	6	16	16	16	6.400	6.400
GeForce 7900 GTX	2006	8	24	24	16	10.400	15.600
GeForce 8800 Ultra	2007	128		32	24	14.700	39.200
GeForce 9800 GTX	2008	128		64	16	10.800	43.200
GeForce GTS 150	2009	128		64	16	20.736	51.840
GeForce GTX 285	2009	240		80	32	11.808	47.232
GeForce GT 340	2010	96		32	8	4.400	17.600
GeForce GTX 480	2010	480		60	48	33.600	42.000
GeForce GTX 560 Ti	2011	384		60	48	29.280	32.210
GeForce GTX 680	2012	1.536		128	32	32.200	128.800
GeForce GTX Titan	2013	2.688		224	48	40.200	187.500
GeForce GTX Titan Black	2014	2.880		240	48	42.700	213.400
GeForce GTX Titan X	2015	3.072		192	96	96.000	192.000
Nvidia Titan X Pascal	2016	3.584		224	96	136.000	317.400
Nvidia Titan V	2017	5.120		320	96	153.600	384.000
Nvidia Titan RTX	2018	4.608		288	96	129.600	388.800
GeForce RTX 2080 Super	2019	3.072		192	64	105.600	316.800
GeForce RTX 3090	2020	10.496		328	112	134.400	459.200

Geometría en la CPU o función fija

Primeras GPUs programables

Shaders diferenciados

Shaders unificados

**VS:** Vertex Shaders

**FS:** Fragment Shaders

**TU:** Texture Units

**ROP:** Render Output Units  
(Pixel Operations)



# Los nuevos tiempos: NVIDIA Titan RTX

GPU
Familia: Turing
Nombre: TU102
Tecnología: 12 nm
Transistores: $18\text{-}600\cdot10^6$
Die Size: 754 mm <sup>2</sup>
GPU Clock: 1350 MHz
Boost Clock: 1770 MHz

Tarjeta
Lanzamiento: Dec 2018
Bus: PCIe x16 3.0
Output: 1 HDMI, 3 DisplayPort, 1 USB-C
Power Input: 2x 8-pin



API
DirectX: 12.0
OpenGL: 4.6
OpenCL: 2.0
Shader Model: 6.3

Rendimiento
Pixel rate: 169.9 Gpix/s
Texture rate: 509,8 Gtex/s
GFLOPs FP32: 16.312
GFLOPs FP64: 509,8
GFLOPs FP16: 32.625
Consumo: 275 W
G3D mark: 20.091

Memoria
Tamaño: 24 GB
Tipo: GDDR6, 1.75GHz
Bandwidth: 672 GB/s
Bus: 384 bits

Configuración
SMM: 72
Shaders: 4608
TMUs: 288
ROPs: 96
Tensor Cores: 576
RT Cores: 72

# GPU Integrada

- Desde finales de los 90 muchos chipset incluían entre sus capacidades un adaptador de video estándar de nivel medio-bajo.

Fabricante	Chipset	CPU soportada	Tipo Adaptador	Soporte	Memoria Vídeo
Intel	865G	Pentium 4, Celeron 4	Intel Extrem Graphics 2	AGP 8x	32MB, 64MB
Intel	915G	Pentium 4, Celeron 4	Intel Extrem Graphics 3	PCIe x16	32MB, 64MB
ATI	IGP 9100	Pentium 4, Celeron 4	ATI RADEON 9000	AGP 8x	Up to 128MB
NVIDIA	NForce2	Athlon, Duron	NVIDIA GeForce MX420	AGP 8x	Up to 64MB
SiS	SiS760	Athlon 64, Opteron	SiS Ultra-AGPII Mirage	AGP 8x	Up to 64MB
VIA	PM880	Pentium 4, Celeron 4	S3 Graphics Unichrome	AGP 8x	Up to 64MB
Intel	865G	Pentium 4, Celeron 4	Intel Extrem Graphics 2	AGP 8x	32MB, 64MB
Intel	915G	Pentium 4, Celeron 4	Intel Extrem Graphics 3	PCIe x16	32MB, 64MB

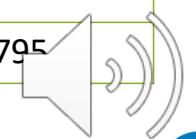
- La tendencia actual es incluir la GPU integrada dentro de la CPU.
- Siempre se puede desactivar e insertar una tarjeta más potente.



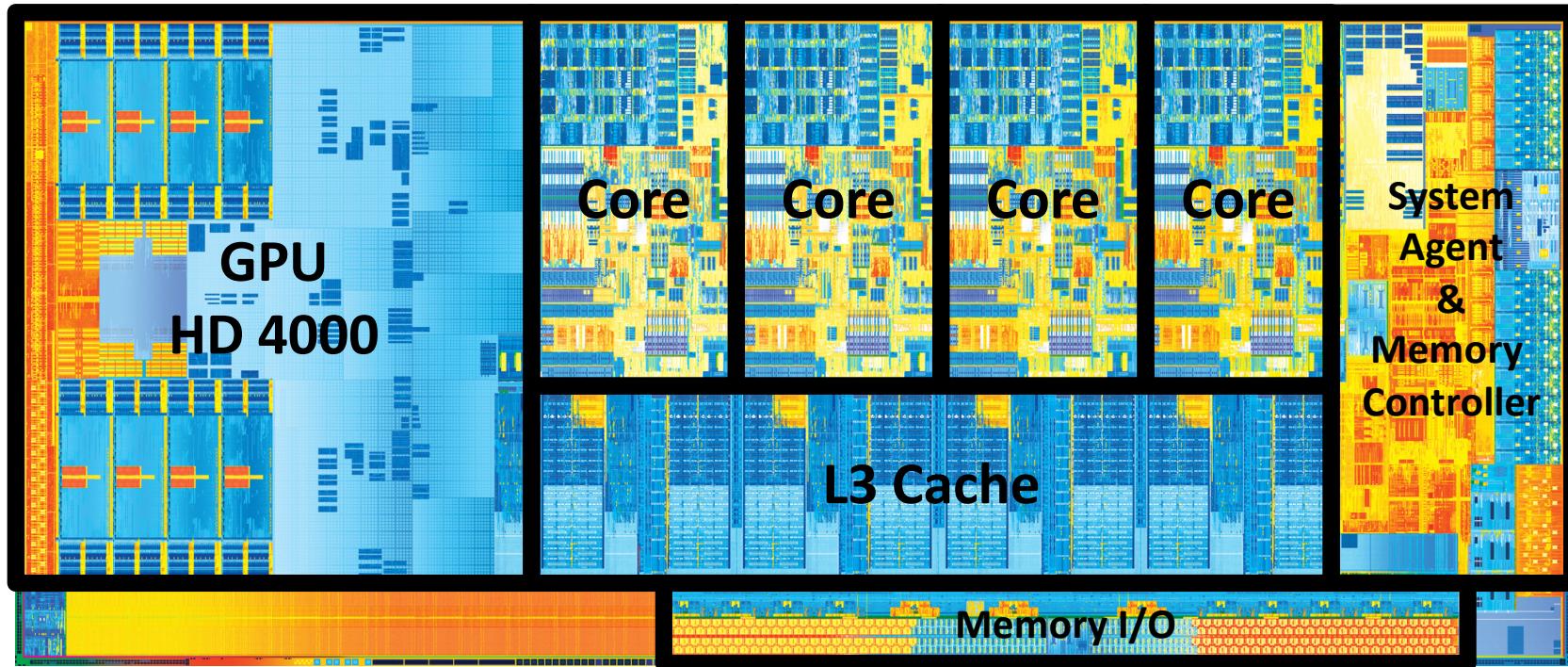
# GPUs Integradas Intel

	HD 2000	HD 3000	HD 2500	HD 4000	HD 4600	HD 5000
Processor Architecture	Sandy Bridge		Ivy Bridge		Haswell	
Technology	32 mm		22 mm		22 mm	
DirectX support	DirectX 10		DirectX 11		DirectX 11	
OpenGL support	OpenGL 3.0		OpenGL 3.1		OpenGL 4.2	
OpenCL support	-		OpenCL 1.1		OpenCL 1.2	
Execution Units	6	12	6	16	20	40
Clock Frequency	650 – 850 GHz	850 GHz	650 GHz	650 GHz	650 GHz	650 GHz
Clock Frequency (Turbo)	1 – 1.1 GHz	1.1 – 1.35 GHz	1,05 – 1,15 GHz	1,15 GHz	Up to 1350 GHz	
Theoretical GFLOPS	31,2 – 52,8	81,6 – 129,6	62,4 – 110,4	166,4 – 294,4	432	704
Memory bus			128 bits			
Memory frequency	1066/1333 MHz			1333/1600 MHz		
Memory type			DDR3			
Memory Bandwidth	17 – 21,3 GB/s		21,3 – 25,6 GB/s		25,6 GB/s	
G3D Mark	Intel HD Family: 299		472	636	795	

Todos estos datos pueden variar en función del procesador en el que esté integrada la GPU.

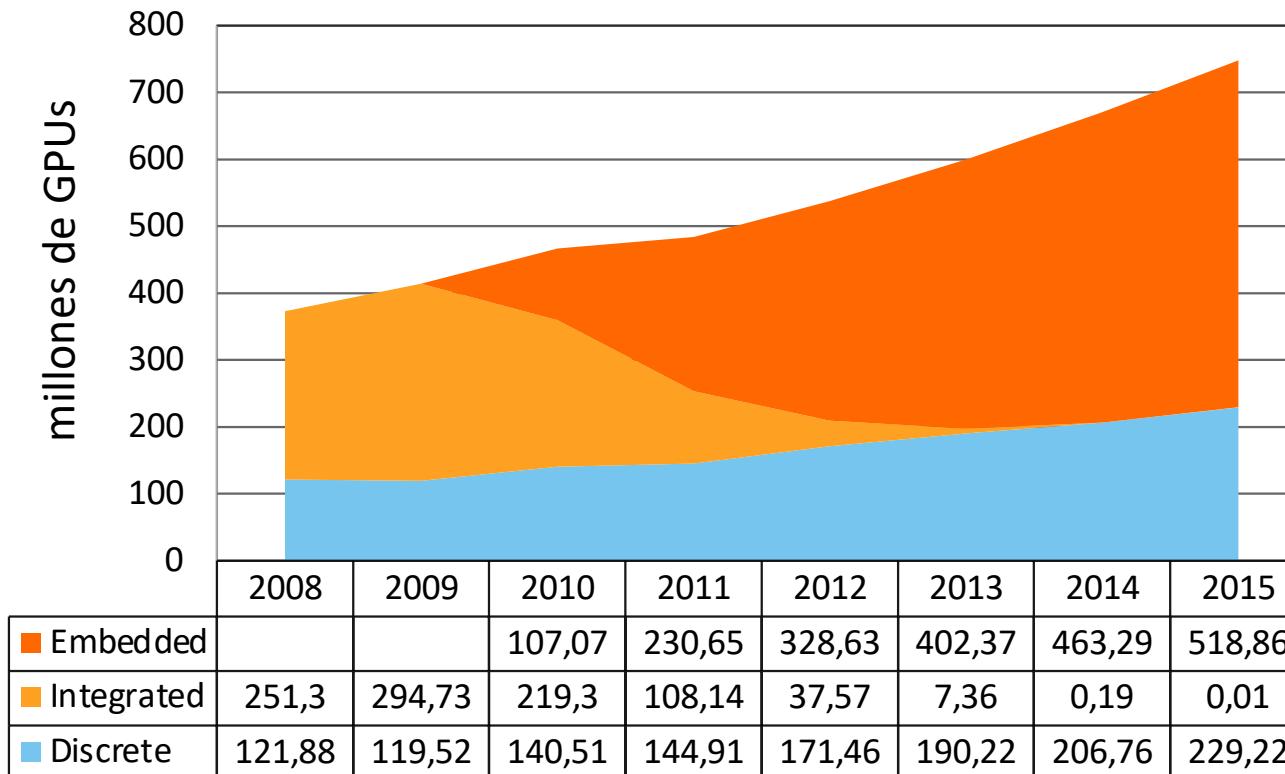


# Ivy Bridge Architecture



# GPUs Integradas

## □ Cuota de mercado entorno PC (previsiones hechas en 2011)



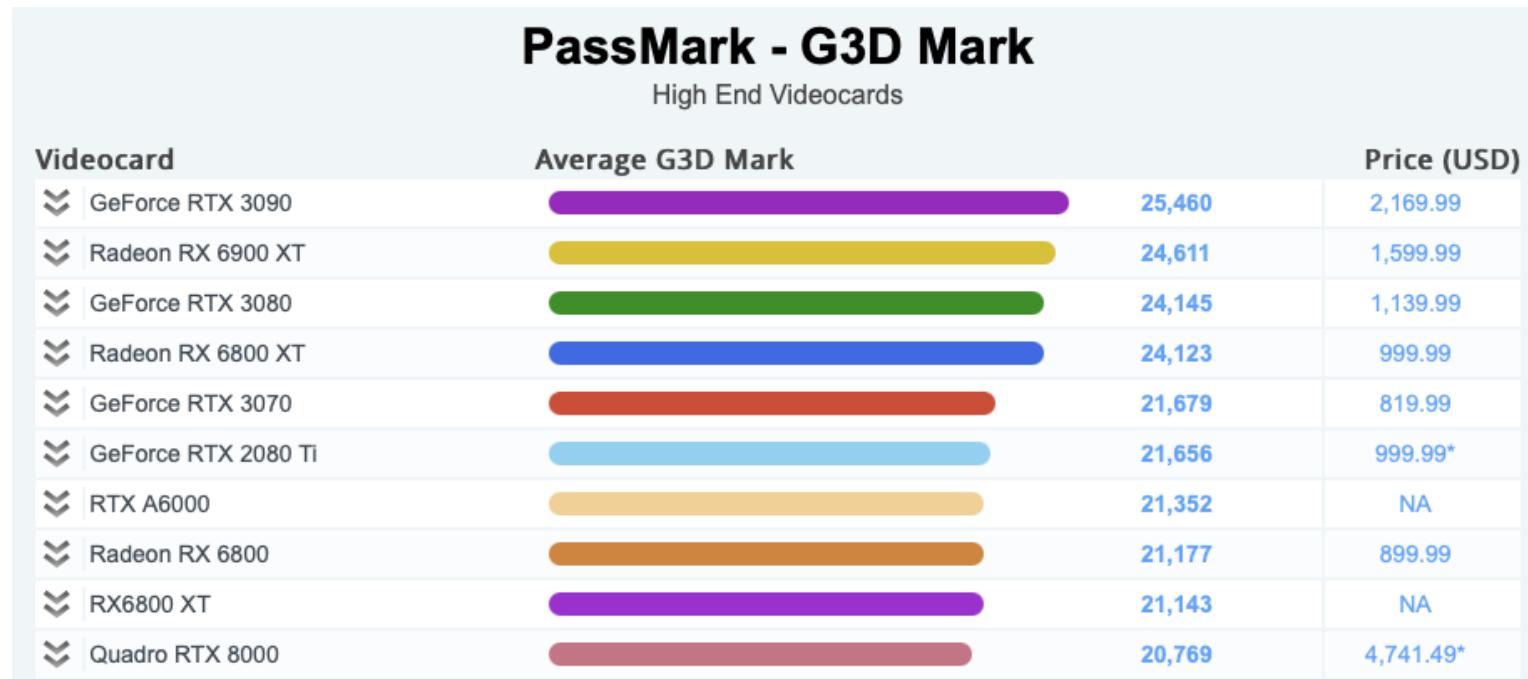
Embedded,  
GPU integradas  
con la CPU

Integrated,  
GPU integradas  
con la placa base



# ¿Es buena mi tarjeta gráfica?

- PassMark – G3D Mark
- [www.videocarbenchmark.net](http://www.videocarbenchmark.net)
- Actualizado diariamente (datos 12/Mar/2020)



# ¿Es buena mi tarjeta gráfica?

- PassMark – G3D Mark
- [www.videocardbenchmark.net](http://www.videocardbenchmark.net)
- Actualizado diariamente (datos 12/Mar/2021)



# Passmark

## Video Card Test

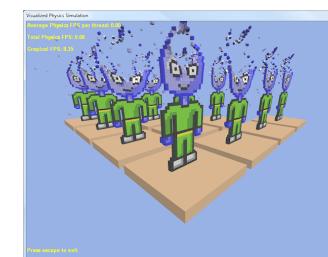
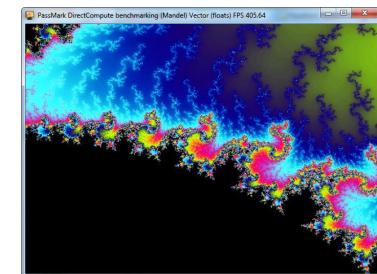
- DirectX 9 o superior
- Múltiples opciones
  - ✓ Efectos shader
  - ✓ Número de objetos
  - ✓ Iluminación
  - ✓ Transparencias
  - ✓ Resolución
  - ✓ Tipos texturas
  - ✓ Duración

## Test de Física

- DirectX 10 o superior
- Provocar una “explosión” (gravedad, resistencia aire, ...)

## Direct Compute Benchmark

- DirectX 11
- Bitonic Sort
- Fluid Test
- Julia Test
- Mandelbrot Test



# La Guerra de las Consolas

**El negocio de las consolas/juegos mueve mucho dinero. Consolas vendidas (enero 2014):**

- PS3:** 82 millones (desde Nov 2006)
- Xbox 360:** 81 millones (desde Nov 2005)
- Wii:** 100 millones (desde Nov 2006)
- PS4:** 4,38 millones (desde Nov 2013)
- Xbox One:** 3,11 millones (desde Nov 2013)

**Está previsto que en 2016 se hayan vendido:**

- PS4:** 38 millones
- Xbox One:** 29 millones

**El negocio también está en los juegos. Videojuegos vendidos (enero 2014):**

- PS3:** 803 millones (Grand Theft Auto V, 16,72 millones)
- Xbox 360:** 872 millones (Kinetic Adventures, 20,89 millones)
- Wii:** 913 millones (Wii Sports, 81,79 millones; Mario Kart Wii, 34,28 millones)



# La Guerra de las Consolas

## PS3

- **CPU:** Cell (1 PPE & 6 SPEs), 3.2 GHz
  - PPE, procesador propósito general, tipo PowerPC, 4,8 GFLOPs rendimiento de pico
  - SPE, procesador vectorial de propósito específico con 256 KB memoria local, 25,6 GFLOPs
  - 153 GFLOPs rendimiento de pico.
- **Memoria:** 256 MB XDR DRAM
  - Diseño derivado de Rambus DRAM
- **GPU:** NVIDIA RSX 550 MHz + 256 MB GDDR3

## Xbox 360

- **CPU:** PowerPC Tri-core Xenon 3.2 GHz
  - 3 cores, con 2 threads por core, 14.4 GFLOPs rendimiento de pico
  - Los cores son una variación del PPE de Cell
- **Memoria:** 512 MB GDDR3
- **GPU:** ATI Xenos 500 MHz + 10 MB eDRAM
  - Primera GPU con los shaders unificados
  - 240 GFLOPs rendimiento de pico



# La Guerra de las Consolas

## PS4

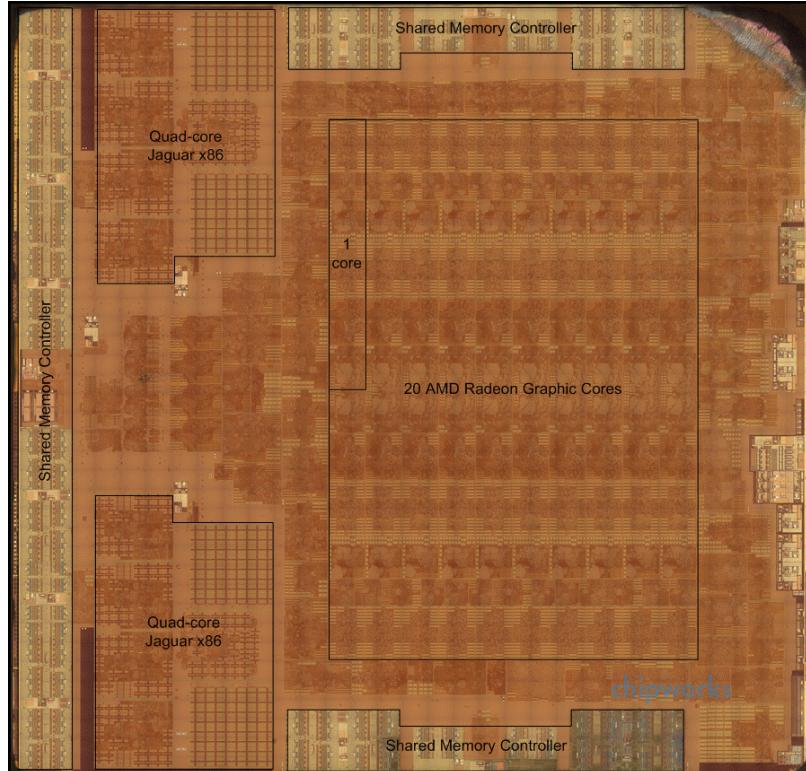
- **CPU + GPU:** integradas, chip fabricado por AMD bajo especificaciones de Sony.
  - CPU: 2 quad core Jaguar x86-64 a 1,8 GHz
  - GPU: AMD Radeon con 18 GCN, 1152 shaders, 800 MHz, 1.81 TFLOPS
  - Además dispone de un procesador ARM + 256 MB DDR3 (standby, descargas, ...) [en la placa]
- **Memoria:** 8 GB GDDR5 DRAM
  - Unificada
  - Ancho de banda 176 GB/s, 4 canales, 256 bits, 2750 MHz

## Xbox One

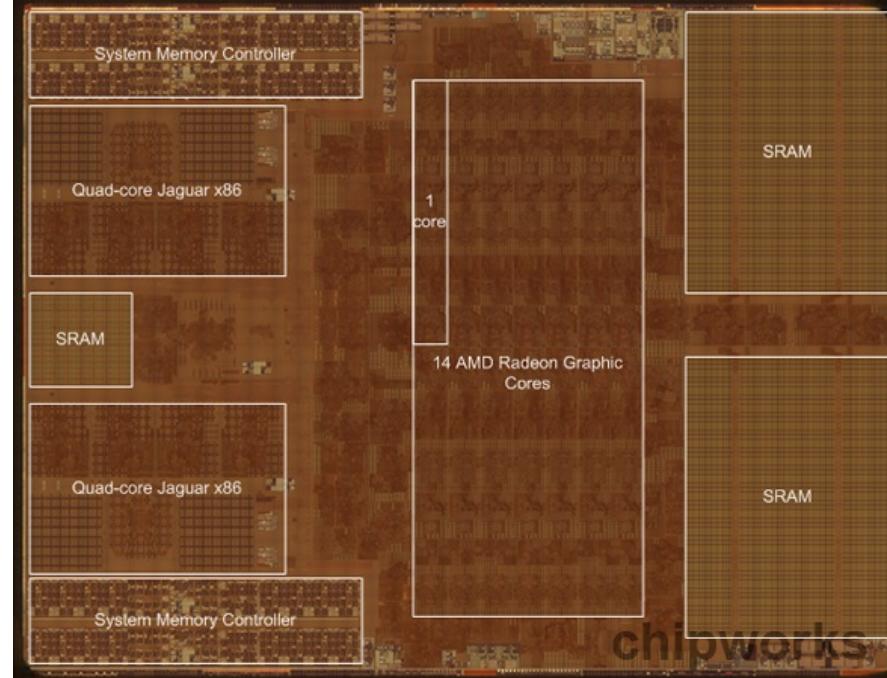
- **CPU + GPU:** integradas, chip fabricado por AMD bajo especificaciones de Microsoft.
  - CPU: 2 quad core Jaguar x86-64 a 1,6 GHz
  - GPU: AMD Radeon con 12 GCN, 768 shaders, 853 MHz, 1.31 TFLOPS
  - 32 MB de SRAM, 204 GB/s
- **Memoria:** 8 GB GDDR3 DRAM
  - Unificada
  - Ancho de banda 68 GB/s, 4 canales, 256 bits, 1066 MHz



# La Guerra de las Consolas



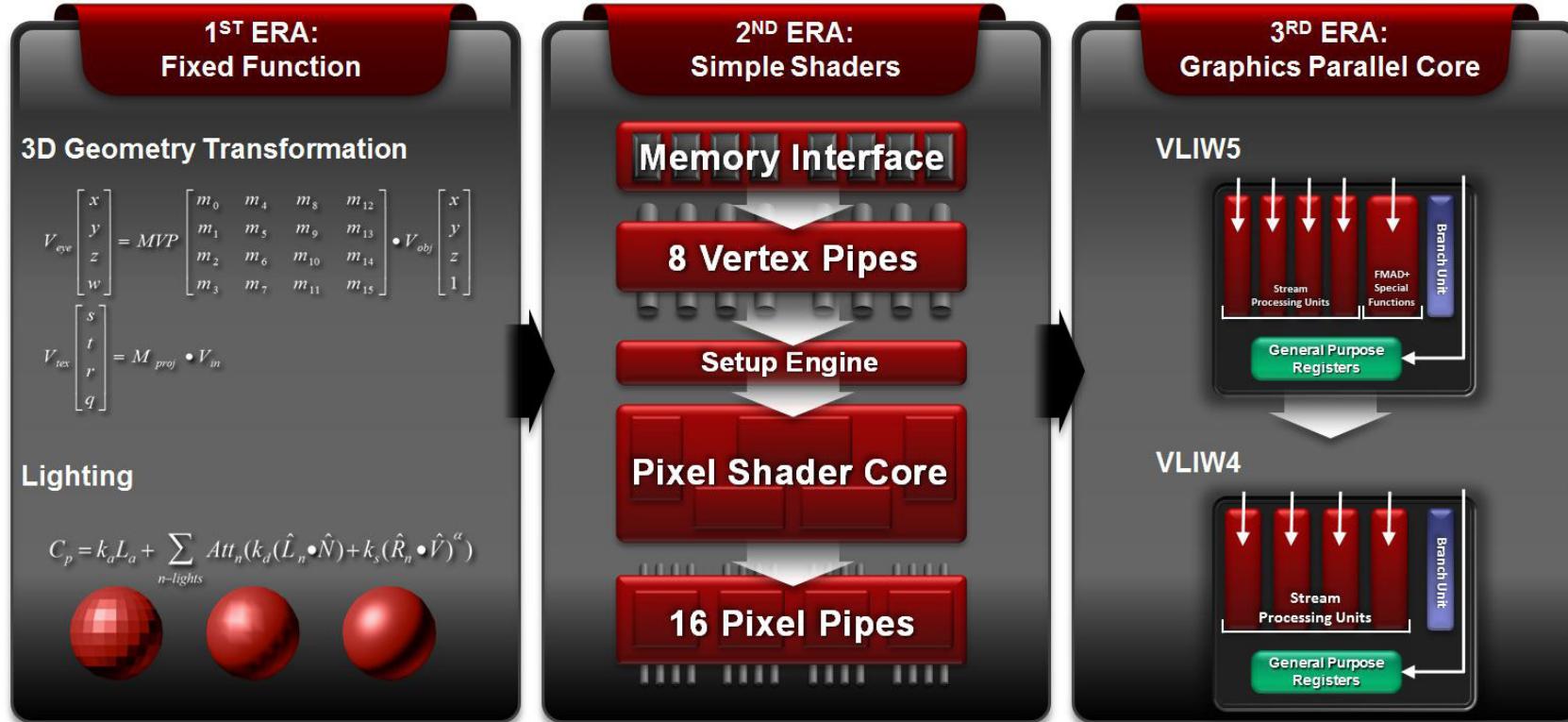
PS4: 8 cores + GPU con 18 GCNs



Xbox One: 8 cores + GPU con 12 GCNs + 32 MB SRAM



# Conclusiones Finales



# Lectura Complementaria

- John Montrym and Henry Moreton  
“The GeForce 6800”  
IEEE Micro, vol 25, issue 2, 2005
  
- “AMD Graphics Cores Next (GCN) Architecture”  
AMD White Paper, 2012





UNIVERSITAT POLITÈCNICA DE CATALUNYA  
BARCELONATECH

Departament d'Arquitectura de Computadors

# Tarjetas Gráficas y Aceleradores

## Ejemplos Comerciales

### Agustín Fernández

Departament d'Arquitectura de Computadors

Facultat d'Informàtica de Barcelona

Universitat Politècnica de Catalunya

