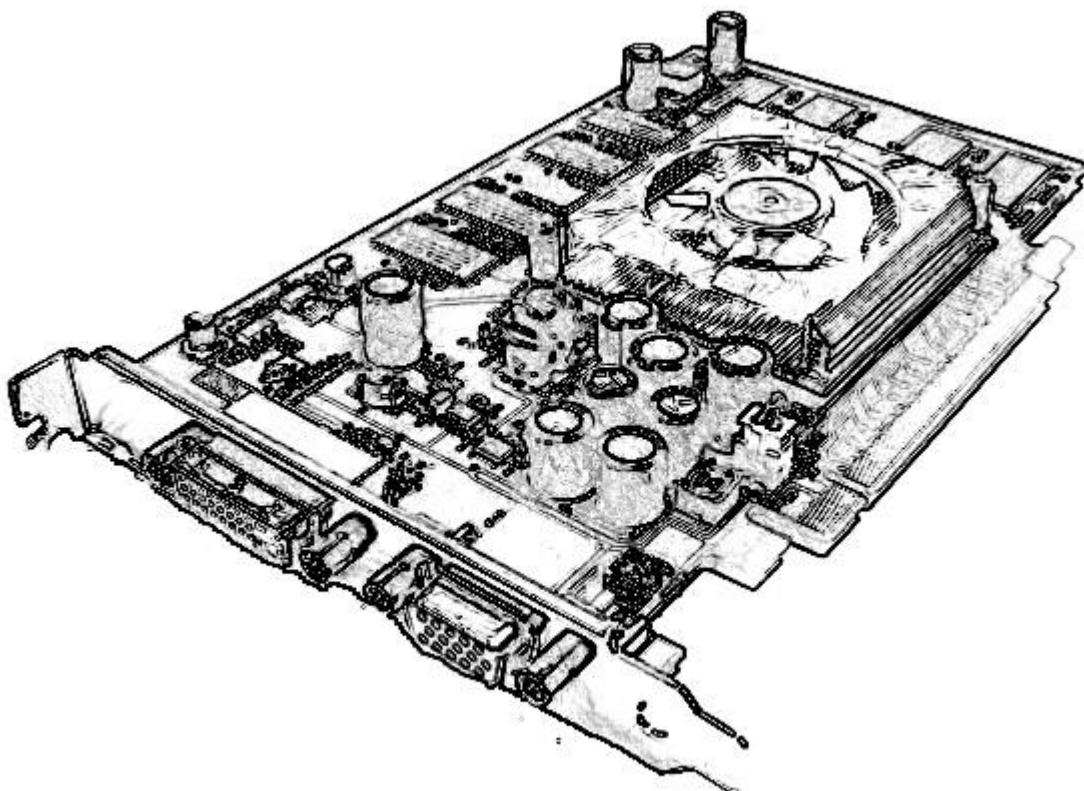


AB

Architektura systemów komputerowych

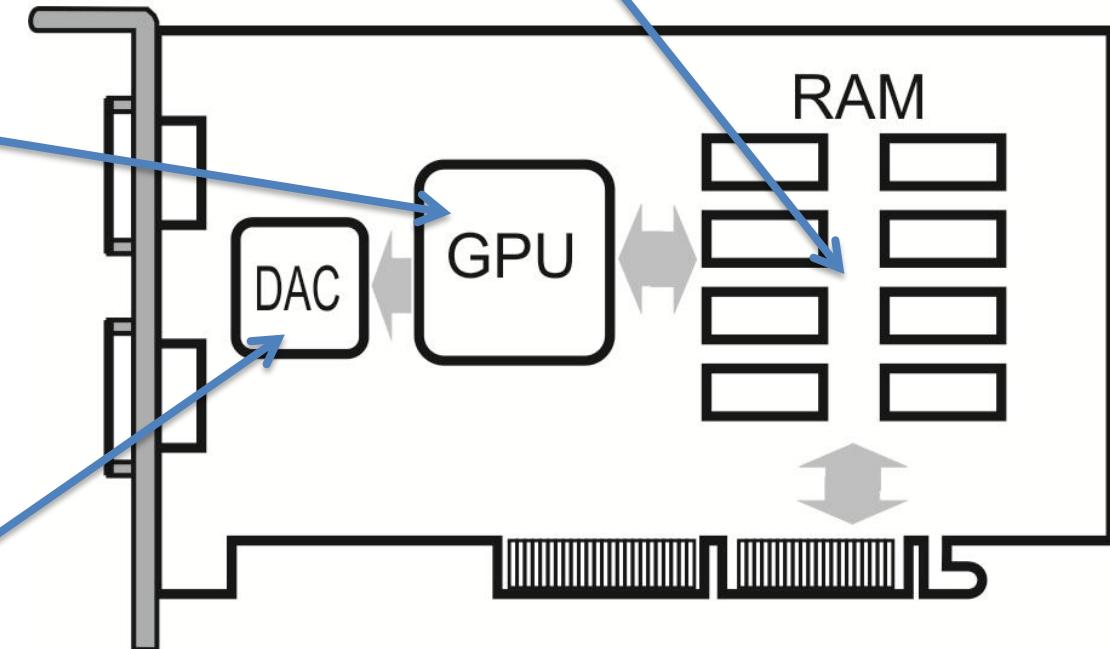
Budowa karty graficznej



Akceleratory graficzne

Graphics Processing Unit – jednostka obliczeniowa specjalizowana do wspomagania obliczeń grafiki 3D

Pamięć karty – służy do przechowywania bitmapy przygotowanej do wyświetlenia oraz jako pamięć operacyjna dla procesora GPU



Digital-analog converter – układ zamieniający bitmapę obrazu zawartą w pamięci RAM na sygnał sterujący monitorem (analogowy lub cyfrowy)

Akceleratory graficzne

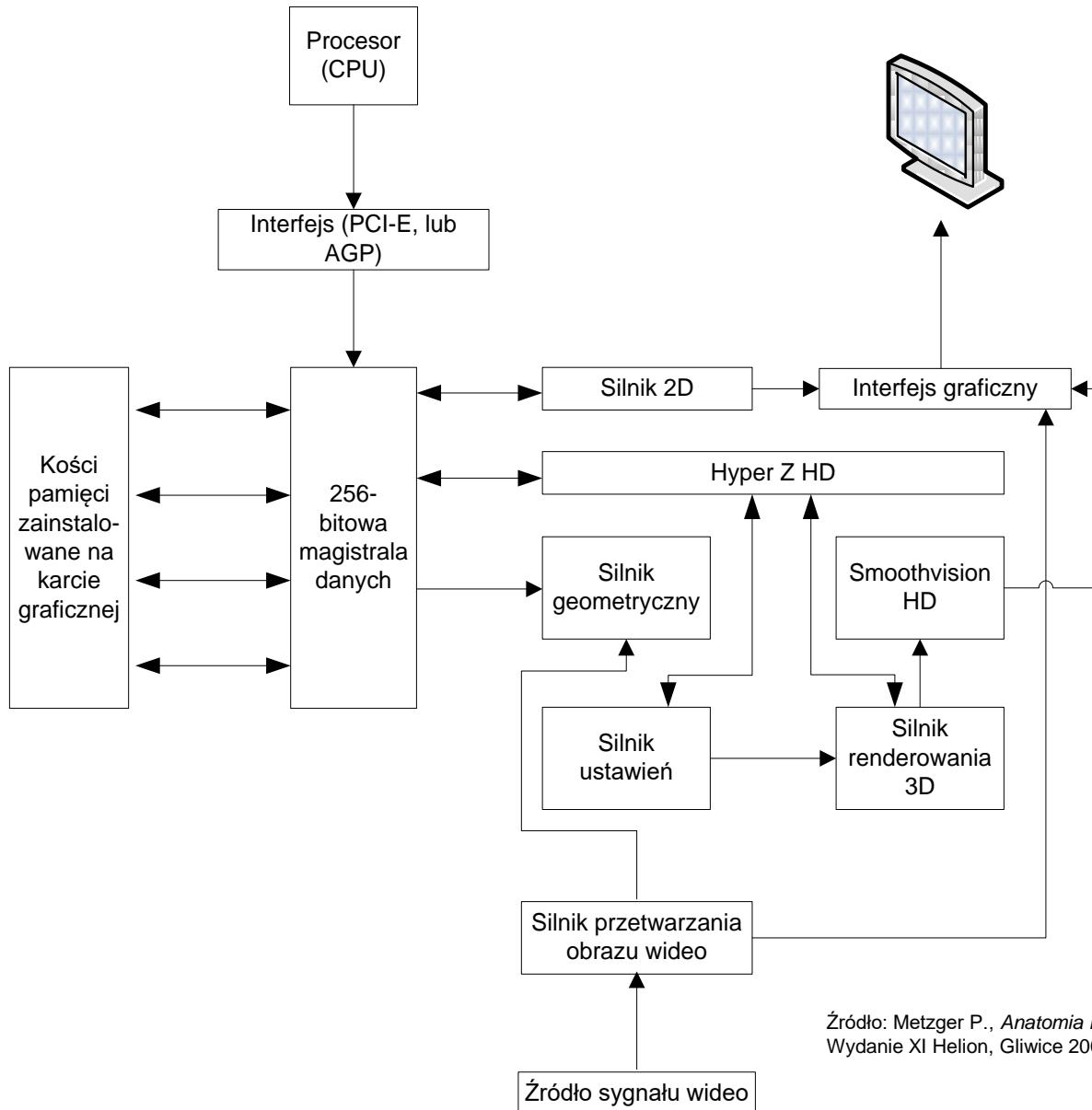
Współczesne karty graficzne zwykle składają się z pięciu komponentów:

1. Interfejsu systemowego,
2. Pamięci graficznej,
3. Procesora graficznego GPU (Graphics Processing Unit),
4. Bufora ramki
5. Układu RAMDAC (Random Access Digital/Analog Converter).

Akceleratory graficzne

- ✓ **Układy takie lokalizuje się** (w odróżnieniu od koprocesora arytmetycznego) nie w rejonie procesora ale blisko przetwarzanych danych (pamięci obrazu) - na karcie graficznej. Biorą one na siebie fragment procesu dekodowania obrazu video, przez co zwalniają jednostkę centralną (CPU) z ogromnej ilości operacji.
- ✓ **Odzyskana moc obliczeniowa** stawiana jest do dyspozycji systemu operacyjnego a odciążone magistrale systemowe umożliwiają szybszy dostęp do zasobów: pamięci RAM, dysków i peryferii.

Budowa akceleratora graficznego



Źródło: Metzger P., *Anatomia PC.*,
Wydanie XI Helion, Gliwice 2007, ISBN: 978-83-246-1119-5

Budowa akceleratora graficznego

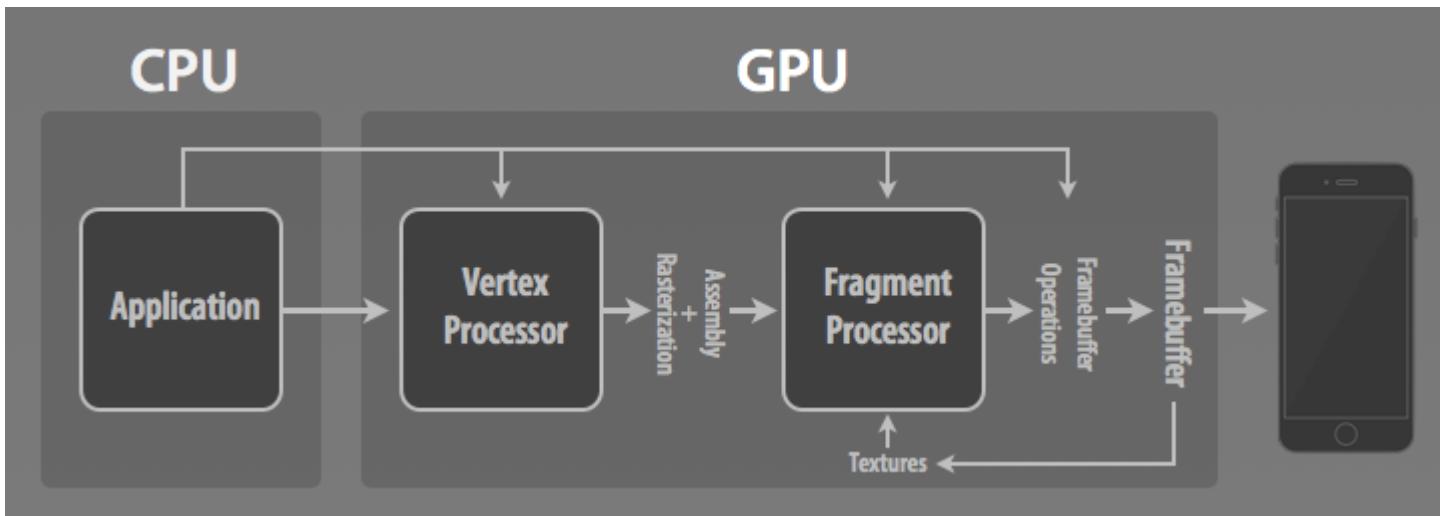
1. Procesor wysyła do układu graficznego żądanie wyświetlenia w danym miejscu na monitorze grupy obiektów trójwymiarowych.
2. Sygnał kierowany jest do magistrali danych.
3. Elementy, które są niewidoczne trafiają do jednostki Hyper Z.
4. Pozostałe dane wędrują do silnika geometrycznego (Vertex Engine), po czym zostają poddane wstępnej obróbce i trafiają do silnika ustawień (Setup Engine).
5. Dane obliczeniowe obiektów, które na tym etapie także zostały uznane za niewidoczne, kierowane są do Hyper Z.
6. Pozostałe elementy poddawane są ostatecznej obróbce w jednostce nazwanej Smooothvision HD czyli jednostce która poprawia jakość obrazu, np. jednostka anti-aliasingu.
7. Następnie gotowe już obiekty trafiają do interfejsu graficznego, w którym na podstawie wyliczonych danych generowany jest obraz.
8. Stąd przesyłany jest on do urządzenia wyświetlającego (np. monitora lub telewizora) przez złącze VGA, DVI, S-VIDEO czy HDMI.

Karta graficzna ze zintegrowanymi shaderami

Shader – krótki program komputerowy często napisany w specjalnym języku (*shader language*), który w grafice trójwymiarowej opisuje właściwości pikseli oraz wierzchołków.

Pod tą nazwą rozumiemy też **programowalną jednostkę** wykonującą wyżej wymienione operacje.

Technologia ta zastąpiła stosowaną wcześniej jednostkę **T&L** - Transform & Lighting – (ang. transformacji i oświetlenia)



Karta graficzna ze zintegrowanymi shaderami

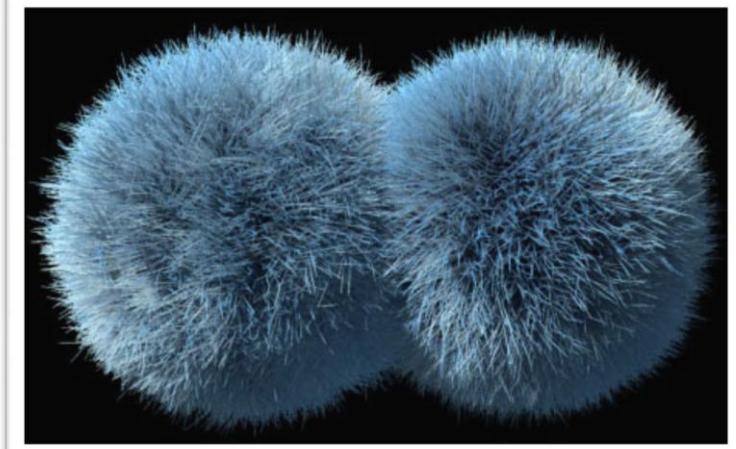
Obecnie wyróżniamy trzy klasy shaderów:

- ✓ **Vertex Shader** - Cieniowanie wierzchołkowe – uruchamiane jest raz dla poszczególnych przetwarzanych wierzchołków. Jego zadaniem jest transformacja położenia wierzchołka w wirtualnej przestrzeni 3D na współrzędne 2D na ekranie.
- ✓ **Pixel Shader** lub **Fragment Shader** - Cieniowanie pikseli – jest programowalną jednostką odpowiadającą za wyliczanie koloru pikseli.
- ✓ **Geometry Shader** (cieniowanie geometryczne) - w procesorach obsługujących DirectX 10 oprócz dotychczas stosowanych vertex- i pixel shaderów dodano jeszcze kolejną warstwę modelowania shaderowego. Pozwala na dodawanie lub usuwanie wierzchołków z siatki wierzchołków (ang. mesh). Może być używane do proceduralnego tworzenia obiektów geometrycznych albo do dodawania objętościowych detali istniejących siatek wierzchołków

Karta graficzna ze zintegrowanymi shaderami

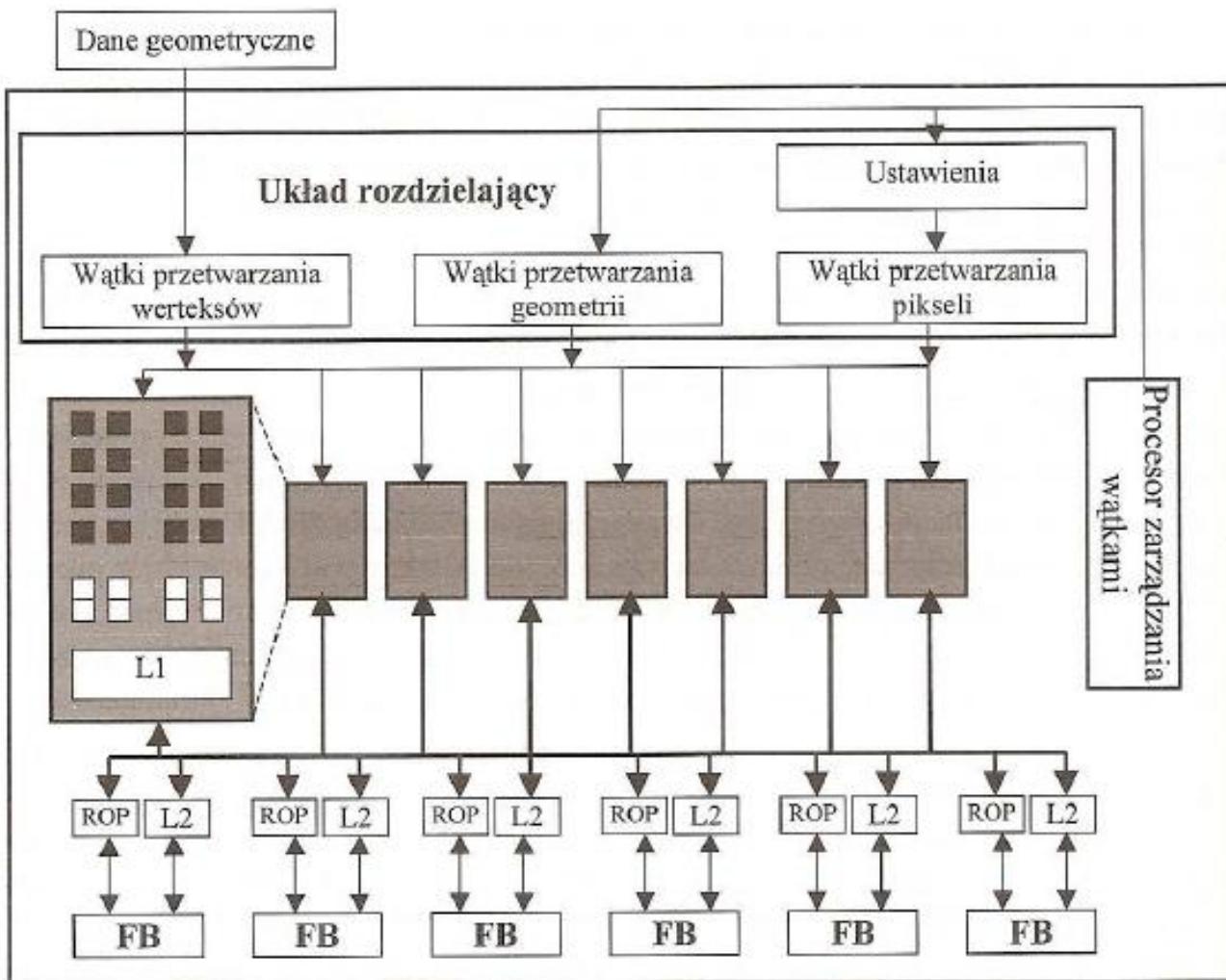
Możliwości Geometry Shadera stanowią doskonałe uzupełnienie funkcjonalności dotychczasowych jednostek. Potrafi on między innymi: przeprowadzać operacje łączenia wierzchołków,

- symulacje fizyczne (np. poruszające się włosy),
- tworzyć paski trójkątów i linii,
- automatycznie generować cienie szablonowe.



https://www.youtube.com/watch?v=T2Wo9eFw_el

Karta graficzna ze zunifikowanymi shaderami



■ – procesory potoków SP

ROP – układy wyjściowe (Render Out Pixel)

□ – układy teksturujące

L1, L2 – wewnętrzne pamięci cache

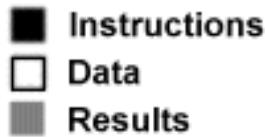
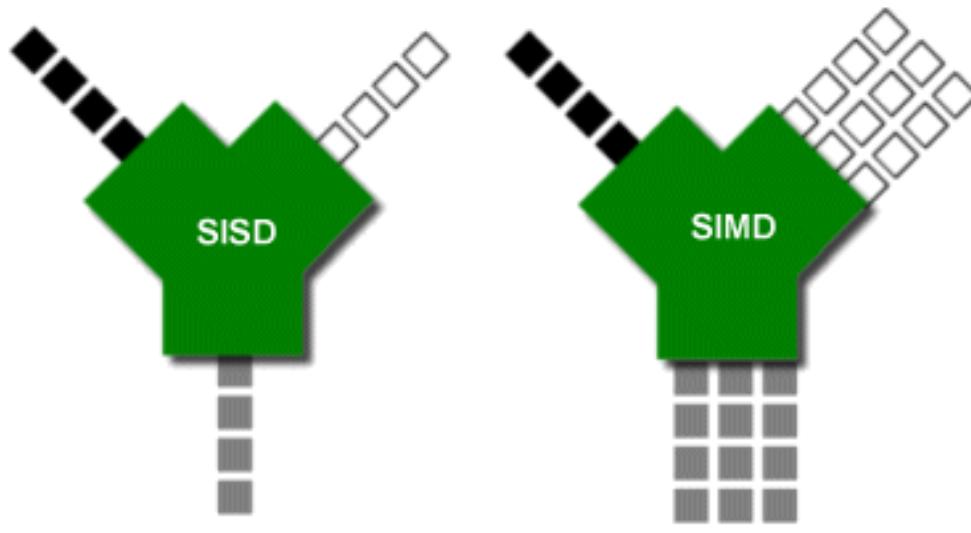
FB

– pamięć karty graficznej i bufory ramki

Karta graficzna ze zintegrowanymi shaderami

We wcześniejszych kartach rodzaje shakerów różniły się sprzętowo. Współcześnie specjalizowane shadery zastąpione zostały przez **shader-y zunifikowane**.

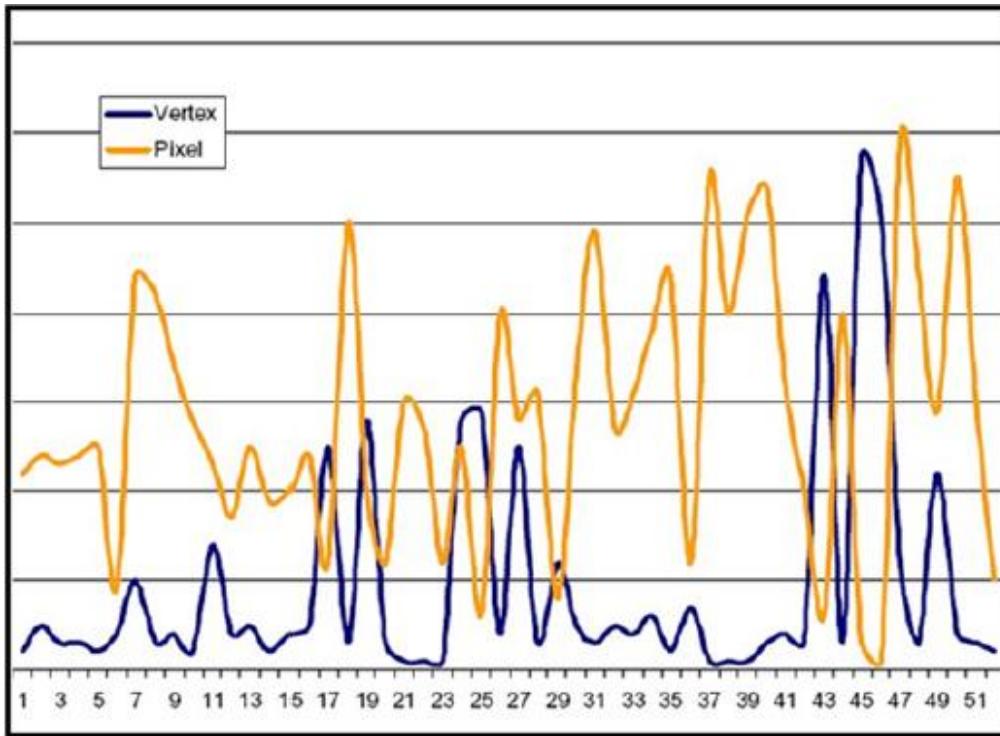
Jednostki zunifikowanych shaderów są wektorowe, obsługują w trybie SIMD (ang. Single Instruction Multiple Data).



- SISD (ang Single Instruction Single Data)
- SIMD (ang. Single Instruction Multiple Data).

Karta graficzna ze zunifikowanymi shaderami

Przykładowe obciążenie shaderów w grze.



Źródło: <https://pclab.pl/art23133.html>

Na powyższym wykresie zauważać można, że w wielu przypadkach maksymalne zapotrzebowanie na jeden typ shaderów występuje w tym samym czasie, gdy jednostki shaderów drugiego typu nie mają zbyt wiele do roboty

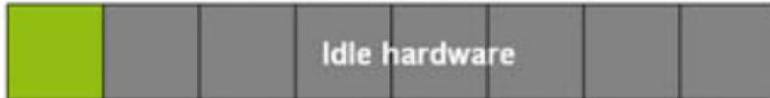
Karta graficzna ze zunifikowanymi shaderami

Why unify?

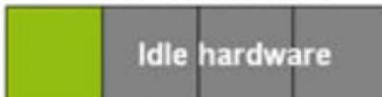
Vertex Shader



Pixel Shader



Vertex Shader



Pixel Shader



Heavy Geometry

Workload Perf = 4



Heavy Pixel

Workload Perf = 8

Karta graficzna ze zunifikowanymi shaderami

Why unify?

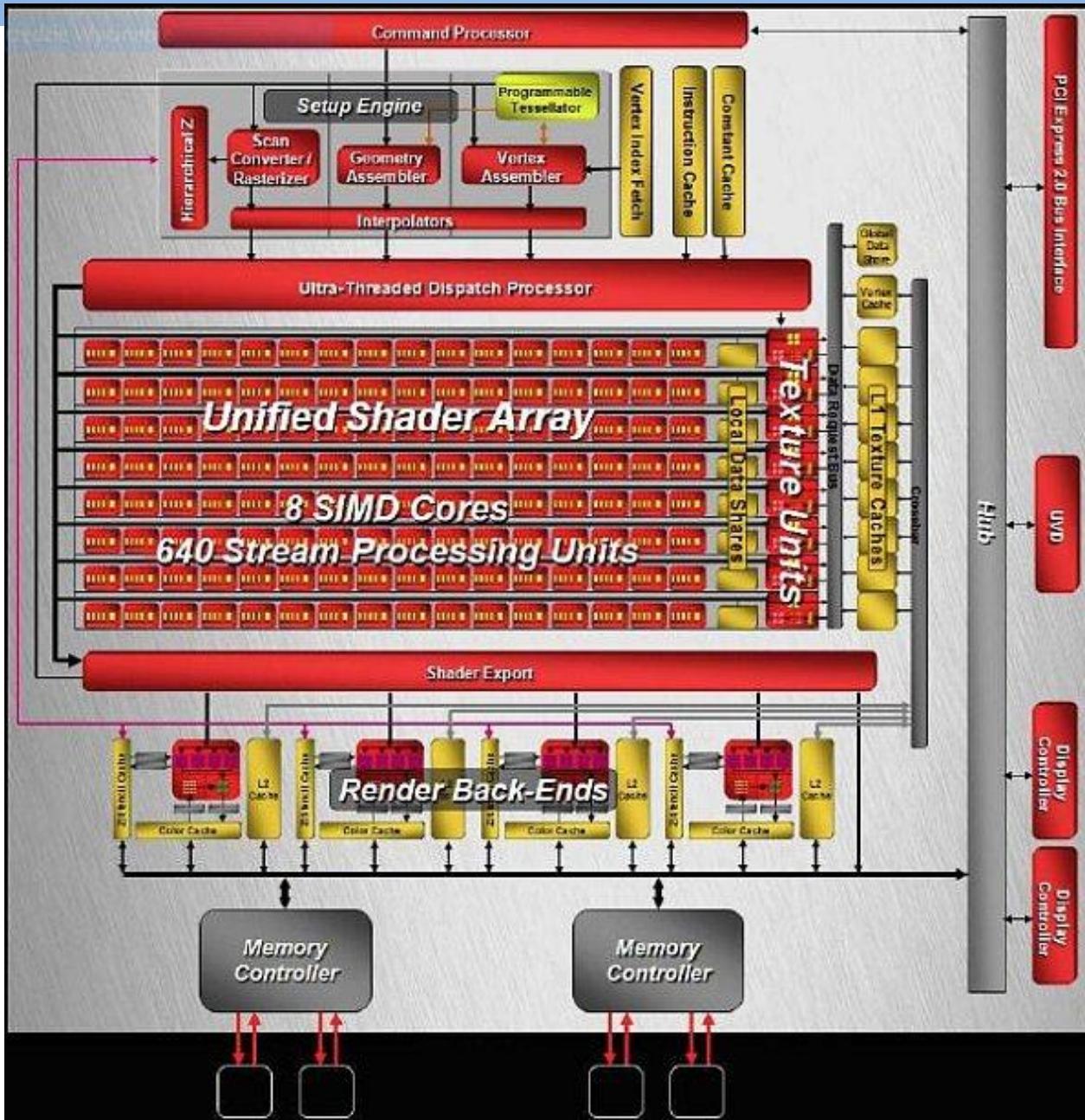


**Heavy Geometry
Workload Perf = 12**



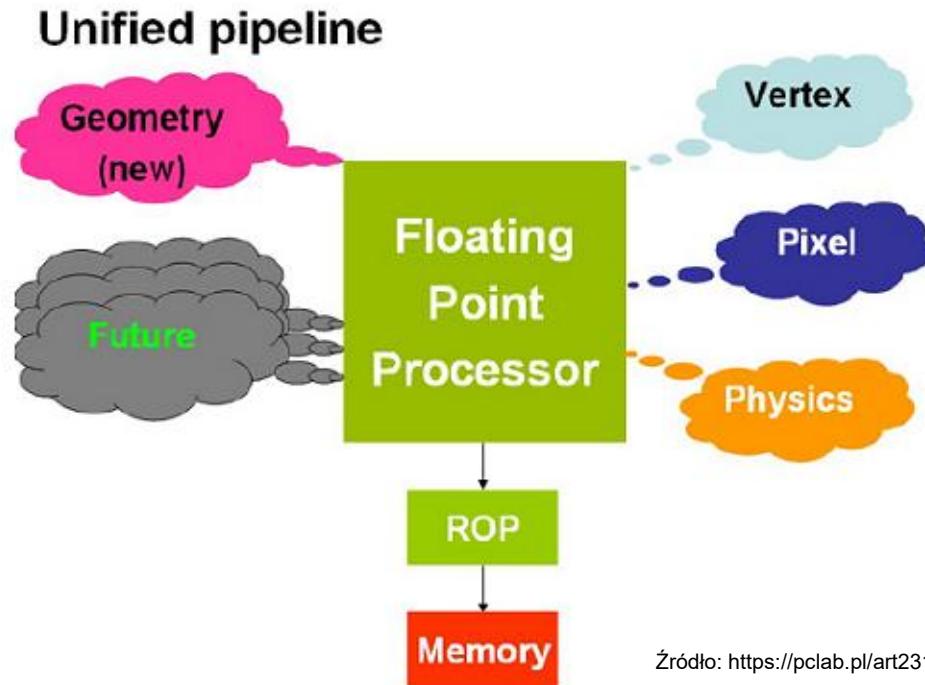
**Heavy Pixel
Workload Perf = 12**

Karta graficzna ze zunifikowanymi shaderami



Źródło:
<http://images.google.pl/imgres?imgurl=http://>

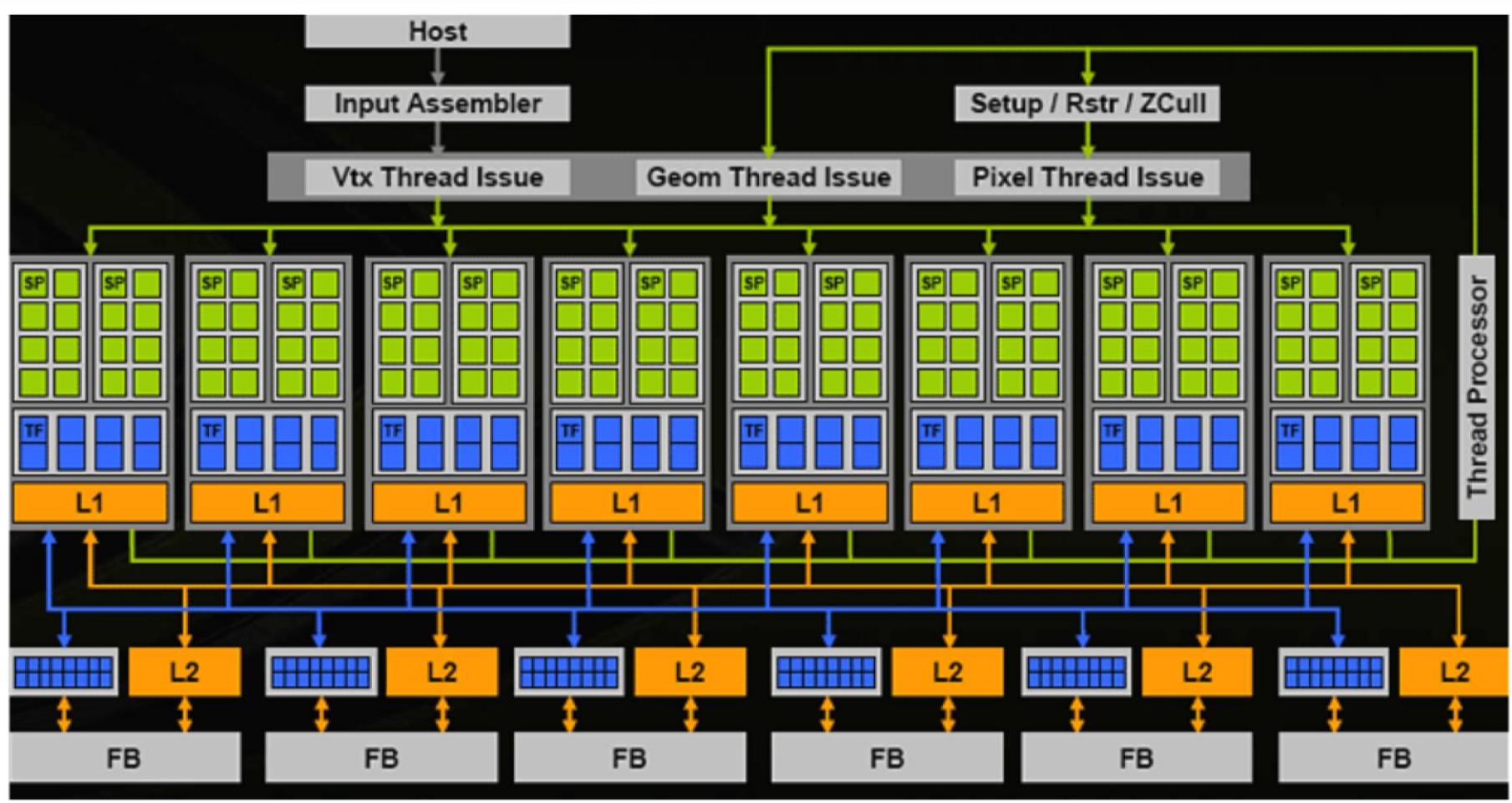
NVIDIA CUDA (ang. Compute Unified Device Architecture).



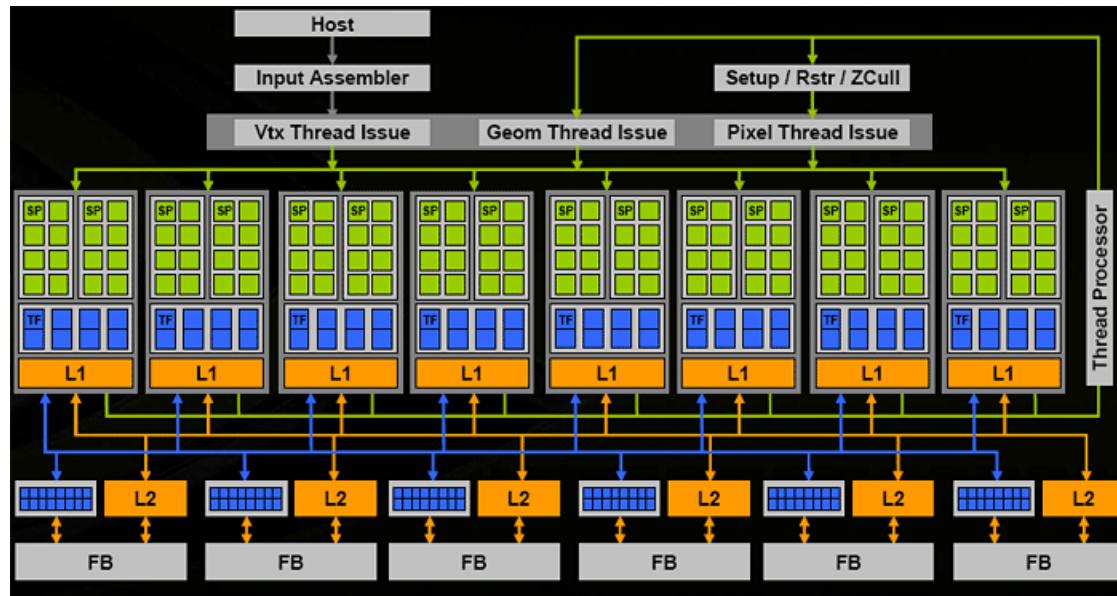
Źródło: <https://pclab.pl/art23133-2>

Vertex-, pixel- i geometry shader przetwarzane w tym samym układzie shadera, z „wolnym miejscem” na przetwarzanie także warstwy fizycznej gier i innych, przyszłych zastosowań

CUDA



CUDA



Źródło: <https://pclab.pl/art23133-2>

W tej architekturze GPU posiada niezależne, zunifikowane jednostki cieniowania (Shadery), z których każda składa się z jednostek obliczeniowych SP - Streaming Processor. Są to 32-bitowe skalarne jednostki zmiennoprzecinkowe (zamiast wektorowych zunifikowanych shaderów).

W shaderze, znajduje się także zestaw modułów TF odpowiedzialnych za operacje na teksturach (na schemacie oznaczone kolorem niebieskim). Moduły zoptymalizowano pod kątem obliczeń HDR

ATI vs. NVIDIA

Klaster z 24 procesorami

Klaster TPC tworzą 3 bloki po 8 procesorów strumieniowych z własną pamięcią.

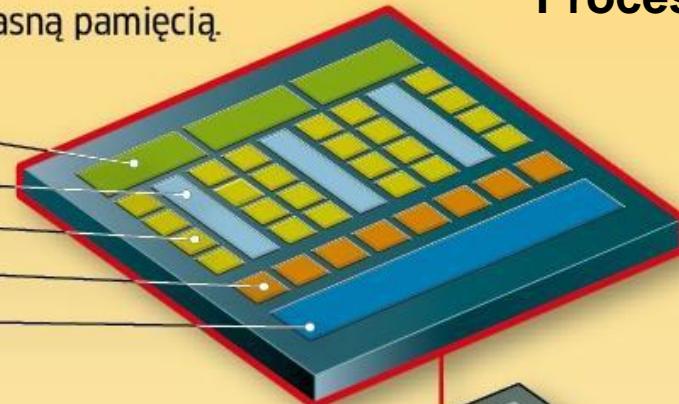
Jednostka sterująca

Pamięć lokalna

Procesory strumieniowe

Jednostki teksturowania

Pamięć podrzczna L1



TPC - Texture /
Processor Cluster

Procesor graficzny

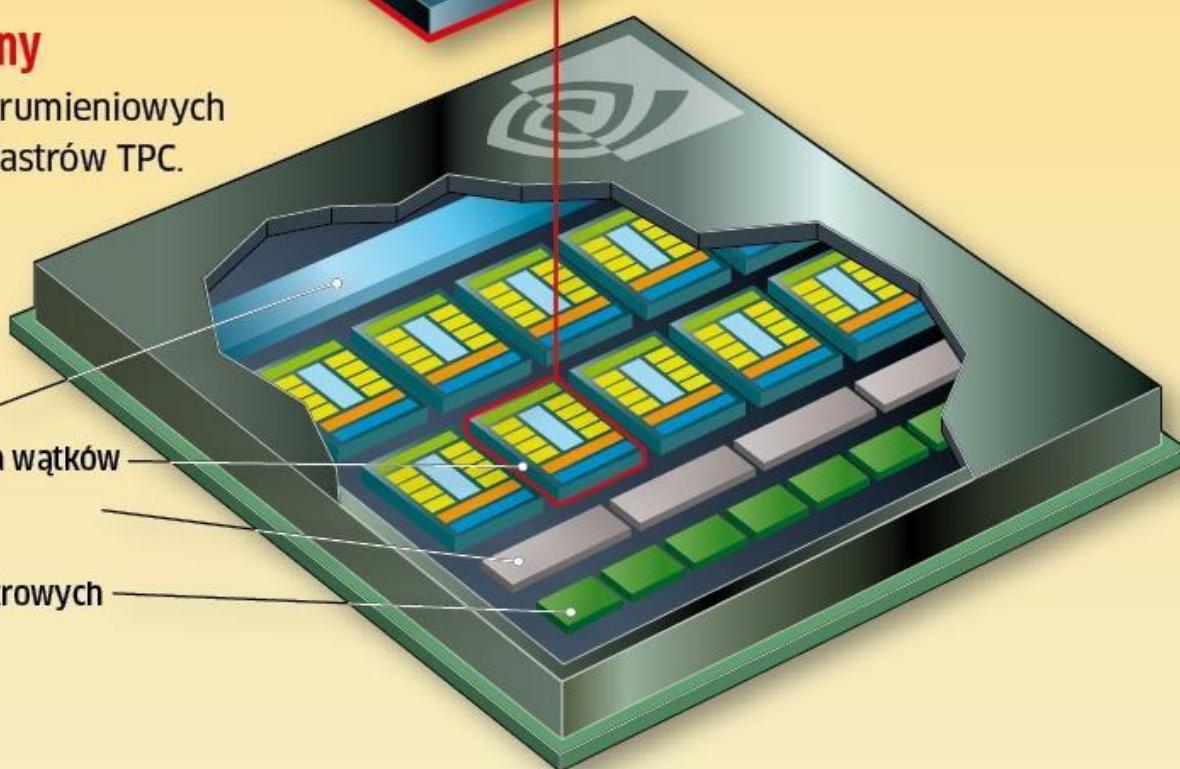
240 procesorów strumieniowych
składa się na 10 klastrów TPC.

Rozdzielacz wątków

Klaster przetwarzania wątków

Pamięć podrzczna L2

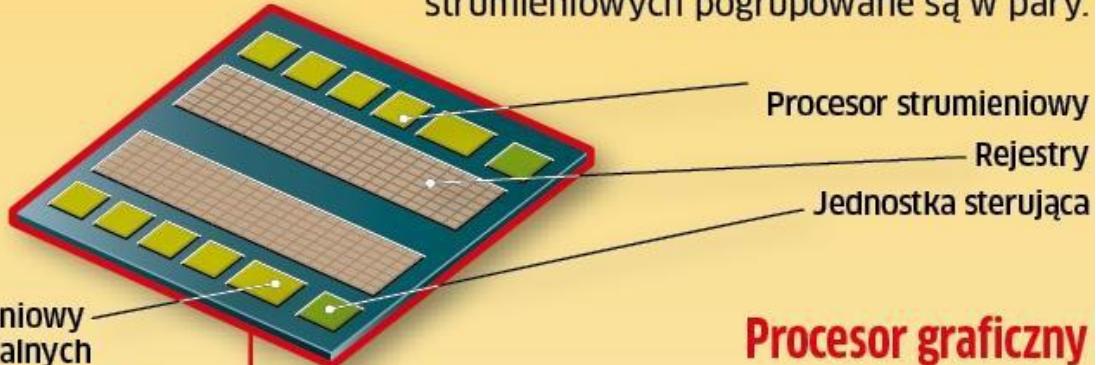
Procesor operacji rastrowych



ATI vs. NVIDIA

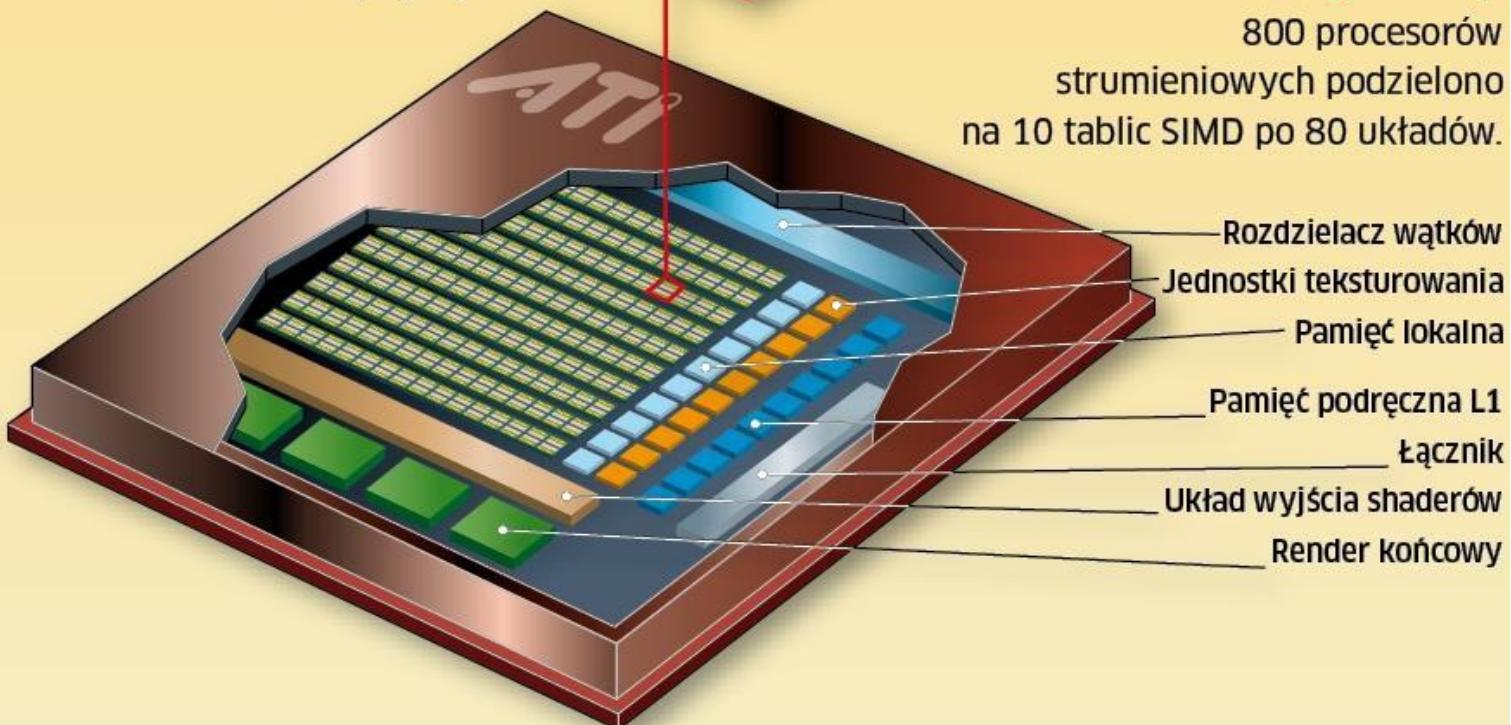
Pary 5-procesorowych jednostek

Jednostki zawierające po 5 procesorów strumieniowych pogrupowane są w pary.



Procesor graficzny

800 procesorów strumieniowych podzielono na 10 tablic SIMD po 80 układów.



SLI

MULTI-GPU
TECHNOLOGY



SLI

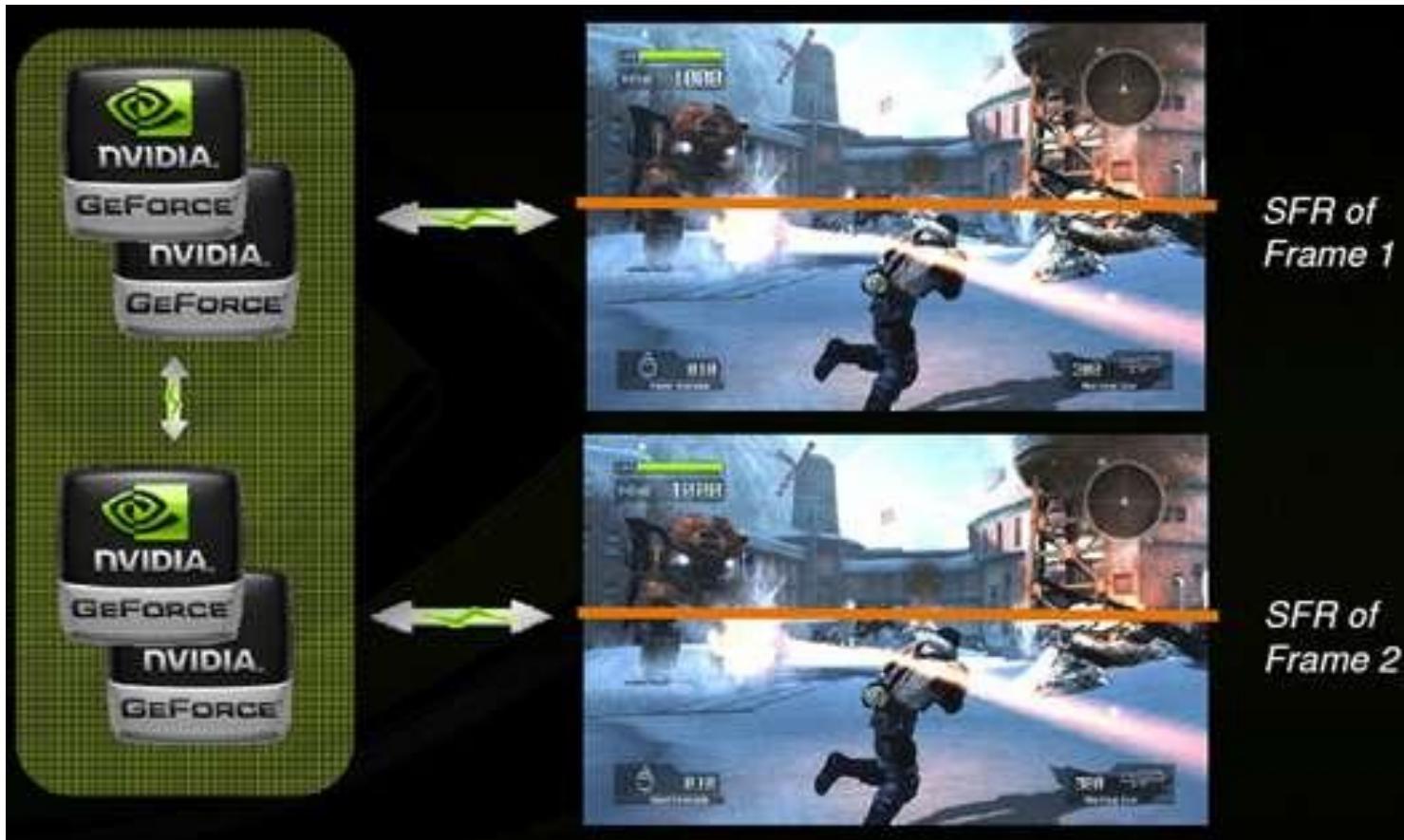
SLI (ang. Scalable Link Interface) rozwiązanie firmy NVIDIA, w którym obraz generowany jest przez co najmniej dwie karty graficzne połączone mostkiem SLI.

Aby korzystać z trybu SLI, należy posiadać płytę główną oraz karty graficzne wspierające tą technologię.

Rodzaje SLI to:

- Quad SLI (dwie karty graficzne mające po dwa układy GPU),
- 3-Way SLI (generowanie obrazu przez 3 karty graficzne)
- Hybrid SLI (generowanie obrazu przez kartę graficzną w złączu PCIe we współpracy z chipsetem graficznym na płycie głównej).

SLI – podział obrazu



SLI



CrossFire



CrossFire

CrossFire – konkurencyjne dla SLI rozwiązanie firmy ATI. Aby skorzystać z trybu CrossFire należy posiadać płytę główną oraz karty graficzne wspierające tę technologię.

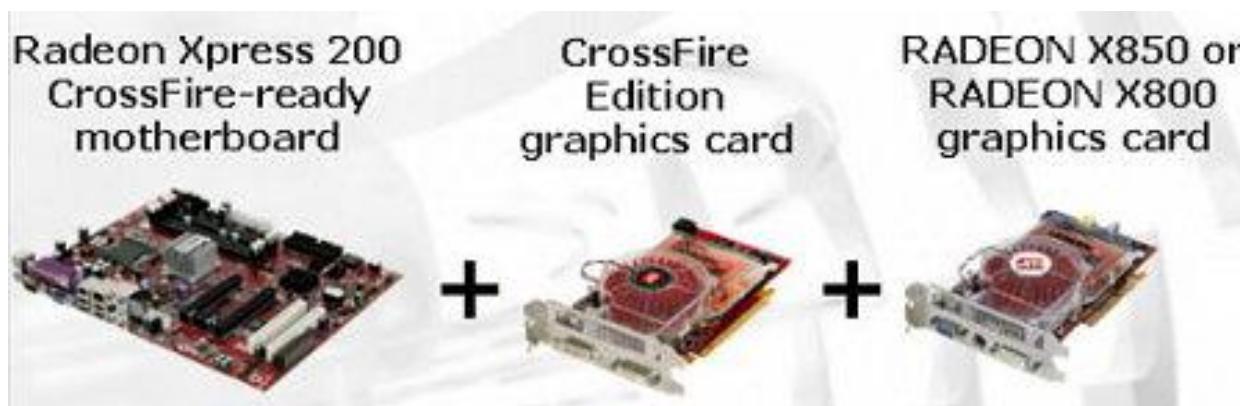
Początkowo karty działały w trybie master-slave, nowe płyty główne pozwalają łączyć karty bez konieczności korzystania z karty master. Karty rozszerzeń również połączone są mostkiem.

Rodzaje CrossFire to:

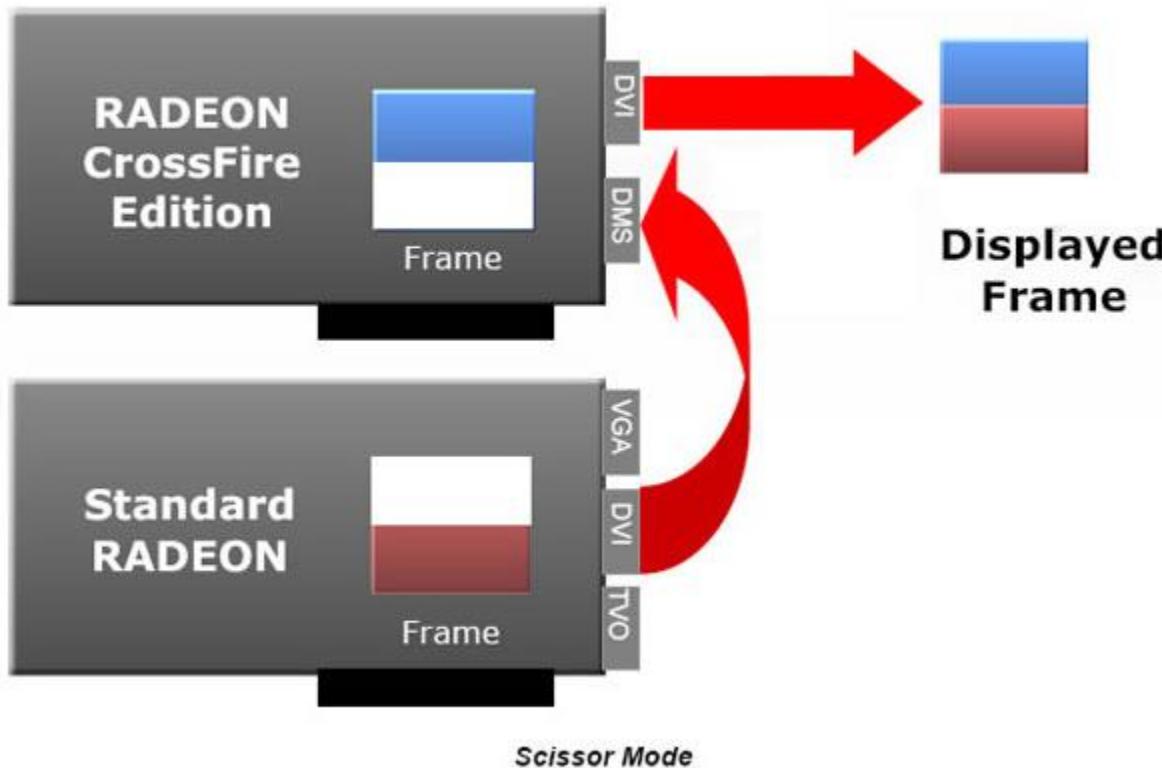
- CrossFire (generowanie obrazu przez dwie karty graficzne),
- CrossFireX (generowanie obrazu nawet przez 4 karty graficzne),
- Hybrid CrossFireX (generowanie obrazu przez kartę graficzną w złączu PCI-E we współpracy z chipsetem graficznym na płycie głównej).

Crossfire

W większości wypadków potrzebna jest karta **CrossFire Edition**. Zawiera ona w sobie układ, który będąc przekąźnikiem TDMS łączy obraz w całość. W tym przypadku musimy też łączyć karty dodatkowym kablem.

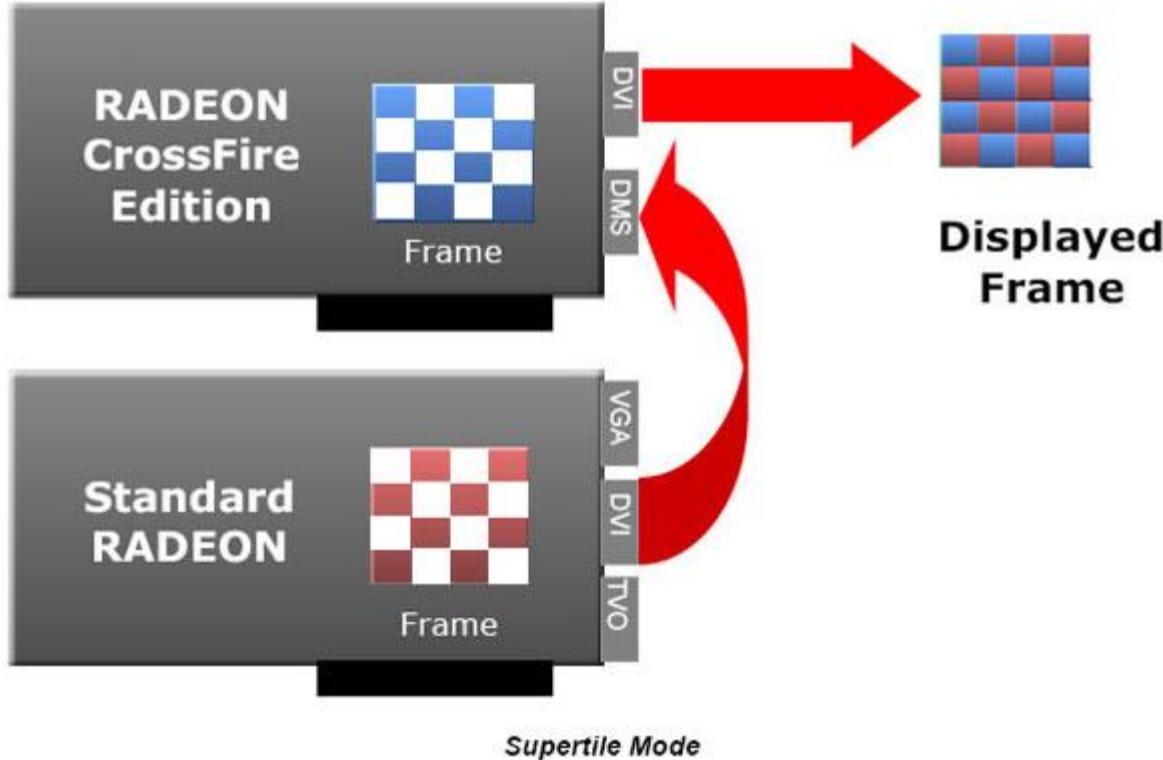


Crossfire - trzy możliwości renderingu



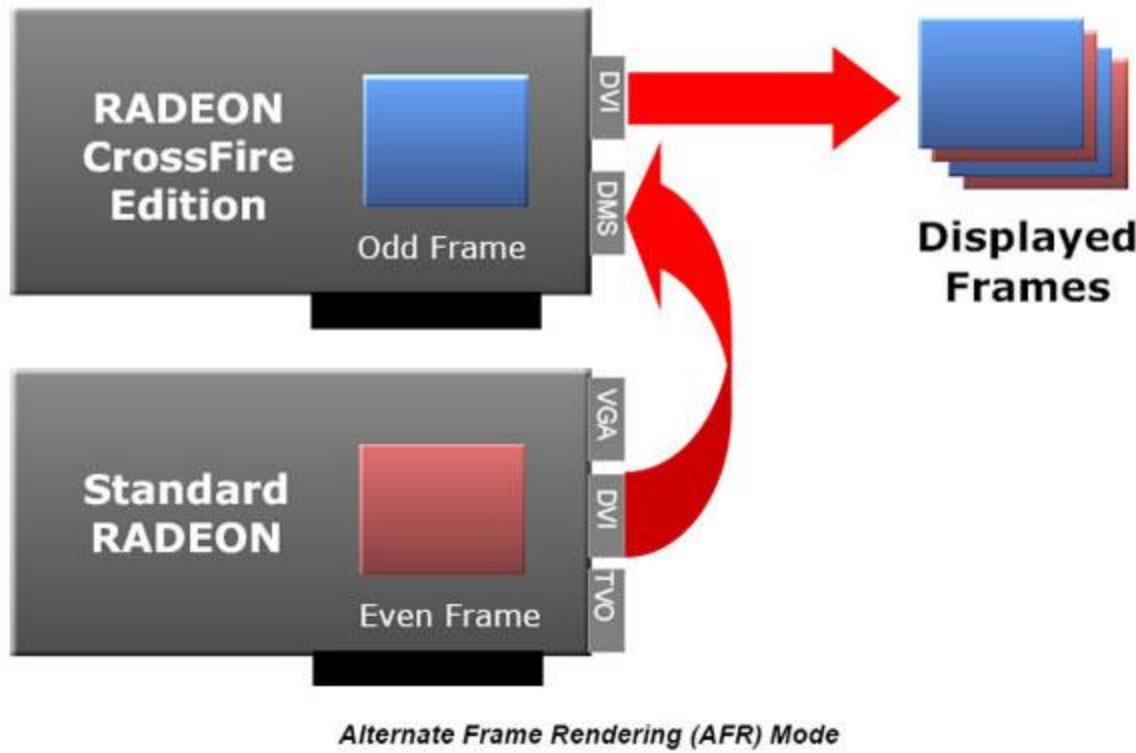
Scissor: Obraz jest dzielony w poziomie na 2 połowy. Podobnie jak w przypadku nVidii linia podziału przesuwa się w zależności od obciążenia obu kart. Pierwsza karta renderuje górną część ramki, a druga dolną.

Crossfire - trzy możliwości renderingu



Alternate Frame Rendering: karty renderują obraz naprzemiennie, podobnie jak u nVidii.

Crossfire - trzy możliwości renderingu



CrossFire działa najlepiej jeżeli funkcja **Catalyst A.I** jest aktywna. Jeżeli jest wyłączona domyślnie aktywowane jest tryb **SuperTiling** dla gier **Direct3D** lub **Scissor** dla gier **OpenGL**. Jeżeli jest włączona sterowniki mogą dodatkowo włączyć tryb **AFFR**.

Trochę historii

http://pclab.pl/coreclub/art1428_Historia_kart_graficznych_1975-2010.html



API dla grafiki 3D



VS

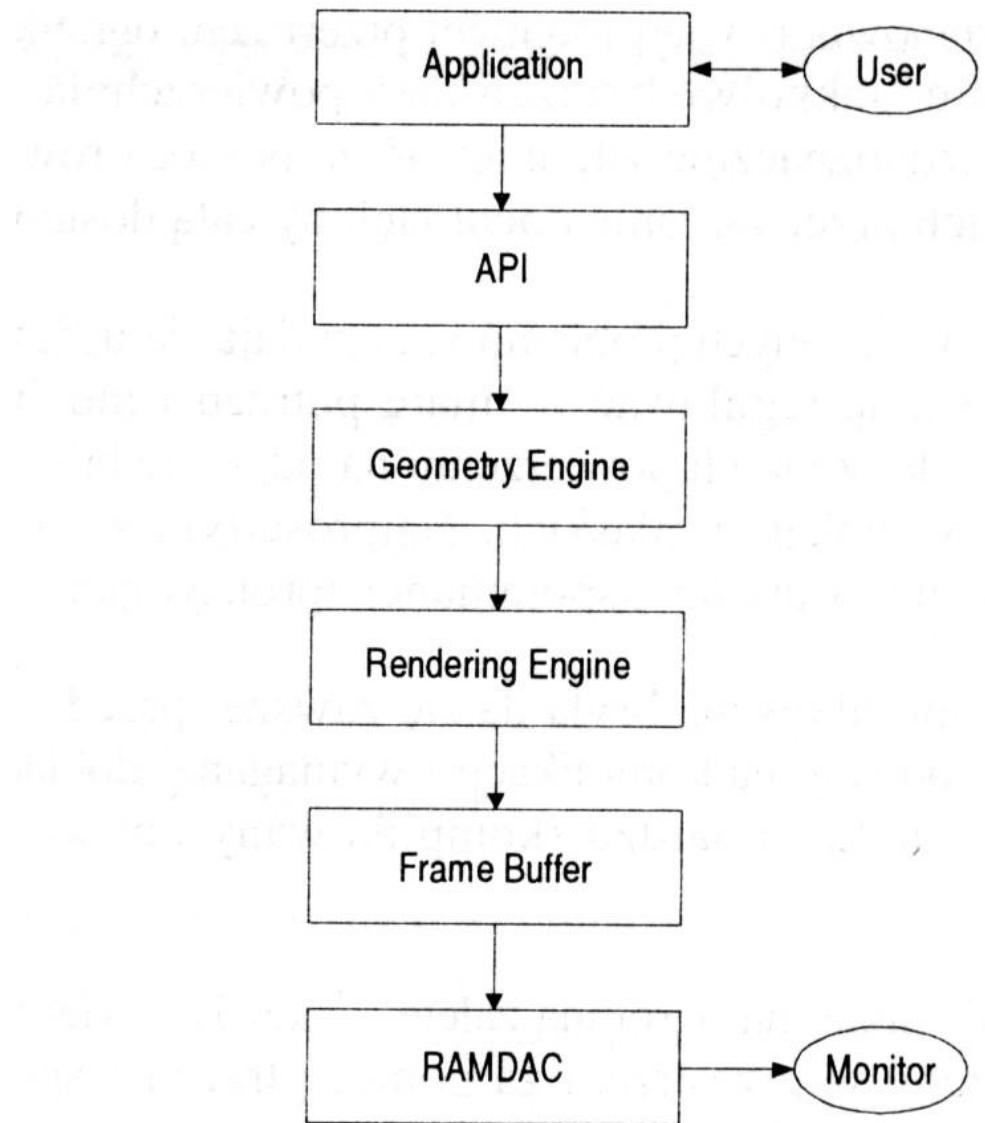


API (ang. Application Programming Interface) biblioteki stanowiące pomost pomiędzy oprogramowaniem multimedialnym a sprzętem typu karta graficzna, karta dźwiękowa.

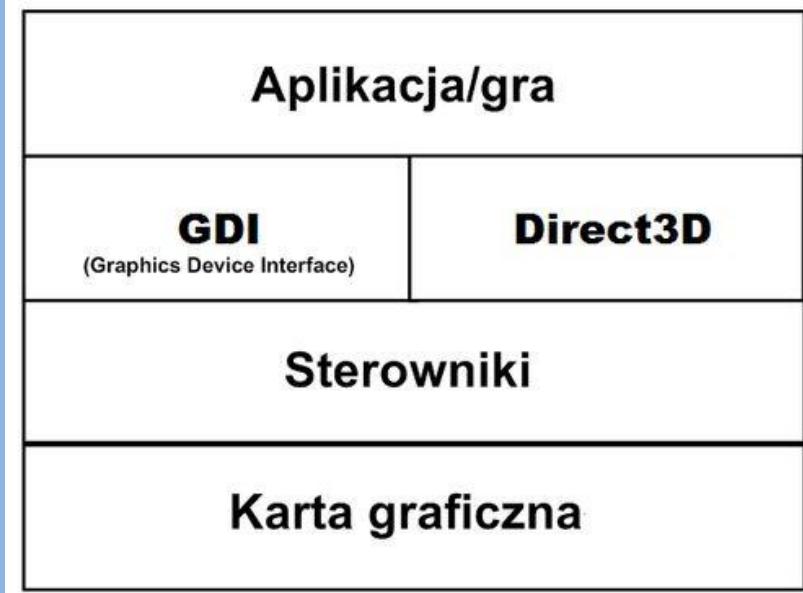
Dominującą rolę na rynku mają dwa API:

- **DirectX** jest produktem firmy Microsoft, dostępnym dla systemów Windows oraz konsoli XBox.
- **OpenGL** to otwarta biblioteka graficzna - jest to zestaw funkcji umożliwiających budowanie złożonych trójwymiarowych scen z podstawowych figur geometrycznych. OpenGL dostępny jest w wielu systemach operacyjnych.

Potok przetwarzania grafiki 3D



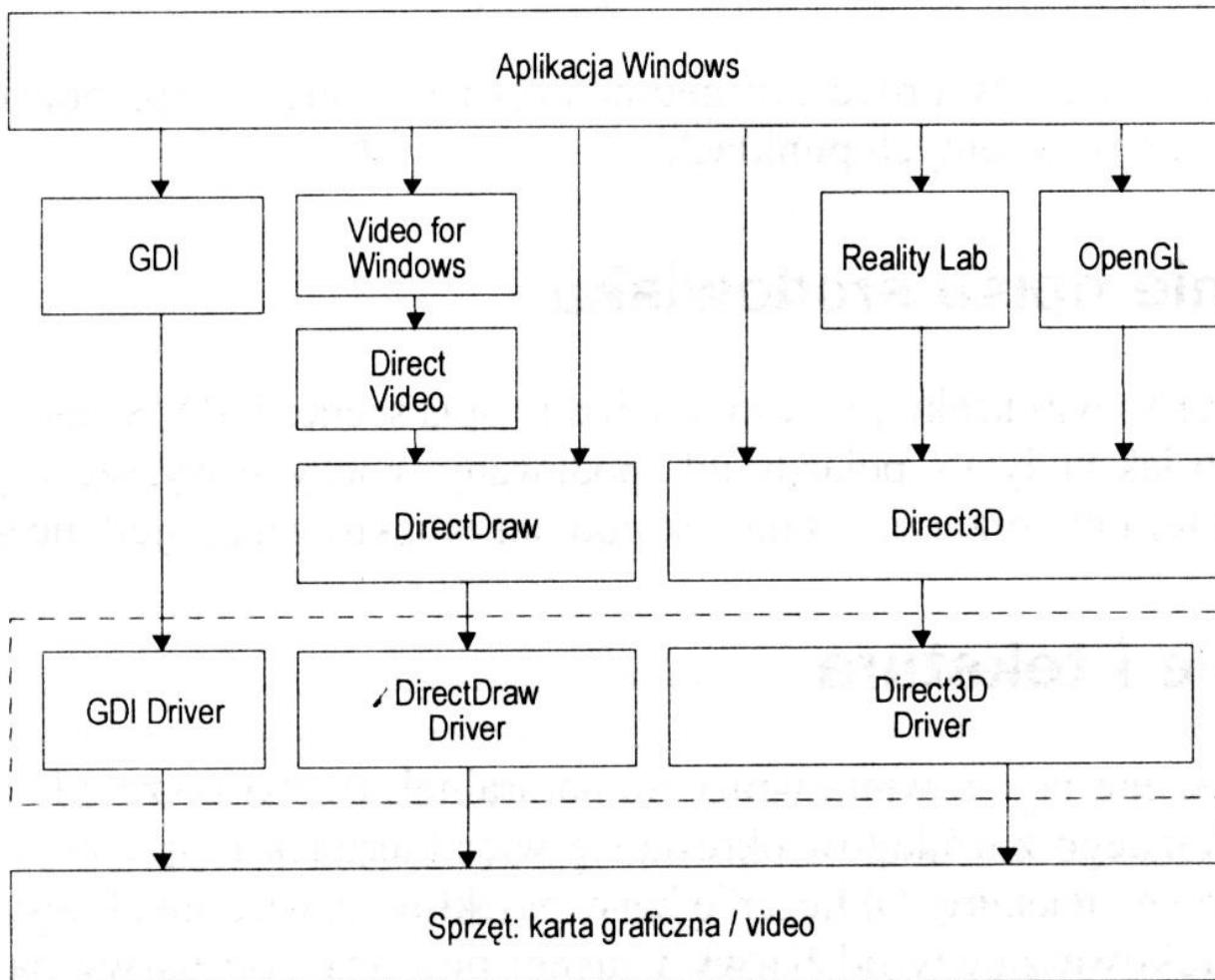
DirectX



Interfejs programowania aplikacji API (ang. Application Programming Interface) umożliwia twórcom oprogramowania pisanie programów pracujących wydajnie i niezawodnie na różnych platformach, a procedurom sprzętu pisanie sterowników współpracujących z dowolnym oprogramowaniem.

- ✓ API nadzoruje wymianę informacji pomiędzy kartą graficzną zainstalowaną w komputerze a oprogramowaniem.
- ✓ Podstawowe interfejsy API dla grafiki 3D to OpenGL firmy SGI i Direct3D wchodzący w skład pakietu DirectX firmy Microsoft.,.
- ✓ Warstwę oprogramowania bliżej związaną z konkretną kartą są **sterowniki**. Głównym zadaniem sterowników jest zamiana instrukcji i danych płynących z interfejsu API na informacje zrozumiałe bezpośrednio dla danego urządzenia.

DirectX



Mapowanie środowiska (DirectX 7)



Mapowanie środowiska (DirectX 9)



Microsoft
Flight Simulator X

Mapowanie środowiska (DirectX 10)



Microsoft
Flight Simulator X

DirectX 11-12



DirectX 11



DirectX 12

Obliczenia „fizyki”

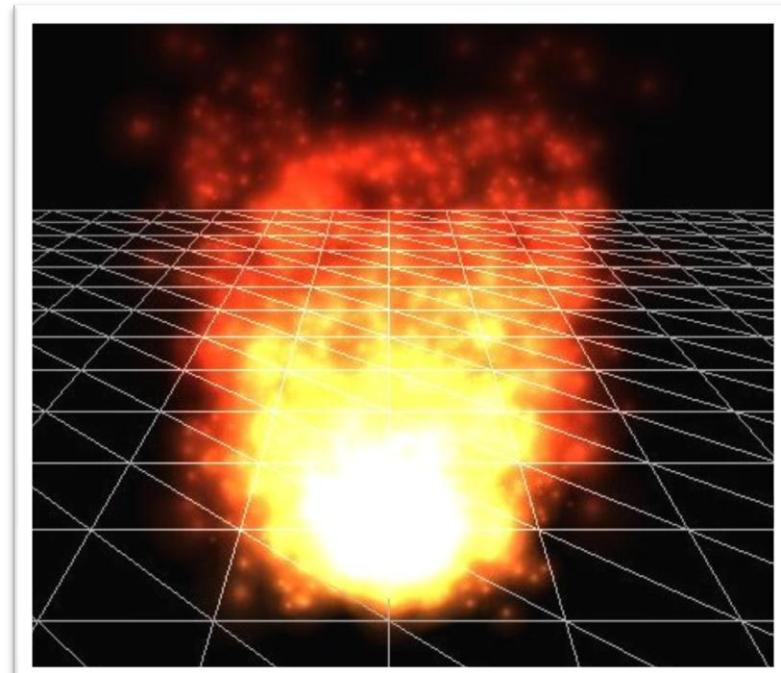
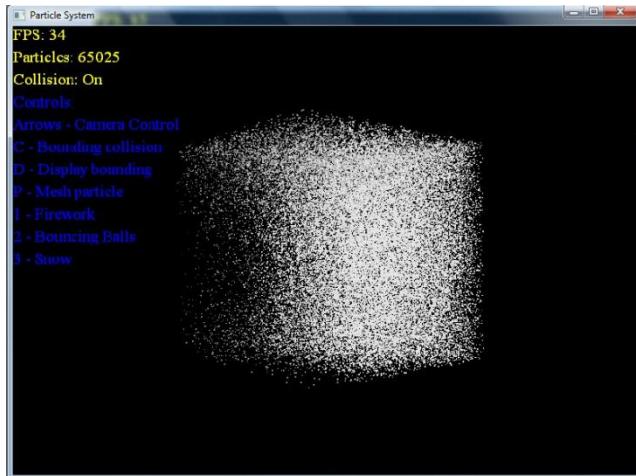
Ogień, dym jest dużym wyzwaniem dla obecnych kart graficznych i CPU. Karta graficzna (np. technologia „Nvidia Quantum Effects”) zwalnia procesor z tego zadania. Dzięki niej wszystkie efekty fizyczne są renderowane przez GPU.



Do tego rodzaju efektów wykorzystywane są (wsparcie sprzętowe poczawszy od DX11) obliczenia fizyczne - **efekty cząsteczkowe**

Obliczenia „fizyki”

Efekty cząsteczkowe - symulacje złożonych zjawisk (takich jak opady śniegu, czy deszczu, dym, pył, płomień), w których podstawowym obiektem jest duża grupa wirtualnych cząsteczek, traktowanych jak obiekty punktowe. Cząsteczki mają zwykle ograniczony czas istnienia, podlegają interakcji z otoczeniem, tj. odbijają się od przeszkód, ulegają również wpływom sił zewnętrznych (np. grawitacji, czy sile wiatru).



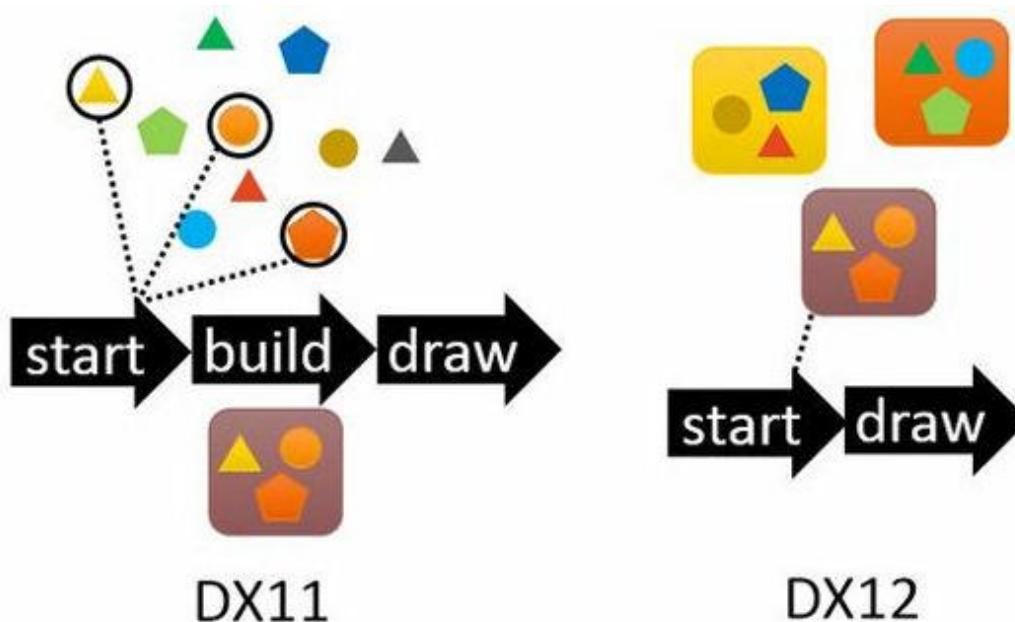
Źródło: wikipedia

DirectX 12

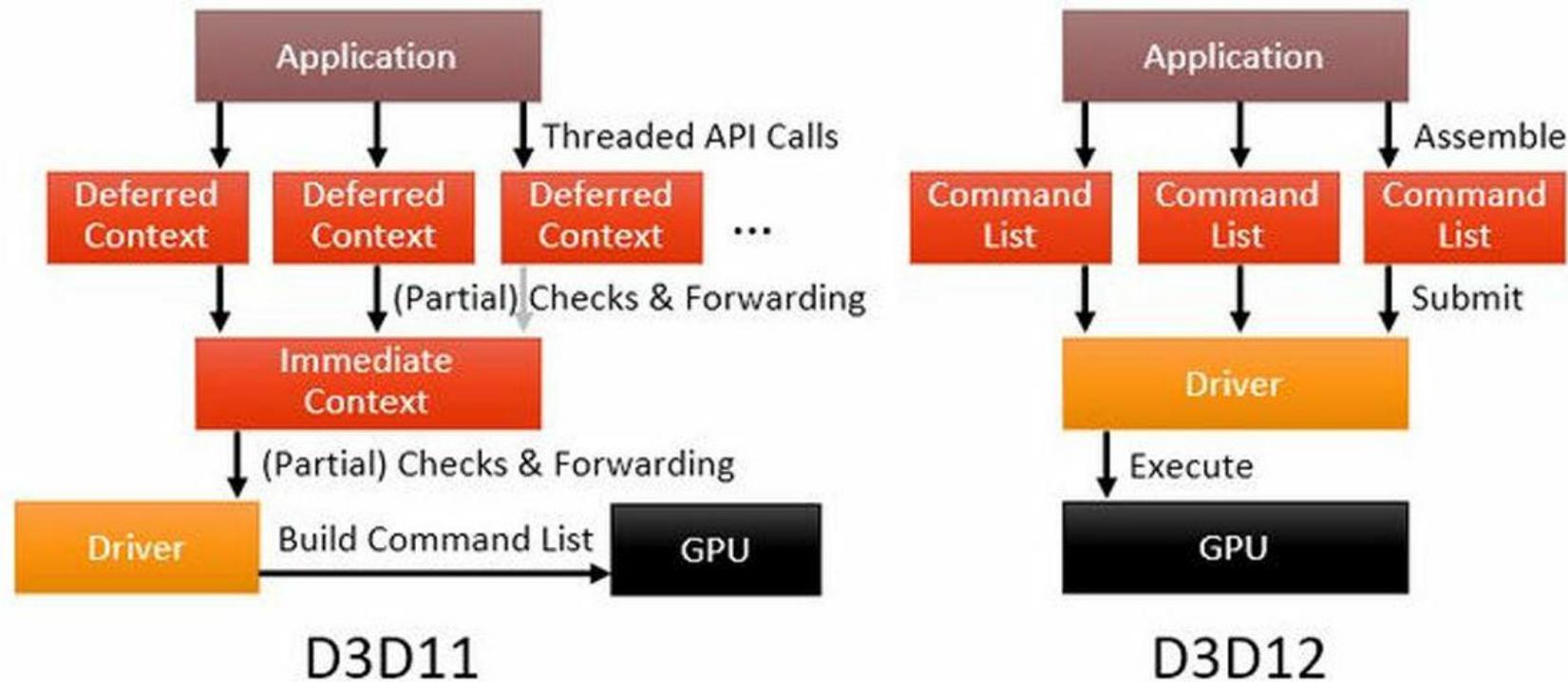
Największą zaletą DirectX 12 jest to, że ma dostęp do sprzętu na niższym poziomie niż poprzednia wersja.

Mechanizm, w którym DirectX 12 przechowuje informacje podczas wysyłania instrukcji do GPU, nazywany jest listą poleceń, a GPU odbierający to listę poleceń renderuje wynikowy obraz.

W DirectX 11 i wcześniejszych, konieczne było przejście przez wiele etapów przed stworzeniem listy poleceń.



DirectX 12



Złącza kart graficznych



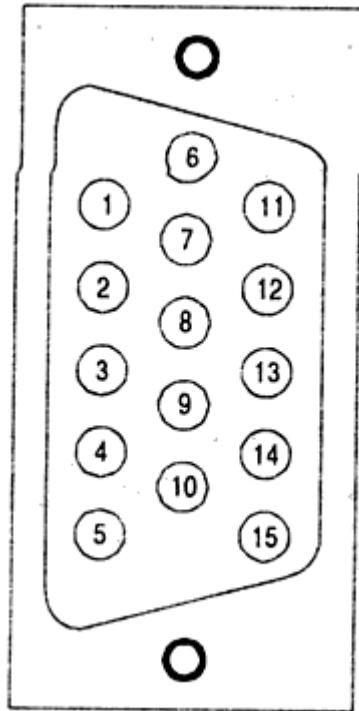
Złącze HDMI
(High Definition
Multimedia Interface)

Złącze analogowe
DSUB (15 pinów)

Złącze cyfrowe DVI
(Digital Video Interface)



Złącza kart graficznych - CRT



1 = RED 6 = GND-RED
2 = GREEN 7 = GND-GREEN
3 = BLUE 8 = GND-BLUE

4, 5, 11 = GND
9 = +5V

13 = SYNC-H 14 = SYNC-V
10 = GND-SYNC-H/V

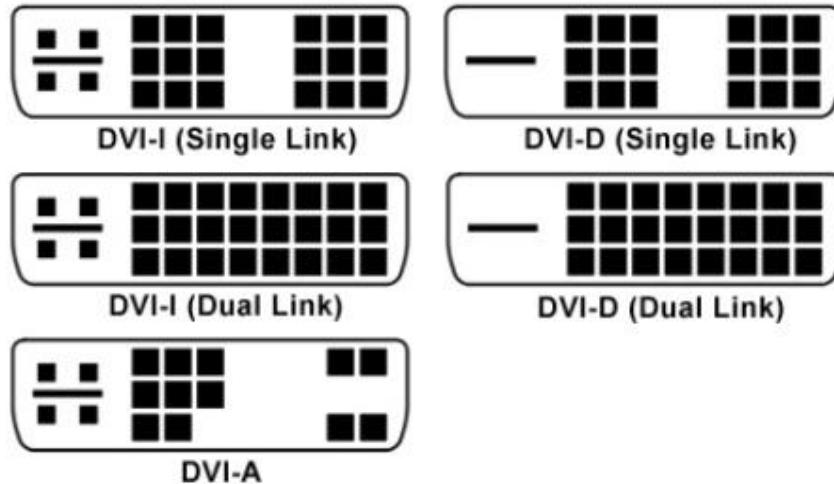
12 = DDC DATA
15 = DDC CLK

D-SUB (CRT)

Złącze popularne w czasach monitorów CRT - analogowe.

Niepożądane w monitorze LCD, ponieważ zachodzi wówczas dwukrotna konwersja sygnału: zamiana sygnału cyfrowego na analogowy w karcie graficznej oraz zamiana sygnału analogowego na cyfrowy w monitorze. Podwójna konwersja powoduje straty w jakości obrazu i zwiększa podatność na zakłócenia.

Złącza kart graficznych - DVI



DVI (ang. Digital Video Interface) - cyfrowy standard przesyłania sygnału wideo umożliwia także przesyłanie sygnału analogowego.

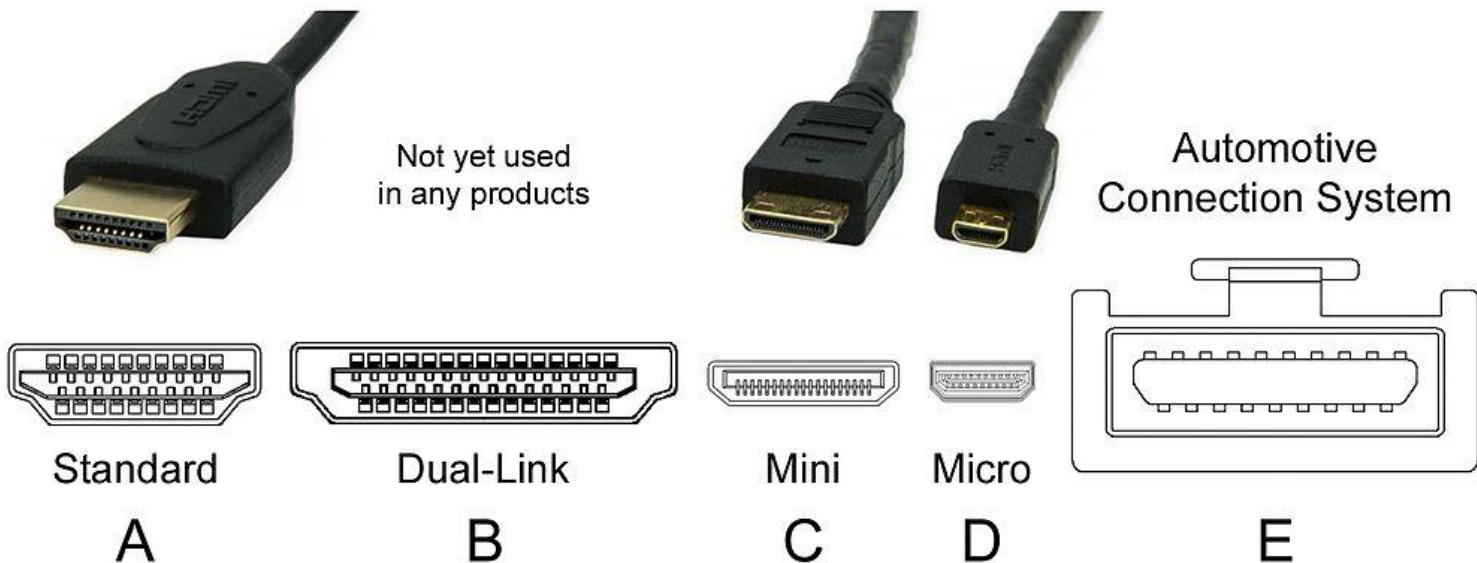
Rodzaje złącza DVI:

- DVI-D - obraz prawdziwie cyfrowy, przesyłany między cyfrowym wyjściem karty graficznej a cyfrowym wejściem monitora
- DVI-A - używany do podłączenia nowoczesnej karty graficznej z wyjściem DVI do monitora z wejściem analogowym
- DVI-I - zapewnia możliwość przesyłu sygnału analogowego źródła do analogowego monitora jak również cyfrowego sygnału źródła do cyfrowego monitora

Złącza kart graficznych - HDMI

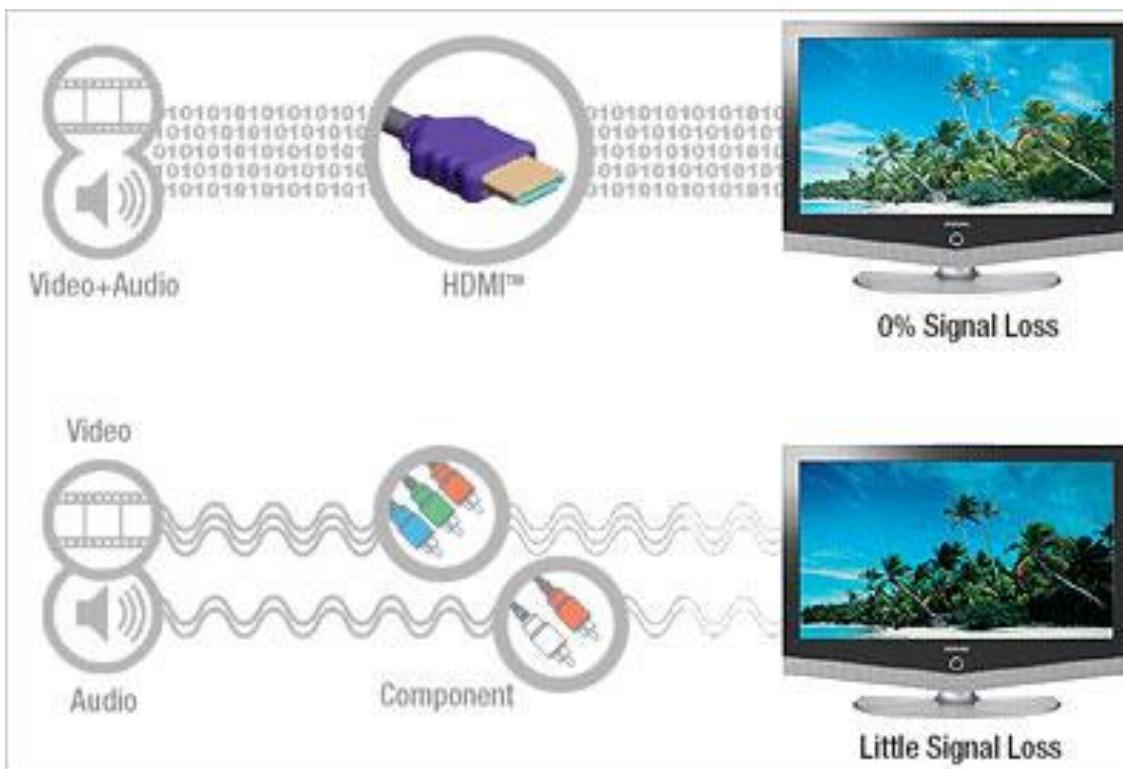
High Definition Multimedia Interface (HDMI) cyfrowy interfejs dla sygnału audio/wideo. Można przesyłać pełen strumień danych bez kompresji – obraz wysokiej rozdzielczości i dźwięk wielokanałowy.

Dane wideo przesyłane są z wykorzystaniem technologii TMDS.



Źródło: Wikipedia.en

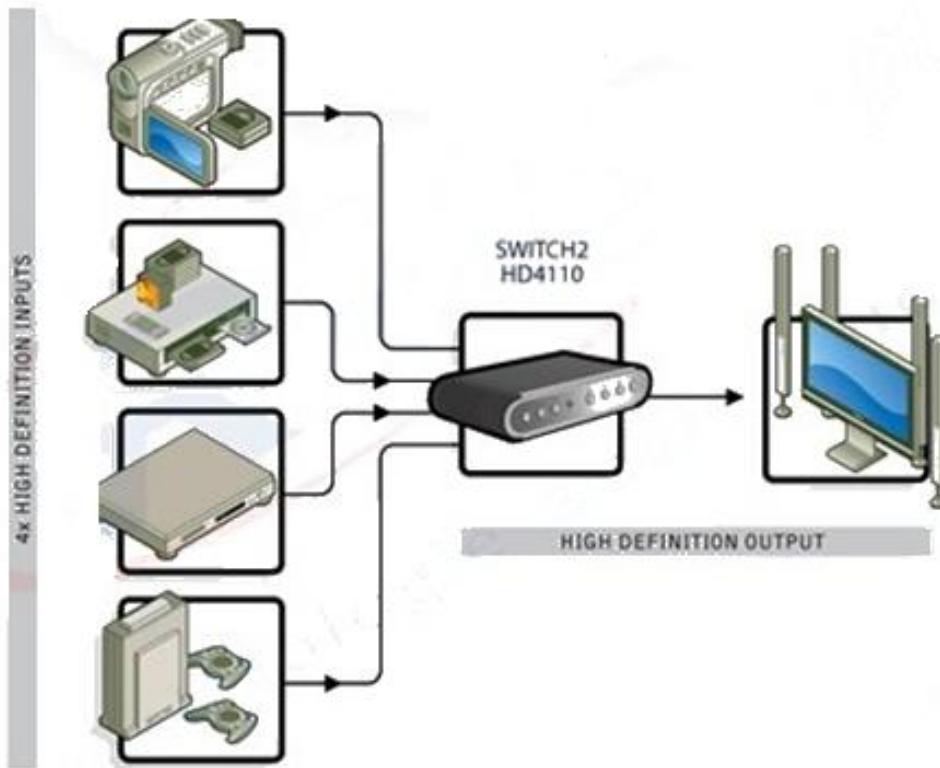
Złącza kart graficznych - HDMI



- **High Definition Multimedia Interface (HDMI)** jest cyfrowym interfejsem dla sygnału audio/wideo zdolnym przesyłać pełen strumień danych bez kompresji.
- Dane wideo przesyłane są z wykorzystaniem technologii TMDS.

Złącza kart graficznych - HDMI

- HDMI pozwala łączyć ze sobą dowolne, zgodne ze standardem, urządzenia audio/video takie jak odtwarzacze DVD, konsole gier, komputery z monitorem lub telewizorem cyfrowym.
- Obecnie zaczyna wypierać starsze standardy zarówno analogowe (np.: S-Video, SCART, VGA, DVI-A) jak i cyfrowe.



Złącza kart graficznych

DisplayPort - Uniwersalny interfejs

Głównym zamierzeniem nowego standardu jest połączenie komputer-monitor lub komputer-system kina domowego (w tym np. projektor, telewizory). DisplayPort już w swojej pierwszej wersji 1.0 osiągnął to, co HDMI udało się osiągnąć dopiero w wersji 1.3b czyli: maksymalna rozdzielcość 2560x1600, 24 bitowa głębia kolorów, maksymalna przepustowość sygnału w granicach około 10 Gb/s. Pełną specyfikację jesteśmy w stanie osiągnąć na kablu o długości 15 metrów, przy dłuższych kablach rozdzielcość zostaje zmniejszona do 1920x1080.



DisplayPort



Mini DisplayPort
/ Thunderbolt



Literatura:

- Metzger Piotr - Anatomia PC, wydanie XI, Helion 2007
- Wojtuszkiewicz Krzysztof - Urządzenia techniki komputerowej, część I: Jak działa komputer, MIKOM, Warszawa 2000
- Wojtuszkiewicz Krzysztof - Urządzenia techniki komputerowej, część II: Urządzenia peryferyjne i interfejsy, MIKOM, Warszawa 2000
- Mirosław Zelent. <http://miroslawzelent.pl/informatyka/karty-graficzne-grafika-3d/>
- Źródło: <https://pclab.pl/art23133-2.html>