

Titan X: la sorpresa di primavera da Nvidia

Con 8 miliardi di transistor, Big Maxwell è la Gpu più potente e più grande di sempre per il settore consumer e in modo particolare per i videogiochi.

Di **Michele Braga**

Con l'introduzione della GeForce GTX 980 in molti hanno pensato che Nvidia avrebbe aspettato le mosse della rivale Amd prima di rilanciarsi sul mercato con un nuovo prodotto. Invece, durante il GDC 2015, Jen-Hsun Huang – Ceo di Nvidia – ha estratto l'asso dalla manica e sorpreso tutti con l'annuncio di una nuova ammiraglia per la famiglia GeForce GTX: la Titan X.

Come ci si aspetta da un prodotto che si proclama primo della classe, i numeri della Titan X impressionano, soprattutto quando li si mette a confronto con quelli del precedente modello top di gamma in casa Nvidia. La Gpu GM200 conta 8 miliardi di transistor, prodotti con tecnologia Tsmc a 28 nanometri e stipati in 601 millimetri quadrati di silicio. Si tratta del processore grafico per il settore consumer e videoludico con il maggior numero di transistor prodotto sino a

oggi; inoltre Nvidia ha sfruttato quasi fino al limite le capacità del processo produttivo Tsmc che permette di produrre chip a 28 nanometri con una superficie massima di 630 millimetri quadrati.

Per questo motivo il GM200 è definito dalla stessa Nvidia come "Big Maxwell" è verosimilmente sarà il più grande e l'ultima implementazione di

questa architettura fino all'introduzione di un processo produttivo più avanzato da parte di Tsmc o all'introduzione della prossima architettura Nvidia.

Come analizzeremo tra breve, il GM200 integra il 50% in più di risorse hardware e di calcolo rispetto al processore grafico GM204-400 alla base delle schede grafiche GeForce GTX 980: 3.072 Cuda Core contro 2.048, rispettivamente 192 e 96 unità di texture e unità Rop contro 128 e 64, un controller di memoria a 6 canali e 384 bit di ampiezza di banda contro i 4 canali e 256 bit. Insomma, la Titan X aggiunge molta potenza di calcolo che si traduce in prestazioni maggiori, ma soprattutto nella capacità di poter gestire qualunque tipo di gioco a risoluzione 4K e in modalità multi fotogramma, sia che si tratti di multi monitor, tecnologie stereoscopiche o visori per la realtà virtuale.

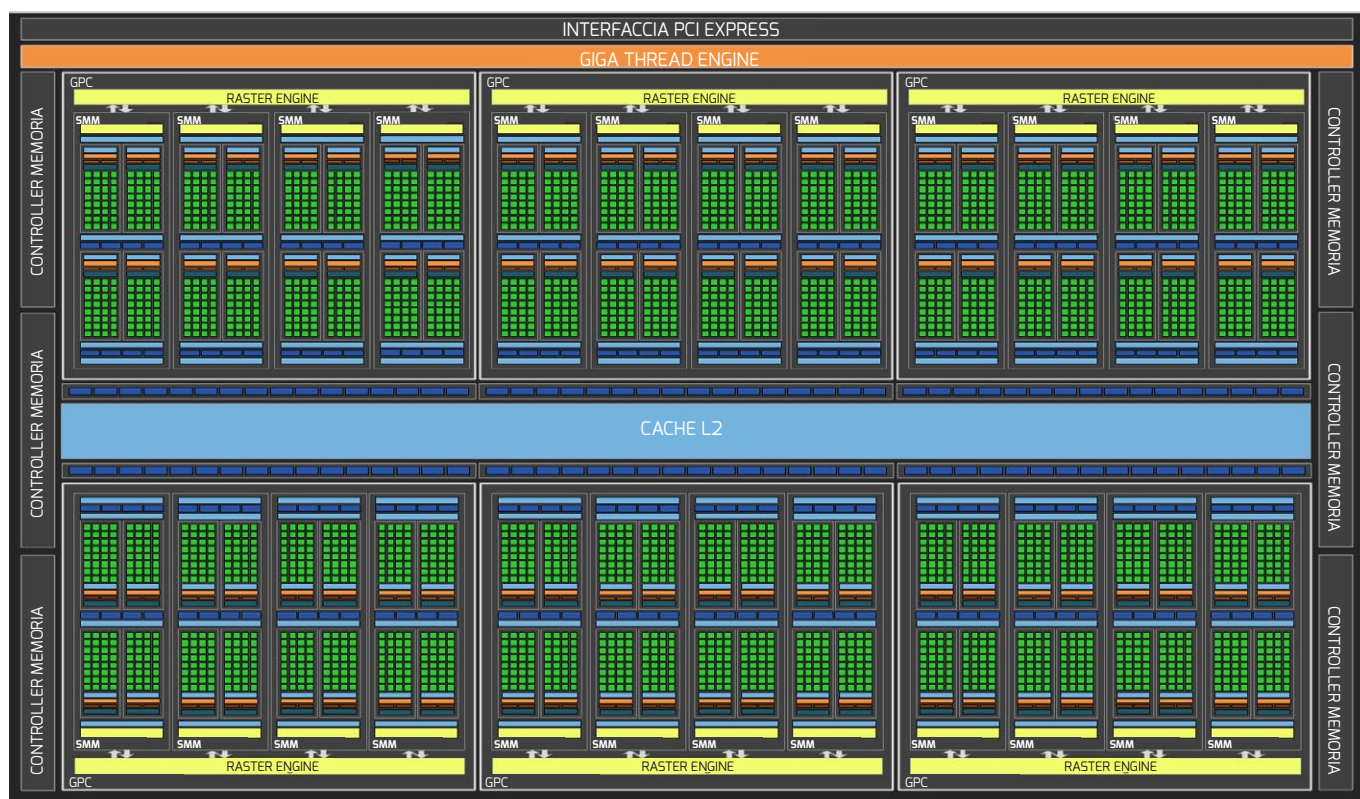
“

La Titan X dispone del 50% in più di hardware rispetto alla precedente top di gamma Nvidia GeForce GTX 980.

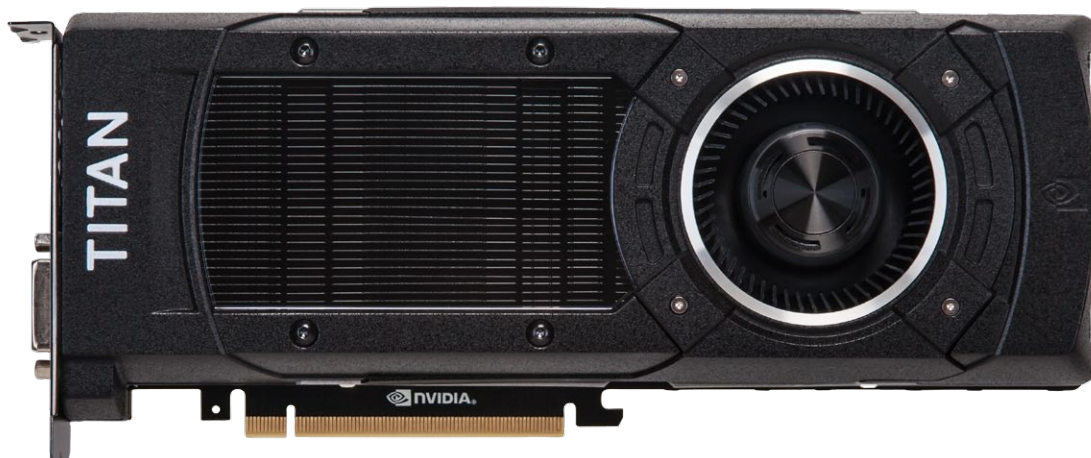
CARATTERISTICHE TECNICHE

MODELLO	GEFORCE GTX 960	GEFORCE GTX 970	GEFORCE GTX 980	GEFORCE GTX TITAN X	GEFORCE GTX TITAN Z	GEFORCE GTX TITAN BLACK	GEFORCE GTX TITAN
Gpu	GM206-300	GM204-200	GM204-400	GM200	2 x GK110	GK110	GK110
Dimensione die (mm²)	227	398	398	601	561	561	521
Numero di transistor (milioni)	2.940	5.200	5.200	8.000	14.200 (2x 7.100)	7.100	7.100
Tecnologia produttiva (nm)	28	28	28	28	28	28	28
Frequenza operativa (MHz)	1.126	1.050	1.126	1.000	705	889	837
Frequenza Gpu Boost (MHz)	1.178	1.178	1.216	1.075	876	980	876
Blocchi Gpc	2	4	4	6	10 (2x 5)	5	5
Moduli SMM / SMX	8	13	16	24	30 (2x 15)	15	14
Cuda Core (singola precisione)	1.024	1.664	2.048	3.072	5.760 (2x 2.880)	2.880	2.688
Cuda Core (doppia precisione)	32	52	64	96	1.920 (2x 960)	960	896
Unità di texture	64	104	128	192	480 (2x 240)	240	224
Unità Rop	32	64	64	96	96 (2x 48)	48	48
Memoria cache L2 (Mbyte)	1	2	2	3	3 (2x 1,5)	1,5	1,5
Supporto DirectX	12	12	12	12	11.1	11.1	11.1
Supporto OpenCL	1.2	1.2	1.2	1.2	1.3	1.3	1.3
Supporto OpenGL	4.5	4.5	4.5	4.5	4.4	4.4	4.4
Frequenza memoria (MHz)	7.010	7.000	7.000	7.010	7.000	7.000	6.008
Ampiezza del bus di memoria (bit)	128	256	256	384	768 (2x 384)	384	384
Tipo di memoria	Gddr5	Gddr5	Gddr5	Gddr5	Gddr5	Gddr5	Gddr5
Quantità di memoria (Mbyte)	2.048	4.096	4.096	12.288	12.288	6.144	6.144
Banda di memoria (Gbyte/s)	112,2	224,0	224,0	336,5	672,0 (2x 336,0)	336,0	288,4
Potenza mass. della scheda (watt)	120	145	165	250	375	250	250

LO SCHEMA INTERNO DEL PROCESSORE GRAFICO NVIDIA GM200



L'organizzazione interna della Gpu GM200 prevede sei blocchi Gpc; ciascuno di questi raggruppa otto moduli SMM da 192 Cuda Core ognuno. L'intera Gpu dispone di 3 Mbyte di cache condivisa di secondo livello e di un controller a sei canali con una ampiezza complessiva di 384 bit.



La scheda è prodotta con una copertura in lega di alluminio e magnesio per dare maggiore rigidità e per migliorare la dissipazione termica.

L'ARCHITETTURA

Come abbiamo avuto modo di illustrare in più riprese su queste pagine durante i mesi scorsi, l'architettura Maxwell è composta – in modo molto simile a quelle che l'hanno preceduta – da una batteria di blocchi logici denominati Gpc (*Graphics Processing Cluster*), all'interno dei quali sono organizzate le unità SMM (*Maxwell Streaming Multiprocessor*), affiancati da una struttura di supporto che comprende l'unità Giga Thread Engine, una cache di secondo livello (L2), il controller di memoria e quello Pci Express 3.0.

Per quanto riguarda i blocchi costitutivi, il processore grafico G200 (Titan X) presenta le stesse caratteristiche dei modelli GM204 (GTX 980 e GTX 970) e GM206 (GTX 960). Il primo stadio vero e proprio della Gpu è costituito dal Giga Thread Engine che si occupa di ricevere le istruzioni e di distribuire il carico di lavoro sui blocchi di calcolo Gpc presenti nel processore grafico. Ogni blocco o cluster Gpc contiene un motore di rasterizzazione dedicato che presiede alle fasi di generazione

delle primitive della scena 3D e di ripartizione del carico computazionale sui quattro moduli SMM di propria competenza. In ogni modulo SMM sono presenti un PolyMorph Engine 3.0, un totale di quattro warp scheduler, 128 Cuda Core, 32 unità di Load/Store, 32 unità per funzioni speciali, otto unità di texture, una cache per le istruzioni una cache di primo livello (L1) e 96 Kbyte di memoria condivisa.

Il Polymorph Engine 3.0 – evoluzione della versione 2.0 presente in Kepler – integra le unità che servono all'acquisizione delle informazioni relative ai vertici della struttura geometrica da elaborare e quelle per le operazioni di tessellation; le unità di elaborazione supportano in modo completo le funzioni delle librerie DirectX 11.2 e DirectX 12 che debutteranno in modo ufficiale con il rilascio di Windows 10. Scendendo ulteriormente nella struttura dei moduli SMM incontriamo i warp scheduler e i Cuda Core. Questi sono i mattoni fondamentali dell'architettura in quanto le unità di calcolo elementari servono a costruire il complesso motore adatto sia alla grafica,

sia alle applicazioni Gpgpu proprio attraverso la tecnologia Cuda. I quattro warp scheduler gestiscono in modo indipendente e dedicato un blocco di 32 Cuda Core così come 8 unità di Load/Store e 8 unità per le funzioni speciali. All'interno di un modulo SMM le uniche risorse a essere rimaste condivise sono quindi le unità Cuda Core in doppia precisione (FP64) e le unità di texture. Nel processore GM200 – come per tutte le Gpu Maxwell dedicate al mercato consumer – le unità in doppia precisione sono in rapporto 1:32 con quelle a singola precisione: 4 unità FP64 per modulo SMM moltiplicate per i 24 moduli SMM presenti nel processore GM200 fanno un totale di 96 Cuda core capaci di lavorare in doppia precisione.

Si tratta di un numero esiguo per una Gpu di questa classe, ma è una scelta strategica adottata da Nvidia per evitare che la Titan X vada ad insidiare il mercato dei prodotti professionali Quadro e Tesla; questi ultimi dispongono, infatti, di un maggior numero di unità in doppia precisione.

Per quanto riguarda le unità Rop, ovvero le unità che sovrintendono

EFFICIENZA

La maggiore granularità con la quale sono stati progettati i moduli SMM dell'architettura Maxwell è un elemento fondamentale per ridurre i consumi: se da un lato le risorse di tipo condiviso sono utili quando è presente una mole di lavoro in grado di riempire tutte le unità di calcolo, dall'altro hanno delle controindicazioni; le interconnessioni necessarie al funzionamento di una architettura condivisa richiedono infatti maggiore spazio a livello di silicio. Inoltre, in presenza di carichi di lavoro che non occupano l'intero blocco di calcolo, gran parte delle risorse consumano energia sebbene non siano effettivamente utilizzate. Stando alle informazioni rilasciate da Nvidia, un modulo SMM da 128 Cuda Core con architettura Maxwell è in grado di fornire prestazioni pari a circa il 90% di quelle fornite da un modulo SMX da 192 Cuda Core con architettura Kepler, a fronte di un ingombro di superficie molto minore.

alle ultime fasi di generazione dei pixel, all'interno del processore grafico GM200 ve ne sono 96, con un rapporto di 16 per ogni gruppo Gpc e per ogni canale di memoria. A differenza delle schede grafiche GeForce GTX 980 e GTX 970, infatti, la Titan X dispone di un controller di memoria crossbar con ampiezza complessiva di 384 bit attraverso 6 canali a 64 bit. Si tratta dello stesso tipo di controller impiegato sulle precedenti soluzioni Titan basate sui processori GK110 con architettura Kepler; questa scelta permette di garantire un flusso dati adeguato alla maggiore quantità di Cuda Core

presenti nella Gpu GM200 rispetto a quelle GM204. Sulla scheda sono presenti 24 chip di memoria Hynix H5GC(Q)4H24MFR con capacità di 512 Mbyte ciascuno e per un totale di 12 Gbyte a livello del die.

ILLUMINAZIONE VXGI

La tecnologia Vxgi (*Voxel accelerated Global Illumination*) – concepita e sviluppata dal ricercatore Cyril Crassin – è un metodo per calcolare in modo realistico all'interno di una scena sintetica tridimensionale. Questa tecnologia approccia il problema attraverso una

mappatura dell'illuminazione nella scena attraverso una griglia di elementi tridimensionali chiamati voxel; questi non sono altro che elementi cubici – la loro dimensione può essere scalata in funzione della potenza di calcolo di cui si dispone – che approssimano le geometrie presenti nella scena e ai quali sono assegnati valori di intensità di luce, opacità e riflessione in funzione dei materiali delle porzioni di oggetto che rappresentano.

L'utilizzo della tecnologia Vxgi permette di ottenere risultati fotorealistici nell'illuminazione di scene complesse in tempo reale.

PRESTAZIONE

	GEFORCE GTX TITAN X	GEFORCE GTX 980	GEFORCE GTX 780 TI
Futuremark 3DMark (patch 1.5.884)			
Fire Strike	15.107	11.584	10.250
Fire Strike Extreme	7.740	5.877	4.980
Fire Strike Ultra	4.112	3.042	2.514
Unigine Heaven 4.0 (tessellation Normal)			
<i>No AA / MSAA4X</i>			
1.680 x 1.050	162,4	128,7	95,5
1.920 x 1.080	144,1	112,2	85,2
2.560 x 1.440	89,1	65,4	50,7
3.840 x 2.160	64,2	43,8	25,2
Crysis 3 (impostazioni Very High)			
<i>No AA / MSAA4X</i>			
1.680 x 1.050	113,8	85,6	61,4
1.920 x 1.080	104,2	78,4	54,5
2.560 x 1.440	66,4	49,5	33,8
3.840 x 2.160	43,2	32,8	22,6
Metro Last Light (impostazioni High)			
<i>No AA / MSAA4X</i>			
1.680 x 1.050	158,8	141,9	74,0
1.920 x 1.080	145,1	126,3	114,1
2.560 x 1.440	95,1	80,6	73,9
3.840 x 2.160	51,6	43,8	36,4
Tomb Raider (impostazioni Ultra)			
<i>No AA / MSAA4X</i>			
1.680 x 1.050	196,7	164,5	166,8
1.920 x 1.080	175,2	146,2	150,5
2.560 x 1.440	108,6	90,5	95,7
3.840 x 2.160	80,4	70,5	69,1
Tessmark 0.3.0			
<i>Set 3 / Set 4</i>			
Tessellation level 16	91.629	73.619	60.839
Tessellation level 32	51.531	39.382	36.042
Tessellation level 64	23.261	18.161	15.773
LuxMark 3.0 - Gpu			
Neumann TLM-102 SE	7.169	5.245	4.183
Hotel lobby	1.931	1.607	1.014
Configurazione - Processore: Intel Core i7 5960X; Scheda madre / chipset: Asus X99 Deluxe / Intel X99;			
Memoria: 4 da 4 Gbyte Kingston Ddr4; Disco: OCZ ARC 100 SSD / 240 Gbyte; Sistema operativo: Microsoft Windows 8.1			
Professional 64bit; Driver: Nvidia Forceware 347.88			

DSR E MFAA

L'architettura Maxwell dispone di due opzioni di rendering che permettono di bilanciare la qualità dell'immagine e le prestazioni: si tratta delle tecnologie Dsr (Dynamic Super Resolution) e Mfaa (Multi-Frame Sampled Antialiasing).

La prima è stata studiata per i videogiochi che con la potenza delle attuali Gpu sono eseguiti a un numero elevato di fps (fotogrammi per secondo), ma che offrono una qualità dei dettagli e dell'immagine bassa o scadente. In modalità Dsr la scheda grafica forza il rendering della scena a risoluzione più alta per poi scalare l'immagine alla risoluzione di visualizzazione, ottenendo così una maggiore qualità finale. Questa tecnica può essere applicata quando il videogioco è piuttosto leggero in termini computazionali. La seconda è stata studiata per ottenere maggiori prestazioni con i titoli che presentano un drastico calo di prestazioni quando si attivano i filtri antialiasing. La tecnica Mfaa utilizza due schemi differenti per il calcolo dell'antialiasing a fotogrammi alternati e fonde tutto in un fotogramma mediato. In questo modo, applicando due diversi schemi antialiasing 2X si ottiene il risultato visivo di un approccio 4X, ma con un livello di prestazioni superiore a quello di un 4X puro.

DIRECT VR

Come sottolineato da Jen-Hsun Huang – Ceo di Nvidia – nel corso della presentazione al GDC 2015, le caratteristiche della GeForce Titan X sono state pensate per offrire in una sola scheda grafica la potenza di calcolo necessaria non solo a giocare su un monitor ad altissima risoluzione, ma anche per i visori di realtà aumentata che sono attualmente in sviluppo. Grazie all'elevata potenza di calcolo, all'ampia banda verso i 12 Gbyte di memoria locale, la Titan X è in grado di gestire due fotogrammi in simultanea a risoluzioni molto elevate.

ACCELERAZIONE VIDEO

Il motore Nvenc di ultima generazione – presente in tutte le Gpu di classe Maxwell – offre supporto alla codifica

Hecv (H.265) che dovrebbe rimpiazzare quella H.264 e che vanta un maggior rapporto di compressione a parità di qualità visiva; sebbene al momento la funzione di accelerazione in hardware di questo formato non sia supportata da software proprietario o di terze parti, le schede grafiche Nvidia si

Potenza che costa

Per acquistare la nuova ammiraglia è necessario spendere più di 1.200 euro

presentano già pronte per i contenuti multimediali del prossimo futuro. La potenza di calcolo offerta dal motore di codifica è impiegata all'intero dei driver Nvidia dalla funzione ShadowPlay che permette di codificare, registrare (1440p60 e 4Kp60 con bitrate fino a 130 Mbps) o inviare in streaming le immagini delle proprie sessioni di gioco.

LA SCHEDA

Per quanto riguarda la costruzione vera e propria, la Titan X ricalca tutte le precedenti top di gamma Nvidia delle ultime generazioni. Il processore grafico è raffreddato per mezzo di un dissipatore in rame con camera di vapore combinato con un radiatore con alettatura in alluminio che occupa lo spazio di due slot di espansione all'interno del telaio. L'aria è mossa attraverso il radiatore per mezzo di una ventola centrifuga che spinge il flusso dall'intero all'esterno del telaio attraverso le griglie presenti sulla staffa posteriore della scheda. Quest'ultima è chiusa all'interno di una scocca realizzata in lega di alluminio e magnesio con lo scopo di fornire maggiore rigidità alla struttura e maggiore capacità di dissipazione termica.

La Titan X consuma fino a un massimo

Sul lato posteriore sono presenti una uscita Dvi, due Displayport e due Hdmi.

di 250 watt e il circuito di alimentazione è composto da sei fasi per la Gpu e da due fasi per la memoria. Grazie all'impiego di condensatori di tipo Poscaps, Nvidia è riuscita a minimizzare il rumore dovuto al circuito di alimentazione quando la scheda è sotto carico.

Nvidia dichiara inoltre che la Titan X – il modello di riferimento – ha un margine del 10% per quanto riguarda le possibilità di overclock.

Senza dubbio la Titan X è tra i desideri di ogni videogiocatore e dispone di potenza e caratteristiche sufficienti anche per eseguire i videogiochi più esigenti del prossimo futuro. L'introduzione e lo sviluppo di titoli DirectX 12 potrebbe inoltre fornire un ulteriore incremento di prestazioni grazie alla maggiore efficienza delle nuove librerie Microsoft, ma bisognerà attendere Windows 10 e soprattutto titoli che sfruttano queste librerie. Il costo su strada della nuova ammiraglia Nvidia supera i 1.200 euro; insomma un prodotto non per tutti, ma in grado di dare grandi soddisfazioni, a patto di avere una configurazione hardware all'altezza della situazione.

NVIDIA GEFORCE GTX TITAN X
Euro 1.250 Iva inclusa

VOTO 9,0

+ PRO

Potenza di calcolo • 12 GB di memoria • Adatta per stereoscopia e visori VR

- CONTRO

Difficile sfruttarne la potenza • Ridotta capacità di calcolo in doppia precisione

Produttore: Nvidia, www.nvidia.it.