

NVIDIA FERMI GRAPHICS PROCESSOR

GTX - 580.

fermi processor has 32768 registers. divided into lanes, each simp thread is limited to By registers.

the simp thread has upto (64 vector registers of 32-but 32 elements)

32 vector register of 32 64-bit elements.

formi has 16 physical SIMD lanes, each containing 2048 registers

DRAM.

contains stack frame, spilling registers, and private variables.

memory.

memory shared memory by SIMD procedsor in GIPU memory host can read and write GIPU memory.

- 2 SIMD thread schedulers, two the truction dispatch units.

	16 SIMD lanes, (SIMD wedth = 32, Chime = 2 cycles)
	16 load - store units, 4 Stownits.
	thus, 2 threads of SIMD instructions are scheduled every
-	2 clock cycles.
-	IEEE 754 bast double prec. fast 2-level caches, 64-614
	addressing and unified address space. AMY
	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
	error correcting codes
Company of	baster context switching
	faster atomic instructions.
	16 sms 32 cores/sm !
	512 cores,
	8 billion transistors
	768 KB shared 12 cache
	6 DRAM channels.
	host interfaces.
	Grega thread scheduler.
4	2 targe warp schedulers.
-	16 KB register files.
-	Shared instruction cache.
-	dota
-	64 KB Shared LI cache. and Shared Cloral) memory.
+	each sm processor ongs got.
-	LD/ST 16 units each, (2 cycle latency)
	4 SFUS (8 cycle latency)
-	Speparate INT FP white in each core

Keplen gk 110 7.1 billion transistors

1 Tflop double precision throughput.

power efficiency, up to 3x performance per watt of fermi.

DYNAMIC PARALLELISM

on results, and controlling the scheduling of the work via definated accelerated hardware paths, all without involving the CPU.

au with soil involving the cro

HYPER-Q

glitch! enables multiple CPU cores to lounch work on a single CrPU simultaneously thereby dramatically chareasing CrPU utilization and significantly reducing CPU calle time.

Cik 110 gpu by allowing 32 simultaneous hardware-

grid-management unit, manages and prioritizes grids

printout.

NVIDIA GIPU direct.

this is a capability that enables apply withing a single computer, or apply in different servers located across the network to directly thange the data conthout heeding to go to apply system memory

it will consume significantly less power and generate much less heat output.

and 6 subst memory controllers. TSMC. - 28 nm

each of the kepler GK-110 SMX units feature 192 Single precision GUDA cores, and each core has bully pripelined floating point and integer arithmetic logic Units, ou DP units, 32 SFU.

full IEEE 754 2008 compliant SP and DP compliant anithmetic - fermi.

instruction dispatch units. allowing a worps to be issued and executed concurrently.

to share data, there are also atomic ops.

NVIDIA PASCAL GP100 GPU

and many more Crpo computing areas.

NV lines, NVIDIAS new high speed, high bandwidth inter-connected for max app. stability.

(His connected to HBM2

fastest, haigh capacity, extremely stacked or DU memory

high bu mem

architectore,

UNIFIED MEMORY AND COMPUTE PREEMPTION

significantly improved programming model.

16 nm FINFET, enables more features, high perf.

and improved power efficiency.

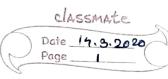
extreme performance for HPC and deep learning.

5.3 TFlops of DP FP performance (FP64).

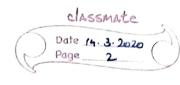
21.2 TELOPS OF HP EP (EPIG) performance,

NV-link extraordinary bandwidth for molti-copu-to-

upto 166 C18 Cs of bidirectional bandwidth.



			Page
	GPU GKIIO CKEPLEN)	Tesla Pioo	
		GPloo (Pascal)	
SMS	15	56	
TPC	(5	28	
FP32 CUDA/S	M 192	64	
FP32 CODA CO		3584	
PET CUDALS	64	32	
FPG4 CUDA CO	red 960	1792	
base clock	745 MHz	1328 MHz	
was clack.	810/875 MHZ	1480 MHZ	
			3000
Peak FP32 G	1FWB 5040	10600	
peak FP64 G	FLORS 1680	5300	
tenture unib		224	
mem interf	are 284-6+6,00R5	4096-bit HBM2.	
mem sze		16 GB.	
L2 cachesize	1536 KB	4096 KB	
Top	235 W	300 W	
transistor	7.1 billion	(5.3 billion	
GAU disc	8-20 551 mm2	6.10 ma	
manufac proce	28 nm	16 nm FINFET	
Company			
Compute apar	3.5 x bermi	6.0	
management	32	82	
manthy	tiproc 64	64	
manthreade	multiproc. 2048	2048	



Compete		
maxtheadslocks.	16	32
max 32- by regs. (SM	658536	6 <i>55</i> 36
max.regl. /block	6.5536	65536
max. rogs. Athread	255	255
max. thread blesize	1024	1024
Shored mean Size (gm	16K/82K/48K	64 K