

NVIDIA GPU コンピューティング エヌビディア ジャパン Tesla Quadro 事業部 マーケティング マネージャー 林 憲一



### NVIDIA について

# ビジュアル コンピューティング テクノロジの世界的リーダー





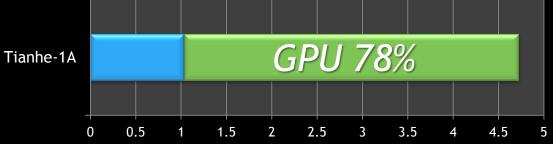
# Top500: 2010年11月リスト

Rank	Site	Computer
1	National Supercomputing Center in Tianjin China	Tianhe-1A - NUDT TH MPP, X5670 2.93Ghz 6C, NVIDIA GPU, FT- 1000 8C NUDT
2	DOE/SC/Oak Ridge National Laboratory United States	Jaguar - Cray XT5-HE Opteron 6-core 2.6 GHz Cray Inc.
3	National Supercomputing Centre in Shenzhen (NSCS) China	Nebulae - Dawning TC3600 Blade, Intel X5650, NVidia Tesla C2050 GPU Dawning
4	GSIC Center, Tokyo Institute of Technology Japan	TSUBAME 2.0 - HP ProLiant SL390s G7 Xeon 6C X5670, Nvidia GPU, Linux/Windows NEC/HP
5	DOE/SC/LBNL/NERSC United States	Hopper - Cray XE6 12-core 2.1 GHz Cray Inc.

### 世界最速システム:中国 Tianhe-1A



- ❖中国国防科学技術大学(NUDT)
- ❖7,168 個の NVIDIA Tesla M2050
- ❖14,336 個の Intel CPU



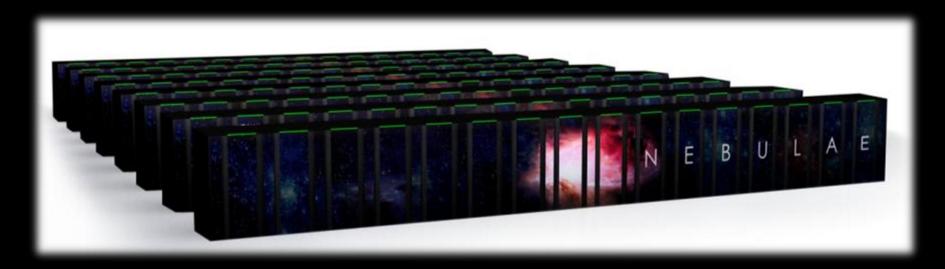
- **❖Linpack 2.507** ペタフロップス
- ❖消費電力 4.04 メガワット
  - ❖ もし同じ性能を CPU だけで実現し ようとすると 12 メガワット必要

### 世界第3位:中国 Nebulae

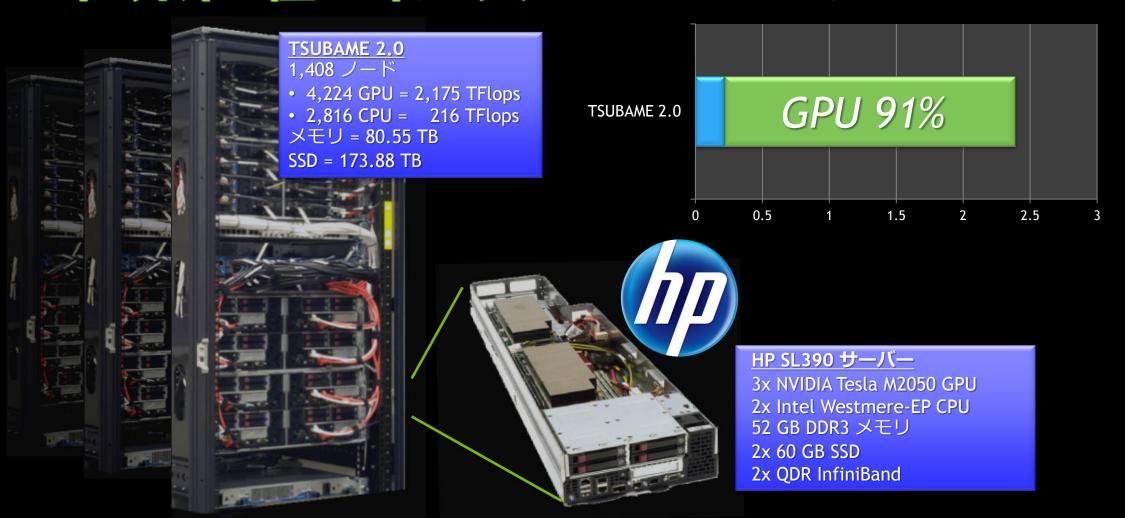
- **❖** Linpack 1.27 ペタフロップス
- ❖ 4,640 個の NVIDIA Tesla C2050 を使用



# CPU ベースの Jaguar に比べて消費電力当り2倍の性能を実現

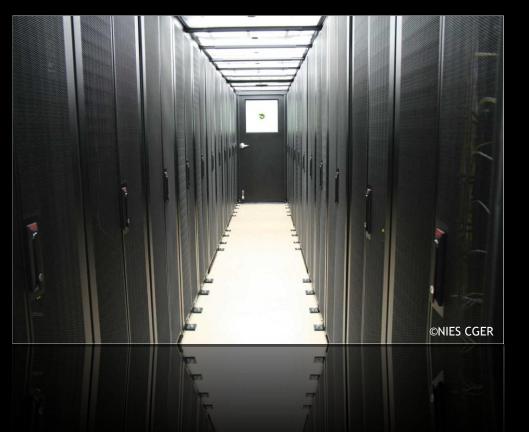


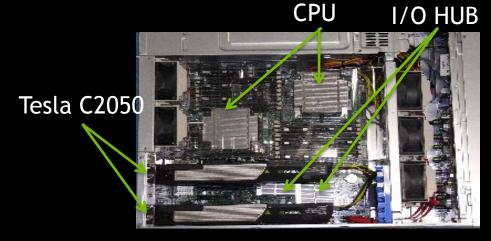
## 世界第4位:東工大 TSUBAME 2.0



### 世界101位:環境研地球環境研究センター







- ●計算ノード: SGI Asterism ID318 x 160 ノード
- ► CPU Intel Xeon E5530 2.4 GHz x 2
- ► GPU NVIDIA Tesla C2050 x 2
- ► SSD 80GB MLC
- ●ディスク: DDN 9000SA, 100 TB
- ▶ ファイルシステム = Lustre (一部 NFS)
- 倍精度浮動小数点演算性能
- ▶ 177 TFLOPS
- ▶ 消費電力 < 170 KVA

### NVIDIA GPU の歴史



1995 NV1 1 Million

**Transistors** 

1999 GeForce 256

22 Million

Transistors

GeForce4
63 Million
Transistors

2002

- 0

GeForce FX 130 Million Transistors

2003

2004

GeForce 6 222 Million Transistors 2005

GeForce 7 **302 Million** Transistors 2006-2007

GeForce 8 **754 Million** Transistors

2008

GeForce GTX 200

**1.4 Billion** Transistors

2010

Fermi
3 Billion
Transistors



DX7 HW T&L 1999 - Test Drive 6 DX8 Pixel Shaders 2001 - Ballistics



DX9 Prog Shaders 2004 - Far Cry



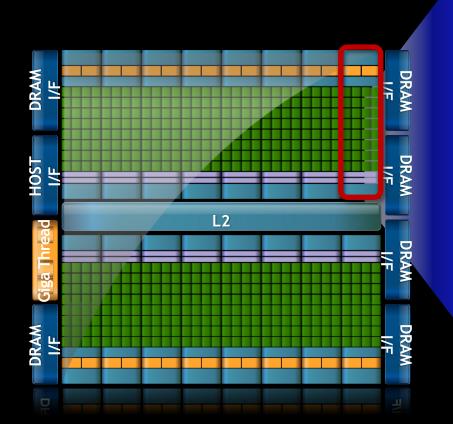
DX10 Geo Shaders 2007 - Crysis

**GPU** 

統合シェーダ + CUDA

### Fermi\* コア アーキテクチャ

\*NVIDIA の最新 GPU アーキテクチャのコードネーム



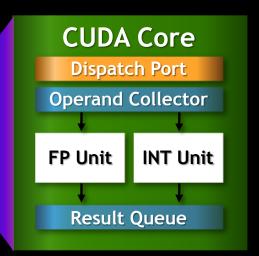
### Scheduler Scheduler Dispatch Dispatch Register File Core Load/Store Units x 16 Special Func Units x 4 Interconnect Network 64K Configurable

Cache/Shared Mem

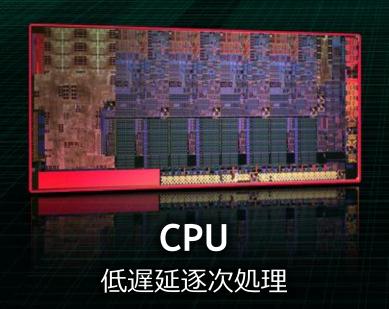
**Uniform Cache** 

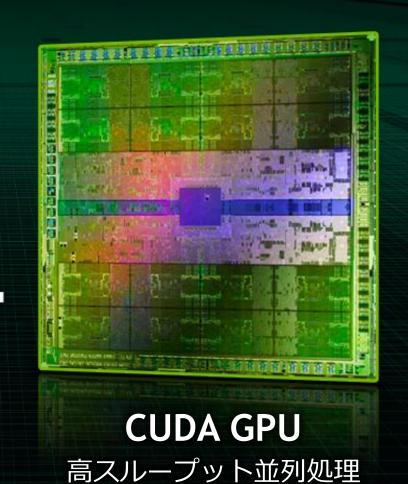
Instruction Cache

### ストリーミング マルチプロセッサ



## GPU コンピューティング時代の幕開け





### NVIDIA Tesla 製品





Tesla S シリーズ S2050 S1070





M2070 M2050 M1060

OEM CPU-GPU 統合 サーバー及びブレード



OEM CPU サーバー + Tesla S シリーズ 1U ラック



ワークステーション 2 ~ 4 Tesla 搭載 パーソナル スーパーコンピュータ

### GPU 搭載サーバーはメインストリームへ



















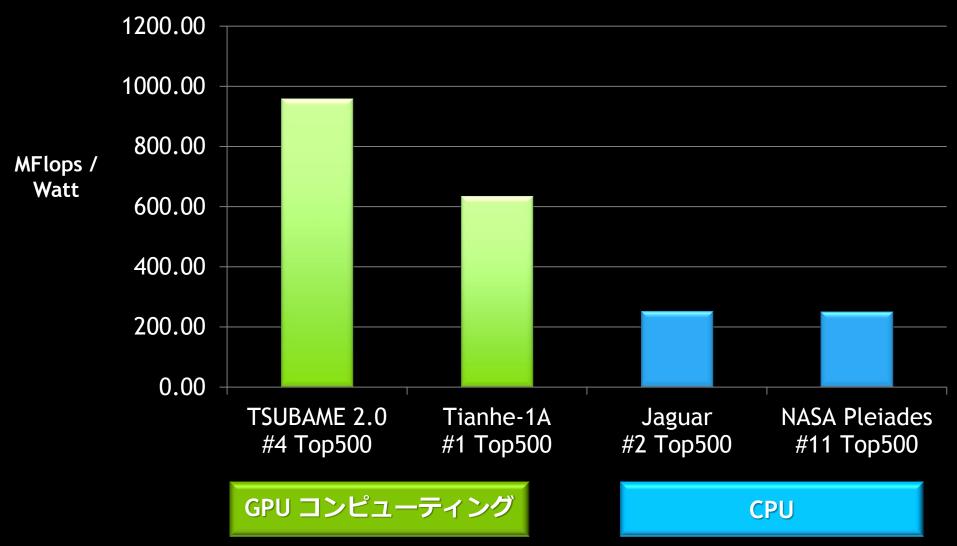








## 圧倒的な消費電力当りの性能



### CUDA C/C++ の継続的なイノベーション

2007 2008 2009 2010

### **CUDA Toolkit 1.x**

- C Compiler
- C Extensions
- Single Precision
- BLAS
- FFT
- SDK w/ 40 samples
- Win XP 64
- Atomics support
- Multi-GPU support

### CUDA Toolkit 2.x

• DP FFT

- Double Precision
- Parallel Nsight (beta)
- Visual Profiler

• cuda-gdb

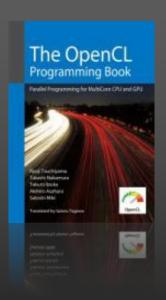
- 16-32 Conversion intrinsics
- Compiler Optimizations
- Performance enhancements
- Vista 32/64
- Mac OSX
- 3D Textures
- HW Interpolation

### **CUDA Toolkit 3.x**

- Fermi arch support
- C++ Class Templates
- C++ Class Inheritance
- Tools updates
- cuda-memcheck
- GPUDirect™
- 16-way concurrency
- Function pointers & recursion

### New in 3.2

- New cuSPARSE Library
- New cuRAND Library (Sobol)
- Support for 6GB Tesla & Quadro
- Multi-GPU Debugging
- Math Library Perf Improvements
- Cluster Management Features
- Integrated TCC Mode











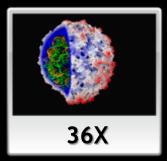




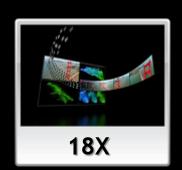


医療画像

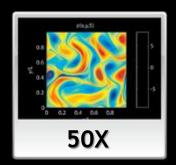
ユタ大学



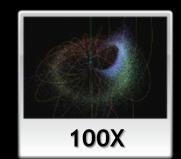




ビデオトランスコード **Elemental Tech** 



MATLAB 演算 **AccelerEyes** 



宇宙物理学 理研



金融シミュレーション オックスフォード



線形計画法 **Universidad Jaime** 



3D 超音波解析 **Techniscan** 



量子化学 イリノイ大学

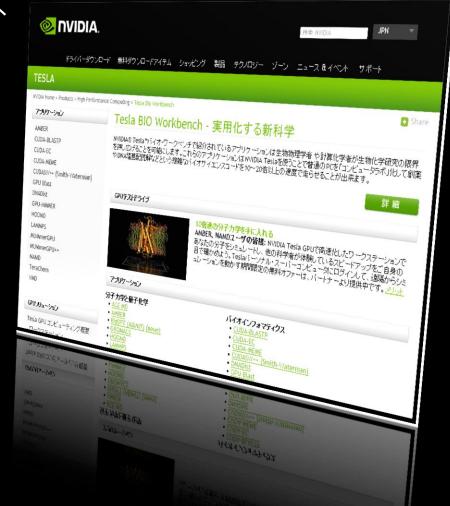


遺伝子配列解析 メリーランド大学

### Tesla BIO Workbench

- ▶分子力学と量子化学
  ▶バイオインフォマティクス
  - > ACE MD
  - > AMBER
  - ➤ BigDFT (ABINIT)
  - ➢ GROMACS
  - > HOOMD
  - > LAMMPS
  - > NAMD
  - > TeraChem
  - > VMD

- CUDA-BLASTP
- > CUDA-EC
- > CUDA-MEME
- ➤ CUDASW++
- > DNADist
- ➤ GPU Blast
- ➤ GPU-HMMER
- HEX Protein Docking
- Jacket (MATLAB Plugin)
- ➤ MUMmerGPU
- ► MUMmerGPU++





**MISYS**®

#1 Engineering Simulation #1 3D DCC

3ds Max

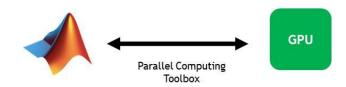


# Mathworks Matlab Parallel Computing Toolbox™ 5.0



MATLAB&SIMULINK

GPU Solution in R2010b シームレスな GPU へのアクセス



>> GX = gpuArray(X); % GPUメモリに送信
>> GY = fft2(GX); % fft2をGPU上で実行

>> Y = gather(GY); % 結果をメインメモリへ回収

MATLAB&SIMULINK

### GPU 対応関数

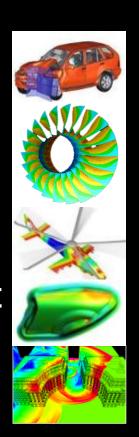
123の関数が GPU 上で計算が可能

abs	complex	gamma	log10	reshape
acos	coni	gammaln	log1p	round
acosh	cos	gather	log2	sec
acot	cosh	ge	logical	sech
acoth	cot	gt	1t	sign
acsc	coth	hypot	1u	sin
acsch	csc	ifft	max	single
all	csch	ifft2	min	sinh
any	ctranspose	imag	minus	size
arrayfun	diag	int16	mldivide	sqrt
asec	disp	int32	mod	sum
asech	display	int64	mrdivide	tan
asin	dot	int8	mtimes	tanh
asinh	double	isempty	ndims	times
atan	eg	isequal	ne	transpos
atan2	erf	isequalwithequalnans	numel	tril
atanh	erfc	isfinite	plus	triu
bitand	erfcinv	isinf	power	uint16
bitcmp	erfinv	islogical	prod	uint32
bitor	exp	isnan	rdivide	uint64
bitshift	expm1	isreal	real	uint8
bitxor	fft	ldivide	reallog	uminus
cast	fft2	le	realpow	uplus
ceil	fix	length	realsqrt	
classUnderlying	floor	log	rem	

7

### CAE での GPU コンピューティングの価値

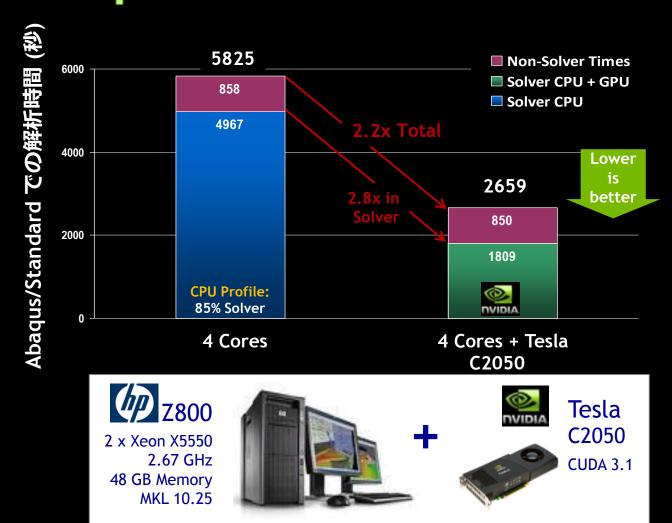
- ❖ 製品品質の向上
  - **❖ シミュレーション回数の増加による品質向上**
- ❖ より速い市場投入
  - ⇒ 高速なシミュレーションによるデザインサイクルの短縮
- ◇ 不可能を可能に
  - **❖ CPU だけではシミュレーションが不可能であった課題が解決可能に**



### CAE アプリケーションの進捗状況

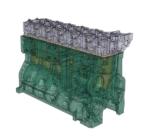
GPU 対応状況	構造解析	流体解析	電磁場解析
利用可能	ANSYS Mechanical IMPETUS AFEA	Acusolve Autodesk Moldflow  FluiDyna Culises (OpenFOAM)  Prometech Particleworks	Agilent EMPro  CST MS  REMCOM XFdtd  S P e a g SEMCAD X
2011年中に リリース予定	LS-DYNA implicit  MSC Software Marc  MSC Software Marc  Abaqus/Standard	Metacomp CFD++	ANSYS HFSS  SAIC Xpatch
製品化評価中	RADIOSS PAM-CRASH implicit MSC Software MD Nastran SIEMENS NX Nastran	CFD-ACE+  GMBRIC: FloEFD  Abaqus/CFD	
研究評価中	LS-DYNA  STREET, Abaqus/Explicit	ANSYS FLUENT/CFX (ANSYS CFD)  ©CD-adapco STAR-CCM+	

# Abaqus での GPU アクセラレーション







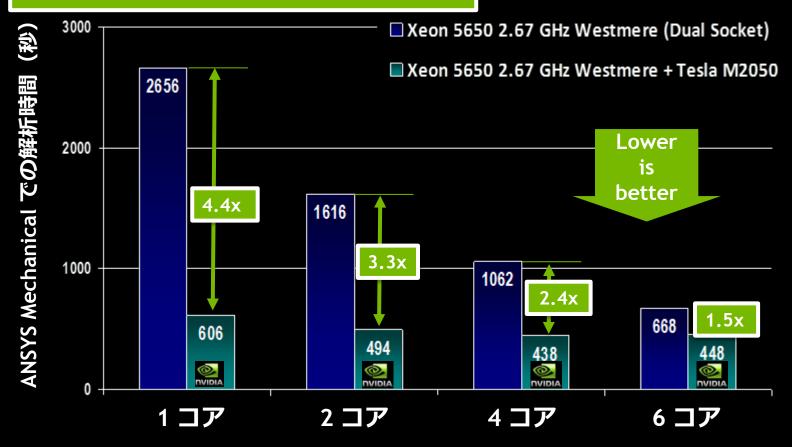


- 150万自由度
- 2 回反復
- 反復当り

5.8e12 Ops

### ANSYS での GPU アクセラレーション

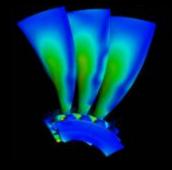
### ANSYS Mechanical R13 スパース直接ソルバー



**NNSYS** 





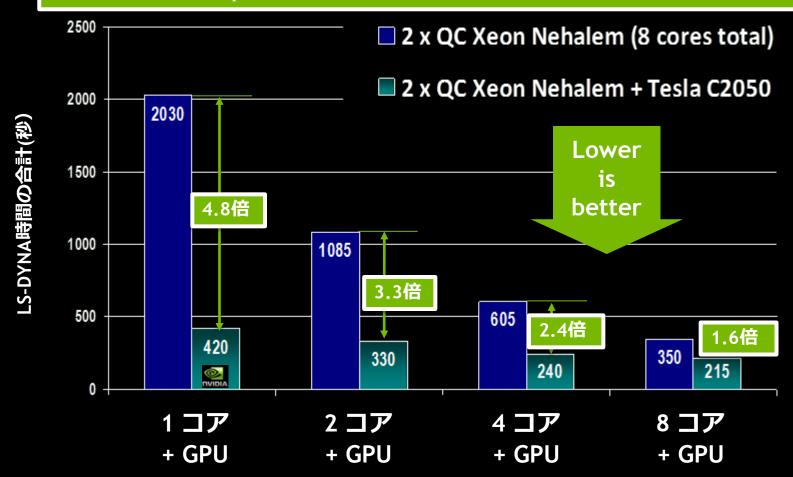


- タービン形状
- 210万自由度
- SOLID187
- 静解析、非線形
- 1荷重ステップ
- スパース直説法

単一の HP-SL390 サーバーノード, 2 x Xeon X5650 2.67GHz CPUs, 48GB メモリ, MKL 10.25, Tesla M2050, CUDA 3.1

### LS-DYNA での GPU アクセラレーション

### 30万自由度の Implicit モデルに対するLS-DYNAの 合計実行時間の比較





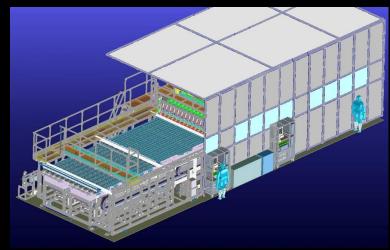
# マシンビジョンにおける GPU の活用







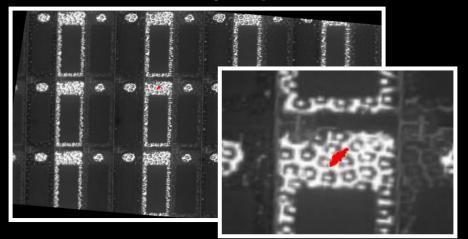




5120pixel

3840
pixel

### 巨大画像の カラー欠陥検出









CPUのみ: 862.9msec GPU活用時: 37.4msec

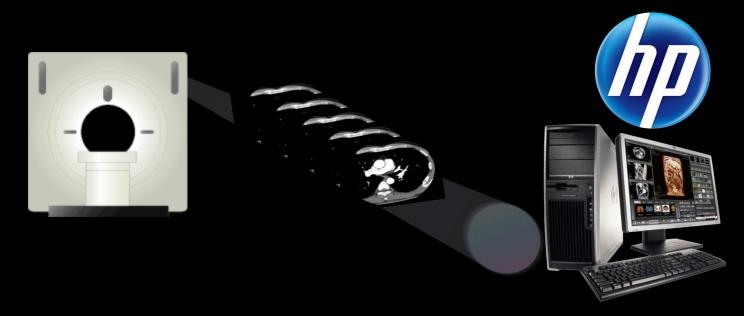
> CPU: Xeon W3860 3.33GHz (12core中2core使用)

GPU:Tesla C2050

### 医用画像処理における GPU の活用



- ❖ 2次元スキャンデータから3次元、4次元イメージの高速生成
- ❖ CUDA 化により画像処理速度を約20倍に高速化

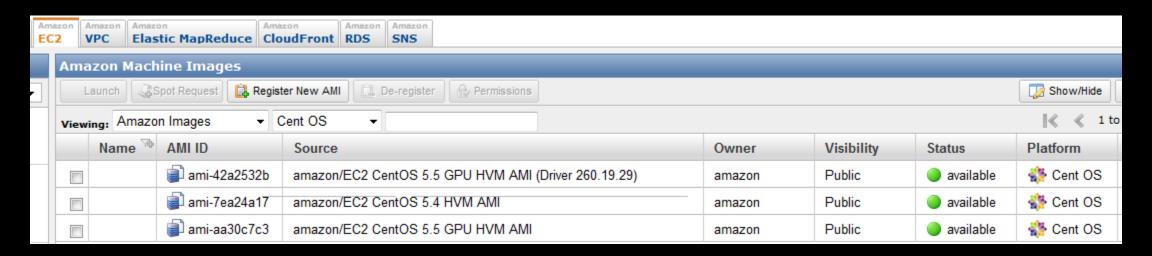




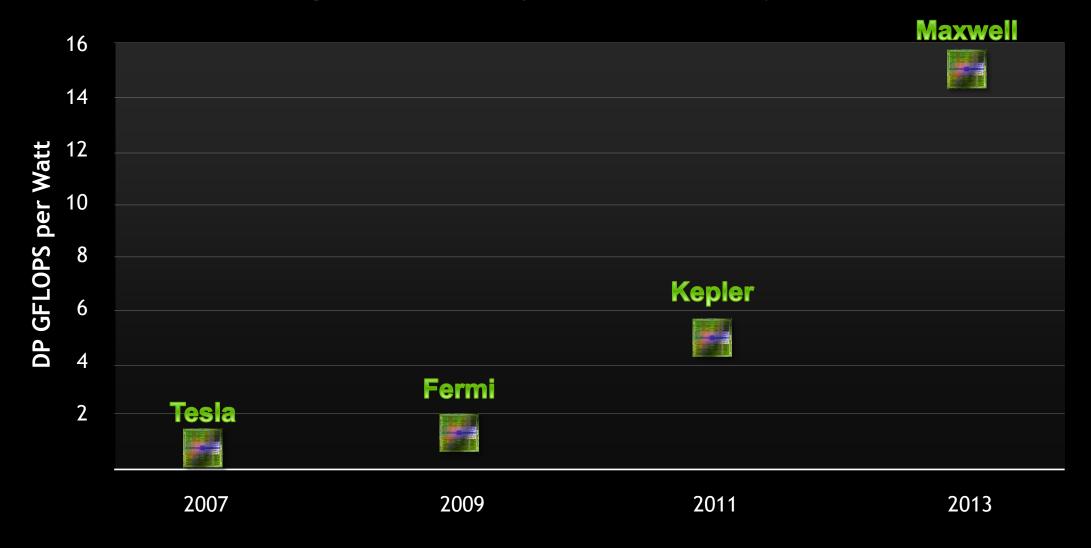
### クラウドでの GPU 利用



- ❖ Amazon Web Services で Tesla M2050 を提供
- ❖ 数分のセットアップで GPU 利用可能



# GPU アーキテクチャ ロードマップ



# Echelon

NVIDIA における超スケールコンピューティング DARPA プロジェクト

Fermi の 100倍の性能目標

## **GPU Computing Solution Finder**

- GPU Computing に対応した
  - ハードウェア
  - アプリケーション
  - ソリューション

を簡単に検索できるポータル サイト

■ 2011年2月提供開始予定



# NVIDIA Japan オフィシャル Twitter



