# **GPU Acceleration Benefits for Scientific Applications**

Axel Koehler Sr. Solution Architect HPC



# **NVIDIA: Parallel Computing Company**



#### **ARM SoCs: Tegra**







## **Continued Demand for Compute Power**



Comprehensive Earth System Model



Simulation of combustion for new high-efficiency, lowemision engines.



Coupled simulation of entire cells



Predictive calculations for supernovae

#### And the Power Crisis in (Super) Computing



#### **Accelerated Computing**

# Add GPUs: Accelerate Applications

**CPUs:** designed to run a few tasks quickly.



**GPUs:** designed to run many tasks *efficiently*.

#### Energy efficient GPU Performance = Throughput

- Fixed function hardware
  - Transistors are primarily devoted to data processing
  - Less leaky cache
- SIMT thread execution
  - Groups of threads formed into warps which always executing same instruction
  - Some threads become inactive when code path diverges
- Cooperative sharing of units with SIMT
  - eg. fetch instruction on behalf of several threads or read memory location and broadcast to several registers
- Lack of speculation reduces overhead
- Minimal Overhead
  - Hardware managed parallel thread execution and handling of divergence



# **Overarching Goals for GPU Computing**







Power Efficiency Ease of Programming And Portability Application Space Coverage



# **Power Efficiency**



# **KEPLER**

SMX Hyper-Q Dynamic Parallelism

(power efficiency)

(programmability and application coverage)

#### Kepler GK110 SMX vs Fermi SM

SM.											
	- 10	-	11.00								
Para Scheduler Web Scheduler											
			Distant Line								
Designer The CALTER & Street											
	1000	10.0	and a	and a second							
÷		i.		SUIT							
			- Care	LUNCT							
-	Care .	-	(Cards)	mat							
				LINE							
Care	Care.	Gara	Cere	LUNC							
				LUIT	-						
Dee	dare.	Carlor,	Certe	A STREET							
		-	-	Labour							
C.e.e	Com.		Cere								
				LINE							
Carrier I.			Care	10.00							
-	-	-	Carte	LUNE							
-				LINE	100						
Date	-	Garren	Cere	LINK							
		La la la		THOMAS IN COLUMN							
40000	1008.4	-	a laster								
-			-	I Carlin							
-			(Dent)								
Tes		Tes.	Tes	100	**						
	-	Tenters	STATUS.								

#### 3x sustained perf/W

Ground up redesign for perf/W 6x the SP FP units 4x the DP FP units Significantly slower FU clocks

Processors are getting wider, not faster

SMX																			
Warp Beheduler			Warp Scherhuler				-	w	191 Dict	with the	- 3		Wayi Bihemiler						
	-	11	Departure	114	10.00	-		(inservit)	uan	1	autors the		Disentation	0.00		eren la		The office	-
								egister	File (	65,536	× 32-4	415							
1	-	÷.	-	1		1	-	1	-	-	-	-	-	-	the second	-	-		-
			_										-				-		
	GUIN			Gara	CORE OF		-	LINET	100	Guite	-	-		car	Citte				-
5.00	ean	6.00	or ten	Gape	Gom	terre .	a the same	LDIEY	SEU.	See	Gom	Gue	Cold Bar	Gere	Garri	C.e.a		LOTT	-
Car	<b>C</b>	-	BP UNI	6.6	Garr	din.	CIP Losse	LDIST	SFU.	Sine	8.ex	4.00	Cill Lines	6-m	6m+	C	Con Million	LOWF	atu
Cure	Gile	<b>C</b>	<b>BP</b> Mili	Cois:	Gain	0111	DP Lost	10.01	SFV.	0.6	Core	<b>C</b> 1110	<b>12</b> 100	Con	Clim	Cim	an tinn	10 H	SFU
Gnite	Cont	<b>6</b> 444	EP-Main	Con	Core	Chre	and local	ubst	sru	See	Con	Com	Per la const	Core	Com	Cost	Die Land	LDIST	sru
Curr	6.m	dire.	-	Circ	<b>C</b> hre	Gim	CP Cost	im	SFU	Curs	6.em	Circ	-	C	cóni	Circ	-	wint	seu
Curr	Cam	-	DP:Det	Gen	Gom	farm.	TP one	LDET	aru	Gore	Con	C.m.	Diff. sales	<b>G</b> 100	Gam	Core	Container of	10.01	-
Carde	cim		DF Unit	Cate	Gara	Corre	TT Law	LINT	aru.	Certe	Core	Circ	Distant.	Cirre	Citra	Cere	Der Mere	LUNIT	-
			-										_				-		
							_						_				-	1000	
Cure	Com	-	0.00	Com	Gone	Const		1534	BPU	344	Cont	Cum-	Carl Colors	Core	Curr	Care		LOSI	380
<b>Fire</b>	Gire	eine.	and block	Gam	Gire	Birt	CTP LINH	6.Detty	SEU	Gotte	Com	Gore	CO- MAR	Ger	Gerre	C.im	an trea	LONT	SFU
Gate	Com	eine		Care	Gare	Gime	218 Sanat	ADD	sru	Sore	Core	Gine	Distant.	Con	Chris	Care	Distant Street	LOW	sru
Circ	Con	Gara	00 1000	Com	6pm	6.00	THE LOCAL	LDBT	SFU	Con	Con	Core	<b>12</b> (14)	Core	60m	Cen	00.000	LOINT	58U
Curt	<b>6</b>	-	mellon	810	Core	6.	CF CHE	LDET	SFU	6	<b>C</b> 278	-	Contraction of	C-110	6.em	6.em	CH LINE	LOST	570
<b>Gene</b>	Gare	d in a	De Linit	Č.	Gare	<u>Gun</u>	In the second	inat	SFÚ	For	Core	Core	D <sup>ar</sup> Line	Cim	Gim	<b>6</b> 111	Der Menn	Abier	BFU
Internet Automatic																			
							-	40.00	area M		IL I CA	-							
	Tex		Ter			Tex		Tex			Тех	T	Tex			Tex	- it	Tex	
	Tex		Tex			Tex		Tex			Tex		Tex			Tex		Tex	

## **Better Utilization with Hyper-Q**





 Grid Management Unit selects most appropriate task from up to 32 hardware queues (CUDA streams)

Improves scheduling of concurrently executed grids

KEPLER 32 Concurrent Work Queues



Particularly interesting for MPI applications when combined with Multi Process Server, but not limited to MPI applications

## Hyper-Q with Multiple MPI Ranks



Hyper-Q with multiple MPI ranks leads to 2.5X speedup over single MPI rank using the GPU

Blog post by Peter Messmer of NVIDIA

http://blogs.nvidia.com/2012/08/unleash-legacy-mpi-codes-with-keplers-hyper-q/

#### **Focus on Power Efficiency**



#### Kayla Development Platform

CUDA 5 | OpenGL 4.3

#### Kick starts ARM + CUDA Ecosystem

#### NAMD Ported in 2 Days



https://developer.nvidia.com/kayla-platform



Quad ARM + Kepler GPU



Quad ARM + Any CUDA GPU



# Ease of Programming and Portability

## **Parallel Computing Platform**



#### **GPU Accelerated Libraries** "Drop-in" Acceleration for your Applications



#### **Dynamic Parallelism** Simpler Code, More General, Higher Performance



#### Launch new kernels from the GPU

- Dynamically based on run-time data
- Simultaneously from multiple threads at once
- Independently each thread can launch a different grid

#### Better load balancing for dynamic workloads

 when work-per-block is data-dependent ( e.g. Adaptive Mesh CFD )



## Kepler Enables Full NVIDIA GPUDirect<sup>™</sup> RDMA



## **NVIDIA GPUDirect™ RDMA**

- True RDMA support for GPU memory
  - NIC performs DMA (GPU DMA engines remain available for CUDA use)
- No hardware changes in NIC
- System BIOS should support Large BARs
- GPU and NIC have to be installed on the same IO Hub (QPI doesn't support it)
- GPUDirect ™ RDMA for communication with other PCI devices (eg. Flash memory devices)
  - Requires adopting GPUDirect-Interop API in vendor software stack
  - Documentation "Developing a Linux Kernel Module using RDMA for GPUDirect" is available at <u>http://docs.nvidia.com/cuda/gpudirect-rdma/index.html</u>
- GPUDirect TM RDMA support available on Tesla and Quadro Kepler class hardware with CUDA 5 and later

# Mellanox Infiniband with GPUDirect<sup>™</sup> RDMA

- Alpha release of Mellanox GPUDirect (GDR) MLNX\_OFED driver is available
  - Alpha release works with CUDA 5.0 or CUDA 5.5
  - Final release will be based on CUDA 6.0 (Beta later in 2013)
  - Supported on any ConnectX adapter that use the MLX4 driver
- MVAPICH2-GDR (based on MVAPICH2 1.9) release can be used with this IB driver release (see later slides)
- Request Mellanox GDR driver and MVAPICH2-GDR via email to hpc@mellanox.com

#### **MPI support for GPUDirect™ RDMA**

- MVAPICH2-GDR (based on MVAPICH2 1.9) supports GPUDirect<sup>™</sup> RDMA
  - Hybrid design takes advantage of GPU Direct <sup>™</sup> RDMA
  - Uses host based buffered design in current MVAPICH2 for reads (Alleviates Sandybridge chipset bottleneck)
- MVAPICH2 team is working on multiple enhancements (collectives, datatypes, one-sided) to exploit the advantages of GPUDirect<sup>™</sup> RDMA
- As Mellanox GDR driver matures, successive versions of MVAPICH2-GDR with enhancements will be made available to the community

### **MVAPICH2** Performance with GPUDirect<sup>™</sup> RDMA

#### Performance of MVAPICH2 with GPU-Direct-RDMA



**GPU-GPU Internode MPI Bi-directional Bandwidth** 

Latency

#### **Bi-Directional Bandwidth**

Performance of MVAPICH2 with GPU-Direct-RDMA

Slides courtesy of DK Panda

### **CUDA Compiler Contributed to Open Source LLVM**

Developers want to build front-ends for Java, Python, R, DSLs

Target other processors like ARM, FPGA, GPUs, x86





# **Enabling More Programming Languages**

## **CUDA Python**





@cuda.jit(restype=uint8, argtypes=[f8, f8, uint32], device=True)
def mandel(x, y, max\_iters):
 zr, zi = 0.0, 0.0
for i in range(max\_iters):
 newzr = (zr\*zr-zi\*zi)+x
 zi = 2\*zr\*zi+y
 zr = newzr
 if (zr\*zr+zi\*zi) >= 4:
 return i
 return 255

```
gimage = np.zeros((1024, 1024), dtype = np.uint8)
d_image = cuda.to_device(gimage)
mandel_kernel[(32,32), (32,32)](d_image, -2.0, 1.0, -1.0, 1.0, 20)
d_image.to_host()
```

#### **Domain-specific Languages**





#### R Statistical Computing Language



#### Liszt

A DSL for solving mesh-based PDEs

# **OpenACC Directives**



Your original Fortran or C code

#### Easy, Open, Powerful

- Simple Compiler hints
- Works on multicore CPUs & many core GPUs
- Compiler Parallelizes code

#### http://www.openacc.org



# **Additions for OpenACC 2.0**

OpenACC. DIRECTIVES FOR ACCELERATORS

- Procedure calls
- Separate compilation
- Nested parallelism
- Device-specific tuning, multiple devices
- Data management features and global data
- Multiple host thread support
- Loop directive additions
- Asynchronous behavior additions
- New API routines for target platforms

(CUDA, OpenCL, Intel Coprocessor Offload Infrastructure)

See http://www.openacc.org/sites/default/files/OpenACC-2.0-draft.pdf





# Application Space Coverage

#### Wide Adoption of Tesla GPUs



#### **Catalog GPU-Accelerated Applications**

#### POPULAR OPU-ACCELERATED APPLICATIONS

Separated Features

Fenerled.

Hulti-GPU Release State

#### http://www.nvidia.com/teslaapps/

Malecular Dynamics		POPULAR GP	U-ACCELERATED APPLICA	TIONS, Continue	al		544										
Abalune	Models molecular dynamics of biopol simulations of proteins, DNA and liga	Application Description Supported Features Espected Multi-GPU Release Status															
ACEMD	Simulation of mechanics force fields, & explicit setwert on CUDA	Weather & Climate	Facecasting	POPULAR GE	PU-ACCELERATED APPLICA	TIONS, Castinue	đ										
AMBER	Suite of programs to simulate molecu dynamics on biamolecules	ASUCA	Weather forecasting model hilly optimized for SPUs	Application	Description	Suppo	rted Feetures Expected H	Aulti-GPU Release									
DL-POLY	Simulate macromotecules, polymers, systems, etc on a distributed memory	CAM/SE	Community Atmospheric Model is a g atmosphere model for weather and c research	Dactuale Design	Automation and CEM	POPULAR OPU-ACCELERATED APPLICATIONS, Cuntinued  Application  Excepted Excepted Excepted Excepted Excepted Excepted											
GROWACS	Simulation of biochemical molecules complicated bord interactions	GE09-5	Weather modeling and ferecasting ap used by NASA	Technologies ADS	and high speed digital circuits	Approxim	a concreta a	DODULI AR CI		NG Continued							
UNDER DEVE		HIRLAM	Weather forecasting model fully optim for GPUs	Technologies	analyzing 30 EM effects of high speed HE/Mcrowave components	Accelerate il TH	Seismic Processing	Association	Bescription	Supported Features	Expected	Melti-GP	U Release Sta				
HOUND-UNIN	ransce dynamics package written gr up for GPUs	HOMME	Weather modeling tool for atmospher scientists	ANS/5 Nacion	Circuit sensition engine for RF/asals	CGG/Veritas RTH	Seamic Processing		The second design		Speed Up*	Support					
NAMD	Designed for high-performance simu	INCOM	Weather forecasting model using icos	CET Meanwork	speedup with GPU computing	Headerane Suite	Seismic Imaging	Chroma	General purpose LOCD application	Wilson-clover fermions, Krylav	5-6x	Yes	Available t				
	of large molacular systems	Hilgon	Numerical model designed for study	Studio (MWS)	Reid senulation	Cesteric Paradigm	Seismic Processing/toterpretation Reservoir Mobiling	MILC	General purpose LOCD application	solvers, Domain-decomposition Staggered fermions, Krylov	15-da	Res	Available n				
Gamester Chamberry GAMESTP US	Computational chemistry suite used t	NIM	Weather forecasting model using icos	COUCH CRO, DAY	computational lithography running of Sauda handware platform	EarthStudy360 Paradigm	Seismic Processing			solvers, Gauge-Enk fattening							
	simulate atomic and molecular electr structure	WIE	Nordonical gnit. Weather and Gozan modeling applica	Remcom XF-did	3D EM modeling and simulation	Eches RTM Paradiem SKUA	Reservoir Modeling	Computational Flu Atom Acufative	id Dynamics General purpose CFD flow spher	Linear opuzion tabler	21	105	Available r				
MWCherts	Computational chemistry package de		-	Rocketsimi	Ventog simulation	Paratism	Saistric Interpretation	Autorienie Malatione	Optimize design of plastic parts and injection molds	Linear equation solver	2x	Single Or	ily Available r				
D-CHEM	for HPC clusters Computational chemistry package de	Adular Francers Dro	Video adming	CAD	30 EM modeling and simulation	Vasellee Schlunderger	Seismic Processing	FEFLO IGNU-Latoert	Namer-Stokes flow solver based on unstructured grids for modeling both	Eighteit solver	10x	Ves	lin Developm				
TeraDiam	for HPC clusters Quantum chemistry software designe	Avid Media Composer	Video editing	CATIA Vo Line Rendering	Photorealistic rendering	Gmega2 RTM	dament free and a	FluiDyna 1 Buttra	Computing physical flows in and around unlift hodies	LBM, particle CFD	20x	Yes	Available r				
	tu run en MADIA GPU	GenArts Suppliere Same Varias Pro	Effects plug-in for video editing	Bunkspeed Pro Suite	Easy to use photorealists; rendering s	Prestack Interpretation	Seemic Processing	FileOyna Culsas OpenFGAM	Computing physical flows with Cullsas — a software library with spocial algorithms for	Linear equation solvers	3x Sober	Single On	ily Available t				
Materiale Science.	Materials code for investigating the et	Animation	Monologi denovita d	HTT DeltaGen 10.s	Photonexistic rendering used for des	Spectra5eis	Seismic Processing / Imaging	Pronotech Particlaworks	Fluid simulation for free surface flow like Tsunami, material processing and liquids	MPS, Particle CFD	48.98	Yes	Available t				
GMCFACK	of temperature on magnetism Solves the many-body Schrödinger ec	Autotesk 30s Mat	3D modeling, animation, and rendering	HTT Columb	Diama di Angela	Stonaridge Reservair	Reservoir Simulation	530 Sanda NL 530	Masswely parallel direct numerical selver IDNSI for the full compressible Navier-Stoke	Chamistry liamel	Bx SP, Sx DP karnet	Yes	In Developm				
WORK PROVIN	for electronic structures using a guan	Autodesk Maya	30 moteling, animation, and	HIT Senarts	TeamCenter and RTT formats	Simulation 5714	-	Turbentream	Utratast CFD solver for turbomochines	Explicit salver	19x.	Tes	Available r				
Quantum-	An integrated suite of computer coder		rendering	Numerical Analysis		Summer Hild	Seidmic Processing	Viratis SpeedIT- OpenEGAM Solve	Set of accelerated solvers for aparse lanear systems of equations	Linear equation solvers	3x Solver	Yes	Available r				
Contraction ( New York, C. ).	modeling at the tunescale			Jacket According Lines	GPU acceleration for MATLAB	Computational Pre		Considerate Discount Manhatry									
WASP	First principles materials code that or electronic structures and guaraum- mechanical molecular dynamics	Advanced Onthe Serves	Geospatial Visualization	Mathematica	Sembolic math analysis	Association MATE AD	Real-time options analytical engine IN	AbaganyStandard	Simulation and analysis tool for structural	Linear equation solver	15-25	Single Or	ly Available r				
		Eterrox Blaze Terra	Orespatul Visualization	Wolfram	Technical computing language and	Mathworks.	PCT, MOCSI Pct, mocsi Relationation (MACS)	ANERS North	Simulation and analysis tool for structural	Linear equation solver	21	Segie De	lly Available r				
Youwikation & Dack	A multifaceted software platform far	ExaMp (117) ENVI	Geospatial Visualization	Mathemata	integrated development environment (MATLAB PCT, NDCS)	Numerical Alexandrea	Random Number Generators	Impetus Alea	Predicts large deformations of structures, and components another to according indiced	Linear equation colver, SPH	10x SPH, 2v Total	Yes	Astatable (				
	visualizing, manipulating, and unders life sciences and bio-medical data	GenEpe Analytics Signature Analysis	Geospatial Visualization			SciComp, Inc	Derivative pricing (SciFinance)	15-00M Implicit	conditions Multiphysics simulation parkage uport	Trace encition toker	31	Vis.	In Developer				
Core Hopping	Rapid screening at novel cores to imp drug preparties	GeoWeb3d Desitors	Geospatial Visualization			Welliam	Mathematical Development Environm	MSC Nastran	Simulation and analysis tool for	Linear equation solver	1.4-21	Yes	Available t				
FastROCS VMD	3D molecular shape comparison Visualizing and analyzing large biolog	Ancogna GIS	Geospatial Visualization			*SPU performance o	emparted against modificante eB6 CPU socket. I	Hart	Simulation and analysis tool for structural mechanics	Linear equation solver	1.52	185	In Developm				
	systems in 3-D graptics	Intergraph Motion Video Analyst	Video fitters and mosaicing Geo-N FMV analytics with intelligence data			nernel is kernel per	formance comparteen. Performance resulte o	RADIOSS Implica	Used to maximize durability, NVH, crash, safety, manufacturability and fluid-structure interaction pactemance	Linear equation solver	24	Single On	ly In Developm				
		Paneptes 1.0	Video Analytica						CONTRACTOR DESIGNATION	1		-					
		and the second s	And the second se														

## **Recent Scientific Breakthroughs using GPUs**

#### Breakthrough in HIV research



Discover the chemical structure of HIV's capsid to build more effective drugs

Run at NCSA Blue Waters (3000 GPUs) Fastest simulation for Silicon for Solar Cells



More efficient & cost-effective solar cells

Gordon Bell Prize Stronger, Lighter Metals



Lighter, Stronger Metals for More Fuel-Efficient Cars

1.87 Petaflop / sec perf on 7168 GPUs on Tianhe-1A,

4224 GPUs at Tokyo Tech, Japan

### **GPUs for control systems**

GPUs are used in many experiments for controlling

- Examples:
- Triggering and tracking for CERN experiments
- Signal processing for Lofar or Square Kilometre Array (SKA)





## **GPUs and Big Data Analytics**

#### salesforce **GPUs** Today **Analyzing Twitter** Computational acceleration for Big Data Visualization ۲ 🕝 sнаzam Searching Audio Accelerating the Cloud + Mobile transformation **GPUs** Tomorrow **Visual Shopping** Converged architecture for Big Data and Compute ELEMENTAL 🕑 **Real-time** Video Delivery

CORTEXICA

ebay

#### Integration of Compute and Visualisation





- GPU Operation Mode "All\_On" enables graphics capabilities for K20/K20X server GPUs
  - nvidia-smi --gom=0

- NVIDIA indeX Scalable Big Data
   Visualization
- Remote visualization tools like
   ParaView



# **The Future**

## DARPA Study Identifies Four Challenges for ExaScale Computing

ExaScale Computing Study: Technology Challenges in Achieving Exascale Systems

Peter Koppe, Editor & Study Lead Keyes Bergman Shekhar Borkar Dan Campbell William Carbon William Dally Monty Denneau Paul Franzon William Harrod Keery Hill Jun Hiller Sheeman Karp Stephen Keckler Dran Klein Robert Locar Mark Richards Al Sevenalli Steven Scott Allan Snavely Thuman Sterling R. Stanley Williams Katherine Yelick

September 28, 2008

The work non-quantered by DARPA DTO is the EnaState Computing Study with Dr. William Harrod on Program Menager. APIE: contrast manufer EAM88-47-C-7734. This report is published in the interest of noisenfile and technical information exchange and its publication does not constitute the Generations' apprecial in this provide of the ideas or Endings.

#### NOTICE

Using Generational develops, specifications, or other data included in this document for any propose other than Generations for momentum document in any may relighter the U.S. Generations. The fast that the Generations throughout the applied the document, specifications, or other data does not limited the labels of any other particle or corporation, or convey suggists or permissions to manufacture, or or still or galacterial corrections the outperformance to faust.

APPROVED FOR PUBLIC RELEASE, DISTRIBUTION UNLIMITED

## Report published September 28, 2008:

#### Four Major Challenges

- Energy and Power challenge
- Memory and Storage challenge
- Concurrency and Locality challenge
- Resiliency challenge
- Number one issue is power
  - Extrapolations of current architectures and technology indicate over 100MW for an Exaflop!
  - Power also constrains what we can put on a chip

#### Which Takes More Energy?

Performing a 64-bit floating-point FMA: 893,500.288914668 × 43.90230564772498 = 39,226,722.78026233027699 + 2.02789331400154 = 39,226,724.80815564 Or moving the three 64-bit operands 20

mm across the die:



This one takes over 4.7x the energy today (40nm)! It's getting worse: in10nm, relative cost will be 17x! Loading the data from off chip takes >> 100x the energy.

## Communication Takes More Energy Than Arithmetic



#### What is important for the future?

- Its not about the FLOPS
- Its about data movements
- Algorithms should be designed to perform more work per unit data movement
- Programming systems should further optimize this data movement
- Architectures should facilitate this by providing an exposed hierarchy and efficient communication

#### 2018 Vision: Compute Node & System



Key architectural features:

- Malleable memory hierarchy
- Hierarchical register files
- Hierarchical thread scheduling
- Place coherency/consistency
- Temporal SIMT & scalarization
- PGAS memory
- HW accelerated queues
- Active messages
- AMOs everywhere
- Collective engines
- Streamlined LOC/TOC interaction

## Conclusion

- Power is the main HPC constraint
  - Vast majority of work *must* be done by cores designed for efficiency
- NVIDIA GPU's are already designed for energy efficiency
- Data movement dominates the power
  - Locality at all levels and reduction of overhead is necessary
- GPU computing has a sustainable model
  - Aligned with technology trends, supported by consumer markets
- GPUs are the path to the tightly-coupled hybrid processor future









# Thank you.

#### Axel Koehler akoehler@nvidia.com

