GPU Computing with Fornax

Dr. Christopher Harris
iVEC@UWA

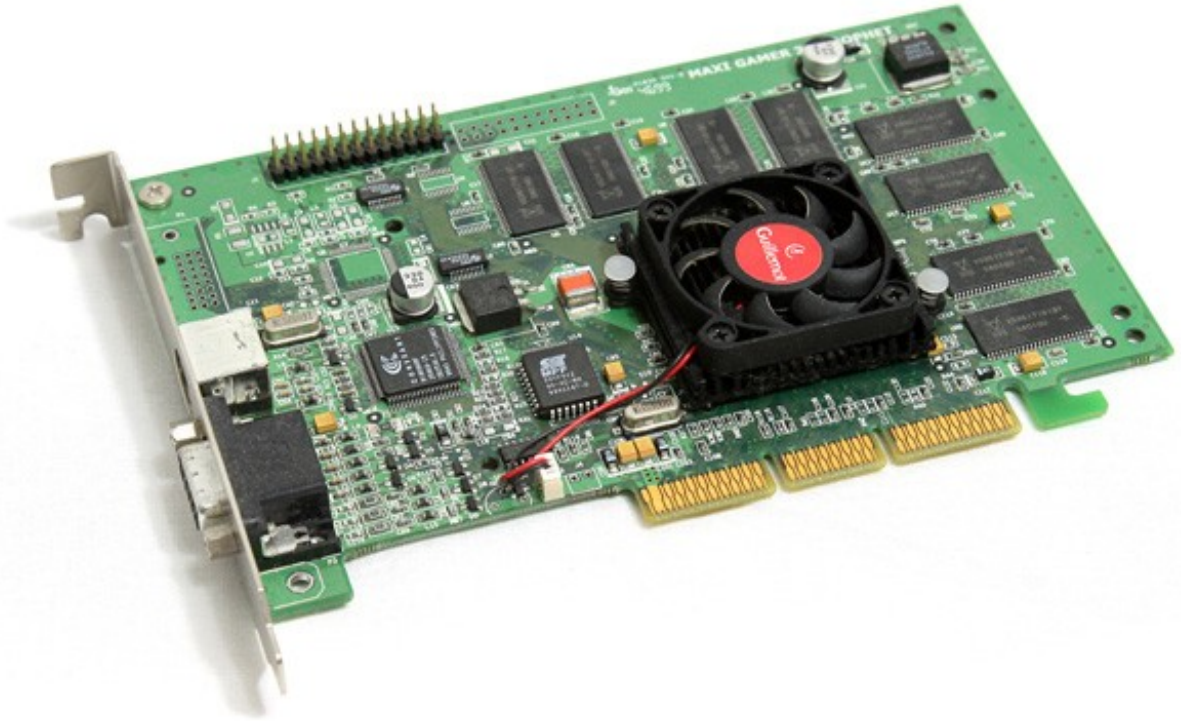


CAASTRO GPU Training Workshop
8-9 October 2012

Introducing the Historical GPU

Graphics Processing Unit (GPU) *n* :

A specialised electronic circuit designed to rapidly manipulate and alter memory in such a way as to accelerate the building of images in a frame buffer intended for output to a display.



Introducing the Modern GPU

Graphics Processing Unit (GPU) *n* :

A general purpose electronic circuit designed to rapidly manipulate and alter memory in such a way as to accelerate computational algorithms that have fine-grained parallelism.





GPU Computing Motivation

But what about...

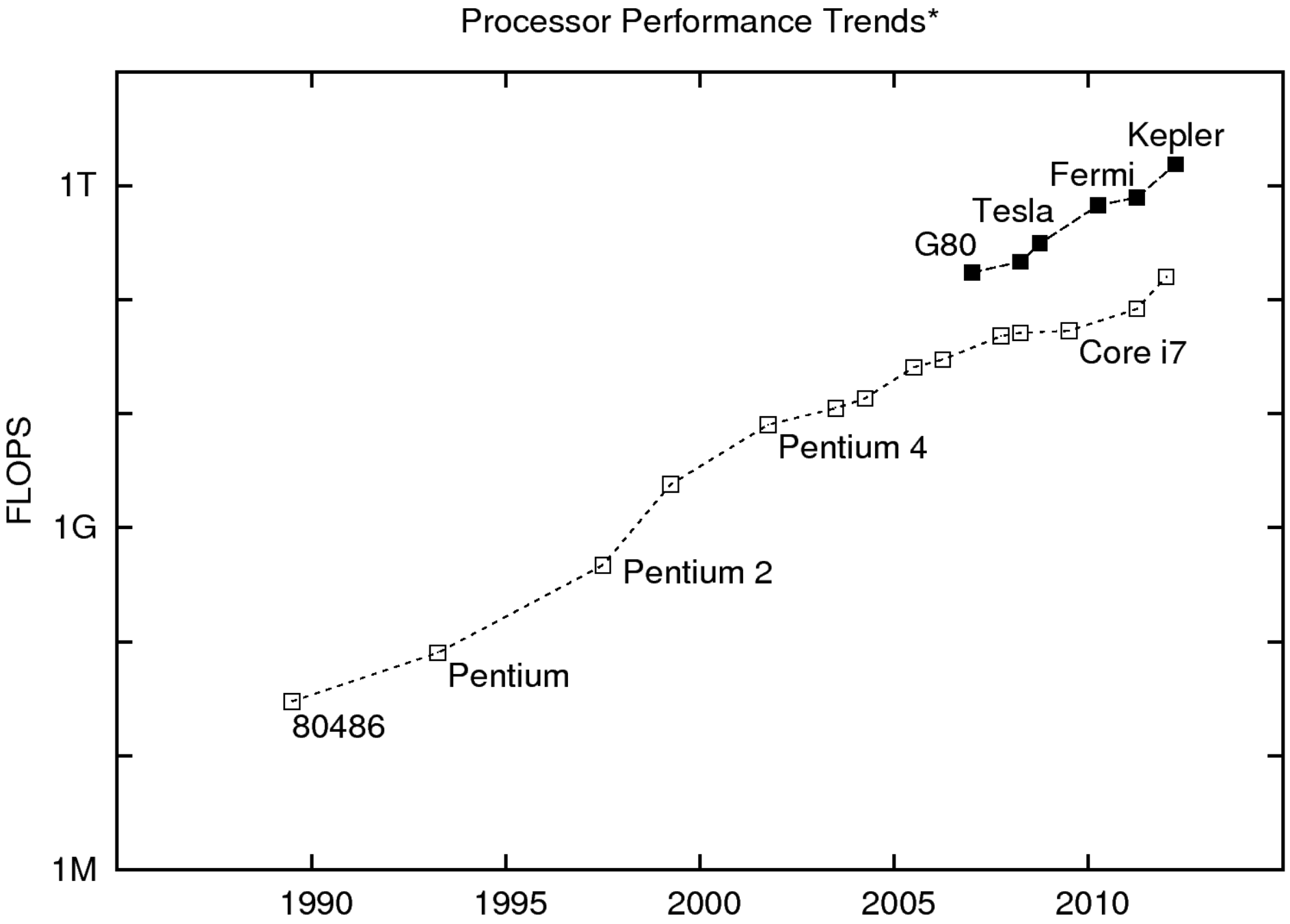
Central Processing Unit (CPU) *n* :

> **the portion of a computer system that carries out the instructions of a computer program, to perform the basic arithmetical, logical, and input/output operations of the system**





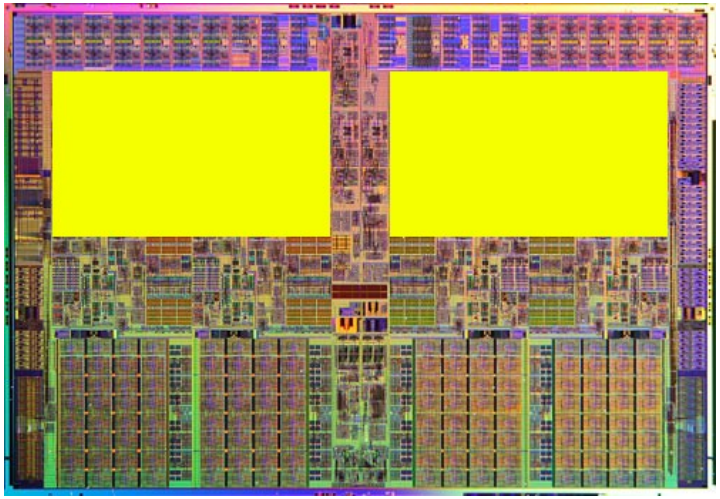
GPU Computing Motivation : Arithmetic Performance



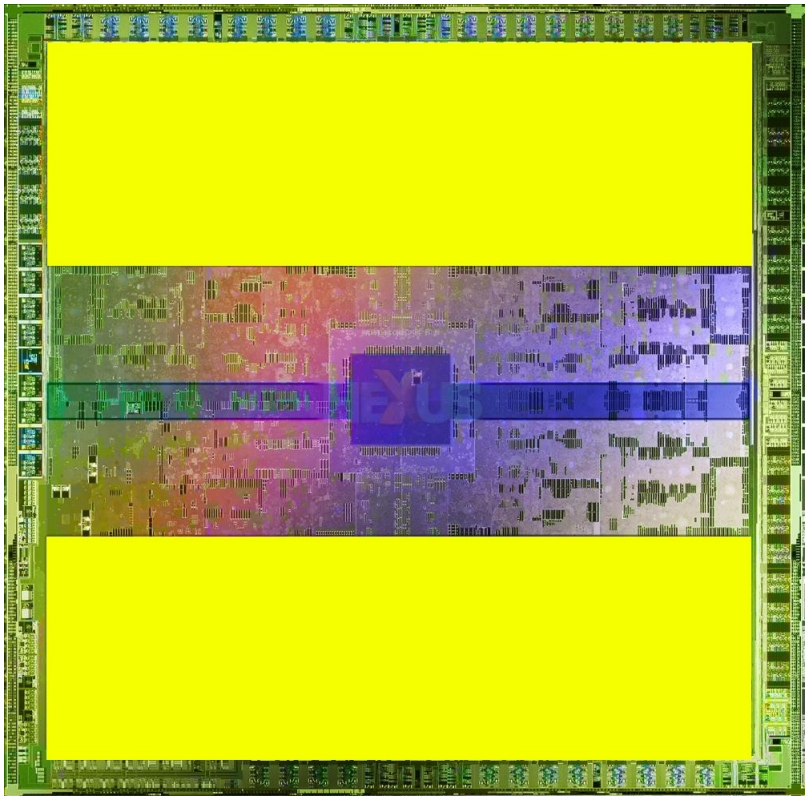
GPU Computing Motivation

How is that possible?

- **GPUs have less cache and logic, more arithmetic**
- **GPUs execute lots of simple threads**
- **GPUs are physically bigger**

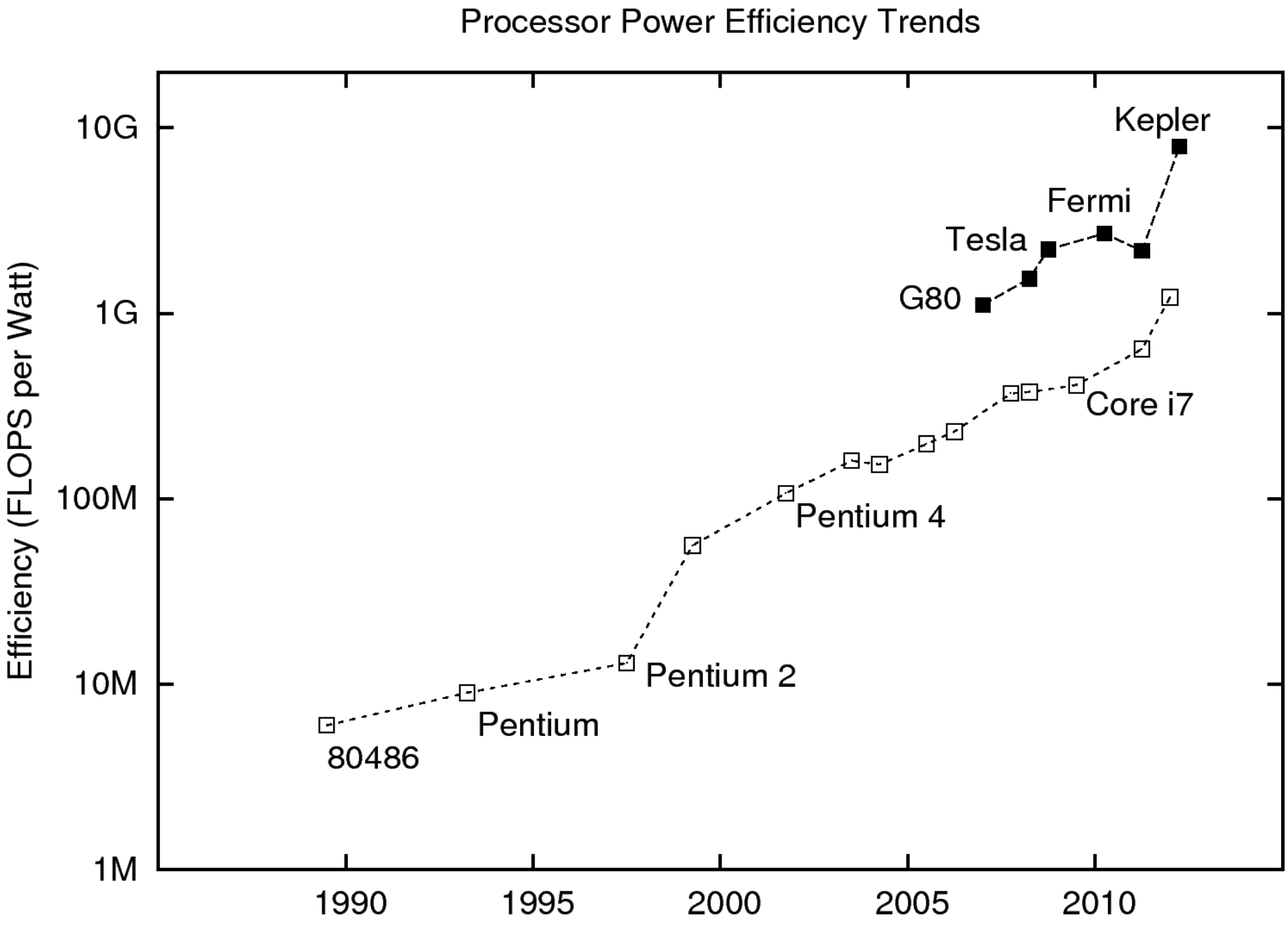


Nehalem Die (263mm^2)

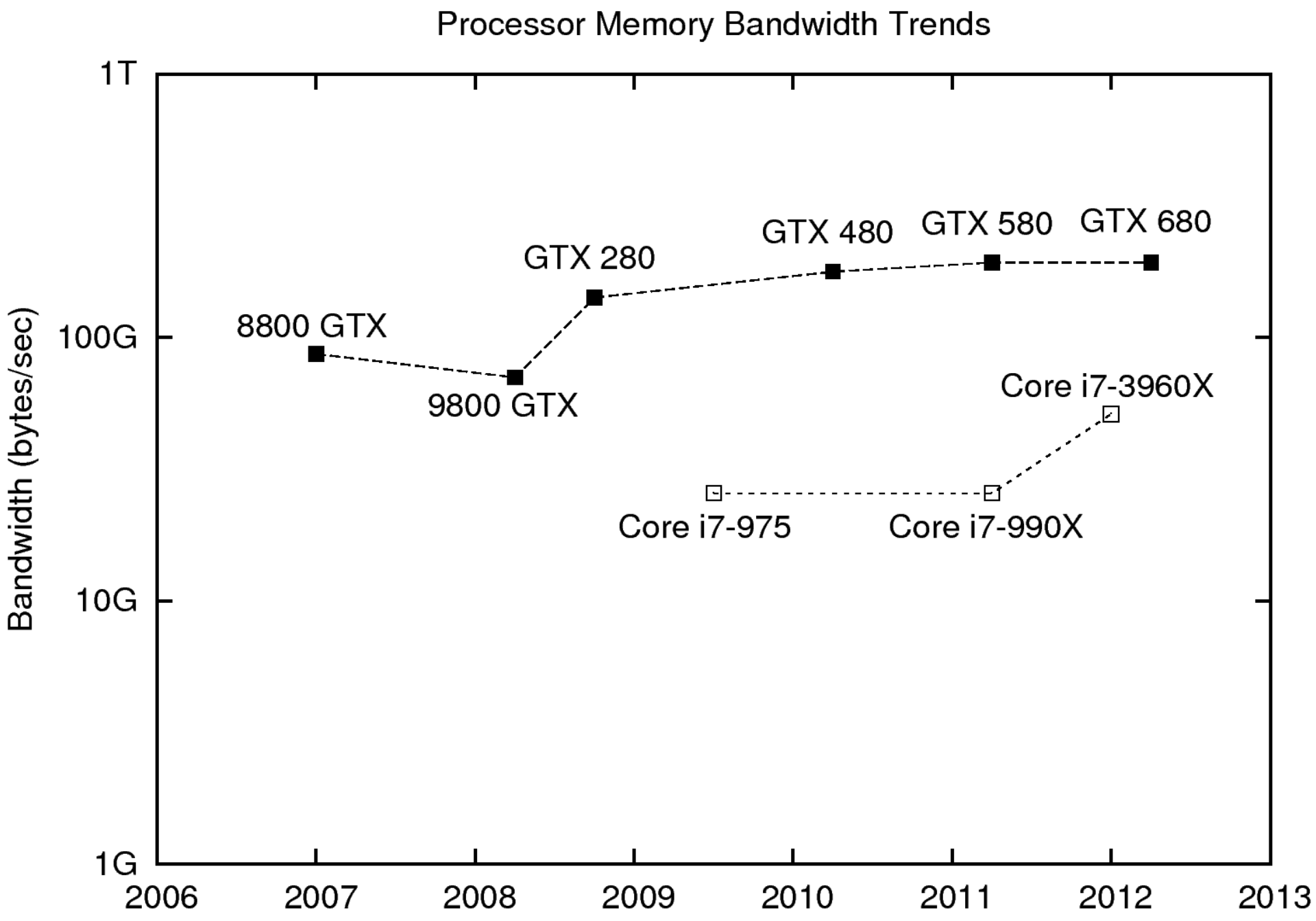


Fermi Die (530mm^2)

GPU Computing Motivation : Power Efficiency



GPU Computing Motivation : Memory Bandwidth



GPU Computing Motivation : Summary

Pros :

- high arithmetic performance
- high memory bandwidth
- power efficient computation

Cons :

- separated by PCI Express bus
- requires fine grain parallelism
- requires high arithmetic intensity
- requires light weight threads
- requires additional software development

For the right algorithm, additional time spent on code development can yield significant performance gains.

Introducing iVEC

Encourages Western Australian research using:

- **supercomputing**
- **large scale data storage**
- **visualisation**

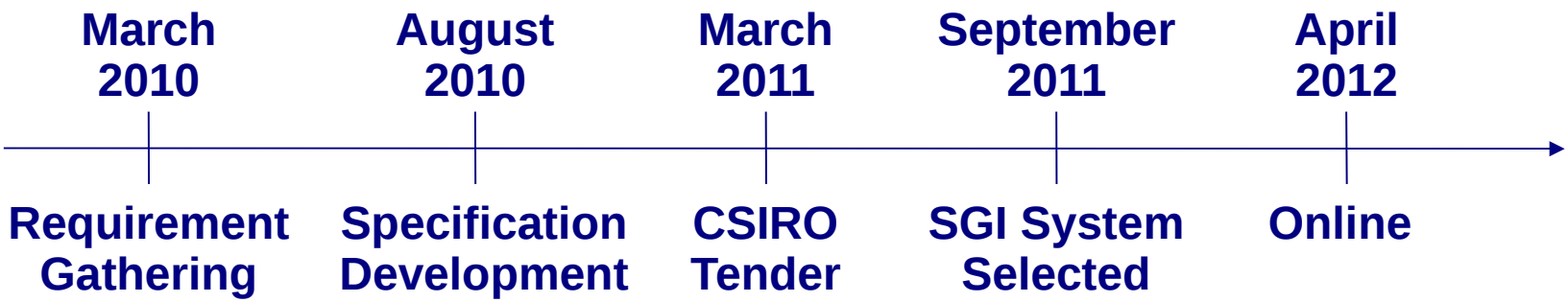
Provides resources to research, education, and industrial communities.

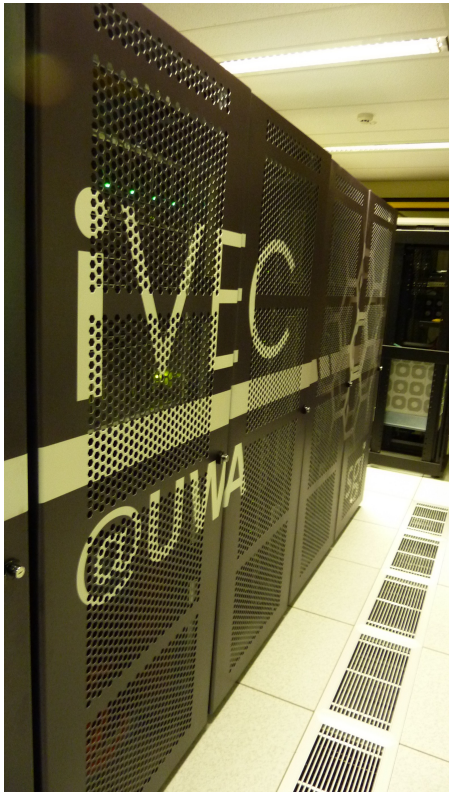
Application areas such as:

- **nanotechnology**
- **radio astronomy**
- **high energy physics**
- **medicine**
- **mining**
- **architecture**
- **multimedia**
- **urban planning**

Pawsey Phase 1B : Fornax Supercomputer

1A	1B	Building	Phase 2





96 compute nodes + 4 development nodes

2 hex-core Intel Xeon X5650 CPUs per blade (2304 cores)

72GB RAM per node (~7TB total)

single NVIDIA Tesla C2075 GPU per node

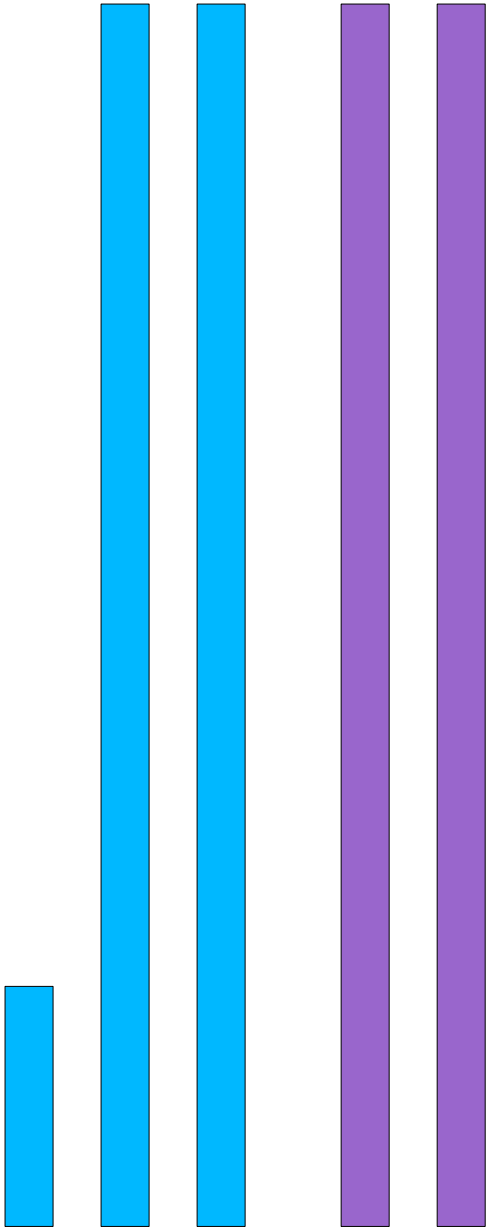
Dual Infiniband QDR 40Gbps interconnect

700TB local, 500 TB scratch, 10Gbps external

~100TF GPU performance

located at iVEC@UWA

Fornax: System Architecture Overview





Head Node

fornax-hn

http://fornax-monitor.ivec.org/gweb

VM Nodes

fornax-xen | ldap | license | monitor

Login Nodes

login1 | login2

Copy Nodes

io1 | io2

Compute Nodes

f001 | f002 | . . . . | f095 | f096

Development Nodes

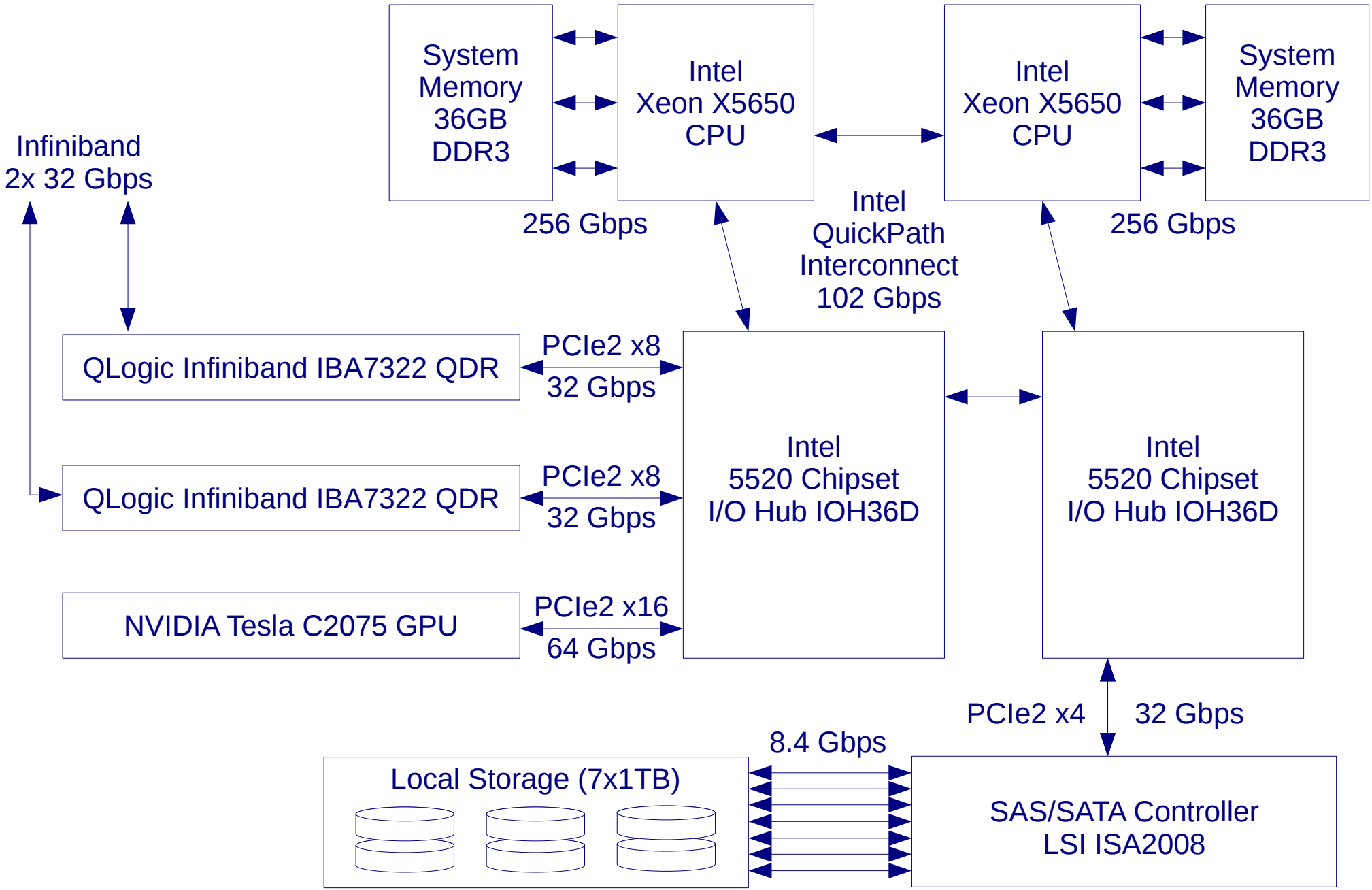
f097 | f098 | f099 | f100

Lustre Nodes

fmds1 | foss1 | foss2 | foss3 | foss4

fmds2 | foss5 | foss6 | foss7 | foss8

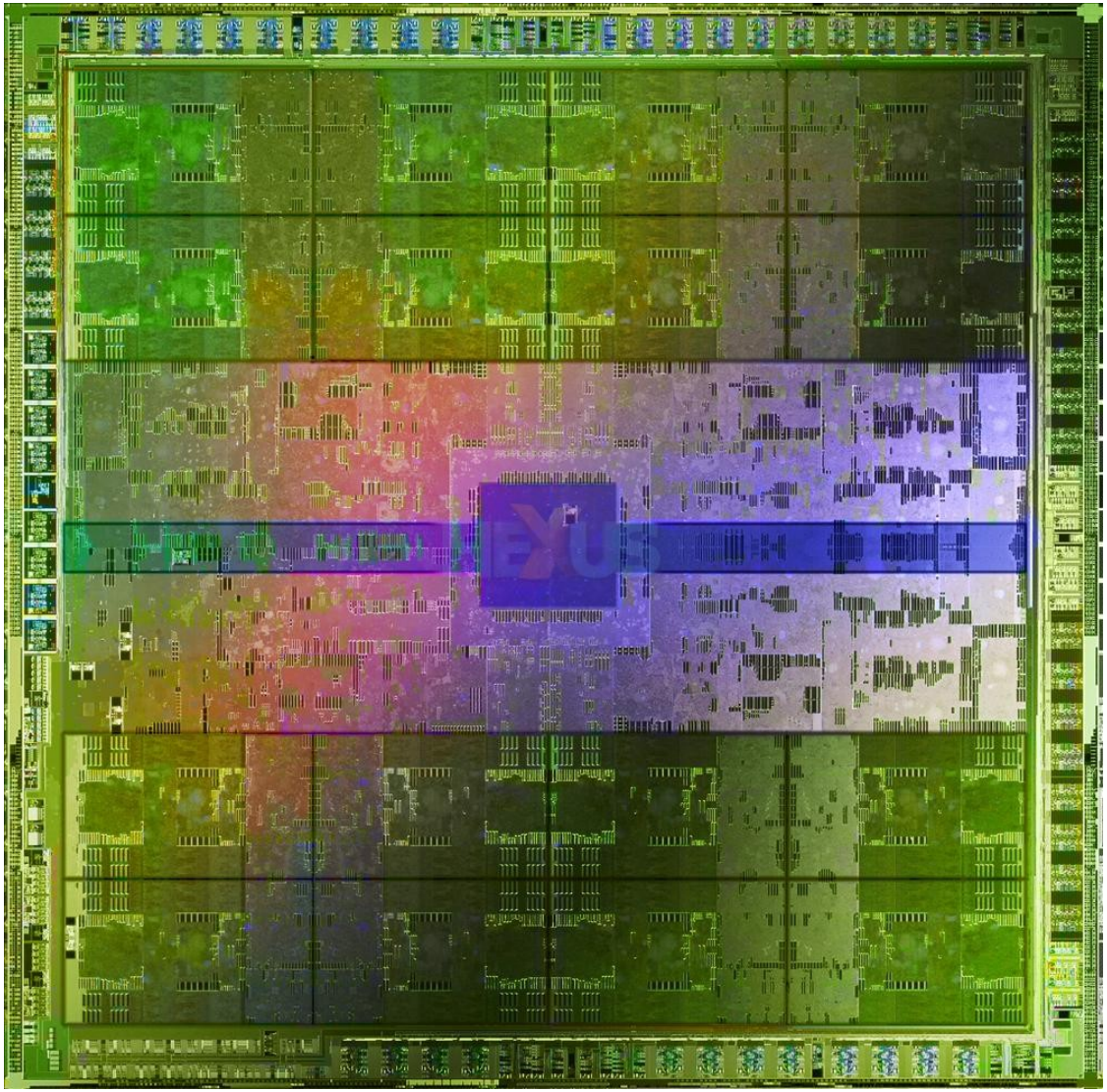
Fornax: Compute Node Architecture (X8DAH)



Fornax: NVIDIA Fermi GPU Architecture

PCIe2 x16

64 Gbps



1152 Gbps

GPU Memory

DDR5

ECC

5.25 GB

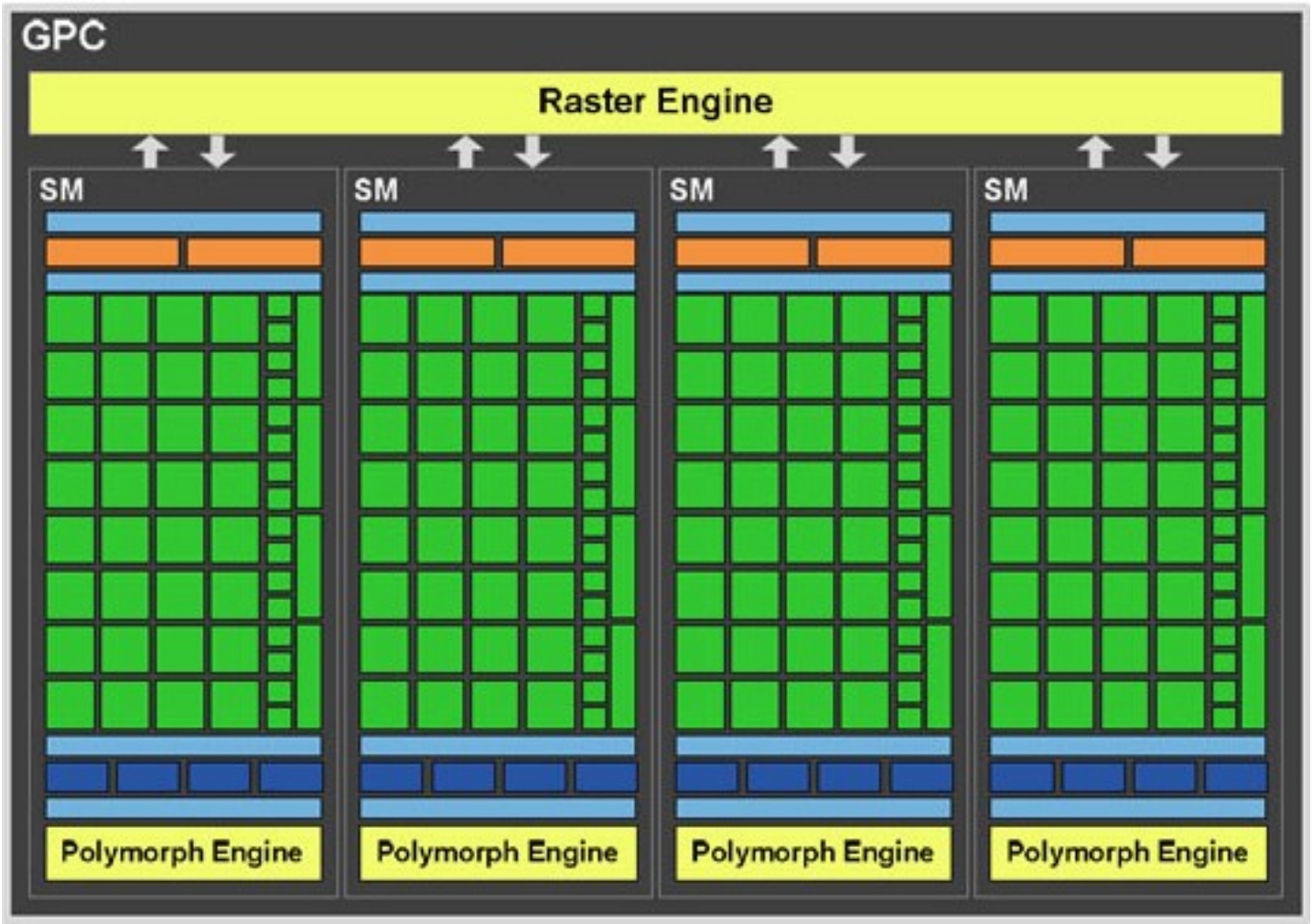
Fornax: NVIDIA Fermi GPU Architecture



Fermi C2070 GPU consists of :

- **Host interface**
- **GigaThread Engine**
- **6x 64 bit DDR5 1566MHz memory controllers**
- **786kB L2 Cache**
- **Four Graphics Processing Clusters (GPC)**

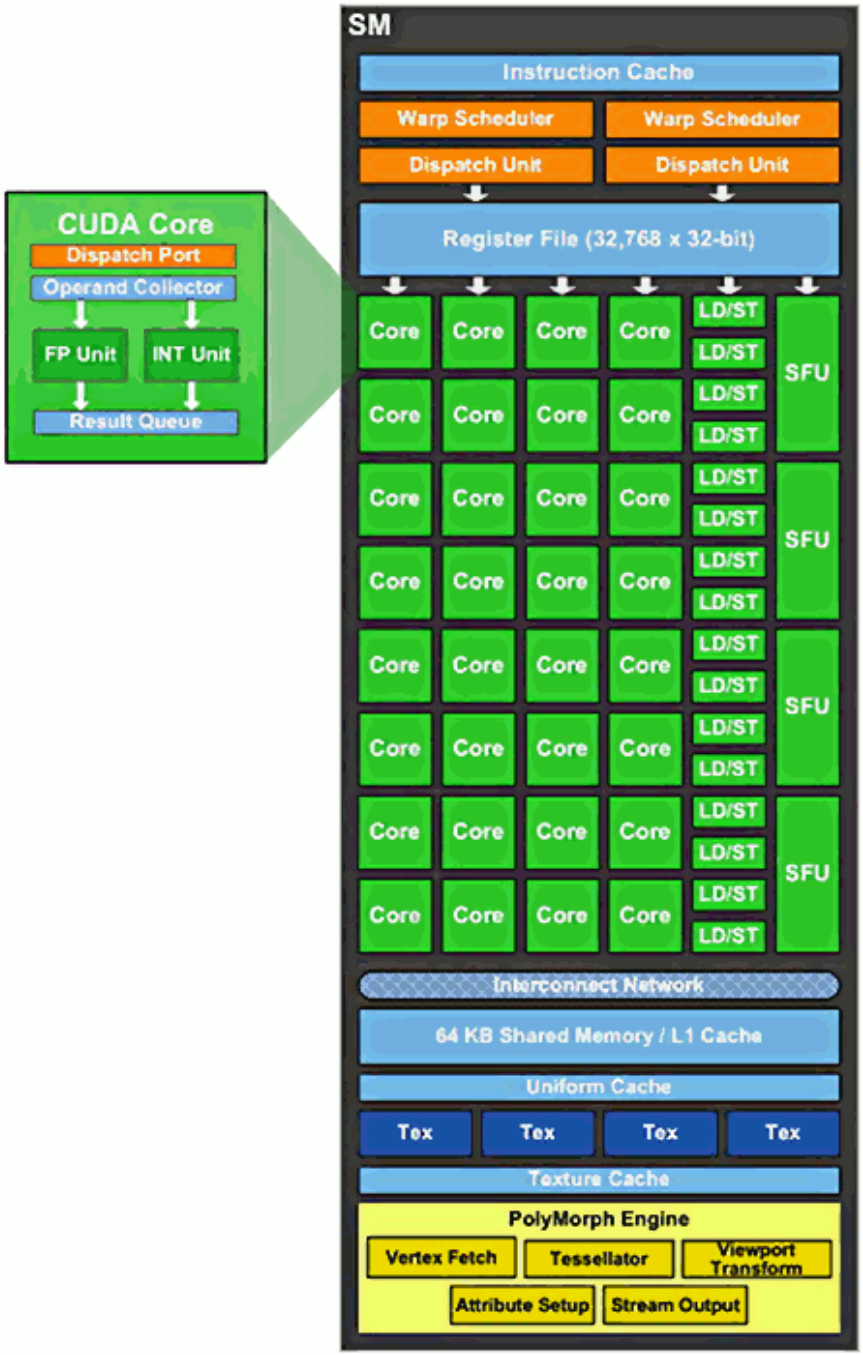
Fornax: NVIDIA Fermi GPU Architecture



Graphics Processing Cluster (GPC) :

- **Raster Engine : triangle setup, rasterization, z-cull**
- **Four Streaming Multiprocessors (SM)**

Fornax: NVIDIA Fermi GPU Architecture



Streaming Multiprocessor (SM) :

- Instruction Cache
- Warp Schedulers
- Dispatch Unit
- Register File (32k 32bit registers)
- 32 CUDA Cores (FP and INT unit each)
- 16 LD/ST Units
- 4 Special Function Units (SFU)
- Interconnect Network
- 64 KB Shared Memory / L1 Cache
  (1:3 or 3:1 configurable)
- Uniform Cache
- Texture Units
- Texture Cache
- Polymorph Engine

NVIDIA Tesla C2075 has 14 SMs enabled

- 448 CUDA cores @ 1.15 GHz
- Fused Multiply Add (FMA) per clock
- 1.030 SP TFLOPS per GPU

Fornax Accounts

Our facilities are available under the iVEC Conditions of Use, you can find them online at:

http://www.ivec.org/services/apply-account/condition-use

Temporary course accounts will be provided, which will last for the duration of the course.

Data associated with this account could be deleted at any time, ensure you have a copy of any files you wish to keep.

Access to iVEC resources is project based, and obtained via regular allocation processes. Find out more at:

http://www.ivec.org/services/apply-account

Logging on to Fornax

Logging in:

```
ssh username@fornax.ivec.org

ssh -u username fornax.ivec.org
```

Beware the DenyHosts intrusion detection daemon!

Repeatedly entering the wrong username or password will get your IP address blacklisted, resulting in the following error:

```
ssh_exchange_identification:
  Connection closed by remote host
```



If this happens, send an email to help@ivec.org to get your IP address reinstated

PBS Scripts on Fornax

PBS Script Path and Filename:

/scratch/*projectname/username/etc/subQuery*

PBS Script Contents:

```
#!/bin/bash
#PBS -W group_list=projectname
#PBS -q workq
#PBS -l walltime=00:02:00
#PBS -l select=1:ncpus=1:ngpus=1:mem=64gb
#PBS -l place=excl

module load cuda
module load cuda-sdk

cd /scratch/projectname/username/etc
deviceQuery
```

Example: Running a script on Fornax

Get template script:

```
cp /scratch/courses01/subQuery .
```

Edit template script:

```
vim subQuery
```

Submitting the job:

```
qsub subQuery
```

Viewing the queue:

```
q
qstat
```

Viewing the output:

```
cat subQuery.o#####
cat subQuery.e#####
```

Running GPU programs on Fornax

deviceQuery.o#####:

```
deviceQuery Starting...

 CUDA Device Query (Runtime API) version (CUDART static linking)

Found 1 CUDA Capable device(s)

Device 0: "Tesla C2075"
  CUDA Driver Version / Runtime Version          4.1 / 4.1
  CUDA Capability Major/Minor version number:    2.0
  Total amount of global memory:                 5375 MBytes (5636554752 bytes)
  (14) Multiprocessors x (32) CUDA Cores/MP:     448 CUDA Cores
  GPU Clock Speed:                               1.15 GHz
  Memory Clock rate:                             1566.00 Mhz
  Memory Bus Width:                              384-bit
  L2 Cache Size:                                 786432 bytes
  ...
  Total amount of constant memory:               65536 bytes
  Total amount of shared memory per block:       49152 bytes
  Total number of registers available per block: 32768
  ...
```

Running GPU programs on Fornax

deviceQuery.e#####:

```
[deviceQuery] starting...
[deviceQuery] test results...
PASSED
> exiting in 3 seconds: 3...2...1...done!
============================================================
Job ID: 4790.fornax
User ID: charris
Group ID: director100
Job Name: subQuery
Session ID: 9656
Resource List: mem=64gb,ncpus=1,partition=1,place=excl,walltime=00:02:00
Resources Used:
cpupercent=0,cput=00:00:03,mem=1396kb,ncpus=1,vmem=106288kb,walltime=00:00:06
Queue Name: workq
Account String: null
Exit Code: 0
```

Common Problems

Missing qsub

```
[username@login1 ~]$ qsub
-bash: qsub: command not found
```

solution

```
module load pbs
```

Missing cuda_runtime.h (or cl.h)

```
[username@login1 ~]$ make
gcc -Wall filename.c -o outputname -lcudart
filename.c:#:#: error: cuda_runtime.h: No such file or directory
etc
```

solution

```
module load cuda

(or manually provide the path to your software, 'module display
cuda' may be helpful for this)
```

Common Problems

Job sitting in queue forever

```
[username@login1 ~]$ qstat
Job id              Name             User              Time Use S Queue
----------------  ---------------- ----------------  -------- - -----
3720[].fornax     jobX             otherperson              0 B workq
4962.fornax       jobY             somebodyelse      49:32:42 R workq
4967.fornax       jobZ             somebodyelse             0 H workq
4970.fornax       myjob            username                 0 Q workq
```

solutions

- *double check the PBS directives in your submission script*
- *wait until the scheduler decides it is your turn to run*
- *contact help@ivec.org if necessary*

Missing CUDA libraries (or OpenCL libraries)

```
[username@login1 device]$ cat myjob.e#####
/home/username/path/to/binary: error while loading shared libraries:
libcudart.so.4: cannot open shared object file: No such file or directory
```

solution

```
module load cuda
```

Common Problems

No CUDA devices detected

```
[username@login1 ~]$ ./device
no CUDA-capable device is detected
```

solutions

- *don't try to run on nodes that don't have GPUs (login,io/copyq)*
- *contact help@ivec.org if necessary*


iVEC