



GPU Clouds – IGT 2009

Cloud Computing Summit
Mordechai Butrashvily, CEO
moti@hoopoe-cloud.com



Agenda

- Introduction to GPU Computing
- Future GPU architecture
- GPU on a Cloud:
 - Visualization
 - Computing
- Summary



About Us

- Former members of IMOD
- Head of GPGPU working group @ IGT
- Developing industrial GPGPU solutions
- Formed in 2008, after years of experience
- Large experience with graphics, HPC and distributed systems/programming

What is a GPU?

- **Graphics Processing Unit** – previously served for graphic transformations



- **Features:**

Computing Power:	Up to 2+ TFLOPS
Dedicated Memory:	Up to 6 GB
Cores	Up to 1600
Memory bandwidth:	Up to 250 GB/s

- **Vendors: Intel, AMD, NVIDIA**
- **Actually, a GPU is like a small GRID/Cluster**

What is a GPU?

- **Graphics Processing Unit** – previously
serv... ons
- **Features**

Computing	100%
Dedicated I/O	100%
Cores	100%
Memory bandwidth	100%
- Vendor **DRAM**
- Actually, a GPU is like a small GRID/Cluster



A GPU core at glance

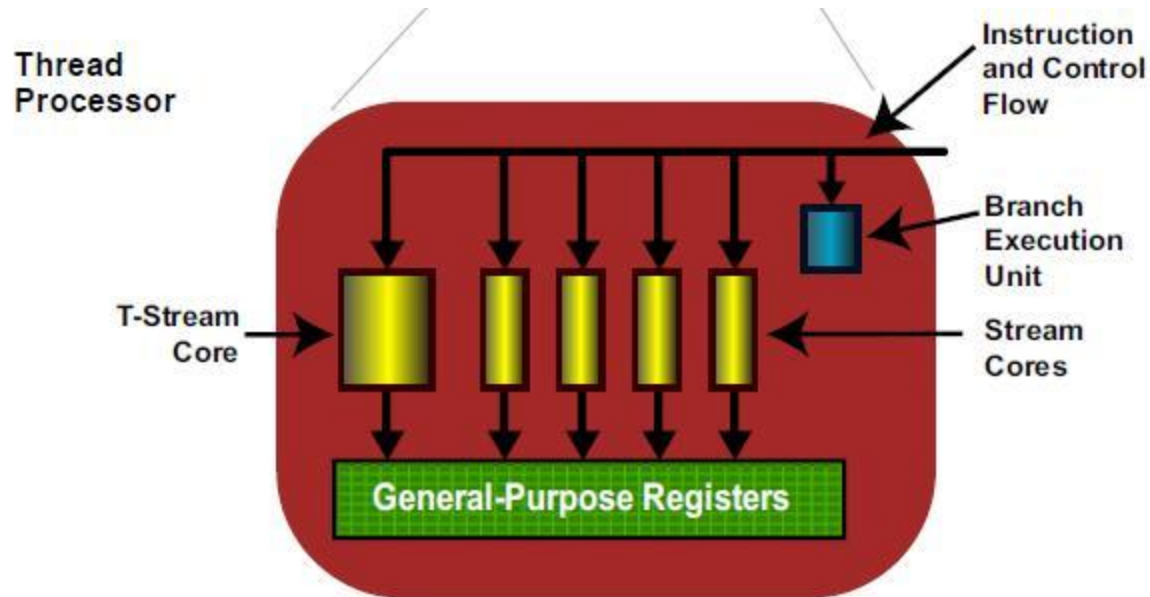


Figure 1.9 Simplified Block Diagram of the Stream Processor¹



GPU Features

- GPU = Co-processor
- CPU compatible arithmetic operations
- Single/double precision, integer operations
- Integrates with C/C++, .NET, FORTRAN, Java, Python, Ruby and more
- OS support (32/64 bit):
 - Windows
 - Linux
- Especially good for distributed algorithms, but not just
- Can do FFT, BLAS, LAPACK, OpenCV and more



Maps to Every Scale





Software Environments

- CUDA
 - C based language, supports C++
 - Targeting NVIDIA products
 - Windows / Linux / Mac support
- OpenCL
 - Emerging standard
 - Supported across vendors
 - Heterogeneous programming (CPU + GPU + DSP...)
 - Windows / Linux / Mac support



GPU Computing

- Using graphics processors for general tasks
- **Successful** in many applications:

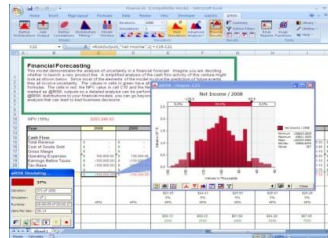
Oil & Gas



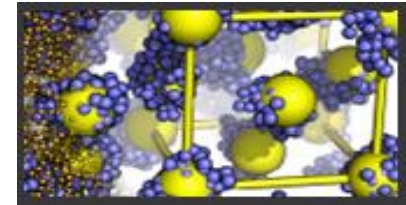
Biology / Medics



Finance



Simulations



- Provides **strong** computing power





Future GPU architecture

- NVIDIA “Fermi” (Additional presentation)



GPU on a Cloud

- Cloud services are based on virtualization
- Achieving better resource usage
- The GPU is not yet virtualized
- Other solutions exist – solving large domain problems:
 - **Visualization** – RealityServer® by mental-images®
 - **Computing** – Hoopoe™ GPU Cloud Computing



GPU Visualization Cloud

- mental-images RealityServer®
- Share design workbench with colleagues
- Perform photorealistic simulations
- (Additional Presentation)



GPU Cloud Computing

A milestone in cloud systems &
(super)computing



Agenda

- GPU Performance
- Introduction
- Motivation
- Services
- Virtualization / Security
- Examples
- Summary

GPU Performance

- **1 GPU hour = (~)**

Computation	Amount (hr)	Details
FFT 2D	36,000,000	128x128 transform
Image Erode	10,800,000	1280x1024 pixels
Image smooth (Gaussian)	3,600,000	1280x1024 pixels
Black & Scholes (option pricing)	1,080,000,000	Derivative options





Introduction to Hoopoe™

- Named after “Dukhifat”, the official bird of Israel for the 60 years celebrations
- A computational “cloud”, also for GPUs
- Bringing the power of thousands of GPUs to consumers
- Already having registrations (> 100’s organizations)
- **SaaS/PaaS** – Web Service based
- Stage - Alpha(α) phase





Motivation

- Providing **easier access** to computing resources and well known libraries/algorithms
- Not an ordinary batch system (SGE, LSF, Condor)
- Provides **real-time** distribution performance
- **Efficiently** using GPU hardware in cluster environments
- **Security & Virtualization** – Achieving better security by hiding the cluster from the user
- All cluster management and orchestration are performed by a the Hoopoe infrastructure



Service Highlights

- **Computing** - Access GPU resources for computing or graphics tasks (“Hoopoe”)
 - **CUDA™** – NVIDIA C language, intuitive
 - **OpenCL™** - new standard, compatible across vendors
 - Additional **Apps** – FFT, BLAS, OpenCV and more...
- **Storage** - Host data for computing
 - “*HoopoeFS*” - an internal storage service
 - Can also connect with **Amazon S3** for extended storage support, or integrate with **EC2**
- Can integrate with any technology (.NET, Java ...)



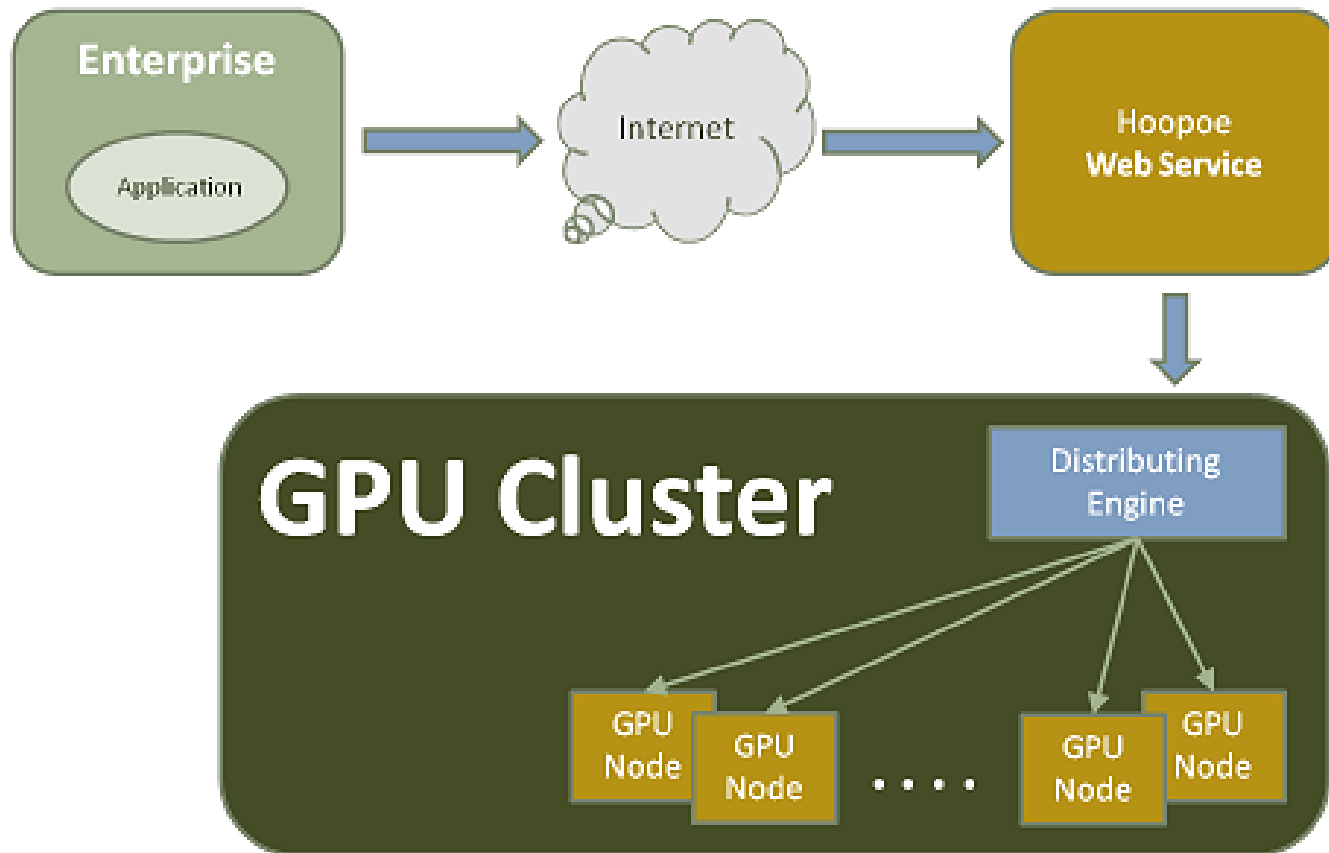


New Hoopoe™ Service

- **Encodeo™** - video transcoding services
- Self developed library and service
- Accelerated video transcoding to HD standards and beyond (x10 + speedup)
- Adapt original video to multiple devices or convert between video codecs/formats
- Pay-per-use = 1/10 the time, 1/10 the price
- Available for media content providers
- Web service based



System Overview





Virtualization

- A key feature of cloud environments, yet missing for GPUs
- Not like with general cloud computing (time sharing of CPU)
- We provide the virtualization level on top of the GPU
- Based on user supplied details
- Achieving finer-grained parallelism and better resource utilization



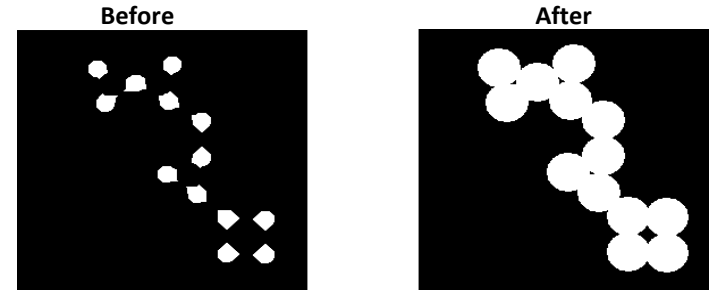
Security

- A problem with all cloud systems where data enters the public domain (Internet)
- In Hoopoe the user has no direct access to the cluster
- Our code manages resources and computing
- The user only runs code inside the GPU
- The GPU is passive, with a special architecture – **not harmful**
- Every user has isolated storage & execution environment
- Actions audit is provided to track activities



Example applications

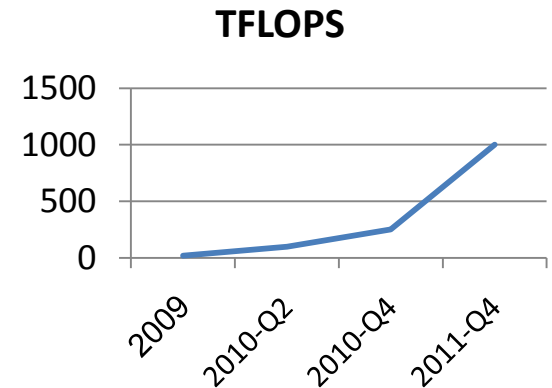
- Image processing
- Finance
- Seismic analysis/processing
- AOI (e.g. wafer inspection)
- Digital printing
- Other simulations (biology, physics, chemistry and more)





Plans for the future

- Extending total computing power (reaching 1 **PetaFLOPS** around 2011)
- Extending available Apps and libraries (Finance, simulations etc.)
- Provide Map-Reduce capabilities
- Integrate with more cloud services (Azure)
- Deploy Hoopoe in private clusters





Examples

1. Submitting an FFT task
 - Using the web
 - Programmatically
2. A custom algorithm – Image Erode



Web FFT example

- Submitting an FFT task using the web
- Process:
 1. Login
 2. Define FFT task parameters and metadata
 3. Click “Submit”
- Wait for results and notification in the email



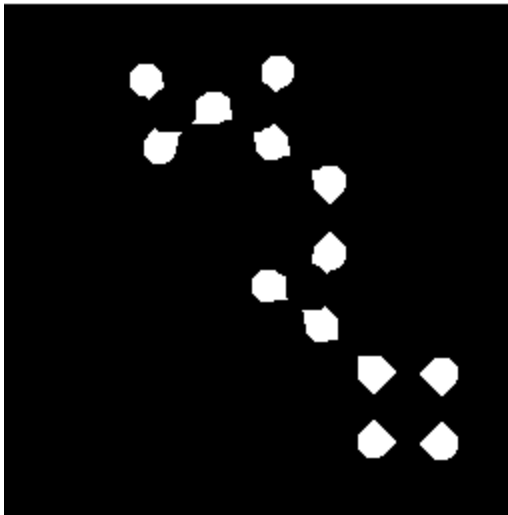
FFT program example

- Integrating with Hoopoe Apps interface
- A simple .NET application
- Defining FFT parameters
- Submitting the task

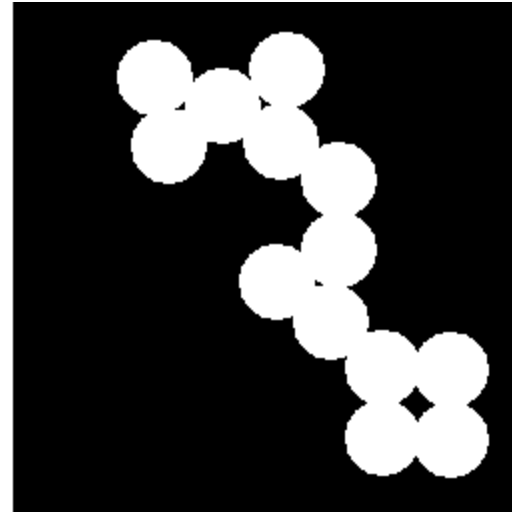
Erode example

- Submitting a custom algorithm to Hoopoe
- Used for image processing operations
- Image: 1280x1024, 32 bit

Before



After





Summary

- GPU's are a mature platform for computing purposes
- Outperforms existing solutions by: Price, performance, easy of use / integration
- **Hoopoe™** is our vision for how computing should be
- GPU based cloud services are just beginning!
- **Contact us for more details**

Questions

