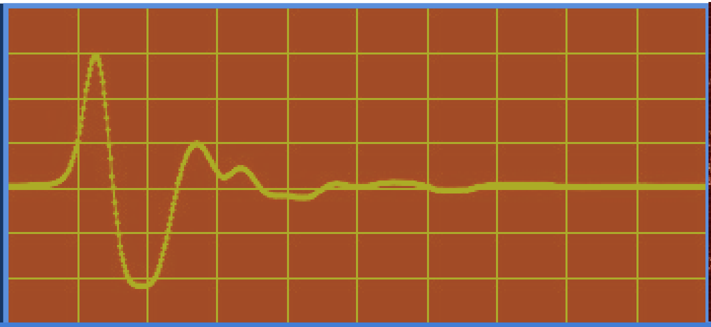
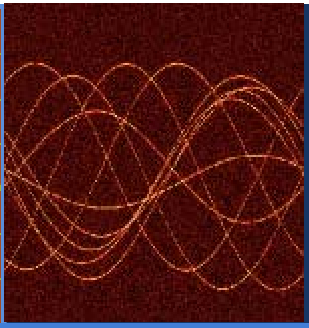
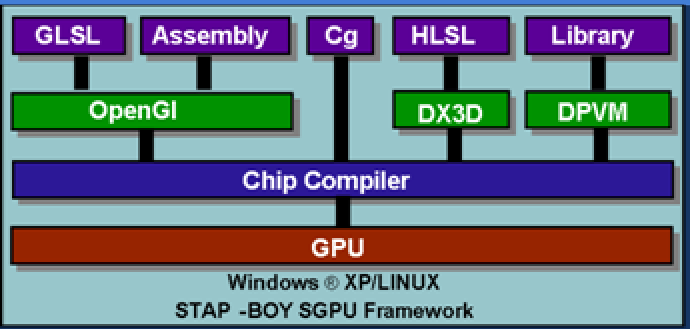






GLSL
Assembly
Cg
HLSL
Library
OpenGl
DX3D
DPVM
Chip Compiler
GPU
Windows ® XP/LINUX
STAP -BOY SGPU Framework

# Power Consumption of Desktop and Mobile GPUs for IRSTAP Applications

Michael Roeder, Jeremy Furtek, Nolan Davis,
Cezario Tebcherani, Masatoshi Tanida and Dennis Braunreiter

High Performance Embedded Computing (HPEC) Workshop, 23-25 September 2008

# STAP-BOY: Concept

## Problem:

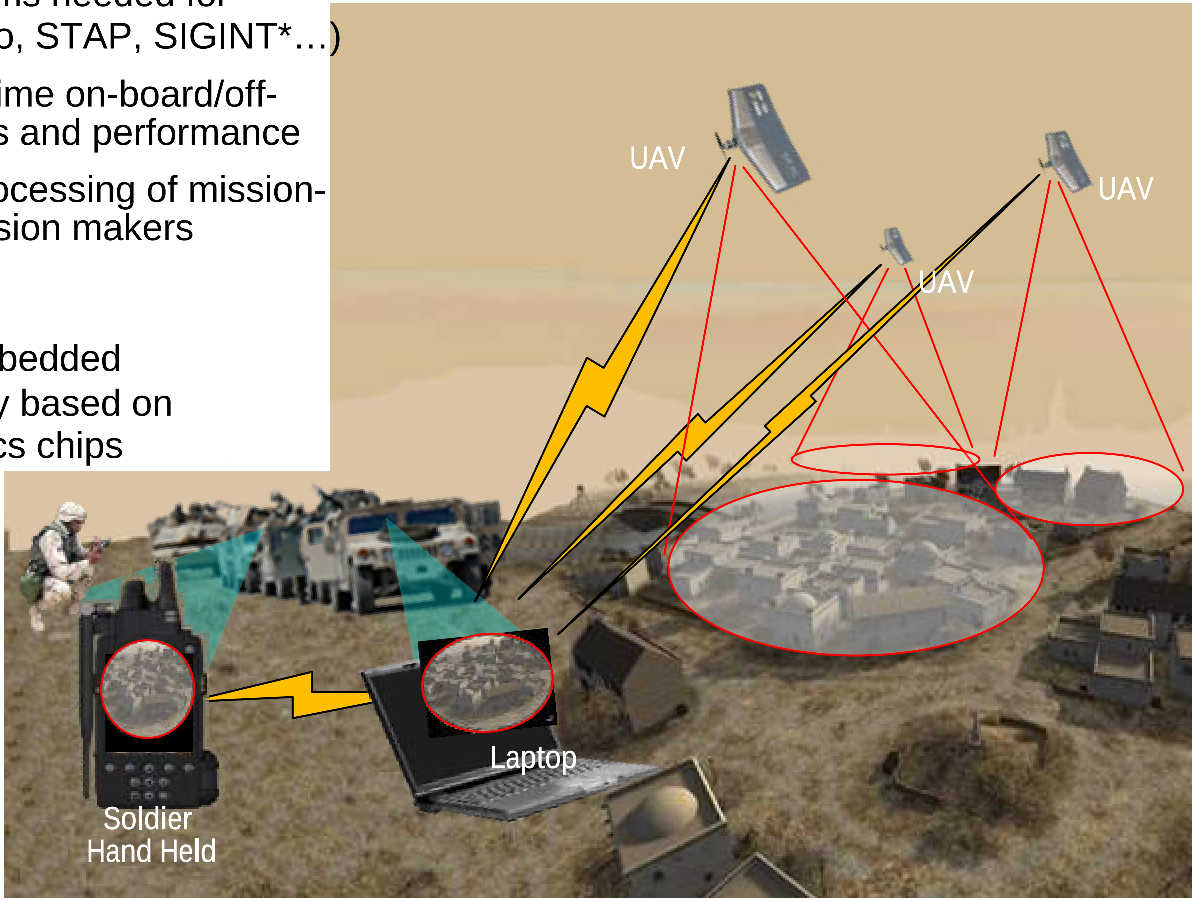
- Complex sensor modalities and algorithms needed for smaller platforms (SAR, 3D-motion video, STAP, SIGINT*…)
- Low-cost platform constraints limit real-time on-board/off-board and distributed sensing algorithms and performance
- Timely distribution, visualization, and processing of mission-critical data not available to tactical decision makers

## STAP-BOY goal:

- Develop low-cost, scalable, teraflop, embedded multi-modal sensor processing capability based on commercial off-the-shelf (COTS) graphics chips

## STAP-BOY approach:

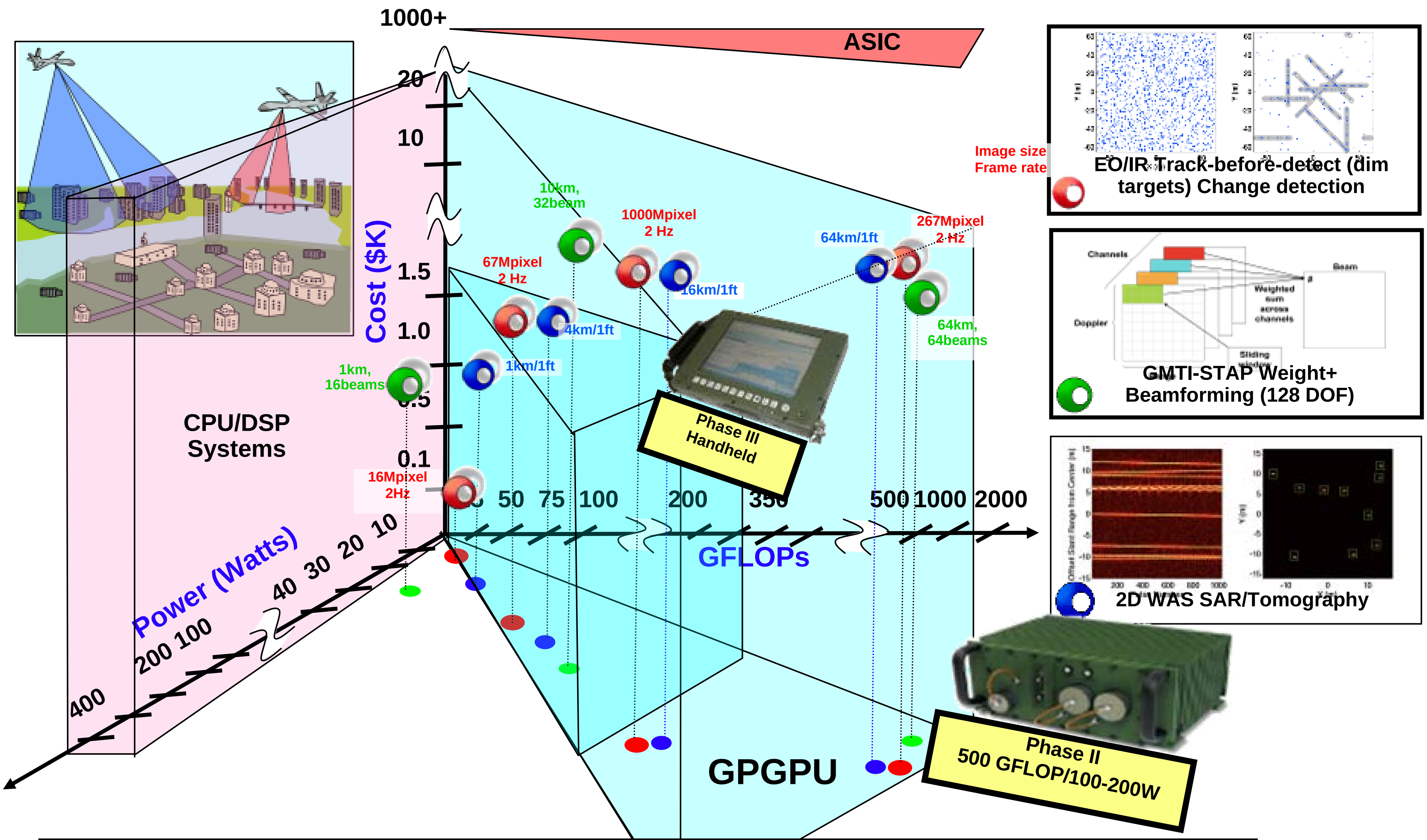
- Map complex algorithms to COTS graphics chips with open source graphics languages
- Prototype scalable parallel embedded computing architecture for handhelds to teraflop single card
- Demonstrate on available tactically representative sensor systems



**Goal:**

**½ Teraflop**
**10 Mobile GPUs**
**100W Total Power**
**$<15K**

# Applications Pull



**Increasing Advanced Sensor Resolution Driving Processing Throughput Beyond CPU/DSP Capability In Communications and Cost-constrained Platforms**

# NVIDIA® GPU Performance/Power Study

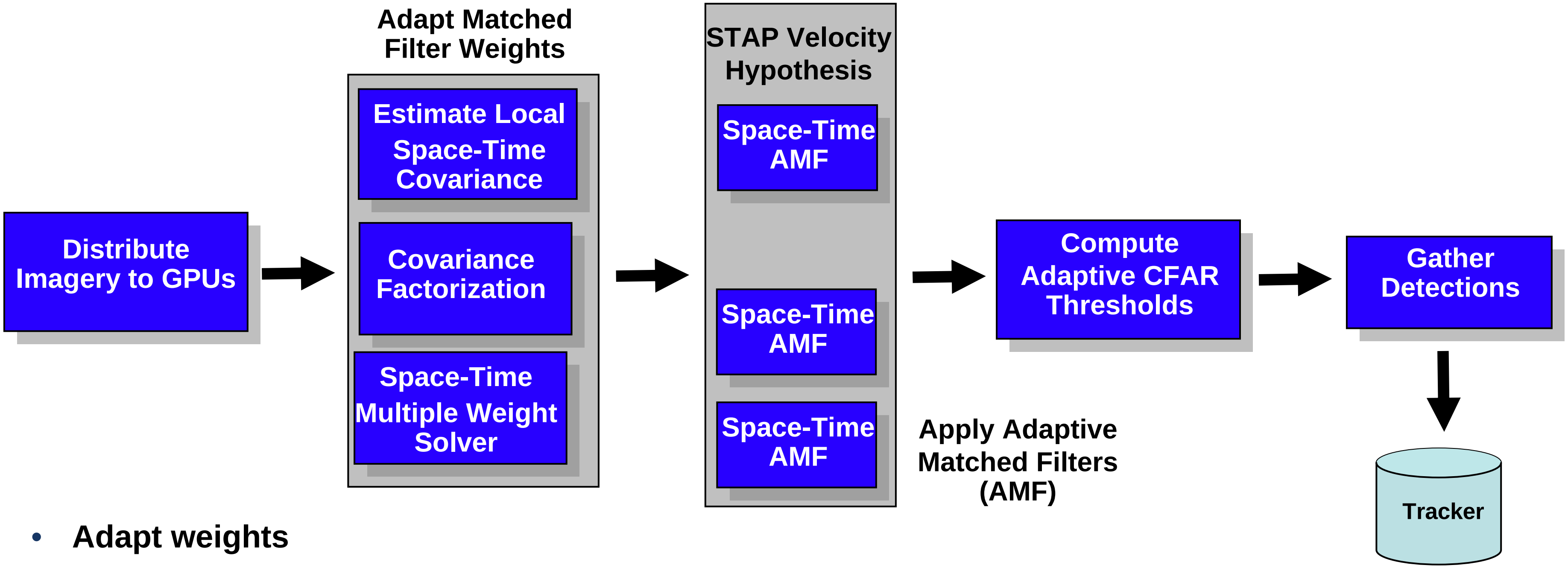
| | GeForce™ 280 GTX | GeForce™ 8800 GTX | GeForce™ 8800M GTX |
|---|---|---|---|
| | | | |
| Category | Desktop | Desktop | Mobile |
| Process (nm) | 65 | 90 | 65 |
| Transistor Count (Millions) | 1400 | 681 | 754 |
| Stream Processors | 240 | 128 | 96 |
| Shader Clock (GHz) | 1.30 | 1.35 | 1.25 |
| Memory Clock (MHz) | 1107 | 900 | 800 |
| Memory Amount (MB) | 1024 | 768 | 512 |
| Thermal Design Power (W) | 236 | 177 | 65 |
| Release Date | June 2008 | November 2006 | November 2007 |

# CPUs vs. GPUs: Head to Head (Floating Point)

| Intel® quad-core Q9650 | | NVIDIA® 280 GTX |
|---|---|---|
| **820 million** | Transistors | **1400 million** |
| **3.0 GHz** | Clock Speed | **1.3 GHz** |
| **4** | Number of Cores | **240** |
| **Serial** | Programming Model | **Highly parallel** |
| **Minimize latency** | Design Goal | **Maximize throughput** |
| **Complex cores:**<br>**• Branch prediction**<br>**• Out-of-order execution** | Design Approach | **Simple cores:**<br>**• Smaller caches**<br>**• In-order execution** |
| **130 W** | Thermal Design Power (TDP) | **236 W** |
| **96 GFLOPS** | Theoretical Max. Computation Rate<br>(single precision) | **933 GFLOPS** |

Intel is a registered trademark of Intel Corporation in the United States and/or other countries.
NVIDIA is a registered is a registered trademark of NVIDIA Corporation in the United States and/or other countries.

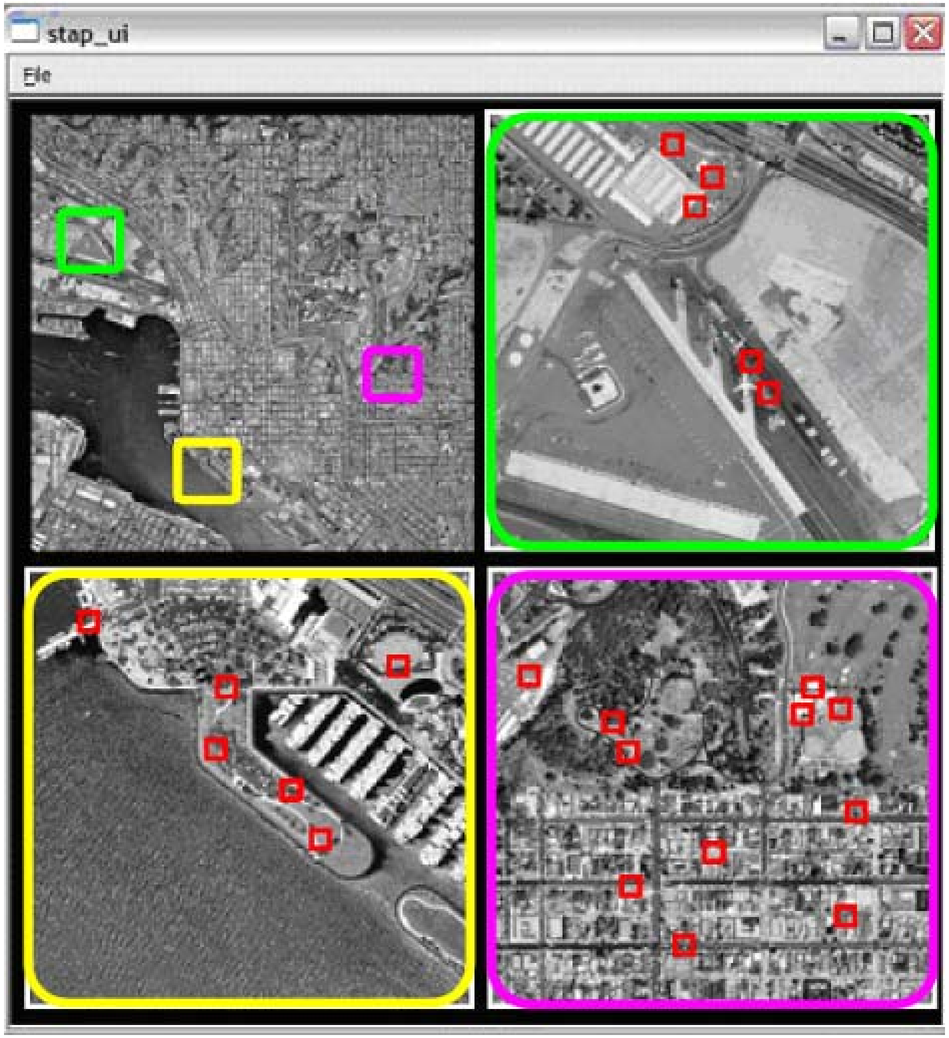
# IRSTAP Application


Adapt Matched Filter Weights
Estimate Local Space-Time Covariance
Covariance Factorization
Space-Time Multiple Weight Solver
STAP Velocity Hypothesis
Space-Time AMF
Space-Time AMF
Space-Time AMF
Distribute Imagery to GPUs
Apply Adaptive Matched Filters (AMF)
Compute Adaptive CFAR Thresholds
Gather Detections
Tracker

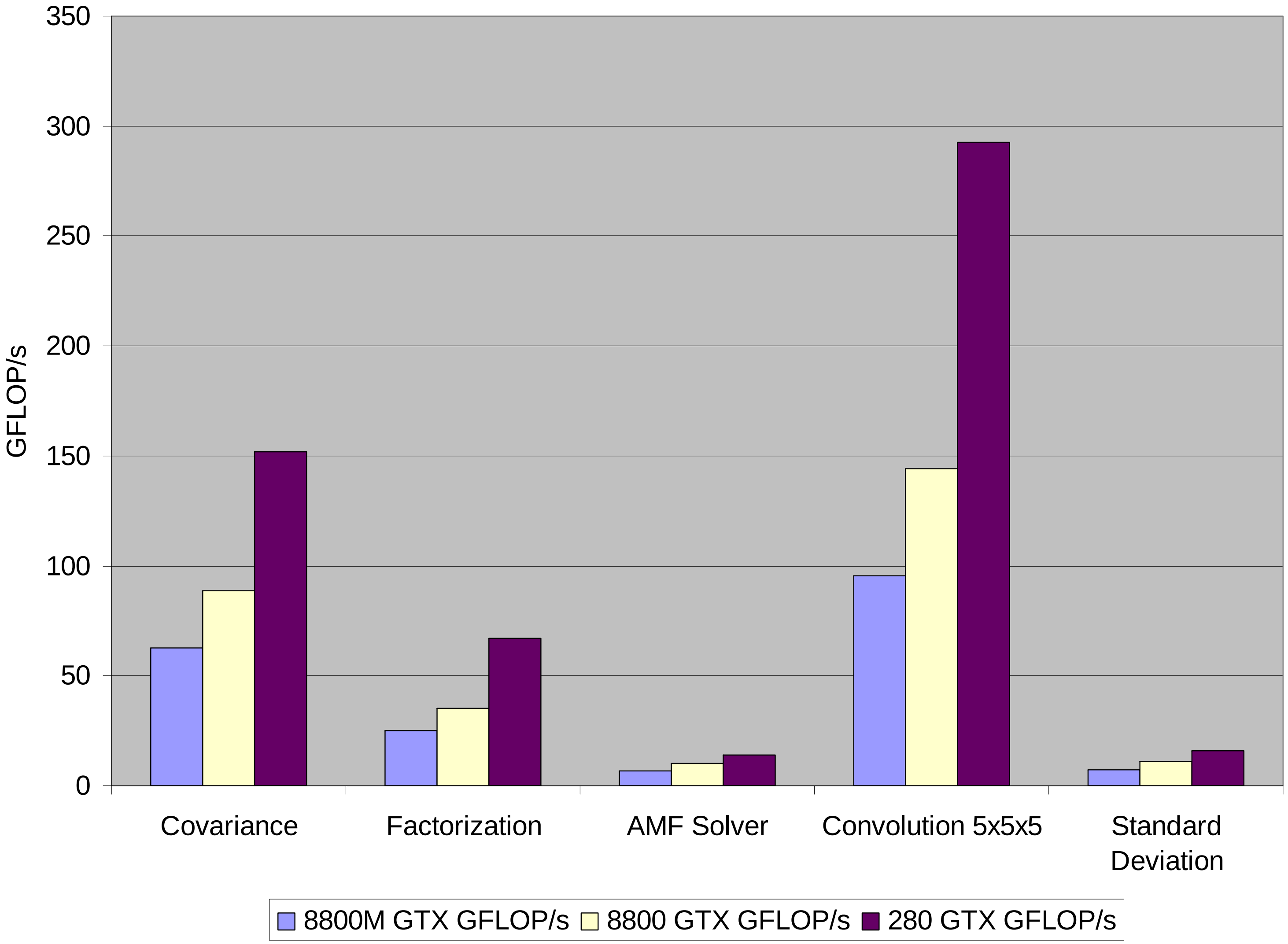
- **Adapt weights**
  - Covariance estimation
  - Covariance factorization
  - Find LMS solution
- **Apply AMF - convolution**
- **Compute CFAR thresholds**
  - Estimate standard deviation for each block
  - Adapt local standard deviations, excluding CUT
- **Gather detections into dense detection list**

# IRSTAP Kernels

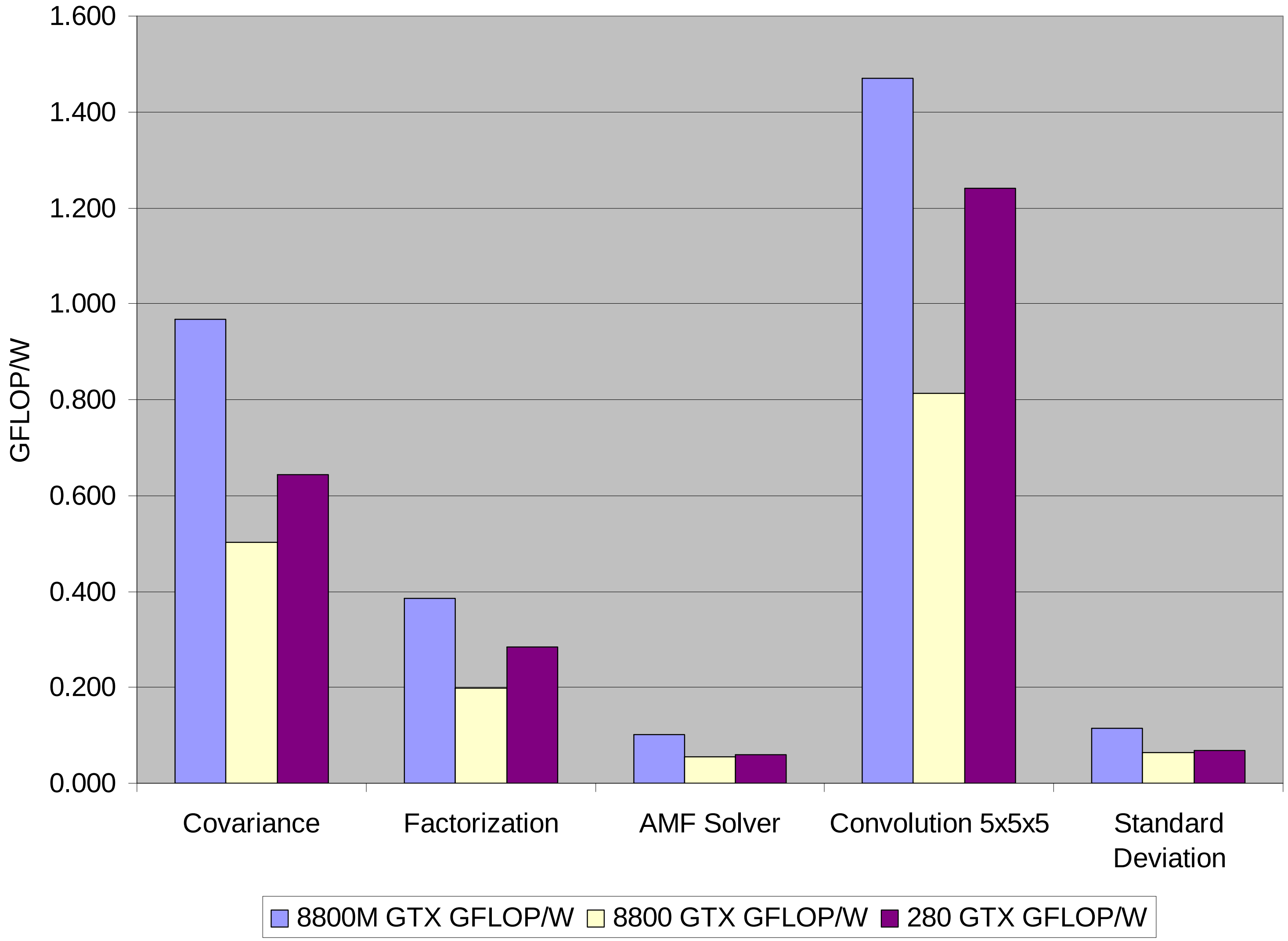
- **Image preprocessing**
  - Registration
  - Clutter subtraction
  - Spatial demeaning
- **Covariance calculation**
  - Ensemble generation
  - Covariance calculation
- **Weight adaptation**
  - Cholesky Factorization
  - Back and forward substitution
- **Weight application**
  - Convolution
- **Adaptive CFAR threshold adaptation**
  - Localized standard deviation
  - Complex mean kernel
- **Detection reduction**



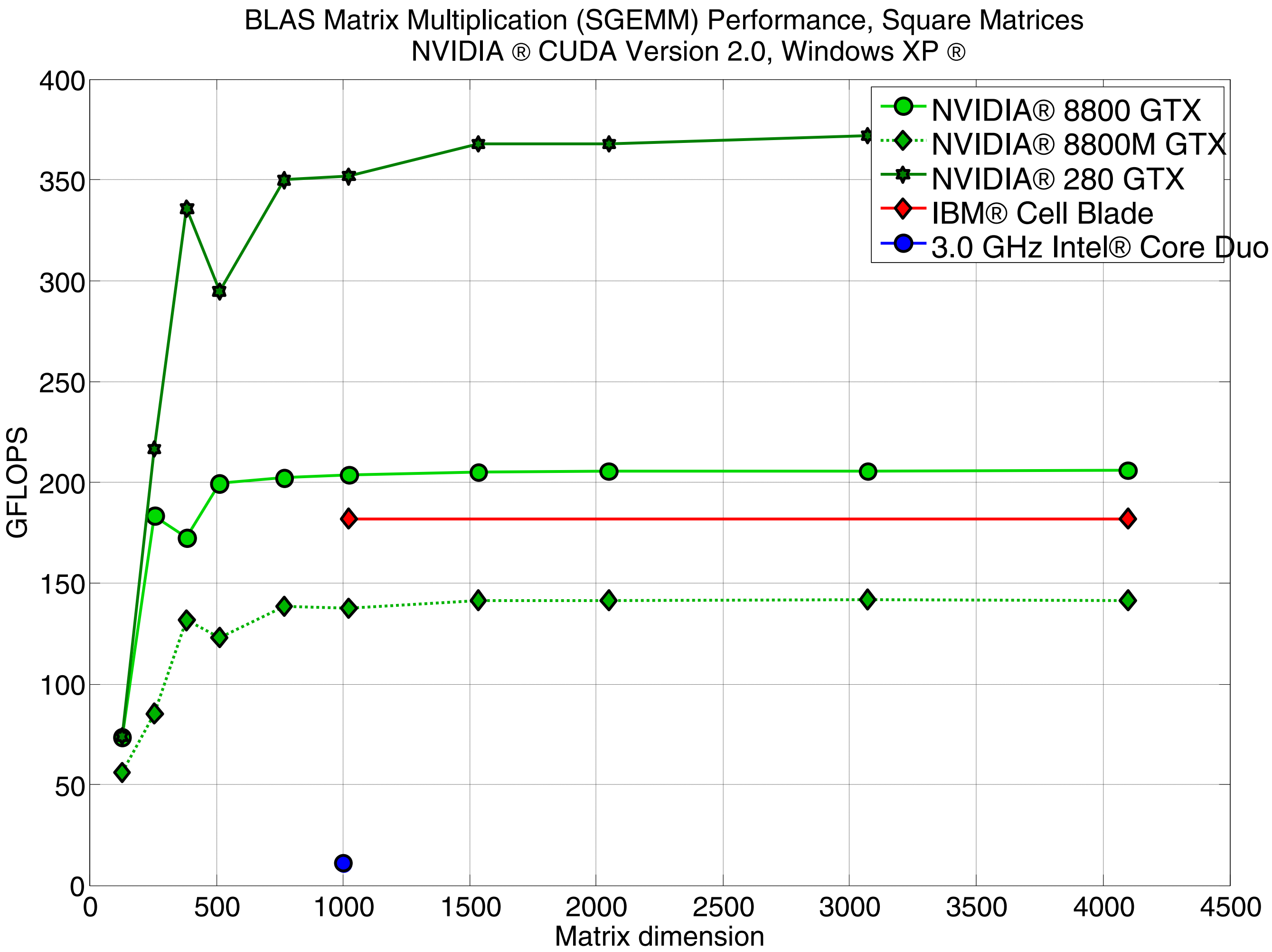
# IRSTAP Benchmarks GFLOP/s



# IRSTAP Benchmarks GFLOP/W



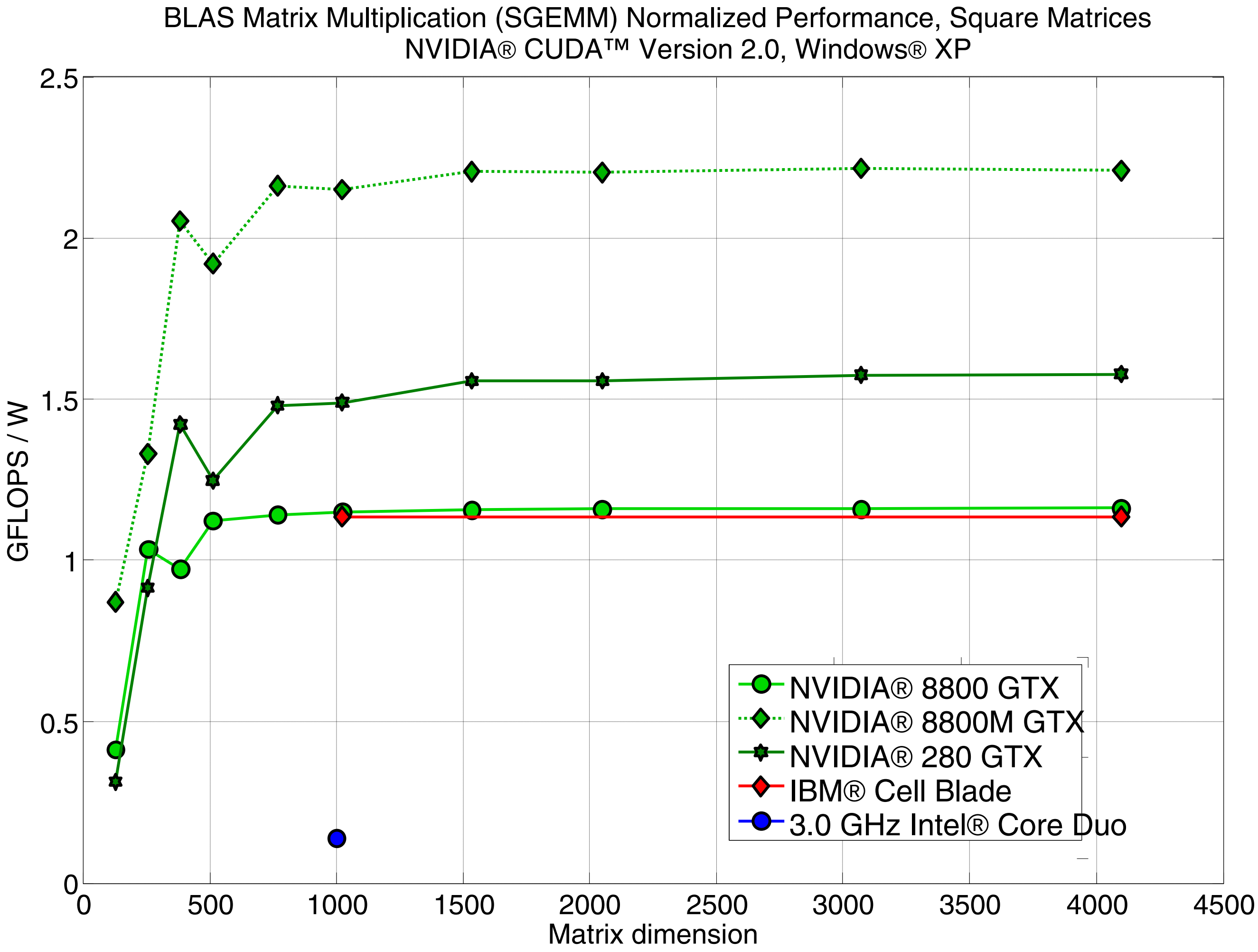
# Matrix Multiplication GFLOP/s



**GPU Multiplication is SGEMM is from the NVIDIA® library**

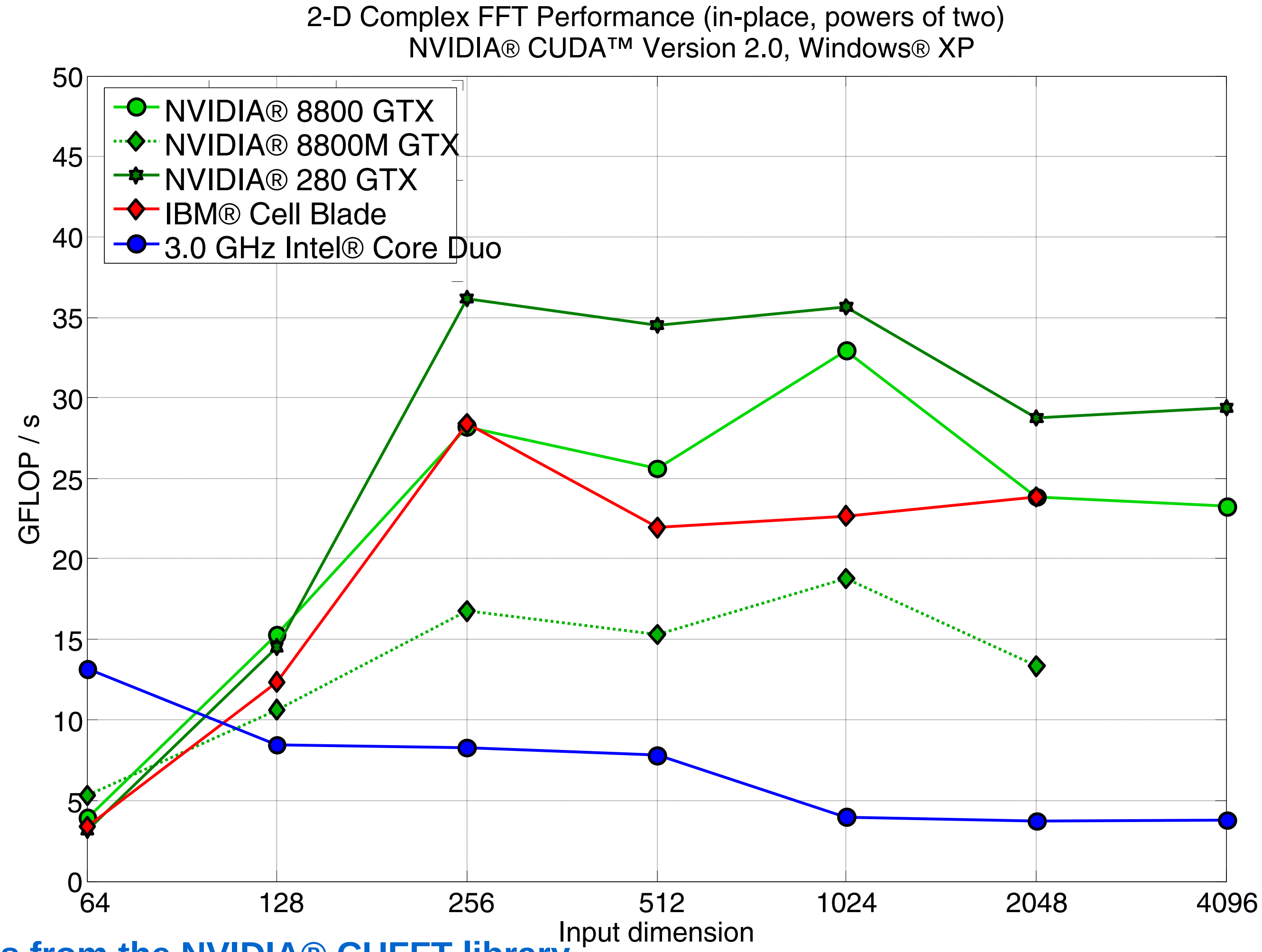
NVIDIA and CUDA are trademarks or registered trademarks of NVIDIA Corporation in the United States and/or other countries. Windows is a registered trademark of Microsoft Corporation in the United States and/or other countries. IBM is a registered trademark of International Business Machines Corporation in the United States and/or other countries. Intel is a registered trademark of Intel Corporation in the United States and/or other countries.

# Matrix Multiplication GFLOP/W



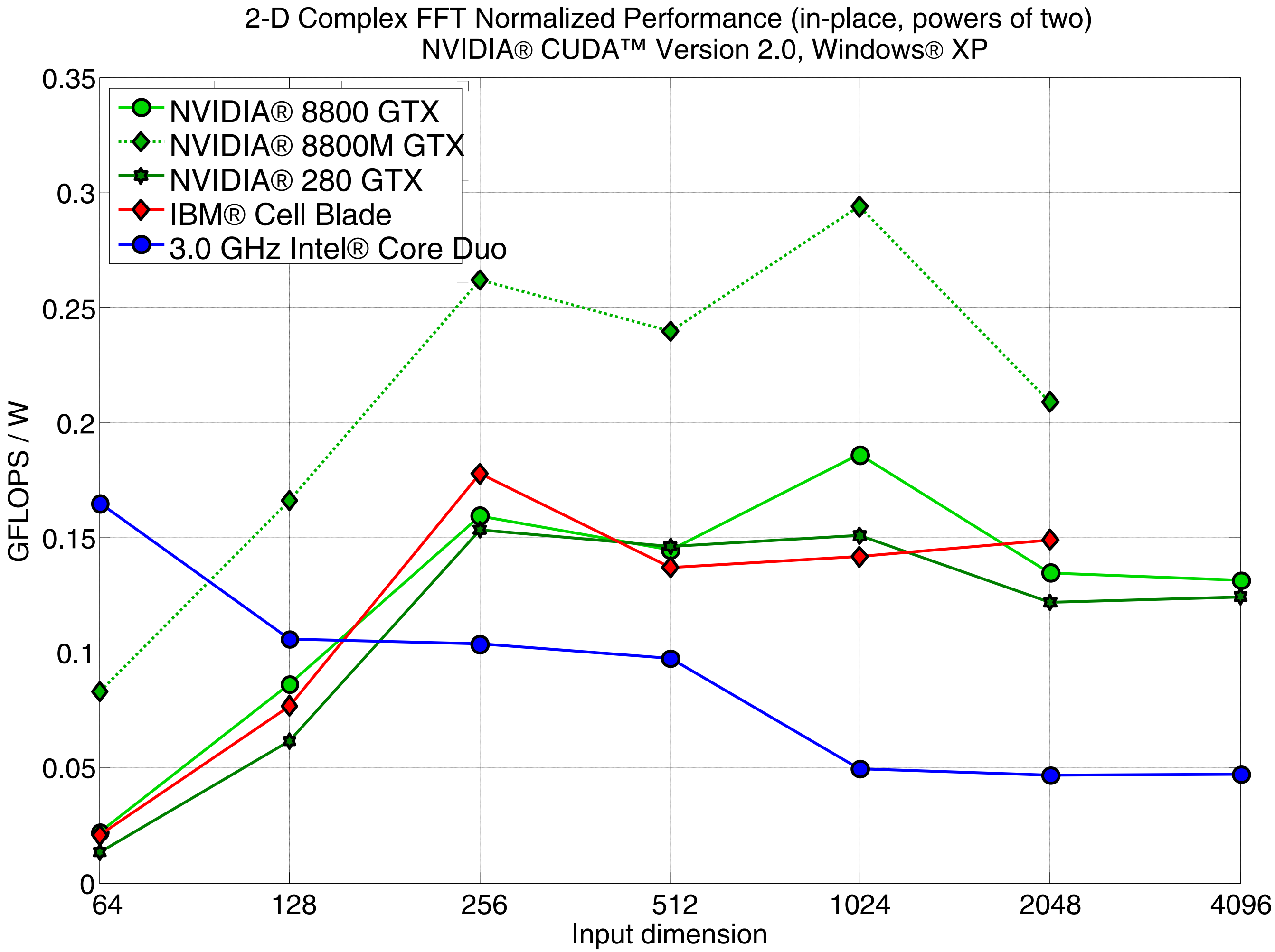
NVIDIA and CUDA are trademarks or registered trademarks of NVIDIA Corporation in the United States and/or other countries. Windows is a registered trademark of Microsoft Corporation in the United States and/or other countries. IBM is a registered trademark of International Business Machines Corporation in the United States and/or other countries. Intel is a registered trademark of Intel Corporation in the United States and/or other countries.

# 2D Complex FFT GFLOP/s



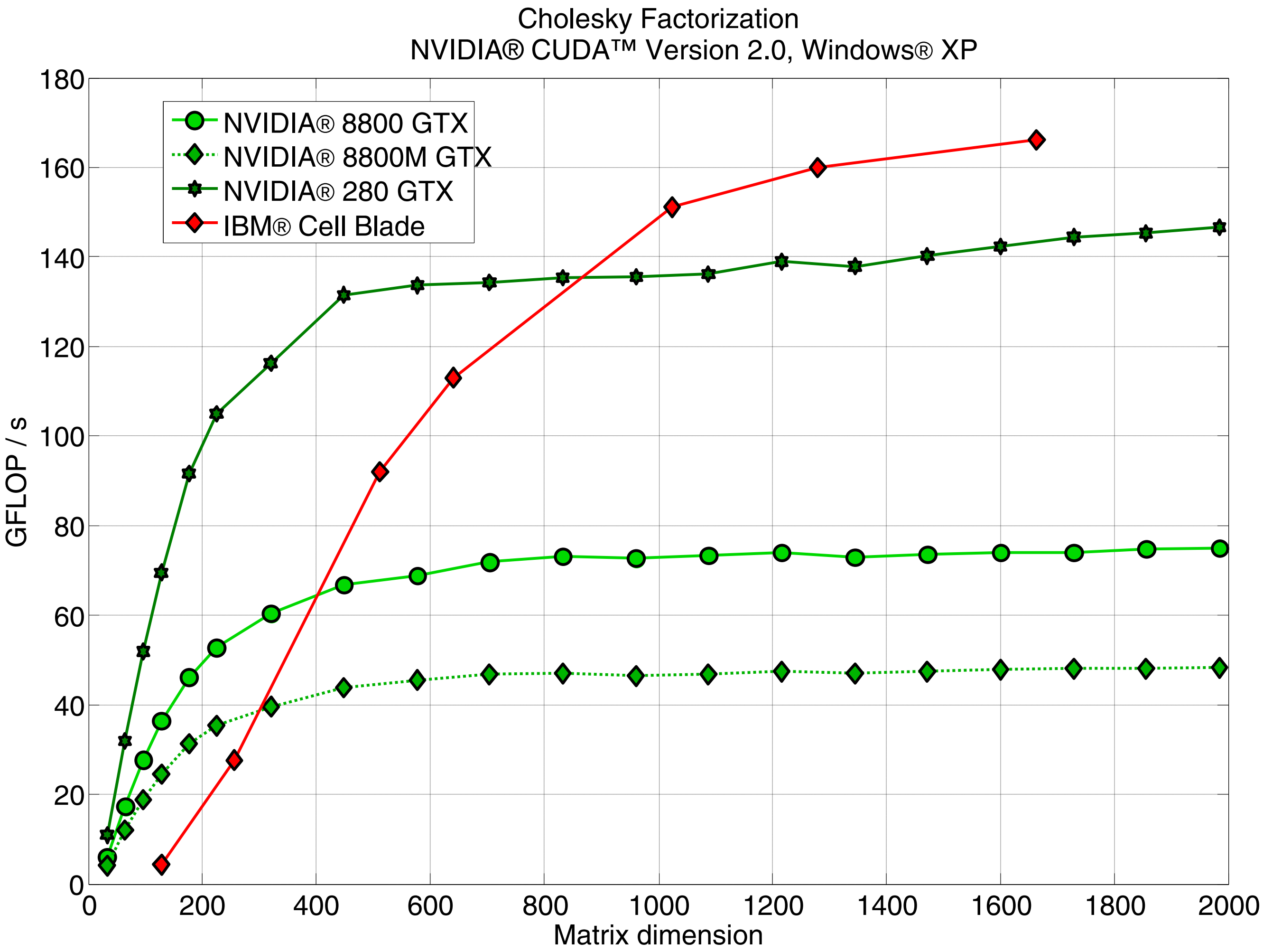
**GPU FFT is from the NVIDIA® CUFFT library**

# 2D Complex FFT GFLOP/W



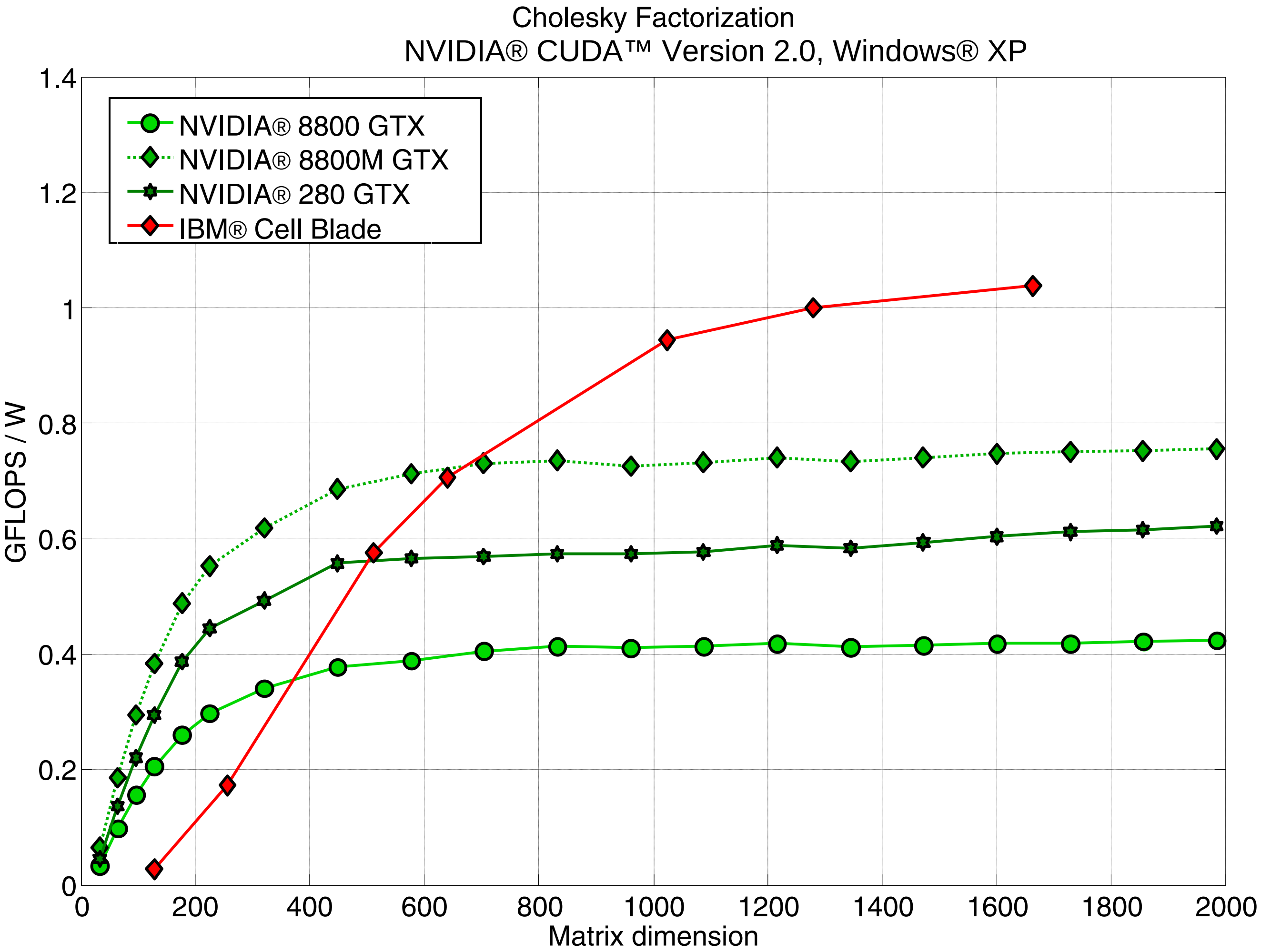
NVIDIA and CUDA are trademarks or registered trademarks of NVIDIA Corporation in the United States and/or other countries. Windows is a registered trademark of Microsoft Corporation in the United States and/or other countries. IBM is a registered trademark of International Business Machines Corporation in the United States and/or other countries. Intel is a registered trademark of Intel Corporation in the United States and/or other countries.

# Cholesky GFLOP/s



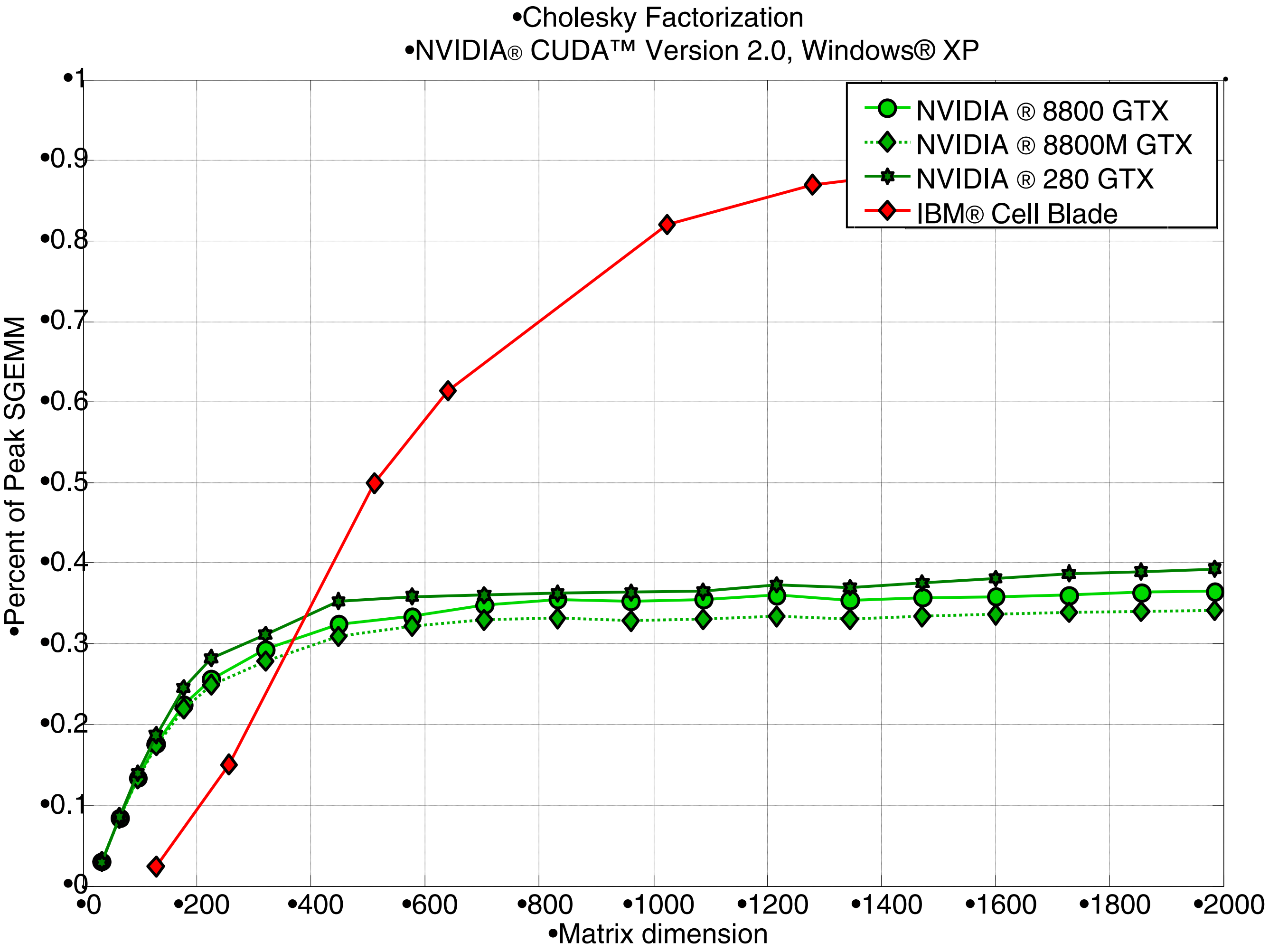
NVIDIA and CUDA are trademarks or registered trademarks of NVIDIA Corporation in the United States and/or other countries. Windows is a registered trademark of Microsoft Corporation in the United States and/or other countries. IBM is a registered trademark of International Business Machines Corporation in the United States and/or other countries.

# Cholesky GFLOP/W



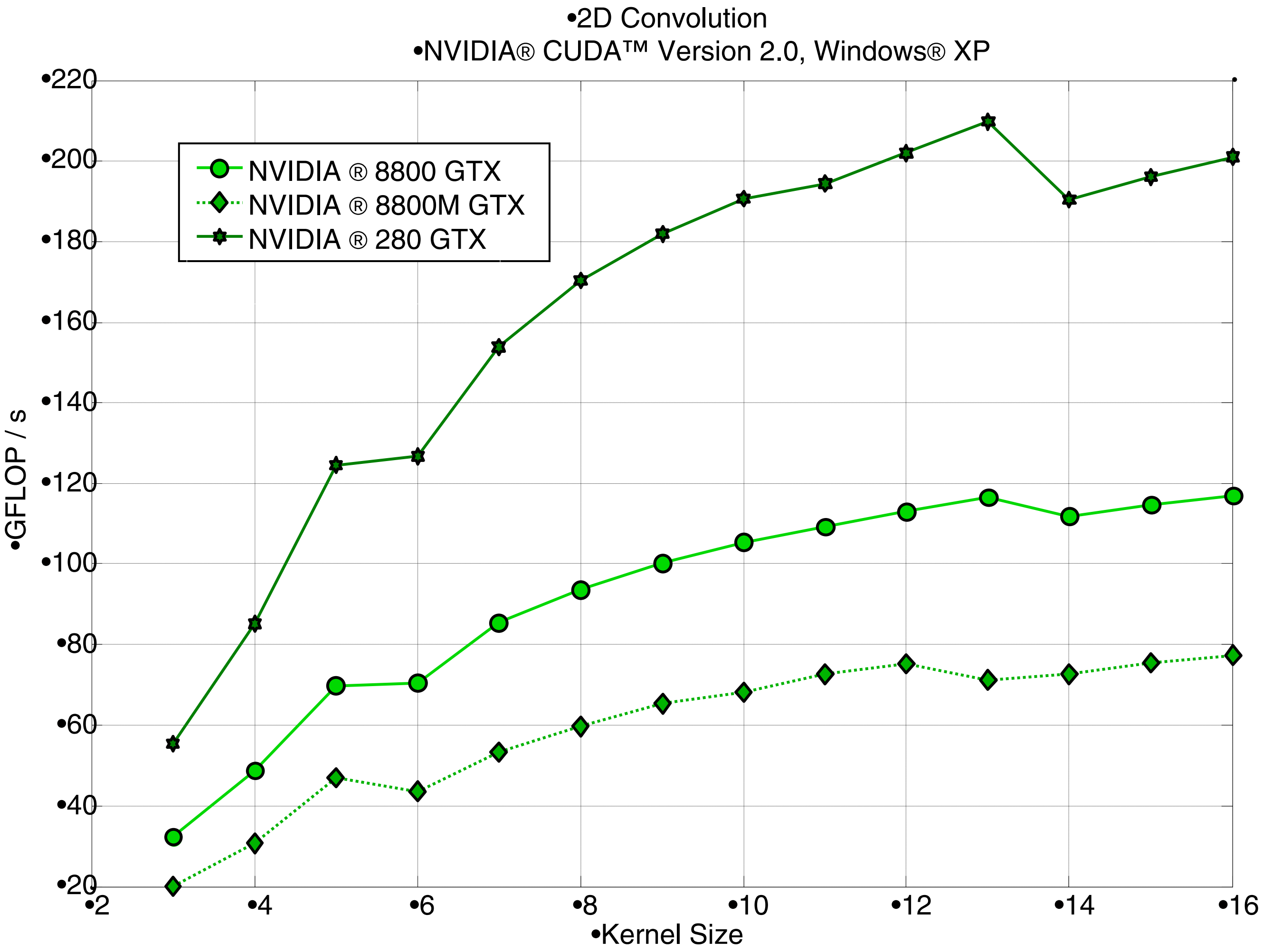
NVIDIA and CUDA are trademarks or registered trademarks of NVIDIA Corporation in the United States and/or other countries. Windows is a registered trademark of Microsoft Corporation in the United States and/or other countries. IBM is a registered trademark of International Business Machines Corporation in the United States and/or other countries.

# Cholesky Efficiency



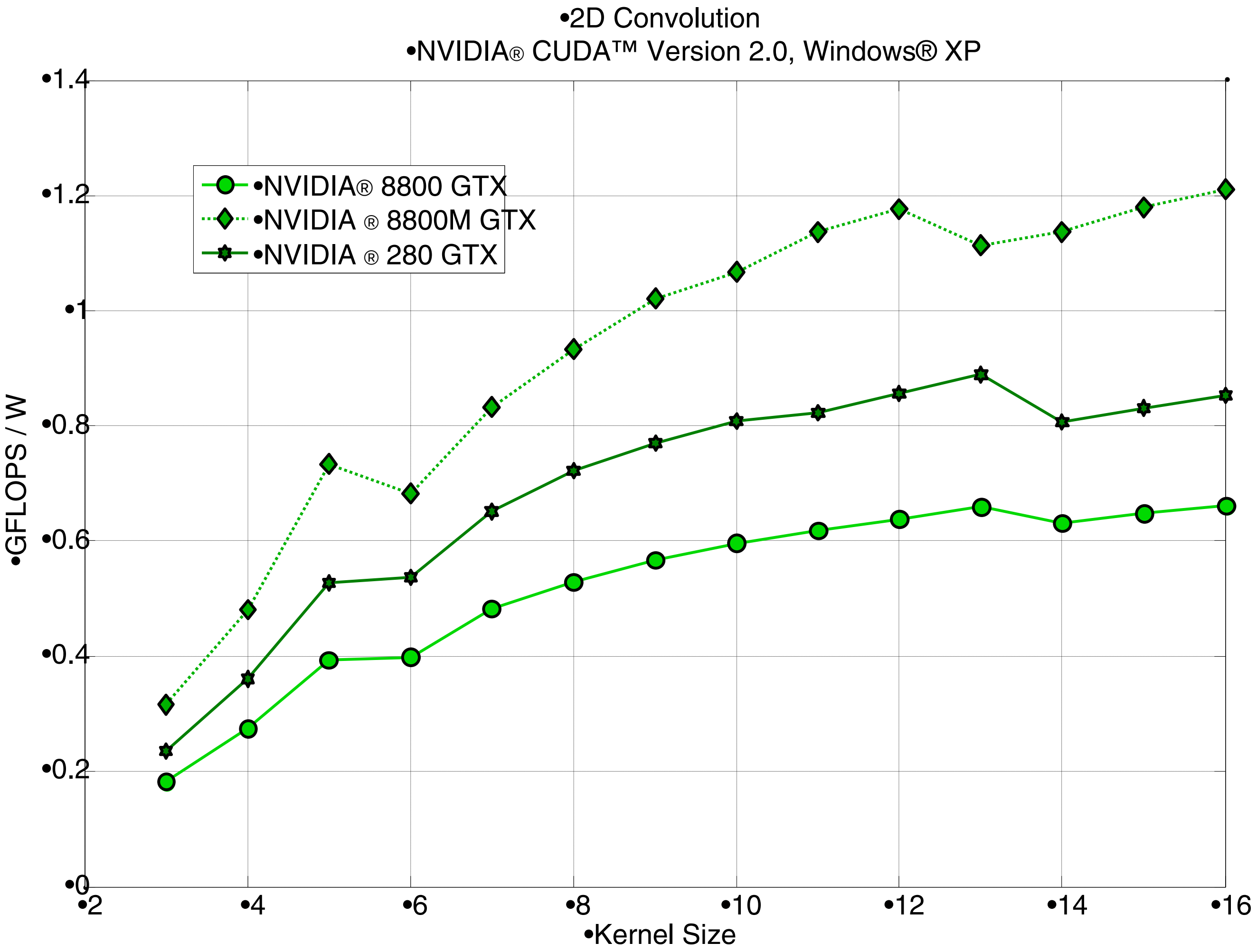
NVIDIA and CUDA are trademarks or registered trademarks of NVIDIA Corporation in the United States and/or other countries. Windows is a registered trademark of Microsoft Corporation in the United States and/or other countries. IBM is a registered trademark of International Business Machines Corporation in the United States and/or other countries.

# 2D Convolution GFLOP/s



NVIDIA and CUDA are trademarks or registered trademarks of NVIDIA Corporation in the United States and/or other countries. Windows is a registered trademark of Microsoft Corporation in the United States and/or other countries.

# 2D Convolution GFLOP/W



NVIDIA and CUDA are trademarks or registered trademarks of NVIDIA Corporation in the United States and/or other countries. Windows is a registered trademark of Microsoft Corporation in the United States and/or other countries.

# Power Estimation

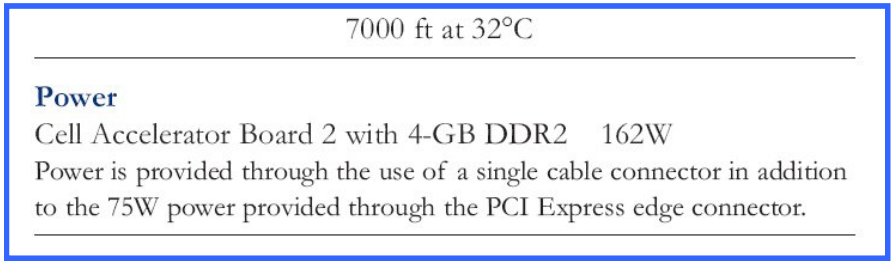
- **Using thermal design power (TDP) for each device**
  - NVDIA® 8800 GTX 128SP 1.35Ghz – TDP: 177W
  - NVDIA® 8800M GTX 96SP 1.25Ghz – TDP: 65W
  - NVDIA® 280 GTX 240SP 1.30Ghz – TDP: 236W
- **Cell – estimating TDP: 160W with 90nm process**

# Questions ?

# Just for comparison…

Assume Cell Processor power consumption is approximately 160 W

2 Cell Processors per blade

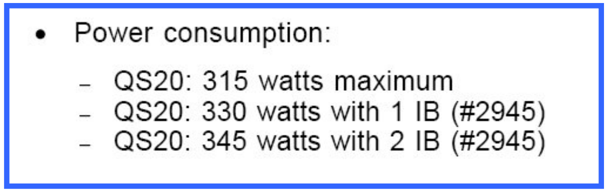


7000 ft at 32°C

**Power**

Cell Accelerator Board 2 with 4-GB DDR2 162W

Power is provided through the use of a single cable connector in addition to the 75W power provided through the PCI Express edge connector.



- Power consumption:
  - QS20: 315 watts maximum
  - QS20: 330 watts with 1 IB (#2945)
  - QS20: 345 watts with 2 IB (#2945)