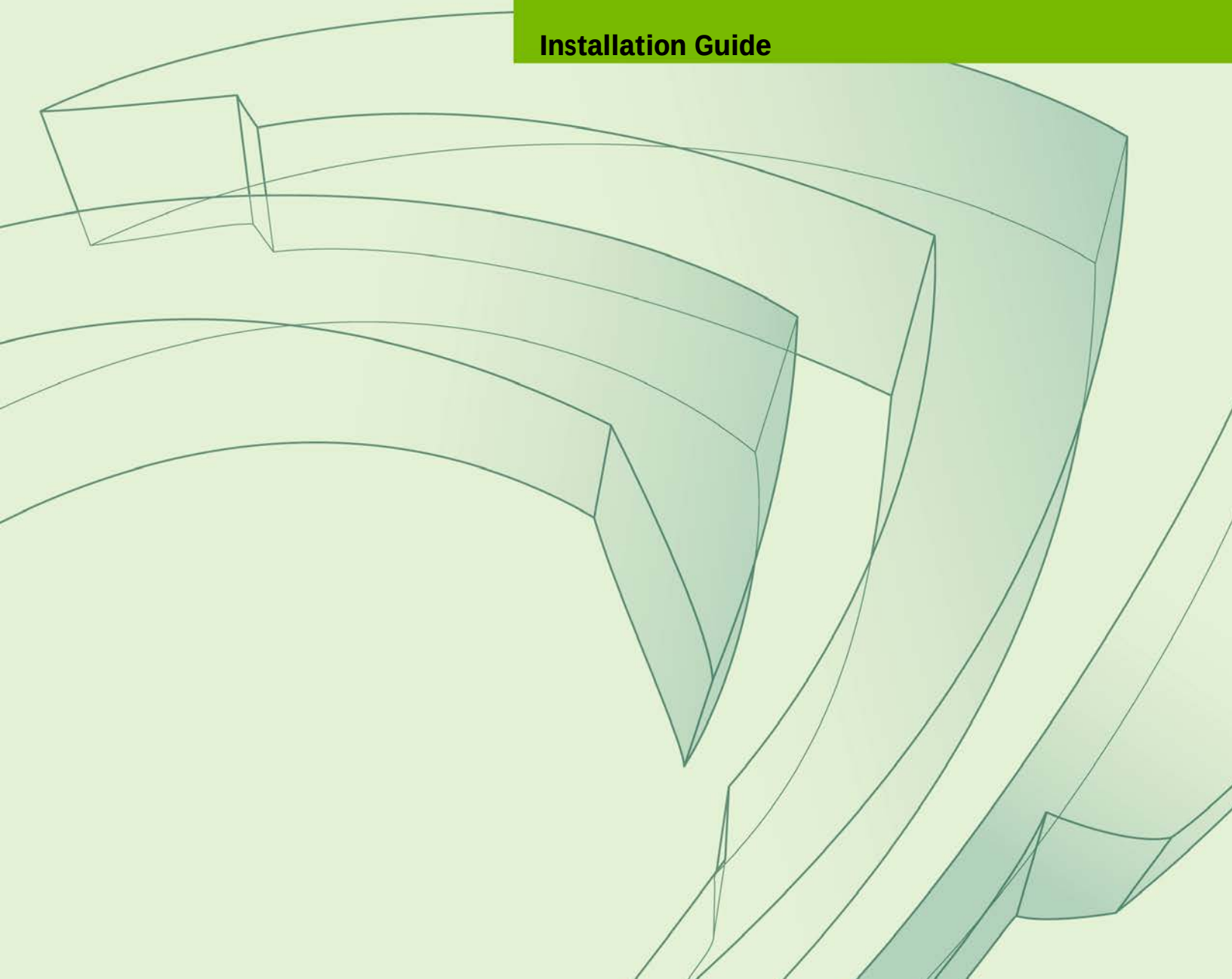
NVIDIA MAXIMUS SYSTEM BUILDER'S GUIDE FOR MICROSOFT WINDOWS 7 OR WINDOWS 8-64

DI-06471-001_v03 | January 2013

Installation Guide



DOCUMENT CHANGE HISTORY

DI-06471-001_v03

Version	Date	Authors	Description of Change
01	August 4, 2012	OO, TS	Initial Release
02	September 6, 2012	EK, TS	Minor Edits
03	January 22, 2013	EKK, SM	Updated to incorporate 2nd Generation Maximus

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

NVIDIA 2ND GENERATION MAXIMUS SYSTEM BUILDER'S GUIDE

CAUTION: NVIDIA® Maximus™ platforms are available only through select OEMs. For a list of OEMs qualified by NVIDIA, refer to the Maximus web page at http://www.nvidia.com/maximus.

This document explains how to install, configure, and verify the primary hardware and software components of an NVIDIA 2nd Generation Maximus platform for those who already have a platform from a qualified OEM (qualified OEMs listed at http://www.nvidia.com/maximus) and would like to upgrade their existing system.

This document *does not* cover vendor-specific graphics processing unit (GPU)-enabled applications. For vendor-specific details about GPU-enabled applications, refer to the documentation provided by the software vendor.

AUDIENCE

This document is intended for professionals who are capable of installing cards on a workstation and making any system upgrades that may be necessary.

PREREQUISITE SKILLS

To complete the tasks described in this document, you should have an understanding of the following concepts and mastery of the skills:

- Personal computer hardware, terminology, and power connectivity
- Personal computer thermals and cooling
- Hardware and software combination troubleshooting
- Software driver installation
- Microsoft Windows 7 and Windows 8 configuration and administration
- Microsoft Windows 7 and Windows 8 device manager and driver installation

CONTENTS

The sections in this document describe the required enabling components for a Maximus platform and how to enable Maximus technology and verify that your installation is correct.

The topics of the sections are as follows:

- "Required Components to Enable Maximus Technology" on Page 4 identifies the components that you need to enable a system for Maximus technology.
- "Implementing the Maximus Platform" on Page 6 describes the installation of the NVIDIA Quadro® and NVIDIA Tesla® cards, the Quadro® driver and the system enhancements that may be required to support a Maximus-enabled system.
- "Verifying System Installation" on Page 13 describes how to check that the Maximus technology is correctly installed and correctly recognized by the system.
- "Advanced Topics" on Page 17 briefly discusses more advanced Maximus configurations, tuning of applications that use Maximus and a link to a list of GPU-enabled applications that use Maximus technology.
- "Using the NVIDIA Maximus Configuration Utility" on Page 21 describes the Maximus Configuration Utility, which you can use to check what GPU cards are installed on a system and to tune your Maximus configuration.

BENEFITS OF 2ND GENERATION MAXIMUS TECHNOLOGY

For an overview of Maximus technology, its benefits, and how it is being used, go to http://www.nvidia.com/maximus.

INSTALLATION

This section identifies the components that are needed to enable a system for 2nd Generation Maximus technology. It also describes the installation of the Quadro and Tesla cards, the Quadro driver, and the system-enhancements that may be required to support a Maximus-enabled system and provides a procedure to verify your installation.

REQUIRED COMPONENTS TO ENABLE MAXIMUS TECHNOLOGY

Following is a list of the critical components needed for an NVIDIA Kepler™ based Maximus-enabled workstation:

- At least one NVIDIA Quadro K600, Quadro K2000, Quadro K4000, Quadro K5000, or Quadro K6000 graphics card (for computation and graphics). Refer to “Quadro Professional Graphics” on Page 5.
- At least one NVIDIA Tesla K20 card (for computation). Refer to “Tesla K20” on Page 5.
- A compatible professional NVIDIA software driver. Refer to “NVIDIA Quadro Professional Graphics Driver” on Page 6
- Microsoft Windows 7 or Windows 8—64-bit
- A host chassis from a qualified OEM with space for the Quadro and Tesla cards and the capacity to support the power demands.

The Maximus technology components are described in the following sections. System enhancements needed to support a Maximus-enabled workstation are described in “Implementing the Maximus Platform” on Page 6.

Quadro Professional Graphics

The NVIDIA Quadro family of professional graphics solutions provides the visualization component of a Maximus-enabled workstation. A Quadro graphics card can also perform GPU computational processing, therefore providing another alternative for users to best utilize their system in demanding workflows. NVIDIA Maximus platform supports the following NVIDIA Quadro graphics cards (Figure 1):

- Quadro K600
- Quadro K2000
- Quadro K4000
- Quadro K5000
- Quadro K6000 (offered later in 2013)

Figure 1. NVIDIA Quadro Family for 2nd Generation Maximus

Note: Some retail Quadro cards may look different than those shown in the illustration. However, these cards are designed, built and supported by NVIDIA and have exactly the same functionality.

Tesla K20

An NVIDIA Maximus platform requires the Tesla K20 card. This is the critical component of a 2nd Generation Maximus platform.

The Tesla K20 is the latest addition to the Tesla series of compute companion processors, which are based on the Kepler architecture. Tesla K20 features 2,496 NVIDIA CUDA® cores and 5 GB of high-speed GDDR5 memory (Figure 2).

Figure 2. NVIDIA Tesla K20

NVIDIA Quadro Professional Graphics Driver

The NVIDIA Quadro Professional Driver is another critical component. Support for both the Quadro and Tesla cards are unified in this one NVIDIA driver.

The NVIDIA Quadro Professional driver enables runtime GPU processing tasks to operate among multiple GPUs in a system. The driver serves as the interface between professional GPU-enabled applications and the GPU hardware itself. The minimum driver which should be used is Version 307.45.

Implementing the Maximus Platform

As stated in the preceding section, a 2nd Generation Maximus-enabled platform is a combination of an NVIDIA Quadro card, Tesla card, and software driver technology. To host this combination of GPUs, a capable chassis is required along with whatever other workflow-specific components are necessary. The recommendations in this section serve only as a guide and can cover only the components noted here. It is the system builder's responsibility to verify that proper power, cooling, thermals, and connectivity requirements are met for the total system.

! **CAUTION:** DISCONNECT ALL MAIN POWER BEFORE PERFORMING ANY SYSTEM HARDWARE WORK.

Requirements for the Host Chassis

You must ensure that the host chassis satisfies the requirements to host your particular configuration of Quadro and Tesla cards. For example, the host chassis must have sufficient physical space for each card and satisfy the cooling and power requirements for your card configuration.

Table 1 outlines the specifications of each supported Tesla and Quadro GPU.

Table 1. Tesla and Quadro GPU Specifications

Model	Onboard Memory	PCIe Slot Width	Maximum Power Consumption	PCIe Slot Type	Auxiliary Power Requirements
Tesla K20	5 GB	Double	225 W	×16 Gen2	Yes
Quadro K5000	4 GB	Double	122 W	×16 Gen2	Yes
Quadro K4000	3 GB	Single	80 W	×16 Gen2	Yes
Quadro K2000	2 GB	Single	51 W	×16 Gen2	No
Quadro K600	1 GB	Single	41 W	×16 Gen2	No

A Maximus-enabled workstation contains at least one NVIDIA Quadro card and at least one NVIDIA Tesla K20 card. A typical configuration is one Quadro K5000 and one Tesla K20 with 5 GB of memory on each board. This configuration provides maximum memory and application flexibility.

Using ×16 Electrical PCI Express Slots

Peripheral Component Internet Express (PCIe) is a data and physical interconnect standard for attaching peripherals (such as graphics cards or other I/O devices) to the main data bus of a host computer system board. The Quadro and Tesla cards conform to the PCIe standard and are specifically built for PCIe ×16 electrical slots.

There are several types of PCIe slots in modern workstations. Typically, high-end workstations from major OEMs such as HP, Dell, and Lenovo feature two or three PCIe ×16 slots. Some specialized workstations may contain more PCIe ×16 slots for larger configurations.

Slots on a workstation system board may appear physically similar, but can be electrically different in that the pin slots may not be connected to the data bus. Consult the documentation for your particular system board to identify the ×16 electrical PCIe slots. For example, use slots that are labeled **SLOT2 PCIe2 ×16 (×16).** Such slots are ×16 mechanical and ×16 electrical.

Try to avoid using slots that are labeled **SLOT4 PCIe2 ×16 (×8)** for example. Such slots are ×16 mechanical and ×8 electrical. Due to reduced bandwidth on these slots, you may not be able to achieve full performance for your application.

The PCIe ×16 slot often has a single spring-like plastic tab (commonly green) at one end. Take care when using these tabs because they can break easily. Typically, PCIe slots are labeled on the system board for convenient identification. Figure 3 shows an HP Z820 workstation system board; the PCIe ×16 slots are shown outlined in red.

Figure 3. PCIe ×16 Slots in an HP Z820 Workstation

Figure 4 shows a typical Quadro and Tesla combination. The Quadro card in the first ×16 electrical PCIe slot should be installed in the slot nearest to the system board I/O connections (this card drives the displays for visualization). The Tesla board should be installed in the next available ×16 electrical PCIe slot nearest the Quadro card.

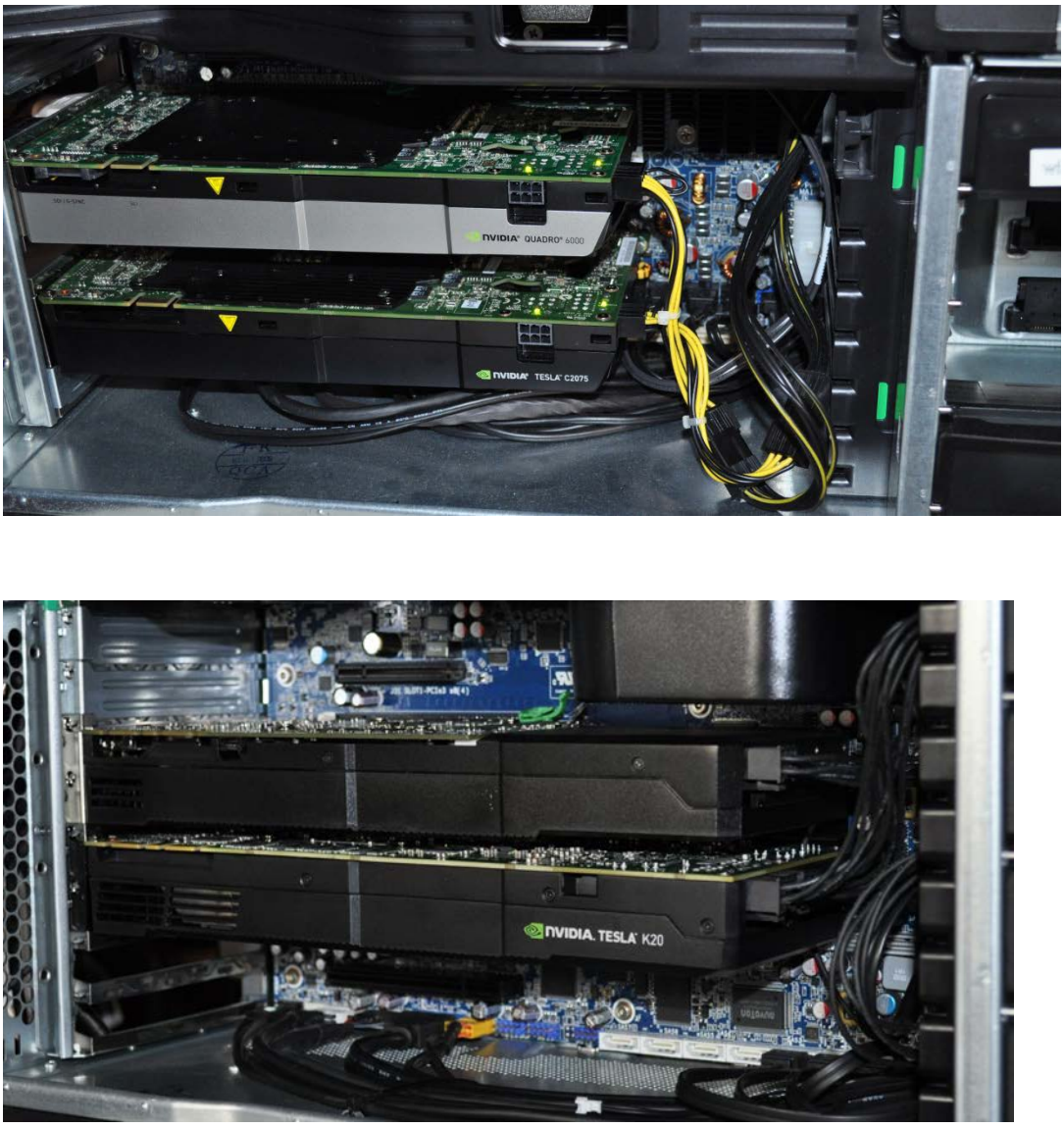


Figure 4. Typical Quadro and Tesla Combinations

REQUIREMENTS FOR HOST MEMORY

For best results, ensure that your Maximus-enabled system has approximately three times as much system memory as aggregate onboard GPU memory. For example; if a system has one Quadro K5000 (4 GB) and one Tesla K20 (5 GB) installed, the host system requires 27 GB (3 × 9 GB = 27 GB) of system memory to accommodate GPU operations. Note that this recommendation is intended for general guidance and will change by application; for system configurations recommended by NVIDIA, go to http://nvidia.com/maximus.

REQUIREMENTS FOR POWER SUPPLIES AND AUXILIARY PCI EXPRESS CABLING

The NVIDIA Tesla card and many of the Quadro cards require auxiliary power within the host chassis. A suitable power supply is necessary to maintain system integrity under computational load. In the event the power supply does not provide enough 6-pin connectors, an 8-pin connector may be used with an 8 to 6-pin adapter. Ensure that your system can deliver the necessary wattage and auxiliary power connectors for all cards in the system. If the supplied power is not adequate, the cards will not function properly. Table 2 specifies the power requirements and power connector options for the NVIDIA Tesla and Quadro cards.

Table 2. Power Consumption and Auxiliary Connector Requirements

Model	Maximum Power Consumption	Required Auxiliary PCIe Power Connectors
Tesla K20	225 W	2/ea. 6-pin
Quadro K5000	122 W	1/ea. 6-pin
Quadro K4000	80 W	1/ea. 6-pin
Quadro K2000	51.1 W	N/A
Quadro K600	41 W	N/A

Power adapters can be obtained to connect 8-pin PCIe chassis power cables to 6-pin device receptacles from several online retailers. Figure 5 shows one type of 8-pin-to-6-pin PCIe power adapter.

Figure 5. 8-Pin to 6-Pin PCIe Power Adapter

Note: If the auxiliary power cabling is not correct, the following message will display at boot time: Power down and connect the PCIe power cable(s) for this graphics card.

CONNECTING DISPLAY MONITORS

CAUTION: IN A MAXIMUS SYSTEM, DO NOT CONNECT A MONITOR TO A TESLA CARD.

In a Maximus-enabled system, connect a monitor to the *Quadro* card only. NVIDIA Quadro graphics cards provide several options for display connections.

In scalable visualization configurations (for an overview go to http://www.nvidia.com/object/quadro_svs_overview.html), dedicate Quadro cards to graphics and Tesla cards to computation. Intermixing graphics and compute tasks may yield undesirable frame rate and display synchronization artifacts.

USING SCALABLE LINK INTERFACE (SLI) RIBBON CABLES

CAUTION: DO NOT CONNECT A QUADRO CARD AND A TESLA CARD WITH AN NVIDIA SLI® RIBBON CABLE.

SLI ribbon cables are not required to enable a system for multiple GPU computational operations. In general, use SLI ribbon cables to connect *two Quadro* graphics cards for scalable visualization or other frame-rate sensitive applications.

QUADRO DRIVER

Prerequisites

- Make sure all applications are closed before installing or updating the driver.
- Ensure the Quadro and Tesla cards are already installed.

Installation

Install the NVIDIA Professional Quadro Driver 307.45 or greater. This is the recommended mainstream driver for enabling Maximus technology. You can download this driver from the NVIDIA Web site at http://www.nvidia.com.

CAUTION: YOU CANNOT USE A TESLA DRIVER TO ENABLE A MAXIMUS ENABLED WORKSTATION.

For specific applications, it is recommended to use ISV-certified drivers. Go to http://www.nvidia.com/page/partner_certified_drivers.html for a list of drivers.

Use the following procedure to install the graphics driver:

1. Download the NVIDIA Professional Quadro Driver 307.45 or greater from the NVIDIA Web site at http://www.nvidia.com and run the installer package
2. Accept the license agreement.
3. Select the **Custom (Advanced)** radio button under Installation options.
4. Click **NEXT** to begin installation.

5. Select **Perform a clean installation** checkbox on the Custom installation menu.

6. Click NEXT.

The screen(s) may go blank, flash, and/or show varying resolutions as the display driver software is unloaded and updated. This is normal as the display driver, operating system, and monitor(s) negotiate screen modes. It may be necessary to reboot the system.

Verifying System Installation

Use the following procedure to verify the system installation:

1. Ensure all Quadro and Tesla cards are properly powered. For more information, refer "Checking Power Connections" on Page 14.

   Hint: If the cards have green LEDs, they should be illuminated. If the Quadro cards do not have LEDs illuminated, check your display for system messages.

2. Check that the correct driver is installed. For more information, refer to "Checking the Quadro Driver Version" on Page 15.

   Hint: Go to NVIDIA Control Panel→System Information to check the driver version. In a standard configuration, a Quadro card is shown.

3. Check to make sure Microsoft Windows 7 or Windows 8 recognizes the Quadro and Tesla cards as display adapters. For more information, refer to "Check GPU Recognition and Quadro Driver Installation" on Page 16.

   Hint: Check for alerts; there should be none.

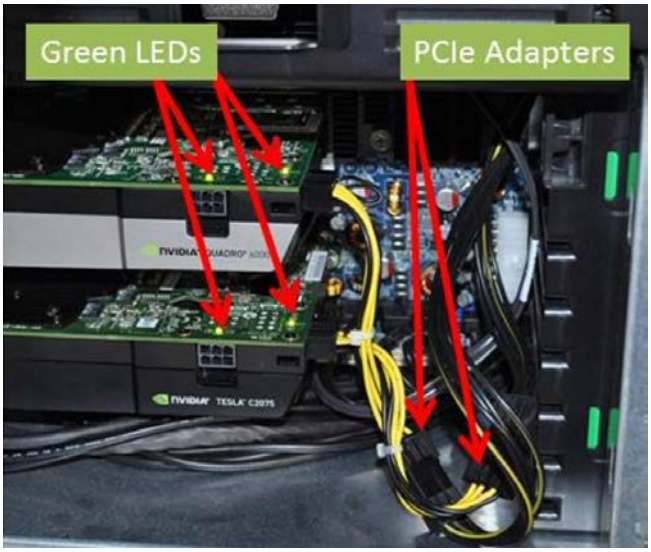
4. Check that all GPUs are physically installed in the system from the Maximus Configuration Utility (MCU). For more information, refer to "Checking Driver Recognition of Quadro and Tesla Cards" on Page 17.

   Hint: All GPUs installed on the system should be listed in the MCU window.

Checking Power Connections

To ensure all Quadro and Tesla cards are properly powered, perform the following system checks:

1. Power on the system.
2. Check to make sure all the cards have green LEDs illuminated (if applicable).



3. Check your display for system messages if the Quadro cards do not have LEDs illuminated,.

   Example: The displayed message **`PLEASE POWER DOWN AND CONNECT THE PCIe POWER CABLE(S) FOR THIS GRAPHICS CARD`** indicates that the correct power was not provided for an NVIDIA card. Check the power connections to all NVIDIA cards.

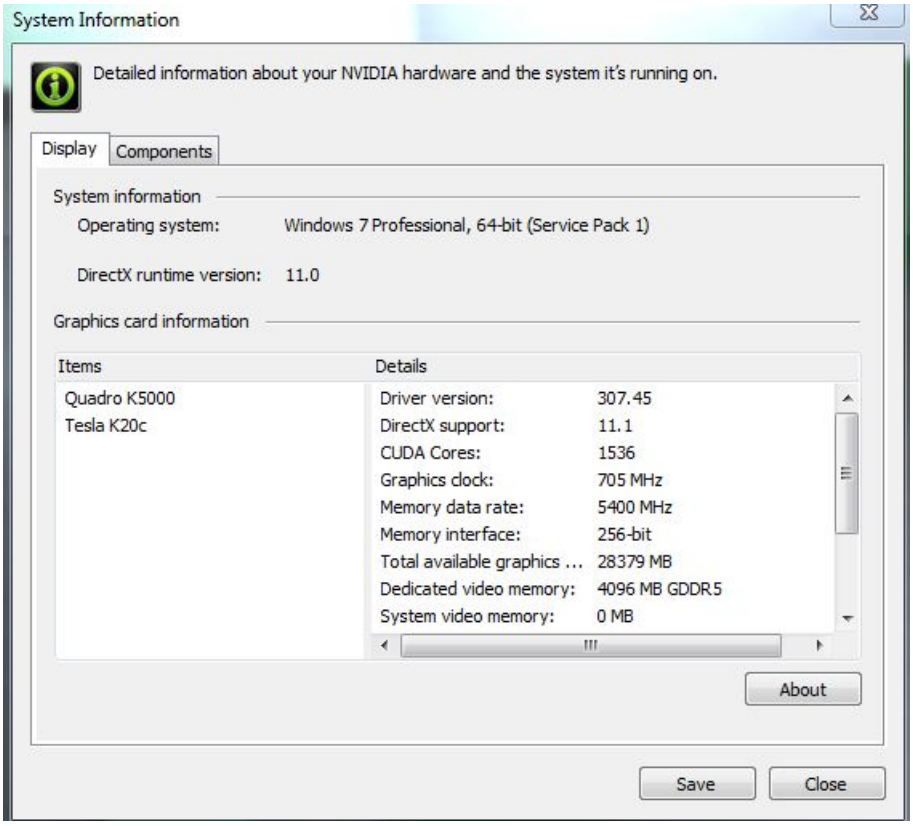
Note: Do NOT use split power connectors as the proper wattage may not be supplied.

4. Check that all power cables are connected at both ends (card side and host-system side) if the green LEDs are not illuminated or a system message is displayed.

Checking the Quadro Driver Version

Use the following procedure to check the NVIDIA Quadro driver version displayed in the NVIDIA Control Panel:

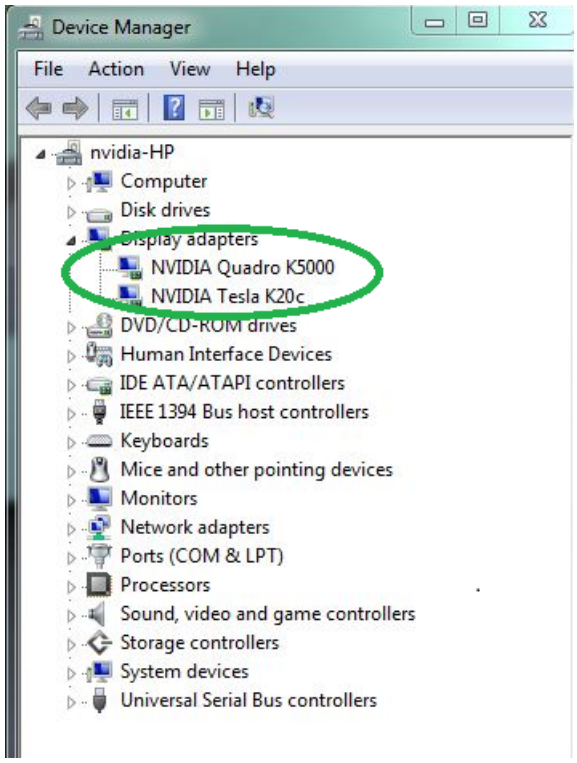
1. Open the System Information menu from the NVIDIA Control Panel.
2. Check the driver version on the System Information menu. If the driver version is older than 307.45, reinstall the driver (refer to "Installation" on Page 12).



Check GPU Recognition and Quadro Driver Installation

Use the following procedure to check that the Quadro and Tesla cards are recognized by the operating system and the Quadro driver is correctly installed:

1. Open the Windows 7 or Windows 8 Device Manager and check that the Quadro and Tesla cards are recognized as display adapters.



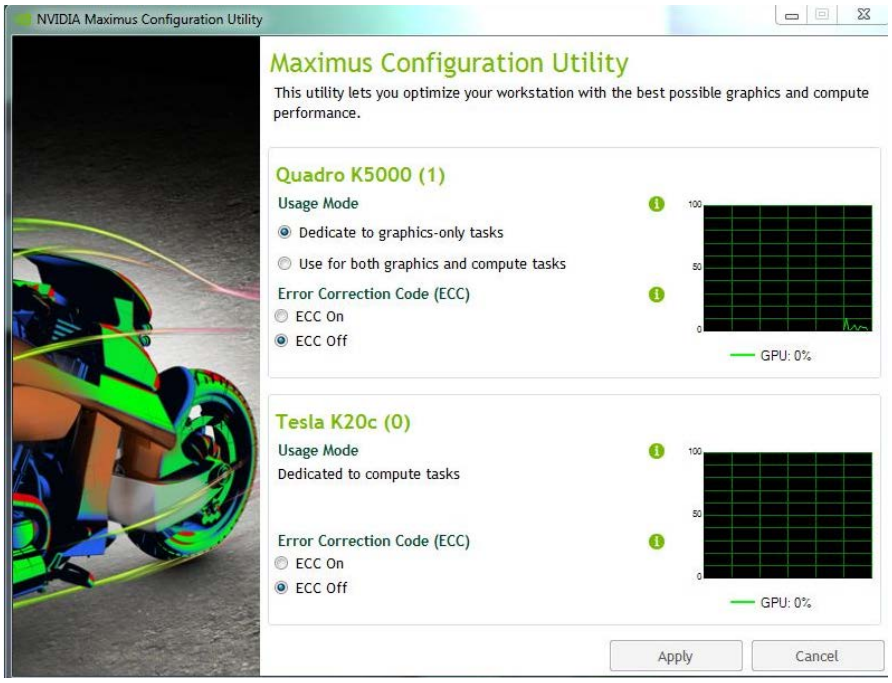
2. Check the driver versions, chassis slots, and power connections to the Quadro and Tesla cards if alerts are shown.
3. Check that the Quadro driver is correctly installed.

Note: Microsoft Windows Update will not necessarily provide the proper NVIDIA Quadro driver version. Do not use the Microsoft Windows Update to upgrade your Quadro driver.

Checking Driver Recognition of Quadro and Tesla Cards

Use the Maximus Configuration Utility (MCU) to verify that all the GPUs that are physically installed in your system are listed in the MCU dialog window. The MCU is described on Page 21.

1. Open the **NVIDIA Maximus Configuration Utility**.



2. Check that all Quadro and Tesla cards installed in your system are listed in the dialog window.

ADVANCED TOPICS

This sub-section contains information about advanced Maximus configurations and advanced tuning of the GPUs in a Maximus system.

Advanced Maximus Configuration

Note: The upper limit of GPUs that can be installed in a single workstation is often limited by the physical space and power consumption.

It is possible to build Maximus-enabled systems with multiple Quadro cards and multiple Tesla cards. Some OEMs offer them as standard configurations. The main

principles still apply to system implementation, verification, and configuration, as described in this document.

Figure 6 shows a workstation configuration with one Quadro card and two Tesla cards, where the Quadro card is installed in the PCI Express ×16 mechanical, ×8 electrical slot of the workstation. This configuration was selected because the workflow and application requires less graphics visualization and more numerical computation.

Figure 6. Workstation Configured with One Quadro and Two Tesla Cards

The workflow and application requires less graphics visualization and more numerical computation. When the Maximus Configuration Utility is run on this system, all Quadro and Tesla cards in the system will be listed (Figure 7).

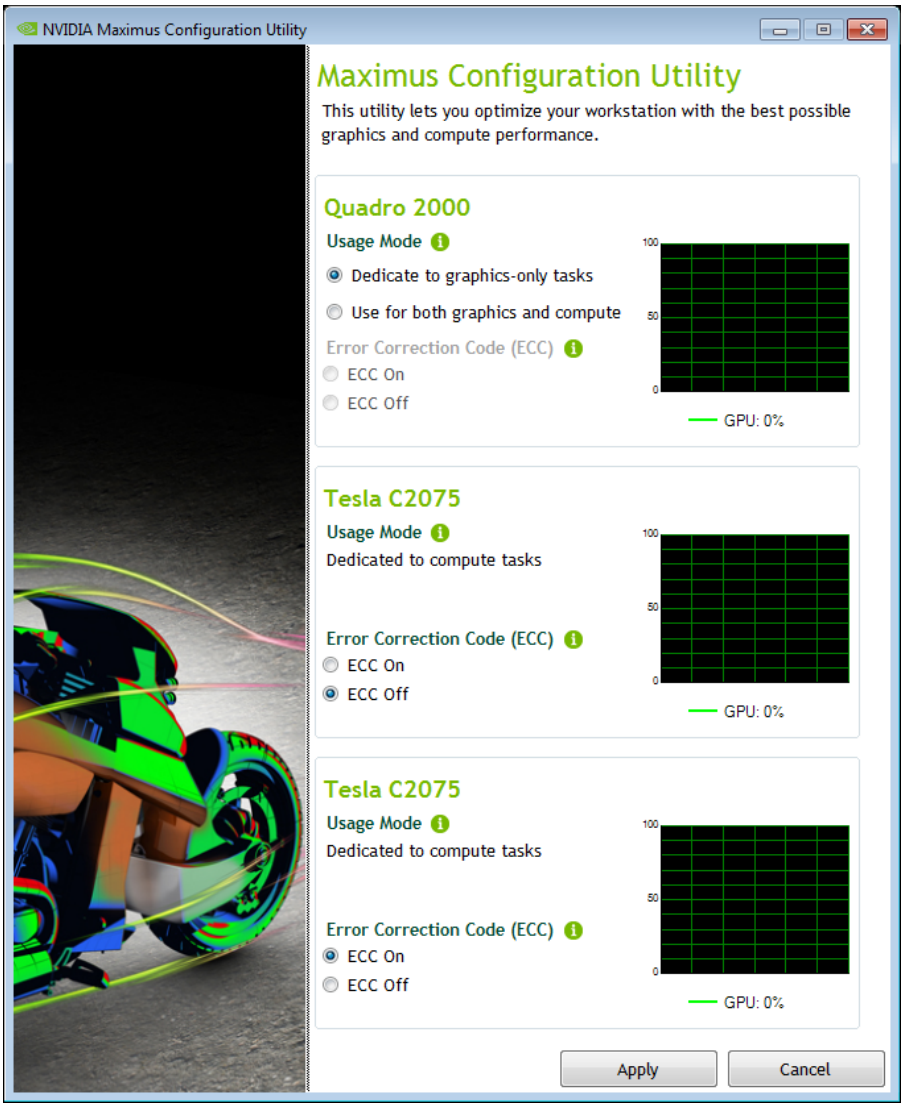


Figure 7. Maximus Configuration Utility Screen

Tuning Applications

A wide variety of GPU-enabled applications can leverage the power of an NVIDIA 2nd Generation Maximus-enabled workstation. There are many different methods by which software vendors can author their applications to leverage the power. The methods are beyond the scope of this document; however, there are tuning adjustments that can be made to a Maximus-enabled system to accommodate certain software applications. Consult the software vendors if there are questions regarding specific tuning needs or contact your NVIDIA representative.

Note: The MCU is a useful tool to monitor GPU activity.

GPU-Enabled Applications

The availability of GPU-enabled applications is continuously expanding. Many software vendors are making use of not only one, but multiple GPUs in a system. These applications enable greater productivity and return on investment for users and their organizations. The benefits, which can be substantial, are dependent on workflow and organization.

A list of GPU-enabled applications are listed on the following NVIDIA Web sites:

- http://www.nvidia.com/object/gpu-accelerated-applications.html
- http://www.nvidia.com/object/vertical_solutions.html

USING THE NVIDIA MAXIMUS CONFIGURATION UTILITY

There are several third-party GPU monitoring software tools available. These tools are not provided by NVIDIA and are not officially supported.

The NVIDIA Maximus Configuration Utility (MCU) is a separate graphical software utility that provides convenient GPU processing controls (Figure 8). The MCU provides GPU memory and utilization monitors for all supported GPUs in a Maximus-enabled system.

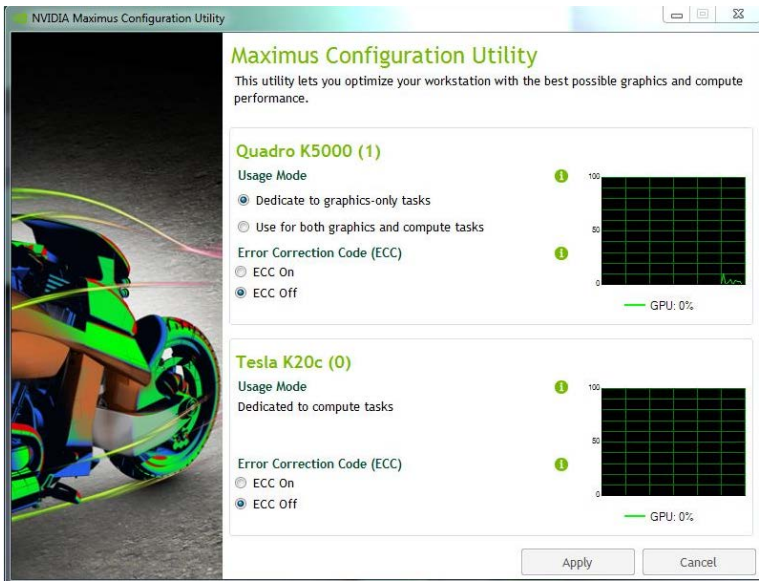


Figure 8. NVIDIA Maximus Configuration Utility Screen

Notice

Trademarks

Copyright