CST STUDIO SUITE® 2016

GPU Computing Guide

Contents

1 Nomenclature

The following section explains the nomenclature used in this document.

command   Commands you have to enter either on a command prompt (cmd on MS Windows or your favorite shell on Linux) are typeset using typewriter fonts.

<...>   Within commands the sections you should replace according to your environment are enclosed in "<...>". For example "<CST_DIR>" should be replaced by the directory where you have installed CST STUDIO SUITE (e.g. "c:\Program Files\CST STUDIO SUITE").

Those icons indicate that the following section of the text applies only to a certain operating system:

= MS Windows   = Linux

2   Supported Solvers and Features

- Transient Solver (T-solver/TLM-solver)
- Integral Equation Solver (direct solver and MLFMM only)
- Multilayer solver (M-solver)
- Particle-In-Cell (PIC-solver)
- Asymptotic Solver (A-solver)

Co-simulation with CST CABLE STUDIO is also supported.

2.1   Limitations

All solver except the T-solver require Tesla 20 (or newer) GPUs. The PIC-solver is currently limited to a single GPU.

2.2   Unsupported Features

The following features are currently not supported by GPU Computing. This list is subject to change in future releases or service packs of CST STUDIO SUITE.

Solver	Unsupported Features on GPU
Transient Solver	• Subgridding
Particle In Cell Solver	• Modulation of External Fields • Open Boundaries

3   Operating System Support

CST STUDIO SUITE is continuously tested on different operating systems. For a list of supported operating systems please refer to

<div align="center">

https://updates.cst.com/downloads/CST-OS-Support.pdf

</div>

In general, GPU computing can be used on any of the supported operating systems.

4   Supported Hardware

CST STUDIO SUITE currently supports up to 8 GPU devices in a single host system, meaning each number of GPUs between 1 and 8 is supported.[1]

---

[1]It is strongly recommended to contact CST before purchasing a system with more than four GPU cards to ensure that the hardware is working properly and is configured correctly for CST STUDIO SUITE.

The following tables contain some basic information about the GPU hardware currently supported by the GPU Computing feature of CST STUDIO SUITE, as well as the requirements for the host system equipped with the hardware. To ensure compatibility of GPU hardware and host system please check the list of qualified servers. Please note that a 64 bit computer architecture is required for GPU Computing.

A general hardware recommendation can be found in the FAQ section of the CST support website (FAQ No.3908).

List of supported GPU hardware for CST STUDIO SUITE 2016 [2] [3]

Card Name	Series	Platform	Min. CST Version
Tesla M60 [4]	Maxwell	Servers/Workst.	2016 SP 4
Tesla M40 [4]	Maxwell	Servers	2016 SP 4
Quadro M6000 24GB [4]	Maxwell	Workstations	2016 SP 4
Quadro M6000 [4]	Maxwell	Workstations	2015 SP 4
Tesla K80	Kepler	Servers	2014 SP 6
Tesla K40 m/c/s/st/d/t	Kepler	Servers/Workst.	2013 SP 5
Quadro K6000	Kepler	Workstations	2013 SP 4
Tesla K20X	Kepler	Servers	2013 release
Tesla K20m/K20c/K20s	Kepler	Servers/Workst.	2013 release
Tesla K10 [4]	Kepler	Servers	2013 release
Quadro 6000	Fermi	Workstations	2012 SP 6
Tesla Fermi M-Series	Tesla 20/Fermi	Servers	2011 SP 6
Tesla Fermi C-Series	Tesla 20/Fermi	Workstations	2011 SP 6
Tesla M1060[5]	Tesla 10	Servers	2010 SP 4
Tesla C1060[5]	Tesla 10	Workstations	2010 release
Quadro FX5800[5]	Tesla 10	Workstations	2010 release
Quadro Plex 2200D2[5]	Tesla 10	Workstations	2010 release

---

[2]Please note that cards of different series (e.g. "Kepler" and "Fermi") can't be combined in a single host system for GPU Computing.

[3]Platform = Servers: These GPUs are only available with a passive cooling system which only provides sufficient cooling if it's used in combination with additional fans. These fans are usually available for server chassis only!

Platform = Workstations: These GPUs provide active cooling, so they are suitable for workstation computer chassis as well.

[4] **Important:** The double precision performance of this GPU device is poor, thus, it can't be recommended for PIC-solver and I-solver (double precision) simulations.

[5] **Important:** This hardware is marked as deprecated and won't be supported in upcoming MWS versions (2017 and newer).

	Hardware Type	NVIDIA Tesla K20c/K20m/K20s (for Workst./Servers)	NVIDIA Tesla K20X (for Servers)
GPU Hardware Properties	**Min. CST version required**	**2013 release**	**2013 release**
	Number of GPUs	1	1
	Max. Problem Size (Transient Solver)	approx. 50 million mesh cells	approx. 60 million mesh cells
	Size	4.36"x10.5" 11.1cm x 26.67cm two slot width	4.36"x10.5" 11.1cm x 26.67cm two slot width
	Memory	5 GB GDDR5	6 GB GDDR5
	Power Consumption	225 W (max.) requires two auxiliary power connectors	235 W (max.)
Host System Requirements	PCI Express Requirements	1x PCIe Gen 2 (x16 electrically)	1x PCIe Gen 2 (x16 electrically)
	Power Supply of Host System [1]	min. 750 W	min. 750 W
	Min. RAM of Host System [2]	24 GB	24 GB
	Recommended Host System for the Use with the Hardware	Please ask your hardware vendor!	Please ask your hardware vendor!
	Recommended OS	RedHat EL 6/7 Windows 7/8 Win Server 2008 R2/2012 R2	RedHat EL 6/7 Win Server 2008 R2/2012 R2
	Vendor Information	www.nvidia.com	www.nvidia.com

[1]**Important:** The specifications shown assume that only one adapter is plugged into the machine. If you would like to plug in two or more adapters you will need a better power supply (1000W or above) as well as more RAM. Additionally, you need to provide sufficient cooling for the machine. Each Tesla card takes power from the PCI Express host bus as well as the 8-pin and the 6-pin PCI Express power connectors. This is an important consideration while selecting power supplies.

[2]The host system requires approximately 4 times as much memory as is available on the GPU cards. Although it is technically possible to use less memory than this recommendation, the simulation performance of larger models will suffer.

CST assumes no liability for any problems caused by this information.

	Hardware Type	NVIDIA Kepler K10[1] (for Servers)	NVIDIA Quadro K6000
GPU Hardware Properties	**Min. CST version required**	**2013 release**	**2013 SP 4**
	Number of GPUs	2	1
	Max. Problem Size (Transient Solver)	approx. 80 million mesh cells	approx. 120 million cells
	Size	4.376"x9.75" 11.11cm x 24.77cm two slot width	4.376"x10.5" 11.11cm x 26.67cm two slot width
	Memory	8 GB GDDR5	12 GB GDDR5
	Power Consumption	225 W (max.)	225 W (max.)
Host System Requirements	PCI Express Requirements	1x PCIe Gen 3 (x16 electrically)	1x PCIe Gen 3 (x16 electrically)
	Power Supply of Host System [2]	min. 750 W	min. 750 W
	Min. RAM of Host System [3]	32 GB	48 GB
	Recommended Host System	Please ask your hardware vendor!	Please ask your hardware vendor!
	Recommended OS	RedHat EL 6/7 Win Server 2008 R2/2012 R2	RedHat EL 6/7 Windows 7/8 Win Server 2008 R2/2012 R2
	Vendor Information	www.nvidia.com	www.nvidia.com

---

[1] **The double precision performance of this GPU device is poor, thus, it can't be recommended for PIC-solver, A-Solver and I-solver (double precision) simulations.**

[2] **Important:** The specifications shown assume that only one adapter is plugged into the machine. If you would like to plug in two or more adapters you will need a better power supply (1000W or above) as well as more RAM. Additionally, you need to provide a sufficient cooling for the machine. Each Tesla card takes power from the PCI Express host bus as well as the 8-pin and the 6-pin PCI Express power connectors. This is an important consideration while selecting power supplies.

[3] The host system requires approximately 4 times as much memory as is available on the GPU cards. Although it is technically possible to use less memory than this recommendation, the simulation performance of larger models will suffer.

CST assumes no liability for any problems caused by this information.

	Hardware Type	NVIDIA Tesla K40m/K40c (for Servers/Workst.)	NVIDIA Tesla K80 (for Servers)
GPU Hardware Properties	**Min. CST version required**	**2013 SP 5**	**2014 SP 6**
	Number of GPUs	1	2
	Max. Problem Size (Transient Solver)	approx. 120 million mesh cells	approx. 240 million mesh cells
	Size	4.376"x9.75" 11.11cm x 24.77cm two slot width	4.376"x9.75" 11.11cm x 24.77cm two slot width
	Memory	12 GB GDDR5	24 GB GDDR5
	Power Consumption	225 W (max.)	300 W (max.)
Host System Requirements	PCI Express Requirements	1x PCIe Gen 3 (x16 electrically)	1x PCIe Gen 3 (x16 electrically)
	Power Supply of Host System	min. 750 W	min. 750 W
	Min. RAM of Host System [1]	48 GB	96 GB
	Recommended Host System	Please ask your hardware vendor!	Please ask your hardware vendor!
	Recommended OS	RedHat EL 6/7 Windows 7/8 Win Server 2008 R2/2012 R2	RedHat EL 6/7 Windows 7/8 Win Server 2008 R2/2012 R2
	Vendor Information	www.nvidia.com	www.nvidia.com

[1]The host system requires approximately 4 times as much memory as is available on the GPU cards. Although it is technically possible to use less memory than this recommendation, the simulation performance of larger models will suffer.

CST assumes no liability for any problems caused by this information.

	Hardware Type	NVIDIA Tesla M60[1] (for Servers/Workst.)	NVIDIA Tesla M40[1] (for Servers)
GPU Hardware Properties	**Min. CST version required**	**2016 SP 4**	**2016 SP 4**
	Number of GPUs	2	1
	Max. Problem Size (Transient Solver)	approx. 160 million mesh cells	approx. 120 million mesh cells
	Form Factor	Dual-Slot PCI-Express	Dual Slot PCI-Express Passive Cooling
	Memory	16 GB GDDR5 (8 GB x 2)	12 GB GDDR5
	Power Consumption	300 W (max.)	250 W (max.)
Host System Requirements	PCI Express Requirements	1x PCIe Gen 3 (x16 electrically)	1x PCIe Gen 3 (x16 electrically)
	Power Supply of Host System	min. 750 W	min. 750 W
	Min. RAM of Host System [2]	64 GB	48 GB
	Recommended Host System	Please ask your hardware vendor!	Please ask your hardware vendor!
	Recommended OS	RedHat EL 6/7 Windows 7/8 Win Server 2008 R2/2012 R2	RedHat EL 6/7 Windows 7/8 Win Server 2008 R2/2012 R2
	Vendor Information	www.nvidia.com	www.nvidia.com

[1] **The double precision performance of this GPU device is poor, thus, it can't be recommended for PIC-solver, A-Solver and I-solver (double precision) simulations.**

[2] The host system requires approximately 4 times as much memory as is available on the GPU cards. Although it is technically possible to use less memory than this recommendation, the simulation performance of larger models will suffer.

CST assumes no liability for any problems caused by this information.

	Hardware Type	NVIDIA Quadro M6000[1] (for Workstations)	NVIDIA Quadro M6000 24 GB[1] (for Workstations)
GPU Hardware Properties	**Min. CST version required**	**2015 SP 4**	**2016 SP 4**
	Number of GPUs	1	1
	Max. Problem Size (Transient Solver)	approx. 120 million mesh cells	approx. 240 million mesh cells
	Form Factor	Dual-Slot PCI-Express	Dual Slot PCI-Express
	Memory	12 GB GDDR5	24 GB GDDR5
	Power Consumption	300 W (max.)	300 W (max.)
Host System Requirements	PCI Express Requirements	1x PCIe Gen 3 (x16 electrically)	1x PCIe Gen 3 (x16 electrically)
	Power Supply of Host System	min. 750 W	min. 750 W
	Min. RAM of Host System [2]	48 GB	96 GB
	Recommended Host System	Please ask your hardware vendor!	Please ask your hardware vendor!
	Recommended OS	RedHat EL 6/7 Windows 7/8 Win Server 2008 R2/2012 R2	RedHat EL 6/7 Windows 7/8 Win Server 2008 R2/2012 R2
	Vendor Information	www.nvidia.com	www.nvidia.com

[1] **The double precision performance of this GPU device is poor, thus, it can't be recommended for PIC-solver, A-Solver and I-solver (double precision) simulations.**

[2]The host system requires approximately 4 times as much memory as is available on the GPU cards. Although it is technically possible to use less memory than this recommendation, the simulation performance of larger models will suffer.

CST assumes no liability for any problems caused by this information.

5 NVIDIA Drivers Download and Installation

An appropriate driver is required in order to use the GPU hardware. Please download
the driver appropriate to your GPU hardware and operating system from the NVIDIA
website. The driver versions listed below are verified for use with our software. Other
driver versions provided by NVIDIA might also work but it is highly recommended to use
the versions verified by CST.
We recommend the following driver versions for all supported GPU cards:

Windows:	**Tesla GPUs:**	**Version 354.74**
	Quadro GPUs:	**Version 364.72**
Linux:	**Tesla GPUs:**	**Version 352.79**
	Quadro GPUs:	**Version 361.42**

5.1 GPU Driver Installation

5.1.1 Installation on Windows

After you have downloaded the installer executable please start the installation procedure
by double clicking on the installer executable. After a quick series of pop-up windows, the
NVIDIA InstallShield Wizard will appear. Press the "Next" button and driver installation
will begin (The screen may turn black momentarily.). You may receive a message indicat-
ing that the hardware has not passed Windows logo testing. In case you get this warning
select "Continue Anyway".
If you are updating from a previously installed NVIDIA driver, it's recommended to select
"clean installation" in the NVIDIA Installshield Wizard. This will remove the current
driver prior to installing the new driver.
The "Wizard Complete" window will appear as soon as the installation has finished. Select
"Yes, I want to restart my computer now" and click the "Finish" button.
**It is recommended that you run the HWAccDiagnostics tool after the instal-
lation to confirm that the driver has been successfully installed. Please use
HWAccDiagnostics_AMD64.exe which can be found in the AMD64 directory
of the installation folder.**

5.1.2 Installation on Linux

1. Login on the Linux machine as root.

2. Make sure that the adapter has been recognized by the system using the command
   `/sbin/lspci | grep -i nvidia`
   If you do not see any settings try to update the PCI hardware database of your system using the command
   `/sbin/update-pciids`

3. Stop the X-Server by running in a terminal the command (You may skip this step if you are working on a system without X-server)
   `init 3`

4. Install the NVIDIA graphics driver. Follow the instructions of the setup script. In most cases the installer needs to compile a specific kernel module. If this is the case the gcc compiler and Linux kernel headers need to be available on the machine.

5. Restart the X-server by running the command (You may skip this step if you are working on a system without X-server)
   `init 5`

   Note: In case you're using the CST Distributed Computing system and a DC Solver Server is running on the machine where you just installed the driver you need to restart the DC Solver Server as otherwise the GPUs cannot be detected properly.

   Note: The OpenGL libraries should not be installed on a system which has no rendering capabilities (like a pure DC Solver Server or a pure cluster node). This can be accomplished by starting the NVIDIA installer using the option `"--no-opengl-files"`.

6. You may skip this step if a X-server is installed on your system **and** you are using a NVIDIA graphics adapter (in addition to the GPU Computing devices) in your system. If no X-server is installed on your machine or you don't have an additional NVIDIA graphics adapter, the NVIDIA kernel module will not be loaded automatically. Additionally, the device files for the GPUs will not be generated automatically. The following commands will perform the necessary steps to use the hardware for GPU Computing. It is recommended to append this code to your `rc.local` file such that it is executed automatically during system start.

```
# Load nvidia kernel module
modprobe nvidia

if [ "$?" -eq 0 ]; then

  # Count the number of NVIDIA controllers found.
  N3D=$(/sbin/lspci | grep -i nvidia | grep "3D controller" | wc -l)
  NVGA=$(/sbin/lspci | grep -i nvidia | grep "VGA compatible controller" | wc -l)

  N=$(expr $N3D + $NVGA - 1)
  for i in $(seq 0 $N); do
    mknod -m 666 /dev/nvidia$i c 195 $i;
  done

  mknod -m 666 /dev/nvidiactl c 195 255

fi
```

Please note:

- If you encounter problems during restart of the X-server please check chapter 8 "Common Problems" in the file README.txt located at `/usr/share/doc/NVIDIA_GLX-1.0`. Please also consider removing existing sound cards or deactivating onboard sound in the BIOS. Furthermore, make sure you are running the latest BIOS version.

- After installation, if the X system reports an error like `no screen found`, please check Xorg log files in `/var/log`. Open the log files in an editor and search for `"PCI"`. According to the number of hardware cards in your system you will find entries of the following form: `PCI: (0@7:0:0)`. In `/etc/X11`, open the file `xorg.conf` in an editor and search for `"nvidia"`. After the line `BoardName "Quadro FX 5800"` (or whatever card you are using) insert a new line that reads `BusID "PCI:7:0:0"` according to the entries found in the log files before. Save and close the `xorg.conf` file and type `startx`. If X still refuses to start, try the other entries found in the Xorg log files.

- You need the installation script to uninstall the driver. Thus, if you want to be able to uninstall the NVIDIA software you need to keep the installer script.

- Be aware of the fact that you need to reinstall the NVIDIA drivers if your kernel is updated as the installer needs to compile a new kernel module in this case.

5.2 Verifying Correct Installation of GPU Hardware and Drivers

As a final test to verify that the GPU hardware has been correctly installed, the following test can be executed: Log in to the machine and execute the HWAccDiagnostics_AMD64 program found in the AMD64 subfolder of your CST installation (Windows) or in the folder LinuxAMD64 on a Linux system. The output of the tool should look similar to the following picture if the installation was successful.

Figure 1: Output of HWAccDiagnostics_AMD64.exe tool.

5.3 Uninstalling NVIDIA Drivers

5.3.1 Uninstall Procedure on MS Windows

To uninstall NVIDIA drivers, select "NVIDIA Drivers" from the "Add or Remove Programs" list and press the "Change/Remove" button (see fig. 2). After the uninstall process has finished you will be prompted to reboot.

Figure 2: "Add or Remove Programs" dialog on Windows

5.3.2 Uninstall Procedure on Linux

Start the installer with the "--uninstall" option. This requires root permissions.

6 Switch On GPU Computing

6.1 Interactive Simulations

GPU Computing needs to be enabled via the acceleration dialog box before running a simulation. To turn on GPU Computing:

1. Open the dialog of the solver.

2. Click on the "Acceleration" button.

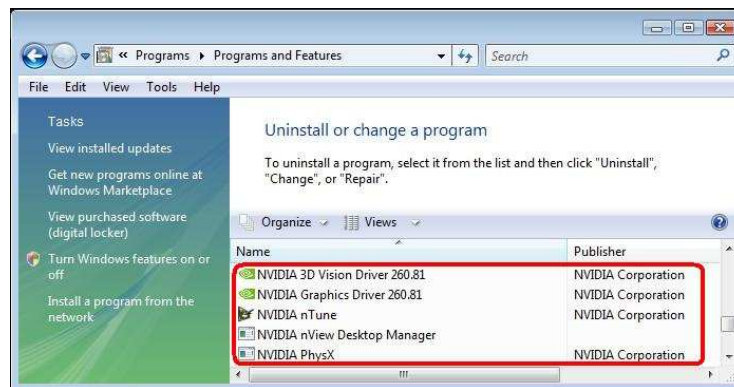3. Switch on "GPU Computing" and specify how many GPUs should be used for this simulation. Please note that the maximum number of GPUs available for a simulation depends upon the number of tokens in your license.

6.2 Simulations in Batch Mode

If you start your simulations in batch mode (e.g. via an external job queuing system) there is a command line switch (`-withgpu`) which can be used to switch on the GPU Computing feature. The command line switch can be used as follows:[6]

In Windows:

```
"<CST_INSTALL_DIR>/CST Design Environment.exe" -m -r -withgpu=<NUMBER_OF_GPUs> "<FULL_PATH_TO_CST_FILE>"
```

In Linux:

```
"<CST_INSTALL_DIR>/cst_design_environment" -m -r -withgpu=<NUMBER_OF_GPUs> "<FULL_PATH_TO_CST_FILE>"
```

---

[6]This example shows the batch mode simulation for the transient solver (`-m -r`). To learn more about the command line switches used by CST STUDIO SUITE please refer to the online help documentation in the section 'General Features', subsection 'Command Line Options'.

7 Usage Guidelines

7.1 The Error Correction Code (ECC) Feature

The recent GPU hardware (Tesla 20 and newer) comes with an ECC feature. ECC can detect and eventually correct problems caused by faulty GPU memory. Such GPU memory errors typically cause unstable simulations. However, this feature deteriorates the performance of the GPU hardware. Therefore, we recommend disabling the feature. In case simulations running on GPU hardware are getting unstable it's recommended to switch ECC back on at least temporarily to check whether the problems are caused by a GPU memory defect. Please also refer to section 10.
The ECC feature can be managed by using either the Nvidia Control Panel or the command line tool `nvidia-smi`. Please note, that on Windows 7, Windows Server 2008 R2, and newer version of Windows the following commands have to be run as administrator.

7.1.1 Managing the ECC Feature via Command Line

This procedure works on all versions of Windows 7/Server 2008 R2 and newer and on all Linux distributions.

1. Locate the file nvidia-smi. This file is typically found in
   "c:\`Program Files\NVIDIA Corporation\NVSMI`" or in `/usr/bin` on Linux.

2. Open up a command prompt/terminal window and navigate to this folder.

3. Execute the following command:
   `nvidia-smi -L`

4. Please note down how many GPUs are found.

5. To disable ECC: Please execute the following command for each of the GPUs:
   `nvidia-smi -i <number_of_the_GPU_card> -e 0`

6. To enable ECC: Please execute the following command for each of the GPUs:
   `nvidia-smi -i <number_of_the_GPU_card> -e 1`

7. Reboot.

7.1.2 Managing the ECC Feature via Nvidia Control Panel

This procedure works on all versions of Windows.

1. Start the Control Panel via the Windows start menu.

2. Start the Nvidia Control Panel.

3. Search for the term "ECC State" in the navigation tree of the dialog and open the "ECC State" page of the dialog by clicking on the tree item.

4. Disable or enable the ECC feature for all Tesla devices (see fig. 3).
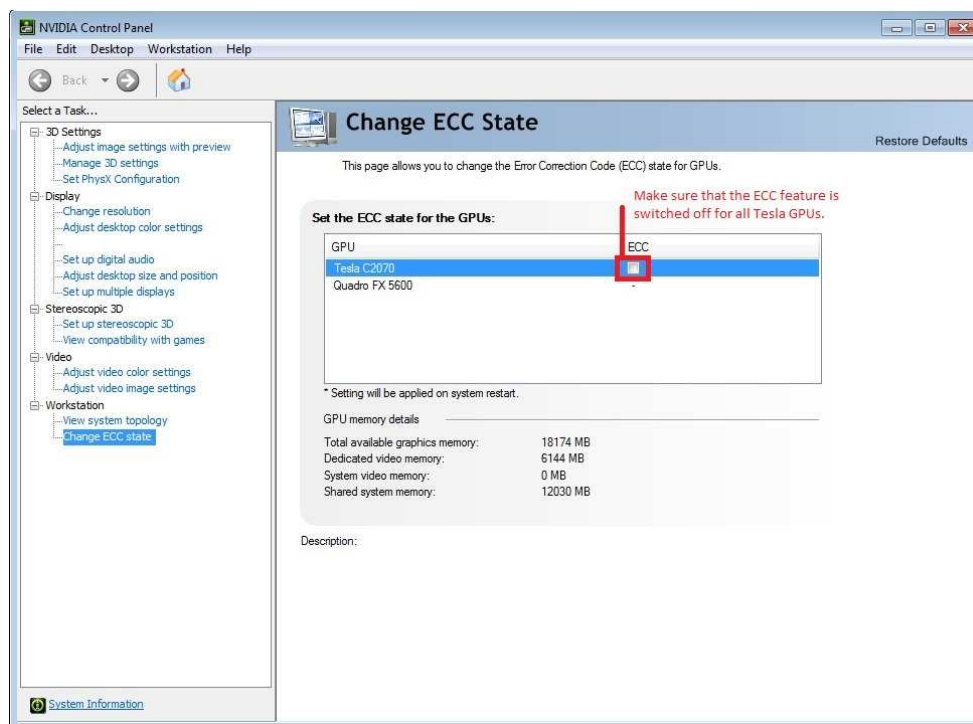

Figure 3: Switch off the ECC feature for all Tesla cards.

7.2   Enable Tesla Compute Cluster (TCC) mode

7.2.1   Enable the TCC Mode

When available, the GPUs have to operate in TCC mode[7]. Please enable the mode, if not yet enabled. Please note, that on Windows Vista, Windows 7 and Windows Server 2008 R2, the following commands have to be run "as administrator".

1. Locate the file nvidia-smi. This file is typically found in
   "c:\Program Files\NVIDIA Corporation\NVSMI".

2. Open up a command prompt and navigate to this folder.

3. Execute the following command:
   nvidia-smi -L

4. Please note down how many GPUs are found.

5. For each of the GPUs, please execute the following command:
   nvidia-smi -i <number_of_the_GPU_card> -dm 1

6. Reboot.

7.2.2   Disabling the TCC mode

If available, this feature should always be enabled. However, under certain circumstances you may need to disable this mode.

1. Locate the file nvidia-smi. This file is typically found in
   "c:\Program Files\NVIDIA Corporation\NVSMI".

2. Open up a command prompt and navigate to this folder.

3. Execute the following command:
   nvidia-smi -L

4. Please note down how many GPUs are found.

5. For each of the GPUs, please execute the following command:
   nvidia-smi -i <number_of_the_GPU_card> -dm 0

6. Reboot.

---

[7]The TCC mode is available on all Tesla cards. Graphics cards, such as the NVIDIA Quadro cards (e.g., Quadro FX5800, Quadro 6000), might not have this mode. Also, this mode is not available for Quadro cards which are connected to a display/monitor.

7.3 Disable the Exclusive Mode

This mode has to be disabled in order to use CST STUDIO SUITE.
To test if this mode is switched on, please do the following:

1. Locate the file nvidia-smi. This file is typically found in
   "c:\Program Files\NVIDIA Corporation\NVSMI" or in /usr/bin on Linux.

2. Open up a command prompt and navigate to this folder.

3. Execute the following command:
   `nvidia-smi -q`

Search for the term "Compute Mode" in the output of the tool. If the setting for "Compute Mode" is not "default" or "0", then the card is being operated in an exclusive mode. In this case, please execute the following commands in order to disable this mode:

1. Locate the file nvidia-smi. This file is typically found in
   "c:\Program Files\NVIDIA Corporation\NVSMI" or /usr/bin on Linux.

2. Open up a command prompt and navigate to this folder.

3. Execute the following command:
   `nvidia-smi -L`

4. Please note down how many GPUs are found.

5. For each of the GPUs, please execute the following command:
   `nvidia-smi -g <number_of_the_GPU_card> -c 0`

6. There is no need to reboot.

7.4 Display Link

Some cards of the Tesla 20 and the Kepler series provide a display link to plug in a monitor. Using this display link has the following implications:

- The TCC mode of the card cannot be used. This deteriorates the performance.

- A remote desktop connection to the machine is no longer possible (if the machine is running Windows).

Because of these limitations we recommend using an additional graphics adapter for the graphics output, or if available, an onboard graphics chipset.

July 8, 2016   20

7.5   Combined MPI Computing and GPU Computing

For combined MPI Computing and GPU Computing the TCC mode of the GPU hardware must be enabled (see 7.2).

7.6   Service User

If you are using GPU Computing via the CST Distributed Computing system and your DC Solver Server runs Windows 7/Windows Server 2008 R2 or a more recent Windows version then the DC Solver Server service must be started using the Local System account (see fig. 4). The CST STUDIO SUITE installer installs the service by default using the correct account.
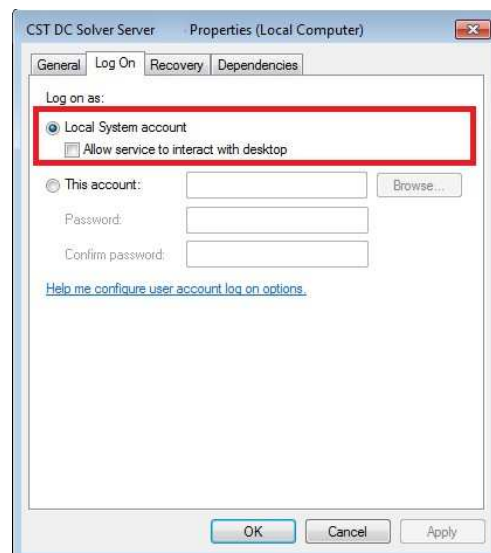
Figure 4: Local System Account.

7.7   GPU Computing using Windows Remote Desktop

For users with a LAN license, GPU computing using RDP can be used in combination with Tesla GPU cards if the TCC mode has been enabled (see section 7.2). Please note that this feature is only supported on newer operating systems (Windows 7/Windows Server 2008 R2 or newer). Additionally, only Tesla cards can be accessed from within RDP sessions, i.e. the Quadro FX5800, Quadro 6000, Quadro K6000 or other graphics cards can't be used for GPU computing in remote desktop sessions.

7.8   Running Multiple Simulations at the Same Time

Running multiple simulations in parallel on **the same** GPU card will deteriorate the performance. Therefore we recommend to run just one simulation at a time. If you have a system with multiple GPU cards and would like to assign simulations to specific GPU cards please refer to section 7.13.

7.9   Video Card Drivers

Please use only the drivers recommended in this document or by the hardware diagnostics tool (See section 5.2). They have been tested for compatibility with CST products.

7.10   Operating Conditions

CST recommends that GPU Computing is operated in a well ventilated temperature controlled area. For more information, please contact your hardware vendor.

7.11   Latest CST Service Pack

Download and install the latest CST Service Pack prior to running a simulation or HWAccDiagnostics.

7.12   GPU Monitoring/Utilization

Locate the file nvidia-smi. This file is typically found in
"c:\Program Files\NVIDIA Corporation\NVSMI" or in /usr/bin on Linux. If you start this tool with the command line switch -l or --loop it will show the utilization and other interesting information such as the temperatures of the GPU cards. The -l option makes sure that the tool runs in a loop such that the information gets updated every couple seconds. For more options please run nvidia-smi -h. If you want to check the GPU utilization only, you can also run the graphical tool NvGpuUtilization (Windows only). This file is typically found in
"c:\Program Files\NVIDIA Corporation\Control Panel Client".

7.13 Select Subset of Available GPU Cards

If you have multiple GPU cards supported for GPU computing in the same machine you may want to specify the cards visible to the CST software such that your simulations are only started on a subset of the cards. This can be accomplished in two different ways.

7.13.1 Environment Variable `CUDA_VISIBLE_DEVICES`

The environment variable `CUDA_VISIBLE_DEVICES` which contains a comma separated list of GPU IDs will force a process (such as a CST solver) to use the specified subset of GPU cards only).[8] If this variable is set in the environment of the CST software or globally on your system the simulation will be started on the cards listed in the `CUDA_VISIBLE_DEVICES` list only.
Example: Open a shell (cmd on Windows, or bash on Linux) and enter

- `set CUDA_VISIBLE_DEVICES=0`

on Windows or

- `export CUDA_VISIBLE_DEVICES=0`

on Linux to bind all CST solver processes started from this shell to the GPU with ID 0. To make the setting persistent for all CST instances started on the system you may add the variable to the global system environment variables.

7.13.2 Distributed Computing

The CST Distributed Computing (DC) system can be used to assign the GPU cards of a multi-GPU system to different DC Solver Servers. The solver processes executed by a certain DC Solver Server will only be able to access the GPU cards assigned to this Solver Server (see fig. 5). Please refer to the online help documents of CST STUDIO SUITE (section "Simulation Acceleration", subsection "Distributed Computing") to learn more about the setup and configuration of the DC system.

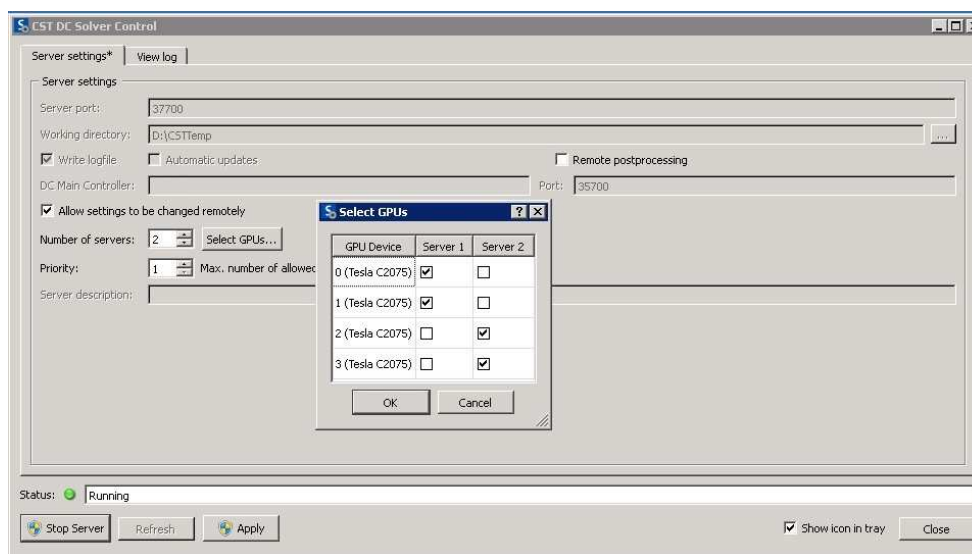[8]Execute the command `nvidia-smi -L` to get the GPU IDs of your cards.

Figure 5: Assignment of GPUs to specific DC Solver Servers.

8 NVIDIA GPU Boost

NVIDIA GPU Boost™is a feature available on the recent NVIDIA Tesla products.
This feature takes advantage of any power and thermal headroom in order to boost performance by increasing the GPU core and memory clock rates. The Tesla GPUs are designed with a specific Thermal Design Power (TDP). Frequently HPC workloads do not come close to reaching this power limit, and therefore have power headroom. A performance improvement can be expected when using the GPU Boost feature on the CST solvers. The Tesla GPUs come with a "Base clock" and several "Boost Clocks" which may be manually selected for compute intensive workloads with available power headroom. The Tesla GPUs give full control to end-users to select one of the core clock frequencies via the NVIDIA System Management Interface (nvidia-smi). For the K40 card, figuring out the right boost clock setting may require some experimentation to see what boost clock works best for a specific workload. NVIDIA GPU Boost on the Tesla K80 is enabled by default and dynamically selects the appropriate GPU clock based on the power headroom. The GPU Boost feature can be employed by using either the NVIDIA Control Panel or the command line tool nvidia-smi. The nvidia-smi file is typically found in
"c:\\`Program Files`\\`NVIDIA Corporation`\NVSMI" in Microsoft Windows or `/usr/bin` in Linux. The following are common commands for setting the GPU Boost feature and checking GPU performance. To display the current application clock in use execute the following command:

<div align="center">

`nvidia-smi -q -d CLOCK`

</div>

```
Default Applications Clocks
        Graphics                       : 745 MHz
        Memory                         : 3004 MHz
Max Clocks
        Graphics                       : 875 MHz
        SM                             : 875 MHz
        Memory                         : 3004 MHz
```

Before making any changes to the clocks, the GPU should be set to Persistence Mode. Persistence mode ensures that the driver stays loaded and does not revert back to the default clock once the application is complete and no CUDA or X applications are running on the GPU. To enable persistence mode use the following command:

<div align="center">

`nvidia-smi -pm 1`

</div>

To view the clocks that are supported by the Tesla board:

<div align="center">

`nvidia-smi -q -d SUPPORTED_CLOCKS`

</div>

```
GPU 0000:02:00.0
        Supported Clocks
                Memory                 : 3004 MHz
                        Graphics       : 875 MHz
                        Graphics       : 810 MHz
                        Graphics       : 745 MHz
                        Graphics       : 666 MHz
                Memory                 : 324 MHz
                        Graphics       : 324 MHz
```

Please note that the supported graphics clock rates are tied to a specific memory clock rate so when setting application clocks you must set both the memory clock and the graphics clock[9]. Do this using the `-ac <MEM clock, Graphics clock>` command line option.

<div align="center">

`nvidia-smi -ac 3004,875`

</div>

```
Applications clocks set to "(MEM 3004, SM 875)" for GPU 0000:02:00.0
Applications clocks set to "(MEM 3004, SM 875)" for GPU 0000:03:00.0
Applications clocks set to "(MEM 3004, SM 875)" for GPU 0000:83:00.0
Applications clocks set to "(MEM 3004, SM 875)" for GPU 0000:84:00.0
```

Execute the following command to reset the application clocks back to default settings.

<div align="center">

`nvidia-smi -rac`

</div>

Changing application clocks requires administrative privileges. However, a system administrator can remove this requirement to allow non-admin users to change application clocks by setting the application clock permissions to 'UNRESTRICTED' using the following command:

<div align="center">

`nvidia-smi -acp UNRESTRICTED`

</div>

---

[9]The memory clock should remain at 3 GHz for the Tesla K40.

```
Applications clocks commands have been set to UNRESTRICTED for GPU 0000:02:00.0
Applications clocks commands have been set to UNRESTRICTED for GPU 0000:03:00.0
Applications clocks commands have been set to UNRESTRICTED for GPU 0000:83:00.0
Applications clocks commands have been set to UNRESTRICTED for GPU 0000:84:00.0
```

Please be aware that the application clock setting is a recommendation. If the GPU cannot safely run at the selected clocks, for example due to thermal or power reasons, it will automatically lower the clocks to a safe clock frequency. You can check whether this has occurred by typing the following command while the GPU is active:

<div align="center">

nvidia-smi -a -d PERFORMANCE

</div>

9   Licensing

The GPU Computing feature is licensed by so called "Acceleration Tokens", i.e. your CST license must contain at least one "Acceleration Token" if you want to accelerate your simulations using a GPU. The CST Licensing Guide, which can be downloaded from the CST homepage, contains more information about the licensing of the GPU Computing feature. Please open the link https://www.cst.com/Company/Terms in your browser to find the recent version of this document.

10   Troubleshooting Tips

The following troubleshooting tips may help if you experience problems.

- If you experience problems during the installation of the NVIDIA driver on the Windows operating system please try to boot Windows in "safe mode" and retry the driver installation.

- Please note that CST STUDIO SUITE cannot run on GPU devices when they are in "exclusive mode". Please refer to section 7.3 on how to disable this mode.

- If you are using an external GPU device ensure that the PCI connector cable is securely fastened to the host interface card.

- Uninstall video drivers for your existing graphics adapter prior to installing the new graphics adapter.

- Make sure the latest motherboard BIOS is installed on your machine. Please contact your hardware vendor support for more information about the correct BIOS version for your machine.

- Use the HWAccDiagnostics tool to find out whether your GPU hardware and your driver is correctly recognized.

- GPU temperatures are crucial for the performance and overheating of GPU devices can lead to hardware failures. Please refer to section 7.12 for details.

- A faulty GPU device can be responsible for seemingly random solver crashes. To ensure that your GPU is working properly please run the stress test of the HWAccDiagnostics tool found in the CST installation (`HWAccDiagnostics_AMD64 --stresstest -duration=2000 -percentage=99`)

- In case simulations are getting unstable when running on the GPU it's recommended to check the GPU memory by switching on the ECC feature on the GPU (see 7.1).

- If a GPU is not recognized during the installation please check if `Memory Mapped I/O above 4GB` is enabled in your bios settings.

If you have tried the points above with no success please contact CST technical support (info@cst.com).

11   History of Changes

The following changes have been applied to the document in the past.

Date	Description
Nov. 11 2015	first version of this document
Dec. 1 2015	update list of supported OS
Dec. 7 2015	update list of unsupported features on GPU, add A-solver to the list of supported solvers
Mar. 30 2016	update list of supported drivers, mark deprecated hardware, add comment for machines with more than 4 GPUs, fix typos and errors
June. 13 2016	update list of unsupported features, add link to list of qualified GPU server
July. 9 2016	add Tesla M60/M40 and Quadro M6000 24GB to the list of supported hardware