

# iray EDUCATIONAL

iray® PERFORMANCE TIPS

APRIL 2011

---

## CONTENTS

|   |          |
|---|----------|
| <b>HARDWARE: GPU SET UP</b>                     | <b>3</b> |
| Performances with CUDA Devices .....            | 3        |
| ECC .....                                       | 3        |
| SLI.....  | 4        |
| Multiple GPUs .....                             | 4        |
| Memory.....                                     | 5        |
| <b>MATERIALS AND LIGHTS</b>                     | <b>6</b> |
| General Performance Tips.....                   | 6        |
| Avoiding Fireflies / High Frequency Noise ..... | 6        |
| <b>MORE INFORMATION</b>                         | <b>8</b> |

---

## HARDWARE: GPU SET UP

### PERFORMANCES WITH CUDA DEVICES

iray achieves its maximum performance levels on CUDA devices when the amount of samples it processes per rendering pass is high enough.

If the resolution of the image to be rendered is low, the workload on the CUDA device might not be high enough to achieve top performance.

When the workload is not sufficiently high on multiple CUDA device systems, the resulting performance drop can be as dramatic as 90%.

To prevent such situations, iray tries to automatically balance the workload on the available GPU(s). In some cases, the automatic workload balancing might yield uninvited timeout situations.

Both the Operating System and NVIDIA drivers use timeouts to prevent a deadlocked or crashed GPU from debilitating the system. If a CUDA kernel runs for too long, the operation simply gets killed and an unspecified error is provided, allowing the system to regain control of the GPU and continue its operations. There might be cases when the GPU is neither deadlocked or crashed but the Operating System believes it is, while the device is simply busy with the rendering process.

If such timeouts are experienced, it is possible to force the maximum amount of work allowed on a CUDA device by using the string option "iray max cuda fragment size", please refer to the "iray string options" guide for more information.

Alternatively, it is possible to either modify or disable the Operating System's timeout mechanism.

On **Windows Vista SP1+** and **Windows 7**, the timeout is set to 2 seconds by default. It is possible to override such value or even disable the timeout mechanism. Please see:

[http://www.microsoft.com/whdc/device/display/wddm\\_timeout.mspx](http://www.microsoft.com/whdc/device/display/wddm_timeout.mspx)

On **Windows XP**, the timeout is set to 5 seconds. Apparently, it is not possible to override or disable it. Please see:

[http://developer.download.nvidia.com/compute/cuda/2\\_3/toolkit/docs/cudatoolkit\\_release\\_notes\\_windows.txt](http://developer.download.nvidia.com/compute/cuda/2_3/toolkit/docs/cudatoolkit_release_notes_windows.txt)

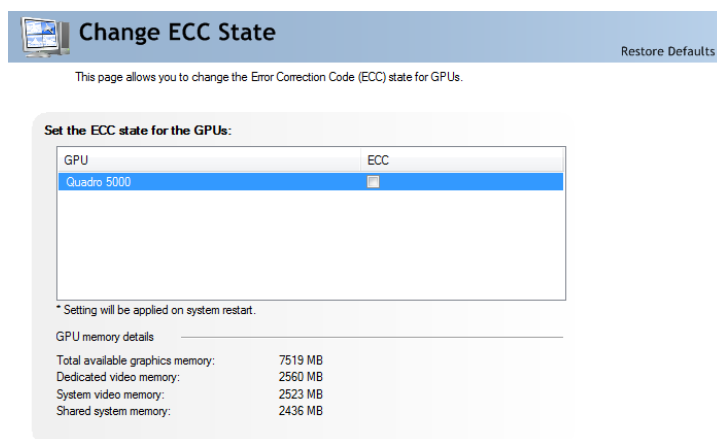
Excerpt:

Individual GPU program launches are limited to a run time of less than 5 seconds on a GPU with a display attached. Exceeding this time limit usually causes a launch failure reported through the CUDA driver or the CUDA runtime. GPUs without a display attached are not subject to the 5 second runtime restriction. For this reason it is recommended that CUDA be run on a GPU that is NOT attached to a display and does not have the Windows desktop extended onto it. In this case, the system must contain at least one NVIDIA GPU that serves as the primary graphics adapter.

### ECC

Fermi-class Quadro and Tesla boards feature "Error Checking and Correction" (ECC) memory modules. However, ECC reportedly introduces a performance drop between 20% and 30% when rendering with iray, therefore, it is recommended to keep ECC disabled.

ECC can be disabled within the NVIDIA Control Panel.



## SLI

iray does not use NVIDIA Scalable Link Interface (SLI). However, performance tests showed that SLI can significantly slow down iray, hence it should be disabled.

SLI can be disabled within the NVIDIA Control Panel.



## MULTIPLE GPUS

iray will try to take advantage of all NVIDIA CUDA GPUs available on the system, regardless of their model. For example it is possible to use systems hosting both Fermi Tesla and Fermi Quadro cards. However, systems hosting models with great performance differences (either memory-wise or compute-wise) should be avoided. Significantly slower models (or models with much less memory available) should be excluded from the rendering process, limiting their use to Display Device, otherwise they might slow down the work of the more recent models.

To prevent specific GPUs from contributing to the rendering process can be achieved in two different ways:

- **Using the string option "iray devices"** (see "iray string options" for more information).
- Disabling specific CUDA devices via the system variable "CUDA\_VISIBLE\_DEVICES":  
"Specific GPUs can be made invisible with the CUDA\_VISIBLE\_DEVICES

---

environment variable. Visible devices should be included as a comma-separated list in terms of the system-wide list of devices. For example, to use only devices 0 and 2 from the system-wide list of devices, set `CUDA_VISIBLE_DEVICES` equal to `"0,2"` before launching the application. The application will then enumerate these devices as device 0 and device 1." (for more information: [http:// developer.nvidia.com](http://developer.nvidia.com)).

## MEMORY

iray requires the whole scene to be loaded into the memory of each participating GPU. Memory issues will therefore occur when the size of the scene (including textures) is greater than the memory available on the GPUs.

When multiple GPUs are used, differences with memory capabilities might yield the automatic exclusion of one or more GPUs. For example, if the system used for rendering hosts one GPU with 1.5 GB available memory, and one other GPU with 3.5 GB memory available, a scene the size of 2 GB will cause the exclusion of the first GPU from rendering, because it has not enough memory to store the scene. However, the second GPU will be used, because its 3.5 GB will be able to store the scene.

When the scene is too large for all the available GPUs, iray will fall back to using the system's CPUs alone.

---

## MATERIALS AND LIGHTS

### GENERAL PERFORMANCE TIPS

In order to keep noise levels low and achieve faster convergence, it is recommended to honor the following suggestions:

- Adjust texture resolutions to the scene requirements.
  - Textures compete with geometry for the GPU memory. The higher the resolution of textures, the less the memory available for geometry, and vice-versa.
  - High frequency textures at high resolutions may become a source of further noise and require more time to converge.
- If object lights are required, the light source material should have no "accident" settings such as Diffuse Color, Reflectivity, Transparency etc... the Glow Color is all that is needed there.

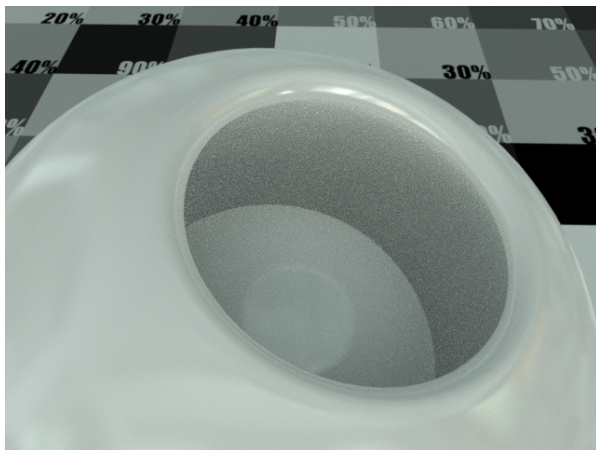
### AVOIDING FIREFLIES / HIGH FREQUENCY NOISE

The first step to do in order to eliminate or attenuate the so called fireflies (high frequency noise that comes in the form of bright spots) is to identify its source.

Most of the times, fireflies are due to hard caustics.

Once the source is identified, the following actions can be taken:

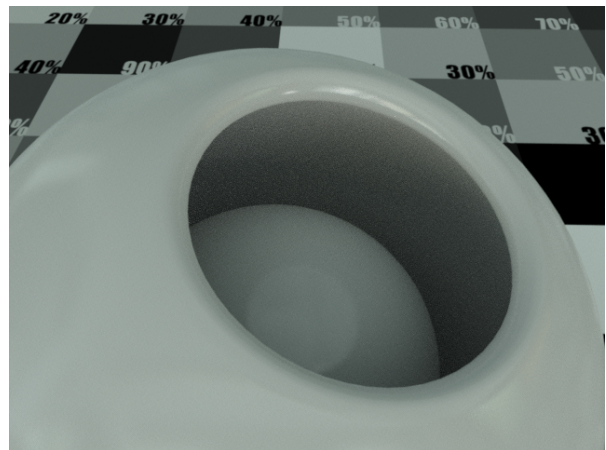
- Reduce the reflectivity. As a general rule, real-life materials rarely exceed a reflectance ratio of 0.7. Highly reflective materials will also slow down convergence.
- Avoid pure white colors.



**Reflectivity = 1.0, Color = [1.0, 1.0, 1.0]**

Render results after 200 iterations on a single Quadro 5000

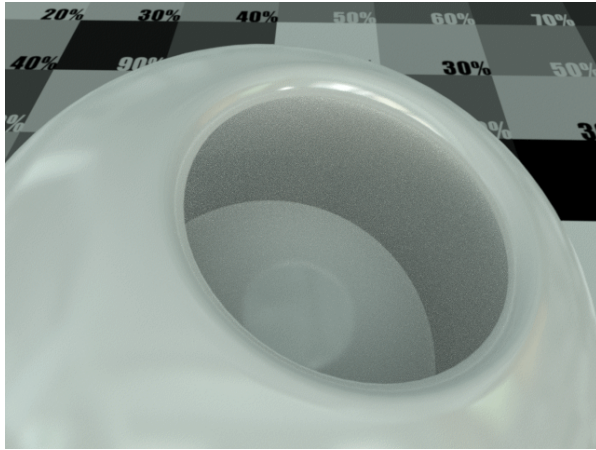
Time: 1 minute, 22 seconds



**Reflectivity = 0.7, Color = [0.7, 0.7, 0.7]**

Render results after 200 iterations on a single Quadro 5000

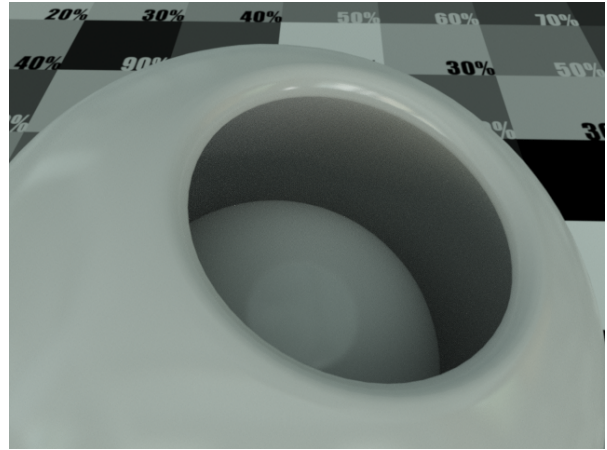
Time: 54 seconds



**Reflectivity = 1.0, Color = [1.0, 1.0, 1.0]**

Render results after 500 iterations on a single Quadro 5000

Time: 2 minutes, 58 seconds

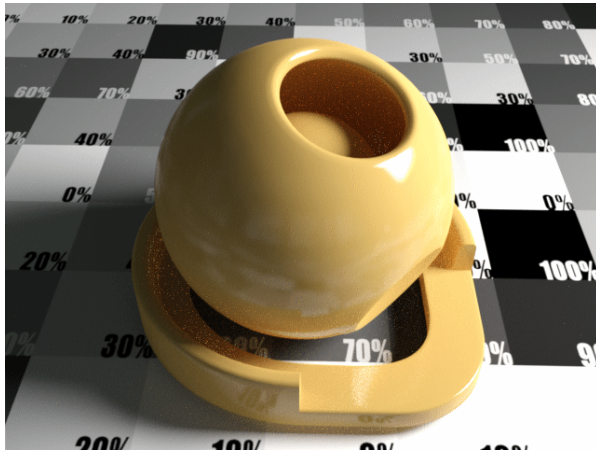


**Reflectivity = 0.7, Color = [0.7, 0.7, 0.7]**

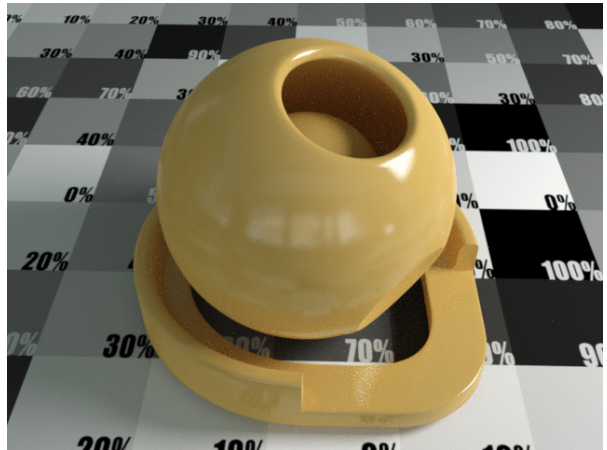
Render results after 500 iterations on a single Quadro 5000

Time: 1 minute, 40 seconds

- Slightly modify the lights direction and/or intensity.



Bright light after a few iterations. Notice the presence of fireflies.



Same light with less intensity, after a few iterations. Fireflies are still present but in less number and with less intensity. Overall lighting was improved by using the "mr photographic exposure".

- Glowing objects may introduce further noise, hence they should be used only when necessary. For all other situations, shape lights are preferable.

---

## MORE INFORMATION

For more tutorials, tips and hints

[www.irayrender.com](http://www.irayrender.com)

For technical questions

[forum.mentalimages.com](http://forum.mentalimages.com)

To stay up to date

[Twitter](#)

[Autodesk Area](#)