Maxwell

NEXT LIMIT

Porting Maxwell to the GPU

Top Challenges

Juan Cañada
Head of Visualization
Next Limit Technologies

GPU TECHNOLOGY CONFERENCE

Maxwell

Agenda

- Maxwell overview
- Why porting to the GPU was challenging
- Performance considerations
- Using the CPU to improve the GPU engine
- Summary

GPU TECHNOLOGY CONFERENCE

Maxwell

Agenda

- Maxwell overview
- Why porting to the GPU was challenging
- Performance considerations
- Using the CPU to improve the GPU engine
- Summary

GPU TECHNOLOGY CONFERENCE

Maxwell Overview

GPU TECHNOLOGY CONFERENCE

Maxwell

RealFlow

CaronteFX

Maxwell

Maxwell Overview

GPU TECHNOLOGY CONFERENCE

MAXWELL

- First physically based render in the market (2004)
- Ground-truth reference render
- Predictive rendering tool
- Light analysis tool

Maxwell in use

- Animation & VFX
- Architecture
- Industrial Design
- Science
- Others

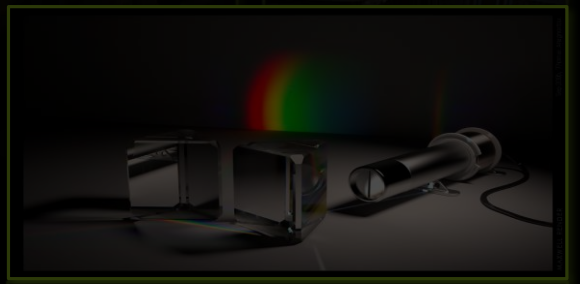
Maxwell in use

- Animation & VFX
- Architecture
- Industrial Design
- Science
- Others

Maxwell in use

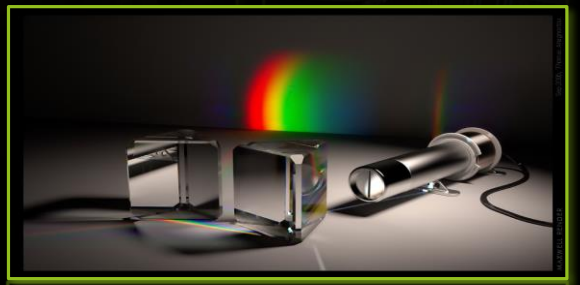
- Animation & VFX
- Architecture
- Industrial Design
- Science
- Others



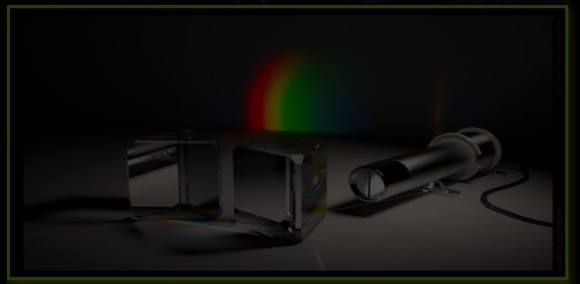

Maxwell in use

- Animation & VFX
- Architecture
- Industrial Design
- Science
- Others



Maxwell in use

- Animation & VFX
- Architecture
- Industrial Design
- Science
- Others




GPU TECHNOLOGY CONFERENCE
Maxwell

Agenda

- Maxwell Render overview
- Why porting to the GPU was challenging
- Performance considerations
- Using the CPU to improve the GPU engine
- Summary

GPU TECHNOLOGY CONFERENCE

Maxwell

Challenges

- Keep pixel accuracy
- Use GPU for predictive rendering
- Improve performance
- Spectral, unbiased, accurate PBR
- Support CPU & GPU resuming & merging
- …

GPU TECHNOLOGY CONFERENCE

Maxwell

Predictive Rendering

GPU TECHNOLOGY CONFERENCE

Photo.
Nikon D70. Focal lenght 24mm. CCD sensor size 23.7x15.6mm. Manual settings
ISO 200 / fStop 8 / Shutter speed 90 / WB set with 6500K lamp in Cornell box
Cornell box dimensions 50x50x50cm
Compact fluorescent lamp Philips Master PL electronic 865 (6500k daylight white)
27W / 1700 lumen / efficacy 62,96 lumen/W

Maxwell render 1.0
Maxwell camera. Focal lenght 24mm. Film size 23.7x15.6mm.
ISO 200 / fStop 8 / Shutter speed 90 / Burn 1 / Gamma 2.2
Scene box dimensions 50x50x50cm
Maxwell emitter 6500k / 27W / 1700 lumen / efficacy 62,96 lumen/W

GPU TECHNOLOGY CONFERENCE

Correct → Fast ☺

Fast → Correct ☹


GPU TECHNOLOGY CONFERENCE
Maxwell

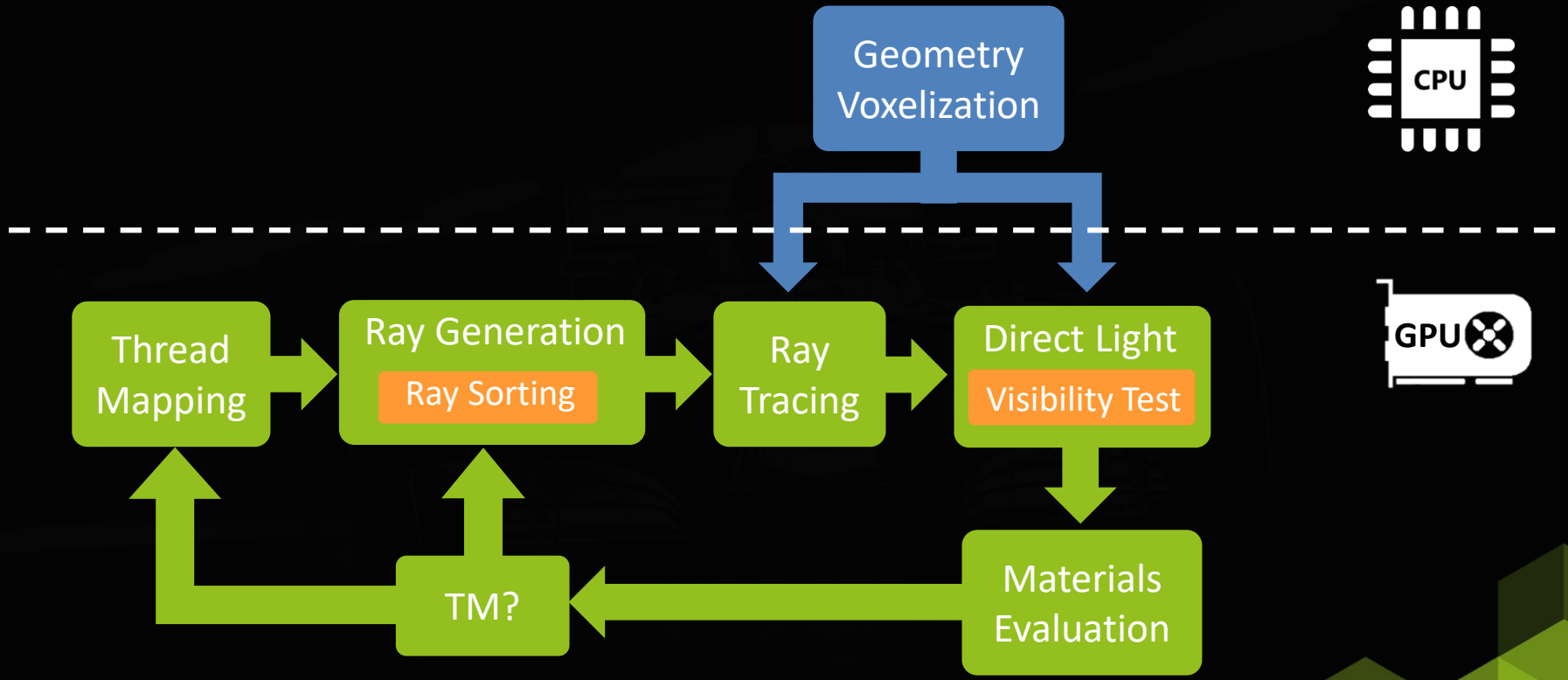
Agenda

- Maxwell overview
- Why porting to the GPU was challenging
- Performance considerations
- Using the CPU to improve the GPU engine
- Summary

GPU TECHNOLOGY CONFERENCE

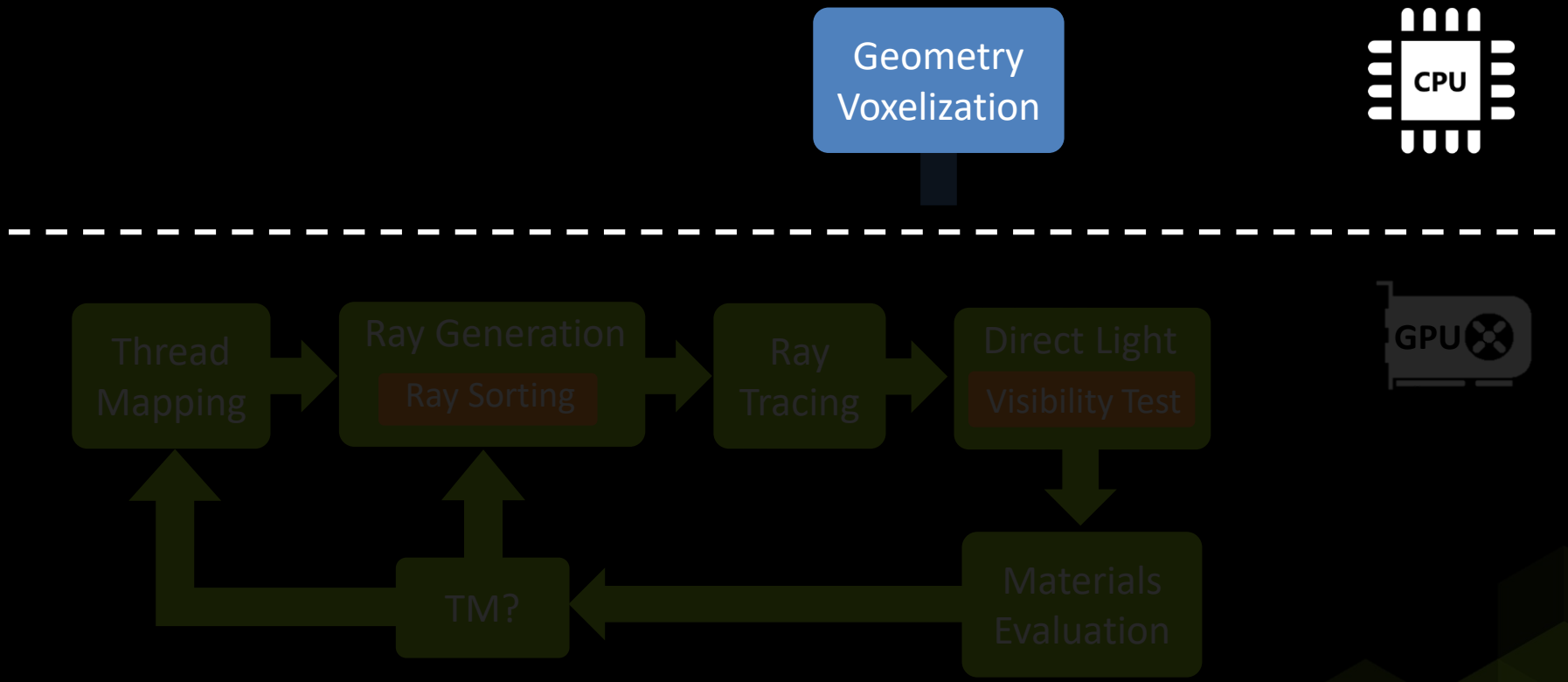
Maxwell GPU Architecture

GPU TECHNOLOGY CONFERENCE



GPU Maxwell

GPU TECHNOLOGY CONFERENCE



GPU Maxwell

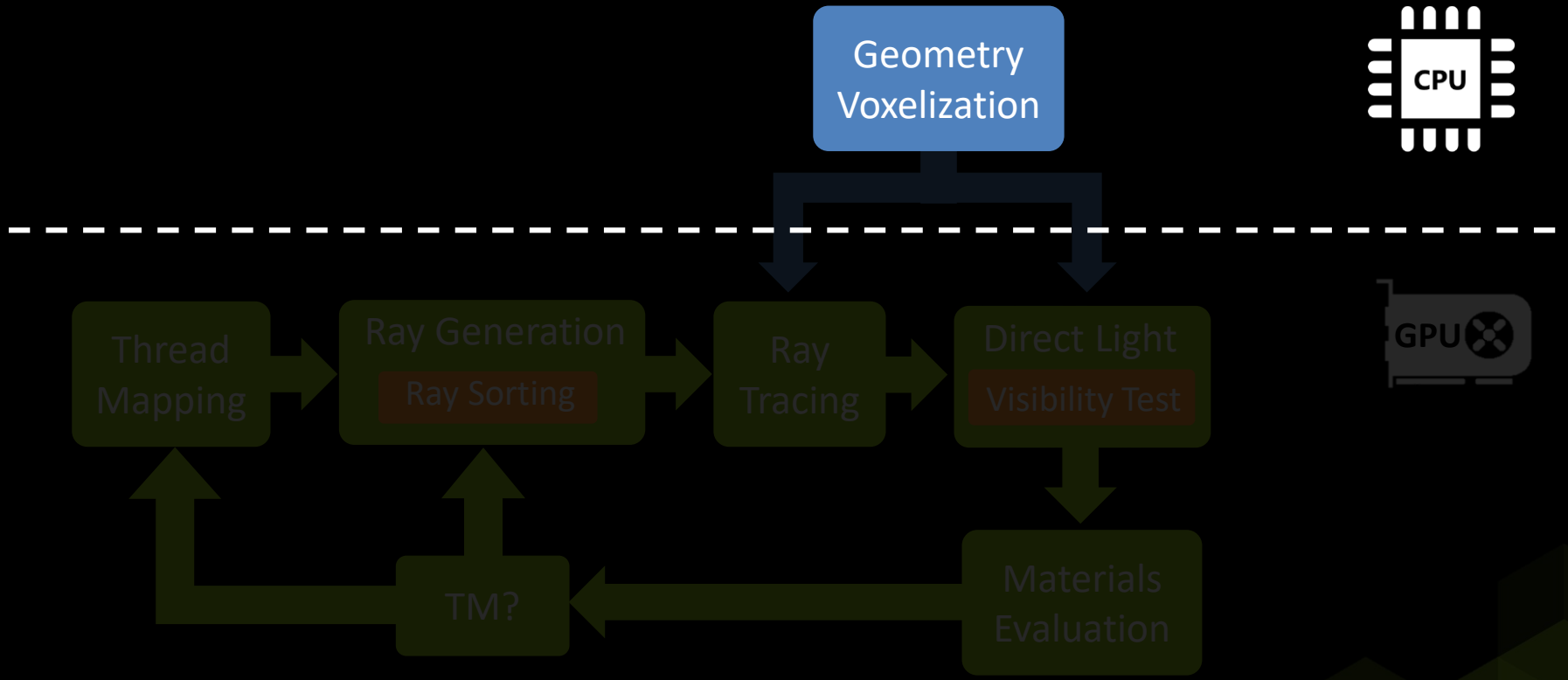
GPU TECHNOLOGY CONFERENCE

- Voxelization
  - Same Voxelization system as the CPU render
  - Currently performed in CPU just once
  - BVH
    - Binary tree (each node has 2 childs)
    - Coherent traversal

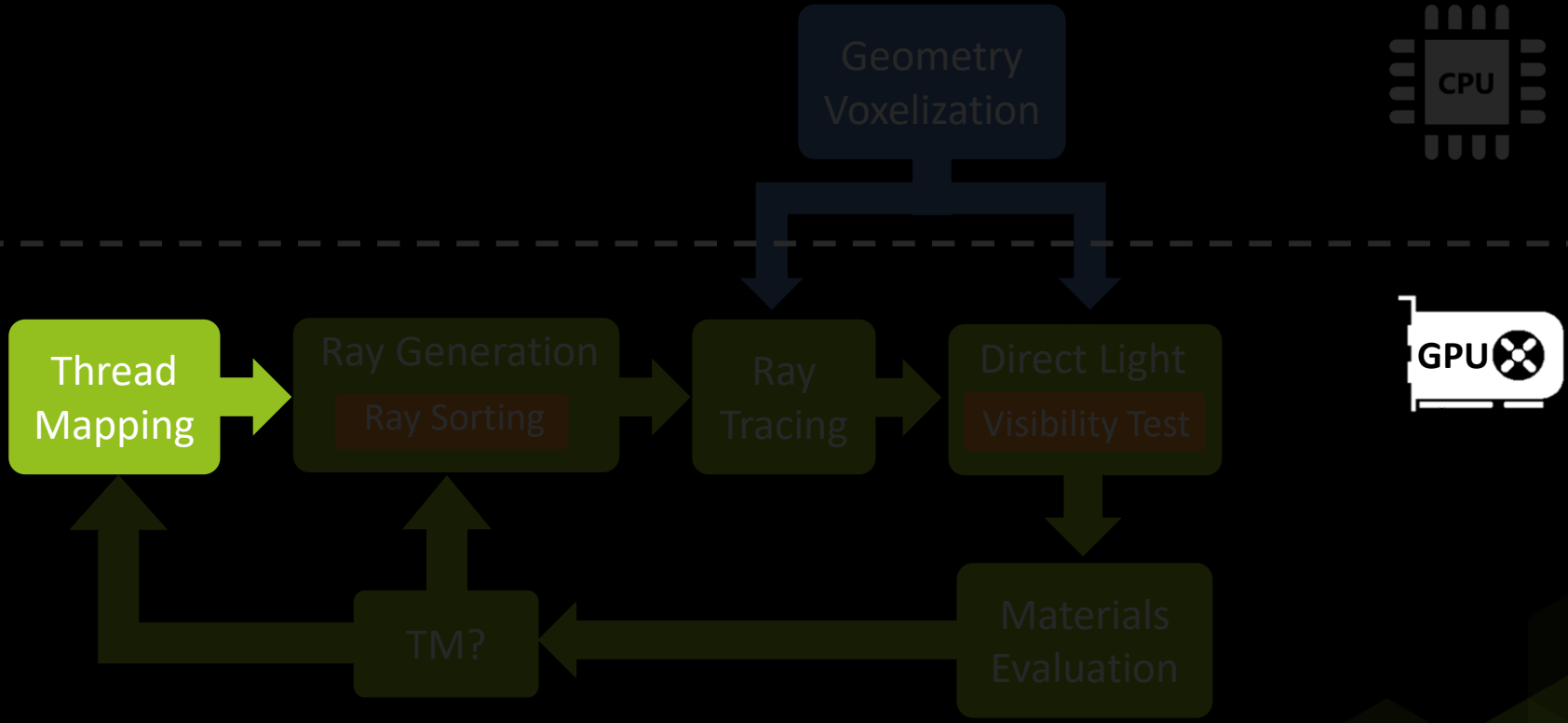
- \+ All threads fetch same amount of data / node
- \+ Increase coherence in performance
- \- Trees become bigger

GPU Maxwell

GPU TECHNOLOGY CONFERENCE

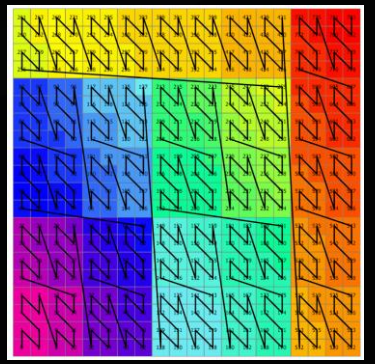


GPU Maxwell


Geometry
Voxelization
CPU
Thread
Mapping
Ray Generation
Ray Sorting
Ray
Tracing
Direct Light
Visibility Test
GPU
TM?
Materials
Evaluation

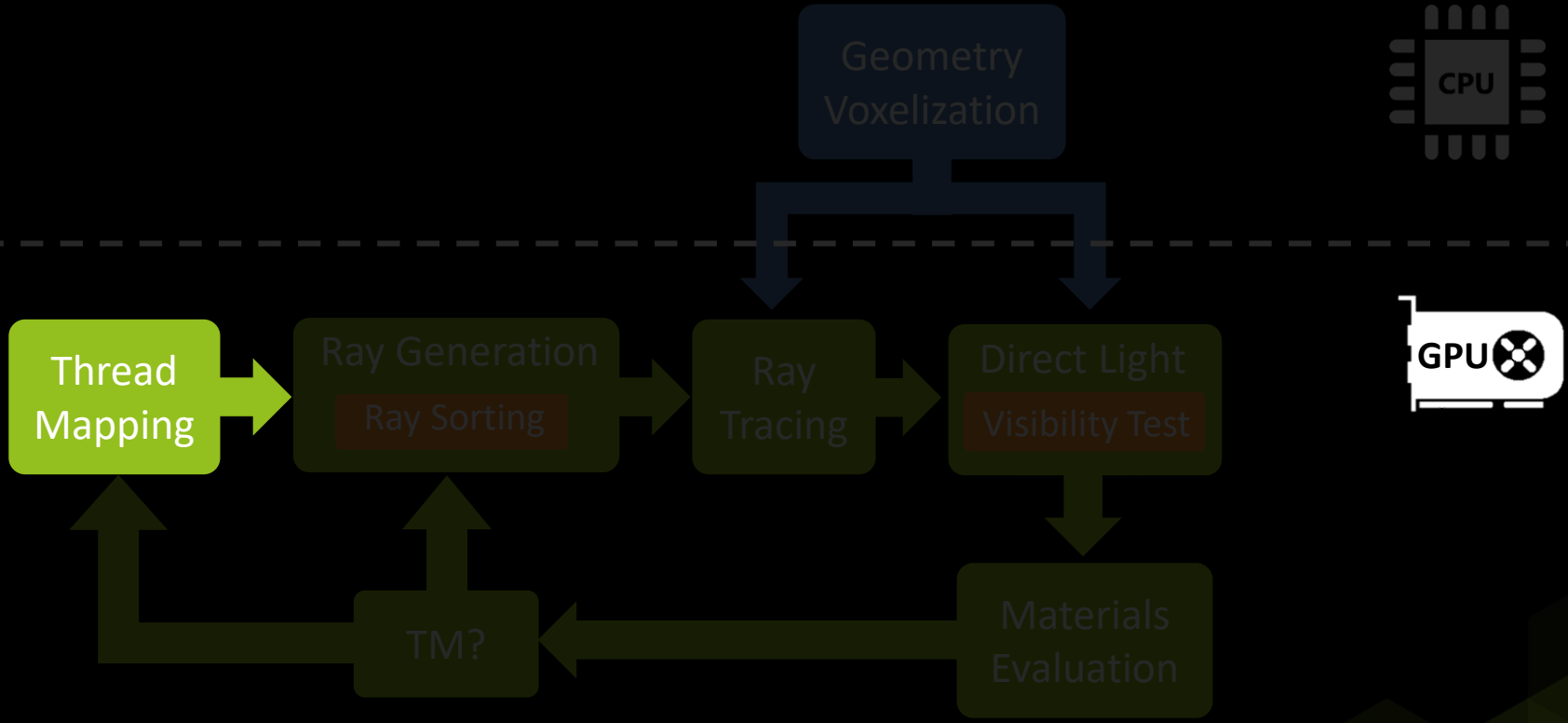
GPU Maxwell

Morton Curve

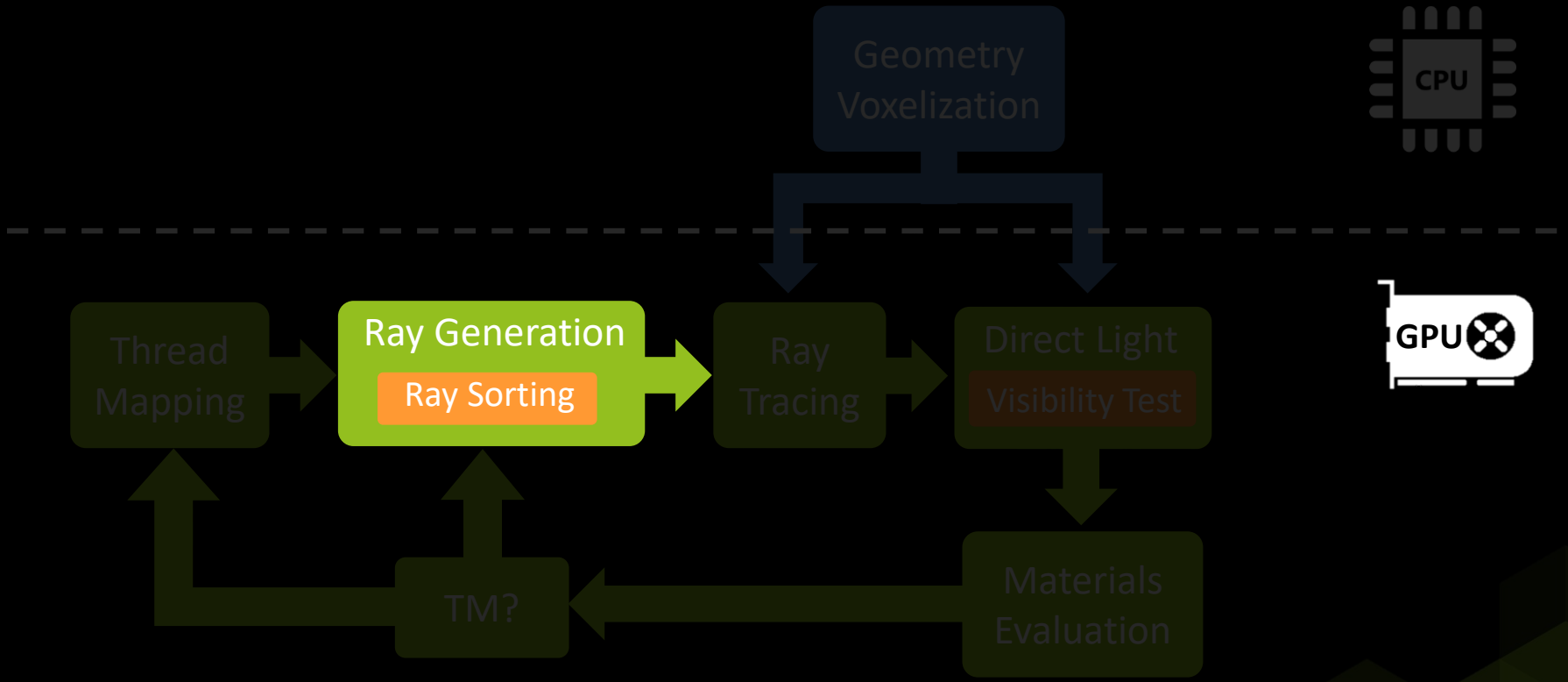


- Thread Mapping
  - Module that manages THREAD / PIXEL mapping
  - Sampling Level (SL)
    - Low Morton → Curve
    - Medium → Balances SPP
    - High → Uses Variance

GPU Maxwell


Geometry
Voxelization
CPU
Thread
Mapping
Ray Generation
Ray Sorting
Ray
Tracing
Direct Light
Visibility Test
GPU
TM?
Materials
Evaluation

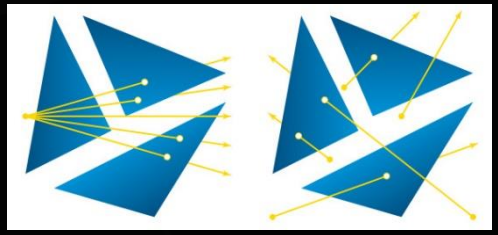
GPU Maxwell


Geometry Voxelization
CPU
Thread Mapping
Ray Generation
Ray Sorting
Ray Tracing
Direct Light
Visibility Test
GPU
TM?
Materials Evaluation

GPU Maxwell

GPU TECHNOLOGY CONFERENCE

- Ray Generation Module



  - Primary Rays (PR)
    - Rays shot from camera
    - High degree of coherence
    - Two neighboring rays will hit nearby similar objects
  - Secondary Rays (SR)
    - Rays shot from surfaces
    - No coherence
    - Two neighbouring rays might hit different objects

GPU Maxwell

- Ray Generation Module
  - Thread blocks with just PR
    - High degree of coherence
    - Best performance situation
  - Thread blocks with just SR
    - All will take much more time than PR
    - The worst SR will drive the performance
  - Thread blocks with PR and SR
    - SR will hurt PR performance

GPU TECHNOLOGY CONFERENCE

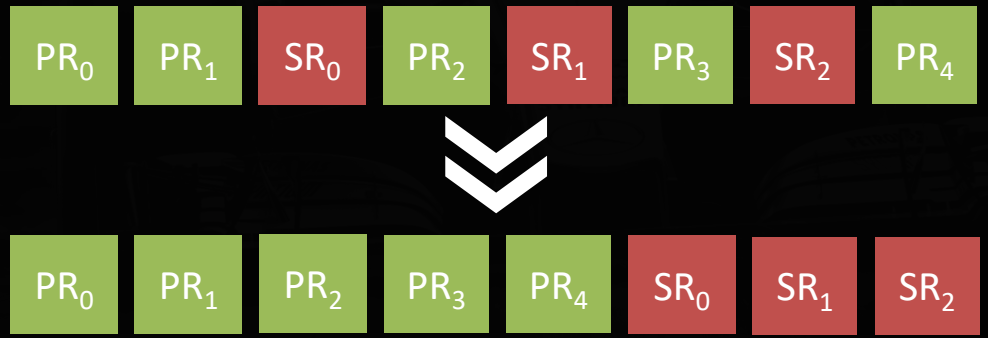
GPU Maxwell

- Ray Generation Module
  - How do we handle it?
    - GPU Ray sorting by Ray Type

GPU TECHNOLOGY CONFERENCE
Maxwell

GPU Maxwell

- Ray Generation Module
  - How do we handle it?
    - GPU Ray sorting by Ray Type



GPU TECHNOLOGY CONFERENCE
Maxwell

GPU Maxwell

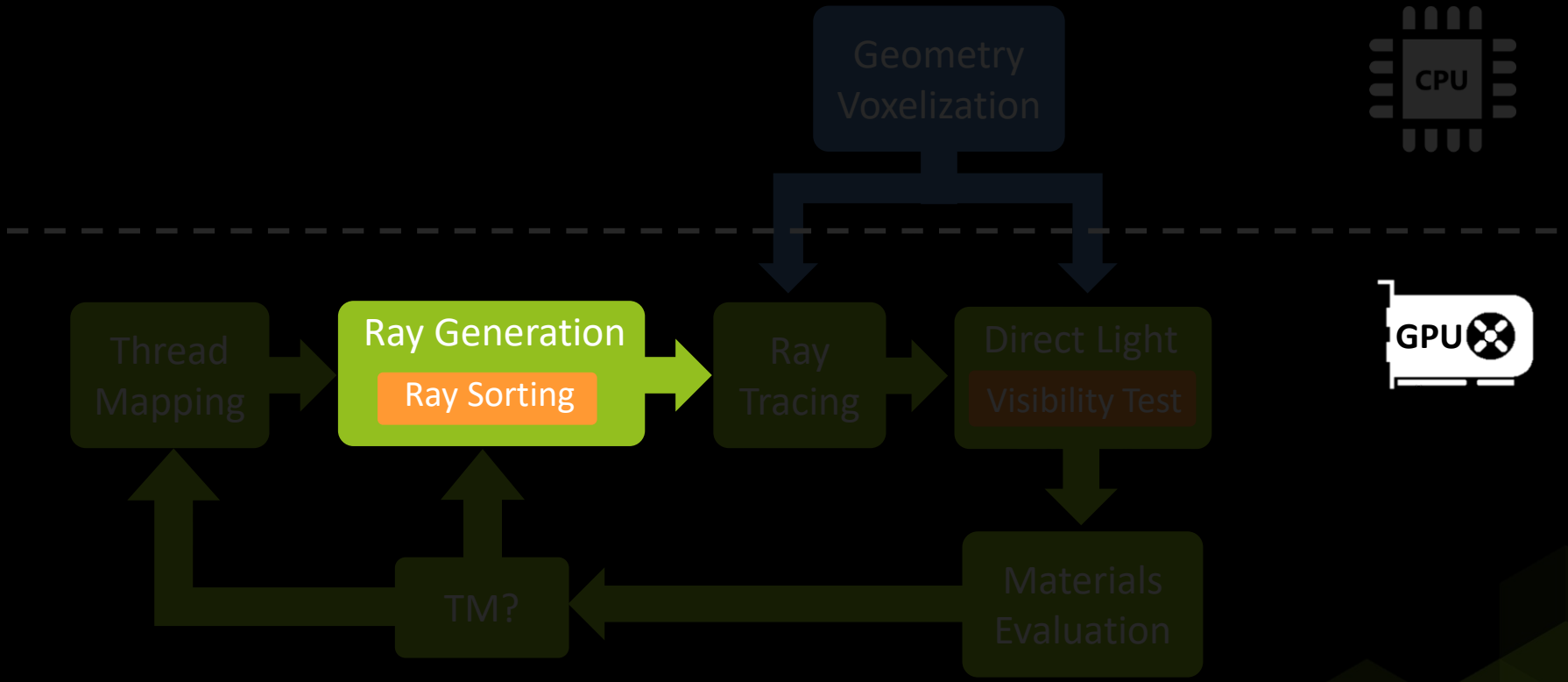
- Ray Generation Module
  - How do we handle it?
    - GPU Ray sorting by Ray Type
      - Sorting is really fast
      - Simple, yet powerful
      - Do it just after 2nd bounce
        - Not needed for PR
      - Performance boost is scene dependant

GPU Maxwell

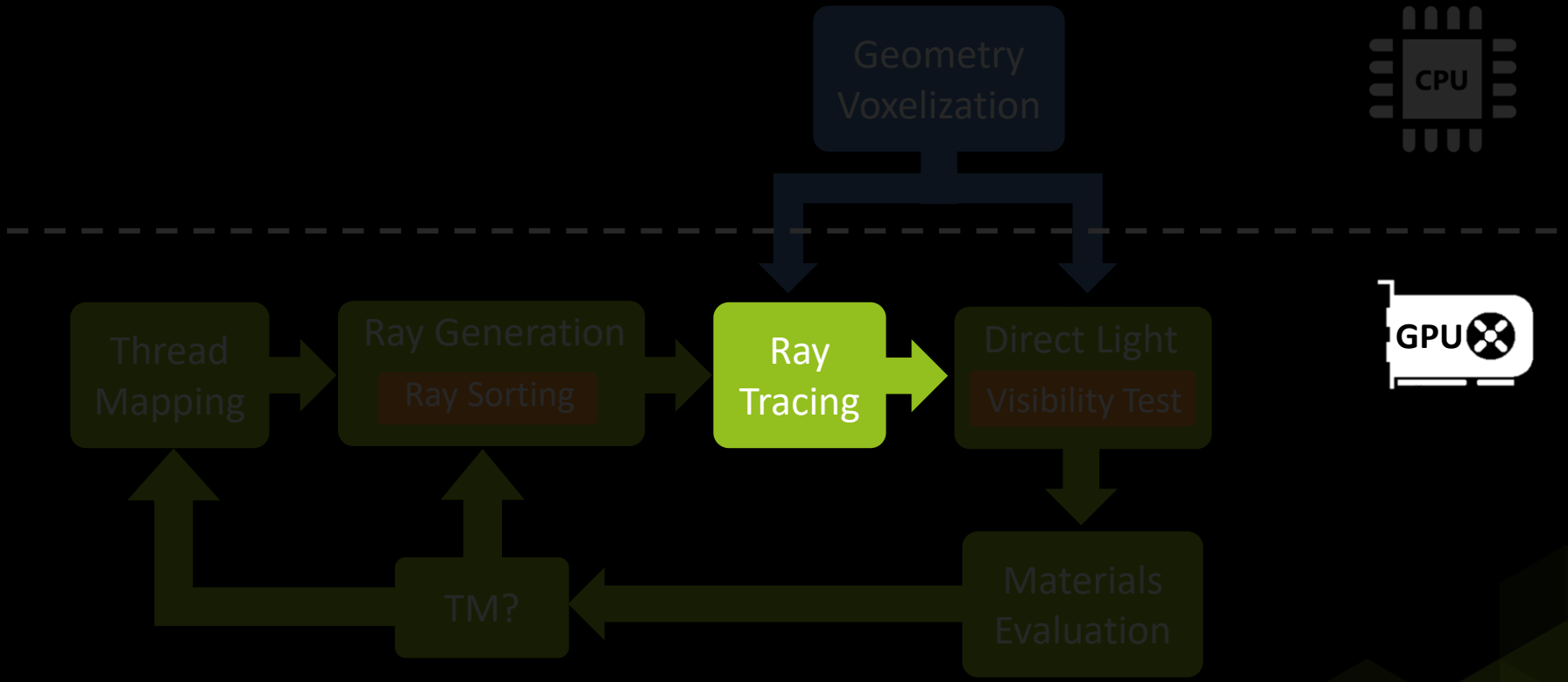
- Ray Generation Module
  - How do we handle it?
    - GPU Ray sorting by Ray Type
      - **Considerations**
        - Not useful for medium to small-res images
        - Use an indirection buffer
          - Cleaner code
          - Avoids moving global data
          - Much better performance

GPU TECHNOLOGY CONFERENCE

GPU Maxwell


Geometry Voxelization
CPU
Thread Mapping
Ray Generation
Ray Sorting
Ray Tracing
Direct Light
Visibility Test
GPU
TM?
Materials Evaluation

GPU Maxwell


Geometry
Voxelization
CPU
Thread
Mapping
Ray Generation
Ray Sorting
Ray
Tracing
Direct Light
Visibility Test
GPU
TM?
Materials
Evaluation

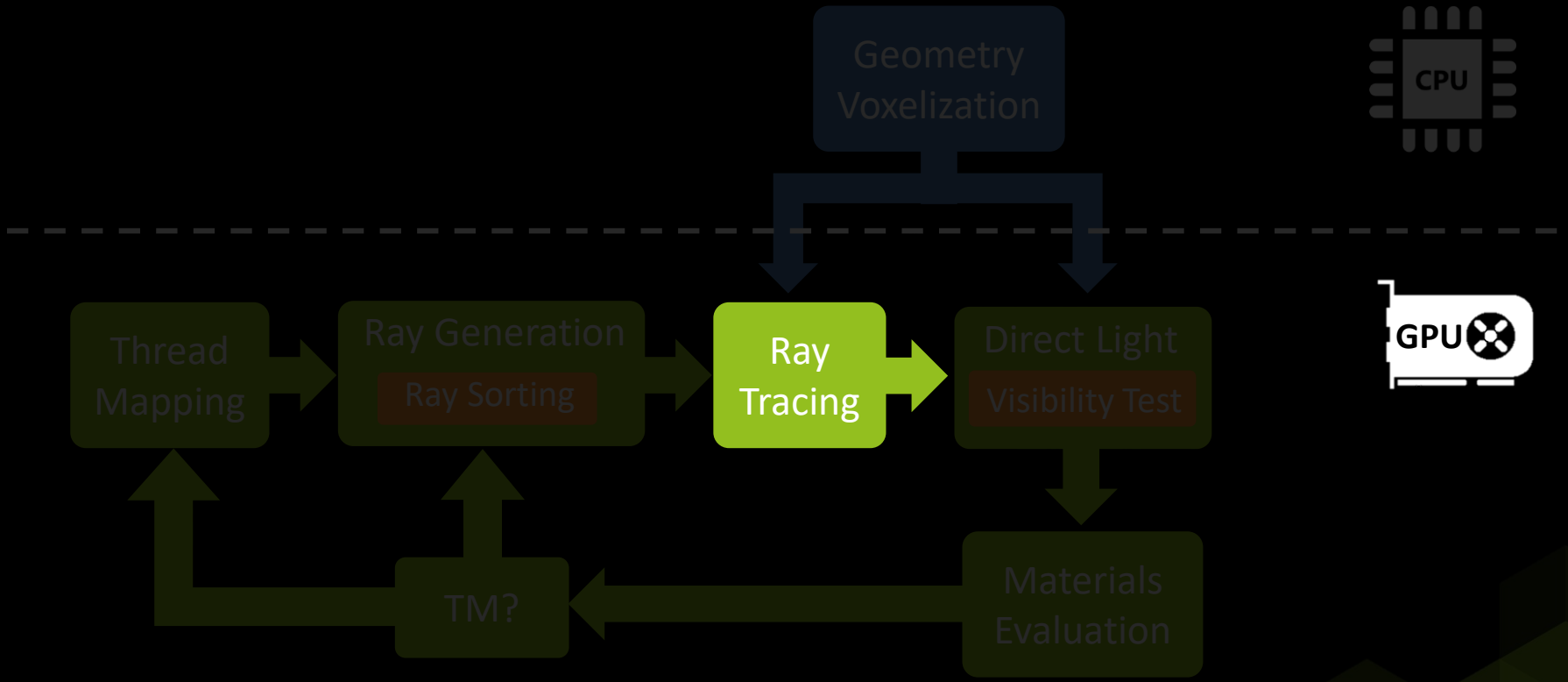
GPU Maxwell

GPU TECHNOLOGY CONFERENCE

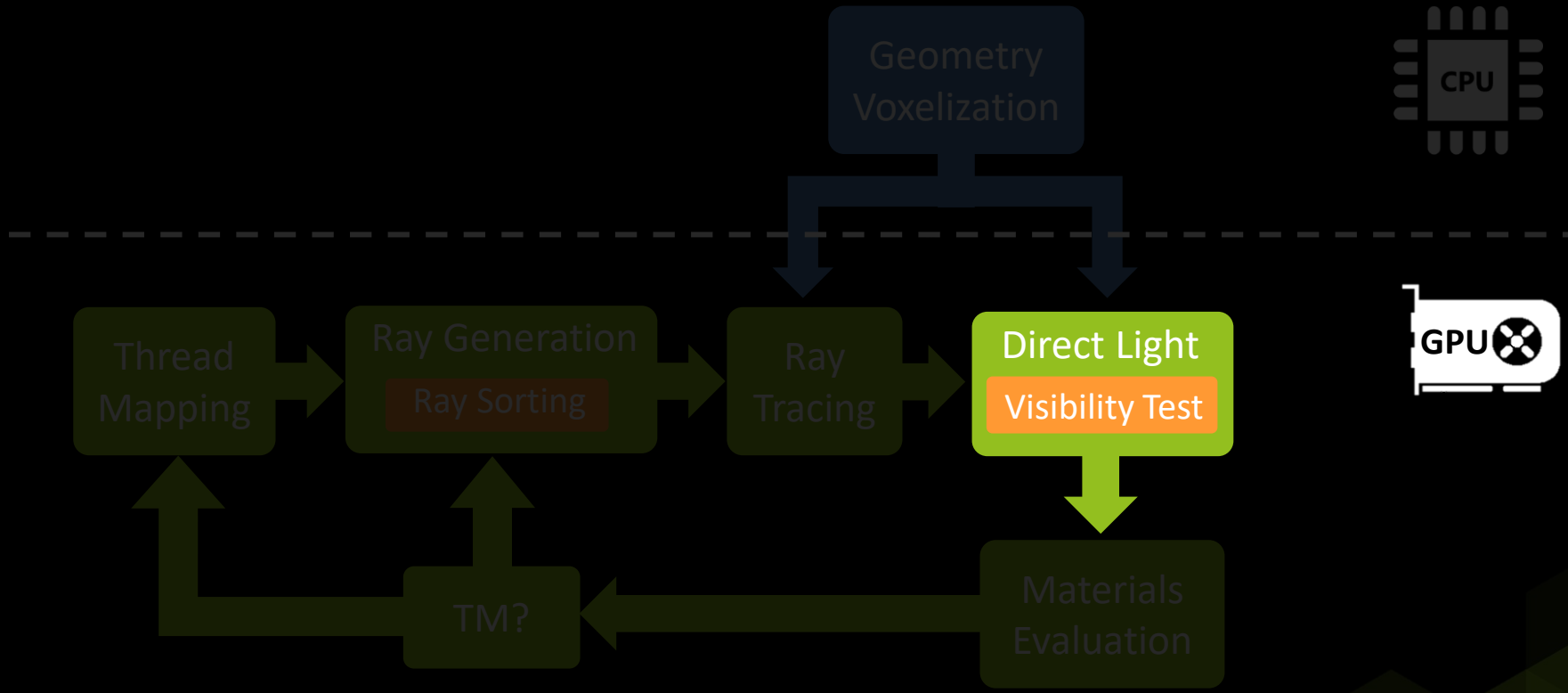
- Ray Tracing Module
  - GPU architecture dependent kernels
    - Fermi, Kepler, Maxwell
    - Use every architecture strengths

Maxwell

GPU Maxwell


Geometry
Voxelization
CPU
Thread
Mapping
Ray Generation
Ray Sorting
Ray
Tracing
Direct Light
Visibility Test
GPU
TM?
Materials
Evaluation

GPU Maxwell Render



GPU Maxwell

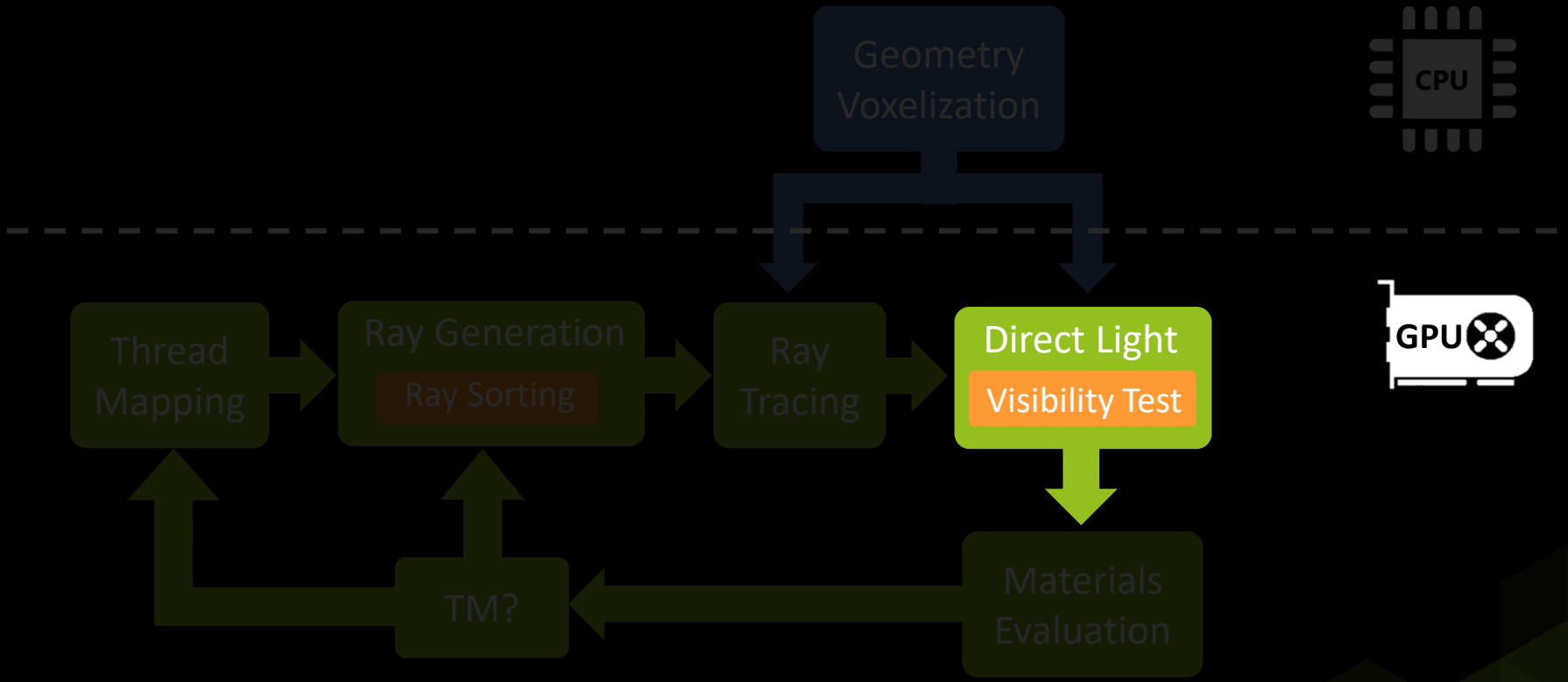
Direct Light Module

1. Sample scene emitters at each path node
   - Two strategies
     - Sample 1 random emitter / sample
     - Sample all emitters / sample

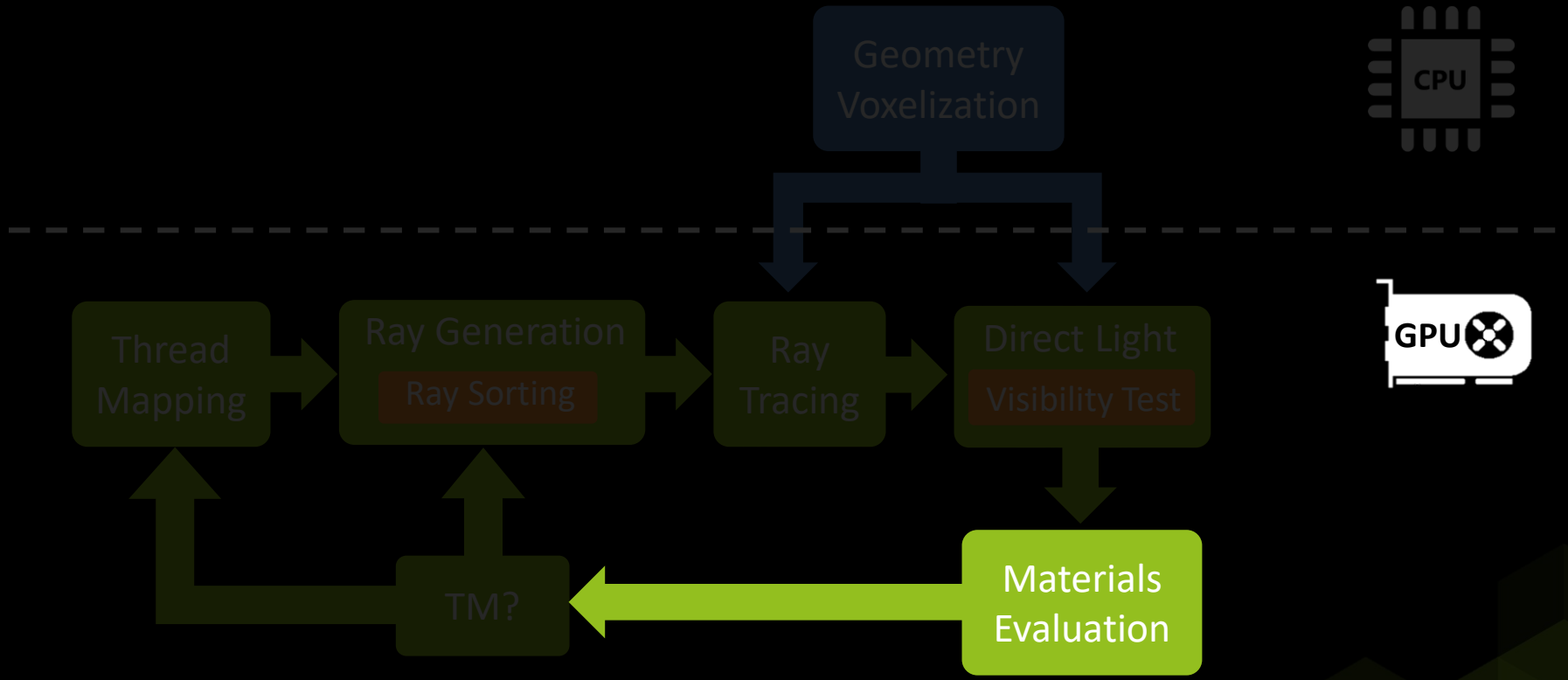
2. Visibility test
   - Trace shadow rays
   - Incoherent rays → Ray sorting does not help

3. Many other optimizations

GPU Maxwell


Geometry
Voxelization
CPU
Thread
Mapping
Ray Generation
Ray Sorting
Ray
Tracing
Direct Light
Visibility Test
GPU
TM?
Materials
Evaluation

GPU Maxwell


Geometry
Voxelization
CPU
Thread
Mapping
Ray Generation
Ray Sorting
Ray
Tracing
Direct Light
Visibility Test
GPU
TM?
Materials
Evaluation

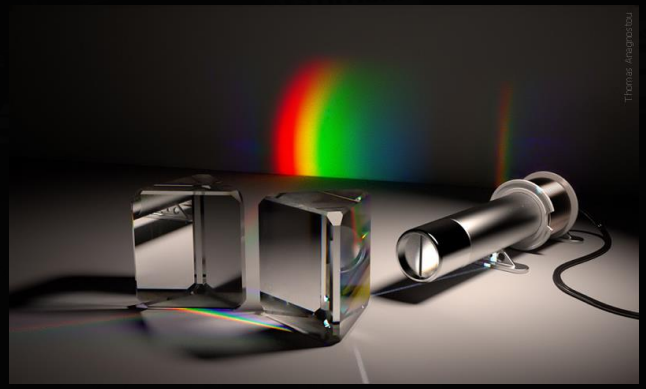
GPU Maxwell

- Materials Evaluation Module
  - Maxwell materials are complex
  - Many layers and many BSDFs / layer → very generic

GPU Maxwell

Materials Evaluation Module

- Bbig kernels are harmful
- Samples evaluating different materials
  - Access different data
  - Execute different code



GPU Maxwell

- Materials Evaluation Module
  - Materials Group Queue System (MGQS)
    1. Every material is assigned a Material Group ID
    2. Queue system for Material Groups (MG)
    3. Every queue has specific kernels
       - Avoid big kernels
    4. Samples are queued to the corresponding MG Queue
    5. All samples evaluating the same MG are executed together
       - Increased coherence in execution time
       - Increased coherence in data access

GPU Maxwell

- Materials Evaluation Module
  - Materials Group Queue System (MGQS)
    1. Every material is assigned a Material Group ID
    2. Queue system for Material Groups (MG)
    3. Every queue has specific kernels
       - Avoid big kernels
    4. Samples are queued to the corresponding MG Queue
    5. All samples evaluating the same MG are executed together
       - Increased coherence in execution time
       - Increased coherence in data access

GPU Maxwell Render

- Materials Evaluation Module
  - Materials Group Queue System (MGQS)
    1. Every material is assigned a Material Group ID
    2. Queue system for Material Groups (MG)
    3. Every queue has specific kernels
       - Avoid big kernels
    4. Samples are queued to the corresponding MG Queue
    5. All samples evaluating the same MG are executed together
       - Increased coherence in execution time
       - Increased coherence in data access

GPU Maxwell

- Materials Evaluation Module
  - Materials Group Queue System (MGQS)
    1. Every material is assigned a Material Group ID
    2. Queue system for Material Groups (MG)
    3. Every queue has specific kernels (Avoid big kernels)
    4. Samples are queued to the corresponding MG Queue
    5. All samples evaluating the same MG are executed together

Increased coherence in execution time
Increased coherence in data access

GPU Maxwell

GPU TECHNOLOGY CONFERENCE

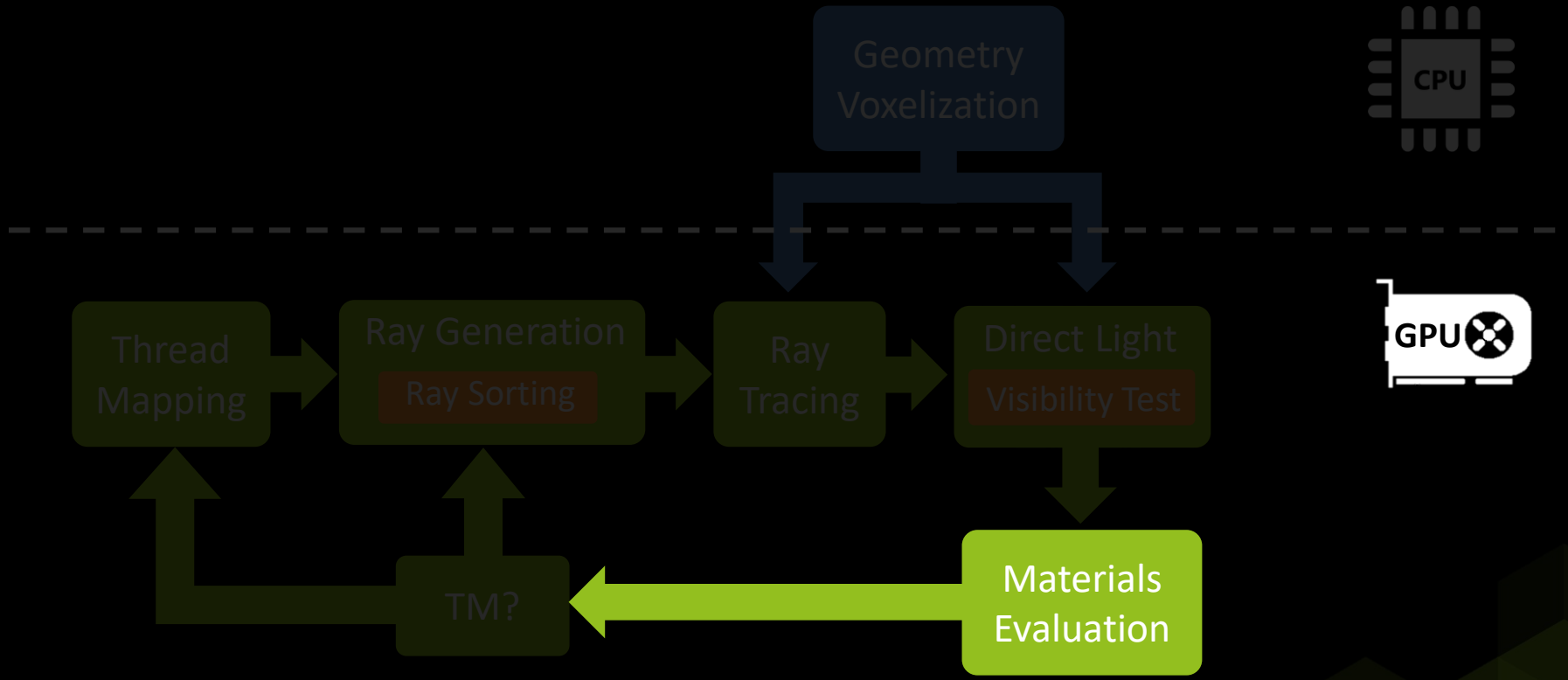
- Materials Evaluation Module
  - Materials Group Queue System (MGQS)
    1. Every material is assigned a Material Group ID
    2. Queue system for Material Groups (MG)
    3. Every queue has specific kernels (Avoid big kernels)
    4. Samples are queued to the corresponding MG Queue
    5. All samples evaluating the same MG are executed together
       - Increased coherence in execution time
       - Increased coherence in data access

GPU Maxwell

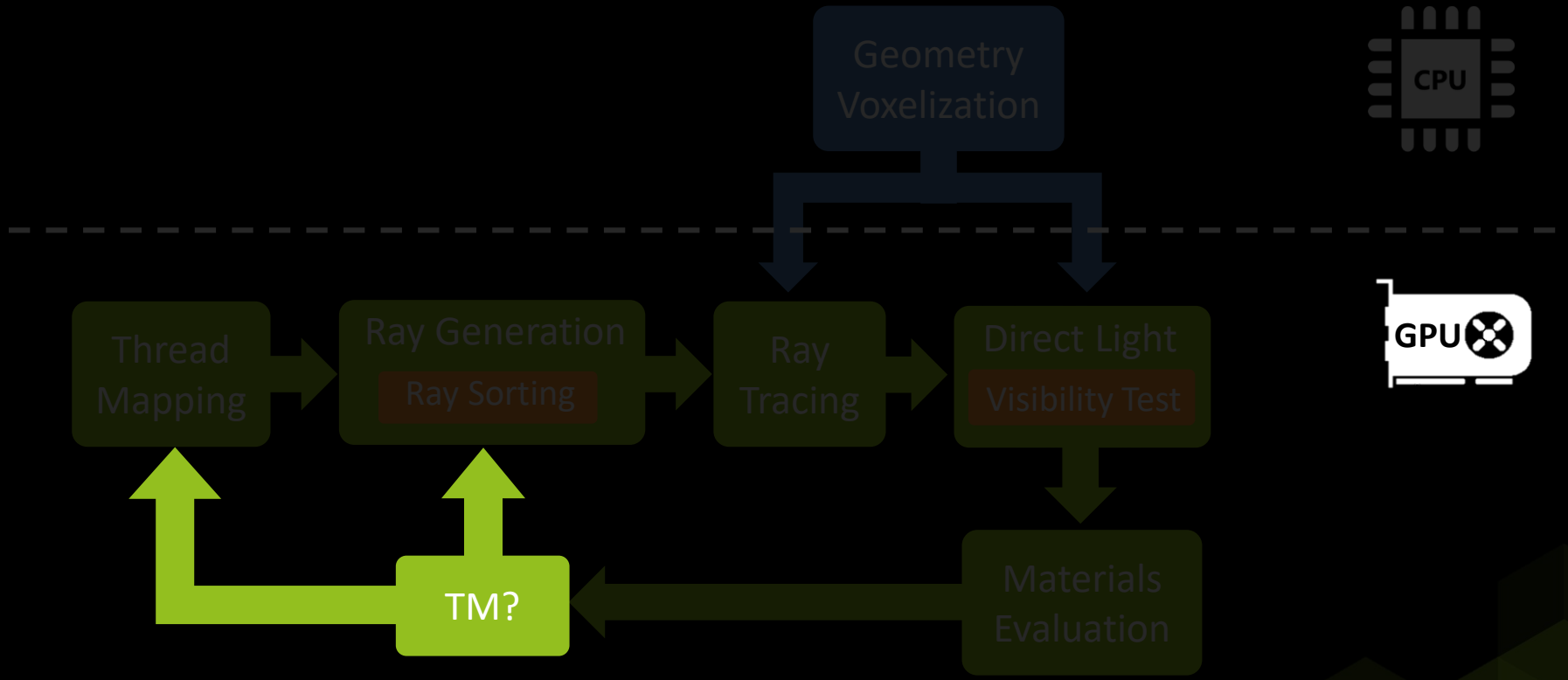
- Materials Evaluation Module
  - Materials Group Queue System (MGQS)
    1. Every material is assigned a Material Group ID
    2. Queue system for Material Groups (MG)
    3. Every queue has specific kernels (Avoid big kernels)
    4. Samples are queued to the corresponding MG Queue
    5. All samples evaluating the same MG are executed together
       - Increased coherence in execution time
       - Increased coherence in data access

GPU TECHNOLOGY CONFERENCE

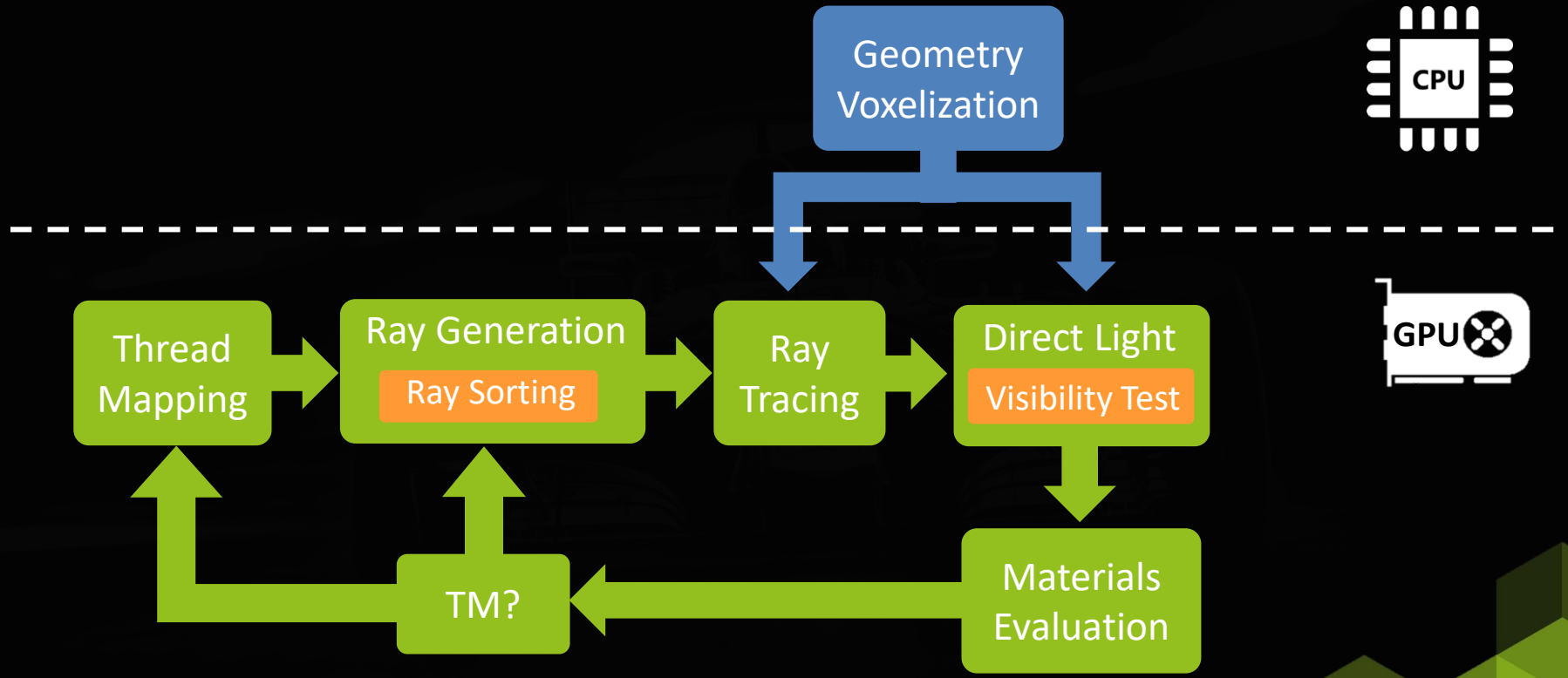
GPU Maxwell


Geometry
Voxelization
CPU
Thread
Mapping
Ray Generation
Ray Sorting
Ray
Tracing
Direct Light
Visibility Test
GPU
TM?
Materials
Evaluation

GPU Maxwell


Geometry
Voxelization
CPU
Thread
Mapping
Ray Generation
Ray Sorting
Ray
Tracing
Direct Light
Visibility Test
GPU
TM?
Materials
Evaluation

GPU Maxwell



GPU TECHNOLOGY CONFERENCE


GPU TECHNOLOGY CONFERENCE
Maxwell

Agenda

- Maxwell overview
- Why porting to the GPU was challenging
- Performance considerations
- Using the CPU to improve the GPU engine
- Summary

GPU TECHNOLOGY CONFERENCE

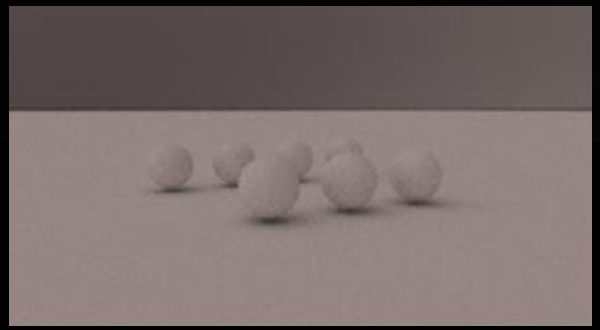
Maxwell

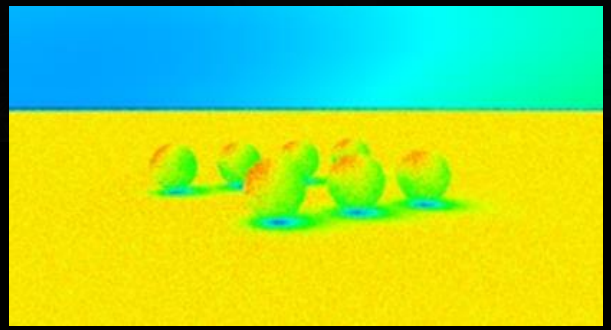
Using the CPU to improve the GPU engine

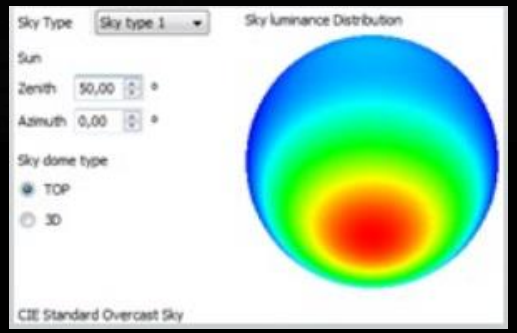
GPU TECHNOLOGY CONFERENCE

Why using our CPU engine as ground truth?

- 12 years old → Stable & Robust
- Used many times for validation purposes







CPU vs GPU Case Studies

GPU TECHNOLOGY CONFERENCE

Guggenheim scene

Teapot scene

Maxwell

Guggenheim Scene

GPU TECHNOLOGY CONFERENCE

Maxwell

Guggenheim Scene

GPU TECHNOLOGY CONFERENCE

Maxwell

Guggenheim Scene

GPU TECHNOLOGY CONFERENCE

ISSUES

- Slight differences in intensity
- Noise in some areas
- Subtle changes in glossy surfaces

Maxwell

Guggenheim Scene

GPU TECHNOLOGY CONFERENCE

STRATEGY

- Simplifying & Isolating (surprise :P)
- Automated numerical comparisons
    - -Raytracing text output
- Ray viewer

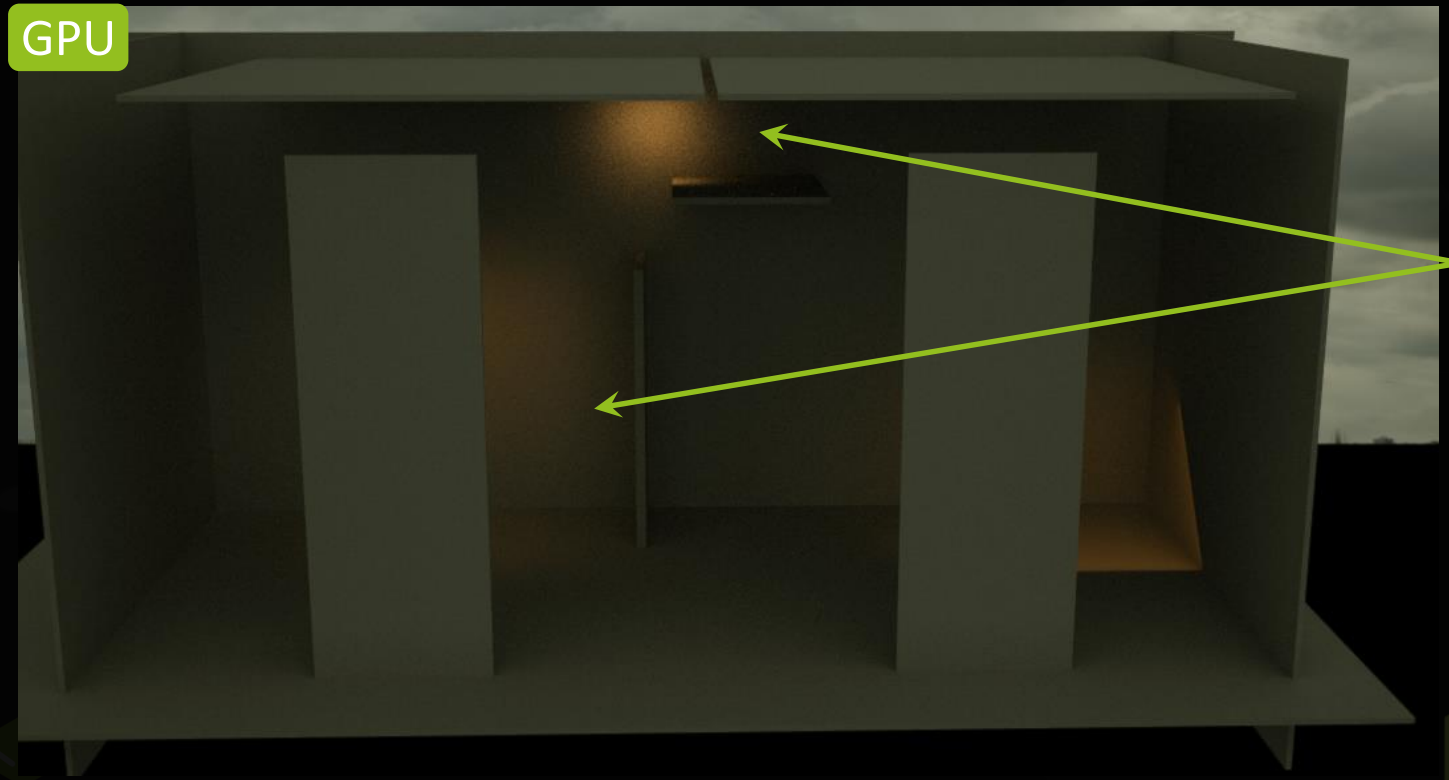
Maxwell

Guggenheim Scene

GPU TECHNOLOGY CONFERENCE

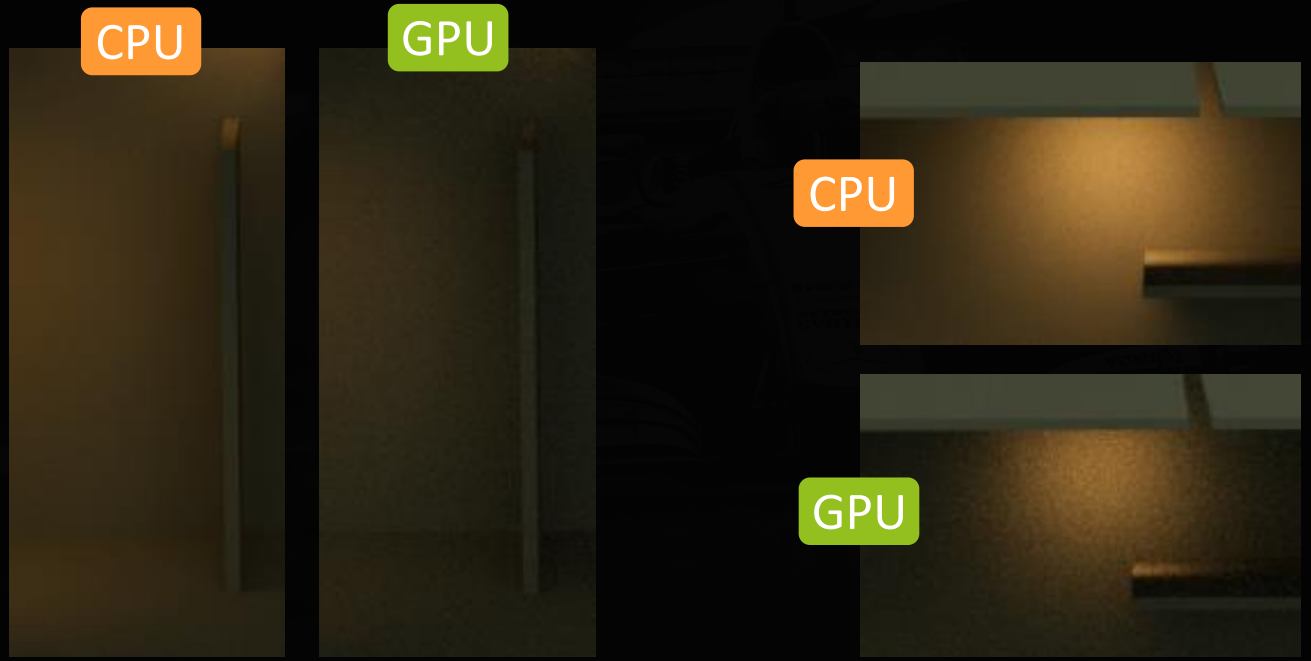
Guggenheim Scene

GPU TECHNOLOGY CONFERENCE



Different Intensity
+
Noise Problems

Guggenheim Scene – Intensity & Noise



GPU TECHNOLOGY CONFERENCE

Maxwell

Guggenheim Scene

GPU TECHNOLOGY CONFERENCE

FINDINGS

- Emitters intensity
  - Hidden property of emitters was not working properly
  - Non-visible emitters were causing occlusions
  - Loss of energy
- Noise
  - QMC had some problems for higher dimensions

Maxwell

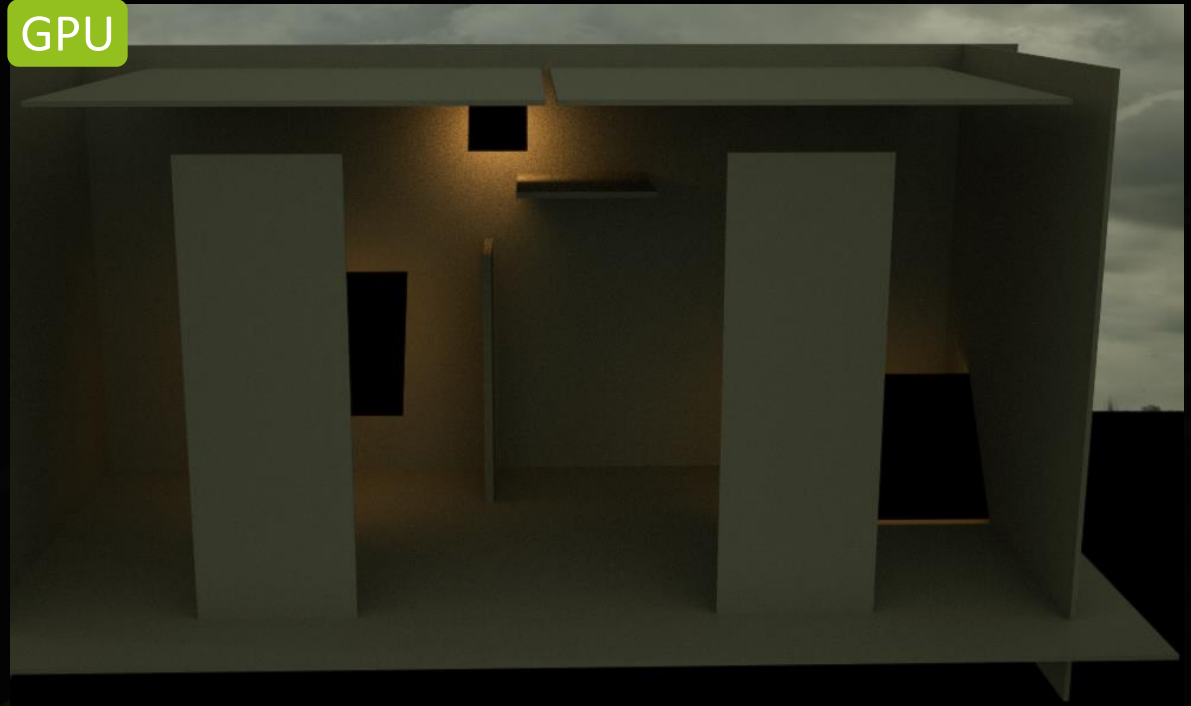
Guggenheim Scene

GPU TECHNOLOGY CONFERENCE

FIXED

Guggenheim Scene

GPU TECHNOLOGY CONFERENCE

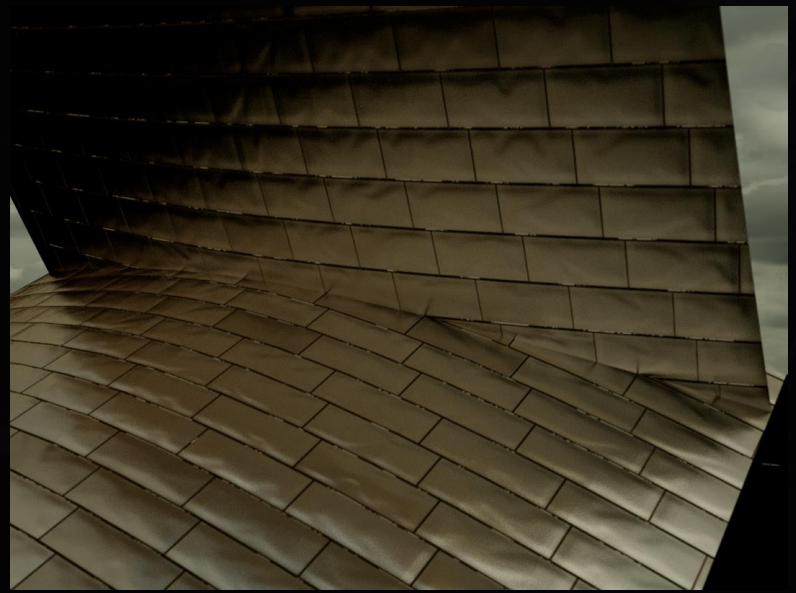


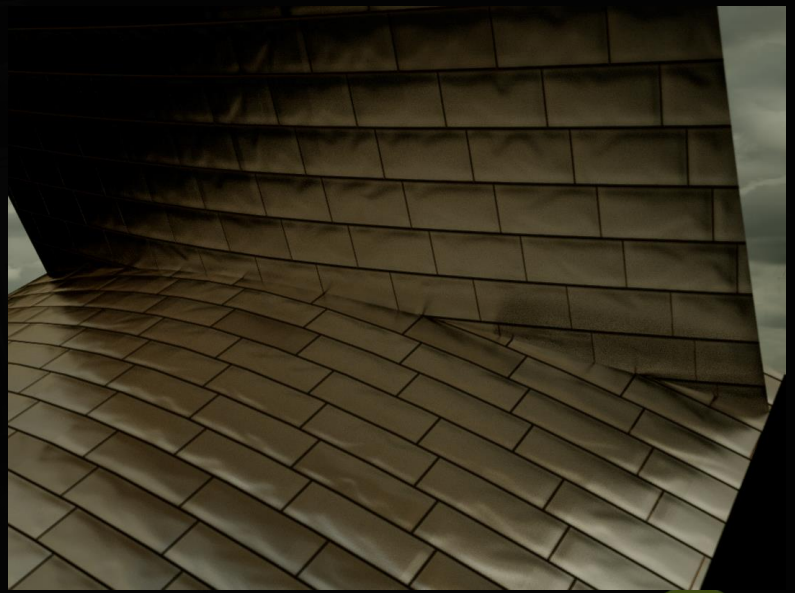
FIXED

Guggenheim Scene

Maxwell

Guggenheim Scene – Differences in glossies

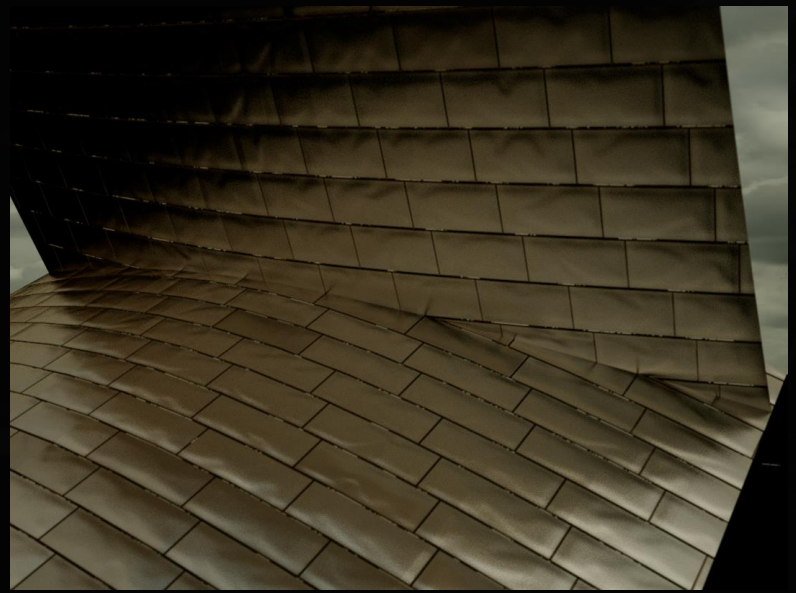




GPU TECHNOLOGY CONFERENCE

Maxwell

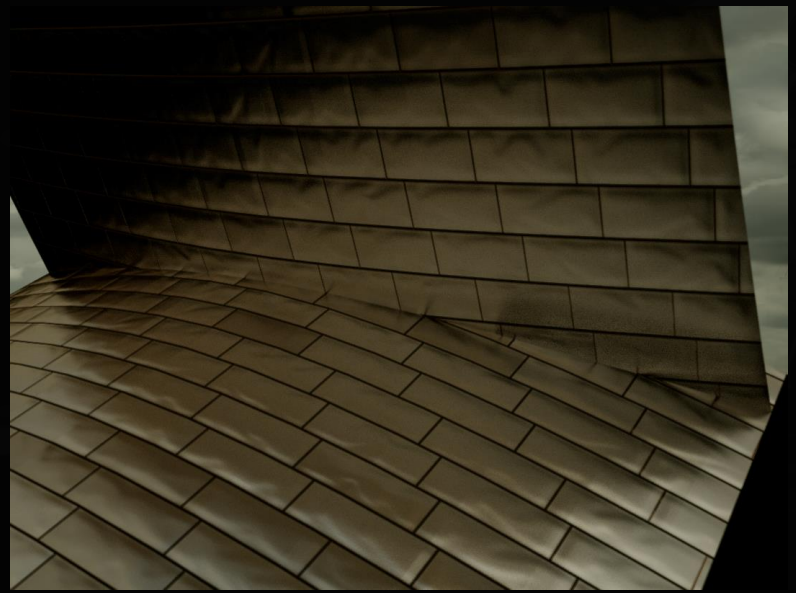
Guggenheim Scene – Differences in glossies



GPU TECHNOLOGY CONFERENCE

Maxwell

Guggenheim Scene – Differences in glossies

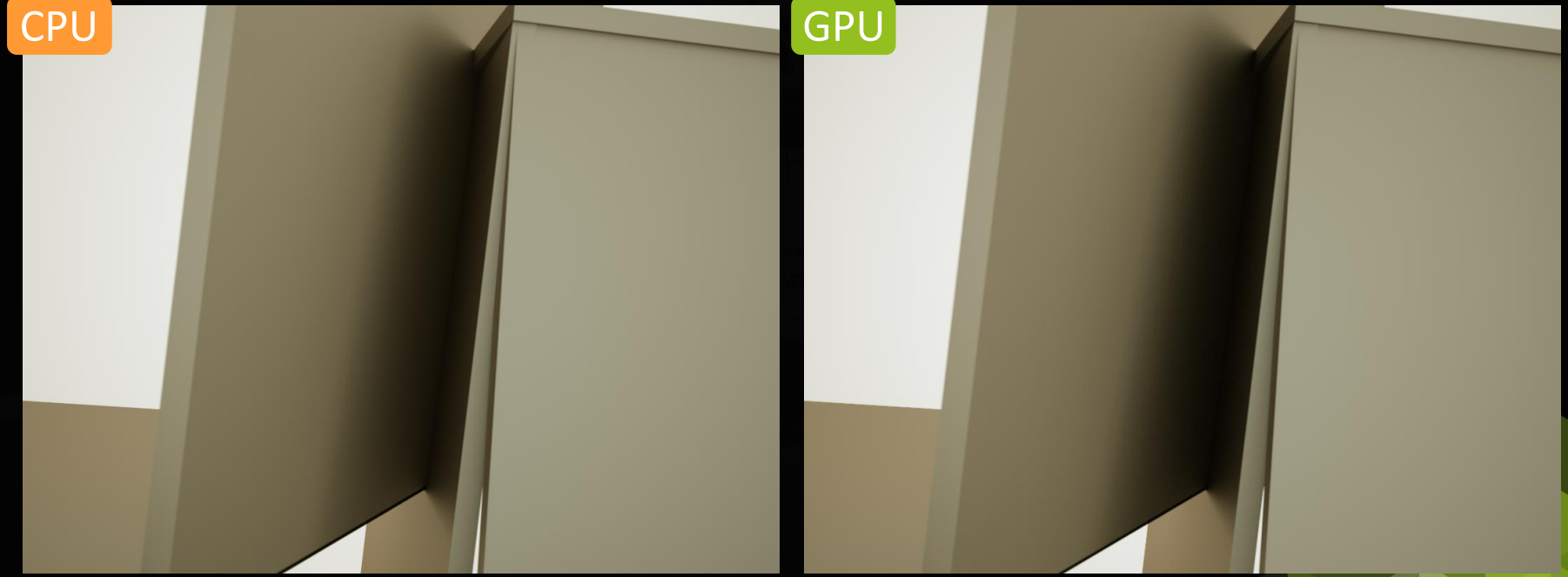


GPU TECHNOLOGY CONFERENCE

Maxwell

Guggenheim Scene – Differences in glossies

GPU TECHNOLOGY CONFERENCE

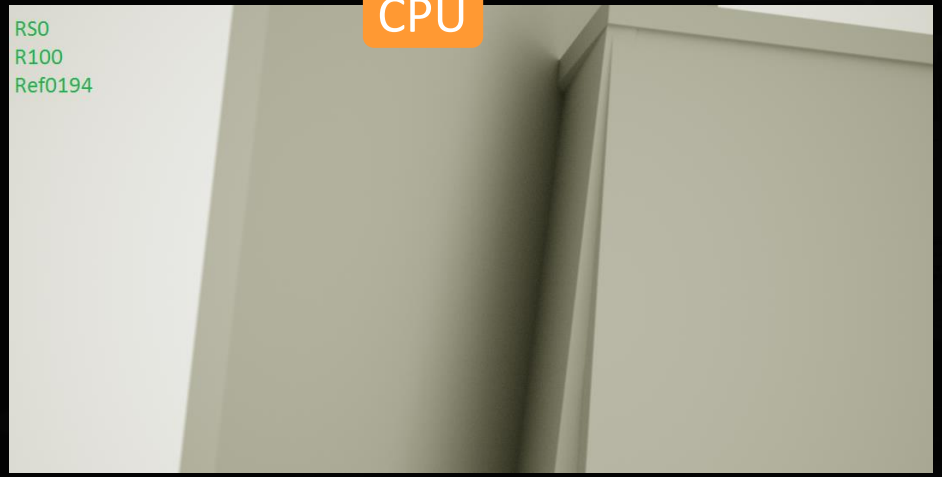


Maxwell

Guggenheim Scene – Differences in glossies

GPU TECHNOLOGY CONFERENCE

CPU

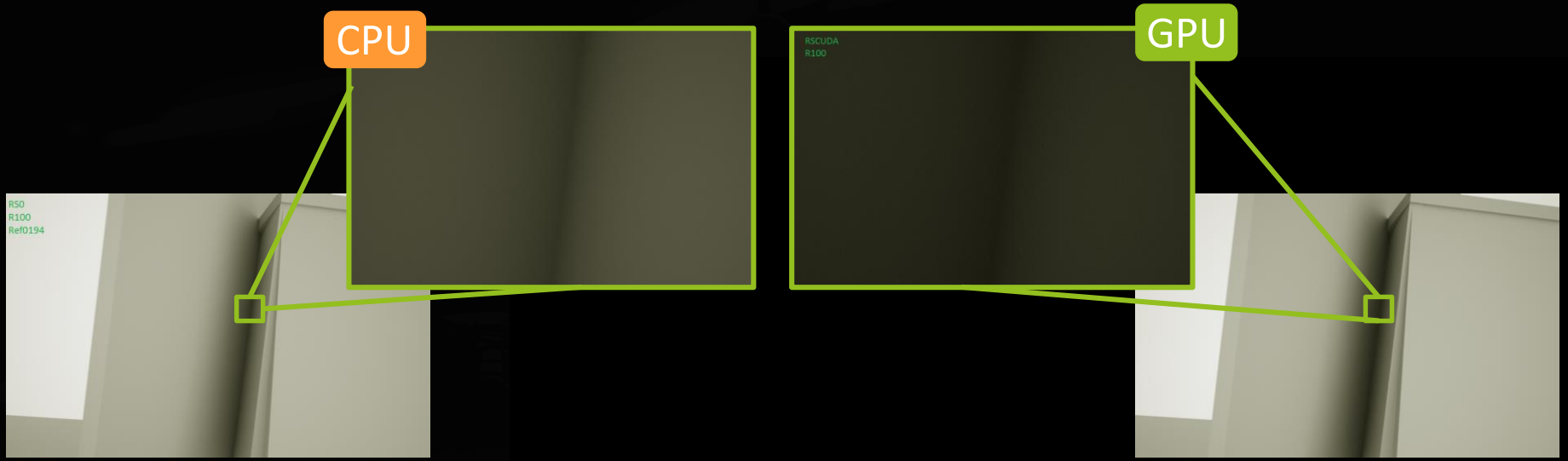


GPU

- Simplify the material → Lambert

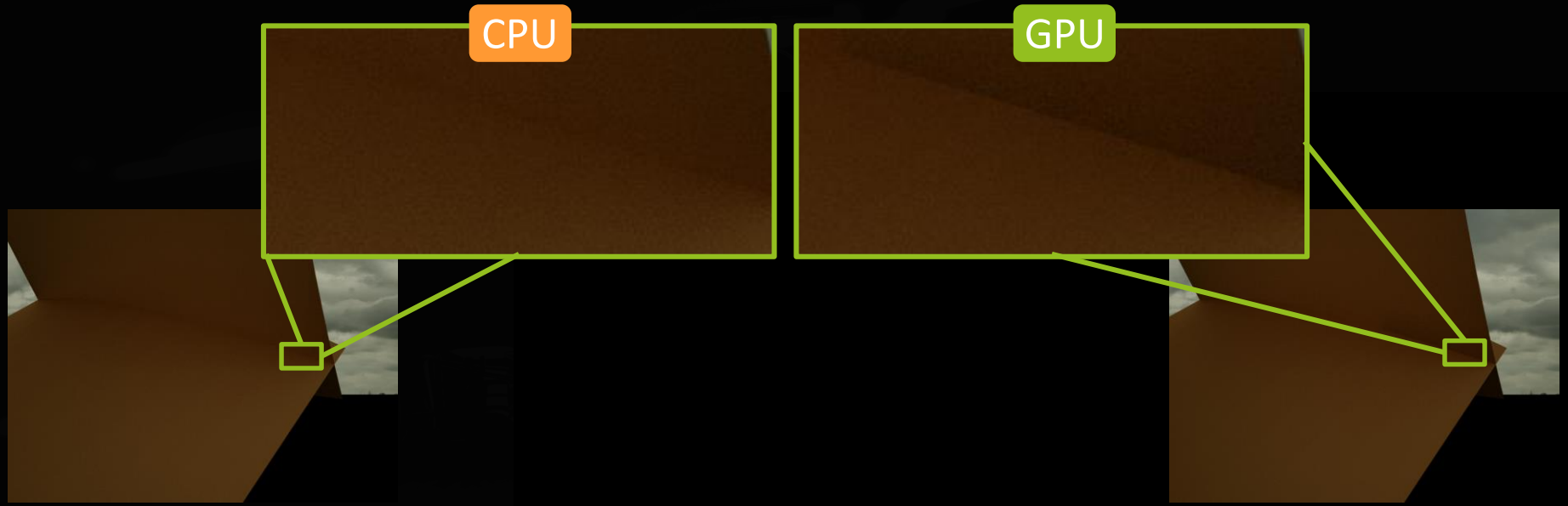
Maxwell

Guggenheim Scene – Differences in glossies



Maxwell

Guggenheim Scene – Differences in glossies



GPU TECHNOLOGY CONFERENCE
Maxwell

Guggenheim Scene – Differences in glossies

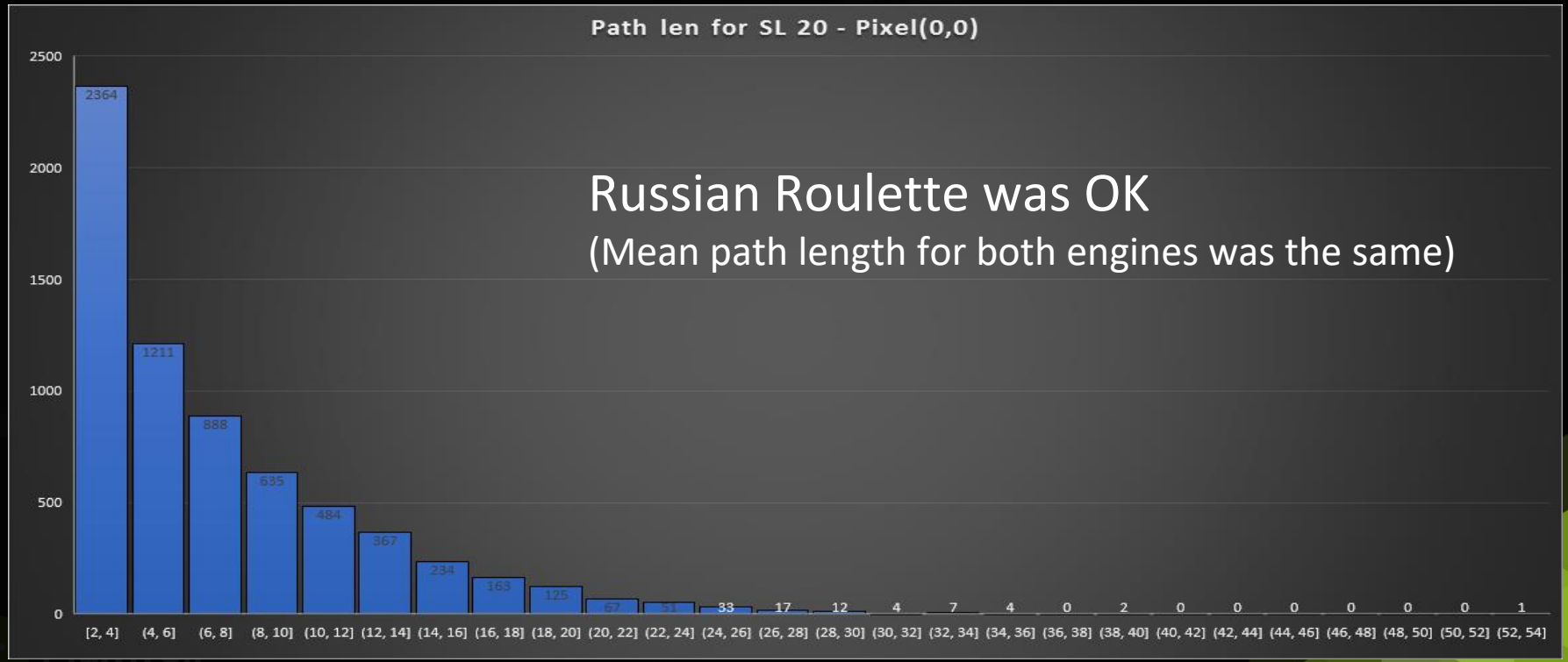
GPU TECHNOLOGY CONFERENCE

- It turned out it was not related to materials
  - Both glossy and lambert have the same problem
  - Difficult to isolate

- Possible problems
  - QMC numbers bug?
  - Russian Roulette bug?
  - Ray / triangle intersection issues with indirect bounces?
  - Energy accumulation problem?
  - Precision issues?
  - …

Maxwell

Guggenheim Scene – Differences in glossies



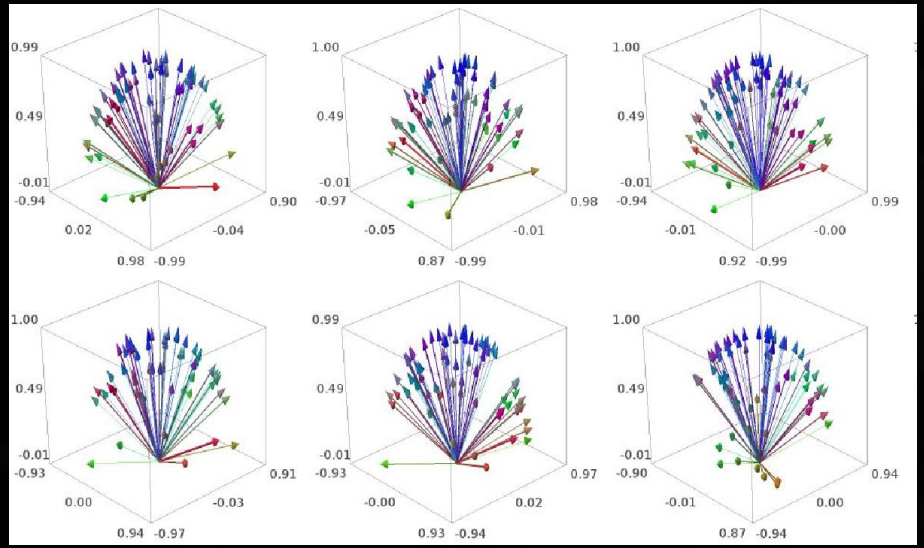
Russian Roulette was OK

(Mean path length for both engines was the same)

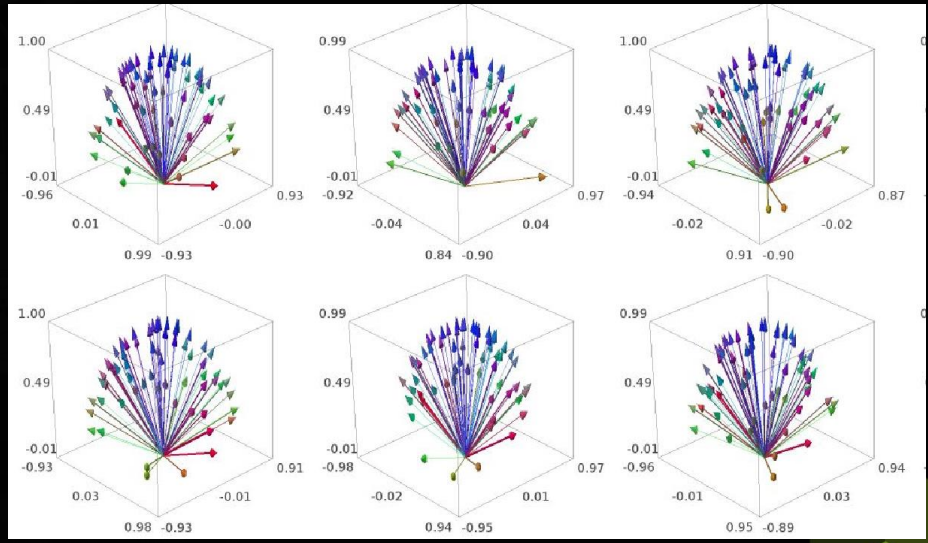
Guggenheim Scene – Differences in glossies

GPU TECHNOLOGY CONFERENCE

CPU – QMC Distributions



GPU – QMC Distributions

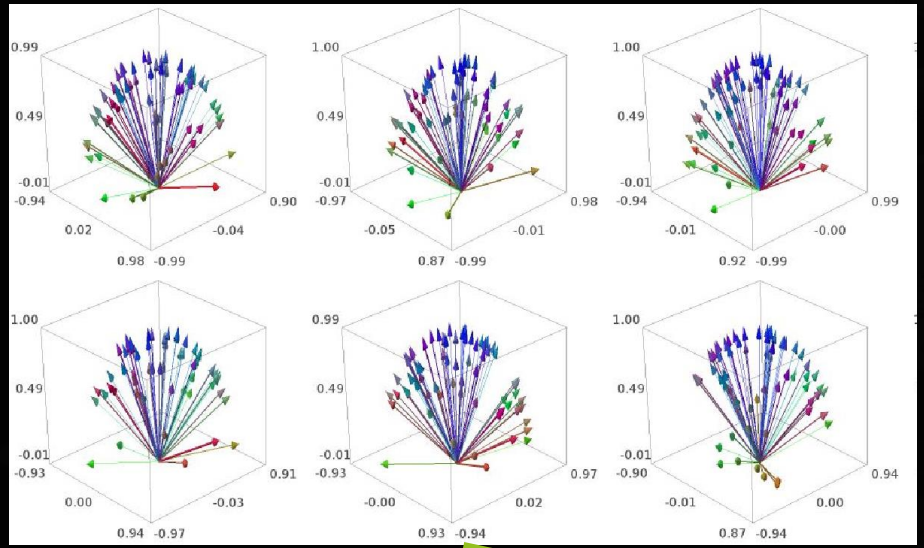


Maxwell

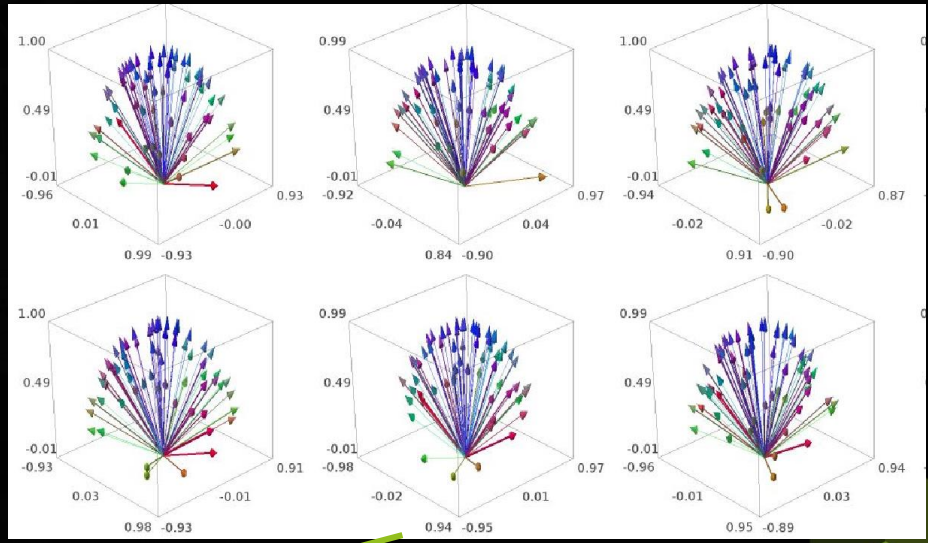
Guggenheim Scene – Differences in glossies

GPU TECHNOLOGY CONFERENCE

CPU – QMC Distributions



GPU – QMC Distributions



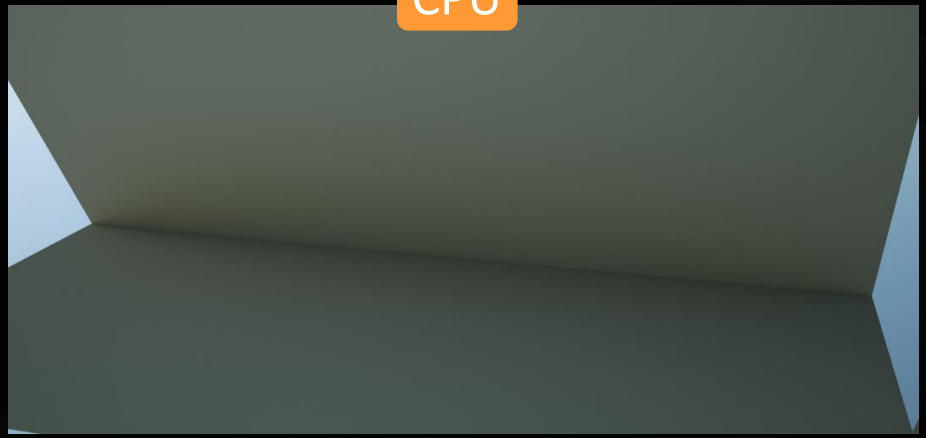
Automated tests detected differences!

Maxwell

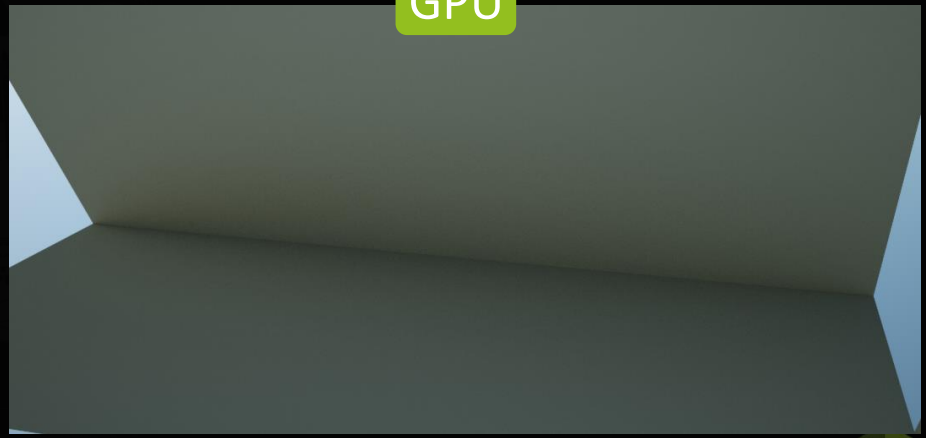
Guggenheim Scene – Differences in glossies

GPU TECHNOLOGY CONFERENCE

CPU



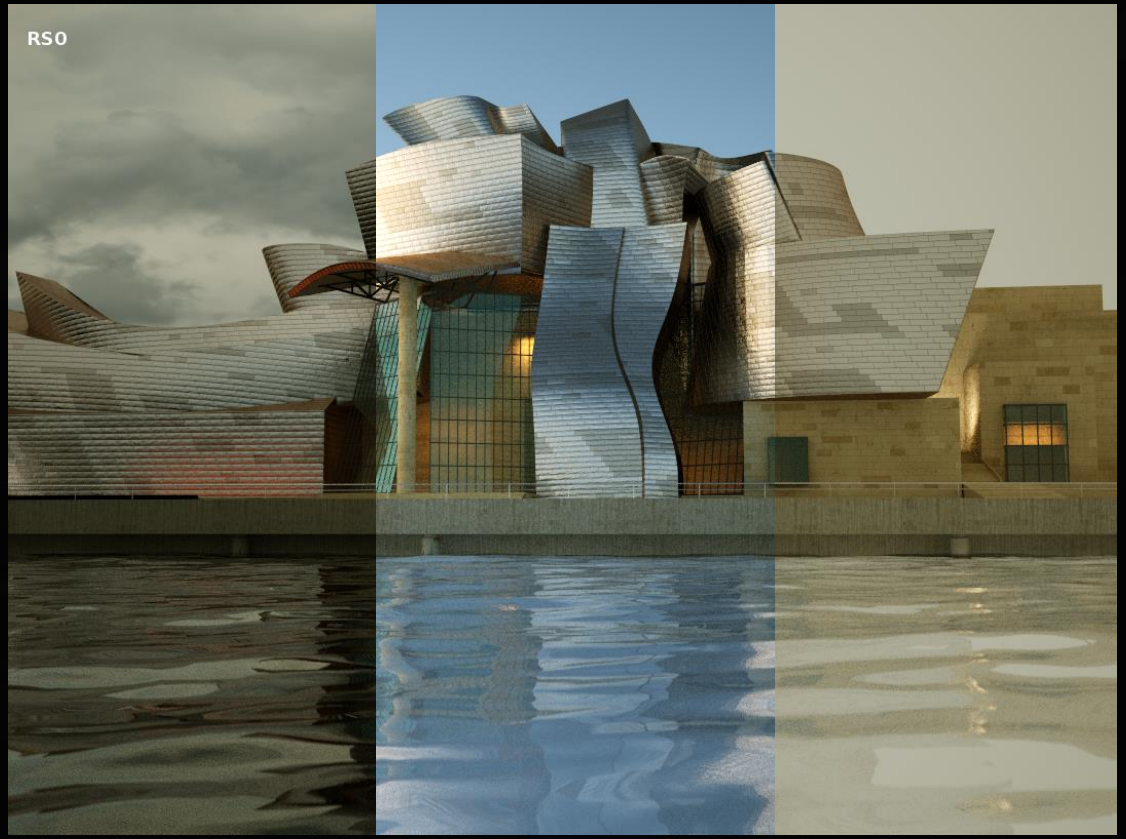
GPU



SOLVED

Maxwell

Guggenheim Scene



CPU == GPU ☺

GPU TECHNOLOGY CONFERENCE

Maxwell


GPU TECHNOLOGY CONFERENCE
Maxwell

Teapot Scene

GPU TECHNOLOGY CONFERENCE

Maxwell

Teapot Scene

GPU TECHNOLOGY CONFERENCE

Maxwell

Teapot Scene

GPU TECHNOLOGY CONFERENCE

Maxwell

Teapot Scene

GPU TECHNOLOGY CONFERENCE

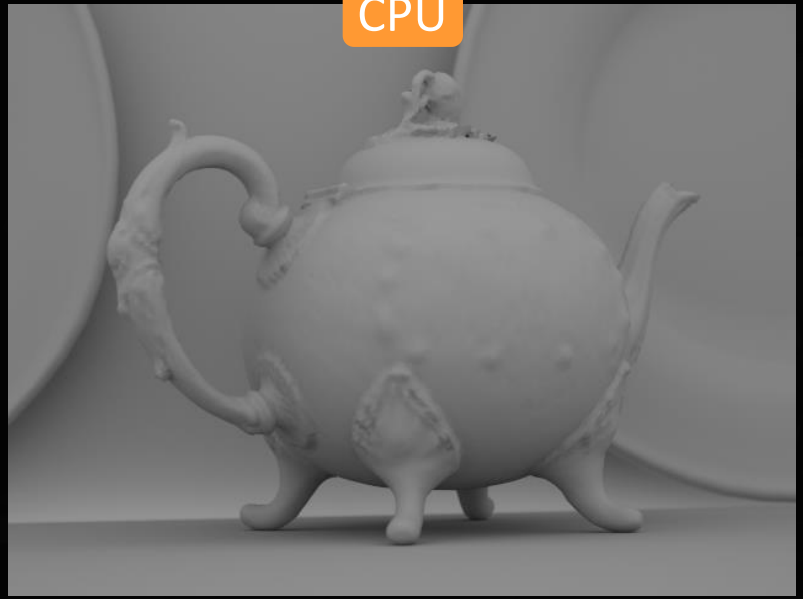
ISSUES

- Subtle differences in bump/normal mapping
- Differences in materials with many layers/bsdfs
- Small changes in intensity

Maxwell

Use cases where CPU Maxwell helped… A LOT!!!

GPU TECHNOLOGY CONFERENCE



Test 1 : Lambert materials + Constant Sky → OK

Maxwell

Teapot Scene

GPU TECHNOLOGY CONFERENCE

Test 2 : Added textures + Normal maps → WRONG

Maxwell

Teapot Scene

GPU TECHNOLOGY CONFERENCE

Test 3 : Added multilayered materials → WRONG

Maxwell

Teapot Scene

GPU TECHNOLOGY CONFERENCE

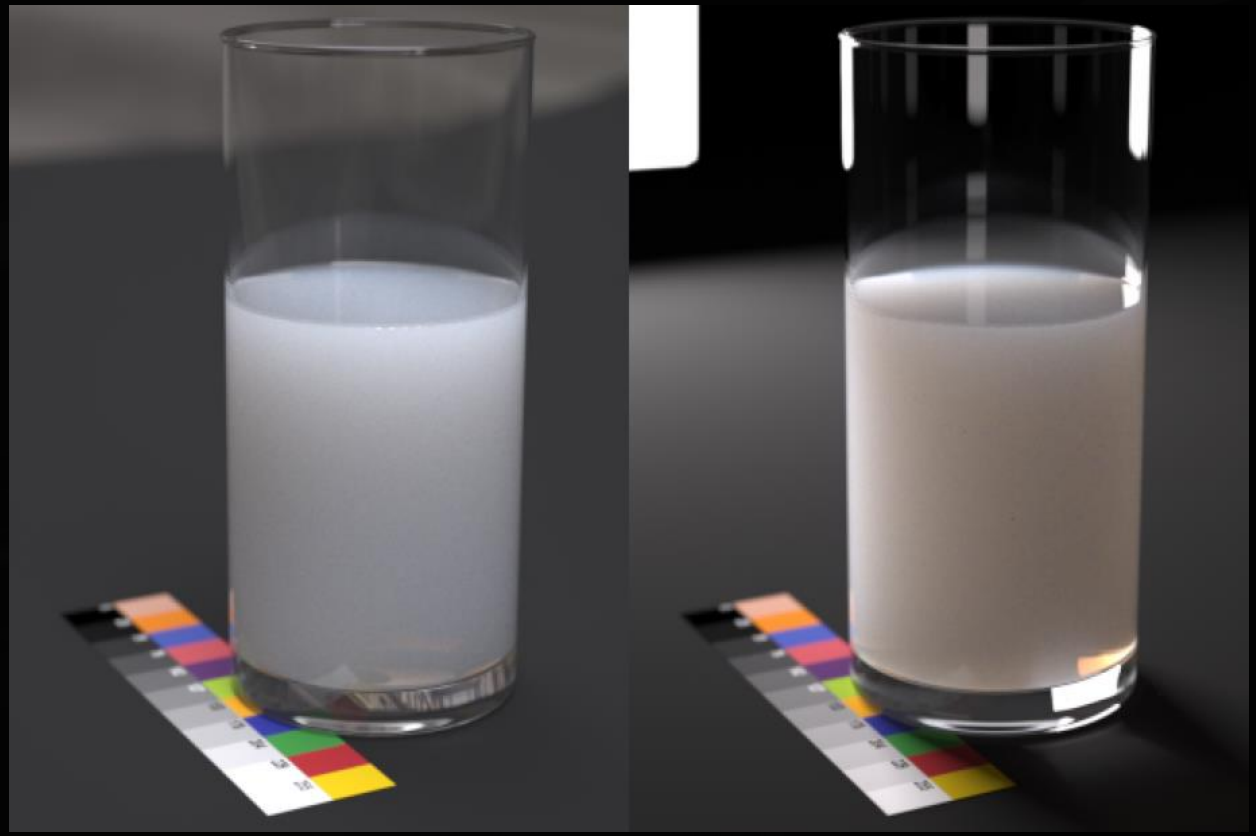
FINDINGS

- Automated CPU vs GPU numerical comparisons were **key**
  - Rays reaching IBL were not accumulating energy properly
  - Multilayered weights were not properly computed
    - Bug introduced when porting CPU **optimized** code
  - Precision issues creating TBN bases (Affected bump/normal mapping)

Maxwell


GPU TECHNOLOGY CONFERENCE
Maxwell

Next Steps Unbiased, GPU friendly SSS



Agenda

- Maxwell Render overview
- Why porting to the GPU was challenging
- Performance considerations
- Using the CPU to improve the GPU engine
- Summary

GPU TECHNOLOGY CONFERENCE

Summary

Main sources of bugs:

- CPU optimized code not easy to port
- Refactoring to make code GPU friendly
- Precision issues with some math operators

GPU TECHNOLOGY CONFERENCE

Maxwell

Summary

GPU TECHNOLOGY CONFERENCE

- **90% of the complexity of Maxwell already ported**
  - -Very happy with the results: Speed boost: 5x-15x
  - - CUDA made it possible
- **Validating using a ground truth renderer**
  - - Was painful
  - - 100% worth in the long run (quality first, speed second)

Maxwell

Maxwell

NEXT LIMIT

Thanks!

Juan Cañada
Head of Visualization
Next Limit Technologies

GPU TECHNOLOGY CONFERENCE

Maxwell