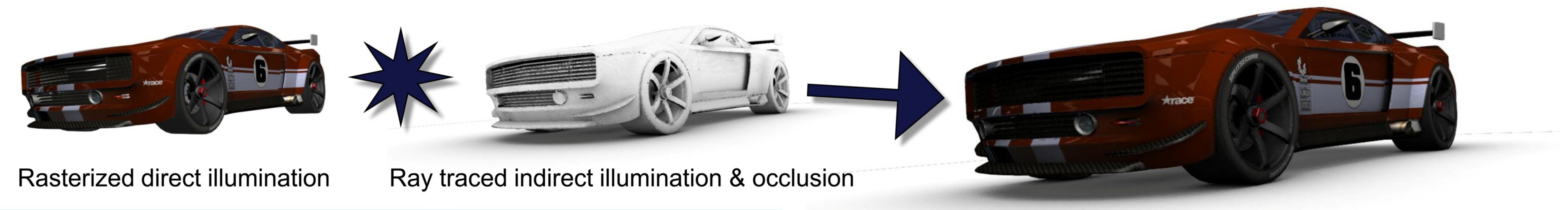
Bridging Ray and Raster Processing on GPUs

Kenny Mitchell, Black Rock Studio
Christian Oberholzer, ETH Zurich
Jan Kautz, University College London
Peter-Pike Sloan, Disney Interactive Studios



Rasterized direct illumination

Ray traced indirect illumination & occlusion

Combined result rendered at 19.8Hz (50.5ms) on a GeForce 280 GTX

ABSTRACT

Introduction

Real-time graphics is dominated by hardware accelerated rendering pipelines based on rasterization. Such pipelines are the result of many focused years of development exploiting the locality and coherence of frame buffer aligned pixel quad operations with many parallel processing units. Local illumination and material shading models are handled effectively with this model, optimized for consistent scene traversals in image space. Real-time global illumination (GI), however, requires rapid traversal of scene structures beyond image space, often incoherent with poor locality in divergent sampling directions.

……

We combine many carefully selected measures to develop a system for bridging ray tracing and rasterization hardware processing. In passing the results of a global illumination ray tracing step using OptiX [1], to a rasterization step, we adhere to the strengths of each. Above we see the results of our application of each step and their combined outcome. In a scalable framework, we employ secondary ray visibility sampling for ambient occlusion (AO) and indirect illumination for color bleeding effects.

DESIGN

System

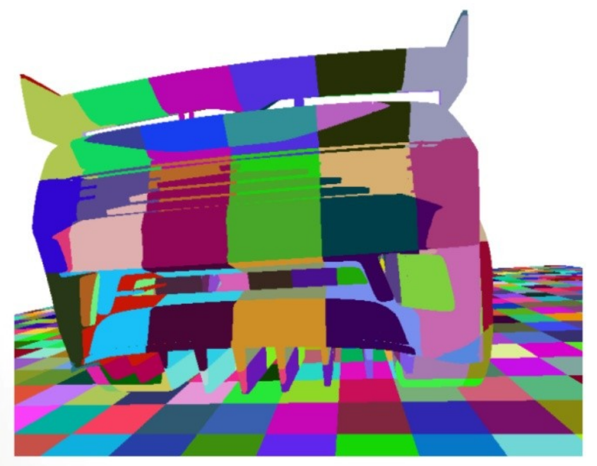
A deferred shading process [2], which is often used in video game rendering engines to reduce lighting and shading calculations to only visible pixels, generates G-Buffer albedo, positions and normals for use in both visible ray spawn cell location and the raster shading processes.

……

A low-frequency meshless 3D volume grid of cells acts as an irradiance cache containing values ranging from scalar occlusion and tri-stimulus color bleeding terms to a directional spherical harmonics irradiance basis.

……

Each cell's ray samples incur a cost that must be minimized. For real-time frame rates we apply an exponential weighted averaging of samples that amortizes ray tracing costs over a number of frames. Indeed, the size of cache, length of traced rays, sampled indirect surface illumination calculation and mesh geometric level of detail may each be scaled to contribute further accelerated processing at the cost of reduced illumination accuracy.



Depiction of cache ray spawn cells

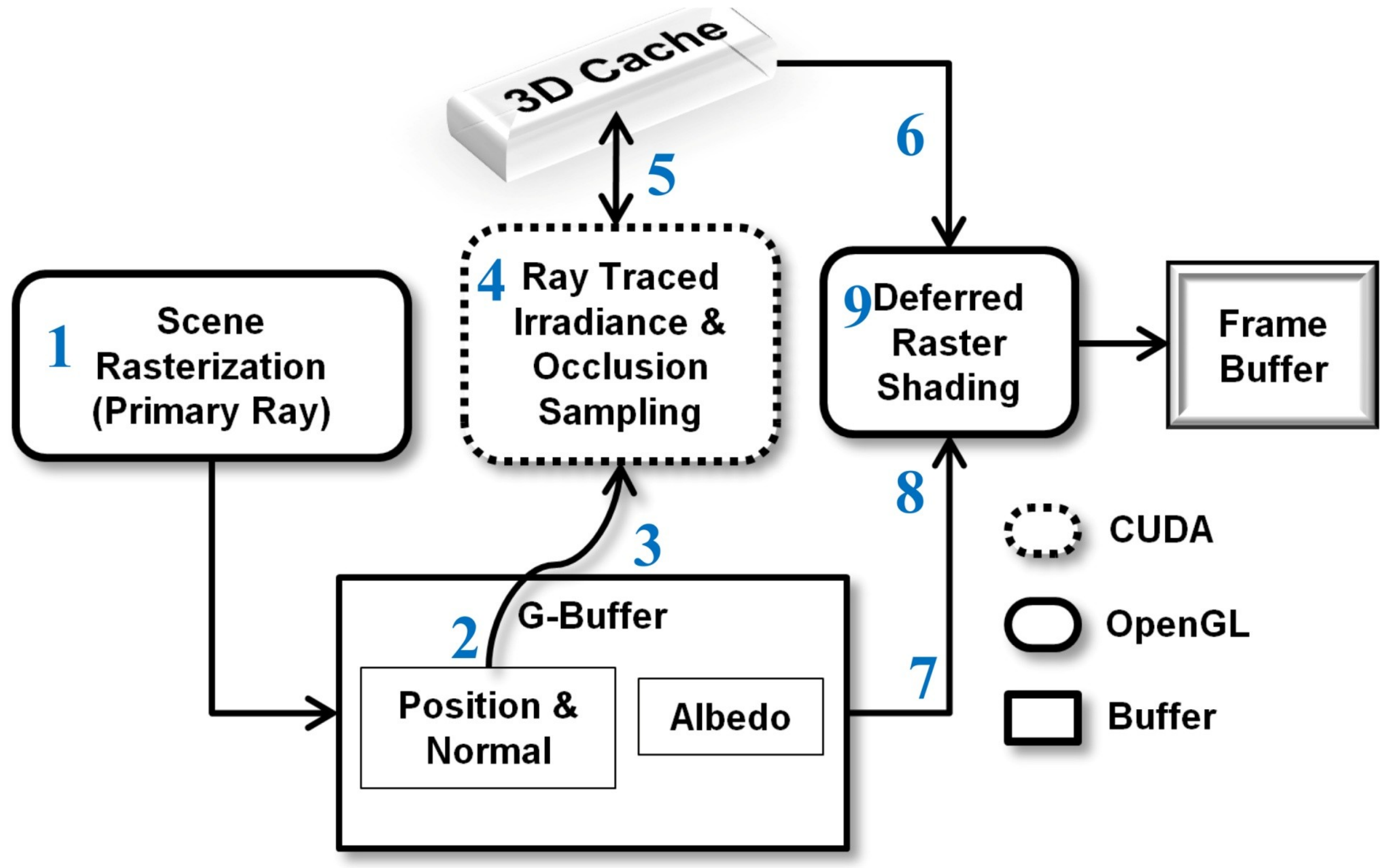
HYBRID

Parallel Processing Steps

We perform parallel processing in the selection of only visible ray spawn cells and sampling from the 3D cached illumination into 2D image space for the deferred rasterization step. Upsampling the 2D irradiance texture with an edge aware cross bilateral filter [3] retains geometric detail in the process of achieving a smooth filtering of noisy irradiance cache entries. Instead of two additional raster passes, we apply a single horizontal filter pass and vertical sampling is incorporated into the deferred shading pass.

Hybrid Ray and Raster Processing Steps

```
1 rasterize scene into G-buffer
2 downsample G-buffer to lower resolution
3 select only visible cells for update
4 forall cell in visible set of 3D cells do
5     ray trace secondary irradiance sampling from cell
6 collect 3D visible cell results into 2D image space
7 rasterize image space global illumination
8 upsample combined blend of 6 and 7 results
9 rasterize deferred local shading with G-buffer and 8
```

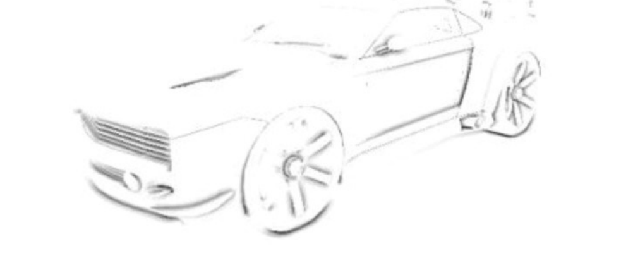


A 3D irradiance cache bridges ray and raster processing steps

Ray and Screen Space Global Illumination

We apply a further image space illumination pass developed from screen space ambient occlusion techniques [4] to accomplish high-frequency near field occlusion and color bleeding. We combine this with the low frequency ray traced illumination results using MIN and MUL blending operators to realize images with both wide and near field illumination effects.

Wide ray occlusion



Near screen space occlusion

SCALABILITY

Level of Detail

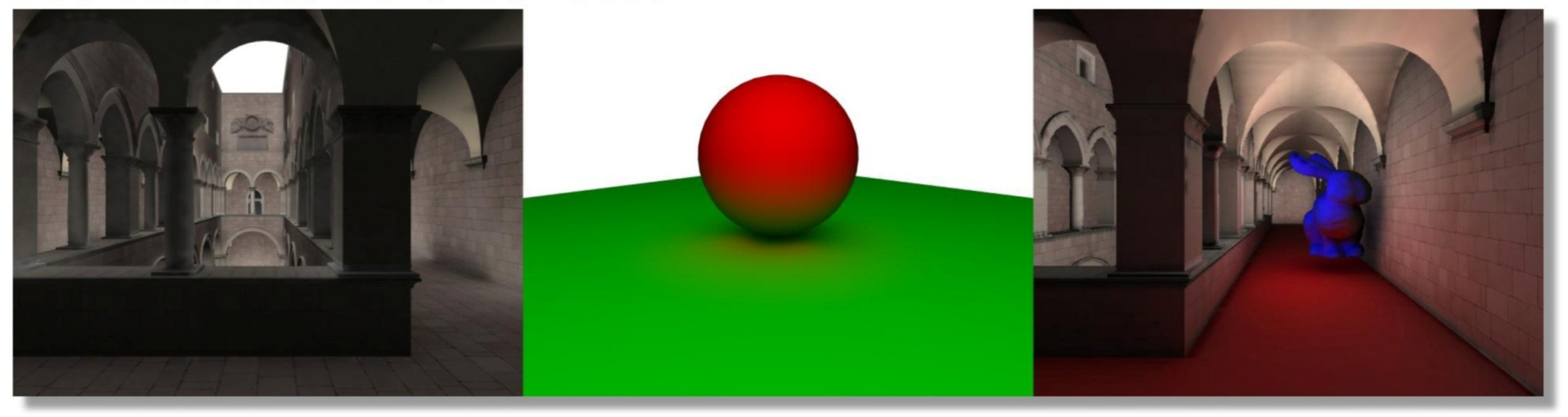
Processing only the level of geometric sophistication necessary to recover low-frequency indirect illumination is sufficient for real-time application of this method as shown opposite. Differences are discernable upon close inspection, but may be improved with mesh reduction tuned towards preservation of topology and volume. Level of refinement adjustments may also be applied to the number of ray samples applied, with fewer ray samples resulting either in increased temporal noise or delay in visual response time for illumination updates.

40k triangles in both ray & raster meshes

4k triangles in ray traced mesh

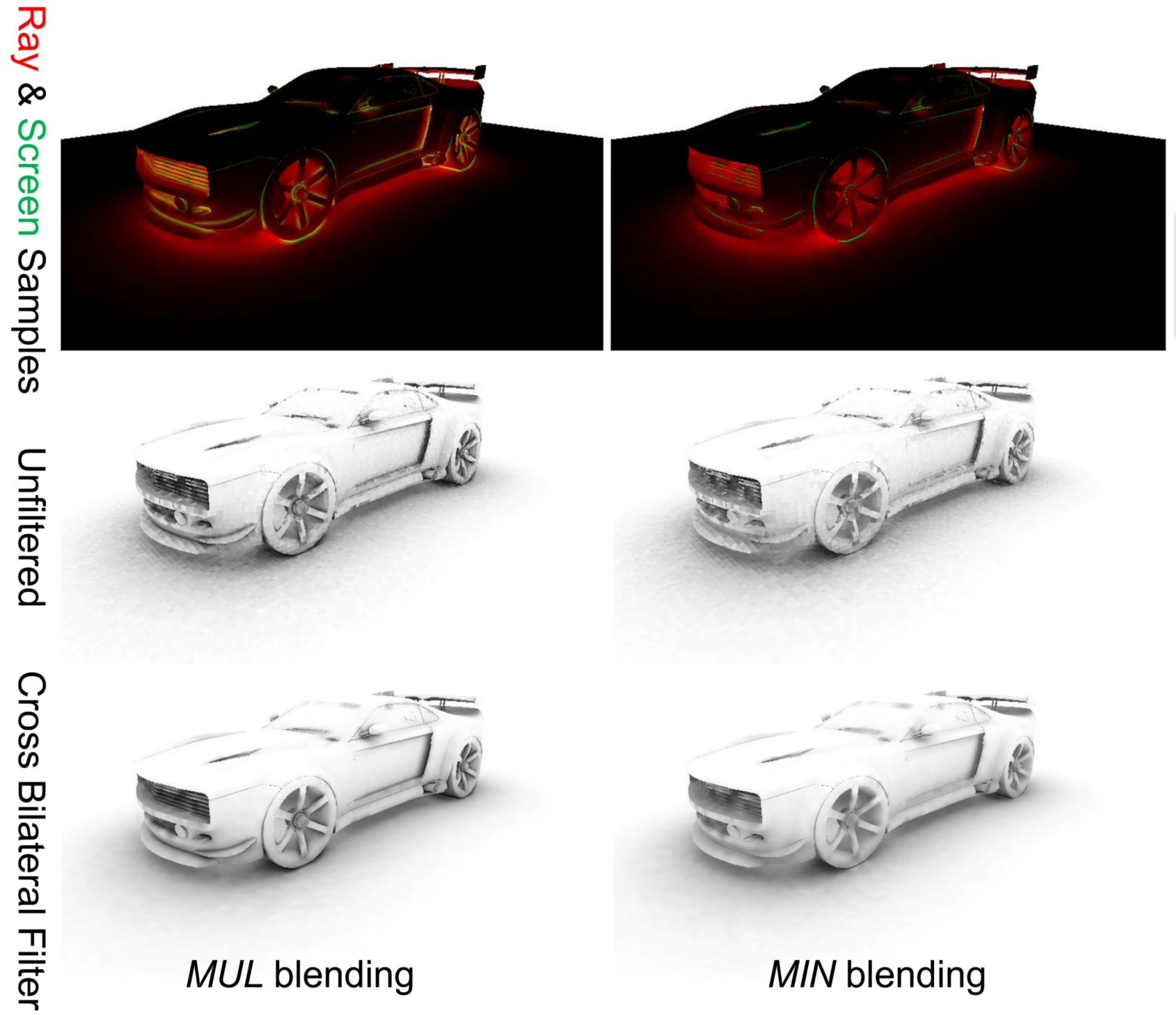
4k triangles in both ray & raster meshes

ANIMATION



Localized Cache Updates

As global illumination sampling is continuously updated, the system is applicable to scenes with animated light sources and deformable geometry. By artificially limiting the range of indirect illumination to a quadratic fall-off, a reduced volume of animated cache updates may be applied as in the squashing ball and bouncing bunny in Sponza animations above.



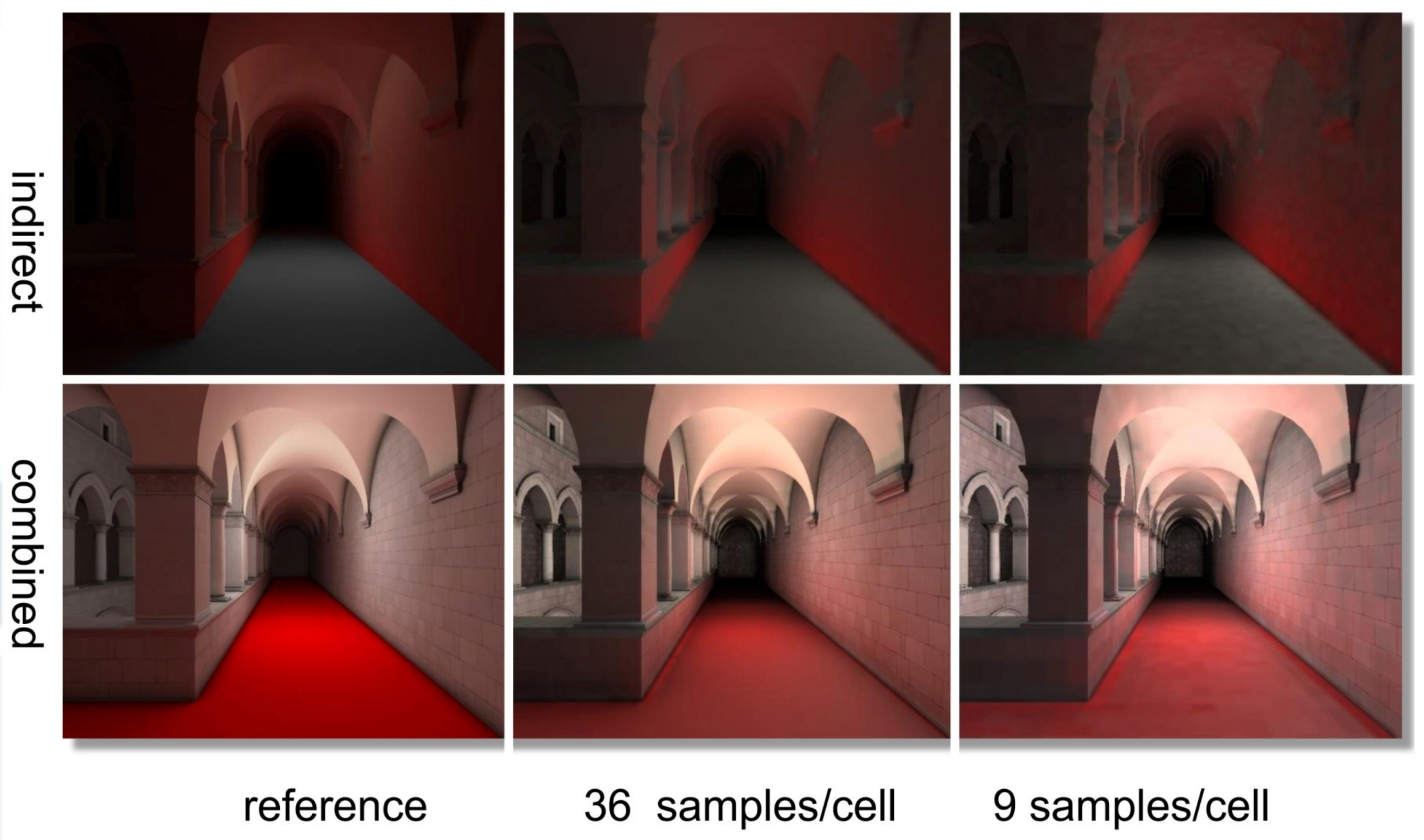
MUL blending

MIN blending

RESULTS

Comparison

Our Sponza renders exhibit some faults at this stage. A sampling discontinuity on the near right archway is due to high contrast light and dark surfaces within the cell. Such artifacts are addressable with irradiance gradients or adaptive resolution of the cache subject to efficient hardware methods. Further, our irradiance fall-off factor illuminates the far end of the corridor and walls greater than the reference render. We expect such over estimation of light transport to occur due to a direct albedo sampling optimization. Other sources of error arise from discrete irradiance cache records, bilateral smoothing, and image space global illumination. However, such measures contribute significantly to the real-time performance of the system.



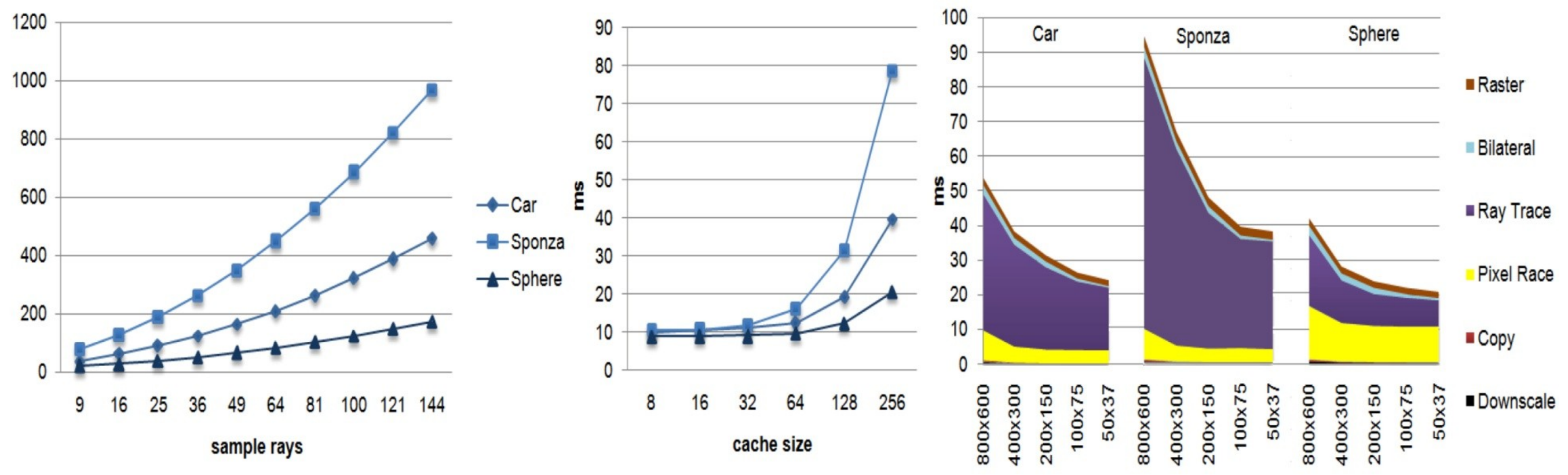
reference

36 samples/cell

9 samples/cell

Measurements

Extensive timing capture data permits us to plot graphs of performance according to number of sample rays, cache size and ray traced resolution for each of our test car, Sponza and sphere scenes. These show a majority of time spent in ray processing, scaling linearly with the number of ray samples.



REFERENCES

[1] PARKER S., BIGLER J., DIETRICH A., FRIEDRICH H., HOBEROCK J., LUEBKE D., MCALLISTER D., MCGUIRE M., MORLEY K., ROBISON A., STICH M.: *OptiX: A General Purpose Ray Tracing Engine.* In ACM Transactions on Graphics (SIGGRAPH) 29, 3, (2010).

[2] THIBIEROZ N.: *Deferred Shading with Multiple-Render Targets.* In ShaderX2 (2003), 251-269.

[3] SLOAN P.-P., GOVINDARAJU N., NOWROUZEZAHRAI D., AND SNYDER J.: *Image-Based Proxy Accumulation for Real-Time Soft Global Illumination.* In Pacific Graphics (2007), vol. 25, 8, 97–105.

[4] KAJALIN V.: *Screen Space Ambient Occlusion.* In ShaderX7 (2009), 413-424.

This work was supported by a hardware donation from NVIDIA.