

Compute Shader 入门精要



Optimize your game using compute shader

凯奥斯

Compute Shader 入门精要

- 概念
- 语法
- 用途

概念

介绍一下背景知识

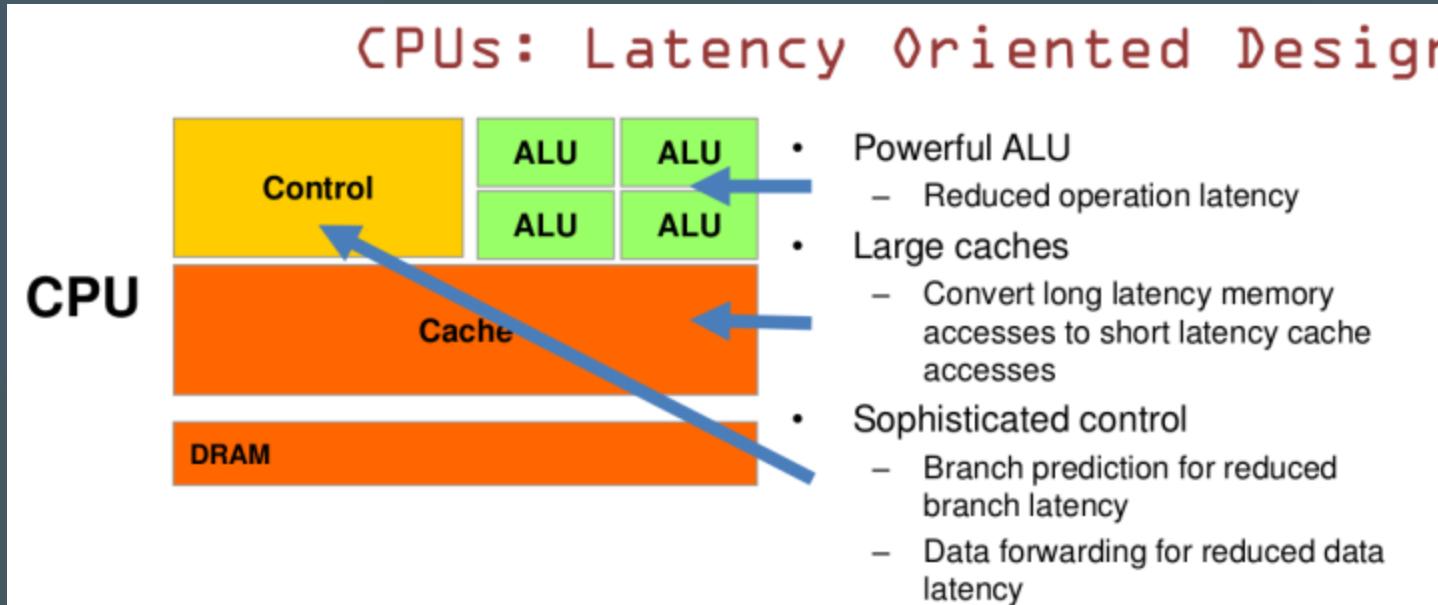
GPGPU

Plain Old
CPU Processing

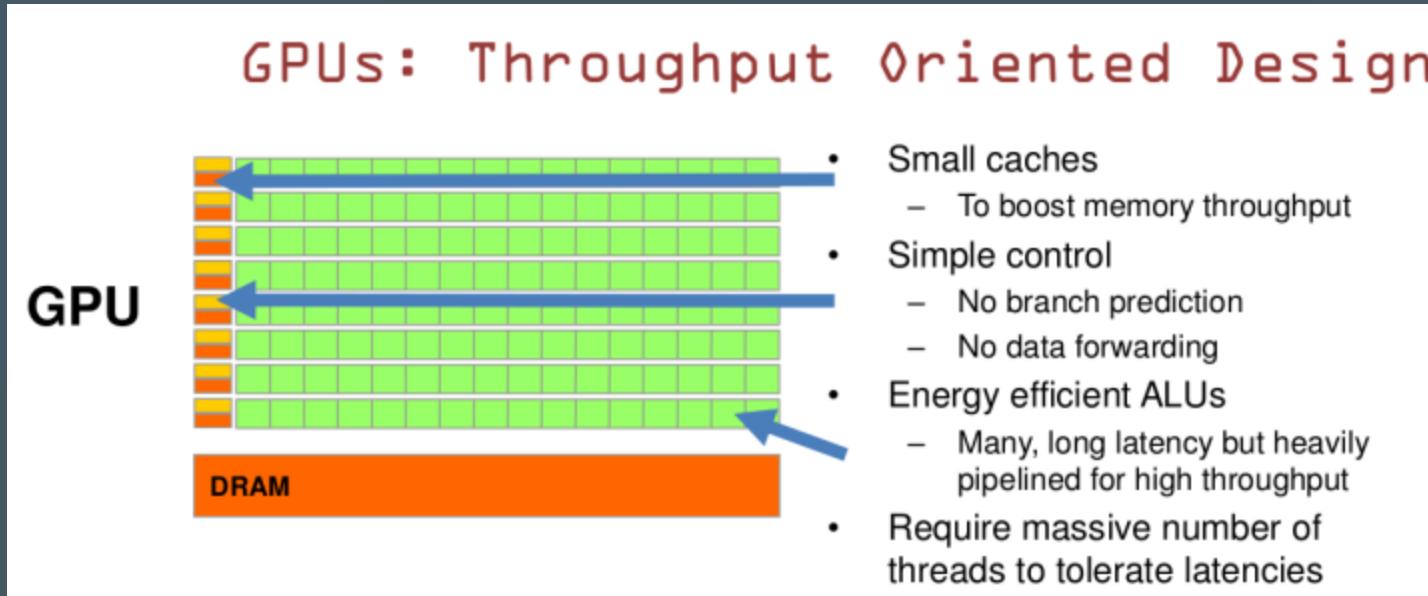
GPGPU
Processing



CPU是基于低延迟的设计



GPU是基于大吞吐量的设计



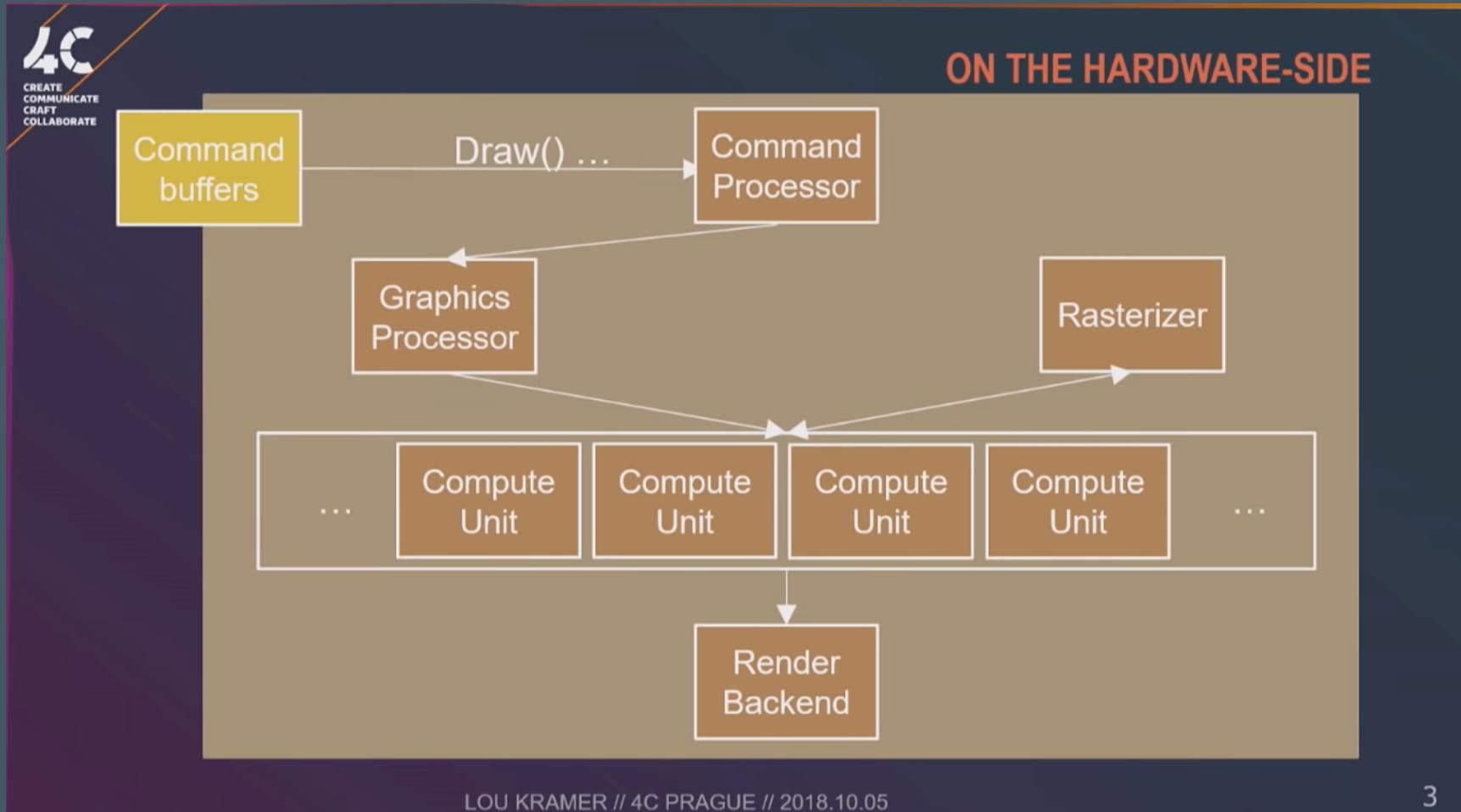
支持Compute Shader的图形API



Compute管线与图像管线的对比

Graphics pipeline	Compute pipeline
One to several shader stages (VS, HS, DS, GS, PS) Input assembler Tessellation Rasterizer ...	CS shader stage

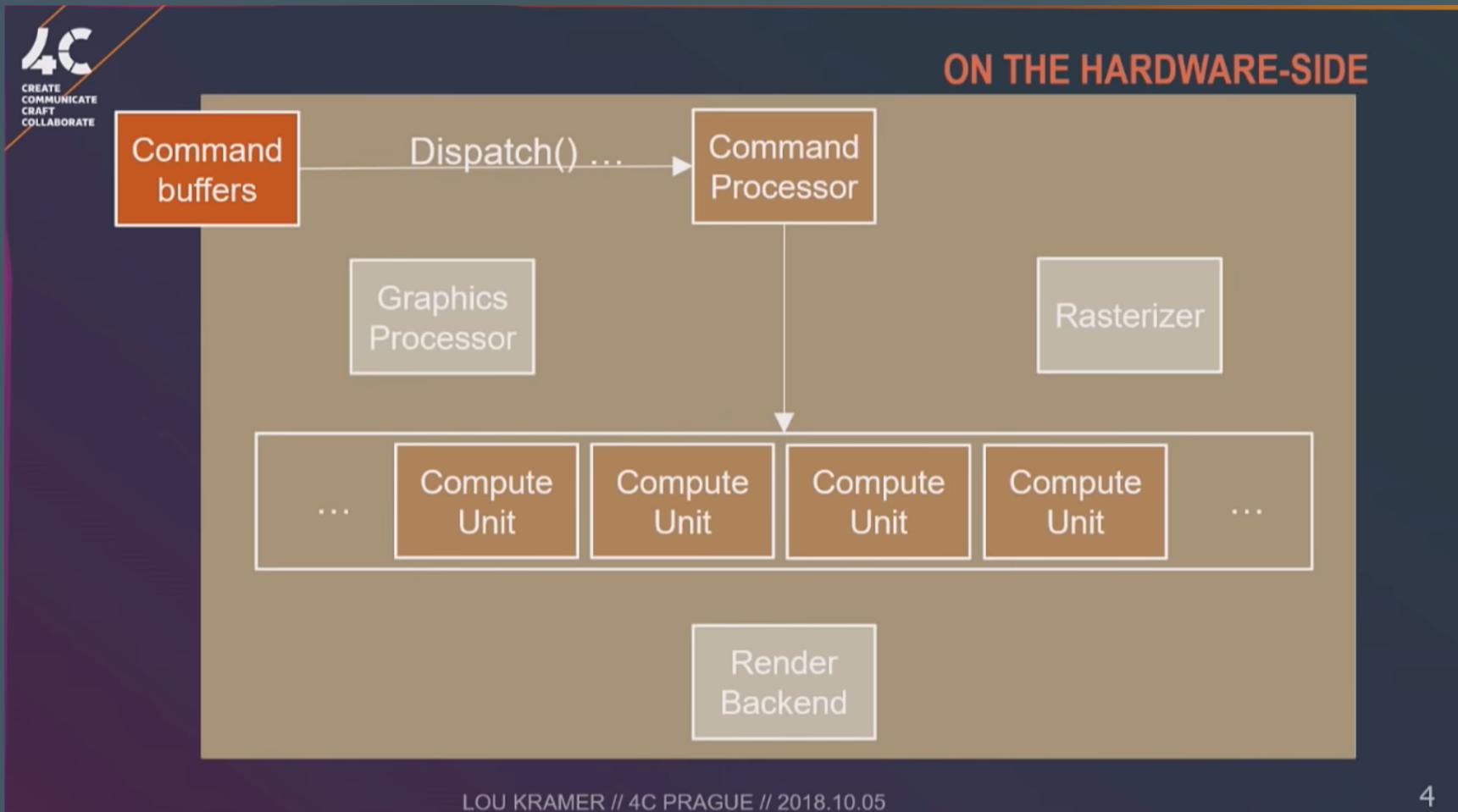
渲染管线（硬件端）



LOU KRAMER // 4C PRAGUE // 2018.10.05

3

计算管线（硬件端）



LOU KRAMER // 4C PRAGUE // 2018.10.05

4

语法

如何在**Unity**里使用**Compute Shader**

kernel

```
// test.compute
#pragma kernel FillWithRed

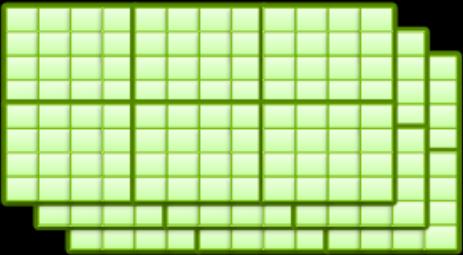
RWTexture2D<float4> res;

[numthreads(8,8,1)]
void FillWithRed (uint3 dtid : SV_DispatchThreadID)
{
    res[dtid.xy] = float4(1,0,0,1);
}
```

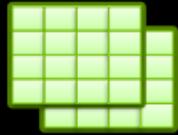
Dispatch

```
public void Dispatch(int kernelIndex,  
                     int threadGroupsX,  
                     int threadGroupsY,  
                     int threadGroupsZ);
```

numthreads



Dispatch: 3D grid of thread groups. Hundreds of thousands of threads.



Thread Group: 3D grid of threads. Tens or hundreds of threads.

numThreads nX, nY, nZ

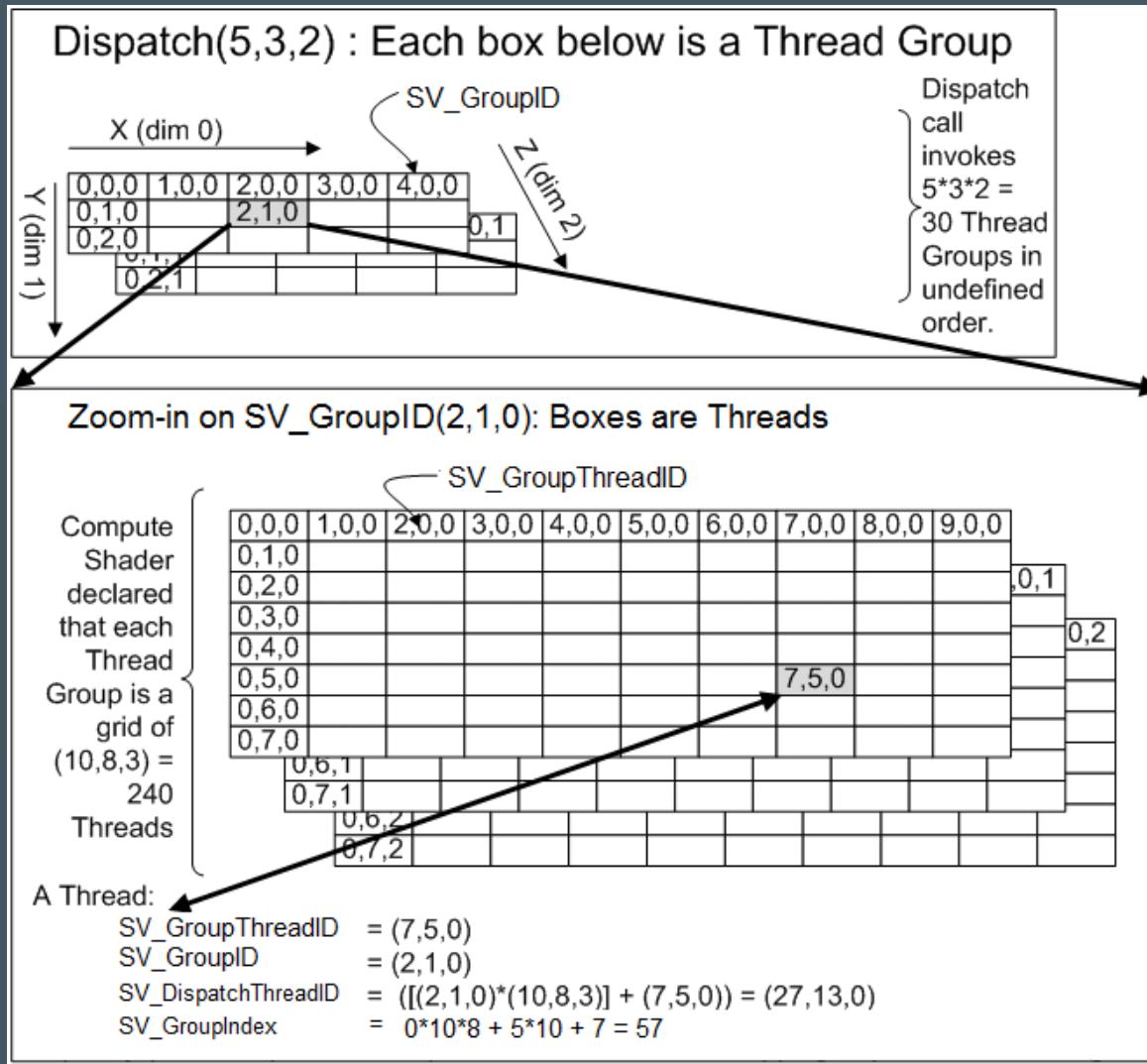


Thread: One invocation of a shader.

`SV_DispatchThreadID`,
`SV_GroupThreadID`,
`SV_GroupID`

PRESENTED BY  NVIDIA.

thread groups



Buffer & Texture

GPU Side	CPU Side
*StructuredBuffer	ComputeBuffer
Texture*D	Texture
RWTexture*D	RenderTexture

groupshared

使用groupshared可以将一个变量标记为组内共享。
(又叫TGSM)

Barrier

当我们在不同线程访问同一个资源的时候，我们需要使用barrier来进行阻塞。

GroupMemoryBarrier

GroupMemoryBarrierWithGroupSync

DeviceMemoryBarrier

DeviceMemoryBarrierWithGroupSync

AllMemoryBarrier

AllMemoryBarrierWithGroupSync

Interlocked

原子操作， 不会被线程调度机制打断。

- InterlockedAdd
- InterlockedAnd
- InterlockedCompareExchange
- InterlockedCompareStore
- InterlockedExchange
- InterlockedMax
- InterlockedMin
- InterlockedOr
- InterlockedXor

但是只能用于int/uint

平台差异

- 数组越界， DX上会返回0， 其它平台会出错。
- 变量名与关键字/内置库函数重名， DX无影响， 其他平台会出错。
- 如果SBuffer内结构的显存布局要与内存布局不一致， DX可能会转换， 其他平台会出错。
- 未初始化的SBuffer或Texture，在某些平台上会全部是0， 但是另外一些可能是任意值， 甚至是NaN。
- Metal不支持对纹理的原子操作， 不支持对SBuffer调用**GetDimensions**。
- ES 3.1在一个CS里至少支持4个SBuffer（所以， 我们需要将相关联的数据定义为struct）。

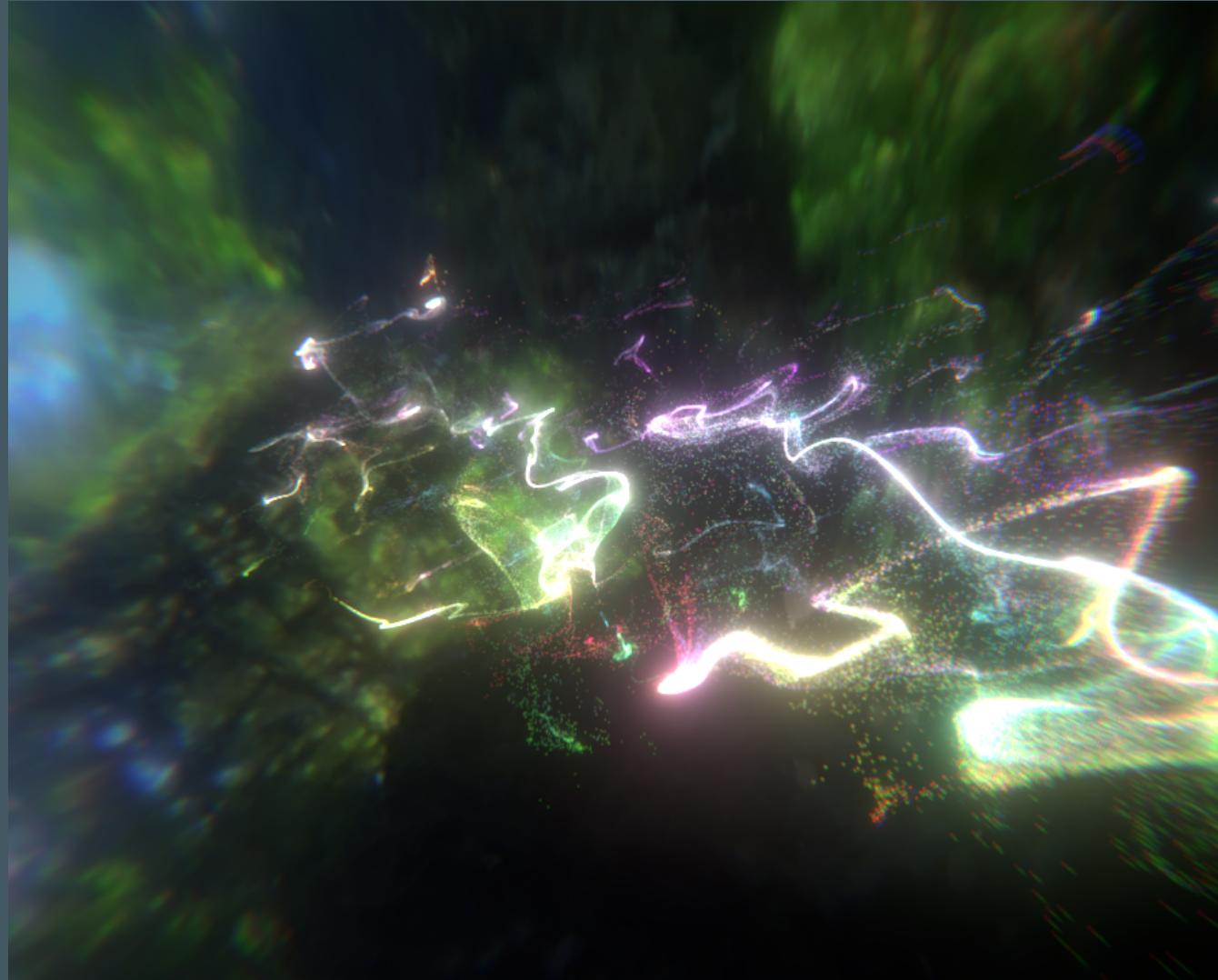
性能

- 尽量减少Group之间的交互
- GPU一次调用64（AMD）或32（NVIDIA）个线程，所以，尽量使numthreads的乘积是这个值的整数倍。
- 避免回读
- 避免分支，重点避免在thread group中间的分支
- 尽量保证内存连续性
- 使用[unroll]来打开循环，有些时候需要手动unroll

应用

目前有哪些应用

GPU Particle System



GPU Simulation

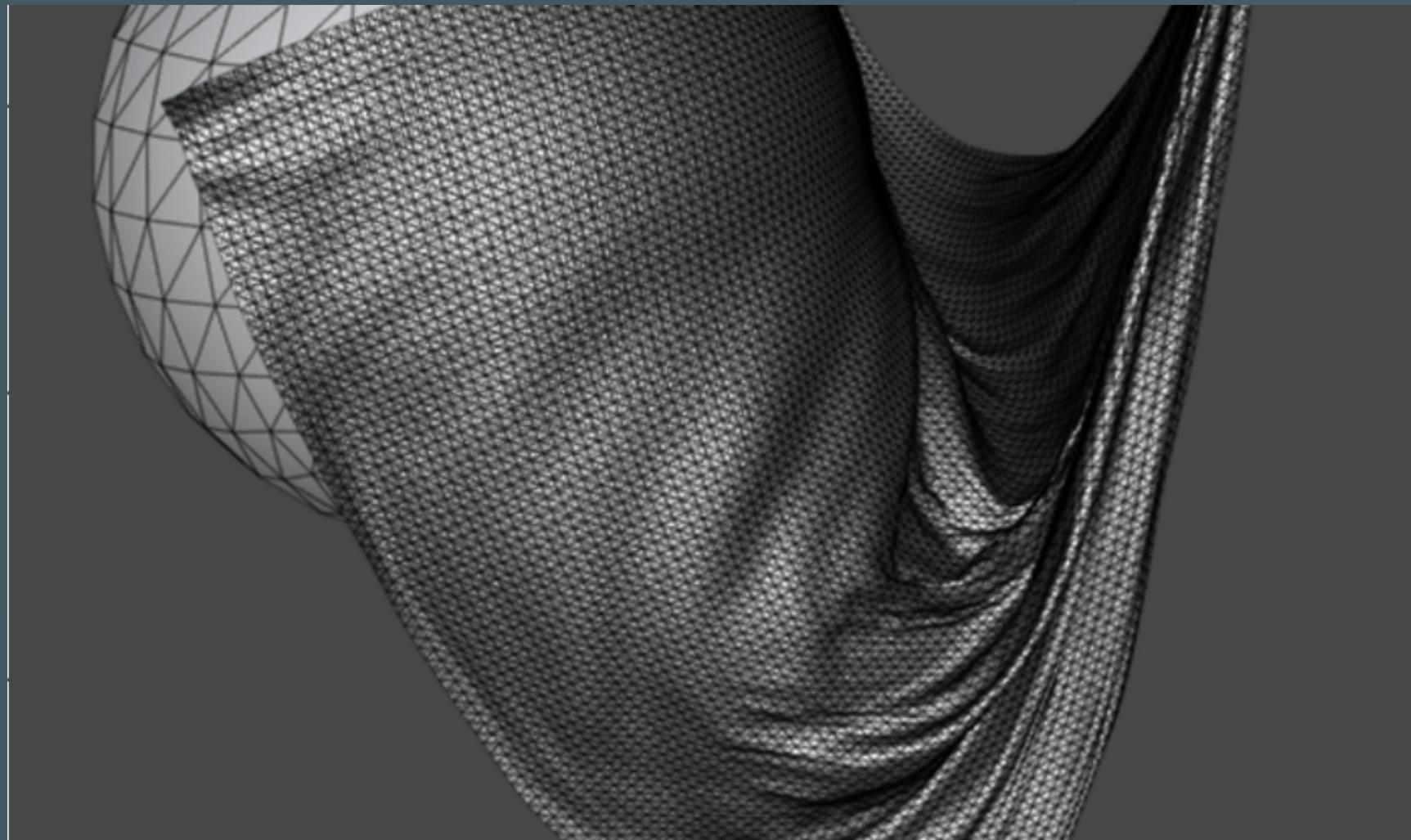


Image Processing

F1 - Desaturate

F2 - Circles

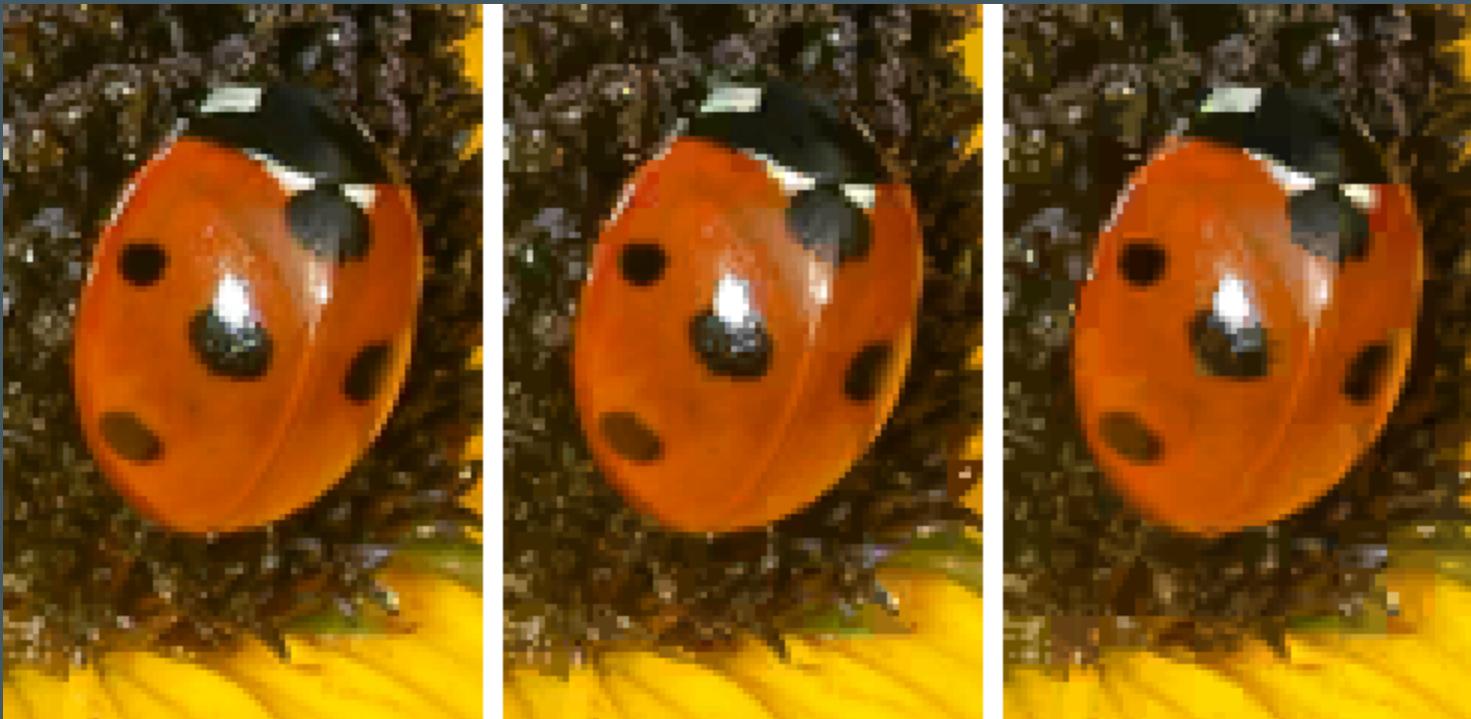


F1 - Desaturate

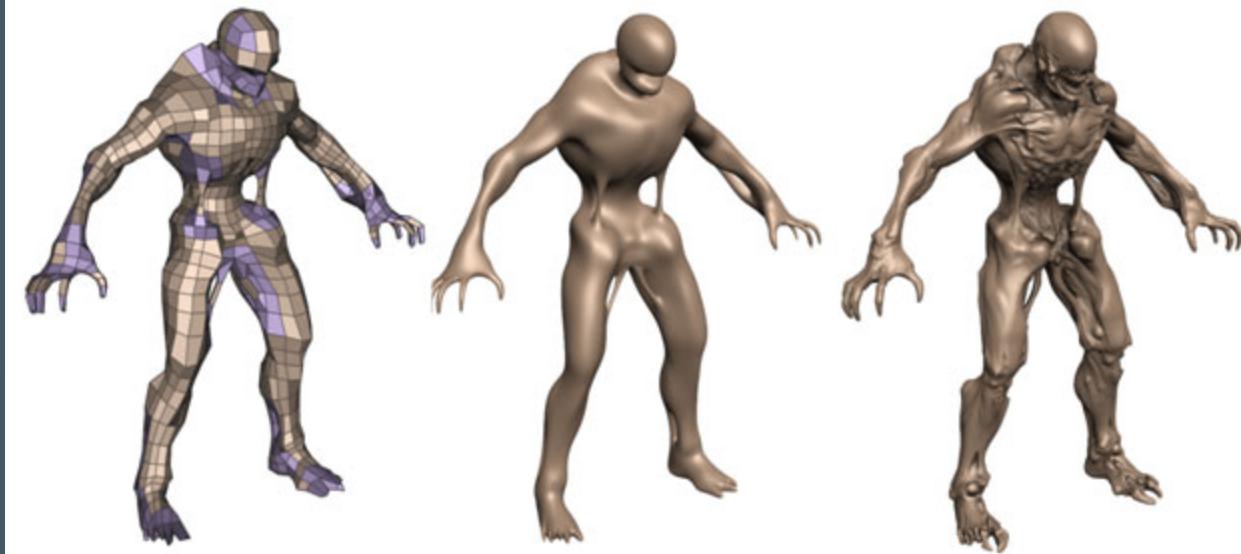
F2 - Circles



Image Compression



Tessellation



Local lights culling



Occlusion culling



GPU Driven Rendering Pipeline



还有很多很多.....

引用

1. <https://slideplayer.com/slide/9636742/>
2. <http://on-demand.gputechconf.com/gtc/2010/presentations/S12312-DirectCompute-Pre-Conference-Tutorial.pdf>
3. <https://www.youtube.com/watch?v=0DLOJPSxJEg>
4. https://arm-software.github.io/vulkan-sdk/basic_compute.html
5. https://www.khronos.org/opengl/wiki/Compute_Shader

引用

6. <https://docs.microsoft.com/en-us/windows/desktop/direct3d11/direct3d-11-advanced-stages-compute-shader>
7. <https://developer.apple.com/metal/>
8. https://static.docs.arm.com/100614/0302/arm_mali_gpu_opencl_developer_guide_100614_0302_00_en.pdf
9. Real-Time Rendering 3rd Edition. Chapter 18
10. <https://github.com/Robert-K/gpu-particles>

引用

11. <https://www.shpakivnia.com/cloth-tool>
12. [http://www.codinglabs.net/tutorial compute shaders filters.aspx](http://www.codinglabs.net/tutorial_compute_shaders_filters.aspx)
13. https://en.wikipedia.org/wiki/Adaptive_Scalable_Texture_Compression
14. Introduction to 3D Game Programming with DirectX 11
15. <https://www.nvidia.com/object/tessellation.html>

引用

16. <https://www.slideshare.net/DICEStudio/directx-11-rendering-in-battlefield-3>
17. <https://zhuanlan.zhihu.com/p/47615677>
18. http://advances.realtimerendering.com/s2015/aaltonenhaar_siggraph2015_combined_final_footer_20dpi.pdf
19. <https://docs.unity3d.com/Manual/class-ComputeShader.html>
20. <https://forum.unity.com/threads/compute-shaders.148874/>

引用

21. <https://docs.unity3d.com/ScriptReference/ComputeShader.Dispatch.html>
22. http://www.humus.name/Articles/Persson_LowlevelShaderOptimization.pptx
23. <https://www.gdcvault.com/play/1020352/Low-Level-Shader-Optimization-for>
24. 数字图像处理（冈萨雷斯）
25. https://en.wikipedia.org/wiki/General-purpose_computing_on_graphics_processing_units