

# Optimizing Unity Games for Mobile Platforms

Angelo Theodorou

Software Engineer

*Brains Eden, 28<sup>th</sup> June 2013*



# Agenda

---

- Introduction
  - The author
  - ARM Ltd.
  - What do you need to have
  - What do you need to know
- Identify the bottleneck
  - CPU
  - Vertex processing
  - Fragment processing
  - Bandwidth
- The Unity Profiler (Pro only)
- Example application overview
  - Geometry, LOD, Particles, Texture, Overdraw, Lights scenes

# The Author

---

- Angelo Theodorou
- Software Engineer
- Media Processing Division, ARM<sup>®</sup> Ltd.
  - Mali Ecosystem Use-Cases team
- B.Sc. In Computer Science
- Studied at Università degli Studi di Napoli “*Federico II*”, Italy
  
- In my previous company I have worked on From Cheese
  - Website: <http://www.fromcheese.com>
  - Grand Prize winner at the Samsung Smart App Challenge 2012
  - Made with Unity 3.5
  - Responsible for SPen integration and additional programming

# ARM Ltd.

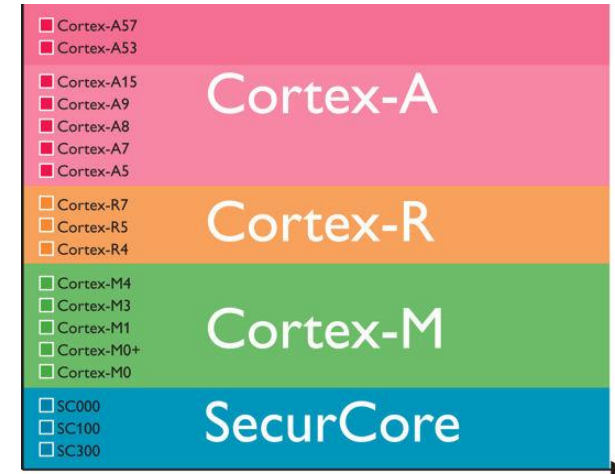
---

- Founded at the end of 1990, 13 engineers and a CEO
- Main goal was to design low power embedded 32bit processors, but to never build them
- Now more than 2500 employees
  - More than 1000 in Cambridge, the headquarters
- Hundreds of partners, a fundamental part of our business
- The business model is based on licenses and royalties
- ARM technology is everywhere: mobile, embedded, enterprise, home

# ARM

# ARM – Cortex™ CPUs

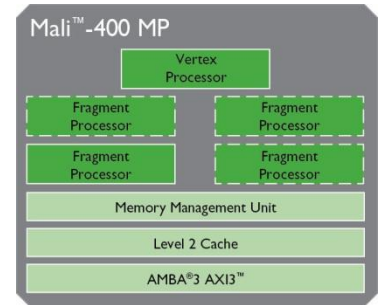
- Cortex-A
  - High performance application processors
  - Cortex-A50 series enables 64bit processing
- Cortex-R
  - Real-time embedded processors
- Cortex-M
  - Embedded processors for deterministic microcontroller applications
- SecurCore™
  - Specialist processors utilized within the security markets



# ARM – Mali™ GPUs

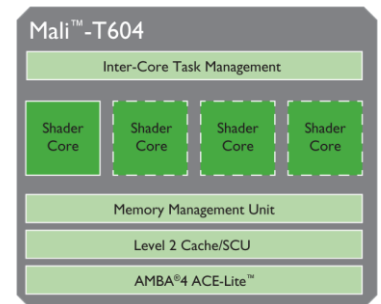
## ■ Mali-400 MP

- OpenGL® ES 2.0
- 1 vertex processor
- Up to 4 fragment processors
- Deployed in many devices, like Galaxy S2 and S3



## ■ Mali-T604

- OpenGL ES 3.0, DirectX® 11
- OpenCL™ 1.1 full profile compute capability
- Up to four unified cores
- Inside the Nexus 10, the Samsung Chromebook, the Arndale board















# What do you need to have

---

- A computer with:
  - Java SE Development Kit (JDK)
  - Android SDK
  - ADB drivers (u20gt\_adb.zip)
  - Unity Pro
- An Android device connected to the computer via ADB
- The example application APK

# Android SDK

- Website: <http://developer.android.com/sdk>
- You don't need Eclipse + ADT for Unity
  1. Click on "Use an Existing IDE"
  2. Click on "Download the SDK Tools for Windows"
  3. Run the "Android SDK Manager" as Administrator
  4. Install the Android SDK platform and build tools
  5. Install at least one SDK Platform (API16 if you want to match the Android version on the Cube U18GT-S tablet)
- Unity supports r22 only since version 4.1.4

 Name	API	Rev.	Status
 Tools			
 Android SDK Tools		22.0.1	 Installed
 Android SDK Platform-tools		17	 Not installed
 Android SDK Build-tools		17	 Not installed
 Android 4.2.2 (API17)			
 Android 4.1.2 (API16)			
 SDK Platform	16	4	 Not installed

# ADB Driver

---

- The Android Debug Bridge (ADB) consists of a server running on the device and a client accessed through a CLI
- Add the platform-tools directory of the SDK to your path
- Extract the u20gt\_adb.zip archive
  1. Enable USB debugging on the device (Settings -> Developer options -> USB debugging)
  2. Copy the “.android” directory in your home directory
  3. Go to the Windows Device Manager
  4. Locate the U18GT-S in “Other devices”
  5. Install the driver for it pointing to the “usb\_driver” subdirectory
  6. After a correct installation you would see an “Android Composite ADB Interface” in “Android Device”
  7. Restart ADB and check the connection (`adb kill-server` and `adb devices` from the command line)

# Unity Pro

---

- Unity is a cross-platform game engine
  - You can deploy on PC, consoles, mobile devices, web browsers
- Scripting based on Mono, supported languages:
  - JavaScript-like, Boo, C#
- Pro version has support for:
  - Profiler
  - Level of Detail
  - Occlusion Culling
  - Light Probes
  - Static Batching
- Latest version: 4.1.5 (June 2013)
- Website: <http://unity3d.com/>



# Android Device

Cube U18GT-S (<http://www.51cube.com/en>)

- Rockchip RK3066
  - Dual-core Cortex™-A9 at 1.6GHz
  - Quad-core Mali™-400 at 250MHz
- 1GB DDR3 RAM
- 8GB Internal Storage
- 1024x600 7" 5-points capacitive screen
- HDMI 1.4 interface
- Android™ Jelly Bean 4.1.1



# What do you need to know

---

- What is a *draw call*
  - A call to a function of the underlying API (e.g. OpenGL ES) to draw something on the screen
- What is a *fragment*
  - A candidate pixel, it may or may not end up on screen for different reasons (e.g. being discarded, being overwritten, etc.)
- Differences between opaque and transparent render queue
  - Objects in the first queue are rendered in a front to back order to minimize the overdraw thanks to the depth test, transparent ones are rendered afterwards and in a back to front order
- What does *batching* mean
  - Group similar draw calls in a single one operating on the whole data
- Why mobile platforms are different
  - *Immediate vs deferred* rendering

# Why mobile platforms are different

---

## ■ Desktop platforms

- *Immediate mode*: graphics commands are executed when issued
- Huge amount of bandwidth available between GPU and dedicated video memory (>100 GB/s)
- No strict limits in size, power consumption or heat generation

## ■ Mobile platforms

- *Deferred mode*: graphics commands are collected by the driver and issued later
- *Tile based*: rendering occurs in a small on-chip buffer before being written to memory
- Bandwidth is severely reduced (~5 GB/s) and transferring data needs a great amount of power
- Memory is unified and shared between CPU and GPU

# Identify the bottleneck

## ■ CPU

- Too many draw calls
- Complex scripts or physics

## ■ Vertex processing

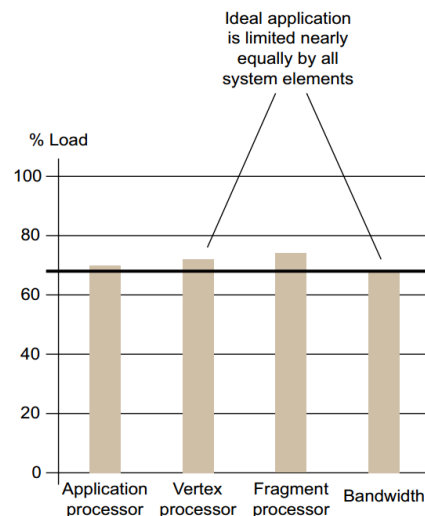
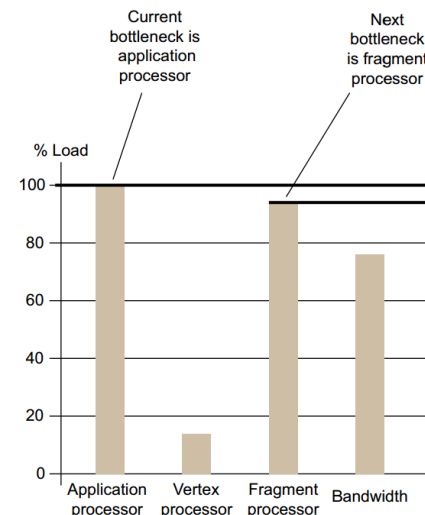
- Too many vertices
- Too much computation per vertex

## ■ Fragment processing

- Too many fragments, overdraw
- Too much computation per fragment

## ■ Bandwidth

- Big and uncompressed textures
- High resolution framebuffer



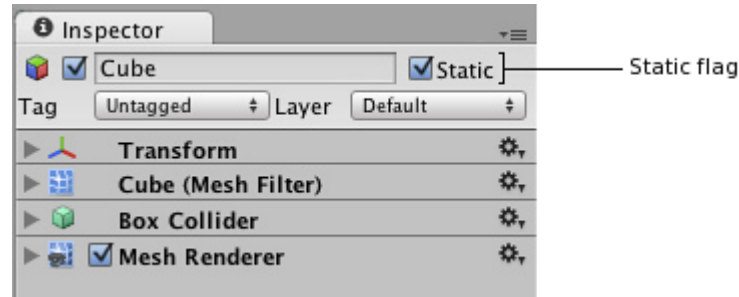
# CPU Bound

---

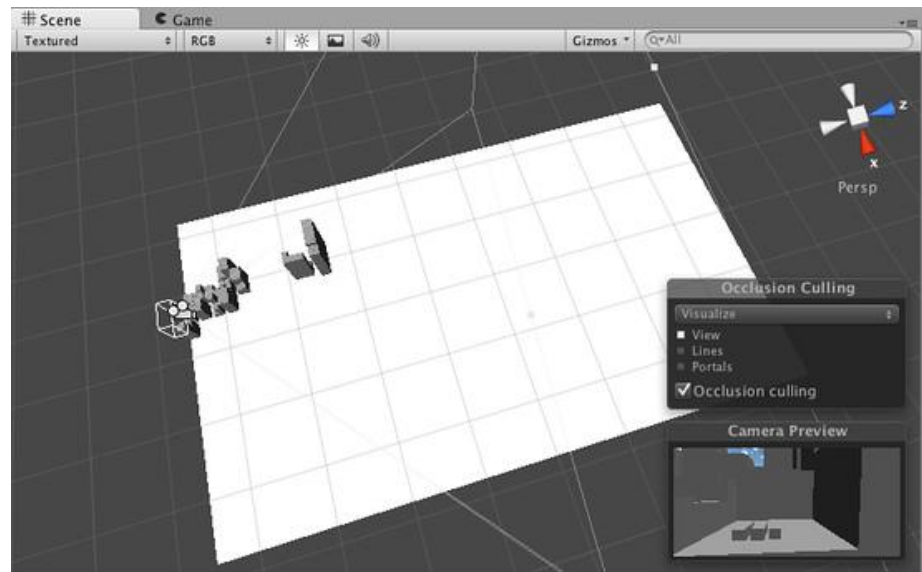
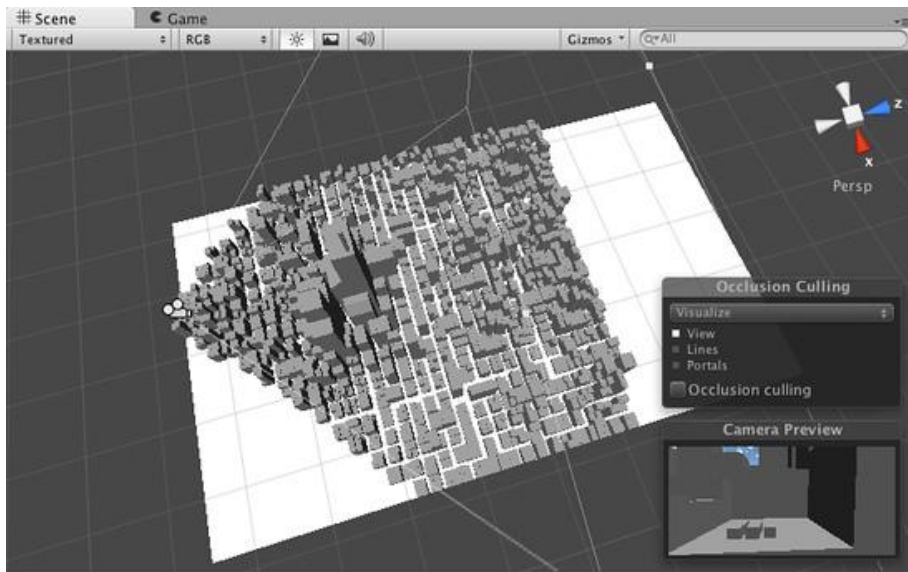
- Too many draw calls
  - Static batching (Unity Pro only)
  - Dynamic batching (automatic)
  - Frustum culling (automatic)
  - Occlusion culling (Unity Pro only)
- Complex scripts
  - Component caching
  - Pool of objects
  - Reduce Unity GUI calls
- Complex physics
  - Compound colliders instead of mesh ones

# CPU Bound – Batching and Culling

- Make an object static to enable static batching (Unity Pro)

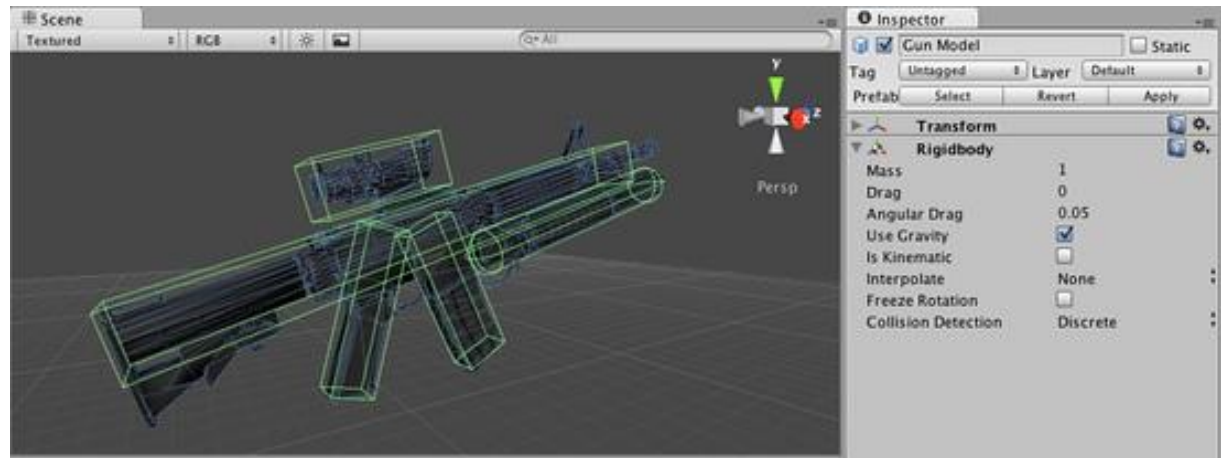
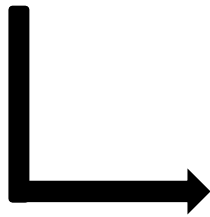
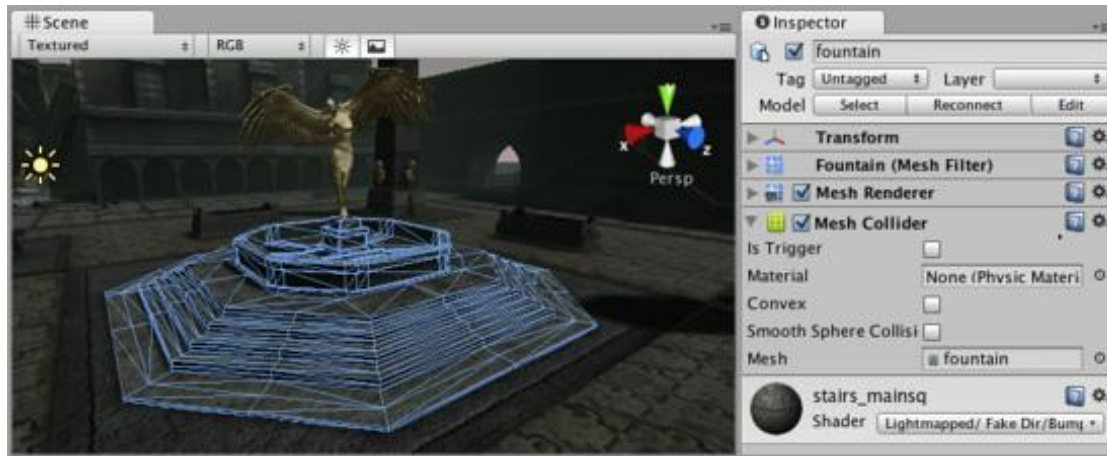


- Occlusion culling (Unity Pro only)



# CPU Bound – Compound colliders

- Use compound colliders instead of mesh colliders if possible



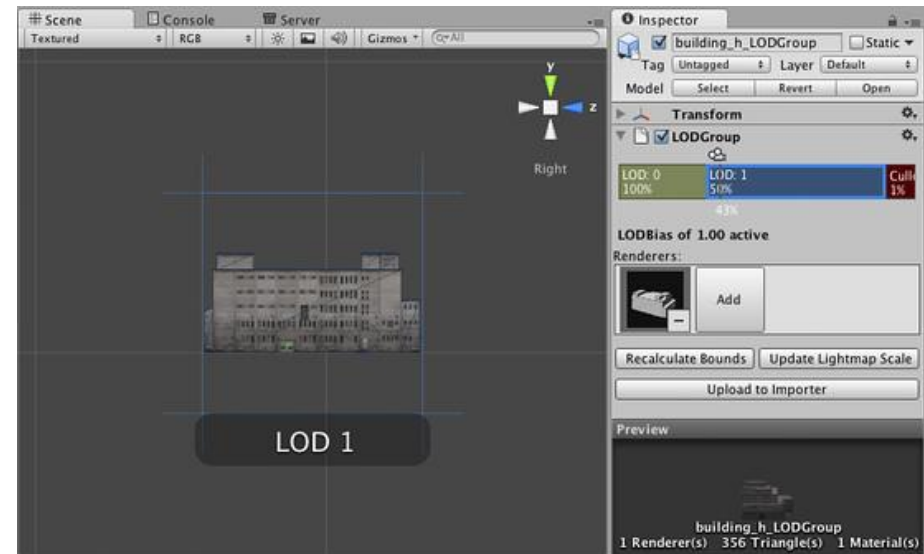
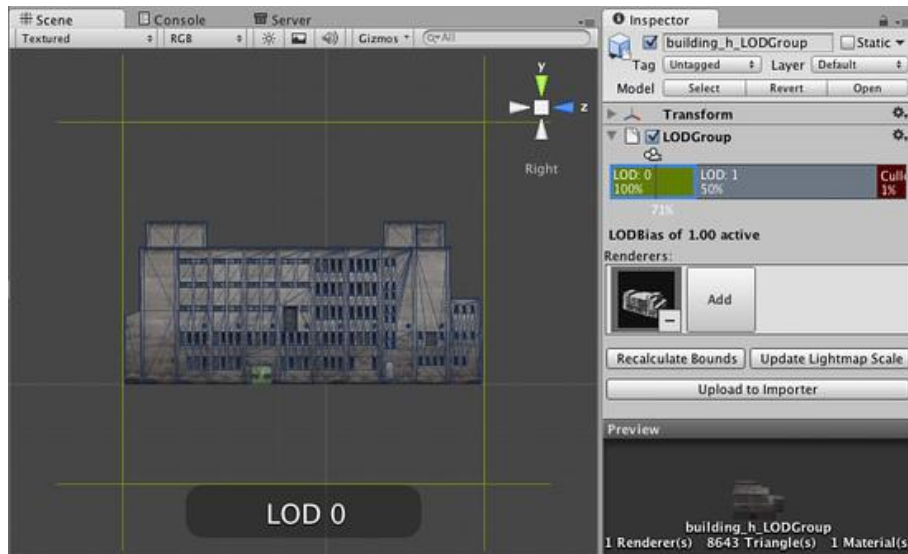
# Vertex Bound

---

- Too many vertices in geometry
  - Remove unnecessary vertices
  - Use LOD switching through LODGroup (Unity Pro only)
  - Frustum culling (automatic)
  - Occlusion culling (Unity Pro only)
- Too much computation per vertex
  - Use the mobile version of Unity shaders whenever you can

# Vertex Bound – Level of Detail

- Use LODGroup to limit vertex count when geometric detail is not strictly needed (very far or small objects)



# Fragment Bound

---

## ■ Overdraw

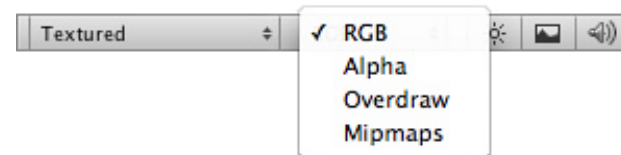
- When you are drawing to each pixel on the screen more than once
- Drawing your objects front to back instead of back to front reduces overdraw, thanks to depth testing
- Limit the amount of transparency in the scene (beware of particles!)
- Unity has a *render mode* to show the amount of overdraw per pixel

## ■ Too much computation per fragment

- Bake as much as you can (lightmaps, light probes, etc.)
- Contain the number of per-pixel lights
- Limit real-time shadows (only high end mobile devices, Unity 4 Pro)
- Try to avoid full screen post-processing
- Use the mobile version of Unity shaders whenever you can

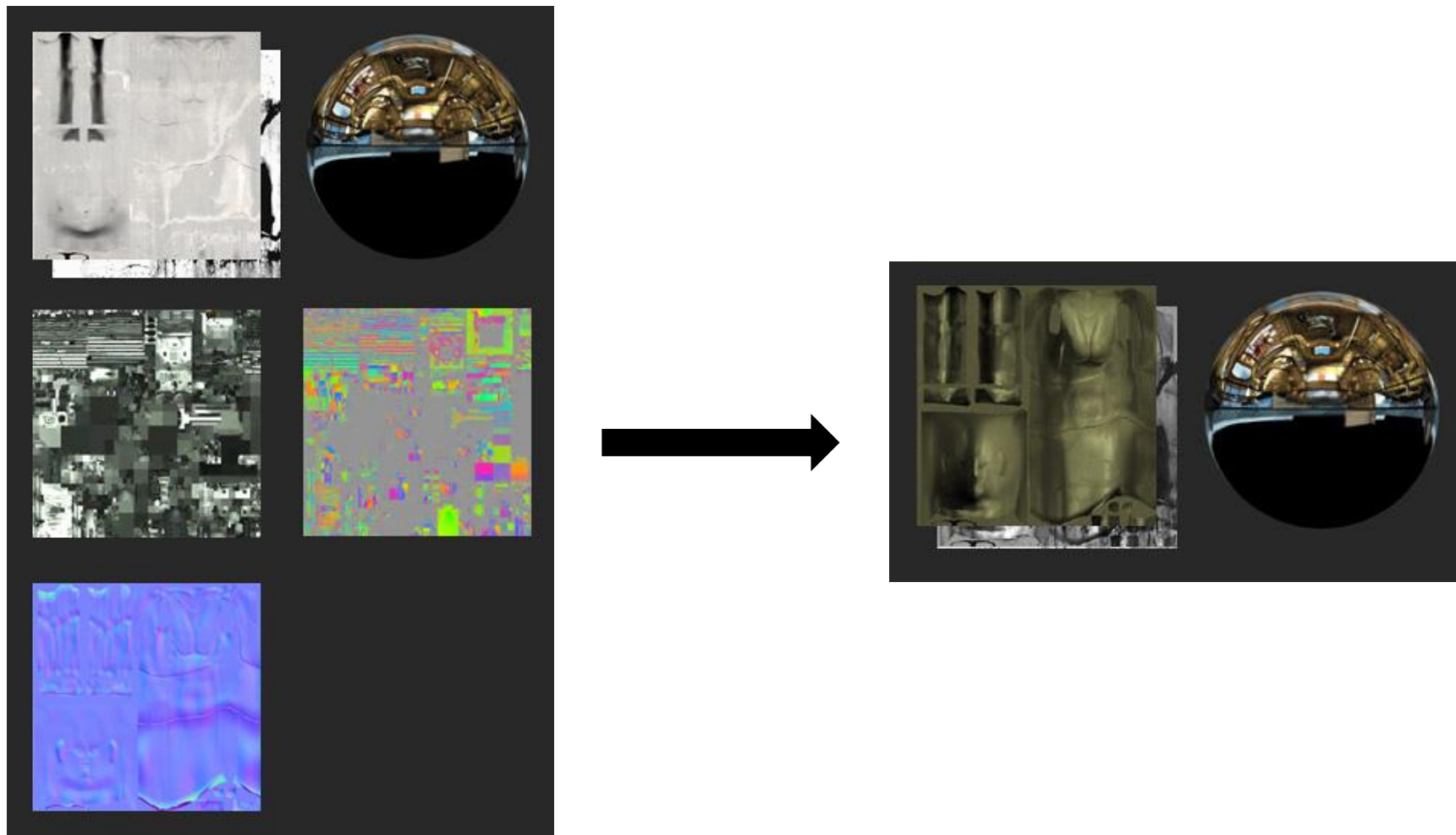
# Fragment Bound – Overdraw

- Use the specific render mode to check the overdraw amount



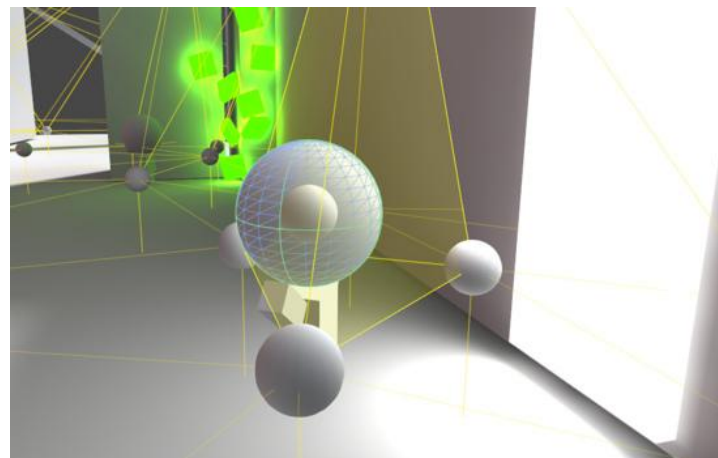
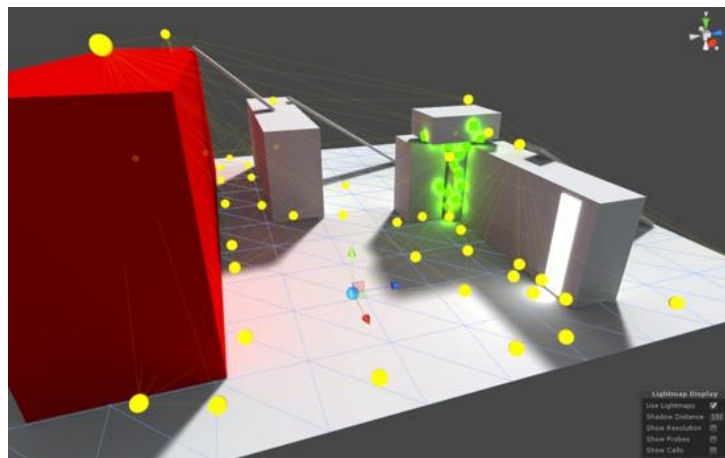
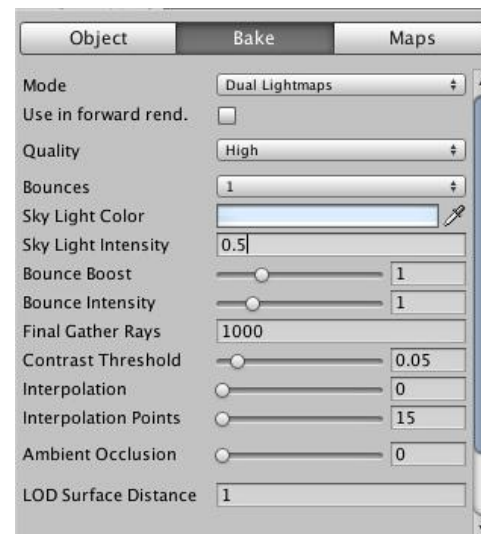
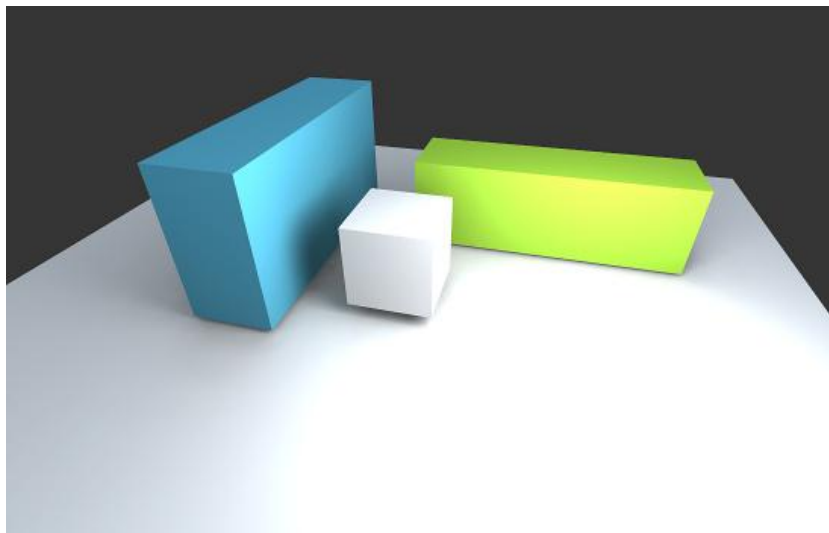
# Fragment Bound – Bake to textures

- Try to bake dynamic effects in textures



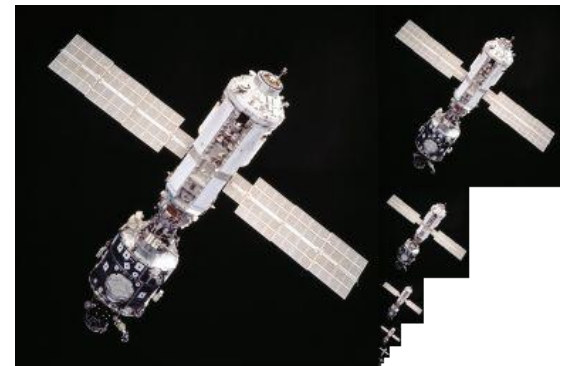
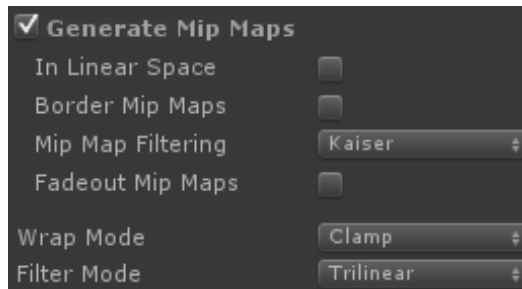
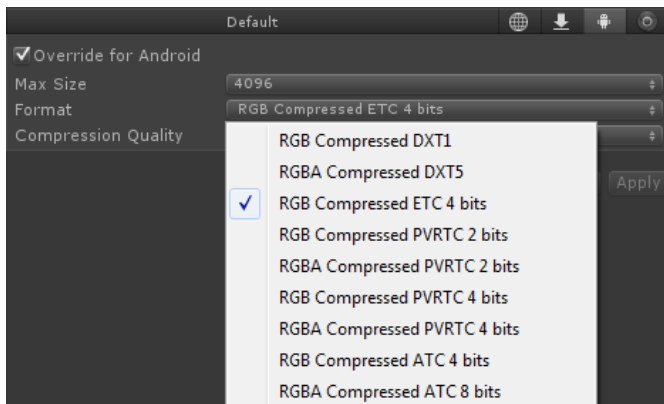
# Fragment Bound – Lightmaps and probes

- Bake static illumination to lightmaps and exploit probes



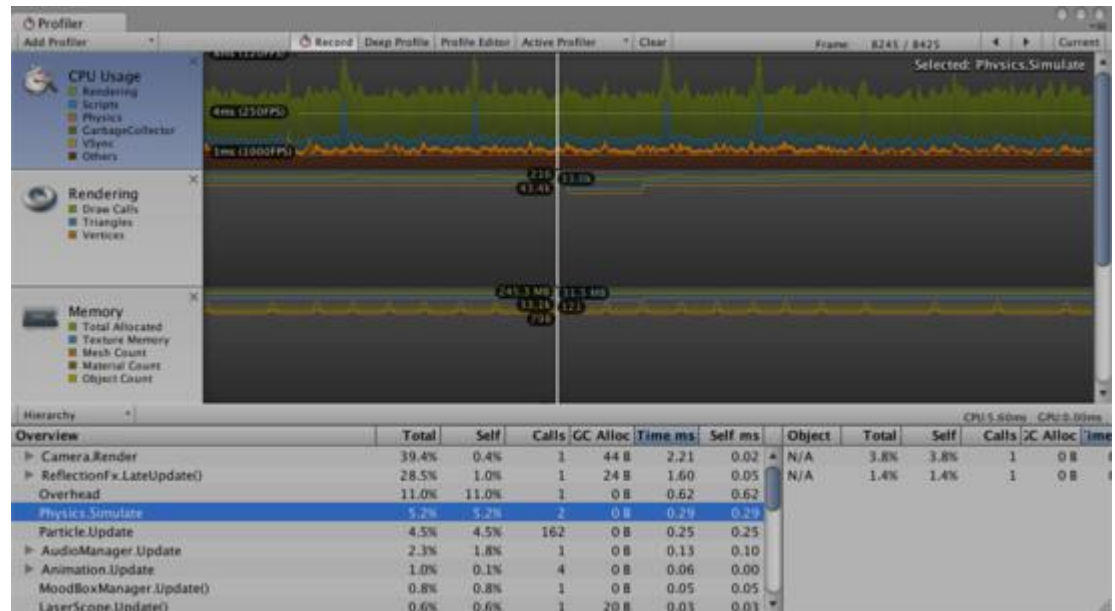
# Bandwidth Bound

- Use texture compression
  - ETC is the standard compressed format that comes with OpenGL ES
  - ASTC is the new Khronos standard soon to be supported by Unity
- Use MIP maps
  - More memory but improved image quality (less aliasing artefacts)
  - Optimized memory bandwidth (when sampling from smaller maps)
- Use trilinear and anisotropic filtering in moderation
- Use Level of Detail techniques for geometry



# Unity Profiler

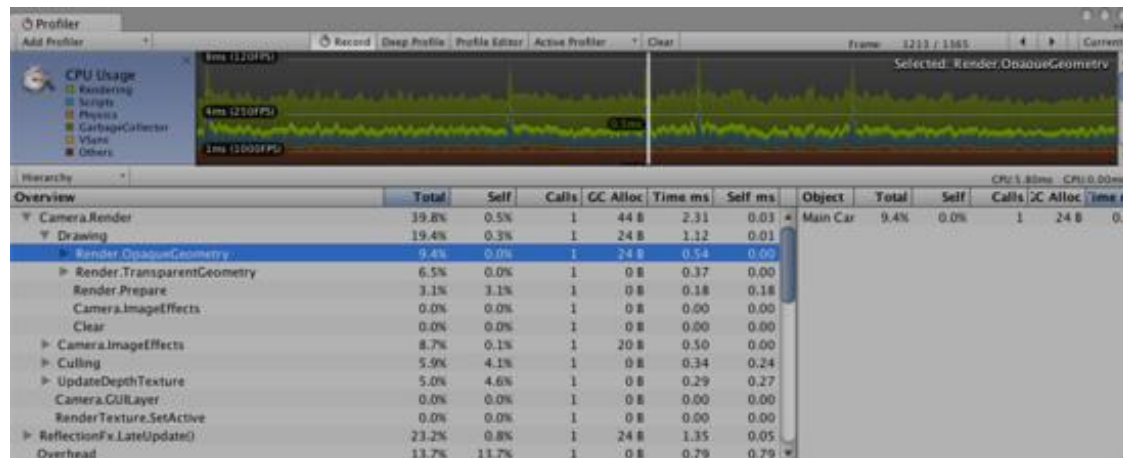
- It instruments the code to provide detailed per-frame performance data
- It provides specific data about:
  - CPU Usage
  - Rendering
  - GPU Usage
  - Memory
  - Physics
  - Audio



- Can profile content running on mobile devices
- Only available in Unity Pro

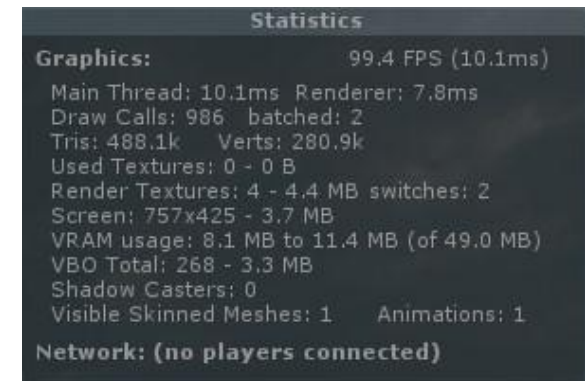
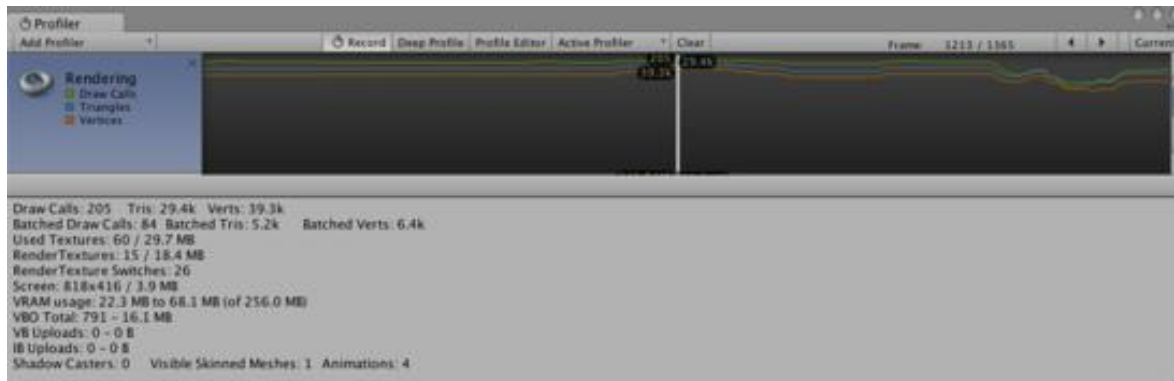
# Unity Profiler – CPU Usage

- It shows CPU utilization for rendering, scripts, physics, garbage collection, etc.
- It shows detailed info about how the time was spent
- You can enable *Deep Profile* for additional data about all the function calls occurring in your scripts
- You can manually instrument specific blocks of code with `Profiler.BeginSample()` and `Profiler.EndSample()`



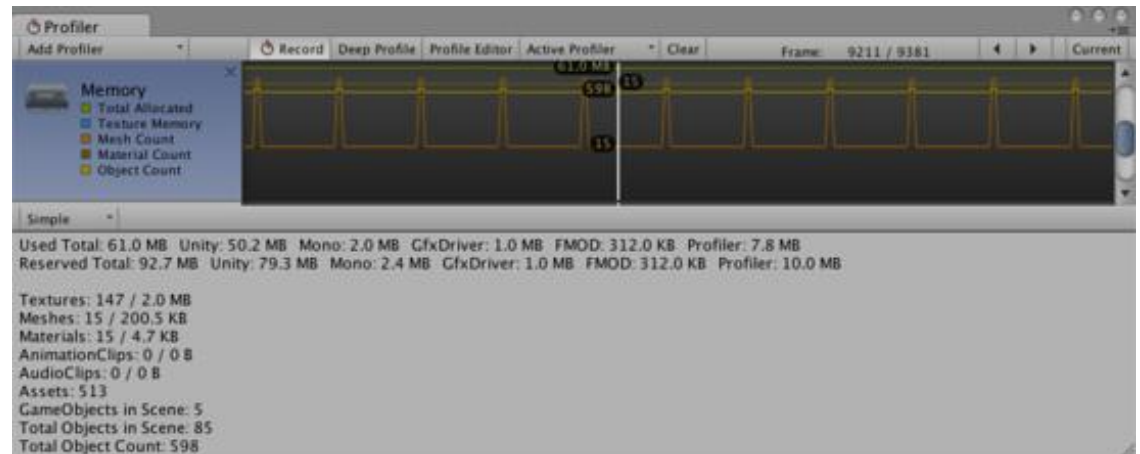
# Unity Profiler – Rendering

- Statistics about the rendering subsystem
  - Draw calls
  - Triangles
  - Vertices
- The lower pane shows data similar to the rendering statistics window of the editor



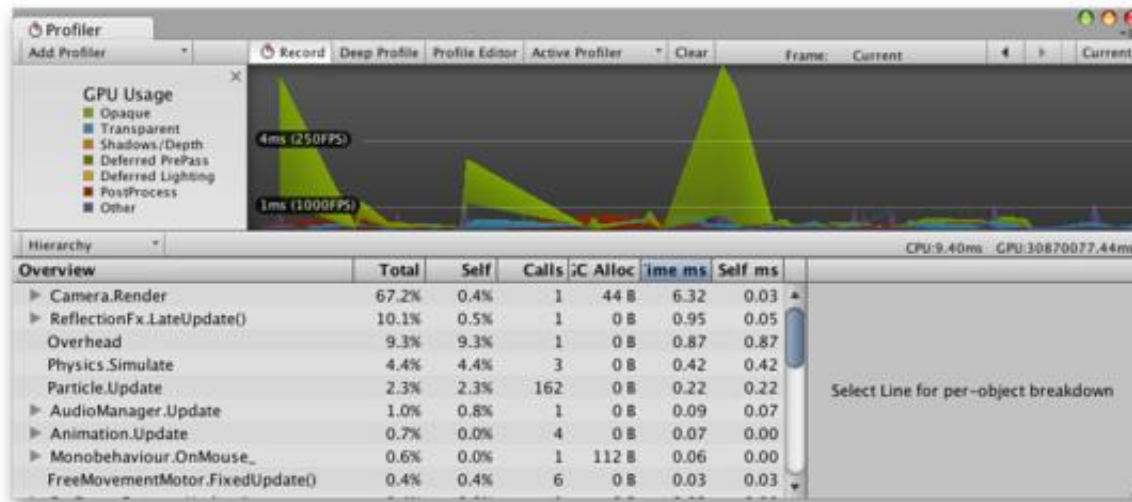
# Unity Profiler – Memory

- It shows memory used/reserved on a higher level
  - Unity native allocations
  - Garbage collected Mono managed allocations
  - Graphics and audio total used memory estimation
  - Memory used by the profiler data itself
- It also shows memory used by assets/objects
  - Textures
  - Meshes
  - Materials
  - Animation clips
  - Audio clips



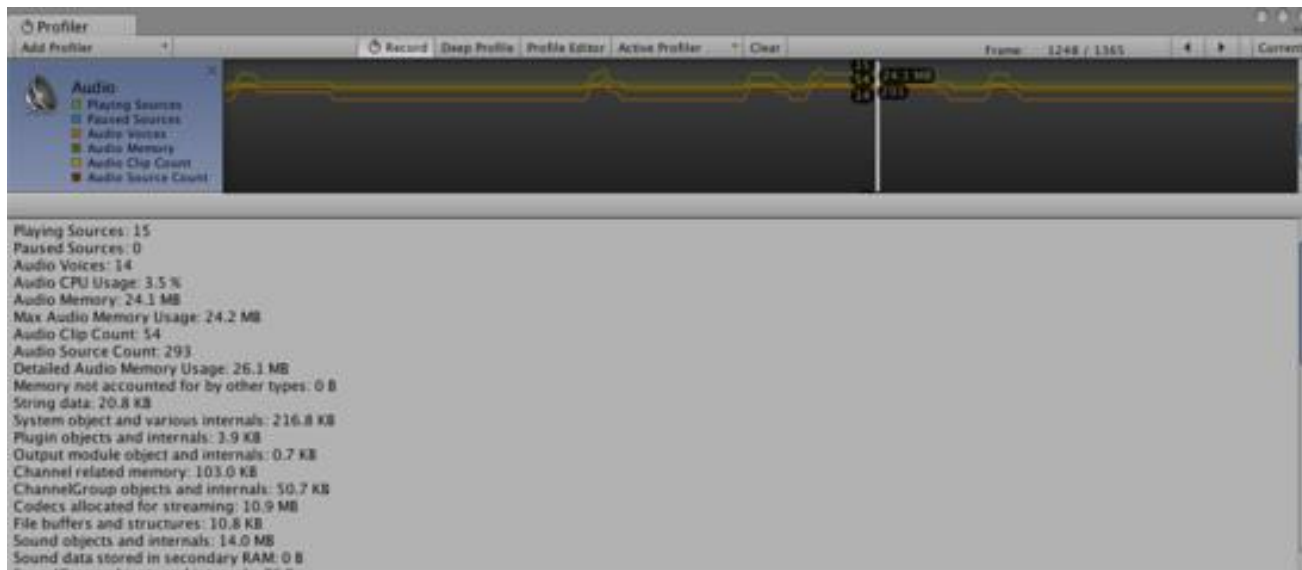
# Unity Profiler – GPU Usage

- It shows a contribution breakdown similar to the CPU profiler
  - Rendering of opaque objects
  - Rendering of transparent objects
  - Shadows
  - Deferred shading passes
  - Post processing
- Not yet available on mobile platforms ☹️



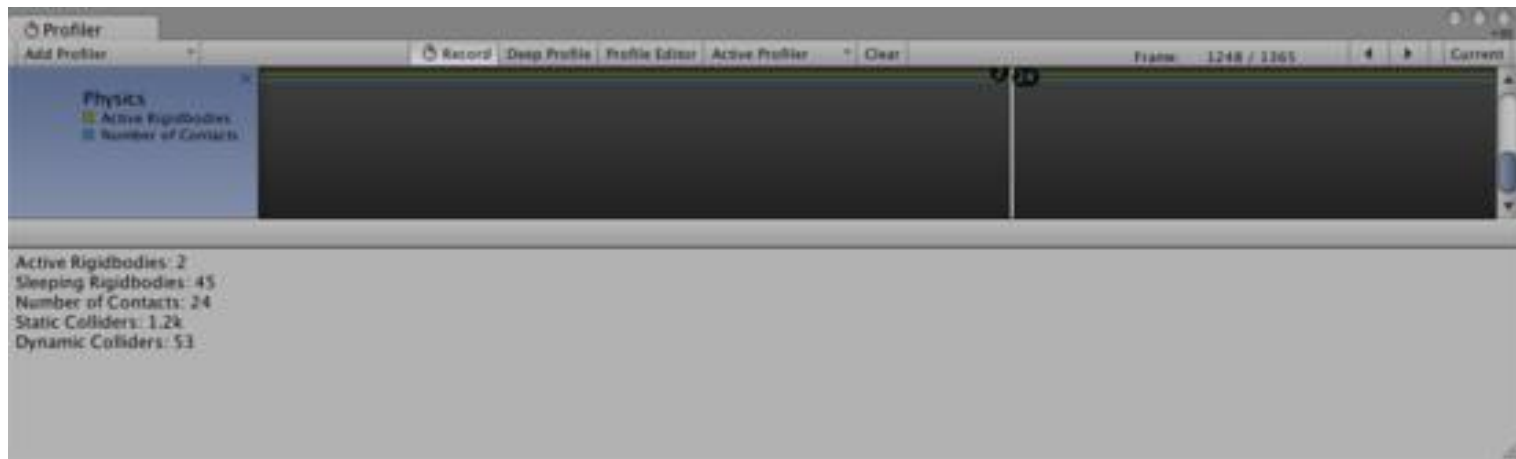
# Unity Profiler – Audio

- Statistics about the audio subsystem
  - Total number of audio clips and sources
  - Number of playing/paused audio sources
  - Audio channels in use
  - Total amount of memory used by the audio engine



# Unity Profiler – Physics

- Statistics about the physics simulation
  - Active rigid bodies (moving)
  - Sleeping rigid bodies (not processed by the physics engine)
  - Points of contact between all colliders
  - Number of static colliders
  - Number of dynamic colliders (attached to rigid bodies)



# Unity Profiler on Android

## ■ WiFi profiling

1. Make sure the Android device and the host are on the same subnet
2. Check the *Autoconnect Profiler* checkbox in build settings dialog
3. Attach your device to host via cable and hit *Build & Run*
4. When the app starts open the profiler window and select the appropriate device from the *Active Profiler* drop down menu

## ■ ADB profiling

1. Attach your device via cable and make sure ADB recognizes it
2. `adb forward tcp:54999 localabstract:Unity-<insert bundle identifier here>`
3. Check the *Development Build* checkbox and hit *Build & Run*
4. When the app starts open the profiler window and select `AndroidProfiler (ADB@127.0.0.1:54999)`
5. Ports from 54998 to 55511 should be open in your firewall rules

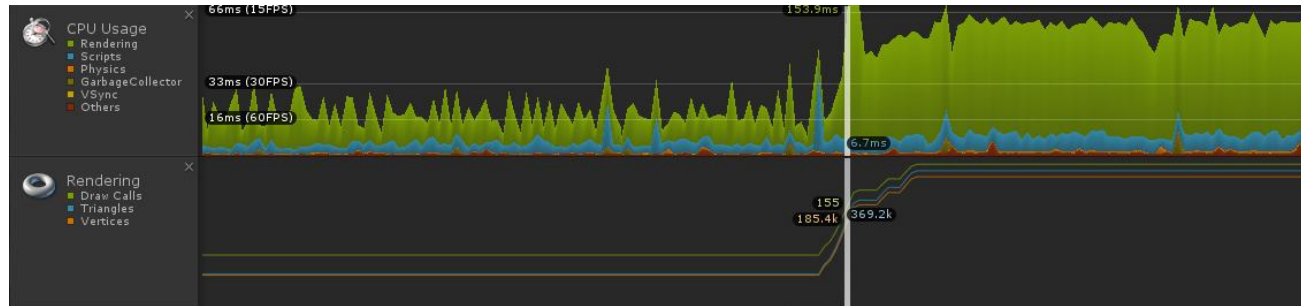
# Example application overview

---

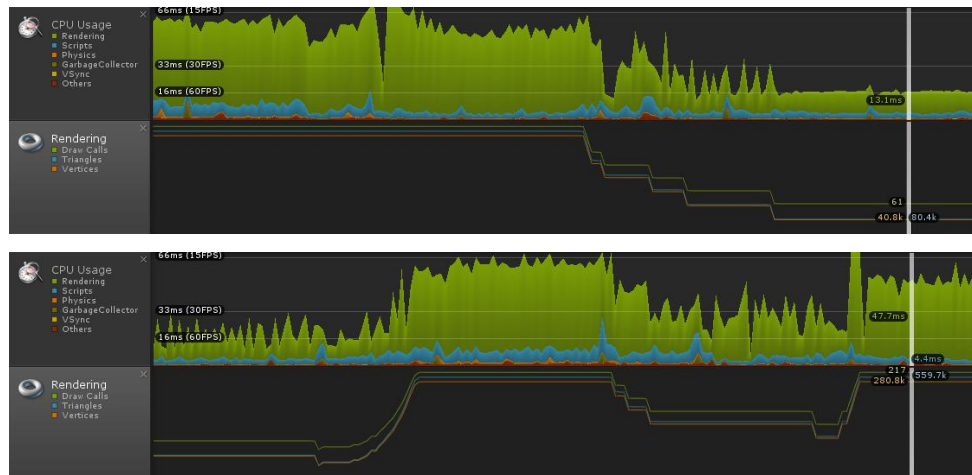
- Made up of six different scenes
  - Geometry (high vertex count, frustum culling)
  - LOD (using level of detail through LODGroup)
  - Particles (transparent rendering)
  - Texture (MIP mapping and compression)
  - Overdraw (alpha blending and alpha test)
  - Lights (vertex and pixel lighting)
- Most of them are based on a Prefab and `Instantiate()`
  - All instances are created at once in the beginning and deactivated
  - Number of active instances can adapt to maintain a target FPS value
  - Instantiation is made in screen space and based on camera aspect

# Geometry Scene

- Stressing the vertex units with lots of high poly objects
  - You can set the amount of visible objects or a target FPS value

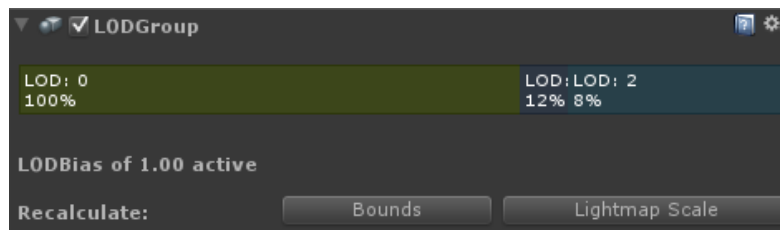


- It is possible to rotate and translate the camera to show frustum culling in action (and lack of occlusion culling)



# LOD Scene

- LODGroup with three different levels

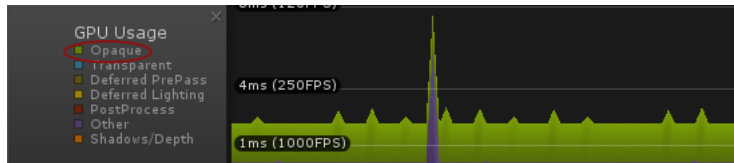


- Triangles and vertices spikes match LOD level switch, draw calls are dependent on frustum culling

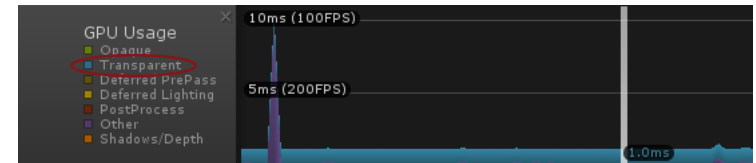


# Particles Scene

- All the rendering is done in the “Transparent Queue”

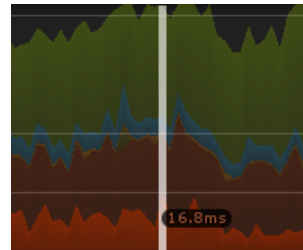


Geometry Scene



Particles Scene

Overview	Total
▼ Camera.Render	67.0%
▼ Drawing	66.2%
▶ Render.TransparentGeometry	65.6%
Render.Prepare	0.2%
Clear	0.1%
▶ Render.OpaqueGeometry	0.0%
Camera.ImageEffects	0.0%



Overview	Total
▶ Camera.Render	49.8%
▶ GUI.Repaint	17.3%
ParticleSystem.WaitForUpdateThreads	14.9%
ParticleSystem.Update	9.3%
Overhead	3.3%
▶ Graphics.PresentAndSync	2.2%
Physics.Simulate	1.0%

- Perfect case of geometry batching (one draw call per emitter)

Draw Calls: 28 Tris: 8.2k Verts: 16.3k  
Batched Draw Calls: 39 Batched Tris: 7.8k Batched Verts: 15.6k

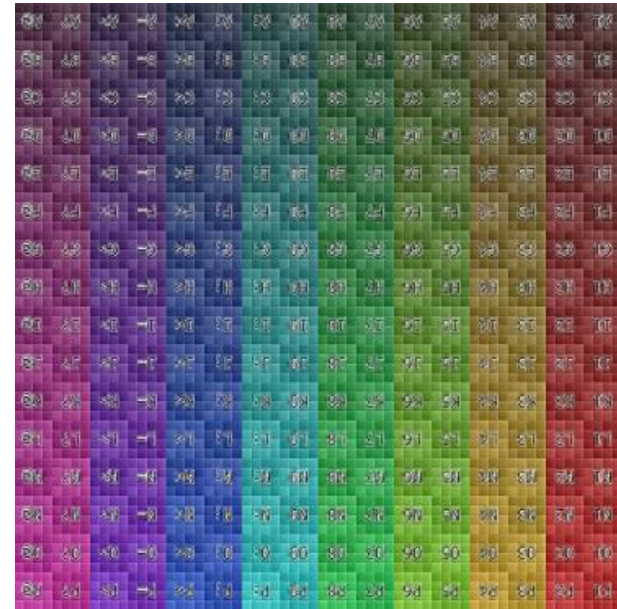
# Texture Scene

- Compressed texture with MIP maps (right) takes less memory and looks better than uncompressed one (left)



Uncompressed, bilinear filtering

```
Tris: 606 Verts: 733
Used Textures: 9 - 12.1 MB
Render Textures: 0 - 0 B switches: 0
Screen: 1442x595 - 9.8 MB
VRAM usage: 9.8 MB to 22.0 MB (of 1.95 GB)
VBO Total: 9 - 92.0 KB
```

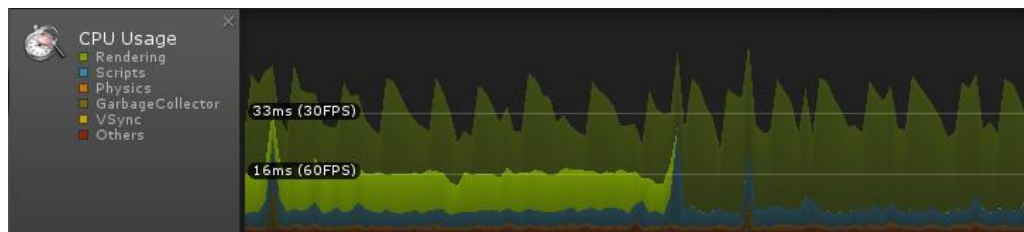


ETC compressed, trilinear filtering, MIP mapped

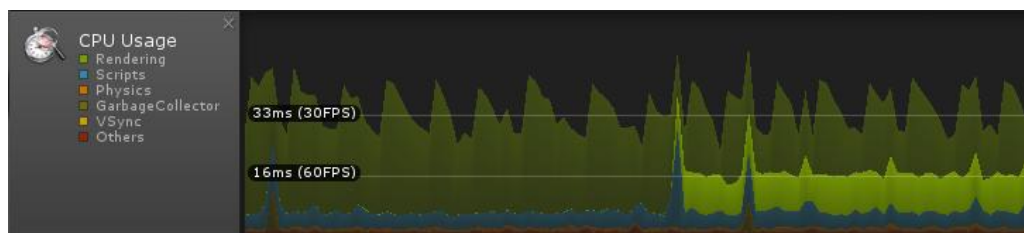
```
Tris: 606 Verts: 733
Used Textures: 10 - 2.7 MB
Render Textures: 0 - 0 B switches: 0
Screen: 1442x595 - 9.8 MB
VRAM usage: 9.8 MB to 12.6 MB (of 1.95 GB)
VBO Total: 9 - 92.0 KB
```

# Overdraw Scene

- Alpha blended quads are rendered in the transparent queue, alpha tested ones in the opaque queue, same performances

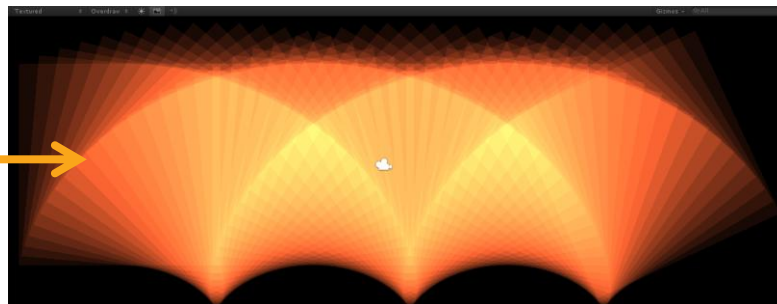


Overview	Total	Time ms
▶ Graphics.PresentAndSync	42.2%	16.35
▼ Camera.Render	31.1%	12.04
▼ Drawing	29.6%	11.49
▶ Render.TransparentGeometry	28.2%	10.95
Render.Prepare	0.6%	0.27
Clear	0.3%	0.12
▶ Render.OpaqueGeometry	0.1%	0.05
Camera.ImageEffects	0.0%	0.00



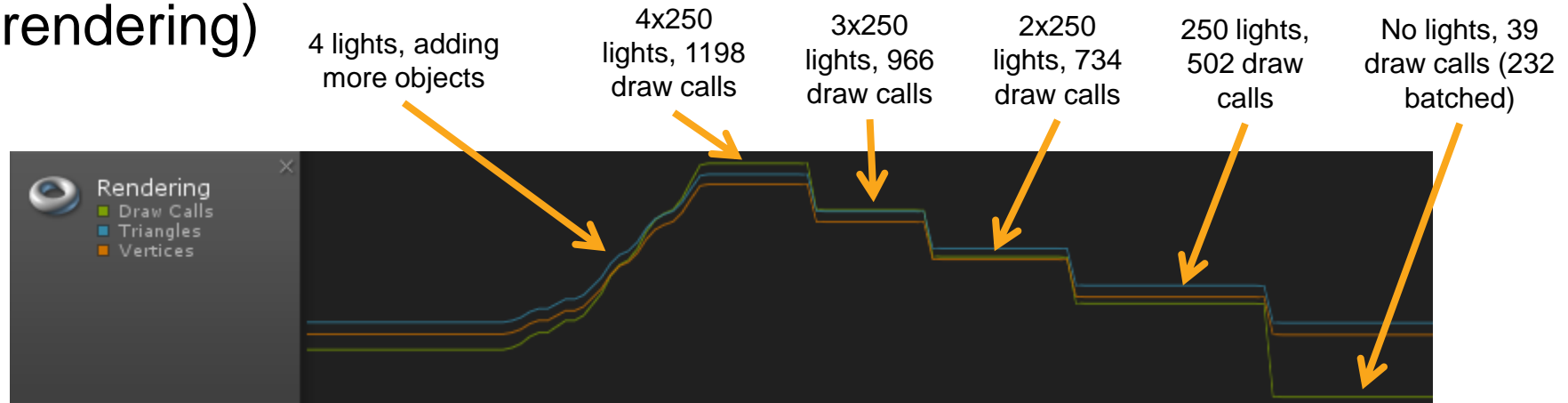
Overview	Total	Time ms
▶ Graphics.PresentAndSync	42.4%	16.09
▼ Camera.Render	31.3%	11.88
▼ Drawing	29.9%	11.32
▶ Render.OpaqueGeometry	28.5%	10.80
Render.Prepare	0.6%	0.26
Clear	0.3%	0.11
▶ Render.TransparentGeometry	0.1%	0.04
Camera.ImageEffects	0.0%	0.00

Overdraw  
render mode

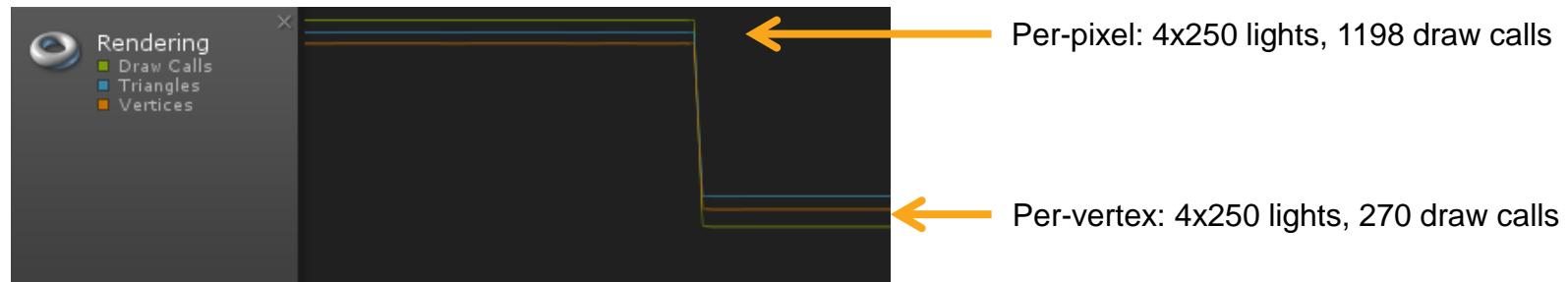


# Lights Scene (1/2)

- Per-pixel lighting uses a draw call per light (forward rendering)



- Per-vertex lighting uses one geometric pass per object, lighting information is a vertex attribute



# Lights Scene (2/2)

- `MeshRenderer.Render` takes 5 times the amount of *ms* when performing per-pixel lighting (one pass per light)

Overview	Total	Time ms
▼ Camera.Render	94.2%	73.11
▼ Drawing	93.6%	72.64
▼ Render.OpacityGeometry	84.6%	65.69
▼ RenderForwardOpaque.Render	48.2%	37.43
▶ MeshRenderer.Render	32.5%	25.29
Shader.SetPass	3.0%	2.40
RenderForwardOpaque.Prepare	36.0%	27.97
RenderForwardOpaque.Sort	0.3%	0.25

← Per-pixel lighting

Overview	Total	Time ms
▼ Camera.Render	67.2%	11.68
▼ Drawing	59.0%	10.24
▼ Render.OpacityGeometry	54.7%	9.50
▼ RenderForwardOpaque.Render	51.9%	9.02
▶ MeshRenderer.Render	28.7%	4.99
Shader.SetPass	2.9%	0.50
RenderForwardOpaque.Prepare	1.9%	0.34
RenderForwardOpaque.Sort	0.6%	0.10

← Per-vertex lighting

# Links

---

- [Practical Guide to Optimization for Mobiles](#) (Unity Manual)
- [Optimizing Graphics Performance](#) (Unity Manual)
- [Profiler](#) (Unity Manual)
  
- [ShadowGun: Optimizing for Mobile Sample Level](#) (Unity Blog)
- [“Fast Mobile Shaders” talk at SIGGRAPH 2011](#) (Unity Blog)
  
- [ASTC Texture Compression: ARM Pushes the Envelope in Graphics Technology](#) (ARM Blogs)

# The End

---

Any questions?