Lecture 9

Memory Management: High-Level Overview
Gaming Memory (Last Generation)

- Playstation 3
  - 256 MB RAM for system
  - 256 MB for graphics card

- X-Box 360
  - 512 MB RAM (unified)

- Nintendo Wii
  - 88 MB RAM (unified)
  - 24 MB for graphics card

- iPhone/iPad
  - 1 GB RAM (unified)
Gaming Memory (Current Generation)

- Playstation 4
  - 8 GB RAM (unified)

- X-Box One
  - 8 GB RAM (unified)
  - 5 GB for games

- Nintendo Wii-U
  - 2 GB RAM (unified)
  - 1 GB only for OS

- iPhone/iPad
  - 1 GB RAM (unified)
Why Not Virtual Memory?

- **Secondary storage** exists
  - Consoles have 500 GB HD
  - iDevices have 64 GB Flash
- But **access time** is slow
  - HDs transfer at \(~160\) MB/s
  - Best SSD is \(~500\) MB/s
- Recall **16 ms** per frame
  - At best, can access 8 MB
  - Yields uneven performance
Memory Usage: Images

- Pixel color is 4 bytes
  - 1 byte each for r, b, g, alpha
  - More if using HDR color

- Image a **2D array** of pixels
  - 1280x1024 monitor size
  - 5,242,880 bytes ~ 5 MB

- More if using **mipmaps**
  - Graphic card texture feature
  - Smaller versions of image
  - Cached for performance
  - But can double memory use
Memory Usage: Images

- Pixel color is 4 bytes
  - 1 byte each for r, b, g, alpha
  - More if using HDR color

- Image a **2D array** of pixels
  - 1280x1024 monitor size
  - 5,242,880 bytes ~ 5 MB

- More if using **mipmaps**
  - Graphic card texture feature
  - Smaller versions of image
  - Cached for performance
  - But can double memory use
But My JPEG is only 8 KB!

- Formats often **compressed**
  - JPEG, PNG, GIF
  - But not always TIFF
- Must **uncompress** to display
  - Need space to uncompress
  - In RAM or graphics card
- Only load when needed
  - Loading is primary I/O operation in AAA games
  - Causes “texture popping”
But My JPEG is only 8 KB!

- Formats often **compressed**
  - JPEG, PNG, GIF
  - But not always TIFF
- Must **uncompress** to display
  - Need space
  - In RAM or graphics card
- Only load when needed
  - Loading is primary I/O operation in AAA games
  - Causes “texture popping”

**Sounds** have a similar problem
Rebel Mages

WHEREAS THE CIRCLE WAS ESTABLISHED NOT MERELY TO PROTECT THE WORLD FROM MAGES, BUT ALSO TO ALLOW MAGES TO PRACTICE THEIR ART SAFELY AND WITHOUT FEAR, AND,

WHEREAS UNDER LORD SEEKER LAMBERT'S COMMAND, THE TEMPLARS SWORN TO PROTECT ALL PEOPLE—INCLUDING MAGES—FROM THE HARMFUL EFFECTS OF MAGIC, HAVE INSTEAD PERSECUTED MAGES WITH SUCH BIASED JUDGMENT AS TO WORSEN THE PROBLEMS THEY WERE MEANT TO MITIGATE, AND,

WHEREAS THE RITE OF TRANQUILITY, INTENDED AS A TOOL OF LAST RESORT TO STOP UNCONTROLLED MAGES FROM HURTING THEMSELVES OR OTHERS, HAS INSTEAD BEEN USED FOR PUNITIVE AND POLITICAL PURPOSES TO SILENCE DISSENT AND INHIBIT CIVILIZED DISCOURSE, AND,

WHEREAS ANDRASTE HERSELF INTENDED THE RELATIONSHIP BETWEEN MAGE AND TEMPLAR TO BE ONE OF PRACTITIONER AND PROTECTOR, NOT PRISONER AND
Problems with Asset Loading

- How to load assets?
  - May have a lot of assets
  - May have large assets

- Loading is **blocking**
  - Game stops until done
  - Cannot draw or animate

- May need to **unload**
  - Running out of memory
  - Free something first
Problems with Asset Loading

- How to load assets?
  - May have a lot of assets
  - May have large assets

- Loading is blocking
  - Game stops until done
  - Cannot draw or animate

- May need to unload
  - Running out of memory
  - Free something first

Memory Management
Loading Screens

Minimal animation/feedback while loading assets
**Solution: Asynchronous Loader**

- **Game Thread**
  - **Update**
  - **Draw**

- **Second Thread**
  - **Asset Loader**

Diagram:
- Update -> Specify Asset -> Notify done -> Asset Loader -> Update and draw simple animations until assets loaded.

**Diagram Details**:
- The diagram illustrates the process of loading assets asynchronously in a game. The game thread updates, and the second thread manages the asset loader. The process involves specifying assets, loading them, and notifying the game thread to update and draw simple animations when the assets are loaded.
**Solution: Asynchronous Loader**

- Also an asset manager
  - Each asset given a key
  - Can access asset by key
  - Works like hash table
**Solution: Asynchronous Loader**

- Not always a good idea
  - May need OpenGL utils
  - **Example**: Textures
  - Limited to main thread
Alternative: Iterative Loader

Game Thread

Update

Draw

Asset Manager

Initialize

Update

Access
Alternative: Iterative Loader

- Uses a time budget
  - Give set amount of time
  - Do as much as possible
  - Stop until next update
- Better for OpenGL
  - Give time to manager
  - Animate with remainder
  - No resource contention
- LibGDX approach
  - CUQL is asynchronous
Alternative: Iterative Loader

- Uses a time budget
  - Give set amount of time
  - Do as much as possible
  - Stop until next update
- Better for OpenGL
  - Give time to manager
  - Animate with remainder
  - No resource contention
- LibGDX approach
  - CUGL is asynchronous
Assets Beyond Images

- AAA games have a lot of 3D geometry
  - Vertices for model polygons
  - Physics bodies **per polygon**
  - Scene graphs for organizing this data

- When are all these objects created?
  - At load time (filling up memory)?
  - Or only when they are needed?

- We need to understand memory better
Traditional Memory Organization

High Address

Stack

Free Space

Heap

Program Data

Program Code

Static Variables

Memory Management
Traditional Memory Organization

- Stack
  - Function parameters
  - Local variables
  - Return values
- Heap
- Program Data
  - Program Code
  - Static Variables
- Free Space
- High Address
- Low Address
Traditional Memory Organization

- Stack
  - Function parameters
  - Local variables
  - Return values

- Heap
  - Objects created via `new`
    - (e.g. Every object in Java)

- Program Data
  - Program Code
  - Static Variables

High Address

Low Address

Memory Management
Traditional Memory Organization

High Address

Stack

Function parameters
Local variables
Return values

Easy to Handle

Free Space

Objects created via new
(e.g. Every object in Java)

Low Address

Heap

Program Data

Program Code
Static Variables

Easy to Handle

Easy to Handle

Memory Management
Traditional Memory Organization

- Stack: Function parameters, Local variables, Return values
  - Easy to Handle
- Heap: Objects created via `new` (e.g., Every object in Java)
  - Problems!
- Free Space
- Program Data: Program Code, Static Variables
  - Easy to Handle

High Address

Low Address

Memory Management
Traditional Memory Organization

Dedicated to process.

Consists of machine addressable space.

Leverages Virtual Memory

High Address

Stack

Heap

Free Space

Program Data

Low Address

Function parameters
Local variables
Return values

Objects created via new Allocations with malloc

Program Code
Static Variables

Memory Management
Mobile Memory Organization

Device Memory

Heap

Stack
Program Data

Stack
Program Data

Stack
Program Data

Device Memory
How Do Apps Compete for Memory?

- Active app takes what it can
  - Cannot steal from OS
  - OS may *suspend* apps

- **App Suspension**
  - App quits; memory freed
  - **iOS**: 5 min (or so) on exit
  - **Android**: If needed

- Suspend apps can *recover*
  - OS allows limited paging
  - Page out on suspension
  - Page back in on restart
How Do Apps Compete for Memory?

- Active app takes what it can
  - Cannot steal from OS
  - OS may *suspend* apps

- **App Suspension**
  - App quits; memory freed
  - **iOS**: 5 min (or so) on exit
  - **Android**: If needed

- Suspend apps can *recover*
  - OS allows limited paging
  - Page out on suspension
  - Page back in on restart
How Do Apps Compete for Memory?

- Active app takes what it can
  - Cannot steal from OS
  - OS may *suspend* apps

- **App Suspension**
  - App quits; memory freed
  - **iOS**: 5 min (or so) on exit
  - **Android**: If needed

- Suspend apps can *recover*
  - OS allows limited paging
  - Page out on suspension
  - Page back in on restart

---

You must code this!
Otherwise, data is **lost**.

Can override in **iOS 7**
State Management in iOS 7+

- **Active**
  - Running & getting input

- **Inactive**
  - Running, but no input
  - Transition to suspended

- **Background**
  - Same as inactive
  - But apps can stay here
  - **Example**: Music

- **Suspended**
  - Stopped & Memory freed
State Management in iOS 7+

- **Active**
  - Running & getting input

- **Inactive**
  - Running, but no input
  - Transition to suspended

- **Background**
  - Same as inactive
  - But apps can stay here
  - **Example**: Music

- **Suspended**
  - Stopped & Memory freed

Write handlers to process entering, leaving each state
iOS State Handling

- **applicationDidBecomeActive:**
  - Your app became (resumed as) the foreground app.
  - Use this to recover memory state.

- **applicationWillResignActive:**
  - Your app will switch to inactive or background.
  - Stop the game loop and page out memory.

- **applicationDidEnterBackground:**
  - Your app is in the background and may be suspended.

- **applicationWillEnterForeground:**
  - Your app is leaving the background, but is not yet active.
Android State Handling

Activity launched

onCreate()
onStart()
onResume()
onPause()
onStop()
onDestroy()
Activity shut down

onRestart()

Activity launched

onCreate()
onStart()
onResume()
onPause()
onStop()
onDestroy()
Activity running

Activity launched

onCreate()
onStart()
onResume()
onPause()
onStop()
onDestroy()
Activity running

All methods in Application class

Memory Management
Android State Handling

All methods in **Application** class

Reload memory
Android State Handling

All methods in Application class

Page out memory
CUGL is Simplified Android Model

- **onStartup()**
  - Initialized and now active

- **onSuspend()**
  - Sent to background
  - Gives you chance to save
  - Also time to pause music

- **onResume()**
  - Returns to app to active
  - Allows you to restore state

- **onShutdown()**
  - Stopped & memory freed
CUGL is Simplified Android Model

- **onStartup()**
  - Initialized and now active

- **onSuspend()**
  - Sent to background
  - Gives you chance to save
  - Also time to pause music

- **onResume()**
  - Returns to app to active
  - Allows you to restore state

- **onLowMemory()**
  - Warning memory is low
  - Gives you chance to unload
  - Else app will shut down

- **onShutdown()**
  - Stopped & memory freed

Memory Management
Memory Organization and Games

- **Inter-Frame Memory**
  - Carries over across frame boundaries

- **Update**

- **Intra-Frame Memory**
  - Recovered each frame

- **Draw**
Memory Organization and Games

Inter-Frame Memory
Carries over across frame boundaries

Heap or Stack?
Does it matter?

Intra-Frame Memory
Recovered each frame

Draw
# Distinguishing Data Types

## Intra-Frame
- **Local computation**
  - Local variables (managed by compiler)
  - Temporary objects (not necessarily managed)
- **Transient data structures**
  - Built at the start of update
  - Used to process update
  - Can be deleted at end

## Inter-Frame
- **Game state**
  - Model instances
  - Controller state
  - View state and caches
- **Long-term data structures**
  - Built at start/during frame
  - Lasts for multiple frames
  - May adjust to data changes
Distinguishing Data Types

Intra-Frame

- Local computation
  - Local variables
    (managed by compiler)
  - Temporary objects
    (not necessarily managed)
- Transient data structures
  - Built at the start of update
  - Used to process update
  - Can be deleted at end

Inter-Frame

- Game state
  - Model instances
  - Controllers
  - View state and caches
- Long-term data structures
  - Built at start/during frame
  - Lasts for multiple frames
  - May adjust to data changes
Distinguishing Data Types

**Intra-Frame**
- Local computation
  - Local variables
    (managed by compiler)
  - Temporary objects
    (not necessarily managed)
- Transient data structures
  - Built at the start of update
  - Used to process update
  - Can be deleted at end

**Inter-Frame**
- Game state
  - Model instances
  - Controller state and caches
- Long-term data structures
  - Built at start or during frame
  - Lasts for multiple frames
  - Adjust to data changes

---

Memory Management
# Handling Game Memory

<table>
<thead>
<tr>
<th>Intra-Frame</th>
<th>Inter-Frame</th>
</tr>
</thead>
</table>
| • Does not need to be paged  
  • Drop the latest frame  
  • Restart on frame boundary | • Potential to be paged  
  • Defines current game state  
  • May just want level start |
| • Want size reasonably **fixed**  
  • Local variables always are  
  • Limited # of allocations  
  • Limit new inside loops | • Size is more **flexible**  
  • No. of objects is variable  
  • Subsystems may turn on/off  
  • User settings may affect |
| • Often use **custom allocator**  
  • GC at frame boundaries | • **OS allocator** okay, but…  
  • Recycle with **free lists** |
Handling Game Memory

**Intra-Frame**
- Does not need to be paged
- Drop the latest frame
- Restart on frame boundary
- Want size reasonably fixed
- Local variables always are
- Limited # of allocations
- Limit new inside loops
- Often use **custom allocator**
- GC at frame boundaries

**Inter-Frame**
- Potential to be paged
- Defines current game state
- Size is more flexible
- No. of objects is variable
- Subsystems may turn on/off
- User settings may affect
- **OS allocator** okay, but…
- Recycle with **free lists**

**Topic of Next Lecture**
Advanced: Spatial Loading

- Most game data is **spatial**
  - Only load if player nearby
  - Unload as player moves away
  - Minimizes memory used

- Arrange memory in **cells**
  - Different from a memory pool
  - Track player visibility radius
  - Load/unload via outer radius

- **Alternative**: loading zones
  - Elevators in *Mass Effect*
Most game data is spatial
- Only load if player nearby
- Unload as player moves away
- Minimizes memory used

Arrange memory in cells
- Different from a memory pool
- Track player visibility radius
- Load/unload via outer radius

Alternative: loading zones
- Elevators in Mass Effect
Advanced: Spatial Loading

- Most game data is *spatial*
  - Only load if player nearby
  - Unload as player moves away
  - Minimizes memory used

- Arrange memory in *cells*
  - Different from a memory pool
  - Track player visibility radius
  - Load/unload via outer radius

- **Alternative**: loading zones
  - Elevators in *Mass Effect*
Spatial Loading in *Assassin’s Creed*
Implementing Spatial Loading

- Part of serialization model
  - Level/save file has the cells
  - Cell *addresses* in memory
  - Load/page on demand
- Sort of like virtual memory
  - But paging strategy is spatial
Spatial Loading Challenges

• **Not same** as virtual memory
  • Objects unloaded do not exist
  • Do not save state when unload
  • Objects loaded are new created

• Can lead to *unexpected states*
  • “Forgetful” NPCs
  • Creative *Assassin’s Creed* kills

• **Workaround**: Global State
  • Track major game conditions
  • **Example**: Guards Alerted
  • Use to load objects in standard, but appropriate, configurations

See **Piazza** for
There is No Spoon
Next Time: Low-Level Details