Data Buffers

Mike Bailey
mjb@cs.oregonstate.edu

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License
Vulkan Pipeline Diagram

- Draw
- Input Assembler
- Vertex Shader
  - Tessellation Control Shader
  - Tessellation Primitive Generator
  - Tessellation Evaluation Shader
- Geometry Shader
  - Vertex Post-Processing
  - Rasterization
  - Early Per-Fragment Tests
- Fragment Shader
  - Late Post-Fragment Tests
- Blending

- Indirect Buffer
- Index Buffer
- Vertex Buffer

Descriptor Sets
- Push Constants
- Uniform Buffer
- Uniform Texel Buffers
- Sampled Images
- Storage Buffers
- Storage Texel Buffers
- Storage Images

Depth/Stencil Attachments
Input Attachments
Color Attachments

Some Vulkan commands specify geometric objects to be drawn or computational work to be performed, while others specify state controlling how objects are handled by the various pipeline stages, or control data transfer between memory organized as images and buffers. Commands are effectively sent through a processing pipeline, either a graphics pipeline or a compute pipeline.
A Vulkan **Data Buffer** is just a group of contiguous bytes in GPU memory. They have no inherent meaning. The data that is stored there is whatever you want it to be. (This is sometimes called a “Binary Large Object”, or “BLOB”.)

It is up to you to be sure that the writer and the reader of the Data Buffer are interpreting the bytes in the same way!

Vulkan calls these things “Buffers”. But, Vulkan calls other things “Buffers”, too, such as Texture Buffers and Command Buffers. So, I have taken to calling these things “Data Buffers” and have even gone to far as to override some of Vulkan’s own terminology:

```c
typedef VkBuffer VkDataBuffer;
```
Vulkan: Buffers

- `vkCreateBuffer()`
  - `VkBufferCreateInfo`
    - `bufferUsage`
    - `queueFamilyIndices`
    - `size (bytes)`

- `vkGetBufferMemoryRequirements()`
  - `Buffer`
  - `VkMemoryAllocateInfo`
    - `memoryType`
    - `size`

- `vkAllocateMemory()`
  - `LogicalDevice`

- `vkBindBufferMemory()`
  - `bufferMemoryHandle`

- `vkMapMemory()`
  - `gpuAddress`
VkBufferCreateInfo vbci;  
vbci.sType = VK_STRUCTURE_TYPE_BUFFER_CREATE_INFO;  
vbci.pNext = nullptr;  
vbci.flags = 0;  
vbci.size = << buffer size in bytes >>  
vbci.usage = <<or’ed bits of: >>  
   VK_USAGE_TRANSFER_SRC_BIT  
   VK_USAGE_TRANSFER_DST_BIT  
   VK_USAGE_UNIFORM_TEXEL_BUFFER_BIT  
   VK_USAGE_STORAGE_TEXEL_BUFFER_BIT  
   VK_USAGE_UNIFORM_BUFFER_BIT  
   VK_USAGE_STORAGE_BUFFER_BIT  
   VK_USAGE_INDEX_BUFFER_BIT  
   VK_USAGE_VERTEX_BUFFER_BIT  
   VK_USAGE_INDIRECT_BUFFER_BIT  
vbci.sharingMode = << one of: >>  
   VK_SHARING_MODE_EXCLUSIVE  
   VK_SHARING_MODE_CONCURRENT  
vbci.queueFamilyIndexCount = 0;  
vbci.pQueueFamilyIndices = (const iont32_t) nullptr;  

VkBuffer Buffer;  

result = vkCreateBuffer ( LogicalDevice, IN &vbci, PALLOCATOR, OUT &Buffer );
VkMemoryRequirements
result = vkGetBufferMemoryRequirements( LogicalDevice, Buffer, OUT &vmr );

VkMemoryAllocateInfo
vmai.sType = VK_STRUCTURE_TYPE_MEMORY_ALLOCATE_INFO;
vmai.pNext = nullptr;
vmai.flags = 0;
vmai.allocationSize = vmr.size;
vmai.memoryTypeIndex = FindMemoryThatIsHostVisible( );

... VkDeviceMemory
result = vkAllocateMemory( LogicalDevice, IN &vmai, PALLOCATOR, OUT &vdm );

result = vkBindBufferMemory( LogicalDevice, Buffer, IN vdm, 0 );  // 0 is the offset

... result = vkMapMemory( LogicalDevice, IN vdm, 0, VK_WHOLE_SIZE, 0, &ptr );

<< do the memory copy >>

result = vkUnmapMemory( LogicalDevice, IN vdm );

Vulkan: Allocating Memory for a Buffer, Binding a Buffer to Memory, and Writing to the Buffer
Finding the Right Type of Memory

```c
int FindMemoryThatIsHostVisible()
{
    VkPhysicalDeviceMemoryProperties vpdmp;
vkGetPhysicalDeviceMemoryProperties( PhysicalDevice, OUT &vpdmp );
    for( unsigned int i = 0; i < vpdmp.memoryTypeCount; i++ )
    {
        VkMemoryType vmt = vpdmp.memoryTypes[ i ];
        if( ( vmt.propertyFlags & VK_MEMORY_PROPERTY_HOST_VISIBLE_BIT ) != 0 )
        {
            return i;
        }
    }
    return -1;
}
```
Finding the Right Type of Memory

```c
int FindMemoryThatIsDeviceLocal()
{
    VkPhysicalDeviceMemoryProperties vpdmp;
vkGetPhysicalDeviceMemoryProperties( PhysicalDevice, OUT &vpdmp );
for( unsigned int i = 0; i < vpdmp.memoryTypeCount; i++ )
{
    VkMemoryType vmt = vpdmp.memoryTypes[ i ];
    if( ( vmt.propertyFlags & VK_MEMORY_PROPERTY_DEVICE_LOCAL_BIT ) != 0 )
    {
        return i;
    }
}
return -1;
}
```
# Finding the Right Type of Memory

```
VkPhysicalDeviceMemoryProperties vpdmp;
vkGetPhysicalDeviceMemoryProperties( PhysicalDevice, OUT &vpdmp );
```

<table>
<thead>
<tr>
<th>Memory Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>0: DeviceLocal</td>
</tr>
<tr>
<td>1: DeviceLocal</td>
</tr>
<tr>
<td>2: DeviceLocal</td>
</tr>
<tr>
<td>3: HostVisible HostCoherent</td>
</tr>
<tr>
<td>4: HostVisible HostCoherent</td>
</tr>
<tr>
<td>5: HostVisible HostCoherent</td>
</tr>
<tr>
<td>6: HostVisible HostCoherent</td>
</tr>
<tr>
<td>7: HostVisible HostCoherent</td>
</tr>
<tr>
<td>8: HostVisible HostCoherent</td>
</tr>
<tr>
<td>9: HostVisible HostCoherent</td>
</tr>
<tr>
<td>10: HostVisible HostCoherent</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Memory Heaps</th>
</tr>
</thead>
<tbody>
<tr>
<td>0: size = 0xb7c00000 DeviceLocal</td>
</tr>
<tr>
<td>1: size = 0xfac00000</td>
</tr>
</tbody>
</table>
I find it handy to encapsulate buffer information in a struct:

```c
typedef struct MyBuffer {
    VkDataBuffer buffer;
    VkDeviceMemory vdm;
    VkDeviceSize size;
} MyBuffer;

MyBuffer MyMatrixUniformBuffer;
```

It’s the usual object-oriented benefit – you can pass around just one data-item and everyone can access whatever information they need.
It’s the usual object-oriented benefit – you can pass around just one data-item and everyone can access whatever information they need.

```c
VkResult Init05DataBuffer( VkDeviceSize size, VkBufferUsageFlags usage, OUT MyBuffer * pMyBuffer )
{
    ..
    vbci.size = pMyBuffer->size = size;
    ..
    result = vkCreateBuffer( LogicalDevice, IN &vbci, PALLOCATOR, OUT &pMyBuffer->buffer );
    ..
    pMyBuffer->vdm = vdm;
    ..
}
```
Here’s a C struct to hold some uniform variables

```c
struct matBuf
{
    glm::mat4 uModelMatrix;
    glm::mat4 uViewMatrix;
    glm::mat4 uProjectionMatrix;
    glm::mat3 uNormalMatrix;
} Matrices;
```

Here’s the shader code to access those uniform variables

```glsl
layout( std140, set = 0, binding = 0 ) uniform matBuf
{
    mat4 uModelMatrix;
    mat4 uViewMatrix;
    mat4 uProjectionMatrix;
    mat4 uNormalMatrix;
} Matrices;
```
uint32_t Height, Width;
const double FOV = glm::radians(60.); // field-of-view angle

glm::vec3 eye(0.,0.,EYEDIST);
glm::vec3 look(0.,0.,0.);
glm::vec3 up(0.,1.,0.);

Matrices.uModelMatrix = glm::mat4(); // identity
Matrices.uViewMatrix = glm::lookAt( eye, look, up );

Matrices.uProjectionMatrix = glm::perspective( FOV, (double)Width/(double)Height, 0.1, 1000. );
Matrices.uProjectionMatrix[1][1] *= -1.; // account for Vulkan's LH screen coordinate system

Matrices.uNormalMatrix = glm::inverseTranspose( glm::mat3( Matrices.uModelMatrix ) );
This C struct is holding the actual data. It is writeable by the application.

```
struct matBuf Matrices;
```

The MyBuffer does not hold any actual data itself. It just represents the collection of data buffer information that will be used by Vulkan.

```
MyBuffer MyMatrixUniformBuffer;
```

The Data Buffer in GPU memory is holding the actual data. It is readable by the shaders.

```
uniform matBuf Matrices;
```

There is one more step in here—Descriptor Sets. Here’s a quick preview…
The Descriptor Set for the Buffer

We will come to **Descriptor Sets** later, but for now think of them as the link between the BLOB of uniform variables in GPU memory and the block of variable names in your shader programs.

```c
VkDescriptorBufferInfo vdbi0;
    vdbi0.buffer = MyMatrixUniformBuffer.buffer;
    vdbi0.offset = 0;       // bytes
    vdbi0.range = sizeof(Matrices);

VkWriteDescriptorSet vwds0;
    // ds 0:
    vwds0.sType = VK_STRUCTURE_TYPE_WRITE_DESCRIPTOR_SET;
    vwds0.pNext = nullptr;
    vwds0.dstSet = DescriptorSets[0];
    vwds0.dstBinding = 0;
    vwds0.dstArrayElement = 0;
    vwds0.descriptorCount = 1;
    vwds0.descriptorType = VK_DESCRIPTOR_TYPE_UNIFORM_BUFFER;
    vwds0.pBufferInfo = &vdbi0;
    vwds0.pImageInfo = (VkDescriptorImageInfo *)__nullptr;

vkUpdateDescriptorSets( LogicalDevice, 1, IN &vwds0, IN 0, (VkCopyDescriptorSet *)__nullptr );
```
Filling the Data Buffer

```c
typedef struct MyBuffer
{
    VkDataBuffer buffer;
    VkDeviceMemory vdm;
    VkDeviceSize size;
} MyBuffer;

// ...

MyBuffer MyMatrixUniformBuffer;

Init05UniformBuffer( sizeof(Matrices), &MyMatrixUniformBuffer );
Fill05DataBuffer( MyMatrixUniformBuffer, (void*) &Matrices );

glm::vec3 eye(0.0,0.0,EYEDIST);
glm::vec3 look(0.0,0.0);
glm::vec3 up(0.0,1.0);

Matrices.uModelMatrix = glm::mat4();   // identity
Matrices.uViewMatrix = glm::lookAt( eye, look, up );

Matrices.uProjectionMatrix = glm::perspective( FOV, (double)Width/(double)Height, 0.1, 1000.0 );
Matrices.uProjectionMatrix[1][1] *= -1.;

Matrices.uNormalMatrix = glm::inverseTranspose( glm::mat3( Matrices.uModelMatrix ) );
```
Creating and Filling the Data Buffer – the Details

VkResult
Init05DataBuffer( VkDeviceSize size, VkBufferUsageFlags usage, OUT MyBuffer * pMyBuffer )
{
    VkResult result = VK_SUCCESS;
    VkBufferCreateInfo vbci;
    vbci.sType = VK_STRUCTURE_TYPE_BUFFER_CREATE_INFO;
    vbci.pNext = nullptr;
    vbci.flags = 0;
    vbci.size = pMyBuffer->size = size;
    vbci.usage = usage;
    vbci.sharingMode = VK_SHARING_MODE_EXCLUSIVE;
    vbci.queueFamilyIndexCount = 0;
    vbci.pQueueFamilyIndices = (const uint32_t *)nullptr;
    result = vkCreateBuffer ( LogicalDevice, IN &vbci, PALLOCATOR, OUT &pMyBuffer->buffer );

    VkMemoryRequirements vmr;
    vkGetBufferMemoryRequirements( LogicalDevice, IN pMyBuffer->buffer, OUT &vmr ); // fills vmr

    VkMemoryAllocateInfo vmai;
    vmai.sType = VK_STRUCTURE_TYPE_MEMORY_ALLOCATE_INFO;
    vmai.pNext = nullptr;
    vmai.allocationSize = vmr.size;
    vmai.memoryTypeIndex = FindMemoryThatIsHostVisible( );

    VkDeviceMemory vdm;
    result = vkAllocateMemory( LogicalDevice, IN &vmai, PALLOCATOR, OUT &vdm );
    pMyBuffer->vdm = vdm;

    result = vkBindBufferMemory( LogicalDevice, pMyBuffer->buffer, IN vdm, 0 ); // 0 is the offset
    return result;
}
Creating and Filling the Data Buffer – the Details

```
VkResult
Fill05DataBuffer( IN MyBuffer myBuffer, IN void * data )
{
    // the size of the data had better match the size that was used to Init the buffer!
    void * pGpuMemory;
    vkMapMemory( LogicalDevice, IN myBuffer.vdm, 0, VK_WHOLE_SIZE, 0, OUT &pGpuMemory );
        // 0 and 0 are offset and flags
    memcpy( pGpuMemory, data, (size_t)myBuffer.size );
    vkUnmapMemory( LogicalDevice, IN myBuffer.vdm );
    return VK_SUCCESS;
}
```

Remember – to Vulkan and GPU memory, these are just bits. It is up to you to handle their meaning correctly.