Outline

- About
    - A GPU is a graphics processing unit, primarily designed to render 3-D objects. It is an independent processor designed specifically for graphics calculations, which can occur in parallel. If these calculations were done on a CPU, it would be too intensive, as a CPU would have to serially calculate each pixel result.
    - A GPGPU is a general purpose GPU that can compute more than just graphics. Utilizing a GPU's highly parallel nature, a GPGPU can do calculations for physics, encryption.
    - CPUs vs GPUs
        - Latency Intolerance versus Latency Tolerance
        - Task Parallelism versus Data Parallelism
        - Multi-threaded Cores versus SIMT (Single Instruction Multiple Thread) Cores
        - 10s of Threads versus 10,000s of Threads
- Rendering pipeline
    - A rendering pipeline is a series of steps a GPU takes to render objects. Here, we look at a basic GPU, following the OpenGL process. 3-D models are converted to vertices and (triangular) primitives before getting rasterized and render onto a 2-D screen.
    - Definitions
        - Vertex specification - setting up necessary objects for rendering
        - Vertex shader - program that handles individual vertices processing
        - Tessellation - vertex data are divided into smaller primitives
        - Geometry shader - program that processes primitives
        - Vertex post-processing - setup for primitive assembly and rasterization
        - primitive assembly. Primitives are clipped.
        - rasterization - individual primitives are broken into fragments
        - fragment shader - program that processes fragments into a set of colors and a single depth value
        - per-sample operations - processes fragments and saves to various buffers
    - GPUs are a relatively new phenomena. Originally, rendering logic were embedded in hardware and have individual roles. More recently, GPUs are

leaning more towards a **unified shader model**, where a single compute unit runs various programmable shaders/kernels.

These more-programmable and unified compute units allow for more general-purpose use, allowing us to have a GPGPU.

- Evolution of GPU
    - The first GPU, Nvidia 256, can process over 10 million polygons per second.
- MIAOW
    - Specific parts - in own words
    - Process (how does a instruction get processed)

Citations

- http://www.webopedia.com/TERM/G/GPU.html
- http://whatis.techtarget.com/definition/GPGPU-general-purpose-graphics-processing-unit
- https://www.opengl.org/wiki/Rendering_Pipeline_Overview
- https://github.com/VerticalResearchGroup/miaow