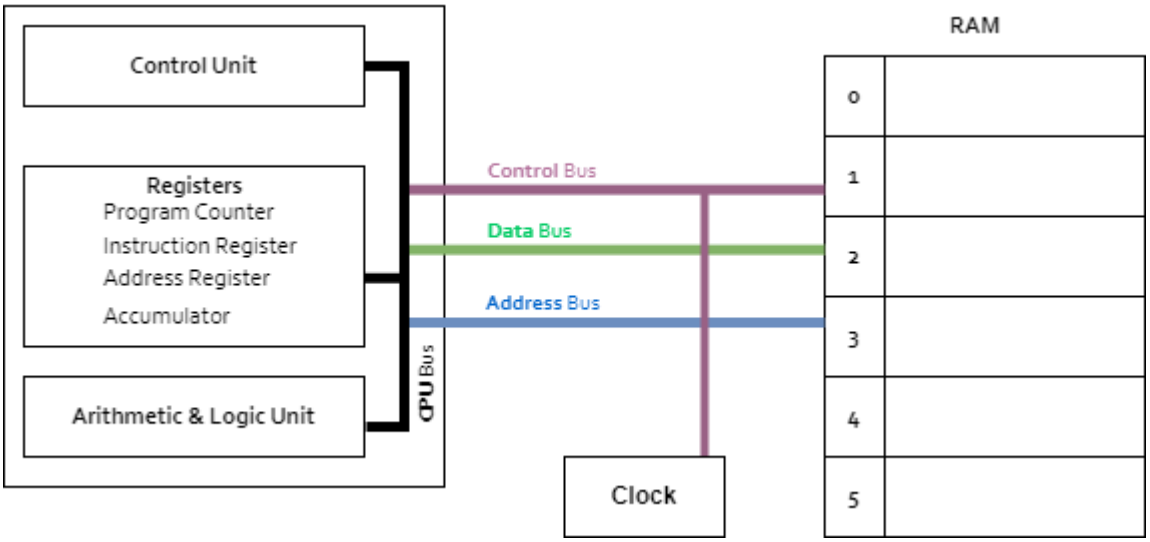
# // Assignment 2

## / The Data Driver

// by George Hotten

# // How the key components interact




## // Clock

The Clock is located near the CPU and is responsible for synchronising the processor and when the FDE cycle starts. The speed of the clock is determined by an oscillator crystal. The clock speed is measured in hertz.

The clock sends pulses through the control bus to reach the CPU and its components.

## // The CPU and Memory

The CPU is responsible for performing the FDE cycle, however the instructions needed are stored in memory so the CPU must use busses to communicate. The CPU requests data along the address bus and memory sends data back along the data bus. The CPU then stores the instruction in the appropriate register and proceeds with the cycle.

## // Registers

### / Program Counter

The program counter stores the location in memory of the next instruction to be executed.

### / Instruction Register

The instruction register stores the current instruction being executed. For example, the operand code 5 may be stored indicating the current instruction is load from memory into the accumulator.

### / Address Register

The address register stores the memory location that is in-use by the instruction. For example, if the instruction was 506, this would mean load from location 6 in memory into the accumulator. This means 06 would be stored in this register.

### / Accumulator

The accumulator stores the output of any operation by the Arithmetic Logic Unit.

# // Memory attached to the processor

## // Direct Memory Access

Direct Memory Access is a feature that allows other devices, such as I/O and storage, to access RAM without going through the CPU. This allows operations to be drastically quicker as it frees up the CPU to perform other actions whilst the DMA controller performs the data transfer. The DMA controller is attached to the processor via I/O bus and the system bus.

The use the DMA controller, the device must request permission from the CPU. Once the device has got permission, it can use the DMA controller to read and write data from memory. Once the transfer is complete, an interrupt it sent to the CPU to notify it of the completion.

## // Read Only Memory

Read Only Memory is storage that is located on the motherboard which is usually responsible for storing the system's BIOS. The ROM chip is not writable and is considered non-volatile meaning data is not lost when power is. The ROM chip communicates with the CPU via a I/O or storage bus.

## // Cache

Cache is usually found in levels built directly into the CPU. Level 1 is the fastest and smallest whilst level 3 is the slowest and largest. Cache holds frequently used instructions to speed up the time needed to complete a cycle; this is because it is quicker to access the cache which is directly on the CPU compared to accessing data in RAM.

## // Random Access Memory

Random Access Memory (RAM) is located next to the CPU and is accessed via the address and data buses. RAM stores the operating system, instructions, and data. To request data, the address bus is used and to receive data the data bus is used.

### / Dynamic RAM

Dynamic RAM (DRAM) is the most common form of memory. DRAM cells contain a transistor and a capacitor and can contain 1 bit. Data is stored in the capacitor by filling it with electrons. However, capacitors leak electrons very quickly causing the capacitor to be empty within milliseconds. This means that the CPU or the memory controller must constantly refresh and re-write data by filling the capacitor up again.

### / Static RAM

Static RAM (SRAM) uses a form of boolean logic called flip-flops to store data. This means SRAM cells contain four to six transistors and some wiring to store data. However, this takes up a lot of space meaning that you often get lower capacity SRAM per stick compared to DRAM.

# // Comparing different types of memory

## / *Volatile vs Non-volatile memory*

Volatile memory requires power to retain data. If power is lost, data is lost.
Non-volatile memory doesn't need power to retain data.

## // Read Only Memory

Read Only Memory (ROM) is **non-volatile** memory and can hold **4-8MB** per chip. ROM can typically read data at x mb/s. ROM's primary purpose in general computers is to hold the BIOS needed for the system to operate.

## // Random Access Memory

### / Dynamic RAM

Dynamic RAM (DRAM) is **volatile** memory and can hold up to **256GB** per stick. On modern DDR4 (Double Data Rate 4) sticks of DRAM, data can be transferred at up to **4800 mega-transfers per second**. DRAM's primary purpose in general computers is to hold the operating system, instructions, and other application data whilst the system is running.

### / Static RAM

Static RAM (SRAM) is **volatile** memory and can hold around **1-2MB** per chip. SRAM can typically read and write data at **100 megabytes per second**. The purpose of SRAM in general computers is to act as small high-speed memory banks, such as registers on the CPU or frame buffers on a GPU.

## // Cache

Cache is **volatile** memory, and its capacity varies by CPU. In this example, I will use data from a Ryzen 7 5800X.

### / L1 Cache

L1 cache has the most variance between CPUs. On our Ryzen chip, there is **32KB** per core, for a total of **256KB**, and operates at the same frequency as the CPU, typically around **4GHz**. The purpose of L1 cache is to hold frequently used instructions and data. L1 cache has a latency of around **1 nano second**.

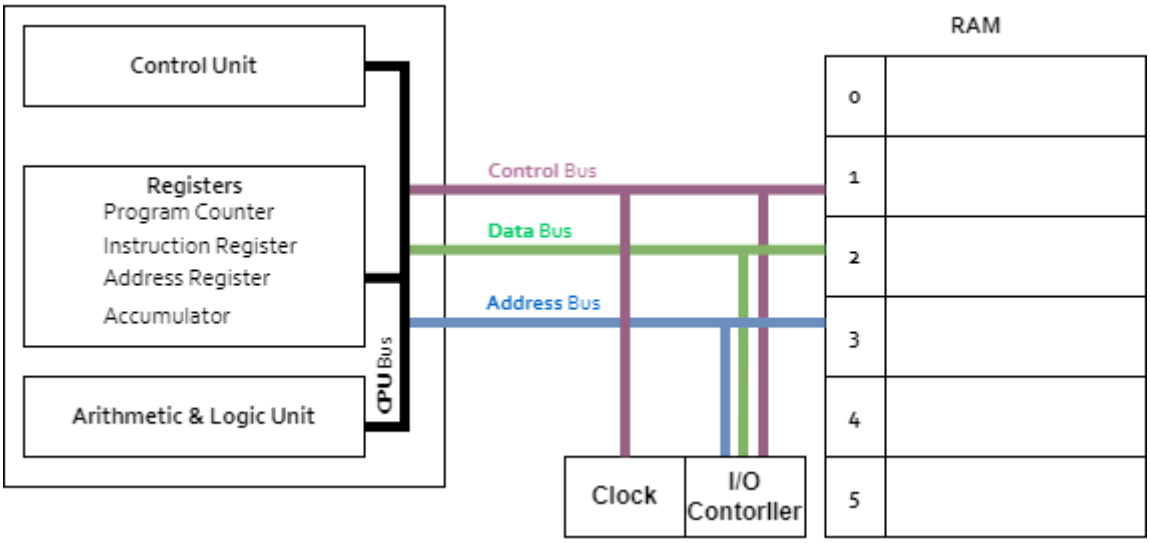
### / L2 Cache

L2 cache holds between **256KB to 8MB** per core. On our Ryzen chip, there is **512KB** per core, for a total of **4MB**, and operates around the same frequency of the CPU but not as fast as L1. L2 cache holds similar data to the L1 cache such as frequently used instructions. L2 cache has a latency of around **7 nano seconds**.

### / L3 Cache

L3 cache can hold up to **64MB** per CPU. Unlike L1 and L2, L3 cache is shared between all cores. Despite being slower than L1 and L2 cache, it is still faster than system memory and often feeds data down into the lower caches when needed. L3 cache holds instructions or data that isn't required frequently enough to need the top speeds of L1 or L2 cache. L3 cache has a latency of around **30 nano seconds**.

# // How the processor is connected with system buses




## // Definitions

### / Control Bus

The control bus is used to carry signals to different components within a computer. The bus is bidirectional and helps the CPU synchronise components such as the clock. It also regulates which direction data is being written to or read from. Before data is written to memory a signal is sent along the control bus so the CPU and memory controller know what action is about to happen.

### / Address Bus

The address bus is a unidirectional bus that carries the memory address needed for a read/write operation to other components such as system memory or secondary storage.

### / Data Bus

The data bus is a bidirectional bus that carries data between the CPU and other components such as system memory.

## // How the CPU connects to system memory

To read or write data from system memory, the CPU must first use the control bus to indicate what operation it intendents to complete. In this example we will read data from memory, meaning the read line will be activated via the control bus. Once that signal has been sent, the CPU can then use the address bus to send a signal to memory containing the address it wants to read data from. Memory then sends the requested data back along the data bus to the CPU.

## // How the CPU connects to the I/O via Direct Memory Access

### / What is Direct Memory Access?

Direct Memory Access (DMA) is a controller on the motherboard that allows for components, such as I/O, to access system memory without going through the CPU. This allows the CPU to continue executing instructions without being overran by requests to read/write data to memory.

## / Completing Transfers

To complete transfers, the DMA must know whether it is performing a read or write operation, the address of the I/O device, and the address of memory to be written or read to. The DMA is connected to the I/O bus allowing only one cycle to be taken up per request.

Using similar buses from reading from memory, the control bus regulates which action will be performed. The address bus and data bus are used to request and receive data respectively. Once the I/O device has finished its action, an interrupt is sent to notify the CPU of the completion. This interrupt is sent by the control bus.