

Introduction to Microcontrollers Notes

James Gowans

August 6, 2014

Licence

This work is licensed under the Creative Commons Attribution-ShareAlike 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-sa/3.0/> or send a letter to Creative Commons, 444 Castro Street, Suite 900, Mountain View, California, 94041, USA.

Contents

1	System Overview	4
1.1	What is a Microcontroller?	4
1.1.1	Development board block diagram	5
2	Memory Model	9
2.1	Data Types and Endianness	9
3	The ARM Cortex-M0	12
3.1	Programmer's Model of the CPU	12
3.2	CPU Architecture	12
3.3	Program Counter	13
3.3.1	Three stage pipeline	14
3.4	Reset Vector	14
4	Coding	16
4.1	Assembly	16
4.1.1	Instruction Sets	16
4.2	Linking	17
4.2.1	Executing Code	17
5	Loading and Storing	18
5.1	Immediate Offset Loading	18
5.1.1	Offset restrictions	18
5.2	Program Counter Relative Loading	19
5.3	Register Offset Loading	20
5.4	Storing	20

1 System Overview

1.1 What is a Microcontroller?

The microcontroller can be understood by comparing it to something you are already very familiar with: the computer. Both a microcontroller and a computer can be modeled as a black box which takes in data and instructions, performs processing, and provides output. In order to do this, a micro has some of the same internals as a computer, shown graphically in [Figure 1.1](#) and discussed now:

- CPU: The section of the microcontroller which does the processing. It executes instructions which allows it to do arithmetic and logic operations, amongst other forms of operations.
- Volatile memory (RAM:): This is general purpose memory. It can be used for storing whatever you want to store in it. Typically it stores variables which are created or changed during the course of execution of a program.
- Non-volatile memory (Flash): This non-volatile memory is used to store any data which must not be lost when the power to the micro is removed. Typically this would include the program code and any constants or initial values of data.
- Ports: Interfaces for data to move in and out of the micro. This allow it to communicate with the outside world.

These resources are typically orders of magnitude smaller or a micro than on a conventional computer. A micro makes up for this lack of resources with a small size, low power and low cost. A comparison of the characteristics can be seen in [Table 1.1](#). A computer is typically defined as a multi-purpose, flexible unit able to do computation. A microcontroller on the other hand typically is hard-coded to do one specific job.

The terms *microcontroller* and *microprocessor* are different and should not be used interchangeably. A *microprocessor* is an IC which is able to perform computation, but requires external memory and peripherals to function. A *microcontroller* has the memory and peripherals built into it, allowing it to be fully independent. Furthermore, the interface in and out of a microprocessor is mainly just an address and data bus. In a microcontroller, these busses are internal to the device. The interfaces in and out of a microcontroller are configurable to be a wide variety of communication standards. This self-contained nature and ability to deal with a wide variety of signals allows a microcontroller to (as the name suggest) be embedded in a larger

	CPU	RAM	Non-volatile	Power	Size/Mass	Cost
Computer	Dual, 3 GHz	4 GiB	500 GB	100 W	Large	R 3000
Micro	48 MHz	8 KiB	32 KiB	50 mW	Small	R 15

Table 1.1: Comparison of specs of entry level computer to STM32F051C6.

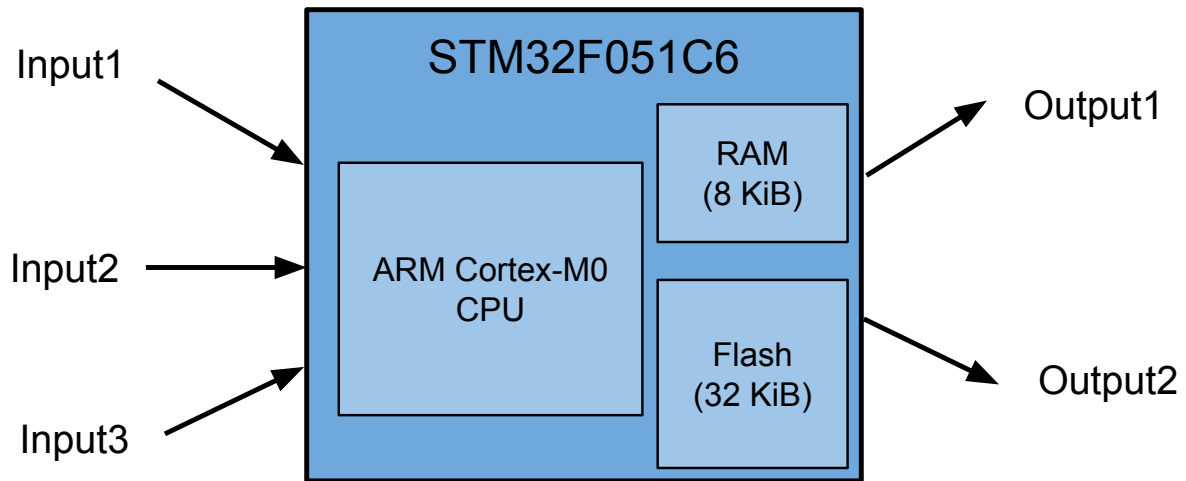


Figure 1.1: The most simplified view of the internals of the STM32F051

system and perform control and monitoring functions.

The micro we will be using is the STM32F051C6. It is manufactured by ST Microelectronics, but has an ARM Cortex-M0 CPU. ARM designed the CPU (specified how the transistors connect together). ST then takes this CPU design, adds it to their design for all of the other bits of the micro (flash, RAM, ports and much much more) and then produces the chip.

1.1.1 Development board block diagram

The development board consists of modules which connect to the microcontroller. Most of these modules are optional in that they are not required for the microcontroller to run. We will develop code later in the course to interface with some of these modules. Those which are not optional are the voltage regulator and the debugger. Following is a brief discussion of the purpose of each of the dev board modules (peripherals).

- STM32F051C6: This is the target microcontroller. It is connected to everything else on the board and it is where the code which we develop will execute.
- Debugger: this is essentially another microcontroller running special code on it which allows it to be able to pass information between a computer and the target microcontroller. The interface to the computer is a USB connection, and the interface to the target is a protocol called Serial Wire Debug (SWD) which is similar to JTAG. The specific type of debugger which we have is a ST-Link.
- Regulator: A MCP1702-33/T0 chip. This converts the 5 V provided by the USB port into 3.3 V suitable for running most of the circuitry on the board.
- LEDs: One byte of LEDs, active high connected to the lower byte of port B.
- Push buttons: Active low push buttons connected to the lower nibble of port A.
- Pots: 2 x 10K (or thereabouts) potentiometers connected to PA5 and PA6.
- LCD Screen: A 16x2 screen connected to the micro in 4-bit mode. Used to display text.

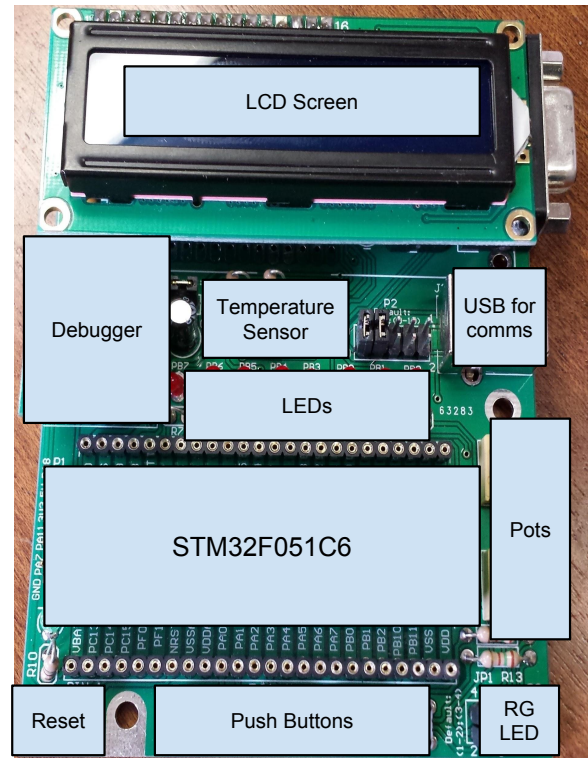
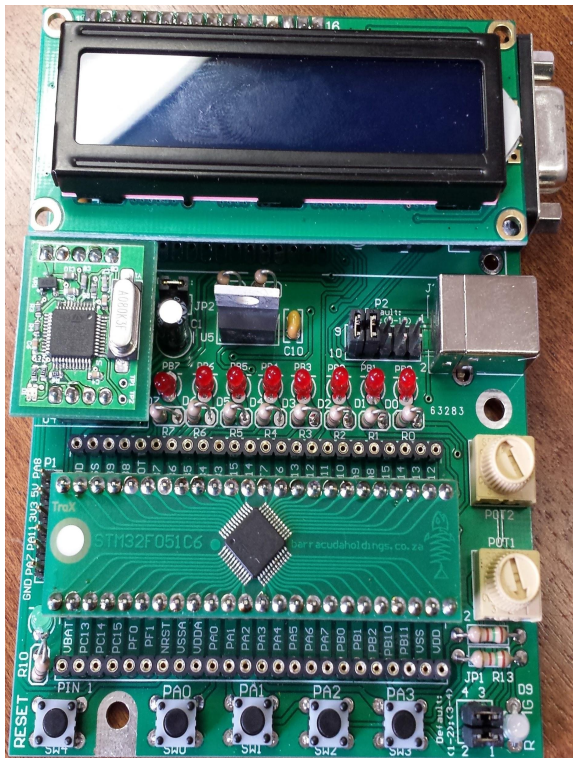
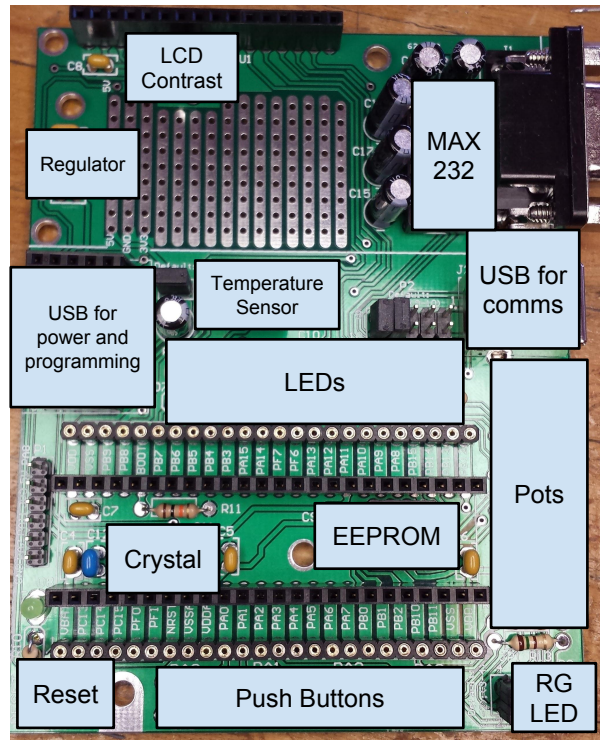
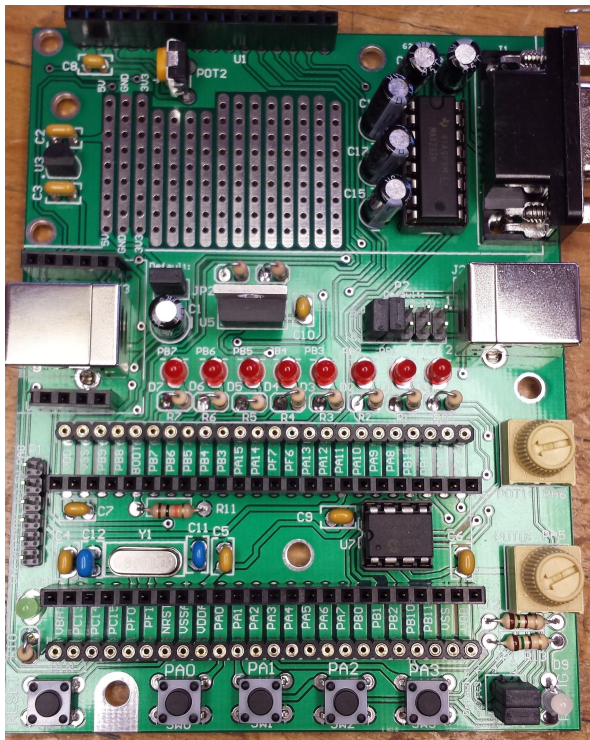


Figure 1.2: Modules on the dev board as seen when top boards unplugged or plugged in.

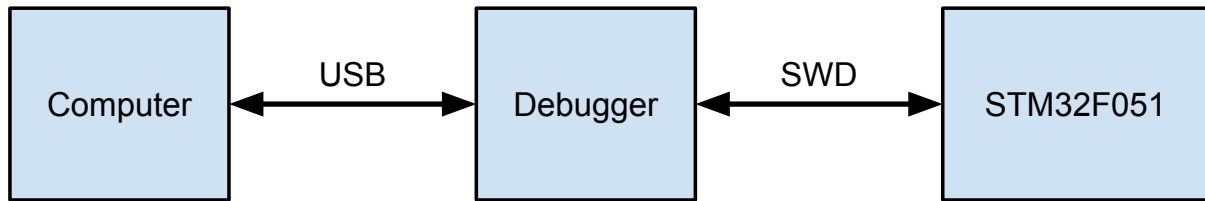
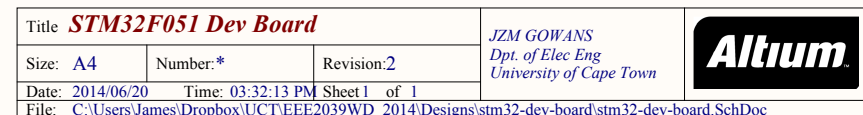


Figure 1.3: Highly simplified diagram showing how micro and computer communicate

- LCD contrast pot: The output of this potentiometer connects to the contrast pin of the LCD screen, hence allowing contrast adjustment.
- MAX232: This chip translates between TTL or CMOS logic level UART traffic and bi-polar higher voltage RS-232 traffic. Used for industrial communications links.
- USB for comms: The header allows intercepting of the UART traffic before it gets to the MAX232 and converting it to USB traffic through a small board which plugs into that header. When this facility is not being used, the jumpers on the header should be placed to allow the UART traffic to make its way to the MAX232.
- Temperature sensor: A TC74-A0 I^2C temperature sensor.
- Crystal: 8 MHz quartz oscillator with 10 pF caps for removing high frequency harmonics.
- EEPROM: A 25LC640A 64Kb Electronically Erasable and Programmable Read Only Memory (EEPROM) chip which communicates over SPI.
- RG LED: Common cathode Red/Green LED.

The full circuit schematic for the board follows. For now, we will forget about all of the other modules on the dev board and consider our system to be a computer talking to a debugger talking to a target micro, as shown in [Figure 1.3](#). This is the most basic system which must be understood to allow us to load code onto the target microcontroller.



2 Memory Model

We will now begin to expand on some of the block in [Figure 1.1](#). Before starting to explore how the CPU works, it's useful to have an understanding of how memory is laid out. We will start looking at the flash and RAM blocks. Together with another block called peripherals (which we will explore later), these blocks make up memory. It's important to note that this memory is located *outside* of the CPU, but still inside the microcontroller IC.

The memory of a device can be thought of as a very long row of post boxes along a street. Each post box has an address, and each post box can have data put into it or taken out. The amount of data that each post box can hold is 8 bits, or one byte. Therefore, each memory address is said to address one byte. The address of each post box is 32 bits long, meaning that addresses range from 0 (0x00000000) to just over 4 billion (0xFFFFFFFF). In actual fact, the *vast* majority of these addresses do not have a post box at them. These addresses are said to be unimplemented. Only very small sections of this address space are implemented and can actually be read from or written to. Flash and RAM are contiguous blocks of memory, with a start address and an end address. A simplified memory map of the STM32F051 is shown in [Figure 2.1](#). From this, we can see that if we want to use changeable variables in our programs, the variables should be located at addresses between 0x2000 0000 and 0x2000 1FFF. If we want to load code onto the micro which should not be lost when the device loses power, the code should be loaded into addresses between 0x0800 0000 and 0x0800 7FFF.

2.1 Data Types and Endianness

Very often we will need to work with clumps of data which are larger than 1 byte. ARM defines datatypes for a 32 bit CPU as follows:

- byte: 8 bits
- halfword: 16 bits
- word: 32 bits
- doubleword: 64 bits

Each memory address only addresses one byte of memory, so how can something like a word (four bytes) be stored in memory? Obviously, the four bytes have to come after each other to form a four byte block, or word. However, it is not obvious which order they should come in. For example, consider the case of wanting to store the word 0xAABBCCDD in address 0. The two possible ways of doing it are shown in [Table 2.1](#). It doesn't really matter which one of these schemes is used - they each have their pros and cons and different processors use different methods. It is important to know which one our processor has chosen to use. Our processor uses little endian. A more abstract view of how data is stored in our processor is given in [Figure 2.2](#)

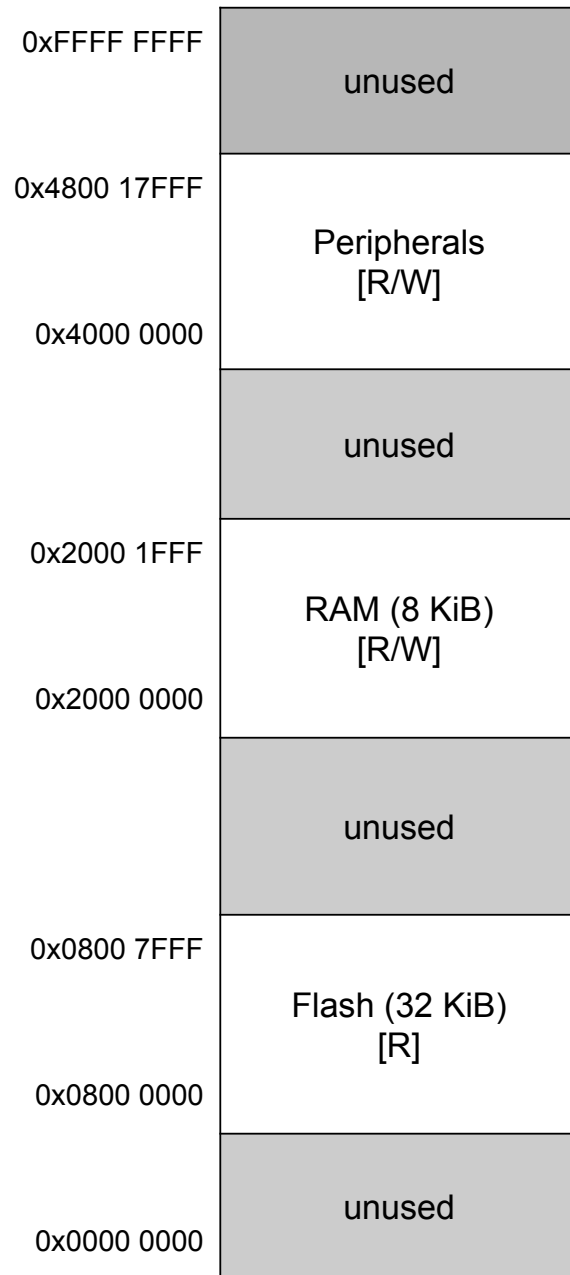


Figure 2.1: Simplified STM32F051C6 memory map. Note how all addresses are 32 bits. The blocks are very much not to scale. Source: datasheet, Figure 9

Little Endian		Big Endian	
Address	Data	Address	Data
3	0xAA	3	0xDD
2	0xBB	2	0xCC
1	0xCC	1	0xBB
0	0xDD	0	0xAA

Table 2.1: Layouts of the word 0xAABBCCDD in memory at effective address 0, according to little or big endian format

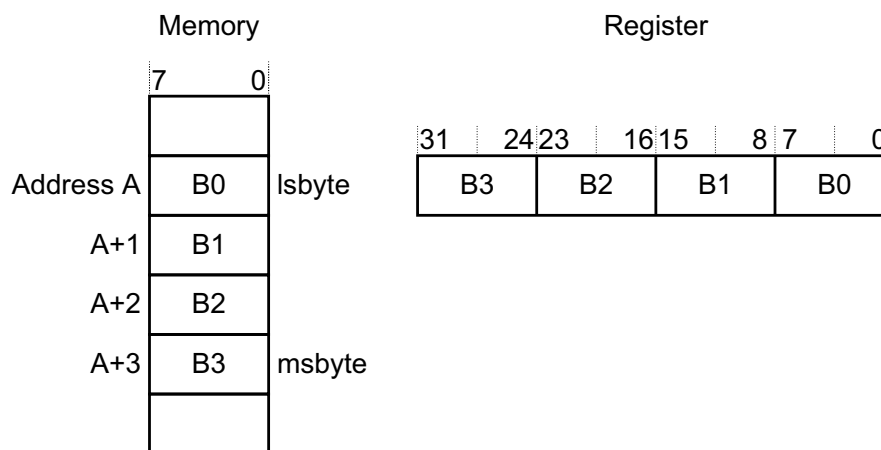


Figure 2.2: More abstract view of little endian layout. Source: Prog Man, page 28

3 The ARM Cortex-M0

At the core of a microcontroller is the CPU. Our CPU is called the Cortex-M0 and is designed by Advanced RISC Machines (ARM). The ARM Cortex-M0 CPU is certainly the most interesting block inside the STM32F051C6. This is where all processing happens, hence this is where the instructions which we write will run. It is therefore essential that we have an intricate understanding of the CPU so that we may write useful code for it. This chapter seeks to explore the CPU in some detail.

3.1 Programmer's Model of the CPU

A programmer's model is a representation of the inner workings of the CPU with sufficient detail to allow us to develop code for the CPU, but no unnecessary detail. The expanded view of the CPU which will now be discussed can be seen in [Figure 3.1](#). This simple model of a CPU is a set of CPU registers, an Arithmetic and Logic Unit (ALU) and a control Unit. The CPU registers are blocks of storage each 32 bits wide which the CPU has the ability to operate on. Only data which is inside a CPU register can be operated on by the CPU. The ARM Cortex-M0 has 16 such registers.

The ALU is that which performs the operations on the registers. It can take data from registers as inputs, do very basic processing and store the result in CPU registers.

The control unit manages execution by telling the ALU what to do. Together, the registers, ALU and control are able to execute instructions. Examples of instructions which the CPU is able to execute:

1. adding the contents of R0 and R1 and storing the result in R6
2. copying the contents of R3 into R0
3. doing a logical XOR of the contents of R3 with the contents of R4 and storing the result in R3
4. moving the number 42 into R5

3.2 CPU Architecture

This section will explore some CPU architectures and compare them to the architecture of the Cortex-M0.

The Cortex-M0 makes use of a Von Neumann architecture. This means that there is a single bus which connects all of the parts (such as CPU, RAM, flash) inside the microcontroller. The implication of this is that the CPU cannot fetch an instruction from flash at the same time as it moves data in or out of RAM. This limitation allows for a much simpler architecture, but at the expense of performance.

Other microcontrollers (even others in the Cortex-M series like the Cortex-M3) follow a Harvard architecture, meaning that there are separate buses used for fetching instructions and

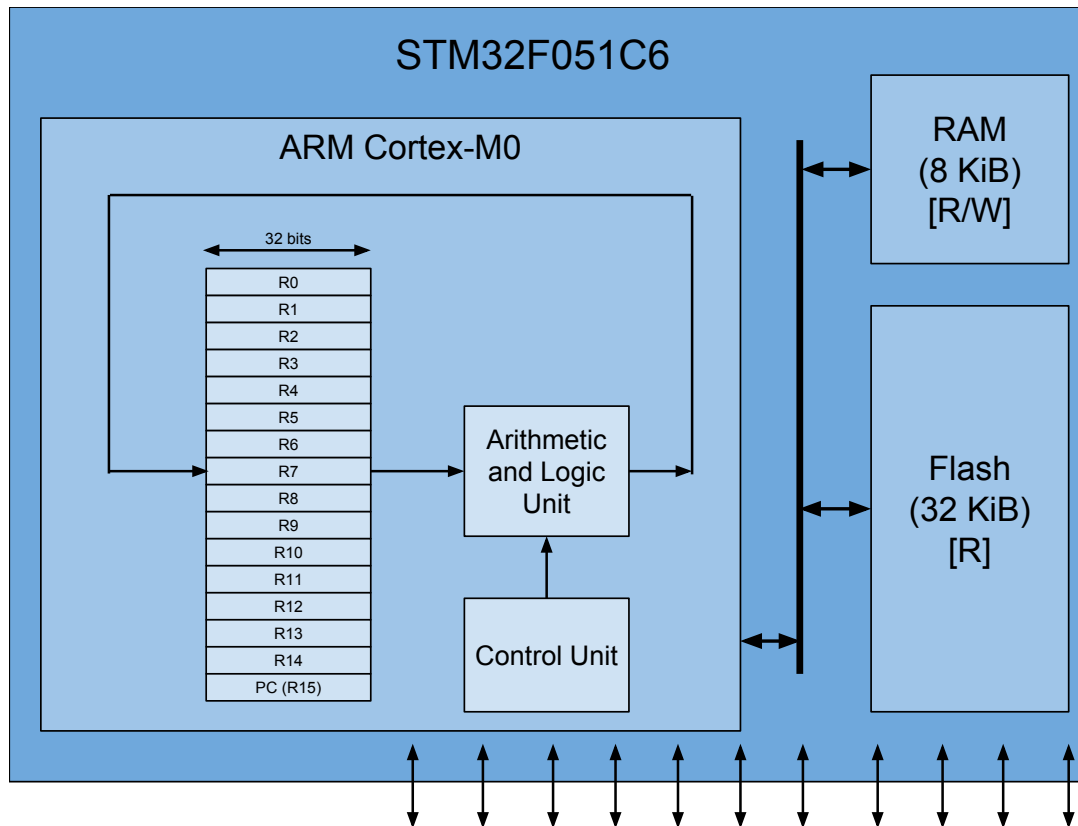


Figure 3.1: A view of the internals of the STM32F051 with the ARM Cortex-M0 expanded

moving data around. This allows faster execution as instructions can be fetched at the same time as data is loaded or stored. However, it necessitates greater complexity and more transistors.

It's been said that the ARM Cortex-M0 is a 32-bit processor. For comparison, the processor which we used in this course previously (MC9S08GT16A) was an 8-bit processor. Your personal computer probably has a 64-bit CPU. 16-bit CPUs are also quite common. So what exactly does it mean when we say that the processor is 32-bits? Essentially, the number of bits which a processor is said to be refers to the size of the data bus. In other words: the amount of data which the processor is able to move around internally or perform arithmetic and logic operations on. Hence, with a 32-bit processor, we can move 32 bits of data from one spot in memory to another in just once instruction. If you had a 8-bit processor, it would cost 4 instructions to move 32 bits of data around.

3.3 Program Counter

The Program Counter is a special register in the CPU, specifically: R15. It's called "special" because it has a specific, fixed purpose and cannot be used as a general purpose register like the other registers can. Its purpose is keeping track of where we are in the execution of a program. All instructions which need to be executed are laid out sequentially in flash, each instruction occupying a halfword of memory. Hence, each instruction has a defined address. The PC points to (ie: hold the address of) the instruction which is about to be fetched from flash and executed.

Typically, the value of the PC is simply incremented by 2 in order to cause it to point to the next instruction in memory. However, it's possible to alter the flow of execution of a program by issuing a *branch* instruction which will cause the PC to be incremented or decremented by a different amount.

3.3.1 Three stage pipeline

There is a bit more complexity to the program counter than initially apparent. It's worth understanding this extra intricacy as it affects how other instructions which depend on the program counter work. The ARM Cortex-M0 implements a three stage pipeline. This means that an instruction is broken up into three parts, and executed over the course of three clock cycles. The parts are:

- **fetch:** the instruction which the program counter points to is pulled into the CPU.
- **decode:** the CPU control unit "looks" at the 16 bits which represent the instruction, and figures out what action it must take.
- **execute:** the CPU runs the instruction, causing data to be modified.

The fact that the CPU is pipelined means that different instructions can be going through different phases *at the same time*. In other words, one instruction can be being fetched while another is being decoded while another is being executed. As an example, assume we have three instructions which we want to execute, instruction A, instruction B and instruction C. The three instructions being run through the pipeline is shown graphically in [Figure 3.2](#). It's critical to note how the program counter is always pointing to the instruction being *fetched*. This makes sense as the job of the program counter after all is to facilitate keeping track of which instruction must be fetched. For this reason, when an instruction is being executed, the PC is actually pointing to two instructions (four bytes) further ahead in memory, and *not* at the address of the instruction in execution. Hence, when an instruction in execution uses the PC, the value which will be used is the address of the instruction plus four.

3.4 Reset Vector

Initial address of PC loaded from 0x0800 0004.

PC	Cycle 1	Cycle 2	Cycle 3	Cycle 4	Cycle 5
Instruc. A	Fetch	Decode	Execute		
Instruc. B		Fetch	Decode	Execute	
Instruc. C			Fetch	Decode	Execute

	Cycle 1	Cycle 2	Cycle 3	Cycle 4	Cycle 5
Instruc. A	Fetch	Decode	Execute		
PC Instruc. B		Fetch	Decode	Execute	
Instruc. C			Fetch	Decode	Execute

	Cycle 1	Cycle 2	Cycle 3	Cycle 4	Cycle 5
Instruc. A	Fetch	Decode	Execute		
Instruc. B		Fetch	Decode	Execute	
PC Instruc. C			Fetch	Decode	Execute

Figure 3.2: Showing three instructions being run through a three stage pipeline, as well as where the PC is pointing every cycle

4 Coding

4.1 Assembly

In order to get the CPU to do some of what we've discussed above, it needs to have code loaded onto it to run. We write code in a language called assembly. Assembly is a human-readable language. A program is made up of a sequence of instructions; each instruction gets executed by the CPU. It's quite easy to see what each instruction does by reading the program. The complete instruction set is located in the Programming Manual. You must be familiar with this document! Examples of instructions which carry out the tasks listed above are:

1. `ADDS R6, R0, R1`
2. `MOV R0, R3`
3. `EORS R3, R3, R4`
4. `MOVS R5, #42`

The CPU does not have the ability to understand our nice English words like *ADD* or *MOV*. The CPU only has the ability to understand binary data. Assembly code must be compiled to machine code. A machine code instruction is a binary string, 16 bits long consisting of the operation code (opcode) and the data which it must operate on (operand). For example, assume that we wanted to ascertain the machine code representation of the instruction `ADDS R6, R0, R1`. An extract from the ARMv6-M Architecture Reference Manual is shown in Figure 4.1 where *Rd* is the destination register and *Rm* and *Rn* are the source registers of the add. It can easily be seen that the instruction would compile to `0001100 001 000 110 = 0x1846`. The opcodes for each instruction are detailed in the ARMv6-M Architecture Reference Manual. All of the instructions in the program are 16 bits long and are stored sequentially after one another in flash memory.

4.1.1 Instruction Sets

An instruction set is the collection of all of the instructions which a processor can execute. The ARM Cortex-M0 uses the ARMv6-M architecture and this architecture supports the Thumb instruction set (as opposed to Thumb-2 or ARM). Thumb contains about XXX instructions,

Encoding T1 All versions of the Thumb instruction set.
`ADDS <Rd>, <Rn>, <Rm>`

15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
0	0	0	1	1	0	0	Rm			Rn			Rd		

Figure 4.1: An encoding of the `ADDS` instruction

each of which is 16 bits long.

Higher end ARM processors such as the Cortex-M3 or Cortex-M4 support the ARMv7-M architecture which allows multiple instruction sets to be supported by the processor. The ability to support multiple instruction sets requires *interworking*. Interworking is the ability to specify to the CPU which instruction set to use. While our ARM Cortex-M0 only supports the Thumb instruction set, there is no need for interworking, yet the capability has still been incorporated into the architecture to allow for compatibility to other processors. This means that although our processor only supports one instruction set (Thumb), we have to explicitly tell it that we are using that instruction set.

4.2 Linking

Once our assembly code has been written and compiled to machine code, the computer which loads the code onto the micro has to be told what addresses to place the code at. The code should be placed starting at the beginning of flash.

4.2.1 Executing Code

The PC always points to the instruction which is about to be fetched. Hence, when your micro boots up, before it has executed anything, the PC will point to the first instruction to be fetched/decoded/executed. By "point to" we mean that it holds the address of the instruction.

As each instruction in the ARM Cortex-M0 instruction set is 16 bits (aka: half a word) long, ARM have implemented a rule that all instructions must be half word aligned. In other words, the address of the instruction must be divisible by 2 bytes. Legal addresses for instructions are hence, 0x02, 0x04, 0x06, 0x08 ... etc. This means that the least significant bit (bit 0) of the PC register is unused in specifying the address of an instruction. Hence, it has been assigned another use. Specifically, to indicate the instruction set which is being executed.

5 Loading and Storing

Loading is the process of getting data from somewhere in the memory space into the CPU registers so that it can be used in processing. Storing is the process of getting data which is in the CPU registers into memory. Remember that seeing as flash is read-only memory, we cannot store data to flash address, but we can store to RAM.

The general format for a load is that a destination register, a register containing a base address, and an offset are supplied. An effective address is then calculated as the base address plus the offset. The contents of memory at the effective address are then copied from memory into the destination CPU register.

A store operation is very similar. Again, a register containing a base address and an offset are supplied, but this time it is a source register not a destination register which is supplied. Again, an effective address of base plus offset is calculated. The contents of the source register is copied into the effective address.

Note that most of the load/store operations which we will be doing are 32-bit (word) load or stores. This is because the CPU registers are 32 bits. So far we have only spoken of a single effective address. As you know, each address can only hold 8 bits. Hence, in order to load or store 32 bits, four sequential addresses are used. The effective address specifies the *lowest* in the sequence of the addresses. For example, if we wanted to store the contents of R0 in 0x20000000, the word would be placed into the address range 0x20000000, 0x20000001, 0x20000002 and 0x20000003. Remember that our processor uses little endian format, so the LSB is placed at 0x20000000 and the MSB at 0x20000003.

We will now explore some implementations of loading and storing.

5.1 Immediate Offset Loading

In this format, the base address is supplied in one of high CPU registers (R0 - R7), and the offset is supplied as an immediate number. The instruction format for loading data into a register is

LDR Rt, [Rn, #imm]

where Rt is the target register for the load, Rn contains the base address and #imm is the offset from the base address.

The way that this instruction works is that it calculates an *effective address* which is equal to the contents of the base address register plus whatever number is supplied as an immediate operand. There is, however, a slight complexity in how the offset is dealt with.

5.1.1 Offset restrictions

Remember that all instructions are limited to 16 bits. The format of the LDR instruction in machine code is shown in [Figure 5.1](#). We can see that after 5 bits of opcode and $2 \times 3 = 6$ bits of register specifications, we are only left with 5 bits of offset. Normally, these 5 bits would only allow us to provide an offset of $2^5 - 1 = 31$ bytes. This is not very much! In order to extend the range of the 5 offset bits, the actual offset used is equal to the 5 bit immediate number multiplied

Encoding T1 All versions of the Thumb instruction set.

LDR <Rt>, [<Rn>{, #<imm5>}]

15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
0	1	1	0	1	imm5					Rn		Rt			

```
t = UInt(Rt); n = UInt(Rn); imm32 = ZeroExtend(imm5:'00', 32);
```

Figure 5.1: Machine Code representation of LDR instruction. Source: ARMv6-M Architecture Reference Manual

by four. This multiplication by four is the same as appending two zeros to the end of the binary value, which you can see is being done in [Figure 5.1](#). This means that the amount which we are able to offset a base address by is now $(2^5 - 1) \times 4 = 124$, which is significantly more useful. However, seeing as we are multiplying to immediate number by four to get the actual offset, the implication is that all offsets *must* be a multiple of four. The compiler automatically takes care of dividing whatever offset we supply in our assembly instruction by four in order to get it to fit into the 5 bit immediate number, and the CPU then multiplies the immediate number by four to get the offset.

For example: if we wanted an offset of 12, the immediate number which would be placed in the instruction by the compiler would be 3.

5.2 Program Counter Relative Loading

There is another format of the LDR instruction which takes the Program Counter as a base register, and allows for an 8-bit immediate offset. If you wish to load data from flash into a CPU register, it makes sense to use the PC as a base register due to the fact that the PC is already initialised to be pointing to an address in flash. Specifically, it is pointing to the instruction which is being fetched (not executed - remember the three stage pipeline!). The format of the LDR instruction for PC relative loading can either be specified in the same way as the general LDR instruction, or it can have a label provided as an operand, as follows:

LDR Rt, [PC, #imm] LDR Rt, <label>

If one supplies a label as an operand, all that the compiler does is calculate the correct immediate offset value to insert, and compiles the instruction as if it were in the first format. It's important to note that these instructions are exactly equivalent: all that using a label does is cause the compiler to do the hard work of calculating the correct offset so you don't have to. It would really be a lot of hard work; every time you changed something in the structure of your program which caused instructions to be moved to different memory addresses (link writing a new line of code!) you'd potentially have to re-calculate your offsets. The ability to use labels is one of the most useful features of the compiler.

5.3 Register Offset Loading

So far all offsets have been supplied as immediate numbers to the load instructions. However, there is another format of the load instruction called a register-offset load. Here, the offset is contained in another register. This is useful as the offset can be set at run-time by modifying the contents of a register, rather than at compile time. In this case, the effective address is calculated as the contents of the base register (**Rn**) plus the contents of the offset register (**Rm**).

LDR Rt, [Rn, Rm]

5.4 Storing

The storing commands are so similar to the loading that they will barely be discussed. One difference is that there is no PC-relative store, as there would be no point trying to store data to read-only memory. The store instruction takes moves the contents of a source register, **Rt**, and places it at the effective memory address equal to the base address, **Rn**, plus an offset either supplied as a 5-bit immediate number, **#imm5**, or in an offset register, **Rm**.

STR Rt, [Rn, #imm5] STR Rt, [Rn, Rm]
