Module 1-C

Addressing modes and conditionals

Reading

This slide set:

- Textbook, Chapter 6, "Branch and Conditional Execution", pages 111 132.
 - Talking about this in this lecture.
- Textbook, Chapter 5, "Load and Store", pages 97 110.
 - Talking about this in the second half of the lecture.

More Reading

End of the week:

- Textbook, Chapter 7,"Structured Programming", pages 133 160.
 - We'll talk about this soon.
- Textbook, Chapter 8, "Subroutines", pages 161 202.
 - We'll talk about this soon.

- ARMv6-M Architecture Reference Manual (436 pages)
 - Get familiar with sections A5.2 (16-bit Thumb encoding) and A6.7 (Alphabetical list of ARMv6-M Thumb instructions)

Learning Outcome #1

"An ability to program a microcontroller to perform various tasks"



A. Architecture and Programming Model

B. Instruction Set Basics

How?

- C. Addressing modes and conditionals
- D. Practical Assembly Language Programming
- E. Stacks and Recursive Functions*

Objective

"Addressing modes and conditionals"

Typically

High-level lang. code

```
1 int array1[100];
2 int array2[100];
3 ...
4 for(n=0; n<100; n++)
5     array1[n] = array2[n] + 5;</pre>
```

Why?

Compiler Eg. gcc,vscc, llvm

Assembly Code

```
main:
            ldr r1, addr1
            ldr r^2, addr^2
11 add5:
            movs r0, #0
                            // n = 0
12 check: cmp r0, #100
                            // n < 100?
13
            bge done
                            // if not, done
14
            lsls r3, r0, #2 // r3 = r0 * 4
15
            ldr r4, [r2, r3] // array2[n]
16
            adds r4, #5
17
            str r4, [r1, r3] // array1[n]
18
            adds r0, #1
                            // n = n + 1
19
                            // again!
            b check
    done:
            b somewhere_else
    .balian 4
    addr1: .word array1
```

Right now

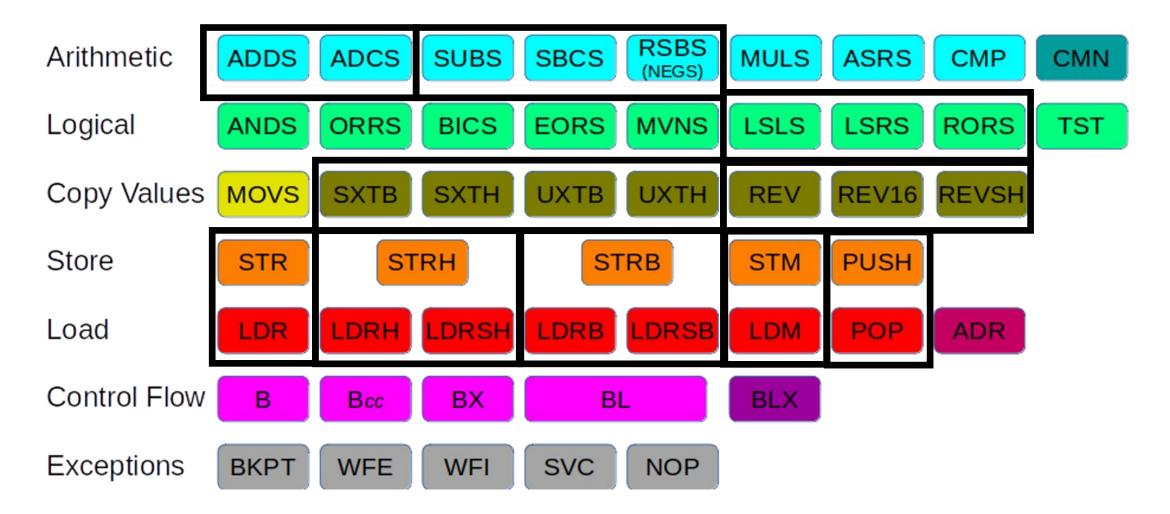


Still useful to think about high-level language constructs: loops, conditionals, etc...

addr2: .word array2



Rick's Handy Cortex-M0 Summary



The Branch

Control Flow B Bcc BX BL BLX

- Most instructions:
 - PC = PC + sizeOf(Insn) // typically PC = PC + 2
- Sometimes we don't want that:

```
LOOPS

1 int array1[100];
2 int array2[100];
3 ...
4 for(n=0; n<100; n++)
5 array1[n] = array2[n] + 5;
```

```
If conditions

if ( a < 5 ) {
    ...
} else {
    ...
}</pre>
```

```
Function Calls

a = func();
int func() {
    ...
}
```

Two classes of branches

- Unconditional
 - PC will **always** go to the target when executed

Control Flow

В

Bcc

BX

BL

BLX

- Conditional
 - PC will go to the target based on a condition

Computing the target: (1) PC-Relative Addressing Mode

- PC Relative Addressing
 - TargetAddress = PC + SignExt(imm) *

- * -- Not quite
 - All instructions at lest 2 bytes, always 2-byte aligned.
 - Bottom bits of instruction address always zero
- * -- One more Weirdness...
 - Offsets start 2 instructions past branch.

Decimal	Binary
2	10
4	10 <mark>0</mark>
1026	100000010

instr	instr	instr	B (0)	instr	instr	instr	instr	instr

How to use the branch

- When coding: Just use labels ©
- Thank your local assembler

```
.cpu cortex-m0
                          Example 0
    .thumb
    .syntax unified
    .fpu softvfp
    .text
    .global main
    main:
            movs r0, #5
                           0x2005
            movs r1, #1
                           0x2101
10
    L0:
            subs r0, #1
                           0x3801
11
            beq done
                           0xd000
12
                 L2
                           0xe001
13 L1:
                 LØ
                           0xe7fb
14 done:
            bkpt
                           0xbe00
            lsls r1, r1, #1 0x0049
15 L2:
16
                 L1
                           0xe74b
```

Conditional Branches

Needed to do anything useful

A6.7.10 B

Branch causes a branch to a target address.

Encoding T1 All versions of the Thumb instruction set. B<c> <|abel>

15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
1	1	0	1		CO	nd					im	m8	3		

if cond == '1110' then UNDEFINED; if cond == '1111' then SEE SVC; imm32 = SignExtend(imm8:'0', 32); if InITBlock() then UNPREDICTABLE;

Encoding T2 All versions of the Thumb instruction set.

B<c> <|abe|>

15 14 13 12 11 10 9 8 7 6 5 4 3 2 1 0 1 1 1 0 0 imm11

imm32 = SignExtend(imm11:'0', 32);
if InITBlock() && !LastInITBlock() then UNPREDICTABLE;

Assembler syntax

 $B{<c>}{<q>}$ <label>

where:

See Standard assembler syntax fields on page A6-98.

— Note —

- · Encoding T1 is conditional.
- · For encoding T1, <c> must not be AL or omitted.
- For ARMv6-M, for encoding T2, <c> must be AL or omitted.

<label>

The label of the instruction that is to be branched to. The assembler calculates the required value of the offset from the PC value of the B instruction to this label, then selects an encoding that sets imm32 to that offset.

Permitted offsets are even numbers in the range -256 to 254 for encoding T1 and -2048 to 2046 for encoding T2.

Operation

if ConditionPassed() then
 EncodingSpecificOperations();
 BranchWritePC(PC + imm32);

Exceptions

None.

BNE "PC+12" 1101 0001 00000110 0xD106

> B "PC-2" 11100 11111111111 0xE7FF

cond	Mnemonic extension	Meaning	Condition flags
0000	EQ	Equal	Z == 1
0001	NE	Not equal	Z == 0
0010	CS a	Carry set	C == 1
0011	CC p	Carry clear	C == 0
0100	MI	Minus, negative	N == 1
0101	PL	Plus, positive or zero	N == 0
0110	VS	Overflow	V == 1
0111	VC	No overflow	V == 0
1000	HI	Unsigned higher	C = 1 and $Z = 0$
1001	LS	Unsigned lower or same	C == 0 or Z == 1
1010	GE	Signed greater than or equal	N == V
1011	LT	Signed less than	N != V
1100	CT	Signed greater than	Z == 0 and $N == V$
1101	LE	Signed less than or equal	Z == 1 or N != V
1110c	None (AL) d	Always (unconditional)	Any

- a. HS (unsigned higher or same) is a synonym for CS.
- b. L0 (unsigned lower) is a synonym for CC.
- This value is never encoded in any ARMv6-M Thumb instruction.
- d. AL is an optional mnemonic extension for always.

A Basic Conversion

- First loop conversion: C do-while loop
 - Simpler in assembly than while: always do it at least once

C "do-while" loop

```
int r0 = 5;
int r1 = 0;
do {
    r1 += 1;
    r0 -= 1;
} while ( r0 != 0 );
```

```
Example 1
    .cpu cortex-m0
    .thumb
  .syntax unified
    .fpu softvfp
    .text
    .global main
    main:
            movs r0, #5
10
            movs r1, #0
11
    loop:
12
            adds r1, #1
13
            subs r0, #1
14
            bne loop
15
            bkpt
   done:
```

More "Exotic" conditions?

- Generally, execute an instruction that mods flags "S" suffix
 - Then do Branch
 - Mnemonics all assume a subtraction right before

Signed vs. Unsigned

- This only matters when you are working with very large integers. For instance: 0xfffffff
 - Is this a negative number?
 - It depends on whether we choose to interpret it as a signed or unsigned number.
 - If we want to treat it as a signed integer, it is -1
 - If we want to treat it as unsigned, it is 4,294,967,295

Unsigned Integers

- Recall:
 - n bits give rise to 2ⁿ combinations
 - let us call a string of 32 bits as " b_{31} b_{30} . . . b_3 b_2 b_1 b_0 "
- $f(b_{31} ... b_0) = b_{31} \times 2^{31} + ... + b_1 \times 2 + b_0 \times 2^0$
- Treat as normal binary number
 - e.g., 0 . . .011010101 = $1 \times 2^{7} + 1 \times 2^{6} + 0 \times 2^{5} + 1 \times 2^{4} + 0 \times 2^{3} + 1 \times 2^{2} + 0 \times 2^{1} + 1 \times 2^{0}$ = 128 + 64 + 16 + 4 + 1 = 213
- max f $(111...11) = 2^{32} 1 = 4,294,967,295$
- min f(000...00) = 0
- range $[0, 2^{32}-1] => \# \text{ values } (2^{32}-1) 0 + 1 = 2^{32}$

Signed Integers

- 2's complement
- $f(b_{31} b_{30} ... b_1 b_0) = -b_{31} x 2^{31} + ... + b_1 x 2 + b_0 x 2^0$
 - max $f(0111...11) = 2^{31} 1 = 2147483647$
 - min $f(100...00) = -2^{31} = -2147483648$ (asymmetric)
- range $[-2^{31}, 2^{31}-1] => \text{#values} (2^{31}-1 -2^{31} + 1) = 2^{32}$
- E.g., -6
- 000 . . . 0110 --> 111 . . 1001 + 1 --> 111 . . .1010

Signed or unsigned data types

- In C, we can explicitly declare variables that are never negative (always positive or zero) like this:
 - unsigned int big32 = 0xffffffff; // AKA uint32_t
 - unsigned short big16 = 0xffff; // AKA uint16_t
 - unsigned char big8 = 0xff; // AKA uint8_t
- We'll look at this again when we talk about converting C into assembly language.

Back to branches: unsigned values

- Subtraction operation is the same, regardless of sign
 - Beauty of 2's compliment!
- If we know that R2 and R5 are unsigned, we just interpret the flag differently
 - BLE: Signed, BLS: Unsigned

```
SUBS R6, R2, R5 // R2 – R5
BLS error // Branch if R2 is <= R5 (unsign)

1001 LS

Unsigned lower or same
```

C == 0 or Z == 1

Avoiding the subtraction

- Can use the CMP instruction
 - Sets flags based on the subtraction.
 - Code branches to error if R2 > R5
 - After CMP insn R2 is unchanged.
- No "S" on CMP, because it's only used to set flags always sets them.

CMP R2, R5 // R2 – R5: set flags, don't store result BGT error // Branch if R2 > R5 (sign)

1100 GT Signed greater than Z == 0 and N == V

Unsigned greater than or equal to

- No single instruction for unsigned gte
- Can do it with 2 branch instructions
 - Only need 1 CMP
 - Branch instructions don't edit flags can reuse results.

```
CMP R2, R5 // R2 – R5: set flags, don't store result

BHI error // Branch if R2 is > R5 (unsign)

BEQ error // Branch if R2 equal to R5
```

1000	HI	Unsigned higher	C == 1 and $Z == 0$
0000	EQ	Equal	Z == 1



(forgotten)

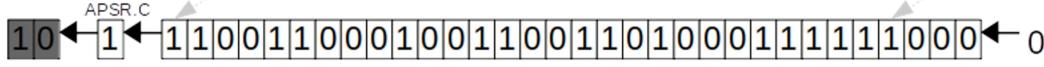
Logical Shift Left

- Can shift the values in 32-bit registers
 - << operator in C
 - Last bit shifted out into APSR.C
 - Zeros shifted in

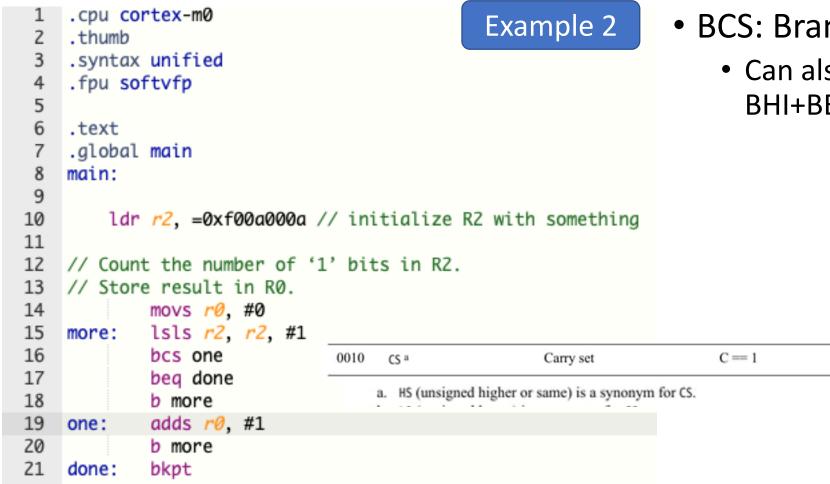
LSLS R6, R5, #3

Initial R5

0 New R6



Can use LSL to test bits



- Leftmost bit in the carry flag
- BCS: Branch if carry set
 - Can also use this instead of BHI+BEQ from 2 slides ago

Shift and Rotate



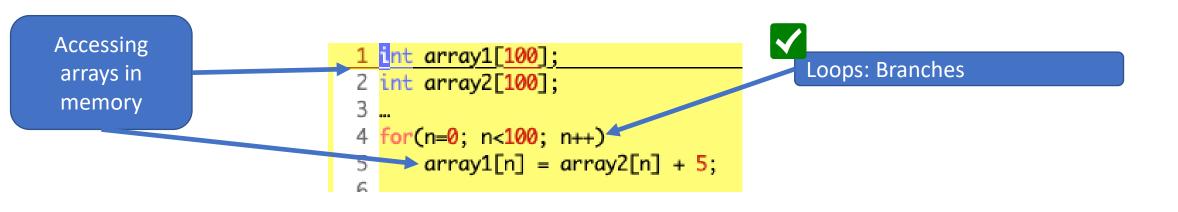
Bitwise logical instructions

```
ANDS R0, R1
 |0|1|1|1|0|0|1|1|0|0|0|1|0|0|1|1|0|0|1|1|0|0|0|1|1|1|1|1|1
                                                      Initial R0
                                                       AND
00001000000011111111111000001111100 (mask)
                                                        R1
New R0
ORRS R0, R1
                                                      Initial R0
  |1|1|1|0|0|1|1|0|0|0|1|0|0|1|1|0|0|1|1|0|1|0|0|0|1|1|1|1|1|1
                                                        OR
00000100000000011111111111100000011111000 (bits to set)
                                                        R1
1011100110001111111111111111111111111
                                                      New R0
BICS R0, R1
                                                      Initial R0
  |1|1|1|0|0|1|1|0|0|0|1|0|0|1|1|0|0|1|1|0|0|0|1|1|1|1|1|1|1
                                                      AND NOT
(bits to clear)
                                                        R1
101100011100000000000000001000000011
                                                       New R0
```

Bitwise logical instructions

```
EORS R0, R1
 |1|0|1|1|0|0|1|1|0|0|0|1|0|0|1|1|0|0|1|1|0|1|0|0|1|1|1|1|1|1|1 \\
                                                   Initial R0
                                                    AND
0000010000000011111111111000001111100 (bits to toggle)
                                                     R1
New R0
MVNS R0, R1 // Also known as NOT
|1|0|1|1|0|0|1|1|0|0|0|1|0|0|1|1|0|0|1|1|0|1|0|0|0|1|1|1|1|1|1
                                                   Initial R1
New R0
```

Next Piece: Need to better understand memory



Recall: Our addressing modes



• Immediate: small value contained in instruction



Register: src/dst register in instruction

- Offset: value is at an address that's a relative (constant) offset from a register
- <u>Indexed</u>: value is at address specified by register plus a second register



The LDR (Load Register) Instruction

- Let's say R0 holds the value 0x20000000.
- We want to load the four-byte value contained in addresses
- 0x20000000 0x20000003 into register R1.
- This is the instruction:
 - LDR R1, [R0]
- "Use the value in RO as an address. Read 4-bytes starting from that address, and put the result in R1."
- Caveat: R0 must be evenly divisible by 4.
 - If not, it will invoke a fault handler. You don't want that.



The STR (Store Register) Instruction

- Let's say R0 holds the value 0x20000000.
- We want to store the four-byte value contained in R1 to the addresses 0x20000000 0x20000003.
- This is the instruction:
 - STR R1, [R0]
- "Use the value in RO as an address. Write the value of R1 into the 4-bytes starting at that address."
- Caveat: R0 must be evenly divisible by 4.
 - If not, it will invoke a fault handler. You don't want that.

Simple Store/Load Example

```
Example 3
   cpu cortex-m0
   .thumb
   .syntax unified
    .fpu softvfp
 5
    .text
    .global main
   main:
           movs r0, #5

    Initializing R2 to an explicit

10
           movs r1, #0
11
           1dr r^2, =0x20000000
                                    address
12
    loop:
13
           str r0, [r2]
14
           adds r1, r0, #1
           ldr r1, [r2]
15
           subs r0, #1
16
17
           bne loop
           bkpt
   done:
```

Another simple example

```
.cpu cortex-m0
                                        Example 4
    .thumb
    .syntax unified
    .fpu softvfp
    .text
    .global main
    main:
 9
           movs r0, #5
           movs r1, #0

    This time we used a label,

11
           ldr r2, =storage
   loop:
                                  instead of an absolute address
13
           str r0, [r2]
14
           adds r1, r0, #1
           ldr r1, [r2]
           subs r0, #1
           bne loop
18
    done:
           bkpt
19
    .data
   .word 5 // some miscellaneous data here...
   .word 2 // ... just to avoid putting the storage
    .word 27 // we care about at 0x20000000
    storage: .word 0
```

Offset Addressing

- All these LDR/STRs are using <u>offset addressing</u> with an offset of zero.
- Generally:
 - LDR/STR Ry, [Rx, #<someConstant>]
 - i.e.: LDR R1, [R0, #4] "Take the value of R0, add 4 to it, use the sum as an address, go to that address, read four contiguous bytes, and put it in R1"
 - For LDR, can use offset from 0-124
 - Caveat: R0+offset must be evenly divisible by 4.

Common use of offset addressing

• Structs in C

```
struct Position {
    int x;
    int y;
    int z;
};
Position pos;
```

```
// Assume R5 holds address of pos
LDR R0, [R5, #0] // Load pos.x
LDR R1, [R5, #4] // Load pos.y
LDR R1, [R5, #8] // Load pos.z
```

Recall: Our addressing modes

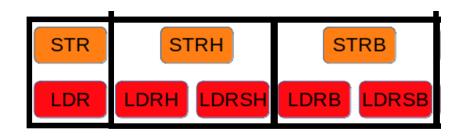
- **✓**
- Immediate: small value contained in instruction
- Register: src/dst register in instruction
- **✓**
- Offset: value is at an address that's a relative (constant) offset from a register
- Indexed: value is at address specified by register plus a second register

Indexed Addressing

- LDR/STR R1, [R0, R2]
- "Take the value of R0, add it to the value of R2, use the sum as an address, read four contiguous bytes, and place them in R1."
- Similar specification for STR.
- Caveat: R0+R2 must be evenly divisible by 4.

Different data sizes

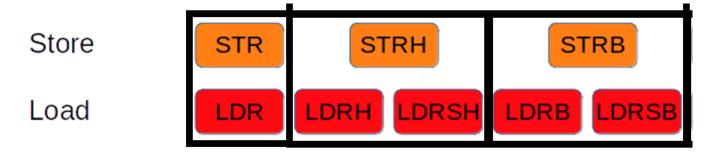
Store Load



- There are also
 - STRH: store a halfword (2 bytes) [index is a reg or an #imm multiple of 2 from 0 62]
 - LDRH: load an unsigned halfword (2 bytes) [index is the same as STRH]
 - LDRSH: load a signed halfword (sign-extend bit 15) [index must be a reg]
 - Caveat: Halfword load/stores must use an address that is evenly divided by
 2.
 - STRB: store a byte [index is a reg or #imm from 0 31]
 - LDRB: load an unsigned byte [index is the same as STRB]
 - LDRSB: load a signed byte (sign-extend bit 7) [index must be a reg]
 - A byte load or store has no alignment requirements.

Why 2 load flavors (signed and unsigned)

- Only one store type per size
- Because registers are always 32-bits. Operations always on 32-bits.



- When loading less than 32-bits from memory need to decide how to fill the register
 - i.e. If we are loading a byte (0xff) from memory LDRB (unsigned) puts 0x000000ff in register, LDRSB (signed) puts 0xffffffff.
- Storing always takes the lowest bits (16 for half, 8 for byte)

Load/Store Bytes

```
Example 5
    .cpu cortex-m0
  .thumb
  syntax unified
   .fpu softvfp
 5
    .text
    .global main
    main:
            movs r0, #0xfc
10
            ldr
                  r2, =0x20000000
11
            movs r3,#0
12
    loop:
13
            strb r0, [r2]
14
            // must use reg index
            ldrsb r0, [r2, r3]
15
16
            adds r0, #1
17
                  loop
            bne
18
            bkpt
   done:
```

- R0 starts out as 0xfc
- One byte stored in memory
- Loaded from memory as a signed byte, then incremented

Putting it all together

```
1 int array1[100];
2 int array2[100];
3 ...
4 for(n=0; n<100; n++)
5     array1[n] = array2[n] + 5;
6</pre>
```

```
For now assume:
"n" is in R0
R1 is the address for array1
R2 is the address for array2
```

```
.cpu cortex-m0
    .thumb
    .syntax unified
    .fpu softvfp
 5
    .data
   array1: .space 400
   array2: .space 400
    .text:
   add5:
           movs r0, #0 // n = 0
11
    check: cmp r0, #100 // n < 100?
           bge done // if not, done
13
           lsls r3, r0, #2 // r3 = r0 * 4
14
           ldr r4, [r2, r3] // array2[n]
15
           adds r4, #5 // add 5
16
17
           str r4, [r1, r3] // array1[n]
18
           adds r0, #1 // n = n + 1
19
           b check
                          // again!
   done:
           b somewhere_else
```

Last Piece: Getting values in memory

- We can load things relative to the PC (i.e. LDR R1, [PC, #12])
 - Cannot perform store instructions relative to a PC instructions are in readonly memory – range not far enough.

 Bonus: PC (rounded up)

+12 is automatically evenly divisible by 4

- So we can load something from a label LDR R1, label
 - Called a literal load
- Caveat: The item must be no further than 1020 bytes beyond the PC in the text segment.

Literal load example

```
6 .text
7 LDR R1, bigvalue
8 LDR R2, value2
9 ADDS R3, R1, R2
10 ...
11 B somewhere
12
13
14 .balign 4
15 bigvalue: .word 0xfedcba98
16 value2: .word 0xdecaf123
```

- Max offset: 1020 bytes
- Words must be in the text segment

Need to make sure the words are aligned.

A Label is an address

Address in memory

```
0x20000000 7 array1: .space 400
0x20000190 8 array2: .space 400
9
10
11 .text
0x08000400 12 addr1: .word array1
0x08000404 13 addr2: .word array2
```

What value is put into the memory word allocated for addr1?

A: 400

B: 100

C: 0x20000000

D: 0x08000400

E: Undefined, not initialized

- addr1/addr2 are locations in text memory where the <u>4-byte address</u> of array1/array2 are stored
 - They are *pointers*

Accessing an array from assembly

```
1 int array1[100];
2 int array2[100];
3 ...
4 for(n=0; n<100; n++)
5     array1[n] = array2[n] + 5;
6</pre>
```

```
7 .data
8 array1: .space 400
9 array2: .space 400
```

```
.text
           ldr r1, addr1
          1dr r^2, addr2
          movs r0, #0 // n = 0
   add5:
           cmp r0, #100 // n < 100?
   check:
           bge done // if not, done
11
12
          lsls r3, r0, #2 // r3 = r0 * 4
13
          ldr r4, [r2, r3] // array2[n]
14
          adds r4, #5 // add 5
15
           str r4, [r1, r3] // array1[n]
16
          adds r0, #1 // n = n + 1
17
           b check // again!
           b somewhere_else
   done:
   balian 4
20 addr1: .word array1
   addr2: .word array2
```

So common – assembler has a shortcut

 Just use LDR with a =symbol, and it will automatically build a literal pool at the end of the block of code...

```
7 .data
8 array1: .space 400
9 array2: .space 400
```

```
.text
           ldr r1, = array1
           ldr r2, =array2
           movs r0, #0 // n = 0
   add5:
   check:
           cmp r0, #100 // n < 100?
           bge done // if not, done
11
           lsls r3, r0, #2 // r3 = r0 * 4
13
           [r^{2}, r^{3}] // array2[n]
           adds r4, #5 // add 5
14
15
           str r4, [r1, r3] // array1[n]
           adds r0, #1 // n = n + 1
17
           b check
                          // again!
           b somewhere_else
   done:
```