

B2

A P P E N D I X

ARM and Thumb Instruction Encodings

Andrew Sloss,
ARM; *Dominic Symes,* ARM;

Chris Wright,
Ultimodule Inc.

B2.1	ARM Instruction Set Encodings	B-3
B2.2	Thumb Instruction Set Encodings	B-9
B2.3	Program Status Registers	B-11

This appendix gives tables for the instruction set encodings of the 32-bit ARM and 16-bit Thumb instruction sets. We also describe the fields of the processor status registers *cpsr* and *spsr*.

B2.1 ARM Instruction Set Encodings

Table B2.1 summarizes the bit encodings for the 32-bit ARM instruction set architecture ARMv6. This table is useful if you need to decode an ARM instruction by hand. We've expanded the table to aid quick manual decode. Any bitmaps not listed are either unpredictable or undefined for ARMv6.

To use Table B2.1 efficiently, follow this decoding procedure:

- Look at the leading hex digit of the instruction, bits 28 to 31. If this has a value $0xF$, then jump to the end of Table B2.1. Otherwise, the top hex digit represents a condition *cond*. Decode *cond* using Table B2.2.

TABLE B2.1 ARM instruction decode table. (Continued.)

Instruction classes (indexed by *op*)

31 30 29 28 27 26 25 24 23 22 21 20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1 0

STRH LDRH <i>pre</i>	0 0 0 1	U 0	W op	Rn	Rd	0 0 0 0 1 0 1 1	Rm
STRH LDRH <i>pre</i>	0 0 0 1	U 1	W op	Rn	Rd	immed [7:4]	immed [3:0]
LDRD STRD LDRSB LDRSH <i>pre</i>	0 0 0 1	U 0	W op	Rn	Rd	0 0 0 0 1 1 op 1	Rm
LDRD STRD LDRSB LDRSH <i>pre</i>	0 0 0 1	U 1	W op	Rn	Rd	immed [7:4]	immed [3:0]
AND EOR SUB RSB ADD ADC SBC RSC	0 0 1 0	op	S	Rn	Rd	rotate	immed
MSR cpsr, #imm MSR spsr, #imm	0 0 1 1	op 1 0	f s x c	1 1 1 1	1 1	rotate	immed
TST TEQ CMP CMN	0 0 1 1	op 1 1	Rn	0 0 0 0	0 0	rotate	immed
ORR BIC	0 0 1 1	op 0 S	Rn	Rd	rotate	immed	
MOV MVN	0 0 1 1	op 1 S	0 0 0 0	Rd	rotate	immed	
STR LDR STRB LDRB LDRB post	0 1 0 0	U op T op	Rn	Rd	rotate	immed12	
STR LDR STRB LDRB pre	0 1 0 1	U op W op	Rn	Rd	rotate	immed12	
STR LDR STRB LDRB post	0 1 1 0	U op T op	Rn	Rd	rotate	immed12	
{ S Q SH U UO UH ADD16	0 1 1 0	op	Rn	Rd	rotate	immed12	Rm
{ S Q SH U UO UH ADDSUBX	0 1 1 0	op	Rn	Rd	rotate	immed12	Rm
{ S Q SH U UO UH SUBADDX	0 1 1 0	op	Rn	Rd	rotate	immed12	Rm
{ S Q SH U UO UH SUB16	0 1 1 0	op	Rn	Rd	rotate	immed12	Rm
{ S Q SH U UO UH ADD8	0 1 1 0	op	Rn	Rd	rotate	immed12	Rm
{ S Q SH U UO UH SUB8	0 1 1 0	op	Rn	Rd	rotate	immed12	Rm
PKHBT PKHTB	0 1 1 0	op 0 0	Rn	Rd	rotate	immed12	Rm
{ S U SAT	0 1 1 0	op 1 1	immed5	Rd	rotate	immed12	Rm
{ S U SAT16	0 1 1 0	op 1 0	immed4	Rd	rotate	immed12	Rm
SEL	0 1 1 0	op 0 0	Rn	Rd	rotate	immed12	Rm
REV REV16 REVSH	0 1 1 0	op 1 1	1 1 1 1	Rd	rotate	immed12	Rm
{ S U XTAB16	0 1 1 0	op 0 0	Rn!=1111	Rd	rotate	immed12	Rm
{ S U XTAB16	0 1 1 0	op 1 0	1 1 1 1	Rd	rotate	immed12	Rm
{ S U XTAB	0 1 1 0	op 1 0	Rn!=1111	Rd	rotate	immed12	Rm
{ S U XTB	0 1 1 0	op 1 0	1 1 1 1	Rd	rotate	immed12	Rm
{ S U XTAH	0 1 1 0	op 1 1	Rn!=1111	Rd	rotate	immed12	Rm
{ S U XTH	0 1 1 0	op 1 1	1 1 1 1	Rd	rotate	immed12	Rm
STR LDR STRB LDRB pre	0 1 1 1	U op W op	Rn	Rd	rotate	immed12	Rm
SMLAD SMLSD	0 1 1 1	op 0 0	Rd	Rn!=1111	Rs	0 op X	1 Rm
SMUAD SMUSD	0 1 1 1	op 0 0	Rd	1 1 1 1	Rs	0 op X	1 Rm
SMLALD SMLSLD	0 1 1 1	op 1 0	RdHi	RdLo	Rs	0 op X	1 Rm

TABLE B2.1 ARM instruction decode table. (Continued.)

	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Instruction classes (indexed by <i>op</i>)																																
SMMLA SMMLS	<i>cond</i>	0	1	1	1	1	0	1	0	1	0	0	1	Rd	Rn != 1111	Rs	op	R	1	Rm												
SMMUL	<i>cond</i>	0	1	1	1	0	1	0	1	0	1	0	1	Rd	1	1	1	1	Rs	0	0	R	1	Rm								
USADA8	<i>cond</i>	0	1	1	1	1	0	0	0	0	0	0	0	Rd	Rn != 1111	Rs	0	0	0	1	Rm											
USAD8	<i>cond</i>	0	1	1	1	1	0	0	0	0	0	0	0	Rd	1	1	1	1	Rs	0	0	0	1	Rm								
Undefined and expected to stay so	<i>cond</i>	0	1	1	1	1	1	1	1	1	1	1	1		x																x	
STMDA LDMDA STMIA LDMIA	<i>cond</i>	1	0	0	0	op	^	W	op					Rn																		
STMDB LDMDB STMIB LDMIB	<i>cond</i>	1	0	0	1	op	^	W	op					Rn																		
B to instruction_address+8+4*offset	<i>cond</i>	1	0	1	0										signed 24-bit branch offset																	
BL to instruction_address+8+4*offset	<i>cond</i>	1	0	1	1										signed 24-bit branch offset																	
MCR MRR	<i>cond</i>	1	1	0	0	0	1	0	op					Rn	Rd	copro	op	1	Cm													
STC{L} LDC{L} <i>unindexed</i>	<i>cond</i>	1	1	0	0	1	L	0	op					Rn	Cd	copro			option													
STC{L} LDC{L} <i>post</i>	<i>cond</i>	1	1	0	0	U	L	1	op					Rn	Cd	copro			immed8													
STC{L} LDC{L} <i>pre</i>	<i>cond</i>	1	1	0	1	U	L	W	op					Rn	Cd	copro			immed8													
CDP	<i>cond</i>	1	1	1	0									Cn		copro	op2	0	Cm													
MCR MRC	<i>cond</i>	1	1	1	0	op1								Cn	Rd	copro	op2	1	Cm													
SWI	<i>cond</i>	1	1	1	1										immed24																	
CPS CPSIE CPSID	1	1	1	0	0	0	1	0	0	0	0	0	0	op	M	0	0	0	0	0	0	0	0	0	0	0	a	i	f	0	mode	
SETEND LE SETEND BE	1	1	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
PLD <i>pre</i>	1	1	1	0	1	0	1	U	1	0	1			Rn						immed12												
PLD <i>pre</i>	1	1	1	0	1	1	U	1	0	1				Rn						immed12												
RFEDA RFEIA RFEDB RFEIB	1	1	1	1	0	0	op	0	W	1				Rn	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	
SRSDA SRSDA SRSDB SRSDB	1	1	1	1	0	0	op	1	W	0				Rn	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	
BLX instruction+8+4*offset+2*a	1	1	1	1	0	0	op	1	W	0				1	1	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	
MRR2 MRR2	1	1	1	1	0	1	a								signed 24-bit branch offset																	
STC2{L} LDC2{L} <i>unindexed</i>	1	1	1	1	1	0	0	0	1	0	op			Rn	Rd	copro	op1		Cm													
STC2{L} LDC2{L} <i>post</i>	1	1	1	1	1	0	0	U	L	1	op			Rn	Cd	copro			option													
STC2{L} LDC2{L} <i>pre</i>	1	1	1	1	1	0	1	U	L	W	op			Rn	Cd	copro			immed8													
CDP2	1	1	1	1	1	1	0	op1						Cn	Cd	copro	op2	0	Cm													
MCR2 MRC2	1	1	1	1	1	1	0	op1						Cn	Cd	copro	op2	1	Cm													

TABLE B2.2 Decoding table for *cond*.

Binary	Hex	<i>cond</i>	Binary	Hex	<i>cond</i>
0000	0	EQ	1000	8	HI
0001	1	NE	1001	9	LS
0010	2	CS/HS	1010	A	GE
0011	3	CC/LO	1011	B	LT
0100	4	MI	1100	C	GT
0101	5	PL	1101	D	LE
0110	6	VS	1110	E	{AL}
0111	7	VC			

- Index through Table B2.1 using the second hex digit, bits 24 to 27 (shaded).
- Index using bit 4, then bit 7 or bit 23 of the instruction where these bits are shaded.
- Once you have located the correct table entry, look at the bits named *op*. Concatenate these to form a binary number that indexes the | separated instruction list on the left. For example if there are two *op* bits value 1 and 0, then the binary value 10 indicates instruction number 2 in the list (the third instruction).
- The instruction operands have the same name as in the instruction description of Appendix B1.

The table uses the following abbreviations:

- *L* is 1 if the L suffix applies for LDC and STC operations.
- *M* is 1 if CPS changes processor mode. *mode* is defined in Table B2.3.

TABLE B2.3 Decoding table for *mode*.

Binary	Hex	<i>mode</i>
10000	0x10	<i>user mode</i> (<i>_usr</i>)
10001	0x11	<i>FIQ mode</i> (<i>_fiq</i>)
10010	0x12	<i>IRQ mode</i> (<i>_irq</i>)
10011	0x13	<i>supervisor mode</i> (<i>_svc</i>)
10111	0x17	<i>abort mode</i> (<i>_abt</i>)
11011	0x1B	<i>undefined mode</i> (<i>_und</i>)
11111	0x1F	<i>system mode</i>

TABLE B2.4 Decoding table for *shift*, *shift_size*, and *Rs*.

<i>shift</i>	<i>shift_size</i>	<i>Rs</i>	Shift action
00	0 to 31	N/A	LSL # <i>shift_size</i>
00	N/A	<i>Rs</i>	LSL <i>Rs</i>
01	0	N/A	LSR #32
01	1 to 31	N/A	LSR # <i>shift_size</i>
01	N/A	<i>Rs</i>	LSR <i>Rs</i>
10	0	N/A	ASR #32
10	1 to 31	N/A	ASR # <i>shift_size</i>
10	N/A	<i>Rs</i>	ASR <i>Rs</i>
11	0	N/A	RRX
11	1 to 31	N/A	ROR # <i>shift_size</i>
11	N/A	<i>Rs</i>	ROR <i>Rs</i>
N/A	0 to 31	N/A	The <i>shift</i> value is implicit: For PKHBT it is 00. For PKHTB it is 10. For SAT it is 2* <i>sh</i> .

- *op1* and *op2* are the opcode extension fields in coprocessor instructions.
- *post* indicates a postindexed addressing mode such as $[Rn], Rm$ or $[Rn], \#immed$.
- *pre* indicates a preindexed addressing mode such as $[Rn, Rm]$ or $[Rn, \#immed]$.
- *register_list* is a bit field with bit *k* set if register *Rk* appears in the register list.
- *rot* is a byte rotate. The second operand is Rm ROR ($8 * rot$).
- *rotate* is a bit rotate. The second operand is $\#immed$ ROR ($2 * rotate$).
- *shift* and *sh* encode a shift type and direction. See Table B2.4.
- *U* is the up/down select for addressing modes. If $\bar{U} = 1$, then we add the offset to the base address, as in $[Rn], \#4$ or $[Rn, Rm]$. If $U = 0$ then we subtract the offset from the base address, as in $[Rn, \#-4]$ or $[Rn], -Rm$.
- *unindexed* indicates an addressing mode of the form $[Rn], \{option\}$.
- *R* is 1 if the R (round) instruction suffix is present.
- *T* is 1 if the T suffix is present on load and store instructions.
- *W* is 1 if ! (writeback) is specified in the instruction mnemonic.
- *X* is 1 if the X (exchange) instruction suffix is present.
- *x* and *y* are 0 for the B suffix, 1 for the T suffix.
- $\hat{}$ is 1 if the $\hat{}$ suffix is applied in LDM or STM instructions.

B2.2 Thumb Instruction Set Encodings

Table B2.5 summarizes the bit encodings for the 16-bit Thumb instruction set. This table is useful if you need to decode a Thumb instruction by hand. We've expanded the table to aid quick manual decode. The table contains instruction definitions up to architecture THUMBv3. Any bitmaps not listed are either unpredictable or undefined for THUMBv3.

To use the table efficiently, follow this decoding procedure:

- Index through the table using the first hex digit of the instruction, bits 12 to 15 (shaded).
- Index on any shaded bits from bits 0 to 11.
- Once you have located the correct table entry, look at the bits named *op*. Concatenate these to form a binary number that indexes the | separated instruction list on the left. For example, if there are two *op* bits value 1 and 0, then the binary value 10 indicates instruction number 2 in the list (the third instruction).

TABLE B2.5 Thumb instruction decode table.

Instruction classes (indexed by <i>op</i>)	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
LSL LSR	0	0	0	0	<i>op</i>	<i>immed5</i>			<i>Lm</i>	<i>Ld</i>						
ASR	0	0	0	1	0	<i>immed5</i>			<i>Lm</i>	<i>Ld</i>						
ADD SUB	0	0	0	1	1	0	<i>op</i>	<i>Lm</i>		<i>Ln</i>	<i>Ld</i>					
ADD SUB	0	0	0	1	1	1	<i>op</i>	<i>immed3</i>		<i>Ln</i>	<i>Ld</i>					
MOV CMP	0	0	1	0	<i>op</i>	<i>Ld/Ln</i>		<i>immed8</i>								
ADD SUB	0	0	1	1	<i>op</i>	<i>immed8</i>										
AND EOR LSL LSR	0	1	0	0	0	0	0	0	<i>op</i>	<i>Lm/Ls</i>		<i>Ld</i>				
ASR ADC SBC ROR	0	1	0	0	0	0	0	1	<i>op</i>	<i>Lm/Ls</i>		<i>Ld</i>				
TST NEG CMP CMN	0	1	0	0	0	0	1	0	<i>op</i>	<i>Lm</i>	<i>Ld/Ln</i>					
ORR MUL BIC MVN	0	1	0	0	0	0	1	1	<i>op</i>	<i>Lm</i>	<i>Ld</i>					
CPY <i>Ld, Lm</i>	0	1	0	0	0	1	1	0	0	0	<i>Lm</i>	<i>Ld</i>				
ADD MOV <i>Ld, Hm</i>	0	1	0	0	0	1	<i>op</i>	0	0	1	<i>Hm & 7</i>		<i>Ld</i>			
ADD MOV <i>Hd, Lm</i>	0	1	0	0	0	1	<i>op</i>	0	1	0	<i>Lm</i>		<i>Hd & 7</i>			
ADD MOV <i>Hd, Hm</i>	0	1	0	0	0	1	<i>op</i>	0	1	1	<i>Hm & 7</i>		<i>Hd & 7</i>			
CMP	0	1	0	0	0	1	0	1	0	1	<i>Hm & 7</i>		<i>Ln</i>			
CMP	0	1	0	0	0	1	0	1	1	0	<i>Lm</i>		<i>Hn & 7</i>			
CMP	0	1	0	0	0	1	0	1	1	1	<i>Hm & 7</i>		<i>Hn & 7</i>			
BX BLX	0	1	0	0	0	1	1	1	<i>op</i>	<i>Rm</i>			0	0	0	
LDR <i>Ld, [pc, #immed*4]</i>	0	1	0	0	1	<i>Ld</i>		<i>immed8</i>								
STR STRH STRB LDRSB <i>pre</i>	0	1	0	1	0	<i>op</i>	<i>Lm</i>		<i>Ln</i>	<i>Ld</i>						
LDR LDRH LDRB LDRSH <i>pre</i>	0	1	0	1	1	<i>op</i>	<i>Lm</i>		<i>Ln</i>	<i>Ld</i>						
STR LDR <i>Ld, [Ln, #immed*4]</i>	0	1	1	0	<i>op</i>	<i>immed5</i>			<i>Ln</i>	<i>Ld</i>						
STRB LDRB <i>Ld, [Ln, #immed]</i>	0	1	1	1	<i>op</i>	<i>immed5</i>			<i>Ln</i>	<i>Ld</i>						
STRH LDRH <i>Ld, [Ln, #immed*2]</i>	1	0	0	0	<i>op</i>	<i>immed5</i>			<i>Ln</i>	<i>Ld</i>						

TABLE B2.5 Thumb instruction decode table. (Continued.)

Instruction classes (indexed by <i>op</i>)	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
STR LDR Ld, [sp, #immed*4]	1	0	0	1	<i>op</i>	<i>Ld</i>		<i>immed8</i>								
ADD Ld, pc, #immed*4 ADD Ld, sp, #immed*4	1	0	1	0	<i>op</i>	<i>Ld</i>		<i>immed8</i>								
ADD sp, #immed*4 SUB sp, #immed*4	1	0	1	1	0	0	0	0	<i>op</i>	<i>immed7</i>						
SXTH SXTB UXTH UXTB	1	0	1	1	0	0	1	0	<i>op</i>	<i>Lm</i>		<i>Ld</i>				
REV REV16 REVSH	1	0	1	1	1	0	1	0	<i>op</i>	<i>Lm</i>		<i>Ld</i>				
PUSH POP	1	0	1	1	<i>op</i>	1	0	<i>R</i>	<i>register_list</i>							
SETEND LE SETEND BE	1	0	1	1	0	1	1	0	0	1	0	1	<i>op</i>	0	0	0
CPSIE CPSID	1	0	1	1	0	1	1	0	0	1	1	<i>op</i>	0	<i>a</i>	<i>i</i>	<i>f</i>
BKPT immed8	1	0	1	1	1	1	1	0	<i>immed8</i>							
STMIA LDMIA Ln!, {register-list}	1	1	0	0	<i>op</i>	<i>Ln</i>		<i>register_list</i>								
B<cond> instruction_address+ 4+offset*2	1	1	0	1	<i>cond</i> < 1110			signed 8-bit offset								
Undefined and expected to remain so	1	1	0	1	1	1	1	0	<i>x</i>							
SWI immed8	1	1	0	1	1	1	1	1	<i>immed8</i>							
B instruction_address+4+offset*2	1	1	1	0	0	signed 11-bit offset										
BLX ((instruction+4+ (poff<<12)+offset*4) &~ 3) This must be preceded by a branch prefix instruction.	1	1	1	0	1	unsigned 10-bit offset									0	
This is the branch prefix instruction. It must be followed by a relative BL or BLX instruction.	1	1	1	1	0	signed 11-bit prefix offset <i>poff</i>										
BL instruction+4+ (poff<<12)+ offset*2 This must be preceded by a branch prefix instruction.	1	1	1	1	1	unsigned 11-bit offset										

- The instruction operands have the same name as in the instruction description of Appendix B1.

The table uses the following abbreviations:

- *register_list* is a bit field with bit *k* set if register *Rk* appears in the register list.
- *R* is 1 if *lr* is in the register list of PUSH or *pc* is in the register list of POP.

B2.3 Program Status Registers

Table B2.6 shows how to decode the 32-bit program status registers for ARMv6.

TABLE B2.6 *cpsr* and *spsr* decode table.

31 30 29 28 27 26 25 24 23 22 21 20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1 0														
<i>N</i>	<i>Z</i>	<i>C</i>	<i>V</i>	<i>Q</i>	<i>Res</i>	<i>J</i>	<i>Res</i>	<i>GE</i> [3:0]	<i>Res</i>	<i>E</i>	<i>A</i>	<i>I</i>	<i>FT</i>	<i>mode</i>
Field	Use													
<i>N</i>	Negative flag, records bit 31 of the result of flag-setting operations.													
<i>Z</i>	Zero flag, records if the result of a flag-setting operation is zero.													
<i>C</i>	Carry flag, records unsigned overflow for addition, not-borrow for subtraction, and is also used by the shifting circuit. See Table B1.3.													
<i>V</i>	Overflow flag, records signed overflows for flag-setting operations.													
<i>Q</i>	Saturation flag. Certain operations set this flag on saturation. See for example QADD in Appendix B1 (ARMv5E and above).													
<i>J</i>	<i>J</i> = 1 indicates Java execution (must have <i>T</i> = 0). Use the BXJ instruction to change this bit (ARMv5J and above).													
<i>Res</i>	These bits are reserved for future expansion. Software should preserve the values in these bits.													
<i>GE</i> [3:0]	The SIMD greater-or-equal flags. See SADD in Appendix B1 (ARMv6).													
<i>E</i>	Controls the data endianness. See SETEND in Appendix B1 (ARMv6).													
<i>A</i>	<i>A</i> = 1 disables imprecise data aborts (ARMv6).													
<i>I</i>	<i>I</i> = 1 disables IRQ interrupts.													
<i>F</i>	<i>F</i> = 1 disables FIQ interrupts.													
<i>T</i>	<i>T</i> = 1 indicates Thumb state. <i>T</i> = 0 indicates ARM state. Use the BX or BLX instructions to change this bit (ARMv4T and above).													
<i>mode</i>	The current processor mode. See Table B2.4.													