

COMET IIの扱う数字と文字

山本昌志*

2004年5月10日

1 前回の復習と本日の内容

1.1 前回の復習

前回の講義内容は以下の通りで、教科書の P.5~P.11 でした。課題ができていれば、十分です。

- 10進数と2進数、8進数、16進数の関係を示した。
- 基数の変換方法を示した。
- 負の数の表現方法(2の補数)を示した。

1.2 本日の内容

本日の講義内容は、以下の通りです。

- COMET IIの取り扱う数値
 - COMET IIのメモリー
 - 整数表現とその最大値と最小値
 - Pentiumのメモリー
- COMET IIの取り扱う文字
 - 文字の取り扱い方
 - 文字と数値の違い

*国立秋田工業高等専門学校 電気工学科

2 COMET IIの取り扱い数値

2.1 COMET IIのメモリー

以前も述べたように情報処理技術者試験では、COMET II という仮想的なコンピューター (ハードウェア) 上で動作する CASL II というアセンブラ言語が使われます。わざわざ、世の中で使われているアセンブラ言語を使用しないで、仮想的なものを使うのは、以下の理由によります。

- アセンブラ言語は、機械語と 1 対 1 に対応している。機械語は CPU 毎に異なる。そのため、世の中には CPU の種類だけ、アセンブラ言語がある。C 言語や FORTRAN が 1 種類しかなく、コンパイルすればどの CPU でも、動作するのと状況は全く異なる。
- 情報処理試験でも、アセンブラ言語を出題する必要がある。CPU の種類毎に出題するのは大変だし、実際に市販されている CPU は高機能すぎて、このような試験に似つかわしくない。
- そのため、機能が限定された COMET II というコンピューターを仮想的に作り、その上で実行されるアセンブラ言語が決められた。

要するに、単純な機能の試験を行いやすいアセンブラ言語が必要だったのです。そのようなわけで、初心者がアセンブラ言語を学習するにはうってつけです。

この COMET II のメインメモリー (主記憶装置) は、アドレスが 16 ビット、メモリー 16 ビットです。すなわち、図 1 のようになっています。

- アドレスが 16 ビット、メモリーが 16 ビットになっていることを確認しましょう。
- アドレス 1 個あたり、1 個のデータがあります。

この図を見ても分かるように、2 進数表示は紙面の面積が必要で紙の無駄です。もう少し紙を節約し、更に分かりやすくするために、通常は図 2 の 16 進数表示が使われます。

COMET II では、データは 16 ビット単位で扱われます。この 16 ビットの単位を 1 ワード (1 語) と呼びます。この 1 ワードは、教科書に書かれているように、便宜上、上位 8 ビットと下位 8 ビットに分けています。そうして、最下位のビットから番号がつけられています。最下位のビット番号が 0 で、最上位が 15 です。コンピューターの世界では、整数は 0 から数えることが多いので、それに慣れましょう。こうすると便利なことは、整数を表す場合、ビット番号が 2 進数の指数を表します。すなわち、第 0 ビットは 2^0 、第 7 ビットは 2^7 、第 15 ビットは 2^{15} の桁を表します。便利でしょう。

2.2 整数表現とその最大値と最小値

CASL II のプログラムではよほどのことが無い限り、整数しか扱いません。ひとつの整数は、メモリーのひとつのアドレスに格納されます。2 進数で表現した整数がそのまま、メモリー上のデータとなります。メモリーに格納された整数の様子を図 3 に示します。ただし、符号付整数か、符号無し整数かの判別は、プログラム次第です。

その整数の範囲については、教科書の通りです。電卓で自分で確認してみよう。

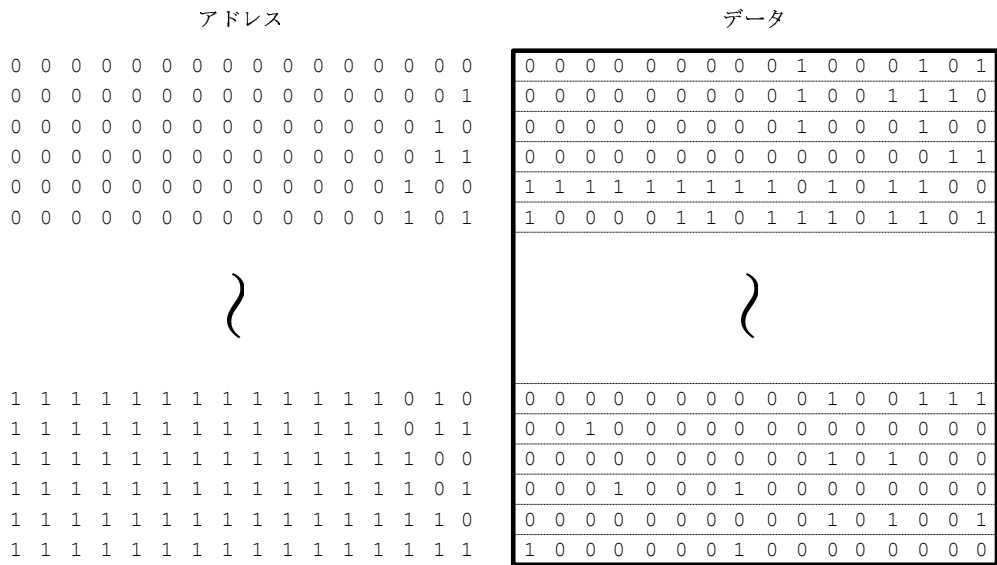


図 1: メモリーのモデル (2 進数表示)

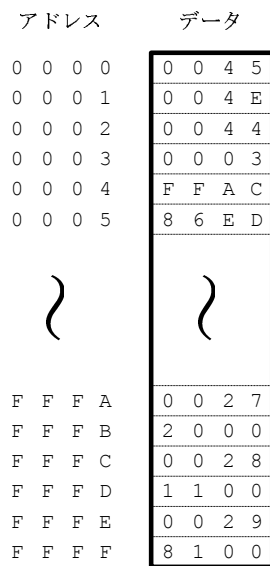


図 2: メモリーのモデル (16 進数表示)

アドレス	データ	符号付整数	符号無整数
9 F F F	? ? ? ?		
A 0 0 0	0 0 0 A	$(10)_{10}$	$(10)_{10}$
A 0 0 1	0 A 0 B	$(2571)_{10}$	$(2571)_{10}$
A 0 0 2	7 F F F	$(32767)_{10}$	$(32767)_{10}$
A 0 0 3	8 0 0 0	$(-32768)_{10}$	$(32768)_{10}$
A 0 0 4	A B C D	$(-21555)_{10}$	$(43981)_{10}$
A 0 0 5	F F F F	$(-1)_{10}$	$(65535)_{10}$
A 0 0 6	? ? ? ?		

図 3: 整数を格納しているメモリの状態 (16 進数表示)

2.3 Pentium のメモリ

COMET II のメモリは先ほど述べたとおりです。皆さんが使っている Intel の Pentium のメモリはどうなっているのでしょうか?。まず、アドレスバスは 32 ビットです。通常のコンピュータのデータは、8 ビット単位で扱われ、その単位を 1 Byte といいます。一つのアドレスに 1 Byte(8 bits) でデータが格納されます。Pentium のアドレスバスは 32 ビットなので、それが取り扱うことができるデータ数は、 2^{32} です。 2^{30} で G(ギガ)¹なので、4 GBytes のデータを扱うことができます。

また、データバスも 32 bits です。一度に 32 bits のデータを取り扱うことができるから、32bit CPU と呼ばれています。ここで、ひとつ疑問があります。1つのアドレスには 8 bits のデータしか格納されないのに、32 bits のデータを一度に取り扱うのは変です。どうやっているかということ、pentium では一度に 4 つのアドレスのデータを読み書きできるのです。

ひとつのアドレスに、32 bits を格納できるようにすれば、問題がなくなるように考える人も居るでしょう。そうすると他の問題が生じます。今まで、8 bits 単位でデータを取り扱ってきたので、過去のデータやソフトウェアとの互換性がなくなる可能性があります。

3 COMET II の取り扱う文字

3.1 文字の取り扱い方

一般に、情報を記号によって表現することをコード化(符号化)と呼びます。表現されたものをコード(符号)と呼びます。COMET II では、JIS X0201 ラテン文字・片仮名用 8 単位符号を用います(教科書 p.13 の表 2.3)。表を見て分かるように、各文字が 8 ビットの数字で表現されているだけです。たとえば、

¹ コンピューターの世界の補助単位は、 2^{10} で k(キロ)、 2^{20} で k(メガ)、 2^{30} で G(ギガ) となります。

文字	2進数	16進数
!	00100001	21
+	00101011	2B
3	00110011	33
K	01001011	4B
t	01110100	74

のようです。これは、JIS X0201 コードです。次の問いに答えなさい。

- このコードのビット数は、いくつでしょうか?。それは、何バイトでしょうか?。
- このコードで表現できる文字数は、いくつでしょうか?
- 漢字やカタカナ、ひらがながある日本語の文字はこのコードで表現できるか?

文字は8ビットで表していますが、COMET IIは16ビット単位でデータを取り扱います。一つのアドレスあたり2文字を格納することも考えられますが、そうすると1文字を取り扱うのが大変になります。実際、COMET IIでは、上位8ビットを0にして、下位8ビットを使うと決められています。それについては、教科書P.211の真ん中あたりの文字定数の部分に書かれています。

図4にCOMET IIのアドレス#A000から'Akita kosen'と文字が格納された場合のメモリーを示します。

- 図4で、上位8ビットが0になっていることを確認しよう。

アドレス	データ	文字
9 F F F	? ? ? ?	
A 0 0 0	0 0 4 1	'A'
A 0 0 1	0 0 6 B	'k'
A 0 0 2	0 0 6 9	'i'
A 0 0 3	0 0 7 4	't'
A 0 0 4	0 0 6 1	'a'
A 0 0 5	0 0 2 0	' '
A 0 0 6	0 0 6 B	'k'
A 0 0 7	0 0 6 F	'o'
A 0 0 8	0 0 7 3	's'
A 0 0 9	0 0 6 5	'e'
A 0 0 A	0 0 6 E	'n'
A 0 0 B	? ? ? ?	

図4: 文字の格納状態。データの'????'は不明を示す。

3.2 文字と数値の違い

教科書に書いてある通り (P.14)。