L. Mc Henry

Digital Computer Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts

Subject: SIGNED TERNARY ARITHMETIC

To: J. W. Forrester

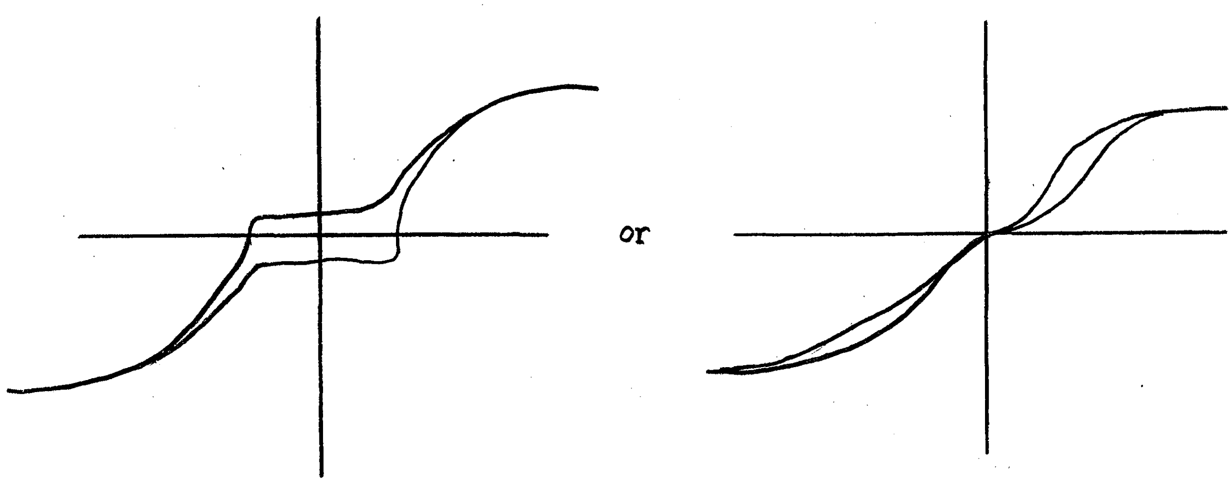
From: H. R. J. Grosch

Date: May 22, 1952

Introduction

Large digital calculating machines have been built to exploit the various advantages of pure decimal, binary-coded decimal, biquinary, and pure binary arithmetics. The availability of high-speed electromechanical and electronic elements having two stable states has been largely responsible for the swing away from the decimal number system.

Now we find ourselves increasingly interested in ferromagnetic and ferroelectric storage elements. There is a good possibility that these can be jockeyed into more than two stable, or better, "insensitive", states. Papian has shown me some B-H curves having the general form:



Such a material may be thought of as having three stable states. I have been interested for some years in the possibilities of a rather special sort of ternary arithmetic. This report is intended to give a resumé of its properties, and to advocate detailed consideration of its use in Whirlwind II.

Definitions

A real number is conventionally represented by a sequence of symbols.

$$A_4A_3A_2A_1A_0 \cdot A_{-1}A_{-2}A_{-3}$$

This is defined as having the value

$$A_4n_4 + A_3n_3 + A_2n_2 + A_1n_1 + A_0n_0 + A_{-1}n_{-1} + A_{-2}n_{-2} + A_{-3}n_{-3}$$

where the n's are the radices of the system. A convenient arithmetic requires that the A's and n's be rational numbers and that the representation possible therewith be both complete and unique over the real number field: that is, that every real number give rise to one and only one sequence of A's (not necessarily closed at the right, of course). The term "convenient" should be understood as referring to use on digital machines, which have a built-in prejudice in favor of the rational numbers. Unless at least one of the A's or n's is negative, it will be necessary to prefix the sequence with a minus sign to get over into the left half of the real number axis. No complementation convention can be adopted to get around this if a long repetition of arithmetic operations such as multiplication is required to yield answers consistent with the system.

The simplest forms of representation arise when all the A's are integers, and when the n's are monotonic non-decreasing. The radix point is placed to the right of an A (called A_0) whose corresponding base number n_0 is ± 1. Still more simplicity is gained when the n's are products of powers of a few integers. The abacus uses powers of two and five:

$$n_0 = 1$$

$n_1 = 5$	$n_{-1} = 2^{-1}$
$n_2 = 2.5$	$n_{-2} = 2^{-1} \cdot 5^{-1}$
$n_3 = 2.5^2$	$n_{-3} = 2^{-2} \cdot 5^{-1}$
$n_4 = 2^2.5^2$	$n_{-4} = 2^{-2} \cdot 5^{-2}$
etc.	etc.

This is also used in Bell and IBM relay calculators; it is called biquinary arithmetic. The A_i are either 0 or + 1 for odd i, or from 0 to + 4 for even i.

Finally, one may adopt a single positive integer radix r, so that $n_i = r^i$ and $A_i = 0, 1, \ldots, r-1$. The values $r = 2$ and $r = 10$ give rise to the conventional binary and decimal systems, and $r = 8$ is familiar as an output system for "base two" machines. The unit base $r = 1$ is a degenerate counting system which cannot represent fractional numbers between -1 and $+1$, and in which many other features of the positional notation, such as rounding, become meaningless.

In conventional ternary arithmetic $r = 3$, $n_i = 3^i$, $A_i = 0, 1,$ or 2. There are interesting possibilities, however, in using $A_i = -1, 0,$ and $+1$; this "signed ternary" system fulfills the conditions of completeness and uniqueness, and possesses several properties which can be very useful in a large computer and which are not present in conventional positive-symbol arithmetics.

To simplify the writing of signed ternary numbers, special symbols can be helpful:

$$\Lambda \quad (\text{"lam"}) \quad \equiv +1$$

$$0 \quad (\text{"oh"}) \quad \equiv 0$$

$$V \quad (\text{"vee"}) \quad \equiv -1$$

Thus $\Lambda VV0.0V = +27-9-3-1/9 = +14\ 8/9$ (decimal),

and $V0.0VV\Lambda VVV\Lambda 000 = -3-1/9-1/27 + 1/81 - 1/243 - 1/729 - 1/2187 + 1/6561$
$= -3.14159......$

In the signed ternary system defined above, there are exactly 3^p different combinations of p symbols $A_{p-1}......A_3A_2A_1A_0$. A demonstration that the system is unique will therefore prove completeness, or vice versa. By ordering all possible sequences of symbols, and by noticing that

$$\Lambda\Lambda\Lambda\Lambda\Lambda\ ..... = \sum_{i=0}^{p-1} 3^i = 1/2(3^p - 1),$$ either one of these proofs can be carried out.

Finally, it may prove amusing to recall the old mathematical recreation of finding the minimum number of (integer) weights which, when placed in either pan of a balance, will weigh any (integer) unknown up to a certain limit. There is a perfect correspondence with signed ternary arithmetic!

Arithmetic

For practical two-quantity addition the following tables are required:

1. If there is a carry ∧ from previous position,

			ADDEND	
		∧	0	V
AUGEND	∧	∧ 0	∧ V	0 ∧
	0	∧ V	0 ∧	0 0
	V	0 ∧	0 0	0 V

Right-hand symbol is digit of sum.
Left-hand symbol is carry to left position.

2. If the carry from previous position is 0,

	∧	0	V
∧	∧ V	0 ∧	0 0
0	0 ∧	0 0	0 V
V	0 0	0 V	V ∧

3. If the carry is V,

	∧	0	V
∧	0 ∧	0 0	0 V
0	0 0	0 V	V ∧
V	0 V	V ∧	V 0

A practical addition example looks like this:

V0.0VV∧VVV∧	(-3.1416.....)
∧∧.∧∧∧∧∧∧∧∧	(+4.4999.....)
0∧.∧0∧V00∧V	(+1.3583.....)

Subtraction will almost always be accomplished by complementing the subtrahend and adding. Complementation is performed by changing all

Λ's to V's and vice versa.

Multiplication is accomplished by shifting, complementation, and addition in a way very similar to multiplication in pure binary. No over-and-over addition is required. If the current digit of the multiplier is Λ , the multiplicand is added; if 0, ignored; if V, subtracted (complemented and then added); the partial product is then shifted one position to the right and the next digit of the multiplier examined in like manner.

```
      ΛV0Λ.      + 19
      V0ΛV       -  2.78....
     ------
      VΛ0V
     ΛV0Λ
    0000
   VΛ0V
   ---------
   VΛ00Λ.ΛV      - 52.78....
```

In case single-digit multiplication is required, the table is as follows:

Λ x Λ = V x V = Λ

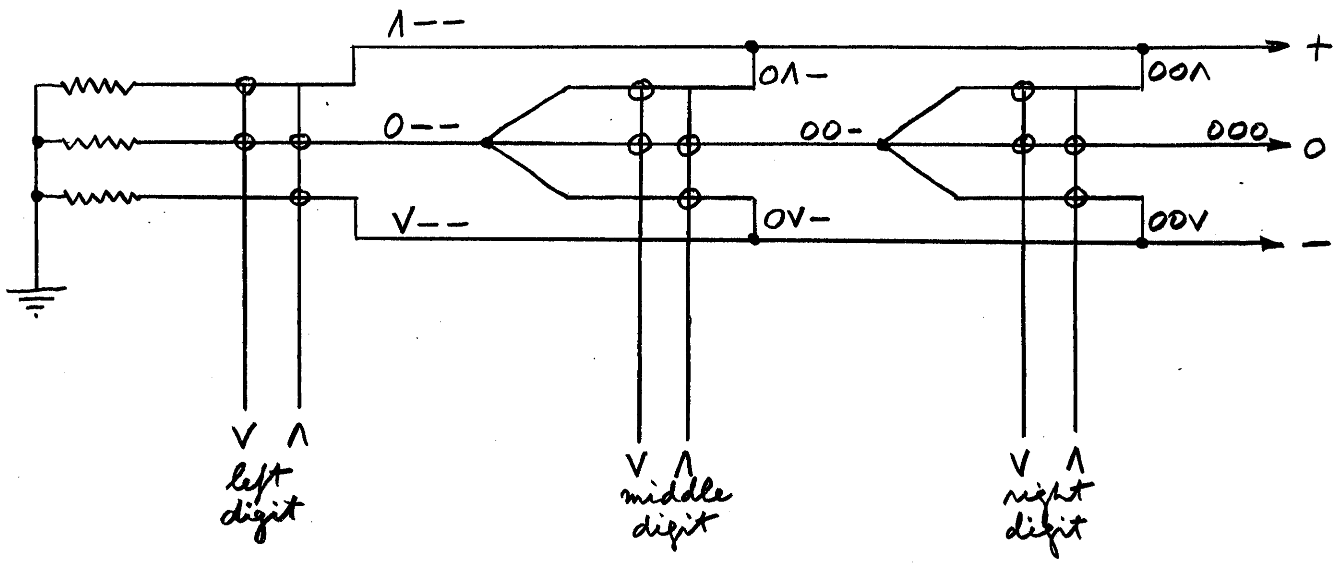
Λ x 0 = 0 x Λ = 0 x 0 = V x 0 = 0 x V = 0

Λ x V = V x Λ = V

The arithmetic of division is by no means so easy. In fact, it probably would not be too pessimistic to say that the introduction of signed symbols makes most arithmetic operations simpler at the expense of more complicated division. Since division occurs so infrequently, there is a temptation not to build it into a machine; this temptation is very strong in the present signed ternary case.

One elementary but wasteful method of division is to start at the extreme left quotient position (assuming as usual that the dividend and divisor have been suitably positioned). This exposes the fundamental difficulty of signed-symbol division clearly: is the divisor to be added or subtracted? Once the process has started correctly, so that the remainder tends toward zero, an alternation of adding and subtracting will be sufficient; the catch is to choose the first "direction" correctly.

In order to transform signed ternary to conventional binary or decimal output, some sort of sign determination is required. In a fast machine this determination should be parallel, and by determining successively the signs of the dividend and division we can set up the start of the division process. One diode network capable of doing the sign determination requires four diodes per digit, but this can probably be reduced to three cores if a long string of cores can be made to flip domino-fashion. The diode array for three digits is



In a two- , three- , or four-address machine two of these networks would be required.

Given the sign of the dividend and of the division, a simple logical combination embodies the rule: if the two signs agree, start by subtracting (complementing and adding) the divisor; if they disagree, start by adding. Shift the divisor one position to the right and "reverse" (add instead of subtract, or subtract instead of add) whenever the remainder changes sign. In detail, then,

```
                   V Λ
            VΛ.V Λ       =  VO.ΛV   (-2 7/9)
(+19)  ΛVOΛ. | VΛOOΛ.ΛV    (-52 7/9)
               ΛVOΛ        (+57)
               OOOΛΛ.ΛV    (+4 2/9)
                VΛOV       (-19)
               OVΛΛO.ΛV    (-14 7/9)
                 ΛVOΛ      (+6 1/3)
               OOVOΛ.VV    (-8 4/9)
                 ΛVOΛ      (+6 1/3)
               OOOVΛ.OV    (-2 1/9)
                 ΛVOΛ      (+6 1/3)
               OOOΛΛ.ΛV    (+4 2/9)
                   VΛOV    (-2 1/9)
               OOOΛV.OΛ    (+2 1/9)
                   VΛOV    (-2 1/9)
               OOOOO.OO
```

It will be evident that for randomly distributed n-digit quotients 2n additions and subtractions will be required, three times as many as the irreducible minimum 2/3 n. Undoubtedly much simpler processes can be evolved.

Given the arithmetic operations and a sign-determining network, and reserving three ternary digits to represent a decimal digit, transformation into and out of the conventional decimal system poses no special problems. After the sign of the number is determined and stored, the number is complemented to positive form if necessary. Three signed ternary digits cover the range -13 to +13 decimal, and only 0 to +9 decimal is required.

Advantages

It may be easily shown that three is the most economical radix of all single-integer-radix systems. If two elements can store a binary digit, and three a ternary digit, the ratio of equipment is 1.500 and the ratio of information stored is $\log_2 3$ or 1.585. Comparing quaternary and ternary on the same basis, one finds an equipment ratio of 1.333 and an information ratio of $\log_3 4$ or 1.262. This argument of economy applies to components such as tubes, diodes, and electromechanical gadgets. If core materials with three "insensitive" states are feasible, the gain of 1.585 in information stored compares with an equipment ratio not much over 1.000 for a large memory.

Is it really true that the outside world is best represented by dichotomies? Computing machines are basically logical, the argument runs, and the logic of the real universe is two-valued. My idea, on the contrary, is that computing machines operate on the basis of a logic concerned not with the real world, but rather with a special representation of the world by rational real numbers. And there are <u>three</u> relationships between pairs of rational numbers: X less than Y, X greater than Y, and X equal to Y. Putting it another way, zero is a member of the class of rational real numbers but is not a member of either the positive or negative subclass. Thus there is some real support for the theory that, while computer <u>arithmetic</u> may be done in any radix system, computer <u>logic</u> is basically <u>ternary</u>.

In real-time control applications the original conversion from continuous to digital representation is the place where the special representation by rational real numbers is introduced. There is of course no reason why the fundamental counters or other transformation devices cannot be ternary, with the same favorable equipment ratio as the central computer. As for the transmission of digital information, there is little doubt that more data can be transmitted over a given channel. Consider the old system of transmitting sixteen binary digits, frequency multiplexed, over a single phone line. If a single frequency is reserved for a fundamental continuous reference sine wave, the other fifteen frequencies can be phase-compared with this reference to give signed ternary digits: in phase represents Λ, 180° out of phase represents V, absence of signal represents 0. These fifteen ternary digits are equivalent to 23.77 bits, a gain of 48.6 per cent. It may be useful here to give the following tabulation:

1 binary digit (bit) = 0.631 stits = 0.301 dits

1 signed ternary digit (stit) = 1.585 bits = 0.477 dits

1 decimal digit (dit) = 3.322 bits = 2.096 stits

So much for what might be called the philosophical advantages of the signed ternary system. There are even more attractive arithmetical advantages. One, which has been mentioned earlier, is the elimination of a special "sign" representation. This makes it unnecessary to decide whether to shift the sign when shifting the rest of the number. In a sense, the left non-zero digit may be thought of as the sign, and this is convenient for the operator or customer. There is no longer any difference between a "negative" number (that is, a positive number with minus sign prefixed) and a complement: also there is no such thing as an end-around carry, and no difference between "nines" and "tens" (or rather, twos and threes) complements.

These advantages would accrue to any practicable signed radix system where zero is near the center of the ordered set of symbols A. Signed ternary, however, is the largest radix which will permit multiplication to be done without over-and-over addition.

Since the base of the system is an odd number, the representation of 1/2 is a repeating ternary - that is, 1/2 is not one of the class of rational real numbers representable by a finite computer. This has inconveniences, but it eliminates the problem of ambiguous rounding. In the decimal system, for instance, one rounds 6.49 to 6 and 6.51 to 7, but what does one do with 6.50? Adding 0.50 will often introduce a statistical bias and render error estimates based on random rounding errors unsatisfactorily optimistic.

In fact, signed ternary arithmetic completely eliminates the operation of rounding: rounding and dropping are identical operations. Thus ΛΛ.VV(+ 3 5/9) rounds to ΛΛ (+ 4) by dropping the last two digits; Λ0.ΛΛ (+ 3 4/9) rounds to Λ0 (+ 3) similarly. Any leftmost portion of a number is therefore the correctly rounded less-accurate representation of that number, a profound advantage in floating-point calculations.

Conclusion

This short introduction to the features of signed ternary arithmetic glosses over many points of interest. There has been no attempt to tie in three-valued logic, nor has there been any attempt to point out the possible advantages of ternary-coded decimal, nonary, or other bastard form

The next stage is obviously a discussion of hardware: core circuitry, ring-of-three counter elements (flip-o-flops?), magnetic tape storage, and so on.

H. R. J. Grosch

H. R. J. Grosch

HRJG:ajg