

A U T O N E T I C S

A DIVISION OF NORTH AMERICAN AVIATION, INC.
INDUSTRIAL PRODUCTS
3400 E. 70th Street, Long Beach 5, Calif.

RECOMP II TECHNICAL BULLETIN NO. 23

TITLE: The use of the RECOMP II floating point instructions for logical operations.

PURPOSE: To show how the floating add instruction of RECOMP II can be used to perform operations such as finding the integral portion of a floating point number or in changing a floating point number to a fixed point number at a given binary point.

EFFECTIVE DATE: February 2, 1962

CONTENTS: 1. Introduction

2. Technique for finding the integral portion of a floating point number.

3. Technique for converting a floating point to a fixed point number.

REFERENCES: None

INFORMATION TO: All Concerned

WRITTEN BY: L. Laubscher

Copyright 1962
Autonetics Industrial Products
A Div. of North American Aviation, Inc.
Long Beach, Calif.

1. Introduction

When performing a floating point addition, if neither of the numbers is zero, the binary point of the smaller number is increased to equal the binary point of the larger, and the smaller number is shifted right the corresponding number of places. Both numbers are then shifted right 1 bit position to avoid overflow. The numbers are added and the result is normalized reducing the exponent of the result respectively.

2. Technique for finding the integral portion of a floating point number.

If a non-zero floating point number, N , with a binary point of 38 is added to a number X , N and X will be shifted to a binary point of 39 and then added. Since the RECOMP II has a 39 bit register, the fractional portion of X will be shifted off. If N is then subtracted from the sum, the integral portion of X will be in the A and R registers as a normalized floating point number.

This technique may also be used to round a floating point number X to the nearest integer by adding .5 to X before finding the integral portion of X .

3. Technique for converting a floating point to a fixed point number.

It is possible to convert a floating point number X , with a binary point ≤ 38 to a fixed point number at a given binary point by the addition of a non-zero floating point number N . The first word of N need not be normalized and should contain only 1 bit. Since the result of an addition is always normalized, the sum of X and N will have a 1 in the high order bit. This may be extracted off or ignored as circumstances determine.

The first word of N will determine the binary point of the fixed point value of X as follows:

<u>First Word of N</u>	<u>Binary Point of X</u>
1@1	38
1@2	37
1@3	36
...	...
1@21	18
...	...
1@38	1

RECOMP II TECHNICAL BULLETIN NO. 23

The exponent of N (second word) determines the number of significant places to the right of the binary point as follows:

<u>Exponent of N</u>	<u>Significant Places</u>
0	38
1	37
...	...
36	2
37	1
38	0

If the exponent of N is greater than 38, the result, X+N, will have low order bits of its integral portion shifted off.

This technique is particularly useful for converting from floating to fixed point notation for address modification. Since the floating point number may be put at a b of 18 or 38 and the fractional portion (if any) can be eliminated simultaneously. If the result is used with a STA command, the high order bit may be ignored.