A STATISTICAL STUDY OF THE ACCURACY OF FLOATING POINT NUMBER SYSTEMS*

by

H, Kuki, The University of Chicago
Chicago, Illinois 60637

W. J. Cody, Argonne National Laboratory
Argonne, Illinois 60439

Abstract

This paper presents the statistical results of tests of the accuracy of certain arithmetic systems in evaluating sums, products and inner products. The arithmetic systems tested include 6-digit hexadecimal and 22-digit binary floating point number representations combined with the usual chop and round modes of arithmetic with various numbers of guard digits, and with a modified round mode with guard digits. In a certain sense the two number representations are shown to be statistically equivalent. Further, the usual round mode with guard digits is shown to be statistically superior in accuracy to the usual chop mode in all cases save one. The modified round mode is found to be superior to the chop mode in all cases.

*Work performed in part under the auspices of the United States Atomic Energy Commission.

A STATISTICAL STUDY OF THE ACCURACY OF FLOATING POINT NUMBER SYSTEMS*

by

H. Kuki, The University of Chicago
Chicago, Illinois 60637

W. J. Cody, Argonne National Laboratory
Argonne, Illinois 60439

*Work performed in part under the auspices of the United States Atomic Energy Commission.

Introduction

There are two major considerations in specifying computer floating point arithmetic systems. The first is the machine representation of floating point numbers, and the second is the design and implementation of the algorithms for the arithmetic operations. While these two considerations are not completely independent when it comes to details of design, there is a certain independence in gross characteristics. This paper presents the results of statistical tests of the accuracy of various *arithmetic systems*, i.e., combinations of number representation and arithmetic modes. Among the questions posed for these tests were the following:

1) is six hexadecimal digit precision equivalent to twenty-two binary digit precision?

2) how effective is rounding as opposed to chopping (or truncation) in preserving accuracy?

3) does the presence of guard digits improve accuracy? If so, by how much?