Fixed Point Library Documentation

Michael J. Malek

August 28, 2017

Contents

1 Introduction

This library implements basic functions necessary for 32-bit fixed-point operations. This document explains the purposes of all functions and their implementations. Supporting files including "parameters.h" and "fixedpoint.h" are discussed as well.
Model for the framework was drawn from the following source:
"code.google.com/archive/p/libfixmath/source/default/source"
OVERALL FEATURES:

i.

ii. Provides

2 Parameters

3 Basic Arithmetic

3.1 fp32_add

Addition is the same as for regular integers.

3.2 Subtraction:

Subtraction is the

3.3 Multiplication

There are three primary issues which can arise in our multiplication:

i. Overflow

ii. Underflow

iii. Loss of Precision

The solution implemented deals with all three of these problems at once, and hinges upon the
Proposition. Multiplication with double the precision guarantees an exact answer
Proof.

3.4 Division

4 Basic Conversions

5 Saturating Arithmetic

6 Min/Max

7 Basic Exponential Functions

8 Basic Trigonometric Functions