

BF20 Floating Point Format Specification

1. Overview

BF20 (Brain Float 20-bit) is a proposed floating point format designed to strike a balance between the compactness of BF16 and the precision of FP32. It uses 20 bits to represent a number, allowing for efficient memory use while preserving sufficient dynamic range and precision for many AI, graphics, signal processing, and embedded workloads.

2. Why BF20

BF20 is introduced to optimize performance and memory efficiency. It maintains an 8-bit exponent like FP32 for compatibility, while using an 11-bit mantissa for more precision than BF16. This balance provides:

- Improved accuracy over BF16
- Compatibility with FP32 exponent range
- Compact data packing (3 per 64-bit)
- Better memory bandwidth and power usage
- Applicability to AI, graphics, DSP, and embedded systems

3. Bit Layout

The 20-bit format includes:

- 1 sign bit
- 8-bit exponent (biased by 127)
- 11-bit mantissa with implicit leading 1

Bit layout: S | EEEEEEEE | MMMMMMMMMMMM

4. Format Encoding

BF20 format encodes values similarly to IEEE 754:

- Sign (1 bit): 0 = positive, 1 = negative
- Exponent (8 bits): Bias = 127
- Mantissa (11 bits): Encodes fraction with implicit leading 1 for normalized numbers

5. Use Cases

BF20 Floating Point Format Specification

- AI Inference: Reduced memory and power, better precision
- Graphics: Ideal for shader calculations with a wide dynamic range
- Signal Processing: Balanced format for accuracy and efficiency
- Embedded Systems: Low power, low memory footprint

6. Hardware Feasibility

Supporting 20-bit numbers is not problematic for modern hardware:

- Most registers are 32 or 64 bits; 20-bit values can be packed and operated efficiently
- SIMD support allows bulk operations on BF20 data
- Masking and shifts enable software fallbacks if needed
- BF20 can be supported with minor extensions to existing ALUs

7. Conclusion

BF20 offers a middle ground between BF16 and FP32, combining efficiency with precision. It is well-suited for the growing needs of edge computing, AI, graphics, and embedded workloads. With minor hardware changes, BF20 can bring substantial improvements in data throughput, memory savings, and energy efficiency.