Binary Arithmetic

Integer Addition

Summary: Binary addition involves adding two binary numbers. Carrying over occurs when the sum of two bits is 2 or greater.

Definitions:

- **Bit**: A binary digit, either 0 or 1.
- **Carry**: The value carried to the next column when two bits sum to 2 or more.

Example: Adding $1101 + 0011$:

$$\begin{array}{r} 1111 \text{ } \text{↰ Remainder} \\ 1101 \\ +0011 \\ \hline 10000 \end{array}$$

Tricky Things to Remember:

- Don't forget to carry over when the sum of two bits is 2 or greater.

Integer Subtraction

Summary: Binary subtraction involves borrowing from the next column when the subtrahend is greater than the minuend.

Definitions:

- **Borrow**: Taking value from the next column when the subtrahend is greater than the minuend.

Example: Subtracting $1101 - 0011$:

$$\begin{array}{r} 1101 \\ -0011 \\ \hline 1010 \end{array}$$

Tricky Things to Remember:

- Don't forget to borrow from the next column when needed.

Integer Division

Summary: Binary division involves dividing the dividend by the divisor to find the quotient and remainder.

Definitions:

- **Dividend**: The number to be divided.
- **Divisor**: The number by which division is to be performed.
- **Quotient**: The result of the division.
- **Remainder**: The remaining part after division.

Example: Dividing $1101 \div 11$:

$$\begin{array}{r} 100R1 \\ 11\overline{)1101} \\ \underline{1100} \\ 0001 \end{array}$$

Quotient: (100), Remainder: (10)

Tricky Things to Remember:

- Carefully follow the division steps to obtain the correct quotient and remainder.

Integer Multiplication

Summary: Binary multiplication involves adding partial products obtained by multiplying the multiplicand by each bit of the multiplier.

Definitions:

- **Multiplicand**: The number to be multiplied.
- **Multiplier**: The number by which multiplication is to be performed.

Example: Multiplying 1101×0011:

$$\begin{array}{r} 1101 \\ \times 0011 \\ \hline 1101 \\ +1101 \\ \hline 100111 \end{array}$$

Tricky Things to Remember:

- Ensure correct alignment of partial products to obtain the correct result.

Floating Point Addition

Summary: Floating-point addition in binary involves aligning the exponents, adding the mantissas, and normalizing the result.

Definitions:

- **Sign Bit**: Indicates the sign of the number.
- **Exponent**: Represents the power of 2.
- **Mantissa**: The fractional part of the number.

Example: Certainly! Here are the examples for floating-point arithmetic using half-precision, single-precision, and double-precision formats according to the IEEE 754 standard.

Half-Precision Example:

1. Convert both numbers to their IEEE 754 format.
2. Align the exponents by shifting the mantissa of the number with the smaller exponent.
3. Add mantissas.
4. Normalize the result by shifting the mantissa and adjusting the exponent accordingly.
5. Round the result to fit within the format's precision.

Tricky Things to Remember:

- Aligning exponents and normalizing the result is crucial.

Floating Point Subtraction

Summary: Floating-point subtraction is similar to addition but involves aligning the exponents, subtracting the mantissas, and normalizing the result.

Definitions:

- **Sign Bit**: Indicates the sign of the number.
- **Exponent**: Represents the power of 2.
- **Mantissa**: The fractional part of the number.

Example:

1. Convert both numbers to their IEEE 754 format.
2. Align the exponents by shifting the mantissa of the number with the smaller exponent.
3. Subtract the mantissas.
4. Normalize the result by shifting the mantissa and adjusting the exponent accordingly.
5. Round the result to fit within the format's precision.

Tricky Things to Remember:

- Aligning exponents and normalizing the result is crucial.

Floating Point Division

Summary: Floating-point division involves dividing the dividend's mantissa by the divisor's mantissa and subtracting the exponents, followed by normalizing.

Definitions:

- **Sign Bit**: Indicates the sign of the number.
- **Exponent**: Represents the power of 2.
- **Mantissa**: The fractional part of the number.

Example:

1. Convert both numbers to their IEEE 754 format.
2. Subtract the exponents and add the bias.
3. Divide the mantissas, including the implicit leading 1.
4. Normalize the result, ensuring that the mantissa fits within the format's precision.
5. Round if necessary, and apply the correct sign.

Tricky Things to Remember:

- Careful handling of the exponent and mantissa is required.

Floating Point Multiplication

Summary: Floating-point multiplication involves multiplying the mantissas and adding the exponents, followed by normalizing.

Definitions:

- **Sign Bit**: Indicates the sign of the number.
- **Exponent**: Represents the power of 2.
- **Mantissa**: The fractional part of the number.

Example:

1. Convert both numbers to their IEEE 754 format.
2. Add the exponents and subtract the bias (e.g., 127 for single precision).
3. Multiply the mantissas, including the implicit leading 1.
4. Normalize the result, ensuring that the mantissa fits within the format's precision.
5. Round if necessary, and apply the correct sign.

Tricky Things to Remember:

- Proper handling of the exponent and mantissa is required.
