# *CSE 1400 Applied Discrete Mathematics*
# *Naming and Number Systems*

*Department of Computer Sciences*

*College of Engineering*

*Florida Tech*

*Spring 2012*

Problems labeled

⚠ are important and must be mastered.

are harder and cover interesting advanced material that is not properly part of this course.

are tangential to the course but cover material that is useful elsewhere.

## *Problems on Naming Systems*

**Background**

Names are made from strings of characters chosen from an alphabet. Numbers are an important class of names, and the binary $\mathbb{B}$, octal $\mathbb{O}$, decimal $\mathbb{D}$, and hexadecimal $\mathbb{H}$ numerals are common alphabets for naming numbers.

Given an alphabet and a string length, an important goals is to be able compute the number of distinct names that can be constructed.

1. ⚠ (Know the terms used to describe languages.)

   (a) `True` or `False`: An alphabet is any finite set.
   Answer: This is `True`. The Greek letter $\Sigma$ is used by many authors to denote a language. Don't confuse it with the summation operator.

   (b) `True` or `False`: The elements of an alphabet are called characters.
   Answer: This is `True`. When the alphabet is for a natural language they are called letters; when the alphabet is for a numbers they are called numerals

   (c) `True` or `False`: A finite sequence of characters is called a string.
   Answer: This is `True`.

   (d) `True` or `False`: A string can be defined recursively by the rules:

   i. The empty string $\langle\rangle$ with no characters is a string.

   ii. If $c$ is character and $s$ is a string, then $cs$ is a string.
   Answer: This is `True`, and an elegant way to define strings recursively.

   (e) A language is a set of strings.
   Answer: This is `True`. The set of all strings over a alphabet $\Sigma$ is denoted $\Sigma^*$ and a language $\mathbb{L}$ is a subset of $\Sigma^*$.

   (f) A string belonging to a language is called a word.
   Answer: This is `True`.

2. ⚠ (Be able to count strings of given lengths.) Given the alphabets below, count the number of strings of a given length

   (a) Consider the alphabet of lower case English characters

   $$\mathbb{E} = \{a, b, c, \ldots, z\}$$

   i. How many strings of length 0 are there over $\mathbb{E}$?
   Answer: There is one, the empty string.

ii. How many strings of length 1 are there over $\mathbb{E}$?
Answer: There are 26, the characters in $\mathbb{E}$.

iii. How many strings of length 2 are there over $\mathbb{E}$?
Answer: There are $26^2$.

iv. How many strings of length 0 or 1 or 2 or $\ldots n-1$ are there over $\mathbb{E}$?
Answer: There are

$$\sum_{k=0}^{n-1} 26^k = \frac{26^n - 1}{25}$$

English strings with these lengths.

(b) Consider the alphabet of lower case Greek characters.

$$\mathbb{G} = \{\alpha, \beta, \gamma, \delta, \epsilon, \zeta, \eta, \theta, \iota, \kappa, \lambda, \mu, \nu, \xi, o, \pi, \rho, \sigma, \tau, \upsilon, \phi, \chi, \psi, \omega\}$$

i. How many strings of length 0 are there over $\mathbb{G}$?
Answer: There is one, the empty string.

ii. How many strings of length 1 are there over $\mathbb{G}$?
Answer: There are 24, the characters in $\mathbb{G}$.

iii. How many strings of length 2 are there over $\mathbb{G}$?
Answer: There are $24^2$.

iv. How many strings of length 0 or 1 or 2 or $\ldots n-1$ are there over $\mathbb{G}$?
Answer: There are

$$\sum_{k=0}^{n-1} 24^k = \frac{24^n - 1}{23}$$

Greek strings with these lengths.

(c) The DNA alphabet is the set

$$\mathbb{DNA} = \{\mathrm{A, C, G, T}\}$$

where A, C, G, and T are nucleic acids called "bases."

i. How many DNA strings of length $n$ are there?
Answer: There are $4^n$ strings.

ii. How many DNA strings are there of length 0 or 1 or 2 or $\ldots n-1$?
Answer: There are

$$\sum_{k=0}^{n-1} 4^k = \frac{4^n - 1}{3}$$

DNA strings with these lengths.

(d) The protein alphabet is the set

$$\mathbb{PRO} = \{\mathrm{A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y, B, Z, X}\}$$

In how many different strings of length five, six, or seven proteins are there?
Answer:

$$23^5 + 23^6 + 23^7$$

(e) Devices on the Internet are assigned an IP address. Today IPv4 remains the most common address, but over time it will be replaced by IPv6.

i. IPv4 identifies a device by an eight hexadecimal numerals. If any combination of numerals is allowable, how many devices can be named using IPv4?
Answer: There are $16^8 = 2^{32} \approx 4 \times 10^9$ different eight character hexadecimal strings.

ii. IPv6 identifies a device by an thirty-two hexadecimal numerals. If any combination of numerals is allowable, how many devices can be named using IPv6?
Answer: There are $16^{32} = 2^{128} \approx 256 \times 10^{36}$ different thirty-two character hexadecimal strings.

(f) A Media Access Control address (MAC address) is a unique identifier assigned to network interfaces for communications on the physical network segment. A MAC address has 12 hexadecimal digits. If any combination of numerals is allowable, how many MAC addresses are there?
Answer: There are $16^{12} = 2^{48} \approx 256 \times 10^{12}$ MAC addresses.

(g) Of the 128 ASCII characters, 94 are printable. How many printable strings of ASCII characters are there of the given length?

i. Length 0 strings.
Answer: There is $94^0$ string of length 0, the empty string.

ii. Length 1 strings.
Answer: There are $94^1$ strings of length 1.

iii. Length 2 strings.
Answer: There are $94^2$ strings of length 2.

iv. Length 3 strings.
Answer: There are $94^3$ strings of length 3.

v. Length $n-1$ strings.
Answer: There are $94^{n-1}$ strings of length $n-1$.

vi. Strings of length 0 or 1.
Answer: There are

$$94^0 + 94^1 = 95 = \frac{94^2 - 1}{94 - 1}$$

strings of length 0 or 1.

vii. Strings of length 0 or 1 or 2.
Answer: There are

$$94^0 + 94^1 + 94^2 = \frac{94^3 - 1}{94 - 1}$$

strings of length 0 or 1 or 2.

viii. Strings of length 0, 1, 2, . . . , $(n-1)$.
Answer: There are

$$94^0 + 94^1 + 94^2 + \cdots + 94^{n-1} = \frac{94^n - 1}{94 - 1}$$

strings of length 0, 1, 2, . . . , or $(n-1)$.

3. ⚠ The following natural numbers are written using their binary names. What are their decimal and hexadecimal names?

(a) 1111
Answer: The binary number 0000 1111 is decimal 15 and $F$ in hexadecimal.

(b) 1111 1111
Answer: The binary number 1111 1111 is decimal 255 and $FF$ in hexadecimal.

(c) 011 1100 1100.
Answer: The binary number 011 1100 1100 is decimal 972 and $3CC$ in hexadecimal.

(d) 110 1011 1010
Answer: The binary number 110 1011 1010 is decimal 1722 and $6BA$ in hexadecimal.

4. Explain how to covert from binary-to-octal and octal-to-binary.
Answer: Construct, or know, a binary-to-octal look-up table.

| Bin | Oct | Bin | Oct |
|---|---|---|---|
| 000 | 0 | 100 | 4 |
| 001 | 1 | 101 | 5 |
| 010 | 2 | 110 | 6 |
| 011 | 3 | 111 | 7 |

Given a binary string, group bits 3 at a time from right-to-left and look-up each octal numeral. For instance $(10\ 101\ 111\ 110)_2$ is $(2576)_8$. On the other hand, given an octal string, expand each numeral to 3 bits.

5. Explain how to covert from binary-to-hexadecimal and hexadecimal-to-binary.
Answer: Construct, or know, a binary-to-hexadecimal look-up table.

| Bin | Hex | Bin | Hex |
|---|---|---|---|
| 0000 | 0 | 1000 | 8 |
| 0001 | 1 | 1001 | 9 |
| 0010 | 2 | 1010 | $A$ |
| 0011 | 3 | 1011 | $B$ |
| 0100 | 4 | 1100 | $C$ |
| 0101 | 5 | 1101 | $D$ |
| 0110 | 6 | 1110 | $E$ |
| 0111 | 7 | 1111 | $F$ |

Given a binary string, group bits 4 at a time from right-to-left and look-up each hexadecimal numeral. For instance $(101\ 0111\ 1110)_2$ is $(57E)_{16}$. On the other hand, given a hexadecimal string, expand each numeral to 4 bits.

6. Expand the hexadecimal names below to their binary equivalent.

(a) $FE$
Answer: The hexadecimal number is 1111 1110 in binary.

(b) $ACED$
Answer: The hexadecimal number is 1010 1100 1110 1101 in binary.

(c) $DEADC0DE$
Answer: The hexadecimal number is 1101 1110 1010 1101 1100 0000 1101 1110 in binary.

(d) 577
Answer: The hexadecimal number is 0101 0111 0111 in binary.

(e) 3141
Answer: The hexadecimal number is 0011 0001 0100 0001 in binary.

7. How many strings of length $n$ are there over the following alphabets?

(a) The binary alphabet $\mathbb{B}$.
Answer: There are $2^n$ binary strings of length $n$. Notice the logarithmic-exponential relationship between the cardinality of the alphabet (the base), the string length $n$, and the number of THINGS named $b^n$.

(b) The decimal alphabet $\mathbb{D}$.
Answer: There are $10^n$ decimal strings of length $n$.

(c) The octal alphabet $\mathbb{O}$.
Answer: There are $8^n$ octal strings of length $n$.

(d) The hexadecimal alphabet $\mathbb{H}$.
Answer: There are $16^n$ hexadecimal strings of length $n$.

(e) The (lower case) English alphabet $\mathbb{E}$.
Answer: There are $26^n$ strings of length $n$ over the lower case letters in the English alphabet.

(f) The union of the decimal digits and the English alphabet $\mathbb{D} \cup \mathbb{E}$.
Answer: There are $36^n$ strings of length $n$ over the union of the lower case letters in the English alphabet and the numerals in the decimal alphabet.

(g) The (lower case) Greek alphabet $\mathbb{G}$.
Answer: There are $24^n$ strings of length $n$ over the lower case letters in the Greek alphabet.

8. How many strings of length 0 or 1 or 2 or $\ldots n-1$ are there over the following alphabets?

(a) The binary alphabet $\mathbb{B}$.
Answer: There are

$$\sum_{k=0}^{n-1} 2^k = 2^n - 1$$

binary strings with length 0 through $n-1$.

(b) The decimal alphabet $\mathbb{D}$.
Answer: There are

$$\sum_{k=0}^{n-1} 10^k = \frac{10^n - 1}{9}$$

decimal strings with length 0 through $n-1$.

(c) The octal alphabet $\mathbb{O}$. or $\ldots n-1$?
Answer: There are

$$\sum_{k=0}^{n-1} 8^k = \frac{8^n - 1}{7}$$

octal strings with length 0 through $n-1$.

(d) The hexadecimal alphabet $\mathbb{H}$.
Answer: There are

$$\sum_{k=0}^{n-1} 16^k = \frac{16^n - 1}{15}$$

hexadecimal strings with length 0 through $n-1$.

9. ⚠ How many characters are needed to name the given number of things?
   (a) 5 things in binary.
   Answer: At least $3 = \lfloor \lg 5 \rfloor + 1$ bits are needed. $5 = (101)_2$.
   (b) 15 things in binary.
   Answer: At least $4 = \lfloor \lg 15 \rfloor + 1$ bits are needed. $15 = (1111)_2$.
   (c) 16 things in binary.
   Answer: At least $5 = \lfloor \lg 16 \rfloor + 1$ bits are needed. $16 = (1\ 0000)_2$.
   (d) 50 things in decimal.
   Answer: At least $2 = \lfloor \log 50 \rfloor + 1$ digits are needed.
   (e) 99 things in decimal.
   Answer: At least $2 = \lfloor \log 99 \rfloor + 1$ digits are needed.
   (f) 100 things in decimal.
   Answer: At least $3 = \lfloor \log 100 \rfloor + 1$ digits are needed.
   (g) 50 things in hexadecimal.
   Answer: At least $2 = \lfloor \log_{16} 50 \rfloor + 1$ hexadecimal digits are needed: $(50)_{10} = (32)_{16} = 3 \cdot 16 + 2$.
   (h) 255 things in hexadecimal.
   Answer: At least $2 = \lfloor \log_{16} 255 \rfloor + 1$ hexadecimal digits are needed: $(255)_{10} = (FF)_{16} = 15 \cdot 16 + 15$.
   (i) 256 things in hexadecimal.
   Answer: At least $3 = \lfloor \log_{16} 256 \rfloor + 1$ hexadecimal digits are needed: $(256)_{10} = (100)_{16} = 1 \cdot 16^2 + 0 \cdot 16 + 0$.
   (j) $n$ things in binary.
   Answer: At least $\lfloor \lg n \rfloor + 1$ bits are needed.
   (k) $n$ things in decimal.
   Answer: At least $\lfloor \log n \rfloor + 1$ digits are needed.
   (l) $n$ things in hexadecimal.
   Answer: At least $\lfloor \log_{16} n \rfloor$ hexadecimal digits are needed.

10. Pretend you need to black-box test a program by entering each and every input combination to verify the output is correct for each input. Pretend the set of input combinations is a set of strings of length 1 through 10 drawn from an alphabet of 94 symbols. How many input combinations are there?
   Answer: There are

$$94 + 94^2 + 94^3 + \cdots + 94^{10} = 94\frac{94^{10} - 1}{94 - 1} = 54,440,667,446,151,152,650$$

different input combinations. Too many to ever test all of them.

## *Problems on Numbers*

**Background**

Numbers are fundamental to mathematics. The natural numbers are fundamental to discrete mathematics. When arithmetic computations occur the integers and rationals are important. In calculus and analysis, the real and complex numbers become the important sets.

1. Place the sets $\mathbb{B}$, $\mathbb{D}$, $\mathbb{H}$, $\mathbb{N}$, $\mathbb{P}$, $\mathbb{C}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{Z}$ in subset order.

   Answer:
$$\mathbb{B} \subset \mathbb{D} \subset \mathbb{H} \subset \mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$$

   Notice $\mathbb{P} \subset \mathbb{N}$, but none of the listed sets are subsets of $\mathbb{P}$. The subset relation is a "partial" order and not a total order, subsets can only be partially ordered.

2. Which proposition is `True` and which is `False`?

   (a) $5 \in \mathbb{D}$.
   Answer: `True`. The numeral 5 is a digit.

   (b) $-5 \in \mathbb{N}$.
   Answer: `False`. The natural numbers do not include negative integers.

   (c) $5 \in \mathbb{Q}$.
   Answer: `True`. The natural numbers are rational numbers too.

   (d) $\sqrt{5} \in \mathbb{Q}$.
   Answer: `False`. The square root of 5 is irrational.

   (e) $3/5 \in \mathbb{Q}$.
   Answer: `True`.

   (f) $3/5 \notin \mathbb{Z}$.
   Answer: `True`. $3/5$ is not an integer.

3. Equality ($=$) less than ($<$), and divides ($|$) are fundamental relations on the real numbers $\mathbb{R}$. Which of the following propositions are `True` and which are `False`.

   (a) For all real numbers $x$, $x = x$.
   Answer: This is `True`. Equality is reflexive.

   (b) For all real numbers $x$, $x < x$.
   Answer: This is `False`. Inequality is irreflexive.

   (c) For all natural numbers $n$, $n \mid n$.
   Answer: This is `True`. Divides is reflexive.

   (d) For all real numbers $x$ and $y$, if $(x = y)$, then $(y = x)$.
   Answer: This is `True`. Equality is symmetric.

   (e) For all real numbers $x$ and $y$, if $(x < y)$, then $(y > x)$.
   Answer: This is `False`. Inequality is not symmetric.

   (f) For all real numbers $x$ and $y$, if $((x < y)$ and $(y < x))$, then $(y = x)$.
   Answer: This is `True`. It is `True` by the form of the statement: The premise is `False`. Inequality is antisymmetric.

   (g) For all natural numbers $m$ and $n$, if $(m \mid n)$, then $(n \mid m)$.
   Answer: This is `False`. As a counterexample, 5 divides 35 but 35 does not divide 5.

   (h) For all real numbers $x$ and $y$, if $(x \leq y)$, then $(y \not\leq x)$.
   Answer: The $\leq$ relation is not strict, for instance a counterexample to the statement is $x = y = 1$. Then $(1 \leq 1)$ is `True`, but $(1 \not\leq 1) \equiv (1 > 1)$ is `False`.

   (i) For all real numbers $x$ and $y$, if $x < y$, then $y \not< x$.
   Answer: The $<$ relation is strict: If $(x < y)$ then $(y \not\leq x) \equiv (y > x)$.

(j) For all real numbers $x$, $y$ and $z$, if $((x = y)$ and $(y = z))$, then $(x = z)$.
Answer: This is `True`.

(k) For all real numbers $x$ and $y$, if $(x < y)$, then $(y > x)$.
Answer: This is `False`.

(l) For all natural numbers $m$ and $n$, if $(m \mid n)$, then $(n \mid m)$.
Answer: This is `False`. As a counterexample, 5 divides 35 but 35 does not divide 5.

4. What is the smallest natural number?
Answer: 0 is the smallest natural number.

5. What is the smallest positive integer?
Answer: 1 is the smallest positive integer.

6. What is the smallest positive rational number?
Answer: I suppose this would be classified as a "trick" question, but its meant to encourage thinking about numbers. There is no smallest positive rational number. Pretend there is a smallest positive rational $q$. But $q/2 < q$ is a smaller positive rational, contradicting that $q$ is the smallest.

7. Evaluate $\exp(x, y) = x^y$ at the given values.

(a) $x = -3$ and $y = -3$
Answer:
$$(-3)^{-3} = \frac{1}{(-3)^3} = \frac{1}{-27}$$

(b) $x = 11$ and $y = 1, 2, 3, 4$
Answer:
$$11^1 = 11,\ 11^2 = 121,\ 11^3 = 1331,\ 11^4 = 14641$$
Do you see a relationship with the values in Pascal's triangle?

(c) $x = 128$ and $y = 7/9$
Answer: $128^{7/9} = (2^7)^{7/9} = 2^{49/9} = 2^{5+4/9} = 32\sqrt[9]{16} \approx 43.5452$.

8. Evaluate $\log(x, b) = \log_b(x)$ at the given values.

(a) $x = 1024$ and $b = 2$.
Answer: $\log_2 1024 = 10$

(b) $x = \sqrt[3]{128}$ and $b = 8$.
Answer: $\log_8 \sqrt[3]{128} = \log_8 2^{7/3} = \log_8 (2^3)^{7/9} = 7/9$

9. Evaluate $\mod(a, n) = a \pmod{n}$ at the given values.

(a) $a = 45$ and $n = 7$?
Answer: Since $45 = 3 + 6 \cdot 7$, $45 \pmod 7 = 3$.

(b) $a = -45$ and $n = 7$?
Answer: Since $-45 = 4 - 6 \cdot 8$, $-45 \pmod 7 = 4$. Or, since $x + (-x) = 0$ and $45 \pmod 7 = 3$, $-43 \pmod 7$ must equal 4 since $3 + 4 = 0 \pmod 7$.

10. Describe the residue (equivalence) classes for the integers mod $n$ listed below.

(a) Integers mod 4.
Answer:

$$[0] = \{0, \pm 4, \pm 8, \pm 12, \ldots\} = \{0, 4, -4, 8, -8, 12, -12, \ldots\} = \{4k : k \in \mathbb{Z}\}$$
$$[1] = \{1, 1 + \pm 4, 1 + \pm 8, 1 + \pm 12, \ldots\} = \{1, 5, -3, 9, -7, 13, -11, \ldots\} = \{4k + 1 : k \in \mathbb{Z}\}$$
$$[2] = \{2, 2 + \pm 4, 2 + \pm 8, 2 + \pm 12, \ldots\} = \{2, 6, -2, 10, -6, 14, -10, \ldots\} = \{4k + 2 : k \in \mathbb{Z}\}$$
$$[3] = \{3, 3 + \pm 4, 3 + \pm 8, 3 + \pm 12, \ldots\} = \{3, 7, -1, 11, -5, 15, -9, \ldots\} = \{4k + 3 : k \in \mathbb{Z}\}$$

(b) Integers mod 5.
Answer:

$$[0] = \{5k : k \in \mathbb{Z}\}$$
$$[1] = \{5k + 1 : k \in \mathbb{Z}\}$$
$$[2] = \{5k + 2 : k \in \mathbb{Z}\}$$
$$[3] = \{5k + 3 : k \in \mathbb{Z}\}$$
$$[4] = \{5k + 4 : k \in \mathbb{Z}\}$$

11. Construct an addition table for the modular numbers.

(a) Integers mod 4.
Answer:

| + | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| 0 | 0 | 1 | 2 | 3 |
| 1 | 1 | 2 | 3 | 0 |
| 2 | 2 | 3 | 0 | 1 |
| 3 | 3 | 0 | 1 | 2 |

(b) Integers mod 5.
Answer:

| + | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| 0 | 0 | 1 | 2 | 3 | 4 |
| 1 | 1 | 2 | 3 | 4 | 0 |
| 2 | 2 | 3 | 4 | 0 | 1 |
| 3 | 3 | 4 | 0 | 1 | 2 |
| 4 | 4 | 0 | 1 | 2 | 3 |

12. Construct a multiplication table for the modular numbers.

(a) Integers mod 4.

Answer:

| $\times$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 1 | 2 | 3 |
| 2 | 0 | 2 | 0 | 2 |
| 3 | 0 | 3 | 2 | 1 |

(b) Integers mod 5.
Answer:

| $\times$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 1 | 2 | 3 | 4 |
| 2 | 0 | 2 | 4 | 1 | 3 |
| 3 | 0 | 3 | 1 | 4 | 2 |
| 4 | 0 | 4 | 3 | 2 | 1 |

13. The size of a computer's address space is governed by $n$, the number of bits in an address. The machine's memory addresses are the values in the set

$$\mathbb{M}_n = \{0, 1, 2, 3, \ldots, 2^n - 1\}$$

Assume the computer is "byte-addressable" so that each address (number) in $\mathbb{M}_n$ names the location of a byte of memory. How large must $n$ be to support an address space of the given sizes.

(a) A byte
Answer: $n = 1$ can address one byte at location 0.

(b) A kilobyte
Answer: $n = 10$ can address a kilobyte of memory from locations 0 to $2^{10} - 1 = 1023 \approx 10^3$.

(c) A megabyte
Answer: $n = 20$ can address a megabyte of memory from locations 0 to $2^{20} - 1 = 1,048,575 \approx 10^6$.

(d) A gigabyte
Answer: $n = 30$ can address $2^{30} \approx 10^9$ bytes of memory (a gigabyte).

(e) A terabyte
Answer: $n = 40$ can address $2^{40} \approx 10^{12}$ bytes of memory (a terabyte).

(f) A petabyte
Answer: $n = 50$ can address $2^{50} \approx 10^{15}$ bytes of memory (a petabyte).

(g) A exabyte
Answer: $n = 60$ can address $2^{60} \approx 10^{18}$ bytes of memory (an exabyte).

(h) A zettabyte
Answer: $n = 70$ can address $2^{70} \approx 10^{21}$ bytes of memory (a zettabyte).

14. $\aleph_0$ is the symbol for the cardinality of $\mathbb{N}$, the count of elements in the set of natural numbers. $\aleph_0$ is pronounced aleph-naught. Cantor called $\aleph_0$ a "transfinite" number. Cantor called the cardinality of $\mathbb{R}$, the set of real numbers, $\aleph_1$ and he proved $\aleph_0 \neq \aleph_1$.

(a) What is the cardinality of the integers $\mathbb{Z}$?
Answer: The integers have the same cardinality as the natural numbers, $\aleph_0$.

$$f(n) = \begin{cases} n/2 & \text{if } n \text{ is even} \\ -\frac{n+1}{2} & \text{if } n \text{ is odd} \end{cases}$$

is a one-to-one map of the natural numbers onto the integers.

(b) What is the cardinality of the rational numbers Q?
Answer: The rationals have the same cardinality as the natural numbers, $\aleph_0$.

(c) Why is transfinite a good word to use?
Answer: The prefix "trans" means "beyond" and these numbers are beyond finite.

(d) Are there transfinite numbers between $\aleph_0$ and $\aleph_1$?
Answer: Cantor hypothesized there are no sets with cardinality between $\aleph_0$ and $\aleph_1$. This is known as the continuum hypothesis. Interestingly, it turns out that the hypothesis can be proved or disproved using the standard axioms of set theory, provided set theory is consistent (does not contain contradictions).

(e) Read the Wikipedia page on natural numbers. The last time I checked, it provided a good summary of useful ideas.

## *Problems on Converting a Number from Decimal Notation*

**Background**

Decimal notation is the lingua franca for communicating about numbers: When we write a number such as 3.1459 or 2.71828 it is understood that it is written in decimal notation. When the context does not make it obvious we will use syntactic sugar and write $(3.1459)_{10}$ or $(2.71828)_{10}$ to denote that number is written in decimal notation. Likewise, we will write $(1001.1011)_2$, $(777.7)_8$, or $(BEEF)_{16}$ to denote that the number is written in binary, octal, or hexadecimal notation.

1. (Be able to use repeated division by 2 to convert a number from decimal to binary.) Convert the following natural numbers from decimal to binary.

(a) $q = 13$.
Answer: Repeatedly divide 13 and following quotients by 2 pushing the remainder bits onto a STACK.

| Repeated Remaindering Mod 2 | | | | |
|---|---|---|---|---|
| Quotients | 13 | 6 | 3 | 1 |
| Remainders | 1 | 0 | 1 | 1 |

$\therefore\ (13)_{10} = (1101)_2$

(b) $q = 32$.
Answer:

| Repeated Remaindering Mod 2 | | | | | | |
|---|---|---|---|---|---|---|
| Quotients | 32 | 16 | 8 | 4 | 2 | 1 |
| Remainders | 0 | 0 | 0 | 0 | 0 | 1 |

$\therefore\ (32)_{10} = (10\ 0000)_2$

(c) $q = 123$.

Answer:

| Repeated Remaindering Mod 2 | | | | | | | |
|---|---|---|---|---|---|---|---|
| Quotients | 123 | 61 | 30 | 15 | 7 | 3 | 1 |
| Remainders | 1 | 1 | 0 | 1 | 1 | 1 | 1 |

$\therefore\ (123)_{10} = (111\ 1011)_2$

(d) $q = 134$.

Answer:

| Repeated Remaindering Mod 2 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Quotients | 134 | 67 | 33 | 16 | 8 | 4 | 2 | 1 |
| Remainders | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 |

$\therefore\ (144)_{10} = (1000\ 0110)_2$

(e) $q = 145$.

Answer:

| Repeated Remaindering Mod 2 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Quotients | 145 | 72 | 36 | 18 | 9 | 4 | 2 | 1 |
| Remainders | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |

$\therefore\ (145)_{10} = (1001\ 0001)_2$

(f) $q = 257$.

Answer:

| Repeated Remaindering Mod 2 | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Quotients | 257 | 128 | 64 | 32 | 16 | 8 | 4 | 2 | 1 |
| Remainders | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |

$\therefore\ (257)_{10} = (1\ 0000\ 0001)_2$

2. ⚠ (Know that $\lfloor \lg n \rfloor + 1$ bits are required to represent the integer $n$.) How many bits are needed to represent each of the numbers in problem 1. What is the relationship to the logarithm base 2?

   Answer: 13 can be written in 4 bits; 32 can be written in 6 bits; 123 can be written in 7 bits; 134 can be written in 8 bits; 145 can be written in 8 bits; 257 can be written in 9 bits. In each case, $\lfloor (\lg 13) \rfloor + 1 = 4$, $\lfloor (\lg 32) \rfloor + 1 = 6$, $\lfloor (\lg 123) \rfloor + 1 = 7$, $\lfloor (\lg 134) \rfloor + 1 = 8$, $\lfloor (\lg 145) \rfloor + 1 = 8$, $\lfloor (\lg 257) \rfloor + 1 = 9$.

3. Convert the following natural numbers from decimal to hexadecimal.

   (a) $q = 13$.

   Answer: Using the answer to problem ,

$$(13)_{10} = (1101)_2 = (D)_{16}$$

(b) $q = 32$.
Answer: Using the answer to problem ,

$$(32)_{10} = (10\ 0000)_2 = (20)_{16}$$

(c) $q = 123$.
Answer: Using the answer to problem ,

$$(123)_{10} = (111\ 1011)_2 = (7B)_{16}$$

(d) $q = 134$.
Answer: Using the answer to problem ,

$$(134)_{10} = (1000\ 0110)_2 = (86)_{16}$$

(e) $q = 145$.
Answer: Using the answer to problem ,

$$(145)_{10} = (1001\ 0001)_2 = (91)_{16}$$

(f) $q = 257$.
Answer: Using the answer to problem ,

$$(257)_{10} = (1\ 0000\ 0001)_2 = (101)_{16}$$

4. How many hexadecimal numerals are needed to represent each of the numbers in problem 3. What is the relationship to the logarithm base 16?
Answer: 13 can be written in 1 hexadecimal numeral; 32 can be written in 2 hexadecimal numerals; 123 can be written in 2 hexadecimal numerals; 134 can be written in 2 hexadecimal numerals; 145 can be written in 2 hexadecimal numerals; 257 can be written in 3 hexadecimal numerals; In each case, $\lfloor(\log_{16} 13)\rfloor + 1 = 1$, $\lfloor(\log_{16} 32)\rfloor + 1 = 2$, $\lfloor(\log_{16} 123)\rfloor + 1 = 2$, $\lfloor(\log_{16} 134)\rfloor + 1 = 2$, $\lfloor(\log_{16} 145)\rfloor + 1 = 2$, $\lfloor(\log_{16} 257)\rfloor + 1 = 3$.
5. Convert the following rational numbers, written in decimal, to binary fixed point notation.
(a) $29/8$.
Answer: $29/8 = 3.625$. The integer 3 is written as 11. The fraction 0.625 is equal to $5/8 = 1/2 + 1/8$ which is binary $(0.101)_2$. Therefore $(29/8)_{10} = (11.101)_2$
(b) 0.3125.
Answer:

| Repeated Multiplying Mod 2 | | | | | |
|---|---|---|---|---|---|
| Fraction | 0.3125 | 0.625 | 1.25 | 0.50 | 1.0 |
| Overflow | 0. | 0 | 1 | 0 | 1 |

$\therefore\ 0.3125 = 0.0101$

(c) 5.1

Answer: The integer 5 is written as 101. The fraction one-tenth, 0.1, is converted to binary using repeated multiplication by 2.

| Repeated Multiplying Mod 2 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Fraction | 0.1 | 0.2 | 0.4 | 0.8 | 1.6 | 1.2 | 0.4 | 0.8 |
| Overflow | 0. | 0 | 0 | 0 | 1 | 1 | 0 | 0 |

$\therefore$ $5.1 = 101.0001100110011\cdots$. (Notice the bit pattern will repeat indefinitely.)

(d) $1/3$.

Answer:

| Repeated Multiplying Mod 2 | | | | | | | |
|---|---|---|---|---|---|---|---|
| Fraction | $1/3$ | $2/3$ | $4/3$ | $2/3$ | $4/3$ | $2/3$ | $4/3$ |
| Overflow | 0. | 0 | 1 | 0 | 1 | 0 | 1 |

$\therefore$ $1/3 = (0.01010101\cdots)_2$.

6. Convert the following rational numbers, written in decimal, to normalized binary floating point notation.

(a) $29/8$

Answer: From problem 5a $(29/8)_{10} = (11.101)_2$. Normalizing this gives $1.1101 \times 2^1$. Using a 8 bit representation with 3 exponent bits with a bias of 4 gives $1.1101 \times 2^1 = (0\ 101\ 1101)_{\text{fp}} = 29/8$. Notice this is a positive (first bit is 0), with biased exponent ($101 = 5$) giving a real exponent of $5 - 4 = 1$, and fraction 1011.

(b) $-29/8$

Answer: From the previous problem $-29/8 = (1\ 101\ 1101)_{\text{fp}}$.

(c) 0.3125

Answer: From problem 5b $(0.3125)_{10} = (0.0101)_2$. Normalizing this gives $1.01 \times 2^{-2}$. Using a 8 bit representation with 3 exponent bits with a bias of 4 gives $1.01 \times 2^{-2} = (0\ 010\ 0100)_{\text{fp}} = 0.3125$. Notice this is a positive (first bit is 0), with biased exponent ($010 = 2$) giving a real exponent of $2 - 4 = -2$, and fraction 01.

(d) 5.1

Answer: From problem 5c $(5.1)_{10} = (101.0001100110011\cdots)_2$. Normalizing this gives $1.010001100110011\cdots \times 2^2$. Using a 8 bit representation with 3 exponent bits with a bias of 4 gives

$$1.010001100110011\cdots \times 2^2 \approx (0\ 110\ 0100)_{\text{fp}}$$

Notice that a round-off error of $0.000001100110011\cdots$ occurred.

## *Problems on Converting a Number to Decimal Notation*

**Background**

An efficient algorithm to convert a number written in binary, octal, hexadecimal or another base to decimal notation is Horner's rule. Numbers are written in a positional (or polynomial) format

$$a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + a_{n-3}x^{n-3} + \cdots + a_2x^2 + a_1x^1 + a_0x^0$$

where $x$ is the base and the coefficients are numerals in the base alphabet. Horner's rule evaluates this expression from left-to-right in the parenthesized form

$$((\cdots((a_{n-1}x + a_{n-2})x + a_{n-3})x + \cdots + a_2)x + a_1)x + a_0$$

which can be organized in a tableaux

| Horner's Rule | | | | |
|---|---|---|---|---|
| $a_{n-1}$ | $a_{n-2}$ | $a_{n-3}$ | $\cdots$ | $a_0$ |
| | $a_{n-1}x$ | $(a_{n-1}x + a_{n-2}x$ | $\cdots$ | $a_{n-1}x^{n-1} + \cdots + a_1x$ |
| $a_{n-1}$ | $a_{n-1}x + a_{n-2}$ | $(a_{n-1}x + a_{n-2})x + a_{n-3}$ | $\cdots$ | $a_{n-1}x^{n-1} + \cdots + a_0$ |

1. (Be able to use Horner's rule to convert a number from some base to decimal.) Use Horner's rule to convert the following unsigned numbers into decimal notation.

   (a) $(0000\ 1001)_2$
   Answer:

| Horner's Rule | | | | | | | |
|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| | 0 | 0 | 0 | 0 | 2 | 4 | 8 |
| 0 | 0 | 0 | 0 | 1 | 2 | 4 | $\boxed{9}$ |

   $\therefore (1000\ 1001)_2 = (9)_{10}$.

   (b) $(1111\ 1111)_2$
   Answer:

| Horner's Rule | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | 2 | 6 | 14 | 30 | 62 | 126 | 254 |
| 1 | 3 | 7 | 15 | 31 | 63 | 127 | $\boxed{255}$ |

   $\therefore (1111\ 1111)_2 = (255)_{10}$.

   (c) $(1100\ 1100)_2$
   Answer:

| Horner's Rule | | | | | | | |
|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 |
| | 0 | 2 | 4 | 8 | 18 | 38 | 76 |
| 0 | 1 | 2 | 4 | 9 | 19 | 38 | $\boxed{76}$ |

   $\therefore (0100\ 1100)_2 = (76)_{10}$.

(d) $(1010\ 1010)_2$
Answer:

| Horner's Rule | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |
| | 2 | 4 | 10 | 20 | 42 | 84 | 170 |
| 1 | 2 | 5 | 10 | 21 | 42 | 85 | $\boxed{170}$ |

$\therefore (1010\ 1010)_2 = (170)_{10}$.

(e) $(1201)_3$
Answer:

| Horner's Rule | | | |
|---|---|---|---|
| 1 | 2 | 0 | 1 |
| | 3 | 15 | 45 |
| 1 | 5 | 15 | $\boxed{46}$ |

$\therefore (1201)_3 = (46)_{10}$.

(f) $(123)_4$
Answer:

| Horner's Rule | | |
|---|---|---|
| 7 | 4 | 7 |
| | 28 | 128 |
| 7 | 32 | $\boxed{135}$ |

$\therefore (747)_4 = (135)_{10}$.

(g) $(123)_8$
Answer:

| Horner's Rule | | |
|---|---|---|
| 1 | 2 | 3 |
| | 8 | 80 |
| 1 | 10 | $\boxed{83}$ |

$\therefore (123)_8 = (83)_{10}$.

(h) $(747)_8$
Answer:

| Horner's Rule | | |
|---|---|---|
| 7 | 4 | 7 |
| | 56 | 480 |
| 7 | 60 | $\boxed{487}$ |

$\therefore (747)_8 = (487)_{10}$.

(i) $(A47)_{16}$
Answer:

| Horner's Rule | | |
|---|---|---|
| 10 | 4 | 7 |
| | 160 | 2624 |
| 10 | 164 | $\boxed{2631}$ |

$\therefore (A47)_{16} = (2631)_{10}$.

(j) $(CAB)_{16}$
Answer:

| Horner's Rule | | |
|---|---|---|
| 12 | 10 | 11 |
| | 192 | 3232 |
| 12 | 202 | $\boxed{3243}$ |

$\therefore (A47)_{16} = (2631)_{10}$.

## *Two's Complement Notation for Integers*

**Background**

In computers, signed integers are represented using two's complement notation. This is because the fundamental arithmetical operations (addition, subtraction, multiplication) can be implemented simply by Boolean logic gates.

1. Convert the following signed decimal integers into two's complement notation.

(a) $-13$
Answer: Convert $|-13| = 13$ to unsigned binary 1101. Append a leading 0 to write $+13$ as 0 1101. Perform the two's complement operation to return 1 0011 as $-13$.

(b) $+134$
Answer: Convert $|134| = 134$ to unsigned binary 1000 0110. Append a leading 0 and return $+134$ as 0 1000 0110.

(c) $-145$
Answer: Convert $|-145| = 145$ to unsigned binary 1001 0001. Append a leading 0 to write $+145$ as 0 1001 0001. Perform the two's complement operation to return 1 0110 1111 as $-145$.

(d) 128
Answer: Convert 128 to unsigned binary

$$(128)_{10} = (1000\ 0000)_2$$

Append a leading 0 to write $+128$ as 0 1000 0000.

(e) $-128$

Answer: Convert $|-128| = 128$ to unsigned binary

$$(128)_{10} = (1000\ 0000)_2$$

Append a leading 0 to write $+128$ as 0 1000 0000. Perform the two's complement operation to return 1 1000 000 as $-128$.

(f) $-257$

Answer: Convert $|-257| = 257$ to unsigned binary

$$(257)_{10} = (1\ 0000\ 0001)_2$$

Append a leading 0 to write $+257$ as 01 0000 0001. Perform the two's complement operation to return 10 1111 1111 as $-257$.

(g) $+76$.

Answer: First convert unsigned 76 to binary: Quotients and remainders in the table below are computed from left-to-right.

| Repeated Remaindering Mod 2 | | | | | | | |
|---|---|---|---|---|---|---|---|
| Quotients | 76 | 38 | 19 | 9 | 4 | 2 | 1 |
| Remainders | 0 | 0 | 1 | 1 | 0 | 0 | 1 |

Prefixing string 100 1100 with a sign bit gives the two's complement representation

$$+76 \equiv (0100\ 1100)_{\text{tc}}$$

(h) $-76$.

Answer: Using the result from problem 1g we know $+76 \equiv 0100\ 1100$. Negate this value, copy the bits from right-to-left up to and including the first 1. Flip the remaining bits. This gives

$$-76 \equiv (1011\ 0100)_{\text{tc}}$$

Notice that

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Carry | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| | | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 |
| | | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 |
| Sum | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

(i) $-137$.

Answer: First convert unsigned 137 to binary: Quotients and remainders in the table below are computed from left-to-right.

| Repeated Remaindering Mod 2 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Quotients | 137 | 68 | 34 | 17 | 8 | 4 | 2 | 1 |
| Remainders | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |

Prefixing the bit string 1000 1001 with a sign bit gives the two's complement representation

$$+137 \equiv 0\ 1000\ 1001$$

Negate $+137 \equiv 0\ 1000\ 1001$ to get the answer

$$-137 \equiv (1\ 0111\ 0111)_{\text{tc}}$$

(j) $+177$.
Answer: First convert unsigned 177 to binary using repeated remaindering.

| Repeated Remaindering Mod 2 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Quotients | 177 | 88 | 44 | 22 | 11 | 5 | 2 | 1 |
| Remainders | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 |

Prefixing the bit string 1011 0001 with a sign bit gives the two's complement representation

$$+177 \equiv 0\ 1011\ 1011$$

2. Write each of the numbers in problem 1 using 12 bits.
   (a) $-13$
   Answer: Pad negative numbers with 1's on the left. $-13 = 1111\ 1111\ 0011$.
   (b) $q = +134$
   Answer: Pad positive numbers with 0's on the left. $+134 = 0000\ 1000\ 0110$.
   (c) $-145$
   Answer: Pad negative numbers with 1's on the left. $-145 = 1111\ 0110\ 1111$.
   (d) $-257$
   Answer: Pad negative numbers with 1's on the left. $-257 = 1110\ 1111\ 1111$.
   (e) 76
   Answer: Pad negative numbers with 1's on the left. $76 = 0000\ 0100\ 1100$.
   (f) $-76$
   Answer: Pad negative numbers with 1's on the left. $-76 = 1110\ 1011\ 0100$.
   (g) $-137$
   Answer: Pad negative numbers with 1's on the left. $-137 = 1111\ 0111\ 0111$.
   (h) 177
   Answer: Pad negative numbers with 1's on the left. $177 = 0000\ 1011\ 1011$.
3. Negate the two's complement integers below.
   (a) 0100 1100.
   Answer: The negative of 0100 1100 is 1011 0100. Notice the two numbers sum to 1 0000 0000 which is truncated to 8 bits, all zero.
   (b) 1100 0000.
   Answer: The negative of 1100 0000 is 0100 0000.

(c) 1010 0100.

Answer: The negative of 1010 0100 is 0101 1100.

(d) 1000 0000.

Answer: Note that the negative of 1000 0000 cannot be written in 8 bits, so extend the sign by writing 1000 0000 = 1 1000 0000 and then negate this string to get 0 1000 0000.

4. Convert the following two's complement integers into decimal notation.

(a) 0100 1100.

Answer: The leftmost sign bit, 0, declares the number is positive. Use Horner's rule to compute

| Horner's Rule | | | | | | | |
|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 |
| | 0 | 2 | 4 | 8 | 18 | 38 | 76 |
| 0 | 1 | 2 | 4 | 9 | 19 | 38 | $\boxed{76}$ |

Since the two's complement number is positive and 0100 1100 represents 76

(b) 1100 0000.

Answer: The leftmost sign bit, 1, declares the number is negative. Use Horner's rule to compute

| Horner's Rule | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 2 | 6 | 12 | 24 | 48 | 96 | 192 |
| 1 | 3 | 6 | 12 | 24 | 48 | 96 | $\boxed{192}$ |

Since the two's complement number is negative and $192 + 64 = 256$, 1100 0000 represents $-64$. An alternative algorithm is to two's complement 1100 0000 to get 0100 0000, convert this number to decimal 64, and return $-64$ as the answer.

(c) 1010 0100.

Answer: The leftmost sign bit, 1, declares the number is negative. Use Horner's rule to compute

| Horner's Rule | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| | 2 | 4 | 10 | 20 | 40 | 82 | 164 |
| 1 | 2 | 5 | 10 | 20 | 41 | 82 | $\boxed{164}$ |

Since the two's complement number is negative and $164 + 92 = 256$, 1010 0100 represents $-92$. An alternative algorithm is to two's complement 1010 0100 to get 0101 1100, convert this number to decimal 92, and return $-92$ as the answer.

(d) 0010 1000.

Answer: The leftmost sign bit, 0, declares the number is positive. Use Horner's rule to compute

| Horner's Rule | | | | | | | |
|---|---|---|---|---|---|---|---|
| 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 |
| | 0 | 0 | 2 | 4 | 10 | 20 | 40 |
| 0 | 0 | 1 | 2 | 5 | 10 | 20 | $\boxed{40}$ |

Since the two's complement number is positive 0010 1000 represents $+40$.

(e) $62A3$.

(f) $C01B$.

5. Using two's complement notation what range of integers from most negative to most positive can be represented using

(a) 2 bits?
Answer: The range is from $-2^1 = -2$ to $2^1 - 1 = 1$.

(b) 8 bits?
Answer: The range is from $-2^7 = -128$ to $2^7 - 1 = 127$.

(c) 32 bits?
Answer: The range is from $-2^{31} = -2\ 147\ 483\ 648$ to $2^{31} - 1 = 2\ 147\ 483\ 647$.

(d) 64 bits?
Answer: The range is from $-2^{63}$ to $2^{63} - 1$. How big is $2^{63}$? Approximate it by a power of 10.

6. Consider ten's complement notation. An integer and its complement should sum to zero. Writing in 4 digits the integer 1234 has complement 8766 since $1234 + 8766 = 1\ 0000$. The smaller numbers from 0000 to 4999 are positive and have their "normal" value. The larger values 5000 to 9999 are negative and defined by the complement rule illustrated in the example.

(a) What are the complements of (a) 3141? (b) 7316? (c) 9999? (d) 5000?
Answer: The complements are (a) 6859; (b) 2684; (c) 0001; (d) 5000 (note 5000 represents decimal $-5000$ and its negative 5000 cannot be represent with 4 digits in ten's complement notation. $(5000)_{10} = (05000)_{10c}$)

(b) What are the decimal numbers for the ten's complements numbers (a) 3141? (b) 7316? (c) 9999? (d) 5000?
Answer: The decimal number are (a) 3141; (b) $-2684$; (c) $-1$; (d) $-5000$.

## *Biased and Floating Point Notation*

**Background**

Biased notation is an alternative way to encode positive and negative integers. It is used to represent the exponent in floating point notation. A base 10 floating point number is familiarly called scientific notation, for example,

$$5.772 \times 10^{-1}$$

where base 2 floating point numbers are written in binary, for example

$$-1.01101 \times 2^{-4}$$

The "point" floats in that we can write

$$-1.01101 \times 2^{-4} = -10.1101 \times 2^{-5} = -0.101101 \times 2^{-3}$$

and other equivalent representations of the same number. The "normalized" representation has one non-zero digit before the decimal point. For instance

$$-1.01101 \times 2^{-4} \qquad \text{and} \qquad 3.141 \times 10^{0}$$

are normalized binary and decimal floating point numbers.

A biased integer $(n)_{\text{bias}=b}$ is a non-negative number that represents the value $n - b$, for example, $(42)_{\text{bias}=64}$ stands for $42 - 64 = -22$.

The IEEE 754 Standard for Floating Point Arithmetic is what (most) computer vendors use in their floating point processors. It is too involved to describe in this class. In most examples, we use a pidgin version that uses 8 bits to encode a floating point number: Read left-to-right 1 sign bit, 3 exponent bits, and 4 fraction bits. For instance,

$$(1\ 101\ 0101)_{fp} = (-1)^1 \times 1.0101^{(101)_{\text{bias}=4}} = -\frac{21}{16} \times 2^1 = \frac{21}{8}$$

1. Convert the following signed integers from decimal to biased notation.

   (a) $-13$ with bias $b = 32$.
   Answer: Add the biased $b = 32$ to $-13$. The biased number $(19)_{\text{bias}=32}$ represents the decimal integer $-13$.

   (b) 134 with bias $b = 256$.
   Answer: Add the bias 256 to 134. The biased number $(390)_{\text{bias}=256}$ represents the decimal integer $+134$.

   (c) $-145$ with bias $b = 256$.
   Answer: Add the bias 256 to $-145$. The biased number $(111)_{\text{bias}=256}$ represents the decimal integer $-145$.

   (d) $-257$ with bias $b = 512$.
   Answer: Add the bias 512 to $-257$. The biased number $(255)_{\text{bias}=512}$ represents the decimal integer $-257$.

2. Convert the following biased numbers to decimal signed integers.

   (a) $(18)_{\text{bias}=16}$
   Answer: Subtract the biased 16 from 18. The biased number $(18)_{\text{bias}=16}$ represents the decimal integer $2 = 18 - 16$.

   (b) $(7)_{\text{bias}=16}$
   Answer: Subtract the biased 16 from 7. The biased number $(7)_{\text{bias}=16}$ represents the decimal integer $-9 = 7 - 16$.

   (c) $(45)_{\text{bias}=32}$
   Answer: Subtract the biased 32 from 45. The biased number $(45)_{\text{bias}=32}$ represents the decimal integer $13 = 45 - 32$.

   (d) $(45)_{\text{bias}=128}$
   Answer: Subtract the biased 128 from 45. The biased number $(45)_{\text{bias}=128}$ represents the decimal integer $-83 = 45 - 128$.

3. Convert the following biased binary numbers to two's complement number.

(a) $(1100)_{\text{bias}=16}$
Answer: Subtract the biased $16 = (1\ 0000)_2$ from $12 = (1010)_2$ to get $-4$. The biased number $(1010)_{\text{bias}=16}$ represents decimal value $-4$ which is $(1100)_{2c}$ in two's complement notation.

(b) $(11\ 1001)_{\text{bias}=16}$
Answer: Subtract the biased $16 = (1\ 0000)_2$ from $29 = (11\ 1001)_2$ to get 13. The biased number $(11\ 1001)_{\text{bias}=16}$ represents decimal value 13 which is $(0\ 1101)_{2c}$ in two's complement notation.

(c) $(1001\ 1001)_{\text{bias}=128}$
Answer: Subtract the biased $128 = (1000\ 0000)_2$ from $153 = (1001\ 1001)_2$ to get 25. The biased number $(1001\ 1001)_{\text{bias}=128}$ represents decimal value 25 which is $(01\ 1001)_{2c}$ in two's complement notation.

(d) $(1000\ 0000)_{\text{bias}=128}$
Answer: Subtract the biased $128 = (1000\ 0000)_2$ from $128 = (1000\ 0000)_2$ to get 0. The biased number $(1000\ 0000)_{\text{bias}=128}$ represents decimal value 0 which is $(0)_{2c}$ in two's complement notation.

4. The following binary strings are floating point numbers where the first (leftmost) bit is a sign bit, the next three bits are exponent bits written in biased notation with bias $b = 3$, and the last four bits are fraction bits. These floating point are normalized. What are decimal names for these floating point numbers?

(a) 1 001 1000
Answer: This is a negative number (the sign bit is 1) with exponent $(001)_{b=3} = 1 - 3 = -2$ and fixed point value $(1.1000)_2 = 3/2$.

$$\therefore 1\ 001\ 1000 = -\frac{3}{2} \times 2^{-2} = -\frac{3}{8}$$

(b) 0 101 1011
Answer: This is a positive number (the sign bit is 0) with exponent $(101)_{b=3} = 5 - 3 = 2$ and fixed point value $(1.1011)_2 = 27/16$.

$$\therefore 0\ 101\ 1011 = \frac{27}{16} \times 2^{2} = -\frac{27}{4}$$

(c) 1 111 1111
Answer: This is a negative number (the sign bit is 1) with exponent $(111)_{b=3} = 7 - 3 = 4$ and fixed point value $(1.1111)_2 = 31/16$.

$$\therefore 1\ 111\ 1111 = -\frac{31}{16} \times 2^{4} = -31$$

(d) 0 000 0001
Answer: This is a positive number (the sign bit is 0) with exponent $(000)_{b=3} = 0 - 3 = -3$ and fixed point value $(1.0001)_2 = 17/16$.

$$\therefore 0\ 000\ 0001 = \frac{17}{16} \times 2^{-3} = -\frac{17}{128}$$

5. The following binary strings are floating point numbers where the first (leftmost) bit is a sign bit, the next 5 bits are exponent bits written in biased notation with bias $b = 16$, and the last 6 bits are fraction bits. These floating point are normalized. What are decimal names for these floating point numbers?

   (a) 1 00100 100001

   Answer: This is a negative number (the sign bit is 1) with exponent $(00100)_{b=16} = 4 - 16 = -12$ and fixed point value $(1.100001)_2 = 97/64$.

$$\therefore 1\ 00100\ 100001 = -\frac{97}{64} \times 2^{-12} = -\frac{97}{2^{18}}$$

   (b) 0 10100 111000

   Answer: This is a positive number (the sign bit is 0) with exponent $(10100)_{b=16} = 20 - 16 = 4$ and fixed point value $(1.111000)_2 = 15/8$.

$$\therefore 0\ 10100\ 111000 = \frac{15}{8} \times 2^4 = 30$$

6. Using an 8 bit normalized floating point notation explain why 17/128 is the smallest positive floating point number and why $1/8 = 16/128$ is not.

   Answer: The smallest non-zero number is represented by the string 0 000 0001. This string is interpreted as

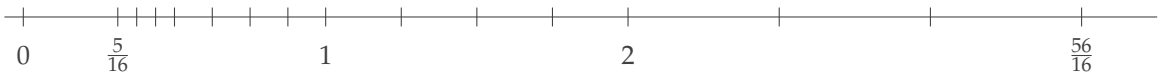
$$(-1)^0\,(1.0001)_2 \times 2^{0-3} = \frac{17}{16}\frac{1}{8} = \frac{17}{128}$$

   On the other hand, $1/8 = (1.0000)_2 \times 2^{-3}$ and that would be written as the all zero string, which is reserved for the value zero.

7. Plot, along a number line, the positive values that can be represented using 1 sign bit, 2 exponent bits with bias 2, and 2 normalized fraction bits.

   Answer: The values that can be represented are

$$0\ 00\ 00 = 0$$
$$0\ 00\ 01 = \frac{5}{4} \times 2^{-2} = \frac{5}{16}$$
$$0\ 00\ 10 = \frac{6}{4} \times 2^{-2} = \frac{6}{16}$$
$$0\ 00\ 11 = \frac{7}{4} \times 2^{-2} = \frac{7}{16}$$
$$0\ 01\ 00 = 1 \times 2^{-1} = \frac{8}{16}$$
$$0\ 01\ 01 = \frac{5}{4} \times 2^{-1} = \frac{10}{16}$$
$$0\ 01\ 10 = \frac{6}{4} \times 2^{-1} = \frac{12}{16}$$
$$0\ 01\ 11 = \frac{7}{4} \times 2^{-1} = \frac{14}{16}$$
$$0\ 10\ 00 = 1 \times 2^{0} = \frac{16}{16}$$
$$0\ 10\ 01 = \frac{5}{4} \times 2^{0} = \frac{20}{16}$$
$$0\ 10\ 10 = \frac{6}{4} \times 2^{0} = \frac{24}{16}$$
$$0\ 10\ 11 = \frac{7}{4} \times 2^{0} = \frac{28}{16}$$
$$0\ 11\ 00 = 1 \times 2^{1} = \frac{32}{16}$$
$$0\ 11\ 01 = \frac{5}{4} \times 2^{1} = \frac{40}{16}$$
$$0\ 11\ 10 = \frac{6}{4} \times 2^{1} = \frac{48}{16}$$
$$0\ 11\ 11 = \frac{7}{4} \times 2^{1} = \frac{56}{16}$$


0
5/16
1
2
56/16