Chapter 1

Negative numbers representation

§1. Complement and its calculation

§1.1. Concept of complement

- Complement of a digit

$$\left.\begin{aligned} &r = radix \\ &d_i \in \{0,1,\ldots,(r-1)\} \\ &\overline{d} = (r-1) - d \end{aligned}\right\} \quad (1.1.1)$$

- *Example*

$$\left.\begin{aligned} &r = 10 \\ &d_i \in \{0,1,\ldots,9\} \\ &\overline{d} = 9 - d \\ &d = 7 \\ &\overline{d} = 9 - 7 = 2 \end{aligned}\right\} \quad (1.1.2)$$

- *r*' s complement of a number (radix complement)

$$\left.\begin{aligned} &N(n,m) \\ &N = d_{n-1}d_{n-2}\ldots d_1 d_0 . d_{-1} d_{-2} \ldots d_{-m} \\ &d_i \in \{0,1,\ldots,(r-1)\} \\ &N = d_{n-1}\cdot r^{n-1} + d_{n-2}\cdot r^{n-2} + \ldots + d_1 \cdot r^1 + \\ &\quad + d_0 \cdot r^0 + d_{-1}\cdot r^{-1} + \ldots + d_{-m}\cdot r^{-m} \\ &\overline{N} = (r^n - N) \bmod r^n \end{aligned}\right\} \quad (1.1.3)$$

- *Example*

$$\left.\begin{aligned} &r = 10 \\ &d_i \in \{0,1,\ldots,9\} \\ &N(3,2) \\ &N = 327.45 \\ &\overline{N} = 10^3 - 327.45 = 672.55 \end{aligned}\right\} \quad (1.1.4)$$

- Property

$$\left.\begin{aligned} &r = radix \\ &N(n,m) \\ &\overline{N} = (r^n - N)\, mod\ r^n \\ &\overline{(\overline{N})} = N \\ &\overline{(\overline{N})} = \overline{((r^n - N)\, mod\ r^n)} = \\ &\quad = (r^n - (r^n - N)\, mod\ r^n)\, mod\ r^n = N \end{aligned}\right\} \quad (1.1.5)$$

- *(r-1)*'s complement of a number (diminished radix complement)

$$\left.\begin{aligned} &N(n,m) \\ &d_i \in \{0,1,\ldots,(r-1)\} \\ &N = d_{n-1}\ldots d_1 d_0 d_{-1}\ldots d_{-m} \\ &\overline{\overline{N}} = r^n - N - r^{-m} \end{aligned}\right\} \quad (1.1.6)$$

- *Example*

$$\left.\begin{aligned} &r = 10 \\ &d_i \in \{0,1,\ldots,9\} \\ &N(3,2) \\ &N = 327.45 \\ &\overline{\overline{N}} = 10^3 - 327.45 - 10^{-2} = \\ &\quad = 1000 - 327.45 - 0.01 = 672.54 \end{aligned}\right\} \quad (1.1.7)$$

- Property

$$\left.\begin{array}{l} r = radix \\ \overline{\overline{(\overline{\overline{N}})}} = N \\ N(n,m) \\ \overline{\overline{N}} = r^n - N - r^{-m} \\ \overline{\overline{(\overline{\overline{N}})}} = r^n - \left(r^n - N - r^{-m}\right) - r^{-m} = N \end{array}\right\} \quad (1.1.8)$$

- Property

$$\left.\begin{array}{l} r = radix \\ N(n,m) \\ \overline{N} = \left(r^n - N\right) \\ \overline{\overline{N}} = r^n - N - r^{-m} \\ \overline{N} = \overline{\overline{N}} + r^{-m} \end{array}\right\} \quad (1.1.9)$$

$$\left.\overline{\overline{N}} = \overline{N} - r^{-m}\right\} \quad (1.1.10)$$

§1.2. Methods for complement calculation

1.2.1 Radix 10

1. *10's complement*

- a) According to definition:

$$\left.\begin{array}{l} N_{10}(n,m) = d_{n-1}\ldots\ldots d_1 d_0 d_{-1}\ldots\ldots d_{-m} \\ \overline{N} = \left(10^n - N\right)\text{mod}10^n \end{array}\right\} \quad (1.1.1)$$

- *Example*:

$$\left.\begin{array}{l} n = 3,\ m = 0 \\ N = 371 \\ \overline{N} = \left(10^3 - 371\right)\text{mod}10^3 = \\ \quad = \left(1000 - 371\right)\text{mod}1000 = 629 \end{array}\right\} \quad (1.1.2)$$

- b) According to the following rule:

$$\left.\begin{array}{l} N = d_{n-1}\ldots\ldots\ldots d_0 d_{-1}\ldots\ldots\ldots\ldots\ldots\ldots\ldots d_{-k+1} d_{-k}\, 0\ldots 0 \\ \quad\downarrow \qquad\qquad \downarrow\ \ \downarrow \qquad\qquad\qquad \swarrow \qquad \downarrow\ \downarrow\ \ \downarrow \\ \left(9 - d_{n-1}\right)\ldots\left(9 - d_0\right)\left(9 - d_{-1}\right)\ldots\left(9 - d_{-k+1}\right)\left(10 - d_{-k}\right)0\ldots 0 \\ d_{-k} \neq 0 \end{array}\right\} \quad (1.1.3)$$

- *Example*:

$$\left.\begin{array}{l} N_{10} = 35.4700 \\ \qquad \downarrow\downarrow\ \downarrow\downarrow\downarrow\downarrow \\ \overline{N_{10}} = 64.5300 \end{array}\right\} \quad (1.1.4)$$

- c) Based on the general property

$$\left.\overline{N_r} = \overline{\overline{N_r}} + r^{-m}\right\} \quad (1.1.5)$$

If $r = 10$, then

$$\overline{N_{10}} = \overline{\overline{N_{10}}} + 10^{-m}$$

Add "1" to the rightmost position of the 9's complement representation. (1.1.6)

2. *9's complement*

- a) According to definition

$$\left.\begin{array}{l} N_{10}(n,m) = d_{n-1}\ldots\ldots\ldots d_0 d_{-1}\ldots\ldots\ldots d_{-m} \\ \overline{\overline{N_{10}}} = 10^n - N_{10} - 10^{-m} \end{array}\right\} \quad (1.2.1)$$

- *Example*:

$$\left.\begin{aligned} &N_{10}(3,2)\\ &N_{10} = 325.49\\ &\overline{\overline{N_{10}}} = 10^3 - 325.49 - 10^{-2} =\\ &\quad = 1000 - 325.49 - 0.01 = 674.50 \end{aligned}\right\} \quad (1.2.2)$$

- b) According to the following rule:

$$\left.\begin{aligned} &N_{10} = d_{n-1}\ldots\ldots\ldots\ldots d_0 d_{-1}\ldots\ldots\ldots\ldots d_{-m}\\ &\qquad\downarrow\qquad\qquad\swarrow\ \searrow\qquad\qquad\downarrow\\ &\overline{\overline{N_{10}}} = (9-d_{n-1})\ldots(9-d_0)(9-d_{-1})\ldots(9-d_{-m}) \end{aligned}\right\} \quad (1.2.3)$$

- *Example*:

$$\left.\begin{aligned} &N_{10}(3,2)\\ &N_{10} = 325.49\\ &\overline{N_{10}} = (9-3)(9-2)(9-5).(9-4)(9-9)\\ &\quad = 674.50 \end{aligned}\right\} \quad (1.2.4)$$

- c) According to the general property

$$\left.\overline{\overline{N}}_r = \overline{N}_r - r^{-m}\right\} \quad (1.2.5)$$

If r =10, then

$$\overline{\overline{N_{10}}} = \overline{N_{10}} - 10^{-m} \quad (1.2.6)$$

Subtract "1" from the rightmost position of the 10's complement representation.

1.2.2 Radix 2

1. ***2's complement***

- According to definition

$$\left.\begin{aligned} &N_2(n,m) = b_{n-1}\ldots\ldots\ldots b_1 b_0\ b_{-1}\ \ldots\ldots\ldots\ldots\ b_{-m}\\ &\overline{N_2} = 2^n - N \quad or \qquad 0 - N \end{aligned}\right\} (2.1.1)$$

$$\left.\begin{aligned} &2^n = 1[00\ldots\ldots\ldots\ldots\ldots 00]\\ &\quad(n)(n-1)(n-2)\ldots 10 \end{aligned}\right\} (2.1.2)$$

- b) According to the following rule:

$$\left.\begin{aligned} &N_2(n,m) = b_{n-1}\ldots\ldots\ldots b_1 b_0\ b_{-1}\ \ldots\ldots\ldots\ b_{-m}\\ &N_2 = \quad b_{n-1}\ldots\ldots\ldots b_0\ b_{-1}\ \ldots\ldots\ldots\ b_{-k+1}\ b_{-k}\quad 0..0\\ &\qquad\quad\downarrow\qquad\quad\downarrow\ \ \downarrow\qquad\qquad\downarrow\quad\ \downarrow\ \ \downarrow\ \downarrow\\ &\overline{N_2} = (1-b_{n-1})..(1-b_0)(1-b_{-1})..(1-b_{-k+1})(2-b_{-k})0..0\\ &b_{-k} \neq 0 \end{aligned}\right\} (2.1.3)$$

- *Example*:

$$\left.\begin{aligned} &N_2(3,4)\\ &N_2 = 1 \quad 0 \quad 1. \quad 1 \quad 1 \quad 0 \quad 0\\ &\quad (1-1)(1-0)(1-1)(1-1)(2-1)\ \downarrow\ \ \downarrow\\ &\overline{N_2} = 0 \quad 1 \quad 0. \quad 0 \quad 1 \quad 0 \quad 0 \end{aligned}\right\} (2.1.4)$$

- c) According to the general property

$$\left.\overline{N_r} = \overline{\overline{N_r}} + r^{-m}\right\} (2.1.5)$$

If r =2, then

$$\overline{N_2} = \overline{\overline{N_2}} + 2^{-m} \quad (2.1.6)$$

Add "1" to the rightmost position of the 1's complement representation.

2. *1's complement*

- a) According to definition

$$\left.\begin{aligned} &N_2(n,m)=b_{n-1}\ldots\ldots b_0 b_{-1}\ldots\ldots b_{-m} \\ &\overline{\overline{N_2}} = 2^n - N_2 - 2^{-m} \quad \text{or} \quad 0 - N_2 - 2^{-m} \end{aligned}\right\} \quad (2.2.1)$$

- b) According to the following rule:

$$\left.\begin{aligned} &N_2 = b_{n-1}\ldots\ldots b_0 b_{-1}\ldots\ldots b_{-m} \\ &\quad\downarrow \qquad\qquad \downarrow\ \downarrow \qquad\quad \downarrow \\ &\overline{N_2} = (1-b_{n-1})\ldots(1-b_0)(1-b_{-1})\ldots(1-b_{-m}) \end{aligned}\right\} \quad (2.2.2)$$

It results that each bit of N_2 is complemented,
since $\overline{b_i} = 1 - b_i$ } (2.2.3)

- Therefore:

$$\overline{\overline{N}}_2 = \overline{b_{n-1}}\ \overline{b_{n-2}}\ldots\overline{b_0}\ \overline{b_{-1}}\ldots\overline{b_{-m}} \quad (2.2.4)$$

- *Example*:

$$\left.\begin{aligned} &N_2(3,4) \\ &N_2 = 101.1101 \\ &\overline{N_2} = \overline{101}.\overline{1101} = 010.0010 \end{aligned}\right\} \quad (2.2.5)$$

- c) According to the general property:

$$\overline{\overline{N}}_r = \overline{N}_r - r^{-m} \quad (2.2.6)$$

If r =2, then

$$\overline{\overline{N}}_2 = \overline{N}_2 - 2^{-m} \quad (2.2.7)$$

Subtract "1" from the rightmost position of the 2's complement representation

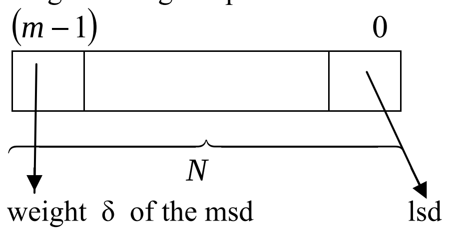
§2. Special codes for negative numbers representation

§2.1. General formulation of the problem

- A general numbering system with radix r:

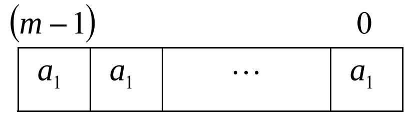
$$\left.\begin{aligned} &r = radix > 1 \\ &a_i \in \{a_1 \ldots a_n\} \\ &a_n = a_1 + (r-1) \\ &0 \text{ must be considered in the set } \{a_1 \ldots\ldots a_n\} \end{aligned}\right\} \quad (2.1.1)$$

- Range of integer representable numbers on m positions:



(2.1.2)

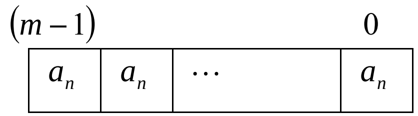
- $N_{\min}$



(2.1.3)

$$\left.\begin{aligned} \bullet\ N_{\min} &= a_1\cdot\delta + a_1\cdot\delta\cdot r^{-1} + \ldots + a_1\cdot\delta\cdot r^{-m+1} = \\ &= a_1\cdot\delta\cdot\left(1 + r^{-1} + \ldots + r^{-m+1}\right) = a_1\cdot\delta\cdot\frac{1-r^{-m}}{1-r^{-1}} \end{aligned}\right\} \quad (2.1.4)$$

- $N_{\max}$



(2.1.5)

- $N_{\max} = a_n \cdot \delta + a_n \cdot \delta \cdot r^{-1} + \ldots + a_n \cdot \delta \cdot r^{-m+1} =$
  $= a_n \cdot \delta \cdot \left(1 + r^{-1} + \ldots + r^{-m+1}\right) = a_n \cdot \delta \cdot \frac{1 - r^{-m}}{1 - r^{-1}}$ (2.1.6)

- Analysis of the sign:

$$\begin{cases} \frac{1 - r^{-m}}{1 - r^{-1}} > 0 \\ r > 1 \end{cases} \Rightarrow \text{Sign of } \{N_{\min}, N_{\max}\} \text{ depends on the sign of } a_1 \text{ and } a_n \quad (2.1.7)$$

- Symmetry condition
  $a_1 = -a_n$
  Therefore only *odd radices* would comply:
  $B_3 = \{-1, 0, +1\}$

  $B_5 = \{-2, -1, 0, +1, +2\}$ (2.1.8)

  In practice, the even radices are used, like:
  $B_2 = \{0, 1\}$
  $B_{10} = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$

- *Conclusion*:
  Natural representations are not possible, thereby necessity of special codes, with extra positions to identify the sign. (2.1.9)

- Three major types of special representations (codes): (2.1.10)
  - 1. Sign-Magnitude / Direct Code
  - 2. Radix Complement / Two's Complement Code/ / Complementary Code
  - 3. Diminished Radix Complement / Inverse Code / One's Complement Code/

§2.2. Sign Magnitude representation

- SM or Direct Code (2.2.1)
- Additional position containing the sign bit (b_s):
  if $N>0 \longrightarrow b_s = 0$
  if $N<0 \longrightarrow b_s = 1$ (2.2.2)

- b_s without weight, therefore placed in any position (2.2.3)

  $N(n, m)$

  $|N| = b_{n-1} b_{n-2} \ldots\ldots b_0 b_{-1} \ldots\ldots b_{-m} = \sum_{i=-m}^{n-1} b_i 2^i$

  - $+|N| \rightarrow 0 \left| \sum_{-m}^{n-1} b_i 2^i \right.$
  - $-|N| \rightarrow 1 \left| \sum_{-m}^{n-1} b_i 2^i \right.$ (2.2.4)

- *Example*:
  $|N| = 110.011$
  $+|N| = 0 \,|\, 110.011$
  $-|N| = 1 \,|\, 110.011$ (2.2.5)

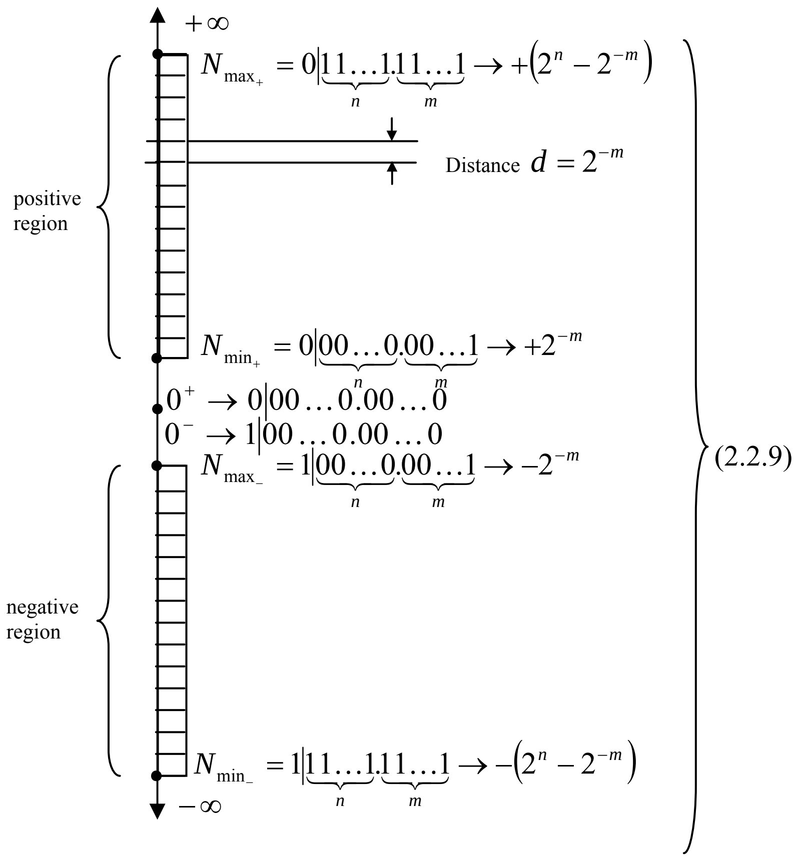
- If N is a fractional number:

  $|N| = b_{-1} b_{-2} \ldots\ldots b_{-m} = \sum_{i=-m}^{-1} b_i 2^i$ (2.2.6)

- Then:
  - $+|N| \rightarrow 0 \left| \sum_{-m}^{-1} b_i 2^i \right.$
  - $-|N| \rightarrow 1 \left| \sum_{-m}^{-1} b_i 2^i \right.$ (2.2.7)

- Representation of 0:
  - clean zero: $0^+ \to 0|\underbrace{00\ldots\ldots0}_{n}.\underbrace{00\ldots\ldots0}_{m}$
  - dirty zero: $0^- \to 1|\underbrace{00\ldots\ldots0}_{n}.\underbrace{00\ldots\ldots0}_{m}$ (2.2.8)

- Range of representation:



§2.3. Two's Complement Representation

2.3.1 General considerations

- $N(n,m)$ $\quad \overline{N} = 2^n - N$
- Sign bits with weight
- In case of negative numbers → the magnitude represented in two's complement code.
- Three variants: variant 1, variant 2, variant 3. (3.1.1)

2.3.2 Variant 1

- $N(n,m),\ r = 2$
- $$|N| = \underbrace{b_{n-1}b_{n-2}\ldots\ldots b_0}_{n} \cdot \underbrace{b_{-1}\ldots\ldots b_{-m}}_{m} = \\ = b_{n-1}\cdot 2^{n-1} + \ldots\ldots b_0 \cdot 2^0 + b_{-1}\cdot 2^{-1} \ldots\ldots b_{-m}\cdot 2^{-m} \quad (3.2.1)$$
- $b_s =$ sign bit with weight δ_s
- $\delta_s = -2^n \rightarrow$ **The weight of the sign bit is negative** (3.2.2)
- General representation of $\pm N$
$$N = b_s\left(-2^n\right) + \sum_{i=-m}^{n-1} b_i 2^i = b_s\left(-2^n\right) + N^* \quad (3.2.3)$$

where:

- if $N > 0 \rightarrow b_s = 0$ and N^* is represented in direct code
  - $N = 0\cdot\left(-2^n\right) + \sum_{i=-m}^{n-1} b_i^* 2^i$ (3.2.4)

- if $N < 0 \rightarrow b_s = 1$ and N^* is represented in two's complement code $\overline{N^*}$ (3.2.5)
  - $$N = 1 \cdot \left(-2^n\right) + \underbrace{\sum_{-m}^{n-1} \overline{b_i} 2^i + 2^{-m}}_{Two's\ complement\ of\ N^*}$$

- In case of fractional numbers, $n = 0$

  $|N| = \sum_{-m}^{-1} b_i 2^i$, b_s in position 0 (3.2.6)

- $$+|N| \rightarrow 0\left(-2^0\right) + \sum_{-m}^{-1} b_i 2^i \quad (3.2.7)$$

- $$-|N| \rightarrow 1\left(-2^0\right) + \underbrace{\sum_{-m}^{-1} \overline{b_i} 2^i + 2^{-m}}_{Two's\ complement\ of\ N^*} \quad (3.2.8)$$

- Representation of zero:

  adopted only clean zero: (3.2.9)

  $$0^+ \rightarrow \overset{b_s}{0}\underbrace{00\ldots0}_{n}.\underbrace{00\ldots0}_{m}$$

- *Property*:

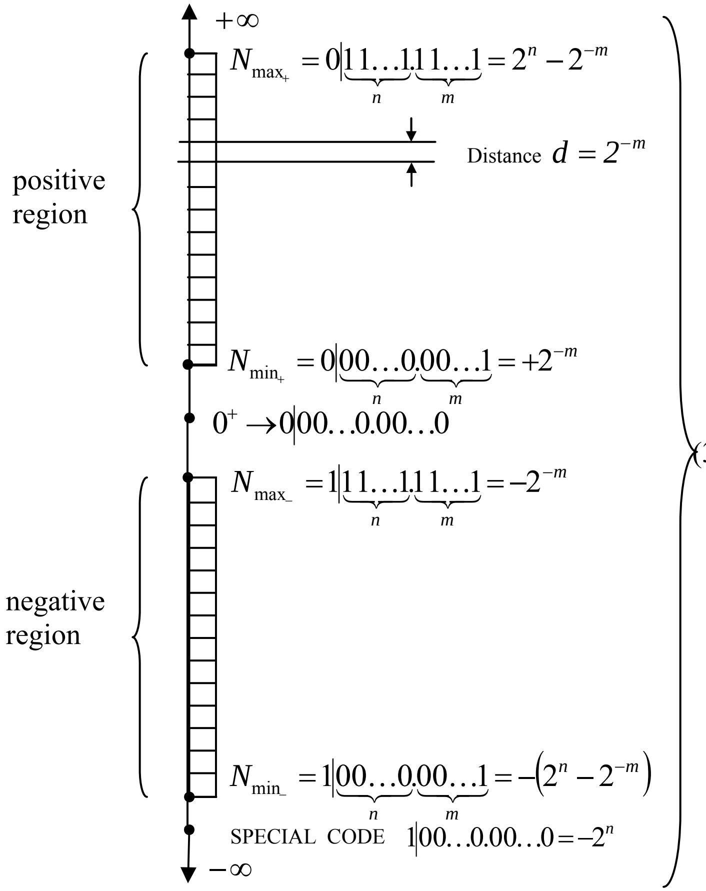
  By considering the negative weight for b_s it is obtained the effective value of the number *N*. (3.2.10)

- *Proof*:

  $$N > 0 \rightarrow N = 0 \cdot \left(-2^n\right) + \sum_{-m}^{n-1} b_i 2^i = +\sum_{-m}^{n-1} b_i 2^i \quad (3.2.11)$$

  $$\begin{aligned} N < 0 \rightarrow N &= 1 \cdot \left(-2^n\right) + \sum_{-m}^{n-1} \overline{b_i} 2^i + 2^{-m} = \\ &= 1 \cdot \left(-2^n\right) + \sum_{-m}^{n-1} (1 - b_i) 2^i + 2^{-m} = \\ &= -2^n + \sum_{-m}^{n-1} 2^i - \sum_{-m}^{n-1} b_i 2^i + 2^{-m} = \\ &= -2^n + \left(2^n - 2^{-m}\right) - \sum_{-m}^{n-1} b_i 2^i + 2^{-m} = \\ &= -\sum_{-m}^{n-1} b_i 2^i \end{aligned} \quad (3.2.12)$$

- Range of representation:



(3.2.13)

- The combination $1|\underbrace{00\ldots\ldots0}_{n}.\underbrace{00\ldots\ldots0}_{m}$ is used

as a particular value giving the minimal negative number. (3.2.14)

2.3.3 Variant 2

- $b_s =$ sign bit with the weight δ_s
- $|N| = \sum_{-m}^{n-1} b_i 2^i \rightarrow b_{n-1} \ldots b_0 b_{-1} \ldots b_{-m}$ (3.3.1)
- δ_s **has positive weight** $+2^n$ (3.3.2)
- General representation of $\pm N$

$$N = b_s\left(2^n\right) + \sum_{i=-m}^{n-1} b_i 2^i = b_s\left(2^n\right) + N^* \quad (3.3.3)$$

- if $N > 0 \rightarrow b_s = 1$ and the magnitude N^* is represented in direct code (3.3.4)
- if $N < 0 \rightarrow b_s = 0$ and the magnitude N^* is represented in two's complement code (3.3.5)

- Therefore,
  - $+|N| \rightarrow 1 \times 2^n + \sum_{-m}^{n-1} b_i 2^i$
  - $-|N| \rightarrow \underbrace{0 \times 2^n + \sum_{-m}^{n-1} \overline{b_i} 2^i + 2^{-m}}_{Two's\ complement}$

(3.3.6)

- In case of fractional numbers, n=0:
  - $+|N| \rightarrow 1 \cdot 2^0 + \sum_{-m}^{-1} b_i 2^i$
  - $-|N| \rightarrow \underbrace{0 \cdot 2^0 + \sum_{-m}^{-1} \overline{b_i} 2^i + 2^{-m}}_{Two's\ complement}$

(3.3.7)

- To obtain the effective value of the number, an *additive correction* is required by subtracting the constant $\delta_s = 2^n$. (3.3.8)

- if $N > 0$ then

$$N = 1\cdot 2^n + \sum_{-m}^{n-1} b_i 2^i - \left(2^n\right) = +\sum_{-m}^{n-1} b_i 2^i \qquad (3.3.9)$$

- if $N < 0$ then

$$\begin{aligned}
N &= 0\cdot 2^n + \sum_{-m}^{n-1} \overline{b_i} 2^i + 2^{-m} - \left(2^n\right) = \\
&= \sum_{-m}^{n-1} (1-b_i) 2^i + 2^{-m} - 2^n = \\
&= \sum_{-m}^{n-1} 2^i - \sum_{-m}^{n-1} b_i 2^i + 2^{-m} - 2^n = \\
&= 2^n - 2^{-m} - \sum_{-m}^{n-1} b_i 2^i + 2^{-m} - 2^n = \\
&= -\sum_{-m}^{n-1} b_i 2^i
\end{aligned} \qquad (3.3.10)$$

2.3.4 Variant 3

- Two sign bits designated b'_s and b_s with the following weights: (3.4.1)
  - Weight of $b_s = \delta_s$
  - Weight of $b'_s = 2\cdot\delta_s$ (3.4.2)
- General representation for $N(n,m)$:

$$N = b'_s b_s N^* = b'_s \cdot 2^{n+1} + b_s \cdot 2^n + N^* \qquad (3.4.3)$$

- if $N > 0 \rightarrow$ b'_s=1, b_s=0, N^* is represented in direct code
  - $N = 1\cdot 2^{n+1} + 0\cdot 2^n + \sum_{-m}^{n-1} b_i 2^i$ (3.4.4)
- if $N < 0 \rightarrow$ b'_s=0, b_s=1, N^* is represented in two's complement code, $\overline{N^*}$
  - $N = 0\cdot 2^{n+1} + 1\cdot 2^n + \underbrace{\sum_{-m}^{n-1} \overline{b_i} 2^i + 2^{-m}}_{Two's\ complement}$ (3.4.5)
- In case of fractional numbers, $n = 0$:

$$|N| = \sum_{-m}^{-1} b_i 2^i$$

  - $+|N| = 1\cdot 2^1 + 0\cdot 2^0 + \sum_{-m}^{-1} b_i 2^i$
  - $-|N| = 0\cdot 2^1 + 1\cdot 2^0 + \underbrace{\sum_{-m}^{-1} \overline{b_i} 2^i + 2^{-m}}_{Two's\ complement}$ (3.4.6)

- To obtain the effective value of N it must be subtracted the constant 2^{n+1} representing the additive correction($\mathscr{C}$): (3.4.7)

- $N > 0 \rightarrow 1\cdot 2^{n+1} + 0\cdot 2^n + \sum_{-m}^{n-1} b_i 2^i - \left(2^{n+1}\right) = +\sum_{-m}^{n-1} b_i 2^i$ (3.4.8)

- $N < 0 \rightarrow \; N = 0 \cdot 2^{n+1} + 1 \cdot 2^{n} + \sum_{-m}^{n-1} \overline{b_i} 2^{i} + 2^{-m} - \left(2^{n+1}\right)$

$$\left.\begin{aligned}
& N = 0 \cdot 2^{n+1} + 1 \cdot 2^{n} + \sum_{-m}^{n-1} \overline{b_i} 2^{i} + 2^{-m} - \left(2^{n+1}\right) \\
& = 1 \cdot 2^{n} + \sum_{-m}^{n-1} (1 - b_i) 2^{i} + 2^{-m} - 2^{n+1} = \\
& = 1 \cdot 2^{n} + \sum_{-m}^{n-1} 2^{i} - \sum_{-m}^{n-1} b_i 2^{i} + 2^{-m} - 2^{n+1} = \\
& = 1 \cdot 2^{n} + 2^{n} - 2^{-m} - \sum_{-m}^{n-1} b_i 2^{i} + 2^{-m} - 2^{n+1} = \\
& = 2 \cdot 2^{n} - \sum_{-m}^{n-1} b_i 2^{i} - 2^{n+1} = \\
& = 2^{n+1} - \sum_{-m}^{n-1} b_i 2^{i} - 2^{n+1} = \\
& = -\sum_{-m}^{n-1} b_i 2^{i}
\end{aligned}\right\} \quad (3.4.9)$$

- *Examples*:

Given $N_2(3,3)$:

$|N| = 101.001$

Variant 1 of representation

$+|N| = 0|101.001$

$-|N| = 1|010.111$

Variant 2 of representation

$+|N| = 1|101.001$

$-|N| = 0|010.111$

Variant 3 of representation

$+|N| = 10|101.001$

$-|N| = 01|010.111$

(3.4.10)

§ 2.4. One's Complement Representation.

2.4.1. General considerations

- *N(n,m)* $\Rightarrow$ $\overline{\overline{N}} = 2^n - N - 2^{-m} = (2^n - 2^{-m}) - N$
- Additional weighted sign bit (bits)
- Advantages, drawbacks. In case of negative numbers it is used inverse code (diminished radix code).
- Three variants: Variant 1, Variant 2, Variant 3.

(4.1.1)

2.4.2. Variant 1

- *N(n,m)*
- $|N| = b_{n-1}....b_1 b_0.b_{-1}....b_{-m} = \sum_{-m}^{n-1} b_i \cdot 2^i$
- $N = \underbrace{b_S}_{\text{sign bit}} \; | \underbrace{b^*_{n-1}....b^*_1 b^*_0.b^*_{-1}....b^*_{-m}}_{\text{magnitude bits}}$

  where the sign bit has the weight δ_S (4.2.1)

- If $N > 0 \;\; \rightarrow \;\; b_S = 0, \;\; N^* = \sum_{-m}^{n-1} b_i 2^i$

  $N = 0 \,|\, b_{n-1}......b_1 b_0.b_{-1}........b_{-m}$ (4.2.3)

- If $N < 0 \rightarrow b_S = 1, \;\; N^* = \sum_{-m}^{n-1} \overline{b_i} \cdot 2^i$ (1's complement)

  where $\overline{b_i} = 1 - b_i$

  Then, $N = 1 \,|\, \overline{b}_{n-1}.........\overline{b}_1 \overline{b}_0.\overline{b}_{-1}........\overline{b}_{-m}$ (4.2.4)

- The **weight of** b_S **is negative,** $\delta_S = (-2^n + 2^{-m})$ (4.2.5)

- *Example:*

  $|N| = 101.1101$

  $+|N| = 0 \,|\, 101.1101$

  $-|N| = 1 \,|\, 010.0010$ (4.2.6)

- If $n = 0$, fractional numbers with b_S in position 0 with $\delta_S = -(2^0 - 2^{-m})$ (4.2.7)
- $|N| = .b_{-1}b_{-2}.....b_{-m} = \sum_{-m}^{-1} b_i \cdot 2^i$ (4.2.8)
  - $+|N| = 0.b_{-1}b_{-2}.......b_{-m} = 0 \cdot 2^0 + \sum_{i=-m}^{-1} b_i \cdot 2^i$ (4.2.9)
  - $-|N| = 1.b_{-1}b_{-2}.......b_{-m} = 1 \cdot 2^0 + \sum_{i=-m}^{-1} \overline{b_i} \cdot 2^i$ (4.2.10)
- Two representations for 0:
  - Positive zero: $0 \,|\, 00.....0.00....0$ (4.2.11)
  - Negative zero: $\underbrace{1}_{\text{sign bit}} | \; \underbrace{11.....1}_{n}.\underbrace{1......1}_{m}$ (4.2.12)
- *Property*:

  By adopting the negative weight for b_S it is obtained the effective value of the number *N*: (4.2.13)

- a) If $N > 0 \;\; \rightarrow \;\; b_S = 0, \;\; N^* = \sum_{-m}^{n-1} b_i 2^i$

  $N = 0 \cdot 2^n + \sum_{-m}^{n-1} b_i \cdot 2^i = + \sum_{-m}^{n-1} b_i \cdot 2^i$ (4.2.14)

- b) If $N < 0 \;\; \rightarrow \;\; b_S = 1, N^* = \sum_{-m}^{n-1} \overline{b_i} \cdot 2^i$

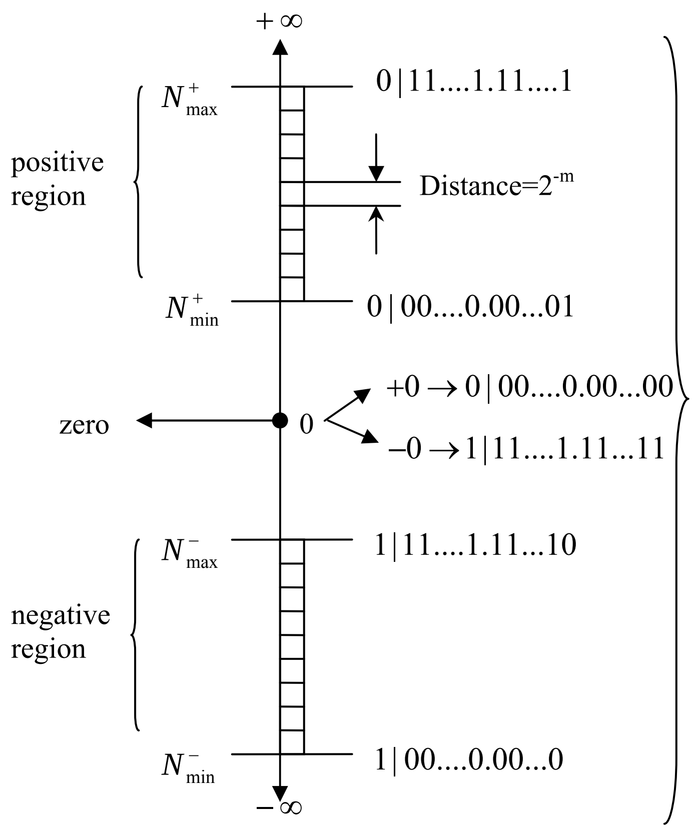
$$\left.\begin{aligned}
& N = 1(-2^n + 2^{-m}) + \sum_{-m}^{n-1} \overline{b_i} \cdot 2^i = \\
& = -2^n + 2^{-m} + \sum_{-m}^{n-1} (1 - b_i) \cdot 2^i = \\
& = -2^n + 2^{-m} + \sum_{-m}^{n-1} 2^i - \sum_{-m}^{n-1} b_i \cdot 2^i = \\
& = -2^n + 2^{-m} + 2^n - 2^{-m} - \sum_{-m}^{n-1} b_i \cdot 2^i = \\
& = -\sum_{-m}^{n-1} b_i \cdot 2^i
\end{aligned}\right\} \quad (4.2.15)$$

- Range of representation:

Extreme values:

$$\left.\begin{aligned}
& N^+_{\max} = 0 \,|\, \underbrace{11...1}_{n}.\underbrace{11...1}_{m} = 2^n - 2^{-m} \\
& N^+_{\min} = 0 \,|\, 00....0.0..01 = +2^{-m} \\
& N^-_{\max} = 1 \,|\, 1...11.1...10 = -2^{-m} \\
& N^-_{\min} = 1 \,|\, 0...00.0..00 = -(2^n - 2^{-m})
\end{aligned}\right\} \quad (4.2.16)$$

- Representation on the real axis



(4.2.17)

2.4.3. Variant 2

- *N(n,m)*

$$|N| = b_{n-1}......b_1 b_0.b_{-1}........b_{-m} = \sum_{-m}^{n-1} b_i \cdot 2^i \quad (4.3.1)$$

- $N = \underbrace{b_s}_{sign\ bit} \mid \underbrace{b^*_{n-1}\ldots\ldots b^*_1 b^*_0 b^*_{-1}\ldots\ldots b^*_{-m}}_{magnitude\ bits}$

b_s = **the sign bit with positive weight** $\delta_S = +(2^n - 2^{-m})$ (4.3.2)

- If $N > 0 \rightarrow b_S = 1,\ \ N^* = \sum_{-m}^{n-1} b_i \cdot 2^i$

$N = 1 \mid b_{n-1}.....b_1 b_0.b_{-1}.....b_{-m}$ (4.3.3)

Hence, $b_i^* = b_i \quad i \in [-m,(n-1)]$

- If $N < 0 \rightarrow b_S = 0,\ \ N^* = \sum_{-m}^{n-1} \overline{b_i} \cdot 2^i$ (1's complement)

$\overline{b_i} = 1 - b_i$ (4.3.4)

$N = 0 \mid \overline{b}_{n-1}.......\overline{b}_1\overline{b}_0.\overline{b}_{-1}.......\overline{b}_{-m}$

Hence, $b_i^* = \overline{b}_i \quad i \in [-m,(n-1)]$

- *Example*:

$|N| = 101.1101$

$+|N| = 1 \mid 101.1101$ (4.3.5)

$-|N| = 0 \mid 010.0010$

- Case of fractional numbers

$$n = 0 \rightarrow |N| = \sum_{-m}^{-1} b_i \cdot 2^i \quad (4.3.6)$$

- If $N > 0 \rightarrow b_s = 1$

and $N^* = \sum_{-m}^{-1} b_i \cdot 2^i$ (4.3.7)

- If $N < 0 \rightarrow b_s = 0$

and $N^* = \sum_{-m}^{-1} \overline{b_i} \cdot 2^i$ (one's complement) (4.3.8)

where $\overline{b_i} = 1 - b_i$

- To obtain the effective value of the number an *additive correction (* $\mathscr{C}$*)* is required:

$\mathscr{C} = -(2^n - 2^{-m})$ (4.3.9)

$N_{ef} = b_S(2^n - 2^{-m}) + N^* + \mathscr{C}$

- a) If $N > 0 \rightarrow b_S = 1$ and $N^* = \sum_{-m}^{n-1} b_i \cdot 2^i$

$N_{ef} = 1\cdot(2^n - 2^{-m}) + \sum_{-m}^{n-1} b_i \cdot 2^i + (-2^n + 2^{-m}) =$ (4.3.10)

$= 2^n - 2^{-m} + \sum_{-m}^{n-1} b_i \cdot 2^i - 2^n + 2^{-m} = +\sum_{-m}^{n-1} b_i \cdot 2^i$

- b) If $N < 0 \rightarrow b_S = 0$ and $N^* = \sum_{-m}^{n-1} \overline{b_i} \cdot 2^i$,

where $\overline{b_i} = 1 - b_i$

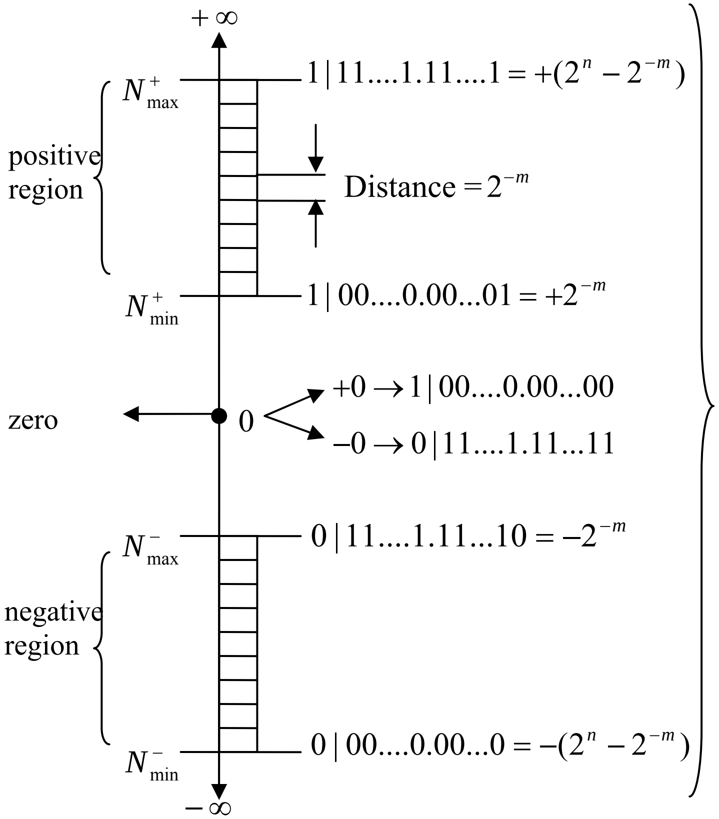
$N_{ef} = 0\cdot(2^n - 2^{-m}) + \sum_{-m}^{n-1} \overline{b_i} \cdot 2^i + \mathscr{C} =$

$= 0 + \sum_{-m}^{n-1} (1 - b_i) \cdot 2^i - 2^n + 2^{-m} =$ (4.3.11)

$= 0 + \sum_{-m}^{n-1} 2^i - \sum_{-m}^{n-1} b_i \cdot 2^i - 2^n + 2^{-m} =$

$= 0 + (2^n - 2^{-m}) - \sum_{-m}^{n-1} b_i \cdot 2^i - 2^n + 2^{-m} = -\sum_{-m}^{n-1} b_i \cdot 2^i$

- Range of representation:



(4.3.12)

2.4.4. Variant 3

- Two weighted sign bits: b'_S and b_S with weights δ'_s and δ_s (4.4.1)
- If $N>0 \to b'_S=1,\ b_S=0$ and $N^*=\sum_{-m}^{n-1} b_i \cdot 2^i$ (4.4.2)

  If $N<0 \to b'_S=0,\ b_S=1$

  and $N^*=\sum_{-m}^{n-1} \overline{b_i} \cdot 2^i$ (one's complement),

  where $\overline{b_i}=1-b_i \quad (i \in [-m,(n-1)])$ (4.4.3)

- The **weights of sign bits**:

  $\delta_S=2^n$

  $\delta'_S=2^{n+1}-2^{-m}$ (4.4.4)

- *Example*:

  $|N|=110.011$

  $+|N|=10|110.011$

  $-|N|=01|001.100$ (4.4.5)

- For fractional numbers *n*=0:

  $|N|=\sum_{-m}^{-1} b_i \cdot 2^i, \quad N=b'_S b_S\,|\,b_{-1}b_{-2}........b_{-m}$ (4.4.6)

- $+|N|=10\,|\,b_{-1}b_{-2}...b_{-m}$
- $-|N|=01\,|\,\overline{b_{-1}}\,\overline{b_{-2}}....\overline{b_{-m}}$ (4.4.7)

- To obtain the effective value of the number an *additive correction* ($\mathscr{C}$) is required:

$$\left.\begin{aligned} &\mathscr{C} = -(2^{n+1} - 2^{-m}) \\ &N_{ef} = b'_S(2^{n+1} - 2^{-m}) + b_S \cdot 2^n + N^* + \mathscr{C} \end{aligned}\right\} \quad (4.4.8)$$

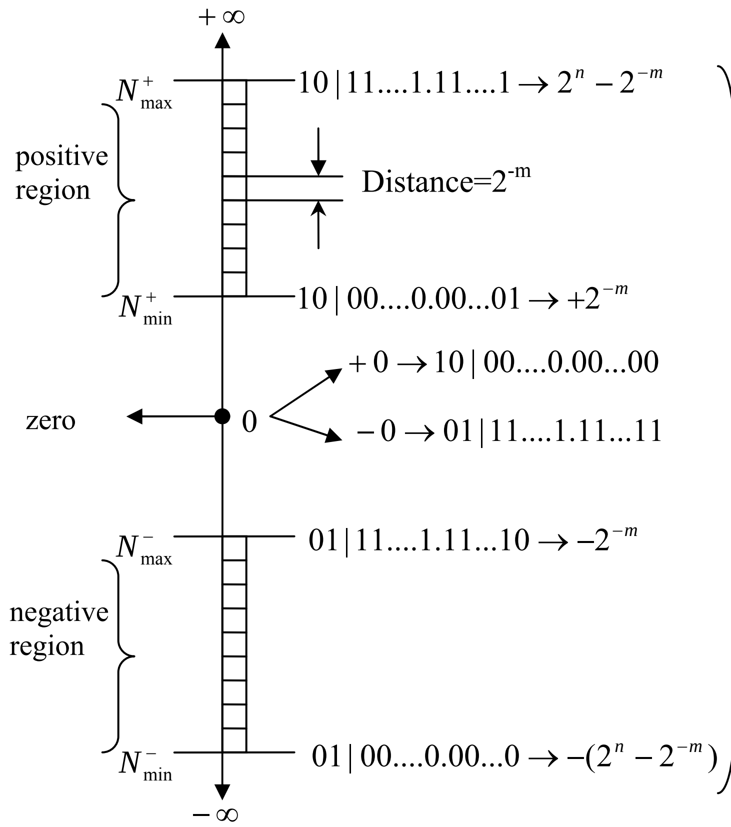
- If $N > 0$:

$$\left.\begin{aligned} N_{ef} &= 1 \cdot (2^{n+1} - 2^{-m}) + 0 \cdot 2^n + \sum_{-m}^{n-1} b_i \cdot 2^i + \mathscr{C} \\ &= 2^{n+1} - 2^{-m} + 0 + \sum_{-m}^{n-1} b_i \cdot 2^i - 2^{n+1} + 2^{-m} = \\ &= +\sum_{-m}^{n-1} b_i \cdot 2^i \end{aligned}\right\} \quad (4.4.9)$$

- If $N < 0$

$$\left.\begin{aligned} N_{ef} &= 0 \cdot (2^{n+1} - 2^{-m}) + 1 \cdot 2^n + \sum_{-m}^{n-1} \overline{b_i} \cdot 2^i + \mathscr{C} = \\ &= 1 \cdot 2^n + \sum_{-m}^{n-1} (1 - b_i) \cdot 2^i - 2^{n+1} + 2^{-m} = \\ &= 1 \cdot 2^n + 2^n - 2^{-m} - 2^{n+1} + 2^{-m} - \sum_{-m}^{n-1} b_i \cdot 2^i = \\ &= 2 \cdot 2^n - 2^{n+1} - \sum_{-m}^{n-1} b_i \cdot 2^i = \\ &= 2^{n+1} - 2^{n+1} - \sum_{-m}^{n-1} b_i \cdot 2^i = \\ &= -\sum_{-m}^{n-1} b_i \cdot 2^i \end{aligned}\right\} \quad (4.4.10)$$

- Range of representation:



(4.4.11)

§2.5. Shifting of signed binary numbers

- In general:

$$\left.\begin{aligned} &N(n,m)\\ &r = radix\\ &N_r = d_{n-1}....d_1 d_0.d_{-1}......d_{-m} =\\ &\quad = d_{n-1}\cdot r^{n-1} + ... + d_1\cdot r^1 + d_0 r^0 + d_{-1}\cdot r^{-1} + ...\\ &\qquad + d_{-m}\cdot r^{-m}\end{aligned}\right\} (5.1)$$

- *Multiplying by r:*

$$\left.\begin{aligned} &r\cdot N_r = r(d_{n-1}\cdot r^{n-1} + ... + d_0\cdot r^0 + d_{-1}\cdot r^{-1} + ...\\ &\qquad + d_{-m}\cdot r^{-m}) =\\ &\quad = d_{n-1}\cdot r^{n} + ... + d_0\cdot r^1 + d_{-1}\cdot r^0 + d_{-2}\cdot r^{-1} + ...\\ &\qquad + d_{-m}\cdot r^{-m+1}\end{aligned}\right\} (5.2)$$

The binary point between positions r^0 and r^{-1} is placed between digits d_{-1} and d_{-2}, hence the number was **shifted to the left** with one position. } (5.3)

- *Dividing by r:*

$$\left.\begin{aligned} &\frac{N_r}{r} = d_{n-1}\cdot r^{n-2} + d_{n-2}\cdot r^{n-3} + ... + d_1\cdot r^0 + d_0\cdot r^{-1} +\\ &\qquad + d_{-1}\cdot r^{-2} + ... + d_{-m}\cdot r^{-m-1}\end{aligned}\right\} (5.4)$$

The binary point was shifted between digits d_1 and d_2, so that the number was **shifted one position to the right**. } (5.5)

- If $r = 2$, then after a multiplication by 2 a *left* shifting with one position is derived, while after a division by 2 a *right* shifting with one position is derived. } (5.6)
- By extending to r^k:

$r^k\cdot N_r \rightarrow$ left shifting of N_r with k positions } (5.7)

$\frac{N_r}{r^k} \rightarrow$ right shifting of N_r with k positions } (5.8)

- If $r = 2,\ N_2(n,m)$

$$N_2 = b_{n-1}\cdot 2^{n-1} + ... + b_0\cdot 2^0 + b_{-1}\cdot 2^{-1} + ... + b_{-m}\cdot 2^{-m} \quad\} (5.9)$$

- After a multiplication by 2^k:

$$\left.\begin{aligned} &N_2\cdot 2^k = b_{n-1}\cdot 2^{n-1}\cdot 2^k + ... + b_0\cdot 2^k + b_{-1}\cdot 2^{-1}\cdot 2^k + ...\\ &\qquad + b_{-m}\cdot 2^{-m+k} =\\ &\qquad = b_{n-1}\cdot 2^{n+k-1} + .. + b_0\cdot 2^k + .. + b_{-k}\cdot 2^0 + b_{-k-1}\cdot 2^{-1}\\ &\qquad\quad + .. + b_{-m}\cdot 2^{-m+k}\end{aligned}\right\} (5.10)$$

A left shifting of N_2 with k positions was derived. } (5.11)

- After a division by 2^k:

$$\left.\begin{aligned} &\frac{N_2}{2^k} = \frac{b_{n-1}\cdot 2^{n-1} + ...b_0\cdot 2^0 + b_{-1}\cdot 2^{-1} + ... + b_{-m}\cdot 2^{-m}}{2^k} =\\ &\quad = b_{n-1}\cdot 2^{n-k-1} + ... + b_0\cdot 2^{-k} + b_{-1}\cdot 2^{-k-1} + ...\\ &\quad + b_{-m}\cdot 2^{-m-k} =\\ &\quad = b_{n-1}\cdot 2^{n-k-1} + ... + b_k\cdot 2^0 + b_{k-1}\cdot 2^{-1} + ... + b_0\cdot 2^{-k} + ...\\ &\quad + b_{-m}\cdot 2^{-m-k}\end{aligned}\right\} (5.12)$$

A right shifting of N_2 with k positions was derived. } (5.13)

- In case of *signed binary numbers* there are formulated the following 3 questions:
  1. What happens to the sign bit in connection with shifting.
  2. The direction of shifting.
  3. The values of bits that are inputted after shifting.

} (5.14)

- The method of realizing the shift depends on the *adopted negative number representation code.*
- For $N > 0$ the representations are identical

A) Multiplication by 2

I. b_S remains unchanged
II. one left shifting occurs
III. the new inputted bit is 0 } (5.15)

B) Division by 2

I. b_S remains unchanged
II. a right shifting occurs
III. the new inputted bit is 0 } (5.16)

- If $N < 0$

I. For Sign Magnitude representation

A) Multiplication by 2

I. b_S remains unchanged
II. a left shifting occurs
III. the new inputted bit is 0 } (5.17)

B) Division by 2

I. b_s remains unchanged
II. a right shifting occurs
III. the new inputted bit is 0 } (5.18)

II. For two's complement representation

A) Multiplication by 2

I. b_S remains unchanged
II. a left shifting occurs
III. the new inputted bit is 0 } (5.19)

B) Division by 2

I. b_S remains unchanged
II. a right shifting occurs
III. the new inputted bit is 1 } (5.20)

III. For one's complement representation

A) Multiplication by 2

I. b_s remains unchanged;
II. a left shifting occurs;
III. the new inputted bit is 1; } (5.21)

B) Division by 2

I. b_s remains unchanged;
II. a left shifting occurs;
III. the new inputted bit is 1; } (5.22)

- *Examples*

a) For positive numbers:

$$|N| = 101.11$$

$$+N \rightarrow 0\,|\,101.11$$

$$2 \cdot N \rightarrow 0\,|\,(1)011.1\underline{0}$$

$$2^2 \cdot N \rightarrow 0\,|\,(10)111.\underline{00}$$

$$\frac{N}{2} \rightarrow 0\,|\,\underline{0}10.11(1)$$

$$\frac{N}{2^2} \rightarrow 0\,|\,\underline{00}1.01(11)$$

(5.23)

Observation:
(b)=the lost bit
$\underline{b}$= the new bit } (5.24)

b) For negative numbers:

1) Sign-Magnitude Code

$$\left.\begin{aligned}
&|N| = 101.11 \\
&-N = 1\,|\,101.11 \\
&-2N = 1\,|\,(1)011.1\underline{0} \\
&-2^2 N = 1\,|\,(10)111.\underline{0}\underline{0} \\
&-\frac{N}{2} = 1\,|\,\underline{0}10.11(1) \\
&-\frac{N}{2^2} = 1\,|\,\underline{0}\underline{0}1.01(11)
\end{aligned}\right\} \quad (5.25)$$

2) Two's Complement Code

$$\left.\begin{aligned}
&-N = 1\,|\,010.01 \\
&-2N = 1\,|\,(0)100.1\underline{0} \\
&-2^2 N = 1\,|\,(01)001.\underline{0}\underline{0} \\
&-\frac{N}{2} = 1\,|\,\underline{1}01.00(1) \\
&-\frac{N}{2^2} = 1\,|\,\underline{1}\underline{1}0.10(01)
\end{aligned}\right\} \quad (5.26)$$

3) One's Complement Code

$$\left.\begin{aligned}
&-N = 1\,|\,010.00 \\
&-2N = 1\,|\,(0)100.0\underline{1} \\
&-2^2 N = 1\,|\,(01)000.\underline{1}\underline{1} \\
&-\frac{N}{2} = 1\,|\,\underline{1}01.00(0) \\
&-\frac{N}{2^2} = 1\,|\,\underline{1}\underline{1}0.10(00)
\end{aligned}\right\} \quad (5.27)$$