

* Integral promotion:- In any exp., all char and short int values are elevated to int. This is called as integral promotion.

Type conversion - Example:

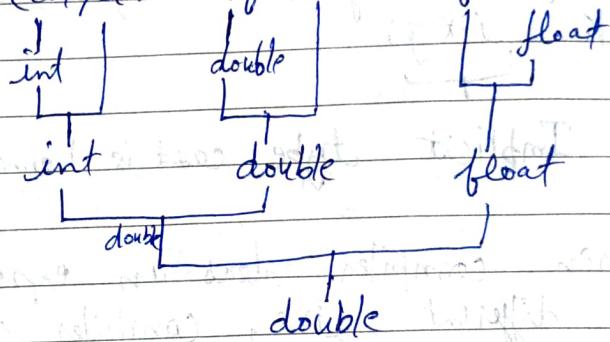
char ch;

int i;

float f;

double d;

$$\text{result} = (\text{ch}/\text{i}) + (\text{f} \times \text{d}) - (\text{f} + \text{i});$$



Example:-

unsigned int i = 2000000000, j = 8;

unsigned long int k;

$k = (\text{unsigned long int})(i * j)$; // Here j is implicitly type casted to i.

- 'i' is explicitly type casted into long int.

- Here i and j are not physically changed

- j is converted or explicit type casted only for that expression

$$k = (\text{unsigned long int}) \underbrace{(i * j)}_T$$

So type casting is here the result is truncated not works here ingrouping. this point only

classmate
Date _____
Page _____

Signed is treated as unsigned in type casting.

Session :- 11

→ Operators

→ Relational operators ($>$ $=<$ \leq \geq \neq)

Logical or Boolean operators (AND OR NOT)

Relational operators :-

- All the relational operators are binary operators
 - Required two operands to perform the task
- All the relational operators gives the result $0/1$. Zero and 1 (one)
- All these operators will not modify the operands.

\square	$>$	\square	greater than
\square	$<$	\square	less than
\square	\geq	\square	
\square	\leq	\square	
\square	\equiv	\square	
\square	\neq	\square	

- The operands can be variables/constants

Example :-

```
#include <stdio.h>
Void main()
{
```

```
int i=10, j=20, k;
```

$k = i > j$; // 10 > 20 the result is FALSE so '0'
 $\text{printf}("i=%d \ j=%d \ K=%d\n", i, j, k);$

O/P:- i=10 j=20 K=0

- Arithmetic operators have more priority than the relational operators.

→ void main()

{

int i=10, j=20, k;

K = i > j - 30;

printf("i=%d j=%d K=%d\n", i, j, k);

}

O/p:- i=10 j=20 K=1

→ i=10, j=20, k;

K = (i > j) - 30; // (10 > 20) - 30

// 0 - 30 = -30

O/p K = -30

→ int i=10, j=30, k;

K = i == j;

Here == is higher priority

O/p K=0 // because i==j is false, so 0

→ Consider the same input

i == j = k

for this statement, 'lvalue error'.
because i == j is constant. (i.e. 0)

Example:- Expressions:-

int i=10; j=20, k;

K = i > j; // 0

K = i > j + 20; // 0

K = i > (j + 20) // 0

K = (i > j) + 20 // 20

K = i == j; // 0

K == i = 10; // lvalue error

$k = i > j;$ // 1

→ Example:-

{
→ int $i = -1, j = 1;$
printf("%d\n", i > j); // 0 ($-1 > 1$) = 0
}

{
→ int $i = -1;$ // signed i > unsigned j
unsigned int $j = 1;$ // compiler will treat this signed
printf("%d\n", i > j); as unsigned
}

Note:- Type promotion :- In any expression, if one operand is signed and another operand is unsigned, signed is treated as unsigned.

Signed i > Unsigned j // Compiler treated
46 > j signed i as unsigned;

because, in signed the binary value of -1 is

(\downarrow 111111 | 1111111) | 1111111 | 1111111) = i
signbit Date bits

But compiler considers this as unsigned
so all one extra bits
- i.e. value of i is 46

⇒ 46 > 1 is checked

Logical operators:-

& Logical AND

|| Logical OR

! Logical NOT

- &&, || are binary operators (2 operands)
 - ! is unary operator
 - All these logical operators will not modify the operands i.e we can supply a constant or variable.
- All these logical operators will give us the result '0' or '1'.

		AND	OR
a	b	a&b	a b
0	0	0	0
0	1	0	1
1	0	0	1
1	1	1	1

0 (zero), 1 (means non-zero)

Ex:- void main()

{

int i=10, j=20, k;

K = i && j;

printf("i=%d j=%d k=%d", i, j, k);

}

- If the first operand is '0' the compiler will not check the second operand for Logical AND. It will decide the result zero.

$i = 10, j = 20$

$$\begin{aligned} k &= i \& 44 \quad (j = 200) \\ &= 0 \quad 44 \quad (\text{exp}) \\ \Rightarrow k &= 0 \end{aligned}$$

But if $i = 10 ; j = 20$

$$\begin{aligned} K &= i \& 44 \quad (j = 200) \\ &= 10 \& 44 \quad 200 \quad (\text{exp}) \\ &= 10 \& 44 \quad 200 \quad k = 1 \end{aligned}$$

O/P $i = 10, j = 20, K = 1;$

→ void main()

```
{  
    int i = 10;  
    !i;  
    printf("i = %d\n", i);  
}
```

Here the logical NOT will not change the operand
Consider:-

```
int i = 10;  
i = !i;  
printf("i = %d\n", i);
```

O/P = $i = 0$

Complimented and NOT are different

↓ ↓
Bitwise (n) !

→ Logical AND has the higher priority than
Logical OR

→ Priority

- (i) NOT
- (ii) Logical AND
- (iii) Logical OR

Session: 12

Operators :-

- Logical (or) Boolean operators (`&& || !`)
- Bitwise operators (`& | ^ ~ << >>`)

Consider,

$$m = i + j - k * l;$$

From a user's view,

$$m = [i + j] - [k * l]$$

from a compiler's view

Depending upon the priority, the compiler will group it implicitly

$$m = i + j - (k * l);$$

$$m = i + j - \text{exp.}$$

$$m = \text{exp} - \text{exp.}$$

$$m = \text{exp2}$$

when a bigger exp. to the compiler, it will compress the expression to the smallest, depending upon the priority, until it becomes one op. Then compiler start solving it.

→ Expressions:-

- int $i=10, j=20, k; // o/p K=0$

$$K = i \& (j = 200)$$

- $i=0, j=20, k=30, l; // o/p l=1$

$$l = i \& (j = 200) \& (k = 300);$$

- $l = i \& (j = 200) \& (k = 300) \& l; // o/p l=0$

- $l = i \& (j = 200) \& (k = 300) \& l \& l; // o/p l=1$

o/p $i=10, j=200, k=30, l=1$

Bitwise operators:-

& → Bitwise AND

| → Bitwise OR

^ → Bitwise XOR

~ → Bitwise NOT
(compliment)

>> → shift right

<< → shift left

P	q	p&q	p q	p^q	~p	up
0	0	0	0	0	1	1
0	1	0	1	1	1	0
1	0	0	1	1	0	1
1	1	1	1	0	0	0

→ Except bitwise compliment, remaining operators are binary.

- All these operators will not modify the operands, so a constant or variable can be supplied as the operands.
- These bitwise operators are not applicable to real numbers. (float and double).
- These operands should be converted into binary to decide the result.

Example:-

→ void main()

{

int i=10, j=15, k;

k=i&j; // 'i' & 'j' are converted into binary
bitwise & is used.

printf("k=%d", k);

}

Explanation:- i = 10 i.e. $\begin{array}{r} 0000 \\ \times 10 \\ \hline 0000 \end{array} \rightarrow 10$

j = 15 i.e. $\begin{array}{r} 0000 \\ \times 1111 \\ \hline 0000 \end{array} \rightarrow 15$

4 $\begin{array}{r} 0000 \\ \times 10101 \\ \hline 0000 \end{array}$

result of bitwise AND $10 \& 15 = 10$

K = 10

$$\rightarrow \text{If } k \neq i l j; \quad \begin{array}{r} 0000 \\ 0000 \\ \hline 0000 \end{array} \rightarrow 16$$

$\frac{0000}{0000} \overline{1111} \rightarrow 15$

$K = 15$

$$\rightarrow K = i^1 j; \quad \begin{array}{r} 0000 \\ 0000 \\ \hline 0000 \end{array} \begin{array}{l} 1010 \\ 1111 \\ 0101 \end{array} \rightarrow \begin{array}{l} 10 \\ 15 \\ 5 \end{array}$$

$\rightarrow K = \text{int}$ // as the given data type is int
(4 bytes = 32 bits), for its compliment,
entire 32 bits have to be considered

00000000 | 00000000 | 00000000 | 0000 | 010 → 1.
 1111111 1111111 1111111 111 010 → 11.
 9 2's complement +

After taking Compliment
Here sign bit is 1 → so it is a -ve no.

→ To decide which one is the again
Q's compliment.

$$\begin{array}{r}
 111111111 | 11111111 | 11111111 | 11110101 \\
 00000000. 00000000 00000000 00001010 \\
 \hline
 + 1 \\
 \hline
 1011 \\
 8421 \\
 \hline
 -11
 \end{array}$$

→ Both `!` and `n` are unary operators
Here `Not` operator inverts the zero/non zero
directly

Where as

Compliment (\sim) operator inverts the value of the operand in bitwise.

`int j=10, j;`

$$j = \sim i$$

$$j = -11$$

$$\begin{array}{l} j = \sim i \\ j = -11 \end{array}$$

Examples:-

→ By converting
i & j into
binary *

- `int i=100, j=150, k;`
- $k = i \& j; // K=4$
- $K = i | j; // K=246$
- $K = i ^ j; // K=742$
- $K = \sim i \& j // K=-101$

Session 13 :-

30-6-90

→ Operators

- Bitwise operators ($& | ^ \sim << >>$)
- Swapping of 2 numbers using bitwise operators
- How to set a bit in a given no.
- How to clear a bit in a given no.

→ Swapping of two no. using bitwise operators:

{

$$a = 10, b = 15$$

$$a = a \& b; // a=5 \quad \begin{array}{r} 0000 \\ 1010 \end{array} \rightarrow 10$$

$$b = a \& b; // b=10 \quad \begin{array}{r} 0000 \\ 1111 \end{array} \rightarrow 15$$

$$a = a \& b; // a=0000 \quad \begin{array}{r} 0000 \\ 0101 \end{array} \rightarrow 5 \quad \text{XOR} \rightarrow 5$$

}

$$b = \underline{\underline{0000}} \quad \begin{array}{r} 1111 \end{array} \rightarrow 15$$

$$b = \underline{\underline{0000}} \quad \begin{array}{r} 1010 \end{array} \rightarrow 10 \quad \text{XOR} \rightarrow 10$$

$$a = \underline{\underline{0000}} \quad \begin{array}{r} 0101 \end{array} \rightarrow 5 \quad \text{XOR} \rightarrow 5$$

$$b = \underline{\underline{0000}} \quad \begin{array}{r} 1111 \end{array} \rightarrow 15 \quad \text{XOR} \rightarrow 15$$

before swap

$$a = 10, b = 15$$

After swap

$$a = 15, b = 10$$

→ Shift operators:

- ($<<$) left shift
- ($>>$) Right shift.

- These are binary operators
- These operators will not modify the operands
- The operand can be constant/variable.

Ex:- Left shift

int i=5, j=2, k;

K=i<< j;

here, it means i is left shifted by j times

i << j
the operand no. of times to be shifted 1st operand is shifted by 2nd operand times

Note:- In shift operators (both left and right) the 1st operand is converted into binary and operand can not be converted into binary

Ex:- int i=5, j=2, k;

K=&i << j;

i = 5

1010101010101010

2 times shift so

This zero is so

out when

shift

00010100 = 20

K = 20 Ans

Ex:- int i=5, j=1, k
 $k = i \ll j;$

$$\begin{array}{r} 0000\ 0101 \rightarrow 5 \\ 0000\ 1010 \rightarrow 10 \\ \hline \end{array}$$

Rule \rightarrow [Result = Number $\times 2^{\text{shift}}$]
 Left shift.

~~* Note:- In shift operators, the 2nd operand should not be negative numbers. The no. should be 0-31 while dealing with int datatype.
 → But if give int number no error will come but unexpected value comes.~~

→ And even if the 2nd operand exceeds the limit (e.g. 0 to 31) the result is unspecified.

The 1st operand may be -ve or +ve no.

Ex:-

int i=-10, j=1, k;
 $K = i \ll j;$
 $K = -10 \ll 1;$

Now take binary of +10

00000000/00000000/00000000/00001010

Taking 2's compliment for -10

11111111/11111111/11111111/11110101

+1

11111111/11111111/11111111/11110110

This no. will left shifted 1

11111111/11111111/11111111/11101100

Left shifted

To find what value it is again 2's complement

$$\begin{array}{r}
 00000000 | 00000000 | 00000000 | 000 | 00 \\
 + \\
 00000000 | 00000000 | 00000000 | 000 | 00
 \end{array}$$

$$\Rightarrow k = -20$$

→ Right shift operator ($>>$)

$$i = 10, j = 1, k$$

$$k = i >> j; \quad // 10 is right shifted by 1 times$$

$$K = 10 >> 1;$$

$i = 1000001010$ thrown out (gone of memory)

$$\begin{matrix} \text{New} \\ \text{zero added} \end{matrix} [k = 5]$$

Rule:-

Number
2 shift

$$\text{int } = i = -1, j = 1, k; \\ k = i >> j;$$

→ -ve no. given
so sign bit
copy mechanism
used.

Note:- Sign bit copy mechanism used here.

~~From old~~ $\begin{array}{r} 11111111 | 11111111 | 11111111 | 11111111 \\ 11111111 | 11111111 | 11111111 | 11111111 \end{array}$ thrown now

sign bit copied again after shift.

This is called sign bit copy mechanism

Note:- In right shift case, though the sign bit is shifted, it is copied to that shifted bit position.

[K = -1]

Here generally

int i=0, j=1, k;

i = i-1;

i = -1

unsigned int i=0, j=1, k;

i = i-1; j = 0-i;

$\Rightarrow [i = 4h]$

printf("%d", i);

0 \leftrightarrow 4h

after 0, it again goes
to 4h, so the value

of i = 4h

\Rightarrow all 32 bits are '1'

All bits are data bits so

after shift

→ 0 111111 11111111 11111111 11111111 /

New zero added

Here the value is 2h

The op is

j = -1, j = 1 & K = 2148748391

-1 because %d is given to take the opp

→ But in signed data type sign bit is copied again. In case of -ve we give

Session 14

How to set a number using bitwise operator.

We need to take 2 inputs from the user

- ① Number
- ② position

1 means set

0 means clear

Set = num = num | 1 << pos;

num = 10 pos = 2

1	0	0	0	0	1	0	1	0
0	0	0	0	1	1	1	0	

↑
bit is set
Ans = 110

<< left shift have higher priority

→ num = num | 1 << pos;
 num = num | Result
 num = Result

Note:- Position should not be +ve numbers.

→ Now, num = 10, pos = 2
 10 binary → 00001010
 As per expression [num = num | 1 << pos]
 num = 10 | 1 << 2

take 1 << 2 as it has higher priority

000001010
 Now num = 10 | ¹ result.
 00001010
 $\underline{000000100}$ OR
 $\underline{00001110}$ 110 Ans.

→ If Num = 15, pos = 2
 [Nothing change already set]

00001111
 00000100 OR
 $\underline{00011111}$ 1111

→ How to clear a bit using bitwise operator

Clear:- $\text{num} = \text{number} \& (1 \ll \text{pos})$

$\text{num} = \text{num} \& \text{~}(1 \ll \text{pos})$

$\text{num} = \text{num} \& \text{~}(\text{result})$

$\text{num} = \text{num} \& \text{~}(\text{result})$

$\text{num} = \text{result}$

$\text{num} = \text{result}$

Consider:- $\text{num} = 10, \text{pos} = 3$

0	0	0	0	1	0	1	0
---	---	---	---	---	---	---	---

 $\rightarrow 10 \cdot (\text{pos})$

0	0	0	0	1	0	0	0
---	---	---	---	---	---	---	---

 $1 \ll 3$

1	1	1	1	0	1	1	1
---	---	---	---	---	---	---	---

 $\sim (1 \ll 3)$

0	0	0	0	1	0	1	0
---	---	---	---	---	---	---	---

 $\&$ $\text{num} = 10$

1	1	1	1	0	1	1	1
---	---	---	---	---	---	---	---

 $\text{AND} \& \sim (1 \ll 3)$

0	0	1	0	0	0	1	0
---	---	---	---	---	---	---	---

 $= \boxed{\text{Num} = 2 \text{ Ans}}$

→ Complementing the bits (Toggle)

This formula toggle 1 bit

Toggle - $\text{num} = \text{num} ^ 1 \ll \text{pos};$

$\text{num} = 10, \text{pos} = 2;$

0	0	0	0	1	0	1	0
---	---	---	---	---	---	---	---

 $\rightarrow 10$

0	0	0	0	0	1	0	0
---	---	---	---	---	---	---	---

 $\rightarrow (1 \ll 2)$

0	0	0	0	1	1	1	0
---	---	---	---	---	---	---	---

 $\text{num} ^ 1 \ll 2$

Toggled successfully

Set \rightarrow num = num | 1 << pos;

Clear \rightarrow num = num & ~ (1 << pos);

Complement \rightarrow num = num ^ 1 << pos;

→ Sizeof() operator :-

- It is a unary operator sizeof()
- The operand can be constant/ variable/ data type
- This operator gives the size in the form of bytes.

Ex:- void main() // a=97 in ASCII
{

char ch = 'a';

printf("%ld %ld %ld\n", sizeof(char), sizeof(ch),
sizeof('a'));

O/P:-

1 4

→ The sizeof() character will give 4 bytes because compiler consider int constant.

→ The size of "a" constant is 4 bytes
the compiler will treat 97 as an integer.

Ex:- void main() -- 1
{

int i=10;

printf("%ld %ld %ld\n", sizeof(int), sizeof(i), sizeof(10));

O/P:- 4 4 4

→ Example:- void main()

{

 float f = 23.5;

 printf("%ld %ld %ld\n", sizeof(f), sizeof(4), sizeof(23.5));

O/P:- 4 4 8

A 23.5 compiler treats as double.

Voidmain()

{

 printf("%ld\n", sizeof("abcd"));

O/P:- 5 // string ("")

Expressions should not be passed in a sizeof() operator.

If written, the expression is not going to be solved.

{ int i=10, j=20, k;
 printf("K=%d\n", k);
 sizeof(k=i+j);
 printf("K=%d\n", k);

}

O/P:- k=0
 K=0

Session 15:-

→ Conditional operators (. ? :) / Ternary

→ How to check in a given number given bit position set or clear.

→ Introduction to IEEE 754 real number representation

→ Ternary / Conditional Operator :-

- It requires 3 operands to perform the task.
- The symbol for conditional operator is
 $(\quad ? \quad : \quad)$
- Syntax :- $\text{exp1 ? exp2 : exp3}$
- Compiler will solve the first expression.
- Depends upon the first expression result may be the 2nd operand or 3rd operand.
- If exp1 is non zero then result becomes exp2 , if exp1 is zero the result becomes exp3 .

Example:-

void main ()

{

int i=10, j;

j = i ? 100 : 200;

printf("i=%d j=%d\n", i, j);

}

1/ O/P :-

i=10 j=100

j=100 because the

1st operand is

non zero.

→ void main ()

{ int i=0, j;

j = i ? 100 : 200;

printf(" %d\n", j);

O/P :- j= 100

Here j=200 bcoz

1st operand is zero

Here 1st operand is variable

We can take variable, expression, constant etc.

Eg:- k = i ? j : 100 : 200; // i=10, j=20

{ printf("%d", k); }

O/P :- K=200.

i < i?j=False

- Write a program to check whether the given no is positive or negative.

int num;

$=$

```
num >= 0 ? printf(" +ve \n") : printf(" -ve \n");
```

- Odd / even program :-

int num;

$=$

```
① num % 2 == 0 ? printf("Even \n") : printf("odd \n");
```

```
② num % 2 ? printf("odd \n") : printf("Even \n");
```

- Write a program to find the biggest of two no.:-

int a, b;

$=$

```
a > b ? printf("a is greater \n") : printf("b is greatest \n");
```

Note:- The conditional operator can be nested.

- Find the biggest of 3 numbers:-

int a, b, big;

```
big = a > b ? (a > c ? a : c) : (b > c ? b : c);
```

- If the 3rd operand is not grouped the compiler thrown 1 value required error.

\nwarrow correct.
(i >= 10) ? j = 20 : (j = 30)

- If not grouped compiler consider j as 3rd operand.

- How to check in a given number given bit is set or clear?

$$r = \underline{\text{num}} + \underline{1} \ll \underline{\text{pos}};$$

result
result

$$r = \underline{\text{num}} \gg \underline{\text{pos}} + \underline{1};$$

result
result

num	pos	$r = \text{num} + 1 \ll \text{pos}$	$r = \text{num} \gg \text{pos} + 1;$
15	0	1	1
15	1	2	1
15	2	4	1
15	3	8	1
15	4	0	1

int num, pos;

scanf - -

result 0/1

num >> pos + 1 ? printf("set\n") : printf("clear\n");
/num & 1 << pos ? printf("set\n") : printf("clear\n");

Y result zero/non zero

O/P:- Enter the number
15

Enter the pos
3

Set.

Session :- 16

→ Operators

- IEEE754 real number representation
- Inc and Dec operator.

- Size of float and double is not decided by the OS, it depends on the standard IEEE754.
- Size of the float is 4 bytes.
- Size of the double is 8 bytes.
- unsigned / signed float / double can't be declared. By default both are signed.
- There's no direct way to operate on the float by the mc/mp. So, a floating point unit [FPU] is used. This is a co-processor. [There are no assembly instructions to deal with the floating points].
- Few processors have inbuilt FPU
- Most of the FPU's internally follows the standard IEEE754.
- Ex:- void main()
 - {
 - ↓
 - float f = 22.5; ← Here double value is given
 - printf ("%f %e %g\n", f, f, f);
 - }
- %e prints the data in standard decimal exponent format.

O/P:- 22.750000 2.275000e+0 2.275

∴ When %f is used, it prints 6 digits after standard exponent format.
Standard exponent format → digit.xe^{±y}

- Double is treated as long float [half]

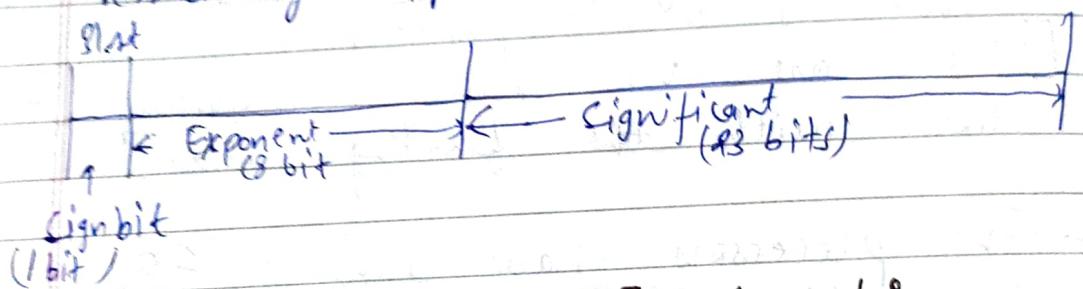
Representation in IEEE754 Standard :-

	Sign	Exponent	Significant (mantissa)
float 32 bits	1	8	23
double 64 bits	1	11	52

Now when float is declared

for

float f = 1 and float f = -1
only the sign bit is changed (varied)
remaining exponent & significant are same.



→ Steps to get IEEE754 format :-
(Real number)

① Convert the float into normal binary format.

e.g. $f = 23.5$

1st convert .5 and then convert 23

Ex: 7.7 binary
 $7 = 111$

$$\begin{array}{r} 1011.10 \\ \hline 2 | 11 \end{array} \quad \begin{array}{r} 2171 \\ 2 | 3 -1 \\ \hline 1 | -1 \end{array} \quad \boxed{11} = 7$$

Now $0.7 \times 2 = 1.4$

$$0.4 \times 2 = 0.8$$

$$0.8 \times 2 = 1.6$$

$$0.6 \times 2 = 1.2$$

$$0.2 \times 2 = 0.4$$

$$0.4 \times 2 = 0.8 \rightarrow \text{Here repeating from 0.8}$$

$7.7 \rightarrow 111.10110$ (bar is given over the nos because they are repeated)

- ③ Convert the binary into standard binary exponent format.

$$\text{Bit. } x^{e+y} \quad \text{where } x \rightarrow \text{significant}$$

Take float f = 23.5;
 $\Rightarrow 1.0111.10 \rightarrow 1.0\underset{\oplus 4}{111.0}e^{\oplus 4}$ → significant/mantissa

- ④ Add or subtract exponent part with standard bias number (IEEE number)

For float 7F for double 3FF

$$(7F \rightarrow 127) \text{ Hexa binary.} \quad (3FF \rightarrow 1023)$$

Here, addition or subtraction depends on the exponential sign

Now,

$$1.01110e^{+4}, \quad 7F \rightarrow 127_{10}$$

$$\begin{array}{r} 127 \\ - 1023 \\ \hline 124 \end{array}$$

(decimal)

- ④ Exponent part 8 bits:
 10000011

$$10000011 \quad \begin{matrix} 8 \\ 7 \end{matrix}$$

- ⑤ Significant bits are 92 here $1.01110e^{+4}$
(Fill the significant part from left to Right)

- ⑥ Sign bit → 0 (bcz. the no +ve)

$$\begin{array}{ccccccccc} 0 & \underline{10000011} & 01100 & 0000000 & 0000000 \\ \downarrow & (8) & & & & (8) \\ \text{Sign}^+ & \text{exponent} & & & & \text{significant} \end{array}$$

→ Example:- Consider

$$\text{float} = -7.7;$$

① Converting -7.7 into binary

$$\rightarrow 7.7_{10} = 111.1011_0$$

② Standard binary format

$$111.1011_0 \Rightarrow 1.111011_0 e^{+2}$$

$$③ 7F + 2 \Rightarrow 81 \text{ (Hexa)}$$

④ 81 into binary

$$\begin{array}{r} 1000 \\ 8 \overline{) 0001} \\ \hline 1 \end{array} \quad (8 \text{ bits})$$

⑤ Significant part (23 part)

$$111\ 0110\ 0110\ 0110\ 0110 \text{ (23 bits)}$$

⑥ Sign bit (1) given no -ve

$$\text{float_b} = -7.7.$$

$$\begin{array}{c|c|c|c} 1 & 0000001 & 111011001100110011001100 & \\ \hline \text{Sign} & \text{exponent} & \text{significant part} & \downarrow \end{array}$$

→ Conversion of double into IEEE 754 format.

① double d = 23.5;

② Convert the no. into its binary
 $\Rightarrow 10111.10$

③ Convert the binary formats into standard
 binary exponent format
 $1.01110e^{+4}$

④ Bias + exponent

$$\Rightarrow 2FF + 4 = 402$$

(Hexa)

⑤ Convert the exponent part into 11 bits

$$\begin{array}{r} 100 \\ 4 \overline{) 0000} \\ \hline 0 \end{array} \quad \begin{array}{r} 0011 \\ 3 \end{array}$$

⑥ Significant part (52 bits)