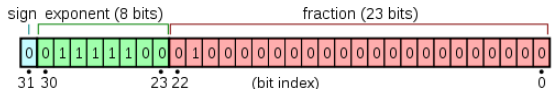
## Cấu trúc

* Để biểu diễn một số không nguyên chúng ta có nhiều cách để thể hiện như:
* Fixed-point
* Logarithmic number systems (LNSs)
* Interval arithmetic
* Đến năm 1985, tính toán dấu phẩy động được thành lập bởi Institute of Electrical and Electronics Engineers (IEEE), tiêu chuẩn này giải quyết được nhiều vấn đề bất cập trước đó.
* Trong tiêu chuẩn IEEE 754, các số được chia làm hai nhóm: đ**ộ chính xác đơn(single precision)** và **độ chính xác kép (double precision).**

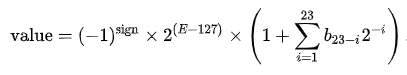
### Cấu trúc của số độ chính xác đơn



* Cấu trúc của số không nguyên với độ chính xác đơn được xác định bởi dãy 32 bit. Trong đó từ trái qua bit đầu tiên sẽ được thể hiện là bit dấu với giá trị 1 là số âm – 0 là số dương, kế tiếp sau đó là 8 bit cho phần xác định số mũ, 23 bit tiếp theo được sử dụng cho phần định trị.
* Giá trị của phần mũ(exponent) để biểu diễn được phần mũ âm và mũ dương người ta qui định giá trị thực tế của phần mũ sẽ được tính với công thức như sau:

Giá trị số mũ thực tế = E - 127

* Giá trị thể hiện của một số không nguyên theo tiêu chuẩn IEE 754 với độ chính xác đơn từ một chuỗi 32 bit sẽ được tính như sau:



Trong đó:

sign: giá trị của phần mũ (1 hoặc 0)

E: giá trị phần mũ

b23-i: là giá trị ở các vị trí trong phần định trị

* Ví dụ:

Cho dãy 32 bit: 1 1000 0001 011 0000 0000 0000 0000 0000

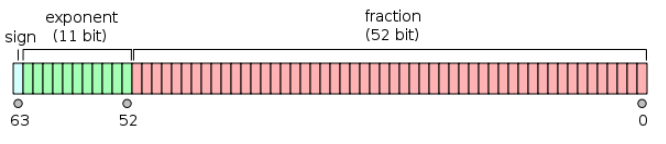
* + - Sign: 1
    - Exponent: 10000001
    - Fraction(mantissa): 011 0000 0000 0000 0000 0000

value = (-1)1 x 2(129-127) x (1+ 1\*2-2 + 1\*2-3)

= (-1) x 4 x (1.375) = 5.5D

* Dựa vào cấu trúc ở trên, ta thấy với độ chính xác đơn ta có 23 bit để thể hiện giá trị của một sô nguyên tức số lượng phần tử nguyên có thể chứa được là **223– 1**  số. nhưng theo cup cấp từ trang của Microsoft thì khoảng giá trị lớn nhất mà số float thể hiện được là 3.402823466.E+38 = 3402823466. E+29 lớn hơn rất nhiều so với số lượng số nguyên có thể chưa được. Từ đó có một lỗi mà mỗi kiểu thể hiện theo dấu chấm động điều mắc phải là tính tràn nhảy cóc giữa các số liên típ nhau. Tức giá trị nguyên sẽ không được lưu tất cả với các số liên tiếp nhau chênh nhau 1 đơn vị. Mà nó sẽ có thể chênh nhau vài đơn vị đến vài trăm vài nghìn đơn vị. Chinh tiết sẽ được thể hiện rõ ở những phần sau.
* Dựa vào phần tích trên thì giá trị các số từ 223 – 1 sẽ được tính toán và thể hiện đúng giá trị của chuỗi số.

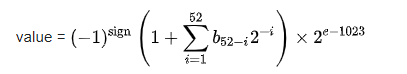
### Cấu trúc của số độ chính xác kép



* Tương tự với độ chính xác đơn, ở độ chính xác kép chúng ta có số lượng tăng lên gấp đôi với tổng số bit trong chuỗi là 64.
* Trong đó bit đầu tiên từ trái sẽ là bit dấu với 1 là số âm, 0 là số dương, 11 bít liền kề kế tiếp sẽ là số bít thể hiện giá trị phần mũ, 53 bit liền sau đó số bit cho phần định trị.
* Giá trị cho phần mũ thật sự được tính theo công thức:

Giá trị số mũ thực tế = E - 1023

* Giá trị thể hiện của một số không nguyên theo tiêu chuẩn IEE 754 với độ chính xác kép từ một chuỗi 64 bit sẽ được tính như sau:



Trong đó:

sign: giá trị của phần mũ (1 hoặc 0)

E: giá trị phần mũ

b23-i: là giá trị ở các vị trí trong phần định trị

* Ví dụ:

Cho dãy 64 bit: 0 10000000001 0100000000000000000000000000000000000000000000000000

* + - Sign: 0
    - Exponent: 10000000001
    - Fraction: 0100000000000000000000000000000000000000000000000000

value = (-1)0 x 2(1025-1023) x (1+ 1\*2-2)

= (1) x 4 x (1.25) = 5D

* Cũng tương tự như số với độ chính xác đơn, với cách biểu diễn với độ chính xác kép giá trị trong miền an toàn để có thể thể hiện được giá trị chính xác với cận trên là

**253– 1.**

### Cách chuyển đổi một số không nguyên từ một số hệ thập sang kiểu thể hiện dấu chấm động với độ chính xác đơn

* Giả định số không nguyên với định dạng : xxx.yyy
* Với xxx là phần nguyên, yyy là phần thập phân
* **Bước 1**: Xác định phần nguyên và phần thập phân
* **Bước 2**: Chuyển đổi phần nguyên và phần thập phân sang chuỗi nhị phân
* **Bước 3**: Nối hai phần nguyên và phần nhị phân đã chuyển đổi qua chuỗi nhị phân lại với nhau với ngăn cách bằng dấu chấm.
* **Bước 4** : Xác định bit dấu (âm là 1,dương là 0)
* **Bước 5** : chuyển chuổi nhị phân ở bước 3 về định dạng **1.xxxx \* 2q**
* **Bước 6** : Xác định phần số trị exponent.
* **Bước 7** : Chuyển đổi exponent sang chuỗi nhị phân.
* **Bước 8** : đặt 3 thành phần sign, exponent, fraction vào đúng vị trí.

Ví dụ:

Cho số ở hệ thập phần là 85.125

**Bước 1:**

Phần nguyên : 85

Phần thập phân : 0.125

**Bước 2:**

Phần nguyên: 1010101

Phần thập phân: 001

**Bước 3:** 1010101.001

**Bước 4:** Do là số dương nên sign : 0

**Bước 5:** 1.010101 \* 26

**Bước 6:** exponent = 127 + 6 (vì 6 = expoment - 127)

**Bước 7:** exponent = 10000101

**Bước 8:** sign = 0, exponent = 10000101, fraction= 010101

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Biểu diễn nhị phần theo độ chính xác đơn của 85.125