ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΥΠΟΛΟΓΙΣΤΩΝ

ΑΡΙΘΜΟΙ ΚΙΝΗΤΗΣ ΥΠΟΔΙΑΣΤΟΛΗΣ

(ΠΡΑΓΜΑΤΙΚΟΙ ΑΡΙΘΜΟΙ)

Γ. Τσιατούχας

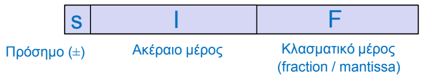
Παράρτημα Β

Διάρθρωση

1. Το σύστημα κινητής υποδιαστολής
2. Αναπαράσταση πραγματικών δυαδικών αριθμών
3. Το πρότυπο 754 της IEEE

Αριθμοί Σταθερής Υποδιαστολής

Στην αναπαράσταση σταθερής υποδιαστολής χρησιμοποιείται ένας συγκεκριμένος αριθμός από bits για το ακέραιο μέρος και ένας συγκεκριμένος αριθμός από bits για το κλασματικό μέρος.



Έστω ότι χρησιμοποιούμε 32 bits για την αναπαράσταση του πραγματικού αριθμού, από τα οποία 1 bit είναι το πρόσημο, 16 bits είναι το ακέραιο μέρος και 15 bits είναι το κλασματικό μέρος. Η υποδιαστολή υπονοείται.

Τότε η τιμή της ακόλουθης 32 bits λέξης υπολογίζεται ως:

$$\underset{\downarrow\,(-)}{1} \mid \underbrace{0000\ 0000\ 0011\ 0101}_{I\,=\,53_{10}} \mid \underbrace{0100\ 0000\ 0000\ 000}_{F\,=\,2^{-2}\,=\,0{,}25_{10}} = -\,53{,}25_{10}$$

Το Σύστημα Κινητής Υποδιαστολής

Η αναπαράσταση ενός πραγματικού αριθμού (n) με βάση το 10 στο σύστημα *κινητής υποδιαστολής* (*floating point*) βασίζεται στον επιστημονικό συμβολισμό ως ακολούθως:

$$n = f \times 10^e$$

όπου το f ονομάζεται *κλασματικό μέρος* (*fraction / mantissa*) και το e είναι ένας θετικός ή αρνητικός ακέραιος που ονομάζεται *εκθέτης* (*exponent*).

Στην αναπαράσταση χρησιμοποιείται μόνο ένα ψηφίο για το ακέραιο μέρος.

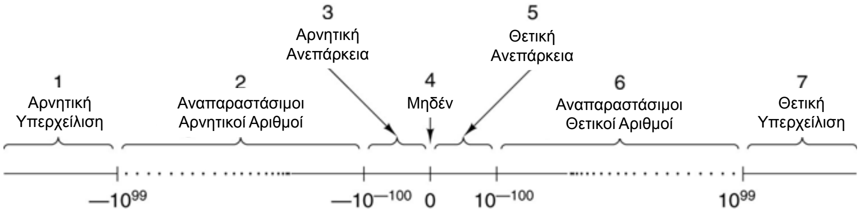
Παράδειγμα:

	κανονικοποιημένη μορφή	
3.14	$= 0.314 \times 10^1$	$= 3.14 \times 10^0$
0.000001	$= 0.1 \times 10^{-5}$	$= 1.0 \times 10^{-6}$
1941	$= 0.1941 \times 10^4$	$= 1.941 \times 10^3$

Ο επιστημονικός συμβολισμός δεν δίνει μια μοναδική αναπαράσταση.

Η αναπαράσταση γίνεται μοναδική αν δεχθούμε ως κανόνα το ακέραιο μέρος να είναι μηδέν και το πιο σημαντικό ψηφίο του κλασματικού μέρους να μην είναι μηδέν, οπότε έχουμε την *κανονικοποιημένη* (*normalized*) μορφή.

Αναπαράσταση Κινητής Υποδιαστολής



Έστω τριψήφιο κλασματικό μέρος f με: $0,1 \leq f \leq 1$ή f=0 και e = διψήφιος εκθέτης

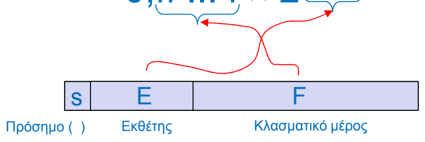
- Αριθμοί στις περιοχές 1, 3, 5, 7 δεν μπορούν να αναπαρασταθούν.
- Οι αριθμοί κινητής υποδιαστολής δεν αποτελούν ένα συνεχές σύνολο!
- Τα αποτελέσματα ενός υπολογισμού δεν μπορούν να αναπαρασταθούν ακριβώς. Λύση η *στρογγύλευση*.
- Η απόσταση μεταξύ διαδοχικών αριθμών που μπορούν να αναπαρασταθούν δεν είναι σταθερή. Όμως, το *σχετικό σφάλμα* που εισάγεται με τη στρογγύλευση είναι περίπου σταθερό.
- Η αύξηση των ψηφίων του f αυξάνει την ακρίβεια των προσεγγίσεων.
- Η αύξηση των ψηφίων του e αυξάνει το μέγεθος των περιοχών 2 & 6.

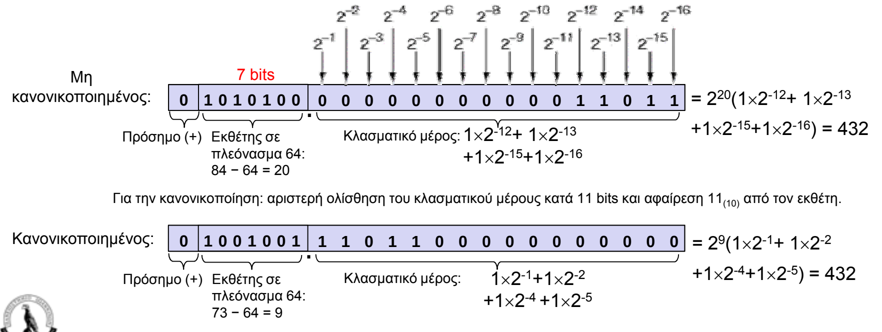
Αναπαράσταση Κινητής Υποδιαστολής και Δυαδικοί Αριθμοί

- Χρήση της κανονικοποιημένης επιστημονικής αναπαράστασης:

$$\pm\, \underbrace{0,ff \ldots f} \times 2^{\underbrace{ee \ldots e}} \qquad \text{όπου} \quad 0 \leq 0,ff \ldots f < 1$$

- Αποθήκευση:





$= 2^{20}(1\times2^{-12}+ 1\times2^{-13} +1\times2^{-15}+1\times2^{-16}) = 432$

Κλασματικό μέρος: $1\times2^{-12}+ 1\times2^{-13} +1\times2^{-15}+1\times2^{-16}$

Για την κανονικοποίηση: αριστερή ολίσθηση του κλασματικού μέρους κατά 11 bits και αφαίρεση $11_{(10)}$ από τον εκθέτη.

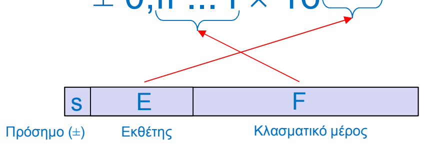
$= 2^{9}(1\times2^{-1}+ 1\times2^{-2} +1\times2^{-4}+1\times2^{-5}) = 432$

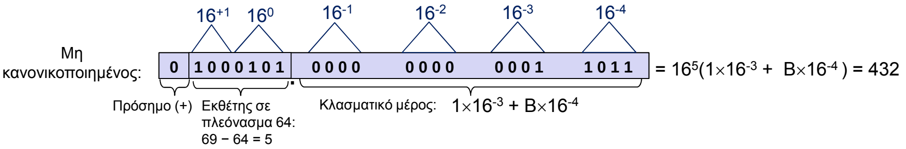
Κλασματικό μέρος: $1\times2^{-1}+1\times2^{-2} +1\times2^{-4} +1\times2^{-5}$

Αναπαράσταση Κινητής Υποδιαστολής με Βάση το 16

- Κανονικοποιημένη επιστημονική αναπαράσταση:

$$\pm\ \underbrace{0,ff\ ...\ f} \times 16^{\underbrace{ee...e}} \qquad \text{όπου} \quad 0 \le 0,ff\ ...\ f < 1$$

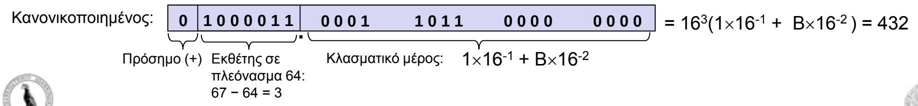




Μη κανονικοποιημένος: 0 | 1 0 0 0 1 0 1 | 0 0 0 0   0 0 0 0   0 0 0 1   1 0 1 1 $= 16^5(1\times16^{-3} + B\times16^{-4}) = 432$

Πρόσημο (+) | Εκθέτης σε πλεόνασμα 64: 69 − 64 = 5 | Κλασματικό μέρος: $1\times16^{-3} + B\times16^{-4}$

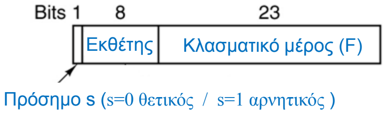
Για την κανονικοποίηση: αριστερή ολίσθηση του κλασματικού μέρους κατά 2 hex ψηφία και αφαίρεση $2_{(10)}$ από τον εκθέτη.



Κανονικοποιημένος: 0 | 1 0 0 0 0 1 1 | 0 0 0 1   1 0 1 1   0 0 0 0   0 0 0 0 $= 16^3(1\times16^{-1} + B\times16^{-2}) = 432$

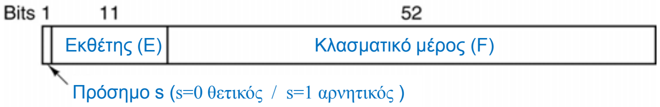
Πρόσημο (+) | Εκθέτης σε πλεόνασμα 64: 67 − 64 = 3 | Κλασματικό μέρος: $1\times16^{-1} + B\times16^{-2}$

Πρότυπο IEEE 754



Πρόσημο s (s=0 θετικός / s=1 αρνητικός)

Αναπαράσταση Απλής Ακρίβειας
(single precision – 32bit)



Πρόσημο s (s=0 θετικός / s=1 αρνητικός)

Αναπαράσταση
Διπλής Ακρίβειας
(double precision – 64bit)

- Η αριθμητική βάση είναι το 2 και χρησιμοποιείται συμβολισμός πλεονάσματος στον εκθέτη (κατά 127 και 1023 αντίστοιχα).
- Καθώς στην κανονικοποιημένη μορφή το αριστερότερο κλασματικό ψηφίο είναι πάντα 1 μπορεί να μεταφερθεί μια θέση αριστερότερα (στο ακέραιο μέρος και να παραληφθεί στην αναπαράσταση (υπονοείται).

Υπολογισμός αριθμού: $(-1)^s \times \overbrace{(1+F)}^{\text{σημαντικό μέρος (significant)}} \times 2^{E-\text{bias}}$

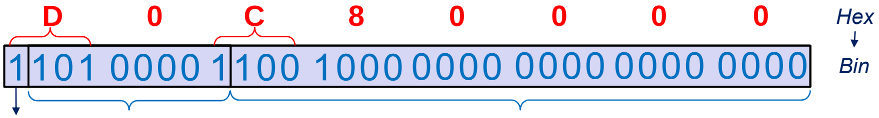
όπου bias = 127 στην αναπαράσταση απλής ακρίβειας και
1023 στην αναπαράσταση διπλής ακρίβειας

Παράδειγμα

Έστω η 32bit λέξη D0C80000$_{16}$ η οποία αντιπροσωπεύει αριθμό σε αναπαράσταση κινητής υποδιαστολής σύμφωνα με το πρότυπο IEEE απλής ακρίβειας. Ποιά είναι η τιμή στο δεκαδικό σύστημα;

⁂

Απάντηση:



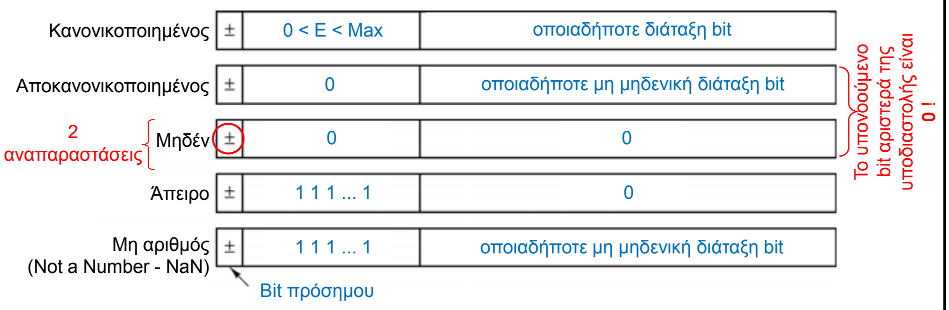
(−) E = 10100001$_2$ = 161$_{10}$

F = ,10010000000000000000000$_2$ = $2^{-1} + 2^{-4}$ = 0.5 + 0.0625 = 0,5625$_{10}$

Συνεπώς η τιμή του αριθμού είναι:

$$(-1)^1 \times (1+0,5625) \times 2^{161-127} = -\,1,5625 \times 2^{34} = -\,26.843.545.600_{10}$$

Αριθμητικοί Τύποι Πρότυπου IEEE



Κανονικοποιημένος	±	0 < E < Max	οποιαδήποτε διάταξη bit
Αποκανονικοποιημένος	±	0	οποιαδήποτε μη μηδενική διάταξη bit
Μηδέν	±	0	0
Άπειρο	±	1 1 1 ... 1	0
Μη αριθμός (Not a Number - NaN)	±	1 1 1 ... 1	οποιαδήποτε μη μηδενική διάταξη bit

Το υπονοούμενο bit αριστερά της υποδιαστολής είναι **0 !**

Με στόχο τη μείωση του προβλήματος ανεπάρκειας (underflow) η IEEE εισήγαγε τους αποκανονικοποιημένους (denormalized) αριθμούς. Σε αυτή την περίπτωση το υπονοούμενο bit αριστερά της υποδιαστολής από 1 γίνεται 0.

- Ο μικρότερος θετικός κανονικοποιημένος αριθμός απλής ακρίβειας έχει E=1 και F=0 και είναι ο $1,0\times2^{-126}$.
- Ο μεγαλύτερος θετικός αποκανονικοποιημένος αριθμός απλής ακρίβειας έχει E=0 και F=111...1 και είναι ο $0,9999999\times2^{-127}$.
- Ο μικρότερος θετικός αποκανονικοποιημένος αριθμός απλής ακρίβειας έχει E=0 και F=00...01 και είναι ο $2^{-23}\times2^{-127}= 2^{-150}$.

Χαρακτηριστικά του IEEE 754

Item	Single precision	Double precision
Bits in sign	1	1
Bits in exponent	8	11
Bits in fraction	23	52
Bits, total	32	64
Exponent system	Excess 127	Excess 1023
Exponent range	–126 to +127	–1022 to +1023
Smallest normalized number	2^{-126}	2^{-1022}
Largest normalized number	approx. 2^{128}	approx. 2^{1024}
Decimal range	approx. 10^{-38} to 10^{38}	approx. 10^{-308} to 10^{308}
Smallest denormalized number	approx. 10^{-45}	approx. 10^{-324}