10η Εβδομαδα - Τριτη 13 Δεκεμβριου 2005

ΣΤ. Εκτελεση πραξεων στον Η.Υ.

1. Εισαγωγή

Στο πλαίσιο της άποψης ότι ο Η.Υ είναι **«μετασχηματιστής πληροφοριών»** που **«επεξεργάζεται σύμβολα»**, κεντρική σπουδαιότητα κατέχει η εκτέλεση πράξεων για την επεξεργασία αριθμητικών λογικών και αλφαριθμητικών δεδομένων.

Στην παράγραφο αυτή θα ασχοληθούμε με την εκτέλεση αριθμητικών πράξεων. Τις πράξεις αυτές τις διακρίνουμε σε 3 κατηγορίες.

– **Πράξεις σταθερής υποδιαστολής** (fixed point arithmetic) στο δυαδικό σύστημα (binary arithmetic).

- **Πράξεις σταθερής υποδιαστολής στο δεκαδικό σύστημα** (decimal arithmetic) μεταξύ δυαδικών αριθμών που έχουν κωδικοποιηθεί σε δυαδική μορφή **με χρήση ενός κώδικα BCD**.

- **Πράξεις κινητής υποδιαστολής** (floating point arithmetic).

Οι **πράξεις σταθερής υποδιαστολής** χρησιμοποιούν αριθμητικά δεδομένα σταθερής υποδιαστολής **που είναι κατάλληλα για παραστάσεις μεταβλητών με απλή δομή** (που παρουσιάζονται σε προβλήματα εμπορικών εφαρμογών, στατιστικής φύσης κλπ.). Επιπλέον εάν **οι πράξεις γίνονται στο δυαδικό σύστημα** τότε οι αριθμοί σταθερής υποδιαστολής έχουν απλή εσωτερική μορφή και μήκος σταθερό. Τέλος, εάν οι πράξεις γίνονται **στο δεκαδικό σύστημα**, τότε οι αριθμοί σταθερής υποδιαστολής παρουσιάζονται σαν σειρά δεκαδικών ψηφίων, που το κάθε ένα έχει κωδικοποιηθεί σε δυαδική μορφή και **το μήκος τους είναι μεταβλητό και εξαρτάται από το εκάστοτε πλήθος των δεκαδικών ψηφίων τους**.

Οι **πράξεις κινητής υποδιαστολής** χρησιμοποιούν μόνο δυαδικά αριθμητικά δεδομένα και είναι κατάλληλες σε υπολογισμούς μαθηματικής φύσεως **με μεταβλητές που έχουν σύνθετη δομή**, καθόσον η κινητή υποδιαστολή απλοποιεί την εκτέλεση των πράξεων.

Παραδείγματα τρόπου παράστασης του κάθε ένα από τα παραπάνω είδη αριθμητικών δεδομένων σε διάφορους Η.Υ δίδει ο παρακάτω πίνακας 1.

Πίνακας 1. Περιγραφή αριθμητικών δεδομένων σε διάφορους Η.Υ.

Δεδομένα σταθερής υποδιαστολής	Μοντέλο Η.Υ. (τρόπος παράστασης)
α. **Δυαδικά** - Μήκος λέξης Η.Υ. (bits)	
0 | 1 ← Μέγεθος → 31	IBM-360 (πρόσημο, συμπλήρωμα 2)
0 | 1 ← Μέγεθος → 35	PDP-10 (πρόσημο, συμπλήρωμα 2)
0 | 1 ← Μέγεθος → 35	UNIVAC-1106 (πρόσημο, συμπλήρωμα 1)
0 | 1 ← Μέγεθος → 59	CDC-6600 (πρόσημο, συμπλήρωμα 1)
β. Δεκαδικά BCD (8, 4, 2, 1)	
0 0 1 0 | 0 1 0 0 (1 byte) 3 4	
0001 | 1001 | 1000 | 0101 (2 bytes) 1 9 8 5	
Δεδομένα κινητής υποδιαστολής	Η.Υ.
0 | 1 ← εκθέτης → 7 | 8 ← εκθέτης → 31	IBM-360 (Πρόσημο κλάσματος, εκθέτης πλεόνασμα των 64, κλάσμα συμπλήρωμα του 2)
0 | 1 ← εκθέτης → 8 | 9 ← εκθέτης → 35	PDP-10 (Πρόσημο κλάσματος, εκθέτης πλεόνασμα του 128, κλάσμα συμπλήρωμα του 2)
0 | 1 ← εκθέτης → 8 | 9 ← εκθέτης → 35	UNIVAC 1106 (Πρόσημο κλάσματος, εκθέτης πλεόνασμα του 128, κλάσμα συμπλήρωμα του 1)
0 | 1 ← εκθέτης → 11 | 12 ← εκθέτης → 59	CDC-6600 (Πρόσημο κλάσματος, εκθέτης πλεόνασμα του 1024, κλάσμα συμπλήρωμα του 1)

2. Συμπληρώματα αριθμών

Προτού εξετάσουμε την εκτέλεση των συγκεκριμένων πράξεων στις προηγούμενες περιπτώσεις, θα είναι χρήσιμο να δώσουμε μερικούς ορισμούς που θα διευκολύνουν αργότερα την παράσταση των προσημασμένων αριθμών.

α. Ορίζουμε συμπλήρωμα $\bar{\psi}_k$ **ενός ψηφίου** ψ_k ενός αριθμητικού συστήματος (Α.Σ.) με βάση β, **τη διαφορά του** ψ_k **από το μέγιστο ψηφίο του συστήματος·** δηλαδή έχουμε:

$$\bar{\psi}_k = (\beta - 1) - \psi_k. \tag{1}$$

Προφανώς, το συμπλήρωμα του συμπληρώματος του ψ_k θα είναι το ίδιο το ψ_k. Εξάλλου, στο δυαδικό σύστημα το συμπλήρωμα του 1 θα είναι το 0 και το συμπλήρωμα του 0 θα είναι το 1, αφού:

$$\bar{1} = (2-1) - 1 \quad \text{και} \quad \bar{0} = (2-1) - 0 = 1.$$

β. Συμπλήρωμα αριθμού ως προς βάση

Ορίζουμε συμπλήρωμα $\bar{A}_{<\beta>}$ **ενός αριθμού** $A_{<\beta>}$, ως προς μία βάση β, **που αποτελείται από μ ακέραια και ν κλασματικά ψηφία**, τον αριθμό:

$$\bar{A}_{<\beta>} = \beta^{\mu} - A_{<\beta>}. \tag{2}$$

Έτσι, το συμπλήρωμα του $A_{<10>} = 1985.15$ ως προς 10 είναι

$$\bar{A} = 10^4 - 1985.15 = 8014.85,$$

ενώ του $B_{<2>} = 110111.010$ είναι:

$$\bar{B} = 001000.110 (= 001000.101 + 0.001),$$

δηλαδή **συμπληρώνονται τα ψηφία, ένα προς ένα, και προστίθεται και μία μονάδα τελευταίας τάξεως**.

Σημείωση:

Για την εύρεση του συμπληρώματος ενός δυαδικού, γράφουμε όπου 1 τον αριθμό 0 και όπου 0 το 1, και προσθέτουμε και μια δυαδική μονάδα, της τελευταίας τάξεως, όπως δίδεται στην παρένθεση του προηγούμενου παραδείγματος.

γ. Συμπλήρωμα αριθμού ως προς μειωμένη βάση

Ορίζουμε συμπλήρωμα $\underline{A}_{<\beta>}$ ενός αριθμού $A_{<\beta>}$ ως προς την μειωμένη βάση β-1, τον αριθμό:

$$\underline{A}_{<\beta>} = \beta^{\mu} - A_{<\beta>} - \beta^{-\nu} \tag{3}$$

όπου μ το πλήθος των ακεραίων ψηφίων του και ν το πλήθος των κλασματικών ψηφίων του αριθμού Α.

Έτσι το συμπλήρωμα του $A_{<10>} = 1985.15$ ως προς 9 είναι:

$$\underline{A} = 10^4 - 1985.15 - 10^{-2} = 8014.84.$$

Ενώ το συμπλήρωμα του δυαδικού $B_{<2>} = 110111.010$ ως προς 1 είναι:

$$\underline{B}_{<2>} = 001000.101,$$

δηλαδή, **συμπληρώνονται τα ψηφία ένα προς ένα**.

Σημείωση:

Το συμπλήρωμα ενός δυαδικού αριθμού ως προς την μειωμένη βάση βρίσκεται εάν θέσουμε όπου 1 το 0 και όπου 0 το 1 στον αριθμό· δηλαδή, το **συμπλήρωμα ενός αριθμού ως προς μια βάση είναι το μεγαλύτερο από το συμπλήρωμα του αριθμού ως προς την μειωμένη βάση κατά μία μονάδα της τελευταίας τάξης του ($\bar{A}_{<\beta>} = \underline{A}_{<\beta>} + \beta^{-\nu}$).**

3. Παράσταση προσημασμένων δυαδικών αριθμών.

α. Παράσταση «Πρόσημο και μέγεθος» (Π&Μ)

Υπάρχουν πολλοί τρόποι παράστασης προσημασμένων δυαδικών αριθμών. Τον πρώτο τον έχουμε ήδη αναφέρει· Είναι η παράσταση «**πρόσημο και μέγεθος**» (Sign and Magnitude) όπου το πιο σημαντικό bit της λέξης του Η.Υ συμβολίζει το πρόσημο του (0 για θετικούς και 1 για αρνητικούς), ενώ τα υπόλοιπα bits συμβολίζουν την απόλυτη τιμή του αριθμού.

Έτσι, εάν πάρουμε ένα μικρουπολογιστή με λέξη 1 byte (8 bits) τότε η παράσταση των αριθμών 30 και -30 είναι

$$\begin{aligned} A_{<10>} &= 30 \xrightarrow{\text{Π \& Μ}} A_{<2>} = 00011110 \\ B_{<10>} &= -30 \xrightarrow{\text{Π \& Μ}} B_{<2>} = 10011110. \end{aligned}$$

Σημείωση:

Σε Η.Υ με λέξη μ bits ο ελάχιστος και μέγιστος αριθμός που μπορούν να παρασταθούν είναι: $2^{\mu-1} - 1$ και $-(2^{\mu-1} - 1)$ ενώ το πλήθος των αριθμών που δύνανται να παρασταθούν είναι 2^{μ}($2^{\mu-1} - 1$ θετικοί $2^{\mu-1} - 1$ αρνητικοί και οι **δύο** παραστάσεις του μηδενός 000...0 και 1000...0).

β. Παράσταση «Πρόσημο και Συμπλήρωμα του 2» (ΠΣ2)

Ένας άλλος τρόπος είναι το «**πρόσημο και συμπλήρωμα του** 2» (ΠΣ2), όπου και πάλι το πιό σημαντικό bit της λέξεως του Η. Υ συμβολίζει το πρόσημο και τα υπόλοιπα bits το μέγεθος του αριθμού όταν είναι θετικός· ενώ **όταν είναι αρνητικός ο αριθμός απομνημονεύεται ως συμπλήρωμα του 2 της απόλυτης τιμής του**, οπότε το πιο σημαντικό bit θα είναι ίσο με τη μονάδα (ένδειξη αρνητικού αριθμού). Έτσι για τους αριθμούς 30 και -30 θα έχουμε:

$$\begin{aligned} A_{<10>} &= 30 \xrightarrow{\Pi\Sigma 2} A_{<2>} = 00011110 \\ B_{<10>} &= -30 \xrightarrow{\Pi\Sigma 2} B_{<2>} = 11100010. \end{aligned}$$

Παρατήρηση:

Η παράσταση του -30 προκύπτει από την παράσταση του 30 με συμπλήρωση ως προς 2.

Τέλος, οι αληθινές τιμές των δυαδικών $\Delta_1 = 01101010$ και $\Delta_2 = 10101100$ που είναι γραμμένοι στο ΠΣ2 είναι $\Delta_1 = 106$ (ως θετικός) και $\Delta_2 = -(01010011+1) = -84$ (αφού το πιο σημαντικό του ψηφίο είναι 1, δηλαδή αρνητικός).

Σημείωση: Στην παράσταση ΠΣ2 για Η.Υ. με μ δυαδικά ψηφία, ο μέγιστος κι ο ελάχιστος αριθμός που μπορούν να παρασταθούν είναι αντίστοιχα οι αριθμοί: $2^{\mu-1} - 1$ και $-2^{\mu-1}$, ενώ **υπάρχει μια παράσταση του 0:** αυτή με όλα τα ψηφία της λέξης μηδενικά. Τέλος, μπορούν να παρασταθούν 2^{μ} διαφορετικοί αριθμοί, $2^{\mu-1}$ θετικοί (συμπεριλαμβάνοντας το 0) και $2^{\mu-1}$ αρνητικοί αριθμοί.

γ. Παράσταση «Πρόσημο και Συμπλήρωμα του 1» (ΠΣ1)

Τέλος, ένας τρίτος τρόπος παράστασης προσημασμένων δυαδικών αριθμών είναι **«πρόσημο και συμπλήρωμα του 1»** (ΠΣ1), που είναι παρόμοιος με τον ΠΣ2, πλην όμως **οι αρνητικοί βρίσκονται με συμπλήρωση ως προς 1, της απόλυτης τιμής τους**.

Στην παράσταση ΠΣ1 σε Η.Υ με μήκος λέξης μ ψηφία, έχουμε $2^{\mu-1}$ θετικούς αριθμούς (συμπεριλαμβάνεται και το 0) και $2^{\mu-1}$ αρνητικούς αριθμούς (συμπεριλαμβάνεται και το -0). Ας πάρουμε και πάλι ένα Η.Υ με λέξη 1 byte, τότε οι αριθμοί 30 και -30 θα παρίστανται στο ΠΣ1 ως εξής:

$$\begin{aligned} A_{<10>} &= 30 \xrightarrow{\Pi\Sigma 1} &= 00011110 \\ B_{<10>} &= -30 \xrightarrow{\Pi\Sigma 1} &= 11100010. \end{aligned}$$

Τέλος, εάν θέλουμε να βρούμε την αληθινή τιμή του δυαδικού $\underline{\underline{\Delta}} = 10000000$ που είναι γραμμένος στο ΠΣ1 θα έχουμε $\Delta = -01111111 = -127$. Όμοια για τον $\underline{\underline{\Delta}}_1 = 11111111$ θα έχουμε $\Delta_1 = -00000000$, δηλαδή -0.

Σημείωση:

Στην παράσταση ΠΣ1 σε Η.Υ με μήκος λέξης μ ψηφία, ο μέγιστος αριθμός είναι ο $2^{\mu-1}$ και ελάχιστος είναι ο -($2^{\mu-1}$-1). Οι λέξεις με όλα τα ψηφία 0 και όλα 1 δίνουν τις **δύο** εκφράσεις του μηδενός 0 και -0.

4. Γενικά για τις πράξεις

Απο τις 4 βασικές πράξεις **η πρόσθεση και η αφαίρεση είναι θεμελιώδεις**, αφού ο πολλαπλασιασμός των δυαδικών υλοποιείται με προσθέσεις και ολισθήσεις, η δε διαίρεση με προσθέσεις, αφαιρέσεις και ολισθήσεις. Η πρόσθεση και η αφαίρεση αποτελούν πράξεις εντελώς διαφορετικές μεταξύ τους και κατά συνέπεια η υλοποίηση τους στον Η.Υ. θα απαιτούσε δύο ξεχωριστές μονάδες. Πλην όμως **η χρήση των συμπληρωμάτων** των αριθμών, όπως θα δούμε παρακάτω, **μετατρέπει την αφαίρεση σε πρόσθεση**· έτσι όλες οι αριθμητικές πράξεις μπορούν να υλοποιηθούν με χρήση προσθέσεων συμπληρωμάτων και ολισθήσεων.

Τέλος, η χρήση συμπληρωμάτων ΠΣ1 και ΠΣ2 απαιτούν διαφορετικές διαδικασίες, έτσι σε κάθε Η.Υ. ενδείκνυται η χρήση μιάς μόνο από τις δύο παραστάσεις ΠΣ1 και ΠΣ2. **Η χρήση της ΠΣ1 είναι ευκολότερη παρά της ΠΣ2, ο αλγόριθμος όμως εκτέλεσης των πράξεων στο ΠΣ2 είναι ευκολότερος από τον αλγόριθμο του ΠΣ1.**

5. Πρόσθεση στο ΠΣ2

Ας υποθέσουμε δύο ν-ψήφιους ακέραιους δυαδικούς αριθμούς ψ_1 και ψ_2, με $|\psi_2| > |\psi_1|$, και ας εξετάσουμε τις διάφορες περιπτώσεις που μπορούν να παρουσιαστούν. Έτσι, έχουμε διαδοχικά τις εξής 4 περιπτώσεις για το άθροισμα $\psi_1 + \psi_2$:

α. Οι αριθμοί ομόσημοι θετικοί

Οι αριθμοί ψ_1 και ψ_2 προφανώς θα ικανοποιούν την σχέση:

$$\psi_1,\ \psi_2 > 0. \tag{4}$$

Αφού οι αριθμοί είναι θετικοί, **η παράστασή τους είναι κανονική** και η πρόσθεση γίνεται με το γνωστό τρόπο, χωρίς πρόβλημα.

β. Οι αριθμοί ετερόσημοι με τον αρνητικό απολύτως μεγαλύτερο

Οι αριθμοί ψ_1 και ψ_2 είναι ετερόσημοι και θα ικανοποιούν τις σχέσεις:

$$\psi_1 > 0 \qquad \text{και} \quad \psi_2 < 0, \qquad \text{με} \quad |\psi_2| > |\psi_1|. \tag{5}$$

Τότε, ο ψ_1 εκφράζεται κανονικά με την απόλυτη τιμή του, ενώ ο ψ_2 ως αρνητικός εκφράζεται με το συμπλήρωμά του ως προς 2, δηλαδή $\bar{\psi}_2 = 2^\nu - |\psi_2|$. Άρα $\psi_1 + \bar{\psi}_2 = |\psi_1| + (2^\nu - |\psi_2|) = 2^\nu - (|\psi_2| - |\psi_1|)$· δηλαδή, το αποτέλεσμα της πρόσθεσης είναι εκπεφρασμένο με το συμπλήρωμα του $|\psi_2| - |\psi_1|$ ως προς 2, **που σημαίνει αρνητικό άθροισμα με απόλυτη τιμή** $|\psi_2| - |\psi_1|$, όπως πρέπει.

Παράδειγμα:

Σαν παράδειγμα, έστω η άθροιση των αριθμών 30 και -45, για τους οποίους έχουμε:

$$\begin{array}{rcrcl} \psi_{1_{<10>}} & = & 30 \stackrel{\Pi\Sigma 2}{\longrightarrow} & = & 00011110 \\ \psi_{2_{<10>}} & = & -45 \stackrel{\Pi\Sigma 2}{\longrightarrow} & = & 11010011. \\ \hline \psi_1 + \bar{\psi}_2 & = & -15 \stackrel{\Pi\Sigma 2}{\longrightarrow} & = & 11110001 \stackrel{\sigma^*}{\rightarrow} -(00001110 + 1) = -15. \end{array}$$

γ. Οι αριθμοί ετερόσημοι με τον θετικό απολύτως μεγαλύτερο

Οι αριθμοί ψ_1 και ψ_2 στην προκειμένη περίπτωση θα ικανοποιούν τις σχέσεις:

(6) $$\psi_1 < 0 \ \text{ και } \ \psi_2 > 0, \quad \text{με} \quad |\psi_2| > |\psi_1.|$$

Βάση των προηγουμένων θα έχουμε, προφανώς:

$$\bar{\psi}_1 + \psi_2 = 2^\nu - |\psi_1| + |\psi_2| = 2^\nu + (|\psi_2| - |\psi_1|)$$

---

*Όπου σ σημαίνει συμπλήρωση ως προς τον αντίστοιχο τρόπο πράστασης

Στο παραπάνω άθροισμα το 2^{ν} θα αποτελέσει το ν+1 ψηφίο του α-θροίσματος που σαν υπερχείλισμα (overflow) αγνοείται, με συνέπεια **να παραμείνει ως άθροισμα ο αριθμός** $|\psi_2| - |\psi_1|$**, που είναι το ζητούμενο**.

Παράδειγμα:

Σαν παράδειγμα ας πάρουμε τους αριθμούς -30 και 45, για τους οποίους έχουμε:

$$\begin{array}{rclcl} \psi_{1_{<10>}} & = & -30 \xrightarrow{\Pi\Sigma 2} & = & 11100010 \\ \psi_{2_{<10>}} & = & 45 \xrightarrow{\Pi\Sigma 2} & = & 00101101 \\ \hline \psi_1 + \psi_2 & = & 15 \xrightarrow{\Pi\Sigma 2} & = & \boxed{1}\,00001111 \rightarrow 15_{<10>} \end{array}$$

αγνοείται

δ. Οι αριθμοί ομόσημοι αρνητικοί

Οι αριθμοί ψ_1 και ψ_2 είναι τώρα ομόσημοι αρνητικοί, δηλαδή:

(7) $$\psi_1,\ \psi_2 < 0.$$

Βάση των προηγουμένων στην προκειμένη περίπτωση:

$$\bar{\psi}_1 + \bar{\psi}_2 = 2^{\nu} - |\psi_1| + 2^{\nu} - |\psi_2| = 2^{\nu} + (2^{\nu} - |\psi_1| - |\psi_2|)$$

Στο παραπάνω άθροισμα το πρώτο 2^ν θα αγνοηθεί ως υπερχείλισμα, ενώ η παρένθεση θα αποτελέσει το άθροισμα που λόγω της έκφρασής του στο ΠΣ2, σημαίνει ότι είναι αρνητικό, με απόλυτη τιμή: $|\psi_1| + |\psi_2|$, όπως πρέπει.

Παράδειγμα:

Σαν παράδειγμα ας πάρουμε τους αριθμούς -30 και -45, έτσι έχουμε

$$\begin{array}{rcccl} {\psi_1}_{<10>} & = & -30 \xrightarrow{\Pi\Sigma 2} & = & 11100010 \\ {\psi_2}_{<10>} & = & -45 \xrightarrow{\Pi\Sigma 2} & = & 11010011 \\ \hline \psi_1 + \psi_2 & = & -75 \xrightarrow{\Pi\Sigma 2} & = & \underbrace{\boxed{1}}_{\textbf{αγνοείται}}10110101 \xrightarrow{\sigma^*} -(01001010 + 1) = -75, \end{array}$$

το αναμενόμενο.

6. Πρόσθεση στο ΠΣ1

Θα υποθέσουμε στη συνέχεια τις ίδιες προϋποθέσεις όπως και στο ΠΣ2 και θα διακρίνουμε, επίσης, τις ίδιες περιπτώσεις. Έτσι, θα έχουμε διαδοχικά τις παρακάτω αντίστοιχες 4 περιπτώσεις:

α. Αριθμοί ομόσημοι θετικοί

Οι αριθμοί ψ_1 και ψ_2 είναι:

$$\psi_1 > 0 \text{ και } \psi_2 > 0, \tag{8}$$

οπότε το άθροισμά τους θα είναι το αναμένομενο (χωρίς μετατροπές):

$$\psi_1 + \psi_2 = |\psi_1| + |\psi_2|.$$

β. Αριθμοί ετερόσημοι με τον αρνητικό απολύτως μεγαλύτερο

Οι αριθμοί ψ_1 και ψ_2 ως ετερόσημοι θα ικανοποιούν τις σχέσεις:

$$\psi_1 > 0 \text{ και } \psi_2 < 0, \ |\psi_2| > |\psi_1|. \tag{9}$$

Στην περίπτωση αυτή θα έχουμε μετατροπή στον αρνητικό και άθροισμα το:

$$\psi_1 + \underline{\psi}_2 = |\psi_1| + 2^\nu - |\psi_2| - 1 = 2^\nu - (|\psi_2| - |\psi_1|) - 1$$

δηλάδή, **αποτέλεσμα αρνητικό, με απόλυτη τιμή** $|\psi_2| - |\psi_1|$**, όπως πρέπει**.

Παράδειγμα:

Σαν παράδειγμα, ας πάρουμε τους αριθμούς $\psi_1 = 30$ και $\psi_2 = -45$. Τότε έχουμε:

$$\begin{array}{rcccl} \psi_{1_{<10>}} & = & 30 \xrightarrow{\Pi\Sigma 1} & = & 00011110 \\ \psi_{2_{<10>}} & = & -45 \xrightarrow{\Pi\Sigma 1} & = & 11010010 \\ \hline \psi_1 + \psi_2 & = & -15 \xrightarrow{\Pi\Sigma 1} & = & 11110000 \end{array} \xrightarrow{\sigma} -(00001111) \rightarrow -15_{<10>}.$$

γ. Αριθμοί ετερόσημοι με τον θετικό απολύτως μεγαλύτερο

Οι αριθμοί ψ_1 και ψ_2 θα ικανοποιούν τις σχέσεις:

$$\psi_1 < 0,\ \psi_2 > 0, \quad \text{με} \quad |\psi_2| > |\psi_1|, \tag{10}$$

ο αρνητικός θα υποστεί μετατροπή και το άθροισμα θα είναι:

$$\underline{\psi}_1 + \psi_2 = 2^\nu - |\psi_1| - 1 + |\psi_2| = 2^\nu + (|\psi_2| - |\psi_1| - 1).$$

Το 2^ν του αθροίσματος αποτελεί το υπερχείλισμα, ενώ το υπόλοιπο είναι το ζητούμενο άθροισμα, αλλά κατά 1 μικρότερο. Έτσι, μπορούμε **στο ΠΣ1** να θεωρήσουμε ότι **το υπερχείλισμα δεν αγνοείται αλλά προστίθεται ως σημείο μονάδας στο τελικό άθροισμα** (ανακύκλωση υπερχείλισης).

Παράδειγμα:

Σαν παράδειγμα ας πάρουμε τους αριθμούς $\psi_1 = -30$ και $\psi_2 = 45$, για τους οποίους έχουμε:

$$
\begin{array}{rcrcr}
\underline{\psi}_{1_{<10>}} & = & -30 & \xrightarrow{\Pi\Sigma 1} & 11100001 \\
\psi_{2_{<10>}} & = & 45 & \xrightarrow{\Pi\Sigma 1} & 00101101 \\
\hline
\underline{\psi_1} + \psi_2 & = & 15 & \xrightarrow{\Pi\Sigma 1} & \boxed{1}\,00001110 \\
 & & & & +1 \\
\cline{5-5}
 & & & & 00001111
\end{array}
\;\longrightarrow 15_{<10>},
$$

όπως πρέπει.

δ. Αριθμοί ομόσημοι αρνητικοί

Οι αριθμοί ψ_1 και ψ_2 είναι:

$$\psi_1,\ \psi_2 < 0, \tag{11}$$

οπότε θα έχουμε, μετατροπές σε αμφότερους και άθροισμα το:

$$\underline{\psi}_1 + \underline{\psi}_2 = 2^\nu - |\psi_1| - 1 + 2^\nu - |\psi_2| - 1 = 2^\nu - (2^\nu - (|\psi_1| - |\psi_2| - 1) - 1.$$

Το 2^ν εκτός παρενθέσεως αποτελεί το υπερχείλισμα που θα προστεθεί σαν μονάδα στο αποτέλεσμα, που γίνεται τελικά:

$$\underline{\psi}_1 + \underline{\psi}_2 = 2^\nu - (|\psi_1| + |\psi_2|) - 1,$$

δηλαδή αριθμός αρνητικός, **όπως πρέπει**.

Παράδειγμα:

Σαν παράδειγμα ας πάρουμε τους αριθμούς -30 και -45· έτσι έχουμε

$$\begin{array}{rcrcr}
\underline{\psi}_{1<10>} & = & -30 & \xrightarrow{\Pi\Sigma 1} & 11100001 \\
\underline{\psi}_{2<10>} & = & -45 & \xrightarrow{\Pi\Sigma 1} & 11010010 \\
\hline
\underline{\psi}_1 + \underline{\psi}_2 & = & -75 & \xrightarrow{\Pi\Sigma 1} & \boxed{1}\,10110011 \\
 & & & & +1 \\
\cline{5-5}
 & & & & 10110100
\end{array} \left.\vphantom{\begin{array}{c}1\\1\\1\end{array}}\right\} \xrightarrow{\sigma} -(01001011) \rightarrow -75.$$

7. Πράξεις στο δεκαδικό με μορφή BCD (Decimal Arithmetic)

Όταν σ' ένα Η.Υ. οι αριθμοί παριστάνονται σε μορφή BCD, οι αριθμητικές πράξεις μπορούν να εκτελεστούν είτε με μετατροπή των BCD μορφών σε δυαδικές και εκτέλεση των πράξεων, όπως στην προηγούμενη παράγραφο, είτε με απ' ευθείας εκτέλεση των πράξεων στις BCD εκφράσεις.

Στη δεύτερη περίπτωση **απαιτείται ο Η.Υ. να διαθέτει ιδιαίτερη μονάδα εκτέλεσης πράξεων σε BCD μορφές**. Στην μονάδα αυτή δεκαδικής αριθμητικής είναι δυνατόν, γενικά, να εκτελεστούν οι 4 πράξεις της αριθμητικής, καθώς και η μετατροπή δυαδικών σε BCD και αντίστροφα. Τέλος, η χρήση μονάδας δεκαδικής αριθμητικής σ᾽ ένα Η.Υ. έχει φυσικά επιπτώσεις στην ταχύτητα λειτουργίας του και στο κόστος του.

Στο κεφάλαιο 4 της κωδικοποίησης, έχουν δοθεί διάφοροι τρόποι παράστασης των δεκαδικών αριθμών σ᾽ ένα Η.Υ. σε μορφή BCD. Στο φυσικό κώδικα BCD 8,4,2,1 κάθε ψηφίο 0-9 του δεκαδικού αριθμού παρίσταται στον Η.Υ. με 4 δυαδικά ψηφία βάσει του πίνακα:

0 → 0000	1 → 0001	2 → 0010	3 → 0011	4 → 0100
5 → 0101	6 → 0110	7 → 0111	8 → 1000	9 → 1001

Οι συνδυασμοί από 1010 (10) έως 1111 (15) δεν είναι επιτρεπτοί στον κώδικα 8, 4, 2, 1. Εξ΄ άλλου, είναι γνωστό ότι οι Η.Υ. χρησιμοποιούν εσωτερικά μόνο δυαδικούς αριθμούς, ενώ η επικοινωνία τους με τους χρήστες γίνεται στο δεκαδικό σύστημα. Έτσι, είναι φανερό ότι υπάρχει συνεχής ανάγκη μετατροπών μεταξύ δυαδικών και δεκαδικών μορφών των αριθμών.

Στην μορφή BCD κάθε ψηφίο του δεκαδικού αριθμού μετατρέπεται ξεχωριστά στο ισοδύναμο του δυαδικό, των 4 bits. Προφανώς, για έναν αριθμό, **η BCD μορφή του δεν είναι η ίδια με την δυαδική μορφή του**· π.χ. για τον αριθμό 1985 θα έχουμε:

$$1985_{<2>} = 11111000001$$
$$1985_{<BCD>} = \underbrace{0001}_{1} \quad \underbrace{1001}_{9} \quad \underbrace{1000}_{8} \quad \underbrace{0101}_{5} = 0001100110000101.$$

Στη συνέχεια, **η αναφορά στις δεκαδικές πράξεις υπονοεί πράξεις που γίνονται μεταξύ αριθμών σε μορφή BCD**. Έτσι, οι αριθμοί θα είναι απομνημονευμένοι σε μορφή BCD, θα ανακαλούνται στην ίδια μορφή, θα υφίστανται επεξεργασία σε δεκαδική αριθμητική και τα αποτελέσματά της θα αποθηκεύονται στην ίδια μορφή, **με συνέπεια να αποφεύγεται ολοκληρωτικά η μετατροπή από δεκαδικούς σε δυαδικούς και αντίστροφα**.

Τέλος, ένα πρόσθετο πλεονέκτημα της δεκαδικής αριθμητικής είναι ότι οι αριθμοί BCD έχουν μεταβλητό μήκος (μέχρι 512 ψηφία το πολύ) με αποτέλεσμα **η ακρίβεια που μπορεί να επιτευχθεί να είναι πάρα πολύ ικανοποιητικη**, στις εφαρμογές.

Οι δεκαδικές πράξεις λαμβάνουν χώρα είτε μεταξύ αριθμών χωρίς πρόσημο, είτε μεταξύ προσημασμένων αριθμών.

Οι αριθμοί χωρίς πρόσημο θεωρούνται πάντοτε θετικοί, ενώ οι προσημασμένοι παρίστανται όπως και στους δυαδικούς, με μία από τις τρεις μορφές, όπου το πιο σημαντικό BCD ψηφίο έχει το ρόλο του προσήμου, που για τους θετικούς είναι το 0 ($= 0000_{<BCD>}$) και για τους αρνητικούς το μέγιστο ψηφίο του δεκαδικού συστήματος το 9 ($= 1001_{<BCD>}$).

Έτσι, **ισχύει ο γνωστός κανόνας του πιο σημαντικού δυαδικού ψηφίου που είναι 0 για τους θετικούς και 1 για τους αρνητικούς**.

Τέλος, οι δεκαδικές πράξεις σε μορφές BCD μοιάζουν με τις δυαδικές πράξεις και ο τρόπος εκτέλεσης τους εξαρτάται από τον τρόπο παράστασης των αρνητικών αριθμών. Ποιο αναλυτικά έχουμε:

α. Πρόσθεση/Αφαίρεση στο BCD

Η πρόσθεση και η αφαίρεση στη δεκαδική αριθμητική γίνεται με χρήση συμπληρωμάτων όπως και στους δυαδικούς, μόνο που εδώ τα συμπληρωμάτα είναι ΠΣ9 και ΠΣ10. **Στο τέλος το άθροισμα πρέπει να υποστεί κατάλληλες μετατροπές** είτε γιατί εμφανίστηκαν μη επιτρεπτά BCD ψηφία (10-15), είτε διότι υπήρξε κρατούμενο. Η μετατροπή για μεν το κρατούμενο είναι απλή, αφού αυτό μεταφέρεται στις μονάδες του ψηφίου της αμέσου υψηλότερης τάξης, ενώ πρόστίθεται στο ψηφίο αυτό ο 6· **στην παρουσία ενός μη αποδεκτού ψηφίου απαιτείται πρόσθεση στο ψηφίο αυτό του** 6.

Παράδειγμα (πρόσθεσης):

Το παρακάτω παράδειγμα δίνει τη πλήρη εικόνα της διαδικασίας:

0698	0000	0110	1001	1000	Δυαδική
+ 0586	0000	0101	1000	0110	Πρόσθεση
1284	0000	1011	$\boxed{1}$ 0001	1110	Κρατούμενο
		+1			Τροποποιήσεις
		0110	0110	0110	Πρόσθεση του 6
	0000	$\boxed{1}$ 0010	0111	$\boxed{1}$ 0100	
	+1		+1		Μεταφορά
	$\underbrace{0001}_{1}$	$\underbrace{0010}_{2}$	$\underbrace{1000}_{8}$	$\underbrace{0100}_{4}$	Κρατουμένου

Δεκαδική Πρόσθεση σε μορφές BCD

Παράδειγμα 2 (Αφαίρεση)

Σαν παράδειγμα αφαίρεσης, ας πάρουμε την περίπτωση:

$$0836_{<10>} - 0226_{<10>}.$$

Επιλέγουμε το ΠΣ10, οπότε έχουμε την ισοδύναμη πράξη:

$$0836 + 9774 = \boxed{1}\,0610 = 610$$

απαλείφεται

Σε μορφή BCD θα έχουμε:

(0836)	+	0000	1000	0011	0110	Δυαδική
(9774)		1001	0111	0111	0100	Πρόσθεση
		1001	$\underline{1111}$	$\underline{1010}$	$\underline{1010}$	$\rightarrow$ Υπογραμμίζονται τα μη αποδεκτά ψηφία BCD
			0110	0110	0110	$\rightarrow$ Τροποιήσεις (Πρόσθεση 6)
		1001	$\boxed{1}$ 0101	$\boxed{1}$ 0000	$\boxed{1}$ 0000	
		+1	+1	+1		Μεταφορά Κρατουμένου
		$\underline{1010}$	0110	0001	0000	
		0110				Τροποποίηση (πρόσθεση 6)
	$\boxed{1}$	$\underbrace{0000}_{0}$	$\underbrace{0110}_{6}$	$\underbrace{0001}_{1}$	$\underbrace{0000}_{0}$	

απαλείφεται

δηλαδή το αποτέλεσμα της αφαίρεσης είναι ο BCD 610, όπως πρέπει.

β. Πολλαπλασιασμός στο BCD

Για τον πολλαπλασιασμό σε μορφή BCD πρέπει να παρατηρήσουμε ότι κατά την εύρεση των μερικών γινομένων, μπορεί να έχουμε γινόμενο δύο ψηφίων BCD μεγαλύτερο του 19 μέχρι και το 81, ούτε οι απαιτούμενες τροποποιήσεις θα πρέπει να καλύψουν και τις νέες περιπτώσεις, βάσει του παρακάτω πίνακα 2.

Πίνακας 2. Τροποποιήσεις των ψηφίων του γινομένου από 10-81.

Γινόμενο $\gamma = \psi_1 \cdot \psi_2$	Τροποποίηση (Πρόσθεση του αριθμού:)
$9 < \gamma \leq 19$	6 (0110)
$19 < \gamma \leq 29$	12 (01110)
$29 < \gamma \leq 39$	18 (010010)
$39 < \gamma \leq 49$	24 (011000)
$49 < \gamma \leq 59$	30 (011110)
$59 < \gamma \leq 69$	36 (0100100)
$69 < \gamma \leq 79$	42 (0101010)
$79 < \gamma \leq 81$	48 (0110000)

Παράδειγμα:

Σαν παράδειγμα ας πάρουμε την περίπτωση $04763_{<10>} * 091_{<10>}$, ή σε BCD μορφή:

× 04763	0000	0100	0111	0110	0011	
091			0000	1001	0001	
4763		0100	0111	0110	0011	**α΄ μερικό γινόμενο**
42867	0100	0111	0110	0011		β΄ μερικό γινόμενο
433433	0000	0000	0000	0000		
	0000	0000	0000	0000		
	0100	0111	0110	0011		
	100100	111111	110110	11011		Τροποποίηση μη
	10010	100100	11110	1100		αποδεκτών ψηφίων
	$\boxed{11}$ 0110	$\boxed{110}$ 0011	$\boxed{101}$ 0100	$\boxed{10}$ 0111		
+11	+ 110	+ 101	+10			Μεταφορά κρατουμένου
	1100	1000	0110	0111		
	+ 0110					Τροποποιήσεις
11	$\boxed{1}$ 0010	1000	0110	0111		
+1						Μεταφορά κρατουμένου
100	0010	1000	0110	0111		**β΄ μερικό γινόμενο**

Το τελικό γινόμενο θα είναι:

		0100	0111	0110	0011	α΄ μερικό γινόμενο
100	0010	1000	0110	0111		β΄ μερικό γινόμενο
100	0010	1100	1101	1101	0011	
		110	110	110		Τροποποιήσεις
100	0010	$\boxed{1}$ 0010	$\boxed{1}$ 0011	$\boxed{1}$ 0011	0011	
	+1	+1	+1			Μεταφορά κρατουμένου
$\underbrace{100}_{4}$	$\underbrace{0011}_{3}$	$\underbrace{0011}_{3}$	$\underbrace{0100}_{4}$	$\underbrace{0011}_{3}$	$\underbrace{0011}_{3}$	**Τελικό αποτέλεσμα**

δηλαδή ο αριθμός 433433, όπως πρέπει.

8. Πράξεις κινητής υποδιαστολής (Floating point arithmetic)

Στις πράξεις που αναλύσαμε στις προηγούμενες παραγράφους, οι αριθμοί που χρησιμοποιήθηκαν **ήταν ακέραιοι**, γι΄ αυτό και η αντίστοιχη αριθμητική ονομάζεται **αριθμητική σταθερής υποδιαστολής** (η υποδιαστολή εννοείται πάντα στο τέλος του αριθμού).

Το **βασικό μειονέκτημα** της αριθμητικής σταθερής υποδιαστολής **είναι η περιορισμένη ακρίβειά της**, που εκδηλώνεται έντονα στις πράξεις του πολλαπλασιασμού και της διαίρεσης, όπου παρουσιάζεται ανεπάρκεια διαθέσιμων ψηφίων για την παράσταση των γινομένων και των πηλίκων· π.χ. για τους εξαψήφιους αριθμούς 43.7198 και 21.231 ενώ το γινόμενό τους είναι 928.5150738 αυτό θα αποδοθεί σε στρογγυλευμένη μορφή ως 928.515 λόγω έλλειψης ψηφίων.

Τέλος, **στους επιστημονικούς υπολογισμούς συχνά παρουσιάζεται η ανάγκη για χρήση αριθμών που ανήκουν σε μια ευρεία περιοχή** (π.χ. χρήση αριθμών με απόλυτη τιμή πολύ μεγάλη και πολύ μικρή συγχρόνως, όπως παρουσιάζεται στην περίπτωση της αφαίρεσης δύο πολύ μεγάλων αριθμών που είναι σχεδόν ίσοι).

Η αριθμητική κινητής υποδιαστολής παρέχει ακριβώς αυτή τη δυνατότητα για χρήση αριθμών που ανήκουν σε μια διευρημένη περιοχή και αυτό με τη χρήση ακριβώς της ίδιας λέξης του Η.Υ. που χρησιμοποιείται στην αριθμητική σταθερής υποδιαστολής, αλλά με άλλη δομή, **τη δομή του επιστημονικού συμβολισμού**, των πραγματικών αριθμών:

$$a = m \cdot \beta^e$$

όπου:

m κλασματικός στην περιοχή $-1 < m < 1$,

e ακέραιος στην περιοχή $-(\beta^\psi - 1)_{<10>} \leq e \leq (\beta^\psi - 1)_{<10>}$

με ψ το πλήθος των ψηφίων του εκθέτη χωρίς πρόσημο. Π.χ. οι αριθμοί: $a_1 = 0.0837 \cdot 10^{-3}$ ή $a_2 = 0.9254 \cdot 10^6$.

Η παραπάνω παράσταση των αριθμών ονομάζεται **παράσταση κινητής υποδιαστολής, γιατί με αυξομείωση του εκθέτη** e **μετακινείται η υποδιαστολή στον** m, έτσι ώστε ο αριθμός a να μην αλλάζει, όπως φαίνεται στις ακόλουθες περιπτώσεις για τον αριθμό 1985:

$$\begin{array}{lcl} 0.1985 & * & 10^4 \\ 1.985 & * & 10^3 \\ 19.85 & * & 10^2 \\ 198.5 & * & 10^1 \\ 19850 & * & 10^{-1} \\ 198500 & * & 10^{-2} \end{array}$$

Είναι προφανές ότι στους αριθμούς κινητής υποδιαστολής μπορούμε με κατάλληλη προσαρμογή του εκθέτη να έχουμε **το πρώτο δεκαδικό ψηφίο του κλασματικού μέρους του αριθμού διάφορο από το 0**, οπότε **ο αριθμός ονομάζεται κανονικοποιημένος** (normalized). **Κανονικοποιημένες μορφές των αριθμών κινητής υποδιαστολής χρησιμοποιούν οι Η.Υ.** Στο προηγούμενο παράδειγμα ο a_1 δεν είναι κανονικοποιημένος, ενώ ο a_2 είναι.

Τέλος, για τη διεύρυνση του πεδίου του εκθέτη, υπάρχει πάντοτε **ένα σταθερό πλεόνασμα σ᾽ αυτόν**, πράγμα που δίνει τη δυνατότητα χρήσης και αρνητικών εκθετών χωρίς τη χρήση bit προσήμου. Έτσι, **ο πραγματικός εκθέτης** e^* **ισούται με τον εμφανιζόμενο** e **μείον το πλεόνασμα** (που είναι το μισό του εύρους του εκθέτη)· π.χ. στον UNIVAC 1106 ο εκθέτης παρίσταται με 8 bits, άρα έχει εύρος 256 και το μισό τους είναι 128· κατά συνέπεια το πεδίο μεταβολής του εκθέτη τους θα είναι από -128 έως 127. Στον πίνακα 1 της περιγραφής των αριθμητικών δεδομένων έχουμε τη δομή των αριθμών κινητής υποδιαστολής για 4 γνωστά υπολογιστικά συστήματα.

Οι κανόνες εκτέλεσης των τεσσάρων πράξεων είναι:

$$\begin{aligned}
(\mathbf{m_1} \cdot \beta^{\mathbf{e_1}}) \cdot (\mathbf{m_2} \cdot \beta^{\mathbf{e_2}}) &= (\mathbf{m_1} \cdot \mathbf{m_2}) \cdot \beta^{\mathbf{e_1+e_2}} \\
(\mathbf{m_1} \cdot \beta^{\mathbf{e_1}}) : (\mathbf{m_2} \cdot \beta^{\mathbf{e_2}}) &= (\mathbf{m_1}/\mathbf{m_2}) \cdot \beta^{\mathbf{e_1-e_2}} \\
(\mathbf{m_1} \cdot \beta^{\mathbf{e_1}}) \pm (\mathbf{m_2} \cdot \beta^{\mathbf{e_1}}) &= (\mathbf{m_1} \pm \mathbf{m_2}) \cdot \beta^{\mathbf{e_1}}
\end{aligned}$$

Προφανώς για την εκτέλεση των προσθέσεων και των αφαιρέσεων οι εκθέτες πρέπει να είναι ίσοι· σε διαφορετικές περιπτώσεις γίνεται «**ευθυγράμμιση**» (alignement), που με κατάλληλη προσαρμογή του κλασματικού μέρους του μικρότερου μπορούμε να επιτύχουμε την ισότητα των εκθετών. Π.χ. για την πρόσθεση των αριθμών $0.48 \cdot 10^5$ και $0.25 \cdot 10^6$, μπορούμε να τροποποιήσουμε τον πρώτο και να πάρουμε τον ισοδύναμό του $0.048 \cdot 10^6$, οπότε πλέον η πρόσθεση εκτελείται χωρίς δυσκολία αφού:

$$0.048 \cdot 10^6 + 0.25 \cdot 10^6 = (0.048 + 0.25) \cdot 10^6 = 0.298 \cdot 10^6.$$