ΕΣ 08 – Επεξεργαστές Ψηφιακών Σημάτων

Αναπαράσταση Δεδομένων σε Επεξεργαστές Ψ.Ε.Σ

Τμήμα Επιστήμη και Τεχνολογίας Τηλεπικοινωνιών

Πανεπιστήμιο Πελοποννήσου

Βιβλιογραφία Ενότητας

- Gilmore [2006]: Κεφάλαια 2 & 3
- Kehtarnavaz [2005]: Chapter 6
- Kuo [2005]: *Chapter 3, Sections 3.1-3.3*
- Lapsley [2002]: *Chapter 3*
- Σημειώσεις διδάσκοντα

Εισαγωγή

- Τα ψηφιακά συστήματα, φίλτρα, αλγόριθμοι κλπ υλοποιούνται σε υλικό στο οποίο τα δεδομένα (τιμές δειγμάτων, συντελεστές) αποθηκεύονται ως δυαδικοί αριθμοί με πεπερασμένο μήκος λέξης.
  - Η αναπαράσταση δεδομένων (αριθμών) μπορεί να γίνει είτε με κινητή υποδιαστολή (floating point) είτε με σταθερή υποδιαστολή (fixed point)
  - Ο τρόπος αναπαράστασης επηρεάζει: (α) το εύρος των αριθμών που μπορούν να αναπαρασταθούν, (β) την ακρίβεια της αναπαράστασης, και (γ) τον τρόπο με τον οποίο πραγματοποιούνται οι πράξεις μεταξύ δυαδικών αριθμών (αριθμητική)
- Το πεπερασμένο μήκος λέξης πρέπει να λαμβάνεται υπ' όψιν δεδομένου ότι:
  - Επηρεάζει την αναπαράσταση των δειγμάτων
  - Επηρεάζει την αναπαράσταση των συντελεστών των ψηφιακών συστημάτων επηρεάζοντας την απόδοση τους
  - Δημιουργεί προβλήματα υπερχείλισης ή/και υποχείλισης με αποτέλεσμα τη δημιουργία ανακριβών ή και εσφαλμένων αποτελεσμάτων

Αριθμοί και Αριθμητική Σταθερής Υποδιαστολής

- Στους επεξεργαστές Ψ.Ε.Σ η αναπαράσταση αριθμών γίνεται (όπως σε κάθε ψηφιακό σύστημα) από μια σειρά από δυαδικά ψηφία (bits).
  - Το πλήθος των bits που χρησιμοποιείται για την αναπαράσταση των αριθμών σε έναν επεξεργαστή ονομάζεται μέγεθος λέξης αναπαράστασης δεδομένων ή μήκος λέξης (data width ή word length)
  - Συνήθως οι επεξεργαστές Ψ.Ε.Σ έχουν μήκος λέξης πολλαπλάσιο του 8. Οι σύγχρονοι επεξεργαστές Ψ.Ε.Σ έχουν μήκος λέξης που φτάνει τα 64 bits αν και το σύνηθες είναι 32 bits.
- Με δεδομένο το μήκος λέξης η αναπαράσταση αριθμών μπορεί να γίνει είτε με σταθερή είτε με κινητή υποδιαστολή. Οπότε αντίστοιχα έχουμε:
  - επεξεργαστές κινητής υποδιαστολής (floating point processors)
  - επεξεργαστές σταθερής υποδιαστολής (fixed point processors)
- Οι επεξεργαστές κινητής υποδιαστολής παρέχουν μεγαλύτερη ακρίβεια αποτελεσμάτων, δεν παρουσιάζουν εύκολα φαινόμενα υπερχείλισης, και προγραμματίζονται ευκολότερα. Από την άλλη πλευρά απαιτούν πιο σύνθετα κυκλώματα και κατά συνέπεια είναι πιο ακριβοί
  - Συνήθως σε μαζική παραγωγή απλών προϊόντων στα οποία οι επεξεργαστές Ψ.Ε.Σ εκτελούν συγκεκριμένους αλγορίθμους (π.χ. Modems, mp3 players, κλπ) χρησιμοποιούνται επεξεργαστές σταθερής υποδιαστολής

Τρόποι αναπαράστασης αριθμών σταθερής υποδιαστολής

- Υπάρχουν τρεις βασικές μορφές αναπαράστασης αριθμών σταθερής υποδιαστολής:
  - Μορφή πρόσημο-μέτρο (sign-magnitude)
    - Για παράδειγμα με μήκος λέξης 4 bits οι αριθμοί 7, -6 θα αναπαρασταθούν ως 0111 (MSB 0=>θετικός αριθμός), και 1110 (MSB 1=>αρνητικός αριθμός) αντίστοιχα
    - MSB = Most Significant Bit και είναι το αριστερότερο δυαδικό ψηφίο στη δυαδική συμβολοσειρά αναπαράστασης του αριθμού
    - Για την αναπαράσταση ενός αριθμού με τη μορφή πρόσημο-μέτρο πρώτα μετατρέπουμε το μέτρο του αριθμού σε δυαδικό και στη συνέχεια προσθέτουμε το MSB ανάλογα με το πρόσημό του
  - Μορφή συμπλήρωμα ως προς 1 (one's complement)
    - Στο παραπάνω παράδειγμα οι αριθμοί 7, -6 θα αναπαρασταθούν ως 0111 και 1001 (το οποίο είναι το συμπλήρωμα του 0110 που αντιστοιχεί στο 6)
    - Για την αναπαράσταση ενός αρνητικού αριθμού με τη μορφή συμπλήρωμα ως προς 1, πρώτα μετατρέπουμε τον αντίστοιχο θετικό αριθμό σε δυαδικό και στη συνέχεια παίρνουμε το συμπλήρωμα του αποτελέσματος. Για τους θετικούς αριθμούς η αναπαράσταση είναι ίδια με αυτήν της μορφής πρόσημο-μέτρο

ΕΣ 08: Επεξεργαστές Ψηφιακών Σημάτων

☑ Εισαγωγή
★ Αριθμοί και Αριθμητική Σταθερής Υποδιαστολής
☐ Αριθμοί και Αριθμητική Κινητής Υποδιαστολής
☐ Οι Συνέπειες του Πεπερασμένου Μήκους Λέξης

Τρόποι αναπαράστασης αριθμών σταθερής υποδιαστολής (II)

- Μορφή συμπλήρωμα ως προς 2 (two's complement)
  - Στη μορφή πρόσημο-μέτρο και συμπλήρωμα ως προς 1 έχουμε διπλή αναπαράσταση του 0 (μηδέν), είτε ως -0 είτε ως +0
  - Το συμπλήρωμα ως προς 2 αποφεύγει αυτή την περίπτωση
  - Στο προηγούμενο παράδειγμα οι αριθμοί 7, -6 θα αναπαρασταθούν ως 0111 και 1010 (το οποίο είναι το συμπλήρωμα ως προς ένα του 0110 που αντιστοιχεί στο 6 +0001)
  - Για την αναπαράσταση ενός αρνητικού αριθμού με τη μορφή συμπλήρωμα ως προς 2 προχωράμε ως εξής:
    - Πρώτα μετατρέπουμε τον αντίστοιχο θετικό αριθμό σε δυαδικό
    - Παίρνουμε το συμπλήρωμα του αποτελέσματος
    - Προσθέτουμε το δυαδικό 1 στο αποτέλεσμα
  - Σημειώνεται ότι για τους θετικούς αριθμούς η αναπαράσταση είναι ίδια με αυτήν της μορφής πρόσημο-μέτρο, και συμπλήρωμα ως προς 1.
  - Η μορφή αναπαράστασης συμπλήρωμα ως προς δύο είναι η μορφή που χρησιμοποιείται στην πράξη στους επεξεργαστές Ψ.Ε.Σ

Ακέραιοι και κλασματικοί αριθμοί σταθερής υποδιαστολής

- Για την αναπαράσταση κλασματικών αριθμών με τη μορφή συμπλήρωμα ως προς δύο χρησιμοποιούνται διάφορα formats τα οποία δηλώνονται με τη μορφή Qm.n
  - Ισχύει N = m+n+1 (όπου N το μήκος λέξης, m το πλήθος των bits για την αναπαράσταση του ακέραιου μέρους, n το πλήθος των bits για την αναπαράσταση του κλασματικού μέρους και 1 bit για το πρόσημο)
  - Για παράδειγμα το format Q3.4 αναπαριστά αριθμούς χρησιμοποιώντας 8 bits, 3 από τα οποία διατίθενται για το ακέραιο μέρος και 4 για το κλασματικό.
  - Το format Q0.15 τις περισσότερες φορές συμβολίζεται απλά με Q.4 όπως και κάθε format που δεν αποθηκεύει το ακέραιο μέρος ενός αριθμού.
  - Για να αποφεύγονται προβλήματα υπερχείλισης στον πολλαπλασιασμό οι επεξεργαστές Ψ.Ε.Σ χρησιμοποιούν format της μορφής Q.x (π.χ Q.15, Q.31 κλπ). Αυτό όμως πρέπει να λαμβάνεται υπ' όψιν από τον προγραμματιστή ώστε τα δεδομένα να κανονικοποιούνται (μετατρέπονται έτσι ώστε να παίρνουν τιμές στο διάστημα [-1 1]) πριν την επεξεργασία τους.
  - Η τιμή ενός κλασματικού αριθμού σε *Qm.n* δίνεται από τη σχέση:

$$x = -b_{N-1} \cdot 2^{m} + \sum_{k=0}^{N-2} b_k \cdot 2^{k-n}$$

όπου b_{N-1} είναι η τιμή του MSB bit και b_0 είναι η τιμή του LSB

Παράδειγμα

- Για το format Q3.4 να βρεθούν:
  - Ο μέγιστος αριθμός που μπορεί να αναπαρασταθεί
    ΑΠ.: 0111 1111 => 7.9375
  - Ο ελάχιστος θετικός αριθμός που μπορεί να αναπαρασταθεί
    ΑΠ.: 0000 0001 => 0.0625
  - Ο ελάχιστος αρνητικός αριθμός που μπορεί να αναπαρασταθεί
    ΑΠ.: 1000 0000 => -8
  - Ο μέγιστος αρνητικός αριθμός που μπορεί να αναπαρασταθεί
    ΑΠ.: 1111 1111 => -0.0625
  - Το διάστημα κβαντισμού (για το κλασματικό μέρος)
    ΑΠ.: Δ =1/(2^4) = 0.0625
  - Το μέγιστο σφάλμα κβαντισμού
    ΑΠ.: e_{max} =Δ/2 = 0.0625/2
  - Η αναπαράσταση του αριθμού 3.67 (ΑΠ.: 0011 1011)
  - Η αναπαράσταση του αριθμού -3.67 (ΑΠ.: 1100 0101)
  - Η αναπαράσταση του αριθμού 8 (ΑΠ.: 0111 1111 – υπερχείλιση)
  - Η αναπαράσταση του αριθμού -0.03 (ΑΠ.: 0000 0000 – σφάλμα κβαντισμού)

Δυναμικό Εύρος και Ακρίβεια Αναπαράστασης

- Το δυναμικό εύρος (dynamic range) της αναπαράστασης αριθμών σε ένα δεδομένο format ορίζεται ως ο λόγος του μέγιστου (σε μέτρο) αριθμού που μπορεί να αναπαρασταθεί προς τον ελάχιστο (σε μέτρο), μη μηδενικό αριθμό
- Ο παραπάνω λόγος συνήθως εκφράζεται στη λογαριθμική κλίμακα ως:

$$DR(db) = 20\log_{10}\left(\frac{Max}{Min}\right)$$

- Το δυναμικό εύρος της αναπαράστασης αριθμών στο format Qm.n είναι:
  - Max = 2^m, Min = 2^{-n}, επομένως:

$$DR(db) = 20\log_{10}\left(\frac{2^m}{2^{-n}}\right) = 20\cdot(m+n)\cdot\log_{10}(2) = (N-1)*6$$

  - Παρατηρούμε ότι το δυναμικό εύρος είναι ανεξάρτητο των παραμέτρων m, n και εξαρτάται αποκλειστικά από το μήκος λέξης N που χρησιμοποιείται για την αναπαράσταση των δεδομένων

Δυναμικό Εύρος και Ακρίβεια Αναπαράστασης (II)

- Η ακρίβεια (precision) της αναπαράστασης αριθμών σε ένα δεδομένο format ορίζεται ως η (κατά απόλυτη τιμή) διαφορά δύο διαδοχικών αριθμών
  - Από τον παραπάνω ορισμό προκύπτει ότι στο format Qm.n η ακρίβεια εξαρτάται αποκλειστικά από τον αριθμό *n* και ισούται με 2^{-n}
  - Η ακρίβεια ισούται με το διάστημα κβαντισμού (Δ). Επομένως το μέγιστο σφάλμα προσέγγισης ή σφάλμα κβαντισμού είναι ίσο με το μισό της ακρίβειας, $e = \Delta/2 = 2^{-(n+1)}$
- Παράδειγμα: Για το format Q3.4 να βρεθούν
  - Το δυναμικό εύρος

    DR = (N-1)*6 = 42 db
  - Το εύρος τιμών

    $[-2^3 \;\; 2^3-2^{-4}] = [-8 \;\; 7.9375]$
  - Η ακρίβεια

    $p = 2^{-n} = 1/16$
  - Το σφάλμα προσέγγισης

    $e = 2^{-(n+1)} = 1/32$

Πρόσθεση και πολλαπλασιασμός αριθμών σταθερής υποδιαστολής

- Η πρόσθεση δύο αριθμών οι οποίοι λαμβάνουν τιμές στο διάστημα $[-a\ b]$ παίρνει τιμές στο διάστημα $[-2a\ 2b]$. Αυτό σημαίνει ότι το εύρος τιμών του αποτελέσματος της άθροισης διπλασιάζεται (από b+a γίνεται 2(b+a))
- Επομένως για την αναπαράσταση του αθροίσματος δύο δυαδικών αριθμών χρειαζόμαστε ένα επιπλέον bit (N+1 αντί για N) για να μην έχουμε φαινόμενο υπερχείλισης
- Ο πολλαπλασιασμός δύο αριθμών οι οποίοι λαμβάνουν τιμές στο διάστημα $[-a\ a]$ παίρνει τιμές στο διάστημα $[-a^2\ a^2]$.
  - Αν το $|a|\leq 1$ το εύρος τιμών του γινομένου δεν αλλάζει. Αντίθετα αν $|a|>1$ για την αναπαράσταση του γινομένου δύο δυαδικών αριθμών χρειαζόμαστε επιπλέον bits για να μην έχουμε φαινόμενο υπερχείλισης
- Παράδειγμα:
  - Δύο αριθμοί σε format Qm.n (α) προστίθενται, (β) πολλαπλασιάζονται. Να βρεθεί το μήκος λέξης που απαιτείται για την αποθήκευση του (α) αθροίσματος και (β) γινομένου
  - Η μέγιστη (κατά μέτρο) τιμή που μπορεί να αποθηκευτεί με βάση το format Qm.n είναι Max = 2^m. Η μέγιστη τιμή του αθροίσματος θα είναι $2*2^m=2^{m+1}$. Επομένως για το άθροισμα χρειαζόμαστε M = (m+1)+n+1 = N+1 bits

ΕΣ 08: Επεξεργαστές Ψηφιακών Σημάτων

☑ Εισαγωγή
★ Αριθμοί και Αριθμητική Σταθερής Υποδιαστολής
☐ Αριθμοί και Αριθμητική Κινητής Υποδιαστολής
☐ Οι Συνέπειες του Πεπερασμένου Μήκους Λέξης

Αφαίρεση και διαίρεση αριθμών σταθερής υποδιαστολής

- Η μέγιστη τιμή του γινομένου θα είναι $2^m*2^m=2^{2m}$. Επομένως για το γινόμενο χρειαζόμαστε M = (2m)+n+1 = N+m bits για να μην έχουμε υπερχείλιση
  - Για format της μορφής Q.x το γινόμενο δεν δημιουργεί φαινόμενα υπερχείλισης (εξαίρεση η περίπτωση πολλαπλασιασμού (-1)x(-1)=1). Αυτός είναι και ο λόγος που οι επεξεργαστές Ψ.Ε.Σ χρησιμοποιούν format αυτής της μορφής
  - Στην πράξη οι συσσωρευτές στους επεξεργαστές Ψ.Ε.Σ έχουν μήκος 2*N για να αντιμετωπίζουν προβλήματα υπερχείλισης λόγω διαδοχικών αθροίσεων. Επειδή και οι συσσωρευτές αναπαριστούν τα δεδομένα σε Q.x κάποια από τα 2N bits κρατούνται ως φρουροί (guard bits) για την αποθήκευση του ακέραιου μέρους του αποτελέσματος
  - Για format της μορφής Qx. απαιτούνται 2N-1 bits για την αποθήκευση του γινομένου
- Η αφαίρεση δυαδικών αριθμών δεν παρουσιάζει κάποια ιδιαιτερότητα γιατί είναι παρόμοιας λογικής με την πρόσθεση
- Η διαίρεση δυαδικών σε επεξεργαστές Ψ.Ε.Σ υλοποιείται με διαδοχικές αφαιρέσεις. Σε αντίθεση με τον πολλαπλασιασμό δεν υλοποιείται κατευθείαν σε υλικό

Αριθμοί και Αριθμητική Κινητής Υποδιαστολής

- Οι επεξεργαστές Ψ.Ε.Σ αναπαριστούν αριθμούς κινητής υποδιαστολής στη μορφή: πρόσημο – εκθέτης - δεκαδικό μέρος (sign-exponent-mantissa(fraction)).
- Σύμφωνα με το πρότυπο IEEE-754 οι αριθμοί κινητής υποδιαστολής απλής ακρίβειας (single) αναπαριστώνται με 1 bit για το πρόσημο, 8 bit για τον εκθέτη, και 23 bit για το δεκαδικό μέρος. Επομένως η τιμή x ενός αριθμού κινητής υποδιαστολής δίνεται από τη σχέση:

$$x = (-1)^{sign} \times 2^{e-127} \times 1.f$$

A single-precision binary floating-point number is stored in a 32 bit word:

Width in bits		
1	8	23
Sign	Exponent	Fraction
bit index (0 on right)		
31	30 29 23	22 21 ..... 1 0

Μορφές Αναπαράστασης

Class	*Exponent Value*	*Fraction Value*
Zeroes	*0*	*0*
Denormalised numbers	*0*	*non zero*
Normalised numbers	*1-254*	*any*
Infinities	*255*	*0*
NaN (Not a Number)	*255*	*non zero*

- Αριθμοί κινητής υποδιαστολής διπλής ακρίβειας (double)

Width in bits		
1	11	52
Sign	Exponent	Fraction
bit index (0 on right) (exponent bias is +1023 $=2^{11-1}$)		
63	62 61 52	51 ..... 1 0

Παράδειγμα

- Ένας επεξεργαστής Ψ.Ε.Σ κινητής υποδιαστολής έχει μήκος λέξης N=8 bits και αναπαριστά τους αριθμούς με 1 bit για το πρόσημο, 3 bit για τον εκθέτη, και 4 bit για το δεκαδικό μέρος. Να βρεθούν:
  - Ποια τιμή αναπαρίσταται με τη συμβολοσειρά 10010110
    sign=1 => αρνητικός, exp = 001 =1-3=-2, mantissa=0110 =>
    $-2^{-2}x1.375 = -0.3438$
  - Η δυαδική αναπαράσταση του αριθμού -12.138
    δεκαδική αναπαράσταση (πρόσημο 1) 1100.0001, μετακίνηση προς τα αριστερά κατά 3 θέσεις => exp =3+3 =6 => 110, mantissa = 1000.
    Τελικά -12.138 => 1 110 1000
  - Ο μέγιστος αριθμός που μπορεί να αναπαρασταθεί
    ΑΠ.: 0111 1111 => 31
  - Ο ελάχιστος θετικός αριθμός που μπορεί να αναπαρασταθεί
    ΑΠ.: 0000 0001 => exp = -3, fraction =0.0001 =>0.1328
  - Ο ελάχιστος αρνητικός αριθμός που μπορεί να αναπαρασταθεί
    ΑΠ.: 1000 0001 => exp = -3, fraction =0.0001 =>-0.1328
  - Ο μέγιστος αρνητικός αριθμός που μπορεί να αναπαρασταθεί
    ΑΠ.: 1111 1111 => -31

ΕΣ 08: Επεξεργαστές Ψηφιακών Σημάτων

☑ Εισαγωγή
☑ Αριθμοί και Αριθμητική Σταθερής Υποδιαστολής
★ Αριθμοί και Αριθμητική Κινητής Υποδιαστολής
☐ Οι Συνέπειες του Πεπερασμένου Μήκους Λέξης

Παράδειγμα (ΙΙ)

- Ένας επεξεργαστής Ψ.Ε.Σ κινητής υποδιαστολής έχει μήκος λέξης N=8 bits και αναπαριστά τους αριθμούς με 1 bit για το πρόσημο, 3 bit για τον εκθέτη, και 4 bit για το δεκαδικό μέρος. Να βρεθούν:
  - Το δυναμικό εύρος τιμών σε db
  - Η ακρίβεια της αναπαράστασης των τιμών
  - Το μέγιστο σφάλμα προσέγγισης

Πρόσθεση και πολλαπλασιασμός αριθμών κινητής υποδιαστολής

◇ Η πρόσθεση δύο αριθμών κινητής υποδιαστολής, απλής ακρίβειας x και y

$$x = -1^{sign_x} \cdot 2^{(\exp_x - 127)} \cdot 1.man_x$$

$$y = -1^{sign_y} \cdot 2^{(\exp_y - 127)} \cdot 1.man_y$$

δίνεται από τη σχέση:

$$z = x + y = \begin{cases} -1^{sign_x} \cdot 1.man_x + \left(-1^{sign_y} \cdot 1.man_y \cdot 2^{-(\exp_x - \exp_y)}\right) \cdot 2^{(\exp_x - 127)} & if\ |x| \geq |y| \\ -1^{sign_y} \cdot 1.man_y + \left(-1^{sign_x} \cdot 1.man_x \cdot 2^{-(\exp_y - \exp_x)}\right) \cdot 2^{(\exp_y - 127)} & if\ |y| > |x| \end{cases}$$

◇ Ο πολλαπλασιασμός των αριθμών x και y δίνεται από τη σχέση:

$$z = x \cdot y = \left(-1^{sign_x} \cdot 1.man_x\right) \cdot \left(-1^{sign_y} \cdot 1.man_y\right) \cdot 2^{(\exp_x + \exp_y - 254)}$$

Παράδειγμα

◇ Ένας επεξεργαστής Ψ.Ε.Σ κινητής υποδιαστολής έχει μήκος λέξης N=8 bits και αναπαριστά τους αριθμούς με 1 bit για το πρόσημο, 3 bit για τον εκθέτη, και 4 bit για το δεκαδικό μέρος. Να βρεθούν το άθροισμα και το γινόμενο των αριθμών x=2.44 και y =-12.16

◇ Οι αριθμοί x και y αναπαρίστανται ως ακολούθως:
x = 0 100 0100, y = 1 110 1000

◇ Ισχύει |y|>|x|

◇ $\exp_x$=4, man_x=0.25, $\exp_y$=6, man_y=0.5

◇ Άθροισμα:
{-1.5+(1.25*2^{-2})}*2^3=-9.5

◇ Γινόμενο:
1.875*2^4 = -30

Οι Συνέπειες του Πεπερασμένου Μήκους Λέξης

- Το πεπερασμένο μήκος λέξης για την αναπαράσταση των δεδομένων (αριθμητικών τιμών) στους επεξεργαστές Ψ.Ε.Σ επηρεάζει την ακρίβεια των αποτελεσμάτων της επεξεργασίας εξαιτίας:
  - Της μειωμένης ακρίβειας στην αναπαράσταση των τιμών του αναλογικού σήματος που έχει ψηφιοποιηθεί και κβαντισθεί
    - Το σφάλμα αναπαράστασης των τιμών λόγω χρήσης πεπερασμένου αριθμού από bits για την αναπαράσταση των δειγμάτων στον αναλογικό-ψηφιακό μετατροπέα (ADC) είναι γνωστό ως σφάλμα κβαντισμού
  - Της μειωμένης ακρίβειας στην αναπαράσταση των συντελεστών των ψηφιακών συστημάτων (π.χ. ψηφιακών φίλτρων) που υλοποιούνται στους επεξεργαστές
  - Των φαινομένων υπερχείλισης που παρουσιάζονται λόγω του πεπερασμένου δυναμικού εύρους σε ενδιάμεσα αποτελέσματα πράξεων όπως η πρόσθεση και ο πολλαπλασιασμός
  - Της αποκοπής των LSB bits για την αποθήκευση ενδιάμεσων και τελικών αποτελεσμάτων στη μνήμη των επεξεργαστών Ψ.Ε.Σ μετά την μεταφορά τους από το συσσωρευτή

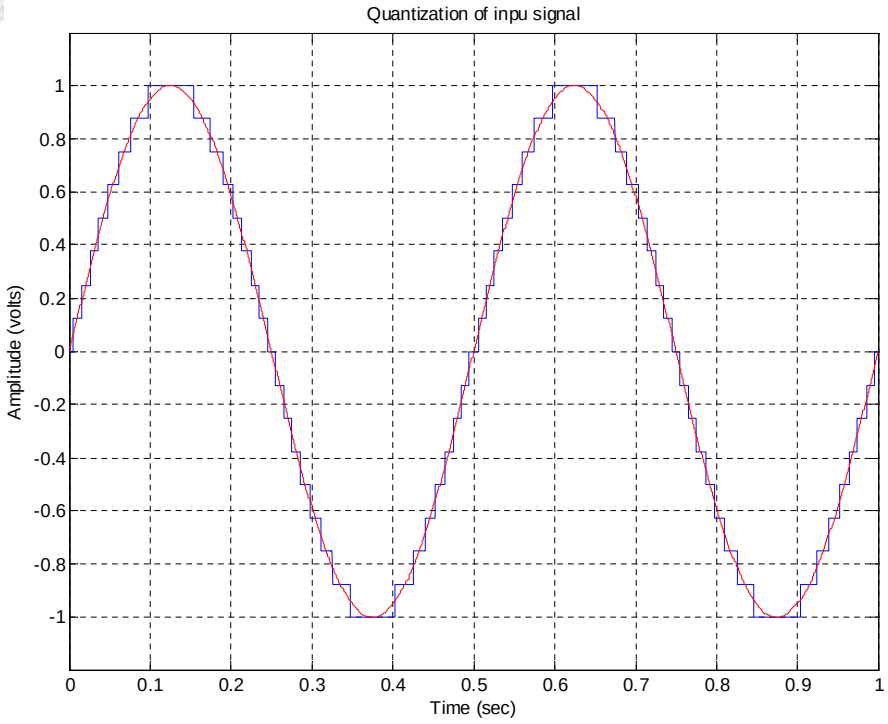
Σφάλμα Κβαντισμού

- Σε πολλές περιπτώσεις οι επεξεργαστές Ψ.Ε.Σ συμπεριλαμβάνουν και αναλογικό-ψηφιακούς μετατροπείς (ADC – Analog to Digital Converters) έτσι ώστε να μπορούν να χειρίζονται απευθείας και αναλογικά σήματα.
- Η διαδικασία της ψηφιοποίησης εισάγει ένα σφάλμα προσέγγισης της πραγματικής τιμής του αναλογικού σήματος από την ψηφιακή αναπαράστασή του (δυαδική σειρά από bits) εξαιτίας του πεπερασμένου αριθμού από bits που διατίθενται για την αναπαράσταση των ψηφιοποιημένων δειγμάτων
  - Το ανωτέρω σφάλμα (δηλαδή η διαφορά της πραγματικής τιμής από την ψηφιοποιημένη) ονομάζεται σφάλμα κβαντισμού
  - Το σφάλμα κβαντισμού είναι μια στοχαστική ποσότητα και επομένως η ισχύς του χαρακτηρίζεται από τη μέση τιμή και τη διασπορά του: $P_e=(m_e)^2 + (\sigma_e)^2$
- Αν V_{FS} είναι το εύρος διακύμανσης του αναλογικού σήματος τότε το διάστημα κβαντισμού Δ δίνεται από τη σχέση:

$$\Delta = \frac{V_{FS}}{2^N}$$

όπου N ο αριθμός των bits που διατίθενται για την αναπαράσταση των ψηφιοποιημένων δειγμάτων

Σφάλμα Κβαντισμού



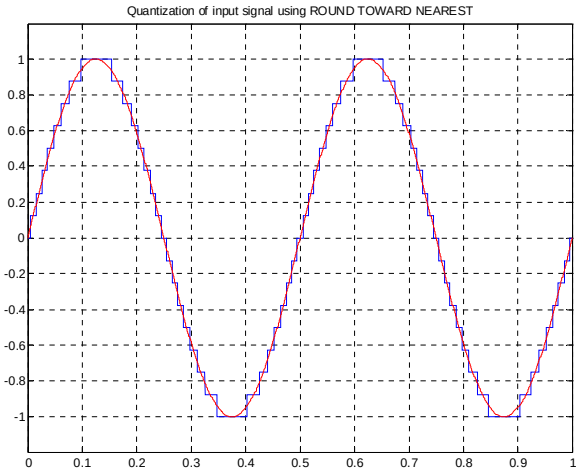
Σφάλμα Κβαντισμού (ΙΙ)

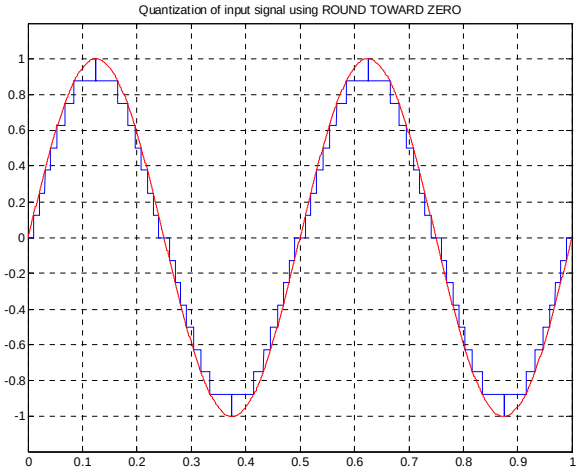
- Υπάρχουν τέσσερις διαφορετικές μεθοδολογίες προσέγγισης της πραγματικής τιμής του σήματος από την ψηφιοποιημένη:
  - Προς την πλησιέστερη στάθμη (round toward nearest - *round*)
  - Προς το μηδέν (round toward zero - *fix*)
  - Προς μικρότερες τιμές (round toward floor - *floor*)
  - Προς μεγαλύτερες τιμές (round toward ceil - *ceil*)
- Από τις παραπάνω μεθοδολογίες οι δύο πρώτες δίνουν σφάλμα κβαντισμού με μέση τιμή m_e=0 ενώ οι άλλες δύο δημιουργούν τάση (bias) προς αρνητικές ή θετικές τιμές καθώς η μέση τιμή του σφάλματος κβαντισμού είναι διάφορη του μηδενός και αρνητική ή θετική αντίστοιχα
  - Με βάση τα παραπάνω προκύπτει ότι η ισχύς του σφάλματος κβαντισμού είναι μικρότερη όταν χρησιμοποιείται η πρώτη μεθοδολογία πράγμα που συμβαίνει στη πράξη στους επεξεργαστές Ψ.Ε.Σ
  - Αποδεικνύεται ότι η ισχύς του σφάλματος κβαντισμού για την προσέγγιση προς την πλησιέστερη στάθμη είναι:

$$P_e = \sigma_e^2 = \frac{\Delta^2}{12}$$

- ☑ Εισαγωγή
- ☑ Αριθμοί και Αριθμητική Σταθερής Υποδιαστολής
- ☑ Αριθμοί και Αριθμητική Κινητής Υποδιαστολής
- ★ Οι Συνέπειες του Πεπερασμένου Μήκους Λέξης

Σφάλμα Κβαντισμού (II)

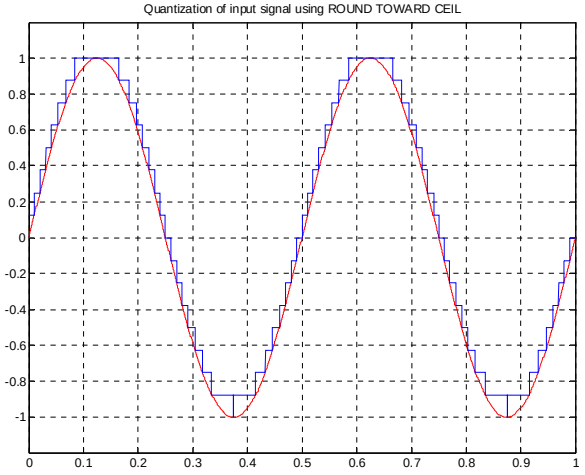


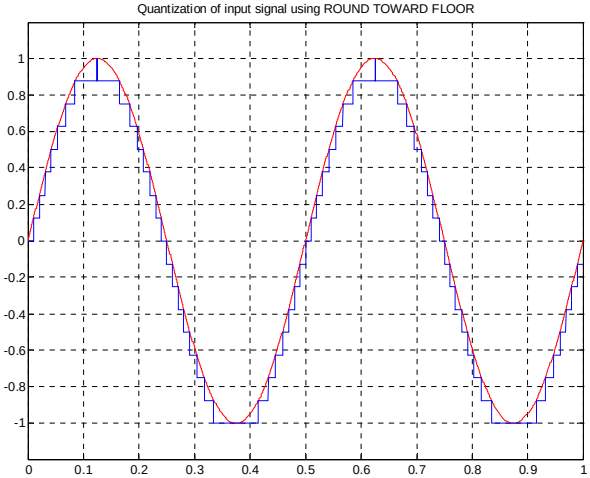


ΕΣ 08: Επεξεργαστές Ψηφιακών Σημάτων

- ☑ Εισαγωγή
- ☑ Αριθμοί και Αριθμητική Σταθερής Υποδιαστολής
- ☑ Αριθμοί και Αριθμητική Κινητής Υποδιαστολής
- ★ Οι Συνέπειες του Πεπερασμένου Μήκους Λέξης

Σφάλμα Κβαντισμού (II)





Σφάλμα Κβαντισμού (ΙΙΙ)

◇ Για ημιτονοειδή αναλογικά σήματα εισόδου η ισχύς τους δίνεται από τη σχέση:

$$P_s = \frac{\left(\frac{V_{FS}^2}{2}\right)^2}{2} = \frac{V_{FS}^2}{8} = \frac{\Delta^2 \cdot 2^{2N}}{8}$$

◇ Ένα μέτρο της ποιότητας κβαντισμού του αναλογικού σήματος είναι ο λόγος σήματος προς σφάλμα κβαντισμού (SQNR – Signal to Noise Quantization Ratio) το οποίο δίνεται από τη σχέση:

$$SQNR = 10\log_{10}\left(\frac{P_s}{P_e}\right) = 10\log_{10}\left(\frac{\frac{\Delta^2 \cdot 2^{2N}}{8}}{\frac{\Delta^2}{12}}\right) = 10\log_{10}\left(2^{2N} + \frac{3}{2}\right) = (6.02N + 1.76)dB$$

◇ Η παραπάνω σχέση δηλώνει ότι με κάθε ένα bit έχουμε αύξηση του SQNR κατά 6 περίπου dB.

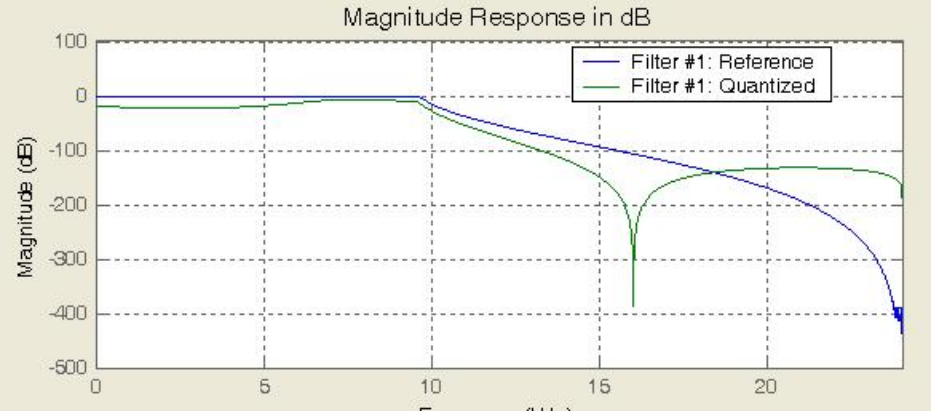
ΕΣ 08: Επεξεργαστές Ψηφιακών Σημάτων

☑ Εισαγωγή
☑ Αριθμοί και Αριθμητική Σταθερής Υποδιαστολής
☑ Αριθμοί και Αριθμητική Κινητής Υποδιαστολής
★ Οι Συνέπειες του Πεπερασμένου Μήκους Λέξης

Κβαντισμός Συντελεστών Ψηφιακών Συστημάτων

◇ Εκτός από τα δεδομένα (δείγματα) με πεπερασμένο αριθμό από bits αναπαριστώνται και οι συντελεστές των ψηφιακών συστημάτων (π.χ οι συντελεστές των ψηφιακών φίλτρων).

◇ Αυτό έχει ως αποτέλεσμα ορισμένες φορές να αλλοιώνονται οι προδιαγραφές και η επίδοση των ψηφιακών συστημάτων. Είναι πιθανόν επίσης να παρουσιαστούν φαινόμενα αστάθειας (ταλαντώσεις της εξόδου)



Υπερχείλιση

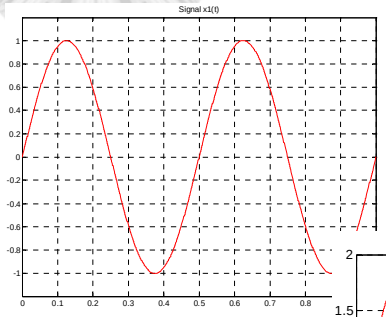
- Υπερχείλιση (overflow) συμβαίνει όταν μετά από αριθμητικές πράξεις όπως άθροιση και πολλαπλασιασμό τα ενδιάμεσα αποτελέσματα δεν μπορούν να αποθηκευτούν (υπερβαίνουν τη μέγιστη τιμή –κατά μέτρο- που μπορεί να αναπαρασταθεί με δεδομένο αριθμό από bits) χωρίς σφάλμα
- Υποχείλιση (underflow) συμβαίνει όταν τα ενδιάμεσα αποτελέσματα λαμβάνουν τιμές μικρότερες από την ελάχιστη κατά μέτρο τιμή που μπορεί να αποθηκευτεί
- Επειδή υπερχείλιση και υποχείλιση συμβαίνουν συνήθως στον συσσωρευτή λαμβάνονται τα εξής μέτρα:
  - Ο συσσωρευτής έχει μήκος λέξης τουλάχιστον διπλάσιο από τον αριθμό των bits που χρησιμοποιούνται για τα δεδομένα
  - Κάποια από τα bits του συσσωρευτή κρατούνται ως φρουροί (guard bits). Αυτό βέβαια έχεις ως συνέπεια μικρότερη ακρίβεια όσον αφορά τα ενδιάμεσα αποτελέσματα των πράξεων: Αν π.χ ο συσσωρευτής έχει μήκος λέξης 32 bits και τα 4 κρατούνται ως φρουροί τότε η ακρίβεια αναπαράστασης για τα ενδιάμεσα αποτελέσματα μειώνεται σε αυτή που αντιστοιχεί στα 28 bits αντί για 32.

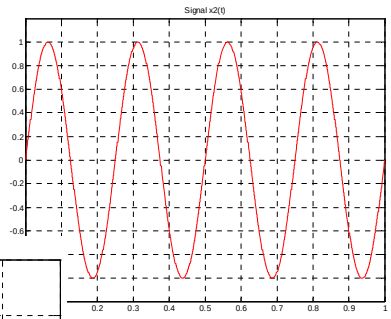
Υπερχείλιση (ΙΙ)

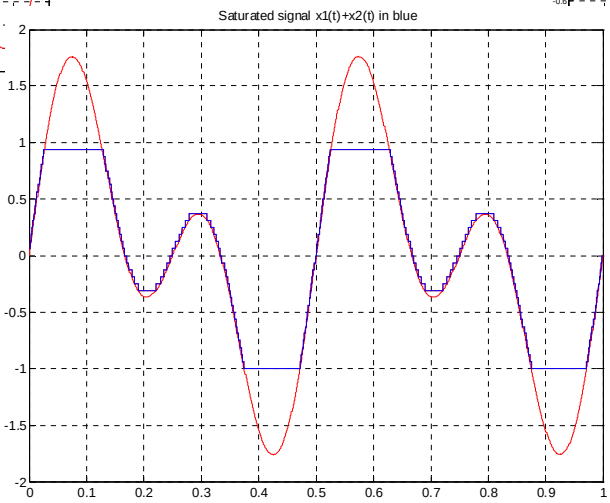
- Επειδή υπερχείλιση είναι πιθανότερο να συμβεί κατά τον πολλαπλασιασμό είναι σύνηθες η αναπαράσταση των δεδομένων και συντελεστών σε επεξεργαστές Ψ.Ε.Σ να γίνεται με format κλασματικής μορφής (π.χ Q.31, Q.15 κλπ).
  - Αυτό συνεπάγεται βέβαια κανονικοποίηση των σημάτων εισόδου (διαίρεση με τη μέγιστη τιμή που μπορούν να πάρουν)
  - Επιπλέον δημιουργεί φαινόμενο υπερχείλισης στην περίπτωση του πολλαπλασιασμού (-1) x (-1) …
- Οι επεξεργαστές Ψ.Ε.Σ αντιμετωπίζουν την υπερχείλιση με τη μέθοδο του κορεσμού σε αντίθεση με τους μικροεπεξεργαστές γενικού σκοπού που δεν λαμβάνουν καμία πρόνοια για την υπερχείλιση με αποτέλεσμα να έχουμε το φαινόμενο περιτυλίγματος (wrap around)
  - Το επόμενο σχήμα επιδεικνύει το πλεονέκτημα της μεθόδου saturation όσον αφορά την αντιμετώπιση της υπερχείλισης

- ☑ Εισαγωγή
- ☑ Αριθμοί και Αριθμητική Σταθερής Υποδιαστολής
- ☑ Αριθμοί και Αριθμητική Κινητής Υποδιαστολής
- ★ Οι Συνέπειες του Πεπερασμένου Μήκους Λέξης

Υπερχείλιση (III)



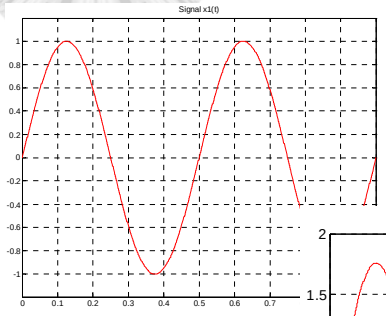


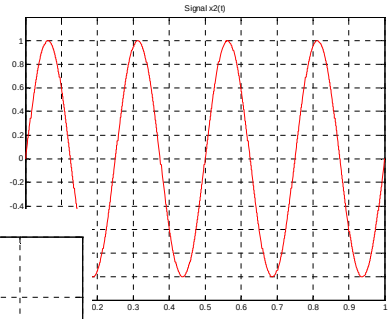


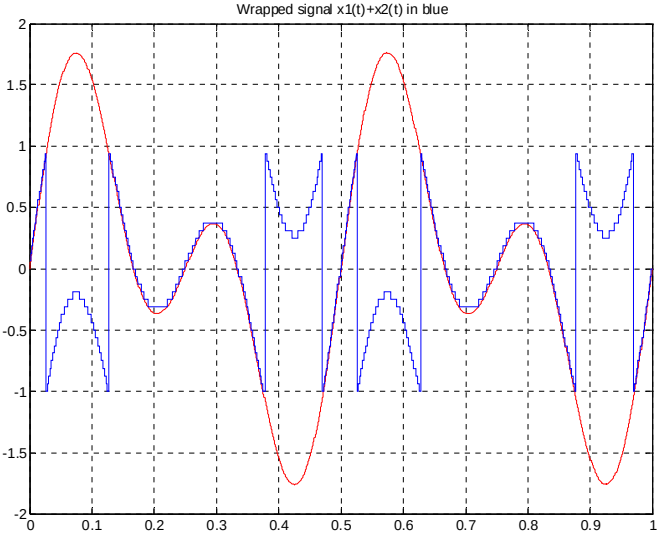
ΕΣ 08: Επεξεργαστές Ψηφιακών Σημάτων

- ☑ Εισαγωγή
- ☑ Αριθμοί και Αριθμητική Σταθερής Υποδιαστολής
- ☑ Αριθμοί και Αριθμητική Κινητής Υποδιαστολής
- ★ Οι Συνέπειες του Πεπερασμένου Μήκους Λέξης

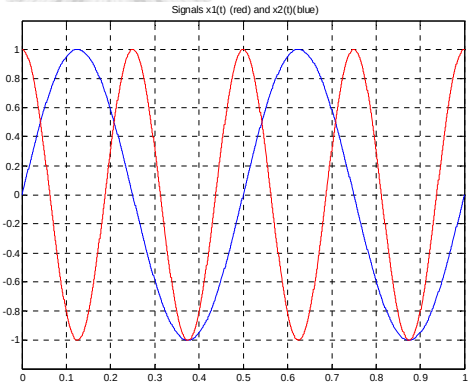
Υπερχείλιση (IV)

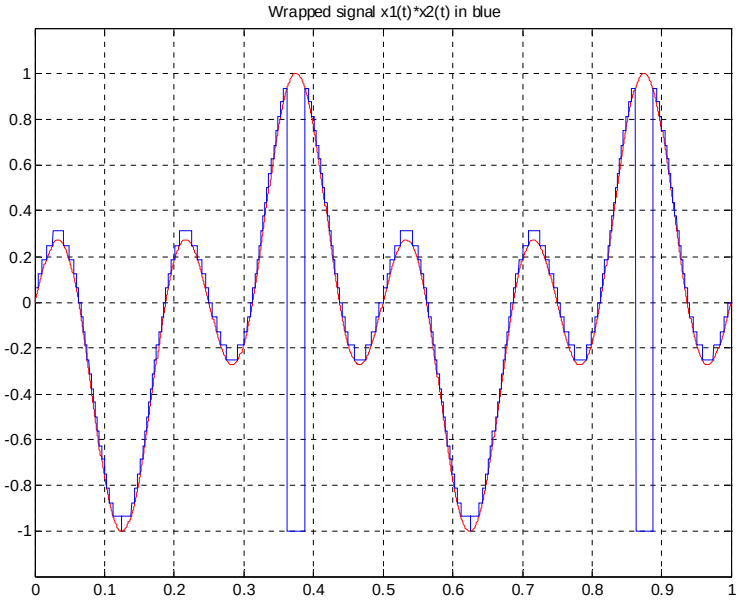






Υπερχείλιση σε πολλαπλασιασμό





Στρογγυλοποίηση και Αποκοπή

- Όπως ήδη αναφέρθηκε το μήκος λέξης του συσσωρευτή είναι πολύ μεγαλύτερο (συνήθως διπλάσιο) από το μήκος λέξης για την αποθήκευση αποτελεσμάτων στη μνήμη
  - Για παράδειγμα ένας επεξεργαστής Ψ.Ε.Σ μήκους λέξης N = 16bits μπορεί να έχει συσσωρευτές με μήκος λέξης 2N = 32
  - Επομένως κατά την μεταφορά αποτελεσμάτων από το συσσωρευτή στη μνήμη χρειάζεται η μείωση της ακρίβειας αναπαράστασης τους από 2N σε N
- Ο συνήθης τρόπος για την αντιμετώπιση του παραπάνω προβλήματος είναι η προσέγγιση προς τον πλησιέστερο ακέραιο (χρησιμοποιώντας προφανώς τα N bits αντί των 2N)
- Ο ευκολότερος τρόπος επίλυσης είναι η αποκοπή των LSB bits. Η μεθοδολογία αυτή είναι γνωστή ως truncation
  - Δυστυχώς για αναπαράσταση αριθμών με βάση το συμπλήρωμα ως προς δύο (δηλαδή η συνηθισμένη μορφή αναπαράστασης σε επεξεργαστές σταθερής υποδιαστολής) η αποκοπή δημιουργεί εσφαλμένα αποτελέσματα (εξαιτίας και των guard bits που χρησιμοποιούνται στο συσσωρευτή) όπως φαίνεται στο επόμενο παράδειγμα:

Παράδειγμα Αποκοπής

- Έστω ότι σε κάποιο στάδιο της επεξεργασίας του σήματος *x(n)* χρειάζεται να εκτελεστεί η πράξη MAC (Multiply Add Computation)

  y = a*x(k)+b*x(k-1)

  Όπου a = 0.667, b=0.333, x(k)=-0.9, x(k-1)=-0.7

- Ο επεξεργαστής έχει μήκος λέξης 4 bits και ο συσσωρευτής έχει μέγεθος 8 bits
  - Οι συντελεστές a, b θα αναπαρασταθούν ως a = 0101 (δηλαδή a =0.625), b = 0011 (δηλαδή b =0.375)
  - Οι τιμές x(k) και x(k-1) θα αναπαρασταθούν ως x(k) = 1001 (δηλαδή x(k) = -0.875), και x(k-1) = 1010 (δηλαδή x(k-1) = -0.750)
  - Το αποτέλεσμα της πράξης θα είναι y = -1.0156, το οποίο σε συσσωρευτή 8 bits με ένα guard bit θα αναπαρασταθεί ως y = 10111111 (δηλαδή y = -1.0156)
  - Εφαρμόζοντας αποκοπή στα τελευταία 4 bits θα έχουμε y = 1011 το οποίο ισοδυναμεί με τιμή y = -0.625 όταν χρησιμοποιήσουμε το format Q.3 με το οποίο αποθηκεύονται οι τιμές στη μνήμη του επεξεργαστή
  - Προφανώς αν χρησιμοποιούσαμε προσέγγιση στη πλησιέστερη τιμή θα είχαμε y =-1 (y = 1000) το οποίο είναι πολύ πιο κοντά στο πραγματικό αποτέλεσμα