

Τεχνολογία Ήχου

Διάλεξη 08: «Αναπαράσταση και κωδικοποίηση ψηφιακού ήχου»

Φλώρος Ανδρέας
Επικ. Καθηγητής

Κωδικοποίηση του ήχου...

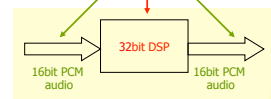
- Η αναγκαιότητα αναπαράστασης του ήχου σε κάποια διαχειρίσιμη και αναγνώσιμη μορφή...



Κωδικοποίηση και αριθμητική αναπαράσταση

- Η κωδικοποίηση των ηχητικών δεδομένων είναι απαραίτητη για
 - τη μετάδοσή τους
 - Π.χ. Μέσω καλωδίου ή ασύρματα
 - την αποθήκευσή τους
- Η επεξεργασία των ψηφιακών δεδομένων στηρίζεται στην αριθμητική αναπαράστασή τους
 - Πρέπει να είναι ανεξάρτητη της κωδικοποίησης
 - Εξαρτάται από την πλατφόρμα υλοποίησης της επεξεργασίας

Κωδικοποίηση ανεξάρτητη της αριθμητικής αναπαράστασης



Τύποι αριθμητικής αναπαράστασης

- Αριθμητική σταθερού σημείου (fixed point)
 - Αναπαράσταση συμπληρώματος του 2
 - Γραμμική αναπαράσταση
 - Χαμηλό κόστος και πολυπλοκότητα
 - Σφάλμα αποκοπής
- Αριθμητική κινητής υποδιαστολής (floating point)
 - Υψηλό κόστος και πολυπλοκότητα
- Η επιλογή του τύπου αριθμητικής αναπαράστασης εξαρτάται από τις ανάγκες της εφαρμογής, όπως:
 - Το επιθυμητό εύρος τιμών
 - Ανάγκες συγκράτησης των τιμών εντός συγκεκριμένων ορίων
 - Πλατφόρμα υλοποίησης



Αριθμητική αναπαράσταση σταθερού σημείου

- Έστω ότι το εύρος τιμών των κβαντιζμένων δειγμάτων εισόδου $s(k)$ είναι $-S_{\max}$ έως $S_{\max}$

- Κανονικοποίηση των τιμών των δειγμάτων εισόδου:

$$\hat{s}(k) = \frac{s(k)}{|S_{\max}|}$$

- Εύρος τιμών κανονικοποιημένου σήματος από -1 έως 1
- Έκφραση κανονικοποιημένων τιμών ως ζυγισμένο άθροισμα πεπερασμένου μήκους:

$$\hat{s}(k) = -b_0 + b_1 2^{-1} + b_2 2^{-2} + \dots + b_{N-1} 2^{-(N-1)} = -b_0 + \sum_{i=1}^{N-1} b_i 2^{-i}$$

- $b_0, b_1, \dots, b_{N-1}$ τα bit της δυαδικής αναπαράστασης



Αριθμητική αναπαράσταση σταθερού σημείου (συν.)

- b_0 : Most Significant Bit (MSB)
 - = 0, για θετικές κανονικοποιημένες τιμές
 - = 1, για αρνητικές κανονικοποιημένες τιμές
- b_{N-1} : Least Significant Bit (LSB)
 - Βήμα κβαντισμού $\Delta = 2^{-(N-1)}$
- Το εύρος τιμών της αναπαράστασης είναι:

$$-1 \leq \hat{s}(k) \leq 1 - 2^{-(N-1)}$$

- Η δυναμική περιοχή (σε dB) της αναπαράστασης είναι:

$$D = 20 \log_{10}(2^{N-1} - 1)$$



Παράδειγμα αναπαράστασης σταθερού σημείου (1)

- Έστω κβαντιστής με $N=3$ bit
 - Ελάχιστο βήμα αναπαράστασης: $\Delta = 2^{-(3-1)} = 0.25$
 - Μέγιστη τιμή αναπαράστασης: $1 - 2^{-(3-1)} = 0.75$
 - Ελάχιστη τιμή αναπαράστασης: -1
 - Δυναμική περιοχή αναπαράστασης: $D = 20 \log_{10}(2^{N-1} - 1) = 9.54$ dB
 - Δυαδική τιμή μέγιστης τιμής: $0.75 = -0 + 1 \times 0.5 + 1 \times 0.25$

$$\begin{array}{ccc} \downarrow & \downarrow & \downarrow \\ b_0 & b_1 & b_2 \end{array} = 0 \ 1 \ 1$$

«Αναπαράσταση συμπληρώματος του 2»

- Κατά τον πολλαπλασιασμό δύο τιμών δεν έχουμε ποτέ υπερχείλιση
- Κατά την πρόσθεση τιμών ενδέχεται να εμφανιστεί clipping
 - Εφαρμογή scaling



Γραμμική αριθμητική αναπαράσταση σταθερού σημείου

- Έστω Nbit αναπαράστασης ($b_0 \dots b_{N-1}$)
 - b_0 : LSB, b_{N-1} : MSB
- Μέγιστη τιμή αναπαράστασης: $\text{MAX} = 2^{(N-1)} - 1$
- Ελάχιστη τιμή αναπαράστασης: $\text{MIN} = -2^{(N-1)}$
- Βήμα κβαντισμού $\Delta = (\text{MAX} - \text{MIN}) / (2^N - 1)$
- Π.χ. για $N=3$, οι δυνατές τιμές αναπαράστασης είναι

Decimal	Binary ($b_2 \ b_1 \ b_0$)
0	000
1	001
2	010
3	011
4	100
5	101
6	110
7	111

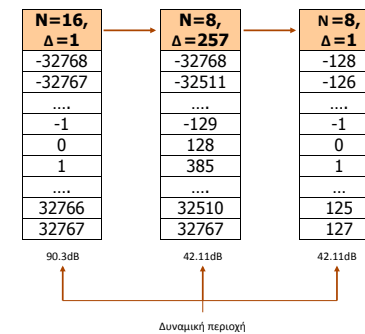


Παράδειγμα αναπαράστασης σταθερού σημείου (2)

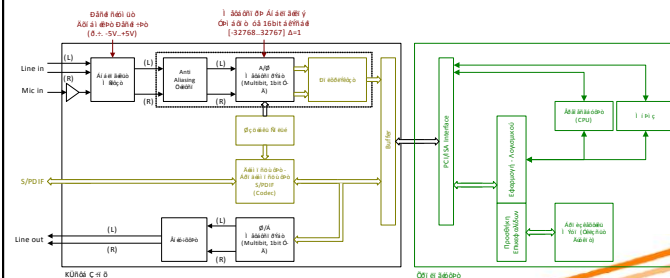
- Έστω κβαντιστής με $N=16\text{bit}$
 - Μέγιστη τιμή αριθμητικής αναπαράστασης: $2^{(N-1)} - 1 = 32767 = (1111111111111111)_2$
 - Ελάχιστη τιμή αριθμητικής αναπαράστασης: $-2^{(N-1)} = -32768 = (0111111111111111)_2$
 - Βήμα κβαντισμού: $\Delta=1$
 - Δυναμική περιοχή αναπαράστασης: $20\log_{10}(32767) = 90.3\text{dB}$
- Έστω κβαντιστής με $N=8\text{bit}$
 - Μέγιστη τιμή αριθμητικής αναπαράστασης 32767
 - Ελάχιστη τιμή αριθμητικής αναπαράστασης -32768
 - Βήμα κβαντισμού: $\Delta = 65535 / (2^8 - 1) = 257$
 - Δυναμική περιοχή αναπαράστασης: $20\log_{10}(32767/257) = 42.11\text{dB}$



Παράδειγμα αναπαράστασης σταθερού σημείου (3)



Παράδειγμα: Σύστημα κάρτας ήχου



Αριθμητική αναπαράσταση κινητής υποδιαστολής

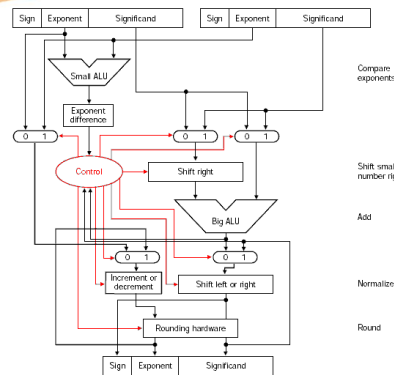
- Η αριθμητική αναπαράσταση ενός δείγματος $s(k)$ έχει τη μορφή:

$$s(k) = M_G 2^{E_G} \begin{cases} 0.5 \leq M_G < 1 \text{ (} N_M \text{ bits, fixed point)} \\ -2^{N_E-1} + 2 \leq E_G \leq 2^{N_E-1} - 1 \text{ (} N_E \text{ bits, ακέραιος)} \end{cases}$$

- Μη σταθερό βήμα κβαντισμού
 - Εξαρτάται από την περιοχή τιμών που πρόκειται να αναπαρασταθούν
- Πολλαπλασιασμός
 - Πρόσθεση E_G , πολλαπλασιασμός M_G , και ρύθμιση αθροίσματος εκθετικών ώστε το γινόμενο των M_G να ανήκει στην περιοχή $[0.5, 1]$
- Πρόσθεση
 - Το εκθετικό του μικρότερου αριθμού γίνεται ίδιο με του μεγαλύτερου
 - Πρόσθεση M_G
 - Κανονικοποίηση τελικού αποτελέσματος



Πολυπλοκότητα αναπαράστασης κινητής υποδιαστολής



Κύκλωμα
πρόσθεσης δύο
αριθμών κινητής
υποδιαστολής



Κωδικοποίηση ψηφιακών ηχητικών δεδομένων



Ψηφιακή κωδικοποίηση ήχου

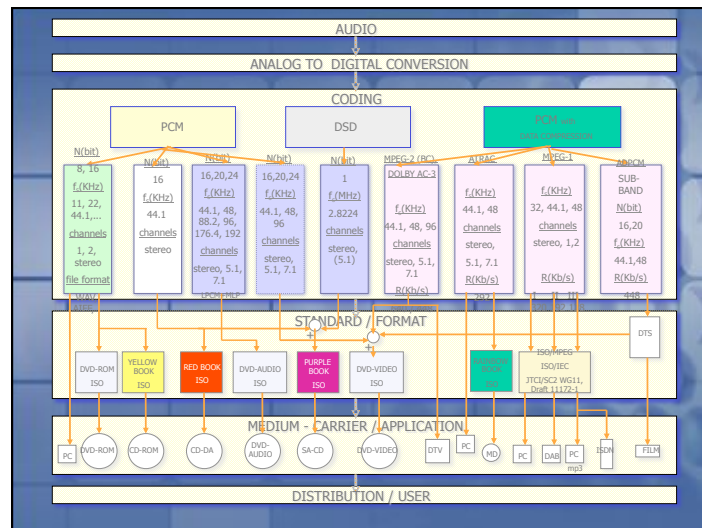
- Στο πρότυπο CD, τα ηχητικά δεδομένα κωδικοποιούνται κατά PCM
 - Παλμοκωδική διαμόρφωση (Pulse Code Modulation)
 - Ανάλυση 16bit, συχνότητα δειγματοληψίας 44.1kHz
 - 2 κανάλια ήχου
 - Ρυθμός μετάδοσης δεδομένων = $2 \times 44100 \times 16 = 1411200 \text{ bps}$
- Η PCM γρήγορα καθιερώθηκε σε ένα πλήθος εφαρμογών και συστημάτων
 - Επέκταση της τυποποίησης CD σε αρχεία υπολογιστή
 - Βελτιώσεις της τυποποίησης CD
 - Π.χ. DVD-Audio
 - Αύξηση συχνότητας δειγματοληψίας, ανάλυσης κβαντισμού
- Αύξηση του ρυθμού μετάδοσης δεδομένων
 - Αυξημένες απαιτήσεις επεξεργασίας και αποθήκευσης



Ψηφιακή κωδικοποίηση ήχου (συν.)

- Ανάπτυξη νέων μεθόδων κωδικοποίησης (εκτός της PCM)
 - Μέθοδοι συμπίεσης
 - Χρήση κυρίως σε εφαρμογές «δικτύωσης» και «αποθήκευσης»
 - MPEG1, 2, 4,
 - Dolby Digital
 - Κωδικοποιήσεις νέων τυποποιήσεων
 - Direct Stream Digital (DSD)
 - Κωδικοποιήσεις νέων ψηφιακών βαθμίδων
 - Pulse Width Modulation (PWM) για βαθμίδες ψηφιακών ενισχυτών
 - Multibit Σίγμα/Δέλτα κ.λ.π.



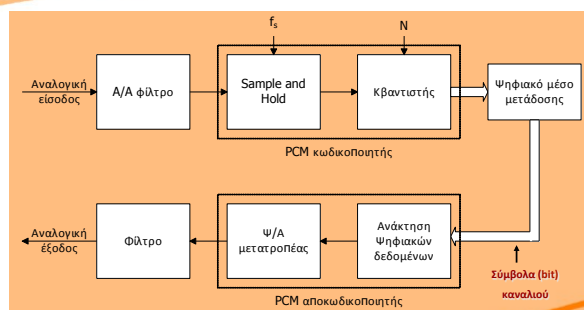


Η κωδικοποίηση PCM (N-bit)

- Π α λ μ ο κ ω δ ι κ ή δ ι α μ ό ρ φ ω σ η (Pulse Code Modulation, PCM)
 - Η πλέον διαδεδομένη μορφή ψηφιακής κωδικοποίησης
 - Απλή υλοποίηση
 - Μικρή (σχεδόν μηδενική) καθυστέρηση στον κωδικοποιητή
 - Ιδιαίτερα ευαίσθητη σε σφάλματα μετάδοσης (bit error rate)
- Β α σ ι κ έ ς π α ρ ά μ ε τ ρ ο ι κ ω δ ι κ ο π ο ι η σ η ς
 - Συχνότητα δειγματοληψίας ($f_s - \text{Hz}$)
 - Αριθμός δειγμάτων ανά δευτερόλεπτο
 - Ευκρίνεια κβαντισμού (N - bit)



Η κωδικοποίηση PCM (συν.)

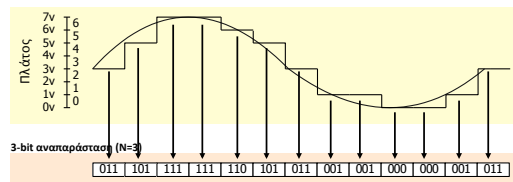


Χαρακτηριστικά PCM κωδικοποίησης

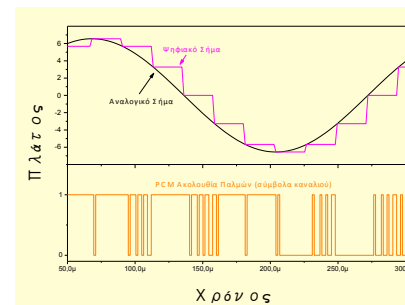
- Ρυθμός μετάδοσης $I = (f_s \times N)$ bit per second (bps)
- Για C ανεξάρτητα κανάλια $I = (C \times f_s \times N)$ bps
- Παράδειγμα: Ο PCM ρυθμός μετάδοσης της τυποποίησης CD
 - $C = 2$ κανάλια
 - $f_s = 44100$ δείγματα/second
 - $N = 16\text{bit}$
 - $I = 1411200\text{bps} \sim 1.4\text{Mbps}$
- Για $C=6, f_s=44.1\text{kHz}, N=16 \rightarrow I=4.2\text{Mbps}$
- Για $C=6, f_s=96\text{kHz}, N=24 \rightarrow I=13.8\text{Mbps}$



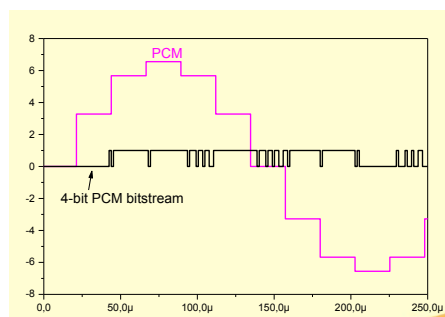
Παράδειγμα PCM κωδικοποίησης



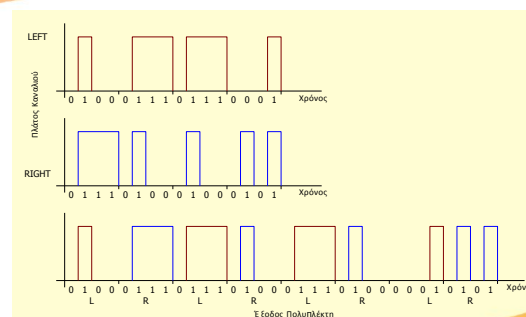
Παράδειγμα PCM κωδικοποίησης (συν.)



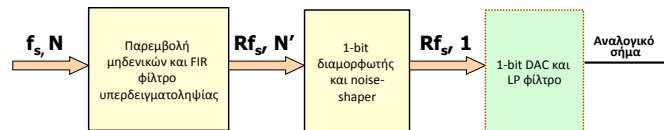
Παράδειγμα PCM κωδικοποίησης (συν.)



Πολύπλεξη PCM καναλιών



Γενική μορφή 1-bit κωδικοποιητών



- Κατηγορίες 1bit κωδικοποιητών / διαμορφωτών
 - Σίγμα/Δέλτα (Σ/Δ)
 - Απλού/Πολλαπλών σταδίων (MASH)
 - PWM
 - Class D, BD



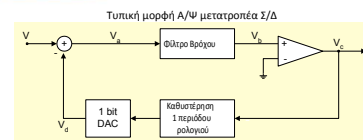
Κωδικοποίηση Σίγμα-Δέλτα (1-bit)

- Χρησιμοποιείται ευρύτατα σε υλοποιήσεις Α/Ψ και Ψ/Α μετατροπών
 - Υψηλή ευκρίνεια σε σχέση με multi-bit μετατροπείς (>16bit)
 - Δεν απαιτούν κυκλώματα υψηλής ακρίβειας
 - Απαιτούν υπερδειγματοληψία
 - Επεξεργαστική ισχύς
- Υλοποίηση ψηφιακών ενισχυτών
- Βασικά μειονεκτήματα
 - Απουσία μαθηματικής μοντελοποίησης για επεξεργασία
 - Η υψηλή ευαισθησία σε θόρυβο jitter
 - Υψηλή ενέργεια θορύβου σε υψηλές συχνότητες
 - Η δημιουργία φασματικών παραμορφώσεων
 - Εμφάνιση συγκεκριμένων δυαδικών ακολουθιών (bit patterns)
 - Π.χ. {1, 1, -1, -1, 1, 1, -1, -1, ...}



Κωδικοποίηση Σίγμα-Δέλτα (συν.)

- Το φίλτρο βρόχου λειτουργεί ως ολοκληρωτής
- Η έξοδος του συγκριτή είναι 1, εάν $V_i > 0$



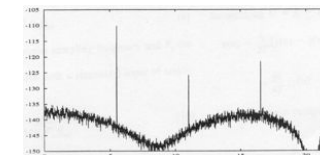
CLOCK	V_a	V_b	V_c	V_d
0	0	0.6	0.6	1
1	1	-0.4	0.2	1
2	1	-0.4	-0.2	0
3	-1	1.6	1.4	1
4	1	-0.4	1.0	1
5	1	-0.4	0.6	1
6	1	-0.4	0.2	1
7	1	-0.4	-0.2	0
8	-1	1.6	1.4	1
9	1	-0.4	1.0	1
10	1	-0.4	1.0	1
11	1	-0.4	0.2	1
12	1	-0.4	-0.2	0

Παράδειγμα λειτουργίας για τάση εισόδου ίση με 0.6V και δυναμική περιοχή λειτουργίας =1V



Κωδικοποίηση Σίγμα-Δέλτα (συν.)

Έξοδος Σ/Δ Α/Ψ μετατροπέα για dc σήμα εισόδου πλάτους 1/512



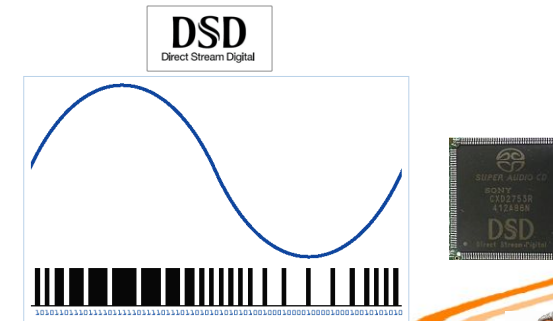
Σύγκριση PCM και Σ/Δ κωδικοποίησης

Απαιτούμενη χωρητικότητα για αποθήκευση στερεοφωνικού σήματος διάρκειας 1 λεπτού

Μέσο αποθήκευσης	CD-DA $f_s=44.1\text{kHz}$ $N=16\text{bit}$	DVD-V $f_s=96\text{kHz}$ $N=24\text{bit}$	DVD-A $f_s=192\text{kHz}$ $N=24\text{bit}$	DSD $f_s=2.82\text{MHz}$ $N=1\text{bit}$
Χωρητικότητα	10.1Mbytes	33Mbytes	66Mbytes	40.4Mbytes



Η κωδικοποίηση Direct Stream Digital (DSD)



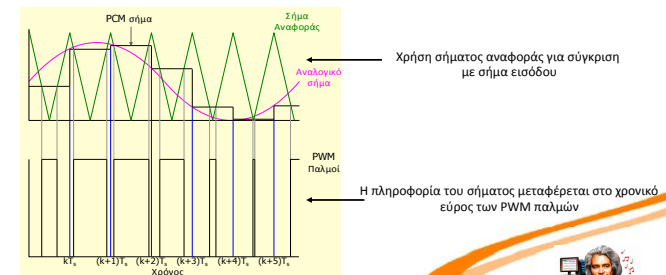
Η κωδικοποίηση PWM

- Διαμόρφωση Παλμών κατά Πλάτος (Pulse Width Modulation)
- Χρησιμοποιείται ευρύτατα σε υλοποιήσεις ψηφιακών ενισχυτών
 - Πολύ υψηλή απόδοση
 - Υψηλότερη από Σίγμα/Δέλτα υλοποιήσεις
 - Υψηλές ταχύτητες λειτουργίας
 - Δεν απαιτείται μετατροπή σε αναλογικό σε οποιοδήποτε στάδιο της αλυσίδας αναπαραγωγής ήχου
- Περιορισμένη χρήση σε υλοποιήσεις 1-bit μετατροπών
 - Προτιμούνται οι Σίγμα/Δέλτα μετατροπής λόγω μειωμένης ταχύτητας λειτουργίας

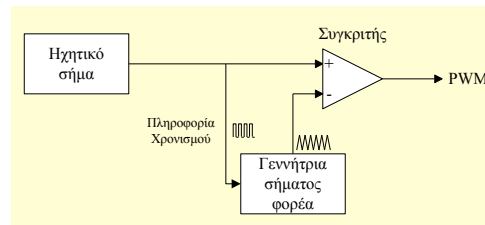


Η PWM κωδικοποίηση στο «αναλογικό» πεδίο

- Πρώτη εφαρμογή στις τηλεπικοινωνίες
 - Υψηλή ανοχή σε προσθετικό θόρυβο



Ο PWM διαμορφωτής

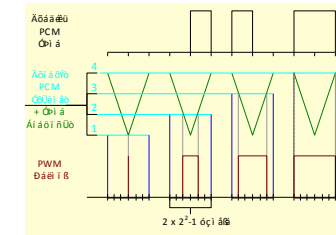


Η PWM κωδικοποίηση στο «ψηφιακό» πεδίο

- Η συχνότητα επανάληψης των παλμών ίση με f_s (Hz)
- Η διατήρηση της πληροφορίας των 2^N PCM σταθμών απαιτεί

$$T_p = \frac{T_s}{2(2^N - 1)}$$

- Άρα, η PWM είναι μια 1-bit κωδικοποίηση με ρυθμό μετάδοσης $f_p = 2(2^N - 1)f_s$ (Hz)



Παράδειγμα PCM-to-PWM mapping για N=2

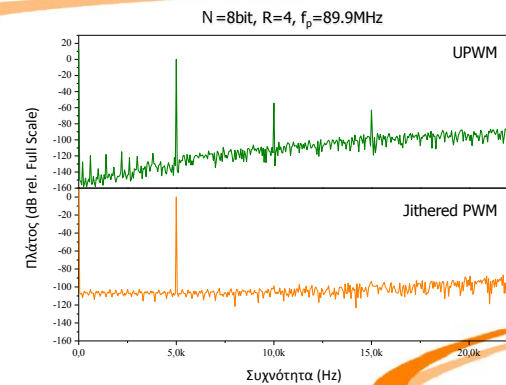


Προβλήματα PWM διαμόρφωσης

- Υψηλοί ρυθμοί μετάδοσης
 - Για $f_s = 44.1\text{kHz}$, $N=16 \rightarrow f_p = 5.78\text{GHz}$
 - Πολύ υψηλή πολυπλοκότητα
 - Κβαντισμός σε $N=8\text{-bit}$
 - $f_p < 100\text{MHz}$
 - Αυξημένος θόρυβος κβαντισμού
 - Λύση η υπερδειγματοληψία
- Φασματική παραμόρφωση
 - Παραμορφώσεις συσχετισμένες με το σήμα
 - Μείωση των παραμορφώσεων με αύξηση της f_s
 - Μερική λύση η υπερδειγματοληψία



Προβλήματα PWM διαμόρφωσης (συν.)



Η PWM κωδικοποίηση στο ψηφιακό πεδίο (συν.)

Demonstration

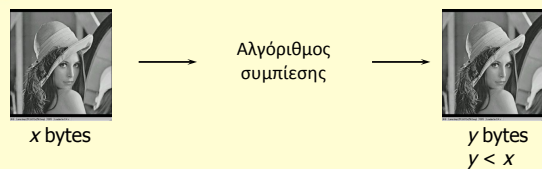
Αρχικό PCM σήμα	$N=16, f_s=44.1\text{kHz}, I=720\text{kbps}$
16-bit PWM	$N=16\text{bit}, f_s=44.1\text{kHz}, f_p=I=5.78\text{GHz}$
8-bit PWM	$N=8\text{bit}, f_s=44.1\text{kHz}, f_p=I=22.491\text{MHz}$
4-bit PWM	$N=4\text{bit}, f_s=44.1\text{kHz}, f_p=I=1.323\text{MHz}$
8-bit Jittered PWM	$N=8\text{bit}, f_s=176.4\text{kHz}, f_p=I=89.96\text{MHz}$



Συμπίεση ψηφιακών ηχητικών δεδομένων



Τι είναι συμπίεση δεδομένων γενικά;



- Λόγος συμπίεσης $C = y/x$
 - Π.χ. Εάν $y=5\text{MB}$ και $x=10\text{MB}$, $C=2:1$
- Γιατί συμπίεση;
 - Μειωμένες απαιτήσεις αποθήκευσης και ταχύτητας μετάδοσης
- Απωλεστική (lossy) ή μη απωλεστική (lossless) συμπίεση;
 - Εξαρτάται από την εφαρμογή (audio/video, data, κ.λ.π.)



Μέθοδοι συμπίεσης ηχητικών δεδομένων

- Μη απωλεστικές μέθοδοι (lossless audio compression)
 - Ακριβής (σε επίπεδο bit) ανακατασκευή των δεδομένων στο δέκτη
 - Μικρός βαθμός συμπίεσης
 - $\sim 2:1$ (π.χ. μέθοδος MLP για το DVD-Audio)
- Απωλεστικές μέθοδοι (lossy audio compression)
 - Απώλεια σημαντικού τμήματος των δεδομένων στο δέκτη
 - Ηχητικά ακριβής ή αποδεκτή αναπαραγωγή
 - Υψηλός βαθμός συμπίεσης
 - $\sim 10:1$ (π.χ. μέθοδος .mp3)



Τεχνικές μη-απωλεστικής συμπίεσης

- Συνήθως υλοποιούνται σε 3 στάδια:
- Γραμμική πρόβλεψη
 - Στόχος η ελαχιστοποίηση της διαφοράς του σήματος εισόδου από την εκτίμηση της τιμής του
 - Η εκτίμηση αυτή παράγεται από τις κβαντισμένες τιμές M δειγμάτων εισόδου
- Κωδικοποίηση εντροπίας
 - Το σήμα διαφοράς κβαντίζεται
 - Οι τιμές με μεγαλύτερη πιθανότητα εμφάνισης κβαντίζονται με μικρότερες ψηφιακές λέξεις
- Πολύπλεξη και δημιουργία πακέτων

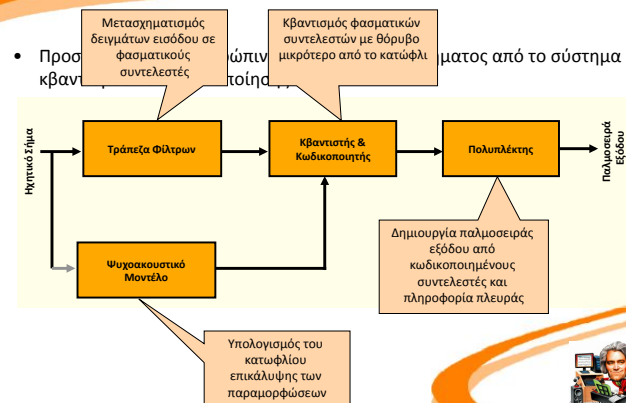


Τεχνικές απωλεστικής συμπίεσης

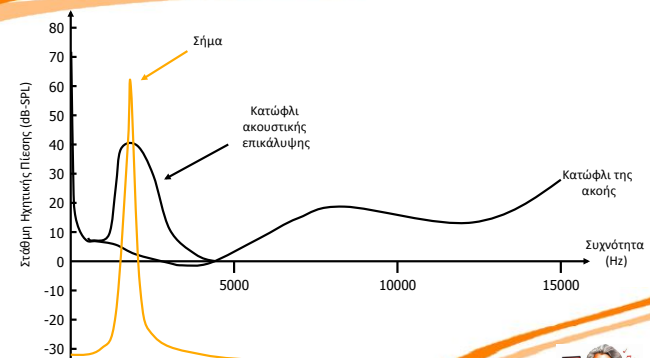
- Συνεχής εξέλιξη από τη δεκαετία του 1960
- Αφαίρεση πλεονασμού από τα δεδομένα
 - Πλεονασμός λόγω θεωρήματος Nyquist
- Μείωση της υποκειμενικά μη ακουστικής πληροφορίας
 - Τι ακούμε πραγματικά;
- Τεχνικές στο πεδίο του χρόνου (time domain)
- Τεχνικές στο πεδίο της συχνότητας (frequency domain)
 - Υποκειμενικές κωδικοποιήσεις
 - Λόγος συμπίεσης μεγαλύτερος από 10:1



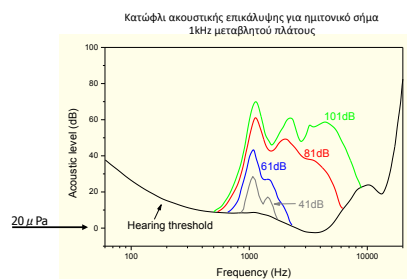
Τεχνικές υποκειμενικής κωδικοποίησης



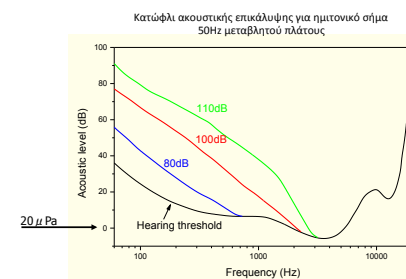
Το φαινόμενο της ακουστικής επικάλυψης



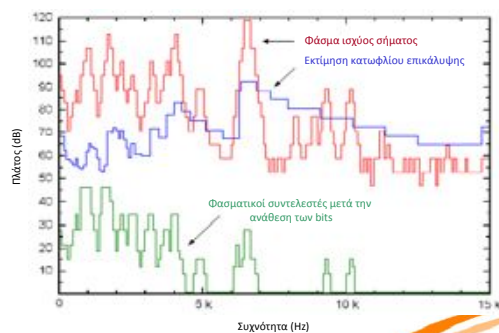
Μοντέλα ακουστικής επικάλυψης



Μοντέλα ακουστικής επικάλυψης (συν.)



Γενικό μοντέλο ψυχοακουστικής συμπίεσης



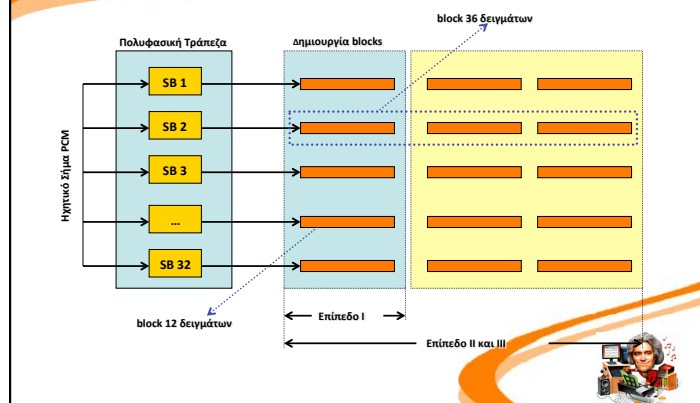
Το πρότυπο MPEG-1

- Πρότυπο ISO/IEC11172-3
 - Χρονοσυχντική ανάλυση
 - Ψυχοακουστικό μοντέλο
 - Δυναμική ανάθεση bits
 - Αρχιτεκτονική 3 επιπέδων (Layer I, II και III)
 - Αυξανόμενη πολυπλοκότητα
 - Καλύτερη ποιότητα και συμπίεση (4:1, 8:1, και 10:1 αντίστοιχα)
 - Συχνότητες δειγματοληψίας 32, 44.1 και 48kHz
 - Μονοφωνικός, στερεοφωνικός και συνδυασμένα στερεοφωνικός ήχος
 - Ρυθμός δεδομένων
 - 32-448kbps (Layer I)
 - 32-384kbps (Layer II)
 - 32-320kbps (Layer III)

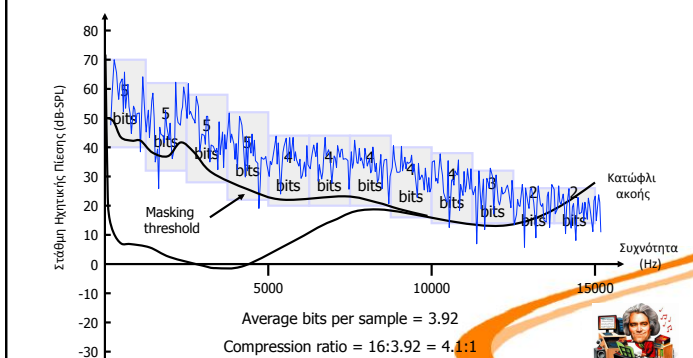
Fraunhofer
Institut
Integrierte Schaltungen



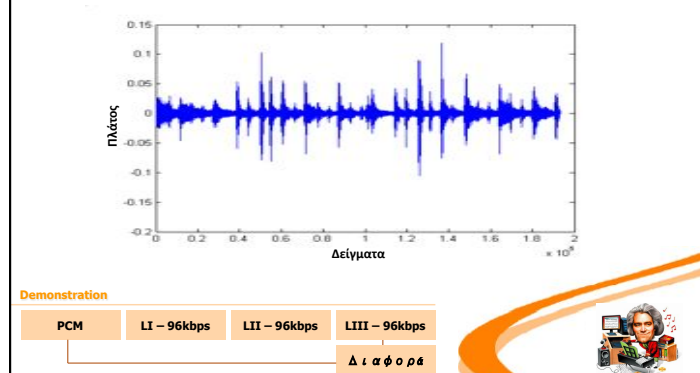
Λειτουργία του προτύπου MPEG-1



Εφαρμογή ψυχοακουστικού μοντέλου



Παράδειγμα συμπίεσης MPEG-1 (Layer I, II και III)



Άλλες τεχνικές υποκειμενικής κωδικοποίησης

- Dolby AC-2 (1989)
 - Εφαρμογές καλωδιακής τηλεόρασης
 - Ρυθμοί 128-192kbps (εύρος ζώνης 20kHz)
 - Χαμηλή πολυπλοκότητα και καθυστέρηση (7-60msec)
- Dolby AC-3
 - Εφαρμογές: HDTV, DVD-Video
 - Υποστήριξη πολυканάλου ήχου
 - Ρυθμοί 32-640kbps
- ATRAC
 - Minidisk
 - Ρυθμός 140kbps ανά κανάλι



