ΔΙΑΦΑΝΕΙΕΣ

ΜΑΘΗΜΑΤΟΣ

ΘΕΩΡΙΑ ΠΛΗΡΟΦΟΡΙΑΣ-ΚΩΔΙΚΕΣ

(ΘΕΩΡΙΑ)

Δρ. Βαρζάκας Παναγιώτης

Επίκουρος Καθηγητής

Τ.Ε.Ι. ΛΑΜΙΑΣ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΝΙΚΗΣ

ΛΑΜΙΑ 2007

- P_i **:** πιθανότητα εμφάνισης του i γεγονότος
- **ποσότητα πληροφορίας που μεταφέρει το γεγονός I**:

$$I \propto \frac{1}{P}$$

$$I = -\log_b P = \log_b\left(\frac{1}{P}\right)$$

Μονάδες ποσότητας πληροφορίας I:

b=2: μονάδα μέτρησης ποσότητας πληροφορίας το *bit (binary unit)*
b=10: μονάδα μέτρησης ποσότητας πληροφορίας το *Hartley*
b=e: μονάδα μέτρησης ποσότητας πληροφορίας το *nats*.

Μέση πληροφορία ανά σύμβολο ενός αλφαβήτου (Εντροπία πηγής)

(αλφάβητο με q ισοπίθανα σύμβολα)

- **ποσότητα πληροφορίας που παρουσιάζει μία φράση από *n* σύμβολα**:

$$I_n = -n \log_2 P = n \log_2 \left(\frac{1}{\frac{1}{q}} \right) = n \log_2 (q) \quad \text{(bits)}$$

- **Μέση πληροφορία ανά σύμβολο Η** (bits/σύμβολο):

(**στατική πηγή*:*** πιθανότητες εμφάνισης όλων των συμβόλων ανεξάρτητες χρόνου**)**

$$H = \bar{I} = -\sum_{i=1}^{q} P_i \cdot \log_2 \left(P_i \right) \quad \text{(bits/σύμβολο)}$$

- Μέγιστη τιμή της εντροπίας H_{max}:

$$H_{\max} = \bar{I}_{\max} = \log_2(q)$$

$$P_i = 1/q, \qquad για \;\; i = 1, 2, ..., q$$

(ισοπίθανα σύμβολα πηγής)

- Ελάχιστη τιμή της εντροπίας H_{min}:

$$H_{\min} \text{ όταν } P_n = 1, \text{ για κάποιο από τα } i = 1,2,...q$$

Πλεονασμός π μιας πηγής πληροφορίας

(ποσό της "άχρηστης πληροφορίας")

$$\pi = \frac{\mathrm{H}_{\max} - H}{H_{\max}} = 1 - \frac{H}{H_{\max}} \qquad (\%)$$

- Εκπομπή συμβόλων με ρυθμό r (r_s) (σύμβολα/sec), τότε ο **ρυθμός παροχής πληροφορίας από την πηγή R:**

$$R = r \cdot H \qquad \text{(bits/sec)}$$

- Αν το i-οστό σύμβολο της πηγής έχει διάρκεια t_i τότε η ***μέση διάρκεια*** των συμβόλων είναι:

$$\bar{\tau} = \sum_{i=1}^{q} P_i \cdot t_i \qquad \text{(sec/σύμβολο)}$$

Χωρητικότητα καναλιού (Channel Capacity)

- **Εντροπία εισόδου X** στο κανάλι επικοινωνίας

(μέση ανά σύμβολο, ποσότητα πληροφορίας στην είσοδο του καναλιού)

$$H(X) = -\sum_{i=1}^{M} P_i^t \cdot \log_2\left(P_i^t\right)$$ (bits/σύμβολο)

(M: πλήθος διακριτών συμβόλων από μία πηγή πληροφορίας)

- Αν ο **ρυθμός εκπομπής συμβόλων** στο κανάλι r_s (σε σύμβολα/sec) τότε ο *μέσος ρυθμός πληροφορίας* D_{in} στην είσοδο του καναλιού επικοινωνίας:

$$D_{in} = H(X) \cdot r_s$$ (bits/sec)

• **Εντροπία υπό συνθήκη** (αμφιβολία αντιστοιχίας των συμβόλων από την έξοδο του καναλιού προς την είσοδο του καναλιού):

$$H(X/Y) = -\sum_{i=1}^{M}\sum_{j=1}^{M} P(X=i,\ Y=j)\cdot \log_2 P(X=i/Y=j)$$

- Ο **μέσος ρυθμός εκπομπής** D_t διαμέσου του καναλιού:

$$D_t = \left[H(X) - H(X/Y)\right]\cdot r_s \qquad \text{(bits/sec)}$$

• **Χωρητικότητα του καναλιού** ***C*** ορίζεται ως η μέγιστη δυνατή τιμή του D_t για όλα τα σύνολα πιθανοτήτων εμφάνισης της εισόδου του καναλιού X:

$$C = \max_{p(x)}\left\{D_t\right\} \qquad \text{(bits/sec)}$$

Χωρητικότητα καναλιού συνεχών μηνυμάτων λευκού προσθετικού θορύβου κατανομής Gauss (AWGN)

- Αν το κανάλι παρουσιάζει λευκό, προσθετικό θόρυβο κατανομής Gauss (Additive White Gaussian Noise, AWGN) τότε αποδεικνύεται ότι η χωρητικότητα C του συγκεκριμένου καναλιού: (Θεωρητικό όριο)

(*Θεώρημα Shannon-Hartley*)

$$C = B \cdot \log_2\left(1 + \frac{S}{N}\right) \quad \text{(bits/sec)}$$

- Ο ρυθμός R της εκπεμπόμενης πληροφορίας στο κανάλι θα πρέπει να ικανοποιεί την επόμενη ανισότητα:

$$R \le C = B \cdot \log_2\left(1 + \frac{S}{N}\right) \quad \text{(bits/sec)}$$

- **B: εύρος ζώνης συχνοτήτων καναλιού επικοινωνίας** (channel bandwidth)

- **S/N ή SNR (Signal-to-Noise Ratio)**
  Λόγος ισχύος σήματος S προς ισχύ θορύβου N στην έξοδο του καναλιού:

S/N=Ισχύς σήματος (σε W)/ Ισχύς θορύβου καναλιού (σε W)

- **C: χωρητικότητα καναλιού (channel capacity) (bits/sec)**

(**μέγιστος δυνατός ρυθμός μετάδοσης πληροφορίας** στο κανάλι ώστε ο ρυθμός λαθών, με μία "πολύπλοκη" κωδικοποίηση, να τείνει σε μηδενική τιμή)

- **Πρακτικά**: R<C/5 ή R<C/4

$$\frac{S}{N}(dB) = 10 \cdot \log_{10}\left(\frac{S}{N}\right)(\text{καθαρός αριθμός})$$

$$\frac{S}{N}(\text{καθαρός αριθμός}) = 10^{\frac{SNR(dB)}{10}}$$

$$\log_2 x = \frac{\log_{10} x}{\log_{10} 2} = 3.32 \cdot \log_{10} x$$

Συμπεράσματα

- Αύξηση της χωρητικότητας C μπορεί να πραγματοποιηθεί με αύξηση του εύρους ζώνης συχνοτήτων του καναλιού B

ή/και

- Αύξηση της χωρητικότητας C μπορεί να πραγματοποιηθεί με αύξηση του λόγου **SNR**

- **Αποδεικνύεται:**

$$\lim_{B\to\infty} C = 1.44 \cdot \frac{S}{N_0}$$

N_0:φασματική πυκνότητα θορύβου καναλιού

(power noise spectral density)(σταθερή για AWGN κανάλι) (Watt/Hz)

(δεδομένη τιμή για ένα κανάλι)

Προσοχή: η αύξηση στο άπειρο του εύρους ζώνης συχνοτήτων B του καναλιού επικοινωνίας, δεν οδηγεί σε αντίστοιχη αύξηση προς το άπειρο της χωρητικότητας C του καναλιού επικοινωνίας

Εντροπία πηγής με μνήμη

- **Πηγή πληροφορίας με μνήμη 1ης τάξης**

(πηγή στην οποία η εκπομπή ενός συμβόλου εξαρτάται από το προηγούμενο σύμβολο που εκπέμφθηκε):

$$H_{\text{πηγή με μνήμη 1ης τάξης}} = \sum_i \sum_j P_i \cdot P(j/i) \cdot \log_2 \left(\frac{1}{P(j/i)} \right)$$

- $P(j/i) = P_{ij}$: πιθανότητα να εκπεμφθεί το j σύμβολο της πηγής δεδομένου ότι έχει σταλεί ήδη το i σύμβολο της πηγής

(μνήμη ενός συμβόλου, *m*=1)

- ο ρυθμός εκπομπής πληροφορίας στο κανάλι για **μία πηγή με μνήμη**:

$$R = r \cdot H_{\text{πηγή με μνήμη 1ης τάξης}}$$

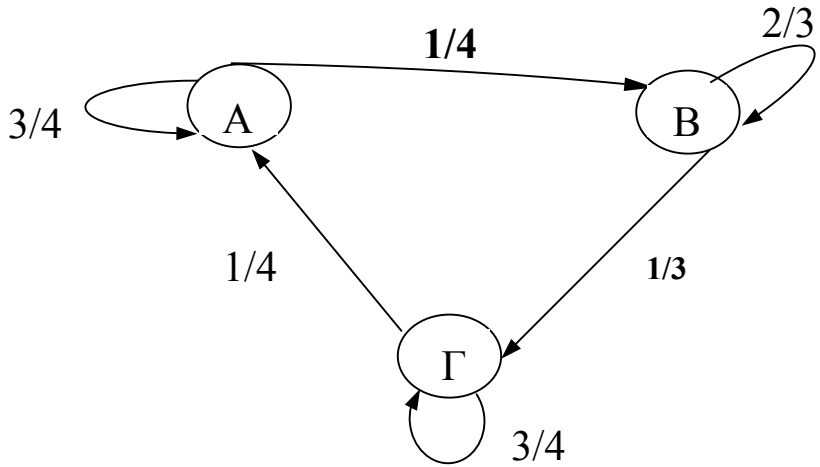
- Οι μεταπτώσεις της μπορούν να περιγραφούν με τη ***μήτρα πιθανοτήτων μετάπτωσης*** δηλαδή τη μήτρα των: $P(j/i)=P_{ij}$
- Μία πηγή με μνήμη q συμβόλων μπορεί να περιγραφτεί με μήτρα P τα στοιχεία της οποίας είναι τιμές πιθανοτήτων μετάπτωσης:

$$P=\begin{bmatrix} p_{1,1} & p_{1,2} & \cdot & \cdot & \cdot p_{1,q} \\ p_{2,1} & p_{2,2} & \cdot & \cdot & \cdot p_{2,q} \\ & & & & \\ p_{q,1} & p_{q,2} & \cdot & \cdot & \cdot p_{q,q} \end{bmatrix}$$

$$P=\begin{bmatrix} & \text{A} & \text{B} & \Gamma \\ \text{A} & 3/4 & 0 & 1/4 \\ \text{B} & 1/4 & 2/3 & 0 \\ \Gamma & 0 & 1/3 & 3/4 \end{bmatrix}$$

[π.χ. η πιθανότητα 3/4 της πρώτης γραμμής και πρώτης στήλης είναι η πιθανότητα να σταλεί το σύμβολο Α δεδομένου ότι έχει ήδη σταλεί το σύμβολο Α, η πιθανότητα 1/3 της τρίτης γραμμής και δεύτερης στήλης ερμηνεύεται ως η πιθανότητα να σταλεί το σύμβολο Γ δεδομένου ότι έχει σταλεί το σύμβολο Β]

Η προηγούμενη μήτρα μεταπτώσεων ισοδυναμεί με το επόμενο ***διάγραμμα καταστάσεων*** της πηγής:



$$P_A = P_{A/A} \cdot P_A + P_{A/B} \cdot P_B + P_{A/\Gamma} \cdot P_\Gamma$$

$$P_B = P_{B/A} \cdot P_A + P_{B/B} \cdot P_B + P_{B/\Gamma} \cdot P_\Gamma$$

$$P_\Gamma = P_{\Gamma/A} \cdot P_A + P_{\Gamma/B} \cdot P_B + P_{\Gamma/\Gamma} \cdot P_\Gamma$$

[σύστημα 3 εξισώσεων με 3 αγνώστους και με αγνώστους τα P_A, P_B, και P_Γ]

[μπορεί να λυθεί είτε με τη μέθοδο της αντικατάστασης είτε με τη μέθοδο των οριζουσών]

Είδη κωδικοποίησης δεδομένων

- *κωδικοποίηση πηγής* (source coding)
- *κωδικοποίηση καναλιού* (channel coding)

• **κωδικοποίηση πηγής:** διαδικασία που ακολουθεί αμέσως μετά την έξοδο μιας πηγής πληροφορίας. Σκοπός ελαχιστοποίηση του απαιτούμενου μήκους της κωδικής λέξης που απαιτείται για να αναπαρασταθούν κατά μέσο όρο τα σύμβολα που παράγει η πηγή της πληροφορίας. (μείωση πλήθους ψηφίων του κώδικα που απαιτείται για να αναπαρασταθούν τα σύμβολα της πηγής)

• **κωδικοποίηση καναλιού:** σκοπός η όσο το δυνατό αξιόπιστη μετάδοση δεδομένων σε ένα κανάλι με θόρυβο (**ενθόρυβο**)
(μείωση της επίδρασης του θορύβου του καναλιού στα εκπεμπόμενα δεδομένα με την ανίχνευση ή/και τη διόρθωση των λαθών (μείωση πιθανότητας λάθους)

Κωδικοποίηση πηγής Huffman

Βήματα:

1. τοποθετούμε σε κατακόρυφη διάταξη τα εκπεμπόμενα σύμβολα της πηγής με σειρά **φθίνουσας πιθανότητας εμφάνισης** των συμβόλων (το άθροισμα των πιθανοτήτων των εκπεμπόμενων συμβόλων θα πρέπει να είναι **πάντα** ίσο με 1)
2. ξεκινώντας από τη μικρότερη πιθανότητα συμβόλου και την αμέσως μικρότερη από αυτή, τις προσθέτουμε και η νέα τιμή πιθανότητας που προκύπτει την τοποθετούμε στη σωστή θέση στην κατακόρυφη διάταξη φθίνουσας πιθανότητας εμφάνισης συμβόλων παράγοντας έτσι δίπλα στη προηγούμενη μία νέα κατακόρυφη διάταξη
3. η διαδικασία αυτή συνεχίζεται έως να έχουμε μόνο δύο τιμές πιθανοτήτων στην κατακόρυφη διάταξη
4. ακολουθούμε αντίστροφη διαδικασία τοποθετώντας δίπλα από τις δύο τελευταίες τιμές πιθανότητας το 0 και το 1

5. τα ψηφία 0 ή 1 (που έχουμε τοποθετήσει δίπλα στις δύο τελευταίες τιμές πιθανοτήτων) οδηγούνται με **αντίστροφη πορεία** και σημειώνονται δίπλα στις τιμές των πιθανοτήτων από τις οποίες οδηγηθήκαμε σε αυτά

6. δίπλα στα ψηφία 0 ή 1 που έχουμε μεταφέρει, τοποθετούμε το 0 και το 1

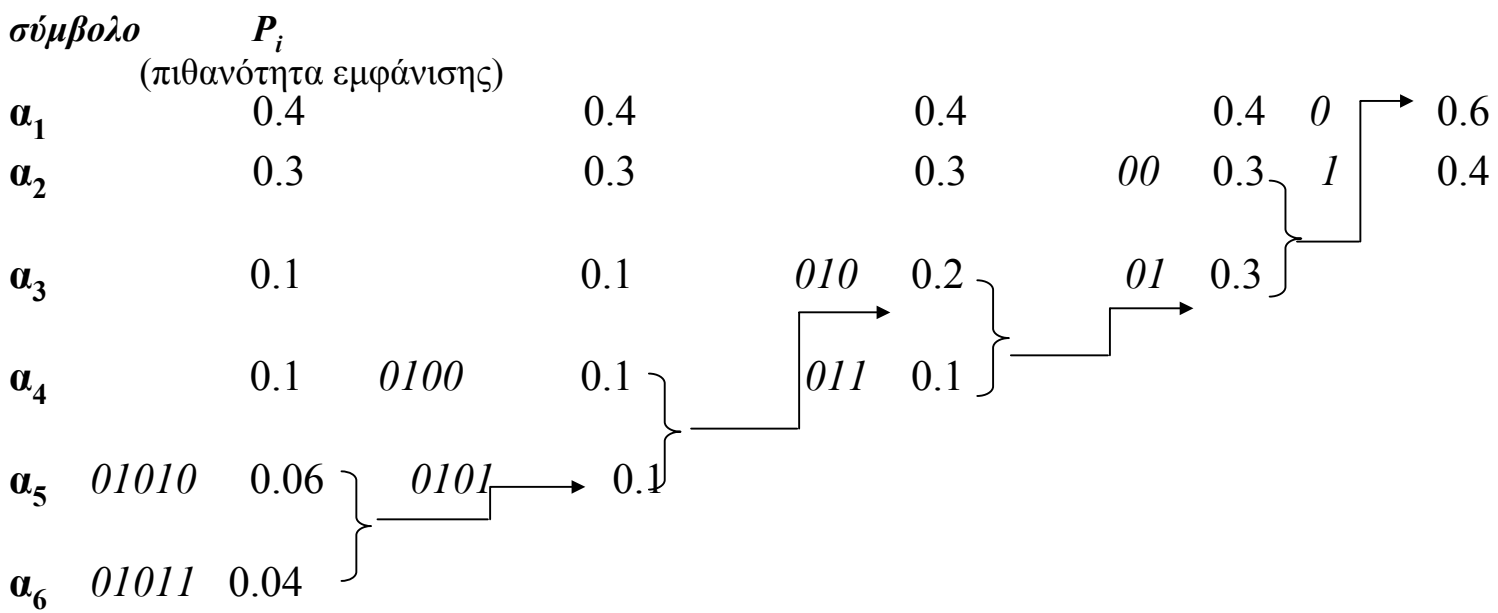
7. ο συνδυασμός των 0 και 1 που έχει τώρα εμφανιστεί επαναλαμβάνεται με την ίδια διαδικασία όπως και πριν έγινε για το 0 και το 1

8. τα προηγούμενα βήματα επαναλαμβάνονται μέχρι να καταλήξουμε στην αρχική διάταξη φθίνουσας πιθανότητας εμφάνισης των συμβόλων

9. οι κωδικές λέξεις που αντιστοιχούν στην κωδικοποίηση είναι αυτές που έχουν απομείνει από τους προηγούμενους συνδυασμούς, χωρίς να έχουν μεταφερθεί προς τα πίσω (προς τα αριστερά), με τη διαδικασία που περιγράφηκε στα προηγούμενα βήματα.

Παράδειγμα

- πηγή 6 συμβόλων τα οποία παρουσιάζουν τις επόμενες πιθανότητες εμφάνισης: α_1:0.4, α_2:0.3, α_3:0.1, α_4:0.1, α_5:0.06, α_6:0.04.



$\alpha_1 \longrightarrow 1$

$\alpha_2 \longrightarrow 00$

$\alpha_3 \longrightarrow 011$

$\alpha_4 \longrightarrow 0100$

$\alpha_5 \longrightarrow 01010$

$\alpha_6 \longrightarrow 01011$

- οι συνδυασμοί που έχουν απομείνει χωρίς να έχουν μεταφερθεί αριστερά (αυτοί οι οποίοι δεν προέρχονται από ένα άθροισμα πιθανοτήτων δηλαδή δεν καταλήγουμε σε αυτούς με την πορεία ενός βέλους) και οι οποίοι τελικά είναι οι κωδικές λέξεις που αντιστοιχούν στα σύμβολα της πηγής

Κωδικοποίηση πηγής Shannon-Fano

Βήματα:

1. καταγράφουμε τα σύμβολα της πηγής κατά σειρά **μειούμενης** πιθανότητας

2. χωρίζουμε το σύνολο σε 2 ομάδες που να είναι όσο το δυνατό πλησιέστερα σε ίσες πιθανότητες (δηλαδή το άθροισμα σε κάθε μία ομάδα) και θέτουμε 0 στην ανώτερη ομάδα και 1 στην κατώτερη ομάδα

3. συνεχίζουμε χωρίζοντας κάθε φορά τις ομάδες με όσο το δυνατό ίσες πιθανότητες μέχρι να μην είναι δυνατός ο περαιτέρω διαχωρισμός τους.

• **Μειονέκτημα: αβεβαιότητα** που υπάρχει στο "**μοίρασμα**" των 2 ομάδων.

Παράδειγμα

- πηγή 6 συμβόλων τα οποία παρουσιάζουν τις επόμενες πιθανότητες εμφάνισης: x_1:0.3, x_2:0.25, x_3:0.2, x_4:0.12, x_5:0.08, x_6:0.05.

(πιθανότητα εμφάνισης)

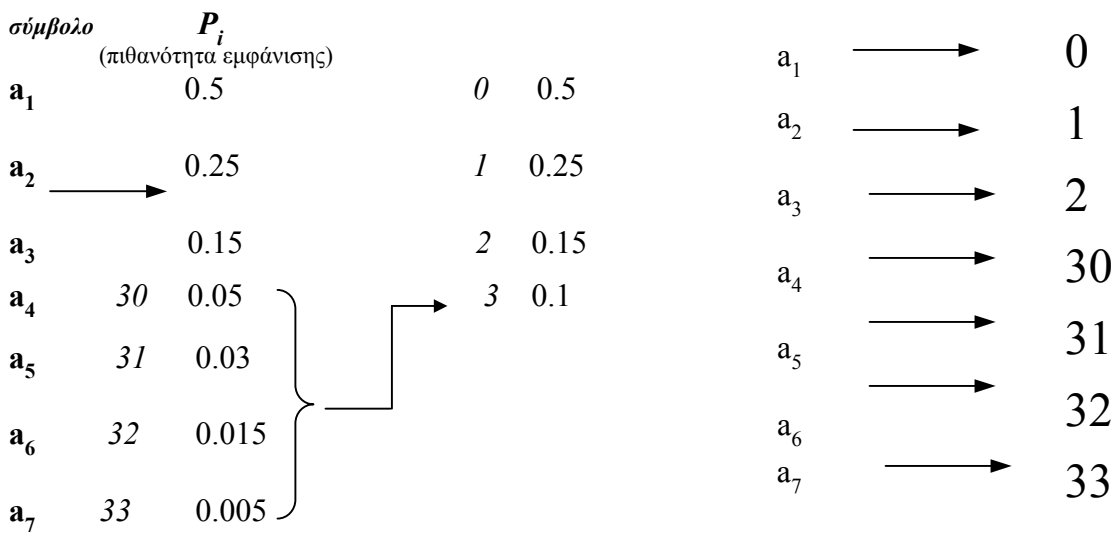
σύμβολο	P_i	**Βήμα 1ο**	**Βήμα 2ο**	**Βήμα 3ο**	**Βήμα 4ο**	**Βήμα 5ο**
x_1	0.3	0	0			**00**
x_2	0.25	0	1			**01**
x_3	0.20	1	0			**10**
x_4	0.12	1	1	0		**110**
x_5	0.08	1	1	1	0	**1110**
x_6	0.05	1	1	1	1	**1111**

$x_1 \longrightarrow$ **00**
$x_2 \longrightarrow$ **01**
$x_3 \longrightarrow$ **10**
$x_4 \longrightarrow$ **110**
$x_5 \longrightarrow$ **1110**
$x_6 \longrightarrow$ **1111**

Τετραδική Κωδικοποίηση πηγής Huffman

- κωδικοποίηση πηγής Huffman στο τετραδικό σύστημα το οποίο έχει ως σύμβολα τα: 0,1,2,3 (**τέσσερα σύμβολα**)

- διαδικασία που η ίδια όπως και στο δυαδικό σύστημα αλλά οι πιθανότητες αθροίζονται στη κατακόρυφη διάταξη ανά **τέσσερις**



Μέσο μήκος κωδικής λέξης

Σύμβολο πηγής ⟶ *κωδική λέξη (codeword)*

$$L = \sum_{i=1}^{q} P_i \cdot l_i$$

(bits/κωδική λέξη του κώδικα)

- l_i είναι το μήκος (σε bits) της i-οστής κωδικής λέξης του κώδικα

- P_i : πιθανότητα εμφάνισης i-οστού συμβόλου (κωδικής λέξης)

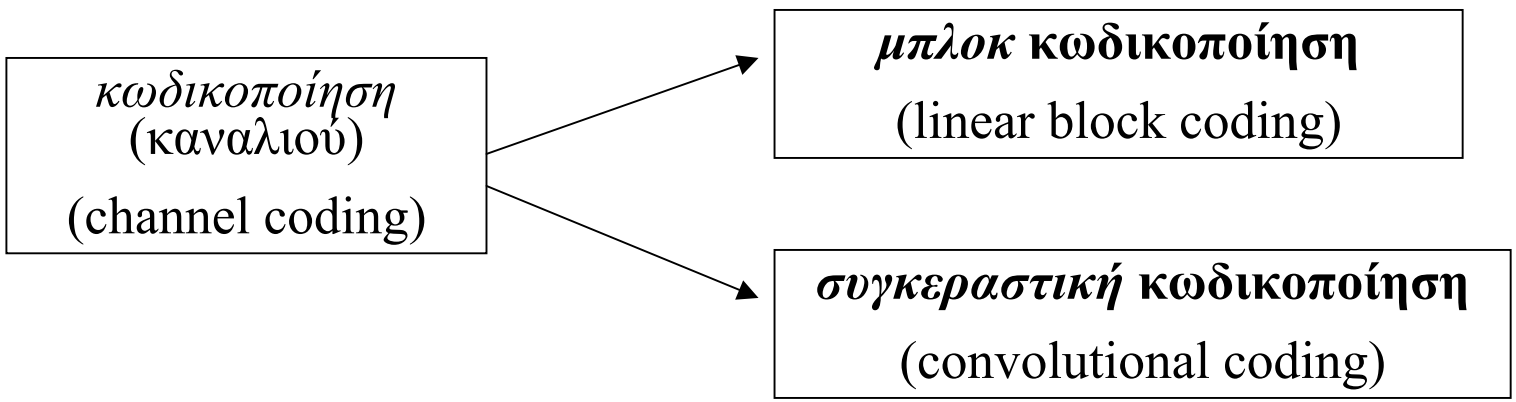
Κωδικοποίηση (καναλιού)

(channel coding)

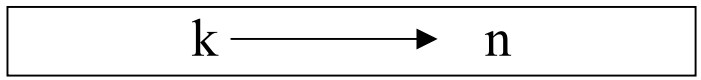
- μετάδοση δεδομένων μέσω ενός καναλιού επικοινωνίας με θόρυβο έχει ως αποτέλεσμα την εμφάνιση λαθών (bit error)

- **Λύση: κωδικοποίηση: πρόσθεση επιπλέον bits**

- **συνέπεια:** μείωση του ρυθμού εκπομπής δεδομένων



- μήνυμα πληροφορίας μήκους k απεικονίζεται σε δυαδικές ακολουθίες (**κωδικές λέξεις**) μήκους n:



(πρόσθεση επιπλέον bits: bits ελέγχου ισοτιμίας)

- ρυθμός ή **απόδοση κώδικα**: $\frac{k}{n}$

(ποσοστό "χρήσιμης" πληροφορίας που εκπέμπεται στα συνολικά n bits κώδικα)

- **κώδικας *(n,k)***: αποτελείται από 2^k κωδικές λέξεις με μήκος n

Απλοί Κώδικες Επανάληψης

(Simple Repetition Codes)

- αντί εκπομπή “0” και “1”, **εκπομπή ακολουθίας** από “0” και “1” στη θέση **αντίστοιχα** του “0” και του “1” (**επανάληψη** του ίδιου bit περιττές φορές)

- μήκος των ακολουθιών (πλήθος bits του κώδικα) επιλέγεται να είναι ένας **περιττός** αριθμός *n*

$$0 \rightarrow \overbrace{00 \ldots 00}^{n \text{ περιττός}}$$

$$1 \rightarrow \overbrace{11 \ldots 11}^{n \text{ περιττός}}$$

Πιθανότητα λάθους αποκωδικοποίησης για απλό κώδικα επανάληψης (*n*,*k*)

$$p_e = \sum_{k=(n+1)/2}^{n} \binom{n}{k} \cdot \varepsilon^k \cdot (1-\varepsilon)^{n-k}$$

- **ε: πιθανότητα λάθους στο εκπεμπόμενο bit (σε απλά εκπεμπόμενο bit στο κανάλι επικοινωνίας)**

Παράδειγμα: απλός κώδικας επανάληψης με n=5, k=3 και ε=0.001=10^{-3}, πιθανότητα λάθους αποκωδικοποίησης:

$$p_e = \sum_{k=3}^{5} \binom{5}{k} \cdot 0.001^k \cdot (0.999)^{5-k} = 9.99 \times 10^{-10} \cong 10^{-9}$$

$$\binom{n}{k} \overset{\text{ορ}}{=} \frac{n!}{k!\,(n-k)!}$$

$$n! = 1 \cdot 2 \cdot \ldots \quad \text{n}$$

Γραμμικοί κώδικες Μπλοκ

- ***γραμμικός:*** ένας οποιαδήποτε γραμμικός συνδυασμός δύο κωδικών λέξεων του κώδικα αποτελεί, κωδική λέξη του κώδικα

- πίνακας γεννήτορας G: $k \times n$ δυαδικός πίνακας
- κωδική λέξη ***c*** του κώδικα: $c = uG$

u: ακολουθία εισόδου μήκους k (**μήνυμα**) (είσοδος κωδικοποιητή)

πράξη modulo-2: (πράξη XOR) (αποκλειστικό Η)

- **δυνατότητα διόρθωσης λαθών**: *ελάχιστη απόσταση* (απόσταση Hamming) του κώδικα

- **ελάχιστη απόσταση (απόσταση Hamming) του κώδικα:** ελάχιστη απόσταση Hamming μεταξύ δύο οποιονδήποτε κωδικών λέξεων του κώδικα

- **ελάχιστη απόσταση του κώδικα:** $d_{\min}$

$$d_{\min} = \min_{i \neq j} d_H\left(c_i, c_j\right)$$

- **γραμμικοί κώδικες:** $d_{\min} \longrightarrow$ $$w_{\min} = \min_{C_i \neq 0} w_H\left(c_i\right)$$

- **βάρος** w_H μιας κωδικής λέξης: **πλήθος των "1" της κωδικής λέξης**

- **ανίχνευση ή/και διόρθωση λαθών στη λαμβανόμενη κωδική λέξη:**

$$d_{\min} = \begin{Bmatrix} e+1, \text{για ανίχνευση } e \text{ σφαλμάτων ανά κωδική λέξη} \\ 2e+1, \text{για διόρθωση } e \text{ σφαλμάτων ανά κωδική λέξη} \end{Bmatrix}$$

κανόνας αποκωδικοποίησης: απόφαση ποια είναι η σωστή εκπεμπόμενη κωδική λέξη:

"*Επέλεξε ως σωστή εκείνη την κωδική λέξη που παρουσιάζει τη μικρότερη ελάχιστη απόσταση Hamming από την υπό κρίση κωδική λέξη που έχει ληφθεί*"

("**αποκωδικοποίηση αυστηρής απόφασης**")

(Hard-decision Decoding)

(κριτήριο ελάχιστης απόφασης Hamming)

- **γραμμικός κώδικας μπλοκ σε συστηματική μορφή:**

$$G=\begin{bmatrix} 1\ 0\ \ldots.\ 0 & p_{1,1} & p_{1,2} & \cdot\ \cdot & p_{1,n-k} \\ 0\ 1\ \ldots.\ 0 & p_{2,1} & p_{2,2} & \cdot\ \cdot & p_{2,n-k} \\ & & & & \\ 0\ 0\ \ldots..1 & p_{k,1} & p_{k,2} & \cdot\ \cdot & p_{k,n-k} \end{bmatrix}$$

$$G=[I_k \mid P]$$

I_k : $(k\times k)$ μοναδιαίος πίνακας

P: $k\times(n-k)$ πίνακας

- **συστηματικό κώδικα**: τα πρώτα k bits της κωδικής λέξης είναι τα bits της μηνύματος και τα υπόλοιπα $(n-k)$ bits είναι τα ***bits ελέγχου της ισοτιμίας***

- **πίνακας ελέγχου της ισοτιμίας (parity check matrix) H:**

$$cH^t = 0$$

↓

(εύρεση κωδικής λέξης c)

H^t : ***ανάστροφος*** *πίνακας* του πίνακα H (Transpose Matrix)
(γραμμές του πίνακα H τις κάνουμε στήλες και αντίστροφα)

Ισχύει: $GH^t = 0$

Αν ο πίνακας γεννήτορας G είναι σε συστηματική μορφή ισχύει:

$$H=[-P^t|I_{n-k}]$$

Κώδικες Hamming

- **κώδικες Hamming:** είναι γραμμικοί κώδικες μπλοκ, διαστάσεων (2^m-1, 2^m-m-1) που παρουσιάζουν ελάχιστη απόσταση ίση με 3

- **πίνακας ελέγχου ισοτιμίας H:** πίνακας $m \times (2^m-1)$ πίνακας, με στήλες όλες τις δυαδικές ακολουθίες με μήκος m, εκτός από την μηδενική ακολουθία

Παράδειγμα: για m=3 έχουμε έναν (7,4) κώδικα του οποίου ο πίνακας ελέγχου της ισοτιμίας, σε συστηματική μορφή, είναι:

$$H=\begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \longrightarrow G=\begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

(σε συστηματική μορφή)
(δηλαδή $G=[I_4 \mid P]$)

- **κώδικας Hamming (*n*,*k*)**

Αν ο **πίνακας ελέγχου της ισοτιμίας** του κώδικα H είναι στην μορφή:

$$H=[-P^t| I_k]$$

τότε ο **πίνακας γεννήτορας *G*** του κώδικα έχει αντίστοιχα τη μορφή:

$$G=[I_k | P]$$

Σύνδρομο S

Χρησιμότητα συνδρόμου: με δεδομένο το σύνδρομο μπορούμε να ανιχνεύσουμε αν υπάρχει **απλό λάθος** στη λαμβανομένη κωδική λέξη και να διορθώσουμε

- Το σύνδρομο που υπολογίζεται αποτελεί μία στήλη του πίνακα ελέγχου ισοτιμίας H
- Η συγκεκριμένη στήλη δείχνει τη θέση στη λαμβανόμενη λέξη που έχει συμβεί το απλό λάθος

- *R:* **λαμβανόμενη κωδική λέξη**

Εύρεση συνδρόμου: ➡ $R \cdot H^t = S$

- **"σωστή" κωδική λέξη** (Corrected Codeword):

⬇

Corrected Codeword= $R \oplus S$

$\oplus$ *:modulo-2* (αποκλειστικό ή)

Κυκλικοί κώδικες

- **κυκλικός** (γραμμικός κώδικας) μία οποιαδήποτε ***κυκλική ολίσθηση*** μιας κωδικής του λέξης να αποτελεί και αυτή κωδική λέξη του κώδικα

$$C = \left[c_{n-1}\ c_{n-2}\ \ldots c_1 c_0\right]$$: **κωδική λέξη** του κυκλικού κώδικα

- αντιστοιχούμε σε κάθε κωδική λέξη ένα πολυώνυμο $C(p)$ βαθμού $\leq n-1$

$$C(p) = c_{n-1}p^{n-1} + c_{n-2}p^{n-2} + \ldots + c_1 p + c_0$$

- **πολυώνυμο γεννήτορας** βαθμού (n-k) : $g(p)$

$$\boxed{g(p) = p^{n-k} + g_{n-k-1}p^{n-k-1} + \ldots + g_1 p + 1}$$

- **πολυώνυμο του μηνύματος:** $X(p) = x_{k-1}p^{k-1} + x_{k-2}p^{k-2} \ldots + x_1 p + x_0$

όπου το $[x_{k-1}\ x_{k-2} \ldots x_1 x_0]$ αναπαρασταίνει τα k bits του μηνύματος

- **γινόμενο των πολυωνύμων** : $X(p) \cdot g(p)$

αναπαρασταίνει μία κωδική λέξη του κυκλικού κώδικα

Παράδειγμα

- κώδικας με μήκος n=7 (σύνολο των bits μετά την κωδικοποίηση) και k=4 (bits μηνύματος)

Το πολυώνυμο p^7+1 μπορεί να γραφτεί ως **γινόμενο παραγόντων**:

$$p^7+1=(p+1)\cdot(p^3+p^2+1)\cdot(p^3+p+1)$$

- θεωρούμε **ένα** από τα επόμενα πολυώνυμα **ως πολυώνυμο γεννήτορα**:

$$g_1(p)=(p^3+p^2+1)$$

ή

$$g_2(p)=(p^3+p+1)$$

(οι κώδικες οι οποίοι μπορούν να παραχθούν από τα δύο προηγούμενα πολυώνυμα, είναι *ισοδύναμοι*)

Για παράδειγμα, τα μηνύματα [0001] και [1110] κωδικοποιούνται αντίστοιχα μέσω του πολυωνύμου $g_1(p)=(p^3+p^2+1)$ ως [0001101] και [1000110]

Η πρόσθεση '+' μεταξύ των διαφόρων παραγόντων των πολυωνύμων είναι **πράξη modulo 2** (λογική πράξη αποκλειστικού Η, EXOR). Συνεπώς, όταν μετά το συνήθη πολλαπλασιασμό εμφανίζονται **δύο ίδιοι παράγοντες τότε αυτοί απαλείφονται μεταξύ τους ανά δύο** π.χ. $p^3+p^3=0$.

- κυκλικός κώδικας (n,k). Τότε ο πίνακας γεννήτορας G, δίνεται σε **συστηματική** μορφή:

$$G=[I_k|\ P]$$

- πίνακας ελέγχου ισοτιμίας H του κυκλικού κώδικα (n,k), δίνεται σε **συστηματική** μορφή:

$$H=[I_{n-k}\ |P^t]$$

BCH Κώδικες

- **BCH** (Bose-Chaudhuri-Hocquenghem) κώδικες: κατηγορία κυκλικών κωδίκων που περιλαμβάνουν **δυαδικά** και **μη δυαδικά αλφάβητα**

- **BCH κώδικες (*n*,*k*):**

$$n = 2^m - 1$$
$$n - k \le mt$$
$$d_{\min} = 2t + 1$$

όπου m ($m \ge 3$) και t είναι αυθαίρετοι θετικοί αριθμοί.

$$t < (2^m - 1)/2$$

- δυνατότητα επιλογής από ένα **μεγάλο σύνολο** από μήκη κωδίκων και ρυθμών κωδίκων

- πολλές εφαρμογές στα **τηλεπικοινωνιακά συστήματα**

(**κυψελωτά συστήματα κινητής τηλεφωνίας**)

(**σήματα σηματοδοσίας**: την ισχύ που πρέπει να εκπέμψει ο κινητός σταθμός και σε ποια συχνότητα του συστήματος)

Κώδικες Reed-Solomon

- **μη δυαδικοί κώδικες** (1960)
- μεγάλη σημασία για τηλεπικοινωνιακά συστήματα (σφάλματα λόγω θορύβου καναλιού επικοινωνίας κατά ριπές) και για συστήματα ακουστικών CD
- **μπλοκ κώδικες** οι οποίοι χρησιμοποιούν αλφάβητα εισόδου και εξόδου με πλήθος συμβόλων 2^m
- **μήκος της κωδικής λέξης *n***: ακέραιες τιμές μεταξύ 3 και 2^m-1
- διορθώνουν e_0 **λάθη** σε ένα μπλοκ από n σύμβολα
- είναι δυνατό **να διορθώσει μέχρι** $t=(n-k)/2$ λάθη
- **bits ελέγχου ισοτιμίας**: $(n-k)=n-2e_0=2^m-1$
- **ελάχιστη απόσταση**: $d_{min}=(n-k+1)$

Ταξινόμηση κωδίκων

x_i	Κώδικας 1	Κώδικας 2	Κώδικας 3	Κώδικας 4	Κώδικας 5	Κώδικας 6
x_1	00	00	0	0	0	1
x_2	01	01	1	10	01	01
x_3	00	10	00	110	011	001
x_4	11	11	11	111	0111	0001

- **Κώδικες σταθερού μήκους**: είναι ο κώδικας που κάθε κωδική του λέξη έχει σταθερό μήκος. Οι κώδικες 1 και 2 έχουν σταθερό μήκος 2
- **Κώδικες μεταβλητού μήκους**:κώδικας μεταβλητού μήκους είναι ο κώδικας του οποίου το μήκος της κωδικής λέξης δεν είναι σταθερό. Όλοι οι κώδικες του πίνακα, εκτός από τους κώδικες 1 και 2, είναι μεταβλητού μήκους
- **Ευκρινείς κώδικες**:ένας κώδικας ονομάζεται ευκρινής αν κάθε κωδική λέξη του ξεχωρίζει από τις άλλες κωδικές λέξεις. Όλοι οι κώδικες του πίνακα εκτός από τον κώδικα 1 είναι ευκρινείς

- **Κώδικες χωρίς πρόθεμα:** κώδικας στον οποίο δεν σχηματίζεται μία κωδική λέξη με πρόσθεση κωδικών συμβόλων σε άλλη κωδική λέξη ονομάζεται κώδικας χωρίς πρόθεμα. Οι κώδικες 2, 4 και 6 του πίνακα είναι κώδικες χωρίς πρόθεμα
- **Μοναδικά αποκωδικοποιούμενοι κώδικες**:ένας κώδικας είναι μοναδικά αποκωδικοποιούμενος αν η αρχική ακολουθία πηγής μπορεί να αναδομηθεί τέλεια από την κωδικοποιημένη δυαδική ακολουθία. Στον πίνακα, ο κώδικας 3 δεν είναι μοναδικά αποκωδικοποιούμενος κώδικας γιατί π.χ. η δυαδική ακολουθία 1001 μπορεί να αντιστοιχεί στις ακολουθίες πηγής $x_2\ x_3\ x_2$ ή $x_2\ x_1\ x_1\ x_2$. Στον πίνακα ο κώδικας 5 είναι μοναδικά αποκωδικοποιούμενος επειδή το bit 0, δείχνει την αρχή κάθε κωδικής λέξης του κώδικα
- **Στιγμιαίοι κώδικες**:ένας μοναδικά αποκωδικοποιούμενος κώδικας ονομάζεται στιγμιαίος κώδικας αν το τέλος οποιασδήποτε κωδικής λέξης αναγνωρίζεται χωρίς να εξεταστούν επόμενα κωδικά σύμβολα. Οι στιγμιαίοι κώδικες έχουν την ιδιότητα ότι καμία κωδική λέξη δεν είναι πρόθεμα κάποιας άλλης κωδικής λέξης
- **Βέλτιστοι κώδικες**: ένας κώδικας είναι βέλτιστος αν είναι στιγμιαίος και έχει ελάχιστο μέσο μήκος για δεδομένη κατανομή πιθανοτήτων για τα σύμβολα της πηγής πληροφορίας

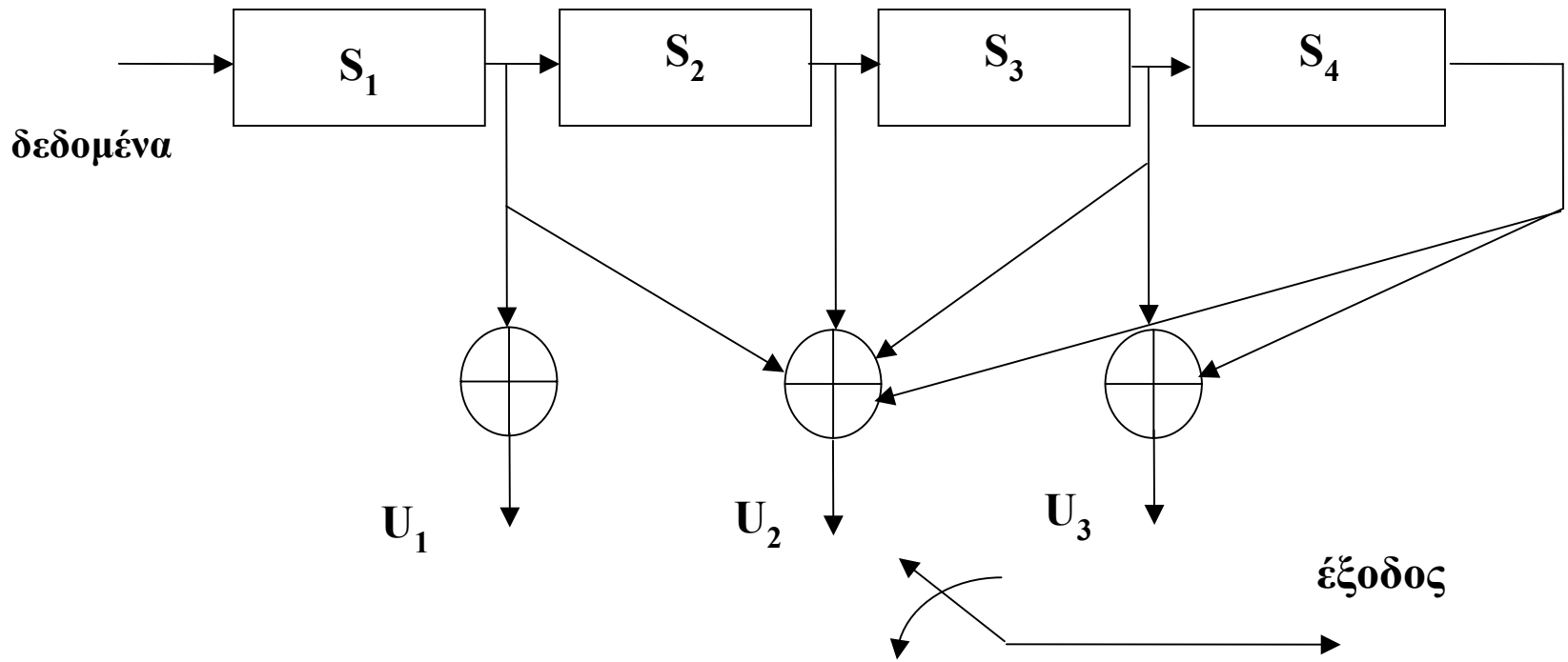
Συγκεραστικοί κώδικες

(Convolutional Codes)

- Στους συγκεραστικούς κώδικες, η κωδικοποίηση πραγματοποιείται πάνω σε ένα **ολόκληρο διάστημα της ροής των συμβόλων του μηνύματος** που ονομάζεται **διάστημα εξαναγκασμού**.

- Ένας συγκεραστικός κώδικας με διάστημα εξαναγκασμού k δημιουργείται με το **συνδυασμό των k εξόδων ενός ολισθητή k-βαθμίδων** και με τη βοήθεια **υ αθροιστών modulo-2**.
- Οι έξοδοι υ_1 υ_1, υ_2,…, υ_υ των αθροιστών δειγματοληπτούνται από έναν κατάλληλο διακόπτη. Έτσι **παράγονται υ ψηφία εξόδου για κάθε ένα ψηφίο εισόδου.**

- διάταξη συγκεραστικού κωδικοποιητή με **k=4 και υ=3**



⊕: **Πύλη αποκλειστικού Η (XOR)**

- **εξισώσεις παραγωγής ψηφίων εξόδου** (δοσμένες):

$$U_1 = S_1$$
$$U_2 = S_1 \oplus S_2 \oplus S_3 \oplus S_4$$
$$U_3 = S_1 \oplus S_3 \oplus S_4$$

- **Π.χ.** αν έχουμε είσοδο το μήνυμα (1011), τότε για κάθε ένα από τα τέσσερα bits παράγονται 3 bits ψηφίων εξόδου

Bit 1: U_1=1, U_2=1, U_3=1 (έξοδος 111)
Bit 0: U_1=0, U_2=1, U_3=0 (έξοδος 010)
Bit 1: U_1=1, U_2=0, U_3=0 (έξοδος 100)
Bit 1: U_1=1, U_2=1, U_3=0 (έξοδος 110)

Κώδικες διόρθωσης καταιγισμού σφαλμάτων

- γραμμικοί κώδικες μπλοκ → διόρθωση **τυχαίων** σφαλμάτων

Σε κανάλια επικοινωνίας (π.χ. κανάλι με διαλείψεις (fading channel) ή σε φθορά ενός CD τα σφάλματα εμφανίζονται κατά ακολουθίες (ριπές)

(**καταιγιστική συμπεριφορά**)

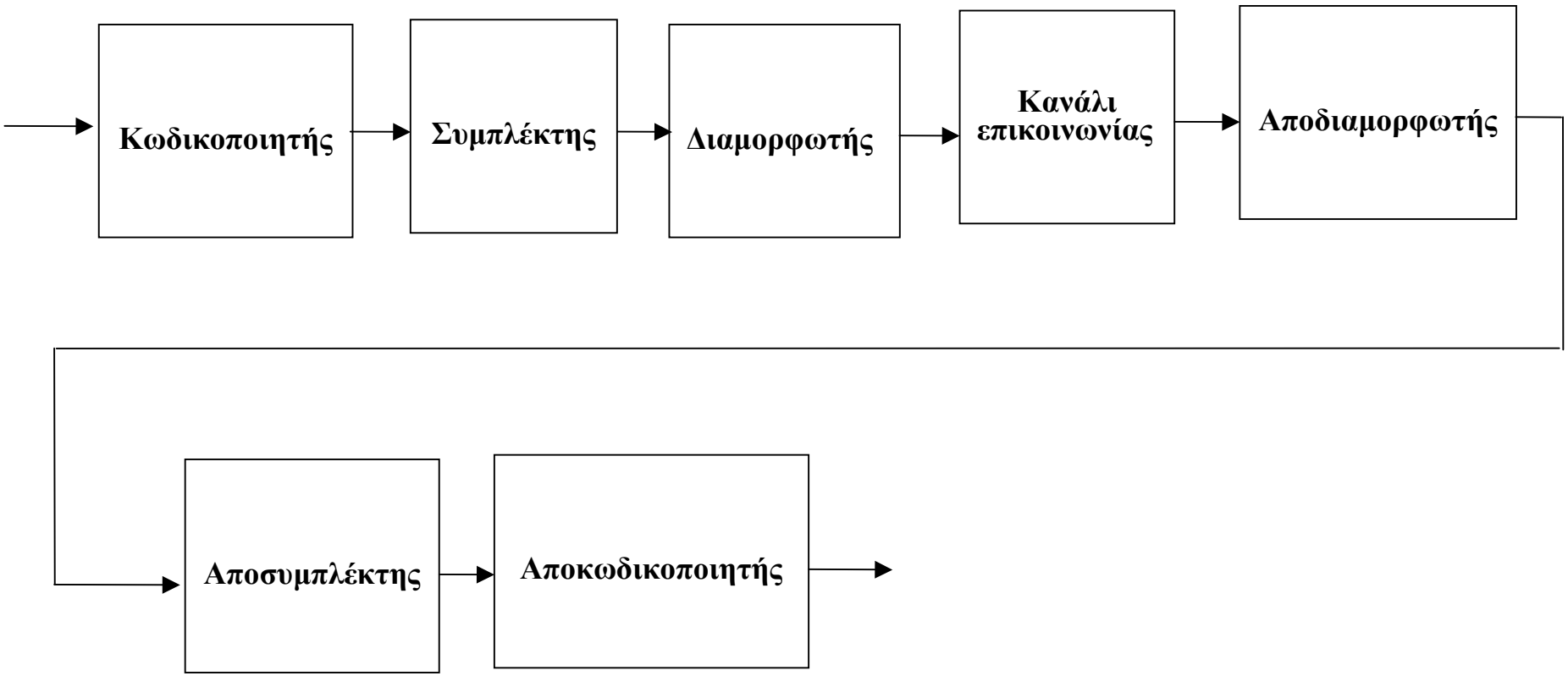
↓

μεθόδοι διόρθωσης καταιγιστικών σφαλμάτων

↓

"***σύμπλεξη***" των κωδικών λέξεων (**interleaving**)

- διάταξη κωδικοποίησης, σύμπλεξης των κωδικών λέξεων, διαμόρφωσης, εκπομπής και **αντίστροφα** αποδιαμόρφωσης, αποσύμπλεξης και αποκωδικοποίησης



Εφαρμογές κωδίκων

- **Κώδικες μπλοκ**: σφάλματα παρουσιάζονται ομοιόμορφα και τυχαία στα εισερχόμενα μπλοκ πληροφορίας ((κανάλι με AWGN θόρυβο) (χερσαίες τηλεφωνικές ζεύξεις)
- **Κώδικες διόρθωσης καταιγισμού σφαλμάτων:** επικοινωνίες κινητών
- **Κώδικες Reed-Solomon:** κινητές επικοινωνίες, τμήματα μηχανισμών διόρθωσης σφαλμάτων των CD

  (συνδυάζονται σε σειρά με έναν δυαδικό κώδικα (π.χ. με κώδικα μπλοκ ή συνελικτικό κώδικα)
- **Κώδικες για μακρινές διαστημικές επικοινωνίες:** χαμηλή τιμή του SNR, μικρή τιμή εκπομπής, θόρυβος AWGN (κώδικες μπλοκ, συνελεκτικοί κώδικες)
- **Κωδικοποίηση για κανάλια περιορισμένου εύρους ζώνης συχνοτήτων:** η κωδικοποίηση οδηγεί σε αύξηση του εύρους ζώνης συχνοτήτων του εκπεμπόμενου σήματος. Στην πράξη όμως έχουμε περιορισμούς στο διαθέσιμο εύρος ζώνης συχνοτήτων π.χ. στη σχεδίαση των modem των τηλεφωνικών καναλιών

  (συνδυασμός μιας μεθόδου κωδικοποίησης και διαμόρφωσης που ονομάζεται ***trellis-κωδικοποιημένη διαμόρφωση*** (trellis coded modulation))

Κώδικες συστημάτων διάχυτου φάσματος

- ακολουθία κωδικοποίησης που χρησιμοποιείται για τη διάχυση του φάσματος του σήματος πληροφορίας πρέπει **τυχαία**, **απείρου μήκους** και **υψηλού ρυθμού**

στη πράξη

ψευδοτυχαίες* ή *ακολουθίες ψευδοθορύβου

(Pseudo-random, Pseudo-noise sequences, PN)

ιδιότητες:

- είναι εύκολο να παραχθούν
- έχουν τυχαίες ιδιότητες
- έχουν μεγάλες περιόδους επανάληψης
- είναι δύσκολο να αναπαραχθούν από ένα μικρό τμήμα τους.

Στη πράξη: ***γραμμικές ακολουθίες μεγίστου μήκους***

(Linear Maximal Length Sequence, LMLS)

ή

ψευδοτυχαίες ακολουθίες μεγίστου μήκους

(m-ακολουθίες)

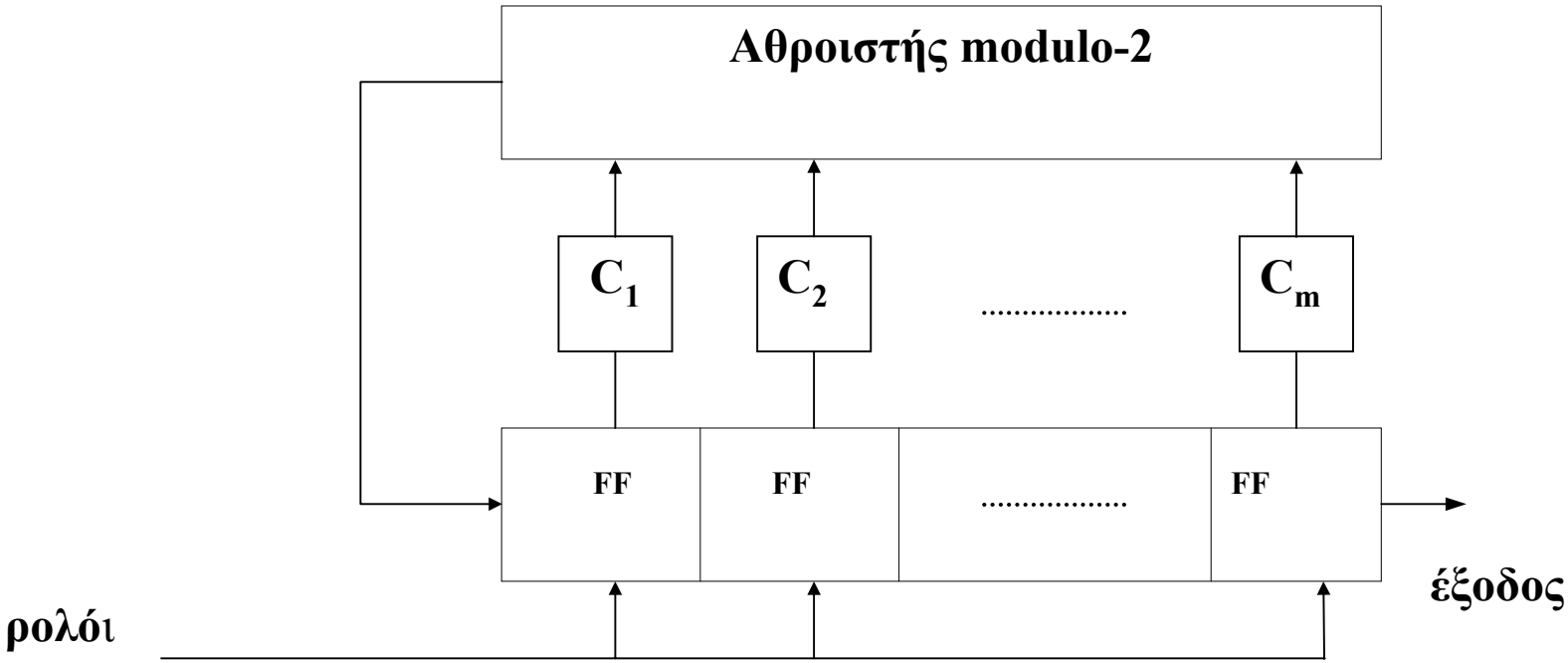
(Pseudo-noise maximal length sequence, m-sequences)

Παραγωγή:

απλή διάταξη γεννήτριας

(Pseudo-random Generator, PRG)

(*καταχωρητής ολίσθησης με γραμμική ανατροφοδότηση*
(linear feedback shift register))



• αριθμός των δυαδικών ψηφίων μετά από τον οποίο η ακολουθία επαναλαμβάνεται λέγεται **περίοδος** *N*:

$$N = 2^m - 1$$

όπου ***m*** είναι ο αριθμός των χρησιμοποιούμενων FF

Ιδιότητες ακολουθιών μεγίστου μήκους

- "**ιδιότητα ισορροπίας**" (balance property): Ο αριθμός των λογικών "1" είναι πάντα κατά ένα μεγαλύτερος του αριθμού των λογικών "0"
- "***ιδιότητα εμφάνισης διαδοχικών "1" και"0"***" (run property): στη διάρκεια της περιόδου N κάθε ακολουθίας μεγίστου μήκους, το πλήθος εμφάνισης q διαδοχικών λογικών "1" ή "0" είναι $2^{m-(q-2)}$

- "***ιδιότητα αυτοσυσχέτισης***" (autocorrelation property): η συνάρτηση αυτοσυσχέτισης $R_x(\tau)$ (autocorrelation function) μιας ακολουθίας μεγίστου μήκους παίρνει δύο τιμές: για μηδενική ολίσθηση παίρνει τη μέγιστη τιμή της, δηλαδή, $N=2^m-1$ ενώ για οποιαδήποτε άλλη ολίσθηση μεγαλύτερη του ενός bit παίρνει την τιμή -1

ΕΝΔΕΙΚΤΙΚΗ ΒΙΒΛΙΟΓΡΑΦΙΑ ΜΑΘΗΜΑΤΟΣ

- J.G.Proakis,M.Salehi, *Μετάφραση*: Κ.Καρούμπαλος, Ζέρβας Ε., Καραμπογιάς Σ.,Σαγκριώτης Ε., *Συστήματα Τηλεπικοινωνιών*, Ε.Κ.Π.Α., Αθήνα 2002
- J.G.Proakis, *Digital Communications*, 3rd Edit., *McGraw-Hill*, 1995
- Δ.Χ.Βούκαλης, *Θεωρία Πληροφοριών και Κωδίκων,* Εκδόσεις ΙΩΝ, Περιστέρι, 1994
- Η.P.Hsu, *Αναλογικές και Ψηφιακές Επικοινωνίες*, Σειρά *Schaum*, *Μετάφραση*:Ι.Βαρδιάμπασης, Εκδόσεις Τζιόλας, 2002
- T.Cover and J.Thomas, *Elements of Information Theory*,New York: *Wiley*, 1991
- Ν.Σ.Τζάννες, *Θεωρία Μετάδοσης Πληροφοριών*, Τόμος ΙΙ, *Εισαγωγή στις Θεωρίες Shannon και Κωδίκων*, Πάτρα, 1981