Théorie de l'information

corinne.mailhes@n7.fr

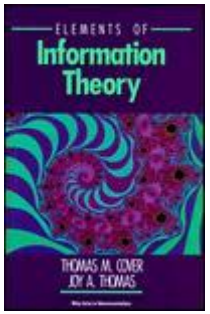
I - Système de communication et codage de source

II- La théorie de l'information

- Définition de la quantité d'information
- Les quantités d'informations d'un système de communication
- Les sources
- Entropie : définitions, différentes entropies
- Capacité d'un canal - Exemple du Canal Binaire Symétrique
- Un « bon » code de source
- Théorème du codage dans un canal bruité (NCC theorem)
- Théorème du codage sans bruit (codage bloc)

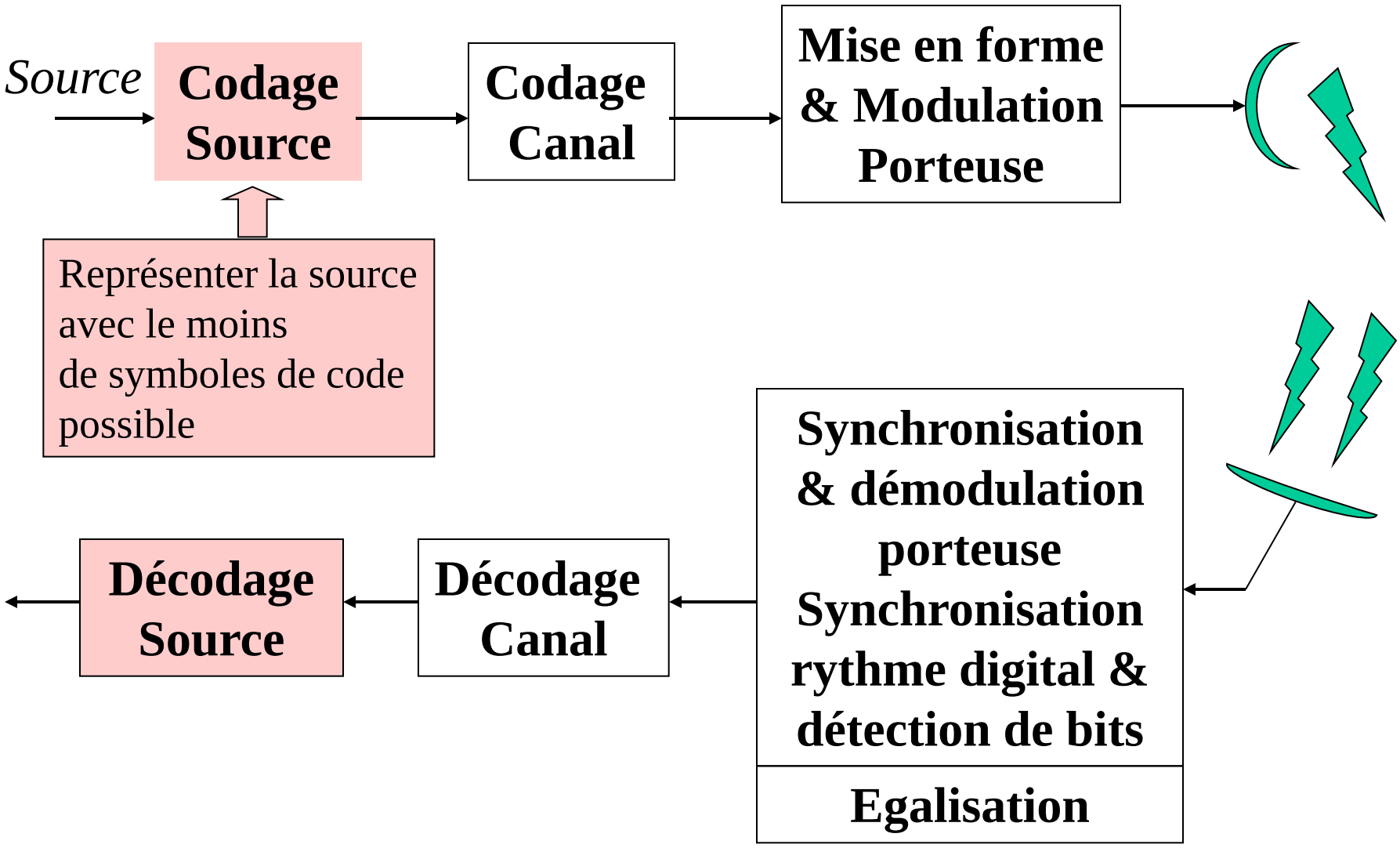
III- Codage sans perte

- Huffman

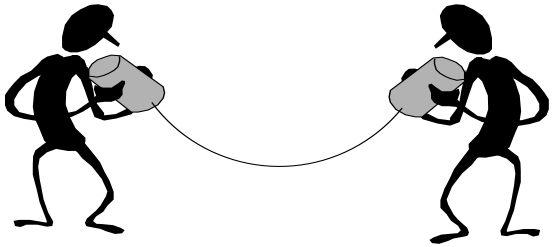


T.M.Cover, J.A.Thomas,
«Elements of Information Theory »,
Wiley Series Telecom, 1991.

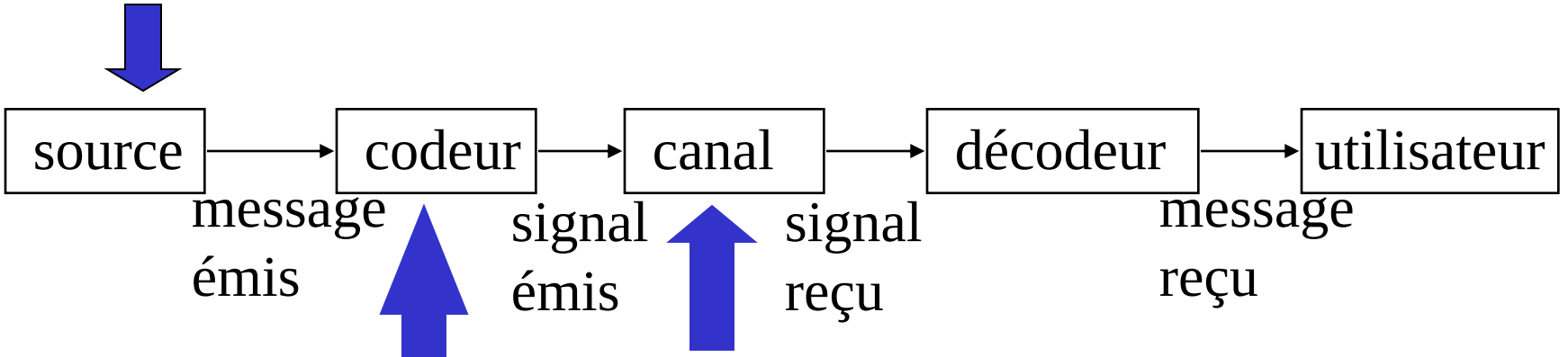
Chaîne de Transmission



I - Système de communication et codage de source



Continue
ou discrète ?



Modèle ?
Bruit additif ? Multiplicatif ?
Discret ? Continu ?

Rôle économique ? Codage source
Lutte contre le bruit ? Codage canal
Codage source-canal conjoint, l'avenir...

II- La théorie de l'information

Définition de la quantité d'information

➔ = quantité de doute, liée à la probabilité de l'événement : $i(x)=F(p(x))$

avec $F(1)=0$

➔ quantité additive : $i(xy)=i(x)+i(y)$ si x et y indépendants

d'où ***i(x)= - a.log(p(x))***

unité : ***Binary Unit*** :

associée à l 'expérience la plus simple possible (pile ou face équiprobable)

i(pile)=i(face)= - a.log(1/2) = 1 binary unit = 1 binit = 1 bit

en choisissant log en base 2, a=1 !

d 'où

$$i(x)= - \log_2 (p(x)) \text{ bits}$$

Claude Elwood Shannon (1916 – Febr 2001)

autres unités existent mais « bit » la plus utilisée

C. E. Shannon, « A mathematical theory of communication »,
Bell System Technical Journal, vol. 27, pp. 379-423 and 623-656, July and October, 1948.
Disponible sur le web

http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf

II- La théorie de l'information

Les sources : simples ? De Markov ?

Andrei Andreyevich Markov 1856 - 1922

1. Zero-order approximation (symbols independent and equiprobable).

   XFOML RXKHRJFFJUJ ZLPWCFWKCYJ FFJEYVKCQSGHYD QPAAMKBZAACIBZL-HJQD.

2. First-order approximation (symbols independent but with frequencies of English text).

   OCRO HLI RGWR NMIELWIS EU LL NBNESEBYA TH EEI ALHENHTTPA OOBTTVA NAH BRL.

3. Second-order approximation (digram structure as in English).

   ON IE ANTSOUTINYS ARE T INCTORE ST BE S DEAMY ACHIN D ILONASIVE TU-COOWE AT TEASONARE FUSO TIZIN ANDY TOBE SEACE CTISBE.

4. Third-order approximation (trigram structure as in English).

   IN NO IST LAT WHEY CRATICT FROURE BIRS GROCID PONDENOME OF DEMONS-TURES OF THE REPTAGIN IS REGOACTIONA OF CRE.

5. First-order word approximation. Rather than continue with tetragram, . . . , n-gram structure it is easier and better to jump at this point to word units. Here words are chosen independently but with their appropriate frequencies.

   REPRESENTING AND SPEEDILY IS AN GOOD APT OR COME CAN DIFFERENT NAT-URAL HERE HE THE A IN CAME THE TO OF TO EXPERT GRAY COME TO FURNISHES THE LINE MESSAGE HAD BE THESE.

6. Second-order word approximation. The word transition probabilities are correct but no further struc-ture is included.

   THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITER THAT THE CHAR-ACTER OF THIS POINT IS THEREFORE ANOTHER METHOD FOR THE LETTERS THAT THE TIME OF WHO EVER TOLD THE PROBLEM FOR AN UNEXPECTED.

II- La théorie de l'information

Entropie : définitions, différentes entropies

Caractérise la source :

$$H(X) = H(p_1,p_2,\ldots,p_N) = E[i(x_k)] = \sum_k p_k\, i(x_k) = -\sum_k p_k \log_2 (p_k) \text{ (bits)}$$

Propriétés :

❶ Entropie positive ou nulle, fonction symétrique du jeu de probabilités

❷ Entropie maximale : loi équiprobable

❸ Décomposition augmente l'entropie

$$H(p_1,p_2,\ldots,p_N) > H(P,Q)$$

Les quantités d'informations d'un système de communication

Source discrète : $X=\{x_1,x_2,\ldots,x_N\}$
Quantité d'incertitude a priori : $i(x_k)=-\log_2(p_k)$ avec $p_k=P[X=x_k]$
Récepteur : $Y=\{y_1,y_2,\ldots,y_M\}$ on reçoit y_n : qui a été émis ? x_k ?
Quantité d'incertitude a posteriori : $i(x_k/y_n)=-\log_2(P[X=x_k/Y=y_n])$
« Réduction d 'incertitude » : apport de la transmission :
Information mutuelle élémentaire :

$$i(x_k, y_n) = i(x_k) - i(x_k/y_n)$$ (Positif ? Négatif ?…)

Exemples : canal parfait et canal « poubelle »

<u>Différentes entropies</u>
Entropie de la source H(X)
Entropie du récepteur H(Y)
Entropies conditionnelles :

ambiguïté : $H(X/Y) = -\Sigma_k \Sigma_n\ p(x_k, y_n) \log2\ (p(x_k / y_n))$

erreur moyenne du canal : $H(Y/X) = -\Sigma_k \Sigma_n\ p(x_k, y_n)\ \log2\ (p(y_n / x_k))$

Entropie conjointe :

$$H(X,Y) = -\Sigma_k \Sigma_n\ p(x_k, y_n) \log2\ (p(x_k, y_n))$$

II- La théorie de l'information

Entropie : définitions, différentes entropies

Transinformation :

$I(X,Y) = E[i(x_k, y_n)] = \sum_k \sum_n p(x_k, y_n) \log_2 \left(\frac{p(y_n / x_k)}{p(y_n)} \right)$

toujours positive (ou nulle) !

Relations entre toutes ces entropies : égalités et inégalités :

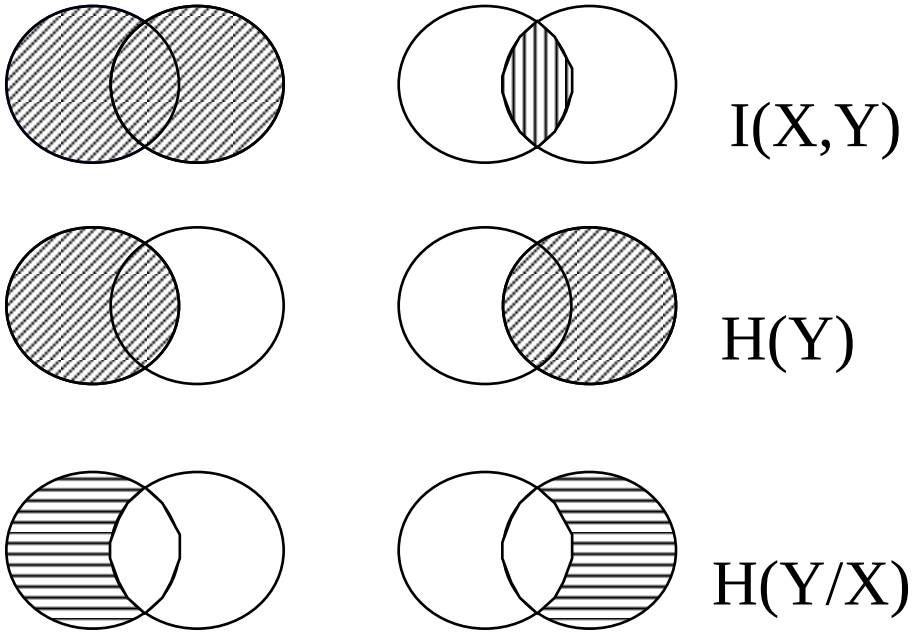
Exemples :

$H(X,Y)=H(X/Y)+H(Y)$

$H(X/Y) \leq H(X)$

…

on s'aide de l'isométrie sur les ensembles ci-contre

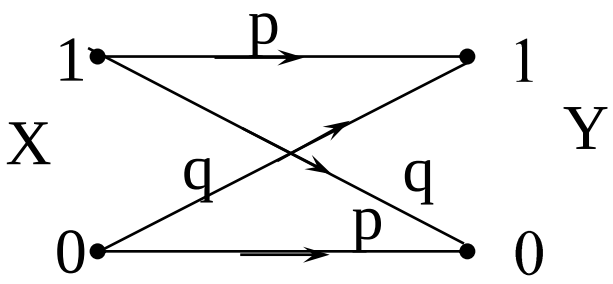


Exemples canal parfait et canal « poubelle »

II- La théorie de l'information

Capacité d'un canal - Exemple du Canal Binaire Symétrique

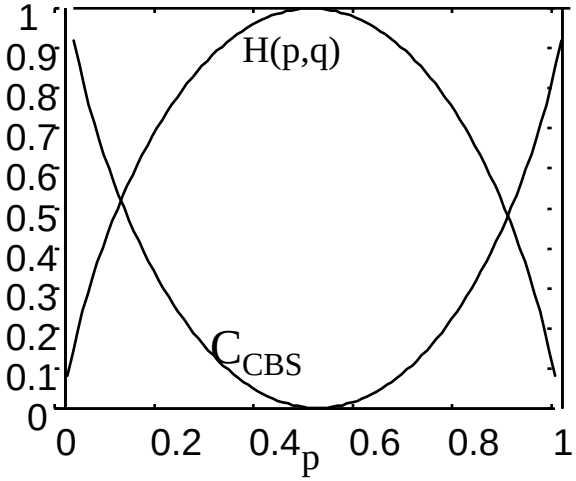
Caractérise le canal :
C = Max(I(X,Y)) sur toutes les lois possibles de X
exprimée en bits



Exemple du Canal Binaire Symétrique (CBS)
I(X,Y) = H(Y) - H(Y/X)
avec
H(Y / X=0) = H(p,q) = H(Y / X=1) = H(Y /X)
d'où

$$C_{CBS} = 1 - H(p,q)$$

atteinte pour
loi d'émission équiprobable



II- La théorie de l'information

Un « bon » code de source

Source discrète avec alphabet : X={x_1,x_2,…,x_N}
Entropie H(X) (bits)

Canal discret avec alphabet : U={u_1,u_2,…,u_D}
Capacité C (bits)

En général N > D
Codage : x_k => mot-code : $m_k = u_{n_1} u_{n_2} \ldots u_{n_k}$ n_k : longueur du mot-code

Paramètre du code : la longueur moyenne

$\overline{n} = \sum_k p_k n_k$

Aussi petit que voulu ?...

Entropie des symboles de code

⇩ Messages H(X) avec $\overline{n}$ symboles de code en moyenne : $H(X) / \overline{n} \leq log_2 (D)$

soit

$$H(X) / log_2 (D) \leq \overline{n}$$

Exemple code binaire

Efficacité $E = H(X) / \overline{n}\, log_2 (D)$ Redondance $\rho = 1 - E$

propriétés : régulier, déchiffrable, irréductible (*aucun mot-code n'est le début d'un autre*)

II- La théorie de l'information

Théorème du codage dans un canal bruité (NCC theorem)

Transmission d'une source à travers un canal possible avec probabilité d'erreur aussi faible que voulue ssi :

$$H(X) / T_S < C / T_C$$

avec T_S temps entre émission de deux messages, T_C temps minimal entre deux symboles de code

Débit de la source : $1/ T_S$ messages par seconde

Taux d 'émission de la source : $H(X) / T_S$ bits/seconde

Capacité du canal : $C ' = C / T_C$ bits/seconde

Si la transmission est possible, reste à trouver le bon code tel que :

$$\bar{n} / T_S < 1 / T_C$$ (en termes de symboles de code par seconde)

⇩ nécessaire d'avoir un code suffisamment efficace

Théorème du codage sans bruit (codage bloc)

Il existe au moins un code irréductible tel que $H(X) / log_2 (D) \leq \bar{n} \leq H(X) / log_2 (D) +1$

Au lieu de coder directement la source $X=\{x_1,x_2,...,x_N\}$

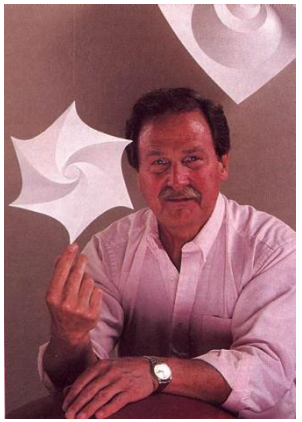
on code son extension d'ordre m constituée de blocs de m messages de X (N^m messages)

alors

$$H(X) / log_2 (D) \leq \bar{n} \leq H(X) / log_2 (D) +1/m$$

III- Codage sans perte

Code d'Huffman



Code d'Huffman (1952) : codage optimal

David A. Huffman (9 août 1925 - 7 octobre 1999)

D.A. Huffman, "A method for the construction of minimum-redundancy codes", Proceedings of the I.R.E., sept 1952, pp 1098-1102

http://compression.ru/download/articles/huff/huffman_1952_minimum-redundancy-codes.pdf

Cas binaire :

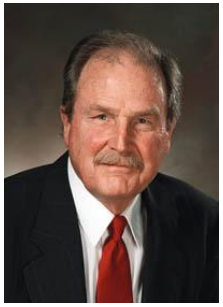
1. on classe les messages par ordre de probabilité décroissante,
2. on regroupe les 2 messages les moins probables, en leur affectant une probabilité égale à la somme des probabilités. Les 2 messages auront le même code sauf la fin : le 1er se verra affecter un symbole « 0 » et le 2ème un symbole « 1 »,
3. on refait 1 jusqu 'à épuisement.
4. La lecture des mots-code se fait en lisant le tableau ainsi constitué de gauche à droite : on lit les mots-code à l'envers (de la fin vers le début)

Cas D différent de 2 :

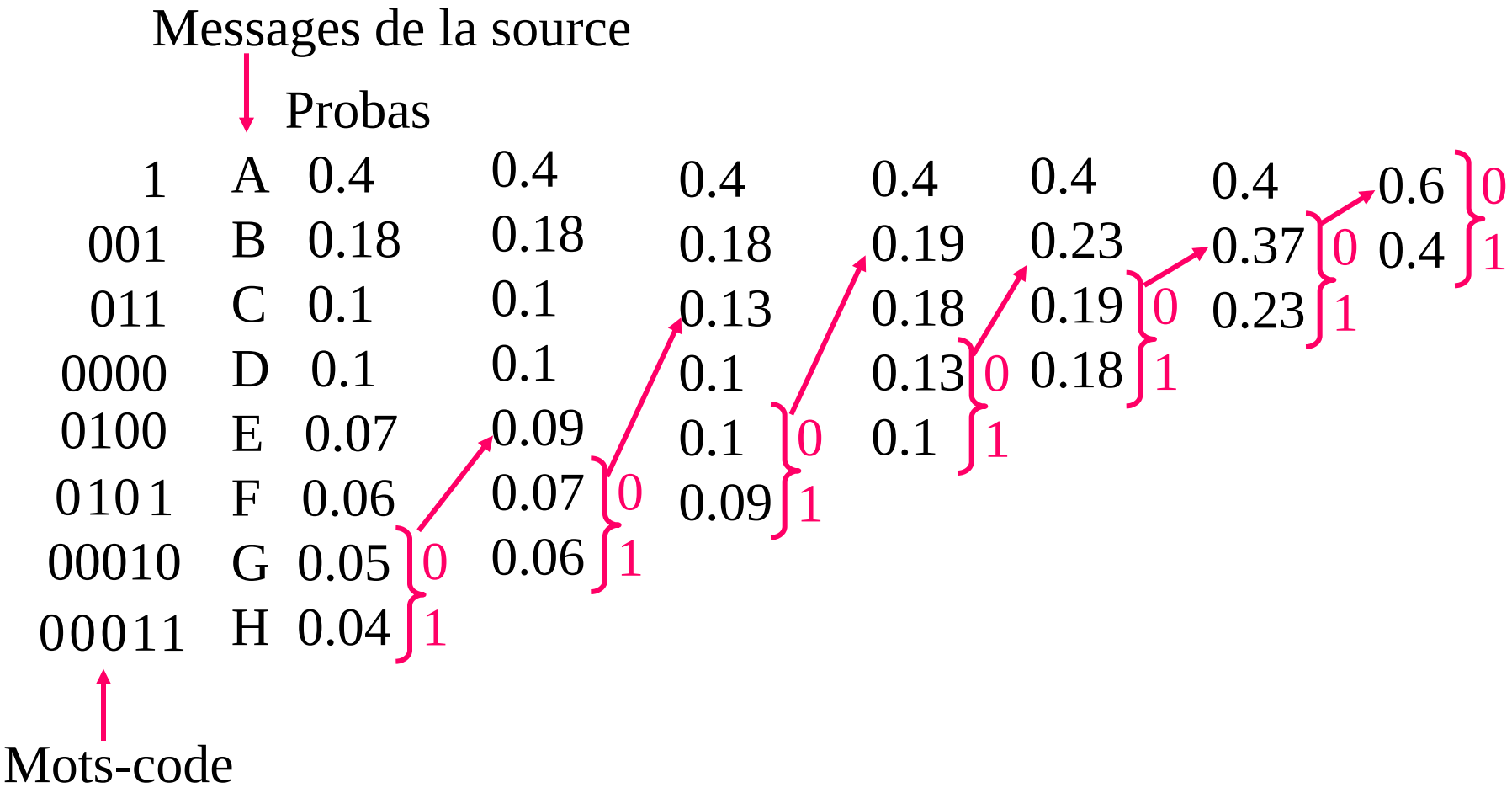
On fait de même en remplaçant « 2 » par « D » et les symboles 0 et 1 par $u_1,\ldots,u_D$

☠💣☠ Initialisation : on regroupe non pas D symboles à la 1ère itération mais :

2 + reste de la division de N-2 par D-1



Code d'Huffman : un exemple



Longueur moyenne atteinte par Huffman : 2.61
H(X)=2.55 bits soit une efficacité de E=97.8%