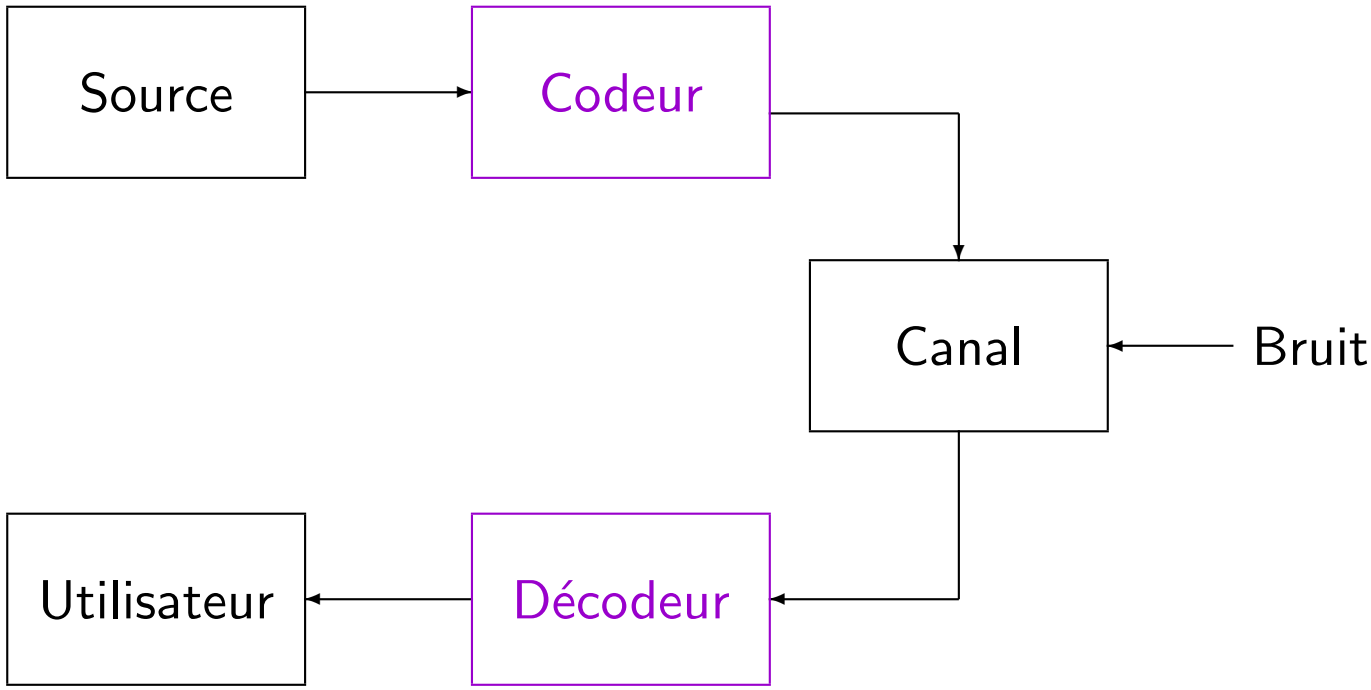
Contenu du cours

- limites théoriques du taux de compression d'une source et du débit de transmission dans un canal bruité,
- algorithmes permettant d'atteindre de manière efficace ces limites.

Applications de la théorie de l'information

- Transmission/stockage des données numériques
- Cryptographie
- Théorie des jeux
- Bioinformatique

Système de communication

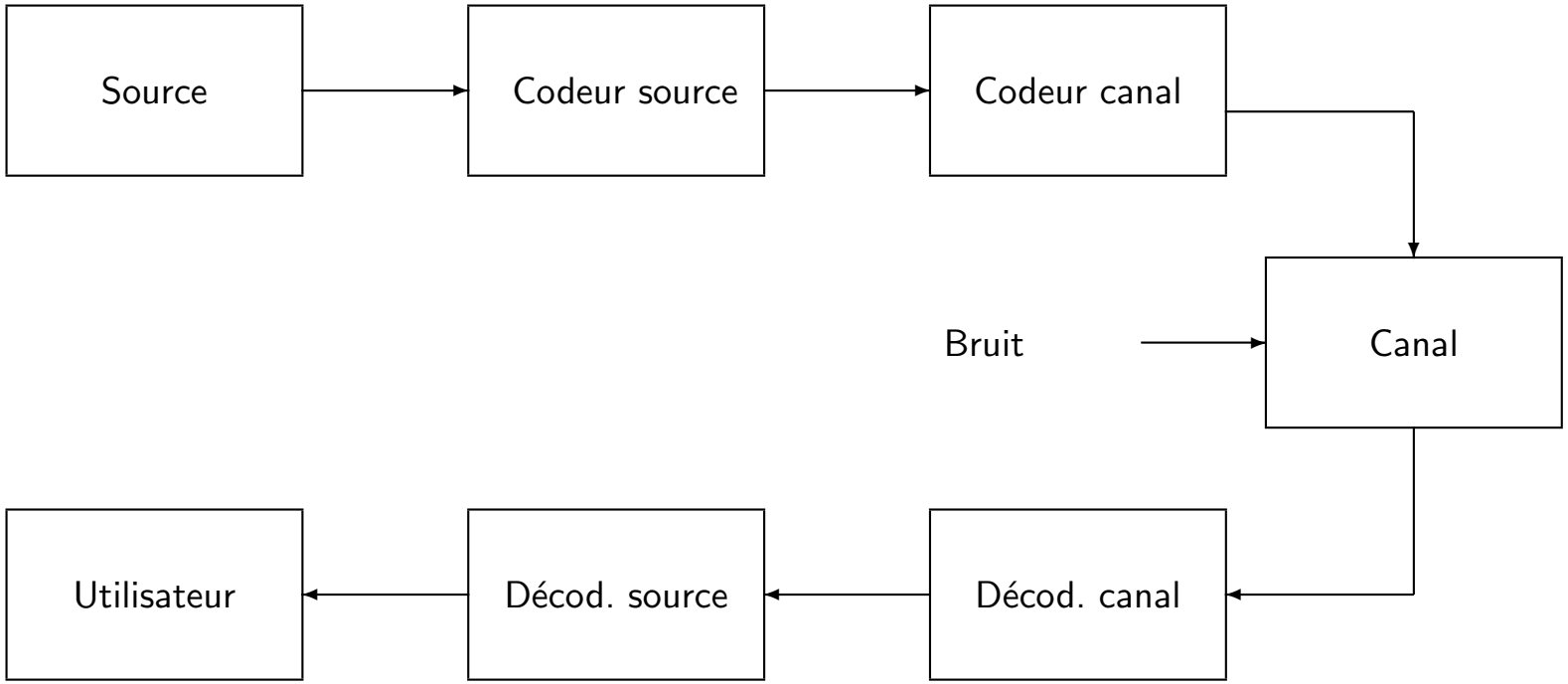


Source : voix, musique, image (fixe ou animée), texte, . . .

Canal : radio, fil, fibre optique, support magnétique/optique,

Bruit : perturbations electromagnétiques, rayures, . . .

Codage de source et de canal



Efficacité : Pour faire parvenir une quantité donnée d'information à l'utilisateur, utiliser le minimum de ressources.

Fiabilité : Restituer à l'utilisateur une information suffisamment fidèle à celle produite par la source.

Codage source/canal

Problématique :

— Codage source : compresser efficacement une source donnée à un taux de compression maximal. Ex :

$$\mathbf{x} = x_1 \dots x_n,\ \mathbf{Prob}(x_i = 1) = p.$$

— Codage canal : transmettre efficacement le maximum d'information à travers un canal bruité. Ex :

$$\mathbf{x} = x_1 \dots x_n \overset{\text{canal}}{\rightsquigarrow} \mathbf{y} = y_1 \dots y_n,\ \mathbf{Prob}(y_i \neq x_i) = p.$$

Une même quantité sert à quantifier cette limite : l'entropie.

Codage source/canal

Problématique :

— Codage source : compresser efficacement une source donnée à un taux de compression maximal. Ex :

$$\mathbf{x} = x_1 \dots x_n,\ \mathbf{Prob}(x_i = 1) = p.$$

$\Rightarrow$ compresser en une séquence de taille $\approx nh(p)$ bits.

— Codage canal : transmettre efficacement le maximum d'information à travers un canal bruité. Ex :

$$\mathbf{x} = x_1 \dots x_n \overset{\text{canal}}{\rightsquigarrow} \mathbf{y} = y_1 \dots y_n,\ \mathbf{Prob}(y_i \neq x_i) = p.$$

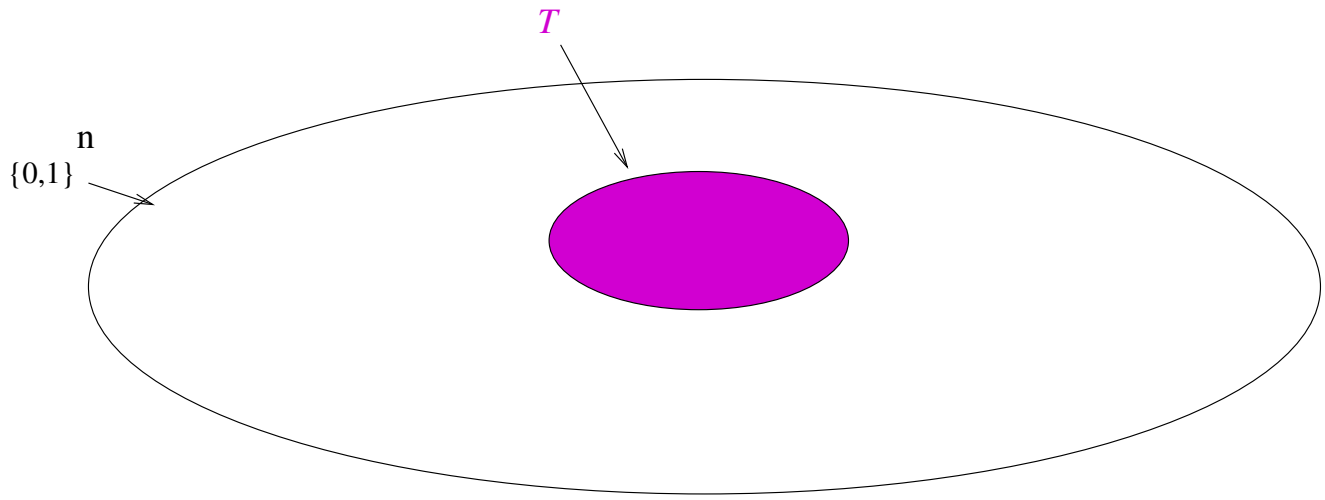
$\Rightarrow$ transmettre $\approx n(1 - h(p))$ bits d'information.

Une même quantité sert à quantifier cette limite : l'entropie.

$$h(p) \overset{\text{def}}{=} -p \log_2 p - (1-p) \log_2 (1-p)$$

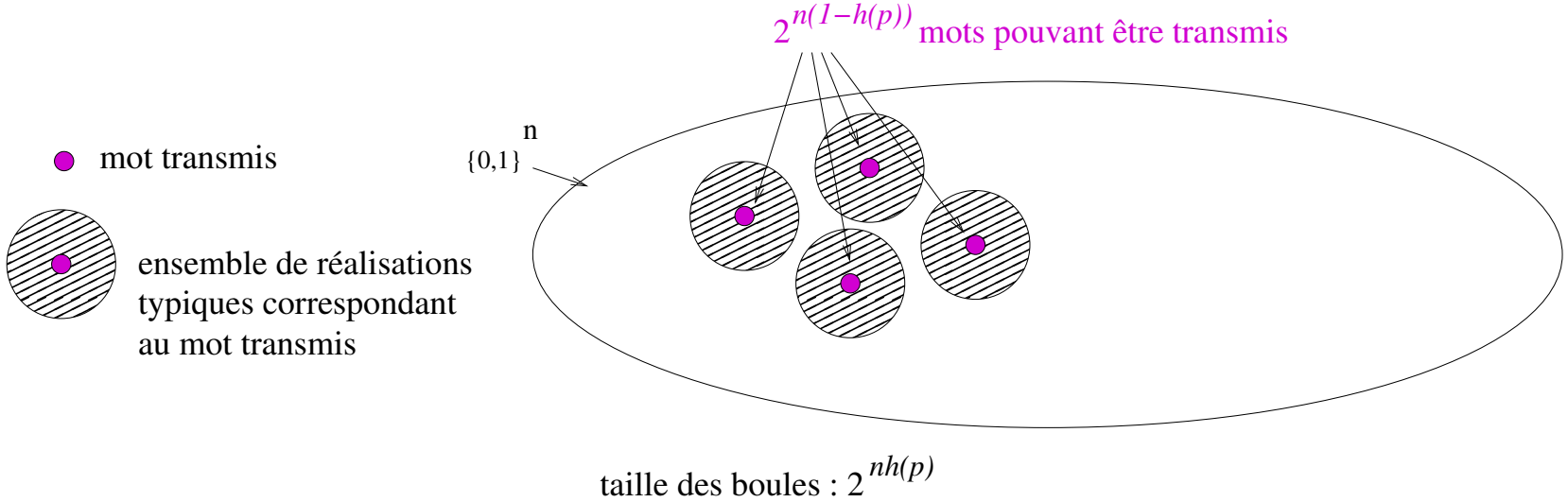
Entropie et séquences typiques

Un principe commun : se concentrer sur les réalisations typiques



$$
\begin{aligned}
T &= \{\mathbf{x}; |\mathbf{x}| \approx pn\} \\
\mathbf{Prob}(\mathbf{x} \in T) &\approx 1 \\
|T| &\approx 2^{nh(p)} \\
\log_2 |T| &\approx \text{Entropie}
\end{aligned}
$$

- Codage source : numéroter avec $nh(p)$ bits les éléments de T et ne rien faire pour les autres.
- Codage canal :



$\log$ (nombre de mots pouvant être transmis) $=$ nombre de bits d'information reçus

Entropie

Formule s'explique par deux faits

(i) $\log$ transforme un produit en somme,

(ii) concentration de la somme de v.a. i.i.d. autour de son espérance.

$$\begin{aligned}
\log \mathbf{Prob}(\mathbf{x}) &\overset{(i)}{=} \log \mathbf{Prob}(x_1) + \cdots + \log \mathbf{Prob}(x_n) \\
&\overset{(ii)}{\approx} n\left(p \log p + (1-p)\log(1-p)\right) = -nh(p) (p.s.) \\
\Rightarrow \mathbf{Prob}(\mathbf{x}) &\approx 2^{-nh(p)} (p.s.)
\end{aligned}$$

De manière générale pour une v.a.d. X prenant ses valeurs dans $\mathcal{A}$:

$$\text{Entropie}(X) \overset{\text{def}}{=} -\sum_{a \in \mathcal{A}} \mathbf{Prob}(X = a) \log \mathbf{Prob}(X = a).$$

Code à répétition

Pour combattre les effets du bruit on ajoutera de la redondance. Par exemple, le code à répétition de longueur 3 :

$$\begin{aligned} 0 &\mapsto 000 \\ 1 &\mapsto 111 \end{aligned}$$

ou, plus généralement le code à répétition de longueur $2m+1$.

$$\begin{aligned} 0 &\mapsto \overbrace{0\ldots0}^{2m+1} \\ 1 &\mapsto \overbrace{1\ldots1}^{2m+1} \end{aligned}$$

Code à répétition

Si la probabilité d'erreur du canal $p = 0.01$ pour chaque symbole transmis, il se produira 0 ou 1 erreur avec une probabilité

$$(1-p)^3 + 3p(1-p)^2 \approx 0.9997$$

et il se produira 2 ou 3 erreurs avec une probabilité

$$3p^2(1-p) + p^3 \approx 3 \times 10^{-4}$$

Le symbole sera mal transmis avec une probabilité 3×10^{-4}. Avec un code à répétition de longueur 5, cette probabilité tombe à 10^{-5}

$$10p^3(1-p)^2 + 5p^4(1-p) + p^5 \approx 10^{-5}$$

Ce code à un taux de transmission 0.2.

Rendement d'un code

Le code à répétition de longueur 3 a un taux de transmission $1/3 = 0.33$ et corrige une erreur.

Le code à répétition de longueur 5 a un taux de transmission $1/5 = 0.2$ et corrige deux erreurs.

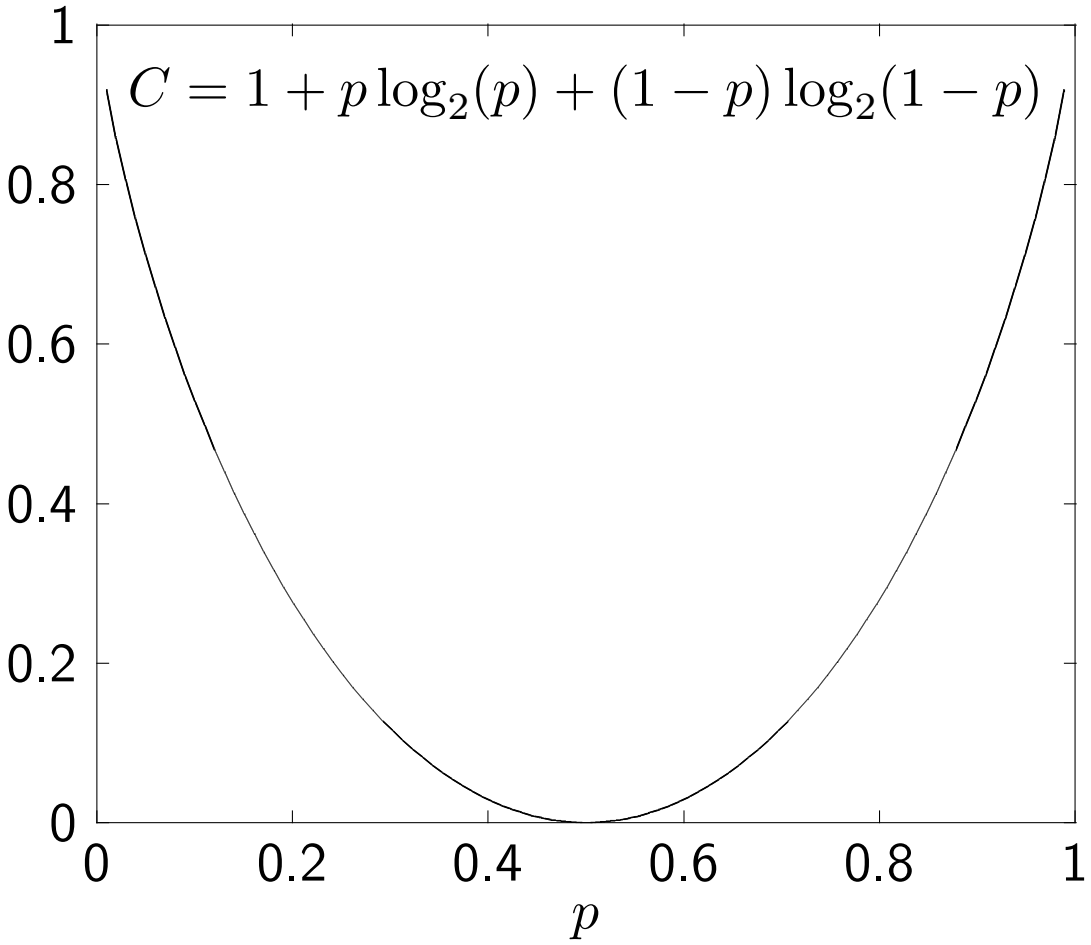
En diminuant le taux de transmission, on fait baisser la probabilité d'erreur après décodage.

Recherche du taux de transmission optimal : doit-il tendre vers 0 ?

Non ! Deuxième théorème de Shannon.

Notion de capacité C d'un canal.

Capacité du canal binaire symétrique



La capacité est le **taux de transmission maximal** du code à utiliser pour transmettre de l'information « dans de bonnes conditions ».

Par ex. $C(0.01) = 0.919$. Il y a donc moyen de faire (beaucoup) mieux que le code à répétition ! ! !

Résultats importants du cours

Premier théorème de Shannon (Codage de source)

1. On peut coder toute source en utilisant un nombre de bits par lettre aussi proche que l'on veut de son entropie.
2. On ne peut pas faire mieux.

Second théorème de Shannon (Codage de canal)

1. On peut transmettre de l'information de façon fiable en utilisant un code correcteur d'erreur de taux de transmission inférieur à la capacité du canal utilisé.
2. On ne peut pas faire mieux.

TD

Première séance : exercices sur feuille

Deuxième séance : TD en java ou langage de votre choix.