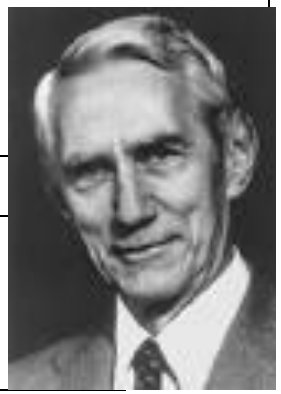
Initiation à la théorie de l'information

(Claude Shannon – 1948)



Emergence Paris - Santacafé

Avant propos

- La théorie de l'information, sans faire appel - du moins dans ses théorèmes de base - à des math très avancées (*) est abstraite et assez difficile;
- L'information de Kolmogorov est une notion plus récente faisant le lien entre information et complexité;
- La présente présentation se propose simplement de faire « sentir » les principes fondamentaux de cette théorie ;
- Quelques rappels et compléments sont donnés en annexe ;

(*) Contrairement à la mécanique quantique qui s'appuie sur un ensemble de formalismes beaucoup plus vaste et avancé;

Quelques aperçus sur la théorie de l'information

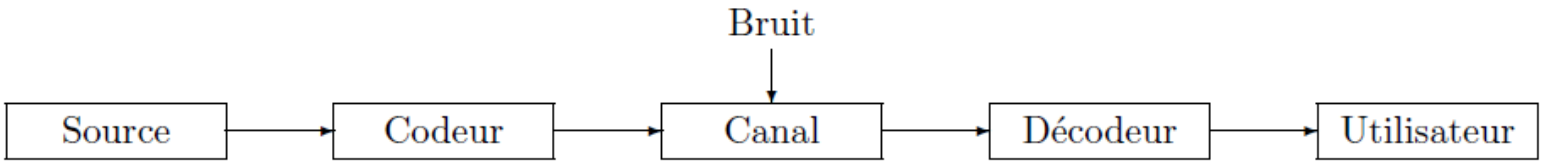
- Théorie créée en 1948 par un ingénieur des Bell Labs, **Claude Shannon** ;
- Est une théorie apparentée à la physique mathématique, utilisant des outils classiques (probabilités, algèbre, transformation de Fourier, etc.) ;
- Objectifs: modéliser les communications, sans ou avec bruit, pour optimiser la transmission des données. Les concepts de base débouchent sur les techniques de protection contre les erreurs, la compression des données et la cryptographie ;
- L'information telle que traitée dans cette théorie est de nature statistique. Il ne s'agit pas de mesurer la valeur du contenu !!
- Beaucoup de techniques modernes de télécom sont expliquées par cette théorie;

Quelques aperçus sur la théorie de l'information

- **Résultats de base:**
  - Avec un codage ad' hoc, il est possible d'atteindre la capacité théorique d'un canal parfait sans bruit ;
  - Même résultat pour un canal avec bruit "idéalisé"
- **Théories liées:**
  - **Théorie du signal:**
    - Théorème de l'échantillonnage (Shannon-Nyquist)
    - Capacité d'un canal physique (Shannon)
    - Filtrage et extraction du signal
  - **Physique statistique (Boltzmann)**
    - Même forme mathématique de l'entropie d'une source d'information et l'entropie en physique où l'entropie mesure le désordre d'un système;

Bases

- **Schéma général**



- **Concepts principaux:**
  - **Mesure d'une quantité d'information : entropie d'une source d'information ;**
  - **Capacité d'un canal de transmission ;**
  - **Bruit et taux d'erreur ;**
  - **Codage de l'information ;**
  - **Redondance, compression ; distance ;**

Mesure de l'information

- La mesure de l'information d'un message est liée à son aspect inattendu ou improbable:
  - « ***Demain à minuit il fera nuit*** » n'apporte pas d'info;
  - « ***Demain à midi il fera nuit à cause d'une éclipse*** » apporte plus d'info;
- Si ***p*** est la probabilité d'un message et **H** l'info qu'il transporte, on comprend que:
  - l'information d'un évènement certain est nulle: **H**(***p***=1)=0
  - l'information croît si la proba. décroit**:** **H** (***p***)↗ si ***p*** ↘

Mesure de l'information

Questions binaires pour découvrir dans quelle case est la boule ?
Chaque réponse est **D**roite ou **G**auche.

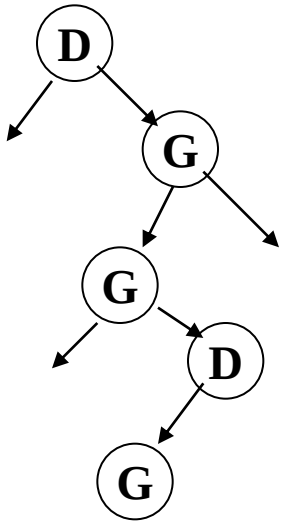
Q1

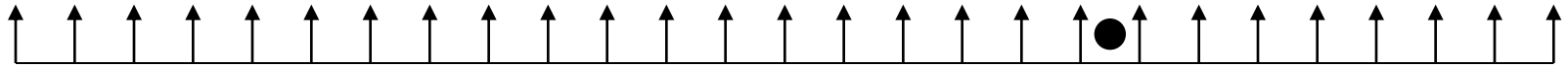
Q2

Q3

Q4

Q5





La boule est dans la 19eme case: $18=16+2=2^4+2^1=10011$ en binaire

Mesure de l'information

- Combien de questions ***q*** (tas gauche ou tas droite) pour découvrir où est une boule dans l'une des ***n*** cases?
- Par exemple si 26 cases (a,b,c,..z) il faudra 5 questions ($26<2^5=32$) pour découvrir où est la boule;
- La probabilité pour la boule d'être dans une case déterminée sera ***p***=1/26

Cas général

- Le nombre de questions sera tel que $2^q \geq n$ ce qui peut s'écrire $q \geq \log_2(n)$
- La probabilité pour une boule d'être dans une case est $p=1/n$ c'est à dire $q \geq \log_2(1/p)$ ou encore $q \geq -\log_2(p)$

Mesure de l'information

- Il paraît naturel de dire que chaque question binaire apporte une unité d'information. L'information contenue dans un message indiquant la case où se trouve une boule sera $\mathbf{H} = -\log_2(\boldsymbol{p})$
- Si le message contient des évènements avec probabilités variées, l'information contenue dans le message (appelée **entropie H**) sera la somme pondérée des entropies des divers évènements:

$$\mathbf{H} = \Sigma - \boldsymbol{p}_i \times \log_2(\boldsymbol{p}_i)$$

- Appliquons cette formule au langage.

Entropie du langage

Fréquences des lettres en Français

Lettre	%	Lettre	%
A	8,11	N	7,68
B	0,81	O	5,20
C	3,38	P	2,92
D	4,28	Q	0,83
E	17,69	R	6,43
F	1,13	S	8,87
G	1,19	T	7,44
H	0,74	U	5,23
I	7,24	V	1,28
J	0,18	W	0,06
K	0,02	X	0,53
L	5,99	Y	0,26
M	2,29	Z	0,12

Avec des lettres au hasard, l'entropie moyenne est de $-\log_2(1/26) \approx$ **4,70**

En appliquant la formule précédente, à un texte français, l'entropie moyenne par caractère est ≈ **3,95**

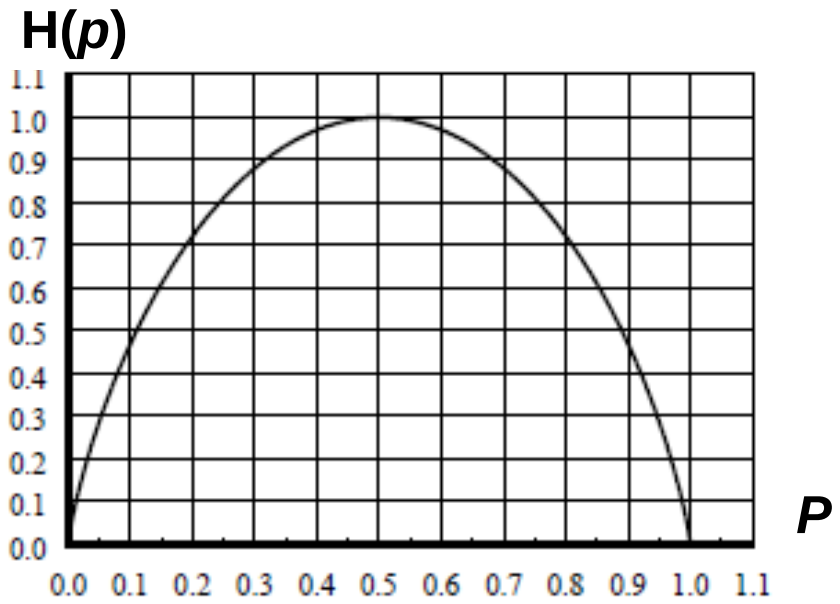
Conséquences:

- Le texte est **redondant** (voir ci-après)
- Dans l'histoire cette redondance a facilité le décodage des cryptages simples (simples substitutions: cas de Marie Stuart);

Canal sans bruit

- Supposons un canal de transmission qui émet 1 élément binaire par seconde (capacité):
  - S_1 avec probabilité p
  - S_2 avec probabilité $q=(1-p)$
- Selon la formule d'entropie, la capacité du canal est $H(p)=-p \times \log(p)+(1-p) \times \log(1-p)$

Par exemple, avec **p**=20%,
H≈0,7219
Il y a donc perte d'efficacité puisque avec un canal de **1** bit/s, on ne transmet que **0,72** unité d'information par seconde. On parle de **redondance**



Canal sans bruit : vers le codage optimal

- Il y a contradiction:
  - Canal qui a une capacité de 1 symbole binaire par seconde ;
  - 0.7219 élément d'information transmis
- Le moyen de résoudre la contradiction:
  - Grouper les symboles (par 3 dans cet exemple)
  - Code à longueur variable (longueur ↗ si probabilité ↘)

symboles à coder	probabilité du triplet	codage du triplet	longueur du code
$s_1s_1s_1$	$0.8^3 = 0.512$	0	1
$s_1s_1s_2$	$0.8^2 \times 0.2 = 0.128$	100	3
$s_1s_2s_1$	$0.8^2 \times 0.2 = 0.128$	101	3
$s_2s_1s_1$	$0.8^2 \times 0.2 = 0.128$	110	3
$s_1s_2s_2$	$0.2^2 \times 0.8 = 0.032$	11100	5
$s_2s_1s_2$	$0.2^2 \times 0.8 = 0.032$	11101	5
$s_2s_2s_1$	$0.2^2 \times 0.8 = 0.032$	11110	5
$s_2s_2s_2$	$0.2^3 = 0.008$	11111	5

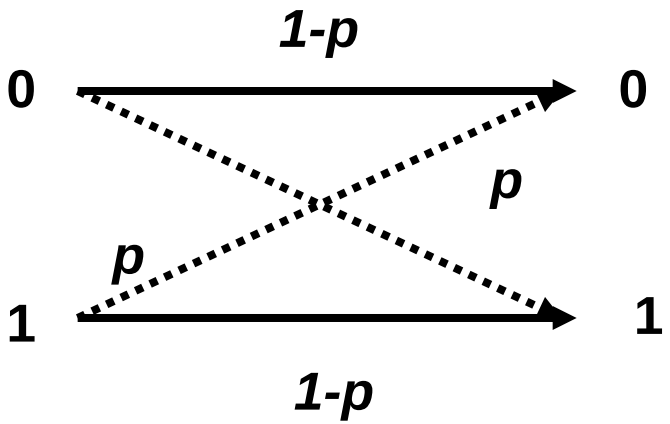
Canal sans bruit: 1er théorème de Shannon

- Avec la méthode de codage précédent, la longueur moyenne sera de **0,7280** élément binaire par symbole et approche de très près la valeur théorique de l'entropie de **0,7219**;
- L'écart entre longueur moyenne et valeur de l'entropie s'appelle la **redondance** (inverse du rendement). Le codage a permis de réduire la redondance de 28%(1-0,7219) à 0,20% (0,7280- 0,7219);

- Le **1er théorème de Shannon** dit que, à la limite, l'entropie transmise dans un canal peut en égaler la capacité en groupant des symboles dans des blocs arbitrairement longs et avec des codes optimisés (on pourrait dire avec compression sans perte d'information ou encore **compression entropique**)

Canal avec bruit: 2eme théorème de Shannon

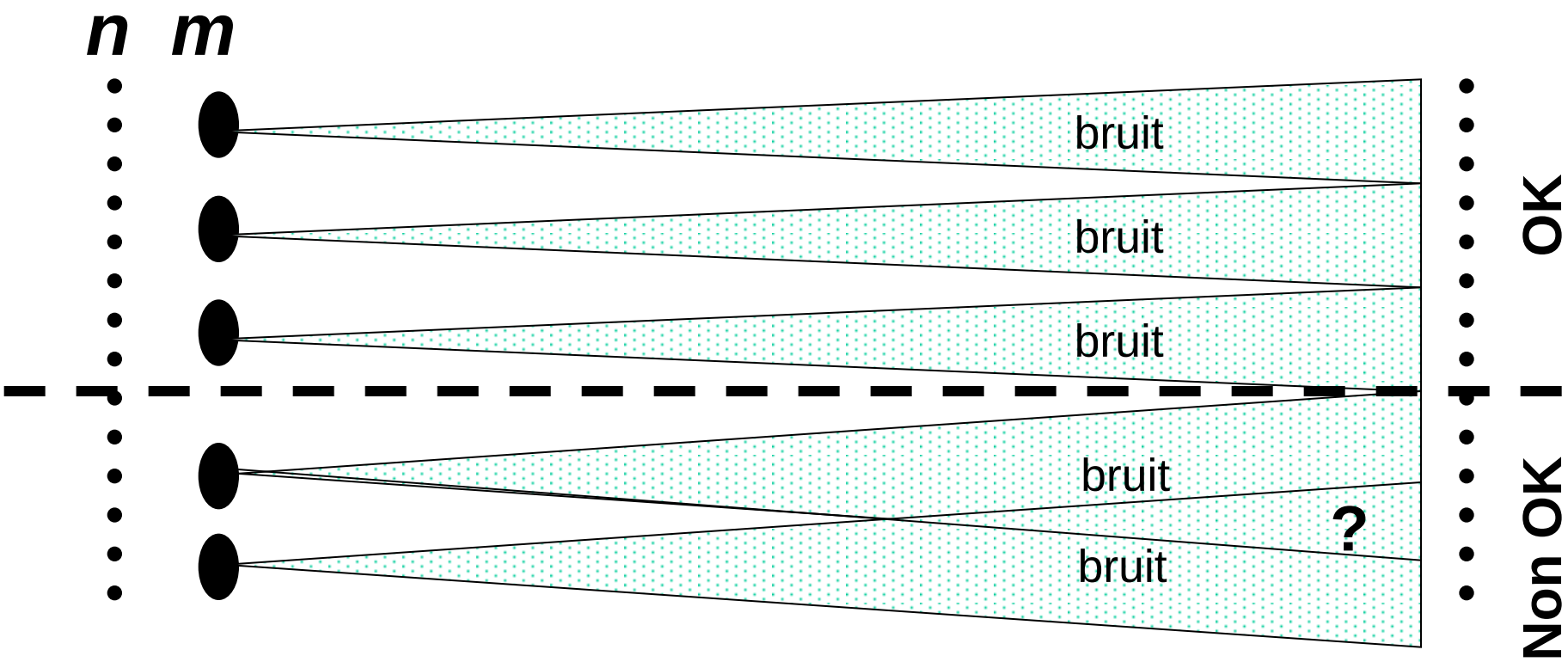
Un canal binaire transmet avec une probabilité d'erreur ***p*** (taux d'erreur). Sa capacité de 1 bit/s va être réduite par le bruit



- Le bruit introduit sur le canal une déperdition d'entropie de $\mathbf{H_b}$=-***p*** x log (***p***)+(***1-p***) x log(***1-p***)

- **Le 2eme théorème de Shannon** dit que:
  1. La capacité maximale d'un canal avec bruit est $\mathbf{1-H_b}$
  2. Cette capacité peut être atteinte à la limite en transmettant des blocs d'information arbitrairement grands et en introduisant la redondance permettant de corriger les erreurs sur les blocs ;

Canal avec bruit: 2eme théorème de Shannon

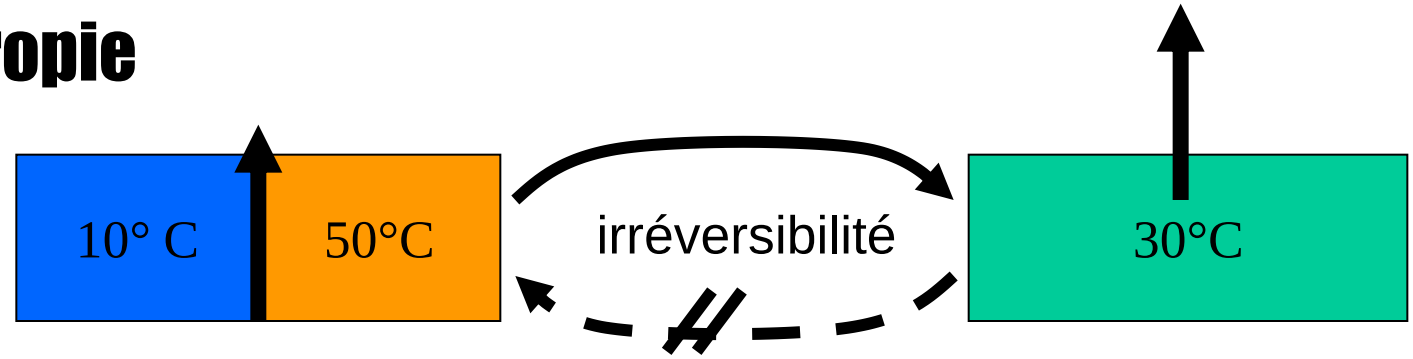


- Blocs de ***n*** bits:
  - 2^n messages possibles
  - ***m*** messages significatifs (***m*** < 2^n) séparés par une distance suffisante
- Malgré le bruit, les messages à la réception doivent être discernables, grâce à la « distance » entre les ***m*** messages possibles ;

Redondance et correction d'erreurs

- Le principe de la correction d'erreurs est donc d'introduire une redondance permettant de détecter et de corriger (en probabilité) les erreurs ;
- Une méthode « bovine » pourrait être par exemple de répéter n fois chaque symbole: peu efficace ;
- L'approche plus avancée consiste à grouper les symboles dans des longs blocs; on y ajoute des symboles de contrôle de telle façon que les groupes significatifs de symboles soient suffisamment séparés par ce que l'on appelle la distance de **Hamming** : cette distance doit être plus grande que les perturbations dues au bruit pour pouvoir corriger les erreurs ;
- Les **turbo-codes**, inventés dans les années 1990 par deux chercheurs de l'ENST de Brest permettent d'approcher les limites théoriques de Shannon. Ils sont utilisés pour la correction des erreurs dans l'UMTS et autres systèmes modernes ;

Entropie



- Un bac d'eau est séparé en deux par une cloison mobile, avec des températures différentes ;
- On enlève la cloison: progressivement les températures s'égalisent;
- Interprétations:
  - L'état final est beaucoup plus probable que celui de la coexistence de deux masses d'eau à températures différentes: le désordre augmente ;
  - Le phénomène est irréversible ;
- En termes scientifiques, on dira que **l'entropie** du système a augmenté;
- Le même schéma avec des gaz a donné lieu au paradoxe du démon de Maxwell: un démon capable de trier les molécules selon leur vitesse (ou leur température) pourrait, sans dépense d'énergie, diminuer l'entropie du système! On démontre que c'est impossible

Entropie en mécanique statistique

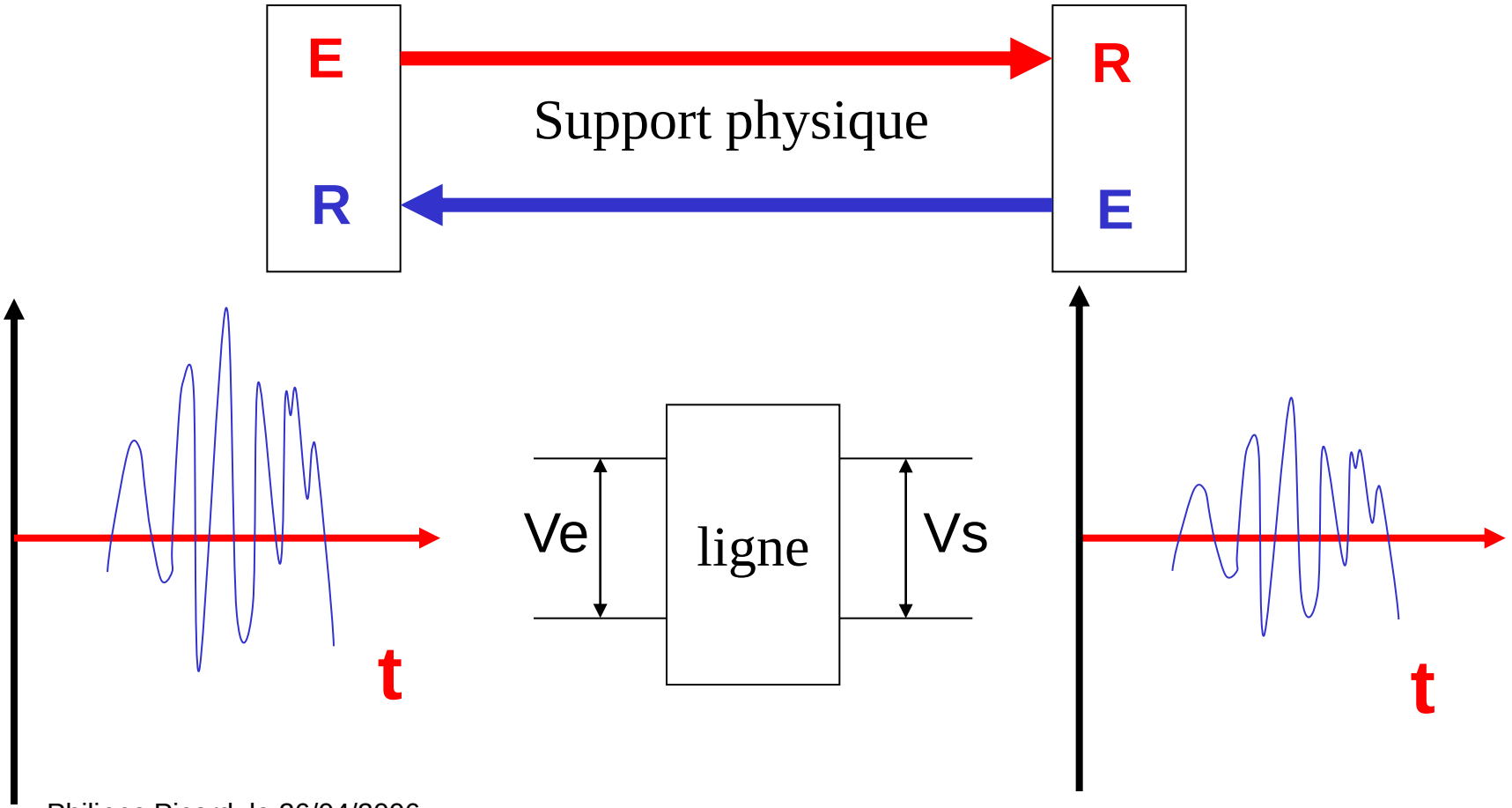
- La notion d'entropie, inventée par Clausius étend le 2eme principe de Carnot. Ce principe établit que le rendement d'une machine à vapeur dépend de la différence de température entre sources chaude et froide ;
- L'entropie mesure la dégradation de l'énergie dues aux transformations d'énergie et est associée à l'irréversibilité de ces transformations ;
- Ludwig Boltzmann, qui développé la mécanique statistique, a démontré que cette variable s'exprime par une formule similaire à celle définie ultérieurement par Claude Shannon pour mesurer l'information. En thermodynamique, elle mesure le désordre des configurations ;
- Léon Brillouin a démontré le lien rigoureux entre les deux concepts;

Théorie de l'information et canaux physiques

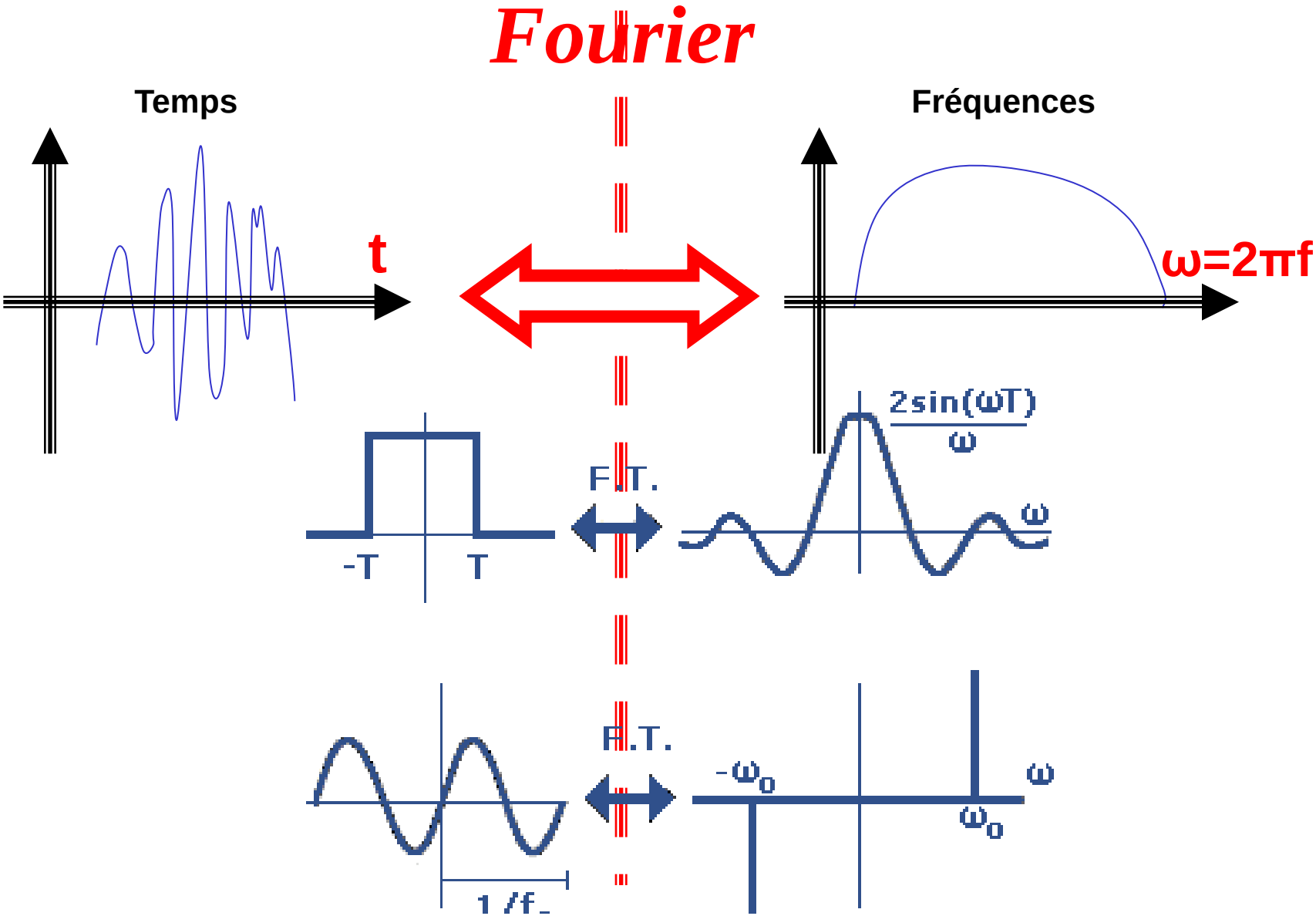
- Un canal physique sera caractérisé par la façon dont il transmet les signaux (bande de fréquences, bruit);
- Shannon et Nyquist ont démontré deux théorèmes fondamentaux:
  - Théorème d'échantillonnage, applicable à toutes les informations analogiques (voix et musique, images et vidéo);
  - L'expression de la capacité maximale d'un canal physique en fonction du bruit et de la largeur de bande de fréquences du canal ;
- C'est la base des télécom numériques d'aujourd'hui !

La transmission physique

Quel que soit le système de télécom, on transporte l'information par un champ électromagnétique (par fil métallique, fibre optique ou onde hertzienne).



La transmission physique



Transformations analogiques / numériques

Il s'agit soit de:

1. **Faire passer des informations numériques sur un canal physique** (cuivre, FO, radio) c'est le rôle des **modems**

2. **Représenter et transmettre l'information "analogique" sous forme numérique** (voix, fax & images fixes, CD, TNT, etc.). La numérisation est pratiquée depuis plus de 30 ans dans le cœur du réseau téléphonique.

Transformations analogiques / numériques

1: envoyer une information numérisée dans un canal physique de largeur de bande B

- **Capacité théorique** d'un canal analogique (exprimée en bits/s, éléments d'information par seconde): **Formule de Shannon**

Capacité (bits /s) < B (Hz) x $\log_2$ (1+ Signal / Bruit)

- **Rapidité de modulation** maximale sur un canal de largeur de bande B: **fréquence de Nyquist**

R (Bauds) < 2 x B(Hz)

- Ces deux principes de base guident la technologie et les performances des transmissions numériques

Modems

Largeur de bande

Bruits

$R_{Bauds} < 2 \times B_{Hz}$

V (valence)

$D\ (bits/s) = R \times \log_2 V$

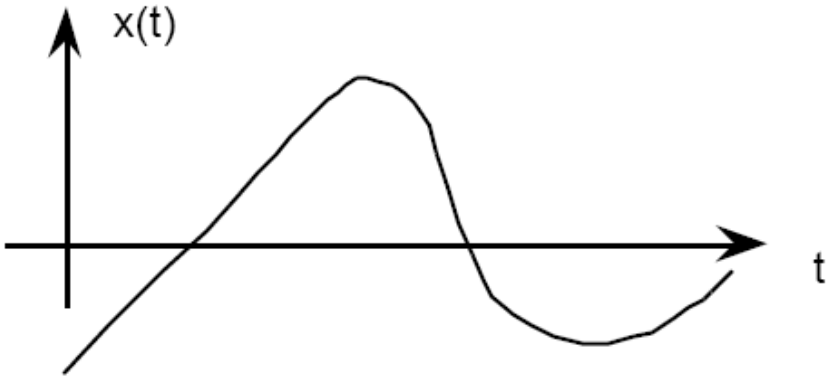
Transformations analogiques / numériques

2: transmettre une information analogique via un canal numérique

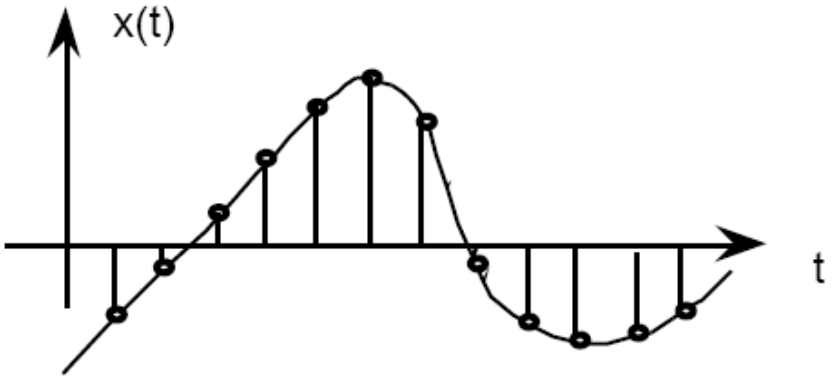
- **Echantillonnage** d'un signal de largeur de bande B: **Théorème de Shannon – Nyquist:** il faut que la **fréquence d'échantillonnage** soit au moins **2 x B** pour pouvoir représenter le signal sans perte d'information

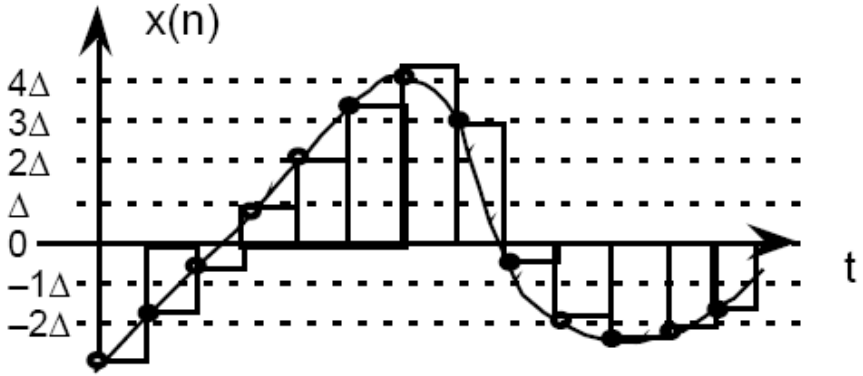
La numérisation d'un signal

Signal à numériser



Echantillonnage





Quantification

Annexes

- Lectures
- Rappels:
  - les log
  - Probabilités conditionnelles

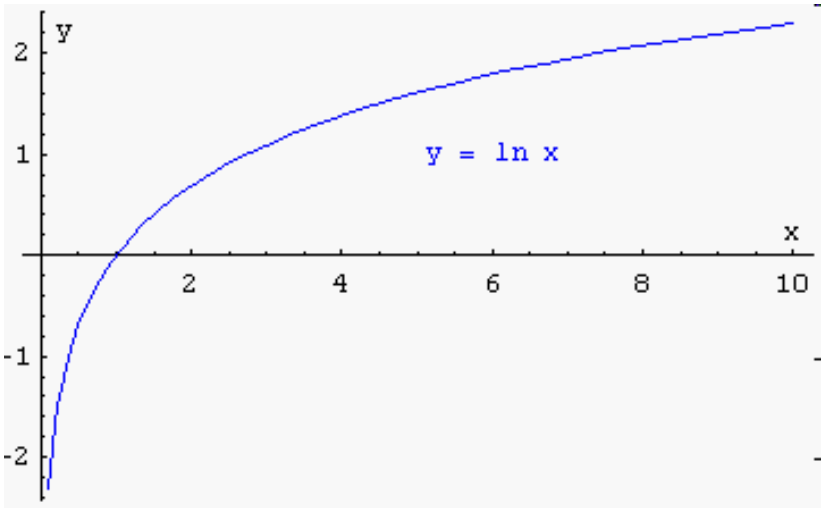
Lectures

- Les probabilités par Albert Jacquard
  - Que sais-je N°1571
- Histoire des codes secrets (Simon Singh)
- Pourquoi le tout est plus que la somme de ses parties, pour une approche scientifique de l'émergence (Jacques Ricard)
  - Chapitre 7, communication et information
- Wikipedia:
  - Théorie de l'information, Shannon
  - Entropie
- Mes livres de semi-vulgarisation sont anciens (60's)

Rappel sur les logarithmes

$$y=b^x \Leftrightarrow x=\log_b(y)$$

- **b** est la base:
  - **Base 10**: la table « Bouvard et Ratinet » de notre jeunesse!!! Utilisé en télécom: les décibels: $A_{db} = 10 \log_{10} (P_{sortie}/P_{entrée})$
  - **Base e**: logarithme népérien (voir cours de math)
  - **Base 2**: numérotation binaire
- Propriétés:
  - log (AxB)=log (A)+log (B)
  - log (1/A)=-log (A)
  - log (A^B)=B x log (A)
  - log (1)=0
  - log (X→0)→-∞
  - si x→0, x.log(x)→0



Formule de Bayes et Information mutuelle

- En toute rigueur, la théorie de l'information fait appel à la notion d'information mutuelle, basée sur le théorème de Bayes.
- Etant donné deux évènements A et B, le théorème de Bayes permet de déterminer la probabilité de A sachant B, si l'on connaît les probabilités de A, de B, de B sachant A. Cela s'écrit:
  - En français: la probabilité de A **et** B est égale à la probabilité de A multipliée par la probabilité de B connaissant A
  - En formule:

$$p(A \cap B) = p(A) \times p(B|A)$$

Formule de Bayes et Information mutuelle

- L'information mutuelle mesure la quantité d'information apportée en moyenne par une réalisation de X sur les probabilités de réalisation de Y.
- En considérant qu'une distribution de probabilité représente notre connaissance sur un phénomène aléatoire, on mesure l'absence d'information par l'entropie de cette distribution. En ces termes, l'information mutuelle s'exprime par:

$$I(X,Y) = H(X) - H(X \mid Y) = H(Y) - H(Y \mid X) = H(X) + H(Y) - H(X,Y)$$

où H(X) et H(Y) sont des entropies, H(X|Y) et H(Y|X) sont des entropies conditionnelles, et H(Y, X) est l'entropie conjointe entre X et Y.