Physique Statistique (PHY432)

Amphi 7

Physique Statistique
et
Théorie de l'Information

Théorie de l'Information

Une branche de la science:

- Récente (60 ans)
- Très importante du point de vue technologique
- Aux multiples ramifications

- ... et intimement reliée à la Physique Statistique

Théorie de l'Information

Données (nombres, texte, images,...)

- Quantifier l'information qu'elles contiennent.
- Comment les comprimer (gif,jpeg,mp3...)?
- Comment les transmettre à travers des canaux de transmission bruités?

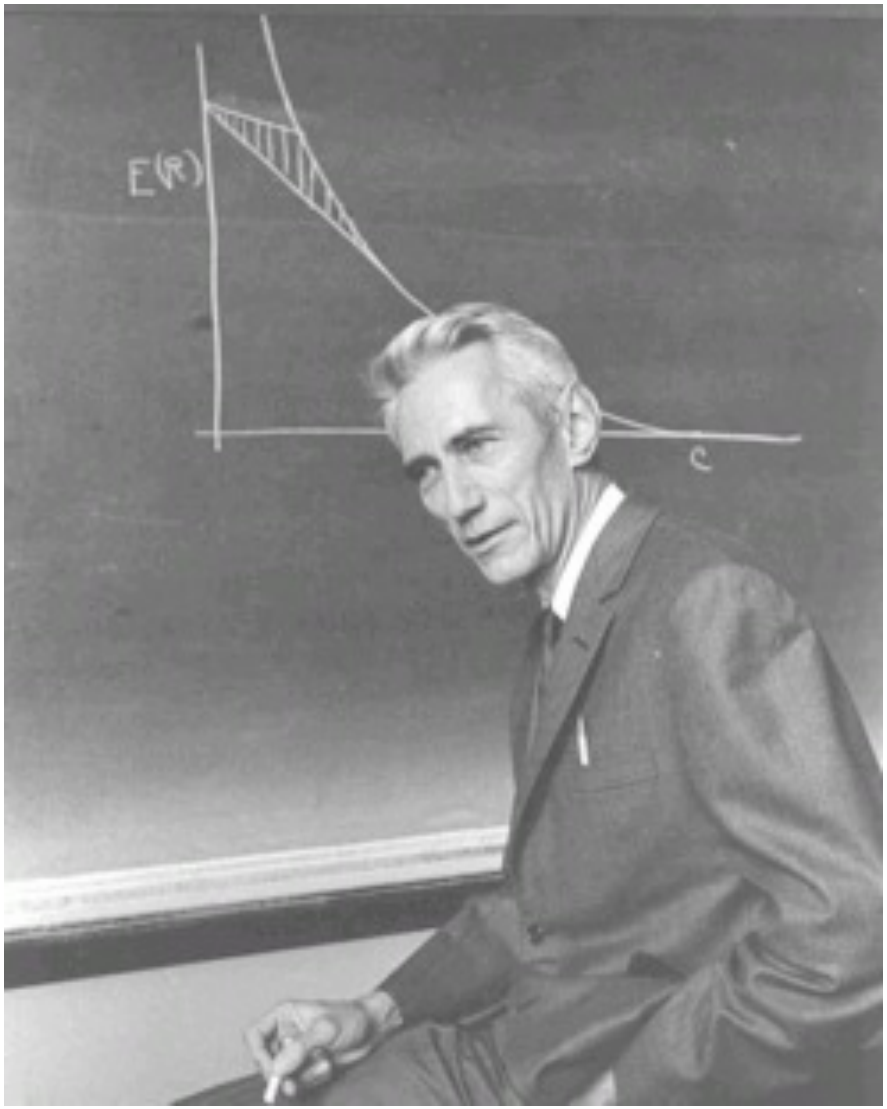
Exemples:

- Communication téléphonique
- Communication radio avec sonde orbitale Cassini
- Reproduction de cellule: réplication de l'ADN
- Ecriture/Lecture sur disque dur

Théorie de l'Information

Claude Shannon
(1916-2001)

Pendant la guerre:
services secrets,
cryptographie



→ "A mathematical theory of communications" (1948)

1948: 1800 conversations sur un câble
2005: 6 400 000 conversations sur fibre optique

L’Information, qu’est-ce que c’est?

Information

On apprend une nouvelle. Si cette nouvelle était a priori improbable: gain d'une grande information. Si cette nouvelle était a priori très probable: gain d'une faible information

Shannon: si la nouvelle avait probabilité p , l'information gagnée lorsqu'on reçoit la nouvelle est $\log_2(1/p)$ (postulat)

Soient N événements, probabilités $p_1, \ldots, p_N$, avec $\sum_{i=1}^{N} p_i = 1$

L'information moyenne reçue si on fait un tirage au hasard vaut

$$H = \sum_{i=1}^{N} p_i \log_2(1/p_i) = -\sum_{i=1}^{N} p_i \log_2 p_i$$

Information

Postulat. L'information manquante associée à la loi de probabilité $p_1, \ldots, p_N$ vaut:

$$H = -\sum_{i=1}^{N} p_i \log_2 p_i$$

Expression générale de l'entropie, pour tous les ensembles (canonique, microc., grand-canonique..) (A3):

$$S = -k \sum_{i=1}^{N} p_i \log p_i$$

L'entropie est l'information manquante de la loi de probabilité physique sur les configurations (avec un changement d'unité de mesure

H en "bits" ← $k = 1/\log 2$ → $k = 1.4\ 10^{-23} J/K$)

Information

$$H = -\sum_{i=1}^{N} p_i \log_2 p_i$$

Exemples

Dé à 4 faces: $p_1 = p_2 = p_3 = p_4 = 1/4$

$$H = -4[(1/4)\log_2(1/4)] = 2$$

$H = 2$ bits

Dé à 4 faces pipé: $p_1 = 1/2,\ p_2 = 1/4,\ p_3 = p_4 = 1/8$

$$H = (1/2)\log_2 2 + (1/4)\log_2 4 + 2[(1/8)\log_2 8]$$

$$= 1/2 + 1/2 + 2(3/8) = 7/4$$

$H = 7/4$ bits

(résultat moins incertain)

Information

$$H = -\sum_{i=1}^{N} p_i \log_2 p_i$$

Propriétés:

- $H \geq 0\,;\ H = 0$ si et seulement si événement certain: $p_i = \delta_{i,n}$

- Paire d'événements indépendants

"Le cheval 12 *a gagné dans la troisième course;*
Le numéro gagnant du loto est 137*"*

$$H \overset{?}{=} H_1 + H_2$$

Information

$$H = -\sum_{i=1}^{N} p_i \log_2 p_i$$

Propriétés:

- $H \geq 0$; $H = 0$ si et seulement si év. certain: $p_i = \delta_{i,n}$

- Paire d'événements indépendants $P(n,a) = p_n q_a$

$$H(\{p_n q_a\}) = -\sum_{n,a} p_n q_a \log_2 (p_n q_a)$$

$$= -\sum_{n,a} p_n q_a (\log_2 p_n + \log_2 q_a)$$

$$H(\{p_n q_a\}) = H(\{p_n\}) + H(\{q_a\})$$

Information

$$H = -\sum_{i=1}^{N} p_i \log_2 p_i$$

Propriétés:

- $H \geq 0$; $H = 0$ si et seulement si év. certain: $p_i = \delta_{i,n}$

- Paire d'événements indépendants $P(n, a) = p_n q_a$

$$H(\{p_n q_a\}) = H(\{p_n\}) + H(\{q_a\})$$

- Paire d'événements dépendants

"Le cheval 12 *a gagné dans la troisième course;*
Le cheval 3 *est arrivé deuxième"*

Information

$$H = -\sum_{i=1}^{N} p_i \log_2 p_i$$

Propriétés:

- $H \geq 0$; $H = 0$ si et seulement si év. certain: $p_i = \delta_{i,n}$

- Paire d'événements indépendants $P(n,a) = p_n q_a$

$$H(\{p_n q_a\}) = H(\{p_n\}) + H(\{q_a\})$$

- Paire d'événements dépendants $P(n,a)$

Lois marginales: $p_n = \sum_a P(n,a)$; $q_a = \sum_n P(n,a)$

$$H = -\sum_{n,a} P(n,a) \log_2 P(n,a) < H(\{p_n\}) + H(\{q_a\})$$

Information

$$H = -\sum_{i=1}^{N} p_i \log_2 p_i$$

Propriétés:

- $H \geq 0$; $H = 0$ si et seulement si év. certain: $p_i = \delta_{i,n}$

- Paire d'événements indépendants $P(n,a) = p_n q_a$

$$H(\{p_n q_a\}) = H(\{p_n\}) + H(\{q_a\})$$

- Paire d'événements dépendants

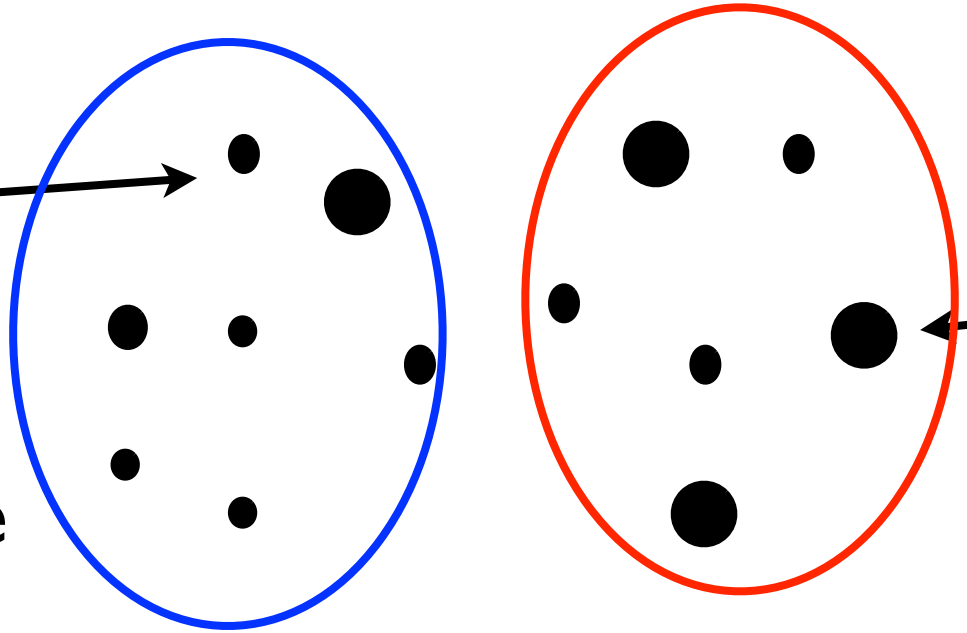
$$H < H(\{p_n\}) + H(\{q_a\})$$

"Le cheval 12 *a gagné dans la troisième course;*
Le cheval 3 *est arrivé deuxième"*

Information

$$H = -\sum_{i=1}^{N} p_i \log_2 p_i$$

Propriétés:



$p + q = 1$

Proba p_n

Proba $p_{n'}$

Proba conditionnelle $r_n = p_n/p$

Proba conditionnelle $s_{n'} = p_{n'}/q$

Proba p

Proba q

$$H(\{p_n\}) = -\sum_n p_n \log_2 p_n$$

$$= -p \log_2 p - q \log_2 q + pH(\{r_n\}) + qH(\{s_n\})$$

Ces propriétés définissent l'information manquante d'une loi de probabilité de façon unique (à la normalisation près)

$$H = -\sum_{i=1}^{N} p_i \log_2 p_i$$

Information

$$S = -k \sum_{i=1}^{N} p_i \log p_i$$

Entropie

Même expression. En quel sens l'entropie mesure-t-elle une absence d'information sur un système physique?

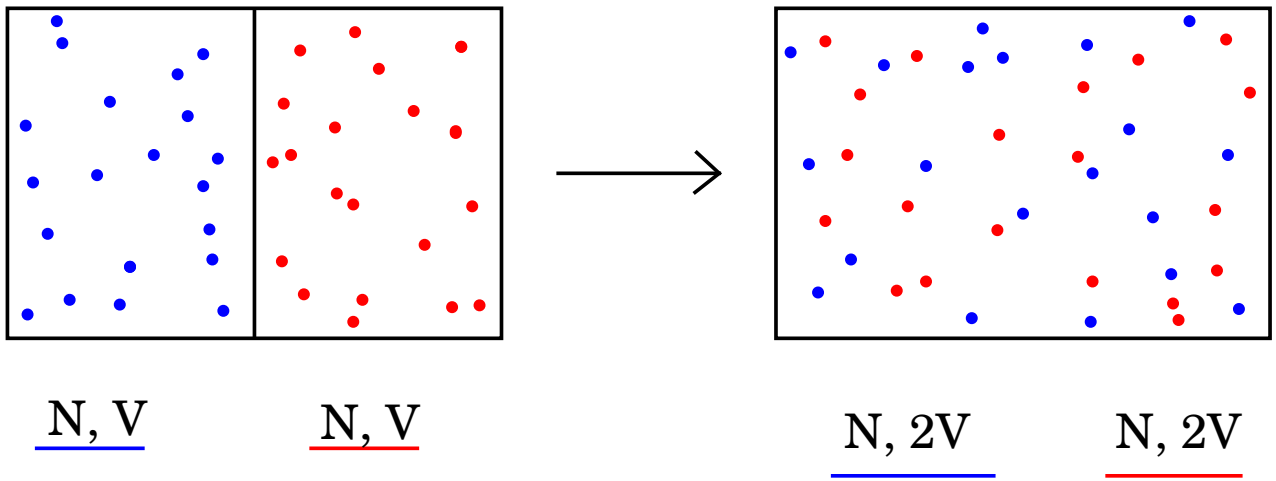
$$S = k \log W$$

$$S = k \log W$$

Système physique: W microétats accessibles. On ignore dans lequel il est. S'ils sont tous équiprobables (proba $1/W$): information manquante sur le microétat = $\log_2 W$

Information et entropie

Entropie de mélange :



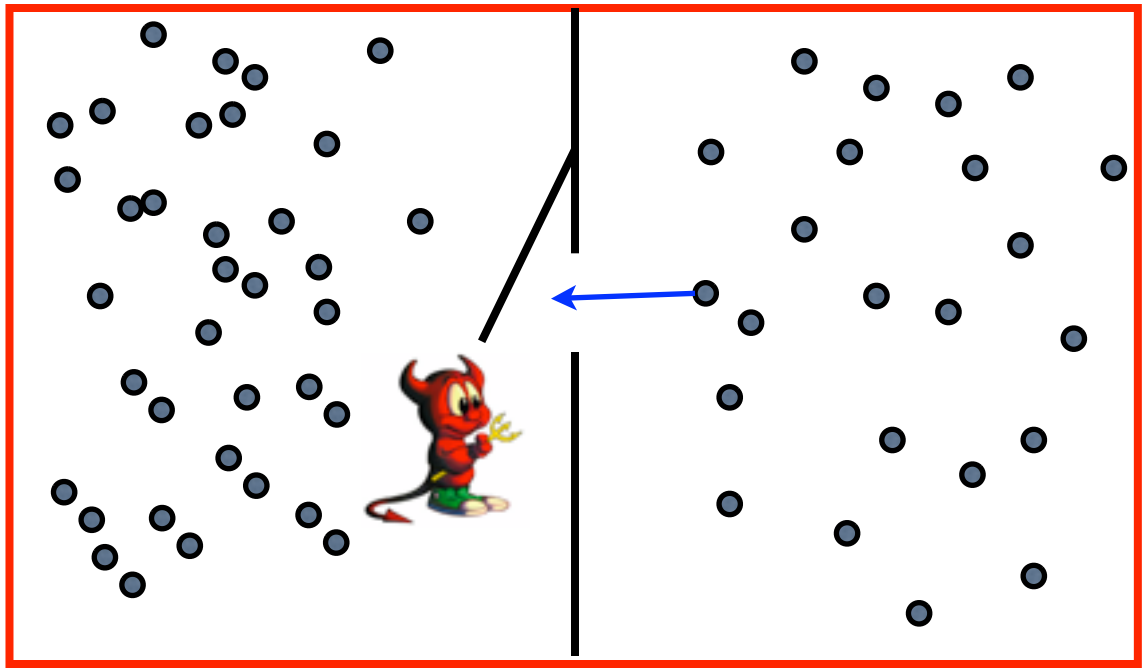
$$\Delta S = 2[S(N, 2V) - S(N, V)] = 2Nk \log 2$$

Pour chaque particule, incertitude sur la couleur :
perte d'information $\log 2$ (en bits) ou $k \log 2$

Information et entropie : Machines thermiques et démons

Démon de Maxwell

Crée une différence de pression: du travail à partir de la chaleur ?



Information et entropie : Machines thermiques et démons

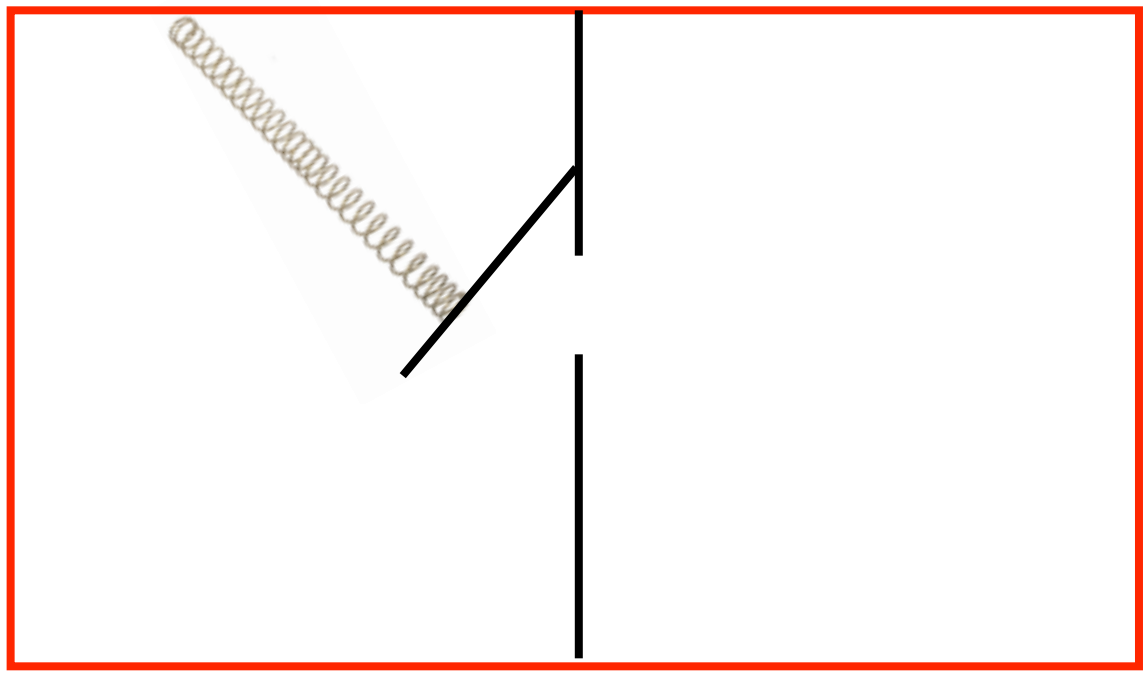
Trappe pour sélectionner les molécules rapides (crée une différence de pression):

Trappe de taille moléculaire. Equilibre

⇒ mouvement Brownien de la porte

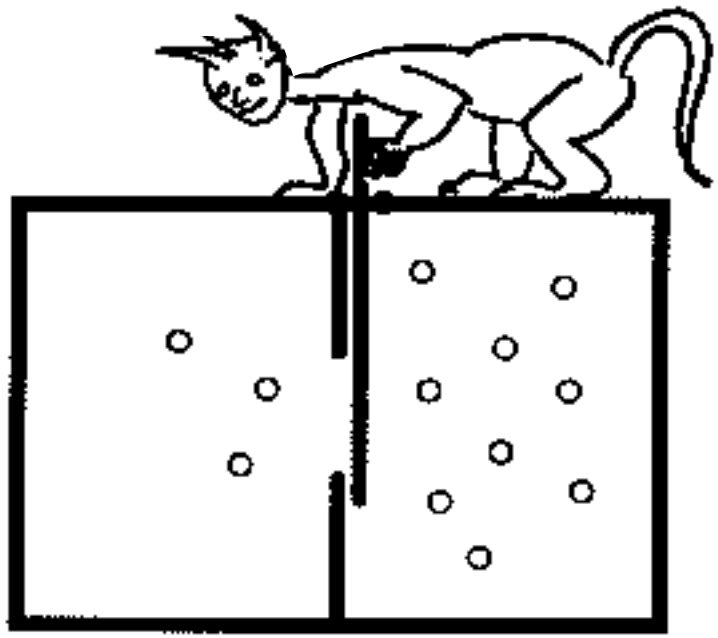
⇒ ne fonctionne pas

(Smoluchowski 1914)



Information et entropie : Machines thermiques et démons

Démon de Maxwell microscopique:



Observation d'un atome: diffusion de la lumière, absorption d'un photon par cellule photoélectrique, processus dissipatif qui crée de l'entropie

(Brillouin 1950)

Démon de Maxwell

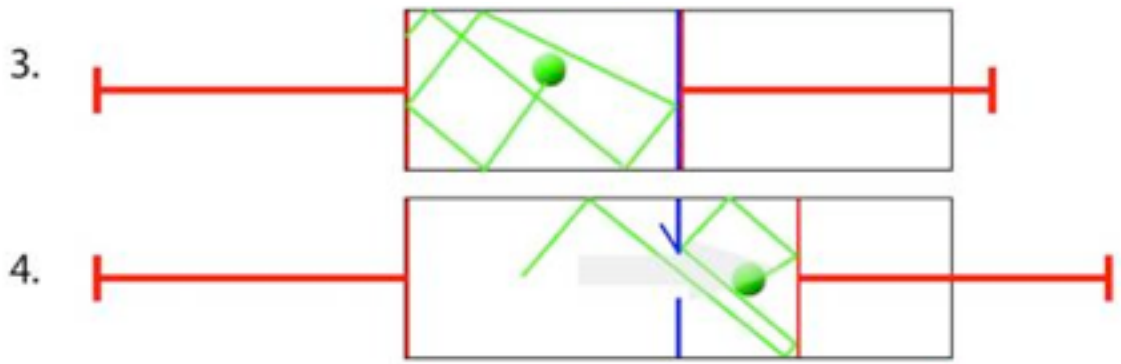
Machine de Szilard (1929)

Travail fourni dans un cycle de la machine

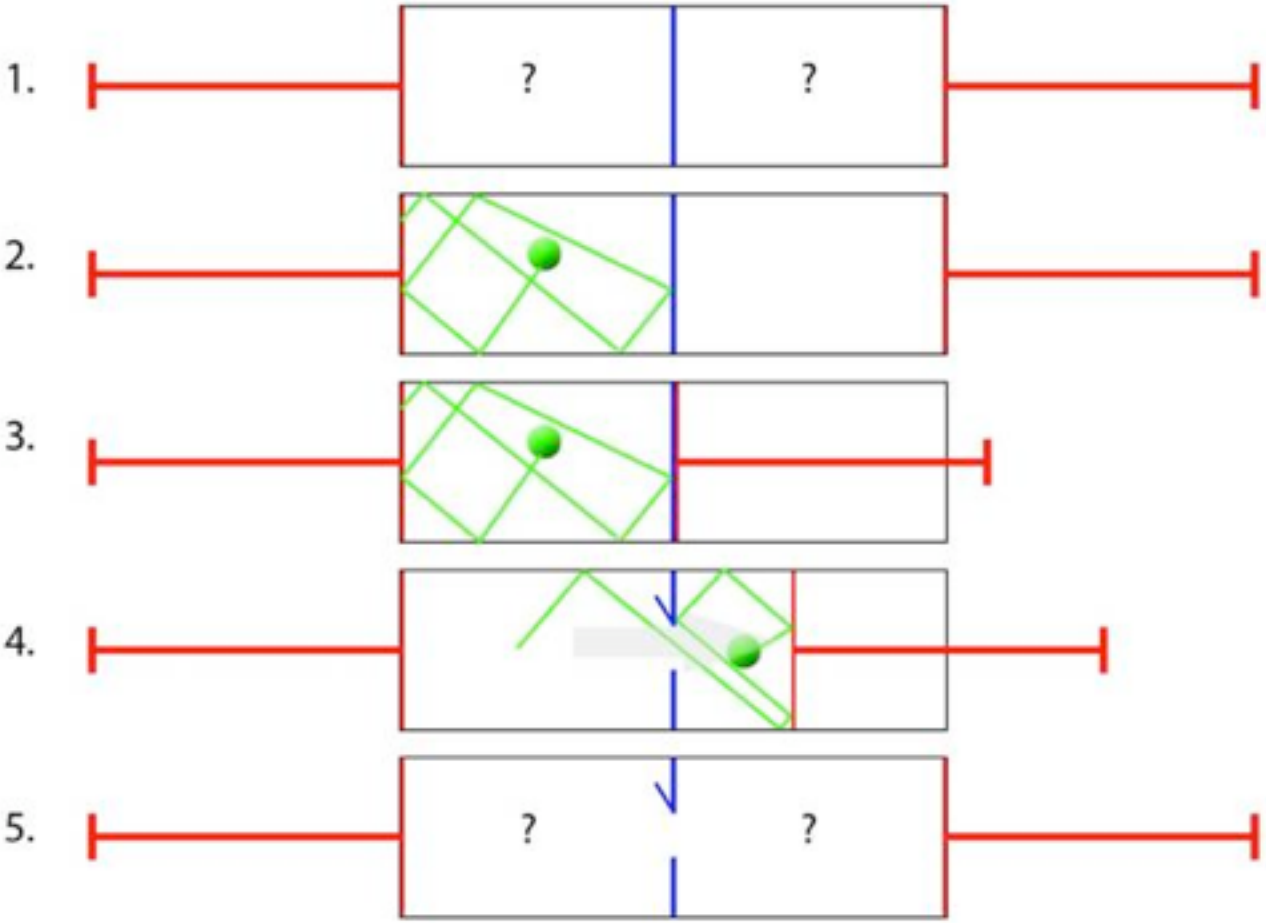
$$p(t) = \frac{kT}{V(t)} \qquad \langle F(t) \rangle = \frac{kT}{V(t)} S = \frac{kT}{L(t)}$$

$$dW(t) = \langle F(t) \rangle dL = kT \frac{dL}{L(t)}$$

$$W = kT \ln 2$$



Démon et information



Démon = acquisition d'information

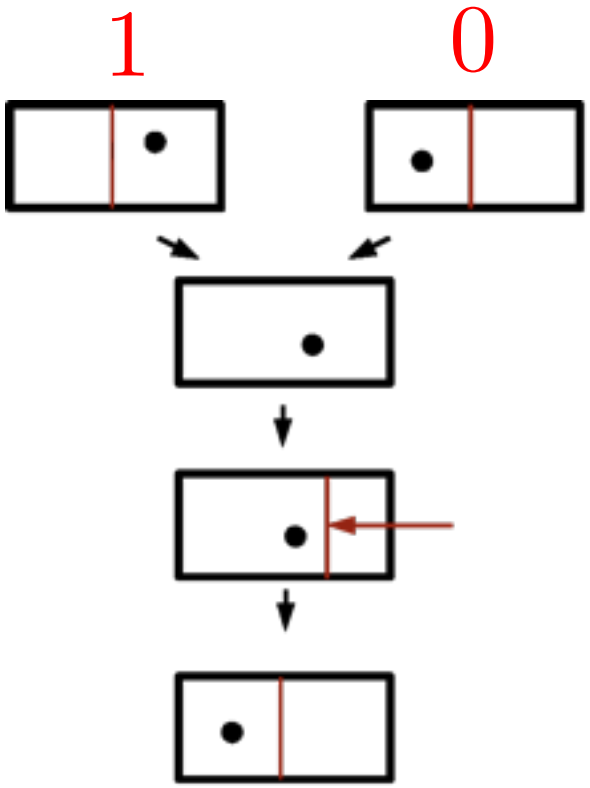
La machine doit savoir où est la particule

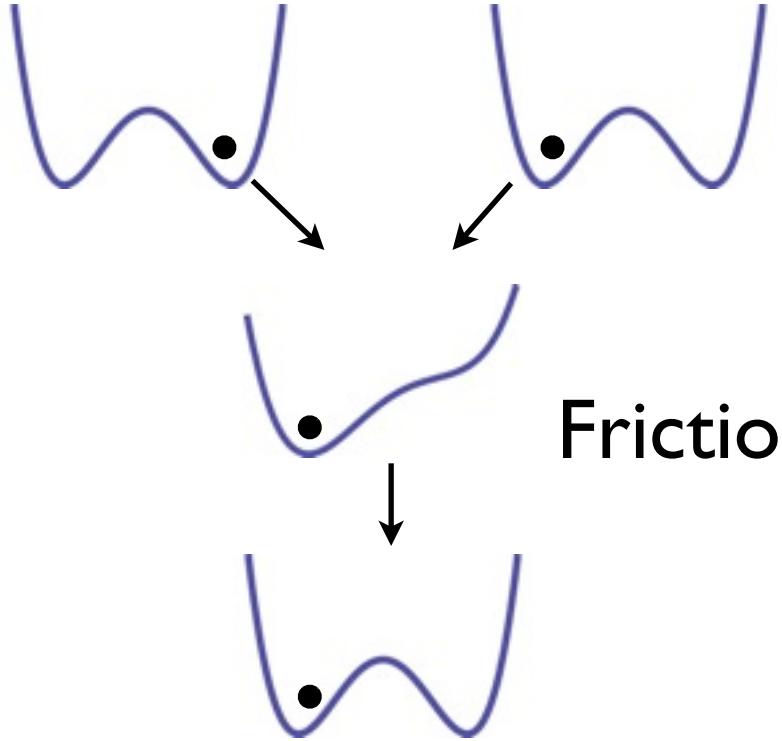
Argument de Landauer

La machine doit savoir où est la particule: 1 bit d'information.

Cycle parfait ⇨ il faut pouvoir effacer ce bit d'information

Reset= ramener tous les bits à 0



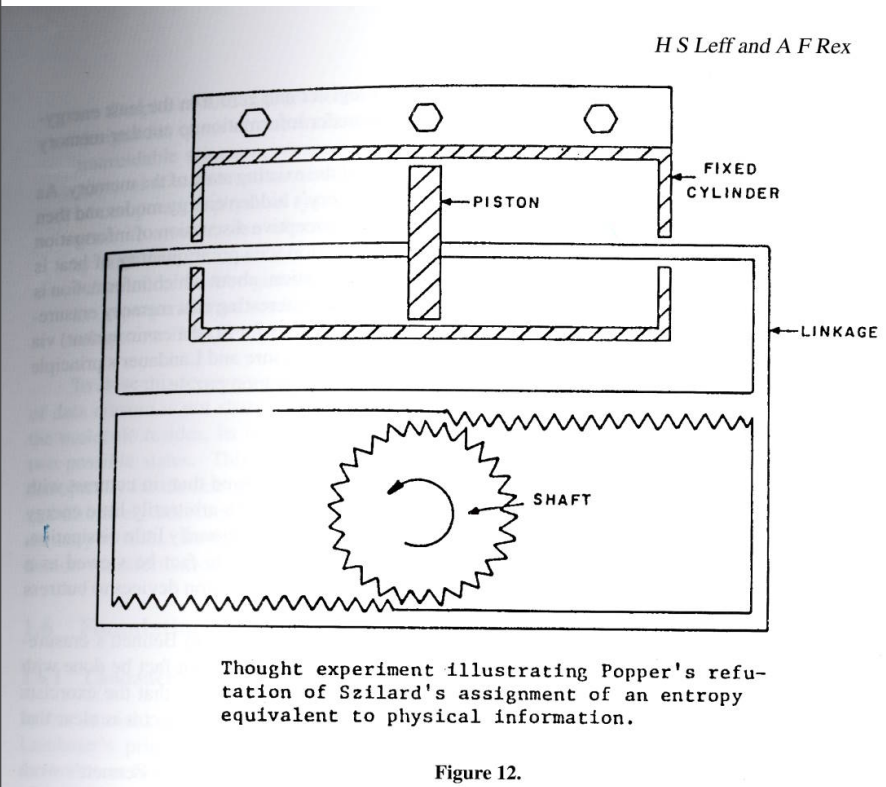


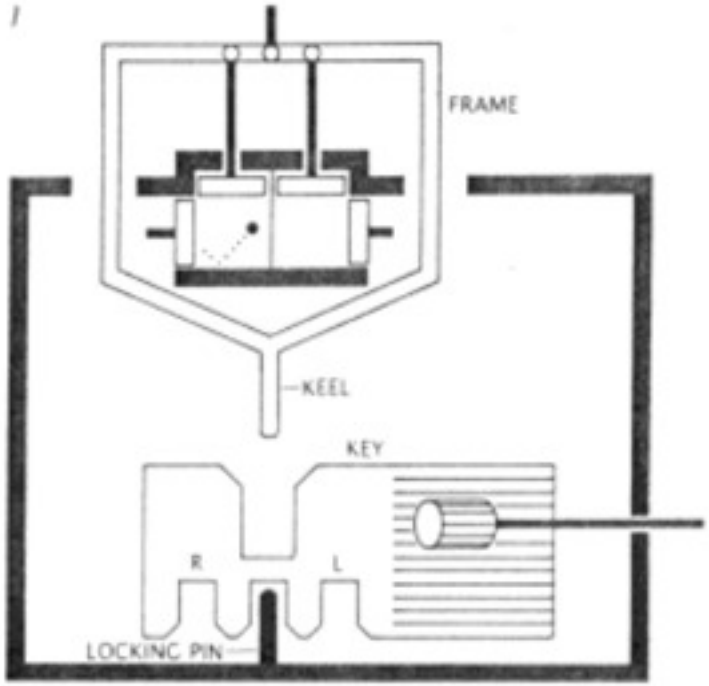
Travail pour le reset: $kT \ln 2$

Quelle que soit la machine, en un cycle:

Travail fourni par la machine: $W \leq kT \ln 2$

Travail fourni par le réservoir pour effacer le bit d'information $W \geq kT \ln 2$





Information manquante et physique statistique

Une autre approche de la physique statistique: construire la loi de probabilité des états, $\{p_n\}$qui maximise l'entropie compte tenu de ce qu'on sait sur le système.

"Loi la moins biaisée"

Exemple: système échangeant de l'énergie avec un thermostat. On impose deux contraintes:

$$\sum_n p_n = 1 \qquad \sum_n p_n E_n = U$$

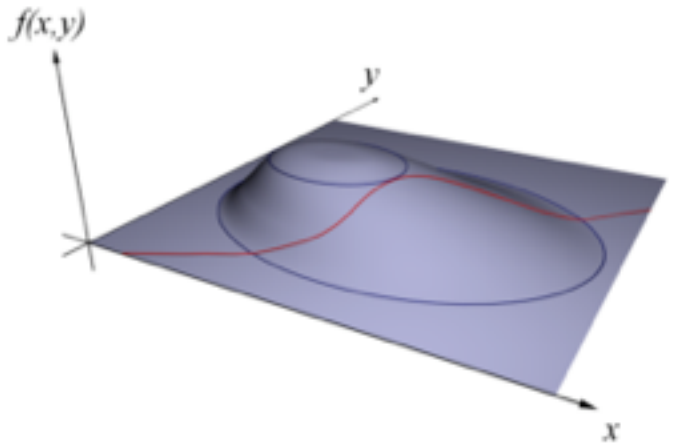
Pb: trouver la loi $\{p_n\}$ qui maximise $H = -\sum_n p_n \log_2 p_n$

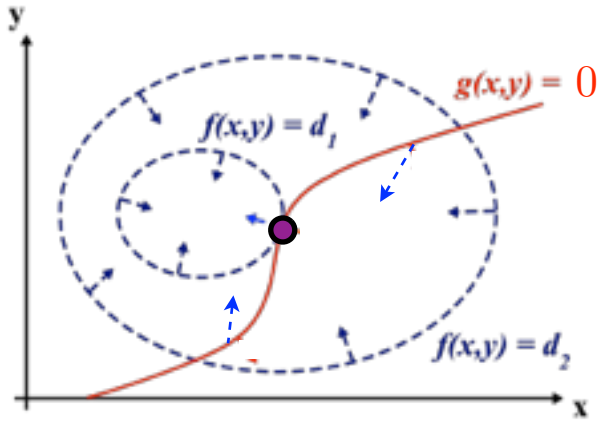
compte tenu des contraintes

Maximisation sous contrainte: multiplicateurs de Lagrange

Pb: maximiser $f(x_1, \ldots, x_N)$ sous la contrainte

$$g(x_1, \ldots, x_N) = 0$$





quand on se déplace sur la surface S: $g(x_1, \ldots, x_N) = 0$

il faut trouver le point où $\vec{\nabla} f$ est perpendiculaire à S

$\Leftrightarrow$ $\vec{\nabla} f \;//\; \vec{\nabla} g$ $\Leftrightarrow$ $\exists \lambda \,/\quad \vec{\nabla} f - \lambda \vec{\nabla} g = 0$

Maximisation sous contrainte: multiplicateurs de Lagrange

Pb: maximiser $f(x_1, \ldots, x_N)$ sous la contrainte

$$g(x_1, \ldots, x_N) = 0$$

$$\Leftrightarrow \quad \vec{\nabla} f \;/\!/\; \vec{\nabla} g \quad \Leftrightarrow \quad \exists \lambda \,/ \quad \vec{\nabla} f - \lambda \vec{\nabla} g = 0$$

Chercher les points stationnaires de

$$\hat{f}(x_1, \ldots, x_N, \lambda) = f(x_1, \ldots, x_N) - \lambda g(x_1, \ldots, x_N)$$

Si plusieurs contraintes: $g_1(x_1, \ldots, x_N) = 0$

$$g_2(x_1, \ldots, x_N) = 0$$

...

Chercher les points stationnaires de $f - \lambda_1 g_1 - \lambda_2 g_2 \ldots$

Information manquante et physique statistique

Maximiser $H = -\sum_n p_n \log_2 p_n$ avec les deux contraintes:

$$g_1(p_1, \ldots, p_N) = \sum_n p_n - 1 = 0$$

$$g_2(p_1, \ldots, p_N) = \sum_n p_n E_n - U = 0$$

Lagrange:

$$\exists \lambda_1, \lambda_2 \; / \;\; \vec{\nabla} H - \lambda_1 \vec{\nabla} g_1 - \lambda_2 \vec{\nabla} g_2 = 0$$

Ensemble canonique

... $\longrightarrow$ $p_n = \frac{1}{Z} e^{-\beta E_n}$

où β, Z sont tels que $\sum_n p_n = 1$ et $\sum_n p_n E_n = U$

Information manquante et physique statistique

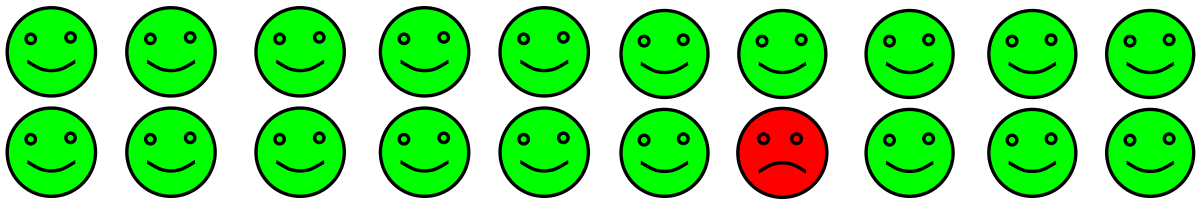
Une autre approche de la physique statistique: construire la loi de probabilité des états, $\{p_n\}$ qui maximise l'entropie compte tenu de ce qu'on sait sur le système.

"Loi la moins biaisée"

Retrouve tous les ensembles statistiques (microcanonique, canonique, grand canonique,...) à partir de ce seul postulat

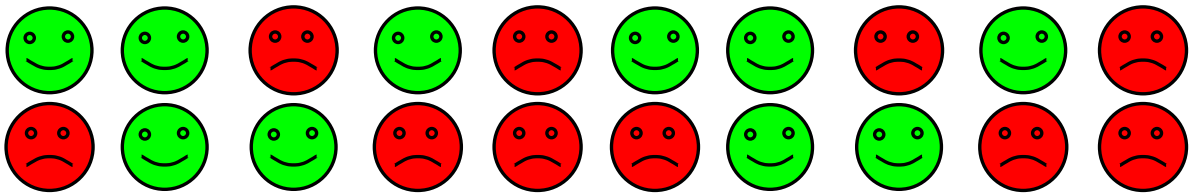
Contenu intuitif de l'entropie

Amphi



$S \sim \mathrm{Log}_2 N$

ou



$S \sim N$

S (ou H) donne le nombre de bits nécessaires pour enregistrer un événement

Compression de données

Symboles ou "mots" à mémoriser: $M_1, M_2, \ldots, M_N$

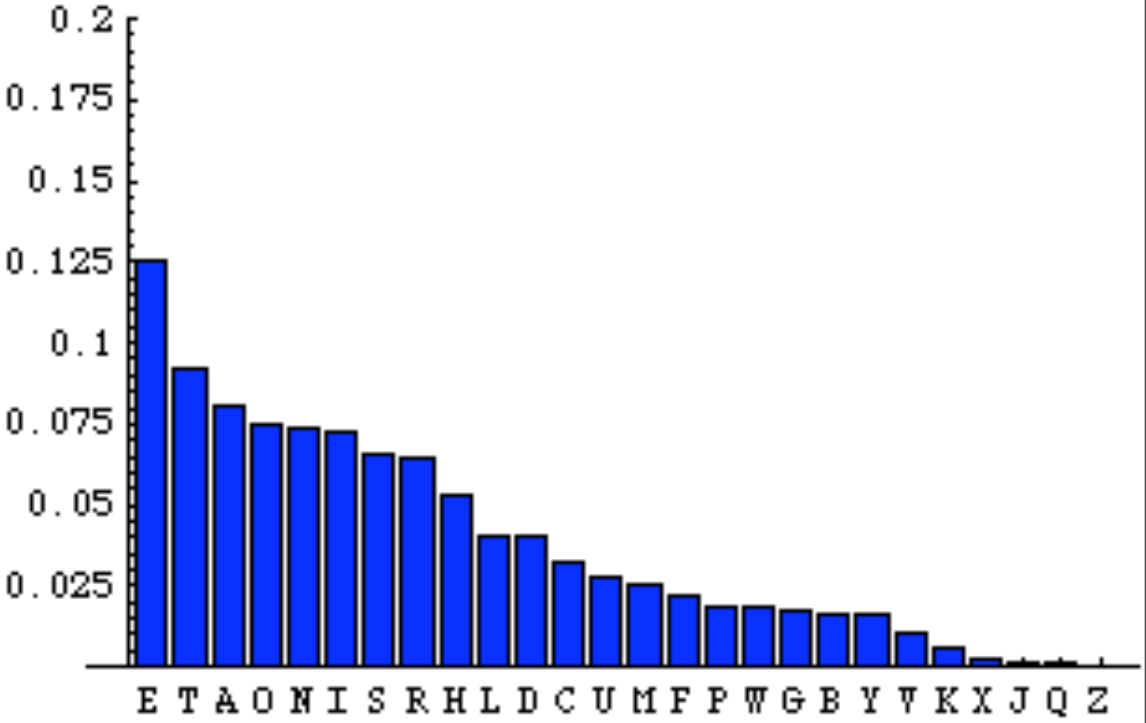
Probabilités (fréquence d'apparition): $p_1, p_2, \ldots, p_N$

Codage: $M_n \rightarrow C_n = 0110010110$

Pb: trouver un codage ($= \{C_n\}$) tel que le nombre de bits utilisé en moyenne soit le plus petit possible.

Principe: utiliser des C_n courts pour les mots fréquents

Exemple: alphabet morse

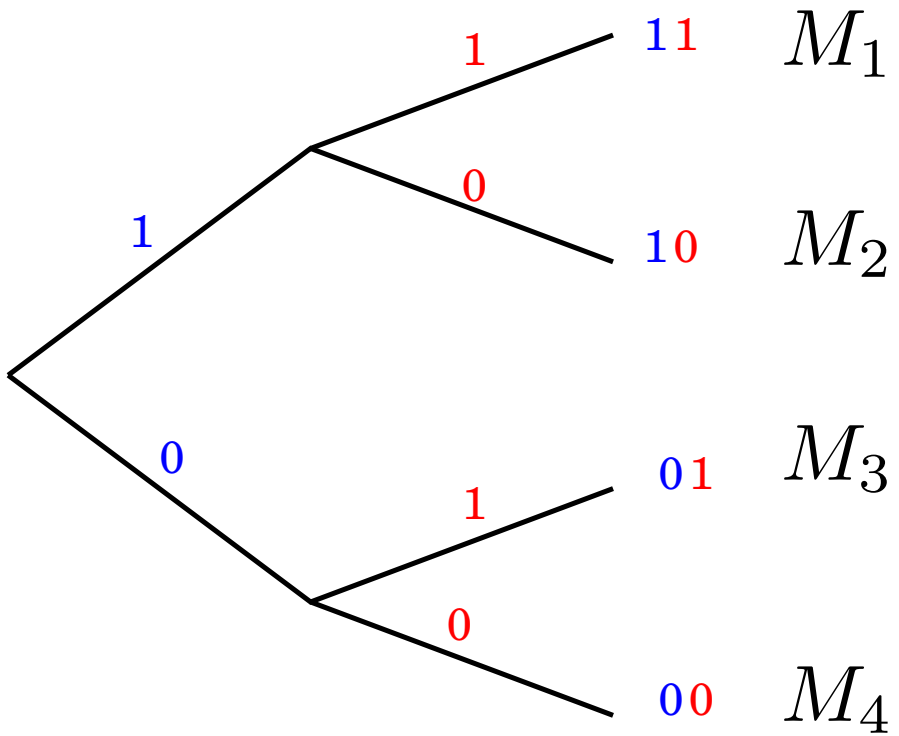


Fréquence des lettres en anglais

Compression de données

Codage: $M_n \to C_n = 0110010110....$

Code A:



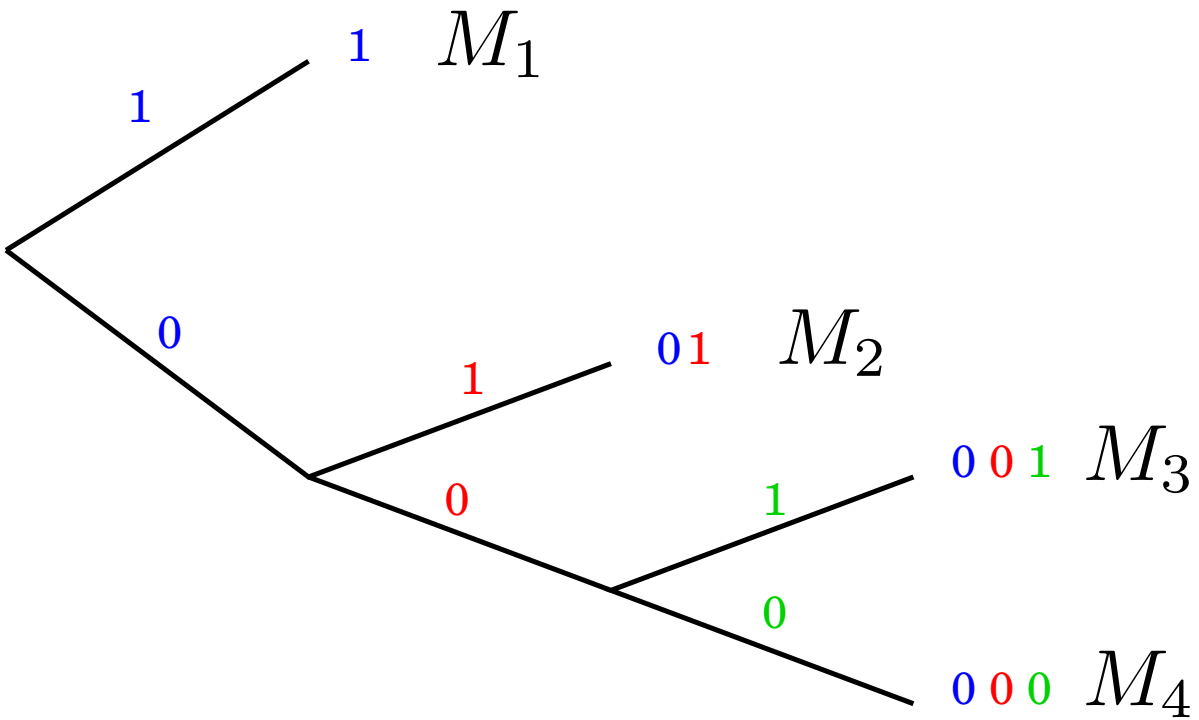
$11\,01\,100010$

$M_1\,M_3\,M_2\,M_4\,M_2$

Compression de données

Codage: $M_n \rightarrow C_n = 0110010110$

Code B:

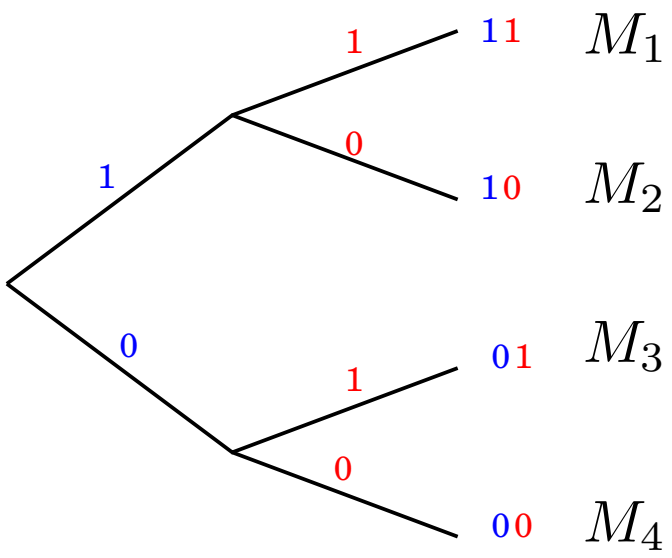


1011101000

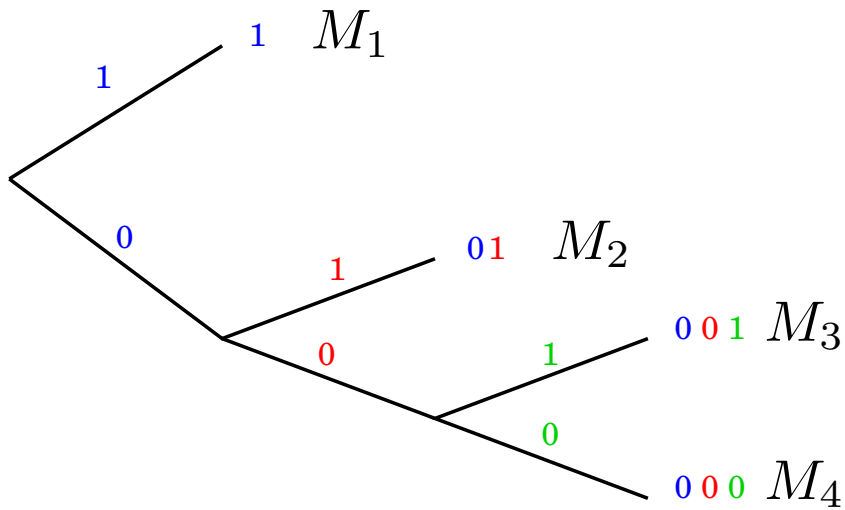
$M_1 M_2 M_1 M_2 M_4$

Compression de données

Code A:



Code B:



Lequel est le meilleur?

$$\langle L \rangle = \sum_n p_n L(M_n)$$

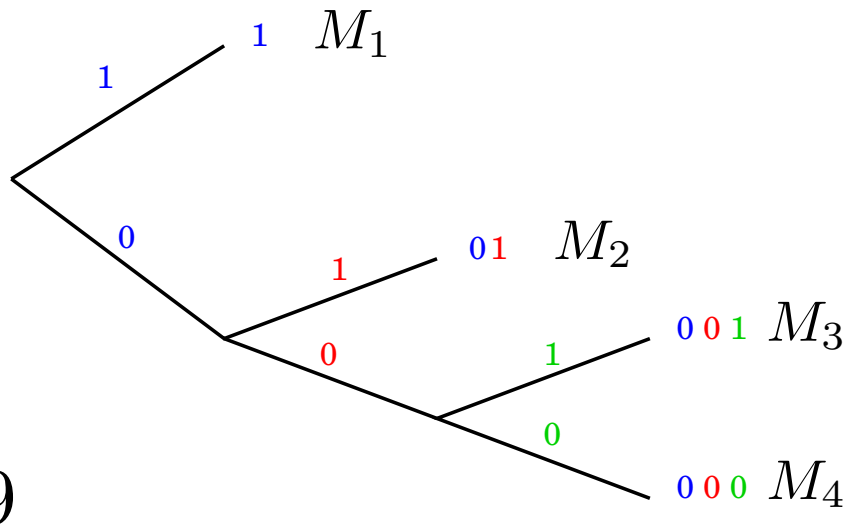
Code A: $\langle L \rangle = 2$

Code B: Dépend des p_n

Compression de données

Code B:

$$\langle L \rangle = \sum_n p_n L(M_n)$$



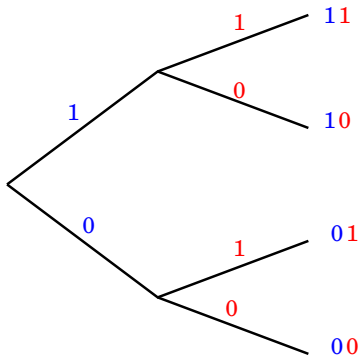
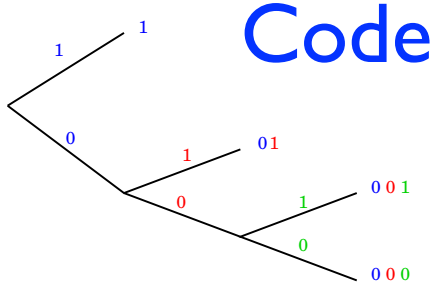
$$p_1 = p_2 = p_3 = p_4 = 1/4$$

$$\langle L \rangle = \frac{1}{4}(1+2+3+3) = \frac{9}{4}$$

$$p_1 = 1/2,\ p_2 = 1/4,\ p_3 = p_4 = 1/8$$

$$\langle L \rangle = 1 \times \frac{1}{2} + 2 \times \frac{1}{4} + 3 \times \frac{1}{8} + 3 \times \frac{1}{8} = \frac{7}{4}$$

Compression de données

	H	Code A: 	Code B: 
 Normal	2	$\langle L \rangle = 2$	$\langle L \rangle = \frac{9}{4}$
 Pipé	$\frac{7}{4}$	$\langle L \rangle = 2$	$\langle L \rangle = \frac{7}{4}$

$$\langle L \rangle_{\text{mini}} = H$$

Compression de données

$$\langle L \rangle_{\text{mini}} = H$$

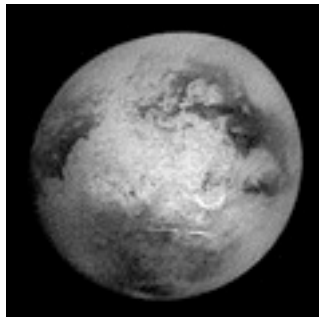
Théorème de Shannon: La longueur moyenne minimale du message, par mot envoyé, est égale à l'entropie H de la loi de probabilité $\{p_n\}$ donnant la fréquence des mots

Plus précisément: paquets de k mots successifs

$H(\text{paquet}) = kH$. Th: $\langle L \rangle_{\text{mini}} \in [kH, kH+1]$

$$\frac{\langle L \rangle_{\text{mini}}}{k} \to H \qquad \text{quand} \qquad k \to \infty$$

Transmission d'information: correction d'erreurs








Transmission d'information: autres exemples

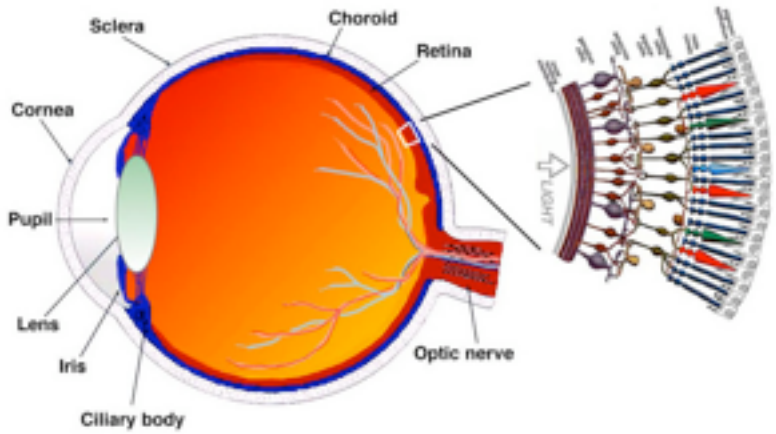
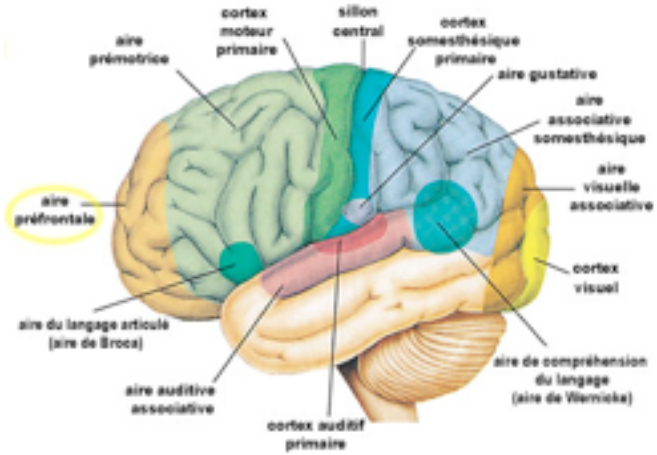
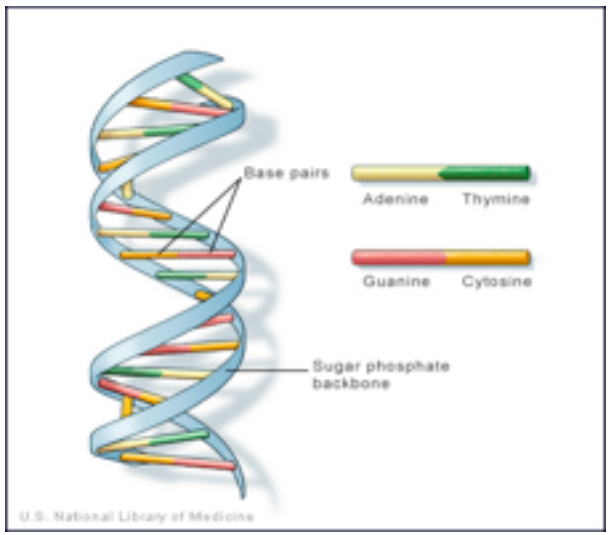
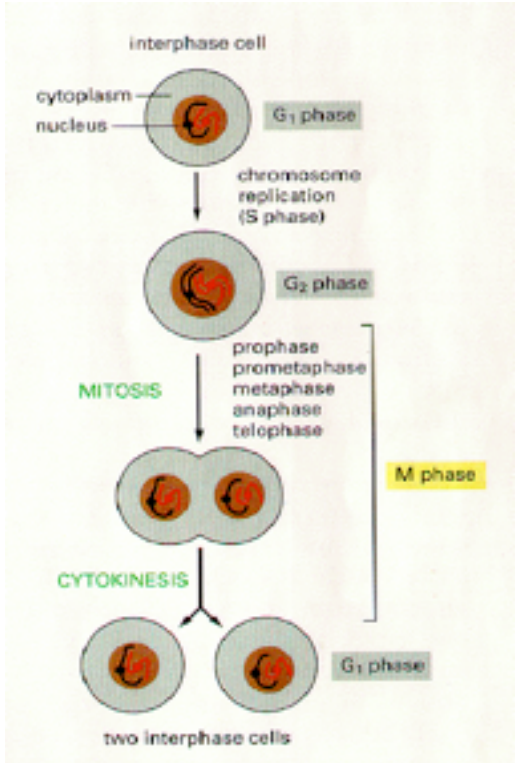


Fig. 1.1. A drawing of a section through the human eye with a schematic enlargement of the retina.

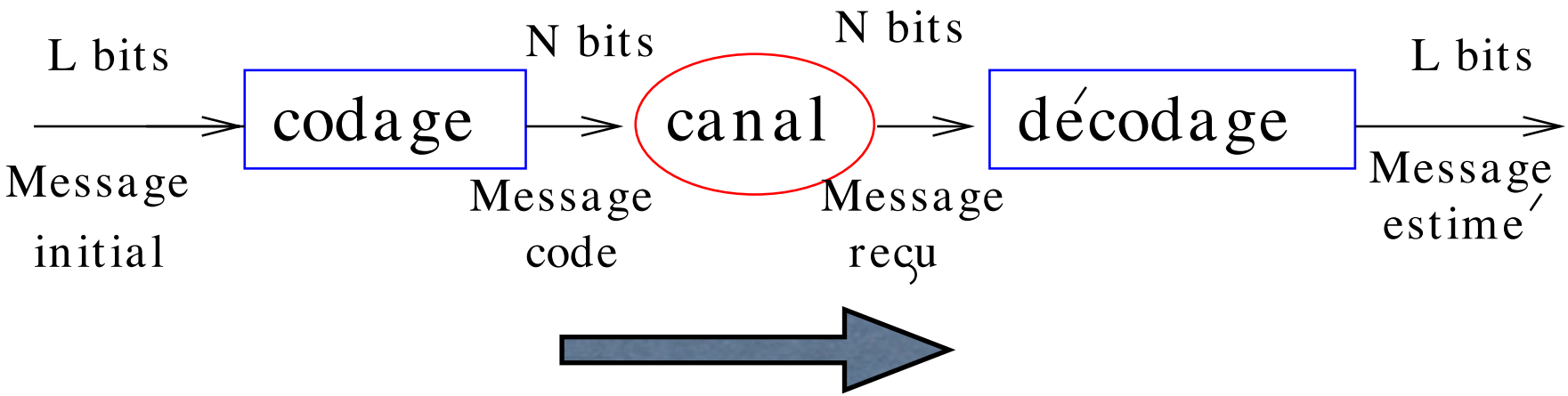






Transmission d'information: correction d'erreurs

Principe: codage, transmission, décodage



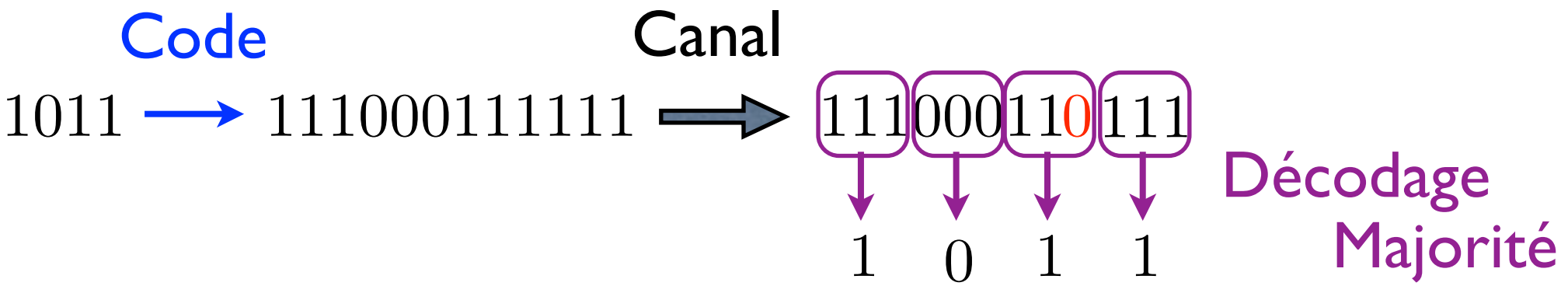
Codage: introduire de la redondance $N > L$

Transmission d'information: correction d'erreurs

Codage: introduire de la redondance

Exemple le plus simple: code par répétition

$0 \to 000 \qquad 1 \to 111$



Canal "binaire symétrique": retourne chaque bit avec proba p

Probabilité d'erreur: $p^3 + 3p^2(1-p)$

Transmission d'information: correction d'erreurs

Code par répétition $0 \to \underbrace{00\ldots0}_{2k+1}$ $1 \to \underbrace{11\ldots1}_{2k+1}$

Probabilité d'erreur à p petit: $\sim \binom{2k+1}{k+1} p^{k+1}$

Taux de transmission $=$ $\dfrac{\text{bits initiaux}}{\text{bits envoyés}} = \dfrac{1}{2k+1}$

Compromis redondance-performance: envoyer $2k+1$ fois plus de bits pour avoir une probabilité d'erreur $\sim p^{k+1}$

Transmission d'information: correction d'erreurs

Code par répétition, taux $\frac{1}{2k+1}$, erreur $\sim \binom{2k+1}{k+1} p^{k+1}$

Mais on peut faire beaucoup mieux. Exemple: Taux $\frac{1}{3}$

★Il existe des codes avec erreur $= 0$ pour $p < 0.17$

★Il existe des codes utilisables avec erreur $= 0$ pour $p < 0.12$

Transition de phase! phase $p < p_d$:décodage sans aucune erreur. Phase $p > p_d$: erreurs.

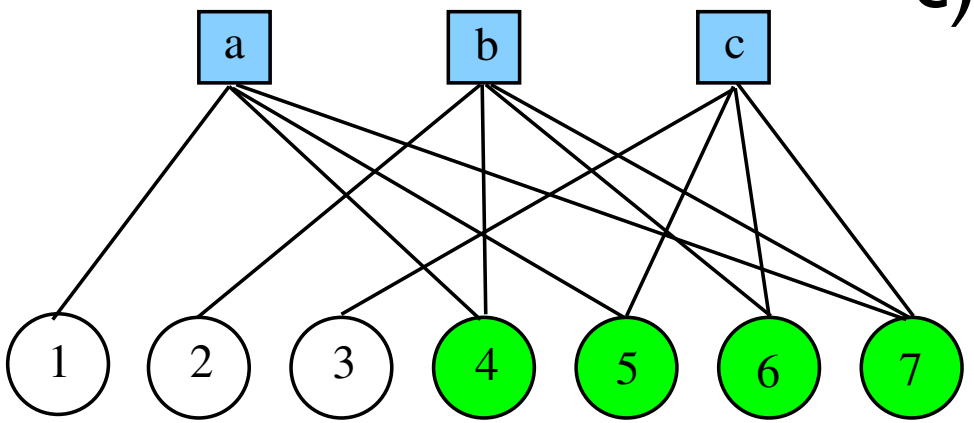
Codes à vérificateurs de parité

Construction du dictionnaire: les mots de codes sont les séquences satisfaisant certaines équation de parité. Ex. $N = 7$:

a) $x_1 + x_4 + x_5 + x_7 = 0 \pmod 2$

b) $x_2 + x_4 + x_6 + x_7 = 0 \pmod 2$

c) $x_3 + x_5 + x_6 + x_7 = 0 \pmod 2$



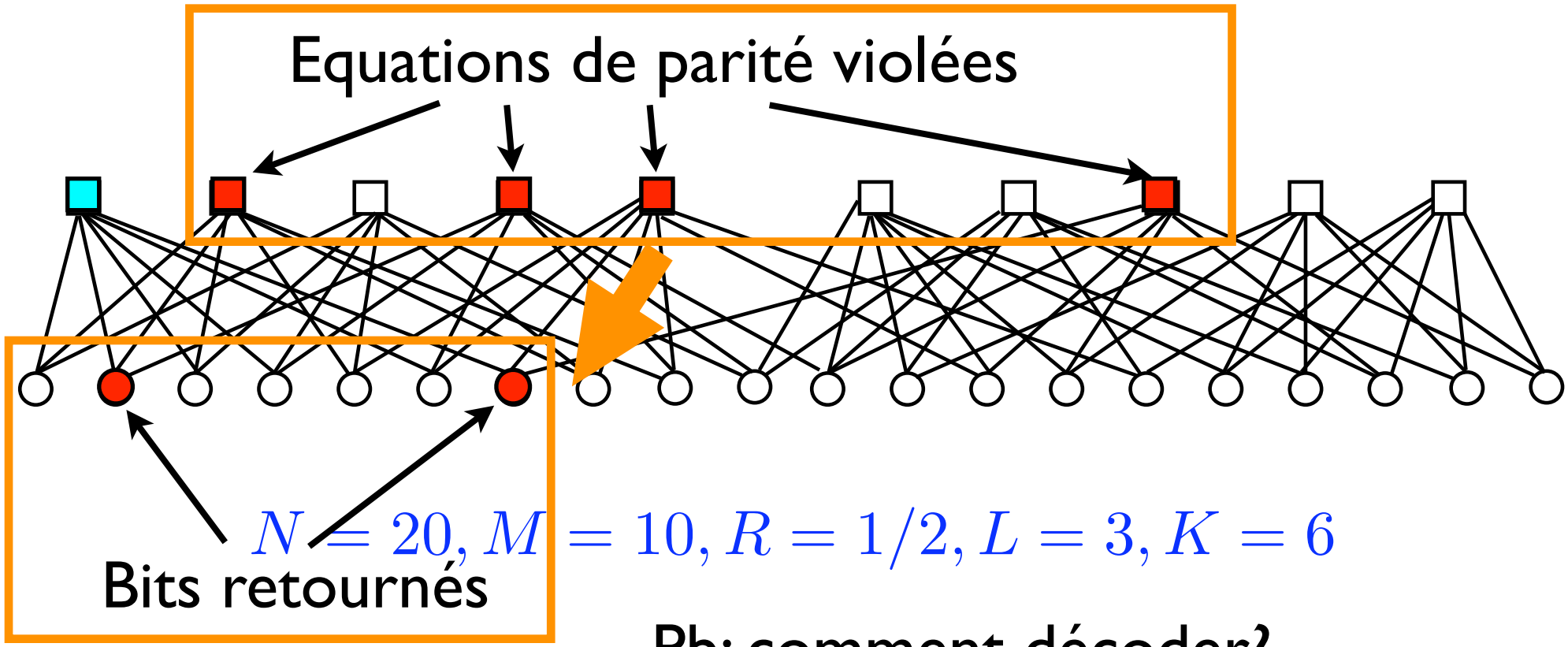
2^4 mots de code, parmi les

2^7 configurations possibles

Codes à vérificateurs de parité

En pratique: $N \gg 1$ variables, $M = (1-R)N$ équations

Ex. Construction aléatoire où K variables par équation
L équations par variable



$N = 20, M = 10, R = 1/2, L = 3, K = 6$

Pb: comment décoder?

Avec une méthode de champ moyen!

Codes à vérificateurs de parité

Décodage par champ moyen BP des codes à vérificateurs de parité: une des méthodes les plus performantes

Transition de phase: pour un niveau de bruit du canal $p < p_d$

on retrouve le signal envoyé sans aucune erreur

Entropie d'un langage (Shannon 1951)

Anglais + espace: 27 symboles. Entropie d'une chaîne de N symboles:

$$I = -\sum_{x_1,...,x_N} P(x_1, ..., x_N) \log_2 P(x_1, ..., x_N).$$

Symboles indépendants:

$P(x_1, ..., x_N) = p(x_1)...p(x_N)$ et $I/N = -\sum_x p(x) \log_2 p(x)$.

Anglais, $p(x) =$ fréquence observée, $I/N = 4.03$ bits/lettre.

$\rightarrow$ OCRO HLI RGWR NMIELWIS EU LL NBNESE...

Symboles corrélés

Approx. d'ordre 2: $P(x_1, ..., x_N) = p(x_1)p(x_2|x_1)...p(x_N|x_{N-1}$ donnée par $p(x)$ et $p(x|y) = p(x, y)/p(y)$.

Alors: $lim_{N\to\infty} I/N = -\sum_{x,y} p(x, y) \log_2 p(x|y)$.

$\rightarrow$ ON IE ANTSOUTINYS ARE T INCTORE ST BE...

NB1: Estimation 'par le jeu': A partir d'un texte, deviner la prochaine lettre, mesurer la fréquence de réponses justes: $I \sim 1.4$ bits/lettre.

NB2: Ecriture aléatoire en utilisant les mots du dictionnaire avec les probabilités de transitions mot-mot exactes:

$\rightarrow$ THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITER THAT THE CHARACTER OF THIS POINT IS THEREFORE ANOTHER METHOD FOR THE LETTERS THAT THE TIME OF WHO EVER TOLD THE PROBLEM FOR AN UNEXPECTED...