

Cours 2 — 18 janvier

Enseignant : Philippe Grangier – Frédéric Magniez

Rédacteur : Charrel - Toubhans

2.1 Notion de théorie de l'information

2.1.1 Introduction informelle

Ce paragraphe n'a pas vocation à être rigoureux mais à présenter brièvement les idées de la théorie classique de l'information. Le problème que l'on se pose est le suivant : on suppose que l'on dispose d'une source émettant un 0 avec une probabilité p et un 1 avec une probabilité $1-p$. On veut compresser au mieux un message de n lettres émis par cette source.

Compresser, cela veut dire trouver un code qui à chaque message de longueur n associe un message et tel que la longueur moyenne des messages ainsi codés soit la plus petite possible.

On représente le code par une fonction :

$$C : \{0, 1\}^n \longrightarrow \{0, 1\}^* = \bigcup_{m>0} \{0, 1\}^m$$

et, si $P(x)$ est la probabilité d'obtenir le message x , la longueur moyenne des messages compressés est définie par la formule :

$$L = \sum_{x \in \{0, 1\}^n} P(x) |C(x)|$$

Comme $P(0) = p$ et $P(1) = 1-p$, on a $P(x) = C_n^m p^m (1-p)^{n-m}$ si x est un message de longueur n contenant m zéros et $(n-m)$ uns.

L'idée que l'on va utiliser ici est le fait qu'un message de longueur n émis par la source a de grande chance d'avoir np zéros et $n(1-p)$ uns. Le nombre de "messages typiques" de cette forme est :

$$C_n^{np} = \frac{n!}{(np)!(n-np)!}$$

En prenant le log et en appliquant la formule de Stirling $\log(n!) = n \log n - n + O(\log n)$, on obtient :

$$\begin{aligned}
 \log(C_n^{np}) &\simeq n \log(n) - n - np \log(np) + np - (n - np) \log(n - np) + n - np \\
 &= n \log(n) - np \log(np) - (n - np) \log(n - np) \\
 &= -n (p \log(p) - (1 - p) \log(1 - p)) \\
 &= nH(p)
 \end{aligned}$$

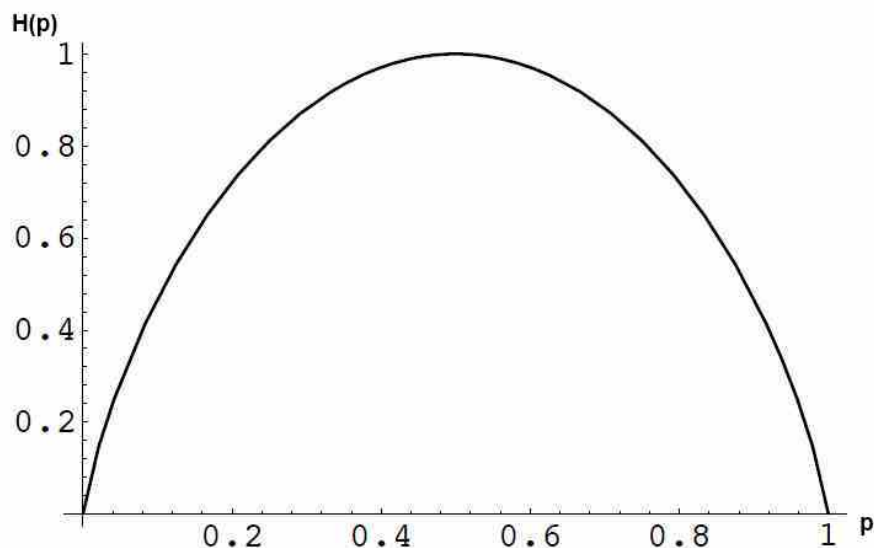
où l'on définit l'entropie (binaire) de Shannon $H(p) = -p \log(p) - (1 - p) \log(1 - p)$.

Ainsi, pour de grandes valeurs de n , on a environ $2^{nH(p)}$ messages typiques. On peut coder chacun d'entre eux par un mot de longueur $nH(p)$ et supposer que les messages "atypiques" n'apparaîtront presque jamais. Dans ce cas la longueur moyenne des mots codés est de $nH(p)$ bits et le taux de compression (taille des mots à l'entrée divisée par la taille moyenne des mots à la sortie) vaut $H(p)$.

L'entropie de Shannon introduite ci-dessus est le point central de la théorie de l'information. On peut l'interpréter comme une quantification du désordre :

- une entropie nulle correspond à aucun désordre, c'est-à-dire la certitude du résultat. Par exemple dans le cas de la source binaire ci-dessus, si $p = 1$ alors l'entropie est nulle et le seul message de longueur n émis par la source est $00\dots 0$.
- a contrario, une entropie maximale correspond au désordre maximal, c'est-à-dire à l'équiprobabilité de tous les cas. Dans le cas de la source binaire, l'entropie est maximale pour $p = 1/2$ (voir graphique ci-dessous) qui correspond au cas où tous les messages sont équiprobables.

Tracé de l'entropie binaire $H(p) = -p \log_2(p) - (1 - p) \log_2(1 - p)$:



2.1.2 Théorie de l'information classique

On va maintenant étendre la notion d'entropie vue dans l'introduction à un alphabet A quelconque (autre que $\{0, 1\}$).

Définition 2.1. On appelle *Entropie de Shannon* de la variable aléatoire X la quantité $H(X)$ définie comme suit :

$$H(X) = - \sum_{x \in A} p_X(x) \log_2(p_X(x))$$

Lemme 2.2. L'entropie de Shannon vérifie les inégalités suivantes :

$$0 \leq H(X) \leq \log_2(|A|)$$

où $|A|$ désigne le nombre de lettres de l'alphabet. De plus,

- le cas $H(X) = 0$ se produit ssi toute l'information se concentre sur une lettre (i.e. il existe une lettre a telle que $p(a) = 1$).
- le cas $H(X) = \log(|A|)$ se produit ssi la distribution p est uniforme.

Preuve: Pour tout x , $0 \leq p(x) \leq 1$ d'où $-p(x) \log(p(x)) \geq 0$ et donc $H(X) \geq 0$.

Si $H(X) = 0$, comme l'entropie est une somme de termes positifs, tous les termes doivent être nuls. On déduit donc pour tout x , $p(x) \log p(x) = 0$ d'où $x = 0$ ou 1 . On conclut facilement qu'il existe une lettre a telle que $p(a) = 1$ et $p(x) = 0$ pour $x \neq a$.

On montre la deuxième inégalité par la concavité de la fonction $f : x \mapsto -x \log(x)$. f est concave car $f''(x) = -1/x < 0$.

Comme f est concave, on a pour tout $p_1, \dots, p_n, a_1, \dots, a_n$ tels que $\sum_i p_i = 1$:

$$\sum_i p_i f(a_i) \leq f\left(\sum_i p_i a_i\right)$$

On a égalité ssi tout les a_i sont égaux entre eux.

Si on applique cela avec les $(p(x))_x$ à la place des a_i et tous les p_i égaux à $1/|A|$, on trouve :

$$\sum_x 1/|A| f(p(x)) = H(X)/|A| \leq f\left(\sum_x 1/|A| p(x)\right) = \log(|A|)/|A|$$

On trouve donc l'inégalité voulue avec égalité ssi tous les $p(x)$ sont égaux entre eux, c'est-à-dire ssi p est une distribution uniforme. \square

Définition 2.3. On appelle *code* une application :


$$C : A^n \mapsto \{0, 1\}^*$$

où A désigne un alphabet. C'est une application qui transforme un mot de longueur n sur l'alphabet en une séquence de bits.

Théorème 2.4. *Premier théorème de Shannon.*

Soit une source qui produit une lettre x avec une probabilité $p(x)$. Alors :

- pour tout code injectif (i.e. deux mots différents se codent vers deux séquences différentes) des mots de longueur n , la longueur moyenne L des mots codés vérifie $L \geq nH(X)$.
- pour tout $\delta > 0$ aussi petit que l'on souhaite, on peut trouver un codage injectif des mots de longueur n tels que la longueur moyenne L des mots codés vérifie $nH(X) \leq L \leq n(H(X) + \delta)$.

 Le premier théorème de Shannon nous dit donc que l'on peut coder un mot de longueur n d'une source d'entropie $H(X)$ en quasiment $nH(X)$ bits. On dit aussi qu'une lettre "transporte" en moyenne $H(X)$ bits d'information.

On peut regarder ce que donnerait le premier théorème de Shannon appliqué à l'alphabet français :

- si les 26 lettres de l'alphabet étaient parfaitement aléatoires, chaque lettre représenterait $\log_2(26) = 4.7$ bits ;
- en fait les probabilités sont inégalement réparties (par exemple le "e" est plus probable que le "w") et fortement corrélées (à la suite de la séquence "ea" le "u" est très probable), si bien qu'une lettre transporte seulement 1.1 bit environ, (pour plus d'information sur le sujet, voir <http://math.ucsd.edu/crypto/java/entropy/>) ;
- il suffit donc en théorie de $n(1.1 + \delta)$ bits pour encoder un message de n lettres sans erreurs asymptotiquement,
- par contre le premier théorème de Shannon nous dit aussi qu'avec $n(1.1 - \delta)$ bits il y aura forcément des erreurs (code non injectif).

On va maintenant définir les notions d'entropie d'une variable aléatoire relative à une autre variable aléatoire et d'information mutuelle.

On travaille sur deux alphabets A et B . On note $p(x, y)$ la probabilité d'avoir la lettre $x \in A$ et la lettre $y \in B$. On note $p(x) = \sum_y p(x, y)$ la probabilité d'avoir la lettre $x \in A$ et symétriquement $p(y) = \sum_x p(x, y)$ la probabilité d'avoir la lettre $y \in B$.

On a les probabilités conditionnelles classiques $p(x|y) = p(x, y)/p(y)$ et $p(y|x) = p(x, y)/p(x)$. Remarquons que l'on peut passer de l'une à l'autre grâce à la formule de Bayes : $p(x|y) = p(x, y) / \sum_x p(y|x)p(x)$.

Définition 2.5. On définit naturellement l'entropie de X et de Y comme l'entropie de la variable aléatoire produit (X, Y) :

$$H(X, Y) = - \sum_{x, y} p(x, y) \log(p(x, y))$$

Définition 2.6. On définit l'entropie conditionnelle partielle $H(X|Y = y)$ par :

$$H(X|Y = y) = - \sum_x p(x|y) \log(p(x|y))$$

Puis on définit l'entropie conditionnelle comme l'espérance de $H(X|Y = y)$:

$$H(X|Y) = \sum_y p(y)H(X|Y = y) = - \sum_{x,y} p(x,y) \log(p(x|y))$$

On peut également définir $H(X|Y)$ par :

$$H(X|Y) = H(X, Y) - H(Y)$$

Définition 2.7. On peut maintenant définir l'information mutuelle de X et de Y :

$$I(X; Y) = H(X) - H(X|Y)$$

Lemme 2.8. L'information mutuelle de X et de Y vérifie :


- (i) $I(X; Y) = H(Y) - H(Y|X)$
- (ii) $I(X; Y) = H(X) + H(Y) - H(X, Y)$
- (iii) $I(X; Y) = I(Y; X)$
- (iv) $I(X; Y) = \sum_{x,y} p(x,y) \log\left(\frac{p(x,y)}{p(x)p(y)}\right)$
- (v) $I(X; Y) \geq 0$
- (vi) $I(X; Y) = 0$ ssi X et Y sont indépendantes.

Preuve: Les points (i) à (iv) s'obtiennent facilement grâce à l'expression de $I(X; Y)$ dans laquelle on remplace $H(X)$ et $H(X|Y)$ par leur expression.

Le point (v) s'obtient par convexité de la fonction $x \mapsto -\log(x)$:

$$\begin{aligned} I(X; Y) &= \sum_{x,y} p(x,y) \log\left(\frac{p(x,y)}{p(x)p(y)}\right) \\ &= \sum_{x,y} p(x,y) \left(-\log\left(\frac{p(x)p(y)}{p(x,y)}\right)\right) \\ &\geq -\log\left(\sum_{x,y} p(x,y) \frac{p(x)p(y)}{p(x,y)}\right) = 0 \end{aligned}$$

Pour le point (vi), il suffit de remarquer que l'inégalité de convexité ci-dessus est une égalité ssi tous les $p(x)p(y)/p(x,y)$ sont égaux entre eux, c'est-à-dire qu'il existe un λ indépendant de x et y tel que $p(x)p(y) = \lambda p(x,y)$. Pour des questions de normalisation $\lambda = 1$, donc pour tout x et y , $p(x)p(y) = p(x,y)$, i.e. X et Y indépendant. □

 Le point (v) du lemme 2.8 peut se voir comme l'inégalité $H(X|Y) \leq H(X)$. Cela s'interprète simplement en terme d'entropie : la variable aléatoire X présente nécessairement un désordre plus grand que la v.a. $X|Y$ car on rajoute de l'information. L'entropie de X est donc forcément plus grande que celle de $X|Y$.

Définition 2.9. On appelle canal bruité, un alphabet d'entrée A , un alphabet de sortie B et des probabilités de transition $(p(y|x))_{x,y}$ qui désignent la probabilité d'avoir la lettre $y \in B$ en sortie du canal sachant que l'on avait la lettre $x \in A$ en entrée.

Remarques :

- on peut inverser l'entrée et la sortie du canal en calculant les $p(x|y)$ à partir de $p(y|x)$ grâce à la formule de Bayes.
- si l'on connaît la distribution $p(x)$, on peut calculer la distribution à la sortie du canal grâce à la formule : $p(y) = \sum_x p(y|x)p(x)$.

Définition 2.10. On a vu que pour un canal donné, si l'on connaît la distribution de probabilité d'entrée $X = (x, p(x))_x$, on connaît automatiquement la distribution Y de sortie. La quantité suivante est donc bien définie et s'appelle la capacité du canal :

$$C = \text{Max}_{(p(x))_x} I(X; Y)$$

Théorème 2.11. Deuxième théorème de Shannon.

Le nombre maximal de bits par symbole que l'on peut transmettre sans erreur asymptotiquement (i.e. tel que la probabilité d'erreur tende vers 0 lorsque la longueur du message tend vers l'infini) dans un canal bruité est sa capacité C .

2.1.3 Entropie de Von Neumann

En information quantique, les lettres de l'alphabet sont remplacées par des états quantiques que l'on représente par leurs matrices densités $(\rho_x)_x$. A chaque état, on peut associer une probabilité p_x . L'état vu par l'observateur est donc :

$$\rho = \sum_x p_x \rho_x$$

ρ est hermitienne, on peut donc la diagonaliser dans une base orthonormée $\{|a\rangle\}_a$:

$$\rho = \sum_a p_a |a\rangle \langle a|$$

$$(\text{avec } \forall a, b, \langle a|b\rangle = \delta_{a,b})$$

Comme ρ est définie positive ou nulle et de norme 1, on a pour tout a , $0 \leq p_a \leq 1$. On peut donc parler de l'entropie de Shannon de la distribution $(p_a)_a$, $H(A) = -\sum_a p_a \log(p_a)$. On aimerait maintenant définir l'entropie de Von Neumann $S(\rho)$ comme étant égale à la quantité $H(A)$, mais cette quantité dépend a priori de la base dans laquelle ρ a été diagonalisée.

Lemme 2.12. *La quantité $H(A)$ définie ci-dessus ne dépend pas de la base dans laquelle ρ a été diagonalisée.*

Preuve: Supposons ρ diagonalisée dans 2 bases orthonormées $\{|a\rangle\}$ et $\{|b\rangle\}$.

$$\rho = \sum_a p_a |a\rangle\langle a| = \sum_b p_b |b\rangle\langle b|$$

Montrons l'égalité des ensembles $\{p_a\}$ et $\{p_b\}$.

En utilisant $\rho|a\rangle = p_a|a\rangle$ et la décomposition de $|a\rangle$ sur la base $\{|b\rangle\}$, $|a\rangle = \sum_{b'} \mu_{b'}|b'\rangle$, on obtient :

$$\begin{aligned} \rho|a\rangle &= \left(\sum_b p_b |b\rangle\langle b|\right) \left(\sum_{b'} \mu_{b'} |b'\rangle\right) \\ &= \sum_{b,b'} p_b \mu_{b'} |b\rangle\langle b|b'\rangle \\ &= \sum_b p_b \mu_b |b\rangle \\ &= p_a |a\rangle \\ &= \sum_b p_a \mu_b |b\rangle \end{aligned}$$

D'où :

$$\forall b, \mu_b(p_a - p_b) = 0$$

Comme $|a\rangle$ est non nul, il existe un μ_b non nul et donc un p_b tel que $p_a = p_b$.

On a donc montré que tout p_a est égal à un certain p_b , c'est-à-dire $\{p_a\} \subseteq \{p_b\}$. L'égalité s'obtient en inversant le rôle de p_a et p_b .

Comme $\forall a, b, 0 \leq p_a, p_b \leq 1$, on a $\forall a, b, -p_a \log(p_a), -p_b \log(p_b) \geq 0$. On somme donc dans $H(A)$ et $H(B)$, les mêmes termes positifs ou nuls (mais pas forcément dans le même ordre). On peut conclure de cela que $H(A) = H(B)$. \square

Définition 2.13. *On appelle entropie de Von Neumann de la matrice densité ρ , et on note $S(\rho)$ la quantité :*

$$S(\rho) = - \sum_a p_a \log(p_a) = H(A)$$

où les $\{p_a\}$ sont tels que ρ s'écrive $\sum_a p_a |a\rangle\langle a|$ dans une certaine base orthonormée $\{|a\rangle\}$.

Lemme 2.14. *L'entropie de Von Neumann vérifie les propriétés suivantes :*

- (i) *l'entropie d'un état pur $\rho = |\psi\rangle\langle\psi|$ est nulle.*
- (ii) *si ρ a D valeurs propres non nulles, alors $S(\rho) \leq \log_2(D)$*
- (iii) *pour $p_i \geq 0$ et $\sum_i p_i = 1$, $S(\sum_i p_i \rho_i) \geq \sum_i p_i S(\rho_i)$
(l'entropie augmente avec l'ignorance de la façon dont l'état est préparé)*
- (iv) *Préparation : pour un état $\rho = \sum_x p_x |\phi_x\rangle\langle\phi_x|$, on a
 $H(X) = -\sum_x p_x \log(p_x) \geq S(\rho)$ (égalité si les $|\phi_x\rangle$ sont orthogonaux)*
- (v) *pour un système bipartite AB , on a*

$$|S(\rho_A) - S(\rho_B)| \leq S(\rho_{AB}) \leq S(\rho_A) + S(\rho_B)$$

- (vi) *Subadditivité forte : pour un système tripartite ABC , on a*

$$S(\rho_{ABC}) + S(\rho_B) \leq S(\rho_{AB}) + S(\rho_{BC})$$

⚡ Attention à la différence entre l'entropie de Shannon et l'entropie de Von Neumann.

⊥ Par exemple, la propriété (v) du lemme précédent n'est pas vérifiée pour l'entropie classique.

On a pour l'entropie de Shannon :

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

alors que l'on a seulement pour l'entropie de Von Neumann :

$$|S(\rho_A) - S(\rho_B)| \leq S(\rho_{AB})$$

En théorie de l'information classique, tout savoir sur l'état de (X, Y) (i.e. $H(X, Y) = 0$) implique tout savoir sur X et Y séparément (i.e. $H(X) = H(Y) = 0$).

En revanche, en théorie de l'information quantique, avoir $S(\rho_{AB}) = 0$ implique seulement $S(\rho_A) = S(\rho_B)$. On peut avoir deux particules intriquées dans un état parfaitement déterminé (i.e. $S(\rho_{AB}) = 0$) et ne pas connaître l'état de chaque particule séparément (i.e. $S(\rho_A) = S(\rho_B) \neq 0$).

2.1.4 Compression de données quantiques

On considère une matrice densité $\rho = \sum_x p_x |\phi_x\rangle\langle\phi_x|$ et on définit $X = \{|\phi_x\rangle, p_x\}$. Un message de n symboles est alors associé à la matrice densité $\rho^n = \rho \otimes \dots \otimes \rho$. On voudrait, comme dans le cas classique, compresser ce message en un nombre minimal de qubits.

Le *théorème de Schumacher* qui est l'analogue du *théorème de Shannon* pour la théorie quantique nous donne la réponse.

Théorème 2.15. *Théorème de Schumacher*

Pour tout message de n symboles tirés d'un état dont la matrice densité est ρ , on peut trouver un code injectif tel que la longueur moyenne des messages codés sera de $nS(\rho)$ qubits.

Rappelons que la dimension de l'espace de Hilbert \mathcal{E} de N qubits est 2^N . Le principe de la démonstration est l'utilisation du premier théorème de Shannon :

- on se place dans une base orthonormée qui diagonalise ρ qui s'écrit alors $\rho = \sum_a p_a |a\rangle\langle a|$,
- on applique alors le théorème de Shannon à l' "alphabet probabilisé" $\{|a\rangle\langle a|, p_a\}$ qui se comporte comme un alphabet classique,
- on montre ensuite que pour de grands n , ρ^n a son support dans un sous espace de dimension $2^{nS(\rho)}$ de l'espace global,
- on finit la démonstration en utilisant les mêmes arguments que ceux de l'introduction formelle de ce chapitre, à savoir on code les messages les plus probables sur $nS(\rho)$ qubits sans se soucier des autres messages.

Quelques remarques :

- en cryptographie, si Alice code une information classique sous forme de qubits et l'envoie à Bob, celui-ci ne peut en général pas remonter simplement à cette information classique.
- dans le cas le plus général, les symboles eux-mêmes ne sont pas des états purs. C'est dans ces cas que Bob ne peut pas remonter à l'information classique envoyée par Alice. On a alors besoin d'évaluer la quantité maximale d'information classique extractible du message quantique. Le *théorème de Holevo* répond à cette question.

Définition 2.16. *Information de Holevo*

On appelle information de Holevo d'un "alphabet quantique" $\{\rho_x, p_x\}$ la quantité :

$$\chi(\{\rho_x, p_x\}) = S\left(\sum_x p_x \rho_x\right) - \sum_x p_x S(\rho_x)$$

Remarque :

L'information de Holevo est positive ou nulle en vertu de la concavité de l'entropie de Von Neumann énoncé (iii) du *lemme 2.14*.

Théorème 2.17. *Théorème de Holevo*

Le maximum de l'information classique accessible sur toutes les mesures est borné par $\chi(\{\rho_x, p_x\})$:

$$\text{Max}_{Y \text{ mesuré}} I(X; Y) \leq \chi(\{\rho_x, p_x\})$$

2.2 La cryptographie quantique

2.2.1 Rappels sur la distribution classique de clés

Le cryptage classique

Le système de cryptage par clé publique classique le plus utilisé et reconnu est le cryptage RSA (Rivest, Shamir and Adleman, 1978). Le principe repose sur l'incapacité des machines classiques à trouver les diviseurs de grands nombres.

Supposons qu'Alice veuille envoyer des informations à Bob :

- Bob commence par générer une clé publique ainsi que quelques intermédiaires de calcul :
 - Il génère 2 grand nombres premiers a et b .
 - $p = ab$ (Clé publique)
 - $q = (a - 1)(b - 1)$
 - s vérifiant $\text{pgcd}(q, s) = 1$
 - r vérifiant $rs \equiv 1[q]$
- Bob envoie la clé publique p ainsi que r à Alice.
- Alice code son message x en $y = x^r[p]$ et envoie ouvertement y .
- La théorie des nombres permet facilement à Bob de déchiffrer le message : $x = y^s[p]$.

L'intérêt de cette méthode est que l'espion éventuel ne connaît pas a, b, s, q , et ne peut donc rien faire avant d'avoir décomposé p en nombres premiers (en pratique impossible si p a plus de 200 chiffres).

La compagnie RSA propose souvent des énigmes dont le but est d'augmenter l'efficacité de leur cryptage. Par exemple, en février 1999, le décryptage d'une clé 512 bits a pris 5 mois.

$p = 0941738641570527421809707322040357612003732945449205990913842131476349984288934784717997257891267332497625752899781833797076537244027146743531593354333897$

La factorisation de ce nombre s'est faite en plusieurs parties :

- Préparation : 9 semaines
- Criblage : 3.5 mois
- Résultat : 3.7 Go de données
- Filtrage : 9.5 jours
- Factorisation : 39 heures

Le résultat (facilement vérifiable) est :

$a = 102639592829741105772054196573991675900716567808038066803341933521790711307779$

$b = 106603488380168454820927220360012878679207958575989291522270608237193062808643$

Inconvénients intrinsèques au cryptage par clé publique

Tout d'abord, de par le caractère statistique du cryptage, il est tout à fait possible de tomber par hasard sur des grands nombres facilement factorisables. Ce problème tente d'être résolu par RSA qui met à jour sa base de données de recommandations pour le choix des nombres premiers a et b .

Ensuite, il n'y a pas de démonstration absolue de la sécurité du cryptage par clé publique. En effet, le seul garant aujourd'hui de la sécurité des clés est la relative lenteur des ordinateurs à calculer les facteurs. (Le temps de calcul est exponentiel en le nombre de chiffres de la clé). Ainsi, si quelqu'un a découvert un ordinateur et/ou un algorithme très puissant, il pourra décrypter plus facilement les clés, et personne ne pourra le savoir.

Enfin, la découverte et la recherche en physique quantique ont amené Peter Shor à découvrir, en 1994, qu'un "ordinateur quantique" serait capable de factoriser une clé en temps polynomial. La comparaison entre l'hypothétique algorithme de Shor et le meilleur algorithme classique (Number Field Sieve) est la suivante :

$$NFS(n) = \exp(1.9 \log(n)^{1/3} \log(\log(n))^{2/3})$$

$$Shor(n) = \log(n)^3$$

Ainsi, il faut $6 * 10^6$ fois plus de temps à NFS pour décrypter un nombre de 1024 chiffres que pour décrypter un nombre de 512 chiffres, alors qu'il ne faut que 8 fois plus de temps pour Shor.

Cryptographie à clé secrète

La cryptographie à clé secrète est sans doute la plus fiable, mais nécessite aux deux protagonistes de s'être échangé la clé de façon totalement sûre avant de communiquer. Dans ce cas, la sécurité est **TOTALEMENT** démontrable dans les conditions suivantes :

- la clé doit être aléatoire
- elle doit être aussi longue que le message
- elle ne doit être utilisée qu'une seule fois (Shannon)

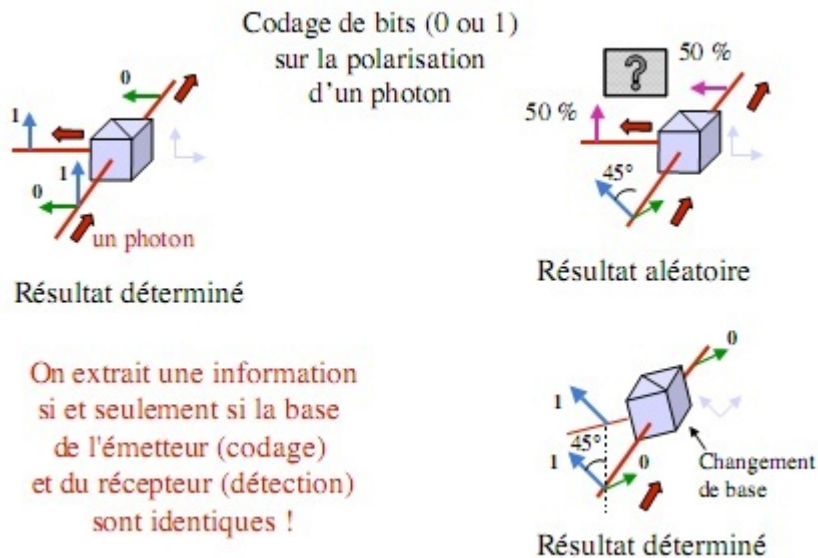
Le problème reste donc de pouvoir disposer d'une clé commune sécurisée. En théorie classique, il n'y a aucun moyen certain de savoir que la clé n'a pas été espionnée. C'est là que rentre en jeu la théorie quantique : on va voir que l'on peut créer une clé qui est sécurisée intrinsèquement.

2.2.2 Le protocole quantique BB84 (Bennett et Brassard, 1984)

Polarisation d'un photon unique

En physique quantique, la mesure de la polarisation d'un photon ne peut donner que 2 valeurs (propres). A l'inverse d'une impulsion lumineuse, la polarisation d'un photon inconnu n'est pas mesurable avec certitude. Le photon est donc le candidat parfait pour porter et sécuriser l'information binaire d'un bit. D'autre part, la mesure de la polarisation du photon se fait nécessairement dans une base définie par l'opérateur effectuant la mesure. Il va sans

dire que la mesure est directement liée à la base utilisée, comme le montre la figure ci-dessous. Si on ne mesure pas dans la bonne base, le résultat est aléatoire :

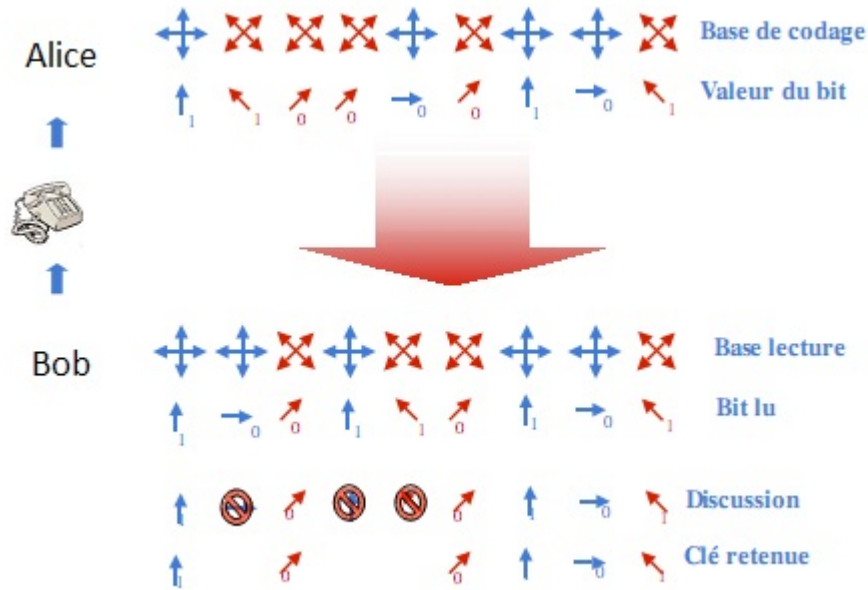


Enfin, une propriété supplémentaire -non des moindres- du photon quantique est que la mesure affecte définitivement l'état par projection. Ainsi, un photon espionné sera aisément repéré par le destinataire.

Création de la clé

Le principe est redoutablement simple : Alice envoie une série de photons polarisés dans une des 2 bases préalablement choisies, le tout de façon aléatoire. Bob mesure les photons sur une des 2 bases, de façon aléatoire également.

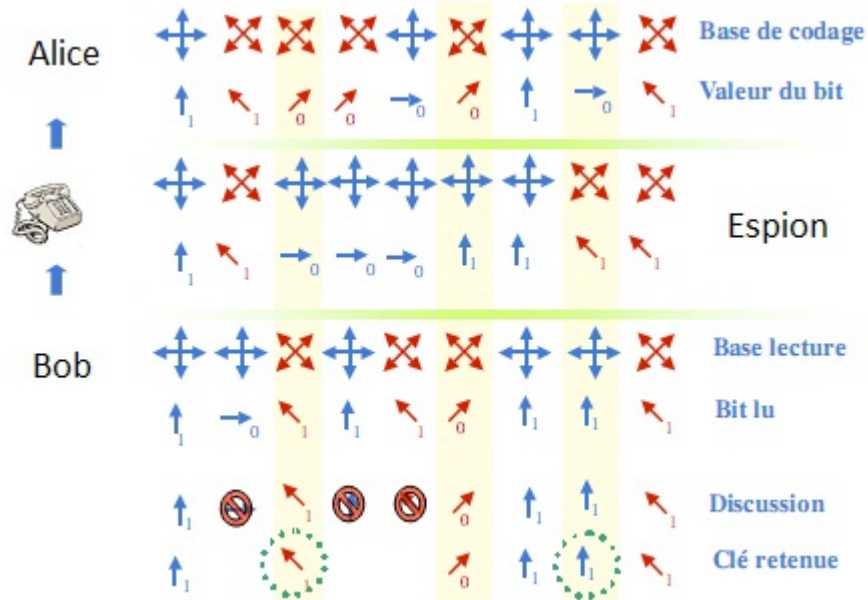
Ensuite, Bob communique à Alice ses choix consécutifs de bases. Alice renvoie alors à Bob la liste de bits à retenir.



Sécurisation de la clé

La clé créée doit donc être intrinsèquement la même du côté de Bob que de celui d'Alice. Le problème maintenant consiste à s'assurer que personne n'a pu espionner la clé pendant sa création.

L'espion peut connaître les 2 bases choisies, mais ne peut jamais deviner celle que Alice utilise pour chaque photon. Ainsi, le seul moyen qu'a l'espion est d'utiliser aléatoirement une des deux bases pour la mesure (on rappelle qu'il ne peut effectuer qu'une seule mesure sur le photon). L'espion va donc se tromper et occasionnellement modifier de la polarisation du photon. A la fin de l'échange, Alice et Bob vont publiquement s'annoncer leurs clés pour pouvoir les comparer. Si l'espion est actif, Bob et Alice trouveront des différences dans leurs mesures, ce qui sera la preuve infaillible de la présence de l'espion :



2.2.3 Principes de la cryptographie quantique

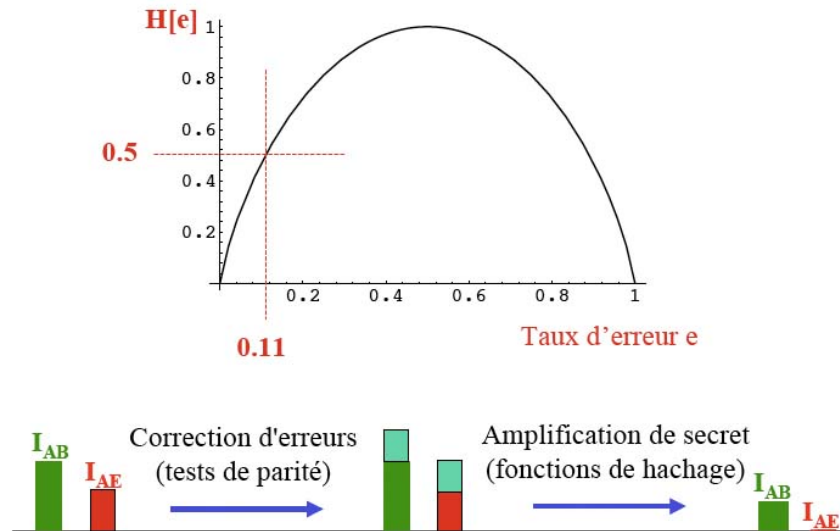
Le point fondamental dans la description du protocole quantique BB84 est que plus Eve acquiert d'information, plus elle crée d'erreurs dans la transmission. On peut ainsi démontrer qu'en mesurant le taux d'erreur (par comparaison publique des résultats de Bob avec les émissions d'Alice), il est possible de borner supérieurement la quantité d'information de l'espion. Alice et Bob peuvent alors utiliser des algorithmes classiques pour corriger les erreurs, et pour produire une clé (plus petite) totalement inconnue de l'espion. La longueur de la clé produite est d'autant plus petite que le taux d'erreur initial est plus grand (maximum tolérable : 11%!).

Entropie de Shannon binaire et sécurité quantique

Une analyse un peu plus poussée des entropies et informations mutuelles en jeu permettent d'extraire un critère inviolable de sécurité de la cryptographie quantique. Ce critère est :

$$H(e) < 1/2, \text{ avec } e \text{ le taux d'erreur.}$$

On peut ainsi réduire le taux d'information mutuelle entre Alice et Bob, en supprimant totalement celle entre Alice et l'espion :



Le théorème de non clonage

La mécanique quantique se révèle ainsi très efficace pour repérer et neutraliser les failles de sécurité. Cependant, le protocole de sécurisation ne serait pas sûr s'il était possible de cloner les particules (L'espion pourrait renvoyer une particule et analyser l'autre, se rendant ainsi entièrement invisible aux yeux d'Alice et Bob). Ce problème est adressé par l'impossibilité fondamentale en mécanique quantique de copier un état arbitraire de polarisation d'un photon parmi un ensemble d'états orthogonaux. La démonstration de ce théorème est proposée ci-dessous :

Théorème 2.18. *Le clonage d'une particule quantique contredit la linéarité de la mécanique quantique*

Preuve: Supposons qu'il existe une machine capable de cloner une particule (en l'occurrence copier un état sur un photon "blanc" générique ψ_0) :

$$|\phi_1\rangle \otimes |\psi_0\rangle \rightarrow |\phi_1\rangle \otimes |\psi_1\rangle$$

$$|\phi_2\rangle \otimes |\psi_0\rangle \rightarrow |\phi_2\rangle \otimes |\psi_2\rangle$$

Les copies sont notées $|\phi_1\rangle$ et $|\phi_2\rangle$ car elles peuvent à priori être effectuées sur des systèmes physiques différents (ex : spin 1/2).

Plaçons maintenant dans cette machine un état $|\phi_3\rangle = (|\phi_1\rangle + |\phi_2\rangle)/\sqrt{2}$.

Par linéarité, le résultat est : $|\phi_3\rangle \otimes |\psi_0\rangle \rightarrow \frac{1}{\sqrt{2}}(|\phi_1\rangle \otimes |\psi_1\rangle + |\phi_2\rangle \otimes |\psi_2\rangle)$

Ce n'est pas un état factorisable sous la forme : $|\phi_3\rangle \otimes |\psi_3\rangle$, d'où l'impossibilité du clonage. \square

Théorème 2.19. *Le clonage d'une particule quantique contredit l'unitarité de la mécanique quantique*

Preuve: Une transformation quantique se doit d'être unitaire et donc de conserver les produit scalaires. Egalisons les produit scalaire des deux états, avant et après clonage :

$$\langle \phi_1 | \phi_2 \rangle \times \langle \psi_0 | \psi_0 \rangle = \langle \phi_1 | \phi_2 \rangle \times \langle \psi_1 | \psi_2 \rangle$$

soit encore (pour $\langle \phi_1 | \phi_2 \rangle \neq 0$)

$$1 = \|\psi_0\|^2 = \langle \psi_1 | \psi_2 \rangle$$

puis

$$\| |\psi_1\rangle - |\psi_2\rangle \|^2 = \|\psi_1\|^2 + \|\psi_2\|^2 - 2\langle \psi_1 | \psi_2 \rangle = 0$$

Ainsi, $|\psi_1\rangle = |\psi_2\rangle$, ce qui est absurde. □

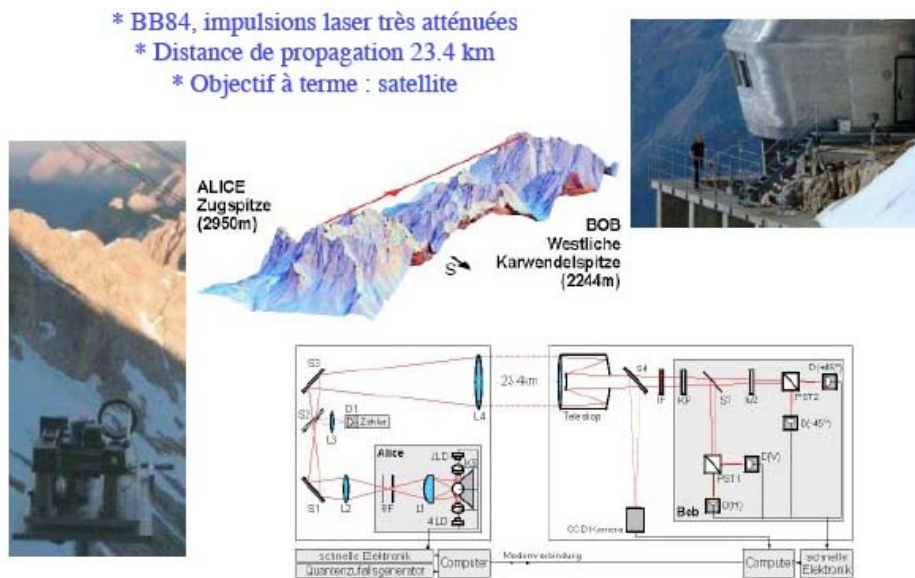
Ceci étant dit, il est possible, en raisonnant sur les matrices densité et les probabilités des états, de réaliser une copie imparfaite d'un état. Le taux d'erreur ne peut pourtant pas descendre en dessous de 16 %, ce qui rend cette copie repérable facilement par Bob !

Au-delà des conséquences en terme de sécurité quantique, le clonage aurait des implications inacceptables pour la théorie :

- Violation des inégalités d'Heisenberg
- Conflit entre Mécanique quantique et Relativité restreinte

2.2.4 Exemples de mise en pratique

Démonstration en espace libre



Potentiel industriel à moyen et long terme

*** Actuellement 2 startups commercialisent des systèmes (fibres optiques, 50 km)**



**IdQuantique
(Genève)**



**MagiQ Technologies
(New York)**

*** Intense activité aux USA (surtout militaire) et au Japon (NEC, Fujitsu...)**

*** « Projet Intégré » Européen SECOQC :** 

« Secure Communication based on Quantum Cryptography ».

... validation des objectifs et des méthodes en cours, conclusion en 2008 !