

# REVUE DE STATISTIQUE APPLIQUÉE

J. OSWALD

## **Essai sur la théorie de l'information (logodynamique statistique)**

*Revue de statistique appliquée*, tome 20, n° 2 (1972), p. 37-70

[http://www.numdam.org/item?id=RSA\\_1972\\_\\_20\\_2\\_37\\_0](http://www.numdam.org/item?id=RSA_1972__20_2_37_0)

© Société française de statistique, 1972, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques  
<http://www.numdam.org/>

# ESSAI SUR LA THÉORIE DE L'INFORMATION (Logodynamique statistique)

J. OSWALD

Directeur des Laboratoires de MARCOUSSIS (C.G.E.)

(2<sup>e</sup> partie)<sup>(1)</sup>

CHAPITRE V

INFORMATION INCERTAINE

CORRECTION ET DETECTION DES ERREURS

Dans tout ce qui précède, on a supposé que l'information reçue était certaine : le message permet d'identifier sans équivoque la configuration du système parmi toutes les configurations possibles. L'énergie diacritique est la mesure de l'information transmise, c'est-à-dire du nombre moyen de configurations réalisées avec N symboles émis successivement, déduit d'un examen portant sur un grand nombre d'épreuves, et parmi lesquelles on a pu faire un choix.

Or le message reçu peut être incorrect parce qu'il a été perturbé pendant la transmission (bruit ou parasite de nature quelconque). Il en résulte que le destinataire n'a qu'une connaissance incertaine de la configuration réalisée, ou ce qui revient au même, du message qui a été émis.

Dans un tel cas, on est amené à répondre à deux questions :

- Quelle quantité d'information "utilisable" trouve-t-on à la réception ?

- Quelle quantité d'information supplémentaire faut-il transmettre pour compenser la perte provoquée par le bruit ?

Avant de faire cette analyse, il convient de préciser les grandeurs qui interviendront dans les formules et qui caractérisent les perturbations apportées au message.

Considérons donc un système (S) susceptible de prendre  $l$  états différents, désignés par les symboles  $s_1, s_2, \dots, s_l$ . Dans une longue séquence comportant N symboles, on émettra  $a_1$  fois l'état  $s_1, a_2$  fois l'état  $s_2, \dots, a_l$  fois l'état  $s_l$ , avec :

$$a_1 + a_2 + \dots + a_l = N \quad (65)$$

Lorsque N devient très grand,  $a_i/N$  tend, en probabilité, vers la probabilité a priori de l'état  $s_i$ , soit  $p_i$ .

Pour simplifier l'exposé, nous supposons que l'effet d'une perturbation est de substituer au symbole émis  $s_i$  un autre symbole  $s_k$  ; plus précisément, le récepteur interprète tout signal reçu comme un des symboles de la liste possible, mais son interprétation est correcte ou incorrecte suivant que le message n'a pas été perturbé ou a été perturbé.

-----  
(1) Pour la première partie de cette étude, voir "Revue de statistique Appliquée, 1972 vol. XX, n°1, pp. 25-65.

Dans ces conditions, le récepteur va reconnaître  $b_1$  fois l'état  $s_1$ ,  $b_2$  fois l'état  $s_2$ , ...,  $b_l$  fois l'état  $s_l$ , avec :

$$b_1 + b_2 + \dots + b_l = N \quad (66)$$

Nous pouvons écrire :

$$\left\{ \begin{array}{l} b_1 = \beta_{11} + \beta_{21} + \dots + \beta_{l1} \\ b_2 = \beta_{12} + \beta_{22} + \dots + \beta_{l2} \\ \dots \\ b_l = \beta_{1l} + \beta_{2l} + \dots + \beta_{ll} \end{array} \right. \quad (67)$$

Le premier indice désigne le symbole qui a été émis et le second celui qui a été reçu :  $\beta_{11}$  est le nombre de fois où le symbole  $s_1$  a été correctement transmis,  $\beta_{21}$  est le nombre de fois où  $s_1$  a été reçu alors que  $s_2$  a été émis, etc...

En supposant  $N$  très grand, donc les  $a_i$  très grands, on peut écrire :

$$\left\{ \begin{array}{l} \lim. \frac{\beta_{11}}{a_1} = \overline{\omega}_{11} \text{ (probabilité de recevoir } s_1 \text{ si } s_1 \text{ a été émis)} \\ \lim. \frac{\beta_{21}}{a_2} = \overline{\omega}_{21} \text{ (probabilité de recevoir } s_1 \text{ si } s_2 \text{ a été émis)} \\ \dots \\ \lim. \frac{\beta_{ik}}{a_i} = \overline{\omega}_{ik} \text{ (probabilité de recevoir } s_k \text{, si } s_i \text{ a été émis)} \end{array} \right. \quad (68)$$

Les  $\overline{\omega}_{ik}$  sont les probabilités conditionnelles de réception des symboles.

Puisque, par hypothèse, l'émission du symbole  $s_i$  se traduit nécessairement par la réception d'un des  $l$  symboles de la liste, on a :

$$\left\{ \begin{array}{l} \overline{\omega}_{11} + \overline{\omega}_{12} + \dots + \overline{\omega}_{1l} = 1 \\ \dots \\ \sum_k \overline{\omega}_{ik} = 1 \end{array} \right. \quad (69)$$

Par application de la formule de BAYES, on détermine les probabilités à posteriori :

$$\left\{ \begin{array}{l} \lim. \frac{\beta_{11}}{b_1} = p_{11} = \frac{p_1 \overline{\omega}_{11}}{p_1 \overline{\omega}_{11} + p_2 \overline{\omega}_{21} + \dots + p_l \overline{\omega}_{l1}} \\ \lim. \frac{\beta_{21}}{b_1} = p_{21} = \frac{p_2 \overline{\omega}_{21}}{p_1 \overline{\omega}_{11} + p_2 \overline{\omega}_{21} + \dots + p_l \overline{\omega}_{l1}} \\ \dots \\ \lim. \frac{\beta_{ik}}{b_k} = p_{ik} = \frac{p_i \overline{\omega}_{ik}}{\sum_i p_i \overline{\omega}_{ik}} \end{array} \right. \quad (70)$$

$p_{11}$  est la probabilité que  $s_1$  ait été émis si  $s_1$  est reçue, et plus généralement  $p_{ik}$  la probabilité que  $s_i$  ait été émis lorsque  $s_k$  est reçu.

Les probabilités de recevoir les divers symboles  $s_1, \dots, s_l$  sont :

$$\left\{ \begin{array}{l} \lim. \frac{b_1}{N} = p_1 \bar{w}_{11} + p_2 \bar{w}_{21} + \dots + p_l \bar{w}_{l1} \\ \lim. \frac{b_2}{N} = p_1 \bar{w}_{12} + p_2 \bar{w}_{22} + \dots + p_l \bar{w}_{l2} \\ \dots \\ \lim. \frac{b_l}{N} = \sum_i p_i \bar{w}_{i_l} \end{array} \right. \quad (71)$$

De (68) on déduit :

$$\left\{ \begin{array}{l} \lim. \frac{\beta_{11}}{N} = p_1 \bar{w}_{11} \\ \lim. \frac{\beta_{21}}{N} = p_2 \bar{w}_{21} \\ \dots \\ \lim. \frac{\beta_{ik}}{N} = p_i \bar{w}_{ik} \end{array} \right. \quad (72)$$

Ce sont les probabilités qu'un symbole  $s_i$  ait été émis et qu'un symbole  $s_k$  soit reçu.

Ces formules seront utilisées dans les développements qui suivent.

Dans l'évaluation de l'information "utilisable" (en dépit de la présence d'erreurs) deux attitudes sont possibles. On sait que l'émetteur envoie un certain nombre de symboles  $s_i$  de probabilité  $p_i$  ; donc l'énergie diacritique émise est :

$$NH_0 = N \sum_i p_i \log \frac{1}{p_i} \quad (73)$$

D'autre part, l'énergie diacritique apparente reçue est celle d'un système fictif qui prendrait  $b_1$  fois l'état  $s_1$ ,  $b_2$  fois l'état  $s_2$ , ...,  $b_l$  fois l'état  $s_l$ , au lieu de  $a_1$  fois l'état  $s_1$ , ...,  $a_l$  fois l'état  $s_l$ . On peut donc écrire cette énergie apparente, en tenant compte des formules (71), sous la forme :

$$NH_r = N \sum_k \left( \sum_i p_i \bar{w}_{ik} \right) \log \frac{1}{\sum_i p_i \bar{w}_{ik}} \quad (74)$$

$NH_r$  n'est pas l'énergie utilisable, à cause de l'incertitude résultant des erreurs.

L'observateur placé à la réception doit donc ou bien chercher à déterminer le nombre moyen d'états identifiés correctement, pour en déduire la procédure à suivre pour corriger les fautes et c'est le problème de la correction des erreurs - ou bien déterminer le nombre de cas où le symbole reçu est erroné, sans chercher pour autant à l'identifier - et c'est le problème de la détection des erreurs. Nous allons examiner successivement ces deux problèmes.

## CORRECTION DES ERREURS

Les erreurs peuvent être corrigées si on possède une connaissance complète de la corrélation entre le symbole émis et le symbole reçu. Ceci revient à substituer au système (S) qui constitue la source de données, un système virtuel (S') qui pourrait prendre  $l^2$  états distincts  $s_{11}, s_{12} \dots s_{ll}$ .

L'état  $s_{ij}$  du système (S') est caractérisé par la présence du symbole  $s_i$  à l'émission et du symbole  $s_j$  à la réception. La connaissance simultanée de l'état du récepteur et de celui de l'émetteur permet en effet de déterminer la transformation qui, dans le récepteur, permettra de substituer le symbole correct du symbole erroné ; c'est bien ainsi que l'on procède, sinon avec une certitude totale, tout au moins avec une très grande probabilité de succès, dans les systèmes à code correcteur, puisque c'est à partir du symbole reçu que l'on détermine le symbole émis ; la partie principale du mot émis indique la réception d'un certain symbole  $s_j$ , mais l'analyse de l'ensemble (partie principale du mot et complément autocorrecteur) montre qu'avec une très grande probabilité, c'est  $s_i$  qui a été émis et non  $s_j$ .

Nous sommes donc conduits à rechercher l'énergie diacritique  $NH'$  du système (S'), c'est-à-dire la quantité globale d'information qui est nécessaire pour le définir sans ambiguïté. (1) Il s'agit de dénombrer les configurations discernables d'un système à  $l^2$  états ; ce nombre est évidemment :

$$\frac{N!}{\beta_{11}! \beta_{12}! \dots \beta_{ll}!} \quad (75)$$

$$\left( \sum_{ij} \beta_{ij} = N \right)$$

Le calcul habituel donne sans difficulté, en tenant compte des formules (72) :

$$H' = \sum_i \sum_j p_i \bar{w}_{ij} \log \frac{1}{p_i \bar{w}_{ij}} \quad (76)$$

On peut écrire la formule (75) sous la forme :

$$\frac{N!}{\beta_{11}! \beta_{12}! \dots \beta_{ll}!} = \frac{N!}{a_1! a_2! \dots a_l!} \cdot \frac{a_1! a_2! \dots a_l!}{\beta_{11}! \beta_{12}! \dots \beta_{ll}!} =$$

$$\frac{N!}{b_1! b_2! \dots b_l!} \cdot \frac{b_1! b_2! \dots b_l!}{\beta_{11}! \beta_{12}! \dots \beta_{ll}!} \quad (77)$$

soit, en prenant les logarithmes et en divisant par  $N$  :

$$H' = H_0 + H'_c = H_r + H'_i \quad (78)$$

On a déjà défini  $H_0$ , énergie diacritique du système (S).  $H_r$  est l'énergie diacritique apparente du message reçu, liée au nombre de configurations que l'on trouve, en moyenne, à la réception, en raison des perturbations que subit le message ;

-----  
 (1) En pratique, nous ne pouvons évidemment connaître le système (S'), puisque nous sommes placés soit près de l'émetteur, soit près du récepteur. Mais c'est l'analyse de ce système virtuel qui permet d'évaluer les quantités d'information correspondant à la correction intégrale des erreurs.

$H'_c$  est l'énergie diacritique corrective : c'est celle qui doit être fournie, en supplément, pour obtenir une connaissance complète du système (S') à  $l^2$  états. Supposons qu'on établisse, par un moyen quelconque, un message complémentaire permettant de corriger les erreurs : la correction ne peut être faite que si l'énergie diacritique de ce message auxiliaire est au moins égale à  $NH'_c$  :  $H'_c$  est ainsi une mesure de l'information transmise par un système de correction du message, donc une estimation de l'énergie diacritique minimale à transmettre pour que la correction soit possible.

$H'_i$  est l'énergie diacritique d'incertitude, qui caractérise l'ambiguïté du choix que doit faire le récepteur. En effet, lorsque l'on reçoit  $b_1$  fois l'état  $s_1$ , il faut décider  $\beta_{11}$  fois d'afficher  $s_1$ ,  $\beta_{21}$  fois d'afficher  $s_2$ , ...  $\beta_{l_1}$  fois d'afficher  $s_l$ .

Mais il y a évidemment doute, car il existe  $\frac{b_1!}{\beta_{11}! \beta_{21}! \dots \beta_{l_1}!}$  configurations répondant à ce critère. Pour un message reçu comportant  $b_1$  fois  $s_1$ , ...  $b_l$ , fois  $s_l$ , le récepteur est incapable, en l'absence d'un système de correction d'erreur de choisir parmi  $\frac{b_1! b_2! \dots b_l!}{\beta_{11}! \beta_{21}! \dots \beta_{l_1}! \beta_{l_2}! \dots \beta_{l_l}!}$  configurations, ce qui correspond bien au dernier facteur de (77), dont le logarithme est  $H'_i$ . Nous verrons plus loin que  $H'_i$  est lié à l'information utilisable transmise par le système.

Evaluons maintenant  $H'_c$  et  $H'_i$ . Par le calcul élémentaire on trouve :

$$H'_i = \sum_i \sum_j p_i \varpi_{ij} \log \frac{\sum_i p_i \varpi_{ij}}{p_i \varpi_{ij}} \quad (79)$$

$$H'_c = \sum_i \sum_j p_i \varpi_{ij} \log \frac{1}{\varpi_{ij}} \quad (80)$$

Il est intéressant d'examiner ce que donnent les formules (76), (79), (80) dans les cas-limites les plus simples, pour lesquels l'interprétation des résultats est aisée.

#### 1er Cas - Transmission sans erreur.

On a alors  $\varpi_{ij} = \delta_{ij}$  ( $\delta_{ij}$  = delta de Kronecker). Dans les formules les termes rectangles sont nuls. On trouve simplement :

$H' = \sum_i p_i \log \frac{1}{p_i} = H$ ,  $H'_c = H'_i = 0$ . L'énergie d'incertitude est nulle, ce qui est évident a priori.

#### 2ème Cas - Transcodage systématique.

Chaque symbole  $s_i$  est systématiquement transformé en un autre symbole  $s_j$ , suivant une règle immuable. Ceci revient à dire que dans chaque ligne de la matrice des  $\varpi_{ij}$  un terme et un seul n'est pas nul. On aboutit exactement au même résultat que dans le premier cas, ce qui n'est guère surprenant, puisqu'un transcodage systématique n'introduit aucune perte d'information.

### 3ème Cas - Incertitude totale.

Elle se produit lorsque les probabilités conditionnelles sont égales :  $\omega_{ij} = 1/l$  ; dans ce cas on a autant de chances de recevoir un symbole donné, quel que soit celui qui a été émis. On trouve alors :

$$H' = \log l + H_0, H'_c = \log l, H'_i = H_0$$

Ce résultat n'est pas surprenant : en effet  $\omega_{ij} = 1/l$  signifie que la probabilité de recevoir un symbole donné  $s_i$  ne dépend pas de  $i$ , donc ne dépend pas de l'état de l'émetteur. L'énergie diacritique correspond à l'identification d'un état parmi  $l$  possibles et équiprobables est effectivement  $\log l$ , et cette énergie s'ajoute à celle  $H_0$ , qui caractérise la suite des états de l'émetteur. Si, par surcroît, on a  $p_i = 1/l$ , quel que soit  $i$ , on trouve  $H' = 2 \log l$ , énergie diacritique (par état) d'un système (S') possédant  $l^2$  états équiprobables. On voit dans un tel cas que la quantité d'information que doit transmettre un message de correction des erreurs est  $\log l$ , ce qui est évident puisque la connaissance de l'état de l'émetteur apportée par le message "principal" est strictement nulle.

On montre dans l'Annexe IV que le maximum relatif de  $H'_i$  obtenu pour  $\omega_{ij} = 1/l$ , est effectivement  $H_0$ , le maximum absolu étant  $\log l$ .

Energie diacritique utilisable. Dans un système de ce genre, le récepteur reçoit une certaine configuration des symboles, exacte ou erronée, et doit afficher la configuration correcte, ou tout au moins indiquer le groupe de configurations parmi lesquelles doit se trouver la configuration correcte, identique à celle affichée par l'émetteur. La connaissance a priori de l'émetteur permet de dire qu'il y a (en moyenne)  $e^{NH_0}$  configurations possibles. Mais on a vu qu'il y a  $e^{NH_i}$  configurations (en moyenne) entre lesquelles le récepteur est incapable de choisir, pour chaque configuration reçue. On peut donc dire que la seule information utilisable concerne le nombre de "groupes" de configurations entre lesquels on peut faire un choix, soit  $e^{NH_0}$  ;  $e^{NH_i} = e^{N(H_0 - H'_i)}$ .

L'énergie diacritique utilisable, par symbole, est donc :

$$H'_u = H_0 - H'_i = H_0 \left(1 - \frac{H'_i}{H_0}\right) = H_0 \left[1 - \frac{H'_i}{(H'_i)_m}\right]$$

en désignant par  $(H'_i)_m$  le maximum absolu de  $H'_i$ .

La formule (81) donne bien  $H'_u = H_0$  pour la transmission sans erreur,  $H'_i = 0$  pour l'incertitude totale.

De plus, le facteur  $1 - \frac{H'_i}{(H'_i)_m}$  apparaît comme un facteur de rendement qui fournit la proportion de l'énergie diacritique émise qui est encore "utilisable" en dépit des erreurs apparues dans la transmission.

## DETECTION DES ERREURS

La détection des erreurs met en oeuvre un processus plus simple que la correction. En effet, dans ce cas, lorsqu'un symbole  $s_i$  est reçu, on ne cherche pas à savoir quel symbole  $s_j$  parmi les  $l$  possibles a été effectivement émis, mais simplement si  $s_i$  peut être accepté ou refusé.

Une telle procédure correspond à la connaissance complète d'un système virtuel ( $S''$ ) susceptible de prendre  $2l$  états différents : pour chaque symbole  $s_i$ , réception correcte de ce symbole, ou réception incorrecte qui doit conduire à afficher une erreur.

La procédure peut être calquée sur celle du système à correction d'erreurs. Le système ( $S''$ ), à  $2l$  états, est caractérisé par  $\beta_{i1}$  réceptions correctes du symbole  $s_i$  et  $b_i - \beta_{i1}$  réceptions incorrectes de ce même symbole. Pour  $N$  symboles émis, le nombre de configurations discernables est donc :

$$\frac{N}{\beta_{11} ! (b_1 - \beta_{11}) ! \beta_{22} ! (b_2 - \beta_{22}) ! \dots \beta_{ll} ! (b_l - \beta_{ll}) !} \quad (82)$$

On trouve ainsi sans difficulté :

$$H'' = \sum_i p_i \bar{\omega}_{i1} \log \frac{1}{p_i \bar{\omega}_{i1}} + \sum_i \left( \sum_{k \neq i} p_k \bar{\omega}_{ki} \right) \log \frac{1}{\sum_{k \neq i} p_k \bar{\omega}_{ki}} \quad (83)$$

Dans (83) la somme  $\Sigma'$  est étendue à tous les indices  $k$  distincts de  $i$  ( $l - 1$  termes), alors que  $\Sigma$  s'étend de  $i = 1$  à  $i = l$ .

Comme dans le cas précédent, on peut développer (82) sous les formes équivalentes suivantes :

$$\frac{N !}{\beta_{11} ! (b_1 - \beta_{11}) ! \dots \beta_{ll} ! (b_l - \beta_{ll}) !} = \frac{N !}{a_1 ! a_2 ! \dots a_l !} \quad (84)$$

$$\frac{a_1 ! a_2 ! \dots a_l !}{\beta_{11} ! (b_1 - \beta_{11}) ! \dots \beta_{ll} ! (b_l - \beta_{ll}) !} = \frac{N !}{b_1 ! b_2 ! \dots b_l !} \frac{b_1 ! b_2 ! \dots b_l !}{\beta_{11} ! (b_1 - \beta_{11}) ! \dots \beta_{ll} ! (b_l - \beta_{ll}) !}$$

En prenant les logarithmes et en divisant par  $N$ , on obtient :

$$H'' = H_0 + H_d'' = H_r + H_i'' \quad (85)$$

$H_0$  et  $H_r$  sont les énergies diacritiques réelle et apparente de ( $S$ ),

$H_d''$  est l'énergie diacritique de détection d'erreur,

$H_i''$  est l'énergie diacritique d'incertitude.

On en déduit l'expression :

$$H_i'' = \sum_i p_i \bar{\omega}_{i1} \log \frac{\sum_k p_k \bar{\omega}_{ki}}{p_i \bar{\omega}_{i1}} + \sum_i \left( \sum_{k \neq i} p_k \bar{\omega}_{ki} \right) \log \frac{\sum_k p_k \bar{\omega}_{ki}}{\sum_{k \neq i} p_k \bar{\omega}_{ki}} \quad (86)$$



(Rappelons que la somme  $\Sigma$  est étendue à tous les indices, mais  $\Sigma'$  aux indices  $k \neq i$ ).

Examinons maintenant quelques cas-limites permettant d'interpréter ces formules.

1er Cas - Transmission sans erreur.

On a alors  $\varpi_{ij} = \delta_{ij}$ . Les termes rectangles sont nuls, et on trouve facilement :

$$H'' = \sum_i p_i \log \frac{1}{p_i} = H_0, \quad H_d'' = H'' - H_0 = 0, \quad H_i'' = 0$$

Ce résultat est évident a priori.

2ème Cas -

$$\varpi_{ij} = 1/l \text{ (pour tout couple } i, j)$$

C'est le cas où la probabilité de recevoir un symbole quelconque ne dépend pas de l'état de l'émetteur.

Pour chaque symbole émis, la probabilité de réception incorrecte est beaucoup plus grande que la probabilité de réception correcte : en effet, c'est  $\varpi_{ii} = 1/l$  pour le deuxième cas et  $\sum_j \varpi_{ij} = \frac{l-1}{l}$  pour le premier. On a donc beaucoup moins de chances de se tromper en affichant une erreur qu'en acceptant le symbole reçu. Il en résulte que l'énergie diacritique d'incertitude n'est certainement pas maximale dans ce cas.

On trouve facilement les valeurs de  $H''$ ,  $H_d''$ ,  $H_i''$  relatives à ce cas :

$$H'' = \log l + \frac{1}{l} \left[ \sum_i \frac{1}{p_i} \log \frac{1}{p_i} + \sum_i q_i \log \frac{1}{q_i} \right]$$

$$H_d'' = H'' - H_0 = \log l - \sum_i p_i \log \frac{1}{p_i} + \frac{1}{l} \left[ \sum_i p_i \log \frac{1}{p_i} + \sum_i q_i \log \frac{1}{q_i} \right]$$

$$H_i'' = \frac{1}{l} \left[ \sum_i p_i \log \frac{1}{p_i} + \sum_i q_i \log \frac{1}{q_i} \right]$$

On a posé ici :  $q_i = \sum_k' p_k = 1 - p_i$

L'interprétation des formules est intéressante. Le terme  $\log l$  correspond simplement à l'énergie diacritique associée à un système comportant  $l$  états équiprobables : c'est-évidemment la valeur prise par  $H_r$  (voir par exemple formule 71).

La valeur de  $H_i''$  représente l'énergie diacritique moyenne, par état, correspondant à l'éventualité qu'un symbole donné, s'il a été reçu, a effectivement été émis ou au contraire qu'il n'a pas été émis. En effet, à la réception, on est, pour chaque symbole devant l'alternative simple : accepter le symbole reçu ou afficher une erreur. Il en était tout autrement dans le cas de la correction des erreurs où il fallait en effet, pour chaque symbole reçu, déterminer celui qui avait été émis, parmi les  $l$  symboles possibles.

On vérifie facilement que le maximum relatif de  $H''$  s'obtient pour  $p_i = 1/l$ ,  $q_i = \frac{l-1}{l}$  : il vient alors  $H_i'' = \log l - \frac{l-1}{l} \log (l-1)$ , qui est une fonction décroissante de  $l$ , inférieure à  $\log 2$  pour  $l > 2$ . Ceci montre que l'incertitude est d'autant plus faible que  $l$  est plus grand, et en tous cas (sauf pour  $l = 2$ ) toujours inférieure à l'incertitude maximale ( $\log 2$ ). Cette constatation correspond, comme il a été dit plus haut, au fait qu'on a beaucoup moins de chances de se tromper, surtout si  $l$  est grand, en refusant le symbole reçu qu'en l'acceptant.

### 3ème Cas -

$$\bar{\omega}_{i,i} = \frac{1}{2}, \quad \bar{\omega}_{i,j} = \frac{1}{2(l-1)} \quad (i = j)$$

On obtient alors dans le système à détection d'erreur, l'incertitude totale. En effet, il y a autant de chances, pour chaque symbole reçu, qu'il ait été effectivement émis ou qu'il n'ait pas été émis :  $\bar{\omega}_{ii} = \sum_j \bar{\omega}_{ij} = 1/2$

On trouve sans difficulté :

$$\left\{ \begin{array}{l} H'' = \log 2 + \frac{1}{2} \left[ H_0 + \log (l-1) + \frac{1}{l-1} \sum_i q_i \log \frac{1}{q_i} \right] \\ H_d'' = H'' - H_0 \\ H_i'' = \frac{1}{2} \left[ \sum_i p_i \log \frac{p_i + \frac{q_i}{l-1}}{p_i} + \sum_i \frac{q_i}{l-1} \log \frac{p_i + \frac{q_i}{l-1}}{\frac{q_i}{l-1}} \right] \end{array} \right.$$

Il est facile de vérifier que  $H_i''$  est l'énergie diacritique moyenne correspondant à l'éventualité qu'un symbole reçu a effectivement été émis ou au contraire qu'il n'a pas été émis.

L'expression de  $H_i''$  est un maximum relatif (par rapport aux  $\omega_{ij}$ ) ; le maximum absolu s'obtient pour  $p_i = \frac{1}{l}$ ,  $q_i = \frac{l-1}{l}$  ; on trouve alors :

$$H'' = \log l + \log 2$$

$$H_i'' = H_d'' = \log 2$$

C'est l'incertitude totale : lorsqu'on reçoit un symbole, il y a autant de raisons de l'accepter que de le refuser. Il en résulte que l'information "utilisable" est strictement nulle, puisqu'aucune décision raisonnable ne peut être prise par le récepteur, en l'absence d'un dispositif de détection d'erreur.

### Energie diacritique utilisable.

Revenons rapidement sur la signification des grandeurs qui interviennent dans les formules précédentes.

Comme dans le cas de la correction d'erreur,  $H_d''$  est l'énergie diacritique (par symbole) que doit véhiculer un "message complémentaire" pour détecter les erreurs de transmission, sans pour autant identifier le symbole

correct.  $H_d''$  est donc une mesure de la dimension minimale à donner à un "canal auxiliaire" pour assurer une détection des erreurs.

$H_1''$ , énergie diacritique d'incertitude, mesure l'ambiguïté du choix que doit faire le récepteur, qui doit, à chaque moment, accepter ou refuser un symbole reçu. En effet, si l'on reçoit  $b_1$  fois le symbole  $s_1$ , il faut l'accepter  $\beta_{11}$  fois et le refuser  $b_1 - \beta_{11}$  fois, mais il y a  $\frac{b_1!}{\beta_{11}! (b_1 - \beta_{11})!}$  façons de le faire.

D'où l'expression (86), logarithme du dernier facteur de (84).

Or nous avons vu que le maximum absolu de  $H_1''$  est  $\log 2$ , ce qui est d'ailleurs évident, puisque pour chaque configuration reçue (de  $N$  symboles), il y a  $2^N$  interprétations possibles, suivant qu'on accepte ou qu'on refuse les divers symboles de cette configuration, et il y a ambiguïté totale lorsqu'elles sont équiprobables.

En nous référant à la formule (81), on voit donc que le "facteur de rendement" qui intervient à la réception est égal à :

$$1 - \frac{H_1''}{(H_1'')_m} = 1 - \frac{H_1''}{\log 2}$$

D'où l'expression de l'énergie diacritique utilisable :

$$H_u'' = H_0 \left[ 1 - \frac{H_1''}{\log 2} \right]$$

(87) donne bien  $H_u'' = H_0$  pour une transmission sans erreur, et  $H_u'' = 0$  lorsque l'incertitude est totale, comme dans le 3ème cas traité dans ce paragraphe.

## CHAPITRE VI

### APPLICATION DE LA THEORIE DES ERREURS AUX MESSAGES BINAIRES

#### SYSTEME A DEUX ETATS.

Considérons un système (S) susceptible de prendre deux états, dont les probabilités respectives sont p et q ( $p + q = 1$ ). On lui associe normalement un message binaire. Dans la transmission du message peuvent apparaître certaines erreurs : cherchons à évaluer leur effet sur l'information transmise : énergies diacritiques d'incertitude et de correction, travail sémiodynamique supplémentaire à mettre en jeu pour corriger les erreurs.

Nous partons d'un système (S) de pression diacritique  $H_0 : H_0 = p \log \frac{1}{p} + q \log \frac{1}{q}$  ; le travail sémiodynamique minimal nécessaire pour transmettre l'information est  $U_m = \varepsilon_1$ , si on emploie un message bivalent ( $\varepsilon_1 = \varepsilon_2 = 0,69$  logon). Par exemple, si  $p = q = 0,5$ , on a évidemment :

$$U_m = H_0 = 0,69 \text{ logon/symbole (1 e.b./symbole)}$$

L'apparition d'erreurs se traduit par les probabilités conditionnelles  $\overline{w}_{ij}$  de recevoir le symbole  $s_j$  si  $s_i$  a été émis = (on admet, pour simplifier, le même taux d'erreurs  $\eta$  pour les deux valences). On a vu, au chapitre précédent, que la connaissance des erreurs est équivalente à celle d'un système virtuel ('S') susceptible de prendre 4 états, dont les probabilités respectives sont :

$$\begin{cases} \overline{w}_{11} = 1 - \eta, \overline{w}_{12} = \eta \\ \overline{w}_{21} = \eta, \overline{w}_{22} = 1 - \eta \end{cases}$$

$$\begin{cases} p_1 = p & (1 - \eta) \text{ (émission de } s_1, \text{ réception de } s_1) \\ p_2 = q & (1 - \eta) \text{ (émission de } s_2, \text{ réception de } s_2) \\ p_3 = p\eta & \text{(émission de } s_1, \text{ réception de } s_2) \\ p_4 = q\eta & \text{(émission de } s_2, \text{ réception de } s_1) \end{cases}$$

Notons au passage que dans ce cas, il y a évidemment identité complète entre correction et détection des erreurs : puisqu'il n'y a que deux états possibles pour S, identifier une erreur, c'est évidemment la corriger.

Il suffit d'appliquer les formules générales données au chapitre V pour déterminer les diverses énergies diacritiques. On trouve :

$$H' = \overline{p\overline{\omega}}_{11} \log \frac{1}{p\overline{\omega}_{11}} + p\overline{\omega}_{12} \log \frac{1}{p\overline{\omega}_{12}} + q\overline{\omega}_{21} \log \frac{1}{q\overline{\omega}_{21}} + q\overline{\omega}_{22} \log \frac{1}{q\overline{\omega}_{22}}$$

$$= H_o + (1 - \eta) \log \frac{1}{1 - \eta} + \eta \log \frac{1}{\eta}$$

$$H'_c = \eta \log \frac{1}{\eta} + (1 - \eta) \log \frac{1}{1 - \eta}$$

$$H'_i = p(1 - \eta) \log \frac{p(1 - \eta) + q\eta}{p(1 - \eta)} + p\eta \log \frac{p\eta + q(1 - \eta)}{p} + q\eta \log \frac{p(1 - \eta) + q\eta}{q\eta}$$

$$+ q(1 - \eta) \log \frac{p\eta + q(1 - \eta)}{q(1 - \eta)}$$

Prenons comme exemple numérique  $p = q = 0,50$ ,  $\eta = 0,01$ . Il vient :

$$H' = 0,746 \text{ logon/symbole } (1,08 \text{ e.b./symbole})$$

$$H'_c = 0,056 \text{ logon/symbole } (8.10^{-2} \text{ e.b./symbole})$$

$$H'_i = 0,056 \text{ logon/symbole } (8.10^{-2} \text{ e.b./symbole}).$$

L'égalité de  $H'_i$  et de  $H'_c$ , donc de  $H_r$  et de  $H_o$ , vient de l'hypothèse  $p = q = 0,50$ . On voit donc que l'énergie diacritique (ou quantité d'information) correspondant à la correction des erreurs est équivalente à 8 éléments binaires pour 100 "utiles", ou encore que l'énergie diacritique utilisable est  $H_o - H'_i = 0,634$  au lieu de 0,69 logon/symbole.

Examinons maintenant le travail sémiodynamique minimal à mettre en jeu (1). Nous pouvons, pour ce faire, considérer la fonction d'état  $Z$  du système (S') :

$$Z = e^{-\mu\varepsilon_1} + e^{-\mu\varepsilon_2} + e^{-\mu\varepsilon_3} + e^{-\mu\varepsilon_4}$$

où les  $\varepsilon_i$  désignent les énergies sémiodynamiques associées à chaque état.

$$Z = e^{-\mu\varepsilon_1} \left[ 1 + e^{-\mu(\varepsilon_2 - \varepsilon_1)} + e^{-\mu(\varepsilon_3 - \varepsilon_1)} + e^{-\mu(\varepsilon_4 - \varepsilon_1)} \right]$$

Mais, à l'état d'équilibre on a :

$$\mu(\varepsilon_2 - \varepsilon_1) = \log \frac{p_1}{p_2} = \log \frac{p}{q}$$

$$\mu(\varepsilon_3 - \varepsilon_1) = \log \frac{p_1}{p_3} = \log \frac{1 - \eta}{\eta}$$

$$\mu(\varepsilon_4 - \varepsilon_1) = \log \frac{p_1}{p_4} = \log \frac{p}{q} \frac{1 - \eta}{\eta}$$

D'où :

$$Z = e^{-\mu\varepsilon_1} \frac{1}{p(1 - \eta)}$$

(1) Cette évaluation repose sur l'association, au système virtuel (S'), d'un message fictif qui transmettrait toute l'information relative aux états successifs pris par ce système.

On en déduit :

$$\left\{ \begin{aligned} U_0 &= -\frac{1}{\mu} \log Z = \varepsilon_1 - \frac{1}{\mu} \log \frac{1}{p(1-\eta)} \\ &= U_0 + \frac{1}{\mu} H' = H_0 + \frac{1}{\mu} [H_0 + H'_c] \end{aligned} \right.$$

U est donc déterminé dès que le choix de  $\varepsilon_1$  et de  $\mu$  est fait.

Il est évident qu'on a intérêt à choisir  $\varepsilon_1$  aussi petit que possible, pour réduire le travail sémiodynamique. Or, l'information normale qu'on peut transmettre avec un système bivalent (1 e.b., soit  $\varepsilon_1 = 0,69$  logon) est de 0,60 logon/symbole (pression diacritique maximale, 1 e.b. symbole) ; ici la pression diacritique est  $H' = 0,746 > 0,69$  logon. Plus généralement, quel que soit le taux d'erreur, on a évidemment  $H' = H_0 + H'_c > H_0 = 0,69$ . Il est impossible de prendre la valence 2, il faut donc prendre  $v = 3$ . C'est d'ailleurs évident, car on a vu à l'annexe II que le système bivalent ne permet tout au plus de définir que 3 états, si l'un des deux correspond à un symbole de durée minimale  $\tau$ .

Or, le système (S') possède 4 états. Il faut donc passer à  $v = 3$ , et il est recommandable de prendre deux symboles de durée minimale  $\tau$ , pour les deux états de grande probabilité :

$$\varepsilon_1 = \varepsilon_2 \log 3 = 1,10 \text{ logon (pour } p_1 = p_2 = 0,495).$$

C'est en tout état de cause, la plus petite valeur que l'on peut assigner à l'énergie sémiodynamique d'un symbole. Mais ici, deux solutions peuvent être comparées, conformément aux considérations développées au chapitre III.

#### 1ère possibilité :

On fixe  $\mu_0$ , et non pas l'alphabet (problème de Shannon)

Rappelons que dans l'application numérique, avec  $p = q = 0,50$ ,  $\eta = 0,01$ , il vient :

$$p_1 = p_2 = 0,495, \log \frac{1}{p_1} = 0,70, p_3 = p_4 = 0,005, \log \frac{1}{p_3} = 5,29.$$

Donc :

$$\mu_0 \varepsilon_1 = \log \frac{1}{p_1} = 0,70, \mu_0 = \frac{0,70}{1,10} = 0,63.$$

$$\varepsilon_3 \text{ se déduit de } \mu_0 : \quad \varepsilon_3 = \frac{1}{\mu_0} \log \frac{1}{p_3} = \frac{5,29}{0,63} = 8,30.$$

$$\text{D'où :} \quad U_{\min} = \frac{H}{\mu_0} = \frac{0,746}{0,63} = 1,17$$

Remarquons d'emblée que cette solution est toute théorique, car le rapport  $\varepsilon_3/\varepsilon_1$  doit être entier ; or ici,  $\varepsilon_3/\varepsilon_1 = 7,56$ . Il faudrait prendre en réalité  $\varepsilon_3 = 8 \varepsilon_1$ , ce qui augmente légèrement le travail par symbole U (1,18 logon environ).

## 2ème possibilité :

On choisit l'alphabet, a priori. On va naturellement prendre les énergies sémiodynamiques minimales :

$$\begin{cases} \varepsilon_1 = \varepsilon_2 = \log 3 = 1,10 \text{ logon} \\ \varepsilon_3 = \varepsilon_4 = 2 \log 3 = 2,20 \text{ logons} \end{cases}$$

Ce choix correspond exactement au cas 2 b/, valence 3, traité dans l'annexe II.

Employons la méthode décrite dans l'annexe V ; le polynôme caractéristique est  $f(\lambda) = 2(\lambda + 1)$

$$\text{d'où} = \lambda_1 = \frac{1}{2 p_1} - 1 = \frac{1}{0,99} - 1, \log \frac{1}{\lambda_1} = \mu(\varepsilon_3 - \varepsilon_1) = 4,59.$$

Le point (U, H) est sur la courbe ( $\Gamma$ ), puisque  $p_3 = \lambda_1 p_1 = 0,005$ .

$$\text{On trouve sans difficulté : } \begin{cases} U_0 = 0,932, \frac{1}{\mu} = 0,24, \mu = 4,17. \\ U = 2(p_1 \varepsilon_1 + p_3 \varepsilon_3) = 1,11 \text{ logon.} \end{cases}$$

On est donc dans le cas d'adaptation optimale ; le travail sémiodynamique par symbole (1,11 logon) est inférieur à celui trouvé précédemment (1,18). C'est évident, a priori, puisque  $\varepsilon_3 = 2,20$  et non 8 comme dans le cas précédent où  $\mu_0$  est fixé. Cet exemple met en évidence le caractère trop théorique du "problème de Shannon" : l'alphabet est pratiquement imposé, parce que l'énergie sémiodynamique minimale  $\varepsilon_1$  ne peut être inférieure à une valeur donnée pour le symbole de plus grande probabilité, et parce que celle des autres symboles est nécessairement un multiple de  $\varepsilon_1$ .<sup>(1)</sup>

Remarquons que dans le cas qui vient d'être traité, la courbe caractéristique de l'alphabet a pour extrémités : ( $H = 0, U_{\min} = 1,10$ ) et ( $H = 1,38, U_{\max} = 1,65$ ). Tant que U reste inférieur à 1,38 logon, on a intérêt à choisir l'alphabet donné ci-dessus. En revanche, dès que les taux d'erreur sont assez élevés pour que les probabilités des 4 états de (S') soient voisines, le travail minimal s'obtient pour  $\varepsilon_1 = \varepsilon_2 = \varepsilon_3 = \varepsilon_4 = 1,38$ , d'où  $U = 1,38$  (2 e.b./symbole). On vérifie sans peine que l'alphabet trivalent ( $\varepsilon_1 = 1,1$ ) est optimal jusqu'au point  $M_0$  de la courbe ( $\Gamma$ ) de rendement maximal ; en un tel point, on a précisément :  $U_0 = 0, \mu_0 = 0,91, 1/\mu_0 = 1,10, U = 1,38, H = 1,274, p_1 = p_2 = 0,366, p_3 = p_4 = 0,134, \eta = 0,268$ . Pour des probabilités d'erreur encore plus fortes, l'alphabet à choisir est l'alphabet quadrivalent, qui équivaut d'ailleurs à un alphabet bivalent comportant quatre symboles de durée  $2\tau$ .

Revenons maintenant à l'interprétation des résultats que nous avons trouvés. Nous avons vérifié qu'un taux d'erreur de 1 % sur les éléments binaires conduit à une perte de pression diacritique de 0,056 logon, soit environ  $8.10^{-2}$  e.b./symbole (92 e.b. d'informations "utiles" sur 100 émis). Mais le travail sémiodynamique minimal nécessaire pour transmettre un message corrigé est incomparablement plus important : 1,11 logon/symbole, soit

-----  
(1) Nous n'examinons pas ici le groupement en "mots" de longue durée qui permet théoriquement la dépense minimale,

encore 1,61 e.b./symbole. On peut encore dire, grossièrement, qu'à chaque bloc de 100 éléments binaires il faut ajouter 61 e.b. pour obtenir une connaissance complète du système (S'). Cet accroissement considérable est lié, non au taux d'erreur, mais à la nécessité d'identifier 4 états différents au lieu de 2. (En effet, la formule  $dU = 1/\mu dH'$  montre que U augmente seulement de 0,01 si H' augmente de 0,04, ce qui correspond au passage de  $\gamma = 0,01$  à  $\eta = 0,02$ ).

Un raisonnement assez grossier, mais intéressant, permet de s'en convaincre. Imaginons un observateur qui pourrait surveiller simultanément le message émis et le message reçu, donc avoir une connaissance permanente et parfaite du système (S'). Cet observateur, tel le démon de Maxwell qui ne laisse passer les particules que dans un seul sens, marque au passage tous les symboles qui vont être altérés dans la transmission : ceci revient, en pratique, à ajouter à chaque symbole un deuxième qui, suivant sa polarité, signifie "bon" ou "mauvais". Si, dans une séquence de 100 symboles successifs on était assuré qu'il n'y avait qu'un seul symbole erroné, il suffirait d'ajouter à la séquence 7 e.b. pour désigner le rang du symbole erroné (puisque  $2^7 = 128$ ). Mais, si en moyenne il y a bien 1 symbole faux sur 100 émis, il y a une probabilité non négligeable qu'il y en ait 2, ou même 3, ou même davantage. La probabilité qu'ils soient tous faux n'est pas nulle, aussi petite qu'elle soit. Suivant que le démon utilise 7, 14, 21, 28, ... e.b. supplémentaires (ou plus précisément 7, 13, 18, etc.), il pourra corriger les messages de 100 symboles ayant 1, 2, 3, ou 4 erreurs. Mais si le démon veut corriger intégralement les erreurs, il sera conduit, de proche en proche, à utiliser jusqu'à 2 e.b. par symbole au lieu d'un seul. En allant jusqu'à cette limite, on aura doublé la longueur de la séquence. Seulement, un démon qui procéderait de la sorte ne serait pas intelligent : en effet, même pour corriger intégralement les longues séquences, il est inutile d'en doubler la durée : un allongement de 61 % suffit, comme le montrent les calculs effectués plus haut. C'est qu'en effet le démon n'a pas tenu compte de la probabilité des erreurs : s'il y a autant de chances qu'un symbole soit altéré que transmis correctement, il faut effectivement doubler la longueur de la séquence, mais il faut seulement l'allonger de 61 % si la probabilité d'erreur n'est que de 1 % - On voit ainsi la signification réelle du travail sémiodynamique minimal nécessaire pour corriger un message.

En pratique, bien sûr, le démon n'existe pas et la correction d'erreurs repose sur l'emploi d'éléments binaires redondants. Le travail réellement dépensé ne pourrait être inférieur au travail minimal évalué ci-dessus, si le codage était totalement efficace ; mais le codage ne peut être totalement efficace car toute altération qui substitue, dans la transmission, un mot autorisé à un autre mot autorisé, ne peut être détectée. On se contente de rendre cette circonstance très improbable. Donc un système réel de correction (ou de détection) d'erreurs, dans la mesure où il ne fait que réduire le nombre de fautes, peut parfaitement correspondre à un travail sémiodynamique dépensé inférieur à celui que nous avons calculé. Mais ce dernier constitue une limite intéressante, puisqu'il correspond à la dépense minimale associée à un système de correction parfait (1).

-----  
(1) En pratique, un code détecteur (ou correcteur) de bonne efficacité correspond toujours à un travail sémiodynamique très supérieur à la valeur nominale théorique.



Remarque : Système "réflex".

Dans un système de transmission de type réflex, le message binaire est envoyé de A à B, mais conservé en mémoire en A, puis B émet à son tour vers A qui compare ce qu'il reçoit à ce qu'il a émis. S'il n'y avait aucune erreur de transmission de B vers A, on aurait une connaissance parfaite du système (S') dont on a parlé ci-dessus. Un canal de correction reliant A à B comporterait un message dont le travail sémiodynamique serait au moins égal à celui, U, que nous avons calculé. Mais dans tout système réel, il y a également des erreurs dans la retransmission de B vers A : un système binaire réflex est en fait un système de 8 états de probabilités inégales.

SYSTEME A l ETATS DIFFERENTS.

Considérons un système (S) susceptible de prendre  $l = 2^m$  états différents, on lui associe un alphabet composé de mots de même longueur (m éléments binaires). Un message composé de la sorte est équivalent à celui qu'on obtiendrait avec un alphabet de valence l comportant des symboles de même durée unitaire. Si l'on désigne par  $p_1, p_2, \dots, p_l$ , les probabilités des états de (S), sa pression diacritique est  $H_0 = \sum p_i \log \frac{1}{p_i}$ . Le travail sémiodynamique dépensé (par état) est  $U_m = \log l = m \log 2$ .

Si  $p_1 = p_2 = \dots = p_l = \frac{1}{l}$ , on a  $\mu_b = 1$ ,  $H_0 = U_m = \log l$ .

Nous nous placerons dans ce cas simple, dans tout ce qui suit.

Comme exemple numérique, prenons  $m = 7$   $H_0 = U_m = 7 \times 0,69 = 4,83$  logons (7 e.b.)

$$l = 128$$

Définition des erreurs :

Faisons l'hypothèse simple d'une répartition d'erreurs selon une loi binominale ; si  $\eta$  est la probabilité qu'un e.b. quelconque d'un mot soit erroné, la probabilité conditionnelle qu'un mot soit transmis sans erreur est  $\varpi_{11} = (1 - \eta)^m$ . Rangeons les l mots en en choisissant un quelconque  $s_1$  au départ (par exemple celui qui ne comporte que des zéros). A partir de  $s_1$ , on trouve m mots ne différant de  $s_1$  que par un élément binaire ; appelons-les  $s_2, s_3, \dots, s_{m+1}$  ; la probabilité conditionnelle  $\varpi_{12} (= \varpi_{13} = \varpi_{14} = \dots = \varpi_{1, m+1})$  a pour valeur  $\varpi_{12} = \eta (1 - \eta)^{m-1}$  ; on trouve ensuite  $C_m^2$  mots qui diffèrent de  $s_1$  par 2 e.b. ( $s_{\frac{m+2}{2}}, \dots, s_{\frac{m+1}{2}}$ ), avec  $\varpi_{1, m+2} = \eta^2 (1 - \eta)^{m-2}$ , et ainsi de suite.

Les probabilités conditionnelles de recevoir  $s_j$  si  $s_1$  est émis, sont donc de la forme :

$$\varpi_{1j} = \eta^d (1 - \eta)^{m-d} = \varpi_d$$

en désignant par d la distance de Hamming entre les mots  $s_1$  et  $s_j$  (nombre d'éléments binaires différents dans les deux mots).

Comme exemple numérique, prenons  $m = 7$ ,  $\eta = 0,01$  ; on obtient le tableau suivant :

d	$\bar{\omega}_d$	$\log \frac{1}{\bar{\omega}_d}$	$\bar{\omega}_d \log \frac{1}{\bar{\omega}_d}$
0	0,932	0,07	0,065
1	0,0094	4,66	0,0438
2	$0,95 \cdot 10^{-4}$	9,25	$8,8 \cdot 10^{-4}$
3	$0,96 \cdot 10^{-6}$	13,84	$1,38 \cdot 10^{-5}$
4	$0,97 \cdot 10^{-8}$	18,43	$1,8 \cdot 10^{-7}$
5	$0,98 \cdot 10^{-10}$	23,02	$2,2 \cdot 10^{-9}$
6	$0,99 \cdot 10^{-12}$	27,61	$2,7 \cdot 10^{-11}$
7	$10^{-14}$	32,2	$3,2 \cdot 10^{-13}$

### Correction des erreurs :

La connaissance des erreurs équivaut à celle d'un système (S') susceptible de prendre  $l^2$  états définis par l'émission de  $s_i$  et la réception de  $s_j$ . Supposons les probabilités a priori de (S) égales à  $1/l$ , on a donc pour (S') :

$l$	états de probabilité	$p_0 = \frac{1}{l} \bar{\omega}_0$	(128 états si $m = 7$ )
$ml$	" "	$p_1 = \frac{1}{l} \bar{\omega}_1$	(7 × 128 états)
$C_m^2 l$	" "	$p_2 = \frac{1}{l} \bar{\omega}_2$	(21 × 128 états)
....			
$C_m^m l$	" "	$p_m = \frac{1}{l} \bar{\omega}_m$	(1 × 128 états)

ce qui fait bien au total  $l^2$  états et une probabilité globale égale à 1.

Les formules (76), (79) et (80) fournissent sans difficulté les pressions diacritiques :

$$H' = \log l + \bar{\omega}_0 \log \frac{1}{\bar{\omega}_0} + C_m^1 \bar{\omega}_1 \log \frac{1}{\bar{\omega}_1} + \dots + C_m^m \bar{\omega}_m \log \frac{1}{\bar{\omega}_m}$$

$$= H_0 + H_c' = 4,83 + 0,390 = 5,22 \text{ logon (7,56 e.b.)}$$

$$H_1' = 0,39 \text{ logon (0,56 e.b./symbole)}$$

$$H_1' = H_c' = 0,39 \text{ logon (1).}$$

L'énergie diacritique (ou quantité d'informations) correspondant à la correction des erreurs, équivaut donc à 56 e.b. environ tous les 100 symboles donc tous les 700 e.b.). L'énergie diacritique utilisable est de 4,44 au lieu de 4,83 logons.

-----  
 (1) On a  $H_r = H_0$ , donc  $H_1' = H_c'$ , parce que les  $p_i$  sont égaux à  $1/l$  et que la matrice des  $\bar{\omega}_{ij}$  est symétrique, puisqu'ils ne dépendent que de la distance de Hamming entre 2 états  $s_i$  et  $s_j$ .

Evaluons le travail sémiodynamique minimal à mettre en jeu. Nous devons faire le choix d'un alphabet pour le système (S'). Les  $l^2$  états sont répartis dans 8 classes de probabilités différentes :  $p_0, p_1, \dots, p_7$ , de sorte qu'on aura au minimum 8 énergies sémiodynamiques différentes :  $\varepsilon_0, \varepsilon_1, \dots, \varepsilon_7$ .

Compte tenu du fait que la probabilité  $p_0'$  (correspondant aux états  $s_{ii}$  associés à la transmission sans erreur) est très supérieure aux autres probabilités, on a évidemment intérêt à rendre l'énergie sémiodynamique  $\varepsilon_0$  aussi faible que possible ; or, s'il faut pouvoir distinguer entre 8 classes différentes, on est conduit à augmenter l'énergie des symboles  $s_{ii}$  d'une quantité équivalente à  $\log_2 8 = 3$  e.b., ce qui est considérable ( $7 + 3 = 10$  e.b.). A l'extrême opposé, si on donne à tous les symboles la même énergie sémiodynamique, il faut l'équivalent de  $2m = 14$  e.b. pour chacun d'entre eux, car il y a  $l^2$  états au total, ce qui est évidemment défavorable. Le travail sémiodynamique minimal sera finalement obtenu en ne considérant que 2 classes seulement : celle des  $l$  symboles de forte probabilité ( $s_{ii}$ ) et celle des  $l(l-1)$  symboles de faible probabilité ( $s_{ij}$  avec  $i \neq j$ ). Dans la représentation binaire, ceci conduit à prendre  $m+1$  e.b. pour les  $l$  mots associés aux états  $s_{ii}$  (le premier élément binaire indiquant par exemple qu'on a affaire à la première classe) et  $2m+1$  e.b. pour ceux  $s_{ij}$  de la deuxième classe (le premier élément indique qu'on a affaire à la 2e classe).

On prendra donc :  $\varepsilon_0 = (m+1) \log 2 = 8 \times 0,69 = 5,52$  logons

$$\varepsilon_1 = (2m+1) \log 2 = 15 \times 0,69 = 10,35 \text{ logons.}$$

Il y a  $l$  éléments de probabilité  $p_0' = \frac{1}{l} \overline{p}_0$ ,  $m$  états de probabilité  $p_1' = \frac{1}{l} \overline{p}_1$ ,

etc. Le travail sémiodynamique dépensé par symbole est donc :

$$\begin{aligned} U &= l p_0' \varepsilon_0 + l \varepsilon_1 (m p_1' + C_m^2 p_2' + \dots + C_m^m p_m') \\ &= \overline{p}_0 \varepsilon_0 + (1 - \overline{p}_0) \varepsilon_1 = 0,932 \times 5,52 + 0,068 \times 10,35 = 5,848 \text{ logons} \end{aligned}$$

$$U = 5,848 \text{ logons} = 8,475 \text{ e.b.}$$

C'est le travail sémiodynamique minimal associé à la pression diacritique  $H' = 5,22$  logons (7,56 e.b.) du système (S').

En s'appuyant sur les résultats de l'annexe V, on vérifie facilement que le point ( $H = 5,22$ ,  $U = 5,848$ ) du plan ( $H$ ,  $U$ ) est situé nettement au-dessus de la courbe caractéristique de l'alphabet choisi. En effet, cet alphabet permet, pour un travail sémiodynamique  $U = 5,848$ , une pression diacritique  $H = 5,43$  logons. Le gradient sémiodynamique, au voisinage du point de fonctionnement, est  $1/\mu = 0,65$  ; il en résulte que lorsque l'adaptation est optimale, un travail de 5,71 logon/symbole suffit pour une pression diacritique de 5,22 logon. Le défaut d'adaptation correspond donc à un excès de 0,14 logon/symbole environ pour le travail sémiodynamique.

En définitive, le travail minimal que l'on doit effectuer pour corriger les erreurs correspond à une augmentation de  $5,848 - 4,83 = 1,018$  logon/symbole, ou encore 1,47 e.b./symbole, ou encore 0,21 e.b. par e.b. Il faudra donc dépenser au moins 147 e.b. supplémentaires tous les 100 symboles (donc tous les 700 e.b.), alors que l'énergie diacritique n'a augmenté, pour la même séquence, que de 56 e.b. soit environ 3 fois moins. Ce mau-

vais rendement est essentiellement dû à la nécessité "d'enrichir" l'alphabet pour discerner  $l^2$  états différents au lieu de  $l$ . En pratique, tous les systèmes réels de correction ont encore un rendement plus faible, et ils ne peuvent, par surcroît, assurer une correction intégrale des erreurs.

#### Détection des erreurs :

Examinons maintenant le cas de la détection des erreurs : on doit considérer le système virtuel ( $S''$ ) à  $2l$  états (symbole reçu correct ou erroné).

La pression diacritique de ce système vaut, d'après la formule (83) :

$$\begin{aligned} H'' &= \sum_1^l \frac{\varpi_0}{l} \log \frac{l}{\varpi_0} + \sum_1^l \sum' \frac{1}{l} C_m^k \varpi_k \log \frac{1}{\sum' C_m^k \varpi_k} \\ &= \varpi_0 \log l + \varpi_0 \log \frac{1}{\varpi_0} + (1 - \varpi_0) \log \frac{l}{1 - \varpi_0} \\ &= \log l + \varpi_0 \log \frac{1}{\varpi_0} + (1 - \varpi_0) \log \frac{1}{1 - \varpi_0} = H_0 + H_d'' \end{aligned}$$

On trouve sans difficulté :

$$H'' = 5,077 \text{ logons (7,36 e.b.)}$$

$$H_d'' = H_1'' = 0,247 \text{ logon (0,36 e.b.)}$$

Ces valeurs sont un peu inférieures à celles  $H'$ ,  $H_1'$ ,  $H_c'$ , obtenues pour la correction des erreurs.

L'alphabet à associer est évidemment celui qui comporte, pour tous les symboles, 8 e.b. au lieu de 7 ; on obtient ainsi exactement  $2l = 2 \times 128$  symboles, et de toute façon c'est l'énergie minimale possible pour les symboles les plus probables.

Le travail sémiodynamique est donc 5,52 logons (8 e.b.). Alors que la "perte d'information" équivaut à 36 e.b. tous les 100 symboles, il faudra dépenser en réalité 1 e.b./symbole, soit 100 tous les 100 symboles, pour détecter les erreurs.

Ces valeurs, données à titre d'exemple, montrent bien que le travail minimal nécessaire est plus important s'il s'agit de corriger les erreurs que s'il s'agit de les détecter. Mais de plus, il est bien clair qu'un système réel exige un travail bien plus considérable : en effet, un élément binaire supplémentaire par symbole ne permet en pratique de détecter que les erreurs de parité, ce qui est très rudimentaire.

Les deux exemples qui ont été traités dans ce chapitre montrent, en outre, l'intérêt d'une détection d'une correction d'erreurs par mots de durée suffisamment longue. Ce travail, associé à la détection, en particulier, n'exige qu'un élément supplémentaire par symbole, ce qui est notable pour un mot de 7 e.b., mais tout à fait négligeable pour un mot qui en comporte quelques dizaines ou quelques centaines. Cette remarque a une portée générale ; elle s'applique à tous les systèmes réels de détection et de correction.

En guise de conclusion, il convient d'insister sur la portée exacte de l'analyse logodynamique de la détection et de la correction des erreurs. La

théorie permet d'évaluer la "perte d'information" ou plus précisément la perte d'énergie diacritique liée à la transmission d'un message erroné : à titre d'exemple, dans un système binaire à deux états, la perte est de 8 % pour un taux d'erreur de 1 % sur les éléments binaires. La théorie permet en outre de mesurer l'énergie diacritique et le travail sémiodynamique supplémentaire qui correspondent à une compensation complète (idéale) des erreurs de transmission. Mais elle ne donne aucune indication sur le moyen pratique d'effectuer la détection ou la correction de ces erreurs. Elle permet simplement d'évaluer quelle quantité d'information, quel travail minimal on a mis en jeu lorsqu'on a su détecter ou corriger intégralement les erreurs des messages.

Ces limites théoriques permettent, d'une certaine façon, d'évaluer le "rendement" des messages codés redondants munis d'une "clef" (codes à détection ou autocorrection). Si l'on considère en effet un message codé correspondant au travail sémiodynamique minimal qu'implique la détection (ou la correction) intégrale des erreurs, pour une probabilité de faute donnée sur les éléments binaires, on peut évaluer la probabilité d'erreurs résiduelles qui résulte de son emploi. Plus cette probabilité est faible, plus le code est efficace, puisqu'il ressemble de plus en plus à cet observateur idéal fictif, qui, connaissant à l'avance ce qui va se passer à l'émission et à la réception, compenserait intégralement les fautes en dépensant le même travail.

ANNEXE I

$$\text{MAXIMUM DE } H = \sum p_i \log \frac{1}{p_i}$$

Soit  $H = \sum p_i \log \frac{1}{p_i}$ , fonction des variables  $p_i$ , liées elles-mêmes par la relation :

$$\sum p_i = 1.$$

Pour chercher le maximum de  $H$ , on peut employer la méthode du multiplicateur de Lagrange. On pose ;

$$Y = H - \lambda \sum p_i$$

D'où, en différentiant par rapport aux variables  $p_i$  :

$$\sum_i (\log \frac{1}{p_i} - 1 - \lambda) \delta p_i = 0$$

Tous les coefficients des  $\delta p_i$  doivent être nuls, donc :

$$\begin{cases} p_1 = p_2 = \dots = p_l = e^{-(1+\lambda)} \\ l e^{-(1+\lambda)} = 1. \end{cases}$$

Soit :

$$\begin{cases} 1 + \lambda = \log l \\ p_1 = p_2 = \dots = p_l = \frac{1}{l} \end{cases}$$

L'extremum est donc obtenu pour  $p_i = 1/l$ , quel que soit  $i$ .

Il s'agit bien d'un maximum, car si l'un des  $p_i$ , par exemple  $p_1$ , tend vers zéro, on peut écrire :

$$\begin{aligned} p_1 &= \eta && \eta \text{ arbitrairement petit positif} \\ \sum_{i=2}^l p_i &= 1 - \eta \\ H &= - p_1 \log p_1 + \sum_{i=2}^l p_i \log \frac{1}{p_i} \end{aligned}$$

Le premier terme vaut  $-\eta \log \eta$  qui tend vers zéro, et le deuxième est au maximum égal à  $\log (l - 1) < \log l = H_n$ .

De même, si l'un des  $p_i$ , par exemple  $p_1$ , tend vers 1, les autres tendent vers zéro, et  $H$  tend manifestement vers zéro.

## ANNEXE II

### DEMONSTRATION DE L'INEGALITE $\mu_0 \leq 1$

Considérons un alphabet possédant  $l$  symboles distincts ; ces symboles  $s_1, \dots, s_l$  se différencient soit par leur "niveau" seulement (amplitude, polarité, phase, etc.) et dans ce cas ils ont la même durée, donc la même énergie sémiodynamique  $\varepsilon_1$  et leur nombre est limité à  $v$ ,  $v$  étant la valence du message : soit par leur durée. Dans le cas général, on peut trouver, parmi les  $l$  symboles,  $\alpha_1$  symboles de même durée  $t_1$ ,  $\alpha_2$  de durée  $t_2$ ,  $\dots, \alpha_p$  de durée  $t_p$ , avec  $\sum_p \alpha_i = l$ .

L'hypothèse faite est qu'il n'y a aucune restriction dans les conditions d'assemblage. On obtient évidemment ce résultat lorsque les symboles ne se différencient que par leur niveau ; dans ce cas trivial, le message aura une puissance diacritique maximale en leur donnant la même durée minimale  $\tau$ , telle que :

$$w\tau = 1 \quad (\varepsilon_1 = \varepsilon_2 = \dots = \varepsilon_l = w\tau \log.v = \log v = \log l).$$

Dans ce cas très simple, la fonction d'état

$$Z = \sum_1^l e^{-\mu_0 \varepsilon_i} = l \exp(-\mu_0 \log v) = 1$$

d'où :

$$\mu_0 = \frac{\log l}{\log v} = 1 \text{ puisque } l = v$$

Dans le cas plus général où certains symboles ont même niveau, et ne se différencient que par leur durée, l'absence de restriction dans l'assemblage impose qu'un des niveaux soit réservé à un symbole particulier de séparation. En effet, s'il en était autrement, dans une suite de deux symboles de durées différentes mais de même niveau, il serait impossible de savoir lequel des deux a été émis le premier ; donc pour que des messages distincts soient discernables, il faut apporter des restrictions aux règles d'assemblage ce qui est contraire à l'hypothèse. Il y a plusieurs façons d'opérer cette séparation entre symboles qui évite toute indétermination sur l'assemblage.

La méthode la plus évidente - peu performante - consiste à faire une séparation systématique : un des niveaux sert uniquement à indiquer l'apparition d'un nouveau symbole, ce qui revient à dire qu'aucun symbole n'a

de durée inférieure à  $2\tau$ ,  $\tau$  étant l'intervalle élémentaire. On obtient ainsi  $v - 1$  symboles possibles de durée  $2\tau$ ,  $(v - 1)^2$  symboles possibles de durée  $3\tau$ ,  $(v - 1)^3$  symboles de durée  $4\tau$ , etc.

Si l'on classe donc les symboles en fonction de leur durée ( $\alpha_j$  symboles de durée  $j\tau$ ) on a :

$$\begin{aligned} \alpha_1 &= 0 \\ \alpha_2 &\leq v - 1 \\ \alpha_3 &\leq (v - 1)^2 \\ &\dots \\ \alpha_p &\leq (v - 1)^{p-1} \end{aligned}$$

avec 
$$\sum_1^p \alpha_p = l$$

Considérons maintenant la fonction d'état :  $Z = \sum_{i=1}^l e^{-\mu \epsilon_i} = Z(\mu)$ .

$Z$  est une fonction décroissante de  $\mu$ . Nous allons démontrer l'inégalité :

$$Z(1) = \sum_{i=1}^l e^{-\epsilon_i} < 1$$

Dans ces conditions, si  $Z(\mu_0) = 1$  (par définition de  $\mu_0$ ), on aura nécessairement :  $\mu_0 < 1$ . Les  $\alpha_2$  symboles de durée  $2\tau$  ont une énergie sémiodynamique :

$$\epsilon_2 = 2w\tau \log v = 2 \log v, \text{ donc } e^{-\epsilon_2} = \frac{1}{v^2}$$

de même :  $e^{-\epsilon_3} = \frac{1}{v^3}, \dots, e^{-\epsilon_p} = \frac{1}{v^p}$ .

On en déduit :

$$Z(1) = \sum_1^l e^{-\epsilon_i} = \alpha_2 e^{-\epsilon_2} + \alpha_3 e^{-\epsilon_3} + \dots + \alpha_p e^{-\epsilon_p} < \frac{v-1}{v^2} + \frac{(v-1)}{v} + \dots + \frac{(v-1)^{p-1}}{v^p}$$

$$Z(1) \leq \frac{v-1}{v} \left[ 1 - \left(\frac{v-1}{v}\right)^{p-1} \right] < 1$$

Dans ce premier cas, la proposition est donc démontrée : on a même nécessairement  $\mu_0 < 1$ , ce qui montre que la puissance diacritique obtenue n'est pas maximale.

Le principe de séparation des symboles qui vient d'être examiné interdisait tout symbole de durée  $\tau$ . Or, on peut vérifier que l'absence de restriction sur les règles d'assemblage équivaut au principe suivant :

- s'il y a  $v$  symboles de durée  $\tau$ , il ne peut y avoir d'autre symbole de durée supérieure à  $\tau$  ;
- s'il y a  $v - 1$  symboles de durée  $\tau$ , il y a au plus  $v$  symboles de durée  $2\tau$  (ou de durée supérieure à  $2\tau$ ) ;



- s'il y a  $v - 2$  symboles de durée  $\tau$ , il y a au plus  $2 v$  symboles de durée  $2 \tau$ , ou  $v$  de durée de  $2 \tau$  et  $v$  de durée  $3 \tau$ .

Plus généralement, s'il y a  $v - h$  symboles de durée  $\tau$  (et ne différant donc que par leur niveau) il y a au plus  $h_1 v$  symboles de durée  $2 \tau$ ,  $h_2 v^2$  symboles de durée  $3 \tau$ ,  $h_3 v^3$  symboles de durée  $4 \tau$ , ...,  $h_{p-1} v^{p-1}$  symboles de durée  $p \tau$ , avec  $h_1 + h_2 + \dots + h_{p-1} = h$ .

Remarquons au passage que cette règle est encore applicable pour  $h = v$  ; dans ce cas, on a par exemple au maximum  $h v = v^2$  symboles de durée  $2 \tau$  (nombre de doublets réalisables avec des signaux de valence  $v$ ).

Reprenons maintenant le raisonnement précédent, en montrant qu'on a encore, avec ces règles :

$$Z(1) = \sum_{i=1}^l e^{-\epsilon_i} < 1.$$

Les énergies sémiodynamiques des symboles sont  $\epsilon_1, \epsilon_2, \dots, \epsilon_p$  avec :

$$e^{-\epsilon_1} = \frac{1}{v}, e^{-\epsilon_2} = \frac{1}{v^2}, \dots, e^{-\epsilon_p} = \frac{1}{v^p}$$

Si on a  $\alpha_1 = v - h$  symboles d'énergie  $\epsilon_1, \epsilon_2 < h_1 v$  d'énergie  $\epsilon_2, \dots, \alpha < h_{p-1} v^{p-1}$  d'énergie  $\epsilon_p$  :

$$\begin{aligned} Z(1) &= \sum_{i=1}^l e^{-\epsilon_i} = \alpha_1 e^{-\epsilon_1} + \alpha_2 e^{-\epsilon_2} + \dots + \alpha_p e^{-\epsilon_p} \leq \frac{v-h}{v} + \frac{h_1 v}{v^2} + \frac{h_2 v^2}{v^3} + \dots + \frac{h_{p-1} v^{p-1}}{v^p} \\ &= \frac{v-h + h_1 + \dots + h_{p-1}}{v} = 1. \end{aligned}$$

On a par conséquent :

$$Z(1) \leq 1$$

et

$$\mu_0 \leq 1.$$

Cette démonstration est générale. Les exemples simples qu'on va donner maintenant montreront que la puissance diacritique obtenue est généralement maximale, contrairement au cas d'une séparation systématique des symboles.

ALPHABETS SIMPLES DE PUISSANCE DIACRITIQUE MAXIMALE ( $\mu_0 = 1$ ).

1er cas :

$v = 2$  - Dans un système bivalent, 2 possibilités :

a) 1 symbole pour chaque niveau (cas trivial) - Donc 2 symboles seulement ( $\epsilon = w \tau \log 2 = \log 2$ )

$$Z = 2 e^{-\mu_0 \epsilon} = 1, \mu_0 \epsilon = \log 2, \mu_0 = 1.$$

b) 1 niveau pour un symbole de durée  $\tau$  (+)

2e niveau associé au premier, pour former 2 symboles (- +, --)

(L'apparition de la valence négative indique que le symbole à prendre en considération est un doublet de durée  $2 \tau$ ).

on a :  $Z = e^{-\mu_0 \varepsilon} + 2 e^{-2\mu_0 \varepsilon} = 1$ , donc  $\mu_0 = 1$  puisque  $\varepsilon = \log 2$ ,

3 symboles distincts, de durée moyenne  $\frac{5}{3} \tau$ .

Remarque :

Bien entendu, nous ne considérons ici que le cas où un symbole au moins a la durée minimale  $\tau$ . Si  $v = h$ , on a évidemment  $h v = v^2$  - ici 4 - symboles possibles d'égale durée  $2 \tau$ ; on a encore  $\mu_0 = 1$ , et les quatre états possibles sont caractérisés par les combinaisons ++, +-, -+, --, etc. Ce cas est en réalité la représentation, sous forme bivalente, du cas quadrivalent "trivial" (Cf plus loin 3e cas a/).

2e cas :

$v = 3$  - Dans un système trivalent comportant au moins un symbole de durée minimale  $\tau$ , 4 possibilités :

a) 1 symbole par niveau - Donc : 3 symboles distincts (cas trivial)

$$Z = 3 e^{-\mu_0 \varepsilon} = 1 \quad \mu_0 \varepsilon = \log 3, \quad \mu_0 = 1 \text{ puisque } \varepsilon = w \tau \log v = \log 3.$$

b) 2 symboles de durée minimale, caractérisés par les niveaux  $a_1$  et  $a_3$   
 3 symboles de durée  $2 \tau$  :  $a_3 a_1, a_3 a_2, a_3 a_3$ .

$$Z = 2 e^{-\mu_0 \varepsilon} + 3 e^{-2\mu_0 \varepsilon} = 1, \quad \mu_0 = 1 \text{ puisque } \varepsilon = \log 3.$$

Donc, au total : 5 symboles distincts, de durée moyenne  $\frac{8}{5} \tau$ .

c) 1 symbole de durée minimale (niveau  $a_1$ )

6 symboles de durée  $2 \tau$  :  $a_2 a_1, a_2 a_2, a_2 a_3, a_3 a_1, a_3 a_2, a_3 a_3$ ,

$$Z = e^{-\mu_0 \varepsilon} + 6 e^{-2\mu_0 \varepsilon} = 1 \quad \mu_0 = 1 \text{ puisque } \varepsilon = \log 3$$

Au total : 7 symboles distincts (durée moyenne  $\frac{13}{7} \tau$ )

Remarque :

Ce troisième cas suppose que l'on identifie les symboles doublet par doublet, aucun doublet significatif ne pouvant commencer par le symbole  $a_1$

d) 1 symbole de durée minimale ( $a_1$ )

3 symboles de durée  $2 \tau$  :  $a_2 a_1, a_2 a_2, a_2 a_3$ .

9 symboles de durée  $3 \tau$  :  $a_3 a_1 a_1, a_3 a_1 a_2, \text{ etc.}$

$$Z = e^{-\mu_0 \varepsilon} + 3 e^{-2\mu_0 \varepsilon} + 9 e^{-3\mu_0 \varepsilon}, \quad \mu_0 = 1 \text{ puisque } \varepsilon = \log 3.$$

Au total, 13 symboles distincts (durée moyenne  $\frac{34}{13} \tau$ ).

L'identification se fait ainsi : au synchronisme, l'apparition de  $a_1$  indique un symbole de durée  $\tau$ , celle de  $a_2$  l'apparition d'un doublet, celle de  $a_3$  l'apparition d'un triplet.

3e Cas :

$v = 4$  8 possibilités.

a) 1 symbole/niveau : (cas trivial).

$$Z = 4 e^{-\mu_0 \varepsilon} = 1, \quad \varepsilon = 2 \log 2, \quad \mu_0 = 1 : 4 \text{ symboles.}$$

b) 3 symboles de durée  $\tau$ , 4 de durée  $2\tau$  ( $a_4 a_1$ ,  $a_4 a_2$ ,  $a_4 a_3$ ,  $a_4 a_4$ ).

$$Z = 3 e^{-\mu_0 \varepsilon} + 4 e^{-2\mu_0 \varepsilon} = 1, \quad \mu_0 = 1 : 7 \text{ symboles.}$$

c) 2 symboles de durée  $\tau$ , 8 de durée  $2\tau$  ( $a_3 a_1$ ,  $a_4 a_1$ ).

$$Z = 2 e^{-\mu_0 \varepsilon} + 8 e^{-2\mu_0 \varepsilon} = 1, \quad \mu_0 = 1 : 10 \text{ symboles.}$$

d) 1 symbole de durée  $\tau$ , 12 de durée  $2\tau$  ( $a_2 a_1$ ,  $a_3 a_1$ ,  $a_4 a_1$ ).

$$Z = e^{-\mu_0 \varepsilon} + 12 e^{-2\mu_0 \varepsilon} = 1, \quad \mu_0 = 1 : 13 \text{ symboles.}$$

e) 2 symboles de durée  $\tau$ , 4 de durée  $2\tau$  ( $a_2 a_1$ ), 16 de durée  $3\tau$  ( $a_4 a_1 a_j$ ).

$$Z = 2 e^{-\mu_0 \varepsilon} + 4 e^{-2\mu_0 \varepsilon} + 16 e^{-3\mu_0 \varepsilon} = 1, \quad \mu_0 = 1 : 22 \text{ symboles.}$$

f) 1 symbole de durée  $\tau$ , 8 de durée  $2\tau$  ( $a_2 a_1$ ,  $a_3 a_1$ ), 16 de durée  $3\tau$  ( $a_4 a_1 a_j$ ).

$$Z = e^{-\mu_0 \varepsilon} + 8 e^{-2\mu_0 \varepsilon} + 16 e^{-3\mu_0 \varepsilon} = 1, \quad \mu_0 = 1 : 25 \text{ symboles.}$$

g) 1 symbole de durée  $\tau$ , 4 de durée  $2\tau$  ( $a_2 a_1$ ), 32 de durée  $3\tau$  ( $a_3 a_j a_1$ ,  $a_4 a_1 a_j$ ).

$$Z = e^{-\mu_0 \varepsilon} + 4 e^{-2\mu_0 \varepsilon} + 32 e^{-3\mu_0 \varepsilon} = 1, \quad \mu_0 = 1 : 37 \text{ symboles.}$$

h) 1 symbole de durée  $\tau$ , 4 de durée  $2\tau$  ( $a_2 a_1$ ), 16 de durée  $3\tau$  ( $a_3 a_1 a_j$ ), 64 de durée  $4\tau$  ( $a_4 a_1 a_j a_k$ ).

$$Z = e^{-\mu_0 \varepsilon} + 4 e^{-2\mu_0 \varepsilon} + 16 e^{-3\mu_0 \varepsilon} + 64 e^{-4\mu_0 \varepsilon} = 1, \quad \mu_0 = 1 : 85 \text{ symboles.}$$

Tous les cas examinés ici correspondent à la puissance diacritique maximale possible ( $\mu_0 = 1$ ,  $C = w \log v$ ). Pour une valence  $v$  donnée, et lorsqu'il y a au moins un symbole de durée  $\tau$ , on obtient le nombre maximal de symboles possibles pour  $h = v - 1$ , le maximum est donc égal à

$$1 + v + v^2 + \dots + v^{v-1} = \frac{v^v - 1}{v - 1}.$$

### ANNEXE III

#### ADAPTATION DU MESSAGE AU SYSTEME LORSQUE LE NOMBRE DE SYMBOLES N'EST PAS IMPOSE

Soit un système (S) prenant  $a_1$  fois l'état 1,  $a_2$  fois l'état 2, ...,  $a_l$  fois l'état  $l$ , et tel que :

$$\left\{ \begin{array}{l} a_1 = N \\ a_1 \varepsilon_1 = E \end{array} \right.$$

$E$  est le travail sémiodynamique total dépensé dans une configuration à  $N$  états ;  $E$  est imposé, mais non pas  $N$ .

On cherche le maximum de  $P(N)$ , nombre de configurations possibles :

$$P(N) = \frac{N!}{a_1! a_2! \dots a_l!}$$

En employant la méthode du multiplicateur de Lagrange, on cherche le maximum de :

$$\log P - \mu \sum a_1 \varepsilon_1$$

En remplaçant  $\log P$  par son développement de Stirling, il vient

$$\delta N \log N - \sum \log a_i \delta a_i - \mu \sum \varepsilon_i \delta a_i = 0$$

ou encore

$$\sum \delta a_i \left[ \log \frac{N}{a_i} - \mu \varepsilon_i \right] = 0$$

D'où l'on tire :

$$p_i = \frac{a_i}{N} = e^{-\mu \varepsilon_i}$$

$$\sum p_i = \sum e^{-\mu \varepsilon_i} = 1.$$

On reconnaît l'équation (31) du chapitre II ( $\mu = \mu_0$ ).

L'adaptation optimale correspond donc à :

$$H = \sum p_i \log \frac{1}{p_i} = \mu_0 \sum p_i \varepsilon_i = \mu_0 \frac{E}{N} = C \frac{T}{N}$$

Ou encore :  $N H = C T$

(égalité entre les énergies diacritiques du système et du message).

ANNEXE IV

$$\text{MAXIMUM DE } H_1' = \sum_{ij} p_i \bar{\omega}_{ij} \log \frac{\sum_i p_i \bar{\omega}_{ij}}{p_i \bar{\omega}_{ij}}$$

Les  $\bar{\omega}_{ij}$  ne sont pas indépendants, car on a, d'après (69) :

$$\sum_j \bar{\omega}_{ij} = 1 \quad (l \text{ relations}).$$

On va employer la méthode des multiplicateurs de Lagrange, en formant :

$$X = H_1' - \lambda_1 \sum_j \bar{\omega}_{1j} - \lambda_2 \sum_j \bar{\omega}_{2j} \dots - \lambda_l \sum_j \bar{\omega}_{lj}$$

En dérivant par rapport aux  $\bar{\omega}_{ij}$ , on trouve :

$$\begin{aligned} \frac{\partial X}{\partial \bar{\omega}_{11}} &= p_1 \log \frac{\sum_i p_i \bar{\omega}_{i1}}{p_1 \bar{\omega}_{11}} - \lambda_1 \\ \frac{\partial X}{\partial \bar{\omega}_{12}} &= p_1 \log \frac{\sum_i p_i \bar{\omega}_{i2}}{p_1 \bar{\omega}_{12}} - \lambda_1 \\ &\dots \\ \frac{\partial X}{\partial \bar{\omega}_{1l}} &= p_1 \log \frac{\sum_i p_i \bar{\omega}_{il}}{p_1 \bar{\omega}_{1l}} - \lambda_1 \end{aligned}$$

D'où l'on déduit :

$$\frac{\bar{\omega}_{21}}{\bar{\omega}_{11}} = \frac{\bar{\omega}_{22}}{\bar{\omega}_{12}} = \frac{\bar{\omega}_{23}}{\bar{\omega}_{13}} = \dots = \frac{\bar{\omega}_{2l}}{\bar{\omega}_{1l}} = \frac{\bar{\omega}_{2j}}{\bar{\omega}_{1j}} = \frac{1}{1} = 1.$$

et les relations homologues pour les autres indices.

Donc :

$$\bar{\omega}_{ji} = \bar{\omega}_{ii} \text{ quel que soit } i.$$

Tous les  $\bar{\omega}_{ij}$  sont égaux à  $1/l$  ; dans ces conditions, on trouve :

$$(H_1')_{\max} = \sum p_i \log \frac{1}{p_i} = H_0$$

## ANNEXE V

### DETERMINATION DU GRADIENT SEMIODYNAMIQUE

Soit un système (S) susceptible de prendre  $l$  états de probabilités  $p_1, p_2, \dots, p_m$ , rangés par valeurs décroissantes. On dénombre ainsi  $\alpha_1$  états de probabilité  $p_1$ ,  $\alpha_2$  de probabilité  $p_2$ , ...  $\alpha_m$  de probabilité  $p_m$ . On a donc :

$$\left\{ \begin{array}{l} p_1 > p_2 > \dots > p_m \\ \alpha_1 p_1 + \alpha_2 p_2 + \dots + \alpha_m p_m = 1 \\ \alpha_1 + \alpha_2 + \dots + \alpha_m = l \text{ (nombre total d'états possibles).} \end{array} \right.$$

On est amené à associer à (S) un message composé à l'aide d'un alphabet ; on choisit donc  $\alpha_1$  symboles d'énergie sémiodynamique  $\varepsilon_1$ ,  $\alpha_2$  d'énergie  $\varepsilon_2, \dots, \alpha_m$  d'énergie  $\varepsilon_m$ , rangés par valeurs croissantes. Mais il y a des contraintes pratiques dans ce choix. Tout d'abord  $\varepsilon_1$  doit être assez grand pour permettre de définir le nombre d'états voulu, comme le montre l'annexe II ; ensuite les différences d'énergie entre symboles sont multiples d'un "quantum"  $\delta\varepsilon = w\tau \log v = \log v$ . (par définition de l'intervalle de temps élémentaire  $\tau$ ). Naturellement, on s'efforcera en général de choisir l'énergie minimale possible pour les symboles. On aura par exemple (ce choix n'est d'ailleurs pas restrictif) :

$$\left\{ \begin{array}{l} \varepsilon_1 < \varepsilon_2 < \dots < \varepsilon_m \\ \varepsilon_2 - \varepsilon_1 = \delta\varepsilon \\ \varepsilon_3 - \varepsilon_1 = 2 \delta\varepsilon \\ \varepsilon_m - \varepsilon_1 = (m - 1) \delta\varepsilon \end{array} \right.$$

On posera :

$$\varepsilon_1 = v \delta\varepsilon \quad (v \text{ entier } \geq 1).$$

En procédant de la sorte, il est clair que le travail sémiodynamique dépensé (par symbole) vaut :

$$U = \alpha_1 p_1 \varepsilon_1 + \alpha_2 p_2 \varepsilon_2 + \dots + \alpha_m p_m \varepsilon_m = \delta\varepsilon \left[ v + \alpha_2 p_2 + \dots + (m - 1) \alpha_m p_m \right]$$

alors que la pression diacritique de (S) est

$$H = \alpha_1 p_1 \log \frac{1}{p_1} + \alpha_2 p_2 \log \frac{1}{p_2} + \dots + \alpha_m p_m \log \frac{1}{p_m}$$

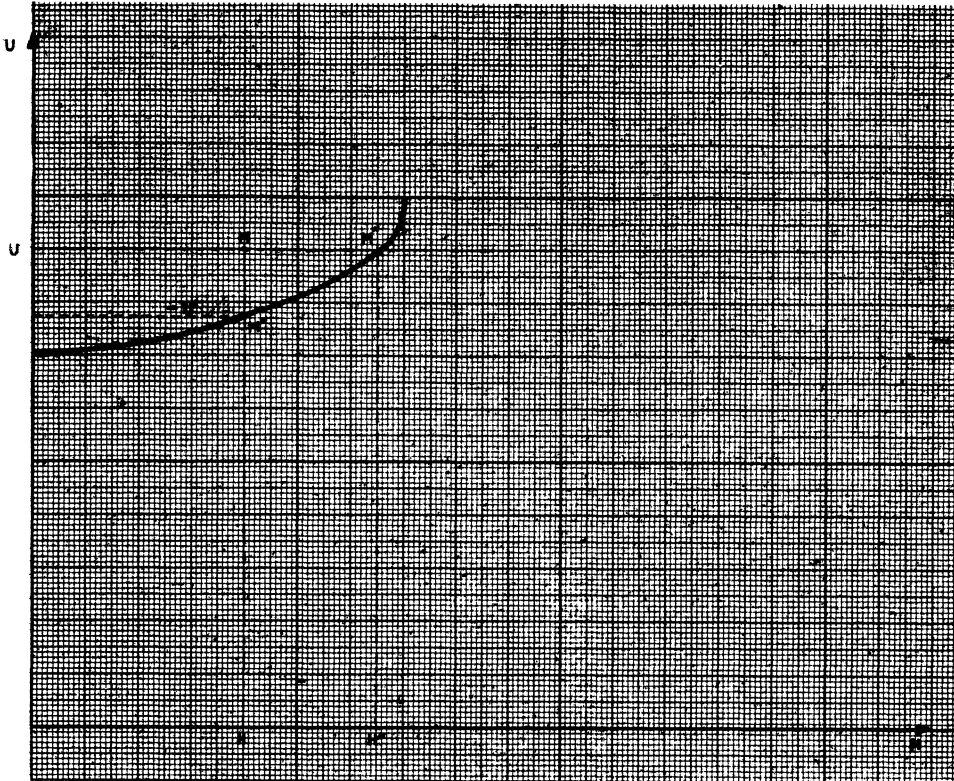


Figure 5 - Comparaison entre le travail sémiodynamique minimal et le travail dépensé

Si l'on considère dans le plan (H, U) la courbe ( $\Gamma$ ) caractéristique de l'alphabet choisi, il est clair que le point M (H, U) défini précédemment n'a aucune raison d'être situé sur cette courbe ; il est généralement situé au-dessus, et il ne peut être au-dessous, car sinon c'est que l'alphabet choisi est incompatible avec le système (S). L'alphabet permet en réalité de transmettre la pression diacritique (H) avec un travail  $U' < U$  (point M' de la figure 5), mais les probabilités associées  $p'_1, p'_2, \dots, p'_m$ , diffèrent en général de  $p_1, p_2, \dots, p_m$  ; on a seulement :

$$\sum \alpha_k p_k \log \frac{1}{p_k} = \sum \alpha_k p'_k \log \frac{1}{p'_k} = H.$$

De même, la pression diacritique maximale associée au travail dépensé U est  $H'' > H$  (point M'' de la figure 4), mais là encore les probabilités  $p''_1, p''_2, \dots, p''_m$ , sont différentes de  $p_1, p_2, \dots, p_m$  ; on a seulement :

$$\sum \alpha_k p_k \varepsilon_k = \sum \alpha_k p''_k \varepsilon_k = U.$$

Il s'agit donc de vérifier si le point M du plan (H, U) s'écarte notablement de M' et M'' avec lesquels il se confond lorsque l'optimum est réalisé.



Cette évaluation s'effectue commodément en utilisant la fonction d'états  $Z$  et le gradient sémiodynamique  $1/\mu$

Soit, en effet :

$$Z = \alpha_1 e^{-\mu \epsilon_1} + \alpha_2 e^{-\mu \epsilon_2} \dots + \alpha_m e^{-\mu \epsilon_m} = e^{-\mu \epsilon} [\alpha_1 + \alpha_2 e^{-\mu \delta \epsilon} + \dots + \alpha_m e^{-(m-1)\mu \delta \epsilon}]$$

Posons :

$$\lambda = e^{-\mu \delta \epsilon}, \text{ on a :}$$

$$Z = e^{-\mu \epsilon_1} f(\lambda) = \lambda^v f(\lambda)$$

avec :

$$f(\lambda) = \alpha_1 + \alpha_2 \lambda + \alpha_3 \lambda^2 + \dots + \alpha_m \lambda^{m-1}$$

$f(\lambda)$  est le polynôme caractéristique de l'alphabet.

Tout point courant  $(h, u)$  de la courbe caractéristique  $(\Gamma)$  satisfait à :

$$u = -\frac{\partial}{\partial \mu} \log Z = \delta \epsilon \left[ v + \lambda \frac{d}{d\lambda} (\log f) \right] = \delta \epsilon \left[ v + \lambda \frac{f'}{f} \right]$$

$$h = \log f + \frac{f'}{f} \lambda \log \frac{1}{\lambda}$$

Tous les autres paramètres liés à la courbe  $(\Gamma)$  s'en déduisent :

$$\frac{1}{\mu} = \frac{\delta \epsilon}{\log \frac{1}{\lambda}}, \mu_0 = \delta \epsilon \left[ v - \frac{\log f}{\log \frac{1}{\lambda}} \right] = -\frac{\delta \epsilon}{\log \frac{1}{\lambda}} \log (\lambda^v f)$$

et enfin :

$$\left\{ \begin{array}{l} p_1 = \frac{1}{f(\lambda)} \\ p_2 = \frac{\lambda}{f(\lambda)} = \lambda p_1 \\ \dots \\ p_m = \frac{\lambda^{m-1}}{f(\lambda)} = \lambda^{m-1} p_1 \end{array} \right.$$

Nous avons donc substitué au gradient sémiodynamique  $1/\mu$  le paramètre  $\lambda$ , plus commode : toutes les grandeurs sont évaluées en fonction de  $\lambda$  et du polynôme  $f(\lambda)$  ou de leurs logarithmes. ( $\lambda$  varie de 0 à 1 lorsque  $1/\mu$  varie de 0 à  $\infty$ ).

Il est maintenant aisé de résoudre les problèmes posés.

1/ Vérification de la position de M (H, U) en dehors de la courbe (Γ).

Si le point M (H, U) se trouve sur la courbe (Γ), les  $p_1$  vérifient les relations qui viennent d'être établies. Dans ces conditions,  $\lambda$  a pour valeur  $\lambda_1$ , racine de  $f(\lambda) - 1/p_1 = 0$ .

(A noter que le polynôme  $f(\lambda)$  a bien une racine unique  $\lambda_1$  comprise entre 0 et 1, car  $f(0) - \frac{1}{p_1} = \alpha_1 - \frac{1}{p_1} = \frac{1}{p_1} (\alpha_1 p_1 - 1) < 0$ ,  $f(1) - \frac{1}{p_1} = 1 - \frac{1}{p_1} > 0$  puisque  $p_1$  est par hypothèse supérieur à la moyenne  $\frac{1}{l}$  des  $p_i$ )

On doit alors trouver  $p_2 = \lambda_1 p_1, \dots, p_m = \lambda_1^{m-1} p_1$ .

En général, ces relations ne sont pas vérifiées ; toutefois, elles le sont s'il n'y a que deux probabilités au total ( $m = 2$ ) ; on a évidemment

$$\alpha_1 p_1 + \alpha_2 p_2 = \frac{1}{f(\lambda)} \left[ \alpha_1 + \alpha_2 \lambda \right] = 1, \text{ car choisir } p_1 \text{ c'est choisir } H.$$

Lorsqu'on a  $p_k = \lambda_1^{k-1} p_1$ , la différence  $p_k - \lambda_1^{k-1} p_1$  (avec  $f(\lambda_1) = 1/p_1$ ) constitue une mesure de l'écart par rapport à l'optimum ; cet écart est certainement faible lorsque  $p_1$  est très supérieur à  $p_2, p_3, \dots, p_m$ , car l'essentiel du travail sémiodynamique est constitué par le premier terme dans la somme  $\sum \alpha_i p_i \varepsilon_i$ .

Géométriquement, le point de la courbe (Γ) auquel on compare M (H, U) est le point de contact de la tangente à (Γ) passant par  $\log \frac{1}{p_1}, \varepsilon_1$ .

2/ Ecart par rapport au travail minimal et à la pression maximale.

Ces évaluations sont immédiates en se référant à la figure 4. Le point (M') a pour abscisse

$$H = \sum \alpha_k p_k \log \frac{1}{p_k}$$

On cherche donc  $\mu$  (ou plutôt son logarithme  $\lambda$ ) en évaluant la racine  $\lambda'_1$  de l'équation :

$$\log (f(\lambda)) + \lambda \log \frac{1}{\lambda} \frac{f'(\lambda)}{f(\lambda)} = H = \sum \alpha_k p_k \log \frac{1}{p_k}$$

Si  $\lambda'_1$  est cette racine, le travail minimal correspondant à H vaut :

$$U' = U_{\min} = \delta \varepsilon \left[ v + \lambda'_1 \frac{f'(\lambda'_1)}{f(\lambda'_1)} \right]$$

L'écart recherché est  $U - U'$ .

Le point M' de la figure 4 a pour ordonnée U ; il correspond à la racine  $\lambda'_1$  de l'équation :

$$\delta \varepsilon \left[ v + \lambda \frac{f'(\lambda)}{f(\lambda)} \right] = U = \delta \varepsilon \left[ v + \alpha_2 p_2 + 2\alpha_3 p_3 + \dots + (m-1) \alpha_m p_m \right]$$

qui s'écrit encore, sous forme polynomiale :

$$\lambda f'(\lambda) - \left[ \alpha_2 p_2 + \dots + (m-1) \alpha_m p_m \right] f(\lambda) = 0$$

De la racine  $\lambda''$  on déduit :

$$H'' = H_{\max} = \log f(\lambda_1'') + \lambda_1'' \log \frac{1}{\lambda_1''} \frac{f'(\lambda_1'')}{f(\lambda_1'')}$$

$H'' - H$  mesure l'écart recherché.

#### BIBLIOGRAPHIE

- [1] R.V.L. HARTLEY - The Transmission of Information ; (Bell System Techn. Journal, 3, 1928, p. 535)
- [2] C.E. SHANNON - The Mathematical Theory of Communication (Bell System Techn. Journal, 3, 1948, p. 379 et 4, 1948, p. 623)
- [3] L. BRILLOUIN - La Science et la Théorie de L'Information (Masson Editeur, 1959)
- [4] E. SCHRODINGER - Statistical Thermodynamics (Cambridge - University Press, 1948)

NOTA. Ce dernier livre, qui résume les cours professés par E Schrodinger à Dublin pendant la dernière guerre, est un chef d'oeuvre de clarté et d'élégance dans son exposition. L'auteur lui doit l'essentiel des raisonnements qui figurent dans le présent travail.

POST-SCRIPTUM. Le présent travail était rédigé et en cours d'impression lorsque l'auteur a eu connaissance d'une publication de B. MANDELBROT "Contribution à la théorie mathématique des jeux de communication" (Thèse Paris 1952) où la même analogie avec la Thermodynamique Statistique était mise en évidence. Il est de stricte justice d'adjoindre cette référence à la liste donnée ci-dessus, car elle comporte sous une autre forme un bon nombre de résultats et formules du présent article.