

COLLOQUE NATIONAL SUR LE TRAITEMENT DU SIGNAL ET SES APPLICATIONS

88/1



NICE du 26 au 30 AVRIL 1977

Un exemple d'application du décodage de Viterbi

An application of Viterbi decoding

Jean-Claude BIC

Jean-Claude IMBEAUX

Centre National d'Etudes des Télécommunications
38-40, Avenue du Général Leclerc 92131 Issy-Les-Moulineaux

RESUME

SUMMARY

L'utilisation du décodage suivant l'algorithme de Viterbi d'un code récurrent est de plus en plus fréquemment envisagée pour les transmissions numériques par satellite.

Les contraintes de largeur de bande et de débit imposent de choisir des codes ayant un rendement relativement élevé $2/3$ ou $3/4$. Le paramètre "longueur de contrainte" qui mesure la complexité du décodeur est limité pour ces rendements à une valeur de 3 ou 4.

Un algorithme donnant les paramètres nécessaires à l'estimation des performances d'un code a été établi ce qui a permis la recherche de "bons" codes répondant aux conditions précédentes.

Les problèmes posés par l'introduction du système codeur-décodeur dans une chaîne de transmission utilisant une modulation numérique par déplacement de phase sont ensuite évoqués, ainsi que les performances escomptées pour le code choisi.

Viterbi decoding for convolutional codes is widely considered for digital satellite transmission.

Bandwidth constraints imply choosing codes with a comparatively high rate $2/3$ or $3/4$ for example. For these rates, "constraint length" which defines decoder complexity is limited to about 3 or 4.

An algorithm giving the parameters needed to estimate the code performances has been designed allowing the research of "good" codes satisfying the aforementioned conditions.

Problems arising from insertion of coding decoding system in a transmission chain using PSK modulation are then considered, and at last expected performances are given for the selected code.



Un exemple d'application du décodage de Viterbi

An application of Viterbi decoding

INTRODUCTION

Un canal de transmission numérique par satellite présente avec une bonne approximation les caractéristiques d'un canal à bruit additif gaussien. Les erreurs apparaissant après une démodulation de phase numérique cohérente sont pratiquement indépendantes. On peut alors attendre de l'utilisation d'une procédure de décodage correcteur d'erreurs aléatoires telle que l'algorithme de Viterbi une sensible amélioration des performances.

La conception du système codeur-décodeur doit faire entrer en ligne de compte les contraintes imposées par la transmission. Pour l'application envisagée la largeur de bande limitée conduit à choisir un code de rendement 2/3 ou 3/4. La levée d'ambiguïté de phase à la réception propre à la démodulation cohérente est facilitée par le choix d'une certaine catégorie de codes dits transparents.

Après quelques rappels sur l'algorithme de Viterbi et ses performances, les problèmes de sélection de codes et d'introduction du codage décodage dans la chaîne de transmission sont examinés.

I QUELQUES RAPPELS SUR L'ALGORITHME DE VITERBI

L'algorithme de Viterbi est une procédure puissante de décodage d'un code récurrent correcteur d'erreurs aléatoires, fondée sur le principe de maximum de vraisemblance (1).

Le code est défini par son rendement $R = k_0/n_0$ et ses $k_0 \times n_0$ polynômes générateurs $g_{ij}(D)$ de degré inférieur ou égal à m . On constitue un treillis formé de $2^{k_0(m-1)}$ noeuds représentatifs des états des registres du codeur, reliés entre eux par des branches représentatives des transitions entre ces états. A chaque branche sont associés un k_0 -uplet à coder et un n_0 -uplet codé. L'opération de codage se traduit par la construction d'un chemin dans ce treillis.

L'opération de décodage consiste à retrouver ce chemin parmi tous les chemins possibles.

L'hypothèse fondamentale supposant le canal de transmission sans mémoire permet de décomposer l'opération de façon à progresser pas à pas sans pour autant renoncer au principe du maximum de vraisemblance (1).

A la réception, on calcule pour chaque branche du treillis une métrique de branche

$$\lambda_t(i, j) = f[\log P_k(\lambda_t, b(i, j))] \text{ avec}$$

- $t = \ell T$, T période de transmission d'un n_0 -uplet
- λ_t le n_0 -uplet reçu
- $b(i, j)$ le n_0 -uplet associé à la branche (i, j) joignant le noeud i au noeud j .

On calcule ensuite une métrique de chemin ou de noeud.

$$\mu_t(i, j) = \text{Min}_{i \in S(j)} (\mu_{t-T}(i) + \lambda_t(i, j)) \quad j = 1, \dots, 2^{k_0(m-1)}$$

où $S(j)$ est l'ensemble des 2^{k_0} noeuds antécédents du noeud j .

On sélectionne ainsi à chaque instant t une branche arrivant à chacun des noeuds et donc un chemin. On retient ainsi pour chaque noeud le chemin formé de N k_0 -uplets de métrique minimale y aboutissant. Il ne restera plus à la fin de la transmission qu'à

sélectionner le noeud J du chemin tel que $\mu_N(J) = \text{Min}_j \mu_N(j)$

On aura au total effectué $N \times 2^{k_0(m-1)} \times 2^{k_0}$ comparaisons au lieu des 2^{Nn_0} calculs nécessaires dans une application directe du principe de maximum de vraisemblance. Le nombre d'opérations croît linéairement en fonction de N et non plus exponentiellement.

Les cas habituels de métrique de branche sont (2) :

- la métrique de Hamming correspondant à une démodulation fournissant des 0 et des 1

$$\lambda_t(i, j) = \sum_{h=1}^{n_0} \lambda_t^{(h)} \otimes b^{(h)}(i, j)$$

- la métrique quantifiée correspondant à une démodulation quantifiée (soft-décision) à 2^q niveaux.

$$\lambda_t(i, j) = \sum_{h=1}^{n_0} \lambda_t^{(h)}(i, j) \quad \text{avec}$$

$$\lambda_t^{(h)}(i, j) = \lambda_t^{(h)} \text{ si } b^{(h)}(i, j) = 0; \lambda_t^{(h)}, \lambda_t^{(h)} \in \{0, 1, \dots, 2^q - 1\}$$

$$\lambda_t^{(h)}(i, j) = (2^q - 1) - \lambda_t^{(h)} \text{ si } b^{(h)}(i, j) = 1$$

II ESTIMATION DES PERFORMANCES

Les transitions entre états du codeur sont représentées cette fois sous forme d'un graphe dont les 2 états "tout 0" constituent les noeuds de départ et d'arrivée. On affecte alors à chaque branche (?)

- 1) une variable de distance D dont l'exposant i est égal au poids de Hamming du n_0 -uplet codé
- 2) une variable de poids N dont l'exposant j est égal au poids du k_0 -uplet à coder

La fonction de transfert s'écrit alors

$$T(D, N) = \sum_{i,j} \sum_{i,j} \lambda_{ij} D^i N^j$$

Le plus petit indice i pour lequel il existe j tel que le coefficient λ_{ij} est non nul est appelé distance libre d .

On démontre les majorations suivantes de la probabilité d'erreur sur les bits

a) Canal binaire symétrique (3)

$$P_b < \frac{\Gamma_{d_0}}{2k_0} \frac{d}{dN} \left\{ T(D, N) + T(-D, N) + D [T(D, N) - T(-D, N)] \right\}_{N=1, D=2\sqrt{P}}$$

$$d_0 = \begin{cases} d/2 & \text{si } d \text{ pair} \\ d+1/2 & \text{si } d \text{ impair} \end{cases} \quad \Gamma_{d_0} = C_{2d_0-1}^{d_0} \cdot 2^{-2d_0}$$

soit encore

$$P_b < \frac{\Gamma_{d_0}}{2k_0} \left\{ D^d \cdot \epsilon(d) \sum_{x_j \in E_d} x_j + \sum_{i=d+1}^{\infty} D^i \cdot \epsilon(i) \left[\sum_{x_j \in E_{i-1}} x_j + \sum_{x_j \in E_i} x_j \right] \right\}$$

ou E_i est l'ensemble des chemins de distance i et x_j le poids d'un chemin x_j et $\epsilon(i) = 1 + (-1)^i$

b) Canal à bruit additif gaussien infiniment quantifié (1)

$$P_b < \frac{1}{2} \cdot \text{erfc} \left(\frac{d E_b}{N_0} \right) \cdot e^{d E_b / N_0} \left. \frac{dT(D, N)}{dN} \right|_{N=1, D=e^{-\frac{E_b}{N_0}}}$$

$$P_b < \frac{1}{2} \cdot \text{erfc} \left(\frac{d E_b}{N_0} \right) \cdot e^{d E_b / N_0} \left[\sum_{i=d}^{\infty} \left(\sum_{x_j \in E_i} x_j \right) D^i \right]_{D=e^{-\frac{E_b}{N_0}}}$$

où $\frac{E_b}{N_0}$ est le rapport énergie par bit à densité monolatérale de bruit.



Ces approximations mettent en évidence le rôle de la distance libre d puisque asymptotiquement c'est le premier terme des développements qui sera prépondérant. La connaissance de la fonction de transfert et de la distance d est importante et fait l'objet du paragraphe suivant.

III ALGORITHME DE RECHERCHE DES K PLUS COURTS CHEMINS D'UN CODE RECURRENT

On utilise une extension de l'algorithme de recherche du plus court chemin dans un graphe. (4)

Soit K le nombre de plus courts chemins entre le noeud de départ "tout 0" et le noeud d'arrivée "tout 0" que l'on désire obtenir.

A chaque noeud i on affecte un vecteur de distance

$$\vec{\pi}_i = (\pi_i^{L_1})_{L_1=1, K} \text{ avec } \pi_i^{L_1} \leq \pi_i^{L_2} \text{ si } L_1 < L_2$$

A chaque branche (i, j) on associe le poids de Hamming d_{ij} du n_0 -uplet.

Les vecteurs $\vec{\pi}_i$ sont initialisés à

• d_{0i} pour les noeuds ayant le noeud "tout 0" origine pour antécédent

• $+\infty$ pour les autres.

On définit deux opérations.:

- une opération scalaire $*$: $\vec{\pi}_i * d_{ij} = (\pi_i^{L_1} + d_{ij})$
- une opération interne \perp : $\vec{\pi}_i \perp \vec{\pi}_j = \vec{\pi}_k$

$\vec{\pi}_k$ est formé des K plus petites composantes de l'ensemble des composantes de $\vec{\pi}_i$ et $\vec{\pi}_j$.

L'algorithme procède alors comme suit : soit I un des $2^{k_0(m-1)}$ noeuds et J un des 2^{k_0} noeuds de l'ensemble $S(I)$. On effectue pour chaque noeud I une suite de 2^{k_0} opérations correspondant à chacun des noeuds J . Chaque opération se décompose de la façon suivante :

$$\begin{aligned} \vec{\pi}_J * d_{JI} &= \vec{\pi}_I' \\ \vec{\pi}_I \perp \vec{\pi}_I' &= \vec{\pi}_I'' \\ \vec{\pi}_I &= \vec{\pi}_I'' \end{aligned}$$

Lorsqu'on a terminé ces opérations pour tous les noeuds I on recommence le même processus. L'algorithme est arrêté lorsque tous les vecteurs $\vec{\pi}_I$ restent invariants.

L'un des problèmes posés par l'utilisation de cette procédure est le risque de prendre en compte plusieurs fois le même chemin. Si deux composantes d'un vecteur de distance sont différentes, elles correspondent à 2 chemins différents. Si elles sont égales, elles peuvent correspondre au même chemin. On compare alors les chemins ayant même distance ce qui nécessite la mise en mémoire des chemins de distance donnée (chaque chemin est repéré par la valeur binaire de la séquence d'éléments binaires dont il est issu par codage).

A chaque étape le vecteur $\vec{\pi}_i$ associé au noeud i a pour composantes K distances de chemins issus de l'origine aboutissant à ce noeud classées dans un ordre croissant. A chaque étape on remplace s'il y a lieu les distances les plus élevées par d'autres plus faibles. Lorsqu'il n'y a plus de changement d'une étape à l'autre, c'est bien que l'on a obtenu les K plus courts chemins issus de l'origine et aboutissant à un noeud quelconque. Pour obtenir les paramètres de

la fonction de transfert, il suffira de considérer le noeud "tout 0" extrémité.

On notera que la complexité du calcul croît en K^2 .

Exemple 1

code 2/3 $m = 3$

$$\begin{aligned} g_{11}(D) &= 1 & g_{21}(D) &= 1 + D^2 & g_{31}(D) &= D + D^2 \\ g_{12}(D) &= 1 + D^2 & g_{22}(D) &= D & g_{32}(D) &= 1 + D + D^2 \\ T(D, N) &= D^5 (N + 2N^2) + D^6 (N + 3N^3 + 3N^4 + 2N^5 + N^6 + N^7) + \dots \end{aligned}$$

Exemple 2

code 2/3 $m = 4$

$$\begin{aligned} g_{11}(D) &= D^2 & g_{21}(D) &= 1 + D^2 + D^3 & g_{31}(D) &= 1 + D^2 + D^3 \\ g_{12}(D) &= 1 + D + D^3 & g_{22}(D) &= D + D^2 & g_{32}(D) &= 1 + D + D^3 \\ T(D, N) &= D^7 (N + 3N^2 + 2N^3 + N^4 + 2N^5 + 2N^6 + 2N^7 + N^8 + 2N^9 + N^{10}) + D^8 (N + 6N^3 + 11N^4 + 6N^5 + 5N^6 + 7N^7 + 4N^8 + 3N^9 + 4N^{10} + 2N^{11} + 2N^{12} + N^{15} + N^{16}) + \dots \end{aligned}$$

IV RECHERCHES DES "BONS" CODES (AU SENS DE LA DISTANCE LIBRE)

Les meilleurs codes de rendement 1/2 sont bien connus (5) mais il n'en est pas de même pour les codes de rendement plus élevé. (6)

Avant d'effectuer la recherche, notons l'existence de limites sur la distance libre obtenues en établissant une analogie avec les codes en blocs linéaires.

pour un code non systématique

$$d_{ns} \leq \text{Min}_{L \geq 1} \left\{ \frac{n_0}{2} \cdot \frac{2^{kL}}{2^{kL} - 1} (L + m - 1) \right\}$$

pour un code systématique

$$d_s \leq \text{Min}_{L \geq 1} \left\{ \frac{2^{kL}}{2(2^{kL} - 1)} [n_0(L + m - 1) - h_0(m - 1)] \right\}$$

Notons que les codes non systématiques fourniront des performances supérieures à celles d'un code systématique.

Nous devons également éliminer les codes dits catastrophiques (1) qui peuvent entraîner une propagation catastrophique d'erreurs. Deux types de critère peuvent être utilisés, l'un s'appliquant aux polynômes générateurs (7), l'autre au graphe (1).

Une recherche systématique est alors entreprise. Compte tenu de l'ordre de complexité croissant en $2^{k_0(m-1)}$ on s'est limité à des valeurs de m de 3 et 4. Les polynômes générateurs $g_{ij}(D)$ constituant un code sont rangés suivant leur valeur octale. On vérifie le caractère non catastrophique. On calcule la distance libre du code puis d'autres termes de la fonction de transfert pour les codes ayant une distance optimale. On constate qu'il n'existe pas de codes (2,3,3) non catastrophiques de distance 6 ni de codes (2,3,4) non catastrophiques de distance 8. Les meilleurs codes ont respectivement pour distance 5 et 7.

V INTRODUCTION DU SYSTÈME CODEUR DECODEUR DANS LA CHAÎNE DE TRANSMISSION

La modulation la plus employée pour les transmissions numériques par satellite est la modulation par déplacement de phase à 2 ou 4 états. L'hypothèse d'indépendance des erreurs prescrit l'usage d'une démodulation différentielle. La démodulation de phase cohérente



pose le problème de levée d'ambiguïté. La solution consistant à envoyer périodiquement des mots de reconnaissance s'oppose au caractère continu propre à la transmission par codage récurrent.

La solution habituelle est alors l'utilisation d'un codage par transition. Mais on sait alors que le décodage par transitions entraîne des erreurs corrélées. Pour ne pas subir une trop sévère dégradation des performances, il faut s'imposer de placer le système codeur-décodeur par transition à l'extérieur du système codeur-décodeur convolutionnel.

On sait en effet obtenir des codes dits transparents insensibles à un changement de phase de π sur le message reçu c'est-à-dire à une inversion du message (si le message codé reçu est le complément du message codé émis, le message après décodage est le complément du message avant codage). Le caractère linéaire du code fait que la condition de transparence s'exprime simplement par

$$\sum_{j=1}^k \oplus \sum_{k=0}^{m-1} \oplus g_k(i, j) = 1 \quad \text{pour } i = 1, \dots, n_0$$

avec :

$$g_{ij}(D) = \sum_{k=0}^{m-1} g_k(i, j) D^k$$

La figure 1 représente le diagramme d'une liaison effectuée avec un code transparent et une modulation MDP2. Avec une modulation MDP4, la solution la plus simple consiste à coder séparément chacun des 2 trains à l'aide du même code transparent et d'utiliser deux décodeurs identiques ce qui est illustré à la figure 2.

VI CHOIX D'UN CODE POUR LA REALISATION PRATIQUE

La recherche et les considérations précédentes nous ont conduit à choisir le code transparent défini à l'exemple 1. La figure 3 représente le codeur correspondant.

La figure 4 donne pour ce code le taux d'erreur sur les éléments binaires après décodage en fonction du rapport E_b/N_0 où E_b est l'énergie par élément binaire d'information et N_0 la densité de bruit.

L'amélioration que l'on peut obtenir par rapport à une transmission non codée effectuée en MDP2 est à 10^{-5} d'environ 3 dB pour une démodulation non quantifiée (hard-décision) et de 4,5 dB pour une démodulation quantifiée à 8 niveaux.

CONCLUSION

L'amélioration des performances obtenue grâce à l'algorithme de Viterbi semble intéressante en regard de la complexité relativement faible du dispositif de codage/décodage. Il a été constaté par ailleurs (8) que l'application sur des canaux réels donnait des résultats proches de ce qui était attendu. Le décodage suivant l'algorithme de Viterbi semble promis à un bel avenir.

BIBLIOGRAPHIE

- (1) A. J. VITERBI : "Convolutional codes and their performances in communications systems "

IEEE Trans Vol COM - 19 , Octobre 1971
- (2) J. A. HELLER, I. M. JACOBS "Viterbi decoding for satellite and space communications"

id.
- (3) L. VAN DE MEEBERG " A tightened upper bound on the error probability of binary convolutional codes with Viterbi decoding"

IEEE Trans on Information theory Mai 1974
- (4) E. MINIEKA " On computing sets of shortest paths in a graph "

Communications of the ACM Vol 17 N°6 Juin 1974
- (5) K. J. LARSEN " Short convolutional codes with maximal free distances for rates 1/2, 1/3, 1/4 "

IEEE Trans Vol IT - 19 Mai 1973
- (6) E. PAASKE "Short binary convolutional codes with maximal free distances for rates 2/3 and 3/4 "

IEEE Trans Vol IT-20 Septembre 1974
- (7) J. L. MASSEY M. K. SAIN "Inverses of linear sequential circuits"

IEEE Trans Vol C - 17 Avril 1968
- (8) I. R. JACOBS "Practical Applications of coding"

IEEE Trans Vol IT - 20 N°3 Mai 1974

Un exemple d'application du décodage de Viterbi

An application of Viterbi decoding

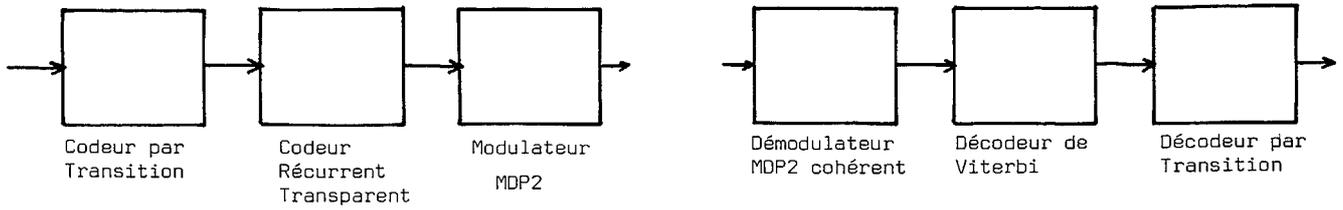


Fig 1 Schéma de la liaison en MDP2

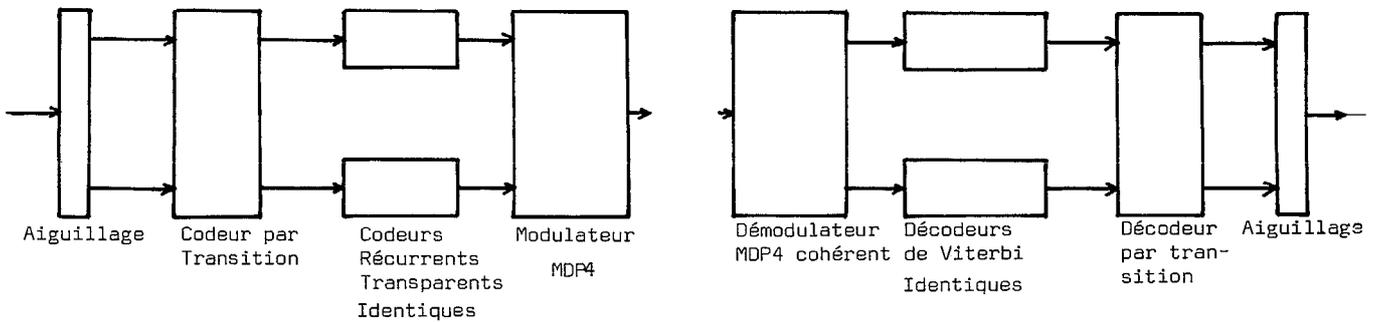


Fig 2 Schéma de la liaison en MDP4

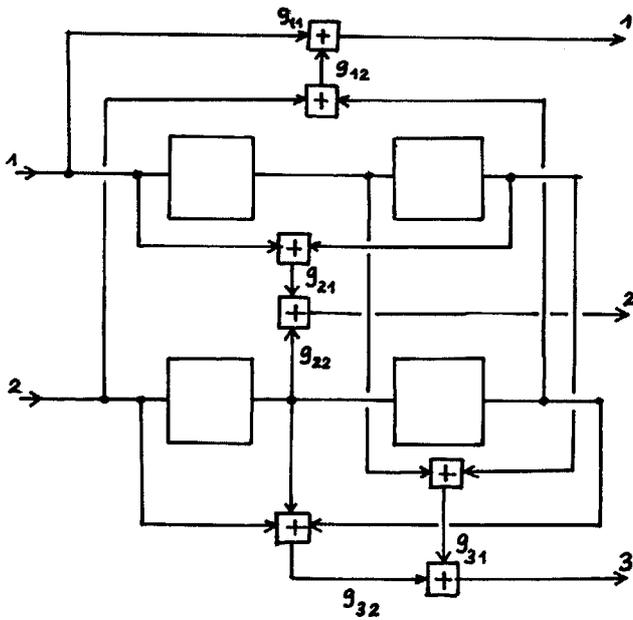


Fig 3 Schéma du codeur $R = 2/3$ $m = 3$ Transparent



Registre à décalage



Additionneur Modulo 2

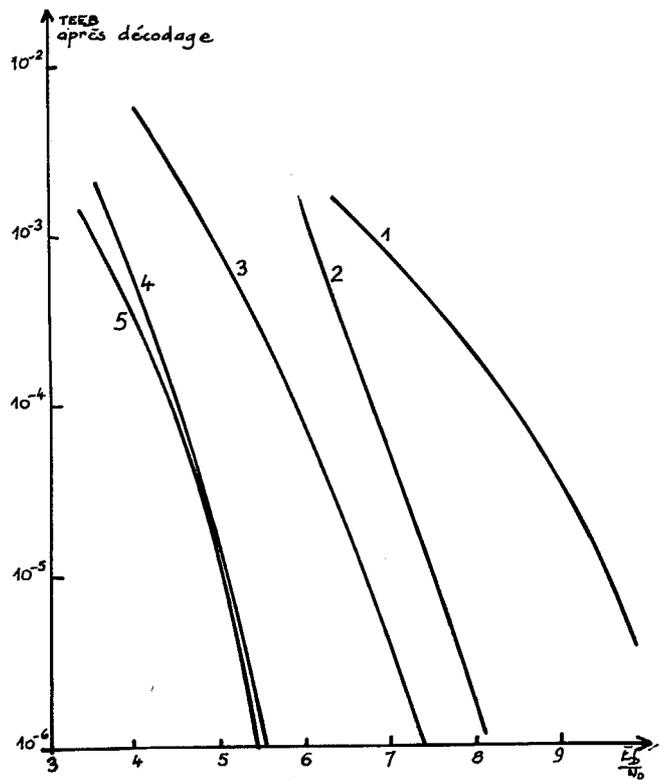


Fig 4 Performances du code $R = 2/3$ $m = 3$

- 1 - Modulation MDP2 sans codage
- 2 - Canal binaire symétrique approximation
- 3 - Canal binaire symétrique simulation
- 4 - Canal infiniment quantifié approximation
- 5 - Canal quantifié à 8 niveaux simulation

