

# COLLOQUE NATIONAL SUR LE TRAITEMENT DU SIGNAL ET SES APPLICATIONS

NICE du 16 au 21 JUIN 75



## LA REDUCTION DE LA REDONDANCE DANS LES TELEMESURES

T.A. HAWKES<sup>\*</sup>, B. COLIN<sup>\*</sup>, P. ANGLADE<sup>\*\*</sup>.

<sup>\*</sup> THOMSON-CSF, Division AVS, 52 rue Guynemer, 92132 ISSY-LES-MOULINEAUX.  
<sup>\*\*</sup> S.E.C.T., Fort de Montrouge, 94110 ARCUEIL.

### RESUME

Pour la réduction de la redondance dans les télémesures des recherches ont montré l'intérêt du prédicteur d'ordre zéro par la simplicité de sa mise en oeuvre et ses performances relativement bonnes. On décrit une méthode de codage qui améliore l'efficacité de ce prédicteur.

On remplace les mots habituels d'amplitude et d'identification à longueurs fixes par des codes à longueurs variables qui donnent la valeur de la différence et de la longueur de l'intervalle entre échantillons non redondants. Ces codes sont très simples, étant de la forme suivante pour valeurs croissantes :

1  
01  
001  
0001  
etc...

Au-delà d'une certaine valeur, on emploie des codes à longueurs fixes.

On donne les résultats de quelques essais sur bandes de télémesures qui montrent une amélioration de deux environ par rapport au prédicteur aux codages fixes. Les essais montrent aussi que l'emploi de codes à longueurs variables pour la longueur des intervalles réduit la probabilité de débordement de la mémoire tampon dans une liaison de transmission.

### SUMMARY

Various investigators have shown that the zero order predictor can be a relatively efficient computational algorithm for redundancy reduction in telemetry data. In this paper we describe a coding method for use with that predictor which permits a substantial improvement in performance.

The usual fixed length words representing amplitude and identification are replaced by variable length codes which give the values of the differences and the interval lengths between non redundant samples. The codes used are very simple and have the following form for increasing values :

1  
01  
001  
0001  
etc...

Beyond a certain limit fixed length codes are used, but these correspond to infrequently occurring values. An additional bit is required to indicate coding mode.

Results of a limited number of tests on telemetry recordings indicate an improvement factor of about two with respect to the predictor using fixed length coding. Tests also show that the use of variable length interval codes reduces the probability of buffer overflow in a transmission link.



1 - INTRODUCTION

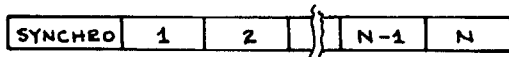
Il est très naturel que l'augmentation dans la transmission et le traitement de données numériques, qui s'est produit pendant ces dernières années, soit accompagnée d'efforts d'effectuer les codages avec un meilleur rendement. Le domaine des télémesures, du point de vue du stockage des informations aussi bien que leur transmission, n'est pas une exception.

Ces codages ont l'objet de réduire la redondance dans les données, permettant donc une réduction dans les moyens de stockage, dans la largeur de bande de transmission, ou dans les puissances d'émission. Malgré des efforts considérables, les réalisations ne sont pas très nombreuses, et les recherches continuent pour des algorithmes de codage plus efficaces.

On décrit ici une application nouvelle de méthodes de codage déjà connues, et l'on donne quelques résultats de tests qui montrent les améliorations obtenues.

2 - ORGANISATION DES TRANSMISSIONS

Lorsque l'on transmet des télémesures, il s'agit de la transmission d'un ensemble de variables multiplexées dans le temps. La transmission est organisée en format (figure 1), où l'on trouve, par exemple, un mot codé de synchronisation en tête suivi des mots d'informations correspondant aux amplitudes des variables échantillonnées. Après l'émission du



CYCLE DE TRANSMISSIONS

FIGURE 1

N<sup>i</sup>ème mot, on recommence le cycle avec le mot de synchronisation. La durée d'un cycle correspond à la période d'échantillonnage de chaque variable.

L'ordre de la transmission des échantillons est immuable, ce qui permet d'identifier une variable à la réception, par simple comptage à partir du code de synchronisation. Lorsque l'on effectue un procédé de réduction de la redondance dans une liaison de transmission, certains mots dans le cycle disparaissent et le démultiplexage ne peut plus se faire par comptage ; il faut ajouter, à chaque échantillon non redondant, des bits d'identification de la variable échantillonnée.

3 - ALGORITHMES DEJA UTILISES

Dans la littérature [(1) à (5)], on trouve un grand nombre d'algorithmes que l'on a déjà essayés.

Parmi eux, on peut identifier les classes principales suivantes :

1° Prédicteurs

A partir d'un ou plusieurs échantillons précédents, on fait une prédiction de la valeur du prochain échantillon. Si la valeur prédite est à l'intérieur d'une fourchette autour de la valeur réelle, on considère l'échantillon redondant et l'on ne le transmet pas. Lorsque l'on trouve un échantillon redondant, on continue la prédiction. Aussi longtemps que les valeurs prédites se trouvent à l'intérieur de la fourchette autour des valeurs réelles, on ne transmet rien. Lorsque l'on trouve un échantillon non redondant, en dehors de la fourchette, on transmet sa valeur accompagnée de bits d'identification.

La prédiction peut être plus ou moins complexe. Les équations pour les prédicteurs polynomiaux se basent sur le calcul aux différences finies et permettent de passer un polynôme d'ordre n par n + 1 pointes. On fait une extrapolation du polynôme trouvé pour estimer les prochaines valeurs des échantillons. La forme générale d'estimation de la valeur du k<sup>i</sup>ème échantillon est :

Y\_k\_hat = Y\_{k-1} + Delta Y\_{k-1} + Delta^2 Y\_{k-1} + ... + Delta^n Y\_{k-1} (1)

ou Delta Y\_{k-1} = Y\_{k-1} - Y\_{k-2}

...

Delta^n Y\_{k-1} = Delta^{n-1} Y\_{k-1} - Delta^{n-1} Y\_{k-2}

Le cas le plus simple est obtenu pour n = 0 ; le prédicteur est alors d'ordre zéro. La formule (1) devient :

Y\_k\_hat = Y\_{k-1}

Dans une forme plus générale, on fait la prédiction suivante :

Y\_k\_hat = sum\_{j=1}^M a\_{k-j} Y\_{k-j} (2)



## LA REDUCTION DE LA REDONDANCE DANS LES TELEMEURES

Les coefficients  $a_{n-j}$  seront choisis de façon à minimiser l'erreur efficace. On obtient un système de M équations simultanées dont la solution nécessite un nombre important de calculs.

En ce qui concerne les performances des différents types de prédicteurs, on a trouvé le prédicteur d'ordre zéro supérieur aux autres prédicteurs polynominaux caractérisés par l'équation (1). Les résultats obtenus confirment ceux des références (1) et (2). En ce qui concerne le prédicteur à erreur efficace minimale, la référence (2) lui confère un avantage par rapport au prédicteur d'ordre zéro, mais le volume de calculs à effectuer, qui sont à multiplier par le nombre de variables dans le format de télémesures, réduit son intérêt sur le plan pratique.

### 2° Interpolateurs

On se limite aux ordres zéro et un. Avec des points initial et final, on espère représenter un nombre de valeurs intermédiaires qui seront à l'intérieur d'une fourchette autour d'une ligne entre ces deux points.

En principe, l'interpolateur donne des performances supérieures à celles d'un prédicteur du même ordre parce que l'on fait une meilleure utilisation de la fourchette mais, pour une fourchette nulle, il n'y a pas de différence entre les deux.

### 3° Codages différentiels

Il s'agit des codages par delta ou par modulation aux impulsions codées en différentiel (DPCM), utilisés respectivement pour les codages de la voix et des images dans la plupart des cas. Les méthodes de codage envisagées normalement ne semblent pas adaptées au problème des télémesures où l'on veut maintenir une précision donnée quelles que soient les évolutions du signal à coder.

Dans la transmission des images, par exemple, on obtient une compression parce que l'oeil n'est pas sensible aux erreurs lorsqu'il y a un grand changement de niveau, ce qui permet une précision très grossière pour ces changements, avec une réduction correspondant dans le nombre de bits de code. De telles méthodes ne peuvent généralement pas être admises dans le codage des télémesures.

### 4° Transformations

Récemment, on a beaucoup étudié les possibilités d'emploi des transformées (Fourier, Hadamard-Walsh, Harr, etc...), surtout pour la transmission des images. Le nombre de calculs à effectuer est onéreux, mais le développement des nouvelles technologies semble permettre leur emploi dans l'avenir. Cependant, leur application aux télémesures semble lointaine ; le nombre de calculs est vraiment très grand puisqu'il faut calculer la transformée pour chaque variable dans le format ; aussi, la fidélité est difficile à contrôler.

#### 4 - PREDICTEUR A CODAGES VARIABLES

Vues les considérations évoquées ci-dessus, on s'est décidé sur le développement d'un algorithme basé sur le prédicteur d'ordre zéro auquel on apporte les modifications suivantes :

- lorsqu'il y a un échantillon non redondant, on code la différence de sa valeur par rapport à la valeur de l'échantillon précédent, au lieu de coder l'amplitude de l'échantillon. Pour ce codage, on utilise des codages à longueurs variables, ce qui prête à la distribution des probabilités de différences (voir paragraphe 5). On a utilisé deux types de codes. Le premier, un code de Huffman (référence 6) se base sur les probabilités mesurées. Il est donc optimal du point de vue performances. Le second est un code simplifié avec la composition suivante :

Différence	Mot de code		
1	0	1	
- 1	0	01	
2	0	001	
- 2	0	0001	
3	0	00001	
- 3	0	000001	
4	0	0000001	
- 4	0	00000001	
5	1	xxxxxxx	
.	.	.	longueurs
.	.	.	constantes
.	.	.	1 + 8 bits
- 255	1	xxxxxxx	



Les mots à longueurs variables commencent avec le bit 0. Après il vient un nombre de zéros correspondant à la valeur de la différence, ces mots se terminent par le bit 1. Pour des différences supérieures à  $\pm 4$ , on emploie des mots à longueurs fixes (9 bits), qui permettent le codage des valeurs allant jusqu'à  $\pm 255$ . Ces mots commencent avec le bit 1. Ce code est facile à traiter, du point de vue génération et aussi en ce qui concerne le décodage. Il est intéressant aussi parce que sa longueur maximale de 9 bits a seulement un bit de plus que la longueur du code fixe. Cela minimise la probabilité d'une décompression, phénomène possible lorsque l'on utilise ces codages.

- au lieu d'identifier les échantillons non redondants par un code classique à longueur fixe, on indique le nombre d'échantillons depuis le dernier échantillon non redondant. Cette indication est donnée par un code à longueur variable. On a utilisé deux codes : code de Huffman, utilisant les distributions mesurées et un code simplifié de la composition suivante :

Longueur de l'intervalle	Code	
0	0	1
1	0	01
2	0	001
3	0	0001
4	0	00001
5	0	000001
6	1	xxxxxx
.	.	longueurs
.	.	constantes
.	.	1 + 6 bits
63	1	xxxxxx

On a donc des mots à longueurs variables pour les intervalles 0, 1, 2, 3, 4 et 5 ; pour les autres intervalles, on a utilisé un code à longueur fixe (7 bits), seulement 1 bit de plus que les 6 bits nécessaires pour coder l'intervalle 63 que l'on a pris comme valeur maximale.

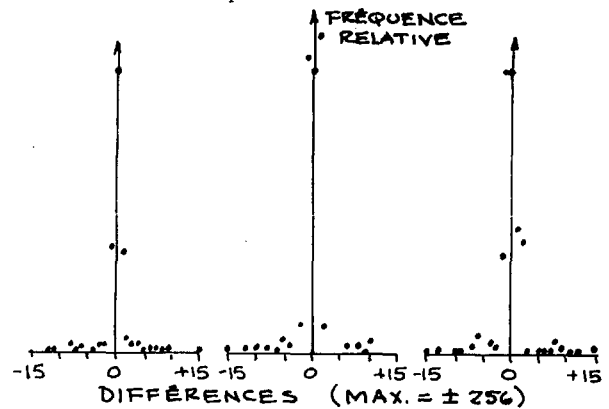
5 - TESTS EFFECTUES

On a effectué une série de tests sur deux bandes de télémesures. Le nombre de mesures obtenues étant limité, il faut regarder les résultats obtenus comme préliminaires, mais ils donnent une indication des performances que l'on peut espérer réaliser.

Les bandes ont les caractéristiques suivantes :

BANDE 1

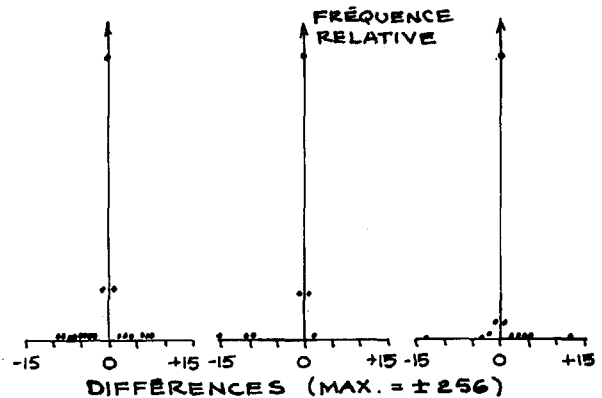
480 cycles, 14 canaux, 4 200 échantillons au total. Les variables sur cette bande évoluent relativement vite. Lorsque l'on mesure la fréquence des différences entre échantillons successifs, on obtient les courbes données sur la figure 2 pour les trois variables les plus actives.



DISTRIBUTION DES DIFFÉRENCES ENTRE ÉCHANTILLONS SUCCESSIFS - BANDE 1  
FIGURE 2

BANDE 2

6 130 cycles, 15 canaux, 92 000 échantillons. Les variables sur cette bande évoluent plus lentement. Pour les trois variables les plus actives, on obtient les distributions de différences entre échantillons successifs donnés sur la figure 3.



DISTRIBUTION DES DIFFÉRENCES ENTRE ÉCHANTILLONS SUCCESSIFS - BANDE 2  
FIGURE 3



Il faut noter qu'il n'y a pas de variable "tout ou rien" sur les bandes testées. Les informations sont quantifiées à 256 niveaux (8 bits).

On a effectué des tests de taux de compression et de longueur de la file d'attente. Pour les premières, on a mesuré les taux de compression pour les cas où l'on admet seulement les erreurs d'approximation nulle, c'est-à-dire, la valeur zéro de la fourchette.

#### 6 - RESULTATS DE COMPRESSION OBTENUS

On définit le taux de compression par l'expression :

$$R = \frac{\text{Nombre d'échantillons avant compression}}{\text{Nombre d'échantillons après compression}} \quad (3)$$

Dans le paragraphe 2, on a indiqué le besoin d'inclure avec les valeurs des échantillons, des informations permettant l'identification des échantillons après avoir appliqué le procédé de compression. Cela est nécessaire dans une liaison de transmission de télémesures. Le taux réel de compression sera, par conséquent, inférieur à celui donné par l'expression précédente. La différence dépend du format des télémesures (nombre de variables à identifier), ce qui fixe le nombre de bits d'identification qu'il faut ajouter à chaque échantillon non redondant. Comme cette caractéristique peut varier d'une situation à une autre, on pense que la formule donnée est la plus générale. Pour modifier les résultats donnés en fonction du nombre de canaux, on applique la formule :

$$R_r = R \left[ \frac{N_i}{N_i + \log_2 N_c} \right]$$

où  $R_r$  = taux réel de compression

$N_i$  = nombre de bits/mot d'information

$N_c$  = nombre de canaux.

Cette formule peut être modifiée pour prendre en compte le nombre de variables supercommutées.

Pour l'application que l'on envisage actuellement, réduction dans le nombre de bandes nécessaires au stockage des télémesures et non au problème de la transmission, les variables à compresser sont déjà accompagnées de mots d'identification et la formule (3) peut être utilisée directement.

Les remarques précédentes s'appliquent au prédicteur d'ordre zéro. Lorsque l'on effectue des codages à longueurs variables, il faut calculer le taux de compression en prenant en compte le nombre de bits utilisés par échantillon. Les résultats sont donnés par rapport aux mots de 14 bits (8 bits d'amplitude et 6 bits d'identification).

Le tableau ci-dessous donne les résultats obtenus sur les deux bandes avec une fourchette = 0.

ALGORITHME	TAUX DE COMPRESSION	
	BANDE 1	BANDE 2
Prédicteur zéro classique	2,5	13,7
Prédicteur zéro, codage Huffman des différences	3,6	24,0
Prédicteur zéro, codage variable simplifié des différences	3,2	20,3
Prédicteur zéro, codages Huffman des différences et des intervalles	6,1	33,2
Prédicteur zéro, codages variables simplifiés des différences et des intervalles	4,0	24,9

Pour séparer les effets des différents types de codage, on a fait d'abord des essais avec un prédicteur d'ordre zéro classique. Ensuite, on a obtenu des résultats en limitant les codages à longueurs variables aux différences entre les échantillons non redondants, et finalement on a appliqué ces codages aux différences et aux intervalles entre échantillons non redondants.

Dans le cas du codage des intervalles, on a limité l'intervalle maximal à 63 échantillons redondants. Le résultat obtenu est à comparer avec un prédicteur à codages classiques utilisé avec un format à 64 canaux de télémesures.

On voit que les codages à longueurs variables apportent une nette amélioration par rapport au prédicteur classique. On obtient de bons taux de réduction, et cela sans dégradation dans la fidélité de représentation des variables traitées.

La différence de performances pour les deux

bandes est très nette, ce qui montre combien les télémesures peuvent avoir des caractéristiques variables. C'est une des difficultés principales dans l'emploi des méthodes de réduction de la redondance dans ce domaine.

7 - FILE D'ATTENTE

Dans l'application d'un système de la compression des télémesures, on sera souvent obligé de sortir les données non redondantes sur une liaison à cadence constante. Pour assurer la continuité de la liaison, il faut prévoir une mémoire tampon après l'algorithme de compression.

Le choix de la taille de la mémoire tampon pose toujours un problème : on veut qu'elle soit suffisante pour minimiser les pertes dues aux débordements lorsqu'il y a des périodes passagères pendant lesquelles le taux de compression devient plus faible que sa valeur moyenne, mais on ne veut pas une capacité excessive qui réduirait l'interêt du procédé de compression.

Pour traiter le problème de débordement, les méthodes utilisées jusqu'ici ont été :

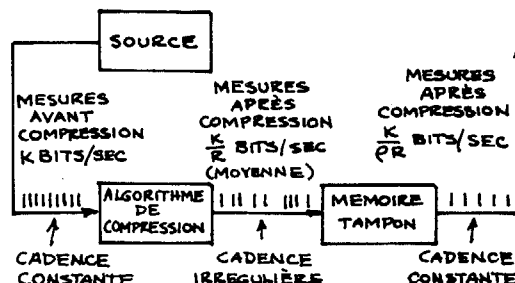
- augmentation de la valeur de la fourchette. Cela augmente le taux de compression ; on peut espérer ou minimiser les pertes en élargissant la fourchette lorsque le niveau de remplissage de la mémoire tampon arrive à une valeur élevée
- élimination tout simplement des échantillons par ordre de priorité.

Ces méthodes ont des inconvénients évidents et l'on préfère minimiser leur emploi. Les tests effectués montrent que l'emploi de mots d'identification (longueurs d'intervalles entre échantillons non redondants) de longueur variable réduit les fluctuations dans la longueur de la file d'attente, ce qui réduit l'emploi des méthodes de limitation de sa longueur.

L'idée de base est de coder les intervalles courts en utilisant les mots courts et les intervalles longs avec les mots longs, comme on l'a indiqué dans le paragraphe 4. La mémoire tampon commence à se remplir lorsque le taux de compression baisse, ce qui implique des intervalles courts entre échantillons non redondants. Le fait d'employer des mots plus courts à ce moment là devrait agir de façon à freiner le remplissage de la mémoire.

Pour mieux définir le rôle de la mémoire tampon

dans une liaison de transmission, considérons le schéma de la figure 4.



SCHEMA DU CODEUR  
FIGURE 4

Après compression avec un taux  $R$ , on obtient en moyenne  $\frac{K}{R}$  bits/s à l'entrée de la mémoire tampon, la cadence des échantillons étant irrégulière. A la sortie de la mémoire, à cadence constante, on a  $\frac{K}{\rho R}$  bits/s ; la valeur  $\rho (< 1)$  est choisie de façon à trouver un compromis entre la capacité de la mémoire et la probabilité de débordement de la mémoire.

Avec la bande 2, on a effectué plusieurs tests dont les résultats sont donnés figures 5 et 6, pour valeurs de  $\rho$  égale à 0,87 et 0,95.

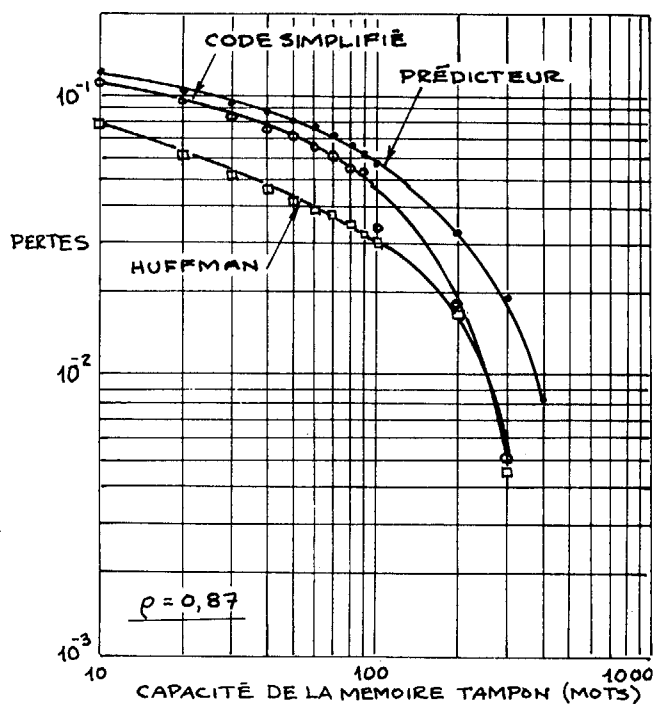
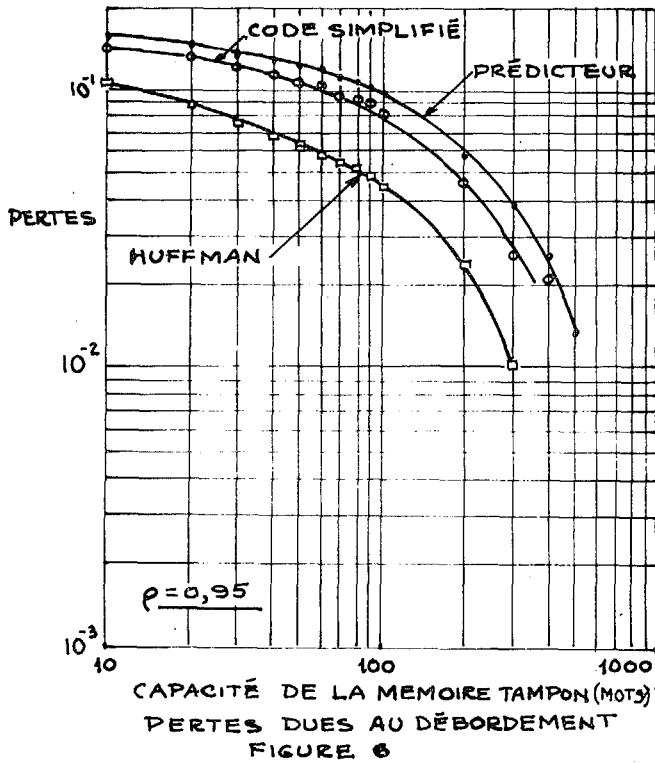


FIGURE 5



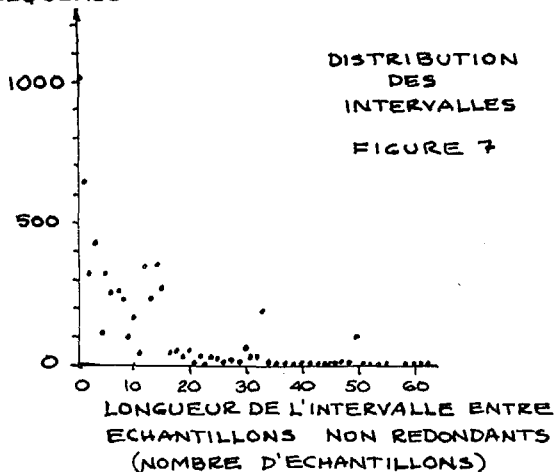


LA REDUCTION DE LA REDONDANCE DANS LES TELEMESURES



On a fait deux séries de tests avec un code de Huffman et le code simplifié déjà décrit. On voit que les codages variables donnent toujours une amélioration dans le comportement de la file d'attente. Pour le cas actuel, le code de Huffman donne des résultats nettement supérieurs à ceux obtenus avec le code simplifié. Cela vient, en partie, du fait que l'on a effectué les essais avec un format à 16 voies de télémesures, et la longueur d'intervalle 16 apparaît très fréquemment (figure 7).

**FRÉQUENCE**



Comme le code simplifié donne un mot relativement long pour cette valeur, une partie de son efficacité est perdue. Avec un format plus grand, à 64 voies par exemple, ce problème ne devrait pas se présenter.

8 - CONCLUSIONS

On a montré qu'un bon taux de réduction de la redondance dans les télémesures peut être obtenu par l'emploi d'un prédicteur d'ordre zéro avec un codage simple à longueur variable des différences et des intervalles entre échantillons non redondants.

Le codage utilisé permet aussi de réduire la longueur de la file d'attente à l'émission dans une liaison de transmission de données comprimées. Cela permet, pour une capacité de mémoire donnée, d'obtenir un taux de débordement plus faible ou de réduire l'emploi de méthodes normalement prévues pour la limitation de la longueur de la file d'attente.

9 - REFERENCES

- (1) "PCM telemetry data compression study, phase II", Lockheed Missiles and Space Co, January 1967,
- (2) "Adaptive compressive telemetry techniques", IBM, Federal Systems Division, 31 August 1966, NASA CR-65498.
- (3) Capellini, V et al, "A study of data compression techniques", ESRO Report N° CR-93, April 1974.
- (4) "Final Report for OAO data transmission study", Adcom Inc., July 1967, NASA CR-90109.
- (5) Turner, L.F et al, "Data compression for archiving", ESOC Report N° 417/72/AR.
- (6) D.A. Huffman, "A method for the construction of minimum redundancy codes", Proc. IRE, September 1952.