Sequential Decoding in the Presence of a Noisy Carrier Reference

J. R. Lesh
Network Operations

A new model for predicting the computational performance of a sequential decoder operating in a noisy carrier reference environment is described. The major difference between this model and previous models is that the new model characterizes the number of computations per frame as the sum of the computations resulting from a number of independent searches. This number of independent searches can then be considered as an effective frame length. When this computational model is averaged over noisy reference phase errors using a medium-rate interpolation scheme, the results are found to agree quite favorably with experimental measurements.

I. Introduction

Characterizing the computational behavior of a sequential decoder in the presence of a noisy carrier reference has long been recognized as a difficult problem. The first difficulty one encounters when undertaking such a study is to characterize the computational behavior of an ideal (noiseless carrier reference case) decoder. It was conjectured by Savage (Ref. 1) and subsequently verified many times experimentally that the number of computations required to decode one bit (or branch) behaves as a Pareto random variable. However, when using sequential decoding one must necessarily group the incoming data into blocks or frames, in which case the random variable of interest is the number of computations per frame rather than the number of computations per bit. Unfortunately, no satisfactory model for the distribution of computations per frame has been proposed to date. Some progress has been reported by Layland (Refs. 2 and 3) using curve fitting techniques of data produced by simulations. Such a technique has the obvious disadvantage of being valid only when the simulation conditions are reproduced, and furthermore, one often loses the insight afforded by more analytical models. In this paper we develop directly a model for the distribution of computations per frame which is based not on simulations but on convolutions of Pareto distributions (Section II). In Section III, we compare the results of this model to experimental data and link the two by considering a quantity called the effective frame length.

In Section IV we attack the second major difficulty associated with actual decoder performance modeling: i.e., the effects of a noisy carrier reference. We consider both high- and low-rate phase error effects upon the ideal decoder model developed previously. The final "real" decoder model results using a medium-rate technique developed by Tausworthe (Ref. 4). This final model is compared with experimental data in Section V, followed by conclusions in Section VI.

II. Convolutions of Pareto Distributions

Let C_1 represent the number of computations required for the decoder to decode (or advance) one bit. As previously stated, it is well known that the random variable C_1 approximately obeys the Pareto law; i.e.,

$$P_r\{C_1 > L\} \approx L^{-\alpha} \tag{1}$$

where α is the Pareto exponent determined by

$$E_0(\alpha) = \alpha R_N \tag{2}$$

where $E_0(\alpha)$ is the random coding bound exponent function (Ref. 5) which depends implicitly on the signal-to-noise ratio and R_N is the code rate in bits per channel symbol. It is customary to plot computational distributions on log-log paper, in which case Eq. (1) represents a straight line with slope $-\alpha$. Now let C_F denote the number of computations required to decode a frame of F bits. Two models for the frame computational distribution which have been proposed in the past are

$$P_r\{C_F \geq L\} = FL^{-\alpha} \tag{3}$$

and

$$P_r\{C_F \geq L\} = \left(\frac{L}{F}\right)^{-\alpha} \tag{4}$$

Equation (3) assumes that the accumulation of L computations occurs from a single long search, whereas Eq. (4) assumes that every bit requires exactly the same number of computations. Both of these expressions are clearly wrong since $P_r\{C_F \geq F\}$ is incorrectly predicted by Eq. (3) (except when $F = 1$ or $\alpha = 1$), and the assumption for Eq. (4) is clearly invalid. However, it is important to note that Eq. (3) represents a vertical shift of Eq. (1) by log F (on log-log paper), whereas Eq. 4 represents a horizontal shift of log F. Both of these observations will become useful later.

With this background, let us consider the following construction. We assume that the computational distribution per bit is given by Eq. (1) (with equality). That is to say:

$$P_r\{C_1 = L\} = L^{-\alpha} - (L+1)^{-\alpha} \tag{5}$$

Now, let us assume that the number of computations required to decode a frame of F_e bits (C_{F_e}) is given by the sum of F_e independent and identically distributed (i.i.d.) random variables each distributed according to Eq. (5). (This assumption will be modified later.) Then it is well known that the probability density function of C_{F_e} is given by the (F_e-1) fold convolution sum of Eq. (5). Figure 1 illustrates the results of such convolutions when the Pareto exponent $\alpha = 1.5$ and for various values of F_e. It is interesting to note that as the frame length F_e increases, the "Pareto-like" characteristic of C_{F_e} rapidly disappears with the appearance of a low end knee. Such a knee has been observed in experimentally determined distributions of computations per frame. Also of significance is the fact that all of the curves tend asymptotically to the result given in Eq. (3), with F replaced by F_e. Such behavior was predicted earlier by Sussman (Ref. 6). Figure 2 illustrates the behavior of the distribution of C_{F_e} as α varies (i.e., signal-to-noise ratio (SNR) varies) for $F_e = 192$.

In order to utilize these results, it is necessary to characterize the distribution of C_{F_e} without having to perform the convolutions each time. For the purpose of characterizing these distributions, a large number of Pareto distribution convolutions were performed at different values of α and F_e and the results studied. It soon became apparent that a pattern was emerging. In particular, it was noticed (at least for reasonably large F_e) that if one fixed the value of α and varied F_e, the point on each C_{F_e} distribution which had a tangent line parallel to the original Pareto distribution occurred at a value of L which was a constant multiple of F_e. In other words, given α and F_e sufficiently large, there exists a constant $K(\alpha)$ such that the tangent line to the distribution C_{F_e} at $L = K(\alpha)F_e$ has slope $-\alpha$. This point appears to be quite useful since it is quite close to the value of L for which the distribution of C_{F_e} begins to rapidly drop. Several values of $K(\alpha)$ are shown in Fig. 3. We note that $K(\alpha)$ can be well approximated by

$$K(\alpha) \exp \frac{1.54}{\alpha^{2.1}} \tag{6}$$

This approximation is also shown in Fig. 3.

Let us now consider the following model for the distribution of C_{F_e}. Given α and F_e let $P_r\{C_{F_e} \geq K(\alpha)\cdot F_e\} = 1$. At $K(\alpha)\cdot F_e$ let the model distribution begin dropping in a straight-line manner which best describes the rapid descent region of the convolved Pareto distribution. This amounts to letting the distribution be characterized as behaving initially as a Pareto distribution with exponent (say) α'. At the point where this straight-line distribution crosses the asymptotic distribution given by Eq. (3) we begin following the asymptotic distribution. By performing such constructions, we again find that the value of α' is also reasonably independent of F_e (for sufficiently large F_e). Several values of α' are shown in Fig. 4 along with an approximation given by

$$\alpha' \approx \alpha + \frac{3\,\alpha^{2.91}}{4} \tag{7}$$

Combining these results we obtain the model for the distribution of C_{F_e} given by

$$P_r\{C_{F_e} \geq L\} = \begin{cases} 1\ ;\ L < K(\alpha)F_e \\ \left[\dfrac{L}{K(\alpha)F_e}\right]^{-\alpha'}\ ;\ K(\alpha)F_e \leq \mathrm{L} < \mathrm{L}^* \\ F_e\,L^{-\alpha}\ ;\ L^* \leq L \end{cases} \tag{8}$$

where

$$L^* = exp\left\{\frac{\alpha' \ln\left[K(\alpha)F_e\right] - \ln F_e}{(\alpha' - \alpha)}\right\} \tag{9}$$

Figure 5 compares the results of the model with the corresponding Pareto distribution convolutions for $F_e = 64$ and $\alpha = 1.0$ and $\alpha = 1.5$. It is interesting to note that a similar type of model (i.e., Pareto distribution with discontinuous α) has been used in the past by Berger and Mandelbrot (Ref. 7) for characterizing the distribution of sequences of intererror gaps in telephone lines.

We will find it useful to scale the results of our model by the frame length F_e. In particular we let

$$N = \frac{L}{F_e} \tag{10}$$

represent the average number of computations per bit (when a frame of F_e is decoded using L computations). It should be noted that N is quite different from the ensemble average of the random variable representing the number of computations per bit. For example, for any values of L and F_e, N is defined as their ratio. On the other hand, the average of the number of computations per bit is a fixed number and may be either finite or infinite depending on the value of α. In terms of N, our model for the distribution of C_{F_e} becomes

$$P_r\{C_{F_e} \geq N\cdot F_e\} = \begin{cases} 1\ ;\ N < K(\alpha) \\ \left[\dfrac{N}{K(\alpha)}\right]^{-\alpha'}\ ;\ K(\alpha) \leq N < N^* \\ F_e^{1-\alpha}\,N^{-\alpha}\ ;\ N^* \leq N \end{cases} \tag{11}$$

where

$$N^* = \exp\left\{\frac{\alpha' \ln K(\alpha) + (\alpha - 1)\ln F_e}{\alpha' - \alpha}\right\} \tag{12}$$

III. The Concept of Effective Frame Length

Now that we have a model for the distribution of C_{F_e} let us see if it can be used to predict sequential decoder performance. To accomplish this, the model in Eq. (11) was compared with experimentally determined frame computation distributions provided by Layland (Ref. 8). Figure 6 illustrates this comparison where we notice immediately that the value of F_e for which the model approximates the experimental result is very much smaller than the frame length F used in the experiment. In fact, it appears in Fig. 6 (as well as in essentially all other comparisons made but not included in this paper) that the value of F_e that one should use is

$$F_e \approx \frac{F}{50} \tag{13}$$

Indeed, this result appears to have been a significant factor in improving the accuracy of the model.

Let us reflect upon what Eq. (13) is telling us. Recall that in Section II F_e was used to determine the number of convolutions which we performed. Consequently, F_e, which we shall interpret as the "effective frame length," represents the number of independent searches made by the decoder. Recall also that a characteristic of the Pareto distribution is that a single long search of length L is more likely than two searches of length $L/2$. However, a single

long search covering the entire distance L is also very unlikely. Consequently, one would expect that within a frame of reasonable length there should be several long searches. Equation (13) tells us that, within a frame, approximately 2% of the bits in the frame result in long searches.

It can be argued that F_e should depend on α. Indeed this surely seems plausible and certainly a model for F_e showing this dependence would be an improvement. However, there are certain factors which make the exact selection of F_e less critical. At large α there is a good separation between lines of different F_e, and consequently Eq. (13) can be visually justified. As α approaches unity, all of the lines of F_e collapse into the same line. Consequently, any value of F_e will work in this region. In the interval $0 < \alpha < 1$, the different F_e curves again separate, except this time with the larger values of F_e on top. However, probabilities are always constrained to be not more than one, so that there is a limit to the separation that can occur for this region of α's. Furthermore, operation of decoders at these values of α usually occurs with relatively small probability, so that if one is interested in average performance, the error contributions resulting from the region $0 < \alpha < 1$ are usually quite small.

IV. The Effects of a Noisy Carrier Reference

When sequential decoders are used in data links involving phase-coherent carrier tracking, one must not only determine the operating characteristics of the decoder but must also determine the effect that the carrier tracking loop has on the decoder as well. There are two cases where these effects can be quite easily determined. The first of these, called the low-rate model, occurs when the data rate is so small relative to the carrier tracking loop bandwidth W_L (two-sided) that one can consider that the tracking loop phase error process $\phi(t)$ varies very rapidly over a sequential decoder bit (branch). In this case, one can compute the effective or degraded symbol energy to noise ratio $\overline{R}$ from

$$\overline{R} = R\left[E\left\{\cos\phi\,(t)\right\}\right]^2 \tag{14}$$

where R is the input symbol SNR. Lindsey (Ref. 9) has shown that Eq. (14) can be expressed as

$$\overline{R} = R\left[\frac{I_1(\rho_L)}{I_0(\rho_L)}\right]^2 \tag{15}$$

where $I_\nu(\cdot)$ is the ν^{th} order modified Bessel function and ρ_L is the carrier loop signal to noise ratio. Now, using $\overline{R}$ one can determine the appropriate random coding error exponent function $E_0(\cdot)$, and then from Eq. (2) the value of the Pareto exponent can be determined. Since we recognize that α depends on the signal-to-noise ratio we shall designate the low-rate Pareto exponent as $\alpha\,(\overline{R})$. Finally, the low-rate model is obtained by using Eq. (11) with α replaced by $\alpha\,(\overline{R})$.

At the opposite extreme, consider the case where the data rate is so high relative to W_L that one can consider the phase error process $\phi(t)$ as being constant over the entire sequential decoder frame. In this case one creates a high-rate model by computing $\alpha(R\cos^2\phi)$, where ϕ is a random variable distributed as $\phi(t)$. Viterbi (Ref. 10) has shown that this ϕ has a density function given by

$$p(\phi) = \frac{\exp(\rho_L\cos\phi)}{2\pi I_0(\rho_L)} \tag{16}$$

Thus, the high-rate model becomes

$$P_r(C_F \geq N\cdot F) = \int_{-\pi}^{\pi} P_r\{C_F \geq N\cdot F|\phi\}\,P(\phi)\,d\phi \tag{17}$$

Unfortunately, in many cases the data rate is such that neither the high or low rate assumptions are justified. When this situation occurs one uses some type of interpolation scheme to interpolate between the high- and low-rate models. The one we shall use is the method developed by Tausworthe (Ref. 4).

The usual problem one encounters when attempting to use this method or any other interpolation model originally developed for uncoded or block coded data, is determining the effective integration time T_M of the sequential decoder. It appears that a great deal of insight into the characterization of T_M can be gained by using the effective frame length developed in the previous section. To understand this, recall that L represents the total number of computations required to decode a frame of length F. Furthermore, we assume that there are F_e long (and independent) searches. For the remaining $(F - F_e)$ bits in the frame let us assume that each bit is decoded using only one computation. Thus, for these $(F - F_e)$ bits the effective integration time T_{M_1} is simply the time per bit T_b. If this is true, then the F_e long searches must accumulate

a total of $L-F+F_e$ computations. Thus, for the bits in this group there is an average number of computations per bit (N') given by

$$N' = \frac{L-F+F_e}{F_e} \tag{18}$$

In order to convert N' to a time we will use the "full tree" assumption suggested by Layland (Ref. 3). This assumption essentially says that if a long search occurs, the search pattern is more likely to look like a full tree search rather than a long search along a single wrong path. Since the average branch depth in a full binary tree containing N' branches is $\log_2 (1 + N'/2)$ the average integration time T_{M_2} for the long search bits is

$$T_{M_2} = T_b \log_2 \left(\frac{L-F+3F_e}{2F_e}\right) \tag{19}$$

Let us now apply the Tausworthe interpolation separately to the short and long searches. Toward this end let P_L and P_H represent the results computed from the low- and high-rate models respectively. Then we have

$$P_i(C_F \geq N \cdot F) = (1-a_i)\, P_L + a_i\, P_H; i = 1,2 \tag{20}$$

where

$$a_i = \frac{\delta_i}{4}\left[1 - \frac{\delta_i}{8}\left(1-e^{-\frac{8}{\delta_i}}\right)\right] \tag{21}$$

$$\delta_i = \frac{2}{W_L T_{M_i}} \tag{22}$$

and where $P_i\,(C_F \geq N \cdot F)$, $i = 1,2$ represent the medium-rate estimates of the computations distribution for the short and long search bits respectively. Then, if we average $P_i\,(C_F \geq N \cdot F)$, $i = 1,2$ over the times during which each result applies we obtain the final estimate $\hat{P}\,(C_F \geq L)$ given by

$$\hat{P}(C_F \geq L) = \left(\frac{F-F_e}{L}\right)[P_1\,(C_F \geq L) - P_2\,(C_F \geq L)] + P_2\,(C_F \geq L) \tag{23}$$

V. Comparison of Predicted and Experimental Results

The real test of a model is its ability to predict performance under real decoding conditions. In order to make such a comparison, data from DSS 71 (Ref. 11) as well as data from DSN System Performance Tests (SPTs), were compared with our model. Actually, a slightly more complex version of the model was used. In particular, the independent search computations distribution model given by Eq. (11), which is a two-straight-line model, was replaced by a three-line model constructed from the two-line result. The construction procedure is illustrated in Fig. 7 and results in a model given by

$$P_r\{C_{F_e} \geq N \cdot F_e\} = \begin{cases} 1\,; N < N_0 \\ \left(\dfrac{N}{N_0}\right)^{-\alpha_1}; N_0 \leq N < N_1 \\ \left[\dfrac{N^*}{K(\alpha)}\right]^{-\frac{\alpha'}{2}}\left(\dfrac{N}{N_1}\right)^{-\alpha_2}; N_1 \leq N < N_2 \\ F_e{}^{1-\alpha} N^{-\alpha}\,; N_2 \leq N \end{cases} \tag{24}$$

where

$$N_0 = \sqrt{K(\alpha)} \tag{25}$$

$$N_1 = \sqrt{K(\alpha)N^*} \tag{26}$$

$$N_2 = \left[\frac{(N^*)^3}{K(\alpha)}\right]^{1/2} \tag{27}$$

$$\alpha_1 = \frac{\alpha' \log\left[\dfrac{N^*}{K(\alpha)}\right]}{\log N^*} \tag{28}$$

and

$$\alpha_2 = \frac{-\log\left[F_e{}^{1-\alpha}\,(N^*)^{-\alpha}\right]}{\log\left[\dfrac{N^*}{K(\alpha)}\right]} + \frac{\alpha - \alpha'}{2} \tag{29}$$

However, the use of Eq. (24) produces only slight differences in the predictions relative to the two-line model.

Whenever comparing theoretical predictions with data taken in the DSN, one must be careful to separate the loss or degradation resulting from the carrier tracking loop from those occurring in the rest of the system (specifically the subcarrier tracking loop and symbol tracking loop). To accomplish this, a telemetry analysis computer program created by Dunn (Ref. 12) was used to determine system and subsystem losses. Those losses which did not

occur as a result of the carrier tracking loop were considered as degradations on the symbol energy-to-noise ratio.

Figure 8 illustrates the closeness of the predicted values to data provided from Ref. 11. These data were obtained using the Helios frame length of 1152, and each curve is characterized by the bit rate, modulation index and total-power-to-noise-density ratio P_T/N_0. In Fig. 9 we see the comparison of our model with data provided by the SPTs, also for a frame length of 1152. In this figure, the separate curves are characterized by the bit rate, modulation index and the symbol error rate (SER). It is believed that the separate curves can be more accurately characterized by using the SER since this quantity is directly measurable.

In Fig. 10 we see the comparison of the predicted result with experimental results for the Pioneer frame length of 192. Here we see a reasonable separation between the theoretical and experimental values. It is believed that this difference is a result of characterizing the distribution of C_{F_e}. Recall that if we use the 2% figure to relate the actual and effective frame lengths, then for Pioneer we are using an effective frame length of less than 4. However, it was stated that Eq. (11) (or, equivalently, Eq. 24) was a good approximation only for sufficiently large F_e. It is therefore believed that a more accurate model for the distribution of C_{F_e} is needed at the Pioneer effective frame lengths if one is to perform better estimates than are indicated in Fig. 10.

VI. Conclusions

We have seen that the performance of an ideal sequential decoder can be quite accurately predicted by considering a number of convolutions of Pareto distributions. We then found that the number of such convolutions could be interpreted as an effective frame length. The effective frame length proved not only useful in characterizing the ideal decoder but also provided much insight into determination of the medium-rate interpolation parameters used in characterizing real decoders operating in a noisy carrier reference environment. Although predictions for Pioneer frame lengths were found to be somewhat in error, the accuracy at longer frame lengths appears to be quite satisfactory.

Acknowledgment

The author wishes to express his appreciation for the comments, advice and experimental data provided by J. W. Layland.

References

1. Savage, J. E., "Sequential Decoding – The Computational Problem," *Bell System Tech. J.*, Vol. 45, pp. 149–174, 1966.
2. Layland, J. W., "Sequential Decoding with a Noisy Carrier Reference," in *The Deep Space Network Progress Report*, Technical Report 32-1526, Vol. XII, pp. 167–175, Jet Propulsion Laboratory, Pasadena, Calif., Dec. 15, 1972.
3. Layland, J. W., "A Sequential Decoding Medium Rate Performance Model," in *The Deep Space Network Progress Report*, Technical Report 32-1526, Vol. XVIII, pp. 29–40, Jet Propulsion Laboratory, Pasadena, Calif., Dec. 15, 1973.
4. Tausworthe, R. C., "Communications Systems Development: Efficiency of Noisy Reference Detection," in *The Deep Space Network*, Space Programs Summary 37-54, Vol. III, pp. 195–201, Jet Propulsion Laboratory, Pasadena, Calif., Dec. 15, 1968.
5. Gallager, R. G., *Information Theory and Reliable Communication*, John Wiley, New York, 1968.
6. Sussman, S. M., "Analysis of the Pareto Model for Error Statistics on Telephone Circuits," *IEEE Trans. Comm. Systems*, pp. 213–221, June 1963.
7. Berger, J. M., and Mandelbrot, B., "A New Model for Error Clustering in Telephone Circuits," IBM J. Res. Dev., Vol. 7, pp. 224–236, July 1963.
8. Layland, J. W., "Performance of an Optimum Buffer Management Strategy for Sequential Decoding," in *The Deep Space Network Progress Report*, Technical Report 32-1526, Vol. IX, pp. 88–96, Jet Propulsion Laboratory, Pasadena, Calif., June 15, 1972.
9. Lindsey, W. C., "Performance of Phase-Coherent Receivers Preceded by Bandpass Limiters," *IEEE Trans. Comm. Tech.*, Vol. Comm-16, No. 2, April 1968, pp. 245–251.
10. Viterbi, A. J., *Principles of Coherent Communication*, McGraw Hill, New York, 1966.
11. Layland, J. W., "DSS Tests of Sequential Decoding Performance," in *The Deep Space Network Progress Report 42-20*, pp. 69-77, Jet Propulsion Laboratory, Pasadena, California, April 15, 1974.
12. Dunn, G. L., "Conversational Telemetry Analysis Program," IOM 421E-74-108, May 1974 (JPL internal document).

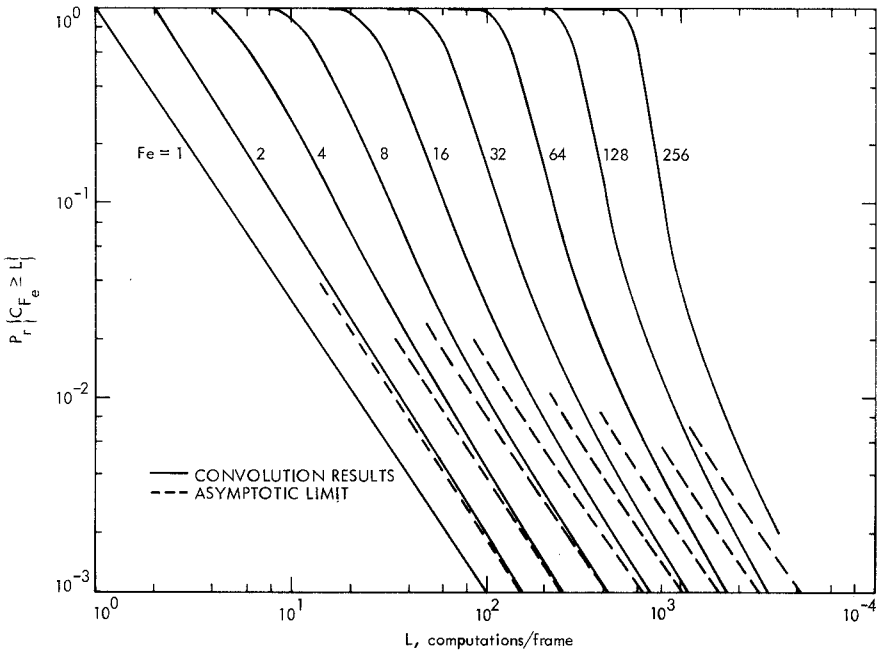

Fig. 1. Distribution of computations resulting from (F_e-1) convolutions of Pareto distribution $\alpha = 1.5$ (dashed lines indicate asymptotic result of Eq. 3)

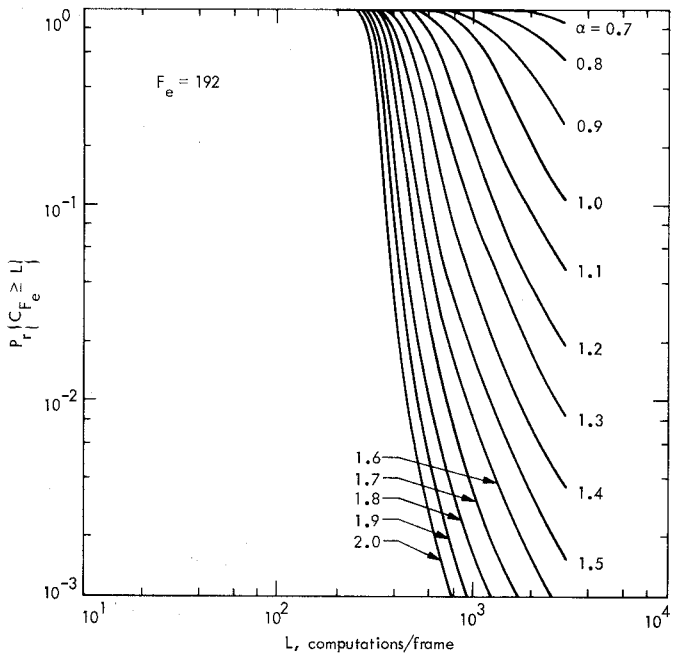

Fig. 2. Distribution of computations for independent searches, $F_e = 192$

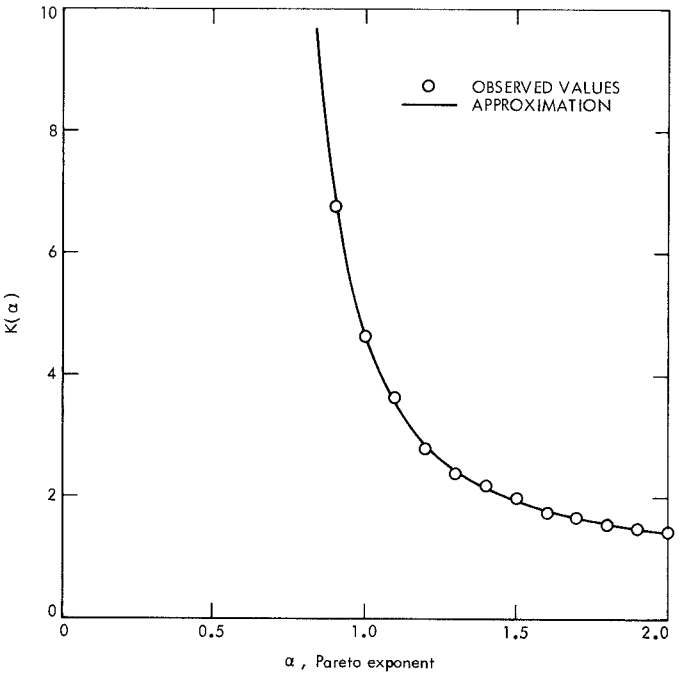

Fig. 3 Characterization of $K(\alpha)$

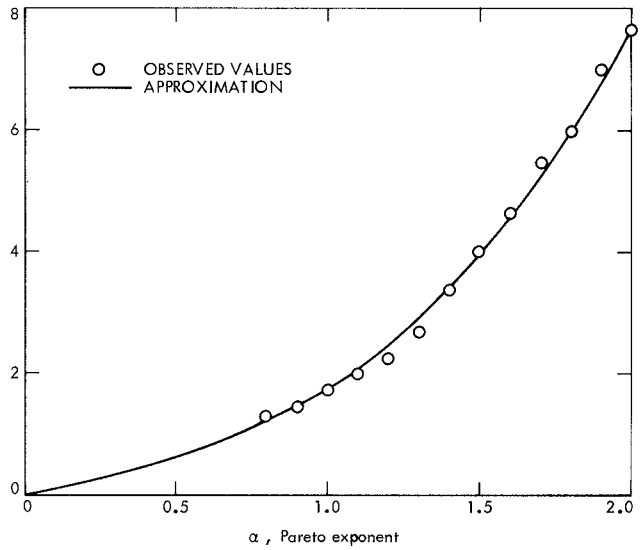

Fig. 4. Characterization of α'

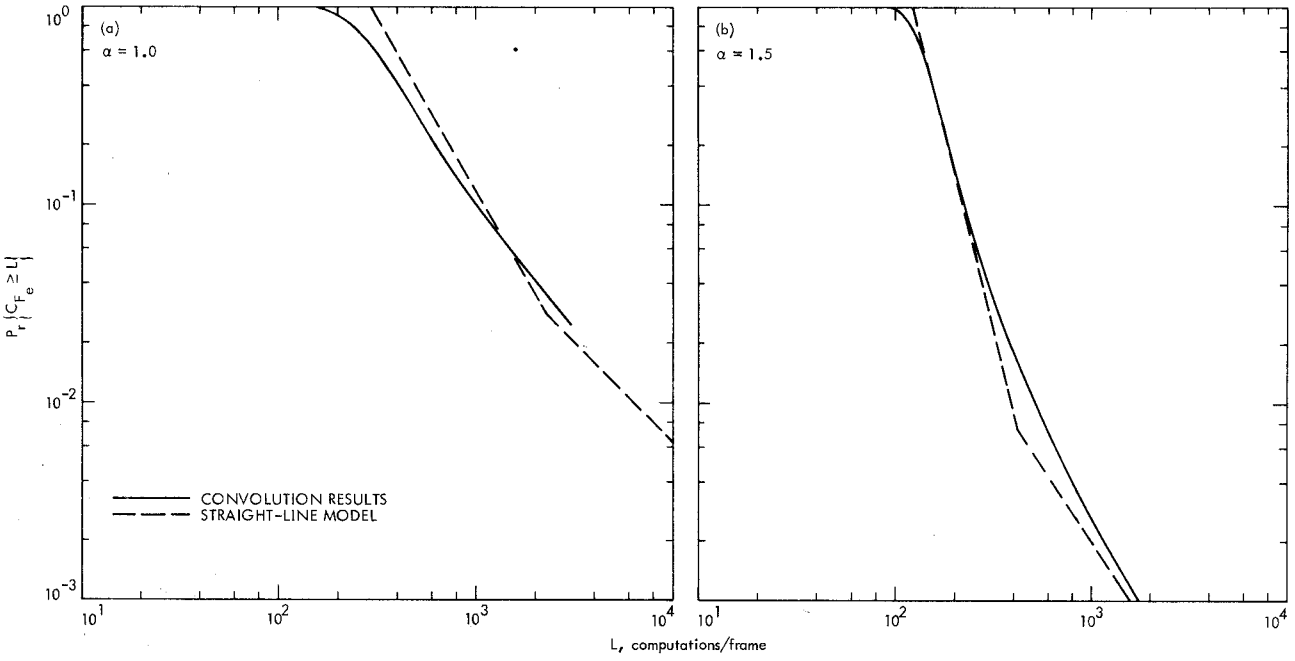

Fig. 5. Comparison of Pareto convolutions and straight-line model for $F_e = 64$ and (a) $\alpha = 1.0$, (b) $\alpha = 1.5$

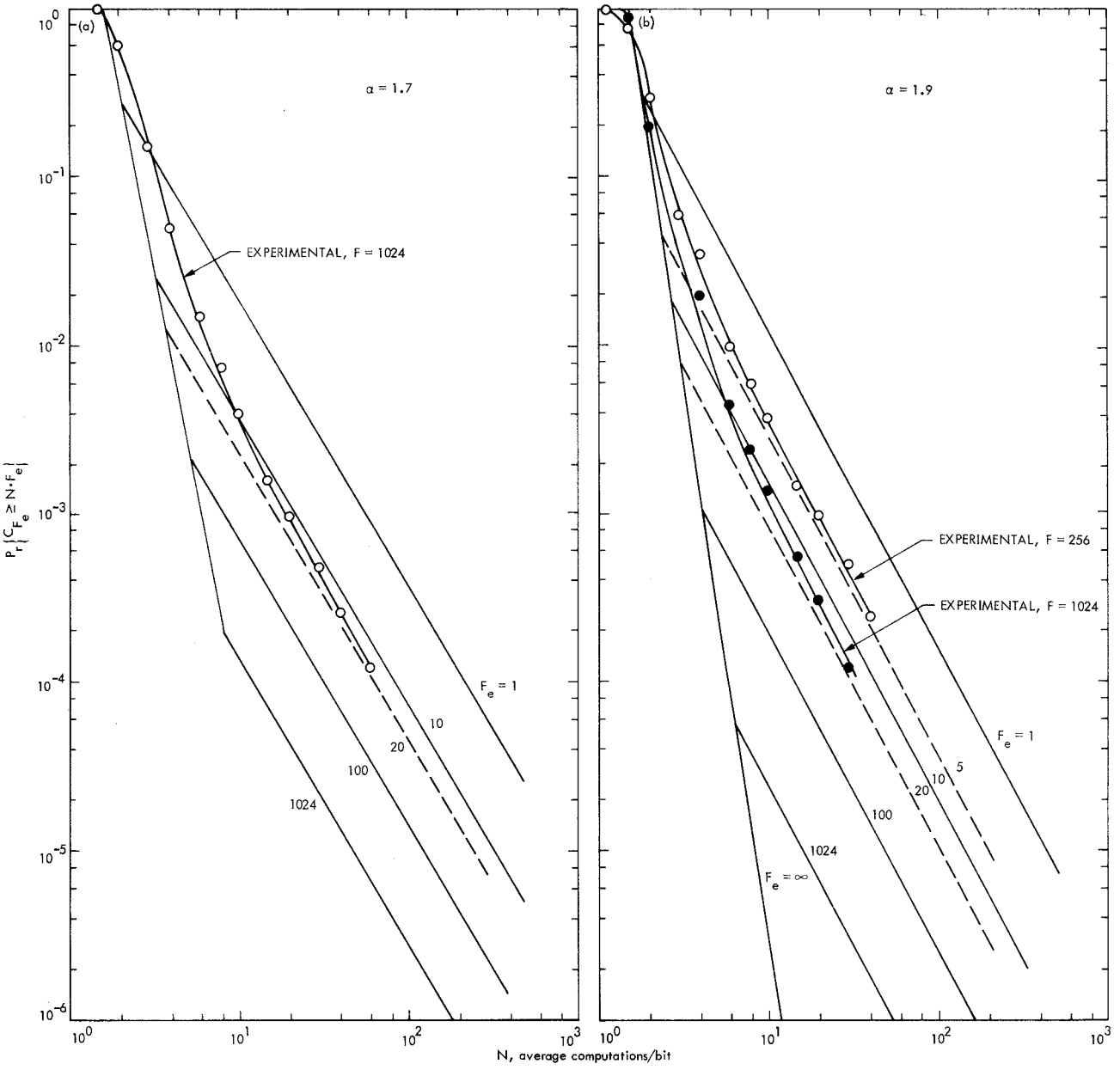

Fig. 6. Comparison of actual decoder computations distribution with straight-line model for (a) $F = 1024$, $\alpha = 1.7$, (b) $F = 256$ and $F = 1024$, $\alpha = 1.9$

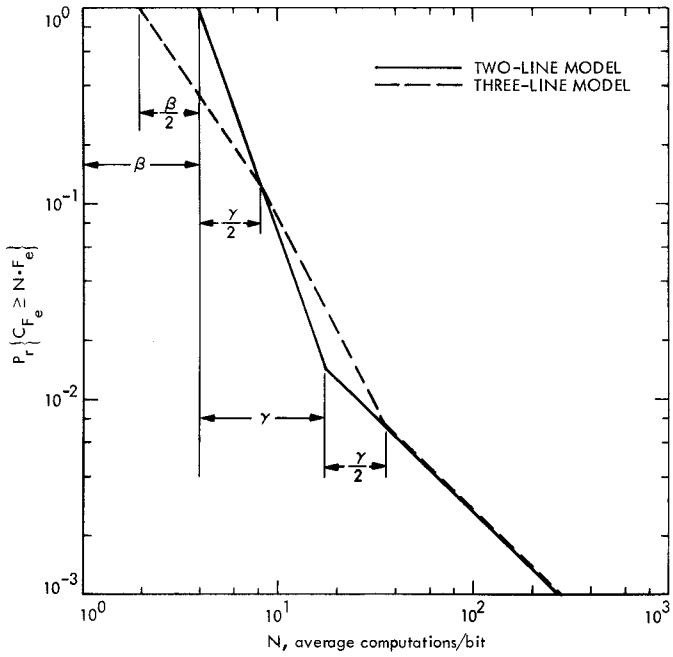

Fig. 7. Construction method for three-line model

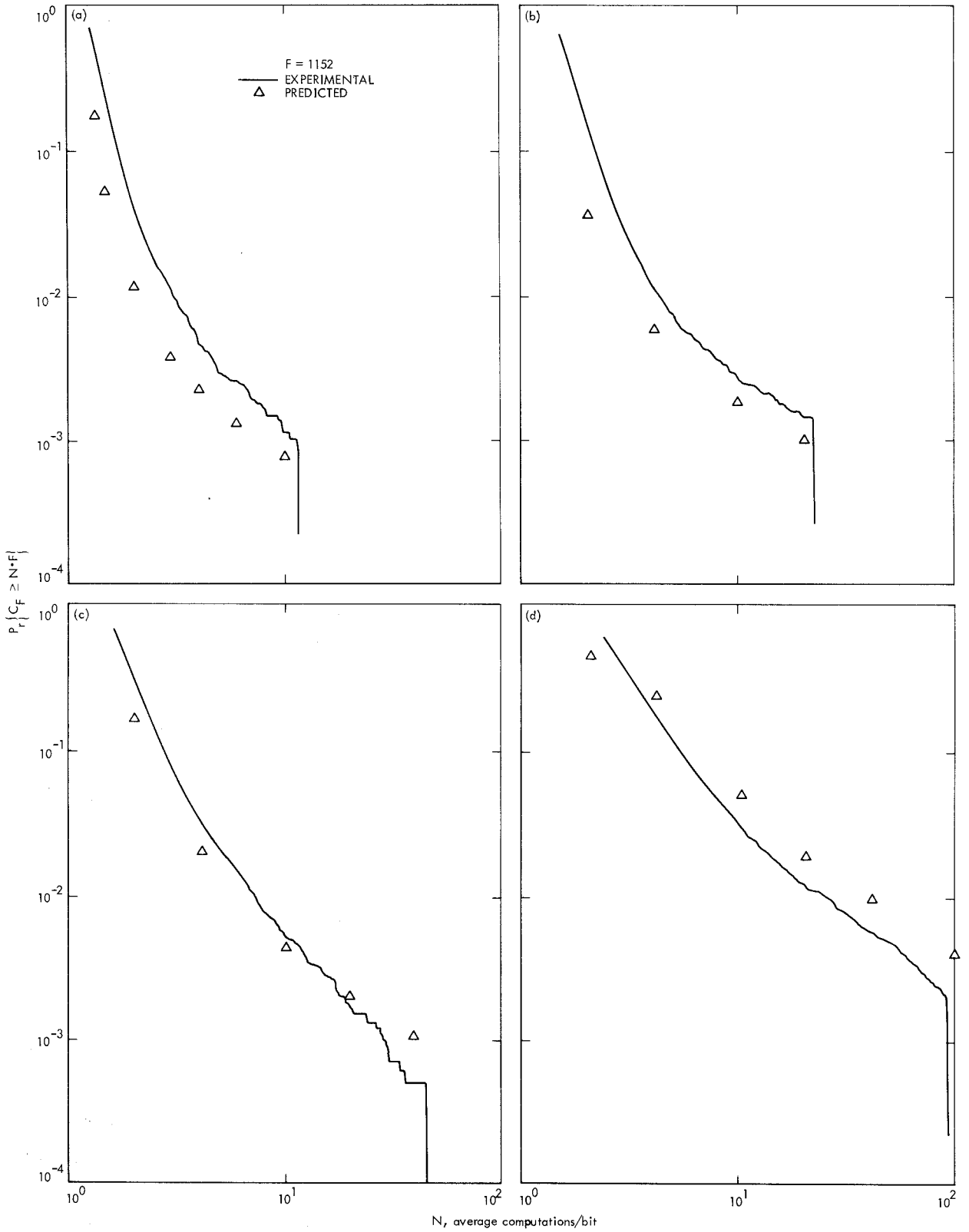

Fig. 8. Comparison of medium-rate model with actual noisy reference decoder for frame length 1152 and (a) 2048 bps, P_T/N_0 = 39.1 dB, mod index = 67.5 deg, (b) 1024 bps, P_T/N_0 = 36.3 dB, mod index = 60.0 deg, (c) 512 bps, P_T/N_0 = 34.1 dB, mod index = 48.0 deg, (d) 256 bps, P_T/N_0 = 31.4 dB, mod index = 42.0 deg

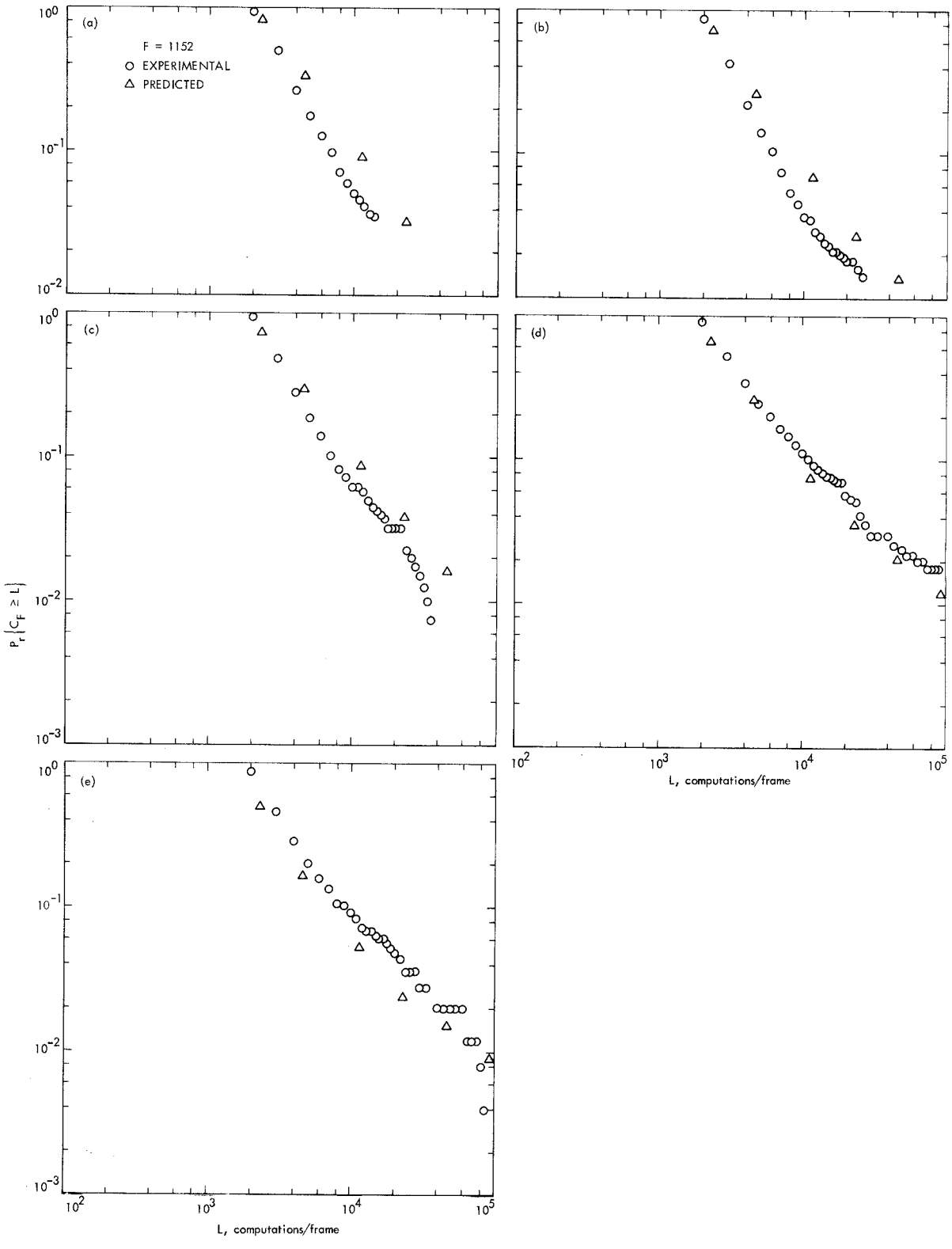

Fig. 9. Comparison of medium-rate model with actual noisy reference decoder for frame length 1152 and (a) 2048 bps, SER = 8.0%, mod index = 54.6 deg, (b) 1024 bps, SER = 7.7%, mod index = 54.6 deg, (c) 512 bps, SER = 7.9%, mod index = 54.6 deg, (d) 256 bps, SER = 7.9%, mod index = 54.6 deg, (e) 128 bps, SER = 7.5%, mod index = 54.6 deg

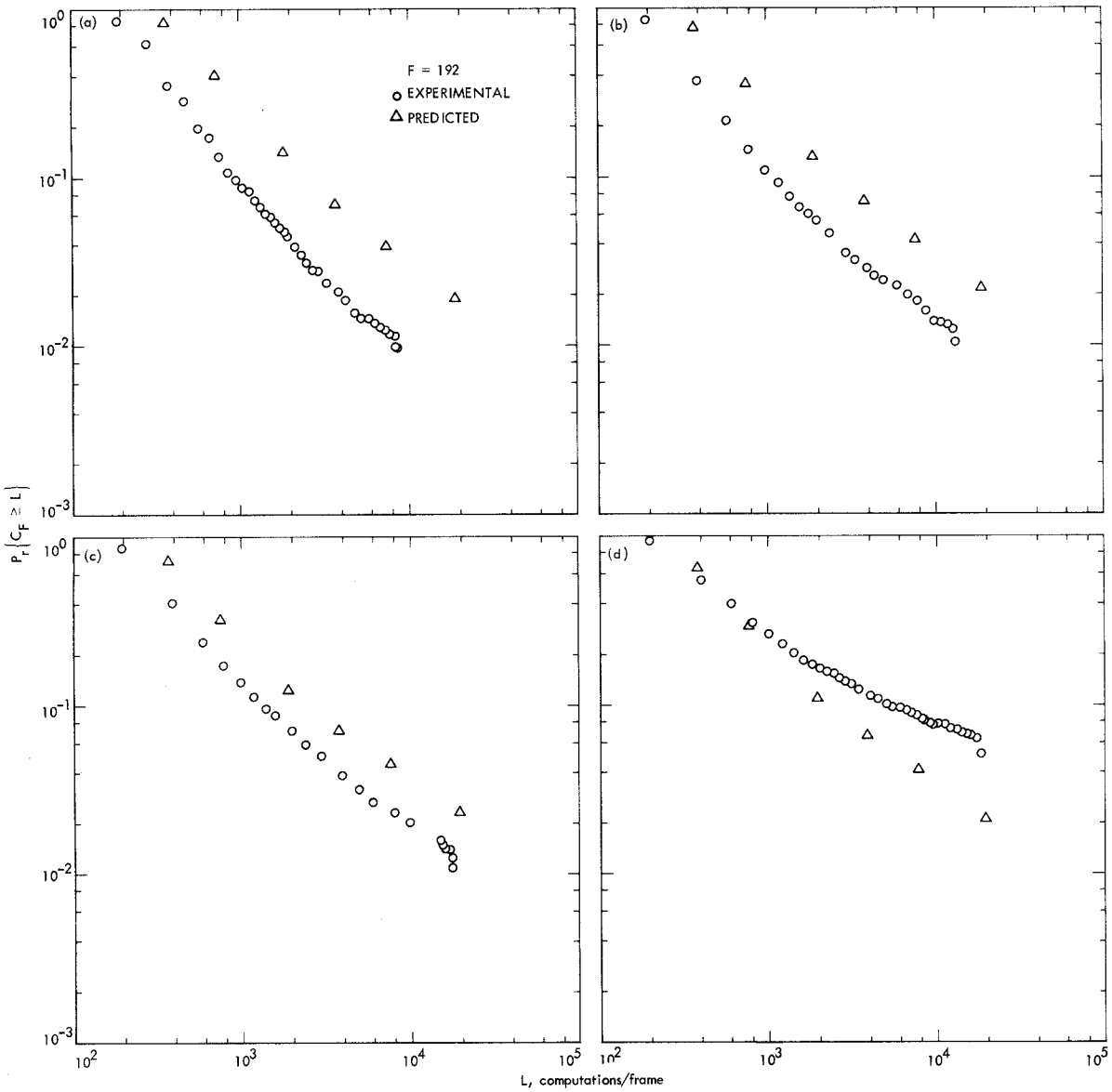

Fig. 10. Comparison of medium rate model with actual noisy reference decoder for frame length 192 and (a) 2048 bps, SER = 8.2%, mod index = 66.5 deg, (b) 1024 bps, SER = 8.1%, mod index = 66.5 deg, (c) 512 bps, SER = 8.2%, mod index = 66.5 deg, (d) 256 bps, SER = 8.5%, mod index = 66.5 deg