Finally, we note that by taking $M = L + T$, the ensemble of $R, L, T, M$ trellis codes becomes exactly the ensemble of $R, L, T$ tree codes. We have already noted that, for $M = T$, the ensemble of $R, L, T, M$ trellis codes becomes the ensemble of trellis codes defined by Massey [1]. Hence our Theorem is a generalization from which upper bounds on $P_e$ for both these ensembles follow as special cases.

III. RESULTS OF SIMULATIONS

Although the above theory was developed for true maximum-likelihood (i.e., Viterbi) decoding where one almost never uses a tail, its practical application is to sequential decoding where a tail is often used. The undetected error phenomenon is more complex for sequential decoding and, hence, we have to be careful with our conclusions. Nevertheless, it is well-known [5], [6] that, with the appropriate bias term, the exponent of error probability for sequential decoding is the same as that for true maximum-likelihood or Viterbi decoding. Thus we have conducted sequential decoding simulations to test the dependence of $P_e$ on $T$ and $M$.

The particular sequential decoding algorithm employed was the stack algorithm [7], [8]. The simulations were all performed with rate $R = 1/2$ optimum distance profile codes [9], [10]. The simulated binary symmetrical channel (BSC) had "crossover probability" $p = 0.045$, which corresponds to $R = 0.45$. For three different code memory lengths, a very large number (100,000) of received "frames," i.e., complete received sequences of length $n(L + T)$, were decoded so that the decoding error probability could be accurately inferred.

In Fig. 1, we give the simulation results for the sequential decoding undetected error probability $P_e$ as a function of the tail length $T$ of the convolutional code. Because of the extreme variability of the computation in sequential decoding when $M$ is large, there were occasions where the decoding had to be stopped, and hence, the frames had to be erased because the computation exceeded the allotted maximum. The number of erased frames is indicated in Fig. 1 and had negligible effect on the curves. These curves show that the actual $P_e$ decreases exponentially with $T$ having an exponent very close to that of the bound (5) for the range $T < M - [nE_{UL}(R)]^{-1} \log_2 L + 1$, while further increases in $T$ beyond this point have virtually no effect on $P_e$.

The range of $T$ for which the bound becomes independent of $L$, viz., $T < M - [nE_{UL}(R)]^{-1} \log_2 L$, is close to the range where the true $P_e$ becomes independent of $L$. Hence, relation (6) can be taken as a slightly conservative design rule for choosing $M$ so that $P_e$ is reduced to the minimum possible for the tail length $T$ that can be allocated to an encoded frame.

IV. REMARK

Finally, we should remark that, if we wanted solely to minimize the undetected error probability with sequential decoding for a given memory length and were not concerned with holding the tail size to a minimum to maximize the true rate of the trellis code, then the optimal value of the tail length is, of course, the memory length, i.e., $T = M$. Probably this fact has caused some investigators to ignore the distinction between the tail and the memory so that the memory length came to be honored for work actually done by the tail.

ACKNOWLEDGMENT

The author is greatly indebted to Prof. James L. Massey for his help in presenting these results in an easily understandable manner. Furthermore, the use in principle of convolutional codes with memory length greater than tail length to remove the dependence of $P_e$ on true length has been suggested independently by Prof. Massey [11].
define
\[ \delta(R) = \limsup_{n \to \infty} \delta(n, R) \]
\[ \hat{\delta}(R) = \liminf_{n \to \infty} \delta(n, R) \]
\[ \bar{E}(R) = \limsup_{m \to \infty} \frac{1}{n} \left( -\log_2 P_e(n, R) \right) \]
\[ \underline{E}(R) = \liminf_{m \to \infty} \frac{1}{n} \left( -\log_2 P_e(n, R) \right). \]

It is widely believed that the limits in (3) and (4) exist. However, for \( \delta(R) \), this is known only for \( R = 0 \) and \( 1: \delta(0) = \delta(1) = 0 \). For \( \bar{E}(R) \), the limit is known to exist, and its value is known at \( R = 0 \) and \( R \geq R_{\text{crit}} \) where \( R_{\text{crit}} \) being a number to be defined below. We shall now briefly survey the known upper and lower bounds on \( \bar{E}(R) \), and indicate how the new upper bound \( \delta^*(R) \) obtained in [1] can be used to improve the known upper bounds on \( \bar{E}(R) \) for small values of \( R \).

First, we have the sphere-packing bound \( E_{\text{sph}}(R) \) and the random coding bound \( E_{\text{rc}}(R) \), both valid for all rates \( R \) less than channel capacity [2], [3]:
\[ E_{\text{sph}}(R) \leq E(R) \leq \bar{E}(R) \leq E_{\text{rc}}(R). \]

The two bounds in (5) are equal for sufficiently large \( R \), and in fact the number \( R_{\text{crit}} \) cited above is the point where these two bounds meet. (Formulas for \( E_{\text{sph}} \) and \( E_{\text{rc}} \) for binary symmetric and binary erasure channels are given in the Appendix.)

Next, we have bounds which depend on the Bhattacharya parameter [4], [5] for the channel, which is defined by
\[ \alpha = -\log_2 \sum_{y \in B} (p(y|0)p(y|1))^{1/2}. \]

These bounds are
\[ \alpha D \leq E(R) \leq \bar{E}(R) \leq \alpha \bar{E}(R), \]
where \( 0 \leq D \leq \frac{1}{2} \) is defined implicitly by \( R = 1 - H_2(D) \), where \( H_2(x) \) is the binary entropy function. (The lower bound in (6) is called the expurgated bound \( E_{\text{sph}}(R) \); it is valid only for \( 0 \leq R \leq R^* \), where \( R^* \) is the rate at which the expurgated bound meets the random coding bound.) As mentioned, the function \( \delta(R) \) is unknown, so the upper bound in (6) is ineffective. However, by using the bound \( \delta(R) \leq \delta^*(R) \) obtained in [1] (for numerical values of \( \delta^*(R) \), see Table 1 in [1]), we obtain an upper bound
\[ \bar{E}(R) \leq \alpha \delta^*(R) \]
which can be evaluated, and which is already better than any previously known upper bound for small values of \( R \).

Finally, Shannon et al. [3] have shown that if \( E_o(R) \) is any upper bound to \( \bar{E}(R) \), then so is the convex hull of the curves \( E_o(R) \) and \( E_{\text{rc}}(R) \). As mentioned, the function \( \delta(R) \) is unknown, so the upper bound in (6) is ineffective. However, by using the bound \( \delta(R) \leq \delta^*(R) \) obtained in [1] (cf. (7)), we can obtain an upper bound which is significantly better than \( \min\{E_{\text{sph}}(R), \bar{E}(R)\} \) for a considerable range of \( R \). We illustrate this in Fig. 1 with a binary symmetric channel with crossover probability \( t = 0.01 \), \( \gamma^* = -\log_2 t(1-t) = 2.329 \), and in Fig. 2 with a binary erasure channel with erasure probability \( t = 0.01 \), \( \gamma = -\log_2 t = 6.644 \). In both figures, the unknown region for \( 0 \leq R \leq R_{\text{crit}} \) bounding \( (\delta(R), \bar{E}(R)) \) is shaded. A final point worth mentioning is that the new upper bound (7) on \( \bar{E}(R) \) always matches the expurgated bound \( E_{\text{rc}}(R) \) at \( R = 0 \). (Both slopes are \( -\alpha \); this slope is well-known for the expurgated bound, and follows for the bound (7) from the results of [1].) This fact supports the conjecture that \( \bar{E}(R) = \bar{E}(R) = E_{\text{rc}}(R) \) for \( R \leq R_{\text{crit}} \) for binary-input channels.

**APPENDIX**

**\( E_{\text{sph}}(R) \) AND \( E_{\text{rc}}(R) \) FOR BINARY SYMMETRIC AND BINARY ERASURE CHANNELS**

For a binary symmetric channel with crossover probability \( \epsilon \), the random coding exponent is given by
\[ E_{\text{rc}}(R) = \begin{cases} 1 - R_{\text{crit}}(D(1 + \sqrt{\epsilon - 2\epsilon(1 - \epsilon)})) & \text{if} \ 0 \leq R \leq 1 - H_2(\sqrt{\epsilon(1 - \epsilon)}), \\ (T_{\epsilon}(D) - H_2(D)) & \text{if} \ 1 - H_2(\sqrt{\epsilon(1 - \epsilon)}(1 - \epsilon)) \leq R \leq 1 - H_2(\epsilon), \end{cases} \]
where \( D = H_2(\epsilon) \), \( D = H_2(\epsilon) \) is the binary entropy function, and \( T_{\epsilon}(D) \) is the highest rate at which the binary symmetrical channel can be operated without error. For a binary erasure channel with erasure probability \( \epsilon \), the random coding exponent is given by
\[ E_{\text{rc}}(R) = \begin{cases} 1 - R_{\text{crit}}(D(1 + \sqrt{\epsilon - 2\epsilon(1 - \epsilon)})) & \text{if} \ 0 \leq R \leq 1 - H_2(\sqrt{\epsilon(1 - \epsilon)}), \\ (T_{\epsilon}(D) - H_2(D)) & \text{if} \ 1 - H_2(\sqrt{\epsilon(1 - \epsilon)}(1 - \epsilon)) \leq R \leq 1 - H_2(\epsilon), \end{cases} \]
where \( T_1(D) = -D \log_2 \epsilon - (1 - D) \log_2 (1 - \epsilon) \), and where \( D \) satisfies (7). The sphere packing exponent is

\[ E_{sp}(R) = T_1(D) - H_2(D), \quad 0 \leq R \leq 1 - H_2(\epsilon). \]

(Hence, \( R_{crit} = 1 - H_2(\sqrt{\epsilon + \sqrt{1 - \epsilon}}) \) and \( E(R) = E_1(R) = E_{sp}(R) \) for \( R \geq R_{crit} \).) For the binary erasure channel with erasure probability \( \epsilon \),

\[ E_r(R) = \begin{cases} 1 - R - \log_2 (1 + \epsilon), & 0 \leq R \leq 1 - 2\epsilon/(1 + \epsilon) \\ \log_2 (1 - \epsilon) + 2\epsilon, & R \geq 1 - \epsilon, \end{cases} \]

where

\[ E_{sp}(R) = \frac{\rho(2^p)}{(1 - \epsilon) + 2\epsilon} - \log_2 ((1 - \epsilon) + 2\epsilon), \]

where \( \rho \) is determined by \( R = 1 - c^2/(1 - \epsilon + 2\epsilon) \).