AT&T Technical Journal
Vol. 64, No. 10, December 1985
Printed in U.S.A.

Union Bounds on Viterbi Algorithm Performance

By W. TURIN*

(Manuscript received January 30, 1985)

In the present paper we use transform methods (characteristic function techniques) and contour integrals to derive a closed-form expression for the performance union bound of a general discrete-time system. We show that previously published results may be derived as particular cases of the general formulation developed in this paper. It is well known that the maximum-likelihood Viterbi algorithm may be employed not only for decoding of convolutional codes but also for optimal detection in other situations. Examples include bandwidth-efficient demodulation, optimal accommodation for intersymbol interference and cross-channel coupling, text recognition, simultaneous carrier phase recovery and data demodulation, digital magnetic recording, nonlinear estimation and smoothing. The union bound is a useful measure of the performance of the Viterbi algorithm. Past closed-form expressions for the union bound have usually involved considerable approximation.

I. INTRODUCTION

A Maximum-Likelihood Receiver (MLR) is optimal when the input signal is distorted not only by noise but also by some deterministic factors. The MLR compares the received signal with all possible signals distorted by the same deterministic factors but not by the noise. The latter comparison signals must be available at the receiver. Possible deterministic impairments include intersymbol interference, cross-channel coupling, modem-implementation errors, channel misequalization, signal distortion by the channel nonlinearities, etc. Guided by some metric, the MLR searches for the comparison signal

* AT&T Bell Laboratories.

that is closest to the signal actually received and asserts that this signal, as it existed before being subject to the deterministic impairment, was the transmitted message.

The MLR performance depends on how well we can model the deterministic distortions of the signal and also on distances between signals. The signal separation may be increased by coding.

Technical problems arise in MLR implementation. Strictly speaking, we need to store the whole transmission history and generate all possible comparison sequences. However, if the system may be modeled by a Markov process, then the Viterbi algorithm may be used to realize a recursive MLR.[1]

In designing the MLR it is very important to be able to accurately evaluate the receiver performance. Because of a very large number of possible comparison signals, it is difficult to find the exact formula for an MLR performance characteristic. The performance characteristic upper bound (so-called union bound) is more easily found. Viterbi introduced transfer function techniques to evaluate the union bound for some performance characteristics of binary convolutional codes.[1] These methods have been extended to obtain performance bounds of the general finite-state system.[2] However, the original union bound was loosened to simplify a series summation. Using the transform methods developed in this paper, the original union bound is expressed in closed form.

II. SYSTEM MODEL

Consider a discrete-time system[2]

$$\begin{aligned} x_k &= f(w_k), \\ s_{k+1} &= g(w_k), \\ w_k &= (u_k, s_k), \quad -\infty < k < \infty, \end{aligned} \tag{1}$$

where u_k is a source symbol, x_k is the transmitted channel symbol, and s_k is the corresponding system (transmitter) state. Symbols x_k are transmitted over a noisy memoryless channel that outputs symbols

$$y_k = h(x_k, n_k), \tag{2}$$

where n_k are independent identically distributed variables. The receiver outputs symbols $\hat{w}_k = (\hat{s}_k, \hat{u}_k)$ which minimize the sum

$$M(y, w) = \sum_k m(y_k, w_k),$$

where $m(y_k, w_k)$ is the so-called branch metric and may be treated as a cost function for making a decision that the transmitter superstate was w_k if y_k was received. This metric is usually a measure of the signal degradation due to noise. For example,

$$m(y_k, w_k) = \ln \Pr\{y_k/f(w_k)\}$$
for the maximum-likelihood receiver,

$$m(y_k, w_k) = \ln \Pr\{u_k\}\Pr\{y_k/f(w_k)\}$$
for the maximum a posteriori receiver,

$$m(y_k, w_k) = -\| y_k - f(w_k) \|^2$$
for the minimum mean-square receiver.

The optimal solution may be found using the Viterbi algorithm. A sequence $\hat{w}_0, \hat{w}_1, \cdots, \hat{w}_{j-1}, w_j$, which terminates at time j, is called a survivor if it minimizes the metric accumulated to this time:

$$\sum_{k=0}^{j} m(y_k, \hat{w}_k) = \min_{(w_0, \cdots, w_{j-1})} \sum_{k=0}^{j} m(y_k, w_k),$$

where $w_j = \hat{w}_j$. It is obvious that the sequence that minimizes the total sum $M(y, w)$ must begin with one of the survivors. The survivors may be determined recursively via the Viterbi algorithm: $\hat{w}_0, \hat{w}_1, \cdots, \hat{w}_j, w_{j+1}$ is a survivor if and only if $\hat{w}_0, \hat{w}_1, \cdots, \hat{w}_{j-1}, w_j$ is a survivor and $\hat{w}_j$ satisfies the equation:

$$\sum_{k=0}^{j+1} m(y_k, \hat{w}_k) = \min_{wj} \sum_{k=0}^{j+1} m(y_k, w_k),$$

where $w_0 = \hat{w}_0, \cdots, w_{j-1} = \hat{w}_{j-1}$.

If all the survivors have a common part, then this part also belongs to the sequence that minimizes the total metric $M(y, w)$ and the corresponding symbols are output by the Viterbi receiver.

Depending on the application, we may want to evaluate the Viterbi receiver performance using some distortion measure $\bar{d} = \mathbf{E}\{d(w_k, \hat{w}_k)\}$, where $d(w_k, \hat{w}_k)$ is the distortion characteristic of the symbol w_k which the receiver identifies as $\hat{w}_k$. For example, if we wish to find a symbol error probability, then the distortion characteristic $d(w_k, \hat{w}_k)$ is equal to zero if there are no errors in the symbol ($u_k = \hat{u}_k$) and is equal to one otherwise. If we wish to evaluate a bit error probability, then $d(w_k, \hat{w}_k) = m_k/b$, where m_k is the number of bit errors in the symbol $\hat{u}_k$ (the Hamming distance between u_k and $\hat{u}_k$), and b is the total number of bits in the symbol.

III. PERFORMANCE UNION BOUND

In order to find the distortion measure union bound, consider an error event of length L (see Ref. 1) that is a pair of the correct path

$\sigma_L = \{w_k\}$ and an incorrect path $\hat{\sigma}_L = \{\hat{w}_k\}$ such that $s_k \neq \hat{s}_k$ for $k = 1, 2, \cdots, L-1$ and $s_k = \hat{s}_k$ otherwise.

Then the distortion measure is upper bounded by the union bound:[2]

$$\bar{d} \leq \mathbf{E}_u \sum_{L=1}^{\infty} \sum_{\hat{\sigma}_L} P_L(\hat{\sigma}_L/\sigma_L) d_L(\sigma_L, \hat{\sigma}_L),$$

where the average is taken over all source sequences, $P_L(\hat{\sigma}_L/\sigma_L)$ is the conditional probability that the incorrect path $\hat{\sigma}_L$ has a larger metric than the correct path σ_L and 0.5 the probability of equality of metrics if a tie is resolved randomly:

$$P_L(\hat{\sigma}_L/\sigma_L) = \Pr\left\{\sum_{k=0}^{L-1} \Delta m(y_k, w_k, \hat{w}_k) > 0\right\} + 0.5 \Pr\left\{\sum_{k=0}^{L-1} \Delta m(y_k, w_k, \hat{w}_k) = 0\right\}, \quad (3)$$

$$\Delta m(y_k, w_k, \hat{w}_k) = m(y_k, \hat{w}_k) - m(y_k, w_k),$$

$d_L(\sigma_L, \hat{\sigma}_L)$ is the total distortion along the incorrect path:

$$d_L(\sigma_L, \hat{\sigma}_L) = \sum_{k=0}^{L-1} d(w_k, \hat{w}_k).$$

Consider first a discrete memoryless channel. Define a generating function of a variable $\Delta m(y_k, w_k, \hat{w}_k)$:

$$D(z; w_k, \hat{w}_k) = \sum_{y_k} \Pr\{y_k/w_k\} z^{\Delta m(y_k, w_k k, \hat{w}_k)}. \quad (4)$$

The generating function of the sum of variables $\Delta m(y_k, w_k, \hat{w}_k)$ is equal to the product of the generating functions (4), and the probability $P_L(\hat{\sigma}_L/\sigma_L)$ is equal to the sum of the coefficients of the positive power terms of the product and one half of the zero power term. Using contour integrals, we may express the sum as

$$P_L(\hat{\sigma}_L/\sigma_L) = \frac{1}{2\pi j} \int_{|s|=\rho} \left(\frac{1}{z-1} - \frac{1}{2z}\right) \prod_{k=0}^{L-1} D(z; w_k, \hat{w}_k) dz, \quad (5)$$

where $\rho > 1$. The total distortion along the incorrect path $\hat{\sigma}_L$ may also be expressed using generating functions:

$$\frac{d}{dz} \prod_{k=0}^{L-1} z^{d(w_k, \hat{w}_k)} \Big|_{z=1}, \quad (6)$$

and therefore the average distortion is bounded by

$$\bar{d} \leq \frac{1}{2\pi j} \frac{d}{dz} \left\{ \int_C \left(\frac{1}{v-1} - \frac{1}{2v}\right) G(z, v) dv \right\} \Bigg|_{s=1}, \quad (7)$$

where

$$G(z, v) = \mathbf{E}_u \sum_{L=1}^{\infty} \sum_{\hat{\sigma}_L} \prod_{k=0}^{L-1} D(v;\, w_k,\, \hat{w}_k) z^{d(w_k, \hat{w}_k)}, \tag{8}$$

$C \in \{v{:}R_G \cap |v| > 1\}$, R_G is the region of convergence of (8). The right-hand side of the inequality (7) is a new expression of the union bound of the average distortion $\bar{d}$.

The generating function $G(z, v)$ may be found from the system state transition graph with branch weights

$$D(z;\, w_k,\, \hat{w}_k) z^{d(w_k, \hat{w}_k)} \tag{9}$$

or the corresponding matrix equation.[1-3] The system symmetry simplifies the construction of the generating function.

In the case of a continuous channel we may also use a transform (characteristic function) technique to obtain a union bound similar to (7). For example, if $h(x, n) = x + n$, where n is a zero-mean Gaussian variable with one-sided spectral density N_0 (AWGN) and Euclidean metric is used by a Viterbi algorithm, then (3) takes on the form

$$P_L(\hat{\sigma}_L/\sigma_L) = 0.5 \,\mathrm{erfc}(\beta\delta), \tag{10}$$

where $\beta^2 = E_s/N_0$ is the signal-to-noise ratio,

$$\delta^2 = \sum_{k=0}^{L-1} \| f(\hat{w}_k) - f(w_k) \|^2. \tag{11}$$

Using Laplace transform,[4]

$$\int_a^{\infty} \frac{e^{-pt}dt}{t\sqrt{t-a}} = \frac{\pi}{\sqrt{a}} \,\mathrm{erfc}(\sqrt{ap}),$$

which after substitutions $\sqrt{a} = \beta$, $\sqrt{p} = \delta$, $\sqrt{t-a} = v$, takes on the form

$$\mathrm{erfc}(\beta\delta) = \frac{2\beta}{\pi} \int_0^{\infty} \frac{1}{(\beta^2 + v^2)} e^{-\delta^2(\beta^2+v^2)} dv.$$

We obtain from (10) and (11)

$$P_L(\hat{\sigma}_L/\sigma_L) = \frac{\beta}{\pi} \int_0^{\infty} \frac{1}{(\beta^2 + v^2)} \prod_{k=0}^{L-1} D_1(v;\, w_k,\, \hat{w}_k) dv, \tag{12}$$

where

$$D_1(v;\, w_k,\, \hat{w}_k) = e^{-\| f(\hat{w}_k) - f(w_k) \|^2 (\beta^2 + v^2)}. \tag{13}$$

Equation (12) is similar to (5), and therefore we may derive a new union bound for the analog case that is similar to (7):

$$\bar{d} \leq \frac{\beta}{\pi} \frac{d}{dz} \left\{ \int_0^\infty \frac{1}{(\beta^2 + v^2)} G_1(z, v) dv \right\} \Bigg|_{z=1}, \tag{14}$$

where

$$G_1(z, v) = \mathbf{E}_u \sum_{L=1}^{\infty} \sum_{\hat{\sigma}_L} \prod_{k=0}^{L-1} D_1(v; w_k, \hat{w}_k) z^{d(w_k, \hat{w}_k)}.$$

IV. EXAMPLES

4.1 Hard-decision convolutional code

Consider the binary convolutional code shown in Fig. 1 with generator polynomials $\{1 + D + D^2, 1 + D^2\}$ (Ref. 1), and the binary symmetric channel with the bit error probability $p = 1 - q < 0.5$. For each input bit u_k the encoder generates two channel bits $x_k = (x_k', x_k'')$ and the encoder state is defined by the two input bits $s_k = (u_{k-1}, u_{k-2})$ and therefore $w_k = (u_k, u_{k-1}, u_{k-2})$ is simply the contents of the shift registers. According to Fig. 1 the equation $x_k = f(w_k)$ is equivalent to the following two equations:

$$x_k' = u_k + u_{k-1} + u_{k-2}$$

$$x_k'' = u_k + u_{k-2},$$

where "+" denotes *mod*2 addition (exclusive-or).

The second equation of the system (1) $s_{k+1} = g(w_k)$ may also be expressed as the system

$$s_{k+1}' = u_k$$

$$s_{k+1}'' = u_{k-1}.$$

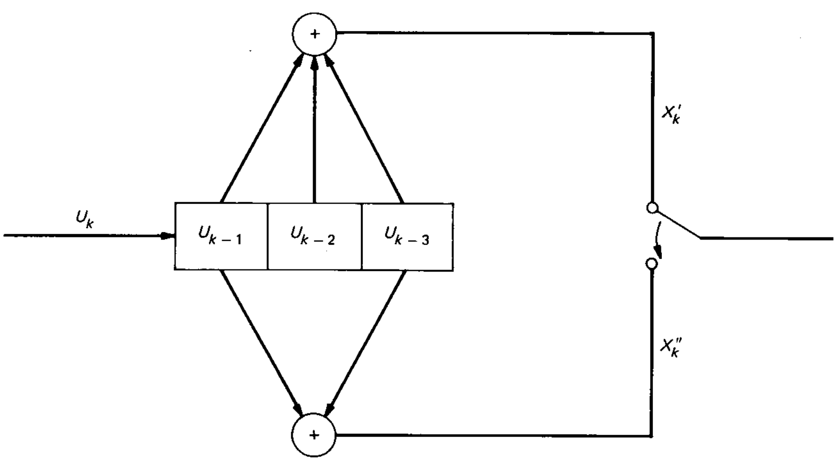

Fig. 1—Convolutional encoder.

However, usually it is represented by the system state graph whose nodes correspond to the system states, and the symbols along the transition lines indicate the encoder input symbols (see Fig. 2).

The next step is to define the algorithm branch metric. Suppose that we received a sequence $y_0, y_1, \cdots, y_N$. Then the MLR will output the sequence $\hat{u}_0, \hat{u}_1, \cdots, \hat{u}_N$, which maximizes the likelihood probability

$$\Pr\{y_0, \cdots, y_N/u_0, \cdots, u_N\} = q^{2N}(p/q)^z,$$

where z is the number of bit errors in the sequence $x_0, x_1, \cdots, x_N$ or, in other words, the Hamming distance (HD) between the sequences: $z = \text{HD}\{(y_0, \cdots, y_N); (x_0, \cdots, x_N)\}$. Since $0 < (p/q) < 1$, maximization of the likelihood is equivalent to minimization of the HD. Therefore we may define the algorithm metric as $m(\hat{w}, w) = \text{HD}\{y; x\}$. This metric depends only on the signal difference.

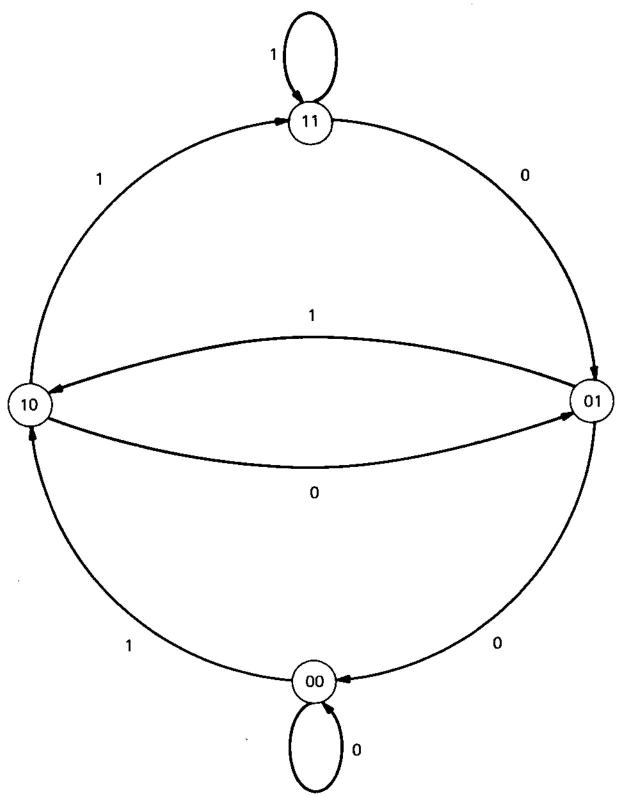

Fig. 2—Encoder state graph.

The channel model may be expressed by eq. (2), which takes the form $y_k = x_k + n_k$, where $n_k = (n_k', n_k'')$ has the following distribution:

$$\Pr\{n_k = (0, 0)\} = p^2, \quad \Pr\{n_k = (1, 0)\} = pq,$$

$$\Pr\{n_k = (0, 1)\} = pq, \quad \Pr\{n_k = (1, 1)\} = q^2.$$

We wish to evaluate the decoder output error probability. Therefore the distortion characteristic is

$$d(\hat{w}, w) = \begin{cases} 0 & \text{if} \quad u = \hat{u} \\ 1 & \text{if} \quad u \neq \hat{u} \end{cases}.$$

This distortion characteristic depends only on the signal difference.

We see that the convolutional code is described in terms of the discrete-time system defined above, and therefore the bit error probability union bound may be found from (7). Because the code is linear, the branch metric and the distortion characteristic depend only on the signal difference. Therefore the generating function (8) is independent of the encoder input sequence and the averaging in (7) is not needed.

We assume that the all-zero sequence is transmitted. The generating function (4) now takes on the form

$$D(v; w, \hat{w}) = (pv + qv^{-1})^{\mu(\hat{w})},$$

where $w = (0, 0, 0)$, $\mu(\hat{w}) = \text{HD}\{(0, 0); \hat{x}\}$ is the Hamming weight of $\hat{x} = f(\hat{w})$. The distortion measure $d(\hat{w}, w) = \hat{u}$; therefore $z^{d(w,\hat{w})} = z^{\hat{u}}$ and

$$D(z; w, \hat{w})z^{d(w,\hat{w})} = z^{\hat{u}}(pv + qv^{-1})^{\mu(\hat{w})}.$$

If we use these values as branch weights on the system state graph, where the all-zero state is split into an initial and final state[1] as shown in Fig. 3, we obtain

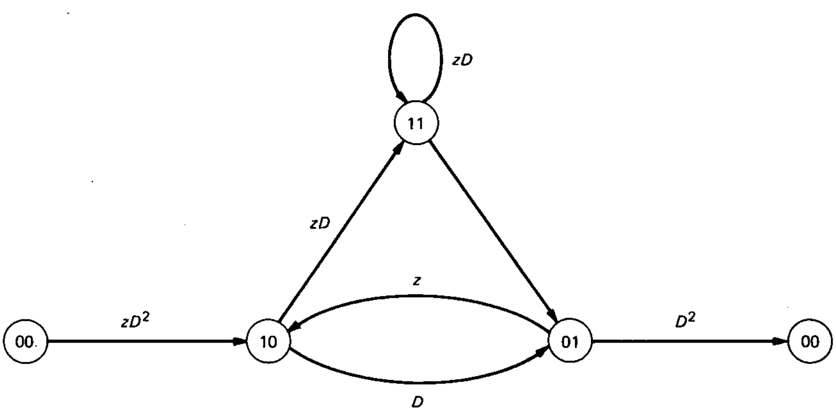

Fig. 3—Signal flow graph of the encoder. $D = (pv + qv^{-1})$.

$$G(z, v) = \frac{z(pv + qv^{-1})^5}{1 - 2z(pv + qv^{-1})},$$

$$|2z(pv + qv^{-1})| < 1$$

as the transition from the initial state to the final state.

The union bound for the bit error probability is found from (7):

$$p_b \le \frac{1}{2\pi j} \int_C \left(\frac{1}{v - 1} - \frac{1}{2v} \right) \frac{(pv + qv^{-1})^5}{[1 - 2(pv + qv^{-1})]^2} \, dv$$

$$C \in \{v: |2(pv + qv^{-1})| < 1 \cap |v| > 1\}.$$

Applying the residue theorem to the previous integral, we obtain

$$p_b \le 2^{-5}[5 - (q - p)(5 - 96pq)(1 - 16pq)^{-1.5} + 14p + 12p^2 - 8p^3],$$

which coincides with the bound derived in Ref. 5.

4.2 Soft-decision convolutional code

In this example we find the union bound on the bit error probability for the same convolutional code but with soft-decision decoding. Suppose that symbols x_k are BPSK modulated. The receive demodulator is followed by a soft-decision sampler with infinite precision.

The likelihood density of receiving $(y_0, \cdots, y_N)$ if $(x_0, \cdots, x_N)$ was transmitted is

$$\psi(y_0, \cdots, y_N/x_0, \cdots, x_N) = \left(\frac{\beta}{\sqrt{\pi}} \right)^{2N} e^{-\delta^2 \beta^2},$$

where

$$\delta = \sqrt{\sum_{k=0}^{N} (y_k' - x_k')^2 + (y_k'' - x_k'')^2}$$

is the Euclidean distance (11) between the signals and $\beta^2 = E_s/N_0$ is the signal-to-noise ratio.

It is clear that the maximum likelihood corresponds to the minimum of the Euclidean distance and vice versa. Therefore we may define the MLR metric as $m(\hat{w}, w) = \| x - \hat{x} \|^2 = (x' - x'')^2 + (x'' - \hat{x}')^2$. This metric depends only on the signal difference.

Using the same arguments as in the previous example, we may assume that the all-zero sequence was transmitted. According to (10) and (13)

$$D_1(v; w, \hat{w}) = e^{-(\beta^2 + v^2)\mu(\hat{w})}.$$

The union bound for p_b is found from (14):

$$p_b \le \frac{\beta}{\pi} \int_0^\infty \frac{e^{-5(\beta^2 + v^2)} \, dv}{(\beta^2 + v^2)[1 - 2e^{-(\beta^2 + v^2)}]^2}.$$

If we use the inequality

$$\frac{1}{1 - 2e^{-(\beta^2+v^2)}} \le \frac{1}{1 - 2e^{-\beta^2}},$$

we obtain

$$p_b < 0.5(1 - 2e^{-\beta^2})^{-2}\mathrm{erfc}(\beta\sqrt{5}),$$

which coincides with the well-known result.[1] Using the more general inequality

$$\frac{1}{(a + q)^2} \le \frac{1}{a^2} \sum_{k=0}^{2n} (k + 1) \left(\frac{-q}{a}\right)^k,$$

where $a = 1 - 2e^{-\beta^2}$, $(\beta \ge \sqrt{\ln 2})$, and $q = 2e^{-\beta^2}(1 - e^{-v^2})$, we may obtain better approximation. For $n = 2$

$$p_b < 0.5a^{-2}[A\ \mathrm{erfc}(\beta\sqrt{5}) + B\ \mathrm{erfc}(\beta\sqrt{6}) + C\ \mathrm{erfc}(\beta\sqrt{7})],$$

where $A = 1 - 2a^{-1}e^{-\beta^2} + 4a^{-2}e^{-2\beta^2}$, $B = 2a^{-1} - 8a^{-2}e^{-\beta^2}$, and $C = 4a^{-2}$.

V. SUMMARY

Using contour integrals we have derived a closed-form expression of the union bound on Viterbi algorithm performance. The transform methods developed in this paper may be used for some other applications. For example, using eq. (12), a sum

$$S = \sum_k a_k\ \mathrm{erfc}(\beta\sqrt{k})$$

may be expressed as

$$S = \frac{2\beta}{\pi} \int_0^\infty \frac{F(e^{-(\beta^2+v^2)})}{(\beta^2 + v^2)}\, dv,$$

where $F(z) = \sum_k a_k z^k$ is the sequence generating function.

The union bound on the performance characteristic of hard- and soft-decision codes was found as an illustration.

VI. ACKNOWLEDGMENTS

The author wishes to thank D. Leed and the referees for their valuable suggestions and comments.

REFERENCES

1. A. J. Viterbi, "Convolutional Codes and Their Performance in Communications Systems," IEEE Trans. Commun. Tech., *COM-19* (October 1971), pp. 751–72.
2. J. K. Omura, "Performance Bounds for Viterbi Algorithms," ICC'81 Conf. Record, Denver, June 1981, pp. 2.2.1–.5.

3. G. D. Forney, Jr., "Maximum-Likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interference," IEEE Trans. Inform. Theory., *IT-18* (May 1972), pp. 363–78.
4. F. Oberhettinger and L. Badii, *Tables of Laplace Transform*, New York: Springer-Verlag, 1973, p. 16, eq. 2.34.
5. K. A. Post, "Explicit Evaluation of Viterbi's Union Bounds on Convolutional Code Performance for the Binary Symmetric Channel," IEEE Trans. Inform. Theory., *IT-23* (May 1977), pp. 403–4.

AUTHOR

William Turin, M.S. (Mathematics), 1958, Odessa State University; Ph.D., (Mathematics), 1966, Institute for Problems of Mechanics, Academy of Sciences, USSR; AT&T Bell Laboratories, 1981—. Mr. Turin has worked at New York University and was an Associate Professor at the Moscow Institute of Electrical Engineering and Telecommunications. His research activities include modeling satellite communication channels, simulation, design and analysis of error-correcting codes, and analysis of Markov processes. He is the author of three books. Senior member, IEEE.