Intraframe Sequential Picture Coding

John A. Stuller
Missouri University of Science and Technology, stuller@mst.edu

Bernd Kurz

Intraframe Sequential Picture Coding

JOHN A. STULLER AND BERND KURZ

Abstract—**This paper generalizes time-discrete autoregressive source coding results of rate-distortion theory to two dimensions. A 2-D discrete autoregressive source is defined and shown to produce a 2-D wide-sense Markovian field. The rate distortion function of the source is then obtained under assumption of Gaussian field statistics and a squared error fidelity criterion. A procedure for generating an ensemble of 2-D codewords whose statistics satisfy the variational equations for *R*(*D*) is given. These 2-D codewords are, by space-time mappings, 1-D *tree* codes, and it is noted that a tree coding theorem of Jelinek, Berger, Davis and Hellman applies. The problem of instrumenting nearly optimum 2-D sequential encoding is discussed briefly. The paper stresses potential application to image coder design.**

1. INTRODUCTION

EFFORTS to apply Shannon's rate-distortion theory to the derivation of optimum intraframe image coders have heretofore dealt primarily with two-dimensional (2-D) block coders [1-3]. These have the necessary characteristic for any optimum image coder of exploiting two-dimensional field redundancy in order to reduce transmitter data rate [4]. It would appear that block codes provided the first application of rate-distortion theory to image coder design partly because, historically, corresponding results for 1-D block codes were relatively well known [5]. Moreover, the conceptual bridge between block coding from one to two dimensions is an easy one to cross. Matters are no longer so clear when one attempts to generalize one-dimensional *sequential* coding to two dimensions. What, after all, is a two-dimensional "sequence" or a two-dimensional "code tree"?

On the other hand, the first (and still most prevalent) intraframe image coders were sequential [6]. These originated relatively independently of information-theoretic analyses: PCM, DPCM, and ADPCM are all in this class. Their primary merit is a simplicity of instrumentation that, with few exceptions [7], has been achieved by making no effort to exploit image redundancy in the direction orthogonal to the line scans. Rate-distortion derived sequential coders for 1-D processes have been devised [8, 9], but no attempt seems to have been made to apply these to the line scan processes of imagery. The reason for this may be partly because of claims that the simpler intuitive coders perform nearly optimally [10] among the class of processors that act independently on each picture line. It is interesting to note, however, that Cutler's ad-hoc delayed source encoder [11] exploits a 1-D code-tree search not unlike that of the 1-D information-theoretic derived coders. Here, however, redundancy among the picture lines is ignored.

To help advance future picture coding research, it would appear useful to have available a theory that describes optimum sequential coding of 2-D discrete processes. Before outlining our approach to this theory, we summarize the pertinent definitions and results that have been established in the 1-D sequential coding case [12, pp. 207-241]. These begin with an autoregressive source model.

A discrete time *autoregressive source* of order M is defined by the 1-D sequence $\{x_t\}$ generated by

$$x_t = -\sum_{k=1}^{M} a_k x_{t-k} + z_t, \qquad t = 1, 2, \cdots \tag{1}$$

where $\{z_t\}$ is a white random sequence, $a_1, \cdots a_M$ are autoregression constants, and $x_0, x_{-1}, \cdots x_{1-M}$ are initial conditions. One can view the sum in (1) as the linear minimum mean square error estimate of x_t given all x_s, $s < t$, and the term z_t as the resulting estimation error, uncorrelated with all x_s, $s < t$. Thus the random sequence $\{x_t\}$ is wide-sense Markov–M [13]. Gray [14] found that for independent $N(0, \sigma_z^2)$ random variables z_t the mean square error (MSE) rate distortion function of $\{x_t\}$ is given parametrically by

$$D_\theta = \frac{1}{2\pi}\int_{-\pi}^{+\pi} \min\,(\theta, S_x(\omega))\, d\omega$$

$$R(D_\theta) = \frac{1}{2\pi}\int_{-\pi}^{+\pi} \max\left(0, \frac{1}{2}\log_2 \frac{S_x(\omega)}{\theta}\right) d\omega$$

where $S_x(\omega)$ is the effective power spectrum of $\{x_t\}$

$$S_x(\omega) = \sigma_z^2 \left| \sum_{k=0}^{M} a_k e^{-jk\omega} \right|^{-2}$$

where $a_0 \triangleq 1$. Berger [12, Thm. 6.3.4] has shown that for distortion $D \leqslant \inf S_x(\omega)$, $R(D)$ can be (approximately) achieved by a code having tree structure of the form illustrated in Figure 1. Moreover, he has shown how to generate an ensemble of such tree codes in which at least one member achieves $R(D)$. The optimum source encoder compares the entire source output sequence $\{x_t\}$ with every word y of its code tree. When the particular codeword, y_o, that minimizes the squared error fidelity criterion is found, the encoder transmits the corresponding path map digits to the receiver from which y_o is recovered. The tree has α path digits per node and β codeword letters per branch. Thus y_o is transmitted at rate $R = \beta^{-1} \log_2 \alpha$ bits per t. Optimum source encoding as such is not instrumentable since an exhaustive search must

Paper approved by the Editor for Communication Theory of the IEEE Communications Society for publication without oral presentation. Manuscript received June 1, 1976; revised December 21, 1976. This work was supported by the National Research Council of Canada under Grant A8251.

The authors are with the Department of Electrical Engineering, University of New Brunswick, Fredericton, N.B., Canada.

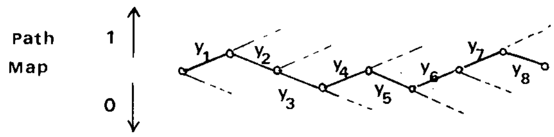

Fig. 1. A Typical Code Tree. In the tree shown there are $\alpha = 2$ branches per node and $\beta = 1$ letters per branch. Quantities $y_1, y_2, y_3, \cdots, y_8$ shown denote one of the 2^8 words of this tree. This word is represented by path map sequence 10010110.

be made over every word of the tree. Anderson and Bodie [9] consider the design of a nearly optimum instrumentable encoder for the 1-D autoregressive source.

The present paper generalizes the theory outlined above to two-dimensional fields. A two-dimensional discrete autoregressive source (8) is defined in Section 2. Section 3 shows that the field it produces is wide-sense 2-D Markov [15]. The MSE rate distortion $R_x(D)$ for Gaussian field statistics is obtained in Section 4. Section 5 gives the explicit generation procedure for the 2-D code tree ensemble leading to $R_x(D)$, and discusses a nearly optimum instrumentable coder. Most of the material in Sections 4-5 rely upon the techniques of Gray and Berger as described in [12, pp. 207-241] of which we assume the reader's familiarity.

2. AUTOREGRESSIVE SOURCE MODEL OF LINE SCANNED PICTURE

Figure 2 illustrates the raster index labeling convention that is used in the source model definition. The raster is $M + I$ rows (or picture lines) by $N + J + K$ columns, with row and column indices (m, n) elements of the set

$$R = \{(m, n) \colon 1 - I \leqslant m \leqslant M,\ 1 - J \leqslant n \leqslant N + K\} \tag{2}$$

where I, J, and K are non-negative integers. The raster is partitioned into two regions:

(i) a *line scan* region defined by index set

$$L = \{(m, n) \colon 1 \leqslant m \leqslant M,\ 1 \leqslant n \leqslant N\}, \tag{3a}$$

and

(ii) a *border* region defined by index set

$$B = R \cap L^c = R - L. \tag{3b}$$

Two-dimensional indices (m, n) in L are associated with a one-dimensional time index $t = 1, 2, \cdots, MN$ to represent the one-to-one mapping of space onto time attendant with line scanning in L. This mapping is given formally by

$$(m, n)_t = (r(t), c(t)); \qquad 1 \leqslant t \leqslant MN, \tag{4a}$$

where

$$r(t) = \mathrm{Int}\left(\frac{t-1}{N}\right) + 1^* \tag{4b}$$

and

*The symbol Int $[\alpha]$ denotes the integer part of α.

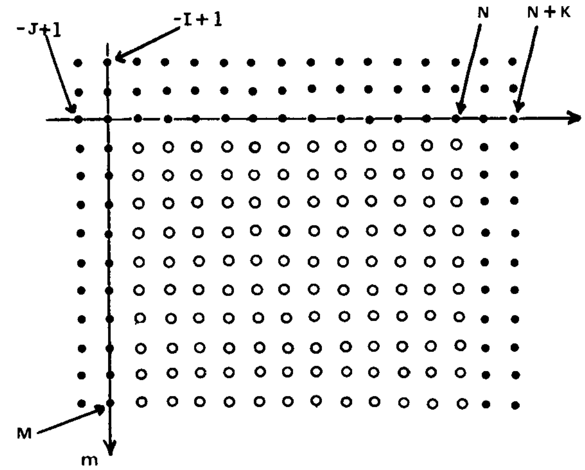

Fig. 2. Raster Sample Center Labeling Convention. Elements denoted by • belong to the border set B of (3b). Those denoted by ○ belong to the line scan set L (3a). In the illustration $I = 3, J = 2, K = 2$.

$$c(t) = (t - 1) \bmod N + 1. \tag{4c}$$

Quantities $r(t)$ are $c(t)$ are, respectively, the rth line and cth column in L at t. The inverse of (4) is

$$t = (m - 1)N + n; \qquad (m, n) \in L. \tag{5}$$

At this point it is convenient to introduce index sets S_W, S_E, S and S_0 as illustrated in Figure 3:

$$S_W = \{(i, j) \colon \quad 0 \leqslant i \leqslant I,\ 0 \leqslant j \leqslant J\} \tag{6a}$$

$$S_E = \{(i, j) \colon \quad 1 \leqslant i \leqslant I,\ -K \leqslant j \leqslant -1\} \tag{6b}$$

$$S = S_E \cup S_W \tag{6c}$$

and

$$S_0 = S - (0, 0). \tag{6d}$$

Source autoregression constants are now defined as subscripted constants a_{ij} with (i, j) in S_0. In the sequel we refer to the spatial configuration of autoregression constants as the *autoregression mask*. Later notation will be simplified by also defining constants a_{ij} for $(i, j) \notin S_0$ as

$$a_{00} = 1 \tag{7a}$$

and

$$a_{ij} = 0 \qquad (i, j) \notin S. \tag{7b}$$

Quantities I, J and K in (2) and (6) are the smallest non-negative integers for which all non-zero values of a_{ij} have indices (i, j) in S_0.

We now define the discrete 2-D field producing autoregressive source by the equation that produces its 2-D output field $[x_{mn}] \equiv X_L$:

$$x_{mn} = -\sum_{(i,j) \in S_0}\sum a_{ij} x_{m-i,n-j} + z_{mn}; \qquad (m, n) \in L, \tag{8}$$

where

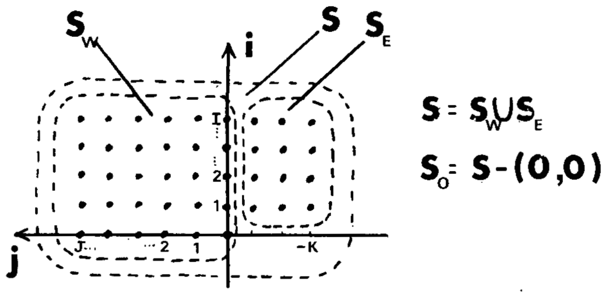

Fig. 3. Sets S_E, S_w, S and S_0. These sets are defined in (6). Set S_0 comprises the indices of the autoregression constants a_{ij}. The spatial configuration of a_{ij}, $(i,j)\epsilon S_0$, is called the *autoregression mask*.

(i) $[z_{mn}]$ is an $M \times N$ array of zero mean uncorrelated random variables (r.v.'s), each having variance $\sigma_z{}^2$;

(ii) the index (m, n) is related to $t = 1, 2, \cdots, MN$ according to (4); and

(iii) boundary values of the x_{pq} are given for all (p, q) in $\mathcal{B}$. In the sequel the set of boundary values $\{x_{pq} : (p, q) \in \mathcal{B}\}$ is called $X_{\mathcal{B}}$.

Insight into the generation of source outputs in *time* $x_{r(t),c(t)} \equiv x_t$ according to (8) follows from visualizing the raster as skewed and then rolled into the form of a cylinder, the raster lines now forming a helix (Figure 4a). There are N values of t per turn in this helix corresponding to the N indices (m, n) per line in L. Boundary region $\mathcal{B}$ comprises the first I turns of the helix as well as a vertical band with $J+K$ vertices per turn running the length of the cylinder. In the process of field generation, the autoregression mask encounters boundary elements $x_{pq} \in X_{\mathcal{B}}$ periodically. It is this fact that makes source (8) distinct from (1).

The close connection between sources (8) and (1) is illustrated in Figure 4b in which an (analog) shift register is used to store quantities x_s needed to produce x_t. Note that the source does not produce outputs in region $\mathcal{B}$ but uses the given boundary values to periodically load the first $J + K$ register cells. One can visualize the 1-D shift register as spiralling down the cylinder as time proceeds, with the individual register contents remaining geometrically adjacent to their field element values on the cylinder. In this configuration, the 1-D register has $(N + J + K)I + J$ storage elements. The memory of the source, however, is actually $NI + J$ since no more than $NI + J$ quantities stored in the register are previously *produced* source outputs. By visualizing a register length equal to the larger value $(N + J + K)I + J$, we obtain a simple way to visualize the dependency of x_t upon both the previously generated outputs and the *given* elements in $X_{\mathcal{B}}$.

Stability tests for two-dimensional equations similar to (8) are described in [16-19]. In the sequel we assume that the process $[x_{mn}]$ becomes asymptotically stationary for (m, n) sufficiently removed from the raster boundaries for $M \to \infty$, $N \to \infty$. In this case a standard calculation reveals that the process is characterized by an effective 2-D power spectrum:

$$S_x(\omega_1, \omega_2) = \frac{1}{g(\omega_1, \omega_2)} \tag{9a}$$

where

$$g(\omega_1, \omega_2) = \frac{1}{\sigma_z{}^2} \left| \sum_{k=0}^{I} \sum_{l=-K}^{J} a_{kl} e^{-j(k\omega_1 + l\omega_2)} \right|^2. \tag{9b}$$

Analysis of the statistical structure of X_L is facilitated by use of vector notation. Let x be the MN element column vector of the x_t:

$$\boldsymbol{x} \triangleq (x_1 x_2 \cdots x_{MN})^T \tag{10a}$$

and similarly, with $z_t = z_{r(t),c(t)}$,

$$\boldsymbol{z} \triangleq (z_1 z_2 \cdots z_{MN})^T. \tag{10b}$$

Equation (8) then assumes the matrix form

$$A\boldsymbol{x} = \boldsymbol{z} + \boldsymbol{b} \tag{11}$$

where A is a lower triangular $MN \times MN$ matrix having block form

$$A = \begin{bmatrix} A_0 & & & & \bigcirc \\ A_1 & A_0 & & & \\ \vdots & A_1 & \ddots & & \\ A_I & \vdots & \ddots & \ddots & \\ & A_I & \ddots & \ddots & \\ \bigcirc & & A_I \cdots & A_1 & A_0 \end{bmatrix}, \tag{12a}$$

$$A_i = \begin{bmatrix} a_{i0} & \cdots & a_{i,-K} & & \bigcirc \\ \vdots & \ddots & & \ddots & \\ a_{iJ} & & \ddots & & a_{i,-K} \\ & \ddots & & & \vdots \\ \bigcirc & & a_{iJ} & \cdots & a_{i0} \end{bmatrix} (N \times N), \tag{12b}$$

and $\boldsymbol{b}$ is a vector of linear combinations of the elements in $X_{\mathcal{B}}$. Since the boundary elements are given constants and z is zero mean, $\boldsymbol{b}$ determines the mean of $\boldsymbol{x}$, η, according to

$$\eta = A^{-1}\boldsymbol{b} \tag{13}$$

where A^{-1} exists since det $A = a_{00}{}^{MN}$ equals unity. The covariance matrix of $\boldsymbol{x}$ is

$$K_x = E\{(x - \eta)(x - \eta)^T\} = \sigma_z{}^2 [A^T A]^{-1}. \tag{14}$$

3. MARKOVIAN PROPERTIES OF X_L

Discrete 2-D Markovian fields are described by Woods [15]. In this section it is shown that the x_{mn} produced by (8) satisfy Woods' conditions for a 2-D Markovian field. It is this property that makes (8) an intuitively reasonable linear model for sampled images.

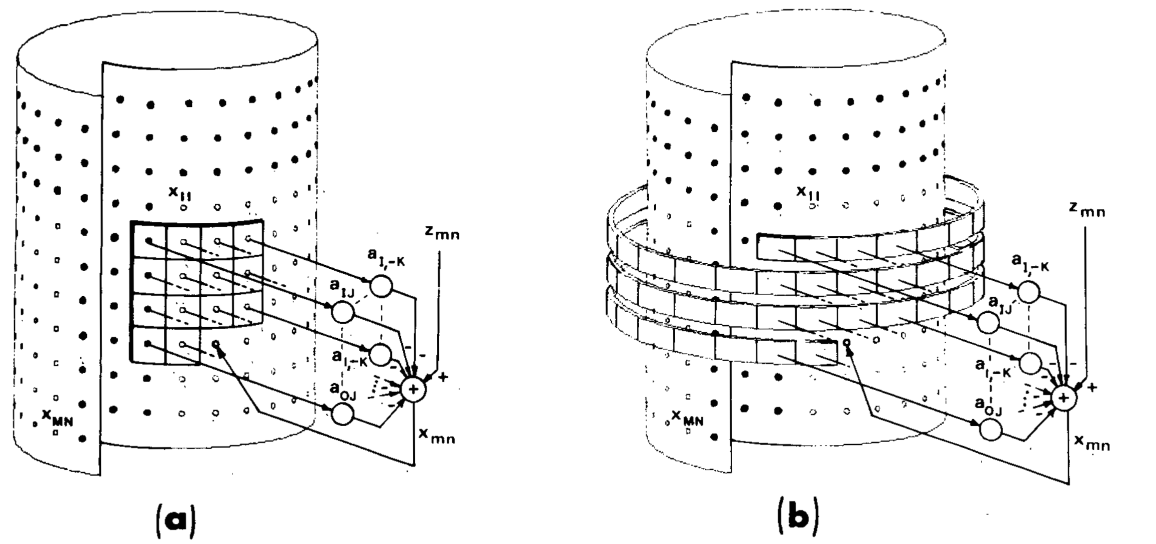

Fig. 4. Conceptualizations of Equation (8): (a) in terms of the autoregression mask; (b) equivalent using 1-D shift register of length $(N + J + K)I + J$. In the illustration, $I = 3, J = 2, K = 1$.

3.1 Preliminaries

Denote the (i, j)th scalar entry of A^TA by β_{ij}, and the (u, v)th scalar entry of A by ξ_{uv}. Some thought will reveal that element ξ_{uv} is the (n, q)th entry of the (m, p)th $N \times N$ submatrix of A, where n, q, m and p are respectively $c(u)$, $c(v)$, $r(u)$ and $r(v)$. Inspection of (12a-b) then indicates that

$$\xi_{uv} = a_{r(u)-r(v),\, c(u)-c(v)} \tag{15}$$

and therefore, by the rule of matrix multiplication

$$\beta_{ij} = \sum_{k=1}^{MN} a_{r(k)-r(i),c(k)-c(i)}a_{r(k)-r(j),c(k)-c(j)}. \tag{16}$$

As k takes values $1, 2, \cdots, MN$, $(r(k), c(k))$ takes values $(1, 1)$, $(1, 2), \cdots, (M, N)$ in L according to mappings (4b-c). Therefore (16) may be rewritten as

$$\beta_{ij} = \sum_{m=1}^{M} \sum_{n=1}^{N} a_{m-r(i),n-c(i)}a_{m-r(j),n-c(j)}. \tag{17}$$

Computation of β_{ij} is shown graphically in Figure 5, where β_{ij} is seen as the two-dimensional autocorrelation of the autoregression mask (including a_{00}). The sum of lagged products in computing β_{ij} is taken over indices $(m, n) \in L$ and for this reason β_{ij} is in general not a function of only $\Delta r \equiv r(i) - r(j)$ and $\Delta c \equiv c(i) - c(j)$. It can be seen from the figure, however, that for (i, j) sufficiently interior to the "borders" of L, β_{ij} assumes the simpler form

$$\gamma_{ij} = \sum_{u=0}^{I} \sum_{v=-K}^{J} a_{uv}a_{u+\Delta r,v+\Delta c} \triangleq \phi(\Delta r, \Delta c). \tag{18}$$

The function $\phi(p, q)$ equals zero for (p, q) outside the region $\{|\,p\,| \leqslant I, |\,q\,| \leqslant J + K\}$ and can also equal zero for other (p, q) depending upon the values of the autoregression constants. In subsequent work it will be convenient to denote the set of (p, q) for which ϕ is non-zero as $C \equiv \{(p, q) : \phi(p, q) \neq 0\}$. It will also be convenient to let the least integer upper bound of the set $\{\sqrt{p^2 + q^2} : (p, q) \in C\}$ be denoted by P.

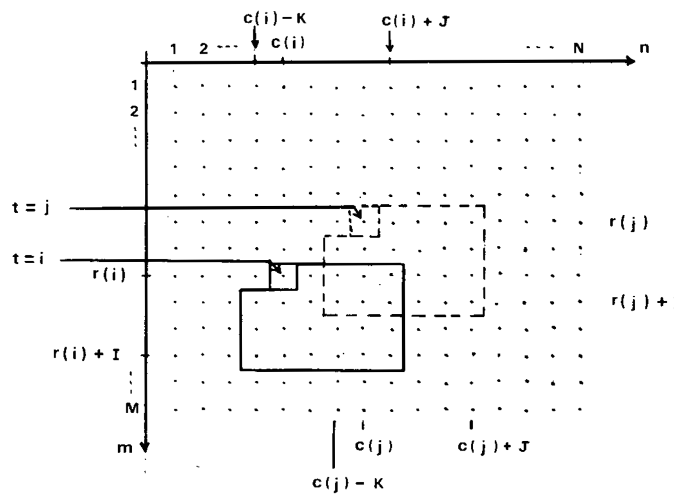

Fig. 5. Graphical Interpretation of Equation (17) for the (i, j)th element β_{ij} in A^TA. The sum of lagged products is taken over all (m, n) in L.

3.2 Two Sided Representation of X_L

Define an $MN \times MN$ matrix H by

$$H = I - \sigma_u^2 K_x^{-1}$$

$$= I - \left(\frac{\sigma_u}{\sigma_z}\right)^2 A^TA \tag{19}$$

where

$$\sigma_u^2 = \sigma_z^2 \left(\sum_{i=0}^{I} \sum_{j=-K}^{J} a_{ij}^2\right)^{-1}. \tag{20}$$

Define an MN component random vector $\boldsymbol{u}$ according to

$$\boldsymbol{x} = H\boldsymbol{x} + \boldsymbol{u}. \tag{21}$$

It follows from definitions (19)-(21) that

$$K_{xu} = K_x [I - H]^T = \sigma_u^2 I \tag{22}$$

and

$$K_u = \sigma_u^2 [I - H] = \sigma_u^4 K_x^{-1} \tag{23}$$

where K_{xu} and K_u are, respectively, cross and autocovariance matrices of $\boldsymbol{u}$.

Direct evaluation of the matrix product in (21) reveals that, except for a width P border interior to the line scan region, field elements x_{mn} are given by

$$x_{mn} = \sum\sum_{(i,j) \in \mathcal{C}} h_{ij} x_{m-i,n-j} + u_{mn} \tag{24a}$$

where

$$h_{pq} = \delta_{pq} - \left(\frac{\sigma_u}{\sigma_z}\right)^2 \phi(p, q), \qquad (p, q) \in \mathcal{C}, \tag{24b}$$

with $\phi(p, q)$ and σ_u^2 as in (18) and (20). The u_{mn} appearing in (24a) are the elements in $\boldsymbol{u}$ of (21) corresponding to mapping (4b-c). Equation (24a) expresses x_{mn} in terms of a linear combination of its immediate surrounds to a depth P plus the non-white "noise" u_{mn}. This is Woods' "two-sided" representation of a discrete 2-D Markov-P process, and we have consequently shown that the process generated by (8) is 2-D (wide-sense) Markov-P, [15, 20]. The "one-sided" representation of the process is, of course, that of equation (8).

It has been noted [20] that an arbitrary Markov-P 2-D discrete field requires a one-sided representation in which element x_{mn} depends upon every x_{ij} in previously generated lines to a depth P lines above x_{mn}. It has been shown in [20] that Markov processes having separable K_x require autoregression constants a_{ij} that are non-zero only in $S_W - (0, 0)$. By including S_E in our definition (8), we have provided a mechanism by which arbitrary Markov processes can be approximated by a one-sided representation in which the dependency of x_{mn} on past x_{ij} does not extend to the raster borders.

4. THE RATE DISTORTION FUNCTION

This section derives the rate distortion function $R_x(D)$ of source (8) for Gaussian z_{mn} under a squared error fidelity criterion. As is well known, $R_x(D)$ is the least number of bits per source letter necessary and sufficient to reproduce $X_{\mathcal{L}}$ with an average MSE D. For discussions of fidelity criteria and the Gaussian assumption in application to picture coding, see [1, 21-23].

Let $\boldsymbol{y}$ denote an MN element random vector of reproduced picture elements corresponding to y_{mn}, $(m, n) \in \mathcal{L}$. The arithmetic average squared error resulting when random $\boldsymbol{x}$ of (11) is reproduced as $\boldsymbol{y}$ is

$$\rho_{MN}(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{MN} \sum_{t=1}^{MN} (x_t - y_t)^2. \tag{25}$$

Let $p(\boldsymbol{x} \mid (\cdot))$ and $q(\boldsymbol{y} \mid (\cdot))$ denote respectively the (conditional) probability density functions of $\boldsymbol{x}$ and $\boldsymbol{y}$ and $p(\boldsymbol{x}, \boldsymbol{y}) = p(\boldsymbol{x})q(\boldsymbol{y} \mid \boldsymbol{x})$ denote their joint density function. We define the rate distortion function of (8) by

$$R_x(D) = \lim_{\substack{M \to \infty \\ N \to \infty}} R_{MN}(D) \tag{26a}$$

where

$$R_{MN}(D) = \frac{1}{MN} \inf_{q \in Q_D} I(q) \tag{26b}$$

with

$$Q_D = \left\{ q(\boldsymbol{y} \mid \boldsymbol{x}, X_{\mathcal{B}}): \iint d\boldsymbol{x}\, d\boldsymbol{y}\, p(\boldsymbol{x} \mid X_{\mathcal{B}}) q(\boldsymbol{y} \mid \boldsymbol{x}, X_{\mathcal{B}}) \rho_{MN}(\boldsymbol{x}, \boldsymbol{y}) \leqslant D \right\} \tag{26c}$$

$$I(q) = \iint d\boldsymbol{x}\, d\boldsymbol{y}\, p(\boldsymbol{x} \mid X_{\mathcal{B}}) q(\boldsymbol{y} \mid \boldsymbol{x}, X_{\mathcal{B}}) \log_2 \frac{q(\boldsymbol{y} \mid \boldsymbol{x}, X_{\mathcal{B}})}{q(\boldsymbol{y} \mid X_{\mathcal{B}})}. \tag{26d}$$

Paralleling Berger [12, p. 225], the initial step in obtaining an explicit form for $R_x(D)$ is to show that $R_{MN}(D)$ does not depend upon the border values $X_{\mathcal{B}}$. This is done by introducing a new MN vector $\boldsymbol{w}$, given (in analogy to (11)) by

$$A\boldsymbol{y} = \boldsymbol{w} + \boldsymbol{b}. \tag{27}$$

In (27), the statistics of $\boldsymbol{y}$ are governed by $q(\boldsymbol{y} \mid X_{\mathcal{B}})$, and $\boldsymbol{b}$ is the vector of linear combinations of elements in $X_{\mathcal{B}}$. Since A is invertible, it follows that for given $X_{\mathcal{B}}$, the mutual information between $\boldsymbol{x}$ and $\boldsymbol{y}$ equals that between $\boldsymbol{z}$ (in (11)) and $\boldsymbol{w}$. By steps identical to those in [12, p. 266] one can now write equivalent expressions for Q_D and $I(q)$ in which $X_{\mathcal{B}}$ does not appear, thereby arriving at the desired conclusion that $R_{MN}(D)$ does not depend upon $X_{\mathcal{B}}$.

Since $X_{\mathcal{B}}$ has no influence on $R_{MN}(D)$, subsequent analyses can assume $\boldsymbol{b} = \boldsymbol{0}$ for which (11) becomes

$$A\boldsymbol{x} = \boldsymbol{z}. \tag{28}$$

$R_{MN}(D)$ for a Gaussian source of the form (28) under a MSE fidelity criterion is well known [12, p. 277]:

$$D_\theta = \frac{1}{MN} \sum_{k=1}^{MN} \min\,[\theta, \lambda_k] \tag{29a}$$

$$R_{MN}(D_\theta) = \frac{1}{MN} \sum_{k=1}^{MN} \max\left[0, \frac{1}{2}\log_2\left(\frac{\lambda_k}{\theta}\right)\right] \tag{29b}$$

where the λ_k are the eigenvalues of K_x. To obtain $R_x(D)$ of (26a) we must take the limit $(M, N) \to (\infty, \infty)$. The reciprocals of the λ_k are by (14) the eigenvalues of $\sigma_z^{-2} A^T A$. Paralleling

[12, pp. 277-231] the solution is to find a (here *block*) Toeplitz matrix Γ having eigenvalues asymptotically equally distributed to those of A^TA. As will be shown, the required matrix is given by

$$\Gamma = [\gamma_{ij}], \qquad MN \times MN, \tag{30}$$

where the γ_{ij} are given by (18) for *all* $1 \leqslant i \leqslant MN$, $1 \leqslant j \leqslant MN$. By (4), $\Delta r = r(i) - r(j)$ is constant for given $i - j$ and $qN + 1 \leqslant i \leqslant qN + N$, $q'N + 1 \leqslant j \leqslant q'N + N$; and $\Delta c = c(i) - c(j)$ is constant for given $i - j$. Since γ_{ij} is a function of only $(\Delta r, \Delta c)$, it follows that Γ is indeed block Toeplitz in M^2 blocks of size $N \times N$.

We obtain the eigenvalues, say α_n, of Γ by considering the ith row of the eigenvector equation $\alpha\xi = \Gamma\xi$:

$$\alpha\xi_i = \sum_{j=1}^{MN} \gamma_{ij}\xi_j, \qquad i = 1, 2, \cdots, MN \tag{31a}$$

from which

$$\alpha\xi_{r(i), c(i)} = \sum_{j=1}^{MN} \phi[r(i) - r(j), c(i) - c(j)]\, \xi_{r(j), c(j)}, \tag{31b}$$

and

$$\alpha\xi_{mn} = \sum_{p=m-1}^{m-M} \sum_{q=n-1}^{n-M} \phi(p, q)\xi_{m-p, n-q}, \qquad (m, n) \in L. \tag{31c}$$

By defining $\xi_{ij} \triangleq 0$ for $(i, j) \notin L$ this becomes

$$\alpha\xi_{mn} = \sum_{p=-I}^{+I} \sum_{q=-(J+K)}^{+(J+K)} \phi(p, q)\xi_{m-p, n-q}. \tag{32}$$

The 2-D version [24] of the Toeplitz distribution theorem [12, p. 112] can now be invoked to reveal that the eigenvalues of Γ are distributed asymptotically as

$$\left| \sum_{m=0}^{I} \sum_{n=-K}^{J} a_{mn} e^{-j(m\omega_1 + n\omega_2)} \right|^2 \equiv \sigma_z^2 g(\omega_1, \omega_2), \qquad (M, N) \to (\infty, \infty), \tag{33}$$

for ω_1, ω_2 in the square $[-\pi, \pi]^2$. The right hand side of (33) follows from (9b). It follows from Appendix 1 that the eigenvalues of A^TA are also distributed asymptotically as in (33). This, (9a) and (29) in turn imply that rate distortion function (26a) has the parametric form

$$D_\theta = \left(\frac{1}{2\pi}\right)^2 \iint_{-\pi}^{+\pi} \min\,[\theta, S_x(\omega_1, \omega_2)]\; d\omega_1\; d\omega_2 \tag{34a}$$

$$R_x(D_\theta) = \left(\frac{1}{2\pi}\right)^2 \iint_{-\pi}^{+\pi} \max\left(0, \frac{1}{2}\log_2 \frac{S_x(\omega_1, \omega_2)}{\theta}\right) d\omega_1\, d\omega_2. \tag{34b}$$

Gray [14] has shown that for 1-D autoregressive source (1), $R_x(D) = R_z(D)$ for $0 \leqslant D \leqslant D_0 \equiv \inf S_x(\omega)$, and $R_x(D) > R_z(D)$ for $D > D_0$. We prove in Appendix 2 that a similar result applies to 2-D autoregressive source (8), namely

$$R_x(D) = \frac{1}{2}\log_2 \frac{\sigma_z^2}{D}; \qquad 0 \leqslant D \leqslant D_0 \equiv \inf S_x(\omega_1, \omega_2) \tag{35a}$$

and

$$R_x(D) > \frac{1}{2}\log_2 \frac{\sigma_z^2}{D}; \qquad D > D_0. \tag{35b}$$

5. IDEAL TREE ENCODING

A random ensemble of tree codes y that achieves $R(D)$ for the 1-D source (1) for $D \leqslant D_0$ was found by Berger [12]. In this section we outline a nearly parallel analysis for the 2-D field. The analysis for the 2-D case encounters two difficulties that do not arise for 1-D fields, and these are emphasized in the sequel.

5.1 The Ideal Codeword Ensemble

A calculation identical to that in [12, p. 237] reveals that codewords $y = A^{-1}w + A^{-1}b$ (27), having statistics solving the variational equation for $R_{MN}(D_\theta)$ (29), $\theta \leqslant \min \lambda_k$, result for

$$K_w = \sigma_z^2 I - DAA^T. \tag{36}$$

Since the eigenvalues of K_w are $\sigma_z^2(1 - D/\lambda_k)$ it follows that K_w is positive definite for $D < \min \lambda_k$. This, in turn implies [25] that a unique invertible lower triangular matrix B exists such that

$$BB^T\sigma_v^2 = K_w \tag{37}$$

for arbitrary $\sigma_v^2 > 0$. Therefore the process w can be obtained by a linear causal transformation on a white Gaussian vector v, with var $\{v_t\} = \sigma_v^2$, $t = 1, 2, \cdots, MN$:

$$w = Bv. \tag{38}$$

Efforts to obtain an exact implementation of B for the 2-D case at hand encounter certain difficulties that do not arise for 1-D fields. The problem can be seen by rewriting (36) in terms of 2-D indices $t \to (r(t), c(t))$. A calculation similar to that of (31) gives the following expression for the covariance $\mu(w_{mn}, w_{ij})$ of w_{mn} and w_{ij}, $(m, n) \in L$, $(i, j) \in L$:

$$\mu(w_{mn}, w_{ij}) = \sigma_z^2 \delta_{mi}\delta_{nj} - D \sum_{p=1}^{M} \sum_{q=1}^{N} a_{m-p, n-q} a_{i-p, j-q}. \tag{39}$$

The last term of (39) has a graphical interpretation similar to that for β_{ij} (17) of Figure 5. Except for certain (m, n) and (i, j) in a width P border interior to the line scan region, (39) assumes the simpler form

$$\mu(w_{mn}, w_{ij}) = \sigma_z{}^2 \delta_{mi}\delta_{nj} - D\phi(m-i, n-j). \tag{40}$$

The 2-D Fourier Transform of (40) gives the effective power spectrum of the w_{mn}:

$$S_w(\omega_1, \omega_2) = \sigma_z{}^2(1 - Dg(\omega_1, \omega_2)) \tag{41}$$

where $g(\omega_1, \omega_2)$ is given by (9b) and ω_1 and ω_2 are in $[-\pi, \pi]$. One can interpret the problem of finding a practical filter to implement (38) as that of obtaining a 2-D filter $B(\omega_1, \omega_2)$ which when driven with white noise yields the process described by (41), i.e.,

$$|B(\omega_1, \omega_2)|^2 \sigma_v{}^2 = \sigma_z{}^2(1 - Dg(\omega_1, \omega_2)). \tag{42}$$

A problem in deriving "one-sided" $B(\omega_1, \omega_2)$ from (42) is that there is no factorization theorem for 2-D polynominals [16]. However, it appears possible to obtain simple causal implementations that closely approximate $|B(\omega_1, \omega_2)|^2$ of (42). An example of this is given by the popular scene covariance model

$$\mu(x_{mn}, x_{ij}) = \sigma_x{}^2 \rho^{|m-i|+|n-j|} \tag{43}$$

for which [20] $S_0 = \{(0, 1), (1, 0), (1, 1)\}$, $a_{01} = a_{10} = -\rho$, $a_{11} = \rho^2$, and

$$\sigma_x{}^2 = (1-\rho^2)^{-2}\sigma_z{}^2. \tag{44}$$

The parameter D_0 of (35a) resulting from (43) is

$$D_0 = \frac{\sigma_z{}^2}{(1+\rho)^4}. \tag{45}$$

Four and five tap 2-D transversal filter approximations to $B(\omega_1, \omega_2)$ for $\rho = .95$ are shown in Figure 6. Here the tap gains b_{ij} were calculated to minimize the sum squared error ϵ^2 between the stationary 2-D covariance terms of (40) and that of the process provided by the filter approximation. For sophisticated design techniques for 2-D filters, see [18, 26].

It is interesting to compare the rate $R_x(D) \equiv R_2(D)$ for the 2-D process characterized by (43) with that (say $R_1(D)$) of an encoder that operates independently on successive line scans. An *individual* line scan of the 2-D process is characterized by the equation

$$x_\tau = \rho x_{\tau-1} + Z_\tau, \qquad \tau = 1, 2, \cdots \tag{46}$$

in which $\{Z_\tau\}$ is a sequence of independent $N(0, \sigma_Z{}^2)$ r.v.'s with

$$\sigma_Z{}^2 = (1-\rho^2)\sigma_x{}^2. \tag{47}$$

Process x_τ of (46) has MSE rate distortion [12, p. 233]

$$R_1(D) = \frac{1}{2}\log\frac{\sigma_Z{}^2}{D}, \qquad 0 \leq D \leq D_1, \tag{48}$$

where

FILTER COEFFICIENTS

$b_{11} = -0.0611$ $b_{10} = 0.1263$

$b_{01} = 0.1263$ $b_{00} = 0.8611$

CORRELATION REQUIRED

-0.0564	0.1131	-0.0564
0.1131	0.7734	0.1131
-0.0564	0.1131	-0.0564

CORRELATION OBTAINED

-0.0526	0.1010	0.0159
0.1010	0.7772	0.1010
0.0159	0.1010	-0.0526

TOTAL SQUARED ERROR $\epsilon^2 = 0.01114$

(a)

FILTER COEFFICIENTS

$b_{11} = -0.0651$ $b_{10} = 0.1441$ $b_{1,-1} = -0.0892$

$b_{01} = 0.1570$ $b_{00} = 0.8466$

CORRELATION REQUIRED

-0.0564	0.1131	-0.0564
0.1131	0.7734	0.1131
-0.0564	0.1131	-0.0564

CORRELATION OBTAINED

0.0	-0.0551	0.1118	-0.0529	-0.0140
0.0058	0.1107	0.7745	0.1107	0.0058
-0.0140	-0.0529	0.1118	-0.0551	0.0

TOTAL SQUARED ERROR $\epsilon^2 = 0.00051$

(b)

Fig. 6. Four and Five Tap 2-D Transversal Filter Approximations to $B(\omega_1, \omega_2)$: (a) Four tap, (b) Five tap. The quantity ϵ^2 is the total squared error in the corresponding covariance approximation. In each case, $\rho = 0.95$, $\sigma_z = \sigma_v = 1$, and $D = 0.0626$.

$$D_1 = \frac{\sigma_Z{}^2}{(1+\rho)^2}. \tag{49}$$

Therefore from (35), (44), (47) and (48) the increase in minimum transmission rate for independent coding of line scans with distortion $D \leq \min(D_0, D_1)$ is $R_1(D) - R_2(D) = 1/2 \log_2 (1-\rho^2)^{-1}$. For $.90 < \rho < .99$ this is an increase of 1.2 to 2.7 bits per pel. This result should be compared to Davisson's in [1].

An ideal tree-code ensemble for source (8) with $(M, N) \to (\infty, \infty)$ and $D \leq D_0$ can be obtained by a procedure identical to that for 1-D source (1) [12, p. 239]. One first chooses integers α and β for rate $\beta^{-1}\log_2 \alpha = R_x(D)$ for some $D \leq D_0$. A tree having α branches per node and β letters per branch is then populated with independent $N(0, \sigma_v{}^2)$ r.v.'s. Each word of the resulting *white code tree* is then transformed by (38) to produce a corresponding non-white vector $\boldsymbol{w}$. This in turn is transformed by A^{-1} (as in (27)) to produce a corresponding codeword $\boldsymbol{y}$. This final operation can be implemented in two dimensions by using the w_{mn} as inputs to the same recursion relation (8) used to generate the x_{mn}. The set of codewords $\{\boldsymbol{y}\}$ so obtained then constitutes a typical tree code in the ensemble for which (35) applies. (Figure 7 illustrates the codeword generation procedure for $\alpha = 2, \beta = 1$.) The rate required to transmit a word from any tree in this ensemble is clearly $\beta^{-1}\log_2 \alpha = R_x$. The proof that at least one code tree $\{y\}$ in this ensemble exists for which $\rho(x, y)$

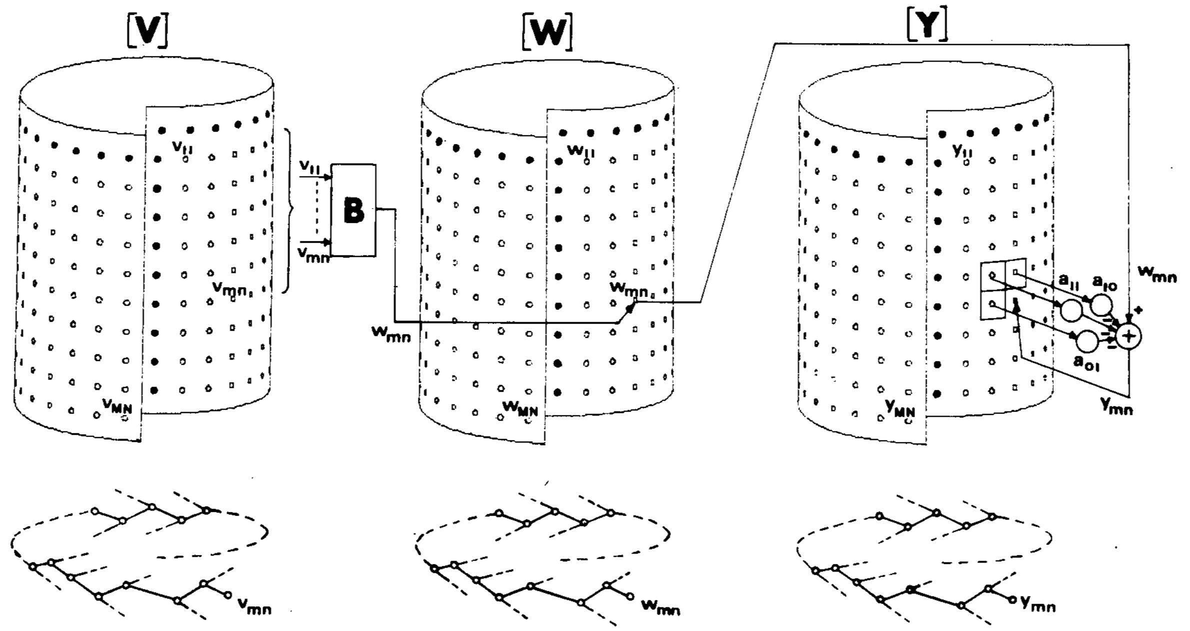

Fig. 7. Optimum Codeword Generation. ($I = 1, J = 1, K = 0$.) Since B is lower triangular, the system it represents is causal. Variates v_{mn} and w_{mn} are not defined in $\mathcal{B}$.

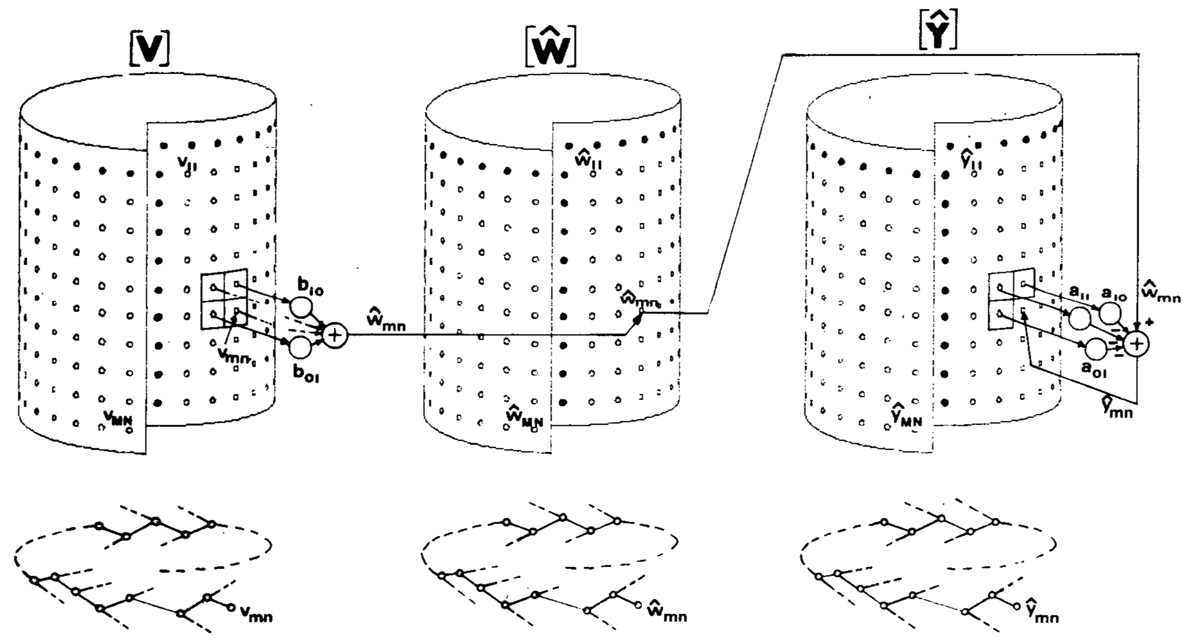

Fig. 8. Suboptimum Codeword Generation. ($I = 1, J = 1, K = 0$.) The transformation B is approximated by a space-invariant 2-D transversal filter. Quantities v_{mn}, $(m, n) \epsilon \mathcal{B}$, are taken as zero.

converges in probability to $D = R_x^{-1}$ as $(M, N) \to (\infty, \infty)$ results from considerations similar to those for 1-D source (1) [12, p. 240 and 27, 28].

Code trees $\{\hat{y}\}$ having statistics nearly equal to those of $\{y\}$ can be generated by using the approximation to B described previously. Figure 8 illustrates the generation of a single word $\hat{y}$ of such a tree by filtering white array $[v_{mn}]$ with 2-D transversal filter b_{ij} to produce $[\hat{w}_{mn}]$. The statistics of $[\hat{w}_{mn}]$ will nearly equal those of $[w_{mn}]$ except for (m, n) within a width P border interior to $\mathcal{L}$. This implies that the increased distortion resulting from $\{\hat{y}\}$ will appear primarily near the raster boundaries. Therefore for large M and N the average distortion $\rho(x, \hat{y})$ will approach D of (35). The effective power spectrum of the codeword process can be deduced from Figure 8 as

$$S_y(\omega_1, \omega_2) = S_w(\omega_1, \omega_2) \left| \sum_{k=0}^{I} \sum_{l=-K}^{J} a_{kl} e^{-j(k\omega_1 + l\omega_2)} \right|^{-2} \tag{50}$$

which is by (9) and (41)

$$S_y(\omega_1, \omega_2) = S_x(\omega_1, \omega_2) - D. \tag{51}$$

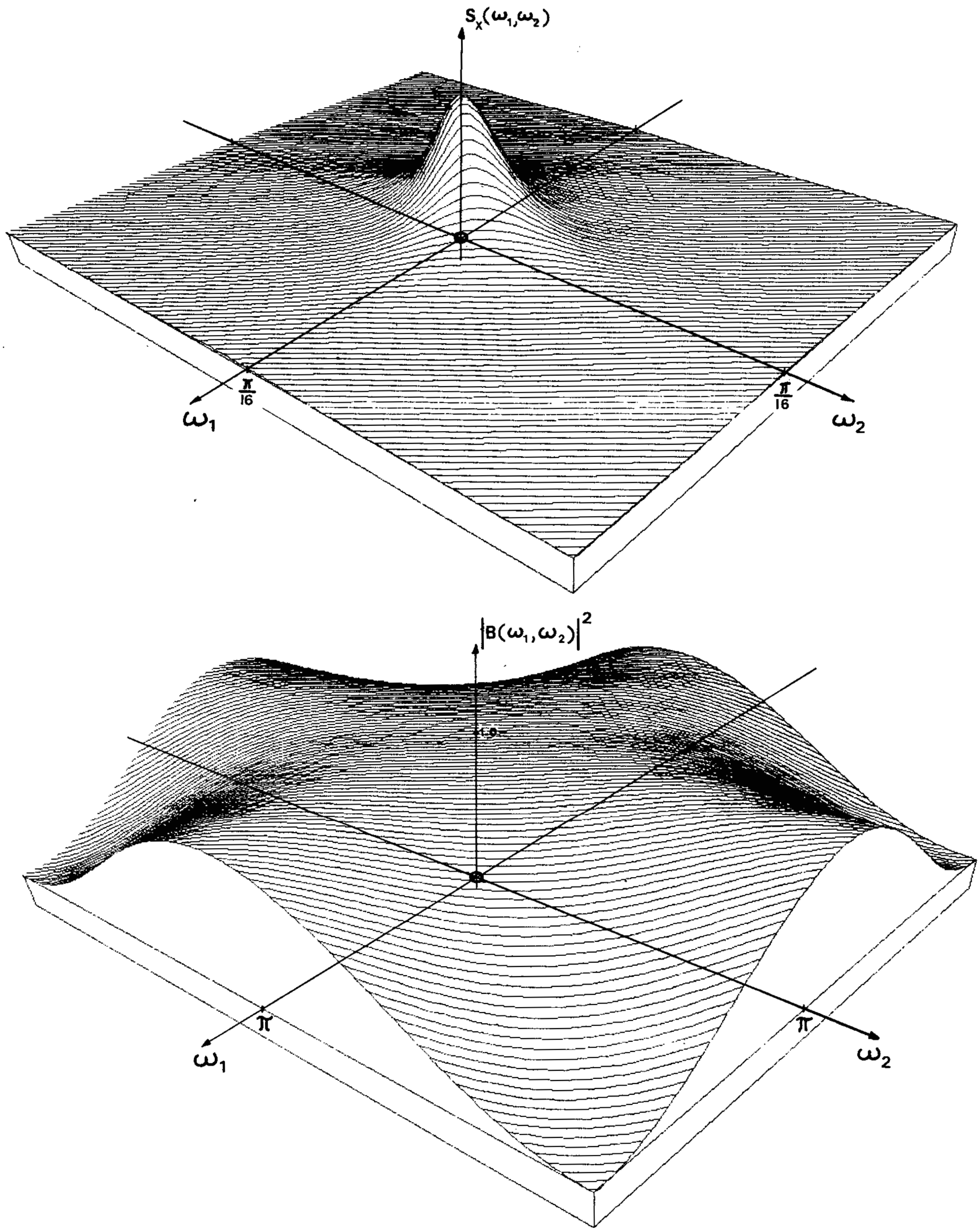

Fig. 9. Perspective Plots of S_x and $S_w \sim |B|^2$ corresponding to Equation (50), for $\rho = 0.95$, $D = 0.0626\sigma_z^2$. S_y is given by $S_y = S_x - D$.

This result parallels that for the 1-D case [12, p. 238]. From (51) follows that in its asymptotically stationary region, process x_{mn} can be expressed as the sum of optimum y_{mn} plus an independent white process having power D. Figure 9 illustrates typical S_x, S_y and S_w for the source characterized by (43). Note that filter b_{ij} has the effect of suppressing the higher frequency components that would otherwise be present in the code tree. Conventional tree encoders such as DM and DPCM are well known to cause "granularity" noise in their scene estimates. This noise may be viewed as a "jitter" at $f_s/2$ in their codeword options. Filter b_{ij} may be viewed heuristically as optimally smoothing (in two dimensions) a similar jitter that would otherwise also be present in $[y_{mn}]$.

5.2 Instrumentation Considerations

Anderson and Bodie [9] consider the problem of instrumentable encoding of a 1-D autoregressive source. Their means to an instrumentable encoder include: (a) quantization of $N(0, \sigma_v^2)$ variate v_t into α possible levels $q_1, q_2, \cdots, q_\alpha$ (one method of choosing these is given by Max [29]); and (b) replacement of exhaustive tree search with a suboptimum but effective method of exploring the tree. The q_i resulting from

(a) then become effectively the α path letters per node of the code tree. Generation (and reconstruction) of any codeword is accomplished simply by applying the path map sequence to a digital filter. Anderson's (M, L) algorithm [12, pp. 216-219] is one means of exploring the tree.

It appears possible to apply similar ideas to the encoding of 2-D sources. In fact, with variates v_{mn} replaced by their quantized versions an encoder bearing some resemblance to the 2-D predictor encoder of Connor *et al.* [7] begins to emerge. Connor found that the pictures resulting from his 2-D schemes were "markedly improved" over those which used only a 1-*D* prediction. The theory of this paper suggests that further improvement will result by the incorporation of a tree search algorithm. The potential of a 1-D tree search in picture coding applications has been stressed on intuitive and experimental grounds by Cutler [11]. In order to exploit the statistical dependencies in images, its seems clear that exploration into the code tree should somehow proceed in both the horizontal and vertical directions. Methods by which this might be accomplished are presently being explored.

APPENDIX 1

The eigenvalues of matrices $A^TA = [\beta_{ij}]$ (17) and $\Gamma = [\gamma_{ij}]$ (18) are shown to be asymptotically equally distributed as $(M, N) \to (\infty, \infty)$. The proof is based upon Theorem 6.3.1 and associated definitions in Berger [12, pp. 228-229].

Let $\mathcal{E}'$ and $\mathcal{E}$ denote the largest magnitude entries of A^TA and Γ respectively:

$$\mathcal{E}' \triangleq \max_{i,j} |\beta_{ij}| = \sum_{i=0}^{I} \sum_{j=-K}^{J} a_{ij}^2, \tag{A.1a}$$

$$\mathcal{E} \triangleq \max_{i,j} |\gamma_{ij}| = \mathcal{E}'. \tag{A.1b}$$

Let $\{\alpha_n'\}$ and $\{\alpha_n\}$ be the sets of MN eigenvalues of A^TA and Γ respectively. The magnitude of the eigenvalues can be upper-bounded by the row norms of the corresponding matrices [30]:

$$|\alpha_n'| \leqslant \max_i \sum_j |\beta_{ij}| \leqslant 4I(J+K)\mathcal{E}' = B, \tag{A.2a}$$

$$|\alpha_n| \leqslant \max_i \sum_j |\gamma_{ij}| \leqslant 4I(J+K)\mathcal{E} = B. \tag{A.2b}$$

From (A.1)-(A.2) one has:

(i) strong norms

$$\|A^TA\| \triangleq \max_n |\alpha_n'| \leqslant B, \tag{A.3a}$$

$$\|\Gamma\| \triangleq \max_n |\alpha_n| \leqslant B, \tag{A.3b}$$

(ii) weak norms

$$|A^TA| \triangleq \sqrt{\frac{1}{NM} \sum_{i=1}^{MN} \sum_{j=1}^{MN} |\beta_{ij}|^2} \leqslant 2\mathcal{E}' \sqrt{I(J+K)} \leqslant B, \tag{A.4a}$$

$$|\Gamma| \triangleq \sqrt{\frac{1}{NM} \sum_{i=1}^{MN} \sum_{j=1}^{MN} |\gamma_{ij}|^2} \leqslant 2\mathcal{E}\sqrt{I(J+K)} \leqslant B. \tag{A.4b}$$

Let $E = [e_{ij}]$ be the $MN \times MN$ difference matrix $E = A^TA - \Gamma$. Denote the largest magnitude entry by $\mathcal{E}_E$. Then

$$\mathcal{E}_E \triangleq \max_{i,j} |e_{ij}| \leqslant \mathcal{E}' + \mathcal{E} = 2\mathcal{E}. \tag{A.5}$$

From (17), (18), and Fig. 5, elements e_{ij} can be non-zero only for points $(r(i), c(i))$, $(r(j), c(j))$ within a width K left (column) border interior to L, a width I bottom (row) border interior to L, and a width J right (column) border interior to L. There are $M(K+J) + (N-K-J)I$ points $(r(i), c(i))$ in this region, and β_{ij} and γ_{ij} are both zero for $(\Delta r, \Delta c)$ outside the $(2I+1)(2J+2K+1)$ element region $\{|\Delta r| \leqslant I, |\Delta c| \leqslant J+K\}$. The weak norm of E is accordingly

$$|E| \triangleq \sqrt{\frac{1}{NM} \sum_{i=1}^{MN} \sum_{j=1}^{MN} |e_{ij}|^2}$$

$$< 2\mathcal{E} \sqrt{\frac{(2I+1)(2J+2K+1)[M(K+J)+(N-K-J)I]}{MN}}, \tag{A.6a}$$

and

$$\lim_{\substack{M\to\infty \\ N\to\infty}} |E| = 0. \tag{A.6b}$$

Properties (A.3), (A.4), and (A.6) establish that the sequence of matrices $\{A^TA\}$ and $\{\Gamma\}$ exhibit mutual approximation $\{A^TA\} \sim \{\Gamma\}$ for both M and N approaching ∞. By use of Theorem 6.3.1 in [12] we find, with $\lim_{MN\to\infty} |A^TA| \leqslant B < \infty$ from (A.4a), that the eigenvalues of A^TA and Γ are asymptotically equally distributed.

APPENDIX 2

To establish (35a), take $\theta \leqslant \inf S(\omega_1, \omega_2)$ for which (34) becomes $D = \theta$, and

$$R_x(D) = \frac{1}{2} \log_2 (\sigma_z^2/D) - \frac{1}{2}\left(\frac{1}{2\pi}\right)^2 \cdot \iint_{-\pi}^{+\pi} \log_2 (\sigma_z^2/S(\omega_1, \omega_2))\, d\omega_1\, d\omega_2. \tag{B.1}$$

$S(\omega_1, \omega_2)$ is related to the autoregression constants by (9). Upon defining $Z_1 = e^{j\omega_1}$, $Z_2 = e^{j\omega_2}$, and $a(Z_1, Z_2) = \sum_{m=0}^{I} \sum_{n=-K}^{J} a_{mn} Z_1^{-m} Z_2^{-n}$, the integral in (B.1) becomes

$$I = \oint_{|Z_2|=1} \frac{dZ_2}{jZ_2} \oint_{|Z_1|=1} \frac{\log_2 |a(Z_1, Z_2)|^2}{jZ_1}\, dZ_1. \tag{B.2}$$

The function $a(Z_1, Z_2)$ has zeros at values of Z_1 equal to say, $Z_{11}, Z_{12}, \cdots, Z_{1I}$, which are themselves functions of Z_2. Some straightforward analysis then yields

$$I = \sum_{i=1}^{I} \oint_{|Z_2|=1} \frac{dZ_2}{jZ_2} \oint_{|Z_1|=1} \frac{\log_2 |Z_1 - Z_{1i}|^2}{jZ_1} dZ_1. \quad \text{(B.3)}$$

A basic theorem of Shanks [16-18] can now be used to deduce that for a stable recursion relation (8), all Z_{1i} have magnitudes less than unity for all Z_2 on $|Z_2| = 1$. Because of this, the contour integral around $|Z_1| = 1$ vanishes exactly as in [12, eq. 6.3.60]. $R_x(D)$ of (B.1) accordingly reduces to (35a).

Inequality (35b) follows directly as in [12, p. 233] upon repeated application of Jensen's inequality.

ACKNOWLEDGMENT

The authors wish to thank John Anderson, Toby Berger, and Robert Gray for carefully reading and commenting on an early draft of this paper. Robert Gray brought to our attention references [2, 23, 28]. A thorough reviewer discovered a technical misstatement concerning the form of Γ in the original submission, and brought reference [19] to our attention. Finally, we thank Ali Habibi for his careful consideration of the paper.

REFERENCES

[1] L. D. Davisson, "Rate-distortion theory and application", *Proc. IEEE*, Vol. 60, pp. 800-808, July 1972.

[2] L. D. Davisson and R. M. Gray, *Data Compression*, Dowden, Hutchinson and Ross, 1976. In particular, see papers 29 (Pratt), and 30 (Habibi and Wintz).

[3] M. Tasto and P. A. Wintz, "A bound on the rate-distortion function and application to images", *IEEE Trans. Inform. Theory*, Vol. IT-18, pp. 150-159, Jan. 1972.

[4] D. J. Sakrison and V. R. Algazi, "Comparison of line-by-line and two-dimensional encoding of random images", *IEEE Trans. Inform. Theory*, Vol. IT-17, pp. 386-398, July 1971.

[5] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968, Thm. 9.7.1.

[6] D. J. Connor, R. C. Brainard, and J. O. Limb, "Intraframe coding for picture transmission", *Proc. IEEE*, Vol. 60, pp. 779-791, July 1972.

[7] D. J. Connor, R. F. W. Pease, and W. G. Scholes, "Television Coding Using Two Dimensional Spatial Prediction", *The Bell System Technical Journal*, Vol. 50, No. 3, pp. 1049-1061, March, 1971.

[8] F. Jelinek and J. B. Anderson, "Instrumentable tree encoding of information sources", *IEEE Trans. Inform. Theory*, Vol. IT-17, pp. 118-119, Jan. 1971.

[9] J. B. Anderson and J. B. Bodie, "Tree encoding of speech", *IEEE Trans. Inform. Theory*, Vol. IT-21, pp. 379-387, July, 1975.

[10] J. B. O'Neal, Jr., "Differential pulse-code modulation (PCM) with entropy coding", *IEEE Trans. Inform. Theory*, Vol. IT-22, pp. 169-174, March 1976.

[11] C. C. Cutler, "Delayed encoding: stabilizer for adaptive coders", *IEEE Trans. Commun. Tech.*, Vol. COM-19, pp. 898-906, Dec. 1971.

[12] T. Berger, *Rate Distortion Theory, A Mathematical Basis for Data Compression*, Englewood Cliffs, N.J.: Prentice-Hall 1971.

[13] A Papoulis, *Probability, Random Variables and Stochastic Processes*, McGraw-Hill 1965, pp. 420-423.

[14] R. M. Gray, "Information rates of autoregressive sources", *IEEE Trans. Inform. Theory*, Vol. IT-16, pp. 412-421, July 1970.

[15] J. W. Woods, "Two dimensional discrete Markovian fields", *IEEE Trans. Inform. Theory*, Vol. IT-18, pp. 232-240, March 1972.

[16] T. S. Huang, W. F., Schreiber and A. J. Tretiak, "Image processing", *Proc. IEEE*, Vol. 59, pp. 1586-1609.

[17] T. S. Huang, "Stability of two-dimensional recursive filters", *IEEE Trans. Audio and Electroacoustics*, Vol. AU-20, pp. 158-163, June 1972.

[18] J. L. Shanks and J. H. Justice, "Stability and synthesis of two-dimensional recursive filters," *IEEE Trans. Audio and Electroacoustics*, Vol. AU-20, pp. 115-128, June 1972.

[19] M. Ekstrom and J. Woods, "Two-dimensional spectral factorization with applications in recursive digital filtering", *IEEE Trans. Acoustics, Speech, and Signal Processing*, Vol. ASSP-24, pp. 115-128, April 1976.

[20] J. Stuller and B. Kurz, "Two dimensional Markov representations of sampled images", *IEEE Trans. Commun.*, Vol. COM-24, pp. 1148-1152, Oct. 1976.

[21] T. G. Stockham, Jr., "Image processing in the context of a visual model", *Proc. IEEE*, Vol. 60, pp. 828-842, July 1972.

[22] J. L. Mannos and D. J. Sakrison, "The effects of a visual fidelity criterion on the encoding of images", *IEEE Trans. Inform. Theory*, Vol. IT-20, pp. 525-536, July 1974.

[23] Z. L. Budrikis, "Visual fidelity criterion and modeling", *Proc. IEEE*, Vol. 60, pp. 771-779, July 1972.

[24] D. J. Sakrison, "An extension of the theorem of Kac, Murdock, and Szego to N dimensions", *IEEE Trans. Inform. Theory* (Corresp.), Vol. IT-15, pp. 608-610, Sept. 1969.

[25] A. Whalen, *Detection of Signals in Noise*. New York and London: Academic Press, 1971, pp. 184-188.

[26] J. V. Hu and L. R. Rabiner, "Design techniques for two-dimensional digital filters", *IEEE Trans. Audio and Electroacoustics*, Vol. AU-20, pp. 249-257, Oct. 1972.

[27] F. Jelinek, "Tree encoding of memoryless time-discrete sources with a fidelity criterion", *IEEE Trans. Inform. Theory*, Vol. IT-15, pp. 584-590, Sept. 1969.

[28] C. R. Davis and M. Hellman, "On tree coding with a fidelity criterion", *IEEE Trans. Inform. Theory*, Vol. IT-21, pp. 373-378, July 1975.

[29] J. Max, "Quantizing for minimum distortion", *IRE Trans. Inform. Theory*, Vol. IT-6, pp. 7-12, Mar. 1960.

[30] Ortega, *Numerical Analysis*, Acadmeic Press 1972, p. 22.

★

John A. Stuller was born in Norwalk, CT, on January 16, 1941. He received the B.S. degree in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1963, the M.S. degree in electrical engineering from the University of Southern California, Los Angeles, in 1967, and the Ph.D. degree in electrical engineering from the University of Connecticut, Storrs, in 1971.

From 1963 to 1965 he was an electrical engineer with the Inertial Guidance Electronics Group of the Nortronics Division of the Northrop Corporation, in Palos Verdes, CA. From 1965 to 1968 he was a member of the technical staff in the Sensors Laboratory of TRW Systems, Redondo Beach, CA. From 1970 to 1971 he was a senior engineer in the Advanced Systems Analysis Group of the Perkin-Elmer Corporation, in Danbury, CT. He joined the University of New Brunswick, Fredericton, N.B., Canada in 1971 where he is presently an Associate Professor in the Department of Electrical Engineering. He was the Director of Graduate Studies in that Department from 1973 to 1975. He has also been a consultant to the Communications Research Centre in Shirley Bay, Ont. His current interests are in the areas of information theory, source encoding, and communications.

Dr. Stuller is a member of Sigma Xi.

★

Bernd Kurz was born in Stuttgart, Germany on January 24, 1943. He received the degree of Dipl.-Ing. in electrical engineering from the University of Stuttgart, Stuttgart, Germany, in 1969 and the M.Sc.E. degree in electrical engineering from the University of New Brunswick, Fredericton, N.B., Canada in 1971. He is presently working for the Ph.D. degree in the electrical engineering Department of the University of New Brunswick.

Since 1971 he has been employed by the Computing Centre of the University of New Brunswick, as Computer Systems Engineer. His technical experience has been in digital systems design and digital signal processing.