

Canonical Coordinates and the Geometry of Inference, Rate, and Capacity

Louis L. Scharf, *Fellow, IEEE*, and Clifford T. Mullis

Abstract—Canonical correlations measure cosines of principal angles between random vectors. These cosines multiplicatively decompose concentration ellipses for second-order filtering and additively decompose information rate for the Gaussian channel. Moreover, they establish a geometrical connection between error covariance, error rate, information rate, and principal angles. There is a limit to how small these angles can be made, and this limit determines channel capacity.

Index Terms—Canonical coordinates, canonical correlations, channel capacity, filtering, information rate.

I. INTRODUCTION

THE STANDARD view of estimation theory and communication is illustrated in Fig. 1. The m -dimensional message \mathbf{x} and the n -dimensional measurement \mathbf{y} are components of the source vector \mathbf{z} . We think of \mathbf{x} as Mother Nature's message and \mathbf{y} as Father Nature's measurement. In the Shannon picture [1], the measurement \mathbf{y} is a "noisy" version of the message \mathbf{x} .

The problems we consider in the context of Fig. 1 are as follows.

- How accurately can the message be estimated from the measurement?
- What is the linear dependence between message and measurement?
- What is the rate as which the measurement carries information about the message?
- What is the capacity of the measurement to carry information about the message?

Our aim in this paper is to answer these questions by showing how the *cosines* for *principal angles* between the message and the measurement determine *error covariance*, *information rate*, and *capacity*. These cosines are just the *canonical correlations* between the *canonical coordinates* of the message and the measurement. This suggests that the system of *canonical coordinates* is the appropriate coordinate system for analyzing the Gaussian channel. As a preview of our results, we offer Fig. 2, which is a redrawing of Fig. 1 in coordinates \mathbf{u} and \mathbf{v} . The trick will be to determine the transformations \mathbf{F} and \mathbf{G} that make \mathbf{u} and \mathbf{v} canonical.

Manuscript received September 25, 1997; revised March 23, 1999. This work was supported by the National Science Foundation under Contracts MIP-9529050 and ECS 9979400 and by the Office of Naval Research under Contracts N00014-89-J-1070 and N00014-00-1-0033. The associate editor coordinating the review of this paper and approving it for publication was Dr. José C. Principe.

The authors are with the Department of Electrical and Computer Engineering, University of Colorado, Boulder, CO 80309-0425 USA (e-mail: scharf@colorado.edu).

Publisher Item Identifier S 1053-587X(00)01534-8.

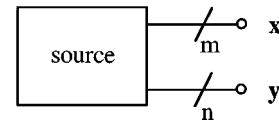


Fig. 1. Source of message and measurement in standard coordinates.

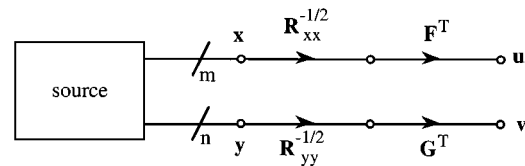


Fig. 2. Source of message and measurement in canonical coordinates.

In the canonical coordinate system, the Gauss–Markov theorem decomposes the MMSE estimator of the message into a *transform coder*, an *equalizer filter* for estimating canonical coordinates, and a *transform decoder*. The error covariances for the canonical coordinates are determined by cosines of principal angles. These cosines also decompose the information rate into a sum of canonical rates, each of which measures the rate at which a canonical coordinate of the measurement carries information about a canonical coordinate of the message. Capacity is determined by the maximum canonical rates that can be achieved, and these are determined by the maximum direction cosines or minimum principal angles that can be achieved.

This paper is a companion to [2]. Our aim is to further explore the algebraic, geometric, and statistical properties of the Shannon experiment [1]. Since completing this paper, we have discovered a relatively obscure paper by Gel'fand and Yaglom [3], which contains some of our results.

II. GEOMETRY AND CANONICAL COORDINATES

We begin our development by defining the source vector \mathbf{z} consisting of the message \mathbf{x} and the measurement \mathbf{y}

$$\mathbf{z} = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \in \mathbb{R}^{m+n}. \quad (1)$$

We will assume that \mathbf{x} and \mathbf{y} have zero means, in which case the second-order characterization of \mathbf{z} is determined by the covariance matrix

$$\mathbf{R}_{zz} = E \left[\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \begin{bmatrix} \mathbf{x}^T & \mathbf{y}^T \end{bmatrix} \right] = \begin{bmatrix} \mathbf{R}_{xx} & \mathbf{R}_{xy} \\ \mathbf{R}_{yx} & \mathbf{R}_{yy} \end{bmatrix}. \quad (2)$$

Whenever we need to assign a probability distribution to \mathbf{z} , we will do so by assuming it to be Gaussian, and we will denote this distribution as $\mathbf{z} : N[\mathbf{0}, \mathbf{R}_{zz}]$. In this case \mathbf{x} and \mathbf{y} are marginally Gaussian, that is, $\mathbf{x} : N[\mathbf{0}, \mathbf{R}_{xx}]$. It is customary to think of the

elements of the cross-covariance matrix \mathbf{R}_{xy} as inner products in the Hilbert space of second-order random variables:

$$(\mathbf{R}_{xy})_{ij} = E[x_i y_j] : \text{inner product between } x_i \text{ and } y_j. \quad (3)$$

If \mathbf{x} and \mathbf{y} are now replaced by their corresponding ‘‘white’’ or ‘‘unit’’ vectors, then the whitened source vector is

$$\underline{\zeta} = \begin{bmatrix} \underline{\mu} \\ \underline{\nu} \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{xx}^{-1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{yy}^{-1/2} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \quad (4)$$

where $\mathbf{R}_{xx}^{-1/2} \mathbf{R}_{xx} \mathbf{R}_{xx}^{-T/2} = \mathbf{I}$, and $\mathbf{R}_{xx}^{1/2} \mathbf{R}_{xx}^{T/2} = \mathbf{R}_{xx}$. The covariance matrix for this whitened vector is

$$\begin{aligned} \mathbf{R}_{\underline{\zeta}\underline{\zeta}} &= E \left[\begin{bmatrix} \underline{\mu} \\ \underline{\nu} \end{bmatrix} \begin{bmatrix} \underline{\mu}^T & \underline{\nu}^T \end{bmatrix} \right] = \begin{bmatrix} \mathbf{I} & \mathbf{C} \\ \mathbf{C}^T & \mathbf{I} \end{bmatrix} \\ \mathbf{C} &= \mathbf{R}_{xx}^{-1/2} \mathbf{R}_{xy} \mathbf{R}_{yy}^{-T/2} \end{aligned} \quad (5)$$

where \mathbf{C} is called the *coherence* matrix. The elements of the coherence matrix are cosines in the Hilbert space of second-order random variables:

$$(\mathbf{C})_{ij} = E[\mu_i \nu_j] : \text{cosine of angle between unit variance random variables } \mu_i \text{ and } \nu_j. \quad (6)$$

This language is evocative, but until we resolve the coherence matrix into an appropriate coordinate system, we have no concrete picture for the underlying geometry. In order to develop this picture, we now determine the singular value decomposition (SVD) of the coherence matrix, namely

$$\mathbf{C} = \mathbf{F} \mathbf{K} \mathbf{G}^T \quad \text{and} \quad \mathbf{K} = \mathbf{F}^T \mathbf{C} \mathbf{G}$$

$$\mathbf{F}^T \mathbf{F} = \mathbf{F} \mathbf{F}^T = \mathbf{I}(m) \quad \text{and} \quad \mathbf{G}^T \mathbf{G} = \mathbf{G} \mathbf{G}^T = \mathbf{I}(n). \quad (7)$$

We then use the orthogonal matrices \mathbf{F} and \mathbf{G} to transform the unit source vector $\underline{\zeta}$ into the *canonical* source vector \mathbf{w}

$$\begin{aligned} \mathbf{w} &= \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} = \begin{bmatrix} \mathbf{F}^T & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^T \end{bmatrix} \begin{bmatrix} \underline{\mu} \\ \underline{\nu} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{F}^T & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^T \end{bmatrix} \begin{bmatrix} \mathbf{R}_{xx}^{-1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{yy}^{-1/2} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}. \end{aligned} \quad (8)$$

The covariance matrix for the canonical source vector is

$$\mathbf{R}_{ww} = E \left[\begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} \begin{bmatrix} \mathbf{u}^T & \mathbf{v}^T \end{bmatrix} \right] = \begin{bmatrix} \mathbf{I} & \mathbf{K} \\ \mathbf{K}^T & \mathbf{I} \end{bmatrix} \quad (9)$$

where the cross-covariance matrix \mathbf{K} is the diagonal matrix of singular values determined from the SVD:

$$\mathbf{K} = \begin{cases} [\mathbf{K}(r) & \mathbf{0}], & m \leq n \\ \begin{bmatrix} \mathbf{K}(r) \\ \mathbf{0} \end{bmatrix}, & m \geq n \end{cases}$$

$$\mathbf{K}(r) = \text{diag}[k(1) \quad \dots \quad k(r)]; \quad r = \min(m, n). \quad (10)$$

The matrix \mathbf{K} is called the *canonical correlation matrix* of *canonical correlations* $k(i)$, and the matrix $\mathbf{K} \mathbf{K}^T$ is called the squared canonical correlation matrix of squared canonical correlations $k^2(i)$ [4], [5]. These squared canonical correlations are eigenvalues of the squared coherence matrix

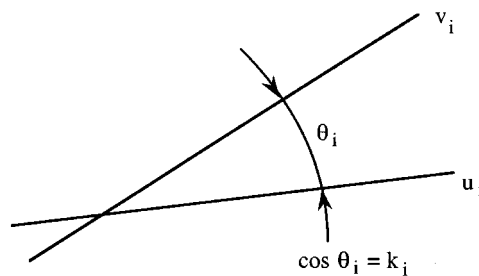


Fig. 3. Geometry of canonical coordinates.

$\mathbf{C} \mathbf{C}^T = \mathbf{R}_{xx}^{-1/2} \mathbf{R}_{xy} \mathbf{R}_{yy}^{-1} \mathbf{R}_{yx} \mathbf{R}_{xx}^{-T/2}$ or, equivalently, of the matrix $\mathbf{R}_{xx}^{-T/2} \mathbf{C} \mathbf{C}^T \mathbf{R}_{xx}^{T/2} = \mathbf{R}_{xx}^{-1} \mathbf{R}_{xy} \mathbf{R}_{yy}^{-1} \mathbf{R}_{yx}$, as the following calculation shows:

$$\mathbf{K} \mathbf{K}^T = \mathbf{F}^T \mathbf{C} \mathbf{G} \mathbf{G}^T \mathbf{C}^T \mathbf{F} = \mathbf{F}^T \mathbf{C} \mathbf{C}^T \mathbf{F}. \quad (11)$$

These eigenvalues are invariant to the choice of a square root for \mathbf{R}_{xx} .

The eigenvalues $k(i)$ are invariant to block-diagonal transformation of \mathbf{R}_{zz} :

$$\begin{bmatrix} \mathbf{T}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_2 \end{bmatrix} \begin{bmatrix} \mathbf{R}_{xx} & \mathbf{R}_{xy} \\ \mathbf{R}_{yx} & \mathbf{R}_{yy} \end{bmatrix} \begin{bmatrix} \mathbf{T}_1^T & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_2^T \end{bmatrix}. \quad (12)$$

In fact, the squared canonical coordinates make up a complete, or maximal, set of invariants for the covariance matrix \mathbf{R}_{zz} under the transformation group

$$\mathcal{T} = \left\{ \mathbf{T} = \begin{bmatrix} \mathbf{T}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_2 \end{bmatrix}, \det(\mathbf{T}) \neq 0 \right\} \quad (13)$$

with group action $\mathbf{R}_{zz} \rightarrow \mathbf{T} \mathbf{R}_{zz} \mathbf{T}^T$. That is, any function of \mathbf{R}_{zz} that is invariant under the transformation $\mathbf{T} \mathbf{R}_{zz} \mathbf{T}$ is a function of $\mathbf{K} \mathbf{K}^T$.

The canonical correlations measure the correlation between the canonical message coordinates and the canonical measurement coordinates. That is, as illustrated in Fig. 3, $k(i)$ is just the cosine of the angle between the canonical message coordinate u_i and the canonical measurement coordinate v_i :

$$\begin{aligned} k(i) &= (\mathbf{K})_{ii} \\ &= E[u_i v_i] : \text{cosine of angle between canonical coordinates } u_i \text{ and } v_i. \end{aligned} \quad (14)$$

The angle between u_i and v_i plays the same role as a principal angle between two linear subspaces. That is, letting $\langle \mathbf{M} \rangle$ and $\langle \mathbf{N} \rangle$ represent m - and n -dimensional orthogonal subspaces of \mathbb{R}^{m+n} , the cosines of the principal angles between $\langle \mathbf{M} \rangle$ and $\langle \mathbf{N} \rangle$ are $k(i)$, which are the diagonal singular values in the SVD of the matrix $\mathbf{M}^T \mathbf{N}$ [6]:

$$\mathbf{M}^T \mathbf{N} = \mathbf{F} \mathbf{K} \mathbf{G}^T. \quad (15)$$

This is the deterministic analog of

$$E[\underline{\mu} \underline{\nu}^T] = E \left[\mathbf{R}_{xx}^{-1/2} \mathbf{x} \mathbf{y}^T \mathbf{R}_{yy}^{-T/2} \right] = \mathbf{C} = \mathbf{F} \mathbf{K} \mathbf{G}^T \quad (16)$$

thereby justifying our interpretation that the canonical correlation $k(i)$ measures the cosine of the i th principal angle between the message \mathbf{x} and the measurement \mathbf{y} . Stated yet another way, the canonical correlations k_i are the cosines of the

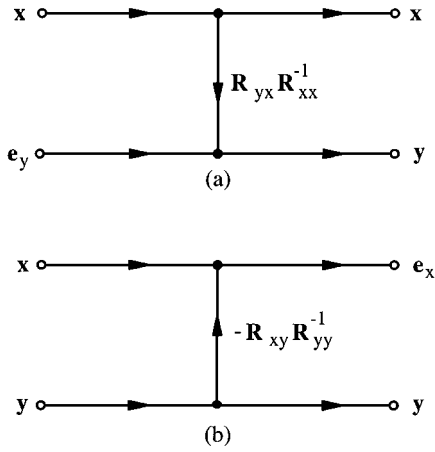


Fig. 4. Source models. (a) Channel model. (b) Filtering model.

canonical angles between the linear subspaces spanned by the canonical message and measurement coordinates $\mathbf{u} = \mathbf{F}^T \underline{\mu}$ and $\mathbf{v} = \mathbf{G}^T \underline{\nu}$. These cosines are invariant to nonsingular transformation of \mathbf{x} by $\mathbf{T}_1 \mathbf{x}$ and \mathbf{y} by $\mathbf{T}_2 \mathbf{y}$. This is consistent with our interpretation of canonical correlations as cosines of principal angles between the message and the measurement: only the principal angles matter, not the internal coordinate systems.

We may now redraw Fig. 1 as Fig. 2 to illustrate the canonical coordinates of the message and the measurement. The connection between \mathbf{z} , the standard coordinates of the source, and \mathbf{w} , the canonical coordinates of the source, is

$$\mathbf{w} = \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} = \begin{bmatrix} \mathbf{F}^T & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^T \end{bmatrix} \begin{bmatrix} \mathbf{R}_{xx}^{-1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{yy}^{-1/2} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \quad (17)$$

and the corresponding connection between their second-order descriptions is

$$\begin{aligned} \mathbf{R}_{wv} &= \begin{bmatrix} \mathbf{F}^T & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^T \end{bmatrix} \begin{bmatrix} \mathbf{R}_{xx}^{-1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{yy}^{-1/2} \end{bmatrix} \begin{bmatrix} \mathbf{R}_{xx} & \mathbf{R}_{xy} \\ \mathbf{R}_{yx} & \mathbf{R}_{yy} \end{bmatrix} \\ &\times \begin{bmatrix} \mathbf{R}_{xx}^{-T/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{yy}^{-T/2} \end{bmatrix} \begin{bmatrix} \mathbf{F} & \mathbf{0} \\ \mathbf{0} & \mathbf{G} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{I} & \mathbf{K} \\ \mathbf{K}^T & \mathbf{I} \end{bmatrix}. \end{aligned} \quad (18)$$

III. FILTERING

The source of Fig. 1 has two equivalent representations. The first is the channel, or signal-plus-noise, model of Fig. 4(a), and the second is the filtering model of Fig. 4(b). In panel Fig. 4(a), the channel noise \mathbf{e}_y has correlation \mathbf{Q}_{yy} , and it is uncorrelated with the message \mathbf{x} . The channel model for the source vector is

$$\mathbf{z} = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{R}_{yx} \mathbf{R}_{xx}^{-1} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{e}_y \end{bmatrix} \quad (19)$$

and the corresponding block Cholesky factorization of the covariance matrix \mathbf{R}_{zz} is

$$\begin{aligned} \mathbf{R}_{zz} &= \begin{bmatrix} \mathbf{R}_{xx} & \mathbf{R}_{xy} \\ \mathbf{R}_{yx} & \mathbf{R}_{yy} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{R}_{yx} \mathbf{R}_{xx}^{-1} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{R}_{xx} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{yy} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{R}_{xx}^{-1} \mathbf{R}_{xy} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \end{aligned}$$

$$\begin{bmatrix} \mathbf{R}_{xx} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{yy} \end{bmatrix} = E \left[\begin{bmatrix} \mathbf{x} \\ \mathbf{e}_y \end{bmatrix} \begin{bmatrix} \mathbf{x}^T & \mathbf{e}_y^T \end{bmatrix} \right] \quad (20)$$

$$\mathbf{Q}_{yy} = \mathbf{R}_{yy} - \mathbf{R}_{yx} \mathbf{R}_{xx}^{-1} \mathbf{R}_{xy}.$$

This factorization produces the model $\mathbf{R}_{yx} \mathbf{R}_{xx}^{-1}$ for the channel filter, the covariance matrix \mathbf{Q}_{yy} for the channel noise, and the following decomposition of $\det[\mathbf{R}_{zz}]$:

$$\det[\mathbf{R}_{zz}] = \det[\mathbf{R}_{xx}] \det[\mathbf{Q}_{yy}]. \quad (21)$$

In Fig. 4(b), the composite source vector is transformed into the filtering error \mathbf{e}_x and the measurement \mathbf{y} . The error has covariance matrix \mathbf{Q}_{xx} , and it is uncorrelated with the measurement \mathbf{y} . The filtering model for the source vector is

$$\begin{bmatrix} \mathbf{e}_x \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & -\mathbf{R}_{xy} \mathbf{R}_{yy}^{-1} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \quad (22)$$

and the corresponding block Cholesky factorization of the covariance matrix \mathbf{R}_{zz} is

$$\begin{aligned} \begin{bmatrix} \mathbf{Q}_{xx} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{yy} \end{bmatrix} &= \begin{bmatrix} \mathbf{I} & -\mathbf{R}_{xy} \mathbf{R}_{yy}^{-1} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \\ &\times \begin{bmatrix} \mathbf{R}_{xx} & \mathbf{R}_{xy} \\ \mathbf{R}_{yx} & \mathbf{R}_{yy} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{R}_{yy}^{-1} \mathbf{R}_{yx} & \mathbf{I} \end{bmatrix} \\ \begin{bmatrix} \mathbf{Q}_{xx} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{yy} \end{bmatrix} &= E \left[\begin{bmatrix} \mathbf{e}_x \\ \mathbf{y} \end{bmatrix} \begin{bmatrix} \mathbf{e}_x^T & \mathbf{y}^T \end{bmatrix} \right] \\ \mathbf{Q}_{xx} &= \mathbf{R}_{xx} - \mathbf{R}_{xy} \mathbf{R}_{yy}^{-1} \mathbf{R}_{yx}. \end{aligned} \quad (23)$$

This factorization produces the model $\mathbf{R}_{xy} \mathbf{R}_{yy}^{-1}$ for the Wiener filter, \mathbf{Q}_{xx} for the error covariance matrix, and the following decomposition of $\det[\mathbf{R}_{zz}]$:

$$\begin{aligned} \det[\mathbf{R}_{zz}] &= \det[\mathbf{Q}_{xx}] \det[\mathbf{R}_{yy}] \\ &= \det[\mathbf{R}_{xx}] \frac{\det[\mathbf{Q}_{xx}]}{\det[\mathbf{R}_{xx}]} \det[\mathbf{R}_{yy}]. \end{aligned} \quad (24)$$

In this decomposition, $\det[\mathbf{R}_{xx}]$ and $\det[\mathbf{R}_{yy}]$ depend only on autocorrelation, and $\det[\mathbf{Q}_{xx}]/\det[\mathbf{R}_{xx}]$ depends on cross-correlation. We will shortly interpret the inverse of this latter quantity as processing gain.

Now let us see how this picture develops in canonical coordinates. The composite canonical source of Fig. 2 has two equivalent representations. The first is the channel, or signal-plus-noise, model of Fig. 5(a), and the second is the filtering model of Fig. 5(b). In Fig. 5(a), the canonical channel noise \mathbf{e}_v has correlation $\mathbf{I} - \mathbf{K}^T \mathbf{K}$, and it is uncorrelated with the canonical message \mathbf{u} . The channel model for the canonical source vector is

$$\mathbf{w} = \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{K}^T & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{e}_v \end{bmatrix} \quad (25)$$

and the corresponding block Cholesky factorization of the covariance matrix \mathbf{R}_{wv} is

$$\begin{aligned} \begin{bmatrix} \mathbf{I} & \mathbf{K} \\ \mathbf{K}^T & \mathbf{I} \end{bmatrix} &= \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{K}^T & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{vv} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{K} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \\ \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{vv} \end{bmatrix} &= E \left[\begin{bmatrix} \mathbf{u} \\ \mathbf{e}_v \end{bmatrix} \begin{bmatrix} \mathbf{u}^T & \mathbf{e}_v^T \end{bmatrix} \right] \\ \mathbf{Q}_{vv} &= \mathbf{I} - \mathbf{K}^T \mathbf{K}. \end{aligned} \quad (26)$$

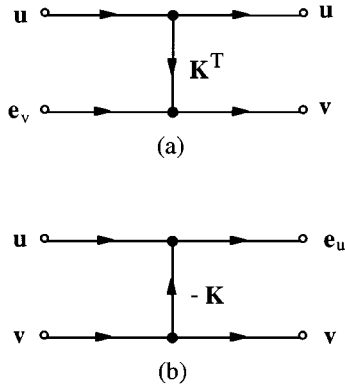


Fig. 5. Canonical source models. (a) Canonical channel model, (b) Canonical filtering model.

This factorization produces the model \mathbf{K}^T for the canonical channel filter, the covariance matrix $\mathbf{Q}_{vv} = \mathbf{I} - \mathbf{K}^T \mathbf{K}$ for the canonical channel noise, and the following decompositions of $\det[\mathbf{R}_{ww}]$ and $\det[\mathbf{R}_{zz}]$:

$$\begin{aligned} \det[\mathbf{R}_{ww}] &= \det[\mathbf{I} - \mathbf{K}^T \mathbf{K}] \quad \text{and} \\ \det[\mathbf{R}_{zz}] &= \det[\mathbf{R}_{xx}] \det[\mathbf{I} - \mathbf{K}^T \mathbf{K}] \det[\mathbf{R}_{yy}]. \end{aligned} \quad (27)$$

In Fig. 5(b), the canonical source vector is transformed into the canonical filtering error \mathbf{e}_u and the canonical measurement \mathbf{v} . The error has covariance matrix $\mathbf{I} - \mathbf{K} \mathbf{K}^T$, and it is uncorrelated with the measurement \mathbf{v} . The filtering model for the canonical source vector is

$$\begin{bmatrix} \mathbf{e}_u \\ \mathbf{v} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & -\mathbf{K} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} \quad (28)$$

and the corresponding block Cholesky factorization of the covariance matrix \mathbf{R}_{ww} is

$$\begin{aligned} \begin{bmatrix} \mathbf{Q}_{uu} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} &= \begin{bmatrix} \mathbf{I} & -\mathbf{K} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{K} \\ \mathbf{K}^T & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{K}^T & \mathbf{0} \end{bmatrix} \\ \begin{bmatrix} \mathbf{Q}_{uu} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} &= E \left[\begin{bmatrix} \mathbf{e}_u \\ \mathbf{v} \end{bmatrix} \begin{bmatrix} \mathbf{e}_u^T & \mathbf{v}^T \end{bmatrix} \right] \\ \mathbf{Q}_{uu} &= \mathbf{I} - \mathbf{K} \mathbf{K}^T. \end{aligned} \quad (29)$$

This factorization produces the model \mathbf{K} for the canonical Wiener filter and $\mathbf{Q}_{uu} = \mathbf{I} - \mathbf{K} \mathbf{K}^T$ for the canonical error covariance matrix.

We may summarize by illustrating the channel and filtering models for the source vector \mathbf{z} in canonical coordinates. These models, which are illustrated in Fig. 6, show that the canonical correlation matrix \mathbf{K} , which may be interpreted as a diagonal equalizer filter, determines the canonical channel filter \mathbf{K}^T and the channel noise covariance $\mathbf{I} - \mathbf{K}^T \mathbf{K}$, as well as the canonical Wiener filter \mathbf{K} and the error covariance matrix $\mathbf{I} - \mathbf{K} \mathbf{K}^T$. With these insights, the standard Shannon picture [1] of Fig. 7(a) may be redrawn as the canonical Shannon picture of Fig. 7(b) to show that the transmitter consists of the whitening transform coder $\mathbf{F}^T \mathbf{R}_{xx}^{-1/2}$, and the receiver consists of the canonical Wiener filter \mathbf{K} followed by the coloring transform decoder $\mathbf{R}_{xx}^{1/2} \mathbf{F}$. The canonical Shannon picture is *automatically* a spread-spectrum picture.

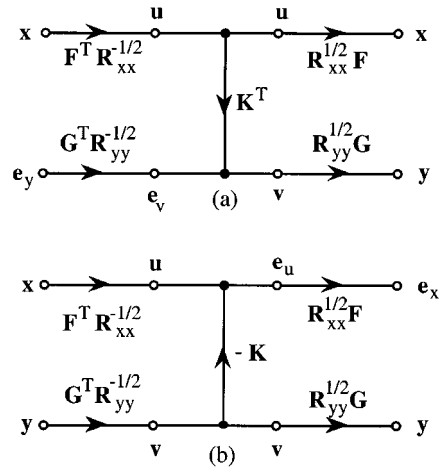


Fig. 6. Source models in canonical coordinates. (a) Channel model. (b) Filtering model.

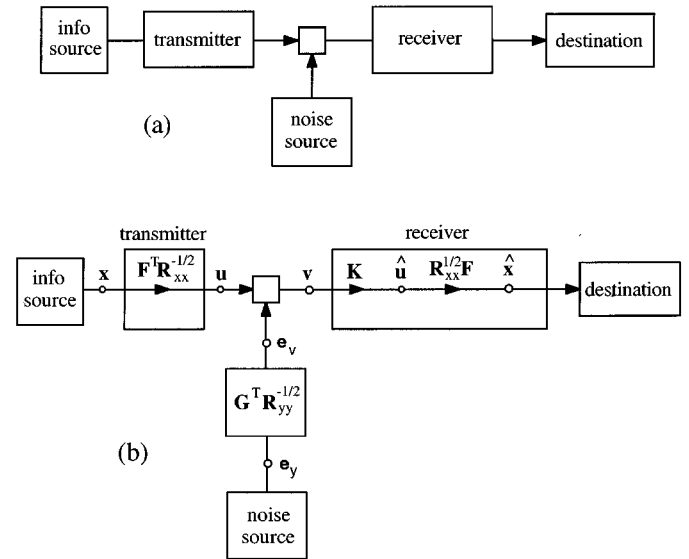


Fig. 7. Shannon's picture. (a) Standard. (b) Canonical.

In canonical coordinates, the Wiener filter and error covariance matrix may be written as

$$\begin{aligned} \mathbf{R}_{xy} \mathbf{R}_{yy}^{-1} &= \mathbf{R}_{xx}^{1/2} \mathbf{F} \mathbf{K} \mathbf{G}^T \mathbf{R}_{yy}^{-1/2} \quad \text{and} \\ \mathbf{Q}_{xx} &= \mathbf{R}_{xx}^{1/2} \mathbf{F} (\mathbf{I} - \mathbf{K} \mathbf{K}^T) \mathbf{F}^T \mathbf{R}_{xx}^{T/2}. \end{aligned} \quad (30)$$

The concentration ellipse for the filtering errors \mathbf{e}_x has volume proportional to $\det[\mathbf{Q}_{xx}]$, and the concentration ellipse for the message \mathbf{x} has volume proportional to $\det[\mathbf{R}_{xx}]$. Their ratio measures the *relative volumes* of these concentration ellipses, and this ratio, which depends only on the canonical correlations or direction cosines, is the same as it is in the canonical coordinate system:

$$\begin{aligned} \frac{\det[\mathbf{Q}_{xx}]}{\det[\mathbf{R}_{xx}]} &= \det[\mathbf{I} - \mathbf{K} \mathbf{K}^T] \\ &= \prod_{i=1}^r (1 - k^2(i)) = \frac{\det[\mathbf{Q}_{uu}]}{\det[\mathbf{R}_{uu}]}. \end{aligned} \quad (31)$$

A physical interpretation is that the canonical coordinate transformation replaces the original composite source by a parallel combination of uncorrelated sources, each of whose error covariance is $1 - k_i^2$. The error covariance for the parallel combination is $\text{diag}[1 - k_1^2, \dots, 1 - k_r^2]$, and the determinant is $\prod(1 - k_i^2)$. In a very real sense, the inverse of the ratio in (31) determines “processing gain,” and it depends only on direction cosines:

$$\text{PG} = \frac{\det[\mathbf{R}_{xx}]}{\det[\mathbf{Q}_{xx}]} = \frac{1}{\prod_{i=1}^r (1 - k_i^2)}. \quad (32)$$

As processing gain is invariant to nonsingular transformation, this is also processing gain for the original experiment.

Example: Signal Plus Noise. The interpretation of canonical coordinates is illuminating when the composite source is a signal-plus-noise source. In this case, the measurement is $\mathbf{y} = \mathbf{x} + \mathbf{n}$ and $E\mathbf{x}\mathbf{n}^T = \mathbf{0}$. Then, the composite correlation matrix is

$$\mathbf{R}_{zz} = \begin{bmatrix} \mathbf{R}_{xx} & \mathbf{R}_{xx} \\ \mathbf{R}_{xx} & \mathbf{R}_{xx} + \mathbf{R}_{nn} \end{bmatrix}. \quad (33)$$

For reasons to become clear, we will define the “signal-to-noise ratio” matrix as

$$\mathbf{S} = \mathbf{R}_{xx}^{T/2} \mathbf{R}_{nn}^{-1} \mathbf{R}_{xx}^{1/2}. \quad (34)$$

Then, with a little algebra, the error covariance matrix may be written as

$$\begin{aligned} \mathbf{Q}_{xx} &= \mathbf{R}_{xx} - \mathbf{R}_{xx}(\mathbf{R}_{xx} + \mathbf{R}_{nn})^{-1} \mathbf{R}_{xx} \\ &= (\mathbf{R}_{xx}^{-1} + \mathbf{R}_{nn}^{-1})^{-1} \\ &= [\mathbf{R}_{xx}^{-T/2} (\mathbf{I} + \mathbf{S}) \mathbf{R}_{xx}^{-1/2}]^{-1} \end{aligned} \quad (35)$$

and the “squared” canonical correlation matrix as

$$\begin{aligned} \mathbf{C}\mathbf{C}^T &= \mathbf{R}_{xx}^{-1/2} \mathbf{R}_{xx} (\mathbf{R}_{xx} + \mathbf{R}_{nn})^{-1} \mathbf{R}_{xx} \mathbf{R}_{xx}^{-T/2} \\ &= (\mathbf{I} + \mathbf{S}^{-1})^{-1}. \end{aligned} \quad (36)$$

This latter identity tells us that the eigenvalues of the SNR matrix \mathbf{S} —call them γ_i^2 —are related to the squared canonical coordinates as

$$\gamma_i^2 = \frac{k_i^2}{1 - k_i^2} \quad \text{or} \quad k_i^2 = \frac{\gamma_i^2}{1 + \gamma_i^2}. \quad (37)$$

This means that the relative volume of concentration ellipses is

$$\frac{\det[\mathbf{Q}_{xx}]}{\det[\mathbf{R}_{xx}]} = \frac{1}{\prod_{i=1}^r (1 + \gamma_i^2)} \quad (38)$$

and the processing gain is $\prod_{i=1}^r (1 + \gamma_i^2)$. The processing gain is 2^r when $\gamma_i = 1$ for all i .

IV. LINEAR DEPENDENCE

The standard measure of linear dependence for the composite random vector \mathbf{z} is the Hadamard ratio inside the inequality

$$0 \leq \frac{\det[\mathbf{R}_{zz}]}{\prod_i (\mathbf{R}_{zz})_{ii}} \leq 1. \quad (39)$$

This ratio takes the value 0 iff there is linear dependence among the z_i ; it takes the value 1 iff \mathbf{R}_{zz} is diagonal, meaning the random variables z_i are all mutually uncorrelated and therefore orthogonal. From the second identity of (27), this ratio may be written as

$$\frac{\det[\mathbf{R}_{zz}]}{\prod_i (\mathbf{R}_{zz})_{ii}} = \frac{\det[\mathbf{R}_{xx}]}{\prod_i (\mathbf{R}_{xx})_{ii}} \det(\mathbf{I} - \mathbf{K}\mathbf{K}^T) \frac{\det[\mathbf{R}_{yy}]}{\prod_i (\mathbf{R}_{yy})_{ii}}. \quad (40)$$

This decomposition of the Hadamard ratio bears comment. The first term measures the linear dependence *among* the random variables $\{x_i\}_1^m$, and the third term measures the linear dependence *among* the random variables $\{y_i\}_1^n$; the middle term measures linear dependence *between* the random variables $\{x_i\}_1^m$ and $\{y_i\}_1^n$. It does so by measuring the error covariance $\mathbf{I} - \mathbf{K}\mathbf{K}^T$ when estimating the canonical message vector $\mathbf{F}^T \mathbf{R}_{xx}^{-1/2} \mathbf{x}$ from the canonical measurement vector $\mathbf{G}^T \mathbf{R}_{yy}^{-1/2} \mathbf{y}$. This error covariance $\det[\mathbf{I} - \mathbf{K}\mathbf{K}^T]$ is also the canonical decomposition of $\det[\mathbf{Q}_{xx}]/\det[\mathbf{R}_{xx}]$.

V. RATE AND CAPACITY

Shannon [1] defines the information rate of the source of Fig. 1 three ways, each of which brings its own interpretations.

- i) $R = H_x - H_{x|y}$: message entropy H_x minus equivocation $H_{x|y}$;
- ii) $R = H_y - H_{y|x}$: measurement entropy H_y minus noise entropy $H_{y|x}$;
- iii) $R = H_x + H_y - H_z$: message entropy H_x plus measurement entropy H_y minus shared entropy H_z .

For the Gaussian source of Fig. 1, entropy is

$$\begin{aligned} H_z &= E[\log f(\mathbf{z})] \\ &= \frac{m+n}{2} \log(2\pi e) + \frac{1}{2} \log \det[\mathbf{R}_{zz}] \end{aligned} \quad (41)$$

and these rate formulas become

i)

$$R = \frac{1}{2} \log \det[\mathbf{R}_{xx}] - \frac{1}{2} \log \det[\mathbf{Q}_{xx}];$$

ii)

$$R = \frac{1}{2} \log \det[\mathbf{R}_{yy}] - \frac{1}{2} \log \det[\mathbf{Q}_{yy}];$$

iii)

$$\begin{aligned} R &= \frac{1}{2} \log \det[\mathbf{R}_{xx}] + \frac{1}{2} \log \det[\mathbf{R}_{yy}] \\ &\quad - \frac{1}{2} \log \det[\mathbf{R}_{zz}]. \end{aligned}$$

Using the determinantal identities of Section III, we may write *equivocation*, *noise entropy*, and *information rate* as

i)

$$\begin{aligned} H_{x|y} &= \frac{m}{2} \log(2\pi e) + \frac{1}{2} \log \det[\mathbf{Q}_{xx}] \\ &= \frac{m}{2} \log(2\pi e) + \frac{1}{2} \log \det[\mathbf{R}_{xx}] - R; \end{aligned}$$

ii)

$$\begin{aligned} H_{y|x} &= \frac{n}{2} \log(2\pi e) + \frac{1}{2} \log \det[\mathbf{Q}_{yy}] \\ &= \frac{n}{2} \log(2\pi e) + \frac{1}{2} \log \det[\mathbf{R}_{yy}] - R; \end{aligned}$$

These are, of course, the spread-spectrum solutions that equalize the signal-plus-noise power across the band. The corresponding capacity is

$$C = \frac{1}{2} \sum_{i=0}^{n-1} \log \frac{P+N}{S_{nn}(i)} \quad (54)$$

and the corresponding error covariance matrix for estimating \mathbf{u} from \mathbf{v} is

$$\mathbf{I} - \mathbf{K}\mathbf{K}^T = \text{diag} \left[\frac{S_{nn}(i)}{P+N} \right]. \quad (55)$$

When the noise is white, meaning $S_{nn}(i) = N$, then the capacity is

$$C = \frac{n}{2} \log \frac{P+N}{N} \quad (56)$$

and the corresponding error covariance matrix is

$$\mathbf{I} - \mathbf{K}\mathbf{K}^T = \text{diag} \left[\frac{N}{P+N} \right]. \quad (57)$$

Under this capacity condition, each canonical measurement coordinate carries information at the same rate $(1/2) \log[(P+N)/N]$, all direction cosines $P/(P+N)$ are equal, and all error variances $N/(P+N)$ are equal.

When only certain DFT frequencies can be used, then n is replaced by r (the dimension of the resulting message), and the capacity formula is

$$C = \frac{r}{2} \log \frac{P+N}{N} \quad (58)$$

which is Shannon's capacity formula.

The asymptotic versions of these formulas are straightforward. For the error covariance matrix $\mathbf{I} - \mathbf{K}\mathbf{K}$, we have

$$\begin{aligned} \lim_{n \rightarrow \infty} \left(\prod_{i=0}^n (\mathbf{I} - \mathbf{K}\mathbf{K}^T)_{ii} \right)^{1/n} &= \int_{-\pi}^{\pi} (1 - |K(e^{j\theta})|^2) \frac{d\theta}{2\pi} \\ &= \int_{-\pi}^{\pi} \frac{S_{nn}(e^{j\theta})}{S_{xx}(e^{j\theta}) + S_{nn}(e^{j\theta})} \frac{d\theta}{2\pi} \end{aligned} \quad (59)$$

where $|K(e^{j\theta})|^2$ is the squared coherence spectrum.

$$\begin{aligned} |K(e^{j\theta})|^2 &= \frac{|S_{xy}(e^{j\theta})|^2}{S_{xx}(e^{j\theta})S_{yy}(e^{j\theta})} \\ &= \frac{S_{xx}(e^{j\theta})}{S_{xx}(e^{j\theta}) + S_{nn}(e^{j\theta})}. \end{aligned} \quad (60)$$

For the rate, we have

$$\begin{aligned} \lim_{n \rightarrow \infty} \left(-\frac{1}{2n} \log \det[\mathbf{I} - \mathbf{K}\mathbf{K}^T] \right) \\ &= \frac{1}{2} \int_{-\pi}^{\pi} \log \frac{1}{1 - |K(e^{j\theta})|^2} \frac{d\theta}{2\pi} \\ &= \frac{1}{2} \int_{-\pi}^{\pi} \log \frac{S_{xx}(e^{j\theta}) + S_{nn}(e^{j\theta})}{S_{nn}(e^{j\theta})} \frac{d\theta}{2\pi}. \end{aligned} \quad (61)$$

If the usable part of the channel has bandwidth $2W$ and the noise power is constant on this band, then the capacity is

$$C = \frac{W}{2\pi} \frac{P+N}{N} \quad (62)$$

TABLE I
SUMMARY OF FORMULAS FOR INFERENCE AND COMMUNICATION.

name	formula
error covariance (standard coordinates)	$\mathbf{R}_{xx}^{1/2} \mathbf{F}(\mathbf{I} - \mathbf{K}\mathbf{K}^T) \mathbf{F}^T \mathbf{R}_{xx}^{T/2}$
error covariance (canonical coordinates)	$\mathbf{I} - \mathbf{K}\mathbf{K}^T$
ratio of concentration volumes	$\det(\mathbf{I} - \mathbf{K}\mathbf{K}^T)$
linear dependence	$\frac{\det[\mathbf{R}_{xx}]}{\prod (\mathbf{R}_{xx})_{ii}} \det(\mathbf{I} - \mathbf{K}\mathbf{K}^T) \frac{\det[\mathbf{R}_{yy}]}{\prod (\mathbf{R}_{yy})_{ii}}$
information rate	$-\frac{1}{2} \log \det[\mathbf{I} - \mathbf{K}\mathbf{K}^T]$

and under this capacity condition, the coherence spectrum, error spectrum, and signal-plus-noise spectra are flat.

$$\begin{aligned} |K(e^{j\theta})|^2 &= \frac{P}{P+N}; \quad 1 - |K(e^{j\theta})|^2 = \frac{N}{P+N} \\ S_{xx}(e^{j\theta}) + S_{nn}(e^{j\theta}) &= P+N. \end{aligned} \quad (63)$$

These formulas illustrate the fundamental role played by canonical coordinates in the computation and interpretation of rate and capacity, and they illustrate the geometry underlying the spectral formulas of [7].

VII. CONCLUSION

Evidently, the canonical coordinate system is the right system for analyzing second-order filtering and communication over the Gaussian channel. In this coordinate system, concentration ellipses are multiplicatively decomposed, and the information rate is additively decomposed into a sum of canonical rates, each of which measures the rate at which a canonical measurement coordinate carries information about a canonical message coordinate. Furthermore, each canonical rate depends only on the direction cosine between a canonical message coordinate and its corresponding canonical measurement coordinate. In the canonical coordinate system, the question of capacity is clarified, and its computation is simplified. In a related paper [2], canonical coordinates are used to solve the rate distortion problem for uniform rounding quantizers.

After all is said and done, the diagonal error covariance matrix $\mathbf{I} - \mathbf{K}\mathbf{K}^T$ determines all performance measures of interest for second-order inference and Gaussian communication. These measures are summarized in Table I.

REFERENCES

- [1] C. E. Shannon, "The mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423; 623–656, 1948, reprinted in C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, Urbana, IL: Univ. of Illinois Press, 1949.
- [2] L. L. Scharf and J. K. Thomas, "Wiener filters in canonical coordinates for transform coding, filtering, and quantizing," *IEEE Trans. Signal Processing*, vol. 46, pp. 647–654, Mar. 1998.
- [3] I. M. Gel'fand and A. M. Yaglom, "Calculation of the amount of information about a random function contained in another such function," in *Amer. Math. Soc. Transl.*, ser. 2, 1959, vol. 12.
- [4] H. Hotelling, "Analysis of a complex pair of statistical variables into principal components," *J. Educ. Psychol.*, vol. 24, pp. 417–441; 498–520, 1933.
- [5] H. Hotelling, "Relations between two sets of variates," *Bimetrika*, vol. 28, pp. 321–377, 1936.

- [6] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 2nd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1989.
- [7] R. A. McDonald and P. M. Schultheiss, "Information rates of Gaussian signals under criteria constraining the error spectrum," *Proc. IEEE*, vol. 52, pp. 415–416, Apr. 1964.
- [8] M. L. Eaton, *Multivariate Statistics: A Vector Space Approach*. New York: Wiley, 1983, ch. 10.

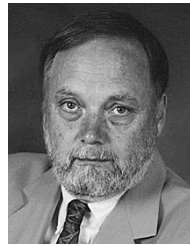


Louis L. Scharf (F'86) received the Ph.D. degree in electrical engineering in 1969 from the University of Washington, Seattle.

From 1969 to 1971, he was a Member of the Technical Staff at Honeywell's Marine Systems Center, Seattle. He served as Professor of Electrical Engineering and Statistics at Colorado State University, Fort Collins, from 1971 to 1981. From 1982 to 1985, he was Professor and Chair of Electrical and Computer Engineering at the University of Rhode Island, Kingston. He is currently Professor of Electrical and

Computer Engineering at the University of Colorado, Boulder, where he teaches and conducts research in signal processing. In 1974, he was Visiting Associate Professor at Duke University, Durham, NC. In 1977, he was a Member of the Technical Staff with the CNRS Laboratoire des Signaux et Systemes, Gif-sur-Yvette, France, and Professeur Associe with the University of South Paris, Orsay, France. In 1981, he was a Visiting Professor at Ecole Nationale Supérieure des Telecommunications, Paris, France, and at the University of La Plata, Buenos Aires, Argentina. He was a Visiting Professor at Institut Eurecom, Sophia-Antipolis, France, in 1992.

Prof. Scharf is a Past Member of the ASSP AdCom. He has served on the Editorial Board of *Signal Processing* and is a Past Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING. He was Technical Program Chairman for the IEEE International Conference on Acoustics, Speech, and Signal Processing in 1980. In 1994, he served as a Distinguished Lecturer for the IEEE Signal Processing Society, and in 1995, he received the Society's Technical Achievement Award.



Clifford T. Mullis received the B.S., M.S., and Ph.D. degrees in electrical engineering from the University of Colorado, Boulder, in 1966, 1968, and 1971, respectively.

He was an Assistant Professor of electrical engineering at Princeton University, Princeton, NJ, from 1971 to 1973. He is now a Professor of electrical engineering at the University of Colorado.