Extended State-Space Model of Discrete-Time Dynamical Systems

JALIL FADAVI-ARDEKANI, MEMBER, IEEE, SANJIT K. MITRA, FELLOW, IEEE, AND BRIAN D. O. ANDERSON, FELLOW, IEEE

Abstract —A model for discrete-time dynamical systems is discussed in which the future values of the internal variables depend on the present and $K-1$ previous time instants, where K is the order of the model. We outline transition matrix and impulse response calculations, equivalences between model and conventional state variable models, methods of constructing such a model, and means for calculating certain quantities relevant to the design of cost-optimal low-roundoff noise digital filters.

I. Introduction

LINEAR dynamical systems are often represented by a state-space description which consists of a set of first-order differential or difference equations involving the state variables and the input, and an output equation given by a linear combination of the state variables and the input. There are, however, a number of dynamical systems which are more naturally described by a set of higher order differential equations involving the internal variables and the input.

When this sort of thinking is carried over to discrete time, one is led to investigate models of the following type:

$$X(n+1)=\sum_{i=0}^{K-1} A_i X(n-i)+\sum_{i=0}^{m} B_i u(n-i) \quad (1.1a)$$

$$y(n)=\sum_{i=0}^{K-1} C_i X(n-i)+\sum_{i=0}^{m} d_i u(n-1) \quad (1.1b)$$

where $X(n)$ is an N-vector of internal variables, A_i's, B_i's are real constant matrices and vectors of appropriate dimensions, and d_i's are real constant scalars.[1] K, the *order* of the model, is an integer greater than or equal to unity; m is also a nonnegative integer independent of K. Without loss of generality, at least one of C_{K-1} and A_{K-1} is nonzero, and at least one of d_m and B_m is nonzero; (note

Manuscript received December 7,1981; March 30, 1982. This work was supported by in part by the National Science Foundation under Grant ECS 79-18028 and by the Australian Research Grants Committee.

J. Fadavi-Ardekani and S. K. Mitra are with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106.

B. D. O. Anderson is with the Department of Systems Engineering, The Australian National University, Canberra ACT 2600, Australia.

[1]There is little interest in digital filtering in studying vector $u(\cdot)$ and $y(\cdot)$. Nevertheless, the theory covers this case provided the obvious changes are made.

0098-4094/82/0800-0547$00.75 ©1982 IEEE

that this assumption does not prevent the actual maximum delays of the input, output and internal variables all being different). Models such as (1.1) can arise from discretization of continuous-time models, but also are of interest in their own right as a basis for optimal digital filter design [1].

When $K=1$, (1.1) is a conventional state-variable model. When $K>1$, we shall term it an extended state (e-state) model.

We investigate the properties of e-state models and our results reduce to the classical case when $K=1$. Moreover, even though we have derived these results for a single-input single-output e-state model, their extension to the multivariate e-state model is straightforward. The organization of the paper is as follows. We first define the transition matrix from a set of recursive equations. The zero-input and zero-state responses are then obtained using the transition matrix, and the transfer function of the model is derived. Questions of minimality and equivalence are considered and a correspondence is set up between conventional and e-state models, and it is shown that any given transfer function has an e-state model realization.

Finally, we outline an application of this model to the design of optimum digital filters which provided the original stimulus for this work.

II. Calculation of the Response of the Model

A. Transition Matrices

Define the *transition matrices* $\mathcal{Q}_{l,j}$ of the e-state model by the recursive equations

$$\mathcal{Q}_{l,j+1}=\sum_{i=0}^{K-1} A_i\mathcal{Q}_{l,j-i}, \qquad l=0,1,\cdots,K-1, \quad j\geqslant 0 \tag{2.1a}$$

with

$$\mathcal{Q}_{l,-j}=I\delta(l-j), \qquad \text{for } l \text{ and } j=0,1,\cdots,K-1. \tag{2.1b}$$

Note that (2.1b) is equivalent to a set of K^2 initial conditions. This definition of the transition matrix reduces to A^j for the conventional state-space model when $K=1$.

Below, we shall indicate a number of ways of viewing the transition matrices which may assist in their calculation. Meanwhile, we indicate their use in describing the response of the e-state model.

B. Zero-Input Response

With the input kept at zero for all the time, the e-state variables at step n in terms of their initial values at instants $k, k-1,\cdots,k-K+1$ are given by

$$X(n)=\sum_{i=0}^{K-1}\mathcal{Q}_{i,n-k}X(k-i), \qquad \text{for } n\geqslant k-K+1. \tag{2.2}$$

This equation provides the basic justification for terming the $\mathcal{Q}_{l,j}$ transition matrices.

We prove the validity of (2.2) by induction. By direct substitution and use of (2.1b) it is easy to show that (2.2) holds for $n=k, k-1,\cdots,k-K+1$. Assume that it holds also for $n, n-1,\cdots,n-K+1$. Then we next show that it holds for $n+1$. From (1.1) and (2.2), we obtain

$$X(n+1)=\sum_{i=0}^{K-1}A_iX(n-i)=\sum_{i=0}^{K-1}A_i\sum_{j=0}^{K-1}\mathcal{Q}_{j,n-i-k}X(k-j)$$
$$=\sum_{j=0}^{K-1}\sum_{i=0}^{K-1}A_i\mathcal{Q}_{j,n-i-k}X(k-j)=\sum_{j=0}^{K-1}\mathcal{Q}_{j,n+1-k}X(k-j)$$

which is identical in form to (2.2). Thus (2.2) is proved.

Substituting (2.2) in (1.5) we obtain the zero-input response at the output as

$$y(n)=\sum_{i=0}^{K-1}C_i\sum_{j=0}^{K-1}\mathcal{Q}_{j,n-k-i}X(k-j). \tag{2.3}$$

C. The Zero-State Response

A zero-state response conventionally is what results when the initial state is zero at $k=0$ and inputs are applied in the interval $k\geqslant 0$. Of course, a zero-state response must have the property that if the input applied in $k\geqslant 0$ is also zero, then the response is identically zero. To obtain this property, it is clearly not enough just to set $X(0)=\mathbf{0}$. One should have also $X(-1)=\cdots=X(-k+1)=\mathbf{0}$ and $u(-1)=\cdots=u(-m)=\mathbf{0}$, save in certain special situations.[2] Under these constraints, consider first the response $h_X(\cdot)$ linking inputs to states, i.e., the impulse response of (1.1a). It is easily shown that

$$h_X(n)=\sum_{i=0}^{K-1}\mathcal{Q}_{0,n-1-i}B_i. \tag{2.4}$$

Then for an input $u(n)$ which is zero for $n<0$

$$X(n)=\sum_{j=1}^{n}h_X(j)u(n-j)=\sum_{j=0}^{n-1}h_X(j+1)u(n-1-j). \tag{2.5}$$

Substituting (2.4) into (2.5) we arrive at

$$X(n)=\sum_{j=0}^{n-1}\left[\sum_{i=0}^{m}\mathcal{Q}_{0,j-i}B_i\right]u(n-1-j), \qquad \text{for } n>0. \tag{2.6}$$

Note that for a k-step delayed input, the zero-state response would have been

$$X(n)=\sum_{j=0}^{n-1-k}\left[\sum_{i=0}^{m}\mathcal{Q}_{0,j-i}B_i\right]u(n-1-j), \qquad \text{for } n>k \tag{2.7}$$

[2]For example, it is possible to get an identically zero response from special nonzero $X(-i)$, $i=1,\cdots,K-1$ if the A_i are singular.

with

$$X(n)=\mathbf{0}, \qquad \text{for } n<k.$$

The zero-state output $y(n)$ in this case will be

$$y(n)=\sum_{l=0}^{K-1} C_l\left[\sum_{j=0}^{n-1-k-l}\left\{\sum_{i=0}^{m} \mathcal{C}_{0,j-1}B_i\right\}u(n-1-j-l)\right] + \sum_{l=0}^{m} d_l u(n-l) \quad (2.8)$$

which because $u(n)=0$ for $n<k$, can be written as

$$y(n)=\sum_{j=0}^{n-1-k}\left[\sum_{l=0}^{K-1} C_l\left\{\sum_{i=0}^{m} \mathcal{C}_{0,j-1}B_i\right\}u(n-1-j-l)\right] + \sum_{l=0}^{m} d_l u(n-l), \qquad \text{for } n\geqslant k. \quad (2.9)$$

D. The Total Response

Due to linearity of the model, we can add the zero-input and the zero-state responses to obtain the total output as

$$y(n)=\sum_{i=0}^{K-1} C_i\left[\sum_{j=0}^{K-1} A_{j,n-k-i}X(k-j)\right] + \sum_{j=0}^{n-1-k}\left[\sum_{l=0}^{k-1} C_l\left\{\sum_{i=0}^{m} \mathcal{C}_{0,j-i}B_i\right\}u(n-1-j-l)\right] + \sum_{l=0}^{m} d_l u(n-l). \quad (2.10)$$

Equation (2.10) gives the total response of the system of the e-state model to an input which is zero for $n<k$, and the initial conditions are given by $X(k),\cdots,X(k-K+1)$.

E. Transfer Function Calculation

From (1.1), it follows at once that

$$H(z)=\left[\sum_{i=0}^{K-1} z^{-i}C_i\right]\left[zI-\sum_{i=0}^{K-1} z^{-i}A_i\right]^{-1}\left[\sum_{i=0}^{m} z^{-i}B_i\right] + \sum_{i=0}^{m} z^{-i}d_i. \quad (2.11)$$

F. Further Properties of the Transition Matrices

The following properties are all fairly easy to prove, and are relevant for calculation of the transition matrix. Proofs will be omitted.

1) *Relation with a Conventional Transition Matrix*: With

$$F=\begin{bmatrix} 0 & I & 0 & \cdots & 0\\ 0 & 0 & I & \cdots & 0\\ & & & \ddots & \\ & & & & I\\ A_{K-1} & A_{K-2} & A_{K-3} & \cdots & A_0 \end{bmatrix} \quad (2.12)$$

then for $j>0$

$$F^j=\begin{bmatrix} \mathcal{C}_{K-1,j-(K-1)} & \mathcal{C}_{K-2,j-(K-1)} & \cdots & \mathcal{C}_{0,j-(K-1)}\\ \vdots & & & \vdots\\ \mathcal{C}_{K-1,j-1} & \mathcal{C}_{K-2,j-1} & \cdots & \mathcal{C}_{0,j-1}\\ \mathcal{C}_{K-1,j} & \mathcal{C}_{K-1,j} & \cdots & \mathcal{C}_{0,j} \end{bmatrix}. \quad (2.13)$$

2) *Frequency-Domain Formulas and Computing the $\mathcal{C}_{l,j}$ from the $\mathcal{C}_{0,j}$*: One has

$$\sum_{j=0}^{\infty} z^{-(j+1)}\mathcal{C}_{l,j+1}=\left[zI-\sum_{i=0}^{K-1} z^{-i}A_i\right]^{-1} \cdot\left[A_l+z^{-1}A_{l+1}+\cdots+z^{-K+1+l}A_{K-1}\right] \quad (2.14)$$

and so

$$\sum_{j=0}^{\infty} z^{-j}\mathcal{C}_{0,j}=z\left(zI-\sum_{i=0}^{K-1} z^{-i}A_i\right)^{-1}. \quad (2.15)$$

It follows that

$$\sum_{j=0}^{\infty} z^{-(j+1)}\mathcal{C}_{l,j+1}=z^{-1}\left(\sum_{j=0}^{\infty} z^{-j}\mathcal{C}_{0,j}\right) \cdot\left(A_l+z^{-1}A_{l+1}+\cdots+z^{-K+1+l}A_{K-1}\right) \quad (2.16)$$

and equating the powers of z^{-1} yields

$$\mathcal{C}_{l,j+1}=\sum_{i=0}^{K-1-l} \mathcal{C}_{0,j-i}A_{l+i}, \qquad j\geqslant 0. \quad (2.17)$$

(Note that with $l=0$, (2.17) gives an alternative recursion for $\mathcal{C}_{0,j}$ to (2.1) and for $l\neq 0$, it implies that only one recursion, for $\mathcal{C}_{0,j}$, need be used to get all $\mathcal{C}_{l,j}$.)

3) *Semigroup Property of Transition Matrices*: Corresponding to the semigroup property $F^{m+n}=F^mF^n$ associated with (2.13), there holds

$$\mathcal{C}_{l,m+n}=\sum_{i=0}^{K-1} \mathcal{C}_{i,m}\mathcal{C}_{l,n-i}. \quad (2.18)$$

G. The Characteristic Polynomial

The characteristic polynomial of (1.1a) is

$$\phi(z)=\det\left[z^K I-\sum_{i=0}^{K-1} A_i z^{K-1-i}\right] \quad (2.19a)$$

$$=\det[zI-F]. \quad (2.19b)$$

If $\phi(z)=\sum_{i=0}^{NK} \alpha_i z^i$, the Cayley–Hamilton theorem and (2.13) imply for $l=0,1,\cdots,K-1$

$$\sum \alpha_i A_{l,i}=0. \quad (2.20)$$

Because of the direct feedthrough term $\sum_{i=0}^{m} d_i u(n-i)$, with

$m > 0$ the system is forced to have a pole at $z=0$. Note that the zeros of $\phi(z)$ may not include this pole.

III. Equivalence and Minimality of e-State Realizations

The ideas underlying equivalence and minimality are set out in books such as [2]–[4].

A. Minimality

In working with (1.1), it is natural from some points of view to work with z^{-1} rather than z in writing down transform quantities. Suppose that we set

$$\begin{aligned}
A_r(z^{-1}) &= I - A_0 z^{-1} - \cdots - A_{K-1} z^{-K}\\
B_r(z^{-1}) &= B_0 z^{-1} + \cdots + B_m z^{-(m+1)}\\
C_r(z^{-1}) &= C_0 - C_1 z^{-1} + C_{K-1} z^{-(K-1)}\\
d_r(z^{-1}) &= d_0 + d_1 z^{-1} + \cdots + d_m z^{-m}
\end{aligned} \tag{3.1}$$

with each expression being regarded as polynomial in z^{-1}, rather than as a finite Laurent series. Then (1.1) has a mixed matrix fraction description (MFD) for its transfer function $H(z)$ as

$$H(z) = C_r(z^{-1}) A_r^{-1}(z^{-1}) B_r(z^{-1}) + d_r(z^{-1}). \tag{3.2}$$

The associated system matrix [3], [4] is

$$P_r(z^{-1}) = \begin{bmatrix} A_r(z^{-1}) & B_r(z^{-1}) \\ -C_r(z^{-1}) & d_r(z^{-1}) \end{bmatrix} \tag{3.3}$$

and the MFD is minimal if and only if $[A_r(z^{-1}) \;\; -B_r(z^{-1})]$ and $[A_r'(z^{-1}) \;\; C_r'(z^{-1})]$ each have full rank for all z^{-1}.

It is on the other hand more conventional to work with MFD's involving polynomials in z rather than z^{-1}. Let us define

$$\begin{aligned}
A(z) &= Iz^K - A_0 z^{K-1} - \cdots - A_{K-1}\\
B(z) &= B_0 z^m + y \cdots + B_m\\
C(z) &= C_0 z^{K-1} + \cdots + C_{K-1}\\
D(z) &= d_0 z^m + \cdots + d_m.
\end{aligned} \tag{3.4}$$

Then it is easily checked that

$$H(z) = C(z) A^{-1}(z) B(z) z^{-m} + D(z) z^{-m}. \tag{3.5}$$

Let $B(z)z^{-m}$ and $C(z)A^{-1}(z)$ be separate transfer functions, with the output of the first, denoted by w, serving as input to the second. Then with ξ_1, ξ_2 denoting partial states, we have, in obvious notation

$$\begin{aligned}
z^m \xi_1 &= u\\
B(z)\xi_1 &= w\\
A(z)\xi_2 &= w\\
y &= C(z)\xi_2 + d_0 u + d_1 z^{-1} u + \cdots + d_m z^{-m} u\\
&= C(z)\xi_2 + d_0 u + d_1 z^{m-1}\xi_1 + \cdots + d_m \xi_1
\end{aligned}$$

whence

$$\begin{bmatrix} A(z) & -B(z) \\ 0 & z^m \end{bmatrix} \begin{bmatrix} \xi_2 \\ \xi_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} u \tag{3.6a}$$

and

$$d_0 u + \begin{bmatrix} C(z) & D(z) - d_0 z^m \end{bmatrix} \begin{bmatrix} \xi_2 \\ \xi_1 \end{bmatrix} = y. \tag{3.6b}$$

Consequently, the system matrix is

$$P(z) = \left[\begin{array}{cc|c} A(z) & -B(z) & 0 \\ 0 & z^m & 1 \\ \hline -C(z) & -D(z) + d_0 z^m & d_0 \end{array}\right]. \tag{3.7}$$

Minimality now holds if

$$K(z) = \begin{bmatrix} A(z) & -B(z) & 0 \\ 0 & z^m & 1 \end{bmatrix} \tag{3.8a}$$

and

$$L(z) = \begin{bmatrix} A(z) & -B(z) \\ 0 & z^m \\ -C(z) & -D(z) - d_0 z^m \end{bmatrix} \tag{3.8b}$$

have full row and column rank for all z. Now $K(z)$ has full row rank for all z if and only if

$$[A(z) \;\; B(z)]$$

has full row rank for all z, while $L(z)$ has full column rank for all z if and only if

$$\begin{bmatrix} A(z) \\ C(z) \end{bmatrix} \text{ and } \begin{bmatrix} -A_{K-1} & -B_m \\ -C_{K-1} & -d_m \end{bmatrix}$$

have full rank for all z (distinguish the cases $z \neq 0$ and $z = 0$).

What is the difference between working with z^{-1} and z in terms of these conditions? Very simple calculations show that for any $z \neq 0$

$$\begin{aligned}
&[A_r(z^{-1}) \;\; B_r(z^{-1})] \text{ has full rank}\\
&\qquad \Leftrightarrow [A(z) \;\; B(z)] \text{ has full rank}
\end{aligned} \tag{3.9a}$$

$$\begin{aligned}
&[A_r'(z^{-1}) \;\; C_r'(z^{-1})] \text{ has full rank}\\
&\qquad \Leftrightarrow [A'(z) \;\; C'(z)] \text{ has full rank.}
\end{aligned} \tag{3.9b}$$

However, differences can arise at $z=0$. The point $z=0$ is not considered in using z^{-1} as a variable, but it is relevant in considering z as a variable. Consequently, we can have minimality with z^{-1} as a variable but not with z as a variable. However, any such nonminimality can only be associated with so called decoupling zeros (the generalization of unobservable or uncontrollable modes appropriate to MFD's) at the origin.

We remark that the dichotomy of choice and subsequent analysis presented by z and z^{-1} is also noted in [5].

B. Removal of Nonminimality: Decoupling Zeros not at Origin

We have just noted that if nonminimality arises in the z-based MFD *other* than with decoupling zeros at the origin, it also arises in the z^{-1} based MFD. We wish to remove it, but retain the structure of the e-state model, including the dimension of the e-state vector. This means that $A_r(0)$ must be the identity. Assume that $[A(z)\quad B(z)]$ has a nontrivial left greatest common divisor with determinantal zero not at $z=0$. The same is true of $[A_r(z^{-1})\quad B_r(z^{-1})]$. Let

$$[A_r(z^{-1})\quad B_r(z^{-1})]=X(z^{-1})[\bar{A}_r(z^{-1})\quad \bar{B}_r(z^{-1})]. \tag{3.10}$$

When $z^{-1}=0$, we see that $X(0)$ is nonsingular. Without loss of generality, $X(0)=I$, so that $\bar{A}_r(0)=0$. Then

$$\bar{X}(n+1)=\sum_{i=0}^{K-1}\bar{A}_i\bar{X}(n-i)+\sum_{i=0}^{m}\bar{B}_iu(n-i)$$
$$y(n)=\sum_{i=0}^{K-1}C_i\bar{X}(n-i)+\sum_{i=0}^{m}d_iu(n-i) \tag{3.11}$$

a similar procedure applies if $[A^t(z)\quad C^t(z)]$ drops in rank for some $z\neq 0$.

C. Removal of Nonminimality: Decoupling Zeros at Origin

Let us now suppose that $[A_r(z^{-1})\quad B_r(z^{-1})]$ and $[A_r^t(z^{-1})\quad C_r^t(z^{-1})]$ have full rank for all z^{-1}, but that $[A(z)\quad B(z)]$ or $L(z)$, see (3.8b), fails to have full rank at the origin. Because of the form of the e-state model (which demands that $A_r(0)=I$ or that $A(z)$ have highest degree term with coefficient I), *we cannot always remove* this nonminimality. We shall now discuss when it can be removed.

Let $V_{11}(z)$ be a greatest common right divisor for

$$\begin{bmatrix} A(z)\\ 0\\ C(z)\end{bmatrix}$$

and let

$$\begin{bmatrix} A(z) & -B(z)\\ 0 & z^m\\ -C(z) & -D(z)+d_0z^m\end{bmatrix}=\begin{bmatrix} \bar{A}(z) & -B(z)\\ 0 & z^m\\ -\bar{C}(z) & -D(z)+d_0z^m\end{bmatrix}\begin{bmatrix} V_{11}(z) & 0\\ 0 & 1\end{bmatrix}.$$

Without loss of generality, we may assume that $\bar{A}(z)$ is column proper. Then, since $CA^{-1}=\bar{C}\bar{A}^{-1}$ is strictly proper, the column degrees of $\bar{C}$ are less than those of $\bar{A}$.

Now suppose that

$$\begin{bmatrix} \bar{A}(0) & -B(0)\\ -\bar{C}(0) & -D(0)\end{bmatrix}$$

fails to have full column rank. Then one can find a polynomial $V_{12}(z)$ and integer $\bar{m}$ such that

$$\begin{bmatrix} A(z) & -B(z)\\ 0 & z^m\\ C(z) & -D(z)+d_0z^m\end{bmatrix}=\begin{bmatrix} \bar{A}(z) & -\bar{B}(z)\\ 0 & z^{\bar{m}}\\ -\bar{C}(z) & -\bar{D}(z)+d_0z^{\bar{m}}\end{bmatrix}\begin{bmatrix} I & V_{12}(z)\\ 0 & z^{m-\bar{m}}\end{bmatrix}$$

with

$$\begin{bmatrix} \bar{A}(0) & -\bar{B}(0)\\ -\bar{C}(0) & -\bar{D}(0)\end{bmatrix}$$

possessing full column rank. This implies that $[\bar{A}(0)\quad -\bar{B}(0)]$ has full row rank ($\bar{D}(0)$ is a scalar), so that the system matrix

$$\left[\begin{array}{cc|c} \bar{A}(z) & -\bar{B}(z) & 0\\ 0 & z^{\bar{m}} & 1\\ \hline -\bar{C}(z) & -\bar{D}(z)+d_0z^{m} & d_0\end{array}\right]$$

corresponds to a minimal MFD. However, it may be that $\bar{A}(z)$ does not have highest degree term with coefficient I. Let the highest degree column have degree $\bar{K}$. Then we can post-multiply $\bar{A}(z)$ and $\bar{C}(z)$ by a matrix with noncontrast determinant $\hat{V}(z)$ such that

$$\hat{A}(z)=\bar{A}(z)\hat{V}(z)$$

and

$$\hat{C}(z)=\bar{C}(z)\hat{V}(z)$$

with $\hat{A}(z)=z^{\bar{K}}I+$lower order terms. Without loss of generality, $\hat{V}(z)$ may be assumed to have all determinantal zeros at the origin. In this way, we get $\hat{A}$, $\bar{B}$, $\hat{C}$, $\bar{D}$ corresponding to an e-state realization of the same e-state dimension as the original. If either $\bar{m}<m$ or $\bar{K}<K$, some nonminimality has been removed in the process.

The above discussion has assumed that the e-state dimension is to be retained. If one drops this assumption, then it is possible to remove all nonminimality. For example, one can find a minimal state-variable realization, which is also an e-state realization of particular dimension.

D. Equivalence

Rosenbrock's strict system equivalence (RSE) is the tool with which to describe and link all minimal MFD descriptions of the same transfer function (or for that matter, classes of nonminimal descriptions in which the nature of the nonminimality is in a rough sense common to the class). Let us check that this notion is appropriate to the e-state models. Using z^{-1} as a variable, RSE dictates that if A_r, B_r, C_r, d_r and $\bar{A}_r$, $\bar{B}_r$, $\bar{C}_r$, and $\bar{d}_r$ define two equivalent realizations, there exist unimodular M_1, M_2 and X, Y, all polynomial in z^{-1}, with

$$\left[\begin{array}{c|c} M_1 & 0 \\ \hline X & 1 \end{array}\right]\left[\begin{array}{cc|c} I_{\bar{N}} & 0 & 0 \\ 0 & A_r & B_r \\ \hline 0 & -C_r & d_r \end{array}\right] = \left[\begin{array}{cc|c} I_N & 0 & 0 \\ 0 & \bar{A}_r & \bar{B}_r \\ \hline 0 & -\bar{C}_r & \bar{d}_r \end{array}\right]\left[\begin{array}{c|c} M_2 & Y \\ \hline 0 & 1 \end{array}\right].$$

(A and $\bar{A}$ are $N \times N$ and $\bar{N} \times \bar{N}$, respectively.)

Now A_r, B_r, C_r, d_r come from an e-state model if and only if $A_r(0) = I$, and z^{-1} divides $B_r(z^{-1})$. Will $\bar{A}_r(0)$ and $\bar{B}_r(z^{-1})$ necessarily have these properties?

The answer is clearly no. However, there is a very closely related set matrices $\tilde{A}_r$, $\tilde{B}_r$, $\tilde{C}_r$, $\tilde{d}_r$, essentially obtained by a type of normalization, which do have the properties:

$$\begin{bmatrix} I_N & 0 & 0 \\ 0 & \tilde{A}_r & \tilde{B}_r \\ 0 & -\tilde{C}_r & \tilde{d}_r \end{bmatrix} = \begin{bmatrix} I_N & 0 & 0 \\ 0 & \bar{A}_r & \bar{B}_r \\ 0 & -\bar{C}_r & d_r \end{bmatrix}\begin{bmatrix} I_N & 0 & 0 \\ 0 & \bar{A}_r^{-1} & -\bar{A}_r^{-1}(0)\,\bar{B}_r(0) \\ 0 & 0 & 1 \end{bmatrix}.$$

(It is easily checked that $\bar{A}_r(0)$ must be nonsingular.)

Hence Rosenbrock's system equivalence, *modulo normalization* as just indicated, allows us to recover *all* e-state realizations which are equivalent using the z^{-1} variable.

Note that if system matrices are set up using the z-variable which correspond to two e-state realizations that are z^{-1}-variable equivalent, the z-variable descriptions may not be equivalent, because of possible differences in coprimeness, etc., at $z = 0$.

We also mention now a second form of equivalence, known as Fuhrmann system equivalence (FSE), which also relates system matrices of the same transfer function [4]. We shall use this in the next section. Though it is not obvious, FSE and RSE define the same equivalence [4]. We say that FSE holds between A, B, C, d, and $\bar{A}$, $\bar{B}$, $\bar{C}$, $\bar{d}$ where $d + CA^{-1}B = \bar{d} + \bar{C}\bar{A}^{-1}\bar{B}$ if

$$\begin{bmatrix} M_1 & 0 \\ X & 1 \end{bmatrix}\begin{bmatrix} A & B \\ -C & d \end{bmatrix} = \begin{bmatrix} \bar{A} & \bar{B} \\ -\bar{C} & \bar{d} \end{bmatrix}\begin{bmatrix} M_2 & Y \\ 0 & 1 \end{bmatrix}$$

where M_1, M_2, X, Y are polynomial, and $[M_1 \;\; \bar{A}]$ and $[A^t \;\; M_2^t]$ have full rank for all z.

IV. The Naturally Associated State-Variable Realization

Fig. 1 shows part of the signal flowgraph which will be normally associated with the realization of (1.1) in a digital signal processing context. The pickoffs, scaling, and summing to produce $y(n)$ are not shown for clarity, but can readily be visualized. The realization of (1.1a) is basically a direct form realization.

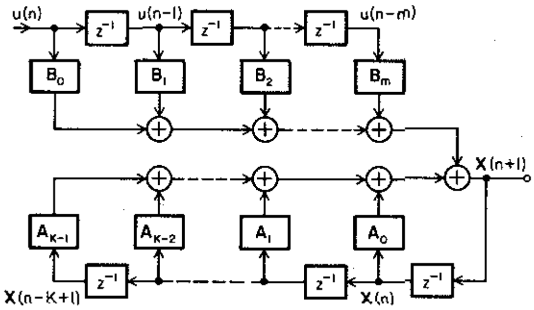

Fig. 1. Realization of (1a). The symbol z^{-1} denotes a (possibly vector) delay.

The "natural" state variable to associate with this realization is

$$\tilde{x}(n) = \begin{bmatrix} u(n-m) \\ u(n-m+1) \\ \vdots \\ u(n-1) \\ X(n-K+1) \\ \vdots \\ X(n) \end{bmatrix}. \tag{4.1}$$

The associated state-variable equations are

$$\begin{aligned} \tilde{x}(n+1) &= \tilde{F}\tilde{x}(n) + \tilde{G}u(n) \\ y(n) &= \tilde{H}\tilde{x}(n) + d_0 u(n) \end{aligned} \tag{4.2}$$

where

$$\tilde{F} = \left[\begin{array}{cccc|cccc} 0 & 1 & \cdots & 0 & & & & \\ \vdots & \vdots & & \vdots & & & & \\ 0 & 0 & \cdots & 1 & & 0 & & \\ 0 & 0 & \cdots & 0 & & & & \\ \hline 0 & 0 & \cdots & 0 & 0 & I & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & I \\ B_m & B_{m-1} & \cdots & B_1 & A_{K-1} & A_{K-2} & \cdots & A_0 \end{array}\right]$$

$$\tilde{G} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ \hline 0 \\ \vdots \\ 0 \\ B_0 \end{bmatrix}$$

$$\tilde{H} = \left[\begin{array}{cccc|cccc} d_m & d_{m-1} & \cdots & d_1 & C_{K-1} & C_{K-2} & \cdots & C_0 \end{array}\right]. \tag{4.3}$$

The first question which arises is whether (4.2) is equivalent (in the sense of Fuhrmann system equivalence) to the MFD described in the previous section.

Theorem 4.1: Let $A(z)$, $B(z)$, $C(z)$, and $D(z)$ be as described in (3.4), and $\tilde{F}$, $\tilde{G}$, $\tilde{H}$, d_0 be as above. Then

$$\begin{bmatrix} M_1 & 0 \\ 0 & 1 \end{bmatrix}\left[\begin{array}{cc|c} A(z) & -B(z) & 0 \\ 0 & z^m & 1 \\ \hline -C(z) & -D(z)+d_0z^m & d_0 \end{array}\right] = \left[\begin{array}{c|c} zI-\tilde{F} & \tilde{G} \\ -\tilde{H} & d_0 \end{array}\right]\begin{bmatrix} M_2 & 0 \\ 0 & 1 \end{bmatrix} \tag{4.4}$$

where

$$M_1 = \left[\begin{array}{cc} 0 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 1 \\ \hline 0 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ I & B_0 \end{array}\right] \begin{array}{l} \left.\vphantom{\begin{array}{c}0\\0\\ \vdots\\0\end{array}}\right\} m \\ \left.\vphantom{\begin{array}{c}0\\0\\ \vdots\\I\end{array}}\right\} K \text{ blocks} \end{array}$$

$$M_2 = \left[\begin{array}{cc} 0 & 1 \\ 0 & z \\ \vdots & \vdots \\ 0 & z^{m-1} \\ \hline I & 0 \\ \vdots & \vdots \\ z^{K-1}I & 0 \end{array}\right]. \tag{4.5}$$

Further, those coprimeness properties hold which imply that (4.4) is a Fuhrmann system equivalence relation.

Proof: Equation (4.4) is easily established by direct verification: both sides evaluate as

$$\left[\begin{array}{cc|c} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0 & z^m & 1 \\ \hline 0 & 0 & 0 \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \\ A(z) & -B(z)+B_0z^m & B_0 \\ \hline -C(z) & -D(z)+d_0z^m & d_0 \end{array}\right].$$

Next, suppose $[M_1(\lambda) \quad \lambda I-\tilde{F}]$ fails to have full row rank for some λ. Then there exists a nonzero w such that $w^t[M_1(\lambda) \quad \lambda I-\tilde{F}]=0$.

Let $w=[w_1 \cdots w_m \quad w_{m+1}^t \cdots w_{m+K}^t]^t$.

Since $w^tM_1(\lambda)=0$,

$$w_{m+K}^t=0, \qquad w_m+w_{m+K}^tB_0=0$$

which imply $w_m=0$. Since $\lambda w^t=w^t\tilde{F}$, we obtain

$$\begin{array}{ll} \lambda w_1 = w_{m+K}^tB_m & \lambda w_{m+1}^t = w_{m+K}^tA_{K-1} \\ \lambda w_2 = w_1 + w_{m+K}^tB_{m-1} & \lambda w_{m+2}^t = w_{m+1}^t + w_{m+K}^tA_{K-2} \\ \vdots & \vdots \\ \lambda w_m = w_{m-1} + w_{m+K}^tB_1 & \lambda w_{m+K}^t = w_{m+K-1}^t + w_{m+K}^tA_0 \end{array}$$

and using the fact that w_m and $w_{m+K}=\mathbf{0}$, $w=\mathbf{0}$ follows.

The argument establishing the full column rank property for

$$\begin{bmatrix} A(z) & -B(z) \\ 0 & z^m \\ M_2 & \end{bmatrix}$$

is trivial since M_2 clearly has full column rank for all z. Q.E.D.

A number of conclusions follow from Theorem 4.1. For example, the only λ for which $[\lambda I-\tilde{F} \quad \tilde{G}]$ fails to have full rank (the uncontrollable modes) are those for which $[A(z) \quad B(z)]$ fails to have full rank. Accordingly, if the original e-state realization is such that $[A_r(z^{-1}) \quad B_r(z^{-1})]$ and $[A_r^t(z^{-1}) \quad C_r^t(z^{-1})]$ have full rank for all z^{-1}, the only uncontrollable and unobservable modes for (4.2) are at $z=0$. It may be possible to eliminate some, or even all, of them while retaining the connection to the e-state structure.

V. e-State Realizations of a Transfer Function

So far, we have not constructed any e-state realizations from a transfer function. Such a construction is however easy to obtain. Let us suppose that

$$H(z)=\frac{a_0+a_1z^{-1}+\cdots+a_Mz^{-M}}{1+b_1z^{-1}+\cdots+b_Mz^{-M}} = a_0+\frac{(a_1-a_0b_1)z^{-1}+\cdots+(a_M-a_0b_M)z^{-M}}{1+b_1z^{-1}+\cdots+b_Mz^{-M}}$$

and suppose we seek an e-state realization with any delay K for the e-state less than M. (If $K \geq M$, then one can identify the e-state with a conventional state vector.) Let $N=\lceil M/K \rceil$ (the least integer greater than or equal to M/K). It is well known that one conventional state-variable realization is provided by

$$x(n+1)=\begin{bmatrix} 0 & 1 & & \\ & & \ddots & \\ & & & 1 \\ -b_M & \cdots & -b_2 & -b_1 \end{bmatrix}x(n) + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}u(n) \tag{5.1a}$$

$$y(n)=[a_M-a_0b_M \quad a_{M-1}-a_0b_{M-1} \quad \cdots \quad a_1-a_0b_1] \cdot x(n)+a_0u(n). \tag{5.1b}$$

If $M \neq KN$, we can form a nonminimal realization of dimension KN, still controllable; as it turns out, all unobservable eigenvalue are at the origin provided the numerator and denominator of $H(z)$ are coprime

$$x(n+1)=\begin{bmatrix} -0 & & 1 & & & \\ & & & \ddots & & \\ & & & & & 1 \\ 0 & \cdots & 0 & -b_M \cdots & -b_2 & -b_1 \end{bmatrix} \cdot x(n)+\begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} u(n) \quad (5.2a)$$

$$y(n)=\begin{bmatrix} 0 & \cdots & 0 & a_M-a_0b_M & \cdots & a_1-a_0b_1 \end{bmatrix} \cdot x(n)+a_0u(n). \quad (5.2b)$$

With a permutation of the entries of $x(n)$, denoted by P, we get

$$Px(n+1)=\begin{bmatrix} 0 & I & 0 & & 0 \\ 0 & 0 & I & & 0 \\ \vdots & & & & \vdots \\ & & & & I \\ A_{K-1} & & \cdots & A_1 & A_0 \end{bmatrix} \cdot Px(n)+\begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} u(n) \quad (5.3a)$$

$$y(n)=\begin{bmatrix} C_{K-1} & C_{K-2} & \cdots & C_0 \end{bmatrix} Px(n)+a_0u(n) \quad (5.3b)$$

where

$$A_i=\left[\begin{array}{cccc} & \mathbf{0} & & \\ \hline -b_{NK-K+i+1} & -b_{NK-2K+i+1} & \cdots & -b_{i+1} \end{array}\right], \quad \text{for } i=0,\cdots,K-2$$

$$A_{K-1}=\left[\begin{array}{c|ccc} \mathbf{0} & & I_{N-1} & \\ \hline -b_{NK} & -b_{NK-K} & \cdots & -b_K \end{array}\right]$$

$$C_i=[a_{NK-K+i+1}-a_0b_{NK-K+i+1} \; a_{NK-2K+i+1}-a_0b_{NK-2K+i+1} \; \cdots \; a_{i+1}-a_0b_{i+1}], \quad \text{for } i=0,1,\cdots,K-1. \quad (5.4)$$

One interprets b_l and a_l for $l>M$ as zero. Then the construction of the last section establishes that these definitions together with evanescing B_i and d_i, $i>0$, and

$$B_0=[0 \quad \cdots \quad 0 \quad 1]^t$$
$$d_0=a_0 \quad (5.5)$$

yield an e-state realization as in (1.1), with $m=0$, with $H(z)$ as transfer function.

VI. Computing Quantities Relevant to Digital Filter Design

As discussed in [1], e-state realizations can often be used to advantage as the basis of a digital filter design. By generalizing work developed in the context of regular state variable realizations [6], [7], [1] showed that the following two quantities are relevant in defining a figure of merit for the digital filter, and in obtaining a scaling property for the internal variables:

$$M=\sum_{i=0}^{K-1}\sum_{j=0}^{\infty} \mathcal{A}_{0,j-i}^t C_i^t C_i \mathcal{A}_{0,j-i} \quad (6.1)$$

$$N=\sum_{i=0}^{m}\sum_{j=0}^{\infty} \mathcal{A}_{0,j-i} B_i B_i^t \mathcal{A}_{0,j-i}^t. \quad (6.2)$$

These infinite sums take finite values if and only if the filter is stable. We now indicate how M, N may be computed. We reuse the definition given earlier in (2.12) for F, and add the following definitions of E_K and H:

$$F=\begin{bmatrix} 0 & I & 0 & \cdots & 0 \\ 0 & 0 & I & \cdots & 0 \\ & & & \ddots & \\ & & & & I \\ A_{K-1} & A_{K-2} & A_{K-1} & \cdots & A_0 \end{bmatrix}$$

$$E_K=\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ I \end{bmatrix} \qquad H^t=\begin{bmatrix} C_{K-1}^t \\ C_{K-2}^t \\ \vdots \\ C_0^t \end{bmatrix}. \quad (6.3)$$

In light of the calculation made for F previously, see (2.13), we have

$$HF^jE_K=\sum_{i=0}^{K-1} C_i A_{0,j-i}$$

and so

$$M=E_K^t\left[\sum_{j=0}^{\infty}(F^j)^t H^t H F^j\right]E_K. \quad (6.4)$$

The infinite sum exists because the stability of the filter implies $|\lambda_i(F)|<1$ for all i. Now as is well known, the quantity

$$P=\sum_{j=0}^{\infty}(F^j)^t H^t H F^j \quad (6.5)$$

can be computed as the solution of

$$P-F^tPF=H^tH \quad (6.6)$$

while also the infinite sum can be approached by a doubling formula. Thus if

$$P_n = \sum_{j=0}^{n} (F^j)^t H^t H F^j \tag{6.7}$$

one computes successively F, F^2, F^4, $F^8, \cdots$ and P_1, P_2, P_4, $P_8, \cdots$ using

$$P_{2n} = P_n + (F^n) P_n F^n. \tag{6.8}$$

So M can be obtained via solution of the linear matrix equation, or by use of a doubling formula to evaluate the infinite sum.

Much the same is true of N. Replace K and m by $L = \max(K, m)$, setting $A_i = 0$ for $i > K-1$, or $B_i = 0$ for $i > L$. With

$$\bar{F} = \begin{bmatrix} 0 & 0 & & A_L \\ I & 0 & & A_{L-1} \\ 0 & I & & A_{L-2} \\ \vdots & & \ddots & \vdots \\ 0 & 0 & I & A_0 \end{bmatrix}$$

$$G = \begin{bmatrix} B_L \\ B_{L-1} \\ B_{L-2} \\ \vdots \\ B_0 \end{bmatrix} \tag{6.9}$$

one has

$$N = E_K^t \sum_{j=0}^{\infty} \bar{F}^j G G^t (\bar{F}^j)^t E_K. \tag{6.10}$$

The way to use a linear matrix equation or a doubling formula should now be clear.

VII. Concluding Remarks

The e-state model, when compared to the classical state-space model with matrices that are not specially structured, requires less storage and less computations for $K > 1$. For instance, for a K-state model whose e-state variable vector is of dimension $N \times 1$, in general $KN^2 + 2KN$ coefficients are needed in contrast to $K^2N^2 + 2KN$ for a classical state-space model neglecting direct passes to the output. That is, a reduction of $[100\,(1-(N+2)/(KN+2)]$ percent in the required storage and number of multiplications is achieved by using the e-state model instead of state-space model.

The amount of reduction for a system of order 30 (transfer function order) is shown in Table I for different orders of e-state model.

Of course, the reduction in computational requirements with an increase in K may be coupled with a deterioration in some other performance measure. For instance, in the case of digital filter design, this reduction is associated with an increase of product roundoff noise in the output. However, in general, between the two extremes of e-state models known and exploited previously, i.e., $K = 1$ (state-space form), and $K = M$, where M is the order of the transfer function (direct form), the optimal solution to any modeling problem can arise with $1 < K < M$. As we have shown, this is true in case of design of low roundoff noise digital filters [1]. Using several implementation techniques we evaluated and compared the cost of building e-state digital filters which realize the same transfer function and have the same roundoff noise power at the output. We have considered the third-order transfer function given by Hwang in [6]. In all the cases the e-state model with $K = 2$ turned out to be the cheapest to build [1].

TABLE I

ORDER OF K-STATE MODEL	DIMENSION OF e-STATE VECTOR, N	REDUCTION (%) OF STORAGE OR NUMBER OF MULTIPLICATIONS
1	30	0
2	15	46
3	10	62
5	6	75
6	5	78
10	3	84
15	2	87
30	1	90

The advantages of e-state models include less computation in terms of storage and arithmetic. The structures provided by the e-state models are ideal for implementation on array-processors. In this case, only arrays of length N, the dimension of the e-state variable vectors, need to be handled. In a large-order system, the problem of bookkeeping of the state-space model implemented on the array-processors is cumbersome, and thus the e-state model provides a simpler techniques.

One might query whether it is fair to compare the e-state model with a state-variable model whose matrices are not specially structured. After all, every e-state model is equivalent to a state-variable model with special structure. The answer is that the comparison is a fair one in the sense that the e-state model provides the tool for highlighting the importance of certain structured state-variable forms, the importance of which may not become apparent or may be simply submerged in a general analysis. For example, the work of [6], [7] on noise basically virtually submerges all special structure of the state variable realizations by introducing unspecialized coordinate basis transformations of the state space.

References

[1] S. K. Mitra and J. Fadavi-Ardekani, "A new approach to the design of cost-optimal low-noise digital filters," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. ASSP-29, pp. 1172–1176, 1981.
[2] W. A. Wolovich, *Linear Multivariable Systems*, New York: Springer-Verlag, 1974, ch. 5.
[3] H. H. Rosenbrock, *State-Space and Multivariable Theory*, England: Thomas Nelson and Sons, 1970.
[4] T. Kailath, *Linear Systems*, Englewood Cliffs, NJ: Prentice Hall, 1980.
[5] B. W. Dickinson, T. Kailath, and M. Morf, "Canonical matrix fraction and state-space descriptions for deterministic and stochastic

systems," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 656–667, 1974.
[6] S. Hwang, "Minimum uncorrelated unit noise in state-space digital filtering," *IEEE Trans Acoustics, Speech, Signal Processing*, vol. ASSP-25, pp. 273–281, 1977.
[7] C. T. Mullis and R. A. Roberts, "Roundoff noise in digital filters frequency transformations and invariants," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. ASSP-24, pp. 538–550, 1976.

Jalil Fadavi-Ardekani (S'78–M'80) was born in Tehran, Iran, in 1953. He recieved the M.S.E.E. degree from Syracuse University, Syracuse, NY, in 1976, and the Ph.D. from the University of California, Santa Barbara, in 1980 in electrcial engineering.

During his study at the University of California, Santa Barbara, he was a Research Assistant in the Department of Electrical and Computer Engineering and the Computer Systems Laboratory. He also worked as a Teaching Assistant, Teaching Associate, and Lecturer at the University of California, successively. While at Santa Barbara, he participated in the development of the Signal Processing Laboratory in the Department of Electrical and Computer Engineering. His current interest includes digital communication systems, digital filtering and system theory.

Sanjit K. Mitra (S'59–M'63–SM'69–F'67) received the B.S. (Hons.) degree in physics from the Utkal University, India, in 1953, the M.Sc. (Tech.) degree in radio physics and electronics from the University of Calcutta, India, in 1956, and the M.S. and Ph.D. degrees in electrical engineering from the University of California, Berkeley, in 1960 and 1962, respectively.

He joined the faculty of the University of California, Davis in 1967 and transferred to the Santa Barbara (UCSB) campus in 1977 as a Professor of Electrical and Computer Engineering. From July 1979 to June 1982 he served as the Chairman of the Department of Electrical and Computer Engineering at UCSB. He is a Consultant to the Lawrence Livermore National Laboratory, Livermore, CA, and is the Consulting Editor for the *Electrical/Computer Science and Engineering Series* of the Van Nostrand Reinhold Company, NY. He has held visiting appointments at the Indian Institute of Technology, New Delhi, India, Kobe University, Japan, and the University of Erlangen-Nuernber, West Germany. From July 1982 to December 1982, he will be a Visiting Fellow at the Australian National University, Canberra. He has published a number of papers in active and passive networks, and digital filters, and is the author of two texts and editor/co-editor of three other books.

Dr. Mitra is a Fellow of the American Association for the Advancement of Science and a member of the American Society for Engineering Education, Sigma Xi and Eta Kappa Nu. He was an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS, and is presently on the editorial boards of the PROCEEDINGS OF THE IEEE and *Circuits, Systems and Signal Processing*. He was a member of the Administrative Committee of the IEEE Circuits and Systems, and a member of the editorial board of the IEEE Press. He is the Technical Program Chairman of the 1983 IEEE International Symposium on Circuits and Systems. He is also the receipient of the 1973 F. E. Terman Award of the American Society for Engineering Education.

Brian D. O. Anderson (S'62–M'66–SM'74–F'75) was born in Sydney, Australia, in 1941. He received the B.S. degrees in pure mathematics and electrical engineering from the University of Sydney, Sydney, Australia, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA.

He is currently Professor and Head of the Department of Systems Engineering at the Australian National University; from 1967 through 1981 he was Professor of electrical engineering at the University of Newcastle. He has also held appointments as a Visiting Professor at Southern Methodist University, University of Massachusetts, University of California at Berkeley and Santa Barbara, and Stanford University. He is coauthor of five books: *Linear Optimal Control*, *Network Analysis and Synthesis Foundation of System Theory*, and most recently *Singular Optimal Control*, with D. J. Clements, and *Optimal Filtering*, with J. B. Moore. His research interests are in control, communication systems, and networks. From 1971 to 1977 he was a member of the Australian Research Grants Committee and since 1977 has been a member of the Australian Science and Technology Council. He is a member of the Australian Academy of Science, Australian Academy of Technological Sciences and Institution of Engineers, Australia, and a member of Sigma Xi, Siam, and Eta Kappa Nu. He is an editor of *Automatica.*