

THE UNIVERSITY OF MANITOBA

MINIMAL WAVE DIGITAL FILTER REALIZATIONS:

THE N-PORT APPROACH

BY

ANTHONY THOMAS ASHLEY

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF ELECTRICAL ENGINEERING

WINNIPEG, MANITOBA

MAY 1978

MINIMAL WAVE DIGITAL FILTER REALIZATIONS:
THE N-PORT APPROACH

BY

ANTHONY THOMAS ASHLEY

A dissertation submitted to the Faculty of Graduate Studies of
the University of Manitoba in partial fulfillment of the requirements
of the degree of

DOCTOR OF PHILOSOPHY

© 1978

Permission has been granted to the LIBRARY OF THE UNIVERSITY OF MANITOBA to lend or sell copies of this dissertation, to the NATIONAL LIBRARY OF CANADA to microfilm this dissertation and to lend or sell copies of the film, and UNIVERSITY MICROFILMS to publish an abstract of this dissertation.

The author reserves other publication rights, and neither the dissertation nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.



ABSTRACT

This thesis is a study of minimal wave digital filters designed using n -port adaptors. A topological characterization of the reactive redundancies in the reference RLC network is used together with a set of constraint equations to simultaneously eliminate excess delays due to both loops and cutsets of capacitance and inductance. The method, which can be applied to prototypes of arbitrary topology, produces n -port adaptors in which the multipliers are restricted to a submatrix, K . A network interpretation of K is given which allows realizations which are canonic in both delays and multipliers to be obtained from ladder prototypes.

Several results regarding the properties of n -port adaptors and the controllability and observability of pseudolossless reciprocal systems are given. The stability of linear wave digital systems and the relationship with controllability and observability is investigated. General system modification schemes which guarantee both state and output stability for nonlinear wave digital systems are presented.

Necessary and sufficient conditions for the existence of diagonal Lyapunov functions for minimal wave digital systems are derived and it is demonstrated that such functions do not exist in a majority of filters. An alternate diagonalization procedure which uses a similarity transformation of the state variables is given.

A technique based upon a form of interval arithmetic is used to bound the errors caused by finite word length effects. These bounds are then used to define signal modifications which guarantee

freedom from parasitic oscillations in n-port filters having diagonal Lyapunov functions. Finally, a type of zeroing arithmetic which inhibits overflow oscillations in canonic realizations is given.

ACKNOWLEDGEMENT

The author wishes to express his sincere appreciation to Dr. G.O. Martens for his guidance and encouragement not only during the preparation of this thesis, but throughout his entire graduate program. The author also wishes to thank his colleagues and those members of the academic staff of the Department of Electrical Engineering who have contributed, either directly or indirectly, to this thesis.

The financial assistance of the National Research Council, Gulf Oil Canada Limited and the University of Manitoba is gratefully acknowledged.

Table of Contents

Chapter		Page
I	Introduction	1
II	Wave Digital Filter Design - n-Port Adaptors	8
	2.1 Introduction to Wave Digital Filters	8
	2.2 n-Port Adaptor Representations	12
	2.3 Reflection-Free n-Port Adaptors	20
	2.4 Illustrative Example	22
III	Canonic Wave Digital Filters: n-Port Adaptor Realizations	25
	3.1 Characterization of Reactive Redundancies in RLC Prototypes	26
	3.2 Voltage Wave Constraint Equations	32
	3.3 n-Port Adaptor Representations for Canonic Wave Digital Filters	36
	3.4 Network Interpretation of K	47
	3.5 Design Procedure - Illustrative Examples	53
IV	Properties of Wave Digital Filters: Controllability, Observability and Stability	87
	4.1 Properties of n-Port Wave Digital Adaptors	87
	4.2 Controllability and Observability of Linear Wave Digital Filters	93
	4.3 Stability of Wave Digital Filters	96
V	Diagonal Lyapunov Functions for Minimal Wave Digital Filters	106
	5.1 Eigenvalues and Eigenvectors of S_{11}	106
	5.2 Generation of Alternate Diagonal Lyapunov Functions	116
	5.3 Transformation of Variables to Diagonalize G_{11}	129
VI	Suppression of Parasitic Oscillations in Nonlinear Wave Digital Filters Using n-Port Adaptors	140
	6.1 Signal Modifications for Stability	140
	6.2 Error Interval Analysis	146
	6.3 Removal of Overflow Oscillations in Minimal Realizations without Diagonalization	163
VII	Concluding Remarks and Suggestions for Future Work	168
	Appendix A	172
	Bibliography	177

CHAPTER I

INTRODUCTION

Digital signal processing is the processing of discrete-time signals with a special-or general-purpose computer. Modern digital processing began with the simulation of complex analog systems on digital computers. As more sophisticated machines became available and the possibility of implementing real-time systems arose, a tremendous interest in developing highly efficient algorithms grew. With the advent of integrated circuit technology which resulted in high-speed circuitry at low cost, it is now practical to build hardware digital signal processors.

Digital systems offer several advantages over analog systems. System specifications can be achieved with a high degree of accuracy and are easily repeatable, high reliability and economy are obtained with IC realizations and time variable or adaptive behaviour is easily implemented.

The applications of digital signal processing are now widespread, including such diverse fields as radar, sonar, geophysical exploration, analysis of biomedical signals and, of course, communication systems. Two recent textbooks [1], [2] provide an excellent introduction to the field of digital signal processing. In addition, a large collection of important papers on the subject is available [3], [4].

Infinite impulse response digital filters can be designed in three distinct, although not independent, steps. These are:

1. From the given performance specification, determine a linear

shift-invariant (LSI) discrete-time system which meets or exceeds the specification.

2. Decide upon a structure in which to realize the LSI system and quantize the coefficients to a fixed word length.
3. Quantize the signals consistent with the word length of the digital system used for implementation.

From the desired performance specification which can be given in the time domain or, as is most often the case, in the frequency domain, two basic techniques are available for determining the necessary LSI system. When the specifications are not standard, mathematical optimization procedures which minimize some specified error criterion can be used to design directly in the z -domain. The majority of the techniques which are available consider a structure consisting of a cascade of second-order sections, the pole and zero locations being determined by the algorithm. If a standard response such as Butterworth, Chebyshev or elliptic is desired, then a more efficient technique is to utilize the well-established theory of RLC filters. Various mappings have been proposed to transform an appropriate analog filter into the required discrete-time filter, the most frequently used methods being the impulse invariant and the bilinear z -transformation [1], [2].

Having specified a suitable discrete-time system, an operational realization consisting of adders, multipliers and delay (memory) elements is required. These elements must be connected in a manner such that the resulting digital structure is computable [1], [5]. Normally, the infinite precision coefficients (multipliers) in the realization must

be modified to a finite word length. The choice of structure is complicated by the sensitivity of the system performance due to this coefficient quantization. The number of components, particularly multipliers and delay elements, and the potential for high-speed operation are other major considerations.

Several standard recursive structures have been developed. The direct forms DI and DII are high-order feedback structures realized directly from the transfer function. The cascade form is obtained by factoring the transfer function into a product of first-and second-order sections while the parallel form is obtained from a partial fraction expansion. In general, these structures all suffer to varying degrees from coefficient sensitivity problems [1] -[4].

A structure called a wave digital filter has been introduced by Fettweis and his co-workers [6]-[10]. This method uses voltage scattering waves together with the bilinear z -transformation to map resistively-terminated LC ladder prototypes into digital structures. Because the multipliers are determined in a one-to-one mapping from the elements in the prototype, the low element sensitivity of the classical filter is transformed into low coefficient sensitivity in the digital filter [8], [11]. Allowing for impedance scaling, a prototype of n elements produces a wave digital filter containing $n-1$ multipliers. A comparison of the complexity of cascade and wave digital realizations of an eighth-order bandpass filter [12] has shown that the number of multiplier elements is essentially the same while the wave digital filter requires almost twice as many adders. The lower coefficient sensitivity of the wave digital structure,

however, allows this realization to be implemented with about 60% of the total number of bits of that required in the cascade form.

A second comparison by Fettweis et al. [13], based upon a seventh-order lowpass filter, shows that the total number of logic circuits required for a serial arithmetic wave digital realization is comparable to that required for a cascade design. The number of delay elements in a standard wave digital realization is equal to the number of reactive elements in the reference filter and thus these wave digital filters are canonic in delays if and only if the reference filter is a minimal realization. By using the dependence of the waves in loops and cutsets of inductances and capacitances, additional hardware can be used to eliminate some of the excess delays caused by these degeneracies [14]. Fettweis et al. [15] and Nouta [16] have both developed a lattice adaptor for realizing symmetrical networks. Wave digital adaptors for the reciprocal and nonreciprocal sections used in classical cascade synthesis have also been obtained by various authors [17] - [21].

Implementation of a digital filter requires that the signals be expressed in binary form. Fixed-point fractional arithmetic is most often used in hardware realizations. Since the memory word length is fixed, the signal values to be stored must lie within a specified interval if they are to be represented accurately. However, as a result of arithmetic operations, numbers may be produced which fall outside the range available. Overflows result when the signal is larger than the maximum value allowable. In this case the most significant bits must be altered, causing large errors in the output and the possibility of zero-input limit cycles, called overflow oscillations. Quantization of the signals is used when necessary to modify the least significant

bits, producing quantization noise and the possibility of zero-input limit cycles, called granularity oscillations. Claassen et al. [22] have an excellent discussion of these problems.

For second-order direct form fixed-point realizations, saturating overflow arithmetic does not produce overflow oscillations [23]. Furthermore, magnitude truncation of the sum of products almost always eliminates granularity oscillations [24]. Fettweis and Meerkötter have used the concept of pseudopower to derive a simple criterion which guarantees the absence of both zero-input overflow and granularity oscillations in wave digital filters [25]. This technique, however, is not directly applicable to those wave digital filters designed by the previously-discussed method which reduces the number of excess delays [14]. Using wave digital concepts, Meerkötter and Wegener [26] have designed a second-order section having no limit cycles while a scheme using controlled rounding has also been found to be effective for certain filters [27], [28].

An alternate wave digital structure in the form of a single n-port adaptor terminated with feedback through memory has been proposed by Martens and Meerkötter [29]. Unlike the standard Sedlmeyer-Fettweis procedure [9], this method is not restricted to the transformation of ladder prototypes.

This thesis is a study of minimal wave digital filters designed using n-port adaptors.

In Chapter II we introduce the basic concepts required for an understanding of wave digital filters. The series-parallel adaptor

method developed by Fettweis is briefly discussed. The n-port adaptor technique introduced by Martens and Meerkötter is described and a derivation of the various n-port adaptor representations is given. Following a brief discussion of reflection-free n-port adaptors, a simple example which illustrates the general n-port procedure is given.

The main purpose of Chapter III is to develop a method of designing minimal wave digital filters using n-port adaptors. First, the reactive redundancies that cause the dimension of the state space of the prototype filter to exceed the minimal dimension are characterized topologically. Constraint equations used to simultaneously eliminate the excess delays due to both loops and cutsets of reactive elements are then formulated and their effect upon the network behaviour is interpreted. The n-port adaptors produced have representations in which the multipliers are restricted to a submatrix K . A network interpretation of K is developed, allowing realizations which are canonic in both multipliers and delays to be obtained from ladder prototypes. The chapter concludes with two illustrative examples.

In Chapter IV some interesting properties of n-port adaptors, including their pseudolossless and reciprocal nature, are established. The controllability and observability of pseudolossless reciprocal systems is studied. Several results concerning the stability of linear wave digital systems are proved using the reference conductance matrix as a Lyapunov function. This matrix is diagonal for non-minimal realizations but becomes nondiagonal as a result of the minimal realization procedure. Finally, the stability of nonlinear wave

digital systems is considered and a general scheme which, in principle, guarantees freedom from both overflow and granularity oscillations in these systems is given.

Chapter V is devoted to the search for diagonal Lyapunov functions for those systems which have a nondiagonal reference conductance matrix. Necessary and sufficient conditions for the existence of alternate Lyapunov functions are derived. Because several examples demonstrate that these conditions can be satisfied in only a limited number of cases, the final section of the chapter presents a technique which uses a similarity transformation of the state variables to produce new systems for which diagonal Lyapunov functions do exist.

In Chapter VI we describe a procedure for implementing signal modifications which inhibit limit-cycles in wave digital filters using n -port adaptors. The errors caused by finite word length constraints are monitored by a form of error interval analysis which produces error bounds on the signals at the outputs of the adaptor. Based upon these bounds, the signals are appropriately modified before being fed back to the adaptor inputs. The final section in this chapter presents a form of zeroing arithmetic which can be used to eliminate overflow oscillations in minimal wave digital filters.

Standard matrix notation is used throughout. Superscripts T and -1 denote transposition and inversion respectively, while U is a unit matrix of appropriate dimensions. In general, time domain vector or scalar signals are denoted by lower-case Latin letters, while upper-case Latin letters identify scalar or vector signals in the complex frequency domain.

CHAPTER II

WAVE DIGITAL FILTER DESIGN - n-PORT ADAPTORS

The design of wave digital filters imitating analog reference networks is carried out via the voltage wave scattering representation of the reference filter structure together with the application of the bilinear z-transformation [6]-[10], [29]. This synthesis procedure transforms the low element sensitivity of doubly terminated LC ladder reference filters into low coefficient sensitivity of the discrete-time realization. In addition, wave digital filters exhibit several interesting properties which can be utilized to guarantee the absence of parasitic oscillations. Chapter II serves as an introduction to wave digital filter design. Included is a discussion of the basic concepts upon which the wave digital approach is based, as well as a brief synopsis of Fettweis' adaptor technique [6]-[9]. The n-port adaptor method of Martens and Meerkötter [29] is reviewed and a derivation of the various n-port adaptor representations is given. Finally, an example is given to illustrate the n-port technique.

2.1 INTRODUCTION TO WAVE DIGITAL FILTERS

The doubly terminated lossless reciprocal network shown in Fig. 2.1 is the most often utilized analog filter structure and hence extensive design tables are available [30]. Such a structure is normally described by either its voltage transfer function

$$T(\psi) = \frac{V_2(\psi)}{E(\psi)} \quad (2.1)$$

or by its transmission coefficient [31]

$$t(\psi) = 2 \left(\frac{R_S}{R_L} \right)^{\frac{1}{2}} T(\psi) \quad (2.2)$$

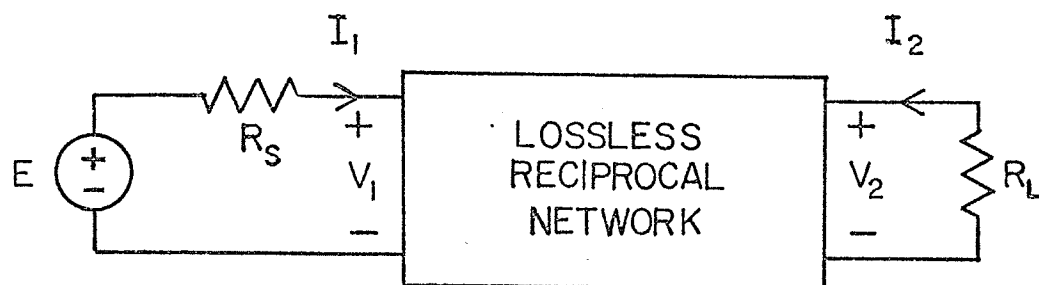


Fig. 2.1 Doubly terminated lossless reciprocal network.

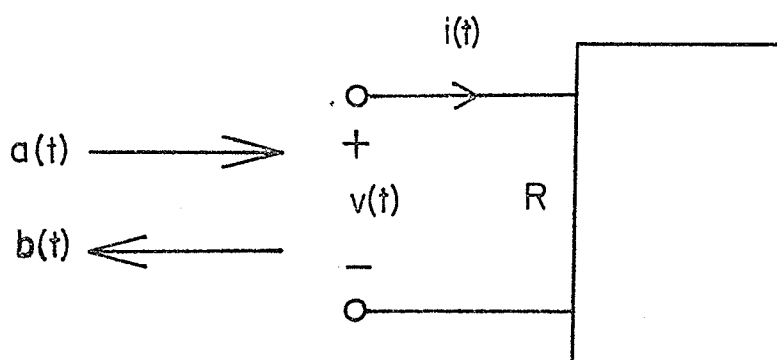


Fig. 2.2 Definition of port variables.

where ψ is the continuous-time domain complex frequency variable.

As an alternative, voltage waves can be used to describe the filter.

For a given port with associated references for the voltage $v(t)$

and the current $i(t)$, (Fig. 2.2), we define the incident and reflected voltage waves $a(t)$ and $b(t)$ respectively by

$$a = v + Ri \quad (2.3a)$$

$$b = v - Ri \quad (2.3b)$$

or, in the complex frequency domain,

$$A = V + RI \quad (2.4a)$$

$$B = V - RI \quad (2.4b)$$

where A , B , V and I are the complex amplitudes of the signals and R is the reference resistance, normally positive, chosen for the port. If we choose the reference resistances for the source and load ports to be equal to R_s and R_L respectively, then

$$A_1 = E \quad (2.5a)$$

$$B_2 = 2V_2 \quad (2.5b)$$

and

$$W(\psi) = \frac{B_2}{A_1} \quad (2.6a)$$

$$= 2T(\psi) \quad (2.6b)$$

$$= \left(\frac{R_L}{R_s} \right)^{\frac{1}{2}} t(\psi) \quad (2.6c)$$

where $W(\psi)$ is the voltage wave transfer function. It is important to note that the magnitude of $T(j\phi)$, $t(j\phi)$ and $W(j\phi)$ differ by, at most, a frequency-independent constant and hence a realization of any

of these functions produces the desired frequency response.

Use of the bilinear z-transformation

$$\psi = \frac{z - 1}{z + 1} \quad (2.7)$$

produces the z-domain transfer function

$$H(z) = W \left(\frac{z - 1}{z + 1} \right) \quad (2.8)$$

where z is the discrete-time domain complex frequency variable. Since the discrete-time frequency response is given by $H(e^{j\omega T})$, the analog frequency ϕ and the digital frequency ω are related by

$$j\phi = \frac{e^{j\omega T} - 1}{e^{j\omega T} + 1} \quad (2.9)$$

or equivalently

$$\phi = \tan \frac{\omega T}{2} \quad (2.10)$$

where T is the sampling period. This nonlinear warping, (2.10), introduced by the bilinear transformation can be compensated by pre-warping the prototype by appropriately changing the element values so that the critical analog frequencies are transformed into the desired critical digital frequencies.

Fettweis' wave digital design technique [6]-[9] takes the voltage wave representation of each element in the reference filter, transforms them into discrete-time equivalents and interconnects these subnetworks using adaptors designed to allow the interconnection of ports with different reference resistances. This procedure represents a departure from the standard recursive filter design techniques where the transformation into the discrete-

time domain is made directly on the transfer function [1], [2].

If the reference resistances are chosen to be equal to R , L and $1/C$ for the resistive, inductive and capacitive branches respectively, the elements are transformed into the discrete-time domain as shown in Fig. 2.3. Since the elements in a ladder structure are arranged in a series-parallel form, series and parallel adaptors are used in the wave digital realization [6]. These instantaneous elements, containing only multipliers and adders, can be designed with reflection-free ports [9]. These special ports allow adaptors to be interconnected with the assurance that no delay-free loops will be introduced. A 3-port series or parallel adaptor contains 2 multipliers and 6 adders. The number of components is reduced to 1 multiplier and 4 adders for reflection-free adaptors [10]. Martens and Meerkötter [29] have proposed an alternate wave digital structure in the form of an n -port adaptor terminated with feedback through memory. This technique can be applied to a network of connections of arbitrary topology.

2.2 n-PORT ADAPTOR REPRESENTATIONS

Consider a lossless reciprocal instantaneous n -port network. Since this network is passive, the port variables can always be partitioned so that a hybrid matrix H exists [32]

$$\begin{bmatrix} v_1 \\ i_2 \end{bmatrix} = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \begin{bmatrix} i_1 \\ v_2 \end{bmatrix} \quad (2.11)$$

where v_1 and v_2 are port voltage vectors, i_1 and i_2 are port current vectors and H is a real constant matrix of appropriate dimension.

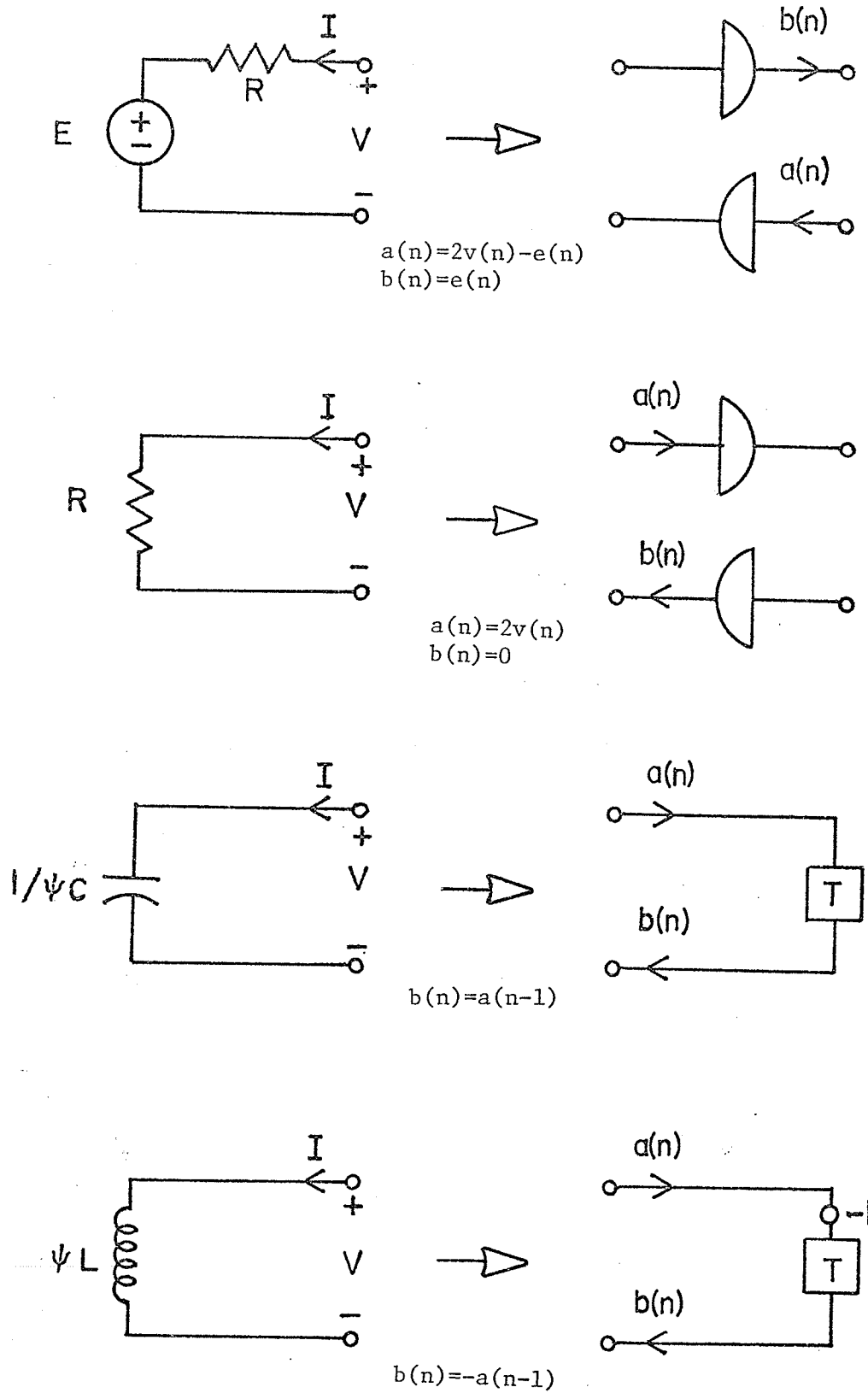


Fig. 2.3 Discrete-time equivalents of the analog elements.

The lossless reciprocal nature of the network allows H to be written in the form

$$H = \begin{bmatrix} 0 & H_{12} \\ -H_{12}^T & 0 \end{bmatrix} \quad (2.12)$$

or, using simpler notation,

$$\begin{bmatrix} v_1 \\ i_2 \end{bmatrix} = \begin{bmatrix} 0 & P^T \\ -P & 0 \end{bmatrix} \begin{bmatrix} i_1 \\ v_2 \end{bmatrix} \quad (2.13)$$

where $P = H_{12}^T$.

The incident and reflected voltage wave vectors are defined

by

$$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} + \begin{bmatrix} R_1 & 0 \\ 0 & R_2 \end{bmatrix} \begin{bmatrix} i_1 \\ i_2 \end{bmatrix} \quad (2.14a)$$

and

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} - \begin{bmatrix} R_1 & 0 \\ 0 & R_2 \end{bmatrix} \begin{bmatrix} i_1 \\ i_2 \end{bmatrix} \quad (2.14b)$$

where $R = \begin{bmatrix} R_1 & 0 \\ 0 & R_2 \end{bmatrix}$ is the diagonal positive definite reference resistance matrix. Then

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} a_1 + b_1 \\ a_2 + b_2 \end{bmatrix} \quad (2.15a)$$

and

$$\begin{bmatrix} i_1 \\ i_2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} G_1 & 0 \\ 0 & G_2 \end{bmatrix} \begin{bmatrix} a_1 - b_1 \\ a_2 - b_2 \end{bmatrix} \quad (2.15b)$$

where $G = R^{-1} = \begin{bmatrix} G_1 & 0 \\ 0 & G_2 \end{bmatrix}$ is the diagonal positive definite reference

conductance matrix.

Substitution of (2.15) into (2.13) yields

$$\begin{bmatrix} U & 0 \\ 0 & G_2 \end{bmatrix} \begin{bmatrix} a_1 + b_1 \\ a_2 - b_2 \end{bmatrix} = \begin{bmatrix} 0 & P^T \\ -P & 0 \end{bmatrix} \begin{bmatrix} G_1 & 0 \\ 0 & U \end{bmatrix} \begin{bmatrix} a_1 - b_1 \\ a_2 + b_2 \end{bmatrix} \quad (2.16)$$

and, upon rearranging,

$$\begin{bmatrix} U & -P^T \\ PG_1 & G_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} -U & P^T \\ PG_1 & G_2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}. \quad (2.17)$$

Inversion of the coefficient matrix on the left-hand side in the form

$$\begin{bmatrix} U & -P^T \\ PG_1 & G_2 \end{bmatrix}^{-1} = \begin{bmatrix} U - P^T Y^{-1} PG_1 & P^T Y^{-1} \\ -Y^{-1} PG_1 & Y^{-1} \end{bmatrix}, \quad (2.18)$$

where $Y = PG_1 P^T + G_2$ is positive definite and hence nonsingular,

produces the following scattering matrix representation of the network:

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 2P^T K - U & 2P^T (U - KP^T) \\ 2K & U - 2KP^T \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \quad (2.19)$$

where $K = Y^{-1} PG_1$. If the network contains only wire connections and ideal transformers, then P contains as its elements only +1, -1, 0 and transformer turns ratios n .

For networks of wire connections only, an obvious partition of the ports is defined by a set of links and twigs of the network graph. Kirchhoff's voltage and current laws

$$\text{KVL: } \begin{bmatrix} U & B_t \end{bmatrix} \begin{bmatrix} v_\ell \\ v_t \end{bmatrix} = 0 \quad (2.20a)$$

$$\text{KCL: } \begin{bmatrix} Q_\ell & U \end{bmatrix} \begin{bmatrix} i_\ell \\ i_t \end{bmatrix} = 0 \quad (2.20b)$$

together with the orthogonality between the fundamental loop and cut-set matrices, $B_t = -Q_\ell^T$, [33] enable P in (2.13) to be replaced by Q_ℓ . We then obtain

$$\begin{bmatrix} b_\ell \\ b_t \end{bmatrix} = \begin{bmatrix} 2Q_\ell^T K - U & 2Q_\ell^T (U - KQ_\ell^T) \\ 2K & U - 2KQ_\ell^T \end{bmatrix} \begin{bmatrix} a_\ell \\ a_t \end{bmatrix} \quad (2.21)$$

where $K = Y^{-1}Q_\ell G_\ell$ and $Y = Q_\ell G_\ell Q_\ell^T + G_t$ is the node-pair admittance matrix of the network of wire connections with each port terminated in its reference resistance. This representation and those which follow are identical to those obtained earlier by Martens and Meerkötter [29].

Alternate forms of the scattering matrix, S, are given by

$$S = \begin{bmatrix} U & Q_\ell^T \\ 0 & U \end{bmatrix} \begin{bmatrix} U & 0 \\ -2K & U \end{bmatrix} \begin{bmatrix} -U & Q_\ell^T \\ 0 & U \end{bmatrix} \quad (2.22)$$

$$= \begin{bmatrix} -U & 0 \\ 0 & U \end{bmatrix} \begin{bmatrix} U & B_t \\ 0 & U \end{bmatrix} \begin{bmatrix} U & 0 \\ 2K & U \end{bmatrix} \begin{bmatrix} U & B_t \\ 0 & U \end{bmatrix} \quad (2.23)$$

$$= \begin{bmatrix} U - Q_\ell^T K & Q_\ell^T \\ -K & U \end{bmatrix} \begin{bmatrix} -U & 0 \\ 0 & U \end{bmatrix} \begin{bmatrix} U & -Q_\ell^T \\ K & U - KQ_\ell^T \end{bmatrix} \quad (2.24)$$

Equation (2.24), which displays the eigenvalues of S, can be rewritten as

$$S \begin{bmatrix} U - Q_\ell^T K & Q_\ell^T \\ -K & U \end{bmatrix} = \begin{bmatrix} U - Q_\ell^T K & Q_\ell^T \\ -K & U \end{bmatrix} \begin{bmatrix} -U & 0 \\ 0 & U \end{bmatrix} \quad (2.25)$$

which then displays the eigenvectors of S .

The representation given by (2.23) is a convenient form for wave digital filter realizations. The modularity of the structures produced can be emphasized by rewriting (2.23) as

$$S = \Theta FMF \quad (2.26)$$

where F describes the topology of the prototype and contains only adders and inverters, M contains the multipliers in the submatrix K and Θ contains only inverters. The general form of a wave digital filter realized in this way is shown in Fig. 2.4. The inverters required by the inductive elements are contained in Σ . It is also apparent from Fig. 2.4 that a set of state equations describing the system can be easily obtained.

If a direct realization is used for F and M , then no internal delay-free loops will appear in the realization of S . In addition, since the connection of the delay elements cannot introduce any delay-free loops, a wave digital filter realized using a single n -port adaptor is always computable.

The matrix K has a simple network interpretation which can be obtained as follows:

Let

$$a_t = 0. \quad (2.27)$$

Then, from (2.21),

$$b_t = 2Ka_\ell. \quad (2.28)$$

If we terminate all of the link ports in their reference resistance in series with a voltage source (Fig. 2.5a) and all of the tree ports in their reference resistances (Fig. 2.5b), then

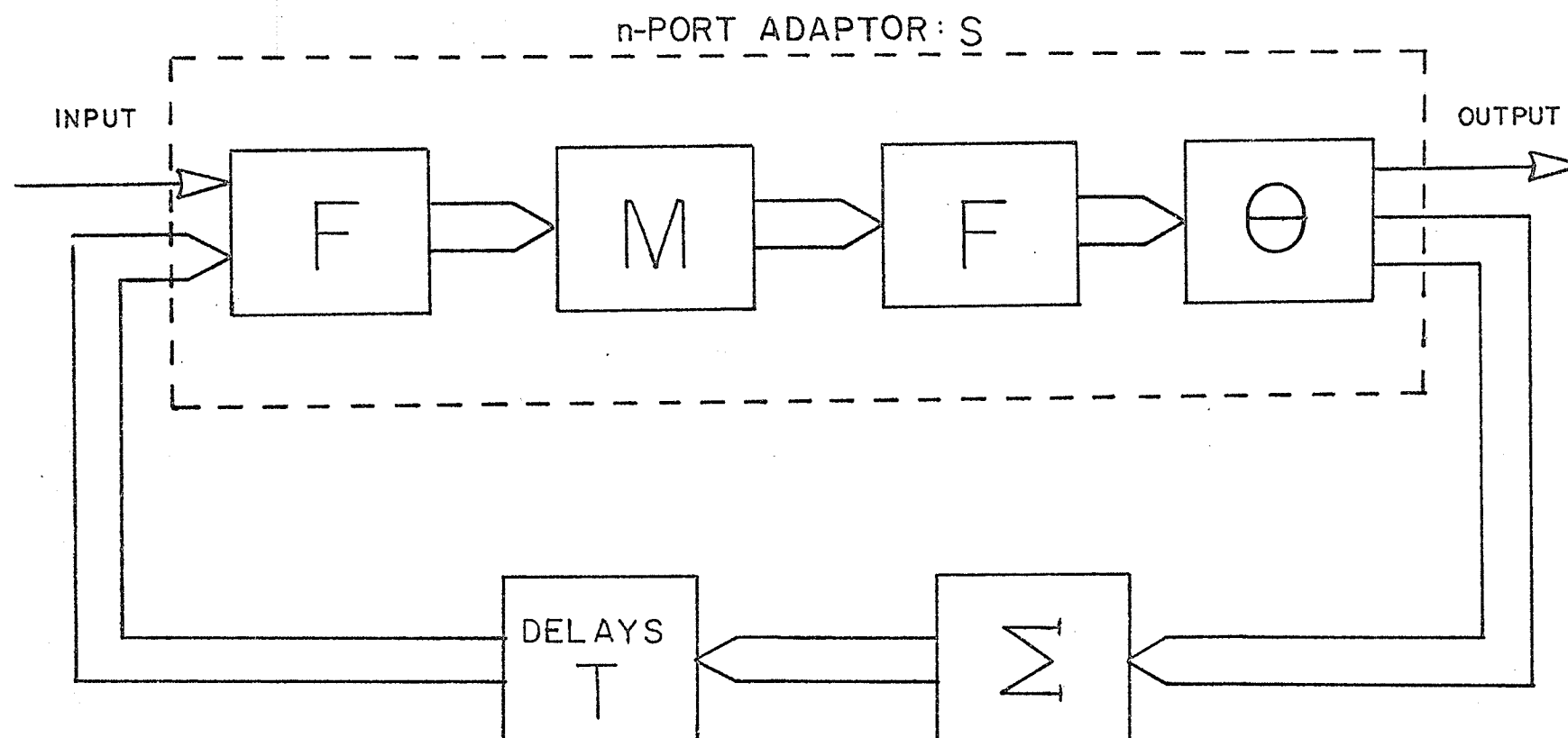


Fig. 2.4 Block flow diagram illustrating the modularity and general form of a wave digital filter based upon the n-port adaptor representation (2.23).

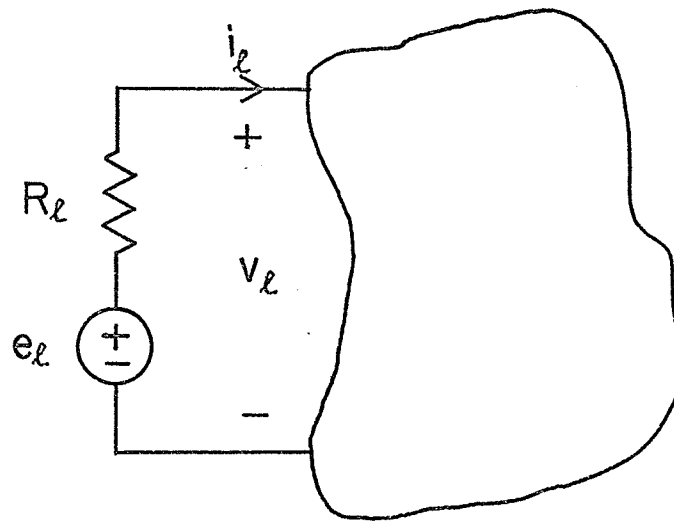


Fig. 2.5a Link termination required for the computation of K.

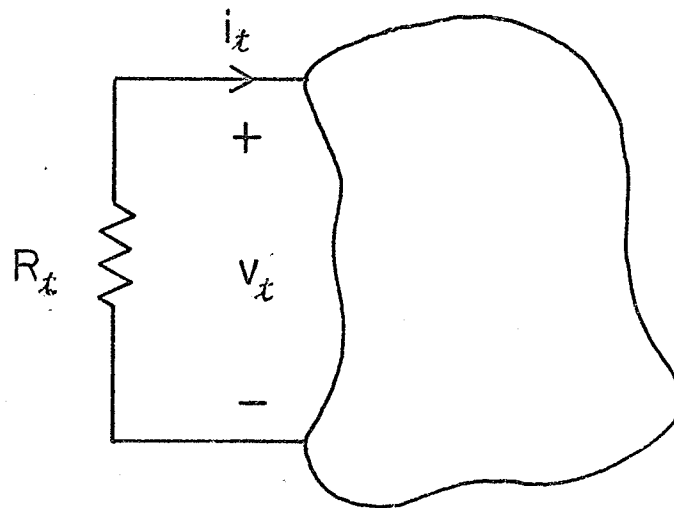


Fig. 2.5b Twig termination required for the computation of K.

$$a_\ell = v_\ell + R_\ell i_\ell = e_\ell$$

$$a_t = v_t + R_t i_t = 0$$

$$b_t = v_t - R_t i_t = 2v_t$$

and finally

$$v_t = K e_\ell \quad (2.29)$$

where e_ℓ is the vector of link voltage sources.

Thus K is the voltage transfer matrix from the link sources to the tree branches and, since the network is resistive, the entries of K are bounded

$$|k_{ij}| \leq 1, \quad i = 1, 2, \dots, t; \quad j = 1, 2, \dots, \ell. \quad (2.30)$$

The dimension of K is $t \times \ell$ and therefore at most $t\ell$ multipliers are needed in a realization. These multipliers can be generated directly by $K = Y^{-1} Q_\ell G_\ell$ or, due to the network interpretation of K , by the application of any suitable network analysis technique. If the prototype filter contains n elements, then, allowing for impedance scaling, there are $n-1$ independent parameters in the transfer function. This implies that a realization of K should be possible with the canonic number, $n-1$, of independent parameters.

2.3 REFLECTION-FREE n-PORT ADAPTORS

If it is desirable to produce a wave digital filter as an interconnection of n -port adaptors, reflection-free ports in some of the adaptors are necessary. These special ports are used to guarantee that delay-free loops cannot occur due to the adaptor interconnections.

In general, one of the two ports at each adaptor connection must be reflection-free.

An adaptor has port m reflection-free if b_m is independent of a_m ; that is, if $S_{mm} = 0$. In order to obtain this condition a particular choice of reference resistance is necessary. If we terminate all ports except port m in their reference resistance, then

$$a_i = 0, i \neq m$$

and

$$b_m = S_{mm} a_m.$$

The reflection-free condition then implies that

$$b_m = v_m - R_m i_m = 0$$

or

$$v_m = R_m i_m.$$

Since the driving point resistance at port m is given by

$$R_{dp} = v_m / i_m$$

the reflection-free condition requires that the reference resistance for port m be equal to the driving point resistance at port m when all other ports are terminated in their reference resistances. The reflection-free condition imposes a constraint upon the entries of K . If port m is chosen as a link, then from (2.21), $S_{mm} = 0$ requires that

$$-1 + 2 \sum_{r=1}^t q_{rm} k_{rm} = 0 \quad (2.31)$$

where q_{ij} are the entries of Q_ℓ . Alternatively, if port m is a twig, then $S_{mm} = 0$ requires that

$$1 - 2 \sum_{r=1}^{\ell} k_{mr} q_{mr} = 0 \quad (2.32)$$

The dependence of the entries of K imposed by the reflection-free condition can be used to reduce the number of multipliers in a reflection-free adaptor [10].

2.4 ILLUSTRATIVE EXAMPLE

The n -port adaptor design procedure can be applied to any topology and hence can be used to obtain the adaptors introduced by Fettweis. As an example, consider the parallel connection of three ports shown in Fig. 2.6a. The corresponding network graph showing the tree chosen for the analysis and the fundamental cutset is given in Fig. 2.6b. We have

$$\begin{aligned} Q &= \begin{bmatrix} Q_{\ell} & U \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \\ G_{\ell} &= \begin{bmatrix} G_1 & 0 \\ 0 & G_2 \end{bmatrix}, \quad G_t = G_3 \end{aligned}$$

and thus

$$\begin{aligned} Y &= \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} G_1 & 0 \\ 0 & G_2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + G_3 \\ &= G_1 + G_2 + G_3. \end{aligned}$$

Then

$$\begin{aligned} 2K &= \frac{2}{G_1 + G_2 + G_3} \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} G_1 & 0 \\ 0 & G_2 \end{bmatrix} \\ &= \begin{bmatrix} \alpha_1 & \alpha_2 \end{bmatrix} \end{aligned}$$

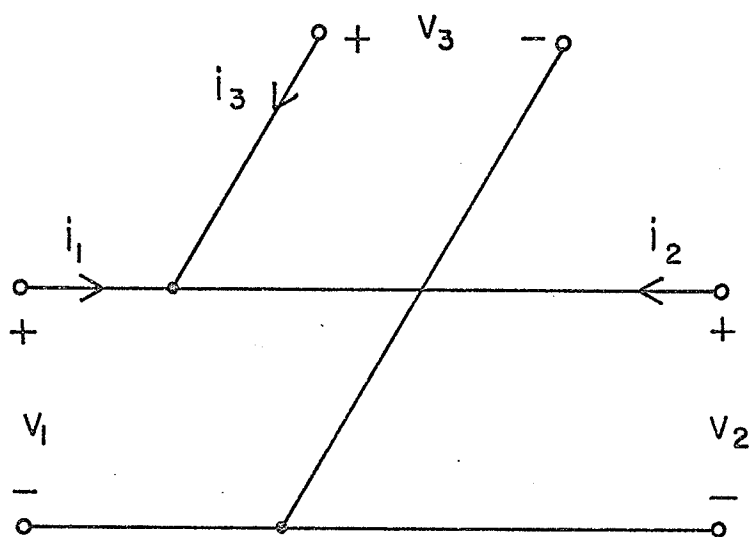


Fig. 2.6a Parallel connection of three ports.

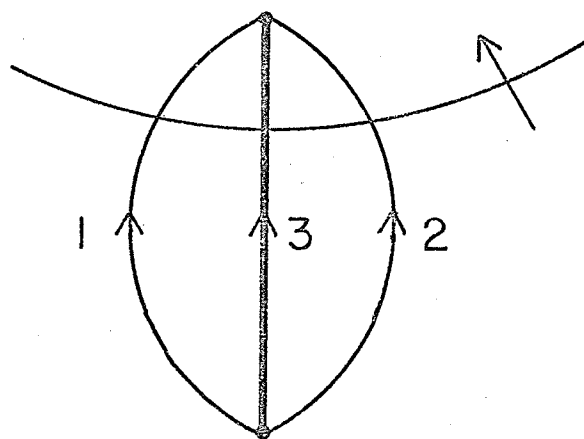


Fig. 2.6b Network graph corresponding to Fig. 2.6a.

where

$$\alpha_1 = \frac{2G_1}{G_1 + G_2 + G_3}, \quad \alpha_2 = \frac{2G_2}{G_1 + G_2 + G_3}.$$

The scattering matrix representation in the form of (2.22)

$$S = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\alpha_1 & -\alpha_2 & 1 \end{bmatrix} \begin{bmatrix} -1 & 0 & 1 \\ 0 & -1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

is identical to that given by Fettweis and Meerkötter[10].

If it is desired to make port 2 reflection-free then since $q_{12}=1$, (2.31) produces

$$-1 + 2k_{12} = 0$$

and thus

$$\alpha_2 = 1.$$

The resulting scattering matrix can then be realized with one less adder and one less multiplier [10].

In the next chapter we develop a procedure for obtaining n-port adaptors for wave digital filter realizations having minimal degree. Since the design examples given there use techniques which are also applicable to the non-minimal filters considered in this chapter, it will not be necessary to give further examples here.

CHAPTER III

CANONIC WAVE DIGITAL FILTERS:

n-PORT ADAPTOR REALIZATIONS

The number of delays contained in a wave digital filter obtained by the n-port adaptor technique of the previous chapter is equal to the total number of reactive elements in the prototype filter. From a system theoretic point of view, such realizations may not be minimal because the degree of the transfer function matrix may be smaller than the number of delays. It is well known that a minimal discrete-time realization can always be constructed from the prototype transfer function. However, such a realization will not generally depend directly upon the structure of the prototype filter and thus the useful properties available in a wave digital realization are not obtained.

In this chapter we describe a procedure for designing n-port adaptors which can be used to obtain realizations with a reduced number of delays. In situations where the excess delays are due to loops and/or cutsets of reactive elements, the realizations obtained will be minimal. A reduction in the order of a realization is not only of academic interest since the delays, unlike the multipliers, cannot be multiplexed.

The technique to be developed can be considered to be a generalization of the procedure used by Fettweis [14] to eliminate some of the excess delays in an adaptor realization. However, unlike Fettweis' procedure, the method given here is simultaneously applicable to both loops and cutsets. In addition, the modularity of the structure

is maintained and the filters can be shown to have all of the interesting properties of wave digital systems. These properties will be discussed in the next chapter.

For ladder prototypes containing n elements, it is possible to obtain realizations using $n-1$ independent multipliers. The number of adders required can be smaller than the number needed by an equivalent Fettweis adaptor realization.

The most often used method of order reduction uses a nonsingular transformation of the state variables. It can be easily demonstrated that such a transformation leaves the transfer function invariant and, if properly chosen, decouples the uncontrollable and/or unobservable parts of the system [34]. Unfortunately, no simple technique exists for determining the required symbolic change of variables. An alternate procedure presented in this chapter solves this problem as it applies to wave digital filter design using n -port adaptors. The specific characteristics of the prototype filter which produce the extra delays are first identified and then used to form constraint equations. These constraints are applied during the formation of the state equations in order to produce a modified state description of lower dimension.

It is important to note, however, that an actual filter design does not require the lengthy and complex proof to be duplicated. A concise set of design rules is given.

The chapter concludes with some illustrative examples.

3.1 CHARACTERIZATION OF REACTIVE REDUNDANCIES IN RLC PROTOTYPES

The minimal degree required in an abstract realization of

an analog transfer function can differ from the number of reactive elements in a concrete RLC realization for two reasons. First, the number of state variables needed to describe the RLC network is not always equal to the number of reactive elements. In fact, it is well known that the dimension of the state space is equal to the total number of reactive elements less the number of independent capacitance - voltage source-only loops and the number of independent inductance-current source-only cutsets [35]. By convention, these classes include capacitance-only loops and inductance-only cutsets as special cases. Secondly, the dimension of the state space of the RLC realization may be excessively large due to the existence of uncontrollable and/or unobservable modes. One such class of modes which can be easily identified is the zero natural frequency due to capacitance-only cutsets and inductance-only loops.

In the following development we shall consider only connected RLC networks. In addition, we shall also assume that all sources have associated resistive elements and that a single edge of the network graph will be assigned to a resistor-source combination and labelled for convenience as a resistive branch. This last restriction also rules out the possibility of capacitance-voltage source loops and inductance-current source cutsets.

A normal tree is defined as a tree having as its twigs the maximum number of capacitive branches and the minimum number of inductive branches [35]. Kirchhoff's voltage and current laws (KVL and KCL), when partitioned with respect to the normal tree, yield

$$\text{KVL: } \begin{bmatrix} U & 0 & 0 & B_{SC} & 0 & 0 \\ 0 & U & 0 & B_{RC} & B_{RG} & 0 \\ 0 & 0 & U & B_{LC} & B_{LG} & B_{L\Gamma} \end{bmatrix} \begin{bmatrix} v_S \\ v_R \\ v_L \\ v_C \\ v_G \\ v_\Gamma \end{bmatrix} = 0 \quad (3.1)$$

$$\text{KCL: } \begin{bmatrix} Q_{CS} & Q_{CR} & Q_{CL} & U & 0 & 0 \\ 0 & Q_{GR} & Q_{GL} & 0 & U & 0 \\ 0 & 0 & Q_{\Gamma L} & 0 & 0 & U \end{bmatrix} \begin{bmatrix} i_S \\ i_R \\ i_L \\ i_C \\ i_G \\ i_\Gamma \end{bmatrix} = 0 \quad (3.2)$$

where the subscripts S, R, L, C, G and Γ denote link: capacitances, resistances, inductances; twig: capacitances, resistances and inductances respectively. Due to the nature of the normal tree, the capacitance-only loops, defined by the link capacitances S, appear explicitly in the first KVL equation while the inductance-only cutsets, defined by the twig inductances Γ , appear explicitly in the last KCL equation. The inductance-only loops and the capacitance-only cutsets do not appear explicitly for this choice of tree.

Consider the network N_{Lo} , obtained by open-circuiting all branches of the original network, N, which are not inductances. Since this procedure cannot create loops, nor can it destroy any inductance-only loops, the number of independent loops in N_{Lo} is equal to the number of independent inductance-only loops in N. Furthermore, since the inductive twigs in N do not, by definition, form any loops in N, these branches will not form any loops in N_{Lo} and therefore can be chosen as part of a tree, T_{Lo} , (or forest if N_{Lo} is not connected) in

N_{Lo} [33]. Any additional twigs required to complete T_{Lo} are chosen from the branches which were links in N , thus inducing a partition in these branches. The fundamental loop equations with respect to T_{Lo} are

$$\begin{bmatrix} U & B_{12} & B_{13} \end{bmatrix} \begin{bmatrix} v_{L1} \\ v_{L2} \\ v_{\Gamma} \end{bmatrix} = 0. \quad (3.3)$$

The original link inductances, denoted by L_1 , define the fundamental loops in N_{Lo} and therefore also define the inductance-only loops in N .

Next consider the network N_{Cs} , obtained by short-circuiting all branches of N which are not capacitances. Since this procedure cannot create cutsets, nor can it destroy any capacitance-only cutsets, the number of independent cutsets in N_{Cs} is equal to the number of independent capacitance-only cutsets in N . Furthermore, since the capacitive links in N do not, by definition, form any cutsets in N , these branches will not form any cutsets in N_{Cs} and therefore can be chosen as part of a cotree, complementary to T_{Cs} , in N_{Cs} [33]. Any additional links required are chosen from the branches which were twigs in N , thus inducing a partition of these branches. The fundamental cutset equations with respect to T_{Cs} are

$$\begin{bmatrix} Q_{11} & Q_{12} & U \end{bmatrix} \begin{bmatrix} i_S \\ i_{C2} \\ i_{C1} \end{bmatrix} = 0 \quad (3.4)$$

where any branches which form self loops are identified by a null

column in the cutset matrix. The original twig capacitances, denoted by C_1 , define the fundamental cutsets in N_{CS} and therefore also define the capacitance-only cutsets in N .

If the branches of N are further partitioned consistent with the partitioning induced in (3.3) and (3.4), KVL and KCL for N become

$$\begin{bmatrix} U & 0 & 0 & 0 & | & B_{SC1} & B_{SC2} & 0 & 0 \\ 0 & U & 0 & 0 & | & B_{RC1} & B_{RC2} & B_{RG} & 0 \\ 0 & 0 & U & 0 & | & B_{L1C1} & B_{L1C2} & B_{L1G} & B_{L1\Gamma} \\ 0 & 0 & 0 & U & | & B_{L2C1} & B_{L2C2} & B_{L2G} & B_{L2\Gamma} \end{bmatrix} \begin{bmatrix} v_S \\ v_R \\ v_{L1} \\ v_{L2} \\ \text{---} \\ v_{C1} \\ v_{C2} \\ v_G \\ v_{\Gamma} \end{bmatrix} = 0 \quad (3.5)$$

and

$$\begin{bmatrix} Q_{C1S} & Q_{C1R} & Q_{C1L1} & Q_{C1L2} & | & U & 0 & 0 & 0 \\ Q_{C2S} & Q_{C2R} & Q_{C2L1} & Q_{C2L2} & | & 0 & U & 0 & 0 \\ 0 & Q_{GR} & Q_{GL1} & Q_{GL2} & | & 0 & 0 & U & 0 \\ 0 & 0 & Q_{\Gamma L1} & Q_{\Gamma L2} & | & 0 & 0 & 0 & U \end{bmatrix} \begin{bmatrix} i_S \\ i_R \\ i_{L1} \\ i_{L2} \\ \text{---} \\ i_{C1} \\ i_{C2} \\ i_G \\ i_{\Gamma} \end{bmatrix} = 0 \quad (3.6)$$

Since there are the same number of equations in the third block row of (3.5) and equation (3.3) and in each case these equations are independent, we can substitute (3.3) into (3.5), obtaining

$$\begin{bmatrix}
 U & 0 & 0 & 0 & | & B_{SC1} & B_{SC2} & 0 & 0 \\
 0 & U & 0 & 0 & | & B_{RC1} & B_{RC2} & B_{RG} & 0 \\
 0 & 0 & U & B_{12} & | & 0 & 0 & 0 & B_{13} \\
 0 & 0 & 0 & U & | & B_{L2C1} & B_{L2C2} & B_{L2G} & B_{L2\Gamma}
 \end{bmatrix}
 \begin{bmatrix}
 v_S \\
 v_R \\
 v_{L1} \\
 v_{L2} \\
 \hline
 v_{C1} \\
 v_{C2} \\
 v_G \\
 v_{\Gamma}
 \end{bmatrix}
 = 0, \quad (3.7)$$

Similarly, the first block row of (3.6) can be replaced by (3.4) to obtain

$$\begin{bmatrix}
 Q_{11} & 0 & 0 & 0 & | & U & Q_{12} & 0 & 0 \\
 Q_{C2S} & Q_{C2R} & Q_{C2L1} & Q_{C2L2} & | & 0 & U & 0 & 0 \\
 0 & Q_{GR} & Q_{GL1} & Q_{GL2} & | & 0 & 0 & U & 0 \\
 0 & 0 & Q_{\Gamma L1} & Q_{\Gamma L2} & | & 0 & 0 & 0 & U
 \end{bmatrix}
 \begin{bmatrix}
 i_S \\
 i_R \\
 i_{L1} \\
 i_{L2} \\
 \hline
 i_{C1} \\
 i_{C2} \\
 i_G \\
 i_{\Gamma}
 \end{bmatrix}
 = 0, \quad (3.8)$$

Equations (3.7) and (3.8) constitute a valid set of KVL and KCL equations for N which explicitly display all of the desired reactive redundancies.

In more compact notation, (3.7) and (3.8) can be written as

$$\begin{bmatrix} B_{\ell} & | & B_t \end{bmatrix} \begin{bmatrix} v_{\ell} \\ \hline v_t \end{bmatrix} = 0, \quad (3.9)$$

and

$$\begin{bmatrix} Q_{\ell} & | & Q_t \end{bmatrix} \begin{bmatrix} i_{\ell} \\ \hline i_t \end{bmatrix} = 0 \quad (3.10)$$

B_ℓ and Q_t are not unit matrices since (3.7) and (3.8) are not fundamental equations; that is, the equations are not all written with respect to the same tree in N .

Using the well-known orthogonality condition

$$Q_\ell B_\ell^T + Q_t B_t^T = 0$$

the following relationships are easily derived:

$$B_{SC_1} = -B_{SC_2} Q_{12}^T - Q_{11}^T, \quad B_{SC_2} = -Q_{C_2S}^T \quad (3.11 \text{ a,b})$$

$$Q_{\Gamma L_1} = -Q_{\Gamma L_2} B_{12}^T - B_{13}^T, \quad Q_{\Gamma L_2} = -B_{L_2\Gamma}^T \quad (3.11 \text{ c,d})$$

$$B_{RC_1} = -B_{RC_2} Q_{12}^T, \quad B_{RC_2} = -Q_{C_2R}^T \quad (3.11 \text{ e,f})$$

$$Q_{GL_1} = -Q_{GL_2} B_{12}^T, \quad Q_{GL_2} = -B_{L_2G}^T \quad (3.11 \text{ g,h})$$

$$B_{L_2C_1} = -B_{L_2C_2} Q_{12}^T, \quad B_{L_2C_2} = -Q_{C_2L_2}^T \quad (3.11 \text{ i,j})$$

$$Q_{C_2L_1} = -Q_{C_2L_2} B_{12}^T, \quad Q_{C_2L_2} = -B_{L_2C_2}^T \quad (3.11 \text{ k,l})$$

3.2 VOLTAGE WAVE CONSTRAINT EQUATIONS

In order to reduce the number of delays in a wave digital realization, the topological descriptions of the reactive redundancies described in the previous section must be transformed into the voltage wave domain and suitable constraint equations must be determined. These constraints can then be used to obtain an n -port adaptor which, when suitably terminated, will yield a discrete-time realization of a lower dimension.

In this section we will assume that all time-domain signals are represented in the form

$$x(t) = X e^{\psi t} \quad (3.12)$$

where X is the complex amplitude and ψ is the complex frequency variable.

For notational convenience the explicit dependence of x upon time will not be shown.

First consider the capacitance-only cutsets in N . A description of these cutsets is available from KCL equation (3.4) or, equivalently, from

$$Q_i = 0 \quad (3.13)$$

where the variable partitioning is no longer explicitly shown in Q and i . Equation (3.13) can be expressed in terms of voltage wave vectors

$$QG(a-b) = 0 \quad (3.14)$$

where the reference conductance matrix G is given by

$$G = \begin{bmatrix} G_S & 0 & 0 \\ 0 & G_{C_2} & 0 \\ 0 & 0 & G_{C_1} \end{bmatrix} = \begin{bmatrix} S & 0 & 0 \\ 0 & C_2 & 0 \\ 0 & 0 & C_1 \end{bmatrix} . \quad (3.15)$$

Using the port voltage-current references of N , the element relationship for the capacitances

$$I = -\psi GV$$

combined with the complex frequency-domain equivalent of (3.13) implies that

$$QGV = 0 \quad \text{for all } \psi \neq 0 . \quad (3.16)$$

Up to this point we have been examining the intrinsic behaviour of the system. Let us now consider the effect of extending (3.16) to include $\psi = 0$; that is,

$$QGV = 0 \quad \text{for all } \psi \quad (3.17)$$

which is equivalent to

$$QGv = 0 . \quad (3.18)$$

Equation (3.18) can be written as

$$QG(a+b) = 0 \quad (3.19)$$

which, together with (3.14), is equivalent to

$$QG_a = 0 \quad (3.20a)$$

$$QG_b = 0. \quad (3.20b)$$

These conditions can be viewed as extensions of the naturally-occurring condition (3.14).

The extension of (3.14) to (3.20) or, equivalently, of (3.16) to (3.17) imposes a restriction upon the network's natural voltage distribution only at $\psi = 0$. Since the only elements which can support a nonzero voltage at this frequency are the capacitances in capacitance-only cutsets, we can short-circuit all non-capacitive elements without altering this voltage distribution. The remaining network, N_{CS} , is characterized by KVL and KCL in the form

$$\text{KVL: } \left[\begin{array}{c|c} U & 0 \\ 0 & U \end{array} \middle| \begin{array}{c} -Q_{11}^T \\ -Q_{12}^T \end{array} \right] \begin{bmatrix} V_S \\ V_{C_2} \\ V_{C_1} \end{bmatrix} = 0 \quad (3.21)$$

$$\text{KCL: } \left[\begin{array}{cc|c} Q_{11} & Q_{12} & U \end{array} \right] \begin{bmatrix} I_S \\ I_{C_2} \\ I_{C_1} \end{bmatrix} = 0. \quad (3.22)$$

Equation (3.21) yields

$$\begin{bmatrix} V_S \\ V_{C_2} \\ V_{C_1} \end{bmatrix} = \begin{bmatrix} Q_{11}^T \\ Q_{12}^T \\ U \end{bmatrix} V_{C_1}$$

$$\text{or, in unpartitioned form} \quad V = Q^T V_{C_1}. \quad (3.23)$$

Substitution of (3.23) into (3.17) yields $Q\bar{G}Q^T V_{C_1} = 0$ and, since $Q\bar{G}Q^T$ is positive definite,

$$V_{C_1} = 0. \quad (3.24)$$

Then, from (3.23),

$$V = \begin{bmatrix} V_S^T & V_{C_2}^T & V_{C_1}^T \end{bmatrix}^T = 0. \quad (3.25)$$

Equation (3.25) shows that the effect of the constraints (3.20), which can be written in the equivalent form

$$Q_{11} G_S a_S + Q_{12} G_{C_2} a_{C_2} + G_{C_1} a_{C_1} = 0 \quad (3.26a)$$

$$Q_{11} G_S b_S + Q_{12} G_{C_2} b_{C_2} + G_{C_1} b_{C_1} = 0, \quad (3.26b)$$

is to inhibit the formation of modes at $\psi = 0$ which could have occurred in the original network due to the capacitance-only cutsets.

By following similar procedures it can be shown that the effect of the constraints

$$a_S + B_{SC_1} a_{C_1} + B_{SC_2} a_{C_2} = 0 \quad (3.27a)$$

$$b_S + B_{SC_1} b_{C_1} + B_{SC_2} b_{C_2} = 0 \quad (3.27b)$$

is to inhibit the formation of modes at $\psi = \infty$, which could have occurred in the original network due to capacitance-only loops,

$$Q_{\Gamma L_1} G_{L_1} a_{L_1} + Q_{\Gamma L_2} G_{L_2} a_{L_2} + G_{\Gamma} a_{\Gamma} = 0 \quad (3.28a)$$

$$Q_{\Gamma L_1} G_{L_1} b_{L_1} + Q_{\Gamma L_2} G_{L_2} b_{L_2} + G_{\Gamma} b_{\Gamma} = 0 \quad (3.28b)$$

is to inhibit the formation of modes at $\psi = \infty$, which could have occurred in the original network due to inductance-only cutsets, and

$$a_{L_1} + B_{12} a_{L_2} + B_{13} a_{\Gamma} = 0 \quad (3.29a)$$

$$b_{L_1} + B_{12} b_{L_2} + B_{13} b_{\Gamma} = 0 \quad (3.29b)$$

is to inhibit the formation of modes at $\psi = 0$, which could have occurred in the original network due to inductance-only loops.

The effect of the constraint equations (3.26) - (3.29) in the discrete-time domain is easily established by use of the bilinear z-transformation. Eliminating modes at $\psi = 0$ in the analog realization eliminates modes at $z = 1$ in the discrete-time realization. Similarly, the removal of modes at $\psi = \infty$ corresponds to the removal of modes at $z = -1$.

3.3 n-PORT ADAPTOR REPRESENTATIONS FOR CANONIC WAVE DIGITAL FILTERS

This section describes the procedure for obtaining wave digital realizations of reduced degree. The application of the constraint equations, used to eliminate redundant variables, together with a change of variables, produces the desired results.

Kirchhoff's voltage and current laws (3.9) and (3.10) can be combined into

$$\begin{bmatrix} B_\ell & 0 \\ 0 & Q_t \end{bmatrix} \begin{bmatrix} v_\ell \\ i_t \end{bmatrix} = \begin{bmatrix} 0 & -B_t \\ -Q_\ell & 0 \end{bmatrix} \begin{bmatrix} i_\ell \\ v_t \end{bmatrix}. \quad (3.30)$$

Introduction of the incident and reflected port voltage waves yields

$$\begin{bmatrix} B_\ell & 0 \\ 0 & Q_t \end{bmatrix} \begin{bmatrix} U & 0 \\ 0 & G_t \end{bmatrix} \begin{bmatrix} a_\ell + b_\ell \\ a_t - b_t \end{bmatrix} = \begin{bmatrix} 0 & -B_t \\ -Q_\ell & 0 \end{bmatrix} \begin{bmatrix} G_\ell & 0 \\ 0 & U \end{bmatrix} \begin{bmatrix} a_\ell - b_\ell \\ a_t + b_t \end{bmatrix} \quad (3.31)$$

which, upon collecting terms, produces a matrix equation describing N in the form

$$\begin{bmatrix} B_\ell R_\ell & B_t \\ Q_\ell & Q_t G_t \end{bmatrix} \begin{bmatrix} G_\ell b_\ell \\ b_t \end{bmatrix} = \begin{bmatrix} -B_\ell R_\ell & -B_t \\ Q_\ell & Q_t G_t \end{bmatrix} \begin{bmatrix} G_\ell a_\ell \\ a_t \end{bmatrix}, \quad (3.32)$$

If the variable partitioning described by (3.7) and (3.8) is shown explicitly, (3.32) becomes

$$\begin{bmatrix}
 R_S & 0 & 0 & 0 & | & B_{SC1} & B_{SC2} & 0 & 0 & | & G_S b_S \\
 0 & R_R & 0 & 0 & | & B_{RC1} & B_{RC2} & B_{RG} & 0 & | & G_R b_R \\
 0 & 0 & R_{L1} & B_{12} R_{L2} & | & 0 & 0 & 0 & B_{13} & | & G_{L1} b_{L1} \\
 0 & 0 & 0 & R_{L2} & | & B_{L2C1} & B_{L2C2} & B_{L2G} & B_{L2\Gamma} & | & G_{L2} b_{L2} \\
 \hline
 Q_{11} & 0 & 0 & 0 & | & G_{C1} & Q_{12} G_{C2} & 0 & 0 & | & b_{C1} \\
 Q_{C2S} & Q_{C2R} & Q_{C2L1} & Q_{C2L2} & | & 0 & G_{C2} & 0 & 0 & | & b_{C2} \\
 0 & Q_{GR} & Q_{GL1} & Q_{GL2} & | & 0 & 0 & G_G & 0 & | & b_G \\
 0 & 0 & Q_{\Gamma L1} & Q_{\Gamma L2} & | & 0 & 0 & 0 & G_{\Gamma} & | & b_{\Gamma}
 \end{bmatrix} = \quad (3.33)$$

$$\begin{bmatrix}
 -R_S & 0 & 0 & 0 & | & -B_{SC1} & -B_{SC2} & 0 & 0 & | & G_S a_S \\
 0 & -R_R & 0 & 0 & | & -B_{RC1} & -B_{RC2} & -B_{RG} & 0 & | & G_R a_R \\
 0 & 0 & -R_{L1} & -B_{12} R_{L2} & | & 0 & 0 & 0 & -B_{13} & | & G_{L1} a_{L1} \\
 0 & 0 & 0 & -R_{L2} & | & -B_{L2C1} & -B_{L2C2} & -B_{L2G} & -B_{L2\Gamma} & | & G_{L2} a_{L2} \\
 \hline
 Q_{11} & 0 & 0 & 0 & | & G_{C1} & Q_{12} G_{C2} & 0 & 0 & | & a_{C1} \\
 Q_{C2S} & Q_{C2R} & Q_{C2L1} & Q_{C2L2} & | & 0 & G_{C2} & 0 & 0 & | & a_{C2} \\
 0 & Q_{GR} & Q_{GL1} & Q_{GL2} & | & 0 & 0 & G_G & 0 & | & a_G \\
 0 & 0 & Q_{\Gamma L1} & Q_{\Gamma L2} & | & 0 & 0 & 0 & G_{\Gamma} & | & a_{\Gamma}
 \end{bmatrix} .$$

Note that if the constraint equations (3.26) - (3.29) were now imposed upon the system, both the right-and left-hand sides of

equations one, three, five and eight in (3.33) would be independently equal to zero. We choose, however, to delay the actual use of the constraints until a more appropriate time.

The coefficient matrices in (3.33) can be made to display a type of hybrid symmetry with the introduction of a nonsingular change of variable together with the corresponding column operation on the coefficient matrices. The variable transformation to be used is given by

$$T_1 = \begin{bmatrix} U & 0 & & & \\ 0 & U & & & \\ \hline & & U & 0 & \\ & & -B_{12}^T & U & \\ \hline & & & & U & 0 \\ & & & & -Q_{12}^T & U \\ \hline & & & & & & U & 0 \\ & & & & & & 0 & U \end{bmatrix} \quad (3.34)$$

where the off-diagonal elements not shown explicitly are all zero.

Insertion of $T_1^{-1}T_1$ into (3.33) together with the use of orthogonality relations (3.11a,c,e,g,i,k), yields

$$\begin{bmatrix}
 R_S & 0 & 0 & 0 & -Q_{11}^T & B_{SC_2} & 0 & 0 & G_S b_S \\
 0 & R_R & 0 & 0 & 0 & B_{RC_2} & B_{RG} & 0 & G_R b_R \\
 0 & 0 & \hat{R}_{L_{11}} & \hat{R}_{L_{12}} & 0 & 0 & 0 & B_{13} & G_{L_1} b_{L_1} \\
 0 & 0 & \hat{R}_{L_{21}} & \hat{R}_{L_{22}} & 0 & B_{L_2 C_2} & B_{L_2 G} & B_{L_2 \Gamma} & G_{L_2} b_{L_2} - B_{12}^T G_{L_1} b_{L_1} \\
 \hline
 Q_{11} & 0 & 0 & 0 & \hat{G}_{C_{11}} & \hat{G}_{C_{12}} & 0 & 0 & b_{C_1} \\
 Q_{C_2 S} & Q_{C_2 R} & 0 & Q_{C_2 L_2} & \hat{G}_{C_{21}} & \hat{G}_{C_{22}} & 0 & 0 & b_{C_2} - Q_{12}^T b_{C_1} \\
 0 & Q_{GR} & 0 & Q_{GL_2} & 0 & 0 & G_G & 0 & b_G \\
 0 & 0 & -B_{13}^T & Q_{\Gamma L_2} & 0 & 0 & 0 & G_{\Gamma} & b_{\Gamma}
 \end{bmatrix} =$$

(3.35)

$$\begin{bmatrix}
 -R_S & 0 & 0 & 0 & Q_{11}^T & -B_{SC_2} & 0 & 0 & G_S a_S \\
 0 & -R_R & 0 & 0 & 0 & -B_{RC_2} & -B_{RG} & 0 & G_R a_R \\
 0 & 0 & -\hat{R}_{L_{11}} & -\hat{R}_{L_{12}} & 0 & 0 & 0 & -B_{13} & G_{L_1} a_{L_1} \\
 0 & 0 & -\hat{R}_{L_{21}} & -\hat{R}_{L_{22}} & 0 & -B_{L_2 C_2} & -B_{L_2 G} & -B_{L_2 \Gamma} & G_{L_2} a_{L_2} - B_{12}^T G_{L_1} a_{L_1} \\
 \hline
 Q_{11} & 0 & 0 & 0 & \hat{G}_{C_{11}} & \hat{G}_{C_{12}} & 0 & 0 & a_{C_1} \\
 Q_{C_2 S} & Q_{C_2 R} & 0 & Q_{C_2 L_2} & \hat{G}_{C_{21}} & \hat{G}_{C_{22}} & 0 & 0 & a_{C_2} - Q_{12}^T a_{C_1} \\
 0 & Q_{GR} & 0 & Q_{GL_2} & 0 & 0 & G_G & 0 & a_G \\
 0 & 0 & -B_{13}^T & Q_{\Gamma L_2} & 0 & 0 & 0 & G_{\Gamma} & a_{\Gamma}
 \end{bmatrix}$$

where

$$\hat{R}_L = \begin{bmatrix} \hat{R}_{L_{11}} & \hat{R}_{L_{12}} \\ \hat{R}_{L_{21}} & \hat{R}_{L_{22}} \end{bmatrix} = \begin{bmatrix} U & B_{12} \\ 0 & U \end{bmatrix} \begin{bmatrix} R_{L_1} & 0 \\ 0 & R_{L_2} \end{bmatrix} \begin{bmatrix} U & 0 \\ B_{12}^T & U \end{bmatrix} \quad (3.36)$$

and

$$\hat{G}_C = \begin{bmatrix} \hat{G}_{C11} & \hat{G}_{C12} \\ \hat{G}_{C21} & \hat{G}_{C22} \end{bmatrix}$$

$$= \begin{bmatrix} U & Q_{12} \\ 0 & U \end{bmatrix} \begin{bmatrix} G_{C1} & 0 \\ 0 & G_{C2} \end{bmatrix} \begin{bmatrix} U & 0 \\ Q_{12}^T & U \end{bmatrix} \quad (3.37)$$

are both positive definite symmetric matrices.

If the constraint equations for the capacitance loops (3.27) and for the inductance cutsets (3.28) are now imposed upon (3.35), then both the right-and left-hand sides of the first and last equations are independently equal to zero. Then, since R_S and G_T are positive definite, variables a_S , a_T , b_S and b_T can be eliminated to produce (using (3.11))

$$\left[\begin{array}{ccc|ccc} R_R & 0 & 0 & 0 & B_{RC2} & B_{RG} \\ 0 & \tilde{R}_{L11} & \tilde{R}_{L12} & 0 & 0 & 0 \\ 0 & \tilde{R}_{L21} & \tilde{R}_{L22} & 0 & B_{L2C2} & B_{L2G} \\ \hline 0 & 0 & 0 & \tilde{G}_{C11} & \tilde{G}_{C12} & 0 \\ Q_{C2R} & 0 & Q_{C2L2} & \tilde{G}_{C21} & \tilde{G}_{C22} & 0 \\ Q_{GR} & 0 & Q_{GL2} & 0 & 0 & G_G \end{array} \right] \left[\begin{array}{c} G_R b_R \\ G_{L1} b_{L1} \\ G_{L2} b_{L2} - B_{12}^T G_{L1} b_{L1} \\ \hline b_{C1} \\ b_{C2} - Q_{12}^T b_{C1} \\ b_G \end{array} \right] =$$

$$\begin{bmatrix}
 -R_R & 0 & 0 & | & 0 & -B_{RC_2} & -B_{RG} \\
 0 & -\tilde{R}_{L_{11}} & -\tilde{R}_{L_{12}} & | & 0 & 0 & 0 \\
 0 & -\tilde{R}_{L_{21}} & -\tilde{R}_{L_{22}} & | & 0 & -B_{L_2 C_2} & -B_{L_2 G} \\
 \hline
 0 & 0 & 0 & | & \tilde{G}_{C_{11}} & \tilde{G}_{C_{12}} & 0 \\
 Q_{C_2 R} & 0 & Q_{C_2 L_2} & | & \tilde{G}_{C_{21}} & \tilde{G}_{C_{22}} & 0 \\
 Q_{GR} & 0 & Q_{GL_2} & | & 0 & 0 & G_G
 \end{bmatrix}
 \begin{bmatrix}
 G_R a_R \\
 G_{L_1} a_{L_1} \\
 G_{L_2} a_{L_2} - B_{12}^T G_{L_1} a_{L_1} \\
 \hline
 a_{C_1} \\
 a_{C_2} - Q_{12}^T a_{C_1} \\
 a_G
 \end{bmatrix}$$

(3.38)

where

$$\begin{aligned}
 \tilde{R}_L &= \begin{bmatrix} \tilde{R}_{L_{11}} & \tilde{R}_{L_{12}} \\ \tilde{R}_{L_{21}} & \tilde{R}_{L_{22}} \end{bmatrix} \\
 &= \begin{bmatrix} B_{13} \\ E_{L_2 \Gamma} \end{bmatrix} R_\Gamma \begin{bmatrix} B_{13}^T & B_{L_2 \Gamma}^T \end{bmatrix} + \hat{R}_L
 \end{aligned}
 \tag{3.39}$$

and

$$\begin{aligned}
 \tilde{G}_C &= \begin{bmatrix} \tilde{G}_{C_{11}} & \tilde{G}_{C_{12}} \\ \tilde{G}_{C_{21}} & \tilde{G}_{C_{22}} \end{bmatrix} \\
 &= \begin{bmatrix} Q_{11} \\ Q_{C_2 S} \end{bmatrix} G_S \begin{bmatrix} Q_{11}^T & Q_{C_2 S}^T \end{bmatrix} + \hat{G}_C
 \end{aligned}
 \tag{3.40}$$

are both positive definite symmetric matrices and thus

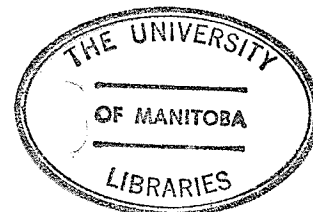
$$\tilde{R}_{L11} = B_{13} R_{13}^T + \hat{R}_{L11} \quad (3.41a)$$

$$\tilde{G}_{C11} = Q_{11} G_{11}^T + \hat{G}_{C11} \quad (3.41b)$$

are both positive definite.

Since (3.38) was produced by a change of variables followed by a variable elimination, if we now impose the remaining constraint equations for inductance loops (3.29) and for capacitance cutsets (3.26), then both the right-and left-hand sides of the second and fourth equations are independently equal to zero. Then, since \tilde{R}_{L11} and \tilde{G}_{C11} are both positive definite, variables a_{L1} , a_{C1} , b_{L1} and b_{C1} can be eliminated. We then have

$$\begin{bmatrix} R_R & 0 & B_{RC_2} & B_{RG} \\ 0 & R_L & B_{L_2 C_2} & B_{L_2 G} \\ \hline Q_{C_2 R} & Q_{C_2 L_2} & G_C & 0 \\ Q_{GR} & Q_{GL_2} & 0 & G_G \end{bmatrix} \begin{bmatrix} G_R b_R \\ G_{L_2} b_{L_2} - B_{12}^T G_{L_1} b_{L_1} \\ \hline b_{C_2} - Q_{12}^T b_{C_1} \\ b_G \end{bmatrix} = \begin{bmatrix} -R_R & 0 & -B_{RC_2} & -B_{RG} \\ 0 & -R_L & -B_{L_2 C_2} & -B_{L_2 G} \\ \hline Q_{C_2 R} & Q_{C_2 L_2} & G_C & 0 \\ Q_{GR} & Q_{GL_2} & 0 & G_G \end{bmatrix} \begin{bmatrix} G_R a_R \\ G_{L_2} a_{L_2} - B_{12}^T G_{L_1} a_{L_1} \\ \hline a_{C_2} - Q_{12}^T a_{C_1} \\ a_G \end{bmatrix} \quad (3.42)$$



where

$$R_{\mathcal{L}} = \tilde{R}_{L_{22}} - \tilde{R}_{L_{21}} \tilde{G}_{L_{11}}^{-1} \tilde{R}_{L_{12}} \quad (3.43a)$$

$$G_{\mathcal{C}} = \tilde{G}_{C_{22}} - \tilde{G}_{C_{21}} \tilde{R}_{C_{11}}^{-1} \tilde{G}_{C_{12}} \quad (3.43b)$$

The positive definite symmetric matrix

$$\begin{bmatrix} U & 0 \\ -\tilde{R}_{L_{21}} \tilde{G}_{L_{11}}^{-1} & U \end{bmatrix} \begin{bmatrix} \tilde{R}_{L_{11}} & \tilde{R}_{L_{12}} \\ \tilde{R}_{L_{21}} & \tilde{R}_{L_{22}} \end{bmatrix} \begin{bmatrix} U & -\tilde{G}_{L_{11}}^{-1} \tilde{R}_{L_{12}} \\ 0 & U \end{bmatrix} = \begin{bmatrix} \tilde{R}_{L_{11}} & 0 \\ 0 & R_{\mathcal{L}} \end{bmatrix}$$

demonstrates that $R_{\mathcal{L}}$ is also positive definite and symmetric.

Similarly,

$$\begin{bmatrix} U & 0 \\ -\tilde{G}_{C_{21}} \tilde{R}_{C_{11}}^{-1} & U \end{bmatrix} \begin{bmatrix} \tilde{G}_{C_{11}} & \tilde{G}_{C_{12}} \\ \tilde{G}_{C_{21}} & \tilde{G}_{C_{22}} \end{bmatrix} \begin{bmatrix} U & -\tilde{R}_{C_{11}}^{-1} \tilde{G}_{C_{12}} \\ 0 & U \end{bmatrix} = \begin{bmatrix} \tilde{G}_{C_{11}} & 0 \\ 0 & G_{\mathcal{C}} \end{bmatrix}$$

shows that $G_{\mathcal{C}}$ is a positive definite symmetric matrix.

The introduction of four new variables

$$b_{\mathcal{L}} = R_{\mathcal{L}} (G_{L_2} b_{L_2} - B_{L_2}^T G_{L_1} b_{L_1}) \quad (3.44a)$$

$$b_{\mathcal{C}} = b_{C_2} - Q_{L_2}^T b_{C_1} \quad (3.44b)$$

$$a_{\mathcal{L}} = R_{\mathcal{L}} (G_{L_2} a_{L_2} - B_{L_2}^T G_{L_1} a_{L_1}) \quad (3.44c)$$

$$a_{\mathcal{C}} = a_{C_2} - Q_{L_2}^T a_{C_1} \quad (3.44d)$$

allows (3.42) to be written as

$$\begin{bmatrix} U & 0 & B_{RC_2} & B_{RG} \\ 0 & U & B_{L_2C_2} & B_{L_2G} \\ \hline Q_{C_2R}^G & Q_{C_2L_2}^G & G_G & 0 \\ Q_{GR}^G & Q_{GL_2}^G & 0 & G_G \end{bmatrix} \begin{bmatrix} b_R \\ b_G \\ \hline b_G \\ b_G \end{bmatrix} = \begin{bmatrix} -U & 0 & -B_{RC_2} & -B_{RG} \\ 0 & -U & -B_{L_2C_2} & -B_{L_2G} \\ \hline Q_{C_2R}^G & Q_{C_2L_2}^G & G_G & 0 \\ Q_{GR}^G & Q_{GL_2}^G & 0 & G_G \end{bmatrix} \begin{bmatrix} a_R \\ a_G \\ \hline a_G \\ a_G \end{bmatrix} \quad (3.45)$$

Equation (3.45) has the same form as equation (2.17) and thus the coefficient matrix on the left-hand side can be inverted, producing

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 2P^T K - U & 2P^T (U - KP^T) \\ 2K & U - 2KP^T \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \quad (3.46)$$

where

$$P = \begin{bmatrix} Q_{C_2R} & Q_{C_2L_2} \\ Q_{GR} & Q_{GL_2} \end{bmatrix} \quad (3.47a)$$

$$-P^T = \begin{bmatrix} B_{RC_2} & B_{L_2C_2} \\ B_{RG} & B_{L_2G} \end{bmatrix} \quad (3.47b)$$

$$K = Y^{-1} P G_1, \quad Y = P G_1 P^T + G_2 \quad (3.48)$$

$$G_1 = \begin{bmatrix} G_R & 0 \\ 0 & G_x \end{bmatrix} \quad (3.49a)$$

$$G_2 = \begin{bmatrix} G_x & 0 \\ 0 & G_G \end{bmatrix} \quad (3.49b)$$

$$a_1 = \begin{bmatrix} a_R^T & a_x^T \end{bmatrix}^T, \quad a_2 = \begin{bmatrix} a_x^T & a_G^T \end{bmatrix}^T \quad (3.50a)$$

and

$$b_1 = \begin{bmatrix} b_R^T & b_x^T \end{bmatrix}^T, \quad b_2 = \begin{bmatrix} b_x^T & b_G^T \end{bmatrix}^T. \quad (3.50b)$$

The structure of (3.46) is identical to that of equation (2.21) and hence (3.46) can be expressed in the various alternate forms given in Chapter 2 (see (2.22) - (2.24)). The matrices P and $-P^T$, (3.47), are submatrices of Q_ℓ and B_t while G_1 and G_2 , (3.49), are no longer diagonal but are now block diagonal. The relationship between K in (3.46) and in (2.21) will be discussed in the next section.

In order to complete the filter, the ports of the adaptor must be suitably terminated. The source and load ports are terminated as discussed in Chapter 2. The remaining terminations can be derived from the definition of the \mathcal{L} and \mathcal{Q} variables together with the standard capacitance and inductance termination equations. Using the waves defined with respect to the n-port rather than the elements

$$\begin{bmatrix} b_{L1} \\ b_{L2} \\ b_\Gamma \end{bmatrix} = -\frac{1}{z} \begin{bmatrix} a_{L1} \\ a_{L2} \\ a_\Gamma \end{bmatrix}$$

together with (3.44a) yields

$$b_{\mathcal{L}} = -\frac{1}{z} a_{\mathcal{L}}. \quad (3.51)$$

Similarly,

$$\begin{bmatrix} b_{C_1} \\ b_{C_2} \\ b_S \end{bmatrix} = \frac{1}{z} \begin{bmatrix} a_{C_1} \\ a_{C_2} \\ a_S \end{bmatrix}$$

together with (3.44b) produces

$$b_{\mathcal{C}} = \frac{1}{z} a_{\mathcal{C}}. \quad (3.52)$$

These terminations, (3.51) and (3.52), correspond to a delay in series with an inverter for the "inductance" (\mathcal{L}) ports and a delay for the "capacitance" (\mathcal{C}) ports.

The dimension of the state space, m , and hence the number of delay elements required in the realization, is given by

$$m = \text{number of } \mathcal{L} \text{ elements} + \text{number of } \mathcal{C} \text{ elements} \quad (3.53)$$

which can be shown to be equivalent to

$$m = (\#L - \#L_1 - \#\Gamma) + (\#C - \#C_1 - \#S) \quad (3.54)$$

where, with respect to the prototype network,

$\#L$ = number of inductances

$\#L_1$ = number of independent inductance loops

$\#\Gamma$ = number of independent inductance cutsets

$\#C$ = number of capacitances

$\#C_1$ = number of independent capacitance cutsets

$\#S$ = number of independent capacitance loops.

In order to demonstrate further that the system obtained by the procedure developed in this section correctly describes the input-output properties of the original system, Appendix A uses the results of this section to prove that a similarity transformation exists which decouples the undesirable modes at $z = -1$ and $z = 1$.

3.4 NETWORK INTERPRETATION OF K

As was the case in Chapter 2, the multipliers in an adaptor realization can be restricted to lie in the submatrix K. The dimensions of K have been reduced from $t \times \ell$ to $\hat{t} \times \hat{\ell}$ where

$$\hat{t} = t - \#C_1 - \#S \quad (3.55a)$$

$$\hat{\ell} = \ell - \#L_1 - \#\Gamma \quad (3.55b)$$

thus reducing the upper bound on the number of multipliers required.

A wide variety of realizations of K is possible. One interesting but comparatively expensive solution would be to perform all of the multiplications in parallel, producing a filter having an extremely short computation time. At the other extreme would be a potentially inexpensive and relatively slow realization using a single multiplexed multiplier. A wave flow diagram of K using $n - 1$ independent multipliers is of interest since, then, if the n - element analog prototype is designed to have maximum transducer power gain at some frequency in the passband, the resulting zero sensitivity property with respect to the element values is maintained in the wave digital filter. The network inter-

pretation of K developed in this section often allows such a set of independent multipliers to be found.

First let us simplify the variable a_x in the adaptor representation (3.46). Substituting for R_x from (3.43) and expanding terms produces

$$a_x = \tilde{R}_{L_{22}} (G_{L_2 a_{L_2}} - B_{12}^T G_{L_1 a_{L_1}}) - \tilde{R}_{L_{21}} \tilde{G}_{L_{11}} \tilde{R}_{L_{12}} (G_{L_2 a_{L_2}} - B_{12}^T G_{L_1 a_{L_1}}). \quad (3.56)$$

Consider first the expression

$$\tilde{R}_{L_{12}} (G_{L_2 a_{L_2}} - B_{12}^T G_{L_1 a_{L_1}}). \quad (3.57)$$

Substituting for $\tilde{R}_{L_{12}}$ from (3.39) produces, upon expansion,

$$B_{13} R_{L_2} B_{L_2}^T G_{L_2 a_{L_2}} - B_{13} R_{L_2} B_{L_2}^T B_{12}^T G_{L_1 a_{L_1}} + B_{12} a_{L_2} - B_{12} R_{L_2} B_{12}^T G_{L_1 a_{L_1}}. \quad (3.58)$$

Now, using the orthogonality conditions (3.11c) and (3.11d) in the second term of (3.58) yields the equivalent form

$$B_{12} a_{L_2} + B_{13} R_{L_2} B_{L_2}^T G_{L_2 a_{L_2}} + B_{13} R_{L_2} B_{L_2}^T G_{L_1 a_{L_1}} - B_{13} R_{L_2} B_{12}^T G_{L_1 a_{L_1}} - B_{12} R_{L_2} B_{12}^T G_{L_1 a_{L_1}}. \quad (3.59)$$

The second and third terms in (3.59) can now be rewritten using constraint (3.28a) producing

$$B_{12} a_{L_2} + B_{13} a_{L_1} - (B_{13} R_{L_2} B_{L_2}^T + B_{12} R_{L_2} B_{12}^T) G_{L_1 a_{L_1}}. \quad (3.60)$$

Use of constraint (3.29a) then yields

$$-(R_{L_1} + B_{12} R_{L_2} B_{12}^T + B_{13} R_{L_2} B_{L_2}^T) G_{L_1 a_{L_1}}. \quad (3.61)$$

Noting that the term in the parentheses in (3.61) is equal to $\tilde{R}_{L_{11}}$

(see (3.36) and (3.39)), we have

$$-\tilde{R}_{L_{11}}^T G_{L_1} a_{L_1} = \tilde{R}_{L_{12}}^T (G_{L_2} a_{L_2} - B_{L_2}^T G_{L_1} a_{L_1}) \quad (3.62)$$

and hence, from (3.56),

$$a_{\chi} = \tilde{R}_{L_{22}}^T (G_{L_2} a_{L_2} - B_{L_2}^T G_{L_1} a_{L_1}) + \tilde{R}_{L_{21}}^T G_{L_1} a_{L_1}. \quad (3.63)$$

Substituting for $\tilde{R}_{L_{22}}^T$ and $\tilde{R}_{L_{21}}^T$ from (3.39) and simplifying produces

$$a_{\chi} = a_{L_2} + B_{L_2}^T \Gamma^R (B_{L_2}^T \Gamma^G a_{L_2} - B_{L_2}^T \Gamma^{B_{L_2}^T} G_{L_1} a_{L_1} + B_{L_3}^T G_{L_1} a_{L_1}). \quad (3.64)$$

Insertion of the orthogonality conditions (3.11c) and (3.11d) yields

$$a_{\chi} = a_{L_2} + B_{L_2}^T \Gamma^R (B_{L_2}^T \Gamma^G a_{L_2} + B_{L_1}^T \Gamma^G a_{L_1}). \quad (3.65)$$

Finally, using constraint (3.28a), a_{χ} becomes

$$a_{\chi} = a_{L_2} + B_{L_2}^T \Gamma^R a_{\Gamma}. \quad (3.66)$$

Following a similar procedure, it can be shown that

$$b_{\chi} = b_{L_2} + B_{L_2}^T \Gamma^R b_{\Gamma}. \quad (3.67)$$

From (3.46), the matrix K can now be defined by

$$\begin{bmatrix} b_{C_2} - Q_{12}^T b_{C_1} \\ b_G \end{bmatrix} = 2K \begin{bmatrix} a_R \\ a_{L_2} + B_{L_2}^T \Gamma^R a_{\Gamma} \end{bmatrix} \quad (3.68)$$

when

$$\begin{bmatrix} a_{C_2} - Q_{12}^T a_{C_1} \\ a_G \end{bmatrix} = 0. \quad (3.69)$$

In order to interpret K with the aid of the original network, it is necessary to obtain terminations which simultaneously guarantee that

(3.69) is satisfied and which allow a simple interpretation for the excitations and responses in (3.68). In addition, the constraint equations (3.26) - (3.29) must be satisfied. We shall, however, consider only those constraints in the incident waves, since it is easily demonstrated from Section 3.2 that, together with the natural behaviour of the system, these equations are sufficient for (3.26) - (3.29).

Since the terminations required at the capacitive, inductive and resistive ports are independent, we shall consider each class in turn. The capacitive constraint equations together with the first equation in (3.69) require that

$$\begin{bmatrix} U & 0 & -Q_{12}^T \\ B_{SC2} & U & B_{SC1} \\ Q_{12}G_{C2} & Q_{11}G_S & G_{C1} \end{bmatrix} \begin{bmatrix} a_{C2} \\ a_S \\ a_{C1} \end{bmatrix} = 0. \quad (3.70)$$

A series of elementary row operations produces an equivalent system in the form

$$\begin{bmatrix} U & 0 & A_{13} \\ 0 & U & A_{23} \\ 0 & 0 & A_{33} \end{bmatrix} \begin{bmatrix} a_{C2} \\ a_S \\ a_{C1} \end{bmatrix} = 0 \quad (3.71)$$

where $A_{33} = G_{C1} + Q_{12}G_{C2}Q_{12}^T + Q_{11}G_SQ_{11}^T$ is positive definite and hence the unique solution is the trivial solution

$$\begin{bmatrix} a_{C2}^T & a_S^T & a_{C1}^T \end{bmatrix}^T = 0. \quad (3.72)$$

This condition, which is similar to that encountered in the previous chapter, is satisfied if all capacitive ports are terminated in their reference resistances. The response $b_{C_2} - Q_{12}^T b_{C_1}$ can be computed from

$$b_{C_2} - Q_{12}^T b_{C_1} = 2(v_{C_2} - Q_{12}^T v_{C_1}) . \quad (3.73)$$

The second condition in (3.69) can be obtained by terminating all resistive twig ports in their reference resistances. Then the response b_G can be computed from

$$b_G = 2v_G . \quad (3.74)$$

The source variable a_R can be obtained by terminating the resistive link ports in their reference resistance in series with a voltage source. Then

$$a_R = e_R . \quad (3.75)$$

Finally, consider the source variable $a_{L_2} + B_{L_2} \Gamma a_\Gamma$ as well as the inductive constraints written in matrix form

$$\begin{bmatrix} U & B_{12} & B_{13} \\ B_{L_1}^T \Gamma^G_{L_1} & B_{L_2} \Gamma^G_{L_2} & -G_\Gamma \end{bmatrix} \begin{bmatrix} a_{L_1} \\ a_{L_2} \\ a_\Gamma \end{bmatrix} = 0 . \quad (3.76)$$

A change of variables allows the following equivalent system to be obtained:

$$\begin{bmatrix} U & B_{L_1} \Gamma \\ 0 & A_{22} \end{bmatrix} \begin{bmatrix} a_{L_1} \\ a_\Gamma \end{bmatrix} = \begin{bmatrix} -B_{12} \\ B_{L_2}^T \Gamma^G_{L_2} - B_{L_1}^T \Gamma^G_{L_1} B_{12} \end{bmatrix} \begin{bmatrix} a_{L_2} + B_{L_2} \Gamma a_\Gamma \end{bmatrix} \quad (3.77)$$

where $A_{22} = G_{\Gamma} + B_{L_1\Gamma}^T G_{L_1} B_{L_1\Gamma} + B_{L_2\Gamma}^T G_{L_2} B_{L_2\Gamma}$ is positive definite. It is evident from (3.77) that the variable $a_{L_2} + B_{L_2\Gamma} a_{\Gamma}$ can be independently specified. Furthermore, from (3.11) and (3.29a),

$$a_{L_1} + B_{L_1\Gamma} a_{\Gamma} = -B_{12}(a_{L_2} + B_{L_2\Gamma} a_{\Gamma}) . \quad (3.78)$$

The variables $a_{L_1} + B_{L_1\Gamma} a_{\Gamma}$ and $a_{L_2} + B_{L_2\Gamma} a_{\Gamma}$ have simple network interpretations. Assume that all of the inductive ports are terminated in their port resistances in series with a voltage source. (Of course, in light of (3.77), these sources cannot be chosen independently.) Then

$$a_{L_1} = \tilde{e}_{L_1} \quad (3.79a)$$

$$a_{L_2} = \tilde{e}_{L_2} \quad (3.79b)$$

$$a_{\Gamma} = \tilde{e}_{\Gamma} \quad (3.79c)$$

where \tilde{e}_{L_1} , \tilde{e}_{L_2} and \tilde{e}_{Γ} are the vectors of the inserted sources. The sources in the twigs Γ can be shifted into the links L_1 and L_2 , producing a total source contribution in L_2 of

$$\begin{aligned} e_{L_2} &= \tilde{e}_{L_2} + B_{L_2\Gamma} \tilde{e}_{\Gamma} \\ &= a_{L_2} + B_{L_2\Gamma} a_{\Gamma} \end{aligned} \quad (3.80)$$

and a total source contribution in L_1 of

$$\begin{aligned} e_{L_1} &= \tilde{e}_{L_1} + B_{L_1\Gamma} \tilde{e}_{\Gamma} \\ &= a_{L_1} + B_{L_1\Gamma} a_{\Gamma} \\ &= -B_{12} e_{L_2} . \end{aligned} \quad (3.81)$$

e_{L_2} can therefore be considered to be a vector of independent voltage sources while e_{L_1} is a vector of dependent voltage sources of value $e_{L_1} = -B_{12}e_{L_2}$. There are no longer any sources in the inductive twigs Γ . The response due to $a_{L_2} + B_{L_2}\Gamma a_{\Gamma}$, which consists of two components due to e_{L_1} and e_{L_2} , can be obtained by superposition.

A summary of the terminations required for the network interpretation of K is given in Fig. 3.1.

If \tilde{K} is defined by

$$\begin{bmatrix} v_{C_1} \\ v_{C_2} \\ v_G \end{bmatrix} = \tilde{K} \begin{bmatrix} e_R \\ e_{L_1} \\ e_{L_2} \end{bmatrix} \quad (3.82)$$

then K is given by

$$K = \begin{bmatrix} -Q_{12}^T & U & 0 \\ 0 & 0 & U \end{bmatrix} \tilde{K} \begin{bmatrix} U & 0 \\ 0 & -B_{12} \\ 0 & U \end{bmatrix}. \quad (3.83)$$

\tilde{K} can be obtained from the K for the non-minimal realization in the previous chapter by simply deleting those rows corresponding to the inductive twigs and those columns corresponding to the capacitive links. K can then be formed by pre- and post-multiplying \tilde{K} by the appropriate matrices, as shown in (3.83).

3.5 DESIGN PROCEDURE - ILLUSTRATIVE EXAMPLES

The design of a wave digital filter can be carried out as follows:

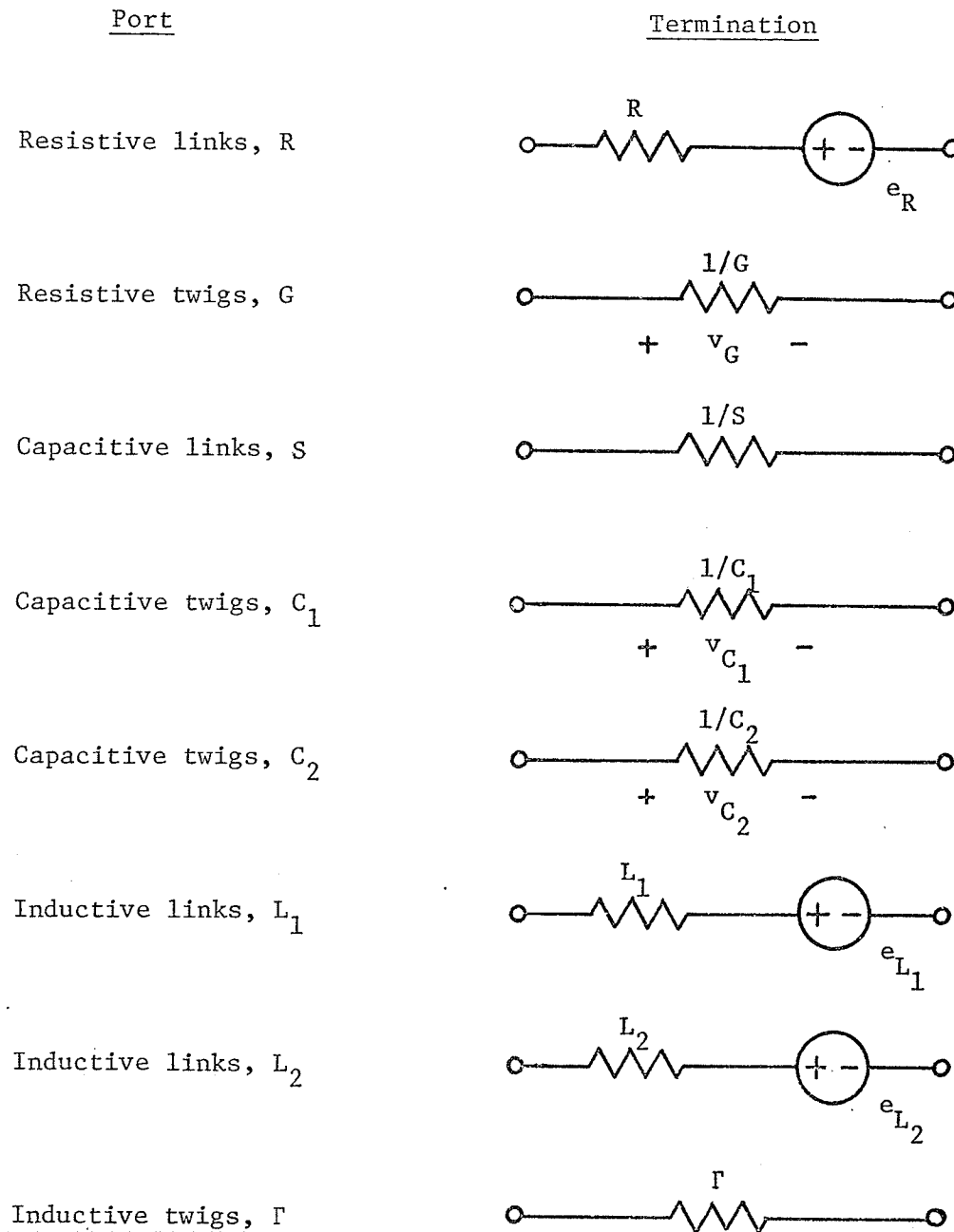


Fig. 3.1 Summary of the terminations required for the network interpretation of K .

1. From design tables or other available sources, choose an analog reference filter which meets the performance specifications.
2. Pre-warp the element values of the reference filter to allow for the nonlinear shift which will be introduced by the bilinear z-transformation.
3. From the network graph, determine the topological matrix P or $-P^T$, (see (3.47)) and, if required, B_{12} and Q_{12} , (see (3.3) and (3.4)).
4. Terminate the ports of the network of connections as shown in Fig. 3.1 and obtain K , (see (3.82) - (3.83)). The warped element values $\{e\}$ produce the multiplier coefficients $\{\alpha\}$.
5. Replace the coefficients $\{\alpha\}$ by a new set $\{\hat{\alpha}\}$ having suitable finite word length. If the realization contains only $n - 1$ independent multipliers, then the low coefficient sensitivity allows relatively drastic modifications to be made in $\{\alpha\}$.
6. If desired, the element values $\{\hat{e}\}$ corresponding to the independent multipliers $\{\hat{\alpha}\}$ can now be calculated. The frequency response of the digital filter with multipliers $\{\hat{\alpha}\}$ is identical to the warped response of the analog filter with element values $\{\hat{e}\}$.

Example 3.1:

This example consists of a double passband filter given by Watanabe [36]. The reference filter shown in Fig. 3.2 is eighth-order with four attenuation poles, two of which are finite. The network contains ten reactive elements, two of which are redundant due to the existence of a capacitance loop and a capacitance cutset. The transformation of structures of this type, where the capacitance loop and

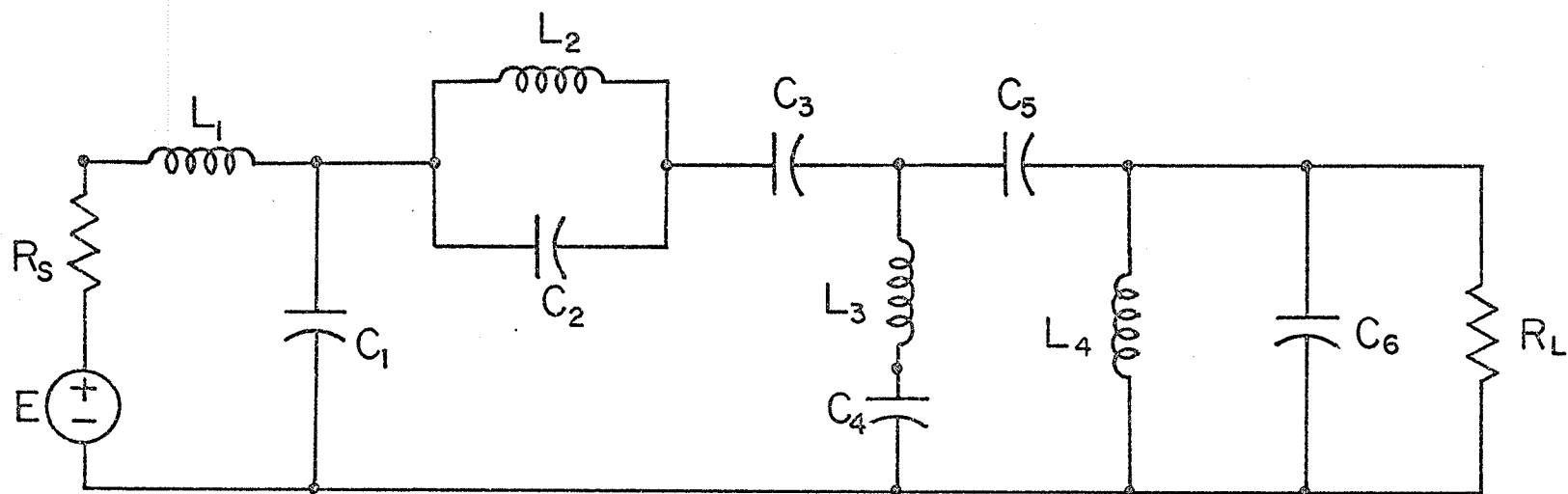


Fig. 3.2 Non-minimal eighth-order double passband reference filter.

cutset have common elements, have not been considered in the literature. The technique developed in this chapter can be used to obtain a minimal realization having the canonic number of multipliers and hence low coefficient sensitivity.

In Watanabe's filter the finite attenuation poles, λ_1 and λ_2 , occur at $\lambda_1^2 = 0.332031$ and $\lambda_2^2 = 0.717096$. A frequency equal to $1/4$ of the sampling frequency, that is $f = F_S/4$, is chosen for the upper attenuation pole in the digital response. The corresponding pole must therefore appear at $\phi = \tan \pi/2 = 1.0$ in the analog filter. The frequency scaling factor $k = \lambda_2$ produces the desired response with the following element values:

$C_1 = 0.223515$	$L_1 = 1.565397$
$C_2 = 0.518856$	$L_2 = 1.927316$
$C_3 = 1.366682$	$L_3 = 6.983054$
$C_4 = 0.309281$	$L_4 = 5.677186$
$C_5 = 0.886196$	$R_S = 1.0$
$C_6 = 0.116416$	$R_L = 8.589851$

As the value of R_L was not given by Watanabe, simulation of the analog filter for various values of R_L was carried out. A value of approximately 8 ohms was found to produce a response similar to Watanabe's. The value of R_L given above, $R_L = 1/C_6$, was chosen since then one of the multipliers becomes equal to $1/2$.

The network graph showing the tree chosen for the analysis is given in Fig. 3.3. The branches are numbered in the order given by equation (3.5), that is, S, R, L_2, C_1, C_2 and G . There are no elements in classes L_1 and Γ . The partition of the capacitive twigs was obtained

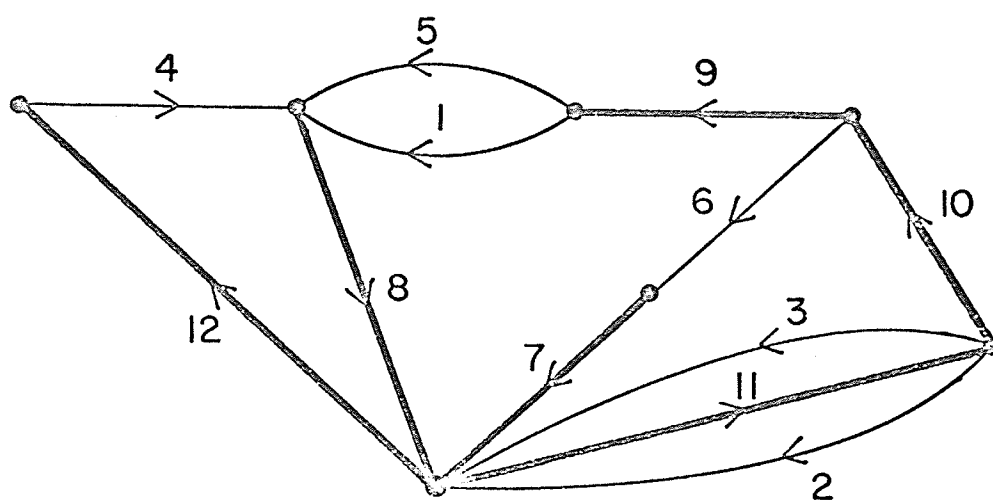


Fig. 3.3 Network graph corresponding to Fig. 3.2.

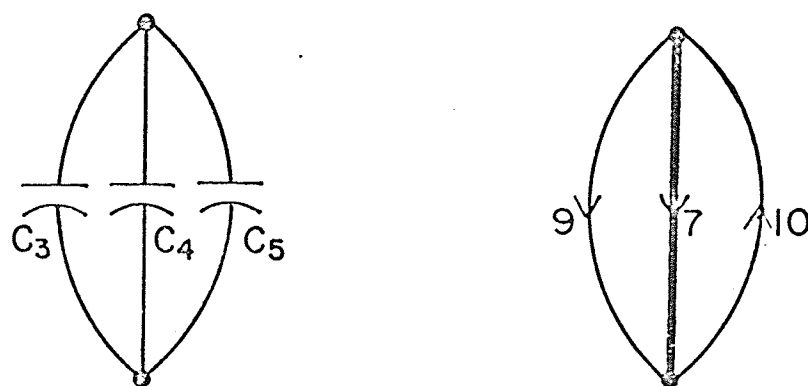


Fig. 3.4 Subnetwork N_{C_S} and corresponding graph.

with the aid of N_{CS} , Fig. 3.4. T_{CS} , and hence class C_1 , consists solely of branch 7.

The non-unit part of the fundamental loop matrix is then

$$B_t = \begin{array}{c} \begin{array}{cccccc} & C_1 & & C_2 & & G \\ & & \underbrace{\hspace{1.5cm}} & & & \\ 7 & 8 & 9 & 10 & 11 & 12 \end{array} \\ \left[\begin{array}{cccccc} 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \end{array} \right] \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{array}{c} S \\ R \\ \\ L_2 \\ \\ \end{array} \end{array}$$

Q_{12} , which is needed to produce K , is obtained from KCL in N_{CS} , (3.4).

$$Q_{12} = [0 \ 1 \ -1 \ 0].$$

A realization in the form of (2.23), Fig. 2.4, requires the topological matrix F

$$F = \begin{bmatrix} U & -P^T \\ 0 & U \end{bmatrix}$$

where $-P^T$ (see(3.47)) is obtained from B_t by striking out both the first row corresponding to the capacitive link in branch 1 and the first column corresponding to the capacitive twig in branch 7.

$$-P^T = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix}.$$

A flow graph realization of F contains only adders and inverters and is trivial.

The network used to obtain K is shown in Fig. 3.5. Thevenin's theorem can be applied successively to obtain a realization containing the canonic number (11) of multipliers. The following sequence of networks is produced:

Fig. 3.6, where

$$\begin{aligned} \alpha_1 &= \frac{R_{11}}{R_2 + R_{11}}, & R_{\alpha_1} &= \frac{R_2 R_{11}}{R_2 + R_{11}}, & e_{\alpha_1} &= \alpha_1 e_2 \\ \alpha_2 &= \frac{R_8}{R_4 + R_8 + R_{12}}, & R_{\alpha_2} &= \frac{R_8 (R_4 + R_{12})}{R_4 + R_8 + R_{12}}, & e_{\alpha_2} &= \alpha_2 e_4 \\ \alpha_3 &= \frac{R_1}{R_1 + R_5}, & R_{\alpha_3} &= \frac{R_1 R_5}{R_1 + R_5}, & e_{\alpha_3} &= \alpha_3 e_5 \end{aligned}$$

Fig. 3.7, where

$$\begin{aligned} \alpha_4 &= \frac{R_3}{R_3 + R_{\alpha_1}}, & R_{\alpha_4} &= \frac{R_3 R_{\alpha_1}}{R_3 + R_{\alpha_1}}, & e_{\alpha_4} &= \alpha_4 e_{\alpha_1} + (1 - \alpha_4) e_3 \end{aligned}$$

Fig. 3.8, where

$$\begin{aligned} \alpha_5 &= \frac{R_6 + R_7}{R_6 + R_7 + R_{10} + R_{\alpha_4}}, & R_{\alpha_5} &= \frac{(R_6 + R_7)(R_{10} + R_{\alpha_4})}{R_6 + R_7 + R_{10} + R_{\alpha_4}}, & e_{\alpha_5} &= \alpha_5 e_{\alpha_4} + (1 - \alpha_5) e_6 \end{aligned}$$

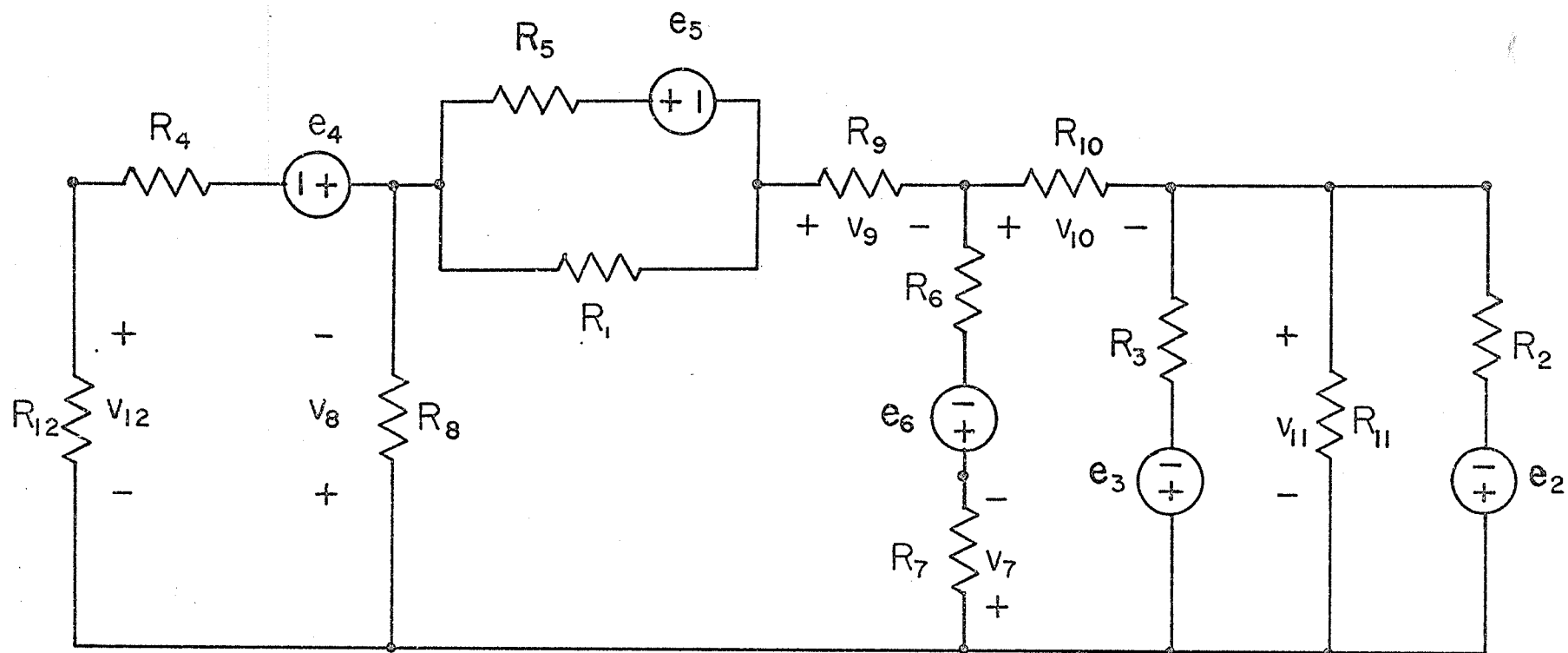


Fig. 3.5 Network used to obtain K for the filter of Fig. 3.2 (see also Figs. 3.6-3.9).

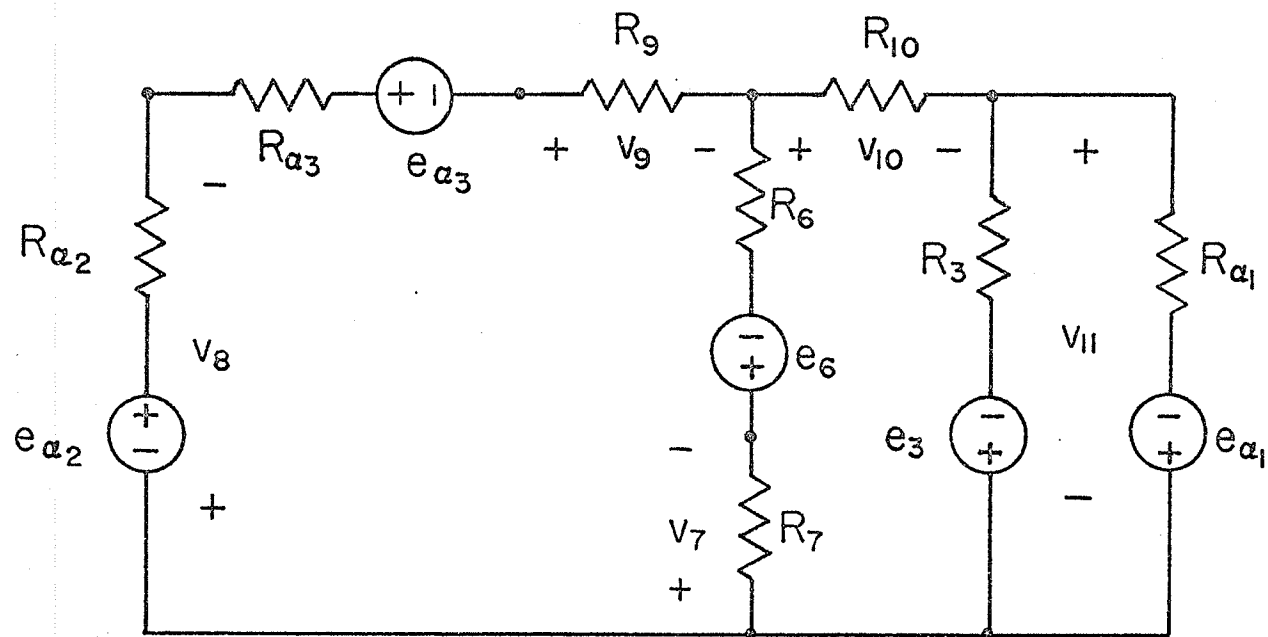


Fig. 3.6

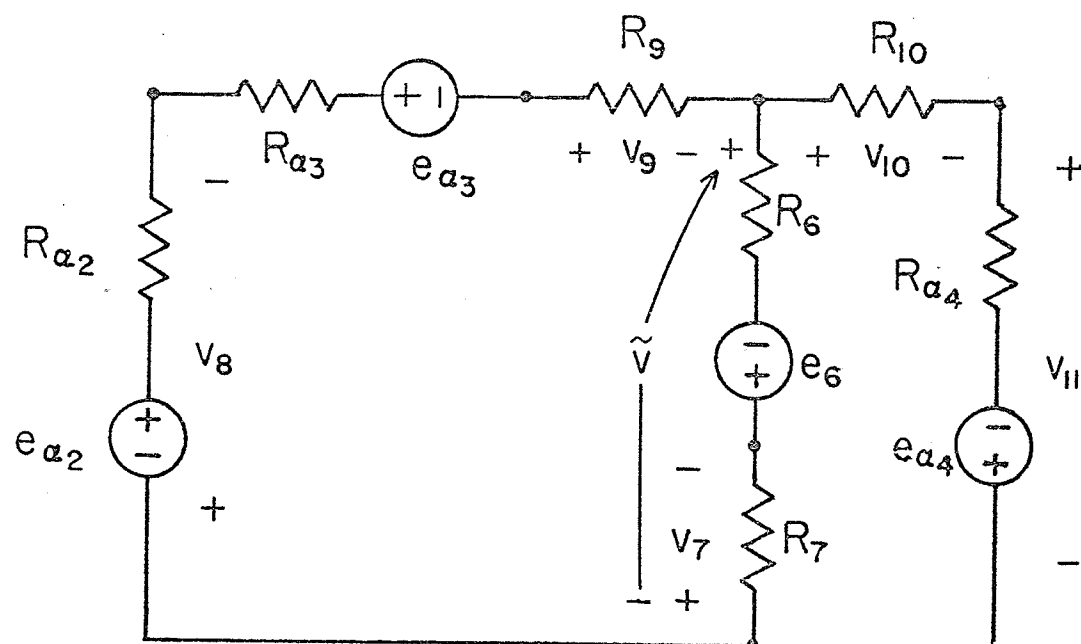


Fig. 3.7

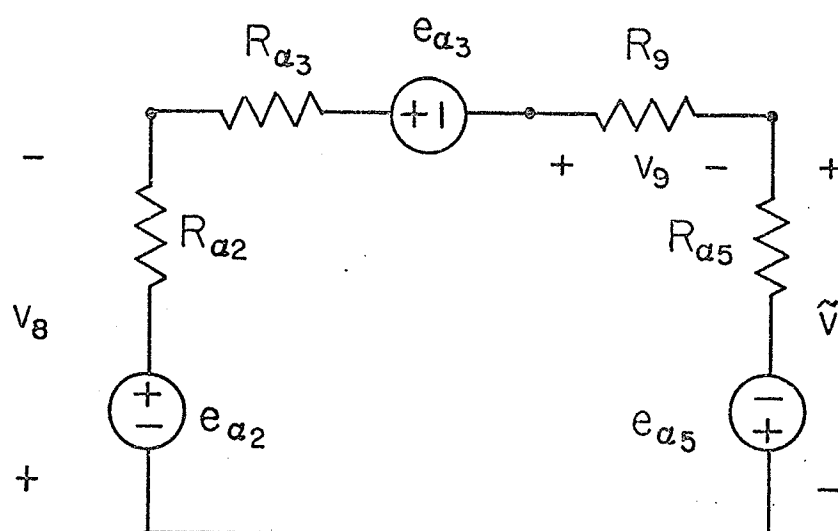


Fig. 3.8

$$\alpha_6 = \frac{R_{\alpha_2}}{R_T}, \quad v_8 = \alpha_6(e_{\alpha_5} - e_{\alpha_3}) + (\alpha_6 - 1)e_{\alpha_2}$$

$$\alpha_7 = \frac{R_9}{R_T}, \quad v_9 = \alpha_7(e_{\alpha_2} + e_{\alpha_5} - e_{\alpha_3})$$

$$\alpha_8 = \frac{R_{\alpha_5}}{R_T}, \quad \tilde{v} = \alpha_8(e_{\alpha_2} - e_{\alpha_3}) + (\alpha_8 - 1)e_{\alpha_5}$$

$$R_T = R_{\alpha_2} + R_{\alpha_3} + R_{\alpha_5} + R_9$$

Finally, Fig. 3.9, where

$$\alpha_9 = \frac{R_7}{R_6 + R_7}, \quad v_7 = -\alpha_9(\tilde{v} + e_6)$$

$$\alpha_{10} = \frac{R_{10}}{R_{10} + R_{\alpha_4}}, \quad v_{10} = \alpha_{10}(\tilde{v} + e_{\alpha_4})$$

$$\alpha_{11} = \frac{R_{12}}{R_{12} + R_4}, \quad v_{12} = -\alpha_{11}(e_4 + v_8)$$

$$v_{11} = \tilde{v} - v_{10}$$

From the above analysis, the responses v_7 through v_{12} can be expressed in terms of the sources e_1 through e_6 . This defines \tilde{K} . K is then formed from (3.83). The flow diagram for $2K$, shown in Fig. 3.10, is arranged so that each section (bounded by the dotted lines) contains the operations required by each of Fig. 3.6

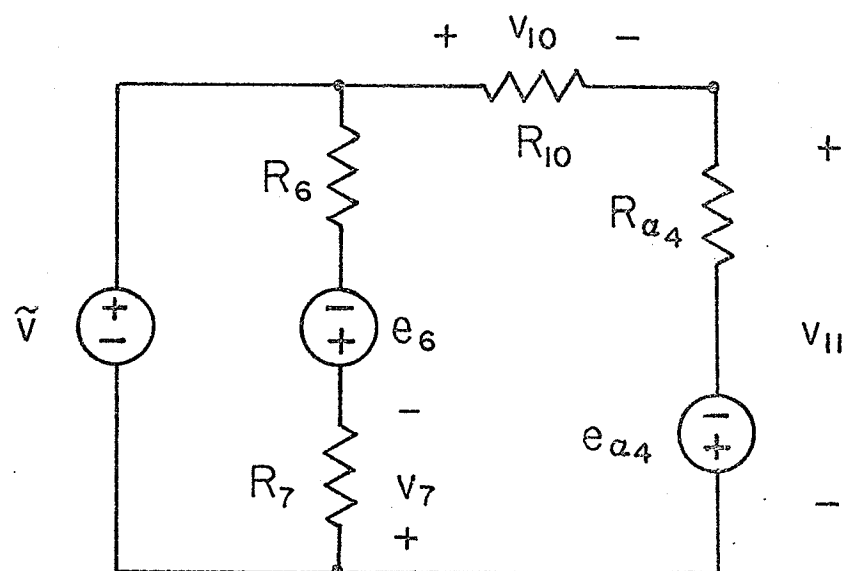
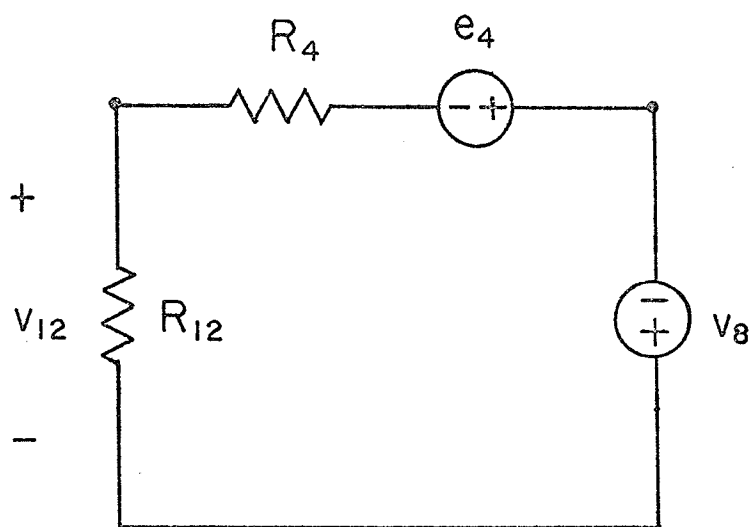


Fig. 3.9

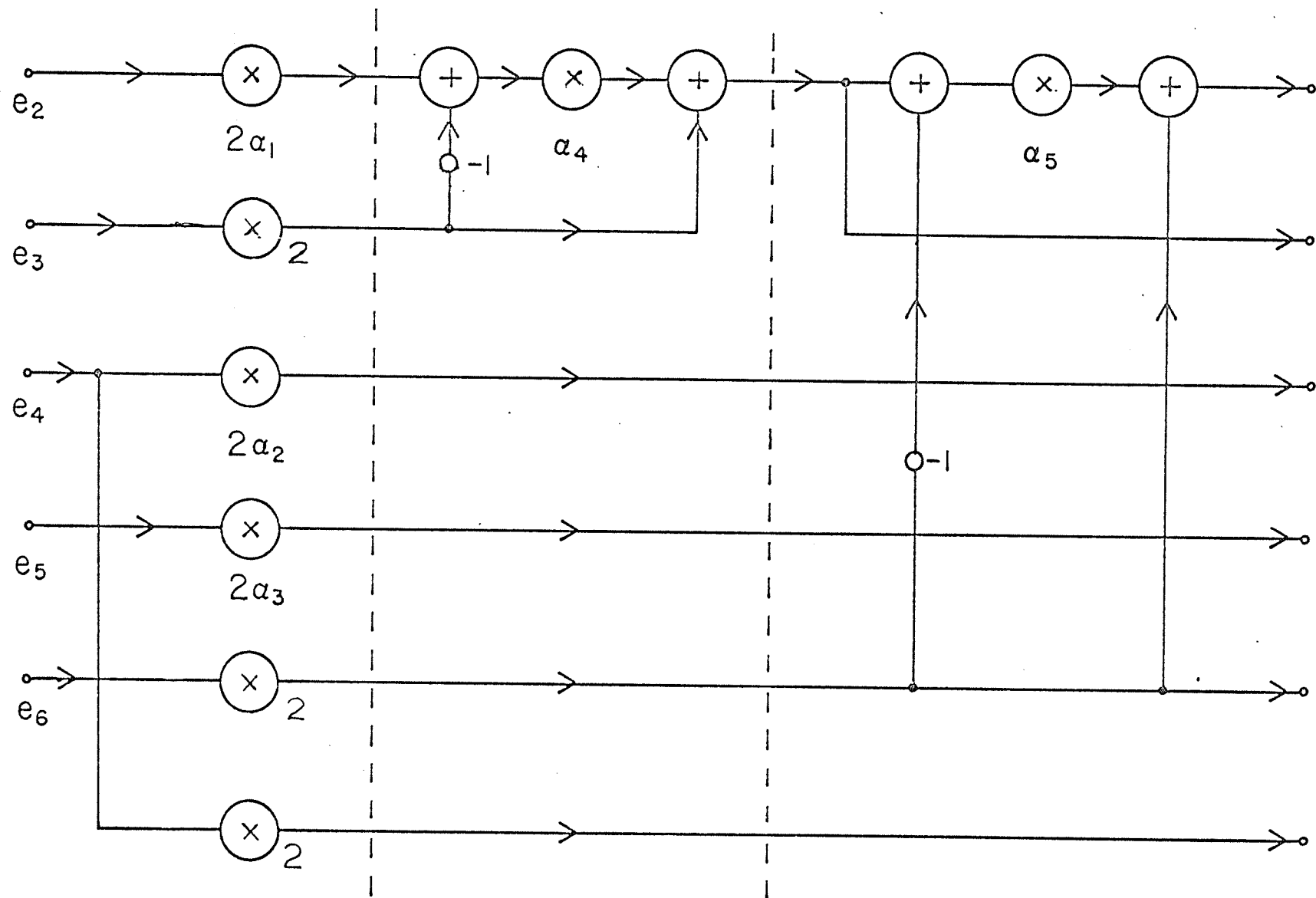


Fig. 3.10 Flow diagram for 2K for Example 3.1

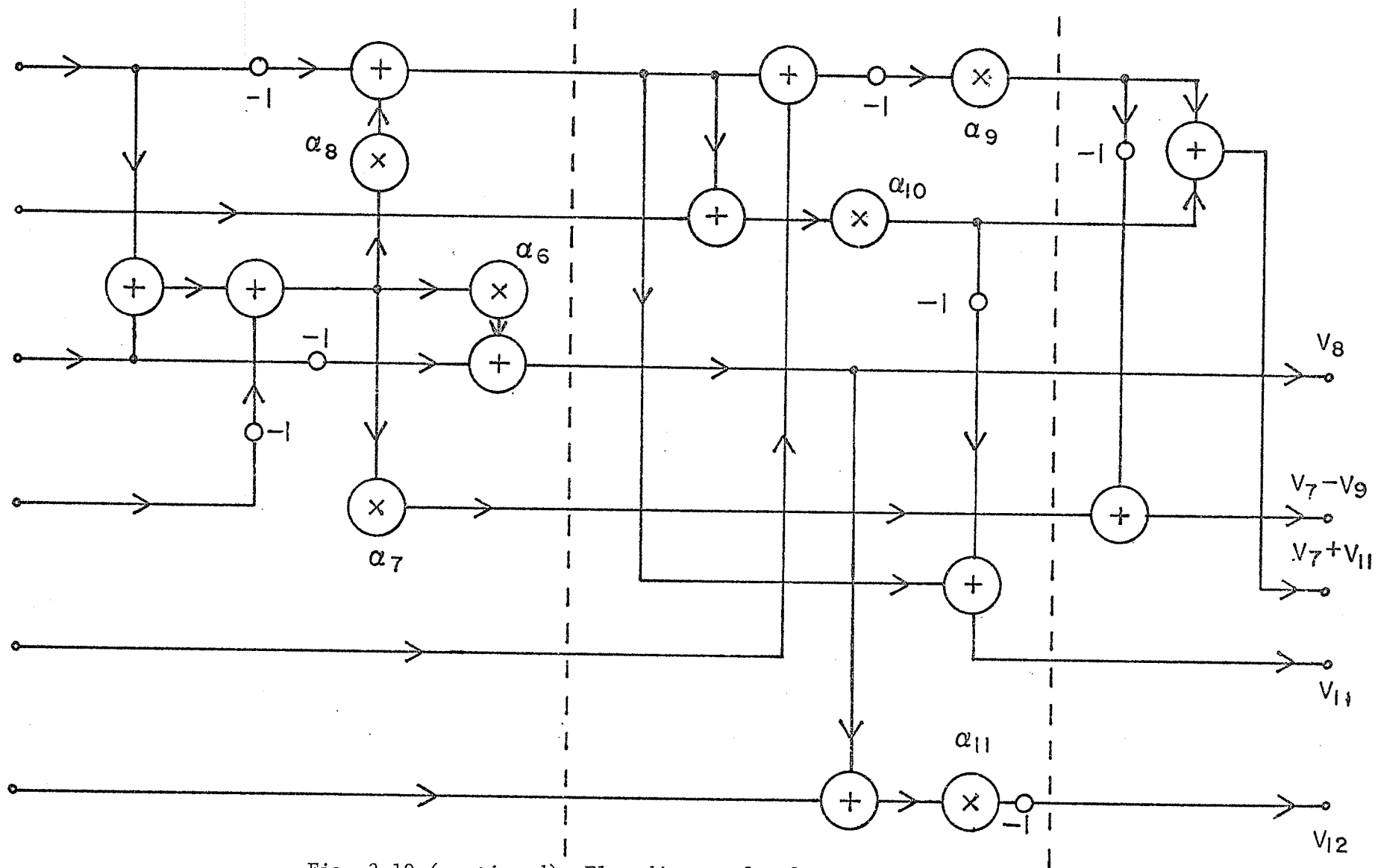


Fig. 3.10 (continued) Flow diagram for 2K

to Fig. 3.9. The last section is due to the effect of the capacitive cutset through Q_{12} . The factor of 2 has been inserted in the first section.

The flow diagram for the entire filter is easily obtained from the interconnection of F , $2K$, θ and Σ , Fig. 2.4. The adaptor input corresponding to branch 12 is the filter input. The adaptor input corresponding to branch 2 must be set equal to zero. The inverters required in θ and Σ cancel, with the exception of the adaptor output corresponding to branch 2, the filter output. This inverter is optional since its only effect is a simple inversion of the response. The adaptor output corresponding to branch 12 is not required. A total of 38 adders, 11 multipliers and 8 delays are required in the flow diagram.

From the equations which define the multiplier coefficients, it follows that for $R_i > 0$, $i = 1, 2, \dots, 12$, $0 < \alpha_j < 1$, $j = 1, 2, \dots, 11$ and $0 < \alpha_6 + \alpha_7 + \alpha_8 < 1$. Conversely, if α_j satisfy these conditions, then a set of $n - 1 = 11$ independent resistance ratios can be obtained from $\{\alpha\}$. The inversion of the equations which define the multipliers produces the following resistance ratios which have been normalized with respect to R_{12} :

$$\frac{R_1}{R_{12}} = \frac{\alpha_2 (1 - \alpha_6 - \alpha_7 - \alpha_8)}{\alpha_6 \alpha_{11} (1 - \alpha_3)},$$

$$\frac{R_2}{R_{12}} = \frac{\alpha_2 \alpha_8 (1 - \alpha_{10})}{\alpha_1 \alpha_4 \alpha_5 \alpha_6 \alpha_{11}}$$

$$\frac{R_3}{R_{12}} = \frac{\alpha_2 \alpha_8 (1 - \alpha_{10})}{\alpha_5 \alpha_6 \alpha_{11} (1 - \alpha_4)},$$

$$\frac{R_4}{R_{12}} = \frac{1 - \alpha_{11}}{\alpha_{11}}$$

$$\frac{R_5}{R_{12}} = \frac{\alpha_2 (1 - \alpha_6 - \alpha_7 - \alpha_8)}{\alpha_3 \alpha_6 \alpha_{11}},$$

$$\frac{R_6}{R_{11}} = \frac{\alpha_2 \alpha_8 (1 - \alpha_9)}{\alpha_6 \alpha_{11} (1 - \alpha_5)}$$

$$\frac{R_7}{R_{12}} = \frac{\alpha_2 \alpha_8 \alpha_9}{\alpha_6 \alpha_{11} (1 - \alpha_5)},$$

$$\frac{R_8}{R_{12}} = \frac{\alpha_2}{\alpha_{11} (1 - \alpha_2)}$$

$$\frac{R_9}{R_{12}} = \frac{\alpha_2 \alpha_7}{\alpha_6 \alpha_{11}},$$

$$\frac{R_{10}}{R_{12}} = \frac{\alpha_2 \alpha_8 \alpha_{10}}{\alpha_5 \alpha_6 \alpha_{11}}$$

$$\frac{R_{11}}{R_{12}} = \frac{\alpha_2 \alpha_8 (1 - \alpha_{10})}{\alpha_4 \alpha_5 \alpha_6 \alpha_{11} (1 - \alpha_1)},$$

The multiplier coefficients $\{\alpha\}$ obtained from $\{e\}$ are

$$\alpha_1 = 0.500000$$

$$\alpha_2 = 0.635564$$

$$\alpha_3 = 0.500000$$

$$\alpha_4 = 0.569306$$

$$\alpha_5 = 0.740853$$

$$\alpha_6 = 0.272960$$

$$\alpha_7 = 0.122495$$

$$\alpha_8 = 0.443219$$

$$\alpha_9 = 0.316483$$

$$\alpha_{10} = 0.315770$$

$$\alpha_{11} = 0.389803$$

For realizability, the coefficients are approximated by finite word length binary numbers as follows:

$$\hat{\alpha}_1 = 0.5 = 2^{-1}$$

$$\hat{\alpha}_2 = 0.6328125 = 2^{-1} + 2^{-3} + 2^{-7}$$

$$\hat{\alpha}_3 = 0.5 = 2^{-1}$$

$$\hat{\alpha}_4 = 0.5703125 = 2^{-1} + 2^{-4} + 2^{-7}$$

$$\hat{\alpha}_5 = 0.7421875 = 1 - 2^{-2} - 2^{-7}$$

$$\hat{\alpha}_6 = 0.2734375 = 2^{-2} + 2^{-5} - 2^{-7}$$

$$\hat{\alpha}_7 = 0.12109375 = 2^{-3} - 2^{-8}$$

$$\hat{\alpha}_8 = 0.44140625 = 2^{-1} - 2^{-4} + 2^{-8}$$

$$\hat{\alpha}_9 = 0.31640625 = 2^{-2} + 2^{-4} + 2^{-8}$$

$$\hat{\alpha}_{10} = 0.31640625 = 2^{-2} + 2^{-4} + 2^{-8}$$

$$\hat{\alpha}_{11} = 0.390625 = 2^{-1} - 2^{-3} + 2^{-6}$$

The corresponding element values $\{\hat{e}\}$ can be obtained exactly in fractional form. Approximate decimal values are given below:

$$\hat{C}_1 = 0.226659$$

$$\hat{L}_1 = 1.560000$$

$$\hat{C}_2 = 0.514403$$

$$\hat{L}_2 = 1.944000$$

$$\hat{C}_3 = 1.393867$$

$$\hat{L}_3 = 6.934091$$

$$\hat{C}_4 = 0.311576$$

$$\hat{L}_4 = 5.605665$$

$$\hat{C}_5 = 0.896960$$

$$\hat{R}_S = 1.00$$

$$\hat{C}_6 = 0.118387$$

$$\hat{R}_L = 8.446893$$

No attempt was made to minimize the coefficient word lengths.

If a discrete optimization procedure were used, more drastic changes in the coefficients would no doubt be possible.

Simulation of the filter with both $\{\alpha\}$ and $\{\hat{\alpha}\}$ was carried out on an IBM 370/158 computer using double precision floating-point arithmetic. The frequency responses shown in Figs. 3.11 and 3.12 were obtained using a 1024 point FFT of the unit sample responses. Due to the low sensitivity of the realization, the two responses are virtually identical. A minor shift in the level of the passband attenuation is the only discernable change.

A realization of this filter could also be obtained by first converting the capacitance cutset into a loop. The resulting structure is not a ladder and thus cannot be handled by the series-parallel adaptor technique. However, the n-port method can again be applied. The realization still only requires 11 multipliers since the bridging capacitor produced by the wye-delta transformation becomes a link in the graph and thus no source need be inserted to obtain K.

Although the form of G_1 and G_2 in (3.49) does not affect the realization, we shall see in the next chapter that these matrices do play an important role in the study of the properties of wave digital filters. G_1 and G_2 can be obtained directly from (3.49) with the aid of (3.43), (3.40) and (3.37). In this example G_R , G_G and G_Z are all diagonal and are easily obtained directly from the corresponding element values in the prototype network. However, G_C is nondiagonal due to the capacitive degeneracies. From N_{CS} , Fig. 3.4, we conclude that $Q_{11} = 0$. Then, using the appropriate formulas, it is straightforward to show that G_C is given by

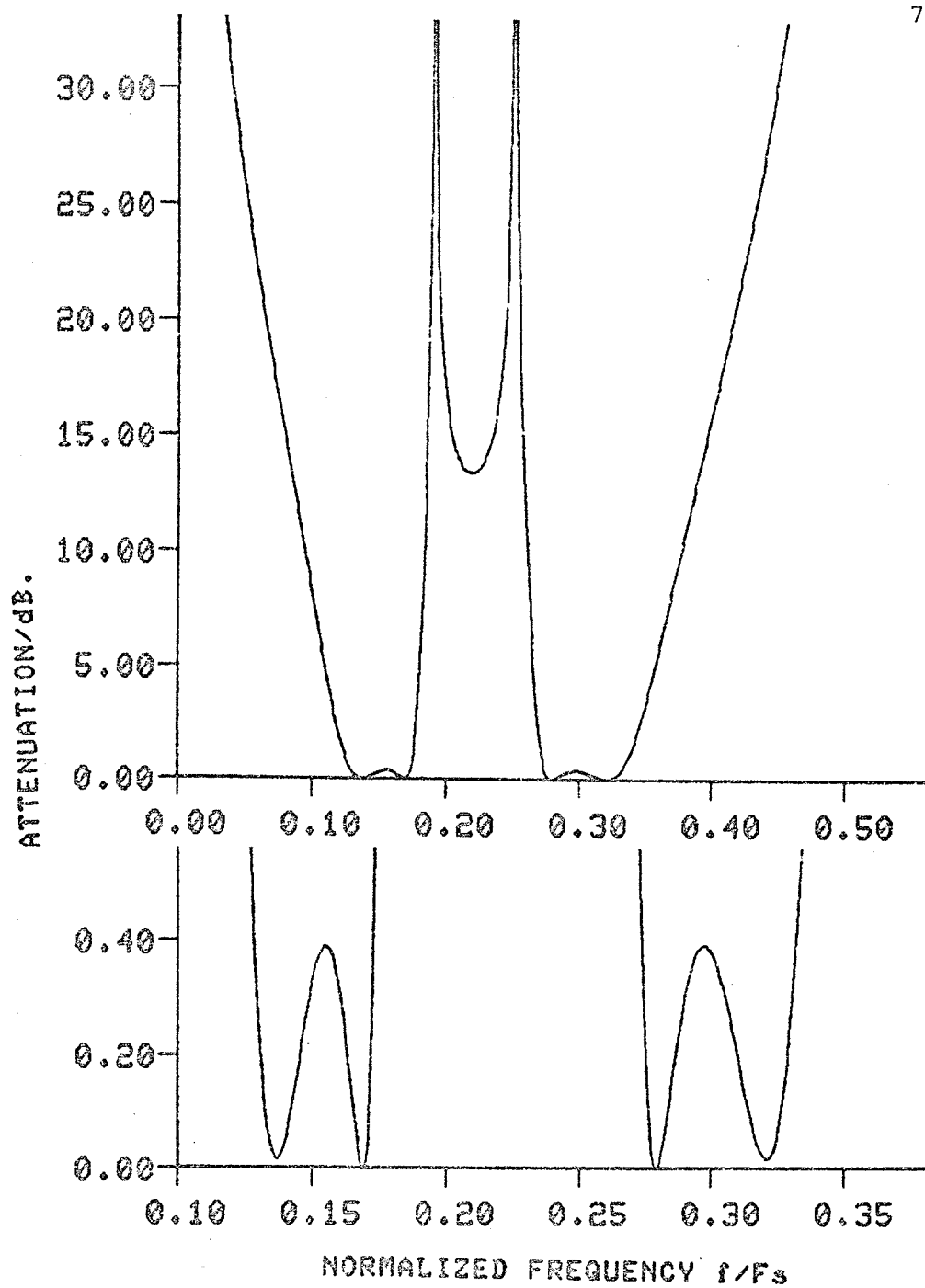


Fig. 3.11 Frequency response of the canonic wave digital realization of Fig. 3.2 using exact floating-point coefficients.

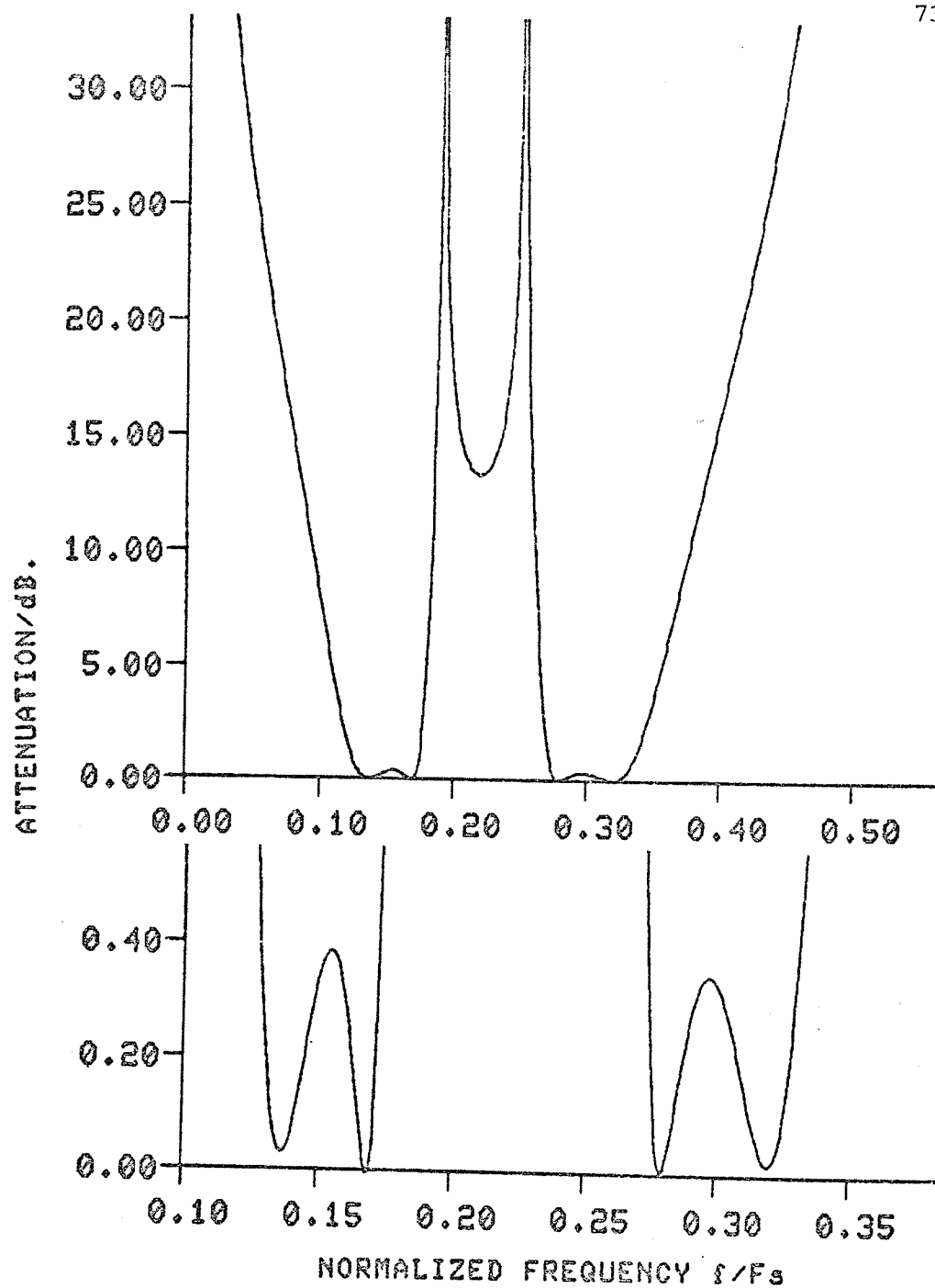


Fig. 3.12 Frequency response of the canonic wave digital realization of Fig. 3.2 using quantized coefficients.

$$G_{\mathcal{L}} = \begin{bmatrix} G_8 + G_1 & G_1 & G_1 & G_1 \\ G_1 & G_{\alpha} + G_{\gamma} & G_{\gamma} & G_1 \\ G_1 & G_{\gamma} & G_{\beta} + G_{\gamma} & G_1 \\ G_1 & G_1 & G_1 & G_{11} + G_1 \end{bmatrix}$$

where

$$G_{\alpha} = \frac{G_7 G_9}{G_7 + G_9 + G_{10}}, \quad G_{\beta} = \frac{G_7 G_{10}}{G_7 + G_9 + G_{10}}$$

$$G_{\gamma} = \frac{G_1 G_7 + G_1 G_9 + G_1 G_{10} + G_9 G_{10}}{G_7 + G_9 + G_{10}}$$

Example 3.2:

The second example consists of a fourteenth-order band-pass filter designed to pass the lower sideband of a telephone signal modulated at 8 kHz while operating at a sampling frequency of $F_S = 24$ kHz. The passband attenuation, which meets the 1/20 C.C.I.T.T. specification, and the stopband attenuation specification are shown in Fig. 3.20. Using Fettweis' adaptor method, Wegener [11] has designed a wave digital filter which meets this performance specification with very short coefficient word lengths. A discrete optimization procedure was used to minimize the multiplier hardware complexity. The lowpass prototype shown in Fig. 3.13 contains the element values which produce the desired response. This network was obtained from Wegener's adaptor realization by reversing his

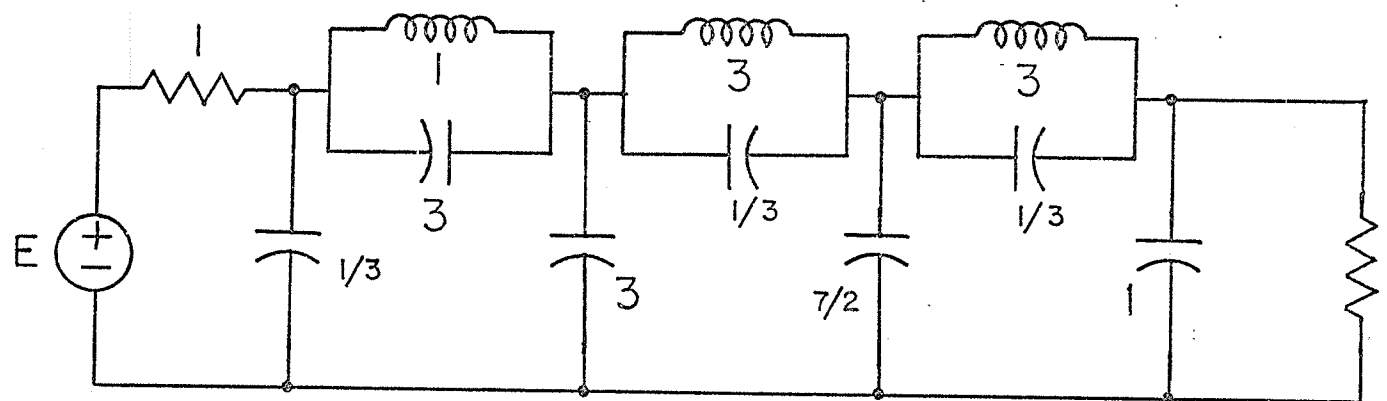


Fig. 3.13 Non-minimal seventh-order lowpass reference filter.

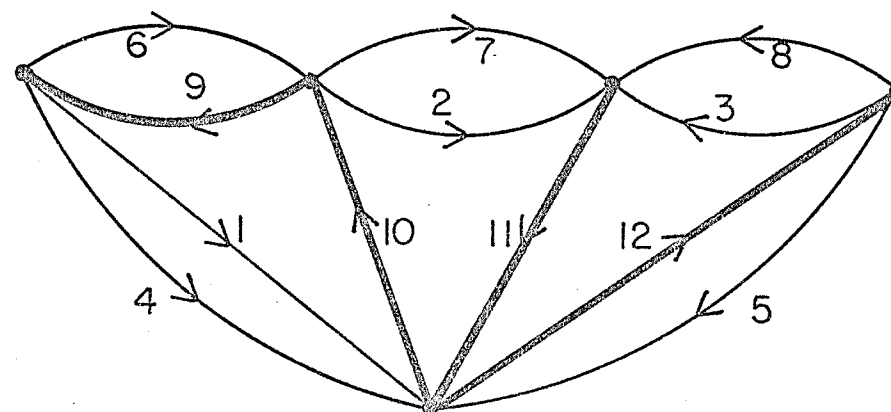


Fig. 3.14 Network graph corresponding to Fig. 3.13.

design procedure. The bandpass filter, which is arithmetically symmetric, is obtained by using a particular form of the standard digital lowpass-bandpass transformation [2] where z is replaced by $-z^2$. This can be accomplished by simply replacing each delay by a double delay in series with an inverter.

Since the same general procedure used in Example 3.1 will be followed here, only the major details of the design will be given.

The network graph showing the tree chosen is given in Fig. 3.14. There are no elements in classes L_1 , C_1 and Γ . $-P^T$ is obtained from the fundamental loop matrix B_t by striking out the first three rows corresponding to capacitive twigs. Thus

$$-P^T = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

The network used to obtain K is shown in Fig. 3.15.

Note that sources are not required in the capacitive links. Successive applications of Thevenin's theorem produces the sequence of networks shown in Figs. 3.16-3.18. The 11 independent multipliers generated are

$$\begin{aligned} \alpha_1 &= 2^{-1} + 2^{-2} & \alpha_2 &= 2^{-1} \\ \alpha_3 &= 2^{-2} & \alpha_4 &= 2^{-1} \\ \alpha_5 &= 2^{-1} & \alpha_6 &= 2^{-2} \end{aligned}$$

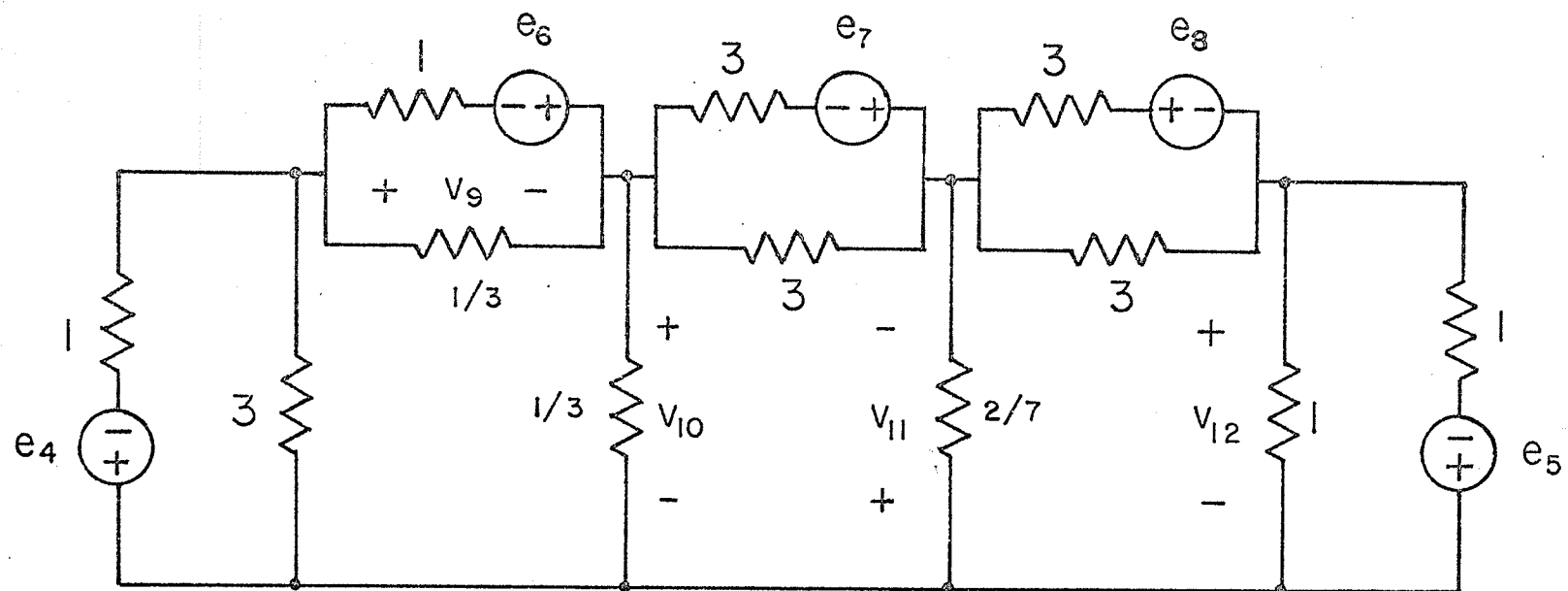


Fig. 3.15 Network used to obtain K for the filter of Fig. 3.13 (see also Figs. 3.16-3.18).

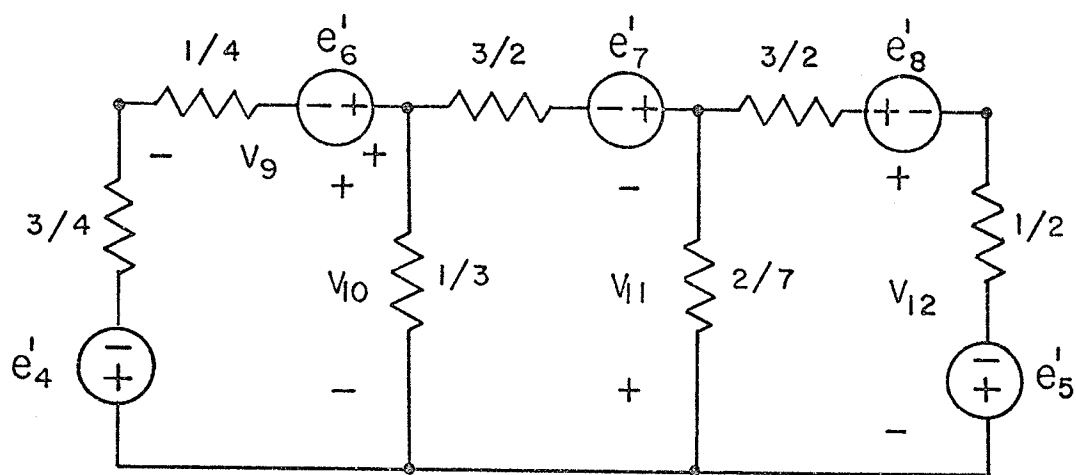


Fig. 3.16

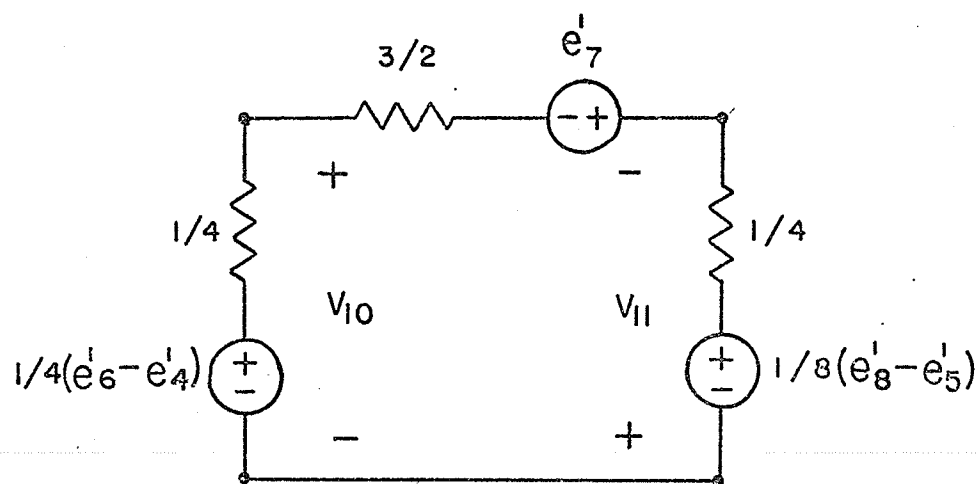


Fig. 3.17

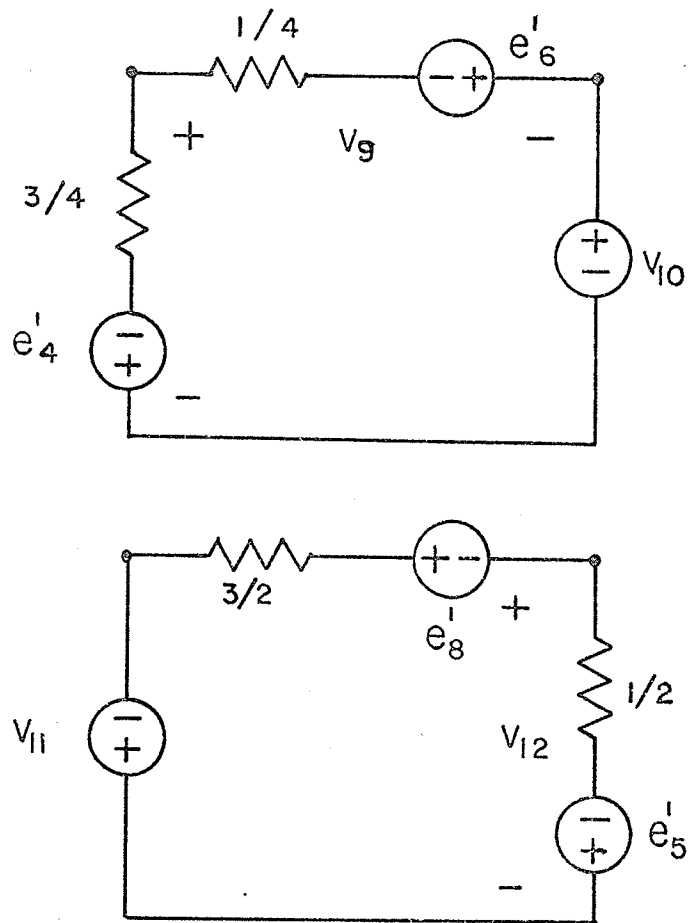


Fig. 3.18

$$\alpha_7 = 2^{-3}$$

$$\alpha_8 = 2^{-3}$$

$$\alpha_9 = 2^{-3}$$

$$\alpha_{10} = 2^{-2}$$

$$\alpha_{11} = 2^{-2}.$$

In the flow diagram for 2K, Fig. 3.19, the factor of 2 in 2K has been combined with α_i , $i=1,2,\dots,5$ to produce $\beta_i = 2\alpha_i$, $i=1,2,\dots,5$

$$\beta_1 = 1 + 2^{-1}$$

$$\beta_2 = 1$$

$$\beta_3 = 2^{-1}$$

$$\beta_4 = 1$$

$$\beta_5 = 1.$$

Thus we actually require only 8 multipliers for this particular case.

The frequency response obtained from a 1024 point FFT of the unit sample response is given in Fig. 3.20.

For any set of prototype element values, the n-port adaptor described above requires 32 adders and 11 multipliers. A noncanonic series-parallel adaptor realization requires from 39 to 42 adders and 11 multipliers, depending upon the technique used to realize the attenuation poles. For an arbitrary set of element values the particular structure used by Wegener requires 39 adders. However, for the element values of the present example, a special case arises and only 33 adders are required. A canonic realization using Fettweis' method [14] requires 3 adders external to the adaptor structure. However, as Fettweis points out, it may be possible to save 3 adders by modifying those adaptors for which all output signals are not necessary. In all of the above cases, more adders may be saved if those adaptor inputs corresponding to the reflected

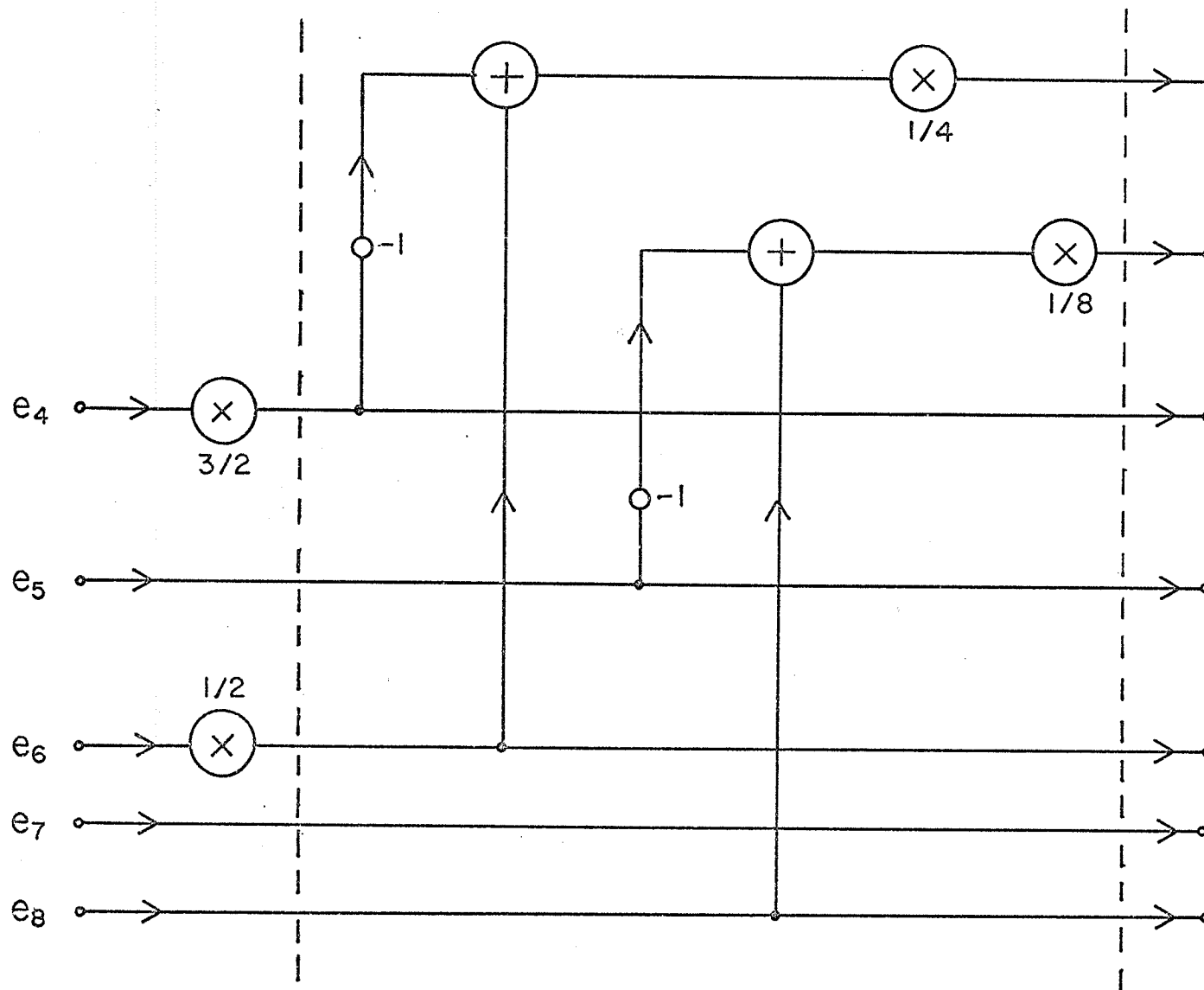


Fig. 3.19 Flow diagram for 2K for Example 3.2

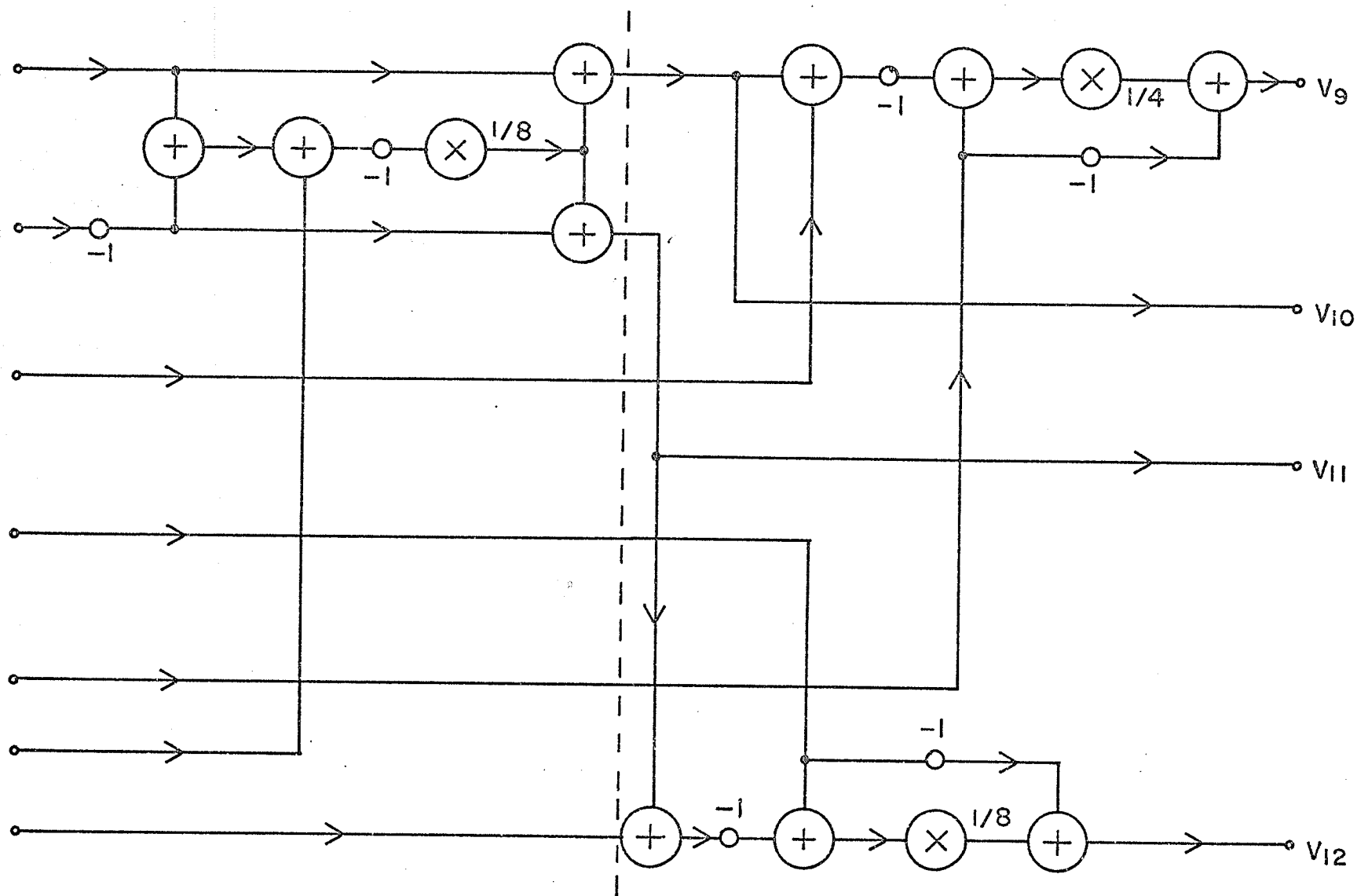


Fig. 3.19 (continued) Flow diagram for 2K

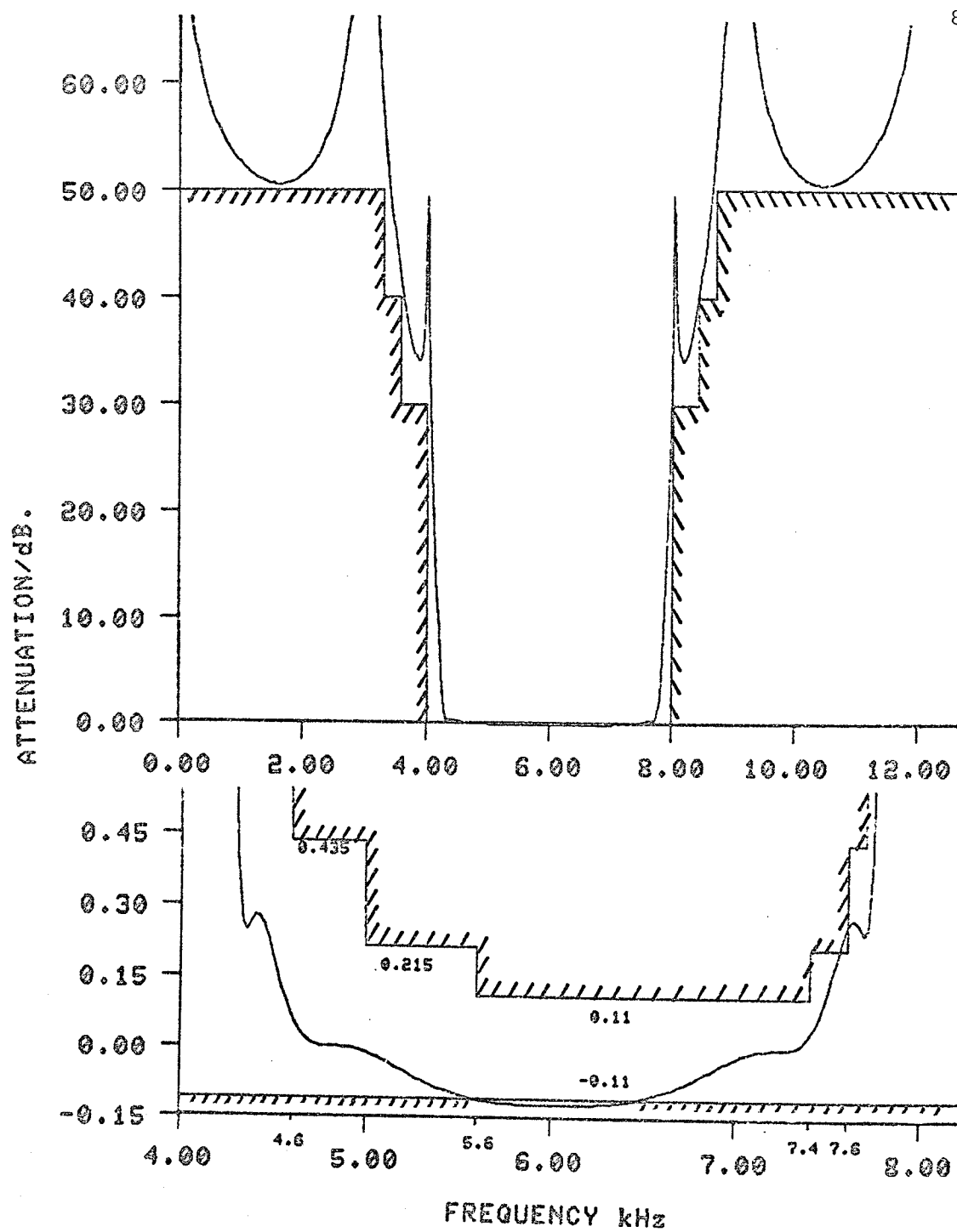


Fig. 3.20 Frequency response of the canonic wave digital realization of Fig. 3.13. Stopband and passband specifications are indicated by cross-hatched region.

wave at the source port and the incident wave at the load port are removed from the realization. For our n-port adaptor 3 adders can always be saved, bringing the total needed to only 29. In conclusion, the n-port adaptor design can save between 7 and 10 adders as compared with the series-parallel adaptor design. These savings are not restricted to the particular example considered here. In fact, savings will always be available if the prototype contains redundant reactive elements and the number of adders saved will increase with the number of redundancies.

The duplication of the topological matrix F in the n-port adaptor representation can also be utilized to advantage. If a hardware realization is of interest, then a single realization of F can be multiplexed between the input and the output. Thus it may be possible to save 8 adders at the expense of a digital multiplexer. In a software or firmware realization, the block of code representing F need not be duplicated. This reduces the amount of memory required to store the program.

The conductance matrices G_1 and G_2 are easily obtained for this example. G_R , G_G and G_X are all diagonal with entries equal to the corresponding element values in the prototype. G_C is nondiagonal due to the capacitance loops, but since there are no cutsets of capacitance, it can be readily shown that

$G_C = G_{C_2} + Q_{C_2S} G_S Q_{C_2S}^T$. This matrix can be interpreted as the node conductance matrix of the capacitive subnetwork obtained by open-circuiting all non-capacitive elements. Thus in this example we obtain

$$G_z = \begin{bmatrix} G_9 + G_1 & G_1 & 0 & 0 \\ G_1 & G_{10} + G_1 + G_2 & G_2 & 0 \\ 0 & G_2 & G_{11} + G_2 + G_3 & G_3 \\ 0 & 0 & G_3 & G_{12} + G_3 \end{bmatrix}. \quad (3.84)$$

Both of the examples presented in this section produced realizations which are canonic in both delays and multipliers. However, if the prototype is not a ladder structure then, although the realizations obtained will generally be canonic in delays (ie minimal), the network interpretation of K may not produce realizations which are canonic in multipliers. In this case there will not exist a one-to-one mapping from the independent resistance ratios in the prototype to the multipliers and thus low coefficient sensitivity and the existence of very short coefficients is not guaranteed. However, the following discussion outlines a design procedure whereby short word lengths can be obtained for the majority of the multipliers. First, it is necessary to partition the multipliers into two disjoint sets, an independent set which can be determined via a one-to-one mapping with the resistance ratios in the prototype and a dependent set which can be related to the independent multipliers by a set of dependency equations. Low element sensitivity in the prototype is transformed into low coefficient sensitivity in the independent multipliers and thus finite word length approximations

for these multipliers can be obtained with a relatively small number of bits. Now, using the dependency equations, the values for the dependent multipliers can be computed. Normally these multipliers will not be representable with a finite number of bits and therefore must be modified for realizability. If the errors introduced by these changes are kept as small as possible by using relatively long word lengths, then the response will not change drastically. Thus, short word lengths are obtainable for the independent multipliers while longer word lengths are required for the dependent multipliers.

CHAPTER IV

PROPERTIES OF WAVE DIGITAL FILTERS:

CONTROLLABILITY, OBSERVABILITY AND STABILITY

In this chapter we examine some important properties of wave digital filters realized using n-port adaptors. These properties not only aid in the characterization of wave digital systems but are also of practical importance. The n-port adaptors of Chapters II and III are shown to be both pseudolossless and reciprocal. Some results concerning the controllability, observability and zero-input stability of linear wave digital systems are presented. A general system modification scheme which guarantees freedom from parasitic oscillations in nonlinear wave digital filters using n-port adaptors is given.

4.1 PROPERTIES OF n-PORT WAVE DIGITAL ADAPTORS

Digital filter realizations are often represented by signal flow graphs. Such representations allow various properties of digital networks to be studied [1], [37], [12]. Wave digital filters or, more generally, wave digital networks belong to a subclass of signal flow networks called port-connected signal flow networks. Fettweis has discussed various properties of these networks in a series of publications [8], [25], [37], [38].

In this section we give a short review of those concepts and definitions required in the remainder of the thesis, followed by the introduction of some properties of wave digital n-port adaptors.

The structure of wave digital network theory follows similar lines to that of classical network theory. A set of elements consisting of wave digital n-ports can be interconnected according to a set of rules

to form large wave digital networks. Each n-port has an associated port weighting matrix which is usually diagonal and positive definite. However, due to the nature of the wave digital n-port adaptor derived in Chapter III, it is necessary to consider nondiagonal positive definite port weighting matrices. For wave digital n-port adaptors the weighting matrix is normally taken to be equal to the port reference conductance matrix

$$G = \begin{bmatrix} G_1 & 0 \\ 0 & G_2 \end{bmatrix} \quad (4.1)$$

where G_1 and G_2 are defined in (3.49).

The instantaneous pseudopower, $p(n)$, absorbed by a wave digital n-port at time n , with respect to the reference conductance matrix G , is given by

$$p(n) = a^T(n)Ga(n) - b^T(n)Gb(n) \quad (4.2)$$

where $a(n)$ and $b(n)$ are the incident and reflected wave vectors at time n . If the n-port is linear and instantaneous, in which case

$$b(n) = Sa(n) \quad (4.3)$$

where the scattering matrix S is constant, then

$$p(n) = a^T(n)(G - S^TGS)a(n). \quad (4.4)$$

The concepts of instantaneous pseudolosslessness and instantaneous pseudopassivity can be defined in terms of the absorbed pseudopower.

With respect to the reference conductance matrix G , a wave digital n-port is

$$(a) \text{ instantaneously pseudolossless if } p(n) = 0 \quad (4.5)$$

$$(b) \text{ instantaneously pseudopassive if } p(n) \geq 0 \quad (4.6)$$

for all n and for all admissible signals $a(n)$ and $b(n)$. For linear instantaneous n -ports we have the following simplified definitions: a linear instantaneous wave digital n -port is

$$(a) \quad \text{instantaneously pseudolossless if } G - S^T G S = 0 \quad (4.7)$$

$$(b) \quad \text{instantaneously pseudopassive if } G - S^T G S \geq 0 \quad (4.8)$$

where the matrix inequality refers to the corresponding quadratic form.

A reciprocity condition can also be developed in the complex frequency domain. A linear instantaneous wave digital n -port is reciprocal with respect to the symmetric reference conductance matrix G if

$$G S = S^T G. \quad (4.9)$$

It is now possible to prove the following:

Theorem 4.1:

For a linear instantaneous wave digital n -port, any two of the following imply the third

- 1) $S^T G S = G$ (pseudolosslessness)
- 2) $S^T G = G S$ (reciprocity)
- 3) $S S^T = U$

Proof:

- a) Given 1) and 2) we have

$$S^T G S = G \rightarrow G S S^T = G \rightarrow S S^T = U.$$

- b) Given 1) and 3) we have

$$S^T G S = G \rightarrow S^T G S S^T = G S \rightarrow S^T G = G S.$$

- c) Given 2) and 3) we have

$$S^T G = G S \rightarrow S^T G S = G S S^T \rightarrow S^T G S = G.$$

Consider the n -port adaptor described by (3.46)-(3.49).

We now show that this class of adaptors, which includes the standard

n-port adaptor of Chapter II as well as those of Fettweis, is both instantaneously pseudolossless and reciprocal.

Theorem 4.2:

An n-port adaptor described by a scattering matrix in the form of (3.46) - (3.49) is instantaneously pseudolossless and reciprocal with respect to its port conductance matrix G .

Proof:

a) Reciprocity

Consider

$$GS = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix}$$

where

$$X_{11} = 2G_1 P^T K - G_1 = 2G_1 P^T Y^{-1} P G_1 - G_1 = X_{11}^T.$$

$$X_{12}^T = [2G_1 P^T (U - K P^T)]^T = 2(U - P K^T) P G_1$$

$$= 2(U - P G_1 P^T Y^{-1}) P G_1 = 2(Y - P G_1 P^T) Y^{-1} P G_1$$

$$= 2G_2 K = X_{21}.$$

$$X_{22} = G_2 (U - 2K P^T) = G_2 (U - 2Y^{-1} P G_1 P^T)$$

$$= G_2 Y^{-1} (Y - 2P G_1 P^T) = G_2 Y^{-1} (-Y + 2G_2)$$

$$= G_2 (-U + 2Y^{-1} G_2) = -(G_2 - 2G_2 Y^{-1} G_2) = X_{22}^T.$$

Hence,

$$GS = (GS)^T = S^T G.$$

b) Pseudolosslessness

Since (3.46) can be written as the product of three self-inverse matrices

$$S = \begin{bmatrix} -U & P^T \\ 0 & U \end{bmatrix} \begin{bmatrix} -U & 0 \\ -2K & U \end{bmatrix} \begin{bmatrix} -U & P^T \\ 0 & U \end{bmatrix}$$

then $SS = U$ and, from Theorem 4.1, $S^T GS = G$. This result is of tremendous importance in the stability studies to follow.

In the design procedure illustrated in Example 3.1 the independent multipliers, and hence K , are modified for realizability. This produces a modified adaptor representation. Since there exists a set of element values for the prototype which corresponds to this new set of multipliers, the modified adaptor is still pseudolossless and reciprocal. However, if a realization of K was not obtained in terms of independent multipliers then, where K is modified for realizability, it may no longer satisfy (3.48). Equivalently, a corresponding set of element values for the prototype does not exist. The following result proves that such an adaptor can no longer be pseudolossless and reciprocal.

Theorem 4.3:

The n -port adaptor described by (3.46) is instantaneously pseudolossless and reciprocal with respect to its port conductance matrix G if and only if (3.48) is satisfied; that is, if and only if $K = Y^{-1}PG_1$ where $Y = G_2 + PG_1P^T$.

Proof:

a) Sufficiency is established by Theorem 4.2.

b) Necessity: since $SS = U$ independent of K , then if $K = Y^{-1}PG_1$ is necessary for $S^TG = GS$, it is also necessary for $S^TGS = G$.

Upon equating the partitioned forms of GS and S^TG , four matrix equations, two of which are identical, are produced. The three independent equations are

$$G_1P^TK = K^TPG_1 \quad (4.10)$$

$$G_1P^T - G_1P^TKP^T = K^TG_2 \quad (4.11)$$

$$G_2KP^T = PK^TG_2 \quad (4.12)$$

Substitution of (4.10) into (4.11) produces

$$G_1P^T - K^TPG_1P^T = K^TG_2$$

which requires that

$$K = (PG_1P^T + G_2)^{-1}PG_1 = Y^{-1}PG_1. \quad (4.13)$$

Equation (4.13), which is the desired condition, can be substituted into (4.10) and (4.12) to check for consistency. Thus, from (4.10),

$$G_1P^TY^{-1}PG_1 = G_1P^TY^{-1}PG_1$$

and, from (4.12),

$$G_2Y^{-1}PG_1P^T = PG_1P^TY^{-1}G_2$$

$$G_2Y^{-1}(Y - G_2) = (Y - G_2)Y^{-1}G_2$$

$$G_2 - G_2Y^{-1}G_2 = G_2 - G_2Y^{-1}G_2.$$

4.2 CONTROLLABILITY AND OBSERVABILITY OF LINEAR WAVE DIGITAL FILTERS

The n-port wave digital filter formulation immediately yields a system description in state equation form. In general, we have

$$\begin{bmatrix} \Sigma x(n+1) \\ y(n) \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} x(n) \\ u(n) \end{bmatrix} \quad (4.14)$$

where x , u and y are vectors of the state, input and output variables respectively, Σ is the diagonal matrix containing the inverters required by the inductive ports, and S (given here in partitioned form) is the scattering matrix representing the n-port adaptor. Since $\Sigma\Sigma = U$, (4.14) becomes

$$\begin{bmatrix} x(n+1) \\ y(n) \end{bmatrix} = \begin{bmatrix} \Sigma S_{11} & \Sigma S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} x(n) \\ u(n) \end{bmatrix}. \quad (4.15)$$

The port reference conductance matrix can always be partitioned so that

$$G = \begin{bmatrix} G_{11} & 0 \\ 0 & G_{22} \end{bmatrix} \quad (4.16)$$

and

$$\Sigma G_{11} = G_{11} \Sigma \quad (4.17)$$

where G_{11} is that component of G associated with the delay-terminated ports of the n-port adaptor and G_{22} is that part associated with the input-output ports. Qualitative properties of a linear discrete-time system expressed in this form can be readily established. The concepts of state controllability and observability and their relationship to

minimal realizations are well documented, as are various procedures which can test a system for these properties [34]. The most commonly used tests, expressed in the notation of (4.15), are as follows:

The m -dimensional linear shift-invariant system described by (4.15) is controllable if and only if the controllability matrix Q has

rank m

$$Q = \begin{bmatrix} \Sigma S_{12}, \Sigma S_{11} \Sigma S_{12}, \dots, (\Sigma S_{11})^{m-1} \Sigma S_{12} \end{bmatrix} \quad (4.18)$$

and is observable if and only if the observability matrix P has

rank m

$$P = \begin{bmatrix} S_{21}^T, (\Sigma S_{11})^T S_{21}^T, \dots, (\Sigma S_{11})^{T(m-1)} S_{21}^T \end{bmatrix}. \quad (4.19)$$

The following theorem and corollary investigate the controllability and observability of a wave digital system built around a pseudolossless reciprocal n -port adaptor. Such a system will be called a pseudolossless reciprocal wave digital system.

Theorem 4.4:

A reciprocal wave digital system is controllable if and only if it is observable.

Proof:

The state description of the system (4.15), together with the reciprocity condition $S^T G = G S$ or $S = R S^T G$ in partitioned form,

yields

$$\begin{bmatrix} x(n+1) \\ y(n) \end{bmatrix} = \begin{bmatrix} \Sigma & 0 \\ 0 & U \end{bmatrix} \begin{bmatrix} R_{11} & 0 \\ 0 & R_{22} \end{bmatrix} \begin{bmatrix} S_{11}^T & S_{21}^T \\ S_{12}^T & S_{22}^T \end{bmatrix} \begin{bmatrix} G_{11} & 0 \\ 0 & G_{22} \end{bmatrix} \begin{bmatrix} x(n) \\ u(n) \end{bmatrix}.$$

Since the controllability and observability are invariant

under a similarity transformation [34], let

$$\tilde{x}(n) = \Sigma G_{11} x(n)$$

and then, using (4.17),

$$\tilde{x}(n) = G_{11} \Sigma x(n).$$

Implementation of this transformation produces

$$\begin{aligned} \begin{bmatrix} \tilde{x}(n+1) \\ y(n) \end{bmatrix} &= \begin{bmatrix} U & 0 \\ 0 & R_{22} \end{bmatrix} \begin{bmatrix} S_{11}^T & S_{21}^T \\ S_{12}^T & S_{22}^T \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & U \end{bmatrix} \begin{bmatrix} U & 0 \\ 0 & G_{22} \end{bmatrix} \begin{bmatrix} \tilde{x}(n) \\ u(n) \end{bmatrix} \\ &= \begin{bmatrix} S_{11}^T \Sigma & S_{21}^T G_{22} \\ R_{22} S_{12}^T \Sigma & R_{22} S_{22}^T G_{22} \end{bmatrix} \begin{bmatrix} \tilde{x}(n) \\ u(n) \end{bmatrix} \end{aligned} \quad (4.20)$$

and thus system (4.15) is controllable (observable) if and only if system (4.20) is controllable (observable). The controllability matrix for system (4.15), Q_1

$$Q_1 = \begin{bmatrix} \Sigma S_{12}, \Sigma S_{11} \Sigma S_{12}, \dots, (\Sigma S_{11})^{n-1} \Sigma S_{12} \end{bmatrix}$$

and the observability matrix for system (4.20), P_2

$$P_2 = \begin{bmatrix} \Sigma S_{12} R_{22}, \Sigma S_{11} \Sigma S_{12} R_{22}, \dots, (\Sigma S_{11})^{n-1} \Sigma S_{12} R_{22} \end{bmatrix}$$

are related by

$$P_2 = Q_1 \text{diag}[R_{22}]$$

where $\text{diag}[R_{22}]$ is a block diagonal matrix with all diagonal elements equal to R_{22} and all off-diagonal elements equal to zero. Since R_{22} is nonsingular, the rank of Q_1 is equal to the rank of P_2 and therefore system (4.15) is controllable if and only if system (4.20) is observable.

But (4.20) is observable if and only if (4.15) is observable and therefore (4.15) is controllable if and only if it is also observable.

Since a system is said to be minimal if it is both controllable and observable, we have the following self-evident corollary to Theorem 4.4.

Corollary 4.4 :

A reciprocal wave digital system is controllable (observable) if and only if it is minimal.

4.3 STABILITY OF WAVE DIGITAL FILTERS

In this section we investigate the zero-input stability of both linear and nonlinear wave digital filters. From the variety of stability criteria which are applicable to discrete-time linear shift-invariant systems, stability in the sense of Lyapunov (i.s.L.) via Lyapunov's direct method is chosen since this method can also be applied to nonlinear systems. The direct method is based upon the existence of a positive definite energy function whose first forward difference is negative semidefinite for systems which are stable i.s.L. and is negative definite for systems which are asymptotically stable i.s.L. Asymptotic stability i.s.L. is also obtained in a system which is stable i.s.L. and for which the first forward difference of the Lyapunov function does not vanish identically on the state trajectory. A complete discussion of this theory is available in Kalman and Bertram [38].

We first consider linear wave digital systems described by (4.15). The Lyapunov function used in the following development is chosen to be $\mathbf{x}^T \mathbf{G}_{11} \mathbf{x}$.

Theorem 4.5 :

A linear pseudolossless wave digital system is stable i.s.L.

Proof:

Consider the positive definite function

$$V(x) = x^T G_{11} x.$$

The first forward difference is given by

$$\begin{aligned} \Delta V(x) &= V(x(n+1)) - V(x(n)) \\ &= x^T(n+1) G_{11} x(n+1) - x^T(n) G_{11} x(n) \\ &= x^T(n) S_{11}^T \Sigma G_{11} \Sigma S_{11} x(n) - x^T(n) G_{11} x(n) \\ &= x^T(n) (S_{11}^T G_{11} S_{11} - G_{11}) x(n). \end{aligned}$$

The pseudolossless property yields

$$S_{11}^T G_{11} S_{11} - G_{11} = -S_{21}^T G_{22} S_{21}$$

and thus

$$\Delta V(x) = -x^T(n) S_{21}^T G_{22} S_{21} x(n).$$

Since S_{21} is, in general, not square, $\Delta V(x)$ is negative semidefinite.

Hence

$$\Delta V(x) \leq 0$$

which is sufficient for stability i.s.L. [38].

This result can be applied immediately to demonstrate that filters obtained from lossless prototypes by either of the n-port adaptor techniques are always stable i.s.L. The pseudolosslessness of such filters is in fact a direct consequence of the lossless nature of the prototypes.

Theorem 4.5 can easily be extended to cover pseudopassive ($S^T G S - G \leq 0$) and strictly pseudopassive ($S^T G S - G < 0$) systems.

Corollary 4.5:

A linear pseudopassive (strictly pseudopassive) wave digital system is stable (asymptotically stable) i.s.L.

The results of Theorem 4.5 can be combined with the observability of the system to investigate asymptotic stability.

Theorem 4.6:

A pseudolossless linear wave digital system is asymptotically stable i.s.L. if and only if it is observable.

Proof:

First the sufficiency. Theorem 4.5 has established stability i.s.L. by showing that

$$\Delta V(x) = -x^T(n) S_{21}^T G_{22} S_{21} x(n) \leq 0, \text{ for all } n.$$

If the system is observable, the observability matrix has rank m and thus

$$\begin{bmatrix} S_{21} \\ S_{21} \Sigma S_{11} \\ \vdots \\ S_{21} (\Sigma S_{11})^{m-1} \end{bmatrix} x \neq 0, \quad \text{for all } x \neq 0. \quad (4.21)$$

Choose $x(0) \neq 0$ as an initial state in the system. Then

$$x(n) = (\Sigma S_{11})^n x(0), \quad n \geq 0$$

defines the succession of states. The observability condition (4.21)

then requires that

$$S_{21} x(p) \neq 0, \quad p = 0, 1, 2, \dots, m-1$$

and hence

$$x^T(p) S_{21}^T G_{22} S_{21} x(p) \neq 0, \quad p = 0, 1, 2, \dots, m-1.$$

Since $x(p)$ is on the state trajectory,

$$\Delta V(x) \not\equiv 0$$

on the trajectory and the system is asymptotically stable i.s.L. [38].

The necessity is most easily established by proving the contrapositive. Assume that the system is not observable, in which case there exists an $x \neq 0$ such that

$$\begin{bmatrix} S_{21} \\ S_{21} \Sigma S_{11} \\ \vdots \\ S_{21} (\Sigma S_{11})^{m-1} \end{bmatrix} x = 0, \quad x \neq 0. \quad (4.22)$$

Let $x(0)$ be a nonzero vector satisfying (4.22). Then, since G_{11} is positive definite

$$x^T(0) G_{11} x(0) \neq 0 \quad (4.23)$$

and, from (4.22),

$$S_{21} x(n) = 0, \quad n = 0, 1, 2, \dots, m-1.$$

This last condition can be extended, using the Cayley-Hamilton Theorem, to all succeeding states

$$S_{21} x(n) = 0, \quad n \geq 0.$$

We then have

$$x^T(n) S_{21}^T G_{22} S_{21} x(n) = 0, \quad n \geq 0$$

which, when combined with the pseudolossless property, produces

$$x^T(n) S_{11}^T G_{11} S_{11} x(n) = x^T(n) G_{11} x(n), \quad n \geq 0$$

or, equivalently,

$$x^T(n+1) G_{11} x(n+1) = x^T(n) G_{11} x(n), \quad n \geq 0. \quad (4.24)$$

Equation (4.23) together with (4.24) yields

$$x^T(n) G_{11} x(n) = x^T(0) G_{11} x(0) \neq 0, \quad n \geq 0$$

which proves that the state does not converge to the equilibrium state, $x_e = 0$, with increasing n . Thus we conclude that if the system is not observable it is not asymptotically stable.

Since most wave digital systems are both pseudolossless and reciprocal, the following corollary is of interest.

Corollary 4.6:

A pseudolossless reciprocal wave digital system is asymptotically stable i.s.L. if and only if it is minimal.

Proof:

From Theorem 4.6, a pseudolossless system is asymptotically stable i.s.L. if and only if it is observable. If the system is also reciprocal, Corollary 4.4 shows that it is observable if and only if it is minimal. The desired result follows immediately.

The filters, which can be obtained by use of the technique of Chapter III, in addition to being pseudolossless and reciprocal, are in most instances minimal. Corollary 4.6 proves that such filters are asymptotically stable.

So far in this section we have investigated the stability of linear wave digital systems. However, due to the finite word length requirements for realizability, any implementation is in reality nonlinear. The stability of such systems can still be investigated by the direct method of Lyapunov. In fact, the results obtained in Theorem 4.5 can be extended to include nonlinear wave digital filters. If the nonlinearities, which can be specified by the filter designer, are such that the nonlinear filter remains pseudolossless, then the filter is still stable i.s.L. The proof of this statement follows directly from the proof of Theorem 4.5 by noting that for a pseudolossless system under zero input the instantaneous pseudopower is given by

$$p(n) = x^T(n+1)G_{11}x(n+1) - x^T(n)G_{11}x(n) + y^T(n)G_{22}y(n) = 0. \quad (4.25)$$

Then, since G_{22} is positive definite,

$$\Delta V(x) = x^T(n+1)G_{11}x(n+1) - x^T(n)G_{11}x(n) \leq 0. \quad (4.26)$$

A similar result is obtained if the nonlinear system is pseudopassive.

A more important practical result is obtained if the nonlinear system is strictly pseudopassive since then it will be asymptotically stable i.s.L. and the state, and subsequently the output, will eventually become permanently zero. In this case limit cycles will not exist.

Because wave digital systems are in fact finite state machines, a more appropriate definition of stability is available [39]. A finite state system is stable under zero-input if the state becomes permanently zero in a finite time. The stability of the output follows immediately by noting that when the states are permanently zero, the outputs, being weighted sums of the states, must also be permanently zero. Fettweis and Meerkötter [25] have shown that pseudopassivity of the nonlinear filter is sufficient for output stability and, if the linear filter has no oscillations, then stability is also guaranteed. In what follows we present some results which demonstrate sufficient conditions for stability, and therefore also for output stability, of wave digital filters using n-port adaptors.

Theorem 4.7:

A nonlinear finite state wave digital system (NL) is stable if it is derived from a linear wave digital system (L) which is asymptotically stable i.s.L. such that, when $x_L(n) = x_{NL}(n)$, the next states $x_L(n+1)$ and $x_{NL}(n+1)$ satisfy either

$$a) \quad x_{NL}(n+1) = x_L(n+1) \text{ and thus } V(x_{NL}(n+1)) = V(x_L(n+1))$$

$$\text{or } b) \quad V(x_{NL}(n+1)) < V(x_L(n+1)) \text{ if } x_{NL}(n+1) \neq x_L(n+1)$$

where $V(x) = x^T G_{11} x$ is the Lyapunov function for the linear system.

Proof:

Assume that a nonlinear system (NL) is obtained from an asymptotically stable linear system (L) according to a) and b). Because system (NL) is a finite state machine, either $x_{NL}(n)$ becomes zero after a finite time, in which case the system is stable, or a cycle begins. Assume that such a cycle exists starting at time n_0 with period N . Then

$$x_{NL}(n_0 + pN) = x_{NL}(n_0) \quad (4.27)$$

and

$$V(x_{NL}(n_0 + pN)) = V(x_{NL}(n_0)) \quad (4.28)$$

where n_0 , p and N are non-negative integers. Let

$$x_L(n_0) = x_{NL}(n_0).$$

Then, due to the asymptotic stability of (L), we have

$$V(x_{NL}(n_0)) = V(x_L(n_0)) \geq V(x_L(n_0 + 1)) \quad (4.29)$$

while, from the design conditions a) and b), we must have

$$V(x_L(n_0 + 1)) \geq V(x_{NL}(n_0 + 1)) \quad (4.30)$$

Equations (4.29) and (4.30) together require that

$$V(x_{NL}(n_0)) \geq V(x_{NL}(n_0 + 1)).$$

If we now set the next state in (L) equal to $x_{NL}(n_0 + 1)$, the same procedure produces

$$V(x_{NL}(n_0 + 1)) \geq V(x_{NL}(n_0 + 2)).$$

Repeated application yields

$$V(x_{NL}(n_0)) \geq V(x_{NL}(n_0 + 1)) \geq \dots \geq V(x_{NL}(n_0 + N)) \geq \dots$$

which, when combined with the cycle condition (4.28), requires that

$$V(x_{NL}(n)) = V(x_{NL}(n_0)), \text{ for all } n \geq n_0.$$

Due to the nonlinear system design criteria a) and b), the above energy condition requires that

$$x_L(n) = x_{NL}(n), \text{ for all } n \geq n_0.$$

This means that the same sequence of states occurs in both the linear and nonlinear systems and thus a cycle must also exist in the linear system. This contradicts the asymptotic stability assumption and thus the nonlinear system cannot support a cycle and must therefore be stable.

If the linear system is not asymptotically stable, but is stable, then a similar result having only slightly modified design conditions can be obtained.

Corollary 4.7:

A nonlinear finite state wave digital system (NL) is stable if it is derived from a linear wave digital system (L) which is stable i.s.L. such that, when $x_L(n) = x_{NL}(n)$, the next states $x_L(n+1)$ and $x_{NL}(n+1)$ satisfy $V(x_{NL}(n+1)) < V(x_L(n+1))$ where $V(x) = x^T G_{11} x$ is the Lyapunov function for the linear system.

Proof:

Using the same arguments as in the previous theorem, it follows that, if a cycle exists starting at n_0 , then

$$V(x_{NL}(n)) = V(x_{NL}(n_0)), \text{ for all } n \geq n_0. \quad (4.31)$$

However, the stability of (L), which requires

$$V(x_{NL}(n_0)) = V(x_L(n_0)) \geq V(x_L(n_0+1))$$

together with the design requirement of (NL), demands that

$$V(x_{NL}(n_0+1)) < V(x_{NL}(n_0)).$$

This contradicts (4.31) and hence the nonlinear system must therefore be stable.

The system modification required by either Theorem 4.7 or Corollary 4.7 can be interpreted as a requirement on the reflected waves at those ports of the n -port adaptor which are connected to delays. For the situation where G_{11} is a diagonal positive definite matrix, we have

$$V(x) = x^T G_{11} x = \sum_{i=1}^m g_i x_i^2 \quad (4.32)$$

where $g_i > 0$, $i=1, \dots, m$. Thus, given $x_{Li}(n) = x_{NLi}(n)$, the condition

$$\left| x_{NLi}(n+1) \right| \leq \left| x_{Li}(n+1) \right| \quad \text{for all } n, i \quad (4.33)$$

is sufficient for stability when the linear system is asymptotically stable i.s.L. Such a linear system could either be a strictly pseudopassive system or a minimal pseudolossless reciprocal system. Similarly, the condition

$$\left| x_{NLi}(n+1) \right| < \left| x_{Li}(n+1) \right| \quad \text{for all } n, i \quad (4.34)$$

is sufficient for stability when the linear system is stable, i.s.L.

Recall that pseudolossless, pseudopassive and strictly pseudopassive systems are all stable i.s.L. A method of implementing (4.33) and (4.34) is discussed in a later chapter.

The conditions (4.33) and (4.34) are similar to those obtained by Fettweis and Meerkötter [25]. The differences are that we need only consider the outputs of the adaptor connected to delays and that stability is guaranteed even for pseudolossless linear filters which contain unobservable modes.

When G_{11} is not diagonal, (4.33) or (4.34) may no longer be sufficient for stability. This situation, which occurs naturally as a result of the design technique of Chapter III, is studied in the next chapter.

CHAPTER V

DIAGONAL LYAPUNOV FUNCTIONS FOR MINIMAL WAVE DIGITAL FILTERS

The criteria developed in the previous chapter which guarantee the absence of parasitic oscillations in nonlinear wave digital filters are most easily implemented for systems with a diagonal reference conductance matrix. The n-port adaptor design outlined in Chapter II automatically produces systems with the appropriate form for G_{11} . However, the minimal realizations obtained in Chapter III have a nondiagonal G_{11} .

The first part of this chapter investigates the eigenvalues and eigenvectors of S_{11} . This information is then utilized in the search for alternate Lyapunov functions which have a diagonal form. The chapter concludes with the presentation of a technique which uses a similarity transformation of the state variables which simultaneously maintains the input-output transfer function and diagonalizes G_{11} to produce a diagonal Lyapunov function.

5.1 EIGENVALUES AND EIGENVECTORS OF S_{11}

In this section we investigate the eigenvalues and eigenvectors of S_{11} . First we present the following lemma.

Lemma 5.1:

Given a real $n \times n$ matrix S , there exists an $n \times n$ symmetric positive definite matrix G satisfying $S^T G = G S$ if and only if S is of simple structure with real eigenvalues.

Proof:

An $n \times n$ matrix is said to be of simple structure if and only if it has a set of n linearly independent eigenvectors [40] or, equivalently, if the modal matrix is nonsingular.

Suppose that $S^T G = GS$, where G is a symmetric positive definite matrix. A congruence transformation $P^T GP$ of G exists for which

$$P^T GP = U \quad (5.1)$$

and thus

$$P^{-1} = P^T G.$$

Consider the similarity transformation $P^{-1}SP$ of S .

$$P^{-1}SP = P^T GSP = P^T S^T GP \quad (5.2)$$

is symmetric and also has the same eigenvalues as S . Further, there exists an orthogonal transformation of (5.2) such that

$$Q^T (P^T GSP) Q = \Lambda_S \quad (5.3)$$

and

$$Q^T Q = U \quad (5.4)$$

where Λ_S is the diagonal matrix of eigenvalues of S .

If we now define

$$W = PQ$$

then, from (5.3),

$$W^T GSW = \Lambda_S \quad (5.5)$$

and, from (5.1) and (5.4),

$$W^T GW = Q^T P^T GPQ = Q^T Q = U. \quad (5.6)$$

Equation (5.6) yields

$$W^{-1} = W^T G$$

which together with (5.5) produces

$$W^T GSW = W^{-1}SW = \Lambda_S$$

and therefore

$$SW = W\Lambda_S. \quad (5.7)$$

Equation (5.7) demonstrates that the nonsingular matrix W is a modal matrix of S and hence S is of simple structure. Furthermore, since W and S are matrices over the field of real numbers, the entries of Λ_S and hence the eigenvalues of S are real.

Suppose now that S is of simple structure with real eigenvalues. Then the modal matrix W is nonsingular and

$$W^{-1}SW = \Lambda_S. \quad (5.8)$$

Since Λ_S is diagonal, and hence symmetric, we have

$$W^{-1}SW = (W^{-1}SW)^T = W^T S^T (W^T)^{-1}$$

from which we obtain

$$(W^T)^{-1}W^{-1}S = S^T (W^T)^{-1}W^{-1}.$$

If we now define

$$G = (WW^T)^{-1} = (W^T)^{-1}W^{-1}$$

then

$$G = G^T, \quad W^T G W = U, \quad GS = S^T G.$$

Thus G is symmetric positive definite and satisfies $S^T G = GS$.

Use of this lemma shows that for reciprocal wave digital systems both S and S_{11} have real eigenvalues and are of simple structure. Martens and Meerkötter [29] have shown that all of the eigenvalues of S are either $+1$ or -1 . The eigenvalues of S_{11} are now investigated.

Theorem 5.1:

Given a pseudolossless reciprocal system, at most two eigenvalues of S_{11} are not equal to $+1$ or -1 .

Proof:

For a pseudolossless reciprocal system we have, from Theorem 4.1,

$$SS = U$$

which, upon partitioning, yields

$$S_{11}S_{11} + S_{12}S_{21} = U$$

or

$$(U - S_{11}^2) = S_{12}S_{21}.$$

Since, in general, S_{12} and S_{21} have dimensions of $n \times 2$ and $2 \times n$ respectively, there exist at least $n - 2$ independent nonzero vectors x_i in the null space of $S_{12}S_{21}$, n being the dimension of S_{11} . That is

$$(U - S_{11}^2) x_i = S_{12}S_{21}x_i = 0 \quad x_i \neq 0, i = 1, 2, \dots, m \quad m \geq n-2.$$

Therefore

$$S_{11}^2 x_i = x_i, \quad x_i \neq 0, i = 1, 2, \dots, m, \quad m \geq n-2$$

which implies that S_{11}^2 has at least $n-2$ unit eigenvalues and hence S_{11} has at least $n-2$ eigenvalues equal to either $+1$ or -1 .

A physical interpretation of the $+1$ and -1 modes of S_{11} is easily obtained. If all of the inductive elements in the prototype are replaced by capacitive elements, an RC network which necessarily has negative real eigenvalues is produced. Such a network may have modes at $\psi = 0$ and $\psi = \infty$ due to the capacitance-only cutsets and loops respectively. These modes are equivalent to modes at $z = 1$ and $z = -1$ in a discrete-time realization produced by the bilinear z -transformation. Further, since $\Sigma = U$, these modes are identical with the eigenvalues of S_{11} describing the original prototype. Some of these eigenvalues, however,

are suppressed during the formation of the reduced S_{11} . The other two modes which must be inside the unit circle (S_{11} is a stable matrix) are a result of the resistive terminations.

During the formation of the reduced order realization, all capacitive and inductive cutsets and loops were effectively removed. The remaining reactive cutsets, which account for the $\lambda = 1$, can be identified by first short-circuiting all capacitances and inductances which define cutsets (classes C_1 and Γ), and then open-circuiting all capacitances and inductances which define loops (classes S and L_1). The remaining network is described by KCL in the form

$$\begin{bmatrix} Q_{C_2R} & Q_{C_2L_2} & | & U & 0 \\ Q_{GR} & Q_{GL_2} & | & 0 & U \end{bmatrix} \begin{bmatrix} i_R \\ i_{L_2} \\ \hline i_{C_2} \\ i_G \end{bmatrix} = 0. \quad (5.9)$$

Any reactance-only cutsets must occur in the first equation but are not shown explicitly. Following similar arguments used in Chapter III, we can partition the elements in class C_2 and then rewrite (5.9) as

$$\begin{bmatrix} 0 & Q_A & | & U & Q_B & 0 \\ Q_{C_2R}^{(2)} & Q_{C_2L_2}^{(2)} & | & 0 & U & 0 \\ Q_{GR} & Q_{GL_2} & | & 0 & 0 & U \end{bmatrix} \begin{bmatrix} i_R \\ i_L \\ \hline i_{C_2}^{(1)} \\ i_{C_2}^{(2)} \\ i_G \end{bmatrix} = 0. \quad (5.10)$$

Elements in the class $C_2^{(1)}$ define the reactance cutsets.

Equation (5.10) can be considered to be the result of a

reordering of the equations and the variables C_2 in (5.9), followed by a sequence of elementary row operations, M_1 , which reduces Q_{C_2R}

$$M_1 Q_{C_2R} = M_1 \begin{bmatrix} Q_{C_2R}^{(1)} \\ Q_{C_2R}^{(2)} \end{bmatrix} = \begin{bmatrix} 0 \\ Q_{C_2R}^{(2)} \end{bmatrix} . \quad (5.11)$$

If the maximum number of reactance cutsets are displayed explicitly in (5.10), then the rows of $Q_{C_2R}^{(2)}$ are independent and thus

$$\text{rank } Q_{C_2R} = \text{rank } Q_{C_2R}^T = \text{number of elements in } C_2^{(2)} . \quad (5.12)$$

However, since

$$\begin{aligned} \text{rank } Q_{C_2R}^T + \text{nullity of } Q_{C_2R}^T &= \text{number of elements in } C_2^{(1)} \\ &\quad + \text{number of elements in } C_2^{(2)} \end{aligned} \quad (5.13)$$

equations (5.12) and (5.13) together yield

$$\text{number of elements in } C_2^{(1)} = \text{nullity of } Q_{C_2R}^T \quad (5.14)$$

and therefore

$$\text{number of } \lambda = 1 \text{ in } S_{11} = \text{nullity of } Q_{C_2R}^T . \quad (5.15)$$

As a direct consequence of this condition, there exists a matrix X having m_1 independent columns such that

$$Q_{C_2R}^T X = 0 \quad (5.16)$$

where m_1 is the number of $\lambda = 1$ in S_{11} .

The eigenvalues $\lambda = -1$ may be treated in a similar manner. After the elements in classes C_1 and I are short-circuited and the elements in classes S and L_1 are open-circuited, the remaining reactive loops account for the $\lambda = -1$. The resulting network is described by KVL in the form

$$\left[\begin{array}{cc|cc} U & 0 & -Q_{C_2R}^T & -Q_{GR}^T \\ 0 & U & -Q_{C_2L_2}^T & -Q_{GL_2}^T \end{array} \right] \begin{bmatrix} v_R \\ v_{L_2} \\ \hline v_{C_2} \\ v_G \end{bmatrix} = 0. \quad (5.17)$$

The reactance-only loops which must occur in the second equation can be displayed explicitly in a manner similar to the previous situation.

Thus, M_2 reduces $Q_{GL_2}^T$

$$M_2 Q_{GL_2}^T = \begin{bmatrix} 0 \\ Q_{GL_2}^{(2)T} \end{bmatrix}$$

and (5.17) becomes

$$\left[\begin{array}{ccc|cc} U & 0 & 0 & -Q_{C_2R}^T & -Q_{GR}^T \\ 0 & U & B_A & B_B & 0 \\ 0 & 0 & U & -Q_{C_2L_2}^{(2)T} & -Q_{GL_2}^{(2)T} \end{array} \right] \begin{bmatrix} v_R \\ v_{L_2}^{(1)} \\ v_{L_2}^{(2)} \\ \hline v_{C_2} \\ v_G \end{bmatrix} = 0 \quad (5.18)$$

where the L_2 variables have been reordered and partitioned appropriately. Elements in class $L_2^{(1)}$ define the reactance loops.

If the maximum number of reactance loops are displayed explicitly in (5.18), then it is easily shown that

$$\text{number of elements in } L_2^{(1)} = \text{nullity of } Q_{GL_2} \quad (5.19)$$

and hence

$$\text{number of } \lambda = -1 \text{ in } S_{11} = \text{nullity of } Q_{GL_2}. \quad (5.20)$$

Furthermore, there exists a matrix Y having m_2 independent columns such that

$$Q_{GL_2}^T Y = 0 \quad (5.21)$$

where m_2 is the number of $\lambda = -1$ in S_{11} .

Having obtained explicit information regarding the +1 and -1 eigenvalues of S_{11} , the matrices X and Y can be used to derive symbolic representations for the corresponding eigenvectors. S_{11} can be obtained directly from equations (3.46) and (2.22) in the form

$$S_{11} = \begin{bmatrix} -U & Q_{C_2L_2}^T \\ 0 & U \end{bmatrix} \begin{bmatrix} -U + 2Q_{GL_2}^T k_{22} & 2Q_{GL_2}^T k_{21} Q_{C_2R}^T \\ -2k_{12} & U - 2k_{11} Q_{C_2R}^T \end{bmatrix} \begin{bmatrix} -U & Q_{C_2L_2}^T \\ 0 & U \end{bmatrix} \quad (5.22)$$

where K has been partitioned conformally

$$K = \begin{bmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{bmatrix}. \quad (5.23)$$

Consider the matrix

$$W_1 = \begin{bmatrix} Q_{C_2L_2}^T \\ U \end{bmatrix} X. \quad (5.24)$$

Then, from (5.22),

$$S_{11} W_1 = \begin{bmatrix} -U & Q_{C_2L_2}^T \\ 0 & U \end{bmatrix} \begin{bmatrix} -U + 2Q_{GL_2}^T k_{22} & 2Q_{GL_2}^T k_{21} Q_{C_2R}^T \\ -2k_{12} & U - 2k_{11} Q_{C_2R}^T \end{bmatrix} \begin{bmatrix} 0 \\ X \end{bmatrix}. \quad (5.25)$$

Using (5.16), the above equation can be reduced to

$$S_{11} W_1 = \begin{bmatrix} -U & Q_{C_2L_2}^T \\ 0 & U \end{bmatrix} \begin{bmatrix} 0 \\ X \end{bmatrix} = \begin{bmatrix} Q_{C_2L_2}^T \\ U \end{bmatrix} X \quad (5.26)$$

from which we have

$$S_{11} W_1 = W_1. \quad (5.27)$$

Equation (5.27) demonstrates that the set of independent vectors contained in W_1 is an independent set of eigenvectors corresponding to eigenvalues at +1.

Now consider the matrix

$$\tilde{W}_2 = \begin{bmatrix} -U \\ Q_{C_2 L_2} \end{bmatrix} Y. \quad (5.28)$$

We then have, using (5.21)

$$S_{11}^T \tilde{W}_2 = \begin{bmatrix} -U & 0 \\ Q_{C_2 L_2} & U \end{bmatrix} \begin{bmatrix} -U + 2k_{22}^T Q_{GL_2} & -2k_{12}^T \\ 2Q_{C_2 R} k_{21}^T Q_{GL_2} & U - 2Q_{C_2 R} k_{11}^T \end{bmatrix} \begin{bmatrix} Y \\ 0 \end{bmatrix} \quad (5.29)$$

$$= \begin{bmatrix} -U & 0 \\ Q_{C_2 L_2} & U \end{bmatrix} \begin{bmatrix} -Y \\ 0 \end{bmatrix} = - \begin{bmatrix} -U \\ -Q_{C_2 L_2} \end{bmatrix} Y \quad (5.30)$$

and thus

$$S_{11}^T \tilde{W}_2 = -\tilde{W}_2. \quad (5.31)$$

Use of the reciprocity condition

$$S_{11}^T G_{11} = G_{11} S_{11} \quad (5.32)$$

together with (5.31) yields

$$S_{11} W_2 = -W_2 \quad (5.33)$$

where

$$W_2 = R_{11} \begin{bmatrix} -U \\ Q_{C_2 L_2} \end{bmatrix} Y = R_{11} \tilde{W}_2. \quad (5.34)$$

Equation (5.33) demonstrates that the independent vectors in W_2 constitute an independent set of eigenvectors of S_{11} corresponding to eigenvalues at -1. We have thus proved the following theorem:

Theorem 5.2:

The number of eigenvalues of S_{11} equal to +1 and -1 is equal to the dimension of the null spaces of $Q_{C_2R}^T$ and Q_{GL_2} respectively. Furthermore, the corresponding sets of independent eigenvectors, W_1 and W_2 , are given by

$$W_1 = \begin{bmatrix} Q_{C_2L_2}^T \\ U \end{bmatrix} X, \quad W_2 = R_{11} \begin{bmatrix} -U \\ Q_{C_2L_2} \end{bmatrix} Y$$

where X and Y are matrices whose columns are linearly independent vectors in the null spaces of $Q_{C_2R}^T$ and Q_{GL_2} respectively.

The complete eigenvalue problem as it relates to S_{11} is given by

$$S_{11}W = W\Lambda \quad (5.35)$$

where

$$\Lambda = \begin{bmatrix} U & & \\ & -U & \\ & & \Lambda_3 \end{bmatrix} \quad (5.36)$$

and

$$W = \begin{bmatrix} W_1 & W_2 & W_3 \end{bmatrix} \quad (5.37)$$

The two eigenvalues which are not on the unit circle comprise Λ_3 ; the corresponding eigenvectors make up W_3 .

5.2 GENERATION OF ALTERNATE DIAGONAL LYAPUNOV FUNCTIONS

It is well known that a Lyapunov function for a particular system is not unique. In fact, for an asymptotically stable linear system there are a large, possibly infinite, number. Thus, for those wave digital filters having a nondiagonal G_{11} , the existence of a diagonal Lyapunov function, D , cannot be ruled out a priori. The major result presented in this section consists of a set of necessary and sufficient conditions for the existence of diagonal Lyapunov functions. The results rely on the eigenvalue and eigenvector analysis of the previous section.

The stability of a linear shift-invariant discrete-time system is normally investigated through the matrix, $A^T DA - D$. For the class of systems presently under investigation, the state transition matrix, A , is given by

$$A = \Sigma S_{11}. \quad (5.38)$$

Furthermore, since we are considering only those potential Lyapunov functions which are diagonal, D and Σ commute and then

$$A^T DA - D = S_{11}^T D S_{11} - D. \quad (5.39)$$

Thus, it is sufficient to investigate the behaviour of $S_{11}^T D S_{11} - D$.

Theorem 5.3:

Given a reciprocal pseudolossless system, there does not exist a positive definite matrix D such that $S_{11}^T D S_{11} - D$ is negative definite for $n \geq 3$, where n is the dimension of the square matrix S_{11} .

Proof:

From Theorem 5.1 we know that there exists at least one nonzero x such that

$$S_{11}x = \pm x.$$

Then

$$x^T D x = x^T S_{11}^T D S_{11} x$$

and thus

$$x^T (S_{11}^T D S_{11} - D) x = 0.$$

This proves that $S_{11}^T D S_{11} - D$ cannot be a definite form.

As a result of this theorem we can confine our search for diagonal Lyapunov functions to those for which $S_{11}^T D S_{11} - D$ is negative semidefinite. It is easy to show that $S_{11}^T D = D S_{11}$ is sufficient for the existence of a diagonal D . The proof proceeds as follows: Since S_{11} satisfies the conditions of Lemma 5.1, a nonsingular modal matrix W exists. After the appropriate multiplication we have

$$W^T S_{11}^T D S_{11} W = W^T D S_{11}^2 W = W^T D W \Lambda^2, \text{ where } W \Lambda = S_{11} W. \quad (5.40)$$

Then

$$W^T (S_{11}^T D S_{11} - D) W = W^T D W (\Lambda^2 - U). \quad (5.41)$$

However, because the magnitude of the eigenvalues of S_{11} is bounded by unity and $W^T D W$ is positive definite, we can conclude that $W^T (S_{11}^T D S_{11} - D) W$ and thus $S_{11}^T D S_{11} - D$ must be negative semidefinite. If S_{11} is given, then $S_{11}^T D = D S_{11}$ is easily checked for solutions. The inability to find a solution, however, does not mean that a suitable D for the system does not exist.

Theorem 5.4:

Given a pseudolossless reciprocal wave digital system there exists a positive definite symmetric matrix D such that $S_{11}^T D S_{11} - D$ is negative semidefinite if and only if there exist nonsingular matrices T_1 and T_2 such that the following three conditions hold:

$$a) \quad D W_1 = G_{11} W_1 T_1$$

$$b) \quad DW_2 = G_{11}W_2T_2$$

$$c) \quad W_3^T(S_{11}^T DS_{11} - D)W_3 \text{ is negative semidefinite}$$

G_{11} is that component of the reference conductance matrix G associated with the delay-terminated ports of S and W_1 , W_2 and W_3 contain the eigenvectors of S_{11} .

Proof:

We first establish two conditions. Using the properties of the eigenvalues and eigenvectors of S_{11} given in (5.35)-(5.37), we have

$$W_1^T(S_{11}^T DS_{11} - D)W_1 = 0 \quad \text{for all } D \quad (5.40)$$

and

$$W_2^T(S_{11}^T DS_{11} - D)W_2 = 0 \quad \text{for all } D. \quad (5.41)$$

Now assume that there exists a positive definite D such that $S_{11}^T DS_{11} - D$ is negative semidefinite. Then, using a result given in [41], (5.40) and (5.41) imply respectively that

$$(S_{11}^T DS_{11} - D)W_1 = 0 \quad (5.42)$$

$$(S_{11}^T DS_{11} - D)W_2 = 0. \quad (5.43)$$

From (5.42) we have

$$(S_{11}^T D - D)W_1 = 0 \quad (5.44)$$

or, equivalently,

$$(S_{11}^T - U)DW_1 = 0. \quad (5.45)$$

Similarly, from (5.43), we obtain

$$(S_{11}^T + U)DW_2 = 0. \quad (5.46)$$

Since it has been previously established by Theorem 4.5 that $S_{11}^T G_{11} S_{11} - G_{11}$ is negative semidefinite, then equation (5.45) must be satisfied when $D = G_{11}$. The set of vectors which constitute the columns of $G_{11}W_1$ are

linearly independent and can therefore be used as a set of basis vectors in the null space of $S_{11}^T - U$. Since DW_1 is in this null space, it is a linear combination of the vectors in $G_{11}W_1$ and hence we obtain a). A similar argument shows that $G_{11}W_2$ can serve as a set of basis vectors in the null space of $S_{11}^T + U$ and thus we obtain b). T_1 and T_2 are nonsingular square matrices of appropriate dimensions. Condition c) is obtained as follows: Pre-multiply (5.42) and (5.43) by W_3^T and take the transpose of each equation to produce

$$W_1^T (S_{11}^T DS_{11} - D) W_3 = 0 \quad (5.47)$$

and

$$W_2^T (S_{11}^T DS_{11} - D) W_3 = 0. \quad (5.48)$$

Similarly, pre-multiply (5.42) by W_2^T and take the transpose to obtain

$$W_1^T (S_{11}^T DS_{11} - D) W_2 = 0. \quad (5.49)$$

Now, using (5.40), (5.41) and (5.47) - (5.49), we have

$$\begin{aligned} W^T (S_{11}^T DS_{11} - D) W &= \begin{bmatrix} W_1^T \\ W_2^T \\ W_3^T \end{bmatrix} \begin{bmatrix} S_{11}^T DS_{11} - D \end{bmatrix} \begin{bmatrix} W_1 & W_2 & W_3 \end{bmatrix} \\ &= \text{diag} \begin{bmatrix} 0 & 0 & W_3^T (S_{11}^T DS_{11} - D) W_3 \end{bmatrix} \end{aligned} \quad (5.50)$$

where "diag" is used to denote a diagonal matrix. Since W is nonsingular, $W^T (S_{11}^T DS_{11} - D) W$ is negative semidefinite and thus, from (5.50), we have condition c). This completes the proof of the necessity.

Consider now the sufficiency. Assume that conditions a), b) and c) are met. For $D = G_{11}$ we know that (5.45) and (5.46) are satisfied.

That is

$$(S_{11}^T - U)G_{11}W_1 = 0 \quad (5.51)$$

$$(S_{11}^T + U)G_{11}W_2 = 0. \quad (5.52)$$

Since T_1 is a nonsingular square matrix, the substitution of condition a) into (5.51) produces

$$(S_{11}^T - U)DW_1T_1^{-1} = 0 \quad (5.53)$$

from which we have

$$(S_{11}^T - U)DW_1 = 0 \quad (5.54)$$

which is identical to (5.45). Equation (5.42) then follows directly.

Similarly, by substituting condition b) into (5.52), we obtain equation (5.43). Now, by following the same procedure used in the necessity, we can arrive at (5.50), from which we conclude, with the use of condition c), that $W^T(S_{11}^T DS_{11} - D)W$ and hence $S_{11}^T DS_{11} - D$, is negative semidefinite.

Theorem 5.4 gives necessary and sufficient conditions for the existence of alternate Lyapunov functions which may be either diagonal or nondiagonal. Alternate equivalent forms of the conditions in the theorem can be stated; however, the form given appears to be the most suitable for the present purpose.

As an example of the application of Theorem 5.4, consider the prototype filter shown in Fig. 5.1. For an appropriate choice of element values, such a structure could be a realization of a third-order elliptic lowpass filter. The network graph, showing the tree chosen for the analysis, is given in Fig. 5.2. We obtain the following matrices:

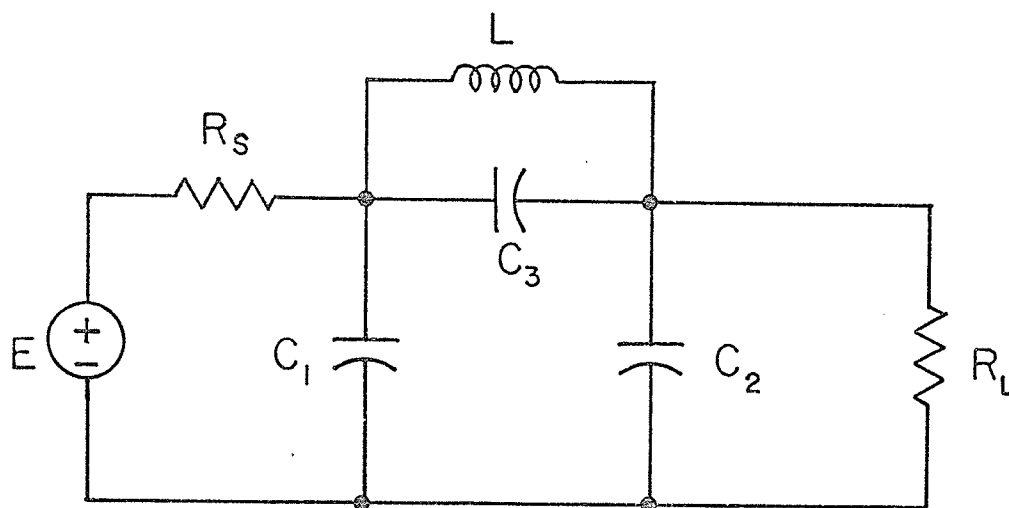


Fig. 5.1 Non-minimal third-order lowpass filter.

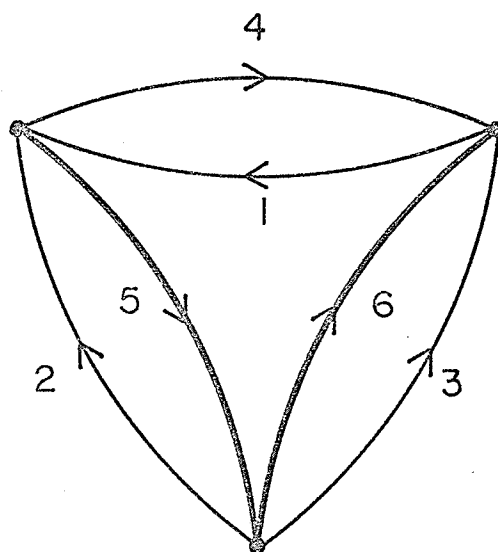


Fig. 5.2 Network graph corresponding to Fig. 5.1.

$$G_{11} = \begin{bmatrix} G_4 & 0 & 0 \\ 0 & G_5 + G_1 & G_1 \\ 0 & G_1 & G_6 + G_1 \end{bmatrix} \quad (5.55)$$

and hence

$$R_{11} = \frac{1}{\Delta} \begin{bmatrix} \Delta R_4 & 0 & 0 \\ 0 & G_1 + G_6 & -G_1 \\ 0 & -G_1 & G_1 + G_5 \end{bmatrix} \quad (5.56)$$

where $\Delta = G_1 G_6 + G_1 G_5 + G_5 G_6$.

$$Q = \begin{bmatrix} -1 & -1 & 0 & 1 & | & 1 & 0 \\ -1 & 0 & 1 & 1 & | & 0 & 1 \end{bmatrix}, \quad Q_{C_2 L_2} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \quad (5.57)$$

There are two loops of reactive elements, hence two eigenvalues equal to -1 in the noncanonic S_{11} . One of these eigenvalues, due to the loop of capacitances, is removed in the formation of the canonic filter; hence, only one eigenvalue at -1 now exists in S_{11} . There are no cutsets of reactive elements; hence there are no eigenvalues equal to +1. The two eigenvalues not on the unit circle which are a result of the resistive terminations can easily be determined if desired.

Since there do not exist any resistive twigs, Q_{GL_2} does not exist. Thus we can choose Y (see (5.21)) to be a convenient nonzero value

$$Y = 1 \quad (5.58)$$

which then produces

$$\tilde{W}_2 = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}. \quad (5.59)$$

Using (5.34), condition b) of Theorem 5.4 can be written as

$$DR_{11}\tilde{W}_2 = \tilde{W}_2 T_2 \quad (5.60)$$

where, in this example, T_2 is a nonzero scalar. If we consider only those solutions D which are diagonal

$$D = \text{diag} \begin{bmatrix} d_1 & d_2 & d_3 \end{bmatrix} \quad (5.61)$$

then (5.60) is equivalent to the following three equations:

$$R_4 d_1 = T_2 \quad (5.62a)$$

$$G_6 d_2 = \Delta T_2 \quad (5.62b)$$

$$G_5 d_3 = \Delta T_2. \quad (5.62c)$$

A solution to (5.62) unique to within the arbitrary constant T_2 can always be obtained. If T_2 is chosen to be unity, then

$$d_1 = G_4 \quad (5.63a)$$

$$d_2 = \frac{G_1 G_6 + G_1 G_5 + G_5 G_6}{G_6} \quad (5.63b)$$

$$d_3 = \frac{G_1 G_6 + G_1 G_5 + G_5 G_6}{G_5}. \quad (5.63c)$$

Since all of the prototype element values are positive, (5.63) will yield positive values for the entries of D .

It now only remains to satisfy condition c) of Theorem 5.4.

However, since D is now unique to within an arbitrary multiplicative

constant, it is more convenient to simply check to see if $S_{11}^T D S_{11} - D$ is negative semidefinite for a particular S_{11} . If, for example, we choose the element values to be

$$\begin{array}{lll} R_S = 1 & R_L = 1 & L = 16/15 \\ C_1 = 1 & C_2 = 1 & C_3 = 1/16 \end{array}$$

then S_{11} is given by

$$S_{11} = \frac{1}{32} \begin{bmatrix} -2 & 18 & 18 \\ 15 & -7 & -7 \\ 15 & -7 & -7 \end{bmatrix} \quad (5.64)$$

and

$$D = \begin{bmatrix} 15/16 & 0 & 0 \\ 0 & 9/8 & 0 \\ 0 & 0 & 9/8 \end{bmatrix}. \quad (5.65)$$

It is readily verified that $S_{11}^T D S_{11} - D$ is negative semidefinite and thus, as discussed previously, a signal modification scheme which reduces the magnitude of the adaptor output signals will produce a realizable filter which is free from parasitic oscillations.

The element values given for this particular filter can be obtained by approximating the values given in [30] for a filter having a CC031517 specification. These element values have also been used by Meerkötter [42].

If all of the extraneous inputs and outputs are removed from the n-port adaptor and the remaining matrix is subjected to a numerical refactorings, the following representation is produced:

$$S = \begin{bmatrix} 1 & 0 & 0 & -1 & -1 \\ 0 & 1 & -1 & -2 & -2 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & \frac{1}{4} & \\ & & & & \frac{15}{32} \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & -1 & -1 \\ 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 1 \\ 0 & 1 & -1 & -1 \end{bmatrix} \cdot (5.66)$$

A realization of (5.66) can be obtained using only 8 adders and 2 multipliers. The multipliers, which have values of $1/4$ and $15/32$, can both be operated in the same time slot. The frequency response obtained from a simulation of this filter is shown in Fig. 5.3. A realization of the same filter, using Fettweis' canonic method [14] and the standard adaptors, requires about 15 adders and 4 multipliers. The multipliers, of which three are equal to $1/2$ and the other is equal to $1/16$, require two time slots for the required computations. A canonic realization using the lattice adaptor [15] requires 7 adders and 3 multipliers having values of $1/2$, $1/2$ and $9/16$. Two multiplier time slots are required.

As a second example, we shall consider the fifth-order elliptic type lowpass structure of Fig. 5.4. Analysis using the tree of Fig. 5.5 produces

$$G_{11} = \left[\begin{array}{cc|ccc} G_5 & 0 & & & \\ 0 & G_6 & & & \\ \hline & & G_1+G_7 & 0 & G_1 \\ & & 0 & G_2+G_8 & G_2 \\ & & G_1 & G_2 & G_1+G_2+G_9 \end{array} \right] \quad (5.67)$$

$$Q_{C_2R}^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad Q_{C_2L_2}^T = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \cdot \quad (5.68)$$

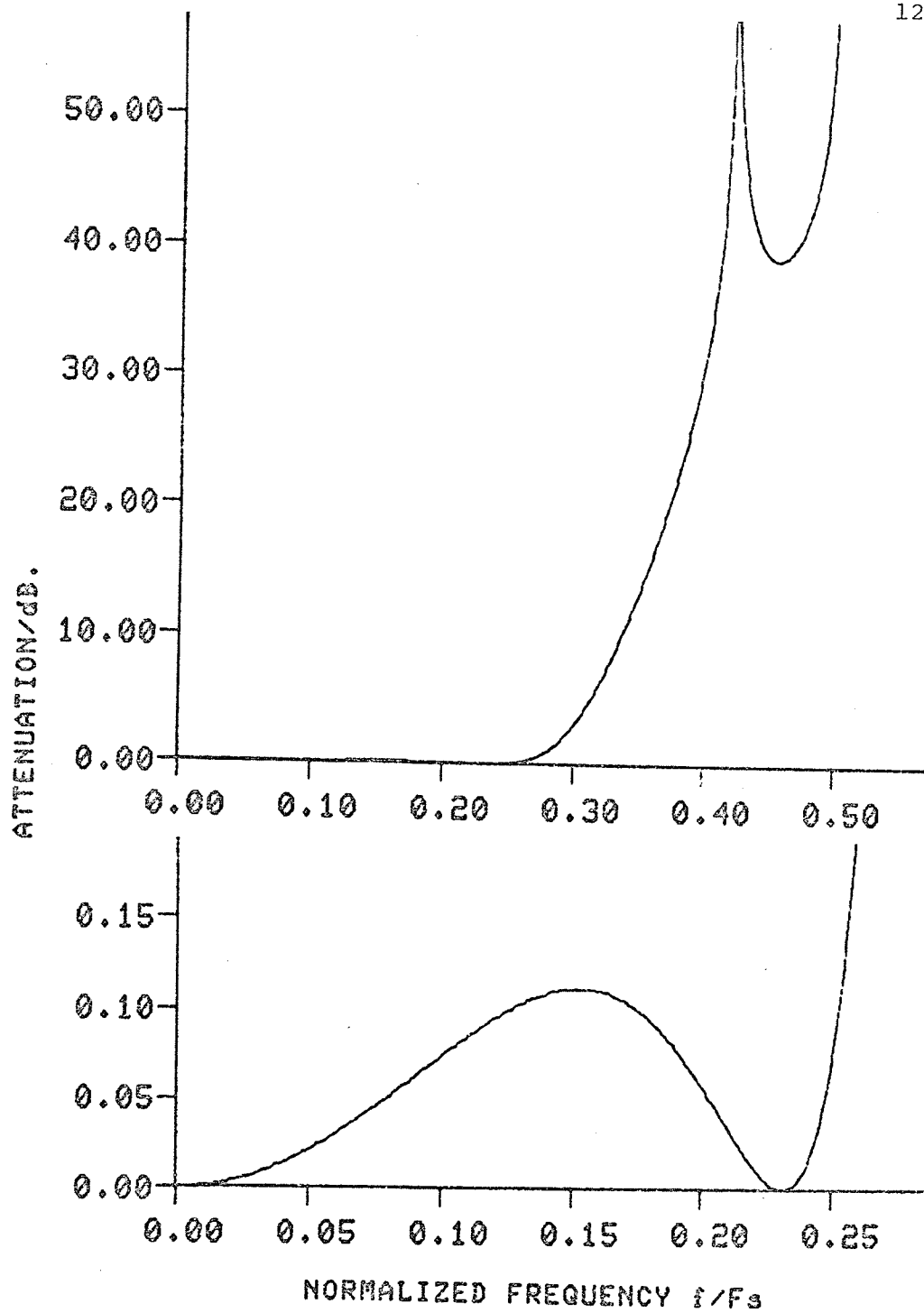


Fig. 5.3 Frequency response of third-order lowpass filter obtained by simulation using (5.66).

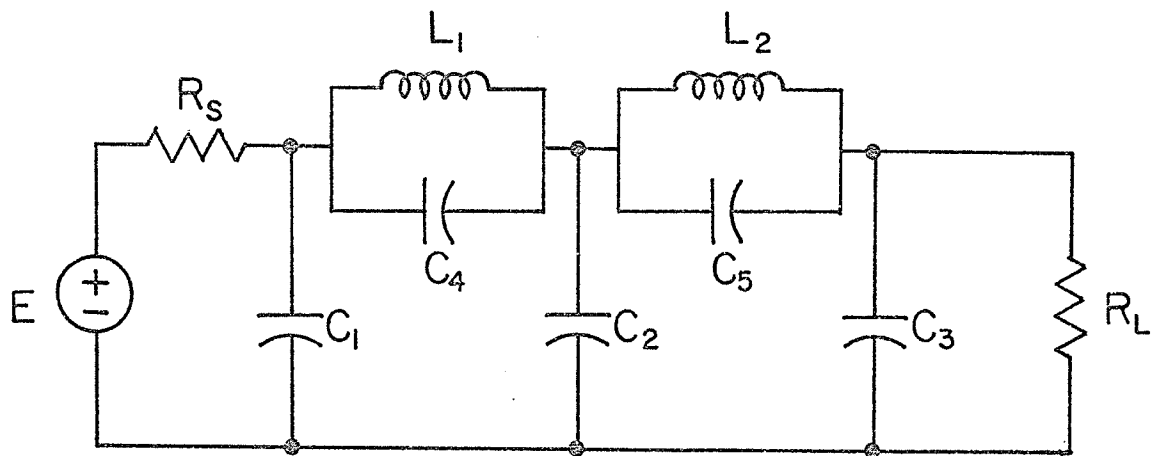


Fig. 5.4 Non-minimal fifth-order elliptic lowpass filter.

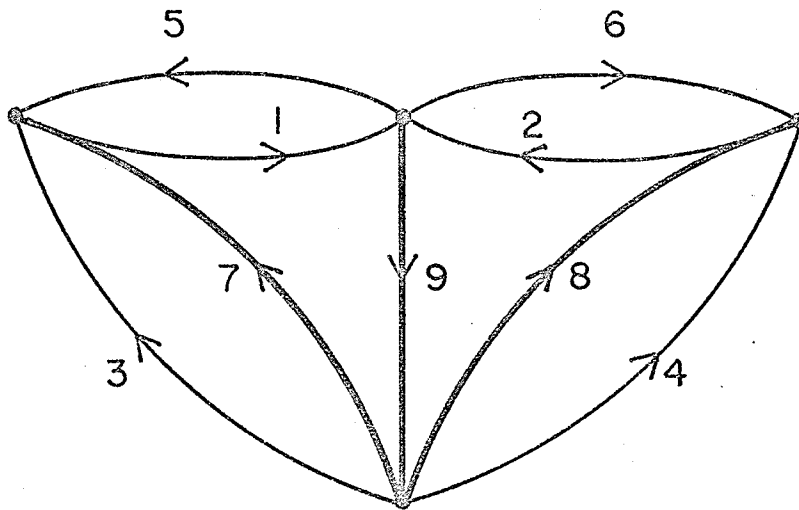


Fig. 5.5 Network graph corresponding to Fig. 5.4.

Of the four reactive loops in the prototype, two are effectively removed by the minimal realization procedure and thus there remain two eigenvalues at -1. The reactive cutset produces one eigenvalue at +1. From $Q_{C_2R}^T$ we obtain

$$X = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad (5.69)$$

while, since Q_{GL_2} does not exist

$$Y = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \quad (5.70)$$

W_1 and \tilde{W}_2 are then found to be

$$W_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \quad \tilde{W}_2 = \begin{bmatrix} -1 & 0 \\ 0 & -1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}. \quad (5.71)$$

Condition a) of the theorem requires that for a diagonal D

$$\begin{bmatrix} d_1 \\ d_2 \\ 0 \\ 0 \\ d_5 \end{bmatrix} = \begin{bmatrix} G_5 \\ G_6 \\ G_1 \\ G_2 \\ G_1 + G_2 + G_9 \end{bmatrix} T_1 \quad (5.72)$$

where T_1 is a nonzero scalar. Since G_1 , G_2 and T_1 are all different from zero, no solution to (5.72) exists and thus we cannot meet the requirements for the existence of a diagonal D. It is interesting to note that the reason for this failure is strictly topological.

The network of Fig. 5.4 contains seven possible normal trees, each one of which will produce a different cutset matrix and thus a different G_{11} , W_1 and W_2 . Unfortunately, the same result is produced in each case and thus we must conclude that there does not exist a diagonal Lyapunov function for a canonical realization of this fifth-order network. Two possible remedies to this problem were considered. The insertion of resistive elements in the prototype, such that the degree of the realization remains unchanged, also failed to produce a desirable G_{11} . Even if one of the modes at -1 was not removed, thus producing a realization of one degree greater than the minimal degree and a G_{11} similar to that in the third-order elliptic case, the results were still negative.

The investigation of other networks, including the Watanabe and seventh-order elliptic type structures used in Chapter III, forces the conclusion that, except in a limited number of cases, diagonal Lyapunov functions do not exist for wave digital filters having a nondiagonal G_{11} .

5.3 TRANSFORMATION OF VARIABLES TO DIAGONALIZE G_{11}

It is well known that a similarity transformation of the state variables of a linear system will produce another system of the same dimension having the same input-output transfer function. Additionally, the new system is controllable (observable) if and only if the original system is controllable (observable). If such a transformation is applied to a wave digital filter for which a diagonal Lyapunov function does not exist, then it is possible that for the new system such a diagonal function will exist.

Consider the nonsingular linear transformation P , relating the original state variables x , to the new set y :

$$x = Py. \quad (5.73)$$

The transformed system will then have a state transition matrix of the form

$$A = P^{-1} \Sigma S_{11} P. \quad (5.74)$$

If we restrict our attention to those transformations which are block diagonal and commutable with Σ , then

$$A = \Sigma \tilde{S}_{11} \quad (5.75)$$

where

$$\tilde{S}_{11} = P^{-1} S_{11} P. \quad (5.76)$$

Next consider the matrix $S_{11}^T G_{11} S_{11} - G_{11}$ which is known to be symmetric and negative semidefinite. A congruence transformation using P produces another negative semidefinite matrix as follows:

$$P^T (S_{11}^T G_{11} S_{11} - G_{11}) P = P^T S_{11}^T P^{T-1} P^T G_{11} P P^{-1} S_{11} P - P^T G_{11} P = \tilde{S}_{11}^T D \tilde{S}_{11} - D \quad (5.77)$$

where

$$D = P^T G_{11} P. \quad (5.78)$$

Because G_{11} is positive definite, the matrix P can always be chosen so that D is a diagonal positive definite matrix and thus can be used as a Lyapunov function for the new system.

A realization of the new system is easily obtained as shown in Fig. 5.6. Since the matrices P and P^{-1} have to be built into this realization, it is important that these matrices are chosen with care.

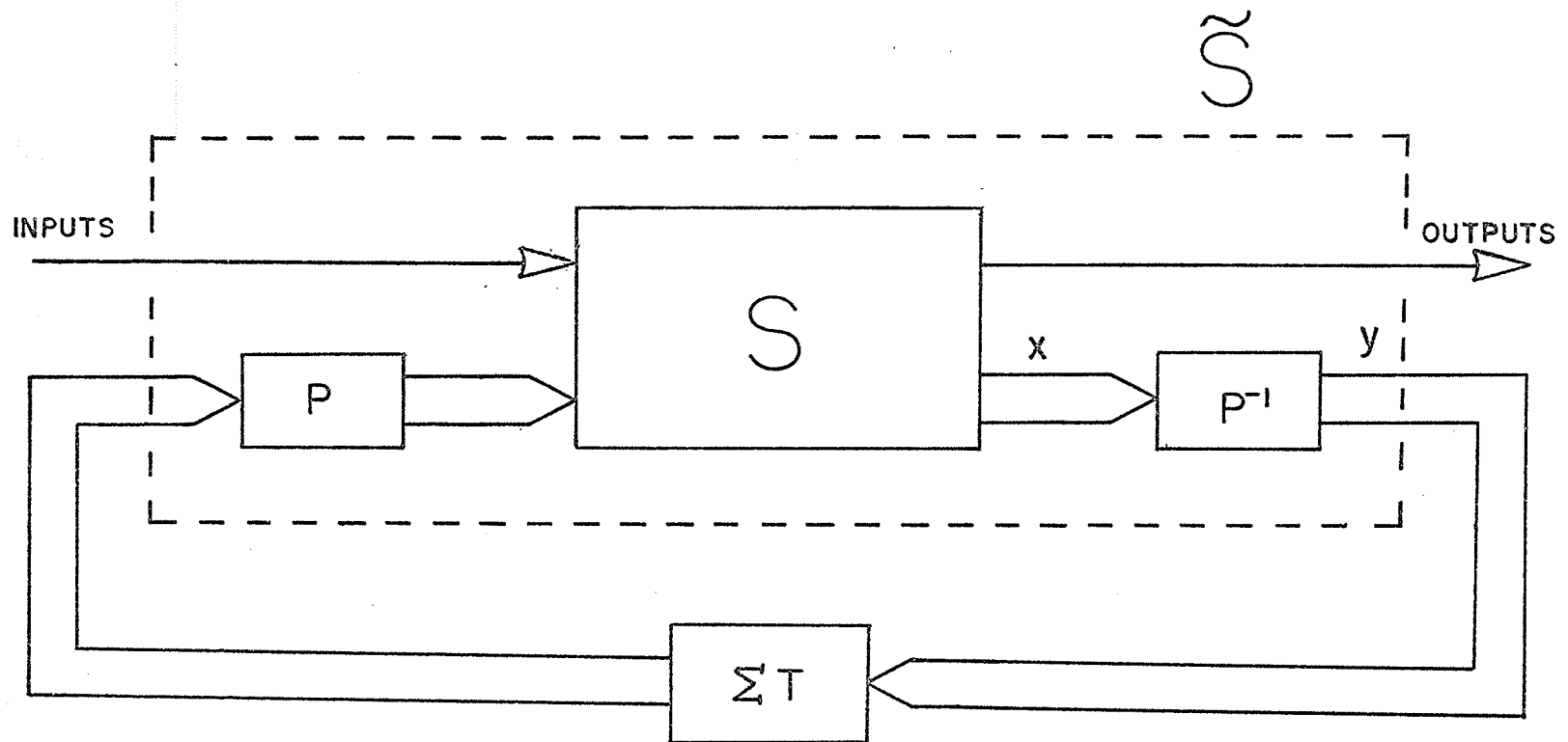


Fig. 5.6 System produced by the introduction of the diagonalization matrices, P and P^{-1} .

We shall show in the next theorem that P can always be chosen so that it is a product of elementary self-inverse matrices. This form is important for three reasons. Firstly, the transformed system cannot suffer from any sensitivity problems with respect to the multipliers in P . No matter which P is implemented, the exact inverse is always used and thus the transformed and original systems have identical transfer functions. Secondly, the elementary product form for P seems to require the minimum number of multipliers and adders for the type of systems which we are considering. Although no proof of this statement is available, experience with several examples supports this claim. Finally, the self-inverse product form of P allows the components of P in the realization to be multiplexed, either in hardware or software, as is the case with the topological matrix F .

Theorem 5.5:

A definite symmetric matrix G can always be reduced to a diagonal form D by a congruence transformation, $P^T G P = D$ where $P = P_1 P_2 \dots P_\ell$ is a product of self-inverse matrices.

Proof:

Let G be an $n \times n$ matrix given by $G = [g_{ij}]$, $i, j = 1, 2, \dots, n$. Then, since G is definite, $g_{ii} \neq 0$ for any i and thus can be used in an elementary column operation to clear out any other nonzero entry in the i^{th} row, say g_{ij} . This operation is equivalent to post-multiplication by the matrix P_1

$$P_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & \\ & & \ddots & \\ & & & 1 & \gamma_1 & \\ & & & & -1 & \\ & & & & & \ddots & \\ 0 & & & & & & 1 \end{bmatrix}, \quad \gamma_1 = \frac{g_{ij}}{g_{ii}}$$

where all diagonal entries are 1 except for the jj entry which is -1 and all off-diagonal entries are 0 except for the ij entry which is equal to γ_1 . P_1^T represents the complementary row operations which will clear out the symmetrical entry in the i^{th} column. It is easily established that $P_1 P_1 = U$. The matrix $P_1^T G P_1$, which now has entries ij and ji equal to zero, is a definite matrix (since P_1 is nonsingular) and hence can be subjected to a similar transformation using P_2 , where $P_2 P_2 = U$. Continuing this procedure using a sequence of self-inverse matrices, it is possible to systematically eliminate all off-diagonal terms to produce the desired diagonal form $D = P^T G P$ where $P = P_1 P_2 \dots P_\ell$. The order and manner in which the entries are cleared out is in general not fixed and thus the sequence of matrices in P is not unique. An upper bound on the number of operations required can be derived as follows: First, clear out the first row and column using the 11 entry. This requires at most $n-1$ operations. Next, using the 22 entry we can clear out the second row and column without altering the first row and column. This requires at most $n-2$ operations. Continuing on in this manner, we finally need at most 1 operation to clear out the $n-1$ row and column. Thus $\ell \leq (n-1) + (n-2) + \dots + 1 = n(n-1)/2$.

In order to illustrate the procedure required to obtain D , consider the form of G_{11} obtained for the fifth-order structure of the

previous section (see equation (5.67)). Since the method used in Chapter III assumed that there is no interaction between the capacitive and inductive redundancies, G_{11} can be written as the direct sum of $G_{\mathcal{L}}$ and $G_{\mathcal{C}}$.

$$G_{11} = \text{diag} [G_{\mathcal{L}}, G_{\mathcal{C}}] \quad (5.79)$$

P can therefore be written as the direct sum of $P_{\mathcal{L}}$ and $P_{\mathcal{C}}$

$$P = \text{diag} [P_{\mathcal{L}}, P_{\mathcal{C}}] \quad (5.80)$$

and thus $P_{\mathcal{L}}$ and $P_{\mathcal{C}}$ are determined independently. For the present example there are no inductive redundancies and thus $G_{\mathcal{L}}$ is diagonal and $P_{\mathcal{L}} = U$. $G_{\mathcal{C}}$ can be diagonalized using the self-inverse matrix

$$P_{\mathcal{C}} = \begin{bmatrix} 1 & 0 & \gamma_1 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \gamma_2 \\ 0 & 0 & -1 \end{bmatrix} = P_1 P_2 \quad (5.81)$$

where

$$\gamma_1 = \frac{G_1}{G_1 + G_7}, \quad \gamma_2 = \frac{G_2}{G_2 + G_8}. \quad (5.82 \text{ a,b})$$

An alternative form for P_2 using $\gamma_2 = G_2/(G_1 + G_2 + G_9)$ is possible.

As a second example, we shall return to the filter introduced in Example 3.2. The lowpass to bandpass transformation used in that example results in a system for which the state transition matrix is not ΣS_{11} and thus the stability theory discussed so far is not directly applicable. The transformation in which z is replaced by $-z^2$ effectively doubles the number of delay terminated ports. Each port which would have been terminated by a delay in the lowpass filter is transformed into two such ports in the bandpass filter, as shown in Fig. 5.7.

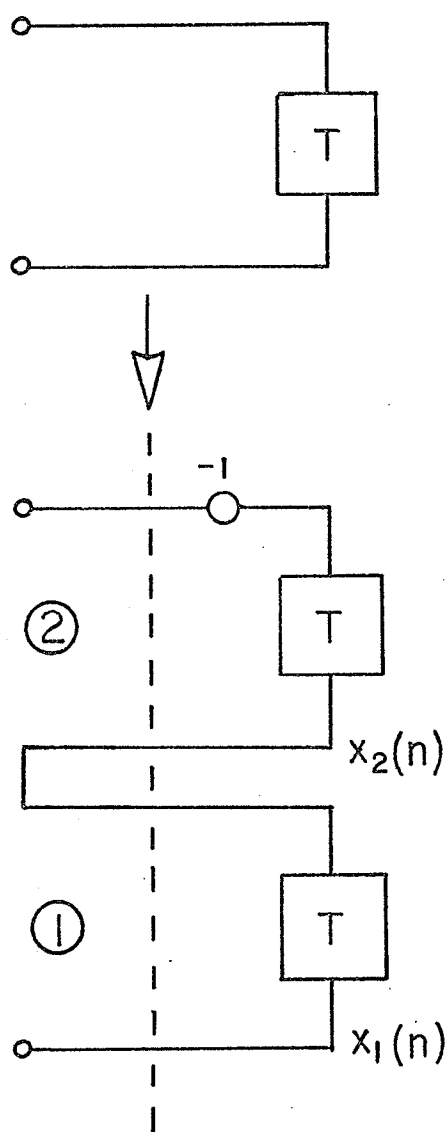


Fig. 5.7 Effect of the lowpass to bandpass transformation on the delay terminated ports.

If the state variables are chosen as shown in Fig. 5.7, then the state equations are

$$\begin{bmatrix} x_1(n+1) \\ x_2(n+1) \\ y(n) \end{bmatrix} = \begin{bmatrix} 0 & U & 0 \\ -\Sigma S_{11} & 0 & -\Sigma S_{12} \\ S_{21} & 0 & S_{22} \end{bmatrix} \begin{bmatrix} x_1(n) \\ x_2(n) \\ u(n) \end{bmatrix}. \quad (5.83)$$

After defining $\hat{\Sigma} = -\Sigma$, these state equations can be associated with an n-port adaptor specified by

$$\hat{S} = \begin{bmatrix} 0 & U & 0 \\ S_{11} & 0 & S_{12} \\ S_{21} & 0 & S_{22} \end{bmatrix}. \quad (5.84)$$

If we consider

$$\hat{G} = \begin{bmatrix} G_{11} & & \\ & G_{11} & \\ & & G_{22} \end{bmatrix} \quad (5.85)$$

where G_{11} and G_{22} are the components of the port conductance matrix for the lowpass filter, then it can be verified that, since the lowpass adaptor is pseudolossless, so is the bandpass adaptor. That is

$$\hat{S}^T \hat{G} \hat{S} - \hat{G} = 0. \quad (5.86)$$

It is interesting to note that this adaptor is not reciprocal and hence not self-inverse. Also, the system (5.83) can be proven to be minimal.

Since the bandpass adaptor is pseudolossless with respect to \hat{G} then, from Theorem 4.5,

$$\hat{G}_{11} = \begin{bmatrix} G_{11} & \\ & G_{11} \end{bmatrix} \quad (5.87)$$

can be used as a Lyapunov function for the system. Noting that $\lambda_{BP} = +(-\lambda_{LP})^{\frac{1}{2}}$ together with the fact that the lowpass minimal realization is asymptotically stable ($|\lambda_{LP}| < 1$), we have $|\lambda_{BP}| < 1$ and thus asymptotic stability is preserved in the bandpass realization. Theorem 4.7 can then be used to guarantee freedom from zero-input limit cycles.

In order to produce a diagonal Lyapunov function, \hat{D} , for the bandpass filter, it is only necessary to diagonalize G_{11} in the lowpass filter. The frequency transformation then produces $\hat{D} = \text{diag} [D, D]$. This means that P and P^{-1} need only be inserted once each in the flow diagram.

Because there are no inductive degeneracies in this seventh-order structure, it is necessary only to diagonalize the capacitive component of G_{11} . From (3.84), $G_{\mathcal{C}}$ is obtained as

$$G_{\mathcal{C}} = \begin{bmatrix} \frac{10}{3} & \frac{1}{3} & 0 & 0 \\ \frac{1}{3} & \frac{11}{3} & \frac{1}{3} & 0 \\ 0 & \frac{1}{3} & \frac{25}{6} & \frac{1}{3} \\ 0 & 0 & \frac{1}{3} & \frac{4}{3} \end{bmatrix} \quad (5.88)$$

The technique which produces the minimum number of multipliers requires that the off-diagonal elements in the first and last row and column be cleared out first. This is accomplished using

$$P_1 = \begin{bmatrix} 1 & \frac{1}{10} & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & \frac{1}{4} & 1 \end{bmatrix} \quad (5.89)$$

Then

$$P_1^T G P_1 = \begin{bmatrix} \frac{10}{3} & 0 & 0 & 0 \\ 0 & \frac{109}{30} & \frac{1}{3} & 0 \\ 0 & \frac{1}{3} & \frac{49}{12} & 0 \\ 0 & 0 & 0 & \frac{4}{3} \end{bmatrix} \quad (5.90)$$

The centre block in (5.90) can be diagonalized in two ways. If the 22 entry is used, then

$$P_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & \frac{10}{109} & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (5.91)$$

and

$$D = \text{diag} \left[\frac{10}{3}, \frac{109}{30}, \frac{5301}{1308}, \frac{4}{3} \right] \quad (5.92)$$

If the 33 entry is used, then

$$P_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & \frac{4}{49} & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (5.93)$$

and

$$D = \text{diag} \left[\frac{10}{3}, \frac{323}{147}, \frac{49}{12}, \frac{4}{3} \right]. \quad (5.94)$$

In either case $P_q = P_1 P_2$ requires 3 multipliers and 3 adders. One of the multipliers has a finite word length binary representation and thus can be realized exactly. The others are not of finite word length and thus must be approximated for realizability. When this is done, D is no longer diagonal. This problem and others regarding the practical implementation of the signal modification scheme are discussed in the next chapter.

CHAPTER VI

SUPPRESSION OF PARASITIC OSCILLATIONS IN NONLINEAR WAVE DIGITAL FILTERS USING n-PORT ADAPTORS

The systems obtained using the standard n-port adaptor method or the minimal systems using the diagonalization procedure of the last chapter both have a diagonal Lyapunov function. In this chapter we develop techniques which can be used to obtain bounds on the errors caused by finite signal word lengths and coefficient truncation in P. These bounds and the results of Theorem 4.7 and Corollary 4.7 can be used to define the signal modifications required to guarantee continued pseudopassivity of the filter and hence freedom from parasitic oscillations.

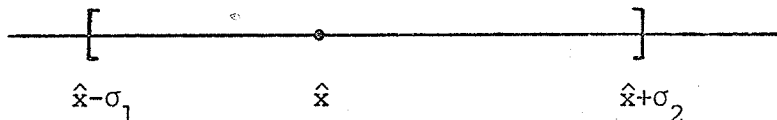
6.1 SIGNAL MODIFICATIONS FOR STABILITY

Finite precision in the signals and coefficients cause the signals in a true digital filter to deviate from the corresponding signals in the associated linear filter. Anticipating the results of the next section, we shall assume that each signal \hat{x} in the finite precision realization has associated with it an error interval such that the corresponding signal, x , in the ideal filter lies in this interval. That is

$$x \in [\hat{x} - \sigma_1, \hat{x} + \sigma_2], \quad \sigma_1 \geq 0, \sigma_2 \geq 0 \quad (6.1)$$

Standard interval notation is used where \in denotes membership.

This interval can be illustrated as shown below.


$$\quad (6.2)$$

As discussed in the closing comments of Chapter IV, parasitic oscillations can be avoided if, for the same input state, the next states of the finite precision filter are smaller in magnitude than those in the ideal filter. Examination of (6.1) or (6.2) shows that this condition is not automatically guaranteed since the ideal value could be anywhere in the interval. However, we shall now show that it is possible to modify the finite precision outputs at the delays so that the magnitude condition is always satisfied. These modifications actually consist of the substitution of a different set of signal values at the delays.

The specific form of the modifications depends upon the manner in which the signal values are represented. We shall assume that the filters under consideration are realized using two's-complement fixed-point arithmetic. This form of arithmetic is used in the majority of hardware filters and is also the basic arithmetic implemented in most minicomputers. If the signal is represented by $m+k$ bits

$$\delta_{-k} \delta_{-k+1} \dots \delta_{-1} \delta_0 \delta_1 \dots \delta_{m-1} \quad (6.3)$$

where the radix point is assumed to be between δ_0 and δ_1 , then the actual signal value is given by

$$x = -\delta_{-k} 2^k + \sum_{i=-k+1}^{m-1} \delta_i 2^{-i}. \quad (6.4)$$

The values of k and m can be different at different points in the filter; however, we will assume that the signals in the main delays have $k = 0$. We also assume that the signals \hat{x} at the output of the adaptor before modification have a value of m consistent with the

number of bits in the delays but that $k > 0$ is temporarily allowed. The signal modification procedure must therefore produce a new set of signals with $k = 0$ which satisfy the magnitude condition. This can be accomplished in two steps. First, a small correction will be applied to produce signals which satisfy the magnitude condition but may not have $k = 0$. The errors introduced by this primary correction can be made arbitrarily small by allowing an increase in the precision of the computations within the filter. Secondly, if $k \neq 0$ after the primary correction, then the signal is simply front-chopped [25]. This is equivalent to discarding the bits to the left of δ_0 . The fact that this procedure reduces the magnitude is easily established. If any of the bits chopped were nonzero, then the original magnitude must have been greater than or equal to one. The front-chopped version, however, is always less than or equal to one in magnitude. Hence, the magnitude is either reduced or stays the same as a result of front-chopping. This type of signal modification can cause large errors in the signals and thus the number of overflows should be limited by proper use of scaling.

The primary correction scheme to be discussed next ideally requires that either $-\sigma_1$ or σ_2 may be added to the signals at the delays. However, in general, σ_1 and σ_2 cannot be represented exactly in the word length allowed at the delays and thus the addition is not realizable. In situations such as this, the least upper bounds on σ_1 and σ_2 must be replaced by weaker bounds which are realizable. Thus, σ_1 is replaced by $\hat{\sigma}_1$ where $\hat{\sigma}_1$ is the smallest number which satisfies $\hat{\sigma}_1 \geq \sigma_1$ and which has an exact representation in the word length allowed. Similarly, σ_2 is replaced by $\hat{\sigma}_2$. For example,

if there are 15 bits after the radix point, then $2^{-15} + 2^{-16}$, which is not realizable, is replaced by 2^{-14} .

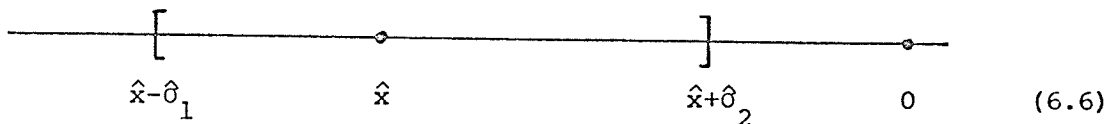
Depending upon the value of \hat{x} , four different situations for the primary corrections arise.

a) $\hat{x} > 0$, $\hat{x} > \hat{\sigma}_1$, which can be represented by



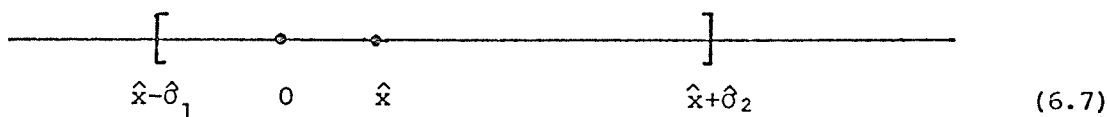
In this case $|\hat{x}-\hat{\sigma}_1| \leq |x|$ and thus the magnitude condition could be satisfied if \hat{x} was replaced by $\hat{x}-\hat{\sigma}_1$. The value $\hat{x}-\hat{\sigma}_1$ can be easily obtained by adding $-\hat{\sigma}_1$ to \hat{x} .

b) $\hat{x} < 0$, $|\hat{x}| > \hat{\sigma}_2$, which can be represented by



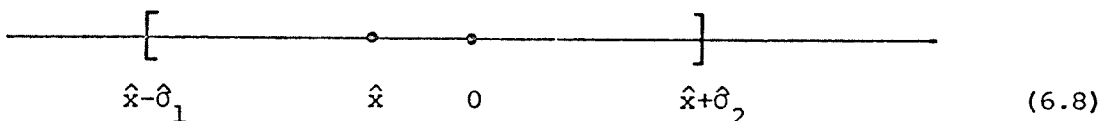
In this case $|\hat{x}+\hat{\sigma}_2| \leq |x|$ and thus the magnitude condition could be satisfied if \hat{x} was replaced by $\hat{x}+\hat{\sigma}_2$. The value $\hat{x}+\hat{\sigma}_2$ can be produced by adding $\hat{\sigma}_2$ to \hat{x} .

c) $\hat{x} \geq 0$, $\hat{x} \leq \hat{\sigma}_1$, which can be represented by



Since x can take on any value in the interval, zero is the only value which is assured of being smaller in magnitude than x . The magnitude condition is satisfied by replacing \hat{x} by zero.

d) $\hat{x} < 0$, $|\hat{x}| \leq \hat{\sigma}_2$, which can be represented by



This situation is similar to that in c) and hence the magnitude

condition is satisfied by replacing \hat{x} by zero.

The compensation just described could be carried out by first determining both the magnitude and the sign of the signal after which the appropriate modification can take place. An equivalent, yet computationally less cumbersome, method can be used if the correction routine is described in the following equivalent form:

If \hat{x} is positive, add $-\hat{o}_1$. Then, if the new signal $\hat{x}_1 - \hat{o}_1$ is still positive, no further action is required. If $\hat{x}_1 - \hat{o}_1$ is now negative, it must be set to zero. If \hat{x} is negative, add \hat{o}_2 . Then, if the new signal $\hat{x}_1 + \hat{o}_2$ is still negative, no further action is required. If $\hat{x}_1 + \hat{o}_2$ is now positive, it must be set to zero.

A straightforward implementation of this scheme could be set up as shown in Fig. 6.1. The sign bit of \hat{x} is used to select the appropriate signal, $-\hat{o}_1$ or \hat{o}_2 , which is added to \hat{x} . The EXCLUSIVE-NOR gate determines whether the sign has changed; if it has changed, the output is 0. The set of AND gates either produces zero or the output of the adder. The leading k bits in the final output are deleted so that $k = 0$.

The data selector can be realized using only a few inverters. If we denote $\delta_i(-\hat{o}_1)$ as the i^{th} bit in $-\hat{o}_1$ and $\delta_i(\hat{o}_2)$ as the i^{th} bit in \hat{o}_2 , then four situations are apparent:

- a) $\delta_i(-\hat{o}_1) = 0, \delta_i(\hat{o}_2) = 0$
- b) $\delta_i(-\hat{o}_1) = 0, \delta_i(\hat{o}_2) = 1$
- c) $\delta_i(-\hat{o}_1) = 1, \delta_i(\hat{o}_2) = 1$
- d) $\delta_i(-\hat{o}_1) = 1, \delta_i(\hat{o}_2) = 0$.

The correct value for the i^{th} bit out of the data selector is obtained as follows: In case

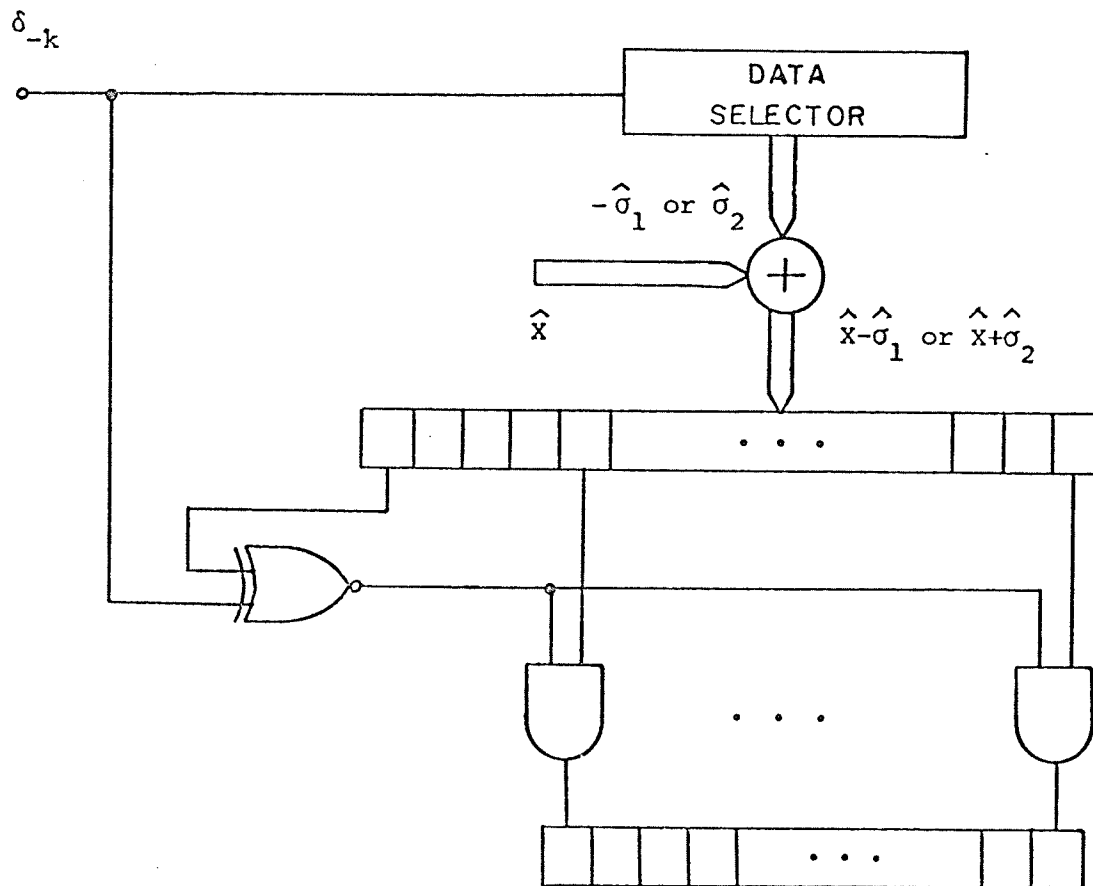


Fig. 6.1 Implementation of the scheme which guarantees freedom from parasitic oscillations.

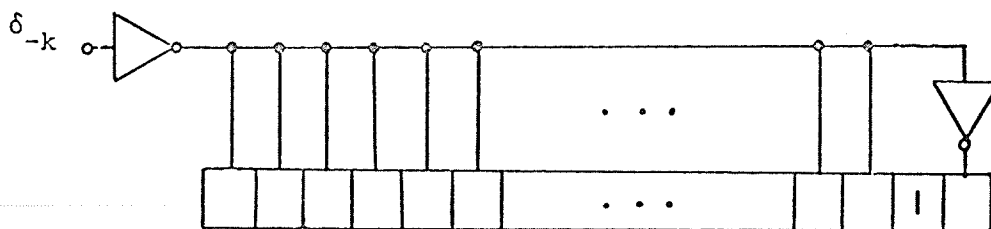


Fig. 6.2 An example of the data selector implementation when $m=15$, $\hat{\sigma}_1 = 2^{-13}$ and $\hat{\sigma}_2 = 2^{-13} + 2^{-14}$.

- a) hardwire a 0,
- b) use the sign bit,
- c) hardwire a 1,
- d) complement the sign bit.

This is illustrated in Fig. 6.2, for the case where $m = 15$, $\hat{o}_1 = 2^{-13}$ and $\hat{o}_2 = 2^{-13} + 2^{-14}$.

The primary and secondary modifications are not generally commutative operations. One exception occurs when $\hat{o}_1 = 0$ and \hat{o}_2 has a value equal to a 1 in the least significant bit of the signal. In this case it may be possible to obtain a reduction in the hardware complexity, since then modulo 2 adders may be used instead of full precision adders at a number of points in the realization. A reduction in the hardware needed to perform the signal modifications can also be achieved. It is necessary only to add a 1 in the least significant bit when the signal is negative. The addition of this bit cannot produce a positive number and therefore no additional checking of the sum is required.

The technique presented in this section can be considered to be a generalization of the method used by Fettweis and Meerkötter [25] as applied to wave digital structures using series and parallel adaptors. In those realizations the above - mentioned special case occurs. A somewhat similar technique has also been used by Lê [20].

6.2 ERROR INTERVAL ANALYSIS

Error analysis using interval algebra has been used to study the errors caused by finite word length effects in computational algorithms [43]. More recently, this technique has been applied,

as an alternative to the commonly used statistical methods, in the study of quantization noise in digital filters [44]. In this section we use a modified form of interval arithmetic to monitor the errors caused by the finite precision realization.

Two sources of error occur. The first is due to the signal word length reduction necessary in a recursive realization. If such reduction is not performed, the signal word length would grow indefinitely as signals propagate inside the feedback loops. The second source is due to coefficient quantization. Since a linear pseudolossless realization with finite word length coefficients can be used as the basis of the filter, the only source of this form of error is in the diagonalization matrices P and P^{-1} which must be approximated for realizability.

Assume that each signal, \hat{x} , in the finite precision realization has an error, e , associated with it such that the corresponding signal, x , in the ideal filter is given by

$$x = \hat{x} + e. \quad (6.9)$$

If the error is known to lie in the interval

$$e \in [-\sigma_1, \sigma_2], \quad (6.10)$$

then the ideal signal x must be in the interval given by

$$x \in [\hat{x} - \sigma_1, \hat{x} + \sigma_2]. \quad (6.11)$$

As these signals pass through the filter, new signals, each having its own error interval, are created. The error at the output of an arithmetic or word length reducing operation depends upon both

the error in the input and any additional error introduced by the operation. If the output error, denoted by e_o , is restricted to an interval similar to (6.10), then the interval in which the ideal signal value lies can be computed from e_o and the realized signal value in a manner identical to that used to produce (6.11) from (6.9) and (6.10).

Using (6.9)-(6.11) to define the input signals, we now consider the output error associated with the individual operations found in a digital filter realization.

(a) Inversion

From (6.9)

$$-x = (-\hat{x}) + (-e). \quad (6.12)$$

Since no additional errors are introduced by an inversion, the interval of the output error $e_o = (-e)$ is obtained from the input error interval by interchanging the absolute value of the end points while maintaining the sign. Thus, from (6.10),

$$e_o = (-e) \in [-\sigma_2, \sigma_1]. \quad (6.13)$$

(b) Addition

If $x_1 = \hat{x}_1 + e_1$ and $x_2 = \hat{x}_2 + e_2$, then

$$x_1 + x_2 = (\hat{x}_1 + \hat{x}_2) + (e_1 + e_2). \quad (6.14)$$

Here we assume that the addition is carried out with full precision which necessitates that the word length at the output of the adder is 1 bit longer than the largest input word length. The input error intervals, which are given by

$$e_1 \in [-\sigma_1(x_1), \sigma_1(x_1)], \quad e_2 \in [-\sigma_1(x_2), \sigma_2(x_2)] \quad (6.15)$$

combine to produce the output error

$$e_o = (e_1 + e_2) \in [-\sigma_1(x_1) - \sigma_1(x_2), \sigma_2(x_1) + \sigma_2(x_2)]. \quad (6.16)$$

(c) Exact multiplication

Using (6.9)

$$\alpha x = \alpha \hat{x} + \alpha e, \quad \alpha > 0. \quad (6.17)$$

In this case we assume that the multiplier coefficient α is positive and can be represented exactly in the word length available. The effect of the multiplier is to scale the input error e into the output error $e_o = \alpha e$

$$e_o = (\alpha e) \in [-\alpha\sigma_1, \alpha\sigma_2]. \quad (6.18)$$

At this point we have assumed that the multiplication $\alpha \hat{x}$ has been carried out in full precision. Truncation or rounding, which will be discussed shortly, is normally used to reduce the output word length.

Multiplication by negative coefficients can be interpreted as the cascade of an inversion and multiplication by a positive coefficient.

(d) Multiplication by a quantized coefficient

The exact multiplier value α is given by

$$\alpha = \hat{\alpha} + \Delta\alpha \quad \alpha > 0, \quad \hat{\alpha} > 0 \quad (6.19)$$

where $\hat{\alpha}$ is the quantized multiplier and $\Delta\alpha$ is the error. Again we assume that both α and $\hat{\alpha}$ are positive. Using (6.19) and (6.9), the ideal output signal is

$$\begin{aligned}
 \alpha x &= (\hat{\alpha} + \Delta\alpha) (\hat{x} + e) \\
 &= \hat{\alpha}\hat{x} + \hat{\alpha}e + \Delta\alpha x.
 \end{aligned} \tag{6.20}$$

The output which appears in the realized system, $\hat{\alpha}\hat{x}$, differs from the ideal output, αx , by the output error e_o .

$$\alpha x = \hat{\alpha}\hat{x} + e_o. \tag{6.21}$$

e_o consists of two terms; $e_1 = \hat{\alpha}e$ which is due to the error in the input and $e_2 = \Delta\alpha x$ which can be attributed to the error in the multiplier. If x , the signal value which would have occurred at the input of the multiplier in an ideal realization is known, then the e_2 is known exactly and can be expressed in interval form as

$$e_2 \in [\Delta\alpha x, \Delta\alpha x]. \tag{6.22}$$

Then, using (6.16) and (6.18), we obtain

$$e_o = (e_1 + e_2) \in [-\hat{\alpha}\sigma_1 + \Delta\alpha x, \hat{\alpha}\sigma_2 + \Delta\alpha x]. \tag{6.23}$$

When x is unknown but can be bounded

$$x_{\min} \leq x \leq x_{\max}$$

that is,

$$x \in [x_{\min}, x_{\max}] \tag{6.24}$$

then, using (6.13) and (6.18), e_2 becomes

$$e_2 \in [\Delta\alpha x_{\min}, \Delta\alpha x_{\max}], \Delta\alpha > 0 \tag{6.25a}$$

or

$$e_2 \in [\Delta\alpha x_{\max}, \Delta\alpha x_{\min}], \Delta\alpha < 0 \tag{6.25b}$$

and e_o is given by

$$e_o \in [-\hat{\alpha}\sigma_1 + \Delta\alpha x_{\min}, \hat{\alpha}\sigma_2 + \Delta\alpha x_{\max}], \Delta\alpha > 0 \quad (6.26a)$$

or

$$e_o \in [-\hat{\alpha}\sigma_1 + \Delta\alpha x_{\max}, \hat{\alpha}\sigma_2 + \Delta\alpha x_{\min}], \Delta\alpha < 0. \quad (6.26b)$$

The intervals associated with the output signal can be obtained if desired from $\hat{\alpha} \hat{x}$ and the appropriate output error, (6.23) or (6.26).

The intervals in (6.26) produced using the bounds on x will generally be larger than the interval in (6.23) obtained from the exact value. However, since these larger intervals can be computed independently of the precise values of the signals and therefore need only be determined once for a particular filter, they will be used exclusively in the remainder of this chapter.

e) Truncation of the two's-complement representation
Truncation of a signal is performed by simply deleting a specified number of bits at the least significant end of the word. If the input signal \hat{x} is truncated so that in the output signal \hat{x}_T there remain $r-1$ bits to the right of the radix point, then, with reference to (6.3) and (6.4), the error introduced is

$$e_T = \hat{x} - \hat{x}_T = \sum_{i=r}^{m-1} \delta_i 2^{-i}. \quad (6.27)$$

Bounds on this truncation error are given by

$$e_T \in [0, \mu_2] \quad (6.28)$$

where

$$\mu_2 = \sum_{i=r}^{m-1} 2^{-i}. \quad (6.29)$$

Since the input signal \hat{x} in general has an associated nonzero error component, then, from (6.9) and (6.27),

$$x = \hat{x} + e = \hat{x}_T + (e_T + e). \quad (6.30)$$

The output error $e_o = e_T + e$ can be computed from (6.28), (6.29) and (6.10) using (6.16)

$$e_o = (e_T + e) \in [-\sigma_1, \sigma_2 + \mu_2]. \quad (6.31)$$

f) Rounding of the two's complement representation

Rounding of a signal to reduce the word length is often used as an alternative to truncation. This signal modification scheme can be modelled as a truncation followed by the addition of the most significant truncated bit into the least significant bit of the truncated signal. Using this model and the interval for e_T , (6.28) and (6.29), the roundoff error, when $r-1$ bits are retained to the right of the radix point, is

$$e_R = \hat{x} - \hat{x}_R \quad (6.32)$$

where

$$e_R \in [-\mu_1, \mu_2] \quad (6.33)$$

$$\text{and } \mu_1 = 2^{-r}, \quad \mu_2 = \sum_{i=r+1}^{m-1} 2^{-i}. \quad (6.34)$$

If the input signal is in error, then

$$x = \hat{x} + e = \hat{x}_R + (e_R + e) \quad (6.35)$$

and the output error $e_o = e_R + e$ is

$$e_o = (e_R + e) \in [-\sigma_1 - \mu_1, \sigma_2 + \mu_2]. \quad (6.36)$$

The width of the error interval for rounding (6.33), (6.34) is the same as the width obtained for truncation (6.28), (6.29). The rounding interval, however, is essentially symmetrically

placed about zero and thus is often preferred to truncation.

The implementation of an interval arithmetic routine can be conveniently carried out on a digital computer using complex arithmetic in which the lower and upper bounds on the intervals are represented by the real and imaginary components of a complex number. All of the required operations, with the exception of inversion, are easily carried out using standard complex addition and multiplication. Inversion can be accomplished by multiplying the conjugate of the complex number representing the input interval by the complex number $(0, -1)$, i.e. $-j$.

The computation of the error bounds used in Section 6.1 can be carried out using interval arithmetic in two ways. The first method is based upon a strictly arithmetic procedure in which the errors are assumed to be independent from signal to signal. The second method exploits the dependent nature of the signals in a combination algebraic and arithmetic routine.

The difference in the intervals produced by these methods can best be illustrated by a simple example. Consider the expression

$$y = x_1 - \alpha(x_1 + x_2), \quad \alpha > 0 \quad (6.37)$$

We shall assume that the filter realization is such that the sum of $x_1 + x_2$ is first computed, followed by multiplication by α and the truncation of the product, which is then subtracted from x_1 . Also assume that the input signals x_1 and x_2 are given by

$$x_1 = \hat{x}_1 + e_1, \quad x_2 = \hat{x}_2 + e_2 \quad (6.38 \text{ a,b})$$

where, for convenience,

$$e_1 = e_2 \in [-\sigma_1, \sigma_2]. \quad (6.39)$$

Using the rules just developed, the ultimate ideal output signal, y , satisfies

$$y = \hat{y} + e_o \quad (6.40)$$

where

$$\hat{y} = \hat{x}_1 - [\alpha(\hat{x}_1 + \hat{x}_2)]_T \quad (6.41)$$

and

$$e_o = e_1 - \alpha(e_1 + e_2) - e_T. \quad (6.42)$$

The symbol $[]_T$ is used to denote the truncated signal.

The difference in the two methods depends upon how the interval for e_o is evaluated. If the error at the multiplier output, $\alpha(e_1 + e_2)$, is thought of as being independent of e_1 , then a straightforward application of the rules to (6.42) produces

$$e_o \in [-2\alpha\sigma_2 - \sigma_1 - \mu_2, 2\alpha\sigma_1 + \sigma_2]. \quad (6.43)$$

The advantage of this technique is the relative ease in which the intervals can be computed. If a simulation of the filter flow diagram is available, then, with only minor modifications, the same program can be used to compute the error intervals. These changes are required to allow the use of complex signals and to provide for the introduction of the intervals due to the finite word length operations.

If, on the other hand, the expression for (6.42) is simplified algebraically to

$$e_o = (1-\alpha)e_1 - \alpha e_2 - e_T \quad (6.44)$$

then the application of the rules produces

$$e_o \in [-(1-\alpha)\sigma_1 - \alpha\sigma_2 - \mu_2, \alpha\sigma_1 + \sigma_2]. \quad (6.45)$$

This interval is smaller than that of (6.43). For example, if $\sigma_1 = \sigma_2$, then (6.43) becomes $[-(1+2\alpha)\sigma_1 - \mu_2, (1+2\alpha)\sigma_1]$ while (6.45) becomes $[-\sigma_1 - \mu_2, (1+\alpha)\sigma_1]$. The difference could be substantial for values of α close to unity. This hypothesis has been confirmed experimentally and thus we shall henceforth only consider intervals produced by the second method.

In order to compute intervals via this method, it is necessary to obtain the output error symbolically in terms of the individual error sources. For a filter containing only a moderate number of multipliers this task, if carried out by hand, becomes extremely time-consuming and prone to error. Fortunately, it is possible to program the entire procedure. The technique closely follows the method used in the example just presented. First, all of the sources of error due to both finite signal word length and coefficient quantization are identified and, at each point in the filter where such an error occurs, an input variable is defined. The transfer matrix from these new inputs to the adaptor (which now includes P and P^{-1}) outputs is computed. This is carried out using a modified version of the filter simulation routine in which all of the adaptor inputs are set to zero and new inputs for the error sources are included as required. Next, based upon the type of arithmetic to be used in the realization, the errors caused by finite signal word length are assigned intervals. Intervals are also computed for the error variables which correspond to coefficient quantization. Finally, the output error intervals can be computed from the transfer matrix and the input error intervals using a matrix interval arithmetic routine based upon the rules given earlier.

The input intervals for the error variables due to coefficient quantization can be obtained using a similar technique. First, using a realization which uses the exact multipliers in P , the transfer matrix from the state variables to the quantized multipliers is computed. Then, with the state variable intervals set to $[x_{\min}, x_{\max}]$, where x_{\min} and x_{\max} are the minimum and maximum signal values allowed in the delays, the intervals at the quantized multipliers can be computed. These intervals represent the range of signal values which are possible in the ideal realization. The desired error intervals are then obtained from (6.25).

Because the above computations are carried out on a finite word length computer, the error interval obtained at the adaptor outputs will itself be in error. However, since most of the calculations involve numbers which have a finite word length that is much shorter than the double precision word length of the computer, these errors are extremely small. Furthermore, since the output interval width will be increased to make the end points realizable numbers as discussed in Section 6.1, these errors are highly unlikely to cause problems.

A summary of the procedure used to compute the error intervals required to implement the signal modifications which guarantee freedom from parasitic oscillations is given in Figure 6.3.

In order to illustrate the theory presented in this section, we will again consider the filter of Example 3.2. The diagonalization matrix, $P = P_1 P_2$, given by (5.89) and (5.91), will be used. The exact multiplier values are given by

1. Simulate the filter using the exact multipliers in P .
2. Compute the transfer matrix from the state variables to the multipliers in P which are to be quantized.
3. With the state variable intervals set to $[x_{\min}, x_{\max}]$, use a matrix interval routine, based upon the interval rules a) to f) and the transfer matrix of step 2 to compute the bounds on the signals at the quantized multipliers.
4. Using (6.25), compute the intervals of the errors caused by the multiplier quantization.
5. Simulate the filter again, this time using the quantized multipliers in P .
6. Identify all sources of error due to word length reduction or coefficient quantization and insert variables into the realization at the appropriate places.
7. Compute the transfer matrix from the error variables to the adaptor outputs. (The adaptor now includes P and P^{-1} .)
8. Depending upon the type of word length reduction to be used (i.e. truncation or rounding), compute the values of the error intervals using (6.27) and (6.28) or (6.32) and (6.33).
9. Using a matrix interval routine together with the results of steps 4, 7 and 8, compute the output error intervals at the outputs of the adaptor.

Figure 6.3 Summary of the procedure for computing the error intervals required to implement the signal modification which guarantee freedom from parasitic oscillations.

$$\gamma_1 = \frac{10}{109}, \quad \gamma_2 = \frac{1}{10}, \quad \gamma_3 = \frac{1}{4}. \quad (6.46)$$

Two of these multipliers, γ_1 and γ_2 must be modified for realizability.

Two different sets of values will be considered.

$$\hat{\gamma}_1 = 2^{-4} + 2^{-5} + 2^{-8} + 2^{-9} + 2^{-12} + 2^{-13} + 2^{-16} + 2^{-17} \quad (6.47a)$$

$$\hat{\gamma}_2 = 2^{-4} + 2^{-5} - 2^{-9} - 2^{-14} + 2^{-17} \quad (6.47b)$$

which will be called the long coefficients, and

$$\hat{\gamma}_1 = 2^{-4} + 2^{-5} + 2^{-8} + 2^{-9} + 2^{-12} + 2^{-13} + 2^{-15} \quad (6.48a)$$

$$\hat{\gamma}_2 = 2^{-4} + 2^{-5} - 2^{-9} - 2^{-14} \quad (6.48b)$$

which will be called the short coefficients.

We have previously assumed that the signals in the delays have a representation as shown in (6.4) with $k = 0$. For the purposes of this example we shall further assume that the delay signal word lengths are 16 bits; thus $m=16$. The minimum and maximum values in each of the delays is therefore

$$x_{\min} = -1, \quad x_{\max} = 1 - 2^{-15}. \quad (6.49)$$

Four different cases were studied. Either rounding or truncation was used exclusively with each of the long and short coefficients. Using the procedure described in Figure 6.3, the resulting output error intervals were obtained. Table 6.1 gives the normalized output intervals, $e \in [-\sigma_1, \sigma_2]$, accurate to 4 decimal places, in each case. The results are normalized such that 1 corresponds to a one in the least significant bit. The unnormalized intervals can be obtained by multiplying by 2^{-15} .

TABLE 6.1

OUTPUT ERROR INTERVALS FOR LONG AND SHORT COEFFICIENTS OF (6.47) AND (6.48)

Long Coefficients		Short Coefficients	
Rounding	Truncation	Rounding	Truncation
[-1.5843, 1.9242]	[-0.9859, 2.5226]	[-1.8871, 2.2269]	[-1.2886, 2.8253]
[-2.6148, 2.2711]	[-2.4523, 2.4336]	[-2.9832, 2.6395]	[-2.8207, 2.8020]
[-1.7005, 1.2122]	[-0.6583, 2.2544]	[-1.7181, 1.2299]	[-0.6760, 2.2720]
[-1.4333, 1.7372]	[-0.8423, 2.3463]	[-1.7131, 2.0169]	[-1.1040, 2.6260]
[-1.6699, 2.0279]	[-1.7814, 1.9165]	[-2.0916, 2.4494]	[-2.2030, 2.3380]
[-0.9339, 0.9496]	[-0.8777, 1.0058]	[-0.9575, 0.9731]	[-0.9013, 1.0294]
[-1.0000, 0.5000]	[-0.0000, 1.5000]	[-1.0000, 0.5000]	[-0.0000, 1.5000]

As expected, the intervals obtained by the exclusive use of rounding are more symmetrical than those produced by truncation. Due to the smaller error produced by the coefficient truncation in P, the intervals for the long coefficients are generally smaller than those for the short coefficients. When the intervals are modified for use in the oscillation suppression scheme the intervals of Table 6.2 are produced. Because the maximum value which appears is 3_{10} , the corrections are in all cases limited to the two least significant bits. In addition, since the actual corrections required when the short coefficients are used are in most cases the same as those required when the long coefficients are used, the extra word length of the long coefficients is unnecessary. A realization using the short coefficients can be easily implemented on a 16 bit machine.

The filter incorporating the short coefficients, rounding to 16 bits after all multiplications and including the signal modifications required for stability, was simulated. The unit sample response became zero after 207 samples. Figure 6.4 was obtained from a 1024 point FFT of this response. The spectrum of the roundoff noise, which is clearly visible in the passband and stopband, accounts for any deviation from the ideal characteristic of Figure 3.20.

As a second example of the application of the interval analysis procedure, we again consider the same prototype filter. This time, however, a non-minimal realization as would be produced by the technique of Chapter 2 is studied. In this case 9 delays are required. Since the port reference conductance matrix is diagonal,

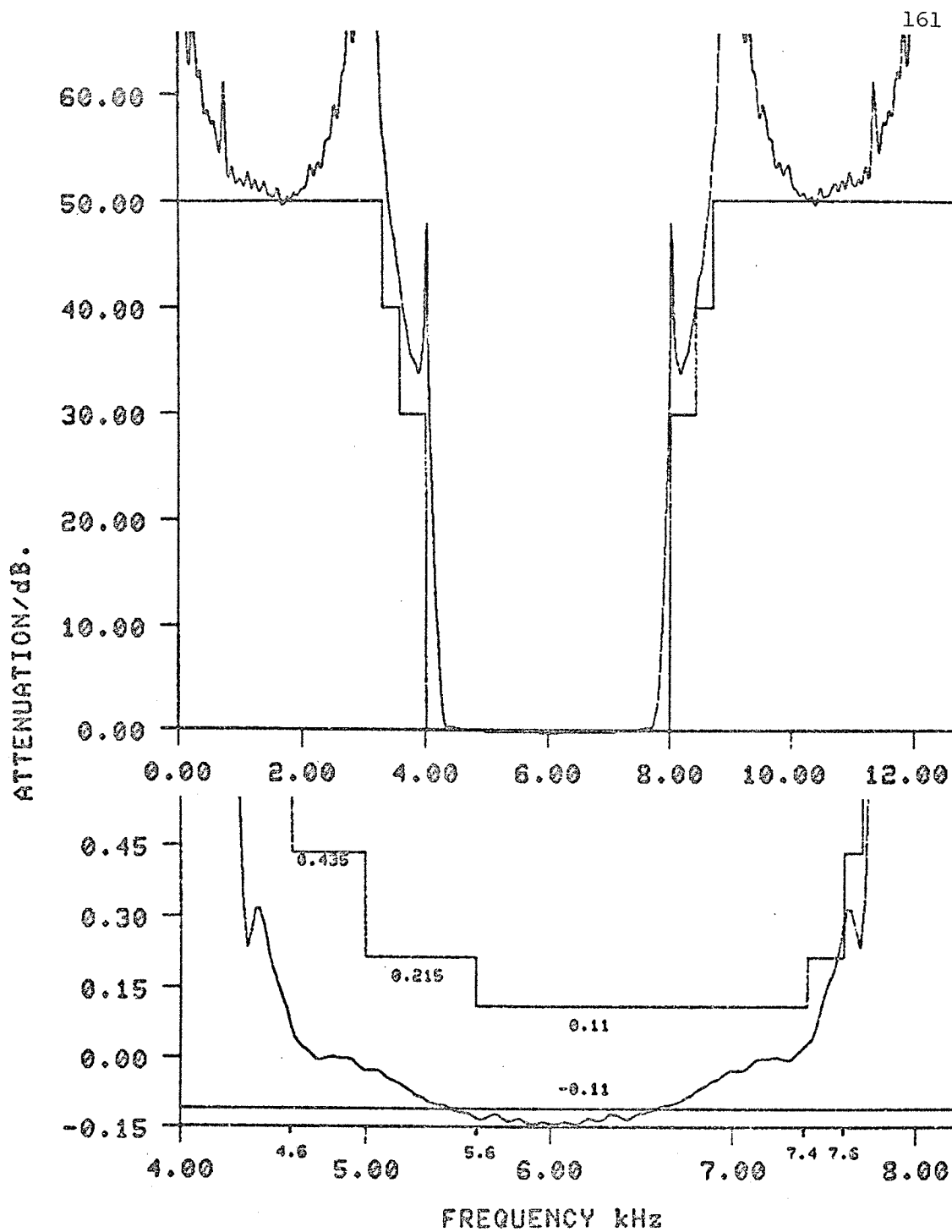


Fig. 6.4 Frequency response of the minimal wave digital realization of Fig. 3.13 after the cycle-suppressing corrections are carried out.

TABLE 6.2

REALIZABLE OUTPUT ERROR INTERVALS FOR LONG AND SHORT COEFFICIENTS OF (6.47) AND (6.48)

Long Coefficients		Short Coefficients	
Rounding	Truncation	Rounding	Truncation
$[-2, 2]$	$[-1, 3]$	$[-2, 3]$	$[-2, 3]$
$[-3, 3]$	$[-3, 3]$	$[-3, 3]$	$[-3, 3]$
$[-2, 2]$	$[-1, 3]$	$[-2, 2]$	$[-1, 3]$
$[-2, 2]$	$[-1, 3]$	$[-2, 3]$	$[-2, 3]$
$[-2, 3]$	$[-2, 2]$	$[-3, 3]$	$[-3, 3]$
$[-1, 1]$	$[-1, 2]$	$[-1, 1]$	$[-1, 2]$
$[-1, 1]$	$[0, 2]$	$[-1, 1]$	$[0, 2]$

no diagonalization matrix is necessary and thus there are no errors due to coefficient quantization. The error intervals obtained are shown in Table 6.3. The corrections required when rounding is used exclusively, are limited to 1_{10} or 2_{10} and are therefore restricted to the two least significant bits. A number of the intervals require corrections only in the least significant bit. The ability to obtain oscillation-free filters in this manner provides an alternative to the method used by Fettweis and Meerkötter for series-parallel adaptor realizations.

As a final note in this section, we emphasize that the interval analysis technique can be used to generate sufficient conditions for stability in nonlinear wave digital filters implemented using any form of arithmetic. The rules presented can be easily modified to include such situations as floating-point arithmetic and sign-magnitude truncation.

6.3 REMOVAL OF OVERFLOW OSCILLATIONS IN MINIMAL REALIZATIONS WITHOUT DIAGONALIZATION

Zero input parasitic oscillations caused by overflows during addition are called overflow oscillations. These oscillations are extremely undesirable in a filter since in some cases the output can oscillate between the maximum amplitude limits [1]. As we have shown, both overflow and granularity oscillations can be avoided in minimal realizations if the diagonalization matrices P and P^{-1} are included in the filter. However, in this section we shall show that it is not necessary to include P and P^{-1} in order to suppress the overflow

TABLE 6.3

OUTPUT ERROR INTERVALS FOR NON-MINIMAL REALIZATION

Rounding		Truncation	
Exact	Realizable	Exact	Realizable
[-1.2500, 1.2305]	[-2, 2]	[-0.5000, 1.9805]	[-1, 2]
[-0.7227, 0.4688]	[-1, 1]	[-0.2227, 0.9688]	[-1, 1]
[-0.6602, 1.0000]	[-1, 1]	[-0.9102, 0.7500]	[-1, 1]
[-1.7500, 1.4063]	[-2, 2]	[-0.7500, 2.4063]	[-1, 3]
[-1.1563, 0.9180]	[-2, 1]	[-0.6563, 1.4180]	[-1, 2]
[-0.6602, 1.0000]	[-1, 1]	[-0.9102, 0.7500]	[-1, 1]
[-1.1094, 0.7500]	[-2, 1]	[-0.1094, 1.7500]	[-1, 2]
[-0.8750, 0.8907]	[-1, 1]	[-0.8750, 0.8907]	[-1, 1]
[-0.7227, 0.4688]	[-1, 1]	[-0.2227, 0.9688]	[-1, 1]

oscillations. This result is of importance if the granularity oscillations have very small amplitudes or can be removed by other means. Since G_{11} is diagonally dominant, compensation for granularity in the signals based upon the diagonal component of G_{11} will reduce the system energy in many cases. This decreases the probability that granularity oscillations will appear.

In the study of overflow oscillations, the granularity of the signal and thus any roundoff or truncation effects are ignored. For the filters which we are considering, this means that only the most significant bits of the adaptor output signals need to be modified before being stored in the delays.

Using the diagonalization method of Section 5.3, the Lyapunov function, $V = x^T G_{11} x$, can be written as a sum of squares weighted with positive coefficients. Thus

$$V = y^T D y = \sum_{i=1}^m y_i^2 d_i \quad (6.50)$$

where $D = \text{diag } [d_i]$, $d_i > 0$, $i = 1, 2, \dots, m$.

Due to the form of (6.50), the conditions of Theorem 4.7 are satisfied if the overflows are removed so that

$$|y_{NL_i}(n+1)| \leq |y_{L_i}(n+1)| \quad i = 1, 2, \dots, m. \quad (6.51)$$

However, the actual signals which appear in the adaptor are the components of x and thus any operation which satisfies (6.51) must be translated into equivalent operations upon the components of x .

A scheme which produces the desired results is illustrated in the following example. We shall again consider the filter first discussed in Example 3.2. Using the inductive element values from the

prototype and the diagonalized capacitive component in (5.92), the Lyapunov function can be written as

$$V = y_1^2 + \frac{1}{3} y_2^2 + \frac{1}{3} y_3^2 + \frac{10}{3} y_4^2 + \frac{109}{30} y_5^2 + \frac{5301}{1308} y_6^2 + \frac{4}{3} y_7^2 \quad (6.52)$$

or in terms of the adaptor outputs numbered according to the edge labels of Figure 3.14

$$V = x_6^2 + \frac{1}{3} x_7^2 + \frac{1}{3} x_8^2 + \frac{10}{3} (x_9 + \frac{1}{10} x_{10} + \frac{1}{109} x_{11})^2 + \frac{109}{30} (-x_{10} - \frac{10}{109} x_{11})^2 + \frac{5301}{1308} x_{11}^2 + \frac{4}{3} (-\frac{1}{4} x_{11} + x_{12})^2. \quad (6.53)$$

If, for example, an overflow in $x_6 = y_1$ occurs, then when x_6 is set to zero, $y_{NL_1}(n+1) = 0$ while $y_{NL_i}(n+1)$, $i = 2, 3, \dots, m$ remain unchanged. Thus, (6.51) is satisfied and no overflow oscillations can occur.

Similar action is required if x_7 or x_8 overflow. If x_9 overflows, $y_4 = x_9 + \frac{1}{10} x_{10} + \frac{1}{109} x_{11}$ can be set to zero by zeroing x_9 , x_{10} and x_{11} . However, the change in these variables not only zeros y_4 , but also changes the value of y_5 , y_6 and y_7 , possibly producing a value which is larger in magnitude. This problem can be avoided by setting y_5 , y_6 and y_7 to zero by also zeroing x_{10} , x_{11} and x_{12} . Further examination shows that if any one of x_9 , x_{10} , x_{11} or x_{12} overflows, then if all four of these signals are zeroed, no overflow oscillations will appear.

The zeroing scheme for each particular filter can be determined by expressing the diagonalized Lyapunov function in the form of (6.50) and following similar steps as outlined above. In general, signals corresponding to capacitive elements which do not appear in capacitance-only loops or capacitance-only cutsets can be

zeroed independently when overflows occur, while all signals which correspond to capacitive elements in capacitive degeneracies must be zeroed simultaneously. The signals corresponding to the inductive elements are treated independently from those corresponding to the capacitive elements but in a similar manner.

CHAPTER VII

CONCLUDING REMARKS AND SUGGESTIONS FOR FUTURE WORK

A technique which simultaneously eliminates redundant delays in wave digital filters caused by loops and/or cutsets of reactive elements has been presented. This technique, which is implemented via an n -port adaptor, can be used to design filters which are canonic in both delays and multipliers from ladder prototypes. Realizations of this type having the canonic number of multipliers retain the low element sensitivity characteristic of properly designed doubly terminated LC prototypes. Such filters should also have the benefit of lower roundoff noise. The design technique can also be used to transform prototypes which do not have ladder structure. The resulting filters will be canonic in delays (ie minimal) but, in general, the network interpretation of K will not produce realizations which are canonic in multipliers. In ladder prototypes where there is no interaction between the loops and cutsets of redundant reactive elements, a canonic realization can also be obtained using Fettweis' method; however, the n -port approach introduced here requires fewer adders.

The n -port adaptor presented is shown to be both pseudo-lossless and reciprocal with respect to a nondiagonal port conductance matrix. This nondiagonal matrix is a direct consequence of the realization procedure. The controllability and observability of wave digital systems using these adaptors was investigated. In particular, it was shown that a pseudolossless reciprocal system is

controllable or observable if and only if it is minimal (ie both controllable and observable simultaneously). Various aspects of the stability of both linear and nonlinear realizations were also studied. The nondiagonal reference conductance matrix was shown to be a Lyapunov function for linear filters and a result which showed that linear pseudolossless reciprocal wave digital filters are asymptotically stable if and only if they are minimal realizations was derived. System modification schemes which can be used to guarantee the total (state) stability as well as the output stability in nonlinear realizations obtained from either stable or asymptotically stable linear filters are also given.

Implementation of the limit cycle suppressing procedure is easily implemented if a diagonal Lyapunov function is available. The possibility of obtaining such functions for minimal wave digital filters was considered. Using the properties of the eigenvalues and eigenvectors of S_{11} , necessary and sufficient conditions for the existence of alternate Lyapunov functions were derived. However, application of these conditions to various prototype structures showed that diagonal solutions exist only in a limited number of cases. Another procedure which utilizes a similarity transformation on the state variables to simultaneously diagonalize G_{11} was developed. Realizations produced by this method are no longer canonic in multipliers but remain canonic in delays. No sensitivity problems are introduced by this procedure since, as long as the transformation matrix P is a product of self-inverse matrices, the input-output

behaviour is independent of the actual multiplier values in P .

The final chapter of the thesis described a method based upon an error interval technique which can be used to implement the system modifications which guarantee freedom from parasitic oscillations. This method is applicable to both non-minimal realizations and also to minimal realizations after the introduction of the diagonalization matrices. In the latter case oscillation-free performance has been gained by essentially trading delays for multipliers. However, multipliers can be multiplexed whereas delays cannot.

The network interpretation of K produces realizations which consist of an interconnection of discrete adders and multipliers. Such realizations are desirable for software, firmware or traditional hardware designs. Distributed arithmetic, which has recently received much attention in the literature, should be considered as one possible alternate way to realize K . The sensitivity and noise performance of such realizations are of interest.

The properties studied in Chapter IV could prove useful in extracting the essential features of wave digital filters so that prototype networks are no longer necessary.

When it was established in Chapter V that diagonal Lyapunov functions for minimal wave digital filters existed only in a few special cases, a similarity transformation was introduced which simultaneously diagonalized G_{11} . It may be possible to find other

simpler transformations which do not immediately diagonalize G_{11} but which transform the system such that a diagonal Lyapunov function now exists.

The cycle suppressing scheme described in this thesis is applicable to both minimal and non-minimal realizations. The stable non-minimal realizations produced offer an alternative to the adaptor technique used by Fettweis. In Fettweis' method correction terms are added at each port of every adaptor. These corrections propagate through the filter to the delay terminated ports where the component in the signal due to the corrections can be substantial. Because Fettweis' method does not take into account the dependent nature of the errors in the various adaptors, it seems likely that the effective corrections at the delay terminated ports are greater than those which are required in the n-port technique. A thorough study of the noise performance of each type of realization should be carried out.

APPENDIX A

In order to verify that the reduced realization of Chapter III correctly describes the input-output properties of the original system, we shall now demonstrate that this realization can be obtained via a similarity transformation which simultaneously decouples the modes at $z = -1$ and $z = 1$. The effort involved in the solution of this problem is considerably reduced by the fact that the procedure already presented in Chapter III reveals the form of the transformed state variables which appear in the reduced order realization.

Equation (3.33) can be written in the form

$$Bb' = VBa' \quad (A.1)$$

where

$$V = \text{diag} \begin{bmatrix} -U & -U & -U & -U & U & U & U & U \end{bmatrix} \quad (A.2)$$

B is the coefficient matrix on the left-hand side of (3.33) and b' and a' are the variables as they appear in that equation. A sequence of nonsingular transformations can be used to obtain

$$\tilde{B}\tilde{b} = \tilde{V}\tilde{B}\tilde{a} \quad (A.3)$$

where

$$\tilde{B} = T_3 T_2 B T_1^{-1} T_3^{-1} T_4^{-1} \quad (A.4)$$

$$\tilde{V} = T_3 T_2 V T_2^{-1} T_3^{-1} \quad (A.5)$$

$$\tilde{b} = T_4 T_3 T_1 b', \quad \tilde{a} = T_4 T_3 T_1 a' \quad (A.6)$$

T_1 is given by (3.34),

$$T_2 = \left[\begin{array}{cc|cc|cc|cc} U & 0 & & & & & & \\ 0 & U & & & & & & \\ \hline & & U & 0 & & & 0 & -B_{13}R_\Gamma \\ & & -\tilde{R}_{L21} & \tilde{G}_{L11} & U & & 0 & E \\ \hline -Q_{11}G_S & 0 & & & U & 0 & & \\ F & 0 & & & -\tilde{G}_{C21} & \tilde{R}_{C11} & U & \\ \hline & & & & & & U & 0 \\ & & & & & & 0 & U \end{array} \right] \quad (A.7)$$

where

$$E = \tilde{R}_{L21} \tilde{G}_{L11} B_{13} R_\Gamma - B_{L2\Gamma} R_\Gamma$$

$$F = \tilde{G}_{C21} \tilde{R}_{C11} Q_{11} G_S - Q_{C2S} G_S,$$

$$T_3 = \left[\begin{array}{cc|cc|cc|cc} 0 & U & 0 & 0 & & & & \\ 0 & 0 & 0 & U & & & & \\ \hline & & & & 0 & U & 0 & 0 \\ & & & & 0 & 0 & U & 0 \\ \hline U & 0 & 0 & 0 & & & & \\ 0 & 0 & U & 0 & & & & \\ \hline & & & & U & 0 & 0 & 0 \\ & & & & 0 & 0 & 0 & U \end{array} \right] \quad (A.8)$$

and

$$T_4 = \left[\begin{array}{cccc|cccc} R_R & & & & & & & \\ & R_x & & & & & & \\ & & U & & & & & \\ & & & U & & & & \\ \hline 0 & 0 & B_{SC2} & 0 & R_S & 0 & -Q_{11}^T & 0 \\ 0 & \tilde{R}_{L12} & 0 & 0 & 0 & \tilde{R}_{L11} & 0 & 0 \\ 0 & 0 & \tilde{G}_{C12} & 0 & 0 & 0 & \tilde{G}_{C11} & 0 \\ 0 & Q_{\Gamma L2} & 0 & 0 & 0 & -B_{12}^T & 0 & G_{\Gamma} \end{array} \right] \quad (A.9)$$

All of the matrices which appear in T_2 , T_3 and T_4 have been defined previously in Chapter III.

Evaluation of \tilde{B} in (A.4) yields a partitioned form

$$\tilde{B} = \text{diag} \begin{bmatrix} C & U \end{bmatrix} \quad (A.10)$$

where C is identical to the coefficient matrix on the left-hand side of (3.45). \tilde{V} is not diagonal. Inversion of \tilde{B} using the same technique previously used for the solution of (3.45), followed by the insertion of the dynamic port terminations produces a set of state equations

$$\begin{bmatrix} \Sigma a_1(n+1) \\ \Sigma a_3(n+1) \\ b_2 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{\alpha} & S_{12} \\ 0 & S_{\beta} & 0 \\ S_{21} & S_{\gamma} & S_{22} \end{bmatrix} \begin{bmatrix} a_1 \\ a_3 \\ a_2 \end{bmatrix} \quad (A.11)$$

The variables a_1 , a_2 and b_2 correspond directly to those in (3.46), while the submatrix S

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \quad (A.12)$$

is the scattering matrix of (3.46). The submatrices S_α , S_β and S_γ are in general nonzero. S_β , which is of particular interest, is given by

$$S_\beta = \begin{bmatrix} -U & 0 & 0 & 0 \\ 0 & -U & 0 & -2B_{13}^R T \\ 2Q_{11}^G S & 0 & U & 0 \\ 0 & 0 & 0 & U \end{bmatrix}. \quad (A.13)$$

Equation (A.11) can be viewed as the result of a similarity transformation using a transformation matrix $P = (T_4 T_3 T_1)^{-1}$. The form of (A.11) clearly shows that the state variables a_3 are decoupled from the remainder of the system. These variables are uncontrollable and can therefore be simply discarded without altering the transfer function. In fact, since these variables are redundant, they can also be arbitrarily assigned a value of zero. This step produces the constraint equations of (3.26) - (3.29). The interpretation of these constraint equations as well as the network interpretation of $2K$ can then be developed in a manner identical to that already used in Chapter III.

Further examination of S_β shows that an additional transformation

$$\tilde{S}_\beta = X^{-1} S_\beta X$$

where

$$X = \begin{bmatrix} U & 0 & 0 & 0 \\ 0 & U & 0 & B_{13}^{R_\Gamma} \\ Q_{12}^{G_S} & 0 & U & 0 \\ 0 & 0 & 0 & U \end{bmatrix}$$

produces

$$S_\beta = \text{diag} \begin{bmatrix} -U & -U & U & U \end{bmatrix}$$

which proves that the decoupled eigenvalues are at $z = -1$ and $z = +1$.

BIBLIOGRAPHY

- [1] A. V. Oppenheim and R. W. Schaffer, Digital Signal Processing. New Jersey: Prentice-Hall, 1975.
- [2] L. R. Rabiner and B. Gold, Theory and Application of Digital Signal Processing. New Jersey: Prentice-Hall, 1975.
- [3] L. R. Rabiner and C. M. Rader, Eds., Digital Signal Processing. New York: IEEE Press, 1972.
- [4] Digital Signal Processing Committee, Eds., Selected Papers in Digital Signal Processing, II. New York: IEEE Press, 1976.
- [5] A. Fettweis, "Realizability of digital filter networks," Arch. Elek. Übertragung, vol. 30, pp. 90-96, Feb. 1976.
- [6] _____, "Digital filter structures related to classical filter networks," Arch. Elek. Übertragung, vol. 25, pp. 79-89, Feb. 1971.
- [7] _____, "Some principles of designing digital filters imitating classical filter structures," IEEE Trans. Circuit Theory, vol. CT-18, pp. 314-316, Mar. 1971.
- [8] _____, "Pseudopassivity, sensitivity, and stability of wave digital filters," IEEE Trans. Circuit Theory, vol. CT-19, pp. 668-673, Nov. 1972.
- [9] A. Sedlmeyer and A. Fettweis, "Digital filters with true ladder configuration," Int. J. Circuit Theory and Applications, vol. 1, pp. 5-10, Mar. 1973.
- [10] A. Fettweis and K. Meerkötter, "On adaptors for wave digital filters," IEEE Trans. Acoust., Speech, and Signal Processing, vol. ASSP-23, Dec. 1975.
- [11] W. Wegener, "Design of wave digital filters with very short coefficient word lengths," in Proc. 1976 IEEE Symp. Circuits and Systems, Munich, pp. 473-476.
- [12] R. E. Crochiere and A. V. Oppenheim, "Analysis of linear digital networks," Proc. IEEE, vol. 63, pp. 581-595, Apr. 1975.
- [13] A. Fettweis, C. J. Mandeville and C-Y. Kao, "Design of wave digital filters for communications applications," in Proc. 1975 IEEE Symp. Circuits and Systems, Boston, pp. 162-165.
- [14] A. Fettweis, "Canonic realization of ladder wave digital filters," Int. J. Circuit Theory and Applications, vol. 3, pp. 321-332, 1975.

- [15] A. Fettweis, H. Levin and A. Sedlmeyer, "Wave digital lattice filters," Int. J. Circuit Theory and Applications, vol. 2, pp. 203-211, June 1974.
- [16] R. Nouta, "The Jauman structure in wave digital filters," Int. J. Circuit Theory and Applications, vol. 2, pp. 163-174, June 1974.
- [17] _____, "Wave digital cascade synthesis," Int. J. Circuit Theory and Applications, vol. 3, pp. 231-247, Sept. 1975.
- [18] M. F. Fahmy, "Digital realization of C- and D- type sections," Int. J. Circuit Theory and Applications, vol. 3, pp. 395-402, Dec. 1975.
- [19] J. O. Scanlan and A. D. Fagan, "Wave digital C- sections," in Proc. 1976 IEEE Symp. Circuits and Systems, Munich, pp. 510-513.
- [20] H. H. Lê, "Wave digital adaptors for Brune, Darlington C and D, and twin - T sections," Ph.D. Thesis, Dept. of Elect. Eng., University of Manitoba, Aug. 1977.
- [21] H. H. Lê and G. O. Martens, "Wave digital adaptors for reciprocal second-order sections," to be published in Proc. 1978 IEEE Symp. Circuits and Systems, New York.
- [22] T. A. Claasen, W. F. G. Mecklenbräuker and J. B. H. Peek, "Effects of quantization and overflow in recursive digital filters," IEEE Trans. Acoust., Speech, and Signal Processing, vol. ASSP-24, pp. 517-529, Dec. 1976.
- [23] P. M. Ebert, J. E. Mazo and M. G. Taylor, "Overflow oscillations in digital filters," Bell Syst. Tech. J., vol. 48, pp. 2999-3020, Nov. 1969.
- [24] T. A. Claasen, W. F. Mecklenbräuker and J. B. Peek, "Second-order digital filter with only one magnitude-truncation quantiser and having practically no limit cycles," Electron. Lett., vol. 9, pp. 531-532, Nov. 1973.
- [25] A. Fettweis and K. Meerkötter, "Suppression of parasitic oscillations in wave digital filters," IEEE Trans. Circuits and Systems, vol. CAS-22, pp. 239-246, Mar. 1975; also "Correction to 'Suppression of parasitic oscillations in wave digital filters,'" IEEE Trans. Circuits and Systems, vol. CAS-22, p. 575, June 1975.
- [26] K. Meerkötter and W. Wegener, "A new second-order digital filter without parasitic oscillations," Arch. Elek. Übertragung, vol. 29, pp. 312-314, 1975.

- [27] H. J. Butterweck, "Suppression of parasitic oscillations in second-order digital filters by means of a controlled-rounding arithmetic," Arch. Elek. Übertragung, vol. 29, pp. 371-374, Sept. 1975.
- [28] G. Verkroost and H. J. Butterweck, "Suppression of parasitic oscillations in wave digital filters and related structures by means of controlled rounding," Arch. Elek. Übertragung, vol. 30, pp. 181-186, May 1976.
- [29] G. O. Martens and K. Meerkötter, "On n-port adaptors for wave digital filters with application to a bridged-tee filter," in Proc. 1976 IEEE Symp. Circuits and Systems, Munich, pp. 514-517.
- [30] A. I. Zverev, Handbook of Filter Synthesis. New York: John Wiley and Sons, 1967.
- [31] L. Weinberg, Network Analysis and Synthesis. New York: McGraw-Hill, 1962.
- [32] J. K. Zuidweg, "Every passive time-invariant linear n-port has at least one 'H matrix'," IEEE Trans. Circuit Theory, vol. CT-12, pp. 131-132, Mar. 1965.
- [33] S. Seshu and M. B. Reed, Linear Graphs and Electrical Networks. Reading, Massachusetts: Addison-Wesley, 1961.
- [34] C. T. Chen, Introduction to Linear System Theory. New York: Holt, Rinehart and Winston, 1970.
- [35] N. Balabanian and T. A. Bickart, Electrical Network Theory. New York: John Wiley and Sons, 1969.
- [36] H. Watanabe, "Approximation theory for filter-networks," IRE Trans. Circuit Theory, vol. CT-8, pp. 341-356, Sept. 1961.
- [37] A. Fettweis, "Some general properties of signal-flow networks," in Network and Signal Theory (Proc. NATO Advanced Study Institute, Bournemouth, 1972; ed. by J. K. Skwirzynski and J.O. Scanlan), pp. 48-59, London: Peter Peregrinus Ltd., 1973.
- [38] R. E. Kalman and J. E. Bertram, "Control system analysis and design via the 'second method' of Lyapunov-II discrete-time systems," Trans. ASME(J. Basic Engrg.), pp. 394-400, June 1960.
- [39] A. Gill, Linear Sequential Circuits. New York: McGraw-Hill, 1967.
- [40] F. R. Gantmacher, Matrix Theory, vol. I. New York: Chelsea Publishing Company, 1959.

- [41] W. Krelle, H.P. Kunzi and W. Oettli, Nonlinear Programming.
- [42] K. Meerkötter, private communication.
- [43] R. E. Moore, Interval Analysis. Englewood Cliffs, N.J.: Prentice-Hall, 1966.
- [44] W. K. Jenkins and B. J. Leon, "An Analysis of Quantization Error in Digital Filters Based on Interval Algebras," IEEE Trans. Circuits and Systems, vol. CAS-22, pp. 223-232, March 1975.