This paper deals with the optimal design of transmission lines and waveguides so as to minimize the reflected power due to mismatch over a frequency band. The problem is formulated on a distributed parameter control problem, and the maximum principle for distributed parameter systems is used to derive the optimal capacitance (permittivity) per unit length; this is evaluated by an algorithm which solves the resultant two-point boundary value problem. Numerical results are presented for a waveguide example.

I. Introduction

This paper deals with the application of the results of the maximum principle for distributed systems to the design of optimal lossless transmission line and waveguide coupling structures. In the case of the waveguide, the solution sought is the optimal choice of dielectric filler, i.e., the permittivity \( \varepsilon(z) \), \( z \in [-d,0] \), when \( \mu(z) = \mu_0 \); in the transmission line, we seek the optimal distribution along the line for the capacity per unit length, \( C(z) \), when the inductance per unit length is constant.

The cost functional of interest in each example will measure how effective a "match" the electromagnetic structure provides between an arbitrary source at \( z = -d \) and an arbitrary load at \( z = 0 \). Specifically, let the source be a transmitter with impedance \( Z_s(\omega) = \Re E_y + jX_y(\omega) \), with an available spectral power density \( S(\omega) \), and let the load at \( z = 0 \) be an arbitrary impedance \( Z_l(\omega) = \Re H_x + jX_h(\omega) \). Denote the reflection coefficient looking into the coupler at \( z = -d \) as \( \rho(\omega, -d) \). Then the total power reflected back into the source (not delivered into the load) is:

\[
J = \int S(\omega) |\rho(\omega, -d)|^2 \, d\omega.
\]

This is the functional which we wish to minimize by finding the optimal parameter distribution \( \varepsilon(z) \) or \( C(z) \), \( z \in [-d,0] \).

The constraint on this minimization is that \( \varepsilon(z) \) is in \( (\varepsilon_1, \varepsilon_2, ..., \varepsilon_r) \) or \( C(z) \) is in \( (C_1, C_2, ..., C_p) \), i.e., some finite discrete set of allowable values. For example, the permittivities can be made to correspond to those of conveniently available dielectric materials.

A similar transmission line problem is treated by Moyer, Wohlers and Kopp (Refs. 1, 2), but the constraints imposed are a fixed electrical line length (rather than physical line length) and \( C(z) \) is only constrained to be in some interval \( [C_{\text{MIN}} < C(z) < C_{\text{MAX}}] \). Rohrer et al, (Refs. 3, 4) consider a lossy transmission line with interval constraints as above, but, they minimize a cost functional which is the integral-squared difference between the output voltage and a desired voltage waveform.

II. Coupling Structures

Waveguide (TE\(_{1,0}\) Structure). A lossless rectangular waveguide of width \( a \) is operated above its low frequency cutoff in the TE\(_{1,0}\) mode. It is characterized by the permittivity, \( \varepsilon(z) \), and the permeability, \( \mu(z) = \mu_0 \), of the material filling it. To examine real power flow in the \( z \)-direction, we need only worry about the components \( E_y \) and \( H_x \) (Ref. 6).

The state vector is defined by:

\[
x(z, \omega) = \begin{bmatrix} [E_y]^2(z, \omega) \\ [H_x]^2(z, \omega) \\ S_R = \Re \{E_y H_x^* / 2\}, \\ \sqrt{S_1}(z, \omega) \\ \sqrt{S_R}(z, \omega) \end{bmatrix}
\]

Then

\[
\frac{3}{2z} = \begin{bmatrix} 0 & 0 & -\omega_0 & 0 \\ 0 & 0 & -\omega_0 & 0 \\ 0 & \omega & 0 & 0 \\ 0 & 0 & 0 & -\omega \end{bmatrix}
\]

where \( \omega(z, \omega) = \varepsilon(z) - \pi^2 / \mu_0 a^2 \omega^2 \). Since \( \varepsilon(z) \) is constrained to lie in some finite discrete set, the parameter \( \varepsilon(z) \), and thus \( C(z, \omega) \) will be piecewise constant over successive intervals in \( z \). The \$/S_1$ component is a constant so the above equation may be reduced to a third order piecewise-constant coefficient linear differential equation in \( z \) for each \( \omega \):
\[ \frac{\partial x(z,\omega)}{\partial z} = A(\omega,\varepsilon) \cdot x(z,\omega). \quad (4) \]

The load at \( z = 0 \) is specified by its wave impedance

\[ Z_L(\omega) = \frac{E_y(0,\omega)}{H_x(0,\omega)} = R_L(\omega) + jX_L(\omega). \quad (5) \]

Thus,

\[ \begin{align*}
\sqrt{S_I}(0,\omega) & = \sqrt{2} \cdot |E_y|^2(0,\omega) \\
\sqrt{S_R}(0,\omega) & = \sqrt{2} \cdot |H_x|^2(0,\omega)
\end{align*} \]

(6)

The wave impedance of the source is \( Z_s(\omega) = R_s(\omega) + jX_s(\omega) \), with available power density \( S_s(\omega) \):

\[ S_s(\omega) = \frac{1}{8\pi} \left[ 1, \left| S_s \right|^2, \sqrt{2} \cdot S_s, -\sqrt{2} \cdot S_s \right] \]

(7)

The reflected power density is given by:

\[ T(\omega) = \frac{1}{8\pi} \left[ 1, \left| S_s \right|^2, \sqrt{2} \cdot S_s, \sqrt{2} \cdot S_s \right] \times
\]

\[ \begin{align*}
\left| E_y \right|^2(-d,\omega) \\
\left| H_x \right|^2(-d,\omega) \\
\sqrt{S_I}(-d,\omega) \\
\sqrt{S_R}(-d,\omega)
\end{align*} \]

(8)

or

\[ T(\omega) = K'(\omega) \cdot x(-d,\omega) \quad (9) \]

II. Transmission Line (TEM Structure)

A parallel problem statement is possible for a lossless TEM (transverse electro-magnetic) structure or transmission line, which is characterized by an inductance per unit length, \( L(z) \), and a capacitance per unit length, \( C(z) \).

The choice for the state vector \( x(z,\omega) \) is:

\[ x(z,\omega) \triangleq \begin{bmatrix} |V|^2(z,\omega) \\
|I|^2(z,\omega) \\
\sqrt{S_I}(z,\omega) \\
\sqrt{S_R}(z,\omega) \end{bmatrix} \]

where

\[ S_I = \text{Im}(VI/2) \]

and

\[ S_R = \text{Re}(VI/2) \quad (10) \]

We omit here the development which parallels Eq. 3 - Eq. 9: the details may be found in Ref. 7, the thesis upon which this paper is based. In short, this system obeys a linear differential equation for each \( \omega \), and the coefficients are constant over successive intervals of \( z \); Eq. 4 is the appropriate description. The available power density and reflected power density are linear in \( x(-d,\omega) \): this parallels Eq. 7 - 9.

III. Application of the Maximum Principle

1. Introduction

In the previous section, we obtained a model for either a lossless transmission line or a lossless waveguide structure; the state differential equation in either case was of the form

\[ \frac{\partial x(z,\omega)}{\partial z} = A(\omega,\varepsilon) \cdot x(z,\omega) \quad (11) \]

where \( u \) is the control parameter (\( C \) in the transmission line and \( E \) in the waveguide). There are linear boundary conditions on the state (Eq. 6 and 7) and the cost \( J \) is a linear functional on the terminal state:

\[ J = \int \omega \cdot T(\omega) d\omega = \int K'(\omega) \cdot x(-d,\omega) d\omega \quad (12) \]

Optimal control of such distributed parameter systems is treated in detail by Wang, (Ref. 5) who develops an appropriate form of the maximum principle. An application of this maximum principle would, however, yield an algebraic condition which is generally ambiguous in \( u \), i.e., the trajectory of the state \( x(z,\omega) \) will generally correspond to a singular extremal. To avoid the computational difficulties which arise in the case of a singular extremal solution, we solve a slightly different problem which uses a modified cost functional and which does not suffer from the difficulties of singularity.

2. Modified Cost Functional

It is desirable to make a small modification to the cost functional

\[ J = \int \omega \cdot T(\omega) d\omega \] indicated in Eq. 12. In the case of the transmission line, we add a
term such that
\[ J = \int \tau(\omega) \, d\omega - \Delta \int (C - C_A)^2 \, dz. \]  
(13)
In the waveguide, the term added is
\[ -\Delta \int (C - C_A)^2 \, dz. \]  
The \( C \) term is some value of the intermediate capacity between the largest and smallest permissible capacities. Since we are concerned about reflected power, but not about \((C - C_A)^2\), \( \Delta \) is selected sufficiently small that the contribution of the \((C - C_A)^2\) term is negligible with respect to the \( \int \tau(\omega) \, d\omega \) term.

This extremely small term is a mathematical artifice that greatly simplifies consideration of the solutions that correspond to singular extremals; we would like to solve the problem in the limit as \( \Delta \to 0 \), but it is necessary to keep \( \Delta \) sufficiently large in the numerical procedures that it does not disappear in round-off and truncation errors.

3. Necessary Conditions

Define the Hamiltonian \( H(x, p, u, z) \) as:
\[ H(x, p, u, z) = \int <p(z, \omega), A(\omega, u)x(z, \omega)>d\omega - \Delta(u - u_a)^2 \]  
(14)
where \( u_a \) represents \( C_A \) or \( c_A \) and \( u \) represents \( C \) or \( z \). It is a necessary condition for optimality that there exist a costate function \( p^*(z, \omega) \) such that
\[ \frac{\partial}{\partial x} p^*(z, \omega) = - \frac{\delta H(x, p^*, u^*, z)}{\delta x} \]  
(15)

(Here the superscript * denotes an optimal quantity.)

and \( p^*(-d, \omega) = \frac{1}{\partial R_S} \begin{bmatrix} 1 \\ |Z_S|^2 \\ \sqrt{2} X_S \\ \sqrt{2} R_S \end{bmatrix} \)  
(16)

It is also necessary that the state obey the differential equation
\[ \frac{\partial}{\partial z} x^*(z, \omega) = A(\omega, u^*)x^*(z, \omega) \]  
(17)
with boundary conditions given in Eq. 6 and 7.

In accordance with the distributed maximum principle, the optimal parameter \( u^*(z) \) must maximize the Hamiltonian:
\[ \int p^*(z, \omega)A(\omega, u^*)x^*(z, \omega)d\omega - \Delta(u^* - u_a)^2 \geq \int p^*(z, \omega)A(\omega, u^*)x^*(z, \omega)d\omega - \Delta(u^* - u_a)^2 \]  
for all \((\omega \in U), (z \in [-d, 0])\).  

(18)

Since \( A \) is linear in \( u \) (Eq. 3), and \( u \) is independent of \( \omega \), \( u \) may be taken outside the first integral: this implies that \( u^* \) must maximize a quadratic function of \( u \).

\[ H(x, p, u^*, z) = \max \left( -\Delta(u - u_a)^2 + u \cdot M + \text{constant}(u) \right) \]  
(19)
where \( M = \int [2wx_j(z, \omega)p_j(z, \omega) + \omega x_1(z, \omega)p_j(z, \omega)]d\omega \)  
(20)
then
\[ H(x, p, u^*, z) = -\Delta(u^* u_a)^2 + u^* \cdot M + \text{constant}(u) \]  
(21)

The optimal value of \( u \) will then be the member of \( (u_1, u_2, \ldots, u_n) \) which is closest to \( u_a + M/\Delta \) as this choice maximizes the Hamiltonian. This maximization will usually provide a unique value of \( u^* \). However, for some values of \( M \) (this set has measure zero), two admissible \( u \)'s, say \( u_k \) and \( u_{k+1} \), \((u_1 < u_2, \ldots < u_n)\) are equally close to \( u_a + M/\Delta \). This condition, i.e.,

\[ H(x, p, u_k, z) = H(x, p, u_{k+1}, z) > H(x, p, u_j, z)V_jK, K+1 \]  
(22)
will be called the singular condition of the modified problem since Eq. 19 does not yield a unique \( u^* \). Note that if we had not used the artifice of the small \( \Delta(u - u_a)^2 \) term, whenever the integral term of Eq. 20 was zero, we would be faced with all possible controls \( u_1, u_2, \ldots, u_n \) as equally good candidates for optimality.

4. Singular Condition of the Modified Problem

In order for the singular condition of the modified problem to hold over a nonzero interval of \( z \), it is necessary that \( \frac{3M}{3z} = \frac{3^2M}{3z^2} = 0 \) over this interval.
Expansion of this second term gives an explicit solution for the $\epsilon(z)$ or $C(z)$ necessary to cause singularity. When these values do not fall in the set of admissible values, the solution cannot be singular; when they do, the ambiguity of the singular extremal has been resolved.

Because of the quadratic term modifying the cost functional, the singular problem no longer presents computational difficulties.

5. Numerical Procedure

The optimal control $u^*$ is known uniquely over $[-d,0]$ (except for a set of $z$ of measure zero) in terms of $X(z,w)$ and $p(z,w)$, so $u^*$ may be eliminated from the state and costate differential equations, Eq. 15 and 17. We now need to solve this set of 8-vector partial differential equations.

Three boundary functions are available at $z = 0$, Eq. 6. Equation 16 provides four boundary functions at $z = -d$, Eq. 7 provides one additional function at $z = -d$. We must solve this 6-state-boundary-function partial differential equation. The procedure used is to guess two functions at $z = -d$: the real and imaginary parts of the impedance presented by the coupler. Then use of Eq. 7, Eq. 16, and Eq. 6 results in a complete boundary specification of $x$ and $p$ at $z = -d$. Solve Eqs. 16 and 17 to $z = 0$. (This doesn't even require numerical integration, since 15 and 17 are piecewise-constant coefficient linear differential equations in $z$; successive multiplications by the transition matrix accomplishes this forward integration); the functions obtained at $z = 0$, say $R(w)$ and $X(w)$, still be exactly equal to $R_0(w)$ and $X_0(w)$. The structure generated above is a control optimisation problem, specifically how to best match between the given source and a load $R(w) + jX(w)$. However, since that is not the problem of interest, we consider this to be just the first step of an iterative procedure which converges in the limit to the boundary condition specified by $R_0(w)$ and $X_0(w)$. For successive steps of the iteration, the initial guess functions must be modified in some consistent way to obtain the desired convergence. The method used was to approximate the guess functions and the desired boundary functions by Tchebycheff polynomial expansions, then measure the variations in result coefficients with respect to guess function coefficients. These partial derivatives formed a sensitivity matrix, which was inverted and used for a Newton-Raphson iteration on the boundary-function coefficients. The convergence of this procedure was rather slow for poor initial guesses, (a gradient method would have been better in such cases) but convergence was quadratically fast (as expected) near the final solution. A complete digital computer program can be found in Ref. 7.

IV. Numerical Results

One problem of particular practical interest was studied extensively: the problem of obtaining a good impedance match between a narrow slot which forms one element of a phased-array receiving antenna, and a quartz-loaded waveguide having the same cross-section connected to the slot. The received power is then conducted to a stripline receiver (also having the same cross-section) by the quartz-loaded waveguide. The mismatch between the wave impedance of the free-space wave impinging on the slot and the wave impedance of a wave in the quartz-loaded waveguide is quite severe, resulting in a VSWR greater than 20:1 and over 7.5 Db of reflection loss in the example studied. The matching problem was further complicated by the fact that the receiving slot end of the waveguide presented an impedance considerably more reactive than resistive, and the impedance match was required over a fairly wide (10 percent) bandwidth from 5 Gc to 5.5 Gc.

The problem was run using the following constraints: $\epsilon_{\text{MAX}} = 3.78$ (quartz dielectric); $\epsilon_{\text{MIN}} = 9.00$ (alumina dielectric); $\epsilon_{\text{MAX}} = 1.15$ one wavelength at center frequency 5.25 Gc in quartz dielectric. The available power was assumed to be uniform at eleven equally spaced frequencies ranging from 5 Gc to 5.5 Gc, and normalized to one watt at each frequency.

An upper limit on the quadratic cost term $[\text{length} \times \Delta(\epsilon_{\text{MAX}} - \epsilon_{\text{MIN}})]^{2/2}$, where $\epsilon_{\text{MIN}} = (\epsilon_{\text{MAX}} + \epsilon_{\text{MIN}})/2$, was selected to be 5 watts, or about 4.5% of the total available power of 118 watts. The solution was then obtained for 5 equally spaced available $\epsilon(z)$ values ranging from $\epsilon_{\text{MIN}}$ to $\epsilon_{\text{MAX}}$. The solution required about 10 minutes of run time on the IBM-360 because the starting guess was considerably different from the final impedance obtained.

The $\epsilon(z)$ solution obtained is shown in Fig. 1. We note the successive quarter-electrical-wavelength sections of essentially "bang-bang" nature (between $\epsilon_{\text{MIN}}$ and $\epsilon_{\text{MAX}}$) beginning from the high impedance source at $z = -d$. At the higher right end of the waveguide ($z = 0$), near the resistive load having the characteristic impedance of a quartz-loaded waveguide, we see a quite a different behavior. Here the
The performance obtained is as follows: the VSWR has been reduced from 20:1 to 5.4:1, reducing reflection losses from 7.5 Db to 2.7 Db, a significant improvement over the performance without an impedance-matching coupler. The total cost obtained is 5.5814 watts of which 1.512 watts is true reflected power and .4294 watts is due to the quadratic cost term. The quadratic cost term took on about 85 percent of the value we specified as its upper limit, because around 3/4 of the line length demonstrated bang-bang or \[ \varepsilon_{\text{MIN}} \rightarrow \varepsilon_{\text{MAX}} \] behavior.

Solution of the problem was repeated for 2, 3, 10, 15 and 20 (equally spaced) admissible values of \( \varepsilon \), still ranging between the previous \( \varepsilon_{\text{MIN}} \), \( \varepsilon_{\text{MAX}} \). The results are shown in Fig. 2: as would be expected, permitting more \( \varepsilon \) values decreases the total cost. However, only a 3% improvement in total cost is obtained as the partition is increased from two available \( \varepsilon \) values to 20. In fact, most of that improvement is obtained in going from 2 to 3 allowable \( \varepsilon \)-values because the addition of the third (intermediate) value permits a significant reduction in the quadratic cost term over the right hand 1/4 of the line. The total cost appears to have converged by the time 20 \( \varepsilon \)-values were allowed so \( J_{\text{MIN}} \approx 5.5812 \) watts. Of this cost 5.154 watts is reflected power and .4272 watts is due to the quadratic cost term. Since the quadratic cost term was quite noticeable compared to the true reflected power cost (about 8 percent in these solutions), it was decided to repeat the solutions for an upper bound of .1 watts on the quadratic cost (QCOST).

(Five times smaller than before). The \( \varepsilon(z) \) solution for 5 \( \varepsilon \)-values, QCOST < 1 is shown in Fig. 3. Note that the "bang-bang" region of the line is unchanged from that of Fig. 1, except that the transitions are five times sharper. (A transition between \( \varepsilon_{\text{MIN}} \) and \( \varepsilon_{\text{MAX}} \) occurs in 1/5 as much distance). The righthand end of the line shows a significant change from Fig. 1; in particular, \( \varepsilon(z) \) is further away from its intermediate value \( \varepsilon_{\text{a}} \) for points \( z \) where \( \varepsilon(z) \approx \varepsilon_{\text{a}} \). This change is reasonable, since the quadratic cost term, \( \delta(\varepsilon-\varepsilon_{\text{a}})^{2}/2 \), is a "centralizing" force which tends to keep \( \varepsilon(z) \) near \( \varepsilon_{\text{a}} \). When QCOST, and thus \( \delta \), is reduced, this centralizing effect is also reduced and the solution becomes nearer to the "bang-bang" or \( \varepsilon(z) = \varepsilon_{\text{MIN}} \) or \( \varepsilon_{\text{MAX}} \) solution with the exception of points where \( \varepsilon = \varepsilon_{\text{a}} \). (Here the quadratic cost term is zero, so the scaling on this term has no effect).

So far, we have only examined changes in the total cost = reflected power + QCOST as the number of available \( \varepsilon \) values, NVALS, and the upper bound on the quadratic cost term QCOST are varied. The engineer using the solution is only interested in the variations of the reflected power, so we examine this in Fig. 4. First, we note that for every condition examined, the reflected power does not vary by more than ~13%. The reflected power for NVALS = 2 is independent of QCOST, as is \( \varepsilon(z) \), so the curves for QCOST < .5 and QCOST < .1 meet at this point. For the QCOST < .1 curve, there is no measurable change in reflected power between NVALS = 2 and NVALS = 20, making it obvious that the coupler can be constructed using only two dielectrics.

A feature that seems surprising at first is that for the QCOST < .5 curve, reflected power increases with NVALS, meaning that with more available \( \varepsilon \) values we do a worse matching job. This behavior is possible because the engineer is minimizing reflected power, but reflected power plus quadratic cost. This function is a decreasing function of NVALS. The explanation for the increasing reflected power is that the intermediate values of \( \varepsilon \) permit a noticeable decrease in the quadratic term (which was fairly large in this problem), and taking on these intermediate values of \( \varepsilon \) reduces the quadratic cost (with increasing NVALS) faster than it increases the reflected power cost. Thus, the system trades some reflected power cost to attain lower cost from the quadratic term. When the quadratic term is made smaller, as on the QCOST < .1 curve, the tendency toward a trade-off is greatly reduced and we see no change in reflected power with increasing NVALS. As QCOST < 0, the plot of reflected power versus NVALS must become a monotone decreasing curve (or at least a monotone non-increasing curve), so one could repeat the solution for NVALS = 5 with successively smaller upper bounds on QCOST to see if a reflected power cost less than 5.147 watts (that obtained for NVALS = 2) can be obtained, or until one is satisfied that the reflected power cost has converged.

Conclusions

The main conclusion possible from these examples is that performance is essentially independent of the number of available dielectrics, so construction can be done by using just two dielectrics. If better performance (less reflected power) is desired, one must relax the problem constraints by either permitting longer structure length or a greater range of \( \varepsilon \).
in the dielectric material.

Acknowledgment

This paper is based in part on a thesis submitted by the first author to the Department of Electrical Engineering, M.I.T., in partial fulfillment of the D. Sc. requirements. The research was supported in part by an NSF graduate fellowship, in part by the M.I.T. Lincoln Laboratory, and in part by NASA under research grant NGR-22-009(124).

References


FIG. 4 VARIATION OF REFLECTED POWER WITH NVALS AND QCOST