Communications in Applied Analysis **12** (2008), no. 3, 297–320

CONTROL THEORETIC FORMULATION OF CAPACITY OF DYNAMIC ELECTRO MAGNETIC CHANNELS

N. U. AHMED, F. REZAEI, AND S. LOYKA

SITE, University of Ottawa, Ottawa, Ontario, CANADA
ahmed@site.uottawa.ca, frezaei@alumni.uottawa.ca
sloyka@site.uottawa.ca

ABSTRACT. In this paper nonhomogeneous deterministic and stochastic Maxwell equations are used to rigorously formulate the capacity of electromagnetic channels such as wave guides (cavities, coaxial cables etc). Both distributed, but localized, and Dirichlet boundary data are considered as the potential input sources. We prove the existence of a source measure, satisfying certain second order constraints (equivalent to power constraints), at which the channel capacity is attained. Further, necessary and sufficient conditions for optimality are presented.

AMS (MOS) Subject Classification. 99Z00, 35Q60, 34H05, 94A40, 93Exx, 49K20

1. INTRODUCTION

Channel capacity for MIMO (multiple input multiple output) channels has been the subject of intense study in recent years. Most of the papers have been concerned with a strictly information theoretic analysis [13],[8]. On the other hand, channel capacity can be treated as an optimization problem subject to the constraint imposed by the Maxwell equations [12]. In this paper, a mathematical framework for MIMO capacity is provided using Electromagnetic constraints of the channel. It is shown that this problem can be rewritten as an optimal control problem where the control is the source measure subject to moment constraints equivalent to transmitter power constraints.

The rest of the paper is organized as follows: The current section ends after a brief list of notations. In Section 2, we present the dynamic models of electro-magnetic channels. Both distributed and boundary sources are considered, and existence and regularity properties of solutions of the dynamic systems are presented. In Section 3, communication problems are formulated. Section 4 deals with the solution of the problems proving existence of control measures from the admissible class at which channel capacity is attained. In Section 5, necessary and sufficient conditions of optimality are presented whereby a numerical algorithm can be developed for capacity computation. The paper is concluded with some comments in Section 6.

Received September 10, 2008

Some notations: Let Ξ denote an arbitrary set and $\mathcal{F}$ the Borel algebra of susbsets of the set Ξ. We call the pair $(\Xi, \mathcal{F})$ a measurable space. Let $\{\mu, \nu\}$ be any two regular Borel measures on the measurable space $(\Xi, \mathcal{F})$. We let $\mu \prec \nu$ to denote the absolute continuity of the measure μ with respect to the measure ν. The Radon-Nikodym derivative, if it exists, of μ with respect to ν is denoted by $\frac{\mu(dx)}{\nu(dx)} \equiv g(x)$, where $g \in L_1(\Xi, \nu)$.

For any pair of Banach spaces X, Y, we let $\mathcal{L}(X, Y)$ denote the space of bounded linear operators from X to Y. For any bounded open connected domain $\Omega \subset R^n$ with sufficiently smooth boundary $\partial\Omega$, $H^s(\Omega, R^m) \subset L_2(\Omega, R^m), s \geq 0$, will denote the standard Sobolev spaces of functions defined on Ω and taking values from R^m whose generalized derivatives up to order s belong to $L_2(\Omega, R^m$. Similarly $H^{-s}(\Omega, R^m), s \geq 0$, will denote the Sobolev spaces with negative exponents. These are distributions and, under some assumptions, are the topological duals of $H^s(\Omega, R^m)$. By Sobolev's embedding theorem, it is known that for $s \geq (n/2) + k$, $H^s(\Omega, R^m) \hookrightarrow C^k(\Omega, R^m)$. Thus the Dirac measure $\delta_\omega(dx)$ with mass concentrated at $\omega \in \Omega$ satisfies $b\delta_\omega \in H^{-s}(\Omega, R^m)$ for any $b \in C(\Omega, R^m)$ and $s > (n/2)$. Note that for $s \geq 0$, a continuous linear functional ℓ on $H^s(\Omega, R^m)$ has the representation

$$\ell(\varphi) = \int_\Omega (\varphi, \psi) dx$$

for some $\psi \in H^{-s}(\Omega, R^m)$.

For example, for any $f \in L_2(\Omega, R^m)$ with a_α being constants, the function ψ, given by $\psi \equiv \sum_{|\alpha| \leq s} a_\alpha D^\alpha f$, is an element of $H^{-s}(\Omega, R^m)$. Here $\alpha \equiv (\alpha_1, \alpha_2, \cdots, \alpha_n)$ stands for the multi index and $|\alpha| = \sum_{i=1}^n \alpha_i, \alpha_i \geq 0$ and $D^\alpha f$ denotes the distributional derivative of f of order $|\alpha|$. For fractional s, the Sobolev spaces are defined by use of Fourier transform.

2. CHANNEL DYNAMICS

In this section we present several models that describe the channel dynamics. The first model is assumed to satisfy homogeneous Neumann boundary condition (no leakage) with input source being a vector of current and charge density. The second model consists of homogeneous wave equation describing the electric field with nonhomogeneous Dirichlet boundary data whereby the input or source is provided.

2.1 Channel with Current and charge as input Sources. First we consider channels with homogeneous Neumann boundary condition. In this case the system is governed by a system of wave equations arising from Lorenz transformation of Maxwell's equations. The electromagnetic waves are generated by input sources such as current and charge densities and are confined in a wave guide. The electrical signals in the wave guide are governed by Maxwell's equations. Using the vector and scalar potentials denoted by (a,φ) and the Lorentz gauge, the Maxwell's equations are given by a system of wave equations:

$$\partial^2 a/\partial t^2 - (1/\mu\epsilon)\triangle a = (1/\epsilon)i,\ t \geq 0,\ \xi \in \Omega \subset R^3, \tag{2.1}$$

$$\partial^2 \varphi/\partial t^2 - (1/\mu\epsilon)\triangle \varphi = (1/\mu\epsilon)\rho, \tag{2.2}$$

where i and ρ are the sources, the first denoting the current density (vector) and the second the charge density. These are the sources that can be controlled to produce desirable field distributions inside the wave guide. The field variables $\{E, B\}$ are related to the potentials by the following equations:

$$E = -(\dot{a} + \nabla\varphi),\ B = \nabla \times a.$$

These models are useful in various fields of communication such as radar, optical fibre etc. [3] (see references therein). We let $\Omega \subset R^3$ denote an open bounded connected domain (representing the waveguide) having piecewise smooth boundary.

Define $H \equiv L_2(\Omega, R^3) \times L_2(\Omega, R) = L_2(\Omega, R^4)$, denote $y \equiv (a,\varphi)$, and define the formal differential operator C by $Cy \equiv (1/\mu\epsilon)(\triangle a, \triangle\varphi)$ and let B denote the Neumann boundary operator and set $B(a,\varphi) = 0$. This operator is simply the outward normal derivative of the arguments at every point on the boundary of the wave guide. Then introduce the operator A as follows:

$$D(A) \equiv \{y \in H : B(y) = 0 \text{ and } Cy \in H\} \subset H^2(\Omega, R^3) \times H^2(\Omega, R)$$
$$\text{and set } Az = Cz \text{ for } z \in D(A).$$

Under the given boundary condition, $-A$ is an unbounded positive self-adjoint operator in H. Define the state space as $\mathcal{H} \equiv D(\sqrt{-A}) \times H$ and the state as $z = (y, \dot{y})$. This is the energy space. Furnished with the scalar product and the associated norm as presented below,

$$(x, z)_{\mathcal{H}} = (\sqrt{-A}x_1, \sqrt{-A}z_1)_H + (x_2, z_2)_H$$
$$\| x \|_{\mathcal{H}} \equiv \left(\| \sqrt{-A}x_1 \|_H^2 + \| x_2 \|_H^2 \right)^{1/2},$$

$\mathcal{H}$ is a Hilbert space. Note that the first term represents the potential energy and the second the kinetic energy (magnetic field energy). Then we define the system operator $\mathcal{A}$ and the control operator $\mathcal{B}$ as follows

$$\mathcal{A} \equiv \begin{pmatrix} 0 & I \\ A & 0 \end{pmatrix},$$

$$\mathcal{B} \equiv \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ (1/\epsilon) & 0 & 0 & 0 \\ 0 & (1/\epsilon) & 0 & 0 \\ 0 & 0 & (1/\epsilon) & \\ 0 & 0 & 0 & (1/\mu\epsilon) \end{pmatrix}.$$

Define the input or the control vector as

$$u \equiv \begin{pmatrix} i_1 \\ i_2 \\ i_3 \\ \rho \end{pmatrix}.$$

Using these notations the system of wave equations given by (1) and (2) can be written as an abstract differential equation on the Hilbert space $\mathcal{H}$ as follows

$$\dot{z} = \mathcal{A}z + \mathcal{B}u, t \geq 0, \tag{2.3}$$

where $\mathcal{A}$ is an unbounded operator with domain and range in $\mathcal{H}$. In practice the input is localized. Let $\Omega_0 \subset \Omega$ be a part of the domain at the input end of the wave guide and consider the Hilbert space $U \equiv L_2(\Omega_0, R^4)$ with the standard scalar product. We may assume that the controls are functions of time taking values from the Hilbert space U. Thus our admissible source is a proper subset $\mathcal{U}_{ad} \subset L_2(I, U)$. It can be shown that on the Hilbert space $\mathcal{H}$ the system operator $\mathcal{A}$ is skew adjoint and hence $i\mathcal{A}$ is self adjoint. Thus it follows from semigroup theory ([1], Theorem 3.1.4, p. 71), in particular Stones theorem, that $\mathcal{A}$ generates a unitary group of operators $\mathcal{S}(t), t \in R$. Using this unitary group of operators we can write the solution (mild) of equation (2.3) as follows

$$z(t) = \mathcal{S}(t)z_0 + \int_0^t \mathcal{S}(t-s)\mathcal{B}u(s)ds, t \geq 0. \tag{2.4}$$

Note that in the absence of external input u, the system is conservative and

$$\parallel z(t) \parallel_{\mathcal{H}} = \parallel \mathcal{S}(t)z_0 \parallel_{\mathcal{H}} = \parallel z_0 \parallel_{\mathcal{H}} \ \ \forall \ t \in R.$$

This can be proved by simply scalar multiplying in $\mathcal{H}$ on either side of the equation

$$\dot{z} = \mathcal{A}z, z(0) = z_0$$

by z and integrating and noting that $(\mathcal{A}\xi, \xi)_{\mathcal{H}} = 0, \forall \ \xi \in D(\mathcal{A})$. We summarize the above results in the following theorem.

Theorem 2.1. *For every input $u \in L_2(I,U)$ and initial state $z_0 \in \mathcal{H}$, the system* (2.3) *has a unique mild solution $z \in C(I,\mathcal{H})$. Further the solution is given by the expression* (2.4). *This in turn implies that the system of wave equations* (2.1)-(2.2) *has a unique mild solution for every given initial state in the energy space and every given finite energy input.*

2.2 Channel with Dirichlet Data as Input Source. In cgs units, the Maxwell equations for electric field E and magnetic field B are given by

$$\nabla \times B = (1/c)\partial E/\partial t + (4\pi/c)i, \tag{2.5}$$

$$\nabla \times E = -(1/c)\partial B/\partial t, \tag{2.6}$$

$$\nabla \cdot E = 4\pi\rho,\ \nabla \cdot B = 0, \tag{2.7}$$

where c denotes the velocity of light and the pair $\{i,\rho\}$ denotes the current density vector and charge density respectively. Here we have used standard notations for $curl\phi \equiv \nabla \times \phi$ and $div\phi \equiv \nabla \cdot \phi$. Using the first identity of equation (2.7), the reader can easily verify that

$$\nabla \times \nabla \times E = -\triangle E + 4\pi(\nabla\rho).$$

Now applying the curl operator on either side of equation (2.6) and using equation (2.5) one can easily verify that the electric field E satisfies the following wave equation

$$\partial^2 E/\partial t^2 - c^2\triangle E = -4\pi(\partial i/\partial t + c^2\nabla\rho). \tag{2.8}$$

Since here we are interested in boundary data, we assume that both the current and charge densities are identically zero. In order to solve such equations in any bounded domain one must specify the initial and boundary conditions. Hence the complete system equation is given by

$$\partial^2 E/\partial t^2 - c^2\triangle E = 0,\ \xi \in \Omega,\ t \geq 0, \tag{2.9}$$

$$E(0,\xi) = E_0(\xi),\ \dot{E}(0,\xi) = E_1(\xi),\ \xi \in \Omega, \tag{2.10}$$

$$E(t,\xi)|_{\partial\Omega} = u(t,\xi),\ \xi \in \partial\Omega,\ t \geq 0. \tag{2.11}$$

This is a initial boundary value problem with nonhomogeneous Dirichlet boundary condition. In general the source u carries the information to be transmitted over the wave guide channel Ω.

There are two possible ways of attacking this problem. One is the semigroup approach [2] and the other is based on the principle of transposition [10].

Method A (Semigroup Approach): The first method is based on a well known technique ([2], p. 59–63) (and the references therein) whereby one can transfer the boundary data to the righthand side of the original differential equation. We write equation (2.9)–(2.11) as a system

$$\partial e/\partial t = Le, \; e(0) = e_0, \tag{2.12}$$

$$Be = Tre_1 = u, \tag{2.13}$$

where $e \equiv (E, \dot{E})$, $Tr\phi \equiv \phi|_{\partial\Omega}$, and

$$L \equiv \begin{pmatrix} 0 & I \\ c^2\triangle & 0 \end{pmatrix}.$$

Note that this is a 6×6 matrix with the elements of the first and the fourth diagonal blocks being all zero and the second block being a 3×3 identity matrix and the third diagonal block is a 3×3 diagonal matrix with the elements being the Laplacian $c^2\triangle$. To avoid introducing new notations, we use the same symbols to define the operators $A, \mathcal{A}, B_0$ by

$$A \equiv c^2 I\triangle|_{kerB} \text{ and } \mathcal{A} \equiv L|_{kerB} \text{ and } B_0 \equiv B|_{kerL}.$$

Note that the operator A is a negative self adjoint unbounded operator on $H \equiv L_2(\Omega, R^3)$. The domain of the operator $\mathcal{A}$ is given by

$$D(\mathcal{A}) = H^2(\Omega, R^3) \cap H_0^1(\Omega, R^3) \times L_2(\Omega, R^3) \subset \mathcal{H}$$

where $\mathcal{H}$ is the energy space,

$$\mathcal{H} \equiv D(\sqrt{-}A) \times H,$$

considered here as the state space. Now returning to the system model, it is not difficult to verify that the operator $\mathcal{A}$ is closed and densely defined and that for any $R \ni \lambda \neq 0$,

$$\| R(\lambda, \mathcal{A} \| \equiv \| (\lambda I - \mathcal{A})^{-1} \|_{\mathcal{H}} \leq \; (1/|\lambda|).$$

Thus by Hille-Yosida theorem ([1], Theorem 2.2.8, p. 27) $\mathcal{A}$ is the infinitesimal generator of a C_0-group $\mathcal{S}(t), t \in R$, of contractions in $\mathcal{H}$. Further, it is easy to verify that the operator $\mathcal{A}$ is skew adjoint and hence by Stones theorem ([1], Theorem 3.1.4, p. 71), it is the infinitesimal generator of a unitary group $\mathcal{S}(t), t \in R$, on $\mathcal{H}$. Our objective is to convert the initial boundary value problem (2.12)–(2.13) into a Cauchy problem (initial value problem). Define

$$\mathcal{W} \equiv H^2(\Omega, R^3) \times L_2(\Omega, R^3) \subset \mathcal{H}$$

and set $\mathcal{W}_1 \equiv KerL, \mathcal{W}_2 \equiv KerB$. For any $\lambda(\in R) \neq 0$, define $P \equiv R(\lambda, \mathcal{A})(\lambda I - L)$, and notice that $P|_{\mathcal{W}_2} = I$, the identity and that $P^2 = P$. Thus $\mathcal{W}$ admits the direct sum decomposition as follows,

$$\mathcal{W} = \mathcal{W}_1 \oplus \mathcal{W}_2.$$

Clearly $R(\lambda, \mathcal{A}) \in \mathcal{L}(\mathcal{H}, \mathcal{W}_2)$. For the source space, let U be a linear subspace of $H^{3/2}(\partial\Omega, R^3)$ carrying the structure of a Banach space such that $B_0 : \mathcal{W}_1 \longrightarrow U$

is surjective and $\Re \equiv (B_0)^{-1} \in \mathcal{L}(U, \mathcal{W}_1)$. Now going back to our original problem (2.12)–(2.13), we can rewrite the first equation in the equivalent form

$$\partial e/\partial t = \mathcal{A}e + (\Pi - (\lambda I - L))e \tag{2.14}$$

with $\Pi \equiv (\lambda I - \mathcal{A})$. For $\lambda \in \rho(\mathcal{A})$, the resolvent set of $\mathcal{A}$, the operator Π has bounded inverse giving the resolvent $R(\lambda, \mathcal{A})$. Using the direct sum decomposition, we can express the solution as the sum given by $e = e^1 + e^2, e^1 \in \mathcal{W}_1, e^2 \in \mathcal{W}_2$. Substituting this in equation (2.14), and following similar steps as presented in ([2], p. 59–62),we arrive at the following abstract Cauchy problem

$$\dot{\zeta} = \mathcal{A}\zeta + \Lambda\Re u, \zeta_0 \equiv \zeta(0) = R(\lambda, \mathcal{A})e_0, \tag{2.15}$$

$$\Lambda \equiv (I - R(\lambda, \mathcal{A})(\lambda I - L)), \tag{2.16}$$

$$e = \Pi\zeta. \tag{2.17}$$

Using the unitary group introduced above, the mild solution of the system (2.15)–(2.16)–(2.17) is given by

$$\zeta(t) = \mathcal{S}(t)\zeta_0 + \int_0^t \mathcal{S}(t-s)\Lambda\Re u(s)ds, \ t \in I, \tag{2.18}$$

$$e(t) = \Pi\zeta(t), \ t \in I. \tag{2.19}$$

Briefly this is the first method. It is clear from the expression (2.18) that, for every $u \in L_2(I, U)$ and $\zeta_0 \in D(\mathcal{A})$, $\zeta \in C(I, D(\mathcal{A}))$ and hence it follows from (2.19) that $e \in C(I, \mathcal{H})$. We collect these facts together in the following theorem.

Theorem 2.2. *For every $e_0 \in \mathcal{H}$ and $u \in L_2(I, U)$, the initial boundary value problem* (2.12)–(2.13) *has a unique mild solution $e \in C(I, \mathcal{H})$, and it is given by the expressions* (2.18) *and* (2.19).

Method B (Principle of Transposition): The second method, which admits much more general boundary data, is the method of transposition ([10], p. 231, p. 283).This method admits $L_2(\partial\Omega)$ data and, more generally, data from Sobolev spaces with negative norm like $H^{-1/2}(\partial\Omega)$. The method consists of constructing a suitable isomorphism and then transposing the isomorphism for the solution of nonhomogeneous Dirichlet problems like (2.9)–(2.11). Consider the homogeneous Dirichlet problem

$$L\psi \equiv \partial^2\psi/\partial t^2 - c^2\triangle\psi = f, \xi \in \Omega, \ t \geq 0, \tag{2.20}$$

$$\psi(T, \xi) = 0, \dot{\psi}(T, \xi) = 0, \ \xi \in \Omega, \tag{2.21}$$

$$\psi(t, \xi)|_{\partial\Omega} = 0, \xi \in \partial\Omega, \ t \geq 0, \tag{2.22}$$

for $f \in L_2(Q, R^3) \equiv L_2(I, L_2(\Omega, R^3))$. Reversing the flow of time, it follows from the results of the previous subsection that for every $f \in L_2(Q, R^3)$ this problem has a unique solution $\psi \in H^2(Q, R^3)$. Now introduce the vector space Ψ by

$$\Psi \equiv \left\{ \psi \in L_2(Q, R^3) : L\psi \in L_2(Q, R^3), \psi(T, \cdot) = 0, \dot{\psi}(T, \cdot) = 0, \psi|_{I\times\partial\Omega} = 0 \right\}$$

and furnish it with the norm topology given by

$$\parallel \psi \parallel_\Psi \equiv \parallel L\psi \parallel_{L_2(Q,R^3)} . \tag{2.23}$$

The reader can easily verify that Ψ is a normed linear space. Since L is a closed operator, it follows that Ψ is a Banach space, in fact a Hilbert space. Thus it follows from the given norm topology that L is an iosmetric isomorphism of Ψ onto $L_2(Q, R^3) \equiv L_2(I, L_2(\Omega, R^3))$. For convenience of notation we may express this fact by stating that

$$L \in Iso(\Psi, L_2(Q, R^3)).$$

This is known as the adjoint isomorphism. Transposing this isomorphism, we can settle the question of existence of solution of our original nonhomogeneous boundary value problem (2.9)–(2.11). This is stated in the following Theorem.

Theorem 2.3. *Consider the system* (2.9)–(2.11) *and suppose* $E_0 \in H^{-1/2}(\Omega, R^3)$, $E_1 \in H^{-3/2}(\Omega, R^3)$ *and* $u \in L_2(I, H^{-1/2}(\partial\Omega, R^3)) \subset H^{-1/2}(I \times \partial\Omega, R^3)$. *Then the system* (2.9)–(2.11) *has a unique solution* $E \in L_2(Q, R^3) = L_2(I, L_2(\Omega, R^3))$.

Proof Formally, scalar multiplying equation (2.9) by any $\psi \in \Psi$ and using Greens formula for integration by parts, one can easily derive the following identity

$$\begin{aligned} &\int_{I\times\Omega} (E, L\psi) d\xi dt \\ &= \int_\Omega (E_1, \psi(0)) d\xi - \int_\Omega (E_0, \dot{\psi}(0)) d\xi - c^2 \int_{I\times\partial\Omega} (\partial\psi/\partial\nu, u(t, \xi) d\sigma(\xi) dt, \end{aligned} \tag{2.24}$$

where $\partial\psi/\partial\nu$ denotes the partial derivative of ψ in the outward direction of the unit normal vector ν at any position on the boundary $\partial\Omega$ and σ denotes the surface (Lebesgue) measure on the boundary. Define the functional

$$\ell(\psi) \equiv \int_\Omega (E_1, \psi(0)) d\xi - \int_\Omega (E_0, \dot{\psi}(0)) d\xi - c^2 \int_{I\times\partial\Omega} (\partial\psi/\partial\nu, u(t, \xi)) d\sigma(\xi)\ dt. \tag{2.25}$$

Clearly this is a linear functional. Since $\psi \in H^2(Q, R^3)$, it follows from standard trace theorems for Sobolev spaces that $\psi(0) \equiv \psi(0, \xi)$, $\xi \in \Omega$, is an element of $H^{3/2}(\Omega, R^3)$, $\dot{\psi}(0) \in H^{1/2}(\Omega, R^3)$ and $\partial\psi/\partial\nu \in H^{1/2}(I \times \partial\Omega, R^3)$. Thus, for the given data $\{E_0, E_1, u\}$ with the regularities as specified in the statement of the theorem, the scalar products on the righthand side of the identity (2.25) have the correct duality pairings. Hence, we conclude that the given data determines a continuous and hence a bounded linear functional ℓ on the Banach space Ψ. Since $L \in Iso(\Psi, L_2(Q, R^3))$, this means that the composition map $(\ell o L^{-1})$ is a continuous linear functional on

$L_2(Q, R^3)$. Hence, by Riesz representation theorem, there exists an unique $E \in L_2(Q, R^3)$ such that

$$(\ell o L^{-1})(f) = (E, f)_{L_2(Q,R^3)} \ \forall \ f \in L_2(Q, R^3). \tag{2.26}$$

Since L is an isomorphism, this is equivalent to saying that

$$\ell(\psi) = (E, L\psi)_{L_2(Q,R^3)} \ \forall \ \psi \in \Psi. \tag{2.27}$$

This also verifies the validity of the formal identity (2.24) obtained by integration by parts. The uniqueness is a consequence of the fact that $L \in Iso(\Psi, L_2(Q, R^3))$. This completes the proof. $\bullet$

Remark 2.4. It is clear from the above result that our nonhomogeneous Dirichlet initial boundary value problem (2.9)–(2.11) has a unique solution $E \in L_2(Q, R^3)$ for a very general set of data form the class of generalized functions

$$\{E_0, E_1, u\} \in H^{-1/2}(\Omega, R^3) \times H^{-3/2}(\Omega, R^3) \times H^{-1/2}(I \times \partial\Omega, R^3).$$

For practical applications we may limit our data from the Hilbert spaces $L_2(\Omega, R^3) \times L_2(\Omega, R^3) \times L_2(I, L_2(\partial\Omega, R^3))$. In this case, of course, we expect our solutions to be much more regular or smooth.

Note that the data to solution map $\{E_0, E_1, u\} \longrightarrow E$, which we denote by G, is a continuous linear map from $H^{-1/2}(\Omega, R^3) \times H^{-3/2}(\Omega, R^3) \times H^{-1/2}(I \times \partial\Omega, R^3)$ to $L_2(Q, R^3) \equiv L_2(I, L_2(\Omega, R^3))$ and hence there exists a constant $K > 0$ such that

$$\begin{aligned} &\| G(E_0, E_1, u) \|_{L_2(Q,R^3)} \leq \\ &\qquad K\Big\{\| E_0 \|_{H^{-1/2}(\Omega,R^3)} + \| E_1 \|_{H^{-3/2}(\Omega,R^3)}\| + \| u \|_{H^{-1/2}(I\times\partial\Omega,R^3)}\Big\}. \end{aligned}$$

Since for $s > 0$, the embeddings $L_2(Q) \hookrightarrow H^{-s}(Q)$ are continuous, it follows from the above result that for $\{E_0, E_1, u\} \in L_2(\Omega, R^3) \times L_2(\Omega, R^3) \times L_2(I, L_2(\partial\Omega, R^3))$

$$\begin{aligned} &\| G(E_0, E_1, u) \|_{L_2(Q,R^3)} \leq \\ &\qquad \tilde{K}\Big\{\| E_0 \|_{L_2(\Omega,R^3)} + \| E_1 \|_{L_2(\Omega,R^3)}\| + \| u \|_{L_2(I,L_2(\partial\Omega,R^3))}\Big\}, \end{aligned}$$

where $\tilde{K}$ depends on K and the embedding constants

$$L_2(\Omega, R^3) \hookrightarrow H^{-1/2}(\Omega, R^3), L_2(I, L_2(\partial\Omega, R^3)) \hookrightarrow H^{-1/2}(I \times \partial\Omega, R^3).$$

In the study of communication problems we will set $E_0 = E_1 = 0$ and consider the boundary data as the input source giving $E = G(u) \in L_2(I, H)$. A complete characterization of the input-output map is given in section 3.

Remark 2.5. It is interesting to note that Theorem 2.2 dealing with the question of existence and regularity properties of solutions also provides a clue to numerical technique for solving the basic problem (2.9)–(2.11). Let $\{f_i\} \subset L_2(Q, R^3)$ be a complete orthonormal set (orthonormality is not essential though linear independence is). Then note that

$$(\ell o L^{-1})(f_i) = (E, f_i)_{L_2(Q,R^3)} \equiv c_i, i \in N.$$

These are precisely the Fourier coefficients of E with respect to the complete set $\{f_i\}$ as indicated by the righthand expression, and hence E is given by $E = \sum_{i=1}^{\infty} c_i f_i$. Further, it is clear that c_i,s are determined entirely by the data $\{E_0, E_1, u\}$ of the problem.

Remark 2.6. Comparing method A (Semigroup Approach) with method B (Principle of Transposition), it is apparent that the later admits much more general data. At least for linear initial-boundary value problems, semigroup theory seems to be less powerful.

3. FORMULATION OF COMMUNICATION PROBLEMS

3.1 Distributed Source:

Transmit End (T): First we consider the system model described by equation (2.3) with the distributed source or control space $U = L_2(\Omega_0, R^4)$, that is,

$$U \equiv \{u \in L_2(\Omega, R^4) : u(\xi) = 0, \forall\, \xi \in \Omega \setminus \Omega_0\}.$$

For application to communication problems we may simplify the source further by taking a finite number of disjoint closed subsets $\{\sigma_i\}_{i=1}^{n} \subset \Omega_0$ and consider input source to be composed of the sum

$$u \equiv \sum_{i=1}^{n} x_i(t)\varphi_i(\xi), t \in I, \xi \in \Omega_0 \tag{3.1}$$

where the functions $\varphi_i \in L_2(\Omega_0, R^4)$ vanishing outside σ_i. In other words, these functions have σ_i as their supports and $x_i \in L_2(I)$ are scalar valued functions which are the signals. These represent message signals radiated by the strategically located n-transmit antennas.

Receiver End (R): Let S_0 denote the receiving end of the wave guide. Sensors are located on this set. Again let $\{\beta_i\}_{i=1}^{m}$ be a family of disjoint closed subsets of the set S_0 where the sensors are located. These sensors are assumed to be able to measure the electric field distribution on these patches. These represent receiving antennas. In terms of the vector and scalar potential $\{a, \varphi\}$ we have already seen that the electric field vector is given by

$$E = -(\dot{a} + \nabla\varphi).$$

Hence in terms of the state variable we have

$$E \equiv -\begin{pmatrix} z_5 + \partial_1 z_4 \\ z_6 + \partial_2 z_4 \\ z_7 + \partial_3 z_4 \end{pmatrix} \equiv \Gamma z, \tag{3.2}$$

where Γ is the matrix of differential operators easily determined by the above relation. Clearly, Γ is a bounded linear operator from $\mathcal{H}$ to $L_2(\Omega, R^3)$. The outputs are the integrals of weighted sensor response to the electric field distribution on the patches. These are given by

$$y_i(t) = \int_0^t \left(\int_{\beta_i} < \alpha_i(\xi), E(s,\xi) >_{R^3} d\sigma(\xi) \right) ds + w_i(t),\ t \in I, i = 1, 2 \ldots, m, \tag{3.3}$$

where α_i is the vector of weight given to the measured electric field distribution on i-th site and w_i represents the measurement noise of this site. The weight vector α_i may be assumed to be supported on the set β_i. We assume that $\{w_i\}_{i=1}^m$ are mutually independent standard Brownian motions with W denoting the corresponding m vector Brownian motion. Throughout the paper we use $(\Xi, \mathcal{F}, \mathcal{F}_t, P)$ to denote the filtered probability space where $\mathcal{F}_t$, $t \geq 0$, is an increasing family of right continuous subsigma algebras of the sigma algebra $\mathcal{F}$. All random processes arising in this paper will be assumed to be based on this complete filtered probability space.

Now returning to our problem and using the source and the output models as described above, the state and the measurement dynamics turn out to be

$$\dot{z} = \mathcal{A}z + \mathcal{C}x, z(0) = z_0, \tag{3.4}$$

$$dy = \mathcal{G}\Gamma z dt + dW, \tag{3.5}$$

where the operators $\{\mathcal{C}, \mathcal{G}\}$ are given by

$$\mathcal{C}x \equiv \sum_{i=1}^n x_i \mathcal{B}\varphi_i, \text{ and } (\mathcal{G}_i E)(t) \equiv \int_{\beta_i} < \alpha_i(\xi), E(t,\xi) > d\sigma(\xi), i = 1, 2, \ldots, m.$$

The reader can easily verify that the operators $\mathcal{C} : R^n \longrightarrow \mathcal{H}$ and $(\mathcal{G}\Gamma) : \mathcal{H} \longrightarrow R^m$ and they are bounded linear operators. Using the semigroup $\mathcal{S}(t),\ t \geq 0$, corresponding to the operator $\mathcal{A}$, it follows from the expression (2.4), that

$$z(t) = K_t(x) \equiv \mathcal{S}(t)z_0 + \int_0^t \mathcal{S}(t-r)\mathcal{C}x(r)dr. \tag{3.6}$$

Define the composition map F with values

$$F_t(x) \equiv (\mathcal{G}\Gamma K_t)(x). \tag{3.7}$$

Clearly F is a nonanticipative (causal) operator mapping $L_2(I, R^n)$ to $L_2(I, R^m) \cap C(I, R^m)$ and, being the composition of bounded linear operators, it is also a bounded linear operator. Thus, the output equations (3.3) can be written as a linear stochastic differential equation in R^m,

$$dy = F_t(x)dt + dW, y(0) = 0, \ t \in I. \tag{3.8}$$

In case of method B, the map F is given by the composition map $F_t(x) = (\mathcal{G}G\mathcal{C})_t(x)$, which is a bounded linear operator from $L_2(I, R^n)$ to $L_2(I, R^m)$.

3.2 Boundary Source:

Next we consider the model described by the boundary value problem (2.9)-(2.11). Here we consider a part $\partial\Omega_0 \subset \partial\Omega$ of the boundary $\partial\Omega$ where the source is active. Then for the source space we take $L_2(I, U)$ where U is a closed linear subspace of $H^{3/2}(\partial\Omega_0, R^3) \equiv \{\varphi \in H^{3/2}(\partial\Omega, R^3) : \varphi(\xi) = 0 \text{ for } \xi \notin \partial\Omega_0\}$. Again we let $\{\sigma_i\}$ denote a family of disjoint subsets of the set $\partial\Omega_0$ and model the input as

$$u(t,\xi) \equiv \sum_{i=1}^{n} x_i(t)\psi_i(\xi), t \in I, \xi \in \partial\Omega_0 \tag{3.9}$$

where $\psi_i \in U$ and supported on the set σ_i and $x_i \in L_2(I)$. The complete system model is then given by

$$\partial^2 E/\partial t^2 - c^2 \triangle E = 0, \ \xi \in \Omega, \ t \geq 0, \tag{3.10}$$

$$E(0,\xi) = 0, \dot{E}(0,\xi) = 0, \ \xi \in \Omega, \tag{3.11}$$

$$E(t,\xi)|_{\partial\Omega} = u(t,\xi) = \sum_{i=1}^{n} x_i(t)\psi_i(\xi), \ \xi \in \partial\Omega, \ t \geq 0. \tag{3.12}$$

For this model, we can use the representations (2.18)-(2.19) and (3.12) to construct the output equation. Define the map

$$K_t(x) \equiv \Pi\left(\mathcal{S}(t)\zeta_0 + \int_0^t \mathcal{S}(t-r)\Lambda\Re\mathcal{C}x(r)dr\right), \ t \in I. \tag{3.13}$$

Let Γ denote the projection map $\Gamma e \equiv e_1$ which projects $e \equiv (e_1, e_2) = (E, \dot{E})$ to the first component $e_1 = E$. Using these maps, again we can write the output equation in the same general form (3.8),

$$\begin{aligned} dy &= (\mathcal{G}\Gamma K_t)(x)dt + dW, \\ &= F_t(x)dt + dW, \ t \in I. \end{aligned} \tag{3.14}$$

The existence of the map F is assured by the expressions (2.18)-(2.19) as presented in the semigroup approach (method A).

3.3 Noisy Source. So far we have assumed that the source is noise free. In order to admit noisy source one must add some compatible additional terms to the evolution equations (2.3) and (2.15). For the distributed source, we replace the evolution equation (2.3) by the stochastic differential equation

$$dz = \mathcal{A}zdt + \mathcal{B}udt + \sigma dW^o, \ t \geq 0, \tag{3.15}$$

on the Hilbert space $\mathcal{H}$, where the operator σ is given by a 8×4 matrix of operators with the first four rows being all zero and the remaining 4×4 matrix is a diagonal matrix of operators $\{\sigma_i, i = 1, 2, 3, 4\}$. The Brownian motion W^o is given by the vector $W^o \equiv \mathrm{col}\{W_1^o, W_2^o, W_3^o, W_4^o\}$ of independent Brownian motions

$$W_i^o(t, \cdot) \equiv \{W_i^o(t, \xi), \ \xi \in \Omega\}, \ i = 1, 2, 3, 4,$$

each taking values possibly from $L_2(\Omega, R)$. In reference to the field equations (2.1) and (2.2), this means adding distributed white noise on the righthand side of each of the equations in the form $\sigma_i \dot{W}_i^o(t) \equiv \sigma_i(\cdot)\dot{W}_i^o(t, \cdot), \ i = 1, 2, 3, 4$. This model allows one to deal with localized as well as distributed noise around the wave guide. Letting V denote any separable Hilbert space, for example a closed linear subspace of $L_2(\Omega, R^4)$, we may assume W^o to be a V valued Brownian motion, independent of the Brownian motion W (receiver noise), with covariance operator denoted by Q^o and $\sigma \in \mathcal{L}(V, \mathcal{H})$ so that $Q_\sigma^o \equiv \sigma Q^o \sigma^*$ is a positive nuclear operator in $\mathcal{H}$. A natural choice for the space V is $L_2(\Omega_0, R^4) \equiv U$, same as the source space, and $\sigma = \mathcal{B}$. In any case, this choice is determined primarily by physical requirements and mathematical simplicities. Since $\mathcal{S}(t), \ t \in R$, is a unitary group, it is easy to verify that $\mathcal{S}(t)Q_\sigma^o S^*(t)$ is a positive nuclear operator in $\mathcal{H}$ for all $t \in R$ given that Q_σ^o is. Thus, for each $u \in L_2(I, U)$, equation (3.15) has a unique mild solution z, which belongs to $C(I, \mathcal{H})$ with probability one, possessing bounded second moments. In this case the map $K_t(x), \ t \geq 0$, is given by

$$K_t(x) \equiv z(t) = \mathcal{S}(t)z_0 + \int_0^t \mathcal{S}(t-r)\mathcal{C}x(r)dr + \int_0^t \mathcal{S}(t-r)\sigma dW^o. \tag{3.16}$$

Considering the boundary value problems, for noisy boundary source, equation (2.15) is replaced by

$$d\zeta = (\mathcal{A}\zeta + \Lambda\Re u)dt + \Lambda\Re\sigma dW^o, \ t \geq 0. \tag{3.17}$$

For the boundary source we had chosen $U \subset H^{3/2}(\partial\Omega_0, R^3)$. Thus, it is necessary that $\sigma \in \mathcal{L}(V, U)$ where V is any separable Hilbert space supporting the Brownian motion W^o. For example, $V \equiv L_2(\partial\Omega_0, R^3)$ or any closed linear subspace thereof. Again for the existence of mild solutions $e \equiv \Pi\zeta \in C(I, \mathcal{H})$, it suffices if $\Lambda\Re\sigma Q^o\sigma^*\Re^*\Lambda^*$ is a positive nuclear operator in $\mathcal{H}$. In this case, equation (3.13) is replaced; and the process $K_t(x),\ t \geq 0$, is given by

$$K_t(x) \equiv \Pi\bigg(\mathcal{S}(t)\zeta_0 + \int_0^t \mathcal{S}(t-r)\Lambda\Re\mathcal{C}x(r)dr + \int_0^t \mathcal{S}(t-r)\Lambda\Re\sigma dW^o\bigg), t \in I. \quad (3.18)$$

Remark 3.1. In case the sensors (receiving antennas) are nonlinear, the operator $\mathcal{G}$ is nonlinear and hence the composition map $F_t, t \in I$, is also nonlinear. The results presented in this paper remain valid provided this nonlinearity is uniformly Lipschitz having at most linear growth.

4. CHANNEL CAPACITY

In view of the preceding discussions, we notice that the input and output spaces are given by $X \equiv L_2(I, R^n)$ and $Y \equiv C(I, R^m)$. Suppose these spaces are furnished with the (topological) Borel algebra turning them into measurable spaces (X, B_X) and (Y, B_Y). Let $M(X)$ and $M(Y)$ denote the space of Borel probability measures on (X, B_X) and (Y, B_Y), respectively. Considering the source space, let $M_2(X) \subset M(X)$ denote the class of probability measures having finite second moments, that is,

$$\mu \in M_2(X) \text{ if and only if } \int_X |x|_X^2\ \mu(dx) < \infty.$$

Since normally the source power is limited, we consider a bounded subset of $M_2(X)$ given by

$$S_r \equiv \{\mu \in M_2(X) : \int_X |x|_X^2\ \mu(dx) \leq rT\}$$

where $r > 0$ is the power constraint and T is the length of the time interval I denoting the duration of the message source. For admissible source measures, we can choose any set $\mathcal{M}_{ad}$ that is a weakly compact and convex subset of the set S_r. For example, $\mathcal{M}_{ad} = wc\ell(\Gamma_r)$ where Γ_r is any convex subset of the set S_r satisfying

$$\lim_{n\to\infty} \sup_{\mu\in\Gamma_r} \int_X \sum_{i\geq n} (x, e_i)^2 \mu(dx) = 0$$

for any orthonormal basis $\{e_i\}$ of the Hilbert space X. Under this assumption, the set Γ_r is uniformly tight and hence conditionally weakly compact. Thus, its weak closure is weakly compact. Necessary and sufficient conditions for weak compactness of subsets of $M_2(X)$ can be found in [[9],Theorem 2, p 377]. For more concrete examples of compact sets $\mathcal{M}_{ad}$, see Remark 4.2 following Theorem 4.1.

Considering the output space (Y, B_Y), let $M(Y)$ denote the space of regular Borel probability measures on it. Let $M(X \times Y)$ denote the space of joint Borel probability measures on the product sigma algebra $B_X \times B_Y$. We have seen in section 3, that the output signal y is related to the input process x through the communication system (3.4)-(3.5) leading to (3.8) for the distributed source; and (3.10)-(3.12) leading to (3.14) for the Dirichlet source. In other words, for a given probability measure $\mu \in S_r \subset M_2(X)$ on the input space, there is a unique measure $\nu \in M(Y)$ on the output space Y induced by the channel. Let $\gamma \in M(X \times Y)$ denote the joint probability measure and $\mu \times \nu$ the product measure with μ and ν being the marginals of γ. The relative entropy of γ with respect to the product measure $\mu \times \nu$, denoted by $I(\mathcal{X}, \mathcal{Y})$, is called the mutual information which is a measure of the amount of information carried by the observable noisy output $\mathcal{Y}$ about the input message (source) $\mathcal{X}$. This is given by the following expression,

$$I(\mathcal{X}, \mathcal{Y}) \equiv \int_{X \times Y} \log\left(\frac{\gamma(dx \times dy)}{\mu(dx) \times \nu(dy)}\right) \gamma(dx \times dy), \tag{4.1}$$

where $\Upsilon(x,y) \equiv \frac{\gamma(dx \times dy)}{\mu(dx) \times \nu(dy)}$ denotes the Radon-Nikodym derivative of γ with respect to the product measure. Clearly, this requires that γ be absolutely continuous with respect to the product probability measure $\mu \times \nu$. Note that the output measure ν is related to the input measure μ through the channel operator and it is given by

$$\nu(D) = \gamma(X \times D) = \int_X q(x, D)\mu(dx), \ \forall \ D \in B_Y \tag{4.2}$$

where

$$q(x, D) = Pr\{y \in D|x\}$$

is the conditional probability of the output y being in $D \in B_Y$ given that the input $x \in X$. This is precisely the action of the channel on the input and, as we have seen in the preceding sections, it is determined by the dynamic models of the channel. A closed form expression for this will follow shortly. Substituting the expression (4.2) into the expression (4.1) we obtain

$$I(\mathcal{X}, \mathcal{Y}) \equiv J(\mu) \equiv \int_{X \times Y} \log\left(\frac{q(x, dy)}{\int_X q(\xi, dy)\mu(d\xi)}\right) q(x, dy)\mu(dx) \tag{4.3}$$

which is a functional of the measures q and μ. Since the channel dynamics is given, this is a functional of the source measure only as indicated above. We have seen in the preceding section that, for both the distributed and the boundary sources, the output equation has the general form given by a linear stochastic differential equation in R^m,

$$dy = F_t(x)dt + dW, y(0) = 0, \tag{4.4}$$

where F is the causal (nonanticipative) map $F : L_2(I, R^n) \longrightarrow L_2(I, R^m)$ as described earlier. This is a continuous linear map. Now it follows from equation (4.4) that for every given $x \in X$, $q(x, \cdot)$ is a Gaussian measure on Y with mean trajectory given by

$$\bar{F}(x) \equiv \{\int_0^t F_s(x)ds, t \in I\} \in Y, \tag{4.5}$$

while the covariance operator Q_1 is given by

$$(Q_1\xi, \xi) \equiv \int_{I^2} (K_1(t, s)\xi(s), \xi(t))dsdt, \xi \in Y,$$

with the kernel of the operator Q_1 being $K_1(t,s) \equiv (t \wedge s)I_m, (t,s) \in I \times I$, and I_m is the identity matrix of dimension m. Thus, the Channel Kernel is given by the conditional Gaussian measure

$$q(x, D) \equiv N_G(\bar{F}(x), Q_1)(D), x \in X, D \in B_Y. \tag{4.6}$$

Since $\bar{F}$ is a continuous linear map from X to Y, it is clear that, for every $D \in B_Y$, the map $x \longrightarrow q(x, D)$ is continuous from X to the interval $[0, 1]$. Note that, by continuity here, we do not mean absolute continuity of Gaussian measures with respect to their means. That is entirely a different question and we are not concerned with this here. However, it may be interesting to note that if $q_2 \in M(Y)$ is the conditional probability measure induced by $x_2 \in X$ and $q_1 \in M(Y)$ is the one induced by $x_1 \in X$ through the output equation (4.4), then the Radon-Nikodym derivative of q_2 with respect q_1 exists and is given by $dq_2 = gdq_1$ where g is given by

$$g \equiv E\Big\{\exp\{\int_I (F_t(x_2) - F_t(x_1), dW) - (1/2)\int_I |F_t(x_2) - F_t(x_1)|^2 dt\}|\mathcal{F}^y\Big\},$$

with $\mathcal{F}^y \subset \mathcal{F}$ denoting the smallest sigma algebra induced by the the random processes $\{y\}$. Exchanging the roles of x_1 and x_2, it is clear from this expression that q_1 and q_2 are actually equivalent measures on Y. Since $x \longrightarrow \bar{F}(x)$ is a continuous map from X to Y, it follows from the above expression for RND that $q_2 \xrightarrow{w} q_1$ (weakly) in $M(Y)$ as $x_2 \longrightarrow x_1$ in X. Thus, $x \longrightarrow q(x, \cdot)$ is also continuous in this sense. Returning to our problem and using the expression (4.6) in the expression for the mutual information given by (4.3) we obtain the following equivalent expression,

$$J(\mu) \equiv \int_{X\times Y} log\left(\frac{N_G(\bar{F}(x), Q_1)(dy)}{\int_X N_G(\bar{F}(\xi), Q_1)(dy)\mu(d\xi)}\right) N_G(\bar{F}(x), Q_1)(dy)\mu(dx), \tag{4.7}$$

which is clearly dependent on the channel operator F. Denoting the convolution

$$\int_X N_G(\bar{F}(x), Q_1)(D)\mu(dx) \equiv \nu_G(D), D \in B_Y,$$

the reader can easily verify that

$$N_G(\bar{F}(x), Q_1)(\cdot) \prec \nu_G(\cdot)$$

for μ almost all $x \in X$. Thus, the Radon-Nikodym derivative of N_G with respect to the measure ν_G exists and hence $J(\mu)$ given by (4.7) is well defined. Now our objective is to determine the capacity of the channel by maximizing the above functional over a set of admissible measures on the source space subject to power constraints. That is, we must find

$$C \equiv \sup\{J(\mu), \mu \in \mathcal{M}_{ad}\} \tag{4.8}$$

where the set $\mathcal{M}_{ad}$, as defined before, is any weakly compact convex subset of the set

$$S_r \equiv \{\mu \in M_2(X) : \int_X |x|_X^2 \ \mu(dx) \leq rT\}. \tag{4.9}$$

The first question that we must address is: does the supremum exist and, if it does, is it attained on the set $\mathcal{M}_{ad}$. Without much additional assumptions we can prove the following result.

Theorem 4.1 Suppose $\mathcal{M}_{ad}$ is a weakly compact and convex subset of the set $S_r \subset M_2(X)$. Then, there exists a unique $\mu^o \in \mathcal{M}_{ad}$ at which J attains its supremum. In other words, capacity is attained.

Proof. For the existence of the supremum on $\mathcal{M}_{ad}$, it suffices to prove that J is weakly upper semicontinuos and bounded away from $+\infty$. For uniqueness we show that J is strictly concave. First we prove that the functional $\mu \longrightarrow J(\mu)$ is concave and weakly upper semicontinuous on $M(X)$. For simplicity of notation we revert back to our original notation and set $N_G(\bar{F}(x), Q_1)(D) \equiv q(x.D), D \in B_Y$. For the first statement, we show that the functional

$$J(\mu) \equiv \int_{X\times Y} log\left(\frac{q(x,dy)}{\int_X q(\xi,dy)\mu(d\xi)}\right) q(x,dy)\mu(dx) \tag{4.10}$$

is concave. Define the measure $\tilde{\mu} \in M(Y)$ by the convolution

$$\tilde{\mu}(D) \equiv \int_X q(x,D)\mu(dx), D \in B_Y.$$

Since, for each $D \in B_Y$, $x \longrightarrow q(x,D)$ is continuous and bounded, with values from $[0,1]$, this is well defined. Choose any $v \in M(Y)$ so that $q(x,\cdot) \prec v(\cdot)$ for all $x \in X$. Such a choice is assured since $q(x,\cdot)$ is a regular Borel probability measure on B_Y induced by the output process $\{y\}$ having continuous sample paths. Clearly, $\tilde{\mu} \prec v$ also. Using this measure we can express (4.10) as the sum of two terms as follows,

$$\begin{aligned} J(\mu) \qquad &\equiv \int_{X\times Y} log\left(\frac{q(x,dy)}{\tilde{\mu}(dy)}\right) q(x,dy)\mu(dx) \\ &= \int_{X\times Y} \left\{ log\left(\frac{q(x,dy)}{v(dy)}\right) - log\left(\frac{\tilde{\mu}(dy)}{v(dy)}\right)\right\} q(x,dy)\mu(dx). \end{aligned}$$

Then using Fubini's theorem and interchanging the order of integration in the second term, we find that

$$\begin{aligned} J(\mu) &= \int_{X\times Y} \left\{ log\left(\frac{q(x,dy)}{v(dy)}\right)\right\} q(x,dy)\mu(dx) - \int_Y log\left\{\left(\frac{\tilde{\mu}(dy)}{v(dy)}\right)\right\}\tilde{\mu}(dy) \\ &= J_1(\mu) - J_2(\mu) \equiv J_1(\mu) - I_2(\tilde{\mu}). \end{aligned} \tag{4.11}$$

Consider the function

$$\eta(x) \equiv \int_Y \log\left(\frac{q(x,dy)}{v(dy)}\right) q(x,dy).$$

We have already noted that for every $D \in B_Y$, $x \longrightarrow q(x,D)$ is continuous on X with values in $[0,1]$. This implies continuity of η. Indeed, η is the uniform limit of the sequence of bounded continuous functions $\{\eta_m\}$ given by

$$\eta_m(x) \equiv \sum_{i=1}^{m} \log\left(\frac{q(x, Y_i)}{v(Y_i)}\right) q(x, Y_i)$$

where $\{Y_i\}$ is a partition of Y by pairwise disjoint members $Y_i \in B_Y$. Thus, $\mu \longrightarrow J_1(\mu)$ is linear and bounded. By definition, the second term is the relative entropy of $\tilde{\mu}$ with respect to the measure v. For an arbitrary but fixed $v \in M(Y)$, $\nu \longrightarrow I_2(\nu)$ is a strictly convex functional ([7], Lemma 1.4.3, p. 36) on the set $\{\nu \in M(Y) : I_2(\nu) < \infty\}$. This can be easily verified by use of Gibb's formula and the strict convexity of the function $\eta(\xi) \equiv \xi \log \xi$, $\xi \geq 0$. Thus I_2 is strictly convex and hence $-I_2$ is strictly concave. Combining these facts we conclude that $\mu \longrightarrow J(\mu)$ is strictly concave proving the first part of the statement. Now we consider the question of continuity. It is clear from the expression (4.11) that J_1 is bounded linear and hence continuous with respect to the weak topology. The functional J_2, or equivalently I_2, gives the relative entropy of $\tilde{\mu}$ with respect to the measure v. Again, it is well known that relative entropy is weakly lower semicontinuous ([7], Lemma 1.4.3, p. 36). Thus, J_2 is weakly lower semicontinuous and hence $-J_2$ is weakly upper semicontinuous. Hence, the functional J given by their sum is weakly upper semicontinuous proving the second part of the statement. Thus we conclude that $\mu \longrightarrow J(\mu)$ is strictly concave and weakly upper semicontinuous. Next we verify that $\sup\{J(\mu),\ \mu \in \mathcal{M}_{ad}\} < \infty$. Clearly it suffices to verify that $\sup\{J(\mu), \mu \in S_r\} < \infty$. An alternative expression for the mutual information, well known in the literature, ([6], Duncan, Theorem 2, p. 269), and ([11], Lipster-Shirayev, Theorem 16.3, p. 174) is given by

$$I(\mathcal{X}, \mathcal{Y}) = (1/2)E\left\{\int_I \left(|F_t(x)|^2_{R^m} - |\hat{F}_t(y)|^2_{R^m}\right) dt\right\}$$

where $\hat{F}_t(y) \equiv E\left\{F_t(x)|\mathcal{F}_t^y\right\}$ with $\mathcal{F}_t^y$ denoting the smallest sigma algebra with respect to which the process $\{y(s), s \leq t\}$ is measurable. This identity holds for all finite dimensional stochastic differential equations like (3.8) and (3.14) with F nonanticipative and F_t taking values from R^m. Under the assumption, $P\left\{\int_I |F_t(x)|^2_{R^m} dt < \infty\right\} = 1$, the proof is identical. We give a brief outline. Let $\gamma(dx \times dy)$ with its marginals, $\mu(dx)$ and $\nu(dy)$, be the measures as introduced earlier and let $\beta(dy)$ denote the Wiener measure on Y. For the system

$$dy = F_t(x)dt + dW, y(0) = 0,$$

it follows from absolute continuity of γ with respect to $\mu \times \beta$, and ν with respect to β, and Girsanov measure substitution that the Radon-Nikodym derivative of γ with respect to $\mu \times \beta$ is given by

$$\begin{aligned}\big(\gamma(dx \times dy)/\mu(dx) \times \beta(dy)\big)(x,y) &= \exp\Big\{\int_I (F_t(x), dy) - (1/2)\int_I |F_t(x)|^2_{R^m} dt\Big\} \\ &= \exp\Big\{(1/2)\int_I |F_t(x)|^2 dt + \int_I (F_t(x), dW)\Big\}.\end{aligned}$$

Similarly, the RND of ν with respect β is given by

$$\begin{aligned}(\nu(dy)/\beta(dy))(y) = \rho(y) &\equiv \exp\Big\{\int_I (\hat{F}_t(y), dy) - (1/2)\int_I |\hat{F}_t(y)|^2 dt\Big\} \\ &= \exp\Big\{\int_I (\hat{F}_t(y), F_t(x))dt - (1/2)\int_I |\hat{F}_t(y)|^2 dt \\ &\qquad + \int_I (\hat{F}_t(y), dW)\Big\},\end{aligned}$$

where $\hat{F}_t(y) = E\{F_t(x)|\mathcal{F}_t^y\}$. Since by assumption $P\Big\{\int_I |F_t(x)|^2_{R^m} dt < \infty\Big\} = 1$, it is clear that $\beta\{\rho(y) = 0\} = 0$ and hence $\beta\{1/\rho(y) < \infty\} = 1$. From these it follows that the RND of γ with respect to the product measure $\mu \times \nu$ is given by

$$\begin{aligned}&\big(\gamma(dx \times dy)/\mu(dx) \times \nu(dy)\big)(x,y) = \\ &= \exp\Big\{(1/2)\int_I |F_t(x) - \hat{F}_t(y)|^2_{R^m} dt + \int (F_t(x) - \hat{F}_t(y), dW)\Big\}.\end{aligned}$$

Using this expression and the definition of mutual information along with standard properties of conditional expectations, we obtain

$$\begin{aligned}&I(\mathcal{X}, \mathcal{Y}) \equiv E\Big\{\log\{(\gamma(dx \times dy)/\mu(dx) \times \nu(dy)\big)(x,y)\}\Big\} = \\ &= (1/2)E\Big\{\int_I |F_t(x)|^2 - \hat{F}_t(y)|^2 dt\Big\}.\end{aligned}$$

This ends the outline. Conditional expectation is a contraction map and so $I(\mathcal{X}, \mathcal{Y}) \geq 0$, as expected, and we have

$$J(\mu) \equiv I(\mathcal{X}, \mathcal{Y}) \leq (1/2)E\int_I |F_t(x)|^2_{R^m} dt \leq (1/2)\int_X \Big\{\int_I |F_t(x)|^2_{R^m} dt\Big\}\mu(dx).$$

Since F is a bounded linear operator from X to Y there exists a finite positive number K, independent of $t \in I$, so that $|F_t(x)|^2_{R^m} \leq K^2 |x|^2_{L_2([0,t],R^n)}, t \in I$. Hence it follows from the above inequality that for $\mu \in M_2(X)$,

$$J(\mu) \leq (K^2 T/2)\int_X |x|^2 \mu(dx) < \infty.$$

Thus on $S_r \subset M_2(X)$, we have

$$\sup_{\mu \in S_r} J(\mu) \leq (K^2 T^2 r/2) < \infty$$

proving that the functional $\mu \longrightarrow J(\mu)$ is bounded away from $+\infty$ on S_r. Since J is weakly upper semicontinuous and bounded away from $+\infty$ and, by our choice, the admissible set $\mathcal{M}_{ad} \subset S_r$ is weakly compact, J attains its supremum on $\mathcal{M}_{ad}$. This proves the existence of a $\mu^o \in \mathcal{M}_{ad}$ at which the capacity C is attained, that is, $C = J(\mu^o)$. The uniqueness follows from strict concavity of $\mu \longrightarrow J(\mu)$ and convexity of $\mathcal{M}_{ad}$. This completes the proof. •

Remark 4.1. (Some Examples of Compact sets): Here we present some simple examples of weakly compact subsets $\mathcal{M}_{ad}$ of the set S_r.

Example 4.2. Let $\{\mu_k\} \subset S_r$ be a family of distinct measures, in the sense that $\mu_k \not\equiv \mu_m$ on B_X for $k \neq m$, and Λ a subset of ℓ_1 satisfying the following properties:

$$(1) : \alpha_k \geq 0, \sum_{k=1}^{\infty} \alpha_k = 1 \text{ for } \alpha \in \Lambda,$$
$$(2) : \lim_{N\to\infty} \sum_{k\geq N} \alpha_k = 0, \text{ uniformly in } \alpha \in \Lambda.$$

Define the set $M_\Lambda \equiv \{\mu \in M(X) : \mu = \sum_{k\geq 1} \alpha_k \mu_k, \alpha \in \Lambda\}$. The reader can easily verify that $M_\Lambda \subset S_r$ and it is uniformly tight, so relatively weakly compact. Since Λ is compact, M_Λ is closed and hence it is weakly compact.

Example 4.3. A variant of this example for which the same conclusion holds is as follows. Let X_0 be a countable dense subset of the closed ball $B_a(X)$ of X of radius $a \leq rT$. Then the set $M_\Lambda(X_0) \equiv \{\mu \in M(X) : \mu = \sum_{k\geq 1} \alpha_k \delta_{x_k}, x_k \in X_0, \alpha \in \Lambda\}$ is a weakly compact subset of S_r.

Example 4.4. If Υ is any uniformly tight subset of S_r, then the set $\mathcal{M}_{ad} \equiv wc\ell(\Upsilon)$ given by the weak closure of Υ is weakly compact.

Example 4.5. Let D be a compact subset of the Hilbert space X and $M(D) = \{\mu \in M(X) : \mu(X \setminus D) = 0\}$. Clearly $M(D)$ is a weakly compact subset of $M(X)$. We may choose D such that $S_r \cap M(D) \neq \emptyset$ and then take $\mathcal{M}_{ad} \equiv S_r \cap M(D)$. Since $X = L_2(I, R^n)$, it follows from Sobolev embedding theorems that for any finite interval I and any $p \in [2, \infty)$ the embedding $W^{1,p}(I, R^n) \hookrightarrow L_2(I, R^n)$ is continuous and compact. Thus a good example of a compact set D of X is any closed bounded subset D of $W^{1,p}(I, R^n)$.

5. MAXIMIZING SOURCE MEASURE

In this section we wish to present necessary conditions that a (maximizing) measure, subject to energy constraints and determining the channel capacity, must satisfy. Such conditions are called necessary conditions of optimality. In fact, in the following result we have both necessary and sufficient conditions of optimality.

Theorem 5.1 In order for $\mu^o \in \mathcal{M}_{ad} \subset M_2(X)$ to be the optimum source measure, it is necessary and sufficient that the following inequality holds

$$\int_{X\times Y}\left\{\log\left(\frac{N_G(\bar{F}(x),Q_1)(dy)}{\int_X N_G(\bar{F}(\xi),Q_1)(dy)\mu^o(d\xi)}\right)\right\}N_G(\bar{F}(x),Q_1)(dy)\mu^o(dx)$$
$$\geq \int_{X\times Y}\left\{\log\left(\frac{N_G(\bar{F}(x),Q_1)(dy)}{\int_X N_G(\bar{F}(\xi),Q_1)(dy)\mu^o(d\xi)}\right)\right\}N_G(\bar{F}(x),Q_1)(dy)\mu(dx) \quad (5.1)$$

for all $\mu \in \mathcal{M}_{ad}$.

Proof Let $\mu^o \in \mathcal{M}_{ad}$ denote the optimizer, that is,

$$J(\mu^o) = C \equiv \sup\{J(\mu), \mu \in \mathcal{M}_{ad}\}.$$

Then for any $\mu \in \mathcal{M}_{ad}$, define $\mu^\varepsilon \equiv \mu^o + \varepsilon(\mu - \mu^o)$ for $\varepsilon \in (0,1)$. Since $\mathcal{M}_{ad}$ is convex, it is clear that $\mu^\varepsilon \in \mathcal{M}_{ad}$. Clearly μ^o being the maximal element, we have $J(\mu^o) \geq J(\mu^\varepsilon)$. Let $DJ(\mu)$ denote the Gateaux gradient of J at μ whenever it exists. Then computing the limit,

$$\lim_{\varepsilon\downarrow 0}\left\{\frac{J(\mu^\varepsilon)-J(\mu^o)}{\varepsilon}\right\} = < DJ(\mu^o), \mu-\mu^o >,$$

and noting that for optimality,

$$< DJ(\mu^o), \mu - \mu^o > \leq \ 0, \quad (5.2)$$

it is easy to verify that

$$\int_{X\times Y}\left\{1-\log\left(\frac{N_G(\bar{F}(x),Q_1)(dy)}{\int_X N_G(\bar{F}(\xi),Q_1)(dy)\mu^o(d\xi)}\right)\right\}N_G(\bar{F}(x),Q_1)(dy)(\mu-\mu^o)(dx) \geq 0,$$

$\forall\ \mu \in \mathcal{M}_{ad}$. The inequality (5.1) now easily follows from the above expression. For the sufficiency, recall that $\mu \to J(\mu)$ is concave. Hence

$$< DJ(\mu), \nu - \mu > \geq \ J(\nu) - J(\mu)\ \forall \mu, \nu \in \mathcal{M}_{ad}. \quad (5.3)$$

Taking $\mu = \mu^o$ it follows from this inequality and (5.1), or equivalently (5.2), that

$$0 \ \geq < DJ(\mu^o), \nu - \mu^o > \geq \ J(\nu) - J(\mu^o)\ \forall\ \nu \in \mathcal{M}_{ad}. \quad (5.4)$$

and hence $J(\mu^o) \geq J(\nu)\ \forall\ \nu \in \mathcal{M}_{ad}$. This proves the sufficiency of condition (5.1) thereby completing the proof. •

Remark 5.1. So far in Sections 4 and 5, we have assumed the source of electro-magnetic fields to be noise free. As seen in section 3.3, to include noisy source, we must consider the evolution equations (3.15) and (3.17) which are the stochastic versions of Maxwell (field) equations. With the source noise included, the results of Theorem 4.1 and Theorem 5.1 remain valid with the replacement of the Channel

Kernel $N_G(\bar{F}(x), Q_1)(\cdot)$ by $N_G(\bar{F}(x), Q_1 + Q_2)(\cdot)$ where Q_2 is the covariance operator associated with the source noise. For the distributed source model (3.15), the covariance operator Q_2 is given by

$$(Q_2\xi, \xi) = \int_{I \times I} < K_2(t,\tau)\xi(t), \xi(\tau) > dt d\tau \tag{5.5}$$

where the kernel K_2 is given in terms of the system parameters as follows:

$$K_2(t,\tau) = \int_0^{t \wedge \tau} dr R(t,r) R^*(\tau, r), \tag{5.6}$$

with $R(t,r), 0 \leq r \leq t \leq T$, given by

$$R(t,r) \equiv \mathcal{G}\Gamma \int_r^t \mathcal{S}(\theta - r)\sigma d\theta = \int_r^t \mathcal{G}\Gamma\mathcal{S}(\theta - r)\sigma d\theta. \tag{5.7}$$

The last identity follows from the fact that $\mathcal{G}\Gamma$ is a bounded linear operator from the state space $\mathcal{H}$ to R^m and hence closed and so commutes with the integral operation. Similar expressions can be derived for the boundary source. These conclusions are based on the properties of stochastic convolutions, see ([5], Theorem 5.2, p. 119) which use stochastic Fubini's theorem ([5], Theorem 4.18, p. 109).

Remark 5.2. (Numerical Algorithm for Computation) Based on the necessary (and sufficient) conditions of optimality as presented above, we can develop a gradient based algorithm for numerical computation. In particular, for simple sources like those of examples, (E1) and (E2), the functional $J(\mu)$ on $\mathcal{M}_{ad}$ can be redefined as being a functional on Λ and, with slight abuse of notation, we may denote it by $J(\alpha)$. Our problem is to find an $\alpha^o \in \Lambda$ at which J attains its supremum. For the source (E1), it follows from the necessary conditions that in order for $\alpha^o \in \Lambda$ to be the optimal distribution of weights assigned to the family of measures $\{\mu^k\}$, it is necessary and sufficient that we have,

$$\sum_{k \geq 1} (\alpha_k - \alpha_k^o) \int_X L_{\alpha^o}(x)\ \mu^k(dx) \geq 0,\ \forall\ \alpha \in \Lambda, \tag{5.8}$$

where

$$L_{\alpha^o}(x) \equiv \int_Y \log\Bigg(q(x,dy) / \sum_{k \geq 1} \alpha_k^o\ \nu^k(dy)\Bigg) q(x,dy),\ \ \nu^k(dy) = \int_X q(x,dy)\mu^k(dx).$$

For the source (E2), this further reduces to

$$\sum_{k \geq 1} (\alpha_k - \alpha_k^o) L_{\alpha^o}(x_k) \geq 0,\ \forall\ \alpha \in \Lambda,$$

where

$$L_{\alpha^o}(x_k) \equiv \int_Y \log\Bigg(q(x_k,dy) / \sum_{k \geq 1} \alpha_k^o\ q(x_k,dy)\Bigg) q(x_k,dy).$$

For these simple sources, one can write a gradient based numerical algorithm for finding the optimum source. Considering the first source, the gradient of J at the n-th iteration is given by

$$\triangle J(\alpha^n) = \{\triangle_k J(\alpha^n) = \int L_{\alpha^n}(x)\mu^k(dx), k = 1, 2, \cdots\}.$$

Choose

$$\alpha_k^{n+1} \equiv \alpha_k^n + \varepsilon_n \triangle_k J(\alpha^n), k \in N,$$

with $\varepsilon_n > 0$ sufficiently small so that $\alpha^{n+1} \in \Lambda$. Using this α^{n+1}, we compute the objective functional giving

$$J(\alpha^{n+1}) = J(\alpha^n) + \varepsilon \sum_{k=1}^{\infty} (\triangle_k J(\alpha^n))^2 + 0(\varepsilon).$$

For $\varepsilon_n > 0$ sufficiently small the series converges guaranteeing improvement of the objective functional at each step. Similar conclusion holds for the source (E2) with the gradient vector given by

$$\triangle J(\alpha^n) = \{\triangle_k J(\alpha^n) = L_{\alpha^n}(x_k), k = 1, 2, \cdots\}.$$

Remark 5.3. The notion of entropy and relative entropy play significant roles not only in communication problems as seen above, but also in min-max games arising in the study of control of uncertain stochastic differential systems on Hilbert spaces ([3], Ahmed and Charalambos) and also in nonlinear filtering ([4], Ahmed and Charalambos).

6. CONCLUSION

We have presented a complete dynamic model for MIMO channels (wave guides, cavities) based on deterministic as well as stochastic Maxwell's equations. To the best of our knowledge, this formulation has not been considered in the literature. Both distributed and boundary sources have been considered. Proof of existence of maximizing source measure subject to power constraints has been presented. Optimality conditions have been developed which can be used for numerical computations as indicated in Remark 5.3 for simple source spaces. The model used for sensors in the above formulation can be extended to cover linear (sensor) dynamics without any difficulty. This requires inclusion of the associated transition operator for the construction of the output map $F_t(x)$. It would be interesting to study similar problems with nonlinear sensor dynamics.

ACKNOWLEDGMENTS

(1): The first author would like to thank his colleague, Professor D. A. McNamara, for many valuable discussions on physical properties of waveguide channels. (2): Also thanks are due to National Science and Engineering Research Council of Canada for partially supporting this research under grant no A7109.

REFERENCES

[1] N. U. Ahmed, Semigroup Theory with Applications to Systems and Control, Pitman Research Notes in Mathematics Series, **246**, Longman Scientific and Technical, Co-published with John Wiley and Sons, Inc. New York, Longman Group U.K Limited, 1991.

[2] N. U. Ahmed, Optimization and Identification of Systems Governed by Evolution Equations on Banach Space, Pitman Research Notes in Mathematics Series, **184**, Longman Scientific and Technical, Co-published with John Wiley and Sons,Inc. New York, Longman Group U.K Limited, 1988.

[3] N. U. Ahmed, C. D. Charalambous, Minimax Games for Stochastic Systems subject to Relative Entropy Uncertainty: Applications to SDE's on Hilbert Spaces, J. Mathematics of Control, Signals and Systems, **19(1)**, (2007), 65–91.

[4] N. U. Ahmed, C. D. Charalambous, Optimal Measurement Strategy for Nonlinear Filtering, SIAM J. Control Optim. **45(2)**, (2006), 519–531.

[5] G. Da Prato & J. Zabczyk, Stochastic equations in Infinite Dimensions, Cambridge University Press,Cambridge, 1992.

[6] T. E. Duncan, Mutual Information for Stochastic Differential equations, *Information and Control*, **19(3)**, (1971), 265–271.

[7] P. Dupuis and R. S. Ellis, A Weak Convergence Approach to the Theory of Large Deviations, John-Wiley and Sons inc., New York, 1997.

[8] G. J. Foschini, M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communication*, **6(3)**, (1998), 311–335.

[9] I. I. Gihman and A. V. Skorohod, The Theory of Stochastc processes I, Springer-Verlag, New York Heidelberg Berlin, 1974.

[10] J.L.Lions and E. Magenes, Nonhomogeneous Boundary Value Problems and applications, **Vol. 1**, (Translated from French by P. Kenneth), Springer-Verlag New York Heidelberg Berlin, 1970.

[11] R. S. Liptser and A. N. Shiryayev, Statistics of Random Processes II: Applications, Springer-Verlag, New York Heidelberg Berlin, 1978.

[12] S. Loyka, "Multiantenna Capacities of Waveguide and Cavity Channels," *IEEE Trans. Vehicular Tech.*, **54(3)**, (2005), 863–872.

[13] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *AT&T Bell Labs. Intern. Tech. Memo.*, June 1995.