State Estimation, Wireless Tropes, Demons and Uncertainty

Christopher Rose *Fellow, IEEE*

***Abstract*—Consider observation of a system with initial state $\mathbf{x}(0)$ through some signal $r(t)$ corrupted by white noise of spectral height $\mathcal{N}_0$. When the system is cast in state-space form and the observations projected onto the relevant orthonormal bases, *completely unbidden*, two well-known wireless communications tropes emerge: a colored noise channel and a multi-access channel wherein elements of system state are associated with different "signatures" defined by the system. That is, from a mathematical perspective, the system *could* communicate to the observer in a well-understood way. Taking these tropes at face value, we investigate the efficiency of conveying a state vector $\mathbf{x}(0)$ through classical estimation to that wherein a "demon" manipulates an identical initially-at-rest system so as to communicate $\mathbf{x}(0)$ to the observer on successive epochs (channel uses). An energy constraint on the initial state $E\left[|\mathbf{x}(0)|^2\right] = \mathcal{E}$ is assumed, and the demon's signaling efforts over the ensemble of epochs are constrained similarly. In all cases, the demon conveys the $\mathbf{x}(0)$ with less error – by orders of magnitude for moderate signal to noise ratio $\frac{\mathcal{E}}{\mathcal{N}_0}$. Furthermore, the demon scenario results in some number of reliably-conveyed bits of information and imposes crisp limits on relative uncertainty of different state element estimates. In fact, the form of these limits is identical to that of the quantum mechanical Uncertainty Principle (although there is no requirement of a momentum-position analog). Nonetheless, the appearance of these tropes raises the question of whether communication and information theory have something deeper to say about physical interactions and the cacophony of system voices in conversation.**

***Index Terms*—state estimation, uncertainty principle, multiple access, colored noise**

I. INTRODUCTION

SYSTEM state estimation is a classical problems of control theory and the measure of quality is often estimator variance. Here we will see that the mathematical description of a classical physical system, when cast in a state-space/signal-space form, results in two telling communication theory tropes – *colored noise channels* and *multiaccess channels*. Under this communication theory lens, crisp limits are imposed on the amount of information that can be reliably conveyed between a sender and receiver under transmission energy constraints and these limits beg an appropriate comparison to the those imposed by classical estimation techniques. The comparison will suggests system state measurement can be viewed as listening to a cacophony of simultaneously emitted "voices" analogous to system state eigenmode mixtures. Furthermore, if the system wishes to convey the vector $\mathbf{x}^*$ where $E\left[|\mathbf{x}^*|^2\right] = \mathcal{E}$, then it is best to employ a "demon" with total energy budget $\mathcal{E}$ which sets the system state at the start of multiple sequential measurement epochs (analogous to channel uses) rather than setting the initial state of the system to $\mathbf{x}(0) = \mathbf{x}^*$ and using classical minimum mean square estimation methods.

In addition, the form of the estimation uncertainty produced by the demon is mathematically identical to that of the Uncertainty Principle – the product of the estimation specificity (variance) is a constant – although the requirement of momentum-position-like eigenmode pairs is unnecessary at the macroscale. Whether this result is simply incidental and amusing or says something deeper about the limits of what a given system can "tell" us about "itself" seems worth further exploration.

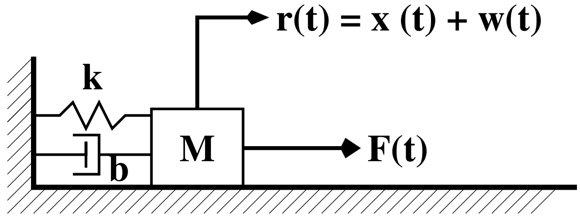

Fig. 1. Spring/Mass "Slinky"

II. MODEL DEFINITION AND ANALYSIS

Consider the Hookean spring and mass system of FIGURE 1 in which a mass, connected by a spring and a dashpot/damper (dissipative element) to a fixed point, slides on a frictionless surface. $F(t)$ is a force applied to the rightmost mass and the observable is the rightmost mass position, $r(t) = x(t) + w(t)$, where $w(t)$ is an observation noise process, assumed Gaussian, white and zero mean for simplicity. In this cartoon we assume the mass, spring, dashpot and anchor boundary never touch, so the system is linear and completely described by the following equations:

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}F(t) \tag{1}$$

$$r(t) = \mathbf{C}\mathbf{x}(t) + w(t) \tag{2}$$

In the case of FIGURE 1 we have

$$\mathbf{x} = \begin{bmatrix} x \\ \dot{x} \end{bmatrix} \tag{3}$$

with x being the position of the mass so that

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -\frac{k}{M} & -\frac{b}{M} \end{bmatrix} \tag{4}$$

$$\mathbf{B} = \begin{bmatrix} 0 \\ \frac{1}{M} \end{bmatrix} \tag{5}$$

$$\mathbf{C} = \begin{bmatrix} 1 & 0 \end{bmatrix} \tag{6}$$

Equation (1) and equation (2) comprise the usual "state-space" description of a dynamical system that can be derived from a Hamiltonian/Lagrangian or from first principles (i.e., $F = ma$ and $F = kx$) [1]–[3].

Using fundamental state-space linear systems methods we have a state transition matrix

$$\Phi(t) = e^{\mathbf{A}t} \tag{7}$$

which is composed of linearly independent functions $\{\psi_n()\}$. Assuming $\mathbf{A}$ is invertible, the $\psi_n(t)$ could be independent linear combinations of $e^{\lambda_n t}$ where the $\{\lambda_n\}$ are the eigenvalues of $\mathbf{A}$. In the undriven (also called homogeneous) case with $F(t) = 0$ and no noise $w(t)$ we then have

$$r(t) = x(t) = \mathbf{C}\mathbf{x}(t) = \mathbf{C}e^{\mathbf{A}t}\mathbf{x}_0 = \sum_{n=1}^{N} a_n(\mathbf{x}_0)\psi_n(t)$$

where $\mathbf{x}_0$ is the initial state of the system.

Now suppose we observe $r(t)$ over $[0, T]$. In general we can derive an orthonormal basis $\{\phi_n()\ \}$ on the interval $[0, T]$ by applying Gram-Schmidt [4], [5] to the $\{\psi_n(t)\}$. We can then project $r(t)$ onto the basis functions (which we will assume real for simplicity) to obtain

$$r_n = \langle r(t), \phi_n(t)\rangle \tag{8}$$

and define the matrix

$$\mathbf{G}_n = \left\langle e^{\mathbf{A}t}, \phi_n(t)\right\rangle \tag{9}$$

as the projection of the transition matrix onto basis function n. We can then define the $N \times N$ matrix $\mathbf{Q}$ as

$$\mathbf{Q} = \begin{bmatrix} \mathbf{C}\mathbf{G}_1 \\ \mathbf{C}\mathbf{G}_2 \\ \vdots \\ \mathbf{C}\mathbf{G}_N \end{bmatrix} \tag{10}$$

We then project the noise similarly

$$\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix} \tag{11}$$

where

$$w_n = \langle w(t), \phi_n(t)\rangle$$

Because the noise $w(t)$ is assumed white, the covariance matrix of $\mathbf{w}$ is

$$\mathbf{K}_{\mathbf{w}} = \mathcal{N}_0\mathbf{I} \tag{12}$$

where $\mathcal{N}_0$ is a constant and $\mathbf{I}$ is the identity matrix. In the case of FIGURE 1, $N = 2$.

The end result of projecting observations onto an orthonormal basis derived from the system eigenspace is

$$\mathbf{r} = \mathbf{Q}\mathbf{x}_0 + \mathbf{w}, \tag{13}$$

A. An Unbidden Colored Noise Channel

Suppose $\mathbf{Q}$ is invertible. We can then rewrite equation (13) as

$$\mathbf{Q}^{-1}\mathbf{r} = \mathbf{x}_0 + \mathbf{Q}^{-1}\mathbf{w},$$

and then

$$\tilde{\mathbf{r}} = \mathbf{x}_0 + \tilde{\mathbf{w}} \tag{14}$$

Equation (14) represents an additive Gaussian colored noise channel in communication theory where $\mathbf{x}_0$ would be the channel input and $\tilde{\mathbf{w}}$ is the noise vector, colored because the covariance, $\mathbf{K}_{\tilde{\mathbf{w}}}$, of $\tilde{\mathbf{w}}$ may no longer be a scaled identity matrix (unless $\mathbf{Q}$ is unitary). In the context of communication, one would use a sequence of different $\mathbf{x}_0[1], \mathbf{x}_0[2], \cdots \mathbf{x}_0[L]$ to convey information to a receiver through a sequence of received vectors $\tilde{\mathbf{r}}_1, \tilde{\mathbf{r}}_2, \cdots, \tilde{\mathbf{r}}_L$.

B. An Unbidden Multiaccess Channel

We can also rewrite equation (13) as

$$\mathbf{r} = \sum_{n=1}^{N} \mathbf{q}_n(\mathbf{x}_0)_n + \mathbf{w}, \tag{15}$$

where $\mathbf{q}_n$ is the n^{th} column of $\mathbf{Q}$. Equation (15) represents a *multiaccess* communication channel with user "symbols" $(\mathbf{x}_0)_n$ and corresponding "signatures" $\mathbf{q}_n$. The relationships between the signatures determine how strongly "users" (the $\{(\mathbf{x}_0)_n\}$) interfere with one another at the receiver. Thus, the user signatures will be received with potentially different strengths and how well a given user (eigenmode corresponding to $(\mathbf{x}_0)_n$) is "heard" at the receiver depends on its average energy relative the others and the relationships between the signatures.

Once again, in the context of communication, one would use a sequence of different $\mathbf{x}_0[1], \mathbf{x}_0[2], \cdots \mathbf{x}_0[L]$ to convey information to a receiver through a sequence of received vectors $\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_L$.

III. MMSE vs. "Demon"-Mediated Estimation

For comparison to the optimal MMSE estimator we will posit a "demon" that uses the channels articulated in equation (14) and equation (15) some number of times L with allowable energy $\mathcal{E}/L$ per channel use on average. Note that the system of FIGURE 1 is lossy owing to the damper/dashpot b, so we can assume sufficiently long epochs $[0, T]$ wherein initial rest is assured at the start of the next channel use. Therefore, the energy input into the channel by the demon during channel use k is $E\left[|\mathbf{x}(k)|^2\right] \leq \mathcal{E}/L$ where $\mathbf{x}(k)$ is the initial state set (slowly to as avoid dissipative loss) by the demon. The total energy used in communication is $\mathcal{E}$. We then assume that in the MMSE case, the initial state is set (again, slowly) to $\mathbf{x}(0)$ where $E\left|\mathbf{x}(0)|^2\right] = \mathcal{E}$. Thus, the total MMSE and demon communication energy budgets are identical.

A. The MMSE Estimator

Suppose we wish to estimate $\mathbf{x}(0)$ under a minimum mean square error (MMSE) criterion. Then the optimal MMSE estimator is the conditional mean [6]

$$\hat{\mathbf{x}}_0(r) = E_{\mathbf{x}(0)}\left[\mathbf{x}(0)|r\right].$$

Assuming zero-mean Gaussian noise with covariance $\mathcal{N}_0\mathbf{I}$, and a zero-mean Gaussian prior on $\mathbf{x}(0)$ with covariance $\mathrm{Trace}[\mathbf{K}_{\mathbf{x}(0)}] = \mathcal{E}$, the optimal MMSE estimate is linear:

$$\hat{\mathbf{x}}_0(r) = \mathbf{Z}\mathbf{r}$$

where

$$\mathbf{Z} = \mathbf{K}_{\{\mathbf{x}(0),\mathbf{r}\}}\mathbf{K}_{\mathbf{r}}^{-1}$$

because we have assumed zero mean $\mathbf{w}$ and $\mathbf{x}(0)$. The estimator error covariance is then

$$\mathbf{K}_e = \mathbf{K}_{\mathbf{x}_0} - \mathbf{K}_{\{\mathbf{x}(0),\mathbf{r}\}}\mathbf{K}_{\mathbf{r}}^{-1}\mathbf{K}_{\{\mathbf{r},\mathbf{x}(0)\}}$$

Derivations of these well-known results can be found in a variety of places (e.g., [6]–[8]). Now we need to adapt them to the "physical plant" specified by equation (13). That is

$$\mathbf{K}_{\mathbf{r}} = \mathbf{Q}\mathbf{K}_{\mathbf{x}(0)}\mathbf{Q}^{\top} + \mathcal{N}_0\mathbf{I}$$

and

$$\mathbf{K}_{\{\mathbf{x}(0),\mathbf{r}\}} = \mathbf{K}_{\mathbf{x}(0)}\mathbf{Q}^{\top}$$

so that

$$\mathbf{K}_{\{\mathbf{r},\mathbf{x}(0)\}} = \mathbf{Q}\mathbf{K}_{\mathbf{x}(0)}$$

Thus,

$$\mathbf{Z} = \mathbf{K}_{\mathbf{x}(0)}\mathbf{Q}^\top\mathbf{K}_{\mathbf{r}}^{-1}$$

and

$$\mathbf{K}_e = \mathbf{K}_{\mathbf{x}_0} - \mathbf{K}_{\mathbf{x}(0)}\mathbf{Q}^\top\mathbf{K}_{\mathbf{r}}^{-1}\mathbf{Q}\mathbf{K}_{\mathbf{x}(0)}$$

B. The Colored Noise Demon

We repeat equation (14) here

$$\tilde{\mathbf{r}} = \mathbf{x}_0 + \tilde{\mathbf{w}}$$

Owing to the assumption of zero mean white Gaussian noise $w(t)$, the capacity of the colored noise channel is attained with a Gaussian distribution on $\mathbf{x}_0$ and can be obtained through the following optimization:

$$C_{\text{color}} = \max_{\text{Trace}[\mathbf{K}_{\mathbf{x}_0}]=\frac{\mathcal{E}}{L}} \frac{1}{2}\left(\log|\mathbf{K}_{\mathbf{x}_0} + \mathbf{K}_{\tilde{\mathbf{w}}}| - \log|\mathbf{K}_{\tilde{\mathbf{w}}}|\right) \quad (16)$$

The specific $\mathbf{K}_{\mathbf{x}_0}$ that maximizes equation (16) is that which renders $\mathbf{K}_{\mathbf{x}_0}\mathbf{K}_{\tilde{\mathbf{w}}}^{-1}$ as close to a scaled identity matrix as possible. To achieve this end, the eigenvectors of $\mathbf{K}_{\mathbf{x}_0}$ must align with those of $\mathbf{K}_{\tilde{\mathbf{w}}}$, and the energy $\mathcal{E}/L$ distributed among those eigendimensions appropriately. This "waterfilling" or "whitening" solution, and the associated maximum C is the "capacity" of the communication channel in nats per channel use (of duration T, the observation interval) [9].

Since we assume many channel uses, L can be large which implies the amount of energy available per channel use decreases. In this case, the waterfilling solution will place all signaling energy along the eigenvector of $\mathbf{K}_{\tilde{w}}$ with the smallest eigenvalue [9]–[11]. The capacity expression of equation (16) then reduces to

$$C_{\text{color}} \approx \frac{1}{2}\log\left(\frac{\mathcal{E}/\mathcal{N}_0}{L\lambda_{min}} + 1\right) \quad (17)$$

where λ_{min} is the smallest eigenvalue of $\mathbf{Q}^{-1}(\mathbf{Q}^{-1})^\top$ since $\mathbf{K}_{\tilde{\mathbf{w}}} = \mathcal{N}_0\mathbf{Q}^{-1}(\mathbf{Q}^{-1})^\top$.

For large L we then have the total amount of information (nats) transferred as

$$B_{\text{color}} = LC_{\text{color}} \approx \frac{1}{2}\frac{\mathcal{E}}{\mathcal{N}_0\lambda_{min}} \quad (18)$$

C. The Multiaccess Demon

We repeat equation (15) here

$$\mathbf{r} = \sum_{n=1}^{N}\mathbf{q}_n(\mathbf{x}_0)_n + \mathbf{w}$$

which represents a multiple access channel with white noise, signatures $\mathbf{q}_n$ and "user" (or demon) symbols $\{(\mathbf{x}_0)_n\}$. The problem here is slightly different than the colored noise channel because demons associated with the $\{(\mathbf{x}_0)_n\}$ are assumed independent and may also interfere with one another owing to the correlation of their "signatures" $\mathbf{q}_n$.

Nonetheless, the problem is well known [12] and the (sum) capacity of the multiple access channel is given by

$$C\text{mac} = \frac{1}{2}\log\left|\mathbf{Q}\mathbf{K}_{\mathbf{x}_{\text{mac}}}\mathbf{Q}^\top\frac{\mathcal{E}}{\mathcal{N}_0 NL} + \mathbf{I}\right| \quad (19)$$

where $\mathbf{K}_{\mathbf{x}_{\text{mac}}}$ is a diagonal matrix with nonzero elements $(\mathbf{x}(0))_n^2$, $n = 1, 2, \cdots, N$ and multiuser detection techniques could be used to decode the information sent by each demon [9]–[12]. However, since from an information-theoretic standpoint, equation (14) and equation (15) are identical (assuming invertible $\mathbf{Q}$) and in the multiple access problem we have less control over where energy is placed in the eigenspace of the noise, in general we have $C_{\text{color}} \geq C_{\text{mac}}$. We will therefore take a more simplistic approach of optimal single-user detection (matched filtering) for each demon, knowing that the colored noise case provides an upper bound on the sum capacity. Assuming energy is allocated equally among the N demons and Gaussian codebooks are used, we then have

$$C_n = \frac{1}{2}\log\left(\frac{|qv_n|^2\frac{\mathcal{E}/L}{N}}{\mathcal{N}_0 + \sum_{i\neq n}\mathbf{q}_n^\top\mathbf{q}_i\frac{\mathcal{E}/L}{N|\mathbf{q}_n|}} + 1\right) \quad (20)$$

and

$$C^{\text{sum}} = \sum_n C_n \quad (21)$$

For L large $\mathcal{N}_0$ dominates in the denominator so we then have the total amount of information (nats) transferred as

$$B_{\text{mac}} \approx LC^{\text{sum}} = \frac{\mathcal{E}}{2N\mathcal{N}_0}\sum_n|\mathbf{q}_n|^2 = \frac{\mathcal{E}}{2\mathcal{N}_0}\overline{|\mathbf{q}_n|^2} \quad (22)$$

D. Distortion

Rate distortion theory [9] allows us to determine the minimum average "distortion" $\mathcal{D}$ achievable under some error measure when the number of nats specifying values (symbols) from a source is fixed to B. To be specific, for a univariate Gaussian random variable $X \sim \mathcal{N}(0, \sigma^2)$ and a square error distortion measure, the minimum distortion imposed by the finite specification of B nats is given by [9].

$$\mathcal{D} = \sigma^2 e^{-2B} \quad (23)$$

For N such i.i.d. Gaussian variables specified by a total of B nats, the minimum distortion (total mean square error) becomes

$$\mathcal{D} = N\sigma^2 e^{-2B/N} \quad (24)$$

IV. Error Comparisons

For MMSE estimation we will assume that $\mathbf{x}(0)$ is a zero mean white Gaussian vector with $\mathbf{K}_{\mathbf{x}(0)} = \frac{\mathcal{E}}{N}\mathbf{I}$. For the colored noise and the multiaccess channels, we will assume that the *message* to be transmitted is drawn from the same white Gaussian distribution with covariance $\mathbf{K}_{\mathbf{x}(0)} = \frac{\mathcal{E}}{N}\mathbf{I}$. Thus, in both demon cases, the value of σ^2 in equation (24) is $\frac{\mathcal{E}}{N}$. However, the *symbols* transmitted by the demon over the colored noise and multiple access channels are assumed to have a per-channel-use energy constraint of $\mathcal{E}/L$.

A. MMSE Estimation Error

The error covariance for MMSE estimation is

$$\mathbf{K}_e = \frac{\mathcal{E}}{N}\left(\mathbf{I} - \mathbf{Q}^\top\left(\mathbf{Q}\mathbf{Q}^\top + \frac{N\mathcal{N}_0}{\mathcal{E}}\mathbf{I}\right)^{-1}\mathbf{Q}\right) \quad (25)$$

and the total error is

$$\epsilon_{\text{MMSE}} = \text{Trace}[\mathbf{K}_e] \quad (26)$$

B. Demon Error

By substituting equation (18) into equation (24) we obtain the total error for the colored noise scenario over L signaling epochs as

$$\epsilon_{\text{color}} = \mathcal{E}\left(e^{-\frac{\mathcal{E}}{\mathcal{N}_0 \lambda_{min}}\frac{1}{N}}\right) \tag{27}$$

remembering that $\lambda_{\min}$ is the minimum eigenvalue of $\mathbf{K}_{\tilde{w}}$.

Likewise, by substituting equation (22) into equation (24) we obtain the total error for the colored noise scenario over L signaling epochs as

$$\epsilon_{\text{mac}} = \mathcal{E}\left(e^{-\frac{\mathcal{E}}{N\mathcal{N}_0}\overline{|\mathbf{q}_n|^2}}\right) \tag{28}$$

C. An Analytic Example for $N = 2$

It is useful to derive a few analytic expressions to help with later numerical calculations. Following equation (4) we have

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -a & -b \end{bmatrix}$$

with $a, b, > 0$ and real. The eigenvalues of $\mathbf{A}$ are

$$\lambda_1 = \tfrac{1}{2}\left(-b + \sqrt{b^2 - 4a}\right) \quad \lambda_2 = \tfrac{1}{2}\left(-b - \sqrt{b^2 - 4a}\right) \tag{29}$$

so that the state transition matrix is

$$\mathbf{\Phi}(t) = \frac{1}{\sqrt{b^2-4a}}\begin{bmatrix} \lambda_1 e^{\lambda_2 t} - \lambda_2 e^{\lambda_1 t} & -e^{\lambda_2 t} + e^{\lambda_1 t} \\ a e^{\lambda_2 t} - a e^{\lambda_1 t} & -\lambda_2 e^{\lambda_2 t} + \lambda_1 e^{\lambda_1 t} \end{bmatrix} \tag{30}$$

To evaluate the projections we *could* define a finite interval $[0, T]$ large enough that at interval's end, the system would be essentially at rest. However, for analytic simplicity we assume each observation interval (channel use) is $[0, \infty]$ and then imagine using $L \to \infty$ identical independent demon-channels in parallel so as to satisfy the assumptions underlying the channel capacity expressions in equation (16) and equation (19) [9].

We can then choose (not necessarily real)

$$\phi_1(t) = -j\sqrt{2\Re\{\lambda_1\}}e^{\lambda_1 t} \tag{31}$$

and via Gram-Schmidt obtain

$$\phi_2(t) = \frac{e^{\lambda_2 t} - \langle e^{\lambda_2 t}, \phi_1(t)\rangle \phi_1(t)}{|e^{\lambda_2 t} - \langle e^{\lambda_2 t}, \phi_1(t)\rangle \phi_1(t)|} \tag{32}$$

where $|\cdot|$ in this case denotes the L^2 function norm and $\langle \cdot, \cdot \rangle$ is the inner product.

We can then define α as

$$\alpha \equiv \langle e^{\lambda_2 t}, \phi_1(t)\rangle = \frac{\left(\sqrt{-b + \Re\{\sqrt{b^2 - 4a}\}}\right)}{\left(-jb + \Im\{\sqrt{b^2 - 4a}\}\right)} \tag{33}$$

and rewrite equation (32) in a slightly more compact form as

$$\phi_2(t) = \frac{e^{\lambda_2 t} - \alpha\phi_1(t)}{\langle e^{\lambda_2 t} - \alpha\phi_1(t), e^{\lambda_2 t} - \alpha\phi_1(t)\rangle} \tag{34}$$

In the next section we evaluate the analytic expressions for an overdamped system with values chosen for analytic simplicity. We also evaluate a near-critically damped ($\sqrt{b^2 - 4a} \approx 0$) and an underdamped (oscillatory) system numerically.

D. Numerical Examples for $N = 2$

Suppose we set

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -9/4 & -5 \end{bmatrix} \quad \mathbf{C} = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

The eigenvalues of $\mathbf{A}$ are $\{-1/2, -9/2\}$ so the transition matrix, $\mathbf{\Phi}(t)$, is

$$\mathbf{\Phi}(t) = \frac{1}{16}\begin{bmatrix} -2e^{-9t/2} + 18e^{-t/2} & -4e^{-9t/2} + 4e^{-t/2} \\ 9(e^{-9t/2} - e^{-t/2}) & 18e^{-9t/2} - 2e^{-t/2} \end{bmatrix} \tag{35}$$

Our orthonormal basis is

$$\phi_1(t) = e^{-t/2}$$

and

$$\phi_2(t) = \frac{15}{4}\left(e^{-9t/2} - \frac{1}{5}e^{-t/2}\right)$$

Projecting the transition matrix onto the $\phi_i()$ as in equation (9) yields

$$\mathbf{G}_1 = \begin{bmatrix} 1.1 & 0.2 \\ -0.45 & 0.1 \end{bmatrix} \quad \mathbf{G}_2 = \begin{bmatrix} -0.0\overline{33} & -0.0\overline{66} \\ 0.015 & 0.3 \end{bmatrix}$$

and thence through equation (10) we have

$$\mathbf{Q} = \begin{bmatrix} 1.1 & 0.2 \\ -0.0\overline{33} & -0.0\overline{66} \end{bmatrix}$$

Notice that in the context of the multiaccess channel, $\mathbf{q}_1$ and $\mathbf{q}_2$ are highly correlated. In addition, $|\mathbf{q}_1|^2 \gg |\mathbf{q}_2|^2$ which bodes ill for the MMSE estimation of $\dot{x}$ – or were separate demons seeking to communicate x and $\dot{x}$ over the multiaccess channel rather than sharing the sum capacity. Put another way, $\mathbf{K}_{\tilde{\mathbf{w}}} = \mathcal{N}_0\mathbf{Q}^{-1}\left(\mathbf{Q}^{-1}\right)^\top$ has large diagonal terms (10 and 272.5) which *amplify the effective observation noise*, especially for $\dot{x}$.

We can set $\mathcal{E} = 1$ with no loss of generality defining a signal-to-noise ratio $\eta = \mathcal{E}/\mathcal{N}_0$. Numerical evaluation of the MMSE, colored noise demon and multiaccess demon errors for $\eta \in \{\frac{1}{10}, 1, 10\}$ are shown for different systems in TABLES I, II and III below: In all cases, demon-mediated estimation results in lower error variance than MMSE estimation. These differences are particularly telling at larger signal to noise ratios, η.

TABLE I
Estimator Error vs. η: overdamped w/ $a = \frac{9}{4}$, $b = 5$.

	$\eta = 0.1$	$\eta = 1$	$\eta = 10$
MMSE	0.97	0.81	0.56
MAC	0.97	0.73	4.3×10^{-2}
Color	0.93	0.53	1.9×10^{-3}

TABLE II
Estimator Error vs. η: near-critically damped w/ $a = 0.99$, $b = 2$.

	$\eta = 0.1$	$\eta = 1$	$\eta = 10$
MMSE	0.96	0.78	0.47
MAC	0.96	0.69	2.3×10^{-2}
Color	0.93	0.48	6.4×10^{-4}

V. Discussion & Conclusion

Two communication theory tropes emerged *unbidden* from the application of state space projection methods to physical systems: colored noise channels and multiaccess channels.

TABLE III
Estimator Error vs. η: under-damped w/ $a = 4$, $b = 2$.

	$\eta = 0.1$	$\eta = 1$	$\eta = 10$
MMSE	0.98	0.89	0.57
MAC	0.98	0.87	0.24
Color	0.96	0.77	6.9×10^{-2}

Taking these tropes at face value and imagining an observed system that *seeks to communicate* rather than passively evolves, we formulated a method by which classical MMSE estimation could be compared to "demon"-mediated communication of state information.

A. Are Demons Fair?

We have shown that demon-mediated communication of information is (potentially *much*) more efficient than simply observing system state over a noisy channel. A simplistic (but essentially correct) explanation for this advantage is that classical estimation gets to use the channel only once whereas the demon uses the channel many times in a structured way. This is the essence of the channel coding theorem [9].

Of course, we must ask: *do our demons have an unfair advantage?* Demon-directed state conveyance certainly provided much better estimates of $\mathbf{x}(0)$. However, perhaps the comparison is unfair since in the classical case, the actual system state, $\mathbf{x}(0)$, is relevant to the observer – as in system control where the observed system must be nudged along some desired trajectory. The demon, in contrast, need only *convey* $\mathbf{x}(0)$ to the observer and leave the system at rest after communication is done. Of course, the demon could then *set* the state to $\mathbf{x}(0)$, but at an additional energy cost of $|\mathbf{x}(0)|^2$ above and beyond that used for communication. So perhaps a more fair question is to ask how much *additional* energy above $|\mathbf{x}(0)|^2$ (to set the state) does the demon require to drive the observer estimation error well below the standard MMSE. The machinery developed in this paper allows such marginal improvement questions to be answered directly although these were not pursued here.

B. An Uncertainty Principle

Our analysis also touches on the relative uncertainty of state variables. We have shown that a system under an energy constraint that deliberately signals an observer through state variation can convey some fixed number of nats, and when those nats are used to code the state vector, that finite number of nats implies some minimum distortion via rate distortion theory. However, perhaps equally interesting is the notion that there is a tradeoff in the precision of different state vector elements *reminiscent* of the Uncertainty Principle from quantum mechanics [13].

Consider that in our canonical formulation of equation (1) we have explicitly made state variables of the position and velocity of the mass (and that any system with masses and springs can be so cast). Thus,

$$B_x + B_{\dot{x}} \leq B_{\text{max}} \tag{36}$$

where B_x and $B_{\dot{x}}$ are the number of nats devoted to specifying x and $\dot{x}$ in the communication and B_{max} is the maximum number of nats carried. The Uncertainty Principle in quantum mechanics [13] is usually stated as

$$\sigma_x \sigma_p = \frac{\hbar}{2} \tag{37}$$

where σ_x^2 and σ_p^2 are the variances of the position observation and momentum, respectively, and $\hbar$ is Planck's constant.

If we then examine equation (23) we see that $\mathcal{D}$, which is exactly σ_x^2 or $\sigma_{\dot{x}}^2$, can be rewritten as

$$-\tfrac{1}{2}\log \mathcal{D} + \log \sigma = B \quad \Rightarrow \quad \sqrt{\mathcal{D}} = \sigma e^{-B}$$

where σ^2 is the variance of the unknown quantity to be specified, x or $\dot{x}$. Since $M^2\sigma_{\dot{x}}^2 = \sigma_p^2$ where M is the mass in FIGURE 1, we can then write

$$\sigma_x \sigma_p = \frac{\sigma^2}{M} e^{-(B_x + B_{\dot{x}})} \tag{38}$$

Thus, equation (37) from quantum mechanics is exactly equivalent *in form* to equation (38) with $\frac{\sigma^2}{M} e^{-(B_x+B_{\dot{x}})}$ replacing $\hbar/2$. Again, *it must be emphasized* that we do not claim *equivalence* with Planck's constant. However, since σ^2, B_x, M and $B_{\dot{x}}$ derive from the physical system, the potential for some relationship or analogy does exist.

C. All Physical Interaction Is Communication

This work was initially prompted by what seemed a simple question – what does one body "say" to another as they interact? At a semi-classical level, Newtonian mechanics, Maxwell's equations and special relativity would seem to provide an answer. But at a fundamental level, such *continuous* partial differential equations are unsatisfying because to a communication/information theorist, the conversations they imply would require infinite signaling energy. Perhaps carefully understanding the limits of the conversations leads to physics of a more discrete nature [13]–[15]

References

[1] F. Szidarovszky and A.Terry Bahill. *Linear Systems Theory*. CRC Press, 1991.

[2] T. Kailath. *Linear Systems*. Prentice-Hall, 1980.

[3] Robert Resnick David Halliday and Jearl Walker. *Fundamentals of Physics Extended*. Wiley, 2010.

[4] Gilbert Strang. *Linear Algebra and Its Applications*. Academic Press, second edition, 1992.

[5] S. Haykin. *Communication Systems*. John Wiley & Sons, 1994.

[6] J.M. Wozencraft and I.M Jacobs. *Principles of Communication Engineering*. Wiley, 1965.

[7] R.G. Gallager. *Information Theory and Reliable Communication*. Wiley, 1968.

[8] Yates, R.D. and Goodman, D.J. *Probability and Stochastic Processes: A Friendly Introduction for Electrical and Computer Engineers*. Wiley, 2014.

[9] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 2006. 2nd Ed.

[10] C. Rose, S. Ulukus, and R. Yates. Interference Avoidance in Wireless Systems. *IEEE Transactions on Wireless Communications*, 1(3):415–428, July 2002.

[11] D. C. Popescu and C. Rose. Interference Avoidance. In *Wiley Encyclopedia of Telecommunications*, 2002. J.Proakis, Ed.

[12] Sergio Verdu. *Multiuser Detection*. Cambridge University Press, 1998.

[13] Griffiths,D.J. *Introduction to Quantum Mechanics*. Prentice Hall, Upper Saddle River, NJ, 1995. ISBN 0-13-111892-7.

[14] J.A. Wheeler. Information, physics, quantum: The search for links. In *Complexity, Entropy, and the Physics of Information*. Addison-Wesley, 1990. Editor, Zurek, W.H.

[15] Erik P. Verlinde. On the Origin of Gravity and the Laws of Newton. *arXiv:1001.0785v1 [hep-th]*, Jan 2010. http://arxiv.org/abs/1001.0785.