Special Issue on Biological Applications of Information Theory in Honor of Claude Shannon's Centennial—Part 2

Inscribed Matter Communication: Part I

Christopher Rose, *Fellow, IEEE*, and I. S. Mian

***Abstract*—We provide a fundamental treatment of the molecular communication channel wherein "inscribed matter" is transmitted across a spatial gap to provide reliable signaling between a sender and receiver. Inscribed matter is defined as an ensemble of "tokens" (biotic/abiotic objects) and is inspired, at least partially, by biological systems where groups of individually constructed discrete particles ranging from molecules to viruses and organisms are released by a source and travel to a target—for example, morphogens or semiochemicals diffuse from one cell, tissue, or organism to another. For identical-tokens that are neither lost nor modified, we consider messages encoded using three candidate communication schemes: 1) token timing (timed release); 2) token payload (composition); and 3) token timing plus payload. We provide capacity bounds for each scheme and discuss their relative utility. We find that under not unreasonable assumptions, megabit per second rates could be supported at 100 femtoWatt transmitter powers. Since quantities such as token concentration or token counting are derivatives of token arrival timing, token timing undergirds all molecular communication techniques. Thus, our modeling and results about the physics of efficient token-based information transfer can inform investigations of diverse theoretical and practical problems in engineering and biology. This paper, Part I, focuses on the information theoretic bounds on capacity. Part II develops some of the mathematical and information theoretic machinery that support the bounds presented here.**

***Index Terms*—Biological communication, molecular channels, molecular communication, molecular information theory, inscribed matter, timing channels.**

I. Introduction

SCALE-APPROPRIATE signaling methods become important as systems shrink to the nanoscale. For systems with feature sizes of microns and smaller, electromagnetic and acoustic communication become increasingly inefficient because energy coupling from the transmitter to the medium and from the medium to the receiver becomes difficult at usable frequencies. Biological systems, with the benefit of lengthy evolutionary experimentation, seem to have arrived at a ubiquitous solution to this signaling problem at small and not so small scales: use of "inscribed matter" (an ensemble of discrete particles) which travels through some material bearing a message from one entity to another. Broad classes of such particles and some illustrative examples include but are not limited to

- *Molecules:* electronically activated species, ions, chemicals, biopolymers, and macromolecular complexes.
- *Membrane-bound structures:* intra- and extracellular vesicles (exosomes, microvesicles, apoptotic bodies, ectosomes, endosomes, lysosomes, autophagosomes, vacuoles) and intracellular organelles (nuclei, mitochondria, chloroplasts).
- *Cells:* stem cells, tumor cells, and hematocytes.
- *Acellular, unicellular and multicellular life forms ("organisms"):* viruses, viroids, phages, plasmids, bacteria, archaea, fungi, protists, plants, and animals.
- *Objects:* matter in the natural world (pollen grains, seeds, proteinaceous aggregates such as prions), and human artifacts (Voyager Golden Record).

Studies of engineered nano-scale communication systems have focused on the encoding, transmission, and decoding of information using patterns of one class of discrete particles. A large portion of this work in "molecular communication" has considered the time-varying concentration profile of molecules as the fundamental signal measurement [1]–[7]. However, concentration is a *collective* property and masks the underlying physics of molecule release by the sender and capture by the receiver. This begs the questions of truly fundamental limits for communication using ensembles of molecules in particular, discrete particles more broadly, and what we term "tokens" in general.

This paper is organized as follows.

First, we discuss communication using inscribed matter from biological and engineering perspectives. We illustrate how scenarios spanning a wide range of spatial and temporal scales and from seemingly disparate disciplines can be understood within a unified framework: the token timing and/or token payload channel, a communication scheme wherein information is carried from sender to receiver by tokens via their timed release, their composition, or both. We will assume tokens always (eventually) arrive, and are removed from circulation upon first seizure by the receiver. This abstraction encompasses not only token timing but also the token concentration and token counting models prevalent in the molecular communication literature [1]–[9].

Next, we describe the token timing channel wherein information is encoded *only* in the release times of *identical* tokens rather than inscribed onto tokens (token payload) or in the number of tokens released (token counting). Though seemingly limited, this pure timing model enables precise analysis

Manuscript received June 15, 2016; revised November 20, 2016; accepted January 9, 2017. Date of publication January 18, 2017; date of current version February 17, 2017. This work was supported by the NSF under Grant CDI-0835592. The associate editor coordinating the review of this paper and approving it for publication was S. M. Moser.

C. Rose is with Brown University, Providence, RI 02912 USA (e-mail: christopher_rose@brown.edu).

I. S. Mian is with University College London, London WC1E 6BT, U.K.

Digital Object Identifier 10.1109/TMBMC.2017.2655025

of all molecular communication schemes with independent stochastic (but asymptotically assured, one-time) token arrivals. The requisite formalized signaling model is carefully derived so that the usual energy-dependent asymptotic sequential channel use coding results based on mutual information between input and output can be employed [10, Ch. 8 and 10]. We then show how these results can be applied to token counting and tokens with payloads. We focus on molecular tokens, particularly DNA and protein sequences since their energy requirements (and information content) are fairly well-understood, but the ideas are generalizable. We find that information transfer using inscribed matter can be extremely energy efficient with potentially megabit per second rates at 100fW transmitter power.

Finally, we explore how our studies and the attendant insights could aid biological understanding of and inform engineering approaches to inscribed matter communication.

II. Inscribed Matter

A. Communication via Discrete Particles: A Natural World Perspective

Networks of intercommunicating biological entities occur at whatever level one cares to consider: (macro)molecules, cells, tissues, organisms, populations, microbiomes, ecosystems, and the Earth. An ancient yet still widespread method for one entity to convey a message to another is via inscribed matter. Typically, information bearing-discrete particles are released by a source, travel through a material, and are captured by a target where they are interpreted. The following are exemplars of the diversity and complexity of such inscribed matter communication (particles are italicized).

- *Electrons* produced as part of a metabolic process in one microorganism travel to another through bacterial nanowires (electrically conductive appendages), bacterial cables (thousands of individuals lined up end-to-end with electron donors located in the deeper regions of marine sediment and electron acceptors positioned in its upper layers where oxygen is more abundant), and an extracellular matrix of self-produced polymers [11].
- *Free radicals* generated by the action of ionizing radiation on molecules in the nucleoplasm diffuse to the genome where they alter/damage nucleotide bases and sugars.
- *Messenger RNA (mRNA) molecules* transcribed from a eukaryotic genome in a nucleus migrate to ribosomes in the cytoplasm where they are translated into proteins.
- *Acetylcholine (ACh) molecules* released by a vertebrate motor neuron diffuse across the synaptic cleft to a muscle cell where their binding to plasma membrane nicotinic ACh receptors triggers fiber contraction.
- A *homing endonuclease* produced from an allele (one of a pair of genes) encoding this "DNA scissor" finds the other allele in the same or a different genome where the "parasitic element" reinserts itself.
- *Ions, molecules, organelles, bacteria and viruses* in one cell travel through a thin membrane channel (tunneling nanotube) to the physically connected cell where they elicit a response.
- *Semiochemicals* (chemical substances or mixtures of volatile molecules) emitted by one individual travel to another of the same (pheromones) or different species (allelochemicals) where they elicit a response – allomones benefit only the sender, kairomones benefit only the receiver, and synomones benefit both.
- *Vesicles* secreted by one cell traverse the extracellular space or body fluids to another cell where their bioactive contents (viruses, molecules) are unloaded.
- *Vesicles* at one location in a cell are transported by molecular motors along a track system of cytoskeletal filaments to another point where their freight (organelles, molecules) are unloaded.
- *Single and clusters of metastatic cells* that have escaped from a primary tumor circulate through the blood or lymph to a secondary organ site where, after extravasation, they can seed a new tumor.
- *Organic particles* such as microorganisms, fungal spores, small insects, and pollen grains associated with one macroorganism, geological site or geographic location relocate to another host or region where they influence the local biochemistry, geochemistry and climate [12]; long distance transport (including movement within and between continents and oceans) can occur via the same meteorological phenomena and processes, such as jetstreams and hurricanes, that translocate non-biological particles such as sea salt and dust.
- *Crustal material* ejected by one Solar System body travels to another where it has the potential to seed life if it carries microbial spores or building blocks such as amino acids, nucleobases and lipids [13], [14].

Regardless of the precise components of the communication system – the particles (information carriers), source (sender), spatial gap (transmission medium), and target (receiver) – two key questions arise: *How reliable is communication?* and *How is useful information conveyed given constraints on resources?* Here, we investigate token timing (particle release and capture times) and token payload (energy required to manufacture particles, to assemble symbolic strings from a set of building blocks – we neglect the energy required for *de novo* synthesis of the building blocks). Although not stated explicitly, our energy model could include also token sequestration, ejection, and transport (see aforementioned vesicle-related exemplars).

Tokens in the timing channel model are neither lost nor modified: the number and makeup of the tokens emitted by the source are the same as the ones arriving at the target. All that differs are their times of emission and their times of arrival. While accommodating tokens that are delayed temporarily, our mathematical model does not consider directly tokens that are detained permanently, removed entirely, never arrive, or are changed en route. In the natural world, particles often interact with the material through which they travel resulting in their immurement and ultimate removal or detention and eventual discharge. For instance, *free radicals* may react chemically with other particles, *mRNAs* may be modified post-transcriptionally, *ACh* can be degraded by the enzyme acetylcholine esterase, and *circulating tumor cells* may be destroyed by the immune system. The random path of

a *semiochemical* diffusing through air, soil or water may result in a trajectory that leads away from the destined individual. *Organic particles* may be immobilized within the mucus that covers the inner linings of the body. *Bacteria* in the atmosphere, particularly plant pathogens, may be diverted because they can nucleate the formation of ice in clouds resulting in snow, rain and hail [12].

Despite such shortcomings, our model provides an organizing principle for all forms of molecular communications because token loss or corruption can only decrease the capacity of the system we analyze. Furthermore, the analysis is "compartmental" in that token corruption and loss can be treated separately without invalidating the fundamental "outer bound" results.

B. Communication via Physical Objects: An Engineering Perspective

Inscribed matter can often be the most energy efficient means of communication when delay can be tolerated. Indeed, a once popular communication networks textbook [15] states,

> *Never underestimate the bandwidth of a station wagon full of tapes hurtling down the highway.*
> – *A.S. Tanenbaum*, Computer Networks, *4th ed., p. 91*

This somewhat tongue-in-cheek "folklore" should come as no surprise. From early antiquity, private persons, governments, the military, press agencies, stockbrokers and others have used carrier pigeons to convey messages. Today, "sneakernets" [16] have been proposed as a low-latency high-fidelity network architecture for quantum computing across global distances: ships carry error-corrected quantum memories installed in cargo containers [17].

Previous work on mobile wireless communication found that network capacity could be increased if delay-tolerant traffic was queued until the receiver and sender were close to one another, perhaps enough to exchange physical storage media [18]–[30]. This observation prompted a careful consideration of the energetics involved in delivery of physical messages, and a series of papers [31]–[33] revealed the surprising results that inscribed matter can be many orders of magnitude more efficient than radiative methods, even with moderate delay constraints and over a variety of size scales. In fact, [33] showed that over interstellar distances (10k light years), inscribed matter could be on the order of 10^{15} times more energy-efficient than radiated messages, suggesting that evidence of extraterrestrial civilizations would more likely come from artifacts than from radio messages if energy requirements are a proxy for engineering difficulty [34].

At the other end of the size scale is biologically inspired inscribed matter communication [**?**] which ranges from timed release of identical signaling agents to specially constructed information carriers [1]–[7], [35]–[38]. Although this nascent field of molecular communication is replete with seemingly futuristic applications such as *in vivo* biological signaling and surgical/medicinal/environmental microbot swarms, the theoretical rates and energy efficiencies (especially through media unfriendly to radiation) are sufficiently large [39] to warrant careful theoretical and practical consideration.

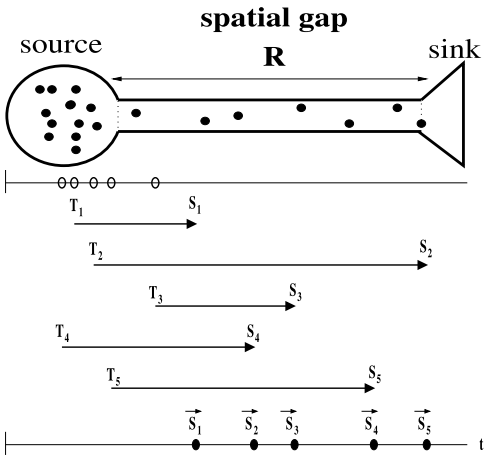

Fig. 1. An abstraction of an inscribed matter communication channel. A sender transmits an ensemble of tokens ("inscribed matter") to a receiver across a spatial gap (of length R in the figure). The tokens are released at (unordered) times $\{T_m\}$, propagate through a transmission medium and are captured at corresponding times $\{S_m\}$. For identical-tokens, the receiver sees ordered arrivals $\{\vec{S}_m\}$ which may differ in index from the unordered arrivals $\{S_m\}$.

From an engineering perspective, the basic idea of inscribed matter communication is very simple (see FIGURE 1). Information is coded in the structure of the signaling agent and/or its release time at the sender. These agents stochastically traverse some spatial gap to the receiver where they are captured and the information decoded. Of course, there are many details and variations on the theme. Like biological systems, the signaling agents (or tokens as we call them) could be identical, implying that timing (which includes time varying concentration) is **the only** way to convey information. Alternatively, tokens could themselves carry data payloads (in addition to, or in lieu of timing) similar to packets traveling the Internet. The "gap" (channel) could be a medium through which tokens diffuse stochastically, or some form of active transport might be employed. In addition, tokens could be deliberately "eaten" by gettering agents injected by the sender, channel or receiver. Tokens could also be corrupted during passage through the channel or might simply get "lost" and never reach the receiver [9], [37].

Upon arrival at the receiver, tokens are captured. If we sought to mimic the reception process in biological systems, noise is just one factor to consider. Typically, the structure of a receptor is matched stereochemically to its ligand (token). The kinetics of receptor-ligand interactions is important as are the number, density and spatiotemporal organization of receptors. While a given receptor may bind preferentially to a particular ligand, there may be other different or identical (but from another source) *interfering* ligands which bind to the same receptor. If one considers networks of molecular transceivers, this sort of "cross talk" or outright interference can be functionally important.

C. Inscribed Matter Communication: Model Distillation

While the various engineering and biological scenarios require slightly different information theoretic formulations,

they can *all* be understood within a unified framework: the identical-token timing channel wherein

- Token release and capture timing is the only mechanism for information transfer.
- Tokens always (eventually) arrive at the receiver.
- Tokens are removed promptly from circulation (or deactivated) after first reception.

The identical-token timing channel abstraction [8], [19], [39]–[41] encompasses token concentration or token counting models since time-varying concentration (or token counts in "bit intervals") at a receiver is a coarse-time approximation to our precise individual token timing model. The timing channel is also important for understanding information carriage via payload-charged tokens whose information packets may need resequencing at the receiver. That is, timing channel results provide tight bounds on resequencing overhead and are especially important if it is technologically difficult to construct tokens with large payloads. The timing channel analysis applies also to channels where the number of tokens sent during signaling intervals is the information carrier [8].

A version of the timing channel has been considered previously in the award-winning paper "Bits Through Queues" by Anantharam and Verdú [42] and later by Sundaresan and Verdú [43], [44]. A key difference is that the molecular communication problem is explicitly an infinite-"server" system where each token enters "service" (transport) immediately upon "arrival" (release), making direct application of single-server [42]–[44] and finite-server [46, Sec. III] results difficult.

Our timing channel analysis provides *outer bounds*. By the data processing theorem [10, Ch. 2], token re-capture arising from ligand-receptor binding/unbinding [37] and token loss (erasure) cannot increase channel capacity. Likewise, processing, corruption and loss of payload bearing tokens cannot increase channel capacity. Thus, the timing channel not only allows upper limits on capacity to be obtained, but also permits the overall channel to be treated as a compartmentalized cascade, each constituent of which can be analyzed separately and compared to identify potential information transfer bottlenecks.

III. Mathematical Formulation

Following [19], [39]–[41], and [45], assume emission of M identical-tokens at times $\{T_m\}$, and their capture at times $\{S_m\}$, $m = 1, 2, \ldots, M$. The duration of token m's first-passage between source and destination is D_m. These D_m are assumed i.i.d. with $f_{D_m}(d) = g(d) = G'(d)$ where $g(\cdot)$ is some causal probability density with mean $\frac{1}{\mu}$ and cumulative distribution function (CDF) $G(\cdot)$. We also assume that $g(\cdot)$ contains no singularities. Thus, the first portion of the channel is modeled as a sum of random M-vectors

$$\mathbf{S} = \mathbf{T} + \mathbf{D} \tag{1}$$

as shown in FIGURE 2 prior to the sorting operation.

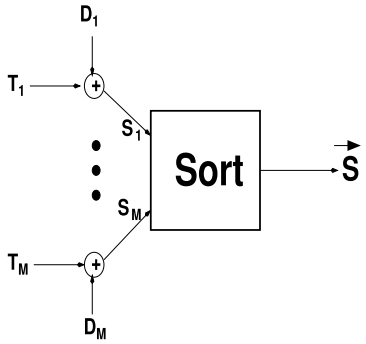

Fig. 2. The token release with reordering inscribed matter communication channel. For token m released at time T_m, the duration of its first-passage between the sender and receiver is D_m so it arrives at time S_m. The $\{S_m\}$ are then sorted by order of arrival. Since the M tokens are identical, the ordered arrival time $\vec{S}_m$ may not correspond to S_m.

We therefore have

$$\begin{aligned} f_{\mathbf{S}}(\mathbf{s}) &= \int_{\mathbf{0}}^{\infty} f_{\mathbf{T}}(\mathbf{t}) f_{\mathbf{S}|\mathbf{T}}(\mathbf{s}|\mathbf{t})\, d\mathbf{t} \\ &= \int_{\mathbf{0}}^{\mathbf{s}} f_{\mathbf{T}}(\mathbf{t}) \prod_{m=1}^{M} g(s_m - t_m)\, d\mathbf{t} \\ &= \int_{\mathbf{0}}^{\mathbf{s}} f_{\mathbf{T}}(\mathbf{t}) \mathbf{g}(\mathbf{s} - \mathbf{t})\, d\mathbf{t} \end{aligned} \tag{2}$$

where

$$\mathbf{g}(\mathbf{s} - \mathbf{t}) = \prod_{m=1}^{M} g(s_m - t_m)$$

We impose a launch (emission) deadline, $T_m \leq \tau$, $\forall m \in \{1, 2, \ldots, M\}$. The associated emission time ensemble probability density $f_{\mathbf{T}}(\mathbf{t})$ is assumed causal, but otherwise arbitrary. Had we imposed a mean launch time constraint instead of a deadline, the channel between $\mathbf{T}$ and $\mathbf{S}$ would be the parallel version of Anantharam and Verdú's *Bits Through Queues* [42]. However, since the tokens are identical we cannot necessarily determine which arrival corresponds to which emission time. That is, the final output of the channel is a reordering of the $\{s_m\}$ to obtain a set $\{\vec{s}_m\}$ where $\vec{s}_m \leq \vec{s}_{m+1}$, $m = 1, 2, \ldots, M-1$, as shown on the right hand side of FIGURE 2 after the sorting operation.

We write this relationship as

$$\vec{\mathbf{S}} = P_{\Omega}(\mathbf{S}) \tag{3}$$

where $P_{\Omega}(\cdot)$, $\Omega = 1, 2, \ldots, M!$, is a permutation operator and Ω is that permutation index which produces ordered $\vec{\mathbf{S}}$ from the argument $\mathbf{S}$. We define $P_1(\cdot)$ as the identity permutation operator, $P_1(\mathbf{s}) = \mathbf{s}$.

We note that the event $S_i = S_j$ $(i \neq j)$ is of zero measure owing to the no-singularity assumption on $g(\cdot)$, Thus, for analytic convenience we will assume that $f_{\mathbf{S}}(\mathbf{s}) = 0$ whenever two or more of the s_m are equal and therefore that the $\{\vec{s}_m\}$ are strictly ordered wherever $f_{\vec{\mathbf{S}}}(\cdot) \neq 0$ (*i.e.,* $\vec{s}_m < \vec{s}_{m+1}$).

Thus, the density $f_{\vec{\mathbf{S}}}(\vec{\mathbf{s}})$ can be found by "folding" the density $f_{\mathbf{S}}(\mathbf{s})$ about the hyperplanes described by one or more of the

TABLE I
A GLOSSARY OF KEY QUANTITIES CONSIDERED/DEFINED IN THE ANALYSIS. FOR CONTINUITY AND CLARITY, THE IDENTICAL TABLE IS INCLUDED IN THE COMPANION PAPER, PART II [46]

Token	A unit released by the transmitter and captured by the receiver
Payload	Physical information (inscribed matter) carried by a token
λ	The average rate at which tokens are released/launched into the channel
$\mathbf{T}$	A vector of token release/launch times
First-Passage	The time between token release/launch and token capture at the receiver
$\mathbf{D}$	A vector of first-passage times associated with launch times $\mathbf{T}$
$G(\cdot)$	The cumulative distribution function for first-passage random variable D
$1/\mu$	Average/mean first-passage time
ρ	λ/μ, a measure of system token "load"
$\mathbf{S}$	A vector of token arrival times, $\mathbf{S} = \mathbf{T} + \mathbf{D}$
$P_k(\mathbf{x})$	A permutation operator which rearranges the order of elements in vector $\mathbf{x}$
Ω	The "sorting index" which produces $\vec{\mathbf{S}}$ from $\mathbf{S}$, *i.e.*, $\vec{\mathbf{S}} = P_\Omega(\mathbf{S})$
$\vec{\mathbf{S}}$	An *ordered* vector of arrival times obtained by sorting the elements of $\mathbf{S}$ (note, the receiver only sees $\vec{\mathbf{S}}$ not $\mathbf{S}$)
$I(\mathbf{S};\mathbf{T})$	The mutual information between the launch times (input) and the arrival times (output)
$I(\vec{\mathbf{S}};\mathbf{T})$	The mutual information between the launch times (input) and the *ordered* arrival times (output)
$h(\mathbf{S})$	The differential entropy of the arrival vector $\mathbf{S}$
$H(\Omega\|\vec{\mathbf{S}},\mathbf{T})$	The *ordering entropy* given the input $\mathbf{T}$ and the output $\vec{\mathbf{S}}$
$H^{\uparrow}(\mathbf{T})$	An upper bound for $H(\Omega\|\vec{\mathbf{S}},\mathbf{T})$
C_q and C_t	The asymptotic per token and per unit time capacity between input and output

s_m equal until the resulting probability density is non-zero only on the region where $s_m < s_{m+1}$, $m = 1, 2, \ldots, M-1$. Analytically we have

$$f_{\vec{\mathbf{S}}}(\vec{\mathbf{s}}) = \sum_{n=1}^{M!} f_{\mathbf{S}}\big(P_n(\vec{\mathbf{s}})\big) \tag{4}$$

Then, since $f_{S|T}(s|t) = g(s-t)$, we can likewise describe $f_{\vec{\mathbf{S}}|\mathbf{T}}(\mathbf{s}|\mathbf{t})$ as

$$f_{\vec{\mathbf{S}}|\mathbf{T}}(\mathbf{s}|\mathbf{t}) = \sum_{n=1}^{M!} \mathbf{g}(P_n(\mathbf{s}) - \mathbf{t}) \prod_{m=1}^{M} u([P_n(\mathbf{s})]_m - t_m) \tag{5}$$

again for $s_1 < s_2 < \cdots < s_m$ and zero otherwise where $u(\cdot)$ is the unit step function. With exponential first-passage, $g(d) = \mu e^{-\mu d} u(d)$, becomes

$$f_{\vec{\mathbf{S}}|\mathbf{T}}(\mathbf{s}|\mathbf{t}) = \mu^M e^{-\mu \sum_{i=1}^{M}(s_i - t_i)} \left(\sum_{n=1}^{M!} \mathbf{u}(P_n(\mathbf{s}) - \mathbf{t}) \right) \tag{6}$$

again assuming $s_1 < s_2 < \cdots < s_m$ and $\mathbf{u}(\cdot)$ the multidimensional step function ($\mathbf{u}(\mathbf{x}) = \Pi_{i=1} u(x_i)$). It is worth mentioning explicitly that equation (6) *does not assume* arguments $s_i \geq t_i$ as might be implicit in equation (2).

Finally, the problem structure will allow us to make use of multi-dimensional function symmetry (hypersymmetry) arguments, $f(\mathbf{x}) = f(P_n(\mathbf{x}))$ $\forall$ permutations n. The following property of expectations of hypersymmetric functions over hypersymmetric random variables will later prove useful.

Theorem 1 (Hypersymmetric Expectation): Suppose $Q(\mathbf{x})$ is a hypersymmetric function, $Q(\mathbf{x}) = Q(P_k(\mathbf{x}))$ $\forall k$, and $\mathbf{X}$ is a hypersymmetric random vector. Then, when $\vec{\mathbf{X}}$ is the ordered version of random vector $\mathbf{X}$ we have

$$E_{\vec{\mathbf{X}}}\Big[Q\Big(\vec{\mathbf{X}}\Big)\Big] = E_{\mathbf{X}}[Q(\mathbf{X})] \tag{7}$$

Proof: Theorem 1 $\tilde{\mathbf{X}}$ is a deterministic function of $\mathbf{X}$; i.e., $\theta(\mathbf{X}) = \tilde{\mathbf{X}}$. Thus,

$$E\Big[Q\Big(\tilde{\mathbf{X}}\Big)\Big] = E[Q(\theta(\mathbf{X}))] = E[Q(\mathbf{X})] \tag{8}$$

where the last equality results from the hypersymmetry of $Q(\cdot)$. ■

With these preliminaries done, we can now begin to examine the mutual information between the unordered emission times $\mathbf{T}$, the unordered arrival times $\mathbf{S}$, and the ordered (sorted) arrival times $\vec{\mathbf{S}}$.

IV. INFORMATION THEORETIC ANALYSIS

A. Formalizing the Signaling Model

To determine whether the mutual information between $\mathbf{T}$ the input and $\vec{\mathbf{S}}$ the output is a measure of channel capacity, it suffices to have a signaling model which patently supports the usual asymptotically large block length and repeated independent sequential channel uses paradigm [10, Ch. 8 and 10]. In addition, we must also pay attention to the channel use energetics since lack of energy constraints can lead to unrealistic results. Thus, we have defined a channel use as the launch and capture of M tokens under an emission deadline constraint, τ, with the further constraint that

$$\lambda\tau = M \tag{9}$$

where λ, the token launch average intensity, has units of tokens per time. Equation (9) is implicitly a constraint on average power assuming a fixed per-token energy cost for construction/sequestration/release/delivery. We also note that the signaling interval τ is now an explicit function of M as in

$$\tau = \tau(M) = \frac{M}{\lambda}$$

So, consider FIGURE 3 where successive M-token transmissions – channel uses – are depicted. We will assume a "guard interval" of some duration $\gamma(M, \epsilon)$ between successive transmissions so that all M tokens are received before the

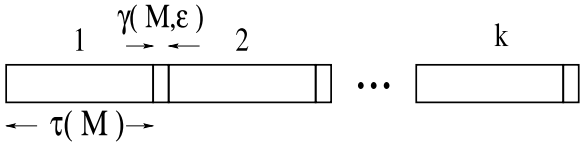

Fig. 3. Successive M-emission channel k successive M-emission channel uses. timing channel, the sender emits M tokens over the transmission interval $\tau(M) = \frac{M}{\lambda}$. $\gamma(M, \epsilon)$ is the waiting period (guard interval) before the next channel use.

beginning of the next channel use with probability $(1-\epsilon)$ for arbitrarily small $\epsilon > 0$. We further require that the average token emission rate, $M/(\tau(M) + \gamma(M, \epsilon))$ satisfies

$$\lim_{\epsilon \to 0} \lim_{M \to \infty} \frac{M}{\tau(M) + \gamma(M, \epsilon)} = \lambda \tag{10}$$

We then require that the last token arrival time $\vec{S}_M$ occurs before the start of the next channel use with probability 1. That is, given arbitrarily small ϵ we can always find a finite M^* such that

$$\text{Prob}\left\{\vec{S}_M \leq \tau(M) + \gamma(M, \epsilon)\right\} > 1 - \epsilon \tag{11}$$

$\forall M \geq M^*$. We now derive a sufficient condition on first-passage time densities for which equation (11) is true.

Calculating a CDF for $\vec{S}_M$ is in general difficult since emission times T_m might not be independent. However, for a fixed emission interval $[0, \tau(M)]$ we can readily calculate a worst case CDF for $\vec{S}_M$ and thence an upper bound on the guard interval duration that satisfies the arrival condition of equation (11). That is, for a given emission schedule $\mathbf{t}$, the $\mathbf{S}$ are conditionally independent and the CDF for the final arrival is

$$F_{\vec{S}_M|\mathbf{t}}(s|\mathbf{t}) = \prod_{m=1}^{M} G(s - t_m) u(s - t_m) \tag{12}$$

so that

$$F_{\vec{S}_M}(s) = \int_{\mathbf{0}}^{\tau(\mathbf{M})} f_{\mathbf{T}}(\mathbf{t}) \prod_{m=1}^{M} G(s - t_m) u(s - t_m)\, d\mathbf{t} \tag{13}$$

However, it is easy to see that

$$F_{\vec{S}_M}(s) \geq G^M(s - \tau(M)) u(s - \tau(M)) \tag{14}$$

since $G(s-t)$ is monotone decreasing in t.

The end of the guard interval is $\tau(M) + \gamma(M, \epsilon)$, so the probability that the last arrival time $\vec{\mathbf{S}}_M$ occurs before the next signaling interval obeys

$$F_{\vec{\mathbf{S}}_M}(\tau(M) + \gamma(M, \epsilon)) \geq G^M(\gamma(M, \epsilon)) \tag{15}$$

And to meet the requirement of equation (11) we must have

$$\lim_{M \to \infty} G^M(\gamma(M, \epsilon)) = 1 \tag{16}$$

which for convenience, we rewrite as

$$\lim_{M \to \infty} M \log G(\gamma(M, \epsilon)) = 0 \tag{17}$$

If rewrite $\log G(\gamma(M, \epsilon))$ in terms of the complementary CDF (CCDF) $\bar{G}(\cdot)$ (which must be vanishingly small in large M if we are to meet the conditions of equation (11)) and note that $\log(1-x) \approx -x$ for x small, we have

$$-\log\big(1 - \bar{G}(\gamma(M, \epsilon))\big) \approx \bar{G}(\gamma(M, \epsilon))$$

for sufficiently large M. Thus, a first-passage distribution whose CCDF satisfies

$$\lim_{M \to \infty} M \bar{G}(\gamma(M, \epsilon)) = 0 \tag{18}$$

with some suitable $\gamma(M, \epsilon)$ will also allow satisfaction of equation (11). However, the satisfaction of equation (18) requires that $1/\bar{G}(\gamma(M, \epsilon))$ be *asymptotically supralinear* in M.

We then note that since all first-passage times are non-negative random variables, the mean first-passage time is given by [47]

$$E[D] = \int_0^\infty \bar{G}(x) dx \tag{19}$$

The integral of equation (19) exists **iff** $1/\bar{G}(x)$ is asymptotically supralinear in x. Thus, the existence of $E[D]$ in turn implies that choosing $\gamma(M, \epsilon) = \epsilon M$ allows satisfaction of equation (18). So, in the limit of vanishing ϵ we then have

$$\lim_{\epsilon \to 0} \lim_{M \to \infty} \frac{M}{\tau(M) + \gamma(M, \epsilon)} \geq \lim_{\epsilon \to 0} \frac{\lambda}{1 + \epsilon} = \lambda$$

and the energy requirement of equation (10) is met in the limit while assuring asymptotically independent sequential channel uses.

The above development proves the following theorem.

Theorem 2 ($E[D] < \infty$ Permits Asymptotically Independent Sequential Channel Uses): Consider the channel use discipline depicted in FIGURE 3 where tokens are emitted on an interval $[0, \tau(M)]$ with $\tau(M) = \frac{M}{\lambda}$ and guard intervals of duration $\gamma(M, \epsilon)$ are imposed between channel uses. If the mean first-passage time $E[D]$ is finite, then guard intervals can always be found such that the sequential channel uses approach asymptotic independence as $\epsilon \to 0$, and the relative duration of the guard interval, $\gamma(M, \epsilon)$ vanishes compared to $\tau(M)$ as $M \to \infty$.

Proof: Theorem 2 See the development leading to the statement of Theorem 2. ∎

Now, suppose the transport process from source to destination has *infinite* first-passage time, implying that $1/\bar{G}(x)$ is linear or sublinear in x. Is asymptotically independent sequential channel use possible? The answer seems to be no.

As a best case, the minimum probability of tokens arriving outside $\tau(M) + \gamma(M, \epsilon)$ is obtained if all emissions occur at $t = 0$ (see equation (13)). Any other token emission distribution must have larger probability of interval overrun. For asymptotically independent sequential channel use we then must have, following equation (10) and equation (18),

$$\lim_{M \to \infty} M \bar{G}\left(\frac{M}{\lambda} + \gamma(M, \epsilon)\right) = 0 \tag{20}$$

We notice that the argument of $\bar{G}(\cdot)$ is at least linear in M, and a linear-in-M argument will not drive $\bar{G}(\cdot)$ to zero faster than $1/M$ because $1/\bar{G}(\cdot)$ is not supralinear. Thus, the argument of $\bar{G}(\cdot)$ must be supralinear in M to drive $\bar{G}(\frac{M}{\lambda} + \gamma(M, \epsilon))$ to

zero faster than $1/M$ which in turn implies that $\gamma(M,\epsilon)$ must be supralinear in M. However, if $\gamma(M,\epsilon)$ is supralinear in M, then equation (10) cannot be satisfied and we have proved the following theorem:

Theorem 3 (E[D] = ∞ Does Not Permit Asymptotically Independent Sequential Channel Uses): Consider the channel use discipline depicted in FIGURE 3 where tokens are emitted on an interval $[0,\tau(M)]$ with $\tau(M)=\frac{M}{\lambda}$ and guard intervals of duration $\gamma(M,\epsilon)$ are imposed between channel uses. If the mean first-passage time $E[D]$ is infinite, then guard intervals can never be found such that the sequential channel uses approach asymptotic independence as $\epsilon\to 0$, and the relative duration of the guard interval, $\gamma(M,\epsilon)$ vanishes compared to $\tau(M)$ as $M\to\infty$.

Proof: Theorem 3 See the development leading to the statement of Theorem 3. ■

To summarize, if the mean first-passage time exists, then asymptotically independent sequential channel uses are possible and the mutual information $I(\vec{\mathbf{S}};\mathbf{T})$ is the proper measure of information transport through the channel. Conversely, if the mean first-passage time is infinite, then asymptotically independent sequential channel uses are impossible and the associated channel capacity problem is ill-posed. It is worth noting that free-space diffusion (without drift) has infinite $E[D]$. However, since all physical systems have finite extent, $E[D]$ is always finite for any realizable ergodic token transport process.

B. Channel Capacity Definitions

The maximum $I(\vec{\mathbf{S}};\mathbf{T})$ is the *channel capacity* in units of bits/nats per channel use. However, we will find it useful to define the maximum mutual information between $\mathbf{T}$ and $\vec{\mathbf{S}}$ *per token*. That is, the channel capacity per token C_q is

$$C_q(M,\tau(M)) \equiv \frac{1}{M}\max_{f_{\mathbf{T}}(\cdot)} I(\vec{\mathbf{S}};\mathbf{T}) \tag{21}$$

Since $\tau(M)=M/\lambda$, it is easy to see that $C_q(M,\tau(M))$ will be monotone increasing in M since concatenation of two emission intervals with durations $\tau(M/2)$ and $M/2$ tokens each is more constrained than a single interval of twice the duration $\tau(M)$ with M tokens. We can thus say that

$$C_q(M,\tau(M)) \geq 2C_q(M/2,\tau(M/2)) \tag{22}$$

We can then define the limiting capacity in nats per token as

$$C_q \equiv \lim_{M\to\infty} C_q(M,\tau(M)) \tag{23}$$

with no stipulation as yet to whether the limit exists or is bounded away from zero.

Now consider the capacity per unit time. The duration of a channel use (or signaling epoch) is $\tau(M)+\gamma(M,\epsilon)$ (see FIGURE 3). Thus, for a given number M of emissions per channel use and a probability $(1-\epsilon)$ that all the tokens are received before the next channel use, we define the channel capacity in nats per unit time as

$$\begin{aligned} C_t(M,\epsilon) &\equiv \max_{f_{\mathbf{T}}(\cdot)} \frac{I(\vec{\mathbf{S}};\mathbf{T})}{\tau(M)+\gamma(M,\epsilon)} \\ &= C_q(M,\tau(M))\left(\frac{M}{\tau(M)+\gamma(M,\epsilon)}\right) \end{aligned}$$

which in the limits of $\epsilon\to 0$ and $M\to\infty$ becomes

$$\lim_{\epsilon\to 0}\lim_{M\to\infty} C_t(M,\epsilon) = \lambda C_q \tag{24}$$

via equation (10) and equation (23).

The above development proves the following theorem:

Theorem 4 (Identical-Token Timing Channel Capacity): If the mean first-passage time $E[D]$ exists, then the channel capacity in nats per unit time obeys

$$C_t = \lambda C_q \tag{25}$$

where C_q is the capacity per token defined in equation (23) and λ is the average token emission rate.

Proof: Theorem 4 See Theorem 2 and the development leading to the statement of Theorem 4. ■

It is worth noting that Theorem (4) is *general* and applies to *any* system with finite first-passage time. Now, we more carefully examine the mutual information $I(\vec{\mathbf{S}};\mathbf{T})$ to determine whether the limits implied by equation (23) and equation (24) exist and are bounded away from zero.

C. Mutual Information Between Input $\mathbf{T}$ and Output $\vec{\mathbf{S}}$

The mutual information between $\mathbf{T}$ and $\mathbf{S}$ is

$$I(\mathbf{S};\mathbf{T}) = h(\mathbf{S}) - h(\mathbf{S}|\mathbf{T}) = h(\mathbf{S}) - Mh(S|T) \tag{26}$$

Since the S_i given the T_i are mutually independent, $h(\mathbf{S}|\mathbf{T})$ does not depend on $f_{\mathbf{T}}(\mathbf{t})$. Thus, maximization of equation (26) is simply a maximization of $h(\mathbf{S})$ which is in turn a maximization of the marginal $h(S)$ over the marginal $f_T(t)$, a problem explicitly considered and solved in closed form for a mean launch time T_m constraint by Anantharam and Verdú [42] and for a launch deadline constraint in [45] and [46], both for exponential first-passage.

The corresponding expression for the mutual information between $\mathbf{T}$ and $\vec{\mathbf{S}}$ is

$$I(\vec{\mathbf{S}};\mathbf{T}) = h(\vec{\mathbf{S}}) - h(\vec{\mathbf{S}}|\mathbf{T}) \tag{27}$$

Unfortunately, $h(\vec{\mathbf{S}}|\mathbf{T})$ now *does* depend on the input distribution and the maximization of $h(\vec{\mathbf{S}})$ is non-obvious. So, rather than attempting a brute force optimization of equation (27) by deriving order distributions $f_{\vec{\mathbf{S}}}(\cdot)$ [37], we explore – *with no loss of generality* – simplifying symmetries.

Consider that an emission vector $\mathbf{t}$ and any of its permutations $P_n(\mathbf{t})$ produce statistically identical outputs $\vec{\mathbf{S}}$ owing to the reordering operation as depicted in FIGURES 1 and 2. Thus, any $f_{\mathbf{T}}(\cdot)$ which optimizes equation (27) can be "balanced" to form an optimizing input distribution which obeys

$$f_{\mathbf{T}}(\mathbf{t}) = f_{\mathbf{T}}(P_n(\mathbf{t})) \tag{28}$$

for $n = 1, 2, \ldots, M!$ and $P_n(\cdot)$ the previously defined permutation operator (see equation (3)). We can therefore restrict our search to hypersymmetric densities $f_{\mathbf{T}}(\mathbf{t})$ as defined by equation (28).

Now, hypersymmetric $\mathbf{T}$ implies hypersymmetric $\mathbf{S}$ which further implies that $f_{\mathbf{S}}(\mathbf{s}) = f_{\mathbf{S}}(P_n(\mathbf{s}))$. The same non-zero corner and folding argument used in the derivation of equation (4) produces the following key theorem:

Theorem 5 (The Entropy of $\vec{\mathbf{S}}$): If $f_{\mathbf{T}}(\cdot)$ is a hypersymmetric probability density function on emission times $\{T_m\}$, $m = 1, 2, \ldots, M$, and the first-passage density $g(\cdot)$ is non-singular, then the entropy of the time-ordered outputs $\vec{\mathbf{S}}$ is

$$h(\vec{\mathbf{S}}) = h(\mathbf{S}) - \log M!$$

Proof: Theorem 5 The hypersymmetry of $f_{\mathbf{S}}(\mathbf{s})$ implies

$$\begin{aligned} h(\vec{\mathbf{S}}) &= -\int_{\vec{\mathbf{s}}} M! f_{\mathbf{S}}(\vec{\mathbf{s}}) \log\big(M! f_{\mathbf{S}}(\vec{\mathbf{s}})\big) d\vec{\mathbf{s}} \\ &= -\log M! - \int_{\vec{\mathbf{s}}} M! f_{\mathbf{S}}(\vec{\mathbf{s}}) \log f_{\mathbf{S}}(\vec{\mathbf{s}}) d\vec{\mathbf{s}} \\ &= -\log M! - \int_{\mathbf{s}} f_{\mathbf{S}}(\mathbf{s}) \log f_{\mathbf{S}}(\mathbf{s}) d\mathbf{s} \\ &= -\log M! + h(\mathbf{S}) \end{aligned} \tag{29}$$

■

It is worth noting that hypersymmetric densities on $\mathbf{T}$ are completely equivalent (from a mutual information maximization standpoint) to their "unbalanced" cousins. Remember that each and every $I(\vec{\mathbf{S}}; \mathbf{T})$-maximizing $f_{\mathbf{T}}(\cdot)$ can be "balanced" and made into a hypersymmetric density without affecting the resulting value of $I(\vec{\mathbf{S}}; \mathbf{T})$. Likewise, any hypersymmetric density has a corresponding ordered density that produces the same $I(\vec{\mathbf{S}}; \mathbf{T})$. So, the assumption of hypersymmetric input densities is simply an analytic aid.

Next we turn to $h(\vec{\mathbf{S}}|\mathbf{T})$. A zero-measure edge-folding argument on the conditional density is not easily applicable here, so we resort to some information-theoretic sleight of hand. As before we define Ω as the permutation index number that produces an ordered output from $\mathbf{S}$ so that $P_\Omega(\mathbf{S}) = \vec{\mathbf{S}}$. We first note the equivalence

$$\left\{\Omega, \vec{\mathbf{S}}\right\} \Leftrightarrow \mathbf{S} \tag{30}$$

That is, specification of $\{\Omega, \vec{\mathbf{S}}\}$ specifies $\mathbf{S}$ and *vice versa* because as in our derivation of $h(\vec{\mathbf{S}})$, this equivalence requires that we exclude the zero-measure "edges" and "corners" of the density where two or more of the s_i are equal. Thus, there is no ambiguity in the $\mathbf{S} \to \vec{\mathbf{S}}$ map.

We then have,

$$h(\mathbf{S}|\mathbf{T}) = h(\Omega, \vec{\mathbf{S}}|\mathbf{T}) = h(\vec{\mathbf{S}}|\mathbf{T}) + H(\Omega|\vec{\mathbf{S}}, \mathbf{T}) \tag{31}$$

which also serves as an *en passant* definition for the entropy of a joint mixed distribution (Ω is discrete while $\vec{S}$ is continuous). We then rearrange equation (31) to prove a key theorem:

Theorem 6 (The Ordering Entropy, $H(\Omega|\vec{\boldsymbol{S}}, \boldsymbol{T})$):

$$h(\vec{\mathbf{S}}|\mathbf{T}) = h(\mathbf{S}|\mathbf{T}) - H(\Omega|\vec{\mathbf{S}}, \mathbf{T}) \tag{32}$$

where $H(\Omega|\vec{\mathbf{S}}, \mathbf{T})$, the *ordering entropy*, is the uncertainty about which S_m corresponds to which $\vec{S}_m$ given both $\mathbf{T}$ and $\vec{\mathbf{S}}$.

Proof: Theorem 6 See equation (31). ■

We note that

$$0 \le H(\Omega|\vec{\mathbf{S}}, \mathbf{T}) \le \log M! \tag{33}$$

with equality on the right for any singular density, $f_{\mathbf{T}}(\cdot)$, where all the T_m are equal with probability 1. We can then, after assuming that $f_{\mathbf{T}}(\cdot)$ is hypersymmetric, write the ordered mutual information in an intuitively pleasing form.

Theorem 7 (The Mutual Information $I(\vec{\boldsymbol{S}}; \boldsymbol{T})$ Relative to the Mutual Information $I(\boldsymbol{S}; \boldsymbol{T})$): For a hypersymmetric density $f_{\mathbf{T}}(\mathbf{t}) = f_{\mathbf{T}}(P_n(\mathbf{t}))$, $k = 1, 2, \ldots, M!$, the mutual information between launch times $\mathbf{T}$ and ordered arrival times $\vec{\mathbf{S}}$ satisfies

$$I(\vec{\mathbf{S}}; \mathbf{T}) = I(\mathbf{S}; \mathbf{T}) - \Big(\log M! - H(\Omega|\vec{\mathbf{S}}, \mathbf{T})\Big) \tag{34}$$

Proof: Theorem 7 Combine Theorem 5 and Theorem 6 with equation (27). ■

Put another way, an average *information degradation* of $\log M! - H(\Omega|\vec{\mathbf{S}}, \mathbf{T}) \ge 0$ is introduced by the sorting operation, $\mathbf{S} \to \vec{\mathbf{S}}$.

Mutual information is convex in $f_{\mathbf{T}}(\mathbf{t})$ and the space $\mathcal{F}_{\mathbf{T}}$ of feasible hypersymmetric $f_{\mathbf{T}}(\mathbf{t})$ is convex. That is, for any two hypersymmetric probability functions $f_{\mathbf{T}}^{(1)}$ and $f_{\mathbf{T}}^{(2)}$ we have

$$\kappa f_{\mathbf{T}}^{(1)}(\mathbf{t}) + (1 - \kappa) f_{\mathbf{T}}^{(2)}(\mathbf{t}) \in \mathcal{F}_{\mathbf{T}} \tag{35}$$

where $0 \le \kappa \le 1$. Thus, we can in principle apply variational [48] techniques to find that hypersymmetric $f_{\mathbf{T}}(\cdot)$ which attains the unique maximum of equation (34). However, in practice, direct application of this method leads to grossly infeasible $f_{\mathbf{T}}(\cdot)$, implying that the optimizing $f_{\mathbf{T}}(\cdot)$ has singular portions which lie along some "edges" or in some "corners" of the convex search space.

D. An Analytic Bound for Ordering Entropy $H(\Omega|\vec{\boldsymbol{S}}, \boldsymbol{T})$

The maximization of equation (34) hinges on specification of $H(\Omega|\vec{\mathbf{S}}, \mathbf{T})$, the *ordering entropy* given $\vec{\mathbf{S}}$ and $\mathbf{T}$. To determine analytic expressions for $H(\Omega|\vec{\mathbf{S}}, \mathbf{T})$, consider that given $\mathbf{t}$ and $\vec{\mathbf{s}}$, the probability that $\vec{\mathbf{s}}$ was produced by the k^{th} permutation of the underlying $\mathbf{s}$ is

$$\text{Prob}\big(\Omega = k\big|\vec{\mathbf{s}}, \mathbf{t}\big) = \frac{f_{\mathbf{S}|\mathbf{T}}\Big(P_k^{-1}(\vec{\mathbf{s}})\big|\mathbf{t}\Big)}{\sum_{n=1}^{M!} f_{\mathbf{S}|\mathbf{T}}\big(P_n(\vec{\mathbf{s}})\big|\mathbf{t}\big)} \tag{36}$$

where $\vec{\mathbf{s}} = P_k(\mathbf{s})$. Some permutations will have zero probability (are *inadmissible*) since the specific $\vec{\mathbf{s}}$ and $\mathbf{t}$ may render them impossible via the causality of $g(\cdot)$.

Using equation (5), the definition of entropy, and equation (36) we have

$$H\big(\Omega|\vec{\mathbf{s}}, \mathbf{t}\big) = -\sum_{n=1}^{M!} \left[\frac{\mathbf{g}\big(P_n(\vec{\mathbf{s}}) - \mathbf{t}\big)}{\sum_{j=1}^{M!} \mathbf{g}\big(P_j(\vec{\mathbf{s}}) - \mathbf{t}\big)}\right] \log \left[\frac{\mathbf{g}\big(P_n(\vec{\mathbf{s}}) - \mathbf{t}\big)}{\sum_{j=1}^{M!} \mathbf{g}\big(P_j(\vec{\mathbf{s}}) - \mathbf{t}\big)}\right] \tag{37}$$

and as might be imagined, equation (37) is difficult to work with in general. Nonetheless, let us define the number of non-zero terms in the sum of equation (37) as $|\Omega|_{\vec{\mathbf{s}},\mathbf{t}}$.

Now, consider that for exponential $g(\cdot)$, we can use equation (6) to write equation (36) as

$$\text{Prob}\big(\Omega = k|\vec{\mathbf{s}}, \mathbf{t}\big) = \frac{\mathbf{u}\Big(P_k^{-1}\big(\vec{\mathbf{s}}\big) - \mathbf{t}\Big)}{\sum_{n=1}^{M!}\mathbf{u}\big(P_n\big(\vec{\mathbf{s}}\big) - \mathbf{t}\big)} \tag{38}$$

where $\mathbf{u}(\cdot)$ is the multidimensional unit step function as previously defined. Equation (38) is a *uniform* probability mass function with $\sum_{n=1}^{M!}\mathbf{u}(P_n(\vec{\mathbf{s}}) - \mathbf{t}) = |\Omega|_{\vec{\mathbf{s}},\mathbf{t}}$ elements – the same as the number of non-zero terms in the sum of equation (37). Thus,

$$H(\Omega|\vec{\mathbf{s}}, \mathbf{t}) \leq \log \sum_{n=1}^{M!}\mathbf{u}\big(P_n\big(\vec{\mathbf{s}}\big) - \mathbf{t}\big) \tag{39}$$

for *all possible causal first-passage time densities*, $g(\cdot)$. In addition, it can be shown that exponential first-passage time is the *only* first-passage density which maximizes $H(\Omega|\vec{\mathbf{s}}, \mathbf{t})$, a result we state as a theorem.

Theorem 8 (A General Upper Bound for $H(\Omega|\vec{\mathbf{s}}, \mathbf{t})$): If we define the number of admissible combinations $\{P_n(\vec{\mathbf{s}}), \mathbf{t}\}$ as

$$|\Omega|_{\vec{\mathbf{s}},\mathbf{t}} \equiv \sum_{n=1}^{M!}\mathbf{u}\big(P_n\big(\vec{\mathbf{s}}\big) - \mathbf{t}\big)$$

where $\mathbf{u}(\cdot)$ is the multidimensional unit step function, then

$$H\big(\Omega|\vec{\mathbf{s}}, \mathbf{t}\big) \leq \log|\Omega|_{\vec{\mathbf{s}},\mathbf{t}}$$

with equality **iff** $g(\cdot)$ is exponential.

Proof: Theorem 8 We have already shown via equation (38) that exponential first-passage renders $\text{Prob}(\Omega = k|\vec{\mathbf{s}}, \mathbf{t})$ uniform.

Now, consider that the probability mass function (PMF) of equation (36) can be written as

$$\text{Prob}\big(\Omega = k|\vec{\mathbf{s}}, \mathbf{t}\big) = \frac{\mathbf{g}(P_k^{-1}\big(\vec{\mathbf{s}}\big) - \mathbf{t})}{\sum_{j=1}^{M!}\mathbf{g}\big(P_j\big(\vec{\mathbf{s}}\big) - \mathbf{t}\big)}$$

This PMF is uniform **iff** for all n and k where $P_n(\vec{\mathbf{s}})$ and $P_k(\vec{\mathbf{s}})$ *are both causal with respect to* $\boldsymbol{t}$ we have

$$\mathbf{g}\big(P_n\big(\vec{\mathbf{s}}\big) - \mathbf{t}\big) = \mathbf{g}\big(P_k\big(\vec{\mathbf{s}}\big) - \mathbf{t}\big) \tag{40}$$

That is, equation (40) must hold for all pairs $(P_n(\vec{\mathbf{s}}), \mathbf{t})$ and $(P_k(\vec{\mathbf{s}}), \mathbf{t})$ that are *admissible*. Since the maximum number of non-zero probability Ω is exactly the cardinality of admissible $(P_n(\vec{\mathbf{s}}), \mathbf{t})$, any density which produces a uniform PMF over admissible Ω thereby maximizes $H(\Omega|\vec{\mathbf{s}}, \mathbf{t})$, which proves the inequality.

We then note that any given permutation of a list can be achieved by sequential pairwise swapping of elements. Thus, equation (40) is satisfied **iff**

$$g(x_1 - t_1)g(x_2 - t_2) = g(x_2 - t_1)g(x_1 - t_2) \tag{41}$$

$\forall$ admissible $\{(x_1, x_2), (t_1, t_2)\}$. Rearranging equation (41) we have

$$\frac{g(x_1 - t_1)}{g(x_1 - t_2)} = \frac{g(x_2 - t_1)}{g(x_2 - t_2)}$$

which implies that

$$\frac{g(x - t_1)}{g(x - t_2)} = \text{Constant w.r.t. } x$$

Differentiation with respect to x yields

$$\frac{g'(x - t_1)}{g(x - t_2)} - \frac{g(x - t_1)g'(x - t_2)}{g^2(x - t_2)} = 0$$

which we rearrange to obtain

$$\frac{g'(x - t_1)}{g(x - t_1)} = \frac{g'(x - t_2)}{g(x - t_2)}$$

which further implies that

$$\frac{g'(x - t_1)}{g(x - t_1)} = c \tag{42}$$

since t_1 and t_2 are free variables. The only solution to equation (42) is

$$g(x) \propto e^{cx}$$

Thus, exponential $g(\cdot)$ is the only first-passage time density that can produce a maximum cardinality uniform distribution over Ω given $\vec{\mathbf{s}}$ and $\mathbf{t}$ – which completes the proof. ■

Now consider that $|\Omega|_{\vec{\mathbf{s}},\mathbf{t}}$, as defined in Theorem 8, is a hypersymmetric function of $\vec{\mathbf{s}}$ and $\mathbf{t}$ and thus invariant under any permutation of its arguments $\vec{\mathbf{s}}$ or $\mathbf{t}$. That is,

$$\begin{aligned}\sum_{n=1}^{M!}\mathbf{u}\big(P_n\big(\vec{\mathbf{s}}\big) - \mathbf{t}\big) &= \sum_{n=1}^{M!}\mathbf{u}\big(P_n\big(\vec{\mathbf{s}}\big) - \vec{\mathbf{t}}\big) \\ &= \sum_{n=1}^{M!}\mathbf{u}\big(P_n(\mathbf{s}) - \vec{\mathbf{t}}\big) \\ &= \sum_{n=1}^{M!}\mathbf{u}(P_n(\mathbf{s}) - \mathbf{t})\end{aligned}$$

because the summation is over all $M!$ permutations. Therefore,

$$|\Omega|_{\vec{\mathbf{s}},\mathbf{t}} = |\Omega|_{\vec{\mathbf{s}},\vec{\mathbf{t}}} = |\Omega|_{\mathbf{s},\vec{\mathbf{t}}} = |\Omega|_{\mathbf{s},\mathbf{t}} \tag{43}$$

We must now enumerate this number of admissible permutations. Owing to equation (43) and Theorem 1 we can assume time-ordered inputs $\vec{\mathbf{t}}$ with no loss of generality. So, let us define contiguous "bins" $\mathcal{B}_k = \{t|t \in [\vec{t}_k, \vec{t}_{k+1})\}$, $k = 1, 2, \ldots, M$ $(\vec{t}_{M+1} \equiv \infty)$ and then define σ_m as bin occupancies. That is, $\sigma_m = q$ if there are exactly q arrivals in $\mathcal{B}_m$. The benefit of this approach is that the σ_m do not depend on whether $\vec{\mathbf{s}}$ or $\mathbf{s}$ is used to count the arrivals. Thus, expectations can be taken over $\mathbf{S}$ whose components are mutually independent given the $\mathbf{t}$ and no order distributions for $\vec{\mathbf{S}}$ need be derived.

To determine the random variable $|\Omega|_{\mathbf{S},\vec{\mathbf{t}}}$ we start by defining

$$\eta_m = \sum_{j=1}^{m}\sigma_j$$

the total number of arrivals up to and including bin $\mathcal{B}_m$. Clearly η_m is monotonically increasing in m with $\eta_0 = 0$ and $\eta_M = M$. We then observe that the σ_m arrivals on $[\vec{t}_m, \vec{t}_{m+1})$ can be assigned to any of the $\vec{t}_1, \vec{t}_2, \ldots, \vec{t}_m$ known emission times

except for those η_{m-1} previously assigned. The number of possible new assignments is $(m-\eta_{m-1})!/(m-\eta_m)!$ which when applied iteratively leads to

$$|\Omega|_{\mathbf{S},\vec{\mathbf{t}}} = \prod_{m=1}^{M} \frac{(m-\eta_{m-1})!}{(m-\eta_m)!} = \prod_{m=1}^{M-1} (m+1-\eta_m) \tag{44}$$

We then define the random variable

$$X_i^{(m)} = \begin{cases} 1 & S_i < \vec{t}_{m+1} \\ 0 & \text{otherwise} \end{cases}$$

for $i = 1, 2, \ldots m$. The PMF of $X_i^{(m)}$ is then

$$p_{X_i^{(m)}}(x) = \begin{cases} G\big(\vec{t}_{m+1} - \vec{t}_i\big) & x = 1 \\ \bar{G}\big(\vec{t}_{m+1} - \vec{t}_i\big) & x = 0 \end{cases} \tag{45}$$

where we note that for a given m, $X_i^{(m)}$ and $X_j^{(m)}$ are independent, $i \neq j$, and as previously defined, $G(\cdot)$ is the CDF of the causal first-passage density $g(\cdot)$. $\bar{G}(\cdot) = 1 - G(\cdot)$ is the corresponding CCDF. We can then write

$$\eta_m = \sum_{i=1}^{m} X_i^{(m)}$$

It is then convenient to define $\bar{X}_i = 1 - X_i$ which allows us to define $\bar{\eta}_m = m - \eta_m$. We can then write

$$|\Omega|_{\mathbf{S},\vec{\mathbf{t}}} = \prod_{m=1}^{M-1} (1 + \bar{\eta}_m) \tag{46}$$

Since we seek the expected value of $|\Omega|_{\mathbf{S},\vec{\mathbf{t}}}$, we can use equation (45) to calculate each individual $E_{\mathbf{S}|\vec{\mathbf{t}}}[\log(1+\bar{\eta}_m)]$ as

$$\sum_{\bar{\mathbf{x}}} \log\left(1 + \sum_{i=1}^{m} \bar{x}_i\right) \prod_{j=1}^{m} \bar{G}^{\bar{x}_j}\big(\vec{t}_{m+1} - \vec{t}_j\big) G^{1-\bar{x}_j}\big(\vec{t}_{m+1} - \vec{t}_j\big) \tag{47}$$

which allows us to define $H^{\uparrow}(\mathbf{t})$, an upper bound on $H(\Omega|\vec{\mathbf{S}}, \mathbf{t})$, as

$$H^{\uparrow}(\mathbf{t}) \equiv \sum_{m=1}^{M-1} \sum_{\bar{\mathbf{x}}} \log\left(1 + \sum_{i=1}^{m} \bar{x}_i\right) \times \prod_{j=1}^{m} \bar{G}^{\bar{x}_j}\big(\vec{t}_{m+1} - \vec{t}_j\big) G^{1-\bar{x}_j}\big(\vec{t}_{m+1} - \vec{t}_j\big) \tag{48}$$

where an ordering permutation on $\mathbf{t}$ provides the $\{\vec{t}_j\}$ in equation (48) which can then be rearranged as

$$H^{\uparrow}(\mathbf{t}) = \sum_{\ell=1}^{M-1} \log(1+\ell) \times \sum_{m=\ell}^{M-1} \sum_{|\bar{\mathbf{x}}|=\ell} \prod_{j=1}^{m} \bar{G}^{\bar{x}_j}\big(\vec{t}_{m+1} - \vec{t}_j\big) G^{1-\bar{x}_j}\big(\vec{t}_{m+1} - \vec{t}_j\big) \tag{49}$$

We then note that

$$\begin{aligned} H\Big(\Omega|\vec{\mathbf{S}}, \mathbf{T}\Big) &\equiv E_{\mathbf{T}}\Big[E_{\vec{\mathbf{S}}|\mathbf{T}}\big[H\big(\Omega|\vec{\mathbf{s}}, \mathbf{t}\big)\big]\Big] \\ &\leq E_{\mathbf{T}}\Big[E_{\vec{\mathbf{S}}|\vec{\mathbf{T}}}\big[\log|\Omega|_{\vec{\mathbf{s}},\mathbf{t}}\big]\Big] \\ &= E_{\vec{\mathbf{T}}}\Big[E_{\mathbf{S}|\vec{\mathbf{T}}}\Big[\log|\Omega|_{\mathbf{s},\vec{\mathbf{t}}}\Big]\Big] \end{aligned}$$

follows from equation (49) in conjunction with Theorem 8 and through hypersymmetric expectations (Theorem 1) of hypersymmetric functions $|\Omega|_{\mathbf{s},\vec{\mathbf{t}}}$ (equation (43)). Adding in the result of Theorem 8 we have proven the following theorem.

Theorem 9 (A General and Computable Upper Bound for $H(\Omega|\vec{\boldsymbol{S}}, \boldsymbol{T})$):

$$E\big[H\big(\Omega|\vec{\mathbf{s}}, \mathbf{t}\big)\big] \equiv H\Big(\Omega|\vec{\mathbf{S}}, \mathbf{T}\Big) \leq H^{\uparrow}(\mathbf{T}) \tag{50}$$

with equality **iff** the first-passage time density $g(\cdot)$ is exponential.

Proof: Theorem 9 See the development leading to the statement of Theorem 9. Theorem 8 establishes equality **iff** the first-passage density is exponential. ∎

Theorem 9 gives us $H^{\uparrow}(\mathbf{T})$, a computable analytic upper bound for $H(\Omega|\vec{\mathbf{S}}, \mathbf{T})$, and an exact expression if the first-passage time is exponential.

E. Capacity Bounds for Timing Channels

Despite significant effort, direct optimization of mutual information, $I(\vec{\mathbf{S}}; \mathbf{T})$ (see equation (34)) remained elusive. The key issue is that $h(\mathbf{S})$ and $H(\Omega|\mathbf{T}, \vec{\mathbf{S}})$ are "conflicting" quantities with respect to $f_{\mathbf{T}}(\cdot)$. That is, independence of the $\{T_m\}$ favors larger $h(\mathbf{S})$ (i.e., $h(\mathbf{S}) \leq \sum_m h(S_m)$) while tight correlation of the $\{T_m\}$ (as in $T_i = T_j$, $i, j = 1, 2, \ldots, M$) produces the maximum $H(\Omega|\vec{\mathbf{S}}, \mathbf{T}) = \log M!$. In light of these difficulties, we sought analytic expressions in the companion to this paper (Part II [46]) for the maximum $h(S)$ and $I(S; T)$ which we restate here as Theorem 10 and Theorem 11 without proof, recalling that $\tau = \tau(M)$.

Theorem 10 (Maximum h(S) for Exponential First-Passage Under a Launch Deadline Constraint): For first-passage time D with density $f_D(d) = g(d) = \mu e^{-\mu d}$, and launch time T constrained to $[0, \tau]$, the maximum entropy of $S = T + D$ is

$$\max_{f_T(\cdot)} h(S) = \log\left(\frac{e + \mu\tau}{\mu}\right) \tag{51}$$

The input density $f_T(\cdot)$ which produces the maximum $h(S)$ is

$$f_T(t) = \delta(t)\frac{1}{e+\mu\tau} + \delta(t-\tau)\frac{1-e}{e+\mu\tau} + \frac{\mu}{e+\mu\tau}(u(t) - u(t-\tau)) \tag{52}$$

Theorem 11 (Maximum I(S; T) for Exponential First-Passage Under a Launch Deadline Constraint): For first-passage time D with density $f_D(d) = g(d) = \mu^{-\mu d}$, and launch time T constrained to $[0, \tau]$, the maximum mutual information between $S = T + D$ and T is

$$\max_{f_T} I(S; T) = \log\Big(1 + \frac{\mu\tau}{e}\Big) \tag{53}$$

The definition of C_q and C_t in Theorem 4 requires we consider the asymptotic value of $H(\Omega|\vec{\mathbf{S}}, \mathbf{T})/M$. A lower bound is provided in Part II [46] assuming exponential first-passage, a result we restate here as Theorem 12 without proof.

Theorem 12 (Asymptotic $H(\Omega|\vec{\boldsymbol{S}},\boldsymbol{T})/M$ for Exponential First-Passage Under a Launch Deadline Constraint $\boldsymbol{T} \in [\boldsymbol{0}, \boldsymbol{\tau}]$): For exponential first-passage with mean $1/\mu$, token launch intensity λ, and i.i.d. input distribution $f_{\mathbf{T}}(\mathbf{t}) =$

$\prod_{m=1}^{M} f_T(t)$ where $f_T(\cdot)$ maximizes $I(S;T)$ as in Theorem 11, the asymptotic ordering entropy per token is

$$\lim_{M\to\infty} \frac{H\left(\Omega|\vec{\mathbf{S}},\mathbf{T}\right)}{M} = \sum_{k=2}^{\infty} e^{-\rho}\frac{\rho^k}{k!}\left(\frac{k}{\rho}-1\right)\log k! \tag{54}$$

where $\rho = \lambda/\mu$ is defined as a measure of system token "load" similar to a queueing system.

We can rewrite the summation term in equation (54) more compactly noting that

$$\sum_{k=1}^{\infty}\frac{\rho^k}{k!}\left(\frac{k}{\rho}-1\right)\log k! = \sum_{\ell=1}^{\infty}\log\ell\sum_{k=\ell}^{\infty}\frac{\rho^k}{k!}\left(\frac{k}{\rho}-1\right) \tag{55}$$

Then

$$\sum_{k=\ell}^{\infty}\rho^k\frac{1}{k!} = e^{\rho} - \sum_{k=0}^{\ell-1}\rho^k\frac{1}{k!} \tag{56}$$

and

$$\sum_{k=\ell}^{\infty}\frac{k}{\rho}\rho^k\frac{1}{k!} = \sum_{k=\ell-1}^{\infty}\rho^k\frac{1}{k!}$$

can be used to obtain

$$\sum_{k=\ell}^{\infty}\frac{\rho^k}{k!}\left(\frac{k}{\rho}-1\right) = \frac{1}{(\ell-1)!}\rho^{\ell-1} = \ell\rho^{\ell}\frac{1}{\ell!\rho}$$

We then note that

$$p_\ell = e^{-\rho}\frac{\rho^\ell}{\ell!} \tag{57}$$

$\ell = 0, 1, \ldots, \infty$ is a Poisson probability mass function with parameter ρ and obtain the more compact

$$\sum_{k=1}^{\infty}\rho^k(k/\rho-1)\frac{\log k!}{k!} = \frac{1}{\rho}E_\ell\big[\ell\log\ell\big] \tag{58}$$

Now turning toward capacity, equation (34) and Theorem 11 are easily combined to show

$$\frac{1}{M}I(\mathbf{S};\mathbf{T}) - \frac{1}{M}\log M! \geq \log\left(1+\frac{\mu\tau}{e}\right) - \frac{1}{M}\log M!$$

Then, since $\tau = M/\lambda$ we have

$$\lim_{M\to\infty}\frac{1}{M}I(\mathbf{S};\mathbf{T}) - \frac{1}{M}\log M! \geq \log\frac{1}{\rho} \tag{59}$$

Noting that $I(\vec{\mathbf{S}};\mathbf{T}) = I(\mathbf{S};\mathbf{T}) - \log M! + H(\Omega|\vec{\mathbf{S}},\mathbf{T})$ and $H(\Omega|\vec{\mathbf{S}},\mathbf{T}) \geq 0$ proves the following theorem:

Theorem 13 (A Simple Lower Bound for C_q under Exponential First-Passage): Given a token launch intensity $\lambda = M/\tau$ and exponential first-passage time distribution with mean $1/\mu$, the identical-token timing channel capacity $C_q(\rho)$ in nats per token obeys

$$C_q(\rho) \geq \max\{-\log\rho, 0\} \tag{60}$$

where $\rho = \frac{\lambda}{\mu}$

Proof: Theorem 13 See the development leading to the statement of Theorem 13. ∎

We can, however, combine equation (59), Theorem 12 and equation (58) to obtain a better lower bound on capacity.

Theorem 14 (Lower Bound for C_q and C_t for Exponential First-Passage): In the limit of large M, with mean $1/\mu$ exponential first-passage, the identical-token channel capacities, C_q and C_t must obey

$$C_q(\rho) \geq \log\frac{1}{\rho} + \frac{1}{\rho}E\big[\ell\log\ell\big] \tag{61}$$

and

$$C_t(\rho) \geq \lambda\left(\log\frac{1}{\rho} + \frac{1}{\rho}E\big[\ell\log\ell\big]\right) \tag{62}$$

where ℓ is Poisson with PMF

$$p_\ell = e^{-\rho}\frac{\rho^\ell}{\ell!} \tag{63}$$

where $\rho = \lambda/\mu$.

Proof: Theorem 14 Combine equation (59), Theorem 12 and equation (58). ∎

Finally, from Part II [41], [46] we have the following upper bound on C_q and the concomitant bound on $C_t = \lambda C_q$ as.

Theorem 15 (Upper Bound for C_q and C_t for Exponential First-Passage): If the first-passage density $f_D(\cdot)$ is exponential with parameter μ and the rate at which tokens are released is λ, then the capacity per token, C_q is upper bounded by

$$C_q \leq \log\left(\frac{1}{\rho}+4\right) \tag{64}$$

and the capacity per unit time is upper bounded by

$$C_t \leq \lambda\log\left(\frac{1}{\rho}+4\right) \tag{65}$$

where $\rho = \frac{\lambda}{\mu}$

Proof: Theorem 15 Theorem 11 in Part II [46] provides the bound for C_q and application of Theorem 4 provides the bound for C_t. ∎

We have now concluded our information-theoretic treatment of identical-token timing channel. In the next sections we show how these results can be applied to molecular communication channels where tokens can carry information payloads.

F. Tokens With Payloads

In Sections IV-A to IV-E we developed all the machinery necessary to provide capacity bounds for channels with identical-tokens where timing is the only means of information carriage. However, one can also imagine scenarios where the token itself carries information, much as a "packet" carries information over the Internet. For example, consider a token that is a finite string of symbols over a finite alphabet. Having constructed tokens from these "building blocks," a sender launches them into the channel and they are captured by a receiver. In this scenario a DNA sequence is a symbolic string drawn from a 4-character alphabet so that each nucleotide could carry 2 bits of information. Similarly, a protein sequence is a symbolic string drawn from a 20-character alphabet so that each amino acid could carry a little over 4 bits of information. Thus, a DNA token constructed from 100 nucleotides would carry 200 bits whereas a corresponding protein token would carry > 400 bits.

However, there are myriad other possibilities for coding information in structure. For example, a third major class of biological macromolecules, carbohydrates (polysaccharides), are linear and branched structures where the polymers are constructed from a much larger alphabet of monosaccharides. In addition to the composition information inherent in the makeup of a linear or non-linear concatenations of building blocks, one could imagine a layer of structural information as well [49] – as is the case with biological macromolecules where the spatiotemporal architecture of a polymer is as important as the order and frequencies of nucleotide, amino acid or monosaccharide residues of which it is composed. However, as the issue of "structural" information (the amount of information contained in an arbitrary 3-dimensional object) is as yet an open problem, we will not consider such constructions in detail. Yet and still, the bounds we will derive are applicable to *any* method of information transfer wherein tokens carry information payload, either as string sequences or in some other structural way.

So, for now consider only string tokens – as exemplified by DNA and protein sequences – where each token in the ensemble released by the sender carries a portion of the message. Thus, irrespective of their individual lengths, such "inscribed matter" tokens must be "strung together" to recover the original message, which implies that *each token must be identifiable*. Just as in human engineered systems like the Internet where information packets could arrive out of order, a sequence number could be appended to each token to ensure proper reconstruction at the destination. Thus, given M tokens per channel use, we could append $\log M$ bits to each token. We will defer detailed discussion of this scenario until Section VI as this approach is asymptotically impractical with $\log M$ tending toward ∞. Alternatively, one could employ gross differences to convey sequence information such as sending tokens of distinct lengths $1, 2, \ldots, K$ where $M = K(K+1)/2$. And there may be other clever ways to embed structural side-information to establish token order. Nonetheless, the myriad possibilities notwithstanding, $H(\Omega|\vec{\mathbf{S}}, \mathbf{T})$ provides the measure of essential token "overhead" or "side-information" (of any form) necessary to maintain proper sequence.

Consider that operation of the timing channel involves construction of "blocks" $\{\mathbf{t}_1, \ldots, \mathbf{t}_N\}$ where each $\mathbf{t}_n$ represents the launch schedule for M tokens (a channel use). These blocks, "codewords" of blocklength N, are launched into the channel. If capacity is not exceeded, the receiver can reliably recover the information embedded in the codewords and since we generally assume the receiver has access to the coding method, a correctly decoded message implies knowledge of the codewords $\{\mathbf{t}_1, \ldots, \mathbf{t}_N\}$. However, the channel imposes residual uncertainty about the mapping $\mathbf{S} \to \vec{\mathbf{S}}$ – the ambiguity about which $\vec{s}_i$ is associated with which s_j. For this reason, the payload-inscribed tokens cannot yet be correctly strung together to recover the message.

However, given the observed arrivals $\vec{\mathbf{s}}$ and the correctly decoded $\mathbf{t}$, $H(\Omega|\vec{\mathbf{s}}, \mathbf{t})$ is *the definition* of the uncertainty about that ordering, Ω. Likewise, the average uncertainty is $H(\Omega|\vec{\mathbf{S}}, \mathbf{T})$. Thus, the source coding theorem implies that at least $H(\Omega|\vec{\mathbf{S}}, \mathbf{T})$ bits must be used, on average, to resolve the mapping ambiguity.

So, consider a message $\mathbf{P}$ to be carried as token payload that we break into equal size ordered submessages p_m, $m = 1, 2, \ldots, M$. We can summarize the previous discussion as a theorem.

Theorem 16 (Sequencing Tokens With Payload): If a message P is broken into equal size "payload" submessages $\{p_m\}$, $m = 1, 2, \ldots, M$ and inscribed into otherwise identical-tokens launched at times $\{t_m\}$, we must provide, on average, additional "sequencing information" $\frac{1}{M}H(\Omega|\vec{\mathbf{S}}, \mathbf{T})$ per token at the receiver to assure recovery of the full payload message $\mathbf{P} = p_1 p_2 \cdots p_M$.

Proof: Theorem 16 Given arrivals $\vec{\mathbf{s}}$ and known departures $\mathbf{t}$, the uncertainty about the mapping between the $\{\vec{s}_m\}$ and the $\{s_m\}$ (and thus the associated $\{t_m\}$) is exactly the ordering entropy $H(\Omega|\vec{\mathbf{s}}, \mathbf{t})$. Considering Ω as a letter from a random i.i.d. source, the source coding theorem [10], [50] requires at least $H(\Omega|\vec{\mathbf{s}}, \mathbf{t})$ bits on average to uniquely specify Ω – or asymptotically over many channel uses, at least $H(\Omega|\vec{\mathbf{S}}, \mathbf{T})$ bits on average. Therefore the information necessary at the receiver to recover the proper sequence and thence the message P is greater than or equal to $\frac{1}{M}H(\Omega|\vec{\mathbf{S}}, \mathbf{T})$ per token. ■

It is important to note that we have not actually provided a *method* for message reconstruction, only a lower bound on the amount of "side information" necessary at the receiver to assure proper reconstruction. However, as a practical matter, the quantity $\frac{1}{M}H(\Omega|\vec{\mathbf{S}}, \mathbf{T})$ does provide some guidance. In the worst case where the order of token arrival is completely random, $H(\Omega|\vec{\mathbf{S}}, \mathbf{T}) = \log M!$ which amounts to each packet carrying a header of size $\frac{1}{M}\log M! \approx \log M$ for large M – essentially numbering the packets from 1 to M. If $\frac{1}{M}H(\Omega|\vec{\mathbf{S}}, \mathbf{T})$ is much smaller, then one could imagine some type of cyclic packet numbers since smaller $\frac{1}{M}H(\Omega|\vec{\mathbf{S}}, \mathbf{T})$ implies that some sets of packets are unlikely to arrive out of order. The sequence header could then be commensurately smaller. In either case, the total amount information necessary to resolve the ordering Ω is $\frac{1}{M}H(\Omega|\vec{\mathbf{S}}, \mathbf{T})$ per packet on average.

G. Energy Costs

System energy is a critical resource which limits capacity in all communication systems. In the case of molecular communication, there are a variety of potential costs, most notably manufacture, launch and transport. So, assume the minimum cost of fabricating a token without a payload is c_0 Joules and with a payload c_1 Joules. Symbolic string tokens incur a "per character" cost which we define as Δc_1 per character per token. For example, adding a nucleotide to double-stranded DNA requires 2 ATP (1.6×10^{-19} J) while adding an amino acid to a protein requires 4 ATP (3.2×10^{-19} J) [51]. Apart from the per residue per token cost, there may be other energy involved in sequestration, release and/or token transport across a gap. However, the key assumption is constant energy use per token. Without considering the details as in [52] we will denote the combination of these and any other relevant energies as c_e Joules per token. Thus, the power required for the timing-only channel is

$$\mathcal{P}_T = \lambda(c_0 + c_e) \tag{66}$$

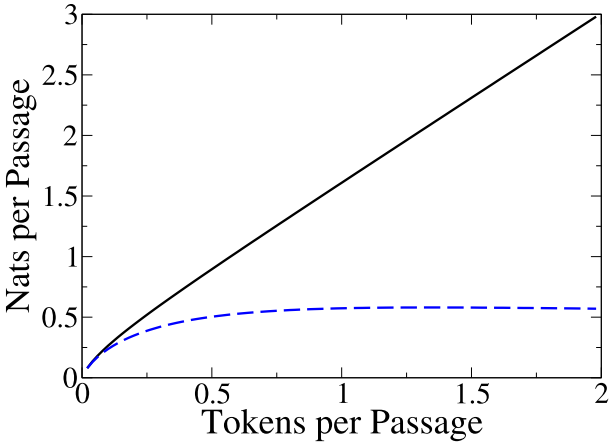

Fig. 4. Lower bound (dashed line: Theorem 14) and upper bound (solid line: Theorem 15) for the identical-token timing channel capacity C_t (in nats per passage time $1/\mu$) as a function of channel load ρ, the ratio of the token launch rate λ to the token uptake rate μ. Exponential first-passage assumed.

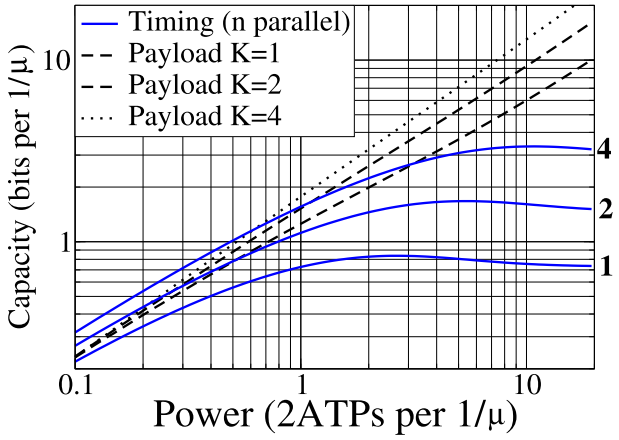

Fig. 5. Lower bounds for the capacities of the token timing (ρC_T) and token timing plus token payload (ρC_{P+T}) channels as a function of power budget ($\mathcal{P}_T$ and $\mathcal{P}_{P+T}$) for DNA string tokens with exponential first-passage times (Theorem 14). Capacity is in units bits per first-passage time $1/\mu$. Power is in units of 2-ATP (1.6×10^{-19} J) per passage time $1/\mu$ and a nucleotide residue is assumed to carry 2-bits of information. Solid lines: aggregate capacity of $n = 1, 2, 4$ separate (independent or parallel) token timing channels where DNA string tokens carry no information payload. Dashed/Dotted lines: aggregate capacity of token timing plus token payload channels for DNA string tokens of different lengths, $K = 1, 2, 4$-residue payloads. Exponential first-passage assumed.

and for the timing-plus-payload channel,

$$\mathcal{P}_{T+P} \leq \lambda\left(c_1 + c_e + \left(\frac{H^{\uparrow}(\mathbf{T})}{\log b} + K\right)\Delta c_1\right) \tag{67}$$

where K is the string length of information-laden tokens, and b is the alphabet size used to construct the strings. The inequality in equation (67) results from the fact that $H^{\uparrow}(\mathbf{T})$ is an upper bound on the ordering entropy $H(\Omega|\vec{\mathbf{S}}, \mathbf{T})$ over all possible first-passage distributions (Theorem 9).

We also note that detailed timing information could be ignored so that information is conveyed by payload using embedded sequencing information. The power budget would be identical to that of equation (67) except that $H^{\uparrow}(\mathbf{T})$ would be replaced by $\lim_{M\to\infty} \min_{\mathbf{t}} H(\Omega|\vec{\mathbf{S}}, \mathbf{t})/M$. However, since $H(\Omega|\vec{\mathbf{S}}, \mathbf{T}) \geq \min_{\mathbf{t}} H(\Omega|\vec{\mathbf{S}}, \mathbf{t})$, equation (67) provides an upper bound for the payload-only power as well.

V. Results

We can now define the capacities for the token-timing (only), token-timing+payload and token-payload (only) channels as follows:

$$\mathcal{C}_T = \lambda C_q(\rho) \tag{68}$$

$$\mathcal{C}_{T+P} = \lambda\big(C_q(\rho) + K\log b\big) \tag{69}$$

and

$$\mathcal{C}_P = \lambda K \log b = \mathcal{C}_{T+P} - \mathcal{C}_T \tag{70}$$

In FIGURE 4 we use Theorem 14 and Theorem 15 (exponential first-passage) to plot lower and upper bounds for $\mathcal{C}_T$ versus ρ, a proxy for power budget, $\mathcal{P}$, assuming some unit cost per token ($c_0 + c_e = 1$). *It is important to note* that first-passage time *variance* (jitter) produces disordered tokens. That is, the mean first-passage time is only a measure of channel *latency* – the "propagation delay" so to speak – and does not itself impact token order uncertainty. However, for exponential first-passage the standard deviation also happens to be the first-passage time $1/\mu$. At small values of ρ the bounds are tight. At larger ρ the bounds diverge and the upper bound offers the tantalizing hint that timing channel capacity increases with increased token load ρ (see also [8], [9]). Unfortunately, we have as yet been unable to find an empirical density $f_{\mathbf{T}}(\mathbf{t})$ which displays capacity growth similar to the upper bound and suspect that timing capacity flattens with increasing ρ owing to a more rapidly increasing probability of token order confusion at the output.

In FIGURE 5, we use Theorem 14 to plot lower bounds in bits per first-passage time, $1/\mu$, as a function of power budget $\mathcal{P}$ assuming DNA-based tokens. For token timing plus token payload signaling we show plots for $K = 1, 2, 4$ DNA-residue tokens. For timing-only signaling we also include plots where different identifiable tokens (different molecule types or physically separate channels) are used (*i.e.,* $n = 1, 2, 4$ parallel timing channels as shown) for comparison with payload channels. We have assumed costs $c_0 = \Delta c_1 = 2$ ATP. Furthermore, we assume $c_1 = c_e = \Delta c_1$ since it seems likely that the absolute minimum energy for token release, c_e, in a purely diffusive channel is probably comparable to the cost of creating (or breaking) the covalent bond used to append a nucleotide residue. If we assume $1/\mu = 1$ms, then the ordinate of FIGURE 5 is in kbit/s and the abscissa is in units of 1.6×10^{-16}W. If $1/\mu = 1\mu s$, (as might be the case for smaller gaps in a nano-system) the ordinate is in $Mbit/s$ and the abscissa is in units of 1.6×10^{-13}W. These data rates are many orders of magnitude larger than the fractional bit/second data rates previously reported for simple demonstrations of molecule communication [53], and the predicted power efficiencies are startling. Comparison of our results to [53] and others would be relatively straightforward if passage time jitter for the experimental setup were provided, although in [53] Avogradrian numbers of molecules were released with each alcohol "puff" so precise timing at the molecular level was not attempted.

Finally, it is worth noting that increasing the rate at which tokens with payload are launched will increase the bit rate but not increase the required energy per bit. Of particular note, at low power, timing-only signaling provides the best rates while at higher power, inscribed matter tokens may be preferred. However, if it is difficult to synthesize long strings (heavily information-laden tokens), even a single bit of payload information (two distinguishable species used in parallel) markedly increases capacity.

VI. Discussion and Conclusion

We have provided a general and fundamental mathematical framework for molecular channels and derived some associated capacity bounds. We now discuss the results in the context of selected prior work and touch upon ideas for further work. First, we examine the engineering implications: how molecular communication can be extended to other known communication scenarios as well where we might look for inspiration from biology that has had eons to evolve solutions. Then, we examine the biological implications: how the results might impact/support known biology and suggest new avenues for investigation.

A. Engineering Implications

1) Capacity Bounds and Coding Methods: Our upper bound on capacity C_t, the timing capacity for identical-tokens, is tight for low token load ρ but diverges for large ρ. However, no empirical distributions which provide rates higher than the lower bound have yet been found. So, does the capacity of the timing-only channel truly flatten with increasing ρ as in FIGURES 5 and 4, or is there a benefit to increasing the intensity of timing token release as suggested in [8] and [9]? Intriguingly, this rate flattening of the lower bound seems to comport almost exactly with a result in [44, Sec. III] which considers a related system where only a finite number of tokens can be simultaneously in flight (a finite "server" system in the parlance of [42] and [44]). More careful exploration of these parallels may reveal launch densities which cause C_t to grow without bound with ρ.

Since exponential first-passage is not the worst case corruption, what *is* the minmax capacity of the molecular timing channel? Likewise, how much better than exponential might other first-passage densities imposed by various physical channels be, and what are good codes for reliable transmission of information over molecular channels?

While we have focused on tokens that are linear strings (DNA and protein sequences), what benefits might strings with a branched structure (exemplified by carbohydrates) confer for token ordering and/or payload? Should we pursue technology to produce large payload tokens (strings with many residue) [54], or should a pool of smaller pre-fabricated payloads be used to deliver information? The bunching seen in FIGURE 5 for payload tokens with increasing K may suggest the latter when rapid token construction is difficult. That is, the capacity per power output does not scale linearly in K owing to the increased power required to add more bases to tokens. This implies a tradeoff between timing only and increasingly larger payloads – completely aside from the fact that payload size, shape and composition can affect token transport and capture.

2) Precise Timing, Fuzzy Timing and Concentration: It is important to quantify the relationship between our fine grain timing model and other less temporally precise ones [1]–[7], [35]–[38]. Our model seems to imply infinitely precise control over the release times $\mathbf{T}$ and infinitely precise measurement of the arrival times $\vec{\mathbf{S}}$. However, imprecision in both times can be incorporated easily into the transit time vector $\mathbf{D}$. Thus, applying our model to the "fuzzier" release and detection times associated with practical/real systems is straightforward. That is, first-passage time jitter already imposes limits on timing precision. So long as launch and capture timing precision is significantly better than passage time jitter, the bounds presented here will be moderately tight. In addition, we are hopeful that the upper bound of Theorem 15 will be useful for evaluating molecular timing channel capacity for arbitrary first-passage time distributions since it requires only knowledge of the timing channel capacity coupled to average properties of the corresponding input distribution.

Concentration is derived from counting arrivals within temporal windows. The data processing theorem indicates that our precise timing model **must** undergird all concentration based methods which, even with perfect concentration detection, *cannot possibly exceed the capacity of the fine grain timing model presented here*. Given the asymptotic nature of our analysis, an individual emission schedule $\mathbf{t}$ for large M is *exactly* a temporal emission concentration profile as time resolution coarsens. Thus, the results here provide crisp upper bounds on the capacities derived from concentration based models.

Consider channel models where information is carried by the *number* of molecules released and received (for recent examples, see [9], [55]). In this case, the capacity per channel use, C_N, is upper bounded by $\log(M+1)$ since between 0 and M tokens can be released during a symbol interval of duration $\tau(M)$. Smaller $\tau(M)$ increases the capacity in bits per second whereas larger M increases the capacity in bits per channel use.

If we assume a fixed signaling interval $\tau(M)$ during which $m = 0, 1, \ldots M$ tokens are emitted, we can use the average token launch rate λ as a proxy for power. Assuming a uniform distribution on the number of tokens sent, we have

$$\tau(M) = \frac{M}{2\lambda} \tag{71}$$

since the average number of tokens released is $M/2$.

Assuming exponential first-passage, the probability that all tokens arrive by $\tau(M)$ is minimized when all tokens are launched at $t = 0$. For exponential first-passage and with arrival probability criterion $1-\epsilon$ as in Section IV-A, we have

$$\tau(M) = -\frac{1}{\mu}\log\left(1-(1-\epsilon)^{\frac{1}{M}}\right) \tag{72}$$

which assures that even when M tokens are emitted, they will all arrive before $\tau(M)$ with probability $1-\epsilon$.

However, equation (72) in combination with the power limit of equation (71) sets λ to

$$\lambda(M) = -\frac{\mu M}{2\log\left(1-(1-\epsilon)^{\frac{1}{M}}\right)} \tag{73}$$

Then, since $C_N \leq \log(M+1)$, after setting a successful channel use criterion of $1-\epsilon$ we have

$$\frac{C_N(M)}{\tau(M)} \leq -\frac{\mu \log(M+1)}{\log\left(1-(1-\epsilon)^{\frac{1}{M}}\right)} \tag{74}$$

which we rewrite as

$$\frac{C_N(M)}{\mu\tau(M)} \leq -\frac{\log(M+1)}{\log\left(1-(1-\epsilon)^{\frac{1}{M}}\right)} \tag{75}$$

with normalized power constraint

$$\mathcal{P}(M) = \frac{\lambda(M)}{\mu} = -\frac{M}{2\log\left(1-(1-\epsilon)^{\frac{1}{M}}\right)} \tag{76}$$

However, except for small ϵ, $\frac{C_N(M)}{\mu\tau(M)}$ is not a reliable indicator of capacity since with increased ϵ, $\tau(M)$ decreases but the probability of intersymbol interference (ISI) increases. Since calculating the capacity of this channel with ISI is difficult, we roughly approximate by normalizing $\frac{C_N(M)}{\mu\tau(M)}$ by the expected number of intervals over which a given emission burst of M tokens will span (thereby potentially corrupting them). Noting that $(1-\epsilon^z)^M$ is the probability that all tokens arrive before the end of the $z+1^{\text{st}}$ interval after emission we have

$$\begin{aligned}\bar{z}(M) &= \sum_{z=0}^{\infty}\left(1-\left(1-\epsilon^z\right)^M\right) \\ &= -\sum_{n=1}^{M}\binom{M}{n}\frac{(-1)^n}{1-\epsilon^n}\end{aligned} \tag{77}$$

and we obtain

$$\begin{aligned}\tilde{C}_N(M) &\equiv \frac{C_N(M)}{\bar{z}(M)\mu\tau(M)} \\ &\leq \frac{\log(M+1)}{-\bar{z}(M)\log\left(1-(1-\epsilon)^{\frac{1}{M}}\right)}\end{aligned} \tag{78}$$

as an approximation to capacity for the number/concentration channel.

In FIGURE 6 we plot the upper bound of $\tilde{C}_N(M)$ in equation (78) versus $\mathcal{P}(M)$ (i.e., parametrized in M) for a range of ϵ as compared to the timing channel lower bound in FIGURE 4. The timing channel capacity lower bound is always significantly greater than $\tilde{C}_N(M)$. Nonetheless, the coding simplicity of the number/concentration channel could make it an attractive option.

3) Identifiable Tokens Without Payload: In Section IV-F we mentioned the possibility of uniquely identifying each of M emitted tokens with a sequence number of length $\log M$ bits. We treat this scenario as distinct from ensemble timing channel coding which resolves residual orderering ambiguity (see Section IV-F) because if the tokens are individually identifiable, the potential emission schedules are not constrained

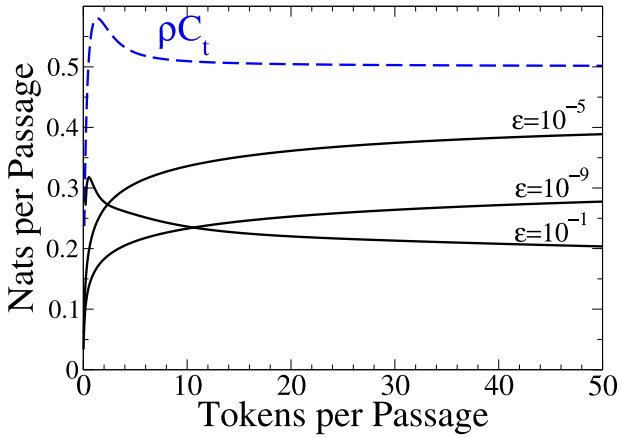

Fig. 6. Lower bound for ρC_t versus tokens per passage (dashed line) compared to $\tilde{C}_N$ vs. tokens per passage for different values of ϵ as shown. Token construction and emission are assumed unit energy for both the timing channel and the number/concentration channel. Exponential first-passage assumed.

to ensemble timing channel coding. Thus, the M identifiable tokens constitute M parallel single-token timing channels, which for exponential first-passage have aggregate capacity $M\log(1+\frac{M}{\rho e})$.

However, because each token requires $\log M$ bits of sequencing information, ρ is limited by the power budget $\mathcal{P}$ (in units of energy per passage, $1/\mu$)

$$\rho \log M \leq \mathcal{P} \tag{79}$$

Following Section IV-A we have $\lambda\tau(M) = M$ so the capacity in nats per passage time is

$$C = \rho\log\left(1+\frac{M}{\rho e}\right) \leq \rho\log\left(1+\frac{e^{\frac{\mathcal{P}}{\rho}}}{\rho e}\right) \tag{80}$$

the last inequality owed to equation (79). However, in the limit of $M\to\infty$ we have $\rho\to 0$ so we have

$$\lim_{\rho\to 0} C = \mathcal{P} \tag{81}$$

in units of nats per passage time (and assuming unit per-bit cost of the token identifer string). Thus, the identifiable token timing channel capacity exceeds the identical-token timing channel lower bound with increasing power budget and scales linearly in power as does the timing channel upper bound (see FIGURE 4) – while still lying below it.

4) Token Corruption and Receptor Noise: The potential for lost or corrupted tokens and potential binding noise at receptor sites must eventually be considered. However, as previously stated, token erasure (tokens which do not arrive) or payload token corruption (tokens which are altered in passage) or receptor noise (tokens bind stochastically to the receptor) cannot increase capacity (data processing theorem). Thus, the results here provide upper bounds. Nonetheless it is worth considering how the analytic machinery developed might be modified to accommodate such impediments.

First, consider alteration of payload-carrying tokens *en route*. If the corruption is i.i.d. for each token, then error correcting codes can be applied individually, or to the token

ensemble. The resulting overall channel capacity will be degraded by the coding overhead necessary to preserve payload message integrity (including the sequencing headers).

Then consider token erasure where a token never arrives (and is assumed to not arrive in a later signaling interval). Since each signaling interval uses M tokens, we will know whether tokens get "lost" in transit and can arbitrarily assign a faux arrival time to such tokens. However, the problem this poses for our analysis is two-fold. First, tokens released later in the signaling interval are more likely to be lost which implies that the first-passage density is not identical for each token. Second, the first-passage density for each token would then contain a singularity equal to the probability of loss, violating one of our key assumptions and making hypersymmetric probability density folding arguments invalid. That said, an erasure channel approach where arriving tokens were deleted randomly could be pursued, providing a worst case scenario since the information associated with erasures being more likely for later emissions would be absent.

Finally, we have previously mentioned that a token may "arrive" multiple times owing to ligand/token binding kinetics. It was previously shown that given the first binding (first-passage) time the information content of subsequent bindings by the same token is nil [37]. In addition, as shown in Section IV-A, information-theoretically patent channel use requires that tokens from a given emission interval be eventually cleared at the receiver. Otherwise, lingering tokens can interfere with subsequent emission intervals.

So, one could imagine the rebinding process producing a characteristic finite-mean arrival "burst" associated with each token which could perhaps be resolved into a single first-arrival time estimate – effectively adding more jitter to the first-passage time D. If so, our model applies directly with appropriate modification. However, we have not attempted to analyze this scenario nor quantify the associated estimation noise. Thus, our results are most appropriately applied to systems where ligands bind tightly or where tokens are removed with high probability after first capture/detection.

5) Interference and Multiple Users: Multi-user communication in a molecular setting is a critical question, and a better understanding of the single-user channel will certainly help with multi-user studies where transmissions interfere. There is some prior work which may provide an information-theoretic foundation [56] similar to how our work builds on [42], but the multi-user molecular signaling problem has not yet been rigorously considered. Of particular interest would be a version of MIMO since FIGURE 5 shows capacity benefits to parallel channels. One could imagine apposed arrays of emitters and receivers which could be engineered to collaborate to encode and decode information in a variety of ways, from parallel non-interfering channels to grossly interfering channels where joint/distributed coding might be employed.

6) Other Applications: Although our work is motivated by molecular communication, the notion of token inscription applies to any system where discrete emissions experience random transport delay between sender and receiver. The most obvious example is the Internet where packets experience variable delay and may arrive out of order. Our results provide crisp bounds on the amount of sequencing overhead necessary for proper message reconstruction and even suggest that (at least for low payload packets traveling over independent routes) timing information could be an interesting adjunct to payload, depending upon the timing jitter between the source and destination.

There is also the potential for cross pollination from biology to communication systems. As a simple example, there may be some selective benefit to hiding information from competitors. So, perhaps biological systems, where signaling chemicals are often detectable by other organisms, convey secrets over molecular communication channels in ways that can be mimicked in engineered systems. An obvious application, biosteganography [57], comes to mind in the context of tokens with payloads, but time-coding could also be used to obscure information.

B. Biological Implications

The natural world offers a dizzying array of processes and phenomena through which the same and different tasks, communication or otherwise, are accomplished (for example, [58]–[63]). Hence, communication theorists have plied their trade heavily in this domain (reviewed relatively recently in [64]). Identifying the underlying mechanisms (signaling modality, signaling agent, signal transport, etc.) as well as the molecules and structures implementing the mechanisms is no small undertaking. Consequently, experimental biologists use a combination of prior knowledge and what can only be called instinct to choose those systems on which to expend effort. Guidance may be sought from evolutionary developmental biology – a field that compares the developmental processes of different organisms to determine their ancestral relationship and to discover how developmental processes evolved. Insights may be gained by using statistical machine learning techniques to analyze heterogeneous data such as the biomedical literature and the output of so-called "omics" technologies – genomics (genes, regulatory, and non-coding sequences), transcriptomics (RNA and gene expression), proteomics (protein expression), metabolomics (metabolites and metabolic networks), pharmacogenomics (how genetics affects hosts' responses to drugs), physiomics (physiological dynamics and functions of whole organisms).

Frequently, the application of communication theory to biology starts by selecting a candidate system whose components and operations have been elucidated to varying degrees using the experimental/computational biology toolbox [65], [66] and then applying communication/information theoretic methods [63], [64], [67]–[72]. However, we believe that communication theory in general and information theory in particular are not mere system analysis tools for biology but new lenses on the natural world [73]. Here, we have sought to demonstrate the potential of *communication theory as an organizing principle for biology*. That is, given energy constraints and some general physics of a problem, an information-theoretic treatment can be used to provide outer bounds on information transfer in a *mechanism-blind* manner. Thus,

rather than simply elucidating and quantifying known biology, communication theory can winnow the plethora of possibilities (or even suggest new ones) amenable to subsequent pursuit.

Examples of the implications of our main communication theoretic results are as follows:

- *We have derived a model and methodology for determining the amount of information a system using chemical signaling can convey under given power constraints.* Do (or How do) biological systems achieve this extremely – even outrageously – low value?
- *Using tokens with large payloads can be very efficient.* Is one exemplar transmission of hereditary material such as a genome over evolutionary time scales (periods spanning the history of groups and species)?
- *If it is difficult to synthesize long strings (information-laden tokens), even a single bit of information (two distinguishable species) increases capacity.* Is one exemplar the alternative versions of a given gene (alleles) that are found at the same position on a chromosome? Different alleles can result in different observable phenotypic traits, for example, the gene that codes for the shape of a pea has two alleles, wrinkled and round.
- *Although the production rate of tokens can be increased, the channel capacity might not increase if tokens are emitted at intervals smaller than the variability of arrival time.* How are release and arrival times modulated by the physicochemical properties of the material through which discrete particles travel, environmental factors such as temperature, pH, pressure and light and crowding at the source and target? In its natural milieu, a molecule lives and operates in an extremely structured, complex and confining environment: one where it is surrounded by other molecules of the same or different chemical nature, the bystanders in the crowd having positive or negative effects on its mobility, biochemistry and cell biology [74], [75]. The plants and animals in agroecosystems exist in similarly challenging environments.

Our results could guide studies aimed at answering three key questions biologists ask of a living organism: *How does it work?*, *How is it built?*, and *How did it get that way?* [76]. This is because our models of token timing, payload and timing plus payload channels are inspired, at least in part, by fundamental "systems" problems about the dynamic and reciprocal communication between diverse biological entities. These include elucidating the origins (evolutionary developmental biology), generation (embryogenesis, and morphogenesis), maintenance (homeostasis, tolerance, and resilience), subversion (infectious and chronic diseases such as cancer and immune disorders), and decline (aging) of complex biological form and function [73].

ACKNOWLEDGMENT

Profound thanks are owed to A. Eckford, N. Farsad, S. Verdú and V. Poor for useful discussions and guidance. The authors are also extremely grateful to the editorial staff and the raft of especially careful and helpful anonymous reviewers.

REFERENCES

[1] A. Einolghozati, M. Sardari, A. Beirami, and F. Fekri, "Capacity of discrete molecular diffusion channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, St. Petersburg, Russia, Jul. 2011, pp. 723–727.

[2] I. F. Akyildiz, J. M. Jornet, and M. Pierobon, "Nanonetworks: A new Frontier in communications," *Commun. ACM*, vol. 54, no. 11, pp. 84–89, 2011.

[3] M. Pierobon and I. F. Akyildiz, "Capacity of a diffusion-based molecular communication system with channel memory and molecular noise," *IEEE Trans. Inf. Theory*, vol. 59, no. 2, pp. 942–954, Feb. 2013.

[4] T. Nakano, A. W. Eckford, and T. Haraguchi, *Molecular Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2013.

[5] A. Einolghozati, M. Sardari, and F. Fekri, "Relaying in diffusion-based molecular communication," in *Proc. IEEE Int. Symp. Inf. Theory*, Istanbul, Turkey, 2013, pp. 1844–1848.

[6] A. Einolghozati, M. Sardari, and F. Fekri, "Design and analysis of wireless communication systems using diffusion-based molecular communication among bacteria," *IEEE Trans. Wireless Commun.*, vol. 12, no. 12, pp. 6096–6105, Dec. 2013.

[7] A. Einolghozati, M. Sardari, and F. Fekri, "Decode and forward relaying in diffusion-based molecular communication between two populations of biological agents," in *Proc. IEEE Inf. Conf. Commun. (ICC)*, Sydney, NSW, Australia, 2014, pp. 3975–3980.

[8] N. Farsad, Y. Murin, A. W. Eckford, and A. Goldsmith, "On the capacity of diffusion-based molecular timing channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Barcelona, Spain, Jul. 2016, pp. 1023–1027.

[9] N. Farsad, Y. Murin, A. W. Eckford, and A. Goldsmith, "Capacity limits of diffusion-based molecular timing channels," *IEEE Trans. Inf. Theory*, Feb. 25, 2016. [Online]. Available: https://arxiv.org/abs/1602.07757

[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 1991.

[11] N. Michelusi and U. Mitra, "Capacity of electron-based communication over bacterial cables: The full-CSI case," *IEEE Trans. Mol. Biol. Multi-Scale Commun.*, vol. 1, no. 1, pp. 62–75, Mar. 2015.

[12] C. Morris and D. Sands. (2012). *From Grains to Rain: The Link Between Landscape, Airborne Microorganisms and Climate Processes*. [Online]. Available: http://bioice.wordpress.com/

[13] J. M. Houtkooper, "Glaciopanspermia: Seeding the terrestrial planets with life?" *Planetary Space Sci.*, vol. 59, no. 10, pp. 1107–1111, 2011. [Online]. Available: http://dx.doi.org/10.1016/j.pss.2010.09.003

[14] M. Küppers *et al.*, "Localized sources of water vapour on the dwarf planet (1) Ceres," *Nature*, vol. 505, no. 7484, pp. 525–527, 2014. [Online]. Available: http://dx.doi.org/10.1038/nature12918

[15] A. S. Tannenbaum, *Computer Networks*. Upper Saddle River, NJ, USA: Prentice-Hall, 2002.

[16] J. Gray, W. Chong, T. Barclay, A. Szalay, and J. vandenBerg, "TeraScale SneakerNet: Using inexpensive disks for backup, archiving, and data exchange," Microsoft Res., Redmond, WA, USA, Tech. Rep. MSR-TR-2002-54, May 2002.

[17] S. J. Devitt, A. D. Greentree, A. M. Stephens, and R. Van Meter, "High-speed quantum networking by ship," *arXiv:1410.3224 [quant-ph]*, accessed on Oct. 13, 2014. [Online]. Available: http://arxiv.org/abs/1410.3224

[18] R. H. Frenkiel and T. Imielinski, "Infostations: The joy of 'many-time, many-where' communications," WINLAB, Rutgers Univ., New Brunswick, NJ, USA, Tech. Rep. WINLAB-TR 119, Apr. 1996.

[19] R. Song, C. Rose, Y.-L. Tsai, and I. S. Mian, "Wireless signaling with identical quanta," in *Proc. IEEE WCNC*, Paris, France, Apr. 2012, pp. 699–703.

[20] D. J. Goodman, J. Borras, N. B. Mandayam, and R. D. Yates, "INFOSTATIONS: A new system model for data and messaging services," in *Proc. IEEE Veh. Technol. Conf.*, vol. 2. Phoenix, AZ, USA, May 1997, pp. 969–973.

[21] J. Irvine, D. Pesh, D. Robertson, and D. Girma, "Efficient UMTS data service provision using infostations," in *Proc. IEEE Veh. Technol. Conf.*, vol. 3. Ottawa, ON, Canada, 1998, pp. 2119–2123.

[22] J. Irvine and D. Pesch, "Potential of DECT terminal technology for providing low-cost wireless Internet access through infostations," in *Proc. IEE Colloquium UMTS Terminal Softw. Radio*, Glasgow, U.K., 1999, pp. 1–6.

[23] R. H. Frenkiel, B. R. Badrinath, J. Borras, and R. D. Yates, "The infostations challenge: Balancing cost and ubiquity in delivering wireless data," *IEEE Pers. Commun.*, vol. 7, no. 2, pp. 66–71, Apr. 2000.

[24] A. L. Iacono and C. Rose, "Bounds on file delivery delay in an infostations system," in *Proc. IEEE Veh. Technol. Conf.*, Tokyo, Japan, 2000, pp. 2295–2299.

[25] A. Iacono and C. Rose, "Infostations: New perspectives on wireless data networks," in *Next Generation Wireless Networks*, S. Tekinay, Ed. Boston, MA, USA: Kluwer Acad., May 2000.

[26] A. L. Iacono, "File delivery delay in and infostations system," Ph.D. dissertation, Dept. Elect. Comput. Eng., Rutgers Univ., New Brunswick, NJ, USA, Jun. 2000.

[27] A. Iacono and C. Rose, "Bounds on file delivery delay in an infostations system," in *Proc. IEEE Veh. Technol. Conf.*, Tokyo, Japan, May 2000, pp. 2295–2299.

[28] A. L. Iacono and C. Rose, "Mine, mine, mine: Information theory, infostation networks, and resource sharing," in *Proc. WCNC*, Chicago, IL, USA, Sep. 2000, pp. 1541–1546.

[29] F. Atay and C. Rose, "Exploiting mobility in multihop infostation networks to decrease transmit power," in *Proc. IEEE WCNC*, Atlanta, GA, USA, Mar. 2004, pp. 2353–2358.

[30] F. Atay and C. Rose, "Exploiting mobility in multihop infostation networks to decrease transmit power," in *Proc. IEEE Wireless Commun. Netw. Conf.*, vol. 4. 2004, pp. 2353–2358, doi: 10.1109/WCNC.2004.1311456.

[31] C. Rose and G. Wright, "Will ET write or radiate: Inscribed mass vs. electromagnetic channels," in *Proc. Conf. Rec. 38th Asilomar Conf. Signals Syst. Comput.*, vol. 1. 2004, pp. 203–207, doi: 10.1109/ACSSC.2004.1399120.

[32] C. Rose, "Write or radiate?" in *Proc. IEEE Veh. Technol. Conf.*, Orlando, FL, USA, Oct. 2003, pp. 2322–2327.

[33] C. Rose and G. Wright, "Inscribed matter as an energy-efficient means of communication with an extraterrestrial civilization," *Nature*, vol. 431, pp. 47–49, Sep. 2004.

[34] S. Goforth. *NSF Discoveries: Getting a Message Across the Universe: Would E.T. Send a Letter?* Accessed on Jan. 31, 2016. [Online]. Available: http://www.nsf.gov/discoveries/disc_summ.jsp?cntn_id=106711&org=CCF

[35] B. L. Bassler, "How bacteria talk to each other: Regulation of gene expression by quorum sensing" *Current Opin. Microbiol.*, vol. 2, no. 6, pp. 582–587, Dec. 1999.

[36] B. L. Bassler, "Small talk: Cell-to-cell communication in bacteria," *Cell*, vol. 109, no. 4, pp. 421–424, May 2002.

[37] A. Eckford, "Nanoscale communication with Brownian motion," in *Proc. CISS*, Baltimore, MD, USA, 2007, pp. 160–165.

[38] "It's the alcohol talking," *The Economist*, Economist, Technology Quarterly Q1, Mar. 2014. [Online]. Available: http://www.economist.com/news/technology-quarterly/21598326-molecular-communications-researchers-are-looking-ways-broadcast-messages

[39] C. Rose and I. S. Mian, "A fundamental framework for molecular communication channels: Timing & payload," in *Proc. IEEE Int. Conf. Commun.*, London, U.K., Jun. 2015, pp. 1043–1048.

[40] C. Rose and I. S. Mian, "Signaling with identical tokens: Lower bounds with energy constraints," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Istanbul, Turkey, Jul. 2013, pp. 1839–1843.

[41] C. Rose and I. S. Mian, "Signaling with identical tokens: Upper bounds with energy constraints," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Honolulu, HI, USA, Jun. 2014, pp. 1817–1821.

[42] V. Anantharam and S. Verdú, "Bits through queues," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 4–18, Jan. 1996.

[43] R. Sundaresan and S. Verdú, "Robust decoding for timing channels," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 405–419, Mar. 2000.

[44] R. Sundaresan and S. Verdú, "Capacity of queues via point-process channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2697–2709, Jun. 2006.

[45] Y.-L. Tsai, C. Rose, R. Song, and I. S. Mian, "An additive exponential noise channel with a transmission deadline," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, St. Petersburg, Russia, Jul. 2011, pp. 718–722.

[46] C. Rose and I. Mian, "Inscribed matter communication: Part II," *IEEE Trans. Mol. Biol. Multi-Scale Commun.*, vol. 2, no. 2, pp. 228–239, Dec. 2016.

[47] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. New York, NY, USA: McGraw-Hill, 1991.

[48] F. B. Hildebrand, *Advanced Calculus for Applications*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1976.

[49] F. P. Brooks, Jr., "Three great challenges for half-century-old computer science," *J. ACM*, vol. 50, no. 1, pp. 25–26, 2003. [Online]. Available: http://dx.doi.org/10.1145/602382.602397

[50] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: Wiley, 1968.

[51] A. L. Lehninger, D. L. Nelson, and M. M. Cox, *Lehninger Principles of Biochemistry*, 4th ed. New York, NY, USA: Freeman, 2005.

[52] M.Ş. Kurana, H. B. Yilmaz, T. Tugcu, and B. Özerman, "Energy model for communication via diffusion in nanonetworks," *Nano Commun. Netw.*, vol. 1, no. 2, pp. 86–95, 2010.

[53] C. Lee *et al.*, "Molecular MIMO communication link," *arXiv:1504.04921v2 [cs.ET]*, Mar. 2015. [Online]. Available: https://arxiv.org/abs/1503.04921

[54] V. Zhirnov, R. M. Zadegan, G. S. Sandhu, G. M. Church, and W. L. Hughes, "Nucleic acid memory," *Nat. Mater.*, vol. 15, pp. 366–370, Apr. 2016.

[55] N. Farsad, H. B. Yilmaz, A. Eckford, C.-B. Chae, and W. Guo, "A comprehensive survey of recent advancements in molecular communication," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1887–1919, 3rd Quart., 2016.

[56] G. C. Ferrante, T. Q. S. Quek, and M. Z. Win, "An achievable rate region for superposed timing channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Barcelona, Spain, Jul. 2016, pp. 365–369.

[57] T. D. Brunet, "Aims and methods of biosteganography," *J. Biotechnol.*, vol. 226, no. 20, pp. 56–64, May 2016.

[58] C. de Joussineau *et al.*, "Delta-promoted filopodia mediate long-range lateral inhibition in drosophila," *Nature*, vol. 426, pp. 555–559, Dec. 2003.

[59] Y. A. Gorby *et al.*, "Electrically conductive bacterial nanowires produced by *Shewanella Oneidensis* strain MR-1 and other microorganisms," *Proc. Nat. Acad. Sci. USA*, vol. 103, no. 30, pp. 11358–11363, 2006.

[60] S. Gurke, J. F. Barroso, and H.-H. Gerdes, "The art of cellular communication: Tunneling nanotubes bridge the divide," *Histochem. Cell Biol.*, vol. 129, no. 5, pp. 539–550, 2008. [Online]. Available: http://dx.doi.org/10.1007/s00418-008-0412-0

[61] X. Wang, M. L. Veruki, N. V. Bukoreshtliev, E. Hartveit, and H.-H. Gerdes, "Animal cells connected by nanotubes can be electrically coupled through interposed gap-junction channels," *Proc. Nat. Acad. Sci. USA*, vol. 107, no. 40, pp. 17194–17199, 2010. [Online]. Available: http://dx.doi.org/10.1073/pnas.1006785107

[62] H. C. Berg and E. M. Purcell, "Physics of chemoreception," *Biophys. J.*, vol. 20, no. 2, pp. 193–219, 1977.

[63] P. Mehta, S. Goyal, T. Long, B. Bassler, and N. Wingreen, "Information processing and signal integration in bacterial quorum sensing," *Mol. Syst. Biol.*, vol. 5, p. 325, Nov. 2009.

[64] O. Milenkovic *et al.*, "Introduction to the special issue on information theory in molecular biology and neuroscience," *IEEE Trans. Inf. Theory*, vol. 56, no. 2, pp. 649–652, Feb. 2010.

[65] A. L. Hodgkin and A. F. Huxley, "A quantitative description of membrane current and its application to conduction and excitation in nerve," *J. Physiol.*, vol. 117, no. 4, pp. 500–544, 1952.

[66] T. Long *et al.*, "Quantifying the integration of quorum-sensing signals with single-cell resolution," *PLoS Biol.*, vol. 7, no. 3, p. e68, 2009. [Online]. Available: http://dx.doi.org/10.1371/journal.pbio.1000068

[67] J. M. Amigó, J. Szczepański, E. Wajnryb, and M. V. Sanchez-Vives, "Estimating the entropy rate of spike trains via Lempel–Ziv complexity," *Neural Comput.*, vol. 16, no. 4, pp. 717–736, 2004.

[68] D. H. Johnson, "Information theory and neural information processing," *IEEE Trans. Inf. Theory*, vol. 56, no. 2, pp. 653–666, Feb. 2010.

[69] R. Barbieri *et al.*, "Dynamic analyses of information encoding in neural ensembles," *Neural Comput.*, vol. 16, no. 2, pp. 277–307, 2004.

[70] Z. Mousavian, J. Díaz, and A. Masoudi-Nejad, "Information theory in systems biology. Part II: Protein–protein interaction and signaling networks," *Seminors Cell Dev. Biol.*, vol. 51, pp. 14–23, Mar. 2016. [Online]. Available: http://dx.doi.org/10.1016/j.semcdb.2015.12.006

[71] Z. Mousaviana, K. Kavousib, and A. Masoudi-Nejad, "Information theory in systems biology. Part I: Gene regulatory and metabolic networks," *Seminars Cell Dev. Biol.*, vol. 51, pp. 3–13, Mar. 2016. [Online]. Available: http://dx.doi.org/10.1016/j.semcdb.2015.12.007

[72] S. Uda and S. Kuroda, "Analysis of cellular signal transduction from an information theoretic approach," *Seminars Cell Dev. Biol.*, vol. 51, pp. 24–31, Mar. 2016. [Online]. Available: http://dx.doi.org/10.1016/j.semcdb.2015.12.011

[73] I. S. Mian and C. Rose, "Communication theory and multicellular biology," *Integr. Biol.*, vol. 3, no. 4, pp. 350–367, Apr. 2011.

[74] S. Mittal, R. K. Chowhan, and L. R. Singh, "Macromolecular crowding: Macromolecules friend or foe," *Biochimica Biophys. Acta Gen. Subjects*, vol. 1850, no. 9, pp. 1822–1831, Sep. 2015. [Online]. Available: http://dx.doi.org/10.1016/j.bbagen.2015.05.002

[75] I. M. Kuznetsova, B. Y. Zaslavsky, L. Breydo, K. K. Turoverov, and V. N. Uversky, "Beyond the excluded volume effects: Mechanistic complexity of the crowded milieu," *Molecules*, vol. 20, no. 1, pp. 1377–1409, 2015. [Online]. Available: http://dx.doi.org/10.3390/molecules20011377

[76] S. Brenner, "Turing centenary: Life's code script," *Nature*, vol. 482, no. 7386, p. 461, 2012. [Online]. Available: http://dx.doi.org/10.1038/482461a

Christopher Rose (S'78–M'86–SM'05–F'07) received the Ph.D. degree in EECS from the Massachusetts Institute of Technology, in 1985. He was with Network Systems Department, AT&T Bell Laboratories Research, as a Technical Staff Member, from 1985 to 1990. He was with the Wireless Information Network Laboratory, Department of Electrical and Computer Engineering, Rutgers University, as a Founding Member, from 1990 to 2015. He joined the Brown School of Engineering in 2015. His current research interests include molecular communication and communication/information theory as a lens on everything.

I. S. Mian is an Honorary Professor with the Department of Computer Science, University College London whose interests lie at the intersection of biology, machine learning and communication theory.