Continuous-Time Directed Information and Its Role in Communication

Haim H. Permuter
Ben Gurion University
Beer-Sheva, Israel
haimp@bgu.ac.il

Young-Han Kim
University of California, San Diego
La Jolla, CA, USA
yhk@ucsd.edu

Tsachy Weissman
Stanford University
Stanford, CA, USA
tsachy@stanford.edu

***Abstract*—The notion of directed information was recently introduced for stochastic processes in continuous time. The key idea of the definition is to consider all possible time partitions of a given interval. Unlike the definition of mutual information of discrete-time random variables with continuous alphabets where the supremum over all possible partitions of the alphabets plays an important role, here the infimum over all possible time-partition plays an important role. We show that the fundamental limit on reliable communication for a wide class of continuous-time channels with feedback are characterized using the notion of continuous-time directed information.**

***Index Terms*—Continuous-time communication, directed information, Duncan's theorem, feedback capacity, Poisson channel.**

I. Introduction

Directed information $I(X^n \to Y^n)$ between two random n-sequences $X^n = (X_1, \ldots, X_n)$ and $Y^n = (Y_1, \ldots, Y_n)$ is a natural generalization of Shannon's mutual information to random objects with causal structures. Introduced by Massey [1], this notion of directed information has been shown to arise as the canonical answer to a variety of problems with causally dependent components. For example, it plays a pivotal role in characterizing the capacity C_{FB} of communication channels with feedback. Massey [1] showed that the feedback capacity is upper bounded by

$$C_{\text{FB}} \le \lim_{n\to\infty} \max_{p(x^n||y^{n-1})} \frac{1}{n} I(X^n \to Y^n),$$

where the definition of directed information $I(X^n \to Y^n)$ is given in Section II and $p(x^n||y^{n-1}) = \prod_{i=1}^n p(x_i|x^{i-1}, y^{i-1})$ is the causal conditioning notation streamlined by Kramer [2]. This upper bound is tight for a certain class of ergodic channels [3]–[5], paving the road to a computable characterization of feedback capacity; see [6]–[8] for examples.

Directed information and its variants also characterize (via multi-letter expressions) the capacity of two-way channels and multiple access channels with feedback [2], [9], and the rate distortion function with feedforward [10], [11]. In another context, directed information also captures the difference in growth rates of wealth in horse race gambling due to *causal* side information [12]. This provides a natural interpretation of $I(X^n \to Y^n)$ as the amount of information about Y^n *causally* provided by X^n on the fly. A similar conclusion can be drawn for other science and engineering problems, in which directed information measures the value of causal side information [13].

In this paper, we use the notion of directed information for continuous-time random processes which we recently introduced [14]. The definition of mutual information of two discrete-time random variables with continuous alphabets is given by the supremum over all possible partitions of the continuous alphabets [15, Ch. 2.5]. To define the directed information between two processes in a given time interval, we consider all possible time partitions. However, differently from the definition of mutual information, the definition of directed information is given by taking the infimum rather than by the supremum over time partitions.

The main contribution of this paper is that we show that the continuous-time directed information characterizes the capacity of continuous-time channels with feedback. This adds to the fact that in [14] we generalized Duncan's theorem which relates the minimum mean squared error (MMSE) of estimating a target signal based on an observation through an additive white Gaussian channel to directed information between the target signal and the observation.

II. Definition of Directed Information in Continuous Time

Let (X^n, Y^n) be a pair of random n-sequences. Directed information from X^n to Y^n is defined as

$$I(X^n \to Y^n) := \sum_{i=1}^{n} I(X^i; Y_i | Y^{i-1}). \tag{1}$$

Note that unlike mutual information, directed information is asymmetric in its arguments, so $I(X^n \to Y^n) \neq I(Y^n \to X^n)$.

For a continuous-time process $\{X_t\}$, let $X_a^b = \{X_s : a \le s < b\}$ denote the process in the time interval $[a, b)$. Throughout this section, equalities and inequalities between random objects, unless explicitly indicated otherwise, are to be understood to hold for all sample paths (i.e., in the sure sense). Functions of random objects are assumed measurable even though not explicitly indicated.

We now develop the notion of directed information between two continuous-time stochastic processes on the time interval $[0, T)$. Let $\mathbf{t} = (t_1, t_2, \ldots, t_n)$ denote an n-dimensional vector with components satisfying

$$0 < t_1 < t_2 < \cdots < t_n = T. \tag{2}$$

Let $X_0^{T,\mathbf{t}}$ denote the sequence of length n resulting from "chopping up" the continuous-time signal X_0^T into consecutive segments as

$$X_0^{T,\mathbf{t}} = \left(X_0^{t_1}, X_{t_1}^{t_2}, \ldots, X_{t_{n-1}}^{t_n} \right). \quad (3)$$

Note that each component of the sequence is a continuous-time stochastic process. Define

$$I_{\mathbf{t}} \left(X_0^T \to Y_0^T \right) := I \left(X_0^{T,\mathbf{t}} \to Y_0^{T,\mathbf{t}} \right) \quad (4)$$

$$= \sum_{i=1}^{n} I \left(Y_{t_{i-1}}^{t_i}; X_0^{t_i} | Y_0^{t_{i-1}} \right) \quad (5)$$

where on the right side of (4) is the directed information between two sequences of length n defined in (1); and in (5) $t_0 = 0$ by convention and the mutual information terms between two continuous time processes, conditioned on a third, are well-defined objects, as developed in [16], [17].

The mutual information of two random variables U, V with continuous alphabets, i.e., $I(U;V)$ is defined as $\sup_P I([U]_P;[V]_P)$ where P is a partition of the continuous alphabets [15, Ch. 2.5]. This is due to the fact that if P' is a refined partition of P then $I([U]_{P'};[V]_{P'}) \geq I([U]_P;[V]_P)$. However, for the continuous-time directed information the definition includes an infimum over all time partitions. This is due to the fact that $I_{\mathbf{t}} \left(X_0^T \to Y_0^T \right)$ is monotone in $\mathbf{t}$ in the following sense:

Proposition 1. *[14] If $\mathbf{t}'$ is a refinement of $\mathbf{t}$, i.e., $\{t_i\} \subset \{t_i'\}$, then $I_{\mathbf{t}'} \left(X_0^T \to Y_0^T \right) \leq I_{\mathbf{t}} \left(X_0^T \to Y_0^T \right)$.*

The following definition is now natural:

Definition 1. [14] *Directed information between X_0^T and Y_0^T is defined as*

$$I \left(X_0^T \to Y_0^T \right) := \inf_{\mathbf{t}} I_{\mathbf{t}} \left(X_0^T \to Y_0^T \right), \quad (6)$$

where the infimum is over all n and $\mathbf{t}$ as in (2).

Note, in light of Proposition 1, that

$$I \left(X_0^T \to Y_0^T \right) = \lim_{\varepsilon \to 0^+} \inf_{\{\mathbf{t}: t_i - t_{i-1} \leq \varepsilon\}} I_{\mathbf{t}} \left(X_0^T \to Y_0^T \right). \quad (7)$$

We extend the notion of directed information to that of conditional directed information $I \left(X_0^T \to Y_0^T \mid V \right)$, where $V \sim F(v)$ is a random object jointly distributed with (X_0^T, Y_0^T), as

$$I \left(X_0^T \to Y_0^T \mid V \right) := \int I \left(X_0^T \to Y_0^T \mid V = v \right) dF(v), \quad (8)$$

where $I \left(X_0^T \to Y_0^T \mid V = v \right)$ on the right hand side of (8) denotes the directed information, as already defined in Definition 1, when the pair (X_0^T, Y_0^T) is jointly distributed according to (a regular version of) the conditional distribution given $\{V = v\}$.

The following proposition collects some properties of directed information in continuous time:

Proposition 2. *[14] Directed information $I \left(X_0^T \to Y_0^T \right)$ has the following properties:*

1) *Monotonicity: $I(X_0^t \to Y_0^t)$ is monotone nondecreasing in t.*
2) *Invariance to time dilation: For $\alpha > 0$, if $\tilde{X}_t = X_{t\alpha}$ and $\tilde{Y}_t = Y_{t\alpha}$, then $I \left(\tilde{X}_0^{T/\alpha} \to \tilde{Y}_0^{T/\alpha} \right) = I(X_0^T \to Y_0^T)$. More generally, if ϕ is monotone strictly increasing and continuous, and $(\tilde{X}_{\phi(t)}, \tilde{Y}_{\phi(t)}) = (X_t, Y_t)$, then*
$$I(X_0^T \to Y_0^T) = I \left(\tilde{X}_{\phi(0)}^{\phi(T)} \to \tilde{Y}_{\phi(0)}^{\phi(T)} \right). \quad (9)$$
3) *Coincidence of directed and mutual information: If the Markov relation $Y_0^t \to X_0^t \to X_t^T$ holds for all $0 < t < T$, then*
$$I \left(X_0^T \to Y_0^T \right) = I \left(X_0^T; Y_0^T \right). \quad (10)$$
4) *Equivalence between discrete time and piecewise constancy in continuous time: Let U^n, V^n be a pair of jointly distributed n-tuples and let $t_0, t_1, \ldots, t_n$ be a sequence of numbers satisfying $t_0 = 0$, $t_n = T$, and $t_{i-1} < t_i$ for $1 \leq i \leq n$. Let the pair (X_0^T, Y_0^T) be defined as the piecewise-constant process satisfying*
$$(X_t, Y_t) = (U_i, V_i) \quad \text{if } t_{i-1} \leq t < t_i \quad (11)$$
for $i = 1, \ldots, n$. Then
$$I \left(X_0^T \to Y_0^T \right) = I \left(U^n \to V^n \right). \quad (12)$$

III. Communication over Continuous-Time Channels with Feedback

Before describing the communication model for which we will prove a coding theorem, let us review the definition of a block-ergodic process as given by Berger [18]. Let $(X, \mathcal{X}, \mu)$ denote a continuous-time process $\{X_t\}_{t \geq 0}$ drawn from a space $\mathcal{X}$ according to the probability measure μ. For $t > 0$, let T^t be a t-shift transformation, i.e., $(T^t x)_s = x_{s+t}$. A measurable set $\mathcal{A}$ is *t-invariant* if it does not change under the t-shift transformation, i.e., $T^t \mathcal{A} = \mathcal{A}$.

Definition 2 (τ-ergodicity)**.** A continuous-time process $(X, \mathcal{X}, \mu)$ is *τ-ergodic* if every measurable τ-invariant set of processes has probability either 1 or 0, i.e., for any τ-invariant set $\mathcal{A}$, in other words, $\mu(\mathcal{A}) = (\mu(\mathcal{A}))^2$.

The definition of τ-ergodicity means that if we take the process $\{X_t\}_{t \geq 0}$ and slice it into time-blocks of length τ then the new discrete-time process $(X_0^\tau, X_\tau^{2\tau}, X_{2\tau}^{3\tau}, ...)$ is ergodic.

Definition 3 (block ergodicity)**.** A continuous-time process $(X, \mathcal{X}, \mu)$ is *block-ergodic* if it is τ-ergodic for every $\tau > 0$.

Berger [18] showed that weak mixing (therefore also strong mixing) implies block ergodicity.

Now let us describe the communication model of our interest (see Fig. 1) and show that the continuous-time directed information characterizes the achievable rate.

Channel and encoding model: Consider a channel that consists of

- the channel input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$, respectively, that are not necessarily finite,

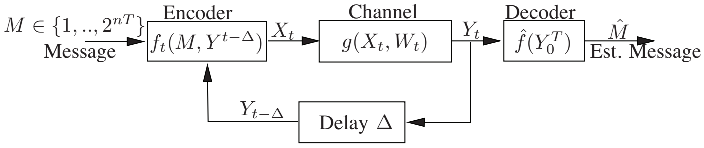

Fig. 1. Continuous-time communication with delay Δ and channel of the form $Y_t = g(X_t, W_t)$, where W_t is a block ergodic process.

- the message M, uniformly distributed on $\{1, 2, \ldots, \lfloor 2^{TR} \rfloor\}$ and independent of the stationary and block-ergodic channel noise process $\{W_t\}$, and
- the channel output process

$$Y_t = g(X_t, W_t). \tag{13}$$

We assume that the conditioned cumulative distribution function (cdf) $F(y_t^{t+\delta}|x_t^{t+\delta}, y^t)$ exists for any $t \geq 0$ and $\delta \geq 0$. Then from (13), we have

$$F(y_t^{t+\delta}|x^{t+\delta}, y_0^t, m) = F(y_t^{t+\delta}|x_0^{t+\delta}, y_0^t). \tag{14}$$

This is analogue to the assumption in the discrete case that $p(y_{n+1}|x^{n+1}, y^n, m) = p(y_{n+1}|x^{n+1}, y^n)$.

The encoder for this communication channel is defined as

$$X_t = f_t(M, Y_0^{t-\Delta}), \tag{15}$$

for $t \geq 0$ and some *feedback delay* $\Delta > 0$ (and set arbitrarily for $t < 0$). From the definition of the encoding function in (15), we note that the conditioned cdf $F(x_t^{t+\delta}|x_0^t, y_0^{t+\delta})$ exists, and for any $t \geq 0$, $\delta \geq 0$, and $\Delta > \delta$,

$$F(x_t^{t+\delta}|x_0^t, y_0^t) = F(x_t^{t+\delta}|x_0^t, y_0^{t+\delta-\Delta}). \tag{16}$$

This is analogue to the assumption in the discrete case that whenever there is feedback of delay $d \geq 1$, $p(x_{n+1}|x^n, y^n) = p(x_{n+1}|x^n, y^{n+1-d})$.

An encoding scheme for the time interval $[0, T)$ is characterized by the family of encoding functions $\{f_t\}_{t=0}^T$. Similar communication settings with feedback in continuous time were studied by [19] for memoryless continuous-time where it is shown that feedback does not increase capacity and by Ihara [20], [21] for the Gaussian case. Our main result in this section is showing that the operational capacity, defined below, can be characterize using the notion of directed information for continuous-time processes. Next we define an achievable rate, the operational feedback capacity, and the information feedback capacity for our setting.

Definition 4. A rate R is said to be *achievable with feedback delay* Δ if for each T there exists a family of encoding functions $\{f_t\}_{t=0}^T$ such that

$$\lim_{T\to\infty} P\{M \neq \hat{M}(Y_0^T)\} = 0, \tag{17}$$

where $\hat{M}(Y_0^T)$ in (17) is the maximum likelihood estimate of M given Y_0^T for the encoding functions $\{f_t\}_{t=0}^T$.

Definition 5. Let

$$C(\Delta) = \sup\{R : R \text{ is achievable with feedback delay } \Delta\} \tag{18}$$

be the *(operational) feedback capacity* with delay Δ.

Definition 6. Let $C^I(\Delta)$ be the information feedback capacity defined as

$$C^I_{\Delta} \lim_{T\to\infty} \frac{1}{T} \sup_{\mathcal{S}_\Delta} I\left(X_0^T \to Y_0^T\right), \tag{19}$$

where the supremum in (19) is over $\mathcal{S}_\Delta$, which is the set of all channel input processes of the form

$$X_t = \begin{cases} g_t(U_t, Y_0^{t-\Delta}) & t \geq \Delta, \\ g_t(U_t) & t < \Delta, \end{cases} \tag{20}$$

some family of functions $\{g_t\}_{t=0}^T$, and some process $\{U_t\}$ which is independent of the channel noise process $\{W_t\}$ (appearing in (13)) and has a finite cardinality that may depend on T.

The limit in (19) is shown to exist in Lemma 1 using the superadditivity property. We now characterize $C(\Delta)$ for the class of channels defined above in terms of $C^I(\Delta)$.

Theorem 1. *For the channel defined in* (13),

$$C(\Delta) \leq C^I(\Delta), \tag{21}$$

$$C(\Delta) \geq C^I(\Delta') \quad \textit{for all } \Delta' > \Delta. \tag{22}$$

Since $C^I(\Delta)$ is a decreasing function in in Δ, (22) may be written as $C(\Delta) \geq \lim_{\delta\to\Delta^+} C^I(\delta)$, and the limit exists because of the monotonicity. Since the function is monotonic then $C^I(\Delta) = \lim_{\delta\to\Delta^+} C^I(\delta)$ with a possible exception of the points Δ on a set of measure zero [22, p. 5]. Therefore $C(\Delta) = C^I(\Delta)$ for almost every $\Delta \geq 0$.

Before proving the theorem we show that the limits in (19) exists.

Lemma 1. *The term* $\sup_{\mathcal{S}_\Delta} I\left(X_0^T \to Y_0^T\right)$ *is superadditive, namely,*

$$\sup_{\mathcal{S}_\Delta} I\left(X_0^{T_1+T_2} \to Y_0^{T_1+T_2}\right) \geq \sup_{\mathcal{S}_\Delta} I\left(X_0^{T_1} \to Y_0^{T_1}\right) + \sup_{\mathcal{S}_\Delta} I\left(X_0^{T_2} \to Y_0^{T_2}\right), \tag{23}$$

and therefore the limit in (19) *exists and is equal to*

$$\lim_{T\to\infty} \frac{1}{T} \sup_{\mathcal{S}_\Delta} I\left(X_0^T \to Y_0^T\right) = \sup_T \frac{1}{T} \sup_{\mathcal{S}_\Delta} I\left(X_0^T \to Y_0^T\right) \tag{24}$$

Proof: First we notice that we do not increase the term $\inf_{\mathbf{t}} I_{\mathbf{t}}\left(X_0^{T_1+T_2} \to Y_0^{T_1+T_2}\right)$ by restricting the time-partition $\mathbf{t}$ to have an interval starting at point T_1. Now fix three time-partitions: $\mathbf{t}_1$ in $[0, T_1)$, $\mathbf{t}_2$ in $[T_1, T_1+T_2)$, and $\mathbf{t}$ in $[0, T_1 + T_2)$ such that $\mathbf{t}$ is a concatenation $\mathbf{t}_1$ and $\mathbf{t}_2$. For $X_0^{T_1}$ and $X_{T_1}^{T_1+T_2}$, fix the input functions of the form of (20) and fix the arguments $U_0^{T_1}$ and $U_{T_1}^{T_1+T_2}$, which corresponds to $X_0^{T_1}$ and $X_{T_1}^{T_1+T_2}$, respectively. The construction is such that the random processes $U_0^{T_1}$ and $U_{T_1}^{T_1+T_2}$ are independent of each other. Let $X_0^{T_1+T_2}$ be a concatenation of $X_0^{T_1}$ and $X_{T_1}^{T_1+T_2}$. For any fixed $\mathbf{t}_1, \mathbf{t}_2, X_0^{T_1}, X_{T_1}^{T_1+T_2}, U_0^{T_1}$, and $U_{T_1}^{T_1+T_2}$ as described above, we have

$$I_{\mathbf{t}_1+t_2}\left(X_0^{T_1+T_2} \to Y_0^{T_1+T_2}\right) \geq I_{\mathbf{t}_1}\left(X_0^{T_1} \to Y_0^{T_1}\right) + I_{\mathbf{t}_2}\left(X_{T_1}^{T_1+T_2} \to Y_{T_1}^{T_1+T_2}\right). \quad (25)$$

This inequality follows since if we have a discrete-time process $\{(X_i, Y_i)\}_{i=1}^{n+m}$, where the Markov relation $X_i \to (X^{i-1}, Y^{i-1}) \to (X_{n+1}^{i-1}, Y_{n+1}^{i-1})$ holds for $i \in \{n+1, n+2, ..., n+m\}$, then

$$I\left(X^{n+m} \to Y^{n+m}\right) \geq I\left(X^n \to Y^n\right) + I\left(X_{n+1}^{n+m} \to Y_{n+1}^{n+m}\right), \quad (26)$$

which is an easy consequence of the identity $I(X^n \to Y^m) = \sum_{i=1}^n I(X_i; Y_i^n | X^{i-1}, Y^{i-1})$; see, for example, [4]. By the stationarity of the noise, (25) implies (23). Finally, using Fekete's lemma [23, Ch. 2.6] and the superadditivity in (23) implies the existence of the limit in (24). ∎

The proof of Theorem 1 consists of two parts: the proof of the converse, i.e., (21), and the proof of achievability, i.e., (22).

Proof of the converse: Fix an encoding scheme $\{f_t\}_{t=0}^T$ with rate R and probability of decoding error, $P_e^{(T)} = P\{M \neq \hat{M}(Y_0^T)\}$. In addition, fix a partition $\mathbf{t}$ of length n such that $t_i - t_{i-1} < \Delta$ for any $i = 1, 2, \ldots, n$, and let $t_n = T$. Consider

$$\begin{aligned}
RT &= H(M) && (27)\\
&= H(M) + H(M|Y_0^T) - H(M|Y_0^T) && (28)\\
&\leq I(M; Y_0^T) + T\epsilon_T && (29)\\
&= I(M; Y_0^{t_1}, Y_{t_1}^{t_2}, \ldots, Y_{t_{n-1}}^{t_n}) + T\epsilon_T && (30)\\
&= \sum_{i=1}^n I(M; Y_{t_{i-1}}^{t_i} | Y_0^{t_{i-1}}) + T\epsilon_T && (31)\\
&= \sum_{i=1}^n I(M, X_0^{t_{i-1}+\Delta}; Y_{t_{i-1}}^{t_i} | Y_0^{t_{i-1}}) + T\epsilon_T && (32)\\
&= \sum_{i=1}^n I(M, X_0^{t_i}, X_{t_i}^{t_{i-1}+\Delta}; Y_{t_{i-1}}^{t_i} | Y_0^{t_{i-1}}) + T\epsilon_T && (33)\\
&= \sum_{i=1}^n I(M, X_0^{t_i}; Y_{t_{i-1}}^{t_i} | Y_0^{t_{i-1}}) \\
&\quad + I(X_{t_i}^{t_{i-1}+\Delta}; Y_{t_{i-1}}^{t_i} | Y_0^{t_{i-1}}, M, X_0^{t_i}) + T\epsilon_T && (34)\\
&= \sum_{i=1}^n I(X_0^{t_i}; Y_{t_{i-1}}^{t_i} | Y_0^{t_{i-1}}) \\
&\quad + I(X_{t_i}^{t_{i-1}+\Delta}; Y_{t_{i-1}}^{t_i} | Y_0^{t_{i-1}}, M, X_0^{t_i}) + T\epsilon_T && (35)\\
&= \sum_{i=1}^n I(X_0^{t_i}; Y_{t_{i-1}}^{t_i} | Y_0^{t_{i-1}}) + T\epsilon_T && (36)\\
&= I_{\mathbf{t}}(X_0^T \to Y_0^T) + T\epsilon_T, && (37)
\end{aligned}$$

where the equality in (27) follows since the message is distributed uniformly, the inequality in (29) follows from Fano's inequality with $\epsilon_T = (1/T) + P_e^{(T)} R \to 0$ as $P_e^{(T)} \to 0$, the equality in (32) follows from the fact that $X^{t_{i-1}+\Delta}$ is a deterministic function of M and $Y_0^{t_{i-1}}$, the equality in (33) follows from the assumption that $t_i - t_{i-1} < \Delta$, the equality in (35) follows from (14), and the equality in (36) follows from (16). Hence, we obtained that for every $\mathbf{t}$

$$R \leq \frac{1}{T} I_{\mathbf{t}}(X_0^T \to Y_0^T) + \epsilon_T. \quad (38)$$

Since the number of codewords is finite, we may consider the input signal of the form $x^{T,\mathbf{t}}$ with $x_{t_{i-1}}^{t_i} = f(u_0^T, y^{t_i - \Delta})$, where the cardinality of u_0^T is bounded, i.e., $|\mathcal{U}^T| < \infty$ for any T, independently of the partition $\mathbf{t}$. Furthermore,

$$\begin{aligned}
R &\leq \inf_{\mathbf{t}} \frac{1}{T} I_{\mathbf{t}}(X_0^T \to Y_0^T) + \epsilon_T, \\
&= \frac{1}{T} I(X_0^T \to Y_0^T) + \epsilon_T. && (39)
\end{aligned}$$

Finally, for any R that is achievable there exists a sequence of codes such that $\lim_{T\to\infty} P_e^{(T)} = 0$, hence $\epsilon_T \to 0$ and we have established (21). ∎

For the proof of achievability we will use the following result for discrete-time channels.

Lemma 2. *Consider the discrete-time channel, where the input U_i at time i has a finite alphabet $\mathcal{U}$, with $|\mathcal{U}| < \infty$, and the output Y_i at time i has an arbitrary alphabet $\mathcal{Y}$. We assume that the relation between the input and the output is given by*

$$Y_i = g(U_i, Z_i), \quad (40)$$

where the noise process $\{Z_i\}_{i\geq 1}$ is stationary and ergodic with an arbitrary alphabet $\mathcal{Z}$. The function $g : \mathcal{U} \times \mathcal{Z} \to \mathcal{Y}$ is a deterministic function. For this discrete time channel, any rate

$$R < \max_{p(u)} I(U; Y) \quad (41)$$

is achievable.

Proof: Fix the pmf $p(u)$ that attains the maximum in (41). Since $I(U;Y)$ can be approximated arbitrarily close by a finite partition of Y [15], assume without loss of generality that $\mathcal{Y}$ is finite. The proof uses the random codebook generation and joint typicality decoding in [24, Lecture 3]. Randomly and independently generate 2^{nR} codewords $u^n(m)$, $m = 1, 2, \ldots, 2^{nR}$, each i.i.d. according to $\prod_{i=1}^n p_U(u_i)$. The decoder finds the unique $\hat{m}$ such that $(u^n(m), y^n)$ is jointly typical. (For the definition of joint typicality, refer [24, Lecture 2].) Now, assuming that $M = 1$ is sent, the decoder makes an error only if $(U^n(1), Y^n)$ is not typical or $(U^n(m), Y^n)$ is typical for some $m \neq 1$. By the packing lemma [24, Lecture 3], the probability of the second event tends to zero as $n \to \infty$ if $R < I(U;Y)$. To bound the probability of the first event, recall from [25, Thm 10.3.1] that if $\{U_i\}$ is i.i.d. and $\{Z_i\}$ is stationary ergodic, independent of $\{U_i\}$, then the pair $\{(U_i, Z_i)\}$ is jointly stationary ergodic. Consequently, from the definition of the channel in (40), $\{(U_i, Y_i)\}$ is jointly stationary ergodic. Thus, by Birkhoff's ergodic theorem, the probability that $(U^n(1), Y^n)$ is not typical tends to zero as $n \to \infty$. Therefore, any rate $R < I(U;Y)$ is achievable. ■

The proof of achievability is based on the lemma above and the definition of directed information for continuous time. It is essential to divide into small time intervals as well as increasing the feedback delay by a small positive value $\delta > 0$.

Proof of achivability for Theorem 1: Let $\delta > 0$ and $\Delta' = \Delta + \delta$. In addition let $\mathbf{t} = (t_1, t_2, \ldots, t_n)$ be such that $t_i - t_{i-1} \leq \delta$ for all $i = 1, 2, \ldots, n$. Let $X_0^{T,\mathbf{t}}$ be of the form

$$X_{t_{i-1}}^{t_i} = \begin{cases} f(u_0^T, y_0^{t_i - \Delta'}) & t_i \geq \Delta', \\ f(u_0^T) & t_i < \Delta', \end{cases} \tag{42}$$

where U_0^T is a random process with finitely many sample paths. Then we show that any rate

$$R < \frac{1}{T} I_{\mathbf{t}}(X_0^{T,\mathbf{t}} \to Y_0^{T,\mathbf{t}}), \tag{43}$$

is achievable. Chop the time into intervals of length T and in each interval $[jT, jT + T)$ which we index by j, fix the relation

$$X_{jT+t_{i-1}}^{jT+t_i} = \begin{cases} f(u_{jT}^{jT+T}, y_{jT}^{jT+t_i-\Delta'}) & t_i \geq \Delta', \\ f(u_{jT}^{jT+T}) & t_i < \Delta'. \end{cases} \tag{44}$$

Note that this coding scheme is possible with feedback delay Δ since $t_{i-1} - \Delta \geq t_i - \Delta'$. This follows from the assumption that $t_i - t_{i-1} \leq \delta$ and $\Delta' - \Delta \geq \delta$. Now, let us define a discrete-time channel where the input at time $j+1$ is $\tilde{U}_{j+1} = U_{jT}^{jT+T}$ (which has a finite alphabet), the output at time $j+1$ is the vector $\tilde{Y}_{j+1} = (Y_{jT}^{jT+t_1}, \ldots, Y_{jT+t_{i-1}}^{jT+t_i}, \ldots, Y_{jT+t_{n-1}}^{jT+T})$ and the noise at time $j+1$ is $\tilde{W}_{j+1} = Z_{jT}^{jT+T}$. Note that since Z_{jT}^{jT+T} is a stationary and block-ergodic the noise process $\{\tilde{W}_{j+1}\}_{j \geq 0}$ is stationary and ergodic. Furthermore the relation $\tilde{Y}_{j+1} = f(\tilde{U}_{j+1}, \tilde{W}_{j+1})$ holds and the alphabet of $\tilde{U}_{j+1}$ is finite. Hence according to Lemma 2 a rate

$$R = \max_{p(\tilde{u})} I(\tilde{U}; \tilde{Y}), \tag{45}$$

is achievable. Now using the definition of the discrete-time channel and the properties of directed information, we obtain

$$\begin{aligned} I(\tilde{U}; \tilde{Y}) &= I(U_0^T; Y_0^T) & (46) \\ &= I(U_0^T; Y_0^{t_1}, Y_{t_1}^{t_2}, \ldots, Y_{t_n-1}^{t_n}) & (47) \\ &= I_{\mathbf{t}}(X_0^{T,\mathbf{t}} \to Y_0^{T,\mathbf{t}}), & (48) \end{aligned}$$

where the equality in (46) follows from the definition of the discrete-time channel and the equality in (48) follows from the same sequence of equalities as in (30)–(37). Since (48) holds for any $\mathbf{t}$ such that $t_i - t_{i-1} \leq \delta$ we obtain that

$$C \geq \frac{1}{T} \inf_{\mathbf{t}} I_{\mathbf{t}}(X_0^{T,\mathbf{t}} \to Y_0^{T,\mathbf{t}}), \tag{49}$$

and finally by the definition of directed information and by the fact that (49) holds for any T we have established (22). ■

IV. Concluding Remarks and Research Directions

We have showed that the notion of directed information characterizes the fundamental limit on reliable communication for a wide class of continuous-time channels with feedback is discussed. The next research direction is to use the results that relates the continuous directed information to estimation problems [14] in order to calculate the feedback capacity of special Poisson and Gaussian channels.

References

[1] J. Massey. Causality, feedback and directed information. *Proc. Int. Symp. Inf. Theory Applic. (ISITA-90)*, pages 303–305, Nov. 1990.

[2] G. Kramer. Capacity results for the discrete memoryless network. *IEEE Trans. Inf. Theory*, IT-49:4–21, 2003.

[3] S. Tatikonda and S. Mitter. The capacity of channels with feedback. *IEEE Trans. Inf. Theory*, 55:323–349, 2009.

[4] Y.-H. Kim. A coding theorem for a class of stationary channels with feedback. *IEEE Trans. Inf. Theory.*, 25:1488–1499, April, 2008.

[5] H. H. Permuter, T. Weissman, and A. J. Goldsmith. Finite state channels with time-invariant deterministic feedback. *IEEE Trans. Inf. Theory*, 55(2):644–662, 2009.

[6] J. Chen and T. Berger. The capacity of finite-state Markov channels with feedback. *IEEE Trans. Inf. Theory*, 51:780–789, 2005.

[7] Y.-H. Kim. Feedback capacity of stationary Gaussian channels. *IEE Trans. Inf. Theory*, 57(1):57–85, Jan. 2010.

[8] H. H. Permuter, P. Cuff, B. Van Roy, and T. Weissman. Capacity of the trapdoor channel with feedback. *IEEE Trans. Inf. Theory*, 54(7):3150–3165, 2009.

[9] G. Kramer. *Directed information for channels with feedback*. Ph.D. dissertation, Swiss Federal Institute of Technology (ETH) Zurich, 1998.

[10] S. Pradhan R. Venkataramanan. Source coding with feed-forward: Rate-distortion theorems and error exponents for a general source. *IEEE Trans. Inf. Theory*, 53(6):2154–2179, 2007.

[11] S. Pradhan. On the role of feedforward in Gaussian sources: Point-to-point source coding and multiple description source coding. *IEEE Trans. Inf. Theory*, 53(1):331–349, 2007.

[12] H. H. Permuter, Y.-H. Kim, and T. Weissman. On directed information and gambling. In *Proc. International Symposium on Information Theory (ISIT)*, Toronto, Canada, 2008.

[13] H. H. Permuter, Y.-H. Kim, and T. Weissman. An interpretation of directed information in portfolio theory, data compression, and hypothesis testing. to appear in *IEEE Trans. Inf. Theory*, June 2010.

[14] H. H. Permuter, Y.-H. Kim, and T. Weissman. Directed information and causal estimation in continuous time. In *Proc. IEEE International Symposium on Information Theory (ISIT)*, Seoul, Korea., June, 2009.

[15] R. G. Gallager. *Information theory and reliable communication*. Wiley, New York, 1968.

[16] A. Kolmogorov. On the shannon theory of information transmission in the case of continuous signals. *IRE Transactions on Information Theory*, 2:102–108, 1956.

[17] M. S. Pinsker. *Information and Information Stability of Random Variables and Processes*. Izv. Akad. Nauk, Moskva, 1960. in Russian.

[18] T. Berger. Rate distortion theory for sources with abstract alphabets and memory. *Information and Control*, 13(3):254–273, 1968.

[19] T. T. Kadota, M. Zakai, and J. Ziv. Capacity of a continuous memoryless channel with feedback. *IEEE Trans. Inf. Theory*, 17:372–378, 1971.

[20] S. Ihara. Coding theorems for a continuous-time Gaussian channel with feedback. *IEEE Trans. Inf. Theory*, 40(6):2041–2044, 1994.

[21] S. Ihara. *Information Theory for Continuous Systems*. River Edge, NJ: World Scientific, 1993.

[22] F. Riesz and B. Sz.-Nagy. *Functional Analysis*. Dover Publications Inc., New York, 2nd edition.

[23] A. Schrijver. *Combinatorial Optimization - Polyhedra and Efficiency*. Springer, 2003.

[24] A. El Gamal and Y.-H. Kim. Lecture notes on network information theory. 2010. availble at http://arxiv.org/abs/1001.3404.

[25] J. Wolfowitz. *Coding Theorems of Information Theory*. Springer, 2nd edition, 1964.