frontiers
in Computational Neuroscience

ORIGINAL RESEARCH
published: 28 August 2015
doi: 10.3389/fncom.2015.00105

Time resolution dependence of information measures for spiking neurons: scaling and universality

Sarah E. Marzen*[1], *Michael R. DeWeese*[1,2] and *James P. Crutchfield*[3]

[1] Department of Physics, University of California, Berkeley, Berkeley, CA, USA, [2] Helen Wills Neuroscience Institute and Redwood Center for Theoretical Neuroscience, University of California, Berkeley, Berkeley, CA, USA, [3] Complexity Sciences Center and Department of Physics, University of California, Davis, Davis, CA, USA

The mutual information between stimulus and spike-train response is commonly used to monitor neural coding efficiency, but neuronal computation broadly conceived requires more refined and targeted information measures of input-output joint processes. A first step toward that larger goal is to develop information measures for individual output processes, including information generation (entropy rate), stored information (statistical complexity), predictable information (excess entropy), and active information accumulation (bound information rate). We calculate these for spike trains generated by a variety of noise-driven integrate-and-fire neurons as a function of time resolution and for alternating renewal processes. We show that their time-resolution dependence reveals coarse-grained structural properties of interspike interval statistics; e.g., τ-entropy rates that diverge less quickly than the firing rate indicated by interspike interval correlations. We also find evidence that the excess entropy and regularized statistical complexity of different types of integrate-and-fire neurons are universal in the continuous-time limit in the sense that they do not depend on mechanism details. This suggests a surprising simplicity in the spike trains generated by these model neurons. Interestingly, neurons with gamma-distributed ISIs and neurons whose spike trains are alternating renewal processes do not fall into the same universality class. These results lead to two conclusions. First, the dependence of information measures on time resolution reveals mechanistic details about spike train generation. Second, information measures can be used as model selection tools for analyzing spike train processes.

Keywords: statistical complexity, excess entropy, entropy rate, renewal process, alternating renewal process, integrate and fire neuron, leaky integrate and fire neuron, quadratic integrate and fire neuron
PACS numbers: 05.45.Tp 02.50.Ey 87.10.Vg 87.19.ll 87.19.lo 87.19.ls

OPEN ACCESS

Edited by:
Guenther Palm,
University of Ulm, Germany

Reviewed by:
Andrey Olypher,
Georgia Gwinnett College, USA
Thomas Wennekers,
University of Plymouth, UK
Tatjana Tchumatchenko,
Max Planck Institute for Brain Research, Germany

Correspondence:
James P. Crutchfield,
Complexity Sciences Center and Department of Physics, University of California, Davis, Davis, CA, USA
chaos@ucdavis.edu

Received: *18 April 2015*
Accepted: *30 July 2015*
Published: *28 August 2015*

Citation:
Marzen SE, DeWeese MR and Crutchfield JP (2015) Time resolution dependence of information measures for spiking neurons: scaling and universality.
Front. Comput. Neurosci. 9:105.
doi: 10.3389/fncom.2015.00105

1. Introduction

Despite a half century of concerted effort (Mackay and McCulloch, 1952), neuroscientists continue to debate the relevant timescales of neuronal communication as well as the basic coding schemes at work in the cortex, even in early sensory processing regions of the brain thought to be dominated by feedforward pathways (Softky and Koch, 1993; Bell et al., 1995; Shadlen and Newsome, 1995; Stevens and Zador, 1998; Destexhe et al., 2003; DeWeese and Zador, 2006; Jacobs et al., 2009; Koepsell et al., 2010; London et al., 2010). For example, the apparent variability of neural responses to repeated presentations of sensory stimuli has led many to conclude that the brain must average across tens or hundreds of milliseconds or across large populations of neurons to extract a meaningful signal (Shadlen and Newsome, 1998). Whereas, reports of reliable responses suggest shorter relevant timescales and more nuanced coding schemes (Berry et al., 1997; Reinagel and Reid, 2000; DeWeese et al., 2003). In fact, there is evidence for different characteristic timescales for neural coding in different primary sensory regions of the cortex (Yang and Zador, 2012). In addition to questions about the relevant timescales of neural communication, there has been an ongoing debate regarding the magnitude and importance of correlations among the spiking responses of neural populations (Meister et al., 1995; Nirenberg et al., 2001; Averbeck et al., 2006; Schneidman et al., 2003, 2006).

Most studies of neural coding focus on the relationship between a sensory stimulus and the neural response. Others consider the relationship between the neural response and the animal's behavioral response (Britten et al., 1996), the relationship between pairs or groups of neurons at different stages of processing (Linsker, 1989; Dan et al., 1996), or the variability of neural responses themselves without regard to other variables (Schneidman et al., 2006). Complementing the latter studies, we are interested in quantifying the randomness and predictability of neural responses without reference to stimulus. We consider the variability of a given neuron's activity at one time and how this is related to the same neuron's activity at other times in the future and the past.

Along these lines, information theory (Shannon, 1948; Cover and Thomas, 2006) provides an insightful and rich toolset for interpreting neural data and for formulating theories of communication and computation in the nervous system (Rieke et al., 1999). In particular, Shannon's mutual information has developed into a powerful probe that quantifies the amount of information about a sensory stimulus encoded by neural activity (Mackay and McCulloch, 1952; Barlow, 1961; Stein, 1967; Laughlin, 1981; Sakitt and Barlow, 1982; Srinivasan et al., 1982; Linsker, 1989; Bialek et al., 1991; Theunissen and Miller, 1991; Atick, 1992; Rieke et al., 1999). Similarly, the Shannon entropy has been used to quantify the variability of the resulting spike-train response. In contrast to these standard stimulus- and response-averaged quantities, a host of other information-theoretic measures have been applied in neuroscience, such as the Fisher information (Cover and Thomas, 2006) and various measures of the information gained per observation (DeWeese and Meister, 1999; Butts and Goldman, 2006).

We take an approach that complements more familiar informational analyses. First, we consider "output-only" processes, since their analysis is a theoretical prerequisite to understanding information in the stimulus-response paradigm. Second, we analyze rates of informational divergence, not only nondivergent components. Indeed, we show that divergences, rather than being a kind of mathematical failure, are important and revealing features of information processing in spike trains.

We are particularly interested in the information content of neural spiking on fine timescales. How is information encoded in spike timing and, more specifically, in interspike intervals? In this regime, the critical questions turn on determining the kind of information encoded and the required "accuracy" of individual spike timing to support it. At present, unfortunately, characterizing communication at submillisecond time scales and below remains computationally and theoretically challenging.

Practically, a spike train is converted into a binary sequence for analysis by choosing a time bin size and counting the number of spikes in successive time bins. Notwithstanding Strong et al. (1998) and Nemenman et al. (2008), there are few studies of how estimates of communication properties change as a function of time bin size, though there are examples of both short (Panzeri et al., 1999) and long (DeWeese, 1996; Strong et al., 1998) time expansions. Said most plainly, it is difficult to directly calculate the most basic quantities—e.g., communication rates between stimulus and spike-train response—in the submillisecond regime, despite progress on undersampling (Treves and Panzeri, 1995; Nemenman et al., 2004; Archer et al., 2012). Beyond the practical, the challenges are also conceptual. For example, given that a stochastic process' entropy rate diverges in a process-characteristic fashion for small time discretizations (Gaspard and Wang, 1993), measures of communication efficacy require careful interpretation in this limit.

Compounding the need for better theoretical tools, measurement techniques will soon amass enough data to allow serious study of neuronal communication at fine time resolutions and across large populations (Alivisatos et al., 2012). In this happy circumstance, we will need guideposts for how information measures of neuronal communication vary with time resolution so that we can properly interpret the empirical findings and refine the design of nanoscale probes.

Many single-neuron models generate neural spike trains that are renewal processes (Gerstner and Kistler, 2002). Starting from this observation, we use recent results (Marzen and Crutchfield, 2015) to determine how information measures scale in the small time-resolution limit. This is exactly the regime where numerical methods are most likely to fail due to undersampling and, thus, where analytic formulae are most useful. We also extend the previous analyses to structurally more complex, alternating renewal processes and analyze the time-resolution scaling of their information measures. This yields important clues as to which scaling results apply more generally. We then show that, across several standard neuronal models, the information measures are universal in the sense that their scaling does not depend on the details of spike-generation mechanisms.

Several information measures we consider are already common fixtures in theoretical neuroscience, such as Shannon's source entropy rate (Strong et al., 1998; Nemenman et al., 2008). Others have appeared at least once, such as the finite-time excess entropy (or predictable information) (Bialek et al., 2001; Crutchfield and Feldman, 2003) and statistical complexity (Haslinger et al., 2010). And others have not yet been applied, such as the bound information (Abdallah and Plumbley, 2009, 2012; James et al., 2011, 2014).

The development proceeds as follows. Section 2 reviews notation and definitions. To investigate the dependence of causal information measures on time resolution, Section 3 studies a class of renewal processes motivated by their wide use in describing neuronal behavior. Section 4 then explores the time-resolution scaling of information measures of alternating renewal processes, identifying those scalings likely to hold generally. Section 5 evaluates continuous-time limits of these information measures for common single-neuron models. This reveals a new kind of universality in which the information measures' scaling is independent of detailed spiking mechanisms. Taken altogether, the analyses provide intuition and motivation for several of the rarely-used, but key informational quantities. For example, the informational signatures of integrate-and-fire model neurons differ from both simpler, gamma-distributed processes and more complex, compound renewal processes. Finally, Section 6 summarizes the results, giving a view to future directions and mathematical and empirical challenges.

2. Background

We can only briefly review the relevant physics of information. Much of the phrasing is taken directly from background presented in Marzen and Crutchfield (2014, 2015).

Let us first recall the causal state definitions (Shalizi and Crutchfield, 2001) and information measures of discrete-time, discrete-state processes introduced in Crutchfield et al. (2009), James et al. (2011). The main object of study is a process $\mathcal{P}$: the list of all of a system's behaviors or realizations $\{\ldots x_{-2}, x_{-1}, x_0, x_1, \ldots\}$ and their probabilities, specified by the joint distribution $\Pr(\ldots X_{-2}, X_{-1}, X_0, X_1, \ldots)$. We denote a contiguous chain of random variables as $X_{0:L} = X_0 X_1 \cdots X_{L-1}$. We assume the process is ergodic and stationary—$\Pr(X_{0:L}) = \Pr(X_{t:L+t})$ for all $t \in \mathbb{Z}$—and the measurement symbols range over a finite alphabet: $x \in \mathcal{A}$. In this setting, the *present* X_0 is the random variable measured at $t = 0$, the *past* is the chain $X_{:0} = \ldots X_{-2}X_{-1}$ leading up the present, and the *future* is the chain following the present $X_{1:} = X_1 X_2 \cdots$ (We suppress the infinite index in these).

As the Introduction noted, many information-theoretic studies of neural spike trains concern input-output information measures that characterize stimulus-response properties; e.g., the mutual information between stimulus and resulting spike train. In the absence of stimulus or even with a non-trivial stimulus, we can still study neural activity from an information-theoretic point of view using "output-only" information measures that quantify intrinsic properties of neural activity alone:

- How random is it? The *entropy rate* $h_\mu = H[X_0|X_{:0}]$, which is the entropy in the present observation conditioned on all past observations (Cover and Thomas, 2006).
- What must be remembered about the past to optimally predict the future? The *causal states* $\boldsymbol{\mathcal{S}}^+$, which are groupings of pasts that lead to the same probability distribution over future trajectories (Crutchfield and Young, 1989; Shalizi and Crutchfield, 2001).
- How much memory is required to store the causal states? The *statistical complexity* $C_\mu = H[\mathcal{S}^+]$, or the entropy of the causal states (Crutchfield and Young, 1989).
- How much of the future is predictable from the past? The *excess entropy* $\mathbf{E} = I[X_{:0}; X_{0:}]$, which is the mutual information between the past and the future (Crutchfield and Feldman, 2003).
- How much of the generated information (h_μ) is relevant to predicting the future? The *bound information* $b_\mu = I[X_0; X_{1:}|X_{:0}]$, which is the mutual information between the present and future observations conditioned on all past observations (Abdallah and Plumbley, 2009; James et al., 2011).
- How much of the generated information is useless—neither affects future behavior nor contains information about the past? The *ephemeral information* $r_\mu = H[X_0|X_{:0}, X_{1:}]$, which is the entropy in the present observation conditioned on all past and future observations (Verdú and Weissman, 2006; James et al., 2011).

The *information diagram* of **Figure 1** illustrates the relationship between h_μ, r_μ, b_μ, and $\mathbf{E}$. When we change the time discretization Δt, our interpretation and definitions change somewhat, as we describe in Section 3.

Shannon's various information quantities—entropy, conditional entropy, mutual information, and the like—when applied to time series are functions of the joint distributions $\Pr(X_{0:L})$. Importantly, for a given set of random variables they define an algebra of *atoms* out of which information measures are composed (Yeung, 2008). James et al. (2011) used this to show that the past and future partition the single-measurement entropy $H(X_0)$ into the measure-theoretic atoms of **Figure 1**. These include those—r_μ and b_μ—already mentioned and the *enigmatic information*:

$$q_\mu = I[X_0; X_{:0}; X_{1:}] \ ,$$

which is the co-information between past, present, and future. One can also consider the amount of predictable information not captured by the present:

$$\sigma_\mu = I[X_{:0}; X_{1:}|X_0].$$

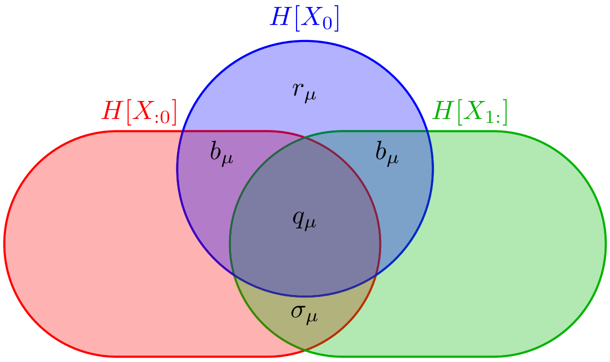

FIGURE 1 | Information diagram illustrating the anatomy of the information $H[X_0]$ in a process' single observation X_0 in the context of its past $X_{:0}$ and its future $X_{1:}$. Although the past entropy $H[X_{:0}]$ and the future entropy $H[X_{1:}]$ typically are infinite, space precludes depicting them as such. They do scale in a controlled way, however: $H[X_{-\ell:0}] \propto h_\mu \ell$ and $H[X_{1:\ell}] \propto h_\mu \ell$. The two atoms labeled b_μ are the same, since we consider only stationary processes. (After James et al., 2011, with permission.)

which is the *elusive information* (Ara et al., 2015). It measures the amount of past-future correlation not contained in the present. It is nonzero if the process has "hidden states" and is therefore quite sensitive to how the state space is "observed" or coarse-grained.

The total information in the future predictable from the past (or vice versa)—the excess entropy—decomposes into particular atoms:

$$\mathbf{E} = b_\mu + \sigma_\mu + q_\mu \ .$$

The process's Shannon entropy rate h_μ is also a sum of atoms:

$$h_\mu = r_\mu + b_\mu \ .$$

This tells us that a portion of the information (h_μ) a process spontaneously generates is thrown away (r_μ) and a portion is actively stored (b_μ). Putting these observations together gives the information anatomy of a single measurement X_0:

$$H[X_0] = q_\mu + 2b_\mu + r_\mu \ . \tag{1}$$

Although these measures were originally defined for stationary processes, they easily carry over to a nonstationary process of finite Markov order.

Calculating these information measures in closed-form given a model requires finding the *ϵ-machine*, which is constructed from causal states. Forward-time causal states $\boldsymbol{\mathcal{S}}^+$ are minimal sufficient statistics for predicting a process's future (Crutchfield and Young, 1989; Shalizi and Crutchfield, 2001). This follows from their definition—a *causal state* $\sigma^+ \in \boldsymbol{\mathcal{S}}^+$ is a sets of pasts grouped by the equivalence relation $\sim^+$:

$$\begin{aligned} x_{:0} &\sim^+ x'_{:0} \\ &\Leftrightarrow \Pr(X_{0:}|X_{:0} = x_{:0}) = \Pr(X_{0:}|X_{:0} = x'_{:0}) \ . \end{aligned} \tag{2}$$

So, $\boldsymbol{\mathcal{S}}^+$ is a set of classes—a coarse-graining of the uncountably infinite set of all pasts. At time t, we have the random variable $\mathcal{S}_t^+$ that takes values $\sigma^+ \in \boldsymbol{\mathcal{S}}^+$ and describes the *causal-state process* $\ldots, \mathcal{S}_{-1}^+, \mathcal{S}_0^+, \mathcal{S}_1^+, \ldots$. $\mathcal{S}_t^+$ is a partition of pasts $X_{:t}$ that, according to the indexing convention, does not include the present observation X_t. In addition to the set of pasts leading to it, a causal state σ_t^+ has an associated *future morph*—the conditional measure $\Pr(X_{t:}|\sigma_t^+)$ of futures that can be generated from it. Moreover, each state σ_t^+ inherits a probability $\pi(\sigma_t^+)$ from the process's measure over pasts $\Pr(X_{:t})$. The forward-time *statistical complexity* is then the Shannon entropy of the state distribution $\pi(\sigma_t^+)$ (Crutchfield and Young, 1989): $C_\mu^+ = H[\mathcal{S}_0^+]$. A generative model is constructed out of the causal states by endowing the causal-state process with transitions:

$$T_{\sigma\sigma'}^{(x)} = \Pr(\mathcal{S}_{t+1}^+ = \sigma', X_t = x | \mathcal{S}_t^+ = \sigma) \ ,$$

that give the probability of generating the next symbol x and ending in the next state σ', if starting in state σ (Residing in a state and generating a symbol do not occur simultaneously. Since symbols are generated during transitions there is, in effect, a half time-step difference in the indexes of the random variables X_t and $\mathcal{S}_t^+$. We suppress notating this.) To summarize, a process's *forward-time ϵ-machine* is the tuple $\{\mathcal{A}, \boldsymbol{\mathcal{S}}^+, \{T^{(x)} : x \in \mathcal{A}\}\}$.

For a discrete-time, discrete-alphabet process, the ϵ-machine is its minimal unifilar hidden Markov model (HMM) (Crutchfield and Young, 1989; Shalizi and Crutchfield, 2001) (For general background on HMMs see Paz, 1971; Rabiner and Juang, 1986; Rabiner, 1989). Note that the causal state set can be finite, countable, or uncountable; the latter two cases can occur even for processes generated by finite-state HMMs. *Minimality* can be defined by either the smallest number of states or the smallest entropy $H[\mathcal{S}_0^+]$ over states (Shalizi and Crutchfield, 2001). *Unifilarity* is a constraint on the transition matrices $T^{(x)}$ such that the next state σ' is determined by knowing the current state σ and the next symbol x. That is, if the transition exists, then $\Pr(\mathcal{S}_{t+1}^+|X_t = x, \mathcal{S}_t^+ = \sigma)$ has support on a single causal state.

3. Infinitesimal Time Resolution

One often treats a continuous-time renewal process, such as a spike train from a noisy integrate-and-fire neuron, in a discrete-time setting (Rieke et al., 1999). With results of Marzen and Crutchfield (2015) in hand, we can investigate how artificial time binning affects estimates of a model neuron's spike train's randomness, predictability, and information storage in the limit of infinitesimal time resolution. This is exactly the limit in which analytic formulae for information measures are most useful, since increasing the time resolution artificially increases the apparent range of temporal correlations as shown in **Figure 3**.

Time-binned neural spike trains of noisy integrate-and-fire neurons have been studied for quite some time (Mackay and McCulloch, 1952) and, despite that history, this is still an active endeavor (Rieke et al., 1999; Cessac and Cofre, 2013). Our emphasis and approach differ, though. We do not estimate statistics or reconstruct models from simulated spike train data using nonparametric inference algorithms—e.g., as done in Haslinger et al. (2010). Rather, we ask how ϵ-machines extracted from a spike train process and information measures calculated from them vary as a function of time coarse-graining. Our analytic approach highlights an important lesson about such studies in general: A process' ϵ-machine and information anatomy are sensitive to time resolution. A secondary and compensating lesson is that the manner in which the ϵ-machine and information anatomy scale with time resolution conveys much about the process' structure.

Suppose we are given a neural spike train with interspike intervals independently drawn from the same interspike interval (ISI) distribution $\phi(t)$ with mean ISI $1/\mu$. To convert the continuous-time point process into a sequence of binary spike-quiescence symbols, we track the number of spikes emitted in successive time bins of size Δt. Our goal, however, is to understand how the choice of Δt affects reported estimates for C_μ, h_μ, $\mathbf{E}$, b_μ, and σ_μ. The way in which each of these vary with Δt reveals information about the intrinsic time scales on which a process behaves; cf., the descriptions of entropy rates in Costa et al. (2002, 2005) and Gaspard and Wang (1993). We concern ourselves with the infinitesimal Δt limit, even though the behavior of these information atoms is potentially most interesting when Δt is on the order of the process' intrinsic time scales.

In the infinitesimal time-resolution limit, when Δt is smaller than any intrinsic timescale, the neural spike train is a renewal process with *interevent count distribution*:

$$F(n) \approx \phi(n\Delta t)\ \Delta t \tag{3}$$

and *survival function*:

$$w(n) \approx \int_{n\Delta t}^{\infty} \phi(t) dt\ . \tag{4}$$

The interevent distribution $F(n)$ is the probability distribution that the silence separating successive events (bins with spikes) is n counts long. While the survival function $w(n)$ is the probability that the silence separating successive events is at least n counts long. The ϵ-machine transition probabilities therefore change with Δt. The mean interevent count $\langle T \rangle + 1$ is *not* the mean interspike interval $1/\mu$ since one must convert between counts and spikes[1]:

$$\langle T \rangle + 1 = \frac{1}{\mu \Delta t}\ . \tag{5}$$

In this limit, the ϵ-machines of spike-train renewal processes can take one of the topologies described in Marzen and Crutchfield (2015).

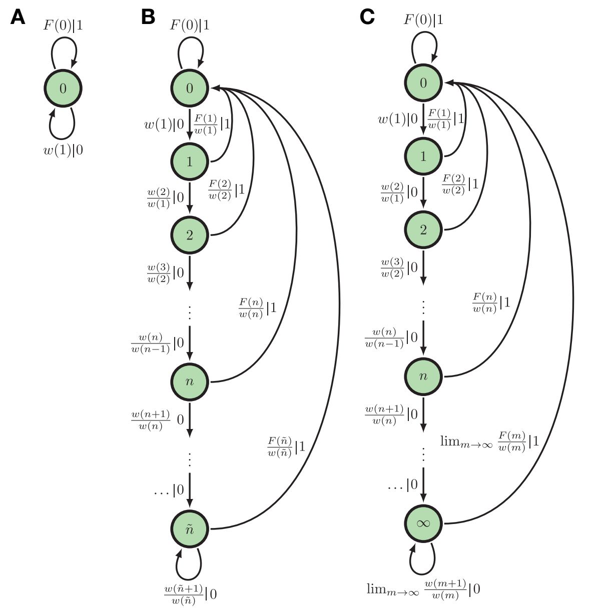

FIGURE 2 | ϵ-Machines of processes generated by Poisson neurons and by integrate-and-fire neurons (left to right): (A) The ϵ-machine for a Poisson process. **(B)** The ϵ-machine for an eventually Poisson process; i.e., a Poisson neuron with a refractory period of length $\tilde{n}\Delta t$. **(C)** The ϵ-machine for a generic renewal process—the not eventually Δ-Poisson process of Marzen and Crutchfield (2015); i.e., the process generated by noise-driven integrate-and-fire neurons. Edge labels $p|x$ denote emitting symbol x ("1" is "spike") with probability p. (Reprinted with permission from Marzen and Crutchfield, 2015.)

Here, we focus only on two of these ϵ-machine topologies. The first topology corresponds to that of an eventually Poisson process, in which the ISI distribution takes the form $\phi(t) = \phi(T)e^{-\lambda(t-T)}$ for some finite T and $\lambda > 0$. A Poisson neuron with firing rate λ and refractory period of time T, for instance, eventually ($t > T$) generates a Poisson process. Hence, we refer to them as *eventually Poisson processes*; see **Figure 2B**. A Poisson process is a special type of eventually Poisson process with $T = 0$; see **Figure 2A**. However, the generic renewal process has ϵ-machine topology shown in **Figure 2C**. Technically, only non-eventually-Δ Poisson processes have this ϵ-machine topology, but for our purposes, this is the ϵ-machine topology for any renewal process *not* generated by a Poisson neuron; see Marzen and Crutchfield (2015).

At present, inference algorithms can only infer finite ϵ-machines. So, such algorithms applied to renewal processes will yield an eventually Poisson topology. (Compare **Figure 2C** to the inferred approximate ϵ-machine of an integrate-and-fire neuron in Figure 2 in Haslinger et al., 2010.) The generic renewal process has an infinite ϵ-machine, though, for which the inferred ϵ-machines are only approximations.

We calculated $\mathbf{E}$ and C_μ using the expressions given in Marzen and Crutchfield (2015). Substituting in Equations (3), (4), and (5), we find that the excess entropy $\mathbf{E}$ tends to:

$$\begin{aligned}\lim_{\Delta t \to 0} \mathbf{E}(\Delta t) = & \int_0^\infty \mu t \phi(t) \log_2 \big(\mu \phi(t)\big) dt \\ & - 2 \int_0^\infty \mu \Phi(t) \log_2 \big(\mu \Phi(t)\big) dt\ ,\end{aligned} \tag{6}$$

where $\Phi(t) = \int_t^\infty \phi(t')dt'$ is the probability that an ISI is longer than t. It is easy to see that $\mathbf{E}(\Delta t)$ limits to a positive and (usually) finite value as the time resolution vanishes, with some exceptions described below. Similarly, using the expression in Marzen and Crutchfield (2015)'s Appendix II, one can show that the finite-time excess entropy[2] $\mathbf{E}(T)$ takes the form:

$$\begin{aligned}\lim_{\Delta t \to 0} \mathbf{E}(T) = & \left(\int_0^T \mu \Phi(t) dt \right) \log_2 \frac{1}{\mu} \\ & - 2 \int_0^T \mu \Phi(t) \log_2 \Phi(t) dt \\ & - \mu \int_T^\infty \Phi(t) dt \log_2 \left(\mu \int_T^\infty \Phi(t) dt \right)\end{aligned}$$

[1] As the subscript context makes clear, the mean count μ is not related to that μ in C_μ and related quantities. In the latter it refers to the measure $\mu(\mathbf{s})$ over bi-infinite sequences $\mathbf{s}$ generated by a process.

[2] In the theoretical neuroscience literature, $\mathbf{E}(T)$ is sometime called the predictive information $I_{pred}(T)$ and is a useful indicator of process complexity when $\mathbf{E}$ diverges (Bialek et al., 2001).

$$+ \int_0^T \mu t F(t) \log_2 F(t) dt$$
$$+ T \int_T^\infty \mu F(t) \log_2 F(t) dt \,. \tag{7}$$

As $T \to \infty$, $\mathbf{E}(T) \to \mathbf{E}$. Note that these formulae apply only when mean firing rate μ is nonzero.

Even if $\mathbf{E}$ limits to a finite value, the statistical complexity typically diverges due to its dependence on time discretization Δt. Suppose that we observe an eventually Poisson process, such that $\phi(t) = \phi(T)e^{-\lambda(t-T)}$ for $t > T$. Then, from formulae in Marzen and Crutchfield (2015), statistical complexity in the infinitesimal time-resolution limit becomes:

$$C_\mu(\Delta t) \sim \left(\mu \int_0^T \Phi(t)dt\right) \log_2 \frac{1}{\Delta t}$$
$$- \int_0^T \left(\mu\Phi(t)\right) \log_2 \left(\mu\Phi(t)\right) dt \tag{8}$$
$$- \left(\mu \int_T^\infty \Phi(t)dt\right) \log_2 \left(\mu \int_T^\infty \Phi(t)dt\right) \,,$$

ignoring terms of $O(\Delta t)$ or higher. The first term diverges, and its rate of divergence is the probability of observing a time since last spike less than T. This measures the spike train's deviation from being Δ-Poisson and so reveals the effective dimension of the underlying causal state space. C_μ's remaining nondivergent component is equally interesting. In fact, it is the differential entropy of the time since last spike distribution.

An immediate consequence of the analysis is that this generic infinitesimal renewal process is highly *cryptic* (Crutchfield et al., 2009). It hides an arbitrarily large amount of its internal state information: C_μ diverges as $\Delta t \to 0$ but $\mathbf{E}$ (usually) asymptotes to a finite value. We have very structured processes that have disproportionately little in the future to predict. Periodic processes constitute an important exception to this general rule of thumb for continuous-time processes. A neuron that fires every T seconds without jitter has $\mathbf{E} = C_\mu$, and both $\mathbf{E}$ and C_μ diverge logarithmically with $1/\Delta t$.

It is straightforward to show that any information measure contained within the present—$H[X_0]$, h_μ, b_μ, r_μ, and q_μ (recall **Figure 1**)—all vanish as Δt tends to 0. Therefore, $\lim_{\Delta t \to 0} \sigma_\mu = \lim_{\Delta t \to 0} \mathbf{E}$ and the entropy rate becomes:

$$h_\mu \sim -\mu \left(\log_2(\Delta t) + \int_0^\infty \phi(t) \log_2 \phi(t) dt\right) \Delta t \,. \tag{9}$$

With $\Delta t \to 0$, h_μ nominally tends to 0: As we shorten the observation time scale, spike events become increasingly rare. There are at least two known ways to address h_μ apparently not being very revealing when so defined. On the one hand, rather than focusing on the uncertainty per symbol, as h_μ does, we opt to look at the uncertainty per unit time: $h_\mu/\Delta t$. This is the so-called *Δt-entropy rate* (Gaspard and Wang, 1993) and it diverges as $-\mu \log \Delta t$. Such divergences are to be expected: The large literature on dimension theory characterizes a continuous set's randomness by its divergence scaling rates (Farmer et al., 1983; Mayer-Kress, 1986). Here, we are characterizing sets of similar cardinality—infinite sequences. On the other hand, paralleling sequence block-entropy definition of entropy rate ($h_\mu =_{\ell \to \infty} H[X_{0:\ell}]/\ell$) (Crutchfield and Feldman, 2003), continuous-time entropy rates are often approached within a continuous-time framework using:

$$h_\mu = \lim_{T \to \infty} H(T)/T \,,$$

where $H(T)$ is path entropy, the continuous-time analog of the block entropy $H(\ell)$ (Girardin, 2005). In these analyses, any $\log \Delta t$ terms are regularized away using Shannon's differential entropy (Cover and Thomas, 2006), leaving the nondivergent component $-\mu \int_0^\infty \phi(t) \log \phi(t) dt$. Using the Δt-entropy rate but keeping both the divergent and nondivergent components, as in Equations (8) and (9), is an approach that respects both viewpoints and gives a detailed picture of time-resolution scaling.

A major challenge in analyzing spike trains concerns locating the timescales on which information relevant to the stimulus is carried. Or, more precisely, we are often interested in estimating what percentage of the raw entropy of a neural spike train is used to communicate information about a stimulus; cf. the framing in Strong et al. (1998). For such analyses, the entropy rate is often taken to be $H(\Delta t, T)/T$, where T is the total path time and $H(\Delta t, T)$ is the entropy of neural spike trains over time T resolved at time bin size Δt. In terms of previously derived quantities and paralleling the well known block-entropy linear asymptote $H(\ell) = \mathbf{E} + h_\mu \ell$ (Crutchfield and Feldman, 2003), this is:

$$\frac{H(\Delta t, T)}{T} = \frac{h_\mu(\Delta t)}{\Delta t} + \frac{\mathbf{E}(T, \Delta t)}{T} \,.$$

From the scaling analyses above, the extensive component of $H(\Delta t, T)/T$ diverges logarithmically in the small Δt limit due to the logarithmic divergence (Equation 9) in $h_\mu(\Delta t)/\Delta t$. If we are interested in accurately estimating the entropy rate, then the above is one finite-time T estimate of it. However, there are other estimators, including:

$$\frac{H(\Delta t, T) - H(\Delta t, T - \Delta t)}{\Delta t} \approx \frac{h_\mu(\Delta t)}{\Delta t} + \frac{\partial \mathbf{E}(T, \Delta t)}{\partial T} \,.$$

This estimator converges more quickly to the true entropy rate $h_\mu(\Delta t)/\Delta t$ than does $H(\Delta t, T)/T$.

No such $\log \Delta t$ divergences occur with b_μ. Straightforward calculation, not shown here, reveals that:

$$\lim_{\Delta t \to 0} \frac{b_\mu}{\Delta t} = -\mu \left(\int_0^\infty \phi(t) \int_0^\infty \phi(t') \log_2 \phi(t + t') dt' dt\right.$$
$$\left. + \frac{1}{\log 2} - \int_0^\infty \phi(t) \log_2 \phi(t) dt\right) \,. \tag{10}$$

Since $\lim_{\Delta t \to 0} b_\mu(\Delta t)/\Delta t < \infty$ and $\lim_{\Delta t \to 0} h_\mu(\Delta t)/\Delta t$ diverges, the ephemeral information rate $r_\mu(\Delta t)/\Delta t$ also diverges as $\Delta t \to 0$. The bulk of the information generated by such renewal processes is dissipated and, having no impact on future behavior, is not useful for prediction.

Were we allowed to observe relatively microscopic membrane voltage fluctuations rather than being restricted to the relatively macroscopic spike sequence, the Δt-scaling analysis would be entirely different. Following Marzen and Crutchfield (2014) or natural extensions thereof, the statistical complexity diverges as $-\log \epsilon$, where ϵ is the resolution level for the membrane voltage, the excess entropy diverges as $\log 1/\Delta t$, the time-normalized entropy rate diverges as $\log \sqrt{2\pi e D\Delta t}/\Delta t$, and the time-normalized bound information diverges as $1/2\Delta t$. In other words, observing membrane voltage rather than spikes makes the process far more predictable. The relatively more macroscopic modeling at the level of spikes throws away much detail of the underlying biochemical dynamics.

To illustrate the previous points, we turn to numerics and a particular neural model. Consider an (unleaky) integrate-and-fire neuron driven by white noise whose membrane voltage (after suitable change of parameters) evolves according to:

$$\frac{dV}{dt} = b + \sqrt{D}\eta(t) \; , \tag{11}$$

where $\eta(t)$ is white noise such that $\langle \eta(t) \rangle = 0$ and $\langle \eta(t)\eta(t') \rangle = \delta(t - t')$. When $V = 1$, the neuron spikes and the voltage is reset to $V = 0$; it stays at $V = 0$ for a time τ, which enforces a hard refractory period. Since the membrane voltage resets to a predetermined value, the interspike intervals produced by this model are independently drawn from the same interspike interval distribution:

$$\phi(t) = \begin{cases} 0 & t < \tau \\ \sqrt{\frac{\lambda}{2\pi(t-\tau)^3}} e^{-\lambda(\mu(t-\tau)-1)^2/2(t-\tau)} & t \geq \tau \end{cases} . \tag{12}$$

Here, $1/\mu = 1/b$ is the mean interspike interval and $\lambda = 1/D$ is a shape parameter that controls ISI variance. This neural model is not as realistic as that of a linear leaky integrate-and-fire neural model (Gerstner and Kistler, 2002), but is complex enough to illustrate the points made earlier about the scaling of information measures and time resolution.

For illustration purposes, we assume that the time-binned neural spike train is well approximated by a renewal process, even when Δt is as large as one millisecond. This assumption will generally not hold, as past interevent counts could provide more detailed historical information that more precisely places the last spike within its time bin. Even so, the reported information measure estimates are still useful. The estimated h_μ is an upper bound on the true entropy rate; the reported $\mathbf{E}$ is a lower bound on the true excess entropy using the Data Processing Inequality (Cover and Thomas, 2006); and the reported C_μ will usually be a lower bound on the true process' statistical complexity.

Employing the renewal process assumption, numerical analysis corroborates the infinitesimal analysis above. **Figure 3** plots $F(n)$—the proxy for the full, continuous-time, ISI distribution—for a given set of neuronal parameter values as a function of time resolution. **Figure 4** then shows that h_μ and C_μ exhibit logarithmic scaling at millisecond time discretizations, but that $\mathbf{E}$ does not converge to its continuous-time value until we reach time discretizations on the order of hundreds of microseconds. Even when $\Delta t = 100\ \mu s$, $b_\mu(\Delta t)/\Delta t$ still has not converged to its continuous-time values.

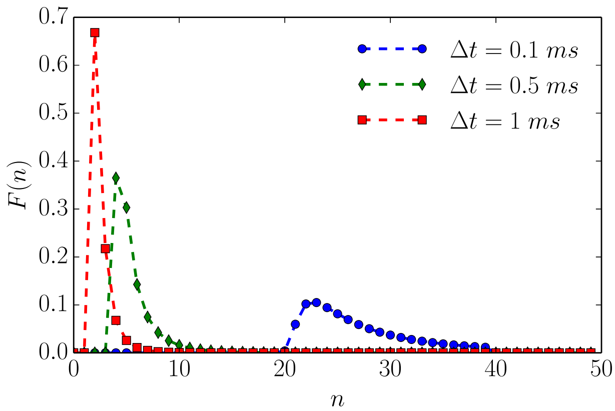

FIGURE 3 | An unleaky integrate-and-fire neuron driven by white noise has varying interevent count distributions *F*(*n*) that depend on time bin size Δ*t*. Based on the ISI distribution $\phi(t)$ given in Equation (12) with $\tau = 2$ ms, $1/\mu = 1$ ms, and $\lambda = 1$ ms. Data points represent exact values of $F(n)$ calculated for integer values of N. Dashed lines are interpolations based on straight line segments connecting nearest neighbor points.

The statistical complexity C_μ increases without bound, as $\Delta t \rightarrow 0$; see the top left panel of **Figure 4**. As suggested in the infinitesimal renewal analysis, h_μ vanishes, whereas $h_\mu/\Delta t$ diverges at a rate of $\mu \log_2 1/\Delta t$, as shown in the top right plots of **Figure 4**. As anticipated, $\mathbf{E}$ tends to a finite, ISI distribution-dependent value when Δt tends to 0, as shown in the bottom left panel in **Figure 4**. Finally, the lower right panel plots $b_\mu(\Delta t)/\Delta t$.

One conclusion from this simple numerical analysis is that one should consider going to submillisecond time resolutions to obtain accurate estimates of $\lim_{\Delta t\rightarrow 0} \mathbf{E}(\Delta t)$ and $\lim_{\Delta t\rightarrow 0} b_\mu(\Delta t)/\Delta t$, even though the calculated informational values are a few bits or even less than one bit per second in magnitude.

4. Alternating Renewal Processes

The form of the Δt-scalings discussed in Section 3 occur much more generally than indicated there. Often, our aim is to calculate the nondivergent component of these information measures as $\Delta t \rightarrow 0$, but the rates of these scalings are process-dependent. Therefore, these divergences can be viewed as a feature rather than a bug; they contain additional information about the process' structure (Gaspard and Wang, 1993).

To illustrate this point, we now investigate Δt-scalings for information measures of alternating renewal processes (ARPs), which are structurally more complex than the standard renewal processes considered above. For instance, these calculations suggest that rates of divergence of the τ-entropy rate smaller than the firing rate, such as those seen in Nemenman et al. (2008), are indicative of strong ISI correlations. Calculational details are sequestered in Appendix A.

In an ARP, an ISI is drawn from one distribution $\phi^{(1)}(t)$, then another distribution $\phi^{(2)}(t)$, then the first $\phi^{(1)}(t)$ again, and so

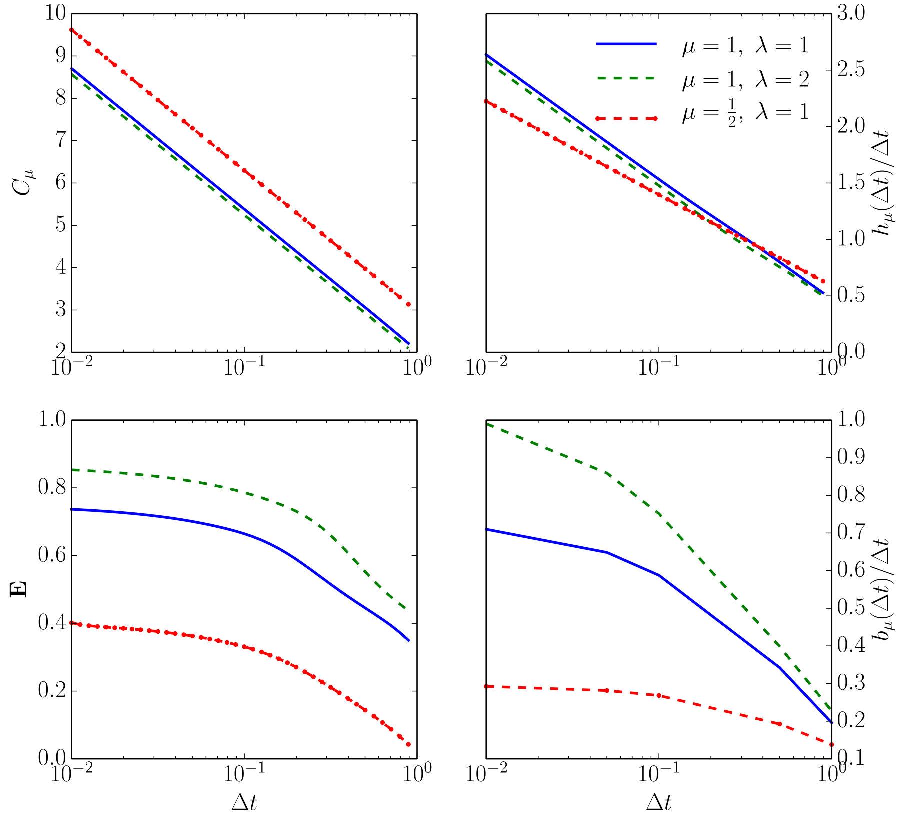

FIGURE 4 | How spike-train information measures (or rates) depend on time discretization Δt for an unleaky integrate-and-fire neuron driven by white noise. Top left: Statistical complexity C_μ as a function of both the ISI distribution shape parameters and the time bin size Δt. The horizontal axis is Δt in milliseconds on a log-scale and the vertical axis is C_μ in bits on a linear scale for three different ISI distributions following Equation (12) with $\tau = 2$ ms. **Top right:** Entropy rate h_μ also as a function of both shape parameters and Δt. Axes labeled as in the previous panel and the same three ISI distributions are used. **Bottom left:** Excess entropy E as a function of both the shape parameters and Δt. For the blue line $\lim_{\Delta t \to 0} E(\Delta t) = 0.75$ bits; purple line, $\lim_{\Delta t \to 0} E(\Delta t) = 0.86$ bits; and yellow line, $\lim_{\Delta t \to 0} E(\Delta t) = 0.41$ bits. All computed from Equation (6). **Bottom right:** Bound information rate $b_\mu(\Delta t)/\Delta t$ parameterized as in the previous panels. For the blue line $\lim_{\Delta t \to 0} b_\mu(\Delta t)/\Delta t = 0.73$ bits per second; purple line, $\lim_{\Delta t \to 0} b_\mu(\Delta t)/\Delta t = 1.04$ bits per second; and yellow line, $\lim_{\Delta t \to 0} b_\mu(\Delta t)/\Delta t = 0.30$ bits per second. All computed from Equation (10).

on. We refer to the new piece of additional information—the ISI distribution currently being drawn from—as the *modality*. Under weak technical conditions, the causal states are the modality and time since last spike. The corresponding, generic ϵ-machine is shown in **Figure 5**. We define the modality-dependent survival functions as $\Phi_i(t) = \int_t^\infty \phi^{(i)}(t')dt'$, the modality-dependent mean firing rates as:

$$\mu^{(i)} = 1 \Big/ \int_0^\infty \phi^{(i)}(t)dt \,, \tag{13}$$

the modality-dependent differential entropy rates:

$$h_\mu^{(i)} = -\mu^{(i)} \int_0^\infty \phi^{(i)} \log_2 \phi^{(i)}(t)dt \,,$$

the modality-dependent continuous-time statistical complexity:

$$C_\mu^{(i)} = -\int_0^\infty \mu^{(i)} \Phi^{(i)}(t) \log_2 \left(\mu^{(i)} \Phi^{(i)}(t)\right) dt \,,$$

and the modality-dependent excess entropy:

$$\begin{aligned}\mathbf{E}^{(i)} &= \int_0^\infty \mu^{(i)} t \phi^{(i)}(t) \log_2 \left(\mu^{(i)} \phi^{(i)}(t)\right) dt \\ &\quad - 2 \int_0^\infty \mu^{(i)} \Phi^{(i)}(t) \log_2 \left(\mu^{(i)} \Phi^{(i)}(t)\right) dt \,. \end{aligned} \tag{14}$$

It is straightforward to show, as done in Appendix A, that the time-normalized entropy rate still scales with $\log_2 1/\Delta t$:

$$\frac{h_\mu(\Delta t)}{\Delta t} \sim \tfrac{1}{2}\mu \log_2 \left(\frac{1}{\Delta t}\right) + \frac{\mu^{(2)} h_\mu^{(1)} + \mu^{(1)} h_\mu^{(2)}}{\mu^{(1)} + \mu^{(2)}}, \tag{15}$$

where $\mu = 2\frac{\mu^{(1)}\mu^{(2)}}{\mu^{(1)}+\mu^{(2)}}$. As expected, the statistical complexity still diverges:

$$\begin{aligned} C_\mu(\Delta t) &\sim 2 \log_2 \left(\frac{1}{\Delta t}\right) + \frac{\mu^{(2)} C_\mu^{(1)} + \mu^{(1)} C_\mu^{(2)}}{\mu^{(1)} + \mu^{(2)}} \\ &\quad + H_b\left(\frac{\mu_1}{\mu_1 + \mu_2}\right) \,, \end{aligned} \tag{16}$$

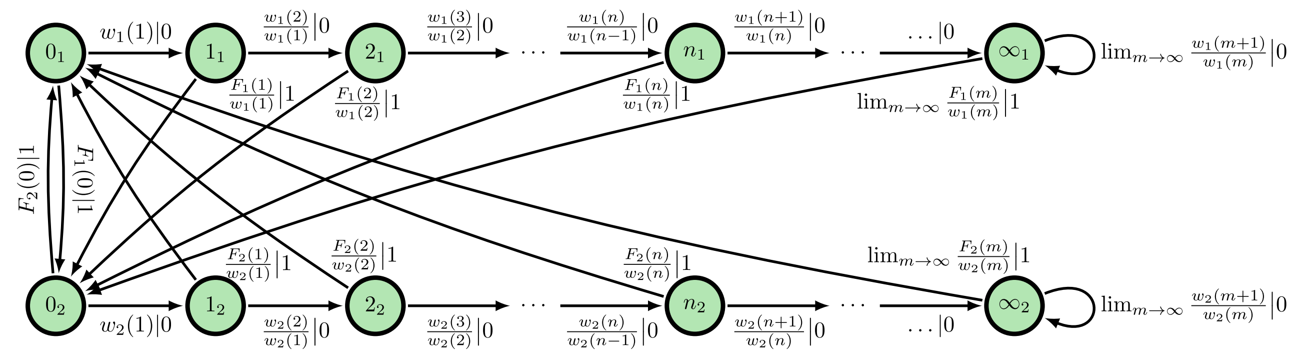
0_1
$w_1(1)|0$
1_1
$\frac{w_1(2)}{w_1(1)}|0$
2_1
$\frac{w_1(3)}{w_1(2)}|0$
$\frac{w_1(n)}{w_1(n-1)}|0$
n_1
$\frac{w_1(n+1)}{w_1(n)}|0$
$\ldots|0$
∞_1
$\lim_{m\to\infty}\frac{w_1(m+1)}{w_1(m)}|0$
$\frac{F_1(1)}{w_1(1)}|1$
$\frac{F_1(2)}{w_1(2)}|1$
$\frac{F_1(n)}{w_1(n)}|1$
$\lim_{m\to\infty}\frac{F_1(m)}{w_1(m)}|1$
$F_2(0)|1$
$F_1(0)|1$
$\frac{F_2(1)}{w_2(1)}|1$
$\frac{F_2(2)}{w_2(2)}|1$
$\frac{F_2(n)}{w_2(n)}|1$
$\lim_{m\to\infty}\frac{F_2(m)}{w_2(m)}|1$
0_2
$w_2(1)|0$
1_2
$\frac{w_2(2)}{w_2(1)}|0$
2_2
$\frac{w_2(3)}{w_2(2)}|0$
$\frac{w_2(n)}{w_2(n-1)}|0$
n_2
$\frac{w_2(n+1)}{w_2(n)}|0$
$\ldots|0$
∞_2
$\lim_{m\to\infty}\frac{w_2(m+1)}{w_2(m)}|0$

FIGURE 5 | ϵ-Machine for an alternating renewal process in which neither interevent count distribution is Δ-Poisson and they are not equal almost everywhere. State label n_m denotes n counts since the last event and present modality m.

where $H_b(p) = -p\log_2 p - (1-p)\log_2(1-p)$ is the entropy in bits of a Bernoulli random variable with bias p. Finally, the excess entropy still limits to a positive constant:

$$\lim_{\Delta t\to 0}\mathbf{E}(\Delta t) = H_b\left(\frac{\mu_1}{\mu_1+\mu_2}\right) + \frac{\mu^{(2)}\mathbf{E}^{(1)} + \mu^{(1)}\mathbf{E}^{(2)}}{\mu^{(1)}+\mu^{(2)}} . \quad (17)$$

The additional terms $H_b(\cdot)$ come from the information stored in the time course of modalities.

As a point of comparison, we ask what these information measures would be for the original (noncomposite) renewal process with the same ISI distribution as the ARP. As described in Appendix B, the former entropy rate is always less than the true h_μ; its statistical complexity is always less than the true C_μ; and its excess entropy is always smaller than the true **E**. In particular, the ARP's h_μ divergence rate is always less than or equal to the mean firing rate μ. Interestingly, this coincides with what was found empirically in the time series of a single neuron; see Figure 5C in Nemenman et al. (2008).

The ARPs here are a first example of how one can calculate information measures of the much broader and more structurally complex class of processes generated by unifilar hidden semi-Markov models, a subclass of hidden semi-Markov models (Tokdar et al., 2010).

5. Information Universality

Another aim of ours is to interpret the information measures. In particular, we wished to relate infinitesimal time-resolution excess entropies, statistical complexities, entropy rates, and bound information rates to more familiar characterizations of neural spike trains—firing rates μ and ISI coefficient of variations C_V. To address this, we now analyze a suite of familiar single-neuron models. We introduce the models first, describe the parameters behind our numerical estimates, and then compare the information measures.

Many single-neuron models, when driven by temporally uncorrelated and stationary input, produce neural spike trains that are renewal processes. We just analyzed one model class, the noisy integrate-and-fire (NIF) neurons in Section 3, focusing on time-resolution dependence. Other common neural models include the linear leaky integrate-and-fire (LIF) neuron, whose dimensionless membrane voltage, after a suitable change of parameters, fluctuates as:

$$\frac{dV}{dt} = b - V + a\eta(t) , \quad (18)$$

and when $V = 1$, a spike is emitted and V is instantaneously reset to 0. We computed ISI survival functions from empirical histograms of 10^5 ISIs; we varied $b \in [1.5, 5.75]$ in steps of 0.25 and $a \in [0.1, 3.0]$ in steps of 0.1 to $a = 1.0$ and in steps of 0.25 thereafter.

The quadratic integrate-and-fire (QIF) neuron has membrane voltage fluctuations that, after a suitable change of variables, are described by:

$$\frac{dV}{dt} = b + V^2 + a\eta(t) , \quad (19)$$

and when $V = 100$, a spike is emitted and V is instantaneously reset to -100. We computed ISI survival functions from empirical histograms of trajectories with 10^5 ISIs; we varied $b \in [0.25, 4.75]$ in steps of 0.25 and $a \in [0.25, 2.75]$ in steps of 0.25. The QIF neuron has a very different dynamical behavior from the LIF neuron, exhibiting a Hopf bifurcation at $b = 0$. Simulation details are given in Appendix B.

Finally, ISI distributions are often fit to gamma distributions, and so we also calculated the information measures of spike trains with gamma-distributed ISIs (GISI).

Each neural model—NIF, LIF, QIF, and GISI—has its own set of parameters that governs its ISI distribution shape. Taken at face value, this would make it difficult to compare information measures across models. Fortunately, for each of these neural models, the firing rate μ and coefficient of variation C_V uniquely determine the underlying model parameters (Vilela and Lindner, 2009). As Appendix B shows, the quantities $\lim_{\Delta t\to 0}\mathbf{E}(\Delta t)$, $\lim_{\Delta t\to 0} C_\mu + \log_2(\mu\Delta t)$, $\lim_{\Delta t\to 0} h_\mu(\Delta t)/\mu\Delta t + \log_2(\mu\Delta t)$, and $\lim_{\Delta t\to 0} b_\mu(\Delta t)/\mu\Delta t$ depend only on the ISI coefficient of variation C_V and not the mean firing rate μ.

We estimated information measures from the simulated spike train data using plug-in estimators based on the formulae in Section 3. Enough data was generated that even naive plug-in estimators were adequate *except* for estimating b_μ when C_V was larger than 1. See Appendix B for estimation details. That said, binned estimators are likely inferior to binless entropy estimators (Victor, 2002), and naive estimators tend to have large biases. This will be an interesting direction for future research, since a detailed analysis goes beyond the present scope.

Figure 6 compares the statistical complexity, excess entropy, entropy rate, and bound information rate for all four neuron types as a function of their C_V. Surprisingly, the NIF, LIF, and QIF neuron's information measures have essentially identical dependence on C_V. That is, the differences in mechanism do not strongly affect these informational properties of the spike trains they generate. Naturally, this leads one to ask if the informational indifference to mechanism generalizes to other spike train model classes and stimulus-response settings.

Figure 6's top left panel shows that the continuous-time statistical complexity grows monotonically with increasing C_V. In particular, the statistical complexity increases logarithmically with ISI mean and approximately linearly with the ISI coefficient of variation C_V. That is, the number of bits that must be stored to predict these processes increases in response to additional process stochasticity and longer temporal correlations. In fact, it is straightforward to show that the statistical complexity is minimized and excess entropy maximized at fixed μ when the neural spike train is periodic. This is unsurprising since, in the space of processes, periodic processes are least cryptic ($C_\mu - \mathbf{E} = 0$) and so knowledge of oscillation phase is enough to completely predict the future. (See Appendix B.)

The bottom left panel in **Figure 6** shows that increasing C_V tends to decrease the excess entropy **E**—the number of bits that one can predict about the future. **E** diverges for small C_V, dips at the C_V where the ISI distribution is closest to exponential, and limits to a small number of bits at large C_V. At small C_V, the neural spike train is close to noise-free periodic behavior. When analyzed at small but nonzero Δt, **E** encounters an "ultraviolet divergence" (Tchernookov and Nemenman, 2013). Thus, **E** diverges as $C_V \to 0$, and a simple argument in Appendix B suggests that the rate of divergence is $\log_2(1/C_V)$. At an intermediate $C_V \sim 1$, the ISI distribution is as close as

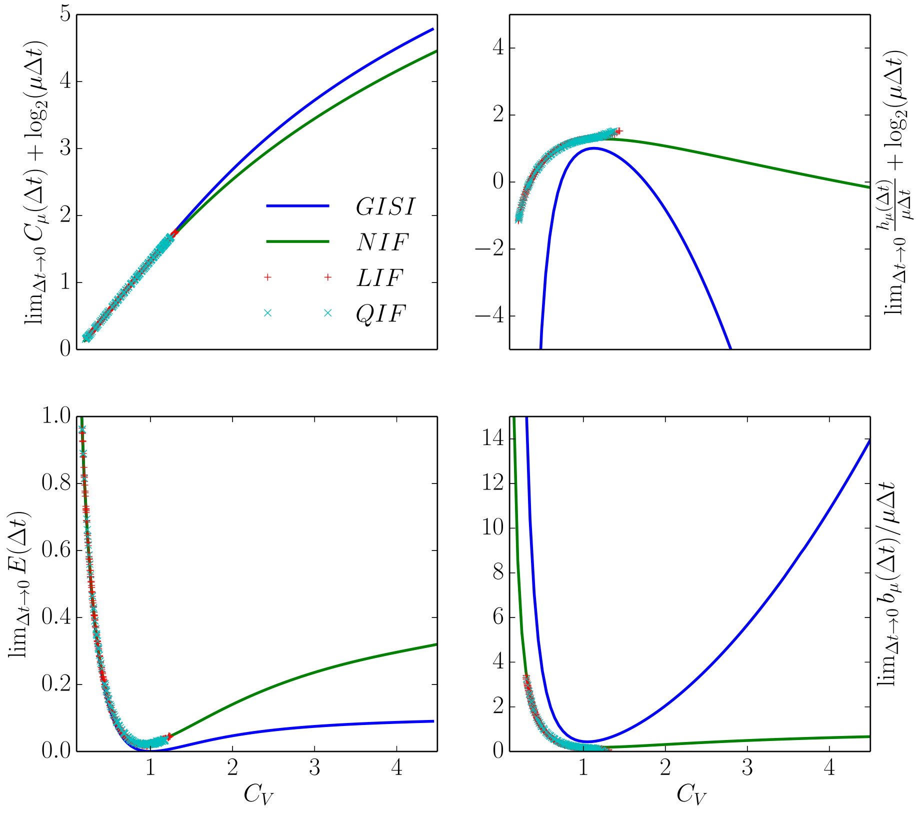

FIGURE 6 | Information universality across distinct neuron dynamics. We find that several information measures depend only on the ISI coefficient of variation C_V and not the ISI mean firing rate μ for the following neural spike train models: (i) neurons with Gamma distributed ISIs (GISI, blue), (ii) noisy integrate-and-fire neurons governed by Equation (11) (NIF, green), (iii) noisy linear leaky integrate-and-fire neurons governed by Equation (18) (LIF, dotted red), and (iv) noisy quadratic integrate-and-fire neurons governed by Equation (19) (QIF, dotted blue). **Top left:** $\lim_{\Delta t \to 0} C_\mu(\Delta t) + \log_2(\mu \Delta t)$. **Top right:** $\lim_{\Delta t \to 0} h_\mu(\Delta t)/\mu \Delta t + \log_2(\mu \Delta t)$. **Bottom left:** $\lim_{\Delta t \to 0} E(\Delta t)$. **Bottom right:** $\lim_{\Delta t \to 0} b_\mu(\Delta t)/\mu \Delta t$. In the latter, ISI distributions with smaller C_V were excluded due to the difficulty of accurately estimating $\int_0^\infty \int_0^\infty \phi(t)\phi(t') \log_2 \phi(t+t') dt dt'$ from simulated spike trains. See text for discussion.

possible to that of a memoryless Poisson process and so **E** is close to vanishing. At larger C_V, the neural spike train is noise-driven. Surprisingly, completely noise-driven processes still have a fraction of a bit of predictability: knowing the time since last spike allows for some power in predicting the time to next spike.

The top right panel shows that an appropriately rescaled differential entropy rate varies differently for neural spike trains from noisy integrate-and-fire neurons and neural spike trains with gamma-distributed ISIs. As expected, the entropy rate is maximized at C_V near 1, consistent with the Poisson process being the maximum entropy distribution for fixed mean ISI. Gamma-distributed ISIs are far less random than ISIs from noisy integrate-and-fire neurons, holding μ and C_V constant.

Finally, the continuous-time bound information (b_μ) rate varies in a similar way to **E** with C_V. (Note that since the plotted quantity is $\lim_{\Delta t \to 0} b_\mu(\Delta t)/\mu \Delta t$, one could interpret the normalization by $1/\mu$ as a statement about how the mean firing rate μ sets the natural timescale.) At low C_V, the b_μ rate diverges as $1/C_V^2$, as described in Appendix B. Interestingly, this limit is singular, similar to the results in Marzen and Crutchfield (2014): at $C_V = 0$, the spike train is noise-free periodic and so the b_μ rate is 0. For $C_V \approx 1$, it dips for the same reason that **E** decreases. For larger C_V, b_μ's behavior depends rather strongly on the ISI distribution shape. The longer-ranged gamma-distribution results in ever-increasing b_μ rate for larger C_V, while the b_μ rate of neural spike trains produced by NIF neurons tends to a small positive constant at large C_V. The variation of b_μ deviates from that of **E** qualitatively at larger C_V in that the GISI spike trains yield smaller total predictability **E** than that of NIF neurons, but arbitrarily higher predictability *rate*.

These calculations suggest a new kind of universality for neuronal information measures *within a particular generative model class*. All of these distinct integrate-and-fire neuron models generate ISI distributions from different families, yet their informational properties exhibit the same dependencies on Δt, μ, and C_V in the limit of small Δt. Neural spike trains with gamma-distributed ISIs did not show similar informational properties. And, we would not expect neural spike trains that are alternating renewal processes to show similar informational properties either. (See Section 4.) These coarse information quantities might therefore be effective model selection tools for real neural spike train data, though more groundwork must be explored to ascertain their utility.

6. Conclusions

We explored the scaling properties of a variety of information-theoretic quantities associated with two classes of spiking neural models: renewal processes and alternating renewal processes. We found that information generation (entropy rate) and stored information (statistical complexity) both diverge logarithmically with decreasing time resolution for both types of spiking models, whereas the predictable information (excess entropy) and active information accumulation (bound information rate) limit to a constant. Our results suggest that the excess entropy and regularized statistical complexity of different types of integrate-and-fire neurons are universal in the sense that they do not depend on mechanism details, indicating a surprising simplicity in complex neural spike trains. Our findings highlight the importance of analyzing the scaling behavior of information quantities, rather than assessing these only at a fixed temporal resolution.

By restricting ourselves to relatively simple spiking models we have been able to establish several key properties of their behavior. There are, of course, other important spiking models that cannot be expressed as renewal processes or alternating renewal processes, but we are encouraged by the robust scaling behavior of the entropy rate, statistical complexity, excess entropy, and bound information rate over the range of models we considered.

There was a certain emphasis here on the entropy rate and hidden Markov models of neural spike trains, both familiar tools in computational neuroscience. On this score, our contributions are straightforward. We determined how the entropy rate varies with the time discretization and identified the possibly infinite-state, unifilar HMMs required for optimal prediction of spike-train renewal processes. Entropy rate diverges logarithmically for stochastic processes (Gaspard and Wang, 1993), and this has been observed empirically for neural spike trains for time discretizations in the submillisecond regime (Nemenman et al., 2008). We argued that the h_μ divergence rate is an important characteristic. For renewal processes, it is the mean firing rate; for alternating renewal processes, the "reduced mass" of the mean firing rates. Our analysis of the latter, more structured processes showed that a divergence rate less than the mean firing rate—also seen experimentally (Nemenman et al., 2008)—indicates that there are strong correlations between ISIs. Generally, the nondivergent component of the time discretization-normalized entropy rate is the differential entropy rate; e.g., as given in Stevens and Zador (1996).

Empirically studying information measures as a function of time resolution can lead to a refined understanding of the time scales over which neuronal communication occurs. Regardless of the information measure chosen, the results and analysis here suggest that much can be learned by studying scaling behavior rather than focusing only on neural information as a single quantity estimated at a fixed temporal resolution. While we focused on the regime in which the time discretization was smaller than any intrinsic timescale of the process, future and more revealing analyses would study scaling behavior at even smaller time resolutions to directly determine intrinsic time scales (Crutchfield, 1994).

Going beyond information generation (entropy rate), we analyzed information measures—namely, statistical complexity and excess entropy—that have only recently been used to understand neural coding and communication. Their introduction is motivated by the hypothesis that neurons benefit from learning to predict their inputs (Palmer et al., 2013), which can consist of the neural spike trains of upstream neurons. The statistical complexity is the minimal amount of historical information required for exact prediction. To our knowledge, the statistical complexity has appeared only once previously in computational neuroscience (Haslinger et al., 2010). The excess entropy, a closely related companion, is the maximum amount of information that can be predicted about the future. When it diverges, then its divergence rate is quite revealing of the underlying process (Crutchfield, 1994; Bialek et al., 2001), but none of the model neural spike trains studied here had divergent excess entropy. Finally, the bound information rate has yet to be deployed in the context of neural coding, though related quantities have drawn attention elsewhere, such as in nonlinear dynamics (James et al., 2014), music (Abdallah and Plumbley, 2009), spin systems (Abdallah and Plumbley, 2012), and information-based reinforcement learning (Martius et al., 2013). Though its potential uses have yet to be exploited, it is an interesting quantity in that it captures the rate at which spontaneously generated information is actively stored by neurons. That is, it quantifies how neurons harness randomness.

Our contributions to this endeavor are more substantial than the preceding points. We provided exact formulae for the above quantities for renewal processes and alternating renewal processes. The new expressions can be developed further as lower bounds and empirical estimators for a process' statistical complexity, excess entropy, and bound information rate. This parallels how the renewal-process entropy-rate formula is a surprisingly accurate entropy-rate estimator (Gao et al., 2008). By deriving explicit expressions, we were able to analyze time-resolution scaling, showing that the statistical complexity diverges logarithmically for all but Poisson processes. So, just like the entropy rate, any calculations of the statistical complexity—e.g., as in Haslinger et al. (2010)—should be accompanied by the time discretization dependence. Notably, the excess entropy and the bound information rate have no such divergences.

To appreciate more directly what neural information processing behavior these information measures capture in the continuous-time limit, we studied them as functions of the ISI coefficient of variation. With an appropriate renormalization, simulations revealed surprising simplicity: a universal dependence on the coefficient of variation across several familiar neural models. The simplicity is worth investigating further since the dynamics and biophysical mechanisms implicit in the alternative noisy integrate-and-fire neural models are quite different. If other generative models of neural spike trains also show similar information universality, then these information measures might prove useful as model selection tools.

Finally, we close with a discussion of a practical issue related to the scaling analyses—one that is especially important given the increasingly sophisticated neuronal measurement technologies coming online at a rapid pace (Alivisatos et al., 2012). How small should Δt be to obtain correct estimates of neuronal communication? First, as we emphasized, there is no single "correct" estimate for an information quantity, rather its resolution scaling is key. Second, results presented here and in a previous study by others (Nemenman et al., 2008) suggest that extracting information scaling rates and nondivergent components can require submillisecond time resolution. Third, and to highlight, the regime of infinitesimal time resolution is exactly the limit in which computational efforts without analytic foundation will fail or, at a minimum, be rather inefficient. As such, we hope that the results and methods developed here will be useful to these future endeavors and guide how new technologies facilitate scaling analysis.

Acknowledgments

The authors thank C. Hillar and R. G. James for helpful comments. They also thank the Santa Fe Institute for its hospitality during visits. JC is an SFI External Faculty member. This material is based upon work supported by, or in part by, the U.S. Army Research Laboratory and the U.S. Army Research Office under contracts W911NF-13-1-0390 and W911NF-13-1-0340. SM was funded by a National Science Foundation Graduate Student Research Fellowship and the U.C. Berkeley Chancellor's Fellowship.

References

Abdallah, S. A., and Plumbley, M. D. (2009). Information dynamics: patterns of expectation and surprise in the perception of music. *Connect. Sci.* 21, 89–117. doi: 10.1080/09540090902733756

Abdallah, S. A., and Plumbley, M. D. (2012). A measure of statistical complexity based on predictive information with application to finite spin systems. *Phys. Lett. A*. 376, 275–281. doi: 10.1016/j.physleta.2011.10.066

Alivisatos, A. P., Chun, M., Church, G. M., Greenspan, R. J., Roukes, M. L., and Yuste, R. (2012). The brain activity map project and the challenge of functional connectomics. *Neuron* 74, 970–974. doi: 10.1016/j.neuron.2012.06.006

Ara, P. M., James, R. G., and Crutchfield, J. P. (2015). The elusive present: hidden past and future dependence and why we build models. Available online at: http://arXiv.org:1507.00672 [cond-mat.stat-mech].

Archer, E., Park, I. M., and Pillow, J. W. (2012). Bayesian estimation of discrete entropy with mixtures of stick-breaking priors. *Adv. Neural Info. Proc. Sys.* 25, 2015–2023.

Atick, J. J. (1992). "Could information theory provide an ecological theory of sensory processing?" in *Princeton Lectures on Biophysics*, ed W. Bialek (Singapore: World Scientific), 223–289.

Averbeck, B. B., Latham, P. E., and Pouget, A. (2006). Neural correlations, population coding and computation. *Nat. Rev. Neurosci.* 7, 358–366. doi: 10.1038/nrn1888

Barlow, H. B. (1961). "Possible principles underlying the transformation of sensory messages," in *Sensory Communication*, ed W. Rosenblith (Cambridge, MA: MIT Press), 217–234.

Bell, A. J., Mainen, Z. F., Tsodyks, M., and Sejnowski, T. J. (1995). *Balancing Conductances may Explain Irregular Cortical Firing*. Technical Report, Institute for Neural Computation, San Diego.

Berry, M. J., Warland, D. K., and Meister, M. (1997). The structure and precision of retinal spike trains. *Proc. Natl. Acad. Sci. U.S.A.* 94, 5411–5416. doi: 10.1073/pnas.94.10.5411

Bialek, W., Nemenman, I., and Tishby, N. (2001). Predictability, complexity, and learning. *Neural Comp.* 13, 2409–2463. doi: 10.1162/089976601753195969

Bialek, W., Ruderman, D. L., and Zee, A. (1991). "Optimal sampling of natural images: a design principle for the visual system?" in *Advances in Neural Information Processing 3*, eds R. P. Lippman, J. E. Moody and D. S. Touretzky (San Mateo, CA: Morgan Kaufmann), 363–369.

Britten, K. H., Newsome, W. T., Shadlen, M. N., Celebrini, S., and Movshon, J. A. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Vis. Neurosci.* 13, 87–100. doi: 10.1017/S095252380000715X

Butts, D. A., and Goldman, M. S., (2006). Tuning curves, neuronal variability, and sensory coding. *PLoS Biol.* 4:e92. doi: 10.1371/journal.pbio.0040092

Cessac, B., and Cofre, R. (2013). Spike train statistics and Gibbs distributions. Available online at: http://arXiv.org:1302.5007.

Costa, M., Goldberger, A. L., and Peng, C. K. (2002). Multiscale entropy analysis of complex physiologic time series. *Phys. Rev. Lett.* 89:068102. doi: 10.1103/PhysRevLett.89.068102

Costa, M., Goldberger, A. L., and Peng, C. K. (2005). Multiscale entropy analysis of biological signals. *Phys. Rev. E* 71:021906. doi: 10.1103/PhysRevE.71.021906

Cover, T. M., and Thomas, J. A. (2006). *Elements of Information Theory, 2nd Edn.* New York, NY: Wiley-Interscience.

Crutchfield, J. P. (1994). The calculi of emergence: Computation, dynamics, and induction. *Physica D* 75, 11–54. doi: 10.1016/0167-2789(94)90273-9

Crutchfield, J. P., Ellison, C. J., and Mahoney, J. R. (2009). Time's barbed arrow: Irreversibility, crypticity, and stored information. *Phys. Rev. Lett.* 103:094101. doi: 10.1103/PhysRevLett.103.094101

Crutchfield, J. P., and Feldman, D. P. (2003). Regularities unseen, randomness observed: levels of entropy convergence. *CHAOS* 13, 25–54. doi: 10.1063/1.1530990

Crutchfield, J. P., and Young, K. (1989). Inferring statistical complexity. *Phys. Rev. Let.* 63, 105–108. doi: 10.1103/PhysRevLett.63.105

Dan, Y., Atick, J. J., and Reid, R. C. (1996). Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory. *J. Neurosci.* 16, 3351–3362.

Destexhe, A., Rudolph, M., and Paré, D. (2003). The high-conductance state of neocortical neurons *in vivo*. *Nat. Rev. Neurosci.* 4, 739–751. doi: 10.1038/nrn1198

Deweese, M. (1996). Optimization principles for the neural code. *Netw. Comp. Neural Sys.* 7, 325–331. doi: 10.1088/0954-898X/7/2/013

DeWeese, M. R., and Meister, M. (1999). How to measure the information gained from one symbol. *Network* 10, 325–340. doi: 10.1088/0954-898X/10/4/303

DeWeese, M. R., Wehr, M., and Zador, A. M. (2003). Binary spiking in auditory cortex. *J. Neurosci.* 23, 7940–7949.

DeWeese, M. R., and Zador, A. M. (2006). Non-Gaussian membrane potential dynamics imply sparse, synchronous activity in auditory cortex. *J. Neurosci.* 26, 12206–12218. doi: 10.1523/JNEUROSCI.2813-06.2006

Farmer, J. D., Ott, E., and Yorke, J. A. (1983). The dimension of chaotic attractors. *Physica* 7D, 153. doi: 10.1007/978-0-387-21830-4/11

Gao, Y., Kontoyiannis, I., and Bienenstock, E. (2008). Estimating the entropy of binary time series: methodology, some theory and a simulation study. *Entropy* 10, 71–99. doi: 10.3390/entropy-e10020071

Gaspard, P., and Wang, X.-J. (1993). Noise, chaos, and (ϵ, τ)-entropy per unit time. *Phys. Rep.* 235, 291–343. doi: 10.1016/0370-1573(93)90012-3

Gerstner, W., and Kistler, W. M. (2002). *Spiking Neuron Models: Single Neurons, Populations, Plasticity*. Cambridge, UK: Cambridge University Press.

Girardin, V. (2005). "On the different extensions of the ergodic theorem of information theory," in *Recent Advances in Applied Probability Theory*, eds R. Baeza-Yates, J. Glaz, H. Gzyl, J. Husler, and J. L. Palacios (New York, NY: Springer), 163–179.

Haslinger, R., Klinkner, K. L., and Shalizi, C. R. (2010). The computational structure of spike trains. *Neural Comp.* 22, 121–157. doi: 10.1162/neco.2009.12-07-678

Jacobs, A. L., Fridman, G., Douglas, R. M., Alam, N. M., Latham, P. E., Prusky, G. T., et al. (2009). Ruling out and ruling in neural codes. *Proc. Natl. Acad. Sci. U.S.A.* 106, 5937–5941. doi: 10.1073/pnas.0900573106

James, R. G., Burke, K., and Crutchfield, J. P. (2014). Chaos forgets and remembers: measuring information creation, destruction, and storage. *Phys. Lett. A* 378, 2124–2127. doi: 10.1016/j.physleta.2014.05.014

James, R. G., Ellison, C. J., and Crutchfield, J. P. (2011). Anatomy of a bit: information in a time series observation. *CHAOS* 21:037109. doi: 10.1063/1.3637494

Koepsell, K., Wang, X., Hirsch, J. A., and Sommer, F. T. (2010). Exploring the function of neural oscillations in early sensory systems. *Front. Neurosci.* 4, 53–61. doi: 10.3389/neuro.01.010.2010

Laughlin, S. B. (1981). A simple coding procedure enhances a neuron's information capacity. *Z. Naturforsch.* 36c, 910–912.

Linsker, R. (1989). "An application of the principle of maximum information preservation to linear systems," in *Advances in Neural Information Processing 1*, ed D. Touretzky (San Mateo, CA: Morgan Kaufmann), 186–194.

London, M., Roth, A., Beeren, L., Häusser, M., and Latham, P. E. (2010). Sensitivity to perturbations *in vivo* implies high noise and suggests rate coding in cortex. *Nature* 466, 123–128. doi: 10.1038/nature09086

Mackay, D. M., and McCulloch, W. W. (1952). The limiting information capacity of a neuronal link. *Bull. Math. Biophys.* 14, 127–135. doi: 10.1007/BF02477711

Martius, G., Der, R., and Ay, N. (2013). Information driven self-organization of complex robotics behaviors. *PLoS ONE* 8:e63400. doi: 10.1371/journal.pone.0063400

Marzen, S., and Crutchfield, J. P. (2014). Information anatomy of stochastic equilibria. *Entropy* 16, 4713–4748. doi: 10.3390/e16094713

Marzen, S., and Crutchfield, J. P. (2015). Informational and causal architecture of discrete-time renewal processes. *Entropy* 17, 4891–4917.

Mayer-Kress, G., (ed.). (1986). *Dimensions and Entropies in Chaotic Systems: Quantification of Complex Behavior*. Berlin: Springer.

Meister, M., Lagnado, L., and Baylor, D. A. (1995). Concerted signaling by retinal ganglion cells. *Science* 270, 1207–1210. doi: 10.1126/science.270.5239.1207

Nemenman, I., Bialek, W., and de Ruyter van Steveninck, R. R. (2004). Entropy and information in neural spike trains: progress on the sampling problem. *Phys. Rev. E* 69, 1–6. doi: 10.1103/PhysRevE.69.056111

Nemenman, I., Lewen, G. D., Bialek, W., and de Ruyter van Steveninck, R. R. (2008). Neural coding of natural stimuli: Information at sub-millisecond resolution. *PLoS Comp. Bio.* 4:e1000025. doi: 10.1371/journal.pcbi.1000025

Nirenberg, S., Carcieri, S. M., Jacobs, A. L., and Latham, P. E. (2001). Retinal ganglion cells act largely as independent encoders. *Nature* 411, 698–701. doi: 10.1038/35079612

Palmer, S. E., Marre, O., Berry, II, M. J., and Bialek, W. (2013). Predictive information in a sensory population. *Proc. Natl. Acad. Sci. U.S.A.* 112, 6908–6913. doi: 10.1073/pnas.1506855112

Panzeri, S., Treves, A., Schultz, S., and Rolls, E. T. (1999). On decoding the responses of a population of neurons from short time windows. *Neural Comp.* 11, 1553–1577. doi: 10.1162/089976699300016142

Paz, A. (1971). *Introduction to Probabilistic Automata*. New York, NY: Academic Press.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications. *IEEE Proc.* 77:257.

Rabiner, L. R., and Juang, B. H. (1986). An introduction to hidden Markov models. *IEEE ASSP Mag.* 4–16. doi: 10.1109/MASSP.1986.1165342

Reinagel, P., and Reid, R. C. (2000). Temporal coding of visual information in the thalamus. *J. Neurosci.* 20, 5392–5400.

Rieke, F., Warland, D., de Ruyter van Steveninck, R., and Bialek, W. (1999). *Spikes: Exploring the Neural Code*. New York, NY: Bradford Book.

Sakitt, B., and Barlow, H. B. (1982). A model for the economical encoding of the visual image in cerebral cortex. *Biol. Cybern.* 43, 97–108. doi: 10.1007/BF00336972

Schneidman, E., Berry, M. J., Segev, R., and Bialek, W. (2006). Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* 440, 1007–1012. doi: 10.1038/nature04701

Schneidman, E., Bialek, W., and Berry, M. J. (2003). Synergy, redundancy, and independence in population codes. *J. Neurosci.* 23, 11539–11553.

Shadlen, M. N., and Newsome, W. T. (1995). Is there a signal in the noise? *Curr. Opin. Neurobiol.* 5, 248–250. doi: 10.1016/0959-4388(95)80033-6

Shadlen, M. N., and Newsome, W. T. (1998). The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding. *J. Neurosci.* 18, 3870–3896.

Shalizi, C. R., and Crutchfield, J. P. (2001). Computational mechanics: Pattern and prediction, structure and simplicity. *J. Stat. Phys.* 104, 817–879. doi: 10.1023/A:1010388907793

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Sys. Tech. J.* 27, 379–423, 623–656. doi: 10.1002/j.1538-7305.1948.tb00917.x

Softky, W. R., and Koch, C. (1993). The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *J. Neurosci.* 13, 334–350.

Srinivasan, M. V., Laughlin, S. B., and Dubs, A. (1982). Predictive coding: A fresh view of inhibition in the retina. *Proc. R. Soc. Lond. Ser. B* 216, 427–459. doi: 10.1098/rspb.1982.0085

Stein, R. B. (1967). The information capacity of neurons using a frequency code. *Biophys. J.* 7, 797–826. doi: 10.1016/S0006-3495(67)86623-2

Stevens, C. F., and Zador, A. (1996). "Information through a spiking neuron," in *Advance Neural Information Processing System*, eds D. Touretzky, M. C. Mozer, and M. E. Hasselmo (Cambridge, MA: MIT Press), 75–81.

Stevens, C. F., and Zador, A. M. (1998). Input synchrony and the irregular firing of cortical neurons. *Nat. Neurosci.* 1, 210–217. doi: 10.1038/659

Strong, S. P., Koberle, R., de Ruyter van Steveninck, R., and Bialek, W. (1998). Entropy and information in neural spike trains. *Phys. Rev. Lett.* 80, 197–200. doi: 10.1103/PhysRevLett.80.197

Tchernookov, M., and Nemenman, I. (2013). Predictive information in a nonequilibrium critical model. *J. Stat. Phys.* 153, 442–459. doi: 10.1007/s10955-013-0833-6

Theunissen, F. E., and Miller, J. P. (1991). Representation of sensory information in the cricket cercal sensory system. ii: information theoretic calculation of system accuracy and optimal tuning curve widths of four primary interneurons. *J. Neurophys.* 66, 1690–1703.

Tokdar, S., Xi, P., Kelly, R. C., and Kass, R. E. (2010). Detection of bursts in extracellular spike trains using hidden semi-Markov point process models. *J. Comput. Neurosci.* 29, 203–212. doi: 10.1007/s10827-009-0182-2

Treves, A., and Panzeri, S. (1995). The upward bias in measures of information derived from limited data samples. *Neural Comp.* 7, 399–407. doi: 10.1162/neco.1995.7.2.399

Verdú, S., and Weissman, T. (2006). "Erasure entropy," in *IEEE International Symposium on Information Theory (ISIT 2006)* (Seattle, WA), 98–102. doi: 10.1109/ISIT.2006.261682

Victor, J. D. (2002). Binless strategies for estimation of information from neural data. *Phys. Rev. E* 66:051903. doi: 10.1103/PhysRevE.66.051903

Vilela, R. D., and Lindner, B. (2009). Are the input parameters of white noise driven integrate and fire neurons uniquely determined by rate and CV? *J. Theo. Bio.* 257, 90–99. doi: 10.1016/j.jtbi.2008.11.004

Yang, Y., and Zador, A. M. (2012). Differences in sensitivity to neural timing among cortical areas. *J. Neurosci.* 32, 15142–15147. doi: 10.1523/JNEUROSCI.1411-12.2012

Yeung, R. W. (2008). *Information Theory and Network Coding*. New York, NY: Springer.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Appendix A

Alternating Renewal Process Information Measures

A discrete-time alternating renewal process draws counts from $F_1(n)$, then $F_2(n)$, then $F_1(n)$, and so on. We now show that the modality and counts since last event are causal states when $F_1 \neq F_2$ almost everywhere and when neither F_1 nor F_2 is eventually Δ-Poisson. We present only a proof sketch.

Two pasts $x_{:0}$ and $x'_{:0}$ belong to the same causal state when $\Pr(X_{0:}|X_{:0} = x_{:0}) = \Pr(X_{0:}|X_{:0} = x'_{:0})$. We can describe the future uniquely by a sequence of interevent counts $\mathcal{N}_i$, $i \geq 1$, and the counts till next event $\mathcal{N}'_0$. Likewise, we could describe the past as a sequence of interevent counts $\mathcal{N}_i$, $i < 0$, and the counts since last event $\mathcal{N}_0 - \mathcal{N}'_0$. Let $\mathcal{M}_i$ be the modality at time step i. So, for instance, $\mathcal{M}_0$ is the present modality.

First, we claim that one can infer the present modality from a semi-infinite past almost surely. The probability that the present modality is 1 having observed the last $2M$ events is:

$$\begin{aligned}\Pr(\mathcal{M}_0 = 1|\mathcal{N}_{-2M:-1} = n_{-2M:-1}) \\ = \prod_{i=-1,\mathrm{odd}}^{2M} F_2(n_i)F_1(n_{i-1}) \, .\end{aligned}$$

Similarly, the probability that the present modality is 2 having observed the last $2M$ events is:

$$\begin{aligned}\Pr(\mathcal{M}_0 = 2|\mathcal{N}_{-2M:-1} = n_{-2M:-1}) \\ = \prod_{i=-1,\mathrm{odd}}^{2M} F_1(n_i)F_2(n_{i-1}) \, .\end{aligned}$$

We are better served by thinking about the normalized difference of the corresponding log likelihoods:

$$Q\!: = \frac{1}{2M} \log \frac{P(\mathcal{M}_0 = 1|\mathcal{N}_{-2M:-1} = n_{-2M:-1})}{P(\mathcal{M}_0 = 2|\mathcal{N}_{-2M:-1} = n_{-2M:-1})} \, .$$

Some manipulation leads to:

$$Q = \frac{1}{2}\Big(\frac{1}{M}\sum_{i=-1,\mathrm{odd}}^{2M} \log \frac{F_2(n_i)}{F_1(n_i)} + \frac{1}{M}\sum_{i=-1,\mathrm{even}}^{2M} \log \frac{F_1(n_i)}{F_2(n_i)}\Big),$$

and, almost surely in the limit of $M \to \infty$:

$$\frac{1}{M}\sum_{i=-1,\mathrm{odd}}^{2M} \log \frac{F_1(n_i)}{F_2(n_i)} \to \begin{cases} D[F_2||F_1] & \mathcal{M}_0 = 1 \\ -D[F_1||F_2] & \mathcal{M}_0 = 2 \end{cases}, \tag{A1}$$

where $D[P||Q]$ is the information gain between P and Q (Cover and Thomas, 2006). And, we also have:

$$\frac{1}{M}\sum_{i=-1,\mathrm{even}}^{2M} \log \frac{F_2(n_i)}{F_1(n_i)} \to \begin{cases} -D[F_1||F_2] & \mathcal{M}_0 = 1 \\ D[F_2||F_1] & \mathcal{M}_0 = 2 \end{cases} .$$

This implies that:

$$\lim_{M\to\infty} Q = \frac{D[F_2||F_1] - D[F_1||F_2]}{2} \begin{cases} 1 & \mathcal{M}_0 = 1 \\ -1 & \mathcal{M}_0 = 2 \end{cases} .$$

We only fail to identify the present modality almost surely from the semi-infinite past if $\lim_{M\to\infty} Q = 0$. Otherwise, the unnormalized difference of the log likelihoods:

$$\log \frac{\Pr(\mathcal{M}_0 = 1|\mathcal{N}_{:-1} = n_{:-1})}{\Pr(\mathcal{M}_0 = 2|\mathcal{N}_{:-1} = n_{:-1})}$$

tends to ±∞, implying that one of the two probabilities has vanished. From the expression, $\lim_{M\to\infty} Q = 0$ only happens when $D[F_2||F_1] = D[F_1||F_2]$. However, equality requires that $F_1(n) = F_2(n)$ almost everywhere.

Given the present modality, we also need to know the counts since the last event in order to predict the future as well as possible. The proof of this is very similar to those given in Marzen and Crutchfield (2015). The conditional probability distribution of future given past is:

$$\Pr(X_{0:}|X_{:0} = x_{:0}) = \Pr(\mathcal{N}_{1:}|\mathcal{N}_0, X_{:0} = x_{:0}) \Pr(\mathcal{N}_0|X_{:0} = x_{:0}) \, .$$

Since the present modality is identifiable from the past $x_{:0}$, and since interevent counts are independent given modality:

$$\Pr(\mathcal{N}_{1:}|\mathcal{N}_0, X_{:0} = x_{:0}) = \Pr(\mathcal{N}_{1:}|M_0 = m_0(n_{:-1})) \, .$$

So, it is necessary to know the modality in order to predict the future as well as possible. By virtue of how the alternating renewal process is generated, the second term is:

$$\Pr(\mathcal{N}_0|X_{:0} = x_{:0}) = \Pr(\mathcal{N}_0|\mathcal{N}'_0 = n'_0, M_0 = m_0(n_{:-1})) \, .$$

A very similar term was analyzed in Marzen and Crutchfield (2015), and that analysis revealed that it was necessary to store the counts since last spike when neither F_1 nor F_2 is eventually Δ-Poisson.

Identifying causal states $\mathcal{S}^+$ as the present modality $\mathcal{M}_0$ and the counts since last event $\mathcal{N}'_0$ immediately allows us to calculate the statistical complexity and entropy rate. The entropy rate can be calculated via:

$$\begin{aligned} h_\mu &= H[X_0|\mathcal{M}_0, \mathcal{N}'_0] \\ &= \pi(\mathcal{M}_0 = 1)H[X_0|\mathcal{M}_0 = 1, \mathcal{N}'_0] \\ &\quad + \pi(\mathcal{M}_0 = 2)H[X_0|\mathcal{M}_0 = 2, \mathcal{N}'_0] \, . \end{aligned}$$

The statistical complexity is:

$$\begin{aligned} C_\mu &= H[\mathcal{S}^+] \\ &= H[\mathcal{M}_0, \mathcal{N}'_0] \\ &= H[\mathcal{M}_0] + \pi(\mathcal{M}_0 = 1)H[\mathcal{N}'_0|\mathcal{M}_0 = 1] \\ &\quad + \pi(\mathcal{M}_0 = 2)H[\mathcal{N}'_0|\mathcal{M}_0 = 2] \, . \end{aligned} \tag{A2}$$

Finally, it is straightforward to show that the modality $\mathcal{M}_1$ at time step 1 and the counts to next event are the reverse-time causal states under the same conditions on F_1 and F_2. Therefore:

$$\begin{aligned}\mathbf{E} &= I[\mathcal{S}^+;\mathcal{S}^-]\\ &= I[\mathcal{M}_0,\mathcal{N}_0';\mathcal{M}_1,\mathcal{N}_0-\mathcal{N}_0']\\ &= I[\mathcal{M}_0;\mathcal{M}_1,\mathcal{N}_0-\mathcal{N}_0']\\ &\qquad + I[\mathcal{N}_0';\mathcal{M}_1,\mathcal{N}_0-\mathcal{N}_0'|\mathcal{M}_0]\ .\end{aligned}$$

One can continue in this way to find formulae for other information measures of a discrete-time alternating renewal process.

These formulae can be rewritten terms of the modality-dependent information measures of Equations (13) and (14) if we recognize two things. First, the probability of a particular modality is proportional to the average amount of time spent in that modality. Second, for reasons similar to those outlined in Marzen and Crutchfield (2015), the probability of counts since last event given a particular present modality i is proportional to $w_i(n)$. Hence, in the infinitesimal time discretization limit, the probability of modality 1 is:

$$\pi(\mathcal{M}_0=1)=\frac{\mu^{(1)}}{\mu^{(1)}+\mu^{(2)}}$$

and similarly for modality 2. Then, the entropy rate out of modality i is:

$$H[X_1|\mathcal{M}_0=i,\mathcal{N}_0']\sim\Delta t\left(\mu^{(i)}\log_2\frac{1}{\Delta t}+h_\mu^{(i)}(\Delta t)\right)\ ,$$

and the modality-dependent statistical complexity diverges as:

$$H[\mathcal{N}_0'|\mathcal{M}_0=i]\sim\log_2 1/\Delta t+C_\mu(\Delta t)\ .$$

Finally, in continuous-time $\mathcal{M}_0$ and $\mathcal{M}_1$ limit to the same random variable, such that:

$$\lim_{\Delta t\to 0}\mathbf{E}(\Delta t)=H[\mathcal{M}_0]+\lim_{\Delta t\to 0}I[\mathcal{N}_0';\mathcal{N}_0-\mathcal{N}_0'|\mathcal{M}_0]\ .$$

Note that $\mathbf{E}^{(i)}=\lim_{\Delta t\to 0}I[\mathcal{N}_0';\mathcal{N}_0-\mathcal{N}_0'|\mathcal{M}_0=i]$.

Bringing these results together, we substitute the above components into Equation (A2)'s expression for C_μ and, after details not shown here, find the expression quoted in the main text as Equation (16). Similarly, for h_μ and $\mathbf{E}$, yielding the the formulae presented in the main text in Equations (15) and (17), respectively.

As a last task, as our hypothetical null model, we wish to find the information measures for the corresponding renewal process approximation. The ISI distribution of the alternating renewal process is:

$$\phi(t)=\frac{\mu^{(2)}\phi^{(1)}(t)+\mu^{(1)}\phi^{(2)}(t)}{\mu^{(1)}+\mu^{(2)}} \tag{A3}$$

and its survival function is:

$$\Phi(t)=\frac{\mu^{(2)}\Phi^{(1)}(t)+\mu^{(1)}\Phi^{(2)}(t)}{\mu^{(1)}+\mu^{(2)}}\ . \tag{A4}$$

Hence, its mean firing rate is:

$$\mu=\frac{1}{1/\mu^{(1)}+1/\mu^{(2)}}\ . \tag{A5}$$

From Section 3, the entropy rate of the corresponding renewal process is:

$$\frac{h_\mu^{\mathrm{ren}}(\Delta t)}{\Delta t}\sim\mu\log_2\frac{1}{\Delta t}+\mu H[\phi(t)]\ ;$$

compare Equation (15). And, the statistical complexity of the corresponding renewal process is:

$$C_\mu^{\mathrm{ren}}(\Delta t)\sim\log_2\frac{1}{\Delta t}+H[\mu\Phi(t)]\ .$$

The rate of divergence of $C_\mu^{\mathrm{ren}}(\Delta t)$ is half the rate of divergence of the true $C_\mu(\Delta t)$, as given in Equation (16). Trivial manipulations, starting from $0\leq\left(\frac{1}{\mu^{(1)}}-\frac{1}{\mu^{(2)}}\right)^2$, imply that the rate of entropy-rate divergence is always less than or equal to the mean firing rate for an alternating renewal process. Jensen's inequality implies that each of the nondivergent components of these information measures for the renewal process is less than or equal to that of the alternating renewal process. The Data Processing Inequality (Cover and Thomas, 2006) also implies that the excess entropy calculated by assuming a renewal process is a lower bound on the true process' excess entropy.

Appendix B

Simplicity in Complex Neurons

Recall that our white noise-driven linear leaky integrate-and-fire (LIF) neuron has governing equation:

$$\dot{V}=b-V+a\eta(t)\ , \tag{A6}$$

and, when $V=1$, a spike is emitted and V is instantaneously reset to 0. We computed ISI survival functions from empirical histograms of 10^5 ISIs. These ISIs were obtained by simulating Equation (A6) in Python/NumPy using an Euler integrator with time discretization of $1/1000$ of $\log b/(b-1)$, which is the ISI in the noiseless limit.

The white noise-driven quadratic integrate-and-fire (QIF) neuron has governing equation:

$$\dot{V}=b+V^2+a\eta(t)\ , \tag{A7}$$

and, when $V=100$, a spike is emitted and V is instantaneously reset to -100. We computed ISI survival functions also from empirical histograms of trajectories with 10^5 ISIs. These ISIs were obtained by simulating Equation (A7) in Python/NumPy using an Euler stochastic integrator with time discretization of $1/1000$ of $\sqrt{\pi/b}$, which is the ISI in the noiseless limit when threshold and reset voltages are $+\infty$ and $-\infty$, respectively.

Figure 6 shows estimates of the following continuous-time information measures from this simulated data as they vary with mean firing rate μ and ISI coefficient of variation C_V. This required us to estimate μ, C_V, and:

$$\begin{aligned}
C_\mu^{CT} &:= \lim_{\Delta t\to 0}(C_\mu(\Delta t)+\log_2 \Delta t),\\
\mathbf{E}^{CT} &:= \lim_{\Delta t\to 0}\mathbf{E}(\Delta t)\ ,\\
h_\mu^{CT} &:= \lim_{\Delta t\to 0}\left(\frac{h_\mu(\Delta t)}{\Delta t}+\mu\log_2 \Delta t\right),\ \text{and}\\
b_\mu^{CT} &:= \lim_{\Delta t\to 0}\frac{b_\mu(\Delta t)}{\Delta t}\ ,
\end{aligned}$$

where the superscript CT is a reminder that these are appropriately regularized information measures in the continuous-time limit.

We estimated μ and C_V using the sample mean and sample coefficient of variation with sufficient samples so that error bars (based on studying errors as a function of data size) were negligible. The information measures required new estimators, however. From the formulae in Section 3, we see that:

$$C_\mu^{CT} = \log_2\frac{1}{\mu}-\mu\int_0^\infty \Phi(t)\log_2\Phi(t)dt\ , \tag{A8}$$

$$\mathbf{E}^{CT} = \int_0^\infty \mu t\phi(t)\log_2(\mu\phi(t))dt - 2\int_0^\infty \mu\Phi(t)\log_2\Phi(t)dt\ , \tag{A9}$$

$$h_\mu^{CT} = -\mu\int_0^\infty \phi(t)\log_2\phi(t)\ ,\ \text{and} \tag{A10}$$

$$b_\mu^{CT} = -\mu\Big(\int_0^\infty \phi(t)\int_0^\infty \phi(t')\log_2\phi(t+t')dt'dt + \frac{1}{\log 2}-\int_0^\infty \phi(t)\log_2\phi(t)dt\Big)\ . \tag{A11}$$

It is well known that the sample mean is a consistent estimator of the true mean, that the empirical cumulative density function is a consistent estimator of the true cumulative density function almost everywhere, and thus that the empirical ISI distribution is a consistent estimator of the true cumulative density function almost everywhere. In estimating the empirical cumulative density function, we introduced a cubic spline interpolator. This is still a consistent estimator as long as $\Phi(t)$ is three-times differentiable, which is the case for ISI distributions from integrate-and-fire neurons. We then have estimators of C_μ^{CT}, $\mathbf{E}^{CT}$, h_μ^{CT}, and b_μ^{CT} that are based on consistent estimators of μ, $\Phi(t)$, and $\phi(t)$ and that are likewise consistent.

We now discuss the finding evident in **Figure 6**, that the quantities $\lim_{\Delta t\to 0}\mathbf{E}(\Delta t)$ and $\lim_{\Delta t\to 0} C_\mu + \log_2(\mu\Delta t)$ depend only on the ISI coefficient of variation C_V and not the mean firing rate μ. Presented in a different way, this is not so surprising. First, we use Marzen and Crutchfield (2015)'s expression for C_μ to rewrite:

$$\begin{aligned}
Q_1 &= \lim_{\Delta t\to 0}\big(C_\mu(\Delta t)+\log_2(\mu\Delta t)\big)\\
&= -\mu\int_0^\infty \Phi(t)\log_2\Phi(t)dt
\end{aligned}$$

and Equation (6) to rewrite:

$$\begin{aligned}
Q_2 &= \lim_{\Delta t\to 0}\mathbf{E}(\Delta t)\\
&= 2Q_1+\int_0^\infty \mu t\phi(t)\log_2(\mu\phi(t))dt\ .
\end{aligned}$$

So, we only need to show that $-\mu\int_0^\infty \Phi(t)\log_2\Phi(t)dt$ and $\int_0^\infty \mu t\phi(t)\log_2(\mu\phi(t))dt$ are independent of μ for two-parameter families of ISI distributions.

Consider a change of variables from t to $t'=\mu t$; then:

$$Q_1 = -\int_0^\infty \Phi\left(t'/\mu\right)\log_2\left(\Phi\left(t'/\mu\right)\right)dt' \tag{A12}$$

and

$$Q_2 = 2Q_1+\int_0^\infty t'\phi\left(t'/\mu\right)\log_2\left(\phi\left(t'/\mu\right)\right)dt'\ . \tag{A13}$$

For all of the ISI distributions considered here, $\phi\left(\frac{t'}{\mu}\right)$ is still part of the same two-parameter family as $\phi(t)$, except that its mean firing rate is 1 rather than μ. Its C_V is unchanged. Hence, Q_1 and Q_2 are the same for a renewal process with mean firing rate 1 and μ, as long as the C_V is held constant. It follows that $\lim_{\Delta t\to 0}\mathbf{E}(\Delta t)$ and $\lim_{\Delta t\to 0} C_\mu+\log_2(\mu\Delta t)$ are independent of μ and only depend on C_V for the two-parameter families of ISI distributions considered in Section 5. Similar arguments apply to understanding the universal C_V-dependence of $\lim_{\Delta t\to 0} b_\mu(\Delta t)/\mu\Delta t$ and $\lim_{\Delta t\to 0} h_\mu(\Delta t)/\mu\Delta t+\log_2(\mu\Delta t)$.

In **Figure 6**, we also see that $\mathbf{E}$ seems to diverge as $C_V\to 0$. Consider the following plausibility argument that suggests it diverges as $\log_2 1/C_V$ as $C_V\to 0$. These two-parameter ISI distributions with finite mean firing rate μ and small $C_V \ll 1$ can be approximated as Gaussians with mean $1/\mu$ and standard deviation C_V/μ. Recall from Equation (6) that we have:

$$\begin{aligned}
\mathbf{E} &= -2\int_0^\infty \mu\Phi(t)\log_2(\mu\Phi(t))dt\\
&\quad+\int_0^\infty \mu t\phi(t)\log_2(\mu\phi(t))dt\\
&= -\log_2\mu-2\mu\int_0^\infty \Phi(t)\log_2\Phi(t)dt\\
&\quad+\mu\int_0^\infty t\phi(t)\log_2\phi(t)dt\ .
\end{aligned}$$

Note that as $C_V\to 0$:

$$\Phi(t)\to\begin{cases}1 & t<\frac{1}{\mu}\\ \frac{1}{2} & t=\frac{1}{\mu}\\ 0 & t>\frac{1}{\mu}\end{cases} \tag{A14}$$

and so:

$$\lim_{C_V\to 0}\int_0^\infty \Phi(t)\log_2\Phi(t)dt = 0\ .$$

We assumed that for small C_V, we can approximate:

$$\phi(t) \approx \frac{1}{\sqrt{2\pi C_V^2/\mu^2}} \exp\left(-\frac{(\mu t-1)^2}{2C_V^2}\right) ,$$

which then implies that:

$$\mu \int_0^\infty t\phi(t) \log_2 \phi(t) dt \approx \log_2 \frac{\mu\sqrt{2\pi}}{C_V} - \frac{1}{2}. \tag{A15}$$

So, for any ISI distribution tightly distributed about its mean ISI, we expect:

$$\mathbf{E} \approx \log_2 \frac{1}{C_V} ,$$

so that **E** diverges in this way. A similar asymptotic analysis also shows that as $C_V \to 0$,

$$\lim_{\Delta t\to 0} \frac{b_\mu(\Delta t)}{\Delta t} \approx \frac{1}{\log 2}(\frac{1}{2C_V^2} - \frac{1}{2}) , \tag{A16}$$

thereby explaining the divergence of $\lim_{\Delta t\to 0} b_\mu(\Delta t)/\Delta t$ evident in **Figure 6**.

Finally, a straightforward argument shows that C_μ is minimized at fixed μ when the neural spike train is periodic. We can rewrite C_μ in the infinitesimal time resolution limit as:

$$C_\mu(\Delta t) \sim \log_2\left(\frac{1}{\mu\Delta t}\right) + \mu \int_0^\infty \Phi(t) \log_2 \frac{1}{\Phi(t)} dt .$$

Note that $0 \leq \Phi(t) \leq 1$, and so $\int_0^\infty \Phi(t) \log_2 \frac{1}{\Phi(t)} dt \geq 0$. We set it equal to zero by using the step function given in Equation (A14), which corresponds to a noiseless periodic process. So, the lower bound on $C_\mu(\Delta t)$ is $\log_2 1/\mu\Delta t$, and this bound is achieved by a periodic process.