On the Biological Signalling, Information and Estimation Limits of Birth Processes

Kris V Parag

***Abstract*—Cellular signalling involves networks of small, interacting molecular populations and requires precise event timing, despite the presence of noise. The Poisson channel is an important communication model for such noisy interactions, providing a fundamental link between timing precision, information transfer and the rate of signalling. Constraints on this rate limit the Poisson channel capacity, which in turn bounds the precision with which molecular networks can solve estimation problems. We investigate these bounds as a function of signalling rate constraints, for problems in which information about a target molecular species, to be estimated, is encoded in the birth rate of a signalling species. Birth-following is a known heuristic encoder that asserts the maximum signalling rate until every target birth is recorded, then deactivates. Here we derive birth-following as a minimum time signalling (bang-bang) code, and prove that it outperforms these precision bounds for general target birth rate functions over arbitrary signalling network architectures. The simplest of these networks commonly models the dynamics of long-lived proteins. Birth-following is therefore an important reference strategy when the maximum signalling rate is high and the mean rate is no smaller than that of the target molecule. Discreteness is important in this regime. We examine the limit of this regime by removing the maximum signalling constraint. This leads to a Poisson channel with infinite capacity, which should allow completely precise timing. However, we find perfect estimation unrealisable unless the mean condition is maintained. The relationship between estimation and information is therefore not as simple or intuitive as in analogous problems on Gaussian channels. Higher Poisson capacities do not always imply better precision and realisable performance is more dependent on finding suitable encoders that capitalise on the information structure of the signalling problem of interest. There is a need for new information-theoretic metrics that can better account for both the gap between achievable and theoretical precision, and the idiosyncrasies of the Poisson channel.**

I. Introduction

Cell biology is characterised by the interactions of networks of small molecular populations. Small populations are intrinsically noisy, undergoing fluctuations due to the random timing of birth (synthesis) and death (turnover) reactions. In cellular signalling or transduction, such populations may represent genetic copy numbers, messenger protein counts or hormone levels which interact with one another by modulating or catalysing their rates of reaction [1]. The signals are the stochastically delayed changes in molecular population size that result from these modulations. A consequence of this framework is that intracellular signalling is, fundamentally, a noisy information theoretic problem [2].

Since biochemical networks appear to attenuate, filter, or effectively utilise this noise to reliably communicate [3], a wealth of research has focussed on demystifying what constitutes effective signalling and on what sets the limits of cellular network performance [4] [5] [6] [7] . These examinations have converged to an understanding that the information capacity of signalling pathways may hold the desired answers [2]. As a result, many recent analyses have attempted to compute these capacities using classical information theoretic measures such as conditional entropy, mutual information and signal to noise ratio [8] [9] [10]. These works are predicated on a 'black box' Gaussian channel description of signalling pathways [5] [6]. While many useful insights have emerged from this approach, the signalling-information-performance link remains, largely, a mystery [8].

KVP did this work in the Control Group, University of Cambridge, Trumpington Street, Cambridge, CB2 1PZ, e-mail: kvpmea@hotmail.com

One reason for this is that Gaussian channels, while easy to work with, do not properly describe the discrete information carrying structures within molecular networks. Under Gaussian communication the received signal is equivalent to the transmitted message corrupted by additive white noise [11]. The parameter defining channel capacity is the signal to noise variance ratio, and the received signal is a continuous waveform [11]. However, in molecular signalling both the information and noise manifests in the timing of discrete events [12]. Noise is also not additive.

A key insight of [13] was the realisation that signalling pathways are actually Poisson channels [14] that convert an input modulating rate into an output Poisson process with intensity equal to that input. Limits on the input rate restrict the Poisson information capacity and hence control the precision of output event timing. The Poisson channel therefore connects discrete signalling to information. Molecular interactions can then be modelled as control or estimation problems under Poisson communication. Using this formulation [13] showed how Poisson capacity constraints lead to a bound on the achievable precision in estimation or control [13], hence revealing a signalling-information-performance relationship.

While these bounds are a significant advance towards properly treating timing information, they are not perfect. Provided non-linear signalling is allowed, a heuristic strategy, known as birth-following, can outperform the bound by exploiting an inherent diffusion approximation [15] [16]. This result holds under certain channel constraints and raises three main questions, which form the motivation and subject matter of this paper.

First, what are optimal yet biologically implementable ways of encoding information in the timing of events? While many analyses are concerned with calculating capacities [8], few studies explore optimal signalling designs. This is surprising because the capacity of a channel may not be biologically actionable. Effective performance in usage depends on how well a chosen encoder matches its channel [11]. We will derive birth-following as the minimum time encoder, establishing it as a natural and effective signalling protocol.

This stimulates the second question. How general is the violation of the bound under birth-following? Mismatches suggests that discreteness is sufficiently dominant for continuity approximations to break down. The more this phenomena is observed the stronger the case for using sharp, non-linear, discrete event based codes like birth-following. We will find that the bound remains violable under both arbitrary signalling networks and generalised target birth rates.

Third, how do channel constraints influence realisable performance? The Poisson capacity and hence timing precision is, under most constraints, a function of the mean and maximum signalling rate [17] (see Eq. (2)). Birth-following outperforms the bound when the maximum rate is high. If we let this maximum become unbounded, the channel capacity becomes infinite. Perfect estimation should then be possible, since we can transmit infinite information. However, we find that this is not the case unless a mean signalling condition is also satisfied. Hence we uproot the belief that higher capacities necessitate better performance [8].

We investigate these questions in the context of pure birth process estimation. Our results therefore apply to stable molecules such as long-lived proteins [1]. We focus on these systems because (i) they are the elementary components of more complex networks, (ii) they are still not completely understood, and most importantly (iii) we are interested in threshold based signalling protocols [18]. Our simple birth model corresponds to that used in a recent study of threshold motifs [12].

Many biological systems activate or signal when proteins or messengers accumulate above some threshold (integrate and fire). Event timing precision in such systems is directly related to how

this threshold is exceeded [12]. In neuroscience, where this motif is extensively studied in an information theoretic framework, it was recently found that a failure to model discreteness and causality in event timing can lead to deceiving conclusions [19]. This reflects the particulate information structure in such systems. We attempt to uncover this embedded structure by designing causal and discrete encoders (signalling strategies) in regions of parameter space where continuity approximations are insufficient.

In this paper Methods recaps Poisson channel, distortion bound and birth-following theory. The time optimality of birth-following is proven in Section III-A of the Results. Section III-B, and Section III-C explore generalisations of birth-following, proving that it outperforms precision bounds when the maximum rate is high and mean rate is no smaller than the target rate. Section III-D examines how death reactions disrupt the problem information structure. When the maximum signalling rate is unconstrained the Poisson channel has infinite capacity. Section III-E and Section III-F find that it is still not always possible to achieve perfect precision under this condition. Biological and information theoretic implications of the results are provided in the Discussion.

II. Methods

A. Information Theoretic Bounds

We define and adapt the information theoretic (distortion) bound of [13] for fundamental birth process estimation. Let $\mathbb{Z}_a^b$ denote the integer set $\{a, a+1, \ldots, b-1, b\}$ with $b > a$. Our target or estimated molecule is X_1 and the signalling molecule is X_j, $j \in \mathbb{Z}_2^+$. The respective populations of these species at time $t \geq 0$ are $x_1(t)$ and $x_j(t)$. We will usually drop the t index for convenience. Populations fluctuate randomly in time due to birth, $x_i \to x_i + 1$ (denoted x_i^+) or death, $x_i \to x_i - 1$ (x_i^-), reactions. All X_i follow Markov birth-death processes ($i = \{1, j\}$). We want to causally estimate x_1, given the signalling history ${x_j}_0^t := \{x_j(s) : 0 \leq s \leq t\}$. Information about X_1 enters X_j via its birth (signalling) rate, f, which is also known as the channel encoder, and has full system knowledge.

$$x_j \xrightarrow{f=f({x_1}_0^t, {x_j}_0^t)} x_j + 1 \tag{1}$$

The channel decoder is some function that results in an estimate of x_1, $\tilde{x}_1 = g\left({x_j}_0^t \,|\, f\right)$. We measure performance with the mean squared error (mse), $J := \mathbb{E}\left[(x_1 - g)^2\right]$ and define the minimum mse (mmse) as $J^* = \min_{\{g\}} J$. The mmse is achieved by the optimal decoder $\hat{x}_1 = g^* = \mathbb{E}\left[x_1 \,|\, f, {x_j}_0^t\right]$ [20]. However, finding g^* is often analytically intractable. We therefore aim to find good f and g pairs (a codec). For convenience we will often drop the explicit dependence of g or g^* on f.

The constraints on the encoder f determine the form of the capacity of the Poisson channel mediating the information flow between X_1 and X_j. We constrain f so that $\mathbb{E}[f] \leq \langle f \rangle$ and $\max(f) \leq f_{\max}$. This leads to the mean-maximum capacity of Eq. (2) [14].

$$C = \langle f \rangle \log\left(\langle f \rangle^{-1} f_{\max}\right) \tag{2}$$

We can relax the maximum constraint to the norm condition $\mathbb{E}\left[f^p\right]^{\frac{1}{p}} \leq f_{\max}$ for $p \in \mathbb{Z}_1^\infty$. In these cases the capacity is $C_p = \frac{p}{p-1} C$ [17]. We only focus on feedforward information flow in this work. However, even if causal feedback was included so that x_1 also modulates x_j, the capacity remains unchanged [17]. These points suggest a fundamental link between timing precision (high fidelity Poisson transmissions) and the mean-maximum properties of the signalling rate.

A major result of [13], was the realisation that this finite capacity implies a lower bound, D, on the achievable mmse distortion (an upper bound on performance). D is derived by approximating the dynamics of X_1 with a suitable diffusion process. If X_1 is subject to the reactions $x_1 \xrightarrow{u} x_1 + 1$, $x_1 \xrightarrow{k_1 x_1} x_1 - 1$ then Eq. (3) gives the appropriate diffusion equation with w as a standard Wiener process. The bound then follows from Eq. (4) [13] [15], with J as the mse under any encoder and $k_1 \langle x_1 \rangle = \langle u \rangle$ due to equilibrium.

$$\mathrm{d}x_1 = (u - k_1 x_1)\, \mathrm{d}t + \sqrt{2 k_1 \langle x_1 \rangle}\, \mathrm{d}w \tag{3}$$

$$D = \langle u \rangle \left(C + k_1\right)^{-1} \leq J^* \leq J \tag{4}$$

This bound holds for all non-linear encoders and is actually a conservative lower limit on the mmse. Tighter bounds do exist under the constraint of linear encoding, but these are likely invoilable and so not treated here [13]. In this work we will modify and adapt Eq. (3) and Eq. (4) to derive performance bounds for various signalling-estimation problems.

B. Birth-Following Estimation

Birth-following encoding was presented in [15] [16] and shown to achieve $J^* < D$, under certain constraint conditions, for several estimation problems. Here we condense the main results of those papers into a theorem for birth processes. Consider the simplest non-trivial reaction scheme in which there are no deaths, u is a constant, and $x_1 \xrightarrow{u} x_1 + 1$ with $j = 2$ in equation 1. Then $X_1 \sim \mathrm{Po}(ut)$ (Poisson distribution) while X_2 admits a Cox process [21] [22]. This reaction set is an elementary motif for more complex signalling networks. The diffusion approximation for this birth-process is given in Eq. (5), by simplifying Eq. (3).

$$\mathrm{d}x_1 = u\, \mathrm{d}t + \sqrt{u}\, \mathrm{d}w \tag{5}$$

To solve the estimation problem we need to design f. Piecewise continuous processes which switch between 0 and $f_{\max}$ achieve Poisson channel capacity [23]. The memoryless birth-following encoder [15], inspired by this, signals the most natural aspect of a birth process. This encoder uses only the difference in the current values of the populations so that $f = f(x_1 - x_2)$. Letting $e = x_1 - x_2$ then $e \xrightarrow{u} e + 1$ and $e \xrightarrow{f=f(e)} e - 1$. Birth-following is then defined as in Eq. (6) with $1(a) = 1$ when a is true and 0 otherwise.

$$f(e) := f_{\max} 1\,(e > 0) \tag{6}$$

Under this encoder our estimation problem conforms to an $M|M|1$ queue with e as the number of customers, and X_1 and X_2 births representing arrivals and departures [15] [24]. The $M|M|1$ forces $\langle f \rangle = \mathbb{E}[u] = u$ so that the queue utilisation, $\rho = \frac{u}{f_{\max}} = \frac{\langle f \rangle}{f_{\max}} < 1$. This equality allows us to simplify the bound of Eq. (4) into Eq. (7).

$$C = -u \log \rho \implies D = (-2 \log \rho)^{-1} \tag{7}$$

We define the performance ratio $\psi := \frac{J}{D}$ with ψ^* obtained when J^* is used. The main theorem follows.

Theorem 1. *The birth-following encoder and the memoryless decoder $\hat{x}_1 = x_2 + \mathbb{E}[e]$ form an asymptotically optimal codec for simple birth processes. It achieves $\psi^* < 1$ in the small ρ regime due to the factor $-\rho \log \rho$.*

We will outline some central elements of the proofs given in [15] [16]. The optimal causal decoder for this problem is $g^* = \mathbb{E}\left[x_1 \,|\, {x_2}_0^t\right]$ and it achieves the mmse, $J^* = \mathbb{E}\left[(x_1 - g^*)^2\right]$. [20]. Expanding and substituting for e we get $J^* = \mathbb{E}\left[\left(e - \mathbb{E}\left[e \,|\, {x_2}_0^t\right]\right)^2\right]$. By Burke's theorem [25] $\mathbb{E}\left[e \,|\, {x_2}_0^t\right] = \mathbb{E}[e]$.

Using the fact that $\mathbb{P}(e=k)=(1-\rho)\rho^k$, $k\in\mathbb{Z}_0^+$ [24], we get Eq. (8).

$$J^*=\mathrm{var}(e)=\rho(1-\rho)^{-2} \tag{8}$$

The by-product of these computations is that the optimal decoder is $g^*=x_2+\mathbb{E}[e]=\hat{x}_1$. The proof that birth-following is an asymptotically optimal encoding is from [16] and shows that as $\rho\to 0$, J^* converges to the true distortion measure of the Poisson channel [26].

Combining Eq. (7) and Eq. (8) gives Eq. (9). The $\lim_{\rho\to 0}\psi=0$ because $\lim_{\rho\to 0}-\rho\log\rho=0$. There exists a $\rho=a$ such that $\psi^*<1$. We can show that $\frac{\partial\psi}{\partial(-\rho)}<0$, for all $\rho\leq a$, so that $\psi^*<1$ over this regime.

$$\psi^*=\frac{2}{(1-\rho)^2}\left[\rho\log\frac{1}{\rho}\right] \tag{9}$$

Numerically we find $a\approx 0.199$. Interestingly both D and J^* become unbounded as $\rho\to 1$, suggesting a limited channel stability. This completes the outline for the main theorem. This line of reasoning will feature in many of the estimation problems solved in this text. We will find that $-\rho\log\rho$ is the key term that allows birth-following to quite generally outperform the bound.

III. Results

A. Minimum Time Encoding

In [16] birth-following is was proven as an asymptotically optimal Poisson channel encoder. As $\rho\to 0$ there is no better way of embedding information about x_1 in x_2. While this suggests that birth-following could be an important signalling scheme, it does not elucidate how it links to the timing problems likely to be faced by biological systems [12]. Here, by examining a class of stochastic optimal control problems, we establish how and why birth-following is a meaningful encoding paradigm, even outside the low ρ region.

We start with some encoder, $f=f(x_1^t,x_2^t)$, that has full knowledge of the history of both target and signalling molecules. Let f be constrained so that $f_{\min}\leq f\leq f_{\max}$ with $\mathbb{E}[f]=\langle f\rangle$. We make the biologically reasonable assumption that f can only change at event times of x_1 or x_2. This makes sense as (i) inter-event decisions would necessitate knowledge of a deterministic clock and (ii) as stated in [27], even with full knowledge we cannot predict exactly when the next jump in either process will occur. This assumption, together with the Markov nature of x_1 and x_2 converts our design problem to a continuous time Markov decision process [28].

This allows us to work in $e=x_1-x_2$. Our decision problem is to define $f(e)$. Since we cannot predict event times (i) $e\in\mathbb{Z}_0^\infty$ and (ii) there is no point in acting when $e=0$ so $f(0)=0$ and $f_{\min}=0$. This means e follows a birth-death process. Markov decision processes admit an estimation-control duality [29]. We can therefore think of our estimation problem as the control of the random walk of e along its states. Our design goal is to force e to 0 as quickly as possible. This is a minimum time problem that is equivalent to controlling the service rate of a queue to minimise congestion [27] [28]. We assume that the service rate is chosen from a discrete set that comprises $[0,\ f_{\max}]$, without loss of generality.

The control law or decision policy f is therefore a mapping from the queue length to the discretised service rate. Framing our encoder design problem in this way allows us to capitalise on the body of stochastic optimal control and queueing theory. We adapt and scale results from [27], which establishes that there exists an optimal policy, f that solves Eq. (10) for $0\leq t\leq T$.

$$f_t^*=\arg\inf\mathbb{E}_{\{f\}}\left[\int_0^T\phi(t,\ f_t,\ e_{t^-})\,\mathrm{d}t+\Phi\left(e_{T^-}\right)\right] \tag{10}$$

Here t^- means infinitesimally before time t, ϕ is a cost on the queue length and Φ is a terminal cost. We use the t subscript to make time dependence explicit. It is known that if $\phi=af+e$, so that it linearly depends on the queue length and the service rate, and $\Phi=0$, then the bang-bang controller of Eq. (11) results [27].

$$f_t^*=f_{\max}1(e_{t^-}>0),\quad \text{for } t\in[0,\ T-a] \tag{11}$$

In our estimation problem, a is vanishingly small (we have no service rate costs) and the mse is a non-decreasing function of e so that minimising a linear cost means we achieve the mmse for a fixed T. As a result Eq. (11) implies birth-following.

We can also consider an average, generalised cost function so that f^* minimises $\lim_{T\to\infty}\frac{1}{T}\mathbb{E}\int_0^T\phi(e_t)+r(f(e_t))\,\mathrm{d}t$ with ϕ as a queue cost and r a service rate charge. As long as $\phi(e)$ is non-decreasing and convex in e with $\phi(0)=0$ and $\sum_{i=0}^\infty\phi(i)\rho^i<\infty$ then the optimal choice of f increases with queue size. This result from [30] and [28] is called a monotone optimal policy. Our mse problem satisfies ϕ and we have $r=0$. The lack of service cost means we can simply choose $f_{\max}$ when $e=1$. Monotone optimality requires that $f_{\max}$ also be chosen for $e>1$. Combining these leads to birth-following once more as the optimal policy. We know from $M|M|1$ theory that this policy also satisfies our mean constraint. Note that none of these results depend on ρ other than requiring a stable queue.

Interestingly, this monotone optimal argument also holds for complex networks of queues [31]. Here if we again apply the zero service rate cost and notate the total queue length cost $\phi=\sum_j\phi_j$ where ϕ_j is the length in the j^{th} queue then we can argue that a network involving birth-following encoders (bang-bang controllers) is optimal [32]. Since this network will consist of purely $M|M|1$ queues then it is a Jackson network [24]. In Jackson networks queues can be treated independently and in isolation. This observation, together with the fact that optimal results for Jackson networks also tend to hold for more arbitrary distributions [33], motivates the investigations of Section III-B and Section III-C.

Thus, birth-following is an optimal bang-bang rate controller. There is a known correspondence between minimum time problems and bang-bang solutions. In [34] the problem of optimally controlling the rate of a point process to maximise the probability of obtaining N events in time T is studied. Using the cost function of Eq. (10), [34] finds that the minimum time solution is in fact bang-bang over $[0,\ f_{\max}]$. When we adapt this result we find that birth-following is the minimum time encoder. Eq. (6) is therefore the quickest way of getting x_2 to achieve some x_1 threshold, at any ratio of u to $f_{\max}$. As a result, birth-following could be an important feedforward modulation scheme for achieving fast and precise event timing in cellular transduction, especially when signals depend on thresholds [12].

B. General Birth-Following

We show that the birth-following solution of theorem 1 is not just a special result for constant rate birth processes. Two generalisations are developed here. First we let the birth rate of X_1 be an arbitrary time varying function, $\lambda(t)$, constrained so that $\frac{u}{\epsilon}\leq\lambda\leq u$, with $\epsilon>1$. X_1 follows an inhomogeneous Poisson process with bounded rate. We let $\rho=\frac{u}{f_{\max}}$ and fix a Poisson channel with constraints as $\langle f\rangle\leq u$ and $\max f\leq f_{\max}$.

Let Y_0 be a hypothetical species with constant birth rate u and population y_0. Then x_1 is a Bernoulli thinning of y_0 by independently accepting each Y_0 event as a new customer in the $X_1 - X_2$ queue with probability $\frac{\lambda(t)}{u}$ [35]. This means that the true $e \leq e_0$ where $e_0 = x_1 - y_0$, and so $J^* \leq J_0^* = \text{var}(e_0) = \rho\,(1-\rho)^{-2}$ (see theorem 1). J^* is the unknown mmse for the $X_1 - X_2$ queue.

We also define a homogeneous Poisson process, Y_1, with rate $\frac{u}{\epsilon}$. Its diffusion approximation is simply a scaled version of Eq. (5). The resulting bound, $D_1 = (-2\epsilon \log \rho)^{-1}$. If the unknown bound for X_1 is D then $D \geq D_1$. This follows as X_1 has both a larger mean and variance, and so would be more difficult to estimate on the same fixed capacity channel. These bound relations also tie into data processing inequalities for thinned Poisson processes [36].

The performance ratio for the $X_1 - X_2$ system is $\psi^* = \frac{J^*}{D} \leq \frac{J_0}{D_1} = \frac{2\epsilon}{(1-\rho)^2}\left[\rho \log \frac{1}{\rho}\right]$. The $-\rho \log \rho$ factor means that $\psi^* < 1$ exists and so birth-following outperforms the bound under finite, arbitrary, time-varying birth rates. We can also obtain these results by taking piecewise constant approximations of $\lambda(t)$ and applying previous $M|M|1$ arguments.

In our second generalisation we allow X_1 to have some generalised birth distribution, with mean rate u. Births must still be independent and identically distributed but we no longer have the Poisson process restriction that forces a mean to variance ratio of 1. Applying birth-following leads to $e = x_1 - x_2$ representing the number of customers in a $G|M|1$ queue. The $G|M|1$ mmse is unknown so we use the mse, $J = \mathbb{E}\left[e^2\right] = \sum_{k=0}^{\infty} k^2 \mathbb{P}(e = k)$. It is known that $\mathbb{P}(e = k) = \rho(1-\sigma)\sigma^{k-1}$ [24], with $0 \leq \sigma < 1$ derived from the chosen generalised arrival distribution. This gives $J = \rho(1+\sigma)(1-\sigma)^{-2}$. At $\sigma = \rho$ we recover the $M|M|1$.

The channel constraints are unchanged so D is from Eq. (7). The gives $\psi = \frac{J}{D} \geq \psi^*$ as shown in Eq. (12). Enumerating over σ characterises the birth-following performance of all possible generalised distributions.

$$\psi = \frac{2(1+\sigma)}{(1-\sigma)^2}\left[\rho \log \frac{1}{\rho}\right] \tag{12}$$

The $-\rho \log \rho$ factor appears and so $\psi^* < 1$ is achievable in some region of low ρ (theorem 1). This results as every possible arrival distribution has $\sigma < 1$. Birth-following is therefore a fundamental encoder for many different birth processes. This even includes the deterministic birth process in which x_1 increments every $\frac{1}{u}$ time units. In this case we get a $D|M|1$ queue [24].

C. Arbitrary Signalling Networks

Having established birth-following as a high performing encoder under arbitrary X_1 synthesis rates, we now generalise the molecular network that is estimating x_1. This allows us to model complex signalling pathways. Information about X_1 is transferred across some molecular cascade involving $n-1$ hidden species, $[X_2, \ldots, X_n]$ in some configuration. We only observe the final output of the network, x_{n+1} which is the population of our signalling species X_{n+1}.

Each X_j for $j \in \mathbb{Z}_2^{n+1}$ observes its input using birth-following. An $M|M|1$ queueing network results. We allow each encoder to have its own constraints so that each queue has a different utilisation. We only consider feedforward networks so that the signalling architecture is composed of splits and joins. The feedforward assumption is not uncommon for intracellular signalling analyses [12]. Burke's theorem states that the output process of any $M|M|1$ queue is Poisson with rate equal to its input process [25]. This implies that x_{n+1} must have rate u, by the principle of local balance [37].

A network involving four intermediate species is given in Fig. 1 to illustrate the types of signalling architectures possible. This network

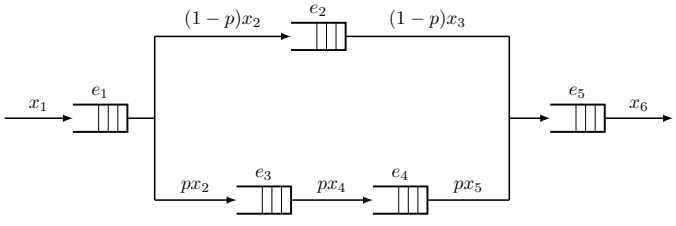

Fig. 1: **Example signalling network.** Information about the target population, x_1, is indirectly transferred to the observed population x_6 by four intermediates. The network splits x_2 (p-thinning) and then joins the branches as an input to the final queue (superposition). Birth-following is applied throughout so each queue is $M|M|1$.

has five queues with a mix of parallel and serial connections. The j^{th} queue has length e_j and utilisation $\rho_j < 1$. The splitting process on the output of the e_1 queue is a random Poisson thinning with probability p. The joining process that serves as the input to the e_5 queue is a Poisson superposition [24]. All molecules conform to homogeneous Poisson descriptions by Burke's theorem [25].

We first derive an appropriate information theoretic bound for an arbitrary network. We can think of our network as a set of effectively serial links with rate u. For example, Fig. 1 features three: e_1, e_5 and an effective link containing the parallel queues. We will always have at least one truly serial queue due to the x_{n+1} species. While the capacity, C_n, of an arbitrary cascade with n queues is unknown for $n > 1$ [38], it is no greater than the that of its most restrictive single, serial Poisson channel [39]. Without loss of generality we assume that the queue with output x_{n+1} has the largest utilisation of the network, ρ. This implies that $C_n < C_1 = -u \log \rho$, so $D_n > D_1 = (-2 \log \rho)^{-1}$ (see Eq. (7)).

The j^{th} queue has stationary distribution $\mathbb{P}(e_j = k_j) = (1-\rho_j)\rho_j^{k_j}$, $k_j \in \mathbb{Z}_0^{\infty}$. We are interested in an effective error between the target and observed species, $e = x_1 - x_{n+1}$. Solving across the network we find that in general $e = \sum_{j=1}^{n} e_j$. By Jackson's theorem [24] the e_j are mutually independent and can be treated in isolation so $\mathbb{P}(e_1 = k_1, \ldots, e_n = k_n) = \prod_{j=1}^{n} \mathbb{P}(e_j = k_j)$. The optimal network decoder is $g^* = \mathbb{E}\left[x_1 \,|\, x_{n+1}{1_0}^t\right]$ [21]. Let the mmse be $J_n^* = \mathbb{E}\left[(x_1 - g^*)^2\right]$. We can expand this to: $\mathbb{E}\left[\left(\sum_{j=1}^{n} e_j - \mathbb{E}\left[\sum_{j=1}^{n} e_j \,|\, x_{n+1}{1_0}^t\right]\right)^2\right]$. Using a corollary of Burke's theorem which states that as e_j is independent of future arrivals in reversed time, it is also independent of past departures in normal time [25], we get that $\mathbb{E}\left[\sum_{j=1}^{n} e_j \,|\, x_{n+1}{1_0}^t\right] = \mathbb{E}\left[\sum_{j=1}^{n} e_j\right]$. This gives $J_n^* = \text{var}\left(\sum_{j=1}^{n} e_j\right) = \sum_{j=1}^{n} \rho_j\,(1-\rho_j)^{-2} < n\rho\,(1-\rho)^{-2}$. This leads to the ψ_n^* inequality of Eq. (13).

$$\psi_n^* < \frac{J_n^*}{D_1} < \frac{2n}{(1-\rho)^2}\left[\rho \log \frac{1}{\rho}\right] \tag{13}$$

Intriguingly, we have recovered a performance ratio that is bounded by a theorem 1 type expression. Hence for any sized network the information theoretic bound will be violated by birth-following codecs at low maximum utilisation ρ. The optimal decoder is $g^* = x_{n+1} + \mathbb{E}\left[\sum_{j=1}^{n} e_j\right]$. This supports the assertion, from Section III-A, that birth-following is a meaningful minimum time encoder across signalling networks.

D. Death Reaction Networks

The pure birth processes we have so far investigated are good models for the accumulation of stable molecules. We now use a standard protein production-degradation model [1], $x_1 \xrightarrow{u} x_1 +$

1, $x_1 \xrightarrow{\mu x_1} x_1 - 1$, to examine how deaths alter the information structure and hence the form of optimal encoders. The chemical master equation of this model suggests that x_1 can be thought of as the number of customers in an $M|M|\infty$ queue with utilisation $\rho_\infty = \frac{u}{\mu} = \mathbb{E}[x_1]$ [24]. We want to estimate x_1 using only our signalling molecules.

We use x_1^+ and x_1^- to indicate target births and deaths and construct the generalised queueing network of Fig. 2. We initially examine the subset of this network in which only x_2 is used. The e_2 queue from Fig. 2 is removed and the e_1 output is replaced with the x_2^+ to x_2^- queue on the right of the main diagram. The x_2^- reaction is $x_2 \xrightarrow{\mu x_2} x_2 - 1$. We use the same rate constant μ because it leads to unbiased estimation.

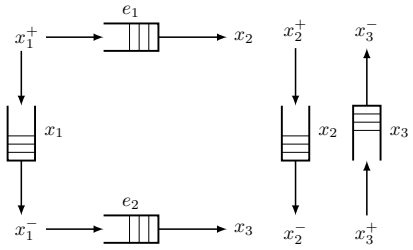

Fig. 2: **Birth-death following network.** Signalling molecules x_2 and x_3 observe the target population births, x_1^+, and deaths, x_1^-, respectively. The x_1 queue is $M|M|\infty$ with number of customers equal to the target population size. The e_1 and e_2 queues are $M|M|1$ by birth-following. If only x_2 is used then a linear death reaction is added with rate constant μ, resulting in an $M|M|\infty$ queue between x_2^+ and x_2^-, with x_2^+ as the output of the e_1 queue. The e_2 queue is then removed. An analogous change occurs if only x_3 is used with the e_1 queue removed. The x_2 and x_3 queues in these cases are shown (disconnected) right.

The signalling molecule death reaction does not alter the bound [40], so that it follows from Eq. (4). Using our queue relations this gives $D = \left(\log\frac{1}{\rho_1} + \frac{1}{\rho_\infty}\right)^{-1}$ with $\rho_1 = \frac{u}{f_{\max}}$ as the utilisation of the $e_1 = x_1^+ - x_2^+$ queue. We do not know the optimum estimator for this problem so instead we define a decoder $g = \tilde{x}_1 = x_2 + \mathbb{E}[x_1 - x_2] = x_2$. This leads to a mse of $J = \text{var}(e)$ with $e = x_1 - x_2$.

To solve for J we must account for the synchronisation between the queue inputs x_1^+ and x_2^+. If the $M|M|1$ connecting them has low ρ_1 then the correlation coefficient, ω will be high and vice versa with $\omega\,|_{\,\rho_1=1} = 0$. Further, ω may also depend on ρ_∞. Expanding the mse and noting that $\text{var}(x_1) = \text{var}(x_2) = \rho_\infty$ gives $J = 2(1-\omega)\text{var}(x_1) = 2(1-\omega)\rho_\infty$. Combining with D we obtain Eq. (14). Simulations suggest that $\max(\omega) \leq 0.5$, which makes sense since the correlation cannot be more than that for a synchronised Flatto-Hahn-Wright queue, which also has $\omega \leq 0.5$ [41] [42]. Using this limit and minimising across ρ_1 gives the inequality in Eq. (14).

$$\psi = 2(1-\omega)\left(1 - \rho_\infty \log \rho_1\right) \geq 1 \tag{14}$$

While we cannot prove that $\psi^* \geq 1$, Eq. (14) insinuates that we have lost the $-\rho_1 \log \rho_1$ structure, and suggests new codes may be needed. If we instead only encode the deaths of x_1 via x_3 so that the e_1 queue is removed and the output of the e_2 queue goes to the x_3 $M|M|\infty$ (right of Fig. 2) then we destroy all correlation. This leads to $\psi = 2\left(1 - \rho_\infty \log \rho_1\right) \geq 2$. This scheme, which leads to a Jackson network [24], can be thought of as death-following. The network results of Section III-C do not hold here due to the $M|M|\infty$ queues.

The above schemes used one encoder and were single channel problems. We now consider the complete network of Fig. 2 which uses two channels by employing birth-following between x_1^+ and x_2 and death-following between x_1^- and x_3. In this case x_2 and x_3 are now pure birth processes. We place the same channel constraints and note that by Burke's theorem on the x_1 $M|M|\infty$, the input process to both the e_1 and e_2 $M|M|1$ queues has rate u. Hence both have utilisation $\rho = \frac{u}{f_{\max}}$. We will call this scheme birth-death following.

As it uses two parallel channels the Poisson capacity doubles [13] [16] leading to $D = \left(2\log\frac{1}{\rho} + \frac{1}{\rho_\infty}\right)^{-1}$. Let $\mathcal{F}_t = [x_{2_0}^t, x_{3_0}^t]$ represent all the observable information. The optimal decoder is then $g^* = \hat{x}_1 = \mathbb{E}[x_1 \,|\, \mathcal{F}_t]$ [20]. No extra information about one $M|M|1$ queue, given its output, is obtained by observing the output of the other $M|M|1$ so $\mathbb{E}[e_i \,|\, \mathcal{F}_t] = \mathbb{E}\left[e_i \,|\, x_{2_0}^t\right] = \mathbb{E}[e_i]$ for $i \in \mathbb{Z}_1^2$. The last equality follows from theorem 1. The mmse is $J^* = \mathbb{E}\left[(x_1 - \mathbb{E}[x_1 \,|\, \mathcal{F}_t])^2\right]$. If $e = x_1 - \hat{x}_1$ then $J^* = \mathbb{E}\left[(e - (\mathbb{E}[e_1 \,|\, \mathcal{F}_t] - \mathbb{E}[e_2 \,|\, \mathcal{F}_t]))^2\right]$. Using previous equalities and the fact that $\mathbb{E}[e_1] = \mathbb{E}[e_2]$ (identical utilisations) gives Eq. (15), with the optimal decoder as the simple population difference $g^* = x_2 - x_3$.

$$J^* = \mathbb{E}\left[e^2\right] = \mathbb{E}\left[(x_1 - (x_2 - x_3))^2\right] \tag{15}$$

We can further expand J^* in terms of the queue variables by using $e_1 - e_2 = x_1 - (x_2 - x_3)$. We obtain $2(1-\omega)\text{var}(e_1) = 2(1-\omega)\rho(1-\rho)^{-2}$ with $\omega = \text{corr}(e_1, e_2)$ and $\text{var}(e_1) = \text{var}(e_2)$. The correlation again derives from the partial synchronisation of the $M|M|1$ inputs. This leads to a Flatto-Hahn-Wright description [41]. The performance ratio can now be written as Eq. (16).

$$\psi^* = \frac{2(1-\omega)}{(1-\rho)^2}\left[2\rho\log\frac{1}{\rho} + \frac{\rho}{\rho_\infty}\right] \tag{16}$$

As synchrony is never perfect then $0 \leq \omega < 1$. Further, $\rho_\infty = \mathbb{E}[x_1] > 0$. For small ρ, $2\rho\log\frac{1}{\rho} \gg \frac{\rho}{\rho_\infty}$ and $\psi^* \approx \frac{4(1-\omega)}{(1-\rho)^2}\left[\rho\log\frac{1}{\rho}\right] \leq \frac{4}{(1-\rho)^2}\left[\rho\log\frac{1}{\rho}\right]$. This generalises the single channel result from theorem 1 and means that birth-death following outperforms the bound.

Comparing our expressions to Eq. (9) we see that birth-death following is analogous to birth-following for a pure birth process. However, the efficiency of the former is reduced due to the death noise, which increases ψ^*. We define efficiency as $\eta = \frac{\psi_b^*}{\psi_{bd}^*}$ with b and d indicating births and deaths. Then $\eta = \frac{1}{2}(1-\omega)^{-1} \geq \frac{1}{2}$. Hence we lose up to 50% of the birth-following performance due to deaths. We cannot get $\eta > 1$ because the Flatto-Hahn-Wright queue condition forces $\omega \leq 0.5$ [42].

Deaths therefore change the information structure of the signalling problem, underscoring the importance of matching the encoder with its channel architecture. In the single channel case birth-following is likely not able to capitalise on the event structure, motivating a need for other coding paradigms [15]. In the two channel case, birth-death following works but at a cost in efficiency.

E. Poisson Sampling at Infinite Capacity

In Section III-B and Section III-C we showed that birth-following quite generally outperforms the bound in a region of small ρ. This corresponds to when the maximum rate of signalling $f_{\max}$ is large. However, birth-following, via its $M|M|1$ description also constrained $\langle f \rangle = u$. We now examine what happens when we let $f_{\max} \to \infty$, so that we are in the region where D is usually bettered, but allow $\langle f \rangle \neq u$. Biologically this models rapid response molecular signalling.

In the limit of these conditions $C \to \infty$ and $D \to 0$ so that $x_1 = x_2 = x$. We will use x_t as short for $x(t)$. Our estimation problem reduces to one of sampling a Poisson population under the mean rate

constraint $\langle f \rangle = \frac{u}{b}$. We let $x_0 = 0$ without loss of generality and define τ as the time to the next sample. Our problem is then to design an optimal sampler or stopping strategy, f, that minimises mse while satisfying $\mathbb{E}[\tau] = \langle f \rangle^{-1} = \frac{b}{u}$.

The sampling time, $\tau < \infty$, is a stopping time with respect to x_0^τ. The compensated Poisson process $M_t = x_t - ut$ is a martingale and therefore obeys the optional stopping theorem [43]. This imposes the condition that all valid sampling schemes must satisfy $\mathbb{E}[M_\tau] = \mathbb{E}[M_0] = \mathbb{E}[x_\tau - u\tau] = 0$, so that $\mathbb{E}[x_\tau] = b$. The last relation comes from our sampling constraint and means that only f functions that sample on average every b births of x_1 can be valid. Moreover, $M_t^2 - ut$ is also a martingale so reapplying the optional sampling theorem gives $\mathbb{E}\left[M_\tau^2\right] = b$. Since $\mathbb{E}[M_\tau] = 0$, then by expanding M_τ we can derive Eq. (17), which is a conservation law. It fixes a linear trade between the variance of our sampling times and sampled population sizes.

$$b = \mathrm{var}(x_\tau) + u^2\,\mathrm{var}(\tau) \tag{17}$$

Our performance index is the mse of reconstruction given by $J = \mathbb{E}_\tau\left[\mathbb{E}\left[(x_\tau - g)^2\right]\right]$, with g as some decoding function. We use the τ subscript on the first expectation to emphasise that we are now averaging across stopped trajectories. This is an ensemble equivalent to the mse expressions used in prior sections. The optimal decoder is $\hat{x} = g^* = \mathbb{E}[x \,|\, \mathcal{F}_\tau]$ with $\mathcal{F}_\tau = [x_0,\, f]$ as the total causal information available, due to the Markov nature of x_1 [20]. The mmse, J^*, results when $g = g^*$.

All valid sampling schemes satisfy Eq. (17). Before examining birth-following, we will look at deterministic stopping protocols for comparison. The most common of these takes a sample at a fixed, constant τ. This is also known as periodic sampling [44]. Let the encoder in this case be f_d with subscript d indicating deterministic. As f_d has no information about x then $\mathcal{F}_\tau = x_0$ and $g^* = \mathbb{E}[x] = ut$. This give $J_d^* = \frac{1}{\tau}\int_0^\tau \mathrm{var}(x_t)\,\mathrm{d}t = \frac{b}{2}$. In this case $b = u\tau$ and $\mathrm{var}(\tau) = 0$. By Eq. (17) this means $\mathrm{var}(x_\tau) = b$. This scheme is on the extreme of maximising the sampled population variance.

Consider a protocol in which we take a sample every b events . We call this b-following. [1] Interestingly, this scheme is on the other extreme of Eq. (17) as it forces $\mathrm{var}(x_\tau) = 0$ so that $\mathrm{var}(\tau) = \frac{b}{u^2}$, is maximised. This means that b-following is an adaptive or event triggered sampler [44]. Adaptive samplers are known to improve upon time triggered schemes, such as the periodic one, by exploiting the information structure of the process to be sampled [45]. At $b = 1$ we recover birth-following as a maximally adaptive sampler.

We calculate the optimal decoder for b-following at $t < \tau$ as $g^* = \mathbb{E}\left[x_t \,|\, x_0^t \leq b-1\right] = \left(\sum_{i=0}^{b-1} i p_i\right)\left(\sum_{i=0}^{b-1} p_i\right)^{-1}$ with $\mathbb{P}(x_t = i) = p_i = \frac{(ut)^i e^{-ut}}{i!}$. We solve to get $\hat{x}_t = ut\left(\sum_{i=1}^{b} \frac{(ut)^i}{i!}\right)\left(\sum_{i=0}^{b-1} \frac{(ut)^i}{i!}\right)^{-1}$, which is a renormalised upper truncated Poisson function. We find that $\hat{x}_t \,|_{\,b=1} = 0 \leq \hat{x}_t \leq ut = \hat{x}_t \,|_{\,b\to\infty}$. Interestingly, these limits recommend a transition in decoding from a zero order hold at $b = 1$ to a linear, uninformative function that matches the periodic sampler decoder as $b \to \infty$. This exemplifies how signalling rate constraints can alter the embedded information structures in discrete problems.

If $t_i = \inf\{t : x_t \geq i\}$ so that $t_b = \tau$ and $t_0 = 0$ then $t_i \sim \mathrm{Erlang}(i, u)$ with density function $\mathbb{P}_{t_i}(t; i, u) = \frac{u^i t^{i-1} e^{-ut}}{(i-1)!}$. We can partition the mmse calculation over t_i so that $J_b^* = \frac{u}{b}\mathbb{E}_{\{t_i\}}[\Omega]$ with expectations taken over $\mathbb{P}_{t_i}(t; i, u)$ and $\Omega = \int_0^{t_b} (b - 1 - \hat{x}_t)^2\,\mathrm{d}t + \sum_{i=1}^{b-1}\left(t_i - 2\int_0^{t_i}(i - \hat{x}_t)\,\mathrm{d}t\right)$. These expressions yield $J_{b\leq 1}^* = 0$. We achieve $D = 0$ at all $b \leq 1$ by simply applying birth-following.

Our interest is in what is achievable at $b \in \mathbb{Z}_2^\infty$, at which the mean encoding rate is smaller than u. For the $b = 2$ case, $\hat{x}_t = ut(1 + ut)^{-1}$ and $\Omega = \int_0^{t_1}\left(\frac{-ut}{1+ut}\right)^2\mathrm{d}t + \int_{t_1}^{t_2}\left(\frac{1}{1+ut}\right)^2\mathrm{d}t = \frac{1}{u}\left(1 + ut_1 - \frac{1}{1+ut_2} - 2\log(1 + ut_1)\right)$. Using $\mathrm{E}_1(y) := \int_y^\infty t^{-1}e^{-t}dt$ gives $J_2^* = \frac{1}{2}(1 - e\mathrm{E}_1(1)) \approx 0.2018$. As J_b^* increases with b, adaptive samplers at $b > 1$ cannot obtain perfect precision, even with infinite capacity. We can use Eq. (8) and Eq. (7) to find that birth-following attains the same mmse as the infinite capacity $b = 2$ case using a channel with $C \approx 0.4616$ nats. Higher capacities do not guarantee improved precision.

F. Removing the Diffusion Approximation

As we have shown, the information theoretic bound, D, does not completely describe the realisable performance limits in biological estimation. This has been largely attributed to the dependence of D on a diffusion approximation of the target molecule [15]. We now examine what happens when this approximation becomes exact, under infinite $f_{\max}$ and hence capacity as in Section III-E. Our estimation problem becomes one of sampling and reconstruction. Let x_t be a Brownian motion with drift that represents our target population at time t, with $x_0 = 0$, so that $\mathrm{d}x_t = u\,\mathrm{d}t + \sqrt{u}\,\mathrm{d}w$ (Eq. (5)) holds. We investigate the design of sampling protocols for x_t under these conditions, at which $D = \frac{u}{2C} \to 0$ [13].

We will focus on optimal reconstruction of x_t as a function of the mean encoding constraint $\langle f \rangle = \frac{u}{b} = \mathbb{E}[\tau]^{-1}$, with τ as the sample time. Using the Markov and ergodic nature of x_t, we can write the usual integral of mmse instead as the ensemble average: $J^* = \frac{u}{b}\mathbb{E}_\tau\left[\int_0^\tau \mathrm{var}\left(x_t \,|\, \mathcal{F}_t\right)\mathrm{d}t\right]$ [46] [47]. Here $\mathcal{F}_t$ represents all the causally observable information about x_t.

We investigate how the mmse changes with our channel and hence sampling constraints. We first examine periodic sampling, which is the deterministic limit of Eq. (17), at which τ is constant so that $\mathrm{var}(\tau) = 0$. Since the encoder is independent of the process in this case then $\mathrm{var}\left(x_t \,|\, \mathcal{F}_t\right) = \mathrm{var}(x_t) = ut$. This leads to $J_d^* = \frac{u^2}{2b}\left(\mathrm{var}(\tau) + \frac{b^2}{u^2}\right) = \frac{b}{2}$. Subscript d indicates deterministic. This result matches that of Section III-E.

Adaptive or event based sampling improves performance by using process information. However, the definition of what constitutes an event is not as clear as when x_t followed a birth-process. It is known that the sampling problem with drift is no different that the one without it [45]. We therefore work with the associated, x_t process, y_t which conforms to $\mathrm{d}y_t = \mathrm{d}w$. We set $u = 1$ here without loss of generality as we can always rescale time t to $\frac{t}{u}$ to recover the $\sqrt{u}$ standard deviation. None of these changes affect the $D = 0$ bound and they allow us to directly use the optimal sampling results from [46] and [44]. These suggest that a sampler which records an event every time y_t crosses a symmetrical threshold at $\pm r$ is mmse optimal. The first exit time of y_t, τ_1, satisfies $\mathbb{E}[\tau_1] = r^2$ [43]. This implies that $r = \sqrt{b}$.

The mmse is given in Eq. (18) and comes from a limit of an expression derived in [46] using optimal stopping theory. We rescale time to get the mmse for x_t, with our sample time $\tau = \frac{\tau_1}{u}$, as $\mathbb{E}_\tau\left[\int_0^\tau u\frac{b}{6}\,\mathrm{d}s\right]\mathbb{E}_\tau[\tau]^{-1} = \frac{b}{6}$. Eq. (18) is hence the mmse for x_t as well.

$$J_b^* = \lim_{l\to 0}\frac{\int_0^r e^{-lz^2}\int_0^z y^2 e^{ly^2}\,\mathrm{d}y\,\mathrm{d}z}{\int_0^r e^{-lz^2}\int_0^z e^{ly^2}\,\mathrm{d}y\,\mathrm{d}z} = \frac{b}{6} \tag{18}$$

[1] Technically we must define our sampling times in b-following as $\tau = \min\{t_b, T\}$ where T is some large but bounded time and t_b is the first time when b events have elapsed. Almost surely boundedness is a requirement of the optional sampling theorem. This does not affect our results so we will not refer to this point in the main text [43].

At finite $\langle f \rangle$, which implies $b > 0$, we cannot realise perfect timing precision. A non-zero mmse suggests an effectively limited information capacity and emphasises the point that channel capacity can be misleading. While the optimal Brownian threshold sampler is not an analogue of birth-following, it is similar to a toggle based encoder developed in [15] which outperforms the bound for birth-death processes like those of Section III-D. This highlights the need to match encoders to the informative structures of biological problems. In this case hard thresholds embed process information.

IV. Discussion

The information-timing-performance link is not well understood for discrete, causal, event based biological mechanisms such as intracellular signalling [13] [19]. Most studies attempting to clarify this link have characterised signalling pathways under the assumption that maximising information implies improved performance, and so have focussed on calculating channel capacities [5] [6] [48]. Little work has been done on examining what coding schemes actually set the achievable performance on a pathway or on how these codes compare to theoretical bounds based on the channel capacity. We fill this niche by taking an in depth look at elementary yet biologically relevant birth process estimation problems with Poisson communication.

Birth processes are not only more amenable to stochastic theory but are also relevant models of protein synthesis, and important components of larger molecular networks [1]. The most basic process that we study, in which a target molecule is produced at a constant rate, is actually a current optimal model for the expression and accumulation of stable proteins such as holin in bacteria [12]. Our more complex birth networks may apply to models of transcription in which rates vary arbitrarily with protein count or to gene promoter autoregulation problems involving intermediate molecules [12].

We work with Poisson channels because they properly describe the discrete structure of information in signalling pathways [13] [19]. Most research into intracellular transduction hinges on Gaussian channel descriptions, which assume continuity. The choice of channel defines the types of optimal coding schemes applicable [49]. For example, normally distributed inputs maximise Gaussian channel performance [11], while random telegraph models achieve capacity on Poisson channels [23]. Moreover, even the parametrisation of codes changes from signal to noise ratio to maximum-mean signal rates [50]. This hints at how important it is to get the encoder-channel relationship right.

It is within this context that, by converting our causal encoding problem to an optimal control problem, we derived birth-following as a minimum time bang-bang encoder. We now have a firm theoretical basis for why this scheme, previously known only as an asymptotically optimal heuristic [16], is important. Minimum time encoders are likely crucial in achieving event precision, since they achieve a desired signalling threshold at the fastest rate possible. Birth-following is also biologically significant since many proteins behave like molecular switches or relays, activating on an event, completing a signalling objective and then deactiving [51]. Both the on and off switching are integral to performance, especially when multiple signals must be made [51]. Birth-following has the fastest switching dynamics achievable under a birth-process description.

We used birth-following to examine mismatches between realisable optimal estimators and slack information theoretic bounds that delimit non-linear coding precision [13]. Mismatches indicate that continuity approximations are insufficient and emphasise the signalling rate regions in which the discreteness of information dominates. We found that if the maximum signalling rate was high enough, birth-following would violate the bound under generalised target synthesis rates and arbitrary signalling network architectures.

The factor $-\rho \log \rho$ was found to be crucial to achieving these mismatches. Closer examination reveals an interesting information theoretic link. The stationary e distribution of the $M|M|1$ is geometric with $\mathbb{P}(e = k)$ describing the probability of k failures under failure parameter ρ [24]. The entropy of this distribution is $\mathcal{H}(\rho) = \frac{1}{1-\rho}\left(-\rho \log \rho - (1-\rho)\log(1-\rho)\right)$. Using a known bound on the binary entropy function [11] and looking at the low $\rho \ll 1$ regime we get Eq. (19). This tells us that estimation precision relative to the bound decays with the square root of ρ. Note that ρ is also the mean queue length, which can be thought of as the average signalling threshold [12]. This relation therefore connects queueing, information and signalling.

$$\mathcal{H}(\rho) \,|_{\,\rho \ll 1} \approx -\rho \log \rho \leq \sqrt{2\rho} \tag{19}$$

Additionally, by applying birth-following we shift our estimation problem into a scheme that depends on a single, dimensionless quantity, ρ. This makes our results scalable to any biological settings satisfying this ratio.

We also examined death reactions to gauge how encoding changes with the additional noise of molecular degradation. We found that previous performance indices could only be reproduced if an additional signalling channel that followed target molecule deaths was introduced. Single channel estimation may therefore require different codes, such as the toggle based one used in [15] for a related control problem. Further, even with the two channel description, recoverable performance suffered up to a halving in efficiency. This correlates well with the observation that deaths often contribute half of the total molecular noise [52], and may suggest that we are approaching the effective noise limits in molecular networks with these approaches.

In all these problems discreteness appeared to dominate when the maximum signalling rate and hence Poisson capacity was high. We let this rate become unbounded to analyse what is achievable when we can transmit infinite information. Under these conditions our estimation problem becomes one of sampling and birth-following reduces to a natural adaptive sampling protocol [44]. We showed that birth-following, in contrast to deterministic periodic sampling, is maximally event triggered and offers sharp molecular coupling. However, it is still unable to achieve perfect timing unless the mean signalling rate is at least as fast as the synthesis rate of the target molecule.

Mismatches with the bound were based on its diffusion description of the target molecule. To cement our conclusions we investigated this sampling problem with an actual diffusion serving as the target. Even in this case a known optimal threshold scheme [45] was unable to get close to the bound (which is 0 here) unless the mean rate also becomes infinite. The consequence of all these sampling results is that the usual assumption, that higher capacity implies better performance [8], must be discarded. It is important to consider the performance that codes can actually achieve. The difference between knowing the capacity and having an implementable code that can realise it is a well known problem in information theory [11].

By concentrating on amenable yet still biologically interesting birth process problems, we have attempted to uncover properties of the information-signalling-performance link for intracellular models involving small molecular populations. We find that when signalling is fast and discreteness prominent, current capacity based bounds do not present the full story on achievable performance. Moreover, realistic performance depends on matching encoders to channels, which unfortunately recommends specialised approaches for different signalling models. For example, birth-following, which encodes the most salient aspect of birth processes and applies to problems

involving molecular switches or timekeeper proteins [12], is unlikely to be optimal when molecular turnover is quick. This encapsulates the idea that noise management motifs are shaped by informational structure [3].

In spite of this our results offer some general insight. Our solutions are based on conditional expectations which minimise the mse. The mse was used because it directly related to known bounds. However, conditional expectations also minimise any Bregman loss function [53] (a broad class that includes Kullback-Liebler divergence) thus making our codes widely applicable. More mechanistically, our work synergises queueing, information, optimal control and stochastic estimation theory, each of which may serve as a point of departure for more searching mathematical analyses. In particular, [54] developed a timing channel which is a queueing equivalent to the Poisson one [38]. This channel more directly emphasises the importance of timing and may present an interesting direction for future research. Biologically, our results also apply to any system characterised by birth processes or requiring single event precision. Approaches developed here may be useful for defining the observational limits of fluorescent proteins, [55] , the physical limits of chemotaxis [56] or even the coding schemes in sensory transduction [19].

Thus the limits imposed by molecular fluctuations and Poisson capacities remain undefined, and misleading, for generalised and sharp non-linear signalling, where discreteness and causality matter. As noted in [10], there is still a need for more inclusive and operational definitions of 'information' and 'uncertainty'. Our work aligns with this drive, suggesting that better metrics may be required for describing the gap between the theoretically possible performance expected from Shannon's channel coding theorem [11] and the effectively achievable precision from implementable codes. One Bregman metric, known as the Poisson loss function [57], is the most natural measure of Poisson channel precision and has the cleanest link to causal estimation and control performance [58]. We recommend this as a possible point for upcoming studies.

Acknowledgements

This work was funded by the Gates Cambridge Trust and the Control Group in the Department of Engineering, University of Cambridge. The author would like to acknowledge Glenn Vinnicombe for his useful advice and comments.

References

[1] D. Wilkinson, Stochastic Modelling for Quantitative Description of Heterogeneous Biological Systems, Nature Reviews: Genetics 10 (128).
[2] A. Rhee, R. Cheong, A. Levchenko, The Application of Information Theory to Biochemical Signaling Systems, Phys. Biol. 9 (2012) 045011.
[3] A. Eldar, M. Elowitz, Functional Roles for Noise in Genetic Circuits, Nature 467.
[4] K. Janes, C. Reinhardt, M. Yaffe, Cytokine-Induced Signaling Networks Prioritize Dynamic Range over Signal Strength, Cell 135 (2008) 343–54.
[5] R. Cheong, A. Rhee, C. Wang, et al., Information Transduction Capacity of Noisy Biochemical Signaling Networks, Science 334 (6054) (2011) 354–58.
[6] C. Waltermann, E. Klipp, Information Theory Based Approaches to Cellular Signaling, Biochimica et Biophsica Acta 1810 (2011) 924–32.
[7] J. Selimkhanov, B. Taylor, J. Yao, et al., Accurate Information Transmission through Dynamic Biochemical Signaling Networks, Systems Biology 346 (6215).
[8] M. Brennan, R. Cheong, A. Levchenko, How Information Theory Handles Cell Signaling and Uncertainty, Science 338 (6105) (2012) 334–335.
[9] S. Uda, S. Kuroda, Analysis of Cellular Signal Transduction from an Information Theoretic Approach, Seminars in Cell and Developmental Biology 51 (2016) 24–31.
[10] Z. Mousavian, J. Diaz, A. Masoudi-Najad, Information Theory in Systems Biology. Part II: Protein–Protein Interaction and Signaling Networks, Seminars in Cell and Developmental Biology (2016) 14–23.
[11] T. Cover, J. Thomas, Elements of Information Theory Second Edition, John Wiley and Sons, 2006.
[12] K. Ghusinga, J. Dennehy, A. Singh, First-passage Time Approach to Controlling Noise in the Timing of Intracellular Events, Proceedings of the National Academy of Science 114 (4) (2017) 693–8.
[13] I. Lestas, G. Vinnicombe, J. Paulsson, Fundamental Limits on the Supression of Molecular Fluctuations, Nature 467 (2010) 174–8.
[14] Y. Kabanov, The Capacity of a Channel of the Poisson Type, Theory of Probability and its Applications 26 (1978) 143–147.
[15] K. Parag, G. Vinnicombe, Single Event Molecular Signalling for Estimation and Control, European Control Conference (2013) 4166–71.
[16] K. Parag, G. Vinnicombe, Event Triggered Signalling Codecs for Molecular Estimation, 52nd IEEE Conference on Decision and Control (2013) 256–61.
[17] M. Frey, Capacity of the Lp Norm Constrained Poisson Channel, IEEE Transactions on Information Theory 38 (2) (1992) 445–50.
[18] N. Perrimon, C. Pitsouli, B. Shilo, Signaling Mechanisms Controlling Cell Fate and Embryonic Patterning, Cold Spring Harb Perspect Biol 4 (2012) a005975.
[19] K. Parag, G. Vinnicombe, Point Process Analysis of Noise in Early Invertebrate Vision, PLoS Computational Biology 13 (10) (2017) e1005687.
[20] D. Snyder, Filtering and Detection for Doubly Stochastic Poisson Processes, IEEE Transactions on Information Theory 18 (1972) 91–102.
[21] D. Snyder, M. Miller, Random Point Procresses in Time and Space, 2nd Edition, Springer-Verlag, 1991.
[22] H. Taylor, S. Karlin, An Introduction to Stochastic Modelling, 3rd Edition, Academic Press, 1998.
[23] M. Davis, Capacity and Cutoff Rates for Poisson Type Channels, IEEE Transactions on Information Theory 26 (1980) 710–715.
[24] L. Kleinrock, Queueing Systems Volume I Theory, John Wiley and Sons, 1975.
[25] P. Burke, The Output of a Queueing System, Journal of Operational Research Society 4 (1956) 699–704.
[26] T. Coleman, N. Kiyavash, V. Subramanian, The Rate-Distortion Function of a Poisson Process with a Queueing Distortion Measure, IEEE Transactions on Information Theory.
[27] R. Boel, P. Varaiya, Optimal Control of Jump Processes, SIAM Journal of Control and Optimization 15 (1) (1977) 92–119.
[28] R. Serfozo, Optimal Control of Random Walks, Birth and Death Processes, and Queues, Advances in Applied Probability 13 (1981) 61–83.
[29] E. Todorov, General Duality between Optimal Control and Estimation, in: Proceedings of the 47th IEEE Conference on Decision and Control, 2008.
[30] T. Crabill, Optimal Control of a Service Facility with Variable Exponential Service Times and Constant Arrival Rate, Management Science 18 (9) (1972) 560–6.
[31] R. Weber, S. Stidham, Optimal Control of Service Rates in Networks of Queues, Advances in Applied Probability 19 (1987) 202–18.
[32] Z. Rosberg, P. Varaiya, J. Walrand, Optimal Control of Service in Tandem Queues, IEEE Transactions on Automatic Control 27 (3) (1982) 600–10.
[33] S. Stidham, Analysis, Design and Control of Queueing Systems, Operations Research 50 (1) (2002) 197–216.
[34] P. Bremaud, Bang-bang Controls of Point Processes, Advances in Applied Probability 8 (1976) 385–94.
[35] P. Lewis, G. Shedler, Simulation of Non-homogeneous Poisson Processes by Thinning, Naval Res. Logistics Quart. 26 (3) (1989) 403–13.
[36] L. Wang, A Strong Data Processing Inequality for Thinning Poisson Processes and Some Applications, in: 2017 IEEE International Symposium on Information Theory (ISIT), 2017, pp. 3180–4.
[37] K. Chandy, The Analysis and Solutions for General Queueing Networks, in: Proceedings of Sixth Annual Princeton Conference on Information Sciences and Systems, Princeton University, NJ, USA, 1972, pp. 224–228.
[38] R. Sundaresan, S. Verdu, Capacity of Queues via Point Process Channels, IEEE Transactions on Information Theory 52 (6) (2006) 2697–709.
[39] R. Sundaresan, S. Verdu, Robust Decoding of Timing Channels, IEEE Transactions on Information Theory 46 (2) (2000) 405–19.
[40] S. Loukas, C. Kemp, The Index of Dispersion Test for the Bivariate Poisson Distribution, Biometrics 42 (4) (1986) 941–8.
[41] L. Flatto, S. Hahn, Two Parallel Queues Created by Arrivals with Two Demands, SIAM Journal of Applied Mathematics 44.
[42] D. Pinotsi, M. Zazanis, Synchonized Queues with Deterministic Arrivals, Operations Research Letters 33.
[43] R. Serfozo, Basics of Applied Stochastic Processes, Springer Science and Business Media, 2009.

[44] K. Astrom, B. Bernhardsson, Comparison of Periodic and Event Based Sampling for First Order Stochastic Systems, in: Proceedings of the 14th IFAC World Congress, 1999.
[45] M. Rabi, G. Moustakides, J. Baras, Adaptive Sampling for Linear State Estimation, SIAM Journal of Control and Optimization 50 (2) (2012) 672–702.
[46] M. Rabi, Packet Based Inference and Control, Ph.D. thesis, University of Maryland (2006).
[47] K. Parag, G. Vinnicombe, Constrained Adaptive Sampling and Causal Estimation of Stochastic Processes, Tech. rep., University of Cambridge (2015).
[48] I. Mian, C. Rose, Communication Theory and Multicellular Biology, Integrative Biology 3 (2011) 350–67.
[49] S. Ghanem, A. Mannujahan, The Poisson Optical Communication Channels: Capacity and Power Allocation, IAENG International Journal of Computer Science 39 (1).
[50] A. Wyner, Capacity and Error Exponent for the Direct Detection Photon Channel-Part I, IEEE Transactions on Information Theory 34 (6) (1988) 1449–61.
[51] B. Alberts, A. Johnson, J. Lewis, et al., Molecular Biology of the Cell, sixth Edition, Garland Science, 2015.
[52] M. Komorowski, J. Miekisz, M. Stumpf, Decomposing Noise in Biochemical Signalling Systems Highlights the Role of Protein Degradation, Biophysical Journal (2013) 1783–1793.
[53] A. Banerjee, X. Guo, H. Wang, On the Optimality of Conditional Expectation as a Bregman Predictor, IEEE Transactions on Information Theory 51 (7) (2005) 2664–69.
[54] V. Anantharam, S. Verdu, Bits through Queues, IEEE Transactions on Information Theory 42 (1) (1996) 4–18.
[55] D. Thorsley, E. Klavins, Estimation and Discrimination of Stochastic Biochemical Circuits from Time Lapse Microscopy Data, PLoS One 7 (11) (2012) e47151.
[56] W. Bialek, S. Setayeshgar, Physical Limits to Biochemical Signalling, Proceedings of the National Academy of Science 102 (29) (2005) 10040–10045.
[57] R. Atar, T. Weismman, Mutual Information, Relative Entropy, and Estimation in the Poisson Channel, IEEE Transactions on Information Theory 58 (3) (2012) 1302–18.
[58] D. Guo, S. Shamai, S. Verdu, Mutual Information and Conditional Mean Estimation in Poisson Channels, IEEE Transactions on Information Theory 54 (5) (2008) 1837–49.