Time as It Could Be Measured in Artificial Living Systems

Andrei D. Robu[1], Christoph Salge[1,2], Chrystopher L. Nehaniv[1] and Daniel Polani[1]

[1]Adaptive Systems Research Group, School of Computer Science
University of Hertfordshire, Hatfield, UK
[2]Game Innovation Lab, Department of Computer Science and Engineering
New York University, New York, USA
a.robu@herts.ac.uk

Abstract

Being able to measure time, whether directly or indirectly, is a significant advantage for an organism. It permits it to predict regular events, and prepare for them on time. Thus, clocks are ubiquitous in biology. In the present paper, we consider the most minimal abstract pure clocks and investigate their characteristics with respect to their ability to measure time. Amongst other, we find fundamentally diametral clock characteristics, such as oscillatory behaviour for local time measurement or decay-based clocks measuring time periods in scales global to the problem. We include also cascades of independent clocks ("clock bags") and composite clocks with controlled dependency; the latter show various regimes of markedly different dynamics.

Introduction

"Bacteria count on probabilistic fingers", to put it crisply, and count they must when they measure time. In the present paper, we study how such counting/time measurement could possibly look like in the most minimal cases which we believe relevant to simple organisms. Klyubin et al. (2007) look at maximising information flow for a navigation task. They find that it is possible to obtain information about the current time. While there time measurement was a side effect, here we make it our primary objective: measure time as the most intrinsic, least environmentally affected processual quantity, with only three assumptions: it is discrete, the tick is global (accessible to the agent) and time has a well-defined beginning. More precisely, we are interested in how much a Markovian agent with limited memory can keep track of the flow of time under these assumptions. It turns out that, in this context, measuring time is effectively counting. To make the best out of limited resources, we need to count probabilistically.

In its conceptually simplest incarnation, the measurement of time would consist essentially of two components: Having a reliable generator of periodic behaviour on the one hand; and being able to count the periods, on the other. To measure larger time intervals precisely, one needs full-fledged counters, which, in turn, require a comparatively complex logical make-up; something which, while not impossible in principle, one would not expect generically in biologically relevant

scenarios and most certainly not in very simple organisms. Rather, one would typically expect to find some less precise, but simpler and more robust solution. The present paper investigates how the most minimal of such models could look and which characteristic properties we expect them to exhibit. In our discussion we include both explicit clocks characterized by a distinct apparatus with a clear function such as the suprachiasmatic nucleus in the mammalian hypothalamus involved in the control of circadian rhythms (Klein and Moore, 1991), but also implicit clocks (which exhibit some aspects of clock-like behaviour without a dedicated mechanism) like the feeding-hunger cycle of an animal. We do not differentiate these classes a priori since our formalism does not discriminate between them, but our study will concentrate on systems which, given their constraints, are maximally able to measure time, not considering any other tasks — thus we study clocks as "pure" as they can be under the circumstances given. We expressly are not studying how an agent could infer time from correlations in the environment (using a sundial for example), but only in how time can be tracked intrinsically. In the same vein as considering Artificial Life as about understanding life-as-we-know-it vs. life-as-it-could-be (Langton, 1989), and in view of how life pervasively makes use of clocks, the present paper studies possible clocks themselves, or, as we say in the title: "time-as-it-could-be-measured".

Temporal Dynamics: The Algebra of Time

We begin with some general comments on the structure of time. Whether in classical or relativistic physics, or in more general models (e.g., ancient Indian notions of cyclical time, or in modern automata networks), there is a commonality to notions and models of time: From the perspective of a single organism or agent, events in time satisfy a grammatical constraint (Rhodes, 2009): If α, β and γ are each sequences of events, then: if β follows γ in an agent's experience, and, prior to both, α occurs, that is exactly equivalent to when β follows α, and γ occurs after both, i.e. $\alpha(\beta\gamma) = (\alpha\beta)\gamma$.

In short, sequences of events in time from the perspective of an individual agent satisfy the *associative law*. Thus, the study of (possibly general) structures satisfying this law becomes the study of models of time. This viewpoint has

deep connections with the theory of discrete or continuous dynamical systems (Rhodes, 2009). In the finite discrete deterministic case it leads to the Krohn-Rhodes theory, a branch of mathematics (algebraic automata theory) where discrete dynamical systems can be decomposed using (non-unique) iterative coarse-graining into a cascade of irreducible components. The composite clocks discussed later constitute a special case of such a decomposition.

The simplest models of time have a single discrete kind of event, a *clock tick*, which drives their dynamics deterministically. The dynamics can be classified into four types (Nehaniv, 1993): *cycles* (after a number n of steps everything repeats), *fuses* (after a number of steps k nothing ever changes again), fuses that end in cycles (fuses that after k transient clock ticks end in a cycle, generalizing the former two cases), or infinite systems with time indexed by the natural numbers (where each moment is different from the others). These types correspond to four different kinds of algebraic structures called *semigroups* with a single generating event. If more than one type of event is allowed (e.g., clock ticks of different types, or sensory input with more than 1-bit), or if the state transitions are not deterministic, much more complicated and interesting dynamics can arise. Here we seek to understand the structure of single tick time "as it could be" for agents experiencing it in a probabilistic or noisy setting, using the language of information theory for selected basic cases. Thus the rest of the paper expands beyond deterministic models into probabilistic ones.

The Cost of Measuring Time

Recent work by Barato and Seifert (2016) highlighted interest in the problem of time measurement by asking if clocks must pay a thermodynamic cost to run. Their stance is grounded in fundamental trade-offs of physics. Our perspective at the level of organisms is far remote from these trade-offs, and physical intuitions and conservation laws and constraints do not apply in a straightforward manner. Instead, we work in a near-macroscopic, classical (non-quantum) Markovian universe without presuming the additional structure of microphysics (microreversibility or Hamiltonian dynamics). In particular, we cannot assume an obvious generalization of the physical concept of energy to a fully general Markovian system. Thus, it is not obvious how to cost computation in terms completely analogous to thermodynamics; the only concepts that carry over are of entropic/information-theoretic nature. Shannon information has been shown to be a highly generalizable measure of information processing cost (Polani, 2009); a universal measure that can even be used to compare systems of a different nature, it does not presuppose any structure on the state space of events. All costs will, therefore, be expressed in this language, specifically information storage and communication costs.

Small State Spaces. We begin by focusing on minimal clocks, for a number of reasons. For one, this will make it easier to explore the full solution space. The second reason is significantly more subtle, and we will only be able to sketch it here: essentially, it is not clear what measure of complexity to utilize for the cost of running a larger single-component counter and/or the complexity of running the transition itself; natural candidates for such costs might be predictive information (Bialek et al., 2001) or statistical complexity (Crutchfield and Young, 1989). However, the possible candidates are not limited to these two measures, and many other plausible information-theoretic alternatives can be conceived[1]. In absence of a canonical measure for the complexity of a single-component clock, here we limit ourselves to investigate the most minimal clocks possible, namely a 2-state (1-bit) clock and we will here not further concern ourselves to take into account the informational cost of actually running this clock.

Information flow between modules. We said above that we prefer minimal clocks by default, to avoid dealing with the complexity of the clock operation. But what if an agent needs a larger clock? In this case, we build it out of smaller clocks. Such a compound clock will perform better if its components "cooperate". For them to be able to cooperate, they must be able to exchange information. Therefore limitations in this clock's internal information flow will reduce the performance of the clock. Yet this information flow will in general be costly (Laughlin, 2001). This will limit how many components can communicate with each other and at which bandwidth. All in all, in our studies, when looking for candidates for clocks, we prefer Markov chains with a small number of states and when we move on to larger clocks we prefer to build them out of small clocks and use the information flow between the components as cost.

Other Relevant Work

Klyubin et al. (2007) consider the maximization of information flows in a very simple agent/environment system. The resulting agent controllers generate a rich set of behaviours. One of the side effects of the controllers' dynamics is that their internal states partially encode location information, but also partially time information. In other words, the resulting agent controllers provide partial information about the point in time in the experiment. Amongst other, the paper studied how space and time are encoded together and how and to which extent they can be separated ("factorized") in informational terms.

Note that in (Klyubin et al., 2007), the measurement of time was a side effect of the overall information flow optimization. The importance of measuring specifically time (as opposed to having this measurement emerge on the side) appears in other scenarios, for instance in 13- or 17-year cicadas (Karban et al., 2000; Sota et al., 2013). On the level of physical limits, Chen and Luo (2010) study the statements that can be obtained by applying Fisher information to the

---

[1]We will investigate this question in a separate paper.

problem of measuring time with quantum clocks and discuss the problem of clock synchronization in the quantum realm.

Karmarkar and Buonomano (2007) note mounting evidence against the view that the brain keeps track of time by counting ticks and instead study the behaviour of simulated SDNs (state-dependent neural networks), showing how neural networks can keep track of time implicitly in their states without counting. Van Wassenhove (2012) discusses not only the possible mechanisms of time perception in the brain but also the difficulty of validating them.

Since we hypothesize that our abstract considerations find functional correspondences in nature, the simplest examples would be expected to be found in microorganisms. We would predict that, if our general hypothesis is correct, the characteristics of our results for minimal clocks will be reflected in very simple organisms. We will preempt here one result detailed in the results section. Under constraints on the memory available to the clock, only two types of clocks are found — local, short-term clocks measuring the time within a cycle (essentially its phase); and, long-term clocks which distinguish large-scale phases within the overall time interval of interest. Thus, we get a dichotomy between local time measurement and global time measurement. We will call the first type cyclic clocks or oscillators, and the second type "drop clocks" (essentially one-off decay-type time measurements). Bacteria show examples of both cyclic and drop clocks.

Hut and Beersma (2011) discuss the reasons why day and night require different behaviours, these forming the selective pressures of evolution for the circadian rhythm of bacteria — the periodic cycle of activity in organisms allows them to adapt to the time of day. This is one example demonstrating the importance of time-keeping to organisms. They note that the circadian rhythm of cyanobacteria has been reproduced and studied in a test tube. Therefore this is an example of a relatively simple, clearly understood explicit cyclic or *alternator* (see below) clock.

Nutsch et al. (2003) study prokaryotic taxis for halobacteria, more specifically signal transduction pathway starting from sensing light and responsible for controlling the switching of the flagellum. This pathway implements a one-off drop clock[2]. They note that while the molecules involved in this pathway are well studied, the way these components behave together dynamically is not well understood and only speculated on. Building upon an earlier model by Marwan and Oesterhelt (1987), they suggest dynamical models to fit experimental findings. These models are synthetic and not based on first principles. Thus, the question about the actual structure and size of the fundamental clocks of bacteria remains currently unanswered.

Measuring time can also help bacteria in spatial tasks. Some bacteria use differences along the length of their body

---

[2]Which is reset only by an external trigger; some mechanisms can be considered drop clocks triggered at conception and never reset until death, such as telomeres.

to measure gradients (Oliveira et al., 2016), however, others instead measure time differences between intensities while moving (Nutsch et al., 2003). This is demonstrating how measurement of space can be converted into measurement of time.

Models of Clocks

[...] time is always encoded in physical systems; [...] any evolving system with nontrivial dynamics can be regarded as a clock. To read the time, we must perform measurements [...]. How precisely we can estimate [time] characterizes the quality of a clock. — Chen and Luo (2010)

Because our clocks are Markov chains, their memory is explicitly encoded in their current state and they are probabilistic. We consider only discrete state space and also advance in time only in discrete steps; importantly, this means that they receive time ticks for free and that they need not be concerned with the challenge of getting accurate ticks (see also discussion). Apart from these global ticks, the clocks do not receive any information from the environment. We furthermore consider first only the most minimal clock designs. We will later consider more complex (composite) clocks.

First, introduce some notation. Random variables are written as capital, their values as lowercase letters. The probability of a random variable X to adopt value x will be written as $P(X = x)$ or, where not ambiguous, as $p(x)$ by abuse of notation. The state of a clock will be denoted by random variables S (possibly subscripted by time t, because the probability distribution over the states of the clock changes in time) which can take on values u and d ("up" and "down"). To model the uncertainty that the agent has about the current time, we treat true time (which the clocks attempt to measure), similarly to Klyubin et al. (2007) also as random variable T which a priori assumes all possible (integer) time values between $t = 0$ and $t = T_{\max}$ with equal probability.

Typical quantities would now be, e.g. $P(S = u|T = t) \equiv p(u|t)$, the probability that the clock state is u ("up") at time t. Or else, $P(T = t|S = d) \equiv p(t|d)$ would be the probability that the time is t, given that the current state of S is d, etc. When we optimize clocks, we quantify our criterium of "performance" as mutual information $I(S; T)$. This mutual information tells how much information a previously uninformed agent would receive about time after looking at the clock (averaged for the possible observations of the clock, "up" or "down"). This quantity, $I(S; T)$ is directly related to the probability of guessing time correctly given one observation of the clock. With this notation, we are ready to define the clocks. In all our experiments, the clocks will be initialized at time $t = 0$ in a fixed known state, specifically u, i.e. $P(S_0 = u) \equiv P(S = u|T = 0) = 1$ and advance once in time per tick according to their respective dynamics.

Alternator

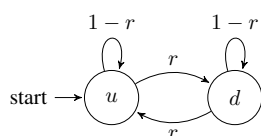We first define the *alternator* as the 2-state Markov chain with a symmetric probability to change state.

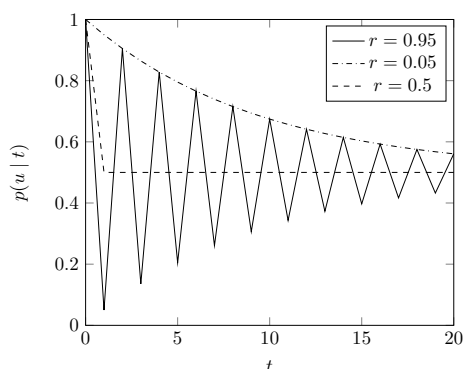Figure 1: The alternator clock. It switches one state to the other with probability r.

Figure 2: The alternator clock. Behaviour classes explained in text.

The plot of $p(u \mid t)$ in Figure 2, shows the distinction between different regimes in which the clock operates, depending on r. We notice the analogy with Damped Harmonic Oscillation: *Stuck*, for $r = 0$, the clock is stuck in state u (not shown in diagram). *Overdamped*, for $0 < r < 0.5$, the probability distribution behaves like an overdamped oscillator. It is interesting to note that, in this regime, the probability distribution of the alternator clock has the same envelope as the drop clock which we discuss below. *Critically Damped*, for $r = 0.5$, the clock acts analogous to a critically damped oscillator: it reaches the equilibrium in the shortest possible amount of time — one time step. *Underdamped*, for $0.5 < r < 1$, the probability distribution behaves like an underdamped oscillator. This can be interpreted as the clock being initially "synchronized" with time like a clock, but that the synchronization gets lost as the clock evolves, until the correlation between t and s disappears. *Undamped*, for $r = 1$, the clock state alternates between u and d. Not shown in figure. These behaviours are taken from the plot of the symmetric alternator, but the asymmetric one (with different transition probabilities $u \rightarrow d$ and $u \leftarrow d$) has the same qualitative behaviour, just with a different equilibrium point. All 2-state Markov chains belong to one these 5 classes. We will refer to this insight in the results section again.

Drop Clock

Consider the thought experiment of an insect that leaves its nest to forage or explore. Even if it does not find anything, the insect should still return at one point, otherwise it risks getting lost. This insect would profit from a clock telling it

if "it's been awhile". An alternator is not well suited for this task; it measures a local phase (an odd or even step), but it does not provide much "larger timescale" information, unless it is overdamped. Instead of the latter, it turns out that a *drop clock* is a more natural model for this task.

More complicated models will exist in nature, but our question is what the simplest ones are. More complex clocks, such as clocks with time-dependent transition laws require more memory under the Markovian. Another example is that of gene regulatory networks which extends beyond discrete time into continuous time.
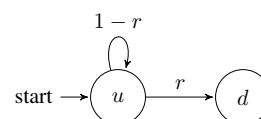

Figure 3: The drop clock, a Markov chain with a probability r to permanently transition to state d and remain there.

The drop clock starts (as all our clocks) in a well defined state, namely u, i.e. $P(S = u | T = 0) = 1$. After each time step, there is a probability r that the clock "decays" (transitions from state u to state d) and a probability $1 - r$ that it does not. Once the clock has decayed, it remains in state d forever. The behaviour that this rule generates is an exponential decay $(1 - r)^t$ in time. An agent infers probable time only from the state of this clock.

Experiments with 1-bit Clocks

Our first experiments are dedicated to the 1-bit (i.e. 2-state) clocks. As discussed earlier, there are only few distinct classes of such clocks. Most notable are the 2-state alternator and the drop clock. The alternator offers the maximum of 1 bit of information about time, namely whether one is in an odd or even time step. However, this information is purely local and cannot distinguish whether one is in an early or a later section of a run. However, even with only 1 bit of state, the drop clock can provide this distinction, albeit imperfectly at less than 1 bit resolution. For this to provide best results, the probabilistic decay rate of the drop clock (unlike the alternator) must be attuned to the length of the total time interval of interest; this rate will be in general acquired by evolution or some learning mechanism — here, we will directly obtain it by optimization of informational costs. This is studied next. Our axis of time is almost featureless except for two features: length of time and the grain. The alternator matches the grain (local information) and the drop clock matches the length (global information). We do not impose any other features on the axis of time such as months or seasons because we only study pure time. For a meaningful link to external events, that would extend to the work of Klyubin et al. (2007).

Measuring Large Time Scales

We now investigate time measurement by drop clocks at different timescales. We tune the clock by finding the drop

probability r maximizing $I(S;T)$ for the particular timescale of the experiment. Optimizing this by grid search gives Fig. 4.
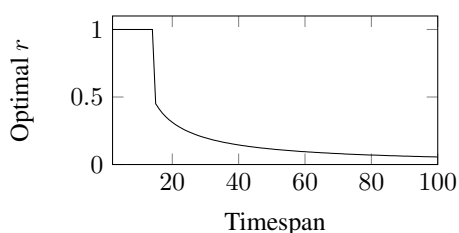

Figure 4: The optimal drop probability r for different timescales. Note that the curve for the drop clock has a discontinuity at $T \approx 15$.

The first result is that the decay rate r that best resolves time with a drop clock clearly depends on the time interval. Such a clock, therefore, must be adapted to the particular time interval to resolve. Strikingly, we find two regimes of solutions, namely one with one fixed decay rate (up to $T \approx 15$), and then a time interval-dependent decay rate. An inspection of Fig. 5 shows a relatively complex landscape where a global maximum of time information at the maximal decay rate $r = 1$ is superseded at larger times by maxima at lower decay rates.
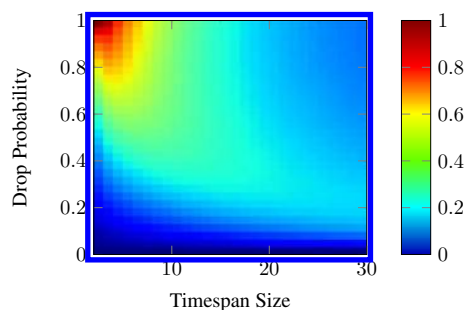

Figure 5: Time information for different drop probabilities and time-spans.

Figure 6 shows with a solid line a vertical slice from Figure 5 at $T = 20$. This information curve does not have a unique maximum, but an inflection. We provide an explanation for the two local maxima by partitioning time into different features: one-way to look at time is to see an "earlier" and a "later" part and the other way is to look at "odd" vs "even" times. Our interpretation is that the different maxima come from information the clock has about different partitionings of time. The maximum around $r \approx 0.3$ appears to come from the global picture the clock has (the "earlier" vs "later" partitioning 🟦🟥) while the second maximum at $r = 1$ comes from the "odd" vs "even" partitioning (🟥🟦🟥🟦).

Lower decay rates prove better at resolving global timing (early or late), and, while the drop clock is generally weak at resolving odd/even times, it still performs this resolution
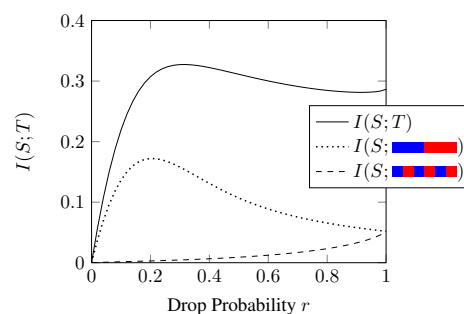

Figure 6: Time information for different drop probabilities at $T = 20$. Apart from total time information, it shows information of clock state S about being in an odd vs. even time step or early vs. late phase of the measured period.

best for a hard decay of r, right at the beginning of the interval. This explains the phase transition in Fig. 4 between the regime of small time-spans to that of large time-spans. This transition occurs because of the inflection in the information curve of the sharply initialized drop clock, when one maximum dips down below the other: different parts of the curve derive from knowing different things and because in some regimes one dominates the other.

We note that these different regimes make the drop clock hard to optimize — small time spans require a different strategy than the long ones: not only must the clock be attuned to a particular time scale to best measure global time, but also does the information curve of the clock show two maxima where one overtakes the other.

Bag of Clocks

Having studied the 2-state-clocks we now design a larger experiment. We would like to keep the clocks simple, but be able to measure time more accurately. For this purpose, we consider "bags" of independent clocks to measure time: none of the clocks in the bag can communicate with any other clock, but to determine time, we consider the state of the whole clock collective.

The experiment is started without any clocks and the collection is built up incrementally, one clock at a time by optimizing the current clock under consideration. We use this incremental process to capture our intuition: that in evolution, existing features tend to be "frozen" because they are intertwined with the rest of the organism and new features tend to be optimized in relation to the existing frozen features. More precisely, assume n clocks are already in the collection ($n = 0, 1, 2\ldots$). Given the state of the clock collection $\mathbf{S}^n = (S_1, \ldots, S_n)$ with $n = 0$ indicating the empty collection, the dynamics of clock n is optimized with grid search as to maximize $I(\mathbf{S}^n; T)$, always for fixed total duration T_{max}; the dynamic parameters of all clocks $k = 1, \ldots, n-1$ are kept fixed during the optimization. Once the optimization is complete, a new clock $n + 1$ is added and the procedure

repeated. In our experiments, we stopped the process, once we reached 10 clocks. Importantly, in this experiment, the optimizer is allowed to choose any parameters of the dynamics, and the clocks that the optimization finds are a pure alternator or else, drop clocks. As the collection grows, so does the achieved time information $I(\mathbf{S}^n; T)$ (Fig. 7). The first clock, with no oscillatory component added to the bag, is the alternator, as intuitively expected, as it resolves 1 full bit of information; but, notably, all subsequent additions to the bag turn out to be pure drop clocks. The first two clocks add ~ 1.5 bits, while every subsequent clock adds significantly less.
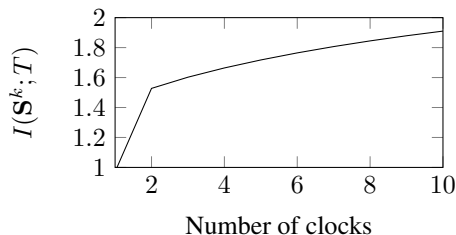

Figure 7: Amount of information as the size of the collection grows for a time interval of length $T_{\max} = 5$.

In detail, we notice that all the clocks except for the first two have very similar dynamics (parameters not shown here). In other words, once the first two clocks are added, all further clocks essentially act together as an increasingly refined binomial process, as more clocks are added[3]. Because the last clocks added to the bag have almost identical parameters, we ran another calculation to explore the behaviour of populations of identical drop clocks. We use a bag of pure drop clocks with the same $r = 0.1$, starting with one of these clocks and adding more until we reach a total of 10 clocks. We compute $I(\mathbf{S}^n; T)$ for a time scale of size 50 to create Figure 8.
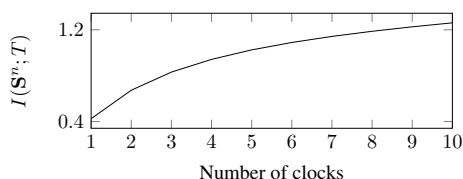

Figure 8: Time information as more identical clocks are added to the set for $T_{\max} = 50$.

Composite Clock

Beyond the simplest clocks and unstructured clock bags listed above, we consider the next more complex clock, a composite clock consisting of two simple 1-bit clocks each, which, however, are permitted to communicate, but with a constraint on how much communication is permitted. This constraint

[3]Special thanks to Nicola Catenacci-Volpi for this observation.

is expressed as a measure of information flow between the composite clocks. This information flow constraint is how we penalize modular systems for their complexity. Consistent with the freezing procedure in the bag-of-clocks experiment we freeze the first clock here as well. Thus, the first (upper) component of the composite clock becomes an alternator. Only the second (lower) component's dynamics will be parametrized and the parameters optimized according to suitable informational criteria. This one-way communication is inspired by the semigroup decomposition of counters (Rhodes, 2009). Future work will allow the first clock to be optimized as well and add a feedback channel.
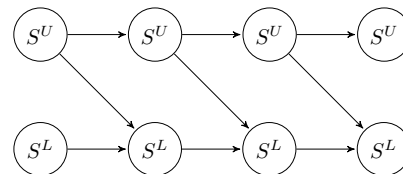

Figure 9: The structure of the composite clock unrolled in time. There is a hierarchy of information flow here. Both clocks send information to themselves but only the upper clock sends information to the lower clock.

To prepare the description of the full system, we write out in detail the first component of the clock as a Markov chain; it is defined by the matrix $A_U = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ and initial state distribution[4] $S_0^U = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. In the next step, we design the dynamics of the lower component S^L, but it is not independent anymore. Rather, it can be influenced by the state of the upper component U. The matrix that will drive the behaviour of the lower component is a conditional probability distribution, a mapping from the joint state of both components (S_t^U, S_t^L) at time t to the future state of S_{t+1}^L at time $t + 1$. There are 4 columns in this matrix because there are four combinations of the input (i.e. condition) states from both components: uu, ud, du, dd (both up, one up and the other down, etc.). In vector notation, the complete probability of the whole system to be in a state is written as:

$$P(S^U, S^L) = \begin{bmatrix} P(S^U = u, S^L = u) \\ P(S^U = u, S^L = d) \\ P(S^U = d, S^L = u) \\ P(S^U = d, S^L = d) \end{bmatrix}, \quad (1)$$

and the transition matrix for the lower clock (we remind that the upper clock is an alternator and that the first coordinate is the probability for u and the second the probability for d) as $A_L = \begin{bmatrix} \theta_1 & \theta_2 & \theta_3 & \theta_4 \\ 1-\theta_1 & 1-\theta_2 & 1-\theta_3 & 1-\theta_4 \end{bmatrix}$ and where we make use of the fact that the resulting probabilities add up to 1 (and we thus need only one parameter to describe each of the conditionals). Combine now the matrix for the upper component (A_U) and the matrix for the lower component (A_L) to obtain the Markov matrix for the whole clock:

[4]We use the random variable name as a proxy notation for the whole distribution.

$$A = \begin{bmatrix} 0 & 0 & \theta_3 & \theta_4 \\ 0 & 0 & 1 - \theta_3 & 1 - \theta_4 \\ \theta_1 & \theta_2 & 0 & 0 \\ 1 - \theta_1 & 1 - \theta_2 & 0 & 0 \end{bmatrix}.$$

We initialize the lower clock S^L, as always, in state u, i.e. with probability $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$. Now we have all required definitions. Expressed in terms of the joint variable $S = (S^U, S^L)$, and following the conventions from (1), the initial state of the complete clock is $S_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}'$ (where $'$ is transpose). Using A, the matrix that ticks the clock forward to its next time step, we can simulate the clock starting at time 0 and into a future time t by repeated matrix multiplication: A^t. For illustration, the first few states look like:

Time dynamics of both parts		
S_0	AS_0	$A^2 S_0$
$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ \theta_1 \\ 1 - \theta_1 \end{bmatrix}$	$\begin{bmatrix} \theta_1 \theta_3 + \theta_4 (1 - \theta_1) \\ \theta_1 (1 - \theta_3) + (1 - \theta_1)(1 - \theta_4) \\ 0 \\ 0 \end{bmatrix}$

Figure 10: The probabilistic state of the composite clock for $t = 0$, $t = 1$ and $t = 2$.

Experiments with the Composite Clock

The current experiments distinguish themselves from the earlier ones by having the participating clocks communicate: we create an information "tap" between the two clocks. Importantly, we will limit the amount of information that may flow from the upper to the lower clock. As discussed earlier, all costs/rewards (e.g. flow vs. time resolution) will be expressed exclusively in terms of Shannon information as unique currency of discourse.

In the same spirit as in the bag-of-clocks experiment, we greedily pre-optimize the upper clock, which results in an alternator whose parameters will be frozen. Thus, the experiment will only optimize the remaining parameters of the clock, i.e. the dynamics of the lower clock and the parameters of its dependence on the upper clock.

While we maximize time information as before, the constraint C is put on the capacity of communication (*transfer entropy*) from the upper clock to the lower one (T in the index denotes marginalization over all valid times occurring in the experiment). Since the system is Markovian, flow only over a single step is considered. Since the upper clock is fixed as the alternator, no feedback channel is included. We compute $\max_{I(S_T^U; S_{T+1}^L | S_T^L) \leq C} I(S; T)$.

To optimize with the constraint, we use the Lagrangian method, i.e. we optimize $I(S; T) - \lambda I(S_T^U; S_{T+1}^L | S_T^L)$. Scanning through the possible values of λ, we cover the spectrum of clocks arising through possible constraints. We first used the DIRECT Lipschitzian method for global optimization (Jones et al., 1993, shown in red in Fig. 11). One finds two distinct regimes: the perfect clock at $C = 1$, and distinctly

suboptimal clocks in the regime below around $C \approx 0.2$. The curve breaks off at the low end at about $C = 0.04$ due to memory limitations of the Lipschitzian optimizer.

The other optimizer used was COBYLA (Powell, 1994), where we used a relaxation method to cover the curve. We started optimization at the permissive information flow bound $C = 1$, optimized, then the bound was slowly tightened, always starting with the clock parameters obtained for the previous constraint C. The results are shown again in Fig. 11, but include the black regions in addition to the red. This optimization being not global, it uncovers additional structure in the solution space. First of all, when, starting at the optimal clock $C = 1$, C is reduced, the trade-off curve between the constraint C and the time information $I(S; T)$ falls below the convex hull of the (globally optimal) red solutions. These black points will thus never be found by a global optimization of the Lagrangian. They correspond to "fuzzy" counters (the optimal clock is a binary counter), but do not trade in information flow and achieved time information in a well-balanced way, although part of that portion of the curve is still Pareto-optimal (is not superseded simultaneously by solutions better in terms of both C *and* $I(S; T)$).

In the range $C \approx 0.5 - 0.7$, no solution is found. Between $C \approx 0.25 - 0.5$, a class of locally optimal solutions is found which is neither findable by the Lagrangian, thus not on the convex hull, nor is it Pareto optimal. Finally, below $C = 0.25$, one regains the lower C regime from the global optimization. The curve continues down to $C = 0$ (continuation of the red to the black curve), not suffering from the memory problems of the global optimizer. The two clock classes below $C \approx 0.5$ look very similar as a whole clock, but distribute the "counting" differently over their component clocks.
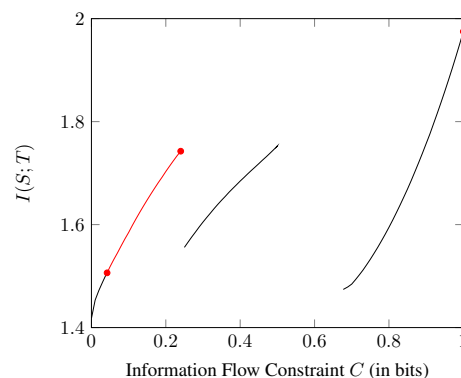

Figure 11: In red, the optimum curve found by the DIRECT global Lipschitz optimizer. In black, the more complete curve found by the COBYLA local optimizer.

Apart from the discovery of different clock "regimes", detailed inspection of the space of possible configurations (not shown here) demonstrate that, in part, finding parameters which achieve high values of time information is difficult

and, despite large permitted flows C, time resolution may still have low values. This makes it clear that measuring time is not an incidental effect that is likely to be found en passant, but one rather will expect good time measurement abilities to have been optimized for, either directly or indirectly (via proxy criteria). In any case, the clocks in the hierarchy need to be attuned to each other for optimal effect.

Final Comments and Future Work

We have studied the ability of minimal clocks to resolve time information. In particular, even 1-bit clocks can, as drop clocks, provide information about global time if the overall time horizon is known when the clock parameters are set. The relation of measured interval and clock parameters is not straightforward and marked by an interplay of global and local properties (Fig. 6). When extending to a bag-of-clocks, the first two clocks are an alternator and a drop clock, but all further drop clocks are nearly identical, so that apart from the first two, the rest of the clocks operate as a nearly binomial process. Finally, when two clocks are stacked together with limited communication, a rich set of regimes opens up, of which just two, the perfect clock, and a very soft clock, can be Lagrange-optimal. These studies provide a spectrum of candidates for behaviours of minimal clocks which one could try to identify in biological systems.

One limitation of the present work is the assumption of a fixed, global tick which drives all the clocks. Future work will, therefore, include the consideration of clocks in continuous time, where the dynamics needs to establish and sustain a synchronization between the subclocks in addition to the coordination of their respective resolution regime.

We conjecture that informationally optimal clocks will exhibit some robustness. Applying a criterium of robustness to them is left for future work.

Acknowledgements

References

Barato, A. C. and Seifert, U. (2016). Cost and precision of brownian clocks. *Phys. Rev. X*, 6:041053.

Bialek, W., Nemenman, I., and Tishby, N. (2001). Predictability, complexity and learning. *Neural Computation*, 13:2409–2463.

Chen, P. and Luo, S. (2010). Clocks and Fisher information. *Theoretical and Mathematical Physics*, 165(2):1552–1564.

Crutchfield, J. P. and Young, K. (1989). Inferring statistical complexity. *Phys. Rev. Lett.*, 63:105–108.

Hut, R. A. and Beersma, D. G. (2011). Evolution of time-keeping mechanisms: early emergence and adaptation to photoperiod. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 366(1574):2141–2154.

Jones, D. R., Perttunen, C. D., and Stuckman, B. E. (1993). Lipschitzian optimization without the lipschitz constant. *Journal of Optimization Theory and Application*, 79(1).

Karban, R., Black, C., and Weinbaum, S. (2000). How 17-year cicadas keep track of time. *Ecology Letters*, 3(4):253–256.

Karmarkar, U. R. and Buonomano, D. V. (2007). Timing in the absence of clocks: encoding time in neural network states. *Neuron*, 53(3):427–438.

Klein, D. C. and Moore, R. Y. (1991). *Suprachiasmatic nucleus: the mind's clock*. Oxford University Press, USA.

Klyubin, A. S., Polani, D., and Nehaniv, C. L. (2007). Representations of space and time in the maximization of information flow in the perception-action loop. *Neural Computation*, 19(9):2387–2432.

Langton, C. (1989). Artificial life. In Langton, C., editor, *Artificial Life*, pages 1–47. Addison-Wesley.

Laughlin, S. B. (2001). Energy as a constraint on the coding and processing of sensory information. *Current Opinion in Neurobiology*, 11:475–480.

Marwan, W. and Oesterhelt, D. (1987). Signal formation in the halobacterial photophobic response mediated by a fourth retinal protein (p480). *Journal of molecular biology*, 195(2):333–342.

Nehaniv, C. L. (1993). The algebra of time. In *Proc. National Conf. of the Japan Society for Industrial and Applied Mathematics*, pages 127–128.

Nutsch, T., Marwan, W., Oesterhelt, D., and Gilles, E. D. (2003). Signal processing and flagellar motor switching during phototaxis of halobacterium salinarum. *Genome research*, 13(11):2406–2412.

Oliveira, N. M., Foster, K. R., and Durham, W. M. (2016). Single-cell twitching chemotaxis in developing biofilms. *Proceedings of the National Academy of Sciences*, page 201600760.

Polani, D. (2009). Information: currency of life? *HFSP journal*, 3(5):307–316.

Powell, M. J. D. (1994). A direct search optimization method that models the objective and constraint functions by linear interpolation. In Gomez, S. and Hennart, J.-P., editors, *Advances in Optimization and Numerical Analysis*, pages 51–67. Kluwer Academic (Dordrecht).

Rhodes, J. (2009). *Applications of automata theory and algebra: via the mathematical theory of complexity to biology, physics, psychology, philosophy, and games*. World Scientific.

Sota, T., Yamamoto, S., Cooley, J. R., Hill, K. B. R., Simon, C., and Yoshimura, J. (2013). Independent divergence of 13- and 17-year life cycles among three periodical cicada lineages. *Proceedings of the National Academy of Sciences*, 110(17):6919–6924.

van Wassenhove, V. (2012). From the dynamic structure of the brain to the emergence of time experiences. *Kronoscope*, 12(2):201–218.