Table of contents

1 Strains

1.1 Plasmids

All plasmids used in this study are listed in Table S1.

The constructs were made using either isothermal assembly (ITA)[37] or 'round-the-horn' site-directed mutagenesis (`http://openwetware.org/wiki/%27Round-the-horn_site-directed_mutagenesis`). Primers and gBlocks were ordered from Integrated DNA technologies (IDT) and PCR were performed with Accuprime Pfx (Life technologies) or Phusion polymerase (New England BioLabs) according to manufacturers' protocols. All plasmids were verified by DNA sequencing (Genewiz). In the text, *laa* is used as a shorthand for the native *ssrA*-(LAA) (AANDENYALAA) and *asv* for the synthetic *ssrA*-(ASV) (AANDENYAASV) degradation tag.

pZE21-GFPasv refers to the original reporter[38], while pZS1-lTlrLLtCL to the original repressilator plasmid. Both were kind gifts of M. Elowitz. The $P_{LlacO-1}$ and $P_{LtetO-1}$ are the hybrid promoters orginally used by switching the binding sites of bacteriophage λ promoter P_L with lacO1 and tetO2 binding sequences, respectively[39]. They are abbreviated P_{Llac} and P_{Ltet} in the text for simplicity. P_R refers to the native bacteriophage λ promoter.

pLPT20 was constructed by assembling *mVenus* (kind gift of the P. Cluzel, Ref.[40] with A206K), P_{RNA1} and *cfp* (kind gift of the P. Cluzel) into a temporary reporter plasmid (pZE21-GFPasv derivative), PCR-amplified and inserted into pZS1-lTlrLLtCL plasmid digested at the unique AatII site.

pLPT41, pLPT42, pLPT43 and pLPT57 were constructed by 'round-the-horn' PCR from pZE21-GFPasv removing different part of the construct. For pLPT57, we merely removed the *asv* degradation tag. The *pep* in pLPT43 refers to keeping a peptide consisting of the N-terminal 40 amino acids fragment of *gfpmut3* (MRKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYG) followed by the *asv* degradation tag. In pLPT41, we kept the promoter, ribosome binding site (RBS) and terminator intact while removing the *gfp*. For pLPT42, we also removed the promoter.

pLPT47 was assembled from a synthesized gBlock and fragment of pLPT1. The P_{const} refers to a synthetic, intermediate strength and constitutive promoter from the iGEM Registry of standard biological parts (`http://parts.igem.org/`, BBa_J23107, tttacggctagctcagccctaggtattatgctagc). We used a Y66L mutation on the GFP to make it non-fluorescent[41] and therefore compatible with the YFP reporter of the integrated repressilator.

pLPT107 was constructed by inserting synthesized gBlocks as well as PCR-amplified fragments into pLPT20 backbone. P_R-*mKate2* was added and the P_{RNA1} promoter was changed to a P_{Llac}. The *mKate2* was a gift from D. Landgraf and consists of the *mCherry* N-terminal 11 amino acids followed by the *mKate2* sequence (used to improve translational efficiency).

pLPT119 was assembled from PCR fragments of pLPT20, simply removing the degradation tags on the repressors.

1.2 Bacterial strains

All *E. coli* strains used in this study are listed in Table S2. The wild-type strain was MC4100.

DHL708 was a kind gift of D. Landgraf[42] and was built by deleting the *clpPX* operon with lambda-Red mediated homologous recombination. The FRT-flanked Kan cassette was then flipped out using the FLP recombinase (pCP20).

Plasmid	Parent	Ori	Antibiotic	Genotype
pDHL474	Gift of D. Landgraf	ColE1	Kan	P_{LtetO1}-*mCherry-asv*
pDHL490	Gift of D. Landgraf	ColE1	Kan	P_{LtetO1}-*mCherry-laa*
pZE21-GFPasv	Gift of M. Elowitz	ColE1	Kan	Reporter, P_{LtetO1}-*gfpmut3-asv*
pZS1-lTlrLLtCL	Gift of M. Elowitz	pSC101	Amp	Repressilator, P_{LtetO1}-*cI-laa*, P_R-*lacI-laa*, P_{LlacO1}-*tetR-laa*
pLPT20	pZS1-lTlrLLtCL	pSC101	Amp	Integrated repressilator, P_{LtetO1}-*cI-laa*, P_R-*lacI-laa*, P_{LlacO1}-*tetR-laa*, P_{LtetO1}-*venus*, P_{RNA1}-*cfp*
pLPT41	pZE21-GFPasv	ColE1	Kan	P_{LtetO1}
pLPT42	pZE21-GFPasv	ColE1	Kan	empty
pLPT43	pZE21-GFPasv	ColE1	Kan	P_{LtetO1}-*pep-asv*
pLPT47	pZE21-GFPasv	ColE1	Kan	P_{const}-*darkgfp-asv*
pLPT57	pZE21-GFPasv	ColE1	Kan	P_{LtetO1}-*gfpmut3*
pLPT107	pLPT20	pSC101	Amp	Triple reporter repressilator, P_{LtetO1}-*cI-laa*, P_R-*lacI-laa*, P_{LlacO1}-*tetR-laa*, P_{LtetO1}-*venus*, P_{LlacO1}-*cfp*, P_R-*mKate2*
pLPT119	pLPT20	pSC101	Amp	Integrated repressilator no degradation tags, P_{LtetO1}-*cI*, P_R-*lacI*, P_{LlacO1}-*tetR*, P_{LtetO1}-*venus*, P_{RNA1}-*cfp*
pLPT144	LPT42	ColE1	Kan	P_{LlacO1}, P_R
pLPT145	LPT41	ColE1	Kan	P_{LtetO1}, P_{LlacO1}, P_R
pLPT146	LPT42	ColE1	Kan	P_{LlacO1}
pLPT147	LPT42	ColE1	Kan	P_R
pLPT148	LPT41	ColE1	Kan	P_{LlacO1}, P_{LtetO1}
pLPT149	LPT41	ColE1	Kan	P_{LtetO1}, P_R
pLPT151	pDHL490	ColE1	Kan	P_{LtetO1}-*mCherry-laa*, P_{const}-*darkgfp-asv*

Table S1: Plasmid list

NDL319 was constructed by inserting the P_{RNA1}-*mKate2* hybrid cassette close to (while preserving) the neutral Tn7 insertion locus[43].

Strain	Parent	Genotype	Plasmids	Period (gen)	Correlation after one period	Generation time (min)
DHL708	MC4100	Δ*clpXP*				
LPT25	MC4100		pLPT20	5.6	0.18	26
LPT44	LPT25		pLPT20, pLPT41	5	0.2	29
LPT45	LPT25		pLPT20, pLPT42	5.5	0.14	27
LPT46	LPT25		pLPT20, pLPT43	4.2	0.16	27
LPT53	LPT25		pLPT20, pLPT47	2.3	0.09	27
LPT54	LPT25		pLPT20, pDHL474	2.6	0.07	27
LPT55	LPT25		pLPT20, pDHL490	5.7	0.2	28
LPT60	MC4100		pZS1-lTlrLLtCL, pLPT57	5	0.18	26
LPT61	DHL708	Δ*clpXP*	pLPT20	10.4	0.09	25
LPT64	LPT61	Δ*clpXP*	pLPT20, pDHL474	13.9	0.5	27
LPT113	DHL708	Δ*clpXP*	pLPT107	14	0.2	26
LPT117	LPT113	Δ*clpXP*	pLPT107, pLPT43	14.8	0.5	28
LPT118	LPT113	Δ*clpXP*	pLPT107, pDHL474	14.9	0.4	29
LPT120	MC4100		pLPT119	10	0.07	25
LPT124	LPT120		pLPT119 pLPT41	12.9	0.3	25
LPT125	LPT120		pLPT119 pLPT43			
LPT127	DHL708	Δ*clpXP*	pLPT107, pLPT41	14.4	0.5	28
LPT143	LPT61	Δ*clpXP*	pLPT20, pLPT41			
LPT150	MC4100	att::Tn7 P_{clpPX}-*clpP*-*msfgfp*				
LPT152	LPT61	Δ*clpXP*	pLPT20, pLPT145	15.6	0.48	26
LPT153	LPT61	Δ*clpXP*	pLPT20, pLPT146			
LPT154	LPT61	Δ*clpXP*	pLPT20, pLPT147			
LPT155	LPT61	Δ*clpXP*	pLPT20, pLPT148			
LPT156	LPT61	Δ*clpXP*	pLPT20, pLPT149	13.3	0.47	25
LPT157	LPT61	Δ*clpXP*	pLPT20, pLPT144			
LPT158	MC4100		pDHL490			

Table S2: Strain list and summary of experimental results

Strain	Parent	Genotype	Plasmids	Period (gen)	Correlation after one period	Generation time (min)
LPT159	LPT150	att::Tn7 P_{clpPX}-*clpP-mgfpmut3*	pLPT20	5.4	0.17	25
LPT165	LPT159	att::Tn7 P_{clpPX}-*clpP-mgfpmut3*	pLPT20, pDHL474	2.8	0.1	26
LPT172	MC4100		pLPT151			
MC4100						
NDL319	MC4100	att::Tn7 P_{RNA1}-*mKate2*				
NDL332	NDL319	att::Tn7 P_{RNA1}-*mKate2*	pZE21-GFPasv, pZS1-lTlrLLtCL	2.4	0.15	25

Table S3: Strain list and summary of experimental results (continued)

2 Microfluidic master fabrication

Fabrication of the *E. coli* mother machine was carried out using standard UV photolithography in a clean room environment. The device was designed using AutoCad, and quartz-chrome photomasks were ordered from Toppan Inc. and from the Center for Nanoscale Systems at Harvard University. We modified the fabrication procedure from the method described in the original report on the mother machine. The fabrication requires three independent layers: an Su8 'base' coat, cell channels, and feeding channels. The process parameters below should be regarded as useful starting points, and all feature dimensions must be empirically verified to ensure that they meet experimental requirements. The tolerances for this device are stringent; a difference of ±0.5 μm in the height or width of a cell channel will compromise function. Note: for all spin coater steps described below, the following shorthand notation is used: speed (rpm)/acceleration (rpm/sec)/time (sec).

2.1 First Layer: Su8 Base Coat

The first layer of the master contains no features. It consists of a thin, uniform layer of completely cured Su8. This layer is intended to serve as an adhesive to improve the retention of the cell channels deposited in the second layer. Without this base, the cell channels are easily removed from the wafer surface during development.

1. Place a new 3'' Si wafer (we used 380 μm TEST grade wafers from University Wafer) in a dish of fresh acetone. Sonicate at high power for 5 minutes.
2. Sequentially rinse the wafer with streams of methyl alcohol (MeOH), isopropyl alcohol (IPA) and H_2O (~ 10 seconds per solvent).
3. Place wafer on 2'' spin chuck and spin seconds at 500 rpm.
4. While spinning, sequentially rinse the wafer with streams of MeOH, IPA and H_2O.
5. Spin wafer 1 minute at 3,000 rpm to dry.
6. Dehydrate wafer for 15 minutes on a hot plate set to 150-200°C.
7. Set spin program to: Step 1: 500/100/10, Step 2: 3000/300/60.
8. Place the dehydrated wafer onto the spin coater chuck and dispense a small (cover ~2/3 of the wafer surface) amount of Su8 2000.5 photoresist (Microchem) with a pipette. Run the spin program. This should result in a coat of ~0.5 μm.
9. Soft bake wafer (in order) for 1 minute at 65°C, 1 minute at 95°C, 1 minute at 65°C.
10. Expose wafer for 5 seconds with no mask at 25 mW/cm^2 (I-line).
11. Post exposure bake the wafer (in order) for 1 minute at 65°C, 1 minute at 95°C, 1 minute at 65°C.
12. Do not develop or hard bake the wafer. Instead, proceed directly to the second layer protocol.

2.2 Second Layer: Cell Channels

This set of steps lays down the channels that house the cells in the final device. The tolerances for this layer are very stringent; the exposure dose and contact between mask and wafer must be optimized. We recommend trying a range of exposure parameters to ensure that a useful device is obtained. We also stress the importance of the very long post exposure bake time in the process below. In our hands, this greatly improves the retention of the cell channels during development.

1. Set spin program to: Step1: 500/100/10, Step 2: 2000/300/60.
2. Place the wafer onto the spin coater chuck and dispense a small (cover 2/3 of the wafer surface) amount of Su8 2001 photoresist with a pipette. Run the spin program. This should result in a coat of ~1.5 μm.
3. Soft bake wafer (in order) for 1 minute at 65°C, 3 minutes at 95°C, 1 minute at 65°C.
4. Expose wafer for 0.75 seconds (25 mW/cm^2, I-line) through cell channel mask in vacuum contact mode.
5. Bake wafer for 1 minute at 65°C, 20 minutes at 95°C, 1 minute at 65°C.
6. Develop wafer for 30 seconds with very gentle agitation in Su8 developer.
7. Rinse wafer for 10 seconds with IPA.
8. Check completeness of development process. If undeveloped Su8 remains on the wafer (other than the desired cell channels) repeat developer treatment for 10 seconds.
9. Hard bake wafer for 10 minutes at 150°C.
10. Verify channel height using a profilometer. The expected height is 1.5 μm. If the channel dimensions lie outside of your expected tolerance bounds, the process must be repeated with modified spin coating parameters.

2.3 Third Layer: Feeding Channels

The final layer of the device forms the medium flow channels. The dimensions of these features are not critical: we have used feeding channels of widely varying dimension to similar effect. The alignment is sensitive to large errors, however. The alignment between feeding channels and cell channels must be accurate (down to a couple of microns) in order to ensure that the cell channels are of the desired final length.

1. Set spin program to: Step 1: 500/100/10, Step 2: 5000/300/60.
2. Place the wafer onto the spin coater chuck and dispense a small (cover ~ 2/3 of the wafer surface) amount of Su8 2025 photoresist with a pipette being careful not to introduce bubbles. Run the spin program. This should result in a coat of ~ 15 μm.

3. Soft bake the wafer (in order) for 1 minute at 65°C, 4 minutes at 95°C, 1 minute at 65°C.
4. With an Su8-developer-soaked swab, clean the newly-deposited photoresist off of the alignment marks to make them visible for the alignment process.
5. Soft bake the wafer (in order) for 1 minute at 65°C, 4 minutes at 95°C, 1 minute at 65°C.
6. Align feeding channel mask to the alignment marks on the wafer. Apply vacuum contact and check alignment again. If the vacuum application skewed the alignment, repeat the alignment process.
7. Expose wafer for 10 seconds (25 mW/cm^2, I-line) through aligned feeding channel mask.
8. Bake wafer for 1 minute at 65°C, 4 minutes at 95°C, 1 minute at 65°C.
9. Develop wafer for 1.5 minutes in Su8 Developer with mild agitation.
10. Rinse wafer for 10 seconds in IPA. Check to ensure that the development is finished. If undesired photoresist remains, develop again for 20 seconds.
11. Hard bake wafer for 15 minutes at 150°C.
12. Verify channel height using a profilometer. The expected height is 15 μm. If the channel dimensions lie outside of your expected tolerance bounds, the process must be repeated with modified spin coating parameters.

3 Supplementary results and discussion

3.1 Interference of the reporter plasmid

3.1.1 ASV-tagged molecules

The results for the period change due to the introduction of ASV-tagged molecules are summarized in Extended Data Fig. 3. The addition of substrates tagged with the native LAA tag slowed down the oscillations (compared to the titration sponge, Extended Data Fig. 3d), as expected if it serves as a competitive substrate of ClpXP. Surprisingly, however, the ASV version of the degradation tag made the periods faster, and the effect became more pronounced the more numerous and stable the ASV tags were. The identity of the ASV-tagged substrate changed the magnitude of the effect: expressing a small peptide-ASV, repressilator-controlled mCherry-ASV or constitutive dark GFP (Y66L), reduced the period from 5.5 to 4.2, 2.6 and 2.4 generations, respectively. Over-expressing a functional ClpP-mGFPmut3 fusion[42] decreased the period slightly, but could not compensate for the presence of ASV-tagged substrates.

To confirm that the proteolysis of the natural, *ssrA*-tagged molecules was faster in the presence of ASV-tagged molecules, we expressed both in the absence of the repressilator (Extended Data Fig. 3e). For the same production rate, the mean abundances of *ssrA*-tagged molecules was reduced four-fold in the presence of ASV-tagged molecules. This response was not observed in the Δ*clpXP* strain (Extended Data Fig. 3f-h): it did not make a significant difference if the titration sponge – in the Δ*clpXP* strain – expressed an ASV-tagged protein or contained only the promoter. Furthermore, the response was also not observed in the repressilator that did not contain degradation tags on the repressors (Extended Data Fig. 6f).

3.1.2 Titration sponge

Addition of the titration sponge had only a minor effect on the fast and medium oscillators: the period became slightly shorter (from 5.5 to 5 generations), and the oscillations slightly better, with a phase drift of 24% per period (Extended Data Fig. 5) and a slight increase in correlation (Extended Data Fig. 3a). This is consistent with expectations since many proteins are not degraded, or degraded more slowly, when bound to DNA[44]. Thus in the strains without repressor degradation, the proteins and repressor sites are diluted out and replaced on the same time-scale, but when repressors are degraded, the repressors turn over after the titration sponges fills up, and the abundance of the free component becomes almost independent of the titration sites (see §4.3).

Trying all possible combinations of titration sponges for the different repressors (Extended Data Fig. 7d) showed that only the sponges containing binding sites for the TetR repressor displayed macroscopic oscillations, and that titration of the other repressors did not influence greatly the oscillations. This suggests that the activity threshold is higher for these repressors, as further supported by our observations of absolute abundances and the amount of dilution required to trigger production in the next step.

3.2 Robustness of the oscillations

The phase of the repressilator oscillations were independent of the cell cycle progression (Extended Data Fig. 9a), as expected. The period of the modified repressilator with degradation (LPT25) increased when grown at lower temperature (up to 10 generations at 25°C). This was expected since degradation of many components ceases during slow growth. The repressilator without degradation (including the titration sponge) was remarkably robust (Fig. 3a). It oscillated at 13.8±0.2 generations at three different temperatures, as well as with conditioned media from early stationary phase culture (OD_{600} ~2), where the cell physiology was radically different (i.e. see Supplementary video 7). Over these conditions, the standard deviation in the average period was ~2%. This robustness across growth conditions is in fact expected for oscillators relying only on dilution, for a combination of reasons.

First, though the total expression rates vary greatly between conditions, *E. coli* to a large extent compensate with changes in size that keep concentrations more constant. For example, under conditions where they make twice as much proteins, they also roughly have twice the volume since much of the mass is in protein form. If this applies to the repressors that form the repressilator, it means that the amplitude peaks may be twice as high, but so is the derepression threshold which is set by the concentration. The number of generations required for dilution, from the peak to the next derepresion, is then the same.

Second, the exponential nature of dilution means that there is a logarithmic dependence between the fold-reduction (from the peak to the derepression threshold) and the time that reduction takes. For example, a two-fold change in the required dilution factor only changes the dilution time with one cell generation, which is much less than a two-fold change when oscillations are slow. Likewise, since the build-up phase approaches a steady state, the height of the peak is insensitive to further increases in the dilution time of the repressor controlling the height of that peak.

Third, protein degradation changes greatly between growth conditions (without a compensating change in concentration because so few proteins are actively degraded), but the 14-generation oscillator does not rely on protein degradation in any way.

3.3 Removing the degradation tags of the repressors

Removing the degradation tags from the repressors in cells that were wild-type for ClpXP produced similar results to having degradation-tagged repressors in the Δ*clpXP* strain (Extended Data Fig. 6). The oscillations were again quite irregular, but the addition of the titration sponge again made the oscillations much more regular, and e.g. enabled the bacterial colony to exhibit ring patterns. Combining the repressilator without degradation tags on the repressors with the Δ*clpXP* strain also gave similar results.

3.4 Stability of the conditions in the microfluidics device

Characterization of the microfluidic device has shown that the conditions experienced by the cells were very uniform both across the device and over time. Division times were stable over multiple days, and growth distributions very similar across different experiments (Extended

Data Fig. 8a). The period of the oscillations was uniform spatially (different cells, position in the field of view, distance to the inlet/outlet, media channel) and temporally (Extended Data Fig. 8b).

3.5 Characterization of the oscillations

3.5.1 Using binomial partitioning to estimate absolute protein abundances

To roughly estimate the average abundances of the repressor proteins in the strains without repressor degradation, we considered the statistical partitioning error at cell division of the corresponding fluorescent proteins, which are expressed under the same promoters from the same plasmid. This allows to approximately convert fluorescence units into number of proteins, as previously demonstrated[45].

Specifically, we measured the root mean squared (RMS) errors between daughters at cell division during the dilution phase of the repressilator oscillation, and confirmed that it scaled binomially with total abundance (Extended Data Fig. 8f), as expected for fluorescent proteins diffusing freely in the cytoplasm. We then used the magnitude of the error to identify the absolute number of molecules. One caveat of this approach, in addition to the fact that the fluorescent proteins are just a proxy for the actual repressors, is that any aggregation of the proteins could increase the error but still produce binomial scaling. For example, if all proteins formed dimers and these segregated independently of each other, the errors would still be binomial but the average abundance would be off by a factor of two. Fluorescent proteins primarily multimerize when tagged to native proteins that multimerize[42] (increasing the local concentration), but since all red fluorescent proteins that we know of have severe multimerization tendencies as well as very slow maturation, including the mCherry used here, we only estimate a lower bound for the abundance of this component.

The results suggest that the YFP and CFP transcriptional reporters peaked at around 2,000-3,000 copies (Extended Data Fig. 8d,g) per cell, while for mCherry we can only say the levels are higher than about 500 copies, which gives a rough estimate of the peak values of the repressors. These numbers should only be seen as a very rough approximation, since some of the repressors form dimers which affects the estimated numbers, but are at least consistent with the dilution times observed, and the fact that the titration sponges had a substantial effect.

3.5.2 Repression by TetR occurs at very low levels

We attempted to measure input-output relations for the fast-maturing CFP and YFP transcriptional reporters (CFP reports levels of TetR and YFP reports on CI), but not for mCherry due to the longer maturation time. We established a detection limit of at least $\sim$ 50 FU for CFP in our imaging conditions, due to combined contributions of auto-fluorescence and shot noise. As shown in Extended Data Fig. 8c, the switch between induction and repression occurs below the detection limit of our transcriptional reporter.

Even if the repression switch occurs below the detection limits, it is possible to get a rough estimate for the location of the derepression threshold by using the protein number at the peak and the duration of the dilution phase obtained from the triple reporter analysis. This suggests a switching point of approximately 5 molecules for the oscillator without titration sponge (LPT61), and approximately 20 molecules with titration sponge (LPT64). Note that the errors in estimation

of the dilution phase and the conversion of fluorescent units to molecules are exponentially amplified, and thus again this should only be considered a rough estimate.

3.5.3 Peak amplitude heterogeneity and dampening

Even in the very precise oscillators, we observe significant peak amplitude heterogeneity, with a CV of about 35% (Extended Data Fig. 8d). However, this heterogeneity is dampened by the exponential dilution, which results in a dependence of the next period in the log of the peak height (e.g. for a peak twice higher than average, the period will be only one generation longer). Using the law of total variance, the fit to a log of the peak height explains 25% of the variance in the period (Extended Data Fig. 8e).

4 Theory

Because key properties of the single-cell kinetics – like the cooperativity or the noise in plasmid copy numbers – are unknown even for the repressilator, we use theory not to validate detailed models, but to identify potential design challenges. First we show that a simple repressilator model can behave either as a harmonic oscillator or a relaxation oscillator, depending on parameters, and that the data show that it operates in the latter regime. The criteria for stable oscillations are then very different, and we show that relaxation repressilators can display regular and sustained oscillations with minimal phase lags or cooperativity. We then analyze how noise arises in the accumulation and decay phases of the oscillations, and how titration sponges affect the dynamics. Finally, in Appendix A we derive exact limits on the accuracy of repressilator-type systems – regardless of control mechanisms – due to information loss and finite numbers of control molecules.

4.1 Harmonic versus relaxation oscillators

We start by considering a minimal model of the repressilator:

$$p_i \xrightarrow{\lambda \frac{K^n}{K^n + p_j^n}} p_i + 1$$

$$p_i \xrightarrow{p_i} p_i - 1$$

where $(i, j) = (1, 3), (2, 1), (3, 2)$, λ is the production rate, K the protein level at half-repression, n the Hill coefficient and p_i the number of repressor proteins. We have set the protein degradation rate constants to 1 without loss of generality, and thus express time in units of average protein lifetimes. Note that the production rates are therefore in units of protein produced per protein lifetime.

Depending on the parameters, this system can behave as a harmonic oscillator, where stability properties are captured by linearizations, or as a relaxation oscillator[46] (Extended Data Fig. 4a) with self-sustained periodic repetitions of an aperiodic phenomenon with a period set by a relaxation time. As opposed to sinusoidal/harmonic oscillators, relaxation oscillators exhibit discontinuous jumps and all-or-none behavior. This well describes the repressilator dynamics

we observe: as shown in the three-color analysis, the production of a repressor starts abruptly and remains constant until it suddenly turns off. In fact, even the seemingly gradual reduction in expression before the turn-off observed for the 14 generation oscillator is a part of the relaxation path rather than a result of control exerted: measuring the production rate and dilution rate shows that the concentration only starts to level off because dilution effects become increasingly important at high numbers. The strongly non-sinusoidal shape thus comes from thresholding, where promoters spend most of their time in either a completely on-state or completely off-state. The mathematical requirements for oscillations are then very different. Classical stability analysis for harmonic oscillators emphasizes the importance of phase lags and high cooperativity, but in the relaxation regime it is possible to oscillate without much of either – in fact for extreme parameters, stochastic models show that oscillations are possible without any phase lag or cooperativity (Extended Data Fig. 4d). Due to the clear separation of the two phases – a build-up phase with constant production and a pure decay phase with no production (Extended Data Fig. 4c) – we next analyze how much noise arises in each phase, with a particular focus on noise in timing rather than in amplitude. The period can either be seen as the sum of three decay phases, or as the sum of three production phases, since the beginning of a production phase of one repressor marks the beginning of a decay phase of another. We find that a key architectural feature of the repressilator – in the observed relaxation mode – is that variation in the build-up occurs in a way has little effect on the overall timing, while timing noise from the decay phase is hard to avoid given the low promoter derepression thresholds observed. Specifically we iterate some known results about noise in gene expression to show how removing degradation tags reduces the noise in the starting value of the decay phase, and then analyze the inherent variation in the decay phase as a function of the starting value and the derepression threshold. Finally we consider the propagation of noise between the build-up and decay phases, i.e., how variation in peak amplitude affect the decay phase and how variation in the decay phase affects the build-up phase.

4.2 Heterogeneity in the accumulation and decay phases

4.2.1 Accumulation phase

Motivated by the experimental observation that the oscillation cycle has a build-up phase with almost constant production rate, we consider the variation in the number of proteins made during a time window T_a during which gene expression is unrepressed. Most of the results presented here are well known from the extensive literature on stochastic gene expression and are only briefly repeated here for convenience.

First, consider a process in which proteins are made constitutively in bursts of size b:

$$p_i \xrightarrow{\lambda} p_i + b$$

The CV (standard deviation over mean) in the number of proteins produced is then $1/\sqrt{\lambda T_a}$, and even if we further allow variation in b, the CV is still inversely proportional to the square root of λT_a. Statistical accuracy can thus be increased either by increasing the production rate, or by integrating over longer time periods by increasing T_a. The latter can be achieved by eliminating repressor degradation and thereby extending the periods.

If the build-up phase is short, so that dilution/degradation is negligible, the number of proteins produced is similar to the peak amplitude (the troughs of the oscillations are virtually at zero so proteins from the previous build-up phase contribute very little). However, for long enough times T_a, decay and dilution eventually become non-negligible, and the system approaches a steady-state (this was directly observed in the data for the slowest oscillator but is likely the case also for the faster oscillator where the actively degraded repressors should approach a steady state much faster than the stable reporter observed). Increasing T_a further thus does not reduce the CV in peak amplitude. Specifically we consider:

$$p_i \xrightarrow{\lambda \mathbf{z}} p_i + b$$

$$p_i \xrightarrow{p_i/\tau_p} p_i - 1$$

where $\mathbf{z}$ is some extrinsic noise variable, b is a geometrically distributed burst size and τ_p the lifetime of the protein. For simplicity, if we assume $\mathbf{z}$ is following its own birth-death Poisson process, then at the end of the build-up phase[47–49]:

$$CV_p^2 = (1 + \langle b \rangle) \frac{1}{\langle p_i \rangle} + \frac{1}{\langle z \rangle} \frac{\tau_z}{\tau_p + \tau_z}$$

where τ_z is the lifetime of z. Therefore, increasing the proteins' lifetimes still reduces noise by increasing the average protein abundance and time-averaging over extrinsic influences.

4.2.2 Decay phase

Motivated by the experimental observation that the decrease in repressor concentration is indistinguishable from exponential decay, we here model the variation expected in the time until the repressor reaches low enough concentrations that expression of the next component in the cycle is derepressed. Specifically, we are interested in the average and standard deviation of the decay time T_d, from an initial peak of N to a threshold S. This time contributes one third of the period, and determines the build-up time of another repressor in the system. We ignore two biological complications. First, for systems with actual degradation of proteins, fluctuations in the proteolysis machinery could produce fluctuating degradation rate constants, i.e., essentially contributing an extrinsic source of noise. However, this noise is unknown, and our primary purpose with this theory is to explain the effects observed for the circuits without active degradation, when molecules are diluted out during the cell cycle and then randomly partitioned between daughters at cell division. For binomial partitioning (as we directly observe, see section 3.5.1), this can be modeled by explicitly accounting for cell growth and division, as we and others have done in other studies[47]. However, because our goal here is to produce insights into experimental design rather than to capture the exact dynamics, we use a common approximation where binomial segregation and dilution are approximated with first order decay, i.e., where the protein decays in individual random steps:

$$p_i \xrightarrow{p_i} p_i - 1$$

Here we again have set the degradation rate constant to one, to express time in units of average protein lifetimes.

Each step is thus probabilistic with an exponentially distributed waiting time, and the total time to decay from N to S molecules is a sum of the random variables for the individual steps. For very high thresholds, where S is almost as high as N, the standard deviation in the total time is substantial compared to the average total time, since few steps are summed. For very low thresholds there are large numbers of steps, but the last few steps are rate limiting and dominate the sum. For example, the average waiting time for going from N to $N-1$ molecules is N times shorter than the average waiting time for going from 1 to 0 molecules. If the derepression threshold is reached when the last molecule is eliminated (S=0), the total variance in the time equals $\sum_{i=1}^{N} 1/i^2$, which is 1 for $N = 1$ and quickly approaches $\pi^2/6$ for higher N. The average time is in turn $\sum_{i=1}^{N} 1/i$, which is approximately $\log(N) + \gamma$ where $\gamma = 0.57721\ldots$ is Euler's constant. More generally, for any threshold $S < N$, the variance is $\sum_{k=S+1}^{N} \frac{1}{k^2}$, which approaches $\frac{1}{S+1} - \frac{1}{N}$ for large N and fixed S/N* (which in turn is approximately $1/(S+1)$ if, in addition, the ratio N/S is sufficiently large†.) The average follows $\sum_{k=S+1}^{N} \frac{1}{k} \approx \log(N/(S+1))$ under the same conditions. Thus, for a thresholding scheme of this type

$$CV \approx \frac{\sqrt{1/(S+1)}}{\log(N/(S+1))}$$

If N/S is large the CV increases very slowly with N, because the logarithm becomes insensitive to changes in N/S. For example, doubling N/S from $1,000$ to $2,000$ only increases $\log(N/S)$ by about 10%. For any given N, there is also an optimal threshold that minimizes the CV at $(S+1)/N = 1/e^2 = 0.13534$, giving a mean of 2 and a variance of e^2/N. Therefore:

$$CV \approx \frac{\sqrt{N/(S+1)}}{\log(N/(S+1))} \frac{1}{\sqrt{N}} \geq \frac{e/2}{\sqrt{N}} = \frac{1.359}{\sqrt{N}}$$

Extended Data Fig. 4e in turn shows the exact CV as a function of N for various values of S, which closely follows these approximations since $N \gg 1$ throughout. For any fixed S, the decrease in CV with N is thus logarithmic, and it is necessary to increase S along with N in order to reduce the CV effectively. However, the CV remains close to its minimum for a rather wide range of S/N values (e.g. between 0.05 and 0.3, see Extended Data Fig. 4e). Thus if S is increased along with N, it should be possible to decrease the CV by the inverse square root of N. These simple theoretical findings could directly explain our experimental observations. The estimate of the TetR threshold suggests that it was lower than optimal even before eliminating repressor degradation. Thus N/S was already very high and we should only expect a very minor reduction in the heterogeneity of the timing, as we indeed observed. However, when the threshold is also increased (Fig. 2b) by adding a titration sponge, the variation in the timing should be more drastically reduced by the increased N, as we also observed.

4.2.3 Transmission of noise between phases

Fluctuations in the decay time will also affect and be affected by fluctuations in amplitude. Here we show how this noise propagation to some extent is naturally damped by the mechanisms of the

*specifically $N \sum_{k=S+1}^{N} \frac{1}{k^2} \to \left(\frac{N}{S+1} - 1\right)$ as $N \to \infty$.

†Note that we sometimes use the approximate N/S ratio (instead of $N/(S+1)$) for simplicity as it is good approximation for the S and N values considered.

system. First, a doubling in the peak amplitude will only result in an increase of one generation of the decay phase, which is small given that the oscillator has a period of 14 generations and each repressor thus spends 4-5 generations in the decay phase. More generally, the time it takes to exponentially dilute out levels from N molecules to S molecules on average approximately equals $\log(N/(S+1))$. The additional variance in the decay period due to the variation in peak amplitude can be estimated using the rule for error propagation for functions of random variables (a small deviations approach), which immediately shows that for $\log(N/(S+1))$ the additional variance in timing is given by $\sigma_N^2/\langle N\rangle^2$, which equals $1/\langle N\rangle$ if the initial number of molecules is Poisson distributed. If the repressor molecules are produced in geometrically distributed bursts of size b this variance increases to $(1+\langle b\rangle)/\langle N\rangle$. In either case, as long as N/S is reasonably high this term will be small in comparison to $\langle 1/(S+1) - 1/N\rangle$ [‡], the expected variance in the decay steps, and so little noise is transferred from the peaks to the periods, as we also observe experimentally (Extended Data Fig. 8e).

Noise is also transmitted from the decay phase to the accumulation phase, as the decay time T_d of one repressor determines the accumulation time T_a of another. In the parameter regime in which the repressor in question does not have time to approach a steady-state before production is turned off again, significant noise is propagated because a doubling in the accumulation time will result approximately in a doubling of the amplitude peak. Even in this regime, little noise propagates back to affect the next decay phase due to the logarithmic principle above. Furthermore, in a regime where the repressor reaches a (quasi-)steady-state before production stops, little noise is transmitted since fluctuations in the accumulation time then only change the time the repressor spends around its steady-state value.

4.3 The role of the titration sponge

The theoretical analysis above shows that low repression thresholds S create large noise in the decay phase of the oscillation even for perfect threshold mechanisms. This seems to explain why titration sponges improve the regularity of oscillations so much. In the next section we include titration in a simulation model, and here we first discuss non-trivial aspects of titration such as the effect of fluctuating thresholds, changes in the effective cooperativity, and the impact of titration when the repressors are degraded.

The perhaps most obvious complication is that the titration sponge not only raises the threshold but also randomizes it because it is encoded on a plasmid that fluctuates in single cells. The effect of such fluctuations is included in the simulation model where we also account for the fact that only some of the sites will have repressor bound at any given time. However, for approximate but simpler results consider a situation where the threshold S varies around an average $\langle S\rangle$, and fluctuations are so slow in time that S can be considered to be fixed during each derepression window. This creates additional variance in the decay period due to variation in the conditional average (in the sense of the Law of Total Variance). As for the discussion of uncertainty in N in the previous section, the rule for error propagation for functions of random variables immediately shows that for $\log(N/(S+1))$ the additional variance in timing is approximately given by $\sigma_S^2/\langle S\rangle^2$, which equals $1/\langle S\rangle$ if the threshold is Poisson distributed. Fluctuations in S then contribute about

[‡]specifically $N/(S+1) >> 1+b$. If the first term is 10 times larger, there will then be a 10% contribution to the variance.

as much to the variance in timing as the inherent randomness of the decay process (this is again a small noise approximation since the 'intrinsic' noise due to random decay events is $\langle 1/(S+1)\rangle$ but not $1/\langle S\rangle$). The experimentally observed improvement of the oscillations with the addition of the titration sponge still makes sense though, because it appears to raise the threshold by an order of magnitude. Intuitively, it may randomize the threshold to some extent, but it also avoids the extreme variation that comes from low thresholds, where the last few decay events greatly randomize control.

The experimental data also shows that the titration sponge only has a small effect when the repressors are degraded (§3.1.2). This is also easily explained by the model above. In fact, none of the conditions for having an effective titration sponge seem satisfied when the repressors are degraded. First, we showed above that there is an optimal threshold S for every peak abundance N that minimizes the relative standard deviation in the decay window. With repressor degradation, N is much lower because of the short repressor half-lives. Based on the numbers observed ($N \approx 600$), we believe that S is still too low to be optimal even under those conditions, but raising S only has a large effect on the noise if S is far below the optimum for a given N (Extended Data Fig. 4e). Second, it is unlikely that repressors are effectively titrated out when the repressors are tagged for degradation. Many proteins appear not to be degraded, or to be degraded much more slowly, when bound to DNA (see for example Ref.[44]). For example imagine a threshold of 50. If there are 51 repressor molecules present but only one available for degradation because the others are titrated out, then the noise in this step is the same as if there were no titration and just one repressor molecule present. Another way of looking at this is that if the repressors are degraded, taking out some of them by titration makes little difference, because they quickly re-adjust to steady state. The fact that both repressor molecules and binding sites are diluted at the same rate is key to ensure effective titration. Finally, for repressors that are actively degraded, the noise in the time needed to decay below some threshold concentration may not be dominated by the last few reaction steps, as in the model above, but by 'extrinsic' noise from fluctuating proteases. Raising the threshold to eliminate the last few steps may then have a marginal impact.

As described above, given the experimental observations of the low TetR threshold, in this context the main effect of the TetR titration sponge seems to be to simply increase the threshold. Most previous models have instead emphasized that titration (and other schemes for zero-order elimination of molecules) create ultra-sensitive switches similar to increasing the effective cooperativity (see e.g. references[50–54]). It is possible that this also helps in this system, but several system-specific features suggest it may not be the case. First, relaxation oscillators where the independent relaxation path is long compared to the time-window for switching are less sensitive to cooperativity – even if the decision to repress was gradual, it is still brief compared to the rest of the dynamics. Indeed, we see that high cooperativity is not needed for the stochastic models to oscillate with high regularity (Extended Data Fig. 4d). Second, we introduced the titration sites on plasmids that fluctuate in copy numbers, and because the binding sites are no different from the regular promoters, not all sites will be filled at the time of the derepression. Thus the titration remove a variable number of molecules, and may not increase the effective cooperativity much.

4.3.1 Simple computational model

We will now show that a simple computational model combining these features can not merely oscillate, but reproduce the amplitude, period and shape of the peaks for the estimated abundances of proteins and plasmids. In fact, we found that many versions of the model and parameter combinations can reproduce the data, and thus our purpose is not to promote one such model over others but merely to demonstrate that the types of models we have in mind are not inconsistent with the key known facts. For example, the absence of definitive data on e.g. binding affinities, repressor dimerization, and repression kinetics, prevents us from usefully modeling the switches in terms of elementary reactions. However, to make the models more realistic than perfectly sharp thresholds we assume Hill functions with reasonably small Hill coefficients (i.e. $\sim 1.5-2$). We further account for translational bursting and for fluctuations in both the repressilator plasmid and the plasmid for the titration sponge, to show that reasonably accurate oscillations (e.g. correlation coefficient after one period $> \sim 0.2$) are possible despite these sources of randomness. Specifically we assume that

$$p_{\mathrm{Tot,i}} \xrightarrow{\lambda N_{\mathrm{o}} \frac{K^n}{K^n + p_{\mathrm{Free,j}}^n}} p_{\mathrm{Tot,i}} + b$$

$$p_{\mathrm{Tot,i}} \xrightarrow{p_{\mathrm{Tot,i}}} p_{\mathrm{Tot,i}} - 1$$

where $(i, j) = (1,3), (2,1), (3,2)$. N_{o} is the number of promoters, $p_{\mathrm{Free,j}}$ is the number of unbound repressor molecules, $p_{\mathrm{Tot,j}}$ is the total number of repressor molecules and b is a geometrically distributed random variable with mean $\langle b \rangle$, representing the number of proteins produced from each mRNA, where the mRNA lifetime is assumed to be short compared to the protein lifetime, which is the case for the oscillators with a ~ 14 generation period. For the number of titration sponge plasmids N_{t} and the number of repressilator plasmids N_{o}, we simply assume

$$N_{\mathrm{o}} \xrightarrow{\langle N_{\mathrm{o}} \rangle} N_{\mathrm{o}} + 1, \quad N_{\mathrm{o}} \xrightarrow{N_{\mathrm{o}}} N_{\mathrm{o}} - 1$$

$$N_{\mathrm{t}} \xrightarrow{\langle N_{\mathrm{t}} \rangle} N_{\mathrm{t}} + 1, \quad N_{\mathrm{t}} \xrightarrow{N_{\mathrm{t}}} N_{\mathrm{t}} - 1$$

This is in fact not so different from some specific plasmid models, where self-replication and non-cooperative negative feedback combine into a constant production rate. The decay of plasmids again approximates the dilution and segregation, and thus has the same 'degradation' rate constant of 1 as the proteins.

Titration effects are harder to model due to several complications. For example, simply removing N_{t} copies from the pool of free proteins (which we found was enough to reproduce key experimental observations) does not account for the fact that repressor binding is not complete, or for the fact that the actual binding sites on the repressilator plasmid are of the same type as on the decoy plasmid. Since we do not know the binding details and only use the Hill function as a generic way to capture nonlinear effects, there is no straightforward way to model titration through elementary reactions. However, motivated by the fact that the dynamics we model are very slow, we assume that the number of proteins bound to promoter sites is in quasi-equilibrium due to fast binding and unbinding. The fraction of promoters which are unbound is $\frac{K^n}{p_{\mathrm{Free,j}}^n + K^n}$. The

number of bound repressor molecules will then be proportional to

$$N_{pro}\frac{p^n_{\mathrm{Free,j}}}{p^n_{\mathrm{Free,j}}+K^n}$$

where N_{pro} is the total number of promoters for the respective repressors

$$N_{pro}=\begin{cases}N_{\mathrm{o}}+N_{\mathrm{t}}, & j=1\\ N_{\mathrm{o}}, & j=2,3\end{cases}$$

where TetR is repressor number 1. For integer n this is consistent with modeling a single elementary reaction of n repressors binding with the promoter in a reaction that comes to equilibrium quickly, where the total number of repressor molecules would then be given as $p_{\mathrm{Free,j}}+nN_{pro}\frac{p^n_{\mathrm{Free,j}}}{p^n_{\mathrm{Free,j}}+K^n}$. However, since we use the Hill function as a generic way to account for cooperativity we instead consider

$$p_{\mathrm{Tot,j}}=p_{\mathrm{Free,j}}+2N_{pro}\frac{p^n_{\mathrm{Free,j}}}{p^n_{\mathrm{Free,j}}+K^n}$$

where the factor of two comes from the two binding sites per promoter. This equation is then solved at each step of the Gillespie algorithm[55] to update the free number of proteins. This is an approximate scheme, but one that accounts for several of the important complications, and we find that many ways of accounting for titration can explain the data.

For the simulation of Extended Data Fig. 4f, the parameters used are $\lambda = 60$, $K = (5, 10, 10)$ for the three repressors, $n = 1.5$ for all repressors, $\langle b\rangle = 10$, $\langle N_{\mathrm{o}}\rangle = 10$ $\langle N_{\mathrm{t}}\rangle = 40$. The mean period for this model is approximately 9 protein lifetimes, or 13 generations, and the CV of the cycle to cycle variation in period (e.g. phase drift) is approximately 10%, which is similar to the experimental data. Extended Data Fig. 4d in turn shows that oscillations are possible as described above, without bursts, biochemical cooperativity (n=1) titration or phase lag contribution from mRNAs (we consider infinitely fast mRNA dynamics). Deterministic linear theory would require a Hill coefficient of over 3 in each component for oscillations in this case.

References

[37] Gibson, D., Young, L. & Chuang, R. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods* **6**, 12–16 (2009).

[38] Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000).

[39] Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res* **25**, 1203–10 (1997).

[40] Nagai, T. *et al.* A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nature biotechnology* **20**, 87–90 (2002).

[41] Rosenow, M. a., Huffman, H. a., Phail, M. E. & Wachter, R. M. The crystal structure of the Y66L variant of green fluorescent protein supports a cyclization-oxidation-dehydration mechanism for chromophore maturation. *Biochemistry* **43**, 4464–72 (2004).

[42] Landgraf, D., Okumus, B., Chien, P., Baker, T. A. & Paulsson, J. Segregation of molecules at cell division reveals native protein localization. *Nat Methods* **9**, 480–482 (2012).

[43] McKenzie, G. J. & Craig, N. L. Fast, easy and efficient: site-specific insertion of transgenes into enterobacterial chromosomes using Tn7 without need for selection of the insertion event. *BMC microbiology* **6**, 39 (2006).

[44] Gora, K. G. *et al.* Regulated proteolysis of a transcription factor complex is critical to cell cycle progression in Caulobacter crescentus. *Molecular Microbiology* **87**, 1277–1289 (2013).

[45] Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S. & Elowitz, M. B. Gene regulation at the single-cell level. *Science* **307**, 1962–1965 (2005).

[46] van der Pol, B. On relaxation-oscillations. *Philosophical Magazine* **2**, 978–992 (1926).

[47] Paulsson, J. & Ehrenberg, M. Noise in a minimal regulatory network: plasmid copy number control. *Q Rev Biophys* **34**, 1–59 (2001).

[48] Paulsson, J. Summing up the noise in gene networks. *Nature* **427**, 415–418 (2004).

[49] Paulsson, J. Models of stochastic gene expression. *Physics of Life Reviews* **2**, 157–175 (2005).

[50] Goldbeter, a. & Koshland, D. E. An amplified sensitivity arising from covalent modification in biological systems. *Proceedings of the National Academy of Sciences of the United States of America* **78**, 6840–6844 (1981).

[51] Berg, O. G., Paulsson, J. & Ehrenberg, M. Fluctuations and quality of control in biological cells: zero-order ultrasensitivity reinvestigated. *Biophys J* **79**, 1228–1236 (2000).

[52] Cherry, J. L. & Adler, F. R. How to make a biological switch. *Journal of theoretical biology* **203**, 117–133 (2000).

[53] Wong, W. W., Tsai, T. Y. & Liao, J. C. Single-cell zeroth-order protein degradation enhances the robustness of synthetic oscillator. *Mol Syst Biol* **3**, 130 (2007).

[54] Buchler, N. E. & Louis, M. Molecular titration and ultrasensitivity in regulatory networks. *Journal of molecular biology* **384**, 1106–19 (2008).

[55] Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem* **81**, 2340–2361 (1977).

[56] Verdú, S. Poisson communication theory. *Invited talk, March* **25** (1999).

[57] Massey, W. A. The analysis of queues with time-varying rates for telecommunication models. *Telecommunication Systems* **21**, 173–204 (2002).

[58] Bar-David, I. Communication under the poisson regime. *Information Theory, IEEE Transactions on* **15**, 31–37 (1969).

[59] Cover, T. M. & Thomas, J. A. *Elements of information theory* (John Wiley & Sons, 2012).

[60] Bar-David, I. Information in the time of arrival of a photon packet: capacity of PPM channels. *JOSA* **63**, 166–170 (1973).

[61] Bar-David, I. Minimum-mean-square-error estimation of photon pulse delay (corresp.). *Information Theory, IEEE Transactions on* **21**, 326–330 (1975).

Appendix A Fundamental limits on the accuracy of the repressilator

In this appendix, we will derive fundamental limits on the phase drift of oscillators that includes a decay step such as the repressilator, where the time at which births are repressed and the decay starts is encoded in the subsequent decay. By using notions from information theory, we will focus on one decay step of the repressilator (e.g. see Extended Data Fig. 4c), and show that there is a bound on the estimation of when this decay step started. Since the repressilator is composed of three steps like this, this will limit the accuracy of the period, regardless of how any switching mechanism might be implemented. Note that the bound is on the variance, not the CV, as nothing is said about the mean period. Indeed, an arbitrarily complex switching mechanism might implement a local timer to obtain a delay of any length. The bound we present is a bound on the precision with which such a timer may be started. The only constraint is that this local mechanism cannot remember anything from the previous period. It is striking that, even allowing for such arbitrarily complex mechanisms between the decay steps, there is still a limit on the phase drift.

We consider an arbitrary molecular species Y, which could be a protein, with exponentially distributed lifetime 1 (i.e. we express time in units of species' lifetime), which is produced with rate N before $t = T$ and then left to degrade from that time. That is

$$y \xrightarrow{u} y + 1$$

$$y \xrightarrow{y} y - 1$$

where

$$u_T(t) = \begin{cases} N & t < T \\ 0 & \text{otherwise} \end{cases}$$

On average there will be N events in total after time T, with increasing (on average) inter-event times and these can be regarded as encoding the time T.

In the specific models above, we simply considered the distribution of times before levels reached some threshold value, and showed that the heterogeneity in the last few steps then can dominate the error. That means the threshold must be set high but not too high (see section 4.2.2), which in turn means that the control system ignores the information in the part of the trace that falls below the threshold. Here we instead consider the problem of estimating the time T, at which births stop, from a knowledge of all the death events without limiting the controller to threshold functions. We thus allow for control systems that can account for the fact that the process slows down. Specifically we will use techniques from information theory (Poisson communications theory[56]) to show that it is impossible to recover the time T with a mean square error better than $1/N$.

To turn this problem into a communications problem, first note that the death events themselves form a Poisson process. This is a result of a version of Burke's theorem which holds for the special case of an M/G/∞ queue with a time varying arrival rate (Ref.[57] Theorem 2.5, part 3). To see that the statement is true, consider sampling the death events as follows: throw points onto the real line according to the birth rate u_T above, one by one, and immediately shift each

point by an independent exponentially distributed amount. The probability of finding a point in a small interval is then the convolution of the nonhomogenous birth rate and the exponential waiting time distribution and is independent of the final resting place of other points. Therefore, the corresponding process is the Poisson counting process

$$x \xrightarrow{\lambda_T(t)} x + 1 \tag{A-1}$$

where

$$\lambda_T(t) = \int_{-\infty}^{t} u_T(\tau) \exp(-(t-\tau))\, d\tau = \begin{cases} \displaystyle\int_{-\infty}^{t} N \exp(\tau - t)\, d\tau = N, & t < T \\ \displaystyle\int_{-\infty}^{T} N \exp(\tau - t)\, d\tau = N e^{-(t-T)}, & t \geq T \end{cases} \tag{A-2}$$

which has precisely the same statistics as the deaths of Y. The times at which Y molecules are eliminated have the same statistics as the times at which x is incremented. An example is given in Figure S1, where the upper diagram shows the birth death process whose deaths were used to provide the Poisson points in the lower plot in Figure S1, with the red lines denoting the event times.

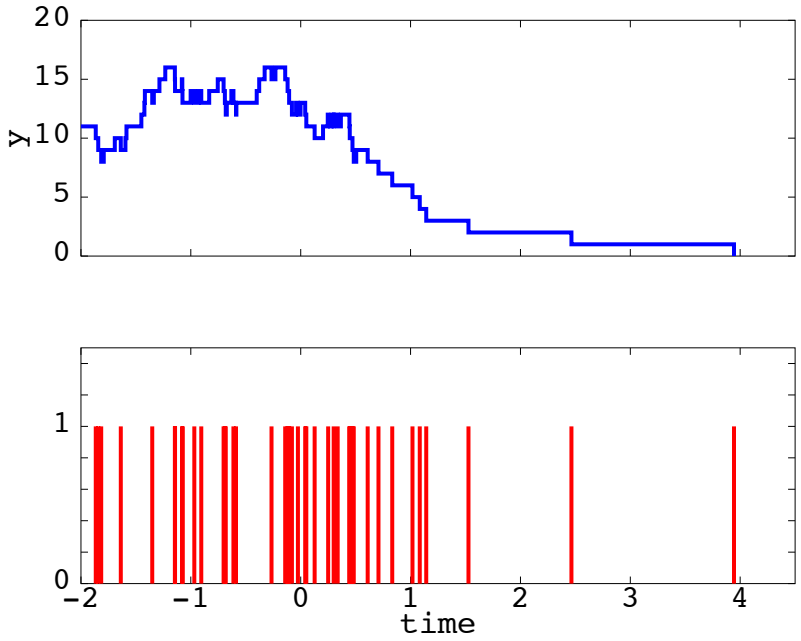

Figure S1: Birth death process (upper panel) with corresponding death events (lower panel). In the upper plot, the birth rate is constant up to a time T (see text) at which point the birth process is switched off and the number of Y molecules allowed to decay. The task is to estimate that time knowing only the death events in the lower plot and the average number of molecules.

Consider now the problem of determining the unknown time T from these event times, which we denote $\{\tau_i\}$. In other words, we imagine that the task of an optimal control system is to

determine when the production was stopped knowing the death times of all molecules and the expected number of molecules at the start of the decay phase (N). For this problem, an estimator is a function $f(\{\tau_i\})$ that maps the set of event times to an estimate of the time T at which the decay started, for one sampling of the Poisson process (A-1). We will derive two estimators, the maximum likelihood (ML) estimator (f_{ML}), for asymptotic results, and the minimum mean square error (MMSE) estimator (f_{MMSE}), for results for finite N. For any N, the MMSE estimator is, by definition, optimal, in that it gives an estimate of the time T with minimum variance about the true value. Calculating this optimal estimate, for a given set of event times $\{\tau_i\}$, is computationally intensive, but feasible for the discrete processes considered here as we shall demonstrate. The ML estimator (f_{ML}) simply returns the most likely T, that is the T for which the probability of observing the given events is highest. The ML estimator for this problem has a simple structure and, whilst it underperforms the optimal MMSE estimator for all values of N, it does approach optimality asymptotically for large N and gives considerable insight into the optimal solution in that limit. Both of these estimators are written in terms of the *likelihood* function $p(\{\tau_i\}|T)$ which is, roughly speaking, the probability of observing the events $\{\tau_i\}$ given the time T.[§]

The $\lambda_T()$ defined in (A-2) has an infinite area, since it is constant for all $t < T$. If instead we replaced $\lambda_T()$ in (A-2) with a pulse $\nu(t)$ of finite area $Q = \int_{-\infty}^{\infty} \nu(t)\,dt$ then the likelihood $p(\{\tau_i\}|T)$ would be given by

$$p(\{\tau_i\}|T) = \exp(-Q) \prod_{k=1}^{K} \lambda_T(\tau_k),$$

with K the total number of events[58]. However, in our case, the area of λ_T is not bounded. Nevertheless, we can regularise the problem, without changing the optimality properties, by setting λ_T to 0 before $t = -X$, where X is a large constant, i.e.

$$\hat{\lambda}_T(t) = \begin{cases} 0 & t < -X \\ N & -X \le t < T \\ Ne^{-(t-T)} & t \ge T \end{cases}$$

Importantly, we shall see that the results do not depend on the choice of X as long as it is sufficiently large. Intuitively, X should be large enough such that at $t = T$ the distribution has reached equilibrium. For this choice of λ_T, it follows that

$$Q = \int_{-\infty}^{\infty} \hat{\lambda}_T(t)\,dt = N(T + X) + N$$

and hence, for $T > -X$, that the likelihood is given by

$$\begin{aligned} p(\{\tau_i\}|T) &= e^{-N(T+X+1)} N^K \prod_{\tau_i > T} e^{-(\tau_i - T)} \\ &\propto e^{-NT} \prod_{\tau_i > T} e^{-(\tau_i - T)} \end{aligned}$$

[§]Strictly speaking, it is the probability distribution function of the event times.

(since N, X and K are constants). Thus, the choice of X, when sufficiently large, only affects the constant of proportionality, which doesn't affect the results in any way.

The first estimator we consider is the ML estimator, $f_{ML}(\{\tau_i\}) = \arg\max_T p(\{\tau_i\}|T)$, that is the estimator that returns the value of T for which the observed sequence is most likely. Note that

$$\log p(\{\tau_i\}|T) = \text{const} - NT + \sum_{\tau_i > T} (T - \tau_i)$$

and so

$$\frac{d \log p(\{\tau_i\}|T)}{dT} = -N + \#\text{ of events after } T. \tag{A-3}$$

Thus the gradient is zero and the likelihood function constant at its maximum value for any T between the Nth last and $(N+1)$th last event time. Thus f_{ML} is ill defined, as any T in that range maximises the likelihood. Any estimator returning a T in that range will enjoy the asymptotic optimality associated with the ML estimator. For convenience, we define as the ML estimator the one that returns the Nth last event time, that is the T that results from counting up to N from the last death event. No unbiased estimator can do better that the Cramér-Rao lower bound $1/\mathcal{I}(T)$ where $\mathcal{I}(T)$ is the Fisher information (Theorem 11.10.1, Ref.[59]). If we impose the constraint that all estimators we consider are time invariant (i.e. simply the fact that if all the death events are shifted by t, the estimate of T will also be shifted by t) , that is

$$f(\{\tau_i - t\}) = f(\{\tau_i\}) - t \tag{A-4}$$

then any biased estimator can be improved by subtracting out the bias, which must necessarily not depend on T. Consequently, the lower bound $1/\mathcal{I}(T)$ is a bound on the mean squared error of *any* estimator satisfying (A-4). The Fisher information can be calculated directly from its definition as

$$I(T) = E\left(\left(\frac{d \log p(\{\tau_i\}|T)}{dT}\right)^2 \middle| T\right) = N$$

using (A-3) and noting that the number of events after time T is Poisson distributed with mean N. It follows that

$$E\left((f(\{\tau_i\}) - T)^2 | T\right) \geq \frac{1}{N}$$

for any estimator $f()$ satisfying (A-4). That is, any time-invariant estimator will have a variance greater or equal to $1/N$. Subject to certain regularity assumptions, the ML estimator is always asymptotically efficient, in that for large N its variance approaches the Cramér-Rao lower bound. In our case, it is relatively straightforward to show directly that the estimator described achieves the bound $1/N$, the utility of the lower bound being that it shows that this is optimal.[¶]

[¶]As described, The ML estimator described above returns the N^{th} last death event. This may have occurred before or after time $t = T$. If before, then there will have been $N - L$ subsequent death events, where L is the number of molecules present at time T, and we need to count back over $N - L$ events, each of which has mean $1/N$ and variance $1/N^2$. The resulting error would have mean $(N - L)/N$ and variance $E_L\left((N - L)/N^2\right)$. If the N^{th} event occurred after $t = T$, we need to count forward over events which are occurring with a decreasing rate. However, the decrease in rate becomes negligible as N increases and the events occur over a shorter and shorter time. In this case, the mean of the error would be $\sum_{k=N}^{L} 1/k \approx (L - N)/N$ and the variance $\sum_{k=N}^{L} 1/k^2 \approx (L - N)/N^2$. By the law of total variance,

To investigate how tight is the bound for finite N we now consider the MMSE estimator, i.e. the estimator with the minimum resulting variance in the estimates of T, which is given by the conditional mean $E(T|\{\tau_i\})$. This has the form

$$f_{MMSE}(\{\tau_i\}) = E(T|\{\tau_i\}) = \frac{\int T p(\{\tau_i\}|T) p(T)\, dT}{\int p(\{\tau_i\}|T) p(T)\, dT} \tag{A-5}$$

where $p(T)$ is the prior distribution of T, which we shall set to be uniform on $[-X, X]$, i.e.

$$p(T) = \begin{cases} \frac{1}{2X} & |T| < X \\ 0 & \text{otherwise} \end{cases}$$

(again, the precise value of X, which matches that in the definition of λ_T, is unimportant when sufficiently large). For the example in Figure S1 the time T, at which the exponential decay in rate commences, is actually 0.5. The optimal estimator (A-5) returns an estimate of 0.57.

From (A-3) it is clear that the logarithm of $p(\{\tau_i\})$ is piecewise linear, and so it is possible to do the piecewise integrations in (A-5) analytically, resulting in a computationally efficient means of calculating the MMSE estimator. Together with using the exact Gillespie algorithm[55] for sampling the event times, this provides an efficient means of determining $E\left((f_{MMSE}(\{\tau_i\}) - T)^2\right)$ using Monte Carlo integration. Figure S2 shows the computed variance approaching its asymptotic value of $1/N$. The upper curve is thus an achievable lower bound on the performance of *any* estimator. For example, when $N = 10$ the variance is approximately $1.20/N$ and when $N = 100$ it is approximately $1.06/N$.

So, it is possible to quantify the amount of timing information encoded in the *elimination* of this molecular species (or, indeed, any process coupled to this). As discussed above, the minimum mean square estimate has a variance no lower than $1/N$. That is, having encoded the timing of the repression of birth events of Y into the deaths of Y, it is impossible to recover the time at which that occurred with accuracy better than this. The repressilator involves 3 such timing events per cycle. If we assume that the repressor distributions reach their equilibrium values before commencing their decay then the variance in the period of the repressilator can be no lower than $3/N$. Any heterogeneity in the buildup phases would only make this worse. The exact derivation above is limited to purely Poisson processes, which precludes the consideration of bursting. Note though that the $1/N$ timing variance we describe, though fundamental, can also be understood in terms of the initial uncertainty in the number of proteins when production is turned off (as is clear from the derivation of the ML estimate and the Cramér-Rao lower bound above). Indeed, the calculation of the variance of the ML estimator (footnote ¶ on page 25) showed that, asymptotically, this approached σ_L^2/N^2, where L is the number of molecules at the time production is switched off, in agreement with the approximations in Section 4.2.3. As stated there, for geometrically distributed bursts of size b the lower bound thus increases to approximately $(1 + \langle b \rangle)/N$.

---

the variance of the error is then $E_L(|N - L|/N^2) + E_L\left((N - L)^2/N^2\right)$. Since L is Poisson, with mean and variance N then the second term is $1/N$. For large N, the first term approaches $\sqrt{2/\pi}/N^{3/2}$ (from results on folded Gaussian distributions) and is dominated by the second. Thus, asymptotically, the variance in f_{ML} is $1/N$ and, asymptotically, no estimator can do better.

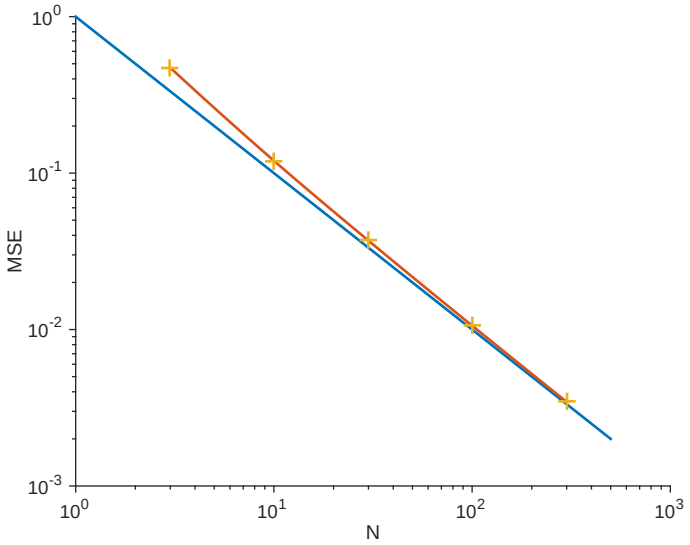

Figure S2: Minimum Mean Square Error of optimal estimator, for given N, (red) vs Asymptotic Lower Bound (blue), showing that the asymptotic bound remains reasonably tight for small N.

We have analysed this example because of its simplicity and its relevance to the repressilator. However, the conclusions do not depend on detailed assumptions. In fact for *any* smooth[‖] $\lambda(t)$ with finite support the asymptotic lower bound on the variance is inversely proportional to N (Ref.[58]), and this result can be extended to more general $\lambda(t)$ in the same manner as we have proceeded above. For example, if the repression switch is not perfect, or if the protein decay rate changes over time, a similar bound will apply, just differing in the constant of proportionality. It is important to emphasise that this is a bound on the variance, not the CV. So it doesn't preclude the CV, or phase shift per cycle, being reduced by increasing the period without increasing N. However, for any given period, the finite number of molecules will limit the accuracy of the oscillations.

[‖]For a bounded pulse $\lambda(t) = Na^2(t)$ where $\int_{-\infty}^{\infty} a^2(t)\,dt = 1$, so the average number of events is N, then the lower bound on the variance is $\frac{1}{4N\int_{-\infty}^{\infty}(\frac{da(t)}{dt})^2\,dt}$ (Ref.[58]). However, if λ is discontinuous (or, to be more precise, if the gradient of λ fails to be square integrable, then the performance improves to $1/N^2$ (Refs.[60,61]). If λ varies quickly, although in a way that allows the integral above to be calculated, then the performance would look like $1/N^2$ for small N but $1/N$ for sufficiently large N, as the number of events during the time of the transition increases. We would argue, though, that sharp changes in rate do not happen in biological circuits, and so these $1/N^2$ results, though interesting, are not relevant here.