Quantifying and correcting estimation bias in transcriptional parameter inference from single-cell data

Ramon Grima*
School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JH, United Kingdom

Pierre-Marie Esmenjaud
Biology department, Ecole polytechnique, Institut Polytechnique de Paris, 91120 Palaiseau, France
(Dated: June 28, 2023)

The snapshot distribution of mRNA counts per cell can be measured using single molecule FISH or single-cell RNA sequencing. These distributions are often fit to the steady-state distribution of the two-state telegraph model to estimate the three transcriptional parameters for a gene of interest: mRNA synthesis rate, the switching on rate (the on state being the active transcriptional state) and the switching off rate. This model assumes no extrinsic noise, i.e. parameters do not vary between cells, and thus estimated parameters are to be understood as approximating the average values in a population. The accuracy of this approximation is currently unclear. Here we develop a theory that explains the size and sign of estimation bias when inferring parameters from single-cell data using the standard telegraph model. We find specific bias signatures depending on the source of extrinsic noise (which parameter is most variable across cells) and the mode of transcriptional activity. If gene expression is not bursty then the population averages of all three parameters are overestimated if extrinsic noise is in the synthesis rate; underestimation occurs if extrinsic noise is in the switching on rate; both underestimation and overestimation can occur if extrinsic noise is in the switching off rate. We find that some estimated parameters tend to infinity as the size of extrinsic noise approaches a critical threshold. In contrast when gene expression is bursty, we find that in all cases, the mean burst size (ratio of the synthesis rate to the switching off rate) is overestimated while the mean burst frequency (the switching on rate) is underestimated. We estimate the size of extrinsic noise from the covariance matrix of sequencing data and use together with our theory to correct published estimates of transcriptional parameters for mammalian genes.

I. INTRODUCTION

The two-state model of gene expression [1, 2], also called the telegraph model, provides a simple framework describing the stochastic nature of gene expression. The model consists of four effective reactions:

$$\begin{aligned} &G_{\rm ON} \xrightarrow{\alpha} G_{\rm ON} + M, \\ &G_{\rm OFF} \xrightarrow{\beta} G_{\rm ON}, \\ &G_{\rm ON} \xrightarrow{\theta} G_{\rm OFF}, \\ &M \xrightarrow{d} \emptyset. \end{aligned} \tag{1}$$

These reactions describe mRNA (M) synthesis in the active promoter state $G_{\rm ON}$ with rate α, switching from the inactive promoter state $G_{\rm OFF}$ to the active state with rate β, switching from the active to the inactive promoter state with rate θ and mRNA degradation with rate d respectively. Assuming Markovian dynamics, the telegraph model can be solved in steady-state leading to an analytical marginal distribution of mRNA counts [1].

It has been shown that the transcript count distributions for many genes, obtained from a snapshot population measurement using smFISH (single molecule fluorescence in situ hybridization) or scRNA-seq (single cell RNA sequencing), can be well fit by the steady-state distributions predicted by the telegraph model [3–5]. This matching also simultaneously leads to estimates of the three main transcriptional parameters normalised by the degradation rate, i.e. α/d, β/d and θ/d. Note that without an explicit measurement of the mRNA degradation rate d, it is not possible to obtain the absolute values of the other parameters. Estimation of these parameters provides useful insight into the mode of transcriptional activity and it is relationship with core promoter elements (such as gene length, TATA and initiator elements)[6]. For example a switching on rate that is much larger than a switching off rate would indicate

* Ramon.Grima@ed.ac.uk

that expression occurs almost incessantly, whereas the reverse condition together with a large synthesis rate suggests expression occurs in large infrequent bursts.

Despite it is simplicity and utility, there are doubts about the telegraph model's ability to capture the major biochemical processes that influence the observed stochasticity in mRNA numbers [7–12]. Gene expression is known to be affected by various processes that are not explicitly modelled by the telegraph model, e.g. cell cycle dynamics [4, 13], coordination of transcription with cell size [14, 15], mitochondrial dynamics [16, 17] and Ca^{2+} signalling [18, 19]. These various processes (and others) can be grouped together under the umbrella of extrinsic noise, since they arise independently of a gene of interest but act on it. Some studies have concluded that for many mammalian genes, the variability of cytoplasmic transcript abundance is determined primarily by the phenotypic cellular state [19, 20]. Thus correcting data for extrinsic noise before it is input to an inference framework based on the telegraph model is important since potentially this type of noise could significantly bias parameter estimation.

In principle, the dual reporter method enables discrimination between extrinsic noise and intrinsic noise (locally determined by the gene sequence and which is modelled by the telegraph model) [21]. This has been done for some organisms using fluorescent protein reporters [22–24] but the experiments are technically challenging and they do not directly enable the study of transcription since mRNA and protein are not generally strongly correlated [25]. There are also parameter inference methods that instead of fitting to population data, they fit to individual cells using single-cell time-lapse protein data [26, 27]. However live-cell imaging approaches which allow visualization of mRNA bursts as they occur in living cells are challenging since they are low-throughput and require genome-editing [28, 29].

An alternative approach to correct the parameter estimates for noise sources not taken into account by the telegraph model is to directly eliminate these sources from the data. For example, to correct for the fact that the telegraph model assumes one gene copy, whereas in reality some cells will have double the gene copy number than others, we can separate the data according to cell-cycle phase. Then we can fit a telegraph model to single allele data from cells in G1 and separately fit the distribution obtained from a convolution of the distributions of two telegraph models to double allele data from cells in G2 [13] (under the simple assumption that copies are independent of each other). This procedure has the added benefit of accounting for cell-cycle dependence of transcriptional parameters [4, 30, 31] which is not described by the telegraph model. While it is still not widespread, this type of correction can nowadays be readily performed since cell-cycle phase classification is possible for both smFISH data (most of the experiments already include a DNA-labelled channel) [13] and scRNA-seq data (deep learning applied to the unspliced-spliced RNA space of cell cycle-related genes) [32].

However these corrections do not go far enough because there are additional noise sources. For example, different cells within the same cell-cycle phase may have different effective synthesis rates of cytoplasmic mRNA because of the coupling of gene expression to cell size for some eukaryotic genes [14, 19] and cell-to-cell variation in the nuclear export rate [33]. While some of these sources could be accounted for by carefully designed experiments to measure extrinsic noise sources of choice (such as cell volume) the fact is that presently we are far from having a complete understanding of which are the most relevant sources. Hence it stands to reason that after any correction we may do to partition the data to exclude one noise source or another, inevitably there will still be cell-to-cell variability in the transcriptional parameters due to unaccounted noise sources [34–38] and fitting the telegraph model to this data may lead to biased estimates. A summary of the standard fitting procedure is illustrated in Fig. 1. *A main question of interest that remains unresolved is: what type (positive / negative) and extent of bias in parameter estimation should we expect when fitting the telegraph model to experimental data?* In the best case scenario, one hopes that the inferred parameters are to a good approximation equal to their true values, i.e. the parameter values averaged over all observed cells.

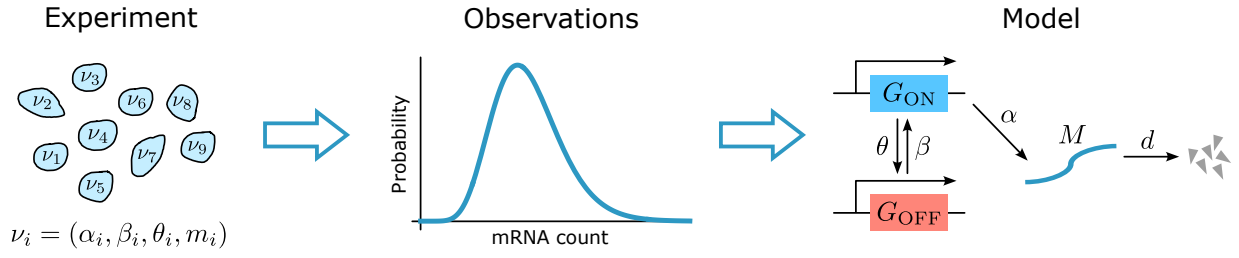

FIG. 1. Illustration of the standard transcriptional parameter inference procedure. In a population of cells, the transcriptional rates for a particular gene vary from cell to cell: α_i, β_i and θ_i are the mRNA synthesis rate, the switching on rate and the switching off rate for cell i, respectively. It is assumed that the degradation rate does not vary. The mRNA count in cell i is denoted by m_i. In a typical inference setup, the steady-state distribution of mRNA (M) per cell is measured and then by fitting this to the analytical steady-state distribution of the standard telegraph model, the three transcriptional rates (normalised by the degradation rate) for the gene are estimated. Since the telegraph model assumes the parameters are the same for each cell, it is unclear what is the relationship of the inferred parameters to the true population averaged parameters.

To address this question, here we develop a moment-based analytical approach to understand how the size and sign of estimation bias depends on the mode of transcriptional activity, the source of extrinsic noise (which parameter varies most from cell-to-cell) and the type of distribution that describes it. We identify specific "bias signatures" which we confirm by inferring parameters, using the method of maximum likelihood, from simulated single cell data. In particular, we show that while the telegraph model distributions can well fit those computed from data, the inferred parameters from the telegraph model can be far away from their true population average. We also use the theory to correct published telegraph model estimates of the mammalian transcriptome using scRNA-seq data [6].

II. THE SOURCE OF EXTRINSIC NOISE AND TIMESCALE SEPARATION DETERMINE THE SIZE AND SIGN OF ESTIMATION BIAS

Generally all of the parameters of the telegraph model may vary from cell to cell and their spread is described by some arbitrary joint distribution. However because of it is generality this is difficult to study and hence we make some simplifying assumptions. We consider three different cases: (i) variability only in the synthesis rate ρ; (ii) variability only in the switching on rate k_{on}; (iii) variability only in the switching off rate k_{on}. Note that we have given the rate parameters different labels than in the reaction scheme (1) to distinguish them from those in the standard telegraph model where there is no parameter variability between cells. Note that we have not considered variability in the mRNA degradation from one cell to another, because this is difficult to study – since the telegraph model solution is a function of the three transcriptional rates normalised by the degradation rate, variability in the degradation rate would imply variability in all three inferred rates simultaneously. In all cases the distribution of each parameter (normalised by the degradation rate) is assumed to be a gamma distribution – this distribution is ideal because it is defined on the positive real line and its mean and variance can be tuned independently. The three circuit is are illustrated on the left hand side of Fig. 2. Our question is: if for a gene of interest the distribution of mRNA counts per cell for each of these three cases is fitted to that of the standard telegraph model, how are the parameters inferred related to the true population averaged parameters? For example if the mRNA distribution comes from a telegraph model where the varying parameter is the synthesis rate then it would be of interest to understand what is the relationship of the population averaged synthesis rate to the synthesis rate inferred from fitting the data to the standard telegraph model.

To answer this question one must precisely define what is meant by fitting the data generated by the telegraph model with extrinsic noise by the standard telegraph model. For the time being, we shall define a matching of the two models to mean that we have found a set of parameter values of the standard telegraph model for which the first three moments of the mRNA counts of this model matches the first three moments of the mRNA counts of the telegraph model with extrinsic noise. Note that we cannot match fewer or more moments than three since we need to solve three equations to exactly determine the values of three unknowns. Let the ith factorial moments of the mRNA counts as given by the standard telegraph model be R_i^T and those of the telegraph model with extrinsic noise be R_i; then the matching procedure is equivalent to setting $R_i^T = R_i$ for $i = 1, 2, 3$. Since R_i^T are functions of the telegraph model parameters $\{\alpha, \beta, \theta\}$, we have three simultaneous equations that can solved for these parameters, leading to

$$\begin{aligned}
\alpha &= \frac{\left(2R_1^2 - R_2\right) R_3 - R_1 R_2^2}{\left(R_1^2 - 2R_2\right) R_2 + R_1 R_3}, \\
\beta &= \frac{2R_1 \left(R_2^2 - R_1 R_3\right)}{\left(2R_1^2 - R_2\right) R_3 - R_1 R_2^2}, \\
\theta &= \frac{2\left(R_1^2 - R_2\right)\left(R_1 R_2 - R_3\right)\left(R_1 R_3 - R_2^2\right)}{\left(\left(R_1^2 - 2R_2\right) R_2 + R_1 R_3\right)\left(\left(2R_1^2 - R_2\right) R_3 - R_1 R_2^2\right)}.
\end{aligned} \tag{2}$$

Exact expressions for R_i for each of the three cases of extrinsic noise (in the synthesis rate, the switching rate or the switching off rate) are derived in Appendix A. Another parameter of interest is the burst size defined as $\lambda = \alpha/\theta$ which is equal to the mean number of transcripts produced while the promoter is in the active state.

To understand systematic bias introduced by the neglection of extrinsic noise in the inference framework, we devise the following procedure: (i) we generated a million random parameter sets with the constraints that the coefficient of variation (CV defined as the standard deviation divided by the mean) of the gamma distribution is less than 1 and the mean of the three transcriptional parameters is in the range $0.1 - 200$ (values normalised by the mRNA degradation rate). The range of these rates covers the transcriptional parameters estimated by previous studies for mouse and human fibroblasts [7]; (ii) for each parameter set, we inferred the effective parameters using Eq. (2). We discarded parameter estimates which were negative since these are non-physical. The statistics of the positive inferred parameters (normalised by the ground truth values) are shown by box-whisker plots in the middle column of Fig. 2.

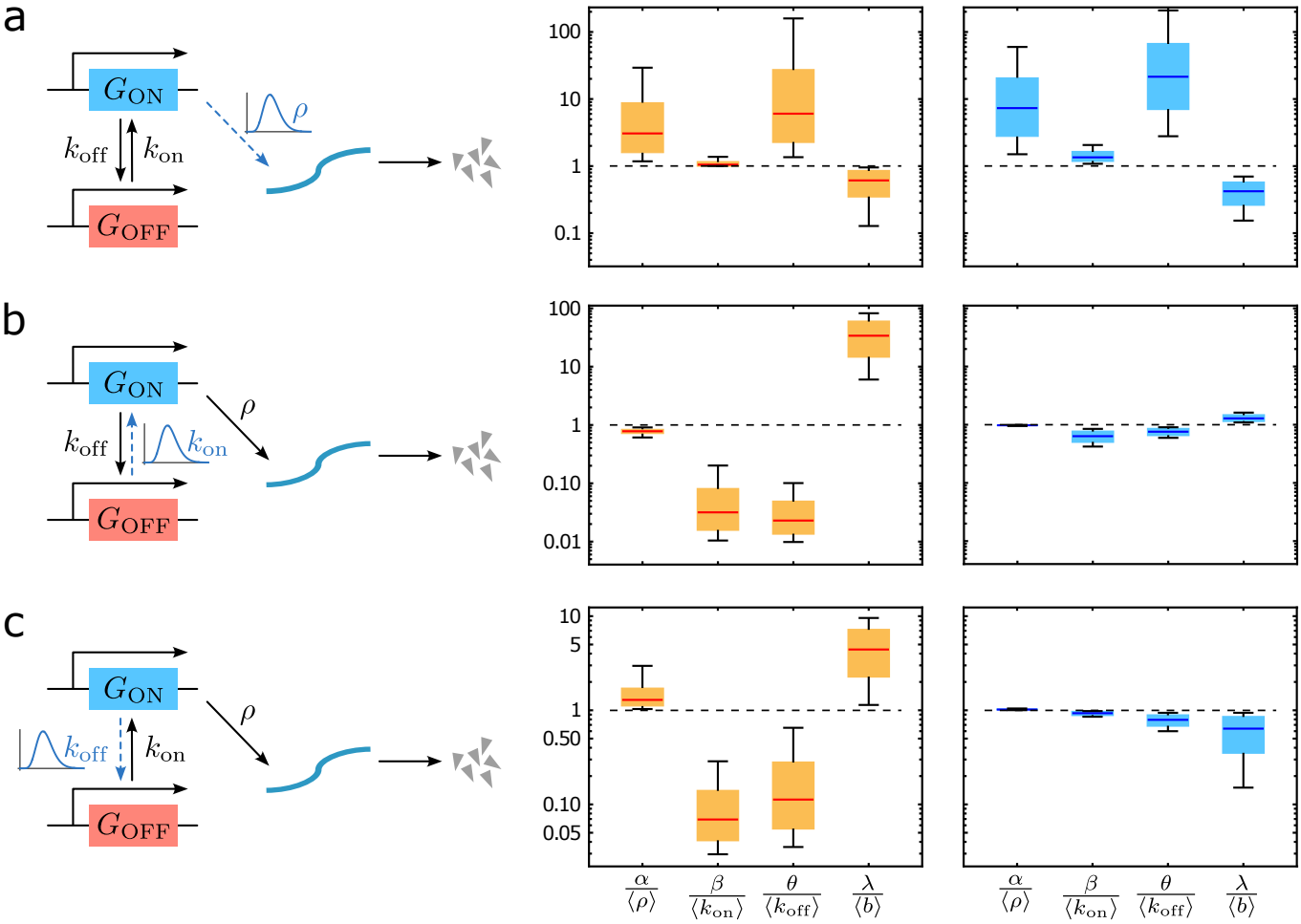

FIG. 2. Systematic bias in parameter estimation due to unaccounted extrinsic noise sources. We consider cell populations with (a) variable synthesis rates, (b) variable switching on rates and (c) variable switching off rates. The variability in a parameter is modelled by a gamma distribution. For each case, we generate a million random parameter sets and estimate the effective parameters of the standard telegraph model (reaction scheme (1)) by exactly matching the first three moments of this model to those of telegraph model with extrinsic noise. For those parameters sets where all inferred parameters are positive, the statistics of their values normalised by the ground truth values (the population averaged parameters) are shown by box-whisker plots in the middle column of the figure (lower and upper fences show the 10th and 90th percentiles, respectively; the yellow bar shows the interquartile range; median is shown by the horizontal red line). The horizontal black dashed line is for reference showing perfectly accurate estimation. The numerical experiments are repeated with the constraint that promoter switching is slow and the results are shown by box-whisker plots in the right column. The plots show that the source of extrinsic noise and timescale separation determine the size and sign of estimation bias. See main text for details.

The ground truth values are the population averaged parameters in the telegraph model with extrinsic noise. We shall use the angled brackets $\langle\rangle$ to denote the average over a distribution of parameters. For noise in the synthesis rates, the ground truth values are the mean synthesis rate $\langle\rho\rangle$ and the mean burst size $\langle b\rangle = \langle\rho\rangle/k_{\text{off}}$. For noise in the switching on rate, the ground truth value is the mean switching on rate $\langle k_{\text{on}}\rangle$. For noise in the switching off rate, the ground truth values are the mean switching off rate $\langle k_{\text{off}}\rangle$ and the mean burst size $\langle b\rangle = \langle\rho/k_{\text{off}}\rangle = \langle k_{\text{off}}\rangle\rho/(\langle k_{\text{off}}\rangle^2 - \sigma^2_{k_{\text{off}}})$ – the latter result was derived by integrating the burst size over a gamma distribution with mean $\langle k_{\text{off}}\rangle$ and variance $\sigma^2_{k_{\text{off}}}$ (under the assumption that $\sigma^2_{k_{\text{off}}} < \langle k_{\text{off}}\rangle$ as required by one of our constraints above).

The main conclusions of our numerical experiments are as follows. Extrinsic noise in the synthesis rate causes an overestimation of the synthesis rate and switching rates but an underestimation of the average burst size (Fig. 2a). Extrinsic noise in the switching on rate leads to the opposite (Fig. 2b). Extrinsic noise in the switching off rate leads to an overestimation of the synthesis rate and the burst size and an underestimation of the switching rates (Fig. 2c). These results show that depending on the source of extrinsic noise, there is a specific "bias signature". Clearly the noise in one parameter causes estimation bias not only in that parameter but in all the others as well. Another difference distinguishing the circuit is based on the source of extrinsic noise is the fraction of estimated parameter

sets that are positive: merely 0.7% for extrinsic noise in the synthesis rate, 100% for extrinsic noise in the switching on rate and 82% for extrinsic noise in the switching off rate. We note that negative estimated parameter values occur when it is not possible to match the first three moments of mRNA counts of the telegraph model with extrinsic noise with those of the standard telegraph model – in these cases, typically it is possible to only match a smaller number of moments (we study this issue in more depth later).

We repeated the numerical experiment but for the case of slow promoter switching where the synthesis rate is in the range $0.1 - 200$ while the switching on and off rates are in the range $0.1 - 1$. The estimated rates as ratios of the ground truth parameters are shown by box-whisker plots in the right column of Fig. 2. By comparison to the previous numerical experiments where all the rates were in the range $0.1 - 200$, it is clear that slow promoter switching has little effect on estimation bias when extrinsic noise is in the synthesis rate but it significantly diminishes estimation bias when extrinsic noise is in the switching rates; in the latter cases the sign of the bias is unchanged except for the burst size which was overestimated previously and now the reverse occurs. In the slow promoter switching regime, inaccuracy in parameter estimation principally stems from extrinsic noise in the synthesis rates. Interestingly we also find a change in the fraction of parameter sets leading to positive estimates: 86%, 100% and 100% for extrinsic noise in the synthesis, switching on and switching off rates, respectively.

Our inference procedure is based on precise moment-matching between models with extrinsic and without extrinsic noise. Previous studies have noted that for certain models of gene expression, matching of the moments to those of experimental data gives different results than matching of the mRNA count distributions [39]. However we find this is not the case for our inference procedure: matching of the first three moments of the standard telegraph model to those of the telegraph model with extrinsic noise, when it resulted in positive parameter estimates, also led to a good match between the distributions of the two types of models. In Fig. 3(a)-(c) we show three specific examples of the excellent fit is, one for each source of extrinsic noise. These are not isolated cases – distribution matching over large swathes of parameters space is confirmed by the high linearity of the P-P probability plots in Fig. 3(d)-(f). Hence we conclude that the results shown in Fig. 1 mirror those that would be obtained by the much more laborious process of matching the distributions of mRNA counts of the standard telegraph model with that from the telegraph model with extrinsic noise. Incidentally this suggests that a model selection algorithm would likely favour the telegraph model since this model has the smaller number of parameters and yet matches the count distributions as well as the extrinsic noise model used to generate the data.

The results derived thus far assume that the distribution of parameters across cells is well approximated by a gamma distribution. In Appendix C we derive results for any distribution of positive-valued parameters provided extrinsic noise is small, i.e. the coefficient of variation of the parameter distributions is small. Surprisingly, we find that when it can be analytically determined, the sign of the estimation bias is the same as in Fig. 2 thus verifying the presence of a universal "bias signature" depending on which rate is the most variable amongst cells.

III. MASSIVE OVERESTIMATION OF PARAMETERS NEAR A CRITICAL EXTRINSIC NOISE THRESHOLD

We now study in detail the case where the gamma-distributed extrinsic noise is present in the synthesis rate. This case is particularly interesting because: (i) it is the only case where Eq. (2) leads to equations for the inferred parameters that are amenable to analytical analysis (the other two cases are in terms of the exponential integral); (ii) this type of noise is the only one which leads to significant estimation bias independent of the speed of promoter switching (Fig. 2); (iii) if promoter switching is not slow, this type of noise is the only one which leads to many instances where the inferred parameters are negative and hence physically meaningless; (iv) the dependence of the synthesis rate on the cell volume has recently been shown to be a major source of extrinsic noise for many eukaryotic genes [14, 15, 19].

Using the expressions derived for the factorial moments of mRNA counts for gamma extrinsic noise in the synthesis rate with mean $\langle\rho\rangle$ and variance σ_ρ^2 in Appendix A, we find that Eq. (2) simplifies to

$$\begin{aligned}
\alpha &= \frac{\langle\rho\rangle\left(xk_{\text{off}}\left(k_{\text{on}}+3\right)+k_{\text{off}}\left(k_{\text{off}}+k_{\text{on}}\right)+x^2\left(k_{\text{on}}+1\right)\left(k_{\text{on}}+2\right)\right)}{(1-x)k_{\text{off}}\left(k_{\text{off}}+k_{\text{on}}\right)},\\
\beta &= \frac{xk_{\text{on}}\left(k_{\text{off}}\left(k_{\text{on}}+3\right)+\left(k_{\text{on}}+1\right)\left(k_{\text{on}}+2\right)\right)+k_{\text{off}}k_{\text{on}}\left(k_{\text{off}}+k_{\text{on}}\right)}{xk_{\text{off}}\left(k_{\text{on}}+3\right)+k_{\text{off}}\left(k_{\text{off}}+k_{\text{on}}\right)+x^2\left(k_{\text{on}}+1\right)\left(k_{\text{on}}+2\right)},\\
\theta &= \frac{\left(k_{\text{off}}+xk_{\text{on}}+x\right)\left(k_{\text{off}}+x\left(k_{\text{on}}+2\right)\right)\left(\left(k_{\text{off}}+k_{\text{on}}\right)\left(k_{\text{off}}+x\left(k_{\text{on}}+3\right)\right)+2x\right)}{(1-x)k_{\text{off}}\left(xk_{\text{off}}\left(k_{\text{on}}+3\right)+k_{\text{off}}\left(k_{\text{off}}+k_{\text{on}}\right)+x^2\left(k_{\text{on}}+1\right)\left(k_{\text{on}}+2\right)\right)},
\end{aligned} \tag{3}$$

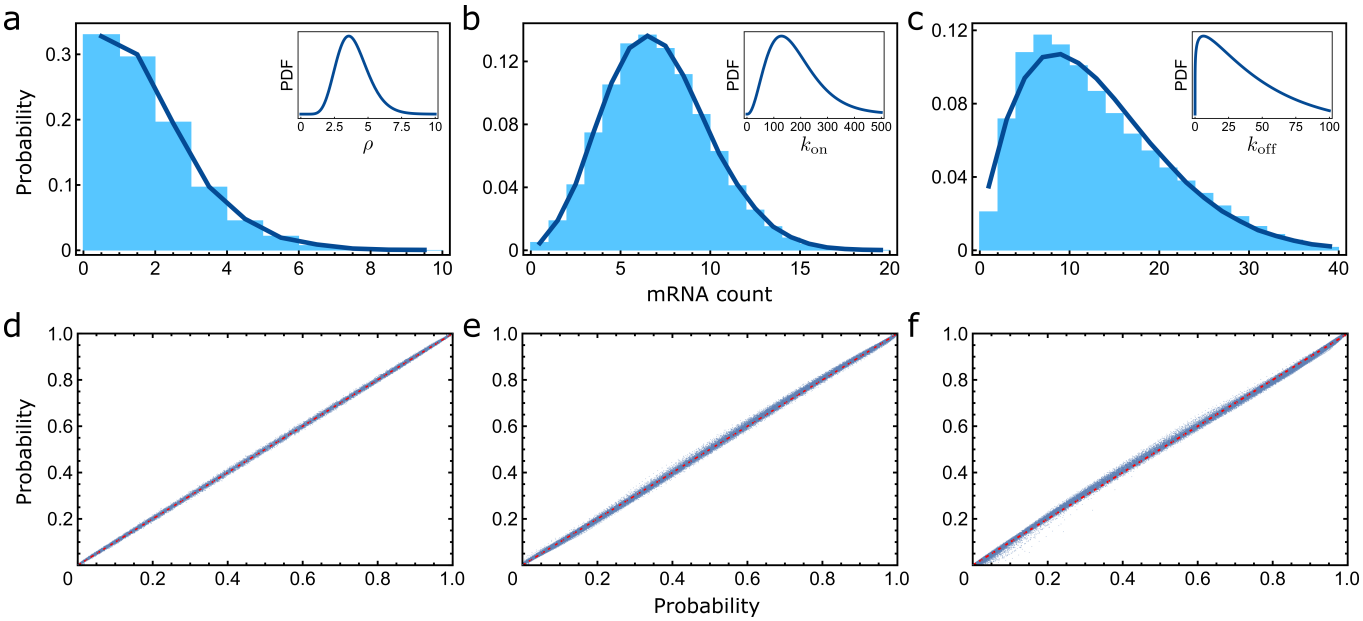

FIG. 3. The steady-state mRNA count distributions of the telegraph model with extrinsic noise and the standard telegraph model are well matched, even when the inferred parameters are very different from the true parameters. (a,b,c) compare the distributions with extrinsic noise (histogram) with those of the fitted telegraph model (solid blue line) where the parameters of the latter are estimated using three-moment-matching procedure given by Eq. (2). The distributions of the parameters that are varying from cell-to-cell are shown in the insets. See Appendix B for the values of the inferred parameters and of the true population averaged parameters. Distribution matching is further confirmed by the high linearity of the P-P probability plots (cumulative distribution function for the telegraph model with extrinsic noise vs cumulative distribution function for the standard telegraph model with parameters from three-moment-matching) for 1000 of the parameter sets used for the plots in the middle column of Fig. 1 for variable synthesis rate (d), variable switching on rate (e) and variable switching off rate (f). Similar results are found for parameter sets estimated under slow promoter switching conditions (right column of Fig. 1).

where x is the normalised coefficient of variation squared of the extrinsic noise in the synthesis rate defined as

$$x = \frac{\sigma_\rho^2}{\langle\rho\rangle^2}(k_{\rm on} + k_{\rm off}). \tag{4}$$

There are several observations that can be made.

1. The inferred mean synthesis rate α is only positive if the size of extrinsic noise is below the critical threshold of $x = 1$. In this parameter regime, the inferred synthesis rate is always larger than the true rate ($\alpha \geq \langle\rho\rangle$) and it increases monotonically with the size of extrinsic noise ($d\alpha/dx > 0$) tending to ∞ as $(1-x)^{-1}$.

2. The inferred mean switching on rate β is positive, independent of the size of extrinsic noise. If the noise is below the critical value of $x = 1$, the inferred rate is always larger than the true rate ($\beta \geq k_{\rm on}$). The estimation error is zero ($\beta = k_{\rm on}$) when $x = 0$ and $x = 1$. The maximum error is reached when

$$x = \frac{k_{\rm off}\,(k_{\rm off} + k_{\rm on})}{k_{\rm off}\,(k_{\rm off} + k_{\rm on}) + \sqrt{k_{\rm off}\,(k_{\rm off} + k_{\rm on})\,(k_{\rm off} + k_{\rm on} + 1)\,(k_{\rm off} + k_{\rm on} + 2)}}.$$

3. The behaviour of the inferred switching off rate θ with x is similar to that of the inferred mean synthesis rate, i.e. it increases monotonically with x tending to ∞ as the critical threshold of $x = 1$ is approached.

4. The inferred burst size ($\lambda = \alpha/\theta$) is positive, independent of the size of extrinsic noise – however note its value is meaningless above the critical threshold because in this case it is the ratio of two negative numbers. If the noise is below the critical value of $x = 1$, the inferred rate is smaller than the true rate ($\lambda < b = \langle\rho\rangle/k_{\rm off}$). At the critical value of $x = 1$, the inferred burst size equals $\langle\rho\rangle/(k_{\rm on} + k_{\rm off})$. When noise is below the critical threshold of $x = 1$, the inferred burst size decreases with the size of extrinsic noise if $k_{\rm off} > k_{\rm off}^*$ otherwise it reaches a minimum for some value of x. The critical value of $k_{\rm off}$ is

$$k_{\rm off}^* = \frac{2 + 3k_{\rm on} + k_{\rm on}^2}{k_{\rm on}}.$$

Of particular importance is the finding that near the critical threshold of $x = 1$ extrinsic noise leads to a very large amplification of the values of the mean synthesis rate and of the mean switching off rate but not of the switching on rate – *this implies that extrinsic noise in the synthesis rate mimics the conditions necessary for transcriptional bursting in the standard telegraph model (without extrinsic noise)* [7].

The results also show that negative parameter values for the synthesis and switching off rates are estimated when noise is larger than the critical threshold of $x = 1$; given the definition of x in Eq. (4) this means that such non-physical estimates occur when the coefficient of variation squared of extrinsic noise $\sigma_\rho^2/\langle\rho\rangle^2$ is greater than $1/(k_{\text{on}}+k_{\text{off}})$. Hence it is unlikely that the size of extrinsic noise is large enough to cause such effects when promoter switching is slow whilst it is very plausible if one or both of the switching rates is large – this explains the previous numerical observation in Section II that when the parameter ranges were $0.1-200$, only 0.7% of the parameter sets of the telegraph model with extrinsic noise in the synthesis rate led to positive values of the standard telegraph model, whereas this percentage rose to 86% when the parameter ranges of the switching rates were restricted to $0.1-1$. We note that while previously it has been reported that estimating parameters by fitting the moments of the standard telegraph model to those of data can lead to negative parameter values, it was not clear why this is the case [40]. One possibility is that this is caused by the finite sample size which leads to errors in moment estimation. However here we have shown that even if the number of cells is infinite (which we simulate by using the exact moments of the telegraph model with extrinsic noise), the inferred parameter estimates from the standard telegraph model can still be negative if the extrinsic noise is larger than the critical threshold.

In Appendix D we show that the estimated parameters show similar qualitative dependence with extrinsic noise strength if the parameter distribution is changed from gamma to lognormal. Since parameter variation between cells is often due to different protein numbers (for e.g. RNA polymerase in the case of synthesis rates) and these are commonly measured to be either gamma or lognormal [41–45] it follows that the results derived here have wide applicability. Interestingly, we find that the critical extrinsic noise strength is smaller for lognormal noise compared to gamma noise, thus suggesting that massive overestimation of the switching off and synthesis rates might be widespread.

A. Maximum likelihood estimates are in good agreement with the moment-based inference theory

Inference using the standard telegraph model is often done using the method of maximum likelihood [6, 13, 46]. This is a different procedure than three-moment matching and hence the question is whether the estimation bias for $x < 1$ predicted by our moment-based theory is also reflected in the maximum likelihood estimates. We directly answer this question as follows. For a given extrinsic noise strength (defined by x), we performed 40 simulation experiments, in each of which sampling of the steady-state distribution of the telegraph model with extrinsic noise in the synthesis rate was used to obtain mRNA counts in 10^4 cells. For each experiment, we infer the rates by maximising the likelihood of the standard telegraph model evaluated using the simulated count data; for details see Appendix E. The statistics of the 40 sets of estimated rates are shown using box-whisker plots in Fig 4(a)-(d). Note that the moment-based theory (blue dashed line) given by Eqs. (3) is close to the median of the rates estimated using maximum likelihood (horizontal red lines). The parameters used for this example are such that the gene spends most of its time in the off state. In Fig. 7(a)-(d) we show the same holds for another parameter set where the gene spends most of its time in the on state – in this case the estimation bias becomes huge – for example the switching off rate and the burst size, as estimated by the standard telegraph model, are about 40 times larger and 5 times smaller respectively than the true values when extrinsic noise is only moderately large $x = 0.45$. In Fig. 7(e)-(h) we also show that for all strengths of extrinsic noise, the mRNA count distribution of the telegraph model with the inferred parameters cannot be distinguished from the mRNA count distribution of the telegraph model with extrinsic noise – this is in line with the findings using moment-based inference in Fig. 3. In summary our results confirm that the bias predictions of our moment-based theory apply also to the commonly used maximum likelihood method when the number of cells for which mRNA counts are measured is not too small (10^4).

While some smFISH and RNA sequencing experiments collect this much data (or more) [13, 47] others do not [6, 48]. If the sample size is very small then the moments may not be well estimated and hence our theory may not accurately capture the magnitude of estimation bias. Hence we used simulations to investigate the extent of bias in the maximum likelihood estimates of the synthesis and switching off rate (the rates displaying the largest bias in Fig. 4 (a,c)) as a function of sample size. In Fig. 4(e)-(f) we show the results for small ($x = 0.2$) and moderately large ($x = 0.6$) extrinsic noise, respectively, for the same parameters as Fig. 4(a)-(d). For a given number N of cells, the positive bias fraction was calculated as follows: we generated 500 samples each with N cells, performed parameter inference for each sample using the maximum likelihood method and then calculated the fraction which overestimated α and θ. When samples had a small number of cells (order of a hundred) and the noise was small, the bias fraction was close to 1/2 indicating that in this case it is difficult to discern any systematic bias since it is overshadowed by the large random variation in parameter estimates from one sample to another. In contrast when the sample size was

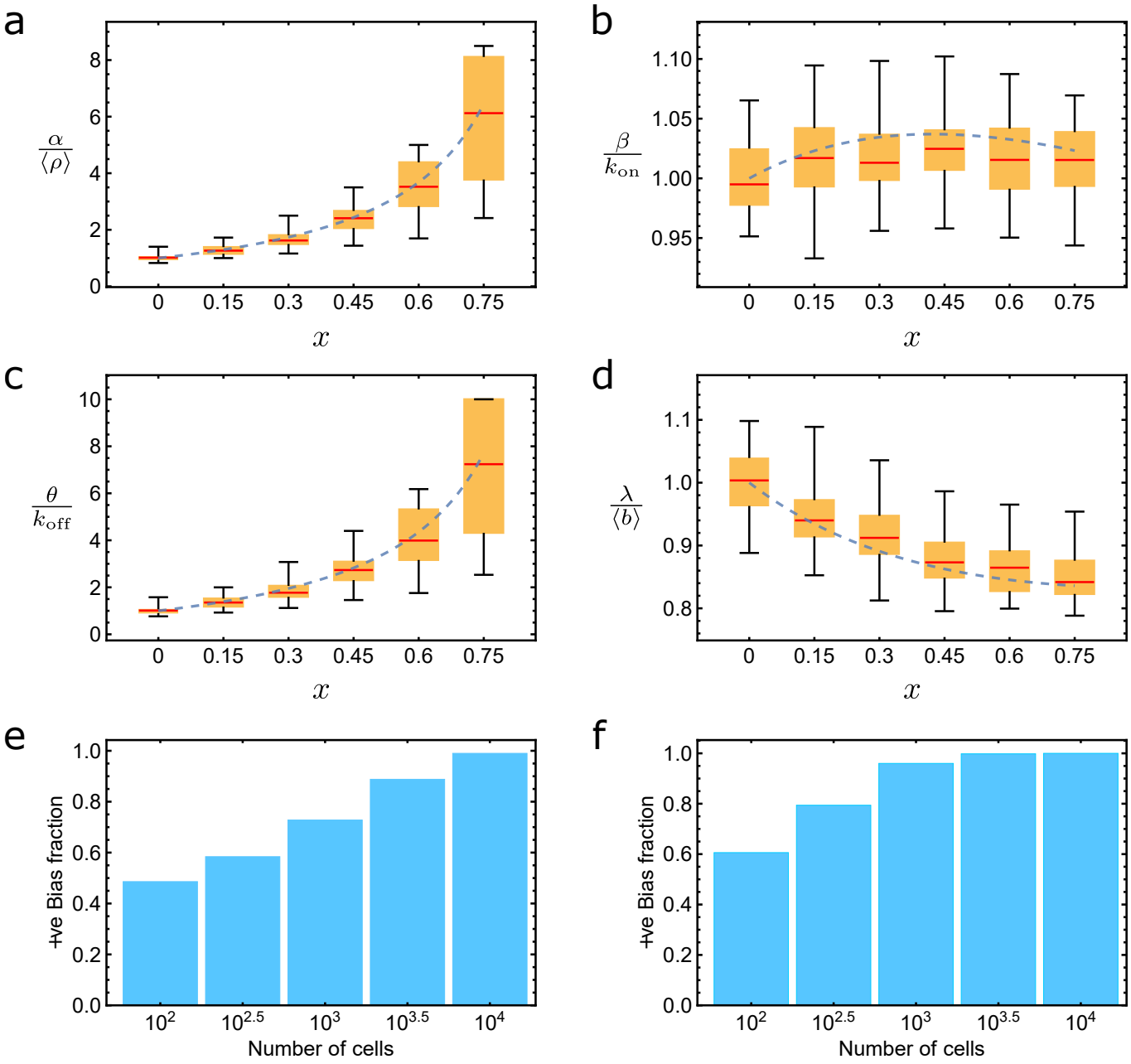

FIG. 4. Comparison of the transcriptional parameters estimated using moment-based theory and those estimated by maximum likelihood fitting of the telegraph model to simulated mRNA count data from a population of cells with cell-to-cell variation in the synthesis rate ρ (gamma distributed). The rates $\alpha, \beta, \theta, \lambda = \alpha/\theta$ are the synthesis rate, switching on rate, switching off rate and burst size of the standard telegraph model, respectively. While $\langle\rho\rangle, k_{\rm on}, k_{\rm off}, \langle b\rangle = \langle\rho\rangle/k_{\rm off}$ have the same meaning but for the telegraph model with extrinsic noise. The latter are the true parameters with which we compare the parameters estimated using the standard telegraph model. Blue dashed lines show the predictions of the moment-matching theory given by Eq. (3). The box-whisker plots (a-d) summarise the statistics of the maximum likelihood rates inferred from a population of 10^4 cells; the distribution for each rate is obtained by repeating the inference for 40 samples, each having 10^4 cells. The lower and upper fences show the min and max, respectively; the yellow bar shows the interquartile range; median is shown by the horizontal red line. The parameters of the telegraph model with extrinsic noise are $\langle\rho\rangle = 20, k_{\rm on} = 1, k_{\rm off} = 5, \langle b\rangle = 4$ (gene spends 5/6 of its time in the inactive state). The size of extrinsic noise x is equal to the normalised coefficient of variation squared as given by Eq. (4). The moment-based theory is close to the median of maximum likelihood estimates. The positive bias fraction (the fraction of parameter estimates for which α and θ are larger than the ground truth values) is generally a function of the number of cells; this is shown in (e) and (f) for $x = 0.2$ and $x = 0.6$, respectively. Bias (positive bias fraction > 0.5) is evident for samples with a number of cells larger than about 10^3 and even for less when the extrinsic noise is strong. Parameter values are the same as for (a)-(d). See main text for details.

of the order of a thousand cells or larger, the positive bias becomes evident. Same occurs for large noise but now the (larger) bias is observable even for samples with only order hundred cells.

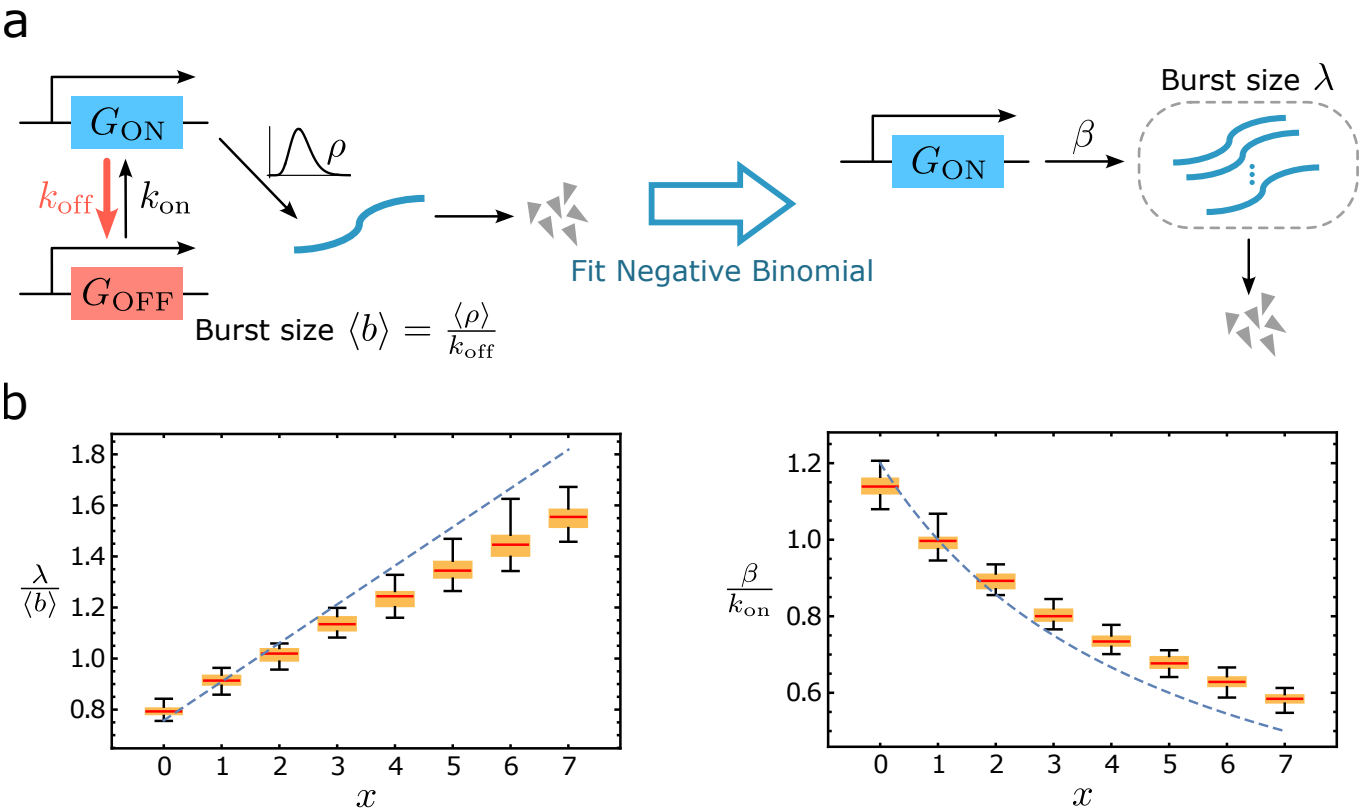

FIG. 5. Estimation bias for extrinsic noise in the synthesis rate in the presence of transcriptional bursting. In (a) we consider the case where parameters are estimated by fitting a negative binomial distribution using maximum likelihood to simulated mRNA count data in a population of cells with cell-to-cell variation in the synthesis rate (gamma distributed) and transcriptional bursting (large switching off and synthesis rates compared to the switching on rate). The negative binomial distribution used for maximum likelihood fitting is given by Eq. (6). In (b) the box-whisker plots (min, max, interquartile range and median) summarise the statistics of the inferred rates using maximum likelihood from 40 simulation experiments, in each of which there are 10^4 cells. The moment-based theory (dashed blue lines) is given by Eq. (8). The trend of maximum likelihood estimates are well described by the moment-based theory – the presence of extrinsic noise causes the burst size to be overestimated and the burst frequency to be underestimated. However as the size of extrinsic noise x increases, there are systematic deviations which derive from deviations of the distribution of mRNA counts from a negative binomial. The parameters of the telegraph model with extrinsic noise are $\langle \rho \rangle = 20, k_{\rm on} = 1, k_{\rm off} = 10, \langle b \rangle = 2$ (gene spends 10/11 of its time in the inactive state). The size of extrinsic noise x is the normalised coefficient of variation squared defined by Eq. (4).

IV. ESTIMATION BIAS IN THE PRESENCE OF TRANSCRIPTIONAL BURSTING: OVERESTIMATION OF BURST SIZE AND UNDERESTIMATION OF BURST FREQUENCY

A. Gamma-distributed burst size

When a gene spends most of its time in the off state, i.e. taking the limit of large $k_{\rm off}$ at constant $k_{\rm on}$, according to Eq. (4) the critical threshold of $x = 1$ corresponds to a coefficient of variation squared of extrinsic noise that is equal to 0. Hence in this case it is generally impossible to match the first three moments of the mRNA counts as predicted by the standard telegraph model to the moments of the telegraph model with extrinsic noise in the synthesis rate (the estimated switching off and synthesis rates will be negative for any extrinsic noise size). In addition, since in this case x can be very large, it follows from Fig. 4(e) and (f) that positive bias will be evident even in small sample sizes of the order of hundred cells. We note that transcriptional bursting is a particular case of the large $k_{\rm off}$ limit; it only manifests when the burst size is also large enough. As we now show, we can still obtain insight into the effects of

extrinsic noise in parameter estimation if we only match the first two moments of mRNA counts to those of a reduced version of the telegraph model.

The simplest two parameter distribution which is most often fitted to mRNA count data, especially that from sequencing, is the negative binomial distribution [49–53]. This mRNA distribution is the steady-state solution of the chemical master equation for the reaction scheme [54]:

$$G_{\text{on}} \xrightarrow{\beta} G_{\text{on}} + rM, \quad M \xrightarrow{1} \emptyset, \tag{5}$$

where r is an integer random number sampled from the geometric distribution supported on the set $0, 1, 2, ...$ Specifically this models the case where a gene produces bursts of mRNA with mean burst size $\lambda = \langle r \rangle$ at a frequency β. The steady-state negative binomial distribution solution of this model is

$$P(m) = \left(1 - \frac{1}{\lambda+1}\right)^m \left(\frac{1}{\lambda+1}\right)^\beta \binom{m+\beta-1}{\beta-1}, \quad m \geq 0, \tag{6}$$

where m is the mRNA count per cell. This distribution is completely defined by its first two factorial moments:

$$\begin{aligned} R_1^b &= \sum_m mP(m) = \lambda\beta, \\ R_2^b &= \sum_m m(m-1)P(m) = \lambda^2\beta(1+\beta). \end{aligned} \tag{7}$$

Equating these to the first and second factorial moments of the telegraph model with gamma distributed noise on the synthesis rate (see expressions for R_1 and R_2 in Appendix A), we find the estimated rates are given by:

$$\begin{aligned} \beta &= \frac{k_{\text{on}}\left(k_{\text{off}} + k_{\text{on}} + 1\right)}{k_{\text{off}} + x(k_{\text{on}}+1)}, \\ \lambda &= \frac{\langle b \rangle k_{\text{off}}\left(k_{\text{off}} + x(k_{\text{on}}+1)\right)}{\left(k_{\text{off}} + k_{\text{on}}\right)\left(k_{\text{off}} + k_{\text{on}} + 1\right)}, \end{aligned} \tag{8}$$

where $\langle b \rangle = \langle \rho \rangle / k_{\text{off}}$ is the true burst size and x is the (normalised) size of extrinsic noise as defined in Eq. (4). Note that a gamma distributed synthesis rate is equivalent to a gamma distributed burst size. The fitting procedure is illustrated in Fig. 5(a). Clearly an advantage of this type of model fitting is that the parameters inferred are always positive, independent of the strength of extrinsic noise. On the other hand, note that the errors made by this type of inference can be significant even when there is no extrinsic noise ($x = 0$). In this case, the burst frequency should be $\beta = k_{\text{on}}$ (the rate at which a burst starts is equal to the rate at which the gene switches on) while the burst size should be $\lambda = \langle b \rangle$. This is only obtained if the condition $k_{\text{off}} \gg 1 + k_{\text{on}}$ is true which makes sense when one considers that the steady-state mRNA distribution solution of the standard telegraph model is well approximated by a negative binomial in this limit [7]. Enforcing the condition $k_{\text{off}} \gg 1 + k_{\text{on}}$, we find Eq. (8) reduces to:

$$\begin{aligned} \lambda &\approx \langle b \rangle (1 + y(1 + k_{\text{on}})), \\ \beta &\approx \frac{k_{\text{on}}}{1 + y(1 + k_{\text{on}})}, \end{aligned} \tag{9}$$

where $y = x/k_{\text{off}}$ is to a very good degree of approximation equal to $\sigma_\rho^2/\langle \rho \rangle^2$, the coefficient of variation of extrinsic noise. From these results, it is clear that *fitting a negative binomial distribution to single-cell data characterised by transcriptional bursting and gamma distributed extrinsic noise in the synthesis rate will overestimate the burst size and underestimate the burst frequency.*

In Fig. 5(b) we compare the predictions of the moment-based theory (dashed blue lines) given by Eqs. (8) with estimates obtained by fitting a negative binomial distribution using maximum likelihood (shown using box-whisker plots) to simulated data for 10^4 cells. There are two main observations: (i) for $x = 0$, the burst size and burst frequencies are not equal to the true parameters. The reason is because the formal condition $k_{\text{off}} \gg 1 + k_{\text{on}}$ necessary for the telegraph model's solution to be well approximated by a negative binomial distribution is only approximately met for the parameters used in this example ($k_{\text{off}} = 10, k_{\text{on}} = 1$). This shows that even in the absence of extrinsic noise, there are typically systematic errors in parameter estimation (underestimation of burst size and overestimation of burst frequency). These differences approach zero as the switching off rate is made to be 100 times or larger than the switching on rate; (ii) the maximum likelihood estimates follow the trends predicted by the moment-based theory, namely that an increase in extrinsic noise leads to an amplification in the observed burst size and an attenuation of

the observed burst frequency. However the differences between the two methods increase with the size of extrinsic noise. This is not a sample size effect – repeating the inference using 100 times more cells (40 experiments each with 10^6 cells) does not remove the differences. We note that the negative binomial is a two parameter distribution that can be estimated completely from knowledge of its two moments, i.e. if the distribution of the telegraph model with extrinsic noise is well approximated by the negative binomial then the moment-based theory and maximum likelihood would be in good agreement. Hence the differences between the two methods, with increasing extrinsic noise size, ensue from the lack of "flexibility" of the negative binomial distribution to describe the distribution of mRNA counts.

B. General case: extrinsic noise in burst size and frequency with arbitrary distributions

We next consider the general case where there is transcriptional bursting and arbitrarily distributed noise on any of the three parameters of the telegraph model (synthesis rate or one of the switching rates). Since we have bursting rather than using the full telegraph model with extrinsic noise, instead we consider the following reaction system as a suitable reduced model describing the data:

$$G_{\text{on}} \xrightarrow{k_{\text{on}}} G_{\text{on}} + rM, \quad M \xrightarrow{1} \emptyset, \tag{10}$$

where $\langle r \rangle$ is the burst size b and k_{on} is the burst frequency. The cell-to-cell variations in these two parameters are described by a burst frequency distribution, $f(k_{\text{on}})$ and a burst size distribution, $g(b)$. It follows that the first two factorial moments of the mRNA counts are given by integrating the factorial moments of the negative binomial distribution solution of the model in the absence of extrinsic noise over the distribution of parameters:

$$\begin{aligned} R_1 &= \int_{k_{\text{on}}} \int_b k_{\text{on}} b f(k_{\text{on}}) g(b) dk_{\text{on}} db = \langle k_{\text{on}} \rangle \langle b \rangle, \\ R_2 &= \int_{k_{\text{on}}} \int_b b^2 k_{\text{on}} (1 + k_{\text{on}}) f(k_{\text{on}}) g(b) dk_{\text{on}} db = \langle b^2 \rangle (\langle k_{\text{on}} \rangle + \langle k_{\text{on}}^2 \rangle). \end{aligned} \tag{11}$$

We simulate the fitting of the data to a non-extrinsic noise model (Fig. 5a) by setting $R_i^b = R_i, i = 1, 2$, i.e. equating Eq. (7) with Eq. (11). Then solving for the estimated mean burst size and the mean burst frequency we find:

$$\begin{aligned} \lambda &= \langle b \rangle \left(1 + \langle k_{\text{on}} \rangle \text{CV}^2_{k_{\text{on}}} + \text{CV}^2_b (1 + \langle k_{\text{on}} \rangle + \langle k_{\text{on}} \rangle \text{CV}^2_{k_{\text{on}}}) \right), \\ \beta &= \frac{\langle k_{\text{on}} \rangle}{1 + \langle k_{\text{on}} \rangle \text{CV}^2_{k_{\text{on}}} + \text{CV}^2_b (1 + \langle k_{\text{on}} \rangle + \langle k_{\text{on}} \rangle \text{CV}^2_{k_{\text{on}}})}, \end{aligned} \tag{12}$$

where CV^2_z is the coefficient of variation squared of parameter z where $z = k_{\text{on}}, b$. In the limit of no extrinsic noise in the switching on rate (no extrinsic noise in the burst frequency), Eq. (12) reduces to Eq. (9). Note that Eq. (12) has a powerful interpretation: *if transcriptional bursting is present, independent of the source of extrinsic noise and the inherent distribution of the parameters, the mean burst size is overestimated while the mean burst frequency is underestimated when these are deduced from fitting a negative binomial distribution (or the mRNA count distribution of the standard telegraph model) to the mRNA count data.* The theory's accuracy is tested and confirmed using maximum likelihood applied to simulated counts data generated using the telegraph model with gamma distributed extrinsic noise for four parameter sets (two with noise in the burst frequency and two with noise in the burst size) in Table I.

C. Application to scRNA-seq data

A recent study has used the method of maximum likelihood to fit the mRNA counts from allele-sensitive single-cell RNA sequencing to the standard telegraph model thus determining the transcription parameters for endogenous mouse genes transcriptome-wide [6]. For a large number of genes, the switching on rate was found to be much less than the switching off rate and the synthesis rate thus leading to the conclusion that transcriptional bursting is ubiquitous in mammalian genes. This suggests that the distributions of mRNA counts for many genes in this dataset are well fit by a negative binomial distribution, i.e. the only two parameters that can be determined accurately are the burst frequency (switching on rate) and the burst size (the ratio of the synthesis and switching off rates). In fact, as we show in Fig. 8, to a good degree of approximation their standard telegraph model estimates (using the method of maximum likelihood) for the burst frequency and size for genes displaying highly bursty expression are in good

$\langle\rho\rangle$	$\langle k_{\text{on}}\rangle$	k_{off}	Noise	$\lambda/\langle b\rangle$	$\beta/\langle k_{\text{on}}\rangle$
600	10	300	$CV_{k_{\text{on}}}=1$	[10.45, 10.55, 10.68]	[0.088, 0.089, 0.090]
200	10	200	$CV_b=1$	[10.30, 10.46, 10.62]	[0.090, 0.091, 0.092]
400	10	600	$CV_{k_{\text{on}}}=1$	[10.59, 10.75, 10.87]	[0.090, 0.091, 0.092]
800	10	500	$CV_b=1$	[10.53, 10.70, 10.82]	[0.091, 0.092, 0.093]

TABLE I. Comparison of maximum likelihood estimation using a negative binomial distribution and Eq. (12). For each set of parameters, we generated 100 independent samples, each with 10^4 cells. The mRNA counts per cell were simulated from the telegraph model with gamma distributed extrinsic noise in the switching on rate (lines 1 and 3) or in the synthesis rates (lines 2 and 4). For each sample we estimated the burst frequency and size by fitting a negative binomial distribution using maximum likelihood. The last two columns show the estimated parameters ([first quartile,median,third quartile]) computed over all 100 samples. The moment-based theory predicts that in all these cases, $\lambda/\langle b\rangle \approx 11$ and $\beta/\langle k_{\text{on}}\rangle \approx 1/11$. The median of the maximum likelihood estimates varies little between the four parameter sets and its value is close to the theoretical predictions.

agreement with those obtained by simply equating the first and second factorial moments of the mRNA count data for each gene to the factorial moments Eq. (7) predicted by the bursty expression model (5) and solving for λ and β. Any deviations between the two estimates are due to the small sample size of a few hundreds of cells.

However since in the transcriptional bursting regime the normalised coefficient of variation of extrinsic noise (Eq. (4)) is large, given the results in Fig. 4(f), we expect that the moment estimates computed from the few hundred cells measured in [6] will be significantly biased due to extrinsic noise. To correct the moment-matching estimates using the theory we previously developed, we need to estimate the coefficient of variation of extrinsic noise. We assume that the data can be well described by the effective reaction scheme (10) with arbitrary distributions of the burst size and burst frequency. Using Eq. (11) it can be shown that the coefficient of variation squared CV_i^2 of the mRNA counts for gene i can be written as:

$$\langle m_i\rangle = \langle b_i\rangle\langle k_{\text{on}}^i\rangle, \tag{13}$$

$$\text{CV}_i^2 = \frac{1+\langle b_i\rangle}{\langle m_i\rangle} + \frac{\langle b_i\rangle \text{CV}_{b_i}^2}{\langle m_i\rangle} + \text{CV}_{b_i}^2 + \text{CV}_{b_i}^2\text{CV}_{k_{\text{on}}^i}^2 + \text{CV}_{k_{\text{on}}^i}^2, \tag{14}$$

where $\langle m_i\rangle$ is the mean number of mRNA counts for gene i, and $\text{CV}_{b_i}^2$ and $\text{CV}_{k_{\text{on}}^i}^2$ are the coefficient of variation squared for the burst size and frequency of gene i, respectively. The means and coefficients of variation of the parameters are calculated for a population of cells. We find that this law is reflected in the sc-RNA seq data (it has also been found in other scRNA-seq datasets [48, 53]). In Fig. 6(a) we plot the coefficient of variation squared as a function of the mean expression for about 10^4 mouse fibroblast genes with mean expression > 1. For both the C57 and the CAST alleles, the law $\text{CV}_i^2 \propto 1/\langle m_i\rangle$ manifests for mean expression below 10 (red line in inset of Fig. 6(a)) followed by a transition to the law CV_i^2 equals some constant (independent of the gene index i) for larger mean expression levels. These observations follow Eq. (14) – the first two terms describe the small expression law and the last three terms explain the high expression law.

The significant scatter about the red line for low mean expression is expected because each gene will have a different value of the mean burst size and hence the first two terms in Eq. (14) can only be approximately inversely proportional to the mean expression. On the other hand, the low scatter about the line CV^2 equals some constant (for high-expression genes) indicates that the coefficient of variation of the burst size and frequency are similar across genes; these measures of the size of extrinsic noise are also independent of the allele. These observations about extrinsic noise can be explained as follows. Since burst frequency changes with cell cycle phase [4, 13, 14], it is expected that its coefficient of variation is not particularly large for this dataset because the population of cells is mostly in the G1 phase [6]. Thus we assume that the variation in the parameters is dominated by the burst size since this is often influenced by the cell volume [14]. Specifically we assume that the burst size b_i of gene i in a cell of volume V is given by $b_i = r_i V$ where r_i is some gene specific constant. This law naturally emerges if there is concentration homeostasis, i.e. transcript numbers scale approximately linearly with the cell volume thus maintaining a constant concentration [55–57]. Then it immediately follows that $\text{CV}_{b_i}^2$ in Eq. (14) equals CV_V^2, i.e. the coefficient of variation in the burst size is the same for each gene and equal to the coefficient of variation of the cell volume (computed for a population of cells). Note that a rough estimate of the size of extrinsic noise is given by the coefficient of variation squared of mRNA counts in the limit of large mean expression as shown by the dashed black line in Fig. 6(a).

An analysis of the covariance between different genes provides strong supporting evidence of extrinsic noise due to a cell-specific factor (such as the volume). In Appendix F we analytically show that if the synthesis rate (and hence the burst size) linearly scale with a cell-specific factor V then the covariance between the mRNA counts of any two

genes i and j ($i \neq j$) is given by

$$\mathrm{Cov}(m_i, m_j) = \langle m_i \rangle \langle m_j \rangle \mathrm{CV}_V^2. \tag{15}$$

This relationship is true also if the burst frequency has a linear dependence on a cell-specific factor provided some additional criteria are met. Hence the prediction is that extrinsic noise induces a positive covariance between the counts of any pair of genes (see also [58] where high gene-gene correlations where observed and explained by burst size coordination). This might be surprising since one would assume that correlations between some genes would be positive/negative due to regulatory effects and many would be uncorrelated. However calculation of the covariance for highly bursty genes shows that most show a positive covariance and the linear relationship predicted by theory Eq. (15) is evident (Fig. 6(b)). Additionally the slope of the best straight line through the data gives us a reliable estimate of the size of extrinsic noise: $\mathrm{CV}_V^2 \sim 0.32$.

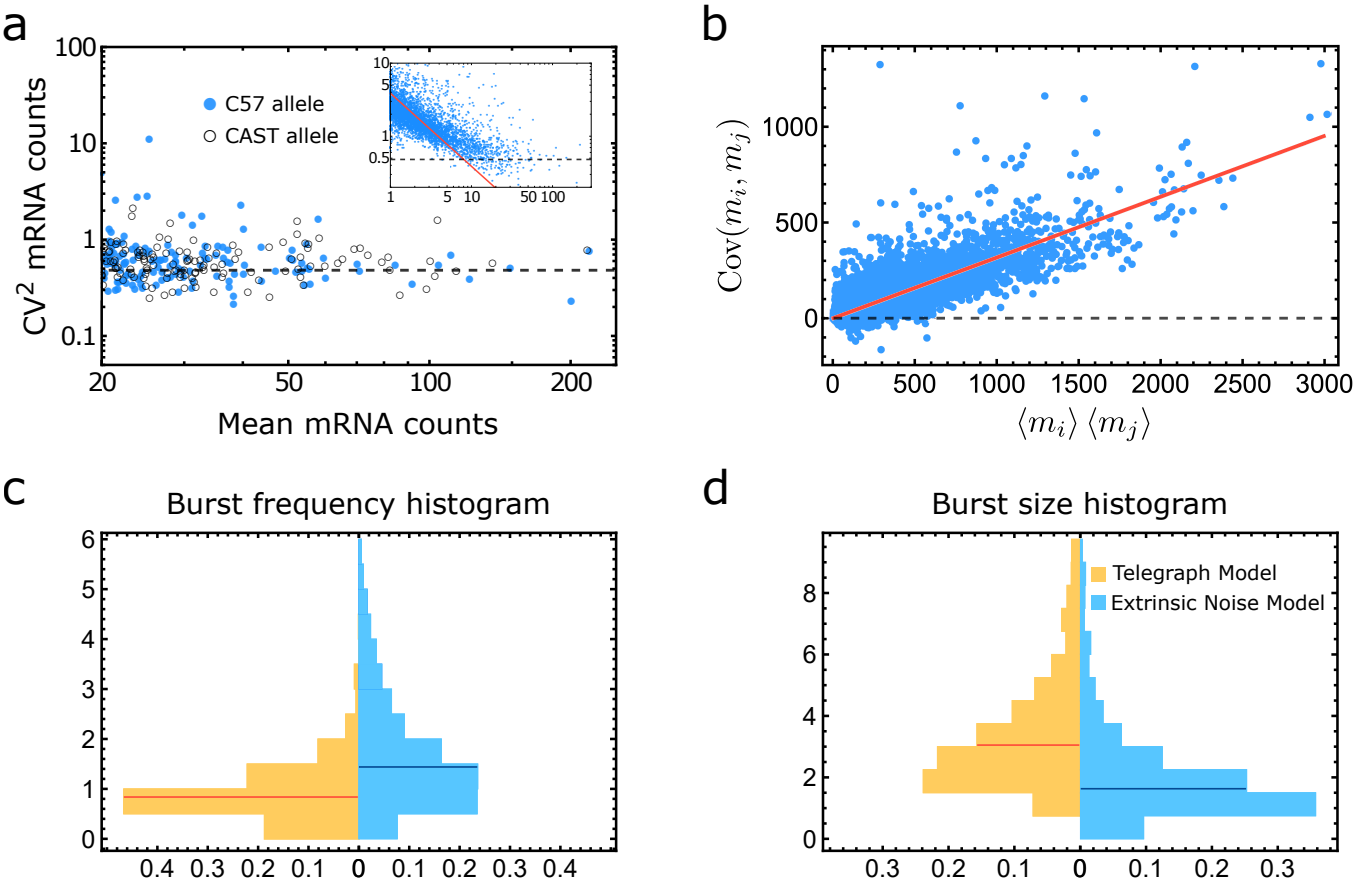

FIG. 6. Estimation bias of burst size and frequency from scRNA-seq data [6]. (a) Plots of the coefficient of variation squared CV^2 of mRNA counts versus their mean for the C57 and cast alleles of high expression genes (defined as those with mean > 20) in 224 primary adult fibroblasts – each dot is a gene. The CV^2 does not vary much with the mean expression (dashed black line) indicating a high degree of extrinsic noise from a cell-specific factor. In contrast for low-expression genes, there is a fuzzy but clear inverse relationship between CV^2 and the mean number of counts (solid red line in the inset), indicating strong contributions from both intrinsic and extrinsic noise. (b) Plot of the covariance between a gene pair (gene i and gene j, where $i \neq j$) versus the product of their mean expression for 1877 highly bursty genes (C57 allele) (see Fig. 8 for their definition). The plot confirms the theoretical prediction Eq. (15); the red regression line has correlation coefficient $= 0.90$ and its slope (0.32) is equal to the coefficient of variation squared of extrinsic noise. 96% of gene pairs have a positive covariance. (c) and (d) Paired histograms for the estimates of the burst frequency and the burst size, obtained by matching the first two moments of the C57 data for each of the highly bursty genes to those of the standard telegraph model assuming transcriptional bursting and separately by correcting it to account for extrinsic noise with size as found in (b). The horizontal lines show the medians of the distributions.

Based on the foregoing arguments we shall assume that for each mouse fibroblast gene displaying apparent bursting behaviour in the sc-RNA seq dataset, the mean expression and the coefficient of variation squared of the mRNA counts are given by Eqs. (13)-(14) with $\mathrm{CV}^2_{k^i_{\mathrm{on}}} = 0, \mathrm{CV}^2_{b_i} = 0.32$. Note that the genes selected for inference are the

subset of the highly bursty genes in Fig. 8 for which $\mathrm{CV}^2_{b_i} > 0.32$. Solving the equations for the mean expression and the coefficient of variation squared simultaneously we obtain estimates of the mean burst size and frequency for each gene. The distribution of these two parameters across the transcriptome are shown in blue in Fig. 6(c) and (d); for comparison, we also show (in yellow) the corresponding distributions if we had to ignore extrinsic noise, i.e. as done in the original analysis of [6]. Clearly, the addition of extrinsic noise leads to an increase in the burst frequency estimates (the mass of the distribution spreads towards higher values) and a corresponding decrease in the burst size estimate (the mass of the distribution concentrates in the small value region) – this is in line with our general theory summarised by Eq. (12). In particular, the median burst frequency increased by about 2 times while the median burst size decreased by the same amount when extrinsic noise was taken into account. We note that what we have referred to as the mean burst size for gene i is the observed mean burst size, i.e. $\langle b_i \rangle = r_i \langle V \rangle$, where the average is taken over the population of cells. If one is interested in deducing r_i then a measurement or estimation of the average cell volume is additionally needed. The difficulty here is that it is likely that what we have called cell volume is not simply that but a variable which depends on various cell-specific factors and also technical variability [59].

In our inference procedure we have made a major assumption: the number of mRNA per gene per cell corresponds to the number of UMIs (unique molecular identifiers) detected. This is the same assumption made in [6]. Comparing their measured distributions of mRNA counts for 5 genes measured using both smFISH and scRNA-seq (Extended Fig. 7), it is clear that the latter display an accumulation of probability mass to the left implying there is a non-negligible fraction of transcripts in each cell that is not captured by scRNA-seq, a well-known property [60]. It is however not clear how to correct for these effects. Hence the burst frequency and size histograms output by our extrinsic noise moment-matching procedure shown in Fig. 6(c) and (d) are not to be interpreted as the true histograms but rather a step in that direction.

Heuristic methods to correct estimates for capture efficiency of scRNA-seq technology have been recently proposed in [61]. The main idea of the moment-based method proposed (referred to as the MME) is that instead of performing parameter inference directly on raw mRNA counts from UMI count-based scRNA-seq data, one should first normalise the counts from each cell by a heuristic cell-specific scaling factor (calculated from the raw data), and subsequently the moments of these renormalised counts can be fitted to those of a standard telegraph model to estimate the transcriptional parameters. Despite its simplicity and ease of computation, it can be analytically shown that this procedure is flawed because the normalisation of count data does not transform the moments of the telegraph model with extrinsic noise to those of the standard telegraph model. We show in Appendix G that inference using this method is only accurate when the mean expression is large enough – in contrast the inference method whose results are shown in Fig. 6(c) and (d) does not suffer from such inaccuracies. A neural-network (NN) method is also developed in [61] to correct for capture efficiency and this maybe useful when it can be ascertained that the major sources of extrinsic noise are in the synthesis rate, e.g. when the scRNA-data is collected from cells in a particular cell cycle phase such that burst frequency variations between cells are minimised [4, 13, 14]. This method also predicts corrected estimates for the burst size and frequency of the same scRNA-seq dataset we analysed in Fig. 6. The burst size is defined differently than in our case (in our notation for gene i it is defined as the parameter r_i) and hence it is not possible to compare our estimate with those from the NN method. The burst frequencies can however be compared: both the NN method and ours predict that taking into account extrinsic noise considerably widens the burst frequency histogram. Our theory additionally predicts that independent of the source of the extrinsic noise, if transcriptional bursting is evident, the median burst frequency of bursty genes should be larger after correcting for it; this might be consistent with the results of the NN study but it is difficult to ascertain from the published data analysis.

V. SUMMARY AND DISCUSSION

In this paper, we developed a theory that explains the size and sign of estimation bias when the standard telegraph model is used to infer transcriptional parameters from single cell data. We note that while it is widely appreciated that this model does not take into account extrinsic noise (it assumes the parameters do not vary between cells which is of course unrealistic), it is still the approach of choice in most studies [3, 5, 6, 40, 62] because of its simplicity and the considerable difficulty in completely correcting for all sources of extrinsic noise in mRNA count data. Hence a main motivation for our study was to understand how well the parameters estimated using the standard telegraph model approximate the true average parameter values in a population, i.e. does the concept of a virtual "mean cell" make sense? [26]

A key assumption of our theory is that single-cell data is accurately approximated by the telegraph model with extrinsic noise in either the synthesis rate or the switching on rate or the switching off rate. By matching the moments of this model with those of the standard telegraph model, we mimic the common procedure of fitting the latter to single-cell data. This method enables us to obtain expressions for the standard telegraph model parameters as a function of the size of extrinsic noise (the coefficient of variation of the parameter that varies the most between cells)

and of the actual population averaged parameters. We show that (i) the estimated parameters are not generally equal to the population average of these parameters; (ii) whenever three moments of the telegraph model can be fitted to those of the data, i.e. none of the estimated parameters are negative, there are unique bias signatures depending on the source of extrinsic noise, i.e. which parameter is the most variable amongst cells. Variability in the synthesis rates leads to an overestimation of the three transcriptional parameters (synthesis rate and the two switching rates) and an underestimation of the burst size. Variability in the switching on rate has the exact opposite effect. Variability in the switching off rate leads to an overestimation of the synthesis rate and underestimation of the switching on rate; the switching off rate is mostly underestimated while the burst size can be over or underestimated depending on the speed of promoter switching; (iii) the size of bias is significantly diminished when extrinsic noise is in one of the switching rates and the promoter switches slowly between inactive and active states compared to the rate of mRNA degradation; (iv) independent of the magnitude of estimation bias, the mRNA count distributions of the standard telegraph model are in most cases practically indistinguishable from those of the telegraph model with extrinsic noise. This implies that given data from only one gene, standard model selection criteria (Akaike and Bayesian information criteria) would likely lead to the conclusion that the standard telegraph model is the model of choice even if the stochasticity in count data is not due to intrinsic sources; (v) extrinsic noise in the synthesis rate leads to a massive overestimation of the synthesis and switching off rates (but has little effect on the switching on rate) when the extrinsic noise size is close to a critical threshold. Beyond this critical size, it is not possible to match the first three moments of the telegraph model to data; (vi) for genes exhibiting transcriptional bursting, only the first two moments can be reliably matched. In this case extrinsic noise, independent of its source or distribution, will lead to an overestimation of the burst size and an underestimation of the burst frequency; (vii) while our analytical results assume moment-matching and an infinite sample size, we showed that similar numerical results are obtained by maximising the likelihood of the telegraph model using data with realistic sample sizes, a commonly used method.

We note that result (i) is not an obvious one. In fact given a parameter that varies about some mean from one cell to another, one would expect this noise to average out when performing inference using the standard telegraph model. This reasoning is implicit in the wide use of this model in smFISH and scRNA-seq studies. But our study shows that this averaging does not occur; rather extrinsic noise leads to a systematic estimation bias which artificially inflates some parameter values while having the opposite effect on others. A particular instance of these deviations was reported in a recent experimental study [13]. The telegraph model parameters were determined for the GAL10 gene for cells in the G1 phase, in the G2 phase and for cells in both phases together (merged data). It was found that the synthesis rate, the switching on rate and the switching off rates estimated using the merged data were larger than the true population averaged values, i.e. the weighted average of the parameters in the G1 and G2 phases. The opposite was true for the burst size. The sign of the bias is exactly the same as in one of our theoretical results (see result (ii) above). Of course one can argue that if the size of the bias is small then there would not be much cause to worry. But by result (v) we have that for common sources of extrinsic noise which affect the synthesis rate (such as the coupling of this rate and cell size to achieve concentration homeostasis [14]), the standard telegraph model estimates of the synthesis and switching off rates go to infinity as the size of extrinsic noise (the variability of the synthesis rates between cells) approaches a critical value. This could explain why a recent study [6] that estimated transcriptome-wide parameters using the standard telegraph model found a significant number of mammalian genes for which the synthesis and the switching off rates are a hundred to a thousand times larger than the switching on rate – examples of such genes are Sdpr, Sf3b1, Hspe1, Cyb5r1, Phlda3 and Prrx1. In fact our reanalysis of the scRNA-seq data for ~ 2000 highly bursty mouse fibroblast genes in this study showed that taking into account extrinsic noise leads to burst frequency and size distributions that are remarkably different to those initially found using the standard telegraph model. While the count distributions for each gene are well fit by both the standard telegraph model and that with extrinsic noise (as found in result (iv)) however we showed that the two models can be told apart using the covariance of mRNA counts between different genes.

The main limitations of our study are: (i) the assumption that the mRNA degradation rate does not vary between cells. Relaxing this assumption makes it difficult to understand the effects of extrinsic noise on parameter estimation because it implies cell-to-cell variability in all three normalised transcriptional parameters of the telegraph model. Some studies have found that degradation rate does vary between cells [31, 63]. Other studies have found no relationship between cell size and degradation rate and thus posited that it is not a major source of extrinsic noise [14, 15]; (ii) the assumption that the data is single-allele. This data is nowadays readily available. However if one is interested in understanding the accuracy of inference from non-allele-specific data then a similar moment-based analysis as we did for one allele can be carried out provided one assumes independent expression from each allele [7, 13]; (iii) the assumption that each cell has a set of time-independent parameters. While generally it is to be expected that parameters vary with time inside a cell, nevertheless if the timescale of variation is slow enough then a cell will alternate between a number of steady states and a measurement will capture a cell in one of these – our theory is valid provided this is true. Moments of the standard model modified to account for dynamic extrinsic noise can be derived in specific cases and under some assumptions [64–66]. However because of the complexity involved, it will

generally be difficult to extract relationships between the true population averaged parameters and those inferred; (iv) the correction of burst frequency and size from scRNA-seq data to account for extrinsic noise assumes that the UMI counts for a gene accurately represents the true counts. Whilst this has been assumed previously [6] it cannot be the case because typically the capture efficiency of sequencing technologies is not very high [53]. The capture efficiency model postulated in [59, 61] suggests that variability in capture efficiency between cells can be modelled by a variable synthesis rate. If this is indeed the case then our theory's corrections for the population averaged burst frequency for highly bursty genes in [6] are accurate. On the other hand, our theory's corrections for the population averaged burst size would depend on both cell-to-cell variation in the synthesis rate and in the capture efficiency and hence would not purely reflect the natural variation in parameters. Thus generally it is possible that the burst frequency is more accurately inferred than the burst size. It is also likely that explicit modelling of technical noise arising due to RNA-sequencing protocols is needed to understand how to properly correct the count data from scRNA-seq studies [67, 68].

The overall implication of our study is to avoid parameter estimation from marginal steady-state distributions of the mRNA counts using the standard telegraph model because the results are meaningless if there is considerable parameter variability between cells (the typical case). This type of estimation is only feasible if the major sources of extrinsic noise can be identified and the data is appropriately corrected for them prior to inference or else if the parameters are corrected *a posteriori*.

ACKNOWLEDGMENTS

We thank Augustinas Sukys for useful feedback and help with the figures. This work was supported by a Leverhulme Trust research award (RPG-2020-327).

Appendix A: Moments of the telegraph model with and without gamma distributed extrinsic noise

Let $x^{(k)} = x(x+1)\cdots(x+k-1)$ (resp. $x_{(n)} = x(x-1)\cdots(x-(k-1))$) be the rising (resp. falling) factorial, and $\left[{n \atop k}\right]$ the absolute value of the corresponding Stirling number of the first kind. These numbers are related by the following identity:

$$x^{(n)} = \sum_{k=0}^{n} \begin{bmatrix} n \\ k \end{bmatrix} x^k. \tag{A1}$$

1. Derivation of the factorial moments of the telegraph model

The steady-state distribution of mRNA counts $P(m)$ of the telegraph model (Eq. (1)) is exactly equal to the Beta-Poisson distribution, i.e. the distribution of a Poisson random variable with parameter $\alpha \cdot r$, where r is a random variable that is distributed according to the Beta(β, θ) distribution [35, 69]. Since the nth factorial moments of a Poisson distribution is the nth power of its parameter, it immediately follows that to obtain the mth factorial moment of the mRNA counts m of the telegraph model, we only have to calculate the following integral:

$$R_n^T = \sum_m m_{(n)} P(m) = \int_{r=0}^{1} (\alpha r)^n \frac{\Gamma(\beta+\theta)}{\Gamma(\beta)\Gamma(\theta)} r^{\beta-1}(1-r)^{\theta-1} dr = \alpha^n \frac{\Gamma(\beta+n)}{\Gamma(\beta)} \cdot \frac{\Gamma(\beta+\theta)}{\Gamma(\beta+\theta+n)}. \tag{A2}$$

2. Derivation of the factorial moments of the telegraph model with gamma-distributed noise

To distinguish the models when there is extrinsic noise, we will consider the setup shown in Fig. 2(a)-(c) left panels where now we call the synthesis rate ρ, the switching on rate k_{on} and the switching off rate k_{off}. We will consider each of these parameters to be distributed according to a gamma distribution with shape parameter a and scale parameter b.

a. A few useful formulas

Using a property of the gamma function, one can write the rising factorial as a quotient of two gamma functions :

$$x^{(k)} = x \cdot (x+1)\cdots(x+k-1) = \frac{\Gamma(x+1)}{\Gamma(x)} \cdot \frac{\Gamma(x+2)}{\Gamma(x+1)} \cdots \frac{\Gamma(x+k)}{\Gamma(x+k-1)} = \frac{\Gamma(x+k)}{\Gamma(x)}. \tag{A3}$$

For $n \geq 1$ the following partial fraction decomposition holds:

$$\frac{1}{x^{(n)}} = \frac{1}{(n-1)!} \sum_{j=1}^{n-1} \binom{n-1}{j} \frac{(-1)^j}{x+j}. \tag{A4}$$

This can be derived by setting $\frac{1}{x(x+1)\cdots(x+n-1)} = \sum_{j=0}^{n-1} \frac{r_j}{x+j}$ where

$$r_j = \lim_{x \to -j} \frac{x+j}{x(x+1)\cdots(x+n-1)} = \frac{1}{(-j)\cdots(-1)\cdot 1\cdots(n-j-1)} = \frac{(-1)^j}{j!(n-1-j)!} = \frac{(-1)^j}{(n-1)!}\binom{n-1}{j}.$$

Kummer's function of the second kind $U(a,c,z)$ can be expressed as [70] :

$$U(a,c,z) = \frac{1}{\Gamma(a)} \int_0^{+\infty} \frac{t^{a-1}\mathrm{e}^{-zt}}{(1+t)^{a+1-c}} dt. \tag{A5}$$

For $a, b, \alpha > 0$ the following equalities hold:

$$\frac{1}{b^{a-1}\Gamma(a)} \int_0^{+\infty} \frac{x^{a-1}\mathrm{e}^{-\frac{x}{b}}}{x+\alpha} dx = \frac{\alpha^{a-1}}{b^{a-1}\Gamma(a)} \int_0^{+\infty} \frac{y^{a-1}\mathrm{e}^{-\frac{\alpha}{b}y}}{y+1} dy, \qquad \text{(where } y = \tfrac{x}{\alpha}\text{)}$$

$$= \left(\frac{\alpha}{b}\right)^{a-1} U\left(a, a, \frac{\alpha}{b}\right), \qquad \text{(using Eq. (A5))}$$

$$= \mathrm{e}^{\frac{\alpha}{b}} E_a\left(\frac{\alpha}{b}\right), \qquad \text{(using Ref. [71])}$$

where $E_a(z) = \int_1^{+\infty} \frac{\mathrm{e}^{-zt}}{t^a} dt$ is the exponential integral function. Hence

$$\frac{1}{b^{a-1}\Gamma(a)} \int_0^{+\infty} \frac{x^{a-1}\mathrm{e}^{-\frac{x}{b}}}{x+\alpha} dx = \mathrm{e}^{\frac{\alpha}{b}} E_a\left(\frac{\alpha}{b}\right). \tag{A6}$$

b. *Gamma noise on the synthesis rate ρ*

To get the factorial moments in this case, we integrate Eq. (A2) (after relabelling the parameters as described in the beginning of this section) over a gamma distribution in the synthesis rate:

$$R_n = \int_{\rho=0}^{+\infty} \left(\frac{\Gamma(k_{\text{on}}+n)\Gamma(k_{\text{on}}+k_{\text{off}})}{\Gamma(k_{\text{on}})\Gamma(k_{\text{on}}+k_{\text{off}}+n)}\rho^n\right) \frac{\rho^{a-1}\mathrm{e}^{-\frac{\rho}{b}}}{b^a\Gamma(a)} d\rho = \frac{\Gamma(k_{\text{on}}+n)\Gamma(k_{\text{on}}+k_{\text{off}})}{\Gamma(k_{\text{on}})\Gamma(k_{\text{on}}+k_{\text{off}}+n)} \int_{\rho=0}^{+\infty} \rho^n \frac{\rho^{a-1}\mathrm{e}^{-\frac{\rho}{b}}}{b^a\Gamma(a)} d\rho.$$

Using the normalisation of the gamma distribution, we immediately have $\int_0^{+\infty} \rho^{n+a-1}\mathrm{e}^{-\frac{\rho}{b}} d\rho = b^{a+n}\Gamma(a+n)$, which gives the following result:

$$R_n = \frac{\Gamma(k_{\text{on}}+n)\Gamma(k_{\text{on}}+k_{\text{off}})}{\Gamma(k_{\text{on}})\Gamma(k_{\text{on}}+k_{\text{off}}+n)} \frac{\Gamma(a+n)}{\Gamma(a)} b^n. \tag{A7}$$

This result has been previously derived in [34].

c. *Gamma noise on the switching on rate k_{on}*

To get the factorial moments in this case, we integrate Eq. (A2) over a gamma distribution in the switching on rate. The integral can be simplified by using Eq. (A3) to write the two quotients of gamma functions in terms of rising factorials:

$$R_n = \int_0^{+\infty} \rho^n \frac{\Gamma(k_{\text{on}}+n)}{\Gamma(k_{\text{on}})} \frac{\Gamma(k_{\text{on}}+k_{\text{off}})}{\Gamma(k_{\text{on}}+k_{\text{off}}+n)} \frac{k_{\text{on}}^{a-1}\mathrm{e}^{-\frac{k_{\text{on}}}{b}}}{\Gamma(a)b^a} dk_{\text{on}} = \rho^n \int_0^{+\infty} (k_{\text{on}})^{(n)} \cdot \frac{1}{(k_{\text{on}}+k_{\text{off}})^{(n)}} \cdot \frac{k_{\text{on}}^{a-1}\mathrm{e}^{-\frac{k_{\text{on}}}{b}}}{\Gamma(a)b^a} dk_{\text{on}}.$$

Next using Eqs. (A1) and (A4) we get rid of the rising factorials

$$\begin{aligned}
R_n =& \rho^n \int_0^{+\infty} \left(\sum_{i=0}^{n} \begin{bmatrix} n \\ i \end{bmatrix} k_{\text{on}}^i\right) \left(\frac{1}{(n-1)!} \sum_{j=0}^{n-1} \binom{n-1}{j} \frac{(-1)^j}{k_{\text{on}}+k_{\text{off}}+j}\right) \frac{k_{\text{on}}^{a-1}\mathrm{e}^{-\frac{k_{\text{on}}}{b}}}{\Gamma(a)b^a} dk_{\text{on}}, \\
=& \rho^n \int_0^{+\infty} \left(\sum_{i=0}^{n}\sum_{j=0}^{n-1} \begin{bmatrix} n \\ i \end{bmatrix} k_{\text{on}}^i \cdot \frac{1}{(n-1)!} \binom{n-1}{j} \frac{(-1)^j}{k_{\text{on}}+k_{\text{off}}+j} \cdot \frac{k_{\text{on}}^{a-1}\mathrm{e}^{-\frac{k_{\text{on}}}{b}}}{\Gamma(a)b^a}\right) dk_{\text{on}}, \\
=& \rho^n \sum_{i=0}^{n}\sum_{j=0}^{n-1} \begin{bmatrix} n \\ i \end{bmatrix} \cdot \frac{1}{(n-1)!} \binom{n-1}{j} (-1)^j \int_0^{+\infty} \left(\frac{k_{\text{on}}^i}{k_{\text{on}}+k_{\text{off}}+j} \cdot \frac{k_{\text{on}}^{a-1}\mathrm{e}^{-\frac{k_{\text{on}}}{b}}}{\Gamma(a)b^a}\right) dk_{\text{on}}, \\
=& \frac{\rho^n}{(n-1)!} \sum_{j=0}^{n-1}\sum_{i=0}^{n} \begin{bmatrix} n \\ i \end{bmatrix} \binom{n-1}{j} (-1)^j \frac{1}{b^a\Gamma(a)} \int_0^{+\infty} \frac{k_{\text{on}}^{a+i-1}\mathrm{e}^{-\frac{k_{\text{on}}}{b}}}{k_{\text{on}}+k_{\text{off}}+j} dk_{\text{on}}, \\
=& \frac{\rho^n}{(n-1)!} \sum_{j=0}^{n-1}\sum_{i=0}^{n} \begin{bmatrix} n \\ i \end{bmatrix} \binom{n-1}{j} (-1)^j \frac{b^{i-1}\Gamma(a+i)}{\Gamma(a)} \left(\frac{1}{b^{a+i-1}\Gamma(a+i)} \int_0^{+\infty} \frac{k_{\text{on}}^{a+i-1}\mathrm{e}^{-\frac{k_{\text{on}}}{b}}}{k_{\text{on}}+k_{\text{off}}+j} dk_{\text{on}}\right).
\end{aligned}$$

Finally using Eq. (A6) we obtain

$$R_n = \frac{\rho^n}{(n-1)!} \sum_{j=0}^{n-1}\sum_{i=0}^{n} \begin{bmatrix} n \\ i \end{bmatrix} \binom{n-1}{j} (-1)^j b^{i-1} \frac{\Gamma(a+i)}{\Gamma(a)} \mathrm{e}^{\frac{k_{\text{off}}+j}{b}} E_{a+i}\left(\frac{k_{\text{off}}+j}{b}\right). \tag{A8}$$

d. Gamma noise on the switching off rate k_{off}

To get the factorial moments in this case, we integrate Eq. (A2) over a gamma distribution in the switching off rate. The integral can be simplified by replacing the quotient $\frac{\Gamma(k_{\text{on}}+k_{\text{off}})}{\Gamma(k_{\text{on}}+k_{\text{off}}+n)}$ by a rising factorial (using Eq. (A3)):

$$R_n = \int_0^{+\infty} \rho^n \frac{\Gamma(k_{\text{on}}+n)}{\Gamma(k_{\text{on}})} \cdot \frac{\Gamma(k_{\text{on}}+k_{\text{off}})}{\Gamma(k_{\text{on}}+k_{\text{off}}+n)} \frac{k_{\text{off}}^{a-1} \mathrm{e}^{-\frac{k_{\text{off}}}{b}}}{\Gamma(a) b^a} dk_{\text{off}} = \rho^n \frac{\Gamma(k_{\text{on}}+n)}{\Gamma(k_{\text{on}})} \int_0^{+\infty} \frac{1}{(k_{\text{on}}+k_{\text{off}})^{(n)}} \cdot \frac{k_{\text{off}}^{a-1} \mathrm{e}^{-\frac{k_{\text{off}}}{b}}}{\Gamma(a) b^a} dk_{\text{off}}.$$

Simplifying further using Eq. (A4) (to get rid of the rising factorial), leads to

$$\begin{aligned}
R_n =& \rho^n \frac{\Gamma(k_{\text{on}}+n)}{\Gamma(k_{\text{on}})} \int_0^{+\infty} \left(\frac{1}{(n-1)!} \sum_{j=0}^{n-1} \binom{n-1}{j} \frac{(-1)^j}{k_{\text{on}}+k_{\text{off}}+j} \right) \frac{k_{\text{off}}^{a-1} \mathrm{e}^{-\frac{k_{\text{off}}}{b}}}{\Gamma(a) b^a} dk_{\text{off}}, \\
=& \frac{\rho^n}{(n-1)!} \frac{\Gamma(k_{\text{on}}+n)}{\Gamma(k_{\text{on}})} \int_0^{+\infty} \left(\sum_{j=0}^{n-1} \binom{n-1}{j} \frac{(-1)^j}{k_{\text{on}}+k_{\text{off}}+j} \cdot \frac{k_{\text{off}}^{a-1} \mathrm{e}^{-\frac{k_{\text{off}}}{b}}}{\Gamma(a) b^a} \right) dk_{\text{off}}, \\
=& \frac{\rho^n}{(n-1)!} \frac{\Gamma(k_{\text{on}}+n)}{\Gamma(k_{\text{on}})} \sum_{j=0}^{n-1} \binom{n-1}{j} (-1)^j \int_0^{+\infty} \left(\frac{1}{\Gamma(a) b^a} \cdot \frac{k_{\text{off}}^{a-1} \mathrm{e}^{-\frac{k_{\text{off}}}{b}}}{k_{\text{on}}+k_{\text{off}}+j} \right) dk_{\text{off}}, \\
=& \frac{\rho^n}{(n-1)!} \frac{\Gamma(k_{\text{on}}+n)}{\Gamma(k_{\text{on}})} \sum_{j=0}^{n-1} \binom{n-1}{j} (-1)^j \frac{1}{b^a \Gamma(a)} \int_0^{+\infty} \frac{k_{\text{off}}^{a-1} \mathrm{e}^{-\frac{k_{\text{off}}}{b}}}{k_{\text{on}}+k_{\text{off}}+j} dk_{\text{off}}, \\
=& \frac{\rho^n}{(n-1)!} \frac{\Gamma(k_{\text{on}}+n)}{\Gamma(k_{\text{on}})} \sum_{j=0}^{n-1} \binom{n-1}{j} (-1)^j \frac{1}{b} \left(\frac{1}{b^{a-1} \Gamma(a+1)} \int_0^{+\infty} \frac{k_{\text{off}}^{a-1} \mathrm{e}^{-\frac{k_{\text{off}}}{b}}}{k_{\text{on}}+k_{\text{off}}+j} dk_{\text{off}} \right).
\end{aligned}$$

Finally by using Eq. (A6) we find

$$R_n = \frac{\rho^n}{(n-1)!} \frac{\Gamma(k_{\text{on}}+n)}{\Gamma(k_{\text{on}})} \sum_{j=0}^{n-1} \binom{n-1}{j} \frac{(-1)^j}{b} \mathrm{e}^{\frac{k_{\text{on}}+j}{b}} E_a \left(\frac{k_{\text{on}}+j}{b} \right). \tag{A9}$$

Appendix B: Parameter details for Figure 3

The telegraph model distributions of mRNA counts (solid blue lines in the figure) are evaluated using the analytical result in [1] for the parameters estimated from three-moment matching procedure.

For (a), the parameters of the telegraph model with extrinsic noise are $\langle \rho \rangle = 3.92872$, $k_{\text{on}} = 2.74898$, $k_{\text{off}} = 5.20897$, $\sigma_\rho^2 = 1.42812$ and the inferred parameters of the telegraph model are $\alpha = 26.2939$, $\beta = 2.87887$ and $\theta = 52.8984$.

For (b), the parameters of the telegraph model with extrinsic noise are $\rho = 10.4132$, $\langle k_{\text{on}} \rangle = 179.703$, $k_{\text{off}} = 86.4225$, $\sigma_{k_{\text{on}}}^2 = 9213.27$ and the inferred parameters of the telegraph model are $\alpha = 8.70773$, $\beta = 4.88633$ and $\theta = 1.56185$.

For (c), the parameters of the telegraph model with extrinsic noise are $\rho = 30.2479$, $k_{\text{on}} = 20.8791$, $\langle k_{\text{off}} \rangle = 49.9974$, $\sigma_{k_{\text{off}}}^2 = 2187.17$ and the inferred parameters of the telegraph model are $\alpha = 50.5625$, $\beta = 2.12659$ and $\theta = 6.47778$.

Appendix C: Estimation bias in the limit of small extrinsic noise

In this section, we study estimation bias when the experimental data can be described by the telegraph model with small extrinsic noise, i.e. the limit of small cell-to-cell variability in parameter values. Our solution strategy is as follows. We will compute the factorial moments of mRNA counts for the telegraph model with small extrinsic noise by means of Taylor's approximation for moments [72]. For simplicity, we will consider only one parameter to vary amongst cells. Let this random parameter $\tilde{z}$ have mean z and a small standard deviation σ_z. We do not make further assumptions on the type of distribution of this parameter. By the aforementioned approximation for some function f of the random parameter, we have

$$\langle f(\tilde{z}) \rangle_{\tilde{z}} \approx f(z) + \frac{f''(z)}{2} \sigma_z^2 = f(z) \left(1 + \frac{f''(z)}{2 f(z)} \sigma_z^2 \right), \tag{C1}$$

where the angled brackets is understood to mean averaging over the distribution of the random parameter. Then we will equate the factorial moments of mRNA counts for the telegraph model with small extrinsic noise with the factorial moments of the standard telegraph (mimicking inference using this model) and solve them for the telegraph model parameters $\{\alpha, \beta, \theta\}$ using Eq. (2). Finally by taking the derivative of these parameters with respect to σ_z^2 when $\sigma_z = 0$, we will be able to determine the sign of estimation bias.

1. General theory

For the sake of generality, the beginning of the derivation will use a tilde over all three parameters. Later on we will consider only one parameter to be noisy, and the other two to be equal to their mean.

The first three factorial moments of mRNA counts as predicted by the standard telegraph model are given by (Appendix A):

$$
\begin{aligned}
f_1(\tilde{\rho}, \tilde{k}_{\text{on}}, \tilde{k}_{\text{off}}) &= \frac{\tilde{k}_{\text{on}}}{\tilde{k}_{\text{on}} + \tilde{k}_{\text{off}}} \tilde{\rho}, \\
f_2(\tilde{\rho}, \tilde{k}_{\text{on}}, \tilde{k}_{\text{off}}) &= \frac{\tilde{k}_{\text{on}}(\tilde{k}_{\text{on}} + 1)}{(\tilde{k}_{\text{on}} + \tilde{k}_{\text{off}})(\tilde{k}_{\text{on}} + \tilde{k}_{\text{off}} + 1)} \tilde{\rho}^2, \\
f_3(\tilde{\rho}, \tilde{k}_{\text{on}}, \tilde{k}_{\text{off}}) &= \frac{\tilde{k}_{\text{on}}(\tilde{k}_{\text{on}} + 1)(\tilde{k}_{\text{on}} + 2)}{(\tilde{k}_{\text{on}} + \tilde{k}_{\text{off}})(\tilde{k}_{\text{on}} + \tilde{k}_{\text{off}} + 1)(\tilde{k}_{\text{on}} + \tilde{k}_{\text{off}} + 2)} \tilde{\rho}^3.
\end{aligned}
$$

By applying Taylor's approximation Eq. (C1) on the functions f_i we obtain the extrinsic noise averaged factorial moments:

$$
R_i = \left\langle f_i(\tilde{\rho}, \tilde{k}_{\text{on}}, \tilde{k}_{\text{off}}) \right\rangle_{\tilde{z}} \approx f_i(\rho, k_{\text{on}}, k_{\text{off}}) \left(1 + \frac{\frac{\partial^2 f_i}{\partial z^2}(\rho, k_{\text{on}}, k_{\text{off}})}{2 f_i(\rho, k_{\text{on}}, k_{\text{off}})} \sigma_z^2 \right),
$$

where z can be ρ, k_{on} or k_{off}. By means of the notation $\hat{R}_i = f_i(\rho, k_{\text{on}}, k_{\text{off}})$, $c_i = \frac{1}{2 f_i(\rho, k_{\text{on}}, k_{\text{off}})} \frac{\partial^2 f_i}{\partial z^2}(\rho, k_{\text{on}}, k_{\text{off}})$, $K_2 = \frac{k_{\text{on}}+1}{k_{\text{on}}} \frac{k_{\text{on}}+k_{\text{off}}}{k_{\text{on}}+k_{\text{off}}+1}$ and $K_3 = \frac{k_{\text{on}}+2}{k_{\text{on}}} \frac{k_{\text{on}}+k_{\text{off}}}{k_{\text{on}}+k_{\text{off}}+2}$, these expressions can be compactly written as

$$
\begin{aligned}
R_1 &\approx \hat{R}_1 (1 + c_1 \sigma_z^2), \\
R_2 &\approx \hat{R}_2 (1 + c_2 \sigma_z^2) = K_2 \hat{R}_1^2 (1 + c_2 \sigma_z^2), \\
R_3 &\approx \hat{R}_3 (1 + c_3 \sigma_z^2) = K_3 \hat{R}_1 \hat{R}_2 (1 + c_3 \sigma_z^2).
\end{aligned}
$$

Substituting these in Eq. (2) we find the telegraph model estimates of the three transcriptional parameters:

$$
\begin{aligned}
\alpha &= \hat{R}_1 \frac{\left(2(1 + c_1\sigma_z^2)^2 - K_2(1 + c_2\sigma_z^2)\right) K_3(1 + c_3\sigma_z^2) - K_2(1 + c_1\sigma_z^2)(1 + c_2\sigma_z^2)^2}{\left((1 + c_1\sigma_z^2)^2 - 2K_2(1 + c_2\sigma_z^2)\right)(1 + c_2\sigma_z^2) + K_3(1 + c_1\sigma_z^2)(1 + c_3\sigma_z^2)}, \\
\beta &= \frac{2(1 + c_1\sigma_z^2)\left(K_2(1 + c_2\sigma_z^2)^2 - K_3(1 + c_1\sigma_z^2)(1 + c_3\sigma_z^2)\right)}{\left(2(1 + c_1\sigma_z^2)^2 - K_2(1 + c_2\sigma_z^2)\right) K_3(1 + c_3\sigma_z^2) - K_2(1 + c_1\sigma_z^2)(1 + c_2\sigma_z^2)^2}, \\
\theta &= \frac{2\left[(1 + c_1\sigma_z^2)^2 - K_2(1 + c_2\sigma_z^2)\right]\left[(1 + c_1\sigma_z^2)(1 + c_2\sigma_z^2) - K_3(1 + c_3\sigma_z^2)\right]}{\left[\left((1 + c_1\sigma_z^2)^2 - 2K_2(1 + c_2\sigma_z^2)\right)(1 + c_2\sigma_z^2) + K_3(1 + c_1\sigma_z^2)(1 + c_3\sigma_z^2)\right]} \\
&\qquad \times \frac{\left[K_3(1 + c_1\sigma_z^2)(1 + c_3\sigma_z^2) - K_2(1 + c_2\sigma_z^2)^2\right]}{\left[\left(2(1 + c_1\sigma_z^2)^2 - K_2(1 + c_2\sigma_z^2)\right)K_3(1 + c_3\sigma_z^2) - K_2(1 + c_1\sigma_z^2)(1 + c_2\sigma_z^2)^2\right]}, \\
\frac{\alpha}{\theta} &= \frac{\frac{\hat{R}_1}{2}\left[\left(2(1 + c_1\sigma_z^2)^2 - K_2(1 + c_2\sigma_z^2)\right)K_3(1 + c_3\sigma_z^2) - K_2(1 + c_1\sigma_z^2)(1 + c_2\sigma_z^2)^2\right]^2}{\left[(1 + c_1\sigma_z^2)^2 - K_2(1 + c_2\sigma_z^2)\right]\left[(1 + c_1\sigma_z^2)(1 + c_2\sigma_z^2) - K_3(1 + c_3\sigma_z^2)\right]\left[K_3(1 + c_1\sigma_z^2)(1 + c_3\sigma_z^2) - K_2(1 + c_2\sigma_z^2)^2\right]}.
\end{aligned}
$$

It is easy to show that if $P(x) = \sum_{i=0}^{m} a_i x^i$, $Q(x) = \sum_{i=0}^{n} b_i x^i$ and $F(x) = \frac{P(x)}{Q(x)}$, then $F'(x) = \frac{P'(x)Q(x) - P(x)Q'(x)}{Q(x)^2}$ and $F'(0) = \frac{a_1 \cdot b_0 - a_0 \cdot b_1}{b_0^2}$.

Using this result, the derivatives with respect to σ_z^2, evaluated at 0, are given by

$$
\left.\frac{\partial \alpha}{\partial \sigma_z^2}\right|_{\sigma_z^2=0} = \frac{\rho}{2}\left(\left[k_{\text{on}}(3k_{\text{on}} + 3k_{\text{off}} + 4) + 2k_{\text{off}} + 2\right] c_1 - \left[k_{\text{on}}(3k_{\text{on}} + 3k_{\text{off}} + 8) + 4k_{\text{off}} + 6\right] c_2 + (k_{\text{on}} + 2)(k_{\text{on}} + k_{\text{off}} + 2) c_3\right),
$$

$$
\begin{aligned}
\left.\frac{\partial\beta}{\partial\sigma_z^2}\right|_{\sigma_z^2=0} &= \frac{k_{\rm on}(k_{\rm on}+1)(k_{\rm on}+2)}{2k_{\rm off}}\Big(\left[3k_{\rm on}+3k_{\rm off}+2\right]c_1-\left[3k_{\rm on}+3k_{\rm off}+4\right]c_2+\left[k_{\rm on}+k_{\rm off}+2\right]c_3\Big),\\
\left.\frac{\partial\theta}{\partial\sigma_z^2}\right|_{\sigma_z^2=0} &= \frac{1}{2(k_{\rm on}+k_{\rm off})}\big(\big[6k_{\rm on}^4+15k_{\rm on}^3(1+k_{\rm off})+2k_{\rm off}(2+3k_{\rm off}+k_{\rm off}^2)\\
&\quad +6k_{\rm on}^2(2+5k_{\rm off}+2k_{\rm off}^2)+k_{\rm on}(4+18k_{\rm off}+17k_{\rm off}^2+3k_{\rm off}^3)\big]c_1\\
&\quad -(k_{\rm on}+k_{\rm off})\left[6k_{\rm on}^3+3k_{\rm on}^2(7+3k_{\rm off})+4(2+3k_{\rm off}+k_{\rm off}^2)+3k_{\rm on}(8+7k_{\rm off}+k_{\rm off}^2)\right]c_2\\
&\quad +(2+k_{\rm on})(k_{\rm on}+2k_{\rm on}^2+k_{\rm off}+3k_{\rm on}k_{\rm off}+k_{\rm off}^2)(k_{\rm on}+k_{\rm off}+2)c_3)\big),\\
\left.\frac{\partial\frac{\alpha}{\theta}}{\partial\sigma_z^2}\right|_{\sigma_z^2=0} &= -\frac{\rho}{2k_{\rm off}^2}\big(\big[6k_{\rm on}^3+4(1+k_{\rm off})+3k_{\rm on}^2(5+2k_{\rm off})+k_{\rm on}(12+11k_{\rm off})\big]c_1\\
&\quad -\left[8+6k_{\rm on}^3+6k_{\rm off}+3k_{\rm on}^2(7+2k_{\rm off})+k_{\rm on}(24+13k_{\rm off})\right]c_2+(k_{\rm on}+2)(2k_{\rm on}+1)(k_{\rm on}+k_{\rm off}+2)c_3\big).
\end{aligned}
\tag{C2}
$$

2. Variable synthesis rate: $\tilde{z}=\tilde{\rho}$

By the definition of the formulae for c_i, we have:

$$
c_1=0, c_2=\frac{1}{\rho^2}, c_3=\frac{3}{\rho^2}. \tag{C3}
$$

Substituting in Eq. (C2) we obtain:

$$
\begin{aligned}
\left.\frac{\partial\alpha}{\partial\sigma_\rho^2}\right|_{\sigma_\rho^2=0} &= \frac{2k_{\rm on}+k_{\rm off}+3}{\rho}\\
\left.\frac{\partial\beta}{\partial\sigma_\rho^2}\right|_{\sigma_\rho^2=0} &= \frac{k_{\rm on}(k_{\rm on}+1)(k_{\rm on}+2)}{\rho^2k_{\rm off}}\\
\left.\frac{\partial\theta}{\partial\sigma_\rho^2}\right|_{\sigma_\rho^2=0} &= \frac{3k_{\rm on}^2+(k_{\rm off}+2)(3k_{\rm on}+k_{\rm off}+1)}{\rho^2}\\
\left.\frac{\partial\frac{\alpha}{\theta}}{\partial\sigma_\rho^2}\right|_{\sigma_\rho^2=0} &= -\frac{2+3k_{\rm on}^2+k_{\rm on}(6+k_{\rm off})}{k_{\rm off}^2\rho}
\end{aligned}
$$

Hence it follows that in the presence of small extrinsic noise in the synthesis rate, use of the telegraph model for inference will lead to an overestimation of the synthesis, switching on and switching off rates and an underestimation of the burst size. This agrees with Fig. 2(a).

3. Variable switching on rate: $\tilde{z}=\tilde{k}_{\rm on}$

By the definition of the formulae for c_i, we have:

$$
\begin{aligned}
c_1 &= -\frac{k_{\rm off}}{k_{\rm on}(k_{\rm on}+k_{\rm off})^2},\\
c_2 &= \frac{k_{\rm off}}{k_{\rm on}}\frac{(k_{\rm on}+k_{\rm off}+1)^3(k_{\rm off}-1)-(k_{\rm on}+k_{\rm off})^3(k_{\rm off}+1)}{(k_{\rm on}+1)(k_{\rm on}+k_{\rm off})^2(k_{\rm on}+k_{\rm off}+1)^2},\\
c_3 &= \frac{k_{\rm off}(k_{\rm off}+1)(k_{\rm on}+k_{\rm off})(k_{\rm on}+k_{\rm off}+1)(k_{\rm on}+k_{\rm off}+2)}{k_{\rm on}(k_{\rm on}+1)(k_{\rm on}+2)}\\
&\quad\times\left[\frac{k_{\rm off}-1}{(k_{\rm on}+k_{\rm off}+1)^3}-\frac{k_{\rm off}+2}{2}\left(\frac{1}{(k_{\rm on}+k_{\rm off})^3}+\frac{1}{(k_{\rm on}+k_{\rm off}+2)^3}\right)\right].
\end{aligned}
$$

Substituting in Eq. (C2) we obtain:

$$
\left.\frac{\partial\alpha}{\partial\sigma_{k_{\rm on}}^2}\right|_{\sigma_{k_{\rm on}}^2=0} = -\frac{\rho k_{\rm off}(k_{\rm off}+1)(2k_{\rm on}+k_{\rm off}+2)}{2k_{\rm on}(k_{\rm on}+1)(k_{\rm on}+k_{\rm off}+1)(k_{\rm on}+k_{\rm off}+2)},
$$

$$\left.\frac{\partial\beta}{\partial\sigma^2_{k_{\rm on}}}\right|_{\sigma^2_{k_{\rm on}}=0} = -\frac{2k_{\rm on}^2(2k_{\rm off}+3)+k_{\rm on}(7k_{\rm off}^2+17k_{\rm off}+12)+3k_{\rm off}^3+11k_{\rm off}^2+12k_{\rm off}+4}{2(k_{\rm on}+k_{\rm off})(k_{\rm on}+k_{\rm off}+1)(k_{\rm on}+k_{\rm off}+2)},$$

$$\left.\frac{\partial\theta}{\partial\sigma^2_{k_{\rm on}}}\right|_{\sigma^2_{k_{\rm on}}=0} = -\frac{k_{\rm off}(k_{\rm off}+1)(6k_{\rm on}^2+6k_{\rm on}(k_{\rm off}+2)+k_{\rm off}^2+5k_{\rm off}+6)}{2k_{\rm on}(k_{\rm on}+1)(k_{\rm on}+k_{\rm off}+1)(k_{\rm on}+k_{\rm off}+2)},$$

$$\left.\frac{\partial\frac{\alpha}{\theta}}{\partial\sigma^2_{k_{\rm on}}}\right|_{\sigma^2_{k_{\rm on}}=0} = \frac{\rho}{k_{\rm off}}\frac{(k_{\rm off}+1)(6k_{\rm on}^2+4k_{\rm on}(3+k_{\rm off})+3(k_{\rm off}+2))}{2k_{\rm on}(k_{\rm on}+1)(k_{\rm on}+k_{\rm off}+1)(k_{\rm on}+k_{\rm off}+2)}.$$

Hence it follows that in the presence of small extrinsic noise in the switching on rate, use of the telegraph model for inference will lead to an underestimation of the synthesis, switching on and switching off rates and an overestimation of the burst size. This agrees with Fig. 2(b).

4. Variable switching off rate: $\tilde{z} = \tilde{k}_{\rm off}$

By the definition of the formulae for c_i, we have:

$$c_1 = \frac{1}{(k_{\rm on}+k_{\rm off})^2},$$

$$c_2 = \frac{(k_{\rm on}+k_{\rm off}+1)^3-(k_{\rm on}+k_{\rm off})^3}{(k_{\rm on}+k_{\rm off})^2(k_{\rm on}+k_{\rm off}+1)^2},$$

$$c_3 = \frac{(k_{\rm on}+k_{\rm off}+1)^3(k_{\rm on}+k_{\rm off}+2)^3-2(k_{\rm on}+k_{\rm off})^3(k_{\rm on}+k_{\rm off}+2)^3+(k_{\rm on}+k_{\rm off})^3(k_{\rm on}+k_{\rm off}+1)^3}{(k_{\rm on}+k_{\rm off})^2(k_{\rm on}+k_{\rm off}+1)^2(k_{\rm on}+k_{\rm off}+2)^2}.$$

Substituting in Eq. (C2) we obtain:

$$\left.\frac{\partial\alpha}{\partial\sigma^2_{k_{\rm off}}}\right|_{\sigma^2_{k_{\rm off}}=0} = \frac{\rho}{2}\cdot\frac{k_{\rm on}+2k_{\rm off}+2}{(k_{\rm on}+k_{\rm off}+1)(k_{\rm on}+k_{\rm off}+2)},$$

$$\left.\frac{\partial\beta}{\partial\sigma^2_{k_{\rm off}}}\right|_{\sigma^2_{k_{\rm off}}=0} = -\frac{k_{\rm on}(k_{\rm on}+1)(k_{\rm on}+2)}{2k_{\rm off}(k_{\rm on}+k_{\rm off}+1)(k_{\rm on}+k_{\rm off}+2)},$$

$$\left.\frac{\partial\theta}{\partial\sigma^2_{k_{\rm off}}}\right|_{\sigma^2_{k_{\rm off}}=0} = \frac{3k_{\rm on}^2(k_{\rm off}-1)+k_{\rm on}(5k_{\rm off}^2-3k_{\rm off}-8)+2k_{\rm off}^3-8k_{\rm off}-4}{2(k_{\rm on}+k_{\rm off})(k_{\rm on}+k_{\rm off}+1)(k_{\rm on}+k_{\rm off}+2)},$$

$$\left.\frac{\partial\frac{\alpha}{\theta}}{\partial\sigma^2_{k_{\rm off}}}\right|_{\sigma^2_{k_{\rm off}}=0} = -\frac{\rho}{2k_{\rm off}^2}\frac{(k_{\rm on}^2(2k_{\rm off}-3)+k_{\rm on}(2k_{\rm off}^2-5k_{\rm off}-8)-2k_{\rm off}^2-8k_{\rm off}-4)}{(k_{\rm on}+k_{\rm off})(k_{\rm on}+k_{\rm off}+1)(k_{\rm on}+k_{\rm off}+2)}.$$

The first and second expressions are respectively positive and negative, implying that the telegraph model overestimates the synthesis rate and underestimates the switching on rate. However the third and the fourth expressions can be positive or negative, depending on the parameter values, and hence both over and underestimation are possible for the switching off rate and the burst size. This agrees with Fig. 2(c). Comparing the box-whisker plots in the middle and right columns, it is clear that the sign of the estimation bias for the burst size changes if promoter switching is constrained to be slow. This is also the case for the switching off rate but cannot be appreciated from the plots (the maximum value of $\theta/\langle k_{\rm off}\rangle$ is slightly greater than 1 but the upper fence of the box-whisker plot only shows the 90th percentile).

The results of the small-noise theory are summarised in Table II.

Appendix D: Estimation bias when the synthesis rate is lognormal distributed

Consider a population of cells where the synthesis rate ρ varies from cell-to-cell. Furthermore we assume that this variation is described by a lognormal distribution with parameters (μ, s). The mean of this distribution is $\langle\rho\rangle = \mathrm{e}^{\mu+\frac{s^2}{2}}$,

	Noise on ρ	Noise on k_{on}	Noise on k_{off}
α	overestimation	underestimation	overestimation
β	overestimation	underestimation	underestimation
θ	overestimation	underestimation	indeterminate
$\frac{\alpha}{\theta}$	underestimation	overestimation	indeterminate

TABLE II. Summary of the small extrinsic noise theory showing the type of bias in the telegraph model parameters $(\alpha, \beta, \theta, \alpha/\theta)$ when these are inferred from data generated by a telegraph model with small extrinsic noise in one of the parameters $(\rho, k_{\text{on}}, k_{\text{off}})$.

and the second moment is $\langle \rho^2 \rangle = \mathrm{e}^{2\mu+2s^2}$. The coefficient of variation of the synthesis rate is equal to $CV_\rho^2 = \mathrm{e}^{s^2} - 1$. The factorials moments of mRNA counts can be straightforwardly computed [34]

$$\begin{aligned}
R_1 =& \frac{k_{\text{on}}}{k_{\text{on}} + k_{\text{off}}} \mathrm{e}^{\mu + \frac{s^2}{2}}, \\
R_2 =& \frac{k_{\text{on}}(k_{\text{on}}+1)}{(k_{\text{on}} + k_{\text{off}})(k_{\text{on}} + k_{\text{off}} + 1)} \mathrm{e}^{2\mu+2s^2} = \frac{k_{\text{on}} + k_{\text{off}}}{k_{\text{on}}} \cdot \frac{k_{\text{on}}+1}{k_{\text{on}} + k_{\text{off}} + 1} R_1^2 (CV_\rho^2 + 1), \\
R_3 =& \frac{k_{\text{on}}(k_{\text{on}}+1)(k_{\text{on}}+2)}{(k_{\text{on}} + k_{\text{off}})(k_{\text{on}} + k_{\text{off}} + 1)(k_{\text{on}} + k_{\text{off}} + 2)} \mathrm{e}^{3\mu + \frac{9s^2}{2}} = \frac{k_{\text{on}} + k_{\text{off}}}{k_{\text{on}}} \cdot \frac{k_{\text{on}}+2}{k_{\text{on}} + k_{\text{off}} + 2} R_1 R_2 (CV_\rho^2 + 1)^2.
\end{aligned} \tag{D1}$$

Substituting in Eq. (2), we obtain the parameters inferred by fitting a telegraph model:

$$\alpha = \langle \rho \rangle \frac{1 + CV_\rho^2 (k_{\text{on}} + 2) + \frac{CV_\rho^4}{2} \frac{(k_{\text{on}}+1)(k_{\text{on}}+2)(k_{\text{on}}+k_{\text{off}})}{k_{\text{off}}}}{1 - CV_\rho^2 (k_{\text{on}} + k_{\text{off}}) - \frac{CV_\rho^4}{2} \frac{(k_{\text{on}}+2)(k_{\text{on}}+k_{\text{off}})(k_{\text{on}}+k_{\text{off}}+1)}{k_{\text{off}}}},$$

$$\beta = k_{\text{on}} \frac{k_{\text{off}} + CV_\rho^2 (k_{\text{on}} + 2)(k_{\text{on}} + k_{\text{off}} + 1)}{k_{\text{off}} + CV_\rho^2 k_{\text{off}} (k_{\text{on}} + 2) + \frac{CV_\rho^4}{2} (k_{\text{on}} + k_{\text{off}})(k_{\text{on}} + 1)(k_{\text{on}} + 2)},$$

$$\begin{aligned}
\theta =& \frac{k_{\text{off}} + CV_\rho^2 (k_{\text{on}} + k_{\text{off}})(2k_{\text{on}} + 3) + \frac{CV_\rho^4}{2} \frac{(k_{\text{on}}+2)(k_{\text{on}}+k_{\text{off}})}{k_{\text{off}}} (2(k_{\text{on}} + 2)(k_{\text{on}} + k_{\text{off}}) + k_{\text{off}}) + \frac{CV_\rho^6}{2} \frac{(k_{\text{on}}+1)(k_{\text{on}}+2)(k_{\text{on}}+k_{\text{off}})^2}{k_{\text{off}}}}{1 - CV_\rho^2 (k_{\text{on}} + k_{\text{off}}) - \frac{CV_\rho^4}{2} \frac{(k_{\text{on}}+2)(k_{\text{on}}+k_{\text{off}})(k_{\text{on}}+k_{\text{off}}+1)}{k_{\text{off}}}} \\
& \times \frac{k_{\text{off}} + CV_\rho^2 (k_{\text{on}} + 2)(k_{\text{on}} + k_{\text{off}} + 1)}{k_{\text{off}} + CV_\rho^2 k_{\text{off}} (k_{\text{on}} + 2) + \frac{CV_\rho^4}{2} (k_{\text{on}} + k_{\text{off}})(k_{\text{on}} + 1)(k_{\text{on}} + 2)}.
\end{aligned} \tag{D2}$$

It is immediately evident that β is always positive, and that the sign of the other two parameters only depends on the sign of their denominator. The discriminant of the denominator is

$$\Delta = (k_{\text{on}} + k_{\text{off}})^2 + 2 \frac{(k_{\text{on}} + 2)(k_{\text{on}} + k_{\text{off}})(k_{\text{on}} + k_{\text{off}} + 1)}{k_{\text{off}}} = (k_{\text{on}} + k_{\text{off}})^2 \left(1 + 2 \frac{(k_{\text{on}} + 2)(k_{\text{on}} + k_{\text{off}} + 1)}{k_{\text{off}}(k_{\text{on}} + k_{\text{off}})} \right) > 0$$

and the two solutions are given by

$$cv_\pm^2 = \frac{-(k_{\text{on}} + k_{\text{off}}) \pm (k_{\text{on}} + k_{\text{off}}) \sqrt{1 + 2 \frac{(k_{\text{on}}+2)(k_{\text{on}}+k_{\text{off}}+1)}{k_{\text{off}}(k_{\text{on}}+k_{\text{off}})}}}{\frac{(k_{\text{on}}+2)(k_{\text{on}}+k_{\text{off}})(k_{\text{on}}+k_{\text{off}}+1)}{k_{\text{off}}}},$$

where cv_-^2 is negative and $cv_+^2 = \frac{k_{\text{off}}}{(k_{\text{on}}+2)(k_{\text{on}}+k_{\text{off}}+1)} \left(\sqrt{1 + 2 \frac{(k_{\text{on}}+2)(k_{\text{on}}+k_{\text{off}}+1)}{k_{\text{off}}(k_{\text{on}}+k_{\text{off}})}} - 1 \right)$ is positive. Hence the parameters α and θ are positive provided the coefficient of variation of the synthesis rate CV_ρ^2 is less than cv_+^2.

If we compare this critical noise threshold with that for gamma-distributed noise (Section III) we find that it is always smaller

$$cv_+^2 = \frac{k_{\text{off}}}{(k_{\text{on}} + 2)(k_{\text{on}} + k_{\text{off}} + 1)} \left(\sqrt{1 + 2 \frac{(k_{\text{on}} + 2)(k_{\text{on}} + k_{\text{off}} + 1)}{k_{\text{off}}(k_{\text{on}} + k_{\text{off}})}} - 1 \right)$$

$$\leq \frac{k_{\text{off}}}{(k_{\text{on}}+2)(k_{\text{on}}+k_{\text{off}}+1)} \cdot \left(\frac{(k_{\text{on}}+2)(k_{\text{on}}+k_{\text{off}}+1)}{k_{\text{off}}(k_{\text{on}}+k_{\text{off}})} \right) = \frac{1}{k_{\text{on}}+k_{\text{off}}} \tag{D3}$$

A summary of the properties of the estimated rates is as follows:

- α is positive only when the extrinsic noise size is below the threshold value cv_+^2. In this case it is an increasing function of CV_ρ^2 and always larger than $\langle\rho\rangle$. It diverges as $(CV_\rho^2 - cv_+^2)^{-1}$ near the critical noise threshold.
- β is always positive, independent of the size of the extrinsic noise. It increases monotonically with the extrinsic noise size on $[0, cv_+^2]$; in this region, it is larger than k_{on} and reaches its maximum when $CV_\rho^2 = cv_+^2$.
- θ is positive only when the extrinsic noise size is below the threshold value cv_+^2. In this case it is an increasing function of CV_ρ^2 and always larger than k_{off}. It diverges as $(CV_\rho^2 - cv_+^2)^{-1}$ near the critical noise threshold.

Appendix E: Maximum likelihood estimation and the generation of simulated data

1. Maximum Likelihood estimation using the standard telegraph model

Given the steady-state distribution of mRNA counts predicted by the standard telegraph model

$$P(m) = \frac{\alpha^m}{m!} \frac{\Gamma(\beta+m)}{\Gamma(\beta)} \frac{\Gamma(\beta+\theta)}{\Gamma(\beta+\theta+m)} {}_1F_1(\beta+m; \beta+\theta+m; -\alpha), \tag{E1}$$

and the data vector $\vec{m} = m_1, m_2, .., m_N$ (where m_i is the mRNA counts in cell i), the procedure consists of finding the parameter set $\{\alpha, \beta, \theta\}$ that maximises the log-likelihood function defined as

$$\mathcal{L} = \log \prod_{i=1}^{N} P(m_i) = \sum_{i=1}^{N} \log P(m_i). \tag{E2}$$

The optimisation was performed using Mathematica's FindMaximum function. Good initial guesses for the optimisation procedure can sometimes speed up estimation [46]; for our numerical experiments, we did not find a large improvement in speed and the results were the same if we simply constrained the optimisation to search over positive parameter values.

2. Simulating mRNA count data

To mimic experimental mRNA count data, we need to sample from the steady-state distribution of the telegraph model with extrinsic noise. For this purpose we made use of the equivalence between the standard telegraph model's solution and the Beta-Poisson compound distribution (discussed in Appendix A). For example if we want to generate data for the telegraph model with gamma distributed synthesis rate (Fig. 2(a)), the mRNA count for each cell was obtained by drawing a Poisson distributed random number Poisson$(r_1 r_2)$ where r_1 is a random number Beta$(k_{\text{on}}, k_{\text{off}})$ and r_2 is a random number Gamma$(\langle\rho\rangle^2/\sigma_\rho^2, \sigma_\rho^2/\langle\rho\rangle)$. Similarly to generate data for the telegraph model with gamma distributed switching on rate (Fig. 2(b)), the mRNA count for each cell can be obtained by drawing a Poisson distributed random number Poisson(ρr_1) where r_1 is a random number Beta(r_2, k_{off}) and r_2 is a random number Gamma$(\langle k_{\text{on}}\rangle^2/\sigma_{k_{\text{on}}}^2, \sigma_{k_{\text{on}}}^2/\langle k_{\text{on}}\rangle)$. Note that $\langle i\rangle$ and σ_i^2 denote the mean and variance of parameter i, respectively.

Appendix F: Covariance matrix analysis of scRNA-seq data

1. Extrinsic noise on the synthesis rate due to coupling of transcription to cell volume

Consider a a population of cells characterized by a distribution $f(V)$ of the cell volume. Furthermore consider the stochastic expression of two genes G_1 and G_2 whose expression dynamics in a cell of volume V is described by the following modified telegraph model:

$$G_{1,\text{OFF}} \underset{k_{\text{off},1}}{\overset{k_{\text{on},1}}{\rightleftharpoons}} G_{1,\text{ON}} \xrightarrow{\rho_1 V} G_{1,\text{ON}} + M_1, \qquad M_1 \xrightarrow{1} \emptyset \tag{F1}$$

$$G_{2,\text{OFF}} \underset{k_{\text{off},2}}{\overset{k_{\text{on},2}}{\rightleftharpoons}} G_{2,\text{ON}} \xrightarrow{\rho_2 V} G_{2,\text{ON}} + M_2, \qquad M_2 \xrightarrow{1} \emptyset, \tag{F2}$$

where M_i is mRNA from gene G_i. Note that the rates for each gene are normalised by the mRNA degradation rate for that gene (hence the latter is set to 1).

We now derive the expression of the covariance $\text{Cov}(m_1, m_2)$ between the mRNA counts m_1 and m_2 of the two genes. Given a cell with volume V, we have the following moments:

$$\begin{aligned}
\langle m_i \rangle_V &= \sum_{m_1}\sum_{m_2} m_i P(m_1, m_2|V) = \frac{k_{\text{on,i}}}{k_{\text{on,i}} + k_{\text{off,i}}}\rho_i V, \\
\langle m_1 m_2 \rangle_V &= \sum_{m_1}\sum_{m_2} m_1 m_2 P(m_1, m_2|V) = \sum_{m_1} m_1 P(m_1|V) \sum_{m_2} m_2 P(m_2|V) = \langle m_1 \rangle_V \langle m_2 \rangle_V, \\
&= \frac{k_{\text{on},1}}{k_{\text{on},1} + k_{\text{off},1}} \frac{k_{\text{on},2}}{k_{\text{on},2} + k_{\text{off},2}} \rho_1 \rho_2 V^2,
\end{aligned} \tag{F3}$$

where we have used the independence of the two genes in the last line. Hence it follows that by taking the average over the extrinsic noise, i.e. integrating over the cell size distribution, we obtain:

$$\begin{aligned}
\text{Cov}(m_1, m_2) &= \int \langle m_1 m_2 \rangle_V f(V) dV - \int \langle m_1 \rangle_V f(V) dV \int \langle m_2 \rangle_V f(V) dV, \\
&= \frac{k_{\text{on},1}}{k_{\text{on},1} + k_{\text{off},1}} \frac{k_{\text{on},2}}{k_{\text{on},2} + k_{\text{off},2}} \rho_1 \rho_2 \left(\langle V^2 \rangle - \langle V \rangle^2 \right), \\
&= \langle m_1 \rangle \langle m_2 \rangle \text{CV}_V^2,
\end{aligned} \tag{F4}$$

where $\langle m_i \rangle$ is the integral of $\langle m_i \rangle_V$ over the volume distribution $f(V)$ and CV_V^2 is the coefficient of variation of the volume distribution. It is interesting that extrinsic noise, i.e. variability in the synthesis rate due to variability in the cell volume, causes the covariance to be always positive. Hence plotting the covariance of two genes versus the product of their mean expression should lead to a straight line with a slope equal to the size of extrinsic noise; in Fig. 6(b) we confirm this using the scRNA-dataset in [6].

2. Extrinsic noise on the burst size or frequency for bursty genes

We consider some source of extrinsic noise (cell-specific) which acts linearly on a bursty gene either via modulation of its burst size or burst frequency. We will denote this global factor by V (does not necessarily mean volume).

For noise in the burst size, the results following directly from those above. The burst size is the synthesis rate divided by the switching off rate hence noise in the synthesis rate will lead to noise in the burst size. Thus Eq. (F4) immediately follows.

For noise in the burst frequency and transcriptional bursting, we consider the reaction scheme (F2) with the changes: $\rho_i V \to \rho_i$, $k_{\text{on,i}} \to k_{\text{on,i}} V$. Then taking the limit of transcriptional bursting, i.e. $k_{\text{on,i}} V_{max} \ll k_{\text{off,i}}$, one finds that $\langle m_i \rangle_V \propto V$ and $\langle m_1 m_2 \rangle_V \propto V^2$ which implies that Eq. (F4) holds. Note that V_{max} is the maximum value of the extrinsic parameter V.

Appendix G: Normalisation of scRNA-seq data can introduce estimation bias

In [61], for a given gene, a telegraph model with extrinsic noise in the synthesis rate is proposed to describe its stochastic expression and to take into account the capture-efficiency of scRNA-seq technology. Expression from cell j is assumed to be described by reaction scheme in Fig. 2(a) left with the change $\rho \to \rho\gamma_j$ where γ_j is a cell-specific factor that is the product of the capture efficiency and the volume for cell j. A recipe is then described for computing γ_j for each cell from the raw scRNA-seq data. Given these cell-specific factors and mRNA count data from each cell their goal is to infer the switching rates k_{on} and k_{off}, and the parameter ρ.

One of the heuristic methods proposed involves a modification of the moment-matching procedure we described in Section II. It is proposed that one uses Eq. (2) with the difference that the ith factorial moment R_i is computed not for the observed counts data but rather for the counts data normalised by the cell-specific factor, i.e. m_i (observed counts in cell i) is replaced by $x_i = m_i/\gamma_i$. Since we can sample from the steady-state distribution of the telegraph

model using the Beta-Poisson distribution, it follows that for cell i the normalised counts are given by

$$x_i \sim \frac{\text{Poisson}(\rho\gamma_i r_i)}{\gamma_i}, \tag{G1}$$

where r_i is a random variable distributed according to $\text{Beta}(k_{\text{on}}, k_{\text{off}})$. The heuristic method then is equivalent to assuming that the moments of x_i are precisely the same as the moments of the random variable

$$y_i \sim \text{Poisson}(\alpha s_i), \tag{G2}$$

where s_i is a random variable distributed according to $\text{Beta}(\beta, \theta)$. Perfect inference by the heuristic method would imply $\alpha = \rho$, $\beta = k_{\text{on}}$ and $\theta = k_{\text{off}}$. By inspection of Eqs. (G1) and (G2), this only occurs if the following statement is true:

$$\frac{\text{Poisson}(\rho\gamma_i r_i)}{\gamma_i} \sim \text{Poisson}(\rho r_i) \tag{G3}$$

It is clear that this cannot be generally true for any distribution of the cell-specific factor γ_i. In fact it is not even true if γ_i is a constant independent of i. However the equivalence is achieved asymptotically in the limit of large $\rho\gamma_i r_i$ and large ρr_i since then the Poisson random variables can be replaced by their mean, i.e. when the mean expression level of a gene is high enough, the heuristic moment-matching method of [61] will give reliable results. In Table III we numerically show that the accuracy increases with the mean expression level thus validating our theoretical results.

We note that the errors in this method can be overcome by using a different moment-matching method. Under the assumption that scRNA-seq data is accurately described by a telegraph model with extrinsic noise on the synthesis rate (as assumed in the beginning of this section and in [61]) then exact parameter inference can be implemented by equating the first three moments of this model (derived for any distribution of extrinsic noise in the synthesis rate in [34]) with the first three raw moments of count data and then solving the three resultant equations simultaneously for $\rho, k_{\text{on}}, k_{\text{off}}$. If gene expression is bursty then this procedure is equivalent to the method described in Section IV.C with the burst size for gene i in cell j proportional to the cell-specific factor γ_j. Of course the accuracy of this method relies on whether scRNA-seq data truly suffers only from noise on the synthesis rate (although generally unlikely this may be true for the data in [6]) and also on the estimation of the cell-specific factors γ_j.

ρ	α/ρ	β/k_{on}	θ/k_{off}	Mean Expression
10	0.51	0.92	0.24	3.35
20	0.72	1.04	0.61	6.66
30	0.81	1.04	0.75	9.97
100	0.94	1.02	0.93	33.31

TABLE III. Numerical investigation of the accuracy of the parameter inference method in [61] based on moment-matching of the telegraph model to the moments of normalised mRNA counts. We simulated data for 10^5 cells by sampling from the Beta-Poisson distribution with synthesis rate $\rho\gamma_i$, switching on rate $k_{\text{on}} = 1$ and switching off rate $k_{\text{off}} = 2$, where γ_i is a gamma-distributed random variable with mean 20 and coefficient of variation squared = 1. The count data from cell i is then divided by γ_i leading to normalised counts. Substituting the first three factorial moments of the latter in Eq. (2) we obtain parameter estimates from the telegraph model (α, β, θ). Perfect parameter inference would mean $\alpha = \rho, \beta = k_{\text{on}}, \theta = k_{\text{off}}$. We vary the mean expression by increasing ρ whilst keeping all other parameters constant. In agreement with our theory (summarised in Eq. (G3)), the accuracy of the heuristic inference method increases with the mean expression.

[1] J. Peccoud and B. Ycart, Markovian modeling of gene-product synthesis, Theoretical population biology **48**, 222 (1995).
[2] A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, and S. Tyagi, Stochastic mrna synthesis in mammalian cells, PLoS biology **4**, e309 (2006).
[3] D. Zenklusen, D. R. Larson, and R. H. Singer, Single-rna counting reveals alternative modes of gene expression in yeast, Nature structural & molecular biology **15**, 1263 (2008).
[4] S. O. Skinner, H. Xu, S. Nagarkar-Jaiswal, P. R. Freire, T. P. Zwaka, and I. Golding, Single-cell analysis of transcription kinetics across the cell cycle, Elife **5**, e12175 (2016).
[5] K. B. Halpern, S. Tanami, S. Landen, M. Chapal, L. Szlak, A. Hutzler, A. Nizhberg, and S. Itzkovitz, Bursty gene expression in the intact mammalian liver, Molecular cell **58**, 147 (2015).
[6] A. J. Larsson, P. Johnsson, M. Hagemann-Jensen, L. Hartmanis, O. R. Faridani, B. Reinius, Å. Segerstolpe, C. M. Rivera, B. Ren, and R. Sandberg, Genomic encoding of transcriptional burst kinetics, Nature **565**, 251 (2019).
[7] Z. Cao and R. Grima, Analytical distributions for detailed models of stochastic gene expression in eukaryotic cells, Proceedings of the National Academy of Sciences **117**, 4682 (2020).
[8] S. Luo, Z. Zhang, Z. Wang, X. Yang, X. Chen, T. Zhou, and J. Zhang, Inferring transcriptional bursting kinetics from single-cell snapshot data using a generalized telegraph model, Royal Society Open Science **10**, 221057 (2023).
[9] R. Perez-Carrasco, C. Beentjes, and R. Grima, Effects of cell cycle variability on lineage and population measurements of messenger rna abundance, Journal of the Royal Society Interface **17**, 20200360 (2020).
[10] C. Jia and R. Grima, Frequency domain analysis of fluctuations of mrna and protein copy numbers within a cell lineage: theory and experimental validation, Physical Review X **11**, 021032 (2021).
[11] E. Tunnacliffe and J. R. Chubb, What is a transcriptional burst?, Trends in Genetics **36**, 288 (2020).
[12] S. Luo, Z. Wang, Z. Zhang, T. Zhou, and J. Zhang, Genome-wide inference reveals that feedback regulations constrain promoter-dependent transcriptional burst kinetics, Nucleic Acids Research **51**, 68 (2023).
[13] X. Fu, H. P. Patel, S. Coppola, L. Xu, Z. Cao, T. L. Lenstra, and R. Grima, Quantifying how post-transcriptional noise and gene copy number variation bias transcriptional parameter inference from mrna distributions, Elife **11**, e82493 (2022).
[14] O. Padovan-Merhar, G. P. Nair, A. G. Biaesch, A. Mayer, S. Scarfone, S. W. Foley, A. R. Wu, L. S. Churchman, A. Singh, and A. Raj, Single mammalian cells compensate for differences in cellular volume and dna copy number through independent global transcriptional mechanisms, Molecular cell **58**, 339 (2015).
[15] X.-M. Sun, A. Bowman, M. Priestman, F. Bertaux, A. Martinez-Segura, W. Tang, C. Whilding, D. Dormann, V. Shahrezaei, and S. Marguerat, Size-dependent increase in rna polymerase ii initiation rates mediates gene expression scaling with cell size, Current Biology **30**, 1217 (2020).
[16] I. G. Johnston, B. Gaal, R. P. d. Neves, T. Enver, F. J. Iborra, and N. S. Jones, Mitochondrial variability as a source of extrinsic cellular noise, PLoS computational biology **8**, e1002416 (2012).
[17] J. Aryaman, I. G. Johnston, and N. S. Jones, Mitochondrial heterogeneity, Frontiers in genetics **9**, 718 (2019).
[18] X. Wu, R. Wollman, and A. L. MacLean, Single-cell ca2+ parameter inference reveals how transcriptional states inform dynamic cell responses, Journal of the Royal Society Interface **20**, 20230172 (2023).
[19] R. Foreman and R. Wollman, Mammalian gene expression variability is explained by underlying cell state, Molecular systems biology **16**, e9146 (2020).
[20] N. Battich, T. Stoeger, and L. Pelkmans, Control of transcript variability in single mammalian cells, Cell **163**, 1596 (2015).
[21] P. S. Swain, M. B. Elowitz, and E. D. Siggia, Intrinsic and extrinsic contributions to stochasticity in gene expression, Proceedings of the National Academy of Sciences **99**, 12795 (2002).
[22] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, Stochastic gene expression in a single cell, Science **297**, 1183 (2002).
[23] I. S. Araújo, J. M. Pietsch, E. M. Keizer, B. Greese, R. Balkunde, C. Fleck, and M. Hülskamp, Stochastic gene expression in arabidopsis thaliana, Nature communications **8**, 2132 (2017).
[24] A. Baudrimont, V. Jaquet, S. Wallerich, S. Voegeli, and A. Becskei, Contribution of rna degradation to intrinsic and extrinsic noise in gene expression, Cell reports **26**, 3752 (2019).
[25] T. Maier, M. Güell, and L. Serrano, Correlation of mrna and protein in complex biological samples, FEBS letters **583**, 3966 (2009).
[26] A. Llamosi, A. M. Gonzalez-Vargas, C. Versari, E. Cinquemani, G. Ferrari-Trecate, P. Hersen, and G. Batt, What population reveals about individual cell identity: single-cell parameter estimation of models of gene expression in yeast, PLoS computational biology **12**, e1004706 (2016).
[27] C. Zechner, M. Unger, S. Pelet, M. Peter, and H. Koeppl, Scalable inference of heterogeneous reaction kinetics from pooled single-cell recordings, Nature methods **11**, 197 (2014).
[28] T. L. Lenstra and D. R. Larson, Single-molecule mrna detection in live yeast, Current protocols in molecular biology **113**, 14 (2016).
[29] B. T. Donovan, A. Huynh, D. A. Ball, H. P. Patel, M. G. Poirier, D. R. Larson, M. L. Ferguson, and T. L. Lenstra, Live-cell imaging reveals the interplay between transcription factors, nucleosomes, and bursting, The EMBO journal **38**, e100809 (2019).
[30] C. Zopf, K. Quinn, J. Zeidman, and N. Maheshri, Cell-cycle dependence of transcription dominates noise in gene expression, PLoS computational biology **9**, e1003161 (2013).
[31] N. Battich, J. Beumer, B. de Barbanson, L. Krenning, C. S. Baron, M. E. Tanenbaum, H. Clevers, and A. van Oudenaarden,

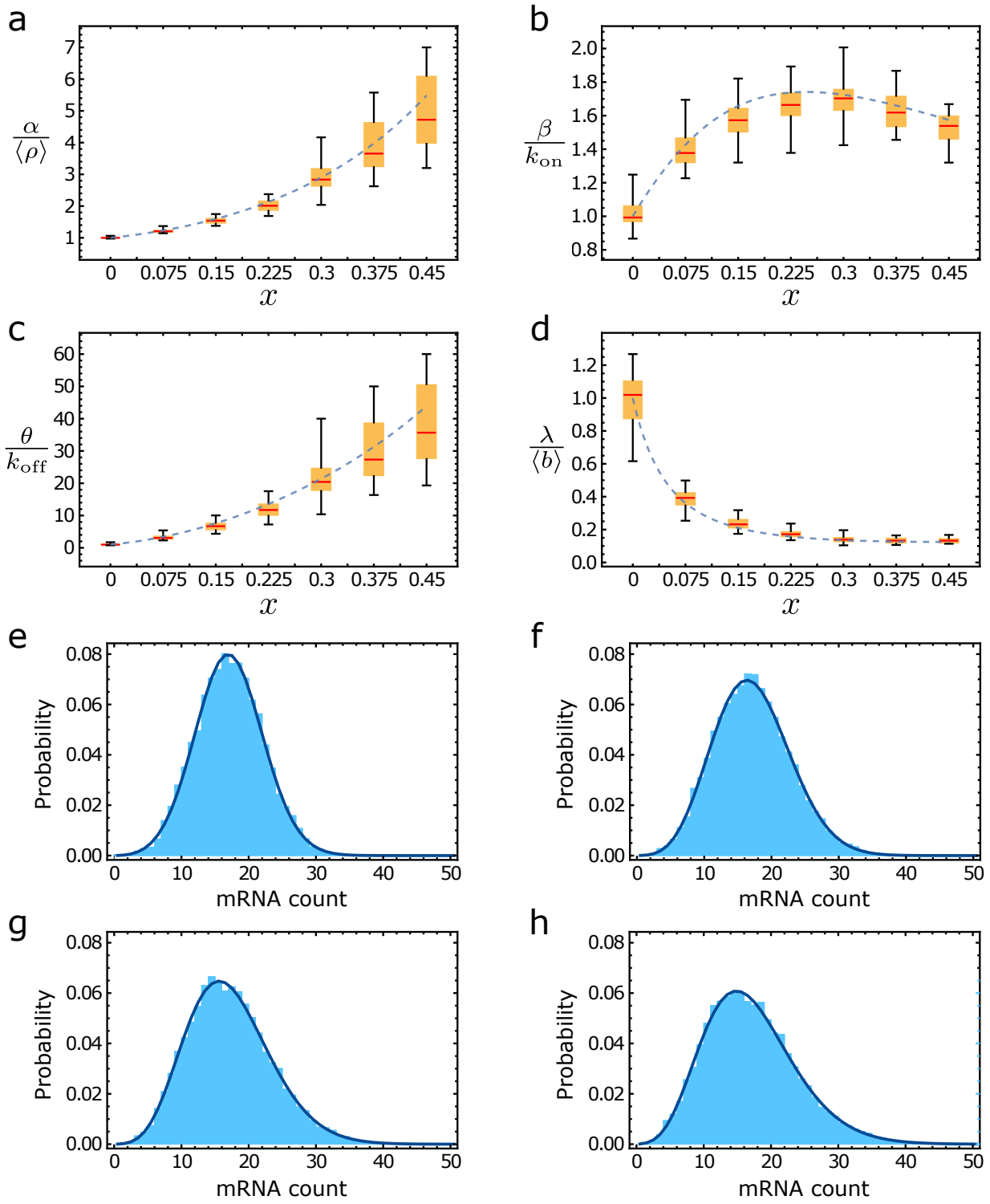

FIG. 7. Comparison of the transcriptional parameters estimated using moment-based theory and maximum likelihood fitting of the telegraph model to simulated mRNA count data with a cell-to-cell variation in the synthesis rate ρ (gamma distributed) and where the gene spends most of its time in the on state. (a)-(d) shows box-whisker plots similar to those in Fig. 4(a)-(d) with the difference that the parameters of the telegraph model with extrinsic noise are $\langle\rho\rangle = 20, k_{\rm on} = 5, k_{\rm off} = 1, \langle b\rangle = 4$ (gene spends 5/6 of its time in the active state). In (e)-(h) we show that the corresponding distributions of the telegraph model fitted by maximum likelihood (solid lines) to the simulated data (histogram) for the cases $x = 0, 0.15, 0.3, 0.45$ respectively – clearly, the telegraph model cannot be distinguished from the telegraph model with extrinsic noise from fitting of the population snapshot distributions of mRNA counts per cell.

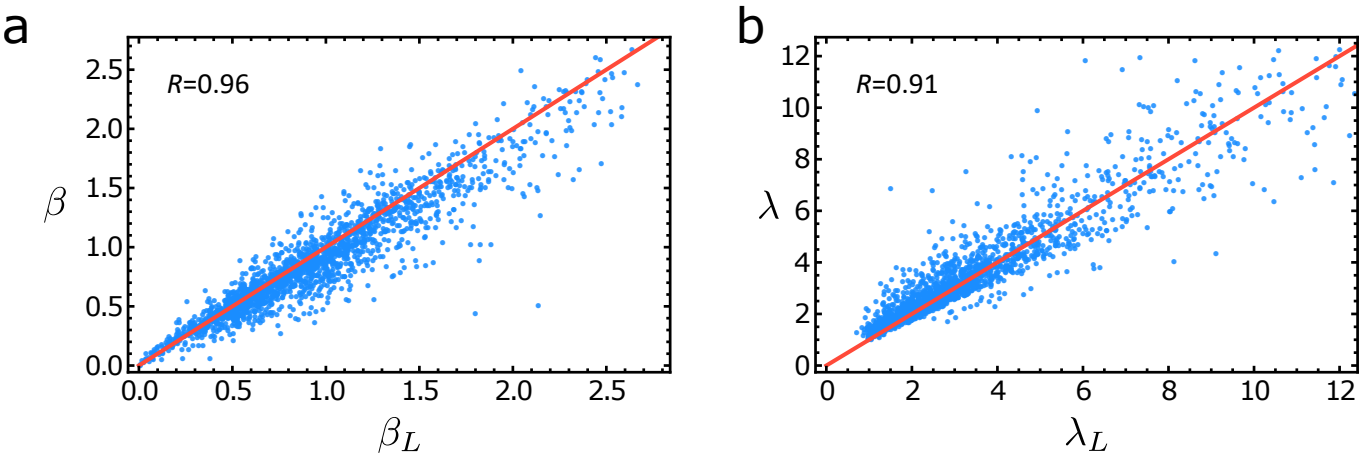

FIG. 8. Plots of the burst frequency (a) and size (b) for 1877 bursty genes estimated by fitting sc-RNA seq data from 224 primary adult mouse fibroblasts (C57 allele) to the standard telegraph model using the method of maximum likelihood (x-axis) and separately estimated using two moment-matching (y-axis). The maximum likelihood estimates are found in the file 41586_2018_836_$MOESM3_ESM.xlsx$ associated with Supplementary Table I in [6] and they are denoted by the subscript L. The moment-matching estimates are obtained by equating the first and second moments of the mRNA count data for each gene to the moments Eq. (7) predicted by the bursty expression model (5) and solving for λ and β. We consider only genes for which the mean expression > 1 and the maximum likelihood estimates of switching off rate is at least 20 times larger than the estimate of the switching off rate, thus enforcing transcriptional bursting. The maximum likelihood and moment-matching estimates are similar as indicated by the Pearson correlation coefficient R and the closeness of the points to the red lines $\beta = \beta_L$ and $\lambda = \lambda_L$.

Sequencing metabolically labeled transcripts in single cells reveals mrna turnover strategies, Science **367**, 1151 (2020).
[32] A. Riba, A. Oravecz, M. Durik, S. Jiménez, V. Alunni, M. Cerciat, M. Jung, C. Keime, W. M. Keyes, and N. Molina, Cell cycle gene regulation dynamics revealed by rna velocity and deep-learning, Nature Communications **13**, 2865 (2022).
[33] D. Lucía, B. Alan, G. Alicia, J. Rikard, J. David, A. Katz, G. Cedersund, and C.-L. Alejandro, Characterization of cell-to-cell variation in nuclear transport rates and identification of its sources, Iscience **26** (2023).
[34] L. Ham, R. D. Brackston, and M. P. Stumpf, Extrinsic noise and heavy-tailed laws in gene expression, Physical review letters **124**, 108101 (2020).
[35] L. Ham, M. Jackson, and M. P. Stumpf, Pathway dynamics can delineate the sources of transcriptional noise in gene expression, Elife **10**, e69324 (2021).
[36] M. S. Sherman, K. Lorenz, M. H. Lanier, and B. A. Cohen, Cell-to-cell variability in the propensity to transcribe explains correlated fluctuations in gene expression, Cell systems **1**, 315 (2015).
[37] J. Dattani and M. Barahona, Stochastic models of gene transcription with upstream drives: exact solution and sample path characterization, Journal of The Royal Society Interface **14**, 20160833 (2017).
[38] C. Zechner, J. Ruess, P. Krenn, S. Pelet, M. Peter, J. Lygeros, and H. Koeppl, Moment-based inference predicts bimodality in transient gene expression, Proceedings of the National Academy of Sciences **109**, 8340 (2012).
[39] B. Munsky, G. Li, Z. R. Fox, D. P. Shepherd, and G. Neuert, Distribution shapes govern the discovery of predictive models for gene regulation, Proceedings of the National Academy of Sciences **115**, 7533 (2018).
[40] J. K. Kim and J. C. Marioni, Inferring the kinetics of stochastic gene expression from single-cell rna-sequencing data, Genome biology **14**, 1 (2013).
[41] C. Furusawa, T. Suzuki, A. Kashiwagi, T. Yomo, and K. Kaneko, Ubiquity of log-normal distributions in intra-cellular reaction dynamics, Biophysics **1**, 25 (2005).
[42] Y. Taniguchi, P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie, Quantifying e. coli proteome and transcriptome with single-molecule sensitivity in single cells, science **329**, 533 (2010).
[43] S. Krishna, B. Banerjee, T. Ramakrishnan, and G. Shivashankar, Stochastic simulations of the origins and implications of long-tailed distributions in gene expression, Proceedings of the National Academy of Sciences **102**, 4771 (2005).
[44] J. Beal, Biochemical complexity drives log-normal variation in genetic expression, Engineering Biology **1**, 55 (2017).
[45] L. Cai, N. Friedman, and X. S. Xie, Stochastic protein expression in individual cells at the single molecule level, Nature **440**, 358 (2006).
[46] T. N. Vu, Q. F. Wills, K. R. Kalari, N. Niu, L. Wang, M. Rantalainen, and Y. Pawitan, Beta-poisson model for single-cell rna-seq data analyses, Bioinformatics **32**, 2128 (2016).
[47] B. Pijuan-Sala, J. A. Griffiths, C. Guibentif, T. W. Hiscock, W. Jawaid, F. J. Calero-Nieto, C. Mulas, X. Ibarra-Soria, R. C. Tyser, D. L. L. Ho, *et al.*, A single-cell molecular map of mouse gastrulation and early organogenesis, Nature **566**, 490 (2019).
[48] P. Brennecke, S. Anders, J. K. Kim, A. A. Kołodziejczyk, X. Zhang, V. Proserpio, B. Baying, V. Benes, S. A. Teichmann,

J. C. Marioni, *et al.*, Accounting for technical noise in single-cell rna-seq experiments, Nature methods **10**, 1093 (2013).
[49] Z. S. Singer, J. Yong, J. Tischler, J. A. Hackett, A. Altinok, M. A. Surani, L. Cai, and M. B. Elowitz, Dynamic heterogeneity and dna methylation in embryonic stem cells, Molecular cell **55**, 319 (2014).
[50] D. Nicolas, B. Zoller, D. M. Suter, and F. Naef, Modulation of transcriptional burst frequency by histone acetylation, Proceedings of the National Academy of Sciences **115**, 7153 (2018).
[51] C. A. Vallejos, J. C. Marioni, and S. Richardson, Basics: Bayesian analysis of single-cell sequencing data, PLoS computational biology **11**, e1004333 (2015).
[52] B. Vieth, C. Ziegenhain, S. Parekh, W. Enard, and I. Hellmann, powsimr: power analysis for bulk and single cell rna-seq experiments, Bioinformatics **33**, 3486 (2017).
[53] D. Grün, L. Kester, and A. Van Oudenaarden, Validation of noise models for single-cell transcriptomics, Nature methods **11**, 637 (2014).
[54] J. Paulsson, O. G. Berg, and M. Ehrenberg, Stochastic focusing: fluctuation-enhanced sensitivity of intracellular regulation, Proceedings of the National Academy of Sciences **97**, 7148 (2000).
[55] C. Jia, A. Singh, and R. Grima, Concentration fluctuations in growing and dividing cells: Insights into the emergence of concentration homeostasis, PLoS Computational Biology **18**, e1010574 (2022).
[56] H. Kempe, A. Schwabe, F. Crémazy, P. J. Verschure, and F. J. Bruggeman, The volumes and transcript counts of single cells reveal concentration homeostasis and capture biological noise, Molecular biology of the cell **26**, 797 (2015).
[57] S. Berry and L. Pelkmans, Mechanisms of cellular mrna transcript homeostasis, Trends in Cell Biology (2022).
[58] G. Gorin and L. Pachter, Modeling bursty transcription and splicing with the chemical master equation, Biophysical Journal **121**, 1056 (2022).
[59] W. Tang, F. Bertaux, P. Thomas, C. Stefanelli, M. Saint, S. Marguerat, and V. Shahrezaei, baynorm: Bayesian gene expression recovery, imputation and normalization for single-cell rna-sequencing data, Bioinformatics **36**, 1174 (2020).
[60] A. M. Klein, L. Mazutis, I. Akartuna, N. Tallapragada, A. Veres, V. Li, L. Peshkin, D. A. Weitz, and M. W. Kirschner, Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells, Cell **161**, 1187 (2015).
[61] W. Tang, A. C. S. Jørgensen, S. Marguerat, P. Thomas, and V. Shahrezaei, Modelling capture efficiency of single cell rna-sequencing data improves inference of transcriptome-wide burst kinetics, bioRxiv (2023).
[62] H. Ochiai, T. Sugawara, T. Sakuma, and T. Yamamoto, Stochastic promoter activation affects nanog expression variability in mouse embryonic stem cells, Scientific reports **4**, 7125 (2014).
[63] H. Sato and R. H. Singer, Cellular variability of nonsense-mediated mrna decay, Nature communications **12**, 7203 (2021).
[64] J. Holehouse, A. Gupta, and R. Grima, Steady-state fluctuations of a genetic feedback loop with fluctuating rate parameters using the unified colored noise approximation, Journal of Physics A: Mathematical and Theoretical **53**, 405601 (2020).
[65] C. Jia and R. Grima, Coupling gene expression dynamics to cell size dynamics and cell cycle events: Exact and approximate solutions of the extended telegraph model, Iscience **26**, 105746 (2023).
[66] G. Gorin, J. J. Vastola, M. Fang, and L. Pachter, Interpretable and tractable models of transcriptional noise for the rational design of single-molecule quantification experiments, Nature Communications **13**, 7620 (2022).
[67] G. Gorin and L. Pachter, Monod: mechanistic analysis of single-cell rna sequencing count data, bioRxiv (2022).
[68] G. Gorin, J. J. Vastola, and L. Pachter, Studying stochastic systems biology of the cell with single-cell genomics data, bioRxiv (2023).
[69] Q. F. Wills, K. J. Livak, A. J. Tipping, T. Enver, A. J. Goldson, D. W. Sexton, and C. Holmes, Single-cell gene expression analysis reveals genetic associations masked in whole-tissue experiments, Nature biotechnology **31**, 748 (2013).
[70] DLMF, *NIST Digital Library of Mathematical Functions*, `https://dlmf.nist.gov/13.4.E4`, Release 1.1.9 of 2023-03-15 (), f. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds.
[71] DLMF, *NIST Digital Library of Mathematical Functions*, `https://dlmf.nist.gov/13.6.E6`, Release 1.1.9 of 2023-03-15 (), f. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds.
[72] K. M. Wolter, Taylor series methods, in *Introduction to Variance Estimation* (Springer New York, New York, NY, 2007) pp. 226–271.