Bursty Gene Expression in Single Cells and Expanding Populations: A Discrete Approach

Jakub Poljovka[1], Iryna Zabaikina[1], Pavol Bokes[1], Abhyudai Singh[2]

[1]Faculty of Mathematics, Physics and Informatics, Comenius University, Slovakia [2]Department of Electrical and Computer Engineering, University of Delaware, USA

BACKGROUND

Gene expression is inherently stochastic, resulting in **cell-to-cell variability in protein levels**, even among genetically identical cells. Such **variability plays a crucial role** in such processes as cell differentiation, stress response, and antibiotic tolerance.

In this work, we model a single cell in which:

- Proteins are synthesized in instantaneous, discrete events – **bursts** – each producing a random number of proteins.
- **Protein levels are decreased by dilution** due to active cell growth.
- **Positive feedback on dilution:** an increase in protein levels imposes a burden on the cell, reducing its growth rate and thereby slowing down dilution.

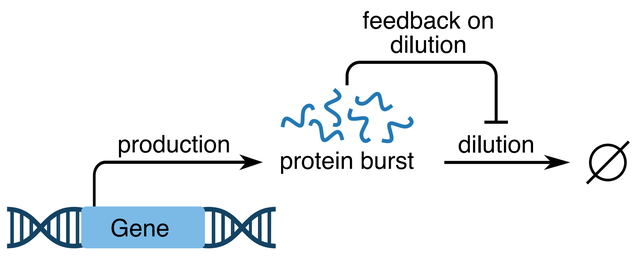

Figure 1. Schematic of the studied model.

MODEL

We model protein dynamics in a single cell as a *continuous-time discrete-value Markov process* $n(t)$ – ***protein count per unit volume*** at time t.

$$R_1: \quad n \xrightarrow{\lambda} n+b,\ b \sim \text{Geom}(\beta) \quad \text{(protein burst)}$$

$$R_2: \quad n \xrightarrow{f(n)} n-1 \quad \text{(protein dilution)}$$

$$R_3: \text{cell} \xrightarrow{r(n)} 2 \text{ cells} \quad \text{(cell division)}$$

- **Bursts occur with frequency** λ, following a *Poisson process*. **The burst size** b follows a *geometric distribution*, with the mean β.
- The **cell growth rate** depends on protein level:
$$r(n) = \frac{\gamma}{1+kn}$$
where k is the *feedback strength*, γ is the maximal growth rate. **Higher protein levels lower the growth rate, modelling the burden effect.**
- **Dilution** is modelled as loss event, occurring at a rate $f(n) = n \cdot r(n)$.
- In the **population model**, **each cell divides independently at rate** $r(n)$.

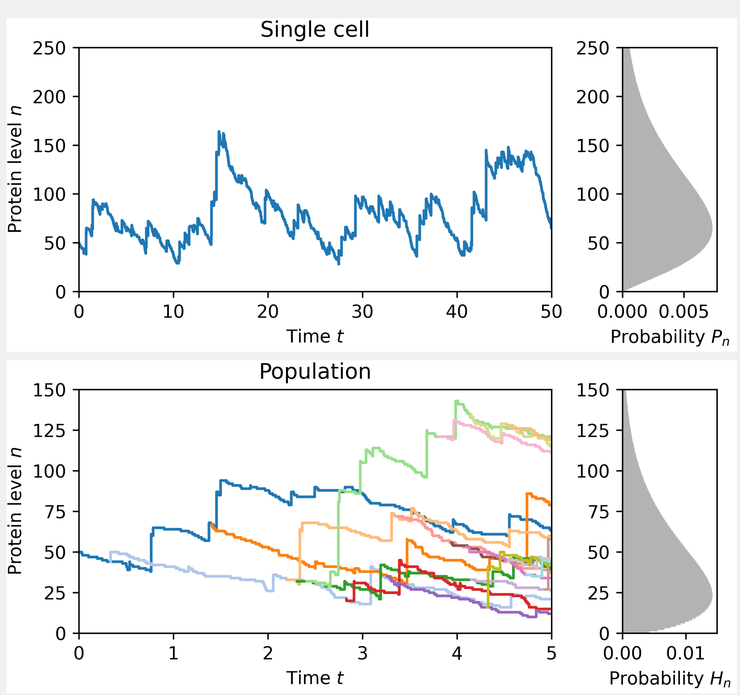

Figure 2. Sample trajectories of single-cell and population processes.

SINGLE-CELL PERSPECTIVE

The probability $P_n(t)$ that the protein level is n at time t satisfies the following master equation:

$$\frac{dP_n(t)}{dt} = \lambda \sum_{i=0}^{n} P_{n-i}(t) b_i + \Delta_n(P_n(t) f(n)) - \lambda P_n(t), \tag{1}$$

where $\Delta_n(\cdot)$ is the forward difference operator.

Let P_n denote the probability that the protein level is n in the stationary state. Then P_n can be expressed as a weighted average of two negative binomial mass functions:

$$P_n = \delta f_{NB}(n-1, \rho+1, q\lambda/\gamma\rho) + (1-\delta) f_{NB}(n, \rho, q\lambda/\gamma\rho),$$

where $f_{NB}(n, r, \rho)$ is *the mass function of the negative binomial distribution* and $\delta = \lambda q k/(\gamma(1-q))$.

POPULATION-LEVEL PERSPECTIVE

The number of cells in population $H_n(t)$ with the protein level n at time t is governed by the following population balance equation:

$$\frac{dH_n}{dt} = \lambda \sum_{i=0}^{n} b_i H_{n-i}(t) + \Delta_n(H_n(t) f(n)) + r(n) H_n(t) - \lambda H_n(t). \tag{2}$$

Similarly to the single-cell model, the stationary distribution H_n can be expressed as the sum of negative binomial mass functions:

$$H_n = (1-\xi) f_{NB}(n, \eta, \theta) + \xi f_{NB}(n-1, \eta+1, \theta), \tag{3}$$

where $f_{NB}(n, r, p)$ is *the mass function of the negative binomial distribution* and $\xi = \lambda q k/(\gamma(1+k-q))$.

The details on derivation of P_n and H_n, as well as parameters ρ, η, θ are given in [1].

SINGLE CELL $\neq$ POPULATION

As the feedback strength k increases:

- **Single-cell level**: dilution becomes less effective and a single cell accumulates more protein.
- **Population level**: fast-dividing cells with low protein proliferate faster, so the mean protein level decreases.

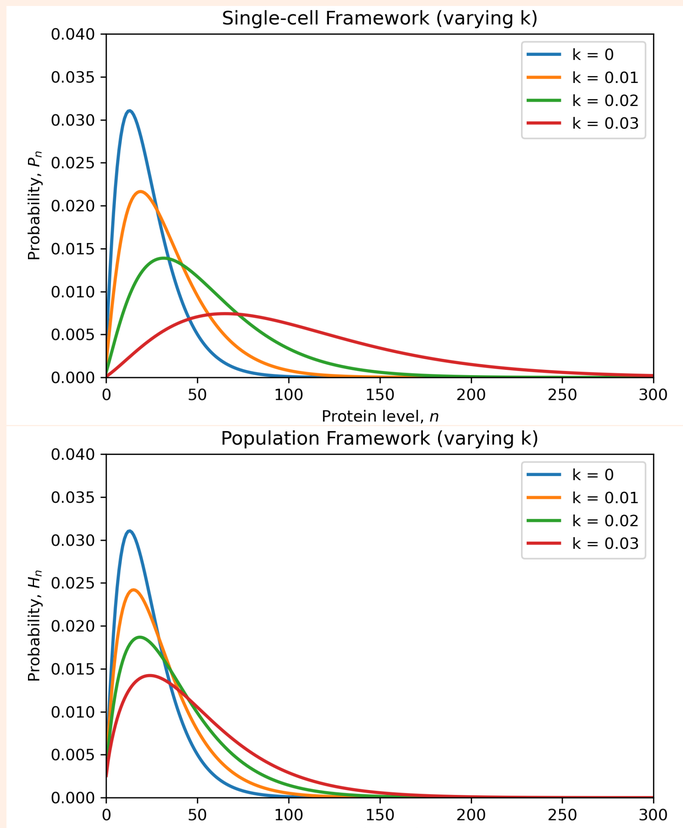

Figure 3. The effect of increasing the positive feedback strength.

TIME-DEPENDENT SOLUTIONS

In general, the master equations (1) and (2) can also be represented in matrix form as:

$$\frac{dP(t)}{dt} = A_s P(t), \quad \frac{dH(t)}{dt} = A_p H(t), \tag{4}$$

where $P(t)$ is probability vector describing the protein level in the single-cell model and $H(t)$ is vector describing the number of cells with given protein level in the population model. The matrices A_s and A_p correspond to the single-cell and population models, respectively, and have the following general form:

$$A_{ij} = \begin{cases} \lambda b_{i-j}, & \text{if } j < i, \\ -\lambda q - f(i) + \chi r(i), & \text{if } j = i, \\ f(i+1), & \text{if } j = i+1, \\ 0, & \text{if } j > i+1, \end{cases} \tag{5}$$

where χ is the indicator parameter: $\chi = 0$ in A_s and $\chi = 1$ in A_p. The solutions of (4) can be expressed as:

$$P(t) = e^{tA_s} P(0), \quad H(t) = e^{tA_p} H(0). \tag{6}$$

The **main challenge in the numerical solution** is that The vectors $P(t)$ and $H(t)$, as well as the matrices A_s and A_p, **are infinite-dimensional**. Thus, to obtain the time-dependent solutions in Fig. 4, we truncate the dimension at a finite level N.

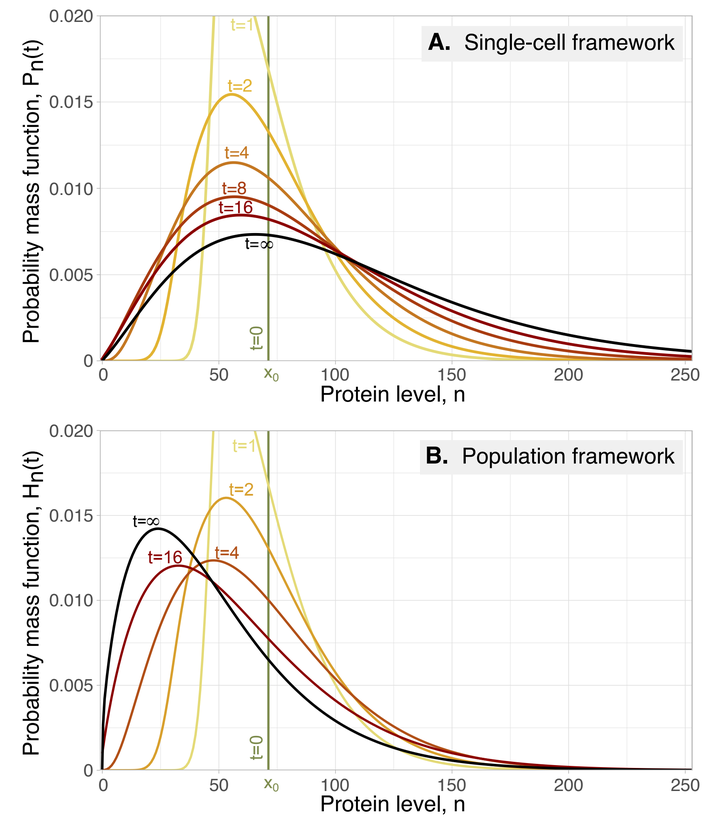

Figure 4. Time evolution of protein distributions at single-cell and population-level perspectives.

RESULTS

- We obtained explicit solutions for the steady-state probability mass function.
- The results are consistent with the continuous model.
- Infinite dimensions are truncated once the tail mass is negligible ($N = 600$).
- Time-dependent numerical solutions **rapidly converge to the explicit steady states**.

References

[1] Jakub Poljovka, Iryna Zabaikina, Pavol Bokes, and Abhyudai Singh. Bursty gene expression in single cells and expanding populations: A discrete approach. *bioRxiv*, 2025.