Supporting Information for "A first–passage time approach to controlling noise in the timing of intracellular events"

Khem Raj Ghusinga, John J. Dennehy, Abhyudai Singh 10.1073/pnas.1609012114

## S1. Computation of first-passage time distribution

Eq. (2) of the main text describes the time evolution of protein count $x(t)$ by the following bursty birth-death process

$$\mathbb{P}\left(x(t+dt)=i+B|x(t)=i\right)=k_i\,dt, \tag{S1.1a}$$

$$\mathbb{P}\left(x(t+dt)=i-1|x(t)=i\right)=i\gamma dt, \tag{S1.1b}$$

where B denotes the protein burst size which is assumed to follow a geometric distribution with mean b. We are interested in computing the first-passage (FPT) described by the following random variable

$$T := \min\{t : x(t) \geq X | x(0) = 0\}. \tag{S1.2}$$
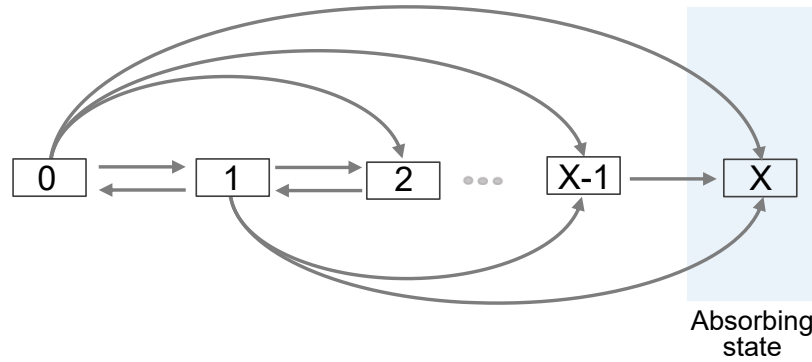
Fig. S1. Illustration of a bursty birth-death process for computing the first-passage time. States $[0, 1, \ldots, X]$ represent the protein population counts, and arrows represent transition between states due to burst and decay events. The destination of a forward jump (a birth event) is decided by the burst size while each degradation event reduces the protein count by one. The process terminates when the protein level reaches the absorbing-state X and the first-passage time is recorded.

One can note that if the protein does not decay, then $x(t)$ accumulates over time and the FPT distribution could be obtained by observing

$$\mathbb{P}\left(x(t) \geq X\right) = \mathbb{P}\left(T \leq t\right). \tag{S1.3}$$

However, with protein degradation, the FPT calculation needs careful consideration so as to avoid counting multiple crossings of the threshold. To this end, we construct an equivalent bursty birth-death process on a finite state-space $[0, 1, \ldots, X]$, where the states represent the protein count (Fig. S1). All states denoting $x(t) \geq X$ are combined into a single absorbing state X. In this model, the probability of the protein level reaching X in the small time window $(t, t+dt)$ is the probability that $x(t) = i$ and a jump of size $X - i$ or larger occurs in $(t, t+dt)$. That is, we have

$$f_T(t)dt = \mathbb{P}\left(T \in (t, t+dt)\right) \tag{S1.4a}$$

$$= \sum_{i=0}^{X-1} k_i \mathbb{P}\left(B \geq X - i\right)\ \mathbb{P}\left(x(t) = i\right)dt, \tag{S1.4b}$$

where $f_T(t)$ is the first-passage time probability density function. Note that Eq. (S1.4b) is valid for any distribution of the burst size B. When the burst size B is geometrically distributed, we can use the following relation in Eq. (S1.4b)

$$\mathbb{P}\left(B \geq X - i\right) = \left(\frac{b}{b+1}\right)^{X-i}. \tag{S1.5}$$

Defining $p_i(t) := \mathbb{P}\left(x(t) = i\right)$, the FPT pdf can be written as product of two vectors

$$f_T(t) = \mathbf{U}^\top \mathbf{P}(t), \qquad \text{[S1.6a]}$$

$$\mathbf{U} = \begin{bmatrix} k_0 \left(\frac{b}{b+1}\right)^X & k_1 \left(\frac{b}{b+1}\right)^{X-1} & \cdots & k_{X-1}\frac{b}{b+1} \end{bmatrix}^\top, \qquad \text{[S1.6b]}$$

$$\mathbf{P}(t) = \begin{bmatrix} p_0(t) & p_1(t) & \cdots & p_{X-1}(t) \end{bmatrix}^\top. \qquad \text{[S1.6c]}$$

To obtain $\mathbf{P}(t)$, we write the Chemical Master Equation (CME) corresponding to the equivalent bursty-birth death process in Fig. S1

$$\dot{p}_0(t) = -k_0 \frac{b}{b+1} p_0(t) + \gamma p_1(t), \qquad \text{[S1.7a]}$$

$$\dot{p}_i(t) = -\left(k_i \frac{b}{b+1} + i\gamma\right) p_i(t) + (i+1)\gamma p_{i+1}(t) + \sum_{n=0}^{i-1} k_n \frac{b^{i-n}}{(b+1)^{i-n+1}} p_n(t), \quad 1 \le i \le X-2, \qquad \text{[S1.7b]}$$

$$\dot{p}_{X-1}(t) = -\left(k_{X-1}\frac{b}{b+1} + (X-1)\gamma\right) p_{X-1}(t) + \sum_{n=0}^{X-2} k_n \frac{b^{X-n}}{(b+1)^{X-n+1}} p_n(t). \qquad \text{[S1.7c]}$$

The CME can be conveniently written as the following linear dynamical system

$$\dot{\mathbf{P}} = \mathbf{A}\mathbf{P}, \qquad \text{[S1.8a]}$$

where an ith row and j^{th} column element a_{ij} of the matrix $\mathbf{A}$ is given by

$$a_{ij} = \begin{cases} 0, & j > i+1 \\ (i-1)\gamma, & j = i+1 \\ -k_{i-1}\frac{b}{b+1} - (i-1)\gamma, & j = i \\ k_{i-1}\frac{b^{i-j}}{(b+1)^{i-j+1}}, & j < i \end{cases} \qquad \text{[S1.8b]}$$

$i, j \in \{1, \ldots, X\}$. Solving Eq. (S1.8a) and using Eq. (S1.6a) yields the following pdf for the first-passage time

$$f_T(t) = \mathbf{U}^\top \mathbf{P}(t) = \mathbf{U}^\top \exp(\mathbf{A}t)\mathbf{P}(0), \qquad \text{[S1.9]}$$

where $\mathbf{P}(0) = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^\top$ is vector of probabilities at $t = 0$ that follows from $x(0) = 0$.

S2. On some properties of the matrix A

In this section, we discuss some properties of the matrix $\mathbf{A}$ given in Eq. (S1.8b). One can note that $\mathbf{A}$ is an "almost" lower triangular matrix: all entries above its first superdiagonal are zero. This kind of matrices are known as lower Hessenberg matrices. Additionally, $\mathbf{A}$ satisfies two important properties that it is a Hurwitz matrix, and its inverse can be explicitly obtained. We'll see in the next section how these properties allow one to compute expressions of FPT moments. Below we give discuss these two properties in detail.

A **is a Hurwitz matrix.** The matrix $\mathbf{A}$ is given by

$$\mathbf{A} = \begin{bmatrix} -\frac{bk_0}{b+1} & \gamma & \cdots & 0 \\ \frac{bk_0}{(b+1)^2} & -\left(\frac{bk_1}{b+1} + \gamma\right) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \frac{b^{X-2}k_0}{(b+1)^{X-1}} & \frac{b^{X-3}k_1}{(b+1)^{X-2}} & \cdots & (X-1)\gamma \\ \frac{b^{X-1}k_0}{(b+1)^X} & \frac{b^{X-2}k_1}{(b+1)^{X-1}} & \cdots & -\left(\frac{bk_{X-1}}{b+1} + (X-1)\gamma\right) \end{bmatrix}. \qquad \text{[S2.1]}$$

In order to prove that $\mathbf{A}$ is a Hurwitz matrix, we show that the following two conditions hold true [1, pp. 48–49]:

1. The diagonal elements $a_{ii} < 0$ for $i = 1, 2, \cdots, X$,

2. $\displaystyle \max_{1 \le j \le X} \sum_{\substack{i=1 \\ j \ne i}}^{X} \left|\frac{a_{ij}}{a_{jj}}\right| < 1.$

The first requirement above is clearly satisfied as $a_{ii} = -\frac{bk_{i-1}}{b+1} - (i-1)\gamma < 0$. To check whether the second requirement is satisfied, for any column $j = 1, 2, \cdots, X$ we have that

$$\sum_{\substack{i=1 \\ j \neq i}}^{X} \left| \frac{a_{ij}}{a_{jj}} \right| = \frac{(j-1)\gamma}{\frac{bk_{j-1}}{b+1} + (j-1)\gamma} + \frac{bk_{j-1}}{b+1} \sum_{i=j+1}^{X} \frac{\left(\frac{b}{b+1}\right)^{i-j}}{\frac{bk_{j-1}}{b+1} + (j-1)\gamma} \qquad [\text{S2.2a}]$$

$$= \frac{(j-1)\gamma}{\frac{bk_{j-1}}{b+1} + (j-1)\gamma} + \frac{\frac{bk_{j-1}}{b+1}\left(1 - \left(\frac{b}{b+1}\right)^{X-j}\right)}{\frac{bk_{j-1}}{b+1} + (j-1)\gamma} \qquad [\text{S2.2b}]$$

$$= \frac{\frac{bk_{j-1}}{b+1}\left(1 - \left(\frac{b}{b+1}\right)^{X-j}\right) + (j-1)\gamma}{\frac{bk_{j-1}}{b+1} + (j-1)\gamma} < 1. \qquad [\text{S2.2c}]$$

Thus, the matrix $\mathbf{A}$ is Hurwitz, i.e., the eigenvalues of $\mathbf{A}$ have negative real part.

Determining inverse of the matrix A. Let us use $\mathbf{A}_0$ to denote the matrix $\mathbf{A}$ when $\gamma = 0$. The lower triangular matrix $\mathbf{A}_0$ is given by

$$\mathbf{A} = \begin{bmatrix} -\frac{bk_0}{b+1} & 0 & \cdots & 0 \\ \frac{bk_0}{(b+1)^2} & -\frac{bk_1}{b+1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{b^{X-2}k_0}{(b+1)^{X-1}} & \frac{b^{X-3}k_1}{(b+1)^{X-2}} & \cdots & 0 \\ \frac{b^{X-1}k_0}{(b+1)^X} & \frac{b^{X-2}k_1}{(b+1)^{X-1}} & \cdots & -\frac{bk_{X-1}}{b+1} \end{bmatrix}. \qquad [\text{S2.3}]$$

We claim that the inverse of $\mathbf{A}_0$ is given by the following matrix

$$\mathbf{A}_0^{-1} = -\frac{1}{b} \begin{bmatrix} \frac{b+1}{k_0} & 0 & \cdots & 0 & 0 \\ \frac{1}{k_1} & \frac{b+1}{k_1} & \cdots & 0 & 0 \\ \frac{1}{k_2} & \frac{1}{k_2} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{1}{k_{X-2}} & \frac{1}{k_{X-2}} & \cdots & \frac{b+1}{k_{X-2}} & 0 \\ \frac{1}{k_{X-1}} & \frac{1}{k_{X-1}} & \cdots & \frac{1}{k_{X-1}} & \frac{b+1}{k_{X-1}} \end{bmatrix}. \qquad [\text{S2.4}]$$

This claim can be quickly verified by multiplying the matrices which results in identity matrix. Next, to determine $\mathbf{A}^{-1}$, we observe that when $\gamma \neq 0$, the matrix $\mathbf{A}$ can be written as

$$\mathbf{A} = \mathbf{A}_0 + \mathbf{A}_e, \qquad [\text{S2.5}]$$

where $\mathbf{A}_e$ is given by

$$\mathbf{A}_e = \begin{bmatrix} 0 & \gamma & \cdots & 0 & 0 \\ 0 & -\gamma & \cdots & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & -(X-2)\gamma & (X-1)\gamma \\ 0 & 0 & \cdots & 0 & -(X-1)\gamma \end{bmatrix}. \qquad [\text{S2.6}]$$

Therefore the inverse of the matrix $\mathbf{A}$ can be written as

$$\mathbf{A}^{-1} = (\mathbf{A}_0 + \mathbf{A}_e)^{-1} = \left(\mathbf{I} + \mathbf{A}_0^{-1}\mathbf{A}_e\right)^{-1} \mathbf{A}_0^{-1}. \qquad [\text{S2.7}]$$

As we have already determined the expression of $\mathbf{A}_0^{-1}$ in Eq. (S2.4), the expression of $\mathbf{A}_0^{-1}\mathbf{A}_e$ can be determined as

$$-\frac{1}{b}\begin{bmatrix} \frac{b+1}{k_0} & 0 & \cdots & 0 & 0 \\ \frac{1}{k_1} & \frac{b+1}{k_1} & \cdots & 0 & 0 \\ \frac{1}{k_2} & \frac{1}{k_2} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{1}{k_{X-2}} & \frac{1}{k_{X-2}} & \cdots & \frac{b+1}{k_{X-2}} & 0 \\ \frac{1}{k_{X-1}} & \frac{1}{k_{X-1}} & \cdots & \frac{1}{k_{X-1}} & \frac{b+1}{k_{X-1}} \end{bmatrix}\begin{bmatrix} 0 & \gamma & \cdots & 0 & 0 \\ 0 & -\gamma & \cdots & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & -(X-2)\gamma & (X-1)\gamma \\ 0 & 0 & \cdots & 0 & -(X-1)\gamma \end{bmatrix} \quad \text{[S2.8a]}$$

$$=\begin{bmatrix} 0 & -\frac{(b+1)\gamma}{bk_0} & \cdots & 0 & 0 \\ 0 & \frac{\gamma}{k_1} & \cdots & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \frac{(X-2)\gamma}{k_{X-2}} & -\frac{(X-1)(b+1)\gamma}{bk_{X-2}} \\ 0 & 0 & \cdots & 0 & \frac{(X-1)\gamma}{k_{X-1}} \end{bmatrix}. \quad \text{[S2.8b]}$$

Thus, the matrix $\mathbf{I}+\mathbf{A}_0^{-1}\mathbf{A}_e$ is a bidiagonal matrix with its diagonal elements $1+\frac{(i-1)\gamma}{k_{i-1}}=\frac{k_{i-1}+(i-1)\gamma}{k_{i-1}}$ for $i=1,2,\cdots,X$. The super diagonal elements are given by $-\frac{j(b+1)\gamma}{bk_{j-1}}$ for $j=1,2,\cdots,X-1$. Using the result for inverse of a bidiagonal matrix [2], we can write an element at the ith row and the jth column of inverse of $\mathbf{E}:=\mathbf{I}+\mathbf{A}_0^{-1}\mathbf{A}_e$ as follows

$$e_{i,j}'=\begin{cases} 0 & \text{if } i>j, \\ \frac{k_{i-1}}{k_{i-1}+(i-1)\gamma}, & \text{if } i=j, \\ \frac{k_{i-1}}{k_{i-1}+(i-1)\gamma}\prod_{l=i}^{j-1}\left(\frac{k_l}{k_l+l\gamma}\frac{l(b+1)\gamma}{bk_{l-1}}\right), & \text{if } i<j. \end{cases} \quad \text{[S2.9]}$$

Alternatively, in matrix form

$$\mathbf{E}^{-1}=\begin{bmatrix} 1 & \frac{k_1}{k_1+\gamma}\frac{(b+1)\gamma}{bk_0} & \cdots & \prod_{l=1}^{X-1}\left(\frac{k_l}{k_l+l\gamma}\frac{l(b+1)\gamma}{bk_{l-1}}\right) \\ 0 & \frac{k_1}{k_1+\gamma} & \cdots & \frac{k_1}{k_1+\gamma}\prod_{l=2}^{X-1}\left(\frac{k_l}{k_l+l\gamma}\frac{l(b+1)\gamma}{bk_{l-1}}\right) \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \frac{k_{X-1}}{k_{X-1}+(X-1)\gamma} \end{bmatrix}. \quad \text{[S2.10]}$$

We can compute $\mathbf{A}^{-1}$ by calculating $\mathbf{E}^{-1}\mathbf{A}_0^{-1}$. Here, we do not give explicit form of $\mathbf{A}^{-1}$ as it is not required for calculations in this document.

S3. Expression of mth moment of first-passage time

In this section, we make use of the properties discussed in the previous section to determine the moments of the first-passage time. As discussed earlier, the distribution of first-passage time (FPT) is given by Eq. (S1.9). Using this, a moment of mth order can be calculated as

$$\langle T^m\rangle=\int_0^\infty t^m\mathbf{U}^\top\exp(\mathbf{A}t)\mathbf{P}(0)dt \quad \text{[S3.1a]}$$

$$=\mathbf{U}^\top\left(\int_0^\infty t^m\exp(\mathbf{A}t)dt\right)\mathbf{P}(0). \quad \text{[S3.1b]}$$

Let us consider $\int_0^\infty t^m\exp(\mathbf{A}t)dt$. Integrating by parts

$$\int_0^\infty t^m\exp(\mathbf{A}t)dt=t^m\mathbf{A}^{-1}\exp(\mathbf{A}t)\bigg|_0^\infty-m\mathbf{A}^{-1}\int_0^\infty t^{m-1}\exp(\mathbf{A}t)dt. \quad \text{[S3.2]}$$

As the matrix $\mathbf{A}$ is a Hurwitz matrix, i.e., eigenvalues of $\mathbf{A}$ have negative real parts, the first term in the above expression goes to zero. Using a_m as a notion to represent $\int_0^\infty t^m\exp(\mathbf{A}t)dt$, we can write the following recursive relationship

$$a_m=-m\mathbf{A}^{-1}a_{m-1}. \quad \text{[S3.3]}$$

Thus, $a_m = (-1)^m m! (\mathbf{A}^{-1})^m a_0$. Further a_0 can be computed as

$$a_0 = \int_0^\infty \exp(\mathbf{A}t) dt = \mathbf{A}^{-1} \exp(\mathbf{A}t) \Big|_0^\infty = -\mathbf{A}^{-1}. \qquad \text{[S3.4]}$$

Therefore $a_m = (-1)^{m+1} m! (\mathbf{A}^{-1})^{m+1}$. Substituting this in Eq. (S3.1b) gives the following for a general m^{th} moment of FPT

$$\langle T^m \rangle = (-1)^m m! \mathbf{U}^\top (\mathbf{A}^{-1})^{m+1} \mathbf{P}(0). \qquad \text{[S3.5]}$$

Calculation of $U^\top A^{-1}$. As we saw in equation Eq. (S3.5), calculation of the moments have a term of the form $\mathbf{U}^\top \mathbf{A}^{-1}$. Here we provide the calculation of this term.

Consider two matrices $\mathbf{G}$ and $\mathbf{H}$ such that $\mathbf{A}_e = \mathbf{GH}$ where $\mathbf{G}$ is a $X \times X - 1$ matrix

$$\mathbf{G} = \gamma \begin{bmatrix} -1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & -2 & 0 & \cdots & 0 & 0 \\ 0 & 2 & -3 & \cdots & 0 & 0 \\ 0 & 0 & 3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & X-2 & -(X-1) \\ 0 & 0 & 0 & \cdots & 0 & X-1 \end{bmatrix}, \qquad \text{[S3.6a]}$$

while $\mathbf{H}$ is a $X - 1 \times X$ matrix

$$\mathbf{H} = \begin{bmatrix} 0 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & -1 \end{bmatrix}. \qquad \text{[S3.6b]}$$

Using the matrix inversion lemma, $\mathbf{A}^{-1}$ can be written as

$$\mathbf{A}^{-1} = (\mathbf{A}_0 + \mathbf{GH})^{-1}, \qquad \text{[S3.7a]}$$
$$= \mathbf{A}_0^{-1} - \mathbf{A}_0^{-1} \mathbf{G} \left(\mathbf{I} + \mathbf{H} \mathbf{A}_0^{-1} \mathbf{G} \right)^{-1} \mathbf{H} \mathbf{A}_0^{-1}. \qquad \text{[S3.7b]}$$

The expression $\mathbf{U}^\top \mathbf{A}_0^{-1} \mathbf{G}$ can be computed as below

$$\mathbf{U}^\top \mathbf{A}_0^{-1} \mathbf{G} = -\begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} -1 & 0 & \cdots & 0 \\ 1 & -2 & \cdots & 0 \\ 0 & 2 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & -(X-1) \\ 0 & 0 & \cdots & X-1 \end{bmatrix}, \qquad \text{[S3.8a]}$$
$$= -\begin{bmatrix} 0 & 0 & \cdots & 0 \end{bmatrix}. \qquad \text{[S3.8b]}$$

Therefore, we can conclude that $\mathbf{U}^\top \mathbf{A}^{-1}$ is in fact equal to $\mathbf{U}^\top \mathbf{A}_0^{-1}$ which could be calculated by multiplying $\mathbf{U}^\top$ and $\mathbf{A}_0^{-1}$.

$$\mathbf{U}^\top \mathbf{A}_0^{-1} = -\frac{1}{b}
\begin{bmatrix}
k_0 \left(\frac{b}{b+1}\right)^X \\
k_1 \left(\frac{b}{b+1}\right)^{X-1} \\
k_2 \left(\frac{b}{b+1}\right)^{X-2} \\
\vdots \\
k_{X-2} \left(\frac{b}{b+1}\right)^2 \\
k_{X-1} \left(\frac{b}{b+1}\right)
\end{bmatrix}^T
\begin{bmatrix}
\frac{b+1}{k_0} & 0 & \cdots & 0 & 0 \\
\frac{1}{k_1} & \frac{b+1}{k_1} & \cdots & 0 & 0 \\
\frac{1}{k_2} & \frac{1}{k_2} & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
\frac{1}{k_{X-2}} & \frac{1}{k_{X-2}} & \cdots & \frac{b+1}{k_{X-2}} & 0 \\
\frac{1}{k_{X-1}} & \frac{1}{k_{X-1}} & \cdots & \frac{1}{k_{X-1}} & \frac{b+1}{k_{X-1}}
\end{bmatrix}
\qquad \text{[S3.9a]}$$

$$= -\frac{1}{b}
\begin{bmatrix}
\frac{b^X}{(b+1)^{X-1}} + \sum_{l=1}^{X-1} \left(\frac{b}{b+1}\right)^l \\
\frac{b^{X-1}}{(b+1)^{X-2}} + \sum_{l=1}^{X-2} \left(\frac{b}{b+1}\right)^l \\
\frac{b^{X-2}}{(b+1)^{X-3}} + \sum_{l=1}^{X-3} \left(\frac{b}{b+1}\right)^l \\
\vdots \\
\frac{b^2}{b+1} + \frac{b}{b+1} \\
b
\end{bmatrix}^T
\qquad \text{[S3.9b]}$$

$$= -\frac{1}{b}
\begin{bmatrix} b & b & b & \cdots & b & b \end{bmatrix}
\qquad \text{[S3.9c]}$$

$$= -\begin{bmatrix} 1 & 1 & 1 & \cdots & 1 & 1 \end{bmatrix}.
\qquad \text{[S3.9d]}$$

Thus $\mathbf{U}^\top \mathbf{A}^{-1} = -\begin{bmatrix} 1 & 1 & 1 & \cdots & 1 & 1 \end{bmatrix}$.

Calculation of the first-two FPT moments. Though the expression in Eq. (S3.5) can be used to compute the first-passage time moments, often one is more interested in the mean FPT and the variance thereof (note that variance can be computed from the second order moment of FPT and the mean FPT). Here we show that the first two moments of FPT can be expressed as the following series summations.

$$\langle T \rangle = \frac{1}{k_0} + \frac{1}{b} \sum_{i=1}^{X} \frac{k_{i-1}}{k_{i-1} + (i-1)\gamma} \left(\frac{1}{k_{i-1}} + \sum_{j=i+1}^{X} \frac{1}{k_{j-1}} \prod_{l=i}^{j-1} \frac{k_l}{k_l + l\gamma} \frac{(b+1)l\gamma}{bk_{l-1}} \right),
\qquad \text{[S3.10a]}$$

$$\langle T^2 \rangle = \frac{2}{b^2} \sum_{i=1}^{X} \frac{k_{i-1}}{k_{i-1} + (i-1)\gamma} \left(\frac{b\,\eta_i + \sum_{j=1}^{i} \eta_j}{k_{i-1}} + \sum_{r=i+1}^{X} \frac{b\,\eta_i + \sum_{j=1}^{r} \eta_j}{k_{r-1}} \prod_{l=i}^{r-1} \frac{k_l}{k_l + l\gamma} \frac{(b+1)l\gamma}{bk_{l-1}} \right),
\qquad \text{[S3.10b]}$$

where the terms η_i are given by

$$\eta_i = \frac{b}{k_0} \delta_{i-1} + \frac{k_{i-1}}{k_{i-1} + (i-1)\gamma} \left(\frac{1}{k_{i-1}} + \sum_{j=i+1}^{X} \frac{1}{k_{j-1}} \prod_{l=i}^{j-1} \frac{k_l}{k_l + l\gamma} \frac{(b+1)l\gamma}{bk_{l-1}} \right),
\qquad \text{[S3.10c]}$$

with δ_{i-1} denoting the Kronecker delta which is one if $i = 1$ and zero otherwise. Below we show detailed calculations that lead to these formulas.

Mean FPT The mean FPT's expression can be written as

$$\langle T \rangle = \mathbf{U}^\top \mathbf{A}^{-2} \mathbf{P}(0) = \mathbf{U}^\top \mathbf{A}_0^{-1} \mathbf{E}^{-1} \mathbf{A}_0^{-1} \mathbf{P}(0).
\qquad \text{[S3.11]}$$

The expression of $\mathbf{A}_0^{-1} \mathbf{P}(0)$ is just the first column of $\mathbf{A}_0$. Therefore

$$\mathbf{E}^{-1} \mathbf{A}_0^{-1} \mathbf{P}(0) = -\frac{1}{b}
\begin{bmatrix}
1 & \frac{k_1}{k_1+\gamma} \frac{(b+1)\gamma}{bk_0} & \cdots & \prod_{l=1}^{X-1} \left(\frac{k_l}{k_l+l\gamma} \frac{l(b+1)\gamma}{bk_{l-1}} \right) \\
0 & \frac{k_1}{k_1+\gamma} & \cdots & \frac{k_1}{k_1+\gamma} \prod_{l=2}^{X-1} \left(\frac{k_l}{k_l+l\gamma} \frac{l(b+1)\gamma}{bk_{l-1}} \right) \\
\vdots & \vdots & \vdots & \vdots \\
0 & 0 & \cdots & \frac{k_{X-1}}{k_{X-1}+(X-1)\gamma}
\end{bmatrix}
\begin{bmatrix}
\frac{b+1}{k_0} \\
\frac{1}{k_1} \\
\vdots \\
\frac{1}{k_{X-1}}
\end{bmatrix}
\qquad \text{[S3.12a]}$$

$$= -\frac{1}{b}
\begin{bmatrix}
\frac{b+1}{k_0} + \frac{1}{k_1} \frac{k_1}{k_1+\gamma} \frac{(b+1)\gamma}{bk_0} + \cdots + \frac{1}{k_{X-1}} \prod_{l=1}^{X-1} \left(\frac{k_l}{k_l+l\gamma} \frac{l(b+1)\gamma}{bk_{l-1}} \right) \\
\frac{1}{k_1} \frac{k_1}{k_1+\gamma} + \frac{1}{k_2} \frac{k_1}{k_1+\gamma} \frac{k_2}{k_2+2\gamma} \frac{2(b+1)\gamma}{bk_1} + \cdots + \frac{1}{k_{X-1}} \frac{k_1}{k_1+\gamma} \prod_{l=2}^{X-1} \left(\frac{k_l}{k_l+l\gamma} \frac{l(b+1)\gamma}{bk_{l-1}} \right) \\
\vdots \\
\frac{1}{k_{X-1}} \frac{k_{X-1}}{k_{X-1}+(X-1)\gamma}
\end{bmatrix}.
\qquad \text{[S3.12b]}$$

Since the vector $\mathbf{U}^\top \mathbf{A}_0^{-1} = -\begin{bmatrix} 1 & 1 & 1 & \cdots & 1 & 1 \end{bmatrix}$, $\mathbf{U}^\top \mathbf{A}_0^{-1}\mathbf{E}^{-1}\mathbf{A}_0^{-1}\mathbf{P}(0)$ is essentially negative sum of the elements of $\mathbf{E}^{-1}\mathbf{A}_0^{-1}\mathbf{P}(0)$. Therefore we have the expression of mean FPT is given by

$$\langle T \rangle = \frac{1}{k_0} + \frac{1}{b} \sum_{i=1}^{X} \frac{k_{i-1}}{k_{i-1}+(i-1)\gamma} \left(\frac{1}{k_{i-1}} + \sum_{j=i+1}^{X} \frac{1}{k_{j-1}} \prod_{l=i}^{j-1} \frac{k_l}{k_l+l\gamma} \frac{(b+1)l\gamma}{bk_{l-1}} \right). \qquad \text{[S3.13]}$$

Second order moment   The second order moment is given by

$$\langle T^2 \rangle = -2\mathbf{U}^\top \mathbf{A}^{-3}\mathbf{P}(0) = -2\mathbf{U}^\top \mathbf{A}_0^{-1}\mathbf{E}^{-1}\mathbf{A}_0^{-1}\mathbf{E}^{-1}\mathbf{A}_0^{-1}\mathbf{P}(0). \qquad \text{[S3.14]}$$

Let us use the notation η_i defined as

$$\eta_i := \frac{k_{i-1}}{k_{i-1}+(i-1)\gamma} \left(\frac{1}{k_{i-1}} + \sum_{j=i+1}^{X} \frac{1}{k_{j-1}} \prod_{l=i}^{j-1} \frac{k_l}{k_l+l\gamma} \frac{l(b+1)\gamma}{bk_{l-1}} \right) + \frac{b}{k_0}\delta_{i-1}, \qquad \text{[S3.15]}$$

where δ_{i-1} denotes the Kronecker delta which is one if $i=1$ and zero otherwise.

Using Eq. (S3.12b), we can write

$$\mathbf{E}^{-1}\mathbf{A}_0^{-1}\mathbf{P}(0) = -\frac{1}{b} \begin{bmatrix} \eta_1 \\ \eta_1 \\ \vdots \\ \eta_{X-2} \\ \eta_{X-1} \end{bmatrix}. \qquad \text{[S3.16]}$$

Therefore

$$\mathbf{A}_0^{-1}\mathbf{E}^{-1}\mathbf{A}_0^{-1}\mathbf{P}(0) = \left(\frac{1}{b}\right)^2 \begin{bmatrix} \frac{b+1}{k_0} & 0 & \cdots & 0 & 0 \\ \frac{1}{k_1} & \frac{b+1}{k_1} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{1}{k_{X-2}} & \frac{1}{k_{X-2}} & \cdots & \frac{b+1}{k_{X-2}} & 0 \\ \frac{1}{k_{X-1}} & \frac{1}{k_{X-1}} & \cdots & \frac{1}{k_{X-1}} & \frac{b+1}{k_{X-1}} \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_{X-1} \\ \eta_X \end{bmatrix} \qquad \text{[S3.17a]}$$

$$= \left(\frac{1}{b}\right)^2 \begin{bmatrix} \frac{(b+1)\eta_1}{k_0} \\ \frac{\eta_1+(b+1)\eta_2}{k_1} \\ \vdots \\ \frac{\sum_{i=1}^{X-2}\eta_i+(b+1)\eta_{X-1}}{k_{X-2}} \\ \frac{\sum_{i=1}^{X-1}\eta_i+(b+1)\eta_X}{k_{X-1}} \end{bmatrix}. \qquad \text{[S3.17b]}$$

Using the notion $\xi_i = \sum_{j=1}^{i-1}\eta_i + (b+1)\eta_i$, we can write $\mathbf{E}^{-1}\mathbf{A}_0^{-1}\mathbf{E}^{-1}\mathbf{A}_0^{-1}\mathbf{P}(0)$ as

$$\left(\frac{1}{b}\right)^2 \begin{bmatrix} 1 & \frac{k_1}{k_1+\gamma}\frac{(b+1)\gamma}{bk_0} & \cdots & \prod_{l=1}^{X-1}\left(\frac{k_l}{k_l+l\gamma}\frac{l(b+1)\gamma}{bk_{l-1}}\right) \\ 0 & \frac{k_1}{k_1+\gamma} & \cdots & \frac{k_1}{k_1+\gamma}\prod_{l=2}^{X-1}\left(\frac{k_l}{k_l+l\gamma}\frac{l(b+1)\gamma}{bk_{l-1}}\right) \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \frac{k_{X-1}}{k_{X-1}+(X-1)\gamma} \end{bmatrix} \begin{bmatrix} \frac{\xi_1}{k_0} \\ \frac{\xi_2}{k_1} \\ \vdots \\ \frac{\xi_X}{k_{X-1}} \end{bmatrix} \qquad \text{[S3.18]}$$

$$= \left(\frac{1}{b}\right)^2 \begin{bmatrix} \frac{\xi_1}{k_0} + \frac{\xi_2}{k_1}\frac{k_1}{k_1+\gamma}\frac{(b+1)\gamma}{bk_0} + \cdots + \frac{\xi_X}{k_{X-1}}\prod_{l=1}^{X-1}\left(\frac{k_l}{k_l+l\gamma}\frac{l(b+1)\gamma}{bk_{l-1}}\right) \\ \frac{\xi_2}{k_1}\frac{k_1}{k_1+\gamma} + \cdots + \frac{\xi_X}{k_{X-1}}\frac{k_1}{k_1+\gamma}\prod_{l=2}^{X-1}\left(\frac{k_l}{k_l+l\gamma}\frac{l(b+1)\gamma}{bk_{l-1}}\right) \\ \vdots \\ \frac{\xi_X}{k_{X-1}}\frac{k_{X-1}}{k_{X-1}+(X-1)\gamma} \end{bmatrix}. \qquad \text{[S3.19]}$$

As $\mathbf{U}^\top \mathbf{A}_0^{-1}$ gives the negative summation of the elements of the column vector it pre-multiplies to, the second order moment of FPT can be given by following explicit formula

$$\langle T^2 \rangle = 2\left(\frac{1}{b}\right)^2 \sum_{i=1}^{X} \frac{k_{i-1}}{k_{i-1}+(i-1)\gamma} \left(\frac{\xi_i}{k_{i-1}} + \sum_{r=i+1}^{X} \frac{\xi_r}{k_{r-1}} \left(\prod_{l=i}^{r-1} \left(\frac{k_l}{k_l+l\gamma}\frac{l(b+1)\gamma}{bk_{l-1}} \right) \right) \right). \qquad \text{[S3.20]}$$

S4. Effect of model parameters on FPT statistics

Having developed the expressions for first two moments of FPT, we can investigate how different model parameters affect the mean and noise in FPT. Perhaps simplest possible case is to consider a constant transcription rate, and that protein does not degrade. For this case, the FPT moments in Eq. (S3.10) lead to

$$\langle T \rangle = \frac{1}{k}\left(\frac{X}{b}+1\right) \approx \frac{X}{bk}, \quad CV_T^2 = \frac{b^2+X+2bX}{(b+X)^2} \approx \frac{1+2b}{X}, \tag{S4.1}$$

where CV_T^2 represents the noise in FPT as quantified by its coefficient of variation squared (variance/mean^2; $\langle T^2 \rangle / \langle T \rangle^2 - 1$). There formulas have been discussed in main text in detail.

The next possibility is to study the case when protein degradation is considered along with constant transcription rate, i.e., $k_i = k$. For this case, the first two FPT moments are given by

$$\langle T \rangle = \frac{1}{k} + \frac{1}{b}\sum_{i=1}^{X}\frac{1}{k+(i-1)\gamma}\left(1+\sum_{j=i+1}^{X}\prod_{l=i}^{j-1}\left(\frac{l\gamma}{k+l\gamma}\frac{b+1}{b}\right)\right), \tag{S4.2a}$$

$$\langle T^2 \rangle = \frac{2}{b^2}\sum_{i=1}^{X}\frac{1}{k+(i-1)\gamma}\left(\xi_i+\sum_{r=i+1}^{X}\xi_r\prod_{l=1}^{r-1}\left(\frac{l\gamma}{k+l\gamma}\frac{b+1}{b}\right)\right), \tag{S4.2b}$$

where

$$\xi_i = \sum_{j=1}^{i-1}\eta_j+(b+1)\eta_i, \tag{S4.2c}$$

$$\eta_i = \frac{1}{k+(i-1)\gamma}\left(1+\sum_{j=i+1}^{X}\prod_{l=i}^{j-1}\frac{l\gamma}{k+l\gamma}\frac{b+1}{b}\right)+\frac{b}{k}\delta_{i-1}. \tag{S4.2d}$$

Some insights obtained from these formulas are discussed at length in the main text. Next we discuss some additional insights obtained from these results.

Independent tuning of mean FPT and noise in FPT. An important result that can be seen from these formulas is that the noise only depends upon the ratio k/γ and not on their individual values. To show this, let us rewrite the formulas with $\theta := k/\gamma$

$$\langle T \rangle = \frac{1}{\gamma}\left(\frac{1}{\theta}+\frac{1}{b}\sum_{i=1}^{X}\frac{1}{\theta+(i-1)}\left(1+\sum_{j=i+1}^{X}\prod_{l=i}^{j-1}\left(\frac{l}{\theta+l}\frac{b+1}{b}\right)\right)\right), \tag{S4.3a}$$

$$\langle T^2 \rangle = \frac{1}{\gamma^2}\left(\frac{2}{b^2}\sum_{i=1}^{X}\frac{1}{\theta+(i-1)}\left(\overline{\xi_i}+\sum_{r=i+1}^{X}\overline{\xi_r}\prod_{l=1}^{r-1}\left(\frac{l}{\theta+l}\frac{b+1}{b}\right)\right)\right), \tag{S4.3b}$$

where

$$\overline{\xi_i} = \sum_{j=1}^{i-1}\overline{\eta_j}+(b+1)\overline{\eta_i}, \tag{S4.3c}$$

$$\overline{\eta_i} = \frac{1}{\theta+(i-1)}\left(1+\sum_{j=i+1}^{X}\prod_{l=i}^{j-1}\frac{l}{\theta+l}\frac{b+1}{b}\right)+\frac{b}{\theta}\delta_{i-1}. \tag{S4.3d}$$

Thus, the noise CV_T^2 can be written as

$$CV_T^2 = \frac{\langle T^2 \rangle}{\langle T \rangle^2} - 1 \tag{S4.4a}$$

$$= \frac{\frac{2}{b^2}\sum_{i=1}^{X}\frac{1}{\theta+(i-1)}\left(\overline{\xi_i}+\sum_{r=i+1}^{X}\overline{\xi_r}\prod_{l=1}^{r-1}\left(\frac{l}{\theta+l}\frac{b+1}{b}\right)\right)}{\left(\frac{1}{\theta}+\frac{1}{b}\sum_{i=1}^{X}\frac{1}{\theta+(i-1)}\left(1+\sum_{j=i+1}^{X}\prod_{l=i}^{j-1}\left(\frac{l}{\theta+l}\frac{b+1}{b}\right)\right)\right)^2} - 1, \tag{S4.4b}$$

which depends only on $\theta = k/\gamma$, and not on the values of k and γ themselves. This result has an important implication that if the noise is in desirable range for a given a set of parameters k, b, γ, X, then the mean FPT can be independently tuned by varying degradation rate and keeping k/γ constant.
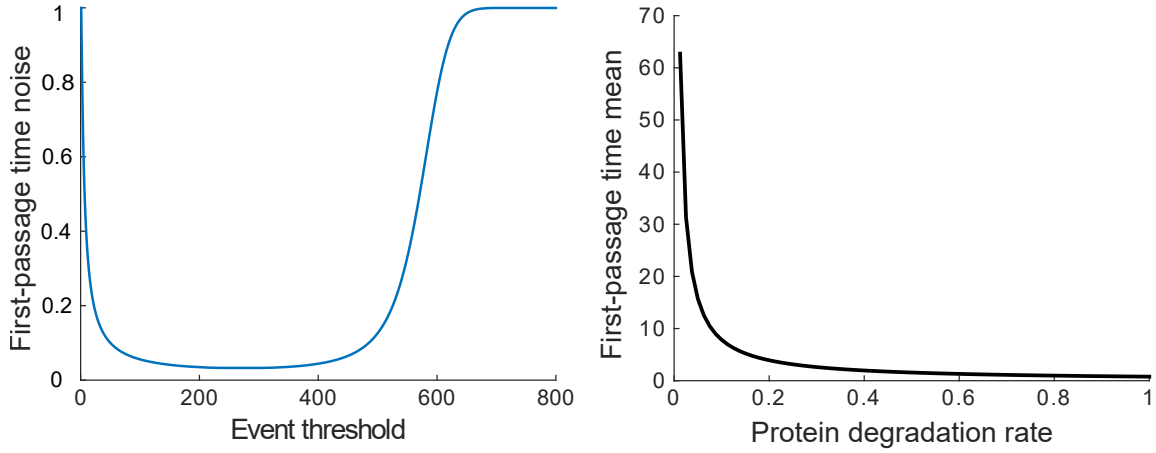
Fig. S2. For a given steady-state protein level, noise in FPT is minimized at an optimal event threshold. *Left:* Noise in FPT (CV_T^2) is plotted as the event threshold (X) is varied. At $X = 1$, CV_T^2 is one which corresponds to an exponential distribution of FPT. As X is increased, CV_T^2 decreases until it hits a minimum and then increases back to one once X gets above the steady-state protein level. This curve is invariant of changes in the transcription rate k or the degradation rate γ as long as k/γ is kept constant. *Right:* The mean first-passage time is plotted as γ is varied while keeping k/γ constant. This allows an appropriate mean FPT to be chosen without affecting the noise.

Effect of model parameters for constant production and degradation. Using the formulas in Eq. (S4.2), we can investigate how different parameters impact mean and noise in FPT. As discussed in the main text, fixing other parameters and varying the event threshold results in a U-shape profile for noise in FPT (Fig. 2 in main text, Fig. S2 (left)). Furthermore, because the noise is minimized at a certain value of X, the cell might choose this value and vary γ while keeping k/γ constant to get a required mean FPT (Fig. S2 (right)). The mean burst size b can also be varied while keeping the steady-state protein level same by appropriately changing k/γ. It is seen that when X is smaller than or comparable to the steady-state protein level, increasing b increases noise in timing. However, when X is higher than the steady-state protein level, the noise in timing is decreased by increasing b (Fig. 2 in the main text).

Remark: The result that CV_T^2 is independent of individual values of k and γ can be straightforwardly extended to the FPT moments with feedback (Eq. (S3.13), Eq. (S3.20)). For this purpose, we define rescaled transcription rates $\mathcal{K}_i = k_i/k_0$, and the ratio $\theta := k_0/\gamma$. Then, the FPT moments can be rewritten as

$$\langle T \rangle = \frac{1}{\gamma} \left(\frac{1}{\theta} + \frac{1}{b} \sum_{i=1}^{X} \frac{\theta \mathcal{K}_{i-1}}{\theta \mathcal{K}_{i-1} + (i-1)} \left(\frac{1}{\theta \mathcal{K}_{i-1}} + \sum_{j=i+1}^{X} \frac{1}{\theta \mathcal{K}_{j-1}} \prod_{l=i}^{j-1} \frac{\mathcal{K}_l}{\theta \mathcal{K}_l + l} \frac{(b+1)l}{b \mathcal{K}_{l-1}} \right) \right), \qquad \text{[S4.5a]}$$

$$\langle T^2 \rangle = \frac{1}{\gamma^2} \left(\frac{2}{b^2} \sum_{i=1}^{X} \frac{\theta \mathcal{K}_{i-1}}{\theta \mathcal{K}_{i-1} + (i-1)} \left(\frac{b \overline{\eta_i} + \sum_{j=1}^{i} \overline{\eta_j}}{\theta \mathcal{K}_{i-1}} + \sum_{r=i+1}^{X} \frac{b \overline{\eta_i} + \sum_{j=1}^{r} \overline{\eta_j}}{\theta \mathcal{K}_{r-1}} \prod_{l=i}^{r-1} \frac{\mathcal{K}_l}{\theta \mathcal{K}_l + l} \frac{(b+1)l}{b \mathcal{K}_{l-1}} \right) \right), \qquad \text{[S4.5b]}$$

where the terms $\overline{\eta_i}$ are given by

$$\overline{\eta_i} = \frac{b}{\theta} \delta_{i-1} + \frac{\theta \mathcal{K}_{i-1}}{\theta \mathcal{K}_{i-1} + (i-1)} \left(\frac{1}{\theta \mathcal{K}_{i-1}} + \sum_{j=i+1}^{X} \frac{1}{\theta \mathcal{K}_{j-1}} \prod_{l=i}^{j-1} \frac{\mathcal{K}_l}{\theta \mathcal{K}_l + l} \frac{(b+1)l}{b \mathcal{K}_{l-1}} \right). \qquad \text{[S4.5c]}$$

It can be seen that the noise only depends upon k_0/γ and not on their individual values. While one could have defined θ as scaling some other transcription rate k_i by γ and made similar observation, the significance of using k_0/γ can be understood by considering Hill function forms of the feedbacks. As in the main text Eq. (16), let us consider a positive feedback implemented as

$$k_i = k_{\max} \frac{r + (ci)^H}{1 + (ci)^H}, \qquad \text{[S4.6]}$$

where $k_{\max}$ is the maximum transcription rate, $k_{\max} r$ is the basal transcription rate with $0 < r < 1$ representing the basal strength, c is the feedback strength, and H is the Hill coefficient. In this case, we have

$$k_0 = k_{\max} r, \qquad \mathcal{K}_i = \frac{r + (ci)^H}{r(1 + (ci)^H)}. \qquad \text{[S4.7]}$$

Typically r is a constant small number less than one. Thus, it can be observed that the noise must only depend only on $k_{\max}/\gamma$ and not on their individual values. One can make similar argument to see that in the negative feedback case (Eq. (15) in main text)

$$k_i = \frac{k_{\max}}{1 + (ci)^H},$$
[S4.8]

as well the noise depends only on $k_{\max}/\gamma$. These results collectively show that the noise in FPT can be tuned by choosing a value of $k_{\max}/\gamma$ and the mean FPT can be independently tuned by choosing a value of γ without affecting the noise.

S5. Optimal feedback when protein does not degrade

As mentioned in the main text, our objective is to find optimal feedback strategy that minimizes $\langle T^2 \rangle$ such that $\langle T \rangle$ is fixed. For calculation purposes, we will denote this constraint as $\langle T \rangle = t_{opt}$. Let m represents the Lagrange's multiplier, then we define the following objective function

$$\phi := \langle T^2 \rangle + m \left(\langle T \rangle - t_{opt} \right).$$
[S5.1]

The optimization problem is solved in two steps. First, we determine the critical points. Second, we find the critical point corresponding to a global minimum.

Determining the critical points requires the following system of equations to be solved

$$\frac{\partial \langle T^2 \rangle}{\partial k_i} = m \frac{\partial \langle T \rangle}{\partial k_i}, 0 \leq i \leq X - 1,$$
[S5.2a]

$$\langle T \rangle = t_{opt}.$$
[S5.2b]

The expressions of the first-two moments of FPT when protein does not degrade can be obtained by substituting $\gamma = 0$ in the moment expressions. These are given by

$$\langle T \rangle = \frac{1}{k_0} + \frac{1}{b} \sum_{i=0}^{X-1} \frac{1}{k_i},$$
[S5.3a]

$$\langle T^2 \rangle = \frac{2}{b^2} \left(\frac{\tau_0}{bk_0} + \sum_{i=0}^{X-1} \frac{\tau_i}{k_i} \right), \text{ where } \tau_i := \frac{b}{k_i} + \sum_{j=i}^{X-1} \frac{1}{k_j}.$$
[S5.3b]

The optimization problem in Eq. (S5.2) requires calculation of the first order derivatives of $\langle T \rangle$ and $\langle T^2 \rangle$. The derivatives of $\langle T \rangle$ with respect to k_i's are given by

$$\frac{\partial \langle T \rangle}{\partial k_0} = \frac{b+1}{b} \left(-\frac{1}{k_0^2} \right); \frac{\partial \langle T \rangle}{\partial k_i} = \frac{1}{b} \left(-\frac{1}{k_i^2} \right), 1 \leq i \leq X - 1.$$
[S5.4]

Similarly, the derivative of $\langle T^2 \rangle$ are

$$\frac{\partial \langle T^2 \rangle}{\partial k_0} = 2 \left(\frac{1}{b} \right)^2 \left(-\frac{2(b+1)^2}{k_0^3} - \frac{b+1}{k_0^2} \sum_{j=1}^{X-1} \frac{1}{k_j} \right),$$
[S5.5a]

$$\frac{\partial \langle T^2 \rangle}{\partial k_i} = 2 \left(\frac{1}{b} \right)^2 \left(\frac{b+1}{k_0} + \frac{2(b+1)}{k_i} + \sum_{\substack{j=1 \\ j \neq i}}^{X-1} \frac{1}{k_j} \right) \left(\frac{-1}{k_i^2} \right), 1 \leq i \leq X - 1.$$
[S5.5b]

Substituting these expressions and assuming $k_0 \neq 0$, $k_i \neq 0$, the system of equations to be solved becomes

$$\frac{2}{b} \left(\frac{2(b+1)}{k_0} + \sum_{j=1}^{X-1} \frac{1}{k_j} \right) = m,$$
[S5.6a]

$$\frac{2}{b} \left(\frac{b+1}{k_0} + \frac{2(b+1)}{k_i} + \sum_{\substack{j=1 \\ j \neq i}}^{X-1} \frac{1}{k_j} \right) = m, 1 \leq i \leq X - 1,$$
[S5.6b]

$$\frac{1}{b} \left(\frac{b+1}{k_0} + \sum_{j=1}^{X-1} \frac{1}{k_j} \right) = t_{opt}.$$
[S5.6c]

Solution to these equations is given by

$$k_0 = \frac{(b+1)(2b+X)}{(2b+1)bt_{opt}} \tag{S5.7a}$$

$$k_i = \frac{2b+1}{b+1}k_0 = \frac{2b+X}{bt_{opt}}, 1 \le i \le X-1, \tag{S5.7b}$$

$$m = \frac{2(b+1)(4b+X+1)}{(2b+1)bk_0} = \frac{2t_{opt}(4b+X+1)}{2b+X}. \tag{S5.7c}$$

We have calculated the critical point for the optimization problem. However, it needs to be checked whether its an minimum or maximum. For this purpose, we consider the bordered Hessian as follows.

$$D_\phi = \begin{bmatrix} \frac{\partial^2 \phi}{\partial m^2} & \frac{\partial^2 \phi}{\partial m \partial k_0} & \frac{\partial^2 \phi}{\partial m \partial k_1} & \cdots & \frac{\partial^2 \phi}{\partial m \partial k_{X-1}} \\ \frac{\partial^2 \phi}{\partial k_0 \partial m} & \frac{\partial^2 \phi}{\partial k_0^2} & \frac{\partial^2 \phi}{\partial k_0 \partial k_1} & \cdots & \frac{\partial^2 \phi}{\partial k_0 \partial k_{X-1}} \\ \frac{\partial^2 \phi}{\partial k_1 \partial m} & \frac{\partial^2 \phi}{\partial k_1 \partial k_0} & \frac{\partial^2 \phi}{\partial k_1^2} & \cdots & \frac{\partial^2 \phi}{\partial k_1 \partial k_{X-1}} \\ \vdots & \vdots & \vdots & \vdots & \\ \frac{\partial^2 \phi}{\partial k_{X-1} \partial m} & \frac{\partial^2 \phi}{\partial k_{X-1} \partial k_0} & \frac{\partial^2 \phi}{\partial k_{X-1} \partial k_1} & \cdots & \frac{\partial^2 \phi}{\partial k_{X-1}^2} \end{bmatrix}. \tag{S5.8}$$

We will show that all the principal minors of this matrix are negative. To start with, let us first determine the second order derivatives of $\langle T \rangle$.

$$\frac{\partial^2 \langle T \rangle}{\partial k_0^2} = \frac{2(b+1)}{bk_0^3}; \quad \frac{\partial^2 \langle T \rangle}{\partial k_i^2} = \frac{2}{bk_i^3}, 1 \le i \le X-1; \quad \frac{\partial^2 \langle T \rangle}{\partial k_i \partial k_j} = 0, 0 \le i, j \le X-1, i \ne j. \tag{S5.9}$$

Similarly, the derivatives for $\langle T^2 \rangle$ are given by

$$\frac{\partial^2 \langle T^2 \rangle}{\partial k_0^2} = 2\left(\frac{1}{b}\right)^2 \left(\frac{6(b+1)^2}{k_0^4} + \frac{2(b+1)}{k_0^3} \sum_{j=1}^{X-1} \frac{1}{k_j} \right), \tag{S5.10a}$$

$$\frac{\partial^2 \langle T^2 \rangle}{\partial k_i^2} = 2\left(\frac{1}{b}\right)^2 \left(\frac{2(b+1)}{k_0 k_i^3} + \frac{6(b+1)}{k_i^4} + \frac{2}{k_i^3} \sum_{j=1}^{X-1} \frac{1}{k_j} \right), 1 \le i \le X-1, \tag{S5.10b}$$

$$\frac{\partial^2 \langle T^2 \rangle}{\partial k_0 \partial k_i} = \frac{\partial^2 \langle T^2 \rangle}{\partial k_i \partial k_0} = 2\left(\frac{1}{b}\right)^2 \frac{b+1}{k_0^2 k_i^2}, 1 \le i \le X-1, \tag{S5.10c}$$

$$\frac{\partial^2 \langle T^2 \rangle}{\partial k_i \partial k_j} = 2\left(\frac{1}{b}\right)^2 \frac{1}{k_i^2 k_j^2}, 1 \le i, j \le X-1, i \ne j. \tag{S5.10d}$$

We can now determine the elements of the bordered Hessian matrix in equation Eq. (S5.8) computed at the solution given by equations Eq. (S5.7)

$$\frac{\partial^2 \phi}{\partial m^2} = \frac{\partial^2 \left(\langle T^2 \rangle + m(\langle T \rangle - t_{opt}) \right)}{\partial m^2} = 0, \tag{S5.11a}$$

$$\frac{\partial^2 \phi}{\partial m \partial k_0} = \frac{\partial^2 \phi}{\partial k_0 \partial m} = \frac{\partial \langle T \rangle}{\partial k_0} = \frac{b+1}{b}\left(-\frac{1}{k_0^2}\right). \tag{S5.11b}$$

$$\frac{\partial^2 \phi}{\partial m \partial k_i} = \frac{\partial^2 \phi}{\partial k_i \partial m} = \frac{\partial \langle T \rangle}{\partial k_i} = \frac{1}{b}\left(-\frac{1}{k_i^2}\right). \tag{S5.11c}$$

$$\frac{\partial^2 \phi}{\partial k_0^2} = \frac{\partial^2 \left(\langle T^2 \rangle + m(\langle T \rangle - t_{opt}) \right)}{\partial k_0^2} = \frac{4(b+1)^2(10b+2X+3)}{(2b+1)b^2 k_0^4} \tag{S5.11d}$$

$$\frac{\partial^2 \phi}{\partial k_i^2} = \frac{\partial^2 \left(\langle T^2 \rangle + m(\langle T \rangle - t_{opt}) \right)}{\partial k_i^2} = \frac{4(b+1)^4(9b+2X+4)}{(2b+1)^4 b^2 k_0^4}. \tag{S5.11e}$$

$$\frac{\partial^2 \phi}{\partial k_0 k_i} = \frac{\partial^2 \phi}{\partial k_i k_0} = \frac{\partial^2 \left(\langle T^2 \rangle + m(\langle T \rangle - t_{opt}) \right)}{\partial k_0 \partial k_i} = 2\left(\frac{1}{b}\right)^2 \frac{b+1}{k_0^2 k_i^2} \tag{S5.11f}$$

$$\frac{\partial^2 \phi}{\partial k_j k_i} = \frac{\partial^2 \phi}{\partial k_i k_j} = \frac{\partial^2 \left(\left\langle T^2 \right\rangle + m \left(\left\langle T \right\rangle - t_{opt} \right) \right)}{\partial k_j \partial k_i} = 2 \left(\frac{1}{b} \right)^2 \frac{1}{k_j^2 k_i^2} \tag{S5.11g}$$

It can be noted elements of D_ϕ are from a set six quantities. Defining

$$q_1 := \frac{b+1}{b} \left(-\frac{1}{k_0^2} \right), \qquad\qquad q_2 := \frac{1}{b} \left(-\frac{1}{k_0^2} \right), \tag{S5.12a}$$

$$q_3 := \frac{4(b+1)^2(10b+2X+3)}{(2b+1)b^2 k_0{}^4}, \quad q_4 := \frac{2(b+1)^3}{(2b+1)^2 b^2 k_0{}^4}, \tag{S5.12b}$$

$$q_5 := \frac{4(b+1)^4(9b+2X+4)}{(2b+1)^4 b^2 k_0{}^4}, \quad q_6 := \frac{2(b+1)^2}{(2b+1)^2 b^2 k_0^4} \tag{S5.12c}$$

we can write D_ϕ as

$$D_\phi = \begin{bmatrix} 0 & q_1 & q_2 & \cdots & q_2 \\ q_1 & q_3 & q_4 & \cdots & q_4 \\ q_2 & q_4 & q_5 & \cdots & q_6 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ q_2 & q_4 & q_6 & \cdots & q_5 \end{bmatrix}. \tag{S5.13}$$

Let us denote by $\mathcal{M}(n)$ the principal minor of the matrix D_ϕ of size $n \times n$. It can be easily seen that $\mathcal{M}(1) = 0$, $\mathcal{M}(2) = 0 - q_1^2 < 0$ and $\mathcal{M}(3) = -q_2^2 q_3 + 2q_1 q_2 q_4 - q_1^2 q_5$. For $4 \leq n \leq X$, we perform the following two elementary operations on D_ϕ

- $col_r = col_r - col_{r-1}$

- $row_r = row_r - row_{r-1}$

for $r = n, n-1, ..., 1$. This yields

$$\mathcal{M}(n) = 2(q_5 - q_6)\mathcal{M}(n-1) - (q_5 - q_6)^2 \mathcal{M}(n-2), \quad 4 \leq n \leq X. \tag{S5.14}$$

The solution to the above recursive equation is given by

$$\mathcal{M}(n) = -(q_5 - q_6)^{n-3} \left((n-2) \left(q_2^2 q_3 - 2q_1 q_2 q_4 + q_1^2 q_6 \right) + \left(q_1^2 q_6 + q_1^2 q_5 \right) \right). \tag{S5.15}$$

It can be easily checked that $\mathcal{M}(n)$ is negative because $q_5 > q_6$, $q_2^2 q_3 - 2q_1 q_2 q_4 + q_1^2 q_6 > 0$ and $q_1^2 q_6 + q_1^2 q_5 > 0$. This proves that the critical point indeed corresponds to a minimum.

Stochastic Simulations considering explicit mRNA dynamics. Note that the stochastic gene expression model considers production of proteins in bursts, which relies upon the assumption that each mRNA degrades instantaneously. Here, we relax this assumption and perform stochastic simulations to see whether a no feedback is still an optimal strategy.

In the Fig. 3 of the main text, we performed simulations assuming a linear form of feedback and our results suggest that the result from the analysis of the bursty approximation holds in this case. As an extension to that, here in Fig. S3 we consider the feedbacks implemented via Hill functions as in Eq. (S4.6)/ Eq. (S4.8).

S6. Optimal feedback strategy in presence of protein degradation

In the previous section, we have derived analytical expression of the optimal feedback strategy that minimizes $\left\langle T^2 \right\rangle$ such that $\left\langle T \right\rangle$ is constant for a stable protein. When the protein decay is considered, the expressions of $\left\langle T \right\rangle$ and $\left\langle T^2 \right\rangle$ in Eq. (S3.13) and Eq. (S3.20) are too convoluted to solve for optimal transcription rates analytically, and we take a numerical approach to solve the optimization problem. For this purpose, we fix the threshold $X = 10$ molecules, and mean burst size $b = 1$ molecules. Using numerical solvers, we search the parameter space of the transcription rates $k_i, i \in \{0, 1, ..., 9\}$ for various values of the protein degradation rate γ. As shown in Fig. S4, when $\gamma = 0$, the transcription rates are equal except for the first one. This is consistent with the expressions in Eq. (S5.7a)-Eq. (S5.7b). Further, as the protein degradation rate is increased, the transcription rates first increase and then decrease, suggesting a mixed feedback strategy. To keep the results biologically meaningful, the optimal feedback strategy is investigated in the main text by considering feedbacks which follow Hill functions forms.

Optimal feedback strategy as X or b are changed. . While in the main text, we showed results for a given threshold $X = 500$ molecules and mean burst size $b = 2$ molecules. Here, we provide results when X and b are changed (Fig. S5). It can be seen that decreasing X does not change the qualitative trend of minimum CV_T^2 as γ is varied. However, the overall noise level increases ((Fig. S5) (Left)). Similar effect is seen as the mean burst size b is increased ((Fig. S5) (Right)).
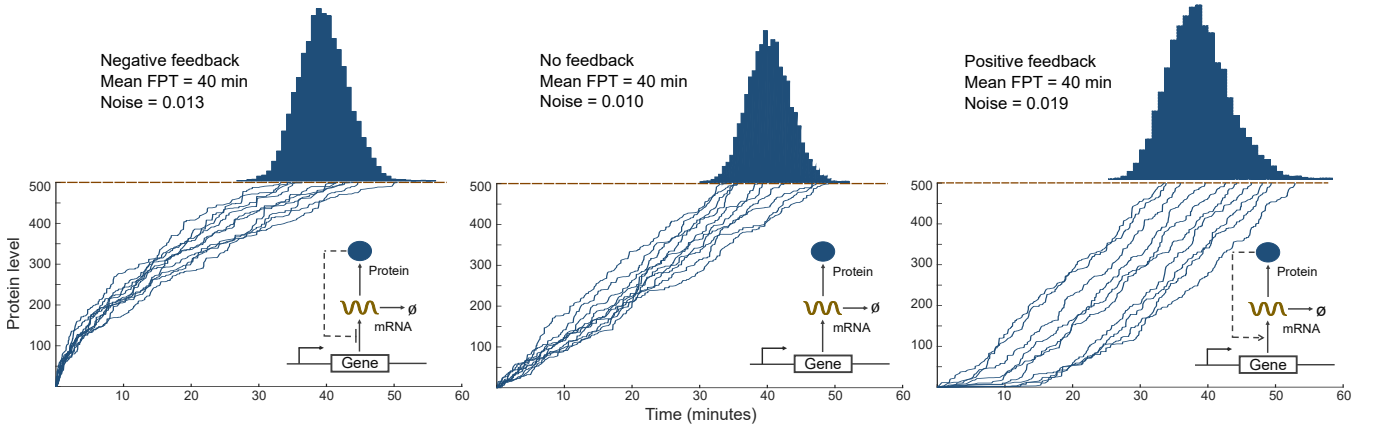
Fig. S3. No feedback provides minimum noise in timing for a stable protein. Protein trajectories obtained using the Stochastic Simulation Algorithm (SSA) for a stochastic gene expression model with negative feedback (left), no feedback (middle), and positive feedback (right) are shown. The threshold is assumed to be 500 molecules and the feedbacks are implemented using Hill function forms of the feedbacks as in Eq. (S4.6)/ Eq. (S4.8). The maximum transcription rate $k_{\max}$ is taken such that the mean FPT is kept constant at 40 minutes while other model parameters are assumed to be fixed as: $c = 0.01$, $H = 1$, $r = 0.05$. The mRNA half-life is assumed to be 2.7 min, and proteins are translated from mRNAs at a rate $0.5\ \mathrm{min}^{-1}$ (these correspond to a mean burst size $b = 2$). Histograms on the top represent distribution of FPT from $10,000$ simulations.
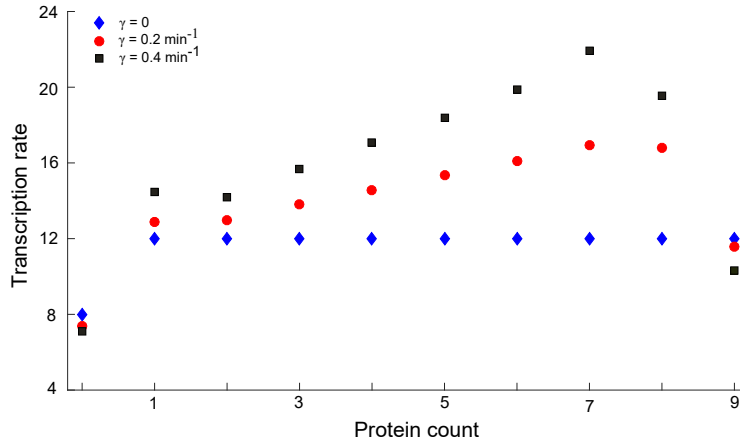
Fig. S4. Optimal feedback strategy for unstable protein. The optimal transcription rates obtained via numerical optimization for different values of degradation rate are shown. The event threshold is assumed to be 10 molecules, and the mean FPT is constrained to be 1 minute. The optimal transcription rates show a non-monotonic pattern as the protein count increases, suggesting presence of a mixed feedback strategy.

Protein trajectories for unstable protein. As shown in Fig. 4 of the main text, a positive feedback strategy gives minimum noise in FPT around a given mean time. Here we show the protein trajectories for different feedback strategies and compare them with those of a no feedback case in Fig. 3 of the main text/Fig. S3.

It is seen that the protein trajectories (Fig. S6/Fig. S7) look similar to those obtained by a no feedback strategy in stable protein case (Fig. 3 of the main text/Fig. S3). Thus, it appears that when the protein is allowed to degrade, a positive feedback attempts to mimic the path taken by the no feedback strategy when protein were stable.

As discussed in the main text, one way to understand this result is to consider the linear feedback form

$$k_i = c_1 + c_2 i \qquad [\mathrm{S6.1}]$$

in which case the mean protein levels evolve according to the following ordinary differential equation

$$\frac{dx(t)}{dt} = b(c_1 + c_2 x) - \gamma x, \quad x(0) = 0. \qquad [\mathrm{S6.2}]$$

While no feedback $(c_2 = 0)$, and negative feedback $(c_2 < 0)$ in Eq. (S6.2) create nonlinear protein trajectories (Fig. S7, choosing a positive value

$$c_2 \approx \frac{\gamma}{b} \qquad [\mathrm{S6.3}]$$

results in linear $x(t)$, and hence minimal noise in event timing. Indeed, the optimal c_2 multiplied by the mean burst size b takes values close to the degradation rate γ as shown in Fig. S8.
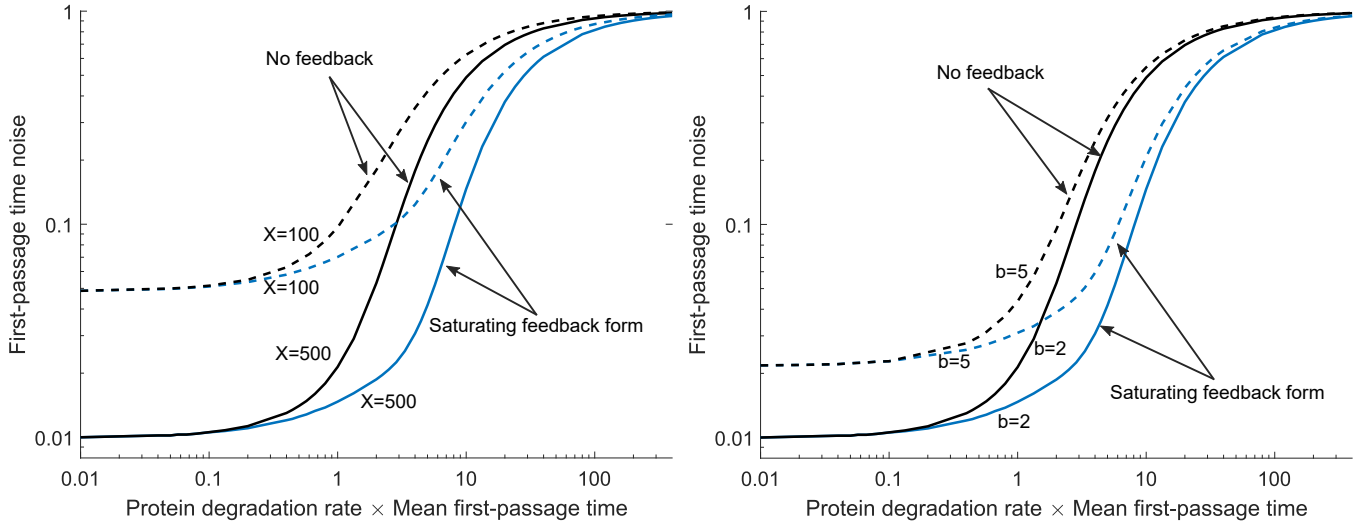
Fig. S5. A positive feedback remains to be optimal as the event threshold or the mean burst size is varied. *Left:* The minimum value of CV_T^2 obtained via positive feedback increases monotonically with the protein degradation rate. Decreasing the event threshold X shifts the curves upwards without changing the qualitative shape. *Right:* Increasing the event threshold X shifts the curves upwards without changing the qualitative shape. The parameters for the positive (optimal) feedback used are $r = 0.05$, and $H = 1$.

S7. Optimal feedback when protein does not degrade in presence of extrinsic noise

In this section, assuming that the protein does not degrade, we investigate how the optimal regulation strategy deviates from a no feedback in presence of a static extrinsic noise. We consider two possibilities here: one, the extrinsic noise affects the mean burst size; two, the extrinsic noise affects the transcription rate. For the first case, we assume that the mean burst size is drawn from an arbitrary positive-valued distribution. The second case is analyzed by assuming that a factor Z multiplies with the transcription rates resulting in an effective transcription rate when $x(t) = i$ to be $k_i Z$.

Optimal feedback when the mean burst size is drawn from a distribution. Let us assume the mean burst size a random variable with probability density function $f_b(\beta)$. Thus, the number of proteins in a burst are geometrically distributed with mean b where b itself is a random variable. The mean and second order moment are given by

$$\langle T | b = \beta \rangle = \frac{1}{\beta} \left(\frac{\beta + 1}{k_0} + \sum_{i=1}^{X-1} \frac{1}{k_i} \right), \tag{S7.1a}$$

$$\implies \langle T \rangle = \left(1 + \left\langle \frac{1}{b} \right\rangle \right) \frac{1}{k_0} + \left\langle \frac{1}{b} \right\rangle \sum_{i=1}^{X-1} \frac{1}{k_i}. \tag{S7.1b}$$

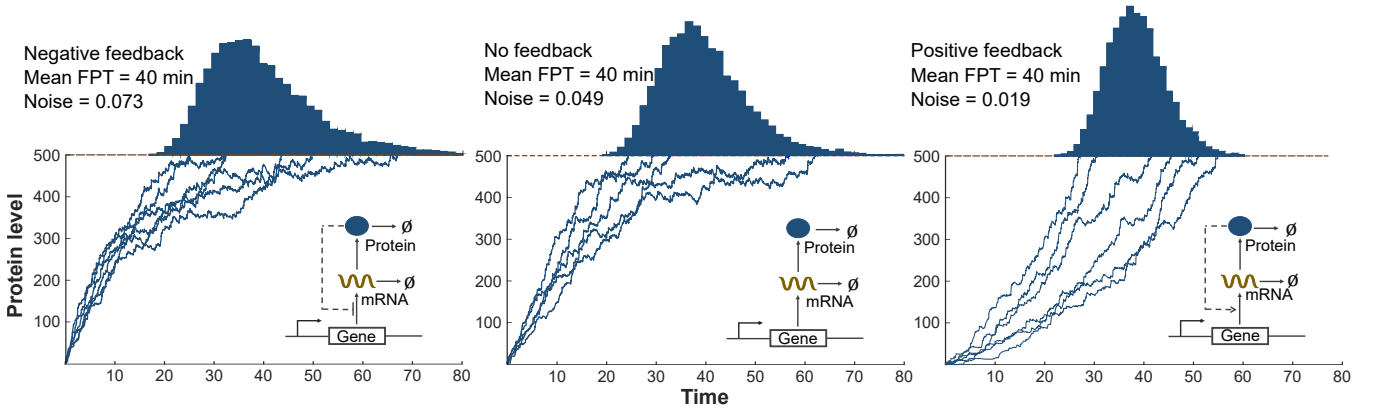$$\langle T^2 | b = \beta \rangle = \left(\frac{1}{\beta} \right)^2 \left(\frac{\beta+1}{k_0} \left(\frac{\beta+1}{k_0} + \sum_{i=1}^{X-1} \frac{1}{k_i} \right) + \sum_{i=1}^{X-1} \frac{1}{k_i} \left(\frac{\beta+1}{k_i} + \sum_{j=i+1}^{X-1} \frac{1}{k_j} \right) \right), \tag{S7.2a}$$

$$\implies \langle T^2 \rangle = \left(\left\langle \frac{1}{b^2} \right\rangle + 2 \left\langle \frac{1}{b} \right\rangle + 1 \right) \frac{1}{k_0^2} + \left(\left\langle \frac{1}{b^2} \right\rangle + \left\langle \frac{1}{b} \right\rangle \right) \left(\frac{1}{k_0} \sum_{i=1}^{X-1} \frac{1}{k_i} + \sum_{i=1}^{X-1} \frac{1}{k_i^2} \right)$$

$$+ \left\langle \frac{1}{b^2} \right\rangle \sum_{i=1}^{X-1} \frac{1}{k_i} \sum_{j=i+1}^{X-1} \frac{1}{k_j}. \tag{S7.2b}$$

Defining $\alpha_1 = \left\langle \frac{1}{b} \right\rangle$ and $\alpha_2 = \left\langle \frac{1}{b^2} \right\rangle$, we have

$$\langle T \rangle = (1 + \alpha_1) \frac{1}{k_0} + \alpha_1 \sum_{i=1}^{X-1} \frac{1}{k_i}, \tag{S7.3a}$$

$$\langle T^2 \rangle = (1 + 2\alpha_1 + \alpha_2) \frac{1}{k_0^2} + (\alpha_1 + \alpha_2) \left(\frac{1}{k_0} \sum_{i=1}^{X-1} \frac{1}{k_i} + \sum_{i=1}^{X-1} \frac{1}{k_i^2} \right) + \alpha_2 \sum_{i=1}^{X-1} \frac{1}{k_i} \sum_{j=i+1}^{X-1} \frac{1}{k_j} \tag{S7.3b}$$

Fig. S6. Shape of protein trajectories for an unstable protein with feedbacks implemented via Hill functions. Protein trajectories obtained using the Stochastic Simulation Algorithm (SSA) for a stochastic gene expression model with negative feedback (left), no feedback (middle), and positive feedback (right) are shown. The threshold is assumed to be 500 molecules and the feedbacks are implemented using Hill function forms of the feedbacks as in Eq. (S4.6)/ Eq. (S4.8). The maximum transcription rate k_{max} is taken such that the mean FPT is kept constant at 40 minutes while other model parameters are assumed to be fixed as: $c = 0.001$, $H = 1$, $r = 0.05$, $\gamma = 0.05\mathrm{min}^{-1}$. The mRNA half-life is assumed to be 2.7 min, and proteins are translated from mRNAs at a rate 0.5 min^{-1} (these correspond to a mean burst size $b = 2$). Histograms on the top represent distribution of FPT from $10,000$ simulations.
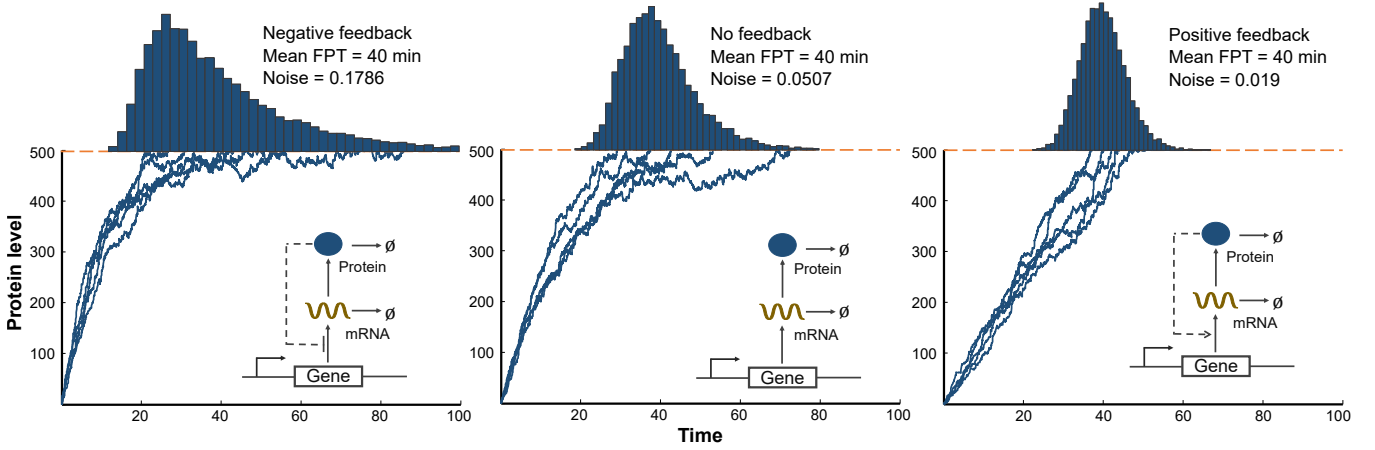
Fig. S7. Shape of protein trajectories for an unstable protein with feedbacks implemented via a linear form. Protein trajectories obtained using the Stochastic Simulation Algorithm (SSA) for a stochastic gene expression model with negative feedback (left), no feedback (middle), and positive feedback (right) are shown. The threshold is assumed to be 500 molecules and the feedbacks are implemented using a linear form $k_i = c_1 + c_2 i$. The parameter c_1 is taken such that the mean FPT is kept constant at 40 minutes while other model parameters are assumed to be fixed as: $c_2 = 0.05\mathrm{min}^{-1}$ (positive feedback), $c_2 = -0.05\mathrm{min}^{-1}$ (negative feedback), $\gamma = 0.05\mathrm{min}^{-1}$. The mRNA half-life is assumed to be 2.7 min, and proteins are translated from mRNAs at a rate 0.5 min^{-1} (these correspond to a mean burst size $b = 2$). Histograms on the top represent distribution of FPT from $10,000$ simulations.
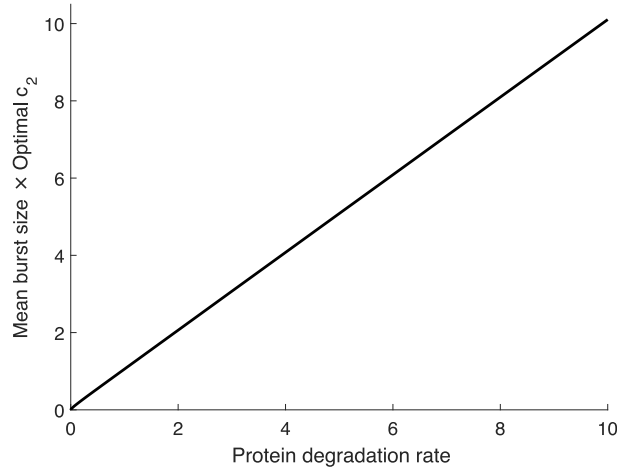
The derivatives of $\langle T \rangle$ with respect to k_i's are given by

$$\frac{\partial \langle T \rangle}{\partial k_0} = (1 + \alpha_1)\left(-\frac{1}{k_0^2}\right); \quad \frac{\partial \langle T \rangle}{\partial k_i} = \alpha_1 \left(-\frac{1}{k_i^2}\right), \; 1 \le i \le X - 1. \qquad [\text{S7.4}]$$

Similarly, the derivative of $\langle T^2 \rangle$ are

$$\frac{\partial \langle T^2 \rangle}{\partial k_0} = 2\left(-\frac{2(1 + 2\alpha_1 + \alpha_2)}{k_0^3} - \frac{(\alpha_1 + \alpha_2)}{k_0^2}\sum_{j=1}^{X-1}\frac{1}{k_j}\right); \qquad [\text{S7.5a}]$$

$$\frac{\partial \langle T^2 \rangle}{\partial k_i} = 2\left(\frac{\alpha_1 + \alpha_2}{k_0} + \frac{2(\alpha_1 + \alpha_2)}{k_i} + \alpha_2 \sum_{\substack{j=1 \\ j \ne i}}^{X-1}\frac{1}{k_j}\right)\left(\frac{-1}{k_i^2}\right), 1 \le i \le X - 1. \qquad [\text{S7.5b}]$$

Fig. S8. Linear increase in c_2 with degradation rate. *Left:* A linear feedback of the form $k_i = c_1 + c_2 i$ is considered. As expected from Eq. (S6.3), $c_2 \approx \gamma/b$. The parameter values are taken to be $X = 500$ molecules, $b = 2$.

To find a critical point, we have to solve the following system of equations

$$2 \left(-\frac{2 \left(1 + 2\alpha_1 + \alpha_2\right)}{k_0^3} - \frac{\left(\alpha_1 + \alpha_2\right)}{k_0^2} \sum_{j=1}^{X-1} \frac{1}{k_j} \right) = m \left(1 + \alpha_1\right) \left(-\frac{1}{k_0^2} \right) \tag{S7.6a}$$

$$2 \left(\frac{\alpha_1 + \alpha_2}{k_0} + \frac{2 \left(\alpha_1 + \alpha_2\right)}{k_i} + \alpha_2 \sum_{\substack{j=1 \\ j \neq i}}^{X-1} \frac{1}{k_j} \right) \left(\frac{-1}{k_i^2} \right) = m\alpha_1 \left(-\frac{1}{k_i^2} \right), 1 \leq i \leq X-1, \tag{S7.6b}$$

$$\left(1 + \alpha_1\right) \frac{1}{k_0} + \alpha_1 \sum_{i=1}^{X-1} \frac{1}{k_i} = t_{opt}. \tag{S7.6c}$$

Assuming that $k_0, k_1, \ldots \neq 0$, we get

$$2 \left(\frac{2 \left(1 + 2\alpha_1 + \alpha_2\right)}{k_0} + \left(\alpha_1 + \alpha_2\right) \sum_{j=1}^{X-1} \frac{1}{k_j} \right) = m \left(1 + \alpha_1\right) \tag{S7.7a}$$

$$2 \left(\frac{\alpha_1 + \alpha_2}{k_0} + \frac{2 \left(\alpha_1 + \alpha_2\right)}{k_i} + \alpha_2 \sum_{\substack{j=1 \\ j \neq i}}^{X-1} \frac{1}{k_j} \right) = m\alpha_1, 1 \leq i \leq X-1, \tag{S7.7b}$$

$$\left(1 + \alpha_1\right) \frac{1}{k_0} + \alpha_1 \sum_{i=1}^{X-1} \frac{1}{k_i} = t_{opt}. \tag{S7.7c}$$

Solution to above system of equations gives

$$k_i = k = k_0 \zeta, 1 \leq i \leq X-1, \tag{S7.8a}$$

$$k_0 = \frac{X\alpha_1 \left(\left(1 + \alpha_1\right)^2 + 2\alpha_2\right) - 3\alpha_1^3 + \alpha_2 + 2\alpha_1^2 + \alpha_1^2\alpha_2 + \alpha_1}{\left(X \left(\alpha_2 - \alpha_1^2\right) + \alpha_1 \left(2 + 3\alpha_1 + \alpha_2\right)\right) t_{opt}}, \tag{S7.8b}$$

$$\zeta = \frac{X(\alpha_2 - \alpha_1^2) + 3\alpha_1^2 + 2\alpha_1 + \alpha_1\alpha_2}{3\alpha_1^2 + \alpha_1 + \alpha_1\alpha_2 - \alpha_2}. \tag{S7.8c}$$

These equations reduce to our previous results of having a constant mean burst size b when $\alpha_1 = 1/b$, $\alpha_2 = 1/b^2$ are used.

Optimal regulation when extrinsic factor affects the transcription rate. We consider an extrinsic factor Z with a positive-valued arbitrary distribution $f_Z(z)$. This factor is assumed to be static, i.e., it does not vary over the time scale of the event of interest. Further we assume that it affects the transcription rates in a multiplicative fashion. The first-passage time mean in this case can be written as

$$\langle T | Z = z \rangle = \frac{1}{z} \left(\frac{1}{k_0} + \frac{1}{b} \sum_{i=0}^{X-1} \frac{1}{k_i} \right) \implies \langle T \rangle = \left\langle \frac{1}{z} \right\rangle \left(\frac{1}{k_0} + \frac{1}{b} \sum_{i=0}^{X-1} \frac{1}{k_i} \right). \tag{S7.9}$$

Likewise the second order moment can be written as

$$\left\langle T^2 \right\rangle = \left\langle \frac{1}{z^2} \right\rangle \frac{2}{b^2} \left(\frac{\tau_0}{bk_0} + \sum_{i=0}^{X-1} \frac{\tau_i}{k_i} \right), \tau_i := \frac{b}{k_i} + \sum_{j=i}^{X-1} \frac{1}{k_j}.$$ [S7.10]

Solving the constrained optimization problem of minimizing $\left\langle T^2 \right\rangle$ constrained to $\langle T \rangle = t_{opt}$ in this case simplifies to solving the following system of equations

$$\frac{\langle 1/z^2 \rangle}{\langle 1/z \rangle} \frac{2}{b} \left(\frac{2(b+1)}{k_0} + \sum_{j=1}^{X-1} \frac{1}{k_j} \right) = m,$$ [S7.11a]

$$\frac{\langle 1/z^2 \rangle}{\langle 1/z \rangle} \frac{2}{b} \left(\frac{b+1}{k_0} + \frac{2(b+1)}{k_i} + \sum_{\substack{j=1 \\ j \neq i}}^{X-1} \frac{1}{k_j} \right) = m, 1 \leq i \leq X-1,$$ [S7.11b]

$$\left\langle \frac{1}{z} \right\rangle \frac{1}{b} \left(\frac{b+1}{k_0} + \sum_{j=1}^{X-1} \frac{1}{k_j} \right) = t_{opt},$$ [S7.11c]

where m represents the Lagrange's multiplier. Solution to these equations gives

$$k_0 = \frac{(b+1)(2b+X)\left\langle \frac{1}{z} \right\rangle}{(2b+1)bt_{opt}}$$ [S7.12a]

$$k_i = \frac{2b+1}{b+1}k_0 = \frac{(2b+X)\left\langle \frac{1}{z} \right\rangle}{bt_{opt}}, 1 \leq i \leq X-1.$$ [S7.12b]

S8. Optimal feedback strategy when burst size is drawn from a Poisson distribution

In the main paper, the mRNA degradation process is assumed to be memoryless (exponential), and consequently the burst of proteins is assumed to follow a geometric distribution. However, in the limit when the mRNA degradation process is deterministic, the burst size distribution becomes Poisson. For this reason, we investigate how the optimal feedback strategy changes in the case when the burst follows a Poisson distribution.
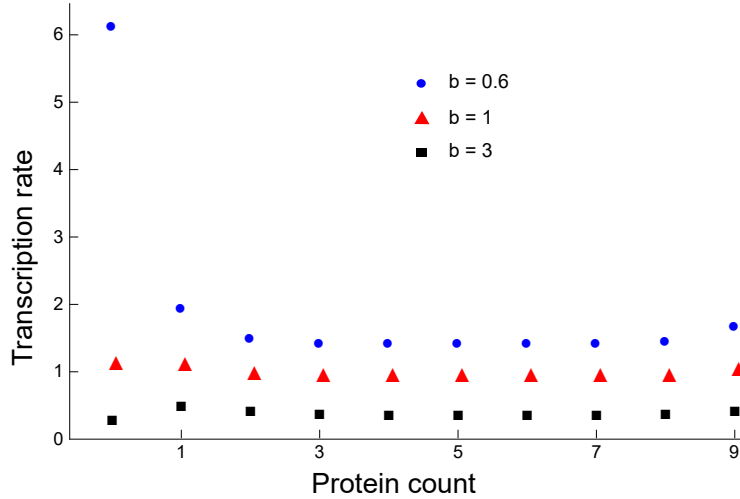
Fig. S9. Optimal feedback strategy for Poisson distributed burst size. The optimal transcription rates obtained via numerical optimization for different values of mean burst size are shown. The event threshold is assumed to be 10 molecules, and the mean FPT is constrained to be 10 minutes.

The probability density function of FPT can be computed by using Eq. (S1.4b). Here the burst size B follows a Poisson distribution given by

$$\mathbb{P}(B = j) = \frac{b^j}{j!}e^{-b}.$$ [S8.1]

Here b, as before, represents the mean burst size, i.e., the average number of protein molecules produced by one mRNA. One can compute the moments of first-passage time in the same manner as done for the geometric burst size case. It turns out that

the first two moments can be compactly written as

$$\langle T \rangle = \sum_{i=0}^{X-1} \frac{\rho_i}{k_i}, \tag{S8.2}$$

$$\langle T^2 \rangle = 2 \sum_{i=0}^{X-1} \frac{\tau_i}{k_i}, \tag{S8.3}$$

where

$$\tau_i = \sum_{j=i}^{X-1} \frac{\rho_{j-i}}{k_j}. \tag{S8.4}$$

The coefficients ρ_i, $i \in \{0, 1, 2, \ldots X-1\}$ are defined as

$$\rho_0 = 1, \quad \rho_i = \frac{e^b}{e^b - 1} \frac{b^i}{i! \, (e^b - 1)^i} \sum_{m=0}^{i-1} a[i, m] e^{mb}, \tag{S8.5}$$

with $a[r, s]$ represents an Eulerian number whose expression is

$$a[r, s] = \sum_{i=0}^{s+1} (-1)^i \binom{r+1}{i} (s + 1 - i)^r. \tag{S8.6}$$

We performed numerical optimization with respect to parameters k_i's for threshold $X = 10$ to see the form of the optimal feedback strategy such that $\langle T^2 \rangle$ is minimized, with constraint $\langle T \rangle = t_{opt} = 10$ minutes. The results show that while the optimal transcription rates are not equal (i.e., no feedback mechanism in strict sense), they are fairly close to each other for mean burst sizes of 1 and 3 molecules. For mean burst size of 0.6 molecules, the first transcription rate when protein count is zero comes out to be significantly higher than other transcription rates which are more or less close to each other (Fig. S9). These results suggest that while the optimal feedback strategy deviates from a no feedback strategy with the underlying distribution of the burst size, it appears to remain close to a no feedback strategy.

S9. Supplemental figures on phage λ lysis

In this section, we provide two figures on phage λ lysis to supplement the discussion in the main text. Fig. S10 shows that the single-cell lysis time data for a wild-type phagel λ shows precision: lysis takes place at about 65 minutes with approximately 5% coefficient of variation. Fig. S11 depicts the holin-antiholin feedforward loop in the lytic pathway of a phage λ. Interestingly for a mutant IN71 which does not have antiholin expression, the lysis time coefficient of variation is higher than the wild-type as shown in Fig. S10.
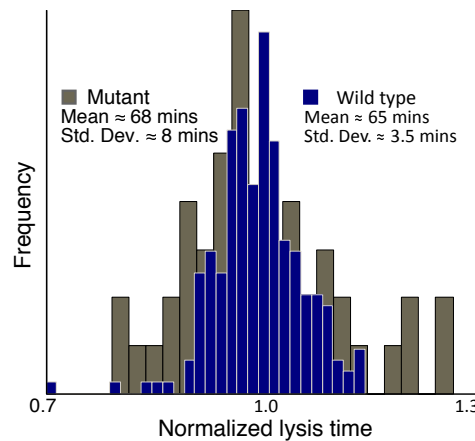
Fig. S10. Single-cell lysis time histograms for bacteriophage λ show precision in timing. Time taken by wild-type λ, and λ mutant IN71, to lyse individual *E. coli* cells infected by a single copy of the virus. For wild-type λ, the coefficient of variation of lysis time is less than 5%. The mutant IN71, does not express antiholin but also has other compensatory mutations to ensure same mean lysis times [4]. The mutant strain has a significantly higher noise in lysis timing as measured by the coefficient of variation (p-value= 0.004 from bootstrapping). Lysis times are normalized by their respective means. See [4] for details on material and methods.

     Khem Raj Ghusinga, John J. Dennehy, Abhyudai Singh 10.1073/pnas.1609012114
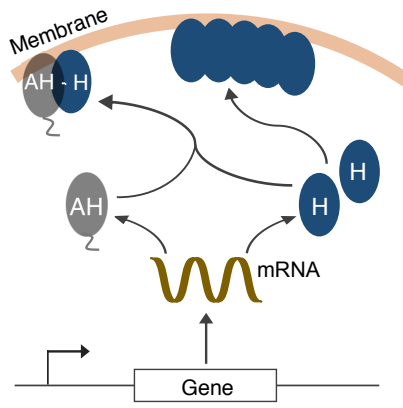
Fig. S11. An incoherent feedforward circuit in phage λ's lytic pathway. *Left:* Bacteriophage λ lysis the infected host cell by expressing a membrane protein, holin (H). The protein slowly accumulates on the cell membrane over time and forms holes when a critical concentration threshold is reached. The mRNA encoding holin also expresses antiholin (AH), which binds to holin and prevents it from participating in hole formation creating a feedforward circuit.

S10. Optimal feedback strategy when the steady-state protein level is fixed

Although achieving precision in timing is an important task, the cost of protein producing may be another factor to consider in deciding upon the optimal feedback strategy. One way to incorporate energy costs, is to find feedback strategies that minimize noise in event timing for a given mean FPT, with the added constraint that they also guarantee a given mean steady-state protein level.

Let x_{ss} denote the steady-state protein level obtained from the ordinary differential equation description of gene expression. Then, constraining x_{ss} to some specific value leads to constraining the ratio $k_{max}b/\gamma$. For instance, for positive feedback

$$k_{\max}\frac{r+(cx_{ss})^H}{1+(cx_{ss})^H}b - \gamma x_{ss} = 0 \qquad [S10.1a]$$

$$\implies \frac{k_{\max}b}{\gamma} = \frac{x_{ss}\left(1+(cx_{ss})^H\right)}{\left(r+(cx_{ss})^H\right)}. \qquad [S10.1b]$$

Similarly, for negative feedback, we obtain

$$\frac{k_{\max}b}{\gamma} = x_{ss}\left(1+(cx_{ss})^H\right). \qquad [S10.2]$$

Note that if we choose fixed values of H, r, b, and γ, then the above constraint on x_{ss} uniquely determines the maximal transcription rate $k_{\max}$. Since the idea is to schedule the event around a given mean FPT, one needs to vary the event threshold X for a given feedback strategy and feedback strength.
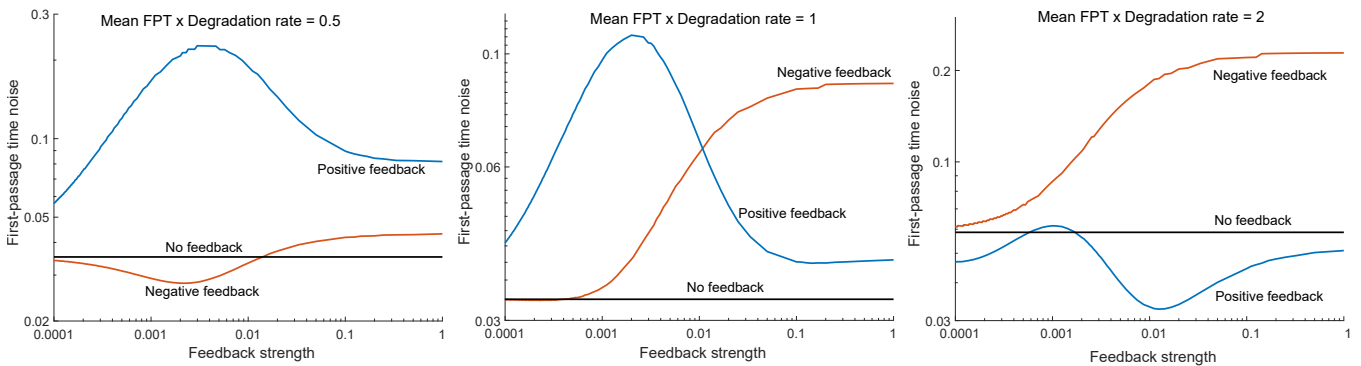
Fig. S12. Optimal feedback strategy when event threshold is varied to schedule event at a given time. The time-scale of event as compared to the protein half-life determines the optimal feedback strategy. *Left:* For events whose time-scale is smaller than protein half-life, a negative feedback provides lowest noise in timing. *Middle:* If time-scale of the event and protein half-life are comparable, both negative and positive feedback are not able to suppress noise for high feedback strengths. For low feedback strengths, a negative feedback gives slightly better attenuation than a no feedback. *Right:* For events whose time-scales are longer than protein half-life, a positive feedback gives minimum noise in timing.

Our preliminary investigation of this related, but different optimization problem, reveals interesting differences with optimal feedback strategies uncovered for the case when the steady-state was not fixed. For example, in the case of a stable protein ($\langle T \rangle \ll 1/\gamma$), a negative feedback provides the best suppression of CV_T^2 for a fixed mean FPT and x_{ss} (Fig S12). Intuitively, for a fixed protein steady-state, the negative feedback has a much faster response time compared to no feedback or positive

feedback [3], and hence has to have a much higher event threshold X to maintain the mean FPT. This leads to lower noise in event timing for negative feedback, as for a stable protein CV_T^2 is inversely proportional to X (Fig. 2 in the main text). Intriguingly, in the case of an unstable protein($\langle T \rangle \gg 1/\gamma$), positive feedback remains the optimal control strategy for buffering fluctuations in event timing, as was the case when the steady-state protein level was not fixed. Hence, depending upon the desired constraints and the parameter regimes, different feedback strategies might be optimal in terms of reducing noise in timing.

It is worth mentioning that there is a range of other possibilities that can realize the constrain on mean FPT. For example, one could vary γ, or b instead of X as done here. Furthermore, a combination of these parameters might also be varied. Clearly, a more systematic study is required to investigate these cases.

References

1. Liao X, Wang L, Yu P (2007) *Stability of Dynamical Systems*. (Elsevier) Vol. 5.
2. Chatterjee G (1974) Negative integral powers of a bidiagonal matrix. *Mathematics of Computation* 28:713–714.
3. Rosenfeld N, Elowitz MB, Alon U (2002) Negative autoregulation speeds the response times of transcription networks. *Journal of molecular biology* 323:785–793.
4. Dennehy JJ, Wang IN (2011) Factors influencing lysis time stochasticity in bacteriophage λ. *BMC Microbiology* 11:174.

Khem Raj Ghusinga, John J. Dennehy, Abhyudai Singh 10.1073/pnas.1609012114