# Derivation of Analytical Expressions

**Moments of Distributions Without Cell Division.** To derive an approximate analytical solution to the model shown in Fig. 1 *Inset*, we at first neglect cell cycle effects. Taking advantage of the constraint on the number of DNA molecules

$$\langle D \rangle + \langle C \rangle = n, \qquad [1]$$

where $n$ is the copy number of the gene of interest on the chromosome, we can write down the master equation for the system as a function of four variables. Let the separate species in Fig. 1, $\{D, C, T, mR, P\}$ be labeled $\{0, 1, 2, 3, 4\}$, respectively. Furthermore, let $p(n_1, n_2, n_3, n_4, t)$ be the probability that there exist, at time $t$, $n_1$ molecules of $C$, $n_2$ molecules of $T$, $n_3$ molecules of mRNA, and $n_4$ protein molecules, then

$$\begin{aligned} \frac{\partial}{\partial t} p(n_1, n_2, n_3, n_4, t) &= f_0 \Big[ (n - n_1 + 1) p(n_1 - 1, n_2, n_3, n_4, t) \\ &\quad - (n - n_1) p(n_1, n_2, n_3, n_4, t) \Big] + \cdots, \end{aligned} \qquad [2]$$

where the dots denote similar terms, one for each rate constant. Rather than write down this long equation, we transform it straight away to an expression for the generating function

$$F(z_1, z_2, z_3, z_4, t) = \sum_{n_1, n_2, n_3, n_4} z_1^{n_1} z_2^{n_2} z_3^{n_3} z_4^{n_4} p(n_1, n_2, n_3, n_4, t), \qquad [3]$$

which can be thought of as a kind of discrete Laplace transform. Defining

$$\begin{aligned} w = z_1 - 1 \quad ; \quad x = z_2 - 1 \\ y = z_3 - 1 \quad ; \quad z = z_4 - 1 \end{aligned} \qquad [4]$$

we arrive at

$$\begin{aligned} \frac{\partial F}{\partial t} &= f_0 n w F - \Big[ f_0 w (1 + w) + b_0 w - k_0 (x - w) \Big] \frac{\partial F}{\partial w} + v_0 (y - x) \frac{\partial F}{\partial x} \\ &\quad + \Big[ v_1' z (1 + y) - d_0' y \Big] \frac{\partial F}{\partial y} - d_1 z \frac{\partial F}{\partial z}. \end{aligned} \qquad [5]$$

Clearly, finding a full solution of Eq. **5** is very difficult. However, from Eq. **3** several properties of the solution are transparent. If all the $z_i$ are set to unity ($w = x = y = z = 0$), then normalization implies that $F =$

1. Differentiating $F$ with respect to a $z_i$, and then setting all $z_i$ to unity, gives $\langle n_i \rangle$, whereas performing the same operation after two derivatives gives $\langle n_i(n_i-1)\rangle$. Because we intend to only calculate the intrinsic noise in protein levels, that is, the variance in $n_4$, an expansion of $F$ around $z_i = 1$ (for small $w$, $x$, $y$, and $z$) should be suitable [this is equivalent to the method of compounding moments (1) and is exact in our case as each moment depends only on moments of lower or equal order].

Because the protein degradation rate is much smaller than all others in Fig. 1 *Inset*,

$$d_1 \ll \{f_0, b_0, k_0, v_0, d_0', v_1'\}, \qquad \textbf{[6]}$$

we assume that all the time dependence in $F$ comes purely from the protein terms. Levels of $C$, $T$, and $mR$ are assumed to be at their steady-state values. In this case, we can write

$$\begin{aligned} F(w,x,t,z,t) \;\simeq\; & 1 + wX_1 + xX_2 + yX_3 + zX_4(t) + \frac{1}{2}\Big[X_{11}w^2 + X_{22}x^2 \\ & + X_{33}y^2 + X_{44}(t)z^2 + 2X_{12}wx + 2X_{13}wy + 2X_{23}xy \\ & + 2X_{14}(t)wz + 2X_{24}(t) + 2X_{34}(t)yz\Big], \end{aligned} \qquad \textbf{[7]}$$

where, for example, $X_3 = \langle mR \rangle$, $X_{11} = \langle C^2 \rangle - \langle C \rangle$, and $X_{34}(t) = \langle mR\ P \rangle$.

Putting Eq. **7** into Eq. **5** and comparing coefficients gives,

$$X_1 = \tfrac{f_0 n}{\ell} \quad ; \quad X_2 = \tfrac{f_0 k_0 n}{v_0 \ell} \quad ; \quad X_3 = \tfrac{f_0 k_0 n}{d_0' \ell} \qquad \textbf{[8]}$$

and

$$\dot{X}_4 = v_1' X_3 - d_1 X_4. \qquad \textbf{[9]}$$

Similarly, it is possible to solve for all the $X_{ij}$. These obey

$$\begin{aligned} d_0' X_{33} &= v_0 X_{23} \\ v_0 X_{22} &= k_0 X_{12} \\ \ell X_{11} &= f_0(n-1)X_1 \\ (\ell + v_0)X_{12} - f_0 n X_2 &= k_0 X_{11} \\ (d_0' + \ell)X_{13} - f_0 n X_3 &= v_0 X_{12} \\ (d_0' + v_0)X_{23} - v_0 X_{22} &= k_0 X_{13}, \end{aligned} \qquad \textbf{[10]}$$

solution of which leads to

$$\eta_{11}^2 = \frac{\langle C^2 \rangle - \langle C \rangle^2}{\langle C \rangle^2} = \frac{1}{\langle C \rangle} - \frac{1}{n} \qquad \textbf{[11]}$$

$$\eta_{22}^2 = \frac{\langle T^2\rangle - \langle T\rangle^2}{\langle T\rangle^2} = \frac{1}{\langle T\rangle} - \frac{v_0}{n(\ell + v_0)} \qquad [12]$$

and

$$\eta_{13}^2 = \frac{\langle C\ mR\rangle - \langle C\rangle\langle mR\rangle}{\langle C\rangle\langle mR\rangle} = -\frac{d_0' v_0}{n(d_0' + \ell)(\ell + v_0)} \qquad [13]$$

$$\eta_{23}^2 = \frac{\langle T\ mR\ \rangle - \langle T\rangle\langle mR\rangle}{\langle T\rangle\langle mR\rangle} = -\frac{d_0' v_0 (d_0' + \ell + v_0)}{n(d_0' + \ell)(\ell + v_0)(d_0' + v_0)}, \qquad [14]$$

as well as the result for the mRNA noise given in the main paper. The cross-correlation functions are negative because of the constraint, **1**, which leads to $\eta_{01}^2 = -1/n$. For example, when $n = 1$, every time $C$ is made, $D$ vanishes and vice versa, giving exact negative correlation. This negative correlation propagates along the chain of different species in Fig. 1 *Inset*, leading to mRNA being made in a pulse-like manner with the number of $C$ molecules increasing, then falling, resulting in a growth in $T$ that falls to produce $mR$.

For the time-dependent cross correlations, we find

$$\begin{aligned}
\dot{X}_{14} &= v_1' X_{13} + f_0 n X_4 - (d_1 + \ell) X_{14} \\
\dot{X}_{24} &= v_1' X_{23} + k_0 X_{14} - (d_1 + v_0) X_{24} \\
\dot{X}_{34} &= v_1' (X_3 + X_{33}) + v_0 X_{24} - (d_0' + d_1) X_{34} \\
\dot{X}_{44} &= 2 v_1' X_{34} - 2 d_1 X_{44}, \qquad [15]
\end{aligned}$$

where the over dots denote differentiation with respect to time. Eq. **9** can be simply integrated (remembering $X_4 = \langle P\rangle$)

$$\langle P(t)\rangle = \frac{v_1' X_3}{d_1}\left(1 - \mathrm{e}^{-d_1 t}\right) + m\,\mathrm{e}^{-d_1 t}, \qquad [16]$$

where $\langle P(0)\rangle = m$. This result is the starting point for the solution of Eq. **15**. In keeping with approximation **6**, we assume sufficient time has passed that the only exponentials that need be considered in the solution (the others are very small) are those in $d_1 t$. In this case, for example,

$$X_{14}(t) = f_0 n \left(\frac{v_1' X_3}{d_1(d_1 + \ell)} + \frac{d_1 m - v_1' X_3}{d_1 \ell}\mathrm{e}^{-d_1 t}\right) + \frac{v_1' X_{13}}{d_1 + \ell}, \qquad [17]$$

with $X_{24}$, $X_{34}$, and $X_{44}$ being given by similar, though more complicated, expressions. Upon simplification and using the definition

$$\lambda = \frac{v_1'}{d_1}\langle mR\rangle = \frac{v_1' f_0 k_0 n}{d_0' d_1 \ell} \qquad [18]$$

the equation for $X_{44}$ gives

$$\begin{aligned}\hat{\sigma}^2_{\rm int}(t) &= \langle P(t)^2\rangle - \langle P(t)\rangle^2 \\ &= \left(1-{\rm e}^{-d_1 t}\right)\left(m\,{\rm e}^{-d_1 t} + \lambda\left[1+\lambda\Omega\left(1+{\rm e}^{-d_1 t}\right)\right]\right), \qquad \textbf{[19]}\end{aligned}$$

with $\Omega$ a measure of the fluctuations in mRNA,

$$\Omega = \frac{d_1}{d_0'+d_1}\left[\eta_{33}^2 + \frac{d_0'}{d_1+v_0}\left(\eta_{23}^2 + \frac{v_0}{d_1+\ell}\eta_{13}^2\right)\right]. \qquad \textbf{[20]}$$

Eq. **19** gives the intrinsic variance in protein number (with all extrinsic variables held fixed) as a function of time given that at $t=0$, $\langle P(0)\rangle = m$ and $\langle P(0)^2\rangle = m^2$. Ideally, $m$ should not be a constant but should be determined by the cell cycle. To facilitate this, let us write down a generating function for just the protein that gives Eq. **16** and Eq. **19** on expansion. Formally, this generating function is

$$Q_m(z,t) = \sum_n q_{n|m}(t) z^n, \qquad \textbf{[21]}$$

where $q_{n|m}(t)$ is the probability of having $n$ proteins at time $t$, given that there were $m$ proteins at time $t=0$. Expanding around $z=1$,

$$Q_m(z,t) \simeq 1 + (z-1)\langle P(t)\rangle + \frac{1}{2}(z-1)^2\Big[\langle P^2(t)\rangle - \langle P(t)\rangle\Big] + \cdots \qquad \textbf{[22]}$$

and so this function is determined, from Eqs. **16** and **19**, up to order $(z-1)^3$.

As

$$\begin{aligned}\langle P(t)\rangle &= \langle P_0(t)\rangle + m\,{\rm e}^{-d_1 t} \\ \langle P^2(t)\rangle &= \langle P_0^2(t)\rangle + {\rm e}^{-d_1 t} m\Big(1+2\langle P_0(t)\rangle\Big) + m(m-1){\rm e}^{-2d_1 t}, \qquad \textbf{[23]}\end{aligned}$$

where the subscript zero denotes evaluation at $m=0$, one can write

$$Q_m(z,t) = Q_0(z,t)\left[1-{\rm e}^{-d_1 t} + z{\rm e}^{-d_1 t}\right]^m, \qquad \textbf{[24]}$$

which also has the desired property, Eq. **22**. This formulation will prove very useful.

In fact, as the gene encoding protein, $P$, is replicated at $t=t_d$, two generating functions need to be considered, $Q_m^{(1)}(z,t)$, which is valid when

the gene copy number is $n$ and $Q_m^{(2)}(z,t)$, which holds after replication when the copy number is $2n$. Defining

$$Y = 1 - \mathrm{e}^{-d_1 t} \qquad [25]$$

then

$$\begin{aligned} Q_m^{(i)}(z,t) &= Q_0^{(i)}(z,t)\Big[Y + z(1-Y)\Big]^m \\ &= \sum_n z^n q_{n|m}^{(i)}(t), \end{aligned} \qquad [26]$$

where $q_{n|m}^{(i)}$ is now the probability of having $n$ proteins at time $t$ given $m$ at $t = 0$ in (copy number) state $i$.

**Including Cell Division.** The number of proteins in the cell will be partly controlled by the cell cycle; dilution due to partition into daughter cells at the end of cell division can play a significant role in keeping protein numbers low. To incorporate this effect into our analysis, let $P_i(n)$ be the probability of finding $n$ proteins at the start of the $i$th division cycle. Then $P_{i+1}(n)$ is related to $P_i(n)$ via a transfer probability $U(n|n')$,

$$P_{i+1}(n) = \sum_{n'} U(n|n') P_i(n'). \qquad [27]$$

In our calculation, just one daughter cell is followed, and we assume that each protein has a 50% probability of being kept in this cell (and so a 50% chance of being discarded into the one not followed). Given $m$ proteins just before cell division, the probability of having $n$ immediately after is just binomial

$$\binom{m}{n} 2^{-m}. \qquad [28]$$

For a cell cycle of length $T$, the transfer probability $U$ is given by

$$U(n|n') = \sum_{m,m'} \binom{m}{n} 2^{-m} q_{m|m'}^{(2)}(T - t_d) q_{m'|n'}^{(1)}(t_d), \qquad [29]$$

where gene replication at time $t_d$ is included, and the definitions of the $q^{(i)}$ have been used (see end of previous section).

After many divisions, the protein number, rather than tending to a steady-state, tends to a limit cycle. Mathematically, as the limit cycle is approached, $P_i(n)$ is expected to tend to $P^*(n)$, which obeys (see Eq. **27**),

$$P^*(n) = \sum_{n'} U(n|n')P^*(n'). \qquad [30]$$

To solve Eq. **30** for $P^*$, we again turn to generating functions. Defining

$$F^*(z) = \sum_{n=0}^{\infty} z^n P^*(n) \qquad [31]$$

multiplying Eq. **30** by $z^n$ and summing over all $n$, gives

$$\begin{aligned} F^*(z) &= \sum_{m,m',n'} \sum_{n=0}^{m} z^n \binom{m}{n} 2^{-m} q^{(2)}_{m|m'}(T-t_d) q^{(1)}_{m'|n'}(t_d) P^*(n') \\ &= \sum_{m,m',n'} \left(\frac{1+z}{2}\right)^m q^{(2)}_{m|m'}(T-t_d) q^{(1)}_{m'|n'}(t_d) P^*(n'), \end{aligned} \qquad [32]$$

where Eq. **29** has been used. From definition **26**, this can be written as

$$\begin{aligned} F^*(z) &= \sum_{m',n'} Q^{(2)}_0\Big((1+z)/2, T-t_d\Big) \Big[Y_1 + \frac{1}{2}(1+z)(1-Y_1)\Big]^{m'} \\ &\quad \times q^{(1)}_{m'|n'}(t_d) P^*(n'), \end{aligned} \qquad [33]$$

with

$$Y_1 = 1 - \mathrm{e}^{-d_1(T-t_d)}. \qquad [34]$$

The power of writing $Q^{(i)}_m(z,t)$ in the form **26** should now be apparent; it also allows the sum over $m'$ to be evaluated similarly,

$$\begin{aligned} F^*(z) &= Q^{(2)}_0\Big((1+z)/2, T-t_d\Big) Q^{(1)}_0\Big(Y_1 + (1+z)(1-Y_1)/2, t_d\Big) \\ &\quad \times \sum_{n'} \Big[Y_2 + (Y_1 + (1+z)(1-Y_1)/2)(1-Y_2)\Big]^{n'} P^*(n'), \end{aligned} \qquad [35]$$

where $Y_2$ is

$$Y_2 = 1 - \mathrm{e}^{-d_1 t_d}. \qquad [36]$$

Finally, Eq. **31** allows the last summation to be carried out

$$\begin{aligned} F^*(z) &= Q^{(2)}_0\Big((1+z)/2, T-t_d\Big) Q^{(1)}_0\Big(Y_1 + (1+z)(1-Y_1)/2, t_d\Big) \\ &\quad \times F^*\Big(Y_2 + (Y_1 + (1+z)(1-Y_1)/2)(1-Y_2)\Big). \end{aligned} \qquad [37]$$

The solution of Eq. **37** will give the generating function related to $P^*(n)$, the probability of finding $n$ proteins at the beginning of the cell cycle given that the bacterium has divided enough times to have reached a limit cycle state. Because we are only interested in calculating the variance in protein number, only the first two moments of $F^*(z)$ are required. Writing

$$F^*(z) = 1 + (z-1)f_1^* + \frac{1}{2}(z-1)^2 f_2^* + \cdots \qquad [38]$$

and then comparing coefficients of $z-1$ in Eq. **37** allows $f_1^*$ and $f_2^*$ to be determined: for example,

$$f_1^* = \frac{2\langle P_0(T-t_d)\rangle + (1-Y_1)\langle P_0(t_d)\rangle}{1+Y_1+Y_2-Y_1Y_2}, \qquad [39]$$

with a similar expression for $f_2^*$.

Because we now know the probability distribution $P^*(n)$, we can write the generating function for the protein number during the two stages of the cell cycle. Given that the protein number has reached a limit cycle state, and defining $t=0$ to be at the beginning of this cycle, i.e., immediately after cell division, then for

**$0 \le t \le t_d$:**

$$\begin{aligned} F^{(1)}(z,t) &= \sum_{n,m} z^n q^{(1)}_{n|m}(t) P^*(m) \\ &= \sum_m Q_0^{(1)}(z,t)\Big[Y + z(1-Y)\Big]^m P^*(m) \\ &= Q_0^{(1)}(z,t) F^*\Big(Y + z(1-Y)\Big), \end{aligned} \qquad [40]$$

with the summations evaluated by using Eq. **26** and Eq. **31** again.

**$t_d \le t \le T$:** In this case, because of gene replication, the expression is a little more complicated. Defining

$$Y' = 1 - \mathrm{e}^{-d_1(t-t_d)}, \qquad [41]$$

one has

$$F^{(2)}(z,t) = \sum_{n,m,m'} z^n q^{(2)}_{n|m}(t-t_d) q^{(1)}_{m|m'}(t_d) P^*(m')$$

$$
\begin{aligned}
&= \sum_{m,m'} Q_0^{(2)}(z, t-t_d)\Big[Y' + z(1-Y')\Big]^m q^{(1)}_{m|m'}(t_d) P^*(m') \\
&= Q_0^{(2)}(z, t-t_d) Q_0^{(1)}\Big(Y' + z(1-Y'), t_d\Big) \\
&\quad \times \sum_{m'} \Big[Y_2 + \Big(Y' + z(1-Y')\Big)(1-Y_2)\Big]^{m'} P^*(m') \\
&= Q_0^{(2)}(z, t-t_d) Q_0^{(1)}\Big(Y' + z(1-Y'), t_d\Big) \\
&\quad \times F^*\Big(Y_2 + \Big(Y' + z(1-Y')\Big)(1-Y_2)\Big). \qquad \textbf{[42]}
\end{aligned}
$$

Differentiation of these two generating functions with respect to $z$ will give the mean and variance of the intrinsic protein number distribution. The protein mean is given in the main paper, and the noise satisfies

$$\hat{\eta}^2_{\rm int}(t) = \frac{1}{\langle P(t)\rangle} + \Omega\Phi_1(t), \qquad \textbf{[43]}$$

with $\Phi_1(t)$ given in the main paper and $\Omega$ by Eq. **20.** Using expressions **13**, **14** and that for the mRNA noise, Eq. **20** simplifies to

$$\Omega \simeq \frac{d_1}{d_0'}\left(1 - \frac{f_0 k_0}{\ell^2}\right) \cdot \frac{1}{\langle mR\rangle} \qquad \textbf{[44]}$$

in the limit of $d_1/d_0' \ll 1$. Eqs. **43** and **44** comprise the expression for the intrinsic protein noise, $\hat{\eta}_{\rm int}$, given in the main paper.

# Parameters Used in Simulations

All parameter values are given in Table 3.

| Process | Parameters |
|---|---|
| RNAP binding to DNA | Free RNAP concentration = 30 nM (2)<br>Binding rate $1.4 \times 10^7$ $M^{-1}s^{-1}$ for $\lambda P_L$ (3)<br>$\Rightarrow f_0 = 0.42\ s^{-1}$<br><br>$b_0 = 0.1\ s^{-1}$ [chosen to give an equilibrium constant of $1.4 \times 10^8$ $M^{-1}$ (2)] |
| Transcription initiation rate | $k_0$; ranges from $0.001\ s^{-1}$ to $0.1\ s^{-1}$ (4)<br>(closed to open complex isomerization) |
| Formation and degradation of RBS on mRNA | $v_0 = 0.3\ s^{-1}$ [RNAP moving at 40 nt $s^{-1}$ (5)]<br><br>$mf_0 = 0.114\ s^{-1}$ (chosen so that the average number of proteins per transcript = 15)<br>$d_0 = 0.1\ s^{-1}$ |
| Binding of ribosome | Free ribosome concentration = 400 nM<br>(order of magnitude larger than RNAP)<br>binding rate $1 \times 10^7$ $M^{-1}s^{-1}$<br>$\Rightarrow\ mf_1 = 4.0\ s^{-1}$<br><br>$mb_1 = 0.4\ s^{-1}$ [chosen to given an equilibrium constant of $2.5 \times 10^7$ $M^{-1}$ (6)] |
| Translation | $k_1 = 0.3\ s^{-1}$ (6)<br>$v_1 = 0.048\ s^{-1}$ [given a 1000 nt protein and a translation rate of 48 nt $s^{-1}$ (7)] |
| Protein degradation | $d_1 = 6.42 \times 10^{-5}\ s^{-1}$ ($t_{\frac{1}{2}} \simeq 3$ hours) |
| Cell cycle time<br><br><br>Gene replication time | $T = 60$ min (chosen for at most two chromosomes per cell)<br><br>$t_d = 0.4\,T$ |
| Cell volume and growth | Linear growth (8)<br>$V(t) = V_0(1 + t/T)$ for $0 \le t \le T$<br>and $V_0 = 2.5 \times 10^{-15}\ell$ |

Table 3. Parameters suitable for constitutive gene expression in *Escherichia coli*. Abbreviations: RNA polymerase (RNAP), ribosome binding site (RBS), nucleotide (nt).

# References


1. Van Kampen, N. G. (1990) *Stochastic Processes in Physics and Chemistry* (Elsevier, New York).

2. McClure, W. R. (1983) in *Biochemistry of Metabolic Processes*, eds. Lennon, D. L. F., Stratman, F. W. & Zahlten, R. N. (Elsevier, New York).

3. Lanzer, M. & Bujard, H. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 8973–8977.

4. Hawley, D. K. & McClure, W. R. (1982) *J. Mol. Biol.* **157**, 493–525.

5. Manor, H., Goodman, D. & Stent, G. S. (1969) *J. Mol. Biol.* **39**, 1–29.

6. Draper, D. E. (1996) in *Escherichia coli and Salmonella : Cellular and Molecular Biology*, ed. Neidhardt, F. C. (Am. Soc. Microbiol., Washington, D.C.).

7. Bremer, H. & Dennis, P. P. (1996) in *Escherichia coli and Salmonella : Cellular and Molecular Biology*, ed. Neidhardt, F. C. (Am. Soc. Microbiol., Washington, D.C.).

8. Kubitschek, H. E. (1990) *J. Bacteriol.* **172**, 94–101.