Translational Crosstalk in Gene Networks: Supporting Material

William H. Mather[1], Jeff Hasty[2,3,4,5], Lev S. Tsimring[2,5], and Ruth J. Williams[2,6,7]

[1] *Department of Physics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061 USA*

[2] *Biocircuits Institute, University of California, San Diego, La Jolla CA 92093 USA*

[3] *Department of Bioengineering, University of California, San Diego, La Jolla CA 92093 USA*

[4] *Molecular Biology Section, Division of Biological Sciences, University of California, San Diego, La Jolla CA 92093 USA*

[5] *San Diego Center for Systems Biology, 9500 Gilman Drive, La Jolla CA 92093 USA*

[6] *Department of Mathematics, University of California, San Diego, 9500 Gilman Drive, La Jolla CA 92093 USA*

[7] *Corresponding Author. Email: williams@stochastic.ucsd.edu*

A Computing moments for protein counts when Markov chain transitions drive protein production

In this section we present a general method for computing time-dependent and steady state moments for multi-type protein distributions where the protein production events occur at transitions of a finite state, continuous time Markov chain. This setup allows for quite general upstream production mechanisms beyond those considered in our main text, such as those involving mRNA transcription, mRNA splicing, multiple ribosomes binding to mRNA, and so forth. Furthermore, this setup could also be used for studying similar production phenomena such as mRNA production. The primary limitation of our methodology is that the upstream process governing protein production events is represented by a continuous time Markov chain with finitely many states.

We suppose that there are Q total species of protein that have respective counts Y_i for $i = 1, \ldots, Q$. Proteins of type i degrade with degradation rate constant $\gamma_i > 0$ (degradation propensity $\gamma_i Y_i$), independently of the counts of other proteins and the state of upstream processes. Protein production reactions are determined by transitions in an upstream continuous time Markov chain X that has a total of $N < \infty$ states; for concreteness we denote these states by $1, \ldots, N$ here, though the specific labels are not important for our analysis. The Markov chain can transition from a state n to state m with rate α_{nm}, making no other change. Additionally, the Markov chain can transition from a state n to state m (m may be equal to n) with rate β_{inm} while simultaneously producing a single protein of type i. In fact, $(X, Y_1, \ldots, Y_Q)$ is also a continuous time Markov chain but with countably many states.

We let $P_m(\vec{Y} = \vec{y}, t)$ denote the joint probability to find, at time t, the Markov chain X in state m and the protein populations Y_i to have respective counts y_i. (We suppress the initial conditions in our notation.) We will primarily be concerned with the derivation of time-dependent and steady state values of averages over these distributions, especially those yielding protein moments. We adopt the notation for averages

$$\langle Z(X,Y) \rangle \equiv \sum_{m,\vec{y}} P_m(\vec{Y} = \vec{y}, t)\, Z(m, \vec{y}) \tag{1}$$

where the sum $\sum_{m,\vec{y}}$ is over all possible states m for X and $\vec{y}$ for $\vec{Y}$, and $Z(m, \vec{y})$ is either a non-negative function of $m, \vec{y}$ or a signed function such that the sum on the right with $|Z|$ in place of Z is finite. We do not explicitly indicate t in this notation as the meaning will usually be clear from the context and this permits us to use the same notation when considering steady state averages.

For a fixed state m of the Markov chain, we define the following partial moments for the protein counts; the sums used in computing these averages are restricted to where the state of X is m:

$$M_{m,ij\cdots k} \equiv \left\langle Y_i Y_j \cdots Y_k \; 1_{\{X=m\}} \right\rangle \equiv \sum_{y_1=0}^{\infty} \cdots \sum_{y_Q=0}^{\infty} P_m(\vec{Y} = \vec{y}, t)\; y_i y_j \cdots y_k. \tag{2}$$

Here for an event A, 1_A takes the value one on A and zero elsewhere. As a matter of notation, the zeroth order partial moment is written as M_m, which is simply the probability for the Markov chain X to be in the state m. The full moments can be obtained from these partial moments by summing over the states of X:

$$M_{ij\cdots k} \equiv \langle Y_i Y_j \cdots Y_k \rangle \;\; = \;\; \sum_{m=1}^{N} M_{m,ij\cdots k}. \tag{3}$$

The partial moments can be obtained from the partial generating function

$$G_m = \sum_{y_1=0}^{\infty} \cdots \sum_{y_Q=0}^{\infty} P_m(\vec{Y} = \vec{y}, t)\, e^{\sum_{\ell=1}^{Q} s_\ell y_\ell} \tag{4}$$

in the usual manner

$$M_{m,ij\cdots k} \equiv \lim_{s_\ell \to 0, \forall\, \ell} \frac{\partial}{\partial s_i}\frac{\partial}{\partial s_j} \cdots \frac{\partial}{\partial s_k} G_m. \tag{5}$$

(The degradation of the proteins and their limited production rate ensure that G_m is finite for all $\vec{s}$.)

Analysis of the chemical master equation defining the full stochastic system reveals through straightforward algebra that the partial generating functions obey the following system of first order linear partial differential equations (PDEs)

$$\begin{aligned}\frac{\partial G_m}{\partial t} &= \left(\sum_{i=1}^{Q} \gamma_i \left(e^{-s_i} - 1\right) \frac{\partial}{\partial s_i}\right) G_m + \sum_{n=1}^{N} \left(\alpha_{nm}\, G_n - \alpha_{mn}\, G_m\right) \\ &\quad + \sum_{i=1}^{Q}\sum_{n=1}^{N} \left(e^{s_i}\, \beta_{inm}\, G_n - \beta_{imn}\, G_m\right).\end{aligned} \tag{6}$$

The dynamics of all partial moments can be found by combining this with Eq. 5 (or by using the chemical master equation directly). In particular, the zeroth order expressions satisfy

$$\frac{\partial M_m}{\partial t} = \sum_{n=1}^{N} \left(\alpha_{nm}\, M_n - \alpha_{mn}\, M_m\right) + \sum_{k=1}^{Q}\sum_{n=1}^{N} \left(\beta_{knm}\, M_n - \beta_{kmn}\, M_m\right) \tag{7}$$

which simply reproduce the Markov chain dynamics. First order expressions have a similar structure

$$\begin{aligned}\frac{\partial M_{m,i}}{\partial t} &= \sum_{n=1}^{N} \left(\alpha_{nm}\, M_{n,i} - \alpha_{mn}\, M_{m,i}\right) + \sum_{k=1}^{Q}\sum_{n=1}^{N} \left(\beta_{knm}\, M_{n,i} - \beta_{kmn}\, M_{m,i}\right) \\ &\quad + \sum_{n=1}^{N} \beta_{inm}\, M_n - \gamma_i M_{m,i}.\end{aligned} \tag{8}$$

Second order expressions satisfy

$$\begin{aligned}\frac{\partial M_{m,ij}}{\partial t} &= \sum_{n=1}^{N} \left(\alpha_{nm}\, M_{n,ij} - \alpha_{mn}\, M_{m,ij}\right) + \sum_{k=1}^{Q}\sum_{n=1}^{N} \left(\beta_{knm}\, M_{n,ij} - \beta_{kmn}\, M_{m,ij}\right) \\ &\quad + \sum_{n=1}^{N} \beta_{jnm}\, M_{n,i} + \sum_{n=1}^{N} \beta_{inm}\, M_{n,j} - \gamma_i M_{m,ij} - \gamma_j M_{m,ij}\end{aligned} \tag{9}$$

when $i \neq j$, and

$$\begin{aligned}\frac{\partial M_{m,ii}}{\partial t} &= \sum_{n=1}^{N} \left(\alpha_{nm}\, M_{n,ii} - \alpha_{mn}\, M_{m,ii}\right) + \sum_{k=1}^{Q}\sum_{n=1}^{N} \left(\beta_{knm}\, M_{n,ii} - \beta_{kmn}\, M_{m,ii}\right) \\ &\quad + \sum_{n=1}^{N} \beta_{inm}\, M_n + \sum_{n=1}^{N} 2\beta_{inm}\, M_{n,i} + \gamma_i M_{m,i} - 2\gamma_i M_{m,ii}\end{aligned} \tag{10}$$

otherwise.

Equations 7–10 provide a closed system of linear first order ordinary differential equations (ODEs). Note that these can be solved in a hierarchical manner starting from the zeroth order and progressing to higher orders. For example, the solution to Eq. 7 can be inserted into Eq. 8 as an inhomogeneous term. The time-independent versions of these equations are systems of linear equations (including eigenfunction problems) that can be used in computing steady state moments when suitable conditions are imposed on the Markov chain to ensure existence of a unique steady state distribution. While these linear equations and ODEs can in principle be solved for specific parameter values using standard numerical analysis methods, the feasibility of this will depend on the dimension and complexity of the underlying Markov chain and the extent of the parameter space to be explored. For the stochastic model introduced in Section 2 of the main text, we do not need to resort to such numerical methods as a further simplification is available which enables us to obtain closed form algebraic expressions for the desired moments. This is the subject of the next section.

B Formulas for protein moments with fixed mRNA and ribosome numbers in the overloaded regime ($T_1 + T_2 > R$)

We now specialize the content in Section A to our stochastic model for translational dynamics introduced in Section 2 of the main text. Recall that it is assumed in Section 2 that the number of ribosomes R and the numbers of mRNAs of type 1 (T_1) and of type 2 (T_2) are all fixed. We focus on the regime where the ribosomes are overloaded ($T_1 + T_2 > R$), since, as indicated in Section 3 of the main text, the underloaded regime ($T_1 + T_2 \leq R$) can be analyzed directly as a system where each mRNA is bound to one ribosome, and proteins of the same type as the mRNA are produced from it, independently of the production from the other mRNAs.

As described in Section 3 of the main text, in the overloaded regime, under our instantaneous binding assumption ($\nu = \infty$), the stochastic process X that tracks the number of ribosomes currently bound to mRNAs of type 1 is a continuous time Markov chain X with $N = (\min(R, T_1) - \max(R - T_2, 0)) + 1$ states. (All remaining $R - X$ ribosomes will be bound to mRNAs of type 2.) This fits the framework of Section A with the transition rates β_{inm} given for $n, m \in \{\max(R - T_2, 0), \ldots, \min(R, T_1)\}$ by

$$\beta_{1nm} = n\mu \left[\delta_{m,n}\, p + \delta_{m,n}\,(1-p) \left(\frac{T_1 - n + 1}{T_1 + T_2 - R + 1} \right) + \delta_{m+1,n}\,(1-p) \left(\frac{T_2 - R + n}{T_1 + T_2 - R + 1} \right) \right] \quad (11)$$

$$\beta_{2nm} = (R-n)\mu \left[\delta_{m,n}\, p + \delta_{m,n}\,(1-p) \left(\frac{T_2 - R + n + 1}{T_1 + T_2 - R + 1} \right) + \delta_{m-1,n}\,(1-p) \left(\frac{T_1 - n}{T_1 + T_2 - R + 1} \right) \right] \quad (12)$$

where $\delta_{m,n} = 1$ if $m = n$ and zero otherwise. When T_1, T_2 are both of size R or greater, the Markov chain X can take any value in $\{0, 1, \ldots, R\}$; otherwise it is more constrained due to limits on the total numbers of mRNAs of each type. Here the states of X are labelled $\max(R - T_2, 0), \ldots, \min(R, T_1)$, rather than $1, \ldots, N$, as in the previous section. This minor notational change does not affect the applicability of those results. The components of the α-matrix for X are all zero as there are no transitions for the Markov chain without the production of a protein.

B.1 ODE's for moments

We seek to find expressions for the means, variances, and covariance of the protein counts for our model. To this end, we seek to fully solve for the moments

$$\begin{aligned}
\langle 1 \rangle &= \sum_m M_m && (13)\\
\langle X \rangle &= \sum_m m M_m && (14)\\
\langle X^2 \rangle &= \sum_m m^2 M_m && (15)\\
\langle Y_i \rangle &= \sum_m M_{m,i} && (16)\\
\langle X Y_i \rangle &= \sum_m m M_{m,i} && (17)\\
\langle Y_i Y_j \rangle &= \sum_m M_{m,ij}. && (18)
\end{aligned}$$

The dynamics of these moments can be generally described by summing appropriate powers of m over Eqs. 7–10 and then simplifying. However, the resulting equations do not close for general models, i.e. the differential equations for moments of a given order depend on the moments of higher order. The equations for our model fortunately do close, as we now show.

Solution of our problem is a somewhat tedious but straightforward endeavor that is assisted by use of an algebraic manipulation package such as Maple 16 (Waterloo Maple Inc.). For convenience, we define

$$\begin{aligned}
T_1 &= T_s + T_a && (19)\\
T_2 &= T_s - T_a && (20)
\end{aligned}$$

which is equivalent to

$$\begin{aligned}
2T_s &= T_1 + T_2 && (21)\\
2T_a &= T_1 - T_2. && (22)
\end{aligned}$$

We require certain summations over the components of β, which simplify to

$$\begin{aligned}
\zeta_{w,1n} &\equiv \sum_m m^w \beta_{1nm} && (23)\\
&= n\mu \left[n^w\, p + n^w\,(1-p)\left(\frac{T_1 - n + 1}{2T_s - R + 1}\right) + (n-1)^w\,(1-p)\left(\frac{T_2 - R + n}{2T_s - R + 1}\right)\right]\\
\zeta_{w,2n} &\equiv \sum_m m^w\, \beta_{2nm} && (24)\\
&= (R-n)\mu \left[n^w\, p + n^w\,(1-p)\left(\frac{T_2 - R + n + 1}{2T_s - R + 1}\right) + (n+1)^w\,(1-p)\left(\frac{T_1 - n}{2T_s - R + 1}\right)\right]
\end{aligned}$$

where $w \in \{0, 1, 2, \ldots\}$. One such situation where these sums arise is in expressions like

$$\sum_k \sum_{m,n} m^w \left(\beta_{knm} V_n - \beta_{kmn} V_m\right) = \sum_n V_n \left(\sum_k \sum_m m^w \beta_{knm} - n^w \sum_k \sum_m \beta_{knm}\right) \quad (25)$$

$$= \sum_n H_{wn} V_n \quad (26)$$

where for our model H_{wn} is a polynomial in n defined as

$$H_{wn} \equiv \sum_k \sum_m m^w \beta_{knm} - n^w \sum_k \sum_m \beta_{knm} = (\zeta_{w,1n} + \zeta_{w,2n}) - n^w (\zeta_{0,1n} + \zeta_{0,2n}) \quad (27)$$

As will become evident, whether our equations close depends critically on whether the order of the polynomials defined by H_{wn} are not too high. This would not have been the case if the translation rate per ribosome was different for the two types of mRNA.

Some explicit expressions for H_{wn} are

$$H_{0n} = 0 \quad (28)$$

$$H_{1n} = \left(\frac{\mu R (T_s + T_a)(1-p)}{2T_s - R + 1}\right) - \left(\frac{2\mu T_s (1-p)}{2T_s - R + 1}\right) n \quad (29)$$

$$H_{2n} = \left(\frac{\mu R (T_s + T_a)(1-p)}{2T_s - R + 1}\right) + \left(\frac{2\mu (1-p)(R(T_s - 1) + (R-1)T_a)}{2T_s - R + 1}\right) n - \left(\frac{2\mu (1-p)(2T_s - 1)}{2T_s - R + 1}\right) n^2 \quad (30)$$

which are sufficient for the calculation of variances and covariances of protein counts.

B.1.1 Zeroth order in protein moments

Lowest order expressions are generated by summing Eq. 7 over various powers of m. The simplest of these is a statement of conservation of probability

$$\frac{\partial}{\partial t} \langle 1 \rangle = \frac{\partial}{\partial t} \sum_m M_m = \sum_n H_{0n} M_n = 0. \quad (31)$$

The next order in m leads to an ODE for the mean ribosome occupancy by type 1 mRNA

$$\frac{\partial}{\partial t} \langle X \rangle = \frac{\partial}{\partial t} \sum_m m\, M_m = \sum_n H_{1n} M_n$$

$$= \left(\frac{\mu R (T_s + T_a)(1-p)}{2T_s - R + 1}\right) - \left(\frac{2\mu T_s (1-p)}{2T_s - R + 1}\right) \langle X \rangle . \quad (32)$$

The next order in m is the average square ribosome occupancy by type 1 mRNA

$$\frac{\partial}{\partial t} \langle X^2 \rangle = \frac{\partial}{\partial t} \sum_m m^2\, M_m = \sum_n H_{2n} M_n \quad (33)$$

$$= \left(\frac{\mu R (T_s + T_a)(1-p)}{2T_s - R + 1}\right) + \left(\frac{2\mu (1-p)(R(T_s - 1) + (R-1)T_a)}{2T_s - R + 1}\right) \langle X \rangle - \left(\frac{2\mu (1-p)(2T_s - 1)}{2T_s - R + 1}\right) \langle X^2 \rangle . \quad (34)$$

As promised, these equations close onto themselves due to the order of the polynomials H_{wn} in n.

B.1.2 First order in protein moments

The next order of protein moments proceeds similarly. These results arise from summing Eq. 8 over various powers of m. The lowest order in m provides

$$\begin{aligned}
\frac{\partial}{\partial t}\langle Y_i\rangle &= \frac{\partial}{\partial t}\sum_m M_{m,i} && (35)\\
&= \sum_k\sum_{m,n}\left(\beta_{knm}\,M_{n,i} - \beta_{kmn}\,M_{m,i}\right) + \sum_{m,n}\beta_{inm}\,M_n - \gamma_i\sum_m M_{m,i} && (36)\\
&= \sum_n H_{0n}M_{n,i} + \sum_n \zeta_{0,in}\,M_n - \gamma_i\langle Y_i\rangle && (37)\\
&= \sum_n \zeta_{0,in}\,M_n - \gamma_i\langle Y_i\rangle && (38)
\end{aligned}$$

which leads to

$$\begin{aligned}
\frac{\partial}{\partial t}\langle Y_1\rangle &= \mu\langle X\rangle - \gamma_1\langle Y_1\rangle && (39)\\
\frac{\partial}{\partial t}\langle Y_2\rangle &= \mu(R - \langle X\rangle) - \gamma_2\langle Y_2\rangle\,. && (40)
\end{aligned}$$

The next order in m provides

$$\begin{aligned}
\frac{\partial}{\partial t}\langle XY_i\rangle &= \frac{\partial}{\partial t}\sum_m mM_{m,i} && (41)\\
&= \sum_n H_{1n}M_{n,i} + \sum_n \zeta_{1,in}\,M_n - \gamma_i\langle XY_i\rangle && (42)\\
&= \left(\frac{\mu R\,(T_s+T_a)(1-p)}{2T_s - R + 1}\right)\langle Y_i\rangle - \left(\frac{2\mu T_s(1-p)}{2T_s-R+1}\right)\langle XY_i\rangle\\
&\quad + \sum_n \zeta_{1,in}\,M_n - \gamma_i\langle XY_i\rangle\,. && (43)
\end{aligned}$$

Thus

$$\begin{aligned}
\frac{\partial}{\partial t}\langle XY_1\rangle &= \left(\frac{\mu R\,(T_s+T_a)(1-p)}{2T_s - R + 1}\right)\langle Y_1\rangle - \left(\frac{2\mu T_s(1-p)}{2T_s-R+1}\right)\langle XY_1\rangle\\
&\quad - \left(\frac{\mu(1-p)(T_s-T_a-R)}{2T_s-R+1}\right)\langle X\rangle + \left(\frac{\mu(2T_s-R+p)}{2T_s-R+1}\right)\langle X^2\rangle\\
&\quad -\gamma_1\langle XY_1\rangle && (44)
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial}{\partial t}\langle XY_2\rangle &= \left(\frac{\mu R\,(T_s+T_a)(1-p)}{2T_s - R + 1}\right)\langle Y_2\rangle - \left(\frac{2\mu T_s(1-p)}{2T_s-R+1}\right)\langle XY_2\rangle\\
&\quad + \left(\frac{\mu R(1-p)(T_s+T_a)}{2T_s-R+1}\right) - \left(\frac{\mu((1-p)\,(T_s+T_a) - pR - 2RT_s + R^2)}{2T_s-R+1}\right)\langle X\rangle\\
&\quad - \left(\frac{\mu(2T_s-R+p)}{2T_s-R+1}\right)\langle X^2\rangle - \gamma_2\langle XY_2\rangle\,. && (45)
\end{aligned}$$

The reader can check that these ODE's can be solved once those in Section B.1.1 are solved.

B.1.3 Second order in protein moments

Summation in m over Eq. 9 provides for $i \neq j$

$$\begin{aligned} \frac{\partial}{\partial t}\langle Y_i Y_j\rangle &= \sum_{m,n}\beta_{jnm}\,M_{n,i} + \sum_{m,n}\beta_{inm}\,M_{n,j} - (\gamma_i+\gamma_j)\langle Y_iY_j\rangle & (46)\\ &= \sum_{n}\zeta_{0jn}\,M_{n,i} + \sum_{n}\zeta_{0in}\,M_{n,j} - (\gamma_i+\gamma_j)\langle Y_iY_j\rangle\,. & (47)\end{aligned}$$

Thus

$$\frac{\partial}{\partial t}\langle Y_1Y_2\rangle = \mu\left(R\langle Y_1\rangle - \langle XY_1\rangle\right) + \mu\langle XY_2\rangle - (\gamma_1+\gamma_2)\langle Y_1Y_2\rangle\,. \quad (48)$$

Summation in m over Eq. 10 provides

$$\frac{\partial}{\partial t}\left\langle Y_i^2\right\rangle = \sum_{n}\zeta_{0,in}\,M_n + \sum_{n}2\zeta_{0,in}\,M_{n,i} + \gamma_i\langle Y_i\rangle - 2\gamma_i\left\langle Y_i^2\right\rangle. \quad (49)$$

Thus

$$\begin{aligned} \frac{\partial}{\partial t}\left\langle Y_1^2\right\rangle &= \mu\langle X\rangle + 2\mu\langle XY_1\rangle + \gamma_1\langle Y_1\rangle - 2\gamma_1\left\langle Y_1^2\right\rangle & (50)\\ \frac{\partial}{\partial t}\left\langle Y_2^2\right\rangle &= \mu(R-\langle X\rangle) + 2\mu(R\langle Y_2\rangle - \langle XY_2\rangle) + \gamma_2\langle Y_2\rangle - 2\gamma_2\left\langle Y_2^2\right\rangle. & (51)\end{aligned}$$

The reader can check that these ODE's can be solved once those in Sections B.1.1–B.1.2 are solved.

With the variances and covariances defined by

$$\begin{aligned} \sigma^2_{Y_i} &= \left\langle Y_i^2\right\rangle - (\langle Y_i\rangle)^2 & (52)\\ \sigma^2_{Y_1,Y_2} &= \langle Y_1Y_2\rangle - \langle Y_1\rangle\langle Y_2\rangle & (53)\\ \sigma^2_{X,Y_i} &= \langle XY_i\rangle - \langle X\rangle\langle Y_i\rangle\,, & (54)\end{aligned}$$

it follows from prior equations that the dynamics of the protein count variance and covariance are given by

$$\begin{aligned} \frac{\partial}{\partial t}\sigma^2_{Y_1} &= \mu\langle X\rangle + 2\mu\sigma^2_{X,Y_1} + \gamma_1\langle Y_1\rangle - 2\gamma_1\sigma^2_{Y_1} & (55)\\ \frac{\partial}{\partial t}\sigma^2_{Y_2} &= \mu(R-\langle X\rangle) - 2\mu\sigma^2_{X,Y_2} + \gamma_2\langle Y_2\rangle - 2\gamma_2\sigma^2_{Y_2} & (56)\\ \frac{\partial}{\partial t}\sigma^2_{Y_1,Y_2} &= \mu(\sigma^2_{X,Y_2} - \sigma^2_{X,Y_1}) - (\gamma_1+\gamma_2)\sigma^2_{Y_1,Y_2}. & (57)\end{aligned}$$

Notice in particular that only the terms σ^2_{X,Y_i}, i.e. correlations between ribosome state and protein number, introduce correlations between proteins.

B.2 Steady state means, variances, and covariance

Solution of Eqs. 31–51 at steady state is straightforward, since it can be done iteratively from lower to higher orders in protein moments. In lowest order, this leads to

$$\langle 1 \rangle = 1 \tag{58}$$

$$\langle X \rangle = \frac{(T_s + T_a)R}{2T_s} \tag{59}$$

$$\langle X^2 \rangle = \frac{R(T_s + T_a)((R-1)T_a + (R+1)T_s - R)}{2T_s(2T_s - 1)}. \tag{60}$$

Protein steady state mean values are

$$\langle Y_1 \rangle = \frac{\mu R(T_s + T_a)}{2T_s\gamma_1} \tag{61}$$

$$\langle Y_2 \rangle = \frac{\mu R(T_s - T_a)}{2T_s\gamma_2}. \tag{62}$$

Protein steady state variances are

$$\begin{aligned} \sigma^2_{Y_1} = {} & \frac{\mu(T_s + T_a)R}{4\gamma_1(2T_s - 1)T_s^2\left[(1-p)2\mu T_s + (2T_s + 1 - R)\gamma_1\right]} \cdot \\ & \cdot \left[(R(R-1) - (2R+2)T_s + 8T_s^2 - 2T_s(2T_s - 2 + R)p)\mu T_s \right. \\ & \left. + 2(2T_s - 1)T_s(2T_s - R + 1)\gamma_1 \; + \; (2T_s - R)(R - 2pT_s - 1)\mu T_a\right] \end{aligned} \tag{63}$$

$$\begin{aligned} \sigma^2_{Y_2} = {} & \frac{\mu(T_s - T_a)R}{4\gamma_2(2T_s - 1)T_s^2\left[(1-p)2\mu T_s + (2T_s + 1 - R)\gamma_2\right]} \cdot \\ & \cdot \left[(R(R-1) - (2R+2)T_s + 8T_s^2 - 2T_s(2T_s - 2 + R)p)\mu T_s \right. \\ & \left. + 2(2T_s - 1)T_s(2T_s - R + 1)\gamma_2 \; - \; (2T_s - R)(R - 2pT_s - 1)\mu T_a\right] \end{aligned} \tag{64}$$

and the protein steady state covariance is

$$\sigma^2_{Y_1,Y_2} = \frac{\mu^2 R(2T_s - R)(R - 2pT_s - 1)(T_s^2 - T_a^2)\left[4(1-p)\mu T_s + (2T_s + 1 - R)(\gamma_1 + \gamma_2)\right]}{4(\gamma_1 + \gamma_2)(2T_s - 1)T_s^2[2(1-p)\mu T_s + (2T_s + 1 - R)\gamma_1]\left[2(1-p)\mu T_s + (2T_s + 1 - R)\gamma_2\right]}. \tag{65}$$

Note that the antisymmetric piece T_a does not appear in the denominator and only appears in integer powers in the numerator. This last fact is useful when performing sums over mRNA distributions, as will be discussed in Section C.2.

The sign of the correlation

$$\mathcal{C}(Y_1, Y_2) \equiv \frac{\sigma^2_{Y_1,Y_2}}{\sqrt{\sigma^2_{Y_1}\sigma^2_{Y_2}}} \tag{66}$$

can be determined from that of the covariance $\sigma^2_{Y_1,Y_2}$. Assuming $T_s > |T_a|$ (positive mRNA levels for both types) and $2T_s > R$ (ribosomes are overloaded), it can be shown that

$$p > \frac{R-1}{2T_s} \iff \mathcal{C}(Y_1, Y_2) < 0 \tag{67}$$

$$p < \frac{R-1}{2T_s} \iff \mathcal{C}(Y_1, Y_2) > 0 \tag{68}$$

$$p = \frac{R-1}{2T_s} \iff \mathcal{C}(Y_1, Y_2) = 0. \tag{69}$$

Thus, in the overloaded regime, there is a critical rebinding probability $p^* = \frac{R-1}{2T_s} \in (0,1)$ above which correlations are negative. This contrasts with the underloaded regime ($2T_s \leq R$), already treated in the main text, where the correlation is always zero.

Fano factors F_1 and F_2 for the two protein types are defined by

$$F_1 \equiv \frac{\sigma^2_{Y_1}}{\langle Y_1 \rangle} \tag{70}$$

$$F_2 \equiv \frac{\sigma^2_{Y_2}}{\langle Y_2 \rangle}. \tag{71}$$

These indicate whether protein count distributions are under ($F < 1$) or over-dispersed ($F > 1$) relative to the Poisson distribution that occurs in the underloaded case. Assuming $T_s > |T_a|$ (positive mRNA levels for both types) and $2T_s > R$ (ribosomes are overloaded), it can be shown that

$$p > \frac{R-1}{2T_s} \iff F_i > 1 \tag{72}$$

$$p < \frac{R-1}{2T_s} \iff F_i < 1 \tag{73}$$

$$p = \frac{R-1}{2T_s} \iff F_i = 1. \tag{74}$$

Notice that these conditions for the Fano factor are closely related to those for correlations (Eqs. 67–69).

In the overloaded regime, the positive correlations between protein counts that arise most noticeably in the absence of rebinding ($p = 0$) can be understood intuitively as follows. When a ribosome binds to a particular type of mRNA, say type 1, this reduces the fraction of free type 1 mRNA's relative to free type 2 mRNA's, thus reducing the probability that a type 1 mRNA will be bound by the next free ribosome. This effect tends to "focus" the dynamics of X, the number of ribosomes bound to mRNAs of type 1, more so than if the pools of mRNA were infinite, in which case the ribosomes operate independently, and the steady state protein counts are uncorrelated for $p = 0$. Evidence of this focusing can be seen by noticing that the variance for X (via Eqs. 59 and 60) is $(T_1 + T_2 - R)/(T_1 + T_2 - 1)$ times the variance for X that is obtained when mRNA populations are infinite but present in the same proportions as in the finite pool case. In the latter case, for $p = 0$, ribosomes act independently, and protein levels are uncorrelated. Since the above factor satisfies $0 \leq (T_1 + T_2 - R)/(T_1 + T_2 - 1) < 1$ for $R > 1$, fluctuations in X are smaller for finite mRNA counts than for infinite counts. This "focusing" leads to an increased probability for ribosomes to bind to mRNAs of type 2 after binding to an mRNA of type 1, thus the production of a protein of type 1 makes a subsequent production of a protein of type 2 more probable than in the "unbiased" independent case. As a result, the respective protein counts become positively correlated.

B.3 Steady state means, variances, and covariance (symmetric case)

The moments for the symmetric case, which are reported in the main text, are trivially obtained from the asymmetric moments in Section B.2 by substituting $T_a = 0$ and $\gamma_1 = \gamma_2 = \gamma$. This leads to the following expressions

$$\langle Y_1 \rangle = \langle Y_2 \rangle = \frac{\mu R}{2\gamma} \tag{75}$$

$$\begin{aligned}\sigma^2_{Y_1} &= \sigma^2_{Y_2} = \frac{\mu R}{4\gamma(2T_s-1)T_s\left[2(1-p)\mu T_s+(2T_s-R+1)\gamma\right]}\cdot \\ &\cdot\ \big[(R(R-1)-(2R+2)T_s+8T_s^2-2T_s(2T_s-2+R)p)\mu T_s \\ &+ 2(2T_s-1)T_s(2T_s-R+1)\gamma\big] \end{aligned} \tag{76}$$

$$\sigma^2_{Y_1,Y_2} = \frac{\mu^2 R(2T_s-R)(R-2pT_s-1)}{4\gamma(2T_s-1)[2(1-p)\mu T_s+(2T_s-R+1)\gamma]}. \tag{77}$$

C Inclusion of mRNA and ribosome fluctuations

The results in Section B, for our stochastic model with fixed numbers of mRNAs and ribosomes, are useful for understanding crosstalk in a translational bottleneck, but translational bottlenecks in living cells exist in the context of both fluctuating mRNAs and ribosome concentrations. In this section, using a time scale separation formalism, we outline how these effects can be layered onto our existing analysis.

C.1 Formulae for total variance and correlation

Assuming that fluctuations in the numbers of mRNAs (T_1, T_2) and ribosomes R occur on a slow time scale relative to the time it takes for the protein counts to reach equilibrium, we will formally approximate steady state statistics for the protein counts for the model with mRNA and ribosome fluctuations by averaging the moment formulas of the previous section over a steady state distribution for (T_1, T_2, R). To this end, we introduce the following notation for averages over distributions of collections $\mathcal{R}$ of such variables

$$\{Z(\mathcal{R})\} = \sum_r Z(r)\,\mathcal{P}(\mathcal{R}=r) \tag{78}$$

where $Z(\cdot)$ is some function of the collection of variables, e.g. the variance of Y_i conditional on given mRNA levels, and $\mathcal{P}(\mathcal{R}=r)$ is the probability to observe a certain value r for the variables $\mathcal{R}$. The sum (rather than an integral) appears here as the random variables T_1, T_2, R are discrete. (In Section C.4, we shall consider further approximations where the sum will be replaced by an integral and time-dependent moments will enter.)

The mean statistics have a particularly simple form with this additional layer of averaging, e.g. the mean protein level for Y_i is the nested average $\{\langle Y_i\rangle\}$. However, by the law of total variance and covariance, the variance $\Sigma^2_{Y_i}$ for Y_i and the covariance $\Sigma^2_{Y_1,Y_2}$ between Y_1 and Y_2 have the more complicated forms

$$\Sigma^2_{Y_i} = \left\{\langle Y_i\rangle^2\right\} - \{\langle Y_i\rangle\}^2 + \left\{\sigma^2_{Y_i}\right\} \tag{79}$$

$$\Sigma^2_{Y_1,Y_2} = \{\langle Y_1\rangle\langle Y_2\rangle\} - \{\langle Y_1\rangle\}\{\langle Y_2\rangle\} + \left\{\sigma^2_{Y_1,Y_2}\right\}. \tag{80}$$

The total variance in Eq. 79 is the sum of a term that is the variance of the conditional mean, $\left\{\langle Y_i\rangle^2\right\} - \{\langle Y_i\rangle\}^2$, and a term that is the mean of the conditional variance, $\left\{\sigma^2_{Y_i}\right\}$. A similar decomposition holds for the total covariance. We will demonstrate that both terms in the decomposition are important when understanding our model for translational crosstalk.

It is a straightforward matter to then show that the total correlation $C_{\text{total}}(Y_1, Y_2)$, which incorporates the variation of mRNA and ribosome levels, is given by

$$C_{\text{total}}(Y_1, Y_2) = \frac{\Sigma^2_{Y_1,Y_2}}{\sqrt{\Sigma^2_{Y_1}\Sigma^2_{Y_2}}} = \rho_m \kappa_m + \rho_\sigma \kappa_\sigma \tag{81}$$

where

$$\kappa_m \equiv \frac{\{\langle Y_1\rangle \langle Y_2\rangle\} - \{\langle Y_1\rangle\}\{\langle Y_2\rangle\}}{\sqrt{\left(\left\{\langle Y_1\rangle^2\right\} - \{\langle Y_1\rangle\}^2\right)\left(\left\{\langle Y_2\rangle^2\right\} - \{\langle Y_2\rangle\}^2\right)}} \tag{82}$$

$$\kappa_\sigma \equiv \frac{\left\{\sigma^2_{Y_1,Y_2}\right\}}{\sqrt{\left\{\sigma^2_{Y_1}\right\}\left\{\sigma^2_{Y_2}\right\}}} \tag{83}$$

$$\eta_i \equiv \frac{\left\{\langle Y_i\rangle^2\right\} - \{\langle Y_i\rangle\}^2}{\left\{\sigma^2_{Yi}\right\}} \tag{84}$$

$$\rho_m \equiv (1+\eta_1^{-1})^{-1/2}(1+\eta_2^{-1})^{-1/2} \tag{85}$$

$$\rho_\sigma \equiv (1+\eta_1)^{-1/2}(1+\eta_2)^{-1/2} \tag{86}$$

and κ_m and κ_σ can be interpreted as variation of the mean and mean variation correlations, respectively. It can be shown that

$$0 \leq \rho_m \tag{87}$$

$$0 \leq \rho_\sigma \tag{88}$$

by the η's being non-negative, and

$$\rho_m + \rho_\sigma = \frac{\sqrt{\eta_1} + \sqrt{\eta_2^{-1}}}{\sqrt{(1+\eta_1)(1+\eta_2^{-1})}} \leq 1 \tag{89}$$

by the Cauchy-Schwarz inequality. The ρ terms can then be interpreted as partial probabilities. Furthermore, in the symmetric case where $\eta_1 = \eta_2$, then

$$\rho_m + \rho_\sigma = 1 \tag{90}$$

so the ρ's can be interpreted as probabilities. The special form of the total correlation $C_{\text{total}}(Y_1, Y_2)$ in Eq. 81 thus looks like a weighted average of the correlations κ_m and κ_σ.

Eq. 81 leads to an important simplification in the case when mRNA levels are fluctuating and ribosome levels are fixed, and the mRNA distributions are such that ribosomes are always overloaded ($T_1 + T_2 > R$). From the previous section, we have that the steady state mean protein levels are deterministically related in the overloaded case by the relationship

$$\gamma_1 \langle Y_1\rangle + \gamma_2 \langle Y_2\rangle = \mu R \tag{91}$$

that is independent of mRNA levels. This immediately leads to the relationship $\kappa_m = -1$ for averaging over steady state means. By Eqs. 82–86, If $\eta_1 \gg 1$ and $\eta_2 \gg 1$, then the total correlation $C_{\text{total}}(Y_1, Y_2)$ is close to -1. Of course, this supposes that the system is purely overloaded, but we find approximations below that are reasonably accurate when the system is occasionally underloaded.

C.2 Exact correlation formula for Poisson distributed mRNA counts

Under the model for mRNA production and degradation associated with Eqs. 19-21, in steady state, T_1, T_2 are independent Poisson distributed random variables with means $\tau_1 = \alpha_1/\delta$, $\tau_2 = \alpha_2/\delta$, respectively. When mRNA copy numbers change slowly relative to the protein dynamics and ribosome numbers are fixed, we expect that using this for the distribution of T_1, T_2 in Eqs. 79–80 will yield good approximations for the steady state variances and covariances for the protein counts. Indeed, the resulting expressions can be written exactly in terms of generalized hypergeometric functions; these are standard functions in computer algebra packages such as Maple which facilitates the evaluation of the correlation. Derivation of these expressions is straightforward but a bit cumbersome, so we only present a sketch of the derivation here.

We suppose that the total mRNA counts T_1, T_2 are independently Poisson distributed with means τ_1, τ_2, respectively, so that their joint probability distribution is given by

$$\mathcal{P}(T_1 = n_1, T_2 = n_2) = \frac{e^{-2\tau}\tau_1^{n_1}\tau_2^{n_2}}{n_1! n_2!} \tag{92}$$

where

$$\tau = \frac{\tau_1 + \tau_2}{2}. \tag{93}$$

In this case, we can simplify the probability ρ_O to be overloaded or balanced ($T_1 + T_2 \geq R$) or the probability ρ_U to be strictly underloaded ($T_1 + T_2 < R$):

$$\rho_O \equiv \mathcal{P}(T_1 + T_2 \geq R) = \sum_{n=R}^{\infty} \frac{e^{-2\tau}(2\tau)^n}{n!} = 1 - \tilde{\Gamma}(R, 2\tau) \tag{94}$$

$$\rho_U \equiv \mathcal{P}(T_1 + T_2 < R) = \sum_{n=0}^{R-1} \frac{e^{-2\tau}(2\tau)^n}{n!} = \tilde{\Gamma}(R, 2\tau) \tag{95}$$

where the normalized incomplete Gamma function $\tilde{\Gamma}(R, 2\tau)$ is defined as

$$\tilde{\Gamma}(R, 2\tau) \equiv \frac{\Gamma(R, 2\tau)}{\Gamma(R)} \tag{96}$$

$$\Gamma(R, 2\tau) \equiv \int_{2\tau}^{\infty} e^{-x}\, x^{R-1}\, dx \tag{97}$$

$$\Gamma(R) \equiv \int_{0}^{\infty} e^{-x}\, x^{R-1}\, dx. \tag{98}$$

Equations 94–95 can also be used to estimate the relevance of underloaded vs. overloaded contributions to statistics, and we will use these again in later sections.

The key remaining steps are to sum the distribution in Eq. 92 over the expressions for $\langle Y_i \rangle$, $\langle Y_i \rangle^2$, and $\left(\langle Y_i^2 \rangle - \langle Y_i \rangle^2 \right)$. The reader can verify that all these summands in both the overloaded and underloaded cases only have asymmetric terms T_a in the numerator of expressions, which suggests splitting the summation into two pieces by the transformation $n_1 \to n$, $n_2 \to N - n$:

$$\sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} W(n_1, n_2) \;\to\; \sum_{N=0}^{R} \sum_{n=0}^{N} W(n, N-n) + \sum_{N=R+1}^{\infty} \sum_{n=0}^{N} W(n, N-n). \tag{99}$$

(Note that we use N here as a dummy variable; it should not be confused with our use of N in Section A for the number of states of the Markov chain.) The transformation in Eq. 99 splits the calculation into underloaded and overloaded pieces. The sum over n in Eq. 99 leaves symmetric terms $(n_1 + n_2)$ constant, such that the sum over n primarily involves the integer powers of $T_a = (n_1 - n_2)/2$ in the numerator. It turns out that the sum over n can be expressed in terms of sums

$$\sum_{n=0}^{N} \frac{x^n y^{N-n} n^m}{n!(N-n)!} \tag{100}$$

for non-negative integer m, and positive real values x and y. These can be simplified by the generating function

$$Q_N(s) = \sum_{n=0}^{N} \frac{x^n y^{N-n} e^{sn}}{n!(N-n)!} = \frac{(y + xe^s)^N}{N!} \tag{101}$$

which provides these sums by computing

$$\sum_{n=0}^{N} \frac{x^n y^{N-n} n^m}{n!(N-n)!} = \lim_{s \to 0} \frac{\partial^m}{\partial s^m} Q_N(s). \tag{102}$$

A few explicit solutions are

$$\begin{aligned}
\sum_{n=0}^{N} \frac{x^n y^{N-n} n^0}{n!(N-n)!} &= \frac{(x+y)^N}{N!} \\
\sum_{n=0}^{N} \frac{x^n y^{N-n} n^1}{n!(N-n)!} &= \frac{(x+y)^{N-1}\, xN}{N!} \\
\sum_{n=0}^{N} \frac{x^n y^{N-n} n^2}{n!(N-n)!} &= \frac{(x+y)^{N-2}\, xN\, (xN+y)}{N!}.
\end{aligned}$$

The remainder of the calculation is a summation over N, which can be separated into overloaded $(N > R)$ and underloaded $(0 \le N \le R)$ pieces.

Here we present the expressions for protein moments obtained by averaging conditional moment expressions against the joint Poisson distribution for the mRNA counts given in Eq. (92). These expressions can be used in Eqs. 79–86 to calculate the correlation. The expressions can be written in terms of generalized hypergeometric functions and special cases of generalized hypergeometric

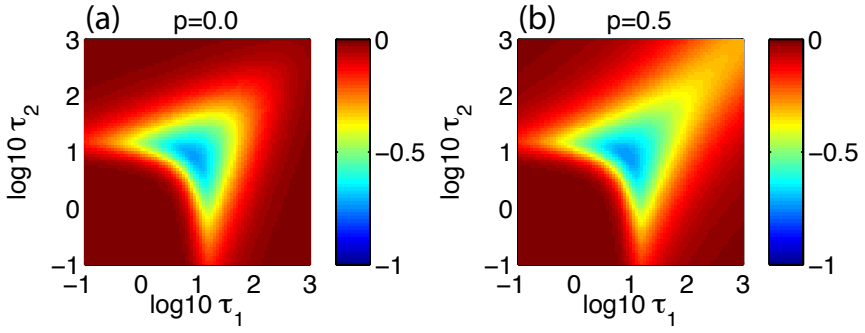

Figure 1: Plot of the protein correlation for mRNA abundances T_i that are independently Poisson distributed with means τ_i, $i = 1, 2$ (see Section C.2 in the S.M.). The correlation is observed to be non-positive up to numerical error. Other parameters are $R = 10$, $\mu = 1$ and $\gamma_1 = \gamma_2 = 0.1$.

functions (exponential, incomplete Gamma function, etc.); we provide a few examples of this later in this section. We suppose that $\gamma_1 = \gamma_2 = \gamma$ for simplicity, though this is not necessary. The mean protein level can be calculated as

$$\{\langle Y_1 \rangle\} = \frac{\mu\tau_1\, e^{-2\tau}}{\gamma}\left[R\sum_{N=R+1}^{\infty}\frac{(2\tau)^{N-1}}{N!} + \sum_{N=0}^{R}\frac{(2\tau)^{N-1}\,N}{N!}\right] \tag{103}$$

The expression for $\{\langle Y_2 \rangle\}$ is obtained by symmetry via interchanging τ_1 and τ_2 (similarly for the results below). The average of the squared conditional mean is needed for computing the variance of the conditional mean, and its value is

$$\left\{\langle Y_1 \rangle^2\right\} = \frac{\mu^2\tau_1 e^{-2\tau}}{\gamma^2}\left[R^2\sum_{N=R+1}^{\infty}\frac{(2\tau)^{N-2}(\tau_2 N^{-1}+\tau_1)}{N!} + \sum_{N=0}^{R}\frac{(2\tau)^{N-2}(\tau_2 N+\tau_1 N^2)}{N!}\right] \tag{104}$$

Finally, the average of the product of the conditional means is given by

$$\{\langle Y_1 \rangle \langle Y_2 \rangle\} = \frac{\mu^2\tau_1\tau_2 e^{-2\tau}}{\gamma^2}\left[R^2\sum_{N=R+1}^{\infty}\frac{(2\tau)^{N-2}(1-N^{-1})}{N!} + \sum_{N=0}^{R}\frac{(2\tau)^{N-2}(N^2-N)}{N!}\right] \tag{105}$$

The average of the conditional variance of the protein is the most complicated expression in this section, being

$$\left\{\left\langle \sigma^2_{Y_1}\right\rangle\right\} = \frac{\mu\tau_1 e^{-2\tau}}{\gamma}\left[R\sum_{N=R+1}^{\infty}\frac{(2\tau)^{N-2}(C_{-1}N^{-1}+C_0+C_1 N)}{N!\,(D_0+D_1 N)} + \sum_{N=0}^{R}\frac{(2\tau)^{N-1}\,N}{N!}\right] \tag{106}$$

$$C_1 \equiv (2\tau - p\tau_1)\mu + 2\tau\gamma \tag{107}$$

$$C_0 \equiv -\tau_2\left((1+p)R-1\right)\mu - 2\tau(R-1)\gamma \tag{108}$$

$$C_{-1} \equiv \mu R\,\tau_2\,(R-1) \tag{109}$$

$$D_1 \equiv \mu(1-p)+\gamma \tag{110}$$

$$D_0 \equiv -(R-1)\gamma \tag{111}$$

while the average of the conditional covariance has a similar structure

$$\left\{\left\langle \sigma^2_{Y_1,Y_2}\right\rangle\right\} = -\frac{\mu^2\tau_1\tau_2 e^{-2\tau}}{\gamma}\left[R\sum_{N=R+1}^{\infty}\frac{(2\tau)^{N-2}(B_{-1}N^{-1}+B_0+B_1N)}{N!\,(D_0+D_1N)}\right] \tag{112}$$

$$B_1 \equiv p \tag{113}$$

$$B_0 \equiv -((1+p)R-1) \tag{114}$$

$$B_{-1} \equiv R(R-1). \tag{115}$$

These latter sums depend on calculating expressions of the general form

$$\sum_{N=R+1}^{\infty}\frac{x^N}{N!}\left(\frac{N^m}{-1+aN}\right) \tag{116}$$

for integer m, and positive real value x, and sufficiently positive real value a. Expressions such as those in Eq. 116 can be readily expressed in terms of generalized hypergeometric functions by comparing Eq. 116 to the series definition of a generalized hypergeometric function. A few explicit solutions are

$$\sum_{N=R+1}^{\infty}\frac{x^N}{N!}\left(\frac{N^{-1}}{-1+aN}\right)$$
$$=\left(\frac{x^{R+1}}{(R+1)!(-1+a(R+1))(R+1)}\right){}_3\mathrm{F}_3\left(\left[1,R+1,\frac{-1+a(R+1)}{a}\right];\left[R+2,R+2,\frac{-1+a(R+2)}{a}\right];x\right)$$
$$\sum_{N=R+1}^{\infty}\frac{x^N}{N!}\left(\frac{N^{0}}{-1+aN}\right)$$
$$=\left(\frac{x^{R+1}}{(R+1)!(-1+a(R+1))}\right){}_2\mathrm{F}_2\left(\left[1,\frac{-1+a(R+1)}{a}\right];\left[R+2,\frac{-1+a(R+2)}{a}\right];x\right)$$
$$\sum_{N=R+1}^{\infty}\frac{x^N}{N!}\left(\frac{N^{1}}{-1+aN}\right)$$
$$=\left(\frac{x^{R+1}\,(R+1)}{(R+1)!(-1+a(R+1))}\right){}_2\mathrm{F}_2\left(\left[1,\frac{-1+a(R+1)}{a}\right];\left[R+1,\frac{-1+a(R+2)}{a}\right];x\right)$$

which use standard notation for generalized hypergeometric functions. Some summations require calculation of the sum in Eq. 116 when $a = 0$, and these sums can similarly be expressed in terms of hypergeometric functions.

C.3 Approximate formula for correlation with small variation in mRNA and ribosome numbers on a slow time scale

Useful approximations for the correlation can be derived for a variety of cases where the quantities T_1, T_2, R vary slowly, such that the averages in Eqs. 79–81 can be used to determine the total correlation $C_{\mathrm{total}}(Y_1, Y_2)$. Consider the quasi-steady state approximation, where mRNA counts T_i and ribosome count R are sampled from respectively independent distributions (the case of dependent variables can also be treated). We focus here on the case where T_1 and T_2 are identically distributed, though the analysis can be readily extended to the non-symmetric case. We will initially assume that the system is overloaded ($T_1 + T_2 > R$), but we will generalize to arbitrary cases in the latter portion of the section. Although our procedures here use formal expansions

and interpolations, they seem to produce reasonably accurate results when compared to simulation when mRNA and ribosome dynamics are slow relative to protein dynamics.

Suppose that the total abundance for each type of mRNA is a random variable with variance σ_T^2, and the number of ribosomes is a random variable with variance σ_R^2. To lowest order in the variances σ_T^2 and σ_R^2, the terms $\left\{\sigma_{Y_1,Y_2}^2\right\}$, $\left\{\sigma_{Y_1}^2\right\}$, and $\left\{\sigma_{Y_2}^2\right\}$ can be approximated by their zeroth order approximation, i.e. the steady state values evaluated at the mean of the parameter distributions. The lowest order expressions for the variances and covariance of the steady state mean values in Eqs. 61–62 can be shown to be

$$\left\{\langle Y_i\rangle^2\right\} - \left\{\langle Y_i\rangle\right\}^2 \approx \frac{\mu^2 R^2 \sigma_T^2}{8T_s^2\gamma_i^2} + \frac{\mu^2\sigma_R^2}{4\gamma_i^2} \tag{117}$$

$$\left\{\langle Y_1\rangle \langle Y_2\rangle\right\} - \left\{\langle Y_1\rangle\right\}\left\{\langle Y_2\rangle\right\} \approx -\frac{\mu^2 R^2 \sigma_T^2}{8T_s^2\gamma_1\gamma_2} + \frac{\mu^2\sigma_R^2}{4\gamma_1\gamma_2} \tag{118}$$

as can be derived by taking linear expansions for the expressions of mean protein count and averaging over joint distributions for ribosome and mRNA copy numbers. This leads to total covariance (see Eqs. 79–80)

$$\Sigma_{Y_i}^2 = \bar{\sigma}_{Y_i}^2 + \frac{\mu^2 R^2 \sigma_T^2}{8T_s^2\gamma_i^2} + \frac{\mu^2\sigma_R^2}{4\gamma_i^2} \tag{119}$$

$$\Sigma_{Y_1,Y_2}^2 = \bar{\sigma}_{Y_1,Y_2}^2 - \frac{\mu^2 R^2 \sigma_T^2}{8T_s^2\gamma_1\gamma_2} + \frac{\mu^2\sigma_R^2}{4\gamma_1\gamma_2}. \tag{120}$$

where $\bar{\sigma}_{Y_i}^2$ and $\bar{\sigma}_{Y_1,Y_2}^2$ are the zeroth order expressions (Eqs. 63–65) for the variance and covariance, with mRNA and ribosome counts evaluated at their mean values. The approximate expression for correlation then is

$$C_{\text{total}}(Y_1,Y_2) \approx \frac{\bar{\sigma}_{Y_1,Y_2}^2 - \frac{\mu^2 R^2\sigma_T^2}{8T_s^2\gamma_1\gamma_2} + \frac{\mu^2\sigma_R^2}{4\gamma_1\gamma_2}}{\Pi_{i=1,2}\sqrt{\bar{\sigma}_{Y_i}^2 + \frac{\mu^2 R^2\sigma_T^2}{8T_s^2\gamma_i^2} + \frac{\mu^2\sigma_R^2}{4\gamma_i^2}}} \tag{121}$$

where $\bar{\sigma}_{Y_1,Y_2}^2$ and $\bar{\sigma}_{Y_i}^2$ are σ_{Y_1,Y_2}^2 and $\sigma_{Y_i}^2$, respectively, evaluated at mean values for mRNA and ribosome counts. Notice that this correlation becomes strongly negative when mRNA fluctuations dominate (σ_T^2 is sufficiently large relative to the $\bar{\sigma}^2$ terms, and $\sigma_T^2 \gg \sigma_R^2$). In contrast, this correlation becomes strongly positive when ribosome fluctuations dominate (σ_R^2 is sufficiently large relative to the $\bar{\sigma}^2$ terms, and $\sigma_R^2 \gg \sigma_T^2$).

A similar argument can be made for the underloaded case ($T_1 + T_2 \leq R$). In this case, the mean of the covariance is precisely zero to all orders. The variance and covariance of the means are also zero in this case, as can be seen from the expression $\langle Y_i\rangle = \mu T_i/\gamma_i$ for the mean in the underloaded case. Thus, the total correlation $C_{\text{total}}(Y_1,Y_2)$ is precisely zero in this case.

There is then the issue that underloaded and overloaded approximations have a discontinuity at the balance point of the system, specifically in the expressions for the variances and covariance of the means $\langle Y_i\rangle$. In order to bridge the approximations in the underloaded and overloaded cases, we consider the probability ρ_O for the system to be overloaded or balanced

$$\rho_O \equiv \mathcal{P}(T_1 + T_2 \geq R) \tag{122}$$

which can be evaluated analytically in certain cases (see Eq. 96). Then $C_{\text{total}}(Y_1, Y_2) \approx 0$ when $\rho_O \approx 0$, and $C_{\text{total}}(Y_1, Y_2)$ is approximately its overloaded value (Eq. 121) when $\rho_O \approx 1$. We thus propose the ansatz

$$C_{\text{total}}(Y_1, Y_2) \approx \rho_O \cdot \left(\frac{\bar{\sigma}^2_{Y_1,Y_2} - \frac{\mu^2 R^2 \sigma_T^2}{8T_s^2 \gamma_1 \gamma_2} + \frac{\mu^2 \sigma_R^2}{4\gamma_1 \gamma_2}}{\Pi_{i=1,2} \sqrt{\bar{\sigma}^2_{Y_i} + \frac{\mu^2 R^2 \sigma_T^2}{8T_s^2 \gamma_i^2} + \frac{\mu^2 \sigma_R^2}{4\gamma_i^2}}} \right) \tag{123}$$

This approximation appears to be good across overloaded and underloaded conditions, as evidenced by comparison to analytic solutions in Fig. 6 (see blue curves).

C.4 Approximate formulae for small mRNA fluctuations on a moderate timescale

When mRNA count fluctuations occur on a time scale commensurate with that of the protein dynamics, we explore a formal generalization of the previous analysis where the mRNA counts $T_1(t)$ and $T_2(t)$ are treated as random, time-dependent parameters in the ODE's in Eqs. 31–57 for the moments. By then approximating each of the mRNA counts as a process that is easy to describe (Ornstein-Uhlenbeck), we can appeal to standard linearization theory to derive approximations for the variability and correlations in protein counts. This approximation, which we outline below, is observed to provide reasonably accurate results for several explored parameter sets (see Fig. 6 in the main text). Comments concerning significant caveats associated with this approximation are presented at the end of this section. Due to limitations of this method, we will ignore ribosome count fluctuations in this discussion (also see comments later in this section).

This approximation procedure is formally described as follows. Suppose we know trajectories for mRNA counts $T_1(t)$ and $T_2(t)$. Treating $T_1(t)$ and $T_2(t)$ as time-dependent parameters in Eqs. 31–57, the trajectory for a moment, e.g. $\langle X \rangle(t)$, follows deterministically. Since this is true for any trajectories $T_1(t)$ and $T_2(t)$, if the trajectories $T_1(t)$ and $T_2(t)$ are then allowed to be random (sampled from some distribution), then this imposes a probability distribution on the trajectories of any conditional moments. In particular, we can in principle know the averages $\{\langle X \rangle(t)\}$, $\{\langle X \rangle(t)^2\}$, etc., at any given time t, where $\{\cdot\}$ now denotes averaging with respect to the distribution of the mRNA trajectories. If $T_1(t)$ and $T_2(t)$ have a steady state distribution, we can expect that the averaged conditional moments approach steady state values as $t \to \infty$. We denote these limits of time-dependent quantities such as $\{\langle X \rangle(t)\}$ and $\{\langle X \rangle(t)^2\}$ by $\{\langle X \rangle\}$ and $\left\{\langle X \rangle^2\right\}$, respectively.

The primary difficulty is then calculating these steady state moments, largely because $T_1(t)$ and $T_2(t)$ enter nonlinearly into Eqs. 31–57. Our primary simplification is linearizing Eqs. 31–57 with respect to $T_1(t)$ and $T_2(t)$. Our second simplification is approximating discrete birth-death process models for $T_1(t)$ and $T_2(t)$ by simple real-valued processes with a timescale set by the death rate δ (1).

If the mRNA counts $T_1(t)$ and $T_2(t)$ are assumed to be independent Ornstein-Uhlenbeck colored noises with identical means T_s, steady state variances σ_T^2, and correlation natural times $1/\delta$, we can form lowest order approximations using standard linearization theory that are numerically accurate for many of the simulations considered. Similarly as in Section C.3, we average over conditional moments to derive new moments for protein counts. This average uses a linearized version (with respect to mRNA counts) of the ODE's Eqs. 31–57 for the protein moments. The linearized ODE's

Eqs. 31–57 thus act like a linear filter for the Ornstein-Uhlenbeck processes approximating the mRNA counts. The combined system is then a multivariate linear system with additive Gaussian white noises (two variables for the mRNA's, one variable for each relevant protein moment), which can be solved by standard techniques (1). The detailed steps for computing second order moments involve constructing a matrix based on the linear coefficients of Eqs. 31–57, and then performing a straightforward but cumbersome matrix calculation. These calculations were done using the software package Maple (Waterloo Maple Inc.), and we only present results here.

We now outline our results using these approximations for the symmetric case, i.e. when the processes for the mRNA counts are identical but independent. As before when approximating the protein correlation, the averages $\left\{\sigma^2_{Y_1,Y_2}\right\}$, $\left\{\sigma^2_{Y_1}\right\}$, and $\left\{\sigma^2_{Y_2}\right\}$ can be well-approximated by their zeroth order values $\bar{\sigma}^2_{Y_1,Y_2}$, $\bar{\sigma}^2_{Y_1}$, and $\bar{\sigma}^2_{Y_2}$, respectively, which are evaluated at mean mRNA counts T_s. We can show using our linearized system derived in the overloaded condition that the variance of the means satisfies

$$\left\{\langle Y_1\rangle^2\right\} - \left\{\langle Y_1\rangle\right\}^2 = \left\{\langle Y_2\rangle^2\right\} - \left\{\langle Y_2\rangle\right\}^2 \approx \frac{2\mu^2(\delta+\gamma+k)B^2\sigma_T^2}{\gamma\,(\gamma+\delta)(\gamma+k)(k+\delta)\,k} \tag{124}$$

$$B \equiv \frac{\mu R(1-p)}{2\,(2T_s - R + 1)} \tag{125}$$

$$k \equiv \frac{2\mu T_s(1-p)}{2T_s - R + 1} \tag{126}$$

and the cross moment of the means satisfies

$$\left\{\langle Y_1\rangle\,\langle Y_2\rangle\right\} - \left\{\langle Y_1\rangle\right\}\left\{\langle Y_2\rangle\right\} = -\frac{2\mu^2(\delta+\gamma+k)B^2\sigma_T^2}{\gamma\,(\gamma+\delta)(\gamma+k)(k+\delta)\,k} \tag{127}$$

Protein correlation in the underloaded case is zero, so as before, we approximate the total correlation using an ansatz similar to that we have already introduced

$$C_{\text{total}}(Y_1,Y_2) \approx \rho_O \cdot \left(\frac{\bar{\sigma}^2_{Y_1,Y_2} - \frac{2\mu^2(\delta+\gamma+k)B^2\sigma_T^2}{\gamma\,(\gamma+\delta)(\gamma+k)(k+\delta)\,k}}{\bar{\sigma}^2_{Y_1} + \frac{2\mu^2(\delta+\gamma+k)B^2\sigma_T^2}{\gamma\,(\gamma+\delta)(\gamma+k)(k+\delta)\,k}}\right) \tag{128}$$

Eq. 128 appears to reasonably approximate stochastic simulations (red) in Fig. 6 of the main text, though this approximation becomes worse as the parameter p is increased (compare Fig. 6a and Fig. 6d).

This inaccuracy with increasing p can be partially explained by the following. Treating the counts $T_1(t)$ and $T_2(t)$ as random, time-dependent parameters in the ODE's Eqs. 31–57 can miss important features of a full stochastic model where the dynamics of total mRNA counts $T_1(t)$ and $T_2(t)$ statistically influence other system variables, e.g. the count X. This is true even though the ODE's Eqs. 31–57 for moment evolution are themselves exact, conditional on fixed mRNA and ribosome counts. For example, suppose we have a system where rebinding always occurs ($p = 1$). We can show that the moment ODE's Eq. 32 for $\langle X\rangle$ predicts that $\langle X\rangle$ is constant in time, i.e. that the mean number of mRNA's of type 1 bound to ribosomes is constant, even if $T_1(t)$ and $T_2(t)$ are time-dependent. Compare this to a full stochastic model that allows mRNA's to be degraded off of ribosomes, even if only slowly. We can show then that $\langle X\rangle$ generally changes as a function of time, in contradiction to blind application of the ODE's Eqs. 31–57. Despite this caveat, we find

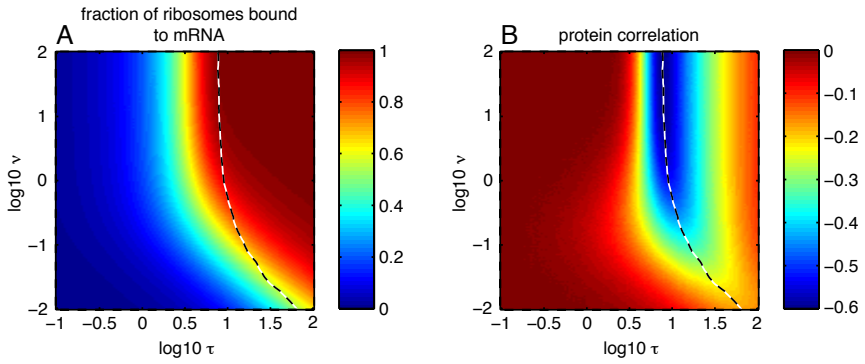

Figure 2: The negative correlation resonance is preserved when the common binding rate ν of mRNA to ribosomes is finite rather than infinite, as evidenced by stochastic simulations where mRNA numbers fluctuate according to a birth-death process (see Eqs. 19–21 of the main text) with mean count τ for each species, while the ribosome count is fixed. (A) The probability that a ribosome is bound to mRNA decreases for constant τ but decreasing ν, due to a lower rate of mRNA binding to ribosomes. This effectively leads to lower competition for free mRNA's to bind to ribosomes. (B) The correlation between protein levels as τ is varied continues to undergo a negative correlation resonance for a range of ν. Both plots have a black and white dashed line where the correlation minimum is reached as τ is scanned. As ν is decreased, the value of the correlation at the resonance slowly decreases in absolute magnitude, and the value of τ where the resonance occurs increases in magnitude. Note that since molecular competition is reduced as ν is reduced, the correlation minimum tends to occur at higher mean mRNA levels (higher τ). In the simulations for this figure, mRNA's have identical production rates $\alpha = \tau\delta$ and degradation rates $\delta = 0.1$. The translation rate per ribosome is $\mu = 1$, there are $R = 10$ ribosomes for the simulation, and there is no rebinding ($p = 0$). Statistics were derived by: (1) simulating ensembles (size 12800) well into steady state (500 time units), (2) time-averaging ensemble statistics over an additional 500 time units.

the approximation works quite well for the case without rebinding ($p = 0$) or with weak rebinding ($p \approx 0$), and based on this accuracy, we conjecture that a rigorous approximation scheme may exist for these cases. However, we leave potential proof of such approximations to future work.

D Simulation results with finite binding rates

All preceding results were obtained in the limit of infinite binding rate of mRNA molecules to ribosomes. To address this limitation, we have conducted extensive stochastic simulations (using highly parallel GPU coprocessors) of models where the binding rates of mRNA molecules to ribosomes are finite, and where the binding rates may differ between mRNA species (ν in Eq. 1 of the main text is then replaced by ν_i). When both binding rates are equal and finite, we find that the correlation resonance is preserved as the mean mRNA count is varied (see sample results in Fig. 2). However, the location of this resonance is shifted to higher mRNA mean count for very small mRNA binding rates. We also observe that the correlation resonance is preserved when the binding rates for different mRNA molecules are unequal (see sample results in Fig. 3).

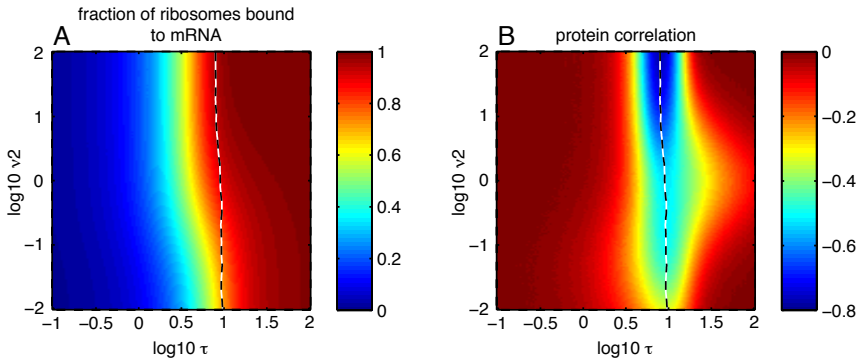

Figure 3: Similar to Fig. 2, except that the rate $\nu_1 = 1$ for mRNAs of type 1 to bind to ribosomes is fixed, while ν_2 is varied. The correlation resonance occurs reliably near the balance point. The location of the minimum of the correlation as τ is varied for fixed ν_1 and ν_2 is indicated by a dashed line.

Supporting Reference

1. Gardiner, C, 2009. *Stochastic Models: A Handbook for the Natural and Social Sciences.* Springer, Berlin.