Supporting Information

Theoretical quantification of interference in the TASEP: application to mRNA translation shows near-optimality of termination rates

Khanh Dao Duc[1,2], Zain H. Saleem[1], and Yun S. Song[1,2,3,4,*]

[1] Department of Mathematics, University of Pennsylvania, PA 19104, USA
[2] Department of Biology, University of Pennsylvania, PA 19104, USA
[3] Computer Science Division and Department of Statistics, University of California, Berkeley, CA 94720, USA
[4] Chan Zuckerberg Biohub, San Francisco, CA 94158

[*] To whom correspondence should be addressed: yss@berkeley.edu

1 Equations satisfied by the correlators in the TASEP

Averaging the master equation associated with the TASEP, the particle densities satisfy the following relations [1]:

$$0 = \langle\tau_1\rangle_N - \langle\tau_1\tau_2\rangle_N - \alpha(1 - \langle\tau_1\rangle_N), \tag{S1}$$

$$0 = \langle\tau_i\tau_{i+1}\rangle_N - \langle\tau_{i-1}\tau_i\rangle_N - \langle\tau_i\rangle_N + \langle\tau_{i-1}\rangle_N, \quad \text{for } 2 \le i \le N-1, \tag{S2}$$

$$0 = \beta\langle\tau_N\rangle_N - \langle\tau_{N-1}\rangle_N + \langle\tau_{N-1}\tau_N\rangle_N. \tag{S3}$$

Note that (S2) implies $\langle\tau_i(1-\tau_{i+1})\rangle_N = \langle\tau_{i-1}(1-\tau_i)\rangle_N$ for all $i = 2,\dots,N-1$. This translation-invariant quantity is called the current (or flux) and is denoted by J. One can also relate the two-point correlators with the three-point correlators as

$$0 = \langle\tau_1\tau_2\tau_3\rangle_N - \langle\tau_1\tau_2\rangle_N(1+\alpha) + \alpha\langle\tau_2\rangle_N, \tag{S4}$$

$$0 = \langle\tau_{i-1}\tau_i\tau_{i+1}\rangle_N - \langle\tau_{i-2}\tau_{i-1}\tau_i\rangle_N - \langle\tau_{i-1}\tau_i\rangle_N + \langle\tau_{i-2}\tau_i\rangle_N, \quad \text{for } 3 \le i \le N-1, \tag{S5}$$

$$0 = \langle\tau_{N-2}\tau_{N-1}\tau_N\rangle_N - \langle\tau_{N-2}\tau_N\rangle_N + \beta\langle\tau_{N-1}\tau_N\rangle_N.$$

2 Description of the matrix Ansatz used in the simple TASEP

To derive analytical expressions for the average densities of the TASEP, Derrida *et al.* [2] showed that the steady state probability of a given configuration can be derived using a matrix formulation as

$$\mathbb{P}(t_1,\dots,t_N) = \frac{f_N(t_1,\dots,t_N)}{\sum_{\theta_1=0,1}\cdots\sum_{\theta_N=0,1} f_N(\theta_1,\dots,\theta_N)},$$

where

$$f_N(t_1, \dots, t_N) = \langle W| \prod_{i=1}^{N} (t_i D + (1 - t_i) E |V\rangle.$$

Here, D and E are infinite dimensional square matrices and $|V\rangle$ and $\langle W|$ are column and row vectors respectively satisfying

$$\begin{aligned} DE &= D + E, \\ D\,|V\rangle &= \frac{1}{\beta}\,|V\rangle, \\ \langle W|\,E &= \frac{1}{\alpha}\,\langle W|. \end{aligned}$$

Using this formulation, the particle density can be derived as

$$\langle \tau_i \rangle_N = \frac{\langle W|C^{i-1} D C^{N-i}|V\rangle}{\langle W|C^N|V\rangle},$$

where $C = D + E$. More generally, for any given index set $i_1, i_2, \dots, i_k$ such that $1 \leq i_1 < \dots < i_k \leq N$, we get

$$\langle \tau_{i_1} \dots \tau_{i_k} \rangle_N = \frac{\langle W|C^{i_1-1} D C^{i_2-i_1-1} \dots C^{i_k-i_{k-1}-1} D C^{N-i_k}|V\rangle}{\langle W|C^N|V\rangle}.$$

Using these algebraic rules, Derrida et al. [2] obtained exact formulas for $\langle \tau_i \rangle_N$. More precisely, for $i \leq N-1$,

$$\langle \tau_i \rangle_N = \sum_{p=0}^{N-i-1} \frac{(2p)!}{p!(p+1)!} \frac{\langle W|C^{N-1-p}|V\rangle}{\langle W|C^N|V\rangle} + \frac{\langle W|C^{i-1}|V\rangle}{\langle W|C^N|V\rangle} \sum_{p=2}^{N-i+1} \frac{(p-1)(2N-2i-p)!}{(N-i)!(N-i+1-p)!} \frac{1}{\beta^p},$$

while for $i = N$,

$$\langle \tau_N \rangle_N = \frac{1}{\beta} \frac{\langle W|C^{N-1}|V\rangle}{\langle W|C^N|V\rangle},$$

where

$$\langle W|C^N|V\rangle = \sum_{p=1}^{N} \frac{p(2N-1-p)!}{N!(N-p)!} \left[\frac{\frac{1}{\beta^{p+1}} - \frac{1}{\alpha^{p+1}}}{\frac{1}{\beta} - \frac{1}{\alpha}} \right],$$

and $\langle W|V\rangle = 1$.

3 Computing the density of isolated particles

Using the matrix Ansatz, we derive here an analytical expression for the average density of isolated particles $\langle \tau'_j \rangle_N$. Our goal is to get $\langle \tau'_j \rangle_N$ as a function of the average densities $\langle \tau_j \rangle_N$. The density of isolated particles inside the lattice ($2 \leq i \leq N-1$) is given by (see equation (2))

$$\langle \tau'_i \rangle_N = \langle \tau_i \rangle_N - \langle \tau_{i-1} \tau_i \rangle_N - \langle \tau_i \tau_{i+1} \rangle_N + \langle \tau_{i-1} \tau_i \tau_{i+1} \rangle_N. \tag{S6}$$

For $2 \le j \le N-1$, we first derive the expression of the two point correlators $\langle \tau_j \tau_{j+1} \rangle_N$ by summing equation (S2) over $i \in \{2, \dots, j\}$ and using the boundary equation (S1)

$$\langle \tau_j \tau_{j+1} \rangle_N = \langle \tau_j \rangle_N - \alpha \left(1 - \langle \tau_1 \rangle_N\right). \tag{S7}$$

Similarly, for $3 \le j \le N-1$, summing equation (S5) from $i = 3$ to j and using boundary equations (S1) and (S4) gives

$$\begin{aligned} \langle \tau_{j-1} \tau_j \tau_{j+1} \rangle_N &= \langle \tau_1 \tau_2 \tau_3 \rangle_N + \sum_{p=3}^{j} \langle \tau_{p-1} \tau_p \rangle_N - \langle \tau_{p-2} \tau_p \rangle_N \\ &= (1+\alpha)^2 \langle \tau_1 \rangle_N - \alpha(1 + \alpha + \langle \tau_2 \rangle_N) + \sum_{p=3}^{j} \langle \tau_{p-1} \tau_p \rangle_N - \langle \tau_{p-2} \tau_p \rangle_N. \end{aligned} \tag{S8}$$

Using the matrix formulation and the identities $DCD = D(DC - DE + ED) = DDC - DC + CD$, we get

$$\begin{aligned} \langle \tau_{p-2} \tau_p \rangle_N &= \frac{\langle W | C^{p-3} DCDC^{N-p} | V \rangle}{\langle W | C^N | V \rangle} \\ &= \langle \tau_{p-2} \tau_{p-1} \rangle_N + J_N \left(\langle \tau_{p-1} \rangle_{N-1} - \langle \tau_{p-2} \rangle_{N-1} \right), \end{aligned} \tag{S9}$$

where $J_N = \frac{\langle W | C^{N-1} | V \rangle}{\langle W | C^N | V \rangle} = \alpha \left(1 - \langle \tau_1 \rangle_N\right)$ is the particle current at steady state [2]. Combining (S9) with (S8) and using (S6) and (S1) yield the result for the three-point correlator

$$\langle \tau_{j-1} \tau_j \tau_{j+1} \rangle_N = \langle \tau_{j-1} \rangle_N - \alpha \left[1 + \alpha + \langle \tau_2 \rangle_N - (2+\alpha) \langle \tau_1 \rangle_N\right] - J_N \left(\langle \tau_{j-1} \rangle_{N-1} - \langle \tau_1 \rangle_{N-1} \right) \tag{S10}$$

for $3 \le j \le N-1$. Using (S1) and (S4), this equation is also true for $j = 2$. Using (S10), (S7) and (2) gives us the formula for the density of isolated particles

$$\langle \tau_i' \rangle_N = \alpha \left[1 - \langle \tau_2 \rangle_N + \alpha(\langle \tau_1 \rangle_N - 1)\right] - J_N \left(\langle \tau_{i-1} \rangle_{N-1} - \langle \tau_1 \rangle_{N-1} \right), \quad \text{for } 2 \le i \le N-1.$$

Finally we can use $J_N = \alpha \left(1 - \langle \tau_1 \rangle_N\right)$ to write the above formula in a more compact notation, as

$$\langle \tau_i' \rangle_N = D_0(\alpha, \beta, N) - D_1(\alpha, \beta, N) \langle \tau_{i-1} \rangle_{N-1}, \tag{S11}$$

where

$$\begin{aligned} D_0(\alpha, \beta, N) &= \alpha \left[1 - \langle \tau_2 \rangle_N + \alpha(\langle \tau_1 \rangle_N - 1)\right] + \alpha \left(1 - \langle \tau_1 \rangle_N\right) \langle \tau_1 \rangle_{N-1}, \\ D_1(\alpha, \beta, N) &= \alpha(1 - \langle \tau_1 \rangle_N). \end{aligned}$$

Similarly, using equations (S1) and (S3) at the boundaries yields

$$\begin{aligned} \langle \tau_1' \rangle_N &= \alpha(1 - \langle \tau_1 \rangle_N), \\ \langle \tau_N' \rangle_N &= \langle \tau_N \rangle_N (1+\beta) - \langle \tau_{N-1} \rangle_N. \end{aligned}$$

4 Asymptotics of the simple TASEP

We provide here the asymptotics for the densities of the TASEP. For large lattice size N, the flux of particles J is given by [2]

$$J \sim \begin{cases} \frac{1}{4}, & \text{if } \alpha > \frac{1}{2},\ \beta > \frac{1}{2} \text{ (MC regime)}, \\ \alpha(1-\alpha), & \text{if } \alpha < \frac{1}{2},\ \beta > \alpha \text{ (LD regime)}, \\ \beta(1-\beta), & \text{if } \beta < \frac{1}{2},\ \beta < \alpha \text{ (HD regime)}. \end{cases}$$

The densities at the boundaries and at positions next to them are given by [2]

$$\begin{aligned} \langle \tau_1 \rangle_N &\sim 1 - \frac{J}{\alpha}, \\ \langle \tau_2 \rangle_N &\sim 1 - J - \left(\frac{J}{\alpha}\right)^2, \\ \langle \tau_{N-1} \rangle_N &\sim J + \left(\frac{J}{\beta}\right)^2, \\ \langle \tau_N \rangle_N &\sim \frac{J}{\beta}. \end{aligned}$$

Out of the boundaries and for large $1 \ll n \ll N$ [2, 3],

$$\langle \tau_{N-n} \rangle_N \sim \begin{cases} \frac{1}{2} - \frac{1}{2\sqrt{\pi n}} + \frac{(2\beta-1)^2+4}{16\sqrt{\pi}(2\beta-1)^2 n^{3/2}}, & \text{if } \alpha > \frac{1}{2},\ \beta > \frac{1}{2} \text{ (MC regime)}, \\ \alpha + \left(\frac{\alpha(1-\alpha)}{\beta(1-\beta)}\right)^{n+1}(1-2\beta), & \text{if } \alpha < \beta < \frac{1}{2} \text{ (LD I regime)}, \\ \alpha + \frac{4^n(\alpha(1-\alpha))^{n+1}}{\sqrt{\pi} n^{3/2}}\left[\frac{1}{(1-2\beta)^2} - \frac{1}{(1-2\alpha)^2}\right], & \text{if } \alpha < \frac{1}{2} < \beta \text{ (LD II regime)}, \\ 1-\beta, & \text{if } \beta < \frac{1}{2},\ \beta < \alpha \text{ (HD regime)}. \end{cases}$$

Using these formulae in (S11) leads to asymptotics for the density of isolated particles.

5 Density of isolated particles in the bulk for the ℓ-TASEP

We compute here an estimate of the density of isolated particles of size ℓ in the bulk ($\langle \tau_i \rangle$, $1 \ll i \ll N - l$). To do so, we use an approximation from Lakatos and Chou [4], assuming that the number of states of n particles of length l, confined to a length of $N' \geq n\ell$ lattice sites, is given by the partition function [5]

$$Z(n, N') = \binom{N' - (\ell-1)n}{n}. \tag{S12}$$

For a given position $i \in \{1, \ldots, \leq N - l\}$, we introduce x_i^- and x_i^+ as the positions of the closest particles to the left and the right of i, respectively, so we get

$$\begin{aligned}\langle \tau_i' \rangle &= \mathbb{P}(\tau_i = 1,\ x_i^- < i - \ell,\ x_i^+ > i + \ell) \\ &= \mathbb{P}(\tau_i = 1)\mathbb{P}(x_i^- < i - \ell,\ x_i^+ > i + \ell \mid \tau_i = 1).\end{aligned} \tag{S13}$$

Assuming x_i^- and x_i^+ being independent yields

$$\langle \tau_i' \rangle = \mathbb{P}(\tau_i = 1)\mathbb{P}(x_i^- < i - \ell \mid \tau_i = 1)\mathbb{P}(x_i^+ > i + \ell \mid \tau_i = 1).$$

Using (S12), the probability $p^+_{n,N'}$ that $x_i^+ > i+\ell$, conditioned on $\tau_i = 1$ and there being n particles in the window $[i + \ell : i + \ell + N' - 1]$ is

$$p^+_{n,N'} = \frac{Z(n, N' - 1)}{Z(n, N')} = \frac{1 - \rho\ell}{1 - \rho(\ell - 1)}, \tag{S14}$$

where $\rho = \frac{n}{N'}$. When n and N' get large and assuming the density of particles in the bulk of the lattice to be approximately constant (denoted $\langle \tau \rangle$), we can replace $p^+_{n,N'}$ and ρ in equation (S14) by $\mathbb{P}(x_i^+ > i + \ell \mid \tau_i = 1)$ and $\langle \tau \rangle$, respectively, which gives

$$\mathbb{P}(x_i^+ > i + \ell \mid \tau_i = 1) = \frac{1 - \langle \tau \rangle \ell}{1 - \langle \tau \rangle (\ell - 1)}.$$

Similarly, we obtain $\mathbb{P}(x_i^- < i - \ell \mid \tau_i = 1) = \frac{1 - \langle \tau \rangle \ell}{1 - \langle \tau \rangle (\ell - 1)}$. Combining these relations and replacing $\mathbb{P}(\tau_i = 1)$ by $\langle \tau \rangle$ in equation (S13), we obtain that the density of isolated particles in the bulk, simply denoted $\langle \tau' \rangle$, is given by

$$\langle \tau' \rangle = \langle \tau \rangle \left(\frac{1 - \ell \langle \tau \rangle}{1 - (\ell - 1)\langle \tau \rangle} \right)^2. \tag{S15}$$

Similarly, for isolation range d, we obtain

$$\langle \tau_i^{(d)} \rangle \sim \mathbb{P}(\tau_i = 1) \left[\frac{Z(n, N' - d)}{Z(n, N')} \right]^2,$$

which simplies to the following expression in the large-N limit:

$$\langle \tau^{(d)} \rangle \sim \langle \tau \rangle \left[\frac{1 - \ell \langle \tau \rangle}{1 - (\ell - 1)\langle \tau \rangle} \right]^{2d}.$$

6 Asymptotics of the ℓ-TASEP

We provide here the asymptotics for the densities and current of the TASEP with extended particles and open boundaries from the mean field model of Lakatos and Chou [4]. The current is given by

$$J \sim \begin{cases} \dfrac{1}{(1+\sqrt{\ell})^2}, & \text{if } \alpha > \alpha^*,\ \beta > \beta^* \text{ (MC regime)}, \\ \dfrac{\alpha(1-\alpha)}{1+(\ell-1)\alpha}, & \text{if } \alpha < \alpha^*,\ \beta > \alpha \text{ (LD regime)}, \\ \dfrac{\beta(1-\beta)}{1+(\ell-1)\beta}, & \text{if } \beta < \beta^*,\ \beta < \alpha \text{ (HD regime)}, \end{cases} \tag{S16}$$

where $\alpha^* = \beta^* = \frac{1}{1+\sqrt{\ell}}$. The density $\langle \tau_i \rangle$ in the bulk (position $i \in [\ell+1 : N-\ell-1]$) is then approximated by

$$\langle \tau \rangle \sim \begin{cases} \dfrac{1}{\sqrt{\ell}(\sqrt{\ell}+1)}, & \text{if } \alpha > \alpha^*,\ \beta > \beta^* \text{ (MC regime)}, \\ \dfrac{1+(\ell-1)J - \sqrt{(1+(\ell-1)J)^2 - 4\ell J}}{2\ell}, & \text{if } \alpha < \alpha^*,\ \beta > \alpha \text{ (LD regime)}, \\ \dfrac{1+(\ell-1)J + \sqrt{(1+(\ell-1)J)^2 - 4\ell J}}{2\ell}, & \text{if } \beta < \beta^*,\ \beta < \alpha \text{ (HD regime)}. \end{cases}$$

Using these formulae in (S15) leads to asymptotics for the density of isolated particles.

References

[1] Derrida B, Domany E, Mukamel D (1992) An exact solution of a one-dimensional asymmetric exclusion model with open boundaries. *Journal of Statistical Physics* 69(3-4):667–687.

[2] Derrida B, Evans MR, Hakim V, Pasquier V (1993) Exact solution of a 1d asymmetric exclusion model using a matrix formulation. *Journal of Physics A: Mathematical and General* 26(7):1493.

[3] Blythe RA, Evans MR (2007) Nonequilibrium steady states of matrix-product form: a solver's guide. *Journal of Physics A: Mathematical and Theoretical* 40(46):R333.

[4] Lakatos G, Chou T (2003) Totally asymmetric exclusion processes with particles of arbitrary size. *Journal of Physics A: Mathematical and General* 36(8):2027.

[5] Buschle J, Maass P, Dieterich W (2000) Exact density functionals in one dimension. *Journal of Physics A: Mathematical and General* 33(4):L41.

[6] Weinberg DE, et al. (2016) Improved ribosome-footprint and mrna measurements provide insights into dynamics and regulation of yeast translation. *Cell Reports* 14(7):1787–1799.

[7] Arava Y, et al. (2003) Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences* 100(7):3889–3894.

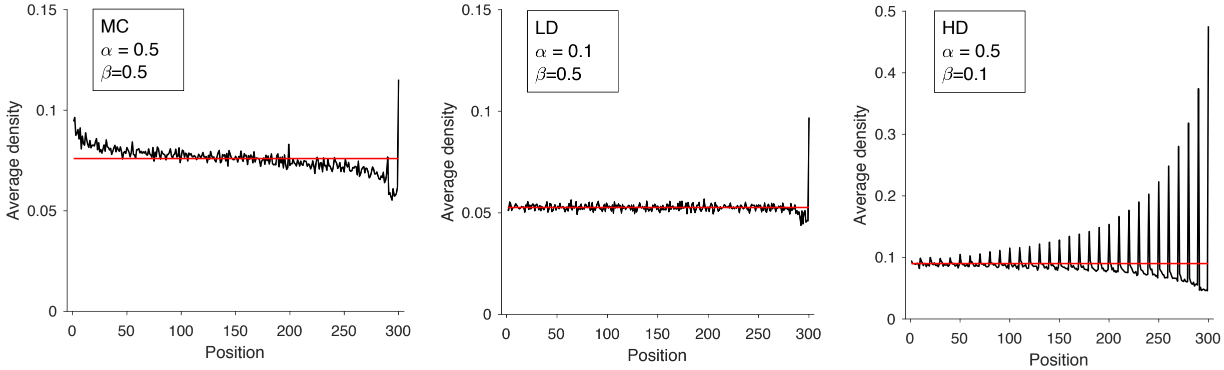

Figure S1: **The density of particles in the ℓ-TASEP model**. We simulated and plot (in black) the density of particles of the ℓ-TASEP ($\ell = 10$) in the different regimes LD, HD and MC. In red, we plot the estimates of the density in the bulk from Lakatos and Chou [4].

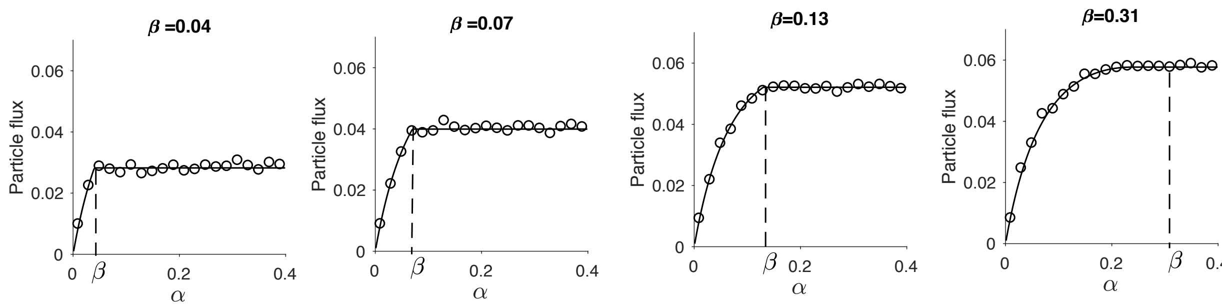

Figure S2: **The particle flux in the ℓ-TASEP model in function of α.** For different values of β, we compare in function of α the flux obtained from Monte Carlo simulations (same as in **Figure 3**B) and asymptotic estimates from Lakatos and Chou [4], given by equation (S16).

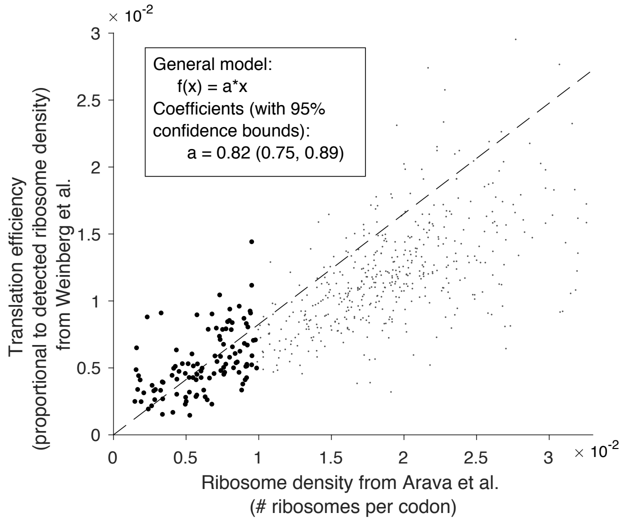

Figure S3: The plot shows the translation efficiency (the number of ribosomes per codon for each mRNA copy, up to a constant) obtained from ribosome profiling data in *S. cerevisiae* (Weinberg *et al.* [6]) against the total ribosome density obtained from polysome profiling (Arava *et al.* [7]). Applying a linear fit $y = ax$ (plotted in dotted line) to genes with total density less than 1 ribosome per 100 codons gives, with 95% confidence interval, $a = 0.82$ $(0.75, 0.89)$.

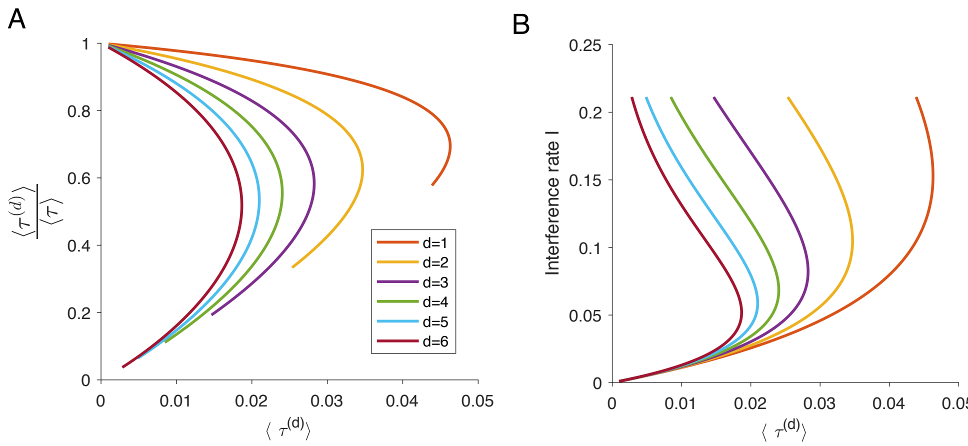

Figure S4: **The fraction of isolated particles and interference rate as a function of $\langle\tau^{(d)}\rangle$. A:** For different isolation ranges $d \in \{1, \ldots, 6\}$, we plot the fraction of isolated particles as a function of the average density of isolated particles $\langle\tau^{(d)}\rangle$, according to (16). Note that for given d, some values of $\langle\tau^{(d)}\rangle$ can lead to two possible fractions of isolated particles. **B:** As in **A**, we plot the isolation rate as a function of the average density of isolated particles $\langle\tau^{(d)}\rangle$, according to (17). Note that for $\langle\tau^{(d)}\rangle \leq 0.02$ and all d, the initiation rates associated with the lower branch are very close.