

Supplementary material

Crowding-induced transcriptional bursts dictate polymerase and nucleosome density profiles along genes

Aafke A van den Berg and Martin Depken

We here give the derivations and arguments that were left out of the main text for clarity. For brevity we will here only keep the 'b' of 'bulk' in the subscript, for example writing v^b instead of v^{bulk} .

1 Heuristic solution for of the hierarchical TASEP model

The average TASEP motor velocity

For the TASEP, gap sizes are geometrically distributed as (21)

$$P_{\text{TASEP}}(g; a) = (1 - a) a^g, \quad (1)$$

for some constant $a < 1$. Unless a motor is blocked by another motor ($g = 0$), it will hop forward with rate k , and the average velocity of motors can be calculated as

$$v^b = k \sum_{g=1}^{\infty} P_{\text{TASEP}}(g; a) = ka. \quad (2)$$

From this it follows that $a = v^b/k$, and we can write

$$P_{\text{TASEP}}(g; v^b/k) = \left(1 - v^b/k\right) \left(v^b/k\right)^g. \quad (3)$$

Intra- and trans-peloton gap sizes

The inclusion of dynamic roadblocks will split the motor dynamics into an intra-peloton and a trans-peloton TASEP describing gap sizes below and above the roadblock shadow Δ . In line with our heuristic argument, we assume there to be no roadblocks within a peloton. Apart from the leading motor, all motors within a peloton thus attempt to hop forward with rate k_{ip} . The pelotons themselves are controlled by the leading motor, which faces a trans-peloton gap filled with roadblocks and thus attempts to hop forward with rate k_{tp} . Assuming that the gap-size distribution is geometric both below and above the roadblock shadow, we can now write our normalized heuristic gap-size distribution for the intra- and trans-peloton regimes as

$$\begin{aligned} P_{\text{ip}}(g; v^b/k_{\text{ip}}) &= \frac{(v^b/k_{\text{ip}})^g}{\sum_{g=0}^{\Delta-1} (v^b/k_{\text{ip}})^g} = \left(1 - v^b/k_{\text{ip}}\right) \left(v^b/k_{\text{ip}}\right)^g + O\left[\left(v^b/k_{\text{ip}}\right)^\Delta\right], & g < \Delta \\ P_{\text{tp}}(g; v^b/k_{\text{tp}}) &= \frac{(v^b/k_{\text{tp}})^g}{\sum_{g=\Delta}^{\infty} (v^b/k_{\text{tp}})^g} = \left(1 - v^b/k_{\text{tp}}\right) \left(v^b/k_{\text{tp}}\right)^{g-\Delta}, & g \geq \Delta. \end{aligned} \quad (4)$$

The above equations are valid when $(k_{\text{ip}}/k_{\text{tp}})^\Delta$ is large, which we refer to as the stable peloton regime (SPR). The condition for the SPR can intuitively be seen as combining the strength of the attraction between motors and roadblocks ($k_{\text{ip}}/k_{\text{tp}}$) and its range (Δ). Due to the SPR conditions exponential dependence on the roadblock shadow size, we expect physiological systems where the roadblock size is substantially larger than the motor step to always be in the SPR. In this limit, we have

$$\begin{aligned} P_{\text{ip}}(g; v^{\text{b}}/k_{\text{ip}}) &= P_{\text{TASEP}}(g; v^{\text{b}}/k_{\text{ip}}), & g < \Delta, \\ P_{\text{tp}}(g; v^{\text{b}}/k_{\text{tp}}) &= P_{\text{TASEP}}(g - \Delta; v^{\text{b}}/k_{\text{tp}}), & g \geq \Delta \end{aligned} \quad (5)$$

With these conditional distributions, we can now calculate the average gaps sizes for both regimes

$$\begin{aligned} \langle g \rangle_{\text{ip}}^{\text{b}} &= \sum_{g=0}^{\infty} g P_{\text{TASEP}}(g; v^{\text{b}}/k_{\text{ip}}) = \frac{v^{\text{b}}}{k_{\text{ip}} - v^{\text{b}}}, \\ \langle g \rangle_{\text{tp}}^{\text{b}} &= \sum_{g=\Delta}^{\infty} g P_{\text{TASEP}}(g - \Delta; v^{\text{b}}/k_{\text{tp}}) = \Delta + \frac{v^{\text{b}}}{k_{\text{tp}} - v^{\text{b}}}. \end{aligned} \quad (6)$$

Defining p^{b} as the probability of a gap in the bulk being a trans-peloton gap, we can write the average gap between motors as $\langle g \rangle^{\text{b}} = (1 - p^{\text{b}})\langle g \rangle_{\text{ip}}^{\text{b}} + p^{\text{b}}\langle g \rangle_{\text{tp}}^{\text{b}}$. Taking the average motor size into account, the average motor density is the inverse of the typical distance between the fronts of neighboring motors,

$$\rho_{\text{m}}^{\text{b}} = \frac{1}{\langle g \rangle^{\text{b}} + \delta_{\text{m}}} = \frac{1}{(1 - p^{\text{b}})\langle g \rangle_{\text{ip}}^{\text{b}} + p^{\text{b}}\langle g \rangle_{\text{tp}}^{\text{b}} + \delta_{\text{m}}}. \quad (7)$$

The relative fraction of trans peloton gaps

Combining the intra- and trans-peloton distributions we can write the complete gap-size distribution of the hierarchical TASEP as

$$P_{\text{hTASEP}}(g; v^{\text{b}}/k_{\text{ip}}, v^{\text{b}}/k_{\text{tp}}, \Delta) = \begin{cases} (1 - p^{\text{b}})P_{\text{TASEP}}(g; v^{\text{b}}/k_{\text{ip}}), & g < \Delta \\ p^{\text{b}}P_{\text{TASEP}}(g - \Delta; v^{\text{b}}/k_{\text{tp}}), & g \geq \Delta \end{cases}. \quad (8)$$

In steady state, the probabilistic flow from intra- to trans-peloton gaps (a peloton is split in two) and from trans- to intra-peloton gaps (two pelotons merge) should balance. The motor ahead of a gap of size $\Delta - 1$ hops with an average rate v and extends the gap to size Δ , inducing the mean-field probabilistic flow $v^{\text{b}}P_{\text{hTASEP}}(\Delta - 1; v^{\text{b}}/k_{\text{ip}}, v^{\text{b}}/k_{\text{tp}}, \Delta) = v^{\text{b}}(1 - p^{\text{b}})P_{\text{TASEP}}(\Delta - 1; v^{\text{b}}/k_{\text{ip}})$. In turn, a motor behind a gap of size hops forward with an average rate and decreases the gap to size $\Delta - 1$, inducing the probabilistic flow $k_{\text{tp}}P_{\text{hTASEP}}(\Delta; v^{\text{b}}/k_{\text{ip}}, v^{\text{b}}/k_{\text{tp}}, \Delta) = k_{\text{tp}}p^{\text{b}}P_{\text{TASEP}}(0; v^{\text{b}}/k_{\text{tp}})$. In the steady state these two flows should balance, and equating these rates gives

$$p^{\text{b}} = \frac{[\langle g \rangle_{\text{tp}}^{\text{b}} - \Delta]}{[\langle g \rangle_{\text{tp}}^{\text{b}} - \Delta] + \langle g \rangle_{\text{ip}}^{\text{b}} \left(\frac{k_{\text{ip}}}{v^{\text{b}}} \right)^\Delta}. \quad (9)$$

Taken together, Equations 6, 7, and 9 relate the average density in the system to the velocity and the microscopic model parameters. In Figure 2 C in the main text we compare the gap-size distribution resulting from our heuristic arguments (solid lines) with ones generated through simulations of the BRM (dots) at different motor densities

2 Relation between heuristic arguments and mean-field solution of BRM

For periodic boundary conditions and irreversible roadblock binding, there exists a mean-field solution for the special case $\delta_m = \delta_{rb} = 1$ [2]. It is informative to consider our more general but less precise heuristic approach in the light of this mean-field solution. We first present a short summary of the mean-field solution, and then show how a few approximations give our heuristic results. The gap-size distribution for the BRM in the mean-field is given by

$$P_{\text{BRM}}(g) = \frac{\prod_{y=1}^g \frac{v^b}{k(y)}}{\sum_{g=0}^{\infty} \prod_{y=1}^g \frac{v^b}{k(y)}}. \quad (10)$$

Here $k(g)$ is the average stepping rate of a motor into a gap of size g . This average stepping rate can be determined by calculating the probability of a roadblock with binding rate k_b being bound to a lattice site after a time t since the last motor passed, $P_{rb}(t) = 1 - e^{-k_b t}$. If we make the mean-field assumption that the motor moves with the average velocity v of the system, the probability of a roadblock being bound can be written in terms of the gap size to the motor ahead $P_{rb}(g) = 1 - e^{-gk_b/v^b}$. The effective stepping rate of a motor as a function of the gap ahead is now given by

$$k(g) = P_{rb}(g)k_{tp} + (1 - P_{rb}(g))k_{ip} = k_{tp} + (k_{ip} - k_{tp})e^{-gk_b/v^b} = k_{tp} + (k_{ip} - k_{tp})e^{-g/(\Delta-1)} \quad (11)$$

where we in the last step introduced the roadblock shadow $\Delta = 1 + v^b/k_b$ for the BRM, and in direct analogy with our heuristic approach. If we now approximate $k(g)$ by a step function

$$k(g) = \begin{cases} k_{ip}, & g < \Delta \\ k_{tp}, & g \geq \Delta \end{cases} \quad (12)$$

the gap-size distribution is given by

$$P_{\text{BRM}}(g) \propto \begin{cases} \left(v^b/k_{ip}\right)^g, & g < \Delta \\ \left(k_{tp}/k_{ip}\right)^{\Delta-1} \left(v^b/k_{tp}\right)^g, & g \geq \Delta \end{cases} \quad (13)$$

Enforcing normalization, Equation 13 implies Equation 4.

3 Observable bulk quantities

In addition to the motor density, there are other interesting observables that can be calculated if we know the average velocity in the bulk. Among them are the current of motors, $J = \rho_m^b v^b$, and the average number of motors in a peloton, $\langle n_p \rangle^b = 1/p^b$. Further, only gaps between pelotons are filled with roadblocks, and then typically only beyond the roadblock shadow. If we let ρ_{rb}^{eq} be the equilibrium roadblock occupancy in the absence of motor activity, we can estimate the average roadblock occupancy as

$$\rho_{rb}^b = \rho_{rb}^{\text{eq}} \frac{p^b [\langle g \rangle_{tp}^b - \Delta]}{\langle g \rangle^b + \delta_m}. \quad (14)$$

In supplementary Figure 1 we illustrate the relationships derived for various observables and check our arguments against simulations of the BRM. In Figure 1 A-D we show the effect of varying the trans-peloton hopping rates for long roadblock equilibration times. As predicted by

our analytical arguments, up to a critical motor density ρ_1 the velocity remains approximately constant (Figure 1 A), and the total current of motors grows linearly with the motor density (Figure 1 B), while the roadblock occupancy decreases linearly with motor density (Figure 1 C). At the critical density, and long before the track is completely covered by motors, all roadblocks are evicted. For motor densities above the critical density, the velocity and motor current follows the relationship for the TASEP without roadblocks (the ipTASEP) (Figure 1 A and B). In Figure 1 D we plot the typical peloton size up to the critical density, after which whole system acts as one large roadblock-excluding peloton. In Figure 1 E-H we vary the roadblock equilibration times. For rapid roadblock equilibration (red curves Figure 1 E-H) the roadblock shadow is small, gaps are largely filled with roadblocks, and the system is well described by a single TASEP with roadblocks in every gap (the tpTASEP) (dashed line in Figure 1 E). In this regime, the total density of roadblocks decreases weakly with motor density (red curve Figure 1 G), as roadblock shadows are small and the motor footprints are all that excludes the roadblocks. For intermediate roadblock equilibration times (the blue curves in Figure 1 E-H) the roadblock shadow is larger, resulting in peloton formation, a velocity that is less sensitive to motor density, and a system that is better at evicting roadblocks. The breakdown of our predictions for roadblock densities and peloton size in the case of fast roadblock binding (red and blue curves in Figure 1 G and H) is not surprising given that we here have small enough roadblock shadows to push the system outside the SPR.

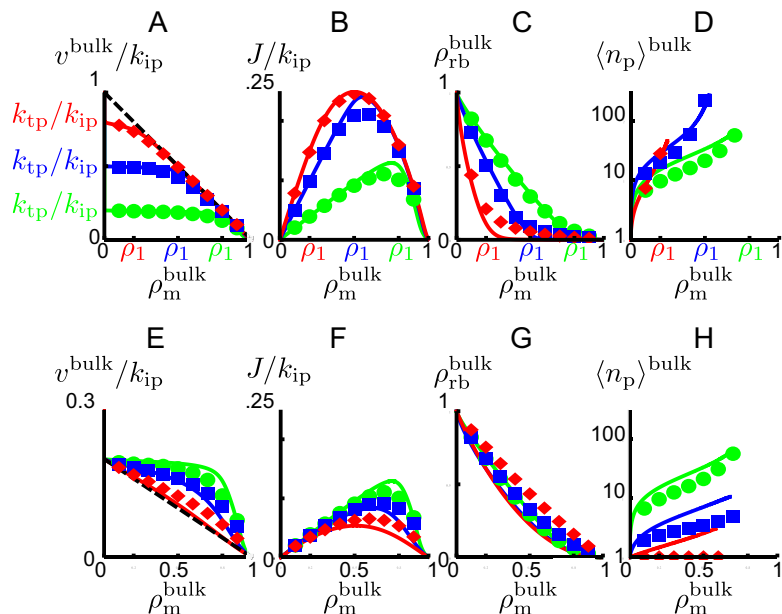


Figure 1: General model captures bulk dynamics of the BRM. Solid lines are analytical predictions and symbols are results from Monte Carlo simulations. Within the rows, each color represents the same parameter values, while green represents the same parameter values across all panels. **(A)-(D)** Systems in the SPR: the parameter values are $k_{ip}\tau = 20$ for all curves, with $k_{tp}/k_{ip} = 0.8$ (red), $k_{tp}/k_{ip} = 0.5$ (blue), and $k_{tp}/k_{ip} = 0.2$ (green). The dashed line in (A) is the velocity relation for the ipTASEP. **(E)-(H)** Sweep from stable to non-stable pelotons: the parameter values are $k_{tp}/k_{ip} = 0.2$ for all curves, and $k_{ip}\tau = 20$ (green), $k_{ip}\tau = 10$ (blue), $k_{ip}\tau = 2$ (red). The dashed line in (E) is the velocity relation for the tpTASEP. In (A) and (E) we show the velocity, in (B) and (F) the motor current, in (C) and (G) the roadblock occupancy and in (D) and (H) the peloton size, all as a function of motor density. In the SPR, it only makes sense to talk about pelotons when there are roadblocks in the system ($\rho_m^{\text{bulk}} < \rho_1$), and we only show the burst size as a function of motor activity for these densities. For small enough roadblock shadows (red and blue line in Figure (E)-(H)) our description breaks down as the system falls outside the SPR.

4 Asymptotic behavior in the SPR

We here show how Equations 6, 7, and 9 can be solved explicitly in terms of the motor density rather than velocity in the limit SPR. For notational convenience we introduce the average excess trans-peloton gap $\langle \tilde{g} \rangle_{\text{tp}}^{\text{b}}$ and the critical density ρ_0 , to write Equation (9) as

$$p^{\text{b}} = \frac{\langle \tilde{g} \rangle_{\text{tp}}^{\text{b}} \rho_0}{1 + \langle \tilde{g} \rangle_{\text{tp}}^{\text{b}} \rho_0}, \quad \langle \tilde{g} \rangle_{\text{tp}}^{\text{b}} = \langle g \rangle_{\text{tp}}^{\text{b}} - \Delta = \frac{v^{\text{b}}}{k_{\text{tp}} - v^{\text{b}}}, \quad \rho_0 = \frac{1}{\langle g \rangle_{\text{ip}}^{\text{b}}} \left(\frac{v^{\text{b}}}{k_{\text{ip}}} \right)^{\Delta}. \quad (15)$$

In the SPR, the transition density ρ_0 is very small by definition. In the limit $\rho_0 \langle \tilde{g} \rangle_{\text{tp}}^{\text{b}} \gg 1$, pelotons are small and p^{b} is close to one, and Equation 3 can be written as

$$1/\rho_{\text{m}}^{\text{b}} - \delta_{\text{m}} \approx \langle g \rangle_{\text{ip}}^{\text{b}} \quad (16)$$

Here we have a system controlled by the tpTASEP. In the limit $\rho_0 \langle \tilde{g} \rangle_{\text{tp}}^{\text{b}} \ll 1$, pelotons are large, p^{b} is small, and Equation 7 can be written as

$$1/\rho_{\text{m}}^{\text{b}} - \delta_{\text{m}} \approx \langle g \rangle_{\text{ip}}^{\text{b}} + \langle \tilde{g} \rangle_{\text{tp}}^{\text{b}} (\langle \tilde{g} \rangle_{\text{tp}}^{\text{b}} + \Delta) \rho_0. \quad (17)$$

If the first two terms on the right hand side dominate, we have the standard ipTASEP. If the last term dominates, we have a composite system. Taken together, there are three limits given by

$$\begin{aligned} \langle g \rangle_{\text{tp}}^{\text{b}} &\approx 1/\rho_{\text{m}}^{\text{b}} - \delta_{\text{m}}, & \rho_{\text{m}}^{\text{b}} &\ll \rho_0 \\ \langle \tilde{g} \rangle_{\text{tp}}^{\text{b}} &\approx \frac{1}{\sqrt{\rho_0 \rho_{\text{m}}^{\text{b}}}}, & \rho_0 &\ll \rho_{\text{m}}^{\text{b}} \ll \rho_1 = \frac{1}{\langle g \rangle_{\text{ip}}^{\text{b}} + \delta_{\text{m}}} \\ \langle g \rangle_{\text{ip}}^{\text{b}} &\approx 1/\rho_{\text{m}}^{\text{b}} - \delta_{\text{m}}, & \rho_1 &\ll \rho_{\text{m}}^{\text{b}}. \end{aligned} \quad (18)$$

In the middle regime is large, and the average velocity (see Equation 6) is close to k_{tp} ,

$$v^{\text{b}} = k_{\text{tp}} \left(1 - \sqrt{\rho_0 \rho_{\text{m}}^{\text{b}}} \right). \quad (19)$$

Using Equations 18 and 19, all observables can be written to leading order in $1/\rho_0$ as

$$\begin{aligned} v^{\text{b}} &= k_{\text{tp}}, & \langle g \rangle_{\text{ip}}^{\text{b}} &= \frac{k_{\text{tp}}}{k_{\text{ip}} - k_{\text{tp}}}, & \langle g \rangle_{\text{tp}}^{\text{b}} &= \frac{1}{\sqrt{\rho_0 \rho_{\text{m}}^{\text{b}}}}, \\ \langle n_{\text{p}} \rangle^{\text{b}} &= \sqrt{\frac{\rho_{\text{m}}^{\text{b}}}{\rho_0}}, & \frac{\rho_{\text{rb}}^{\text{b}}}{\rho_{\text{rb}}^{\text{eq}}} &= 1 - \frac{\rho_{\text{m}}^{\text{b}}}{\rho_1}, & \rho_0 &= \frac{k_{\text{ip}} - k_{\text{tp}}}{k_{\text{tp}}} \left(\frac{k_{\text{tp}}}{k_{\text{ip}}} \right)^{k_{\text{tp}} \tau + \delta_{\text{rb}}} \end{aligned} \quad (20)$$

In this regime, ($\rho_0 \ll \rho_{\text{m}} \ll \rho_1$), the average bulk peloton size is large, and that is why we refer to this as the stable peloton regime.

5 The bulk state is never reached

With the parameter values in Table 1 in the main text, the bulk state is given by

$$\langle g \rangle_{\text{ip}}^{\text{b}} \approx 0.4, \quad \langle g \rangle_{\text{tp}}^{\text{b}} \approx 1 \frac{5 \cdot 10^{82}}{\sqrt{\rho_{\text{m}}}} \gg 1, \quad v^{\text{b}} \approx 3/s, \quad \langle n_{\text{p}} \rangle^{\text{b}} \approx 5 \cdot 10^{82} \sqrt{\rho_{\text{m}}} \gg 1. \quad (21)$$

From the above it is clear that the average steady-state peloton size $\langle n_{\text{p}} \rangle^{\text{b}}$ is in general enormous throughout the experimentally accessible range, and that the true bulk-dynamics will never be reached over a finite gene. Judging by the size of bulk pelotons, polymerases that meet along the gene stay together until termination, invariably producing bursts of RNA production.

6 Initiation limited dynamics

Without roadblocks, and if the initiation rate k_{in} is limiting, the density and the distribution of motors at the start of the lattice are the same as in the bulk [3]. With the inclusion of roadblocks, the microscopic organization among motors and roadblocks can change from the start of the lattice to the bulk. Here we assume that the density of motors is low enough, such that once a motor is initiated it is only slowed down by other motors when it encounters a peloton. This condition means that once the initiation site is cleared, motors step away much faster at the beginning of the lattice than a new motor typically initiates. We will refer to this regime as the slow initiation regime, and we detail its extent below.

The formation of pelotons

A schematic kymograph for the TASEP with roadblocks and open boundary conditions is shown in supplementary Figure 2. At the start of the lattice, the time gaps between newly initiated motors are exponentially distributed with rate k_{in} . As the motors move into the system, those motors that happened to have a roadblock bound ahead will start inducing peloton-forming traffic jams. For convenience we here call these jams proto-pelotons, and they will grow until all motors between roadblocks are absorbed into one peloton. Once all motors are collected into pelotons, we will refer to these as the fully formed pelotons. We here set out to determine the nature of this peloton formation, and what effects it has on both motor and roadblock density profiles.

In all the expressions below, the superscript 'in' refers to the first site after the initiation site for which $x = 0$. For a roadblock to bind to the first site after a motor just left, the motor first has to take δ_{rb} steps and then a roadblock has to bind, all before another motor initiates. Considering the splitting probabilities for each step, we can write the probability of a roadblock binding between two motor initiation events as

$$p^{\text{in}} = \left(\frac{1}{1 + k_{\text{in}}\tau_{\text{m}}^{\text{in}}} \right)^{\delta_{\text{rb}}} \frac{1}{1 + k_{\text{in}}\tau}. \quad (22)$$

Here $\tau_{\text{m}}^{\text{in}}$ is the average time it takes a motor to take a step at the start of the track,

$$\tau_{\text{m}}^{\text{in}} = \frac{1 - p^{\text{in}}}{k_{\text{ip}}} + \frac{p^{\text{in}}}{k_{\text{tp}}}. \quad (23)$$

The definition of the slow initiation regime implies $\tau_{\text{m}}^{\text{in}} \ll 1/k_{\text{in}}$, and Equation 22 can be simplified as

$$p^{\text{in}} = \frac{e^{-\delta_{\text{rb}}\tau_{\text{m}}^{\text{in}}k_{\text{in}}}}{1 + k_{\text{in}}\tau}. \quad (24)$$

Equation 23 and 24 can be used to solve for p^{in} explicitly in the steady state,

$$p^{\text{in}} = \Lambda^{-1} W \left(\frac{k_{\text{in}}\tau}{1 + k_{\text{in}}\tau} \delta_{\text{rb}} \Lambda e^{-k_{\text{in}}\delta_{\text{rb}}/k_{\text{ip}}} \right), \quad \Lambda = \left(\frac{1}{k_{\text{tp}}} - \frac{1}{k_{\text{ip}}} \right) \frac{1}{\tau}, \quad (25)$$

where W is the Lambert W function. In the limit of low initiation rates we can write

$$p^{\text{in}} = \frac{1}{1 + k_{\text{in}}\tilde{\tau}}, \quad \tilde{\tau} = \tau + \delta_{\text{rb}}/k_{\text{tp}}. \quad (26)$$

Next we calculate the typical time it takes for n_{p}^* motors to aggregate into a peloton. Let Δx be the typical distance that the proto-peloton back end moves between two motors joining (supplementary Figure 2). From the geometry of typical times and distances sketched in the kymograph of supplementary Figure 2, we can write

$$\left. \begin{aligned} \Delta t k_{\text{ip}} &= \Delta x \\ \langle g \rangle_{\text{ip}}^{\text{in}} &= \frac{k_{\text{tp}}}{k_{\text{ip}} - k_{\text{tp}}} \\ (1/k_{\text{in}} + \delta_{\text{m}}/k_{\text{ip}} + \Delta t) k_{\text{tp}} &= \Delta x + \delta_{\text{m}} + \langle g \rangle_{\text{ip}}^{\text{in}} \end{aligned} \right\} \Rightarrow \Delta x = \frac{1}{k_{\text{in}} \tau \Lambda} - \delta_{\text{rb}}, \quad (27)$$

where Λ is given in Equation (25). In the last step above, we used the fact that we are considering the slow initiation regime, $k_{\text{in}} \ll k_{\text{ip}}, k_{\text{tp}}$. Proto-pelotons grow as more and more motor catch up, and again ignoring correlations, the final size n_{p}^* of a proto peloton is geometrically distributed as

$$P(n_{\text{p}}^*) = p^{\text{in}}(1 - p^{\text{in}})^{n_{\text{p}}^* - 1} \Rightarrow \langle n_{\text{p}}^* \rangle = 1/p^{\text{in}}. \quad (28)$$

Here the first factor in the probability function accounts for the probability of having a roadblock in a gap, and the following factors accounts for the probability of having no roadblocks in the preceding $n_{\text{p}}^* - 1$ gaps. The probability that a proto-peloton is still growing at position x , or equivalently the probability that $n_{\text{p}}^* > x/\Delta x$, is then given by

$$P_{\text{grow}}(x) = \sum_{n_{\text{p}}^* > x/\Delta x} p^{\text{in}}(1 - p^{\text{in}})^{n_{\text{p}}^* - 1} = (1 - p^{\text{in}})^{x/\Delta x}. \quad (29)$$

Letting $\langle n_{\text{p}}(x) \rangle$ be the size of the average proto-peloton at a distance from the initiation site, we can now write down the discrete evolution equation $\langle n_{\text{p}}(x + \Delta x) \rangle = \langle n_{\text{p}}(x) \rangle + 1 \cdot P_{\text{grow}}$, with $\langle n_{\text{p}}(0) \rangle = 1$, giving

$$\langle n_{\text{p}}(x) \rangle = \sum_{n=1}^{x/\Delta x} (1 - p^{\text{in}})^n + \langle n_{\text{p}}(0) \rangle = \langle n_{\text{p}}^* \rangle - [\langle n_{\text{p}}^* \rangle - 1] e^{-x/x_{\text{p}}}, \quad x_{\text{p}} = -\Delta x / \ln(1 - p^{\text{in}}). \quad (30)$$

Though strictly only true for x in multiples of Δx , we take the above equation to be valid for any position $x \geq 0$.

The macroscopic effects of peloton formation

We now turn to calculate how the gradual growth of the proto pelotons impacts the motor density and velocity. The motor number density $\rho_{\text{m}}(x)$ at any position x is defined as the fraction of time that a site is occupied by (say) the front of a motor. The proto-peloton size at position x tells us that the fraction $\langle n_{\text{p}}(x) \rangle p_{\text{in}}$ motors take an average time $1/k_{\text{tp}}$ to step, while the rest take $1/k_{\text{ip}}$, giving the average stepping time

$$\tau_{\text{m}}(x) = \frac{\langle n_{\text{p}}(x) \rangle p^{\text{in}}}{k_{\text{tp}}} + \frac{1 - \langle n_{\text{p}}(x) \rangle p^{\text{in}}}{k_{\text{ip}}}, \quad \tau_{\text{m}}(0) = \tau_{\text{m}}^{\text{in}}. \quad (31)$$

The total time T between the seeding of two pelotons (supplementary Figure 2) averages over peloton sizes to

$$\langle T \rangle = \langle n_{\text{p}}^* \rangle (1/k_{\text{in}} + \delta_{\text{m}} \tau_{\text{m}}^{\text{in}}). \quad (32)$$

The fraction of time that the track is occupied by motors at position x is then

$$\rho_{\text{m}} = \frac{\langle n_{\text{p}} \rangle \tau_{\text{m}}(x)}{\langle T \rangle} = \frac{\tau_{\text{m}}(x)}{1/k_{\text{in}} + \delta_{\text{m}} \tau_{\text{m}}^{\text{in}}} = \rho_{\text{m}}^* + (\rho_{\text{m}}^{\text{in}} - \rho_{\text{m}}^*) e^{-x/x_{\text{p}}} \quad (33)$$

where the ρ_{m}^* denotes the motor density where the proto pelotons have fully formed. The average motor hopping rate can similarly be written as

$$k_m(x) = p^{\text{in}} \langle n_p(x) \rangle k_{\text{tp}} + (1 - p^{\text{in}} \langle n_p(x) \rangle) k_{\text{ip}} = k^* + (k^{\text{in}} - k^*) e^{-x/x_p}. \quad (34)$$

Though true for the TASEP, note that here the average motor hopping rate (Equation 34) is typically not the same as the average velocity $v(x) = \tau_m^{-1}(x)$ in our system. With the velocity defined this way, it satisfies the standard relation $J = \rho_m(x)v(x)$ where $J = \langle n_p \rangle^* / \langle T \rangle$ is the flux of motors.

In supplementary Figure 2 we indicate the times and positions where there typically are roadblocks in pink. Moving away from the initiation site, the fraction of time a site is occupied by roadblocks grows because motors that have not yet caught up with a proto peloton move faster than the proto pelotons themselves. The moving front of equilibrating roadblocks (intersection of white and pink region in supplementary Figure 2) is typically offset with respect to the last motor of the peloton by a distance δ_{rb} (see dashed square in supplementary Figure 2). Taking the offset δ_{rb} into account, Equation 29 implies that a fraction $P_{\text{grow}}(x + \delta_{\text{rb}})$ of the proto pelotons is still evolving when the front of equilibrating roadblocks is at position x . The increase of the average time a site at position x (see supplementary Figure 2) is occupied by a roadblock then grows with distance as

$$t_{\text{rb}}(x + \Delta x) = t_{\text{rb}}(x) + (1/k_{\text{tp}} - 1/k_{\text{ip}}) \Delta x P_{\text{grow}}(x + \delta_{\text{rb}}), \quad t_{\text{rb}}(0) = 1/k_{\text{in}}. \quad (35)$$

This expression can be summed in the same manner as we previously summed to solve for $\langle n_p(x) \rangle$ in Equation 30, yielding

$$t_{\text{rb}}(x) = [1/k_{\text{tp}} - 1/k_{\text{ip}}] \Delta x e^{-\delta_{\text{rb}}/x_p} (\langle n_p^* \rangle - 1) (1 - e^{-x/x_p}) + 1/k_{\text{in}}. \quad (36)$$

The total fraction of time a site is covered by roadblocks is now

$$\rho_{\text{rb}}(x) = \rho_{\text{rb}}^{\text{eq}} \frac{t_{\text{rb}}(x)}{\langle T \rangle} = \rho_{\text{rb}}^* - (\rho_{\text{rb}}^{\text{in}} - \rho_{\text{rb}}^*) e^{-x/x_p}. \quad (37)$$

Relative changes are often easier to measure experimentally than absolute changes, therefore we here also give the relative changes in velocity and density at the beginning, and compared to well after the pelotons are formed

$$\begin{aligned} \frac{\rho_m^* - \rho_m^{\text{in}}}{\rho_m^*} &= (1 - p^{\text{in}}) \left(1 - \frac{k_{\text{tp}}}{k_{\text{ip}}}\right), & \frac{v^* - v^{\text{in}}}{v^*} &= -(1 - p^{\text{in}}) \left(\frac{k_{\text{ip}}}{k_{\text{tp}}} - 1\right), \\ \frac{\rho_{\text{rb}}^* - \rho_{\text{rb}}^{\text{in}}}{\rho_{\text{rb}}^*} &= \frac{1}{1 + \frac{p^{\text{in}}}{1 - p^{\text{in}}} \frac{\Delta x + \delta_m}{\Delta x} \cdot e^{\delta_{\text{rb}}/x_p}}. \end{aligned} \quad (38)$$

To have an appreciable motor density and velocity evolution, we need only the typical peloton to have a size of a few motors. In the limit of low initiation rates, the above can be written as

$$\begin{aligned} \frac{\rho_m^* - \rho_m^{\text{in}}}{\rho_m^*} &= \frac{k_{\text{in}} \tilde{\tau}}{1 + k_{\text{in}} \tilde{\tau}} \left(1 - \frac{k_{\text{tp}}}{k_{\text{ip}}}\right), & \frac{v^* - v^{\text{in}}}{v^*} &= -\frac{k_{\text{in}} \tilde{\tau}}{1 + k_{\text{in}} \tilde{\tau}} \left(\frac{k_{\text{ip}}}{k_{\text{tp}}} - 1\right), \\ \frac{\rho_{\text{rb}}^* - \rho_{\text{rb}}^{\text{in}}}{\rho_{\text{rb}}^*} &= \frac{k_{\text{in}} \tilde{\tau}}{e^{\delta_{\text{rb}}/x_p} + k_{\text{in}} \tilde{\tau}}. \end{aligned} \quad (39)$$

We see that relative changes along the track grow in magnitude with the initiation rate, but that the effect saturates around $k_{\text{in}} \approx 1/\tilde{\tau}$ for motor-density and velocity changes, while the roadblock density saturates later, around $k_{\text{in}} \approx e^{\delta_{\text{rb}}/x_p}/\tilde{\tau}$. Interestingly, we see that the total shift in motor density and velocity along the track is set solely by the ratio of motor stepping rates with and without roadblocks ahead.

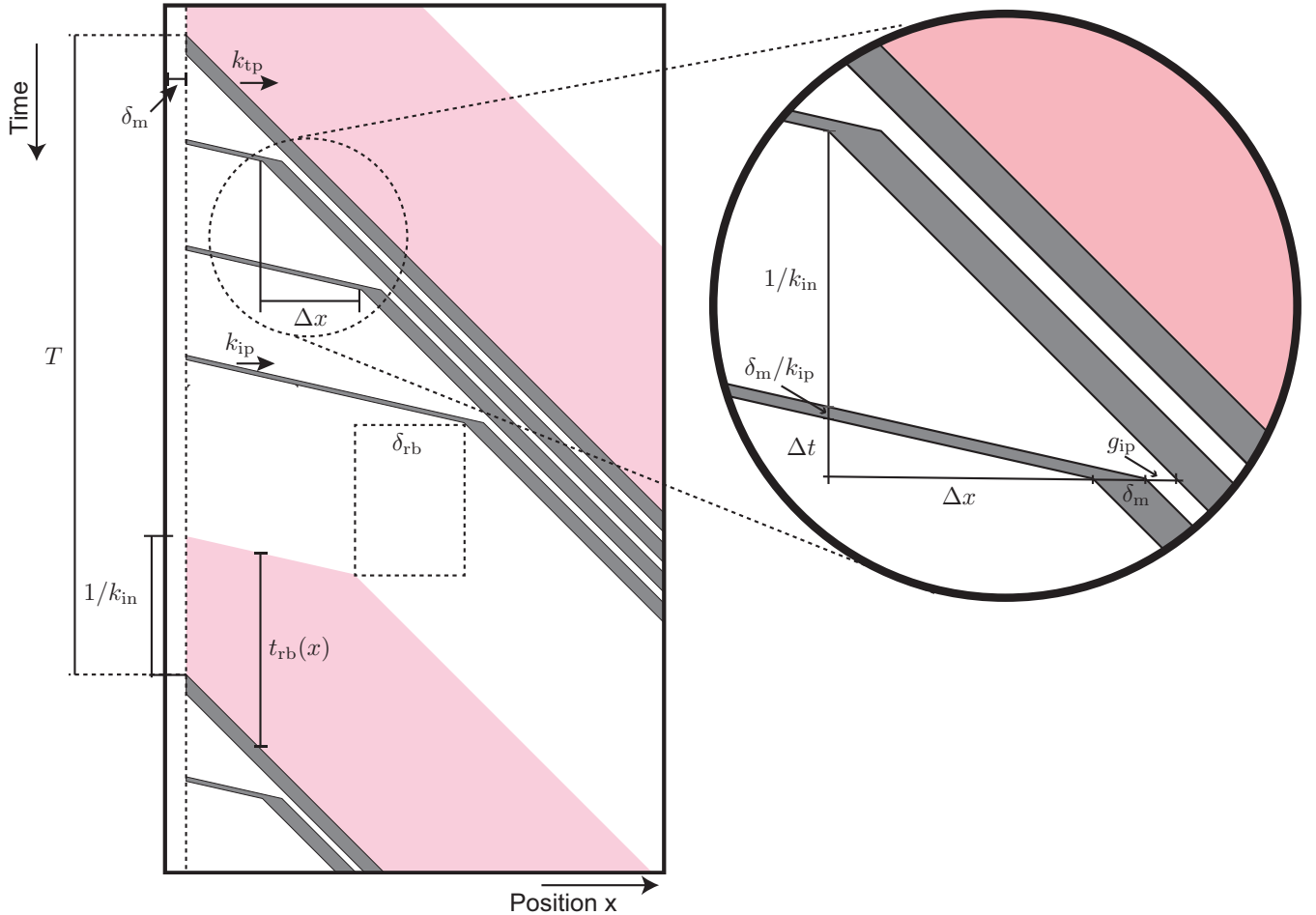


Figure 2: A schematic kymograph of the system with open boundary conditions. Motors (grey) initiate on the left side of the lattice, with an initiation site free of roadblocks. As motors travel into the system, their speed depends on if there is a roadblock (pink) ahead of them. With a roadblock ahead, the motor speed is k_{tp} , while without roadblock ahead it is k_{ip} . The typical time that a roadblock at position x is bound before being evicted by the next peloton is given by $t_{rb}(x)$. The right figure is a magnification of the region defining Δx , which is the typical distance a motor travels to catch up with the proto peloton since the last motor caught up.

7 Bursts from terminating pelotons

The peloton-forming dynamics of our model will manifest as burst of motor activity if viewed from a specific position (for example the transcription-termination site). Using the average velocity of the system, we can translate the average gap-sizes to average time gaps between motors arrivals

$$\tau_{ip} = \frac{\langle g \rangle_{ip} + \delta_m}{v}, \quad \tau_{tp} = \frac{\langle g \rangle_{tp} + \delta_m}{v}. \quad (40)$$

Letting k_{tr} be the rate of reaction when the process is in the on state, k_{off} be the rate at which the system transitions to the off state, and k_{on} be the rate at which the system transitions back to the on state, we can relate our first-principles model to the phenomenological two-state model traditionally used to describe transcriptional bursts (Figure 4 C in the main text) [4]. As both models generate double exponentials, we relate them to each other by equating the time constant and the corresponding relative probabilistic weight. Since we are interested in describing transcriptional bursts, we here consider the limit where the two-state model gives clearly separated bursts, $k_{on} \ll k_{off} + k_{tr}$. In this limit we have, to leading order in k_{on} ,

$$p = \frac{k_{\text{off}}}{k_{\text{off}} + k_{\text{tr}}}, \quad \tau_{\text{ip}} = \frac{1}{k_{\text{off}} + k_{\text{tr}}}, \quad \text{and} \quad \tau_{\text{tp}} = \frac{1}{k_{\text{on}}} \left(1 + \frac{k_{\text{off}}}{k_{\text{tr}}} \right). \quad (41)$$

The relations in Equation 41 can be inverted to give

$$k_{\text{tr}} = \frac{1-p}{\tau_{\text{ip}}}, \quad k_{\text{off}} = \frac{p}{\tau_{\text{ip}}}, \quad \text{and} \quad k_{\text{on}} = \frac{1}{(1-p)\tau_{\text{tp}}}. \quad (42)$$

For the bulk we can in principle calculate the corresponding two-state model using Equation 6, though this state is likely never reached. The physiologically more interesting situation is just after the initial pelotons have fully formed. The relative probability of a gap being between pelotons is then (see Equation 26)

$$p = p^{\text{in}} = \frac{1}{1 + k_{\text{in}}\tilde{\tau}}. \quad (43)$$

Once the initial pelotons have formed beyond x_p , we know that the average velocity is k_{tp} , and we have

$$\langle g \rangle_{\text{ip}}^* = \frac{k_{\text{tp}}}{k_{\text{ip}} - k_{\text{tp}}} \quad \Rightarrow \quad \tau_{\text{ip}}^* = \frac{1}{k_{\text{ip}} - k_{\text{tp}}} + \frac{\delta_m}{k_{\text{tp}}}. \quad (44)$$

The trans-peloton gaps can be written as (see supplementary Figure 2)

$$\begin{aligned} \langle T \rangle &= \tau_{\text{tp}}^* + \tau_{\text{ip}}^* (\langle n_p \rangle^* - 1) \\ \Rightarrow \quad \tau_{\text{tp}}^* &= \langle T \rangle - \tau_{\text{ip}}^* k_{\text{in}} \tilde{\tau} = \left(\frac{1}{k_{\text{in}}} + \frac{\delta_m}{k_{\text{tp}}} \right) + \left(\frac{1}{k_{\text{in}}} - \frac{1}{k_{\text{ip}} - k_{\text{tp}}} \right) k_{\text{in}} \tilde{\tau} \end{aligned} \quad (45)$$

Combining Equations 42-45 we have

$$k_{\text{tr}} = \frac{k_{\text{in}}\tilde{\tau}}{1 + k_{\text{in}}\tilde{\tau}} \frac{1}{\tau_{\text{ip}}^*}, \quad k_{\text{off}} = \frac{1}{1 + k_{\text{in}}\tilde{\tau}} \frac{1}{\tau_{\text{ip}}^*}, \quad \text{and} \quad k_{\text{on}} = \frac{1 + k_{\text{in}}\tilde{\tau}}{k_{\text{in}}\tilde{\tau}} \frac{1}{\tau_{\text{tp}}^*}. \quad (46)$$

In the main text we are interested in the case where motors are large and the initiation rate is low compared to motor stepping rates. In this limit we have

$$k_{\text{tr}} = \frac{k_{\text{in}}\tilde{\tau}}{1 + k_{\text{in}}\tilde{\tau}} \frac{k_{\text{tp}}}{\delta_m}, \quad k_{\text{off}} = \frac{1}{1 + k_{\text{in}}\tilde{\tau}} \frac{k_{\text{tp}}}{\delta_m}, \quad \text{and} \quad k_{\text{on}} = \frac{1}{\tilde{\tau} + \frac{k_{\text{in}}\tilde{\tau}}{1 + k_{\text{in}}\tilde{\tau}} \frac{\delta_m}{k_{\text{tp}}}} \approx \frac{1}{\tilde{\tau}}. \quad (47)$$

Here the approximate relation should be valid in the case of transcription through nucleosomes as here the roadblock are substantially larger than the motors, and $\tilde{\tau}$ is consequently substantially larger than δ_m/k_{tp} .

8 Monte Carlo Simulations

We validate our heuristic arguments using a random-sequential-update Monte Carlo scheme with fixed time step dt to simulate our model. During a Monte Carlo step on a lattice of size $L + 1$ there are $L + 1$ possible events: all the motors on the lattice can make a step forward, bound roadblocks can unbind, roadblocks in solution can bind to an empty lattice site, and a motor can bind at the start of the lattice. The time step is chosen small enough that the probability of any event occurring with rate k in a time dt can be approximated as kdt . In our simulation k_{ip} is the fastest rate, and we choose $k_{\text{ip}}dt = 0.1$. We verified that our results are robust towards changes in dt . Without roadblocks, the time to equilibration for periodic systems scales with the system size as $L^{3/2}$ [1]. With roadblocks, the time to equilibration is expected to be larger due to the slow peloton dynamics. For the simulations with periodic boundary conditions we waited L^2/dt iterations for the system to reach steady state and let simulations run a total amount

of $10L^2/dt$ iterations, and checked that the peloton size did not change for longer equilibration times. The velocities presented in supplementary Figure 1 and Figure 5 in the main text are calculated by averaging over the instantaneous hopping rates of the motors.

References

- [1] R A Blythe and M R Evans. Nonequilibrium steady states of matrix-product form: a solver's guide. *Journal of Physics A: Mathematical and Theoretical*, 40(46):R333–R441, nov 2007.
- [2] O J O Loan, M R Evans, and M E Cates. Jamming transition in a homogeneous one-dimensional system : The bus route model. *Physical Review E*, 58(2):1404–1418, 1998.
- [3] C T MacDonald, J H Gibbs, and A C Pipkin. Kinetics of Biopolymerization on Nucleic Acid Templates. *Biopolymers*, 6:1–25, 1968.
- [4] Brian Munsky, Gregor Neuert, and Alexander van Oudenaarden. Using gene expression noise to understand gene regulation. *Science*, 336(6078):183–7, April 2012.