Stochastic strategies for survival: bacterial competence in Bacillus Subtilis

Sandra H. Dandach Mustafa Khammash

Abstract—Under stressful environments, organisms take actions that help them protect their DNA. An example of such actions is the stochastic switching that Bacillus subtilis undergoes, in which it goes from the vegetative state to a competent state. When in competence, the cell has an increased ability to bind and internalize exogenous DNA. This increases the chances of survival of a cell. Competence is nonetheless an expensive state for the cell to be in, so the decision to switch undergoes a very delicate regulation. A major player in controlling the switching of the cell is the ComK protein. ComK protein is a key regulator which activates hundreds of genes, including the genes encoding the DNA-uptake and recombination systems. In Bacillus subtilis, stress in the environment activates a sequence of chemical reactions that, driven by cellular noise, stochastically increases the level of ComK in some bacterial cells driving them from their original vegetative state into a competent state. In this work, we use the Finite State Projection (FSP) method to analyze stochastic biochemical events and to study the excitable dynamics responsible for competence in Bacillus Subtilis. We compute the probability with which Bacillus subtilis enters in competence. We also present a method to analyze the sensitivity of these stochastic events to various system parameters such as binding affinities, transcription rates, degradation rates, etc.

I. INTRODUCTION

Competence is a state that a bacterium can switch to in order to preserve its DNA under stressful conditions. It allows the cell, to bind and internalize transforming exogenous DNA. Under the same stressful environment, such as nutrient limitations, some cells enter competence while other cells commit irreversibly to sporulation. Entry in competence is a transient probabilistic event that facilitates copying of the exogenous DNA [1], [2]. It has been shown that among a group of cells only a randomly chosen fraction enters in competence [3], [4]. It is crucial to correctly account for the noise driving this stochastic event, in order to understand the underlying biological explanation. When in competence, the cells express a high concentration of the key regulator ComK, which activates hundreds of genes, including the genes encoding the DNA-uptake and recombination systems [5], [6], [7]. Competence is understood as a bistability pattern [8], [4] and the nonlinear system describing the competence regulatory circuit is an excitable dynamical system.

ComK is the main player responsible for the bistable response in competence development. In specific, Auto-activation of ComK is essential and can be sufficient to generate a bistable expression pattern [9], [10], [11], [12].

Although the gene regulatory circuit consists of many proteins, there are two main proteins that play a major role. A deterministic model driven by an additive noise to describe the dynamics of competence regulation was presented in [13]. We use the reduced order Stochastic Differential Equation model (SDE) presented in [13] to develop a discrete stochastic model for competence. When utilizing the CME to account for noise in analyzing gene regulatory circuits that involve two species or more, researchers usually use Monte-Carlo simulations to calculate the distribution of the first passage time (e.g. see [14] and references therein). We propose an alternative approach in this work that aggregates two different regions of the state space into absorbing states: one region represents the state of competence, the other region accounts for states that happen with a low probability. This technique is useful in analytically computing the distribution of the first passage time, by providing a way to deal with the infinite dimension of the state space over which the system evolves.

The contributions of this paper are two folds. First, it uses the Finite State Projection (FSP) method to obtain analytical expressions for the probabilities of biological phenomena where transient behaviors such as competence, which is the topic we chose to study here, occur. Second, it shows how to calculate sensitivities of the probabilities of passing to the transient state with respect to the system’s parameters.

This paper is organized as follows: In Section II we describe the chemical reactions and the deterministic model. In Section III we generate the Chemical Master Equation (CME) of our proposed discrete stochastic model. The CME characterizes the evolution of the probability density of the different discrete states. We simulate it using the Stochastic Simulation Algorithm (SSA), and show how the solution can be approximated using the Finite State Projection method (FSP) [15]. In Section IV we use the analytical solution to the problem obtained using FSP to conduct a sensitivity analysis of entrance in competence with respect to the reaction rate parameters. In Section V we give concluding remarks.

II. CHEMICAL REACTIONS DESCRIBING THE GENE REGULATORY CIRCUIT

Competence is a physiological state that enables cells to bind and internalize transforming DNA. This state is accompanied by blockage of the essential cell’s functions, and since this state is driven by the transcriptional factor ComK, it is no surprise that ComK synthesis is subject to a number of finely tuned regulatory circuits [16]. The gene regulatory model for competence has been presented and described in [13]. Entrance of a cell in competence
is controlled by a set of molecular interactions. The transcriptional factor ComK activates its own expression through positive feedback. Bound to MecA, ComK becomes inactive. In stressful environments, the level of ComS is high and that favors entrance into competence, since ComS competes with ComK to bind to MecA. Inhibition of the binding of ComK to MecA by competitive binding with MecA-ComS allows a higher number of free ComK molecules to be present, which finally triggers the positive feedback that further raises the number of ComK molecules driving the cell in competence. This rise in the number of ComK is specific to competence. Once the number of ComK molecules reaches a certain level, it acts as an inhibitor for ComS through negative feedback. ComS now will be released from MecA-ComS freeing more MecA molecules, which in turn inhibits ComK activity which now can bind to free MecA. The level of ComK decreases until the cell eventually exits the state of competence. The above mentioned molecular interactions are described by the following chemical reactions [13].

\[
\begin{align*}
\text{MecA} + \text{ComK} \xrightarrow{\gamma_a} \text{MecA} - \text{ComK} \xrightarrow{\gamma_1} \text{MecA} \\
\text{MecA} + \text{ComS} \xrightarrow{\gamma_b} \text{MecA} - \text{ComS} \xrightarrow{\gamma_2} \text{MecA}.
\end{align*}
\]

The rate equations describing the dynamics of the molecular reactions between the 5 species model are the following:

\[
\begin{align*}
\frac{dK}{dt} &= \alpha_k + \frac{\beta_k K^n}{k^n + K^n} - \gamma_a M_f K + \gamma_{-a} M_K \quad (1) \\
\frac{dS}{dt} &= \frac{\beta_a}{1 + (\frac{K}{k_s})^p} - \gamma_b M_f S + \gamma_{-b} M_S \quad (2) \\
\frac{dM_K}{dt} &= -(\gamma_{-a} + \gamma_1) M_K + \gamma_a M_f K \quad (3) \\
\frac{dM_S}{dt} &= -(\gamma_{-b} + \gamma_2) M_S + \gamma_b M_f S, \quad (4)
\end{align*}
\]

where \(K, S, M_f, M_K\) and \(M_S\) are the concentrations of ComK,ComS, MecA, MecA-ComK and MecA-ComS respectively. We give in table II the values and the description of each of the parameters in (1)-(4). If one further assumes that the reactions of degradation of \(M_K\) and \(M_S\) are much faster than the other reactions, \(M_K\) and \(M_S\) can then be eliminated through time scale separation and the conservatism law [13][17]:

\[M_f + M_K + M_S = \text{Constant},\]

giving the following reduced model for the dynamics of competence:

\[
\begin{align*}
\frac{dK}{dt} &= \alpha_k + \frac{\beta_k K^n}{k^n + K^n} - \frac{\delta_k K}{1 + \frac{K}{K_s} + \frac{S}{T_s}} \quad (5) \\
\frac{dS}{dt} &= \frac{\beta_a}{1 + (\frac{K}{k_s})^p} - \frac{\delta_a S}{1 + \frac{K}{K_s} + \frac{S}{T_s}}, \quad (6)
\end{align*}
\]

where

\[
\Gamma_k = \frac{\gamma_{-a} + \gamma_1}{\gamma_a}, \quad \Gamma_s = \frac{\gamma_{-b} + \gamma_2}{\gamma_b},
\]

and

\[
\delta_k = \frac{\gamma_1 M_{total}}{\Gamma_k}, \quad \delta_s = \frac{\gamma_2 M_{total}}{\Gamma_s}.
\]

Cells have shown to stochastically enter in competence. In order to properly model this stochastic cell behavior, we need to properly account for the effect of noise on the dynamics of competence. A lot of the times, stochasticity is accounted for by an additive noise. In their analysis, Suel et al. [13] add white gaussian noise terms in equation (6). This drives the excitable dynamical system (5)-(6) into long excursions when the noise magnitude is large enough. Long excursions imply high levels of ComK, a characteristic of competence. The problem with this approach is that reaching a competent state is highly dependent on the magnitude of the additive noise. Suel et al. [13] studied the deterministic dynamical system in (5)-(6) and showed that it has three fixed points: only one fixed point is locally stable; it corresponds to the vegetative state of the cell. The remaining two fixed points are unstable. One is an unstable spiral and corresponds to an intermediate value of the number of ComK molecules. The other is an unstable spiral and corresponds to a high value of ComK, in other words to the competent state. If the cell has an initial number of molecules of ComK and ComS close to the number of molecules of the stable fixed point (vegetative state), the number of molecules remain in the vicinity of that point and the cell stays in a vegetative state. If on the other hand, the number of molecules is driven beyond a threshold (intermediate value of ComK), the dynamical system in (5)-(6) gets excited and the system goes in a long excursion where the number of molecules of ComK reaches a high level leading the cell to a state of competence. In this work we study the probability with which a cell takes these long excursions. We analyze the stochastic behavior of the dynamics of the competence regulatory circuit taking into account the internal noise in the environment of the cell. To do so, we model the stochasticity in the chemical reactions using the CME. We look at the problem at the molecular level and propose 4 reactions to model (5)-(6). The four reactions are:

\[
\phi \xrightarrow{k_1} K, \quad \phi \xrightarrow{k_3} S, \quad (7)
\]

with the following reaction rates:

\[
k_1 = \alpha_k + \frac{\beta_k K^n}{k^n + K^n}, \quad k_2 = \frac{\delta_k}{1 + \frac{K}{K_s} + \frac{S}{T_s}}, \quad k_3 = \frac{\beta_a}{1 + (\frac{K}{k_s})^p}, \quad k_4 = \frac{\delta_a}{1 + \frac{K}{K_s} + \frac{S}{T_s}}.
\]

These reactions will serve as the starting point for developing and simulating a discrete stochastic model for competence in the next section.

III. CHEMICAL MASTER EQUATION: DESCRIPTION AND ANALYSIS TOOLS

The CME describes the evolution of the probability density vector describing the number of molecules of ComK and
ComS. Obtaining a solution for the CME allows calculating the probability of entering into competence. Starting from a number of molecules \((x_0, y_0)\), the probability of being at \((x, y)\) molecules at time \(t\) has the following dynamics:

\[
\dot{p}(x; t) = -p(x; t) \sum_{\mu=1}^{4} a_{\mu}(x) + \sum_{\mu=1}^{4} p(x - \nu_{\mu}; t) a_{\mu}(x - \nu_{\mu}),
\]

where \(\nu_{\mu}\) is the stoichiometry vector and it represents the change that reaction \(\mu\) will have on the number of molecules of each of the species. Reaction 1 increases ComK by one molecule and leaves the number of molecules of ComS unchanged so the propensity vector \(\mu_1\) is \((1, 0)^T\). Written in vector form, the CME becomes

\[
\dot{\mathbf{p}}(x; t) = \begin{bmatrix}
- \sum_{\mu=1}^{4} a_{\mu}(x) & a_1(x - \nu_1) & a_2(x - \nu_2) & \cdots \\
& a_3(x - \nu_2) & a_4(x - \nu_4)
\end{bmatrix}
\begin{bmatrix}
p(x; t) \\
p(x - \nu_1; t) \\
p(x - \nu_2; t) \\
p(x - \nu_4; t)
\end{bmatrix},
\]

where 4 corresponds to the number of reactions that the species would go through. Let \(\mathbf{X} = (x_1, x_2, \ldots)^T\) be a vector of the possible states of the system. Let \(P(\mathbf{X}, t)\) be the corresponding vector of probabilities of the states in \(\mathbf{X}\) computed at time \(t\). \(P(\mathbf{X}, t)\) evolves according to the equation

\[
\dot{P}(\mathbf{X}; t) = \mathbf{A} \cdot P(\mathbf{X}; t).
\]

In general, \(\mathbf{X}\) may be infinite, resulting in an infinite dimensional system.

### A. Stochastic Simulation Algorithm

Getting exact values for the solution to the CME is not generally an easy task. In this part subsection we use the SSA to simulate (9). The SSA is a Monte-Carlo based algorithm that generates sample paths for the underlying stochastic process [18]. We applied SSA to both the full model presented in (1)-(4), as well as the reduced model presented in (7). When running numerical simulations, we set the initial number of molecules to be \((\text{ComK}, \text{ComS}) = (25, 225)\). All runs simulate 40 hours of molecular reactions. The initial starting point roughly corresponds to the mean steady state values of the reduced model. The cell is considered competent, when the number of ComK molecules crosses the threshold \(\text{ComK} = 80\). We study the probability with which the number of molecules cross this threshold.

Figure 1 shows one SSA run where both ComK and ComS concentrations were plotted. Competence is clear in this case, and is detected by both the high level of ComK and the negative correlation between ComK and ComS. The negative correlation corresponds to the negative feedback between ComK and ComS. The threshold ComK corresponds to an approximate probability 0.341. Using SSA to get an estimate of the probability of entering into competence is easy to implement. However, a large number of simulations is required to get a reasonable estimate of the probability. An alternative method is introduced in the next section.

### B. Finite State Projection

In the CME presented in (9), the probability density vector evolves on an infinite lattice and results therefore in an infinite dimensional system (see Figure 2). In [15], Munsky et al. introduces a method to compute an analytical approximate for the probability vector to take values in a given region of the subspace. For our purposes, instead of dealing with the infinite lattice, one can think of aggregating a suitable portion of the state space into one state. This state corresponds to levels of the ComK and ComS proteins and describes a state that the cell is in. We retain all the chemical reactions in the original system and allow the system to change states reversibly inside the original system. The transition to the aggregated region are kept, while the transitions allowing return from the aggregated region are deleted. This makes the new state an absorbing state and we are interested in calculating the probability with which a cell visits the absorbing state within a determined reaction time (see Figure 2 for illustration). The finite state projection method gives the probability of being at any of the states inside the specified region at any point in time. In this problem we are interested in finding the probability with which ComK crosses a certain threshold \((\text{ComK} = 80)\). This corresponds to the probability with which a cell enters in competence. The sum of the probabilities of being at any of the states has to equal one at all times. Moreover the sum of the probabilities of being at any state inside the projected region without ever leaving the region, and that of the probability of leaving the region once within a time \(T\) should also sum to one. These properties, make FSP a very well suited numerical method to solve our problem.

The probability vector at time \(T\) is given as in (10) by

\[
P(\mathbf{X}, T) = \exp(A_{\inf} T) P(\mathbf{X}, 0),
\]

where \(A_{\inf}\) is an infinite matrix and \(P(\mathbf{X}, 0)\) is the initial distribution of the probabilities. We aggregate into one absorbing state, \(X_{exit}\), all the states outside the projection.
region. We can then project the infinite system in (11) into
the following finite system:

\[ P(X_j, T) = \exp(AT)P(X_j, 0). \]

Here, \( A \) is a finite matrix, and \( X_j \) is a finite vector
of projected states. The \( A \) matrix is built as follows

\[ A_{ji} = \begin{cases} \delta_{ij} \sum_{i=1}^{4} a_i(X_i), & \text{for } i = j, \\ a_i(X_i), & \text{for all } j \text{ such that } X_j = X_i + \nu_{i}, \\ 0, & \text{otherwise}. \end{cases} \]

Here, \( \mu, \nu_{i} \) and \( a_i \) are the terms appearing in (8). If \( x(t) \)
denotes the underlying stochastic process, \( P(X_j, T) \) gives
the probability of \( x(t) \) being in any of the states listed in
\( X_j \) during the time \([0, T]\), conditioned on the event of never
leaving the inside region for any time \( t \in [0, T] \). We can rewrite
the probability as the conditional probability

\[ P(X_j, T) | x(t \leq T) \neq X_{exit}) = \exp(AT)P(X_j, 0), \]

where \( X_{exit} \) accounts for the aggregation of the outside
region. The probability of being inside the region \( X_j \) without
ever leaving it during the interval \([0, T]\) and the probability of
visiting \( X_{exit} \) once should sum to one (\( X_{exit} \) is an absorbing
state). Therefore

\[ P(x(t \leq T) = X_{exit}) = 1 - \exp(AT)P(X_j, 0). \]

The equation above gives the probability of exiting the
specified region at least once within a time \( T \). The region
being projected into an absorbing state, corresponds to
\( ComK \geq 80 \). The probability with which the state of the
system reaches \( X_{exit} \) corresponds then to the probability
with which the cell reaches a state of competence. Denoting
\( P(X_j, t) \) by \( P(t) \) and \( P(X_{exit}, t) \) by \( P_{exit}(t) \) we can see
that the probability of competence at time \( t \), \( P_{exit}(t) \), is given
by

\[ \begin{pmatrix} \dot{P} \\ \dot{P}_{exit} \end{pmatrix} = \begin{pmatrix} A & 0 \\ b & 0 \end{pmatrix} \begin{pmatrix} P \\ P_{exit} \end{pmatrix}, \]

where \( b \) is chosen so that the columns of the state transition
matrix add up exactly to zero. Using the above formulation,
we find that the probability of entering in competence at least
once in 40 hours is 0.3339 (compare to 0.341 estimated from
10000 runs of SSA). The slight difference between the SSA
and the FSP results comes from the fact that 10000 runs of
SSA might not be enough to get a proper estimate on the
probability.

C. Reduced model analysis with SSA and FSP

We present in this section a comparison between the full
and reduced models for the competence gene regulatory

circuit. The full model was presented in Equations (1)-(4),
while the reduced model was presented in Equations (5)-(6).
Deterministic characterizations of both models were
presented in [13]. We simulate both models using SSA and
compare the probability of entering in competence as the
parameters presented in Table II were changed. We show
in Figures 3 and 4 the effect of the saturating expression
of ComK and of the repressed expression of ComS on
the probability with which a cell enters in competence.
We also compare in these figures, the results given by the SSA
and the FSP method, when applied to the reduced model.
Our numerical results show that SSA and FSP give similar
results for the reduced model. Finally, our numerical results
shown in these figures also confirm that the reduced model
approximates the full model very well. The plots in Figures 3
and 4 show simulations from the full model using SSA
(black), the reduced model using FSP (blue) and the reduced
model using SSA (red). SSA results were generated by
averaging over 10,000 runs. For each data point, the error
indicated by the errorbar is no larger than \( \pm 0.025 \) with a
certainty no smaller than 0.999. This is to be compared to
an upper bound of \( 10^{-3} \) when using FSP.

We next show how FSP can be used to conduct a sensiti-

Fig. 3: This figure shows the probability of entering in
competence when \( \beta_k \) is varied.
Reduced and full models using SSA and FSP

Reduced and full models using SSA and FSP

β_s (unrepressed ... that are most likely to be taken can be computed using similar approaches to the one discussed in this paper.

Looking at the sensitivity of vector with indicating the dependence of the stochastic switching in bac-

Realizing that $P = AP$, suppose we are interested in looking at the sensitivity of $P$ with respect to a parameter $\lambda$, which represents one of the parameters presented in Table II. The $jth$ entry in $P$ is given by $p = e_jP$, where $e_j$ is an $1 \times n$ vector with 1 in the $jth$ entry and zero everywhere else. We then have from equation (9) that $p = e_jP = e_jAP$. Letting $\lambda$ represent any of the parameters $\{\alpha_k, \beta_k, \beta_s, \delta_k, \delta_s\}$, and using the fact that $\frac{dP}{d\lambda} = \frac{dp}{d\lambda}$, we get $\frac{dp}{d\lambda} = e_jA\lambda + e_jAP\lambda$ where $P\lambda$ is defined to be $\frac{dP}{d\lambda}$. Similar equations hold for $P_{\text{exit}}$. Hence we have the following system:

$$
\begin{pmatrix}
P
P_{\text{exit}}
P_{\lambda}
P_{\text{exit},\lambda}
\end{pmatrix}
= 
\begin{pmatrix}
A & 0 & 0 & 0
b & 0 & 0 & 0
A_\lambda & 0 & A & 0
b_\lambda & 0 & b & 0
\end{pmatrix}
\begin{pmatrix}
P
P_{\text{exit}}
P_{\lambda}
P_{\text{exit},\lambda}
\end{pmatrix}.
\tag{13}
$$

Solving the above linear system, we obtained the sensitivity of the exit probability to all the parameters, evaluated at the nominal values given in Table II.

The results for normalized sensitivities are shown in table I. We next compute the same terms numerically, according to the following formula: $S = \frac{P(\lambda_0 + \delta \lambda) - P(\lambda_0)}{\delta \lambda}$, where $\lambda_0$ is the nominal value of the parameter, and

$$
\lambda = \lambda_0 + \delta \lambda, \quad \delta \lambda = 0.001 \times \frac{\lambda_0}{2^l}, \quad l = 0, ..., 10. \tag{14}
$$

To summarize, the sensitivity presented in table I are calculated in two different ways:

**Analytical derivative:** In the first method we solve the double order system in (13). This results in more accurate answers but is more computationally expensive.

**Finite difference:** In the second method, we use the solutions for the original system describing the evolution of the probabilities of the states presented in (12) in addition to the numerical approximation method presented in (14) with $l = 10$. This method is less accurate than the first but is considerably faster to implement.

V. Conclusion

Competence is an exhaustive state for the cell, nevertheless it is occasionally necessary. The plot in figure 3 shows that an increase in $\beta_k$ increases the probability of entering in competence. Our calculations also show that an increase in $\delta_s$ has a similar effect to an increase in $\delta_k$. Remember that $\delta_k$ is the degradation rate of ComK, and $\delta_s$ is the degradation rate of ComS. Also recall that whenever the number of ComS molecules is smaller, more MecA molecules will be available for ComK binding. And clearly a degradation of ComK will lead to a decrease in the number of ComK molecules. So both a low number of ComS molecules and the degradation of ComK lead to driving the cell back to its vegetative state and decrease the probability with which it enters in competence. This explains the similarity in the effect of $\delta_k$ and $\delta_s$. Although Figures 3 and 4 show that ComK and ComS have similar roles in driving a cell into and back from competence, Table I suggests that changes in ComS affected by the values of $(\beta_s, \delta_s)$ affect the probability of entering and staying in competence more than changes in ComK affected by the values of $(\beta_k, \delta_k)$. This leads to the expectation that the genetic circuits controlling ComS levels need to be much more sophisticated and complex than those regulating ComK in order to keep ComS concentration at specific values.

In this paper we developed a discrete stochastic model for competence in *Bacillus subtilis*. We performed simulations of the model using Monte Carlo based SSA and verified that the reduced order model gave a valid approximation of the full model. We then applied the recently developed FSP method to the reduced model and computed the probability of competence, where competence has been defined in terms of the trajectory leaving a certain region of the state space. Having the analytical solution, we were able to conduct sensitivity analysis of entrance in competence with respect to the model parameters.

This paper presents numerical methods that are applicable to many biological systems that exhibit a transient behavior. In summary these methods were shown to be very useful in studying the competence network in a cell, and in answering questions about the computation of the probabilities of events in a biological network. They were also useful in studying sensitivities of these events when expression for proteins, their degradation rate, repression rates or activation rates are changed. Finally, the methods introduced in this paper showed how expected times for trajectories for return from transient states can be calculated. Many other terms characterizing different transient physiological behaviors, such as the number of molecules that are most likely to enter in the transient states, and the return trajectories that are most likely to be taken can be computed using similar approaches to the one discussed in this paper.
TABLE I: Sensitivity of the system with respect to various system parameters, when the parameters are set to their nominal values as presented in [13].

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>α_k</td>
<td>4.9931</td>
</tr>
<tr>
<td>β_k</td>
<td>8.4417</td>
</tr>
<tr>
<td>δ_k</td>
<td>43.0106</td>
</tr>
<tr>
<td>δ_s</td>
<td>-11.9632</td>
</tr>
<tr>
<td>δ_k</td>
<td>-43.1321</td>
</tr>
</tbody>
</table>

TABLE II: Parameter values as given in [13].

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>α_k</td>
<td>0.0028 nM/s</td>
</tr>
<tr>
<td>β_k</td>
<td>0.049 nM/s</td>
</tr>
<tr>
<td>δ_k</td>
<td>2.50 nM</td>
</tr>
<tr>
<td>δ_s</td>
<td>5.00 nM</td>
</tr>
<tr>
<td>δ_k</td>
<td>110 nM</td>
</tr>
<tr>
<td>Γ_k</td>
<td>500 nM</td>
</tr>
<tr>
<td>n</td>
<td>2.0</td>
</tr>
<tr>
<td>p</td>
<td>0.0014</td>
</tr>
</tbody>
</table>

REFERENCES


