A MATHEMATICAL AND COMPUTATIONAL APPROACH FOR INTEGRATING THE MAJOR SOURCES OF CELL POPULATION HETEROGENEITY

Michail Stamatakis and Kyriacos Zygourakis*
Department of Chemical and Biomolecular Engineering MS-362, Rice University, Houston, TX 77005, USA

Abstract

Several approaches have been used in the past to model heterogeneity in bacterial cell populations, with each approach focusing on different sources of heterogeneity. However, a holistic approach that integrates all the major sources into a generic framework applicable to cell populations is still lacking. We present here the mathematical formulation of a Master Equation for a cell population that considers the major sources of heterogeneity, namely stochasticity in reaction, division, and DNA duplication. The formulation also considers cell growth and accounts for the discrete nature of the molecular contents. We further develop a Monte Carlo algorithm for the simulation of the stochastic processes considered here. Using this approach, we finally demonstrate the effect of each source of heterogeneity on the overall phenotypic variability for the two-promoter system used by Elowitz et al. (2002) to experimentally quantify intrinsic versus extrinsic noise.

Keywords

stochastic, intrinsic noise, extrinsic noise, heterogeneity, cell division, DNA duplication.

Introduction

Until the 1990's, the biological paradigms and many state-of-the-art modeling frameworks neglected cell population heterogeneity. The fundamental assumption of such theoretical investigations was that all cells behave like the average cell and, thus, their behavior can be described by continuum models consisting of Ordinary Differential Equations (ODEs). This is not what is observed *in vivo*, however, since cell populations are inherently heterogeneous (Davidson and Surette 2008). More importantly, even if one is interested only in the average dynamics, it turns out that use of continuum models (Fredrickson 1976) that neglect heterogeneity will result in incorrect predictions (McAdams and Arkin 1998). Hence, one has to use models that explicitly account for the heterogeneous nature of the cell population.

None of the earlier studies, however, has adopted a holistic approach that integrates all the major sources of heterogeneity into a general cell population model. The population balance framework introduced by Fredrickson et al. (1967) to model cell population dynamics takes into account stochasticity in division times and potentially unequal partitioning. However, this approach treats intracellular reaction events deterministically, does not consider intrinsic noise and only involves protein levels, since it does not treat DNA species. The Chemical Master Equation simulated by the Gillespie algorithm (Gillespie 1976) takes into account intrinsic noise, and some variants of this algorithm also incorporate deterministic growth and stochastic partitioning (Lu et al. 2004). However, these algorithms do not consider variability in division or DNA duplication times and can only describe single cells, not

* To whom all correspondence should be addressed

cell populations. The ensemble methods (Domach and Shuler 1984; Henson 2003) simulate populations in which variability is attributed to different initial conditions or kinetic rates. These methods do not follow the dynamics of the cell population and do not account for intrinsic noise. Finally, an algorithm proposed recently by Mantzaris (2007) takes into account intrinsic noise at the cell population level as well as stochastic partitioning. But, it models intrinsic noise using Stochastic Differential Equations (SDEs) that are only valid at limiting cases of large systems. Furthermore, this approach does not explicitly treat DNA species.

This study presents a framework that accounts for all the major sources of heterogeneity, namely stochasticity in reaction, division, and DNA duplication. The mathematical formulation applies to the cell population level, takes into account cell growth, and respects the discrete nature of the molecular contents and cell numbers.

Mathematical Formulation

We assume that each cell can be completely described by a state vector that contains information about the chemical content of the cell and its morphometric characteristics such as length, membrane area or volume. This study will consider only one morphometric characteristic: volume. Thus, the state vector of the cell has length n+1, with *n* components corresponding to the species copy numbers and one component for the volume.

We further assume that reaction dynamics can be captured by a reaction network of the generic form:

$$\left\{\sum_{i=1}^{n}\alpha_{ij}\,\mathrm{S_i}\ \xrightarrow{k_j}\ \sum_{i=1}^{n}\beta_{ij}\,\mathrm{S_i}\right\}_{j=1}^{m} \tag{1}$$

where the species vector is:

$$\mathbf{S}=\underbrace{\{\mathrm{S}_1,\mathrm{S}_2,\ldots\mathrm{S}_n}_{\text{non-chrom. DNA}},\underbrace{\mathrm{S}_{n+1},\mathrm{S}_{n+2},\ldots,\mathrm{S}_{n+s_1}}_{\substack{\text{chrom. DNA species 1}\\ \text{in its various states}}},\ldots\ \mathrm{S}_N\} \tag{2}$$

$$N=n+\sum_{i=1}^{d}s_i$$

where n is the number of non-chromosomal DNA species and d the number of chromosomal DNA species. The necessity for discriminating between chromosomal and non-chromosomal species comes from the fact that, upon division, chromosomal DNA species are partitioned equally in the two daughters. This is not generally true, however, for the other species. Furthermore, each of the DNA species $i = 1,\ldots,d$ may exist in s_i states. For example an operator may exist in three states: the free state O, the repressed state with one repressor molecule bound RO, or the repressed state with two repressor molecules bound R_2O. Thus for this case, $s_1 = 3$ and $(S_{n+1}, S_{n+2}, S_{n+3}) = (O, RO, R_2O)$.

During duplication the chromosomal DNA species are doubled. We assume that the newly produced chromosomal DNA species i exists in a basal state η_i. For example, in the aforementioned example of the operator existing in the free and the two bounded states, the basal state will be the free state. Then the production of the new DNA will schematically be:

$$\left\{\sum_{j=0}^{s_i-1}\mathrm{S}_{\eta_i+j}\longrightarrow\ \mathrm{S}_{\eta_i}\right\}_{i=1}^{d}\quad\text{where:}\qquad \eta_i=n+\sum_{j=1}^{i-1}s_j+1 \tag{3}$$

Single cell growth is modeled as a deterministic process:

$$\frac{dV}{dt}\ =\ g\left(\mathbf{X},V\right) \tag{4}$$

At the cell population level, consider a population that has ν cells existing in states $(\mathbf{X_1}, V_1), \ldots, (\mathbf{X_\nu}, V_\nu)$. $\mathbf{X_i}$ is a vector with the number of molecules per species and V_i is the volume of the cell. Then the dynamics of the probability of finding such a population at time t, will be governed by the following Master Equation:

$$\begin{aligned}\frac{\partial J_\nu}{\partial t}\ =\ &\sum_{\varsigma=1}^{\nu}\sum_{j=1}^{m}\left[\mathbb{E}_{\mathbf{X}_\varsigma}^{-\boldsymbol{\nu}_j}\left(a_j\left(\mathbf{X}_\varsigma,V_\varsigma\right)\cdot J_\nu\right)-a_j\left(\mathbf{X}_\varsigma,V_\varsigma\right)\cdot J_\nu\right]\\ &+\sum_{\varsigma=1}^{\nu}\left[\mathbb{E}_{\mathbf{X}_\varsigma}^{-\boldsymbol{\nu}_s}\left(a_s\left(\mathbf{X}_\varsigma,V_\varsigma\right)\cdot J_\nu\right)-a_s\left(\mathbf{X}_\varsigma,V_\varsigma\right)\cdot J_\nu\right]\\ &-\sum_{\varsigma=1}^{\nu}\frac{\partial}{\partial V_\varsigma}\left[g\left(\mathbf{X}_\varsigma,V_\varsigma\right)\cdot J_\nu\right]\\ &+2\cdot\sum_{\varsigma=1}^{\nu-1}\sum_{\zeta=\varsigma+1}^{\nu}\Big[a_d\left(\mathbf{X}_\varsigma+\mathbf{X}_\zeta,V_\varsigma+V_\zeta\right)\\ &\cdot h\left(\mathbf{X}_\varsigma,V_\varsigma\middle|\mathbf{X}_\varsigma+\mathbf{X}_\zeta,V_\varsigma+V_\zeta\right)\\ &\cdot J_{\nu-1}\left(\left(\mathbf{X_1},V_1\right),\ldots,\left(\mathbf{X}_\varsigma+\mathbf{X}_\zeta,V_\varsigma+V_\zeta\right),\ldots,\right.\\ &\left.\left(\mathbf{X}_{\zeta-1},V_{\zeta-1}\right),\left(\mathbf{X}_{\zeta+1},V_{\zeta+1}\right),\ldots,\left(\mathbf{X_\nu},V_\nu\right);t\right)\Big]\\ &-\sum_{\varsigma=1}^{\nu}a_d\left(\mathbf{X}_\varsigma,V_\varsigma\right)\cdot J_\nu\end{aligned} \tag{5}$$

where a_j and $\boldsymbol{\nu_j}$ are respectively the propensity function for reaction j and the vector of change, a_s and $\boldsymbol{\nu_s}$ are the analogous quantities for DNA duplication (synthesis), g is the single cell growth rate, a_d is the division propensity and $h(\mathbf{x}|\mathbf{y})$ is the partitioning probability density function giving the probability of obtaining a daughter cell of state $\mathbf{x}$ after division of a mother with state $\mathbf{y}$. Furthermore:

$$J_\nu=J_\nu\left(\left(\mathbf{X_1},V_1\right),\ldots,\left(\mathbf{X_i},V_i\right),\ldots,\left(\mathbf{X_\nu},V_\nu\right);t\right) \tag{6}$$

and we have made use of the step operator defined as follows (van Kampen 1992):

$$\mathbb{E}_m^p f(k,l,m,n,\ldots) = f(k,l,m+p,n,\ldots) \quad (7)$$

Simulation Algorithm

We will now outline a computational algorithm that can simulate stochastic paths of Eq. (5) given a reaction network, single cell growth rate, DNA duplication and division propensities, and a partitioning mechanism.

- For every cell in the population create random times for reaction division, duplication.
- For $t < t_{max}$
  - Determine the next event and the affected cell.
  - Simulate the first event in the appropriate cell (additional samplings may be necessary).
  - Update X, V of all cells and time t.
  - For the affected cell(s), update or create random times for reaction division and duplication.

Two-Promoter System

We will now use this algorithm to simulate a genetic network consists of two genes under the influence of two identical repressible promoters. Such a genetic network was used by Elowitz et. al (2002) to decompose the extrinsic and intrinsic contributions of noise to the overall single cell noise. In particular, two GFP variants, a yellow (YFP) and a cyan (CFP), were cloned into the *E. coli* chromosome. Expression of both proteins is driven from identical Lac repressible promoters and the fluorescence intensity of both variants is approximately the same.

Thus, measurements of the fluorescence of the cells in the yellow and cyan channels can reveal the effects of intrinsic and the extrinsic noise. Specifically, differences in the fluorescence intensity of the same cell measured by the two channels originate from intrinsic noise, while differences between distinct cells are the result of extrinsic noise. This decomposition of noise to extrinsic and intrinsic components is rather phenomenological and based on how the protein contents of identical genes correlate.

Using our framework, however, we can identify the contribution of fundamental biological mechanisms on the extrinsic or the intrinsic noise components. This paper will focus on the non-repressed system.

All Sources of Heterogeneity Present

Figure 1a shows the normalized Yfp content versus the normalized Cfp content in a plot similar to that used by Elowitz et al. (2002). Each point in the plot corresponds to one cell of the population.

Table 1. Chemical Species Notation

Symbol	Species denoted
RP	RNA polymerase
RB	ribosome
O_{Yfp}	free operator of *yfp* gene
R_{Yfp}	*yfp* mRNA
Yfp	Yfp protein molecule
O_{Cfp}	free operator of c*fp* gene
R_{Cfp}	*cfp* mRNA
Cfp	Cfp protein molecule
$\varnothing$	Generic source or sink

Table 2. Chemical Reactions and Propensities

(i)	$\varnothing \xrightarrow{k_1} RP$	$k_1 \cdot V_{E.coli} \cdot N_A$
(ii)	$\varnothing \xrightarrow{k_2} RB$	$k_2 \cdot V_{E.coli} \cdot N_A$
(iii)	$O_{yfp} + RP \xrightarrow{k_3} O_{yfp} + RP + R_{yfp}$	$\frac{k_3}{V_{E.coli} \cdot N_A} \cdot O_{yfp} \cdot RP$
(iv)	$R_{yfp} + RB \xrightarrow{k_4} R_{yfp} + RB + Yfp$	$\frac{k_4}{V_{E.coli} \cdot N_A} \cdot R_{yfp} \cdot RB$
(v)	$O_{cfp} + RP \xrightarrow{k_5} O_{cfp} + RP + R_{cfp}$	$\frac{k_5}{V_{E.coli} \cdot N_A} \cdot O_{cfp} \cdot RP$
(vi)	$R_{cfp} + RB \xrightarrow{k_6} R_{cfp} + RB + Cfp$	$\frac{k_6}{V_{E.coli} \cdot N_A} \cdot R_{cfp} \cdot RB$
(vii)	$RP \xrightarrow{k_7} \varnothing$	$k_7 \cdot RB$
(viii)	$RB \xrightarrow{k_8} \varnothing$	$k_8 \cdot RP$
(ix)	$R_{yfp} \xrightarrow{k_9} \varnothing$	$k_9 \cdot R_{yfp}$
(x)	$Yfp \xrightarrow{k_{10}} \varnothing$	$k_{10} \cdot Yfp$
(xi)	$R_{cfp} \xrightarrow{k_{11}} \varnothing$	$k_{11} \cdot R_{cfp}$
(xii)	$Cfp \xrightarrow{k_{12}} \varnothing$	$k_{12} \cdot Cfp$

The observed scatter of the points results from stochasticity, which generates heterogeneity at the population level. For this simulation, all sources of noise that can be captured with our model are present.

More specifically, transcriptional and translational stochasticity is significant due to the low copy numbers of mRNA and protein. These are the intrinsic noise sources and contribute to the spread of points far from the diagonal Cfp = Yfp.

Furthermore, the stochasticity in DNA duplication and division, as well as the fluctuations in the contents of RNA polymerase and ribosomes constitute the extrinsic noise sources and contribute to the elongation of the ellipsoid along the diagonal Cfp = Yfp.

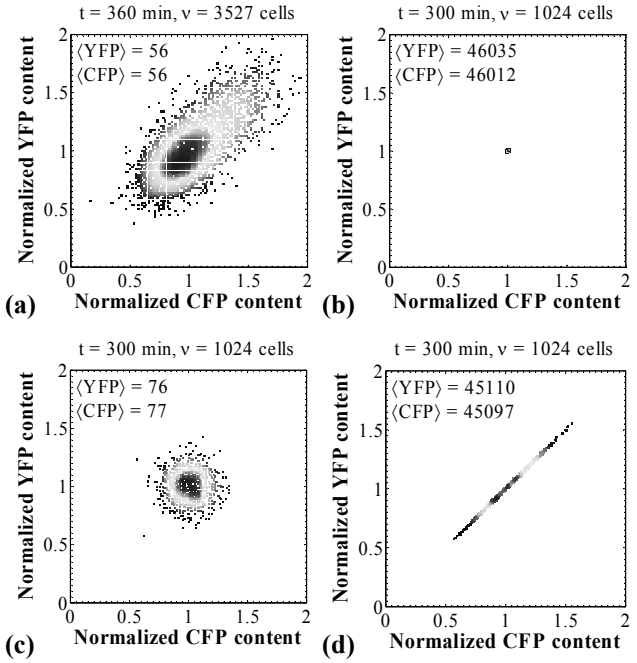

Figure 1: scatter-plots for the Yfp and Cfp: (a) all noise sources present; (b) no heterogeneity; (c) intrinsic noise; (d) extrinsic noise.

Homogeneous Cell Populations

When the population in homogeneous, all cells will behave identically which means that (i) the fluctuations of the species copy numbers due to reactions must be infinitesimally small; (ii) duplication and division events must occur in synchrony; and (iii) the cells must divide in a way that the two daughter cells have equal volumes and contents. Figure 1b shows such a case: all cells express the same amount of proteins and, thus, all the points in the scatter-plot are concentrated to the mean expression levels.

Only Extrinsic or Only Intrinsic Noise

Figure 1c shows the scatter plot graph for the case where only intrinsic noise is present. Stochasticity in the biomolecular reactions is significant, but DNA duplication and symmetric division events occur in synchrony. In this case, the points in the plot form a circular pattern, showing that the variability in the Cfp and Yfp content of a single cell is equal to the variability of Cfp (or Yfp) content between different cells of the cell population.

On the other hand, Figure 1d pertains to a case where only extrinsic noise is present. The latter is brought about by fluctuations in the RNA polymerase. Division is still symmetric in this case, and the duplication and division are synchronized. Moreover, intrinsic noise is negligible because the transcriptional rates of *cfp* and *yfp* are high, keeping mRNA and protein contents high. In this case, the points in the Cfp and Yfp graph are arranged along the line Cfp = Yfp, showing that in any single cell the Cfp and Yfp contents are identical, but there exists variability between different cells of the cell population.

Conclusions

We developed a mathematical and computational framework that can predict the phenotypic distributions observed in cell populations. The framework can be used to systematically study the contribution of fundamental biological mechanisms on the extrinsic and intrinsic noise components that give rise to highly heterogeneous bacterial populations. Therefore, computational studies using this approach can complement experiments in which a fundamental decomposition of noise sources is impossible.

Acknowledgments

The authors would like to gratefully acknowledge financial support from NIH/NIGMS (R01GM071888).

References

Davidson, C. J. and M. G. Surette (2008). "Individuality in Bacteria." Annual Review of Genetics **42**: 253-268.

Domach, M. M. and M. L. Shuler (1984). "A finite representation model for an asynchronous culture of *E. coli*." Biotechnology and Bioengineering **26**(8): 877-884.

Elowitz, M. B., A. J. Levine, E. D. Siggia and P. S. Swain (2002). "Stochastic Gene Expression in a Single Cell." Science **297**(5584): 1183-1186.

Fredrickson (1976). "Formulation of Structured Growth Models." Biotechnology and Bioengineering **XVIII**: 1481-1486.

Fredrickson, A. G., D. Ramkrishna and H. M. Tsuchyia (1967). "Statistics and dynamics of prokaryotic cell populations." Mathematical Biosciences **1**(3): 327-374.

Gillespie, D. T. (1976). "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions." Journal of Computational Physics **22**(4): 403-434.

Henson, M. A. (2003). "Dynamic modeling of microbial cell populations." Current Opinion in Biotechnology **14**(5): 460-467.

Lu, T., D. Volfson, L. Tsimring and J. Hasty (2004). "Cellular growth and division in the Gillespie algorithm." Systems Biology **1**(1): 121-128.

Mantzaris, N. V. (2007). "From Single-Cell Genetic Architecture to Cell Population Dynamics: Quantitatively Decomposing the Effects of Different Population Heterogeneity Sources for a Genetic Network with Positive Feedback Architecture." Biophysical Journal **92**: 4271-4288.

McAdams, H. H. and A. Arkin (1998). "Simulation of prokaryotic genetic circuits." Annual Review of Biophysics and Biomolecular Structure **27**: 199-224.

van Kampen, N. G. (1992). Stochastic processes in physics and chemistry. New York, Amsterdam, North-Holland-Personal-Library.