A Modelling Framework for Catalysing Progress in the Rod-Shaped Bacterial Cell Growth Discourse

Anonymous Author(s)
Affiliation
Address
email

Abstract

The fundamental question of how cells maintain their characteristic size remains open. Cell size measurements made through microscopic time-lapse imaging of microfluidic single-cell cultivations have seriously questioned classical cell growth models and are calling for newer, nuanced models that explain empirical findings better. Yet current models are limited in that they explain cellular growth either only in specific organisms and/or specific micro-environmental conditions. Together with the fact that tools for robust analysis of said time-lapse images are not widely available as yet, the previously mentioned point presents an opportunity to progress the cell growth and size homeostasis discourse through generative (probabilistic) modelling. Our contribution is a novel Model Framework for simulating microfluidic single-cell cultivations of rod-shaped bacteria with 36 different simulation modalities, each integrating dominant cell growth theories and generative modelling techniques. Our framework enables the simulation of diverse microscopic image sequences of the said class of single-cell cultivations as well as the generation of corresponding ground truths. More generally, our framework enables simulations of image sequences that imperfect camera and imaging conditions can produce, along with corresponding segmentation and tracking information. It thus enables the generation of datasets consisting of image sequence inputs and corresponding tabular labels, which can help develop robust machine image analysis networks applicable to real-world microfluidic experiments aimed at progressing the cell growth discourse. We demonstrate the usability of our framework through synthetic experiments and conclude by presenting its limitations as well as opportunities for further work.

1 Introduction

Rod-shaped bacteria, particularly E. coli, are model organisms of tremendous importance to the life sciences [Blount, 2015]. Though touted as the most well-understood organisms, the extent to which growth and size homeostasis of these cells (and more generally, single-cells of several organisms) is understood remains limited [Leygeber et al., 2019]. Microscopic time-lapse imaging of microfluidic single-cell cultivations provides the means for measuring cell size at spatio-temporal resolution [Grünberger et al., 2015]. Learning complex cell growth and homeostasis dynamics directly from image sequences without annotations is a hard problem. Sachs et al. [2022] have developed CellSium, a BSD licensed open-source software for simulating time-lapse images of microfluidic cultivations of single-cells with ground truth when explicitly programmed with cellular birth, growth and reproduction (division) mechanisms. Our contribution is a framework with 36 modalities for the generative (probabilistic) modelling of the said mechanisms, accounting for diverse bio-physical factors impacting cellular growth. We demonstrate here the grounding of our design principles in dominant cellular growth theories and present evidence of the framework's usability.

Submitted to 36th Conference on Neural Information Processing Systems (NeurIPS 2022). Do not distribute.

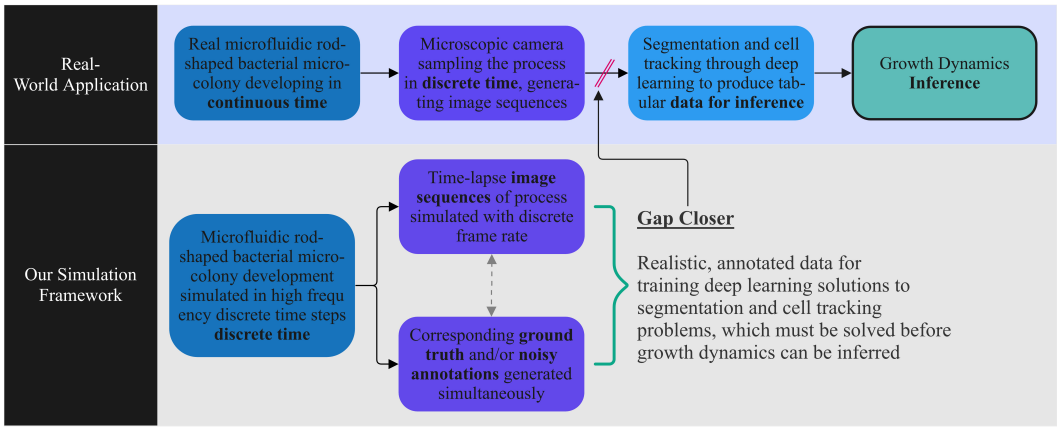

Figure 1: Top - Schematic of a real-world application marred by image-analysis bottleneck. Bottom - Our framework allows the generation of datasets to enable (semi-)supervised machine image analysis

2 Model Framework

2.1 Core Principles

Our framework allows for modelling cell lengths (growth rates) on the basis of scientific hypotheses related to two culture-level (global) considerations, namely the *growth law* cells obey [Schaechter et al., 1958], [Monod, 1949] and the *cell-division paradigm* applicable [Amir, 2014], [Taheri-Araghi et al., 2015], and three other factors, namely the availability of nutrients (growth *medium*) [Schaechter et al., 1958], [Fantes and Nurse, 1977], the basal *expression of genes* regulating growth [Elowitz et al., 2002], [Dong et al., 1995] and intrinsic cell-to-cell *heterogeneity* [Hashimoto et al., 2016], [Ribbe and Maier, 2016]. A foundational premise is that prokaryotic cell-cycles consist of three distinct phases - B-, C- and D-periods - wherein cellular growth takes place only in the second [Wang and Levin, 2009]. For the sake of exposition, we present below one of 36 supported model classes (we intend to summarise the remaining in the appendix should the paper be accepted).

2.2 Instantaneous Cell Lengths

Under the exposed model, cells obey the Linear Growth law [Schaechter et al., 1958]. Our model of instantaneous cell lengths depends on three free parameters – τ_0, the B-period duration of cells, τ_2, the D-period duration of cells, and λ_1, the start length of the first generation cell. The model assumes that the B- and D-periods of all cells are equally long and that the division of a cell births two daughter cells of the same length. Equations characterising instantaneous length, $l_i(a_i)$ of a cell i under the model are given below:

$$l_1(a_1) = \begin{cases} \lambda_1, & \text{if } a_1 \leq \lfloor \tau_0/\gamma \rfloor \text{ i.e., the cell is in its B-period} \\ l_1(a_1-1) + g_1(a_1)\gamma, & \text{if } \lfloor \tau_0/\gamma \rfloor < a_1 \leq n_1 - \lceil \tau_2/\gamma \rceil \text{ i.e., cell in its C-period} \\ l_1(a_1-1) = \Lambda_1, & \text{if } a_1 > n_1 - \lceil \tau_2/\gamma \rceil \text{ i.e., the cell in is its D-period} \end{cases} \quad (1)$$

For all $i > 1$, i.e. for all cells other than the first generation cell:

$$l_i(a_i) = \begin{cases} \lambda_i = \Lambda_{\lfloor i/2 \rfloor}/2, & \text{if } a_i \leq \lfloor \tau_0/\gamma \rfloor \text{ (symmetrical division of parent, } \lfloor i/2 \rfloor) \\ l_i(a_i-1) + g_i(a_i)\gamma, & \text{if } \lfloor \tau_0/\gamma \rfloor < a_i \leq n_1 - \lceil \tau_2/\gamma \rceil \\ l_i(a_i-1) = \Lambda_i, & \text{otherwise, i.e. if } a_i > n_i - \lceil \tau_2/\gamma \rceil \end{cases} \quad (2)$$

Here, $g_i(a_i)$ denotes the growth rate of cell i at age a_i (i.e, the number of simulation time steps that have elapsed between the cell's birth and the current instant), γ denotes the duration of a simulation time-step, and n_i, the number of time-steps between birth and division of the cell.

2.3 Instantaneous Growth Rates

In this model, we include hypotheses about how intrinsic heterogeneity, basal gene expression and growth medium concentration impacts instantaneous growth rates of cells.

Intrinsic heterogeneity [Hashimoto et al., 2016], [Ribbe and Maier, 2016] and basal gene expression [Elowitz et al., 2002], [Dong et al., 1995] are modelled as determining the (finite) maximum growth rate any given cell i could possibly attain, denoted $g_{max}(i)$. We motivate this ceiling on growth rate by the simple observation that in nature bacteria have finite lifespans in which their lengths are limited to orders of micrometers. For any cell i, we compose $g_{max}(i)$ as the sum of maximum growth rate due to heterogeneity, denoted $g_{het}(i)$ and that due to basal gene expression, denoted $g_{gexp}(i)$, i.e. $g_{max}(i) := g_{het}(i) + g_{gexp}(i)$, such that in expectation, the contributions from both these components are equal. Being random and continuous, we desire that $g_{max}(\cdot)$ be distributed around a mean b (a model parameter), which offers the interpretation of a population mean maximum growth rate, with some variance, denoted v.

The Beta distribution, a member of the exponential family, is often used to model continuous variables restricted to the support $(0, 1)$. So for each component $g_{het}(\cdot)$ and $g_{gexp}(\cdot)$, we scale a symmetrical Beta distribution by b, so that the supports of both are restricted to $(0, b)$ whereas that of $g_{max}(\cdot)$ is restricted to $(0, 2b)$. The symmetry ensures that $\mathbb{E}[g_{max}(\cdot)] = \mathbb{E}[g_{het}(\cdot)]+\mathbb{E}[g_{gexp}(\cdot)] = {}^{b}\!/_{2}+{}^{b}\!/_{2} = b$. The variance desideratum is enforced so that $v = v_h + v_g$, where the summands are model parameters representing g_{max} variance due to heterogeneity and gene expression respectively. These parameters are essentially free, to the extent that they respect the integrity of the Beta distributions (i.e. the shapes of the Beta distributions must be positive). Presented below are equations characterising g_{het}:

$$
\begin{gathered}
\alpha_h = {}^{b^2}\!/_{8v_h} - {}^{1}\!/_{2}, \text{ where } v_h \in (0, {}^{b^2}\!/_{4}) \\
Z_h(i) \overset{iid}{\sim} \text{Beta}(\alpha_h, \alpha_h) \\
g_{het}(i) = bZ_h(i)
\end{gathered}
\tag{3}
$$

The model for $g_{gexp}(i)$, in particular, requires that there be a culture-level (global) hypothesis about how many genes $n_g > 0$ regulate growth rates in the same. Under this hypothesis, we model the support of each gene's basal expression level as the set $\{0, 1\}$ (*on* or *off*). We represent the collection of basal expression levels of all the n_g genes of a cell i as its *gene configuration*, $\zeta_i \in \{0, 1, ..., 2^{n_g} - 1\}$. We model this variable and its implication upon $g_{gexp}(i)$ with two parameters $\mathbf{p}_0 \in (0, 1)^{2^{n_g}}$ and $\mathbf{p}_{tr} \in (0, 1)^{2^{n_g} \times 2^{n_g}}$, as shown below. $\mathbf{p}_0$ is a normalised vector of probabilities, so that $\mathbf{p}_0(j)$, where $j \in \{0, 1, ..., 2^{n_g} - 1\}$ is the probability that the first generation cell has gene configuration (i.e., $\zeta_1) = j$. $\mathbf{p}_{tr}$ is a (sparse) matrix, where each row is a vector of normalised conditional probabilities. We can interpret $\mathbf{p}_{tr}$ as a transition matrix which manages the degree to which a daughter cell inherits its gene configuration. Thus we have:

$$
\begin{gathered}
\zeta_1 \sim \text{Categorical}(\mathbf{p}_0) \\
\zeta_i \mid \zeta_{\lfloor i/2 \rfloor} \sim \text{Categorical}(\mathbf{p}_{tr}(\zeta_{\lfloor i/2 \rfloor})) \\
\alpha_g = {}^{b^2}\!/_{8\mathbf{v_g}(\zeta_i)} - {}^{1}\!/_{2}, \text{ where } \mathbf{v_g}(\zeta_i) \in (0, {}^{b^2}\!/_{4}) \text{ for all } \zeta_i \in \{0, 1, ..., 2^{n_g} - 1\} \\
Z_g(i) \overset{iid}{\sim} \text{Beta}(\alpha_g, \alpha_g) \\
g_{gexp}(i) = bZ_g(i)
\end{gathered}
\tag{4}
$$

Instantaneous growth rate $g_i(t)$ of a cell i, limited to remain in the interval $[0, g_{max}(\cdot)]$, is then modelled as a sigmoid function of $g_{max}(i)$, instantaneous medium concentration [Schaechter et al., 1958] $m(t)$, and a factor $A_i(t)$, which can be interpreted as the fraction of space through which the cell is able to *access* media, based on its spatial positioning amongst other competing cells in the culture. To model the latter, we posit an (immaterial) envelope of thickness δ surrounding each cell's membrane. When multiple cells inhabit a dense neighbourhood, their envelopes can overlap, resulting in ground(s) for competition. $A_i(t)$ is then defined as the fraction of cell i's envelope through which it can exclusively access media plus a uniform portion of the fraction of its envelope comprising of competition grounds. Localising competition grounds is a non-trivial problem, and we device an algorithm for the task (see algorithm 1 in appendix).

2.4 Cell Division

The final component of the growth dynamics model is one governing cell division. In this model, we base cell division on the Adder paradigm [Amir, 2014], [Taheri-Araghi et al., 2015], which holds that cell division occurs after it has *added* a certain critical volume since birth. Given that rod-shaped bacteria grow only along the length dimension in diverse environments, we model cell i's division as occurring when it has added a critical length, $\Delta(i)$, modeled as:

$$\Delta(i) \overset{iid}{\sim} \text{Gamma}({\mu_a}^2/\sigma_a^2, \sigma_a^2/\mu_a) \tag{5}$$

3 Experiments

A total of 180 simulations of single-cell cultivations of rod-shaped bacteria in microfluidic apparatus were performed using five random seeds for each of the 36 models supported by our framework. Each simulation resulted a sequence of images resembling microscopic time-lapse images captured in a real microfluidic experiment (see fig 2 for an example) and a tabular set of annotations tracking cells, their lineages, their physical dimensions and spatial positions in each simulated camera frame.

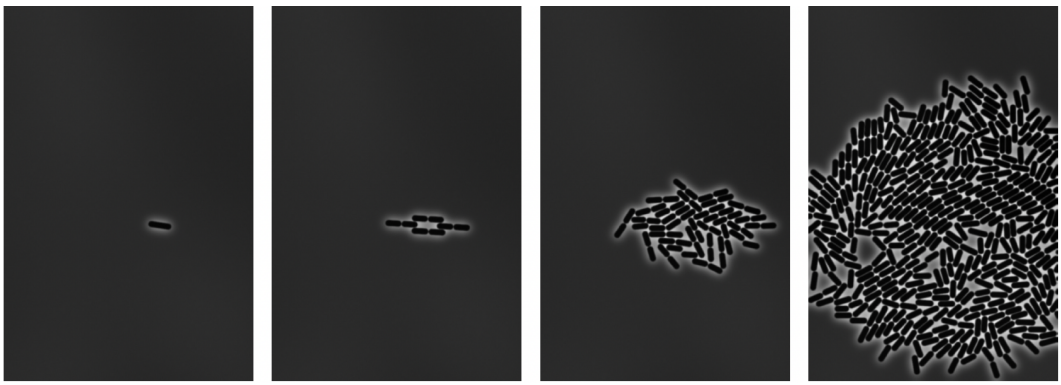

Figure 2: A simulated single-cell cultivation at t = 0, 4, 8, and 12 hrs respectively (left to right).

4 Discussion

Our work would bear fruit when the community develop detection and tracking networks using labelled datasets generated by this framework. However, the quintessential issue of whether learnings from high-quality simulation imagery and annotations would transfer to applications involving noisy real-world microscopy still looms. To alleviate this, the framework allows users the option to add independent white noise to simulated images as well as annotations.

In the same vein, whereas elements of our framework are indeed grounded in theories that have been instrumental in shaping the cell-growth discourse thus far, future work could enable the inclusion of expert knowledge in a Bayesian manner. Expert beliefs can be encoded as priors on model parameters, for example, adding hierarchical layer(s) to the framework, thereby making the framework more flexible, effectively extending its usability for learning more robust tracking networks that do not overfit on fixed parameters.

Furthermore, our work is limited to modelling the development of microfluidic rod-shaped bacterial cultures in a non-negative growth context. Antibiotics research is often concerned with how bacterial growth is contained and future work could address impacts of growth inhibitors on cell lengths and (negative) growth rates. Moreover gene expression levels can be modelled more realistically, as continuous variables where we presently model them as Bernoullis. Finally, future work could extend the framework to also model growth in prokaryotic organisms whose characteristic shapes are different from rod-shaped bacteria, and indeed to model single-cell growth in higher organisms.

References

A. Amir. Cell Size Regulation in Bacteria. *Physical Review Letters*, 112(20):208102, 2014.

Z. D. Blount. The natural history of model organisms: The unexhausted potential of E. coli. *eLife*, 4:e05826, 2015.

H. Dong, L. Nilsson, and C. G. Kurland. Gratuitous overexpression of genes in escherichia coli leads to growth inhibition and ribosome destruction. *Journal of Bacteriology*, 177(6):1497–1504, 1995.

M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain. Stochastic gene expression in a single cell. *Science*, 297(5584):1183–1186, 2002.

P. Fantes and P. Nurse. Control of cell size at division in fission yeast by a growth-modulated size control over nuclear division. *Experimental Cell Research*, 107(2):377–386, 1977.

A. Grünberger, C. Probst, S. Helfrich, A. Nanda, B. Stute, W. Wiechert, E. von Lieres, K. Nöh, J. Frunzke, and D. Kohlheyer. Spatiotemporal microbial single-cell analysis using a high-throughput microfluidics cultivation platform. *Cytometry Part A*, 87(12):1101–1115, 2015.

M. Hashimoto, T. Nozoe, H. Nakaoka, R. Okura, S. Akiyoshi, K. Kaneko, E. Kussell, and Y. Wakamoto. Noise-driven growth rate gain in clonal cellular populations. *Proceedings of the National Academy of Sciences*, 113 (12):3251–3256, 2016.

M. Leygeber, D. Lindemann, C. C. Sachs, E. Kaganovitch, W. Wiechert, K. Nöh, and D. Kohlheyer. Analyzing Microbial Population Heterogeneity—Expanding the Toolbox of Microfluidic Single-Cell Cultivations. *Journal of Molecular Biology*, 431(23):4569–4588, 2019.

J. Monod. The growth of bacterial cultures. *Annual Review of Microbiology*, 3(1):371–394, 1949.

J. Ribbe and B. Maier. Density-dependent differentiation of bacteria in spatially structured open systems. *Biophysical Journal*, 110(7):1648–1660, 2016.

C. C. Sachs, K. Ruzaeva, J. Seiffarth, W. Wiechert, B. Berkels, and K. Nöh. CellSium: versatile cell simulator for microcolony ground truth generation. *Bioinformatics Advances*, 2(1), 08 2022. ISSN 2635-0041. doi: 10.1093/bioadv/vbac053. URL `https://doi.org/10.1093/bioadv/vbac053`. vbac053.

M. Schaechter, O. Maaløe, and N. O. Kjeldgaard. Dependency on Medium and Temperature of Cell Size and Chemical Composition during Balanced Growth of Salmonella typhimurium. *Microbiology*, 19(3):592–606, 1958.

S. Taheri-Araghi, S. Bradde, J. T. Sauls, N. S. Hill, P. A. Levin, J. Paulsson, M. Vergassola, and S. Jun. Cell-Size Control and Homeostasis in Bacteria. *Current Biology*, 25(3):385–391, 2015.

J. D. Wang and P. A. Levin. Metabolism, cell growth and the bacterial cell cycle. *Nature Reviews Microbiology*, 7(11):822–827, 2009.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes]
   (c) Did you discuss any potential negative societal impacts of your work? [N/A]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes]
   (b) Did you include complete proofs of all theoretical results? [N/A]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? **[TODO]**

(b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [N/A]

(c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)?[N/A]

(d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? **[TODO]**

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...

(a) If your work uses existing assets, did you cite the creators? [Yes]

(b) Did you mention the license of the assets? [Yes]

(c) Did you include any new assets either in the supplemental material or as a URL? **[TODO]**

(d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]

(e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]

5. If you used crowdsourcing or conducted research with human subjects...

(a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]

(b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

A Appendix

Algorithm 1 Envelope Partitioning Algorithm

procedure PARTITION(E, M) ▷ E = $(E_1, ..., E_N)$, M = $(M_1, ..., M_N)$, where E_i is an envelope, or part thereof (*envelopelet*), and M_i is the (possibly incomplete) set of identifiers of cells competing for nutrition in E_i (*member set*), for all $i = \{1, 2, ..., N\}$

  if $|E|$ < 2 **then**

    return E, M ▷ Elementary instance of the problem

  else

    X $\leftarrow E_1 \setminus E_2 \setminus E_3 ... \setminus E_N$ ▷ Part(s) of E_1 accessed exclusively

    I $\leftarrow E_2 \cap E_1, E_3 \cap E_1, ..., E_N \cap E_1$ ▷ Intersections of E_1 with others

    R $\leftarrow E_2 \setminus E_1, E_3 \setminus E_1, ..., E_N \setminus E_1$ ▷ Remaining parts of $E_{2:n}$

    $M_X \leftarrow M_1$ ▷ X exclusively accessed by M_1

    $M_I \leftarrow M_2 \cup M_1, M_3 \cup M_1, ..., M_N \cup M_1$ ▷ Intersections member set update

    $M_R \leftarrow M \setminus M_1$ ▷ Remove M_1

    $E', M' \leftarrow$PARTITION($I \cup R, M_I \cup M_R$) ▷ Recursive call

    return $X \cup E', M_X \cup M'$

  end if

end procedure