

# Cell fate determination by Lamarckian molecule-inheritance and chance

Frank J. Bruggeman<sup>1,\*</sup>, Jaap Schouten<sup>2</sup>, Daan H. de Groot<sup>1</sup>, and Robert Planqué<sup>2</sup>

<sup>1</sup>Systems Bioinformatics, AIMMS, VU University, De Boelelaan 1108, 1081 HZ, Amsterdam, The Netherlands

<sup>2</sup>Mathematics, VU University, De Boelelaan 1085, 1081 HV, Amsterdam, The Netherlands

\*Corresponding author: f.j.bruggeman@vu.nl

## Abstract

Single, isogenic cells can differ in their survival and adaptation capacity. This phenotypic diversity is generally due to stochastic molecular events. Since mother cells on average pass half of their molecular content on to their daughters, the states of progeny cells strongly correlate with that of mother cells (Lamarckian inheritance). Why a particular cell deviates qualitatively from others therefore requires consideration of chance events along its ancestral lineage. Here we develop theory to understand cellular heterogeneity in terms of stochastic ancestral events of molecule synthesis, molecule degradation and cell divisions. We find that cell growth stochasticity has profound consequences for molecular heterogeneity in isogenic populations of cells, especially for long-lived molecules such as proteins. For instance, the lower bound on noise in molecule copy numbers that has often been observed experimentally is shown to be solely determined by the probability distribution for the generation times of cells. Thus, copy-number noise is unavoidable,

even in high-copy number circuits. Stochastic cell-fate and cell-differentiation decisions are therefore not necessarily due to noise in genetic circuits. We conclude that consideration of past chance events along cell lineages improves our understanding of how adaptive and maladaptive phenotypic heterogeneity arises in populations of isogenic cells.

## 1 Introduction

Random fluctuations in molecule concentrations may cause a particular cell to deviate in behaviour relative to others, even its close ancestors (Rosenfeld et al., 2005). Such phenotypic heterogeneity emerges spontaneously in isogenic cell populations (Paulsson, 2004). It may lead to the formation of subpopulations of cells, which in some cases spontaneously switch between phenotypic states at random times (Norman et al., 2015).

Whether fluctuations in a specific kind of molecule can have a significant phenotypic influence depends, amongst other factors, on its life time (Choi et al., 2008). Whereas some molecules live for a short period, because of rapid degradation, others exist for a long time and can only effectively reduce in concentration due to cell division events. For such stable molecular species, some copies in a cell were made by this cell itself, while others were inherited from its mother, grandmother, etc. The transition of a cell to another state, due to a deviating molecular content, can therefore be caused by events in its ancestral cells. For a stable molecule, at the end of a cell cycle, on average 50% of the cell's molecules came from its ancestors. If an ancestral cell had a high molecular content because of a stochastic event, such as a higher overall synthesis rate or a longer life time, then this causes the daughter cell to be born with high molecule counts as well. In short, past chance events can have phenotypic consequences for extant cells, a true case of Lamarckian inheritance.

This phenomenon has been illustrated for the *lac* operon in *E. coli* (Choi et al., 2008). When *E. coli* grows in an environment with glucose and lactose, it prefers glucose and grows on lactose only when glucose has been depleted. Since *E. coli* requires a lactose permease to import lactose and lacks a lactose sensor, cells only initiate growth on lactose if a lactose permease was already produced purely by chance during growth on glucose, or after glucose depleted. This mechanism might be functional for all nutrients that cannot pass the membrane freely by diffusion and for which cells lack a sensing mechanism (van Boxtel et al., 2017).

Another example of the importance of past chance events is the occurrence of phenotypic diversification in isogenic cell populations. Phenotypic diversification can be reversible or irreversible. Persister formation in *E. coli* is an example of the reversible case, in which cells randomly switch between a growing and a hardly-growing antibiotic-tolerant phenotype (Norman et al., 2015). If this population is confronted with a sudden

extinction-threatening condition, the persister cells have a higher chance of survival. Such behaviour can result for instance from the stochastic dynamics of toxin/antitoxin systems and fluctuations in the alarmone molecule ppGpp (Maisonneuve et al., 2013). In the irreversible case, spontaneous, random switching between phenotypes does not occur. During growth-supporting conditions one phenotype, e.g., a stress-resistant spore (Levine et al., 2012), is made via a chance event. The spore can survive harsh conditions, during which the growing phenotype dies, and can resuscitate again when conditions improve.

Here we present a theory that describes how the variability in molecular content of a single cell relates to random molecular events in itself and in any of its ancestors. We consider a description of the stochasticity of molecular circuits, cell growth and division; both of these influence the stochasticity of molecule copy numbers. At the level of molecular circuits, random fluctuations arise because of asynchronous molecule synthesis and degradation events that are influenced by fluctuating regulators and feedback circuitry. Copy-number stochasticity also arises because some cells have longer generation times than others, causing them to have more molecules at the end of their lifetime. In addition, random partitioning of molecules over two daughter cells during cell division, causes variation in molecule number between sisters. In order to obtain an ancestral perspective on molecule copy number in a single cell, we have to consider for each of its molecules the probability that they were made a certain number of generations ago, have not been degraded since, and ended up in the focal cell after each cell division. If we would achieve this, we would have a complete understanding of molecular fluctuations in living cells.

The outline of this paper is as follows. We start by illustrating how phenotypic heterogeneity can arise by considering the molecular stochasticity in a single cell that is part of a population of cells in balanced growth. Next, we express the probability that a single cell contains a certain number of molecules in terms of molecular chance events in any of its ancestral cells. We derive a loss rate of molecular memory in terms of molecular and cellular parameters. Then we apply the theory to stable molecules, for which ancestral effects are most important. We characterise the ancestral contribution to the statistics of molecule copy numbers of a single cell for different simple molecular circuits for protein synthesis. In all these cases, we find a non-negligible contribution of the stochasticity of cell growth and division events on the molecular stochasticity of a single cell. The randomness in generation times of cell (i.e., cell cycle durations) causes macroscopic, deterministic models, and the variance estimates deriving from those, to be inexact. We discuss how significant those deviations are for making predictions. Another finding is that molecules display stochasticity that is entirely due to cellular, rather than molecular, chance events. This explains the existence of a lower noise floor that is often observed experimentally. Since the noise floor turns out to be due to cellular stochasticity, it has consequences even for abundant proteins, such as those involved in metabolic networks. Cell-fate decisions are therefore not limited to low-copy number genetic circuits, but can also arise in such high-copy number circuits.

## 2 Results

### 2.1 Cell fate decisions

Phenotypic diversification of a population of isogenic cells, which we study in this paper, is explained in Figure 1A. Imagine a cell that can exist in one of two phenotypes. We shall assume that a cell's fate, i.e., the phenotype it displays, depends on whether a particular molecule  $Y$ , such as a transcription factor, is above or below a critical concentration (Figure 1A). The level  $y$  of  $Y$  is assumed to be entirely set by the level  $x$  of another regulating molecule  $X$ , so  $y = f(x)$  for some suitable function  $f$ . Since fluctuations occur in  $X$ , of which some will propagate to  $Y$ , the noise in  $Y$  is partially due to noise in  $X$ . High noise in  $X$  might then cause some cells to be in phenotypic state 1 and others in state 2 (Figure 1B). To understand this diversification, we have to understand why cells have deviating levels of  $X$ . For proteins, which are generally long-lived, the copy number of molecules found in a cell depends on past molecular and cellular events. Hence, we have to look back in time along the cell lineage to find the ancestral cells that had large enough fluctuations of  $X$  (Figure 1C). An explicit experimental example of this phenomenon can be found in Choi et al. (2008).

### 2.2 Derivation of the molecule copy number distribution for single cells in terms of their ancestral cell histories

The molecular content of single cells is ever-changing during their cell cycle time. For cells that continuously grow and divide in a non-steady state manner, it is generally impossible to determine the probability distribution of finding a certain number of molecules inside a cell of a given age. However, if the population is in a state of balanced growth, all the probability distributions that characterise this population, such as cell volumes at birth and division, cell ages, generation times and molecule copy numbers, are independent of time (Painter and Marr, 1968). In this growth condition, the number of cells,  $N_c$ , grows exponentially in time with a constant specific growth rate  $\mu = \frac{1}{N_c} \frac{dN_c}{dt}$ . Only in balanced growth, therefore, these distributions are constant in time and may be determined. We assume that populations grow balanced throughout this paper.

The probability that a cell in a population contains  $x$  molecules of a certain type can be decomposed into two stochastic components. The first component specifies the probability to observe a cell with a certain age  $a$ , which is described by the statistical theory of cell populations that grow balanced (Collins and Richmond, 1962; Painter and Marr, 1968; Powell, 1964). The second component specifies the probability that a cell at age  $a$  contains  $x$  molecules of that type. This probability derives from the theory about the stochasticity of molecular reactions (Van Kampen, 1992; Gillespie, 1992; McQuarrie, 1967). In this work we will merge those two theories.

We denote by  $G_x(z)$  the probability generating function (PGF) of the probability mass function (PMF) of the molecule counts. The PGF determines the probability distribution that a cell drawn at random from a population in balanced growth contains a particular

number of molecules of a specific type.  $G_x(z)$  is related to the cell-age distribution  $u(a)$  (a probability density function (PDF)) and  $G_{x|a}(z)$ , the PGF of the probability mass function specifying the probability that a cell contains  $x$  molecules at age  $a$  (Schwabe and Bruggeman, 2014a), via

$$G_x(z) = \int_0^\infty u(a)G_{x|a}(z)da. \quad (2.1)$$

To derive the overall PGF for  $G_{x|a}(z)$ , we first need to consider a single ancestral lineage, including the division times of the mother, grandmother, etc. These PGFs are denoted by  $G_{x|a}^{\text{lin}}(z)$ . Later we will consider the possibilities of different lineages. When we consider a cell at age  $a$ , then this cell was born  $a$  time ago. For the ancestral cells, their ages at division matter most for our theory, i.e., their division times. We shall denote by  $T_1$  the division time of the mother, by  $T_2$  the division time of the grandmother, and so on. The sequence  $T_i$  thus marks the sequence of division times of past generations within this specific lineage.

The PGF  $G_{x|a}(z)$  can itself be expressed as a product of two PGFs. One is associated with the probability of the number of molecules synthesised since cell birth ( $a$  time ago), and another with the probability that the molecules were inherited from the mother cell, and have ‘survived’ since birth, a period of duration  $a$ . The latter effect can be considered an example of Lamarckian inheritance. This decomposition is reflected in terms of generating functions as

$$G_{x|a}^{\text{lin}}(z) = F_{x|a}(z)H_{x|a}^{\text{lin}}(z), \quad (2.2)$$

where the PGF  $H_{x|a}^{\text{lin}}(z)$  accounts for the probability of the number of molecules obtained from previous generations, and  $F_{x|a}(z)$  accounts for the probability for the number of newly produced molecules.

The molecules obtained from the mother cell are only degraded in the current cell. Their number at time  $a$  is described by a Bernoulli distribution with age-dependent parameter  $p(a)$  and PGF  $1 - p(a) + p(a)z$ , reflecting the probability that a molecule has not yet been degraded at time  $a$  after birth. This leads to the relation

$$H_{x|a}^{\text{lin}}(z) = H_{x|0}^{\text{lin}}(1 - p(a) + p(a)z). \quad (2.3)$$

At age 0, the probability for zero synthesised molecules is 1, i.e.,  $F_{x|0}(z) = 1$ . Hence,

$$G_{x|0}^{\text{lin}}(1 - p(a) + p(a)z) = H_{x|0}^{\text{lin}}(1 - p(a) + p(a)z).$$

Due to balanced growth, we know that the  $G_{x|a}^{\text{lin}}(z)$  is constant in time: mother and daughter cells have the same distribution. We can therefore relate  $G_{x|0}^{\text{lin}}$  to the molecule count in the mother cell at the time of division  $T_1$ , by keeping track which molecules from the mother end up in which daughter. Assuming that for a particular division this probability is equal to a constant  $q$  for each individual molecule, we have

$$G_{x|0}^{\text{lin}}(z) = G_{x|T_1}^{\text{lin}}(1 - q + qz).$$

Combined with (2.3), we have

$$H_{x|a}^{\text{lin}}(z) = G_{x|T_1}^{\text{lin}}(1 - qp(a)(1 - z)).$$

Here we recognise  $qp(a)$  as the probability of a molecule to be partitioned into one of the daughter cells and to survive up to age  $a$ . Thus, we obtain the following expression

$$G_{x|a}^{\text{lin}}(z) = F_{x|a}(z)G_{x|T_1}^{\text{lin}}(1 - qp(a)(1 - z)), \quad (2.4)$$

which relates the distribution of molecule counts in the current generation to that of the previous generation. The PGF for the previous generation,  $G_{x|T_1}^{\text{lin}}(z)$ , itself satisfies

$$G_{x|T_1}^{\text{lin}}(z) = F_{x|T_1}(z)G_{x|T_2}^{\text{lin}}(1 - qp(T_1)(1 - z)),$$

so that

$$\begin{aligned} & G_{x|T_1}^{\text{lin}}(1 - qp(a)(1 - z)) \\ &= F_{x|T_1}(1 - qp(a)(1 - z))G_{x|T_2}^{\text{lin}}(1 - qp(T_1)(1 - (1 - qp(a)(1 - z)))) \\ &= F_{x|T_1}(1 - qp(a)(1 - z))G_{x|T_2}^{\text{lin}}(1 - q^2p(a)p(T_1)(1 - z)). \end{aligned}$$

Therefore, equation (2.4) becomes

$$G_{x|a}^{\text{lin}}(z) = F_{x|a}(z)F_{x|T_1}(1 - qp(a)(1 - z))G_{x|T_2}(1 - q^2p(a)p(T_1)(1 - z)).$$

and going back  $N$  generations leads to the following relation for a single lineage,

$$\begin{aligned} & G_{x|a}^{\text{lin}}(z) = \\ & F_{x|a}(z) \prod_{i=1}^{N-1} F_{x|T_i} \left[ 1 - q^i(1 - z)p(a) \prod_{j=1}^{i-1} p(T_j) \right] G_{x|T_N} \left[ 1 - q^N p(a) \prod_{j=1}^{N-1} p(T_j)(1 - z) \right]. \end{aligned}$$

(The terms in large brackets are the arguments of the PGFs, not product terms. The products have a meaningful interpretation for  $N \geq 2$ ,  $i \geq 2$ ; for  $N = 1$  and  $i = 1$  they should be read as yielding factors of 1.)

As a last step, we need to consider that a single cell may have many possible lineages, each one specified by the division times of its ancestors,  $T_1, T_2, \dots$ . Let  $f_\tau(T)$  denote the generation time distribution for the average lineage (Thomas, 2019; Wakamoto et al., 2012). A particular sequence of generation times, i.e., the specific lineage, comes with a probability equal to  $\prod_{j=1}^i f_\tau(T_j) dT_j$ . The overall probability that a cell contains  $x$  molecules at age  $a$  can now be calculated by integrating over all lineages,

$$\begin{aligned} & G_{x|a}(z) = \\ & F_{x|a}(z) \times \\ & (\int_0^\infty f_\tau(T_1) F_{x|T_1}[1 - qp(a)(1 - z)] \times \\ & (\int_0^\infty f_\tau(T_2) F_{x|T_2}[1 - q^2p(a)p(T_1)(1 - z)] \times \\ & (\int_0^\infty f_\tau(T_3) F_{x|T_3}[1 - q^3p(a)p(T_1)p(T_2)(1 - z)] \times \\ & \dots \times (\int_0^\infty f_\tau(T_i) F_{x|T_i}[1 - q^i p(a) \prod_{j=1}^{i-1} p(T_j)(1 - z)] dT_N) \dots dT_3) dT_2) dT_1) \quad (2.5) \end{aligned}$$

which may be simplified somewhat to

$$G_{x|a}(z) = F_{x|a}(z) \underbrace{\int_0^\infty \dots \int_0^\infty}_{N} \prod_{i=1}^N f_\tau(T_i) F_{x|T_i} [1 - q^i p(a) \prod_{j=1}^{i-1} p(T_j)(1-z)] dT_N dT_{N-1} \dots dT_1. \quad (2.6)$$

Equation (2.6) is the main equation for this paper. Increasing the number of ancestral generations  $N$  should improve the approximation of the probability distribution of  $x|a$ . As we will see, for molecules that have a short “memory” (they are short-lived) considering even one or two generations should give acceptable accuracies. Fluctuations in such molecules may give rise to fate decisions in cells that are only a one or two generations later in the lineage. For long-lived molecules, explaining deviating values requires looking back multiple generations ago.

## 2.3 The rate of molecular memory loss by cell division and molecule degradation

To obtain a good intuition for how many generations we should look back in time to understand the historical events that have shaped the molecular composition of a cell, we derive an expression for the rate of “memory loss”. To do this, we choose  $q = \frac{1}{2}$ , so each molecule has equal probability to be partitioned into either of the two daughter cells. We also choose a specific function for the probability that a synthesised molecule survives (is not degraded) for a period of length  $t$ ,  $p(t) = e^{-k_d t}$ , where  $k_d$  is a first-order degradation rate constant.

With these two model ingredients, the loss rate can be understood from (2.6) from the term expressing the fraction of molecules that have survived from an ancestor  $i$  generations ago,

$$q^i \prod_{j=1}^{i-1} p(T_j) = e^{-(i \ln 2 + \sum_{j=1}^{i-1} k_d T_j)}.$$

When  $i$  equals a large value  $N$  we obtain

$$e^{-(N \ln 2 + k_d \sum_{j=1}^{N-1} T_j)} \approx e^{-(N \ln 2 + N \langle T \rangle)} = e^{-N \langle T \rangle (\frac{\ln 2}{\langle T \rangle} + k_d)}.$$

A corresponding macroscopic description of changes in the concentration of this molecule due to dilution by growth and degradation is given by  $\dot{x} = -(\mu + k_d)x$  with solution  $x(t) = x(0)e^{-(\mu+k_d)t}$ . We define  $t := N \langle T \rangle$  so that,

$$N \langle T \rangle (\mu + k_d) = N \langle T \rangle \left( \frac{\ln 2}{\langle T \rangle} + k_d \right).$$

Therefore, an estimate of the average loss rate per generation equals

$$\Lambda = e^{-(\ln 2 + k_d \langle T \rangle)} = \left( \frac{1}{2} \right)^{1 + \frac{k_d}{\mu}}. \quad (2.7)$$

and the fraction of molecules from  $N$  generations ago hence equals  $\Lambda^N$ . Thus, the dimensionless parameter  $k_d/\mu$ , which is proportional to the number of cellular generation times per molecule life time, determines how many generation times one should look back in order to explain the molecule copy numbers that a particular cell contains; the probability that a cell contains molecules that are older than that time becomes negligibly small. Short-lived molecules have a high  $k_d/\mu$  value so their memory decays quickly while non-degraded, ‘stable’ molecules have the longest memory since  $k_d/\mu = 0$ .

We note that Equation (2.7) cannot be directly substituted in (2.6), because it only applies when we look at histories that are so long that their average generation times matches the mean generation time of the entire population.

## 2.4 The benchmark case of stable molecules

### 2.4.1 The main equation, (2.6), simplifies for stable molecules

Completely stable molecules, which are never degraded, have the strongest possible lineage effect on copy numbers of an extant cell. This extreme case therefore may serve as a benchmark, against which more realistic scenarios may be compared. The case of stable molecules can be solved explicitly for reasonable assumptions on the mechanism involving molecule synthesis and the generation time distribution  $f_\tau(T)$ , as we will see.

We start from (2.6). If the molecules are not degraded, so that  $p(a) = p(T_j) = 1$ , we obtain, with  $N \rightarrow \infty$ , that

$$G_{x|a}^{stable}(z) = F_{x|a}(z) \cdot \lim_{N \rightarrow \infty} \prod_{i=1}^N \int_0^\infty f_\tau(T_i) F_{x_i|T_i}[1 - q^i(1 - z)] dT_i. \quad (2.8)$$

This equation is easier to analyse than (2.6).

### 2.4.2 The Poisson model for protein synthesis

To gain further intuition, we start by considering the simplest circuit. Let us assume that a molecule is synthesized by a zero-order process, depicted by  $\xrightarrow{k_s} x$ . In this case, the probability mass function for the number of molecules synthesized since birth follows a Poisson distribution with parameter  $k_s a$ . Its probability generating function is described by  $F_{x|a}(z) = e^{k_s a(z-1)}$ , so that

$$F_{x|T_i}(1 - q^i(1 - z)) = e^{k_s T_i(-q^i(1-z))}.$$

Moreover, we assume the distribution of generation times follows an Erlang distribution,

$$f_\tau(T) = \frac{\lambda^k T^{k-1} e^{-\lambda T}}{\Gamma(k)}, \quad (2.9)$$

with mean generation time  $\langle T \rangle = \frac{k}{\lambda}$  and variance  $\langle \delta^2 T \rangle = \frac{k}{\lambda^2}$ , in line with experimental findings (Heerden et al., 2017). A random variable is Erlang distributed with parameters  $k$  and  $\lambda$  if it is the sum of  $k$  exponentially distributed variables, each with parameter  $\lambda$ .

With these two modelling ingredients specified, (2.8) can be explicitly calculated,

$$\begin{aligned}
 G_{x|a}^{stable}(z) &= F_{x|a}(z) \lim_{N \rightarrow \infty} \prod_{i=1}^N \int_0^\infty f_\tau(T_i) e^{k_s T_i (-q^i(1-z))} dT_i \\
 &= F_{x|a}(z) \lim_{N \rightarrow \infty} \prod_{i=1}^N \left( \frac{\lambda}{\lambda + k_s(q^i(1-z))} \right)^k \\
 &= F_{x|a}(z) \underbrace{\lim_{N \rightarrow \infty} \prod_{i=1}^N \left( \frac{1 - r(i)}{1 - r(i)z} \right)^k}_{\text{Ancestral effect}}, \tag{2.10}
 \end{aligned}$$

where

$$r(i) = \frac{k_s q^i}{\lambda + k_s q^i}.$$

We note that the term  $\left(\frac{1-r(i)}{1-r(i)z}\right)^k$  in (2.10) is the PGF of a negative binomial distribution with parameters  $k$  and  $r(i)$ , indicating that in this model the number of molecules in a cell equals the sum of a Poisson-distributed number of molecules made in the current generation and of negative-binomially distributed molecule numbers made by ancestral cells.

Thus the number of molecules,  $x_i$ , in a cell that originate from an ancestral cell  $i$  generations ago is distributed according to a negative binomial distribution with  $i$  influencing one of its parameters. It has mean  $\langle x_i \rangle = \frac{k}{\lambda} k_s q^i = k_s \langle T \rangle q^i$ , which equals the multiplication of the mean number of molecules made in this cell,  $k_s \langle T \rangle$ , by the total dilution due to cell divisions, i.e.,  $q^i$ . Hence, in this way we have formally established that for completely stable molecules, on average a cell that is about to divide contains molecules that are for 50% made by itself, 25% by its mother, 12.5% by its grandmother, etc (as one would have expected, of course).

For a negative binomial distribution with parameters  $k$  and  $r(i)$ , the variance equals the mean divided by  $1 - r(i)$ . Hence, here  $\langle \delta^2 x_i \rangle = k_s \langle T \rangle q^i (1 + \frac{k_s q^i}{\lambda})$ , with  $k_s q^i / \lambda$  as the number of surviving molecules that were made in time interval  $1/\lambda$ . Adding to this that we have  $\langle T \rangle = \frac{k}{\lambda}$ ,  $\langle \delta^2 T \rangle = \frac{k}{\lambda^2}$ ,  $\langle T \rangle / k = \frac{1}{\lambda}$ , we get

$$\begin{aligned}
 \langle \delta^2 x_i \rangle &= k_s \langle T \rangle q^i \left( 1 + \frac{k_s q^i}{\lambda} \right) \\
 &= \langle x_i \rangle (1 + \langle x_i \rangle / k) \\
 &= \langle x_i \rangle \left( 1 + \langle x_i \rangle \frac{\langle \delta^2 T \rangle}{\langle T \rangle^2} \right).
 \end{aligned}$$

Therefore, the noise in that number of molecules equals

$$\begin{aligned} \frac{\langle \delta^2 x_i \rangle}{\langle x_i \rangle^2} &= \frac{1}{\langle x_i \rangle} + \frac{\langle \delta^2 T \rangle}{\langle T \rangle^2} \\ &= \frac{1}{k_s \langle T \rangle} q^{-i} + \frac{\langle \delta^2 T \rangle}{\langle T \rangle^2}. \end{aligned} \quad (2.11)$$

This is an important relation for our intuition. It indicates that: 1. the noise in the generation time distribution codetermines the heterogeneity in the copy numbers; 2. the noise is decreased if cells produce more molecules during single cell cycles, but not to zero; 3. the noise increases with the number of past generations ( $q^{-i}$  increases with  $i$ ). The last conclusion implies that we lose certainty – ‘information’ – when we consider older chance molecular events; they are expressed as a longer product of probabilities. Note however, that while the noise increases, its effect on the current cell could still decrease because the number of molecules decreases.

We show how the ancestral history disappears in a given cell as function of the number of past generations in Figure 2. Figure 2A shows the probability distribution of molecule copy numbers in the entire population of growing cells. In Figure 2B, the probabilities are shown that cells from the same population contain a certain number of molecules from ancestral cells  $i$  generations back. The probabilities of retaining molecules from past generations decreases.

#### 2.4.3 The macroscopic description of molecule copy numbers can be a poor approximation of the stochastic description

Using relation (2.10), we can determine the mean and variance of the copy number of molecules. The mean of a sum of random variables equals the sum of the means; therefore, the mean copy number at a particular cell age equals,

$$\langle x|a; N \rangle = k_s a + \sum_{i=1}^N k_s \langle T \rangle q^i = k_s a + k_s \langle T \rangle \frac{q(q^N - 1)}{(q - 1)}. \quad (2.12)$$

For  $q = 1/2$ , we have that  $q(q^N - 1)/(q - 1) \rightarrow 1$  for  $N \rightarrow \infty$ , so that

$$\langle x|a \rangle = \lim_{N \rightarrow \infty} \langle x|a; N \rangle = k_s(a + \langle T \rangle).$$

The term  $k_s \langle T \rangle$  equals the number of molecules at cell birth and  $k_s a$  equals the number of molecules made since birth.

The copy-number variance at a particular age,  $\langle \delta^2 x|a \rangle$ , equals the sum of the variances of the molecules made in different cell generations since those random events are independent in our model. Since we are considering a zero-order synthesis process, the number of molecules made in the cell follows a Poisson process and therefore has variance equal to the mean,  $\langle \delta^2 x|a \rangle = \langle x|a \rangle = k_s a$  and the variance of the ancestral molecules that

survived  $i$  generations, equals  $k_s \langle \delta^2 T \rangle q^i (k_s q^i + \lambda)$ . Therefore,

$$\begin{aligned} \langle \delta^2 x | a; N \rangle &= k_s a + \sum_{i=1}^N k_s \langle \delta^2 T \rangle q^i (k_s q^i + \lambda) \\ &= k_s a + k_s \langle \delta^2 T \rangle \frac{q (q^N - 1) (\lambda + q (\lambda + k_s q^N + k_s))}{(q^2 - 1)} \end{aligned}$$

so that for  $q = 1/2$ ,

$$\langle \delta^2 x | a \rangle = \lim_{N \rightarrow \infty} \langle \delta^2 x | a; N \rangle = k_s a + k_s \langle T \rangle + \frac{1}{3} k_s^2 \langle \delta^2 T \rangle. \quad (2.13)$$

We can now establish the overall mean and variance by integration over the age distribution. For the mean we find

$$\begin{aligned} \langle x \rangle &= \int_0^\infty \langle x | a \rangle u(a) da \\ &= k_s (\langle a \rangle + \langle T \rangle). \end{aligned} \quad (2.14)$$

The variance needs to be decomposed into

$$\langle \delta^2 x \rangle = \langle \delta^2 \langle x | a \rangle \rangle + \langle \langle \delta^2 x | a \rangle \rangle$$

where

$$\begin{aligned} \langle \delta^2 \langle x | a \rangle \rangle &= \langle \delta^2 (k_s (a + \langle T \rangle)) \rangle = k_s^2 \langle \delta^2 a \rangle \\ \langle \langle \delta^2 x | a \rangle \rangle &= \langle k_s a + k_s \langle T \rangle + \frac{1}{3} k_s^2 \langle \delta^2 T \rangle \rangle = \langle x \rangle + \frac{1}{3} k_s^2 \langle \delta^2 T \rangle. \end{aligned} \quad (2.15)$$

Thus we still need to calculate the mean  $\langle a \rangle$  and the variance  $\langle \delta^2 a \rangle$  of the cell age to be able to calculate the quantities mentioned above.

Following (Painter and Marr, 1968), we can calculate the population growth rate  $\mu$  from  $\int_0^\infty 2e^{-\mu t} f_\tau(t) dt = 1$ , after which we can determine the age distribution from  $u(a) = 2\mu e^{\mu a} \int_a^\infty f_\tau(t) dt$ . For the specific choice of an Erlang distribution for  $f_\tau(T)$  chosen earlier, see (2.9), the growth rate equals  $\mu = \frac{\langle T \rangle}{\langle \delta^2 T \rangle} \left( 2^{\frac{\langle \delta^2 T \rangle}{\langle T \rangle^2}} - 1 \right)$ . This indeed equals the macroscopic estimate  $\frac{\ln 2}{\langle T \rangle}$  when  $\langle \delta^2 T \rangle \rightarrow 0$ . Now that we have found  $\mu$ , we can determine

$$u(a) = \frac{2 (2^{1/k} - 1) \lambda e^{a(- (2^{1/k} - 1)) \lambda} \Gamma(k, a\lambda)}{\Gamma(k)}, \quad (2.16)$$

which has as mean value

$$\langle a \rangle = \frac{(2^{-1/k} - 1) k + 1}{\lambda 2^{1/k} - \lambda}$$

and variance

$$\langle \delta^2 a \rangle = \frac{\langle \delta^2 T \rangle^2}{\langle T \rangle^2 \left( 2^{\frac{\langle \delta^2 T \rangle}{\langle T \rangle^2}} - 1 \right)^2} - 4^{\frac{\langle \delta^2 T \rangle}{\langle T \rangle^2}} (2\langle T \rangle^2 + \langle \delta^2 T \rangle).$$

Using these identities, the values for the mean and variance of the copy number of  $X$  (2.15) become

$$\begin{aligned} \langle x \rangle &= k_s(\langle a \rangle + \langle T \rangle) = k_s \langle T \rangle \left( 1 - 2^{\frac{\langle \delta^2 T \rangle}{\langle T \rangle^2}} + \frac{\langle \delta^2 T \rangle}{\langle T \rangle^2} \frac{1}{2^{\frac{\langle \delta^2 T \rangle}{\langle T \rangle^2}} - 1} \right) \\ \langle \delta^2 x \rangle &= \langle x \rangle + \frac{1}{3} k_s^2 \langle \delta^2 T \rangle + k_s^2 \left( \frac{\langle \delta^2 T \rangle^2}{\langle T \rangle^2 \left( 2^{\frac{\langle \delta^2 T \rangle}{\langle T \rangle^2}} - 1 \right)^2} - 4^{\frac{\langle \delta^2 T \rangle}{\langle T \rangle^2}} (2\langle T \rangle^2 + \langle \delta^2 T \rangle) \right). \end{aligned} \quad (2.17)$$

When we take the limit  $\langle \delta^2 T \rangle \rightarrow 0$  then  $\langle x \rangle = \frac{k_s}{\mu}$  and  $\langle \delta^2 x \rangle = \langle x \rangle + \langle x \rangle^2(1 - 2(\ln 2)^2) \approx \langle x \rangle + 0.039\langle x \rangle^2$ . The mean value agrees with the deterministic model  $\frac{d\langle x \rangle}{dt} = k_s - \mu\langle x \rangle$  for constant synthesis rate  $k_s$  and dilution by growth, which indeed has as steady state  $\langle x \rangle = \frac{k_s}{\mu}$ , indicating that the macroscopic approximation assumes  $\langle \delta^2 T \rangle \rightarrow 0$ .

When this macroscopic ODE system is modelled probabilistically with chemical master equations and then linearised, using a method called linear noise approximation (LNA; see (Van Kampen, 2003)), the mean and variance around this steady state may be determined analytically, showing that they are equal,  $\langle x \rangle = \langle \delta^2 x \rangle$  (Elf and Ehrenberg, 2003; Bruggeman et al., 2009). Thus, we conclude that variance estimation with LNA misses the term  $\langle x \rangle^2(1 - 2(\ln 2)^2)$  when  $\langle \delta^2 T \rangle \rightarrow 0$  because cell growth is not modelled in sufficient detail. Therefore, copy-number noise estimations of biochemical circuits in single cells that are members of cell populations that grow balanced, using LNA, lack the term  $(1 - 2(\ln 2)^2)$ . This amounts to 20% coefficient of variation in molecule copy number and can exceed the value  $1/\langle x \rangle$  by orders of magnitude.

Equation (2.17) also indicates that the variance in molecule copy numbers is increased due to heterogeneity in generation times. (This is because cells having a deviating generation time have had deviating times for molecule synthesis and therefore made different amounts.) The increased variance in copy numbers, due to generation time variation, indicates that *molecular* non-phenotypic heterogeneity is enhanced by the stochasticity of *cellular* processes.

## 2.5 Noise in molecule copy numbers always exceeds a noise floor that is set by the probability distribution of generation times

The equations for the mean and variance in copy numbers, (2.14)–(2.15), indicate the existence of a ‘noise floor’, the lowest possible noise value, for any generation time distri-

bution,

$$\frac{\langle \delta^2 x \rangle}{\langle x \rangle^2} = \frac{1}{\langle x \rangle} + \underbrace{\frac{1}{3} \frac{\langle \delta^2 T \rangle}{(\langle a \rangle + \langle T \rangle)^2}}_{\text{Noise Floor}} + \frac{\langle \delta^2 a \rangle}{(\langle a \rangle + \langle T \rangle)^2}. \quad (2.18)$$

This result indicates that the noise floor, which is also found in experimental data (Wolf et al., 2015; Taniguchi et al., 2010), cannot be explained from biochemical reaction kinetics alone: the stochasticity associated with cell growth is responsible for its existence. Its theoretical value is actually quite close to experimental data (Kempe et al., 2015). We have also seen that, in the case of an Erlang-distributed generation time, when the variance of the generation time becomes zero, the second term in the noise floor equals  $1 - 2(\ln 2)^2$ .

Thus, noise in copy numbers cannot become zero in asynchronously growing cell populations. The second term of the noise floor arises because molecule copy numbers increase with cell age, and a mixture of cell ages is always found in such conditions, so that heterogeneity in copy number immediately follows. Note that molecular noise considered in terms of concentrations does not have this noise term (Schwabe and Bruggeman, 2014a). The first term of the noise floor, however, is due to variation in cell cycle times and will persist even if concentrations are considered.

The noise floor argues against an often used rule of thumb, i.e.  $\frac{\langle \delta^2 x \rangle}{\langle x \rangle^2} = \frac{1}{\langle x \rangle}$ , that suggests that copy number noise in abundant molecules is negligible. The existence of the noise floor implies that the origins of fate decisions can come from deviating copy numbers of abundant molecules. This can underlie cell-fate decisions leading to persister cells in bacterial cultures (since those have been attributed to stochastic events associated with metabolism (Maisonneuve et al., 2013)) provided that those phenomena are sensitive to copy number noise and not to concentration noise, as concentration noise does not appear to have a large noise floor (Schwabe and Bruggeman, 2014a).

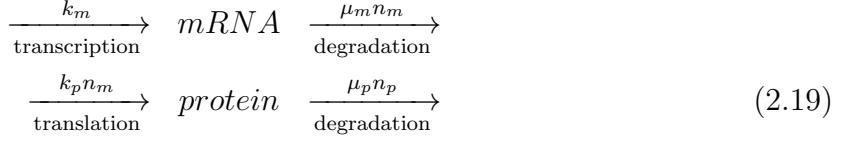
## 2.6 How do random bursts of translational activity in ancestral cells influence cell fate decisions?

Transcription and translation bursts are a potent mechanism to enhance molecular heterogeneity across isogenic cells (Raj and van Oudenaarden, 2008). Here we use the classical model for translation bursts (Shahrezaei and Swain, 2008; Dobrzynski and Bruggeman, 2009) in the context of ancestral history effects on the randomness of molecule copy numbers in single cells obeying balanced growth statistics.

The model describes a gene switching between an OFF and an ON state. Only during the ON state transcription occurs. Due to the switching of the gene, transcription activity will switch at random times from ON to OFF, and back, leading to inactive and active periods of random durations. During the active periods, synthesis and degradation of mRNA and proteins occurs while during the inactive period mRNA synthesis stops.

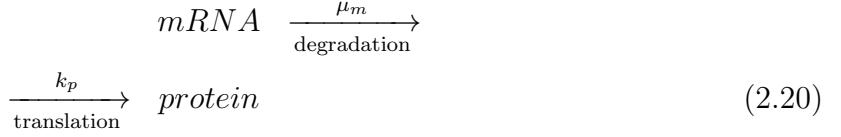
We shall be considering the following molecular circuit, with mRNA and protein having copy numbers  $n_m$  and  $n_p$ , respectively, and assume it to be embedded in a single cell that

is a member of an isogenic population of cells that grows balanced,



### 2.6.1 Stochasticity of translation bursts

A translation burst size is defined as the number of proteins made during the life time of a *single* mRNA. Thus, we are considering (note that  $n_m = 1$ ),



Since the times between consecutive protein production events are exponentially distributed, with mean time  $1/k_p$ , the number of proteins made in a time interval of length  $t$  follows a Poisson distribution (Schwabe et al., 2012) with mean and variance  $k_p t$ ; hence,  $n_p | t \sim \frac{e^{-k_p t} (k_p t)^{n_p}}{n_p!}$ . The duration of the protein production time is random itself and equals the life time of the mRNA, which we assume to follow an exponential distribution with mean  $\mu_m$ . The resulting probability distribution of the burst size,  $b$  therefore equals

$$b \sim \int_0^\infty \frac{e^{-k_p t} (k_p t)^b}{b!} e^{-\mu_m t} \mu_m dt = p(1-p)^b, \quad p := \frac{\mu_m}{k_p + \mu_m}. \tag{2.21}$$

Thus, the burst size follows a geometric distribution with mean

$$\langle b \rangle = \frac{k_p}{\mu_m} = \frac{\text{mean life time of mRNA}}{\text{mean time for a protein synthesis event}}$$

and variance  $\langle \delta^2 b \rangle = \langle b \rangle^2 + \langle b \rangle$ . We note that translation bursting causes large heterogeneity in protein copy numbers per cell because its noise equals  $\frac{\langle \delta^2 b \rangle}{\langle b \rangle^2} = \frac{1}{\langle b \rangle} + 1$ , which should be compared to 0 when bursts occur of a fixed size.

A useful relation for later is that the PGF for a geometric distribution equals  $\frac{p}{1-z+pz} = \frac{1}{1+\langle b \rangle(1-z)}$  and that a sum of  $N$  independent geometrically distributed random variables equals a negative binomial distribution, such that its PGF equals  $\left(\frac{p}{1-z+pz}\right)^N = \left(\frac{1}{1+\langle b \rangle(1-z)}\right)^N$ .

### 2.6.2 Probability distribution of molecule copy numbers

Following (Shahrezaei and Swain, 2008) the PGF for the number of proteins made by the circuit in (2.19) during a time interval of length  $t$  is given by

$$F_{x|t}(z) = \left( \frac{1 + \langle b \rangle e^{-\mu_p t} (1-z)}{1 + \langle b \rangle (1-z)} \right)^{\frac{\mu_m}{\mu_p}}. \tag{2.22}$$

If the cell had  $n_0$  molecules to start with then this PGF should be multiplied with the ‘survival’ term  $(1 + e^{-\mu_p t}(z - 1))^{n_0}$ .

If the proteins are not degraded then the previous PGF (2.22) simplifies to

$$F_{x|t}^0(z) := \lim_{\mu_p \rightarrow 0} F_{x|t}(z) = e^{\frac{-k_m \langle b \rangle (1-z)t}{1 + \langle b \rangle (1-z)}}.$$

Under this assumption the distribution of molecule copy numbers in a growing cell population can be obtained from (2.8). We first calculate the integral,

$$F_x^0(z) = \int_0^\infty f_\tau(t) F_{x|t}^0(z) dt,$$

which, using the choice  $f_\tau(t) \sim Erlang(k, \lambda)$  becomes

$$F_x^0(z) = \left( \frac{1}{1 + \langle b \rangle (1 + \frac{k_m}{\lambda})(1-z)} \right)^k (1 + \langle b \rangle (1-z))^k.$$

We note that this PGF equals the ratio of the PGFs of two negative binomial distributions, but we do not understand fully the physical interpretation of this insight. To determine the probability generating function of the molecule copy numbers in a growing population of cells we would have to calculate  $G_{x|a}^0(z)$ . To obtain the noise floor, however, we only need to derive an expression for the mean and variance of this molecule copy number.

We start by deriving those quantities conditional on the age. Since the distribution of  $x_0|a$  has PGF  $F_{x|a}^0(z)$  and that of  $x_i$  is  $F_{x|T_i}^0[1 - q^i(1-z)]$ , using reasoning identical to that for the Poisson model in Section 2.4.3,

$$\begin{aligned} \langle x|a \rangle &= \langle x_0|a \rangle + \lim_{N \rightarrow \infty} \sum_{i=1}^N \langle x_i \rangle = k_m \langle b \rangle (a + \langle T \rangle), \\ \langle \delta^2 x|a \rangle &= \langle \delta^2 x_0|a \rangle + \lim_{N \rightarrow \infty} \sum_{i=1}^N \langle \delta^2 x_i \rangle = \langle x|a \rangle + \frac{2}{3} b^2 k_m \langle T \rangle + \frac{1}{3} b^2 k_m^2 \langle \delta^2 T \rangle. \end{aligned} \quad (2.23)$$

For the mean and variance we obtain,

$$\begin{aligned} \langle x \rangle &= k_m \langle b \rangle (\langle a \rangle + \langle T \rangle) \\ \langle \delta^2 x \rangle &= \langle x \rangle + k_m^2 \langle b \rangle^2 \langle \delta^2 a \rangle + \frac{2}{3} k_m \langle T \rangle \langle b \rangle^2 + \frac{1}{3} k_m^2 \langle b \rangle^2 \langle \delta^2 T \rangle, \end{aligned} \quad (2.24)$$

which are completely general and independent of the mathematical definition of the interdivision time distribution. The noise in molecule numbers is therefore given by the following expression,

$$\frac{\langle \delta^2 x \rangle}{\langle x \rangle^2} = \frac{1}{\langle x \rangle} + \frac{2}{3} \frac{\langle T \rangle}{k_m (\langle a \rangle + \langle T \rangle)^2} + \frac{\langle \delta^2 a \rangle}{(\langle a \rangle + \langle T \rangle)^2} + \frac{1}{3} \frac{\langle \delta^2 T \rangle}{(\langle a \rangle + \langle T \rangle)^2}. \quad (2.25)$$

To derive the noise floor we have to consider the noise at  $\langle x \rangle \rightarrow \infty$ . If we achieve this by changing the transcription rate  $k_m$ , as one would normally do, then the last two terms in the previous equation remain,

$$\lim_{\langle x \rangle \rightarrow \infty} \frac{\langle \delta^2 x \rangle}{\langle x \rangle^2} = \frac{\langle \delta^2 a \rangle}{(\langle a \rangle + \langle T \rangle)^2} + \frac{1}{3} \frac{\langle \delta^2 T \rangle}{(\langle a \rangle + \langle T \rangle)^2}, \quad (2.26)$$

which is identical to the noise floor (2.18) that we found earlier for a model that did not include transcriptional bursts. The difference between these models is that at low and moderate levels copy numbers the translation bursts model generates more noise.

Thus, we find again that the noise floor is completely determined by the stochasticity associated with cell growth and that therefore cell fate decisions may also depend on molecules with high copy numbers, for which  $\frac{1}{\langle x \rangle} \approx 0$ . (Provided that those decisions depend on copy numbers and not on concentrations.)

Figure 3 summarises the main insights from the analysis of noise floors (eq. 2.25). In Figure 3A the noise floor is shown as function of the mean copy number, which ranges from a few per cell, for transcription, to hundreds for signalling proteins and (tens of) thousands for metabolic proteins. Thus at least about 20% of phenotypic variation is found for molecule copy numbers in growing cell populations that is entirely due to those populations consisting of cells having different sizes, i.e., ages, and that from birth to division the number of molecules doubles, during balanced growth. Figure 3B shows the dependency of the noise floor on the coefficient of variation of the generation time distribution  $f_\tau(t)$ . Large variation in generation times contributes to molecule copy number heterogeneity because cells that live longer make more molecules than cells living shorter.

## 2.7 Influence of chance events associated with cell growth

Growth rate influences many aspects of cell physiology (Klumpp et al., 2009). One of its effects is that it sets the generation time, which limits the time that cells have for molecule synthesis from their birth to division. The analysis leading to (2.25) indicates that if cells would tune mRNA synthesis with growth rate in order to compensate for the reduction in copy numbers of mRNA and protein molecules due to shorter generation times, noise in protein expression would be independent of growth rate. We think it is unlikely that cells can coordinate mRNA and/or protein synthesis with such precision. Molecular noise is therefore expected to be growth rate dependent. For instance, experimental data indicates that noise of a constitutively expressed protein increases with growth rate in *B. subtilis* (Nordholt et al., 2017).

Genes that are turned off by repressors do occasionally make proteins if the repressor falls off its regulatory site and a RNA polymerase binds instead (Choi et al., 2008). This “leaky gene expression” can lead to appreciable protein accumulation at low growth rate, because then generation times are long. This is shown in Figure 4AB. At high growth rate, the probability of a leaky expression event of a repressed gene during a single generation time is small. Few cells will therefore experience such faulty events and the probability for a cell to have zero molecules is then high. At low growth rate, however, leaky gene

expression events become more probable during a single generation time and more cells will experience leaky gene expression events occurring purely by chance. These cells accumulate proteins that they do not require now.

This accumulation of (apparently) unnecessary proteins due to leaky gene expression at low growth rate, can lead to spontaneous adaptation of cells, because cells could accumulate proteins above a threshold level. In this way, a contribution to adaptation to a new future condition appears purely by chance. This is illustrated in Figure 4CD. The probability distributions of molecule copy numbers were calculated at different growth rates. The probability that a cell contains more than 20 molecules, which we set to be the threshold fraction for adaptation, decreases with growth rate. This indicates that some cells have a spontaneous adaptation capacity at low growth rate, due to chance events (Choi et al., 2008), and that this cell fraction decreases with growth rate.

### 3 Discussion

Since mother cells pass on average 50% of their molecular content on to each daughter cell, the inherent randomness of molecular and cellular processes of one cell can affect the fate of its progeny cells. The lineage history of a cell therefore matters when we aim to understand why it displays a qualitatively different phenotype than its kin cells. This ‘Lamarckian inheritance’ mechanism plays an important role in the emergence and maintenance of phenotypic heterogeneity in growing cell cultures. In this paper, we related the molecule content of a cell to that of its ancestors.

We combined stochasticity at the molecular level with stochastic cellular processes. In this way we obtained a full probabilistic description of molecule copy numbers in exponentially growing cells. The condition of balanced growth is a requirement for our theory, because only under those conditions all the probability distributions of intensive characteristics of the growing cell population are time-invariant. The theory relates the current molecular state of the cell to that of its ancestors along its lineage, taking into account molecule turnover and partitioning during cell division. Our main equation (2.6) is completely general and can be used with any stochastic biochemical reaction network. It will not always be analytically solvable, in that case numerical simulations will be required.

The theory considers the statistics of molecular and cellular events along lineages. The generation times of cells play an important role. First, whether a single ancestral molecule survives in offspring cells depends on whether it is degraded before it is partitioned into the daughter cells. Therefore its survival depends on the generation time of the mother. Second, since the generation time sets the total synthesis time for cells to make molecules, cells with long generation times generally make more molecules. Thus, the variation of generation times between cells plays an important role in our theory. If those generation times show great variation then the population will be necessarily more heterogeneous too. We have little molecular understanding at the moment of what sets the generation time of a cell and how its variance is determined.

Our theoretical results mostly concern the case of infinitely stable molecules that do not degrade, but only dilute by cell division. The effect of ancestral stochasticity is greatest for such stable molecules, of which metabolic proteins are good examples.

The theory indicates that variances in copy number and in generation times matter to molecular heterogeneity. It has been known for a long time that nonlinear kinetics gives rise to approximative ordinary differential equation (ODE) models because of the lack of moment closure, i.e., the ODEs for the mean copy number also depend on the variance of the copy number (Goel and Richter-Dyn, 1974). Here, we have shown that deterministic ODE models of the dynamics of molecular circuits inside living cells fail even for the simplest possible zero-order reaction kinetics. Also the variance of the generation time matters: the mean copy number only equals the macroscopic ODE result when the variance of the generation times equals zero.

We found a noise floor that specifies a lower bound to the noise in highly abundant molecules, with copy numbers exceeding a 100 or so. Since 1 molecule per bacterial cell, such as *E. coli*, is approximately equal to a concentration of 1 nM, our result indicate that the noise floor applies to molecules that have concentrations in the  $\mu\text{M}$  range or higher. Those are mostly metabolic proteins and low-molecular weight molecules, e.g., intermediates of metabolism. Transcription factors and signalling proteins are generally lower in concentrations and shall experience higher noise levels than the noise floor.

The noise floor arises for two reasons. Firstly, in a cell population at balanced growth extant cells have different ages: some are just born, others are halfway through their cell cycle and others are about to divide. Since the copy number of each molecule is doubled on average during a single cell cycle, the molecule content of cells increases with their age. Heterogeneity in copy number therefore arises because of the heterogeneity in cell ages. This effect is independent of the mean copy number of a molecule so it persists when the mean copy number approaches infinity. Secondly, the variation in generation times causes variation in the number of molecules made by cells during their cell cycle. We note that the first effect also arises when all cells have the same generation time (Schwabe and Bruggeman, 2014b). Cell ages also have generally greater variance when the generation times are more variable amongst cells.

Copy number variation is higher for cell cultures with longer generation times. The theory therefore indicates a possible mechanism to induce heterogeneity, thereby increasing the adaptive capacities of microbial cell populations. It indicates how leaky gene expression of a repressed gene can lead to a bet-hedging type of adaptive behaviour at low growth rates that diminishes at high growth rate. The relationship between growth rate and protein copy number heterogeneity is important for microbiology. It appears to be associated with the adaptation of microorganisms to new nutrient conditions (Choi et al., 2008; Boulineau et al., 2013; van Heerden et al., 2014; Solopova et al., 2014; Kotte et al., 2014) and likely also the formation of persister cells (Balaban et al., 2004; Harms et al., 2016) as critical ppGpp fluctuations for persister cell formation are likely more probable at low growth rate when the mean ppGpp concentration is higher.

From this work we draw several conclusions. Fluctuations in ancestral cells that led to deviating phenotypes can propagate along a lineage tree of growing cells, causing

the formation of a subpopulation of cells, because mother cells pass on part of their molecular state to their daughters. The variation of generation times between cells plays an important role in copy-number heterogeneity in growing cell populations; it contributes to a copy-number noise floor that exists for each molecule. Finally, Lamarckian inheritance appears to play an important role in setting the heterogeneity and adaptive capacities of growing cell populations.

## 4 Acknowledgements

Daan de Groot was supported by the NWO VICI grant 865.14.005.

## References

Balaban, N., Merrin, J., Chait, R., Kowalik, L., and Leibler, S. (2004). Bacterial persistence as a phenotypic switch. *Science*, 305(5690):1622–1625.

Boulineau, S., Tostevin, F., Kiviet, D. J., ten Wolde, P. R., Nghe, P., and Tans, S. J. (2013). Single-cell dynamics reveals sustained growth during diauxic shifts. *PLoS ONE*, 8(4):e61686.

Bruggeman, F. J., Blüthgen, N., and Westerhoff, H. V. (2009). Noise management by molecular networks. *PLoS Comput Biol*, 5(9).

Choi, P. J., Cai, L., Frieda, K., and Xie, X. S. (2008). A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science*, 322(5900):442–446.

Collins, J. and Richmond, M. (1962). Rate of growth of bacillus cereus between divisions. *Microbiology*, 28(1):15–33.

Dobrzynski, M. and Bruggeman, F. J. (2009). Elongation dynamics shape bursty transcription and translation. *Proc Natl Acad Sci U S A*, 106(8):2583–2588.

Elf, J. and Ehrenberg, M. (2003). Fast evaluation of fluctuations in biochemical networks with the linear noise approximation. *Genome research*, 13(11):2475–2484.

Gillespie, D. T. (1992). A rigorous derivation of the chemical master equation. *Physica A: Statistical Mechanics and its Applications*, 188(1-3):404–425.

Goel, N. S. and Richter-Dyn, N. (1974). *Stochastic Models in Biology*. Blackburn Press.

Harms, A., Maisonneuve, E., and Gerdes, K. (2016). Mechanisms of bacterial persistence during stress and antibiotic exposure. *Science*, 354(6318):aaf4268.

Heerden, J. H., Kempe, H., Doerr, A., Maarleveld, T., Nordholt, N., and Bruggeman, F. J. (2017). Statistics and simulation of growth of single bacterial cells: illustrations with *b. subtilis* and *e. coli*. *Scientific reports*, 7(1):16094.

Kempe, H., Schwabe, A., Cremazy, F., Verschure, P. J., and Bruggeman, F. J. (2015). The volumes and transcript counts of single cells reveal concentration homeostasis and capture biological noise. *Mol. Biol. Cell*, 26(4):797–804.

Klumpp, S., Zhang, Z., and Hwa, T. (2009). Growth rate-dependent global effects on gene expression in bacteria. *Cell*, 139(7):1366–1375.

Kotte, O., Volkmer, B., Radzikowski, J. L., and Heinemann, M. (2014). Phenotypic bistability in escherichia coli's central carbon metabolism. *Molecular systems biology*, 10(7):736.

Levine, J. H., Fontes, M. E., Dworkin, J., and Elowitz, M. B. (2012). Pulsed feedback defers cellular differentiation. *PLoS biology*, 10(1):e1001252.

Maisonneuve, E., Castro-Camargo, M., and Gerdes, K. (2013). (p) ppGpp controls bacterial persistence by stochastic induction of toxin-antitoxin activity. *Cell*, 154(5):1140–1150.

McQuarrie, D. A. (1967). Stochastic approach to chemical kinetics. *Journal of applied probability*, 4(3):413–478.

Nordholt, N., Heerden, J., Kort, R., and Bruggeman, F. J. (2017). Effects of growth rate and promoter activity on single-cell protein expression. *Scientific reports*, 7(1):6299.

Norman, T. M., Lord, N. D., Paulsson, J., and Losick, R. (2015). Stochastic switching of cell fate in microbes. *Annual review of microbiology*, 69:381–403.

Painter, P. R. and Marr, A. G. (1968). Mathematics of microbial populations. *Annu. Rev. Microbiol.*, 22:519–548.

Paulsson, J. (2004). Summing up the noise in gene networks. *Nature*, 427(6973):415–418.

Powell, E. (1964). A note on Koch and Schaechter's hypothesis about growth and fission of bacteria. *Microbiology*, 37(2):231–249.

Raj, A. and van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell*, 135(2):216–226.

Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S., and Elowitz, M. B. (2005). Gene regulation at the single-cell level. *Science*, 307(5717):1962–1965.

Schwabe, A. and Bruggeman, F. J. (2014a). Contributions of cell growth and biochemical reactions to nongenetic variability of cells. *Biophysical journal*, 107(2):301–313.

Schwabe, A. and Bruggeman, F. J. (2014b). Contributions of cell growth and biochemical reactions to nongenetic variability of cells. *Biophys. J.*, 107(2):301–313.

Schwabe, A., Rybakova, K. N., and Bruggeman, F. J. (2012). Transcription stochasticity of complex gene regulation models. *Biophys J*, 103(6):1152–1161.

Shahrezaei, V. and Swain, P. S. (2008). Analytical distributions for stochastic gene expression. *Proc. Natl. Acad. Sci. U.S.A.*, 105(45):17256–17261.

Solopova, A., van Gestel, J., Weissing, F. J., Bachmann, H., Teusink, B., Kok, J., and Kuipers, O. P. (2014). Bet-hedging during bacterial diauxic shift. *Proceedings of the National Academy of Sciences*, 111(20):7427–7432.

Taniguchi, Y., Choi, P. J., Li, G. W., Chen, H., Babu, M., Hearn, J., Emili, A., and Xie, X. S. (2010). Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science*, 329(5991):533–538.

Thomas, P. (2019). Intrinsic and extrinsic noise of gene expression in lineage trees. *Scientific reports*, 9(1):474.

van Boxtel, C., van Heerden, J. H., Nordholt, N., Schmidt, P., and Bruggeman, F. J. (2017). Taking chances and making mistakes: non-genetic phenotypic heterogeneity and its consequences for surviving in dynamic environments. *Journal of The Royal Society Interface*, 14(132):20170141.

van Heerden, J. H., Wortel, M. T., Bruggeman, F. J., Heijnen, J. J., Bollen, Y. J. M., Planqué, R., Hulshof, J., O'Toole, T. G., Wahl, S. A., and Teusink, B. (2014). Lost in transition: Start-up of glycolysis yields subpopulations of nongrowing cells. *Science*, 343(6174).

Van Kampen, N. G. (1992). *Stochastic processes in physics and chemistry*, volume 1. Elsevier.

Van Kampen, N. G. (2003). *Stochastic processes in physics and chemistry*. North-Holland.

Wakamoto, Y., Grosberg, A. Y., and Kussell, E. (2012). Optimal lineage principle for age-structured populations. *Evolution: International Journal of Organic Evolution*, 66(1):115–134.

Wolf, L., Silander, O. K., and van Nimwegen, E. (2015). Expression noise facilitates the evolution of gene regulation. *Elife*, 4:e05856.

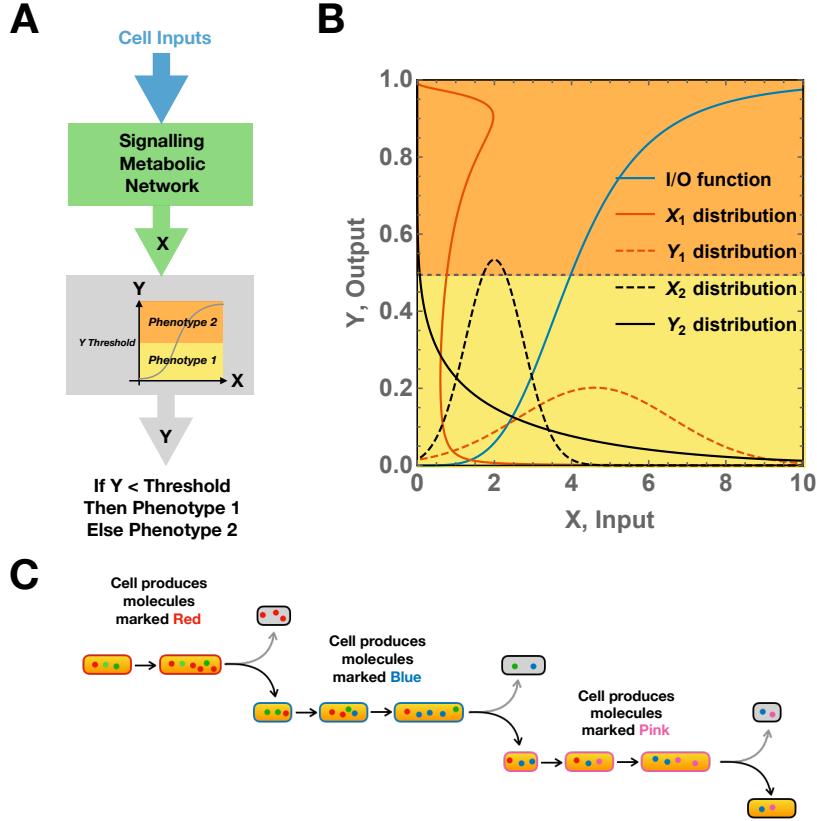
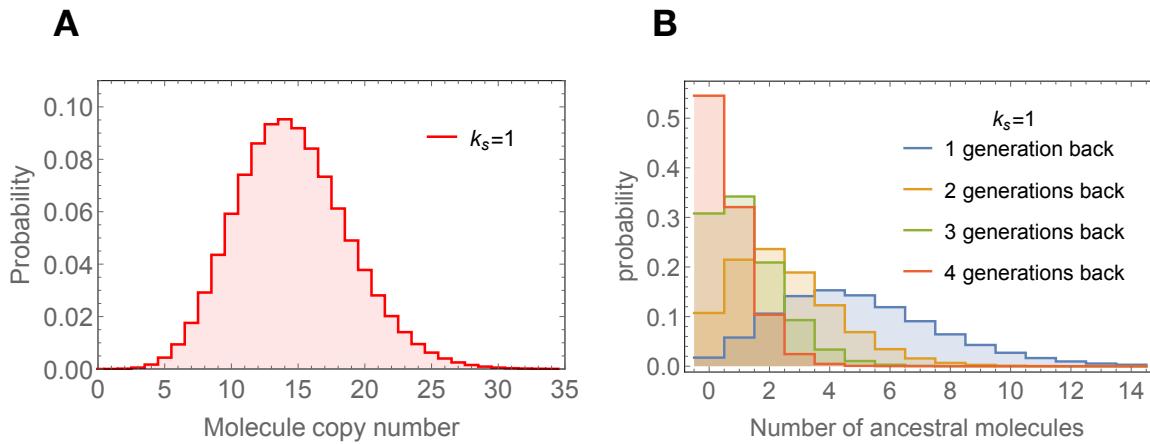


Figure 1: **How variable molecular states can affect a cell's fate.** **A.** A schematic overview of the cellular decision making process of a cell. The regulatory molecule  $X$  decides on a cell's fate, i.e., phenotype 1 or 2, by inducing changes that either cause  $Y$  to exist above or below a threshold level for two phenotypic states. **B.** Fluctuations in the level of  $X$  induces fluctuations in  $Y$  that can cause the cell population to diversify and display two states simultaneously (red scenario) or only one state (black scenario). Noise and the mean level of  $X$  and the dependency of  $Y$  on  $X$  determine whether subpopulations arise or not. **C.** If fluctuations in  $X$  cause a cell to shift phenotype, then chance events in any of its ancestors might be responsible. This is because some of the molecules in the cell are made in ancestral cells. In the figure, we colour-coded the molecules to mark the ancestor in which they were made. The last two sisters contain molecules from their grandmother. Also note, that cell division can cause variation in molecules between sisters, since the pink-producing cell starts with one molecule more than here sister.



**Figure 2: Illustration of how the molecule copy number distribution of a cell can be viewed as a sum of contributions made by its ancestral cells. A.** The molecule copy-number distribution according to (2.10) with parameters  $k_s = 1 \text{ min}^{-1}$ ,  $k = 10$ ,  $\lambda = 1 \text{ min}^{-1}$ , and  $q = \frac{1}{2}$ . This models a cell with a realistic generation times (mean 60 min and coefficient of variation of 30 %) and a low copy number of proteins, e.g. transcription factors. Cases of high protein expression are considered below. Calculation were done at the mean cell age. **B.** Contributions of ancestral cells to the molecule copy-number distribution shown in Figure A, according to (2.10). Calculations were done at the mean cell age.

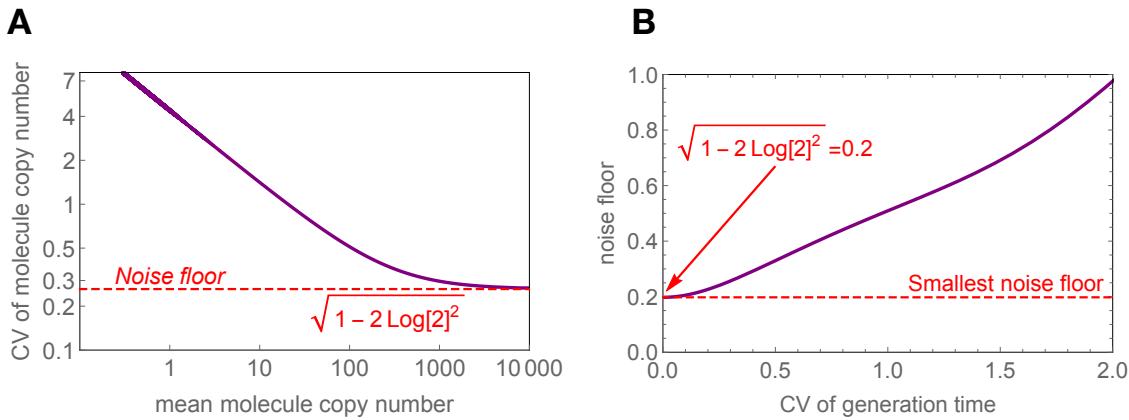
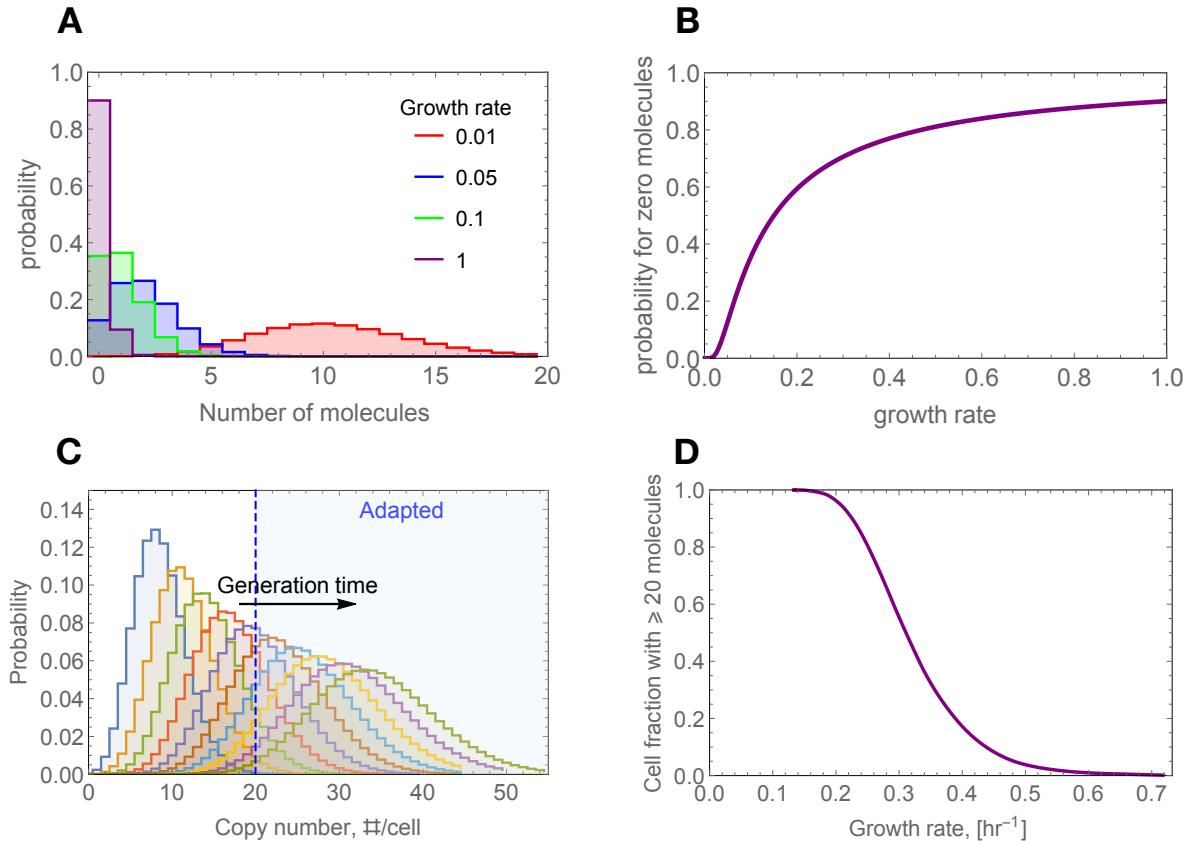


Figure 3: **Noise floor of molecule copy-numbers.** **A.** Calculation of the relation between the coefficient of variation and the mean of the molecule copy-numbers. This relation was calculated by varying the growth rate; alternatively we could have changed the rate of mRNA synthesis. The other parameters were to  $k_s = 2 \text{ mRNA/min}$ ,  $k_p = 4 \text{ proteins/min}$ ,  $\mu_m = 0.1 \text{ min}^{-1}$ . In bacterial cells, transcription factors occur as a few to tens of copies of per cell whereas metabolic proteins occur as hundreds to tens of thousands of copies per cell. **B.** The noise floor is only determined by the coefficient of variation in the cellular generation time, which has experimentally been found to be  $\approx 0.3$  (e.g. Heerden et al. (2017)). Those two plots were generated using equation 2.25 and by assuming that the interdivision time follows an Erlang distribution.



**Figure 4: The influence of growth rate on ‘leaky’ gene expression.** **A.** The growth rate influences the molecule copy number distribution. Since higher growth rate imply shorter generation times, less molecules are made by cell during the period from birth to division. Here we consider a gene that sometimes makes an mRNA under conditions when it is supposed to be off – it is a ‘leaky’ gene. Calculations were done at the mean cell age, using (2.10). **B.** At low growth rate single cells accumulate protein via leaky gene expression while this effect is negligible at high growth rate. Cell physiology can therefore be expected to be more noisy at low growth rate. Calculations were done at the mean cell age.