







# Reducing Bias and Quantifying Uncertainty in Fluorescence Produced by PCR

Robert F. DeJaco<sup>1,2</sup>  · Matthew J. Roberts<sup>1,3</sup> · Erica L. Romsos<sup>4</sup>  · Peter M. Vallone<sup>4</sup>  · Anthony J. Kearsley<sup>1</sup> 

Received: 16 November 2022 / Accepted: 20 June 2023 / Published online: 14 August 2023  
© The Author(s) 2023

## Abstract

We present a new approach for relating nucleic-acid content to fluorescence in a real-time Polymerase Chain Reaction (PCR) assay. By coupling a two-type branching process for PCR with a fluorescence analog of Beer's Law, the approach reduces bias and quantifies uncertainty in fluorescence. As the two-type branching process distinguishes between complementary strands of DNA, it allows for a stoichiometric description of reactions between fluorescent probes and DNA and can capture the initial conditions encountered in assays targeting RNA. Analysis of the expected copy-number identifies additional dynamics that occur at short times (or, equivalently, low cycle numbers), while investigation of the variance reveals the contributions from liquid volume transfer, imperfect amplification, and strand-specific amplification (i.e., if one strand is synthesized more efficiently than its complement). Linking the branching

---

✉ Robert F. DeJaco  
rdejaco@compactmembrane.com

Matthew J. Roberts  
mjrobert@ida.org

Erica L. Romsos  
erica.romsos@nist.gov

Peter M. Vallone  
peter.vallone@nist.gov

Anthony J. Kearsley  
anthony.kearsley@nist.gov

- <sup>1</sup> Applied and Computational Mathematics Division, National Institute of Standards and Technology, 100 Bureau Dr., MS 8910, Gaithersburg, MD 20899-8910, USA
- <sup>2</sup> Department of Chemistry and Biochemistry, University of Maryland, 8051 Regents Dr., College Park, MD 20742-4454, USA
- <sup>3</sup> Cost Analysis and Research Division, Institute for Defense Analyses, 730 E. Glebe Rd., Alexandria, VA 22305-3086, USA
- <sup>4</sup> Biomolecular Measurement Division, National Institute of Standards and Technology, 100 Bureau Dr., MS 8314, Gaithersburg, MD 20899-8314, USA

process to fluorescence by the Beer's Law analog allows for an *a priori* description of background fluorescence. It also enables uncertainty quantification (UQ) in fluorescence which, in turn, leads to analytical relationships between amplification efficiency (probability) and limit of detection. This work sets the stage for UQ-PCR, where both the input copy-number and its uncertainty are quantified from fluorescence kinetics.

**Keywords** Real-time polymerase chain reaction · Stochastic branching process · Uncertainty quantification

## 1 Introduction

Polymerase Chain Reaction (PCR) is a hallmark of molecular biology and applied genetics. When the dynamics of PCR are monitored by a fluorescent probe, the initial amount of target sequence can be quantified (qPCR) by a computational algorithm equipped with a mathematical model and a set of control experiments (Ruijter et al. 2009; Lievens et al. 2012; Zhao and Fernald 2005; Peirson et al. 2003; Tichopad et al. 2003; Boggy and Woolf 2010; Guescini et al. 2008; Ruijter et al. 2013). Quantification by PCR is routinely exploited in many applications, including analysis of forensic evidence (Nicklas and Buel 2003; Bauer 2007), monitoring of food safety (Elizaquível et al. 2014), and clinical diagnostics (Kaltenboeck and Wang 2005; Bustin et al. 2021).

The accuracy and precision of the quantification process is limited by the mathematical model relating DNA content to fluorescence. Current models possess subjective and systematic bias and do not account for the uncertainty in fluorescence that arises from imperfect amplification and pipetting errors.

Systematic bias originates from assuming that the initial DNA type is double-stranded and that the fluorescence increases each time *either* complementary strand is replicated. The former is obviously not true when the initial DNA is produced by reverse-transcription of single-stranded RNA (RT, as in RT-qPCR), as only *one* of the complementary DNA strands is present at the beginning of PCR. The second statement is not true for common probes that possess a fluorophore covalently attached to an oligonucleotide. Since the oligonucleotide only hybridizes to *one* of the complementary strands, the fluorescence only increases when *one* of the complementary strands is replicated. The second assumption also does not appear to be true for fluorescent dyes that bind non-specifically to DNA, as the amount of dye bound to each DNA strand depends on the amount of DNA in solution.

The impact of several of these assumptions was assessed by Ruijter et al. (2014). The authors found that, depending on the initial DNA type and fluorescent probe chemistry, the background-subtracted fluorescence could differ by up to a factor of 2 in the exponential phase. However, the authors' analysis was rooted in the assumption of perfect amplification. They also noted that what actually occurs during the first few cycles of PCR is unknown.

Another source of systematic bias in the mathematical description of fluorescence arises when the initial DNA content is very small. Current approaches are deterministic and do not take into account the fact that the number of DNA strands is an integer. While the kinetics of PCR have been investigated in the framework of stochastic

branching processes (Nedelman et al. 1992; Sun 1995; Weiss and von Haeseler 1995; Stolovitzky and Cecchi 1996; Jacob and Peccoud 1996b, a), the first of which was published in this journal, such models have not been linked to the fluorescence reported by probes. Like the deterministic approaches described above, these stochastic models neither discriminate between complementary strands nor describe initial conditions encountered in RT-qPCR.

A mathematical model that discriminates between complementary DNA strands can investigate another source of bias: the assumption that the efficiency of synthesis is independent of directionality (i.e., reverse or forward). Since primers are complementary to different ends of the target sequence, and are specifically chosen not to be complementary to each other (i.e., avoiding dimerization), the formation of one primer–target complex may be more efficient than the other. In addition, the yield of the strand whose replication is being monitored by the fluorescent probe may be affected by the monitoring process. These arguments are also supported by the fact that optimal concentrations of each primer can be different (see, for example, Bustin (2004), where the two concentrations differ by a factor of 3).

To address these challenges, we present a two-type stochastic branching process model in Sect. 2.1 that differentiates between complementary strands and amplification probabilities. Analysis of the expected value in Sect. 2.2 identifies a new timescale that is prevalent during the first few cycles. This timescale explains some of the unknown behavior that occurs during the first few cycles of PCR, explaining some of the aforementioned unknown behavior. At short times, there is a lag in exponential growth where the ratio of expected strand counts changes from its initial to critical value. The critical ratio is related to the amplification probability of each complementary strand, being unity when the probabilities are identical. The analysis also demonstrates that the popular parameter describing PCR efficiency (or amplification probability) is really the geometric mean of the efficiencies of both complementary strands.

Quantification by real-time PCR is also limited by a subjective and empirical description of the fluorescence that is not associated with amplification, or the background fluorescence. The description of background fluorescence is usually taken *a posteriori* (i.e., after measurements of fluorescence monitoring amplification). Without a clear connection to the chemical and physical processes occurring in solution, the background fluorescence is often assumed to be a linear function of cycle, or a ‘baseline.’

In Sect. 3.1, we address these concerns by using the fluorescence analog of Beer’s Law to relate fluorescence to the concentration of each fluorescent species. While such expressions have often been used to describe the fluorescence of dyes interacting with known amounts of DNA (Biver et al. 2003, 2005), we are not aware of any adaptation to real-time PCR. We discriminate between the fluorescent species by referring to the form present before PCR as the inactive species and the form activated by PCR as the active species. For hydrolysis probes, we show how the relevant parameters can be extracted from a few control experiments (see Sect. 3.2). In contrast to other approaches, we can quantify the validity of the model of background fluorescence. The relevant contributions can be determined without interrogating or adjusting fluorescence data associated with amplification. We find that the model agrees well with

experiment and observe that the incremental increase in fluorescence is not independent of cycle.

A final limitation of real-time PCR is the lack of a mathematical expression relating errors arising from pipetting and imperfect amplification to uncertainty in fluorescence. To quantify the variance in copy number (Peccoud and Jacob 1996), we investigate the stochastic branching process (Sect. 2.3) in a manner similar to previous reports analyzing error in high-throughput sequencing (Kebschull and Zador 2015; Schwabe and Falcke 2022). After validating that the fluorescence parameter (the fluorescence per mole) for each species is approximately constant for each cycle and well (Sect. 3.2), the models for PCR and fluorescence are combined (Sect. 3.3). This yields analytical expressions of the first two central moments of fluorescence in terms of reaction efficiencies, input content, and input type (i.e., double-stranded DNA, forward-stranded RNA, and reverse-stranded RNA).

Together with the parameters determined from experiment, fluorescence curves computed with uncertainty identify regimes under which certain sources of error are more prevalent than others (see Sect. 3.3). When the expected initial-strand-number is sufficiently large, or the cycle number sufficiently small, the error in fluorescence in a specific well is less than the well-to-well variation in expected value. As the initial strand count decreases and the fluorescence rises above initial levels, however, the variance in input copy-number and imperfect amplification become the dominant contributions to error. Finally, in Sect. 3.4, we use the fluorescence model to develop analytical expressions for the limit of detection as a function of amplification efficiency and nucleic acid type. These expressions may be particularly useful for application in epidemic diseases, as false positives or false negatives may be instead termed inconclusive.

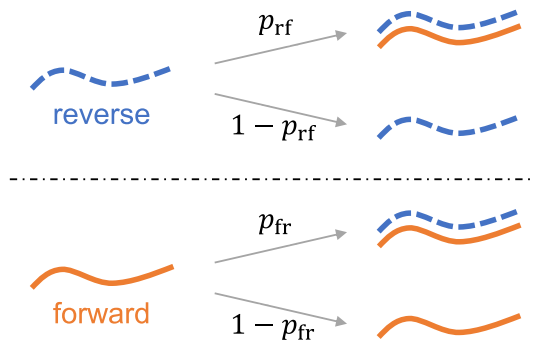
## 2 Strand-Specific Branching Process

In this section, we model PCR as a two-type branching-process. The model distinguishes between complementary DNA strands and amplification efficiencies. We then derive analytical expressions relating the first two central-moments of strand counts before PCR to those after each cycle has completed. Compartmentalizing DNA amplification and fluorescence, the linking of the two phenomena is postponed until Sect. 3.

### 2.1 Mathematical Model

PCR consists of a series of  $n$  cycles, with  $n$  usually ranging between 35 and 50. Each cycle consists of a melting, annealing, and elongation step to synthesize new DNA from existing DNA (i.e., a chain reaction). A variety of resources on PCR are available online for further information (e.g., National Institutes of Health, National Human Genome Research Institute 2023).

To distinguish between the two complementary strands of DNA, we refer to one as the forward strand and the other as the reverse strand. We let the discrete random variables  $X_i$  and  $Y_i$  represent the number of forward and reverse strands, respectively,



**Fig. 1** Illustration of strand-specific model of DNA amplification. (top) After each cycle  $i$ , the outcome of synthesis of a forward strand (orange, solid) from each reverse strand (blue, dashed) present in the previous cycle is modeled as a Bernoulli random-variable with probability of success  $p_{rf}$ . (bottom) Similarly, the outcome of synthesis of a reverse strand (blue, dashed) from each forward strand (orange, solid) present in the previous cycle is modeled as a Bernoulli random-variable with probability of success  $p_{fr}$  (Color figure online)

present after  $i = 0$  to  $n$  cycles have been completed. We represent the strand counts after completion of  $i$  cycles as the random vector

$$U_i = \begin{pmatrix} X_i \\ Y_i \end{pmatrix},$$

and refer to  $U_0$  as the initial strand count. However, we distinguish  $U_0$  from the strand count *input* to the reaction mixture, denoting the latter by the random vector  $\begin{pmatrix} I_X \\ I_Y \end{pmatrix}$ , where  $I_X$  and  $I_Y$  are discrete random variables representing the number of forward and reverse strands input, respectively. This is important to distinguish between assays targeting DNA and RNA sequences, as discussed in Sect. 2.1.2.

### 2.1.1 Relationship Between Consecutive Cycles

After completing  $i - 1$  cycles, the biochemical events occurring during the next cycle involve the attempt to produce one forward strand from each of the  $Y_{i-1}$  reverse strands and the attempt to produce one reverse strand from each of the  $X_{i-1}$  forward strands (see Fig. 1). The outcome of synthesis of a forward strand from each reverse strand is modeled as a Bernoulli random-variable with probability of success  $p_{rf} \in (0, 1)$ . Similarly, the outcome of synthesis of a reverse strand from each forward strand is modeled as a Bernoulli random-variable with probability of success  $p_{fr} \in (0, 1)$ . (The subscripts *rf* and *fr* denote the direction reverse-to-forward and forward-to-reverse, respectively.) This corresponds to the mathematical model

$$U_i = U_{i-1} + \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} B(U_{i-1}; p_{fr}, p_{rf}), \tag{1}$$

after completing  $i$  cycles, where

$$\mathbf{B}(\mathbf{U}_{i-1}; p_{fr}, p_{rf}) := \begin{bmatrix} \mathbf{B}(X_{i-1}; p_{fr}) \\ \mathbf{B}(Y_{i-1}; p_{rf}) \end{bmatrix},$$

and  $\mathbf{B}(a; b)$  denotes a Binomial random-variable of  $a$  trials with probability of success  $b$ . All Bernoulli trials, whether associated with the outcome of synthesis of a forward or reverse strand, are taken to be independent.

To compare to previous approaches that do not discriminate between complementary strands and efficiencies, we will use

$$N_i = X_i + Y_i$$

to denote the total number of strands after  $i$  cycles have been completed. We will see that an appropriate characterization of the average amplification efficiency of both complementary strands is

$$\bar{p} := \sqrt{p_{rf}p_{fr}}, \tag{2}$$

and that an appropriate parameter for the deviation in efficiencies is

$$R := \sqrt{\frac{p_{rf}}{p_{fr}}}. \tag{3}$$

To avoid changing notation, we will subsequently only investigate  $p_{fr} = \frac{\bar{p}}{R}$  and  $p_{rf} = \bar{p}R$  in terms of  $\bar{p}$  and  $R$ , so that

$$\mathbf{B}(\mathbf{U}_{i-1}; p_{fr}, p_{rf}) = \mathbf{B}\left(\mathbf{U}_{i-1}; \frac{\bar{p}}{R}, \bar{p}R\right) = \begin{bmatrix} \mathbf{B}\left(X_{i-1}; \frac{\bar{p}}{R}\right) \\ \mathbf{B}(Y_{i-1}; \bar{p}R) \end{bmatrix}.$$

### 2.1.2 Relationship Between Initial and Input Condition

The relationship between the input number  $\begin{pmatrix} I_X \\ I_Y \end{pmatrix}$  and initial number  $\mathbf{U}_0$  of strands depends on whether the nucleic acids input to the reaction mixture are forward-stranded RNA (fs-RNA, referred to as Case RF), reverse-stranded RNA (rs-RNA, referred to as Case RR), or double-stranded DNA (ds-DNA, which consists of both fs-DNA and rs-DNA, referred to as Case D). If ds-DNA is generated by transferring rs-DNA and fs-DNA separately into the reaction mixture,  $X_0$  and  $Y_0$  can be modeled as independent and identically distributed (i.i.d.). Here,

Case D : 
$$\mathbf{U}_0 = \begin{pmatrix} X_0 \\ Y_0 \end{pmatrix} = \begin{pmatrix} I_X \\ I_Y \end{pmatrix}, \quad I_X \text{ (or } I_Y \text{) represents number of fs-DNA (or rs-DNA) strands.}$$

When RNA is input to the reaction mixture, on the other hand, the nucleic acids all possess the same strandedness (i.e., they are all fs-RNA or rs-RNA). The RT step yields DNA strands that are complementary to the RNA. As all RNA strands are fs-RNA (or rs-RNA), the RT step yields rs-DNA (or fs-DNA). Modeling the outcome of synthesis of each cDNA from each RNA strand as a Bernoulli random-variable with probability of success  $r \in (0, 1)$ ,  $U_0$  is related to  $\begin{pmatrix} I_X \\ I_Y \end{pmatrix}$  via either

$$\text{Case RF : } U_0 = \begin{pmatrix} X_0 \\ Y_0 \end{pmatrix} = \begin{bmatrix} 0 \\ \mathbf{B}(I_X; r) \end{bmatrix}, \quad \begin{array}{l} I_X \text{ represents number of} \\ \text{fs-RNA strands, } I_Y = 0, \end{array}$$

or

$$\text{Case RR : } U_0 = \begin{pmatrix} X_0 \\ Y_0 \end{pmatrix} = \begin{bmatrix} \mathbf{B}(I_Y; r) \\ 0 \end{bmatrix}, \quad \begin{array}{l} I_Y \text{ represents number of} \\ \text{rs-RNA strands, } I_X = 0. \end{array}$$

In comparison to the conventional PCR amplification efficiency  $p := p_{fr} = p_{rf}$  when  $R = 1$ , which is usually between 0.8 and 0.99, the RT efficiency  $r$  can adopt a relatively large range of values (Bustin et al. 2015; Schwaber et al. 2019).

The nucleic acids are input to the reaction mixture by transferring liquids from one container to another using a pipette. Since the process of transferring such liquids is independent of the type of nucleic acids (i.e., independent of whether they are fs-RNA, rs-RNA, fs-DNA, or rs-DNA), the number of strands of each type input to the reaction mixture obey the same distribution. We will let this distribution be obeyed by the discrete random variable  $I$ . As a result, it follows that

$$\begin{cases} I_X, I_Y \stackrel{\text{i.i.d.}}{\sim} I, & \text{Case D,} \\ I_X \stackrel{\text{i.d.}}{\sim} I, \quad I_Y = 0, & \text{Case RF,} \\ I_X = 0, \quad I_Y \stackrel{\text{i.d.}}{\sim} I, & \text{Case RR,} \end{cases}$$

where i.d. denotes identically distributed.

### 2.2 Expected Value

In this section, we derive relationships between  $\mathbb{E}[I]$  and expected copy-numbers after  $i$  cycles have been completed. After using the total law of expectation to obtain a relationship in expected copies between two sequential cycles, we use induction to generate the desired results. We first investigate the conventional branching process for PCR which results from assuming  $R = 1$  and does not distinguish between fs-DNA and rs-DNA. Subsequently, we consider the two-type branching process (1) where  $R$  may not be 1 and fs-DNA is distinguished from rs-DNA.

### 2.2.1 Conventional Branching Process

The conventional branching process model for PCR occurs when  $R = 1$ , implying from (3) that  $p_{\text{rf}} = p_{\text{fr}} =: p$ . In this case, the system (1) can be summed to yield

$$N_i = N_{i-1} + B(N_{i-1}; p), \tag{4}$$

as has been investigated elsewhere (Nedelman et al. 1992; Sun 1995; Weiss and von Haeseler 1995; Peccoud and Jacob 1996; Jacob and Peccoud 1996b, a; Stolovitzky and Cecchi 1996). Using the law of total expectation and (4), one finds that

$$\mathbb{E}[N_i] = \mathbb{E}[\mathbb{E}[N_i | N_{i-1}]] = \mathbb{E}[N_{i-1}](1 + p), \tag{5}$$

for any two consecutive cycles. From induction, it follows that (5) is identical to

$$\mathbb{E}[N_i] = \mathbb{E}[N_0](1 + p)^i, \tag{6}$$

as reported elsewhere (Nedelman et al. 1992; Sun 1995; Weiss and von Haeseler 1995; Peccoud and Jacob 1996; Jacob and Peccoud 1996b, a; Stolovitzky and Cecchi 1996).

### 2.2.2 Strand-Specific Branching Process

Below, we develop relationships between  $\mathbb{E}[I]$  and each  $\mathbb{E}[U_i]$  for the more general case of (1) where forward strands are distinguished from reverse strands and  $R$  may not be 1. The law of total expectation and (1) lead to the relation

$$\begin{aligned} \mathbb{E}[U_i] &= \mathbb{E}[\mathbb{E}[U_i | U_{i-1}]] = \mathbb{E}\left\{U_{i-1} + \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \mathbb{E}\left[\mathbf{B}\left(U_{i-1}; \frac{\bar{p}}{R}, \bar{p}R\right) \mid U_{i-1}\right]\right\} \\ &= \mathbf{A}\mathbb{E}[U_{i-1}] = \mathbf{A}^i\mathbb{E}[U_0], \end{aligned} \tag{7}$$

where  $\mathbf{A}$  is defined as

$$\mathbf{A} = \begin{pmatrix} 1 & \bar{p}R \\ \bar{p}/R & 1 \end{pmatrix},$$

and the last step follows from induction. The relationship between  $\mathbb{E}[U_0]$  in (7) depends on  $\mathbb{E}[I]$  through

$$\mathbb{E}[U_0] = \begin{cases} \begin{pmatrix} \mathbb{E}[I_X] \\ \mathbb{E}[I_Y] \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \mathbb{E}[I], & \text{Case D;} \\ \begin{pmatrix} 0 \\ r\mathbb{E}[I_X] \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} r\mathbb{E}[I], & \text{Case RF;} \\ \begin{pmatrix} r\mathbb{E}[I_Y] \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} r\mathbb{E}[I], & \text{Case RR;} \end{cases} \tag{8}$$



which follows from  $\mathbb{E}[U_0] = \mathbb{E}[\mathbb{E}[U_0 | I_X, I_Y]]$  and the case-by-case relationships presented in Sect. 2.1.2.

We will see that the matrix  $A$  plays a central role in dynamics of the first two central-moments of strand counts.  $A$  has two distinct eigenvalues,

$$\begin{aligned} \lambda_1 &:= 1 + \bar{p}, \\ \lambda_2 &:= 1 - \bar{p}, \end{aligned}$$

and can be decomposed as

$$A = X\Lambda Z = \sum_{j=1}^2 \lambda_j \mathbf{x}_j \mathbf{z}_j^\top, \tag{9}$$

where

$$X := (\mathbf{x}_1 \ \mathbf{x}_2) := \frac{1}{\sqrt{2}} \begin{pmatrix} R & R \\ 1 & -1 \end{pmatrix}, \quad \Lambda := \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix},$$

and

$$Z := X^{-1} = \frac{1}{R\sqrt{2}} \begin{pmatrix} 1 & R \\ 1 & -R \end{pmatrix} =: \begin{pmatrix} \mathbf{z}_1^\top \\ \mathbf{z}_2^\top \end{pmatrix}.$$

If  $R = 1$ , the scale factor  $\frac{1}{\sqrt{2}}$  in the definition of  $X$  implies that

$$\|\mathbf{x}_1\| = \|\mathbf{x}_2\| = \|\mathbf{z}_1\| = \|\mathbf{z}_2\| = 1, \tag{10}$$

where  $\|\cdot\|$  denotes the Euclidean norm. If  $R \neq 1$ , on the other hand, no scale factor can be chosen to satisfy (10).

Since  $Z := X^{-1}$ ,

$$\mathbf{x}_i^\top \mathbf{z}_j = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{otherwise,} \end{cases} \tag{11}$$

for  $i, j \in \{1, 2\}$ , and (after substitution of (9))

$$A^i = \left( \sum_{j=1}^2 \lambda_j \mathbf{x}_j \mathbf{z}_j^\top \right)^i = \sum_{j=1}^2 \lambda_j^i \mathbf{x}_j \mathbf{z}_j^\top. \tag{12}$$

Substituting (12) into (7) leads to

$$\begin{aligned} \mathbb{E}[U_i] &= z_1^\top \mathbb{E}[U_0] \lambda_1^i x_1 + z_2^\top \mathbb{E}[U_0] \lambda_2^i x_2 \\ &= \mathbb{E}\left[\frac{X_0 + RY_0}{2}\right] \lambda_1^i \begin{pmatrix} 1 \\ R^{-1} \end{pmatrix} + \mathbb{E}\left[\frac{X_0 - RY_0}{2}\right] \lambda_2^i \begin{pmatrix} 1 \\ -R^{-1} \end{pmatrix}, \end{aligned} \tag{13}$$

where second expression aids in physical interpretation (below) of Case D, RF, and RR.

Where Equation (6) has  $1 + p$ , Equation (13) has two values  $\lambda_1$  and  $\lambda_2$ . In the case that  $R = 1$ ,  $\lambda_1 := 1 + \bar{p} = 1 + p$ , as is present in (6). The eigenvalue  $\lambda_2$  does not have an analog in (6).

The quantities  $X_i + RY_i$  and  $X_i - RY_i$  in the second expression of (13) arise frequently in the investigation of the first two central-moments of  $U_i$ . The quantity  $X_i + RY_i$  is referred to as the weighted sum of strand counts after  $i$  cycles have been completed, while  $X_i - RY_i$  is referred to as the weighted difference. Multiplying each side of the equation (13) by the row vector  $\begin{pmatrix} 1 \\ \pm R \end{pmatrix}^\top$  demonstrates<sup>1</sup> that

$$\begin{cases} \mathbb{E}[X_i + RY_i] = \mathbb{E}[X_0 + RY_0] \lambda_1^i, \\ \mathbb{E}[X_i - RY_i] = \mathbb{E}[X_0 - RY_0] \lambda_2^i. \end{cases}$$

As  $\lambda_1 > 1$ , the expected weighted-sum exhibits exponential growth. As  $\lambda_2 < 1$ , on the other hand, the expected weighted-difference exhibits exponential decay.

The term in (13) associated with  $\lambda_2$  is always present when the input is RNA (i.e., for Case RF or RR and any  $\bar{p} \in (0, 1)$  and  $R \in (0, \infty)$ ), as  $\mathbb{E}[X_0] \neq R\mathbb{E}[Y_0]$ . However, it may not be present if the input is DNA, as the term vanishes ( $\mathbb{E}[X_0 - RY_0] = (1 - R)\mathbb{E}[I]$ ) for Case D when  $R = 1$ .

When  $\mathbb{E}[X_0] \neq R\mathbb{E}[Y_0]$ , the term involving  $\lambda_2^i$  is usually negligible after a few cycles, as  $\bar{p}$  is usually above 0.8. After the lag time is over ( $i$  is large enough for  $\lambda_2^i$  to be negligible), the *ratio* of expected forward to reverse strand counts reaches its critical value, as

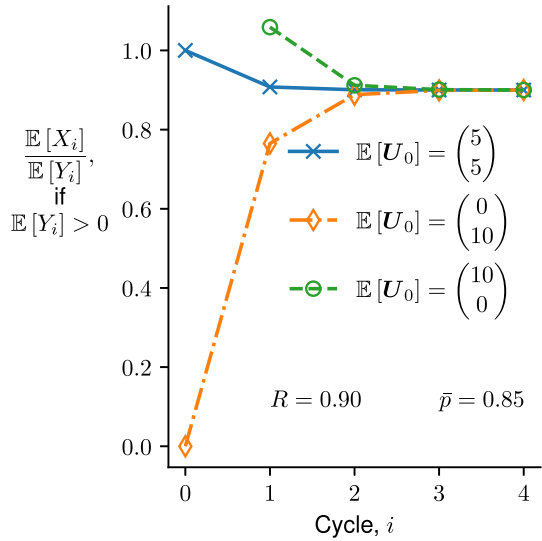
$$\lim_{i \rightarrow \infty} \frac{\mathbb{E}[X_i]}{\mathbb{E}[Y_i]} = \lim_{i \rightarrow \infty} \frac{e_1^\top \mathbb{E}[U_i]}{e_2^\top \mathbb{E}[U_i]} = \frac{e_1^\top x_1}{e_2^\top x_1} = R, \tag{14}$$

where  $e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$  and  $e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$  are the standard unit vectors. An illustration of the transition of  $\frac{\mathbb{E}[X_i]}{\mathbb{E}[Y_i]}$  (when  $\mathbb{E}[Y_i] > 0$ ) to  $R$  is depicted in Fig. 2.

<sup>1</sup> The expression which follows could have also been developed by substituting the first expression of (9) into the final expression of (7) and left-multiplying by  $Z$ , yielding

$$Z\mathbb{E}[U_i] = \Lambda^i Z\mathbb{E}[U_0].$$

**Fig. 2** Visualization of convergence of the ratio of expected-forward to expected-reverse strands to  $R$ . The initial condition  $\mathbb{E}[U_0]$  is taken from (8) with Case D, RF, or RR. The values of  $\bar{p}$  and  $R$  are annotated, while  $r$  and  $\mathbb{E}[I]$  are chosen so that  $\mathbb{E}[N_0] = \mathbb{E}[X_0 + Y_0] = 10$



To compare the two-type branching process to (6), we compute from (13)

$$\begin{aligned} \mathbb{E}[N_i] &= \mathbb{E}[X_i + Y_i] = \begin{pmatrix} 1 \\ 1 \end{pmatrix}^\top \mathbb{E}[U_i] \\ &= \mathbf{z}_1^\top \mathbb{E}[U_0] \lambda_1^i \begin{pmatrix} 1 \\ 1 \end{pmatrix}^\top \mathbf{x}_1 + \mathbf{z}_2^\top \mathbb{E}[U_0] \lambda_2^i \begin{pmatrix} 1 \\ 1 \end{pmatrix}^\top \mathbf{x}_2 \\ &= \mathbb{E}\left[\frac{X_0 + RY_0}{2}\right] \lambda_1^i \left(\frac{R+1}{R}\right) + \mathbb{E}\left[\frac{X_0 - RY_0}{2}\right] \lambda_2^i \left(\frac{R-1}{R}\right), \end{aligned}$$

and, for Case D,

$$\mathbb{E}[N_i] = \mathbb{E}[N_0] \left[ \frac{(R+1)^2}{4R} \lambda_1^i - \frac{(R-1)^2}{4R} \lambda_2^i \right], \tag{15}$$

as (8) implies that  $\mathbb{E}[I] = \mathbb{E}[X_0] = \mathbb{E}[X_0 + Y_0]/2 = \mathbb{E}[N_0]/2$ . Equation (15) demonstrates that the two approaches are equal if  $R = 1$ . Otherwise, the absolute difference between  $\mathbb{E}[N_i]$  calculated from (6) and (15) increases exponentially with increasing  $i$ . In the future, determining  $R$  and using (15) instead of (6) may therefore be important to reduce bias in quantification.

On the other hand, when  $R = 1$ , the expected amount of each strand is eventually independent of the composition at the start of PCR. That is,

$$\lim_{i \rightarrow \infty} \frac{\mathbb{E}[X_i]}{\lambda_1^i} = \lim_{i \rightarrow \infty} \frac{\mathbb{E}[Y_i]}{\lambda_1^i} = \lim_{i \rightarrow \infty} \frac{\mathbb{E}[N_i]}{2\lambda_1^i} = \frac{\mathbb{E}[N_0]}{2}, \quad \text{if } R = 1, \tag{16}$$

only depends on the expected initial-sum, or  $\mathbb{E}[N_0]$ . This result is in contrast to the report of Ruijter et al. (2014), who used a deterministic model with perfect amplification efficiency. When  $R = 1$  and the input is RNA, the directionality of RNA (i.e., fs-RNA or rs-RNA) only dictates the characteristic length of the lag time.

## 2.3 Variance

In this section, we derive the relationship between  $\text{Var}[I]$  and the variance after  $i$  cycles have been completed. The procedure is similar to the previous section, except the law of total variance is used instead of the law of total expectation.

### 2.3.1 Conventional Branching Process

Using the law of total variance and (4), one obtains

$$\begin{aligned}\text{Var}[N_i] &= \text{Var}[\mathbb{E}[N_i | N_{i-1}]] + \mathbb{E}[\text{Var}[N_i | N_{i-1}]] \\ &= \text{Var}[N_{i-1}](1+p)^2 + \mathbb{E}[N_{i-1}]p(1-p).\end{aligned}\quad (17)$$

Induction can be used to show that (17) is equivalent to

$$\text{Var}[N_i] = \text{Var}[N_0](1+p)^{2i} + p(1-p)(1+p)^{2(i-1)} \sum_{j=0}^{i-1} \frac{\mathbb{E}[N_j]}{(1+p)^{2j}}. \quad (18)$$

Substitution of (6), simplification of the resultant geometric series, and rearrangement leads to the expression

$$\begin{aligned}\text{Var}[N_i] &= \left\{ \text{Var}[N_0] + \mathbb{E}[N_0] \left( \frac{1-p}{1+p} \right) \right\} (1+p)^{2i} \\ &\quad - \mathbb{E}[N_0](1-p)(1+p)^{i-1}.\end{aligned}\quad (19)$$

When  $\text{Var}[N_0] = 0$ , Equation (19) becomes identical to what has been reported elsewhere (Sun 1995; Weiss and von Haeseler 1995; Jacob and Peccoud 1996b,a; Stolovitzky and Cecchi 1996) (Nedelman et al. (1992); Peccoud and Jacob (1996) report the leading-order approximation for large  $i$ ). The more general result, Equation (19), demonstrates that  $\text{Var}[N_0]$  can significantly impact  $\text{Var}[N_i]$ , as it is part of the dominant term.

The growth in variance with increasing  $i$  by an exponent twice that of the expected value explains why very large cycles, where the expected copy-number is also very large, are not of interest. It also explains why quantification by real-time PCR is more reproducible than end-point PCR.

### 2.3.2 Strand-Specific Branching Process

The variance–covariance matrix of  $\mathbf{U}_i$  is defined as

$$\text{Var} [\mathbf{U}_i] = \mathbb{E} \left[ \mathbf{U}_i \mathbf{U}_i^\top \right] - \mathbb{E} [\mathbf{U}_i] \mathbb{E} [\mathbf{U}_i]^\top,$$

for each  $i$ . From the law of total variance,

$$\text{Var} [\mathbf{U}_i] = \text{Var} \left[ \mathbb{E} [\mathbf{U}_i \mid \mathbf{U}_{i-1}] \right] + \mathbb{E} \left[ \text{Var} [\mathbf{U}_i \mid \mathbf{U}_{i-1}] \right].$$

In a manner similar to (7), it follows that

$$\text{Var} \left[ \mathbb{E} [\mathbf{U}_i \mid \mathbf{U}_{i-1}] \right] = \text{Var} [\mathbf{A} \mathbf{U}_{i-1}] = \mathbf{A} \text{Var} [\mathbf{U}_{i-1}] \mathbf{A}^\top.$$

In addition, after substitution of (1), it follows that

$$\begin{aligned} \mathbb{E} \left[ \text{Var} [\mathbf{U}_i \mid \mathbf{U}_{i-1}] \right] &= \mathbb{E} \left\{ \text{Var} \left[ \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \mathbf{B} \left( \mathbf{U}_{i-1}; \frac{\bar{p}}{R}, \bar{p}R \right) \mid \mathbf{U}_{i-1} \right] \right\} \\ &= \mathbb{E} \left[ \text{Var} \left\{ \begin{bmatrix} \mathbf{B} (Y_{i-1}; \bar{p}R) \\ \mathbf{B} \left( X_{i-1}; \frac{\bar{p}}{R} \right) \end{bmatrix} \mid \mathbf{U}_{i-1} \right\} \right] \\ &= \begin{bmatrix} \bar{p}R (1 - \bar{p}R) \mathbb{E} [Y_{i-1}] & 0 \\ 0 & \frac{\bar{p}}{R} \left( 1 - \frac{\bar{p}}{R} \right) \mathbb{E} [X_{i-1}] \end{bmatrix}. \end{aligned}$$

Combining the two expressions, we obtain the two-type analog of (17),

$$\begin{aligned} \text{Var} [\mathbf{U}_i] &= \mathbf{A} \text{Var} [\mathbf{U}_{i-1}] \mathbf{A}^\top \\ &\quad + \begin{bmatrix} \bar{p}R (1 - \bar{p}R) \mathbb{E} [Y_{i-1}] & 0 \\ 0 & \frac{\bar{p}}{R} \left( 1 - \frac{\bar{p}}{R} \right) \mathbb{E} [X_{i-1}] \end{bmatrix}. \end{aligned} \tag{20}$$

Before using induction to relate  $\text{Var} [\mathbf{U}_i]$  to  $\text{Var} [\mathbf{U}_0]$ , it is useful to simplify (20) by substituting  $\mathbb{E} [Y_{i-1}] = \mathbf{e}_2^\top \mathbb{E} [\mathbf{U}_{i-1}]$  and  $\mathbb{E} [X_{i-1}] = \mathbf{e}_1^\top \mathbb{E} [\mathbf{U}_{i-1}]$  with  $\mathbb{E} [\mathbf{U}_{i-1}]$  provided by the first expression of (13). Collecting terms multiplying each eigenvalue, we find that

$$\text{Var} [\mathbf{U}_i] = \mathbf{A} \text{Var} [\mathbf{U}_{i-1}] \mathbf{A}^\top + \sum_{\ell=1}^2 \lambda_\ell^{i-1} \mathbf{K}_\ell, \tag{21}$$

where

$$\mathbf{K}_\ell := \begin{bmatrix} \bar{p}R (1 - \bar{p}R) \mathbf{e}_2^\top \mathbf{x}_\ell & 0 \\ 0 & \frac{\bar{p}}{R} \left( 1 - \frac{\bar{p}}{R} \right) \mathbf{e}_1^\top \mathbf{x}_\ell \end{bmatrix} \mathbf{z}_\ell^\top \mathbb{E} [\mathbf{U}_0]. \tag{22}$$

From induction, it follows that (21) is identical to

$$\text{Var}[U_i] = A^i \text{Var}[U_0] (A^\top)^i + \sum_{\ell=1}^2 \sum_{q=0}^{i-1} \lambda_\ell^q A^{i-1-q} \mathbf{K}_\ell (A^\top)^{i-1-q}. \tag{23}$$

Substitution of (12) and simplification of the resultant geometric series leads to (see Sect. B.1)

$$\text{Var}[U_i] = \sum_{j=1}^2 \sum_{k=1}^2 \left\{ \left[ v_{j,k} + \sum_{\ell=1}^2 \eta_{j,k}^\ell \right] (\lambda_j \lambda_k)^i - \sum_{\ell=1}^2 \eta_{j,k}^\ell \lambda_\ell^i \right\} \mathbf{x}_j \mathbf{x}_k^\top, \tag{24}$$

where

$$v_{j,k} := \mathbf{z}_j^\top \text{Var}[U_0] \mathbf{z}_k, \tag{25a}$$

$$\eta_{j,k}^\ell := \frac{\mathbf{z}_j^\top \mathbf{K}_\ell \mathbf{z}_k}{\lambda_j \lambda_k - \lambda_\ell}. \tag{25b}$$

In (24), the dependence of  $\text{Var}[U_i]$  on  $I$  arises through  $\text{Var}[U_0]$  in each  $v_{j,k}$  and  $\mathbb{E}[U_0]$  in each  $\eta_{j,k}^\ell$  (through (22)). While the dependence of  $\mathbb{E}[U_0]$  on  $\mathbb{E}[I]$  is given in (8), the relationship between  $\text{Var}[U_0]$  and  $\text{Var}[I]$  for Case D, RF, or RR is given by

$$\text{Var}[U_0] = \begin{cases} \begin{pmatrix} \text{Var}[I_X] & 0 \\ 0 & \text{Var}[I_Y] \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{Var}[I], & \text{Case D;} \\ \begin{pmatrix} 0 & 0 \\ 0 & \text{Var}[\mathbf{B}(I_X; r)] \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \text{Var}[\mathbf{B}(I; r)], & \text{Case RF;} \\ \begin{pmatrix} \text{Var}[\mathbf{B}(I_Y; r)] & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \text{Var}[\mathbf{B}(I; r)], & \text{Case RR,} \end{cases} \tag{26a}$$

$$\begin{aligned} \text{Var}[\mathbf{B}(I; r)] &= \text{Var}[\mathbb{E}[\mathbf{B}(I; r) | I]] + \mathbb{E}[\text{Var}[\mathbf{B}(I; r) | I]] \\ &= \text{Var}[I] r^2 + r(1-r) \mathbb{E}[I] \\ &= \text{Var}[\mathbf{B}(I_X; r)] = \text{Var}[\mathbf{B}(I_Y; r)], \end{aligned} \tag{26b}$$

where (26b) utilizes the law of total variance. Since  $X_0$  and  $Y_0$  are independent for any case,  $\text{Cov}[X_0, Y_0] = 0$ . As such, we will often use the substitution

$$\text{Var}[U_0] = \begin{pmatrix} \text{Var}[X_0] & 0 \\ 0 & \text{Var}[Y_0] \end{pmatrix} \tag{27}$$

without specifying a particular case.

In contrast to (19), which only has terms proportional to  $(1+p)^{2i}$  and  $(1+p)^{i-1}$ , Equation (24) has terms proportional to  $\lambda_1^{2i} > \lambda_1^i > (\lambda_1 \lambda_2)^i > \lambda_2^i > \lambda_2^{2i}$  (if  $i > 0$ ).

While physical interpretation of all terms can be complicated, it is useful to examine the leading-order expression as  $i \rightarrow \infty$ ,

$$\text{Var}[U_i] = \left( v_{1,1} + \eta_{1,1}^{(1)} + \eta_{1,1}^{(2)} \right) \lambda_1^{2i} \mathbf{x}_1 \mathbf{x}_1^\top + \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} O\left(\lambda_1^i\right), \tag{28}$$

and compute the terms explicitly as

$$v_{1,1} = \text{Var} \left[ \frac{X_0 + RY_0}{2} \right] \frac{2}{R^2}, \tag{29a}$$

$$\eta_{1,1}^{(1)} = \mathbb{E} \left[ \frac{X_0 + RY_0}{2} \right] \left( \frac{\lambda_2}{\lambda_1} \right) \frac{R + 1}{2R^2}, \tag{29b}$$

$$\eta_{1,1}^{(2)} = \mathbb{E} \left[ \frac{X_0 - RY_0}{2} \right] \left( \frac{\bar{p}\lambda_1}{\lambda_1^2 - \lambda_2} \right) \frac{R - 1}{2R^2}, \tag{29c}$$

where we have used (27). Note that the moments of  $(X_0 \pm RY_0)/2$  arise in (28) as they did in (13). The term  $v_{1,1}$  represents the contribution from the variance at the start of PCR. The term  $\eta_{1,1}^{(1)}$  accounts for the variance due to imperfect amplification, as  $\bar{p} \approx 1$  implies that  $\lambda_2 \approx 0 \approx \eta_{1,1}^{(1)}$ . If  $R = 1$ ,

$$\eta_{1,1}^{(1)} = \frac{\mathbb{E}[N_0]}{2} \left( \frac{1 - p}{1 + p} \right),$$

one-half the second term in brackets of (19). The third term  $\eta_{1,1}^{(2)}$  does not have a counterpart in (19). It accounts for differences in strand-specific amplification, as it vanishes when  $R \rightarrow 1$ .

The expressions developed for expected value and variance can be used to produce expressions in the squared coefficients of variation, defined as

$$\text{CV}[U_i]^2 = \text{Var}[U_i] \oslash \mathbb{E}[U_i] \mathbb{E}[U_i]^\top, \tag{30}$$

where  $\oslash$  denotes element-wise division. Substitution of (13) and (28) into (30) yields

$$\text{CV}[U_i]^2 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \left\{ \frac{v_{1,1} + \eta_{1,1}^{(1)} + \eta_{1,1}^{(2)}}{\left( \mathbf{z}_1^\top \mathbb{E}[U_0] \right)^2} + O\left(\lambda_1^{-i}\right) \right\}. \tag{31}$$

As the term  $O\left(\lambda_1^{-i}\right)$  rapidly approaches zero with increasing  $i$ , and  $\lambda_1$  is usually more than 1.8, the dominant term in (31) is a useful approximation. The dominant term is independent of  $i$  and is therefore a very practical tool for estimating the error present

in PCR. As such, it is useful to express (31) as

$$\begin{aligned} \mathbf{e}_j^\top \text{CV} [U_i]^2 \mathbf{e}_k &\sim \text{CV} [X_0 + RY_0]^2 + \frac{1}{\mathbb{E} [X_0 + RY_0]} \left( \frac{\lambda_2}{\lambda_1} \right) \frac{R+1}{2} \\ &\quad + \frac{\mathbb{E} [X_0 - RY_0]}{\mathbb{E} [X_0 + RY_0]^2} \left( \frac{\bar{p}\lambda_1}{\lambda_1^2 - \lambda_2} \right) \frac{R-1}{2} \\ &= \alpha \text{CV} [I]^2 + \frac{\beta}{\mathbb{E} [I]}, \end{aligned} \quad (32)$$

for any  $j, k \in \{1, 2\}$ , where (27) is used in the first step and the second step follows from simplification with (8) and (26). The quantities  $\alpha$  and  $\beta$ , defined as

$$\alpha = \begin{cases} \frac{R^2 + 1}{(R + 1)^2}, & \text{Case D;} \\ 1, & \text{Case RF or RR;} \end{cases} \quad (33a)$$

$$\beta = \begin{cases} \left( \frac{\lambda_2}{\lambda_1} \right) \frac{1}{2} - \frac{\bar{p}\lambda_1}{\lambda_1^2 - \lambda_2} \left( \frac{R-1}{R+1} \right)^2 \frac{1}{2}, & \text{Case D;} \\ \frac{1-r}{r} + \left( \frac{\lambda_2}{\lambda_1} \right) \frac{R+1}{2Rr} - \left( \frac{\bar{p}\lambda_1}{\lambda_1^2 - \lambda_2} \right) \frac{R-1}{2Rr}, & \text{Case RF;} \\ \frac{1-r}{r} + \left( \frac{\lambda_2}{\lambda_1} \right) \frac{R+1}{2r} + \left( \frac{\bar{p}\lambda_1}{\lambda_1^2 - \lambda_2} \right) \frac{R-1}{2r}, & \text{Case RR;} \end{cases} \quad (33b)$$

relate the reaction efficiencies to the coefficient of variation, as  $\alpha = \alpha(R)$  and  $\beta = \beta(R, \bar{p}, r)$ . In Sect. 3.4, we will see that (32) also plays a central role in the limit of detection.

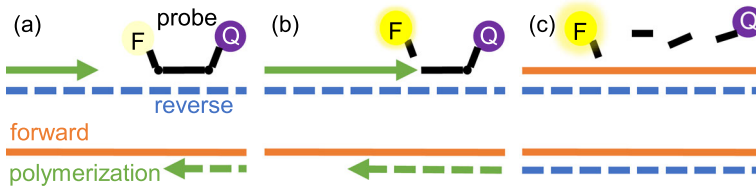
### 3 Fluorescence Dynamics

#### 3.1 Mathematical Model

To adapt the approach to the fluorescence measured in real-time PCR, it is necessary to relate the DNA content in solution to the monitoring chemistry. When a fluorescent probe is used to monitor the kinetics of PCR, the inactive and active probe species usually make significant contributions to fluorescence. With these two fluorescent species, the fluorescence analog of Beer's Law is

$$F_{i,w} = f_{i,w}^- C_{i,w}^- + f_{i,w}^+ C_{i,w}^+, \quad (34)$$





**Fig. 3** Illustration of relationship between successful DNA replication and changes in fluorescence associated with a hydrolysis probe binding to the reverse strand. (a) As polymerization begins, the hydrolysis probe binds to the reverse strand (blue, dashed line). The probe is in its inactive state, where fluorescence emitted by the fluorophore ('F') is quenched by the quencher ('Q') in close proximity. (b) As polymerization (green arrow) reaches the location of the probe, the probe is hydrolyzed. (c) After successful production of a forward strand (orange, solid line) from a reverse strand, the fluorophore is activated, as it is no longer in close proximity to the quencher. When a reverse strand is produced from a forward strand, however, the fluorescence does not change (Color figure online)

after<sup>2</sup> each cycle  $i = 1$  to  $n$  and for each well  $w = 1$  to  $m$ . Here,  $F_{i,w}$  is the fluorescence measured,  $f_{i,w}^-$  (or  $f_{i,w}^+$ ) is a constant representing the fluorescence per mole of inactive (or active) probe, and  $C_{i,w}^-$  (or  $C_{i,w}^+$ ) is the molar concentration of inactive (or active) probes. Each molar fluorescence,  $f_{i,w}^-$  or  $f_{i,w}^+$  (also denoted as  $f_{i,w}^\pm$ ), may depend on  $i$  due to photobleaching. Each may also depend on  $w$  due to spatial variation in electronics and temperature. The terms  $F_{i,w}$ ,  $C_{i,w}^-$ , and  $C_{i,w}^+$  are random variables through their dependence on DNA content (see below).

Assuming that probe is not degraded during cycling,

$$C = C_{i,w}^- + C_{i,w}^+ \tag{35}$$

for all  $i$  and  $w$ , where  $C$  is a known constant representing the total concentration of probe in solution. As DNA is replicated, inactive probe is converted to active probe in a manner that depends on the reaction stoichiometry. For hydrolysis probes, an inactive probe is activated, or hydrolyzed, each time *one* of the complementary strands is replicated. Without loss of generality, we consider the case where the hydrolysis probe binds to the reverse strand (see Fig. 3). After completing  $i$  PCR cycles, the concentration of active probe in each well  $w$  is then

$$C_{i,w}^+ = \frac{\Delta X_{i,w}}{\mathcal{V}N_a}, \tag{36}$$

where  $\mathcal{V}$  and  $N_a$  are constants representing the volume of solution and Avogadro's number, respectively, and  $\Delta X_{i,w} := X_{i,w} - X_{0,w}$ . We assume for each  $i = 0$  to  $n$  that  $X_{i,1}, \dots, X_{i,m}$  are independent and distributed identically to  $X_i$ , and similarly that  $Y_{i,1}, \dots, Y_{i,m}$  are independent and distributed identically to  $Y_i$ . By combining (34), (35), and (36), the fluorescence model becomes

$$F_{i,w} = b_{i,w} + d_{i,w} \Delta X_{i,w}, \tag{37}$$

<sup>2</sup> In this work, we assume that the fluorescence is measured after each PCR cycle. This is typical for assays utilizing hydrolysis probes, but is not always the case.

where

$$b_{i,w} := f_{i,w}^- C \tag{38}$$

represents the contribution of background fluorescence, and

$$d_{i,w} := \frac{f_{i,w}^+ - f_{i,w}^-}{\mathcal{V}N_a} \tag{39}$$

represents the increase in fluorescence per synthesis of forward strand.

From (37), the first two central-moments of  $F_{i,w}$  are

$$\mathbb{E} [F_{i,w}] = b_{i,w} + d_{i,w} \mathbb{E} [\Delta X_i], \tag{40a}$$

$$\text{Var} [F_{i,w}] = d_{i,w}^2 \text{Var} [\Delta X_i], \tag{40b}$$

where  $\Delta X_i := X_i - X_0$ . Here,  $\mathbb{E} [\Delta X_i]$  can be viewed as a function of  $\bar{p}$ ,  $R$ ,  $r$ , and  $\mathbb{E} [I]$  by substituting (8) into (13). Similarly,  $\text{Var} [X_i]$  can be viewed as a function of  $\bar{p}$ ,  $R$ ,  $r$ ,  $\mathbb{E} [I]$ , and  $\text{Var} [I]$  by substituting (8) and (26) into (24). While  $\text{Var} [X_0]$  is a function of  $r$ ,  $\mathbb{E} [I]$ , and  $\text{Var} [I]$  through (26), an expression for  $\text{Cov} [X_i, X_0]$  is needed for (40b). The methods described in Sect. 2.2 and 2.3 can be readily adapted to the cross-covariance matrix (see Sect. B.2) to obtain

$$\text{Cov} [X_i, X_0] = \frac{\text{Var} [X_0]}{2} (\lambda_1^i + \lambda_2^i). \tag{41}$$

From (13), (28), and (41), it follows that (40) can be expressed as

$$\mathbb{E} [F_{i,w}] = \left( \frac{d_{i,w}}{2} \right) \mathbb{E} [X_0 + RY_0] \lambda_1^i + O(1), \tag{42}$$

and

$$\begin{aligned} \text{Var} [F_{i,w}] = \frac{d_{i,w}^2}{4} \left\{ \text{Var} [X_0 + RY_0] + \mathbb{E} [X_0 + RY_0] \left( \frac{\lambda_2}{\lambda_1} \right) \frac{R+1}{2} \right. \\ \left. + \mathbb{E} [X_0 - RY_0] \left( \frac{\bar{p}\lambda_1}{\lambda_1^2 - \lambda_2} \right) \frac{R-1}{2} \right\} \lambda_1^{2i} + O(\lambda_1^i), \end{aligned} \tag{43}$$

as  $i \rightarrow \infty$ . As the dominant term of  $\text{Var} [F_{i,w}]$  is proportional to  $\lambda_1^{2i} \gg \lambda_1^i$ , it is independent of  $\text{Cov} [X_i, X_0] = O(\lambda_1^i)$ . Instead, the dominant term of  $\text{Var} [F_{i,w}]$  arises from  $\text{Var} [X_i]$  as in (28). As in (32), Equations (42) and (43) imply

$$\text{CV} [F_{i,w}]^2 = \alpha \text{CV} [I]^2 + \frac{\beta}{\mathbb{E} [I]} + O(\lambda_1^{-i}), \tag{44}$$

where  $\alpha = \alpha(R)$  and  $\beta = \beta(R, \bar{p}, r)$  were defined in (33) for Case D, RF, and RR.

In contrast to other models (e.g., Ruijter et al. 2013, 2014), the fluorescence model (37) is consistent with stoichiometric reactions involving hydrolysis probes and DNA polymerase. By using the fluorescence analog of Beer’s Law, it provides a physical basis for description of the background fluorescence  $b_{i,w}$ . Finally, unlike conventional approaches, the more mechanistic model demonstrates that  $d_{i,w}$  may depend on cycle  $i$ .

When  $R \neq 1$ ,  $R$  complicates the relationship between DNA content and fluorescence. However, for Case RR, Equations (8), (13), and (40a) lead to

$$\mathbb{E}[F_{i,w}] = b_{i,w} + \left(\frac{rd_{i,w}}{2}\right) \mathbb{E}[I] (\lambda_1^i + \lambda_2^i - 2), \quad \text{Case RR,}$$

which is independent of  $R$ . Since the choice of forward and reverse strands was arbitrary, this demonstrates that the monitoring probe can be chosen so that  $\mathbb{E}[F_{i,w}]$  is independent of  $R$ , and may be a useful design-rule for RT-qPCR assays.

### 3.2 Extraction of Molar Fluorescence

In real-time PCR, control experiments containing all reagents except nucleic acid template are often performed to check for contamination. In this section, we will show how they can also be used to calculate  $f_{i,w}^\pm$ .

Since template is not present (i.e.,  $U_0 = 0$ ), amplification does not occur and the probe cannot be activated. Equation (37) with (38) becomes

$$F_{i,w} = f_{i,w}^- C, \tag{45}$$

where  $F_{i,w}$  is instead deterministic. After filling  $m$  wells of a PCR plate with inactive probe at known  $C$  (and appropriate solvation environment) and measuring  $F_{i,w}$  after each  $i \geq 1$ ,  $f_{i,w}^-$  can be calculated pointwise from (45) by division. However, to get a more realistic estimate of each  $f_{i,w}^-$ , the experiment can be repeated with  $q$  different plates having a total probe concentration  $C^1 < \dots < C^q$  in all  $m$  wells. With

$$C := (C^1, \dots, C^q)^\top,$$

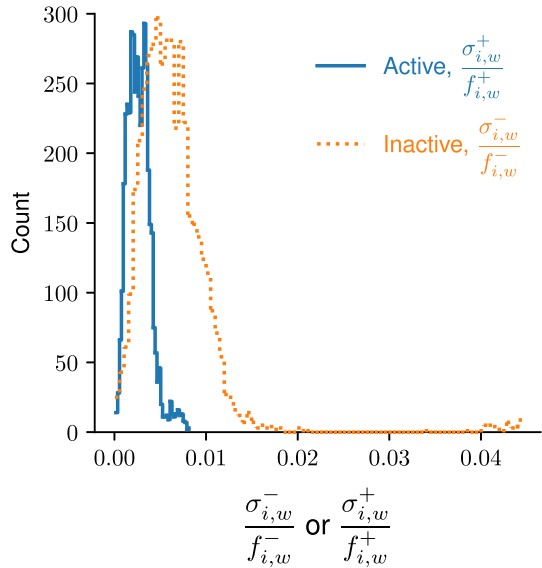
these experiments yield the measurements

$$F_{i,w} := (F_{i,w}^1, \dots, F_{i,w}^q)^\top$$

for each cycle  $i$  and well  $w$ . Under the assumption that  $\frac{F_{i,w}^j}{C^j}$  for  $j = 1$  to  $q$  are i.i.d. to a normal distribution, we can compute  $f_{i,w}^-$  via

$$f_{i,w}^\pm = \underset{f}{\operatorname{argmin}} \|F_{i,w} - fC\|^2 = \frac{F_{i,w}^\top C}{C^\top C}. \tag{46}$$

**Fig. 4** Total number of cycles and wells for all plates (count, vertical axis) possessing values of  $\sigma_{i,w}^{\pm}/f_{i,w}^{\pm}$  in a certain interval (bin, horizontal axis). The bin widths are obtained from the Freedman–Diaconis rule. All counts for  $\sigma_{i,w}^{-}/f_{i,w}^{-} > 0.022$  correspond to  $w = 2$  (well A2, see Fig. S2 and Table S2 in the SI)



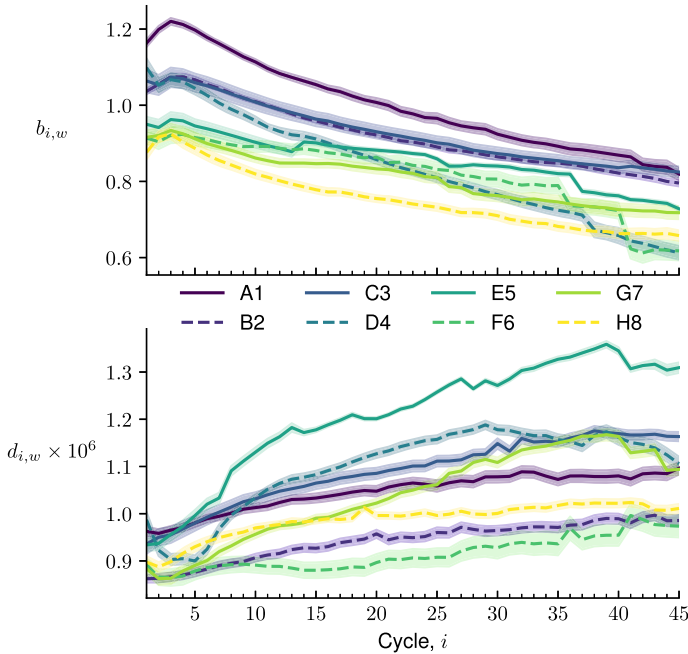
The same procedure can be used to calculate  $f_{i,w}^+$  after performing the identical experiments with active probe instead of inactive probe. The standard deviation in  $f_{i,w}^{\pm}$  can be estimated pointwise by

$$\sigma_{i,w}^{\pm} = \frac{\|F_{i,w} - f_{i,w}^{\pm} C\|}{\sqrt{q - 1}}. \tag{47}$$

Using hydrolysis probes, we selected the fluorophore (‘F’ in Fig. 3) to represent the active probe. We used  $q = 4$  concentrations for the active probe, and  $q = 3$  concentrations for the inactive probe. Additional details of the experimental procedure can be found in Appendix A. Visual comparisons between the model and experimental data can be found in Figs. S1 to S96 in the Supplementary Information (SI). The pointwise values of  $f_{i,w}^{\pm}$  and  $\sigma_{i,w}^{\pm}$  are tabulated in Tables S1 to S96 of the SI.

To assess the validity of approximating the measured fluorescence by  $f_{i,w}^{\pm} C$  for each cycle  $i$  and well  $w$ , the coefficient of variation, or  $\frac{\sigma_{i,w}^{\pm}}{f_{i,w}^{\pm}}$ , was calculated. A histogram of all values (i.e., all  $n$  cycles, all  $m$  wells, and all  $q$  plates) for each probe species is depicted in Fig. 4. For the active probe, the coefficient of variation is often very small, typically much less than 0.01. For the inactive probe, the coefficient of variation can be larger but is still often less than 0.02. This is an indication that the model is realistic.

Having validated the use of the Beer’s Law analog in describing the fluorescence, we used the molar-fluorescence parameters in Fig. 5 to assess how the background  $b_{i,w}$  and incremental increase  $d_{i,w}$  change with cycle  $i$  (for a fixed well  $w$ ). Here, we see that  $b_{i,w}$  is not a linear function of  $i$ . In fact, for several  $w$ ,  $b_{i,w}$  possesses a



**Fig. 5** Background (top) and incremental increase (bottom) as a function of cycle  $i$  for a selection of eight different wells (colors, see legend) with  $C = 0.125$  pmol/L and  $\mathcal{V} = 20$   $\mu$ L. The partially transparent shadings at each  $i$  depict an estimate of a 95 % confidence interval (Color figure online)

maximum in cycle around  $i = 2$  or  $i = 3$ . The decreases in  $b_{i,w}$  with increasing  $i$  are attributed to result from photobleaching.

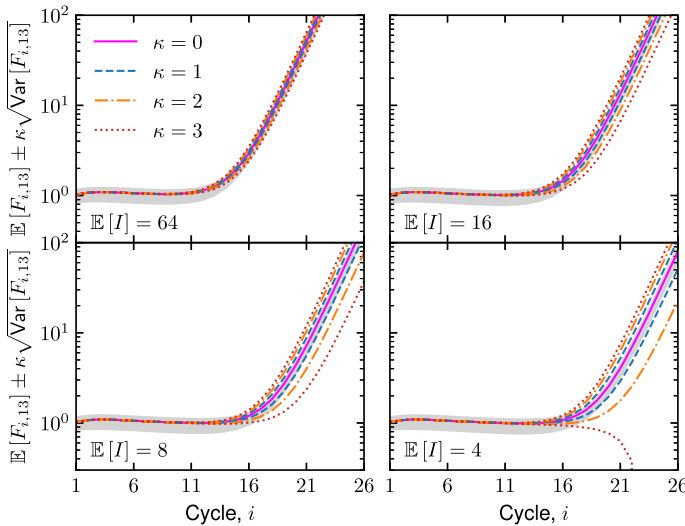
Since the fluorescence of an active probe is larger than the fluorescence of an inactive probe,  $d_{i,w}$  is positive. Figure 5 also illustrates that  $d_{i,w}$  is not independent of cycle; instead it often increases with  $i$ . This reveals another source of systematic error, as most models assume that  $d_{i,w}$  is independent of cycle (see, e.g., Ruijter et al. (2009), Equation (4); Lievens et al. (2012), Equation (7); Liu and Saint (2002), Equation (2)).

### 3.3 Calculation of Fluorescence Profiles

Having calculated  $f_{i,w}^-$  and  $f_{i,w}^+$  pointwise through (46), we leverage (40) with (8), (13), (24), (26), and (41) to compute fluorescence curves with uncertainty. We prescribe common values for assay parameters  $C$ ,  $\mathcal{V}$ ,  $\bar{p}$ , and  $R$  and assume  $\mathbb{E}[I]$  is known. However, we also need to specify the relationship between  $\text{Var}[I]$  and  $\mathbb{E}[I]$ , as well as the distribution of each  $F_{i,w}$ .

We will assume for simplicity that

$$\text{Var}[I] = \chi \mathbb{E}[I] \tag{48}$$



**Fig. 6** Fluorescence curves calculated from (40) with (8), (13), (24), (26), and (41) for  $w = 13$ . Each subplot is associated with a different input copy-number  $\mathbb{E}[I]$ , as depicted in each bottom-left corner, while all subplots have  $C = 0.125$  pmol/L,  $\mathcal{V} = 20$   $\mu$ L,  $\bar{p} = 0.9$ , and  $R = 1$ . The initial distribution is taken as Case D with (48) and  $\chi = 1$ . The light-grey shadings at each  $i$  depict the region between  $\min_{1 \leq w \leq 96} \mathbb{E}[F_{i,w}]$  and  $\max_{1 \leq w \leq 96} \mathbb{E}[F_{i,w}]$

for some constant  $\chi > 0$ . If we make the conventional assumption (Nedelman et al. 1992; Sundberg et al. 2010; Tellinghuisen and Spiess 2015; Ruiz-Villalba et al. 2021) that  $I$  is a Poisson random-variable,<sup>3</sup> then  $\chi = 1$ . However, (48) can also correspond to different probability distributions. If  $I$  obeys a negative binomial distribution with probability of success  $\varphi \in (0, 1)$ , for example, then  $\chi = 1/\varphi > 1$ .

Our assumption on the distribution of each  $F_{i,w}$  is rooted in the characteristic values of the fluorescence parameters  $b_{i,w}$  and  $d_{i,w}$ . As Fig. 5 demonstrates that  $b_{i,w}$  and  $d_{i,w}$  are typically around 1 and  $10^{-6}$ , respectively, this implies with (40a) that  $\mathbb{E}[\Delta X_i]$  should be more than  $10^6$  for  $\mathbb{E}[F_{i,w}] > b_{i,w}$ . That is, the expected number of successful Bernoulli trials over all  $i$  cycles should be more than  $10^6$  for the fluorescence to reach levels above background. With such a large sample size, it is natural to invoke the central limit theorem and assume that  $F_{i,w}$  obeys a normal distribution with mean  $\mathbb{E}[F_{i,w}]$  from (40a) and variance  $\text{Var}[F_{i,w}]$  from (40b).

In Fig. 6, fluorescence curves are computed with uncertainty for ds-DNA for  $w = 13$ . Different  $\mathbb{E}[I]$  ranging from 64 (top-left subplot) down to 4 (bottom-right subplot) are investigated. The expected fluorescence depicts behavior that is characteristic of the background and exponential phase observed in typical measurements. During the

<sup>3</sup> There is some limited experimental evidence for assuming that  $I$  is a Poisson random-variable (Beer et al. 2007); however, its validity is challenging to assess experimentally. (If the initial number of DNA strands could be easily measured, there would be no need for qPCR.) In the future, it would be interesting to compare (48) to a result obtained from error propagation for a conventional experimental procedure (see, e.g., Hedges (2002)).

initial cycles, the term  $b_{i,w}$  is much larger than  $d_{i,w} \mathbb{E} [\Delta X_i]$  and only small changes in fluorescence are observed. Here, the fluorescence is in the background regime. After more cycles are performed, however, the expected value of fluorescence increases exponentially with cycle. The change in  $\mathbb{E} [F_{i,13}]$  with  $\mathbb{E} [I]$  is also in line with typical trends. As  $\mathbb{E} [I]$  is decreased, the expected value of the fluorescence in the exponential phase shifts to the right. In other words, more cycles are required to reach the same expected fluorescence value.

This approach provides quantitative estimates of sources and magnitudes of uncertainty in different regimes, which are difficult to determine from replicate experiments alone. For large  $\mathbb{E} [I]$  and small  $i$ , the well-to-well variation in expected value (light-grey, shaded regions in Fig. 6) is larger or comparable to the error in fluorescence in each well. As such, spatial variation has a significant impact on the error. After many cycles have been completed, on the other hand, the uncertainty in fluorescence is strongly dependent on the expected initial copy-number, increasing drastically with decreasing  $\mathbb{E} [I]$ . At  $\mathbb{E} [I] = 4$ , the fluorescence does not reach values that are larger than the background fluorescence by an amount that is statistically significant (for  $\kappa = 3$ ). This observation suggests that quantifying the uncertainty in fluorescence can provide limitations on the measurement.

### 3.4 Limit of Detection

After performing  $n$  PCR cycles, the fluorescence produced by PCR is only useful if it is larger than background levels by a statistically significant amount. Requiring the increase to be at least some  $0 < \kappa < \infty$  standard deviations, this amounts to the constraint

$$\mathbb{E} [F_{n,w}] - \kappa \sqrt{\text{Var} [F_{n,w}]} \geq b_{n,w}, \tag{49}$$

for some well  $w$ . Equation (49) describes a feasible region of design space for a real-time PCR assay. It can be considered to depend on  $n, b_{n,w}, d_{n,w}, \bar{p}, R, r, \mathbb{E} [I]$ , and  $\text{Var} [I]$  through (8), (13), (24), (26), and (41). However, (49) can be further simplified by substituting (40) and rearranging, leading to

$$\text{CV} [\Delta X_n] \leq \frac{1}{\kappa},$$

an expression that is no longer dependent on  $w$ . Since  $\text{CV} [\Delta X_n] \geq 0$  and  $\kappa > 0$ , it follows that

$$\text{CV} [\Delta X_n]^2 \leq \frac{1}{\kappa^2}. \tag{50}$$

In addition, since

$$\text{CV} [\Delta X_n]^2 = \frac{\text{Var} [X_n] + \text{Var} [X_0] - 2\text{Cov} [X_n, X_0]}{\mathbb{E} [X_n]^2 - 2\mathbb{E} [X_n] \mathbb{E} [X_0] + \mathbb{E} [X_0]^2} = \text{CV} [X_n]^2 + O(\lambda_1^{-n})$$

by (13), (28), and (41), and  $n$  typically ranges from 35 to 50, the error in approximating  $\text{CV}[\Delta X_n]^2$  by  $\text{CV}[X_n]^2$  is extremely small, often less than machine precision. As a result, the left-hand-side of (50) is expressed as the term on the right-hand-side of (32), or

$$\alpha(R) \text{CV}[I]^2 + \frac{\beta(R, \bar{p}, r)}{\mathbb{E}[I]} \leq \frac{1}{\kappa^2}, \quad (51)$$

where  $\alpha$  and  $\beta$  are defined in (33) for Case D, where the input is ds-DNA; Case RF, where the input is fs-RNA; and Case RR, where the input is rs-RNA. If we let  $I$  satisfy (48), Equation (51) can be rearranged to

$$\mathbb{E}[I] \geq (\chi\alpha + \beta)\kappa^2. \quad (52)$$

The limit of detection,  $L$ , or the smallest expected-initial-copy-number that can be detected reliably, is then

$$L = \min \left\{ y \in \mathbb{N} \mid y \geq (\chi\alpha + \beta)\kappa^2 \right\}. \quad (53)$$

The *largest* coefficient of variation in  $I$  that can be detected,  $M$ , is estimated from (53) and (48), or

$$M := \sqrt{\frac{\chi}{L}}. \quad (54)$$

To compute typical values of  $L$  and  $M$ , we evaluated them as in (53) and (54) with  $\chi = 1$ ,  $\kappa = 3$ , for 100 equally-spaced  $\bar{p} \in [0.8, 0.99]$ ,  $R \in [0.9, 1.1]$ , and  $r \in [0.2, 0.99]$  (including endpoints). For  $I$  representing ds-DNA (Case D), we find that  $L$  is either 5 or 6, corresponding to  $M$  of 0.447 and 0.408, respectively. For  $I$  representing RNA as in Case RF or Case RR,  $L$  ranges between 9 and 52, corresponding to  $M$  of 1/3 and 0.139, respectively. The range of  $L$  is the same for fs-RNA and rs-RNA.

## 4 Conclusions and Future Work

In this work, we presented a new model for fluorescence in real-time PCR that reduced bias and quantified uncertainty. Distinguishing between complementary strands provided a stoichiometric description of fluorescence reported by hydrolysis probes and permitted application to initial conditions encountered in RT-qPCR. Viewing the fluorescence as a Beer's Law analog enabled the background fluorescence to be determined without extrapolation or assuming a certain relationship with cycle. It also allowed for measurement and calculation of background fluorescence without adjusting amplification data. Incorporating the variance in copy number into the fluorescence model enabled quantification of fluorescence uncertainty and analytical expressions for the limit of detection.



In addition to their practical utility, the two-type branching-process and repurposed fluorescence-model provided new intuition on the physics and chemistry in PCR. At short times, there is a lag in exponential growth (usually at most 5 cycles) as the ratio of expected strand counts changes from its initial to critical value,  $R$ . The quantity  $R$  represents the square root of the ratio of the two synthesis efficiencies (see (3)). In contrast to a previous report investigating deterministic models, we found that the initial composition only impacts the dynamics after the lag phase if  $R \neq 1$ .

The variance in the fluorescence is dominated by a term that increases exponentially by twice the factor of the expected value. This explains, in part, why quantification by end-point PCR is not reproducible. The three terms dominating the variance were attributed to arise from the initial variance ( $\nu_{1,1}$ , see (29a)), imperfect amplification ( $\eta_{1,1}^{(1)}$ , see (29b)), and deviation in directional efficiencies (i.e.,  $R \neq 1$ ; see  $\eta_{1,1}^{(2)}$  in (29c)). The fluorescence model for hydrolysis probes demonstrated that the background fluorescence originates from the molar fluorescence of the inactive probe times the total concentration of probe. The incremental increase in fluorescence is proportional to the difference in molar fluorescence between active and inactive probe and, like the background fluorescence, is neither independent of cycle nor a linear function of cycle.

The stochastic view of PCR explains, in part, why deterministic methods that use reaction-specific amplification probabilities are generally less accurate (Ruijter et al. 2013). This is because, for each well  $w$  and cycle  $i$ ,  $N_{i,w}/N_{i-1,w} \neq 1 + p$  (see Equation (5)). Even if  $R = 1$  and  $N_{i,1}, \dots, N_{i,m}$  are independent and distributed identically to  $N_i$  for each  $i$ , this is not necessarily true because not every realization of a random variable is equal to its expected value.

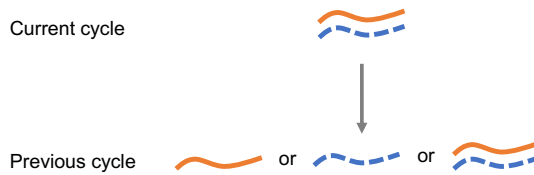
While this work applied the stochastic model of PCR to the fluorescence reported by hydrolysis probes, it can readily be extended to other chemistries. For example, for probes that anneal to forward-stranded DNA, (37) instead becomes

$$F_{i,w} = b_{i,w} + d_{i,w}X_{i-1,w}.$$

For these probes, the fluorescence is measured during the annealing portion of each cycle where only  $i - 1$  cycles of PCR have been completed.

To capture the fluorescence reported by DNA-binding dyes, on the other hand, an extension to the model is needed. This is because the amount of dye bound to a DNA strand depends on the total amount of DNA present in solution (this includes DNA that is not template, like primers (Ruijter et al. 2009)). Application to fluorescent dyes represents an interesting direction for future generalizations of the fluorescence model.

Finally, our approach in this work focused on quantifying the dynamics and uncertainty of fluorescence when the initial amount of each complementary strand is known, as well as their amplification probabilities. (That is, we assumed that  $\mathbb{E}[I]$ ,  $\text{Var}[I]$ ,  $\bar{p}$ ,  $R$ , and  $r$  were known.) However, the ultimate goal of monitoring the kinetics of PCR by fluorescent probes is to infer  $\mathbb{E}[I]$ , the expected input number. As such, it is of interest to extend the approach to UQ-PCR, or uncertainty quantification of the initial amount of DNA. To this end, another direction for future work is the investigation of the probabilistic nature of PCR in backwards time (see Fig. 7).



**Fig. 7** Conceptualization of probabilistic view of PCR in backwards time. Each ds-DNA complex present after  $i$  cycles have been completed may have originated from a forward strand or a reverse strand. Each complex may also be identical to a ds-DNA complex present at the end of the previous cycle

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11538-023-01182-z>.

**Acknowledgements** The authors acknowledge D. Duewer for helpful discussions. R. F. D. and M. J. R. acknowledge support from a National Research Council fellowship. R. F. D. also acknowledges support from the NIST Professional Research Experience Program through University of Maryland.

**Data Availability** The experimental data and software associated with this study are available at <https://github.com/usnistgov/bias-ug-pcr>. Additional parameters computed from experimental data are tabulated in the Supplementary Information.

## Declarations

**Disclaimer** Certain equipment, instruments, software, or materials, commercial or non-commercial, are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement of any product or service by the National Institute of Standards and Technology, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Appendix A Experimental

Reagents were obtained from suppliers for use in experiments. A DNA suspension buffer ( $\text{TE}^{-4}$ ) of pH 8.0 with 10 mmol/L Tris and 0.1 mmol/L EDTA was obtained from Teknova (Hollister, CA). 6-Carboxyfluorescein (6-FAM), single isomer, was purchased in solid form from Thermo Fisher Scientific (Waltham, MA). The working solution of 6-FAM was prepared by dissolving 10.1 mg 6-FAM in 1.75 mL of absolute ethanol (Sigma-Aldrich, St. Louis, MO). Subsequently, 10.0  $\mu\text{L}$  of the ethanolic solution was dissolved in 500 mL of  $\text{TE}^{-4}$ . A TaqMan minor groove binder probe with non-fluorescent quencher (Applied Biosystems, Waltham, MA), possessing the sequence 5'–ACCCCGCATTACGTTTGGTGGACC–3' and reporter (6-FAM) of the National Center for Immunization and Respiratory Diseases (U.S.). Division of Viral

Diseases (2020) 2019-nCoV\_N1 assay, was obtained from Thermo Fisher Scientific. Working solutions of probe were prepared by adding TE<sup>-4</sup> to a portion of the stock solution to yield a solution of concentration 5 μmol/L probe.

In a typical experiment, a solution was prepared with 25 vol. % TE<sup>-4</sup> at a specific concentration of either 6-FAM (active probe) or TaqMan probe (inactive probe). After mixing, 20 μL was transferred into each well of a 96-well plate. The plate was covered with adhesive film and centrifuged. After assessing that air bubbles were not visible, the plate was placed in an Applied Biosystems 7500 HID Real-Time PCR instrument. The thermal cycling protocol consisted of a 2 min holding stage at 55 °C (328 K), followed by 45 cycles. Each cycle consisted of 30 s at 55 °C (328 K), followed by 3 s at 95 °C (368 K). Data collection occurred during the 55 °C step in each cycle. The raw data through filter 1 was exported with HID Real-Time PCR Analysis Software, Version 1.2 (Applied Biosystems). The fluorescence values were divided by 10<sup>6</sup> before subsequent analysis.

## Appendix B Extended Derivations

### B.1 Derivation of (24) from (23)

Consider some  $\mathbf{B} \in \mathbb{R}^{2 \times 2}$ . From (12), we can write

$$\begin{aligned} \mathbf{A}^i \mathbf{B} \left(\mathbf{A}^\top\right)^i &= \left(\sum_{j=1}^2 \lambda_j^i \mathbf{x}_j \mathbf{z}_j^\top\right) \mathbf{B} \left(\sum_{k=1}^2 \lambda_k^i \mathbf{z}_k \mathbf{x}_k^\top\right) \\ &= \sum_{j=1}^2 \sum_{k=1}^2 (\lambda_j \lambda_k)^i \mathbf{x}_j \mathbf{z}_j^\top \mathbf{B} \mathbf{z}_k \mathbf{x}_k^\top, \\ &= \sum_{j=1}^2 \sum_{k=1}^2 (\lambda_j \lambda_k)^i \mathbf{T}_{j,k}(\mathbf{B}), \end{aligned} \tag{B1}$$

where  $\mathbf{T}_{j,k} : \mathbb{R}^{2 \times 2} \mapsto \mathbb{R}^{2 \times 2}$  is the linear operator

$$\mathbf{T}_{j,k}(\mathbf{B}) = \mathbf{x}_j \mathbf{z}_j^\top \mathbf{B} \mathbf{z}_k \mathbf{x}_k^\top = \left(\mathbf{z}_j^\top \mathbf{B} \mathbf{z}_k\right) \mathbf{x}_j \mathbf{x}_k^\top. \tag{B2}$$

Application of (B1) to (23) leads to

$$\begin{aligned} \text{Var}[U_i] &= \sum_{j=1}^2 \sum_{k=1}^2 (\lambda_j \lambda_k)^i \mathbf{T}_{j,k}(\text{Var}[U_0]) \\ &\quad + \sum_{j=1}^2 \sum_{k=1}^2 \sum_{\ell=1}^2 \sum_{q=0}^{i-1} \lambda_\ell^q (\lambda_j \lambda_k)^{i-1-q} \mathbf{T}_{j,k}(\mathbf{K}_\ell). \end{aligned} \tag{B3}$$

For each  $b \in \{\lambda_1^2, \lambda_1\lambda_2, \lambda_2^2\}$ , the geometric series in (B3) can be simplified via

$$\sum_{q=0}^{i-1} \lambda_\ell^q b^{i-1-q} = b^{i-1} \sum_{q=0}^{i-1} \left(\frac{\lambda_\ell}{b}\right)^q = b^{i-1} \left(\frac{1 - \left(\frac{\lambda_\ell}{b}\right)^i}{1 - \frac{\lambda_\ell}{b}}\right) = \frac{b^i - \lambda_\ell^i}{b - \lambda_\ell}.$$

This admits the expression

$$\begin{aligned} \text{Var}[U_i] &= \sum_{j=1}^2 \sum_{k=1}^2 (\lambda_j \lambda_k)^i \mathbf{T}_{j,k} (\text{Var}[U_0]) + \sum_{j=1}^2 \sum_{k=1}^2 \sum_{\ell=1}^2 \left(\frac{(\lambda_j \lambda_k)^i - \lambda_\ell^i}{\lambda_j \lambda_k - \lambda_\ell}\right) \mathbf{T}_{j,k} (\mathbf{K}_\ell), \\ &= \sum_{j=1}^2 \sum_{k=1}^2 (\lambda_j \lambda_k)^i \mathbf{T}_{j,k} \left(\text{Var}[U_0] + \sum_{\ell=1}^2 \frac{\mathbf{K}_\ell}{\lambda_j \lambda_k - \lambda_\ell}\right) \\ &\quad - \sum_{j=1}^2 \sum_{k=1}^2 \sum_{\ell=1}^2 \lambda_\ell^i \mathbf{T}_{j,k} \left(\frac{\mathbf{K}_\ell}{\lambda_j \lambda_k - \lambda_\ell}\right). \end{aligned} \tag{B4}$$

Since, from (25) and (B2),

$$\begin{cases} \mathbf{T}_{j,k} (\text{Var}[U_0]) = \nu_{j,k} \mathbf{x}_j \mathbf{x}_k^\top, \\ \mathbf{T}_{j,k} \left(\frac{\mathbf{K}_\ell}{\lambda_j \lambda_k - \lambda_\ell}\right) = \eta_{j,k}^\ell \mathbf{x}_j \mathbf{x}_k^\top, \end{cases}$$

it is evident that (B4) is identical to (24).

### B.2 Derivation of (41)

In this section, the derivation of (41) is completed by investigating the cross-covariance matrix, defined as

$$\mathcal{K}[X, Y] = \mathbb{E}[XY^\top] - \mathbb{E}[X] \mathbb{E}[Y]^\top$$

for any two random vectors  $X$  and  $Y$ . Notice that  $\mathcal{K}[X, X] = \text{Var}[X]$ .

From the law of total expectation,

$$\mathbb{E}[U_i] \mathbb{E}[U_0]^\top = \mathbb{E}[\mathbb{E}[U_i | U_{i-1}]] \mathbb{E}[U_0]^\top = \mathbf{A} \mathbb{E}[U_{i-1}] \mathbb{E}[U_0]^\top,$$

and

$$\begin{aligned} \mathbb{E}[U_i U_0^\top] &= \mathbb{E}[\mathbb{E}[U_i U_0^\top | U_{i-1}, U_0]] = \mathbb{E}[\mathbb{E}[U_i | U_{i-1}] U_0^\top] \\ &= \mathbf{A} \mathbb{E}[U_{i-1} U_0^\top]. \end{aligned}$$

The latter expression follows because  $U_i$  and  $U_0$  are independent when conditioned on  $U_{i-1}$ . Thus,

$$\begin{aligned}\mathcal{K}[U_i, U_0] &= A\mathcal{K}[U_{i-1}, U_0] = A^i \text{Var}[U_0] \\ &= \sum_{j=1}^2 \lambda_j^i \mathbf{x}_j \mathbf{z}_j^\top \text{Var}[U_0] \\ &= \sum_{j=1}^2 \lambda_j^i \mathbf{x}_j \mathbf{z}_j^\top \begin{pmatrix} \text{Var}[X_0] & 0 \\ 0 & \text{Var}[Y_0] \end{pmatrix},\end{aligned}$$

where the second, third, and fourth step follow from induction, Equation (12), and the independence of  $X_0$  and  $Y_0$ , respectively. In particular, it follows that

$$\begin{aligned}\text{Cov}[X_i, X_0] &= \mathbf{e}_1^\top \mathcal{K}[U_i, U_0] \mathbf{e}_1 \\ &= \mathbf{e}_1^\top \left( \sum_{j=1}^2 \lambda_j^i \mathbf{x}_j \mathbf{z}_j^\top \right) \mathbf{e}_1 \text{Var}[X_0] \\ &= \frac{\text{Var}[X_0]}{2} (\lambda_1^i + \lambda_2^i),\end{aligned}$$

which is the same as (41).

## References

- Bauer M (2007) RNA in forensic science. *Forensic Sci Int Genet* 1:69–74. <https://doi.org/10.1016/j.fsigen.2006.11.002>
- Beer NR, Hindson BJ, Wheeler EK et al (2007) On-chip, real-time, single-copy polymerase chain reaction in picoliter droplets. *Anal Chem* 79:8471–8475. <https://doi.org/10.1021/ac701809v>
- Biver T, Secco F, Tiné MR et al (2003) Equilibria and kinetics of the intercalation of Pt-proflavine and proflavine into calf thymus DNA. *Arch Biochem Biophys* 418:63–70. [https://doi.org/10.1016/S0003-9861\(03\)00384-9](https://doi.org/10.1016/S0003-9861(03)00384-9)
- Biver T, De Biasi A, Secco F et al (2005) Cyanine dyes as intercalating agents: Kinetic and thermodynamic studies on DNA/Cyan40 and DNA/CCyan2 systems. *Biophys J* 89:374–383. <https://doi.org/10.1529/biophysj.105.059790>
- Boggy GJ, Woolf PJ (2010) A mechanistic model of PCR for accurate quantification of quantitative PCR data. *PLoS ONE* 5:e12355. <https://doi.org/10.1371/journal.pone.0012355>
- Bustin S, Dhillon HS, Kirvell S et al (2015) Variability of the reverse transcription step: Practical implications. *Clin Chem* 61:202–212. <https://doi.org/10.1373/clinchem.2014.230615>
- Bustin S, Mueller R, Shipley G et al (2021) COVID-19 and diagnostic testing for SARS-CoV-2 by RT-qPCR—facts and fallacies. *Int J Mol Sci* 22:2459. <https://doi.org/10.3390/ijms22052459>
- Bustin SA (2004) Getting started—the basics of setting up a QPCR assay. *IUL Biotechnol, La Jolla, CA*, chap 13:530–543
- Elizaquível P, Aznar R, Sanchez G (2014) Recent developments in the use of viability dyes and quantitative PCR in the food microbiology field. *J Appl Microbiol* 116:1–13. <https://doi.org/10.1111/jam.12365>
- Guescini M, Sisti D, Rocchi MB et al (2008) A new real-time PCR method to overcome significant quantitative inaccuracy due to slight amplification inhibition. *BMC Bioinform* 9:326. <https://doi.org/10.1186/1471-2105-9-326>
- Hedges AJ (2002) Estimating the precision of serial dilutions and viable bacterial counts. *Int J Food Microbiol* 76:207–214. [https://doi.org/10.1016/S0168-1605\(02\)00022-3](https://doi.org/10.1016/S0168-1605(02)00022-3)

- Jacob C, Peccoud J (1996) Estimation of the offspring mean for a supercritical branching process from partial and migrating observations. *C R Acad Sci Paris Ser I* 322:763–768
- Jacob C, Peccoud J (1996) Inference on the initial size of a supercritical branching process from migrating binomial observations. *C R Acad Sci Paris Ser I* 322:875–880
- Kaltenboeck B, Wang CM (2005) Advances in real-time PCR: application to clinical laboratory diagnostics. *Adv Clin Chem* 40:219–259. [https://doi.org/10.1016/S0065-2423\(05\)40006-2](https://doi.org/10.1016/S0065-2423(05)40006-2)
- Kebschull JM, Zador AM (2015) Sources of PCR-induced distortions in high-throughput sequencing data sets. *Nucleic Acids Res* 43:e143. <https://doi.org/10.1093/nar/gkv717>
- Lievens A, Van Aelst S, Van den Bulcke M et al (2012) Enhanced analysis of real-time PCR data by using a variable efficiency model: FPK-PCR. *Nucleic Acids Res* 40:e10. <https://doi.org/10.1093/nar/gkr775>
- Liu W, Saint DA (2002) A new quantitative method of real time reverse transcription polymerase chain reaction assay based on simulation of polymerase chain reaction kinetics. *Anal Biochem* 302:52–59. <https://doi.org/10.1006/abio.2001.5530>
- National Center for Immunization and Respiratory Diseases (U.S.). Division of Viral Diseases (2020) 2019–Novel Coronavirus (2019-nCoV) Real-time rRT-PCR Panel Primers and Probes. <https://stacks.cdc.gov/view/cdc/84525>
- National Institutes of Health, National Human Genome Research Institute (2023) Polymerase chain reaction (PCR). <https://www.genome.gov/genetics-glossary/Polymerase-Chain-Reaction>
- Nedelman J, Heagerty P, Lawrence C (1992) Quantitative PCR: Procedures and precisions. *Bull Math Biol* 54:477–502
- Nicklas JA, Buel E (2003) Quantification of DNA in forensic samples. *Anal Bioanal Chem* 376:1160–1167. <https://doi.org/10.1007/s00216-003-1924-z>
- Peccoud J, Jacob C (1996) Theoretical uncertainty of measurements using quantitative polymerase chain reaction. *Biophys J* 71:101–108. [https://doi.org/10.1016/S0006-3495\(96\)79205-6](https://doi.org/10.1016/S0006-3495(96)79205-6)
- Peirson SN, Butler JN, Foster RG (2003) Experimental validation of novel and conventional approaches to quantitative real-time PCR data analysis. *Nucleic Acids Res* 31:e73. <https://doi.org/10.1093/nar/gng073>
- Ruijter JM, Ramakers C, Hoogaars WMH et al (2009) Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. *Nucleic Acids Res* 37:e45. <https://doi.org/10.1093/nar/gkp045>
- Ruijter JM, Pfaffl MW, Zhao S et al (2013) Evaluation of qPCR curve analysis methods for reliable biomarker discovery: Bias, resolution, precision, and implications. *Methods* 59:32–46. <https://doi.org/10.1016/j.jymeth.2012.08.011>
- Ruijter JM, Lorenz P, Tuomi JM et al (2014) Fluorescent-increase kinetics of different fluorescent reporters used for qPCR depend on monitoring chemistry, targeted sequence, type of DNA input and PCR efficiency. *Microchim Acta* 181:1689–1696. <https://doi.org/10.1007/s00604-013-1155-8>
- Ruiz-Villalba A, Ruijter JM, van den Hoff MJB (2021) Use and misuse of  $C_q$  in qPCR data analysis and reporting. *Life* 11:496. <https://doi.org/10.3390/life11060496>
- Schwabe D, Falcke M (2022) On the relation between input and output distributions of scRNA-seq experiments. *Bioinformatics* 38:1336–1343. <https://doi.org/10.1093/bioinformatics/btab841>
- Schwaber J, Andersen S, Nielsen L (2019) Shedding light: The importance of reverse transcription efficiency standards in data interpretation. *Biomol Detect Quantif* 17:100077. <https://doi.org/10.1016/j.bdq.2018.12.002>
- Stolovitzky G, Cecchi G (1996) Efficiency of DNA replication in the polymerase chain reaction. *Proc Natl Acad Sci USA* 93:12947–12952. <https://doi.org/10.1073/pnas.93.23.12947>
- Sun F (1995) The polymerase chain reaction and branching processes. *J Comput Biol* 2:63–86. <https://doi.org/10.1089/cmb.1995.2.63>
- Sundberg SO, Wittwer CT, Gao C et al (2010) Spinning disk platform for microfluidic digital polymerase chain reaction. *Anal Chem* 82:1546–1550. <https://doi.org/10.1021/ac902398c>
- Tellinghuisen J, Spiess AN (2015) Absolute copy number from the statistics of the quantification cycle in replicate quantitative polymerase chain reaction experiments. *Anal Chem* 87:1889–1895. <https://doi.org/10.1021/acs.analchem.5b00077>
- Tichopad A, Dilger M, Schwarz G et al (2003) Standardized determination of real-time PCR efficiency from a single reaction set-up. *Nucleic Acids Res* 31:e122. <https://doi.org/10.1093/nar/gng122>
- Weiss G, von Haeseler A (1995) Modeling the polymerase chain reaction. *J Comput Biol* 2:49–61. <https://doi.org/10.1089/cmb.1995.2.49>
- Zhao S, Fernald RD (2005) Comprehensive algorithm for quantitative real-time polymerase chain reaction. *J Comput Biol* 12:1047–1064. <https://doi.org/10.1089/cmb.2005.12.1047>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.