The optimisation of quantitation assays

Paweł Rafał Dębski

Advisor: Prof. dr hab. Piotr Garstecki

The dissertation was prepared within the International PhD Studies at the

Institute of Physical Chemistry of the Polish Academy of Sciences

Kasprzaka 44/52, 01-224 Warsaw

Warsaw, November 2016

Acknowledgements

I would like to thank my supervisor Prof. dr hab. Piotr Garstecki for the guidance in the world of microfluidis and molecular diagnostics. I could always count on his advice and support during my exciting scientific journey at the intersections of chemistry, physics, biology and maths.

The projects presented in this work could be run successfully only in a supporting scientific environment. Therefore, I would like to thank my current and former colleagues from the Microfluidics and Complex Fluids group for their help and friendly atmosphere in the lab. In particular, I would like to thank the co-authors of the projects I was involved in during my doctoral studies: Tomasz Kamiński, Ott Scheler and Artur Ruszczak. Also, I would like to acknowledge Joanna Smyda, Dominika Ogończyk and Karol Makuch. I would like to express my gratitude to Sławomir Jakieła, who was not only my closest collaborator and a brilliant researcher, but also a great friend.

During my PhD studies, I had the opportunity to work with brilliant researchers from outside the Microfluidics group. I would like to acknowledge Magdalena Sulima, Krzysztof Mellem, Kamil Gewartowski, Seweryn Bajer and Marcin Izydorzak from Scope Fluidics Sp. z o.o. and Curiosity Diagnostics Sp. z o.o.

I am grateful to my Parents for their patience in recent years, and for supporting my passion for Science.

I would like to thank my Friends who were always there for me.

Finally, I would like to thank Sylwia for the inspiration and encouragement. You are a special person in my life and your support means a lot to me.

The projects described in this work were supported by the following grants:

- Starting Grant 279647 *Microfluidic Combinatorial on Demand Systems: a Platform for High-Throughput Screening in Chemistry and Biotechnology* financed by the European Research Council

- E'8042 OPTIGENS grant provided by the National Centre for Research and Development within the Eureka Initiative

- Science scholarship *Mazowieckie Doktoraty* for PhD students who successfully cooperate with innovative companies from Mazovia (granted for the cooperation with Curiosity Diagnostics Sp. z o.o.)

Publications

Publications in the field of physical chemistry

1. Debski, P.R., Gewartowski, K., Bajer, S., and Garstecki, P., *Calibration-free assays on standard real-time PCR devices.*, submitted

2. Debski, P.R., and Garstecki, P., *Designing and interpretation of digital assays: concentration of target in the sample and in the source of sample*, Biomolecular Detection and Quantification, Available online 17 May 2016

3. Debski, P.R., Gewartowski, K., Sulima, M., Kaminski, T.S., and Garstecki, P., *Rational design of Digital assays*, Analytical Chemistry, 2015, 87 (16), 8203-8209

4. Jakiela, S.; Debski, P.R.; Dabrowski, B.; Garstecki, P. Generation of Nanoliter Droplets on Demand at Hundred-Hz Frequencies. Micromachines 2014,5, 1002-1011

Earlier publications

1. Aamodt, K.; Abelev, B.; Abrahantes Quintana, A.; et al., *Charged-Particle Multiplicity Density at Midrapidity in Central Pb-Pb Collisions at root s(NN)=2.76 TeV*, Phys. Rev. Lett. 2010, 105, 252301

2. Aamodt, K.; Abelev, B.; Abrahantes Quintana, A.; et al., *Elliptic Flow of Charged Particles in Pb-Pb Collisions at root s(NN)=2.76 TeV*, Phys. Rev. Lett. 2010, 105, 252302

3. Bobinski, T., Debski, P.R., Lisicki, M., Wojtowicz, K., and Zielenkiewicz, M., *Einstein - de Haas effect*, Proceedings of the XVIII International Young Physicists' Tournament, Winterthur 2005

Patent Applications

1. *Method for smoothing of the surface of polycarbonate*, Polish Patent Application P-404975 (filed: 2 August 2013)

2. *Method for performing quantitation assays*, PCT/EP2013/000805 (patent granted in 2016; filed: 15 March 2013)

3. *Method for the on-demand separation of paramagnetic material from droplet and the apparatus for the on-demand separation of paramagnetic material from droplet*, German Patent

Application P-52366 (based on Polish Patent Application P-397837; filed: 18 January 2013)

4. *Method for performing quantitation assays*, PCT/EP2012/004792 (patent granted in 2016; filed: 19 November 2012)

5. *Method for performing quantitation assays*, Polish Patent Application P-399908 (filed: 11 July 2012)

6. *Method for performing quantitation assays*, Polish Patent Application P-399673 (filed: 26 June 2012)

7. *Method for the on-demand separation of paramagnetic material from droplet and the apparatus for the on-demand separation of paramagnetic material from droplet*, Polish Patent Application P-397837 (patent granted in 2014; filed: 18 January 2012)

8. *Method for performing quantitation assays*, Polish Patent Application P-397026 (filed: 17 November 2011)

9. *Method for performing quantitation assays*, Polish Patent Application P-397027 (filed: 17 November 2011)

10. *Method for performing quantitation assays*, Polish Patent Application P-397028 (filed: 17 November 2011)

Conference contributions

1. **Scheler, O.,** Pacocha, N., Debski, P.R., Ruszczak, A., Kaminski, T.S., Garstecki, P., *Digital counting of bacteria over a broad dynamic range of concentrations* (poster), microTAS 2016, Dublin, Ireland, 9-13 October 2016

2. **Scheler, O.,** Pacocha, N., Debski, P.R., Ruszczak, A., Kaminski, T.S., Garstecki, P., *Digital counting of bacteria over a broad dynamic range of concentrations* (poster), EMBL Microfluidics 2016, Heidelberg, Germany, 24-26 July 2016

3. **Scheler, O.,** Pacocha, N., Debski, P.R., Kaminski, T.S., Ruszczak, A., Garstecki, P., *Growing and detecting bacteria in nanoliter "test-tubes"* (poster), GRC conference: 2016 Drug Resistance, Biddeford/Maine, USA, 12-16 June 2016

4. **Debski, P.R.,** Garstecki, P., *Optimisation of digital analytical assays in microfluidic devices* (talk), Chemsession 2016 - 13th PhD Students of Chemistry Seminar, Warsaw,

Poland, 10 June 2016

5. **Scheler, O.,** Kaminski, T.S., Dutka, F., Horka, M., Ruszczak, A., Debski, P.R., Barys, J., Babiak, M., Jakiela, S., Churski, K., Garstecki, P., *Droplet microfluidics for microbiology* (talk), Seminar in Faculty of Biotechnology of the University of Wroclaw, Wroclaw, Poland, 29-30 September 2015

6. **Scheler, O.,** Kaminski, T.S., Dutka, F., Horka, M., Ruszczak, A., Debski, P.R., Barys, J., Garstecki, P., *Culturing, counting and analysis of bacteria using droplet microfluidics* (poster), 6th International Weigl Conference on Microbiology, Gdansk, Poland, 8-10 July 2015

7. **Debski, P.R.,** Garstecki, P., *Optimisation of digital analytical assays in microfluidic devices* (poster), Chemsession 2015 - 12th PhD Students of Chemistry Seminar, Warsaw, Poland, 8 May 2015

8. **Debski, P.R.,** Garstecki, P., *Synergistic assays for microfluidic diagnostics* (poster), Patchy colloids, active matter and nanofluids conference, Geilo, Norway, 16-26 March 2015

9. **Kaminski, T.S,** Debski, P.R., Postek, W., Garstecki, P., *Droplet libraries for digital PCR with improved dynamic range* (poster), MNM 2014 Micro and Nanotechnologies in Medicine ,Kahuku, Hawaii, USA, 8-12 December 2014

10. **Debski, P.R.,** Garstecki, P., *Optimisation of digital analytical assays in microfluidic devices* (poster), Platform for Commercialisation of Knowledge, Warsaw, Poland, 20 May 2014

11. **Postek, W.,** Kaminski, T.S., Debski, P.R., Garstecki, P., *Microfluidic Technique for Point-of-Care Digital PCR* (poster), SLAS2014 3rd Annual Conference and Exhibition, San Diego/California, USA, 18-22 January 2014

12. **Debski, P.R.,** Garstecki, P., *Optimisation of digital analytical assays in microfluidic devices* (poster), International Soft Matter Conference, Rome, Italy, 15-19 September 2013

13. **Debski, P.R.,** Jakiela, S., Garstecki, P., *Separation of paramagnetic beads from droplets in microfluidic devices* (poster), Soft Matter Confinement: From Biology to Physics conference, Geilo, Norway, 11-21 March 2013

Abstract

Quantitation assays are a vital tool in various applications of molecular diagnostics. The demand for new DNA-, RNA- and immuno-diagnostic tests is constantly growing. They have found a number of applications in the diagnostics of genetic, parasite, and infectious diseases, as well as detection and monitoring of cancer, or the paternity testing. Such applications require tests that provide precise and specific quantitative assessment of the concentration of the analyte over a wide range of concentrations.

Quantitative Polymerase Chain Reaction (qPCR) is the central technique of molecular biology and an important solution in medical diagnostics. In was introduced in 1991 by Holland et al. in a format of Real-Time PCR [1]. This technique monitors the intensity of the physical signal (level of fluorescence, turbidity, etc.) connected with the formation of DNA product during PCR. The amount of this product (concentration C_p of the product) increases geometrically with the number n of PCR cycles: $C_p \propto q^n$. It is possible to assess the initial concentration of the analyte by specifying the number of PCR cycle after which the signal reached a given threshold, and comparing it with calibrated references. qPCR provides the assessment within a wide range of initial concentrations. However, the precision and accuracy of the assessment may be compromised by a number of factors, including the quality of samples and reagents, the presence of inhibitors, the quality of thermal cycling, and finally the reliability of signal detection in a qPCR device.

Despite the listed problems, real-time PCR is recognized as a *golden standard* qPCR technique, thanks to the relatively simple sample handling protocols (no partitioning of the sample is required), and well-established mathematical routines for calculating the final result (the comparison with a calibration curve).

However, there is an alternative PCR-based quantitative technique in a format of digital PCR (dPCR). Digital protocols were first introduced in 1915 by McCrady [2]. He described a the limiting-dilution quantitation assay for bacteria counting. Then, in 1992, digital protocols were applied to quantitative PCR by Sykes et al. [3]. This idea was further developed by Vogelstein and Kinzler in 1999 [4]. They introduced the division of the sample into identical partitions which yielded either a positive ($s = 1$) or negative ($s = 0$) signal, depending on the presence of at least one molecule of the analyte in the inspected partition. Digital assays provide an absolute and highly reliable assessment of the initial concentration of the analyte, without any calibration of the experimental set-up. Moreover, the assessment by means of digital assays is usually very

precise and sensitive [5–7].

Still, the popularity of digital techniques is increasing slowly. The main obstacle is the need for new apparata able to perform complicated partitiong and thermal treatment with the readout of tens or millions of partitions. This elevates the complexity and cost of devices.

In this work, we present the optimization of digital assay, and provide the systematic mathematical description. The algorithms we propose help to lower the laboratory requirements for running digital assessment by limiting the number of compartments.

We also describe an analog-digital method that combines the advantages of qPCR and dPCR and can be performed by means of standard qPCR devices. All the protocols presented in this work lower the number of compartments needed for the assessment and provide absolute quantitation. The protocols can be adapted to required parameters of a test: range of detected concentrations and precision of the assessment.

In the first chapter, we give a description of the project of optimisation of quantitation assays. The second chaper contains a short introduction to molecular diagnostics, with a brief description of the PCR-based methods. In the third chapter, we state the aim of our research, while the fourth chapter contains the description of the mathematical, numerical and experimental methods used to describe analytical assays, derive formulas for optimized assays and finally verify their performance. In chapters fifth, sixth, seven and eighth we describe the results of the analysis and the design of classic digital single- and multivolume assays, rationally designed digital assays and synergistic digital-analogue assays. We discuss and summarize the results and performance in the ninth chapter. Also, the instructions of how to prepare optimized assays and how to analyze the outcome from such assays are given in Appendix A, B, and C.

The findings of this work were published in the following research papers:

1. Debski, P.R., Gewartowski, K., Bajer, S., and Garstecki, P., *Calibration-free assays on standard real-time PCR devices.*, submitted

2. Debski, P.R., and Garstecki, P., *Designing and interpretation of digital assays: concentration of target in the sample and in the source of sample*, Biomolecular Detection, and Quantification, Available online 17 May 2016

3. Debski, P.R., Gewartowski, K., Sulima, M., Kaminski, T.S., and Garstecki, P., *Rational design of Digital assays*, Analytical Chemistry, 2015, 87 (16), 8203-8209

Streszczenie

Oznaczenia ilościowe są ważnym narzędziem diagnostyki molekularnej, znajdującymi szerokie zastosowanie w badaniach naukowych i diagnostyce. Zapotrzebowanie na nowe testy do diagnostyki opartej na RNA, DNA i immunodiagnostyki stale wzrasta. Takie testy są powszechnie używane do diagnostyki chorób genetycznych i zakaźnych, wykrywania patogenów, detekcji i monitorowania nowotworów, oraz badania pokrewieństwa. Takie zadania wymagają technik zapewniających precyzyjne i specyficzne ilościowe oznaczanie analitu w szerokim zakresie stężeń.

Ilościowa Łańcuchowa Reakcja Polimerazy (qPCR) jest podstawową metodą biologii molekularnej i znajduje powszechne zastosowanie w diagnostyce medycznej. Została opisana w formacie PCR w czasie rzeczywistym (real-time PCR) w 1991 roku przez Pamelę Holland i jej współpracowaników z firmy Cetus [1]. Wykorzystuje ona stałe monitorowanie natężenia fizycznego sygnału (poziomu fluorescencji, zmiany mętności, i innych) związanego z narastaniem produktu (łańcuchów DNA) reakcji PCR. Ilość produktu (stężenie produktu C_p) wzrasta geometrycznie wraz z numerem n cyklu PCR: $C_p \propto q^n$. Wyznaczenie początkowego stężenia analitu jest możliwe dzięki określeniu numeru cyklu, po którym sygnał osiągnął określony poziom, a następnie porównanie go z wynikami dla wystandaryzowanych próbek. Chociaż qPCR pozwala na określenie początkowego stężenia analitu w szerokim zakresie stężeń, precyzja i dokładność pomiaru mogą być zaburzone przez wiele czynników: jakość próbki i substratów reakcji, obecność inhibitorów reakcji, jakość obróbki termicznej i niezawodnosć systemów detekcji sygnału w urządzeniach do analiz qPCR.

Pomimo powyższych problemów, dzięki uproszczonej obróbce próbki (próbka nie musi być dzielona na kompartmenty), oraz powszechnie znanym metodom matematycznym pozwalającym na wyznaczenie początkowego stężenia analitu (porównanie do krzywej kalibracyjnej), PCR w czasie rzeczywistym jest uznawany za wzorcowy przykład ilościowej reakcji PCR.

Istnieje alternatywna forma ilościowej reakcji PCR, zwana również cyfrową reakcją PCR (dPCR). Protokół cyfrowy w oznaczeniach ilościowych został po raz pierwszy opisany w 1915 roku przez M.H. McCrady'ego [2]. Opisał on rozcieńczeniowe oznaczenie ilościowe służące do określania liczby bakterii. Następnie, w 1992 roku, protokoły cyfrowe zostały zastosowane przez Pamelę Sykes i jej współpracowników w ilościowej reakcji PCR [3]. Pomysł ten został rozwinięty w 1999 roku przez Berta Vogelsteina i Kennetha Kinzlera [4], którzy zaproponowali podział próbki na identyczne kompartmenty, które dawały pozytywny ($s = 1$) lub negatywny ($s = 0$) sygnał w zależności od tego, czy w danym kompartmencie znajdowała się przynajmniej jedna cząsteczka anal-

and the following patents and patent applications:

1. *Method for performing quantitation assays*, PCT/EP2013/000805 (patent granted in 2016; filed: 15 March 2013)

2. *Method for performing quantitation assays*, PCT/EP2012/004792 (patent granted in 2016; filed: 19 November 2012)

3. *Method for performing quantitation assays*, Polish Patent Application P-399908 (filed: 11 July 2012)

4. *Method for performing quantitation assays*, Polish Patent Application P-399673 (filed: 26 June 2012)

5. *Method for performing quantitation assays*, Polish Patent Application P-397026 (filed: 17 November 2011)

6. *Method for performing quantitation assays*, Polish Patent Application P-397027 (filed: 17 November 2011)

7. *Method for performing quantitation assays*, Polish Patent Application P-397028 (filed: 17 November 2011).

itu. Oznaczenia cyfrowe zapewniają absolutne i niezawodne wyznaczanie początkowego stężenia [5–7], nie potrzebując przy tym kalibracji układu doświadczalnego. Oznaczenia te są zwykle precyzyjne i czułe.

Mimo tych zalet, oznaczenia cyfrowe wciąż nie są powszechnie stosowane, głównie ze względu na konieczność stosowania nowych urządzeń pozwalających na podział próbki na ogromną liczbę kompartmentów (tysiące lub miliony), a następnie poddanie ich obróbce cieplnej i odczytanie sygnałów. Te wymagania wymagają bardzo skomplikowanych, a przez to również kosztownych urządzeń.

W tej pracy zostaną zaprezentowane metody optymalizacji oznaczeń diagnostycznych wraz z ich opisem matematycznym. Algorytmy podziału próbki pozwalają na zmniejszenie wymagań oznaczeń cyfrowych dzięki ograniczeniu liczby kompartmentów, na które musi być podzielona próbka.

Ponadto zostaną zaprezentowane metody cyfrowo-analogowe, które w synergiczny sposób łączą zalety qPCR i dPCR i mogą być przeprowadzone z użyciem standardowych urządzeń do qPCR. Wszystkie protokoły opisane w tej pracy pozwalają na ograniczenie liczby kompartmentów, jednocześnie zapewniając absolutny pomiar. Co więcej, opisane oznaczenia diagnostyczne mogą być dostosowane do konkretnych wymagań dotyczących precyzji pomiaru i zakresu badanych stężeń.

W pierwszym rozdziale zawarty został opis projektu dotyczącego optymalizacji oznaczeń diagnostyczych. Rozdział drugi zawiera krótki wstęp do diagnostyki molekularnej, wraz z opisem reakcji PCR. W rozdziale trzecim zostały określone cele pracy, natomiast w rozdziale czwartym przedstawiono metody matematyczne, numeryczne i eksperymentalne, które pozwoliły na znalezienie równań analitycznych pozwalających określić optymalny podział próbki, oraz przetestowanie działania tak zaprojektowanych oznaczeń. Rozdziały piąty, szósty, siódmy i ósmy opisują wyniki analizy oznaczeń cyfrowych o identycznych i różnych kompartmentach, optymalnie zaprojektowane oznaczenia cyfrowe i analogowo-cyfrowe. Podsumowanie wyników i dyskusja są zawarte w rozdziale dziewiątym. Ponadto praca zawiera suplement podzielony na trzy części: A, B i C, które zawieraja instrukcje pozwalające na zaprojektowanie własnego zonaczenia, oraz analizę jego działania.

Wyniki opisane w tej pracy zostały opisane w poniższej serii publikacji naukowych:

1. Debski, P.R., Gewartowski, K., Bajer, S., and Garstecki, P., *Calibration-free assays on standard real-time PCR devices.*, zgłoszone do publikacji

2. Debski, P.R., and Garstecki, P., *Designing and interpretation of digital assays: concentration of target in the sample and in the source of sample*, Biomolecular Detection and Quantification, dostępne online od 17 maja 2016

3. Debski, P.R., Gewartowski, K., Sulima, M., Kaminski, T.S., and Garstecki, P., *Rational design of Digital assays*, Analytical Chemistry, 2015, 87 (16), 8203-8209

oraz poniższych patentach i zgłoszeniach patentowych:

1. *Method for performing quantitation assays*, PCT/EP2013/000805 (patent przyznany w 2016; data zgłoszenia: 15 marca 2013)

2. *Method for performing quantitation assays*, PCT/EP2012/004792 (patent przyznany w 2016; data zgłoszenia: 19 listopada 2012)

3. *Sposób przeprowadzania cyfrowych oznaczeń analitycznych i diagnostycznych*, Polskie zgłoszenie patentowe P-399908 (data zgłoszenia: 11 lipca 2012)

4. *Sposób przeprowadzania cyfrowych oznaczeń analitycznych i diagnostycznych*, Polskie zgłoszenie patentowe P-399673 (data zgłoszenia: 26 czerwca 2012)

5. *Sposób przeprowadzania cyfrowych oznaczeń analitycznych i diagnostycznych*, Polskie zgłoszenie patentowe P-397026 (data zgłoszenia: 17 listopada 2011)

6. *Sposób przeprowadzania cyfrowych oznaczeń analitycznych i diagnostycznych*, Polskie zgłoszenie patentowe P-397027 (data zgłoszenia: 17 listopada 2011)

7. *Sposób przeprowadzania cyfrowych oznaczeń analitycznych i diagnostycznych*, Polskie zgłoszenie patentowe P-397028 (data zgłoszenia: 17 listopada 2011).

Contents

Acknowledgements **2**

Publications **4**

Abstract **7**

Streszczenie **10**

1 Introduction **18**

2 Quantitative assays in molecular diagnostics **20**

2.1 DNA-based Molecular Diagnostics 20

2.2 Model Reaction: Polymerase Chain Reaction 21

2.2.1 History of PCR: the revolution in molecular diagnostics 21

2.2.2 PCR procedure 25

2.2.3 Optimisation of PCR 27

2.2.4 Performance of PCR 28

2.2.5 Conclusions 28

3 The aim of the work **30**

3.1 Introduction 30

3.2 Systematic description and interpretation of digital assays 31

3.3 Optimisation of quantitation assays 31

3.3.1 Analytical formulas for classic digital assays 32

3.3.2 Multivolume digital assays 32

3.3.3 Rational digital assays - the concept of *active stripe* 33

3.3.4 Synergistic assays - combining real-time and digital signals 34

4 Methods **35**

4.1 Mathematical methods . . . 35
4.1.1 The Most Probable Number method . . . 36
4.1.2 Bayes' probability . . . 36
4.2 Model assumptions . . . 37
4.3 Numerical methods: Monte Carlo simulations . . . 38
4.4 Experimental methods . . . 39
4.4.1 Experiment 1 - performance of *rational digital* and *synergistic* algorithms 39
4.4.2 Experiment 2 - synergistic PCR assay is immune for initial sample buffer composition . . . 40

5 Systematic description and interpretation of digital assays 41
5.1 Introduction . . . 41
5.1.1 Applications . . . 41
5.1.2 Analytical procedure . . . 42
5.2 Mathematical routines . . . 42
5.2.1 Mathematical procedures . . . 42
5.2.2 Concentration of the analyte in the sample and in the source of the sample 43
5.2.3 Analytical description of the outcome of a digital assay . . . 46
5.2.4 Translation from dependent to independent analysis - convolution . . . 47
5.2.5 Translation from independent to dependent analysis - deconvolution . . . 48
5.3 Summary . . . 49

6 Optimized design of single-volume and multi-volume digital assays 50
6.1 Introduction . . . 50
6.2 Single-volume digital assays - new analytical solutions . . . 51
6.2.1 Estimate of the initial concentration . . . 51
6.2.2 Precision of the estimate - new analytical formula . . . 52
6.3 Multi-volume digital assays . . . 54
6.3.1 Derivation of the algoritms . . . 54
6.3.2 The outcome of the assay reaching its *optimum mode*. . . . 57
6.3.3 The standard deviation of the estimate and the dynamic range provided by an assay comprising a fixed number of compartments. . . . 58
6.3.4 Shannon Entropy as a measure of a digital assay's performance. . . . 59
6.4 The design of multi-volume digital assays . . . 60
6.5 Performance . . . 61

6.5.1 Performance of single-volume digital assays 61
6.5.2 Performance of multi-volume digital assays 61
6.6 Summary 63

7 Rational design of digital assays 65
7.1 Introduction 65
7.2 Mathematical routines 65
7.2.1 Analytical model of a single compartment of the assay - digital measurement 69
7.2.2 Estimation of concentration $E(C)$ and the uncertainty of this estimation $\sigma(C)$ for the simplest digital assay comprising 2 identical compartments. . 70
7.2.3 Analytical model of the microstate μ of the assay. 71
7.2.4 Distribution of standard deviation of the estimate of concentration from different microstates. 72
7.3 Derivation of the algorithm 73
7.3.1 Derivation of analytical formulas for assay design. 74
7.3.2 Rational design of digital assays 78
7.3.3 Libraries 78
7.3.4 Detailed description of precision-varied assays 79
7.4 Performance of Rational Digital assays 80
7.4.1 Information loss caused by not tracking the compartments 81
7.4.2 Resolution of the assays. 84
7.5 Complete list of design protocols 87
7.6 Experimental and numerical verification of Rational Digital algorithm 88
7.6.1 Numerical verification of the algorithms. 88
7.6.2 Experimental verification of the algorithms. 91
7.7 Summary 93

8 Synergistic analogue-digital design of quantitation assays 94
8.1 Introduction 94
8.2 Mathematical routines 96
8.2.1 Analytical model of one compartment of the assay - combination of the real-time and the digital measurement 96
8.2.2 Analogue signals - difference in time determines the ratios of numbers of molecules 97
8.2.3 Information content of a synergistic signal 99

8.3 Derivation of analytical formulas for assay design 101
8.3.1 The architecture of the synergistic assay 101
8.3.2 Derivation of analytical formulas for assay design 102
8.3.3 Designing synergistic PCR assay 103
8.3.4 Detailed decription of precision-varies synergistic assays 104
8.4 Experimental and numerical verification of the synergistic alorithm 105
8.4.1 Experimental verification of the synergistic assay 105
8.4.2 Numerical verification of the synergistic assay with Monte-Carlo simulations 108
8.4.3 An example of the use of a synergistic PCR assay: synergistic PCR assay is immune for initial sample buffer composition 108
8.5 Summary 110

9 Discussion **112**
9.1 Analytical description of digital assays 112
9.2 Optimized digital assays 113
9.3 Multivolume digital assays 113
9.4 Rational digital assays 114
9.5 Synergistic analogue-digital assays 115
9.6 Practical applications of the algorithms 116
9.6.1 Point-of-Care applications 117
9.6.2 Multiple assays on a single 96-well plate 117
9.6.3 Copy Number Variation 117
9.7 The possible application of the algorithms for digital counting of bacteria 118

A Single-volume digital assays **120**
A.1 Step by step instruction on how to analyze the results of a digital PCR assay . . 120
A.1.1 Independent scheme, Bayes formalism 120
A.1.2 Dependent scheme, Bayes formalism 122
A.1.3 Independent scheme, Most Probable Number method 123
A.1.4 Dependent scheme, Most Probable Number method 125

B Rational digital assays **127**
B.1 Practical guideline how to design a digital assay that provides the required dynamic range and precision of the assessment 127
B.1.1 Design of an assay with tracking of the identity of each of the compartments. 128

B.1.2 Design of an assay, with a limitation on minimum gradation of compartments, with tracking of the identity of each of the compartments. 129
B.1.3 Design of an assay without tracking of the identity of each of the compartments. 131
B.1.4 Design of an assay, with a limitation on minimum gradation of compartments, without tracking of the identity of each of the compartments. . . . 133
B.2 Step by step instruction on how to analyze the results of a rational design PCR assay 135

C Synergistic analogue-digital assays **137**
C.1 Practical guideline how to design a synergistic assay that provides the required dynamic range and precision of the assessment 137
C.1.1 Detailed description of the design of an optimum synergistic assay. 138
C.1.2 Detailed description of the design of an assay, with a limitation on the minimum gradation between the compartments. 139
C.2 Step by step instruction on how to analyze the results of a synergistic PCR assay 141
C.3 Calculation the amplification factor q 143

Bibliography **144**

Chapter 1

Introduction

DNA-, RNA-, and immuno-diagnostics are gaining popularity and therefore there is a growing demand for new solutions that provide reliable and precise qunatitative assessment of the analyte in a wide range of concentrations. The methods based on the Quantitative Polymerase Chain Reaction (qPCR) play an important role in molecular biology and are used commonly in medical diagnostics. One of the formats of qPCR methods is the real-time PCR [3, 4, 8, 9], introduced in 1991 by Holland et al [1]. This technique assesses the concentration of the analyte by tracing the increase of signal (usually the intensity of fluorescence from DNA product provided by PCR) and recording the number of PCR cycle after which the signal reached a threshold value. Then, this number is compared to the number of cycles needed to reach a treshold for calibrated references.

The state-of-art real-time PCR methods provide the assessment of a wide range of concentrations of target DNA. However, the need for calibration with a reference may compromise the accuracy and precision of the measurement. The sources of error root from the quality of PCR reagents and characteristis of the apparatus. Still, despite these problems, qPCR is recognized as a *golden standard*.

An alternative concept called *digital PCR assay* (dPCR) alleviates the need for calibration and provides absolute assessment. It requires the partitioning of the sample into a large number of separate partitions that are thermally treated for PCR, and the initial concentration of target DNA is estimated from the fraction of end-point binary (yes/no) signals from partitions.

Digital assays are expected to replace classic qPCR as they provide absolute, highly sensitive and precise [4, 10] estimates. However, their popularity is limited because the digital design is not flexible (precision and dynamic range of the assessment cannot be tuned independently),

and they require the partitioning of the sample into extremely large numbers of partitions: from tens of thousands to millions, which increases the complexity and the cost of a single test, and therefore narrows the portfolio of applications.

Therefore, the further popularisation of digital assays depends critically on their optimisation. The obvious directions of this development are the simplification of the partitioning of the sample and the possibility of running digital assays in standard well plate format in stock qPCR devices. This requires limiting the number of compartments of an assay (i.e. partitions of the sample).

The optimisation described in this work is achieved by a thorough analysis of the behaviour of classic digital assays and the information gain from digital signals yielded by compartments comprising an assay. This, in turn, allows designing new multi-volume, or multi-dilution schemes that require orders of magnitude less compartments compared to classic single-volume digital schemes to provide the assessment of the initial concentration of the analyte within a requested dynamic range. Such constructed assays are easily tunable and can be executed in a well plate format in stock qPCR devices.

The algorithms described in this work teach how to set the product of volume and dilution of the sample in the compartments comprising an assay in order to provide the requested precision of the assessment within the requested range of concentrations of the analyte.

Finally, we will present a method that synergistically combines the advantages of qPCR and dPCR and bypasses their drawbacks. The synergistic scheme allows futher simplification of the partitioning of the sample, and the improvement of the precision of the assessment.

The optimisation presented in this work may accelerate the spread of digital assays by lowering the technical requirements for running them and providing the flexibility of the assay's design, which widens the protfolio of possible appliactions. Moreover, the multivolume and multidilution schemes presented here can be readily executed in stock qPCR devices, alleviating the need for new hardware.

Chapter 2

Quantitative assays in molecular diagnostics

2.1 DNA-based Molecular Diagnostics

Molecular diagnostics is currently one of the fastest developing areas in medical and biological research [11–14]. It is focused on establishing and elaborating new techniques that can be used to analyse the biological markers in genome, transcriptiome, and proteome. As molecular diagnostics methods are intrinsically targeted at the specific molecular causes of diseases in a patient, their development is a milestone in achieving the idea of the personalized medicine. Moreover, the molecular diagnostics methods have become a vital tool for biological and biochemical research, biotechnological industry, forensic diagnostics and other fields of use.

Diagnostic studies on DNA date back to 1978 and the research on sickle-cell anemia carried by Kan and Dozy [15], and independently by Orkin [16]. At the beginning, the development of the DNA-based molecular diagnostics was possible thanks to the introduction of restrictive enzymes and molecular probes, the development of DNA cloning [17], Southern blot [18], DNA sequencing [19], labeling, and detection, and in further perspective, the Polymerase Chain Reaction [20, 21] and oligonucleotides synthesis. In the 90s, the DNA-based molecular diagnostics was gaining popularity, and nowadays it is recognized as a routine clinical tool. Also, the perspectives of future development are already well-stated.

Further development of molecular diagnostics is now focused on the following challenges:

- limiting the time of analysis (i.e. the turnaround time),
- increasing the throughput,
- lowering the detection limit,
- increasing the accuracy,
- improving quantitative analysis,
- improving Point-of-Care methods, and
- lowering the cost of analysis.

Molecular diagnostics of DNA/RNA can be divided into two areas:

- direct diagnostics, which covers the detection of specific target DNA sequences, and
- indirect diagnostics, which determines the presence of target sequences on the basis of the analysis of conjugations of mutations with a given locus.

Currently, the methods of molecular diagnostics are usually aided with the amplification of specific sequences by means of the Polymerase Chain Reaction (PCR), or with restrictive analysis and hybridization with molecular probes. Some diagnostic tests are based on both amplification and hybridization.

Still, in the nearest future, it is expected that the detection of specific DNA sequences will become routine in various fields, as they are now in the treatment of genetic diseases, cancer, and pathogen detection.

2.2 Model Reaction: Polymerase Chain Reaction

2.2.1 History of PCR: the revolution in molecular diagnostics

The popularization of the Polymerase Chain Reaction (PCR) was a breakthrough in DNA-based molecular diagnostics. It was introduced by Mullis and Faloona in 1986 [8, 22], and provides an almost infinite amplification of a target DNA sequence, facilitating its detection. For his discovery, Mullis was awarded the Nobel Prize in Chemistry in 1993. From the beginning, PCR was expected to be a useful tool for the qualitative detection and, in further perspective, the quan-

titative assessment of the presence of nucleic acids in clinical samples. Its potential applications in medical and clinical analysis range from the diagnostics of genetic [23–25], parasitic [26–28], and infectious diseases [29–37], through the cancer detection and monitoring after chemotherapy, measurement of viral load suppression [38–49] and finally the paternity testing [50–52].

The PCR was able to detect as little as a few copies of DNA in the sample – the possibility of detecting extremely low signals (i.e. concentration of the analyte) [53, 54] boosted the development of techniques aimed at the early-stage cancer detection and monitoring, the monitoring of the development of infections and non-invasive tests using cell-free DNA [55–58].

On the other hand, the excellent sensitivity of the reaction was a problem while developing quantitative PCR-based tests as even a small amount of contamination could severely offset the result. Moreover, as the sensitivity of PCR was much better than of the sensitivity of the common techniques at the time (northern [59] and Southern [18] blots), it was difficult to verify whether PCR-based tests provide accurate quantitation. Therefore, the quantitative PCR-based methods were believed to be unattainable for a long time, and first handbooks on PCR did not even mention the possibility of a quantitative assessment [60–62].

The main problem was the feasibility of the reaction. Initially, the heat-labile Klenow fragment of the *Escherichia Coli* pol-1 DNA polymerase was used [63], and it had to be added in every cycle of PCR. After changing it with the heat-stable polymerase *Taq* isolated from *Thermus Aquaticus* [64, 65], the execution of PCR tests was dramatically simplified. Since then, the *Taq* polymerase was used as the basis or comparator of PCR [66]. This improvement allowed the development of quantitative PCR (qPCR) techniques.

The crucial improvements that led to the development of modern qPCR techniques, were:

- the optimisation of the thermal conditions of a PCR cycle, which let to the development of the rapid-cycle PCR (30s per cycle instead of 3-5 minutes) by Wittwer et al. [67], and
- the development of real-time detection using ethidium bromide ($EtBr$) for the monitoring of the product synthesis [68].

In the meantime, the details of the reaction, and detection schemes have been improved [69]. These improvements are described in the following sections. The PCR techniques became automated and simplified up to stage where they are used for routine tests in a form of the real-time PCR.

The real-time PCR [70–74], also known as the rapid-cycle real-time PCR, was developed in 1991 by the group of Wittwer and Garling [75] at University of Utah and Idaho Technologies Incorporated. The reaction mixture was sealed in 1-10 μL volumes in capillary tubes that were heated and cooled using compressed air. This shortened cycle times to less than 30 seconds.

The first attempt to use a real-time monitoring for quantitative analysis was made in 1993 by Higuchi et al. [68, 76]. The increase of fluorescence from $EtBr$ used as a detector was monitored with a video camera. The authors explicitely stated that the procedure they proposed was an improvement of PCR towards the automation and widespread use in the situations requiring high sample throughput.

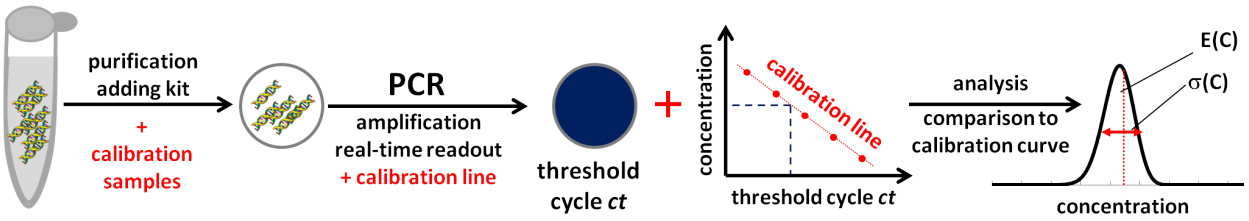


Figure 2.1: A schematic of a real-time PCR procedure.

Analogue assays based on the real-time PCR technique have found a wide variety of applications in biochemistry and diagnostics and are used commonly for the assessment of concentration of target DNA fragments. They present a range of advantages. First of all, they require a very small sample and no (or very simple) partitioning for assessment and easily determine the relative changes of a number of analyte particles or the concentration of the analyte. Analogue PCR techniques are also relatively quick, as the whole analysis can take only up to one hour, and relatively robust thanks to the detection based on the use of molecular probes.

Nowadays a wide range of real-time PCR instruments is available, yet the procedure of estimating the concentration of the target nucleic acids is the same. It is based on recording the number of PCR cycle after which a threshold value of signal was reached and referencing it to the signal from an externally calibrated reference sample containing a known concentration of the target. In practice, due to random and systematic changes in the choice of substrates, analyte particles, the sensitivity of the sensor, the condition of the apparatus, etc., calibration needs to be performed frequently. Importantly, the accuracy of the estimate of concentration obtained via real-time PCR procedure depends on the quality of external calibration and cannot be assessed at the point of measurement.

In spite of these limitations, real-time PCR remains the *golden standard* in quantitative PCR. The

advantages include simple liquid handling protocols that do not require extensive partitioning of the sample and straightforward mathematical procedures for obtaining the final result from the measurements on the sample and from the calibration curve.

An alternative approach towards the quantitative PCR-based assessment, known as the *digital* assay, roots from the works of Phelps in 1908 [77] McCrady in 1915 [2], when they introduced a limiting-dilution assay made of fermentation tubes for counting bacteria. PCR was adopted for digital quantitation in the beginning od 1990s by Sykes et al. [3] and then by Vogelstein and Kinzler [4, 78, 79]. This solution is based on the partitioning of the sample into a large number of partitions. These partitions are then treated separately for PCR (thermocycling) and inspected for the occurrence of a binary (positive or negative) signal. The presence of an *a priori* known threshold number of analyte particles (in case of digital PCR a single particle is enough to give a positive readout), or a threshold concentration of the analyte in the inspected partition is amplified to obtain a measurable *positive* signal (*positive* value). The recorded signal is often in a form of high/low fluorescence from DNA probes. However, the recorded signal may have a different physical nature, including the change of viscosity [80–82] due to the increasing amount of large molecules during the progress of the reaction, change of turbidity [83–85], or scattering parameters of the sample [86–89] etc. The concentration of analyte particles is assessed with the use of appropriate statistical models on the basis of the count of positive signals from the total of partitions of the sample.

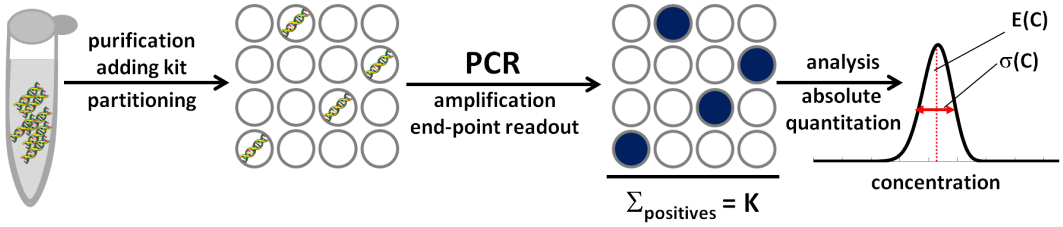


Figure 2.2: A schematic of a digital PCR procedure.

With the development of digital assays, a new paradigm of quantitation was introduced to analytical chemistry. It provides an absolute assessment of even very minute concentrations of the analyte, and does not require calibration of the experimental set-up using standardized samples. Moreover, the laboratory routines are simplified as digital assays benefit from the end-point readout of a binary signal, compared to real-time tracking of signal in *golden standard* qPCR systems, as is the mathematical interpretation of the results. However, the spread of digital techniques is still limited due to the limitations they are affected.

First of all, digital assays require a complicated partitioning of the sample: tens of thousands, or even millions of partitions are needed to cover practically useful range of concentrations [90–92]. Furthermore, the design of a digital assay is fixed and depends solely on the number of partitions, i.e. it is impossible to tune the assay independently to cover a specified dynamic range or to provide a specified precision. For example, it is not possible to design a digital assay that provides a high precision of the assessment in a narrow range of concentrations, or a low precision, but in wide range of concentrations.

Some solutions have been tried to overcome these problems. One possibility is the combination of a number of digital assays, each of them providing assessment within a different dynamic range [5–7, 93]. In practice, such assays use multiple sets of identical compartments, while each set is diluted in order to widen an achievable dynamic range. However, the multivolume (or multidilution) approaches have not been optimized for the information gain, and its dependence on the design of the assay, i.e. the number and parameters of compartments. In the following chapters, this analysis will be presented and will lead to the formulation of equations for the design of optimized digital assays.

2.2.2 PCR procedure

The experimental procedure of PCR comprises the steps of mixing a sample containing the DNA template with DNA fragments that should be replicated, with a buffered solution of PCR reagents, which contains deoxyribonucleotide triphosphates (dNTPs), primers, potassium ions, thermally stable DNA polymerase and polymerase cofactors (mostly Mg^{2+} or Mn^{2+} ions). The primers, called also starters, are short DNA fragments that typically have a length of 15-25 base pairs, and determine the replicated sequence. There are two types of primers:

- the forward primer which has an identical sequence as the start section of the fragment that should be replicated, and
- the reverse primer which has a complementary sequence to the terminal section of the fragment that should be replicated.

The most commonly used enzyme for PCR is the thermally stable *Taq* polymerase, which was isolated from the bacterium *Thermus aquaticus*. However, for some applications other enzymes and their modifications should be used. For example, long DNA fragments can be amplified using *Pfu* polymerase isolated from *Pyrococcus furiosis* [94–96].

The amplification with PCR relies on the cyclic heating and cooling of the sample [75, 97, 98]. A single PCR cycle is composed of three stages:

- DNA denaturation stage - the sample is heated to approximately 96 degrees (depends on the reaction), then the melting of DNA takes place and the double-stranded DNA helix (dsDNA) is unwind to optain a single-stranded DNA (ssDNA). Usually, the first denaturation (i.e. in the first PCR cycle) is longer than in further cycles.

- Annealing stage - the sample is cooled to the temperature of 45-60 degrees (depends on the sequence and length of the primers used) and the primers anneal to the DNA template in specified places at the ends of the DNA fragment that should be replicated. The primers should be designed in such a way that it anneals in only one location to the template.

- Elongation stage - the sample is heated to approximately 72 degrees and the replicated DNA fragment is synthesised. The syntesis starts when the DNA polymerase bonding. Then, new dNTPs are attached and the DNA fragment is extended.

Theoretically, after every PCR cycle, the number of target DNA fragments present in the sample doubles. Therefore, after 30 PCR cycles, a single amplicon is transformed into one billion copies of the replicated DNA fragment. This gigantic level of amplification makes it easy to detect the signal both for quanlitative and quantitative assays.

The number of copies M_n after n cycles of amplification can be given by the equation:

$$M_n = M_0(1 + E)^n,$$

where M_0 is the initial number of copies and E is the efficiency of the amplification. If the amplification is perfectly efficient, it equals unity, i.e. $E = 1$. However, usually it is lower, i.e. $E \approx 0.95$. In this work, instead of the efficiency of the amplification E, the so-called amplification factor q will be used, which equals $q = 1 + E$.

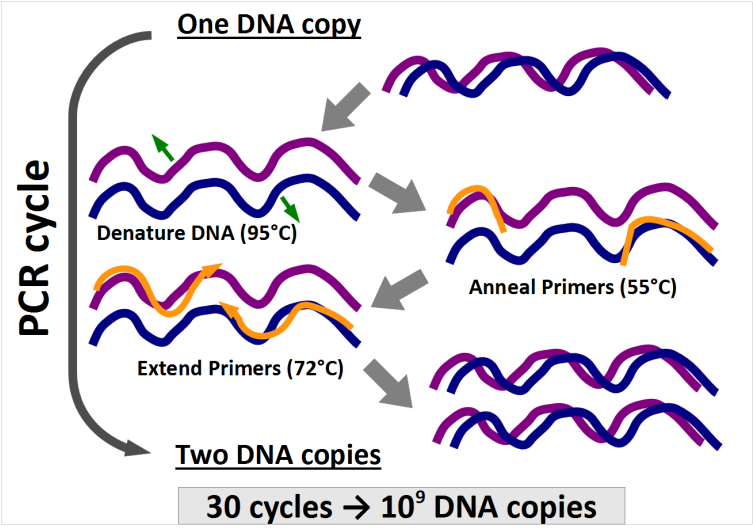


Figure 2.3: Scheme of a single PCR cycle. It comprises 3 phases: denaturation of the double-stranded DNA, annealing of the primers to single-stranded DNA, and elongation of primers with polymerase enzyme. Ideally, each PCR cycle allows to double the number of DNA copies, therefore the increase of signal in time is geometric.

2.2.3 Optimisation of PCR

High efficiency of the amplification is one of the most important requirements for the successful quantitation of the target DNA fragments. The optimisation of the reaction includes [99]:

- choice of target region of DNA and primer binding sites,
- extraction of DNA,
- concentration of buffers (containing Mg^{2+} and Mn^{2+} cations),
- temperature and duration of each stage of a thermocycle, and
- detection (detector probe).

The selected target fragment, as well as the primer and probe binding sites, should be unique and not prone to mutations. Moreover, the target should not have any structures that could influence the amplification, and its structure should not obstruct the dissociation. It is highly recommended that the primers have similar base ratios for similar binding curves, and undesirable if the primers and probes bind to one another and give false signals. However, nowadays the catalogs of popular nucleic sequences are available, and they simplify the right selection of the sites.

Also, the template must not be polluted or degraded [100]. Current methods for DNA purification are focused on fast and efficient isolation of DNA fragments without any proteins or enzyme inhibitors [101, 102]. The most important parts of the process are collecting and marking the sample, its transport, storage, and isolation of DNA.

Enzymes that synthesise DNA play a key role in the reaction. Polymerases vary in efficiency, accuracy, and the proofreading capabilities of the synthesized DNA fragments [103, 104]. As an example, the probability of building in a non-complementary nucleotide is usually between 1/10,000 and 1/100,000 nucleotides in case of *Taq* polymerase, and the probability of getting errors increases with the number of cycles [105]. Also, the half-life of polymerase activity depends on temperature [106, 107], and decreases from 2 hours in 92.5 degrees to 40 minutes at 95 degrees, and even 5 minutes at 97.5 degrees.

2.2.4 Performance of PCR

The proper optimisation of the reaction, including the choice of the target region, binding sites, and reagents, is required for the efficient amplification and quantitative assessment of the presence of the targeted sequence. If the DNA fragments of interest are present in the sample, they are quickly amplified by the PCR reaction; the number of amplicons increases geometrically with the number of PCR cycles. The increase of the number of DNA copies can be monitored in every PCR cycle (real-time PCR), or at the end of the reaction (dPCR) by measuring the level of fluorescence (from fluorescent probes). On the other hand, no fluorescence and therefore no signal is a sign that no DNA fragments of interest were present in the sample at the beginning of the reaction.

2.2.5 Conclusions

In the past 30 years since the introduction of PCR revolutionized the area of molecular diagnostics, and 25 years since first attempts to quantitative PCR were made, the PCR-based molecular techniques has been optimized and popularized to become a common tool in diagnostics. However, there is still a possibility of improvement, especially in the field of digital PCR assays, that despite the obvious advantages, do not gain appropriate recognition as they are difficult to handle and not flexible. Moreover, the commonly used methods of quantitative PCR show complementary features, but they have never been combined. The analogue qPCR techniques gain advantage in simplified partitioning of the sample (they require only a single partition of the

sample), but require calibration of the experimental set-up. On the other hand, the digital PCR requires complicated partitioning, but provide absolute assessment. A synergistic combination of the two techniques may bring a new standard to quantitative assaying.

Chapter 3

The aim of the work

3.1 Introduction

The popularization of DNA-, RNA- and immuno-diagnostics will require new techniques for reliable detection and precise qualitative assessment in a wide range of concentrations. The PCR-based methods play a central role in molecular biology and have become a common tool in medical diagnostics. The digital PCR (dPCR) assays [4, 10] provide precise quantitation and thanks to absolute nature of assessment, they alleviate the need for calibration. They encode the information about the concentration of the analyte in the count of binary signals collected from a large set of compartments (partitions of the sample) [4], simplifying the mathematical and experimental routines. Offering these attractive features, digital techniques are expected to replace the *golden standard* real-time methods (qPCR). However, the popularity of digital techniques is growing slowly because they need complicated partitioning of the sample (a large number of partitions) for the assessment. This requires new complex devices and more consumables compared to traditional methods, increasing the cost and limiting availability. Digital analytical assays are typically parameterized by the precision of the assessment and the dynamic range of assessed concentrations. These parameters depend on the mathematical and statistical interpretation of the signals, which enforces different methods of calculation of the result. The proper understanding of the behavior of digital assays is also a starting point for their optimization. In the following chapters, the analytical procedure will be explained, as well as the sources of errors, which will subsequently lead to the proper (better) understanding of the estimate of the concentration.

Moreover, a proper understanding of the behavior of digital assays will be then a basis for the optimization and the description of new, optimum (rational) design of digital and analogue-digital assays, that will:

- provide for a lower number of compartments needed for the assessment of the concentration, and
- allow the independent adjusting to the requested dynamic range and precision of the assessment.

3.2 Systematic description and interpretation of digital assays

In the first chapter of the analysis (Chapter 5), the behavior of digital assays comprising identical compartments (i.e. partitions of the sample, or test-volumes), which will be also referred as *classic digital assays*, will be discussed. Their design will be rigorously investigated, concerning the full analytical procedure and the sources of errors.

Also, the proper understanding of the estimate of the concentration of the target marker will be given. Since digital PCR assays are primarily directed to the quantitation of small concentrations of diagnostic markers it is important to understand how the estimates provided by the assays represent the concentration in the source. In this chapter, the two alternative methods of analysis of digital estimates and their significant differences will be discussed explicitly, as well as the propagation of errors in the analytical procedures. It will be also shown how to translate the result referring to the concentration in the sample into the concentration in the source of the sample, including the significant change in the breath of the confidence intervals, being the result of different statistical approach. The findings of this research will be then used to propose an optimum design of digital assays to address the requested range of concentrations with the required precision.

3.3 Optimisation of quantitation assays

A portfolio of optimum designs of analytical assays will be discussed in chapters 6-8.

3.3.1 Analytical formulas for classic digital assays

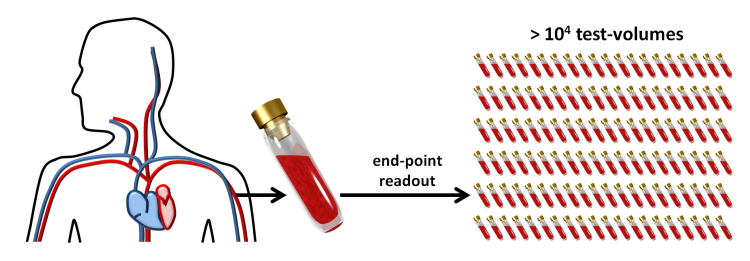


Figure 3.1: Classic digital assay. The sample is split into a large number of identical compartments, which are inspected for the presence of a binary signal.

In this chapter, the design of classic digital assays will be given with a detailed analysis of the response of an assay. The design, including the requirement for the minimum number of partitions and their recommended volume, depends significantly on whether the assay is to assess the concentration of the target analyte in the sample or in the source of the clinical sample (e.g. a patient body), providing a given, requested precision. These results will be used to design multivolume digital assays for a requested dynamic range and precision of the estimate of concentration.

3.3.2 Multivolume digital assays

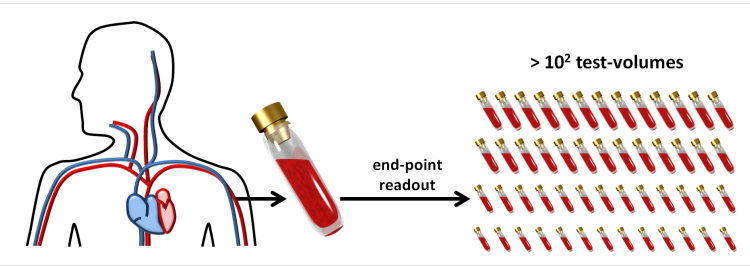


Figure 3.2: Multivolume digital assay. The sample is split into sets of identical compartments, which are inspected for the presence of a binary signal. The compartments belonging to different sets vary in volume and/or dilution.

In the following, the calculated dependencies of dynamic range and standard deviation of the estimate from the *dependent* and *independent* analysis will be used for the derivation of equations that allow designing digital assays that address the requested dynamic range of concentrations with the required precision. In short, this will be achieved by joining many classic digital assays. Such approach will provide assays with the precision and dynamic range tuned independently

and requiring a limited number of compartments. They will be most suitable for the assessment that requires high precision (standard deviation of the estimate smaller than 10%).

3.3.3 Rational digital assays - the concept of *active stripe*

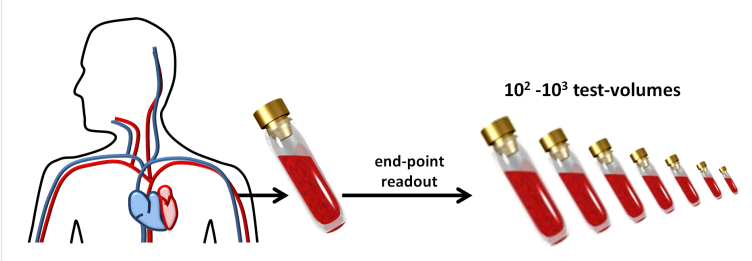


Figure 3.3: Rational design of digital assay. The sample is split into non-identical compartments, which are inspected for the presence of a binary signal. All the compartments vary in volume and/or dilution.

In the next chapter it will be shown that it is possible to design an assay which comprises non-identical compartments that are uniformly distributed on the logarithmic scale (i.e. the expected numbers of molecules of the analyte m_i per compartment are uniformly distributed on the logarithmic scale) [108]. Such approach provides assays with the precision and dynamic range tuned independently and requiring a limited number of compartments. The method described there minimizes the number of partitions and provides explicit formulas for the design of the assay for the requested dynamic range and precision of the estimate. This optimization is achieved via appropriate tuning of the expected number of molecules of the analyte (i.e. DNA molecules) in the partitions. The algorithms there describe explicitly teach how to adjust the modification factor (i.e. the product of volume and dilution of the sample in each of the compartments). In the rational algorithm, the chemical compartments are treated as bits of probabilistic information and are arranged these in a fractional positional system. Thanks to the maximization of information gain, the number partitions required to provide assessment within a requested dynamic range and with a requested precision is reduced by by orders of magnitude.

3.3.4 Synergistic assays - combining real-time and digital signals

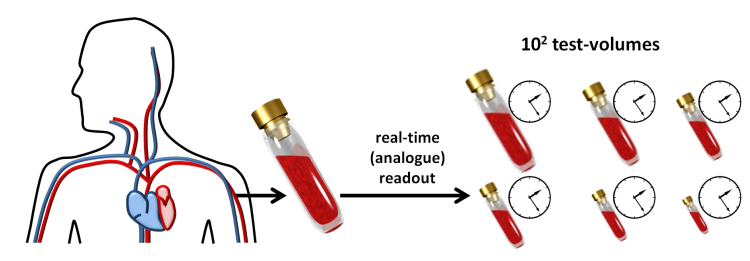


Figure 3.4: Synergistic analogue-digital assay. The sample is split into non-identical compartments, which are inspected for the presence of a real-time (analogue) signal (i.e. the increase of the signal is traced and the number of PCR cycles after which a threshold value was reached, is recorded). All the compartments vary in volume and/or dilution.

At the end, there will be a discussion of an analog-digital method that can be executed on standard real-time PCR devices. The method synergistically combines the advantages of the real-time and of the digital PCR and bypasses their drawbacks. In the synergistic PCR scheme, the digital signals auto-calibrate the measurement while the analog measurements refine the estimate of concentration. We also list a practical recipe how to design the assay for the required dynamic range and precision of the estimate, and how to analyze the signals to extract the estimate of concentration. The combination of the digital and analog information provides for absolute quantitation with adjustable resolution. We also describe the procedure and derive a prescription for designing synergistic digital-analogue PCR assays and for analyzing the results.

Chapter 4

Methods

4.1 Mathematical methods

The outcome of the experimental procedure of digital assessment by means of amplification reactions like PCR, i.e. the ratio of positive and negative test-volumes of the sample, has to be mathematically interpreted in order to provide the final result of the assessment, i.e. the value of initial concentration (or number of copies) of the target analyte.

The first analytical descriptions of digital experiments (i.e. the experiment that provide a set of yes/no results) date back to the first decades of the XXth century and the works of Phelps [77] and McCrady [2]. They dealt with the estimation of the initial number of *Escherichia coli* using tube dilution tests. The calculations based on the outcome from a set fermentation tubes that provided only binary results.

In the Phelps' approach, the number of bacteria in the test was calculated as a reciprocal of the smallest volume of sample (i.e. from the most diluted test-volume) for which the positive result was recorded. However, in this approach, the interpretation of the results was confusing if the sequence of signals from dilution was not regular (i.e. if the signals recorded from the sequence: 1111110110000000, where '1' means positive and '0' means negative readout).

Another approach was presented by McCrady in 1915, who presented a deeper mathematical analysis of the experimental results. He estimated the initial concentration as a concentration that yields the highest probability of obtaining the particular recorded outcome of the experiment (i.e. ratio of positive and negative tubes).

4.1.1 The Most Probable Number method

The method of getting quantitative data on concentrations of discrete items from positive/negative data proposed by McCrady is now called the Most Probable Number (MPN) method [5–7, 109–112], and is commonly used for quantitative analysis of various experiments, including growth of microorganisms, enzyme action, or catalytic reactions. Since 1915 the method was elaborated, although the original tables calculated by McCrady were still in use [113]. The MPN method is usually used to for the quantitative assessment of large numbers of discrete entities (i.e. bacteria, DNA fragments, proteins, etc.) that can be detected, usually via a sort of amplification reaction (PCR or the incubation of bacterial colony), but they are difficult to count. The MPN method requires dividing the sample into sets that vary (are diluted) by orders of magnitude (frequently 10-fold or 2-fold), and assessing presence of the assessed entities in partitions belonging to sets.

The dilution at which no presence of the assessed entities has to be recorded. Then, the initial number of the entities is calculated as the number, for which the probability of obtaining a given readout was the highest.

The MPN method can be also applied for problems of molecular biology and tests involving DNA templates amplified by the Polymerase Chain Reactions (PCR). Another application involves enzyme-based immunoassays including ELISA (Enzyme-Linked ImmunoSorbent Assay) [114] or anly other antibody cascade detection reaction, which allow the assessment of the initial concentaion of a sought enzyme or antigen.

4.1.2 Bayes' probability

An alternative approach to the analysis of experimental data is based on the Bayes reasoning. Bayes' formalism inverts the concentration dependent probability of the recorded result to yield the probability distribution of a concentration having caused the result [115–117]. Such obtained probability distribution provides the estimate of concentration together with its standard deviation.

The Bayesian reasoning starts from the general case of m template molecules randomly distributed within N wells, and the number k is the count of the wells that yielded a positive signal. The conditional probability $p(k|m)$ that k signals were caused by m template molecules can be calculated using combinatorical equations or Poisson statistics. A specific example of digital signals will be discussed in detail in the next two chapters. Then, the procedure is

implemented using the following steps:

1. If we assume a uniform probability distribution for m over a finite range, from 1 to M_{max}, then the probability $p(m)$ of any particular value of m is constant and equals $1/M_{Max}$ for all m within that range. Substituting gives us the following equality: $p(m|k) = p(k|m)/M_{max}$

2. Thus, given a table of $p(k|m)$ values (from numerical simulations), we can readily compute a table of $p(m|k)$ values: for every pair m and k, we divide $p(k|m)$ by the sum of $p(k|i)$ for $i = 1$ to M_{max}; this sum converges for large M_{max}, i.e. the result becomes independent of the choice of M_{max}

3. For any given k, $E[m]$ is the weighted average obtained by summing the product of each m value by the corresponding $p(m|k)$ entry.

4. Confidence intervals on $E[m]$ can be derived by building a cumulative distribution function table from the $p(m|k)$ table.

5. Another useful uncertainty metric, the coefficient of variance (CV), which equals the relative standard deviation of the estimate and will be also called the precision of the assessment, is calculated as follows: $\sigma = CV = \sqrt{E[m^2] - E[m]^2}/E[m]$.

In this work, we use the Bayes' formalism in the process of the formulation of optimum algorithms, unless it is stated otherwise. Still, the conclusions are valid also for the MPN formalism.

4.2 Model assumptions

The analysis that leads to the optimisation of quantitation assays presented in this work is based on three assumptions. First, the Bayes' probability is used to derive probability distribution of the initial concentration C of the analyte. Second, for *rational digital* and *synergistic* algorithms presented in chapters 7 and 8, the most probable outcome from an assay for a given concentration, called also a *microstate*, will be used to calculate the precision of the assessment, which is given as a the standard deviation of the probability distribution of the initial concentration C. Still, we have tried also other approaches, including the determination the precision of the assessment with the tools provided by the Information Theory: by the amount of Fisher Information [5] and Shannon Entropy. In case of Fisher Information, the lower limit of the standard deviation of the estimate is given by the Cramer-Rao bound. Unfortunately, in many cases the existence of an ideal estimator is not determined, therefore our methods were based on the most probable

precision of the estimate. Finally, we assumed a perfect amplification of a single molecule, i.e. the efficiency of this amplification is 100% and is independent form the parmeters of a compartment (volume or dilution). Still, in the real samples the efficiency may be lowered depending on the surface-to-volume ratio the dilution of the sample. Therefore, is is advisable to run run a series of control experiments first.

4.3 Numerical methods: Monte Carlo simulations

We used Monte Carlo simulations to establish the design formulas for quantitation assays. There were run by means of the random number generators provided by the ROOT framework [118]. We used the terms *canonical* and *grand canonical* Monte Carlo simulations analogically to canonical and grand canonical ensembles to differentiate between two situations:

- The *grand canonical* Monte Carlo simulations used as an input the initial concentration C_A of the analyte in the assay of a total volume V_A. Then, each compartment with volume $v = V_A/N$ was treated individually, i.e. the each compartment gained randomly a positive signal with probability $p(s = 1) = 1-e^{-C_A v}$ or negative signal with probability $p(s = 0) = e^{-C_A v}$. Therefore, there was a randomness of the distribution of the number of target molecules M_A in the assay (i.e. Poisson distribution with expected value $\lambda = C_A V_A$) and the distribution of these molecules among the compartments. These simulations were used to produce probability distributions $p(K|C_A)$, where K is the count of positive compartments, that were later used for Bayes' method.

- The *canonical* Monte Carlo simulations used as an input a hard-fixed initial number of molecules of the analyte M_A in the assay. Then, each molecule was randomly distributed among the compartments with uniform probability. At the end, the compartments containing at least one molecule of the analyte were given a positive signal or negative signal otherwise. Therefore, there was a randomness only in the distribution of the molecules among the compartments. These simulations were used to produce probability distributions $p(K|M_A)$, where K is the count of positive compartments, that were later used for Bayes' method.

4.4 Experimental methods

4.4.1 Experiment 1 - performance of *rational digital* and *synergistic* algorithms

Materials

The reaction was performed in the volume of 20 μL, consisting of 4.5 μL of diluted plasmid DNA, 125 nM of forward and reverse primers. The forward primer sequenced TCTTGCCCTCTTTCTGCTTC and reverse primer sequenced GATCGGCTCGAGAATCATTGCG were used. Also 10 μL of SensiFAST SYBR No-ROX mix [119] was added to the reaction mixture.

Methods

We used the pJET1.2 plasmid with a fragment of LepA gene cloned from *Mycobacterium smegmatis*. The initial concentration of DNA was quantified with the use of a NanoDrop device. DNA used for all tests were stored in frozen aliquots.

A three-step amplification protocol was performed in 7500 Fast Real-Time System [120]; an initial denaturation was performed with one cycle at 95 degrees for 10 min. Subsequently, target amplification involved 50 cycles of 15 s at 95 degrees, 25 s at 62 degrees for annealing, then extension for 15 s at 72 degrees. After amplification cycles, PCR products were evaluated for quality using melt curve analysis, which entailed 15 s at 95 degrees, 1 min at 70 degrees, 15 s at 95 degrees and 1 min at 55 degrees.

14 different DNA concentrations were tested from 0.08 to 500,000 DNA particles in first well (from 0.004 to 25,000 particles/μL). The geometric sequences of the modification factors of compartments comprising tested assays were made via multi-dilution approach, i.e. the volume of all the compartments were same and the dilution factor changed geometrically.

4.4.2 Experiment 2 - synergistic PCR assay is immune for initial sample buffer composition

Materials

All experiments were prepared on IVD certificated PCR kit for *Cytomegalovirus* detection [121]. Internal calibrator from the kit was used as a DNA template after 400 times diluted in water or 3 different elution buffers from commercially available DNA isolation kits: AE elution buffer from QIAamp DNA Mini Kit [122], MBL5 elution buffer from NucleoMag Blood [123] and MagJET elution buffer from Whole Blood Genomic DNA Kit [124] to obtain model samples with 25,000 copies of the target DNA per mL.

Methods

To compare the traditional RT-PCR with digital approach three-step amplification protocol was performed in 7500 Fast Real-Time System [120] according to *Cytomegalovirus* PCR kit prescription: UGD decontamination 37 degrees for 2 min an initial denaturation at 95 degrees for 10 min. Subsequently, target amplification involved 45 cycles of 5 s at 95 degrees, 40 s at 60 degrees for annealing, then extension for 20 s at 72 degrees. After amplification cycles, PCR products were evaluated for quality using melt curve analysis, which entailed 15 s at 95 degrees, 1 min at 70 degrees, 15 s at 95 degrees and 1 min at 55 degrees.

Chapter 5

Systematic description and interpretation of digital assays

5.1 Introduction

In a classic digital assay, the sample is split into a large number of compartments, which are then treated separately in order to obtain the signals from end-point readout. Then, the number of target molecules, or the initial concentration of the analyte, is calculated from the fraction of positive signals, which typically indicate a presence of at least one target molecule in the inspected partition.

5.1.1 Applications

Digital assays are widely recognized as a promising method for the quantitative assessment of low DNA concentrations because they represent a number of attractive characteristics [125–131]. First of all, they do not need any calibration with reference standards for the assessment because it is intrinsically absolute. Moreover, they provide excellent precision and high sensitivity (from 1 molecule present in the sample). This made the digital assays a popular tool in diagnostic techniques, including the oncology and any tests where the quantitative detection of extremely low quantities of target molecules (i.e. DNA markers) is needed [125], or the Copy Number Variation tests [132–135]. They are also expected to find use in diagnostic methods in the Point-of-Care (PoC) format, especially for the identification and quantitation of viral and microbial pathogens. In the future, the spread of new chain and avalanche reactions used for the identification of the

presence of very minute quantities of the molecules of analyte may boost the popularization of digital techniques in chemistry, biochemistry, and medicine [136].

5.1.2 Analytical procedure

The aim of the analytical procedure that benefits from digital methods is the assessment of the concentration C of the analyte in the source of the sample (e.q. human body). First, a sample of volume V_S is drawn, which contains M_S molecules of the analyte. The number M_S has a stochastic nature – it is a random variable with Poisson distribution with expected value $\lambda = C \cdot V_S$, i.e. $M_S \propto Pois(CV_S) = (e^{-CV_S} \cdot (CV_S)^{M_S})/M_S!$. Therefore, the local concentration in the sample $C_S = M_S/V_S$ is also a random variable of the global concentration C and of V_S: $C_S = M_S/V_S \propto 1/V_S \cdot Pois(CV_S) = 1/V_S \cdot (e^{-CV_S} \cdot (CV_S)^{M_S}/M_S!$. After the sample is collected, it is purified and the PCR reagents are added in order to run PCR. Therefore, the final, total volume of the assay equals $V_A = V_{elution} + V_{PCR}$. This volume is then partitioned into a usually large number N of identical compartments with volume $v = V_A/N$ that are separately treated for PCR. At the end, the outcome of an assay is collected, i.e. the number K of partitions that yielded positive signals is count and then used to assess the initial local concentration of the analyte in the assay $E(C_A)$ with precision $\sigma(C_A)$ determined as the relative standard deviation of the estimate $E(C_A)$.

5.2 Mathematical routines

5.2.1 Mathematical procedures

As we described earlier in chapter 4, there are two mathematical procedures that are used for the quantitative analysis of the outcome of a digital assay. They are based on the Most Probable Number (MPN) [5–7, 109–112] or Bayes' method [115–117]. Either of them can be applied both for the estimation of the concentration in the source C and the concentration in the sample C_S (the number of molecules in the sample M).

Here and in the following chapters, we base the analysis on the Bayes' formalism is used for the analysis, unless it is stated otherwise. In the Bayes' methods, the information provided by the outcome of an assay is used directly for the analysis. However, the conclusions raised here remain valid for the MPN method of analysis.

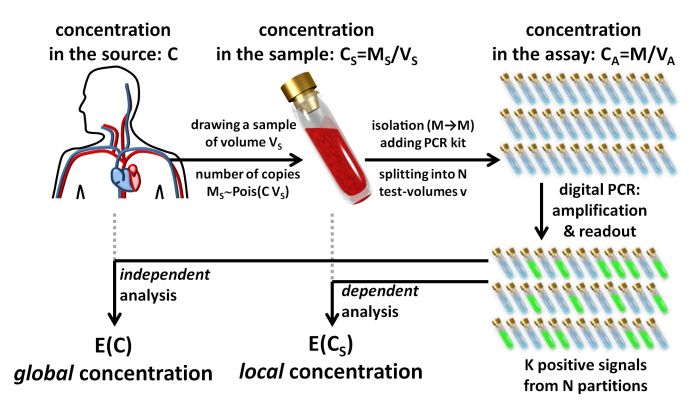


Figure 5.1: The protocol of running a digital assay. It consists of a medical and analytical procedure. The upper section shows the medical procedure of drawing the sample and preparing the PCR reactions. In the source of the sample (e.g. a human body), there is an unknown concentration C of the analyte, which is also called the *global* concentration. Then, a small sample (compared to the source) having a volume V_S is drawn from the source. It contains M_S molecules. The *local* concentration of target in the sample $C_S = M_S/V_S$ is a stochastic variable of C. After the collection of the sample, the nucleic acids are isolated and the final eluate, which contains $M \leq M_S$ target molecules, is mixed with PCR reagents. The final volume of the mixture equals the volume of the assay V_A. Then, the volume of the PCR-ready mixture is partitioned into N partitions. The number of target molecules is stochastically distributed between the compartments. After the treatment for PCR amplification, the raw result of an assay is the yield in the form of K positive signals from N compartments. The raw result is then used for the mathematical analysis shown in bottom section of the figure. It provides the estimate $E_D(C_A)$ which corresponds to the estimate $E(C_S)$ of local concentration within a *dependent* scheme, or the estimate $E_I(C_A)$ which corresponds to the estimate $E(C)$ of the global concentration of target within an *independent*. The mentioned estimated have different confidence intervals.

5.2.2 Concentration of the analyte in the sample and in the source of the sample

There are two approaches to the analysis of the outcome of an assay that will be discussed in this chapter: *dependent* and *independent* approach.

Dependent approach

In the first approach, that will be called *dependent*, the number of particles found in one compartment depends on the content of all the other compartments of the assay. The total num-

ber of molecules M in the assay volume is fixed and distributed among the compartments. This approach provides the assessment of the local concentration of the analyte in the sample $E(C_S) = \alpha E_D(C_A)$ with precision $\sigma(C_S) = \sigma_D(C_A)$. The numerical constant α reflects the change of volume (from the volume of the sample to the volume of the assay) and the efficiency of isolation of DNA $\eta = M/M_S$: $\alpha = (V_A/V_S)/\eta$.

The estimate provided by the *dependent* approach is limited to the local concentration, or the number of target molecules in the sample, the volume of which is small compared to the volume of the source (human body). Therefore, the value of precision from the *dependent* approach, which is most often used by the producers of digital PCR systems, might be misleading if the global concentration (i.e. concentration in the human body) is of interest.

Independent approach

In the other approach, called (*independent*), the number of particles found in one compartment is independent of the content of other compartments, as it was taken directly from the source. Therefore, the total number of molecules M in the assay volume is not fixed but is a random Poisson variable with expected value $\lambda = C \cdot V_S$.

This approach provides an estimate of the global concentration (i.e. concentration in the source): $E(C) = \alpha E_I(C_A)$, and $\sigma(C) = \sigma_I(C_A)$, therefore in reflects the actual situation in the human body.

In this chapter, we discuss of how the two approaches to the analysis of the outcome of a digital assay are used and the difference between the relative standard deviations σ_D and σ_I.

The difference between the two approaches (i.e. between the assessment of the concentration in the sample and in the source) is revealed in the statistical character of a single partition of an assay, which is represented as a binary random variable. In this chapter, we give the interpretation of the outcome of digital assays in two approaches, which we will use later (in the next chapter) to design assays that provide the requested estimate.

The analysis in the *dependent* approach focuses on counting the number of target molecules in the sample (local concentration). The only stochasticity can be found in the location of each of the M target molecules distributed among the compartments. As we mentioned, a fixed number of target molecules forces the dependent probability of the occupancy of a single compartment. It can be well visualized for an extremely low number of molecules in the sample. For example,

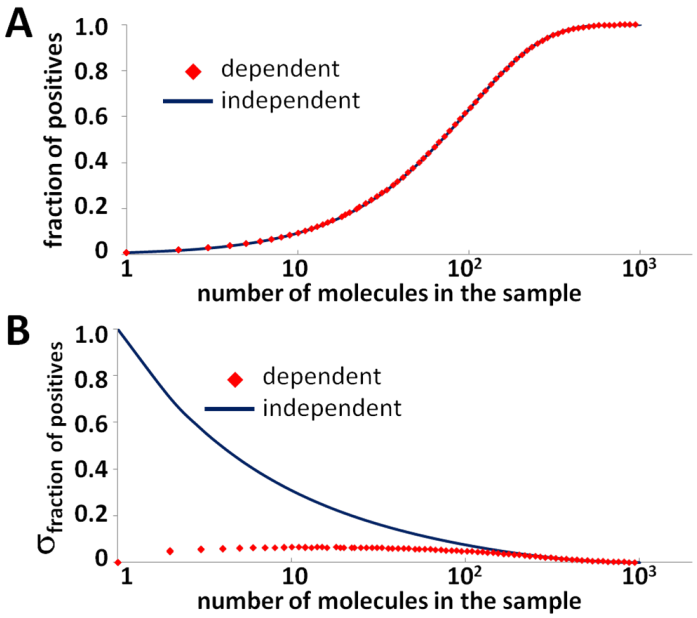


Figure 5.2: (a) The response of a digital assays comprising $N = 100$ identical partitions to a given *global* concentration (or *local* concentration in case of *independent* scheme). The response in the *independent* approach (blue line) was calculated analytically, while the response in the *dependent* (red points) represents an average from 10,000 Monte Carlo simulations. (b) Unlike the average response of an assay, the spread of the results depends on the statistical formalism used for the analysis. In case of low global concentrations, which correspond to a single, or a few target molecules in the sample, the spread reaches 100% of the number of positive signals (for *independent* analysis). In contrast, the spread of the number of positive signals in case of *dependent* analysis converges to zero for low *local* concentration.

if only one target molecule is present in the whole sample (i.e. $C_A = \frac{1}{V_A}$), and it is distributed between the compartments, and one of the compartments happens to contain the molecule, all the others have to remain empty (the probability of finding a target molecule there is exactly zero). However, the knowledge on the global concentration is limited.

The second approach focuses on the concentration of the analyte in the source of the sample. Here, we assume that each compartment is drawn independently from a large reservoir (i.e. source of the sample; patient body) containing a global concentration C_A of target molecules. Hence, the probability of finding particles in a compartment of volume v is independent of the presence of particles in any of the other compartments and is only a function of C_A: $p = 1 - e^{-C_A v}$. For the global concentration equal to one partition per assay $C_A = 1/V_A$, the assay may yield more than one positive signal (on the other hand, the result of no positive signals also cannot be excluded) as well. Therefore, the spread of the outcomes of the assay is wider compared

to *dependent* approach, and the estimate of the concentration has a higher relative standard deviation. This corresponds to the concentration of target molecules in the source C.

5.2.3 Analytical description of the outcome of a digital assay

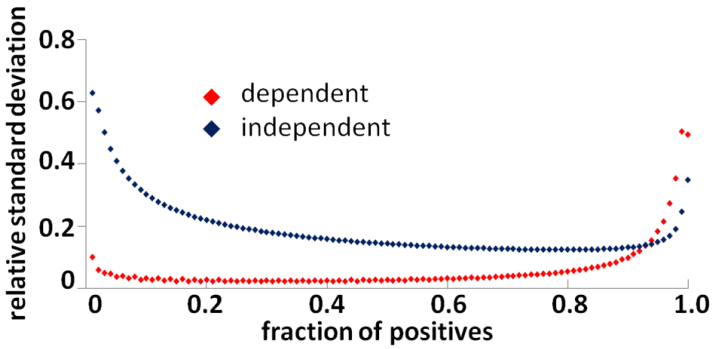


Figure 5.3: The comparison of the standard deviation of the estimate of concentration of the analyte for the two approaches. The *independent* analysis provides the estimate of the global concentration $E_I(C_A)$ with standard deviation $\sigma_I(C_A)$ (blue points), which is the same as the standard deviation $\sigma(C)$ of the estimate $E(C)$ of the concentration in the source. It is significantly higher than the standard deviation $\sigma_D(C_A)$ of the estimate of *local* concentration $E_D(C_A)$ (red points), which equals the standard deviation $\sigma(C_S)$ of the estimate $E(C_S)$ of concentration in the sample.

We also decribed quantitatively the difference between the two approaches. As we noticed earlier, the spread of the outcomes yielded by an assay is bigger for the *independent* approach than for the *dependent* approach. The distribution of the number of positive signals in the *dependent* scheme (i.e. the probability of observing exactly K positive signals from a total of N compartments if there are $M = \alpha \cdot C_S \cdot V_S = C_A \cdot V_A$ target molecules) is given by a combinatorial equation:

$$P(K|M,N) = \frac{\binom{N}{K}\sum_{i=0}^{K-1}((-1)^i\binom{K}{K-i}(K-i)^M)}{N^M}.$$

For the *independent* scheme, the distribution of the outcomes (given the concentration C_A) is representad by:

$$p(K|C_A,N) = \binom{N}{K}(1-e^{-C_A v})^K(e^{-C_A v})^{N-K}.$$

The latter distribution has a bigger variance, and the difference is particularly evident for small values of C_A. As a result, for the *independent* approach we observe a much bigger relative standard deviation of the estimate of $E_I(C_A)$ and $E(C)$, compared to the standard deviation of the estimate of $E_I(C_A)$ and $E(C)$ provided by the *dependent* approach (Fig. 5.3).

5.2.4 Translation from dependent to independent analysis - convolution

It is possible to translate between the two estimates. If the estimate of the local concentration (*dependent* approach) is available, and the volume of the sample is known, we can calculate the estimate of the global concentration, with a higher standard deviation.

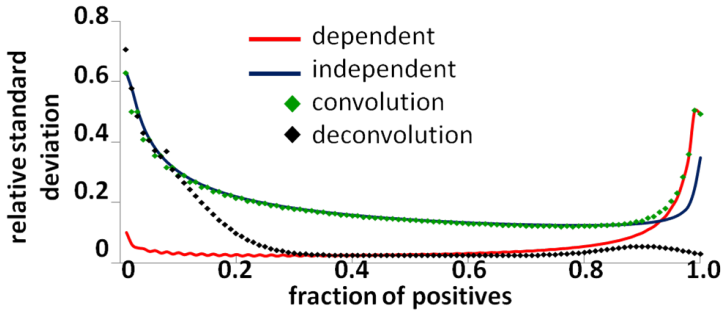


Figure 5.4: The translation between the *dependent* (local concentration) and *independent* (global concentration) analysis calculated by means of mathematical procedure of convolution.

The translation between the estimates, i.e. between the distribution of the local concentration in the sample provided by the dependent *approach* into the distribution of global concentration in the source provided by the *independent* approach can be calculated by means of the procedure of convolution. In short, the distribution of the possible outcomes K from an assay containing M target molecules from *dependent* analysis has to be convoluted with the distribution of finding M molecules in the volume V of the assay for a given global concentration C.

Formally, this can be written as the following summation:

$$p(K|C) = \sum_{M=0}^{\infty} p(K|M) \cdot p(M|C),$$

where $p(K|M)$ is the distribution of the outcomes provided by the *dependent* analysis, and $p(M|C)$ is the distribution of the number of molecules M, usually given by the Poisson distribution with expected value $\lambda = CV_S$. It can be calculated that the estimate derived in this analysis is equal to the outcome provided by the *independent* analysis (Fig. 5.4, green points, and blue line, respectively).

Mathematically, the convolution (marked with the symbol *) requires the following integral transformation of two functions (distributions), while one of the functions is shifted:

$$f(x) = (g * h)(x) = \int_0^{\infty} g(\tau)h(x - \tau)d\tau.$$

For the case of single-volume digital assays we discuss in this chapter, the distribution f represents the probability of the outcome K of positive partitions from an assay given the con-

centration C: $(K|C)$ (*independent* approach), the distribution g represents the probability of observing K positive partitions in the assay given the number of target molecules in the sample M: $P(K|M,N)$ (*dependent* approach), and distribution h represents Poisson distribution $p(M|C) = Pois(M, CV_S)$. It is worth mentioning that the number of molecules M can adopt only positive integer values. Finally, the translation between the estimates is given by:

$$p(K|C) = \sum_{\tau=0}^{\infty} p(\tau|C)p(K|\tau).$$

Here, the symbol of the integral was replaced by the summation because the variables M and τ adopt only integer values.

5.2.5 Translation from independent to dependent analysis - deconvolution

There is also a possibility to translate the concentration in the source (*independent* analysis) into the concentration in the sample (*dependent*). Provided the distribution $p(M|C)$ (i.e. probability that there are M molecules in the sample provided concentration in the source is C) is known, one can recalculate the distribution $p(K|C)$ to $p(K|M)$.

Formally, the translation is done by means of the mathematical procedure of deconvolution. In the previous section, the procedure of convolution from *dependent* to *independent* scheme was shown; here the reverse process is discussed.

The probability $P(s_i|C)$ of obtaining a recorded state of an assay (i.e. the count K of positive signals from a total number N of compartments available: $K = \sum_{i=0}^{N-1} s_i$) given the global concentration C in the *independent* also reflects the probability that M molecules of the analyte were present in the sample at the said global concentration C. In case of stable and uniform solution of target molecules, the latter probability is described by the Poisson distribution with parameter (expected value) $\lambda = CV_S$.

However, the procedure described here does not reproduce the distribution $p(K|M)$ ideally. The difference stems from the fact that the Poissonian distribution $Pois(M|CV)$ and $P(K|M)$ have a non-zero covariance, especially for a small fraction of positive compartments (low M). Hence, the *dependent* analysis should be used if the local concentration is in question (Fig. 5.4, red line).

5.3 Summary

In this chapter we decribed the two approaches to the analysis of the result of a digital PCR assay that leads to the assessment of the concentration of the analyte either in the sample or in the source of the sample. The difference between the approaches is manifested by the character of random variables used for the statistical description of signals. This reflects the qualitative difference between the quantities assessed, i.e. the local and global concentration of the analyte.

This analysis is the starting point for the optimisation of single-volume digital assays for the assessment of the requested dynamic range and providing a requested precision, as well as the design of multi-volume digital assays described in the following chapters.

Chapter 6

Optimized design of single-volume and multi-volume digital assays

6.1 Introduction

The discussion of the response of an assay, i.e. the function $p(K|C)$ describing the response of a digital assay to global concentration and $p(K|M)$ that concerns the local concentration, that we presented in the previous chapter, allows the correct interpretation of the assessment by means of a digital assay.

In this chapter, we will enhance this analysis with the analytical expressions describing the behavior of digital assays, the precision of the assessment in particular. These equations allow designing the digital assays that assess the concentration of the analyte within the requested dynamic range with the required precision.

This subsequently leads to designing multivolume digital assays. Knowing the relation between the dynamic range and the precision of the single volume assay comprising N compartments, we can arrange a set of classic digital assays in a sequence differing in the volume of their compartments. This allows to:

- increase the dynamic range of a classic digital assay, and
- tune the multivolume assay for the desired precision and dynamic range.

The multivolume digital assays were shown earlier by Ismagilov et al. [5–7] and Chiu et al. [137],

although the mathematical procedures that led to presented designs were not discussed. Here we demonstrate explicitely the procedure and show that it leads to optimum designs - i.e. designs that offer uniform precision across the whole dynamic range and require minimum number of compartments.

6.2 Single-volume digital assays - new analytical solutions

6.2.1 Estimate of the initial concentration

Using the equations for the assay's response from the previous chapter, we can estimate the initial global concentration or the number of molecules target molecules in the sample, using Bayes' theorem [115] or Most Probable Number Method [5–7, 109–111]. The latter provides analytical formula for the most probable concentration $E(C)$ that caused a given outcome of an assay.

The analysis is done as following:

1. From every positive partition (i.e. $s_i = 1$) with volume v, a probability function p_i is constructed: $p_i = (s_i = 1|C) = 1 - e^{-Cv}$

2. Also, the probabilities of obtaining negative signals for all negative compartments are constructed: $p_i = (s_i = 0|C) = e^{-Cv}$

3. Then, the probability $P(s_i|C)$ of obtaining the recorded state of a digital assay (e.g. the recorded number of positive compartments $K = \sum_{i=0}^{N-1} s_i$ out of all N compartments constituting and assay) is calculated, which is a product of the probability functions for all the compartments: $P(s_i|C) = (1 - e^{-Cv})^K \cdot (e^{-Cv})^{N-K}$

4. Then, we find the value of concentration in the assay C, for which the probability $P(s_i|C)$ has the largest value. To do that, we have to solve the equation: $dP(s_i|C)/dC = 0$.

5. The solution of this equation is the value of C for which the derivative of probability $(s_i|C_A)$ equals zero, is:

$$E(C) = v^{-1} ln(N/(N - K)).$$

As we showed in the previous chapter, the most probable outcome of an assay K is the same in the *independent* – $K(C)$ and *dependent* – $K(M = C \cdot V)$ approach (Fig. 5.2a). Therefore, the

estimate of the number of molecules in the sample $E(M)$ averages to a similar logarithmic form:

$$E(M) = Nln(N/(N-K)).$$

The above equations are similar despite the different statistical interpretation of signals because here one operates only on the most probable outcomes of the assay and most probable initial concentration (or a number of molecules). Therefore, the stochastic nature of the number M of target molecules is not taken into account.

6.2.2 Precision of the estimate - new analytical formula

The derivation of formulas for the precision of the estimates is not as straightforward, as the estimate of initial concentration, and therefore it was usually calculated numerically or by means of Monte Carlo simulations. Still, it is obvious that the precision should scale with $1/\sqrt{N}$, because each of the compartments is, in fact, an independent test. Also, the lower bound for the standard deviation is determined by the Fisher Information and Cramer-Rao inequality [138–143]: $\sigma(C) \geq \frac{\sqrt{K}}{\sqrt{(N-K)(N+K(N-K))}} \cdot \frac{1}{E(C)}$. The final analytical expression for the precision of the estimate is as following:

$$\sigma(C) = \frac{\sqrt{K}}{\sqrt{N(N-K)}} \cdot \frac{1}{E(C)}.$$

The variance in the independent scheme is, in the first approximation, the sum of the variance in the dependent scheme and variance of Poisson distribution $Pois(\lambda = M/N = ln(N/(N-K)))$, which equals $\lambda = M/N = ln(N/(N-K))$ (still, form small values of C, it is not true because the covariance matrix $P(M|K)$ and Poisson distribution is non-zero). Therefore, the expression for the precision of the estimate, in the dependent scheme, can be approximated with:

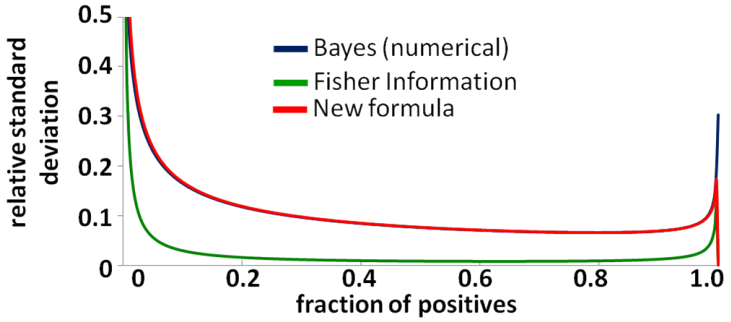
$$\sigma(M) = \sqrt{\sigma^2(C) - 1/M} = \sqrt{\sigma^2(C) - (Nln(N/(N-K)))^{-1}}.$$

We can transform these equations into the expressions that bind the precision with the dynamic range within which this precision is provided. In order to do that, they have to be represented in terms of C and $M = CV$. Therefore, for the *independent* approach, we get the expression for precision:

$$\sigma(C) = \frac{\sqrt{e^{Cv} - 1}}{\sqrt{N}} \cdot \frac{1}{Cv}.$$

For small values of initial concentration, i.e. $Cv \ll 1$, this can be approximated with

$$\sigma(C) \approx \frac{\sqrt{Cv + \frac{(Cv)^2}{2} + \frac{(Cv)^3}{6}}}{\sqrt{N}Cv},$$



$$\sigma(C) = \frac{\sqrt{K}}{\sqrt{(N-K)N} \cdot ln(N/(N-K))}$$

Figure 6.1: Numerical verification of the equations for a single-volume digital assay in the *independent* scheme. The precision of the estimate as the function of initial concentration calculated analytically (red line) and compared to numerical results (blue line) and the value calculated by means of the Fisher Information and Cramer-Rao inequality (green line). The analytical formulas ideally match the numerical results. The value of the standard deviation of the estimate determined by means of the Information Theory resembles the same qualitative character, although quantitatively it is lower in the whole dynamic range; it represents the lower (best) limit of precision for any achievable estimate.

and for high values of initial concentration, i.e. $Cv \geq 2$, it can be approximated with:

$$\sigma(C) \approx \frac{\sqrt{e^{Cv}}}{\sqrt{N}Cv}.$$

These two functions are inversible, therefore one can retrieve the value of C^- being the lower limit of the dynamic range:

$$C^- = (6\sigma^2 N - \sqrt{3}\sqrt{1\sigma^4 N^2 - 12\sigma^2 N - 5} - 3)),$$

and C^+ being the upper limit of the dynamic range:

$$C^+ = -4W_{-1}(\frac{-1}{2\sigma\sqrt{N}}),$$

where W_{-1} is the analytic continuation of product log (Lambert W) function.

Finally, we obtain the expression for the dependence between the dynamic range Ω and precision σ:

$$\Omega(\sigma) = \frac{C^+}{C^-} = -4W_{-1}(\frac{-1}{2\sigma\sqrt{N}})/((6\sigma^2 N - \sqrt{3}\sqrt{1\sigma^4 N^2 - 12\sigma^2 N - 5} - 3)).$$

For *dependent* approach $\sigma(M = CV)$, the lower bound is fixed in first approximation and equals $C^- = 1/N$, while the upper bound C^+ equals, analogically for *independent* approach:

$$C^+ = -4W_{-1}(\frac{-1}{2\sigma\sqrt{N}}).$$

Finally, for the *dependent* approach, one gets:

$$\Omega(\sigma) = \frac{C^+}{C^-} = -4NW_{-1}(\frac{-1}{2\sigma\sqrt{N}}).$$

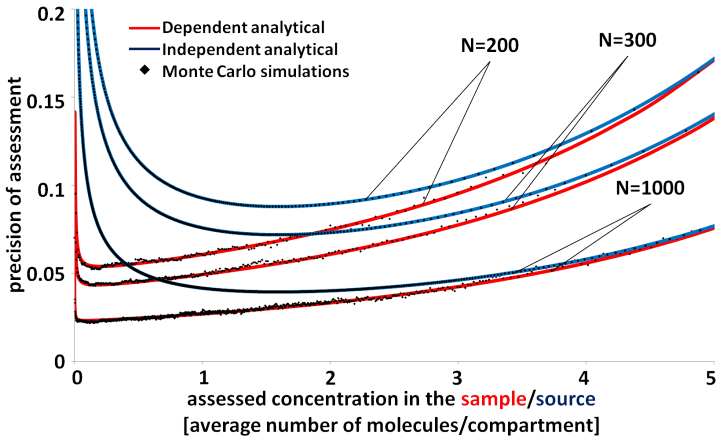


Figure 6.2: Numerical verification of the equations for a single-volume digital assay in the *dependent* (red) and the *independent* (blue) schemes. The precision of the estimate as the function of initial concentration calculated analytically (red line for *dependent* and blue line for *independent*) and compared to numerical results (black dots). The analytical formulas almost ideally match the numerical results for a wide range of digital assays.

6.3 Multi-volume digital assays

6.3.1 Derivation of the algoritms

We can use the newly derived expressions that bind the dynamic range with precision and give us the full information about the behavior of single-volume digital assays as a basis for the design of multi-volume digital assays. However, what remains to be determined, are the conditions under which the assays work in their *optimum mode*. Here, we determine the so-called *optimum mode* as the best compromise between the value of precision of the assessment and dynamic range for which the precision is provided by an assay. In this section we give a reasonable candidate for a definition of such compromise.

The derivation of the design of an assay design starts with introducing the *optimum dynamic range*. Even after a superficial analysis of the functions determining the standard deviation of the estimates shown in Fig. 6.3a, it can be noticed that despite the obvious differences, the both

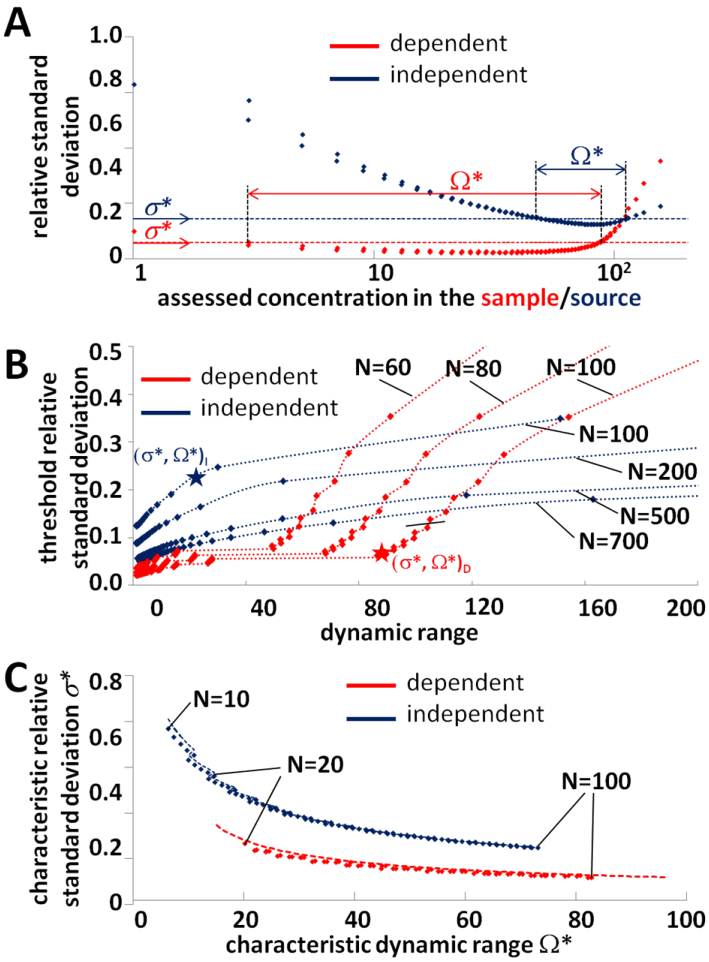


Figure 6.3: (a) The geometrical interpretation of the dependence between the precision and the dynamic range both for *independent* approach (blue dots, global concentration) and *dependent* approach (red dots, local concentration). The threshold values for the relative standard deviation of the estimate ($\sigma_{threshold}$), which is unambiguously paired with the dynamic range $\Omega = C^+/C^-$, where C^- and C^+ determine the lower and upper limits of concentration range where the precision of the assessment $\sigma_{threshold}$) (or better) is provided. (b) The function $\Omega(\sigma_{threshold})$ can be determined for any digital assay. However, it is vital to find a reasonable trade-off between the wide dynamic range and acceptable precision of the assessment. Such a trade-off can be determined by finding a deflection point(Ω^*, σ^*) (marked with stars) of the function $\Omega(\sigma_{threshold})$. (c) The characteristic values $\Omega^*(\sigma^*)$ of an assay shown as a function of the size of an assay (i.e. the number N of compartments). One can determine the analytical dependence (dashed lines) from the numerical results (blue line represents the formula $\Omega^*(\sigma^*) = 0.9925\cot(\sigma^*)^{-2.065}$ and red represents the formula $\Omega^*(\sigma^*) = 0.1697 \cdot (\sigma^*)^{-2.214}$

.

functions have a relative plateau near the minimum value. Then, one can observe a sharp increase of the value of the relative standard deviation in the region of low (*independent* approach) and

high concentrations (both *dependent* and *independent* approach). This plateau determines the spread of the values of concentration for which the assay works in the optimum mode.

Generally, an optimized design of a digital assay provides the assessment in a dynamic range $\Omega^* = C^+/C^-$ (where C^- and C^+ are the lower and upper bounds of the dynamic range) a standard deviation of the estimate smaller or equal to σ^*. As it can be seen in Fig. 6.3b, a wider dynamic translates to worse precision of the assessment and *vice versa*. A candidate for a definition of the optimum dynamic range could be the spread of the deflection points $d^2\Omega/d\sigma^2 = 0 \rightarrow (\Omega^*, \sigma^*)$ (however, other candidates can be given, as we know the functions $\Omega(\sigma)$). For $\sigma > \sigma^*$, the increase of the dynamic range is not justified by the decrease of precision of the estimate. Also, other definitions of the dynamic range, based on the Shannon entropy (Section 6.3.4) or other measures of information should provide qualitatively similar results.

Choosing any unambiguous definition makes it possible to bind Ω^* with σ^* for any size of an assay N, as it is shown in Fig. 6.3c.

Therefore, the characteristics of the assay can be tuned only by changing the number of compartments. To verify this, we tested several digital assays with the new formulas and Monte Carlo simulations. Indeed, we found a simple dependence between the precision σ of the estimate and the size of the said assay N. From a simple algebraic fit to the numerical results, we established the relation $N \geq a\sigma^{-b}$ (Fig. 6.3).

As we mentioned earlier, digital assays behave similarly in *dependent* and *independent* approach. Still, there is a quantitative difference coming from the fact that for the same number of compartments, the estimate of local concentration is more precise than the estimate of global concentration. The same numerical data can be used to find the dependence between the characteristic parameters of a digital assay (Ω^*, σ^*), determined according to the above-mentioned definition of an optimum mode of working.

For the assessment of the global concentration (i.e. *independent* scheme), one gets

$$\Omega^*(\sigma^*) = 0.9925 \cdot (\sigma^*)^{-2.065},$$

while for the local concentration (i.e. *dependent* scheme) it is equal to:

$$\Omega^*(\sigma^*) = 0.1697 \cdot (\sigma^*)^{-2.214}.$$

Finding the relationship between the dynamic range and the standard deviation of the estimate is a basis for the derivation of an algorithm that allows the tuning of the number of compartments

N required to provide estimate of the concentration within a requested dynamic range with a requested precision.

The analytical equations given below use as an explicit input the dynamic range $\Omega = C^+/C^-$ and the relative standard deviation of the estimate σ that the assay should provide. In case of *independent* approach, both values of C^- and C^+ can be freely chosen (the only limitation is that $C^+ > C^-$), and the design equations are as following:

$$N = max(N_\Omega; N_\sigma),$$

$$N_\Omega = 1.447\Omega^{0.985},$$

$$N_\sigma = 1.436\sigma^{-2.033}.$$

For *dependent* scheme, the lower bound of the dynamic range C^- equals one molecule per assay (i.e. $C^- = 1/V_A$), while C^+ is equal to the maximum number of molecules M_{max} that can be found in the sample. Then, the design is given by:

$$N = max(N_\Omega; N_\sigma),$$

$$N_\Omega = 2.154\Omega^{0.857},$$

$$N_\sigma = 0.454\sigma^{-1.912}.$$

N_Ω and N_σ determine the minimum size of an assay to cover the requested dynamic range or to provide the required precision, respectively. In order to prepare a digital assay that provides the requested assessment, the sample should be partitioned into at least the highest of the two number of compartments. However, some digital PCR systems [144, 145] have a hardwired number of compartments the sample is partitioned into. In such case, the equations provide the value of the precision and the dynamic range the system provides in case of *dependent* and *independent* analysis (Section 6.3.3).

6.3.2 The outcome of the assay reaching its *optimum mode.*

It is worth noticing that in the case of *independent* analysis, the assay reaches the optimum mode (i.e. the precision reaches a plateau - Fig. 6.3a) while 20% of compartments yield positive signal, regardless the size of an assay (Fig. 6.4). This observation is vital for the determination of the volume of a single compartment. For a calculated number of compartments N, and the lower bound of dynamic range C^-, the volume of a single compartment equals $v = \frac{log(1.25)}{C^-}$. This determines the volume of the assay $V_A = Nv = N \cdot log(1.25)/C^-$. Here, and in the following

chapters, a simplified situation is assumed that the volume of the sample equals the volume of an assay. In practice, the volume of an assay is the sum the volume of the PCR mix of the eluate and reagents. Therefore, the volume of the sample that is physically drawn has to be recalculated concerning the purification and isolation protocol, and the volume of PCR reagents.

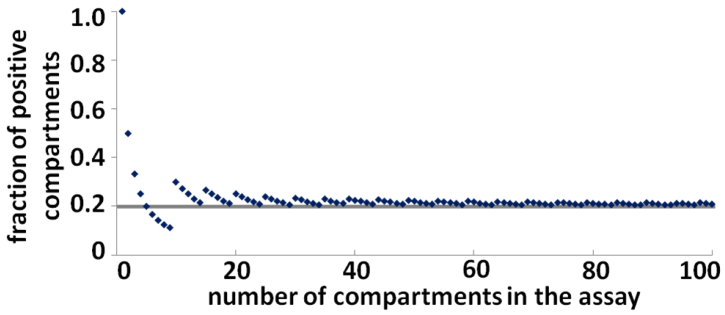


Figure 6.4: The ratio of positive compartments in the assay reaching its optimum mode (i.e. reaches the plateau of the standard deviation of the estimate) in the independent analysis. The ratio saturates at 20% (best fit), which provides the condition for the volume of a single compartment of the assay.

6.3.3 The standard deviation of the estimate and the dynamic range provided by an assay comprising a fixed number of compartments.

If the number of compartments comprising a digital assay is hard-fixed by the dPCR system, we can rearrange the equations for the design so they give the precision of the estimate (relative standard deviation σ of the estimate) and dynamic range Ω provided by this system. The digital assay comprising of N compartments provides, in the independent scheme, the relative standard deviation of the estimate to be:

$$\sigma = 1.195 \cdot N^{-0.492}$$

within the dynamic range:

$$\Omega = 0.687 \cdot N^{1.015}.$$

For the dependent scheme, the relative standard deviation of the estimate to be:

$$\sigma = 0.662 \cdot N^{-0.523}$$

within the dynamic range:

$$\Omega = 0.408 \cdot N^{1.167}.$$

6.3.4 Shannon Entropy as a measure of a digital assay's performance.

Tools provided by the Information Theory [146–149] may be useful to determine the parameters of an assay and can be used in the design procedure. For example, it is possible to determine the amount of information gained from one compartment of an assay by means of the Shannon Entropy, which is determined by

$$H_b = -\sum_i p_i log(p_i),$$

where i counts all the possible outcomes from the compartments, i.e. positive and negative signals. We can calculate that the amount of information given by a single compartment is maximum if the initial concentration is equal to $C = log(2)/v$, where v is the volume of the compartment. If the compartments are treated as independent random variables (independent approach), the Shannon Entropy simply adds up, so it can be easily calculated analytically (Fig. 6.5). The dynamic range of an assay can be determined as the range of concentrations for which the value H_b is higher than a given threshold, which is related to the requested precision of the estimate.

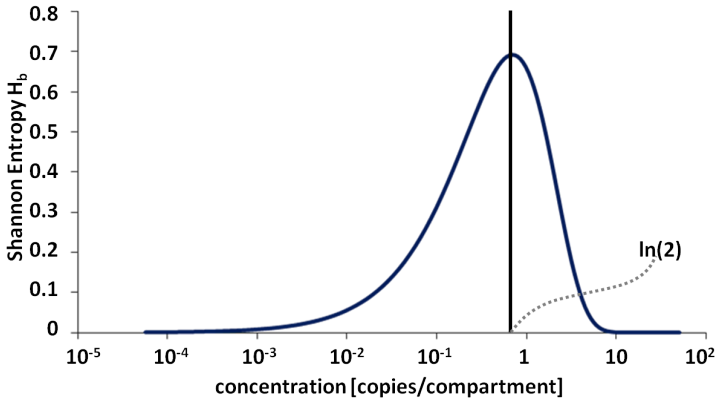


Figure 6.5: The amount of information (Shannon Entropy H_b) gained from a $N = 20,000$ compartments as a function of initial concentration. The higher the value H_b, the more informative is the outcome of an assay. The dynamic range of an assay can be determined as the range of concentrations for which the value H_b is higher than a given threshold.

Another approach is to determine the lower bound for the standard deviation of the estimate by calculating Fisher Information I [138–143]: :

$$I = E[(\partial p/\partial C)^2],$$

where p is the probability of obtaining a given outcome of an assay K:

$$p = \binom{N}{K}(1 - e^{-C_A v})^K \cdot (e^{-C_A v})^{N-K}.$$

According to the Cramer-Rao inequality, the variance of the estimate is higher or equal to $Var(C) = \sigma^2(C) > 1/I$. Therefore it is possible to determine the lower bound of the relative standard deviation of the estimate and base the design of an assay on this relation. However, the upper bound remains unknown. Still, if the number N of compartments is higher than 200, the value of variance given as a reciprocal of the Fisher Information: $Var(C) = \sigma^2(C) = 1/I$ agrees well with numerical results.

6.4 The design of multi-volume digital assays

The analysis of the behaviour of a single-volume digital assay, and the derivation of equations that bind the parameters of the provided assessment (precision and dynamic range) allow designing assays in the multivolume scheme, i.e. arranging a set of relatively small single volume digital assays, called sub-assays ($N' \approx 10^2$) that vary in volume or dilution, in a geometric sequence, to cover a wider dynamic range.

The input parameters for the design procedure are the required total dynamic range of the assay Ω_{tot} and the required standard deviation of the measurement σ_{tot} in the whole dynamic range. The equations that we list below provide the number of compartments in one sub-assay N', ratio of products of volume and dilution factor for two consecutive sub-assays x and the number of sub-assays N_{lib}. The design is performed as follows:

- The standard deviation provided by a single sub-assay σ' must be equal to the required standard deviation σ_{tot}, i.e. $\sigma' = \sigma_{tot}$. Therefore, according to the methodology presented in this chapter and the definition of optimum mode of working of a digital assay, the number of compartments in a single assay is equal to $N' = a\sigma_{tot}^{-b}$.

- Having established the number of compartments in a single assay, we have to calculate its dynamic range $\Omega' = c(N')^d$. The values a, b, c and d are constants of the order of unity, different for *independent* and *dependent* approaches. The ratio of volumes of compartments in the consecutive sub-assays depends on the dynamic range of a single sub-assay: $x = 1/\Omega'$.

- The number of sub-assays N_{lib} depends on the ratio of the required total dynamic range of the assay and the dynamic range provided by a single sub-assay: $N_{lib} = log(\Omega_{tot})/log(\alpha\sigma_{tot}^{-\beta})$.

Such constructed multivolume assay provides standard deviation of the measurement σ_{tot} in the whole dynamic range Ω_{tot}. It comprises N_{lib} single volume sub-assays, each comprising N'

compartments, with the ratio of volumes of consecutive compartments from two consecutive sub-assays x. The values of N_{lib}, N' and x are given by the following formulas:

$$N_{lib} = log(\Omega_{tot})/log(\alpha\sigma_{tot}^{-\beta})$$

$$N' = \gamma\sigma_{tot}^{-\delta}$$

$$x = \Omega_{tot}^{1/N_{lib}}$$

Where α, β, γ and δ are constants: $\alpha - 0.6813$, $\beta = 2.0966$, $\gamma = 0.9925$ and $\delta = 2.065$.

6.5 Performance

6.5.1 Performance of single-volume digital assays

We verified numerically the performance of single-volume assays designed using the equations we listed in the previous sections by means of canonical (*dependent* approach) and grand canonical (*independent* approach) Monte Carlo simulations. We designed digital assays to provide the assessment of the concentration within the dynamic range (*i*) $\Omega_1 = 10^2$, (*ii*) $\Omega_2 = 5 \cdot 10^2$ and (*iii*) $\Omega_3 = 10^3$. This required the partitioning into $N_1 = 135$, $N_2 = 660$ and $N_3 = 1305$ compartments in *independent* scheme and $N_1 = 112$, $N_2 = 443$ and $N_3 = 802$ compartments in *dependent* scheme (Fig. 6.6).

6.5.2 Performance of multi-volume digital assays

The assays that we present here were tested numerically using canonical and grand canonical Monte Carlo simulations. The equations we presented in this chapter were used to design digital assays that comprise (Fig. 6.7):

- $N_{total} = 96$ compartments in 6 sub-assays, $x = 0.087$, that cover dynamic range $\Omega = 10^6$ with precision $\sigma = 30\%$, and
- $N_{total} = 96$ compartments in 2 sub-assays, x=0.027, $\Omega = 10^3$, $\sigma = 20\%$.

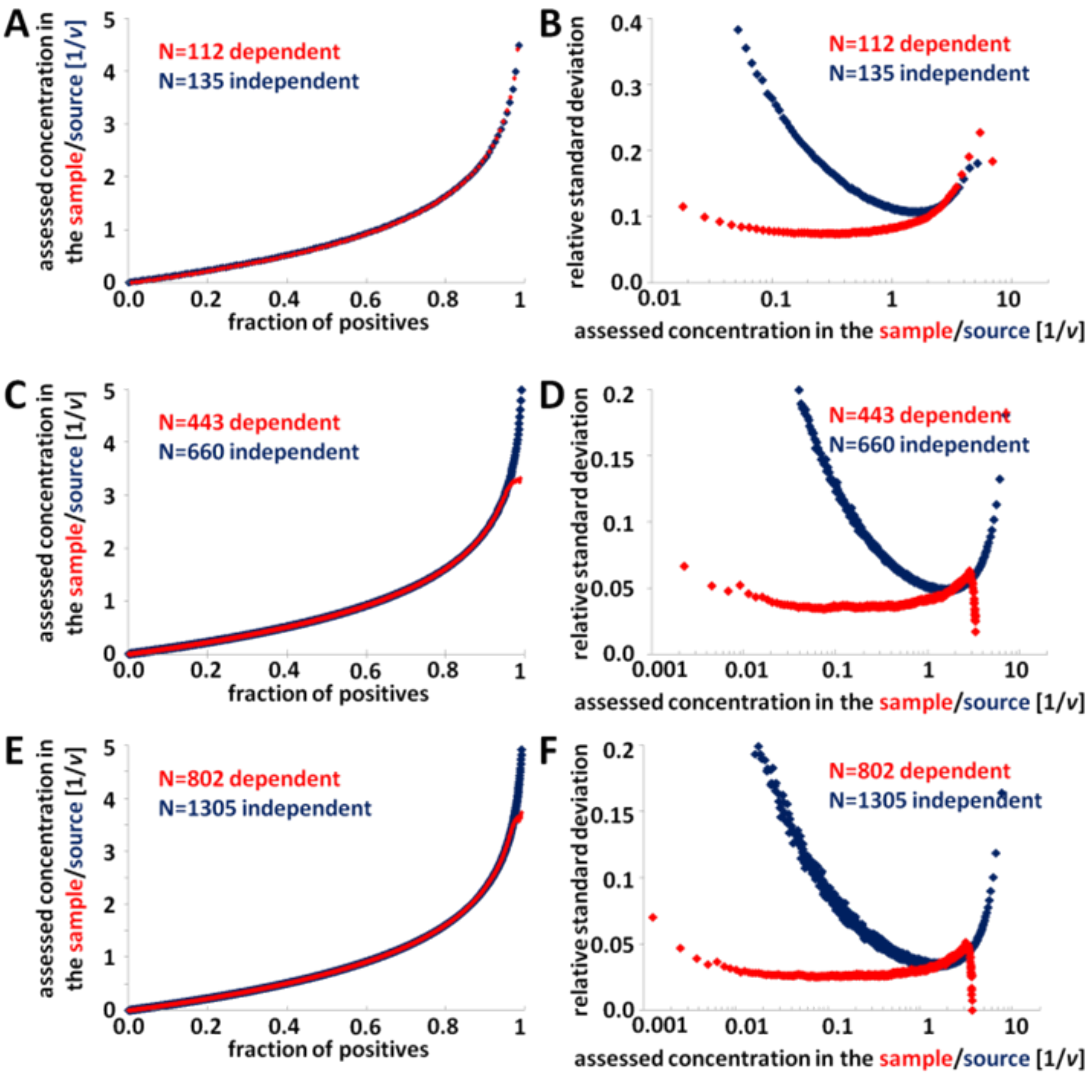


Figure 6.6: The numerical verification of the performance of digital assays in case of *dependent* (red) and *independent* (blue) by means of Monte Carlo simulations. The assays are designed to provide assessment within the same dynamic range (a, b - assays covering $\Omega = 10^2$; c, d - assays covering $\Omega = 5 \cdot 10^2$; and e,f - assays covering $\Omega = 10^3$). (a, c, e) The estimate of the local concentration in the sample, which equals the number of molecules in the sample per the volume of the sample (*dependent* scheme), and the global concentration in the source (*independent* scheme) are gievn as the function of the fraction of positive compartments from an assay. The value of concentration is given in$[1/v]$, where v is the volume of a single compartment $v = V_S/N$. (b, d, f) The precision of the estimate of concentration assessed by digital assays. The standard deviation of the estimate of the *global* concentration in the source is always higher or equal to the standard deviation of the estimate of the *local* concentration the in the sample because of the randomness of the number of molecules actually present in the sample during drawing it from the source.

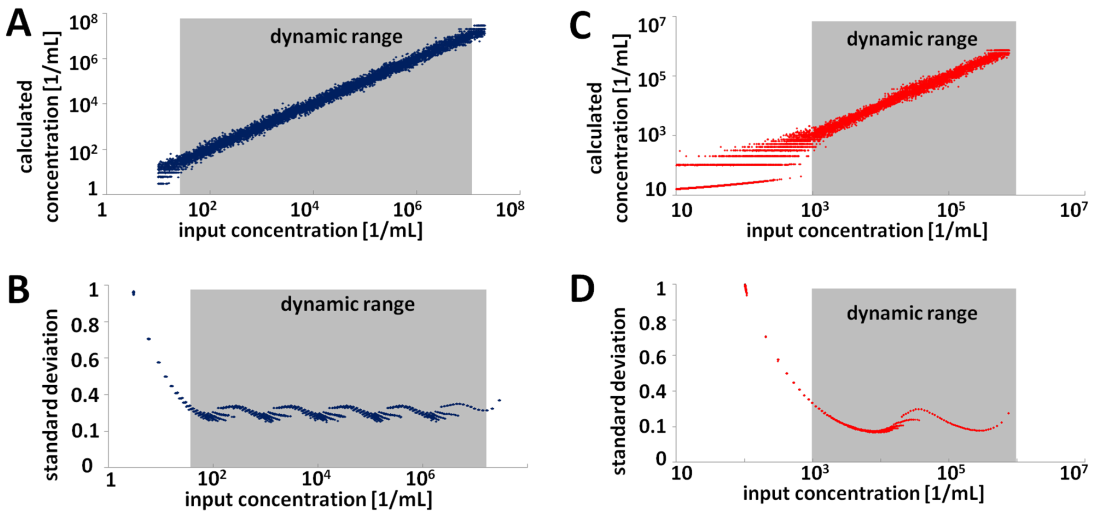


Figure 6.7: Numerical verification of a multivolume digital assay. (a, c) The estimated concentration of DNA as a function of known input concentration in the source C. Data points (blue - multivolume digital assay comprising 96 compartments in 6 sub-assays, covering the dynamic range $\Omega = 10^6$ with requested precision 30%, red - multivolume digital assay comprising 96 compartments in 2 sub-assays, covering the dynamic range $\Omega = 10^3$ with requested precision 20%) represent 10 000 grand canonical Monte Carlo simulations. (b, d) Standard deviation of the estimates within and outside the requested dynamic range The gray areas mark the requested dynamic range of both assays.

6.6 Summary

The proper understanding of the behavior of digital assays allowed the derivation of new analytical formulas for designing analytical tests that provide the assessment of the concentration with the requested precision within the requested dynamic range and with. We can use the mathematical formulas for the design were obtained both for the analysis of the local (*dependent* scheme) and global concentration (*independent* scheme) to tailor digital tests for various applications.

Knowing the response of digital assays allows characterizing any assay by its characteristic parameters: dynamic range and precision. This subsequently leads to the optimization of the assessment by combining many classic, single-volume assays in a multi-volume scheme.

The main advantages of the multi-volume scheme is the ability of simple widening of the dynamic range, thanks to the logarithmic scaling of the number of compartments, and the possibility of tuning the assay for specific user's requirements (given requested precision and dynamic range of the assessment), one can design a multi-volume assay that comprises much fewer compartments than a classic single-volume assay. The reduction of a number of compartments lowers

technical requirements to run digital assays. Therefore, the design of the assay described here can be treated as an alternative to state-of-art techniques that require extremely high numbers of compartments.

The multivolume design that we presented here can be treated as a development of classic single volume digital assays. In the design we described, the sets are large (usually tens or hundreds of compartments), but require only coarse gradation, and therefore are reasonably easy to perform. They usually comprise more compartments, but are easier to prepare and could be recommended for very precise assessments ($\sigma < 10\%$), similarly to classic digital assays.

The presented methodology and design of quantitative assays may lead to the development of new, reliable and high-throughput devices for quantitative assaying and find use in the Point-of-Care applications.

Chapter 7

Rational design of digital assays

7.1 Introduction

The method that we describe in this chapter allows to fully benefit from the information provided by binary signals, which leads to the minimization of the number of partitions required for the assessment. The optimization is physically performed by the tuning of volumes and/or dilutions (more generally, the modulation factors) of the compartments comprising a digital assay. This, in turn, modulates the expected number of target molecules in each of the compartments, and therefore modulates the information gain. Also, we give the explicit analytical formulas for the rational design of the assay providing the assessment of concentration within the requested dynamic range and precision of the estimate (see Appendix B).

7.2 Mathematical routines

The result of a digital assay represented as a set of positive and negative signals yielded by the partitions of the sample represents (i.e. codes) the initial concentration of the analyte or the number of target molecules present in the sample. The simplest way of representing numbers is with equally valued symbols, or bits. For example, the number 157 can be represented by 157 dots. This, in general (with some sophistication from Poisson statistics), is the scheme of coding the concentration C (or the number of molecules) in a classic single-volume digital assay, where positive signals act as bits. This results in the requirement of large number of partitions of an assay to target practically large dynamic range of concentration of the analyte.

The problem of efficient representing large numbers can be solved by arranging bits in positional systems. For example, using binary coding, only 8 digits are required to represent the said value 157. In a chemical assay, the signals yield by partitions (compartments) play the role of bits. The difference is that they do not represent any fixed value, but the likelihood that a given partition contains at least one molecule of the analyte. The signals yielded are positive or negative, so they have Boolean character. This suggests the binary positional system to be used. It is also worth mentioning that the signals in a digital PCR assay are stochastic variables of the initial concentration C, and the parameters of the compartments: the volume (v) and the dilution factor ($d = C_{compartment}/C$) that produce the so-called modification factor $z = d \cdot v$. In this analysis, we assume the perfect efficiency of a single-molecule amplification (100%), and neglect all the effects regarding volume and/or dilution ratio, or surface effects of any particular compartment. This means that the probability of obtaining a particular signal is based on Poisson distribution and depends solely on the initial concentration C and parameters of the compartment v and d. We can easily calculate from the Poisson distribution, that the probability of positive signal equals: $p_{positive} = p(m \geq 1|C) = 1 - e^{-E(m)}$, and the probability of negative signal equals $p_{negative} = p(m \geq 0|C) = e^{-E(m)}$, where $E(m) = dvC$ is the expected number of molecules in the compartment. Therefore, the signal carries only a probabilistic information about the unknown initial concentration of the analyte that can be used for further calculations.

We can use the information retrieved from the probabilistic information by means of the Bayesian formalism to build the probability $\rho(C)$ that a given value of C has caused the recorded outcome of the assay (i.e. the collection of signals s_i yielded by the compartments). No *a priori* knowledge of the distribution of C can be represented by the constant density of probability $\rho_0(C) = 1/C_\infty$, which means a uniform probability of finding any particular value of C between t$C = 0$ and an arbitrary set upper bound C_∞. Therefore, the information about C gained from a single digital signal (with the notation used: $s = 1$ for positive or $s = 0$ for negative) is (Fig. 7.2a):

$$\rho(C|s = 1) = (1 - e^{-Cdv})/(C_\infty - (dv)^{-1}),$$

and

$$\rho(C|s = 0) = dve^{-Cdv}.$$

It is worth mentioning that all the information retrieved from the collection of signals is used in building the probability distribution ρ. However, the density of this information (or the information gain from compartments) is not uniformly distributed for all values of C. One can use the concept of the Shannon Entropy $H_b = \sum_i p_i log(p_i)$, where i counts all the possible

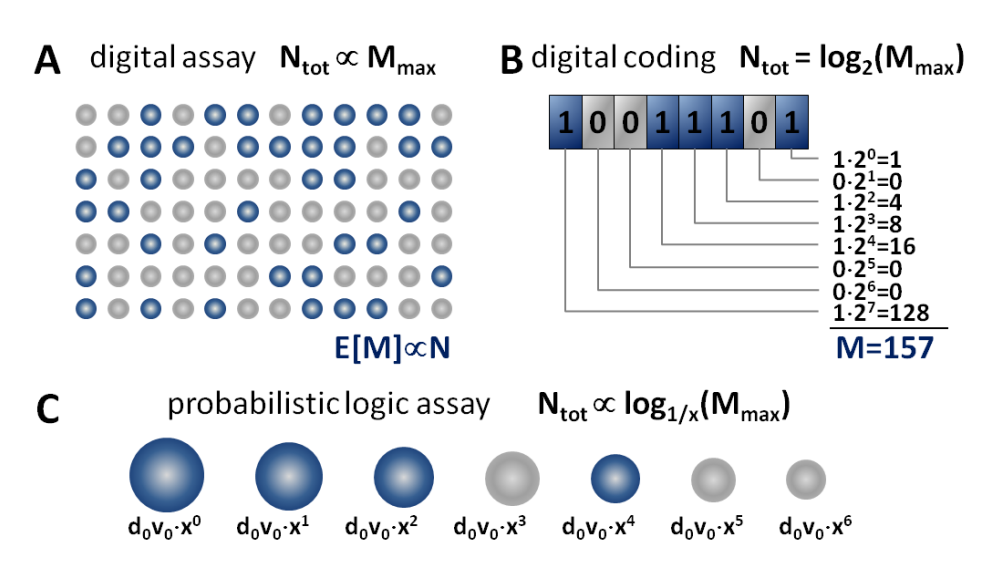


Figure 7.1: (a) In a classic single-volume digital scheme, the number of target molecules is determined using the count of positive signals from the total of the compartments, therefore $M_{max}N$. (b) In binary coding, the positional system comprising N bits allow coding numbers as high as $M = 2^N$. (c) Chemical bits are used as carriers of probabilistic information (each of the two values – positive and negative – can be represented with a probability function) used to build the hypotheses that C is larger (smaller) than some defined characteristic value C^*.

outcomes from the test (in this case, the two outcomes from a single compartment), to measure the information gain. It achieves maximum value $C^* = log(2)/dv$, which depends solely on the physical characteristics of the compartment. Therefore, most information can be extracted from a single signal s for the value of concentration near C^*, where the probability of obtaining positive and negative signal is equal: $p(s = 1|C^*) = 1/2$ (Fig. 7.2b). On the other hand, the same compartment carries much less (or even neglectable) information on much smaller $C \ll C^*$ or much larger $C \gg C^*$ concentrations.

The information provided by all the signals is combined to give the final information about C. The product of probabilities given by the individual signals from compartments is equal (in the *independent* approach used for this analysis) to the probablility of observing a particular sequence of signals $s_1, s_2, \ldots, s_N$, which will be called the microstate μ of an assay:

$$p(s_1, s_2, \ldots, s_N = \mu|C) = \sum_{n=1}^{N} p(s_n|C).$$

Therefore, the probability distribution

$$\rho(C|\mu) = p(\mu|C) / \int_0^\infty p(\mu|C)dC$$

conveys full information about the concentration C.

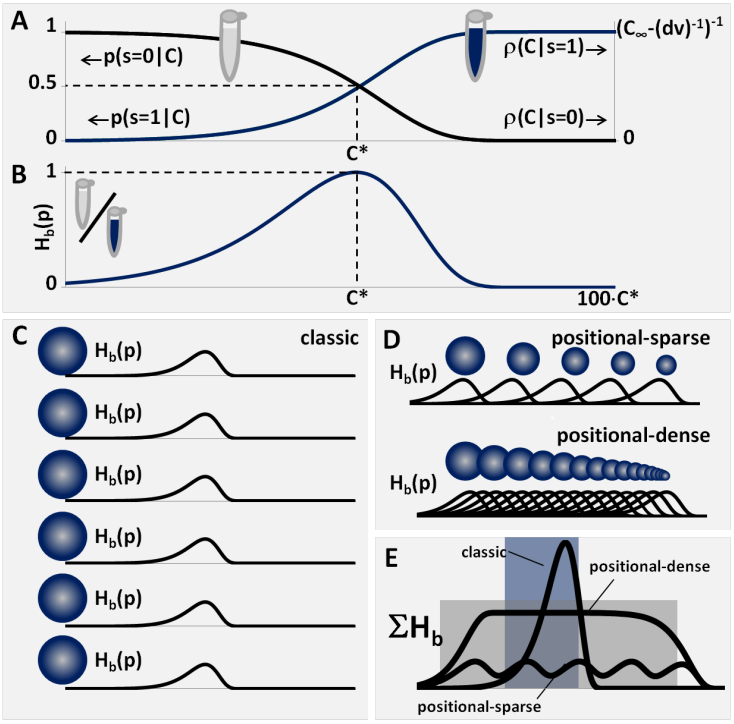


Figure 7.2: (a) The probability $p(s = 1|C)$ of finding at least one target molecule in volume v. Bayesian formalism translates the signal into probabilistic information $\rho(C|s = 1)$ on C. (b) The binary entropy function $H_b(p)$ for a single Bernoulli (positive/negative) trial achieves maximum at C^*: the signal provides most information about $C \approx C^*$. (c) Information from multiple compartments can be combined. For the set of identical compartments, the information gain is large in a very narrow range of concentrations, while (d) for geometrical sequence of volumes of compartments the information gain is almost constant in a wide dynamic range. (e) The geometric sequence of compartments allows to tune the amount of information retrieved from the assay (by tuning the common ratio x) for a wide range of concentrations - a feature that cannot be achieved with identical compartments.

In case of identical compartments, the probability of obtaining a particular outcome, i.e. exactly N_p positive and N_n negative signals from the total of compartments, is given by the Bernoulli distribution:

$$p(N_p; N_n = N - N_p = \mu|C) = \binom{N}{N_p} (p(1|C))^{N_p} (p(0|C))^{N_n}.$$

However, the information about the concentration C provided by the single-volume digital assay is qualitatively the same (has the same functional form) as the information provided by a single compartment (fig. 7.2b,c). Therefore, the precision of the estimate provided by a single-volume assay is not constant in the whole dynamic range. Moreover, obtaining reasonable estimates of the high concentrations (or numbers of target molecules) requires partitioning of the sample into massively large numbers of compartments.

A constant information gain for all the values of concentration C within the dynamic range can be provided by a sequence of compartments (i.e. their C^* values) that are spread uniformly over the axis of C. Typically, the quality of the assessment is determined by the coeffcieint of variance, i.e. the relative standard deviation of the estimate (expressed in the percentage of the estimated value). Hence, the distribution of compartments should be uniform along the logarythmic scale of concentrations, so the values C_i^* should form a geometric sequence.

7.2.1 Analytical model of a single compartment of the assay - digital measurement

In the mathematical model of the behaviour of an assay, we can assume that the molecules of the analyte form a uniform solution. In such a case, the expected number of molecules per one partition (compartment) of an assay is equal to $E(m) = Cdv$, where C is the concentration of the analyte, and v and d are the parameters of a single compartments: volume and dilution respectively.

If the solution is uniform and stable, the probability of finding a molecule of the analyte does not depend on the time and other parameters, apart from the charateristics of the said compartment. Therefore, the probability of finding z target molecules in a given partition is expressed by the Poisson distribution:

$$p(z|C) = (Cdv)^z e^{-Cdv}/z!$$

The PCR amplifies the presence of any non-zero amount of analyte up to a readable signal, therefore one should determine the probability of finding at least one molecule in a given compartment. Again, this can be calculated from the Poisson distribution and equals:

$$p(z \geq 1|C) = \sum_{z=1}^{\infty} p(z|C) = 1 - p(z = 0|C) = 1 - e^{-Cdv}.$$

The probability that no molecules were found in the compartment, which therefore yields a negative signal, is given as:

$$p(z = 0|C) = e^{-Cdv}.$$

The two probability functions that we described earlier are the basis for the mathematical interpeataion of the positive and negative signals from compartments, which is vital for the mathematical model of the response of an assay. Here, we treat the compartments as binomal random variables, and their state independent from each other (so-called *independent* approach decribed in previous chapters). Therefore, the probability that a particular sequence of signals, which

wille be called the *microstate* of this assay, was observed, equals the product of the probabilities for all of the compartments.

Such methodology only allows to specify the conditional probability of observing a particular sequence for a given concentration. To determine the concentration using solely the observed sequence of signals, the conditional probability must be inverted.

In probability theory, the relationship between the inverse conditional probabilities for two events, is described by the Bayes' theorem. Here, as an input the conditional probability $p(s|C)$ of observing the outcome s provided the concentration C is used. The wanted result of the procedure is the conditional probability $\rho(C|s)$ of a concentration C provided the signal s. Obiously, the signal s has a discrete character, while C is a continuous variable, therefore the function $\rho(C|s)$ can be determined as:

$$\rho(C|s) = (p(s|C)f(C))/p(s),$$

where $f(C)$ is the distribution of concentration C. However, no a priori knowledge about the concentration is available, therefore $f(C)$ is set to be uniform within the range of assessment, i.e. $f(C) = 1/C^{\infty}$, for $C \in (0, C^{\infty})$. The function $p(s)$ equals the total probability of obtaining signal s, i.e. the sum of probabilites of obtaining s for all achievable concentration values: $p(s) = \int_0^{C^{\infty}} [p(s|C) \cdot f(C)]dC$. Then, the function $\rho(C|s)$ can be given by:

$$\rho(C|s) = (p(s|C)f(C))/\int_0^{C^{\infty}} [p(s|C) \cdot f(C)]dC = (s|C)/\int_0^{C^{\infty}} p(s|C)dC.$$

7.2.2 Estimation of concentration $E(C)$ and the uncertainty of this estimation $\sigma(C)$ for the simplest digital assay comprising 2 identical compartments.

Because the compartments belonging to an assay are intepreted mathmatically as independent random variables, the probability of obtaining a non-trivial microstate form the simplest set two compartments, one being positive and the other negative, is given by the product of conditinal probabilities:

$$p_{21}(s_2 = 0, s_1 = 1|C) = p(s_2 = 0|C)p(s_1 = 1|C) = e^{-Cdv}(1 - e^{-Cdv}) = e^{-Cdv} - e^{-2Cdv}$$

The inversed probability $\rho_{21}(C|s_2 = 0, s_1 = 1)$, where $C^{\infty} \to \infty$ is, according to the Bayes' therem, equal to:

$$\rho_{21}(C|s_2 = 0, s_1 = 1) = \frac{p_{21}(s_2 = 0, s_1 = 1|C)}{\int_0^{\infty} p_{21}(s_2 = 0, s_1 = 1|C)dC} = \frac{-e^{-Cdv} + e^{-2Cdv}}{(2dv)^{-1}}.$$

This probability distribution is then used to determine the estimate $E(C)$ of the initial concentration of the analyte:

$$E(C) = \int_0^\infty C \cdot \rho_{21}(C|s_2 = 0, s_1 = 1)dC.$$

This, in turn, equals the following:

$$E(C) = \frac{3}{2dv}.$$

The precision of the assessment, which is given as the relative standard deviation of the estimate, equals

$$\sigma(C) = \sqrt{E(C^2) - (E(C))^2)/E(C).}$$

This provides the final result:

$$\sigma(C) = \sqrt{7/(2v^2 - (3/(2v))^2} = \sqrt{(5)}/3.$$

7.2.3 Analytical model of the microstate μ of the assay.

The quantitation of the initial concentation $E(C)$ can be provided by the calculations basing on the microstate $\mu = \{s_i\}$ of an assay, i.e. the sequence of positive and negative signals yielded by compartments. This approach allows to simplify the mathematical routines and helps to reduce the number of compartments actually needed to assess the initial concentration with a given confidence. It consists of calculating the product conditional probability of observing a given microstate $p(\mu|C)$ provided the concentration C:

$$p(\mu|C) = \prod_{i=0}^{N-1} (1 - e^{-v_i d_i C})^{s_i} \cdot (e^{-v_i d_i C})^{1-s_i}.$$

Then, this probability is inverted in order to get the distribution of concentration $\rho(C|\mu)$:

$$\rho(C|\mu) = p(\mu|C) \cdot f(C) / \int_0^\infty p(\mu|C) f(C) dC,$$

and using this distribution to calculate the estimate (expected value) of concentration $E(C)$:

$$E(C) = \int_0^\infty C \cdot \rho(C|\mu) dC$$

and the precision ot estimate, i.e. the relative standard deviation $\sigma(C)$:

$$\sigma = \sqrt{E(C^2) - (E(C))^2}/E(C)$$

where

$$E(C^2) = \int_0^\infty C^2 \cdot \rho(C|\mu) dC.$$

For this calculation, the volumes $\{v_i\}$ and dilution ratios $\{d_i\}$ for all the compartments must be *a priori* known.

These calculations that lead to the estimate of initial concentration function $E(C) = g(\mu)$ can be performed for any set of volumes and dilutions $d_i v_i$. The characteristics of the assay, i.e. the dynamic range and precision of the estimates, depend on the set of values $d_i v_i$, and the total number of partitions.

7.2.4 Distribution of standard deviation of the estimate of concentration from different microstates.

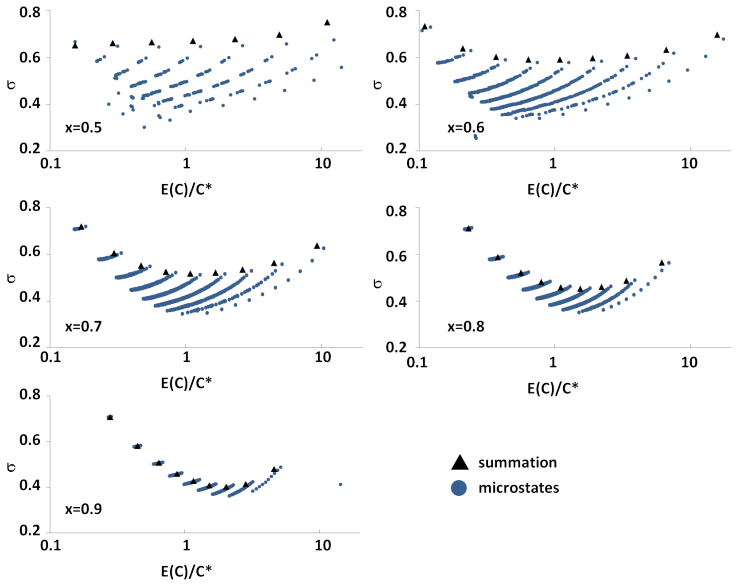


Figure 7.3: (a) The standard deviation of the estimate of concentration calculated using two methods: summation of positive compartments and microstates for assays comprising test-volumes arranged in geometrical sequence with common ratio x. In most cases, the standard deviation of the estimate is higher for the summation procedure than for microstate analysis. Moreover, the microstate approach provides much more frequent probing of concentration, i.e. 2^N times instead of N times for summation algorithms.

Quite intuitively, every allowable outcome of an assay, i.e. the microstate of an assay, yields a different estimate $E(C)$ of the initial concentration and different σ. The distribution of the values of σ from the microstates can be used for a quantitative description of the assay's performance. In this analysis, the highest from the set values of the relative standard deviation of the estimate from microstates defines the precision of the test. However, it is not the only possible solution. Instead of using the maximum value of σ, we can use an average precision. We show the distribution of standard deviation values in Fig. 7.4. We can closely approximate the average

(expected) standard deviation of the estimate as a function of $\{d_i v_i\}$ values that determine the parameters if assay's compartments. For example, if the volumes (or dilutions) of compartments follow a geometric sequence, the average standard deviation can be approximated as a polynomial function of the common ratio x of this sequence:

$$E(\sigma) = -0.868x^2 - 0.3208x + 0.6765.$$

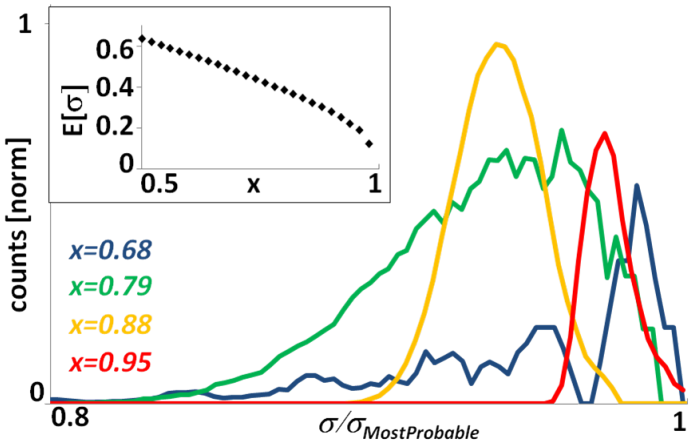


Figure 7.4: The distribution of the standard deviation of the estimate $E(C)$, normalized by the maximum standard deviation of the estimate of the concentration for the most probable microstate, for all microstates of the assay defined by common ratio x. The distributions are calculated using 1024 microstates (i.e. each assay contained $2{\cdot}\Delta N{+}10$ compartments). For each microstate, the probability density $p(\mu|C)$ was calculated and inverted to get $\rho(C|\mu)$ needed to calculate $E(C)$ and $\sigma(C)$. All the values $\sigma(C)$ were then plotted in a histogram. For smaller values of x the distribution of σ is wider, and for $x = 0.68$ or less it is divided into a set of peaks, because of the big difference between 'microstate families', i.e. (100000...), (110000...), (111000...), etc. and (101000...), (110100...), (111010...), etc.. In the inset the expected value of $\sigma(C)$ as a function of the common ratio x is shown.

7.3 Derivation of the algorithm

As we indicated earlier (Fig. 7.2b), the information yielded by single compartment is the largest near the characteristic concentration determined by the properties of the said compartment. Therefore, it should be possible to arrange the compartment of an assay in such a way, that a steady, efficient information gain is provided in the whole dynamic range. The setting of compartments should be prepared with the following problems in mind:

- most efficient probing of unknown concentration with a set of compartment volumes and dilutions, and
- the extraction of information from the signals yielded by the compartments.

The aim of this analysis is to determine the characteristics of the geometric sequence of compartments: the common ratio x, which determines the ratio modulation factor (i.e. ratio of the product of volume and dilution of two consecutive compartments $x = \frac{d_{i+1}v_{i+1}}{d_i v_i}$), the value of the first term $v_0 d_0$, and the number of compartments N for a given requested dynamic range and precision of the estimate of C.

7.3.1 Derivation of analytical formulas for assay design.

In order to find an optimum sequence of compartments, i.e the positional system that will be used to code the concentration, we should choose an arbitrary input concentration C_{input} and set a compartment $d_0 v_0 = log(2)/C_{input}$ (i.e. set $C_0^* = C_{input}$), which is 'centered' at this concentration. The, on each side (i.e. higher and lower concentrations) of the said compartment we add a number ΔN of compartments: $d_i v_i = d_0 v_0 x^i$, $i \in (-\Delta N, \Delta N)$. Half of these compartments is larger and half of them is smaller than the central one, and they are arranged in a geometric sequence of modulation factors $d_i v_i$ parametrized by the common ratio x. The probability that an inspected compartment yields a positive signal depends solely on the volume and dilution of the compartment: $p_i = 1 - e^{-C_{input} d_0 v_0 x^i}$ (Fig. 7.2a). All the relevant about the initial concentration comes from the analysis of signals yielded by compartments surrounding the central one that belong to the *active stripe*. We can notice, that adding more compartments to an assay (or a fixed value of x), and therefore increasing the number ΔN, improves the precision of the assessment (the standard deviation of the estimate decreases). However, this improvement is limited, because at some point, the precision of the assessment saturates because new compartments are too far from the actual concentration and the information increase is negligible. The precision $\sigma(C)$ saturates at a limit $lim_{\Delta N \to \infty}\sigma(x)$, which depends only on the common ratio x (Fig. 7.6a).

The dependence of $lim_{\Delta N \to \infty}\sigma(x)$ on the common ratio x of the sequence of compartments can be closely approximated with a function: $lim_{\Delta N \to \infty}\sigma(x) = a(1-x)^b$, where $a = 1.2739$ and $b = 1.9895$ (Fig. 7.6b) in a practically wide range of x and $\sigma(x)$. In the design formulas, we use the maximum allowable relative standard deviation σ_{max} of the estimate as an input, therefore the term $lim_{\Delta N \to \infty}\sigma(x)$ should be replaced with σ_{max}. Then, the the equation should be inverted

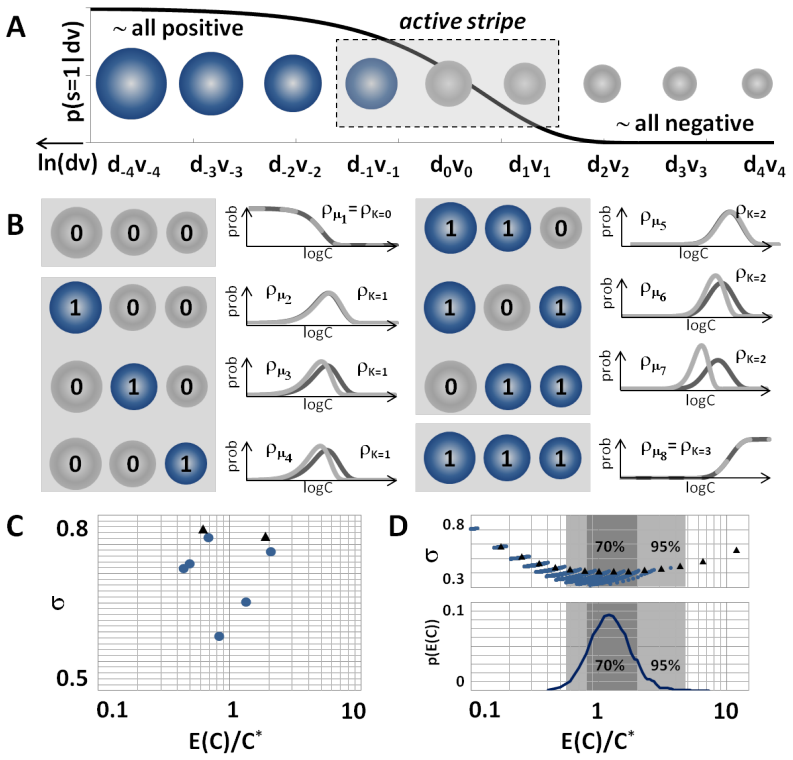


Figure 7.5: (a) For any C_{input} only one compartment (here $v_0 = log(2)/C_{input}$) addresses this concentration most closely. The progressively larger/smaller compartments carry less information as the probability of obtaining a signal from these compartments approaches unity/zero. The active stripe (gray area) comprises a set of volumes with a probability of yielding a given signal significantly different than zero or one. (b) An active stripe yields stochastic combinations of signals, or microstates (μ_i), in every run of the assay. Each microstate yields $\rho(C|\mu_i)$ that provides more information on C than that retrieved from the sole number of signals S. Each of the distributions $\rho(C|\mu_i)$ have different standard deviation. Inset (c) shows the estimated concentrations and the standard deviations based on the analysis of all possible microstates (green dots) of an active stripe comprising three compartments. The black triangles show the estimates and standard deviations retrieved from the sum of positive signals (here either $K = 1$ or $K = 2$). Inset (d) illustrates the same information for an active stripe comprising 15 compartments. The shaded areas indicate the range of the estimates of concentration given by 70% (95%) of most probable outcomes of the assay.

in order to allow finding the common ratio x of the geometric sequence of compartments: $x = 1 - \alpha\sigma_{max}^{\beta}$ for a requested precision.

Knowing the analytical formula for the determination of the common ratio x, we have to calculate the number $2 \cdot \Delta N_x + 1$ of compartments belonging to the *active stripe*. Intuitively, the optimum value of ΔN_x is the lowest integer ΔN, for which the precision of the assessment saturates

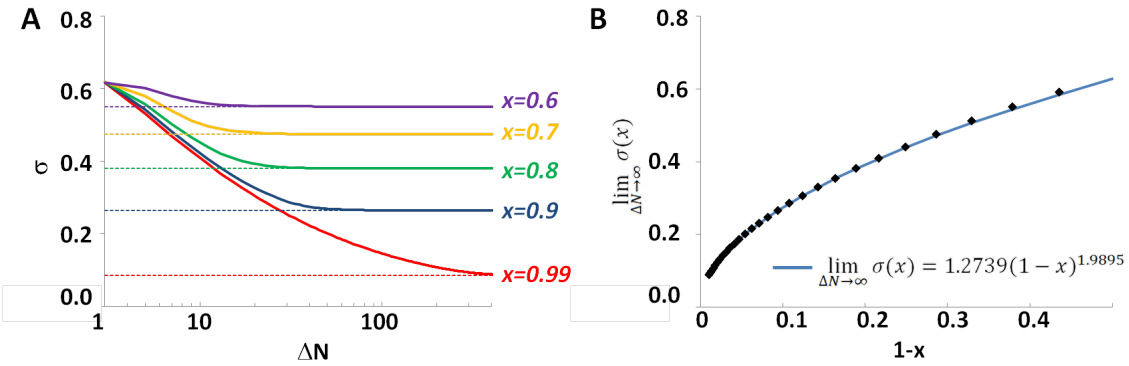


Figure 7.6: (a) The precision (standard deviation) of the estimate of the concentration depends on the number of compartments analyzed within the active stripe. The decrease of the standard deviation is limited to a value given for an optimum count of the active stripe. This count (ΔN_x) is a well defined function of σ, and consequently of x. (b) The limit (for $\Delta N \to \infty$) of standard deviation σ as a function of the common ratio x of the geometric sequence of the factors $d_i v_i$ of compartments in the active stripe. The data points calculated numerically from the analytical model based on the Bayes formalism. The solid line shows the algebraic fit to the data: $lim_{\Delta N \to \infty}\sigma(x) = a(1-x)^b$ with $a = 1.2739$ and $b = 1.9895$.

(i.e. σ reaches $lim_{\Delta N \to \infty}\sigma(x)$). Mathematically, it is the solution to the following equation which the derivative $\frac{d\sigma(\Delta N)}{\Delta N} = 0$. Here, we calculated the value of ΔN_x, and the treshold value for $\frac{d\sigma(\Delta N)}{\Delta N} = 0$ was set to $1/1000$, so the further addition of compartments will not increase significantly the precision of assessment. The value of ΔN_x can be closely approximated with a simple analytical function of x: $\Delta N = \lceil \Delta N_x \rceil = \lceil \delta \sigma_{max}^{-2} \epsilon \sigma_{max}^{-1} - \phi \rceil$. Therefore, the equations that allow determining the parameters of the *active stripe* are found and they depend solely on the requested precision of the assay.

The above formulas allow us to determinine the common ratio of the geometric sequence of modulation factors of compartments that assure the requested precision of the assessment is provided. However, the assessment of concentration should be given withing the limits of dynamic range $\Omega = C^+/C^-$ of concentrations requested. Being arranged in a geometric progression, the sequence of modulation factors of compartments is self-similar, so the previously defined *active stripe* can be just extended to cover the requested dynamic range $C \in (C^-, C^+)$ by simply adding a number $log_x(1/\Omega)$ of compartments arranged in the same geometric sequence. The 'margins' of the active stripe, each having ΔN_x compartments, should be kept in order to provide the precision of assessment at the bounds of the dynamic range (Fig. 7.8). The most remote compartment (from the central one) in the sequence should be shifted from the limit of dynamic range (here, it is shifted from the lower bound, but same could be determined for the upper

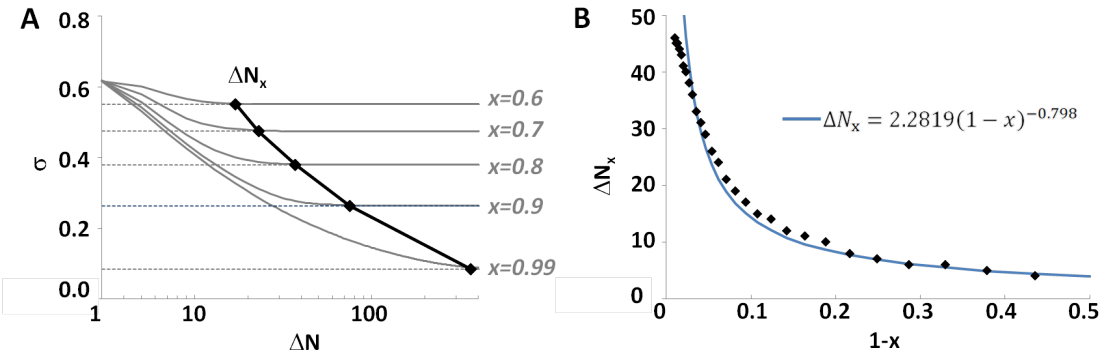


Figure 7.7: (a) The optimum number ΔN_x of compartments in the active stripe defined as the lowest integer value at which the standard deviation $\sigma(x)$ saturates at the limiting value $lim_{\Delta N \to \infty}\sigma(x)$. (b) The graph of ΔN_x as a function of x. The data points calculated numerically from data for relation $\sigma(\Delta N, x)$ and from the condition $d\sigma/d\Delta N = 0.001$. The solid line shows the algebraic fit to the data: $\Delta N_x = \gamma(1-x)^{-\delta}$ with $\delta = 0.798$ and $\gamma = 2.2819$.

bound of concentration range) $d^- v^- = log(2)/C^-$ by the value $x^{-\Delta N}$.

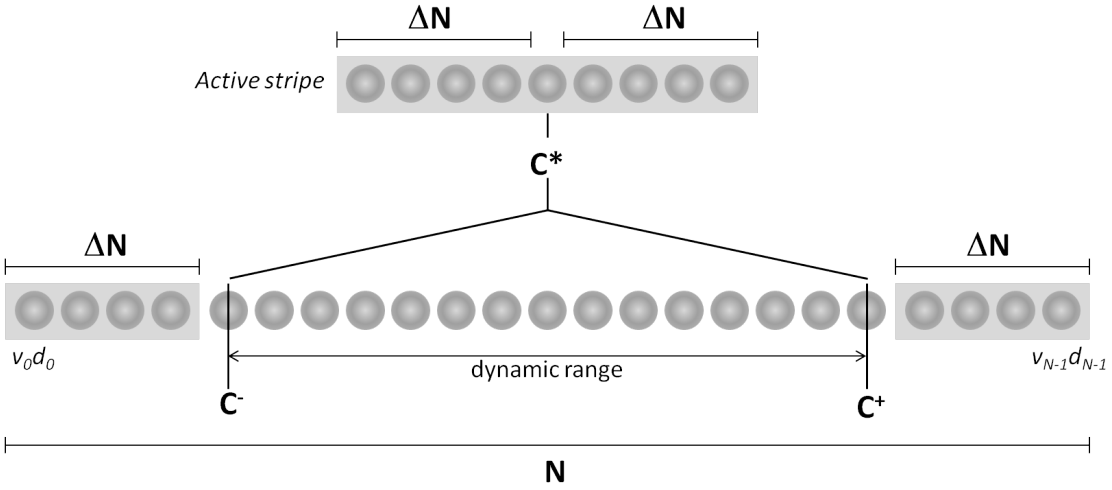


Figure 7.8: Construction of an assay. The number $2\Delta N_x + 1$ determines all the compartments that carry useful information about C^*. If one takes all these compartments into account (the whole active stripe) then the precision depends only on the value of x. Further, as the above is true for any value of v_0 (i.e. for any C^*) and since the geometric progression is self similar, if it is wished to guarantee the same precision for all C's within a required range $C \in (C^-, C^+)$, it is enough to span the assay, keeping the required x and the required 'margins' of compartments outside $(d^- v^-, d^+ v^+)$, with $d^{+/-} v^{+/-} = log(2)/C^{+/-}$.

Finally, a complete design o an assay can be determined using the following equations. As use explicitely as an input the requested dynamic range $\Omega = C^+/C^-$ and precision σ_max of the estimate $E(C)$:

$$x = 1 - \alpha\sigma_{max}^{\beta}$$

$$\Delta N = \lceil \Delta N_x \rceil = \lceil \gamma (1-x)^{-\delta} \rceil$$

$$N = 2\Delta N + \lceil log_x(1/\Omega) \rceil$$

$$d_0 v_0 = log2 \cdot x^{-\Delta N} / C^-$$

where α, β, γ and δ are positive constants:

$\alpha = 1.24$, $\beta = 1.9493$, $\gamma = 2.2815$, and $\delta = 0.798$.

7.3.2 Rational design of digital assays

The analytical expressions we derieved in previous section come from the close approximations to Monte Carlo data, and they demonstrate how the precision of the assement depends on the design of the active stripe and *vice versa*. They can be used for a practically wide range of requested precision $\sigma_{max} < 0.89$.

The procedure of designing the assays comprises the following steps:

1. calculation of N, ΔN and x for a given requested precision σ_max and limits of the dynamic range: C^- and C^+,

2. calculating the modulation factor $d_0 v_0$ of the first compartment in the sequence using x and ΔN, and

3. creating the geometric sequence of N compartments and volumes determined by: volume $d_i v_i = d_0 v_0 x^i$ for $i = 1, \ldots, N-1$.

7.3.3 Libraries

The gradation of the compartments x is usually much finer than the precision of the estimate, which can be formally expressed as $x^{-1} - 1 \ll \sigma$. Therefore, for the most precise assays, i.e. the assays that provide the smallest relative standard deviation of the estimate, the quality of the assessment stems just from a large number of compartments used, rather than the fine gradation. What is more, this fine gradation (i.e. $x > 0.95$) can be technically challenging to perform. A resonable solution is to use coraseer gradation, but supplement each compartment with a set of copies with the same parameters $\{d_i c_i\}$. The sets of identical compartments will be called libraries. This scheme is similar to the multi-volume assays described in previous chapter and some state-of-art solutions, including a SlipChip device [5–7].

Formally, each partition characterized by the dilution factor d_i and volume v_i can be prepared N_i' times, increasing the number of partitions of the assay up to $N_{tot} = \sum_{i=0}^{N-1} N_i'$. Also, the definition of a microstate changes; now it is defined as a set of ratios of positive compartments in each library $d_i v_i$ instead of being just a collection of binary values.

If al the libraries contain the same number of compartments N', the standard deviation σ of the estimate $E(C)$ simply scales with $1/\sqrt{N'}$. Therefore, for the request precision of the assessment σ_{max} the value of x can be calculated from the following equation:

$$x = 1 - \alpha(\sigma_{max}\sqrt{N'})^{\beta}.$$

This also allows tailoring the assays to the capabilities of laboratory equipment, i.e. designing assays that provide a requested precision of the assessment, but use only a preferable gradation x of the compartments: $N'(x, \sigma_{max})$ (the detailed instruction for the design of such assays is given in the Appendix B).

The equations we derive here allow the trade-off oetween the gradation x and the number of copies of compartments N' (Appendix B).

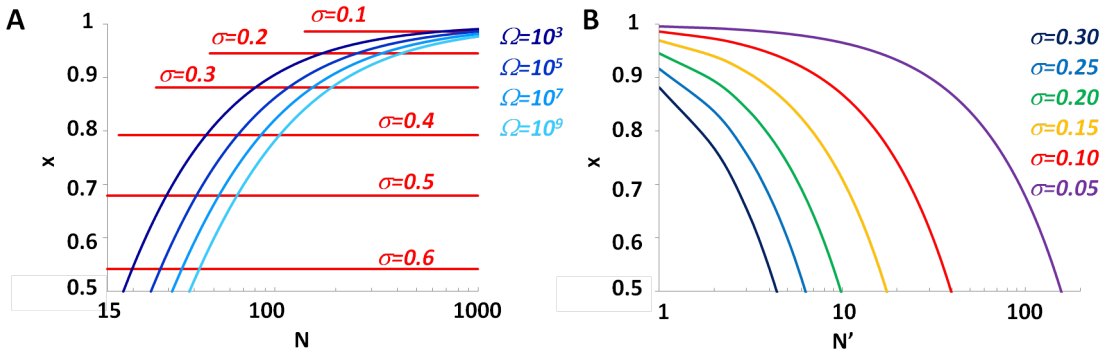


Figure 7.9: (a) The algorithm allows to independently tune the standard deviation σ of the estimate and the dynamic range Ω of the assay as a function of the parameters N and x. The graph shows how σ and Ω depend on N and x and – *vice versa* – can be used to estimate N and x for a given pair of σ and Ω. (b) In addition, for any chosen dynamic range, and precision, the factor x of geometric progression of volumes can be interchanged with the number N' of copies of the compartments.

7.3.4 Detailed description of precision-varied assays

Some applications of digital assays require the assessment that varies in precision for different concentration ranges. A digital assay we describe here can be tuned to meet this requirement. In

order to provide different precision of the estimate, the common ratio x of the geometric sequence modulation factors has to be changed. If the dynamic range of concentration to be assessed (C^-, C^+) is divided into j subranges(C_n^-, C_n^+), the sequence of $d_i v_i$ covering each subrange has a different common ratio x_n, whereas:

For $\sigma_{max_n} < \sigma_{max_{n-1}}$:

$$x_n = 1 - (\sigma_{max_n}/0.8955)^{1/0.513},$$

$$\Delta N_n = \lceil 2.2815(1 - x_n)^{-0.798} \rceil,$$

$$N_n = 2\Delta N_n + \lceil log_{x_n}(C_n^-/C_n^+) \rceil,$$

$$d_{n_0} v_{n_0} = log(2) x_n^{-\Delta N_n}/C_n^-.$$

For $\sigma_{max_n} > \sigma_{max_{n-1}}$:

$$x_n = 1 - (\sigma_{max_n}/0.8955)^{1/0.513},$$

$$\Delta N_n = \lceil 2.2815(1 - x_n)^{-0.798} \rceil,$$

$$N_n = \Delta N_n + \Delta N_{n-1} + \lceil log_{x_n}(C_n^-/C_n^+) \rceil,$$

$$d_{n_0} v_{n_0} = log(2) x_{n-1}^{-\Delta N_{n-1}}/C_n^-,$$

where C^- is the lower limit of the dynamic range of concentrations,

C^+ is the upper limit of the dynamic range of concentrations,

C_n^- is the lower limit of the n-th subrange,

C_n^+ is the upper limit of the n-th subrange,

σ_{max_n} is precision of the estimate $E(C)$ in the n-th subrange,

n is the number of subrange, and

i is the number of compartment of an assay.

7.4 Performance of Rational Digital assays

The design we present here is a powerful analytical tool. It allows an independent tuning of dynamic range and precision (Fig. 7.9a). The number of compartments required for the assessment

is dramatically lowered: the assays require orders of magnitude less compartments than the classic digital assays to cover a given, requested dynamic range Ω with precision σ_{max}. For example, in order to cover the dynamic range $\Omega = 10^4$ or $\Omega = 10^7$ using single-volume digital assays, one needs $N = 2 \cdot 10^3$ or $N = 2 \cdot 10^3$ compartments respectively. However, if the level of precision $\sigma_{max} = 50\%$ is satisfactory, the same dynamic range Ωs can be assessed with just $N = 35$ or $N = 47$ compartments. If bettter precision of the estimate is expected, (e.g. $\sigma_{max} = 25\%$ or $\sigma_{max} = 10\%$) $N = 140$ and $N = 795$, or $N = 192$ and $N = 1120$ test-volumes are required.

The method we described here can be used to design exemplary digital assays that provide:

1. moderate precision of the assessment within a wide dynamic range, using a small numer of compartments (e.g. $\Omega = 10^9$ with $\sigma_{max} = 50\%$ with $N = 65$),

2. high precision of the assessment within a wide dynamic range (e.g. $\Omega = 10^9$, $\sigma_{max} = 10\%$, N=1615), or

3. high precision of the assessment within a narrow dynamic range, again using a small number compartments (e.g. $\Omega = 10^3$ offering $\sigma_{max} = 30\%$ can be realized with $N = 80$ and offering $\sigma_{max} = 10\%$ with $N = 631$).

7.4.1 Information loss caused by not tracking the compartments

So far we used the concept of a *microstate*, i.e. the combination of positive and negative signals coming from identifiable compartments, for the analysis. This requires the tracking of the compartments during the amplification, which is an additional complication compared to classic digital assays, where the tracking is not needed. However, the algotirhms presented in this chapter can be also used with non-identifiable compartments. This is associated with some loss of the information stemming from the fact that there no linking between the signal and respective compaerments. Fortunately, this loss of information is relatively small and can be qunatitatively determined. Fistly, we can observe that the probability of observing a given sum of positive signals equals the sum of probabilities of all the microstates comprising the same number of positive compartments. Secondly, the biggest input to this sum of probabilities, which dominates the final result, is given by by the most probable microstates, which are also most regular (for example, in a simple assay shown in Fig. 7.5b the contribution of the most probable microstate is more than 60% of the final outcome).

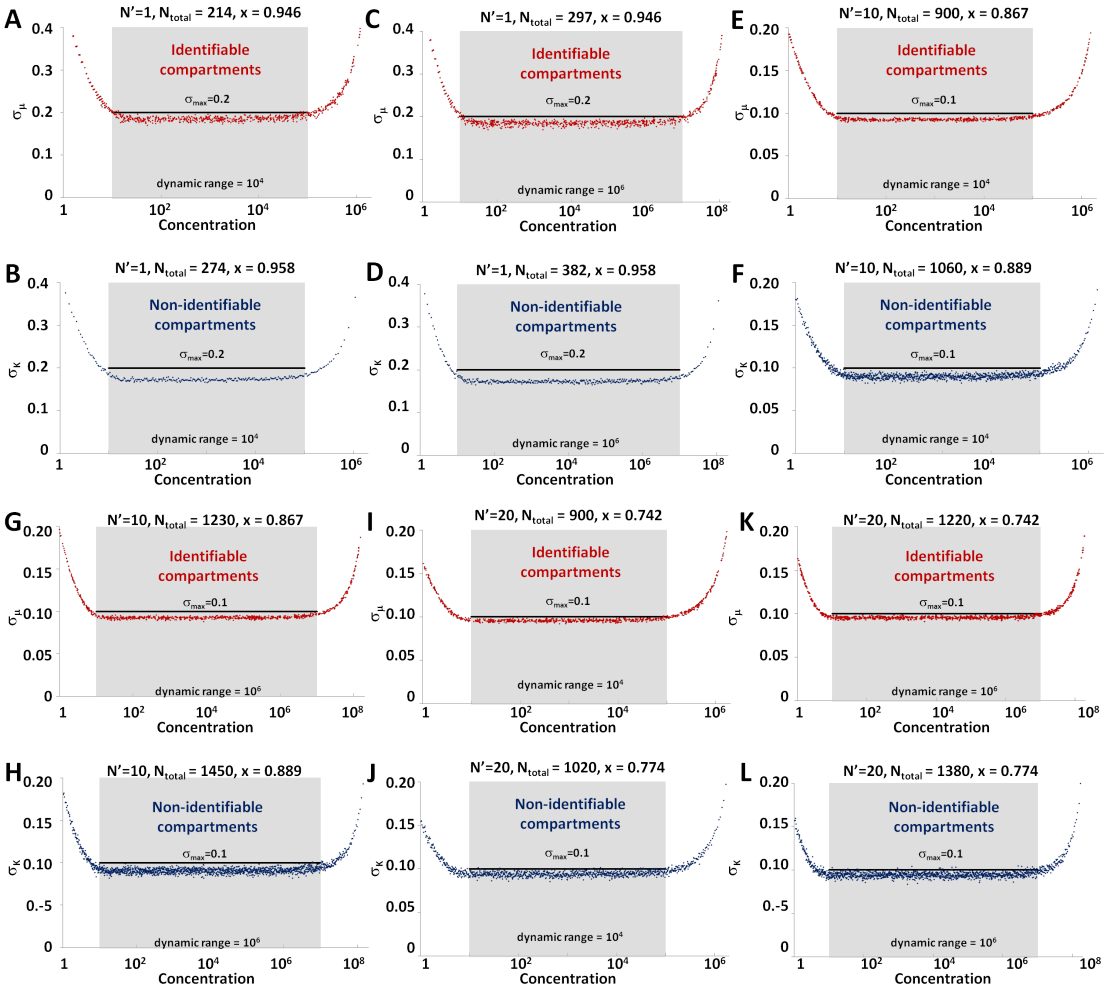


Figure 7.10: The performance of the assays designed to provide the requested precision of the assessment for a given range of concentrations using both microstate analysis and summation of positive signals. Graphs (a-d) show the precision provided multivolume assays (all the compartments have different volume and dilution products) designed for the assessment with $\sigma_max = 0.2$ in the range of concentrations $C \in (10, 10^5)$, i.e. $\Omega = 10^4$ (a, b), and concentrations $C \in (10, 10^7)$, i.e. $\Omega = 10^6$ (c, d). Graphs (e-h) show the precision provided assays (comprising sets, each containing 10 identical compartments) designed for the assessment with $\sigma_{max} = 0.1$ in the range of concentrations $C \in (10, 10^5)$, i.e. $\Omega = 10^4$ (e, f), and concentrations $C \in (10, 10^6)$, i.e. $\Omega = 10^6$ (g, h). Graphs (i-l) show the precision provided assays (comprising sets, each containing 20 identical compartments) designed for the assessment with $\sigma_{max} = 0.1$ in the range of concentrations $C \in (10, 10^5)$, i.e. $\Omega = 10^4$ (i, j), and concentrations $C \in (10, 10^7)$, i.e. $\Omega = 10^6$ (k, l). Red data points represent the results of Monte Carlo simulations for microstate analysis and blue data point represents the results of Monte Carlo simulations for summation analysis. Black lines shows the requested value of σ_{max} and gray regions represent dynamic ranges of the assays.

Concerning these observations, we can conclude that the precision of the assessment provided by

an assay operated on the sum of signals from unidentifiable compartments is only slightly lower than the precision of the estimate provided by assays based on the microstate.

We determined quantitatively the drop of precision of the assessment for the situation when the tracking of the compartments is not available, by the means of numerical calculations. For a set of *active stripes*, we compared the precision of the assessment with the tracking of the compartments σ_{max} (used as an input in design formulas) with the numerically established precision σ_K of the assessment without the tracking of compartments. The relation between the two parameters $\sigma_K(\sigma_{max})$ is can be closely approximated. For $\sigma_{max} < 0.55$, this function is linear and equals $\sigma_K \approx 1.023\sigma_{max} + 0.021$. Therefore, the design formulas of an assay for the assessment based on the sum of positive signals instead of a microstate are qualitatively the same, but the term σ_{max} is replaced by $0.978\sigma_{max} - 0.02$. The performance of assays based on microstates and summation of positive signals is given in Fig. 7.10.

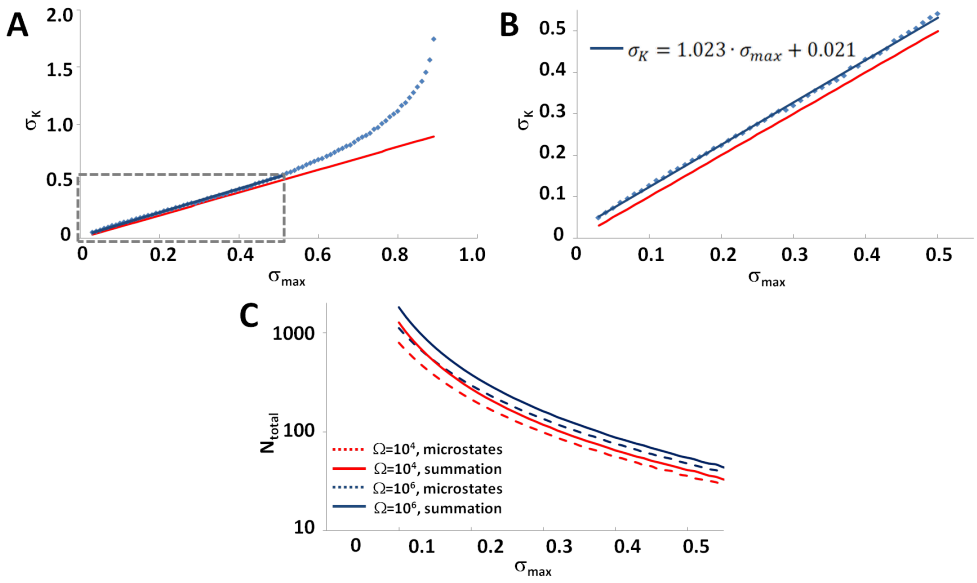


Figure 7.11: The comparison between the analysis based on the microstate of the assay and sum of positive compartments (macrostate of the assay). (a) The precision provided by an assay for summation approach as a function of the precision provided by microstate approach. The red line shows 1:1 relation. The precision σ_K of the assessment based of summation approach was calculated for a set of active stripes designed to provide the assessment with requested precision σ_{max} for microstate approach. For a wide range of values of precision σ_K is a linear function of σ_{max} (b) and can be determined as $\sigma_K \approx 1.023\sigma_{max} + 0.021$. (c) The loss of information due to not tracking the compartments translates into a slightly higher number of compartments needed to assess the initial concentration with the same precision. The graph shows the comparison between the number of compartments needed to assess concentration with the required precision σ_{max} in the range of concentrations $\Omega = 10^4$ (red lines) and $\Omega = 10^6$ (blue lines) using both microstate (dotted lines) and summation (continuous lines) approaches.

In conclusion, the information loss is relatively small. Therefore, in order to keep a precision of the assessment for microstate analysis, the necessary increase the number of compartments is usually much smaller than 100% (Fig. 7.11c). Practically, for many applications, it seems to be easier just to double the number of compartments and alleviate the need to track them, especially in microfluidic systems which allow generating a large number of partitions, thermocycling them together, and simply counting the number of positive compartments.

7.4.2 Resolution of the assays.

The procedure we described here enables the design of assays for a particular precision and dynamic range. However, in many practical applications, it is useful to know the resolution of

an assay, i.e. the ability of an assay to distinguish different input concentrations, with a set confidence. Therefore, it is useful to assess the resolution of an assay. The resolution can be determined as the smallest relative difference between the two concentrations that is noticed with a given confidence: 70% (1-fold resolution), 90% (2-fold resolution), or 95% (3-fold resolution).

The resolution of an assay is therefore closely related to the previously defined precision. The 1-fold resolution equals the change of the standard deviation, so the concentrations are distinguished with 70% confidence. The 2-fold or 3-fold resolution requires the discrimination with 90% and 95% confidence, or two or three standard deviations, respectively.

The performance of an analytical assay can be assessed using many parameters. One of the possibilities is the precision $\sigma(C)$ of the estimate, i.e. the relative standard deviation, which is used in this work. Another is the accuracy of the test, which is determined by the spread of outputs, i.e. the values of concentrations calculated from the set responses of the assay for a constant input concentration. Hence, it determines the repeatability of the assay. The distribution of outputs values ($\delta(C)$) yielded by the assay for the same input concentration can be calculated numerically using Monte Carlo simulations. Also, we can determine the relative difference between the input and calculated concentration $\gamma(C)$.

We carried the analysis of the performance of digital assays designed the formulas derived in previous sections. The tested assays were design to cover the dynamic range $\Omega = 10^6$ and provide the precision of the assessment varying in the range $\sigma(C) \in (15\%, 50\%)$ (Fig. 7.12).

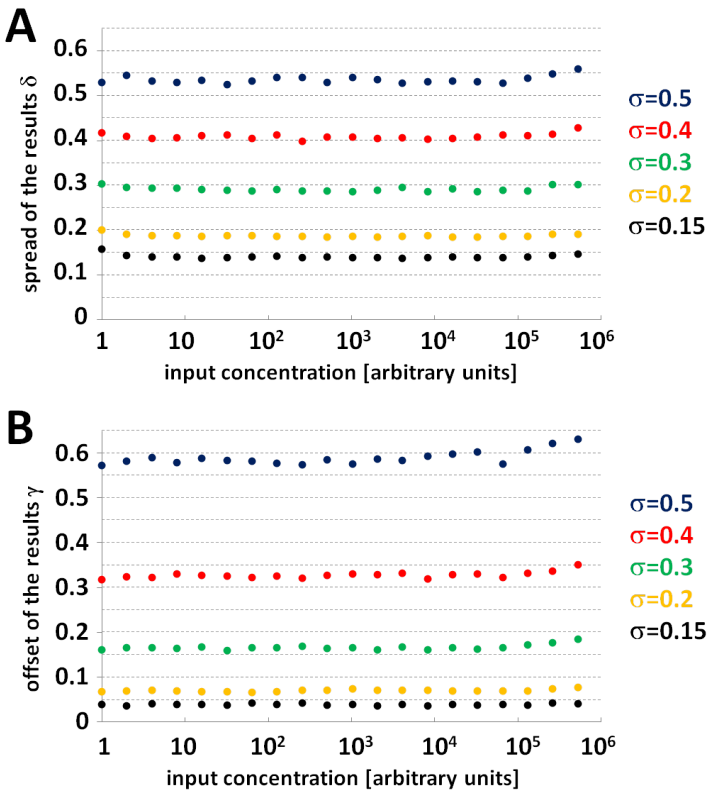


Figure 7.12: (a) The spread $\delta(C)$ of the output values E(C) given by the same assay for the same concentration of the analyte may be used to determine the precision of this assay. The precision of the assay designed using the rational digital PCR algorithm was tested for assays with $\sigma(C) \in (15\%, 50\%)$ via Monte Carlo simulations. The simulations were run 10000 times for each value of concentration for each assay design. The values of the spread are close to predetermined values of $\sigma(C)$, which means that the precision of assay $\delta(C)$ can be determined using the precision $\sigma(C)$ of the estimate of the concentration. (b) The difference $\gamma(C)$ between the estimate of concentration $E(C)$ calculated from the state of the assay and the real value of concentration C is another useful factor that characterizes the accuracy of this assay. The accuracy of the assay designed using the rational digital PCR algorithm was tested for assays with $\sigma(C) \in (15\%, 50\%)$ via Monte Carlo simulations. The simulations were run 10000 times for each value of concentration for each assay design. For the majority of assays, the values of the accuracy are lower $(E(C))$ is closer to the real value of (C) than the predetermined values of $\sigma(C)$ and are constant for the whole dynamic range of the assay.

7.5 Complete list of design protocols

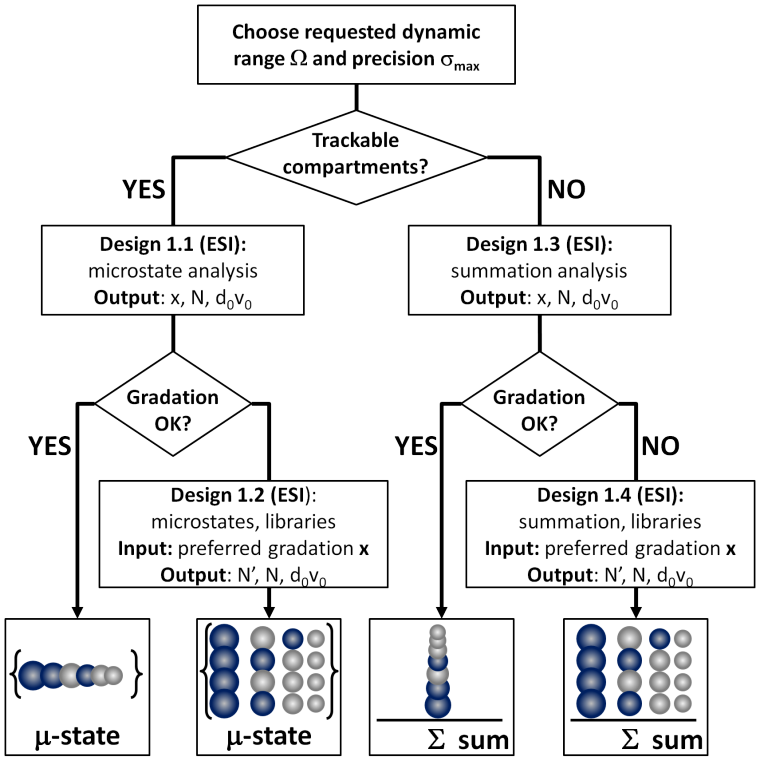


Figure 7.13: Block diagram for effective design of the assay depending on the method of analysis and technical means of partitioning of the sample. If it is possible to construct a device that can track all the compartments during amplification and readout (assays designed for practical purposes can comprise up to hundreds of compartments), the optimal assay would be designed using instructions for identifiable compartments. Otherwise, the user should use the protocols that describe the assays comprising non-tractable compartments . Although assays with the smallest number of compartments are designed using instructions for fine gradation of compartments, they may be challenging in the execution of the fine gradation of volume or dilution factors of compartments (the common ratio of the sequence of volume or dilution of compartments is close to unity). To overcome this difficulty, the user can set the finest (optimum) gradation achievable using available equipment (i.e. set the maximum value of the common ratio of the sequence of volume or dilution of compartments) and enhance the precision of the assessment by making sets of copies of each compartment (libraries).

The algorithms that we described in this chapter provide the formula for the partitioning of the sample for assays that assess the concentration within a requested dynamic range with a requested precision. The digital assays designed using the algorithms can be divided into four groups by different design protocols, depending on the method of analysis (identifiable and non-identifiable compartments) and technical possibilities for the preparing of the partitions (i.e.

limitation in the minimum dilution factor). The procedure of choosing the right design protocol is based on a set of decisions visualized in Fig. 7.13 and concerns:

- tracking of the compartments during amplification (tracking of the compartments increase the information gain from the outcome of the assay, although it may be experimentally challenging), and
- gradation of the modulation factor (the optimum value determined using the formulas for assay design might be too close to unity and therefore difficult to obtain; hence, one can input the limitations on the gradation of modulation factors, which provides design of an assay with more compartments, but easier to realize in practical conditions.

7.6 Experimental and numerical verification of Rational Digital algorithm

7.6.1 Numerical verification of the algorithms.

We verified the design of digital assays presented here whether it meets the requirements concerning the precision of the assessment and dynamic range and compared to state-of-art multivolume digital assays [5–7]. In the Fig. 7.14b-d the performance of a set of assays is shown. The assays are designed to cover the dynamic range of $\Omega = 7650$, providing the precision of the assessment $\sigma_{max} = 17\%$, similarly to the state-of-art multivolume design described by Kreutz [5] (Fig. 7.14a). The design described here offers more uniform precision in the whole dynamic range and uses fewer compartments comparing to Kreutz's solution ($N = 348$,360 or 350, depending on the design). Using the same number of compartments, it is possible to design an assay providing the precision improved to $\sigma_{max} = 11\%$ (Fig. 7.14e-f).

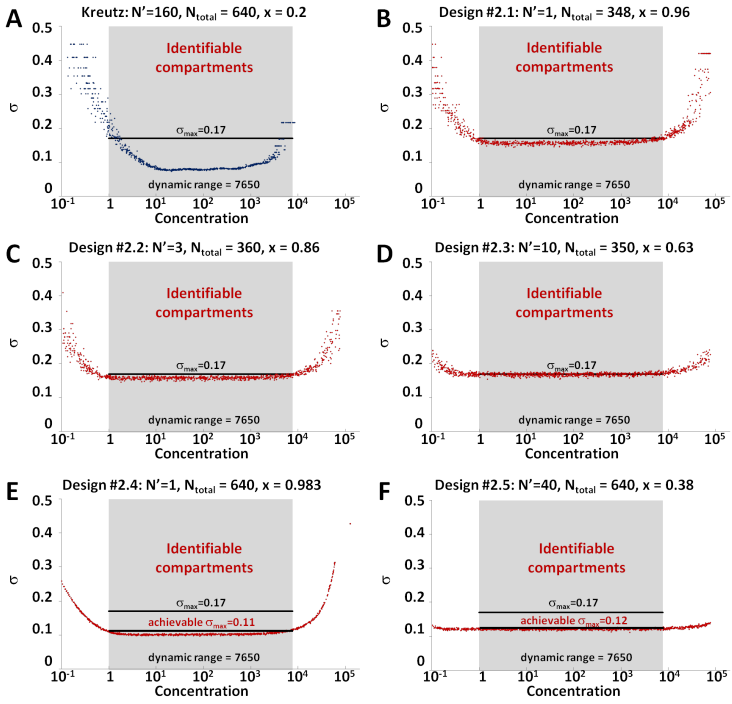


Figure 7.14: Comparison of the performance of assays designed using the described protocol and state-of-art method (Kreutz et al. Slip Chip design [5–7]). The comparison was done via grand canonical Monte Carlo simulations. The SlipChip assay [5] slightly overestimates the dynamic range and underestimates the precision in a fragment of the dynamic range. The distribution of precision is not flat (it is equal to 8% in the middle of the declared dynamic range and more than twice that at both ends). Our model allows designing an assay that covers the same declared dynamic range ($\Omega = 7650$) and precision ($\sigma_{max} = 17\%$) using nearly twice fewer compartments (350) with a flat profile of precision, exactly as requested. The design of SlipChip [5] can also be improved by using the same number of 640 compartments and providing a uniformly precise estimate ($\sigma_{max} = 11\%$) over the entire requested dynamic range.

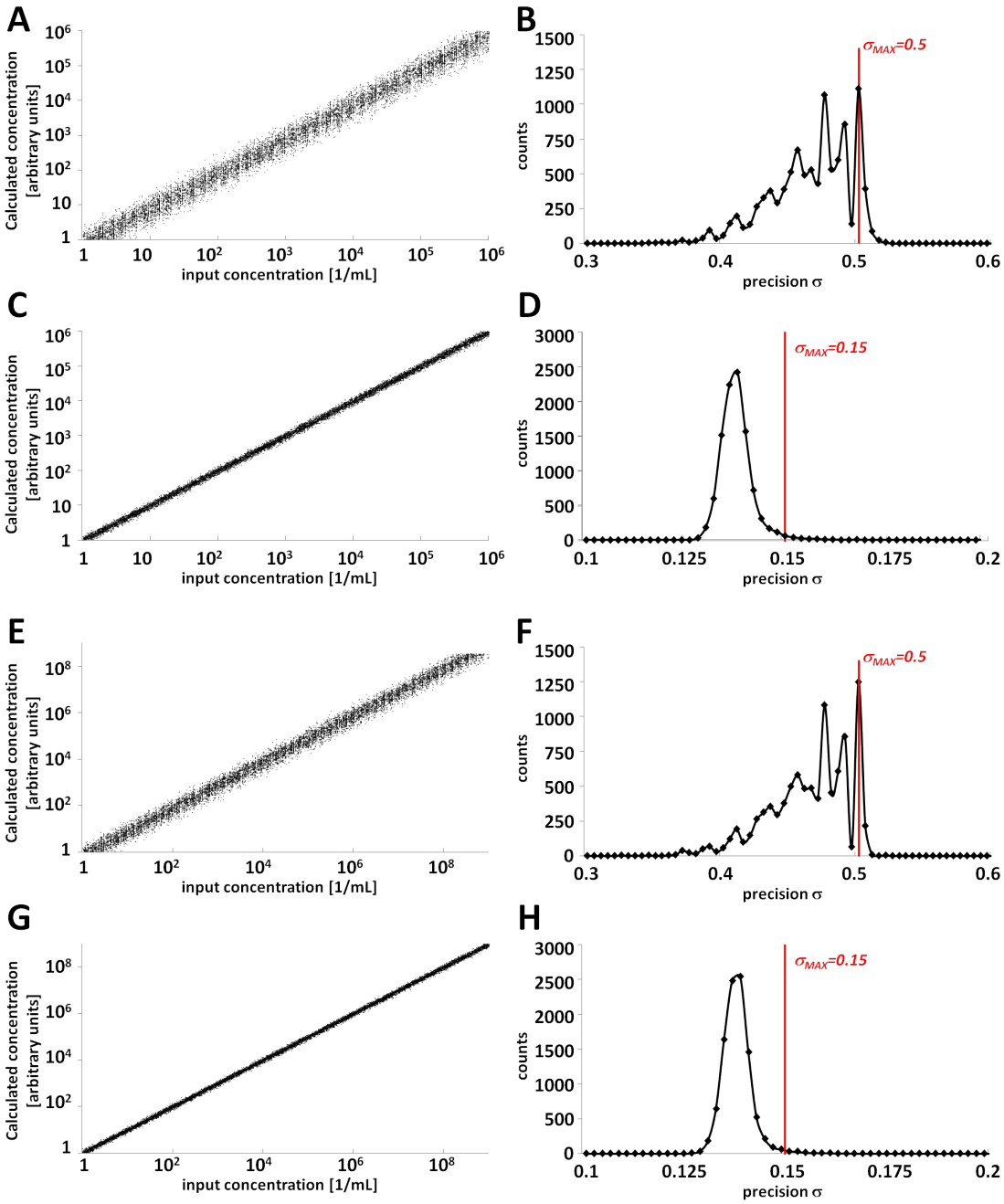


Figure 7.15: Numerical verification of the algorithm. (a, b) an assay built with $N = 45$ compartments with $x = 0.68$ and declared $\Omega = 10^6$ and $\sigma_{max} = 50\%$, (c, d) $N = 297$, $x = 0.95$, $\Omega = 10^6$ and $\sigma_{max} = 15\%$, (e, f) $N = 65$, $x = 0.68$, $\Omega = 10^9$ and $_{max} = 50\%$, (g, h) $N = 422$, $x = 0.95$, $\Omega = 10^9$ and $\sigma_{max} = 15\%$. For each assay 10,000 trials were done with random input concentration $C_{input} \in (C^-, C^+)$. The relation between the input concentration and concentration calculated from the assay response is shown (a, c, e, g) as the distribution of the precision of the concentration estimation (b, d, f, h).

7.6.2 Experimental verification of the algorithms.

We also verified experimentally the performance of the assays(Fig. 7.16 and Fig. 7.17). A set of $N = 16$, 32, 48 and 96 compartment assays were designed. Each of the assays was meant to provide the assessment within the dynamic range $\Omega = 10^4$. The precision σ_{max} varied from 70% to 30%. The results were compared with Monte Carlo simulations. The experimental results, as well as numerical verification, prove that the performance of the assays agrees with analytical predictions.

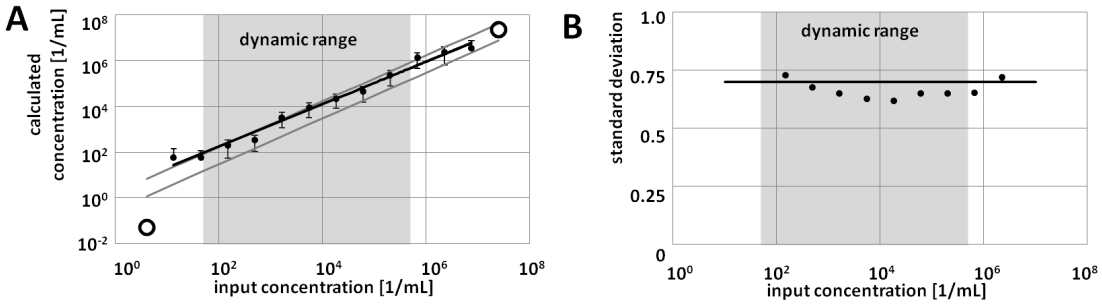


Figure 7.16: Experimental verification of a 16-compartment digital assay. (a) The estimated concentration of DNA as a function of known input concentration. Each data point is averaged over 12 runs of the assay and error bars indicate the standard deviation of the estimates. The solid line is a power fit to the results. The gray lines indicate the region of expected outputs of the digital assay, given by nominal input concentration plus/minus one standard deviation. The green region marks the dynamic range. The circles mark the output of the digital assay when all the signals were either positive or negative indicating the concentration that with 95% confidence is higher (lower) than C_{input}. (b) The standard deviation of the estimates with the solid line showing the expected value within the dynamic range. The assays are shown in Fig. 7.16 and in Fig. 7.17 are designed to cover the same dynamic range with a different precision of the assessment. This feature is accomplished by changing (i.e. increasing, for better precision) the number of copies of each compartment. Fig. 7.16, shows the performance of an assay that has only one copy of each compartment, while Fig. 7.17 shows the performance of the assays comprising 2, 3 or 6 copies of each compartment. The common ratio of the sequence of products of volume and dilution ratios are the same.

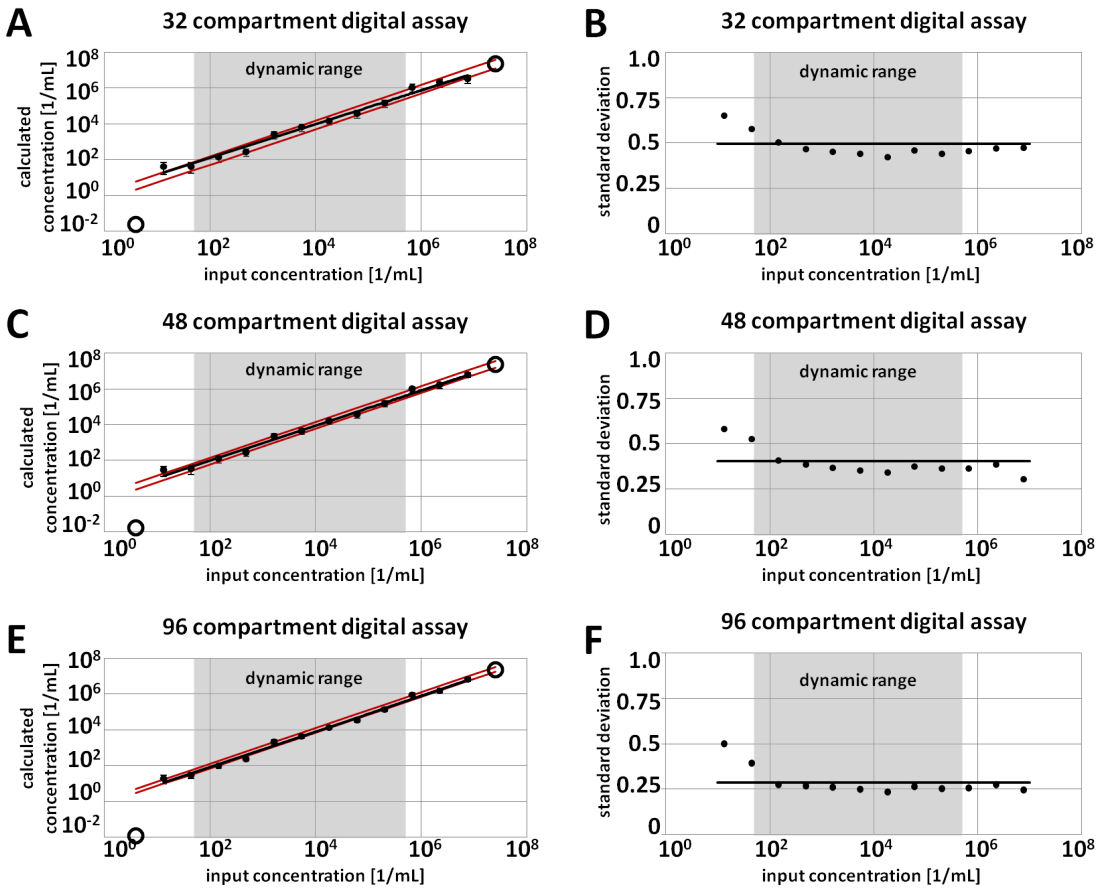


Figure 7.17: Experimental verification of the algorithm. (a, c, e) The estimated concentration of DNA as a function of known input concentration used in the experiments. Each data point is an average of the number of independent runs of the same assay design for each concentration value (e.g. for 32-partition assays, the data point is an average over 6 results, for 48-partitions assays an average over 4 results, and an average over 2 results for the 96-partition assays). Error bars indicate the standard deviation of the estimate, calculated over the set of results, drawn as one standard deviation away from the point (i.e. corresponding to the 68% confidence interval). The solid black line is a power fit to the averaged results. The gray lines indicate the region of expected outputs of the digital assay, given by nominal input concentration plus/minus one standard deviation expected from the assay. The green-shaded region marks the designed dynamic range of the assay. Black circles mark the output of the digital assay for the situation when all the compartment yield positive signal (left side of the graph) or all the compartments yield negative signal. These data points indicate that with 95% probability the input concentration was higher (for all positive) or lower (for all negative) than the value marked. For these cases, the standard deviation is not calculated. (b, d, f) Standard deviation of the estimates from our algorithms as a function of the input concentration, while the black solid line marks the expected standard deviation for a given digital assay, within the dynamic range (again, marked with the gray shade).

7.7 Summary

In this chapter, we presented the algorithm for designing optimized multivolume digital assays. They require orders of magnitude fewer partitions to cover the same dynamic range compared to single-volume digital assays and allow to tune independently key parameters of the assessment. It benefits from the optimized use of information provided by the analysis of all the possible stochastic realizations of digital assays.

Moreover, it is possible to design assay that provides different precision of the assessment for different sub-ranges of the dynamic range, which can be useful in cases where there is no need for precise assessment of very high and very low concentrations, but the exact estimates are clinically informative for intermediate concentrations.

The technical requirements of the digital assays are minimized thanks to a drastic reduction of the number of partitions needed for the assessment. Moreover, the design protocol can be tailored depending on the method of analysis and technical means of partitioning of the sample. Such features are important for the popularization of digital techniques for the reference-free, absolute, quantitation of DNA and increase the availability of digital PCR and digital immunoassays.

This is also the reason why digital methods may be used more frequently in the quantitative identification of viral and microbial pathogens in the Point-of-Care applications.

Chapter 8

Synergistic analogue-digital design of quantitation assays

8.1 Introduction

The existing methods of quantitative PCR present complementary features, that have not yet been combined. The analog real-time PCR techniques offer relatively facile handling of samples. The sample is typically mixed with the PCR kit into one well. The increase of the real-time fluorescent signal is then compared with a calibration curve obtained from a reference sample. The method offers high information gain via analog resolution of the signal yet relies heavily on calibration - an effect that may introduce significant errors in accuracy of the result. The digital assays alleviate the need for calibration, yet due to the intrinsically low (binary) information content of the signal from any single partition, the classic digital assays that use identical compartments require large numbers of them. While the number of required partitions of the sample in a digital assay can be minimized while using only binary (yes/no) signals, further reduction of the number of partitions of the sample or, alternatively, an enhancement of the precision of the assessment, can be obtained via a synergistic analysis of both the digital and analogue signals recorded on a real-time PCR instrument.

In this chapter, we present how the combination of the digital and analog information provides for absolute quantitation with adjustable resolution. Also, we give the description of the procedure and derive a prescription for designing synergistic digital-analogue PCR assays and for analyzing the results.

The analog-digital method that can be executed on standard real-time PCR devices is presented. The method synergistically combines the advantages of the real-time and of the digital PCR and bypasses their drawbacks. The protocols described here provide for a small number of partitions in comparison to digital techniques and for absolute quantitation, therefore they are not influenced by the change of details of reaction procedure (i.e. change of the buffer in the isolation protocol or aging of reagents). In the synergistic PCR scheme, the digital signals auto-calibrate the measurement while the analog measurements refine the estimate of concentration. There is also listed a practical recipe how to design the assay for the required dynamic range and precision of the estimate, and how to analyze the signals to extract the estimate of concentration.

It is demonstrated that it is possible to:

- execute digital assays on standard real-time PCR devices, and to
- synergistically extract the advantages of the analog and digital schemes.

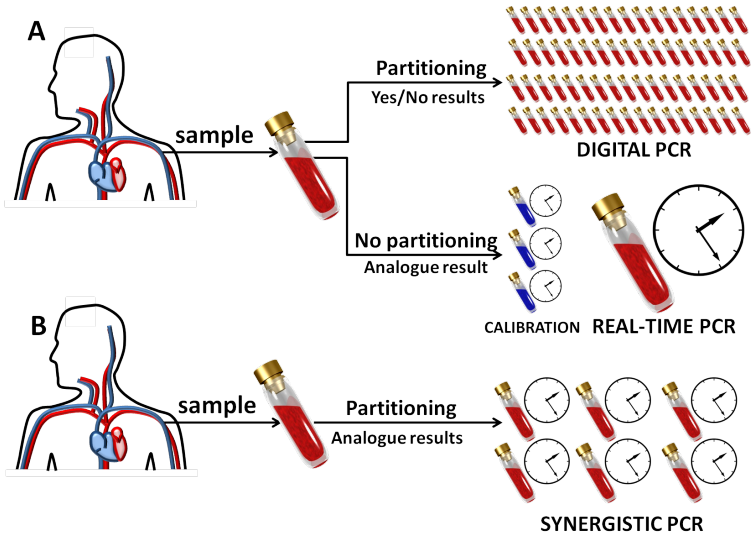


Figure 8.1: A scheme of the analytical procedures. (a) Digital and analog real-time PCR assays. Digital assays require the division of the sample into a large number of partitions, even thousands to millions, but they provide assessment without calibration. Real-time assays require only one partition but need calibration using standardized samples to give an absolute result. (b) Synergistic assays described in this paper require simplified partitioning of the sample - only tens of non-identical partitions are needed - and provide absolute assessment without calibration.

In previous chapters, we described and thoroughly investigated an alternative technique called digital PCR. Digital assays offer an absolute and calibration-free assessment of the concentration. However, they require complicated partitioning of the sample into a large number of partitions. However, in Chapter 7, a method for optimization of information gain from digital signals and

for an optimum (rational) design of assays that minimizes the number of compartments required to obtain a given precision over a requested dynamic range was presented. The protocol offers dramatic reduction (by orders of magnitude in comparison to classic digital methods) of the number of partitions. These end-point digital assays could be run on real-time PCR devices that use the strips (32) or (96, 192 or 384) well plates.

8.2 Mathematical routines

The discussion in this chapter focuses on the application of digital methods to the real-time PCR assays. Digital assays provide only a qualitative information from a single test-volume (i.e the yes/no answer, usually called *digital*) on whether at least a threshold amount of particles (threshold concentration) $m \geq m_{tr}$ was present in the said test-volume (which is reflected by a positive signal). To extract this information, the test-volumes are inspected for the presence of signal after a large number of PCR cycles has been performed (i.e. an end-point signal). In the synergistic assays, the iterative measurements of the signal from test-volumes are performed (real-time readout). Similarly to real-time PCR assays, the number of the cycle after which an inspected test-volume yields signal higher than the threshold is collected.

Therefore, in the synergistic scheme, the information from digital and analog readouts are collected and used to improve assessment of the concentration of the analyte.

8.2.1 Analytical model of one compartment of the assay - combination of the real-time and the digital measurement

The analysis of the behavior of synergistic assays starts similarly to the analysis of rational digital assays given in the previous chapter. It is assumed that the sample contains a stable and uniform solution of the molecules of the analyte. Therefore, the expected number of molecules to be found in a single partition depends only on the parameters of the partition: $E(m) = Cdv$, where C is the concentration of the analyte, and v and d are the parameters of the partition: the volume and the dilution, respectively. Because the distribution of the target molecules is not dependent on any other parameters (i.e. time, temperature, the size of target molecules), the probability of finding a given number z of molecules in this partition is given by the Poisson distribution: $p(m) = \frac{e^{-Cdv}Cdv^m}{m!}$. This provides the analytical formulas for the probability density function that a given initial concentration of the analyte has caused the positive signal yielded by the

partition: $\rho_i(C) = 1 - e^{-Cd_iv_i} \sum_{j=0}^{m_{tr}-1} (Cd_iv_i)^j/j!$, while the negative signal yielded is translated into another probability density function $\rho_i(C) = e^{-Cd_iv_i} \sum_{j=0}^{m_{tr}-1} (Cd_iv_i)^j/j!$. If the threshold number of molecules equals one $m_{tr} = 1$, which is usually the case for PCR, the positive signal from a partition can be calculated represented with the function $\rho_i(C) = 1 - e^{-Cd_iv_i}$, while the negative signal is represented by: $\rho_i(C) = e^{-Cd_iv_i}$. It is worth mentioning, that both probability density functions reach the value 1/2 for the concentration that is determined solely by the parameters of the partition. In this case, it equals $C^* = log(2)/(v_id_i)$.

The above-mentioned probability functions determine the conditional probability of obtaining a recorded outcome of an assay (microstate) for a given concentration. However, the aim is to assess the initial concentration given the recorded microstate, therefore the said probabilities have to be inverted.

The inverse probabilities concerning two events can be related using the Bayes' theorem. Here, the input conditional probability is $p(s|C)$, and the expected output is the probability distribution of concentrations C yielding the signal s: $\rho(C|s)$. If s has a discrete character, and C is continuous, the distribution $\rho(C|s)$ is given by:

$$\rho(C|s) = p(s|C) \cdot f(C)/p(s),$$

where $f(C)$ is the distribution of C. Here, this distribution is unknown because it was assumed that no prior knowledge about concentration is given, therefore $f(C)$ is constant in the arbitrary bounds, i.e. $f(C) = 1/C^{\infty}$, for the range $C \in (0, C^{\infty})$. The function $p(s)$ is the total probability of obtaining a signal s for any concentration and can be given as $p(s) = \int_0^{C^{\infty}} p(s|C)f(C)dC$. Finally, since $f(C) = const$, the distribution of concentration $\rho(C|s)$ can be written as:

$$\rho(C|s) = (p(s|C) \cdot f(C))/\int_0^{C^{\infty}} p(s|C)f(C)dC = p(s|C)/\int_0^{C^{\infty}} p(s|C)dC.$$

8.2.2 Analogue signals - difference in time determines the ratios of numbers of molecules

As we mentioned earlier, the assessment in the synergistic scheme requires a real-time measurement of the signal yielded by the test-volume to determine the number of the cycle after which the signal reaches a higher level than a given threshold, which is usually unknown.

The real-time (analog) measurement allows assigning a real number of cycle (ct_i) to the i-th compartment which yielded positive signal. Knowing the cycle numbers of any two compartments allows calculating the ratio of numbers of molecules initially present in the two compartments,

i.e. $m_i/m_j = q^{ct_i - ct_j}$, where m_i and m_j are the initial numbers of molecules of the analyte found in the two compartments, and q is the amplification factor $q = 1 + E$. However, the measurement of cycle numbers provides only the ratio of the numbers of molecules of the analyte, without the absolute quantitation. This is the reason why classic real-time PCR assays require calibration for the assessment.

In the synergistic scheme, this requirement is omitted. Here, the collection of digital signals is used for the initial auto-calibration, and then the assessment of the initial concentration is refined using the cycle numbers ct_i recorded from the positive compartments. The source of the refinement is the change of probability density functions used to interpret the positive signals, using information from real-time (analog) readout. In essence, the hypotheses on the threshold numbers of target molecules in positive compartments are changed, i.e. $m_i > 1$.

Practically, from the set of recorded ct_i values yielded by positive compartments, we choose the highest one (the corresponding compartments will be indexed ω). This compartment initially contained the smallest number of molecules of the analyte.

It is assumed that in the reference (ω) at least a threshold number of molecules of the analyte was present, i.e. $m_\omega \geq m_{tr}$. Therefore, the i-th compartment contained, which yielded real-time signal ct_i, contained at least $m_i = m_\omega q^{ct_\omega - ct_i}$ target molecules instead of $m_i \geq m_{tf}$. Effectively, this is represented by the shift of the sigmoidal probability density functions $\rho_i(C) = 1 - e^{-Cd_iv_i}$ towards the actual concentration. As a result, the precision and accuracy of the assessment is improved thanks to the improved probabilistic interpretation of the result of the assay: $P(s_0, s_1, \ldots, s_{N-1}|C) = \prod_{i=0}^{N-1} \rho(s_i|C)$.

In essence, the methodology that we present here changes the threshold number of molecules of the analyte contained by any of the positive compartments from $m_i \geq 1$ to $m_i \geq A$, where the threshold value $A = q^{\Delta ct_i}$ is determined using the collection of $\{ct_i\}$ readouts (Fig. 8.3a). The new sigmoid functions for different thresholds differ in shape, although this change can be easily calculated analytically. As they are qualitatively similar, the concept of active stripe presented in the previous chapter can be used to derive the formulas for the synergistic assay design. Again, in order to provide the constant precision of the assessment in the whole dynamic range, the products of volume and dilution d_iv_i of the compartments should be arranged in a geometrical sequence, the parameters of which depending solely on the required precision and dynamic range.

8.2.3 Information content of a synergistic signal

The observation that an inspected compartment contained (or not) at least a threshold number of target molecules yields some probabilistic information about the initial concentration C. We can write the probability density function that represents the positive signal recorded as:

$$\rho_i(C|m_i \geq m_{tr}) = 1 - e^{-Cd_iv_i \sum_{j=0}^{m_{tr}-1} (Cd_iv_i)^j/j!}.$$

In the simplest case, i.e. $m_{tr} = 1$:

$$\rho_i(C|m_i \geq m_{tr}) = 1 - e^{-Cd_iv_i}.$$

The observation of a negative signal can be represented with another probability function:

$$\rho_i(C|m_i < m_{tr}) = e^{-Cd_iv_i \sum_{j=0}^{m_{tr}-1} (Cd_iv_i)^j/j!}.$$

If $m_{tr} = 1$, this probability equals $\rho_i(C|m_i < m_{tr}) = e^{-Cd_iv_i}$. As was described qualitatively in previous sections, if the threshold number of molecules increases thanks to the real-time readout, the resulting probability function is shifted towards the actual value of concentration. Also, the characteristic concentrations C^* is shifted (Fig. 8.2a). However, it still depends solely on the parameters of compartments.

The quantitative measure of the information gain assessed with the readout from the inspected compartment (Bernoulli trial) can be determined with tools provided by the Information Theory. One of the measures applicable here is the Shannon Entropy $H = \sum_i p_i log(p_i)$, where i indexes all the possible possible outcomes (in this case, there are only two outcomes: positive and negative; Fig. 8.2b).

Ideally, the digital PCR provides the 100% efficiency of amplification of even very minute concentrations of the analyte in the given compartment. With this assumption, the information gain measured by the Shannon Entropy H reaches its at the characteristic concentration $C^* = log(2)/(v_id_i)$. The information gain is highest for concentrations close to the characteristic concentration, where $p(s = 1|C = C^*) = 1/2$ (Fig. 8.2a), and much lower for 'far' concentrations (i.e. much smaller or much larger).

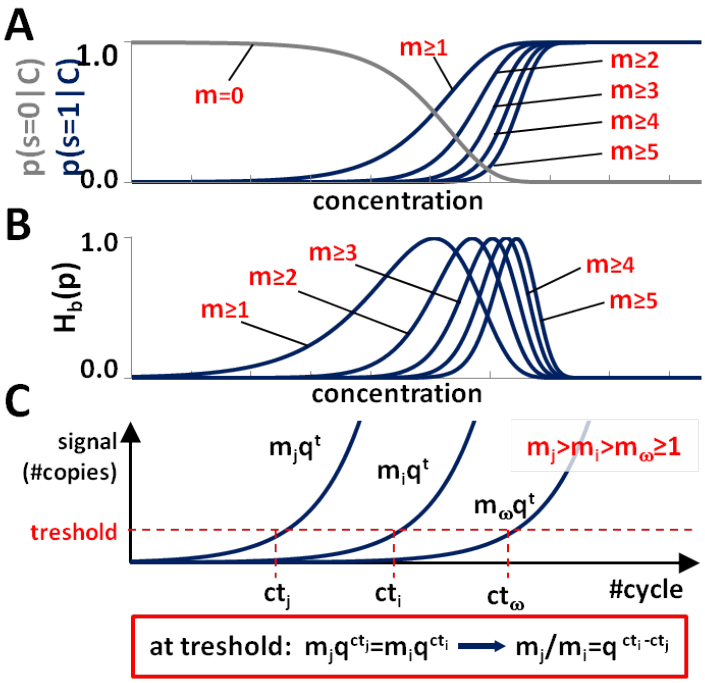


Figure 8.2: (a) Probability $p(s = 0|C)$ (grey line) of not finding any target molecules in a test-volume (negative digital signal $s = 0$) and probabilities $p(s = 1|C)$ (blue lines) of finding at least a threshold number of molecules in a test-volume (positive digital signal $s = 1$). For digital PCR, the threshold number of molecules is typically equal to one ($m \geq 1$). Using synergistically information from digital and analog readout, one can interpret positive signals as the threshold number was higher, which effectively shifts the probability functions towards the real value of initial concentration. Bayesian formalism translates the signal into probabilistic information $\rho(C|s = 0)$ and $\rho(C|s = 1)$ on C, later used for the more precise calculation of the estimate of initial concentration. (b) The Shannon (binary) Entropy function H_b (p) quantifies the information gain from a single Bernoulli trial (i.e. a test with that provides only a positive/negative answer). Each trial provides most information at a specific value of concentration C_i^*, which depends solely on the volume and dilution of a test-volume $d_i v_i$, and a threshold number of molecules m_{tr} (i.e. for $m_{tr} = 1$, $C_i^* = log(2)/(d_i v_i)$. (c) Measurement of the real-time signal. In the real-time PCR, the level of signal (i.e. fluorescence) that is proportional to the number of copies of the amplicon in the compartment is measured at predefined time intervals, preferably at the end of every PCR cycle. The procedure returns an approximation (or estimation) of the real (non-integer) number of the cycle at which the signal from a given compartment crosses the pre-set threshold. Such identified threshold cycle number corresponds to a given, fixed (unknown, but same for all positive compartments) number of copies of the amplicons in that compartment. This allows using simple calculations to determine the ratios of initial numbers of molecules in these compartments as a function of their threshold cycle number ct_i.

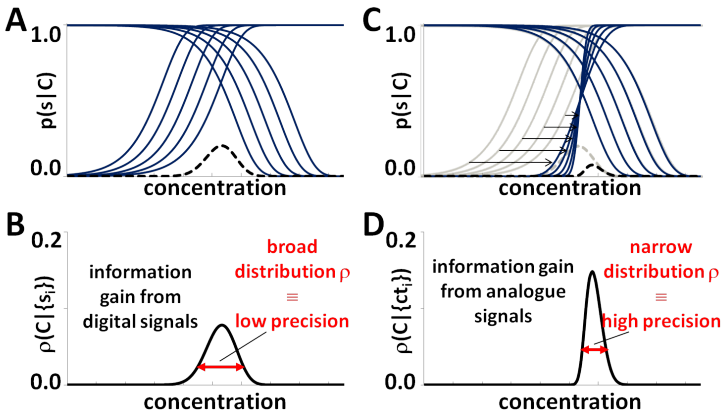


Figure 8.3: (a) The effect of using real-time signal. (a) a set (assay) of compartments with geometrical sequence of modulation factor $d_i v_i$. Larger compartments yield positive digital signals while the other yield negative digital signals. (b) Digital signals from compartments can be used to determine the probability density function of the concentration of the analyte that caused such state of the assay. (c) If the digital measurement is accompanied with analogue measurement, sigmoidal functions of probability for positive compartments can be shifted towards the real value of concentration. Hence, they all contribute to the estimation of initial concentration and provide for higher precision. (d) The new probability density function of concentration can be calculated, which is narrower than the function based solely on digital measurement, and therefore provides higher precision (lower relative standard deviation) of the estimate of concentration.

8.3 Derivation of analytical formulas for assay design

8.3.1 The architecture of the synergistic assay

We start the considerations that lead to the formulation of the design of a synergistic assay from the concept of *active stripe*, which we described in the previous chapter. It determines the number of compartments that actually particitapte in the assessment of the initial concentration and contribute significantly to the information gain. Again, we can observe that the probability density functions related to the outcomes from assay's compartments are centered (i.e. the probability equals 1/2, and the information gain is maximum) at characteristic concentrations that depend on the parameters of compartments (volume and dilution), and thershold number of target molecules. Therefore, only a fraction of the compartments take part in the assessment; the rest is too 'far' to provide significant information (the probability functions given in Fig. 8.3a). This means that highly diluted, or very mall compartments provide no significant information about small concentrations, as they are always negative, and large (or less diluted) comparments

are not useful in the assessment of high concentrations, as they are alwas positive. The fraction of compartments of an assay that actually provide useful information about initial concentration C is called the *active stripe*. The parameters of the *active stripe* have to be determined in order to provide the design of a synergistic assay and test the methodology presented in this chapter.

8.3.2 Derivation of analytical formulas for assay design

The vital point of the derivation of design formulas is finding the gradation of the geometric sequence of mudulation factors (i.e. the products of volume and dilution) of the compartments constituting a synergistic assay. The gradation, i.e. the common ratio x of the geometric sequence should be an unambiguous function of the requested precision of the assessment.

The analysis starts from centering an assay on a known input concentration C_{input}, i.e. seetting one compartment $d_0v_0 = log(2)/C_{input}$ (which is equvalent to setting $C_0^* = C_{input}$). Then, the compartment is assissted with a number ΔN of additional compartment at each side (i.e. adding a number ΔN of larger/less diluted compartments and ΔN smaller/more diluted compartments). Their modulation factors are arranged in a geometric sequence with the common ratio $x = d_{i+1}v_{i+1}/d_iv_i$. The probability that an inspected compartment yields a positive signal depends on the volume and dilution of the said compartment: $p_i(\Delta ct_i|C) = e^{-Cd_iv_i}\sum_{j=0}^{m_i-1}(Cd_iv_i)^j/j!$ (Fig. 8.2). The useful information about the initial concentration is provided by the subset of compartments surrounding the central one that belong to the *active stripe*. For a fixed value of x, adding more compartments to the assay, i.e. increasing the number ΔN allows to improve the quality of the assessment (i.e. lower the standard deviation of the estimate), although this improvement is limited. At some point, adding more and more compartments does not improve the quality of the assessment because the compartments are too small or too large to provide any useful information about C_{input}. Hence, the precision of the assessment $\sigma(C)$ saturates at a limit $lim_{\Delta N\to\infty}\sigma(x)$. This limit depend only on the common ratio x.

We can approximate the relation between the value at which the precision saturates $lim_{\Delta N\to\infty}\sigma(x)$ and the common ratio x in a practically wide range with a function: $lim_{\Delta N\to\infty}\sigma(x) = ax^2 - bx + c$, where $a = 0.439$, $b = 1.3458$ and $c = 0.9960$. For the design formulas, the limit $lim_{\Delta N\to\infty}\sigma(x)$ has to be replaced with the reqired precision of the estimate σ_{max} and the whole equation has to be inverted. This results in the formula that determines the common ratio x of the geometric sequence of compartments as a function of σ_{max} only:

$$x = \alpha\sigma_{max}^2 + \beta\sigma_{max} + \gamma.$$

Then, having established the analytical formula for the common ratio x, what remains to be determined is the number ΔN_x of compartments belonging to the *active stripe*. It equals the lowest integer ΔN for which the standard deviation $\sigma(\Delta N)$ reaches minimum. In other words, the derivative of $\sigma(\Delta N)$ with respect to ΔN equals zero. In this analysis we calculated the value ΔN_x numerically, therefore it was set that the derivative should be no more than 1/1000. Again, the values of ΔN_x can be closely approximated with a simple analytical function: $\Delta N = \lceil \Delta N_x \rceil = \lceil \delta \sigma_{max}^{-2} \epsilon \sigma_{max}^{-1} - \phi \rceil$.

At this point, the analytical formulas allow to design an assay that provides the requested precision of the assessment. The estime $E(C)$ should be provided within the requested dynamic range of concentrations. As the geometric sequence of modulation factors of compartments is self-similar, it is enough to span the previously defined *active stripe* by adding a numer of compartments arranged in the same geometric sequence to cover the requested range of concentrations $C \in (C^-, C^+)$. Also, the 'margins' of the compartments $(d^- v^-, d^+ v^+)$, with $d^{+/-} v^{+/-} = log(2)/C^{+/-}$) outside the dynamic range have to be kept in order to provide the precision of assessment at both ends of the dynamic range.

Finally, in order to design a synergistic assay that provides the assessment of the initial concentration with a requested precision σ_{max} within the requested dynamic range $\Omega = C^+/C^-$, it is enough to follow a simple recipe that uses a set of analytical equations. The requested dynamic range $\Omega = C^+/C^-$ and precision σ_{max} of the estimate are used as an input:

$$x = \alpha \sigma_{max}^2 + \beta \sigma_{max} + \gamma$$

$$\Delta N = \lceil \Delta N_x \rceil = \lceil \delta \sigma_{max}^{-2} \epsilon \sigma_{max}^{-1} - \phi \rceil$$

$$N = 2\Delta N + \lceil log_x(1/\Omega) \rceil$$

$$d_0 v_0 = log(2) x^{-\Delta N} / C^-$$

with α, β, γ, δ, ϵ and ϕ being positive constants:

$\alpha = 0.5540$, $\beta = 1.6504$, $\gamma = 1.1135$, $\delta = 2.0533$, $\epsilon = 1.3220$, and $\phi = 1.9601$.

8.3.3 Designing synergistic PCR assay

The formulas determining the parameters of the sequence of compartments, i.e. x and ΔN, are derieved using close approximations to the numerical data. Therefore, the equations are useful for a practically wide range of requested precision $\sigma_{max} \in (0.06, 1)$. Still, we can improve the

quality of the assessment (i.e. the standard deviation of the extimate can be lowered) by using a number of copies of each compartment. This sets of copies of compartments are called libraries. The recipe for a synergistic library design is given in detail in the following sections and Appendix C.

Finally, the equations we list earlier allows designing a synergistic assay in the following steps:

- determining the parameters of the sequence of compartments: x, N and ΔN for a requested precision σ_{max} of the assessment and dynamic range $\Omega = C^+/C^-$,
- determining the modulation factor of the first compartment in the sequence: d_0v_0, and
- determining a geometric sequence of modulation factors, which comprises N terms $d_iv_i = d_0v_0x^i$ for $i = 0, \ldots, N-1$.

The step-by-step instructions for the design of synergistic assay and the instructions for the analysis of the outcome of such assays are given in the Appendix C.

8.3.4 Detailed decription of precision-varies synergistic assays

Some diagnostic applications of analytical assays require the assessment with varied precision in different concentration subranges. The design of a synergistc assay can be tuned to meet such requirement. In order to provide different precision of the estimate, we have to change the common ratio x of the geometric sequence modulation factors. If the dynamic range of concentration to be assessed (C^-, C^+) is divided into j subranges(C_n^-, C_n^+), the sequence of d_iv_i covering each subrange has a different common ratio x_n, whereas:

For $\sigma_{nmax} < \sigma_{n-1max}$:

$$x_n = \alpha\sigma_{max}^2 + \beta\sigma_{max} + \gamma$$

$$\Delta N_n = \lceil \Delta N_x \rceil = \lceil \delta {\sigma_n}_{max}^{-2} \epsilon {\sigma_n}_{max}^{-1} - \phi \rceil$$

$$N_n = 2\Delta N_n + \lceil log_{x_n}(1/\Omega) \rceil$$

$$d_{0n}v_{0n} = log(2)x_n^{-\Delta N_n}/C^-$$

with α, β, γ, δ, ϵ and ϕ being positive constants:

$\alpha = 0.5540$, $\beta = 1.6504$, $\gamma = 1.1135$, $\delta = 2.0533$, $\epsilon = 1.3220$, and $\phi = 1.9601$.

For $\sigma_{nmax} > \sigma_{n-1max}$:

$$x_n = \alpha\sigma_{max}^2 + \beta\sigma_{max} + \gamma$$

$$\Delta N_n = \lceil \Delta N_x \rceil = \lceil \delta {\sigma_n}_{max}^{-2} \epsilon {\sigma_n}_{max}^{-1} - \phi \rceil$$

$$N_n = \Delta N_n + \Delta N_{n-1} + \lceil log_{x_n}(1/\Omega) \rceil$$

$$d_{0n} v_{0n} = log(2) x_{n-1}^{-\Delta N_{n-1}} / C^-$$

with α, β, γ, δ, ϵ and ϕ being positive constants:

$\alpha = 0.5540$, $\beta = 1.6504$, $\gamma = 1.1135$, $\delta = 2.0533$, $\epsilon = 1.3220$, and $\phi = 1.9601$.

Where C^- is the lower limit of the dynamic range of concentrations,

C^+ is the upper limit of the dynamic range of concentrations,

C_n^- is the lower limit of the n-th subrange,

C_n^+ is the upper limit of the n-th subrange,

σ_n is precision of the estimate $E(C)$ in the n-th subrange,

n is the number of subrange, and

i is the number of compartment of an assay.

8.4 Experimental and numerical verification of the synergistic alorithm

8.4.1 Experimental verification of the synergistic assay

We compared the synergistic assays with purely digital single-volume assays [108] and multi-volume assays described in the previous chapter by means of experimental and numerical data. To test the performance of synergistic multivolume digital assays, we designed and tested a set of $N = 16$ compartments. It is enough to build an assay that should cover the dynamic range $\Omega = 10^4$ and provide the precision of the estimate $\sigma = 70\%$ in the end-point (digital) scheme and $\sigma = 60\%$ in the synergistic scheme, making it a promising tool for Point-of-Care applications. We tested experimentally the performance of the assays for 12 input concentrations within and

outside the dynamic range. Each experiment was repeated 12 times. We proved that the assays offer a linear response, i.e. the linear relationship between the input concentration and the calculated concentration (Fig. 8.4a). Then, we supplemented the assays with sets of copies of each compartment in order to improve the precision of the assessment (Fig. 8.5).

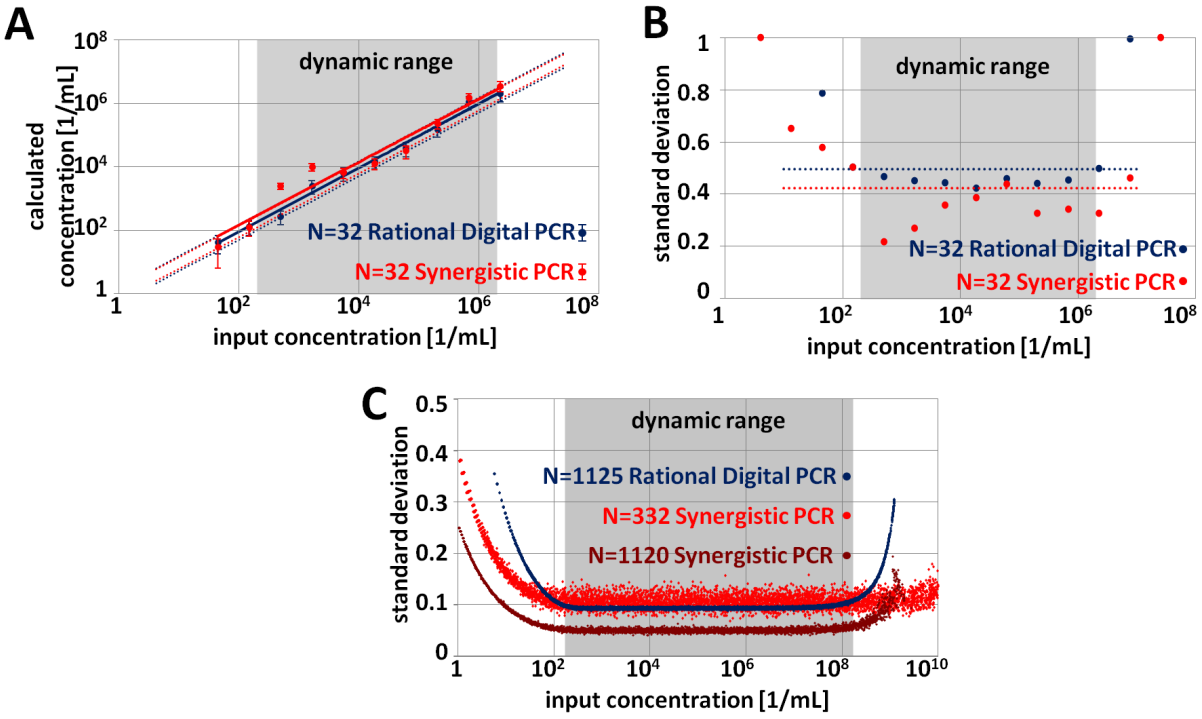


Figure 8.4: Comparison of the synergistic analogue-digital assay and current state-of-art digital methods. (a) Experimental results for 16-compartment multivolume digital (blue points) and synergistic analogue-digital (red points) assays. The value of concentration calculate from the outcome of the assay as a function of a real value of initial concentration is given. Data point shows the averaged results from 12 runs of the assay. The error bars show the standard deviation. The dynamic range of the assays is marked with the gray region. (b) The precision of multivolume digital (blue) and synergistic analogue-digital (red) assays given as the relative standard deviation of the estimates within and outside the dynamic range. The dotted lines show the required precision of the assays. (c) The comparison of the performance of large digital and synergistic assays. To provide 10% precision of the assessment within 6log dynamic range, multivolume digital assays (blue) require 1120 compartments, while using synergistic scheme (red) requires only 332 compartments. If one uses the same number of compartments, the precision of the assessment is improved.

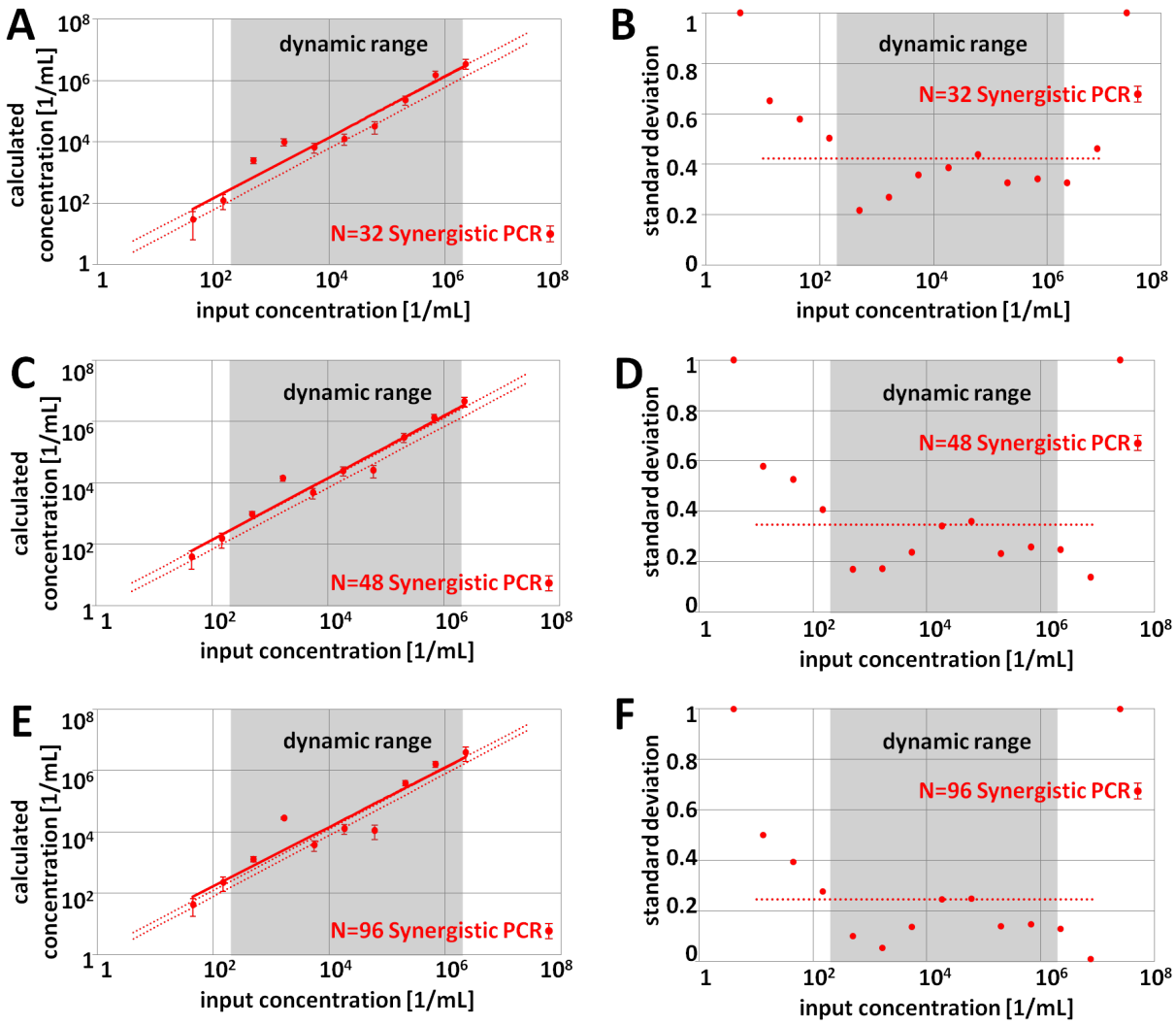


Figure 8.5: Experimental verification of the synergistic algorithm.(a, c, e) The estimated concentration of DNA as a function of known input concentration used in the experiments. Each data point is an average over the number of independent runs of the same assay design for each concentration value (e.g. for 32-partition assays, the data point is an average over 6 results, for 48-partitions assays an average over 4 results, and an average over 2 results for the 96-partition assays). Error bars indicate the standard deviation of the estimate, calculated over the set of results, drawn as one standard deviation away from the point (i.e. corresponding to the 68% confidence interval). The solid red line is a power fit to the averaged results. The dotted red lines indicate the region of expected outputs of the digital assay, given by nominal input concentration plus/minus one standard deviation expected from the assay. The grey-shaded region marks the designed dynamic range of the assay. (b, d, f) Standard deviation of the estimates from our algorithms as a function of the input concentration, while the red dotted line marks the expected standard deviation for a given digital assay, within the dynamic range (again, marked with the gray shade).

8.4.2 Numerical verification of the synergistic assay with Monte-Carlo simulations

We also tested a range of synergistic assays using Monte Carlo simulations. We used the analytical formulas presented earlier to design assays that cover the dynamic range $\Omega = 10^6$, providing the precision of the assessment $\sigma = 10\%$. To meet these requirements, we need a multivolume digital assay that comprises $N = 1125$ compartments or a synergistic assay with only $N = 332$ compartments. Increasing the number of compartments to the said $N = 1125$ would result in a 2-fold improvement in the precision of the assessment. Such reduction simplifies the preparation of the sample and laboratory routines needed to run the assay. The performance of the above-mentioned digital and synergistic assays was tested numerically. The results of this analysis are given in Fig. 8.4c. It is worth mentioning that in order to cover the same dynamic range, a classic single-volume digital assay would require $N = 200,000$ compartments.

As we mentioned earlier, synergistic assays are also a promising tool for Point-of-Care diagnostics. Moreover, they can be easily implemented in stock real-time PCR devices. To test the applicability, the assays that provide the assessment within the dynamic range $\Omega = 10^3$ with precision $\sigma = 40\%$ ($N = 32$, $x = 0.54$), $\Omega = 10^6$ with $\sigma = 52\%$ ($N = 16$, $x = 0.4$) and $\Omega = 10^8$ with $\sigma = 10\%$ ($N = 200$, $x = 0.9$) were designed and tested (Fig. 8.6a; the blue lines visualize the trends approximated from the results of 10,000 Monte Carlo simulations), and they meet the set requirements.

These results are a starting point for a general comparison of the methods described in this work, including classic single-volume and multivolume digital PCR assays. If the number of compartments is fixed, classic digital assays cannot be tuned for precision or the dynamic range of the assessment, which is possible for more flexible multivolume digital and synergistic schemes. Moreover, thanks to the reduction of the number of compartments needed to cover the same dynamic range, the multivolume digital and synergistic assays reduce the cost of assessment.

8.4.3 An example of the use of a synergistic PCR assay: synergistic PCR assay is immune for initial sample buffer composition

The synergistic assays provide a reduction of the number of compartments needed for the assessment compared to digital assays, but compared to the *golden standard* real-time PCR scheme, the partitioning of the sample is still more complicated. The analog techniques, such as real-time

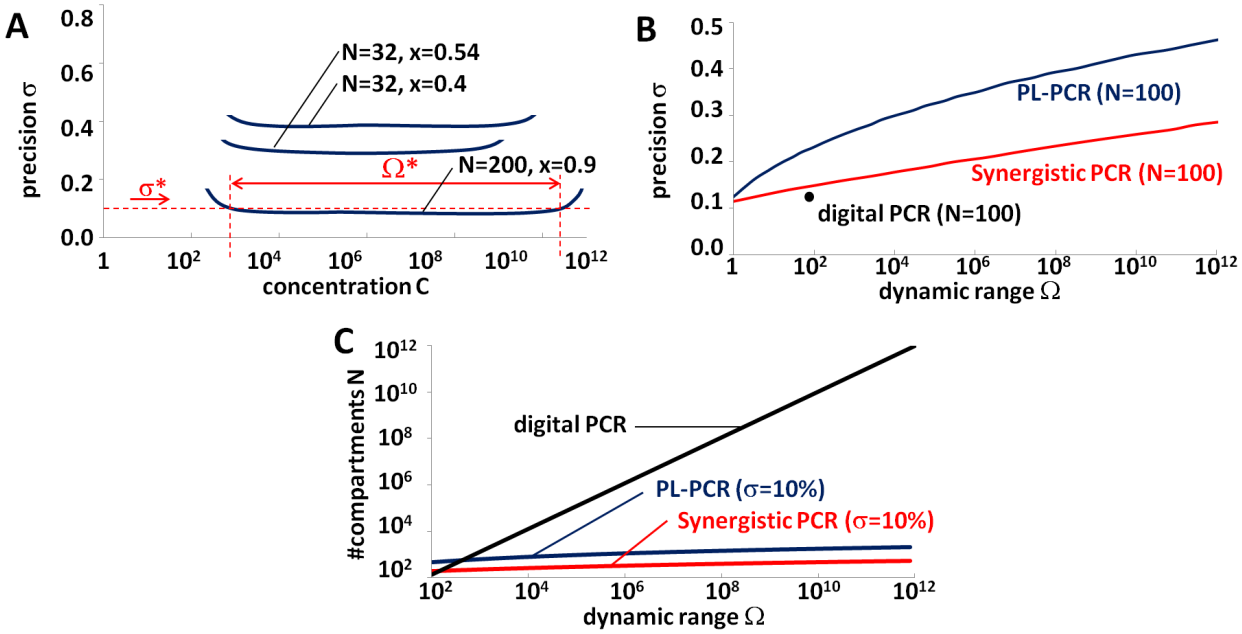


Figure 8.6: Comparison of the synergistic analogue-digital assay and current state-of-art digital methods: single-volume assay and multi-volume rational digital PCR assay. (a) The geometrical sequence of volumes/dilutions of compartments described in the synergistic design provides constant information gain in a wide dynamic range, therefore every synergistic assay offers the constant precision of the assessment. (b) The performance of the 100-compartment assays. Classic digital assay (black point) offers only one value of precision and dynamic range for a given number of compartments and cannot be tuned. On the other hand, the multi-volume (rational digital PCR; blue line) design allows to 'trade' the precision of the assessment for the dynamic range and therefore is more flexible. Synergistic design (red line) offers the same flexibility, but thanks to analog readout, provides a better precision of the assessment. (c) Another advantage of multivolume designs is lower technical requirements in comparison to classic digital assays. In classic digital methods (black line), the number of compartments required for the assessment is directly proportional to the dynamic range, while in digital multivolume (rational digital PCR; blue line) and synergistic (red line) designs this number is proportional to the logarithm of the dynamic range (therefore, it is lowered by orders of magnitude).

PCR, require only one compartment for the assessment. However, the result has to be compared with the calibration curve, which needs additional 15-25 experiments and use of standardized samples. Here, the synergistic assays gain advantage thanks to the absolute character of the measurement. Moreover, the real-time assays are optimized for a set of reagents and are sensitive to even minor changes in the preparation procedure and prone to errors. The synergistic assays should be insensitive to such problems.

In the real-time scheme, even changing the buffer used for the isolation of DNA influences the

reaction which then produces highly scattered results. To quantify this effect, we run four real-time assays in order to determine the initial concentration. The only difference between the reactions were the samples, which were prepared by diluting reference DNA in various elution buffers: water, AE elution buffer QIAamp DNA Mini Kit [122], MBL5 NucleoMag Blood 200uL [123] and MagJET Whole Blood Genomic DNA Kit [124]. The samples contained 25,000 DNA copies, but the real-time assays provided highly scattered results, ranging from 24 967 [1/mL] for the MagJET buffer, 29 684 [1/mL] for the MBL5 buffer, up to 79 368 [1/mL] for the AE buffer and 96 514 [1/mL] for water. Hence, the change of readout was almost 400%.

The spread of the results is caused by PCR inhibitors present in the buffers containing DNA. For the analysis, the standardized samples were diluted 100 times. Then, if the water was used to dilute the sample, the (Fig. 8.7a, blue point), the inhibitors were diluted up to the point they were negligible, which is observed as a higher readout. The inhibiting agent was the ethylenediaminetetraacetic acid (EDTA) [150, 151]. It was added in the buffer in order to reduce the degradation od DNA by neutralizing DNases. It also reduces the amount of free Mg^{2+} ions, affecting the operation of primers.

Synergistic assays seem to this problem thanks to the absolute character of the calibration-free assessment. In order to prove it, a synergistic assay was used to determine the concentration in the samples diluted with buffers that earlier yielded the biggest spread of outputs: water and MagJET buffer (Fig. 8.7b). We used a set of assays comprising $N = 16$ compartments. Each experiment was repeated 12 times, making a total of 24 synergistic PCR runs. The values of initial concentration assessed by a synergistic assay for both samples are similar and close to the actual value of concentration: for the samples diluted with water, the calculated concentration equaled 23 355 copies/mL, which is 6.6% off the reference, while for samples diluted with the MagJET buffer the calculated concentration was 25 163 copies/mL (0.65% off the reference).

8.5 Summary

Analytical assays benefiting from a combination of digital and real-time measurements may become an attractive solution for DNA/RNA diagnostics. They provide an absolute, calibration-free assessment, but allow a simplified sample preparation (reduced number of compartments), minimizing the technical requirements for running quantitative assays.

The synergistic design that we described in this chapter proved to be reliable and insensitive

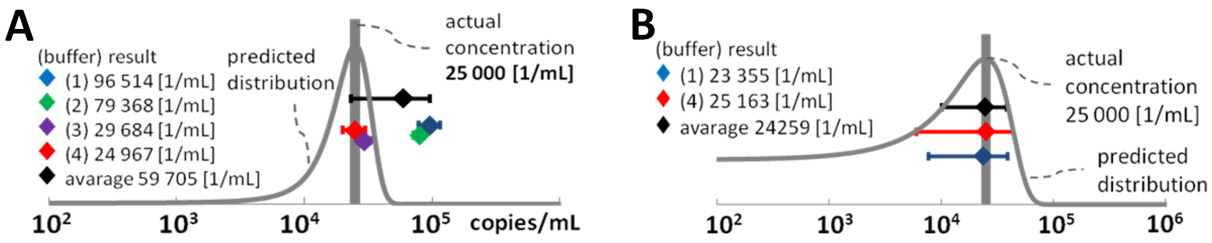


Figure 8.7: Experimental verification of a 16-compartment digital assay with comparison to the performance of real-time PCR assays. (a) The graph shows the result of 24 runs of the synergistic assay, each on 16 partitions of the amplification mix. Two series of 12 assays were conducted on the two elution buffers (1 and 4) that provided the largest difference in the result of the conventional real-time PCR analysis. The gray line shows the expected distribution of results from the synergistic assay used in the experiment, which should provide 30% precision of assessment. This distribution was also verified using 10,000 Monte-Carlo simulations. (b) Same amount of reference DNA suspended in different elution buffers and quantified with conventional real-time PCR and with the synergistic PCR algorithm. Tests performed on Applied Biosystems 7500 Fast RT System on the IVD Cytomegalovirus PCR kit (GeneProof) according to the prescription. Elution buffers: (1) water, (2) AE elution buffer QIAamp DNA Mini Kit (Quiagen), (3) MBL5 NucleoMag Blood 200uL (MACHEREY-NAGEL), (4) MagJET Whole Blood Genomic DNA Kit (Thermo Scientific). The gray line shows the expected distribution of results for real-time assay.

to minor changes in the reaction procedure. Moreover, its flexibility allows tuning for specific requirements – dynamic range and the precision of the assessment – of various scientific and practical applications, including Point-of-Care diagnostics. Importantly, thanks to the reduced number of compartments needed for the assessment, the synergistic assays can be readily run in standard real-time PCR devices, in a standard well plate format. Moreover, a standard 96-well plate can be utilized to run multiple synergistic assays , including Point-of-Care applications. For example, it allows to run:

- 3 assays that provide the assessment within the dynamic range $\Omega = 10^6$ and with precision $\sigma = 50\%$, or
- 4 assays providing $\Omega = 10^4$ and $\sigma = 55\%$, or
- 6 assays providing $\Omega = 10^4$ and $\sigma = 60\%$.

Chapter 9

Discussion

9.1 Analytical description of digital assays

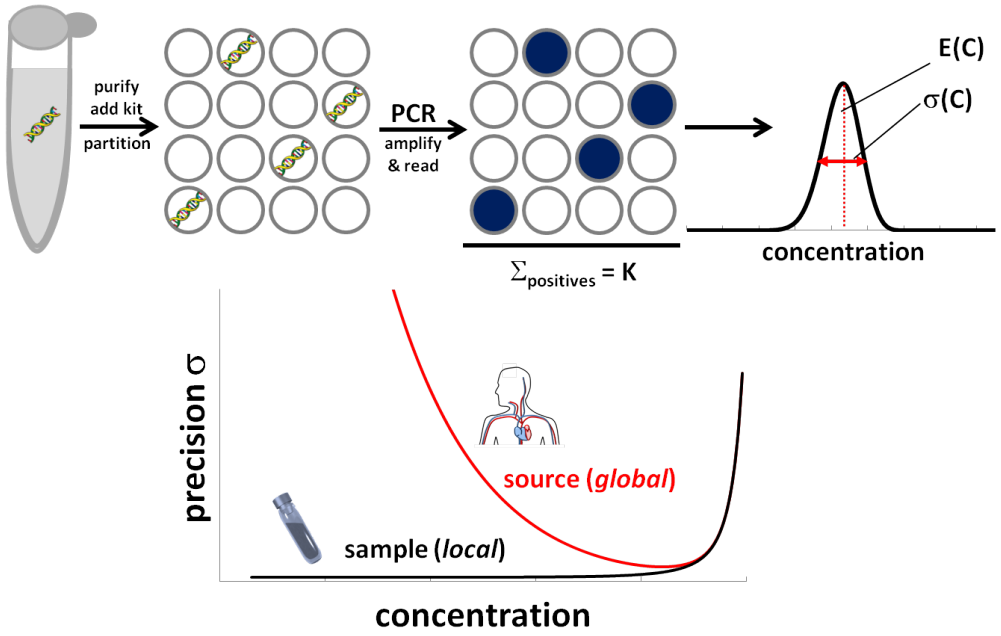


Figure 9.1: A schematic of a analytical procedure with a classic digital assay.

The proper mathematical description and interpretation of the outcome (i.e. the ratio of positive compartments) of an assay is vital for the reliable assessment of the initial concentration of the analyte. In the Chapter 5 we presented the two approaches to the analysis, concerning the assessment of local concentration in the sample (*dependent* approach) and global concentration in the source of the sample(*independent* approach). The two methods differ in the mathematical treatment of the signals yielded by the compartments of an assay and relations between the said

signals. This analysis can be used for the interpretation of the results obtained from state-of-art digital methods and was a starting point for the optimization of quantitation assays described in this work.

9.2 Optimized digital assays

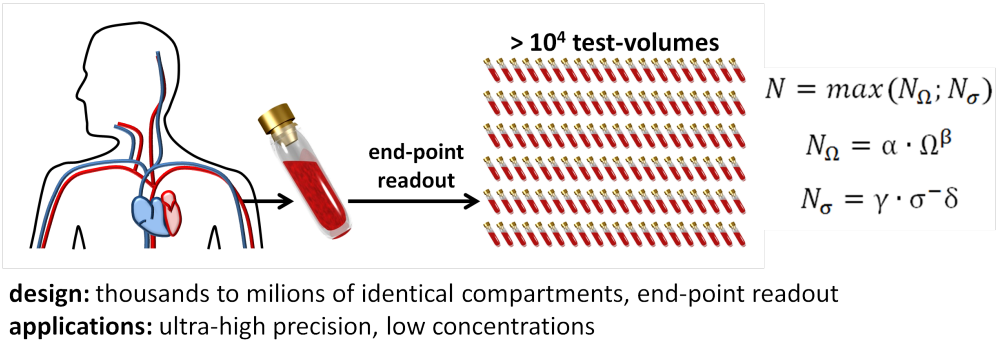


Figure 9.2: A schematic of an optimized classic digital assay.

We used the analysis of the response of single-volume digital assays for the derivation of new analytical formulas that allow retrieving from the outcome of a digital assay the estimate of initial concentration of the analyte and the precision of this estimate, which previously needed time-consuming calculations by means Monte Carlo simulations.

This, in turn, allowed the derivation of analytical formulas that bind the most important characteristics of the assessment – dynamic range and precision – with the size (i.e. the number of compartments) of a digital assay, which was used to design tailor analytical tests for various applications. The derieved the formulas for assay design both for *dependent* and *independent* analysis.

9.3 Multivolume digital assays

The derivation formulas for the design of single-volume digital assays leads to the optimization of the assessment by combining many them in a multi-volume scheme. The subsequent assays differ in volume or dilution, which provides a simple widening of the dynamic range, which is a logarithmic function of the total number of compartments. It is also possible to tune multi-volume assays to meet the specifications needed (dynamic range and precision). This design lowers the technical requirements for running digital assays because it allows to cover wider

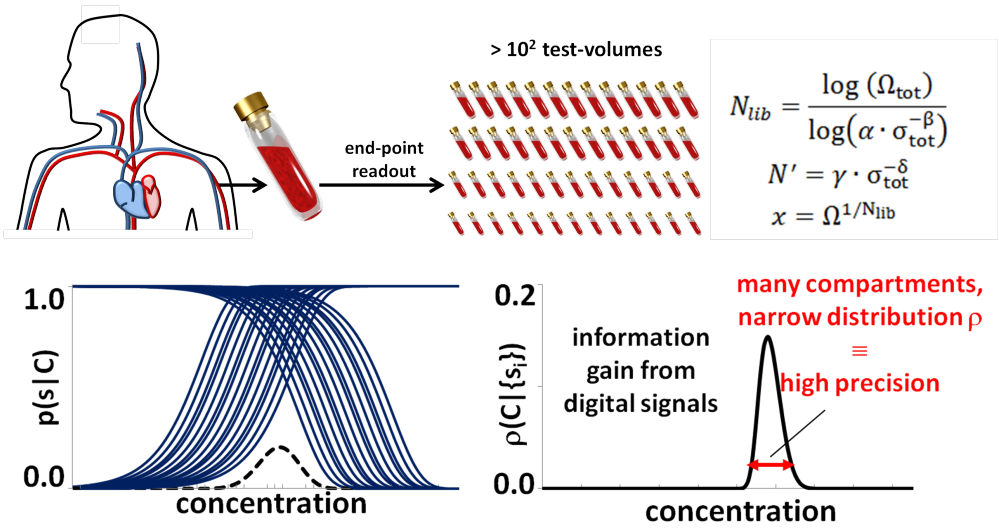


Figure 9.3: A schematic of an optimized multivolume digital assay.

dynamic range with fewer compartments, therefore it is an alternative to currently available systems, where the dynamic range widened by dividing the sample into extremely high numbers of compartments.

In the multi-volume scheme that we described, the sets of compartments (subassays) are relatively large (usually $N > 100$), but the gradation of the subsequent subassays (i.e. the ratio of the volume or dilution) is very coarse, making them easier to prepare. They could be recommended for precise assessments ($\sigma < 10\%$), similarly to classic single-volume digital assays.

9.4 Rational digital assays

In the next chapter, concerning multivolume digital assays, where all the compartments differ in volume or dilution, we showed how to extract all the information from the realization of the result of a digital assay. The analysis of the information gain leads to the further reduction of the number of compartments required to the assessment, compared to the multi-volume and multi-dilution systems. Compared to the single-volume scheme, Rational Digital algorithm offers digital assays with the number of compartments lowered by orders of magnitude. The algorithms that we derived in Chapter 7 allowed designing assays with independently tuned dynamic range and precision of the assessment (the range of acceptable inputs was large compared to multi-volume scheme from Chapter 6, especially for moderate and low precision).

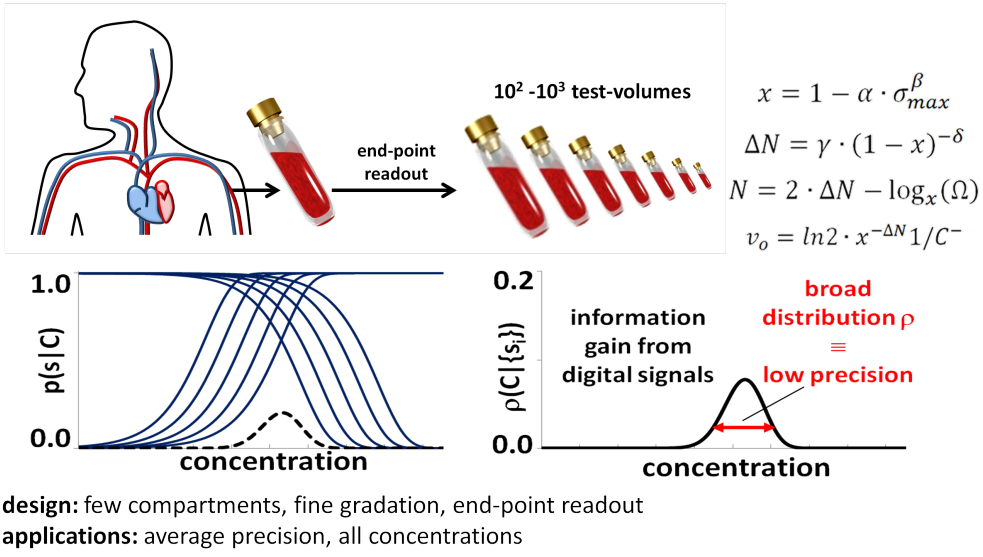


Figure 9.4: A rational design of a digital assay: the compartments are arranged in a geometric sequence of volumes and/or dilutions.

Moreover, it is possible to design assays that provide different precision of the assessment in various subranges of the dynamic range.

Most importantly, the dramatic reduction of the number of compartments allows running digital assays on standard multiwell plates in stock PCR machines, widening the field for practical applications of this technology and adds to the popularization of digital techniques.

9.5 Synergistic analogue-digital assays

The possibility of running digital assays in stock PCR creates opportunity: the use of real-time readout to increase the information gain from compartments and enhance the quality of the assessment. We showed that the use synergistic combination of digital and analog techniques is beneficial and in further perspective may bring a new standard in quantitative assaying. Using information from digital and analog (real-time) signals, it is possible to run quantitative assays that provide an absolute and precise measurement. Also, the technical requirements concerning the partitioning of the sample minimized thanks the reduction of the number of compartments.

The algorithms presented in Chapter 8 also allows to tune independently the dynamic range and precision of the assay, and design synergistic assays that provide different precision of the assessment in various subranges of the dynamic range.

As synergistic PCR assays are run on standard real-time PCR devices, they can find use in

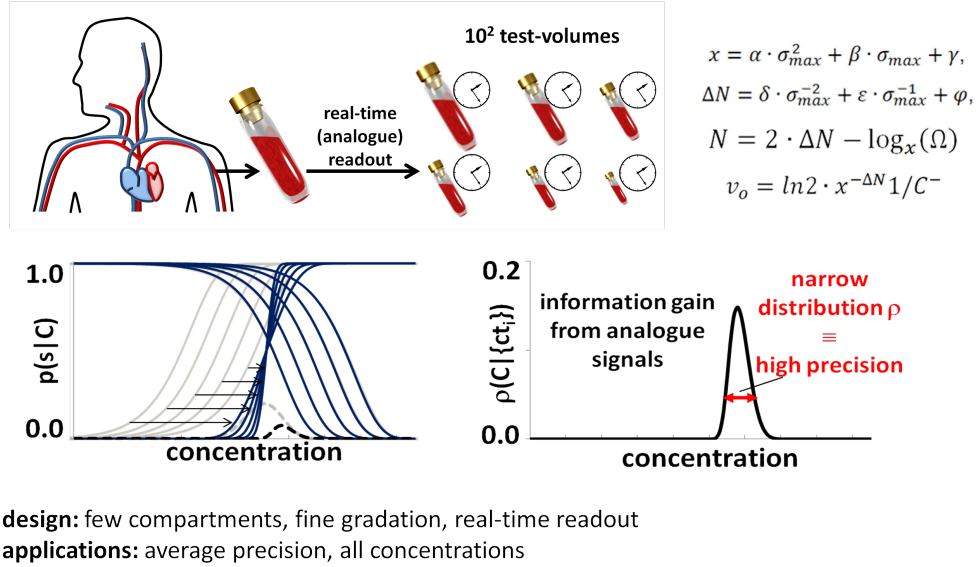


Figure 9.5: A synergisitc analogue-digital design of an assay: the compartments are arranged in a geometric sequence of volumes and/or dilutions. The increas of the signal yielded by each of the compartments is traced.

the assessment of the samples with high accuracy, where the proper calibration is difficult or impossible to achieve.

Looking further, the combination of the methodology of synergistic assays with simplified detection schemes and amplification methods would allow introducing complete system offering precise and absolute quantitation within minutes from the collecting of the sample for Point-of-Care applications, or high-throughput systems [152–155] for the analysis of a large number of samples.

9.6 Practical applications of the algorithms

The dramatic reduction of the number of compartments required to assess the concentration provided by the algorithms presented in this work and the flexibility of the multivolume schemes allows designing digital and synergistic PCR assays that can be used in various practical applications.

9.6.1 Point-of-Care applications

For example, we can use the algorithms to design assays for Point-of-Care applications [156–160] that can be run in stock thermocyclers and PCR devices in a standard 96-well plate format. This allows to run tests that provide the assessment of the DNA concentration:

- with the precision $\sigma = 28\%$ within the dynamic range $\Omega = 10^6$ (*synergistic* algorithm), and
- with the precision $\sigma = 36\%$ within the dynamic range $\Omega = 10^6$ (*rational digital* algorithm).

9.6.2 Multiple assays on a single 96-well plate

If a narrower dynamic range is required, or the confidence intervals of the assessment can be loosened, it is possible to run many multiple independent assays in the same 96-well plate format, lowering the cost of a single assessment. The following sets can be run using 96 PCR reactions:

- **6x** *rational digital* assay which provides the precision of the assessment $\sigma = 70\%$ within the dynamic range $\Omega = 10^4$,
- **4x** *rational digital* assay which provides the precision of the assessment $\sigma = 60\%$ within the dynamic range $\Omega = 10^4$,
- **2x** *rational digital* assay which provides the precision of the assessment $\sigma = 50\%$ within the dynamic range $\Omega = 10^6$,
- **3x** *synergistic* assay which provides the precision of the assessment $\sigma = 50\%$ within the dynamic range $\Omega = 10^4$, and
- **2x** *synergistic* assay which provides the precision of the assessment $\sigma = 50\%$ within the dynamic range $\Omega = 10^6$.

9.6.3 Copy Number Variation

As we mentioned in the Chapter 5, one of the key applications of classic digital assays is the testing of Copy Number Variation [132–135]. This requires high precision of the assessment that is unreachable for real-time PCR. However, it can be achieved with *rational digital* and *synergistic* schemes. For example, the *rational digital* scheme provides for the discrimination

between:

- 2 and 3 copies with 90% confidence within the dynamic range $\Omega = 10^6$ using 730 compartments,
- 9 and 10 copies with 90% confidence within the dynamic range $\Omega = 10^6$ using 4700 compartments, and
- 17 and 18 copies with 90% confidence within the dynamic range $\Omega = 10^6$ using 15000 compartments.

Such assays cannot be run using standard well plated and require tailored PCR devices. The sample handling can be preferably done by means of microfluidic techniques, such as serial dilution systems. Also, the analysis can be simplified by using the end-point readout only and skipping the tracking of the compartments. This results in the 2- or 3-fold increase in the number of compartments needed for the assessment. However, such a facilitated design is easier to run and has been already tested for a similar assessment of the number of bacteria [161].

9.7 The possible application of the algorithms for digital counting of bacteria

The quantitative detection of the presence of specific bacterial strains is widely discussed and well established. As an example of such detection system is the *BacChip* system described by Ho et al. [162]. The method is based on using specific growth media for culturing bacteria. If a colony is observed in one medium, the presence of a specific bacterial strain is confirmed. The method is very sensitive; it is possible to observe positive readout from 1 colony-forming unit (CFU) in the sample. The signal measured is the presence of the colony growing in the medium, therefore it is a two-state variable (the presence of the colony is interpreted as a positive signal, otherwise, the negative value is adopted), and the positive signal refers to the presence of *at least* one CFU.

Such method of detection resembles qualitative measurement based on PCR or other amplification reaction. Therefore, it is possible to use – with minor changes – the algorithms described in this work to create assays for the quantitative assessment of the presence of bacteria. The dynamic range and precision of the assessment should be tunable, similarly to PCR tests. Physically, such tests would be executed in a set of chambers containing growth medium and diluted

partition of the sample. The sequence of diluted partitions of the sample could be introduced to chambers using a pipette or a network of microfluidic channels. Then, the chip would be incubated and inspected for the signal readout. The chambers containing colonies would be attributed with positive signal while the rest would be attributed with a negative signal. On the basis of the sequence of signals, the initial number of CFUs in the sample would be assessed using the methodology described in Chapter 7.

The above procedure is general and can be applied for the quantitative assessment of bacteria, fungi, and other microorganisms.

Moreover, one of the algorithms that we described in this work in Chapter 8 benefits from the digital and analog (real-time) readout. This requires the real-time tracking of the growth of signal (or the growth of bacteria colony). Such methods are also well known and were applied in microfluidic systems.

Appendix A

Single-volume digital assays

A.1 Step by step instruction on how to analyze the results of a digital PCR assay

A.1.1 Independent scheme, Bayes formalism

Here, detailed instructions for the qualitative analysis of the readout from a single-volume digital assay, based on the independent random variables, with the use of the Bayes' formalism is given. A positive signal yielded by an inspected compartment of a digital assay means that the said compartment contained at least threshold number of molecules of the analyte. For PCR-based digital methods, it is assumed that the threshold equals $m_{tr} = 1$. Therefore, the presence of a positive signal if an i-th compartment having a volume v can be translated into the following probability density, which is a function of initial concentration in the assay C_A:

$$p_i = p(s_i = 1|C_A) = 1 - e^{-C_A v}.$$

Then, a negative signal yielded by an inspected compartment of the volume v can be interpreted as a situation, when the said i-th partition did not contain any molecules of the analyte. This can be translated into the following probability distribution, being also a function of initial concentration C_A:

$$p_i = p(s_i = 0|C_A) = e^{-C_A v}.$$

The analysis of the outcome of the assay starts with collecting all the digital readouts s_i from all the compartments of the assay. Then, the procedure is as following:

1. Every partition with volume v, which yields a positive signal (i.e. $s_i = 1$) is assigned with a probability density p_i:

$$p_i = p(s_i = 1|C_A) = 1 - e^{-C_A v}$$

2. Analogically, all the negative compartments are assigned with a respective probability density:

$$p_i = p(s_i = 0|C_A) = e^{-C_A v}$$

3. The probability $P(s_i|C_A)$ of recording a given outcome of a digital assay, i.e. the sum K of positive compartments out of the total of N compartments of the assay $K = \sum_{i=0}^{N-1} s_i$, is a combination product of the probability densities assigned to all the compartments:

$$P(\{s_i\}|C_A) = \binom{N}{K}(1 - e^{-C_A v})^K \cdot (e^{-C_A v})^{N-K}$$

4. Then, this product probability is inverted using the Bayesian formalism in order to obtain the distribution $P(C_A|s_i)$ of the initial concentration C_A, given the recorded outcome of an assay (it is assumed that no a priori information, i.e. the distribution $p(C_A)$, is available):

$$P(C_A|\{s_i\}) = P(\{s_i\}|C_A) / \int_0^\infty P(\{s_i\}|C_A) dC_A$$

   for the numerical calculations of the integral, the upper limit can be finite, although it should be reasonably larger (by at least one order of magnitude) than the highest initial concentration to be assessed

5. Having calculated the distribution $P(C_A|s_i)$, the estimate of the initial concentration can be calculated as the expected value of this distribution:

$$E_I(C_A) = \int_0^\infty C_A \cdot (C_A|s_i) dC_A$$

   again, for the numerical calculations of the integral, the upper limit can be finite, although it should be reasonably larger than the highest initial concentration to be assessed

6. Then, the precision of the estimate is given as a relative standard deviation of the distribution $P(C_A|s_i)$:

$$\sigma_I(C_A) = \sqrt{E_I(C_A^2) - (E_I(C_A)^2}/E_I(C_A),$$

   where

$$E_I(C_A^2) = \int_0^\infty C_A^2 \cdot (C_A|s_i) dC_A$$

7. Finally, knowing the estimate of the initial global concentration $E(C_A)$ and the precision $\sigma_I(C_A)$, the estimate of the concentration C in the source can be calculated as:

$$E(C) = \alpha E_I(C_A)$$

$$\sigma(C) = \sigma_I(C_A),$$

where a factor α is determined by the efficiency of the isolation protocol $\eta = M/M_S$ and the difference between the volume of the sample and the volume of the assay: $\alpha = (V_A/V_S)/\eta$.

A.1.2 Dependent scheme, Bayes formalism

This instruction details how to analyze the readout from a digital assay in the dependent scheme, using Bayes' formalism. In digital methods, the positive signal yielded by a compartment or partition v states, that in the said partition the number of molecules of the analyte is at least one. Unlike independent scheme, here the probability of obtaining a positive signal in one partition affects the probability of obtaining a positive signal in any of the rest of compartments constituting the assay. Therefore, the probability of obtaining a given outcome is not the product of probability functions of the compartments, but gets a combinatorial form.

To analyze the results of the assay, first, the binary values (digital readouts) of s_i for the partitions must be collected.

Then, the analysis is done as follows:

1. We calculate the probability $P(s_i|M)$ of obtaining the recorded state of a digital assay (e.g. the recorded number of positive compartments $K = \sum_{i=0}^{N-1} s_i$ out of all N compartments constituting and assay), which is a combinatorial function of the number of copies in the assay M:

$$P(K|M,N) = \frac{\binom{N}{K} \sum_{i=0}^{K-1} ((-1)^i \binom{K}{K-i} (K-i)^M)}{N^M}$$

2. Then, we use Bayesian formalism to inverse the product probability to the probability distribution $P(M|s_i)$ of the initial number of copies M in the sample (under the condition that any a priori information about the distribution of M is available): $P(M|K,N) = P(K|M,N)/\sum_{M=0}^{\infty} P(K|M,N)$; for practical reasons (numerical calculation), the upper limit of summation can be finite but should be at least one order of magnitude larger than the upper limit of the dynamic range of the assay

3. Knowing the distribution $P(M|K,N)$, one can estimate the initial number of copies in the assay as the expected value of this distribution: $E_D(M) = \sum_{M=0}^{\infty} M \cdot P(M|K,N)$; again, the upper limit of integration can be finite but should be at least one order of magnitude larger than the upper limit of the dynamic range of the assay

4. The precision of the estimate can be calculated as the spread (relative standard deviation) of the distribution $P(M|K,N)$: $\sigma_D(M) = \sqrt{E_D(M^2) - (E_D(M))^2}/E_D(M)$, where
$$E_D(M^2) = \sum_{M=0}^{\infty} M^2 \cdot P(M|K,N)$$

5. Knowing the estimate of the initial number of copies in the assay M and the relative standard deviation $\sigma_D(M)$ of this estimate, one can determine the initial concentration of the analyte in the assay:
$$E_D(C_A) = E_D(M)/(N \cdot v)$$
and
$$\sigma_D(C_A) = \sigma_D(M)$$

6. If the calculated precision of the estimate is worse than required (i.e. $\sigma_D(C_A) > \sigma_{max}$, where the value σ_{max} was used to design the assay), the assay should be repeated. Higher than expected value of $\sigma_D(C_A)$ is a sign of a very improbable outcome of the assay, so the result cannot be trusted.

7. Finally, having calculated the estimate of the concentration in the assay C_A and the relative standard deviation $\sigma_D(C_A)$, one can determine the estimate of the concentration C_S in the sample: $E(C) = \alpha E_D(C_A)$, and $\sigma(C) = \sigma_D(C_A)$, where α is a numerical factor reflecting the change in the volume between the sample and the assay and the efficiency of isolation $\eta = M/M_S$: $\alpha = (V_A/V_S)/\eta$.

A.1.3 Independent scheme, Most Probable Number method

This instruction details how to analyze the readout from a digital assay in the independent scheme, using the Most Probable Number (MPN) method. In digital methods, the positive signal yielded by a compartment or partition v states, that in the said partition the number of molecules of the analyte is at least one. This can be interpreted as a probability of obtaining a positive signal, which is a function of initial concentration in the assay C_A:
$$p_i = p(s_i = 1|C_A) = 1 - e^{-C_A v}$$

On the other hand, the negative signal yielded by a compartment or partition v states, that in the said partition there were no particles of the analyte, which can be interpreted with another probability function of C_A:

$$p_i = p(s_i = 0|C_A) = e^{-C_A v}$$

To analyze the results of the assay, first, the binary values (digital readouts) of s_i for the partitions must be collected.

Then, the analysis is done as follows:

1. From every positive partition (i.e. $s_i = 1$) with volume v we construct probability function p_i:

$$p_i = p(s_i = 1|C_A) = 1 - e^{-C_A v}$$

2. We also construct probabilities of obtaining negative signals for all negative compartments:

$$p_i = p(s_i = 0|C_A) = e^{-C_A v}$$

3. Then, we calculate the probability $P(\{s_i\}|C_A)$ of obtaining the recorded state of a digital assay (e.g. the recorded number of positive compartments $K = \sum_{i=0}^{N-1} s_i$ out of all N compartments constituting and assay), which is a product of the probability functions for all the compartments:

$$P(\{s_i\}|C_A) = (1 - e^{-C_A v})^K \cdot (e^{-C_A v})^{N-K}$$

4. Then, we find the value of concentration in the assay C_A, for which the probability $(s_i|C_A)$ has the largest value. To do that, we have to solve the equation: $(d(\{s_i\}|C_A))/(dC_A) = 0$. The solution of this equation is the value of C_A for which the derivative of probability $(\{s_i\}|C_A)$ equals zero, is:

$$E_I(C_A) = v^{-1} log(N/(N - K))$$

5. Finally, having calculated the estimate of the concentration in the assay C_A, one can determine the estimate of the concentration C in the source: $E(C) = \alpha E_I(C_A)$, where α is a numerical factor reflecting the change in the volume between the sample and the assay and the efficiency of isolation $\eta = M/M_S$: $\alpha = (V_A/V_S)/\eta$.

A.1.4 Dependent scheme, Most Probable Number method

This instruction details how to analyze the readout from a digital assay in the dependent scheme, using the Most Probable Number (MPN) Method. In digital methods, the positive signal yielded by a compartment or partition v states, that in the said partition the number of molecules of the analyte is at least one. Unlike independent scheme, here the probability of obtaining a positive signal in one partition affects the probability of obtaining a positive signal in any of the rest of compartments constituting the assay. Therefore, the probability of obtaining a given outcome is not the product of probability functions of the compartments, but gets a combinatorial form.

To analyze the results of the assay, first, the binary values (digital readouts) of s_i for the partitions must be collected. Then, the analysis is done as follows:

- We calculate the probability $P(s_i|M)$ of obtaining the recorded state of a digital assay (e.g. the recorded number of positive compartments $K = \sum_{i=0}^{N-1} s_i$ out of all N compartments constituting and assay), which is a combinatorial function of the number of copies in the assay M:

$$P(K|M,N) = \frac{\binom{N}{K} \sum_{i=0}^{K-1} ((-1)^i \binom{K}{K-i} (K-i)^M)}{N^M}$$

- Then, we find the value of the number of copies in the assay M, for which the probability $P(K|M,N)$ has the largest value. To do that, we have to solve the equation:

$$dP(K|M,N)/dM = 0.$$

  The solution of this equation is the value of M for which the derivative of probability $P(K|M,N)$ equals zero. It is advisable to calculate this value numerically.

- Knowing the estimate of the initial number of copies in the assay M, one can determine the initial concentration of the analyte in the assay:

$$E_D(C_A) = E_D(M)/(Nv).$$

- Finally, having calculated the estimate of the concentration in the assay C_A and the relative standard deviation $\sigma_D(C_A)$, one can determine the estimate of the concentration C_S in the sample:

$$E(C_S) = \alpha E_D(C_A),$$

  where α is a numerical factor reflecting the change in the volume of the sample and the

assay and the efficiency of isolation $\eta = M/M_S$:

$$\alpha = (V_A/V_S)/\eta.$$

Appendix B

Rational digital assays

B.1 Practical guideline how to design a digital assay that provides the required dynamic range and precision of the assessment

The formulas for the design of rational digital assays described in Chapter 7 use as an explicit input the values of the precision and range of concentrations to be assessed by an assay.

Rational digital assays can be divided into four groups, depending on the way of execution and limitations stemming from the available laboratory equipment. Firstly, the distinction comes from the fact that the compartments can be tracked during amplification or not. Tracing of the compartments means that if each of the compartments can be identified and its parameters, i.e. the volume and dilution factor, are known and can be assigned to the signal recorded. This provides bigger information gain from the signals, and therefore lowers the number of compartments needed to assess the concentration within a requested dynamic range with a requested concentration. However, the tracking of the compartments can be technically demanding, especially for large assays (in practice, these are assays that cannot be executed using a standard 384-well plate).

Alternatively, the compartments can be amplified together and inspected for signal in any order. This may dramatically lower the technical requirements, but in order to provide assessment with requested parameters, a larger number of compartments is needed.

Another distinction is a result of technical limitations of the generation of compartments; especially the generation of fine gradations (i.e. $x \approx 1$), where the consecutive compartments vary

by less than 5%, is demanding. As it was stated, such difficult assays can be simplified by using sets of copies (libraries) of identical compartments, where the gradation of the sets is coarse.

In the following subsections, designs of assays for any combination of the abovementioned criteria are given. In the first subsection, the design for an assay with trackable compartments is given. In the second section, the library scheme of trackable compartments is presented. In the third subsection, the design of assays with compartments that are not tracked is given, and in the last subsection similar assays comprising libraries are shown.

B.1.1 Design of an assay with tracking of the identity of each of the compartments.

Here, the design of a rational digital assay is given in detail. It is not constrained by any technical restrictions on the preparation, including the gradation of volumes, and the compartments are tracked during the process of amplification and signal readout.

In order to design an assay, the parameters of the assessment, i.e. the precision σ_{max} of the estimate and the range of concentration $C \in (C^-, C^+)$ should be determined.

These parameters are used as input for the design formulas. The common ratio x of the geometric sequence of modulation factors of compartments (gradation between the compartments):

$$x = 1 - \alpha\sigma_{max}^{\beta},$$

where $\alpha = 1.24$ and $\beta = 1.9493$.

This equation can be used to determine the common ratio x for the requested standard deviation of the estimate limited by $\sigma_{max} < 0.89$. If the gradation S of compartments calculated using the above equation cannot be performed with the available laboratory equipment, a preferred value of gradation should be chosen and the instructions for library scheme in the next subsection should be used.

Then, the number of compartments comprising an *active stripe* is calculated:

$$\Delta N = \lceil \gamma(1-x)^{-\delta} \rceil,$$

where $\gamma = 2.2815$ and $\delta = 0.798$.

The *active stripe* is span in order to cover the requested dynamic range $\Omega = C^+/C^-$ by adding $\lceil log_x(1/\Omega) \rceil$ compartments in the same geometric sequence. Therefore, the final number of

compartments of the assay is given by:

$$N = 2 \cdot \Delta N + \lceil log_x(1/\Omega) \rceil$$

The first compartment in the geometric sequence is parametrized by the modulation factor $z_0 = d_0 v_0$ given by the equation:

$$d_0 v_0 = log(2) \cdot x^{-\Delta N} / C^-.$$

Having established the parameters of the geometric sequence of compartments, starting from the first compartment $d_0 v_0$, what remains is the preparation of the partitioning of the sample into the said series of compartments. It can be can be obtained in a number of ways:

- Changing only the volumes v_i of partitions in order to form a geometric sequence with common ratio x: $v_i = v_0 z_i = v_0 x^{i-1}$.
- Changing only the dilutions d_i of partitions in order to form a geometric sequence with common ratio x: $d_i = d_0 z_i = d_0 x^{i-1}$.
- In general, any combination of the above can be used as long as a geometric sequence of modulation factors ($z_i = d_i v_i$) with common ratio x is established: $d_i v_i = d_0 v_0 z_i = d_0 v_0 x^{i-1}$.

Such partitioning of the sample for a digital PCR assay provides the assessment of the concentration with a requested standard deviation σ_{max} in the requested dynamic range $C \in (C^-, C^+)$.

B.1.2 Design of an assay, with a limitation on minimum gradation of compartments, with tracking of the identity of each of the compartments.

Here, the design of a rational digital assay, with the compartments tracked during the process of amplification and signal readout, but with a limited gradation of compartments is discussed in detail. The compartments are tracked during the process of amplification and signal readout. The fine gradation of compartments is replaced with a coarse gradation of libraries.

In order to design an assay, the parameters of the assessment, the range of concentrations $C \in (C^-, C^+)$ should be determined. In addition, two out of three following parameters should be fixed:

- precision of the assessment σ_{max} of the initial concentration,

- preferable gradation of compartments comprising consecutive libraries x (common ratio of the geometric progression of modulation factors), given by the laboratory equipment, and
- the number N' of identical compartments in a single library, determined by the laboratory equipment.

After fixing two parameters, the third one has to be determined using the following formulas:

- for a fixed precision σ_{max} and gradation x, the number N' of compartments belonging to a single library has to be determined:

  $$N' = \lceil 1/(1\sigma_{max}^2((1-x)/\alpha)^{2/\beta} \rceil,$$

  where $\alpha = 1.24$ and $\beta = 1.9493$, and this formula is correct for precision $\sigma_{max} < 0.89$ and gradation $x \in (0, 1)$;

- for a fixed precision σ_{max} and number N' of compartments belonging to a single library, the gradation x has to be determined:

  $$x = 1 - \alpha \cdot (\sigma_{max}\sqrt{N'}^{\beta},$$

  where $\alpha = 1.24$ and $\beta = 1.9493$, and this formula is correct for precision σ_{max} and for the number of copies N' which satisfies the requirement $\sqrt{N'} < 0.89/\sigma_{max}$;

- for a fixed gradation x and number N' of compartments belonging to a single library, the maximum precision σ_{max} of the estimate to be determined:

  $$\sigma_{max} = 1/\sqrt{N'}((1-x)/\alpha)^{1/\beta},$$

  where $\alpha = 1.24$ and $\beta = 1.9493$, and this formula is correct for gradation $x < 1$ and the number N' of compartments belonging to a single library $N' \geq 1$.

Then, the number of libraries comprising an *active stripe* is calculated:

$$\Delta N = \lceil \gamma \cdot (1-x)^{-\delta} \rceil,$$

where $\gamma = 2.2815$ and $\delta = 0.798$.

The *active stripe* is span in order to cover the requested dynamic range $\Omega = C^+/C^-$ by adding $\lceil log_x(1/\Omega) \rceil$ libraries in the same geometric sequence. Therefore, the final number of compartments of the assay is given by:

$$N = 2 \cdot \Delta N + \lceil log_x(1/\Omega) \rceil.$$

Compartments belonging to the first library in the geometric sequence are parametrized by the modulation factor $z_0 = d_0 v_0$ given by the equation:

$$d_0 v_0 = log(2) \cdot x^{-\Delta N} / C^-.$$

Having established the parameters of the geometric sequence of libraries, starting from the first library $d_0 v_0$, what remains is the preparation of the partitioning of the sample into the said series of libraries. It can be can be obtained in a number of ways:

- Changing only the volumes v_i of partitions in consecutive libraries in order to form a geometric sequence with common ratio x: $v_i = v_0 z_i = v_0 x^{i-1}$.
- Changing only the dilutions d_i of partitions in consecutive libraries in order to form a geometric sequence with common ratio x: $d_i = d_0 z_i = d_0 x^{i-1}$.
- In general, any combination of the above can be used as long as a geometric sequence of modulation factors ($z_i = d_i v_i$) with common ratio x is established: $d_i v_i = d_0 v_0 z_i = d_0 v_0 x^{i-1}$.

Such partitioning of the sample for a digital PCR assay provides the assessment of the concentration within a requested dynamic range $C \in (C^-, C^+)$ and meeting the additional requirements set by the user.

B.1.3 Design of an assay without tracking of the identity of each of the compartments.

Here, the design of a rational digital assay, with the compartments not being tracked during the process of amplification and signal readout, is discussed in detail. It is not constrained by any further technical restrictions on the preparation, including the gradation of volumes.

In order to design an assay, the parameters of the assessment, i.e. the precision σ_{max} of the estimate and the range of concentration $C \in (C^-, C^+)$ should be determined.

These parameters are used as input for the design formulas. The common ratio x of the geometric sequence of modulation factors of compartments (gradation between the compartments):

$$x = 1 - \alpha \cdot (\beta \sigma_{max} - \gamma)^{\delta},$$

where $\alpha = 1.24$, $\beta = 0.978$, $\gamma = 0.02$ and $\delta = 1.9493$.

This equation can be used to determine the common ratio x for the requested standard deviation of the estimate limited by: $\sigma_{max} < 0.89$. If the gradation S of compartments calculated using the above equation cannot be performed with the available laboratory equipment, a preferred value of gradation should be chosen and the instructions for library scheme in the next subsection should be used.

Then, the number of compartments comprising an *active stripe* is calculated:

$$\Delta N = \lceil \epsilon (1-x)^{-\theta} \rceil,$$

where $\epsilon = 2.2815$ and $\theta = 0.798$.

The *active stripe* is span in order to cover the requested dynamic range $\Omega = C^+/C^-$ by adding $\lceil log_x(1/\Omega) \rceil$ compartments in the same geometric sequence. Therefore, the final number of compartments of the assay is given by:

$$N = 2 \cdot \Delta N + \lceil log_x(1/\Omega) \rceil.$$

The first compartment in the geometric sequence is parametrized by the modulation factor $z_0 = d_0 v_0$ given by the equation:

$$d_0 v_0 = log(2) \cdot x^{-\Delta N}/C^-.$$

Having established the parameters of the geometric sequence of compartments, starting from the first compartment $d_0 v_0$, what remains is the preparation of the partitioning of the sample into the said series of compartments. It can be can be obtained in a number of ways:

- Changing only the volumes v_i of partitions in order to form a geometric sequence with common ratio x: $v_i = v_0 z_i = v_0 x^{i-1}$.
- Changing only the dilutions d_i of partitions in order to form a geometric sequence with common ratio x: $d_i = d_0 z_i = d_0 x^{i-1}$.
- In general, any combination of the above can be used as long as a geometric sequence of modulation factors ($z_i = d_i v_i$) with common ratio x is established: $d_i v_i = d_0 v_0 z_i = d_0 v_0 x^{i-1}$.

Such partitioning of the sample for a digital PCR assay provides the assessment of the concentration with a requested standard deviation σ_{max} in the requested dynamic range $C \in (C^-, C^+)$ without the requirement of tracking the compartments during the procedure.

B.1.4 Design of an assay, with a limitation on minimum gradation of compartments, without tracking of the identity of each of the compartments.

Here, the design of a rational digital assay, with the compartments not being tracked during the process of amplification and signal readout, but with a limited gradation of compartments is discussed in detail. The fine gradation of compartments is replaced with a coarse gradation of libraries.

In order to design an assay, the parameters of the assessment, the range of concentrations $C \in (C^-, C^+)$ should be determined. In addition, two out of three following parameters should be fixed:

- precision of the assessment σ_{max} of the initial concentration,
- preferable gradation of compartments comprising consecutive libraries x (common ratio of the geometric progression of modulation factors), given by the laboratory equipment, and
- the number N' of identical compartments in a single library, determined by the laboratory equipment.

After fixing two parameters, the third one has to be determined using the following formulas:

- for a fixed precision σ_{max} and gradation x, the number N' of compartments belonging to a single library has to be determined:

  $$N' = \lceil 1/(\beta\sigma_{max})^2((1-x)/\alpha)^{1/\delta} + \gamma \rceil,$$

  where $\alpha = 1.24$ and $\beta = 1.9493$, and this formula is correct for precision $\sigma_{max} < 0.55$ and gradation $x \in (0, 1)$;

- for a fixed precision σ_{max} and number N' of compartments belonging to a single library, the gradation x has to be determined:

  $$x = 1 - \alpha \cdot (\beta\sigma_{max}\sqrt{N' - \gamma}^{\delta},$$

  where $\alpha = 1.24$, $\beta = 0.978$, $\gamma = 0.02$ and $\delta = 1.9493$, and this formula is correct for precision σ_{max} and for the number of copies N' which satisfies the requirement $\sqrt{N'} < 0.55/\sigma_{max}$;

- for a fixed gradation x and number N' of compartments belonging to a single library, the

maximum precision σ_{max} of the estimate has to be determined:

$$\sigma_{max} = 1/(\beta\sqrt{N'})(((1-x)/\alpha)^{1/\delta} + \gamma),$$

where $\alpha = 1.24$, $\beta = 0.978$, $\gamma = 0.02$ and $\delta = 1.9493$, and this formula is correct for gradation $x < 1$ and the number N' of compartments belonging to a single library $N' \geq 1$.

Then, the number of libraries comprising an *active stripe* is calculated:

$$\Delta N = \lceil \gamma \cdot (1-x)^{-\delta} \rceil,$$

where $\gamma = 2.2815$ and $\delta = 0.798$.

The *active stripe* is span in order to cover the requested dynamic range $\Omega = C^+/C^-$ by adding $\lceil log_x(1/\Omega) \rceil$ libraries in the same geometric sequence. Therefore, the final number of compartments of the assay is given by:

$$N = 2 \cdot \Delta N + \lceil log_x(1/\Omega) \rceil.$$

Compartments belonging to the first library in the geometric sequence are parametrized by the modulation factor $z_0 = d_0 v_0$ given by the equation:

$$d_0 v_0 = log(2) \cdot x^{-\Delta N}/C^-.$$

Having established the parameters of the geometric sequence of libraries, starting from the first library $d_0 v_0$, what remains is the preparation of the partitioning of the sample into the said series of libraries. It can be can be obtained in a number of ways:

- Changing only the volumes v_i of partitions in consecutive libraries in order to form a geometric sequence with common ratio x: $v_i = v_0 z_i = v_0 x^{i-1}$.
- Changing only the dilutions d_i of partitions in consecutive libraries in order to form a geometric sequence with common ratio x: $d_i = d_0 z_i = d_0 x^{i-1}$.
- In general, any combination of the above can be used as long as a geometric sequence of modulation factors ($z_i = d_i v_i$) with common ratio x is established: $d_i v_i = d_0 v_0 z_i = d_0 v_0 x^{i-1}$.

Such partitioning of the sample for a digital PCR assay provides the assessment of the concentration within a requested dynamic range $C \in (C^-, C^+)$ and meeting the additional requirements set by the user without the requirement of tracking the compartments during the procedure.

B.2 Step by step instruction on how to analyze the results of a rational design PCR assay

Here, detailed instructions for the qualitative analysis of the readout from an optimized digital assay, based on the independent random variables, with the use of the Bayes' formalism is given. A positive signal yielded by an inspected compartment of a digital assay means that the said compartment contained at least threshold number of molecules of the analyte. For PCR-based digital methods, it is assumed that the threshold equals $m_{tr} = 1$. Therefore, the presence of a positive signal if an i-th compartment having a volume v can be translated into the following probability density, which is a function of initial concentration in the assay C_A:

$$p_i = p(s_i = 1|C_A) = 1 - e^{-C_A v}.$$

Then, a negative signal yielded by an inspected compartment of the volume v can be interpreted as a situation when the said i-th partition did not contain any molecules of the analyte. This can be translated into the following probability distribution, being also a function of initial concentration C_A:

$$p_i = p(s_i = 0|C_A) = e^{-C_A v}.$$

The analysis of the outcome of the assay starts with collecting all the digital readouts s_i from all the compartments of the assay. Then, the procedure is as following:

1. Every partition with volume v, which yields a positive signal (i.e. $s_i = 1$) is assigned with a probability density p_i:

$$p_i = p(s_i = 1|C_A) = 1 - e^{-C_A d_i v_i}$$

2. Analogically, all the negative compartments are assigned with a respective probability density:

$$p_i = p(s_i = 0|C_A) = e^{-C_A d_i v - i}$$

3. The probability $P(\{s_i\}|C_A)$ of recording a given outcome of a digital assay, i.e. *microstate* of the assay, is a product of the probability densities assigned to all the compartments:

$$P(\{s_i\}|C) = \prod_{i=0}^{N-1} p_i(s_i|C).$$

Therefore, the final probability equals:

$$P(\{s_i\}|C) = \prod_{i=0}^{N-1} (1 - e^{-Cd_iv_i})^{s_i} \cdot (e^{-Cd_iv_i})^{1-s_i}.$$

4. Then, this product probability is inverted using the Bayesian formalism in order to obtain the distribution $P(C_A|\{s_i\})$ of the initial concentration C_A, given the recorded outcome of an assay (it is assumed that no a priori information, i.e. the distribution $p(C_A)$, is available):

$$P(C_A|\{s_i\}) = P(\{s_i\}|C_A) / \int_0^\infty P(\{s_i\}|C_A) dC_A$$

for the numerical calculations of the integral, the upper limit can be finite, although it should be reasonably larger (by at least one order of magnitude) than the highest initial concentration to be assessed

5. Having calculated the distribution $P(C_A|\{s_i\})$, the estimate of the initial concentration can be calculated as the expected value of this distribution:

$$E_I(C_A) = \int_0^\infty C_A \cdot P(C_A|\{s_i\}) dC_A$$

again, for the numerical calculations of the integral, the upper limit can be finite, although it should be reasonably larger than the highest initial concentration to be assessed

6. Then, the precision of the estimate is given as a relative standard deviation of the distribution $P(C_A|\{s_i\})$:

$$\sigma_I(C_A) = \sqrt{E_I(C_A^2) - (E_I(C_A)^2}/E_I(C_A),$$

where

$$E_I(C_A^2) = \int_0^\infty C_A^2 \cdot P(C_A|\{s_i\}) dC_A$$

7. Finally, knowing the estimate of the initial global concentration $E(C_A)$ and the precision $\sigma_I(C_A)$, the estimate of the concentration C in the source can be calculated as:

$$E(C) = \alpha E_I(C_A)$$

$$\sigma(C) = \sigma_I(C_A),$$

where a factor α is determined by the efficiency of the isolation protocol $\eta = M/M_S$ and the difference between the volume of the sample and the volume of the assay: $\alpha = (V_A/V_S)/\eta$.

Appendix C

Synergistic analogue-digital assays

C.1 Practical guideline how to design a synergistic assay that provides the required dynamic range and precision of the assessment

The formulas for the design of synergistic digital-analogue assays described in Chapter 8 use as an explicit input the values of the precision and range of concentrations to be assessed by an assay.

Synergistic assays can be divided into two groups, depending on the limitations steming from the available laboratory equipment. The distinction is a result of technical limitations of the generation of compartments; especially the generation of fine gradations (i.e. $x \approx 1$), where the consecutive compartments vary by less than 5%, is demanding. As it was stated, such difficult assays can be simplified by using sets of copies (libraries) of identical compartments, where the gradation of the sets is coarse.

In the following subsections, designs of assays for the abovementioned criteria are given. In the first subsection, the design for an assay with non-identical compartments is given, while in the second subsection, the library scheme is presented.

C.1.1 Detailed description of the design of an optimum synergistic assay.

Here, the design of a synergistic assay is given in detail. It is not contrained by any technical restrictions on the preparation, including the gradation of volumes, and the compartments are tracked during the process of amplification and signal readout.

In order to design an assay, the parameters of the assessment, i.e. the precision σ_{max} of the estimate and the range of concentration $C \in (C^-, C^+)$ should be determined.

These parameters are used as input for the design formulas. The common ratio x of the geometric sequence of modulation factos of compartments (gradation between the compartments):

$$x = \alpha\sigma_{max}^2 - \beta\sigma_{max} + \gamma,$$

where $\alpha = 0.5540$, $\beta = 1.6504$ and $\gamma = 1.1135$.

This equation can be used to determine the common ration x for the requested standard deviation of the estimate limited by: $\sigma_{max} \in (0.07, 1)$. If the gradation S of compartments calculated using the above equation cannot be performed with the available laboratory equipement, or a more precise assessment is needed, a preferred value of gradation (and precision) should be chosen and the instructions for library scheme in the next subsection should be used.

Then, the number of compartments comprising an *active stripe* is calculated:

$$\Delta N = \lceil \delta\sigma_{max}^{-\epsilon} \rceil,$$

where $\delta = 2.0532$ and $\epsilon = 1.3220$.

The *active stripe* is span in order to cover the requested dynamic range $\Omega = C^+/C^-$ by adding $\lceil log_x(1/\Omega) \rceil$ compartments in the same geometric sequence. Therefore, the final number of compartments of the assay is given by:

$$N = 2 \cdot \Delta N + \lceil log_x(1/\Omega) \rceil.$$

The first compartment in the geometric sequence is parametrized by the modulation factor $z_0 = d_0 v_0$ given by the equation:

$$d_0 v_0 = log(2) \cdot x^{-\Delta N}/C^-.$$

Having established the parameters of the geometric sequence of compartments, starting from the first compartment $d_0 v_0$, what remains is the preparation of the partitioning of the sample into the said series of compartments. It can be can be obtained in a number of ways:

- Changing only the volumes v_i of partitions in order to form a gemoetric sequence with common ratio x: $v_i = v_0 z_i = v_0 x^{i-1}$.

- Changing only the dilutions d_i of partitions in order to form a gemoetric sequence with common ratio x: $d_i = d_0 z_i = d_0 x^{i-1}$.

- In general, any combination of the above can be used as long as a gemoetric sequence of modulation factors ($z_i = d_i v_i$) with common ratio x is established: $d_i v_i = d_0 v_0 z_i = d_0 v_0 x^{i-1}$.

Such partitioning of the sample for a synergistic PCR assay provides the assessment of the concentration with a requested standard deviation σ_{max} in the requested dynamic range $C \in (C^-, C^+)$.

C.1.2 Detailed description of the design of an assay, with a limitation on the minimum gradation between the compartments.

Here, the design of a synergistic assay, with the compartments tracked during the process of amplification and signal readout, but with a limited gradation of compartments is discussed in detail. The fine gradation of compartments is replaced with coarse gradation of libraries.

In order to design an assay, the parameters of the assessment, the range of concentrations $C \in (C^-, C^+)$ should be determined. In addition, two out of three following parmeters should be fixed:

- precision of the assessment σ_{max} of the initial concentration,

- preferable gradation of compartments comprising consecutive libraries x (common ratio of the geometric progression of modulation factors), given by the laboratory equipement, and

- the number N' of identical compartments in a single library, determined by the laboratory equipment.

After fixing two parameters, the third one has to be determined using the following formulas:

- for a fixed precision σ_{max} and gradation x, the number N' of compartments belonging to a single library has to be determined:

$$N' = \lceil (\frac{\beta - \sqrt{\beta - 4\alpha(\gamma - x)}}{\alpha \sigma_{max}})^2 \rceil,$$

where $\alpha = 0.5540$, $\beta = 1.6504$ and $\gamma = 1.1135$, and this formula is correct for precision $\sigma_{max} < 1$ and gradation $x \in (0, 1)$;

- for a fixed precision σ_{max} and number N' of compartments belonging to a single library, the gradation x has to be determined:

$$x = \alpha(\sigma_{max}\sqrt{N'})^2 - \beta\sigma_{max}\sqrt{N'} + \gamma,$$

where $\alpha = 0.5540$, $\beta = 1.6504$ and $\gamma = 1.1135$;

- for a fixed gradation x and and number N' of compartments belonging to a single library, the maximum precision σ_{max} of the estimatehas to be determined:

$$\sigma_{max} = \frac{\beta - \sqrt{\beta - 4\alpha(\gamma - x)}}{\alpha\sqrt{N'}},$$

where $\alpha = 0.5540$, $\beta = 1.6504$, and this formula is correct for gradation $x < 1$ and the number N' of compartments belonging to a single library $N' \geq 1$.

Then, the number of libraries comprising an *active stripe* is calculated:

$$\Delta N = \lceil \delta\sigma_{max}^{-\epsilon} \rceil,$$

where $\delta = 2.0532$ and $\epsilon = 1.3220$.

The *active stripe* is span in order to cover the requested dynamic range $\Omega = C^+/C^-$ by adding $\lceil log_x(1/\Omega) \rceil$ libraries in the same geometric sequence. Therefore, the final number of compartments of the assay is given by:

$$N = 2 \cdot \Delta N + \lceil log_x(1/\Omega) \rceil.$$

Compartments belonging to the first library in the geometric sequence are parametrized by the modulation factor $z_0 = d_0 v_0$ given by the equation:

$$d_0 v_0 = log(2) \cdot x^{-\Delta N}/C^-.$$

Having established the parameters of the geometric sequence of libraries, starting from the first library $d_0 v_0$, what remains is the preparation of the partitioning of the sample into the said series of libraries. It can be can be obtained in a number of ways:

- Changing only the volumes v_i of partitions in consecuteive libraries in order to form a gemoetric sequence with common ratio x: $v_i = v_0 z_i = v_0 x^{i-1}$.

- Changing only the dilutions d_i of partitions in consecuteive libraries in order to form a gemoetric sequence with common ratio x: $d_i = d_0 z_i = d_0 x^{i-1}$.

- In general, any combination of the above can be used as long as a gemoetric sequence of modulation factors ($z_i = d_i v_i$) with common ratio x is established: $d_i v_i = d_0 v_0 z_i = d_0 v_0 x^{i-1}$.

Such partitioning of the sample for a synergistic PCR assay provides the assessment of the concentration within a requested dynamic range $C \in (C^-, C^+)$ and meeting the additional requirements set by the user.

C.2 Step by step instruction on how to analyze the results of a synergistic PCR assay

Here, detailed instructions for the qualitative analysis of the readout from a synergistic analogue-digital assay, based on the independent random variables, with the use of the Bayes' formalism is given. A positive signal yielded by an inspected compartment of a digital assay means that the said compartment contained at least treshold number of molecules of the analyte. For PCR-based digital methods, it is assumed that the threshold equals $m_{tr} = 1$. Therefore, the presence of a positive signal if an i-th compartment having a volume v can be translated into the following probability density, which is a function of initial concentration in the assay C_A:

$$p_i = p(s_i = 1|C_A) = 1 - e^{-C_A v}.$$

Then, a negative signal yielded by an inspected compartment of the volume v can be interpreted as a situtation, when the said i-th partition did not contain any molecules of the analyte. This can be translated into the following probability distribution, being also a function of initial concentration C_A:

$$p_i = p(s_i = 0|C_A) = e^{-C_A v}.$$

However, in synergistic assays, the analogue readout is used to increase the minimum (i.e. new threshold) number $m_i > 1$ of molecules in the inspected positive compartment. This translates into a modified probability of obtaining a positive signal, which is a function of initial concentration C_A, is given by:

$$p_i(s_i = 1|C_A) = 1 - e^{-C_A d_i v_i} \sum_{j=0}^{m_i - 1} (C_A d_i v_i)^j / j!.$$

The analysis of the outcome of the assay starts with collecting all the digital readouts s_i and analogue readouts ct_u from all the compartments of the assay. They are used to determine the values m_i for positive partition. The procedure is as following:

1. from the analogue readouts from positive partitions ct_i, the highest value, which refers to the latest appearance of the treshold level of signal, is chosen, and the respective compartments is called the reference compartment ω with readout ct_ω

2. the vector of m_i values is build by using the following formula:

$$m_i = q^{\Delta ct_i} = q^{ct_\omega - ct_i},$$

where q is the amplification factor of the reaction, which is a priori known, or can be calulated from experimental data, as it is described in the next section

3. after the calculation of the values m_i, each positive compartment is assigned with a probability density p_i:

$$p_i(\Delta ct_i|C) = 1 - e^{-Cd_iv_i} \sum_{j=0}^{m_i-1} (Cd_iv_i)^j/j!$$

4. Analogically, all the negative compartments are assigned with a respective probability density:

$$p_i = p(s_i = 0|C_A) = e^{-C_A d_i v_i}$$

5. The probability $P(s_i|C_A)$ of recording a given outcome of a synergistic assay, i.e. *microstate* of the assay, is a product of the probability desities assigned to all the compartments:

$$P(\{s_i\}, \{ct_i\}|C) = \prod_{i=0}^{N-1} p_i(s_i|C).$$

6. Then, this product probability is inversed using the Bayesian formalism in order to obtain the distribution $P(C_A|s_i)$ of the initial concentration C_A, given the recorded outcome of an assay (it is assumed that no a priori information, i.e. the distribution $p(C_A)$, is avaialable):

$$P(C_A|s_i) = P(s_i|C_A) / \int_0^\infty P(s_i|C_A) dC_A$$

for the numerical calculations of the integral, the upper limit can be finite, although it should be resonably larger (by at least one order of magnitude) than the highest initial concentration to be assessed

7. Having calculated the distribution $P(C_A|s_i)$, the estimate of the initial concentration can be calculated as the expected value of this distribution:

$$E_I(C_A) = \int_0^\infty C_A \cdot P(C_A|s_i) dC_A$$

again, for the numerical calculations of the integral, the upper limit can be finite, although it should be resonably larger than the highest initial concentration to be assessed.

8. Then, the precision of the estimate is given as a relative standard deviation of the distribution $P(C_A|s_i)$:

$$\sigma_I(C_A) = \sqrt{E_I(C_A^2) - (E_I(C_A)^2}/E_I(C_A),$$

where

$$E_I(C_A^2) = \int_0^\infty C_A^2 \cdot P(C_A|s_i) dC_A$$

9. Finally, knowing the estimate of the initial global concentration $E(C_A)$ and the precision $\sigma_I(C_A)$, the estimate of the concentration C in the source can be calculated as:

$$E(C) = \alpha E_I(C_A)$$

$$\sigma(C) = \sigma_I(C_A),$$

where a factor α is determined by the efficency of the isolation protocol $\eta = M/M_S$ and the difference between the volume of the sample and the volume of the assay: $\alpha = (V_A/V_S)/\eta$.

C.3 Calculation the amplification factor q

The use of real-time signals also allows us to determine the amplification factor q (the average of the ratio of the numbers of particles analyte in the test volume after two subsequent cycles or time intervals), if it is not a priori known. The advantage is, that this factor is determined specifically for the current sample and current substrate or apparatus, and each measurement is treated separately.

The calculation of the amplification factor q is done in the following steps:

1. for a known sequence of $d_i v_i$, the numbers of cycles ct_i, after which threshold signal (level of fluorescence) is observed, are measured

2. observing threshold level of fluorescence is a sign, that the current number of particles in the compartments is equal to some constant value (possibly unknown), hence we can state

that $m_{obs} = m_i q^{ct_i}$, where m_{obs} is the threshold number and m_i is the initial number of particles

3. the expected value of m_i is equal to $d_i v_i C$, where C is constant (it is the real value of concentration) and unknown

4. then, from the measurement of $\{ct_i\}$ and a priori knowledge of $\{d_i v_i\}$, one can plot $ct_i = f(log(d_i v_i))$

5. the gradient of the linear fit to this data is equal to $a = f(q) = 1/log(q)$, hence $q = e^{-1/a}$.

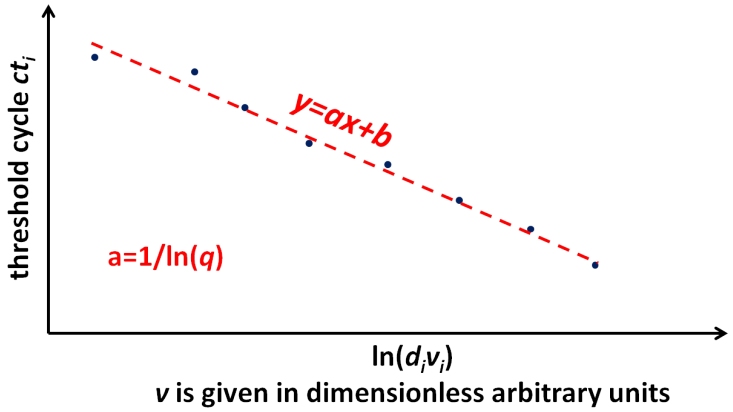


Figure C.1: Determination of the amplification factor $q = 1 + E$. The value of q can be retrieved from the slope of the linear fit to the points $(log(d_i v_i), ct_i)$. This procedure allows to determine the efficiency of the reaction in every run of a synergistic PCR alorithm, therefore it an be used even for poorly optimized reactions.

Bibliography

[1] Pamela M. Holland, Richard D. Abramson, Robert Watson, and David H. Gelfand. Detection of specific polymerase chain reaction product by utilizing the 5'–3'exonuclease activity of Thermus aquaticus DNA polymerase. *Proceedings of the National Academy of Sciences*, 88(16):7276–7280, 1991.

[2] M. H. McCrady. The Numerical Interpretation of Fermentation-Tube Results. *Journal of Infectious Diseases*, 17(1):183–212, July 1915.

[3] P. J. Sykes, S. H. Neoh, M. J. Brisco, E. Hughes, J. Condon, and A. A. Morley. Quantitation of targets for PCR by use of limiting dilution. *BioTechniques*, 13(3):444–449, September 1992.

[4] B. Vogelstein and K. W. Kinzler. Digital PCR. *Proceedings of the National Academy of Sciences*, 96(16):9236–9241, August 1999.

[5] Jason E. Kreutz, Todd Munson, Toan Huynh, Feng Shen, Wenbin Du, and Rustem F. Ismagilov. Theoretical design and analysis of multivolume digital assays with wide dynamic range validated experimentally with microfluidic digital PCR. *Analytical chemistry*, 83(21):8158–8168, 2011.

[6] Feng Shen, Wenbin Du, Jason E. Kreutz, Alice Fok, and Rustem F. Ismagilov. Digital PCR on a SlipChip. *Lab on a Chip*, 10(20):2666, 2010.

[7] Feng Shen, Bing Sun, Jason E. Kreutz, Elena K. Davydov, Wenbin Du, and Rustem F. Ismagilov. Quantification of hiv and hcv viral load with large dynamic range using multivolume digital reverse transcription pcr on a rotational slipchip, Jun 2013.

[8] K. B. Mullis and F. A. Faloona. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods in Enzymology*, 155:335–350, 1987.

[9] Kary B. Muliss. Nobel lecture, dec 1993.

[10] Alexander A. Morley. Digital PCR: A brief history. *Biomolecular Detection and Quantification*, 1(1):1 – 2, 2014.

[11] George Poste. Molecular diagnostics: a powerful new component of the healthcare value chain. *Expert Review of Molecular Diagnostics*, 1(1):1–5, May 2001.

[12] Slomski Ryszard. *Analiza DNA - teoria i praktyka*. Wydawnictwo Uniwersytetu Przyrodniczego w Poznaniu, Poznan, 2008.

[13] Nelson Fausto and Karen Lynn Kaul. Presenting The Journal of Molecular Diagnostics. *The Journal of Molecular Diagnostics*, 1(1):1, November 1999.

[14] Margaret A. Hamburg and Francis S. Collins. The Path to Personalized Medicine. *New England Journal of Medicine*, 363(4):301–304, July 2010.

[15] Y Waikan. Antenatal diagnosis of sickle-cell anaemia by D.N.A. analysis of amniotic-fluid cells. *The Lancet*, 312(8096):910–912, October 1978.

[16] Stuart H. Orkin, Blanche P. Alter, Cigdem Altay, Maurice J. Mahoney, Herbert Lazarus, John C. Hobbins, and David G. Nathan. Application of Endonuclease Mapping to the Analysis and Prenatal Diagnosis of Thalassemias Caused by Globin-Gene Deletion. *New England Journal of Medicine*, 299(4):166–172, July 1978.

[17] Bernard R. Glick, Jack J. Pasternak, and Cheryl L. Patten. *Molecular biotechnology: principles and applications of recombinant DNA*. ASM Press, Washington, DC, 4th ed edition, 2010. OCLC: ocn421361097.

[18] E.M. Southern. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology*, 98(3):503–517, November 1975.

[19] Erik Pettersson, Joakim Lundeberg, and Afshin Ahmadian. Generations of sequencing technologies. *Genomics*, 93(2):105–111, February 2009.

[20] K. B. Mullis. The unusual origin of the polymerase chain reaction. *Scientific American*, 262(4):56–61, 64–65, April 1990.

[21] John M. S. Bartlett and David Stirling. A Short History of the Polymerase Chain Reaction. In *PCR Protocols*, volume 226, pages 3–6. Humana Press, New Jersey, August 2003.

[22] K. Mullis, F. Faloona, S. Scharf, R. Saiki, G. Horn, and H. Erlich. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbor Symposia on Quantitative Biology*, 51 Pt 1:263–273, 1986.

[23] C. D. Boehm. Use of polymerase chain reaction for diagnosis of inherited disorders. *Clinical Chemistry*, 35(9):1843–1848, September 1989.

[24] Scott C. Kogan, Marie Doherty, and Jane Gitschier. An Improved Method for Prenatal Diagnosis of Genetic Diseases by Analysis of Amplified DNA Sequences. *New England Journal of Medicine*, 317(16):985–990, October 1987.

[25] E.J. Sorscher and Z. Huang. Diagnosis of genetic disease by primer-specified restriction map modification, with application to cystic fibrosis and retinitis pigmentosa. *The Lancet*, 337(8750):1115–1118, May 1991.

[26] Tomas Duffy, Margarita Bisio, Jaime Altcheh, Juan Miguel Burgos, Mirta Diez, Mariano Jorge Levin, Roberto Rene Favaloro, Hector Freilij, and Alejandro Gabriel Schijman. Accurate Real-Time PCR Strategy for Monitoring Bloodstream Parasitic Loads in Chagas Disease Patients. *PLoS Neglected Tropical Diseases*, 3(4):e419, April 2009.

[27] S. P. Johnston, N. J. Pieniazek, M. V. Xayavong, S. B. Slemenda, P. P. Wilkins, and A. J. da Silva. PCR as a Confirmatory Technique for Laboratory Diagnosis of Malaria. *Journal of Clinical Microbiology*, 44(3):1087–1089, March 2006.

[28] Noris RodrÃŋguez, Bernardo Guzman, A Rodas, Howard Takiff, Barry R Bloom, and Jacinto Convit. Diagnosis of cutaneous leishmaniasis and species discrimination of parasites by PCR and hybridization. *Journal of Clinical Microbiology*, 32(9):2246–2252, 1994.

[29] R. F. D'Amato, A. A. Wallman, L. H. Hochstein, P. M. Colaninno, M. Scardamaglia, E. Ardila, M. Ghouri, K. Kim, R. C. Patel, and A. Miller. Rapid diagnosis of pulmonary tuberculosis by using Roche AMPLICOR Mycobacterium tuberculosis PCR test. *Journal of Clinical Microbiology*, 33(7):1832–1834, July 1995.

[30] M Clementi, S Menzo, P Bagnarelli, A Manzin, A Valenza, and PE Varaldo. Quantitative PCR and RT-PCR in virology. *PCR methods and applications*, 2:191–191, 1993.

[31] E. M. Elnifro, A. M. Ashshi, R. J. Cooper, and P. E. Klapper. Multiplex PCR: optimization and application in diagnostic virology. *Clinical Microbiology Reviews*, 13(4):559–570, October 2000.

[32] U. Evertsson, H. J. Monstein, and A. G. Johansson. Detection and identification of fungi in blood using broad-range 28s rDNA PCR amplification and species-specific hybridisationNote. *APMIS*, 108(5):385–392, May 2000.

[33] Tomasz Hryniewiecki, Anna Gzyl, Ewa Augustynowicz, and Irena Rawczynska-Englert. Development of broad-range polymerase chain reaction (PCR) bacterial identification in diagnosis of infective endocarditis. *The Journal of Heart Valve Disease*, 11(6):870–874, November 2002.

[34] Ian M. Mackay, Katherine E. Arden, and Andreas Nitsche. Real-time PCR in virology. *Nucleic Acids Research*, 30(6):1292–1305, March 2002.

[35] C. Piersimoni, A. Callegaro, D. Nista, S. Bornigia, F. De Conti, G. Santini, and G. De Sio. Comparative evaluation of two commercial amplification assays for direct detection of Mycobacterium tuberculosis complex in respiratory specimens. *Journal of Clinical Microbiology*, 35(1):193–196, January 1997.

[36] Samuel Yang and Richard E Rothman. PCR-based diagnostics for infectious diseases: uses, limitations, and future applications in acute-care settings. *The Lancet Infectious Diseases*, 4(6):337–348, June 2004.

[37] TimothÃľe Houssin, JÃľrÃľmy Cramer, RÃľbecca Grojsman, Lyes Bellahsene, Guillaume Colas, HÃľlÃĺne Moulet, Walter Minnella, Christophe Pannetier, MaÃńl Leberre, Adrien Plecis, and Yong Chen. Ultrafast, sensitive and large-volume on-chip real-time PCR for the molecular diagnosis of bacterial and viral infections. *Lab Chip*, 16(8):1401–1411, 2016.

[38] D. Brambilla, P. S. Reichelderfer, J. W. Bremer, D. E. Shapiro, R. C. Hershow, D. A. Katzenstein, S. M. Hammer, B. Jackson, A. C. Collier, R. S. Sperling, M. G. Fowler, and R. W. Coombs. The contribution of assay variation and biological variation to the total variability of plasma HIV-1 RNA measurements. The Women Infant Transmission Study Clinics. Virology Quality Assurance Program. *AIDS (London, England)*, 13(16):2269–2279, November 1999.

[39] Alexandra Calmy, Nathan Ford, Bernard Hirschel, Steven J. Reynolds, Lut Lynen, Eric Goemaere, Felipe Garcia de la Vega, Luc Perrin, and William Rodriguez. HIV viral load monitoring in resource-limited regions: optional or necessary? *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, 44(1):128–134, January 2007.

[40] R. M. Gersberg, M. A. Rose, R. Robles-Sikisaka, and A. K. Dhar. Quantitative Detection of Hepatitis A Virus and Enteroviruses Near the United States-Mexico Border and Correlation with Levels of Fecal Indicator Bacteria. *Applied and Environmental Microbiology*, 72(12):7438–7444, December 2006.

[41] C. C. Ginocchio, M. Kemper, K. A. Stellrecht, and D. J. Witt. Multicenter Evaluation of the Performance Characteristics of the NucliSens HIV-1 QT Assay Used for Quantitation of Human Immunodeficiency Virus Type 1 RNA. *Journal of Clinical Microbiology*, 41(1):164–173, January 2003.

[42] Atul Humar, Carlos Paya, Mark D. Pescovitz, Ed Dominguez, Kenneth Washburn, Emily Blumberg, Barbara Alexander, Richard Freeman, Nigel Heaton, and Barbara Mueller. Clinical Utility of Cytomegalovirus Viral Load Testing for Predicting CMV Disease in D+/R- Solid Organ Transplant Recipients. *American Journal of Transplantation*, 4(4):644–649, April 2004.

[43] Olivia Keiser. Switching to second-line antiretroviral therapy in resource-limited settings: comparison of programmes with and without viral load monitoring:. *AIDS*, 23(14):1867–1874, September 2009.

[44] M. L. Landry, R. Garner, and D. Ferguson. Real-Time Nucleic Acid Sequence-Based Amplification Using Molecular Beacons for Detection of Enterovirus RNA in Clinical Specimens. *Journal of Clinical Microbiology*, 43(7):3136–3139, July 2005.

[45] Hubert G.M. Niesters. Quantitation of Viral Load Using Real-Time Amplification Techniques. *Methods*, 25(4):419–429, December 2001.

[46] L. E. Scott, L. D. Noble, J. Moloi, L. Erasmus, W. D. F. Venter, and W. Stevens. Evaluation of the Abbott m2000 RealTime Human Immunodeficiency Virus Type 1 (HIV-1) Assay for HIV Load Monitoring in South Africa Compared to the Roche Cobas AmpliPrep-Cobas Amplicor, Roche Cobas AmpliPrep-Cobas TaqMan HIV-1, and BioMerieux NucliSENS EasyQ HIV-1 Assays. *Journal of Clinical Microbiology*, 47(7):2209–2217, July 2009.

[47] Mickey Urdea, Laura A. Penny, Stuart S. Olmsted, Maria Y. Giovanni, Peter Kaspar, Andrew Shepherd, Penny Wilson, Carol A. Dahl, Steven Buchsbaum, Gerry Moeller, and Deborah C. Hay Burgess. Requirements for high impact diagnostics in the developing world. *Nature*, 444:73–79, November 2006.

[48] ShuQi Wang, Feng Xu, and Utkan Demirci. Advances in developing HIV-1 viral load assays for resource-limited settings. *Biotechnology Advances*, 28(6):770–781, November 2010.

[49] M.-H. Yen, K.-C. Tsao, Y.-C. Huang, C.-G. Huang, Y.-L. Huang, R. Lin, M.-L. Chang, C.-C. Huang, D.-C. Yan, and T.-Y. Lin. Viral Load in Blood Is Correlated with Disease Severity of Neonatal Coxsackievirus B3 Infection: Early Diagnosis and Predicting Disease

Severity Is Possible in Severe Neonatal Enterovirus Infection. *Clinical Infectious Diseases*, 44(10):e78–e81, May 2007.

[50] M. A. Jobling, A. Pandya, and C. Tyler-Smith. The Y chromosome in forensic analysis and paternity testing. *International Journal of Legal Medicine*, 110(3):118–124, April 1997.

[51] Niels Morling and Angel Carracedo. International recommendations for paternity testing standards. *Forensic Science International*, 129(3):147, October 2002.

[52] SÃĺrgio D.J. Pena and Ranajit Chakraborty. Paternity testing in the DNA era. *Trends in Genetics*, 10(6):204–209, June 1994.

[53] P Taberlet. Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research*, 24(16):3189–3194, August 1996.

[54] Yu-Li Tsai and Betty H Olson. Detection of low numbers of bacterial cells in soils and sediments by polymerase chain reaction. *Applied and Environmental Microbiology*, 58(2):754–757, 1992.

[55] Ying Li, Godelieve C. M. L. Page-Christiaens, Johan J. P. Gille, Wolfgang Holzgreve, and Sinuhe Hahn. Non-invasive prenatal detection of achondroplasia in size-fractionated cell-free DNA by MALDI-TOF MS assay. *Prenatal Diagnosis*, 27(1):11–17, January 2007.

[56] Vishnu Swarup and M.R. Rajeswari. Circulating (cell-free) nucleic acids - A promising, non-invasive tool for early detection of several human diseases. *FEBS Letters*, 581(5):795–799, March 2007.

[57] Haruo Takabayashi, Soryu Kuwabara, Toshihiko Ukita, Kazumi Ikawa, Kaoru Yamafuji, and Tatsuhiro Igaras. Development of non-invasive fetal DNA diagnosis from maternal blood. *Prenatal Diagnosis*, 15(1):74–77, January 1995.

[58] C. F. Wright and H. Burton. The use of cell-free fetal nucleic acids in maternal blood for non-invasive prenatal diagnosis. *Human Reproduction Update*, 15(1):139–151, October 2008.

[59] Christopher G. Kevil, Loren Walsh, F.Stephen Laroux, Theodore Kalogeris, Matthew B. Grisham, and J.S. Alexander. An Improved, Rapid Northern Protocol. *Biochemical and Biophysical Research Communications*, 238(2):277–279, September 1997.

[60] Henry A. Erlich. Polymerase chain reaction. *Journal of Clinical Immunology*, 9(6):437–447, November 1989.

[61] Henry A. Erlich, editor. *PCR technology: principles and applications for DNA amplification*. Stockton Press, New York, 1989.

[62] F Ferre. Quantitative or semi-quantitative PCR: reality versus myth. *Genome Research*, 2(1):1–9, August 1992.

[63] R Iakobashvili. Low temperature cycled PCR protocol for Klenow fragment of DNA polymerase I in the presence of proline. *Nucleic Acids Research*, 27(6):1566–1568, March 1999.

[64] A. Chien, D. B. Edgar, and J. M. Trela. Deoxyribonucleic acid polymerase from the extreme thermophile Thermus aquaticus. *Journal of Bacteriology*, 127(3):1550–1557, September 1976.

[65] M. Crescenzi, M. Seto, G. P. Herzig, P. D. Weiss, R. C. Griffith, and S. J. Korsmeyer. Thermostable DNA polymerase chain amplification of t(14;18) chromosome breakpoints and detection of minimal residual disease. *Proceedings of the National Academy of Sciences of the United States of America*, 85(13):4869–4873, July 1988.

[66] R. K. Saiki, S. Scharf, F. Faloona, K. B. Mullis, G. T. Horn, H. A. Erlich, and N. Arnheim. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science (New York, N.Y.)*, 230(4732):1350–1354, December 1985.

[67] C. T. Wittwer, G. C. Fillmore, and D. J. Garling. Minimizing the time required for DNA amplification by efficient heat transfer to small samples. *Analytical Biochemistry*, 186(2):328–331, May 1990.

[68] R. Higuchi, C. Fockler, G. Dollinger, and R. Watson. Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Bio/Technology (Nature Publishing Company)*, 11(9):1026–1030, September 1993.

[69] Marilynn R. Fairfax and Hossein Salimnia. Quantitative PCR. In *Molecular Diagnostics*, pages 3–14. Elsevier, 2010.

[70] Philip S. Bernard and Carl T. Wittwer. Real-time PCR technology for cancer diagnostics. *Clinical Chemistry*, 48(8):1178–1185, August 2002.

[71] S. A. Bustin. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *Journal of Molecular Endocrinology*, 25(2):169–193, October 2000.

[72] S. A. Bustin. Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *Journal of Molecular Endocrinology*, 29(1):23–39, August 2002.

[73] Dieter Klein. Quantification using real-time PCR technology: applications and limitations. *Trends in Molecular Medicine*, 8(6):257–260, June 2002.

[74] Nigel J. Walker. Tech.Sight. A technique whose time has come. *Science (New York, N.Y.)*, 296(5567):557–559, April 2002.

[75] C. T. Wittwer and D. J. Garling. Rapid cycle DNA amplification: time and temperature optimization. *BioTechniques*, 10(1):76–83, January 1991.

[76] R. Higuchi, G. Dollinger, P. S. Walsh, and R. Griffith. Simultaneous amplification and detection of specific DNA sequences. *Bio/Technology (Nature Publishing Company)*, 10(4):413–417, April 1992.

[77] E. B. Phelps. A Method of Calculating the Numbers of B. Coli from the Results of Dilution Tests. *American Journal of Public Hygiene*, 18(2):141–145, May 1908.

[78] D P Chandler. Redefining relativity: quantitative PCR at low template concentrations for industrial and environmental microbiology. *Journal of Industrial Microbiology and Biotechnology*, 21(3):128–140, September 1998.

[79] Friedrich Schuler, Martin Trotter, Marcel Geltman, Frank Schwemmer, Simon Wadle, Elena Dominguez-Garrido, Maria Lopez, Cristina Cervera-Acedo, Paula Santibanez, Felix von Stetten, Roland Zengerle, and Nils Paust. Digital droplet PCR on disk. *Lab Chip*, 16(1):208–216, 2016.

[80] Slawomir Jakiela, Piotr M. Korczyk, Sylwia Makulska, Olgierd Cybulski, and Piotr Garstecki. Discontinuous Transition in a Laminar Fluid Flow: A Change of Flow Topology inside a Droplet Moving in a Micron-Size Channel. *Physical Review Letters*, 108(13), March 2012.

[81] Olgierd Cybulski and Piotr Garstecki. Transport of resistance through a long microfluidic channel. *Physical Review E*, 82(5), November 2010.

[82] Slawomir Jakiela, Sylwia Makulska, Piotr M. Korczyk, and Piotr Garstecki. Speed of flow of individual droplets in microfluidic channels as a function of the capillary number, volume of droplets and contrast of viscosities. *Lab on a Chip*, 11(21):3603, 2011.

[83] Tomasz S. Kaminski, Ott Scheler, and Piotr Garstecki. Droplet microfluidics for microbiology: techniques, applications and challenges. *Lab Chip*, 16(12):2168–2187, 2016.

[84] Krzysztof Churski, Artur Ruszczak, Slawomir Jakiela, and Piotr Garstecki. Droplet Microfluidic Technique for the Study of Fermentation. *Micromachines*, 6(10):1514–1525, October 2015.

[85] Slawomir Jakiela, Tomasz S. Kaminski, Olgierd Cybulski, Douglas B. Weibel, and Piotr Garstecki. Bacterial Growth and Adaptation in Microdroplet Chemostats. *Angewandte Chemie International Edition*, 52(34):8908–8911, August 2013.

[86] Nicole Pamme, Ryuji Koyama, and Andreas Manz. Counting and sizing of particles and particle agglomerates in a microfluidic device using laser light scattering: application to a particle-enhanced immunoassay. *Lab on a Chip*, 3(3):187, 2003.

[87] Demetri Psaltis, Stephen R. Quake, and Changhuei Yang. Developing optofluidic technology through the fusion of microfluidics and optics. *Nature*, 442(7101):381–386, July 2006.

[88] S. A. Rogers, M. Lisicki, B. Cichocki, J. K. G. Dhont, and P. R. Lang. Rotational Diffusion of Spherical Colloids Close to a Wall. *Physical Review Letters*, 109(9), August 2012.

[89] Maciej Lisicki and Gerhard Nagele. Colloidal Hydrodynamics and Interfacial Effects. In Peter R. Lang and Yi Liu, editors, *Soft Matter at Aqueous Interfaces*, pages 313–386. Springer International Publishing, Cham, 2016.

[90] Fluidigm. Biomark hd system: http://www.fluidigm.com/biomark-hd-system.html, 2013.

[91] Raindance. Raindrop system: http://www.raindancetech.com/products/raindrop.asp, 2013.

[92] Quanta. Quanta life system: http://www.quantalife.com, 2013.

[93] Feng Shen, Bing Sun, Jason E. Kreutz, Elena K. Davydova, Wenbin Du, Poluru L. Reddy, Loren J. Joseph, and Rustem F. Ismagilov. Multiplexed Quantification of Nucleic Acids with Large Dynamic Range Using Multivolume Digital RT-PCR on a Rotational SlipChip Tested with HIV and Hepatitis C Viral Load. *Journal of the American Chemical Society*, 133(44):17705–17712, November 2011.

[94] J Cline. PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Research*, 24(18):3546–3551, September 1996.

[95] Kelly S. Lundberg, Dan D. Shoemaker, Michael W.W. Adams, Jay M. Short, Joseph A. Sorge, and Eric J. Mathur. High-fidelity amplification using a thermostable DNA polymerase isolated from Pyrococcus furiosus. *Gene*, 108(1):1–6, December 1991.

[96] Veronique Picard, Eva Ersdal-Badju, Aiqin Lu, and Susan Clark Bock. A rapid and efficient one-tube PCR-based mutagenesis technique using *Pfu* DNA polymerase. *Nucleic Acids Research*, 22(13):2587–2591, 1994.

[97] WJSW Rychlik, WJ Spencer, and RE Rhoads. Optimization of the annealing temperature for DNA amplification in vitro. *Nucleic acids research*, 18(21):6409–6412, 1990.

[98] Randall K Saiki. The design and optimization of the PCR. In *PCR technology*, pages 7–16. Springer, 1989.

[99] Wayne Grody, Robert Nakamura, Charles Strom, and Federick Kiechle. *Molecular Diagnostics - Techniques and Applications for the Clinical Laboratory*. Elsevier Inc., 2010.

[100] RichardW. Cone, AnnC. Hobson, Meei-Liw. Huang, and MarilynnR. Fairfax. Polymerase chain reaction decontamination: the wipe test. *The Lancet*, 336(8716):686–687, September 1990.

[101] S. Aljanabi. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research*, 25(22):4692–4693, November 1997.

[102] JL Cenis. Rapid extraction of fungal DNA for PCR amplification. *Nucleic acids research*, 20(9):2380, 1992.

[103] Bahram Arezi, Weimei Xing, Joseph A. Sorge, and Holly H. Hogrefe. Amplification efficiency of thermostable DNA polymerases. *Analytical Biochemistry*, 321(2):226–235, October 2003.

[104] J.K. Purzycka, I. Olewiecki, I. Soltyszewski, W. Pepinski, and J. Janica. Efficiency comparison of seven different Taq polymerases used in hemogenetics. *International Congress Series*, 1288:719–721, April 2006.

[105] K. A. Eckert and T. A. Kunkel. DNA polymerase fidelity and the polymerase chain reaction. *PCR methods and applications*, 1(1):17–24, August 1991.

[106] New England BioLabs Inc. Thermophilic dna polymerases: https://www.neb.com/tools-and-resources/selection-charts/thermophilic-dna-polymerases, 2016.

[107] F C Lawyer, S Stoffel, R K Saiki, S Y Chang, P A Landre, R D Abramson, and D H Gelfand. High-level expression, purification, and enzymatic characterization of full-length Thermus aquaticus DNA polymerase and a truncated form deficient in 5' to 3' exonuclease activity. *Genome Research*, 2(4):275–287, May 1993.

[108] Pawel R. Debski, Kamil Gewartowski, Magdalena Sulima, Tomasz S. Kaminski, and Piotr Garstecki. Rational design of digital assays. *Analytical Chemistry*, 87(16):8203–8209, August 2015.

[109] Philippe Corbisier, Leonardo Pinheiro, StÃľphane Mazoua, Anne-Marie Kortekaas, Pui Yan Jenny Chung, Tsvetelina Gerganova, Gert Roebben, Hendrik Emons, and Kerry Emslie. DNA copy number concentration measured by digital and droplet digital quantitative PCR using certified reference materials. *Analytical and Bioanalytical Chemistry*, 407(7):1831–1840, March 2015.

[110] Robert M. Dorazio and Margaret E. Hunter. Statistical Models for the Analysis and Design of Digital Polymerase Chain Reaction (dPCR) Experiments. *Analytical Chemistry*, 87(21):10886–10893, November 2015.

[111] Nivedita Majumdar, Thomas Wessel, and Jeffrey Marks. Digital PCR Modeling for Maximal Sensitivity, Dynamic Range and Measurement Precision. *PLOS ONE*, 10(3):e0118833, March 2015.

[112] W. G. Cochran. Estimation of bacterial densities by means of the "most probable number". *Biometrics*, 6(2):105–116, June 1950.

[113] JL Oblinger and JA Koburger. Understanding and teaching the most probable number technique. *Journal of Milk and Food Technology (JMFT)*, 38(9):540–545, 1975.

[114] C. Goodridge, L. Goodridge, D. Gottfried, P. Edmonds, and J. C. Wyvill. A rapid most-probable-number-based enzyme-linked immunosorbent assay for the detection and enumeration of Salmonella Typhimurium in poultry wastewater. *Journal of Food Protection*, 66(12):2302–2306, December 2003.

[115] Luigi A. Warren, Joshua A. Weinstein, and Stephen R. Quake. The digital array response curve, jun 2007.

[116] Simant Dube, Jian Qin, and Ramesh Ramakrishnan. Mathematical Analysis of Copy Number Variation in a DNA Sample Using Digital PCR on a Nanofluidic Device. *PLoS ONE*, 3(8):e2876, August 2008.

[117] Mano Sivaganesan, Shawn Siefring, Manju Varma, Richard A Haugland, and Orin C Shanks. A Bayesian method for calculating real-time quantitative PCR calibration curves using absolute DNA standards. *BMC Bioinformatics*, 9(1):120, 2008.

[118] CERN. Root environment: http://www.root.cern.ch, 2013.

[119] Bioline. Sensifast sybr no-rox kit: http://www.bioline.com/sg/sensifast-sybr-no-rox-kit.html, 2016.

[120] ThermoFisher Scientific. 7500 fast 7500 real-time pcr system: https://www.thermofisher.com/pl/en/home/life-science/pcr/real-time-pcr/real-time-pcr-instruments/7500-fast-real-time-pcr-system.html, 2016.

[121] GeneProof. Geneproof cytomegalovirus (cmv) pcr kit: http://www.geneproof.com/en/products/diagnostic-kits/microbiological-dna-diagnostics/geneproof-cytomegalovirus-cmv-pcr-kit, 2016.

[122] QIAGEN. Qiaamp dna mini kit: https://www.qiagen.com/gb/shop/sample-technologies/dna/qiaamp-dna-mini-kit/orderinginformation, 2016.

[123] Macherey-Nagel. Nucleomag blood: http://www.mn-net.com/tabid/1350/default.aspx, 2016.

[124] ThermoFisher Scientific. Magjet whole blood gdna kit: https://www.thermofisher.com/order/catalog/product/k2741, 2016.

[125] Jim F. Huggett, Simon Cowen, and Carole A. Foy. Considerations for digital PCR as an accurate molecular diagnostic tool. *Clinical Chemistry*, 61(1):79–88, January 2015.

[126] Timothy J. Henrich, Sebastien Gallien, Jonathan Z. Li, Florencia Pereyra, and Daniel R. Kuritzkes. Low-level detection and quantitation of cellular HIV-1 DNA and 2-LTR circles using droplet digital PCR. *Journal of Virological Methods*, 186(1-2):68–72, December 2012.

[127] Kevin A Heyries, Carolina Tropini, Michael VanInsberghe, Callum Doolin, Oleh I Petriv, Anupam Singhal, Kaston Leung, Curtis B Hughesman, and Carl L Hansen. Megapixel digital PCR. *Nature Methods*, 8(8):649–651, July 2011.

[128] Gudrun Pohl and Ie-Ming Shih. Principle and applications of digital PCR. *Expert Review of Molecular Diagnostics*, 4(1):41–47, January 2004.

[129] Rebecca Sanders, Deborah J. Mason, Carole A. Foy, and Jim F. Huggett. Evaluation of Digital PCR for Absolute RNA Quantification. *PLoS ONE*, 8(9):e75296, September 2013.

[130] Alexandra S. Whale, Simon Cowen, Carole A. Foy, and Jim F. Huggett. Methods for Applying Accurate Digital PCR Analysis on Low Copy DNA Samples. *PLoS ONE*, 8(3):e58177, March 2013.

[131] Bernhard G. Zimmermann, Simon Grill, Wolfgang Holzgreve, Xiao Yan Zhong, Laird G. Jackson, and Sinuhe Hahn. Digital PCR: a powerful new tool for noninvasive prenatal diagnosis? *Prenatal Diagnosis*, 28(12):1087–1093, December 2008.

[132] Benjamin J. Hindson, Kevin D. Ness, Donald A. Masquelier, Phillip Belgrader, Nicholas J. Heredia, Anthony J. Makarewicz, Isaac J. Bright, Michael Y. Lucero, Amy L. Hiddessen, Tina C. Legler, Tyler K. Kitano, Michael R. Hodel, Jonathan F. Petersen, Paul W. Wyatt, Erin R. Steenblock, Pallavi H. Shah, Luc J. Bousse, Camille B. Troup, Jeffrey C. Mellen, Dean K. Wittmann, Nicholas G. Erndt, Thomas H. Cauley, Ryan T. Koehler, Austin P. So, Simant Dube, Klint A. Rose, Luz Montesclaros, Shenglong Wang, David P. Stumbo, Shawn P. Hodges, Steven Romine, Fred P. Milanovich, Helen E. White, John F. Regan, George A. Karlin-Neumann, Christopher M. Hindson, Serge Saxonov, and Bill W. Colston. High-Throughput Droplet Digital PCR System for Absolute Quantitation of DNA Copy Number. *Analytical Chemistry*, 83(22):8604–8610, November 2011.

[133] Laura Miotke, Billy T. Lau, Rowza T. Rumma, and Hanlee P. Ji. High Sensitivity Detection and Quantitation of DNA Copy Number and Single Nucleotide Variants with Single Color Droplet Digital PCR. *Analytical Chemistry*, 86(5):2618–2624, March 2014.

[134] J. Qin, R. C. Jones, and R. Ramakrishnan. Studying copy number variations using a nanofluidic platform. *Nucleic Acids Research*, 36(18):e116–e116, September 2008.

[135] A. S. Whale, J. F. Huggett, S. Cowen, V. Speirs, J. Shaw, S. Ellison, C. A. Foy, and D. J. Scott. Comparison of microfluidic digital PCR and conventional quantitative PCR for measuring copy number variation. *Nucleic Acids Research*, 40(11):e82–e82, June 2012.

[136] Marina Cretich, George G. Daaboul, Laura Sola, M. Selim Unlu, and Marcella Chiari. Digital detection of biomarkers assisted by nanoparticles: application to diagnostics. *Trends in Biotechnology*, 33(6):343–351, June 2015.

[137] Daniel T. Chiu, Robert M. Lorenz, and Gavin D. M. Jeffries. Droplets for Ultrasmall-Volume Analysis. *Analytical Chemistry*, 81(13):5111–5118, July 2009.

[138] Harald Cramer. *Mathematical methods of statistics*. Princeton landmarks in mathematics and physics. Princeton University Press, Princeton, 1999.

[139] C. Radhakrishna Rao. Information and the Accuracy Attainable in the Estimation of Statistical Parameters. In Samuel Kotz and Norman L. Johnson, editors, *Breakthroughs in Statistics*, pages 235–247. Springer New York, New York, NY, 1992.

[140] Adam Merberg and Steven J. Miller. Course notes for math 162: Mathematical statistics - the cramer-rao inequality, available online: https://web.williams.edu/mathematics/sjmiller/public_html/brownclasses/162/handouts/cramerraohan May 2008.

[141] Alexander Ly, Maarten Marsman, Josine Verhagen, Raoul Grasman, and Eric-Jan Wagenmakers. A tutorial on fisher information, available online: http://www.ejwagenmakers.com/submitted/lyetaltutorial.pdf, 2016.

[142] John Duchi. Stanford statistics 311/electrical engineering 377: Fisher information, available online: https://web.stanford.edu/class/stats311/lectures/lec-09.pdf, 2016.

[143] Songfeng Zheng. Math 541: Statistical theory ii - fisher information and cramer-rao bound, available online: http://people.missouristate.edu/songfengzheng/teaching/mth541/lecture_notes/fisher_info.pdf, 2016.

[144] Bio-Rad. Qx200 digital pcr system: http://www.bio-rad.com/en-us/product/qx200-droplet-digital-pcr-system, 2016.

[145] RainDance Technologies. Raindrop digital pcr system: http://raindancetech.com/digital-pcr-tech/raindrop-digital-pcr-system, 2016.

[146] C. E. Shannon. A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3):379–423, July 1948.

[147] C. E. Shannon. Prediction and Entropy of Printed English. *Bell System Technical Journal*, 30(1):50–64, January 1951.

[148] Claude Elwood Shannon. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1):3–55, 2001.

[149] Jianhua Lin. Divergence measures based on the Shannon entropy. *IEEE Transactions on Information theory*, 37(1):145–151, 1991.

[150] Promega. Nucleic acid amplification protocols and applications guide, available online: http://www.promega.com/-/media/files/resources/paguide/letter/chap1.pdf, 2016.

[151] P. Markoulatos, N. Siafakas, and M. Moncany. Multiplex polymerase chain reaction: A practical approach. *Journal of Clinical Laboratory Analysis*, 16(1):47–51, 2002.

[152] Angie L. Bookout, Carolyn L. Cummins, David J. Mangelsdorf, Jean M. Pesola, and Martha F. Kramer. High-Throughput Real-Time Quantitative Reverse Transcription PCR. In Frederick M. Ausubel, Roger Brent, Robert E. Kingston, David D. Moore, J.G. Seidman, John A. Smith, and Kevin Struhl, editors, *Current Protocols in Molecular Biology*. John Wiley & Sons, Inc., Hoboken, NJ, USA, February 2006.

[153] H. R. Garner, B. Armstrong, and D. M. Lininger. High-throughput PCR. *BioTechniques*, 14(1):112–115, January 1993.

[154] T. Morrison, J. Hurley, J. Garcia, K. Yoder, A. Katz, D. Roberts, J. Cho, T. Kanigan, S. E. Ilyin, D. Horowitz, J. M. Dixon, and C. J.H. Brenan. Nanoliter high throughput quantitative PCR. *Nucleic Acids Research*, 34(18):e123–e123, September 2006.

[155] Sandra L. Spurgeon, Robert C. Jones, and Ramesh Ramakrishnan. High Throughput Gene Expression Measurement with Real Time PCR in a Microfluidic Dynamic Array. *PLoS ONE*, 3(2):e1662, February 2008.

[156] Angelika Niemz, Tanya M. Ferguson, and David S. Boyle. Point-of-care nucleic acid testing for infectious diseases. *Trends in Biotechnology*, 29(5):240–250, May 2011.

[157] Seungkyung Park, Yi Zhang, Shin Lin, Tza-Huei Wang, and Samuel Yang. Advances in microfluidic PCR for point-of-care infectious disease diagnostics. *Biotechnology Advances*, 29(6):830–839, November 2011.

[158] Ramakrishna Sista, Zhishan Hua, Prasanna Thwar, Arjun Sudarsan, Vijay Srinivasan, Allen Eckhardt, Michael Pollack, and Vamsee Pamula. Development of a digital microfluidic platform for point of care testing. *Lab on a Chip*, 8(12):2091, 2008.

[159] Yujun Song, Yu-Yen Huang, Xuewu Liu, Xiaojing Zhang, Mauro Ferrari, and Lidong Qin. Point-of-care technologies for molecular diagnostics using a drop of blood. *Trends in Biotechnology*, 32(3):132–139, March 2014.

[160] Paul Yager, Gonzalo J. Domingo, and John Gerdes. Point-of-Care Diagnostics for Global Health. *Annual Review of Biomedical Engineering*, 10(1):107–144, August 2008.

[161] Ott Scheler, Natalia Pacocha, Pawel R. Debski, Artur Ruszczak, Tomasz S. Kaminski, and Piotr Garstecki. Digital counting of bacteria over a broad dynamic range of concentrations, October 2016.

[162] Jack Y. Ho, Nate J. Cira, John A. Crooks, Josue Baeza, and Douglas B. Weibel. Rapid Identification of ESKAPE Bacterial Strains Using an Autonomous Microfluidic Device. *PLoS ONE*, 7(7):e41245, July 2012.