Ernst-Moritz-Arndt-Universität Greifswald
Mathematisch-Naturwissenschaftliche Fakultät

PhD Thesis

Mathematical Modeling of Microarray Experiments

Mathematical Modeling of Microarray Experiments

I n a u g u r a l d i s s e r t a t i o n

zur

Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Ernst-Moritz-Arndt-Universität Greifswald

vorgelegt von
Robert Bialowons
geboren am 27.10.1979
in Demmin

Greifswald, 27.01.2010

Dekan: .....

1. Gutachter: .....

2. Gutachter: .....

Tag der Promotion: .....

Acknowledgements

First of all, I would like to thank my supervisor, Prof. Dr. V. Liebscher for his continuous support during the past four and a half years. He was always there to give advice and to listen. He also gave me scientific freedom and supported my ideas. I also have to thank him for tolerating my self-employment and my excursions into the studies of business administration.

A special thank goes to Dr. S. Kläre for proofreading and for many helpful comments.

My colleagues supported me during my research work. Especially, I thank M. Kläre, N. Kosytsina, E. Herrholz, P. Nestler and Dr. B. Alexandrov for fruitful discussions and valuable hints as well as for their help with LaTeX, Mathematica, Matlab and R.
I am grateful to Dr. A. Pototskyy for his comments on the PDE of the hybridization process.

Further, I thank A. Bösel for helping me to understand the physical background and A. Reeder and Dr. D. Höper for explaining some aspects of microarray data analysis.

For the emotional support I thank my family and friends.

Last but not least, I thank the Department of Mathematics and Computer Science, Ernst-Moritz-Arndt-University Greifswald for the financial support.

Contents

List of Figures

Introduction

Microarray experiments have become a major analyzing method in various fields of research such as biology, medicine or pharmaceutics. Even though they are used very often, only little is known about the character of their underlying subprocesses. This is a major objection to the validity of inferences from microarray experiments.

This work is concerned with modeling respective subprocesses of microarray experiments as well as the analysis of the corresponding results including comparisons to usual inferences by common analyzing methods. It shall help to understand the character of the underlying subprocesses and thus give some advice to researchers about the design of microarray experiments and the choice of analyzing methods.

Characterization of microarray experiments

In this section we will characterize the fields of application and the mode of operation of microarray experiments. For a more detailed introduction to this topic see for example [MüNi].

The major purpose of microarray experiments is to discover gene mechanisms in organisms. For example, researchers try to answer questions such as

- Which genes does Bacillus subtilis use to handle salt stress? (biology, [Hahne])
- Which genes are involved in the generation of breast cancer? (medicine, [Welch])
- Which human gene expressions are affected by a new drug? (pharmaceutics, [Chavan])

More precisely, such genes and respective proteins shall be identified which are used by a cell during distinct cell states. Cell states are affected by environmental conditions such as concentrations of chemical substances (salt,

drugs, hormones,...) in the near neighborhood of the cell and by physical conditions (temperature, pressure, light,...) as well as by tissue types (skin, lung, blood,...). Microarray experiments make use of the fact that the same genes are expressed differently in cells at different cell states. Before describing the mode of operation we will explain some terms which are important to understand the biochemical processes linked to microarray experiments.

Genes and DNA

Genes are encoded in DNA (deoxyribonucleic acid) of an organism.[1] Typically, DNA is double-stranded where each strand is a concatenation of nucleotides. DNA nucleotides consist of three portions, a sugar (deoxyribose), a phosphate and one out of four bases. Possible bases in DNA molecules are adenine (A), thymine (T), guanine (G) and cytosine (C). The bases of one strand are able to form two different kinds of pairs (*complementary base pairs*) with the bases of another strand, i.e. the A-T pair and the G-C pair. The former is linked by two and the latter by three hydrogen bonds. Thus, complementary base pairs are responsible for the linkage of two DNA strands. For further details see [Camp] and for illustration see Figure 1. Consequently, DNA molecules with a high percentage of G-C are more sta-

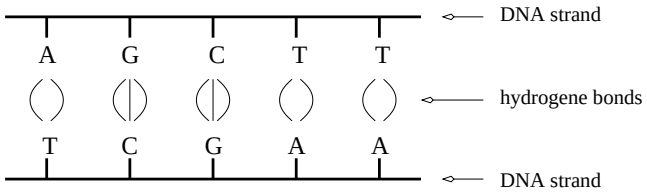


Figure 1: Sketch of the 2-dimensional structure of a DNA molecule.

ble than molecules with lower percentages. This becomes important if one tries to denature DNA molecules (split the two strands). More energy is needed to denature DNA with a higher GC content.

The actual genetic information is encoded in the sequence of the bases within the strands. Since other base pairs than G-C and A-T cannot be formed the sequence of one strand can be inferred by the sequence of the other strand and vice versa.

The decoding of the DNA's information inside an organism is basically achieved by two processes, *transcription* and *translation*. During transcription the DNA is copied into mRNA (messenger ribonucleic acid) by comple-

[1] An exception to this rule are RNA-viruses, whose genetic information is encoded in strands of RNA (ribonucleic acid).

mentary base pairing, i.e. the mRNA has the same sequence like one of the DNA strands with the exception of thymine being substituted by uracil.

At this point it is necessary to give some information on RNA and its connection to DNA. RNA and DNA are both so-called nucleic acids. Summarizing the differences we find RNA to be single-stranded, the sugar in RNA is ribose instead of deoxyribose and as already mentioned thymine is substituted by uracil. An illustration of RNA and DNA nucleotides as well as of the respective bases can be found in Figure 2. During translation the

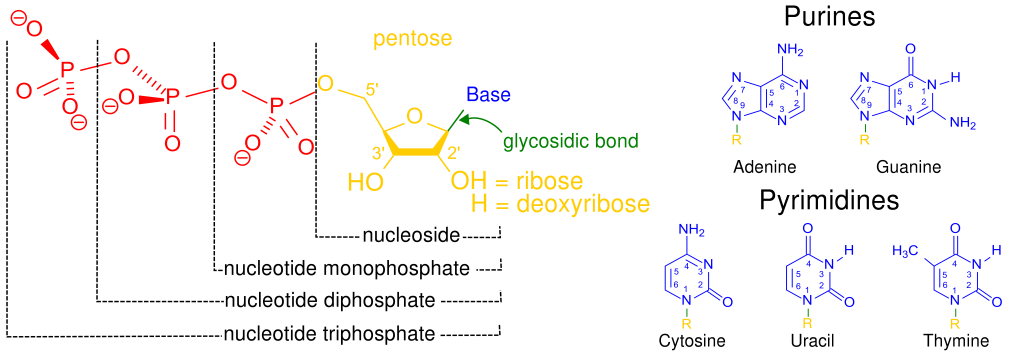


Figure 2: The chemical structure of RNA and DNA nucleotides from [WiNu].

mRNA is translated into a chain of amino acids, the protein. Here, the mRNA's base sequence determines the sequence of amino acids within the protein. Proteins are responsible for the reaction on the change of cell states. So, if an organism reacts to a change with the multiplication of the production of a specific protein, in advance there must be a multiplication of the corresponding mRNA. The change in the amount of mRNA of a certain type is supposed to be directly proportional to the change of the amount of the corresponding protein. Subsequently, we will show how microarrays use the conjecture to infer the change in the amount of proteins an organism synthesizes during a specific cell state. Verifying the conjecture will be one of the major tasks of this work.

Microarrays

In microarray experiments the change in the amount of mRNA is estimated in the following fashion. Typically, either a gene library[2] of the examined

[2]A gene library consists of pieces of an organism's DNA representing the entire genome or the part of the genome one is interested in. The genome is defined as the entire hereditary information encoded in DNA.

organism or artificially synthesized pieces of DNA are fixed to an appropriate surface. Both cases use single strand DNA. See Figure 3 for illustration.

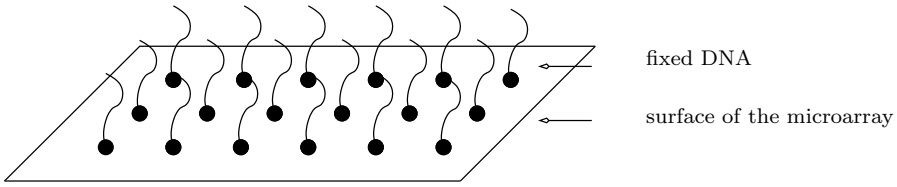


Figure 3: Sketch of a microarray without hybridizations.

The fixed DNA is called the probe and its sequence is well known, which is an essential fact as can be seen later. The surface with the immobilized probes is called the microarray. On the microarray, the probes are organized in circle objects. These objects are called spots. Each spot contains probes of a single kind. On many microarrays spots occur doubly to receive more data and thus grant more reliability. An illustration of the organization of spots on a microarray can be seen in Figure 4. Microarrays can either be prepared by the researcher himself or purchased from specialized companies.

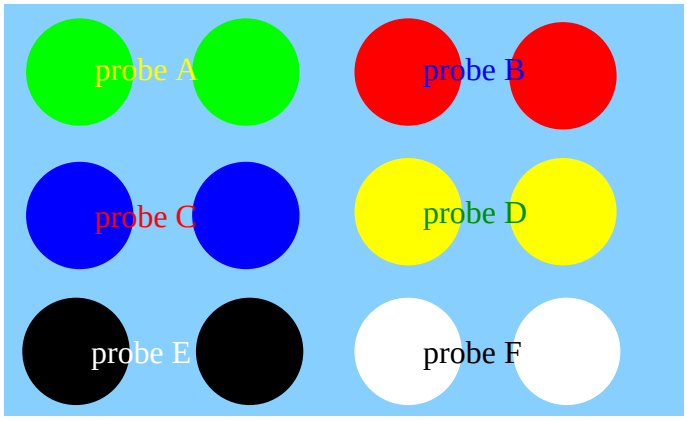


Figure 4: Sketch of the spot organization on a microarray.

The entire mRNA of the organism which has been transcribed during the respective cell states is extracted and reversely transcribed into cDNA (complementary DNA), while feeding labeled nucleotides to the reverse transcriptase. The reverse transcriptase is an enzyme (catalytic protein). It is used to catalyze the transcription process from mRNA to cDNA. cDNA is called the *target*. Often used labels are radioactive substances or the fluorescent dyes Cy3 (green fluorescent) and Cy5 (red fluorescent). For a list

and description of other methods see [Prietz], chapter 4. At this point it is important to mention that the extracts of the cells originating from different cell states are kept separately and then are labeled differently. Sometimes the reverse transcription step is omitted and the $mRNA$ is chosen to be the target. In this work, reverse transcription is the first subprocess and is examined accordingly.

If incubated under sufficient heat, cDNA will become single-stranded and thus during cooling can be hybridized to DNA with a complementary base sequence fixed on the microarray (see Figure 5). This reaction is called hybridization and is the second subprocess we will look at. During this step the targets from the different cell states compete for free probes on the array.

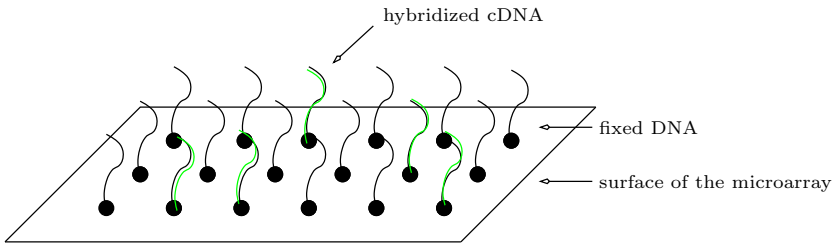


Figure 5: Sketch of a microarray with hybridizations.

Non hybridized targets and other materials are washed off the surface, while hybridized targets stay on the array due to their linkage to the immobilized probes. This reaction characterizes the third subprocess, the washing.

Next, hybridized cDNAs are detected by scanning for the targets. This is accomplished by using the physical characteristics of the labels. Radioactive labels radiate per se whereas fluorescent dyes have to be stimulated by lasers. This work is restricted to detection with fluorescent dyes since this is the most commonly used labeling method. The fluorescent reaction is the fourth subprocess.

The fifth subprocess is the detection itself. It is achieved either by CCD (charge-coupled device) cameras or by photomultiplier tubes. The results are signal intensity values, two for each spot due to the two dyes. These intensities are visualized as shown in Figure 6. This work only deals with the photomultiplier tube. Here, the spots in the picture represent the spots on the microarray, maintaining the spatial information. Different spot colors represent different labels whereas the brightness of spots is directly proportional to the intensity detected. Thus, the genes corresponding to bright spots are supposed to be expressed strongly.

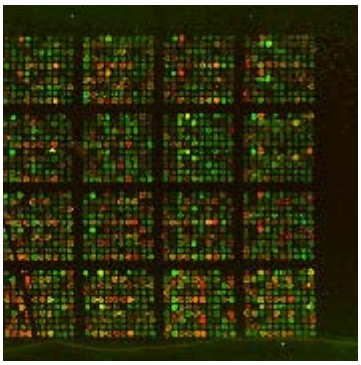


Figure 6: Visualization of a microarray from [Alhad].

Goal and structure of this work

The assumption of many researchers is that signal intensities from microarrays are directly proportional to the amount of the respective mRNA molecules (see for example [Bronch], [ZveBa] or [McLach]). This enables them to use linear models to describe the output signal (see [Speed] for a detailed overview or [Ochs] for an example). There are hints which are contradictory to using linear models. E.g., the dynamic of particles during the hybridization step might behave non linear. Linear models are still a common tool since their analysis is simple.

Many publications distance themselves from the assumption of direct proportionality. They propose more complex models to describe the relationship between the amount of mRNA molecules and detected signal intensities. We will give two examples. [Held] use an ordinary differential equation to model the hybridization and washing step. Their model does not only depend on the amount of mRNA molecules but also on various physical properties. [DaViHa] use stochastic differential equations to describe the long term behavior of binding and release processes such as the hybridization reaction.

The main aim of this work is to mathematically analyze the assumption of direct proportionality by investigating the five subprocesses, i.e. *reverse transcription, hybridization, washing, fluorescence and detection.*

In the first chapter we describe the subprocesses with the help of models taken from the literature if possible or developed by ourselves. We put the main focus on the hybridization process since it is the basic principle behind microarray reactions. The idea to model the hybridization by a Markov process is borrowed from [ReWi]. We extend their model by an additional transition, which describes the release of targets from the spot. The transi-

tion rates are adjusted adequately.

The other subprocesses are summarized as *residual subprocesses*. Here, we start with the reverse transcription which we describe by a Markov process, too. We state a binomial model for the washing process. The fluorescence is divided into two parts, the laser light reaction and the fluorescence itself. The laser light is modeled by an ordinary differential equation with a noise term ([SaTe]) whereas the fluorescence intensity is determined by a heuristic equation ([Schwedt]). Finally, on the one hand, we describe the detection and attached noise sources with heuristic results from [SauWei], [Uiga], [BiSchl] and [SiSu], and on the other hand we use the branching process from [MaTeSa] as alternative description.

In the second chapter we analyze the models from chapter one and try to quantify the noise added to the signal. In the end of each section we will give a short summary of the results.

Again, starting with the hybridization process as the most important process of the microarray experiment, in a first step we discuss its parameter situation. Afterwards, its stationary distribution is calculated for different parameter settings. Then, we use a result from [Kurtz] to approximate the Markov process by a deterministic process. The respective stationary points are determined for the same parameter settings and in general. Then, they are compared to the stationary distribution. In addition, the existence and uniqueness of the stationary point is shown. Afterwards, we apply another limit to approximate the process by a partial differential equation (PDE). We show that the stationary points from the first limit are consistent to the distributional solution of the PDE. In a last step the results are verified by simulating the entire process.

We begin the analysis of the reverse transcription process by investigating its dependency on the parameters involved. A common perturbation approach of the rates of the process is applied. A Taylor approximation and a statistical test are used to examine the impact of the perturbation. In addition we propose the choice of parameter values by minimizing the area under the ROC curve of the test. Finally, an estimator for the amount of input particles of the reverse transcription process is proposed if a certain output is measured. The distribution of the estimator is determined.

The binomial model of the washing step is analyzed by determining the mean and the variance of the particle distribution for realistic parameter situations. We determine these values for increasing detergent intensities and reproduce observations made by biologists in washing experiments (see [Drob]). It can be seen that only within a small range of washing intensities the correct signal can be achieved.

We combine the solution of the ordinary differential equation of the laser light intensity with the heuristic equation of the fluorescence intensity. On this basis we develop a correction factor which is a measure for the noise added by the process if two signals are compared. For illustration, we give an example for two signal intensities and determine the correction factor for this situation.

Last but not least, the detection model is analyzed. Here, we restrict the analysis to the branching process. We determine mean, variance, skewness and kurtosis of the number of output particles with the help of the probability generating function. These characteristics are used to verify the approximation by a normal distribution for realistic parameter situations. We give an example for two signals passing the detection aperture for different parameter values. Finally, we investigate the ratio of two signals and derive its probability distribution.

In chapter 3 we will give an example of the signal passing through all subprocesses using the models from the first chapter and the results of the analysis from the second chapter.

For each subprocess the frequencies of the particles involved and of the respective intensity ratios are determined. In addition the noise due to each subprocess is specified and respective confidence intervals are determined. Finally, a conclusion is drawn.

In the last chapter we discuss the results of this work and point out some weaknesses of the models. Lastly, we give an outlook to future work.

Chapter 1

Modules

In this section the five subprocesses which were described in the introduction are modeled and an overview to sources of noise is given.

As already mentioned, the main focus of attention will be the hybridization process. Its dynamics is responsible for the number of targets hybridized to the spot. Thus if the other processes do not add too much noise, it will mainly determine the final signal.

In a microarray experiment the original gene expression level is transformed by different subprocesses into the final signal which is detected. Knowing only the final signal, a deep understanding of the underlying transformation processes is essential to infer the original gene expression level. In the sequel of this chapter, we will look at the entire process ranging from amplifying the mRNA to detecting the intensities from the microarray. In order to understand this process, it is divided into five subprocesses (modules). The modules are:

- reverse transcription,
- **hybridization**,
- washing,
- fluorescence and
- detection.

In order to motivate the modeling, for each module we will point out some of the problems, which cause noise and inhomogeneities of the detected signal at the end of a microarray experiment.

Firstly, we will look at the reverse transcription reaction. This module is necessary to solve the problem of unstable mRNA by reversely transcribing it into its more stable version, cDNA. Here, reading errors might occur,

which lead to a wrong sequence of nucleotides. Further, for later detection of the amount of cDNA, it must be modified during the reverse transcription. Often used methods employ the incorporation of dyes into the cDNA. Mark, different dyes have different incorporation efficiencies which has to be accounted for.

After reverse transcription, the targets are hybridized to the probes. Only complementary targets are supposed to bind to the probes on one spot. In practice, targets with similar sequences also hybridize. This process is unwanted. It is called cross-hybridization and is considered later. Since hybridization works with hydrogen bonds between the bases adenine and thymine (two bonds) and between the bases cytosine and guanine (three bonds), its strength depends on the number of hydrogen bonds and therefore on the amount of the different bases, the length of the probes and the length of the targets. But also the temperature, the time span and the competition with similar targets might influence the hybridization process.

The next step is the washing procedure. Here, non-hybridized targets and other chemicals are removed from the microarray with the help of detergents. But also hybridized targets might be removed if their binding to probes is not strong enough or the detergent is too strong.

To get a signal from the microarray, the labels which were incorporated during the reverse transcription step are scanned for. In the case of dye labels, a stimulation step by a laser light causes fluorescence of the dyes. Different dyes have different absorption spectra and are therefore stimulated at different wavelengths. But also the laser power, which is a measure for the intensity of the laser is important.

The light signal is transformed into an electron current and often multiplied by a photomultiplier tube (PMT). Afterwards, the current is detected by an amperemeter. The strength of the multiplying effect is supposed to be proportional to the voltage of the PMT. Our modularization of the microarray process is depicted in Figure 1.1.

Above, we listed some of the noise sources which perturb the signal of the microarray experiment. In order to describe their impact, in the following we present models for the modules. Due to its importance to the dynamic of the microarray experiment, we will begin with the hybridization process.

1.1 Hybridization

The DNA probes which are immobilized on a microarray slide are organized in regions of circular shape, each containing only a single kind of DNA. These areas are called *spots*. During the hybridization reaction cDNA molecules

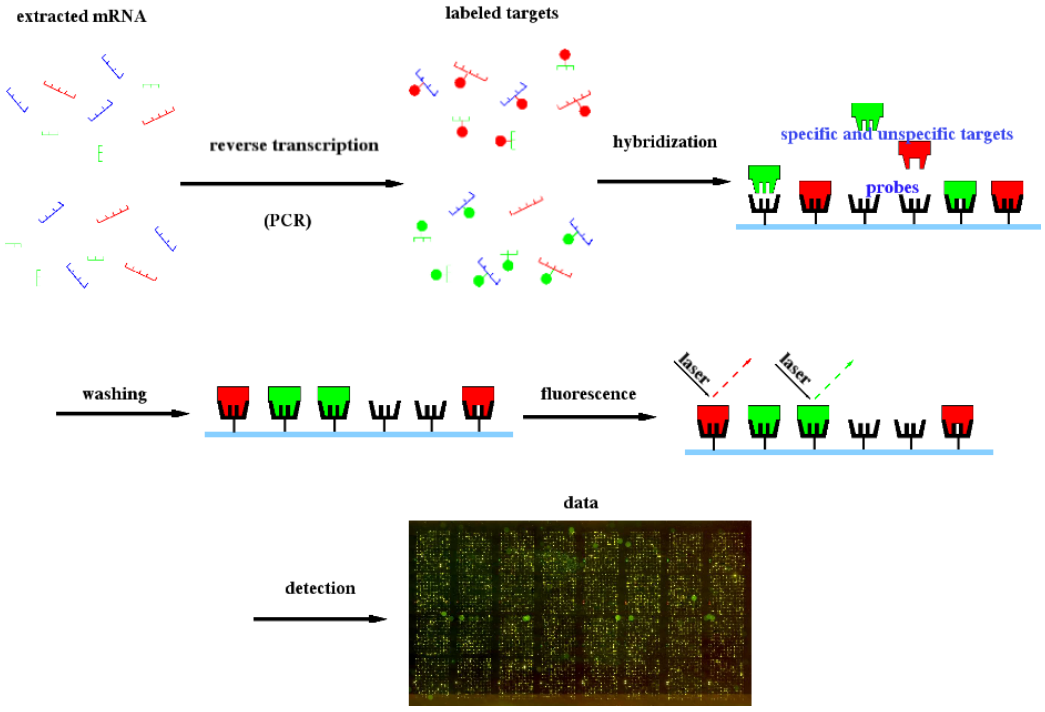


Figure 1.1: Sketch of the microarray process.

approach a small neighborhood of a spot, which allows them to interact with the DNA of the spot. According to the number of complementary bases within the target and probe sequences a target molecule might bind to a probe molecule on the spot, where the probability of binding should increase with the degree of complementarity.

1.1.1 A hybridization model from [ReWi]

Let m denote the number of different target types and S the number of probes per spot. [ReWi] developed a reasonable model which describes the hybridization reaction on a single spot with the help of a continuous-time Markov process $\{N(t) = (N_1(t), N_2(t), \ldots, N_m(t)) \mid t > 0\}$ on the discrete state space

$$\Sigma_{S,m} = \{N = (N_1, N_2, ..., N_m) \in \mathbb{N}_0^m : \sum_{i=1}^{m} N_i \leq S\} \tag{1.1}$$

where $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. Let t_n denote the event time of the nth event. The number of events is countable and the respective event times fulfill $0 < t_1 <$

$t_2 < ....$ In addition, there are finitely many events in a bounded interval. In the context of the hybridization model, $N(t_n) = (N_1(t_n), N_2(t_n), ..., N_m(t_n))$ denote the number of different target species $1, 2, ..., m$ hybridized to the probes on the spot at time t_n. A Markov process is conveniently described by its transition rates which are in our case defined as follows

$$r_{a,b} := \frac{d}{dt_n}\mathbb{P}(N(t_n) = b \mid N(0) = a).$$

for a transition from state a to state b. For a more detailed theory of Markov processes see [ChWa]. Figure 1.2 illustrates the state space in two dimensions including two examples for possible transitions.

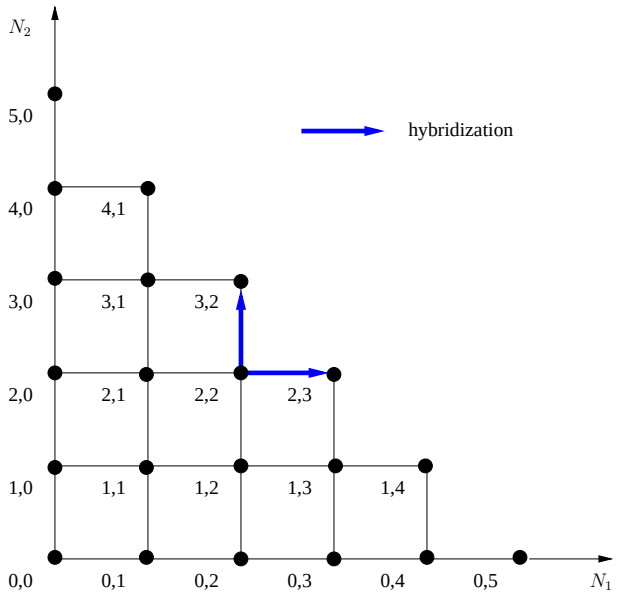


Figure 1.2: A two dimensional lattice ($m = 2$) as state space for $S = 5$ probes on the spot. Blobs and numbers represent the states of the hybridization process and lines represent possible transitions between states. Blue arrows are examples of hybridization events.

Assume, that there are $T_i > S$ cDNAs of type i, $i = 1, 2, ..., m$. The resulting percentages $p_i(n)$ of type i targets within all non-hybridized targets and $q(n)$ of non-hybridized probes on the spot are

$$p_i(n) = \frac{T_i - N_i(t_{n-1})}{\sum_{j=1}^{m}(T_j - N_j(t_{n-1}))} \text{ and}$$

$$q(n) = \frac{S - \sum_{j=1}^{m} N_j(t_{n-1})}{S}.$$

The inter arrival times $t_{n+1} - t_n$ are independently exponentially distributed with parameter

$$r_1 = \lambda \sum_{j=1}^{m} \left(T_j - N_j(t_{n-1})\right),$$

where the sum $\sum_{j=1}^{m} \left(T_j - N_j(t_{n-1})\right)$ denotes the number of free targets and $\lambda > 0$ is the recruitment rate for a single target.

[ReWi] tried to derive λ with the help of collision theory. They defined λ to be the ratio of the mean collision time Θ of a single target within a neighborhood D of the spot and the mean sojourn time $\hat{\tau}(T) > 0$ of this target in D within a time interval of length T. Unfortunately, free parameters like Θ could not be motivated. Thus we might as well restrict our considerations to the model with λ in order to keep the number of free parameters small. We assume that λ only has an effect on the speed of the process rather than on the composition of targets on the spot.

Let $\pi_j > 0$ be the probability of a target of type j binding to a probe on the spot, whenever there is a collision between these two particles. It increases with the number of possible hydrogen bonds with the probe.

There are two possible events which might take place whenever a target comes close to a probe.

1. The probe, which is approached by a target of type i, is not yet hybridized and the target binds to it. The new state will be $N(t_n) = N(t_{n-1}) + e_i$, where $e_i \in \mathbb{R}^m$ is the ith unit vector. The probability for this transition is

   $$\hat{P}_n(i, +1) = \pi_i p_i(n) q(n),$$

   where $(i, +1)$ denotes the event of one additional target of type i hybridizing to the spot.

2. Nothing happens, i.e. the approaching target does not bind. The new state will be $N(t_n) = N(t_{n-1})$. The probability for this transition is

   $$\hat{P}_n(\cdot) = 1 - \sum_{j=1}^{m} \hat{P}_n(j, +1).$$

Since already bound targets might dissociate from the probe, the model of [ReWi] should be extended by the incorporation of dissociation events as follows.

1.1.2 A hybridization model with dissociation

We will start with the model from [ReWi] and will add extra transitions which enable targets to dissociate from the spot. The new process will be a Markov process, too. We can adopt the same state space $\Sigma_{S,m}$ but have to modify the transitions subsequently. The state space in two dimensions including possible transitions is illustrated in Figure 1.3. Let $\gamma_j > 0$ be

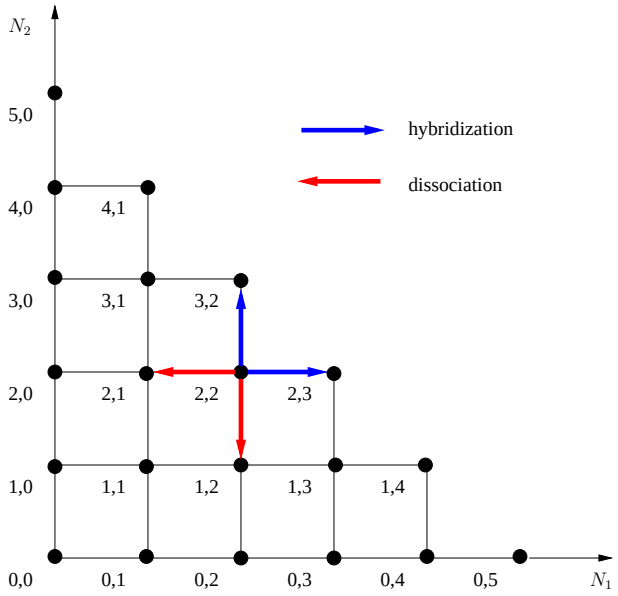


Figure 1.3: A two dimensional lattice ($m = 2$) as state space for $S = 5$ probes on the spot. Blobs and numbers represent the states of the hybridization process and lines represent possible transitions (including dissociation events) between states. Blue and red arrows are examples of hybridization and dissociation events, respectively.

the rate for a bound target of type j to dissociate. The resulting rate for dissociations of any of the targets will be

$$r_2 = \sum_{j=1}^{m} \gamma_j N_j(t_{n-1}).$$

Thus, the rates for the inter arrival times have to be modified in order to account for the additional transition. The new rate r will be the sum of the two rates r_1, r_2 according to superposition of independent Poisson processes

(see [Hueb], chapter 8)

$$\begin{aligned} r &= r_1 + r_2 \\ &= \lambda \sum_{j=1}^{m} (T_j - N_j(t_{n-1})) + \sum_{j=1}^{m} \gamma_j N_j(t_{n-1}). \end{aligned} \tag{1.2}$$

The resulting process is still a Markov process in continuous time but at a higher rate. Assume the process is in state $N(t_{n-1})$ at time t_{n-1}. Three different transitions are possible.

1. A target of type i approaches and binds to a probe, which is not yet hybridized to another target. The new state will be $N(t_n) = N(t_{n-1}) + e_i$, where $e_i \in \mathbb{R}^m$ is the ith unit vector. The probability for this transition is
$$P_n(i, +1) = \hat{P}_n(i, +1)\frac{r_1}{r}. \tag{1.3}$$

2. Nothing happens, i.e. the approaching target does not bind. The new state will be $N(t_n) = N(t_{n-1})$. The probability for this transition is
$$P_n(\cdot) = \hat{P}_n(\cdot)\frac{r_1}{r}. \tag{1.4}$$

3. An already bound target of type i dissociates from the spot. The new state will be $N(t_n) = N(t_{n-1}) - e_i$. The probability for this transition is
$$P_n(i, -1) = \frac{\gamma_i N_i(t_{n-1})}{r}. \tag{1.5}$$

This model shall be analyzed in the 2nd chapter. Even though the hybridization model is stated for an arbitrary number of different target types, we will restrict the analysis to the case of only two types. On the one hand, it can be used to investigate the labeling effect and on the other hand, it is simple enough to do proper numerical analysis.

1.2 Residual subprocesses

1.2.1 Reverse Transcription

After opening the cells of the organisms of interest, the mRNA is extracted as an indicator of the gene expression activity. The extracted mRNA is quite unstable and therefore has to be transformed into a stable molecule, preserving the information of the mRNA sequence. This is achieved by reverse

transcription. Here, the extracted mRNA is reversely transcribed into cDNA via complementary base pairing with the help of a viral enzyme. This enzyme is called reverse transcriptase and is able to merge single deoxy-nucleotides into single stranded cDNA, complementary to the original mRNA sequence.

In microarray experiments one kind of nucleotide is chosen to be labeled with a fluorescence dye in order to detect and quantify cDNA molecules. These nucleotides as well as unlabeled nucleotides (of the same and of the three other kinds) are fed to the reverse transcriptase to be incorporated into the cDNA strand. This leads to a competition of labeled and unlabeled nucleotides. Labeled nucleotides are bigger, i.e. they diffuse slower and are incorporated less efficiently by the reverse transcriptase. See Figure 1.4 for illustration.

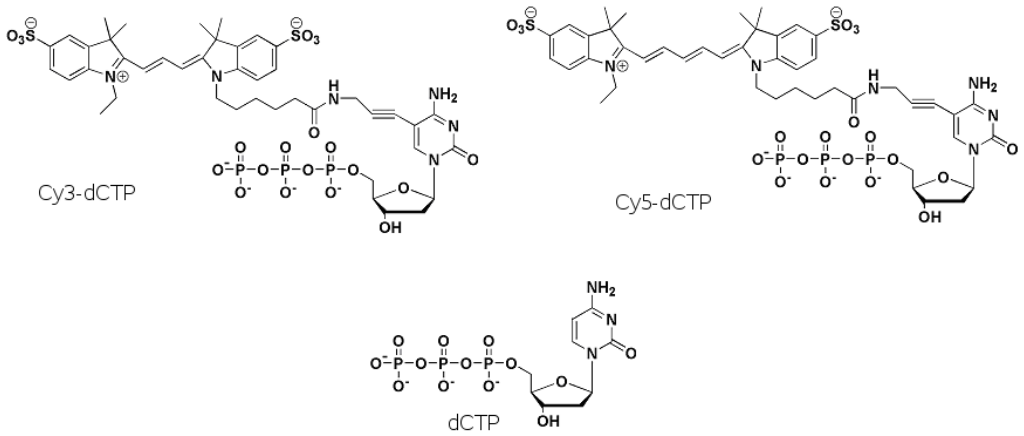


Figure 1.4: Chemical structure of labeled (with Cy3 and Cy5) and unlabeled Deoxycytidine triphosphates (dCTP).

In addition incorporation of labeled nucleotides is forbidden if the incorporation positions are too close to each other. The reason is a steric conflict of the three-dimensional structure of cDNA and the shape of labeled nucleotides. For simplicity, this effect shall be neglected.

1.2.1.1 A reverse transcription model

The following model describes the process of reverse transcription mathematically. We state a continuous time, discrete state Markov process but eventually we look at the embedded Markov chain in discrete time. During detection the signal is caused by fluorescence of labeled nucleotides. Thus, we only have to look at those positions within the mRNA sequence, which are able to serve as a template for labeled nucleotides. Let m be the number

of such positions within a certain mRNA strand. Unlabeled nucleotides are denoted by u and labeled nucleotides by l. The respective state space is $\{0, 1, ..., m\}$. The rates of the process can be derived as follows. Initially, there are V_i nucleotides of type i, $i \in \{u, l\}$ which are free to react with mRNA molecules. Assume $V_i \gg m$, $i \in \{u, l\}$. Let $z(k)$ be the kind of nucleotide incorporated at position k, $k \in \{1, 2, ..., m\}$. Thus, the number of l-nucleotides incorporated up to position k (including k) is

$$Z(k) = \sum_{i=1}^{k} \mathbb{1}_l(z(i))$$

where $\mathbb{1}_l(z(i))$ denotes the indicator function

$$\mathbb{1}_l(z(i)) = \begin{cases} 1 & \text{if } z(i) = l, \\ 0 & \text{else.} \end{cases}$$

At this moment, the resulting number of free nucleotides is $V_u - (k - Z(k))$ for the unlabeled and $V_l - Z(k)$ for the labeled type. Single nucleotides of type i, $i \in \{u, l\}$ are recruited (bound after approaching) by the enzyme with rate r_i. Thus, the rate for recruiting a nucleotide of type i at position $k + 1$ is the product of r_i and the number of free nucleotides of the respective type, i.e. $r_u(V_u - (k - Z(k)))$ for the unlabeled and $r_l(V_u - Z(k))$ for the labeled type. The resulting probability $q_i(k + 1, Z(k)) := P(z(k + 1) = i \mid Z(k))$ of the recruited molecule at position $k + 1$ being of type i conditional on the history $Z(k)$ until position k will be the relative rate, i.e.

$$q_i(k+1, Z(k)) = \begin{cases} \frac{r_u(V_u-(k-Z(k)))}{r_u(V_u-(k-Z(k)))+r_l(V_l-Z(k))} & \text{if } i = u, \\ \\ \frac{r_l(V_l-Z(k))}{r_u(V_u-(k-Z(k)))+r_l(V_l-Z(k))}, & \text{if } i = l \end{cases}$$

with $Z(0) \equiv 0$. Obviously, q_u and q_l are probabilities and fulfill

$$q_u(k + 1, Z(k)) + q_l(k + 1, Z(k)) = 1.$$

Since we are not interested in the time, we will restrict further investigations to the embedded markov chain.

Once a nucleotide of type i, $i \in \{u, l\}$ is recruited by the enzyme, it can either be incorporated into the cDNA sequence with probability p_i or not with probability $1 - p_i$. Note, $r_l < r_u$ as well as $p_l < p_u$ due to the size and structure of the labeled nucleotides which are disadvantageous for the diffusion towards the enzyme and the subsequent incorporation reaction catalyzed by the enzyme. So, whenever the reverse transcriptase prepares for the reaction at position $k + 1$, three different events can occur.

1. A nucleotide of type u is recruited and linked with the cDNA molecule, which happens with probability $q_u(k+1, Z(k))p_u$. The enzyme moves on to position $k+2$ and $Z(k+1) = Z(k)$.

2. A nucleotide of type l is recruited and linked with the cDNA molecule, which happens with probability $q_l(k+1, Z(k))p_l$. The enzyme moves on to position $k+2$ and $Z(k+1) = Z(k)+1$.

3. The recruited nucleotide dissociates and is not attached to the cDNA molecule. The enzyme stays at position k and recruits the next nucleotide. The probability for this event is $q_u(k+1, Z(k))(1-p_u) + q_l(k+1, Z(k))(1-p_l)$.

Therefore, if the cDNA sequence is known until position k, the probability distribution of $z(k+1)$ can be determined by calculating the limit of the geometric series

$$\begin{aligned}
&\mathbb{P}(z(k+1) = i | Z(k)) \\
&\quad = q_i(k+1, Z(k))p_i \sum_{n=0}^{\infty} \left(q_u(k+1, Z(k))(1-p_u) + q_l(k+1, Z(k))(1-p_l)\right)^n \\
&\quad = \frac{q_i(k+1, Z(k))p_i}{1 - \left(q_u(k+1, Z(k))(1-p_u) + q_l(k+1, Z(k))(1-p_l)\right)} \\
&\quad = \frac{q_i(k+1, Z(k))p_i}{q_u(k+1, Z(k))p_u + q_l(k+1, Z(k))p_l}. \qquad (1.6)
\end{aligned}$$

Remark: *Obviously, Equation (1.6) is independent of the third transition. Thus, if we are only interested in the number of labeled nucleotides which are incorporated till position k, this transition can be omitted.*

Let $\mathbb{P}(Z(k+1) = i) \equiv 0$ if $k+1 < i$. Thus, the probability for the number of l-nucleotides incorporated up to position k can be determined recursively as

$$\begin{aligned}
\mathbb{P}(Z(k+1) = i) \quad = \quad & \mathbb{P}(Z(k) = i)\mathbb{P}(z(k+1) = u | Z(k) = i) \\
& + \mathbb{P}(Z(k) = i-1)\mathbb{P}(z(k+1) = l | Z(k) = i-1) \qquad (1.7)
\end{aligned}$$

with initial probabilities

$$\mathbb{P}(Z(1) = i) = \begin{cases} q_u(1,0)p_u \text{ if } i = 0, \\ q_l(1,0)p_l \text{ if } i = 1. \end{cases}$$

This recursion cannot be simplified in order to determine the probability distribution of $Z(m)$ in general. But it will be sufficient to determine the distribution if the parameters of the model are known.

To approximate the distribution of $Z(m)$, it is possible to simplify the model as follows. Since $V_i \gg m$, $i \in \{u, l\}$, assume the number of free nucleotides is constant at all times during reverse transcription. Thus, the probability for the recruited molecule at position k being of type i will also be constant, i.e.

$$q_i(k, Z(k-1)) = \frac{r_i V_i}{r_u V_u + r_l V_l} =: q_i, \ i \in \{u, l\}.$$

As a result the distribution of $Z(m)$ becomes a binomial distribution and thus is given by

$$\mathbb{P}(Z(m) = j) = \binom{m}{j} q_l^j q_u^{m-j}. \tag{1.8}$$

Therefore, $Z(m)$ has mean

$$\mathbb{E}(Z(m)) = m q_l \tag{1.9}$$

and variance

$$\mathbb{V}ar(Z(m)) = m q_l q_u. \tag{1.10}$$

For details see [Kren].

The distribution of $Z(m)$ will be analyzed in Chapter 2.

1.2.2 Washing

Once the DNA has hybridized to the probes on the microarray, non-hybridized DNA and other materials have to be washed off the array.

For this purpose detergents like SDS (sodium dodecyl sulfate) are used. These molecules are amphophilic, i.e. they have a hydrophobic part which binds water-insoluble molecules like DNA and a hydrophilic part which is water soluble. The water-insoluble molecules are covered entirely by SDS and a so-called micelle is formed, which can be washed off the surface of the microarray. See Figure 1.5 for illustration.

Obviously, the degree of efficiency of the washing procedure depends on the concentration of the detergent. Too low concentrations lead to incomplete or missing micelles around the water-insoluble molecules whereas too high concentrations might even dissolve hybridized cDNAs. Either ways are accompanied by an improper signal.

But not only the concentration of the detergent is important. The length of the cDNA molecules and their mixture of the four bases is of major interest to the strength of binding to the immobilized DNA on the surface of the microarray. Since targets hybridize to the probes via hydrogen bonds between

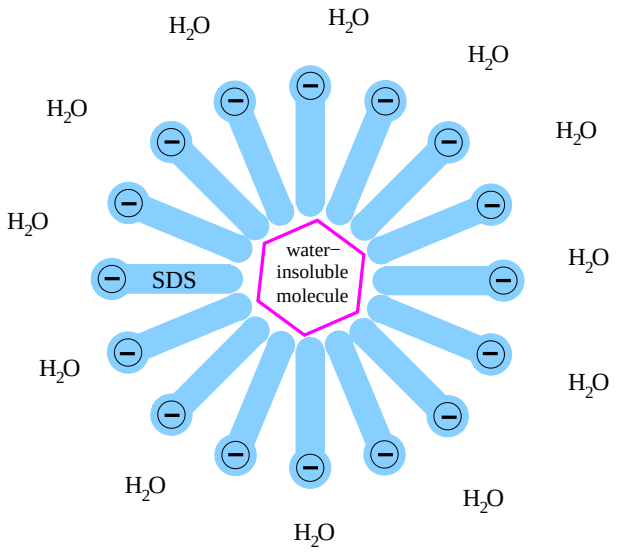


Figure 1.5: A two-dimensional sketch of a micelle covering a water-insoluble molecule. The blue objects represent SDS molecules with a water soluble end marked by a circle with a minus (negative charge) and the opposite end which is water insoluble. The water soluble part sticks to water molecules (H_2O) and the insoluble part to the water insoluble molecule (the pink hexagon). In three dimensions the micelle is a ball.

the bases, it becomes clear that the number of bonds determines the strength of binding. Targets and probes with a high GC-content (i.e. the percentage of G or C within the base sequence) will have more bonds than those with the same length but a smaller GC-content and thus their binding will be stronger. So, if a target has hybridized with only a few number of hydrogen bonds, the concentration of the detergent does not need to be very high in order to allow single detergent molecules to slip between target and probe by canceling hydrogen bonds.

A washing model

The following model shall help to understand the main characteristics of the washing procedure. Acting on the assumption that the detergent concentration is chosen high enough to dissolve all non-hybridized material, we only look at the dynamics of hybridized targets. First of all we introduce some notation. Let $W_i(t)$ be the number of targets of type i which are washed off the surface by the detergent within a period of time t. According to Section 1.1 m is the number of different target types, N_i, $i = 1, 2, ..., m$ the number of targets of type i hybridized to the spot and c the detergent concentration.

Further we assume that within a specified period of time t a certain target is approached by a random number of detergent molecules $\tilde{D}(t)$. Let $\tilde{D}(t)$ be Poisson distributed with intensity $\tilde{\lambda}(c)t$. Obviously, the intensity $\tilde{\lambda}(c)t$ depends on the concentration c and the time t. Only a percentage r_i of the approaching detergent molecules will bind a target of type i. This behavior is governed by the affinity of the detergent to bind a target of type i. Due to the theory of thinning Poisson processes ([Chung], chapter 7), the resulting process for the number of bound detergent molecules $D(t)$ is Poisson, too, with intensity

$$\lambda_i(c) \cdot t = \tilde{\lambda}(c) \cdot r_i t.$$

We assume that at least k_i detergent molecules have to bind a target of type i in order to cancel its hybridization energy to the probe, and thus, wash it off the surface. Let $p_{k_i} := \mathbb{P}(D(t) \geq k_i)$ denote the probability of solving a target of type i. Since $D(t)$ is Poisson, we get

$$\begin{aligned} p_{k_i} &= 1 - \mathbb{P}(D(t) < k_i) \\ &= 1 - e^{-\lambda_i(c)\cdot t} \sum_{j=0}^{k_i-1} \frac{(\lambda_i(c) \cdot t)^j}{j!}. \end{aligned}$$

Further, the probability of solving l targets of type i, conditional on the total number of such targets N_i of this type hybridized to the spot, computes to

$$\mathbb{P}(W_i(t) = l \mid N_i) = \binom{N_i}{l} p_{k_i}^l (1 - p_{k_i})^{N_i - l}. \tag{1.11}$$

Since $W_i(t)$ follows a binomial distribution as can be seen from Equation (1.11), it has mean

$$\mathbb{E}(W_i(t) \mid N_i) = N_i \cdot p_{k_i}$$

and variance

$$\mathbb{V}ar(W_i(t) \mid N_i) = N_i \cdot p_{k_i} \cdot (1 - p_{k_i}).$$

See [Grab] for more details.

We are further interested in the number of targets of type i that stay on the spot, because this is the input of the fluorescence reaction. It shall be denoted by H_i and can be easily determined by subtracting $W_i(t)$ from N_i, i.e.

$$H_i = N_i - W_i(t).$$

The washing model, especially the distribution of H_i and its dependency on the detergent concentration will be investigated in Section 2.2.2.

1.2.3 Fluorescence

In microarray experiments the so-called laser induced fluorescence (LIF) is used. Here, laser light of high intensity and a certain wavelength is used to excite molecules in order to emit light of a different wavelength. To understand this process it is helpful to separately look at the light generated by the laser and the process of fluorescence itself.

1.2.3.1 Light Amplification by Stimulated Emission of Radiation (LASER)

Inside a laser device light is amplified. Therefore an active medium with atoms at two energy levels $E_1 < E_2$ is needed. If a photon strikes an atom which is at level E_2 an additional photon is emitted if the energy of the striking photon $h\nu$ is approximately $E_2 - E_1$, with Planck's constant h and the photon's frequency ν. This process is accompanied by a transition of the atom's energy level from E_2 to E_1 and is called *stimulated emission*. For illustration see Figure 1.6.

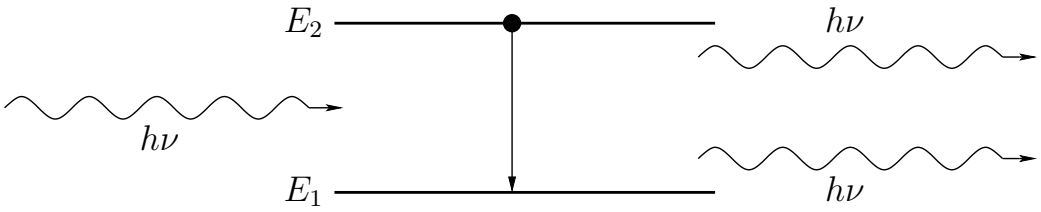


Figure 1.6: A photon of energy $h\nu$ stimulates an atom to emit a clone photon by a transition from energy level E_2 to E_1. (On the basis of figure $12.2 - 4$ of [SaTe], chapter 12.)

Since an atom at level E_2 is lost at each transition, external energy is needed to recover the number of atoms at level E_2 by exiting atoms at level E_1 to undergo an upward transition to level E_2.

The emitted photon has the same characteristics as the striking photon, i.e. the same wavelength, direction and polarization. Thus, it is able to cause further emissions if striking other atoms at level E_2. This amplification process is continued until the photons reach the end of the laser device. In addition, photons striking atoms at level E_1 might be absorbed and transitions to level E_2 take place.

1.2.3.2 A laser model from [SaTe]

Let N_1 and N_2 be the number of atoms at level E_1 and E_2, respectively. According to [SaTe], chapter 13 the resulting photon flux density $\phi(z)$ (i.e.

the number of photons per cm^2 and s, traveling in the z-direction) satisfies the ordinary differential equation

$$\frac{d\phi(z)}{dz} = (N_2 - N_1)\sigma(\nu)\phi(z), \tag{1.12}$$

where $\sigma(\nu)$ is the transition cross section that is a measure for the area in which a photon of frequency ν is able to interact with an atom at energy level E_1 or E_2. The transition cross section can be calculated from Schrödinger's equation but is usually determined experimentally.

$$\phi(z) = \phi_0 e^{(N_2-N_1)\sigma(\nu)z} \tag{1.13}$$

solves Equation (1.12) for $\phi(0) = \phi_0$. Thus, the ratio of $\phi(d)/\phi(0)$ defining the overall gain $G(\nu)$ in a laser device of length d and for photons at frequency ν is

$$G(\nu) = e^{(N_2-N_1)\sigma(\nu)d}.$$

See [SaTe], chapter 12 for more details.

Laser noise: The major noise source in laser devices is *spontaneous emission*. Spontaneous emission is due to atoms in energy level E_2 which undergo a transition to energy level E_1 without any external stimulation in contrast to stimulated emission as described in the previous paragraph. During this process a photon of random direction and polarization is generated. But it still has a frequency of approximately $\nu_0 = (E_2 - E_1)/h$.

It is possible to filter out some of this noise by using a collection aperture, a bandpass optical filter and a polarizer. Let $d\Omega$ be the angle of collection from the aperture and B the frequency band of the bandpass filter centered about the stimulated emission frequency. According to [SaTe], chapter 13, the resulting number of photons added by spontaneous emission from an incremental volume of unit area and length dz is $\epsilon_{sp}(\nu)dz$, where

$$\epsilon_{sp}(\nu) = \frac{N_2 B d\Omega \sigma(\nu) \int_0^\infty \sigma(\nu)\nu^2 d\nu}{c^2 \int_0^\infty \sigma(\nu) d\nu}. \tag{1.14}$$

Since $\sigma(\nu)$ is sharply peaked, it is narrow in comparison with ν^2. Therefore and because of $\sigma(\nu)$ being centered about ν_0 (according to [SaTe], chapter 12), ν^2 might be replaced by ν_0^2. This leads to the following simplification of Equation (1.14):

$$\epsilon_{sp}(\nu) = \frac{N_2 B d\Omega \sigma(\nu)\nu_0^2}{c^2}.$$

In order to account for spontaneous emission in the overall gain, Equation (1.12) has to be modified as follows:

$$\frac{d\phi(z)}{dz} = (N_2 - N_1)\sigma(\nu)\phi(z) + \epsilon_{sp}(\nu).$$

The solution of this equation with the initial value $\phi(0) = \phi_0$ is

$$\phi(z) = -\frac{\epsilon_{sp}(\nu)}{(N_2 - N_1)\sigma(\nu)} + \left(\phi_0 + \frac{\epsilon_{sp}(\nu)}{(N_2 - N_1)\sigma(\nu)}\right) e^{(N_2 - N_1)\sigma(\nu)z}. \quad (1.15)$$

Using Equation (1.15) the overall gain $G(\nu)$ in a laser device of length d is

$$G(\nu) = -\frac{\epsilon_{sp}(\nu)}{(N_2 - N_1)\sigma(\nu)\phi_0} + \left(1 + \frac{\epsilon_{sp}(\nu)}{(N_2 - N_1)\sigma(\nu)\phi_0}\right) e^{(N_2 - N_1)\sigma(\nu)d}.$$

As can be seen here, the gain with spontaneous emission is by the summand

$$-\frac{\epsilon_{sp}(\nu)}{(N_2 - N_1)\sigma(\nu)\phi_0} + \frac{\epsilon_{sp}(\nu)}{(N_2 - N_1)\sigma(\nu)\phi_0} e^{(N_2 - N_1)\sigma(\nu)d}$$

greater than the gain without spontaneous emission. So, the difference of the two gains depends on the initial photon flux ϕ_0 and the rate for spontaneous emission which contributes to $\epsilon_{sp}(\nu)$. Thus, randomness in these two quantities implicates additional randomness in the overall gain.

Nevertheless, laser devices are electronic components. Thus, they exhibit the same noise sources as all electronic components do, including *Johnson-Nyquist noise*, *shot noise* and *Flicker noise*. These noise sources will be considered in Section 1.2.4.1.

After discussing these noise sources it is clear that it is difficult to make a good prediction for the intensity of the light which leaves the laser. The good news about this problem is that the intensity can be measured quite accurately, so it is not necessary to try to calculate the true value. We tried to measure the intensity of a laser used in a microarray scanner but failed to catch enough light to get a significant signal since we were not allowed to open the scanner. The manufacturer of the scanner did not give any information on the noise of the scanner. Thus, we will restrict our investigation to the major noise source, i.e. spontaneous emission.

This model will also be further looked at in Chapter 2.

1.2.3.3 Fluorescence

There are different forms of radiation which emanate from materials. Two major forms are heat radiation and *luminescence*. Heat radiation is due

to kinetic energy of molecules and atoms whereas radiation without release of thermal energy due to excitation of a material is called luminescence. According to the source of excitation, luminescence can be distinguished into (see [Schwedt])

- radioluminescence, i.e. excitation by nuclear radiation,
- electroluminescence, i.e. excitation by alternating electrical fields,
- triboluminescence, i.e. mechanical excitation,
- sonoluminescence, i.e. excitation by sound,
- galvanoluminescence, i.e. excitation by electrolysis,
- thermoluminescence, i.e. excitation by heat,
- chemiluminescence, i.e. excitation by chemical reactions and
- photoluminescence, i.e. excitation by light.

In the following we are interested in photoluminescence. As mentioned above it is caused by light. More precisely, it covers all phenomena where molecules reach an electronically excited state by absorption of a photon and as a result emit another photon [Haßl]. The detailed mechanisms are illustrated in Figure 1.7, where the most important excitation and de-excitation pathways are shown.

Absorption of a photon leads to a transition of an electron from the ground state to the excited state. In general, these states are singlet states, i.e. the promoted electron of a pair of electrons does not change its spin s_1 and the spin quantum number, $S = s_1 + s_2$, where $s_1, s_2 \in \{+\frac{1}{2}, -\frac{1}{2}\}$, remains zero. The term singlet refers to the multiplicity of the total spin quantum number, $M = 2S + 1$, which is 1 at singlet states. The ground singlet state is S_0 and the excited singlet states are $S_1, S_2, ...$ Every singlet state is associated with a number of vibrational states, which are further divided into rotational levels. The following pathways are possible to return from S_1 to S_0.

- *Internal conversion* is non-radiative (i.e. there is no emission of a photon) and leads directly to the ground state S_0.
- *Fluorescence* leads to the ground state by emission of a photon.
- *Intersystem crossing* leads to a transition to the triplet state T_1. This process is non-radiative and accompanied by change in spin of the promoted electron.

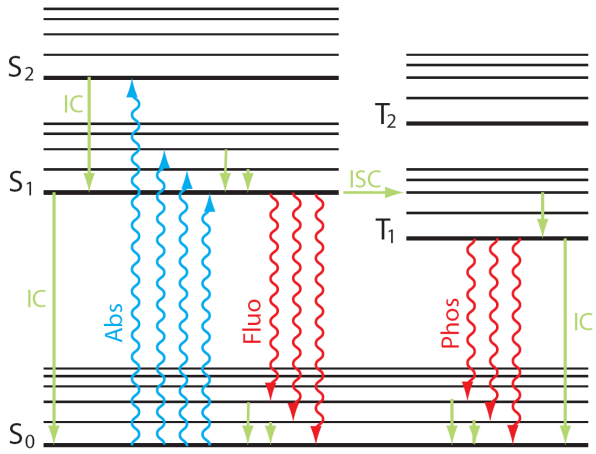


Figure 1.7: Perrin-Jablonski diagram [Haßl]. Electronic (thick horizontal lines) and associated vibrational energy levels (thin horizontal lines) and the most important excitation and de-excitation pathways are shown. Wavy lines denote absorption and radiative pathways. Straight lines with arrow-heads represent possible non-radiative pathways. Abs: absorption; Fluo: fluorescence; Phos: phosphorescence; IC: internal conversion; ISC: intersystem crossing; S: singlet states; T: triplet state.

- Returning to the ground state is either achieved by internal conversion as already mentioned or by *phosphorescence*, i.e. the emission of a photon at a larger wavelength than the fluorescence photon.

If a molecule is excited to higher singlet states than S_1, it will loose the energy difference to S_1 by internal conversion. The de-excitation times of the different pathways are for internal conversion in the range of $10^{-14}s$ for fluorescence between 10^{-9} and $10^{-6}s$ for intersystem crossing $10^{-8}s$ and for phosphorescence 10^{-4}-$100s$. For details see [Schwedt] and [Haßl].

The holding time in the excited state is exponentially distributed with parameter $k = \sum_i k_i$, where the k_i are the individual rate constants of the different de-excitation pathways. The rate constants depend on the molecule itself and on its micro-environment, such as the refractive index, the pH, the oxygen concentration, the ion concentration and the temperature [Haßl]. Therefore, it is essential to keep the micro-environment stable during a measurement.

The fluorescence quantum yield p_f (which is the probability for the event that the excited molecule returns to the ground state via fluorescence) is

$$p_f = \frac{k_f}{k},$$

where k_f is the rate constant for fluorescence. According to [Schwedt] the fluorescence intensity F is given by

$$F = 2.3 p_f I_0 \tau \kappa l, \tag{1.16}$$

where I_0 is the intensity of the laser light, τ is the molar extinction coefficient, κ the molar concentration and l the thickness of the layer. According to [SaTe], the laser light in Equation (1.16) has intensity $I_0 = h\nu\phi(d)$.

In microarray experiments often used fluorescence dyes are the cyanines Cy3 and Cy5. Cy3 has its absorption maximum at 550 nm, emits maximally at 570 nm and has a quantum yield of 0.15. In contrast Cy5 has its absorption maximum at 649 nm, emits maximally at 670 nm and has a quantum yield of 0.28. For details see [Lako], chapter 3, [EGMW] and [South].

The relations between fluorescence dyes, incoming laser light and induced fluorescence will be looked at in Chapter 2.

1.2.4 The Detection

As previously mentioned, the signal is multiplied with a PMT (photomultiplier tube). A typical PMT is shown in Figure 1.8.

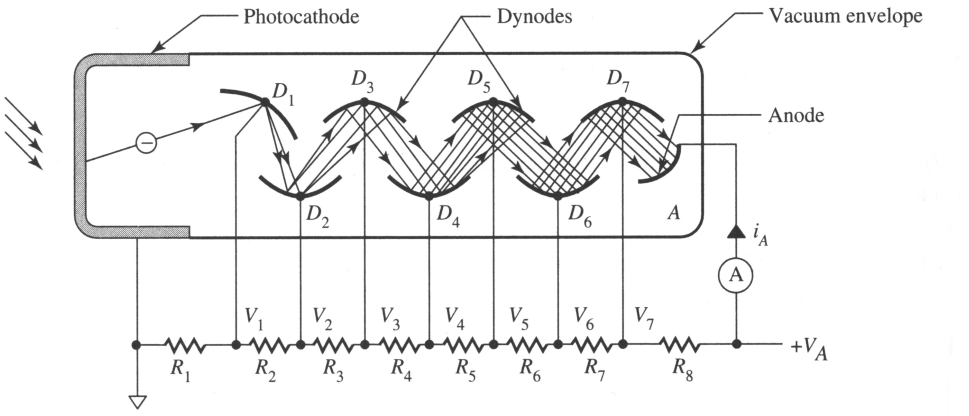


Figure 1.8: Sketch of a photomultiplier tube from [Uiga], chapter 6.

It consists of three main components, the photocathode, several dynodes and an anode. Photons of the light beam have a specific energy $E = h\frac{c}{\lambda}$, where λ is the wavelength of the light, h is Planck's constant and c is the velocity of light. An electron inside the photocathode is able to absorb the photon's energy. If this amount of energy is larger than the work function

W_c of the photocathode, the electron will be emitted. The work function depends on the material. Transferred energy exceeding the work function will be transformed into kinetic energy of the emitted electron according to Einstein's equation:

$$E_{kin} = h\frac{c}{\lambda} - W_c. \tag{1.17}$$

Recapitulating, the theoretic number of emitted electrons (i.e. primary electrons) equals the number of incident photons if their energy exceeds the work function of the photocathode. Therefore, the total number of electrons depends only on the intensity of light and not particularly on its wavelength. In practice the ratio η of emitted electrons to photons is less than one, since photons and electrons might be trapped by interacting with other particles. For details see for example [SauWei]. In addition, the number of striking photons is not a constant. In fact it stochastically fluctuates around its mean. This error is called quantum noise ([SauWei]).

The emitted electron is accelerated by the electric field towards the first dynode with a final energy of $eV_1 + E_{kin}$, where e is the elementary charge and V_1 is the dynode voltage. When hitting the dynode, the kinetic energy is used to emit further electrons according to the dynode's work function W_1. The number Z_1 of emitted electrons (i.e. secondary electrons) after a collision of the electron and the first dynode theoretically is

$$Z_1 = \frac{eV_1 + E_{kin}}{W_1}. \tag{1.18}$$

Subsequent dynodes are under progressively higher potential, so the newly emitted electrons are accelerated towards the next dynode. Further electrons are emitted. Analogously, the number Z_i of emitted electrons after a collision of a single electron and the ith dynode theoretically is

$$Z_i = \frac{eV_i}{W_i}. \tag{1.19}$$

In practice some of the electron's kinetic energy is transformed into infrared radiation, oscillation of the atom lattice (resulting in thermal energy, too) and penetration, i.e. the electron penetrates the dynode and only keeps some of its kinetic energy. Hence, the realistic number of secondary electrons of the ith dynode is smaller than shown in Formula (1.18) or (1.19). This number is denoted by N_i. For reference see [SiSu], chapter 6.

Another problem is the efficiency of an electron in reaching the next dynode. Some electrons get lost. Let α_i be the efficiency of an electron in finding its way from dynode $(i-1)$ to dynode i. Thus, the total number

of emitted electrons by the mth dynode due to one electron emitted by the photocathode is

$$N_m = \alpha_1 Z_1 \cdot \alpha_2 Z_2 \cdot ... \cdot \alpha_m Z_m. \tag{1.20}$$

See [SiSu], chapter 6 for details. These electrons are detected by an anode, where the current is measured.

Considering dynodes of metal, higher voltage leads to increasing penetration of the dynode's surface by the striking electron. If this happens, the efficiency of electron emission will decrease and Equation (1.18) does not hold. In an NEA-type dynode (negative electron affinity) this effect is minimized by using semiconductors as dynodes. Figure 1.9 illustrates the dependance of the number of emitted electrons on the dynode voltage and the dynode material.

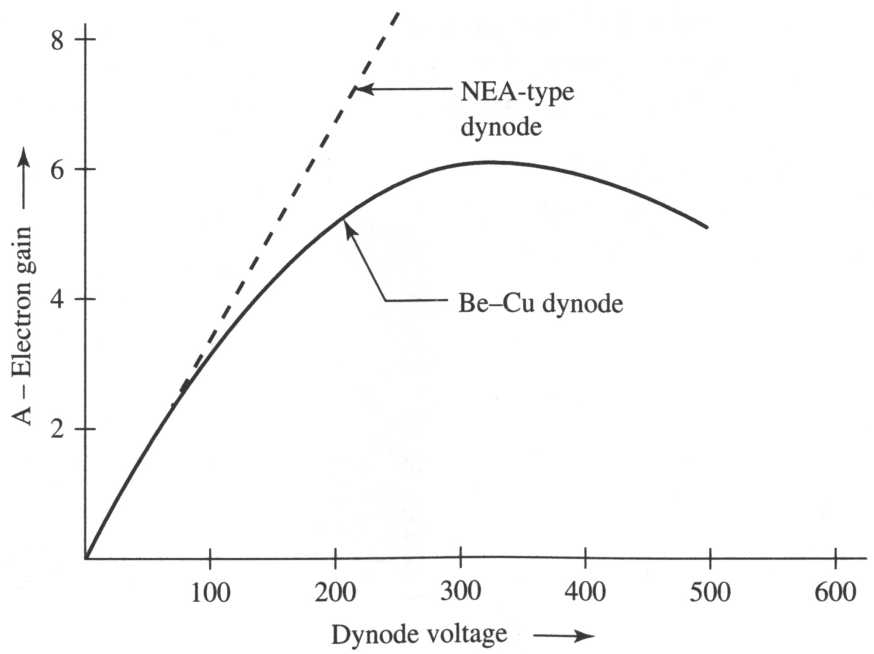


Figure 1.9: Electron gain (Z_e) as function of dynode voltage (V_d) for a NEA-type dynode and a metal dynode (Be-Cu) from [Uiga], chapter 6.

For references see [Uiga], chapter 6 and [BiSchl], chapter 2.

1.2.4.1 Possible sources of noise of electronic devices

During detection, there are many possible noise sources, which perturb the signal. Thus, it is essential to understand the relationship between the original signal entering the detection aperture and the falsified signal which is finally detected. Possible noise sources shall be looked at in the following paragraph. At this point it shall be mentioned, that these sources could be applied to the electronic devices of the other modules, too.

1. **Johnson-Nyquist noise** is produced by the thermal agitation of charged particles in a resistive element. Such resistive elements are R_1, R_2, ..., R_8, the amperemeter itself and the vacuum envelope in Figure 1.8. But every measurement device, including the amperemeter to measure the current is a resistive element, too. The root mean square voltage V_{Jrms} and current I_{Jrms} of disturbance can be calculated as follows

$$\begin{aligned} V_{Jrms} &= \sqrt{4kRT\Delta f}, \\ I_{Jrms} &= \sqrt{4kT\Delta f/R}. \end{aligned}$$

   R is the resistance, T the temperature, k Boltzmann's constant and Δf the bandwidth over which the noise is measured/the electrical device is operated at. The noise voltage and current have mean zero. So the root mean square values are equivalent to the standard deviation. Johnson-Nyquist noise is white noise (mean 0, constant variance) and follows a normal distribution. For details see [Uiga], chapter 6, [SauWei] and [SaTe].

2. **Shot noise** is caused by the electron multiplication process which is usually modeled with a Poisson process. The high degree of multiplication that is achieved, implicates a multiplication of small fluctuations in the electron current (see [SiSu], chapter 7). The shot noise describes the deviation of the current from its mean. The root mean square current I_{Srms} of the shot noise is due to the discrete nature of the generated photoelectrons. It depends on the average current I_{avg} and the bandwidth Δf (see [SiSu], chapter 7 and [Uiga], chapter 6):

$$I_{Srms} = \sqrt{2eI_{avg}\Delta f}.$$

   Shot noise is white noise with standard deviation I_{Srms}. For large numbers of electrons it is approximately normally distributed.

3. **Generation-recombination noise.** The freeing of electrons (i.e. generation), which are associated to an atom and the ensuing trapping

(i.e. recombination) of freed carriers by uncovered acceptors are discrete processes. This leads to a random fluctuation in the number of free electrons. The root mean square of the current I_{GRrms} is

$$I_{GRrms} = 2eG\sqrt{\eta EA\Delta f},$$

where G is the ratio of active electrons to photoelectrons generated, η the quantum efficiency (the ratio of emitted electrons to photons), E the radiant incidence and A the detector receiving area. For details see [Uiga], chapter 6. Generation recombination noise is also supposed to be Gaussian with mean 0 and standard deviation I_{GRrms}.

4. **1/f or Flicker noise** is a phenomenon which occurs in all nonmetal conductors. Its origin is not clarified. Thus, only a heuristic solution for the determination of the root mean square of the current of the flicker noise I_{Frms} exists. It is

$$I_{Frms} = k\sqrt{{I_{dc}}^a \Delta f / f^b},$$

where I_{dc} is the direct current through the conductor, f the operating frequency and k, a, b are arbitrary constants. k depends on the material of the conductor and its treatment, whereas $a \approx 2$ and $b \approx 1$. See [Uiga], chapter 6 for details. This noise source is considered to be Gaussian with mean 0 and standard deviation I_{Frms}.

The resulting total equivalent root mean square noise current is (see [Uiga], chapter 6)

$$I_{Neq} = \sqrt{{I_{Jrms}}^2 + {I_{Srms}}^2 + {I_{GRrms}}^2 + {I_{Frms}}^2}. \tag{1.21}$$

Thus, under the assumption that the shot noise is close to Gaussian we have a total noise current perturbing the signal which is Gaussian, too. It has mean 0 and standard deviation I_{Neq}.

1.2.4.2 Single type Branching process - a model for the PMT

Due to the various error sources described in the previous paragraph, it is difficult to describe the process of detection properly. Using stochastic processes might help to understand tendencies of the underlying process without knowing the parameter situations in detail.

Since the process is governed by multiplication of signal carriers within the PMT, it is reasonable to apply the theory of single type branching processes. Reference for the following paragraph is [MaTeSa].

Let N_m be the random variable denoting the number of secondary electrons at the mth dynode and assume that the number of primary electrons is $N_0 = 1$. Under the assumption that the number of electrons generated at a certain dynode is independently and identically distributed for each striking electron, N_m is described by a Galton-Watson process. Thus N_{m+1} is determined by the sum

$$N_{m+1} = \sum_{k=1}^{N_m} Z_m^{(k)}$$

of N_m i.i.d. random variables $Z_m^{(1)}, Z_m^{(2)}, ..., Z_m^{(N_m)}$, denoting the number of secondary electrons generated by the electrons of the mth dynode. Each of the Z_m^i, $i = 1, ..., N_m$ has a probability distribution

$$\mathbb{P}\left(Z_m^i = k\right) = p_k^{(m)}, \;\; k \in \mathbb{N}_0.$$

The statistical properties of the distribution of N_m can be derived with the help of the probability generating function $G_m(z) = \mathbb{E}\left(z^{N_m}\right)$ which holds the following recursion:

$$\begin{aligned} G_0(z) &= z, \\ G_{m+1}(z) &= G_m\left(Q_m(z)\right), \;\; m \in \mathbb{N}_0, \end{aligned} \tag{1.22}$$

where

$$Q_m(z) = \sum_{k=0}^{\infty} p_k^{(m)} z^k$$

is the probability generating function of $Z_m^{(i)}$.

[MaTeSa] use the Recursion (1.22) to develop expressions for the mean $\mathbb{E}(N_m)$ of the number N_m of secondary electrons at the mth dynode and its variance $\mathbb{V}ar(N_m)$. But at this point, we are only interested in its probability distribution $\mathbb{P}(N_m = n) =: p_m(n)$, which is

$$p_m(n) = \left[\frac{1}{n!}\frac{\partial^n}{\partial z^n} G_m(z)\right]_{z=0}. \tag{1.23}$$

A random number of N_0 primary electrons is considered next. For illustration of the following paragraph see Figure 1.10. Let the number of photons striking the photocathode in a period of time be Poisson distributed with parameter μ.

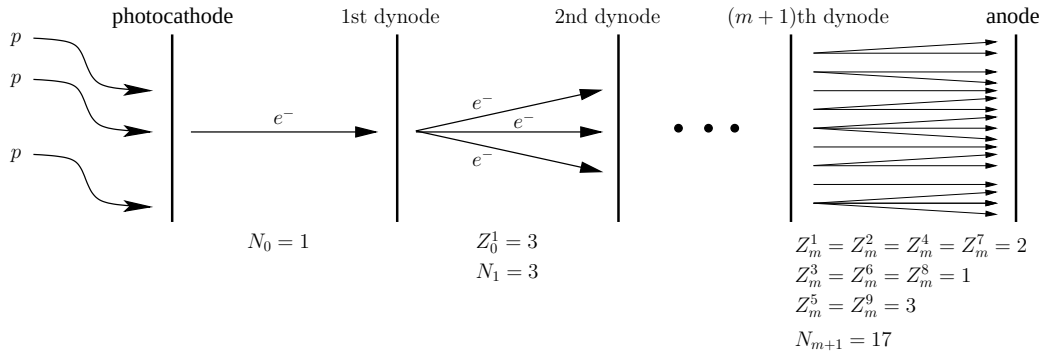


Figure 1.10: Illustration of the branching process in a PMT.

Due to the theory of thinning Poisson processes (see [Chung], chapter 7), the resulting number of primary electrons N_0 is Poisson distributed with parameter $\eta\mu$. Recall, η denotes the ratio of emitted electrons to striking photons.

Let $l \in \mathbb{N}_0$ be the number of primary electrons and N_m^i be the number of electrons at the mth dynode, generated by the ith primary electron. Assuming statistical independence of the different offspring paths, the multiplication process at the dynodes can be considered separately for each of the l primary electrons. Thus, according to Formula (1.23) the probability of k_i, $i = 1, 2, ..., l$, secondary electrons at the mth dynode generated by the ith primary electron is $p_m^i(k_i) := \mathbb{P}(N_m^i = k_i) = p_m(k_i)$.

As a result we can give an expression for the conditional probability of $N_m = \sum_i N_m^i = k$ secondary electrons at the mth dynode, given that there are l primary electrons, which is

$$\mathbb{P}(N_m = k \mid N_0 = l) = \sum_{\substack{k_1,k_2,...,k_l \in \mathbb{N}_0 \\ k_1+k_2+...+k_l=k}} \prod_{i=1}^{l} p_m^i(k_i). \tag{1.24}$$

Formula (1.24) is derived by convolution of the N_m^i. Since we are interested in the distribution of N_m, we have to determine the marginal distribution according to N_0 as follows.

$$\begin{aligned} \mathbb{P}(N_m = k) &= \int \sum_{\substack{k_1,k_2,...,k_l \in \mathbb{N}_0 \\ k_1+k_2+...+k_l=k}} \prod_{i=1}^{l} p_m^i(k_i) d\mathbb{P}(N_0) \qquad (1.25) \\ &= \sum_{l=0}^{\infty} \mathbb{P}(N_0 = l) \sum_{\substack{k_1,k_2,...,k_l \in \mathbb{N}_0 \\ k_1+k_2+...+k_l=k}} \prod_{i=1}^{l} p_m^i(k_i). \end{aligned}$$

In Section 2.2.4 we will use this model to examine the randomness of the number of electrons striking the anode in order to characterize the electron current measured by the amperemeter.

Chapter 2

Examination of introduced modules

In this chapter we will explore the statistical behavior of the models introduced in the previous chapter. This helps to derive the underlying laws which affect the detected signal in the end of the microarray experiment. Thus, it might serve as a basis to develop a reliable test for the fold change in gene expression activities.

2.1 Hybridization

The hybridization process without dissociation of targets has already been simulated by [ReWi]. In Section 1.1 we extended this model by including dissociation events of targets. The resulting hybridization process with dissociation of targets is the basis of the following analysis.

In a first step simulated the process with realistic parameter situations. We soon realized that this effort is limited by computational power due to too large target and probe numbers. For that reason we decided to investigate the process more analytically by determining its stationary distribution with the help of the Markov generator Q. This also leads to computational problems for realistic parameter situations. As a result we applied two different kinds of limits. With the first limit we received a partial differential equation (PDE) which has the stationary distribution for large S as solution. The second limit was applied in order to find a deterministic process which describes the behavior of the hybridization process for large S, too.

We were able to show that the only solution of the PDE is the trivial solution where no targets have hybridized at all. This of course is not of interest. However, we were able to draw closer to the stationary distribution

with the help of the deterministic process. We received a solution which could be directly associated to the distributional solution of the respective PDE. The results have been compared to the outcome of the simulation which corroborated the drawn conclusions.

2.1.0.3 Parameter situation

Before analyzing the process from Section 1.1, we need to clarify its parameter situation. Even though we only model the dynamic of a single spot, we will start with the total number of spots per microarray because it helps to understand other parameters. In the literature, the number of spots ranges from $1,536$ ([Buhl]) to $25,392$ ([Yats]). Most commonly used values are $6,000$ ([Jain]) and $4,800$ ([Alhad]).

On the other hand, the number of different targets types m depends on the organism observed and the metabolic cell situation. Since every spot at least hybridizes to two different kinds of targets (the specific targets of the two colors), the number of targets could be considered about twice the number of spots. At this point it is necessary to mention that on larger microarrays (e.g. [Yats]), some spots occur in multiple copies. Nevertheless, the number of target types stays large. The next important parameter is the number of probes per spot S. An indication could be found in [Chou], i.e. a total number of probes per spot ranging from several millions to hundreds of millions of molecules.

Commonly, the rates of a Markov process are summarized in a quadratic matrix, the Markov generator Q. Its entry $q_{i,j}$ at row i and column j, $i \neq j$, is the transition rate from state j to i. Its diagonal elements $q_{i,i}$ are the negative sum of the entries in the respective columns

$$q_{i,i} = -\sum_{j \neq i} q_{j,i}.$$

Definition 2.1. *Let $dim_{S,m}(Q)$ be the dimension of Markov generator Q which represents the number of states of the hybridization process depending on the total number of probes per spot S and the number of different target types m.*

Note, $dim_{S,m}(Q) = |\Sigma_{S,m}|$.

Theorem 2.2. *The size of the state space of the process from Section 1.1, i.e. $dim_{S,m}(Q)$ increases*

a) in S at least as fast as S^m (i.e. $dim_{S,.}(Q) \in \Omega(S^m)$) and

b) in m at least as fast as m^S (i.e. $dim_{\cdot,m}(Q) \in \Omega(m^S)$).

Proof. *a)* Firstly, the number of states is

$$dim_{S,m}(Q) = \binom{S+m}{m}. \tag{2.1}$$

This can be seen as follows. The dimension of Q equals the number of possible states within the hybridization process. Every state is characterized by the numbers $N_1, N_2, ..., N_m$ of hybridized targets of each type. Additionally, the overall number of hybridized targets cannot exceed S, i.e. $N_1 + N_2 + ... + N_m \leq S$. Thus, the set of states is $\{(N_1, N_2, ..., N_m) \in \mathbb{N}^m : \sum_{i=1}^{m} N_i \leq S\} = \{(N_1, N_2, ..., N_m, N_{m+1}) \in \mathbb{N}^{m+1} : \sum_{i=1}^{m+1} N_i = S\}$. We are interested in the number of elements of this set, i.e. $|\{(N_1, N_2, ..., N_m, N_{m+1}) \in \mathbb{N}^{m+1} : \sum_{i=1}^{m+1} N_i = S\}| = dim_{S,m}(Q)$. Obviously, $dim_{S,1}(Q) = S$ and $dim_{S,m}(Q)$ holds the following recursion:

$$dim_{S,m}(Q) = \sum_{N_{m+1}=0}^{S} dim_{S-N_{m+1},m-1}(Q). \tag{2.2}$$

With Equation (2.2) in mind Equation (2.1) can be proved by induction over m. See [HaHiMo], chapter 2 for detail.

Secondly, the following inequality generally holds for binomial coefficients ([Steger], chapter 1),

$$\left(\frac{n}{k}\right)^k \leq \binom{n}{k}.$$

Applying this to $dim_{S,m}(Q)$ yields

$$\left(\frac{S+m}{m}\right)^m \leq \binom{S+m}{m}.$$

Thus,

$$\begin{aligned} dim_{S,m}(Q) &= \binom{S+m}{m} \geq \left(\frac{S+m}{m}\right)^m = \left(\frac{S}{m}+1\right)^m \\ &> \left(\frac{1}{m}\right)^m S^m \in \Omega(S^m). \end{aligned}$$

b) Since $dim_{S,m}(Q) = \binom{S+m}{m} = \binom{m+S}{S}$, the same arguments as in a) yield $dim_{S,m}(Q) \in \Omega(m^S)$.

□

Thus, the number of states increases too fast. For example, for realistic parameter values of $S \approx 10^6$ and $m \approx 10^3$ it would be greater than $10^{6,000}$. This number is way too large to be considered in a numeric analysis of the process. So, we restrict the analysis to a total number of 500 probe molecules and 2 or 4 different target types per spot. These values turned out to be just manageable in simulations and further analysis. But they are still complex enough to serve as a simple model for basic considerations.

On the one hand, the case of two target types will be able to show the basic dynamics of the process if only specific targets are investigated and we are interested in dye effects. On the other hand, the case of four targets is able to also model cross-hybridization (see page 10 for an explanation of the cross-hybridization term). During cross-hybridization specific and unspecific target types hybridize to the spot. Both types are labeled with the two different dyes. This makes it impossible to distinguish between the types. With respect to [ReWi] the case of $m = 2$ is called the *ideal case* because there is no cross-hybridization.

In real hybridization reactions there are about 150 times as many targets as probes per spot (see [ReWi]). So, we chose the ratio of the sum of targets to the number of probes per spot to be 150, too.

No indications for hybridization probabilities, dissociation probabilities and recruitment rates could be found in the literature. Thus, we decided to investigate the model with rates that are reasonable to our understanding of the underlying process. The chosen parameter situations of the ideal case are shown in Table 2.1 and in Table 2.2 for equal and unequal hybridization and dissociation probabilities and rates, respectively.

Table 2.3 shows the investigated parameter situation for the process in presence of cross-hybridization.

Subsequently, we will refer to the case of equal hybridization probabilities and dissociation rates as *the case of equal probabilities* and to the case of unequal hybridization probabilities and dissociation rates as *the case of unequal probabilities*. Our investigation will start with the ideal case.

2.1.1 The stationary distribution

At the beginning of the hybridization process there are no hybridized targets at all. Thus, the respective Markov process starts in state $(0, 0)$ or in other words, the probability of the process being in state $(0, 0)$ at time $t = 0$ is 1

binding probabilities	
π_1	.7
π_2	.7
dissociation rates	
γ_1	.5
γ_2	.5
initial target numbers	
T_1	$25,000$
T_2	$50,000$
number of probes	
S	500
exponential clock	
λ	2
duration of the experiment	
θ	.5

Table 2.1: Parameter situation for analyzing the ideal case with equal probabilities and rates.

and 0 for all other states. At this point recall the state space $\Sigma_{S,m}$ of the hybridization process from Formula 1.1.

In the course of time, the probabilities of the process being in certain states change due to the character of the Markov generator. Under certain assumptions, which are described later, these probabilities reach an equilibrium as time tends to infinity. The vector of probabilities of the process being in the respective states at equilibrium is called *stationary distribution*. So, if the duration of the hybridization reaction is sufficiently large, the hybridization process will be expected to be close to its stationary distribution.

Thus it might be useful to calculate the stationary distribution to overcome the computational limits of too many simulations and a realistic number of probe molecules per spot (i.e. 6×10^8).

The stationary distribution ρ of a continuous time Markov process satisfies ([YiZha], Chapter 1)

$$Q\rho = 0,$$

where Q is the Markov generator of our process.

2.1.1.1 The ideal case (two target types)

The exact structure of Q for the ideal case is described next.

binding probabilities	
π_1	.6
π_2	.7
dissociation rates	
γ_1	.6
γ_2	.5
initial target numbers	
T_1	$25,000$
T_2	$50,000$
number of probes	
S	500
exponential clock	
λ	2
duration of the experiment	
θ	.5

Table 2.2: Parameter situation for analyzing the ideal case with unequal probabilities and rates.

Let $r_{(M_1,M_2),(L_1,L_2)}$ be the rate of a transition from state (M_1, M_2) to state (L_1, L_2), i.e. the entry of the process' Markov generator Q at position $(L_1, L_2), (M_1, M_2)$. These rates can be derived from Section 1.1 by multiplying the rate r of the process (Equation (1.2)) with the respective transition probability (Equations (1.3), (1.4)). As already mentioned, the condition for the stationary distribution ρ is $Q\rho = 0$. Looking at this equation component wise (here we look at component (L_1, L_2)) implies:

$$\sum_{(M_1,M_2)\neq(L_1,L_2)} r_{(M_1,M_2),(L_1,L_2)}\rho_{(M_1,M_2)} - \rho_{(L_1,L_2)} \sum_{(M_1,M_2)\neq(L_1,L_2)} r_{(L_1,L_2),(M_1,M_2)} = 0$$

Adding

$$\rho_{(L_1,L_2)} \sum_{(M_1,M_2)\neq(L_1,L_2)} r_{(L_1,L_2),(M_1,M_2)}$$

to both sides implies:

$$\sum_{(M_1,M_2)\neq(L_1,L_2)} r_{(M_1,M_2),(L_1,L_2)}\rho_{(M_1,M_2)} = \rho_{(L_1,L_2)} \sum_{(M_1,M_2)\neq(L_1,L_2)} r_{(L_1,L_2),(M_1,M_2)}.$$

Since the number of hybridized targets can only increase or decrease by one,

binding probabilities	
π_1	.7
π_2	.6
π_3	.2
π_4	.15
dissociation probabilities	
γ_1	.3
γ_2	.4
γ_3	.8
γ_4	.85
initial target numbers	
T_1	$50,000$
T_2	$50,000$
T_3	$50,000$
T_4	$50,000$
number of probes	
S	500
exponential clock	
λ	2
duration of the experiment	
θ	.4

Table 2.3: Parameter situation in presence of cross-hybridization.

we end up with the balance equation:

$$\begin{aligned} r_{(L_1-1,L_2),(L_1,L_2)}\rho_{(L_1-1,L_2)} &+ r_{(L_1+1,L_2),(L_1,L_2)}\rho_{(L_1+1,L_2)} \\ +r_{(L_1,L_2-1),(L_1,L_2)}\rho_{(L_1,L_2-1)} &+ r_{(L_1,L_2+1),(L_1,L_2)}\rho_{(L_1,L_2+1)} \\ &= \\ \rho_{(L_1,L_2)}(r_{(L_1,L_2),(L_1-1,L_2)} &+ r_{(L_1,L_2),(L_1+1,L_2)} \\ +r_{(L_1,L_2),(L_1,L_2-1)} &+ r_{(L_1,L_2),(L_1,L_2+1)}). \end{aligned} \tag{2.3}$$

According to Section 1.1, the transition rates of the hybridization process at

state $N = (L_1, L_2)$ are given by

$$\begin{aligned}
r_{(L_1-1,L_2),(L_1,L_2)} &= \begin{cases} \pi_1 \frac{T_1-L_1+1}{T_1-L_1+1+T_2-L_2} \cdot \frac{S-L_1+1-L_2}{S} \lambda(T_1 - L_1 + 1 + T_2 - L_2), & L_1 > 0 \\ 0, & \text{else} \end{cases} \\
&= \begin{cases} \pi_1 (T_1 - L_1 + 1) \frac{S-L_1+1-L_2}{S} \lambda, & L_1 > 0 \\ 0, & \text{else} \end{cases} \\
&=: w(L_1, L_2), \qquad (2.4) \\
r_{(L_1,L_2-1),(L_1,L_2)} &= \begin{cases} \pi_2 \frac{T_2-L_2+1}{T_1-L_1+T_2-L_2+1} \cdot \frac{S-L_1-L_2+1}{S} \lambda(T_1 - L_1 + T_2 - L_2 + 1), & L_2 > 0 \\ 0, & \text{else} \end{cases} \\
&= \begin{cases} \pi_2 (T_2 - L_2 + 1) \frac{S-L_1-L_2+1}{S} \lambda, & L_2 > 0 \\ 0, & \text{else} \end{cases} \\
&=: s(L_1, L_2), \qquad (2.5) \\
r_{(L_1+1,L_2),(L_1,L_2)} &= \begin{cases} \gamma_1(L_1 + 1), & L_1 < S \\ 0, & \text{else} \end{cases} \\
&=: e(L_1, L_2) \qquad (2.6)
\end{aligned}$$

and

$$\begin{aligned}
r_{(L_1,L_2+1),(L_1,L_2)} &= \begin{cases} \gamma_2(L_2 + 1), & L_2 < S \\ 0, & \text{else} \end{cases} \\
&=: n(L_1, L_2). \qquad (2.7)
\end{aligned}$$

Let

$$m(L_1, L_2) := -\left(w(L_1 + 1, L_2) + s(L_1, L_2 + 1) + e(L_1 - 1, L_2) + n(L_1, L_2 - 1)\right)$$

be the negative sum of all rates for exiting the state (L_1, L_2). The notations of the rates can be motivated in the following way. Consider the state space of Figure 1.3 and imagine it was an oriented map. Now, consider state (L_1, L_2). The rates are named after the direction where they come from. For example, $(L_1 - 1, L_2)$ is west of (L_1, L_2). So, the respective rate is called $w(L_1, L_2)$. The negative sum of the rates which leave the middle (L_1, L_2) is denoted by $m(L_1, L_2)$.

Using the rates, matrix Q and vector ρ can be displayed as follows

$$Q = \begin{pmatrix} C_0 & U_0 & & & \\ D_1 & C_1 & U_1 & & \\ & \ddots & \ddots & \ddots & \\ 0 & & D_{S-1} & C_{S-1} & U_{S-1} \\ & & & D_S & C_S \end{pmatrix} \in \mathbb{R}^{\frac{1}{2}(S+1)(S+2)\times\frac{1}{2}(S+1)(S+2)},$$

$$\rho = \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_S \end{pmatrix} \in \mathbb{R}^{\frac{1}{2}(S+1)(S+2)}$$

with submatrices $C_0, C_1, ..., C_S,\ U_0, U_1, ..., U_{S-1},\ D_1, D_2, ..., D_S$ and subvectors $b_0, b_1, ..., b_S$ of the following types

$$C_i = \begin{pmatrix} m(0,i) & e(0,i) & & & \\ w(1,i) & m(1,i) & e(1,i) & & \\ & \ddots & \ddots & \ddots & \\ 0 & & w(S-i-1,i) & m(S-i-1,i) & e(S-i-1,i) \\ & & & w(S-i,i) & m(S-i,i) \end{pmatrix}$$
$$\in \mathbb{R}^{(S-i+1)\times(S-i+1)},\ i = 0, 1, ..., S,$$

$$U_j = \begin{pmatrix} n(0,j) & & & & \\ & n(1,j) & & & \\ & & \ddots & 0 & \\ & 0 & & n(S-j-1,j) & \\ & & & & n(S-j,j) \end{pmatrix}$$
$$\in \mathbb{R}^{(S-j+1)\times(S-j+1)},\ j = 0, 1, ..., S-1,$$

$$D_k = \begin{pmatrix} s(0,k) & & & & & \\ & s(1,k) & & & & \\ & & \ddots & 0 & & \\ & 0 & & s(S-k-1,k) & & \\ & & & & s(S-k,k) & 0 \end{pmatrix}$$
$$\in \mathbb{R}^{(S-k+1)\times(S-k+2)},\ k = 1, 2, ..., S$$

and

$$b_l = \begin{pmatrix} \rho_{(0,l)} \\ \rho_{(1,l)} \\ \vdots \\ \rho_{(S-l-1,l)} \\ \rho_{(S-l,l)} \end{pmatrix} \in \mathbb{R}^{S-l+1},\ l = 0, 1, ..., S$$

Furthermore, from Markov process theory it is well known that a unique stationary distribution exists for irreducible Markov processes (see [ChNg], chapter 1 or [Gant], chapter 13). We will restrict the analysis to the case $T_1, T_2 > 0$. The cases $T_1 = 0 \vee T_2 = 0$ are not of interest since they do not have any interaction between the targets. Further, we will only look at the process with positive hybridization and dissociation rates. On the one hand, the case of $\gamma_i = 0$, $i = 1, .., m$ refers to the model of [ReWi]. On the other hand, the case of $\pi_i = 0$, $i = 1, ..., m$ is not relevant because it shows no hybridizations. The respective Markov processes are not irreducible and almost all states are transient. Therefore, these cases have infinitely many stationary distributions.

Theorem 2.3. *The ideal case hybridization process with Markov generator Q is irreducible for T_1, T_2, π_1, π_2, γ_1, $\gamma_2 > 0$.*

Proof. A Markov process is irreducible if and only if every state of the process can be reached from any other state of the process. Therefor, in a first step we will show that every state can be reached from $(0, 0)$. In a second step we will show that $(0, 0)$ can be reached from any other state in return.
$\underline{S = 0}$:
There is only one state and there is nothing to be proven.
$\underline{S > 0}$:
We will use the principle of complete induction over the states of the process.
Base case:
The states $(1, 0)$ and $(0, 1)$ can be reached from $(0, 0)$ since the respective rates $w(1, 0)$ and $e(0, 1)$ are positive according to Formulas (2.4) and (2.5). Analogously, the states $(2, 0)$, $(1, 1)$ and $(0, 2)$ can be reached from $(1, 0)$ and $(0, 1)$ since the respective rates $w(2, 0)$, $s(1, 1)$, $w(1, 1)$ and $s(0, 2)$ are positive.
Induction step:
Having reached state (i, j), with $i, j \in \mathbb{N}$ the rates for transitions to $(i+1, j)$ and $(i, j + 1)$ are also positive as long as $i + j < S$. Otherwise, no further states can be reached since $i + j = S$ is at the boundary of the state space (compare Figure 1.3).
Thus, all states of the hybridization process can be reached directly and indirectly (via other states) from $(0, 0)$.

On the other hand, for each transition described so far, there is a return path with positive transition rates. This can be seen equivalently to the previous consideration. The same arguments as for the forward direction hold. As a result $(0, 0)$ can be reached from any other state and thus there is a closed path over all states of the hybridization process. □

Although a unique stationary distribution exists, we still do not know whether the process converges to it fast enough. To solve this problem we need to introduce some theory of stochastic processes.

There is a class of Markov processes, where the speed of convergence can be estimated - the class of *reversible* Markov processes. Loosely speaking, reversibility means if we took a film of such a process and then ran the film backwards, the resulting process would be stochastically indistinguishable from the original process. We will briefly describe some theory of reversible Markov processes.

Definition 2.4. *(see [Kelly], Chapter 1) A stochastic process $X(t)$ is said to be reversible if*

$$(X(t_1), X(t_2), \ldots, X(t_n))$$

has the same distribution as

$$(X(\tau - t_1), X(\tau - t_2), \ldots, X(\tau - t_n))$$

for all $t_1, t_2, \ldots, t_n, \tau \in \mathbb{R}$.

Lemma 2.1.1. *A stationary Markov process with state space Σ and Markov generator $Q = (q_{i,j})_{i,j\in\Sigma}$ is reversible iff its transition rates satisfy*

$$q_{i_2,i_1}q_{i_3,i_2} \cdot \ldots \cdot q_{i_n,i_{n-1}}q_{i_1,i_n} = q_{i_n,i_1}q_{i_{n-1},i_n} \cdot \ldots \cdot q_{i_2,i_3}q_{i_1,i_2}$$

for any finite sequence of states $i_1, i_2, \ldots, i_n \in \Sigma$.

Proof. See [Kelly], Theorem 1.8. □

In other words, a stationary process is reversible iff the product of transition rates going forward through a cycle of states is equivalent to the product of going backwards through it.

If the underlying Markov process is reversible, the speed of convergence can be determined by investigating the eigenvalues of e^Q since

$$\frac{\partial}{\partial t}\psi(t) = Q\psi(t),\ \psi(0) = \psi_0$$

has solution $\psi(t) = e^{Qt}\psi_0$ for any probability distribution $\psi(t)$. According to [Mitro], the eigenvalues of Q are real. Let λ_1 and λ_2 be the largest and second largest eigenvalues of Q. Then, $\lambda_1^* = e^{\lambda_1}$ and $\lambda_2^* = e^{\lambda_2}$ are the largest and second largest eigenvalues of e^Q. According to [Klen], the second largest eigenvalue λ_2^* determines the speed of convergence towards the stationary distribution ρ. It is

$$\|e^{Qt}\psi_0 - \rho\| \leq C|\lambda_2^*|^t \tag{2.8}$$

for a positive constant $C < \infty$. Thus, the closer λ_2^* to zero the faster an initial target distribution will converge to the stationary distribution of the hybridization process. The constant C can be determined according to the prove of Theorem 11.2.1 of [Wink].

In the following, the theory of reversible Markov processes shall be used to examine the hybridization process.

Theorem 2.5. *The hybridization process is reversible.*

Proof. For an illustration of the state space recall Figure 1.3. We will proof the theorem by considering all possible sequences $i_1, i_2, ..., i_n \in \Sigma_{S,2}$.

1. Arbitrary states i, j of the hybridization process satisfy $q_{i,j} \neq 0 \Leftrightarrow q_{j,i} \neq 0$ since every dissociation event can be canceled by a hybridization event and vice versa. This is equivalent to $q_{i,j} = 0 \Leftrightarrow q_{j,i} = 0$. Thus, a zero on the left hand side of Equation (2.8) implies a zero on the right hand side and Equation (2.8) holds. For that reason we restrict further considerations to cycles on the lattice graph, where transitions are characterized by adding or subtracting a single target.

2. Looking at trivial cycles, i.e. i_1 and $i_n = i_2$ are adjacent, Equation (2.8) always holds, since $q_{i_1,i_2}q_{i_2,i_1} = q_{i_2,i_1}q_{i_1,i_2}$ no matter which process is considered.

3. Consider 4-cycles $i_1, i_2, i_3, i_n = i_4$ where i_l, $l = 1, 2, 3, 4$ are pairwise different. We will call cycles whose states are pairwise different *disjunct cycles*. For all subsequent considerations we find it useful to transform Equation (2.8) as follows

$$\frac{q_{i_2,i_1}q_{i_3,i_2} \cdot \ldots \cdot q_{i_n,i_{n-1}}q_{i_1,i_n}}{q_{i_n,i_1}q_{i_{n-1},i_n} \cdot \ldots \cdot q_{i_2,i_3}q_{i_1,i_2}} = 1. \tag{2.9}$$

   W.l.o.g. let i_1 be the north western corner of the cycle and let L_p, $p = 1, 2$ denote the number of hybridized targets of type p in state i_1. Further, we define $\omega_p := S^{-1}\pi_p\lambda$, $p = 1, 2$. Using Equations (2.4)-(2.7) yields

$$\begin{aligned} &\frac{q_{i_2,i_1}q_{i_3,i_2}q_{i_4,i_3}q_{i_1,i_4}}{q_{i_4,i_1}q_{i_3,i_4}q_{i_2,i_3}q_{i_1,i_2}} \\ &= \frac{\omega_1(T_1-L_1)(S-L_1-L_2)\gamma_2 L_2\gamma_1(L_1+1)\omega_2(T_2-L_2+1)(S-L_1-L_2+1)}{\gamma_2 L_2\omega_1(T_1-L_1)(S-L_1-L_2+1)\omega_2(T_2-L_2+1)(S-L_1-1-L_2+1)\gamma_1(L_1+1)} \\ &= 1 \end{aligned}$$

   Hence, arbitrary disjunct 4-cycles satisfy Equation (2.9).

4. Consider disjunct cycles $i_1, i_2, \ldots, i_n$ where $n > 4$. Since $\Sigma_{S,2} \subset \mathbb{Z}^2$, we can classify cycles by the area A the corresponding path encloses. For example, each disjunct 4-cycle encloses an area of $A = 1$. Arbitrary disjunct cycles satisfy Equation (2.9). This is now shown by induction over the enclosed area.
Base case $A = 1$: This is the case of disjunct 4-cycles shown in No. 3.
Induction step $A = m \to A = m + 1$: W.l.o.g let i_1 be the northernmost north western corner (largest L_2 component among all north western corners) of the cycle. We distinguish two cases.

First case: State i_3 is southern of i_2.

$$
\begin{aligned}
& \frac{q_{i_2,i_1} q_{i_3,i_2} \cdot \ldots \cdot q_{i_n,i_{n-1}} q_{i_1,i_n}}{q_{i_n,i_1} q_{i_{n-1},i_n} \cdot \ldots \cdot q_{i_2,i_3} q_{i_1,i_2}} \\
= \; & \frac{q_{i_n,i_3}^2 q_{i_3,i_n}^2}{q_{i_n,i_3}^2 q_{i_3,i_n}^2} \cdot \frac{q_{i_2,i_1} q_{i_3,i_2} \cdot \ldots \cdot q_{i_n,i_{n-1}} q_{i_1,i_n}}{q_{i_n,i_1} q_{i_{n-1},i_n} \cdot \ldots \cdot q_{i_2,i_3} q_{i_1,i_2}} \\
= \; & \underbrace{\frac{q_{i_2,i_1} q_{i_3,i_2} q_{i_n,i_3} q_{i_1,i_n}}{q_{i_n,i_1} q_{i_3,i_n} q_{i_2,i_3} q_{i_1,i_2}}}_{\text{disjunct 4-cycle}} \cdot \underbrace{\frac{q_{i_3,i_n} q_{i_4,i_3} q_{i_5,i_4} \cdot \ldots \cdot q_{i_{n-1},i_{n-2}} q_{i_n,i_{n-1}}}{q_{i_{n-1},i_n} q_{i_{n-2},i_{n-1}} \cdot \ldots \cdot q_{i_4,i_5} q_{i_3,i_4} q_{i_n,i_3}}}_{\text{disjunct cycle of area } A = m} \\
= \; & 1 \cdot 1 \\
= \; & 1
\end{aligned}
$$

Second case: State i_3 is eastern of i_2.
We denote the state south of i_2 by j.

$$
\begin{aligned}
& \frac{q_{i_2,i_1} q_{i_3,i_2} \cdot \ldots \cdot q_{i_n,i_{n-1}} q_{i_1,i_n}}{q_{i_n,i_1} q_{i_{n-1},i_n} \cdot \ldots \cdot q_{i_2,i_3} q_{i_1,i_2}} \\
= \; & \frac{q_{i_n,j}^2 q_{j,i_n}^2 q_{i_2,j}^2 q_{j,i_2}^2}{q_{i_n,j}^2 q_{j,i_n}^2 q_{i_2,j}^2 q_{j,i_2}^2} \cdot \frac{q_{i_2,i_1} q_{i_3,i_2} \cdot \ldots \cdot q_{i_n,i_{n-1}} q_{i_1,i_n}}{q_{i_n,i_1} q_{i_{n-1},i_n} \cdot \ldots \cdot q_{i_2,i_3} q_{i_1,i_2}} \\
= \; & \underbrace{\frac{q_{i_2,i_1} q_{j,i_2} q_{i_n,j} q_{i_1,i_n}}{q_{i_n,i_1} q_{j,i_n} q_{i_2,j} q_{i_1,i_2}}}_{\text{disjunct 4-cycle}} \cdot \underbrace{\frac{q_{j,i_n} q_{i_2,j} q_{i_3,i_2} q_{i_4,i_3} \cdot \ldots \cdot q_{i_{n-1},i_{n-2}} q_{i_n,i_{n-1}}}{q_{i_{n-1},i_n} q_{i_{n-2},i_{n-1}} \cdot \ldots \cdot q_{i_3,i_4} q_{i_2,i_3} q_{j,i_2} q_{i_n,j}}}_{\text{disjunct cycle of area } A = m} \\
= \; & 1 \cdot 1 \\
= \; & 1
\end{aligned}
$$

Hence, arbitrary disjunct cycles satisfy Equation (2.9).

5. Finally, we have to look at cycles which are not disjunct. These contain states which are visited twice in the forward direction as well as in the backward direction. Splitting the cycles at these states yields

components which are either trivial cycles or disjunct cycles. Thus, rearranging the rates along a cycle which is not disjunct according to the described split yields Equation (2.9).

Therefore, the hybridization process is reversible. □

Due to the reversibility of the hybridization process we can use Equation (2.8) to analyze the speed of convergence towards its stationary distribution.

The examination of the eigenvalues is summarized in Table 2.4.

	equal probabilities					**unequal probabilities**				
S	λ_1	λ_2	λ_1^*	λ_2^*	τ_{comp}	λ_1	λ_2	λ_1^*	λ_2^*	τ_{comp}
10	0	$-.5002$	1	.6064	< 1	0	$-.5700$	1	.5655	< 1
50	0	$-.5004$	1	.6063	< 1	0	$-.5704$	1	.5653	< 1
100	0	$-.5005$	1	.6062	3	0	$-.5706$	1	.5652	3
200	0	$-.5012$	1	.6058	43	0	$-.5713$	1	.5648	32
500	0	$-.5034$	1	.6045	2,164	0	$-.5732$	1	.5637	1,915
1,000	0	$-.5067$	1	.6025	19,100	0	$-.5766$	1	.5618	18,805
2,000	0	$-.5135$	1	.5984	92,271	0	$-.5832$	1	.5581	90,890

Table 2.4: S denotes the number of probes per spot, λ_1 and λ_2 the largest and second largest eigenvalues of Q , λ_1^* and λ_2^* the largest and second largest eigenvalues of e^Q. τ_{comp} is the computing time in seconds for calculating the eigenvalues in MATLAB on a Pentium III, 3.19 GHz, 3 GB RAM. The case of 2,000 probes has been calculated on an Intel Core 2 Duo, 2.4 GHz, 4 GB RAM. All values in the equal probabilities column have been determined for the parameter situation of Table 2.1 whereas all values in the unequal probabilities column have been determined for the parameter situation of Table 2.2 except for the number of probes which has been modified according to the first column of the table.

Obviously, all eigenvalues λ_2^* are real and decreasing as S increases. They are approximately 0.6 which yields quite a fast convergence towards the stationary distribution. From the investigation of the eigenvalues in Table 2.4 we can also get an idea of how the convergence behaves for increasing probe numbers. Here we find, that λ_2^* decreases as S increases, i.e. the speed of convergence increases, too. This aspect will be important for the limit of Section 2.1.2. Determining the stationary distribution of the hybridization process is the next step.

The system of linear equations $Q\rho = 0$ has been solved numerically in MATLAB with the help of Gauss' algorithm. Unfortunately, common methods which try to take advantage of the sparse structure of Q fail due to the unusual pattern of entries in Q.

As already mentioned, the total number of probes per spot varies between several millions and several hundreds of millions and the sum of all targets is even 150 times larger. Solving such a system of equations is impossible due to computational power. Hence, we solved the system of equations for the simplified parameter situation. As already mentioned, the speed of convergence increases with increasing probe numbers which can be seen from the eigenvalues in Table 2.4. Thus, we do not expect any surprises concerning the convergence for larger probe numbers.

For equal hybridization and dissociation probabilities and target and probe numbers as shown in Table 2.1 the stationary distribution of the process could be calculated in a tolerable time span of about two hours on a Pentium III, 3.19 GHz, 3 GB RAM. It is shown in Figure 2.1. A peak can be

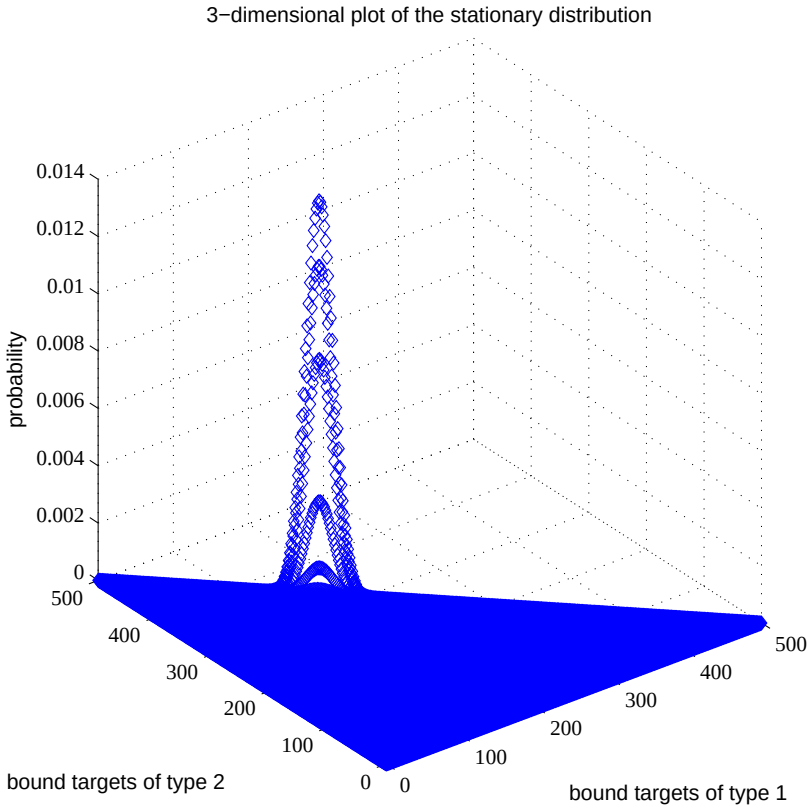


Figure 2.1: The numerical stationary distribution of the hybridization process with parameters as shown in Table 2.1.

seen at about $N_1 = 166$, $N_2 = 333$. It corresponds to a fraction of hybridized

targets of type 1 to those of type 2 of approximately 1/2 which contributes to the idea that the ratio of initial free targets is reproduced in the ratio of hybridized targets.

To calculate ratios is a common method to quantify the number of hybridized targets of the signals coming from the two target types. As can be seen in Table 2.1, the binding and dissociation probabilities are chosen to be equal for both targets. Thus, the ratio of the number of hybridized targets of both types

$$R(t) = \frac{N_1(t)}{N_2(t)},\ t \in \{0, d\}$$

should be close to 1/2 because there are initially twice as many targets of the first type than of the second type. The reason for using ratios is that the intensity values themselves cannot be interpreted in the form of giving an indication for the amount of hybridized targets. But under the assumption of intensity values being directly proportional to the number of hybridized targets, the ratio indicates the fold change between the two intensities. A more commonly used measure is the so-called *log ratio* $R_{log} := \log \frac{N_1(t)}{N_2(t)}$ (see for example [Speed]). It is used to scale down large intensity values. Using the logarithm, the values are scaled down to smaller values. The log ratio is numerically unstable for small intensities (compare [Ultsch]). Thus we concentrate our further analysis on the simple ratio but will also report the log ratio in some of the tables.

Note, so far we considered the case of equal hybridization and dissociation probabilities. Subsequently, the behavior of the process shall be investigated if this assumption does not hold.

Since no realistic values for the hybridization and dissociation probabilities are known we decided to choose the values to be slightly but noticeable different as shown in Table 2.2. The results are shown in Figure 2.2.

The stationary distribution has a sharp peak at $(N_1, N_2) = (131, 368)$. It corresponds to a ratio of $R \approx 0.356$ which is contradictory to the ratio of initial target numbers of 1/2. So, under the assumption of unequal hybridization and dissociation probabilities the process has a stationary distribution which deviates from the one seen in the situation of equal probabilities even though the initial target numbers are the same. Thus, if inferring the initial number of targets, it is important to account for unequal probabilities. Otherwise, drawn conclusions are incorrect and do not reflect the true ratio of initial targets.

The results for the mean μ and the variance σ^2 of both parameter situations are summarized in Table 2.5.

Unfortunately, calculating the solution of the system of linear equations

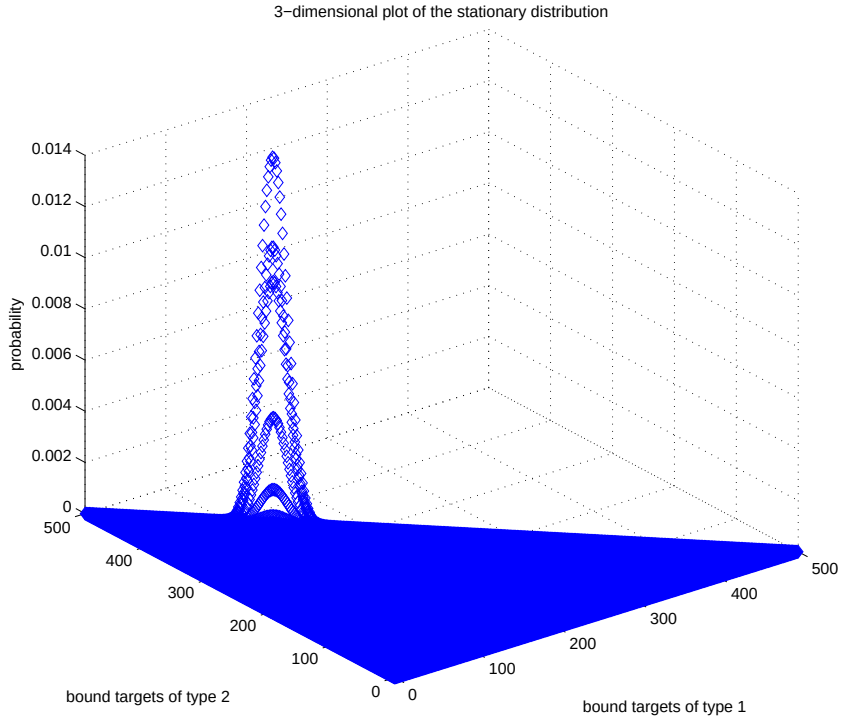


Figure 2.2: The numerical stationary distribution of the hybridization process with parameters as shown in Table 2.2.

is approximately as expensive as the simulation. This is due to the number of states of the process, which increases too fast in m and S as stated in Theorem 2.2. Approaches to use the sparseness of matrix Q in order to improve the computational power of the algorithm failed due to the asymmetric structure of Q. Hence, we have to think of another strategy to come close to realistic parameter values of about 6×10^8 probe molecules per spot.

Substituting the stochastic by a deterministic process and applying the limit $S \to \infty$ might help to draw near to results of realistic parameter situations of very large S.

2.1.2 A deterministic limit for the peak of the stationary distribution

In this section we will use the approach from [Kurtz] who defines a limit for a family of stochastic population processes which approaches a determinis-

	$\mu(N_1)$	$\sigma^2(N_1)$	$\mu(N_2)$	$\sigma^2(N_2)$	$\mu(R)$	$\sigma^2(R)$	$\mu(R_{log})$	$\sigma^2(R_{log})$
equal prob.	166.3	110.2	332.5	110.6	.5015	.0023	$-.6946$	.0090
unequal prob.	131.4	96.3	367.2	96.9	.3589	.0013	-1.0300	.0103

Table 2.5: Results from solving $Q\rho = 0$ for the situations of equal (see Table 2.1) and unequal (see Table 2.2) hybridization and dissociation probabilities).

tic model. The error due to using the deterministic model instead of the stochastic is estimated and shown to be zero as the population size tends to infinity. We will use this limit in order to estimate the peak of the stationary distribution since this can be done much faster than actually calculating the stationary distribution itself as shown above.

According to [Kurtz], if the rates of a Markov jump process $(X_S)_{S\in\mathbb{N}}$ can be rewritten as

$$q^{(S)}_{k,k+l} = S\beta_l(S^{-1}k) \tag{2.10}$$

where $k, l \in \mathbb{Z}^d$ and S is a parameter which is of the same order of magnitude as the population size, under appropriate conditions the family of processes $\{X_S\}$ satisfies $\lim_{S\to\infty} S^{-1}X_S(t) = X(t)$, in probability, where $X(t)$ is a solution of the differential equation

$$\dot{X} = \sum_l l\beta_l(X) =: F(X)$$

with initial value $X(0)$. The precise formulation is the following theorem by [Kurtz].

Theorem 2.6. *Let X_S be a d-dimensional Markov jump process whose rates satisfy Equation (2.10). Suppose for each bounded set $K \subseteq \mathbb{R}^d$,*

$$\sum_l \mid l \mid \sup_{u\in K} \beta_l(u) < \infty$$

and there exists a finite constant $M_K > 0$ such that for all $u, v \in K$

$$\mid F(u) - F(v) \mid \leq M_K \mid u - v \mid . \tag{2.11}$$

If the solution of

$$X(t) = X(0) + \int_0^t F(X(\tau))d\tau$$

exists for all $t \geq 0$ and $X_S(0) \to X(0)$ in probability, then for all $t \geq 0$

$$\lim_{S\to\infty} \sup_{\tau\leq t} \mid X_S(\tau) - X(\tau) \mid = 0$$

in probability, too.

So, in a first step we will have to show that the rates of the hybridization process are of the form of Equation (2.10).

Lemma 2.1.2. *The rates of the hybridization process satisfy Equation (2.10).*

Proof. We will show the lemma for a single hybridization and dissociation rate. The others follow by terms of symmetry.

The hybridization rate $w(L_1, L_2)$ from Equation (2.4) of targets of type 1 can directly be rewritten as

$$w(L_1, L_2) = \begin{cases} S\pi_1 \frac{(T_1 - L_1 + 1)}{S} \frac{S - L_1 + 1 - L_2}{S} \lambda, & L_1 > 0 \\ 0, & \text{else.} \end{cases}$$

which is already the form of Equation (2.10). The dissociation rates $e(L_1, L_2)$ for targets of type 1 from Equation (2.6) can also be directly rewritten as

$$e(L_1, L_2) = \begin{cases} S\gamma_1 \frac{(L_1 + 1)}{S}, & L_1 < S \\ 0, & \text{else.} \end{cases}$$

Hence, the rates are of the demanded form. □

The second step consists of proving whether our process fulfills the conditions of Theorem 2.6. The rates β_l are polynomials. Thus, in a bounded set K they are bounded themselves, i.e. their supremum is finite. Since the number of possible transitions is also finite (≤ 4), the sum $\sum_l \mid l \mid \sup_{u \in K} \beta_l(u)$ is finite, too. On the other hand, the components of $F(u)$ are polynomials and thus $F(u)$ fulfills the Lipschitz condition from Equation (2.11) in K. It remains to prove $X_S(0) \to X(0)$. But this trivially is true because $X_S(0) \equiv 0$ for all S, i.e. the process always starts with an unhybridized spot.

So, the rates are of the expected form and the conditions of Kurtz' Theorem (Theorem 2.6) are satisfied. Thus, we can apply the limit in order to estimate the state at time t of our hybridization process, which gives the initial value problem

$$\dot{X}(t) = \sum_l l\beta_l(X(t)),\ X(0) = 0. \tag{2.12}$$

Solving this equation is as difficult as solving the PDE from Section 2.1.3 as is shown later. But it might be enough to look at large times, since we are interested in the long term behavior of the process.

For this purpose it is necessary that the process converges sufficiently fast to its stationary point. This aspect has been looked at in Section 2.1.1. Increasing S seems to increase the speed of convergence of the hybridization

process towards its stationary distribution and therefore to the limit received with Kurtz' Theorem (Theorem 2.6). Later on, this will be corroborated by the simulation results from Section 2.1.4.

For further investigations the limit which is used here shall be stated more precisely. It is

$$\begin{aligned} & S \to \infty,\ T_1 \to \infty,\ T_2 \to \infty, \\ & \text{such that } \left(\frac{T_1}{S}, \frac{T_2}{S}\right) \to (\alpha_1, \alpha_2) \\ & \text{and } \left(\frac{L_1}{S}, \frac{L_2}{S}\right) \to (x, y). \end{aligned} \tag{2.13}$$

This limit keeps the relative amounts of targets (compared to the number of probes on the spot) constant at the level of the original relative amounts, i.e. α_1 for the first type and α_2 for the second type.

On the one hand, combining the rates $w(L_1, L_2)$ and $s(L_1, L_2)$ with Limit (2.14) yields

$$\beta_{(1,0)}(x, y) \quad = \quad \pi_1(\alpha_1 - x)(1 - x - y)\lambda$$

and

$$\beta_{(0,1)}(x, y) \quad = \quad \pi_2(\alpha_2 - y)(1 - x - y)\lambda.$$

On the other hand, combining the rates $e(L_1, L_2)$ and $n(L_1, L_2)$ with Limit (2.14) yields

$$\beta_{(-1,0)}(x, y) \quad = \quad \gamma_1 x$$

and

$$\beta_{(0,-1)}(x, y) \quad = \quad \gamma_2 y.$$

At this point we can specify Equation (2.12) to the ordinary differential equation of interest

$$\begin{aligned} \frac{\partial}{\partial t}(x(t), y(t)) =& \sum_l l\beta_l(X) \\ =& (1,0)\beta_{(1,0)}(x, y) + (0,1)\beta_{(0,1)}(x, y) + (-1,0)\beta_{(-1,0)}(x, y) \\ & + (0,-1)\beta_{(0,-1)}(x, y) \\ =& (1,0)\pi_1(\alpha_1 - x)(1 - x - y)\lambda + (0,1)\pi_2(\alpha_2 - y)(1 - x - y)\lambda \\ & + (-1,0)\gamma_1 x + (0,-1)\gamma_2 y \\ =& \left(\pi_1 \left(\alpha_1 - x\right)\left(1 - x - y\right)\lambda - \gamma_1 x, \pi_2 \left(\alpha_2 - y\right)\left(1 - x - y\right)\lambda - \gamma_2 y\right). \end{aligned} \tag{2.14}$$

Subsequently, we will investigate the stationary points of Equation (2.14), which we get from $\dot{X} = \frac{\partial}{\partial t}(x(t), y(t)) = 0$ and hence

$$(0,0) = (\pi_1 (\alpha_1 - x)(1 - x - y)\lambda - \gamma_1 x, \pi_2 (\alpha_2 - y)(1 - x - y)\lambda - \gamma_2 y).$$

Thus we have to solve the following system of equations,

$$\begin{aligned} 0 &= \pi_1(\alpha_1 - x)(1 - x - y)\lambda - \gamma_1 x \\ 0 &= \pi_2(\alpha_2 - y)(1 - x - y)\lambda - \gamma_2 y. \end{aligned}$$

Solving the first equation for y and combining it with the second equation leads to finding the roots of a polynomial of third order in x, i.e.

$$\begin{aligned} p(x) := & -\lambda\pi_1^2\alpha_1^2\gamma_2 \\ & +\pi_1\alpha_1\{\lambda\pi_1\gamma_2(2 + \alpha_1) + \gamma_1(-\lambda\pi_2 + \lambda\pi_2\alpha_2 + \gamma_2)\}x \\ & +\{\pi_2\gamma_1^2 + \pi_1\gamma_1(\lambda\pi_2(1 + \alpha_1 - \alpha_2) - \gamma_2) - \lambda\pi_1^2\gamma_2(1 + 2\alpha_1)\}x^2 \\ & +\{\lambda\pi_1(\pi_1\gamma_2 - \pi_2\gamma_1)\}x^3. \end{aligned} \tag{2.15}$$

This is quite difficult as long as the parameters π_1, π_2, α_1, α_2, γ_1, γ_2 and λ are unknown. Here, at least α_1, α_2, $\pi_1\lambda/\gamma_1$ and $\pi_2\lambda/\gamma_2$ are free after scaling. Solving Equation (2.15) yields huge and complex expressions without providing further insight. See Section A in the appendix for the solution provided by MATHEMATICA. Once the parameters are specified, the roots can be found fast with the help of computer algebra programs.

But before, we will investigate some features of the solution. A first observation leads to the simplex of valid solutions

$$\Sigma = \{(x, y) \in \mathbb{R}^2 \mid x + y \leq 1 \text{ and } x, y \geq 0\}. \tag{2.16}$$

The similarity of the notation to $\Sigma_{S,m}$ in Equation (1.1) is intended. Via Limit (2.14) we have $\Sigma = \Sigma_{\infty,2}$.
Solving the first equation of System (2.15) for y and placing it in Simplex (2.16) yields

$$0 \leq y = \frac{(1 - x)\lambda\pi_1(x - \alpha_1) + \gamma_1 x}{\lambda\pi_1(x - \alpha_1)} \leq 1 - x.$$

This inequality was solved in MATHEMATICA. It holds for all $x \in [0, x_{max}]$, α_1, λ, π_1, $\gamma_1 > 0$ with

$$x_{max} = \frac{1}{2}\left(1 + \alpha_1 + \frac{\gamma_1}{\lambda\pi_1} - \sqrt{(\alpha_1 - 1)^2 + \frac{2(\alpha_1 + 1)\gamma_1}{\lambda\pi_1} + \frac{\gamma_1^2}{\lambda^2\pi_1^2}}\right) \in (0, 1].$$

Theorem 2.7. *There is always a unique solution (x^*, y^*) of Equation System (2.15) which is in Σ.*

Proof. In order to prove the existence of a valid solution of (2.15), we have to find roots of (2.15) in $[0, x_{max}]$. According to Bolzano's theorem about roots of a continuous function $f(x)$, there is at least one root $x \in [a, b]$, if the $sign(f(a)) \neq sign(f(b))$ (see [Heus]). Polynomials are always continuous. It remains to prove $sign(p(0)) \neq sign(p(x_{max}))$. It is

$$\begin{aligned} p(0) &= -\lambda\pi_1\alpha_1\gamma_2 < 0 \quad \text{and} \\ p(x_{max}) &= \frac{\pi_2\alpha_2\gamma_1}{2\lambda\pi_1}\Bigg((\lambda\pi_1+\gamma_1)\Bigg(\lambda\pi_1\Bigg(\sqrt{(\alpha_1-1)^2+\frac{2\gamma_1(\alpha_1+1)}{\lambda\pi_1}+\frac{\gamma_1^2}{\lambda^2\pi_1^2}} \\ &\quad -1\Bigg)-\gamma_1\Bigg)+\lambda\pi_1\alpha_1(\lambda\pi_1-\gamma_1)\Bigg) > 0 \end{aligned}$$

for α_1, α_2, π_1, π_2, γ_1, γ_2, $\lambda > 0$. Thus, there is at least one root in $(0, x_{max})$. Subsequently, we will show that there is exactly one root in $(0, x_{max})$. We know that $p(x)$ is a polynomial of third order and thus has at most three real roots. So, showing two other real roots exist outside the interval $[0, x_{max}]$ will imply that there is exactly one root inside the interval since roots with imaginary part unequal to zero always occur in pairs. Looking at the behavior of $p(x)$ as x tends to $\pm\infty$ yields:

$$\begin{aligned} \lim_{x\to\infty} p(x) &= \infty \cdot sign(\pi_2\gamma_1 - \pi_1\gamma_2) \text{ and} \\ \lim_{x\to-\infty} p(x) &= -\infty \cdot sign(\pi_2\gamma_1 - \pi_1\gamma_2). \end{aligned}$$

For $\pi_2\gamma_1 - \pi_1\gamma_2 > 0$ we can use Bolzano's theorem once more in order to show that there is a root in $(-\infty, 0)$ and another root in (x_{max}, ∞) since $p(0) < 0$ and $p(x_{max} > 0)$. But this implies that there is exactly one root in $(0, x_{max})$. Bolzano's theorem can also help us with the case $\pi_2\gamma_1 - \pi_1\gamma_2 < 0$. All we need to do is to switch into the picture for y, which is solving the second equation of (2.15) for x and combining it with the first equation. This yields finding the roots of a polynomial of third order in y. The same arguments as previously used imply that there is exactly one root in $(0, y_{max})$ with

$$y_{max} = \frac{1}{2}\left(1+\alpha_2+\frac{\gamma_2}{\lambda\pi_2}-\sqrt{(\alpha_2-1)^2+\frac{2(\alpha_2+1)\gamma_2}{\lambda\pi_2}+\frac{\gamma_2^2}{\lambda^2\pi_2^2}}\right) \in (0,1]\,.$$

It remains to look at the case $\pi_2\gamma_1 - \pi_1\gamma_2 = 0$. Here, we find two real roots which are

$$\begin{aligned} x^* &= \frac{1}{2\lambda\pi_1(\alpha_1+\alpha_2)}\Big(\alpha_1\big(\lambda\pi_1(1+\alpha_1+\alpha_2)+\gamma_1\big) \\ &\quad -\sqrt{\alpha_1^2(\lambda^2\pi_1^2(-1+\alpha_1+\alpha_2)^2+2\lambda\pi_1(1+\alpha_1+\alpha_2)\gamma_1+\gamma_1^2)}\Big) \end{aligned}$$

and

$$\begin{aligned} x^+ &= \frac{1}{2\lambda\pi_1(\alpha_1+\alpha_2)}\Big(\alpha_1\big(\lambda\pi_1(1+\alpha_1+\alpha_2)+\gamma_1\big) \\ &\quad +\sqrt{\alpha_1^2(\lambda^2\pi_1^2(-1+\alpha_1+\alpha_2)^2+2\lambda\pi_1(1+\alpha_1+\alpha_2)\gamma_1+\gamma_1^2)}\Big) \end{aligned}$$

With the help of MATHEMATICA we verified, that $x^* \in (0, x_{max})$ and $x^+ \notin (0, x_{max})$.

Consequently, there is exactly one root (x^*, y^*) fulfilling x^*, $y^* \geq 0$ and $x^* + y^* \leq 1$. □

Now, we will have a look at the quality of this root, i.e. its stability.

Theorem 2.8. *The root (x^*, y^*) is asymptotically stable.*

Proof. In order to prove stability we will set up the Jacobian matrix for the right hand side of Equation (2.12), i.e.

$$\begin{aligned} J(x,y) &= \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \\ &= \begin{pmatrix} \frac{\partial}{\partial x}\left(\pi_1(\alpha_1-x)(1-x-y)\lambda-\gamma_1 x\right) & \frac{\partial}{\partial y}\left(\pi_1(\alpha_1-x)(1-x-y)\lambda-\gamma_1 x\right) \\ \frac{\partial}{\partial x}\left(\pi_2(\alpha_2-y)(1-x-y)\lambda-\gamma_2 y\right) & \frac{\partial}{\partial y}\left(\pi_2(\alpha_2-y)(1-x-y)\lambda-\gamma_2 y\right) \end{pmatrix} \\ &= \begin{pmatrix} (-1+2x+y)\lambda\pi_1-\lambda\pi_1\alpha_1-\gamma_1 & -\lambda\pi_1\left(-x+\alpha_1\right) \\ -\lambda\pi_2\left(-y+\alpha_2\right) & (-1+x+2y)\lambda\pi_2-\lambda\pi_2\alpha_2-\gamma_2 \end{pmatrix}. \end{aligned}$$

According to [HeusD], Chapter 10, we need negative real parts of the eigenvalues of $J(x^*, y^*)$ for asymptotic stability. This is equivalent to $Tr(J(x,y)) := a_{11} + a_{22} < 0$ and $det(J(x,y)) := a_{11}a_{22} - a_{12}a_{21} > 0$ ([Britt], Appendix B). We will use the latter condition:

$$\begin{aligned} a_{11}+a_{22} &= (-1+2x+y)\lambda\pi_1-\lambda\pi_1\alpha_1-\gamma_1+(-1+x+2y)\lambda\pi_2-\lambda\pi_2\alpha_2-\gamma_2 \\ &= -\lambda\pi_1(1-x-y)-\lambda\pi_2(1-x-y)-\lambda\pi_1(\alpha_1-x)-\lambda\pi_2(\alpha_2-y)-\gamma_1-\gamma_2 \\ &< 0, \end{aligned}$$

as well as

$$\begin{aligned} a_{11}a_{22}-a_{12}a_{21} &= \left((-1+2x+y)\lambda\pi_1-\lambda\pi_1\alpha_1-\gamma_1\right)\left((-1+x+2y)\lambda\pi_2-\lambda\pi_2\alpha_2-\gamma_2\right) \\ &\quad -\left(-\lambda\pi_1\left(-x+\alpha_1\right)\right)\left(-\lambda\pi_2\left(-y+\alpha_2\right)\right) \\ &= \lambda\pi_2\big(\lambda\pi_1(x+y-1)(2x+2y-1-\alpha_1-\alpha_2)-(x+2y-1-\alpha_2)\gamma_1\big) \\ &\quad -\left(\lambda\pi_1\left(2x+y-1-\alpha_1\right)-\gamma_1\right)\gamma_2 \\ &> 0 \end{aligned}$$

since α_1, α_2, γ_1, γ_2, π_1, π_2, λ are positive and $0 \leq x + y \leq 1$, $x \leq \alpha_1$, $y \leq \alpha_2$. □

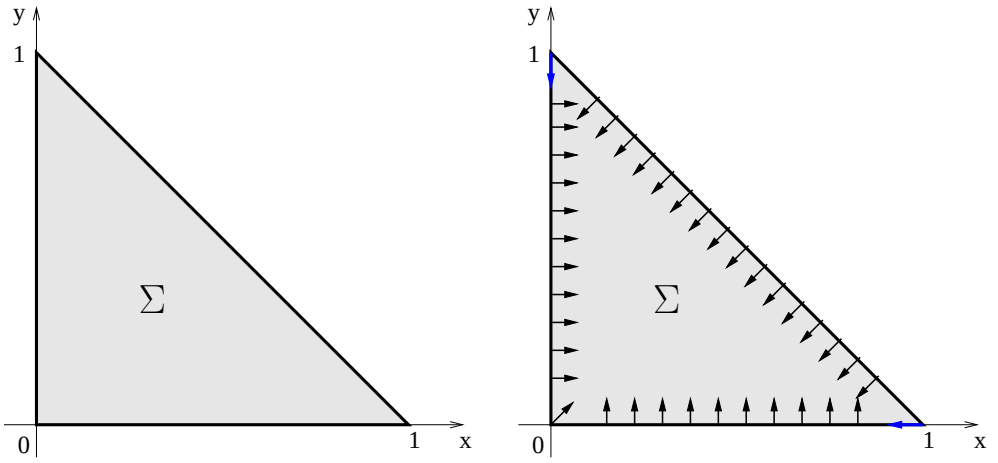
So, within a small neighborhood all solutions of Equation (2.12) are attracted by the stationary point (x^*, y^*). But we need global convergence instead of local which is restricted to a small neighborhood, since it is not obvious that the solution through $(0, 0)$ ever comes close enough to the stationary point.

Theorem 2.9. *Every solution starting in* $(0, 0)$ *converges to* (x^*, y^*).

Proof. As shown above there is only one stationary point within the simplex of valid solutions

$$\Sigma = \{(x, y) \mid x + y \leq 1 \text{ and } x, y \geq 0\}.$$

In a first step we will show that a solution starting in Σ never leaves it. Looking at the right hand side of Equation (2.12) at the boundaries of Σ will tell us whether solutions at the boundaries will leave Σ or will be attracted by the inside area of Σ. The simplex Σ is shown in Figure 2.3(a) We will look at



(a) The simplex of valid solutions.

(b) The simplex including the vector field at the boundaries corresponding to Equation (2.12).

Figure 2.3: Simplex Σ of valid solutions of Equation (2.12).

the boundaries and corners of Σ, separately. The first corner is $x = 0$, $y = 0$. Here, the gradient is

$$\begin{aligned} \begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} &= \begin{pmatrix} \lambda\pi_1(1 - x - y)(\alpha_1 - x) - \gamma_1 x \\ \lambda\pi_2(1 - x - y)(\alpha_2 - y) - \gamma_2 y \end{pmatrix} \\ &= \begin{pmatrix} \lambda\pi_1\alpha_1 \\ \lambda\pi_2\alpha_2 \end{pmatrix}. \end{aligned}$$

Thus, the gradient is positive in both components, i.e. the vector points to the interior of Σ.

At the second corner $x = 0$, $y = 1$ the gradient is

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} 0 \\ -\gamma_2 \end{pmatrix}.$$

Consequently, the gradient points into the direction of boundary $x = 0$, $0 < y < 1$.

Analogously, the gradient at the third corner $x = 1$, $y = 0$ points into the direction of boundary $0 < x < 1$, $y = 0$. It remains to investigate the boundaries. The gradient at the first boundary $x = 0$, $0 < y < 1$ is given by

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} \lambda\pi_1(1-y)\alpha_1 \\ \lambda\pi_2(1-y)(\alpha_2 - y) - \gamma_2 y \end{pmatrix}.$$

The first component of the gradient is greater than zero, i.e the vector field points to the inner area of Σ. The same arguments hold for the second boundary $0 < x_1$, $y = 0$.

At the third boundary $x + y = 1$ the gradient is given by

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} -\gamma_1 x \\ -\gamma_2 y \end{pmatrix},$$

i.e. both components are less than zero and thus the vector field also points to the inner area of Σ. The behavior of the vector field at the boundary of Σ is illustrated in Figure 2.3(b). Consequently, a solution which starts within Σ cannot leave it. Thus, it cannot be attracted by one of the stationary points outside of Σ if such points exist.

In a second step we will show that all solutions which start in Σ are attracted by the stationary point (x^*, y^*). According to the theorem of Poincaré-Bendixson (see [Wigg], Chapter 9), all solutions which start in Σ will be attracted by the stationary point (x^*, y^*), if the following three assumptions hold:

- there is a unique stationary point (x^*, y^*) in Σ,
- the vector field points to the inner of Σ and
- there is no limit cycle in Σ.

The first two assumptions have been shown above. In order to show that there are no limit cycles in Σ we will use Bendixson's negative criterion (see e.g. [JoSm], Chapter 3). It says that there are no limit cycles in Σ if the

divergence $div\, F = \frac{\partial \dot{x}}{\partial x} + \frac{\partial \dot{y}}{\partial y}$ is of one sign for all $(x, y) \in \Sigma$. Indeed the divergence is

$$\begin{aligned} div\, F &= \frac{\partial}{\partial x}\lambda\pi_1(1-x-y)(\alpha_1 - x) - \gamma_1 x + \frac{\partial}{\partial y}\lambda\pi_2(1-x-y)(\alpha_2 - y) - \gamma_2 y \\ &= -(1-x-y+\alpha_1 - x)\lambda\pi_1 - (1-x-y+\alpha_2 - y)\lambda\pi_2 - \gamma_1 - \gamma_2 \\ &< 0 \end{aligned}$$

since $0 \leq x+y \leq 1$, $x \leq \alpha_1$, $y \leq \alpha_2$ and α_1, α_2, γ_1, γ_2, π_1, π_2, λ are positive. So, solutions never leave Σ, there are no limit cycles in Σ and hence the solution through $(0, 0)$ converges to (x^*, y^*). □

On the basis of these results we are able to approximate the peak of the stationary distribution of the stochastic version of hybridization process for a large and realistic number of probes and targets.

This shall be done for respective parameter situations according to the previous examinations. We have to be careful with transforming the parameters in the right way due to the applied limit. Both parameter situation are summarized in Table 2.6. Solving Equation (2.15) yields the stationary

	π_1	π_2	γ_1	γ_2	$\alpha_1 = \frac{T_1}{S}$	$\alpha_2 = \frac{T_2}{S}$	λ
equal probabilities	.7	.7	.5	.5	$\frac{25{,}000}{500} = 50$	$\frac{50{,}000}{500} = 100$	2
unequal probabilities	.6	.7	.6	.5	$\frac{25{,}000}{500} = 50$	$\frac{50{,}000}{500} = 100$	2

Table 2.6: Parameter situations after applying the limit from [Kurtz] to the parameters in Tables 2.1 and 2.2.

point of the initial value problem in Equation (2.12) and thus serves as an approximation of the stationary distribution of the hybridization process. The results are shown in Table 2.7. As can be seen, the solution for equal

	x	$\widehat{=} N_1$	y	$\widehat{=} N_2$	R	R_{log}
equal probabilities	$\approx .333$	≈ 166.3	$\approx .665$	≈ 332.9	$\approx .500$	$-.693$
unequal probabilities	$\approx .263$	≈ 131.4	$\approx .734$	≈ 367.2	$\approx .358$	-1.027

Table 2.7: Results from solving Equation (2.15).

hybridization and dissociation probabilities is as expected $R = .500$, since the process was fed with twice as many targets of type 2 than of type 1. But if we look at the situation of unequal target types we see a deviant value, i.e. $R = .358$. Thus, inferring the initial target concentrations would yield

an overestimation of 39.7% of the second type. This is corroborated by the results from the simulation as can be seen in Section 2.1.4.

Moreover, the results are very close to those received when using the stationary distribution in Section 2.1.1 (compare Table 2.5 and Table 2.7). Hence, the stationary point according to Kurtz serves as a good approximation for the mean of the stationary distribution and vice versa.

In the next section we will try to expand the solution (x^*, y^*) which is a single point to a functional which is defined on entire Σ.

2.1.3 Approximation of the solution of $Q\rho = 0$ with a PDE

In addition to previous approaches, the solution of the system of linear equations $Q\rho = 0$ for large S can be approximated by a partial differential equation as follows.

An appropriate partial differential equation is received via two construction steps. Firstly, we build a difference equation. Secondly we apply Limit (2.14).

Combining Equations (2.4)-(2.7) with Equation (2.3) and simplifying yields

$$\begin{aligned} &\pi_1(T_1 - L_1 + 1)\tfrac{S-L_1+1-L_2}{S}\lambda\rho_{L_1-1,L_2} + \gamma_1(L_1+1)\rho_{L_1+1,L_2} \\ &+\pi_2(T_2 - L_2 + 1)\tfrac{S-L_1-L_2+1}{S}\lambda\rho_{L_1,L_2-1} + \gamma_2(L_2+1)\rho_{L_1,L_2+1} \\ &= \\ &\left((\pi_1(T_1 - L_1) + \pi_2(T_2 - L_2))\tfrac{S-L_1-L_2}{S}\lambda + \gamma_1 L_1 + \gamma_2 L_2\right)\rho_{L_1,L_2} \end{aligned} \tag{2.17}$$

for the case of $L_1, L_2 > 0$, $L_1 + L_2 < S$, i.e. the process is not on the boundary of the state space. Otherwise, the rates which either lead to a state outside the state space or originate from outside are omitted according to Equations (2.4)-(2.7).

Equation (2.17) can be rewritten as an inhomogeneous difference equation, i.e.:

$$\begin{aligned} 0 \;=\; &\pi_1(T_1 - L_1)\frac{S - L_1 - L_2}{S}\lambda(\rho_{L_1-1,L_2} - \rho_{L_1,L_2}) + \gamma_1 L_1(\rho_{L_1+1,L_2} - \rho_{L_1,L_2}) \\ &+\pi_2(T_2 - L_2)\frac{S - L_1 - L_2}{S}\lambda(\rho_{L_1,L_2-1} - \rho_{L_1,L_2}) + \gamma_2 L_2(\rho_{L_1,L_2+1} - \rho_{L_1,L_2}) \\ &+\pi_1\frac{T_1 + S - 2L_1 - L_2 + 1}{S}\lambda\rho_{L_1-1,L_2} + \gamma_1\rho_{L_1+1,L_2} \\ &+\pi_2\frac{T_2 + S - L_1 - 2L_2 + 1}{S}\lambda\rho_{L_1,L_2-1} + \gamma_2\rho_{L_1,L_2+1}. \end{aligned} \tag{2.18}$$

Now, we can apply Limit (2.14). It keeps the relative amount of targets constant. Thus, it is appropriate to give an approximation of the hybridization process with large numbers of targets and probes. We get the following differential equation.

Let $f(x,y)$ be an arbitrary differentiable function, with $f(x,y) = \rho_{(Sx,Sy)} = \rho_{(N_1,N_2)}$. Combining Limit (2.14) with Equation (2.18) and simplifying yields:

$$\begin{aligned} 0 &= (\pi_1(\alpha_1 - 2x - y + 1)\lambda + \pi_2(\alpha_2 - x - 2y + 1)\lambda + \gamma_1 + \gamma_2)f(x,y) \\ &\quad +(\gamma_1 x - \lambda(1-x-y)\pi_1(\alpha_1 - x))\frac{\partial}{\partial x}f(x,y) \\ &\quad +(\gamma_2 y - \lambda(1-x-y)\pi_2(\alpha_2 - y))\frac{\partial}{\partial y}f(x,y) \end{aligned} \tag{2.19}$$

The boundary conditions for this inhomogeneous partial differential equation can be derived similarly. As shown in Section 1.1 the state space is of triangular shape (compare Figure 1.3). We have to take a look at the three sides $L_1 = 0$, $L_2 = 0$ and $L_1 + L_2 = S$ of the triangle.

First corner, $L_1 = 0$, $L_2 = 0$. This is the case of no hybridized targets at all. Here, rates $r_{(L_1-1,L_2),(L_1,L_2)}$, $r_{(L_1,L_2),(L_1-1,L_2)}$, $r_{(L_1,L_2-1),(L_1,L_2)}$ and $r_{(L_1,L_2),(L_1,L_2-1)}$ can be omitted in Equation (2.17). We receive

$$\begin{gathered} \gamma_1(0+1)\rho_{0+1,0} + \gamma_2(0+1)\rho_{0,0+1} \\ = \\ \left((\pi_1(T_1 - 0) + \pi_2(T_2 - 0))\,\tfrac{S-0-0}{S}\lambda\right)\rho_{0,0}. \end{gathered}$$

This equation cannot be rewritten as a difference equation. Instead, after dividing both sides by S the Limit (2.14) can be directly applied, which yields

$$0 = (\pi_1\alpha_1 + \pi_2\alpha_2)\,\lambda\; f(0,0)$$

and thus

$$f(0,0) = 0, \tag{2.20}$$

since λ, α_1, α_2, π_1, $\pi_2 > 0$.

Second corner, $L_1 = 0$, $L_2 = S$. This is the case of all probes being hybridized to targets of type 2. Here, rates $r_{(L_1-1,L_2),(L_1,L_2)}$, $r_{(L_1,L_2),(L_1-1,L_2)}$, $r_{(L_1,L_2+1),(L_1,L_2)}$ and $r_{(L_1,L_2),(L_1,L_2+1)}$ can be omitted in Equation (2.17). We receive

$$\begin{gathered} \gamma_1(0+1)\rho_{0+1,S} + \pi_2(T_2 - S + 1)\tfrac{S-0-S+1}{S}\lambda\rho_{0,S-1} \\ = \\ \left(\pi_1(T_1 - 0)\,\tfrac{S-0-S}{S}\lambda + \gamma_2 S\right)\rho_{0,S}. \end{gathered}$$

This equation cannot be transformed into a difference equation, either. But dividing both sides by S and applying Limit (2.14) yields

$$0 = \gamma_2 f(0,1)$$

and thus

$$f(0,1) = 0 \tag{2.21}$$

since $\gamma_2 > 0$.

Third corner, $L_1 = S$, $L_2 = 0$. This is the case of all probes being hybridized to targets of type 1. The condition for the stationary distribution at this corner can be derived equivalently to the previous case and therefore is

$$f(1,0) = 0. \tag{2.22}$$

Subsequently, we will derive the conditions for the stationary distribution at the inner of the sides of the state space.

First side, $L_1 = 0$, $0 < L_2 < S$. This is the case where only targets of type 2 have hybridized to the spot. Here, rates $r_{(L_1-1,L_2),(L_1,L_2)}$ and $r_{(L_1,L_2),(L_1-1,L_2)}$ can be omitted in Equation (2.17). We receive

$$\begin{gathered}\gamma_1(0+1)\rho_{0+1,L_2} \\ +\pi_2(T_2-L_2+1)\tfrac{S-0-L_2+1}{S}\lambda\rho_{0,L_2-1} + \gamma_2(L_2+1)\rho_{0,L_2+1} \\ = \\ \left((\pi_1(T_1-0)+\pi_2(T_2-L_2))\tfrac{S-0-L_2}{S}\lambda + \gamma_2 L_2\right)\rho_{0,L_2}.\end{gathered}$$

Rewriting it as a difference equation yields:

$$\begin{aligned}0 \;=\; & -\pi_1 T_1 \frac{S-L_2}{S}\lambda\rho_{0,L_2} + \gamma_1\rho_{0+1,L_2} \\ & +\pi_2(T_2-L_2)\frac{S-L_2}{S}\lambda(\rho_{0,L_2-1}-\rho_{0,L_2}) + \gamma_2 L_2(\rho_{0,L_2+1}-\rho_{0,L_2}) \\ & +\pi_2\frac{T_2+S-2L_2+1}{S}\lambda\rho_{0,L_2-1} + \gamma_2\rho_{0,L_2+1}.\end{aligned} \tag{2.23}$$

Dividing by S and applying Limit (2.14) to Equation (2.23) then yields:

$$0 = -\pi_1\alpha_1(1-y)\lambda f(0,y). \tag{2.24}$$

Equation (2.24) holds, iff

$$f(0,y) = 0, \tag{2.25}$$

for $0 < y < 1$ since π_1, α_1, $(1-y) > 0$.

Second side, $0 < L_1 < S$, $L_2 = 0$. This is the case where only targets of type 1 have hybridized to the spot. The condition for the stationary distribution can be derived equivalently to the previous case which yields

$$f(x, 0) = 0 \tag{2.26}$$

for $0 < x < 1$.

Third side, $L_1 + L_2 = S$, $0 < L_1, L_2 < S$. This case describes the situation where each probe on the spot is either covered by a target of type 1 or by a target of type 2. Here, rates $r_{(L_1+1,L_2),(L_1,L_2)}$, $r_{(L_1,L_2+1),(L_1,L_2)}$, $r_{(L_1,L_2),(L_1+1,L_2)}$ and $r_{(L_1,L_2),(L_1,L_2+1)}$ can be omitted in Equation (2.17). Furthermore L_2 can be substituted by $S - L_1$. Thus, we receive

$$\begin{gathered}\pi_1(T_1 - L_1 + 1)\tfrac{S-L_1+1-(S-L_1)}{S}\lambda\rho_{L_1-1,S-L_1} \\ +\pi_2(T_2 - (S - L_1) + 1)\tfrac{S-L_1-(S-L_1)+1}{S}\lambda\rho_{L_1,S-L_1-1} \\ = \\ (\gamma_1 L_1 + \gamma_2(S - L_1))\,\rho_{L_1,S-L_1}.\end{gathered}$$

This equation cannot be rewritten as difference equation but dividing by S and applying Limit (2.14) yields

$$(\gamma_1 x + \gamma_2(1 - x))\, f(x, 1 - x) = 0,$$

and thus

$$f(x, 1 - x) = 0, \tag{2.27}$$

for $0 < x, y < 1$ since γ_1, γ_2, x, $1 - x > 0$.

In summary, Equation (2.19) together with Equations (2.20), (2.21), (2.22), (2.25), (2.26) and (2.27) yield the partial differential equation

$$\begin{aligned}0 \;=\; & (\pi_1(\alpha_1 - 2x - y + 1)\lambda + \pi_2(\alpha_2 - x - 2y + 1)\lambda + \gamma_1 + \gamma_2)f(x, y) \\ & +(\gamma_1 x - \lambda(1 - x - y)\pi_1(\alpha_1 - x))\frac{\partial}{\partial x}f(x, y) \\ & +(\gamma_2 y - \lambda(1 - x - y)\pi_2(\alpha_2 - y))\frac{\partial}{\partial y}f(x, y)\end{aligned}$$

in the region

$$\Sigma = \{(x, y) \mid x, y \geq 0,\, x + y \leq 1\}$$

and with Dirichlet boundary condition

$$f(x, 0) = f(0, y) = f(x, 1 - x) = 0.$$

In order to solve this equation, we will use the method of characteristics as introduced for example in [KamII], Chapter 2. The characteristics of the solution surface satisfy the equations:

$$\begin{aligned} \frac{dx}{dt} &= (\gamma_1 x - \lambda(1-x-y)\pi_1(\alpha_1 - x)) \\ \frac{dy}{dt} &= (\gamma_2 y - \lambda(1-x-y)\pi_2(\alpha_2 - y)) \\ \frac{dz}{dt} &= -(\pi_1(\alpha_1 - 2x - y + 1)\lambda + \pi_2(\alpha_2 - x - 2y + 1)\lambda + \gamma_1 + \gamma_2)z \end{aligned} \tag{2.28}$$

Since the first two equations in System (2.28) do not depend on z, their solution can formally be written in the form

$$x = X(t \mid x(t_0) = x_0),\ y(t) = Y(t \mid y(t_0) = y_0), \tag{2.29}$$

with initial conditions given by $x(t_0) = x_0$, $y(t_0) = y_0$. Now we substitute Equations (2.29) into the third equation of System (2.28), which yields

$$\frac{dz}{dt} = -(\pi_1(\alpha_1 - 2X - Y + 1)\lambda + \pi_2(\alpha_2 - X - 2Y + 1)\lambda + \gamma_1 + \gamma_2)z.$$

This equation can be integrated to yield

$$z(t) = z(t_0) exp\left(-\int_{t_0}^{t} R(\tau)d\tau\right), \tag{2.30}$$

where $R(t) := -(\pi_1(\alpha_1 - 2X(t) - Y(t) + 1)\lambda + \pi_2(\alpha_2 - X(t) - 2Y(t) + 1)\lambda + \gamma_1 + \gamma_2)z$.

Obviously, if $z(t_0) = 0$, the z-coordinate of the solution surface is identically equal to zero, i.e. $z(t) = 0$ for all t. Thus the only possible continuous solution is the trivial solution $f(x, y) = 0$.

The standard approach with power series corroborates this result, since all coefficients vanish. Numeric approaches like the finite elements method also always yielded $f = 0$.

The probable reason is that the normalized (states divided by S) bell shaped curve of the stationary distribution (compare Figures 2.10(a) and 2.11(a)) narrows (in terms of variance) too fast, i.e. with factor $S/S^2 = 1/S$ as observed during the investigation of the stationary distribution when solving the system of linear equations for different probe numbers. So the limit might be a distribution which is zero except for one point, where it has a δ-peak.

As seen in Section 2.1.2 the deterministic limit provides the single point (x^*, y^*) as long term state of the process. This together with the intuition

from investigating the variance of the stationary distribution as the number of states tends to infinity yields an approach for a non trivial solution of the PDE. Therefor, we will need some aspects of functional analysis, especially distribution theory. For reference of the next paragraph see [Dobr], Chapter 9 and [Wern], Chapter 8.

Definition 2.10. *Let*

$$supp(u) := \overline{\{x : u(x) \neq 0\}}$$

be the support *of function u. Then, let C_0^∞ denote the space of infinitely often differentiable functions with compact support.*

Definition 2.11. *$K \subset\subset \Omega$ iff $\overline{K}$ compact and $\overline{K} \subset \Omega$. We say K is* compactly enclosed *in Ω.*

Definition 2.12. *Let $\Omega \subset \mathbb{R}^2$ and (ϕ_k) be a sequence in $C_0^\infty(\Omega)$. We say, (ϕ_k) converges to $\phi \in C_0^\infty(\Omega)$ (denoted by $\phi_k \xrightarrow{\mathcal{D}} \phi$), if there is a compact set $K \subset\subset \Omega$ with $supp(\phi_k), supp(\phi) \subset K$ and if $D^\alpha \phi_k \to D^\alpha \phi$ uniformly in Ω for all multi-indices α. C_0^∞ together with this convergence definition is denoted $\mathcal{D}(\Omega)$. $\phi \in C_0^\infty(\Omega)$ is called* test function.

Definition 2.13. *$T : \mathcal{D}(\Omega) \to \mathbb{C}$ is called a* distribution *if T is linear and it satisfies*

$$\phi_k \xrightarrow{\mathcal{D}} \phi \Rightarrow T(\phi_k) \to T(\phi).$$

For our PDE we set $\Omega = \Sigma \subset \mathbb{R}^2$.

Theorem 2.14. *The delta distribution defined by*

$$\delta_{(x^*,y^*)}(\phi) := \phi(x^*, y^*) \ \forall \phi \in \mathcal{D}(\Sigma)$$

solves the PDE from Equation (2.19), where (x^*, y^*) is the solution of Equation (2.15).

Proof. To prove this theorem we will use the following lemma.

Lemma 2.1.3. *According to [Dobr], Chapter 9, the delta distribution fulfills*

$$p\delta_{(x^*,y^*)}(\phi) = p(x^*, y^*)\phi(x^*, y^*) \text{ for } p \in C^\infty(\Sigma) \tag{2.31}$$

and its derivatives fulfill

$$D^\beta \delta_{(x^*,y^*)}(\phi) = (-1)^{|\beta|} D^\beta \phi(x^*, y^*) \tag{2.32}$$

for all multi indices β, with $\beta = (\beta_1, \beta_2)$, $|\beta| = \beta_1 + \beta_2$ and $\beta_1, \beta_2 \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}$.

Subsequently, we will use the following abbreviations:

$$\begin{aligned} c_1(x,y) &:= (\pi_1(\alpha_1 - 2x - y + 1)\lambda + \pi_2(\alpha_2 - x - 2y + 1)\lambda + \gamma_1 + \gamma_2) \\ c_2(x,y) &:= (\gamma_1 x - \lambda(1 - x - y)\pi_1(\alpha_1 - x)) \\ c_3(x,y) &:= (\gamma_2 y - \lambda(1 - x - y)\pi_2(\alpha_2 - y)). \end{aligned}$$

Rewriting Equation (2.19) in a distributional manner yields

$$\begin{aligned} 0 &= [c_1(x,y)\delta_{(x^*,y^*)}](\phi(x,y)) \\ &\quad +[c_2(x,y)\frac{\partial}{\partial x}\delta_{(x^*,y^*)}](\phi(x,y)) \\ &\quad +[c_3(x,y)\frac{\partial}{\partial y}\delta_{(x^*,y^*)}](\phi(x,y)). \end{aligned}$$

for all test functions ϕ. The coefficients $c_1(x,y)$, $c_2(x,y)$ and $c_3(x,y)$ are polynomials and thus are in $C^\infty(\Sigma)$. Using Lemma 2.1.3 leads to

$$\begin{aligned} 0 &= c_1(x^*,y^*)\phi(x^*,y^*) \\ &\quad - \left[\frac{\partial}{\partial x}(c_2(x,y)\phi(x,y))\right](x^*,y^*) \\ &\quad - \left[\frac{\partial}{\partial y}(c_3(x,y)\phi(x,y))\right](x^*,y^*) \\ \iff \\ 0 &= c_1(x^*,y^*)\phi(x^*,y^*) \\ &\quad - \left[\frac{\partial c_2(x,y)}{\partial x}\phi(x,y) + \frac{\partial \phi(x,y)}{\partial x}c_2(x,y)\right](x^*,y^*) \\ &\quad - \left[\frac{\partial c_3(x,y)}{\partial y}\phi(x,y) + \frac{\partial \phi(x,y)}{\partial y}c_3(x,y)\right](x^*,y^*) \\ \iff \\ 0 &= (\pi_1(\alpha_1 - 2x - y + 1)\lambda + \pi_2(\alpha_2 - x - 2y + 1)\lambda + \gamma_1 + \gamma_2)\phi(x^*,y^*) \\ &\quad -(\gamma_1 + \pi_1\lambda(\alpha_1 - x^*) + \pi_1\lambda(1 - x^* - y^*))\phi(x^*,y^*) \\ &\quad -(\gamma_1 x^* - \lambda(1 - x^* - y^*)\pi_1(\alpha_1 - x^*))\frac{\partial \phi(x,y)}{\partial x}(x^*,y^*) \\ &\quad -(\gamma_2 + \pi_2\lambda(\alpha_2 - y^*) + \pi_2\lambda(1 - x^* - y^*))\phi(x^*,y^*) \\ &\quad -(\gamma_2 y^* - \lambda(1 - x^* - y^*)\pi_2(\alpha_2 - y^*))\frac{\partial \phi(x,y)}{\partial y}(x^*,y^*) \\ \iff \\ 0 &= -(\gamma_1 x^* - \lambda(1 - x^* - y^*)\pi_1(\alpha_1 - x^*))\frac{\partial \phi(x,y)}{\partial x}(x^*,y^*) \\ &\quad -(\gamma_2 y^* - \lambda(1 - x^* - y^*)\pi_2(\alpha_2 - y^*))\frac{\partial \phi(x,y)}{\partial y}(x^*,y^*). \end{aligned} \tag{2.33}$$

Equation (2.33) holds for all test functions ϕ if and only if the coefficients of the derivatives of ϕ vanish. This yields

$$\begin{aligned} 0 &= (\gamma_1 x^* - \lambda(1 - x^* - y^*)\pi_1(\alpha_1 - x^*)) \\ 0 &= (\gamma_2 y^* - \lambda(1 - x^* - y^*)\pi_2(\alpha_2 - y^*)). \end{aligned} \tag{2.34}$$

Equation System (2.34) is equivalent to Equation System (2.15) from the deterministic limit. Since (x^*, y^*) solves Equation System (2.15) it solves Equation System (2.34), too. Thus, the delta distribution solves the PDE from Equation (2.19). □

Thus, Theorem 2.14 confirms the result which we received from the deterministic limit in Section 2.1.2, because $\delta_{(x^*,y^*)}(\phi)$ corresponds to the distribution received via Kurtz' Theorem (Theorem 2.6) where the entire mass is concentrated in (x^*, y^*).

Summarizing the results shows that the stationary distribution of the hybridization process converges to a distribution with the entire mass concentrated in (x^*, y^*) as the number of probes per spot tends to infinity. So, on the one hand, if we wait long enough and the number of probes is sufficiently large, the hybridization process will have a computable stationary distribution. But on the other hand, we still do not know what long enough and sufficiently large really means. For that reason we will investigate the convergence to the stationary distribution by simulating the hybridization process for different parameter situations.

2.1.4 Simulation results

Two different cases have been looked at, i.e. the number of different targets is $m = 2$ or $m = 4$.

The process has been simulated with the help of the *Gillespie algorithm* ([Gill]). Limited by computational power we were able to simulate the process 100 times with parameters as shown in Tables 2.1, 2.2 and 2.3. Already, this almost took 8 hours for the two target case and $2^1/_2$ days for the four targets case in *MATHEMATICA* on a Pentium III, 3.19 GHz, 3 GB RAM.

2.1.4.1 The ideal case

This case shall be looked at for two different parameter situations. On the one hand we are interested in the behavior of the process for equal binding and dissociation probabilities. This is an assumption, most biologists make when analyzing microarrays. On the other hand the behavior of the process for unequal binding and dissociation probabilities is examined in order to account for dye effects.

Equal probabilities Again, we will start with the case of equal hybridization and dissociation probabilities with parameter values from Table 2.1.

As could be seen in earlier investigations where the system has been solved directly for its stationary distribution and a closely related deterministic process has been derived, the scales of the probe and initial target numbers affect the variance of the results but not the ratio of hybridized target types to each other.

The ratio of hybridized targets should be close to 1/2 since there are initially twice as many targets of type 1 than of type 2. In Figure 2.5 the path of $R(t)$ is shown for the time interval $(0, \theta]$ and for a single simulation. Most of the variation in the figure appears during the first $2,500$ iterations. This is illustrated in Figure 2.4.

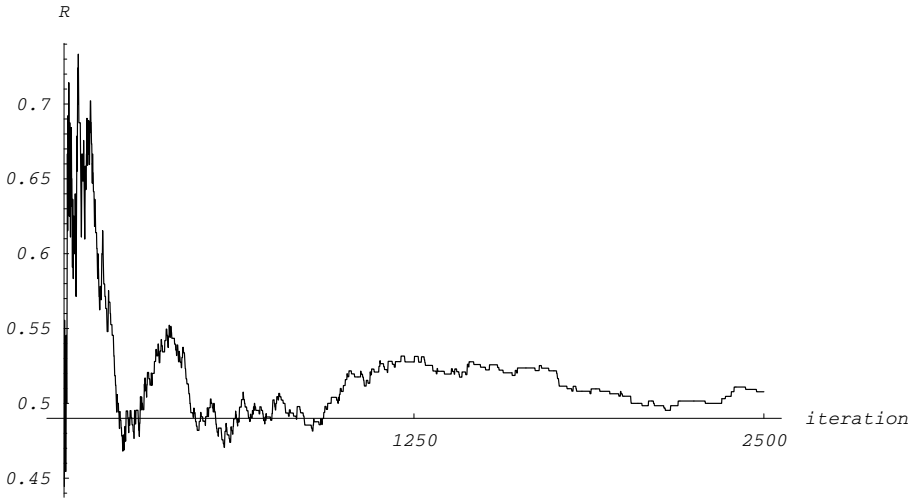


Figure 2.4: The ratio $R(t)$ of hybridized targets for a single simulation of the first $2,500$ iterations according to the ideal case with parameters as shown in Table 2.1

At this moment, approximately all probes are hybridized to a target for the first time. Afterwards, the ratio settles down at about $\frac{1}{2}$. It has mean $\hat{\mu}(R(t)) \approx 0.4892$ and variance $\hat{\sigma}^2(M) \approx 0.00014$. The corresponding graphs for the number of bound targets are shown in Figure 2.7.

All subsequently calculated values for single simulations (those, which depend explicitly on t) correspond to the simulation in Figures 2.5, 2.7(a) and 2.7(b). $N_1(t)$ has mean $\hat{\mu}(N_1(t)) \approx 162.1$ and variance $\hat{\sigma}^2(N_1(t)) \approx 133.9$, whereas $N_2(t)$ has mean $\hat{\mu}(N_2(t)) \approx 331.9$ and variance $\hat{\sigma}^2(N_2(t)) \approx 518.7$. As can be seen, the numbers of hybridized targets of type 1 and of type 2 rapidly reach their characteristic values of $\frac{1}{3}S$ and $\frac{2}{3}S$, respectively.

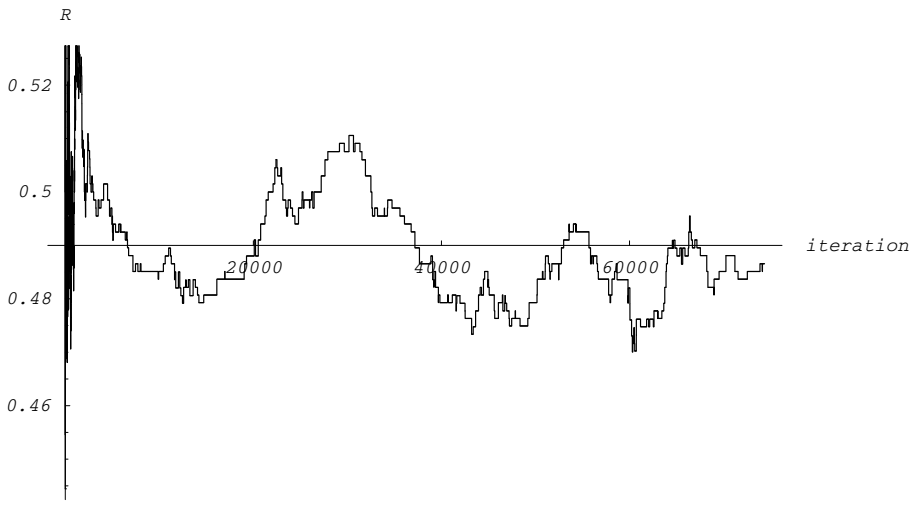


Figure 2.5: The ratio $R(t)$, $t \in (0, \theta)$ of hybridized targets for a single simulation of $74,376$ iterations according to the ideal case with parameters as shown in Table 2.1.

Afterwards, the process seems to have a random walk like structure with equal rates for both directions. Another realization of the process is shown in Figure 2.6. At the beginning large values of R are reached. Afterwards the process decreases slowly to $R = 1/2$. Various other possible paths have been observed. For this reason it is useful to investigate all simulations.

Looking at the result of all simulations in Figures 2.7(c) and 2.7(d) gives an idea of the variation within the process. Here the number of hybridized targets of both types $N_1(\theta)$, $N_2(\theta)$ (Figure 2.7(c)) and the ratio $R(\theta)$ (Figure 2.7(d)) at the end of the experiment θ are shown for each simulation. $N_1(\theta)$ has mean $\hat{\mu}(N_1(\theta)) \approx 164.1$ and variance $\hat{\sigma}^2(N_1(\theta)) \approx 117.7$, whereas $N_2(\theta)$ has mean $\hat{\mu}(N_2(\theta)) \approx 334.5$ and variance $\hat{\sigma}^2(N_2(\theta)) \approx 116.4$. $R(\theta)$ has mean $\hat{\mu}(R(\theta)) \approx 0.4921$ and variance $\hat{\sigma}^2(R(\theta)) \approx 0.00235$. Under the null hypothesis $\hat{\mu}(R(\theta)) = 1/2$, a Student's t-test does not lead to a significant deviation of $R(\theta)$ from its expected value $\frac{1}{2}$. Obviously, in terms of the initial fraction of target concentrations $T_1/T_2 = \frac{1}{2}$, $\hat{\mu}(N_1(\theta))$ and $\hat{\mu}(N_2(\theta))$ are close to the expected values of $\mathbb{E}(N_1(\theta)) = 166.\overline{6}$ and $\mathbb{E}(N_2(\theta)) = 333.\overline{3}$. This contributes to the idea that the fraction of hybridized target numbers is approximately equal to the fraction of initial target numbers, and thus serves as a good approximation of the actual ratio T_1/T_2 and the corresponding mRNA species.

The more interesting case is where the two dyes for labeling cause different

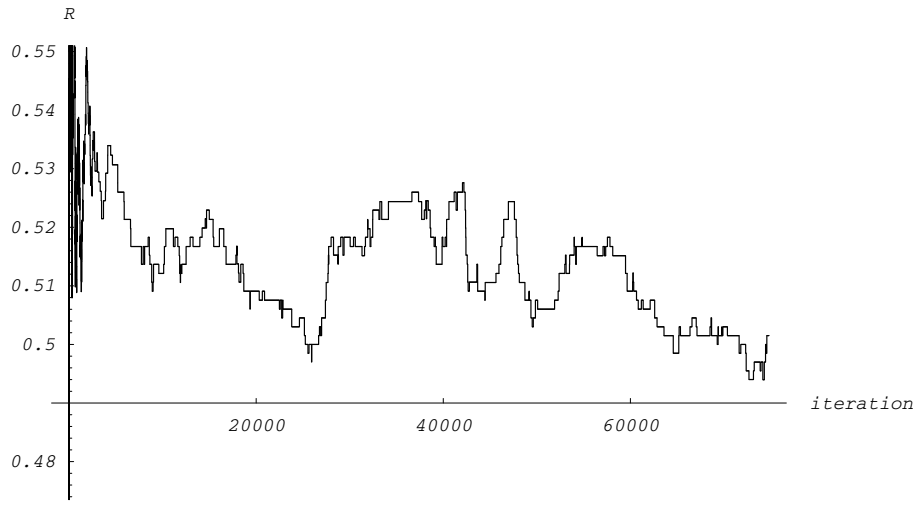


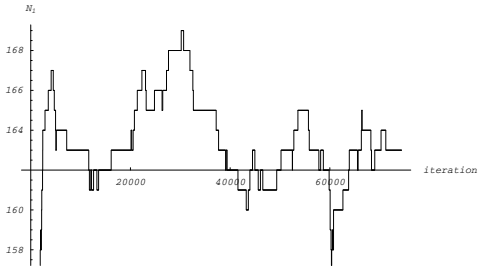
Figure 2.6: The ratio $R(t)$, $t \in (0, \theta)$ of hybridized targets for a single simulation of $74,819$ iterations according to the ideal case with parameters as shown in Table 2.2.

binding and dissociation probabilities. So far researchers assume the results of the hybridization process to show correct ratios, i.e. an x-fold in the ratio of hybridized targets is supposed to correspond to an x-fold in the associated number of the respective mRNAs. For that reason it shall be interesting to examine whether unequal binding and dissociation probabilities affect the ratio of hybridized targets or not. This case shall be looked at in the following paragraph.

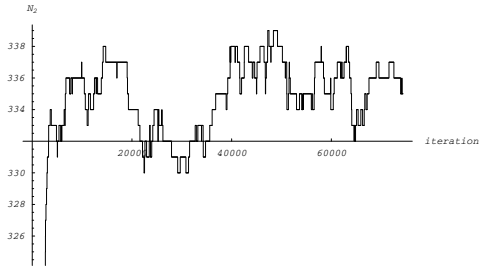
Unequal probabilities Assume the parameters from Table 2.2. The binding probability of type 1 is chosen to be less than of type 2, whereas the dissociation probability of type 1 is greater than of type 2. As a result one would expect less hybridized targets of type 1 for this case than for equal probabilities. This should result in a ratio $R(\theta)$ less than $1/2$.

Looking at the result of the simulation (Figure 2.8) one can see, that indeed $R(t) < 1/2$ for all $t \in (0, \theta)$ for a single simulation. This is due to the number of hybridized targets of type 1 and of type 2. As can be seen in Figure 2.9(a), the number of hybridized targets of type 1, $N_1(t)$, is always less than 149 whereas the number of hybridized targets of type 2, $N_2(t)$, is almost always greater than 348 (see Figure 2.9(b)). The resulting ratio $R(t)$ has mean $\hat{\mu}(R(t)) \approx 0.3865$ and variance $\hat{\sigma}^2(R(\theta)) \approx 0.0004$.

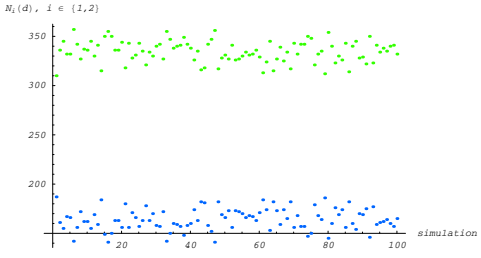
Looking at all 100 simulations corroborates this result. Here we find a



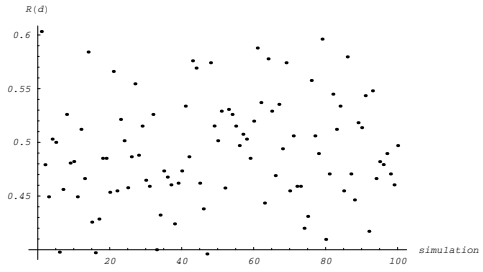
(a) The process $N_1(t)$, $t \in (0, \theta)$ of the number of hybridized targets of type 1 for a single simulation of 74,376 iterations.



(b) The process $N_2(t)$, $t \in (0, \theta)$ of the number of hybridized targets of type 2 for a single simulation of 74,376 iterations.



(c) Number of hybridized targets $N_1(\theta)$ (blue) and N_2 (green) for each of the 100 simulations.



(d) The ratio $R(\theta)$ at the end of the process θ for each simulation.

Figure 2.7: Number of hybridized targets for a single simulation and for 100 simulations with parameters as shown in Table 2.1.

mean $\hat{\mu}(R(\theta)) \approx 0.4095$ and variance $\hat{\sigma}^2(R(\theta)) \approx 0.0014$. For illustration see Figures 2.9(c) and 2.9(d). This is a significant deviation from the null hypothesis $\hat{\mu}(R(\theta)) = 1/2$. $N_1(\theta)$ has mean $\hat{\mu}(N_1(\theta)) \approx 143.7$ and variance $\hat{\sigma}^2(N_1(\theta)) \approx 93.9$. $N_2(\theta)$ has mean $\hat{\mu}(N_2(\theta)) \approx 354.9$ and variance $\hat{\sigma}^2(N_2(\theta)) \approx 90.2$.

In this case, if an analyzing method was used, which does not account for different binding and hybridization probabilities, it would underestimate the actual ratio of targets by almost 20%.

Assume targets of type 1 are generated under normal environmental conditions and targets of type 2 are generated under stress. Without accounting for different probabilities, we would infer, that under stress the respective mRNA was produced approximately 2.5 times as often as without stress instead of 2 times as its actual relation is. For a summary of the results see Table 2.8.

Also noticeable in this context is the way the ratio R behaves. As can

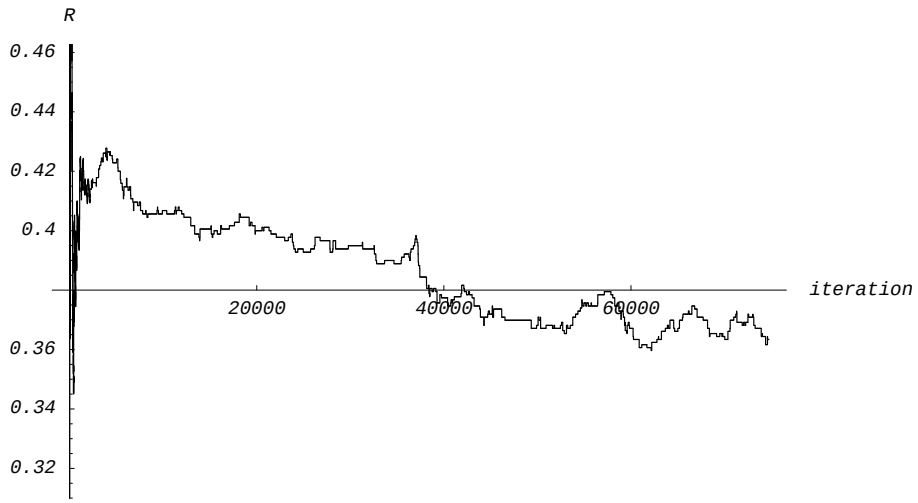
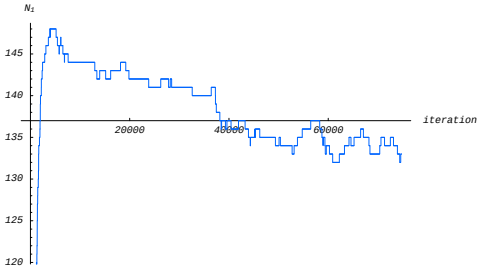


Figure 2.8: The ratio $R(t)$, $t \in (0, \theta)$ of hybridized targets for a single simulation of 74,604 iterations according to the ideal case with parameters as shown in Table 2.1 together with Table 2.2.

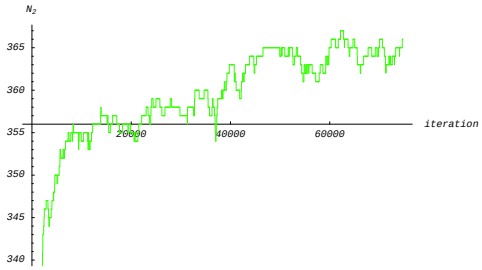
be seen in Figure 2.8, the ratio inclines rapidly. Its actual maximum is $^8/_{11} \approx 0.73$ which corresponds to the state where 8 targets of the first type and 11 targets of the second type are hybridized. This is characteristic for all simulations observed. It is due to the fact, that as long as the hybridization of targets is not restricted by the number of free probes, almost no competition between the target types takes place. Once all probes are covered for the first time, we observe that the ratio is closer to the actual ratio of $^1/_2$ than almost ever after. Then, the dissociation comes into play and the ratio decreases further. As can be seen in Figure 2.8 the process of decrease does not seem to be finished yet. This behavior can be observed in almost all simulations, i.e. the process has not reached its stationary distribution yet. This observation will be corroborated if we compare the stationary distribution with the histogram of the simulation. See Figures 2.10 and 2.11.

For example, looking at the case of unequal probabilities, we find the peak of the histogram at approximately $(N_1(\theta), N_2(\theta)) = (144, 355)$ whereas the peak of the stationary distribution can be seen at about $(N_1(\theta), N_2(\theta)) = (131, 367)$. So, the process has not yet reached its stationary distribution. This phenomenon has also been observed in Figures 2.8, 2.9(a) and 2.9(b).

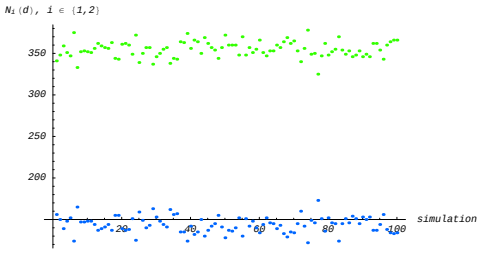
Thus, the time point of stopping the hybridization reaction is important to the experimenter for the situation of unequal probabilities. If it is chosen too early, the process might be far away from the stationary distribution



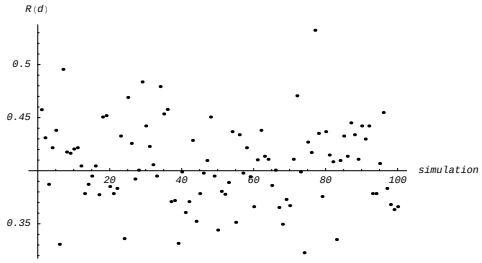
(a) The process $N_1(t)$, $t \in (0, \theta)$ of the number of hybridized targets of type 1 for a single simulation 74, 604 iterations.



(b) The process $N_2(t)$, $t \in (0, \theta)$ of the number of hybridized targets of type 2 for a single simulation 74, 604 iterations.



(c) Number of hybridized targets of type 1 (blue) and of type 2 (green) for each of the 100 simulations.



(d) The ratio $R(\theta)$ at the end of the process θ for each simulation.

Figure 2.9: Number of hybridized targets for a single simulation and for 100 simulations with parameters as shown in Table 2.1 and unequal probabilities as shown in Table 2.2.

and therefore we cannot infer the initial amounts of targets. As already seen during the examination of the eigenvalues of the process' generator Q, this effect will cancel out if waiting long enough. Quantifying this time span is hard due to the unknown size of hybridization and dissociation rates. Once these parameters could be determined we would be able to give advise for the duration time of the hybridization experiment in the case of two target types.

In the next paragraph we will give a short introduction to the case of four targets by looking at its simulation.

2.1.4.2 Investigation in presence of cross-hybridization

As already mentioned, non specific targets might hybridize and thus contaminate the spot signal. Therefore, it is interesting to include cross-hybridization into the model of the previous paragraph. A simple way is to define two

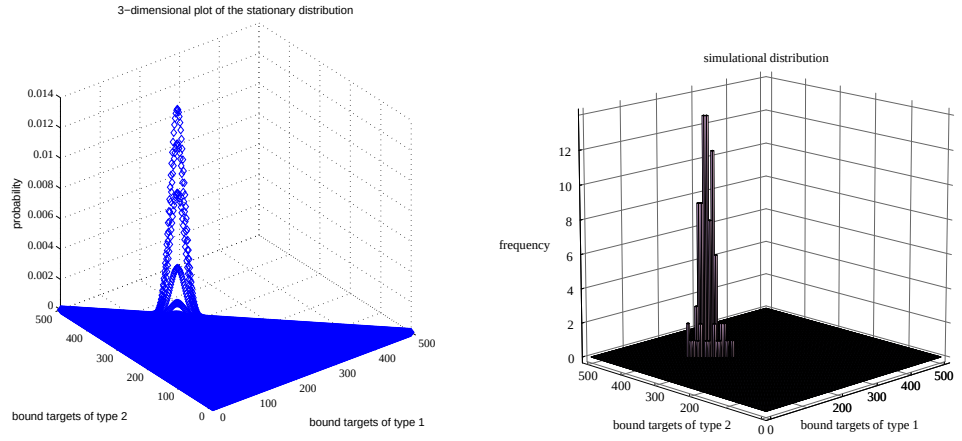
	a single simulation		100 **simulations**	
	equal prob.	unequal prob.	equal prob.	unequal prob.
$\hat{\mu}(N_1)$	162.1	137.5	164.1	143.7
$\hat{\sigma}^2(N_1)$	133.9	101.2	117.7	93.9
$\hat{\mu}(N_2)$	331.9	356.4	334.5	354.9
$\hat{\sigma}^2(N_2)$	518.7	652.9	116.4	90.2
$\hat{\mu}(R)$	.4892	.3865	.4921	.4060
$\hat{\sigma}^2(R)$	1.4×10^{-4}	3.5×10^{-4}	2.3×10^{-3}	1.5×10^{-3}
$\hat{\mu}(R_{log})$	$-.7210$	$-.9515$	-0.7138	$-.8792$
$\hat{\sigma}^2(R_{log})$	9.8×10^{-4}	2.1×10^{-3}	9.7×10^{-3}	9.7×10^{-3}

Table 2.8: Summary of the simulation results for a single simulation and 100 simulations of the ideal case. The parameters are according to Table 2.1 for the case of equal probabilities and Table 2.2 for unequal probabilities. The values for a single simulation are calculated over all $t \in (0, \Theta]$ whereas the values for 100 simulations are calculated over the result of all simulations at time Θ.

pairs of targets. Each pair represents a certain mRNA type, i.e. specific and unspecific targets. The members of each pair are labeled with two different fluorescence dyes according to their respective environmental condition. This model has also been simulated in MATHEMATICA 100 times. The parameters of this simulation can be seen in Table 2.3. The first pair is chosen to be the specific pair and the second pair to be the non-specific pair which might cause cross-hybridization events. This is modeled by greater binding probabilities and smaller dissociation probabilities for the specific pair in comparison to the non specific pair. As can be seen in Table 2.3 the probabilities within each pair are also chosen to be slightly different, which shall model the dye effect on hybridization and dissociation. Starting off with $50,000$ targets of each type, it is going to be interesting to see how the different probabilities will affect the number of hybridized targets. The results of the simulation can be seen in Figure 2.12.

Looking at the number of hybridized targets for a single simulation in Figure 2.12(b) shows that greater hybridization probabilities cause larger numbers of hybridized targets of the respective type. Thus, even starting off with equal initial numbers of free targets results in a considerable difference in hybridized targets.

In contrast, greater dissociation probabilities seem to have a rather long term effect, i.e. the respective targets seem to vanish from the spot. Thus it seems to be important at which time the hybridization process is stopped by



(a) stationary distribution (b) histogram

Figure 2.10: The numerical stationary distribution and the histogram derived from 100 simulations of the hybridization process with parameters as shown in Table 2.1.

the biologist for scanning the signals. Stopping at the middle of the process leads totally different ratios compared to stopping at the end. This can be seen in Figure 2.12(a). Here, R is calculated as follows

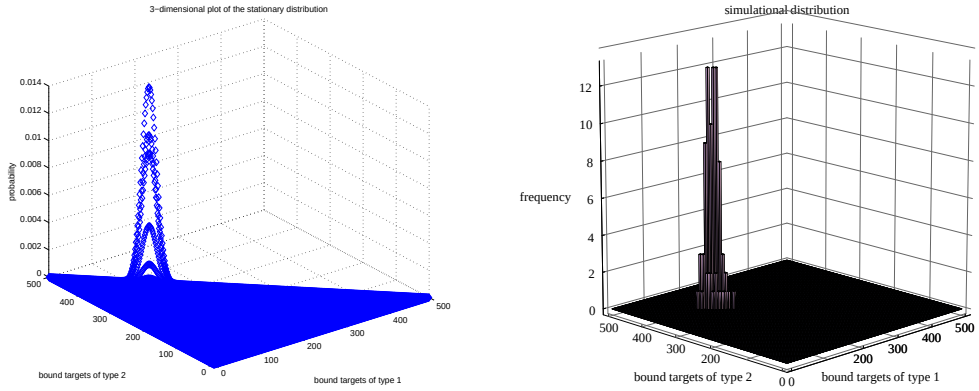
$$R(t) = \frac{N_1(t) + N_3(t)}{N_2(t) + N_4(t)},$$

since $N_1(t), N_3(t)$ are scanned in the first and $N_2(t)$, $N_4(t)$ in the second channel during detection.

The respective stationary distribution and limit according to Kurtz' Theorem (Theorem 2.6) have not been determined. This would be even more difficult than for the case of two target types. Further analysis of the hybridization model is suggested. At this point we will restrict the analysis to the simulation, which at least gives an idea of the behavior of the four targets case.

2.1.5 Résumé

Simulating the hybridization process as well as calculating its stationary distribution are computational expensive. Determining the limit from Kurtz is very fast and approximates the peak of the stationary distribution quite well. By the analysis of the eigenvalues of the Markov generator Q, we were able to show, that the process converges sufficiently fast towards Kurtz'



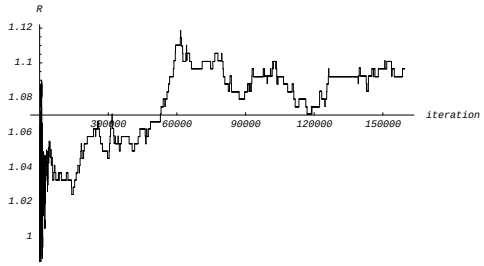
(a) stationary distribution (b) histogram

Figure 2.11: The numerical stationary distribution and the histogram for 100 simulations of the hybridization process with parameters as shown in Table 2.1 together with Table 2.2.

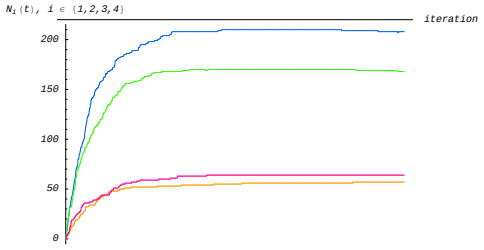
limit. For certain parameter situations, this observation was corroborated by comparing the values of the simulation, the determination of the stationary distribution and of Kurtz' limit. With the help of another limit we received a PDE. Its solution is a distribution which also corresponds to the limit from Kurtz.

Unfortunately, Kurtz' limit is a deterministic process and thus does not yield the variance of the hybridization process. But as already mentioned, we also looked at the behavior of the stationary distribution as the number of probes S increases. We observed a fast narrowing of the stationary distribution which means that the variance almost vanishes for large S. So, inferring the hybridized target numbers via Kurtz' limit is a useful method. Unfortunately, many parameters values are unknown. Thus, further experiments are recommended to determine the parameter values of the model.

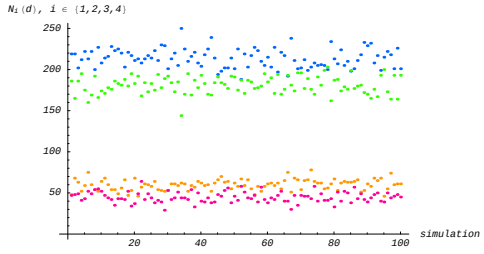
Result: In the parameter situation of unequal hybridization and dissociation probabilities a dye effect as well as a cross-hybridization effect (in the case of four targets) is visible. One is interested in the initial target ratio. Considering Limit (2.14), this ratio is α_1/α_2. From Equations (2.15) and (2.15) we can see that it depends on the ratios of hybridized targets in a nonlinear way. In contrast, most microarray statistical methods use linear models to describe the impact of the parameters on the intensity values (see [Speed] for more details or [Ochs] for an example). Thus, it is desirable to implement analyzing methods which account for these nonlinearities. We see



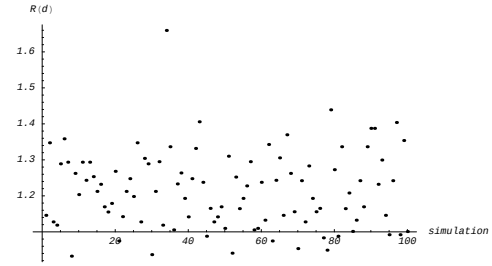
(a) The ratio $R(t)$, $t \in (0, \theta)$ of hybridized targets for a single simulation of 159, 598 iterations.



(b) The process of hybridized targets of types 1 (blue), 2 (green), 3 (orange) and 4 (red) for a single simulation of 159, 598 iterations.



(c) Number of hybridized targets of types 1 (blue), 2 (green), 3 (orange) and 4 (red) for each of the 100 simulations.



(d) The ratio $R(d)$ at the end of the process θ for each simulation.

Figure 2.12: Ratio and number of hybridized targets in presence of cross-hybridization for a single simulation and for 100 simulations with parameters as shown in Table 2.3.

no direct way how to accomplish this.

In addition, the stopping time of the experiment should be chosen sufficiently large in order to be able to infer the initial target relations via the stationary point of Kurtz or the stationary distribution ρ, $Q\rho = 0$.

In practice, the problem of dye effects is known quite well to researchers, who try to cancel it out by using various normalizing methods (see for example [Speed]). We have verified and quantified the dye effect in theory, which might help to improve the normalizing methods, once all parameters of the model are determined.

2.2 Residual subprocesses

2.2.1 Reverse transcription

In a first step we will analyze the original model from Section 1.2.1. In a second step this model is modified by a perturbational approach.

2.2.1.1 Analysis of the original model

Subsequently, the reverse transcription process shall be investigated with the help of the model from Section 1.2.1. Realistic parameters of the model were found in the literature (e.g. [ToHo] or [Sing]). The length of a target molecule ranges from 25 bases on oligonucleotide arrays to complete cDNAs (see [ToHo]). We will look at the range from 25 to 100 bases, since these values are most common. During reverse transcription a mixture of nucleotides is used to build the cDNA. It consists of unlabeled nucleotides of all four kinds and the additional labeled nucleotides of a single kind. We are only interested in those sequence positions of the target which are potentially able to hybridize to a labeled nucleotide. Thus, we roughly divide the probe length by four and add some error which yields a range for the sequence length of about 5 to 40. These values shall be investigated. Looking at typical protocols for reverse transcription reactions we find a ratio of 3 : 2 labeled to unlabeled nucleotides. For details see the protocol of a reverse transcription reaction from [Sing] in Section B in the appendix.

Thus, we have a distinct number of sequence positions on each target able to hybridize to labeled nucleotides. In addition there is a mixture of labeled and unlabeled nucleotides approaching the target.

We are interested in the probability distribution of the number of labeled nucleotides for a target of a certain length. With the help of the model from Section 1.2.1 this distribution shall be determined. We set up the global parameter situation of the model in Table 2.9.

target length	binding sites m	ratio labeled to unlabeled
$25 - 100$ bp	$5 - 40$	3 : 2

Table 2.9: Parameter situation of the reverse transcription process.

In order to use the model from Section 1.2.1 four more parameters are needed: the recruitment rates r_u and r_l as well as the total numbers of nucleotides V_u and V_l. In the literature no values for the recruitment rates could be found. However, we know that on the one hand the basic approach

of researchers working with microarrays is $r_u = r_l$ and that on the other hand the size of the recruitment rates only affects the timescale. For these reasons we chose rates which seemed to be reasonable in size. The exact values are mentioned whenever they come into play.

In addition to this parameter setting we will investigate the case of unequal recruitment rates since we assume the polymerase enzyme to be more likely to recruit an unlabeled nucleotide than a labeled one due to steric problems. The reason for these problems is that labeled nucleotides are larger than unlabeled as shown in Figure 1.4.

In contrast to the recruitment rates the total number of nucleotides is known quite well from the protocol [Sing]. Here we find 15 mmol labeled and 10 mmol unlabeled nucleotides. Multiplying these numbers with *Avogadro's constant* yields the total numbers of labeled and unlabeled nucleotides. According to [EKMPW] *Avogadro's constant* is

$$\eta_A = 6.0221367 \cdot 10^{23} mol^{-1}.$$

The resulting numbers of labeled and unlabeled nucleotides are
$V_l = 9.03320505 \cdot 10^{21}$ and $V_u = 6.0221367 \cdot 10^{21}$. Since V_l, $V_u \gg m$ we are able to restrict the investigation of the model to its binomial approximation in Equation (1.8).

Looking at the distribution of $Z(m)$, $m \in \{5, 6, ..., 40\}$ yields Figure 2.13.

This figure illustrates the dependency of the process on the number of binding sites m. As can be seen, the mean and the variance of the number of hybridized nucleotides $Z(m)$ shifts to larger numbers as m increases. This observation is corroborated by Figure 2.14.

Here, we can even see, that both, mean and variance linearly depend on m. This can also be seen directly from Equations (1.9) and (1.10) since the success probability q_l is independent from m.

In microarray experiments, those expression values are compared, which correspond to one gene and to one target length, respectively. Thus the dependency on the target length is not so important for the examination, since both signals come from probes with the same length. But because it has an effect on the variance it will be looked at nevertheless for further investigation in this work. Another reason is the occurrence of cross-hybridizations. The length of non specific targets is most likely to be different from the length of specific targets.

Subsequently, the dependency on the recruitment rates and thus on the steric structure shall be investigated. See Figures 2.15(a) and 2.15(b) to receive a first impression.

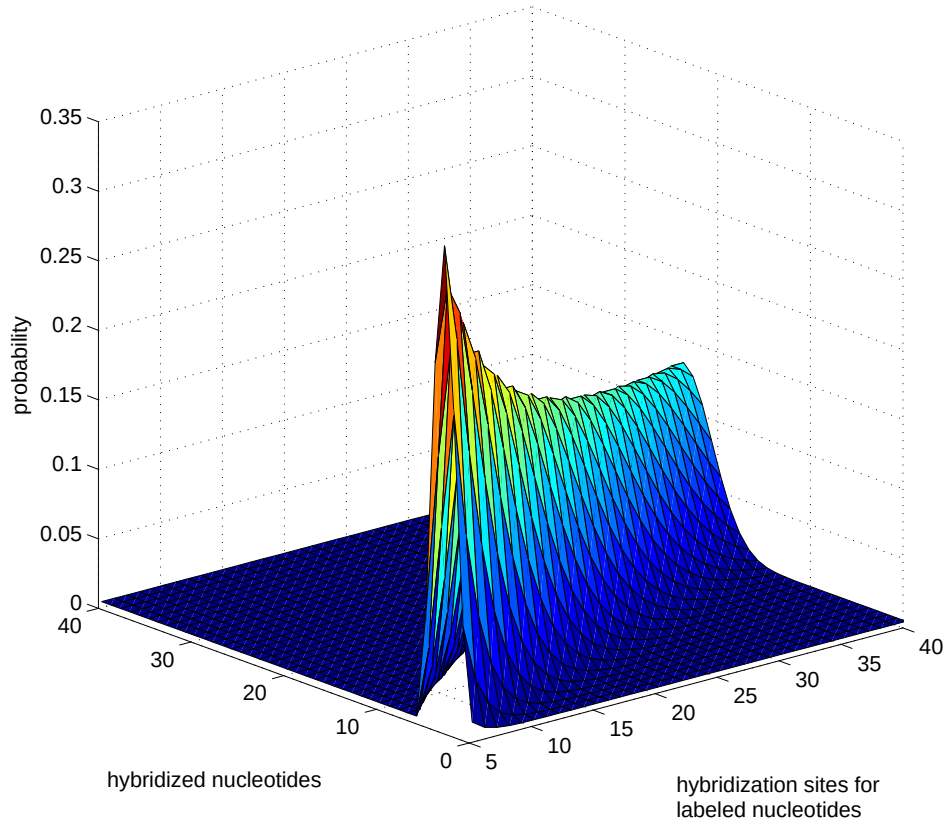


Figure 2.13: Probability distributions of hybridized labeled nucleotides for different numbers of potential binding sites.

For fixed numbers of hybridization sites of labeled nucleotides one observes, that the larger the difference $r_u - r_l$, the smaller the mean of hybridized nucleotides. This effect seems to become stronger as m increases. So, a rough tendency is that with an increasing hybridization rate for unlabeled nucleotides the actual number of hybridized labeled nucleotides decreases. On the other hand, in Figure 2.15(b) the variance almost seems to be independent from the difference $r_u - r_l$ and for fixed m. The exact dependencies shall be investigated next.

The mean seems to be constant on straight lines whereas the variance seems to have a more complex dependency. The exact dependency can be

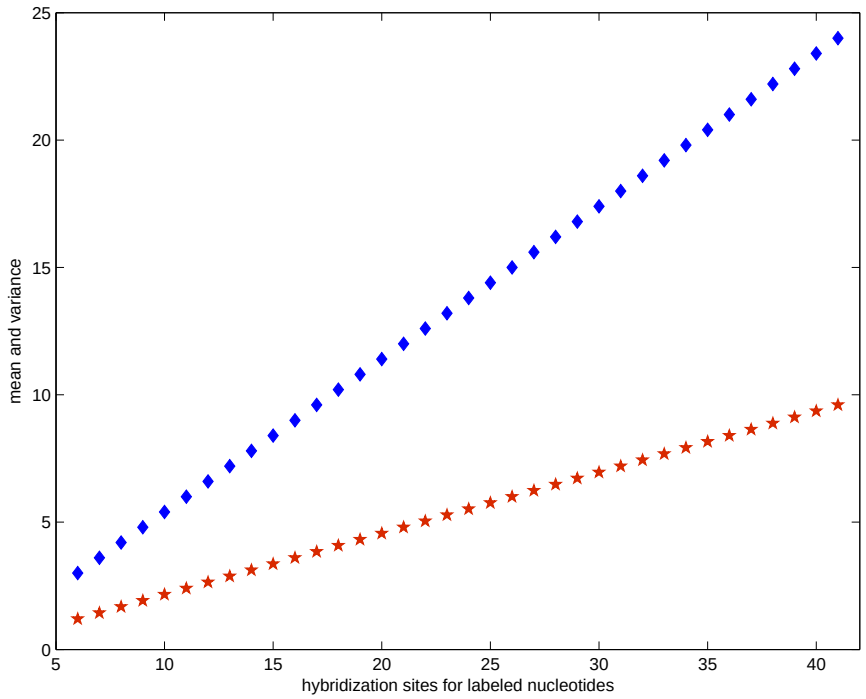


Figure 2.14: Mean (blue diamonds) and variance (red stars) of the number of hybridized nucleotides in dependency on the number of hybridization sites for labeled nucleotides m.

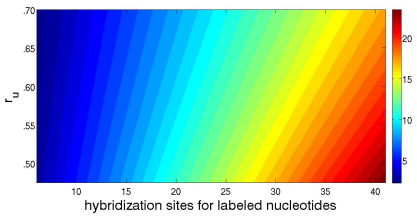
determined as follows. Using Equation (1.8) yields

$$\begin{aligned} \mathbb{E}(Z(m)) = \text{const}_1 &= mq_l = m\frac{r_l V_l}{r_l V_l + r_u V_u} \\ \iff r_u &= \frac{r_l V_l}{\text{const}_1 V_u} m - \frac{r_l V_l}{V_u} \end{aligned}$$

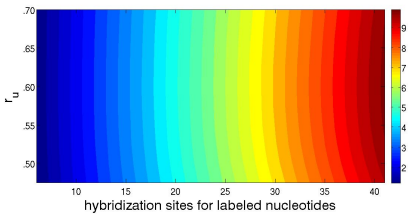
and

$$\begin{aligned} \sigma^2(Z(m)) = \text{const}_2 &= mq_l q_u = m\frac{r_l V_l}{r_l V_l + r_u V_u}\frac{r_u V_u}{r_l V_l + r_u V_u} \\ \iff r_u &= \frac{r_l V_l \left((m - 2\text{const}_2) \pm \sqrt{m\,(m - 4\text{const}_2)}\right)}{2\text{const}_2 V_u}. \end{aligned}$$

So, indeed, the mean is constant on straight lines whereas the variance is constant on branches of root functions. The dependency on the recruitment rates shall be investigated further.



(a) Heatmap of the mean $\mathbb{E}(Z(m))$.



(b) Heatmap of the variance $\sigma^2(Z(m))$.

Figure 2.15: Heatmaps for mean and variance of $Z(m)$ for $r_l = .5$, $r_u \in \{.50, .55, ..., .7\}$ and $m \in \{5, 6, ..., 40\}$.

2.2.1.2 A perturbation approach for the recruitment rates

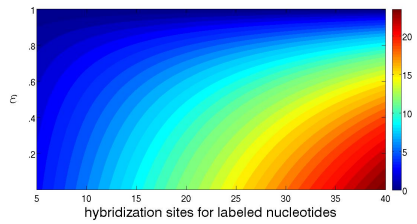
As already mentioned, labeled nucleotides are larger molecules and thus are recruited slower than unlabeled nucleotides. The difference in the recruitment rate is supposed to be small. For this purpose let r_l be a perturbation of r_u such that

$$r_l = (1 - \varepsilon) r_u$$

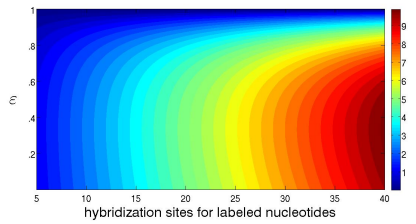
with $1 \gg \varepsilon > 0$. The resulting formulas for the mean and the variance are

$$\begin{aligned} \mathbb{E}_\varepsilon(Z(m)) &= m \frac{(1-\varepsilon) V_l}{(1-\varepsilon) V_l + V_u} := \mu(\varepsilon) \\ \sigma^2_\varepsilon(Z(m)) &= m \frac{(1-\varepsilon) V_l V_u}{((1-\varepsilon) V_l + V_u)^2} := \sigma^2(\varepsilon) \end{aligned} \tag{2.35}$$

Thus, both values only depend on the ratio $\frac{r_l}{r_u} = 1 - \varepsilon$. For an illustration of the dependency on the perturbation ε and the number of hybridization sites m see Figure 2.16.



(a) Heatmap of the mean $\mathbb{E}(Z(m))$.



(b) Heatmap of the variance $\sigma^2(Z(m))$.

Figure 2.16: Heatmaps for mean and variance of $Z(m)$ for $\varepsilon \in [0, 1]$ and $m \in \{5, 6, ..., 40\}$.

In these figures, the mean is constant at

$$\varepsilon = \frac{\text{const}_1 V_u}{(\text{const}_1 - m)V_l} + 1$$

and the variance at

$$\varepsilon = \frac{2\text{const}_2(V_l + V_u) - mV_u \pm V_u\sqrt{m(m - 4\text{const}_2)}}{2\text{const}_2 V_l},$$

where $const_1$ denotes the respective constant value of the mean and $const_2$ denotes the respective constant value of the variance.

2.2.1.3 Taylor expansion of μ and σ

To investigate the dependency on ε more precisely, we use the first-order Taylor expansion of μ and σ of Equation (2.35). This is a common tool to examine small perturbations.

μ and σ still depend on V_u, V_l and m. Hence, absolute changes of μ and σ cannot be interpreted without relating them to the starting point where $\varepsilon = 0$. Therefore, we will look at relative changes, i.e. $\frac{\mu(0+\varepsilon)}{\mu(0)}$ and $\frac{\sigma(0+\varepsilon)}{\mu(0)}$. Here, the division by $\mu(0)$ normalizes the values and thus causes independency from the scale. Mark, $\mu(0)$ and $\sigma^2(0)$ correspond to the case of equal hybridization rates $r_u = r_l$. Starting with absolute changes of μ yields:

$$\mu(\varepsilon) = \mu(0) + \mu'(0)\varepsilon + O(\varepsilon^2) \tag{2.36}$$

Terms of higher order can be neglected since the perturbation ε is assumed to be very small. In order to receive relative changes we divide Equation (2.36) by $\mu(0)$:

$$\frac{\mu(\varepsilon)}{\mu(0)} \approx 1 + \frac{\mu'(0)}{\mu(0)}\varepsilon.$$

Consequently, the relative change of μ is dominated by the coefficient of ε, i.e. $\frac{\mu'(0)}{\mu(0)}$. For this reason it shall be examined further. Using Equation (2.35) yields

$$\frac{\mu'(0)}{\mu(0)} = -\frac{V_u}{V_l + V_u} \in [-1, 0]\,.$$

So the relative change of the mean μ depends on the ratio of unlabeled nucleotides within all nucleotides. Its absolute value is minimal at $V_u = 0$, $V_l > 0$ and maximal either at $V_l = 0$, $V_u > 0$ or for the limit $V_u \to \infty$,

$V_l = \text{const}$. Based on the fact that in microarray experiments almost always $V_u, V_l > 0$, for small V_u the relative change of μ is almost independent from ε, i.e. changing ε does not have any effects on μ. But, if $V_u \gg V_l$, the dependency on ε will be strong. For this reason, it is recommended to keep the number of unlabeled nucleotides as small as possible. Looking at an example might help to understand the effect of ε on μ.

Example: Consider the parameter situation from Table 2.9. Here, $V_u = \frac{2}{3}V_l$. Thus,

$$\frac{\mu(\varepsilon)}{\mu(0)} \approx 1 - \frac{\frac{2}{3}V_l}{\frac{2}{3}V_l + V_l}\varepsilon = 1 - \frac{2}{5}\varepsilon.$$

So, if the perturbation ε of the recruitment rate was 10%, the mean of hybridized labeled nucleotides would decrease by approximately 4%.

In a next step we will investigate the dependency of the coefficient of variation $CV(0) := \frac{\sigma(0)}{\mu(0)}$ on ε. Again, we start with the Taylor expansion of the absolute change but this time with the change of the standard deviation:

$$\sigma(\varepsilon) = \sigma(0) + \sigma'(0)\varepsilon + O(\varepsilon^2)$$

Dividing by $\mu(0)$ and neglecting terms of higher order yields

$$\frac{\sigma(\varepsilon)}{\mu(0)} \approx \frac{\sigma(0)}{\mu(0)} + \frac{\sigma'(0)}{\mu(0)}\varepsilon. \tag{2.37}$$

Combining Equation (2.37) with Equation (2.35) yields

$$\begin{aligned}
\frac{\sigma(\varepsilon)}{\mu(0)} &\approx \frac{V_u}{\sqrt{mV_lV_u}} + \frac{V_u}{\sqrt{mV_lV_u}} \cdot \frac{V_l - V_u}{2(V_l + V_u)}\varepsilon \\
&\approx \frac{V_u}{\sqrt{mV_lV_u}}\left(1 + \frac{V_l - V_u}{2(V_l + V_u)}\varepsilon\right) \\
&\approx CV(0)\left(1 + \frac{V_l - V_u}{2(V_l + V_u)}\varepsilon\right).
\end{aligned}$$

Consequently, the coefficient of variation increases by $CV(0)\frac{V_l - V_u}{2(V_l+V_u)}\varepsilon$. $CV(0)$ and ε are considered to be constant. Thus, it is sufficient to investigate the behavior of $\frac{V_l - V_u}{V_l + V_u}$ to understand the dependency of CV on ε. $\frac{V_l - V_u}{V_l + V_u}$ is continuous, its global minimum is -1, its global maximum is 1 and it has its root at $V_u = V_l$. The minimum is reached at $V_u \neq 0$, $V_l = 0$ while the maximum is reached at $V_u = 0$, $V_l \neq 0$. Both cases are boring since they are due to a parameter situation with only one type of nucleotides. But there are other parameter situations where $\frac{V_l - V_u}{V_l + V_u}$ comes close to its extrema. On the one hand, if we consider the limit $V_l \to \infty$, $V_u = \text{const}$. the function will

approach its maximum and on the other hand if $V_u \to \infty$, $V_l = \text{const.}$ it will approach its minimum. So, if $V_l \gg V_u$ the coefficient of variation will increase rapidly in ε whereas it will decrease with the same velocity if $V_u \gg V_l$. It will not change at all, if $V_u = V_l$. Hence, in contrast to the results from the mean a high number of labeled nucleotides is disadvantageous as it increases the noise. Because of this contradiction it is necessary to look at the process from a different point of view.

2.2.1.4 Testing the impact of the perturbation

Subsequently, we will use a statistical test to examine the impact of a perturbation of the recruitment rates on the dye effect. This is a more sensible approach since the whole distribution comes into play, not only the mean and the variance.

Consider the following setting. Two cell cultures of the same organism have been exposed to different environmental conditions. The mRNA of both cultures has been extracted, separately reverse transcribed and also separately labeled. The result is two fluids, each containing the labeled targets according to one environmental condition but one labeled with color 1 and the other labeled with color 2. Thus, the reverse transcription process in the first fluid depends on a perturbation ε_1 and in the second fluid on a perturbation ε_2. We will look at a single target type. Due to our model, the number of labeled nucleotides within a single target is approximately binomially distributed. Generally, there is not only a single target of a specific type but thousands or even hundreds of thousands. All contribute to the signal during detection if being hybridized to the microarray. Let n_1, n_2 be the number of targets of this specific type in the respective fluid. Furthermore, let Σ_1 be the average number of labeled nucleotides per target in the first fluid and Σ_2 the respective average number in the second fluid. Expressed in a formula:

$$\Sigma_j \;=\; \frac{1}{n_j}\sum_{i=1}^{n_j} Z_i^{(j)}(m),\; j = 1,2 \tag{2.38}$$

where $Z_i^{(j)}(m)$ is the number of hybridized labeled nucleotides to target i of length m in fluid $j = 1, 2$.

According to the Theorem of de Moivre-Laplace, for i.i.d. $Z_i^{(j)}(m)$, the sum $\sum_{i=1}^{n_1} Z_i^{(1)}(m)$ is asymptotically normally distributed with mean $n_1\mu(\varepsilon_1)$ and variance $n_1\sigma^2(\varepsilon_1)$ whereas $\sum_{i=1}^{n_2} Z_i^{(2)}(m)$ is asymptotically normally distributed with mean $n_2\mu(\varepsilon_2)$ and variance $n_2\sigma^2(\varepsilon_2)$ (for details see [Kren],

Chapter 1 or [Grab]). Consequently, for large n_1 and n_2, Σ_1 as well as Σ_2 satisfy approximately

$$\Sigma_j \sim N\left(\mu\left(\varepsilon_j\right), \sqrt{\frac{1}{n_j}}\sigma\left(\varepsilon_j\right)\right), j = 1, 2$$

where $N(a, b)$ denotes the normal distribution with mean a and standard deviation b. Following, we assume these relations to hold exactly.

Let $D := \Sigma_1 - \Sigma_2$ be the difference between the average numbers of nucleotides per target. Obviously, D is a random variable which is normally distributed with mean $\mathbb{E}(D) = \mu(\varepsilon_1) - \mu(\varepsilon_2)$ and variance $\mathbb{V}ar(D) = \frac{1}{n_1}\sigma^2(\varepsilon_1) + \frac{1}{n_2}\sigma^2(\varepsilon_2)$.

At this point we have everything we need for a simple statistical test. We will test the null hypothesis

$$H_0 : \ \mathbb{E}(D) = \mathbb{E}(\Sigma_1) - \mathbb{E}(\Sigma_2) = \mu(\varepsilon_1) - \mu(\varepsilon_2) = 0.$$

In other words H_0 states that the perturbation and the difference in environmental conditions do not have a visible effect on the incorporation of labeled nucleotides.

We are interested in the power of this test in dependency on ε_1 and ε_2, i.e.

$$Power(\varepsilon_1, \varepsilon_2) := 1 - \beta,$$

with type II error β. The power of a test is the probability to make the correct decision of rejecting H_0 if indeed H_0 is false. In other words, our test has a good power whenever the difference in the mean numbers of incorporated labeled nucleotides is likely to be detected by the test. Thus, the effect of a perturbation is visible and influences the ratio of signal intensities.

Let Φ denote the probability distribution of the standardized normal distribution and $d^{(0)}_{\frac{\alpha}{2}}$ and $d^{(0)}_{1-\frac{\alpha}{2}}$ denote the $\alpha/2$ and $(1 - \alpha/2)$ quantiles of the probability distribution of D under H_0, i.e. $\mathbb{E}(D) = 0$. Depending on the level of significance $\alpha \in [0, 1]$ (type I error), the power can be determined as follows

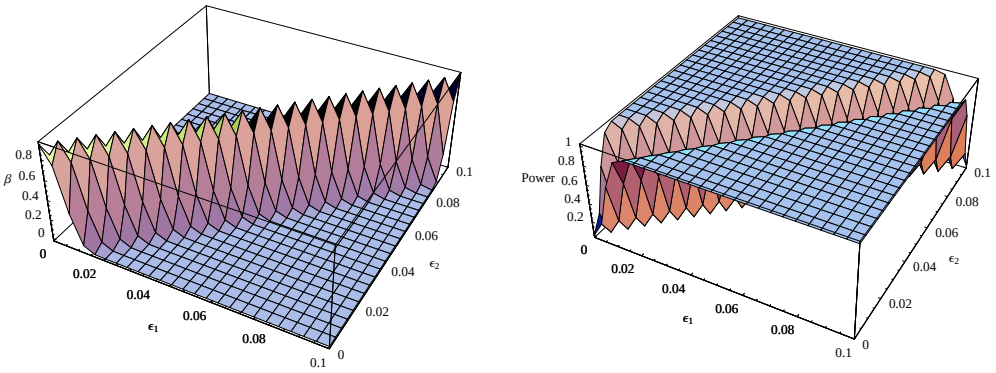
$$\begin{aligned} Power(\varepsilon_1, \varepsilon_2) &= 1 - \beta \\ &= 1 - \mathbb{P}\left(D \in \left[d^{(0)}_{\frac{\alpha}{2}}, d^{(0)}_{1-\frac{\alpha}{2}}\right]\right) \\ &= 1 - \mathbb{P}\left(\frac{D - \mathbb{E}(D)}{\sqrt{\mathbb{V}ar(D)}} \in \left[\frac{d^{(0)}_{\frac{\alpha}{2}} - \mathbb{E}(D)}{\sqrt{\mathbb{V}ar(D)}}, \frac{d^{(0)}_{1-\frac{\alpha}{2}} - \mathbb{E}(D)}{\sqrt{\mathbb{V}ar(D)}}\right]\right) \\ &= 1 - \Phi\left(\frac{d^{(0)}_{1-\frac{\alpha}{2}} - \mathbb{E}(D)}{\sqrt{\mathbb{V}ar(D)}}\right) + \Phi\left(\frac{d^{(0)}_{\frac{\alpha}{2}} - \mathbb{E}(D)}{\sqrt{\mathbb{V}ar(D)}}\right). \end{aligned} \tag{2.39}$$

Subsequently, let $\alpha = 5\%$.

For the parameter situation of Table 2.10 the type II error and the power are illustrated in Figure 2.17 for ε_1, $\varepsilon_2 \in [0, .1]$.

V_l	V_u	m	n_1	n_2
$9.03320505 \times 10^{21}$	6.0221367×10^{21}	20	$25,000$	$50,000$

Table 2.10: Parameter situation for Figure 2.17.



(a) Type II error β.

(b) Power $1 - \beta$.

Figure 2.17: Type II error and power in dependency on perturbations ε_1 and ε_2 for parameters as shown in Table 2.10.

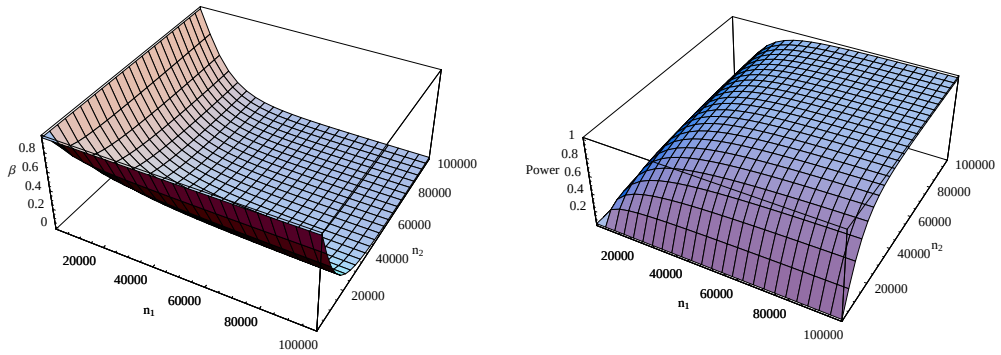
As can be seen, if the perturbations are equal, the power will be zero. This is reasonable since $\varepsilon_1 = \varepsilon_2$ implies equal dye incorporation rates in the two fluids and thus a test cannot see a difference in the labeling efficiency. But, leaving the situation of equality yields an increase in the power of the test. So, if the dyes are incorporated at different rates, the test finds this difference at a higher probability. This means, the perturbations ε_1 and ε_2 have a remarkable impact on the number of labeled nucleotides in the two fluids.

On the other hand, if we fix ε_1 and ε_2 we can have a look at the dependency on n_1, $n_2 \in [1; 100,000]$, i.e. the initial target numbers. The results for the parameter situation of Table 2.11 are shown in Figure 2.18.

Obviously, small target numbers of either type cause a low power of the test whereas large target numbers yield a high power. Thus, for realistic target numbers, the difference in dye incorporation between the fluids is visible in a sense that it has to be considered if inferring the initial target amounts.

V_l	V_u	m	ε_1	ε_2
$9.03320505 \times 10^{21}$	6.0221367×10^{21}	20	.04	.05

Table 2.11: Parameter situation for Figure 2.18.



(a) Type II error β. (b) Power $1 - \beta$.

Figure 2.18: Type II error and power in dependency on the numbers of targets within the fluids for parameters as shown in Table 2.11.

Analogously, we can examine the dependency on the numbers of initial labeled nucleotides in the fluids. See Table 2.12 for the parameter situation and Figure 2.19 for the results in dependency on $V_l, V_u \in [1, 10^{22}]$.

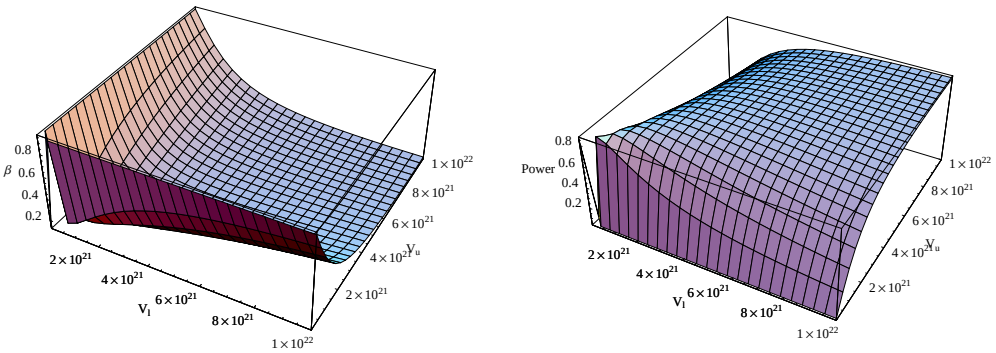
n_1	n_2	m	ε_1	ε_2
$25,000$	$50,000$	20	.04	.05

Table 2.12: Parameter situation for Figure 2.19.

The power is close to one as long as $V_u \approx V_l$ or for large V_u, V_l. Thus, for realistic parameter situations of large numbers of nucleotides there is also a visible dye effect. But, we can also see, that using only labeled nucleotides would lead to a power of zero. Thus, we suggest to use labeled nucleotides only.

Lastly, we will look at the dependency on the number of nucleotides within the specific target type. We will use the parameter situation from Table 2.13. The results are illustrated in Figure 2.20 for $m \in [1, 40]$.

The power of the test increases in m. So, the larger a specific nucleotide the stronger the dye effect in labeling.



(a) Type II error β. (b) Power $1-\beta$.

Figure 2.19: Type II error and power in dependency on the numbers of initial nucleotides within the fluids for parameters as shown in Table 2.12.

n_1	n_2	ε_1	ε_2	V_l	V_u
$25,000$	$50,000$	.04	.05	$9.03320505 \times 10^{21}$	6.0221367×10^{21}

Table 2.13: Parameter situation for Figure 2.20.

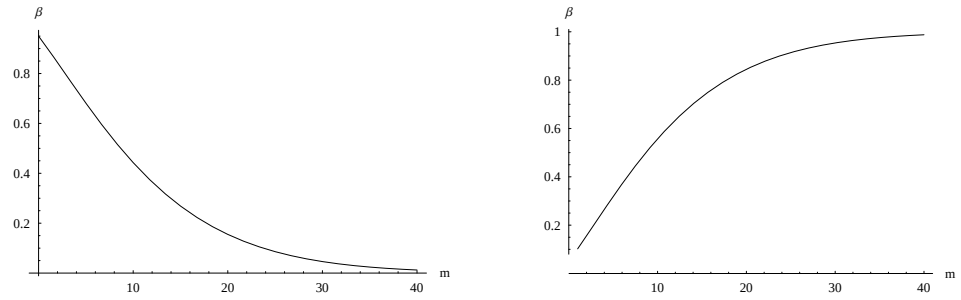
Concluding, we see, that if we have realistic parameter situations, the dye effect is always visible as soon as we have different perturbations associated to the different dyes within the two fluids. So, if one wants to infer the initial number of targets, the dye effect has to be accounted for. This aspect is going to be discussed in the résumé on page 94.

Another tool to rate a test is the ROC curve, which comes from signal detection theory. This topic is looked at in the next paragraph to give advice on improving the parameter choice.

2.2.1.5 Parameter choice due to the ROC curve

For details of the theory of this paragraph see [ZwCa] and [Faw]. The Receiver Operating Characteristic (ROC) curve is a true positive rate (TPR) vs. false positive rate (FPR) plot for the test looked at and for all possible levels of significance α. The true and false positive rates in our model are as follows:

$$\begin{aligned} TPR &= 1-\beta \\ FPR &= \alpha. \end{aligned} \tag{2.40}$$



(a) Type II error β. (b) Power $1 - \beta$.

Figure 2.20: Type II error and power in dependency on the number of nucleotides within the fluids for parameters as shown in Table 2.13.

E.g., for the parameter situation of Table 2.13 and $m = 20$ binding sites for labeled nucleotides within the target, the ROC curve is illustrated in Figure 2.21.

It can be interpreted as follows. The steeper its incline is to the value of 1, the better is the test in means of showing difference in color effects. So, a perfect distinction can be made with a test which has a ROC curve constant to 1. Or, in other words, the area under the ROC curve has to be large (close to 1) for good tests.

For example, the area under the ROC curve in Figure 2.21 is ≈ 0.965. I.e. for this parameter situation the dye effect is clearly visible.

Our aim is not a good test but rather a small color effect. So we are looking for parameter settings, which minimize the area under the ROC curve.

For this purpose, we will look at the situation scientists will be faced with, if performing the reverse transcription step. Once the labels have been chosen, the perturbations ε_1, ε_2 have to be considered constant for a given target. Also, the number of hybridization sites for labeled nucleotides for a certain target type cannot be influenced as well as the number of targets which originate from the two fluids. So, the interesting parameters for minimizing the area under the ROC curve are the numbers of labeled nucleotides and unlabeled nucleotides fed to the reverse transcription reaction. For several parameter situations the minimum and the values of V_l and V_u have been calculated with MATHEMATICA. The results are summarized in Table 2.14.

Having a minimum of .5 is equivalent to guessing the color of a target, i.e. the test cannot distinguish between the colors of the targets. For all four parameter situations of Table 2.14 we discovered a flat minimum,

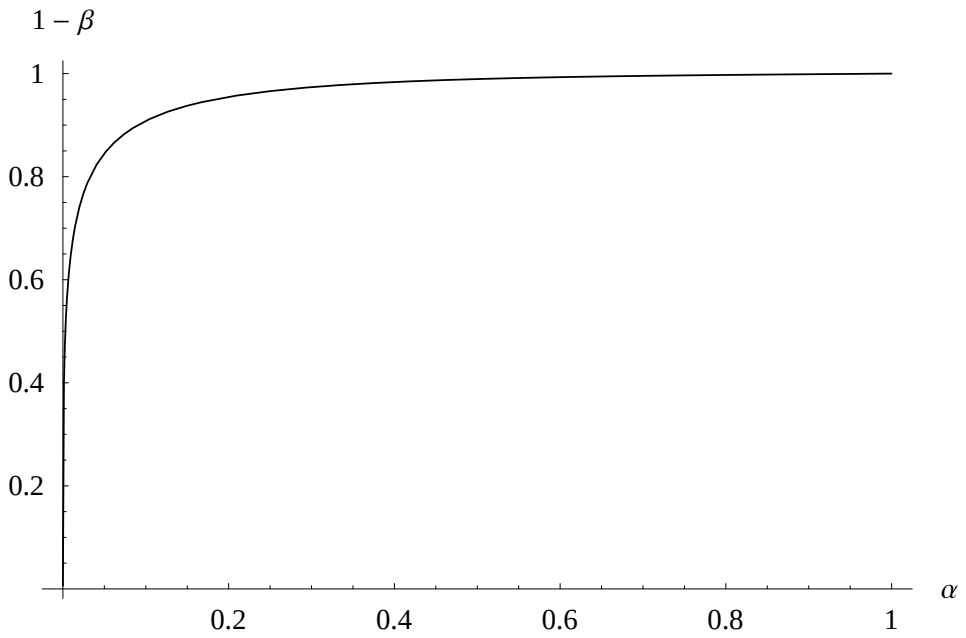


Figure 2.21: ROC curve for the test described in the previous section with parameter values from Table 2.13 and for $m = 20$.

i.e. the value of approximately .5 has been observed for $V_l \in [1, 10^8]$ while $V_u \in [10^{15}, 10^{22}]$. So, as long as there are much more unlabeled than labeled nucleotides, the test does not see any color effect. But, a much larger number of unlabeled than labeled nucleotides is not realistic since we need enough labeled nucleotides to detect the signal. This phenomenon is corroborated by the previous examination of the power of the test. Here, we could also see that very asymmetric numbers of labeled and unlabeled nucleotides led a bad power. Looking at Figure 2.22 which illustrates the area under the ROC curve in dependency on V_l and V_u yields the same conclusion.

The area under the ROC curve as well as the power are small as soon as $V_l \gg V_u$ or $V_u \gg V_l$.

Optimization for the worst parameter situation, i.e. there is a maximally visible color effect or area under the ROC curve, yields the values of Table 2.15.

As can be seen, the optimal parameter values are close to realistic parameters. This means, for realistic parameters, the color effect is visible quite well. So, as already mentioned, it has to be accounted for. A possible approach to estimate the initial numbers of targets including the color effect is discussed in the subsequent paragraph.

n_1	n_2	m	ε_1	ε_2	**Min**	$\mathbf{V_l}$	$\mathbf{V_u}$
$25,000$	$50,000$	20	.04	.05	**.5**	$\approx \mathbf{1}$	$\approx \mathbf{6 \times 10^{21}}$
$25,000$	$50,000$	20	.01	.02	**.5**	**1**	$\mathbf{1 \times 10^{22}}$
2.5×10^6	5×10^6	20	.04	.05	**.5**	$\approx \mathbf{9}$	$\mathbf{1 \times 10^{22}}$
2.5×10^6	5×10^6	20	.01	.02	**.5**	$\approx \mathbf{3}$	$\approx \mathbf{7 \times 10^{20}}$

Table 2.14: Parameter situations for the parameters n_1, n_2, m, ε_1 and ε_2 for determining the minimum **Min** of the area under the ROC curve. Additionally, the table contains **Min** itself and the position V_l, V_u where it is reached. The nucleotide numbers were restricted to V_l, $V_u \in \left[1, 1 \times 10^{22}\right]$.

Max	n_1	n_2	m	ε_1	ε_2	V_l	V_u
1	$\approx 903,327$	$\approx 34,731$	≈ 27	$\approx .097$	$\approx .006$	10^{22}	$\approx 2.6 \times 10^{21}$

Table 2.15: Parameter situation calculated with MATHEMATICA, where the maximum Max of the area under the ROC curve is reached for n_1, $n_2 \in \left[1, 10^6\right]$, ε_1, $\varepsilon_2 \in [0, .1]$, V_l, $V_u \in \left[1, 1 \times 10^{22}\right]$ and $m \in [1, 100]$.

2.2.1.6 A simple approach for estimating the initial target numbers

The usual case in microarray experiments is, that n_1, n_2 are unknown and one tries to estimate their ratio or log ratio. After motivation of the necessity to account for dye effects under the perturbation model, we will try to estimate the number of initial targets n_1, n_2 given the numbers of hybridized nucleotides S_1, S_2 which are realizations of Σ_1 and Σ_2, respectively. Using the method of moments (see [Grab]) results in the following equation

$$\mu(\varepsilon_j) \quad = \quad \frac{1}{n_j} \sum_{i=1}^{n_j} Z_i^{(j)}(m),\ j = 1, 2.$$

Since $\sum_{1=1}^{n_j} Z_i^{(j)}(m) = S_j$, $j = 1, 2$, rearranging yields the estimator

$$\hat{n}_j \quad \approx \quad \frac{S_j}{\mu(\varepsilon_j)},\ j = 1, 2,$$

which is according to the previous considerations of Σ_j, $j = 1, 2$ asymptotically normally distributed with

$$\hat{n}_j \quad \sim \quad N\left(n_j, \sqrt{\frac{n_j}{\mu\left(\varepsilon_j\right)^2}} \sigma\left(\varepsilon_j\right)\right),\ j = 1, 2.$$

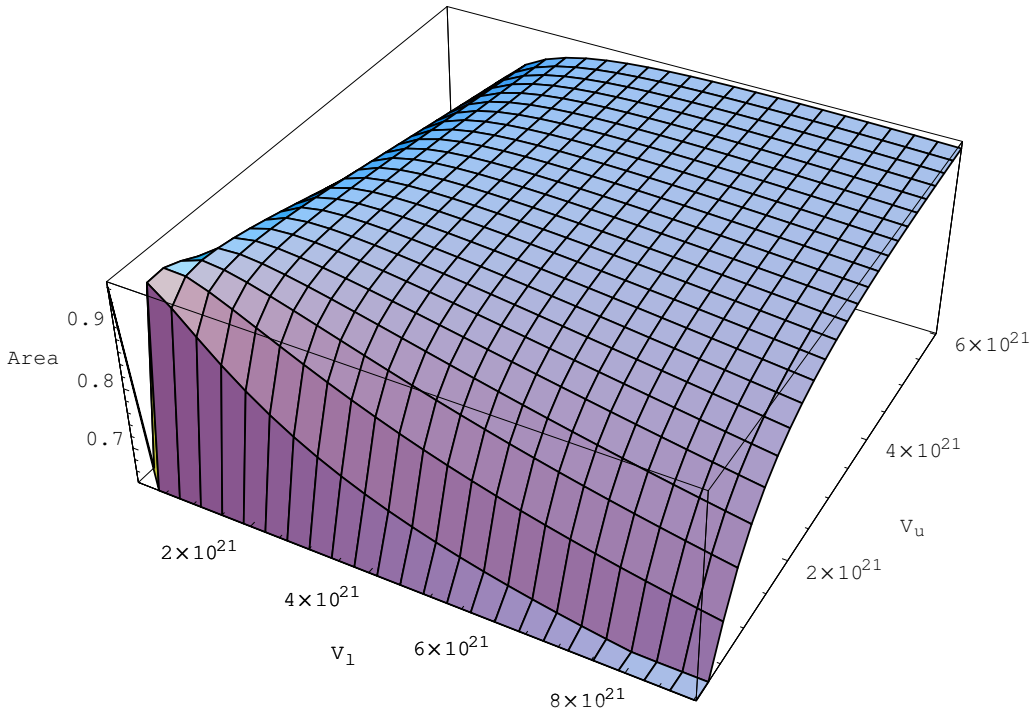


Figure 2.22: Area under the ROC curve for the test described in the previous section with parameter values from Table 2.12.

Thus, once the parameters of the model are known it will be easy to determine the initial number of targets and thus the ratio of the amount of mRNA corresponding to the two different environmental conditions. Note however, the results of the estimation are error prone since the estimators are random variables. So, confidence intervals or similar statistics have to be looked at to draw reliable conclusions.

2.2.1.7 Résumé

In this section we have seen that a color effect is visible for realistic parameter situations in the reverse transcription model from Section 2.2.1. This has been verified by looking at the mean and the variance of the number of hybridized labeled nucleotides for the original model as well as for the model of perturbed recruitment rates. This observation has been corroborated by looking at the Taylor expansion for perturbation ε and by developing a simple

test to analyze the impact of the perturbation on the dye effect. In addition, by determining the minimum of the area under the ROC curve for this test, we were able to recommend a parameter situation for the numbers of initial nucleotides which would yield an almost invisible color effect. Unfortunately, this parameter situation is not realistic. But the investigation of the power led to another suggestion. From Figure 2.19 it is theoretically advisable to use labeled nucleotides only. But practically, the labeling of nucleotides is not perfect. So, we suggest to keep the amount of unlabeled nucleotides as small as possible.

Summarizing the results of this section, we have to say, that for realistic parameter situations, the color effect is visible in a sense of influencing the number of fluorescence molecules attached to the targets. This effect has to be incorporated into the inference of initial target concentrations. State of the art normalizing methods try to overcome this problem by establishing the same mean or median of all the intensities or of a selection of intensities due to one dye compared to the other dye. The selection of genes either corresponds to housekeeping genes (considered to be equally expressed during the different cell states, see [Speed]) or to genes which are artificially added to the microarray (spike-in method, see [Rydén]). Note, using only a selection of genes implies a small bias but a large variance compared to the situation of using all genes.

In contrast, we solved this problem by using the knowledge about the reverse transcription model. For this purpose we proposed a simple approach to estimate the initial target numbers which might be used after determining the parameter values of the model.

2.2.2 Washing

In the following section we will analyze the washing model which has been introduced in Section 1.2.2.

First of all, let us recall some notations from Section 1.2.2 which we will need in this section, too. Let N_i and H_i be the total target numbers of type i on the spot before and after washing and let $W_i = N_i - H_i$ be the number of targets which were washed off. In addition we will use $\lambda_i(c)$ and p_{k_i}. The former is the rate for the event of detergent molecules binding to a target and the latter describes the probability of solving a target of type i. Here, c denotes the detergent concentration and k_i the number of detergent molecules needed to dissolve a target of type i.

We begin with deriving the mean and the variance of H_i.

Fixation of N_i yields

$$\begin{aligned}\mathbb{E}(H_i \mid N_i) &= N_i - \mathbb{E}(W_i(t) \mid N_i) \\ &= N_i - N_i \cdot p_{k_i} \\ &= (1 - p_{k_i})N_i \end{aligned} \tag{2.41}$$

and thus

$$\mathbb{E}(H_i) = (1 - p_{k_i})\mathbb{E}(N_i). \tag{2.42}$$

Using the computational formula for the variance yields

$$\begin{aligned}\mathbb{E}(H_i^2 \mid N_i) &= \mathbb{E}(H_i \mid N_i)^2 + \mathbb{V}ar(H_i \mid N_i) \\ &= (1 - p_{k_i})^2 N_i^2 + \mathbb{V}ar(N_i - W_i(t) \mid N_i) \\ &= (1 - p_{k_i})^2 N_i^2 + \mathbb{V}ar(W_i(t) \mid N_i) \\ &= (1 - p_{k_i})^2 N_i^2 + N_i p_{k_i}(1 - p_{k_i}) \end{aligned}$$

and therefore

$$\mathbb{E}(H_i^2) = (1 - p_{k_i})^2 \mathbb{E}(N_i^2) + \mathbb{E}(N_i) p_{k_i}(1 - p_{k_i}).$$

Again, with the help of the computational formula for the variance we receive

$$\begin{aligned}\mathbb{V}ar(H_i) &= \mathbb{E}(H_i^2) - \mathbb{E}(H_i)^2 \\ &= (1 - p_{k_i})^2 \mathbb{E}(N_i^2) + \mathbb{E}(N_i) p_{k_i}(1 - p_{k_i}) - (1 - p_{k_i})^2 \mathbb{E}(N_i)^2 \\ &= (1 - p_{k_i})^2 \mathbb{V}ar(N_i) + p_{k_i}(1 - p_{k_i})\mathbb{E}(N_i). \end{aligned} \tag{2.43}$$

The mean and the variance of H_i shall be investigated for different parameter situations. Both depend on the output of the hybridization process N_i and through p_{k_i} on the intensity $\lambda_i(c)t$.

First of all we will discuss the nature of $\lambda_i(c)$. The intensity should be increasing with the concentration of detergent molecules. On the one hand it should not be smaller than zero and on the other hand a maximal intensity $\lambda_i^{\max}(c)$ should not be exceeded. Many function satisfy these criteria, especially cumulative distribution functions. We picked the cumulative distribution function of the Maxwell-Boltzmann distribution (see [Papou]) because of its close relation to the movement of particles and multiplied it by $\lambda_i^{\max}(c)$ in order to rescale its values to the interval $[0, \lambda_i^{\max}(c)]$. The result is the following intensity function:

$$\lambda_i(c) = \lambda_i^{\max}(c) \left[2\Phi\left(\frac{c}{a_i}\right) - \sqrt{\frac{2}{\pi}} \frac{c e^{-c^2/(2a_i^2)}}{a_i} \right]$$

with standard normal distribution function

$$\Phi(z) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du$$

and parameter a_i which is useful for calibrating the shape of the intensity function. A larger a_i implies a slower growing $\lambda_i(c)$. See Figure 2.23 for an illustration.

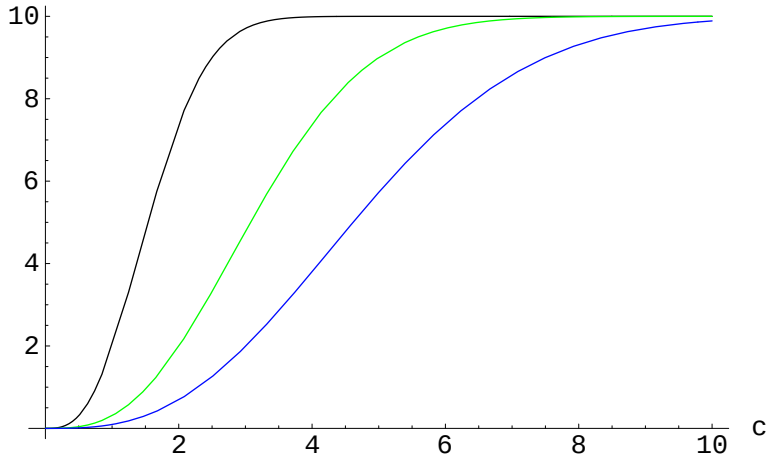


Figure 2.23: $\lambda_i(c)$ for $\lambda_i^{\max}(c) = 10$ and $a_i = 1$ (black), $a_i = 2$ (green) and $a_i = 3$ (blue).

Next, we will have a look at the washing model for different parameter situations following those from the hybridization model. We will start with the setting from Table 2.16, i.e. we have a single cDNA type, labeled with two different colors according to two different cell states. The values of $\lambda_i^{\max}(c)$ and k_i, $i = 1, 2$ are chosen arbitrarily since no values could be found in the literature. Both only affect the time scale. The only difference between the two target types is the attached label. Therefore, they are supposed to behave similarly. For that reason both parameters are chosen to be equal. We will look at two different cases. The first case, where we have no color effect, i.e. $a_1 = a_2 = 1$ and the second with an effect, i.e. $a_1 = 1$, $a_2 = .99$.

Obviously, $\mathbb{E}(N_1)$ and $\mathbb{E}(N_2)$ as well as $\mathbb{V}ar(N_1)$ and $\mathbb{V}ar(N_2)$ are chosen according to the solution of $Q\rho = 0$ with Markov generator Q and stationary distribution ρ for the equal probabilities case from Table 2.5 of the hybridization process.

Figures 2.24 and 2.25 show the mean and the variance of the number of hybridized targets H_i, $i = 1, 2$ in dependency on the concentration of detergent molecules.

maximal detergent intensities	
$\lambda_1^{\max}(c)$	10
$\lambda_2^{\max}(c)$	10
detergent concentration	
c	$\in [0, .22]$
duration of the washing procedure	
t	900
mean number of initially hybridized targets	
$\mathbb{E}(N_1)$	166.3
$\mathbb{E}(N_2)$	332.5
variance of the number of initially hybridized targets	
$\mathbb{V}ar(N_1)$	110.2
$\mathbb{V}ar(N_2)$	110.6
detergent molecules needed for solution	
k_1	10
k_2	10

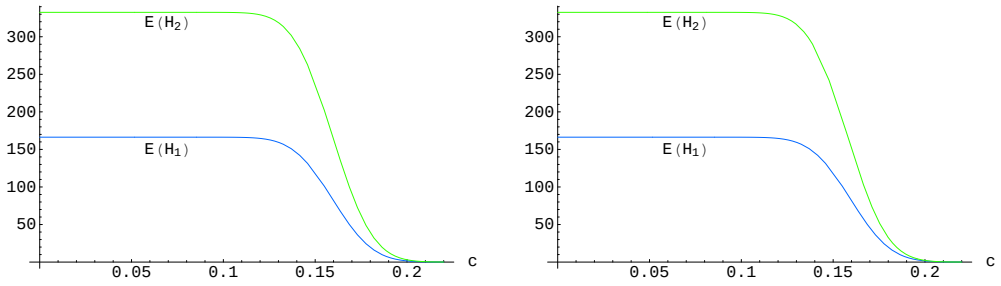
Table 2.16: Parameter situation for the case of one cDNA type labeled with two different colors.

As can be seen all curves are constant at the initial level until reaching a characteristic value of about $c = 0.12$. Here, the targets start to dissociate from the spot and the mean of the number of hybridized targets approaches zero. The variance shows a slightly different behavior. On the one hand $\mathbb{V}ar(H_1)$ also decreases similarly after reaching the characteristic value, but on the other hand, $\mathbb{V}ar(H_2)$ increases approximately until $c = 0.15$ before decreasing to zero as well.

So, besides a small range of concentrations, the variance decreases with the mean, which is good since their ratio (coefficient of variation) characterizes the randomness of the distribution.

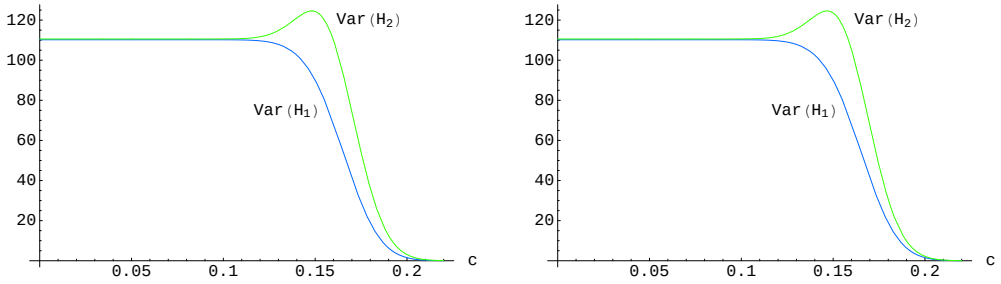
So far we have not yet looked at the dependency on the parameters a_1 and a_2. Even though there is a difference, we cannot see it by comparing Figures 2.24(a) and 2.24(b). We can make it visible by looking at the ratio $R = \frac{\mathbb{E}(H_1)}{\mathbb{E}(H_2)}$ and the log ratio $LR = log\frac{\mathbb{E}(H_1)}{\mathbb{E}(H_2)}$, which are illustrated in Figures 2.26 and 2.27.

Here, one sees a constant ratio and log ratio until the characteristic value of approximately $c = 0.12$ is reached in both cases. Afterwards, these values stay at this constant level in the case of $a_1 = a_2$ but increase in the case of $a_1 \neq a_2$. Thus, as soon as there is a small color effect, there will be an in-



(a) $a_1 = a_2 = 1$. (b) $a_1 = 1$, $a_2 = .99$.

Figure 2.24: The mean of the number of hybridized targets after washing at different detergent concentrations.

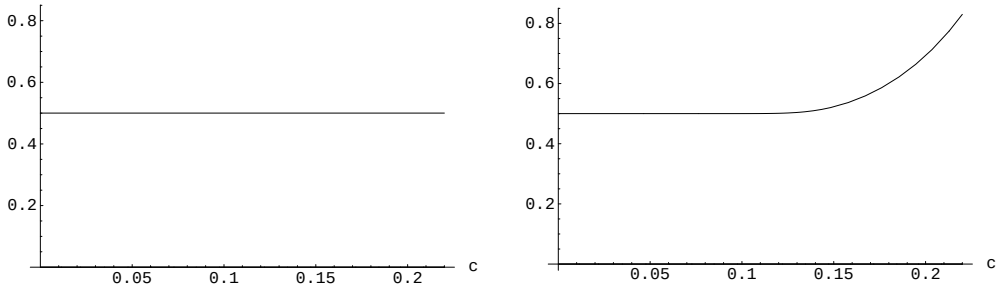


(a) $a_1 = a_2 = 1$. (b) $a_1 = 1$, $a_2 = .99$.

Figure 2.25: The variance of the number of hybridized targets after washing at different detergent concentrations.

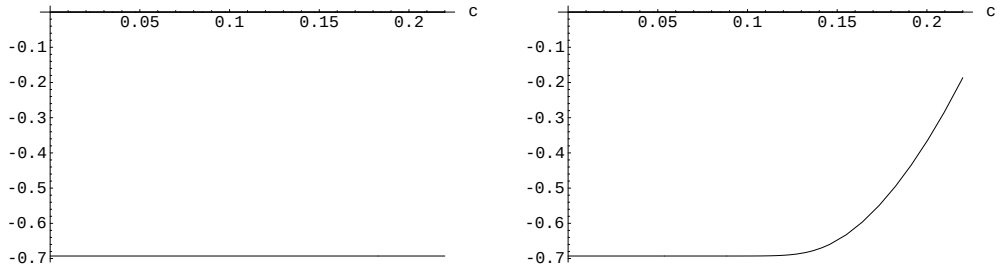
creasing error to the ratio and log ratio of the numbers of hybridized targets if the detergent concentration is past its characteristic value. Hence, washing too stringently will lead to a false signal. Therefore, it is recommended to find out the characteristic values of the detergent concentration for the different target types and as a consequence not to exceed these levels during the washing step.

Washing in presence of cross-hybridized targets The simplicity of the washing model makes it possible to analyze the more complex case of cross-hybridized targets. Assume the most simple case, where we have two different kinds of cDNA, both labeled with two fluorescence dyes. This yields four target types. One of the cDNAs hybridizes to the spot because of a complete consensus sequence to the probe sequence and the other one hybridizes due to a short partial consensus sequence. The corresponding targets to the latter



(a) $a_1 = a_2 = 1$. (b) $a_1 = 1$, $a_2 = 0.99$.

Figure 2.26: The ratio of the mean numbers of hybridized targets of both types after washing at different detergent concentrations.



(a) $a_1 = a_2 = 1$. (b) $a_1 = 1$, $a_2 = 0.99$.

Figure 2.27: The log ratio of the mean numbers of hybridized targets of both types after washing at different detergent concentrations.

case are called cross-hybridized targets. Those corresponding to the former case will have a much stronger binding to the spot than cross-hybridized targets. As a consequence less detergent molecules are needed to dissolve cross-hybridized targets.

We have determined the mean number of hybridized targets for the four types for the parameter situation of Table 2.17. Target types 1 and 2 belong to the first kind of cDNA and types 3 and 4 are targets which have cross-hybridized to the spot.

The parameters of target types 1 and 2 are chosen to be equal to the situation of Table 2.16. Since types 3 and 4 are supposed to be cross-hybridized targets, we assume that they can be dissolved more easily with only three detergent molecules. Further we assume less targets of types 3 and 4 being hybridized to the spot than those of types 1 and 2.

The trend of the mean number of hybridized targets for all four types after washing in dependency on different detergent concentrations is shown

maximal detergent intensities	
$\lambda_1^{\max}(c)$	10
$\lambda_2^{\max}(c)$	10
$\lambda_3^{\max}(c)$	10
$\lambda_4^{\max}(c)$	10
detergent concentration	
c	$\in [0, .22]$
duration of the washing procedure	
t	900
mean number of initially hybridized targets	
$\mathbb{E}(N_1)$	166.3
$\mathbb{E}(N_2)$	332.5
$\mathbb{E}(N_3)$	50
$\mathbb{E}(N_4)$	100
detergent molecules needed for solution	
k_1	10
k_2	10
k_3	3
k_4	3
parameter a	
a_1	1
a_2	.99
a_3	.5
a_4	.45

Table 2.17: Parameter situation for the case of two cDNA types labeled with two different colors.

in Figure 2.28.

Again, one observes a characteristic value where the behavior of the curves changes. It is about $c = 0.12$ for target types 1 and 2 whereas types 3 and 4 start to dissociate at approximately $c = 0.03$. Until reaching their characteristic values the target numbers are constant at the initial level from Table 2.17 and tend to zero afterwards.

While measuring fluorescence intensities of a single dye, there is no distinction between cross-hybridized and not cross-hybridized targets. The resulting curves are displayed in Figure 2.29.

If not accounting for cross-hybridization, we will receive a false signal for a wide range of detergent concentrations. Only the interval $c \in [.08, .12]$ yields

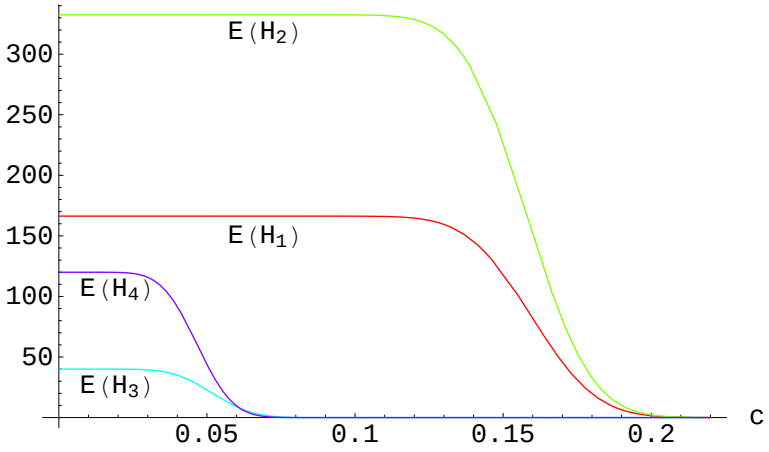
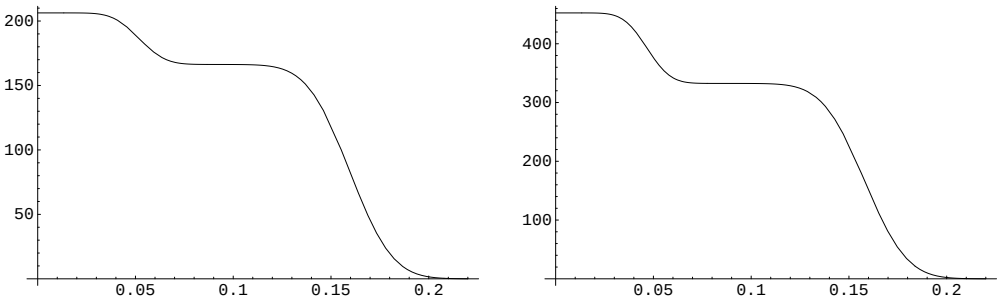


Figure 2.28: The mean numbers of hybridized targets after washing at different detergent concentrations and in presence of cross-hybridization.



(a) $\mathbb{E}(H_1) + \mathbb{E}(H_3)$. (b) $\mathbb{E}(H_2) + \mathbb{E}(H_4)$.

Figure 2.29: The sum of the mean numbers of hybridized targets according to one fluorescence dye after washing at different detergent concentrations.

the correct signal. So, again it is important to find out the behavior of the observed targets at different detergent concentrations and find the interval which yields the correct signal. The behavior described in Figure 2.29 has also been observed by biologists. See Figure 2.30 from [Drob].

The experimentally produced curves in subfigure (F) show a similar shape as those from our model in Figure 2.29, including the characteristic plateau where all cross-hybridized targets are washed off the spot. So, indeed the cross-hybridization effect has a practical meaning, which underlines the importance of washing at the right detergent concentrations. There is a large number of spots on a microarray. Each spot has its own interval of detergent concentrations which yield the characteristic plateau. Since the microarray is washed at a single concentration, we have to intersect the intervals of the

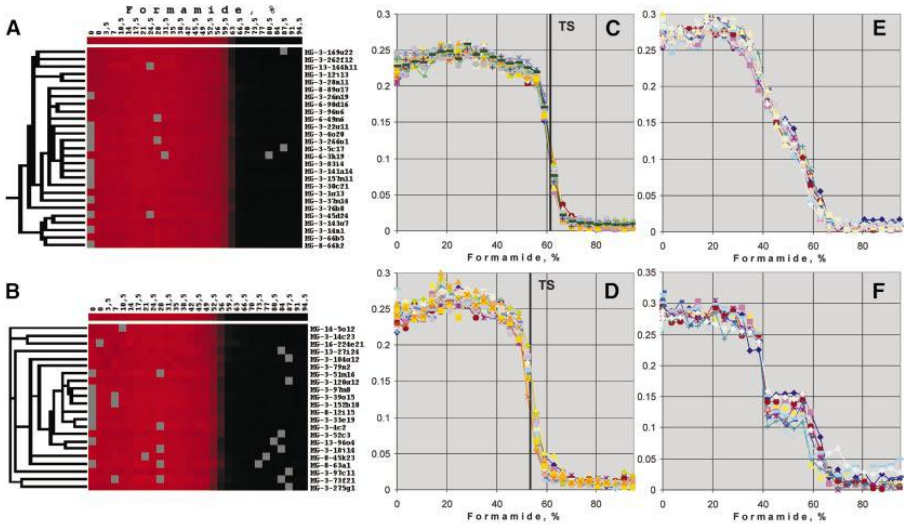
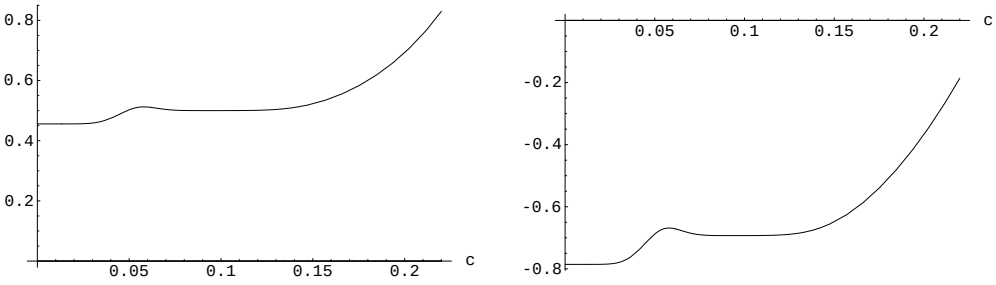


Figure 2.30: ([Drob]) Comprehensive assessment of shapes of fractionation curves from normalized data. Fragments of the cluster tree representing different types of fractionation curves for Cy5-labeled testis cDNA hybridization are shown. (A) Part of the hierarchical tree with genes having sharp transitions from the hybridized to non-hybridized state near 62% formamide that cluster together. (B) As (A) but with genes that have a sharp transition near 55% formamide. (C) Normalized signal intensities (y-axis) over increasing formamide concentrations (x-axis) of the same 27 genes as in (A). The vertical line indicates the transition stringency (TS), the mid-point of the transition from hybridized to dehybridized signal intensities. (D) Fractionation curves (x-axis, normalized signal intensities; y-axis, formamide concentration) of the same 21 genes as in (B). Vertical line indicates the transition stringency (TS) in this cluster of fractionation curves. (E) Cluster of 14 fractionation curves having broad transition regions. (F) Cluster of 10 fractionation curves having a two-step transition from the hybridized to non-hybridized state.

different spots in order to find the concentration which yields a characteristic plateau for all spots. The final interval might be very small or it might even vanish. This aspect should be considered in microarray experiments.

We will also look at the ratio and log ratio of the fluorescence intensities which are illustrated in Figure 2.31.

As can be seen, both start at a constant level, increase at the characteristic detergent concentration of target types 3 and 4, reach a local maximum at about $c = 0.05$, then decrease to another plateau before increasing after passing the characteristic detergent concentration of target types 1 and 2. The second plateau can be found at the level, which characterizes the actual ratio and log ratio of initially hybridized targets of types 1 and 2. Thus it is



(a) $\frac{\mathbb{E}(H_1)+\mathbb{E}(H_3)}{\mathbb{E}(H_2)+\mathbb{E}(H_4)}$. (b) $log\frac{\mathbb{E}(H_1)+\mathbb{E}(H_3)}{\mathbb{E}(H_2)+\mathbb{E}(H_4)}$.

Figure 2.31: The ratio and log ratio of the mean numbers of hybridized targets of both colors after washing at different detergent concentrations.

recommended to wash at a detergent concentration of this plateau in order to correctly infer the initial target levels.

We have to keep in mind, that the interesting ratio of regularly hybridizing targets $\frac{\mathbb{E}(N_1)}{\mathbb{E}(N_2)} = \frac{166.3}{332.5}$ is approximately 1/2. Too stringent washing and cross-hybridization significantly disturb the signal. This can be seen if looking at the respective figures of the ratio of hybridized targets after washing. This effect could be eliminated by washing at the right detergent concentration. So, finding out the right concentration should be of major interest to scientists before committing the washing step. Otherwise, inferring initial target numbers could not be conducted efficiently.

2.2.2.1 Résumé

In this section we have derived expressions for the mean and for the variance of the number of hybridized targets after the washing procedure from Section 1.2.2. These were used to determine the dependency of the model on the stringency of washing. In the case of two target types we could discover a small dye effect which will perturb the signal if the detergent concentration is too large. In addition, the case of four targets revealed a cross-hybridization effect. On the one hand, this effect was shown to be canceled out in the case of washing at a concentration which is strong enough to wash off cross-hybridized targets but leaves regularly hybridized targets on the spot. On the other hand it could be shown, that the cross-hybridization effect will yield completely wrong inferences if washing at other detergent concentrations. Thus, washing at the right concentration should be of major interest to researchers.

2.2.3 Fluorescence

Several noise sources contribute to the output of the fluorescence module, where spontaneous emission is considered to be the main noise source in laser devices. Other noise sources can be neglected. Thus, we will concentrate on spontaneous emission. As described in Section 1.2.3.2 it can be considered to be non random. Thus, it does not cause fluctuations in the intensity of the laser light. From Equation (1.15) we can determine the photon flux density including spontaneous emission. Combining it with

$$I(z) = \phi(z)h\nu$$

yields the intensity of the laser light (see [SaTe])

$$I(z) = h\nu\left(-\frac{\epsilon_{sp}(\nu)}{(N_2-N_1)\sigma(\nu)} + \left(\phi_0 + \frac{\epsilon_{sp}(\nu)}{(N_2-N_1)\sigma(\nu)}\right)e^{(N_2-N_1)\sigma(\nu)z}\right).$$

In our context, the laser noise cannot be valuated without considering its impact on the fluorescence intensity. According to Equation (1.16), the impact of the laser light intensity on the intensity of fluorescence is

$$F = 2.3p_f I(z)\tau\kappa l. \tag{2.44}$$

For the sake of simplicity let

$$a := 2.3p_f\tau l.$$

Further, let

$$I_u(z) = h\nu\phi_0 e^{(N_2-N_1)\sigma(\nu)z}$$

denote the laser light intensity without spontaneous emission according to Equation (1.13). Subtracting $I_u(z)$ from $I(z)$ yields the intensity due to spontaneous emission

$$\begin{aligned} I_{sp}(z) &= I(z) - I_u(z) \\ &= h\nu\frac{\epsilon_{sp}(\nu)}{(N_2-N_1)\sigma(\nu)}\left(e^{(N_2-N_1)\sigma(\nu)z} - 1\right). \end{aligned}$$

So Equation (2.44) can be rewritten as follows:

$$F = a(I_u(z) + I_{sp}(z))\kappa.$$

In general, two different laser devices are used to excite the fluorescence dyes in microarray experiments. Therefor, all symbols will get an additional superscript index if we want to distinguish between the two devices. The

value of interest is the ratio of target concentrations $R = \frac{\kappa^1}{\kappa^2}$. So far, F^1 and F^2 are measured and their ratio $\frac{F^1}{F^2}$ (or log ratio $log\frac{F^1}{F^2}$) is calculated and considered to be equal to R. Obviously, this assumption does not hold, because

$$\frac{F^1}{F^2} = \frac{a^1(I_u^1(z) + I_{sp}^1(z))\kappa^1}{a^2(I_u^2(z) + I_{sp}^2(z))\kappa^2}.$$

So, the ratio has to be corrected by the factor

$$CF = \frac{a^2(I_u^2(z) + I_{sp}^2(z))}{a^1(I_u^1(z) + I_{sp}^1(z))}. \tag{2.45}$$

A general analysis of the impact of CF on the fluorescence intensity is quite difficult due to the large quantity of parameters it depends on. But, it can be determined quite easily once the parameter values are known. In this case the ratio must be corrected by CF. But instead of accounting for it, researchers so far normalize microarray data for example by establishing the same mean or median of fluorescence intensities for the two dyes and all spots. The idea behind this procedure is that a vast majority of genes does not change its expression level and thus the mean or the median should almost stay constant. This works fine as long as the gene expression activity does not change too much between the cell states. Thus, another alternative only uses the intensity values of a selection of genes which are supposed to be constantly expressed during the different cell states (so-called housekeeping genes) or those which are artificially added to the microarray experiment (spike-in method). These two methods are better since they do not include the intensity effects of those genes which are differentially expressed and thus yield a more reliable correction. For details concerning the housekeeping gene methods see e.g. [Speed] and for an overview of spike-in methods see [Rydén]. The advantage of using all genes of the array is that the variance of the correction would be quite small compared to using a selection of genes. On the other hand, the bias due to all genes would be larger than the bias caused by a selection of genes. As a consequence, the mean of a normalization received from a selection of non-differentially expressed genes will be quite good but will have a large variance. In contrast, a normalization due all genes would have a small variance but a biased mean. The results from this section might help to overcome this problem.

Our model enables researchers to normalize microarray data non heuristically with the help of the correction factor CF. We suggest a study which compares the effectiveness of heuristic normalizing methods and our method. Subsequently, we will give an example of how to use our method.

Example: The parameters ϕ_0, N_1, N_2, ν and the functions $\sigma(\nu)$ and $\epsilon_{sp}(\nu)$ vary between different laser devices. Some of them can be found in the literature. Others have to be asked for at the manufacturer of the device of interest. Many scanners use an Argon ion laser exciting Cy3 (green) at a wavelength of $514\,nm$ and a Helium Neon laser exiting Cy5 (red) at a wavelength of $633\,nm$ (see [Ramp]). We will use their parameter values to examine the spontaneous emission error. We assume, target types 1 and 3 to be labeled wit Cy3 and the others with Cy5. The parameter setting is summarized in Table 2.18.

parameter values found in the literature		
	Argon ion laser	**HeNe laser**
ν	$.515\,\mu m$	$.6328\,\mu m$
$\sigma(\nu)$	$3 \times 10^{-12}\,cm^2$	$1 \times 10^{-13}\,cm^2$
parameter values not found in the literature		
ϕ_0	4×10^{15} photons$/cm^2 s$	5×10^{15} photons$/cm^2 s$
N_1	$.4 \times 10^{10}\,cm^{-3}$	$.1 \times 10^{10}\,cm^{-3}$
N_2	$.6 \times 10^{10}\,cm^{-3}$	$.9 \times 10^{10}\,cm^{-3}$
$d\Omega$	$\pi/24$	$\pi/24$
c	$299.792458 \times 10^6\,m/s$	$299.792458 \times 10^6\,m/s$
p_f	$1/3$	$1/4$
τ	$10^4\,m^2/mol$	$10^4\,m^2/mol$
l	$10^{-4}\,m$	$10^{-4}\,m$
d	$.1\,m$	$.1\,m$

Table 2.18: Parameter situation for Argon and Helium Neon lasers. Some of the parameter values could be found in the literature (see [SaTe], Chapter 13). Others have been assigned values which are reasonable to our understanding.

Using the parameter values of Table 2.18 yields $CF = .89$ according to Equation (2.45). Thus, if not modifying the ratio F^1/F^2 by CF would lead to an overestimation of the ratio of target concentrations κ^1/κ^2 of about 12%.

2.2.3.1 Résumé

With the help of the model from Section 1.2.3, we were able to show that using different lasers and dyes affects the fluorescence intensity. We derived the correction factor CF to normalize signal intensities. It uses physical properties of the laser devices to adjust the fluorescence intensities from the two colors. If we do not use the correction factor, the fluorescence intensities will be biased.

Unfortunately, most of the parameters had to be guessed. At this point, further analysis is recommended in order to determine the unknown parameter values of the model and to verify the correction factor by measurements.

2.2.4 Detection

In order to determine the noise due to detection one could measure the noise due to shot, generation-recombination, Johnson-Nyquist as well as Flicker noise and then determine the total noise according to Equation (1.21).

Instead, in this section we will analyze the single type branching process from Section 1.2.4 in order to understand the behavior of the detection process and to quantify the error which is due to this module.

A photon striking the photocathode causes emission of a single primary electron with an efficiency proportional to its frequency (energy). Due to the nature of either causing emission or not, we define the number of primary electrons caused by a single photon as a Bernoulli variable with success probability λ_p. The success probability can be identified with the quantum efficiency which is the number of primary electrons per photon. It takes values between 1% and 20%. The number of striking photons can reach values up to 10^8. For details see [Lako] and [Pawl]. Additionally, let the number of secondary electrons emitted by the ith dynode due to a single striking electron be Poisson distributed with intensity λ_s. This implies equal probability distributions at the different dynodes and thus is a restriction of the more general model from Section 1.2.4. We will only look at this case because we assume equal dynodes within a PMT (photomultiplier tube). Furthermore, we assume to have $l = 10$ dynodes which multiply the electron stream. This is a common number for PMTs ([WeAa], Chapter 14). [WeAa] also suggests $\lambda_s = 5$.

In the end of the multiplication process the secondary electrons are collected by the anode and the current is measured by the amperemeter, which is not able to display every single fluctuation in the current. Moreover it displays the sum of all electrons which hit the anode within a certain period of time. Let $N_{l,k}$ be the number of secondary electrons emitted by the lth dynode which are due to k photons striking the photocathode within this time span. These electrons hit the anode and thus cause the signal in the amperemeter. Our aim is to determine the probability distribution of $N_{l,k}$ in order to be able to estimate the error in the amperemeter due to the branching process. For this purpose, we will determine the probability generating function (p.g.f.) of $N_{l,k}$. According to the assumptions above, the number of

primary electrons due to a single photon has p.g.f.

$$Q_0(z) = (1 - \lambda_p) + \lambda_p z.$$

The number of secondary electrons due to an electron striking the mth dynode has p.g.f.

$$Q_m(z) = \sum_{i=0}^{\infty} \frac{\lambda_s^i}{i!} e^{-\lambda_s} z^i = e^{\lambda_s(z-1)}$$

for $m \in \{1, 2, ..., l\}$. Combining Recursion (1.22) with

$$\begin{aligned} h(z) &:= e^{\lambda_s(z-1)} \text{ and} \\ f(z) &:= (1 - \lambda_p) + \lambda_p z \end{aligned}$$

yields the p.g.f.

$$G_l(z) = \big(f \circ \underbrace{h \circ h \circ \ldots \circ h}_{\text{l times}}\big)(z)$$

for the number of secondary electrons $N_{l,1}$ from the lth dynode. The resulting p.g.f. of $N_{l,k}$ is

$$\begin{aligned} G_l^k(z) &= \mathbb{E}(z^{kN_{l,k}}) = \underbrace{\mathbb{E}(z^{N_{l,1}})\mathbb{E}(z^{N_{l,1}}) \cdot \ldots \cdot \mathbb{E}(z^{N_{l,1}})}_{\text{k times}} \\ &= G_l(z)^k \end{aligned} \tag{2.46}$$

for k independent photons causing k independent multiplication paths (see [Grab]).

In practice, combining Equation (2.46) with Equations (1.24) and (1.26) yields functions which cannot be evaluated efficiently due to computational power. Thus, we will use some properties of the p.g.f. to estimate a distribution of $N_{l,k}$.

We will determine the mean $\mathbb{E}(N_{l,k})$, the variance $\mathbb{V}ar(N_{l,k})$, the skewness $\gamma_1(N_{l,k})$ and the kurtosis $\gamma_2(N_{l,k})$ from the p.g.f. as follows. The mean and the variance can be expressed as (see [Grab])

$$\begin{aligned} \mathbb{E}(N_{l,k}) &= \frac{\partial}{\partial z} G_l^k(z)\bigg|_{z=1} \\ \mathbb{V}ar(N_{l,k}) &= \frac{\partial^2}{\partial z^2} G_l^k(z)\bigg|_{z=1} + \frac{\partial}{\partial z} G_l^k(z)\bigg|_{z=1} - \left(\frac{\partial}{\partial z} G_l^k(z)\bigg|_{z=1}\right)^2. \end{aligned} \tag{2.47}$$

The skewness and kurtosis are the third and fourth standardized moment and are defined as (see [FrHH])

$$\gamma_1(N_{l,k}) := \frac{\mathbb{E}(N_{l,k} - \mathbb{E}(N_{l,k}))^3}{\mathbb{V}ar(N_{l,k})^{3/2}} \text{ and} \tag{2.48}$$

$$\gamma_2(N_{l,k}) := \frac{\mathbb{E}(N_{l,k} - \mathbb{E}(N_{l,k}))^4}{\mathbb{V}ar(N_{l,k})^2}. \tag{2.49}$$

The following theorem will help to express γ_1 and γ_2 in terms of G.

Lemma 2.2.1. *([SchmK], Chapter 16) Let X be a random variable. If there is an $a \in (0, \infty)$ with $M_X(t) = \mathbb{E}(e^{tX}) < \infty$ for $t \in (-a, a)$ the following equation holds*

$$\mathbb{E}(X^k) = \frac{\partial^k}{\partial t^k} M_X(t)\Big|_{t=0}.$$

$M_X(t)$ is called *moment generating function*. Obviously, with $z = e^t$ follows

$$M_X(t) = \mathbb{E}(e^{tX}) = \mathbb{E}(z^X) = G_X(z) \tag{2.50}$$

where $G_X(z)$ is the probability generating function of X. Thus $G_X(z) < \infty$ implies $M_X(t) < \infty$. The probability generating function from Equation (2.46) is a composition of finitely many exponential functions and thus has always finite values within a finite interval $(-a, a)$. So Theorem (2.2.1) is useful to determine the moments of $N_{l,k}$.

For simplicity, let $N_{l,k} =: X$ and $G_l^k(z) = g(z)$. The moments are

$$\begin{aligned} \mathbb{E}(X) &= \frac{\partial}{\partial t} M_X(t)\Big|_{t=0} \\ &= \frac{\partial}{\partial t} g(e^t)\Big|_{t=0} \\ &= e^t g'(e^t)\big|_{t=0} \\ &= g'(1) \end{aligned} \tag{2.51}$$

as already seen above,

$$\begin{aligned} \mathbb{E}(X^2) &= \frac{\partial^2}{\partial t^2} M_X(t)\Big|_{t=0} \\ &= \frac{\partial^2}{\partial t^2} g(e^t)\Big|_{t=0} \\ &= \left(e^t g'(e^t) + g''(e^t)e^{2t}\right)\big|_{t=0} \\ &= g'(1) + g''(1), \end{aligned} \tag{2.52}$$

$$\begin{aligned}\mathbb{E}(X^3) &= \frac{\partial^3}{\partial t^3} M_X(t)\Big|_{t=0} \\ &= \frac{\partial^3}{\partial t^3} g(e^t)\Big|_{t=0} \\ &= \left(e^t g'(e^t) + 3g''(e^t)e^{2t} + g'''(e^t)e^{3t}\right)\big|_{t=0} \\ &= g'(1) + 3g''(1) + g'''(1) \end{aligned} \tag{2.53}$$

and

$$\begin{aligned}\mathbb{E}(X^4) &= \frac{\partial^4}{\partial t^4} M_X(t)\Big|_{t=0} \\ &= \frac{\partial^4}{\partial t^4} g(e^t)\Big|_{t=0} \\ &= \left(e^t g'(e^t) + 7g''(e^t)e^{2t} + 6g'''(e^t)e^{3t} + g^{(4)}(e^t)e^{4t}\right)\big|_{t=0} \\ &= g'(1) + 7g''(1) + 6g'''(1) + g^{(4)}(1). \end{aligned} \tag{2.54}$$

Remark: *The coefficients in front of the derivatives of the p.g.f. seem to be the sterling numbers of second kind. And indeed, this assumption could be verified with the help of the literature (see [Renyi], Chapter* 3*). However, for further considerations this is not relevant.*

Equations (2.51) and (2.52) can be used to derive the expressions for the mean and the variance as displayed above. With the help of all four equations we will derive the formulas for the skewness and the kurtosis. Expanding Equations (2.48) and (2.49) and using the linearity of the mean yields

$$\gamma_1(X) = \frac{\mathbb{E}(X^3) - 3\mathbb{E}(X^2)\mathbb{E}(X) + 3\mathbb{E}(X)\mathbb{E}(X)^2 + \mathbb{E}(X)^3}{\left(\mathbb{E}\left(X^2\right) - \mathbb{E}\left(X\right)^2\right)^{3/2}} \text{ and} \tag{2.55}$$

$$\gamma_2(X) = \frac{\mathbb{E}(X^4) - 4\mathbb{E}(X^3)\mathbb{E}(X) + 6\mathbb{E}(X^2)\mathbb{E}(X)^2 - 4\mathbb{E}(X)\mathbb{E}(X)^3 + \mathbb{E}(X)^4}{\left(\mathbb{E}\left(X^2\right) - \mathbb{E}\left(X\right)^2\right)^2}. \tag{2.56}$$

Combining Equations (2.55) and (2.56) with (2.51)-(2.54) yields

$$\gamma_1(X) = \frac{g'(1) + 2g'(1)^3 + 3g''(1) - 3g'(1)\left(g'\left(1\right) + g''\left(1\right)\right) + g'''(1)}{\left(g'\left(1\right) + g''\left(1\right) - g'\left(1\right)^2\right)^{3/2}}, \tag{2.57}$$

$$\gamma_2(X) = \frac{g'(1) - 3g'(1)^4 + 7g''(1) + 6g'(1)^2\left(g'\left(1\right) + g''\left(1\right)\right) + 6g'''(1) - 4g'(1)\left(g'\left(1\right) + 3g''\left(1\right) + g'''\left(1\right)\right) + g^{(4)}(1)}{\left(g'\left(1\right) + g''\left(1\right) - g'\left(1\right)^2\right)^2}. \tag{2.58}$$

Equations (2.57) and (2.58) enable us to determine the skewness and the kurtosis of the distribution of $N_{l,k}$ with the help of its p.g.f.. Both characteristics help to evaluate the similarity to a normal distribution.

Skewness is a measure of symmetry, where values close to zero indicate symmetric distributions. The peakedness of a distribution can be measured with the help of the kurtosis. The normal distribution has a kurtosis of 3 whereas larger values indicate a higher, more acute peak and fatter tails than those of a normal distribution. Smaller values indicate a lower, wider peak and thinner tails than those of a normal distribution.

The probability distribution of $N_{l,k}$ should only be positive in the first quadrant since $N_{l,k} < 0$ cannot occur. Thus, $\mathbb{E}(N_{l,k})$ must be positive. If we want to identify this distribution with a normal distribution, we must choose a positive mean and a variance that is small enough to keep almost all of the values in the first quadrant. "Almost all" can be verified quite well for normal distributions. 99.7% of the values of a normal distribution can be found within $[\mu - 3\sigma, \mu + 3\sigma]$ where μ is its mean and σ its standard deviation (see [Grab]). Thus, we propose a normal distribution with $\mu = \mathbb{E}(N_{l,k})$ and $\sigma = \sqrt{\mathbb{V}ar(N_{l,k})}$ which has an area under curve of almost 1 within the first quadrant. More precisely, we receive the condition

$$\mu - 3\sigma > 0$$

which is equivalent to

$$CV = CV(N_{l,k}) := \frac{\sigma}{\mu} < \frac{1}{3},$$

where CV denotes the *coefficient of variation*.

Following we will summarize the conditions which must be satisfied by the distribution of $N_{l,k}$:

$$\mu > 0, \tag{2.59}$$

$$CV < \frac{1}{3}, \tag{2.60}$$

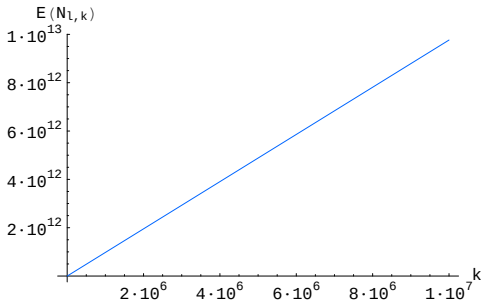
$$\gamma_1(N_{l,k}) \approx 0, \tag{2.61}$$

$$\gamma_2(N_{l,k}) \approx 3. \tag{2.62}$$

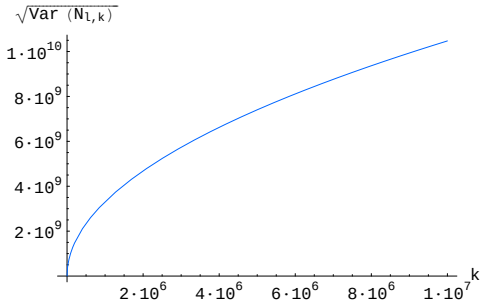
The four characteristics mean, variance, skewness and kurtosis as well as the coefficient of variation have been determined for different parameter situations in order to classify regions where the distribution of $N_{l,k}$ can be substituted by a normal distribution. These regions must satisfy the four Conditions (2.59)-(2.62).

The detection model has four parameters, i.e. the number of striking photons, the quantum efficiency, the intensity of the branching process at the dynodes and the number of dynodes within the PMT. For the entire examination we fix the number of dynodes to be $l = 10$. The characteristics have been investigated by keeping two other parameters constant and the fourth variable.

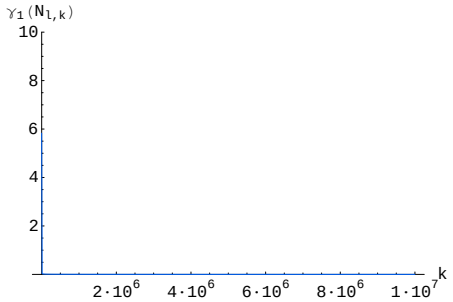
For constant λ_p, λ_s and variable k, the results are shown in Figure 2.32. We observe a fast decline of the CV below the critical value of $1/3$. In fact,



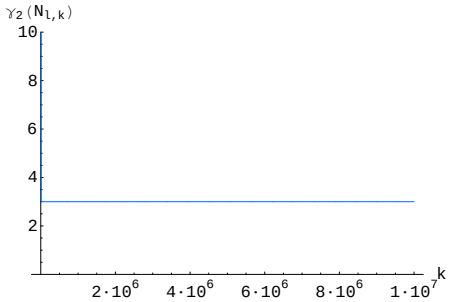
(a) The mean $\mathbb{E}(N_{l,k})$.



(b) The standard deviation $\sqrt{\mathbb{V}ar(N_{l,k})}$.



(c) The skewness $\gamma_1(N_{l,k})$.



(d) The kurtosis $\gamma_2(N_{l,k})$.

Figure 2.32: Mean, standard deviation, skewness and kurtosis for $\lambda_p = .1$, $\lambda_s = 5$ and $k \in [0, 10^7]$.

at $k = 104$ the CV falls below $1/3$ for the first time. Thus, beyond $k = 104$, the normal distribution is almost only concentrated in the first quadrant and thus might serve as a good approximation of $N_{l,k}$ whereas the case of $k < 104$ photons is not of major interest because it corresponds to very small fluorescence intensities.

We also have to analyze the shape of the distribution. As can be seen in Figure 2.32(c), the skewness descends to zero rapidly. This indicates a very symmetric distribution. For example, the value for $k = 10^3$ photons is already .119. From Figure 2.32(d) we also know, that the kurtosis fast descends to 3. This indicates that the peakedness of the distribution is close to a normal

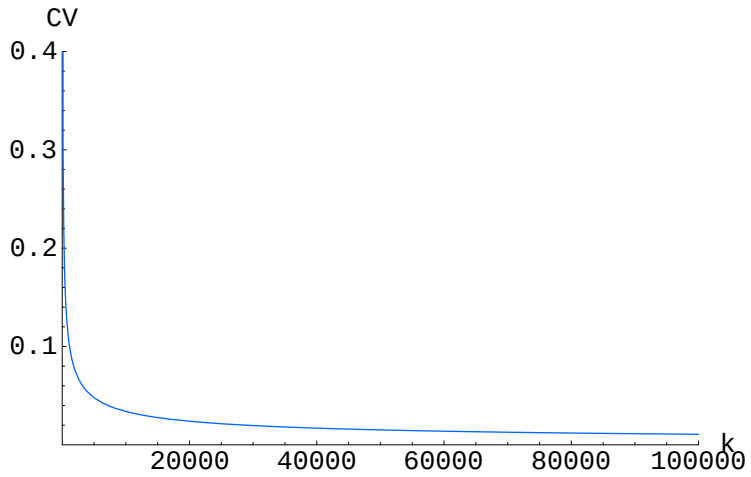


Figure 2.33: The coefficient of variation for $\lambda_p = .1$, $\lambda_s = 5$ and $k \in [0, 10^5]$.

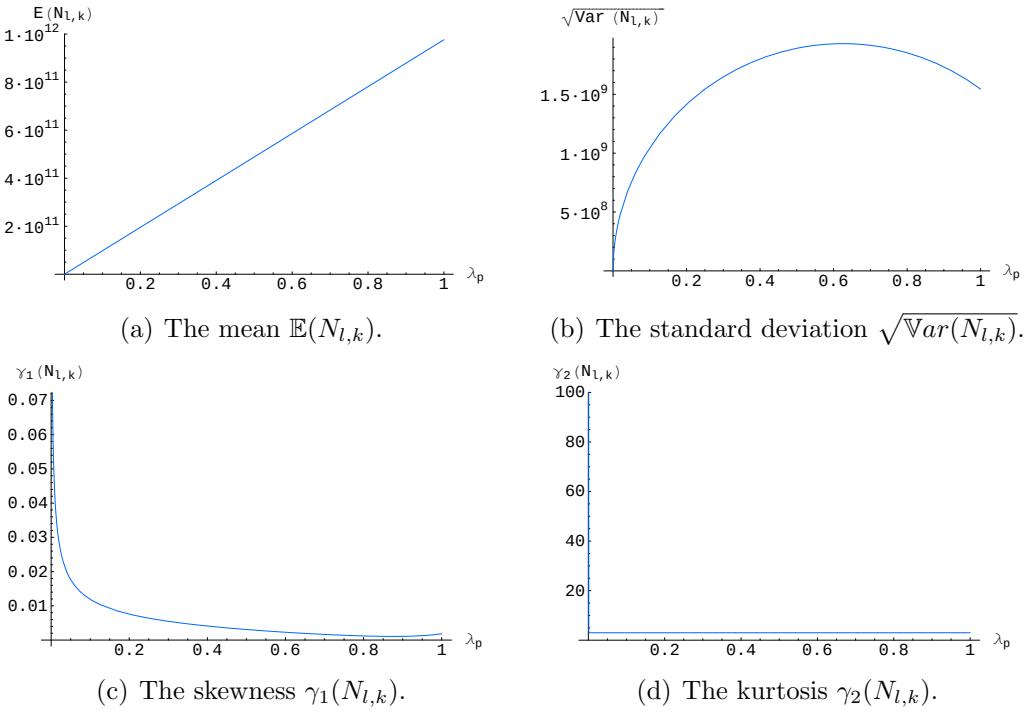
distribution. For example, the kurtosis at $k = 100$ has already descended to 3.147. In summary, the distribution of $N_{l,k}$ can hardly be distinguished from a normal distribution with mean $\mathbb{E}(N_{l,k})$ and standard deviation $\sqrt{\mathbb{V}ar(N_{l,k})}$ for k greater than 10^3 or 10^4. Thus in the case of fixed λ_p and λ_s this normal distribution is a good approximation.

The observation for fixed k and λ_s is shown in Figure 2.34. Again, we see a linear dependency of the mean and a much smaller standard deviation for larger quantum efficiencies λ_p. The coefficient of variation in Figure 2.35 confirms this result. It can be seen that it decreases fast below the critical value of $1/3$. Precisely, it falls below $1/3$ after passing $\lambda_p \approx 1.12 \times 10^{-4}$. So for realistic quantum efficiencies, almost the entire mass of the normal distribution with parameters μ and σ will be concentrated in the first quadrant.

Furthermore, for realistic quantum efficiencies of 0.05 and greater, the skewness is smaller than 0.02. The kurtosis instantaneously declines to 3. Hence, also for fixed k and λ_s the distribution of $N_{l,k}$ can be approximated by a normal distribution with mean $\mathbb{E}(N_{l,k})$ and standard deviation $\sqrt{\mathbb{V}ar(N_{l,k})}$.

The last case of constant k and λ_p looks similar. The results are shown in Figure 2.36. Mean and standard deviation increase exponentially but the variance is again some orders of magnitude smaller. This is corroborated by Figure 2.37. Here, the coefficient of variation falls below $1/3$ for $\lambda_s \geq 1$. Thus for all possible parameter situations, the normal distribution will be concentrated almost entirely in the first quadrant.

Skewness and kurtosis fulfill the Conditions (2.61) and (2.62) for $\lambda_s \geq 1$. Thus, for this case, the normal distribution with mean $\mathbb{E}(N_{l,k})$ and standard



(a) The mean $\mathbb{E}(N_{l,k})$.

(b) The standard deviation $\sqrt{\mathbb{V}ar(N_{l,k})}$.

(c) The skewness $\gamma_1(N_{l,k})$.

(d) The kurtosis $\gamma_2(N_{l,k})$.

Figure 2.34: Mean, standard deviation, skewness and kurtosis for $k = 10^5$, $\lambda_s = 5$ and $\lambda_p \in [0, 1]$.

deviation $\sqrt{\mathbb{V}ar(N_{l,k})}$ is an appropriate approximation of the distribution of $N_{l,k}$, too.

The knowledge of the previous section enables us to approximate the distribution of the number of secondary electrons hitting the anode by a normal distribution. Assuming that these electrons directly cause the signal in the amperemeter and neglecting noise in the amperemeter yields the distribution of the signal which is finally detected.

Example for two signals Subsequently, we will give a simple example. Assume for one gene the fluorescence intensities from the two colors are detected. We look at two different situations. The first situation deals with very small signals whereas the second situation considers large intensities. We will approximate the distribution of $N_{l,k}$ by a normal distribution as described in the previous section. Figures 2.38(a) and 2.38(b) show the two normal distributions due to the parameter situations of the first and of the second row in Table 2.19, respectively.

Obviously, the distributions in Figure 2.38(a) can be distinguished quite well even though they have a remarkable overlap. Here, we looked at very

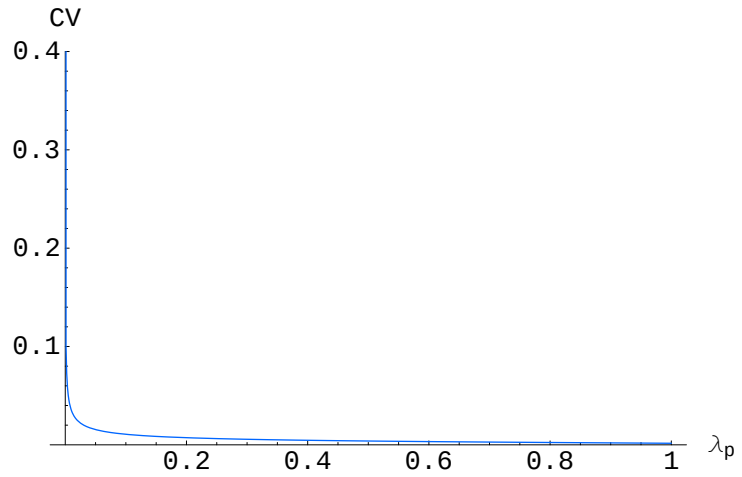


Figure 2.35: The coefficient of variation for $k = 10^5$, $\lambda_s = 5$ and $\lambda_p \in [0, 1]$.

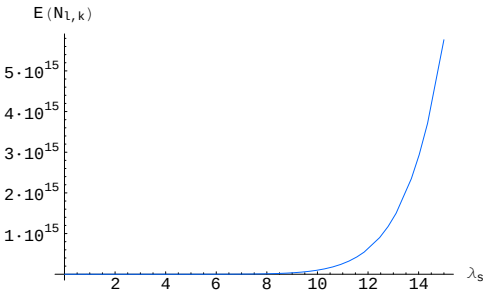
	k_1	k_2	λ_p	λ_s
small intensities	200	400	.1	5
large intensities	2,000	4,000	.1	5

Table 2.19: The parameter situations of small and large fluorescence intensities are displayed in the first and second row, respectively.

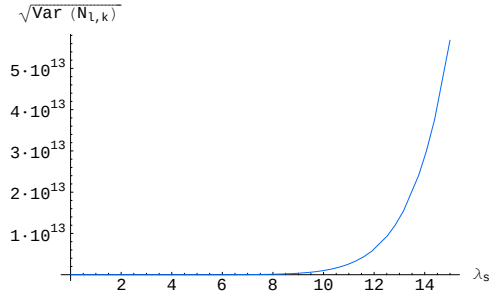
small intensity values which are not of major interest in microarray experiments. On the other hand, Figure 2.38(b) shows two distributions of larger but still quite small intensities. Here, no overlap is visible. A one sided Gauss test at a significance level of 5% has a power of almost 1 for this parameter situation. In Figure 2.39 we can see the power of this test in dependency on k, where the first signal consists of k photons and the second of $2k$ photons.

After passing $k = 300$, the power reaches about 1. Thus, a twofold in signal intensities will be detected almost surely if the smaller intensity is due to a number of striking photons larger than 300.

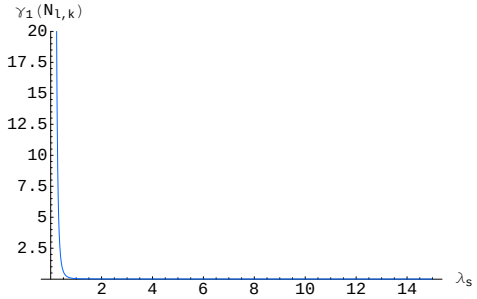
Behavior of the intensity ratio Another question is whether the ratio of striking photons is displayed in the current at the amperemeter? To answer this, we will have a look at the mean $\mathbb{E}(N_{l,k})$ and its dependency on the number of striking photons k.



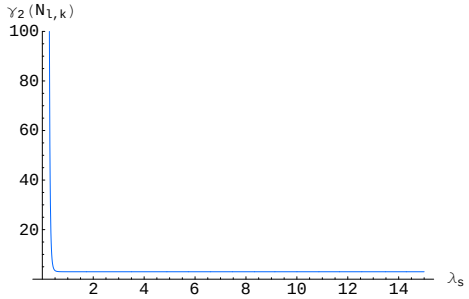
(a) The mean $\mathbb{E}(N_{l,k})$.



(b) The standard deviation $\sqrt{\mathbb{V}ar(N_{l,k})}$.



(c) The skewness $\gamma_1(N_{l,k})$.



(d) The kurtosis $\gamma_2(N_{l,k})$.

Figure 2.36: Mean, standard deviation, skewness and kurtosis for $k = 10^5$, $\lambda_p = .1$ and $\lambda_s \in [0, 15]$.

Let $\zeta_l(z)$ be the composition of $h(z)$ for l times, i.e.

$$\zeta_l(z) := \underbrace{h \circ h \circ \ldots \circ h}_{l \text{ times}}(z).$$

Lemma 2.2.2. *$\zeta_l(z)$ has derivative λ_s^l at $z = 1$.*

Proof. We prove this lemma by induction over l.
$\underline{l = 1:}$

$$\begin{aligned}
\frac{\partial}{\partial z}\zeta_1(z)\Big|_{z=1} &= \frac{\partial}{\partial z}h(z)\Big|_{z=1} \\
&= \frac{\partial}{\partial z}e^{\lambda_s(z-1)}\Big|_{z=1} \\
&= \lambda_s e^{\lambda_s(z-1)}\Big|_{z=1} \\
&= \lambda_s^1
\end{aligned}$$

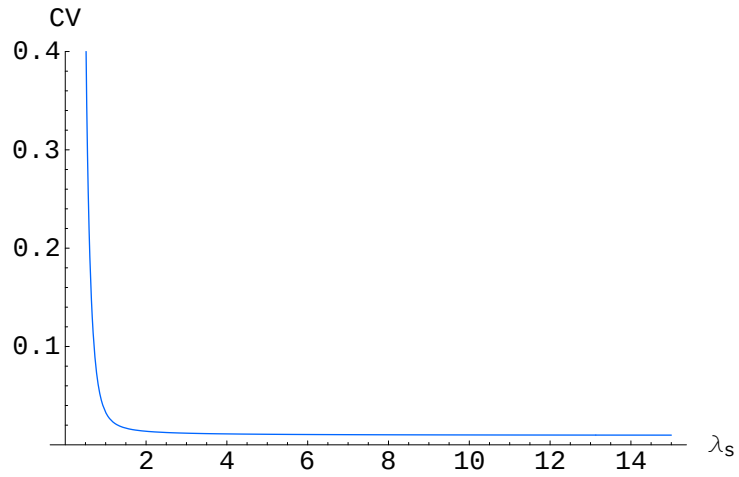
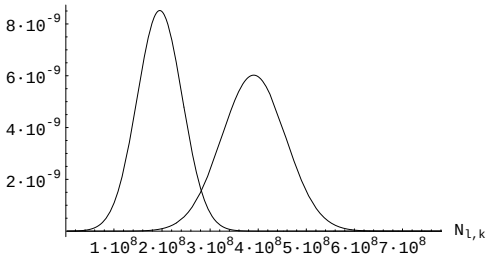
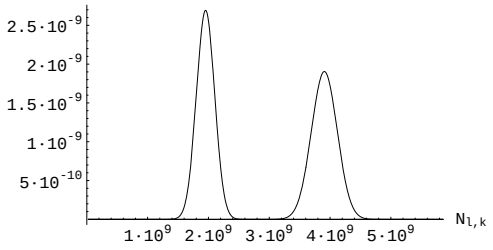


Figure 2.37: The coefficient of variation for $k = 10^5$, $\lambda_p = .1$ and $\lambda_s \in [0, 15]$.



(a) Two detection distributions according to the first row in Table 2.19.



(b) Two detection distributions according to the second row in Table 2.19.

Figure 2.38: Normal distributions to approximate the distribution of $N_{l,k}$.

$\underline{l \to l+1:}$

$$\begin{aligned}
\frac{\partial}{\partial z}\zeta_{l+1}(z)\Big|_{z=1} &= \frac{\partial}{\partial z}h(\zeta_l(z))\Big|_{z=1} \\
&= \frac{\partial}{\partial \zeta_l(z)}h(\zeta_l(z))\frac{\partial}{\partial z}\zeta_l(z)\Big|_{z=1} \\
&= \frac{\partial}{\partial \zeta_l(z)}h(\zeta_l(z))\Big|_{z=1}\ \frac{\partial}{\partial z}\zeta_l(z)\Big|_{z=1} \\
&= \lambda_s e^{\lambda_s(\zeta_l(1)-1)}\lambda_s^l \\
&= \lambda_s\lambda_s^l \\
&= \lambda_s^{l+1}
\end{aligned}$$

□

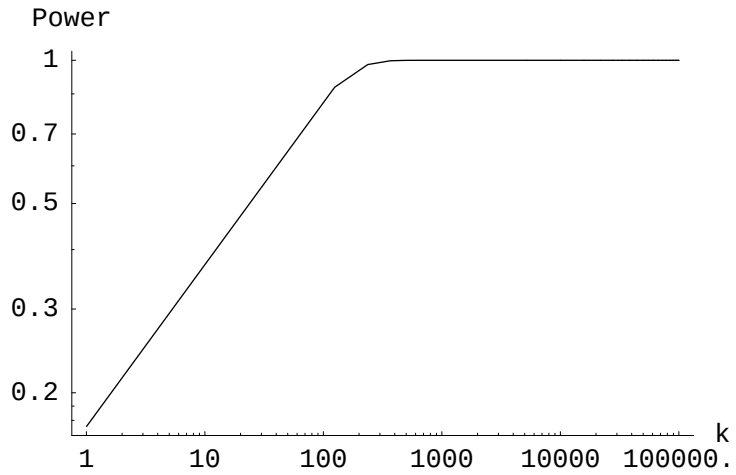


Figure 2.39: The power in dependency on the number of photons k of a one sided Gauss test for two normal distributions generated at a parameter situation of $l = 10$, $\lambda_p = .1$ and $\lambda_s = 5$. The first distribution is due to k and the second to $2k$ striking photons.

Theorem 2.15. *The mean of the number of electrons $N_{l,k}$ after passing the PMT aperture is*

$$\mathbb{E}(N_{l,k}) \;=\; k\lambda_p\lambda_s^l. \tag{2.63}$$

Proof.

$$\begin{aligned}
\mathbb{E}(N_{l,k}) &= \frac{\partial}{\partial z}G_l^k(z)\Big|_{z=1} \\
&= \frac{\partial}{\partial z}G_l(z)^k\Big|_{z=1} \\
&= kG_l(z)^{k-1}\frac{\partial}{\partial z}G_l(z)\Big|_{z=1} \\
&= kG_l(z)^{k-1}\frac{\partial}{\partial z}f(\zeta_l(z))\Big|_{z=1} \\
&= kG_l(z)^{k-1}\frac{\partial}{\partial \zeta_l(z)}f(\zeta_l(z))\frac{\partial}{\partial z}\zeta_l(z)\Big|_{z=1} \\
&= kG_l(1)^{k-1}\lambda_p\zeta_l(1)\lambda_s^l \\
&= k1^{k-1}\lambda_p 1\lambda_s^l \\
&= k\lambda_p\lambda_s^l
\end{aligned}$$

□

Using Formula (2.63) for two different numbers of striking photons k_1, k_2 yields a ratio of the means

$$\frac{\mathbb{E}(N_{l,k_1})}{\mathbb{E}(N_{l,k_2})} = \frac{k_1\lambda_p\lambda_s^l}{k_2\lambda_p\lambda_s^l} = \frac{k_1}{k_2}, \tag{2.64}$$

which is equal to the original ratio. Thus, ratios or even log ratios of the signals from the amperemeter display the ratio of striking photons. Therefore they can be used to infer the ratio of fluorescence intensities. At this point it is important to mention, that Equation (2.64) uses means and not the signals themselves. The signals are realizations from the normal distributions and thus are error prone according to the size of the variance. In the previous paragraphs we saw that in the case of large signals the variance is very small compared to the mean. Hence, the ratio of signal intensities is almost independent of the variance. Therefore, the variance can be neglected.

In the next paragraph we will look at the distribution of the intensity ratio. This includes the case of smaller signals where the variance is not small compared to the mean.

2.2.4.1 The distribution of the ratio of intensities

E. C. Fieller [Fiell] developed probability density functions of ratios of correlated random variables. We will use the more recent version of his results from [Hink] to derive the probability density function of

$$R_N := \frac{N_{l,k_1} + N_{l,k_3}}{N_{l,k_2} + N_{l,k_4}}, \tag{2.65}$$

which is the ratio of signal intensities in the case of cross-hybridization with normally distributed signals N_{l,k_1}, N_{l,k_3} of the first color and signals N_{l,k_2}, N_{l,k_4} of the second color.

The number of striking photons and secondary electrons is supposed to be very small compared to the number of atoms in the photocathode and in the dynodes. Hence, there is almost no competition of these particles for the atoms. Therefore, we will assume independently distributed N_{l,k_1}, N_{l,k_2}, N_{l,k_3} and N_{l,k_4}. Thus, we have $X := N_{l,k_1} + N_{l,k_3}$ and $Y := N_{l,k_2} + N_{l,k_4}$ are normally distributed with means $\mathbb{E}(X) := \mathbb{E}(N_{l,k_1}) + \mathbb{E}(N_{l,k_3})$ and $\mathbb{E}(Y) := \mathbb{E}(N_{l,k_2}) + \mathbb{E}(N_{l,k_4})$ and variances $\mathbb{V}ar(x) := \mathbb{V}ar(N_{l,k_1}) + \mathbb{V}ar(N_{l,k_3})$ and variance $\mathbb{V}ar(Y) := \mathbb{V}ar(N_{l,k_2}) + \mathbb{V}ar(N_{l,k_4})$, respectively.
Before stating the result of interest we will give some definitions.

Definition 2.16. *Let*

$$\begin{aligned}
a(w) &= \left(\frac{w^2}{\mathbb{V}ar(X)} - \frac{2\rho w}{\sqrt{\mathbb{V}ar(X)\mathbb{V}ar(Y)}} + \frac{1}{\mathbb{V}ar(Y)}\right)^{1/2},\\
b(w) &= \frac{\mathbb{E}(X)w}{\mathbb{V}ar(X)} - \frac{\rho(\mathbb{E}(X)+\mathbb{E}(Y)w)}{\sqrt{\mathbb{V}ar(X)\mathbb{V}ar(Y)}} + \frac{\mathbb{E}(Y)}{\mathbb{V}ar(Y)},\\
c &= \frac{\mathbb{E}(X)^2}{\mathbb{V}ar(X)} - \frac{2\rho\mathbb{E}(X)\mathbb{E}(Y)}{\sqrt{\mathbb{V}ar(X)\mathbb{V}ar(Y)}} + \frac{\mathbb{E}(Y)^2}{\mathbb{V}ar(Y)},\\
d(w) &= exp\left(\frac{b(w)^2 - ca(w)^2}{2(1-\rho^2)a(w)^2}\right),\\
\Phi(z) &= \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du,
\end{aligned}$$

where ρ is the coefficient of correlation of the normally distributed random variables X and Y.

Lemma 2.2.3. *The probability density function of the quotient X/Y of two normally distributed random variables X, Y with coefficient of correlation ρ is*

$$\begin{aligned}
f(w) =& \frac{b(w)d(w)}{\sqrt{2\pi\mathbb{V}ar(X)\mathbb{V}ar(Y)}a(w)^3}\left[\Phi\left(\frac{b(w)}{\sqrt{1-\rho^2}a(w)}\right) - \Phi\left(\frac{b(w)}{\sqrt{1-\rho^2}a(w)}\right)\right]\\
&+ \frac{\sqrt{1-\rho^2}}{\pi\sqrt{\mathbb{V}ar(X)\mathbb{V}ar(Y)}a(w)^2} exp\left(-\frac{c}{2(1-\rho^2)}\right).
\end{aligned}$$

([Hink]) Assuming independence of X and Y implies $\rho = 0$. Combining this with Definition 2.16 and Lemma 2.2.3 yields the following simplified definition and lemma.

Definition 2.17. *Let*

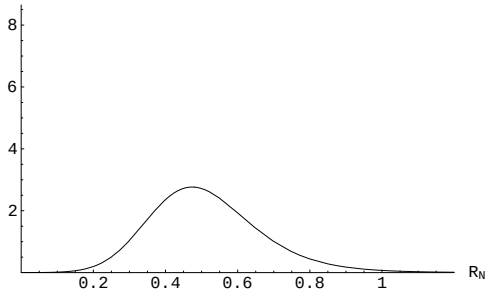
$$\begin{aligned}
a(w) &= \left(\frac{w^2}{\mathbb{V}ar(X)} + \frac{1}{\mathbb{V}ar(Y)}\right)^{1/2},\\
b(w) &= \frac{\mathbb{E}(X)w}{\mathbb{V}ar(X)} + \frac{\mathbb{E}(Y)}{\mathbb{V}ar(Y)},\\
c &= \frac{\mathbb{E}(X)^2}{\mathbb{V}ar(X)} + \frac{\mathbb{E}(Y)^2}{\mathbb{V}ar(Y)},\\
d(w) &= exp\left(\frac{1}{2}\frac{b(w)^2}{a(w)^2} - \frac{1}{2}c\right),\\
\Phi(z) &= \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du
\end{aligned}$$

for independently, normally distributed random variables X and Y.

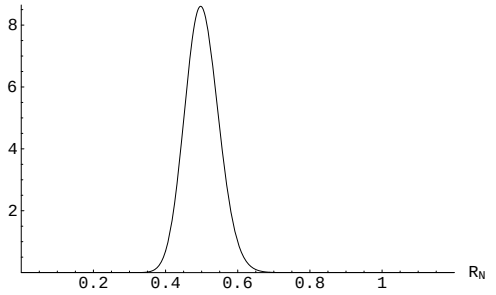
Lemma 2.2.4. *([Hink]) The probability density function of the quotient X/Y of two independent normally distributed random variables X, Y is*

$$f(w) = \frac{b(w)d(w)}{\sqrt{2\pi\mathbb{V}ar(X)\mathbb{V}ar(Y)}a(w)^3}\left[2\Phi\left(\frac{b(w)}{a(w)}\right)-1\right] + \frac{1}{a(w)^2\pi\sqrt{\mathbb{V}ar(X)\mathbb{V}ar(Y)}}exp\left(-\frac{1}{2}c\right).$$

Using this result, we have a look at the probability density functions for the two signals from the first and the second row of Table 2.19 which are illustrated in Figure 2.40.



(a) The ratio distribution according to Lemma 2.2.4 and row one in Table 2.19.



(b) The ratio distribution according to Lemma 2.2.4 and row two in Table 2.19.

Figure 2.40: The ratio distributions of R_N of two detection signals and two different parameter situations according to Lemma 2.2.4.

The distribution of Figure 2.40(a) has a mean of ≈ 0.5158 and variance of ≈ 0.0252. Approximately 99.2% of the values can be found between 0 and 1. The distribution of Figure 2.40(b) has a mean of ≈ 0.5015 and variance of ≈ 0.0022. Here, 99.8% of the values can already be found between 0.35 and 0.65.

On the one hand, both means are close to the initial ratio of $1/2$ which corroborates the previous results. On the other hand it can be seen that the variance strongly depends on the strength of the signal. Smaller numbers of striking photons imply a larger noise. For that reason, the detection scale has to be looked at critically.

2.2.4.2 Résumé

In a first step we looked at the p.g.f. of the number of detected electrons $N_{l,k}$ at the amperemeter and tried to determine the probability distribution of

$N_{l,k}$. This failed for concrete parameter values due to computational power. Thus, we approximated the distribution by a normal distribution. Again with the help of the p.g.f., we derived the four moments mean, variance, skewness and kurtosis as well as the coefficient of variation. We used these values to verify the approximation for realistic parameter values.

Afterwards the respective normal distribution was used to give an example for two different signal intensities striking the photocathode. We could see, that there was a rather large overlap for small intensities compared with larger intensities.

This behavior was corroborated by looking at the power of a simple test, which even helped to find a kind of minimal intensity to separate two signals.

Finally, we derived the distribution of the intensity ratio and could see that its variance depends on the strength of the signals, too.

Summarizing the results from this section, we could see that the signal within a PMT is amplified very well and the noise due to the amplification process does hardly affect the power of distinguishing two different signals as long as the number of striking photons is sufficiently large. In this case, using PMTs is a good method to make photon streams visible without perturbing the ratio of the photon stream intensities. In the case of small signals a PMT is not an advisable detection aperture.

Chapter 3

The compound model

In this chapter we will analyze the way of the signal through the entire microarray process by simulation. We will make use of the previous results in order to state a model which concatenates the modules. This model shall be called the *compound model*.

We will simulate the models of the hybridization including dissociation, the washing and the reverse transcription modules. During the fluorescence step we will use the reciprocal of the correction factor CF from Equation (2.45) to account for fluorescence noise. To simulate the branching process from the detection module is computational too expensive. Thus we will use the results from Section 2.2.4 and approximate the distribution of electrons detected at the amperemeter by a normal distribution. Mean, standard deviation and histograms of the particle distributions within each step shall be determined in order to identify major noise sources. We will finish this chapter by comparing the input signal (mRNA) with the detected output.

Since simulation is possible for four target types we can skip the case of two types and directly look at the case which includes cross-hybridization. Target types 1 and 2 are specific to the spot whereas types 3 and 4 are supposed to be unspecific as summarized in Table 3.1.

	specific target types	unspecific target types
Cy3 labeled	1	3
Cy5 labeled	2	4

Table 3.1: Specification of target types.

The reverse transcription process labels each target with a random number of dye molecules. The result is a large number of target types. Simulating

the hybridization process with such large numbers is impossible due to computational power. Thus, during the hybridization and the washing process as stated in Chapter 1 we summarize all targets into four different classes, the target types, instead of considering each target as a single type. As a result, both processes are de facto independent of the number of dye molecules attached to the targets. Thus, it does not matter at which point in time the dyes are incorporated. Since it is much easier, we will simulate the labeling after the hybridization and washing process. One could imagine, that we randomly labeled the nucleotides bound to the array with the respective color according to their target type. For that reason, in this chapter will use the term *labeling process* instead of reverse transcription process.

In detail, we will investigate the compound model by following order modules:

1. hybridization,
2. washing,
3. labeling process (reverse transcription),
4. fluorescence and
5. detection.

We will determine mean, standard deviation, histogram and intensity ratios of the particles involved. The ratio of interest for all modules is

$$R_P = \frac{P_1 + P_3}{P_2 + P_4}, \tag{3.1}$$

where P_i, $i = 1, 2, 3, 4$ is the placeholder of particle or value i after the respective module. E.g., in the hybridization module it is the placeholder of N_i.
We will look at two different scenarios. During the first scenario, we start the analysis of the compound module with the initial target numbers T_i, $i = 1, 2, 3, 4$ from Table 3.2. The respective initial ratio is $R_T = 1$. The second scenario uses the same parameter situation besides a doubled number of type 2 targets. This yields an initial target ratio of $R_T = 2/3$ considering all targets and an initial ratio of $1/2$ if only considering specific targets.

In addition, for each module we will determine the significance of the distortion of the ratio of particles by looking at its empirical 95% confidence interval.

In the following, we consecutively analyze the modules.

3.1 Hybridization

We will start with 100 simulations of the hybridization process with four target types from Section 1.1 at the parameter setting from Table 3.4.

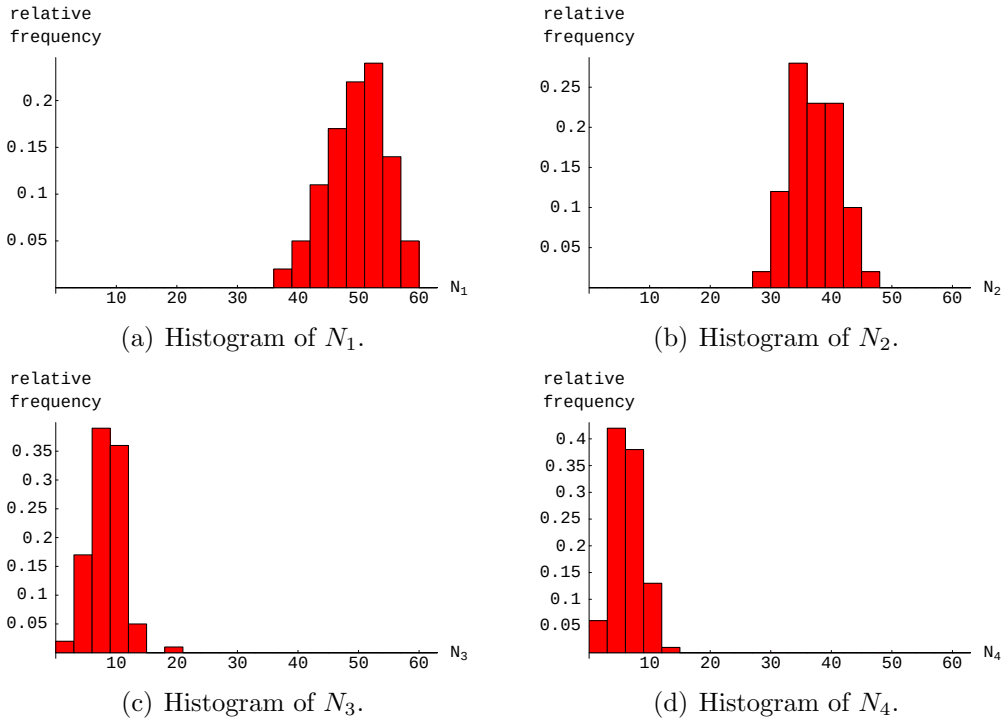
binding probabilities	
π_1	.7
π_2	.6
π_3	.2
π_4	.15
dissociation probabilities	
γ_1	.3
γ_2	.4
γ_3	.8
γ_4	.85
initial target numbers	
T_1	$10,000$
T_2	$10,000$
T_3	$10,000$
T_4	$10,000$
number of probes	
S	100
exponential clock	
λ	2
duration of the experiment	
θ	1.5

Table 3.2: Parameter situation in presence of cross-hybridization for the compound model.

Regarding the situation of Section 2.1.4, we scaled down the probe and the target numbers but increased the duration of the experiment. The downscaling has hardly any effects on the ratio of target types as could be seen in Section 2.1.4 whereas through the extension of the duration the process is supposed to be closer to its stationary distribution. The hybridization reaction has been simulated with Gillespie's Algorithm ([Gill]).

Figure 3.1 shows the histograms of the four hybridized target numbers.

Obviously all four types have hybridized to the spot with different efficiencies. It seems, there are two main factors which influence the efficiencies. On the one hand, specific target types have hybridized better than unspecific. On the other hand, targets labeled with the first color (types 1 and



(a) Histogram of N_1.

(b) Histogram of N_2.

(c) Histogram of N_3.

(d) Histogram of N_4.

Figure 3.1: Histograms for the four target types after the hybridization process at the parameter situation of Table 3.2.

3) seem to have hybridized better than those of the second color (types 2 and 4). This behavior is not surprising, we even expect it. The reasons can be found in the difference in the rates and probabilities of hybridization and dissociation.

The respective means and standard deviations of the four types are summarized in Table 3.3 and underline the observation from the histograms since their means indicate the same ranking of hybridization efficiencies. Recall

$\mu(N_1)$	$\sigma(N_1)$	$\mu(N_2)$	$\sigma(N_2)$	$\mu(N_3)$	$\sigma(N_3)$	$\mu(N_4)$	$\sigma(N_4)$
49.21	4.69	36.86	3.98	7.98	2.62	5.85	2.35

Table 3.3: Mean and standard deviation of the four target types after the hybridization reaction.

that before the hybridization process the ratio of initial target numbers was $R_T = 1$. After the hybridization module we have a mean $\mu(R_N) \approx 1.37$ and a variance $\sigma^2(R_N) \approx 0.067$. Using the $2\sigma-$rule leads to a 95% confidence interval for the real mean of $[0.85, 1.89]$. So, the deviation of the ratio from 1 is

not significant. Using the $2\sigma-$rule implies a normal distribution of the ratio ([Grab]). That this assumption is not too bad can be seen in the respective histogram (see Figure 3.2). In order to avoid errors made by the assumption

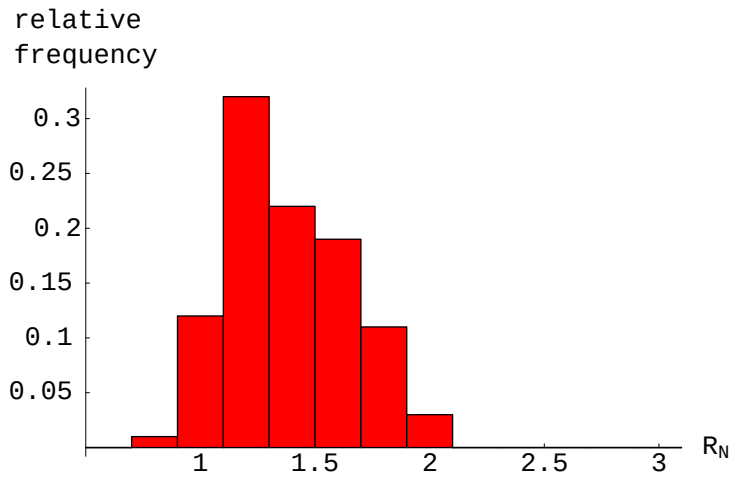


Figure 3.2: Histogram of the ratio R_N of the output of the hybridization reaction.

of a normal distribution, we will derive the confidence interval from the ordered list of the 100 ratio values, empirically. For this purpose we determine the 2.5%- and 97.5%-quantiles. In between we find 95% of all observed ratios. The respective interval is the empirical 95% confidence interval. The analysis of subsequent modules shall be restricted to this interval.

Looking at the ordered list of ratio values after hybridization yields a 2.5%-quantile of about 0.96 and a 97.5%-quantile of about 1.94. This leads the empirical 95% confidence interval of about $[0.96, 1.94]$.

Summarizing the results, the hybridization module adds about 37% error on average to the initial ratio. But, this deviation is not significant. Next, we will look at the washing module.

3.2 Washing

We will simulate the washing model from Section 1.2.2 by drawing a random number of dissolved targets from its probability distribution in Equation (1.11). The input of the washing module shall be the 100 data sets which have been generated by the hybridization module in the previous paragraph. We will use the model from Section 1.2.2 and the parameter setting of Table 3.4 to simulate the washing procedure separately for each data set.

The parameter values are chosen to be equal to the consideration from Section 2.2.2 besides the detergent concentration which is fixated at a value

maximal detergent intensities	
$\lambda_1^{\max}(c)$	10
$\lambda_2^{\max}(c)$	10
$\lambda_3^{\max}(c)$	10
$\lambda_4^{\max}(c)$	10
detergent concentration	
c	.1
detergent molecules needed for solution	
k_1	10
k_2	10
k_3	3
k_4	3
parameter a	
a_1	1
a_2	.99
a_3	.5
a_4	.45

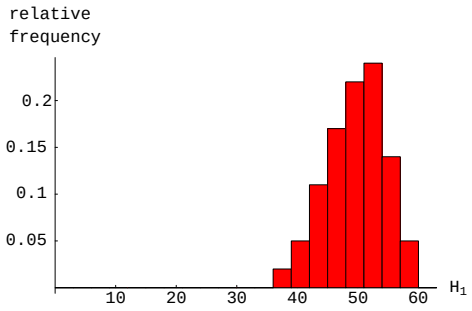
Table 3.4: Parameter situation during the washing module of the compound model. It models the case of two cDNA types labeled with two different colors.

in the interval where unspecific targets dissociate from the spot but specific targets are still hybridized. This implies the assumption that the microarray is always washed at the right detergent concentration. Looking at detergent concentrations outside of this interval either yields a washing process without effect (too low concentrations of the detergent) or a complete removal of all targets (too high concentrations of the detergent). The former case would leave the ratio from the hybridization model unchanged whereas the latter case would imply a ratio of target types which does not display the ratio of initial target numbers at all and thus falsifies the signal entirely. These conclusions correspond to Figure 2.31 from Section 2.2.2 where only a small bandwidth of washing intensities leads the correct ratio.

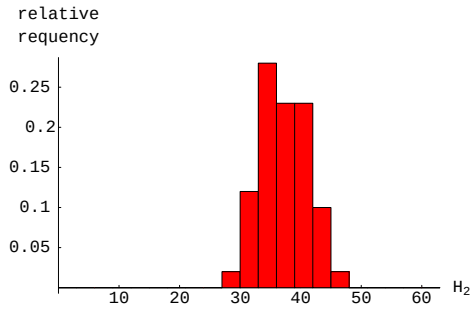
Figure 3.3 shows the histograms of hybridized targets after washing.

The washing step seems to work quite well. As can be seen in the histograms, all cross-hybridized targets are washed off and the histograms of specific targets are left almost unchanged. This observation can also be verified by looking at the means and standard deviations of the target numbers which are summarized in Table 3.5.

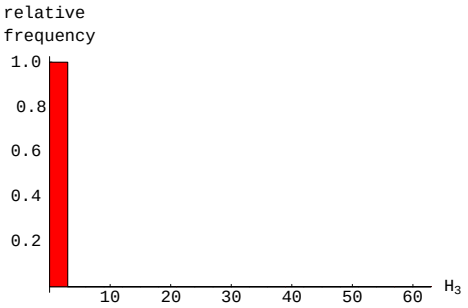
Here, the means and standard deviations of the first two target types do



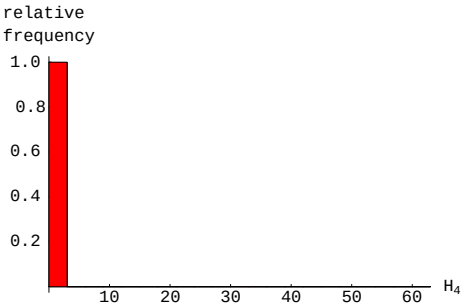
(a) Histogram of H_1.



(b) Histogram of H_2.



(c) Histogram of H_3.



(d) Histogram of H_4.

Figure 3.3: Histograms for the four target types after the washing process at the parameter situation of Table 3.4.

$\mu(H_1)$	$\sigma(H_1)$	$\mu(H_2)$	$\sigma(H_2)$	$\mu(H_3)$	$\sigma(H_3)$	$\mu(H_4)$	$\sigma(H_4)$
49.21	4.70	36.85	3.98	0	0	0	0

Table 3.5: Mean and standard deviation of the four target types after the washing reaction.

hardly change (compare Table 3.3). Those of the types 3 and 4 vanish.

Determining the ratio yields $\mu(R_H) \approx 1.36$ and $\sigma^2(R_H) \approx 0.069$. This implies a deviation of 36% on average from the initial ratio which is a slight improvement compared to the ratio of the previous module. So, indeed, the washing procedure might help to reduce the noise of the signal. Looking at the ordered list of washing ratios yields the empirical 95% confidence interval of approximately $[0.95, 1.90]$ of the real ratio.

So, the deviation of the ratio from 1 is still not significant. See Figure 3.4 for the histogram of ratios.

Due to washing, there are no hybridized targets of types 3 and 4. Thus, there cannot be any labeled nucleotides hybridizing to such targets during

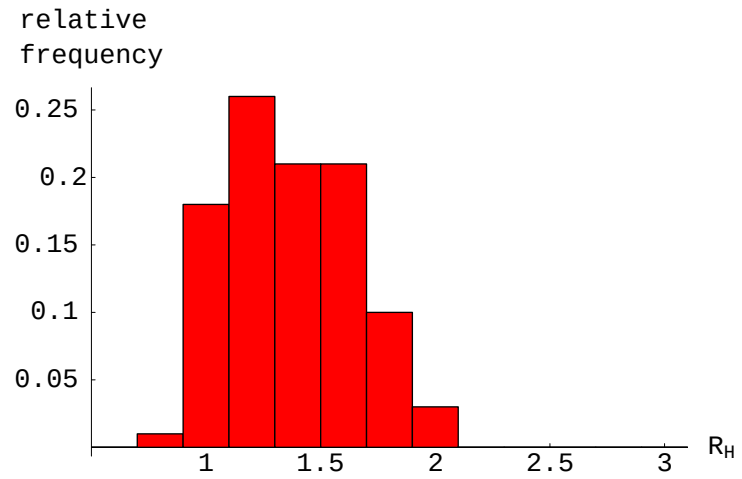


Figure 3.4: Histogram of the ratio R_H of the output of the washing reaction.

the labeling process. Consequently, there will not be any labeled nucleotides at targets of types 3 and 4 either, which might fluoresce in order to be detected. For that reason, the number of particles of types 3 and 4 will be zero in all subsequent modules. Thus, we only keep them in mind to determine the particle ratios.

The 100 data sets from the output of the washing module are used as input for the reverse transcription process.

3.3 Labeling process (reverse transcription)

As mentioned in the introduction of this chapter, we will use the output of the washing module as input for the labeling process. The process will be simulated by drawing a random number of labeled nucleotides from a binomial distribution for each hybridized target according to Equation (1.8). Since there are no hybridized targets of types 3 and 4, after washing we will omit the simulation of these types.

We restrict the simulation to the parameter situation of Table 3.6.

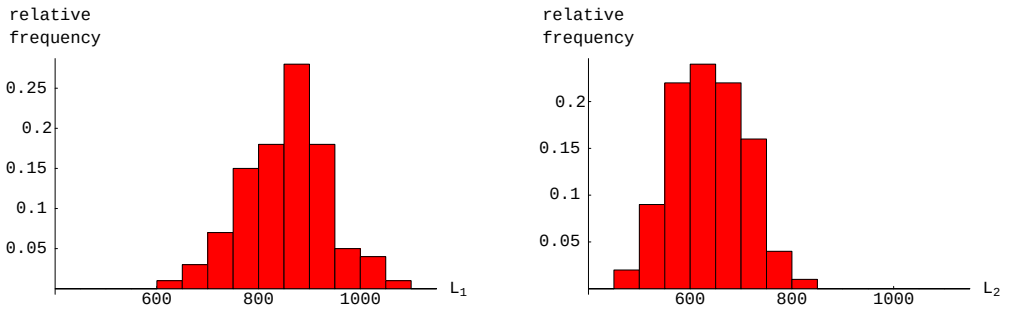
V_l	V_u	m_1	r_l	r_u
$9.03320505 \times 10^{21}$	6.0221367×10^{21}	30	.45	.5

Table 3.6: Parameter situation for the labeling transcription module of the compound model.

The initial numbers of labeled and unlabeled nucleotides are chosen ac-

cording to the protocol from the Appendix (see B). The number of potential binding sites for labeled nucleotides and for the recruitment rates are chosen in agreement with Section 2.2.1.

The detection signal is caused by all labeled nucleotides from one target type. Summing up the number of labeled nucleotides which are attached to targets of a certain type yields the histograms from Figure 3.5



(a) Histogram of L_1. (b) Histogram of L_2.

Figure 3.5: Histograms for the numbers of labeled nucleotides to the targets of remaining types 1 and 2 after the labeling process with the parameters of Table 3.6.

where

$$L_i := \sum_{j=1}^{H_i} Z_i(m)$$

is the total number of hybridized nucleotides to all targets of type i, $i = 1, 2$. The means and the standard deviation of the two remaining types are summarized in Table 3.7.

$\mu(L_1)$	$\sigma(L_1)$	$\mu(L_2)$	$\sigma(L_2)$
851.42	84.69	634.12	70.85

Table 3.7: Mean and standard deviation for the total numbers of labeled nucleotides of the four target types after the labeling reaction.

The respective ratio has mean $\mu(R_L) \approx 1.37$ and variance $\sigma^2(R_L) \approx 0.074$. So, the deviation of the ratio after the modules of labeling, hybridization and washing is about 37% on average. Thus, it hardly changed, compared to the previous modules. This means almost all of the noise so far is due to the hybridization module. The histogram of the ratio is shown in Figure 3.6.

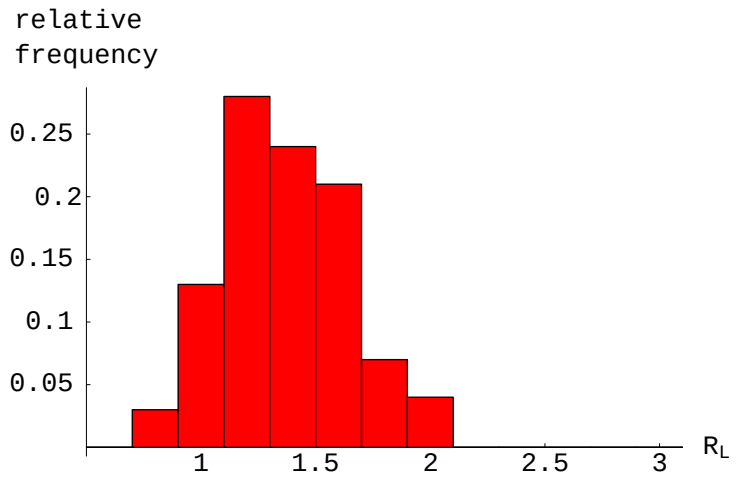


Figure 3.6: Histogram of the ratio R_L of the output of the labeling reaction.

Moreover, we find the empirical 95% confidence interval to be $[0.90, 1.97]$. So, the change still is not significant.

In the next section, we will discover the influence of the fluorescence reaction.

3.4 Fluorescence

We use the numbers of labeled nucleotides as input of the fluorescence module. During Section 1.2.3 we derived the correction factor CF which describes the ratio of output signals if two different lasers and dyes are used to induce fluorescence.

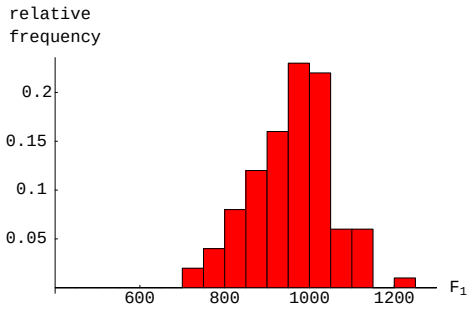
Generally, one should leave the input of types 2 and 4 unchanged whereas the input of types 1 and 3 is multiplied by CF^{-1}. This produces values of all four types which are proportional to the real output and shall be sufficient since we are interested in ratios instead of absolute values. As a results, the output will be dimensionless.

The CF has been determined with the parameter setting from Table 2.18 and Equation (2.45). This yields $CF = 0.89$. Since type 3 has vanished, the correction factor has to be applied to type 1 only.

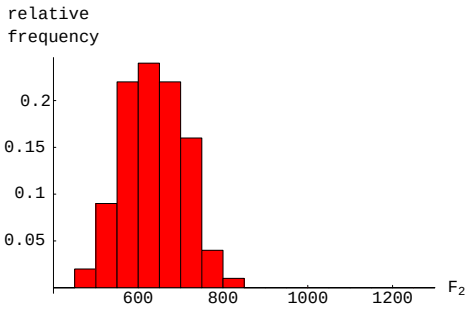
The resulting histograms of the fluorescence values F_i, $i = 1, 2$ of the two remaining types can be found in Figure 3.7.

The respective means and standard deviations are summarized in Table 3.8.

The values of the first type have increased whereas all others stay the same. Thus, the ratio changes as follows. The mean is $\mu(R_F) \approx 1.54$ and the



(a) Histogram of F_1.



(b) Histogram of F_2.

Figure 3.7: Histograms for the fluorescence values of the two remaining types after the fluorescence process at the parameter situation of Table 2.18.

$\mu(F_1)$	$\sigma(F_1)$	$\mu(F_2)$	$\sigma(F_2)$
956.66	95.16	634.12	70.85

Table 3.8: Mean and standard deviation for the fluorescence values of the two remaining target types after the fluorescence reaction.

variance $\sigma^2(R_F) \approx 0.093$. Looking at the ordered list of fluorescence ratios, we find the empirical 95% confidence interval for the ratio to be $[1.01, 2.21]$. This is underlined by the histogram of the fluorescence ratio in Figure 3.8.

The total noise added to the ratio by the previous modules including fluorescence is 54% on average. This implies a total increase of 17% due to fluorescence. Here, for the first time the deviation is significant if looking at the 95% confidence interval.

3.5 Detection

The output of the fluorescence module are values which are no longer particles. But, these values are proportional to the numbers of particles which are some orders of magnitude higher. Since we are only interested in ratios, we will use the fluorescence values as input for the branching process. These values shall be denoted by k_1 and k_2 as previously described in Section 2.2.4.

To directly simulate the branching processes with ten branchings is almost impossible due to computational power. In Section 2.2.4 we have seen that for interesting parameter situations it is possible to approximate the distribution of the number of secondary electrons at the anode $N_{l,k}$ by a normal distribution with mean $\mathbb{E}(N_{l,k})$ and standard deviation $\sqrt{\mathbb{V}ar(N_{l,k})}$.

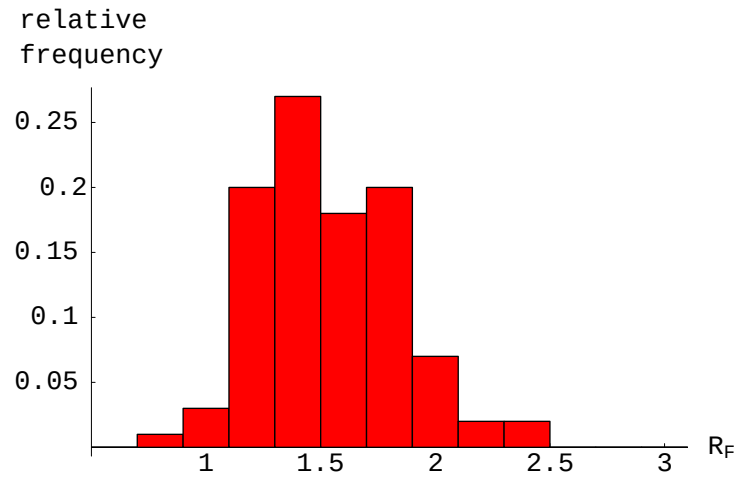


Figure 3.8: Histogram of the ratio R_F of the output of the fluorescence reaction.

The mean and the standard deviation are determined with the help of the p.g.f. of $N_{l,k}$ according to Equations (2.51) and (2.52). For the simulation we will pick a random realizations from the normal distribution for each of the 100 data points according to the parameter setting of Table 3.9.

λ_s	λ_p	l
5	.1	10

Table 3.9: Parameter situation of the detection reaction.

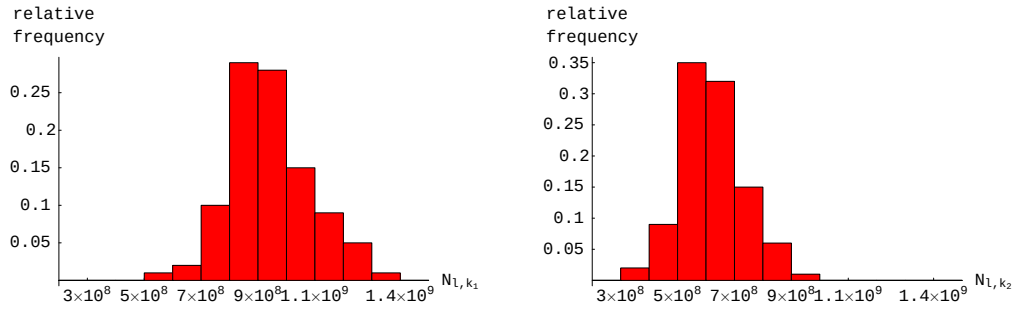
Since the input values for the cross-hybridized targets is zero, we can assume that the respective values in the branching process stay zero. Thus, we only consider types 1 and 2.

Figure 3.9 shows the histograms for the simulations of type 1 and 2. The respective means and standard deviations are summarized in Table 3.10.

$\mu(N_{l,k_1})$	$\sigma(N_{l,k_1})$	$\mu(N_{l,k_2})$	$\sigma(N_{l,k_2})$
9.485×10^8	1.483×10^8	6.210×10^8	1.049×10^8

Table 3.10: Mean and standard deviation for the detection values of types 1 and 2 after the detection reaction. The parameters k_1 and k_2 denote the numbers of photons striking the photocathode as described in Section 2.2.4.

Obviously, the values of the first two types have been multiplied by the branching process. However, the more interesting question is, whether



(a) Histogram of N_{l,k_1}. (b) Histogram of N_{l,k_2}.

Figure 3.9: Histograms for the detection values of types 1 and 2 after the detection process at the parameter situation of Table 3.9. The parameters k_1 and k_2 denote the numbers of photons striking the photocathode as described in Section 2.2.4.

the values of the ratio have been influenced or not. The ratio has mean $\mu(R_N) \approx 1.58$ and variance $\sigma^2(R_N) \approx 0.180$. Thus, the variance has doubled compared to the previous module. This increases the uncertainty attached to the data. The empirical 95% confidence interval underlines this observation. It is approximately $[0.93, 2.59]$. This can also be seen in the histogram of the detection ratio in Figure 3.10.

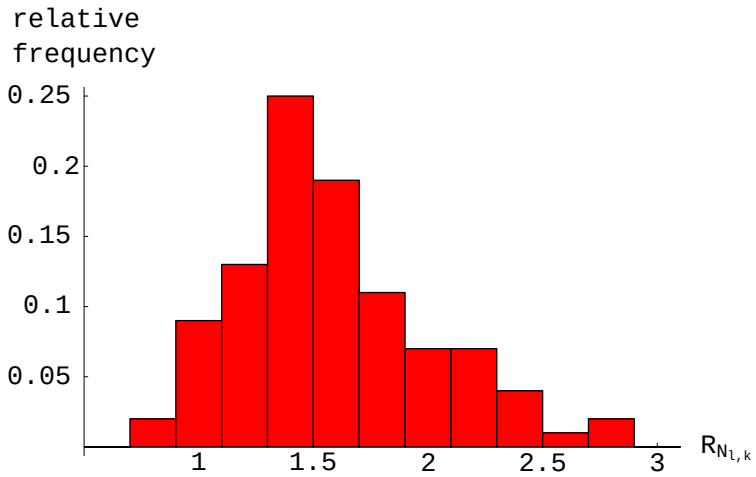


Figure 3.10: Histogram of the ratio R_N of the output of the detection reaction.

So, an extra noise of approximately 4% is added on average. We return to the situation of non-significance due to the increase in variance which influences the confidence interval. As a result, the true ratio could still be 1 or even less.

Looking directly at the histogram, we see that most of the values are

larger than one. For these values, experimenters would infer a change in the expression of the respective gene.

For an overview of the error propagation of the modules see the box plots in Figure 3.11.

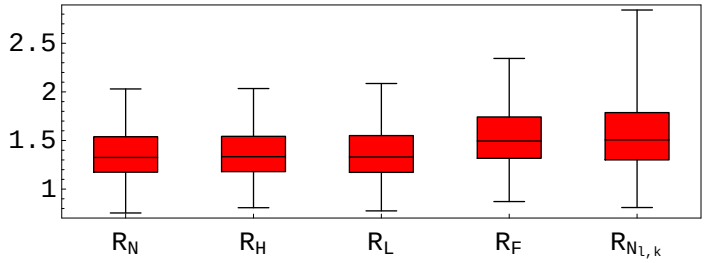


Figure 3.11: Box plots of the ratio values after each module. Lower and upper bounds of the boxes are the 25%- and 75%- quantiles, respectively. Whiskers denote the minimum and maximum of the values.

Obviously, the noise added by the hybridization process hardly changes during the two following modules, washing and labeling. Afterwards, the noise increases fast. This observation underlines the importance of correctly accounting for the shifts in the ratio which are caused by the modules of hybridization, fluorescence and detection. Looking at Figure 3.12 corroborates this result.

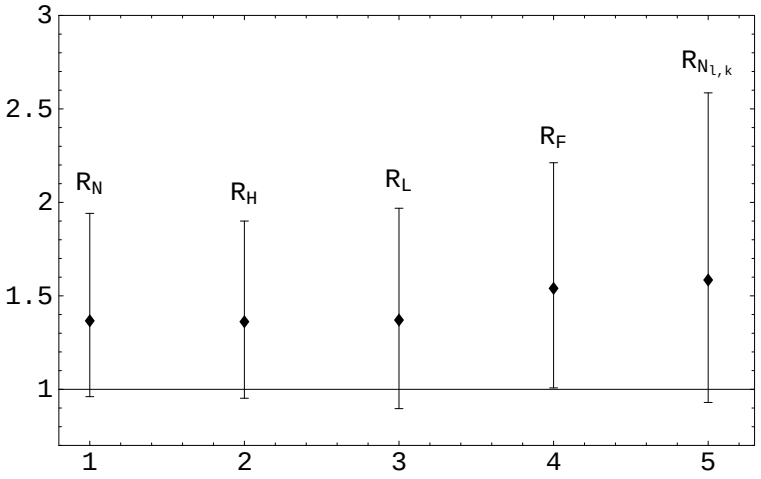


Figure 3.12: Confidence intervals of the ratio values after each module. Upper and lower bounds of the bars represent the bounds of the confidence interval whereas small diamonds on the bars are the means of the respective ratios.

Here, the confidence intervals for the ratios of each module are illustrated. They show a similar behavior as the respective box plots. We also see that

the confidence interval of the fluorescence ratio is the only one which is completely above 1.

3.6 A different starting point

At this point one could ask for the behavior of the compound model in the case of unequal numbers of initial targets. We will briefly look at this case, too, in order to see whether initial fold changes of the ratio might vanish. For this purpose we will ceteris paribus increase the number of initial targets of type two to $T_2 = 20,000$. This yields an initial ratio of $R_T = 2/3$ if accounting for cross-hybridizing targets and a ratio of $R_T = 1/2$ if only considering specific target types 1 and 2. In figures 3.13 and 3.14 we find the box plots and confidence intervals of the five modules.

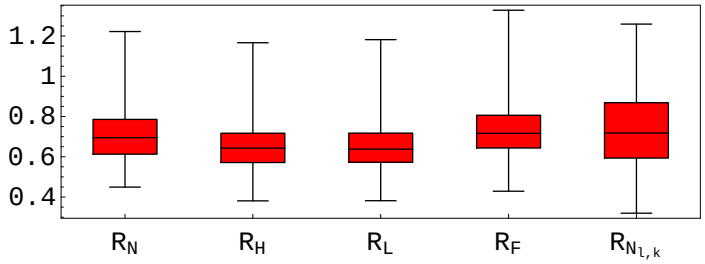


Figure 3.13: Box plots of the ratio values after each module. Lower and upper bounds of the boxes are the 25%- and 75%- quantiles, respectively. Whiskers denote the minimum and maximum of the values.

In both figures we see that the module of hybridization (R_N) adds a lot of noise to the ratio. In Figure 3.14, its confidence interval touches 1. Thus, to a level of significance of 5% a deviation of the ratio from 1 even is not significant. The module of washing (R_H) reduces the noise to a small amount, but, enough to ensure the significance of the deviation by bringing the confidence interval down below 1. During the labeling process (R_L), the noise is hardly enlarged. But the fluorescence (R_F) and the detection ($R_{N_{l,k}}$) modules add a lot of noise and even bias the mean of the ratio. For both modules we find the deviation from one to be insignificant. So, starting with a fold change of two might result in a detected ratio which is close to one. This is equivalent to a constant gene expression comparing the different cell states. Obviously, this is not true.

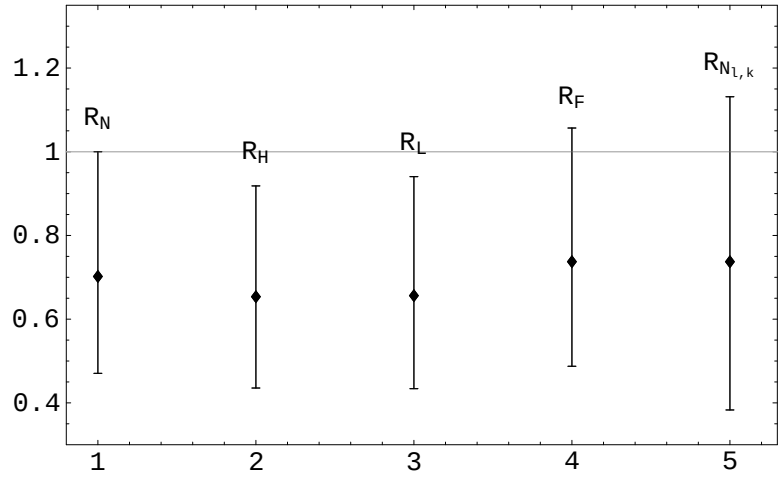


Figure 3.14: Confidence intervals of the ratio values after each module. Upper and lower bounds of the bars represent the bounds of the confidence interval whereas small diamonds on the bars are the means of the respective ratios.

3.7 Résumé

We found that hybridization, fluorescence and detection contributed the most noise in the compound process. We also observed that on average 37%, 17% and 4% are added to the true ratio. We also could observe that the variance of the measured ratios increased within the scope of these three modules. This yields an increasing uncertainty during the inference of initial ratios.

Recall, the initial ratio of target types was 1. By sending the signal through the five modules we have transformed the ratio to an average of 1.58. We found that this increase is not significant to a level of 5%. This also illustrates the amount of noise a microarray experiment can add to a true ratio of initial targets. Assuming a symmetric 95%-confidence region our example also contains a ratio of 2.

As a major result it should be clear, that at least for small x, an $x-$fold in mRNA amounts cannot be inferred precisely by the microarray experiment due to the natural randomness which influences the signal. Inferring initial target ratios is noise afflicted. The noise should be quantified by determining the standard deviation or confidence intervals for the true values of initial target ratios in order to give a measure of reliability of the estimated ratio.

Chapter 4

Discussion and outlook

During this chapter we discuss performance, limitations and weaknesses of each module. We further propose expansions of the modules and suggest directions of future work on our approach.

4.1 Hybridization

The hybridization process was modeled by extending the discrete state, continuous time model proposed in [ReWi]. This was achieved by adding dissociation events (Section 1.1).

This extended version was analyzed in Section 2.1. Firstly, we determined its parameter situation and discussed its complexity. Then we established its stationary distribution for the case of two targets. We soon realized that the computation was limited by too large numbers of probes and targets. Thus, we used a theorem from [Kurtz] to develop a limit which approximates the Markov process by a deterministic process for large numbers. We determined its stationary point and proved its uniqueness. In another approach we approximated the process by a PDE. We proved that the stationary point from the first limit is consistent to the distributional solution of the PDE. In a last step the solution of the limit was compared to the result of simulating the Markov process. We could see that both correspond to each other. All solutions revealed a notable bias added to the initial ratios of targets. This bias has to be accounted for in the analysis of microarray data.

During the analysis of the compound module in Chapter 3, we saw that the hybridization process adds most of the noise to the signal which is eventually detected. For that reason, its importance and preferred position in this work are corroborated.

The hybridization model itself has been stated for an arbitrary number of

target types. But, the investigation of the model has been restricted to two target types with a short preview of the four target case. Thus we suggest to expand the investigation to more complex cases of at least four target types. For this purpose some of the methods we used should be applicable, too. In addition, the choice of dissociation and hybridization parameters as well as the recruitment rate could not be determined realistically. This should also be done in future work.

However, after fixation of the unknown parameters in the two target case, we were able to examine the underlying process to a satisfying amount as described above. With the help of this case we are able to declare the color effect of labels. We could see that the output signal nonlinearly depends on the number of initial targets. This dependency is contradictory to using linear models to describe signal intensities.

In addition, the deterministic limit via Kurtz' theorem provides the right values of the distribution of target types for the long term behavior of the process. The calculation of this limit can be done very fast and enables us to do a simple inference without statistics. To develop the analogous limit for the four target case should be a major goal.

4.2 Residual subprocesses

4.2.1 Reverse transcription

In Section 1.2.1 we modeled the reverse transcription module with a discrete state, continuous time Markov process and reduced it to its embedded Markov chain in discrete time. We simplified the probability distribution of the number of incorporated labeled nucleotides to a binomial distribution by assuming large numbers of initial nucleotides.

This distribution was analyzed in Section 2.2.1. We started the analysis by investigating the dependency of the distribution on the parameters involved. We also applied a perturbation approach to model small differences of the recruitment rates of labeled and unlabeled nucleotides. Here, a Taylor approximation and a test were used to examine the impact of the perturbation. In addition, we proposed the choice of parameter values by minimizing the area under the ROC curve of the test. We could show, that for realistic parameter situations, there will always be a visible dye effect. Using labeled nucleotides, only, might solve this problem. Here, no dye effect is visible. For parameter situations, where the dye effect is visible, we proposed estimators for the numbers of targets fed to the reverse transcription process. In addition, we determined the distributions of the estimators.

Moreover, in Section 2.2.1 we showed, that the reverse transcription process hardly affects the output signal. There are several weaknesses which might distort this result. On the one hand, due to steric problems the incorporation of labeled nucleotides should be forbidden in case of binding sites which are too close to each other. This effect has been neglected in our model. On the other hand, only an approximation for the distribution of labeled nucleotides is used. Incorporating both effects into the model should enable us to get a more detailed insight.

In addition, similar to the hybridization process, crucial parameters like the recruitment rates could not be determined. Finding out their true values should be of major interest.

Also, more input by experimenters is needed. For example, the number of binding sites per target could be determined quite easily. Thus, once we have decided to look at certain target types, it should be possible to specify their parameter situations.

4.2.2 Washing

In Section 1.2.2 we stated a binomial model for the washing process. Here, the probabilities of success were motivated with a Poisson process whose rate depends on the detergent concentration of washing.

Assuming this model, in Section 2.2.2 we could show that the washing process will eliminate falsely hybridized targets if washing at the right detergent concentration. In detail, the model was analyzed by determining the mean and the variance of the particle distribution for interesting parameter situations. We determined these values for increasing detergent intensities and even were able to reproduce observations made by biologists in washing experiments (see [Drob]). It can be seen that only within a small range of washing intensities the correct signal can be achieved. Finding out this range of concentrations should be of major interest to researchers. At this point it is important to mention, that all spots are washed at approximately the same concentration. This narrows the range of the right concentration since there are different targets hybridized to other spots.

This module also makes use of parameters which could not entirely be declared, such as the maximal detergent intensities $\lambda_i^{\max}(c)$ and the number of detergent molecules needed for solution. Also the choice of the type of the intensity function could not be motivated. Thus, a more detailed modeling of the intensity function together with the determination of respective parameter values is recommended. Experiments might help to declare these components of the washing model.

4.2.3 Fluorescence

In Section 1.2.3 we modeled the fluorescence process by dividing it into a laser and a fluorescence model. Both were described deterministically, even though a modeling with random processes is imaginable, too. But, the large numbers of involved particles propose a deterministic modeling. The equations used in this section were found in the literature and neglect some noise sources, such as those shown by all electronic devices (see page 30). The laser light was modeled with an ordinary differential equation from [SaTe] including a noise term whereas the fluorescence intensity was determined by a heuristic equation from [Schwedt].

In Section 2.2.3 we combined the solution of the ordinary differential equation of the laser light intensity with the heuristic equation of the fluorescence intensity. On this basis, we developed the correction factor CF which is a measure for the noise added by the process if two signals are compared.

The investigation in Section 2.2.3 was restricted to the spontaneous emission noise. Still, we were able to show that the fluorescence module might perturb the final signal and with the help of the correction factor we were able to quantify this perturbation.

Some of the parameter values of this module were chosen arbitrarily. Declaring their real values should be done whenever the model is used to quantify the noise added by the fluorescence process in order to normalize the data.

Additionally, as already mentioned, the fluorescence intensity has been determined with the help of a heuristic equation from the literature. A more detailed modeling is suggested at this point. For instance, a model including other noise sources besides spontaneous emission could be subject to future work. The resulting variance would be greater and our variance could provide a lower bound.

4.2.4 Detection

In Section 1.2.4 we used the branching process from [MaTeSa] to describe the signal multiplication during detection. In addition, we described the detection and attached noise sources with heuristic results from [SauWei], [Uiga], [BiSchl] and [SiSu].

The analysis in Section 2.2.4 was restricted to the branching process. We failed to directly determine the probability distribution of the detected signal. But, with the help of the probability generating function we were able to determine mean, variance, skewness and kurtosis of the number of output particles. Afterwards, we used these characteristics to verify the approxi-

mation of the distribution of the detected signal by a normal distribution for interesting parameter situations. In addition, for different parameter values, an example for two signals passing the detection aperture was given. Finally, we investigated the ratio of two signals and derived its probability distribution. This way we were able to approximate the distribution of detected electrons and determine the noise added by this module. However, this model neglects noise sources due to the detection aperture, i.e. those of all electronic devices as described on page 30. Besides that, the distribution of detected electrons at the amperemeter is approximated instead of being determined directly. Incorporating further noise sources and determining the distribution of detected electrons could be of interest to future work.

Furthermore, we looked at the detection with PMTs, only. However, charge coupled device (CCD) cameras are a commonly used method, too. They work very different from PMTs and thus demand a different model.

4.3 Résumé and general discussions

The entire microarray process has been divided into five subprocesses, the modules. This made a modeling and the examination of the models easier. Most of our models are more detailed than the ad-hoc models, which can be found in the literature. Our approach enabled us to separately identify and quantify noise sources of the modules. In addition, by the division into different modules, in future work, we will be able to replace single modules without having to remodel the entire microarray process. E.g., this work is restricted to labeling with fluorescent dyes. But the results of the labeling process, the hybridization model and the washing procedure are applicable to other labeling methods, too.

During the investigation of some of the modules we were restricted by computational power. Parallelizing the computation should improve this problem.

We have seen the general problem of unknown parameter values for most of the modules. Without determining these values we are able to describe tendencies of how the modules will behave under certain assumptions. This is a major result of this work. Determining the values would enable us to give advice to researchers regarding the treatment of their data.

The second major result is that we encountered nonlinearities in the relationship between input and output of the hybridization process. This is contradictory to using linear models which is done by many analyzing methods. We recommend to implement methods accounting for nonlinearities.

Real data are produced by thousands of spots and target types. We ne-

glected this fact by only looking at a single spot and maximal four target types to hybridize to this spot. Certainly, the interactions between spots and between target types affect the dynamic of the reverse transcription, hybridization and washing procedures. This yields a bias which is expected to be relatively small. However, further analysis of these interactions is recommended.

Final conclusion: Microarray experiments are very complex processes. Thus, a very sensible investigation is recommended. We have seen many sources of noise which have to be accounted for. E.g., the difference in hybridization probabilities and dissociation rates, the spontaneous emission or the random multiplication of the signal within the PMT are remarkable. Future work should be concerned with describing the effects of these noise sources and with giving advice to normalization of data. Therefore, we recommend the application of nonlinear models since we have encountered nonlinear relationships between the input and the output of our modules. At this point it is not clear how to practically implement this.

On the other hand, we have also corroborated assumptions made by researchers, e.g. washing at the right detergent concentration dissolves all cross-hybridized targets.

Still, we were able to discover major relationships between each module of the microarray process and the output signal. Using these results should improve the analysis of microarray data by drawing error bounds and developing normalizing methods for detected signal intensities.

Bibliography

[Alhad] Alhadidi, B. and Fakhouri, H.N. and Al Mousa, O.S. (2006). cDNA Microarray Genome Image Processing Using Fixed Spot Position. *American Journal of Applied Science 3, pp. 1730-1734.*

[BiSchl] Bille, J. and Schlegel, W. (2005). Medizinische Physik. *Springer, Berlin, Heidelberg, New York.*

[Britt] Britton, N.F. (2003). Essential mathematical biology. *Springer, London.*

[Bronch] Bronchud, M.H. and Foote, M. and Giaccone, G. and Olopade, O. and Workman, P. (2008). Principles of molecular oncology . *Humana Press.*

[Buhl] Buhler, J. and Ideker, T. and Haynor, D. (2000). Dapple: Improved Techniques for Finding Spots on DNA Microarrays. *UW CSE Technical Report UWTR 2000-08-05.*

[Camp] Campbell, N.A. and Reece, J.B. (2009). Biologie. *Pearson Studium.*

[Chavan] Chavan, P. and Joshi, K. and Patwardhan, B. (2006). DNA Microarrays in Herbal Drug Research. *Evidence-based Complementary and Alternative Medicine 3, No. 4, pp. 447-457.*

[ChNg] Ching, W.-K. and Ng, M.K. (2006). Markov Chains. *Springer, New York.*

[Chou] Chou, C.-C. and Chen, C.-H. and Lee, T.-T. and Peck, K. (2004). Optimization of probe length and the number of probes per gene for optimal microarray analysis of gene expression. *Nucleic Acids Research, Vol. 32 No. 12, Oxford University Press.*

[Chung] Chung, K.L. (1985). Elementare Wahrscheinlichkeitstheorie und stochastische Prozesse. *Springer, Berlin, Heidelberg, New York.*

[ChWa] Chung, K.L. and Walsh J.B. (2005). Markov Processes, Brownian Motion, and Time Symmetry. *Springer, Berlin, Heidelberg, New York.*

[DaViHa] Das, S. and Vikalo, H. and Hassibi, A. (2009). On scaling laws of biosensors: A stochastic approach. *Journal of Applied Physiscs Volume 105, No. 10 pp. 102021-102021-7.*

[Dobr] Dobrowolski, M. (2006). Angewandte Funktionalanalysis. *Springer, Berlin, Heidelberg, New York.*

[Drob] Drobyshev, A.L. and Machka, C. and Horsch, M. and Seltmann, M. and Liebscher, V. and Hrab de Angelis, M. and Beckers, J. (2003). Specificity assessment from fractionation experiments (SAFE): a novel method to evaluate microarray probe specificity based on hybridisation stringencies. *Nucleic Acids Research, Volume 31, No. 2 e1.*

[EGMW] Ernst, L.A. and Gupta, R.K. and Mujumdar, R.B. and Waggoner, A.S. (1989). Cyanine dye labeling reagents for sulfhydryl groups. *Cytometry, Volume 10, pp. 3-10.*

[EKMPW] Erbrecht, R. and König, H. and Martin, K. and Pfeil, W. and Wörstenfeld, W. (1999). Das große Tafelwerk. *Volk und Wissen, Berlin.*

[Faw] Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters 27, pp. 861-874.*

[Fiell] Fieller, E.C. (1932). The distribution of the index in a normal bivariate population. *Biometrika 24, pp. 428-440.*

[FrHH] Franke, J. and Härdle, W and Hafner, C.M. (2004). Einführung in die Statistik der Finanzmärkte. *Springer, Berlin, Heidelberg, New York.*

[Gant] Gantmacher, F.R. (1986). Matrizentheorie. *VEB Deutscher Verlag der Wissenschaften.*

[Gill] Gillespie, D. T. (1977). Exact Stochastic Simulation of Coupled Chemical Reactions. *Journal of Physical Chemistry, Volume 81, No 25, pp. 23402361.*

[Grab] Grabowski, B. (2004). Lexikon der Statistik. *Spektrum, München, Heidelberg.*

[Hahne] Hahne, H. and Mäder, U. and Otto, A. and Bonn, F. and Steil, L. and Bremer, E. and Hecker, M. and Becher, D. (2009). A comprehensive proteomics and transcriptomics analysis of Bacillus subtilis salt stress adaptation. *Journal of Bacteriology.*

[Haßl] Haßler, K. (2006). Single molecule detection and fluorescence correlation spectroscopy on surfaces. *Laboratoire d'Optique Biomedicale, Ecole Poytechnique Federale de Lausanne.*

[HaHiMo] Harris, J.M. and Hirst, J.L. and Mossinghoff, M.J. (2000). Combinatorics and Graph Theory. *Springer, New York, Berlin, Heidelberg.*

[Held] Held, G.A. and Grinstein, G. and Tu, Y. (2006). Relationship between gene expression and observed intensities in DNA microarraysa modeling study. *Nucleic Acids Research Volume 34, No. 9, pp. e70.*

[Heus] Heuser, H. (1991). Lehrbuch der Analysis, Teil 1. *Teubner, Stuttgart.*

[HeusD] Heuser, H. (1995). Gewöhnliche Differentialgleichungen. *Teubner, Stuttgart.*

[Hink] Hinkley, D.V. (1969). On the ratio of two correlated normal randaom variables. *Biometrika 56 (3), pp. 635-639.*

[Hueb] Hübner, G. (2000). Stochastik. *Vieweg, Braunschweig.*

[Jain] Jain, A.N. and Tokuyasu, T.A. and Snijders, A.M. and Segraves, R. and Albertson, D.G. and Pinkel, D. (2002). Fully automatic quantification of microarray image data. *Genome Research, Cold Spring Harbor Laboratory Press, pp. 325-32.*

[JoSm] Jordan, D.W. and Smith, P. (1999). Nonlinear ordinary differential equations - An introduction to dynamical systems. *Oxford University Press, New York.*

[KamII] Kamke, E. (1965). Differentialgleichungen, Lösungsmethoden und Lösungen, Band II. *Akademische Verlagsgesellschaft Geest & Portig K.-G., Leipzig.*

[Kurtz] Kurtz, T.G. (1980). Relationships between stochastic and deterministic population models. *Lecture Notes in Biomathematics 38, Springer, Berlin, pp. 449-467.*

[Kelly] Kelly, F.P. (1994). Reversibility and Stochastic Networks. *Wiley.*

[Klen] Klenke, A. (2008). Wahrscheinlichkeitstheorie. *Springer, Berlin, Heidelberg.*

[Kren] Krengel, U. (2003). Einführung in die Wahrscheinlichkeitstheorie und Statistik. *Vieweg, Wiesbaden.*

[Lako] Lakowicz, J.R. (1999). Principles of flourescence spectroscopy. *Kluwer Academic/Plenum Publishers, New York.*

[MaTeSa] Matsuo, K. and Teich, M.C. and Saleh, B.E.A. (1984). Poisson branching point processes. *Journal of Mathematical Physics, Volume 25, Issue 7, pp. 2174-2185.*

[McLach] McLachlan, G.J. and Do, K.-A. and Ambroise, C. (2004). Analyzing microarray gene expression data. *Wiley.*

[Mitro] Mitrophanov, A. (2004). The spectral gap and perturbation bounds for reversible continuous-time Markov Chains. *Journal of Applied Probabaility 41, pp. 1219-1222.*

[MüNi] Müller, U.R. and Nicolau, D.V. (2005). Microarray technology and its applications. *Spinger, Berlin, Heidelberg, New York.*

[Ochs] Ochsner, S.A and Steffen, D.L. and Hilsenbeck, S.G. and Chen, E.S. and Watkins, C. and McKenna, N.J. (2009). GEMS (Gene Expression Metasignatures), a Web Resource for Querying Meta-analysis of Expression Microarray Datasets: 17β-Estradiol in MCF-7 Cells. *Cancer Research, Volume 69, No. 1, pp. 23-26.*

[Papou] Papoulis, A. (1984). Probability, Random Variables, and Stochastic Processes. *McGraw-Hill, New York.*

[Pawl] Pawley, J. B. (2006). Handbook of biological confocal microscopy. *Springer Science + Business Media.*

[Prietz] Prietz, S. (2003). Expressionsanalysen mit cDNS-Mikroarrays - Aufklärung der Pathogenesemechanismen einer seltenen Augenkrankheit. *Fachbereich Humanmedizin der Freien Universität Berlin.*

[ReWi] Reiner, R. and Wittich, O. (2004). A compound model for hybridization on microarrays. *Preprint, Neuherberg.*

[Renyi] Rényi, A. (1966). Wahrscheinlichkeitsrechnung. *Verlag der Wissenschaften, Berlin.*

[Ramp] Rampal, J.B. (2007). Microarrays: Volume 1: Synthesis Methods. *Humana Press.*

[Rydén] Rydén, P. and Andersson, H. and Landfors, M. and Näslund, L. and Hartmanová, B. and Noppa, L. and Sjösted, A. (2006). Evaluation of microarray data normalization procedures using spike-in experiments. *BMC Bioinformatics.*

[SaTe] Saleh, B.E.A. and Teich, M.C. (1991). Fundamentals of Photonics. *Wiley, New York.*

[SauWei] Sauter, D. and Weinerth, H. (1990). Lexikon Elektronik und Mikroelektronik. *VDI, Düsseldorf.*

[SchmK] Schmidt, K. D. (2009). Maß und Wahrscheinlichkeit. *Springer, Berlin, Heidelberg.*

[Schwedt] Schwedt, G. (1981). Fluorimetrische Analyse. *Verlag Chemie, Weinheim, Deerfield Beach, Basel.*

[Sing] Singer, M. Synthesis of fluorescently labeled cDNA probe for microarrays. Lab of Mitch Singer, Section of Microbiology; Center for Genetics and Development, University of California, Davis. http://micro.mic.ucdavis.edu/singer/protocols/MicroarrayProtocol.pdf

[SiSu] Simon, H. and Suhrmann, R. (1958). Der lichtelektrische Effekt und seine Anwendungen. *Springer, Berlin, Göttingen, Heidelberg.*

[South] Southwick, P.L. and Ernst, L.A. and Tauriello, E.W. and Parker, S.R. and Mujumdar, R.B. and Mujumdar, S.R. and Clever, H.A. and Waggoner, A.S. (1990). Cyanine dye labeling reagents - carboxymethylindocyanine succinimidyl esters. *Cytometry, Volume 11, pp. 418-430.*

[Speed] Speed, T. (2003). Statistical analysis of gene expression microarray data. *Chapman & Hall/CRC, Boca Raton, London, New York, Washington, D.C..*

[Steger] Steger, A. (2002). Diskrete Strukturen. *Springer, Berlin, Heidelberg, New York.*

[ToHo] Tomiuk, S. and Hofman, K. (2001). Microarray probe selection strategies. *Briefings in bioinformatics, Volume 2, No. 4, pp. 329-340.*

[Uiga] Uiga, E. (1995). Optoelectronics. *Prentice-Hall, New Jersey.*

[Ultsch] Ultsch, A. (2003). Is log ratio a good value for identifying differential expressed genes in microarray experiments?. *Databionics Research Laboratory, Universtity of Marburg.*

[Welch] Welch, D.R. (2004). Microarrays bring new insights into understanding of breast cancer metastasis to bone. *Breast Cancer Research, Volume 6, No. 2, pp. 61-64.*

[Wern] Werner, D. (2007). Funktionalanalysis. *Springer, Berlin, Heidelberg, New York.*

[WeAa] Wernick, M.N. and Aarsvold, J. N. (2004). Emission Tomography: the fundamentals of PET and SPECT. *Academic Press.*

[Wigg] Wiggins, S. (2003). Introduction to applied nonlinear dynamical systems and chaos. *Springer, New York.*

[Wink] Winkler, G. (2003). Image analysis, random fields and Markov chain Monte Carlo methods. *Springer, Berlin, Heidelberg, New York.*

[WiNu] http://en.wikipedia.org/wiki/Nucleotide

[Yats] Yatskou, M. and Novikov, E. and Vetter, G. and Muller, A. and Barillot, E. and Vallar, L. and Friederich, E. (2008). Advanced spot quality analysis in two-colour microarray experiments. *BMC Research Notes* 2008.

[YiZha] Yin, G. and Zhang, Q. (1998). Continuous-time Markov chains and applications: a singular perturbation approach. *Springer, New York.*

[ZveBa] Zvelebil, M.J. and Baum, J.O. (2008). Understanding bioinformatics. *Garland Science.*

[ZwCa] Zweig, M.H. and Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry 39, pp. 561-577.*

Appendix A

Solution of Equation (2.16)

The polynomial of third order from Equation (2.15) has the following three solutions:

$$x_1 = \frac{1}{6\lambda\pi_1(\pi_2\gamma_1-\pi_1\gamma_2)}$$
$$(-2\left(-\pi_2\gamma_1^2+\lambda\pi_1^2\left(1+2\alpha_1\right)\gamma_2+\pi_1\gamma_1\left(-\lambda\pi_2-\lambda\pi_2\alpha_1+\lambda\pi_2\alpha_2+\gamma_2\right)\right)+$$
$$\left(22^{1/3}\left(\pi_2^2\gamma_1^4+2\pi_1\pi_2\gamma_1^3\left(\lambda\pi_2+\lambda\pi_2\alpha_1-\lambda\pi_2\alpha_2-\gamma_2\right)+\lambda^2\pi_1^4\left(-1+\alpha_1\right)^2\gamma_2^2+\right.\right.$$
$$\pi_1^2\gamma_1^2\left(\lambda^2\pi_2^2+\lambda^2\pi_2^2\alpha_1^2+\lambda^2\pi_2^2\alpha_2^2+\lambda\pi_2\alpha_1\left(-\lambda\pi_2+\lambda\pi_2\alpha_2-3\gamma_2\right)-4\lambda\pi_2\gamma_2+\right.$$
$$\left.\gamma_2^2+2\lambda\pi_2\alpha_2\left(-\lambda\pi_2+\gamma_2\right)\right)+$$
$$\left.\lambda\pi_1^3\gamma_1\gamma_2\left(-\lambda\pi_2\alpha_1^2+2\left(-\lambda\pi_2+\lambda\pi_2\alpha_2+\gamma_2\right)+\alpha_1\left(3\lambda\pi_2+\lambda\pi_2\alpha_2+\gamma_2\right)\right)\right)\Big)/$$
$$(2\lambda^3\pi_1^3\pi_2^3\gamma_1^3-3\lambda^3\pi_1^3\pi_2^3\alpha_1\gamma_1^3-3\lambda^3\pi_1^3\pi_2^3\alpha_1^2\gamma_1^3+2\lambda^3\pi_1^3\pi_2^3\alpha_1^3\gamma_1^3-6\lambda^3\pi_1^3\pi_2^3\alpha_2\gamma_1^3+$$
$$6\lambda^3\pi_1^3\pi_2^3\alpha_1\alpha_2\gamma_1^3+3\lambda^3\pi_1^3\pi_2^3\alpha_1^2\alpha_2\gamma_1^3+6\lambda^3\pi_1^3\pi_2^3\alpha_2^2\gamma_1^3-3\lambda^3\pi_1^3\pi_2^3\alpha_1\alpha_2^2\gamma_1^3-$$
$$2\lambda^3\pi_1^3\pi_2^3\alpha_2^3\gamma_1^3+6\lambda^2\pi_1^2\pi_2^3\gamma_1^4+3\lambda^2\pi_1^2\pi_2^3\alpha_1\gamma_1^4+6\lambda^2\pi_1^2\pi_2^3\alpha_1^2\gamma_1^4-12\lambda^2\pi_1^2\pi_2^3\alpha_2\gamma_1^4-$$
$$3\lambda^2\pi_1^2\pi_2^3\alpha_1\alpha_2\gamma_1^4+6\lambda^2\pi_1^2\pi_2^3\alpha_2^2\gamma_1^4+6\lambda\pi_1\pi_2^3\gamma_1^5+6\lambda\pi_1\pi_2^3\alpha_1\gamma_1^5-6\lambda\pi_1\pi_2^3\alpha_2\gamma_1^5+$$
$$2\pi_2^3\gamma_1^6-6\lambda^3\pi_1^4\pi_2^2\gamma_1^2\gamma_2+12\lambda^3\pi_1^4\pi_2^2\alpha_1\gamma_1^2\gamma_2-3\lambda^3\pi_1^4\pi_2^2\alpha_1^2\gamma_1^2\gamma_2-3\lambda^3\pi_1^4\pi_2^2\alpha_1^3\gamma_1^2\gamma_2+$$
$$12\lambda^3\pi_1^4\pi_2^2\alpha_2\gamma_1^2\gamma_2-9\lambda^3\pi_1^4\pi_2^2\alpha_1\alpha_2\gamma_1^2\gamma_2-12\lambda^3\pi_1^4\pi_2^2\alpha_1^2\alpha_2\gamma_1^2\gamma_2-6\lambda^3\pi_1^4\pi_2^2\alpha_2^2\gamma_1^2\gamma_2-$$
$$3\lambda^3\pi_1^4\pi_2^2\alpha_1\alpha_2^2\gamma_1^2\gamma_2-18\lambda^2\pi_1^3\pi_2^2\gamma_1^3\gamma_2-3\lambda^2\pi_1^3\pi_2^2\alpha_1\gamma_1^3\gamma_2-12\lambda^2\pi_1^3\pi_2^2\alpha_1^2\gamma_1^3\gamma_2+$$
$$24\lambda^2\pi_1^3\pi_2^2\alpha_2\gamma_1^3\gamma_2+9\lambda^2\pi_1^3\pi_2^2\alpha_1\alpha_2\gamma_1^3\gamma_2-6\lambda^2\pi_1^3\pi_2^2\alpha_2^2\gamma_1^3\gamma_2-18\lambda\pi_1^2\pi_2^2\gamma_1^4\gamma_2-$$
$$15\lambda\pi_1^2\pi_2^2\alpha_1\gamma_1^4\gamma_2+12\lambda\pi_1^2\pi_2^2\alpha_2\gamma_1^4\gamma_2-6\pi_1\pi_2^2\gamma_1^5\gamma_2+6\lambda^3\pi_1^5\pi_2\gamma_1\gamma_2^2-$$
$$15\lambda^3\pi_1^5\pi_2\alpha_1\gamma_1\gamma_2^2+12\lambda^3\pi_1^5\pi_2\alpha_1^2\gamma_1\gamma_2^2-3\lambda^3\pi_1^5\pi_2\alpha_1^3\gamma_1\gamma_2^2-6\lambda^3\pi_1^5\pi_2\alpha_2\gamma_1\gamma_2^2+$$
$$3\lambda^3\pi_1^5\pi_2\alpha_1\alpha_2\gamma_1\gamma_2^2+3\lambda^3\pi_1^5\pi_2\alpha_1^2\alpha_2\gamma_1\gamma_2^2+18\lambda^2\pi_1^4\pi_2\gamma_1^2\gamma_2^2-3\lambda^2\pi_1^4\pi_2\alpha_1\gamma_1^2\gamma_2^2+$$
$$3\lambda^2\pi_1^4\pi_2\alpha_1^2\gamma_1^2\gamma_2^2-12\lambda^2\pi_1^4\pi_2\alpha_2\gamma_1^2\gamma_2^2-6\lambda^2\pi_1^4\pi_2\alpha_1\alpha_2\gamma_1^2\gamma_2^2+18\lambda\pi_1^3\pi_2\gamma_1^3\gamma_2^2+$$
$$12\lambda\pi_1^3\pi_2\alpha_1\gamma_1^3\gamma_2^2-6\lambda\pi_1^3\pi_2\alpha_2\gamma_1^3\gamma_2^2+6\pi_1^2\pi_2\gamma_1^4\gamma_2^2-2\lambda^3\pi_1^6\gamma_2^3+6\lambda^3\pi_1^6\alpha_1\gamma_2^3-$$
$$6\lambda^3\pi_1^6\alpha_1^2\gamma_2^3+2\lambda^3\pi_1^6\alpha_1^3\gamma_2^3-6\lambda^2\pi_1^5\gamma_1\gamma_2^3+3\lambda^2\pi_1^5\alpha_1\gamma_1\gamma_2^3+3\lambda^2\pi_1^5\alpha_1^2\gamma_1\gamma_2^3-$$
$$6\lambda\pi_1^4\gamma_1^2\gamma_2^3-3\lambda\pi_1^4\alpha_1\gamma_1^2\gamma_2^3-2\pi_1^3\gamma_1^3\gamma_2^3+$$
$$\sqrt{\left(4\left(3\lambda\pi_1^2\alpha_1\left(-\pi_2\gamma_1+\pi_1\gamma_2\right)\left(\lambda\pi_1\left(2+\alpha_1\right)\gamma_2+\gamma_1\left(-\lambda\pi_2+\lambda\pi_2\alpha_2+\gamma_2\right)\right)-\right.\right.}$$
$$\left.\left.\left(-\pi_2\gamma_1^2+\lambda\pi_1^2\left(1+2\alpha_1\right)\gamma_2+\pi_1\gamma_1\left(-\lambda\pi_2-\lambda\pi_2\alpha_1+\lambda\pi_2\alpha_2+\gamma_2\right)\right)^2\right)^3+$$
$$\left(2\pi_2^3\gamma_1^6+6\pi_1\pi_2^2\gamma_1^5\left(\lambda\pi_2+\lambda\pi_2\alpha_1-\lambda\pi_2\alpha_2-\gamma_2\right)+2\lambda^3\pi_1^6\left(-1+\alpha_1\right)^3\gamma_2^3-\right.$$
$$\left.3\lambda^2\pi_1^5\left(-1+\alpha_1\right)\gamma_1\gamma_2^2\left(\lambda\pi_2\alpha_1^2-2\left(-\lambda\pi_2+\lambda\pi_2\alpha_2+\gamma_2\right)-\alpha_1\left(3\lambda\pi_2+\lambda\pi_2\alpha_2+\gamma_2\right)\right)+\right.$$

$$
\begin{aligned}
&3\pi_1^2\pi_2\gamma_1^4\left(2\lambda^2\pi_2^2\alpha_1^2+\lambda\pi_2\alpha_1\left(\lambda\pi_2-\lambda\pi_2\alpha_2-5\gamma_2\right)+\right.\\
&\left.2\left(\lambda^2\pi_2^2+\lambda^2\pi_2^2\alpha_2^2-3\lambda\pi_2\gamma_2+\gamma_2^2+2\lambda\pi_2\alpha_2\left(-\lambda\pi_2+\gamma_2\right)\right)\right)-\\
&3\lambda\pi_1^4\gamma_1^2\gamma_2\left(\lambda^2\pi_2^2\alpha_1^3+\lambda\pi_2\alpha_1^2\left(\lambda\pi_2+4\lambda\pi_2\alpha_2-\gamma_2\right)+\right.\\
&\left.2\left(\lambda^2\pi_2^2+\lambda^2\pi_2^2\alpha_2^2-3\lambda\pi_2\gamma_2+\gamma_2^2+2\lambda\pi_2\alpha_2\left(-\lambda\pi_2+\gamma_2\right)\right)+\right.\\
&\left.\alpha_1\left(-4\lambda^2\pi_2^2+\lambda^2\pi_2^2\alpha_2^2+\lambda\pi_2\gamma_2+\gamma_2^2+\lambda\pi_2\alpha_2\left(3\lambda\pi_2+2\gamma_2\right)\right)\right)+\\
&\pi_1^3\gamma_1^3\left(2\lambda^3\pi_2^3\alpha_1^3+3\lambda^2\pi_2^2\alpha_1^2\left(-\lambda\pi_2+\lambda\pi_2\alpha_2-4\gamma_2\right)-\right.\\
&3\lambda\pi_2\alpha_1\left(\lambda^2\pi_2^2+\lambda^2\pi_2^2\alpha_2^2+\lambda\pi_2\gamma_2-4\gamma_2^2-\lambda\pi_2\alpha_2\left(2\lambda\pi_2+3\gamma_2\right)\right)-\\
&\left.2\left(-\lambda^3\pi_2^3+\lambda^3\pi_2^3\alpha_2^3+9\lambda^2\pi_2^2\gamma_2-9\lambda\pi_2\gamma_2^2+\gamma_2^3+3\lambda^2\pi_2^2\alpha_2^2\left(-\lambda\pi_2+\gamma_2\right)+\right.\right.\\
&\left.\left.\left.3\lambda\pi_2\alpha_2\left(\lambda^2\pi_2^2-4\lambda\pi_2\gamma_2+\gamma_2^2\right)\right)\right)^2\right)^{1/3}+\\
&2^{2/3}\\
&(2\lambda^3\pi_1^3\pi_2^3\gamma_1^3-3\lambda^3\pi_1^3\pi_2^3\alpha_1\gamma_1^3-3\lambda^3\pi_1^3\pi_2^3\alpha_1^2\gamma_1^3+2\lambda^3\pi_1^3\pi_2^3\alpha_1^3\gamma_1^3-6\lambda^3\pi_1^3\pi_2^3\alpha_2\gamma_1^3+\\
&6\lambda^3\pi_1^3\pi_2^3\alpha_1\alpha_2\gamma_1^3+3\lambda^3\pi_1^3\pi_2^3\alpha_1^2\alpha_2\gamma_1^3+6\lambda^3\pi_1^3\pi_2^3\alpha_2^2\gamma_1^3-3\lambda^3\pi_1^3\pi_2^3\alpha_1\alpha_2^2\gamma_1^3-\\
&2\lambda^3\pi_1^3\pi_2^3\alpha_2^3\gamma_1^3+6\lambda^2\pi_1^2\pi_2^3\gamma_1^4+3\lambda^2\pi_1^2\pi_2^3\alpha_1\gamma_1^4+6\lambda^2\pi_1^2\pi_2^3\alpha_1^2\gamma_1^4-12\lambda^2\pi_1^2\pi_2^3\alpha_2\gamma_1^4-\\
&3\lambda^2\pi_1^2\pi_2^3\alpha_1\alpha_2\gamma_1^4+6\lambda^2\pi_1^2\pi_2^3\alpha_2^2\gamma_1^4+6\lambda\pi_1\pi_2^3\gamma_1^5+6\lambda\pi_1\pi_2^3\alpha_1\gamma_1^5-6\lambda\pi_1\pi_2^3\alpha_2\gamma_1^5+\\
&2\pi_2^3\gamma_1^6-6\lambda^3\pi_1^4\pi_2^2\gamma_1^2\gamma_2+12\lambda^3\pi_1^4\pi_2^2\alpha_1\gamma_1^2\gamma_2-3\lambda^3\pi_1^4\pi_2^2\alpha_1^2\gamma_1^2\gamma_2-3\lambda^3\pi_1^4\pi_2^2\alpha_1^3\gamma_1^2\gamma_2+\\
&12\lambda^3\pi_1^4\pi_2^2\alpha_2\gamma_1^2\gamma_2-9\lambda^3\pi_1^4\pi_2^2\alpha_1\alpha_2\gamma_1^2\gamma_2-12\lambda^3\pi_1^4\pi_2^2\alpha_1^2\alpha_2\gamma_1^2\gamma_2-6\lambda^3\pi_1^4\pi_2^2\alpha_2^2\gamma_1^2\gamma_2-\\
&3\lambda^3\pi_1^4\pi_2^2\alpha_1\alpha_2^2\gamma_1^2\gamma_2-18\lambda^2\pi_1^3\pi_2^2\gamma_1^3\gamma_2-3\lambda^2\pi_1^3\pi_2^2\alpha_1\gamma_1^3\gamma_2-12\lambda^2\pi_1^3\pi_2^2\alpha_1^2\gamma_1^3\gamma_2+\\
&24\lambda^2\pi_1^3\pi_2^2\alpha_2\gamma_1^3\gamma_2+9\lambda^2\pi_1^3\pi_2^2\alpha_1\alpha_2\gamma_1^3\gamma_2-6\lambda^2\pi_1^3\pi_2^2\alpha_2^2\gamma_1^3\gamma_2-18\lambda\pi_1^2\pi_2^2\gamma_1^4\gamma_2-\\
&15\lambda\pi_1^2\pi_2^2\alpha_1\gamma_1^4\gamma_2+12\lambda\pi_1^2\pi_2^2\alpha_2\gamma_1^4\gamma_2-6\pi_1\pi_2^2\gamma_1^5\gamma_2+6\lambda^3\pi_1^5\pi_2\gamma_1\gamma_2^2-\\
&15\lambda^3\pi_1^5\pi_2\alpha_1\gamma_1\gamma_2^2+12\lambda^3\pi_1^5\pi_2\alpha_1^2\gamma_1\gamma_2^2-3\lambda^3\pi_1^5\pi_2\alpha_1^3\gamma_1\gamma_2^2-6\lambda^3\pi_1^5\pi_2\alpha_2\gamma_1\gamma_2^2+\\
&3\lambda^3\pi_1^5\pi_2\alpha_1\alpha_2\gamma_1\gamma_2^2+3\lambda^3\pi_1^5\pi_2\alpha_1^2\alpha_2\gamma_1\gamma_2^2+18\lambda^2\pi_1^4\pi_2\gamma_1^2\gamma_2^2-3\lambda^2\pi_1^4\pi_2\alpha_1\gamma_1^2\gamma_2^2+\\
&3\lambda^2\pi_1^4\pi_2\alpha_1^2\gamma_1^2\gamma_2^2-12\lambda^2\pi_1^4\pi_2\alpha_2\gamma_1^2\gamma_2^2-6\lambda^2\pi_1^4\pi_2\alpha_1\alpha_2\gamma_1^2\gamma_2^2+18\lambda\pi_1^3\pi_2\gamma_1^3\gamma_2^2+\\
&12\lambda\pi_1^3\pi_2\alpha_1\gamma_1^3\gamma_2^2-6\lambda\pi_1^3\pi_2\alpha_2\gamma_1^3\gamma_2^2+6\pi_1^2\pi_2\gamma_1^4\gamma_2^2-2\lambda^3\pi_1^6\gamma_2^3+6\lambda^3\pi_1^6\alpha_1\gamma_2^3-\\
&6\lambda^3\pi_1^6\alpha_1^2\gamma_2^3+2\lambda^3\pi_1^6\alpha_1^3\gamma_2^3-6\lambda^2\pi_1^5\gamma_1\gamma_2^3+3\lambda^2\pi_1^5\alpha_1\gamma_1\gamma_2^3+3\lambda^2\pi_1^5\alpha_1^2\gamma_1\gamma_2^3-\\
&6\lambda\pi_1^4\gamma_1^2\gamma_2^3-3\lambda\pi_1^4\alpha_1\gamma_1^2\gamma_2^3-2\pi_1^3\gamma_1^3\gamma_2^3+\\
&\sqrt{\left(4\left(3\lambda\pi_1^2\alpha_1\left(-\pi_2\gamma_1+\pi_1\gamma_2\right)\left(\lambda\pi_1\left(2+\alpha_1\right)\gamma_2+\gamma_1\left(-\lambda\pi_2+\lambda\pi_2\alpha_2+\gamma_2\right)\right)-\right.\right.}\\
&\left.\left.\left(-\pi_2\gamma_1^2+\lambda\pi_1^2\left(1+2\alpha_1\right)\gamma_2+\pi_1\gamma_1\left(-\lambda\pi_2-\lambda\pi_2\alpha_1+\lambda\pi_2\alpha_2+\gamma_2\right)\right)^2\right)^3+\right.\\
&\left.\left(2\pi_2^3\gamma_1^6+6\pi_1\pi_2^2\gamma_1^5\left(\lambda\pi_2+\lambda\pi_2\alpha_1-\lambda\pi_2\alpha_2-\gamma_2\right)+2\lambda^3\pi_1^6\left(-1+\alpha_1\right)^3\gamma_2^3-\right.\right.\\
&\left.\left.3\lambda^2\pi_1^5\left(-1+\alpha_1\right)\gamma_1\gamma_2^2\left(\lambda\pi_2\alpha_1^2-2\left(-\lambda\pi_2+\lambda\pi_2\alpha_2+\gamma_2\right)-\alpha_1\left(3\lambda\pi_2+\lambda\pi_2\alpha_2+\gamma_2\right)\right)+\right.\right.\\
&\left.\left.\right.\right.\\
&3\pi_1^2\pi_2\gamma_1^4\left(2\lambda^2\pi_2^2\alpha_1^2+\lambda\pi_2\alpha_1\left(\lambda\pi_2-\lambda\pi_2\alpha_2-5\gamma_2\right)+\right.\\
&\left.2\left(\lambda^2\pi_2^2+\lambda^2\pi_2^2\alpha_2^2-3\lambda\pi_2\gamma_2+\gamma_2^2+2\lambda\pi_2\alpha_2\left(-\lambda\pi_2+\gamma_2\right)\right)\right)-\\
&3\lambda\pi_1^4\gamma_1^2\gamma_2\left(\lambda^2\pi_2^2\alpha_1^3+\lambda\pi_2\alpha_1^2\left(\lambda\pi_2+4\lambda\pi_2\alpha_2-\gamma_2\right)+\right.\\
&\left.2\left(\lambda^2\pi_2^2+\lambda^2\pi_2^2\alpha_2^2-3\lambda\pi_2\gamma_2+\gamma_2^2+2\lambda\pi_2\alpha_2\left(-\lambda\pi_2+\gamma_2\right)\right)+\right.\\
&\left.\alpha_1\left(-4\lambda^2\pi_2^2+\lambda^2\pi_2^2\alpha_2^2+\lambda\pi_2\gamma_2+\gamma_2^2+\lambda\pi_2\alpha_2\left(3\lambda\pi_2+2\gamma_2\right)\right)\right)+\\
&\pi_1^3\gamma_1^3\left(2\lambda^3\pi_2^3\alpha_1^3+3\lambda^2\pi_2^2\alpha_1^2\left(-\lambda\pi_2+\lambda\pi_2\alpha_2-4\gamma_2\right)-\right.\\
&3\lambda\pi_2\alpha_1\left(\lambda^2\pi_2^2+\lambda^2\pi_2^2\alpha_2^2+\lambda\pi_2\gamma_2-4\gamma_2^2-\lambda\pi_2\alpha_2\left(2\lambda\pi_2+3\gamma_2\right)\right)-\\
&2\left(-\lambda^3\pi_2^3+\lambda^3\pi_2^3\alpha_2^3+9\lambda^2\pi_2^2\gamma_2-9\lambda\pi_2\gamma_2^2+\gamma_2^3+3\lambda^2\pi_2^2\alpha_2^2\left(-\lambda\pi_2+\gamma_2\right)+\right.
\end{aligned}
$$

$$3\lambda\pi_2\alpha_2\left(\lambda^2\pi_2^2-4\lambda\pi_2\gamma_2+\gamma_2^2\right)\right)\right)^2\Big)\Big)^{1/3}\Big),$$

$$x_2=\frac{1}{12\lambda\pi_1(\pi_2\gamma_1-\pi_1\gamma_2)}$$
$$(-4\left(-\pi_2\gamma_1^2+\lambda\pi_1^2\left(1+2\alpha_1\right)\gamma_2+\pi_1\gamma_1\left(-\lambda\pi_2-\lambda\pi_2\alpha_1+\lambda\pi_2\alpha_2+\gamma_2\right)\right)-$$
$$\left(2i2^{1/3}\left(-i+\sqrt{3}\right)\left(\pi_2^2\gamma_1^4+2\pi_1\pi_2\gamma_1^3\left(\lambda\pi_2+\lambda\pi_2\alpha_1-\lambda\pi_2\alpha_2-\gamma_2\right)+\lambda^2\pi_1^4\left(-1+\alpha_1\right)^2\gamma_2^2+\right.\right.$$
$$\pi_1^2\gamma_1^2\left(\lambda^2\pi_2^2+\lambda^2\pi_2^2\alpha_1^2+\lambda^2\pi_2^2\alpha_2^2+\lambda\pi_2\alpha_1\left(-\lambda\pi_2+\lambda\pi_2\alpha_2-3\gamma_2\right)-4\lambda\pi_2\gamma_2+\right.$$
$$\gamma_2^2+2\lambda\pi_2\alpha_2\left(-\lambda\pi_2+\gamma_2\right))+$$
$$\left.\lambda\pi_1^3\gamma_1\gamma_2\left(-\lambda\pi_2\alpha_1^2+2\left(-\lambda\pi_2+\lambda\pi_2\alpha_2+\gamma_2\right)+\alpha_1\left(3\lambda\pi_2+\lambda\pi_2\alpha_2+\gamma_2\right)\right)\right)\Big)/$$
$$(2\lambda^3\pi_1^3\pi_2^3\gamma_1^3-3\lambda^3\pi_1^3\pi_2^3\alpha_1\gamma_1^3-3\lambda^3\pi_1^3\pi_2^3\alpha_1^2\gamma_1^3+2\lambda^3\pi_1^3\pi_2^3\alpha_1^3\gamma_1^3-6\lambda^3\pi_1^3\pi_2^3\alpha_2\gamma_1^3+$$
$$6\lambda^3\pi_1^3\pi_2^3\alpha_1\alpha_2\gamma_1^3+3\lambda^3\pi_1^3\pi_2^3\alpha_1^2\alpha_2\gamma_1^3+6\lambda^3\pi_1^3\pi_2^3\alpha_2^2\gamma_1^3-3\lambda^3\pi_1^3\pi_2^3\alpha_1\alpha_2^2\gamma_1^3-$$
$$2\lambda^3\pi_1^3\pi_2^3\alpha_2^3\gamma_1^3+6\lambda^2\pi_1^2\pi_2^3\gamma_1^4+3\lambda^2\pi_1^2\pi_2^3\alpha_1\gamma_1^4+6\lambda^2\pi_1^2\pi_2^3\alpha_1^2\gamma_1^4-12\lambda^2\pi_1^2\pi_2^3\alpha_2\gamma_1^4-$$
$$3\lambda^2\pi_1^2\pi_2^3\alpha_1\alpha_2\gamma_1^4+6\lambda^2\pi_1^2\pi_2^3\alpha_2^2\gamma_1^4+6\lambda\pi_1\pi_2^3\gamma_1^5+6\lambda\pi_1\pi_2^3\alpha_1\gamma_1^5-6\lambda\pi_1\pi_2^3\alpha_2\gamma_1^5+$$
$$2\pi_2^3\gamma_1^6-6\lambda^3\pi_1^4\pi_2^2\gamma_1^2\gamma_2+12\lambda^3\pi_1^4\pi_2^2\alpha_1\gamma_1^2\gamma_2-3\lambda^3\pi_1^4\pi_2^2\alpha_1^2\gamma_1^2\gamma_2-3\lambda^3\pi_1^4\pi_2^2\alpha_1^3\gamma_1^2\gamma_2+$$
$$12\lambda^3\pi_1^4\pi_2^2\alpha_2\gamma_1^2\gamma_2-9\lambda^3\pi_1^4\pi_2^2\alpha_1\alpha_2\gamma_1^2\gamma_2-12\lambda^3\pi_1^4\pi_2^2\alpha_1^2\alpha_2\gamma_1^2\gamma_2-6\lambda^3\pi_1^4\pi_2^2\alpha_2^2\gamma_1^2\gamma_2-$$
$$3\lambda^3\pi_1^4\pi_2^2\alpha_1\alpha_2^2\gamma_1^2\gamma_2-18\lambda^2\pi_1^3\pi_2^2\gamma_1^3\gamma_2-3\lambda^2\pi_1^3\pi_2^2\alpha_1\gamma_1^3\gamma_2-12\lambda^2\pi_1^3\pi_2^2\alpha_1^2\gamma_1^3\gamma_2+$$
$$24\lambda^2\pi_1^3\pi_2^2\alpha_2\gamma_1^3\gamma_2+9\lambda^2\pi_1^3\pi_2^2\alpha_1\alpha_2\gamma_1^3\gamma_2-6\lambda^2\pi_1^3\pi_2^2\alpha_2^2\gamma_1^3\gamma_2-18\lambda\pi_1^2\pi_2^2\gamma_1^4\gamma_2-$$
$$15\lambda\pi_1^2\pi_2^2\alpha_1\gamma_1^4\gamma_2+12\lambda\pi_1^2\pi_2^2\alpha_2\gamma_1^4\gamma_2-6\pi_1\pi_2^2\gamma_1^5\gamma_2+6\lambda^3\pi_1^5\pi_2\gamma_1\gamma_2^2-$$
$$15\lambda^3\pi_1^5\pi_2\alpha_1\gamma_1\gamma_2^2+12\lambda^3\pi_1^5\pi_2\alpha_1^2\gamma_1\gamma_2^2-3\lambda^3\pi_1^5\pi_2\alpha_1^3\gamma_1\gamma_2^2-6\lambda^3\pi_1^5\pi_2\alpha_2\gamma_1\gamma_2^2+$$
$$3\lambda^3\pi_1^5\pi_2\alpha_1\alpha_2\gamma_1\gamma_2^2+3\lambda^3\pi_1^5\pi_2\alpha_1^2\alpha_2\gamma_1\gamma_2^2+18\lambda^2\pi_1^4\pi_2\gamma_1^2\gamma_2^2-3\lambda^2\pi_1^4\pi_2\alpha_1\gamma_1^2\gamma_2^2+$$
$$3\lambda^2\pi_1^4\pi_2\alpha_1^2\gamma_1^2\gamma_2^2-12\lambda^2\pi_1^4\pi_2\alpha_2\gamma_1^2\gamma_2^2-6\lambda^2\pi_1^4\pi_2\alpha_1\alpha_2\gamma_1^2\gamma_2^2+18\lambda\pi_1^3\pi_2\gamma_1^3\gamma_2^2+$$
$$12\lambda\pi_1^3\pi_2\alpha_1\gamma_1^3\gamma_2^2-6\lambda\pi_1^3\pi_2\alpha_2\gamma_1^3\gamma_2^2+6\pi_1^2\pi_2\gamma_1^4\gamma_2^2-2\lambda^3\pi_1^6\gamma_2^3+6\lambda^3\pi_1^6\alpha_1\gamma_2^3-$$
$$6\lambda^3\pi_1^6\alpha_1^2\gamma_2^3+2\lambda^3\pi_1^6\alpha_1^3\gamma_2^3-6\lambda^2\pi_1^5\gamma_1\gamma_2^3+3\lambda^2\pi_1^5\alpha_1\gamma_1\gamma_2^3+3\lambda^2\pi_1^5\alpha_1^2\gamma_1\gamma_2^3-$$
$$6\lambda\pi_1^4\gamma_1^2\gamma_2^3-3\lambda\pi_1^4\alpha_1\gamma_1^2\gamma_2^3-2\pi_1^3\gamma_1^3\gamma_2^3+$$
$$\sqrt{\left(4\left(3\lambda\pi_1^2\alpha_1\left(-\pi_2\gamma_1+\pi_1\gamma_2\right)\left(\lambda\pi_1\left(2+\alpha_1\right)\gamma_2+\gamma_1\left(-\lambda\pi_2+\lambda\pi_2\alpha_2+\gamma_2\right)\right)-\right.\right.}$$
$$\left.\left.\left(-\pi_2\gamma_1^2+\lambda\pi_1^2\left(1+2\alpha_1\right)\gamma_2+\pi_1\gamma_1\left(-\lambda\pi_2-\lambda\pi_2\alpha_1+\lambda\pi_2\alpha_2+\gamma_2\right)\right)^2\right)^3+\right.$$
$$\left(2\pi_2^3\gamma_1^6+6\pi_1\pi_2^2\gamma_1^5\left(\lambda\pi_2+\lambda\pi_2\alpha_1-\lambda\pi_2\alpha_2-\gamma_2\right)+2\lambda^3\pi_1^6\left(-1+\alpha_1\right)^3\gamma_2^3-\right.$$
$$3\lambda^2\pi_1^5\left(-1+\alpha_1\right)\gamma_1\gamma_2^2\left(\lambda\pi_2\alpha_1^2-2\left(-\lambda\pi_2+\lambda\pi_2\alpha_2+\gamma_2\right)-\alpha_1\left(3\lambda\pi_2+\lambda\pi_2\alpha_2+\gamma_2\right)\right)+$$
$$3\pi_1^2\pi_2\gamma_1^4\left(2\lambda^2\pi_2^2\alpha_1^2+\lambda\pi_2\alpha_1\left(\lambda\pi_2-\lambda\pi_2\alpha_2-5\gamma_2\right)+\right.$$
$$\left.2\left(\lambda^2\pi_2^2+\lambda^2\pi_2^2\alpha_2^2-3\lambda\pi_2\gamma_2+\gamma_2^2+2\lambda\pi_2\alpha_2\left(-\lambda\pi_2+\gamma_2\right)\right)\right)-$$
$$3\lambda\pi_1^4\gamma_1^2\gamma_2\left(\lambda^2\pi_2^2\alpha_1^3+\lambda\pi_2\alpha_1^2\left(\lambda\pi_2+4\lambda\pi_2\alpha_2-\gamma_2\right)+\right.$$
$$2\left(\lambda^2\pi_2^2+\lambda^2\pi_2^2\alpha_2^2-3\lambda\pi_2\gamma_2+\gamma_2^2+2\lambda\pi_2\alpha_2\left(-\lambda\pi_2+\gamma_2\right)\right)+$$
$$\left.\alpha_1\left(-4\lambda^2\pi_2^2+\lambda^2\pi_2^2\alpha_2^2+\lambda\pi_2\gamma_2+\gamma_2^2+\lambda\pi_2\alpha_2\left(3\lambda\pi_2+2\gamma_2\right)\right)\right)+$$
$$\pi_1^3\gamma_1^3\left(2\lambda^3\pi_2^3\alpha_1^3+3\lambda^2\pi_2^2\alpha_1^2\left(-\lambda\pi_2+\lambda\pi_2\alpha_2-4\gamma_2\right)-\right.$$
$$3\lambda\pi_2\alpha_1\left(\lambda^2\pi_2^2+\lambda^2\pi_2^2\alpha_2^2+\lambda\pi_2\gamma_2-4\gamma_2^2-\lambda\pi_2\alpha_2\left(2\lambda\pi_2+3\gamma_2\right)\right)-$$
$$2\left(-\lambda^3\pi_2^3+\lambda^3\pi_2^3\alpha_2^3+9\lambda^2\pi_2^2\gamma_2-9\lambda\pi_2\gamma_2^2+\gamma_2^3+3\lambda^2\pi_2^2\alpha_2^2\left(-\lambda\pi_2+\gamma_2\right)+\right.$$
$$\left.\left.\left.3\lambda\pi_2\alpha_2\left(\lambda^2\pi_2^2-4\lambda\pi_2\gamma_2+\gamma_2^2\right)\right)\right)^2\Big)\Big)^{1/3}+$$
$$i2^{2/3}\left(i+\sqrt{3}\right)$$
$$(2\lambda^3\pi_1^3\pi_2^3\gamma_1^3-3\lambda^3\pi_1^3\pi_2^3\alpha_1\gamma_1^3-3\lambda^3\pi_1^3\pi_2^3\alpha_1^2\gamma_1^3+2\lambda^3\pi_1^3\pi_2^3\alpha_1^3\gamma_1^3-6\lambda^3\pi_1^3\pi_2^3\alpha_2\gamma_1^3+$$

$$6\lambda^3\pi_1^3\pi_2^3\alpha_1\alpha_2\gamma_1^3+3\lambda^3\pi_1^3\pi_2^3\alpha_1^2\alpha_2\gamma_1^3+6\lambda^3\pi_1^3\pi_2^3\alpha_2^2\gamma_1^3-3\lambda^3\pi_1^3\pi_2^3\alpha_1\alpha_2^2\gamma_1^3-$$
$$2\lambda^3\pi_1^3\pi_2^3\alpha_2^3\gamma_1^3+6\lambda^2\pi_1^2\pi_2^3\gamma_1^4+3\lambda^2\pi_1^2\pi_2^3\alpha_1\gamma_1^4+6\lambda^2\pi_1^2\pi_2^3\alpha_1^2\gamma_1^4-12\lambda^2\pi_1^2\pi_2^3\alpha_2\gamma_1^4-$$
$$3\lambda^2\pi_1^2\pi_2^3\alpha_1\alpha_2\gamma_1^4+6\lambda^2\pi_1^2\pi_2^3\alpha_2^2\gamma_1^4+6\lambda\pi_1\pi_2^3\gamma_1^5+6\lambda\pi_1\pi_2^3\alpha_1\gamma_1^5-6\lambda\pi_1\pi_2^3\alpha_2\gamma_1^5+$$
$$2\pi_2^3\gamma_1^6-6\lambda^3\pi_1^4\pi_2^2\gamma_1^2\gamma_2+12\lambda^3\pi_1^4\pi_2^2\alpha_1\gamma_1^2\gamma_2-3\lambda^3\pi_1^4\pi_2^2\alpha_1^2\gamma_1^2\gamma_2-3\lambda^3\pi_1^4\pi_2^2\alpha_1^3\gamma_1^2\gamma_2+$$
$$12\lambda^3\pi_1^4\pi_2^2\alpha_2\gamma_1^2\gamma_2-9\lambda^3\pi_1^4\pi_2^2\alpha_1\alpha_2\gamma_1^2\gamma_2-12\lambda^3\pi_1^4\pi_2^2\alpha_1^2\alpha_2\gamma_1^2\gamma_2-6\lambda^3\pi_1^4\pi_2^2\alpha_2^2\gamma_1^2\gamma_2-$$
$$3\lambda^3\pi_1^4\pi_2^2\alpha_1\alpha_2^2\gamma_1^2\gamma_2-18\lambda^2\pi_1^3\pi_2^2\gamma_1^3\gamma_2-3\lambda^2\pi_1^3\pi_2^2\alpha_1\gamma_1^3\gamma_2-12\lambda^2\pi_1^3\pi_2^2\alpha_1^2\gamma_1^3\gamma_2+$$
$$24\lambda^2\pi_1^3\pi_2^2\alpha_2\gamma_1^3\gamma_2+9\lambda^2\pi_1^3\pi_2^2\alpha_1\alpha_2\gamma_1^3\gamma_2-6\lambda^2\pi_1^3\pi_2^2\alpha_2^2\gamma_1^3\gamma_2-18\lambda\pi_1^2\pi_2^2\gamma_1^4\gamma_2-$$
$$15\lambda\pi_1^2\pi_2^2\alpha_1\gamma_1^4\gamma_2+12\lambda\pi_1^2\pi_2^2\alpha_2\gamma_1^4\gamma_2-6\pi_1\pi_2^2\gamma_1^5\gamma_2+6\lambda^3\pi_1^5\pi_2\gamma_1\gamma_2^2-$$
$$15\lambda^3\pi_1^5\pi_2\alpha_1\gamma_1\gamma_2^2+12\lambda^3\pi_1^5\pi_2\alpha_1^2\gamma_1\gamma_2^2-3\lambda^3\pi_1^5\pi_2\alpha_1^3\gamma_1\gamma_2^2-6\lambda^3\pi_1^5\pi_2\alpha_2\gamma_1\gamma_2^2+$$
$$3\lambda^3\pi_1^5\pi_2\alpha_1\alpha_2\gamma_1\gamma_2^2+3\lambda^3\pi_1^5\pi_2\alpha_1^2\alpha_2\gamma_1\gamma_2^2+18\lambda^2\pi_1^4\pi_2\gamma_1^2\gamma_2^2-3\lambda^2\pi_1^4\pi_2\alpha_1\gamma_1^2\gamma_2^2+$$
$$3\lambda^2\pi_1^4\pi_2\alpha_1^2\gamma_1^2\gamma_2^2-12\lambda^2\pi_1^4\pi_2\alpha_2\gamma_1^2\gamma_2^2-6\lambda^2\pi_1^4\pi_2\alpha_1\alpha_2\gamma_1^2\gamma_2^2+18\lambda\pi_1^3\pi_2\gamma_1^3\gamma_2^2+$$
$$12\lambda\pi_1^3\pi_2\alpha_1\gamma_1^3\gamma_2^2-6\lambda\pi_1^3\pi_2\alpha_2\gamma_1^3\gamma_2^2+6\pi_1^2\pi_2\gamma_1^4\gamma_2^2-2\lambda^3\pi_1^6\gamma_2^3+6\lambda^3\pi_1^6\alpha_1\gamma_2^3-$$
$$6\lambda^3\pi_1^6\alpha_1^2\gamma_2^3+2\lambda^3\pi_1^6\alpha_1^3\gamma_2^3-6\lambda^2\pi_1^5\gamma_1\gamma_2^3+3\lambda^2\pi_1^5\alpha_1\gamma_1\gamma_2^3+3\lambda^2\pi_1^5\alpha_1^2\gamma_1\gamma_2^3-$$
$$6\lambda\pi_1^4\gamma_1^2\gamma_2^3-3\lambda\pi_1^4\alpha_1\gamma_1^2\gamma_2^3-2\pi_1^3\gamma_1^3\gamma_2^3+$$
$$\sqrt{\left(4\left(3\lambda\pi_1^2\alpha_1\left(-\pi_2\gamma_1+\pi_1\gamma_2\right)\left(\lambda\pi_1\left(2+\alpha_1\right)\gamma_2+\gamma_1\left(-\lambda\pi_2+\lambda\pi_2\alpha_2+\gamma_2\right)\right)-\right.\right.}$$
$$\left.\left.\left(-\pi_2\gamma_1^2+\lambda\pi_1^2\left(1+2\alpha_1\right)\gamma_2+\pi_1\gamma_1\left(-\lambda\pi_2-\lambda\pi_2\alpha_1+\lambda\pi_2\alpha_2+\gamma_2\right)\right)^2\right)^3+$$
$$\left(2\pi_2^3\gamma_1^6+6\pi_1\pi_2^2\gamma_1^5\left(\lambda\pi_2+\lambda\pi_2\alpha_1-\lambda\pi_2\alpha_2-\gamma_2\right)+2\lambda^3\pi_1^6\left(-1+\alpha_1\right)^3\gamma_2^3-\right.$$
$$3\lambda^2\pi_1^5\left(-1+\alpha_1\right)\gamma_1\gamma_2^2\left(\lambda\pi_2\alpha_1^2-2\left(-\lambda\pi_2+\lambda\pi_2\alpha_2+\gamma_2\right)-\alpha_1\left(3\lambda\pi_2+\lambda\pi_2\alpha_2+\gamma_2\right)\right)+$$
$$3\pi_1^2\pi_2\gamma_1^4\left(2\lambda^2\pi_2^2\alpha_1^2+\lambda\pi_2\alpha_1\left(\lambda\pi_2-\lambda\pi_2\alpha_2-5\gamma_2\right)+\right.$$
$$\left.2\left(\lambda^2\pi_2^2+\lambda^2\pi_2^2\alpha_2^2-3\lambda\pi_2\gamma_2+\gamma_2^2+2\lambda\pi_2\alpha_2\left(-\lambda\pi_2+\gamma_2\right)\right)\right)-$$
$$3\lambda\pi_1^4\gamma_1^2\gamma_2\left(\lambda^2\pi_2^2\alpha_1^3+\lambda\pi_2\alpha_1^2\left(\lambda\pi_2+4\lambda\pi_2\alpha_2-\gamma_2\right)+\right.$$
$$2\left(\lambda^2\pi_2^2+\lambda^2\pi_2^2\alpha_2^2-3\lambda\pi_2\gamma_2+\gamma_2^2+2\lambda\pi_2\alpha_2\left(-\lambda\pi_2+\gamma_2\right)\right)+$$
$$\left.\alpha_1\left(-4\lambda^2\pi_2^2+\lambda^2\pi_2^2\alpha_2^2+\lambda\pi_2\gamma_2+\gamma_2^2+\lambda\pi_2\alpha_2\left(3\lambda\pi_2+2\gamma_2\right)\right)\right)+$$
$$\pi_1^3\gamma_1^3\left(2\lambda^3\pi_2^3\alpha_1^3+3\lambda^2\pi_2^2\alpha_1^2\left(-\lambda\pi_2+\lambda\pi_2\alpha_2-4\gamma_2\right)-\right.$$
$$3\lambda\pi_2\alpha_1\left(\lambda^2\pi_2^2+\lambda^2\pi_2^2\alpha_2^2+\lambda\pi_2\gamma_2-4\gamma_2^2-\lambda\pi_2\alpha_2\left(2\lambda\pi_2+3\gamma_2\right)\right)-$$
$$2\left(-\lambda^3\pi_2^3+\lambda^3\pi_2^3\alpha_2^3+9\lambda^2\pi_2^2\gamma_2-9\lambda\pi_2\gamma_2^2+\gamma_2^3+3\lambda^2\pi_2^2\alpha_2^2\left(-\lambda\pi_2+\gamma_2\right)+\right.$$
$$\left.\left.\left.3\lambda\pi_2\alpha_2\left(\lambda^2\pi_2^2-4\lambda\pi_2\gamma_2+\gamma_2^2\right)\right)\right)^2\right)\Big)^{1/3}\Bigg) \text{ and}$$

$$x_3=-\tfrac{1}{12\lambda\pi_1\left(\pi_2\gamma_1-\pi_1\gamma_2\right)}$$
$$\left(4\left(-\pi_2\gamma_1^2+\lambda\pi_1^2\left(1+2\alpha_1\right)\gamma_2+\pi_1\gamma_1\left(-\lambda\pi_2-\lambda\pi_2\alpha_1+\lambda\pi_2\alpha_2+\gamma_2\right)\right)-\right.$$
$$\left(2i2^{1/3}\left(i+\sqrt{3}\right)\left(\pi_2^2\gamma_1^4+2\pi_1\pi_2\gamma_1^3\left(\lambda\pi_2+\lambda\pi_2\alpha_1-\lambda\pi_2\alpha_2-\gamma_2\right)+\lambda^2\pi_1^4\left(-1+\alpha_1\right)^2\gamma_2^2+\right.\right.$$
$$\pi_1^2\gamma_1^2\left(\lambda^2\pi_2^2+\lambda^2\pi_2^2\alpha_1^2+\lambda^2\pi_2^2\alpha_2^2+\lambda\pi_2\alpha_1\left(-\lambda\pi_2+\lambda\pi_2\alpha_2-3\gamma_2\right)-4\lambda\pi_2\gamma_2+\right.$$
$$\left.\left.\gamma_2^2+2\lambda\pi_2\alpha_2\left(-\lambda\pi_2+\gamma_2\right)\right)+\right.$$
$$\left.\lambda\pi_1^3\gamma_1\gamma_2\left(-\lambda\pi_2\alpha_1^2+2\left(-\lambda\pi_2+\lambda\pi_2\alpha_2+\gamma_2\right)+\alpha_1\left(3\lambda\pi_2+\lambda\pi_2\alpha_2+\gamma_2\right)\right)\right)\Big)/$$
$$\left(2\lambda^3\pi_1^3\pi_2^3\gamma_1^3-3\lambda^3\pi_1^3\pi_2^3\alpha_1\gamma_1^3-3\lambda^3\pi_1^3\pi_2^3\alpha_1^2\gamma_1^3+2\lambda^3\pi_1^3\pi_2^3\alpha_1^3\gamma_1^3-6\lambda^3\pi_1^3\pi_2^3\alpha_2\gamma_1^3+\right.$$
$$6\lambda^3\pi_1^3\pi_2^3\alpha_1\alpha_2\gamma_1^3+3\lambda^3\pi_1^3\pi_2^3\alpha_1^2\alpha_2\gamma_1^3+6\lambda^3\pi_1^3\pi_2^3\alpha_2^2\gamma_1^3-3\lambda^3\pi_1^3\pi_2^3\alpha_1\alpha_2^2\gamma_1^3-$$
$$2\lambda^3\pi_1^3\pi_2^3\alpha_2^3\gamma_1^3+6\lambda^2\pi_1^2\pi_2^3\gamma_1^4+3\lambda^2\pi_1^2\pi_2^3\alpha_1\gamma_1^4+6\lambda^2\pi_1^2\pi_2^3\alpha_1^2\gamma_1^4-12\lambda^2\pi_1^2\pi_2^3\alpha_2\gamma_1^4-$$
$$3\lambda^2\pi_1^2\pi_2^3\alpha_1\alpha_2\gamma_1^4+6\lambda^2\pi_1^2\pi_2^3\alpha_2^2\gamma_1^4+6\lambda\pi_1\pi_2^3\gamma_1^5+6\lambda\pi_1\pi_2^3\alpha_1\gamma_1^5-6\lambda\pi_1\pi_2^3\alpha_2\gamma_1^5+$$
$$2\pi_2^3\gamma_1^6-6\lambda^3\pi_1^4\pi_2^2\gamma_1^2\gamma_2+12\lambda^3\pi_1^4\pi_2^2\alpha_1\gamma_1^2\gamma_2-3\lambda^3\pi_1^4\pi_2^2\alpha_1^2\gamma_1^2\gamma_2-3\lambda^3\pi_1^4\pi_2^2\alpha_1^3\gamma_1^2\gamma_2+$$

$$12\lambda^3\pi_1^4\pi_2^2\alpha_2\gamma_1^2\gamma_2 - 9\lambda^3\pi_1^4\pi_2^2\alpha_1\alpha_2\gamma_1^2\gamma_2 - 12\lambda^3\pi_1^4\pi_2^2\alpha_1^2\alpha_2\gamma_1^2\gamma_2 - 6\lambda^3\pi_1^4\pi_2^2\alpha_2^2\gamma_1^2\gamma_2 -$$
$$3\lambda^3\pi_1^4\pi_2^2\alpha_1\alpha_2^2\gamma_1^2\gamma_2 - 18\lambda^2\pi_1^3\pi_2^2\gamma_1^3\gamma_2 - 3\lambda^2\pi_1^3\pi_2^2\alpha_1\gamma_1^3\gamma_2 - 12\lambda^2\pi_1^3\pi_2^2\alpha_1^2\gamma_1^3\gamma_2 +$$
$$24\lambda^2\pi_1^3\pi_2^2\alpha_2\gamma_1^3\gamma_2 + 9\lambda^2\pi_1^3\pi_2^2\alpha_1\alpha_2\gamma_1^3\gamma_2 - 6\lambda^2\pi_1^3\pi_2^2\alpha_2^2\gamma_1^3\gamma_2 - 18\lambda\pi_1^2\pi_2^2\gamma_1^4\gamma_2 -$$
$$15\lambda\pi_1^2\pi_2^2\alpha_1\gamma_1^4\gamma_2 + 12\lambda\pi_1^2\pi_2^2\alpha_2\gamma_1^4\gamma_2 - 6\pi_1\pi_2^2\gamma_1^5\gamma_2 + 6\lambda^3\pi_1^5\pi_2\gamma_1\gamma_2^2 -$$
$$15\lambda^3\pi_1^5\pi_2\alpha_1\gamma_1\gamma_2^2 + 12\lambda^3\pi_1^5\pi_2\alpha_1^2\gamma_1\gamma_2^2 - 3\lambda^3\pi_1^5\pi_2\alpha_1^3\gamma_1\gamma_2^2 - 6\lambda^3\pi_1^5\pi_2\alpha_2\gamma_1\gamma_2^2 +$$
$$3\lambda^3\pi_1^5\pi_2\alpha_1\alpha_2\gamma_1\gamma_2^2 + 3\lambda^3\pi_1^5\pi_2\alpha_1^2\alpha_2\gamma_1\gamma_2^2 + 18\lambda^2\pi_1^4\pi_2\gamma_1^2\gamma_2^2 - 3\lambda^2\pi_1^4\pi_2\alpha_1\gamma_1^2\gamma_2^2 +$$
$$3\lambda^2\pi_1^4\pi_2\alpha_1^2\gamma_1^2\gamma_2^2 - 12\lambda^2\pi_1^4\pi_2\alpha_2\gamma_1^2\gamma_2^2 - 6\lambda^2\pi_1^4\pi_2\alpha_1\alpha_2\gamma_1^2\gamma_2^2 + 18\lambda\pi_1^3\pi_2\gamma_1^3\gamma_2^2 +$$
$$12\lambda\pi_1^3\pi_2\alpha_1\gamma_1^3\gamma_2^2 - 6\lambda\pi_1^3\pi_2\alpha_2\gamma_1^3\gamma_2^2 + 6\pi_1^2\pi_2\gamma_1^4\gamma_2^2 - 2\lambda^3\pi_1^6\gamma_2^3 + 6\lambda^3\pi_1^6\alpha_1\gamma_2^3 -$$
$$6\lambda^3\pi_1^6\alpha_1^2\gamma_2^3 + 2\lambda^3\pi_1^6\alpha_1^3\gamma_2^3 - 6\lambda^2\pi_1^5\gamma_1\gamma_2^3 + 3\lambda^2\pi_1^5\alpha_1\gamma_1\gamma_2^3 + 3\lambda^2\pi_1^5\alpha_1^2\gamma_1\gamma_2^3 -$$
$$6\lambda\pi_1^4\gamma_1^2\gamma_2^3 - 3\lambda\pi_1^4\alpha_1\gamma_1^2\gamma_2^3 - 2\pi_1^3\gamma_1^3\gamma_2^3 +$$
$$\sqrt{\left(4\left(3\lambda\pi_1^2\alpha_1\left(-\pi_2\gamma_1 + \pi_1\gamma_2\right)\left(\lambda\pi_1\left(2+\alpha_1\right)\gamma_2 + \gamma_1\left(-\lambda\pi_2 + \lambda\pi_2\alpha_2 + \gamma_2\right)\right) - \right.\right.}$$
$$\left.\left.\left(-\pi_2\gamma_1^2 + \lambda\pi_1^2\left(1+2\alpha_1\right)\gamma_2 + \pi_1\gamma_1\left(-\lambda\pi_2 - \lambda\pi_2\alpha_1 + \lambda\pi_2\alpha_2 + \gamma_2\right)\right)^2\right)^3 +$$
$$\left(2\pi_2^3\gamma_1^6 + 6\pi_1\pi_2^2\gamma_1^5\left(\lambda\pi_2 + \lambda\pi_2\alpha_1 - \lambda\pi_2\alpha_2 - \gamma_2\right) + 2\lambda^3\pi_1^6\left(-1+\alpha_1\right)^3\gamma_2^3 -\right.$$
$$3\lambda^2\pi_1^5\left(-1+\alpha_1\right)\gamma_1\gamma_2^2\left(\lambda\pi_2\alpha_1^2 - 2\left(-\lambda\pi_2 + \lambda\pi_2\alpha_2 + \gamma_2\right) - \alpha_1\left(3\lambda\pi_2 + \lambda\pi_2\alpha_2 + \gamma_2\right)\right) +$$
$$3\pi_1^2\pi_2\gamma_1^4\left(2\lambda^2\pi_2^2\alpha_1^2 + \lambda\pi_2\alpha_1\left(\lambda\pi_2 - \lambda\pi_2\alpha_2 - 5\gamma_2\right) +\right.$$
$$\left.2\left(\lambda^2\pi_2^2 + \lambda^2\pi_2^2\alpha_2^2 - 3\lambda\pi_2\gamma_2 + \gamma_2^2 + 2\lambda\pi_2\alpha_2\left(-\lambda\pi_2 + \gamma_2\right)\right)\right) -$$
$$3\lambda\pi_1^4\gamma_1^2\gamma_2\left(\lambda^2\pi_2^2\alpha_1^3 + \lambda\pi_2\alpha_1^2\left(\lambda\pi_2 + 4\lambda\pi_2\alpha_2 - \gamma_2\right) +\right.$$
$$2\left(\lambda^2\pi_2^2 + \lambda^2\pi_2^2\alpha_2^2 - 3\lambda\pi_2\gamma_2 + \gamma_2^2 + 2\lambda\pi_2\alpha_2\left(-\lambda\pi_2 + \gamma_2\right)\right) +$$
$$\left.\alpha_1\left(-4\lambda^2\pi_2^2 + \lambda^2\pi_2^2\alpha_2^2 + \lambda\pi_2\gamma_2 + \gamma_2^2 + \lambda\pi_2\alpha_2\left(3\lambda\pi_2 + 2\gamma_2\right)\right)\right) +$$
$$\pi_1^3\gamma_1^3\left(2\lambda^3\pi_2^3\alpha_1^3 + 3\lambda^2\pi_2^2\alpha_1^2\left(-\lambda\pi_2 + \lambda\pi_2\alpha_2 - 4\gamma_2\right) -\right.$$
$$3\lambda\pi_2\alpha_1\left(\lambda^2\pi_2^2 + \lambda^2\pi_2^2\alpha_2^2 + \lambda\pi_2\gamma_2 - 4\gamma_2^2 - \lambda\pi_2\alpha_2\left(2\lambda\pi_2 + 3\gamma_2\right)\right) -$$
$$2\left(-\lambda^3\pi_2^3 + \lambda^3\pi_2^3\alpha_2^3 + 9\lambda^2\pi_2^2\gamma_2 - 9\lambda\pi_2\gamma_2^2 + \gamma_2^3 + 3\lambda^2\pi_2^2\alpha_2^2\left(-\lambda\pi_2 + \gamma_2\right) +\right.$$
$$\left.\left.\left.3\lambda\pi_2\alpha_2\left(\lambda^2\pi_2^2 - 4\lambda\pi_2\gamma_2 + \gamma_2^2\right)\right)\right)^2\right)\Big)^{1/3} +$$
$$2^{2/3}\left(1 + i\sqrt{3}\right)$$
$$(2\lambda^3\pi_1^3\pi_2^3\gamma_1^3 - 3\lambda^3\pi_1^3\pi_2^3\alpha_1\gamma_1^3 - 3\lambda^3\pi_1^3\pi_2^3\alpha_1^2\gamma_1^3 + 2\lambda^3\pi_1^3\pi_2^3\alpha_1^3\gamma_1^3 - 6\lambda^3\pi_1^3\pi_2^3\alpha_2\gamma_1^3 +$$
$$6\lambda^3\pi_1^3\pi_2^3\alpha_1\alpha_2\gamma_1^3 + 3\lambda^3\pi_1^3\pi_2^3\alpha_1^2\alpha_2\gamma_1^3 + 6\lambda^3\pi_1^3\pi_2^3\alpha_2^2\gamma_1^3 - 3\lambda^3\pi_1^3\pi_2^3\alpha_1\alpha_2^2\gamma_1^3 -$$
$$2\lambda^3\pi_1^3\pi_2^3\alpha_2^3\gamma_1^3 + 6\lambda^2\pi_1^2\pi_2^3\gamma_1^4 + 3\lambda^2\pi_1^2\pi_2^3\alpha_1\gamma_1^4 + 6\lambda^2\pi_1^2\pi_2^3\alpha_1^2\gamma_1^4 - 12\lambda^2\pi_1^2\pi_2^3\alpha_2\gamma_1^4 -$$
$$3\lambda^2\pi_1^2\pi_2^3\alpha_1\alpha_2\gamma_1^4 + 6\lambda^2\pi_1^2\pi_2^3\alpha_2^2\gamma_1^4 + 6\lambda\pi_1\pi_2^3\gamma_1^5 + 6\lambda\pi_1\pi_2^3\alpha_1\gamma_1^5 - 6\lambda\pi_1\pi_2^3\alpha_2\gamma_1^5 +$$
$$2\pi_2^3\gamma_1^6 - 6\lambda^3\pi_1^4\pi_2^2\gamma_1^2\gamma_2 + 12\lambda^3\pi_1^4\pi_2^2\alpha_1\gamma_1^2\gamma_2 - 3\lambda^3\pi_1^4\pi_2^2\alpha_1^2\gamma_1^2\gamma_2 - 3\lambda^3\pi_1^4\pi_2^2\alpha_1^3\gamma_1^2\gamma_2 +$$
$$12\lambda^3\pi_1^4\pi_2^2\alpha_2\gamma_1^2\gamma_2 - 9\lambda^3\pi_1^4\pi_2^2\alpha_1\alpha_2\gamma_1^2\gamma_2 - 12\lambda^3\pi_1^4\pi_2^2\alpha_1^2\alpha_2\gamma_1^2\gamma_2 - 6\lambda^3\pi_1^4\pi_2^2\alpha_2^2\gamma_1^2\gamma_2 -$$
$$3\lambda^3\pi_1^4\pi_2^2\alpha_1\alpha_2^2\gamma_1^2\gamma_2 - 18\lambda^2\pi_1^3\pi_2^2\gamma_1^3\gamma_2 - 3\lambda^2\pi_1^3\pi_2^2\alpha_1\gamma_1^3\gamma_2 - 12\lambda^2\pi_1^3\pi_2^2\alpha_1^2\gamma_1^3\gamma_2 +$$
$$24\lambda^2\pi_1^3\pi_2^2\alpha_2\gamma_1^3\gamma_2 + 9\lambda^2\pi_1^3\pi_2^2\alpha_1\alpha_2\gamma_1^3\gamma_2 - 6\lambda^2\pi_1^3\pi_2^2\alpha_2^2\gamma_1^3\gamma_2 - 18\lambda\pi_1^2\pi_2^2\gamma_1^4\gamma_2 -$$
$$15\lambda\pi_1^2\pi_2^2\alpha_1\gamma_1^4\gamma_2 + 12\lambda\pi_1^2\pi_2^2\alpha_2\gamma_1^4\gamma_2 - 6\pi_1\pi_2^2\gamma_1^5\gamma_2 + 6\lambda^3\pi_1^5\pi_2\gamma_1\gamma_2^2 -$$
$$15\lambda^3\pi_1^5\pi_2\alpha_1\gamma_1\gamma_2^2 + 12\lambda^3\pi_1^5\pi_2\alpha_1^2\gamma_1\gamma_2^2 - 3\lambda^3\pi_1^5\pi_2\alpha_1^3\gamma_1\gamma_2^2 - 6\lambda^3\pi_1^5\pi_2\alpha_2\gamma_1\gamma_2^2 +$$
$$3\lambda^3\pi_1^5\pi_2\alpha_1\alpha_2\gamma_1\gamma_2^2 + 3\lambda^3\pi_1^5\pi_2\alpha_1^2\alpha_2\gamma_1\gamma_2^2 + 18\lambda^2\pi_1^4\pi_2\gamma_1^2\gamma_2^2 - 3\lambda^2\pi_1^4\pi_2\alpha_1\gamma_1^2\gamma_2^2 +$$
$$3\lambda^2\pi_1^4\pi_2\alpha_1^2\gamma_1^2\gamma_2^2 - 12\lambda^2\pi_1^4\pi_2\alpha_2\gamma_1^2\gamma_2^2 - 6\lambda^2\pi_1^4\pi_2\alpha_1\alpha_2\gamma_1^2\gamma_2^2 + 18\lambda\pi_1^3\pi_2\gamma_1^3\gamma_2^2 +$$
$$12\lambda\pi_1^3\pi_2\alpha_1\gamma_1^3\gamma_2^2 - 6\lambda\pi_1^3\pi_2\alpha_2\gamma_1^3\gamma_2^2 + 6\pi_1^2\pi_2\gamma_1^4\gamma_2^2 - 2\lambda^3\pi_1^6\gamma_2^3 + 6\lambda^3\pi_1^6\alpha_1\gamma_2^3 -$$
$$6\lambda^3\pi_1^6\alpha_1^2\gamma_2^3 + 2\lambda^3\pi_1^6\alpha_1^3\gamma_2^3 - 6\lambda^2\pi_1^5\gamma_1\gamma_2^3 + 3\lambda^2\pi_1^5\alpha_1\gamma_1\gamma_2^3 + 3\lambda^2\pi_1^5\alpha_1^2\gamma_1\gamma_2^3 -$$
$$6\lambda\pi_1^4\gamma_1^2\gamma_2^3 - 3\lambda\pi_1^4\alpha_1\gamma_1^2\gamma_2^3 - 2\pi_1^3\gamma_1^3\gamma_2^3 +$$

$$\sqrt{}\left(4\left(3\lambda\pi_1^2\alpha_1\left(-\pi_2\gamma_1+\pi_1\gamma_2\right)\left(\lambda\pi_1\left(2+\alpha_1\right)\gamma_2+\gamma_1\left(-\lambda\pi_2+\lambda\pi_2\alpha_2+\gamma_2\right)\right)-\right.\right.$$
$$\left.\left.\left(-\pi_2\gamma_1^2+\lambda\pi_1^2\left(1+2\alpha_1\right)\gamma_2+\pi_1\gamma_1\left(-\lambda\pi_2-\lambda\pi_2\alpha_1+\lambda\pi_2\alpha_2+\gamma_2\right)\right)^2\right)^3+$$
$$\left(2\pi_2^3\gamma_1^6+6\pi_1\pi_2^2\gamma_1^5\left(\lambda\pi_2+\lambda\pi_2\alpha_1-\lambda\pi_2\alpha_2-\gamma_2\right)+2\lambda^3\pi_1^6\left(-1+\alpha_1\right)^3\gamma_2^3-\right.$$
$$3\lambda^2\pi_1^5\left(-1+\alpha_1\right)\gamma_1\gamma_2^2\left(\lambda\pi_2\alpha_1^2-2\left(-\lambda\pi_2+\lambda\pi_2\alpha_2+\gamma_2\right)-\alpha_1\left(3\lambda\pi_2+\lambda\pi_2\alpha_2+\gamma_2\right)\right)+$$
$$3\pi_1^2\pi_2\gamma_1^4\left(2\lambda^2\pi_2^2\alpha_1^2+\lambda\pi_2\alpha_1\left(\lambda\pi_2-\lambda\pi_2\alpha_2-5\gamma_2\right)+\right.$$
$$2\left(\lambda^2\pi_2^2+\lambda^2\pi_2^2\alpha_2^2-3\lambda\pi_2\gamma_2+\gamma_2^2+2\lambda\pi_2\alpha_2\left(-\lambda\pi_2+\gamma_2\right)\right))-$$
$$3\lambda\pi_1^4\gamma_1^2\gamma_2\left(\lambda^2\pi_2^2\alpha_1^3+\lambda\pi_2\alpha_1^2\left(\lambda\pi_2+4\lambda\pi_2\alpha_2-\gamma_2\right)+\right.$$
$$2\left(\lambda^2\pi_2^2+\lambda^2\pi_2^2\alpha_2^2-3\lambda\pi_2\gamma_2+\gamma_2^2+2\lambda\pi_2\alpha_2\left(-\lambda\pi_2+\gamma_2\right)\right)+$$
$$\alpha_1\left(-4\lambda^2\pi_2^2+\lambda^2\pi_2^2\alpha_2^2+\lambda\pi_2\gamma_2+\gamma_2^2+\lambda\pi_2\alpha_2\left(3\lambda\pi_2+2\gamma_2\right)\right))+$$
$$\pi_1^3\gamma_1^3\left(2\lambda^3\pi_2^3\alpha_1^3+3\lambda^2\pi_2^2\alpha_1^2\left(-\lambda\pi_2+\lambda\pi_2\alpha_2-4\gamma_2\right)-\right.$$
$$3\lambda\pi_2\alpha_1\left(\lambda^2\pi_2^2+\lambda^2\pi_2^2\alpha_2^2+\lambda\pi_2\gamma_2-4\gamma_2^2-\lambda\pi_2\alpha_2\left(2\lambda\pi_2+3\gamma_2\right)\right)-$$
$$2\left(-\lambda^3\pi_2^3+\lambda^3\pi_2^3\alpha_2^3+9\lambda^2\pi_2^2\gamma_2-9\lambda\pi_2\gamma_2^2+\gamma_2^3+3\lambda^2\pi_2^2\alpha_2^2\left(-\lambda\pi_2+\gamma_2\right)+\right.$$
$$\left.\left.\left.\left.\left.3\lambda\pi_2\alpha_2\left(\lambda^2\pi_2^2-4\lambda\pi_2\gamma_2+\gamma_2^2\right)\right)\right)^2\right)\right)^{1/3}\Bigg)$$

Appendix B

A reverse transcription protocol

The following protocol for a reverse transcription experiment is taken from [Sing]. See the URL for details.

http://micro.mic.ucdavis.edu/singer/protocols/MicroarrayProtocol.pdf

for details. The protocol was used to determine the parameter values for the analysis of the reverse transcription model in Section 2.2.1.

SYNTHESIS OF FLUORESCENTLY LABELLED
CDNA PROBE FOR MICROARRAYS

DAY 1
- Note the reverse transcriptase, coupling and hybridization protocols have been adapted from those posted on www.microarrays.org/protocols.html by Joe DeRisi.
- Blocking protocol has been adapted from those posted at the Erie Scientific website.
- Protocol adapted from Gross Lab (Virgil Rhodus)
Block Microarray slides

(Adapted from Erie Scientific)
A/ Stock BSA solution and blocking (remember to etch arrays on back, label side down, before proceeding.
1) Make stock BSA solution. In 1L beaker add 10 g Fraction V BSA to 840 ml MilliQ $h-20$, stir at RT until dissolved (takes several hours).
1) Add 150 ml 20x SSC and filter sterilize with 22 μm filter (can be stored

at 4 °C for up to 3 months).
1) Blocking. Fill dish with RT BSA solution and set on rotator with slides for 10 min.
1) Transfer slides to MilliQ H_2O, plunge ∼30x. Repeat 4x.
1) Transfer slides to boiling (>95 °C) MilliQ H_2O, incubate 2.5 min on bench.
1) Spin slides ∼1 (400k)
1) Stores slides in slide box for up to 1 week.
Reverse transcriptase reaction
(Adapted from Joe De Risi and Holly Baxter, and developed by Rosetta Inpharmatics, Kirkland, WA)

A/ Primer annealing and cDNA synthesis
Note - continue to use RNase free tubes, pipette tips and solutions until the end of step A
1) Annealing step. In 0.5 ml microfuge tubes, mix 20 μg RNA sample with 10 μg random hexamer (10 μl of 1 $\mu g/\mu l$) in H_2O (DEPC treated, RNase free) to give a final volume of 20 μl.
2
1) Incubate mixture of RNA and hexamer at 70 °C for 10 min.
1) Chill on ice for 10 min.
1) cDNA synthesis reaction: Make up following master mix:

Vol. per reaction (μl)	Reagent	
3	10x StrataScript RT Buffer	Stratagene
0.6	50x aa-dUTP/dNTP mix	50x mix = 25 mM each dA/dC/dG,15 mM amino-allyl dUTP, 10 mM dTTP
3	StrataScript RNase H-RT	Stratagene; catalog 600085-51
0.4	RNase Inhibitor (40 U/μl)	Boehringer Mannheim; catalog 799017
3	0.1 M DTT	
10	Total vol/tube	

5) Add 10 μl of the Master Mix to each RNA/hexamer mixture (20 μl) to give 30 μl
final volume.
5) Incubate at 37 °C for 10 min.
5) Incubate at 42 °C for 1 hr 40 min.
5) Incubate at 50 °C for 10 min. (you can freeze samples @ -20°C here)
B/ RNA Hydrolysis
1) Add 10 μl 0.5 M EDTA (pH 8.0) to the 30 μl RNA/cDNA reaction, mix and spin, then add 10 μl 1 N NaOH (freshly prepared or from unopened

frozen aliquots).
1) Incubate at 65°C for 1 hr.
1) Add 25 μl of 1 M Hepes pH 7.5 to neutralize the reaction (or 10 μl 3M NaOAc pH 5.2).
C/ Cleanup using Microcon-30 filters
Note - When using the Microcon filters, try not to let sample spin dry: if this occurs, the sample can be recovered simply by adding ∼30 μl H_2O to the membrane, incubating for a few minutes and then eluting the sample. The spin times are approximate and will vary from batch to batch and sample to sample.
1) Fill microcon-30 tube with 350 μl H_2O, add sample (∼75 μl), and rinse reaction tube with 100 μl H_2O. (Total amount of H_2O added is 450 μl).
1) Spin at 10,000 g for 8.5 min using Beckman centrifuge rotor, F2402.
3
1) Check filter/vol. in upper chamber. Should be between 10-50 μl; if not, spin for additional 1-2 min. Recheck volume. Discard flow-through.
1) Wash 2 times by adding 450 μl H_2O to upper chamber and recentrifuging at 10,000 rpm for 8.5 min. Each time ensure the volume has reduced to 10-50 μl before proceeding.
1) Elute sample by placing the microcon inverted into a fresh microfuge tube. Centrifuge at 5,000 rpm for 30 sec.
1) Dry the sample in a speed vac. (approx. 20 min.) or ∼3 μl remains . Do not over dry.
1) Store dried samples at -20 °C.
NB - the dried samples can be stored at -20 °C for at least 1 month.
4
DAY 2
Coupling reaction of Alexa dyes to aadUTP-cDNA sample and overnight hybridization to microarray
Note - the Alexa dyes are light sensitive. Therefore minimize light exposure where possible during the following procedures. In addition, the Alexa dyes degrade over a few days. Only perform the coupling reaction if it is possible to directly proceed to the hybridization step and then on to scan the microarrays.
A/ Coupling of Alexa dyes to the aadUTP-cDNA
Note - the Alexa fluor 555 dye appears pink, but scans as "green" (comparable to Cy3), and the Alexa fluor 647 dye appears blue, but scans as "red" (comparable to Cy5). By convention, the "wild type" or control sample is labeled with 555 and the experimental sample is labeled with 647.
1) Dissolve cDNA in 5 μl dH_2O, warm @ 42°C∼5".
2) Add 3 μl labeling buffer (25 mg sodium bicarbonate, 1 ml dH_2O) to each

sample, mix well.
3) Add 6 μl RT DMSO to tube of dye and resuspend (1 tube of dye per 3 samples). Add 2 μl DMSO/dye to sample.
4) Mix cDNA and dye together and incubate at room temperature for 1 hr in the dark. B/ Cleanup with QIA-quick PCR kit.
1) Add 90 μl H_2O to each sample to make up to 100 μl.
2) Add 500 μl PB buffer from kit.
3) Apply to QIA-quick column. Spin at 13,000 rpm for 30-60 s.
4) Dump flow through. Add 750 μl PE buffer and spin at 13,000 rpm for 30-60 s.
5) Dump flow through. Repeat PE buffer wash step 4x.
6) Dump flow through. Spin for 1 min at 14,000 rpm. (Filters should look pink for Cy3 and blue for Cy5 reactions at this point).
7) Transfer to a fresh eppendorf tube. Add 30 μl Tris pH 8.5 (EB buffer). Let sit 1 min. Spin 13,000 rpm for 1 min.
8) Add an additional 30 μl to column. Let sit 1 min. Spin 13, 000 rpm for 1 min.
9) Final volume 60 μl. The solutions should be clearly pink for Cy3 and blue for Cy5 at this point. If they are not, the labeling reaction did not work.
5
10) Pool sample pairs to give 120 μl of purple solution.
11) Apply samples to microcons and spin for 1.5-2 min to reduce sample to 2-5 μl. The flow through will be clear and the sample will be strongly visible on the membrane.
12) Invert microcons and elute samples into fresh tubes.
13) Dry samples in Speed Vac. Cover the lid with foil to avoid exposing the sample to light. (Approximately 5-15 min. to dry).
14) Store the dried sample in the dark (wrapped in foil) at 4°C. Stable for 1-2 days.
6
Day 3
A/ Hybridization step
(Hybridization conditions: cDNA from 16 μg total RNA, 15 μg poly(dI-dC), 3x SSC, 25 mM Hepes (pH 7.0), 0.225% SDS)
1) Set up the following hybridization mix in a 0.5 ml microfuge tube.

37.8 μl resuspended cDNA in H_2O (dissolve at 65°C, 1-2 min.)
7.1 μl 20x SSC
1.2 μl 1 M Hepes pH 7.2
1.0 μl 10% SDS
47.1 μl Total volume

Add SDS last after mixing the cocktail; do not chill samples after adding SDS, this will cause the SDS to precipitate.
2) Incubate samples at 95°C in dry heating block for 2 min.
3) Allow samples to cool 5-10 min at room temperature and spin down briefly.
B/ Slide preparation
Use fresh or less than 2 weeks old post-processed slides.
1) Whilst samples are cooling, place slides in hybridisation chamber and remove any dust using compressed air briefly.
2) Clean coverslips using EtOH soaked Kimwipes. Dry and dust with compressed air and carefully place over the top of the microarray using forceps such that the dull white strips (rough side down) are on the long axis of the slide and touching the glass.
3) Add a total of 6-10 2 μl drops of 3x SSC at the two ends of the slides removed from the coverslip.
C/ Sample application
1) After the samples have cooled, apply to the array by placing a pipette tip at one end of the coverslip and allow the sample to move up underneath the coverslip by capillary action. Move the pipette tip repeatedly along the length of the coverslip to avoid any bubbles. Add sample to the other end of the coverslip once completely full underneath, to "top up" both ends.
7
2) Place cover on the hybridization chamber and tighten the lid screws carefully to make water tight. Keep the chamber horizontal at all times so as not to disturb the 3x SSC droplets.
3) Carefully lower the hybridization chamber onto a plastic holder in a water bath.
4) Hybridize at 63-65 °C for at least 5-6 hrs, or overnight (12 hrs max.).
D/ Rinse Step
1) Prepare wash solutions in glass slide dishes, with each dish having its own rack.
Wash solution 1: 340 ml Milli-Q water
10 ml 20x SSC
1 ml 10 % SDS
Wash solution 2: 350 ml Milli-Q water
1 ml 20x SSC
Wash solution 3: 350 ml Milli-Q water
100 μl 20x SSC
Wash solution 4: 350 ml Milli-Q water
10 μl 20x SSC
2) Remove array carefully from the water bath, keeping the chamber level. Dry the chambers with paper towels and "wick" any water from the chamber

seems.
3) Unscrew the chamber and remove array slide.
4) First Rinse: Rinse slide in Wash solution 1. Use forceps to move slide gently up and down in the solution until the coverslip is dislodged. Avoid allowing coverslip to scratch the surface of the array. Once coverslip is off and all the slides are in place, shake in solution by plunging rack up and down 10-20 times. Let incubate for 1 minute.
5) Second Rinse: Individually transfer slides to Wash solution 2, blotting the base of the slide on a paper towel to avoid carrying over too much SDS. Shake gently in solution a few times. Let incubate for 1 minute. Repeat for washes 3 and 4.
6) Remove excess liquid by blotting the rack on a paper towel, and then dry array at room temperature by centrifuging at 600 rpm for 5 min.
7) Scan array soon as the dyes are unstable and degrade differentially.

Eidesstattliche Erklärung

Hiermit erkläre ich, dass diese Arbeit bisher von mir weder an der Mathematisch-Naturwissenschaftlichen Fakultät der Ernst-Moritz-Arndt-Universität Greifswald noch einer anderen wissenschaftlichen Einrichtung zum Zwecke der Promotion eingereicht wurde. Ferner erkläre ich, dass ich diese Arbeit selbständig verfasst und keine anderen als die darin angegebenen Hilfsmittel benutzt habe.