

Dissertation
submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

Put forward by

Dipl. Phys. Jennifer Wagner

born in Dudweiler

Oral examination: October 2011

Quality control for peptide chip array production

Referees:

Prof. Dr. Bernd Jähne
Prof. Dr. Michael Hausmann

Abstract

German

Diese Arbeit baut die Qualitätskontrolle für die selektive Ablagerung von triboelektrisch geladenen, mit Aminosäuren bestückten Mikropartikeln auf einem CMOS-Chip auf, um hohe Peptiddichten in einer modifizierten Merrifield-Peptidsynthese zu erhalten. Sie umfasst den Aufbau und die Implementierung einer auf Bildverarbeitung basierenden, automatisierten Evaluierung der Partikelablagerungen. Letztere gründet sich auf die in dieser Arbeit durchgeführten Forschungen über die Analyse der Partikelcharakteristik, die Modellierung des Partikeltransfers aus einem Aerosol, die Ableitung von optimierten Parametern für den Übertrag und das Auffinden einer Oberflächenanalysemethode, die es erlaubt, halbquantitativ ortsaufgelöst die auf dem Träger gekoppelten Aminosäuren im Mikrometerbereich zu detektieren. Außerdem wird eine flexible Schnittstelle zu der Steuerung des automatisierten Systems implementiert, die unabhängig vom Steuersystem selbst ist.

English

This work establishes quality control for selective deposition of tribo-electrically charged micro-particles containing pre-activated amino acids onto a CMOS-chip in order to achieve high peptide densities in a modified Merrifield peptide synthesis. It encompasses the setup and implementation of an image processing supported automated assessment of the particle depositions based on the analysis of the particles used, modelling the particle transfer out of an aerosol, deducing optimised parameters and finding a surface analysis experiment that allows for semi-quantitative spatially resolved detection of amino acids coupled to the support in the micrometer range. Furthermore, a flexible stand-alone interface to the control program of the automated system is implemented.

Contents

1	Preface	1
1.1	Motivation	1
1.2	Peptide synthesis methods	2
1.2.1	Merrifield synthesis	2
1.2.2	Peptide synthesis on CMOS chips	4
1.3	State of the art quality analysis	5
1.4	Quality analysis by image processing	6
1.4.1	Related work in biology	6
1.4.2	Application to peptide chips	8
1.5	Outline	10
1.6	Publications	11
2	Setup analysis	13
2.1	Requirements	13
2.2	Camera setups	16
2.3	Microscope setups	19
2.4	Image formats	22
3	Aerosol dynamics	23
3.1	Related work	23
3.2	Experimental setup	24
3.3	Prerequisites	25
3.3.1	Characteristic numbers	25
3.3.2	Forces	28
3.4	Models	30
3.4.1	Homogeneous field in stationary air	31
3.4.2	Inhomogeneous field in moving air	34
4	Coupling detection	41
4.1	Motivation	41
4.2	Bio-chemical detection reactions	42
4.2.1	Bromophenol blue	42
4.2.2	Ninhydrin	42

4.2.3	Fmoc determination	43
4.3	Surface analysis techniques	43
4.3.1	Time Of Flight - Secondary Ion Mass Spectrometry	44
5	Quality controlling program	51
5.1	Concept and outline	51
5.2	Image enhancement	52
5.2.1	Noise reduction	52
5.2.2	Flat-field correction	53
5.2.3	Smoothing with a Gaussian kernel	54
5.2.4	Smoothing with a bilateral filter	55
5.3	Image segmentation	56
5.3.1	Global thresholding	57
5.3.2	Histogram bin k-means clustering	59
5.3.3	Histogram bin earth mover's distance clustering	60
5.4	Feature extraction and selection	62
5.4.1	Principle component analysis (PCA)	63
5.4.2	Mean grey value	65
5.4.3	Spot coverage	66
5.4.4	Number and size distributions of particles	67
5.5	Quality measures	70
5.5.1	Expert labelling quality measure	70
5.5.2	Bio-chemically implied quality measure	73
5.6	Training phase	79
5.6.1	Selection of a representative training set	79
5.6.2	Training of the parameters	80
5.7	Testing phase	85
5.7.1	Processing chain	85
5.7.2	Experimental results	87
5.8	Automation	92
5.8.1	Automated peptide synthesis machine	92
5.8.2	Role of the quality analysis program	93
5.8.3	Interface to the quality analysis program	95
5.9	Test on an entire peptide synthesis	98
5.9.1	Particle deposition analysis	98
5.9.2	Fluorescence coupling analysis	104
6	Discussion & outlook	107
6.1	Discussion	107
6.2	Outlook	110
	Acknowledgements	112

A	Coupling detection supplements	121
A.1	Surface coatings	121
B	Algorithmic supplements	123
B.1	Correlation related images	123
B.2	Usage of the training routine	127
B.3	Usage of testing routine	128
B.4	Special parameter specifications	129
B.4.1	Series numbers	129
B.5	Segmentations of synthesis images	130

Chapter 1

Preface

Quality means doing it right when no one is looking.
- Henry Ford

1.1 Motivation

In nature, peptides, short sequences of two to about one hundred amino acids, are synthesised in ribosomes by translation of mRNA or in enzymatic catalysis. They can be sorted into different classes according to their function, e.g. hormones, neuropeptides, alkaloids, antibiotics and toxins. One prominent example is insulin from the class of peptide hormones consisting of 51 amino acids, regulating the glucose metabolism. Neuropeptides usually produced in the brain are involved in signal processing in neural tissue. In plants, fungi or some animals alkaloids contribute to the defence of the organism against being eaten by other organisms. Peptides belonging to the group of toxins are poisonous substances, while antibiotics usually inhibit the growth of micro organisms such as bacterial cells. Hence, understanding the interactions and reactions peptides are involved in can give insights in the functions of cellular processes but can also be used in drug development, when investigating interactions with peptides that lead to a malfunction in cellular processes as, for instance, the immune response. Especially for this purpose, artificial peptide synthesis is necessary to systematically test variants of existing amino acid sequences to find optimum binding partners (ligands) to the epitopes of the malignant molecules to tackle diseases in which peptides play an important role.

Since the first peptide synthesis in 1882 by T. Curtius, peptides can artificially be produced by several methods, the most common ones being liquid phase synthesis and solid phase synthesis. The former assumes that peptides are synthesised in solution, while the latter requires a solid support with coupling sites to which the first amino acids can couple. Solid phase peptide synthesis surpasses the method based on liquids in the percentage of correctly assembled peptides as separation and purification of the desired peptide sequences are easier to perform.

1.2 Peptide synthesis methods

1.2.1 Merrifield synthesis

The state of the art solid phase peptide synthesis which will also be considered in this work is called the Merrifield synthesis [Mer63], after R. Merrifield who developed it in 1963. A solid support is provided with linkers joining it with the first coupling sites. The synthesis is a process of repeating transport, coupling, washing and deprotecting steps. Depending on the medium in which the amino acids are transferred to the growing peptides, liquid transport and particle based transport can be distinguished. Liquid transport is widely used for mass production of peptides, see for instance [Vol09] or [FRP⁺91], while particle based transport is exclusively applied in the collaboration between Heidelberg University and the German Cancer Research Centre and recently became commercially available via the start-up PEPperPRINT [BNB⁺07].

The synthesising process can then be described by the following steps. After reaching the synthesis spot, the amino acids in the transfer medium can couple covalently with their C-terminus to the linker, by which they are immobilised. Subsequently, a washing step with various solvents discards the transfer medium, but does not affect the peptide bond between the linker and the first amino acid. In order to prevent couplings between amino acids in the transfer medium, their N-termini are protected by Fmoc groups¹. Furthermore, the side chains of the amino acids also have to be protected as they can contain chemically active groups that might interfere in the synthesis process as well (s. [Kön10] for an overview of the natural amino acids used in the peptide project). The subsequent capping step couples a permanent capping group to all N-termini of the support to which no amino acid has coupled to prevent couplings to those sites in the next steps. Hence, this procedure drastically reduces the number of incorrectly assembled peptides. After capping, a deprotecting step is necessary to free the N-termini of the coupled amino acids but leave the protection of the side chains (and the capping) intact. Then, the cycle starts anew with the next amino acids that can now couple to the N-termini of the first ones.

Figure 1.1 summarises this process. The solid polymer on the left represents the support, linker and coupling sites. In the first row, an amino acid is already coupled to the polymer. Its side chains are summarised as R_1 and the Fmoc-protection of the N-terminus is marked by the circle around the amino group. In the second row, the deprotection step is shown. The capping step can be omitted here, as only one growing peptide is considered. Then, in the third row, the second amino acid with rest R_2 , which is already shown on the right in the first row, is coupled to the growing peptide consisting of the polymer bead and the first amino acid. After another

¹Another option would be to use Boc as protection group, which has a lower cost of reagents and a higher purity, but needs strong acids and therefore requires special equipment.

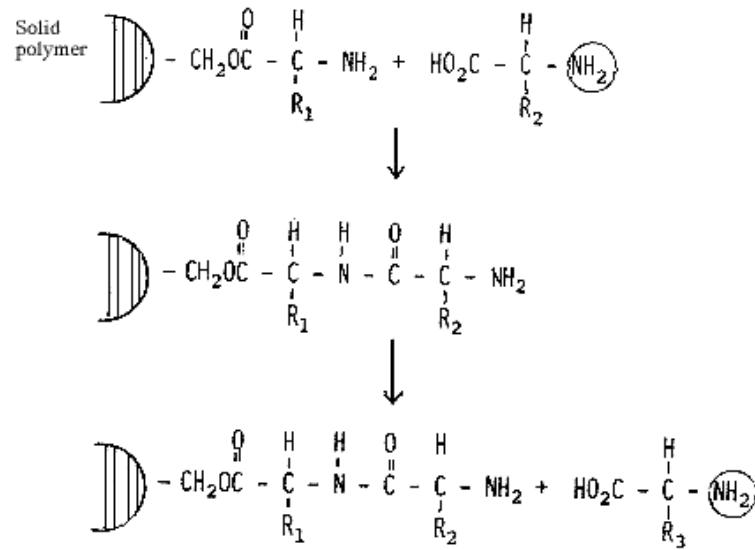


Figure 1.1: Top to bottom: Solid phase Merrifield synthesis: An amino acid coupled to the solid polymer comes into contact with an amino acid with a protected N-terminus (indicated by the circle around NH_2); The two amino acids couple, form a dimer and the N-terminus of the coupled amino acid is deprotected; The dimer comes into contact with a third protected amino acid to be coupled in the next step (source: [Mer84])

deprotection step, the third amino acid, which is shown on the right in the third row, can couple to the peptide consisting of two amino acids (dimer). Using this technique in which the direction of synthesis is opposite to the natural one starting from N- to C-termini 2 to about 100 amino acids can be assembled to build a peptide.

Setting up such a peptide synthesis system requires a proof that it can create specific peptides. This is usually performed by synthesising some short peptides whose behaviour is known in advance. Among the most common ones are the FLAG-epitope, a peptide consisting of eight amino acids (octamer) and the HA-epitope, consisting of nine amino acids (nonamer). Both of them can be detected by coupling them to antibodies (anti-FLAG, anti-HA) which can be marked by fluorescent molecules (for an example see [Kön10]). After binding, the excess antibodies that were not bound to the epitopes are discarded, so that the fluorescence scanner being sensitive to the wavelength of the fluorescent light of the marker molecules can determine those peptides at which antibodies have been bound. As anti-FLAG does not couple to the HA-epitope and vice versa, synthesising both peptides and exposing them to both antibodies, the functional evidence can be adduced by means of a fluorescence scan.

1.2.2 Peptide synthesis on CMOS chips

The liquid transport of the amino acids to their synthesis sites bears some disadvantages. First of all, the form and extension of a liquid drop on a support is hard to control. This means that synthesis spots for different peptides must be positioned on the support in a distance far enough that drops to build different peptides cannot overlap accidentally. Furthermore, there is a minimum size of the drops below which the drop evaporates before it can reach the surface and, at last, there is no control over the amino acids in the drop as they couple to the growing peptide upon contact to the support. This means that incorrectly deposited drops will immediately lead to incorrectly assembled peptides at that spot.

In order to overcome these disadvantages, the particle based transport was developed in [SFB⁺08]. Similar to toners that are used in laser printing, these particles or toners consist of 89% of a solvent (resin) and 10% of the particle mass are preactivated amino acids (Fmoc-amino-acid-OPfp-ester) which are relatively stable and commercially available, see [SFB⁺08] for details. The remaining 1% consists of charge control agents that allow for a tribo-electrical charge of the particles and dye for better visibility, as the resin is transparent. These particles have to be compatible with all steps of the Merrifield synthesis to replace the liquid transfer medium. As the resin is solid at room temperature, using particle based amino acid transport solves the problems occurring with liquid based transport.

Inspired by colour laser printing, a peptide printer using particles with mean diameter of $10\mu\text{m}$ was built to transport the amino acids to their synthesis sites, as described in detail in [SFB⁺08]. By means of this, particle heaps containing one sort of amino acid instead of liquid drops form one synthesis spot. If the deposition of the particles was successful, they are heated at 90°C to liquify the resin such that the amino acids contained can couple to the support. Then, the chemical washing, capping and deprotecting steps of the Merrifield synthesis can be performed before new particles containing the second amino acid are printed on the support.

Yet, as with all printers, the alignment of the printing heads is the limiting factor of the distance between the synthesis sites for different peptides, as the overlap of subsequent particle depositions must be guaranteed in order to correctly assemble the desired peptide. To further reduce the distance between the synthesis spots and to increase deposition precision, the method of selective particle deposition on electrically programmable surfaces was set up and optimised by [Kön10]. Contrary to the printing technique, the method is based on the principle that the synthesis spots themselves attract the tribo-electrically charged toner particles. This was implemented by designing a CMOS chip with squared synthesis spots that can be put on an electric potential and are separated by a chemically inactive grid electrode, see [KBN⁺10] or [Kön10] for details about the chip design. Each spot (also called pixel

or chip cell) can be selectively switched on or off to attract the negatively charged particles. The latter are charged and blown towards the chip surface in a specific aerosol deposition apparatus which is further described in [LWK⁺10]. As soon as each synthesis spot has attracted its required amino acid, the support can be heated so that the resin liquifies and the process already described for the peptide printer can be applied. The process is summarised in Figure 1.2.

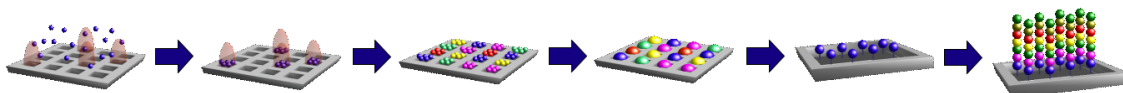


Figure 1.2: Left to right: Solid phase peptide synthesis on CMOS chips: Negatively charged micro-particles containing amino acids are guided to their synthesis sites by the selectively programmable electric fields of the CMOS chip; Particles containing one specific sort of amino acids are deposited; Particles containing all sorts of amino acids for the first layer of the peptide array are deposited; Heating initiates the coupling of the first layer of amino acids to the surface; Remnants of particles and excess amino acids are discarded in chemical washing; Repeating this procedure leads to a complete peptide array (source: [BNB⁺07])

1.3 State of the art quality analysis

Quality checks on each component of the peptide synthesis process are regularly performed. The particle size can be measured by means of a Malvern Mastersizer and its form by raster electron microscopy. The percentages of the particle contents are monitored by high-performance liquid chromatography (HPLC).

The deposition quality is checked by experts that investigate the spots on the support under a light microscope to find contaminations and to determine whether enough particles are deposited on each synthesis spot to guarantee a good coupling efficiency. If the spots are homogenously covered with particles, the experts assume that this deposition will result in a good coupling efficiency, having already achieved good synthesis results with such depositions, as demonstrated in [BNB⁺07].

The coupling efficiency itself can be estimated in the deprotection step, when the Fmoc groups are removed from the amino acid that has just coupled to the growing peptide. The reaction product of the Fmoc group with the washing solution absorbs UV-light, so that UV spectrometry of the solution yields an estimate of coupled amino acids on the entire support. Yet, this gives no information about the number of coupled amino acids per synthesis spot, only averaged over the entire array. Practical experience, documented in [Blo09], furthermore shows that this

estimate is inaccurate and can have error rates of about 50%. Experiments to investigate the error in [Blo09] revealed that the measurements are at the detection limit of the spectrometer and additionally depend on the amino acid the Fmoc group was removed from and on the chemicals used for washing.

After termination of the entire process, a selective antibody with a fluorescing molecule on top can be coupled to each peptide to measure the fluorescence signal as evidence for a correct synthesis. Furthermore, there is the possibility to perform a mass spectrometric analysis to investigate the synthesis quality of the assembled peptides. This was performed in [Bey05] and [Blo09]. The measurements were not spatially resolved, i.e. analogously to the coupling efficiency test, the mass spectrometry performed could only provide results for the entire support.

1.4 Quality analysis by image processing

In order to improve the state of the art quality analysis methods in the *Peptide Chip Project* described in Section 1.3, this thesis focuses on the application of image processing (as a non-destructive evaluation tool) to create quality standards that are more robust. Inspired by the related work in biology, introduced in Section 1.4.1, Section 1.4.2 discusses the requirements and experiments that are necessary to develop such a quality analysis system for the peptide synthesis technique based on CMOS microchips.

1.4.1 Related work in biology

Using image processing tools for the investigation of phenomena in life sciences has become standard in the past few decades [FMG89, TN97, MLM⁺01]. High uncertainty and variability in biological samples, however, require statistical evaluations over large data sets and systematic analyses of similar structures. Both aspects are taken account of in array experiments that allow for the simultaneous creation and test of a large number of similar structures. The miniaturisation of those arrays, leading to micro arrays (sometimes also called *bio-array* experiments), is often inevitable due to the limited amount of test material, e.g. blood of patients, and the production costs.

Depending on the nature of the biological question, *generative* and *discriminative* models have to be distinguished. While the former are of descriptive kind, model entire cellular processes and usually include a vast number of possible variables, parameters and prior information to find the desired solution, the latter can be solved efficiently by training a so-called decision function to check whether a given hypothesis is true or not, which only requires those parameters and variables found relevant to the specific problem during the training procedure.

One prominent micro array experiment that makes use of image processing analysis for both, generative and discriminative models, is the concept of a DNA-array consisting of spots with different genes or other short sequences of DNA in order to investigate gene expression and thereby get information about the activity of genes for different cell types in various states under certain environmental conditions [LW00].

In this setup, image processing can contribute to the analysis in several ways:

- Average array sizes of more than 10,000 cDNAs per cm^2 require an automated readout of the results and structuring of the acquired information as manual evaluation is tedious and often inconsistent [WGG01].
- Having a hypothesis about the activity of specific genes under the experimental conditions, image processing can evaluate relative differences² between gene affinities in a single array experiment, i.e. a discriminative model can be defined that decides, for instance, which gene is the most probable to be expressed under certain circumstances.
- Without a definite hypothesis at hand, image processing, especially (unsupervised) clustering techniques, as for instance, k-means clustering, can help to group similar reactions in an array experiment of various genes in order to find good hypotheses for detailed testing.
- In automated production chains of arrays which are to be used in clinical tests, image processing can help to control the quality of DNA-arrays in order to guarantee the spot density and correct location of the genes, so that the tests to be performed on these arrays will yield a significant and reliable result with certainty.

The role of image processing and the specific algorithms used in DNA-array analysis can serve as an example from which possible applications for peptide arrays can be deduced. This is possible as both cases consist of combinatorial variations of structures in a periodic pattern of similar extend with required densities and both are used either for hypothesis testing or finding a good hypothesis in the first place.

While the concept of DNA-arrays includes image processing in many ways, querying the literature yields only a few image processing supported peptide array experiments. In most cases, mass spectrometry data is analysed by means of pattern matching in order to identify peptide sequences for quality analysis [GWPH93]. Other possible applications can be found in analysing the images of chemical detection reactions, see 5.4. of [Blo09] for an overview.

²absolute differences are often hard to measure with certainty due to the semi-quantitative nature of most bio-chemical detection reactions

Yet, up to now, all image processing tasks are only used in the evaluation step of the array experiment. As this thesis will show by setting up a quality analysis system, image processing can also be used for directly surveilling the array production and its quality (e.g. controlling the relevant quantities in order to obtain high spot density or density of correctly assembled peptides per spot).

1.4.2 Application to peptide chips

The main part of the thesis focuses on the integration of image processing based methods into the process described in Section 1.2.2 to set up a quality analysis system for the peptide array production that uses solid micro-particles as transport medium. So far, the technique based on solid micro-particles, described in Section 1.2.2, is the only one that enables quality analysis already during the in situ synthesis process after each deposition step. Using liquids as transfer medium for the amino acids, the amino acids couple immediately after contact with the support surface. This direct coupling in liquid then requires purification steps that reduce the density of assembled peptides in one spot. Being able to detect contamination before coupling, it is possible to preserve the initial spot density by repeating or repairing the erroneous deposition. Furthermore, controlling the amount of amino acids deposited on each spot, deficiencies in coupling can be detected and corrected in each step. While the former is a unique feature of the solid micro-particle synthesis, the latter also seems feasible for liquid amino acid transfer using the same image processing techniques. Considering bio-chemical or surface analytical methods for direct quality analysis, currently, there are no options for a non-destructive, reversible and significant detection reaction of the amino acid density per spot. Surface analysis techniques often destroy the uppermost molecular layers, bio-chemical reactions are either irreversible or not sufficient for the small molar densities of amino acids per spot.

Investigating the synthesis protocols, the following requirements on the array production process have to be fulfilled in order to guarantee a reliable quality control:

- *Reproducibility* (ensuring that the outcome does not change as long as the production conditions are not altered)
- *Continuity* (maintaining constancy in the production conditions)
- *Robustness* (ensuring that slight changes in the conditions do not yield a totally different outcome)

Under these conditions, quality analysis in the array production process can be modelled as a discriminative decision model that determines the amount of contamination and deficiencies after each toner deposition step and then evaluates the deposition and future coupling by comparing the current toner deposition with a trained deposition for which coupling analyses have been performed. If the quality

of the array in the analysed step is satisfactory, the next step can be started, if not, the step has to be improved or repeated. Being in accordance with [Vap95], "*avoiding the solution of unnecessarily general problems*", the advantage is that this allows for fast and efficient decisions that can be used in automated process chains, yet, this comes at the cost of thorough elementary research in the training procedure, when it comes to determining the important variables and parameters.

Having determined the decisive parameters, training them requires a certain amount of data because this has been proved to yield good results in practice, only if the training data was representative for the underlying data structure [Nal04, LHW06]. Due to the periodic structure and the construction of the image acquisition setup for the quality analysis of particle based peptide array production, the training sample can be obtained by taking one picture only, contrary to most bio-chemical and medical applications of image processing, where thousands of sample images have to be analysed in the training step.

One standard method for training a classifier in image processing is to use supervised (already classified) training samples to decide for the test samples, e.g. by means of a Support Vector Machine (SVM), which is also used for the quality analysis of peptide array production, or a k-nearest-neighbour classification. Hereby, the classification for the training samples is often acquired by expert labelling or some other artificially produced ground truth. In the case of the peptide chips, this implies the creation of a correlation between the deposition of the toner particles and the coupled amino acids to the growing peptide on the support. Before image processing was taken into consideration, state of the art quality analysis was performed by expert analysis of all deposition steps. In this thesis, starting the automated quality control based on image processing, the first training also relies on expert labelling. Yet, using modern surface analysis techniques (space-resolved time of flight secondary ion mass spectrometry) in collaboration with the *Verbundzentrum für Oberflächenanalyse Münster (VOM)*, it will be shown that the training step can be improved by a physical correlation model between the deposited toner particles and the density of coupled amino acids per synthesis spot. This thus enables a non-destructive, significant, fast and reliable quality analysis for each synthesis spot on a sub-spot resolution.

Having an automated processing chain with high-precision positioning or alignment of the support in the image acquisition system³, the density of correctly assembled peptides can also be estimated basing on the toner deposition distributions recorded during synthesis using the quality analysis system set up in this thesis.

³In principle, the alignment of the acquired images could also be done by the quality analysis software, yet, this would imply using contents in the image for alignment to be evaluated, which can lead to systematic errors in the quality measure.

1.5 Outline

After this introductory motivation, the following chapters investigate each synthesis step in detail concerning reproducibility, continuity and robustness. They encompass a description of the status quo, an investigation on necessary and feasible improvements, as well as the modelling and design of suitable experiments with which the quality of each synthesis step can be controlled. Summarising the results obtained in the experiments performed in cooperation with the German Cancer Research Centre and the VOM, an automated quality analysis algorithm is implemented that is capable of evaluating and using the experimental achievements to provide optimum conditions in each synthesis step.

Chapter 2 describes the setup for deposition quality analysis developed in the course of this thesis and the requirements a setup must fulfil to automatically determine the deposition quality by means of image processing. Subsequently, in Chapter 3, models for the transport processes of micro-particles are established that can be used for deposition followed by a critical discussion on the question of quality control by parameter adjustments. Chapter 4 investigates which techniques are appropriate to measure the coupling efficiency of one layer of amino acids. Furthermore, biochemical and surface analysis techniques are tested (in cooperation with the German Cancer Research Centre and at the VOM) concerning their accuracy and spatial resolution to find a spatially resolved quality measure of the coupling efficiency per synthesis spot. Chapter 5 introduces the image processing tools that can be used for the quality control algorithm after toner deposition. Hard- and software-based image enhancement methods are investigated as well as appropriate segmentation methods and decisive features to estimate the toner deposition quality. Using the images of the deposition, quality measures are set up in a supervised training step to correlate the deposition quality with the coupling efficiency by means of the results gained in Chapter 4. Having determined this quality measure, the thereby created quality analysis algorithm is implemented and tested on experimental data, showing its flexibility by analysing images with varying setup parameters and micro-particle depositions. Finally, an interface between the quality analysis algorithm and any peptide synthesis automaton is also built. The interface is designed such that it can be integrated in any framework as a standalone module that communicates via simple ASCII-files with the main control unit of the automaton due to the fact that the synthesis automaton is still in the stage of development at the Karlsruhe Institute of Technology. After summarising the results, Chapter 6 gives an outlook of further possible applications of the quality control algorithm developed in this thesis and starting points for future research towards extensions of the algorithm to tackle the evaluation of biological questions in array experiments as well.

1.6 Publications

Parts of this work have already been published in

- K. König, I. Block, A. Nesterov, G. Torralba, S. Fernandez, T. Felgenhauer, K. Leibe, C. Schirwitz, F. Löffler, F. Painke, J. Wagner, U. Trunk, M. Hausmann, F. R. Bischoff, F. Breitling, V. Stadler and V. Lindenstruth, *Programmable high-voltage CMOS chips for particle-based high-density combinatorial peptide synthesis*, *Sensors and Actuators B* **147** (2010), 418–427
- J. Wagner, F. Löffler, K. König, S. Fernandez, A. Nesterov-Müller, F. Breitling, F. R. Bischoff, V. Stadler, M. Hausmann and V. Lindenstruth, *Quality analysis of selective microparticle deposition on electrically programmable surfaces*, *Review of Scientific Instruments* **81** (2010), 073703-1–073703-6
- J. Wagner and B. Ommer, *Efficient clustering earth mover's distance* in R. Kimmel, R. Klette, and A. Sugimoto (Eds.), *ACCV 2010, Part II*, LNCS 6493, Springer, Heidelberg, pp. 477–488 (2011)
- F. Löffler, J. Wagner, K. König, F. Märkle, S. Fernandez, C. Schirwitz, G. Torralba, M. Hausmann, V. Lindenstruth, F. R. Bischoff, F. Breitling and A. Nesterov, *High-precision combinatorial deposition of micro particle patterns on a microelectronic chip*, *Aerosol Science and Technology* **45** (2011), 65–74
- J. Wagner, K. König, T. Förtsch, F. Löffler, S. Fernandez, T. Felgenhauer, F. Painke, G. Torralba, V. Lindenstruth, V. Stadler, F. R. Bischoff, F. Breitling, M. Hausmann, A. Nesterov, *Microparticle Transfer onto Pixel Electrodes of 45 μ m Pitch on HV-CMOS Chips – Simulation and Experiment*, (accepted in *Sensors and Actuators A: Physics*)

Chapter 2

Setup analysis

2.1 Requirements

Assembling a suitable analysis station to take images of the chips during the particle deposition process requires a thorough analysis of the chip size and the particles deposited thereon.

The chip on which the particles are deposited is constructed as described in [KBN⁺10]. From the viewpoint of quality control imaging, only the size of the device as well as the extent of the substructures are of interest. Figure 2.1 shows these details of the chip.

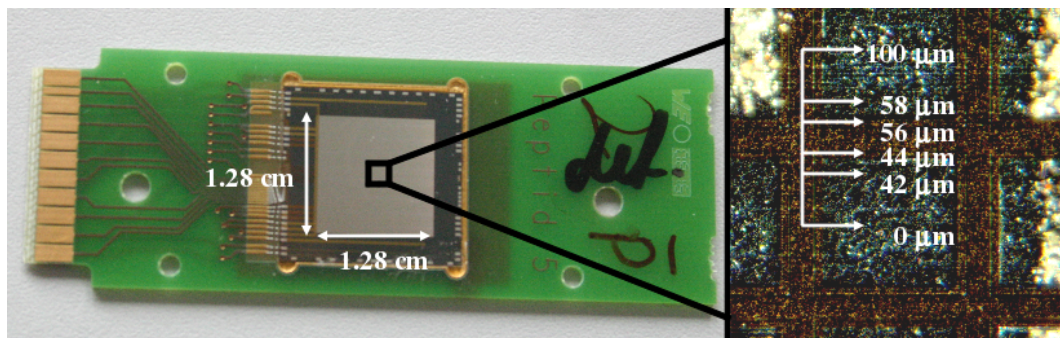


Figure 2.1: Left: Chip on board, 128 x 128 pixels Right: Pixel centre to pixel centre distance: $100\mu\text{m}$, grid width: $16\mu\text{m}$, spacing on each side of the grid: $2\mu\text{m}$

The size distribution of the particles that contain the amino acids can be measured by means of the Malvern Mastersizer. Figure 2.2 shows a typical Mastersizer analysis of a size distribution. The size measurement in this device is performed by Mie-scattering (s. [All97], 401ff for further details) of a red and a blue laser beam from a representative sample set of particles, which allows to measure particle sizes down to $0.02\mu\text{m}$ given the refraction and absorption indices of the particle mate-

rial. Since Mie-scattering bases on the assumption of spherical particles, the plotted results show the spherical equivalent diameter for irregularly shaped particles. Furthermore, the Mastersizer provides a special program to correct the deviations in the size distributions of irregular particles at wide angles, so that a relative measurement accuracy of 1% can be guaranteed.

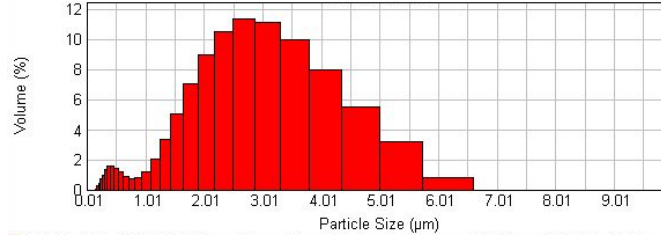


Figure 2.2: Master Sizer measured distribution of a typical sample of particles

The distribution of particle sizes is represented in form of the volume percentage for each size class of particles, which means that small particles (with small volumina) must occur numerously in order to attain the same volume percentage as larger particles.

In order to account for the detector resolution, the particle sizes are divided into logarithmically distributed classes within which the particle sizes can no longer be distinguished.

The fact that the size measurements become imprecise for particles with high aspect ratios (e.g. plates or needles) can be ignored, since the majority of particles used does not belong to this category of particles, as it can be demonstrated in electron microscopy measurements. An image of those measurements taken in the course of the standard particle quality analysis is shown in Figure 2.3.

According to Figure 2.2, a resolution down to $1\mu m$ is necessary to detect approximately 90% of single wrongly deposited particles on the chip surface. In order to reach this resolution, the setup for automated deposition quality analysis is chosen according to its magnification and theoretical diffraction limit.

Apart from requiring a high resolution, one image should display as much chip cells as possible in order to collect statistical information about the particle deposition on the cells, examine the effects near the edges of the chip and survey the overall deposition pattern at the same time. Since it is not possible to analyse an image of the whole chip in the desired resolution, as will be shown below, it is important to analyse a representative detail of the chip when investigating the quality of the particle deposition. Then, by means of the so called test error bound it is possible to find an upper bound for the misclassification probability μ on the entire chip when analysing an array of m chip cells, observing k classification errors. Assuming

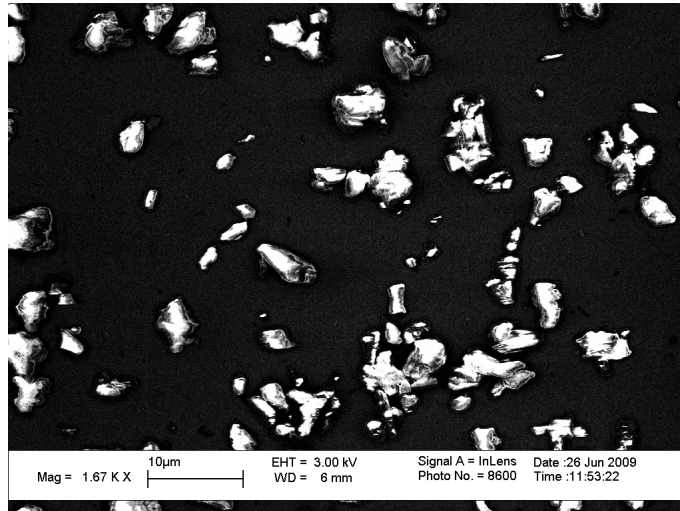


Figure 2.3: Electron microscope picture showing typical particle forms

that less than 5% of the total number of chip cells, i.e. less than 820 out of 16384, are in the sample to be analysed, then a simple model for the misclassification is a Binomial probability distribution¹ $\mathcal{B}(k, m, \mu)$. Then, the probability of observing not more than k errors in the sample is given by

$$\mathcal{CB}(k, m, \mu) = \sum_{j=0}^k \mathcal{B}(j, m, \mu) \quad (2.1)$$

Inverting the statement with respect to the unknown classification error μ

$$\hat{\mu} = \max_{\mu} \{ \mu \mid \mathcal{CB}(k, m, \mu) \geq \delta \} \quad (2.2)$$

one searches for the largest misclassification probability $\hat{\mu}$ such that for all $\mu \geq \hat{\mu}$ the probability of detecting k errors is smaller than δ . Thus, the true μ is smaller than $\hat{\mu}$ with probability of at least $1 - \delta$. For zero classification errors, the misclassification probability can be calculated in closed form as

$$\mu \leq 1 - \sqrt[m]{\delta} \quad (2.3)$$

Detecting for instance no classification errors in a sample of 160 cells, then $\mu \leq 2.84\%$ with a probability of at least 99%.

In the case of one classification error in a sample of m cells, the solution cannot be

¹The correct model for a sampling without putting the analysed samples back in the set would be a hypergeometrical distribution, but the latter does not differ significantly from the much simpler Binomial distribution for sample sizes of less than 5% of the total set size.

written down explicitly. The equation which has to be solved for predefined δ and m reads

$$(1 - \mu)^{m-1} \cdot (1 - \mu(1 - m)) \geq \delta \quad (2.4)$$

Hence, detecting one classification error in the sample of 160 cells, $\mu \leq 4.08\%$ with a probability of at least 99%.

Although this simple model does not take into account the effects at the edges of the chip or systematic misclassifications, it gives a good estimate about the number of chip cells to be analysed, hence a theoretical, lower bound for the misclassification error in order to provide a fast and stable evaluation.

Since the particles are not fixed on the chip surface, the optical imaging should be contactless, without interference of additional material. For instance, electron microscopy cannot be used due to the required sample preparations such as chemical fixation and embedding, although it is capable of reaching the resolution. The same reason also bars the use of oil immersion objectives for microscopes.

Furthermore, the optical setup should be robust in handling, so that all images are taken under comparable conditions. This especially concerns the illumination of the chip, but also the recording device, e.g. the optical alignment of the camera or the constancy of the readout noise of the CCD or CMOS camera during the whole operating time.

It also should be easy to use and provide fast access to the images for human quality monitoring as well. This also means that the data format of the images should be one of the common image formats as PNG, JPEG or TIFF that can be opened by any standard image viewer.

The experimental tests carried out in the course of this thesis and summarised in the next two sections 2.2 and 2.3 will show that there are two possible setup choices that are capable of fulfilling these requirements in theory, namely a consumer digital single-lens reflex camera or a microscope with high-resolution objectives. Their respective data formats will then be discussed in Section 2.4.

2.2 Camera setups

The easiest way of controlling the particle deposition is to take a picture of the chip by means of a consumer camera and a macro objective. In order to achieve a resolution in the range of micrometres, the following theoretical background has to be taken into consideration.

The f-number N is defined as the quotient of the focal length f by the diameter of the entrance pupil D . Together with the magnification m and the wavelength

λ of the light used in the setup the f-number determines the diffraction limit that sets the theoretical lower limit for the particle sizes to be detected. As the size of the chip is only 1.28 centimetres which is far less than the 24 centimetres minimum distance from the objective to the chip, the far-field approximation of the diffraction limit

$$d_C = \lambda \cdot N \quad (2.5)$$

gives a good estimate of the resolution. Using white light as illumination, the wavelengths are roughly in the range from 400 to 700nm. Inserting 3.2 as f-number in Equation 2.5, the resolution is $2.24\mu\text{m}$ for the longest wavelength of the spectrum. Knowing the optical diffraction limit, the camera has to be chosen such that it does not further degrade the resolution. Assuming a maximal uncertainty of one pixel in each direction, the required pixel size is given by

$$p_C = \frac{1}{2} d_C \cdot m \quad (2.6)$$

A magnification $m = 5$, for instance, requires pixel sizes of about $5.6\mu\text{m}$. A setup of a Canon EOS 1000D with a CMOS chip of $5.71\mu\text{m}$ pixel length and a Canon MP-E macro objective with a magnification of 5 comes very close to these requirements.

For practical tests the camera and the objective are connected to a stand with balancing weights as shown in Figure 2.4 at the bottom. The figure additionally shows a typical image detail taken by the camera setup. As can be observed in the upper right corner, the borders of the image are blurry due to optical aberrations. Hence, only parts of the image taken can be used for quality analysis. Since the total amount of chip cells on one image is about 1300, taking only a detail does not deteriorate the analysis but complicates the handling of the setup.

Furthermore, the Canon MP-E macro objective with zoom factor five is the best lens which is currently on the market, i.e. all consumer camera setups are limited by this factor, so there is no further magnification available for more detailed images of fewer chip cells.

The option to increase magnification by means of bellows does not yield good results. Due to the increased path length for the light, the focal plane does not coincide with the camera chip plane anymore. This means that the image deteriorates even more with the length of the bellow and higher resolutions cannot be achieved.

Combining a converter with a magnification factor of two with the macro objective, a magnification factor of ten can be reached theoretically. Measuring the effective resolution², however, shows that the obtained $5\text{-}10\mu\text{m}$ are not better than the resolution of the original setup without converter.

²The effective resolution, which is dependent on the illumination conditions, is defined as the size of a particle that can be **clearly** recognised by a human being investigating the image.

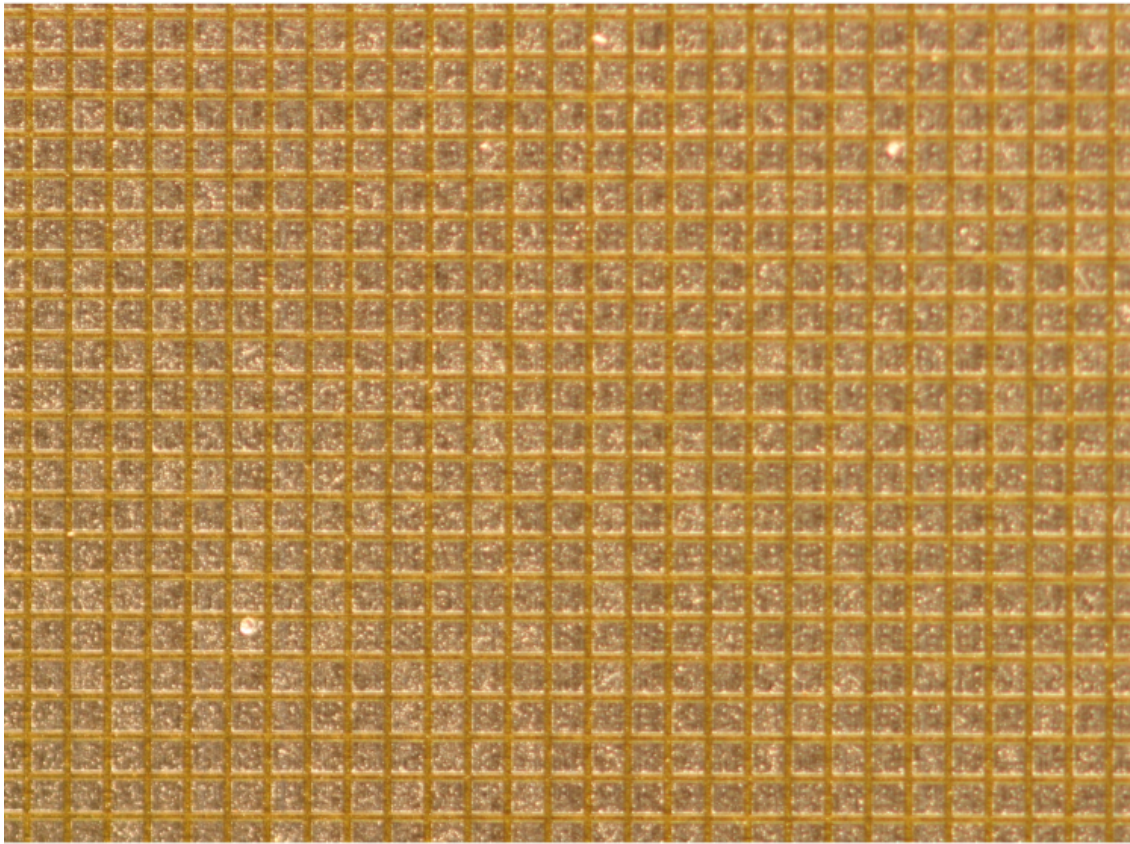


Figure 2.4: Top: Chip detail image taken with this setup Bottom: Camera setup

2.3 Microscope setups

The second option for an image acquisition system is the usage of a microscope and a microscope camera. For the microscope, the theoretical diffraction limit is given by

$$d = \frac{\lambda}{2 \cdot N_A} \quad (2.7)$$

where λ is the wavelength of the light used (in this case white light with wavelengths between 400 and 700nm) and N_A is the numerical aperture. The latter is usually indicated on the microscope objective. For a $1\mu\text{m}$ diffraction limit for light with 700nm wavelength, N_A has to be 0.35.

The microscope available is a Zeiss Axiovert 35 for which the objectives as listed in Table 2.1 can be used.

magnification	N_A	#(cells) per image
5	0.12	475
10	0.25	130
20	0.3	24
40	0.6	6

Table 2.1: Available microscope objectives for the Zeiss Axiovert 35 and the number of chip cells they can picture per image

Apart from the diffraction limit, the number of cells that can be captured in one image is also an important parameter for the setup. In order to increase the number of cells per image, a TV-adaptation of magnification 0.63 is installed on the C-mount where the microscope camera is placed. The number of chip cells per image is listed in Table 2.1 and takes this adaptation already into account.

Table 2.1 shows that the setup with the magnification of 10 capturing 130 chip cells per image is closest to the requirements previously defined. The actual diffraction limit for this objective is $1.4\mu\text{m}$ for light with $\lambda = 700\text{nm}$.

Instead of using a TV-adaptation it is also possible to reduce the magnification in order to observe the required amount of chip cells but the available objective for the magnification of factor 5 has $N_A = 0.12$. Hence, the theoretical diffraction limit for these objectives is comparable to that of the consumer camera and therefore not suitable.

As Figure 2.5 shows, the images acquired by the microscope setup fulfil the requirements previously mentioned in Section 2.1. Yet, for automation the microscope setup is not applicable due to its high purchase costs. So the setup presented here

can only serve as a prototype for scientific testing purposes. Developing an automated peptide synthesis machine in mass production, cheaper constructions have to be considered. One goal of this thesis is to determine such a minimal image acquisition setup given the requirements. From the hardware side of the image acquisition, the microscope could be reduced to a fixed magnification with one objective only. The high-precision mechanical positioners for object adjustment could also be discarded if the machine already provides high-precision positioning and adjustment in the transport process of the chip through the assembly line.

Another very important issue is the illumination of the peptide chip in order to obtain a good image quality. Light sources which are mounted on the microscope perpendicular to the chip surface such that the light directly falls onto the chip are unsuitable. In this case the reflections and scatter light from the chip surface itself impede the analysis of the particle deposition. Using lamps mounted at small angles parallel to the chip surface yields better results.

Furthermore, light intensities should be tuned such that the entire chip surface and the deposited particles are well illuminated but scattered light from reflections of the chip surface is kept as low as possible. The setup described in this thesis includes a gooseneck lamp with a luminous emittance of 18Mlx maximal illuminance at the entrance of the optical fibre. Mounting this one to two centimetres far away almost parallel to the chip surface a satisfactory illumination is achieved as can be seen in Figure 2.5.

Concerning the image acquisition device of the microscope, several cameras can be used. The only restrictions they have to match is that the resolution of the camera chip should be better than the optical resolution limit discussed above and that the camera chip is large enough to capture the required amount of chip cells. Expensive microscope cameras like the DXC-950P from Sony with three CCDs or the Progres C5 from Zeiss yield images of high quality that fulfil all criteria stated above. Consumer cameras like the EOS 1000D from Canon, however, are also capable of producing images of satisfactory quality when integrated in the microscope setup. Yet, using the Axiovert 35, the integration of a microscope camera is more suitable since no accessories like TV-adaptations or lenses are available for consumer cameras which could provide the analysis of the required amount of chip cells per image.

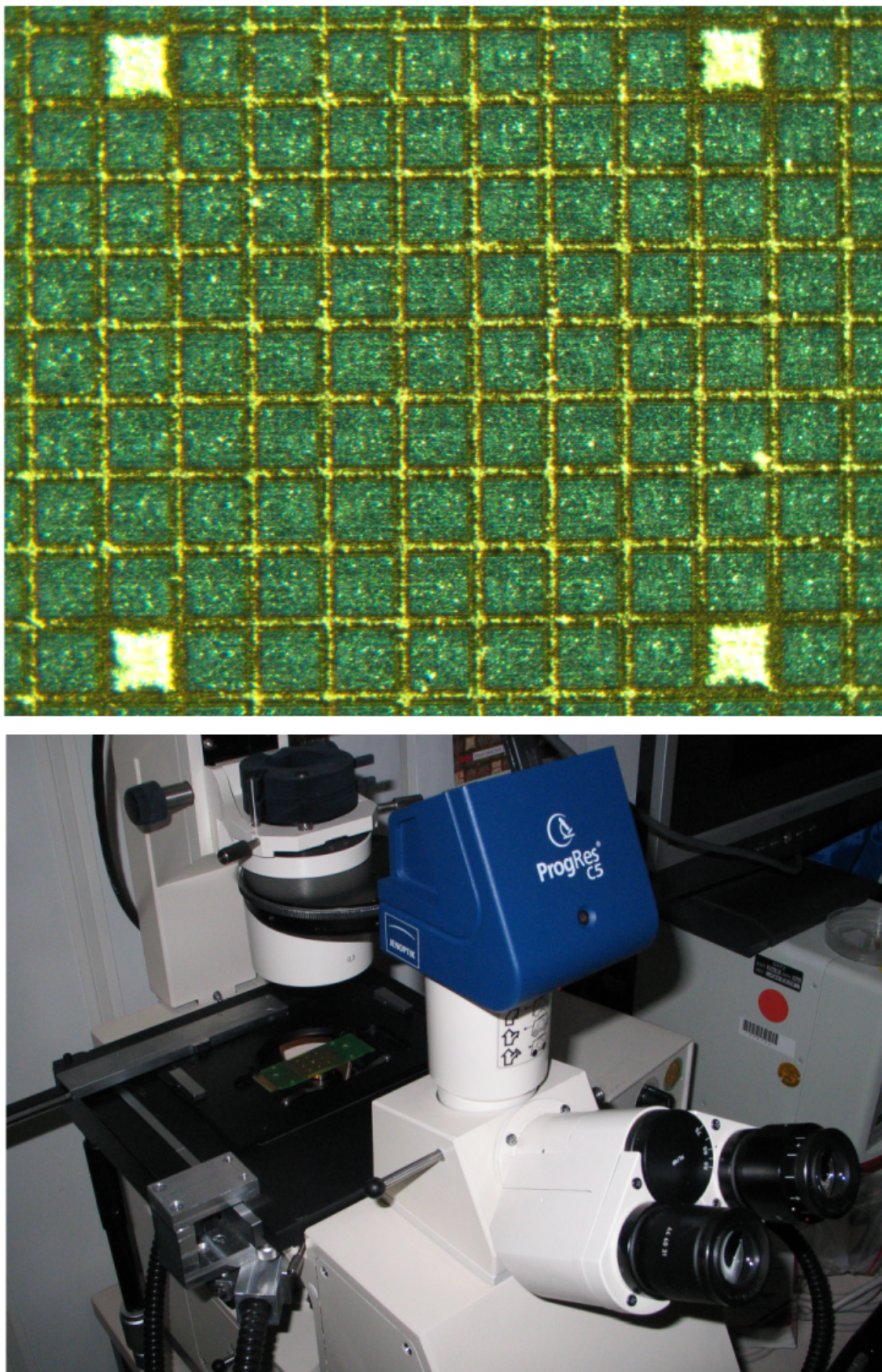


Figure 2.5: Top: Chip detail image taken with this setup Bottom: Microscope setup

2.4 Image formats

Depending on the camera used, several different image formats can be chosen to save the image. Concerning the format, the requirements for a fast high-quality image analysis are the following. The image quality should be kept, such that the quality loss due to data compression is smaller than the resolution set by the optical diffraction limit. This can only be achieved at the cost of providing the required amount of disc space. Simultaneously, the image format should be chosen such as to save as much memory as possible yielding the same image quality.

Considering the worst case scenario, that the combinatorial peptide synthesis to be performed consists of 20 amino acids per peptide and each synthesis layer contains all 20 amino acids, then 400 images have to be taken. Assuming that each of these is of the highest quality in TIFF of 20MB size, then the total amount of disc space needed for this synthesis is less than 10GB.

Other lossless formats to save the images in are the BMP- and the PNG-format. The BMP-standard does not compress the data at all, which results in a high amount of used disc space, similar to the lossless TIFF. Contrary to that, PNG applies a lossless prefiltering and DEFLATE-compression of the data, such that the amount of required disc space can be reduced.

Using the JPEG-format to save the images instead, the lossy compression of this format operates on blocks of 8 x 8 pixels, i.e. quantisation and compression artefacts are expected on the scale of the size of these blocks. If the optical resolution is worse than that, the JPEG-format can be used. As an example, consider the setup with an objective of magnification 10 with $N_A = 0.25$, a TV-adaptation of 0.63 and a Progres C5 camera. As previously calculated, it has an optical resolution of $1.4\mu\text{m}$ which corresponds to a resolution of 2.6 pixels on the camera chip. Ergo, the JPEG-format cannot be used to save the images in this case.

Chapter 3

Aerosol dynamics

3.1 Related work

In general, streaming systems consisting of a combination of several materials in different states can be modelled as multiphase flows, a special case of which are two phase flows. Depending on the states, gas fluid, gas solid or fluid solid two phase flows can be distinguished. All theoretical aspects, characterisations of the flow regimes and solutions to simple example configurations are presented in [Kle03], which gives a good introduction to the topic.

A vast number of examples for multiphase flows can be found in industrial applications. Chemical processes like combustion require multiphase flow models of the reactions or the emission of their products. The latter case is an example of gas solid or gas fluid flow. A comprehensive study of dispersed two phase flows including particles or bubbles, combined with the development of numerical simulation software can be found in [Nie09], examples in which these models are applied to special processes are [WJZ09] or [PL98].

In an environmental and geophysical context, these models contribute to the investigations of dust storms. An overview over this topic is given in [Mar70]. Furthermore, sedimentation processes can also be described by two phase flows as shown in [Gre02].

While the focus of these applications lies on the flow of the materials to determine their paths, including electric fields in two phase flows plays an important role in applications that aim at selectively depositing the material to be transported. These are probably the most similar works to the one underlying this thesis. In [SW72], the flow of a turbulent airstream through a pipe containing charged micro-particles of potassium chloride is investigated experimentally and the results are compared to theoretical expectations. More recent research of charged micro-particles was carried out in [FKS02], in which possibilities of selectively depositing charged micro-

particles on an array of conductor spots out of a solvent are analysed. In the setup described in [FKS02], single particles of $10\mu\text{m}$ diameter are precisely deposited on the attracting spots with an even smaller diameter. Due to these proportions in size, particle-particle interactions at one spot are taken into account. A selective particle deposition method out of an aerosol for nano-particles is described in [LYW⁺09], in which electro-sprayed nano-particles were analysed in view of electrodynamic focussing for a precise deposition.

3.2 Experimental setup

The deposition of the triboelectrically micro-particles onto the peptide chip strongly depends on the characteristics of the particles, i.e. on their form, size and charge or alternatively their q/m -value. The form and size analysis methods were already discussed in Section 2.1 in the context of imaging. As the results showed, the number of irregularly formed particles is small so that the toner particles will be approximated as spherical objects with radius r_p in the following. Concerning their charge, experimental tests performed in [Löf09] revealed that the q/m -value ranges from -2 to -5mC/kg for particles with mean diameters of 5 to $10\mu\text{m}$ with increasing q/m -value for a decreasing particle size. For these particles, a deposition setup has to be constructed such that they can be selectively deposited.

From the experimental side, optimisation in the design of the particle deposition chambers was performed by replacing the deposition chambers used in [NM06] and [Kön10] by smaller ones. Their setup is simpler yet comes at the cost of not being able to recollect particles not deposited on the chip and reuse them. Figure 3.1 shows the working principle on the left, followed by the actual setup and the outline of the particle outlet on the right.

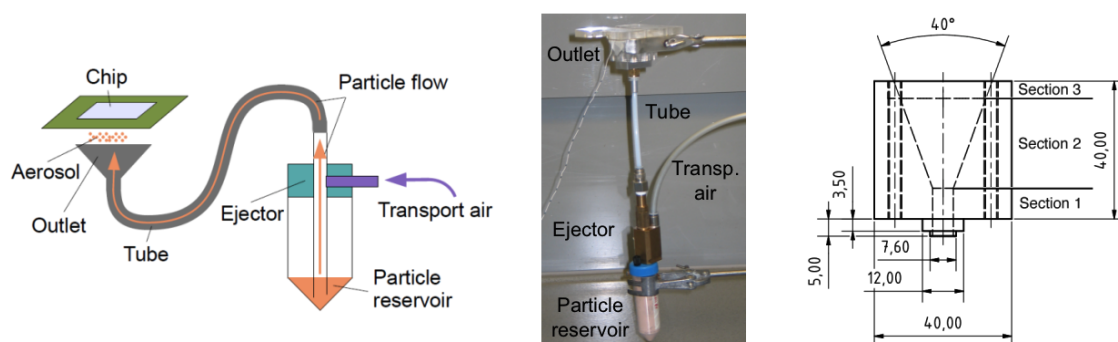


Figure 3.1: Left: Schematic of the particle deposition chamber Centre: Experimental setup of a particle deposition chamber for one sort of amino acid particles Right: Outline of the particle outlet (source [WKF⁺10])

Driven by the pressurised transport air, an air ejector draws amino acid particles out of the reservoir tube and generates the aerosol. The latter is then guided through a tube where the particles become negatively charged due to interactions with the tube walls and the surrounding air. At the transition from the tube to the particle outlet, a fine-meshed sieve is installed and usually set to high negative voltage (approximately -1kV). By means of the sieve, particle agglomerates larger than the mesh size, positively charged particles and negatively charged particles with only a small negative charge are selected out. Furthermore, the negatively charged particles that can transcend the sieve are accelerated towards the chip by the electric field of the sieve, so that less particles are sedimented on the walls of the deposition chamber.

Using these experimental prerequisites, theoretical models can be set up to describe the particle deposition in this geometry. In the course of quality analysis, however, other (simpler) setups are tested as well for their suitability. Therefore, this chapter investigates several different transport models and compares them according to their deposition time or the precision up to which the deposition can be predicted. In contrast to [NM06], [Löf09] and [Kön10], the focus is set on theoretical descriptions to complement their work.

3.3 Prerequisites

3.3.1 Characteristic numbers

1. **Reynolds number:** The Reynolds number Re is defined as the ratio of the inertial forces to the viscous forces. It is a measure of the degree of turbulence for a fluid with density ρ and dynamic viscosity μ streaming at mean velocity v in a pipe with diameter D or streaming with a volumetric flow rate R through a pipe with diameter D and cross-sectional area A

$$Re = \frac{\rho v D}{\mu} = \frac{R D}{\mu A} \quad (3.1)$$

If the vessel does not have the form of a cylindrical pipe, the Reynolds number is computed using the equivalent diameter of that form. The equivalent diameter of a conic vessel with maximum and minimum diameter D_{max} and D_{min} , for instance, can be defined as

$$D = \frac{D_{max} + D_{min}}{2} \quad (3.2)$$

Assuming that the particles are transported to the synthesis sites by air, the Reynolds number for the airstream at room temperature ($T = 293\text{K}$) in the vessel shown in Figure 3.1 on the right can be calculated for the three

sections with their equivalent diameters $D_1 = 7.6\text{mm}$, $D_2 = 16.3\text{mm}$ and $D_3 = 28.5\text{mm}$. The density of air at 293 Kelvin is given as 1.204kg/m^3 and the viscosity can be calculated using Sutherland's formula for an ideal gas

$$\mu = \mu_0 \frac{411.15 \text{ K}}{T + 120 \text{ K}} \left(\frac{T}{291.15 \text{ K}} \right)^{1.5} \quad (3.3)$$

Inserting $\mu_0 = 18.27 \cdot 10^{-6}\text{Pa} \cdot \text{s}$ the formula yields $\mu = 18.36 \cdot 10^{-6}\text{Pa} \cdot \text{s}$. Hence, the Reynolds numbers depending on the airstream velocities can be calculated as

$$Re_1 = 498.4 \frac{\text{s}}{\text{m}} \cdot v \quad (3.4)$$

$$Re_2 = 1068.9 \frac{\text{s}}{\text{m}} \cdot v \quad (3.5)$$

$$Re_3 = 1869.0 \frac{\text{s}}{\text{m}} \cdot v \quad (3.6)$$

which shows that for typical velocities of about 0.1m/s the airstream can be characterised as laminar, while for higher velocities of more than 1m/s , the flow can enter the transition regime between laminar streaming and turbulence¹.

2. **Knudsen number:** The Knudsen number Kn is defined as the molecular mean free path length λ divided by the characteristic length of the vessel L , which is, in the case of the pipe the diameter and for the cone the equivalent diameter, accordingly

$$Kn = \frac{\lambda}{L} \quad (3.7)$$

assuming air to be an ideal gas, the formula reads

$$Kn = \frac{k_B T}{\sqrt{2} \pi \sigma^2 p L} \quad (3.8)$$

where k_B is the Boltzmann constant, T is the absolute temperature, σ the hard shell diameter of the molecules and p the pressure. Comparing the mean free path with the characteristic length of the vessel, it is a measure to decide whether the continuum mechanical or the statistical mechanical approach is the correct one.

Assuming $\sigma = 3.62 \cdot 10^{-10}\text{m}$ as mean size of the air molecules and a pressure of 1bar (i.e. 10^5Pa) at room temperature, 293K , the mean free path of the air molecules can be calculated as $0.0695\mu\text{m}$, which is quite small even compared to the characteristic radius of the particles $r_p = 1$ to $5\mu\text{m}$. This indicates that the air can be modelled as a continuous fluid streaming around the particles. Furthermore, Brownian motion can be neglected as only particles in the sub-micrometer range are significantly affected by this motion.

¹Typical transition numbers lie between $Re = 1000 - 2000$ but depend strongly on the geometry and characteristic lengths.

3. **Mach number:** Being the ratio of the mean velocity v to the velocity of sound v_s , the Mach number Ma decides, whether effects due to the compressibility of the fluid occur or not.

$$Ma = \frac{v}{v_s} \quad (3.9)$$

The velocity of sound for air at room temperature and a pressure of 1bar is 343m/s, which is more than 1000km/h. The particles in the setup considered here that travel with the airstream are much slower than this with certainty. Thus, it can be concluded that compressibility does not play a role for the air in the particle deposition chambers considered here.

Combining the Reynolds, Knudsen and Mach number into one equation yields

$$Re = \frac{Ma}{Kn} \cdot \sqrt{\frac{\gamma\pi}{2}} \quad (3.10)$$

where γ is the adiabatic index, 7/5 in this case.

4. **Euler number:** The Euler number Eu is a measure for the losses of flow due to friction and is defined as

$$Eu = \frac{p_{up} - p_{down}}{\rho v^2} \quad (3.11)$$

where p_{up} and p_{down} are the pressures upstream and downstream, respectively, ρ is the density of the fluid and v its mean velocity.

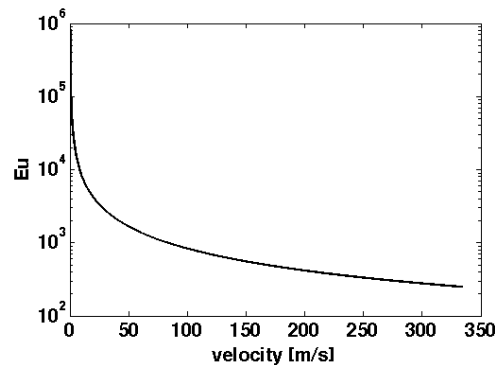


Figure 3.2: Euler number for velocities from 0 to 343m/s for $p_{down} - p_{up} = 1\text{bar}$

The upstream pressure can be determined by the pressure at the outlet valve of the particle pump, e.g. $p_{up} = 2\text{bar}$, which is a commonly used value, while the downstream pressure must be the pressure in the laboratory, $p_{down} = 1\text{bar}$, since the deposition chamber is constructed such that air can escape at the borders of the chip during deposition (see Figure 3.1). Assuming the density

of air to be constant, 1.204kg/m^3 as before, the Euler number quadratically drops with increasing airstream velocity as shown in the graph in Figure 3.2. The graph shows that the loss of energy due to friction in the velocity regime below 100m/s is very large, hence, the airstream at the boundaries of the vessel will be drastically slowed down due to friction, which implies that the flow close to the walls can be assumed to be laminar in good approximation.

5. **Stokes number:** While the former numbers only considered the air flow, the Stokes number St is a characteristic number of the combination of the transport medium and the particles, relating the characteristic response time of the particles τ_p to the characteristic time of the particle flow τ_f :

$$St = \frac{\tau_p}{\tau_f} = \frac{16\rho_p r_p^2}{3C_D Re \mu} \cdot \frac{v_p}{D} \quad (3.12)$$

where τ_p can be derived from the equation of motion for one particle of radius r_p with drag coefficient C_D subject to Newtonian drag forces (s. Equation 3.17) in an airstream that travels at velocity v :

$$m \frac{dv_p}{dt} = \frac{1}{2} C_D \pi r_p^2 (v - v_p) \|v - v_p\| \Rightarrow \frac{dv_p}{dt} = \frac{(v - v_p)}{\tau_p} \quad (3.13)$$

and τ_f is given as the time a particle needs to travel the characteristic length D at velocity v_p . Stokes numbers smaller than one characterise dilute aerosols so that settling due to gravity is unlikely and an influence of the particles on the transporting airstream is negligible.

Given $C_D = 0.1$ for a rough sphere in a laminar flow with $\mu = 18.36 \cdot 10^{-6} \text{Pa} \cdot \text{s}$, $Re = 1869$ for a velocity $v_p = 0.1\text{m/s}$ of a particle with radius $r_p = 1\mu\text{m}$ and $\rho_p = 1.1\text{g/cm}^3$ as density² in the third region of the vessel shown in Figure 3.1 with diameter $D = 28.5\text{mm}$, $St = 6 \cdot 10^{-5}$, hence the particles are diluted in the airstream.

3.3.2 Forces

Many forces can act on the particles in the airstream, as also discussed in [Kön10]:

1. The gravitational force, which is given as

$$F_G = m_p \cdot g = \frac{4}{3} \pi r_p^3 \cdot \rho_p \cdot g \quad (3.14)$$

for a spherical particle of mass m_p and density ρ_p . As analysed in Section 2.1, the particle size distribution ranges around $r_p = 1\mu\text{m}$ equivalent radius. Hence, the gravitational forces acting on a particle with this mean radius are very small, $F_G = 4.5 \cdot 10^{-14}\text{N}$, for $\rho_p = 1.1\text{g/cm}^3$.

²This density estimate bases on the fact that resin, which is the main ingredient, has a density of about this value.

2. The centrifugal force, which acts on the particles of mass m_p following a curved trajectory of transport air is defined as

$$F_Z = \frac{m_p \cdot v_p^2}{r^2} \quad (3.15)$$

where v_p is the velocity of the particles. Calculating the centrifugal force, for instance, for a curve of radius $r = 1\text{cm}$ and a velocity of $v_p = 0.1\text{m/s}$, it is one order of magnitude higher than the gravitational force. Taking the radius $r = 1\text{mm}$, it becomes three orders of magnitude higher, so that the gravitational force is negligible in this case. This force contributes to the deviation of particle trajectories from those of the air stream due to the larger inertia of particles.

3. The drag force induced by the relative motion at velocity v_p of spherical particles with radius r_p and a fluid with dynamic viscosity μ can be calculated as

$$F_D = 6\pi\mu r_p v_p \quad (3.16)$$

in the range of laminar relative particle motion. Inserting the value μ as calculated in Equation 3.3 and $v_p = 0.1\text{m/s}$, $F_D = 3.46 \cdot 10^{-11}\text{N}$ and thus comparable to the centrifugal force for $r = 1\text{mm}$.

In the case of Newtonian drag forces it reads

$$F_D = \frac{1}{2} C_D \cdot \rho \cdot \pi r_p^2 \cdot v_p^2 \quad (3.17)$$

where C_D is the drag coefficient, whose dimensionless value depends on the aerosol conditions, namely the Reynolds number. For spherical particles, it can range from 0.1 under very laminar conditions to 0.47 in turbulent flows. Since the force scales quadratically with an increase in velocity, it becomes important only at high velocities, i.e. in turbulent flows.

4. The lift force that is induced by pressure differences above and below the particles which causes the particles to move orthogonal to the streaming direction of the circumfluent air. Lift plays an important role in turbulent aerodynamics and at high Reynolds numbers and will not be considered further on, as the focus is set on the modelling of the laminar parts.
5. The electrical force between the chip and a particle in the particle cloud can be written as

$$F_E = q_p \cdot E(x, y, z) \quad (3.18)$$

where q_p is the charge of the particle and $E(x, y, z)$ the electrical field which depends on the position of the particle in the aerosol chamber. An average particle charge density can be measured according to the methods described

in [NLK⁺07a] and [NLK⁺07b] to be -2 to -5mC/kg, which was performed in [Löf09]. Yet, as the particle forms and volumes are not regular, these values only provide an estimate of the range of possible particle charges. Assuming a constant electrical field $E = 1\text{V}/\mu\text{m}$, $F_E = -2.30 \cdot 10^{-11}\text{N}$ for a particle of radius $r_p = 1\mu\text{m}$ and $q/m = -5\text{mC}/\text{kg}$.

6. The electrical force between two particles p_1 and p_2 considered as point charges at distance d can be computed as

$$F_{PP} = \frac{q_{p1} \cdot q_{p2}}{4\pi\epsilon_0 d^2} \quad (3.19)$$

which yields, for two particles of the same size and q/m -value of -2mC/kg at distance $d = 10\mu\text{m}$: $F_{PP} = 7.63 \cdot 10^{-15}\text{N}$, which is one order of magnitude smaller than the gravitational force and can thus be neglected at first order.

3.4 Models

Combining the facts and figures discussed in Section 2.1 and Section 3.3, the deposition process of the tribo-electrically charged micro-particles onto the chip can be modelled. At first, a model for the transport of particles onto the chip is defined in which the surrounding air is stationary, i.e. non-moving, and the electric field is constant. The advantage of this model is that it can be solved analytically. It can be used for the description of the far field of the chip, when the distance of the particles to the chip is larger than the extensions of the voltage pattern on its surface so that the charges of the individual chip cells add up to a homogeneously charged conductor plate. An example application for this constellation is the direct particle transfer as discussed in [NM06]. In that case, the model describes the particle trajectories for those particles that overcome the adhesive forces to the particle carrier surface. It is also valid in the proximity of one chip cell, as the extensions of the particles are small compared to the size of the chip cells. Generally, the electric field of a static charge density distribution $\rho(x')$ located at position x' in the closed volume $\Omega \subset \mathbb{R}^3$ that acts on a particle at position x is given as

$$E(x) = -\frac{1}{4\pi\epsilon_0} \nabla \int_{\Omega} \frac{\rho(x')}{\|x - x'\|} d^3x' \quad (3.20)$$

Assuming that the particle is located centrally in front of the charged area of the chip cell of $84 \times 84 \mu\text{m}^2$ as shown in Section 2.1, the distance $\|x - x'\|$ to the first neighbouring chip cell is $58\mu\text{m}$, to the second already $116\mu\text{m}$. Hence, due to the constant electric charge densities per chip cell, the contributions of neighbouring chip cells to the electric field are of higher order, so that the total electric field acting on the particle can be approximated by the one originating from one single chip cell. Concerning possible air flow in that region, the calculations for the Euler

number already showed that the loss of velocity due to friction is very high below 100m/s, hence, the velocity of air less than 84 μ m from the chip can be set to zero in a first approximation.

The second model can then be set up using a laminar flow model to transport the particles to the chip. Since a description of the full process in closed form is difficult and probably not possible at all³, approximations to the problem must be made, in order to be able to formulate a feasible model. Thus, the complete model can only be set up by means of simulation software capable of numerical integration and partial differential equation solving. One possible program, COMSOL, is introduced and the numerical methods with which the transport process can be simulated are discussed. This also allows for the investigation of more complex experimental setups as, for instance, including an approximation to the electric field of the actual field pattern instead of the constant electrical field, which yields the best approximation to the true particle deposition process. Furthermore, COMSOL facilitates the inclusion of other supplementary devices in the models, e.g. the installation of the sieve at the front of the deposition chamber (see Section 3.2), which has been shown to reduce contaminations in [Löf09]. One disadvantage of COMSOL is, however, that the effect of particles deposited onto the positively charged electrode on the particles to be deposited cannot be simulated in an easy way, as particle trajectories can only be calculated independent from each other. Hence, the simulations can only yield a first approximation to the true deposition process and cannot completely replace experimental testing but only support it. This, on the other hand, implies that models alone are not sufficient to understand and control particle transfer and guarantee good deposition qualities.

3.4.1 Homogeneous field in stationary air

The first model assumes particle transport in stationary, incompressible air as surrounding fluid medium. This assumption is supported by the calculations of the Knudsen and Mach numbers. Combining the gravitational force, the drag force and the electric force, the equation of motion for one particle in z-direction towards the chip reads

$$m \cdot \ddot{z}(t) = -m \cdot g - 6\pi\mu \cdot r \cdot \dot{z}(t) - q \cdot E \quad (3.21)$$

where m is the mass of the particle, μ is the dynamic viscosity of air, r is the particle radius \dot{z} is the velocity of the particle, q is the charge of the particle⁴ and E is the constant electric field. Here, the particles are not considered as point masses but assumed to be spherical, as the term of the drag force shows. The centrifugal force, which lies between 10^{-13} and 10^{-11} , is left out in this simple model to avoid nonlinearities. For simplicity, the index p for the parameters and variables of the particle

³see http://www.claymath.org/millennium/Navier-Stokes_Equations/ for the description of the respective Millenium Problem.

⁴Since q is negative, the electric force is antiparallel to the other two forces.

is omitted, as there is no airstream to confuse them with. Figure 3.3 summarises the setting and the forces acting on the particles.

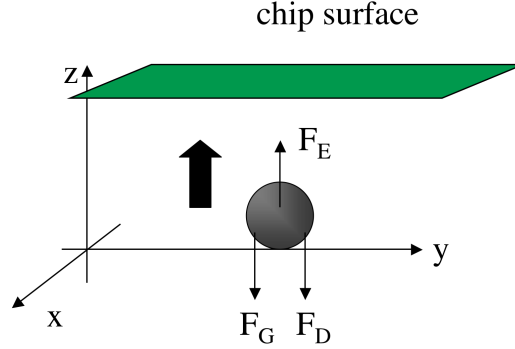


Figure 3.3: Model for the movement of spherical particles in the homogeneous electrical field of the chip in stationary air

Dividing by m and replacing the variables as follows

$$\ddot{z}(t) = - \left(g + \frac{q}{m} E \right) - \frac{6\pi\mu r}{m} \dot{z}(t) := -A - B\dot{z}(t) \quad (3.22)$$

the ansatz for the velocity dependence on time t with initial velocity v_o

$$\dot{z}(t) = -\frac{A}{B} (1 - \exp(-Bt)) + v_o \cdot \exp(-Bt) \quad (3.23)$$

can be used to determine the particle trajectory as

$$z(t) = \left(v_o + \frac{A}{B} \right) \cdot \frac{1}{B} \cdot (1 - \exp(-Bt)) - \frac{A}{B} t + z_0 \quad (3.24)$$

where z_0 is the starting point of the particle under consideration.

A encodes the gravitational and the electrical force. Under the assumption that the gravitational force can be neglected, as the calculations in Section 3.3 show, A is a measure for the electric force only. B is the parameter needed for the drag force. Hence, the fraction A/B determines the trajectory. While B can be easily calculated ($B = 7.51 \cdot 10^4/s$ for $r = 1\mu\text{m}$), A is difficult to estimate due to the unknown exact tribo-electric charge of a single particle and the electric field estimate for the chip.

For B in the range as calculated here, the exponential terms do not play a role for the trajectory. Therefore, the trajectory of the particle can be approximated by the linear and constant terms, i.e. the movement, more precisely the velocity, of the particle is constant and determined by the size of A compared to B . This means that negatively charged particles that are subject to finite drag forces will always reach the surface of the chip in a finite amount of time. As a rough estimate,

A can be assumed to be $5 \cdot 10^{-9} \text{m/s}^2$, resulting from $q/m = -5 \cdot 10^{-3} \text{C/kg}$ and $E = 100 \text{V}/100 \mu\text{m}$, which yields a velocity of $\dot{z} \approx 66 \mu\text{m/s}$ in direction to the chip.

The motion in x- and y-direction can be decoupled from the one in z-direction, because the chip is assumed to be a homogeneously charged conductor plate. This guarantees that the particle moves in an equipotential plane in x- and y-direction parallel to the chip surface for each constant z. The only force acting on the particle is the drag force, so that the simple differential equations

$$\ddot{x}(t) = -\frac{6\pi\mu \cdot r}{m} \cdot \dot{x}(t) := -B \cdot \dot{x}(t) \quad (3.25)$$

$$\ddot{y}(t) = -\frac{6\pi\mu \cdot r}{m} \cdot \dot{y}(t) := -B \cdot \dot{y}(t) \quad (3.26)$$

have to be solved, which yield

$$\dot{x}(t) = v_x(t=0) \cdot \exp(-Bt) \quad (3.27)$$

$$\dot{y}(t) = v_y(t=0) \cdot \exp(-Bt) \quad (3.28)$$

for the velocities in these directions, assuming that the initial velocities are given by v_x and v_y , respectively. Integrating over t , in turn, results in the trajectories

$$x(t) = \frac{v_x(t=0)}{B} \cdot (1 - \exp(-Bt)) \quad (3.29)$$

$$y(t) = \frac{v_y(t=0)}{B} \cdot (1 - \exp(-Bt)) \quad (3.30)$$

Using the parameter B as already calculated above, the transport in x- and y-direction is very small, taking into account that the initial velocities are small compared to B , so that the ratios v_x/B and v_y/B are small.

As these calculations show, a selective particle deposition on the chip surface in stationary air is hardly possible, since this task would require very high electric fields, long deposition times or very short distances between the particles and the chip. One way to improve this method could be to use a medium with a lower density than air to transfer the particles in. Using for example hydrogen, which has the lowest dynamic viscosity of all gases, the dynamic viscosity can be halved. Hence, keeping all other input values the same, a factor of 2 is gained at the cost of having to construct deposition chambers that can be evacuated or risking a hydrogen-oxygen reaction. Another option to reduce the drag force is using particles with a smaller radius r at the same mass as before, i.e. increasing the density of the particle material.

3.4.2 Inhomogeneous field in moving air

Motion of the airstream

Usually, the gaseous phase is described by as an incompressible, isothermal, continuous stream in the Eulerian frame of reference, while each particle of the solid phase is tracked individually by its equations of motion in the Lagrangian frame of reference that moves with the particles. Further assumptions to define a computable model are the following: the estimates of the characteristic numbers derived in Section 3.3 show that the models of particle transport by airstream must lie in the continuous mechanics regime with laminar air flow when the region in the vicinity of the synthesis sites is to be considered. The two phases in their different frames of reference are decoupled, so that the particles do not influence the airstream. (Weak) Coupling can then be introduced over an exchange of energy or momentum.

The decoupling is motivated in Section 3.3 that states that the particles form a dilute phase in the airstream with negligible particle-particle interactions, as Equation 3.19 shows that the Coulomb force for two point charges of distances farther apart than $10\mu\text{m}$ becomes smaller than the gravitational force. The negative charge and resulting repellent force of the particles contribute to the assumption of negligible particle-particle collisions⁵. Further assumptions are that the particle motion can be approximated by the trajectory of their centre of mass and that the gravitational force can be neglected in a first order approximation, being at least one order of magnitude smaller than the drag forces.

The equations of motion for the airstream of constant density are then given by the incompressible, isothermal Navier-Stokes-equations, stating conservation of momentum (first equation) and conservation of mass (second equation)

$$\rho \frac{\partial v}{\partial t} - \mu \cdot \nabla \cdot (\nabla v + (\nabla v)^\top) + \rho \cdot (v \nabla) v = -\nabla p + f_{ext} \quad (3.31)$$

$$\text{div}(v) = 0 \quad (3.32)$$

The first term in the first equation describes the acceleration over time, the second is the viscous net force and evolves as a result of the resistance to the rate of deformation of fluid elements. It can also be written as

$$\mu \cdot \nabla \cdot (\nabla v + (\nabla v)^\top) = \mu \nabla \tau = \begin{pmatrix} \frac{\partial \tau_{xx}}{\partial x} + \frac{\partial \tau_{yx}}{\partial y} + \frac{\partial \tau_{zx}}{\partial z} \\ \frac{\partial \tau_{xy}}{\partial x} + \frac{\partial \tau_{yy}}{\partial y} + \frac{\partial \tau_{zy}}{\partial z} \\ \frac{\partial \tau_{xz}}{\partial x} + \frac{\partial \tau_{yz}}{\partial y} + \frac{\partial \tau_{zz}}{\partial z} \end{pmatrix} \tau_{ij} = \left(\frac{\partial v_i}{\partial x_j} \right) \quad (3.33)$$

where τ is the tensor of shear strain with the entries τ_{ij} for $i, j = 1, 2, 3$. The third term is the convective acceleration, a time independent effect on macroscopic scales

⁵For the case in which the particle-particle interactions via the Coulomb force cannot be neglected, the assumption of a dilute gas is not valid anymore (s. [Nie09]). As a consequence, the back-reaction of the particles on the airstream also has to be included in the model.

caused by a change of velocity in position and given by

$$(v\nabla) v = \left(v_x \frac{\partial}{\partial x} + v_y \frac{\partial}{\partial y} + v_z \frac{\partial}{\partial z} \right) \cdot v \quad (3.34)$$

On the right hand side, the pressure gradient ∇p and external forces per volume are listed. In the special case of the particle deposition, the external forces are zero, as the airstream is not charged and hence not subject to the electric force of the chip. For fully developed flows, the first term can be omitted as the velocity change over time is negligible. This is the case when the time to reach the chip surface is much smaller than the time of the entire deposition. Since the latter is in the range of seconds, the time independent version is supported, as the velocity of the pressurised air containing the particles injected into the chamber is in the range of 1m/s, implying transport times in the range of 40 ms for the geometry shown in Figure 3.1 on the right.

Furthermore, the geometry of the deposition chamber in the vicinity of the chip has to be taken into account, providing the boundary conditions for the Navier-Stokes-equation. Assuming that the adhesion force to the chamber walls and the chip is stronger than any other force at first contact to the boundary, the velocity of the airstream is set to zero there (which is called the *no slip boundary condition*). Taking the estimates of Section 3.3 into account, the boundary and initial conditions shown in Figure 3.4 are reasonable to assume. For the Navier-Stokes-equation, however, only the region in front of the chip surface close to the outlet is relevant, as the airstream in the inlet cylinder and the conic region might be turbulent.

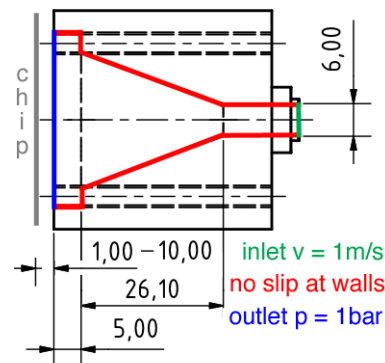


Figure 3.4: Boundary and initial conditions for the cylindric deposition chamber, the velocity profile at the inlet is assumed parabolic with average velocity $v = 1\text{m/s}$

Solving the Navier-Stokes-equation given the boundary conditions is impossible in closed analytical form due to the non-linearity on the left side of Equation 3.33. A very detailed analysis of the problem formulation described above as well as conditions under which stable solutions exist and algorithmic implementations of an

approximated solution can be found in [Ran99]. The method of choice for solving Equation 3.33 according to [Ran99] is based on the discretisation of the geometry by finite elements and an iterative solution of the linearised equation on the single elements.

Finding an approximation to the solution of Equation 3.33, the prerequisites calculated in Section 3.3 allow to select a stationary solver that linearises Equation 3.33 according to the prescriptions detailed in Chapter 6 of [COM]. Then, the linearised equation is solved numerically by a damped Newton method (see [BV04] for implementational details) on the discretised geometry of the finite elements. The latter are created as the standard triangular Lagrange finite elements of secondary order for the velocity and of first order for the pressure by respecting that the mesh size h of the elements must be such that the Péclet number Pe

$$Pe = \frac{\|\beta\| \cdot h}{2c} \quad (3.35)$$

is smaller than one, which provides the necessary regularity of the solution and ensures numerical stability. β is given as the coefficient of the diffusion term and c as the coefficient of the convection in the linearised Navier-Stokes equation. Using 50,868 finite elements with average element area ratio in the range of 10^{-6} , the geometry in the vicinity of the chip in 1cm distance is discretised as shown on the left side of Figure 3.5⁶. UMFPACK⁷ with a relative error below 10^{-6} , then produces the numerical solution to Equation 3.33 in form of the velocity distribution shown on the right side of Figure 3.5.

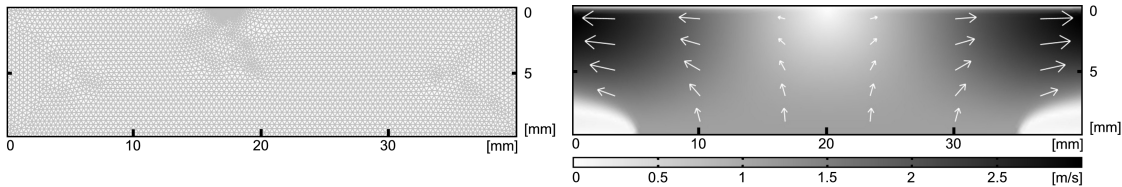


Figure 3.5: Left: Finite element discretisation of the x-z-plane of the geometry Right: Numerical solution in the x-z-plane obtained by COMSOL for the velocity function of Equation 3.33

⁶The finer triangulation in the central part close to the chip is necessary when zooming to a detail of a few chip cells.

⁷s. <http://www.cise.ufl.edu/research/sparse/umfpack/> for implementational details

Electric field configuration

Due to the linearity of the electro-magnetic theory, the electric potential in the proximity of the chip consists of the superposition of the electric potentials of all chip cells at their locations (s_j, t_j) in the x-y-plane for a constant height z_0 . Without loss of generality, z_0 can be assumed to be zero. The electric potential of one chip cell that is assumed to be homogeneously charged can be calculated by formulating Poisson's equation and the constraints at the borders of the cell. Let $\xi \in C_0^\infty(\Omega \subset \mathbb{R}^3, \mathbb{R})$ be a test function and $\phi_j \in C^2(\Omega \subset \mathbb{R}^3, \mathbb{R})$ the electric potential of one chip cell at position $(s_j, t_j, 0)$. Poisson's equation, which must be valid for all ξ , reads

$$\int_{\Omega} \Delta \phi_j(x, y, z) \xi(x, y, z) d\Omega = -4\pi\epsilon_0 \int_{\Omega} \sigma(x, y) \delta(z) \xi(x, y, z) d\Omega \quad (3.36)$$

The boundary conditions are

$$\begin{aligned} \phi_j(x < s_j, y, z = 0) &= 0 & \phi_j(x, y < t_j, z = 0) &= 0 \\ \phi_j(x > s_j + a, y, z = 0) &= 0 & \phi_j(x, y > t_j + b, z = 0) &= 0 \end{aligned}$$

which yield the integral equation

$$\int_{\Omega} \Delta \phi_j(x, y, z) \xi d\Omega = -4\pi\epsilon_0 \sigma \int_{\Omega} \delta(z) \theta(x - s_j) \theta(s_j + a - x) \theta(y - t_j) \theta(t_j + b - y) \xi d\Omega$$

where σ is the constant surface charge density, a the length and b the width of the chip cell. Figure 3.6 shows the configuration of the charge density.

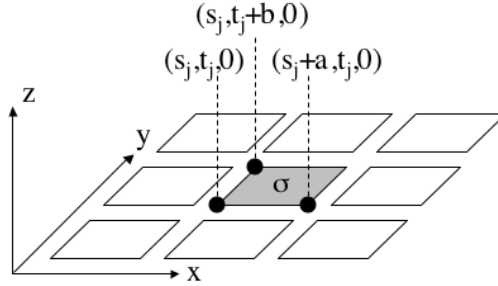


Figure 3.6: Configuration of charge density for the potential of one chip cell

This problem formulation is a typical representative of the class of Dirichlet problems. Yet, in order to prove the existence and uniqueness of any possible solution, $\rho(x, y, z)$ must be Lipschitz continuous. Assuming $\rho(x, y, z)$ to be continuous and bounded, only the existence of a solution can be granted. The respective theorems and proofs can be found in [Alt02]. As $\rho(x, y, z)$ calculated above is not even

continuous, the problem cannot be solved in this way. The solution must be determined by first introducing a (Lipschitz) continuous approximation to $\rho(x, y, z)$. Then, the modified Dirichlet problem can be solved by Fourier transforming both sides of the equation. This simplifies the calculations as the problem formulation for the integrands of the Fourier transformed problem reads

$$\Delta \hat{\phi}_j(k, l, m) = -(k^2 + l^2 + m^2) \hat{\phi}_j(k, l, m) = -4\pi\epsilon_0 \hat{\rho}(k, l, m) \quad (3.37)$$

where k, l, m are the dual variables to x, y, z . The Laplace operator on the left is inverted and applied to the right hand side, so that Fourier transforming the potential $\hat{\phi}_j(k, l, m)$ and the right hand side then leads to the desired result, the potential $\phi_j(x, y, z)$.

Using COMSOL to calculate an approximation to the electric potential of the chip, the result shown in Figure 3.7 can be obtained by formulating Poisson's equation as a linear system of equations. Using about 40 000 quadratic Lagrange elements with element area ratio in the range of 10^{-5} to triangulate the region and solving the linear optimisation problem up to a tolerance of 10^{-6} with UMFPACK, a high degree of accuracy for the approximation is reached. In Figure 3.7 two chip cells with surface charge density σ are separated by seven chip cells that are not charged. The picture is drawn for a profile through the x-z-plane. Apart from setting the potential at the chip surface to 100V for the charged chip cells, the grid between the non charged chip cells is set to 30V, as experiments revealed that charging the grid between the chip cells prevents contamination on the non charged chip cells (s. [Löf09]). From the figure it can be furthermore read off that the electric potential becomes homogeneous in a distance of 300 to 400 μm from the chip, supporting the estimate about the homogeneity of the electric far field obtained in Section 3.4.

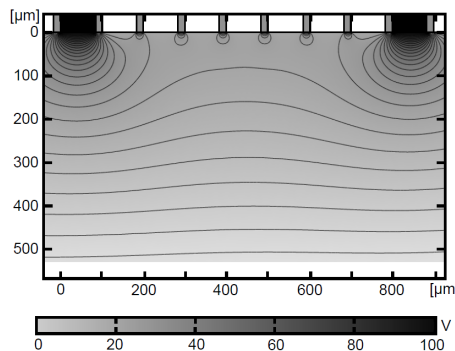


Figure 3.7: x-z-profile of the electric potential of two chip cells separated by seven chip cells, shown are equipotential lines for every 5V

Particle transport

After convergence of the airstream calculation and modelling the electric potential, the traces of the particles can be determined by inserting their parameters (diameter, mass, charge) and initial conditions (start points, initial velocity) in the particle tracking module of COMSOL that simply solves Newton's equation of motion

$$m_p \frac{d^2 x}{dt^2} = F(t, x, \dot{x}) \quad (3.38)$$

where F are the external forces acting on the particle with mass m_p . At first, the electrical force in form of Equation 3.18, i.e. determining the derivative of the electric potential of Section 3.4.2, is added to the right hand side. Second, the Khan-Richardson force as a generalisation of the drag forces discussed in Section 3.3 is also included. In [KR87], the Khan-Richardson force is experimentally investigated and defined as

$$F_K = \pi \cdot r_p^2 \cdot \rho (v - v_p)^2 \cdot (1.84 \cdot (Re_2)^{-0.31} + 0.293 \cdot (Re_2)^{0.06})^{3.45} \quad (3.39)$$

with the particle radius r_p and velocity v_p , the fluid density ρ , dynamic viscosity μ and velocity v and the Reynolds number Re_2 of the two-phase flow given as

$$Re_2 = \frac{2 \cdot r_p \cdot \rho \cdot \|v - v_p\|}{\mu} \quad (3.40)$$

The advantage of this force in comparison to the drag forces of Section 3.3 is that it describes the effect of the airstream motion on the spherical particles over a large range of velocities and Reynolds numbers.

Equation 3.38 can be rewritten as a system of ordinary differential equations that can be solved in COMSOL by using a pair of Runge-Kutta methods of orders four and five, as described in Chapter 7 of [COM].

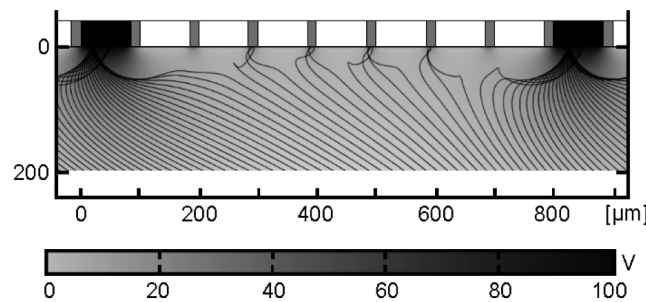


Figure 3.8: x-z-profile of the velocity distribution (s. bar at the bottom) and the particle trajectories (black lines) under the influence of the Khan-Richardson force and the electric field of a checkerboard pattern simulated in COMSOL

Simultaneously visualising all results in the vicinity of the chip surface, Figure 3.8 shows the velocity of the airstream by the grey value profile and the particle trajectories under the influence of the Khan-Richardson force and the electric field of a checkerboard pattern configuration as black lines.

By means of these models, particle deposition for all kinds of possible configurations can be theoretically tested to find those parameters of the particles (size, charge) and of the chip (size of the chip cells, extensions of the grid, applied voltage) that lead to minimal contamination. Further work, as well as comparisons to experiments, based on this aerosol deposition model can be found in [LWK⁺10] and [WKF⁺10].

Chapter 4

Coupling detection

4.1 Motivation

From the bio-chemical point of view, a quality measure in each deposition step is hard to realise, facing that the detection reaction must be non-destructive, reversible and significant. Furthermore, testing each layer of coupled amino acids for their coupling efficiency is very time-consuming. Therefore, the goal of this chapter is to find a bio-chemical quality measure that is able to spatially resolve the coupling of the amino acids and that is then correlated to the images of the particle depositions, such that it is possible to estimate the coupling efficiency by looking at the deposition quality only. This approach also allows to use a destructive detection reaction for determining the coupling efficiency and furthermore avoids undesired side reactions during the actual synthesis, since the quality assessment only uses (non-destructive) image processing for the prediction.

The difficulty of finding such a detection reaction lies in the fact that the density of coupling sites on the support and hence, the maximum density of peptides, is a priori unknown. Most dominantly, it depends on the surface modification, for instance, polymer surfaces with their arborescent structure, as described in [Kön10], are supposed to have more coupling sites than monolayers with an approximate linear structure. Another factor is steric hindrance which inhibits the coupling of amino acids. Taking all these effects into account, there is no theoretical calculation that reliably estimates the range of surface densities of coupling sites and coupled amino acids, in order to narrow the number of possible detection methods. Therefore, the standard chemical detection reactions, Bromophenol blue and the Kaiser test with Ninhydrin, are investigated first with respect to their suitability for spatially resolved amino group detection. Since these methods are known to yield only qualitative results, surface analysis techniques are tested as well.

4.2 Bio-chemical detection reactions

4.2.1 Bromophenol blue

Bromophenol blue (0.01 percent in methanol) is a colour marker that binds to free amino groups. The areal density of amino groups must not fall below 1nmol per cm^2 in order to be made visible. The absorption maximum lies between 590 and 595nm at a pH-value of 4.6. The range of pH-values initiating the reaction is 3.0 to 4.6. The colouration of the N-terminated amino groups after the elimination of the Fmoc groups is reversible due to the ionic, non-covalent bond: three cleaning cycles of five minutes in Dimethylformamide (DMF) suffice to remove the colouration. As described in [Blo09], three additional cleaning cycles of five minutes with methanol (MeOH) and two five minute cycles with ethanol (EtOH) are performed to prepare the peptide chips for the next layer of amino acids. Applying this detection reaction to the amino acids coupled on a CMOS chip, in most cases, no reaction is visible. Hence, the detection reaction is not a reliable tool for quality analysis, let alone that the spatial resolution of the method is by far not satisfactory to determine the coupling efficiency for single chip cells. Moreover, the visibility of this detection reaction decisively depends on the surface coating used for solid phase peptide synthesis. For instance, a heteropolymere consisting of 20% PEGMA and 80% PMMA is compatible with Bromophenol blue while the current standard for peptide synthesis on glass slides consisting of 10% poly(ethylene glycol)-methacrylate (PEGMA) and 90% methyl-methacrylate (MMA) is not [Sch10] (see [Kön10] for details about the surfaces).

4.2.2 Ninhydrin

Ninhydrin is another possible colour marker that couples covalently to free amino groups. For the detection reaction, also called the Kaiser test [KCBC70], to yield a violet colour, also called *Ruhemann's purple*, a Ninhydrin molecule reacts with an amino acid to an amino ketone, which then reacts with another Ninhydrin molecule to produce the target molecule as shown in Figure 4.1.

The best detection limit lies at 50pmol per cm^2 by $3 \cdot 10^{13}$ particles on 1cm^2 , i.e. one order of magnitude lower than the one for Bromophenol blue. Due to the covalent bond, the amino acid is reduced, i.e. the process is non-reversible. The problem with this reaction, however, lies in the fact that *Ruhemann's purple*, indicating the coupling efficiency, is not fixed at the target molecule. Hence, there is no possibility to spatially resolve the locations of coupled amino acids. Hence, this technique is not appropriate for a spatially resolved quality measure, either.

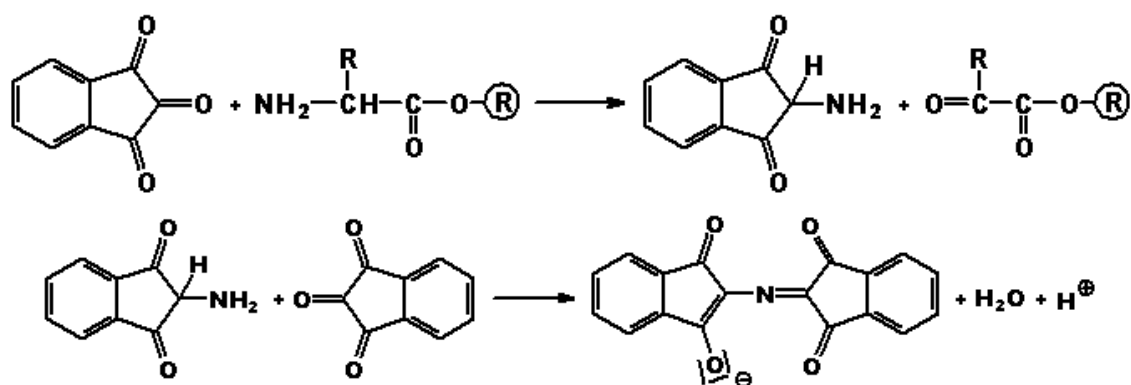


Figure 4.1: Top: First part of Ninhydrin reaction from amino acid to amino ketone Bottom: Second part of Ninhydrin reaction from amino ketone to Ruhemann's purple

4.2.3 Fmoc determination

Another possibility to measure the coupling efficiency uses the protection groups of the amino group on the newly coupled amino acid, as introduced in Section 1.3. The NH_2 end of the amino acid is protected by an Fmoc molecule so that the amino acid cannot react with other free amino acids in the toner matrix when melting. Having the amino acids coupled on the chip and the remnants of the toner discarded, the deprotection step splits the Fmoc groups off the amino acid, providing the next coupling sites for the subsequent amino acid. Hence, measuring the number of Fmoc groups split off in the deprotection step by means of an UV-spectrophotometer as described in Chapter 3 of [CW00] is an estimate for the number of coupled amino acids. Inherent to this detection principle, the method is not spatially resolved and thus not suitable for finding a quality measure as to be defined in Section 5.5. However, the densities of coupling sites obtained by this method should be in accordance to the ones obtained by other means.

4.3 Surface analysis techniques

As Section 4.2 reveals, there are no simple chemical detection reactions to measure the coupling efficiency. Therefore, physical surface analysis techniques are tested for their suitability. Table 4.1 gives a synopsis of possible methods, their detection range and their analytical spot size. Another important parameter is the profile depth which is also listed in Table 4.1.

Requiring an analytical spot size in the range of 1 to $10\mu\text{m}$, time of flight secondary ion mass spectrometry (TOF-SIMS) seems to be the best option. First of all it can

method	detection range	lateral resolution	profile depth
AES	10^{20} atoms/cm ³	1nm – 10nm	< 10nm
AFM	1 atom/cm ³	100pm – 50pm	uppermost atom layer
TOF-SIMS	10^{15} atoms/cm ³	500nm – 50 μ m	< 3nm
XPS	10^{19} atoms/cm ³	10 μ m – 1mm	< 10nm

Table 4.1: Surface analysis techniques that may be able to resolve the density of amino acids on the chip surface: Auger electron spectroscopy (AES), atomic force microscope (AFM), time-of-flight secondary ion mass spectrometry (TOF-SIMS), X-ray photoelectron spectroscopy (XPS) (sources: [EAG09], [VfOM09])

detect down to 10^{15} atoms per cm³, which should be sufficient to detect amino acids at the coupling sites on the surface because the surface density of coupling sites was determined to be in the range of 10^{16} per cm² ([SBK⁺09]). Secondly, its profile depth is small enough to analyse only the uppermost layer of the surface. Intruding deeper into the surface, artefacts coming from the polymer/ monolayer surface activation (see Appendix A.1) or even from the chip might obscure the signal actually coming from the amino acids at the top of the growing peptide. This excludes techniques as, for instance, X-ray photoelectron spectroscopy (XPS) or Auger electron spectroscopy (AES).

Scanning probe microscopy methods (SPM), as atomic force microscopy (AFM), could also be considered due to their shallow profile depth and their capability of analysing spot sizes far less than 10 μ m. Yet, inherent to its functional principle, AFM cannot yield a good spatial resolution for the observation of single amino acids coupled to the surface, as the growing peptide is free to move vertically, leading to distorted images.

Not shown in Table 4.1, spatially resolved matrix-assisted laser deposition/ ionisation (MALDI) could also be considered. Yet, this method is excluded again, as the minimum spot size analysable is 20 μ m.

4.3.1 Time Of Flight - Secondary Ion Mass Spectrometry

The TOF-SIMS method is thus selected as method of choice, further encouraged by literature, demonstrating that single molecule detection, especially amino acid detection is possible with this method, for example in [ARB08] or [Hag00]. Furthermore, it is possible to chemically characterise self-assembled monolayers (SAMs) by means of TOF-SIMS (see [GR94]), which can become important for peptide synthesis on CMOS chips as well, since SAMs are currently being established as chemically active surface coating. Knowing the TOF-SIMS fingerprint of the surface, it is easier to filter out the signals from the amino acids thereon.

The imaging of TOF-SIMS measurements works as depicted in Figure 4.2. The surface of the sample to be analysed is exposed to a pulsed beam of primary ions (usually gallium, bismuth or noble gas ions). The secondary ions that are dissolved from the surface of the sample by the primary ions are accelerated according to their charge and enter a time-of-flight mass spectrometer in which they are separated according to their mass. This procedure is performed as a scan over the entire surface, so that a spatially resolved image of the detected mass distribution is created.

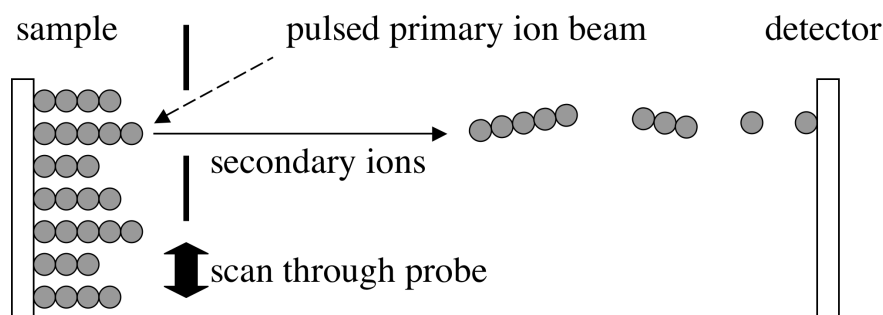


Figure 4.2: TOF-SIMS: The sample is exposed a pulsed beam of primary ions to dissolve secondary ions from the sample surface. The latter enter a time-of-flight mass spectrometer in which they are separated by their mass. Performing the procedure as a scan over the entire surface, a spatially resolved mass distribution is created

The difficulty of recognising the amino acids among all other signals lies in the fact, that, depending on the energy of the primary beam, the fragments ionised out of the sample surface can consist of any sequence of molecules that occurs in the sample. Considering for instance proline with the chemical formula $C_5H_9NO_2$, ionised fragments like C_4H_6N , C_4H_5O or C_4H_8N can occur. The latter is also a configuration that can occur for arginine $C_6H_{14}N_4O_2$, so that the mass spectra of their ions overlap. Hence, the goal is to find fragments that are specific for each sort of amino acid that cannot occur due to any other molecule on the sample surface.

Planning a TOF-SIMS experiment to be performed at the Verbundzentrum für Oberflächenanalyse Münster in Münster, Germany, an analysis chip with a special test pattern is created. It is prepared with a polymer surface coating optimised for on-chip synthesis, as developed in detail in [Bey05] and tested in [Kön10]. After chemical activation of the surface, deposition with three toners containing different amino acids, namely proline (Pro), tryptophan (Trp) and arginine (Arg) in a three-fold redundant pattern as shown in Figure 4.3 is performed. A composition like this allows proof-of-principle detection of any of the three amino acids, as well as a comparison and a test for homogeneity of deposition and coupling across the entire chip.

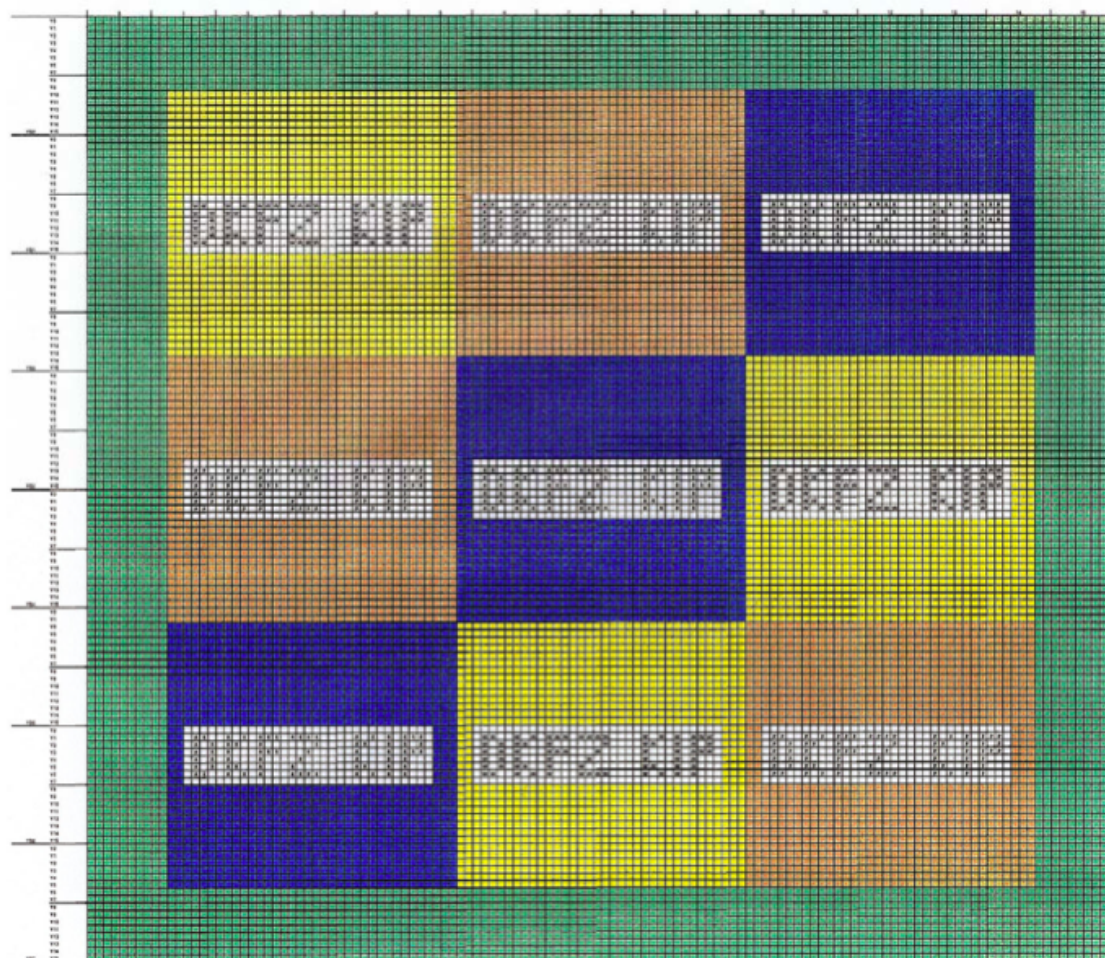


Figure 4.3: Deposition pattern for the chip analysed with TOF-SIMS at the VOM in Münster, Germany, letters in yellow areas covered with Trp, in red areas with Arg, in blue areas with Pro

Melting and coupling subsequently happens placing the chip in the oven at 90°C for 90 minutes. The final step is the chemical washing and deprotection of the amino groups, making the chip ready for the next deposition. Instead of depositing the next layer, the surface is analysed by TOF-SIMS using Bi_3^+ ions yielding a total rate of detected fragments of 3800 counts per second and chip cell as shown in Figure 4.4 top left. Since the method is expensive and laborious, the proof-of-principle experiments shown here are limited to an area of 5×5 chip cells containing spots deposited with Trp at a resolution of 128×128 pixels, i.e. $3.91\mu\text{m}$ per pixel.

The total ion image in Figure 4.4 on the top left, by which all other images are normalised, shows that ions are homogeneously sputtered off the area under investi-

gation. The image of the Si signal right to it reveals that a high percentage of Si is found on the grid and also on the chip cells, where toner was deposited. The former might come from the silanisation of the surface (see [SBK⁺09] for details) or from the chip surface itself, the latter indicates that the silica particles contained in the toner are not completely dissolved from the surface in the chemical washing. The assumption that the weak signals of Si on the grid originate from the chip surface is supported by the image of the Al signal on the right of the top row, showing a rather uniform distribution with some noise. The K signal on the left of the second row also confirms this idea, as potassium is often used in the etching processes of chip production and not applied in the chemical processing of on-chip peptide synthesis. Hence, assuming a profile depth of a few molecular lengths, the surface coating must be broken at the grid, so that chemical activation is limited to the chip cells. This might be caused by the passivation openings with a step in the range of 0.5 to 1 μm going from the active chip cell area to the grid electrode (see [Kön10] for details), which is one order of magnitude larger than the maximum thickness of the surface coating of about 50nm, measured in [Bey05].

Another consistent explanation for the signals in the Si, K and Al images is the possibility that Si originates from the silanisation of the surface and the other two elements are transported from the chip surface into the surface coating. This can happen, as the chemical surface activation is performed in aqueous solution. In order to produce peptide densities that yield fluorescence signals as in [BNB⁺07], the surface coating must be continuous on the chip cell areas, which supports this theory as the coating cannot contain so many holes to produce the Al signal.

The next two images in the centre row of Figure 4.4 containing the iron and silicone signals reveal that the insufficiency of the chemical washing, since both are components of the toner and, thus, should not be present after washing. The reason for their presence is explained by the fact that the chip is washed in horizontal position, so that the remnants of the toner are spread over the surface before the entire solution is discarded. Washing a glass slide in the same manner, a subsequent XPS analysis also yields a large Fe signal, clearly implying that the Fe signal is caused by the toner remnants and not by the chip surface. Analysing a vertically washed glass slide, the Fe signals, and hence the toner remnants, are decreased as found out by [Sch10].

The bottom row of Figure 4.4 shows the signals of the Trp fragments on the left and in the centre of the row. As observed, the $\text{C}_9\text{H}_8\text{N}$ is clearly visible and matches the deposition as shown in Figure 4.3, while the signal of the second fragment, $\text{C}_{11}\text{H}_8\text{NO}$, is weaker. Due to this, the first fragment is used to set up the correlation to the images of the particle depositions as introduced in Section 5.5.2. The last image on the right of this row can be considered as a negative cross-check, as calcium is not supposed to appear in the analysis, which is confirmed by the low signals.

Detecting signals of the amino acid fragments, it is clear that the density of coupling sites per $3.91\mu\text{m}$ must lie above the detection limit shown in Table 4.1. Hence, approximating the density of coupling sites per cm^2 with the number of atoms per cm^3 , which is sensible as the profile depth of the method is only a few atoms thick, there must be about $5 \cdot 10^{15}$ coupling sites per cm^2 , which is in accordance with the measurements of [SBK⁺09] and [Bey05] which determined the number of coupling sites in the range of 10^{16} per cm^2 .

On the whole, these measurements prove the suitability of the method for amino acid detection and already reveal a lot of useful information about the entire synthesis process, starting from toner remnants over the washing process to the coupled amino acids. Despite the low signal to noise ratio of 2 to 1 in the images of Figure 4.4, the $\text{C}_9\text{H}_8\text{N}$ is successfully used for quality evaluation in Section 5.5.2. Additionally, the method can be applied when optimising the synthesis protocols as well. In order to improve image quality, further calibration analyses are required, for instance one in which the characteristic TOF-SIMS signals of the toner matrix are recorded. This is possible by preparing a second analysis chip with a deposition pattern containing a toner with amino acids and the reference toner without any amino acids and immobilising both sorts of particles on the surface. The second calibration can then be performed by analysing all three areas of the amino acids on the first chip, in order to monitor homogeneity.

Having a calibrated TOF-SIMS apparatus, measurements on a third chip with one layer of all amino acids available can be performed, as well as a measurement of coupling efficiencies between all possible pairs of amino acids. With this information it is possible to describe peptide synthesis as a Markov chain model of first order and obtain an estimate¹ how difficult the assembly of a certain peptide is. Furthermore, comparing the TOF-SIMS results with the ones gained by analysing the fluorescence signals as performed in [Bey05] and [Blo09], the accuracy of the latter can be tested.

¹In fact, modelling the peptide synthesis of a n-mer as a Markov chain model of n-th order comes closest to reality but is too laborious to test experimentally.

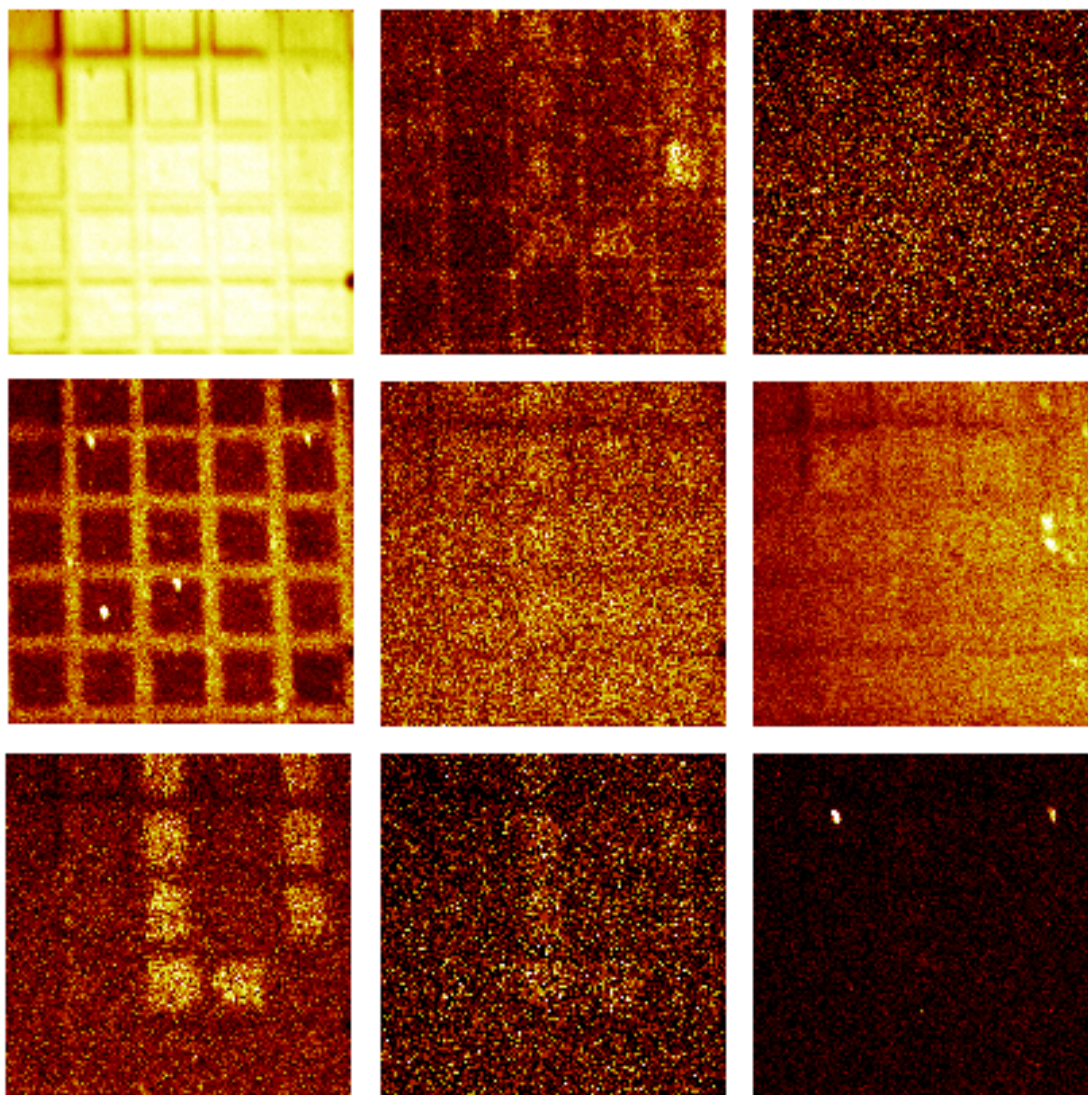


Figure 4.4: Top row: Total ion, Si, Al Centre row: K, Fe, C₅H₁₅Si₂O Bottom row: C₉H₈N, C₁₁H₈NO, Ca, toner deposition looks like the C₉H₈N signal

Chapter 5

Quality controlling program

5.1 Concept and outline

Since the image acquisition setup will be integrated in the peptide synthesis automaton which is still in the state of development and might undergo changes in the future, the program has to be flexible to cope with different microscope and camera resolutions and various image formats. Furthermore, it might also be necessary to exchange parts of the image acquisition setup even during a peptide synthesis. Therefore, the quality controlling program, designed in this chapter, consists of two entities, a training part and a testing part.

The training part is performed prior to the start of a peptide synthesis. It calibrates the setup for image acquisition and determines the parameters required for the subsequent testing part. The latter takes place during peptide synthesis and applies the quality measure to be set up based on the results of Chapter 4 to analyse and classify the deposition quality of the chip after each particle deposition step. Hence, a training step is required only when changing the image acquisition conditions.

Facing hundreds of particle depositions per peptide synthesis, the program should also be run time optimised. As particle depositions are performed within a few seconds, the processing time of the algorithm should be of the same time complexity in order not to stall the synthesis process during quality analysis.

But before automated training and testing phases can be set up, the images taken with the setup described in Chapter 2 have to be investigated concerning required image enhancement. Furthermore, image segmentation and feature selection and extraction with respect to the results obtained in Chapter 4 have to be discussed, in order to find the best classifier for the deposition quality.

5.2 Image enhancement

5.2.1 Noise reduction

Before starting the actual quality check on the images of peptide arrays, the images have to be preprocessed to account for inhomogeneous illumination and image noise. Depending on the device used to take the picture and the time limits to be met, several methods are investigated.

In order to reduce the complexity of the software preprocessing, the hardware should cover the noise reduction as far as possible. As described in Chapter 2.3, the setup can be installed to provide optimal optical conditions for the image acquisition. Furthermore, the chip of the microscope camera can be calibrated to account for the inhomogeneous dark currents of the individual pixels, as well as it can perform white and black balancing for a homogeneous intensity profile over the entire image in all colour channels.

Since the particle clusters that contain the amino acids are always of the same whitish colour contrasting with the darker colour of the chip, it is convenient to take the images in grey values or convert the RGB images into grey values before processing, which enables to apply image processing techniques as edge detection, in the first place.

Using a camera that produces coloured images by means of three CCDs or by a chip with a Bayer pattern, colour interpolation and conversion errors can occur in the subpixel range, but are of minor importance, when compared to the optical resolution limit of the setup.

Figure 5.1 shows the conversion from coloured RGB image to grey scale in the range of $[0, 255]$, or in the normalised version in $[0, 1]$. This step reduces the amount of pixels to be processed down to one third of the original image.

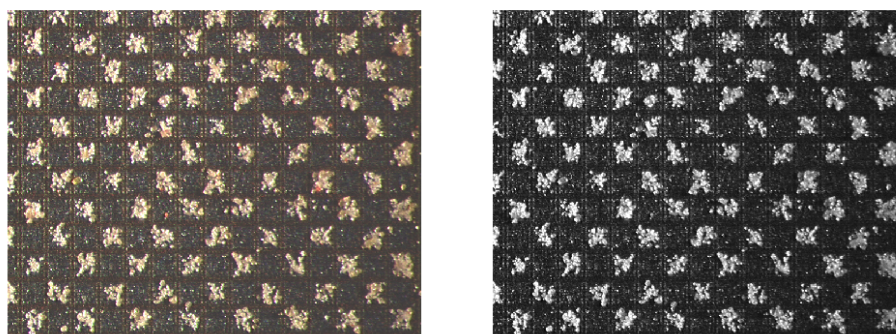


Figure 5.1: Left: RGB image Right: Grey value image of a particle deposition

5.2.2 Flat-field correction

In order to account for dust, inhomogeneous pixel sensitivities or other non-uniformities caused by the optical setup, CCD chips are calibrated by means of a white image taken with the same setup. First, the images to be analysed are pixel-wise divided by this white image and afterwards multiplied with its mean grey value. This procedure is usually performed together with the calibration of the dark current of the individual pixels in hardware.

If the camera does not provide this feature or in case of inadequate illuminations, the flat-field correction can also be calculated retroactively in software to each colour channel of the CCD pictures before converting to grey scale. This can be advantageous for images of past peptide syntheses that do not meet the current standards concerning homogeneity of illumination.

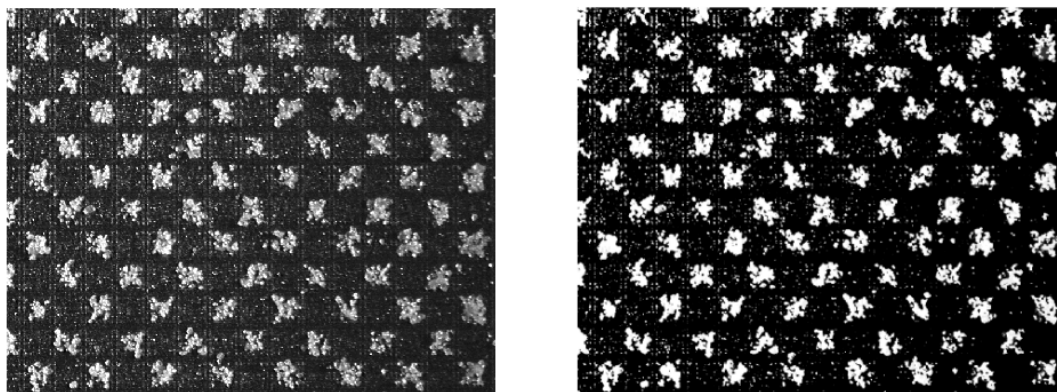


Figure 5.2: Left: Grey level image without flat-field correction Right: Grey level image with flat-field correction

Comparing the left image in Figure 5.2 with the right one, it can be clearly observed that the illumination of the calibrated image is more homogeneous, especially in the right part of the image which is underexposed in the left image. Additionally, the amount of noisy pixels is reduced in the right image. The disadvantage of this method lies in the fact that the white image has to be taken with the same setup and under the same conditions as the image to be analysed. If the lighting conditions are not exactly reproducible, a lot of white images have to be stored in addition to the images to be investigated.

Concerning the computational complexity of the flat-field correction, $\mathcal{O}(MN)$ arithmetic operations are necessary for an image with $M \times N$ pixels per channel: after the three divisions of each RGB channel by the white image, a mean value calculation and three multiplications follow.

5.2.3 Smoothing with a Gaussian kernel

The best way to discard of statistical noise is to convolve the image with a smoothing kernel.

If the image is represented by a discrete function f asserting each of the $M \times N$ pixel coordinates $p_i = (x_i, y_i)$, $i \in \{1, \dots, M \times N\}$, a grey value and the discrete smoothing kernel is called k_d , the discrete convolution is given by

$$(f * k_d)(p_i) = \sum_{p_j \neq p_i} f(p_j) \cdot k_d(p_i, p_j) \quad (5.1)$$

Here, the discrete version of an isotropic Gaussian kernel

$$k(p_i, p_j) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\|p_j - p_i\|_2^2}{2\sigma^2}\right) \quad (5.2)$$

with width σ depending on the Euclidean distance $\|\cdot\|_2$ of the pixel coordinates p_i and p_j is used. Since k is quickly decreasing, its values more than 3σ far away from its mean are practically zero. Therefore, the discretised version k_d of the continuous Gaussian kernel k can be cut of at the distance of 3σ from its mean when implemented in an algorithm. Figure 5.3 shows an example of a discrete Gaussian kernel with a size of 49×49 pixels and a width of 3 pixel on which the statements can easily be verified.

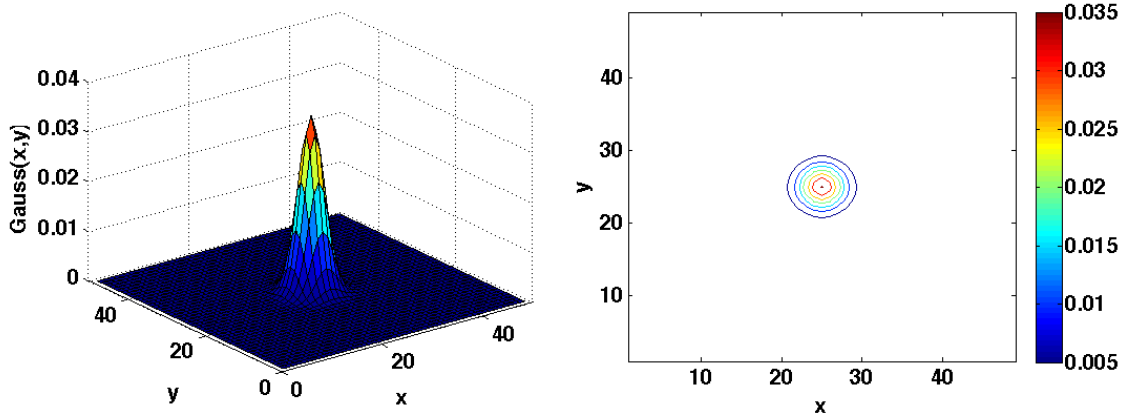


Figure 5.3: Left: 49×49 px k_d with $\sigma = 3$ px in 3D Right: Contour plot of k_d

Figure 5.4 shows three details from an example image convolved with a Gaussian isotropic, discrete kernel with different kernel widths σ . It can be observed that the noise is reduced, but with increasing kernel width smaller particle clusters are

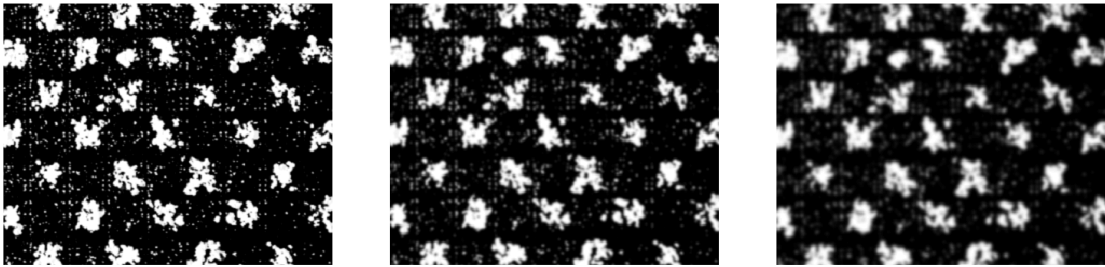


Figure 5.4: Left: 49×49 px k_d with $\sigma = 1$ px Centre: $\sigma = 2$ px Right: $\sigma = 3$ px

smoothed out beyond recognition, as well. Hence, σ should not be chosen larger than the optical resolution limit calculated in Equation 2.7.

The disadvantage of the smoothing method lies in the enlargement of the image by the size of the kernel due to the convolution. Having to process an $(M+S) \times (N+S)$ image instead of an $M \times N$ image slows down the algorithm for quality classification. Yet, as will be shown in Section 5.4, the algorithm analyses only complete chip cells, i.e. the boundaries of the image are not taken into account. Hence, without loss of information relevant for classification, the additional pixels due to the convolution at the boundaries can be discarded.

Due to the associativity and commutativity of the convolution, it is possible to split the time consuming two dimensional convolution of k_d with the image into two separate convolution steps for the two image dimensions. Subsequent application of these two steps saves processing time: If the kernel k_d is of size $S \times S$ pixels, then the two dimensional convolution is a matrix multiplication of the order $\mathcal{O}(S^2)$ while the subsequent one dimensional steps are of the order $\mathcal{O}(S)$. Altogether, this yields a complexity of the order of $\mathcal{O}(MNS)$ for $S < N$ and $S < M$.

5.2.4 Smoothing with a bilateral filter

Instead of smoothing with a simple Gaussian kernel, a non-linear, local bilateral filtering technique as described in [TM98] could be used. The advantage of bilateral filtering is that it does not only smooth the image in the spatial domain but simultaneously correlates the smoothing in the intensity domain to the spatial one. The algorithm bases on the discrete filter transformation

$$b(p_i) = n^{-1} \sum_{p_j \in S \times S(p_i)} f(p_j) r(p_i, p_j) I(f(p_i), f(p_j)) \quad (5.3)$$

with the normalisation constant

$$n = \sum_{p_j \in S \times S(p_i)} r(p_i, p_j) I(f(p_i), f(p_j)) \quad (5.4)$$

$f(p_j)$ is the function that assigns a grey value to each pixel p_j in the $S \times S$ neighbourhood of pixel p_i , $r(p_i, p_j)$ calculates the proximity of p_j to p_i and $I(f(p_i), f(p_j))$ determines the proximity of their grey values. In the simplest case, both functions r and I can be written as Gaussian kernels (s. Equation 5.2) of the Euclidean distance of their input arguments.

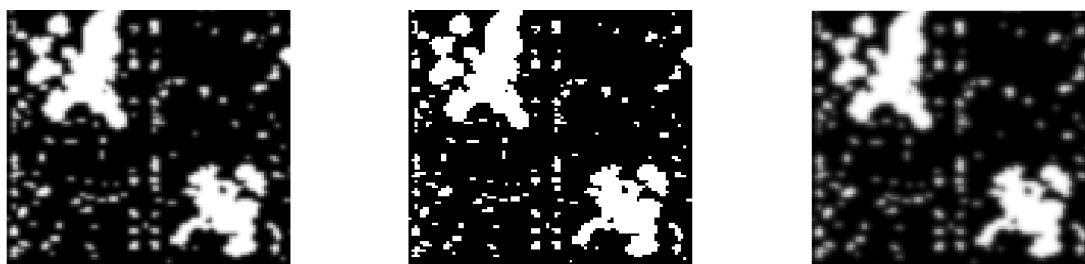


Figure 5.5: Left: Gaussian smoothing: $\sigma = 1\text{px}$ Centre: Bilateral smoothing: $S = 10$, $\sigma_r = 1\text{px}$, $\sigma_I = 1$ grey value Right: Bilateral smoothing: $S = 10$, $\sigma_r = 1\text{px}$, $\sigma_I = 51$ grey values

Having the same complexity as the simple Gaussian smoothing, the execution of the bilateral filter is more time consuming due to the additional multiplication of $I(f(p_i), f(p_j))$ for each pair (p_i, p_j) . Furthermore, as can be seen in Figure 5.5, the filter does not achieve better results than the simple Gaussian smoothing, which is shown on the left. Using the smallest widths of 1px and 1 grey value for the spatial and intensity Gaussians $r(p_i, p_j)$ and $I(f(p_i), f(p_j))$, respectively, the centre image shows that the bilateral filter does not discard of noise, but rather converts the image into a binary one, which could have been expected beforehand, since the bilateral filter is an edge preserving filter that homogenises connected pixel values. Keeping the spatial width 1px as in the simple Gaussian method and increasing the intensity width, as shown on the right hand side of Figure 5.5, the image is blurred in the same way as it is observed for the Gaussian smoothing.

Thus, since the bilateral filter does not lead to a further improvement of the image contents and has a longer execution time, the simple Gaussian smoothing suffices to reduce the noise in the chip images.

5.3 Image segmentation

Having reduced the noise, the image is now divided into fore- and background, i.e. the particles are separated from the background of the chip. From the image processing point of view, this is a conversion into a binary image (each pixel is labelled as either fore- or background). Pattern recognition tools ranging from

simple thresholding over level-sets [OS88] [Set99] to methods using total variation [CCN07] can then be applied to achieve this goal. Yet, not all of these methods are equally suitable. Total variational approaches, for instance, tend to segment convex, contiguous regions by using spatial correlations in the image and, thus, over-estimate the size of the foreground regions in the final segmentation. Furthermore requiring that the image segmentation should be a fast processing step, simple algorithms like global thresholding are preferred. Since this requires a training step to learn the threshold value, two other, unsupervised methods are also tested for suitability: k-means clustering [Ste57] and the efficient clustering earth mover's distance [WO11].

5.3.1 Global thresholding

Assuming that reflections of the chip surface do not exceed the number of pixels that correspond to the optical resolution limit, a simple linear global thresholding can be applied to the image to segment it into particle clusters as foreground and chip surface as background.

The thresholding consists of the projection of all values higher than the threshold to one, and the projection of all pixel values below the threshold to zero. Hence, for an image with M rows and N columns, the processing time is of the order $\mathcal{O}(MN)$, comparing each pixel to the threshold and projecting it either to one or zero. Compared to fancier methods as [OS88] and [CCN07], the global linear thresholding is much faster.

If the reflections are larger than the resolution limit, the imaging artefacts cannot be distinguished from the deposited particles. Yet, this uncertainty can be accounted for in the quality measure when the image contents showing the deposition of particles containing amino acids are correlated with the bio-chemical gain in coupled amino acids to the growing peptide, as described in Chapter 4 and Section 5.5.

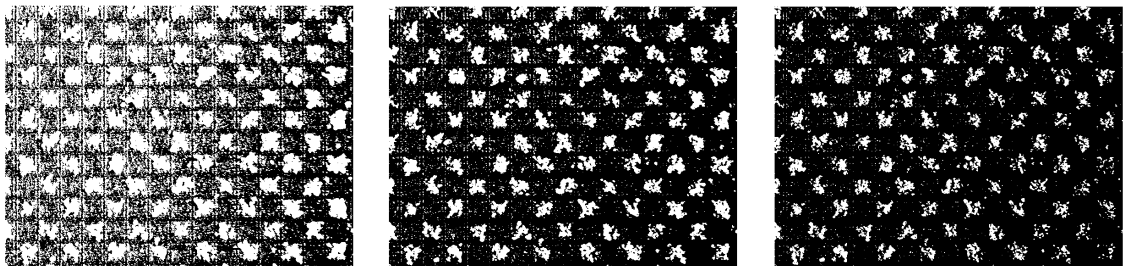


Figure 5.6: Left: Segmentation with threshold 0.31 Centre: Threshold of 0.47 Right: Threshold of 0.63

The disadvantage of this method lies in the subtraction of a *global*, absolute value, which makes it very sensitive to inhomogeneous illumination. Considering Figure 5.1 with normalised grey values in the interval $[0, 1]$ as a negative example image,

Figure 5.6 shows the effect of segmenting the image with a global threshold of various height. In the left image, the upper left part of the chip cells cannot be analysed anymore, while, in rightmost image, this is the case for the lower right part of the image.

On the other hand, ensuring a suitable illumination as described in Chapter 2.3, the global linear thresholding is the fastest method that requires only one parameter, namely the binarisation threshold. In order to find the threshold, the histogram of (normalised) grey values is plotted as shown in Figure 5.7. If the distribution of grey values is bimodal, the threshold can be chosen such as to separate the two peaks in the histogram, e.g. by means of Bayesian inference. Yet, as the grey value distribution of the peptide chip images is not bimodal, this method would not yield a robust binarisation threshold. As will be explained in Section 5.6, the binarisa-

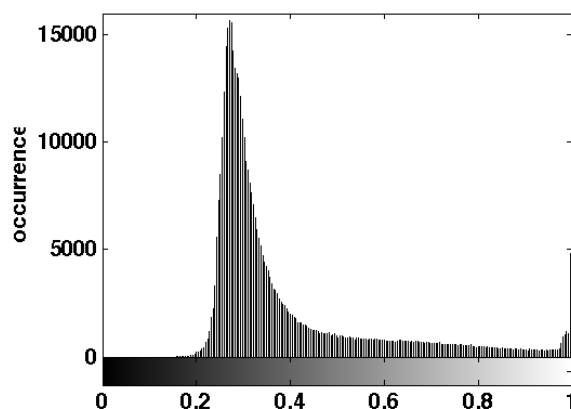


Figure 5.7: Histogram of normalised grey values for the image shown in Figure 5.1

tion threshold can be calculated in a supervised training step using the average of the mean grey values of all chip cells, assuming that the average grey value can be turned into a threshold for one pixel to belong to the foreground.

As an example, a well illuminated chip cell of the image in Figure 5.1 can be considered. Figure 5.8 shows the original chip cell as normalised grey value image (left) and next to it the segmentation in which the mean grey value threshold was applied as described in Section 5.6. Comparing the original image to the segmented, the particle form can be clearly recognised in the segmentation. Contrary to that, the images at the right of Figure 5.8 show segmentation thresholds below and above the trained threshold. While the lower threshold allows for too many background pixels to belong to the foreground, the higher threshold discards too many foreground pixels that a human being would clearly label to be particles. Taking furthermore into account that connected regions smaller than the optical resolution limit are still to be discarded, the trained threshold seems to be an appropriate choice.

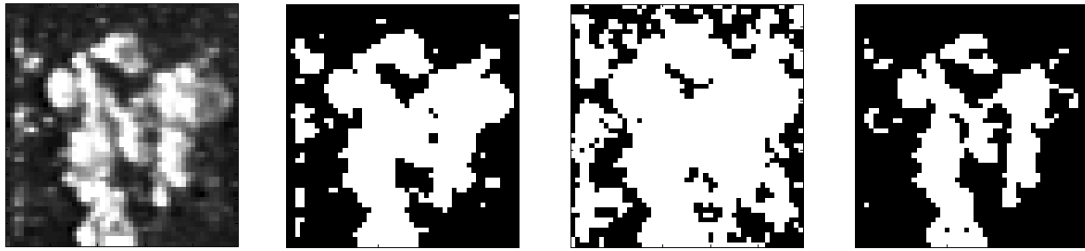


Figure 5.8: Left: Chip cell detail in normalised grey value image Centre left: Segmentation with trained threshold (0.42) Centre right: Segmentation with grey value lower than threshold (0.30) Right: Segmentation with grey value higher than threshold (0.6)

5.3.2 Histogram bin k-means clustering

Instead of separating the intensity values by considering each single pixel, segmentation methods that base on the separation of histogram bins can also be considered. First, a histogram of normalised grey values as in Figure 5.7 is calculated from the image. Dividing the range of 256 grey values into n bins, k-means clustering [Ste57] can be used to cluster all bins such that the intra-class variance of these two cluster classes is minimal

$$\arg \min_C \sum_{i=1}^2 \sum_{b_j \in c_i} \|b_j - \mu_i\|_2^2 \quad (5.5)$$

where $C = \{c_1, c_2\}$ denotes the set of two clusters, b_j with $j = 1, \dots, n$ the n coordinates of the histogram bins and μ_1, μ_2 the mean values of the bin coordinate of the fore- and background clusters. The solution of this non-convex optimisation problem is usually obtained by applying Lloyd's algorithm, which consists of two steps. Given an initial set of mean values $\mu_1^{(0)}, \mu_2^{(0)}$ which is usually a random initialisation, the assignment step at iteration t

$$c_i^{(t)} = \left\{ b_j \mid \|b_j - \mu_i^{(t)}\|_2 \leq \|b_j - \mu_l^{(t)}\|_2 \forall l \neq i \right\} \quad (5.6)$$

asserts each histogram bin to the closest mean. Then, in the updating step

$$\mu_i^{(t+1)} = \frac{1}{|c_i^{(t)}|} \sum_{b_j \in c_i} b_j \quad (5.7)$$

the mean values are calculated anew using the assignments made in the previous assignment step. The two steps are iterated until the cluster assignments do not change anymore. Although the algorithm converges fast, there is the disadvantage of the random initialisation resulting only in local optima which may change for different initialisations.

Compared to the global thresholding, histogram bin k-means clustering is as sensitive to inhomogeneous illumination but being an unsupervised method does not require a training step to determine a threshold value.

To investigate the suitability of the method and the influence of the number of bins on the segmentation, the same example image detail as in Figure 5.8 is considered again. Creating histograms with $n = 256, 64, 16, 4$ bins for the grey values and applying k-means clustering to assign each grey value bin the class label fore- or background leads to the results shown in Figure 5.9. As can be observed in Figure 5.9, the segmentation is not as good as using a global threshold, which is mainly caused by true foreground bins assigned to the background due to a random initialisation. Running the algorithm several times with different initialisations until a sufficient segmentation is obtained is not an option here. This would require an additional training step again, defining what a sufficient segmentation is, apart from the fact, that there is no guarantee that a sufficient segmentation will be found at all using random initialisations.

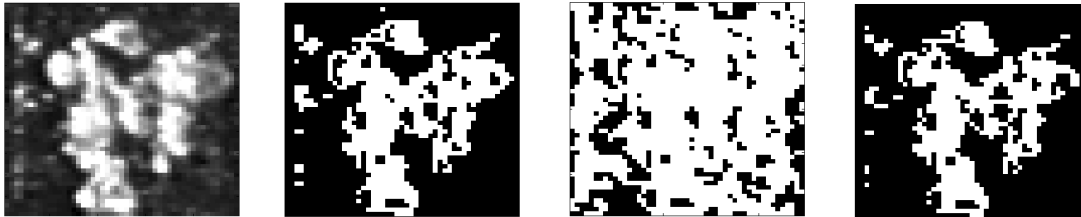


Figure 5.9: Left: Chip cell detail in normalised grey value image Centre left: Segmentation with $n = 16$ bins Centre right: Segmentation with $n = 4$ bins Right: Segmentation with $n = 64$ bins

5.3.3 Histogram bin earth mover's distance clustering

Another histogram based algorithm, the efficient clustering earth mover's distance (ECEMD) uses a deterministic initialisation, which makes it more robust than the k-means ansatz. The implementation bases on the general algorithm of [WO11] but is adapted to the special case of the particle segmentation, which means that the histogram is build from the grey values of the image and that the bin with the highest grey value is the one used to initialise the foreground set of bins.

Let n be the number of histogram bins b_i ($i \in \{1, \dots, n\}$) and denote the class labels $c_i = -1$ or $c_i = +1$ for back- and foreground assignment, respectively. Given a weight w_i for each b_i , assumed to be a pile of earth of height w_i at bin b_i , it is denoted by $w_i^{\mathcal{F}}$ if b_i is in the foreground \mathcal{F} and $w_i^{\mathcal{B}}$ if b_i is in the background. The optimal

class assignments to \mathcal{F} and \mathcal{B} are chosen such that the earth mover's distance (EMD) between these classes is maximal. The EMD itself can be understood as finding the minimum work required to transport all piles from one class to the other, respecting that the *entire* amount of earth has to be moved and that each b_i can only acquire or transport a pile up to its own w_i . This leads to the optimisation problem of Equation 5.8 and Equation 5.9. The first constraint assures that the work, also called flow, f_{ij} between each bin $b_i \in \mathcal{F}$ and $b_j \in \mathcal{B}$ is unidirectional, the second and third that the flow from/ to one bin $b_i \in \mathcal{F}, \mathcal{B}$ does not exceed the weight $w_i^{\mathcal{F}, \mathcal{B}}$ of this bin. The fourth forces all weights of one class to be moved.

$$\max_{c_1, \dots, c_n} \{\text{EMD}(\mathcal{F}, \mathcal{B})\} = \max_{c_1, \dots, c_n} \left\{ \min_f \sum_{i=1}^m \sum_{j=1}^{n-m} d_{ij} \cdot f_{ij} \right\} \quad s.t. \quad (5.8)$$

$$f_{ij} \geq 0 \quad \sum_j f_{ij} \leq w_i^{\mathcal{F}} \quad \sum_i f_{ij} \leq w_j^{\mathcal{B}} \quad \sum_i \sum_j f_{ij} = \min \left\{ \sum_i w_i^{\mathcal{F}}, \sum_j w_j^{\mathcal{B}} \right\} \quad (5.9)$$

The distance d_{ij} between the pairs of b_i, b_j of the opposite classes in the objective function can be calculated as the Euclidean distance

$$d_{ij} = \|b_i - b_j\|_2^2 \quad \forall b_i : c_i = 1, \quad \forall b_j : c_j = -1. \quad (5.10)$$

As can be read off the indices of the sums, m points have been assumed to belong to the foreground class and $n - m$ points to be background for $1 \leq m < n$, m to be determined by the class assignments.

Using the prior information that the particles in the foreground have high grey values, the algorithm starts with assigning the histogram bin with the highest intensity value to the foreground, which yields an initial clustering for which the EMD is calculated. After that, the next lower intensity bins are subsequently added to the foreground until the EMD of the new cluster configuration becomes smaller than the previous one.

Comparing the results of the ECEDM, shown in Figure 5.10 to the previous segmentation methods, ECEDM clearly excels over the k-means clustering in so far that all bright intensities are correctly assigned to the foreground. Since it visually obtains a very similar result as the global thresholding without any training step, it also excels over this method. For the second image detail on the left in Figure 5.10 the bin coordinate above which all grey levels are assigned to foreground is determined as 0.54, which is higher than the optimal threshold 0.42 of the second image detail on the left of Figure 5.8. Hence, ECEDM reduces the number of background pixels incorrectly assigned to foreground. Varying the number of bins from $n = 4$ to $n = 64$, it can be observed that $n = 4$ yields too many false foreground pixels. The best result which is visually closest to the chip cell detail is obtained for $n = 8$. Taking into account the linear dependence of run time and memory usage on the number of bins, this ensures fast and efficient unsupervised segmentation.

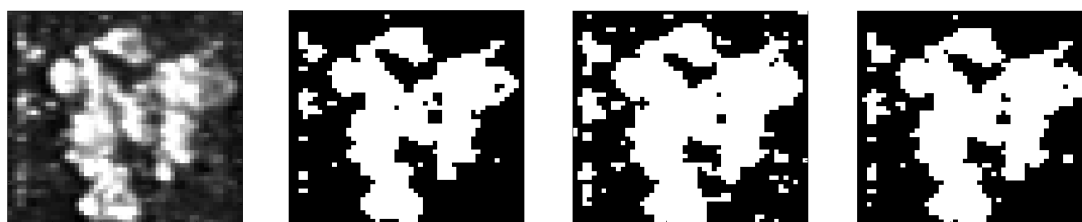


Figure 5.10: Left: Chip cell detail Centre left: Segmentation with $n = 8$ bins Centre right: Segmentation with $n = 4$ bins Right: Segmentation with $n = 64$ bins

5.4 Feature extraction and selection

For later use in the quality classification algorithm (s. Section 5.6), decisive features for the covered and uncovered cells have to be determined. Possible candidates, among others, can be found performing a principle component analysis or analysing the mean grey value of each cell and the number and size distributions of connected non zero grey values after segmentation.

In the computer vision community, more complex features and feature descriptors are in use, like histograms of oriented gradients [DT05] or SIFT feature descriptors [Low04]. Compared to the simple ones mentioned before, these features are often high-dimensional abstract concepts, having no meaning that can be directly correlated to bio-chemical quality measurements on the chip cells. Furthermore, the majority of the fancy features aims at detecting edges and corners, being the decisive parts in images. Yet, analysing the deposition quality of the particles, connected regions of pixels are of paramount importance.

No matter which features are considered for each chip cell, they depend on the image acquisition and preprocessing steps, but they should not be affected by scaling and rotation of the chip cell, since an enlargement of the image or a rotation of the chip in the setup do not change the bio-chemical result. Concerning translation, however, features can be sensitive to the relative position in the chip cell, since, particles deposited in the central spot of the cell are required while a particle deposition in the corners of the cell is welcome but not necessary. Furthermore, dividing the image into the single chip cells, the (semi-)automated method used here (s. Section 5.6) does not account for the grid on which no synthesis is performed. Therefore, the first step in feature extraction is to cut out the frame of pixels belonging to the grid, which is 10% of the length of the chip cell at its borders.

In the following paragraphs, promising features extracted from chip cell examples are presented and after pondering the advantages and disadvantages, the most decisive ones are selected for the discrimination of the chip cell qualities.

5.4.1 Principle component analysis (PCA)

As first noticed in [Pea01] it is possible to transform a set of images into its eigenbasis such that the vectors belonging to the largest eigenvalues, the principal components, point in the directions of the greatest variance in the images. For the analysis of the peptide chip images all N chip cell image details of 5 images of peptide chip depositions like Figure 5.1 are used. Then, each cell image detail of $M \times M$ pixel size is converted into a vector of length M^2 such that all cell images together form a matrix A of size $M^2 \times N$. From all these N image details, the mean image (in form of a vector) is determined and subtracted from each image (vector). Afterwards, a singular value decomposition is applied to the matrix A resulting in

$$A = U D V^T$$

The matrix D is of size $M^2 \times N$ and contains all eigenvalues sorted from high to low values, while the rows of U , which is of size $M^2 \times M^2$, represent the eigenvectors of A . Thus, the eigenimages as shown in Figure 5.11 can then be obtained by multiplying each eigenvector with its respective eigenvalue and reshaping the resulting vector back into an $M \times M$ image. Regions that show very high (red) and very low (blue) values set the range of variance for the respective eigenimage.

As can be observed from the first image in Figure 5.11, the highest variance in grey values for all chip cells can be found in the centre of each cell. This could have been expected beforehand, since the evaluation is performed on both cells with particles and uncovered cells having their largest difference in grey values in the centre of the cell. Analogously, the highest variance in grey values for covered cells (first image of the second row) can be explained: since almost all cells have a covered circle in their centre, the highest variance of coverage can be found around that circle.

The following eigenimages then show the variance between the left and right and the lower and upper parts of the cell, respectively. Eigenimages for smaller eigenvalues, from fifth eigenvalue onwards, are not shown here because they describe smaller details of minor importance in the variances of grey values.

For the uncovered cells, even the highest variance is already of minor importance, since all eigenimages for the first three eigenvalues look very similar, comparing the images of the third row with each other. Only for the fourth eigenimage, it can be observed that there are large incorrectly deposited particles. Analysing more cell details from further images, the eigenvalue to this eigenimage decreases even further, which also decreases its importance. Hence, the variance among the uncovered cell details is very small compared to that of the covered cells. An interesting fact, however, is that the variations of grey value for the uncovered cells mainly lie at the cell detail boundaries, indicating that incorrectly deposited particles on uncovered cells are mostly located at the cell borders.

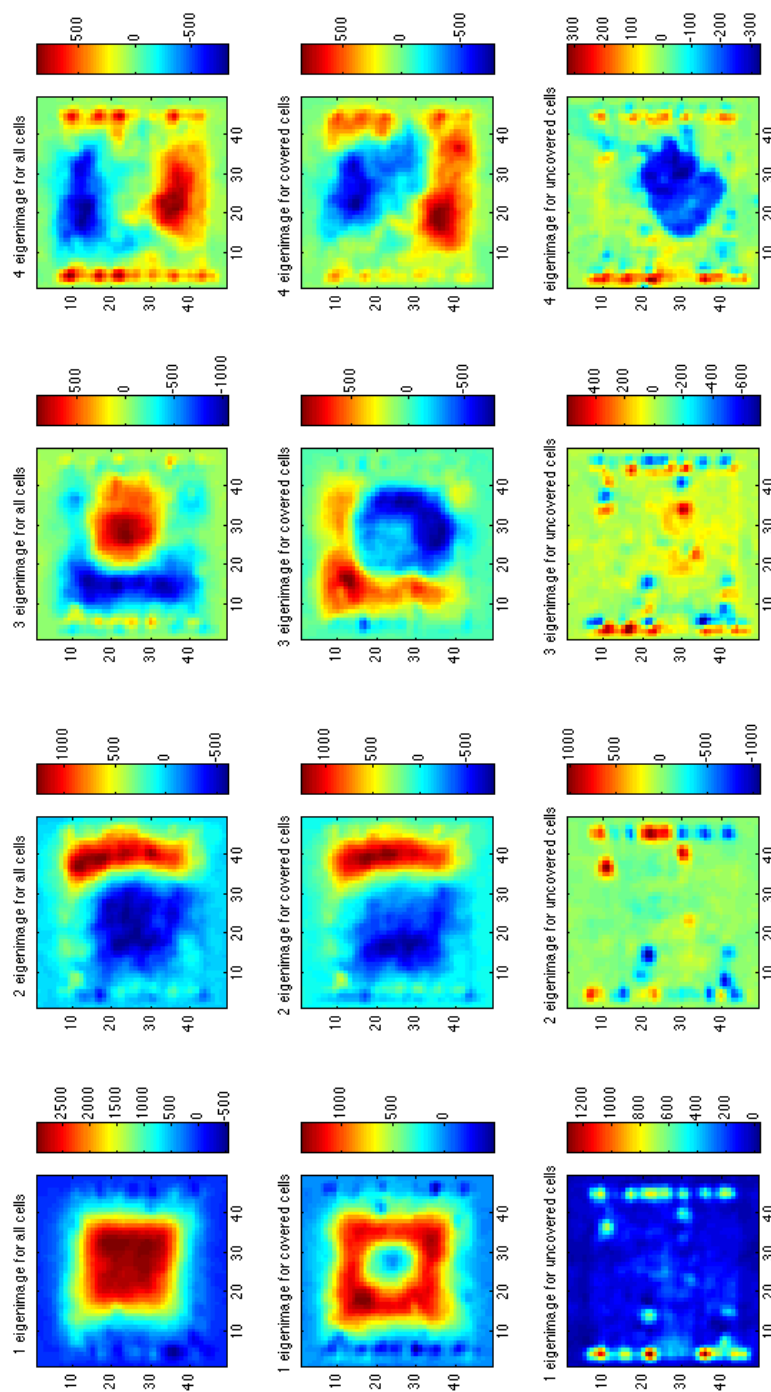


Figure 5.11: First 4 eigenimages (49×49 pixels), i.e. four highest deviations in coverage: Top row: between all 769 cells (including 385 covered and 384 uncovered cells) Centre row: between 385 covered cells only Bottom row: between 384 uncovered cells only

Comparing the first four eigenimages from the first row to those in the second row, a similarity can be noticed, which means that the highest variances in grey values are mainly caused by the covered cells.

Processing time can be saved by calculating only the top N eigenvalues and eigenvectors of U in case if M^2 is larger than N . Consider for example the cell details shown in Figure 5.11 where $N = 769$ and $M = 49$, yielding $M^2 = 2401$.

Although the results look like PCA yields reliable features to distinguish covered from uncovered cells, the method bears problems, as the following arguments will indicate. First of all, the chip cells in the image are extracted by defining the length of one chip cell and the origin of the first chip cell in the image (s. Section 5.6), so that the other cells are subsequently extracted. Since the defined cell length in the image does not fully match the actual cell length, this automated cell extraction does not achieve a perfect alignment of all cells, i.e. the pixelwise operations performed in the PCA are applied to pixels that do not represent the same location in all chip cells. Computationally, this phenomenon can be investigated by calculating the reconstruction error of a chip cell c when representing it as a linear combination of the top n eigenvalues λ_j times the respective eigenvectors v_j :

$$c - \mu - \sum_{j=1}^n \lambda_j \cdot v_j = \sum_{j=n+1}^N \lambda_j \cdot v_j \quad (5.11)$$

where μ is the mean value of all cells c which is subtracted before the actual PCA decomposition of the data. If the last eigenvalues and eigenvectors are less important, as expected by construction of PCA, the reconstruction error should decrease when adding more eigenvalues and eigenvectors to the linear combination. In the example shown above, however, the error does not decrease when adding 100 eigenvalues and eigenvectors to a linear combination of 10 eigenvalues and eigenvectors.

Thus PCA does not result in suitable features for the quality analysis algorithm, as the single chip cells would have to be aligned in order to extract the actual principle components. This furthermore implies that good features should be independent of pixelwise comparisons.

5.4.2 Mean grey value

Analysing the mean grey value distributions for covered and uncovered cells for five image samples (i.e. 773 chip cells analysed in total) yields the results shown in Figure 5.12. Thereby, covered cells are defined as those cells which are to be covered by particles, analogously, uncovered cells are those which should remain free from particles after the particle deposition. Due to deposition errors, it may happen that uncovered chip cells are covered with some particles so that the mean grey value

of these cells increases. On the other side, covered cells can be sparsely covered such that the mean grey value of those cells is decreased. These two phenomena will complicate the distinction of uncovered from covered cells in an unsupervised classification. Therefore, the mean grey value alone will not suffice to achieve a good classification unless the training sample shows a good deposition with no or only a few outliers.

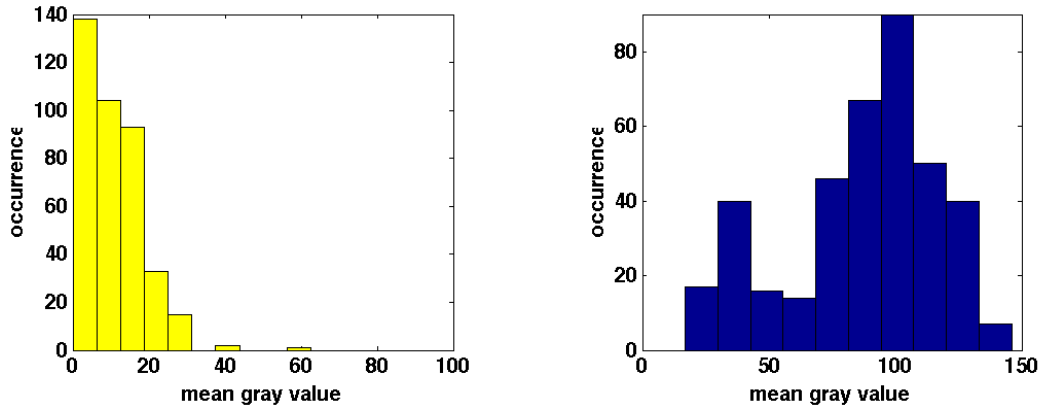


Figure 5.12: Left: Mean grey value histogram for uncovered cells Right: Mean grey value histogram for covered cells

As can be determined from the histograms, the mean grey values for 99% of all analysed 386 uncovered chip cells is below value 30 (0.12 for normalised grey values), while 4% of the 387 covered cells have a grey value below 30. This indicates that the two classes have a small overlap in their data ranges and are therefore separable, allowing for a few outliers. Considering those outliers for the uncovered cells, it turns out that these cells are well covered with particles like covered cells, i.e. the algorithm categorises them correctly according to their degree of coverage and not due to improper illumination conditions. Investigating the lower end of the mean grey values of the covered cells, the same statement can be made.

As will be elaborated in Section 5.7, using this simple feature a quick consistency check can be performed. If the cells classified as covered and uncovered do not match the deposition pattern of the chip, a grave malfunction has occurred and detailed quality analysis is not required anymore, as the deposition has to be redone then.

5.4.3 Spot coverage

After segmentation, the coverage of circular spots of different sizes in the centre of the chip cells can also be extracted as a feature. Example chip cells are shown in Figure 5.13 for a small radius of 0.2 times the cell length, covering the central 15% of the cell and for a large radius of 0.4 times the cell length, covering the central 50%

of the cell. Depending on the quality measure which will be defined in Section 5.5 the percental coverage of the central areas marked by the red circles can serve as a feature that defines thresholds for minimum coverage and maximum contamination, respectively. The radius of the area(s) of interest around the cell centre can also be adjusted by the quality measure defined in Section 5.5. Compared to the mean grey value, the spot coverage emphasises the central part of a chip cell and assures that particles form a large connected region in the part which is most likely to be covered again by particles in the next deposition step.

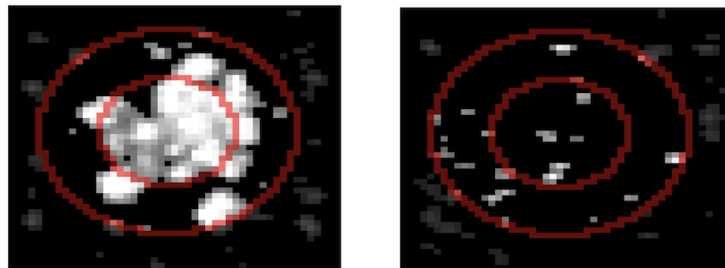


Figure 5.13: Left: Covered cell with marked central areas Right: Uncovered cell with marked central areas

This feature also has a linear time complexity with respect to the number of pixels, first pixel-wise multiplying the circular mask with the chip cell detail, then summing over all non zero entries.

5.4.4 Number and size distributions of particles

Finding the number and size for the particles on the chip decisively depends on the segmentation process, as Figure 5.8, Figure 5.9 and Figure 5.10 indicate. If the segmentation threshold is set too low, the number of particles found will be smaller than the actual amount and the size will be overestimated, leading to a too optimistic evaluation of the spread of amino acids on the chip for covered cells and increasing the amount of contaminations beyond the really existing ones for uncovered cells. Vice versa, if the threshold is too high, contaminations of uncovered cells and amino acid deposition on covered cells are both underestimated.

Assuming that the threshold is set such that the number and size of particles approximate reality well, the image can be segmented. Subsequently, a search for connected regions is performed on the segmentation result by means of the MATLAB intrinsic routine *bwlabel*, which uses the algorithm described in [HS92]. Briefly summarised, as outlined in the MATLAB help, the algorithm calculates the following steps until convergence:

1. run-length encode the input image
2. scan the runs, assigning preliminary labels and recording label equivalences in a local equivalence table
3. resolve the equivalence classes
4. relabel the runs based on the resolved equivalence classes

After convergence, the connected regions (defined by 4- or 8-neighbourhoods of pixels) are labelled with numbers. Finding the maximum of these automatically yields the number of particles in the chip cell detail. In order to account for the limited optical resolution, those connected regions smaller than the resolution limit have to be discarded before counting. The size of one connected region can be calculated by summing all pixels with the same label number.

Determining the degree of contamination in Section 5.5, the number and sizes of particles combined with their location in the chip cell can contribute to define the quality measure for uncovered cells. For the covered cells, the size and location of the largest connected region could be of interest in order to define a quality measure.

As an example, consider the images in Figure 5.14, the left one shows the size of the largest particle per cell (normalised to the total area of one chip cell) and the right one the number of connected pixel regions (i.e. particles) above the optical resolution limit. The threshold for the segmentation in this case is set to 0.4750, which was determined by thresholding in the training step described in Section 5.6.

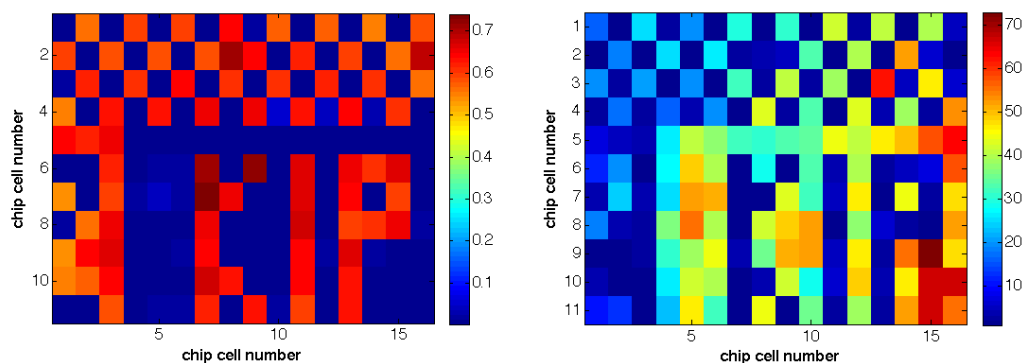


Figure 5.14: Left: Percental size of largest particle per cell Right: Number of particles above resolution per cell

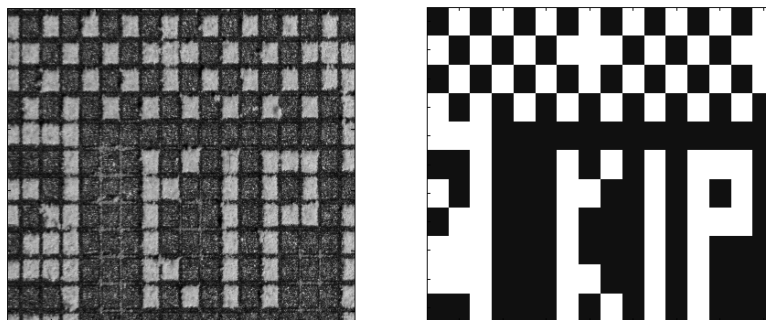


Figure 5.15: Left: Grey valued image of the particle deposition Right: Actual deposition pattern

Comparing the results obtained by the particle size and particle number features with the original image and deposition pattern shown in Figure 5.15, it is obvious that the size of the largest particle could also serve as a separator of covered and uncovered cells. Another option would be to combine the mean grey value feature and the size of the largest particle to a two dimensional feature space, in which covered cells are separated from uncovered by means of a Support Vector Machine as described in Section 5.6.2 for the one dimensional case. As it will be shown in Section 5.6.2, however, the mean grey value utterly suffices to distinguish the two classes of chip cells from each other so that the size of the largest particle can contribute to the quality measure. Hence, the largest connected region of particles can be a measure of the homogeneity of the deposition.

Contrary to that, the number of connected pixel regions is not so good a feature to distinguish covered from uncovered cells because finding the separating threshold is more difficult than in the case of the largest particle or the mean grey value. Consider, for instance, the chip cells in the lower left part of the right image in Figure 5.14 where covered and uncovered cells both show low numbers of particles, i.e. the covered cells are uniformly covered with only a few connected regions and the uncovered cells are contamination free.

For the correlation of the deposition image with the measurements of Chapter 4, these two features are the most powerful ones to predict the coupling of the amino acid layer under investigation to the growing peptide. The run time to extract these features strongly depends on the convergence velocity of the routine that merges the regions of connected pixels and scales with the size of the chip cell, yet is fast enough to meet the requirements in all cases.

5.5 Quality measures

As the quality of the results strongly depends on the underlying peptide synthesis model, the goal of this section is to investigate the strengths and disadvantages of several models, starting from the simplest one, expert labelling, to more involved ones using bio-chemical and deposition-adapted previous knowledge, as set up in Chapter 3 and Chapter 4. The most important issue for the bio-chemical models is the correlation between the particle deposition as shown in the microscope images and the actual amino acid coupling to the growing peptide chain.

5.5.1 Expert labelling quality measure

Intuitive quality measure

The simplest quality measure can be defined by asking experts about the characteristics of good and bad particle depositions, which can be qualitatively summarised as follows:

- The centre of chip cells to be covered should be covered contiguously:
This requirement ensures that the density of correctly assembled peptides in the central spot of each chip cell is approximately the desired one so that the success of the synthesis for this part of a cell can be proved by anti-body coupling and a subsequent fluorescence detection reaction as described in detail in [Blo09].
- The centre of chip cells to be uncovered should be free of particles:
Analogously to the first item, the second one ensures a high density of correctly assembled peptides in the central spot of each chip cell. Concerning the question what free of particles means, expert opinions differ about the maximum particle size that is believed not to cause contamination and therefore allowed on cells which should not be covered.
- The particle deposition in the corners of the cells is not of great importance:
This relaxation of the strong requirements that the entire cell should either be covered or not is motivated by simulations of particle depositions that revealed that the central spots of the cell are most likely to be covered with the probability of deposition in the chip cell corners being much lower.

In order to quantify these statements, the experts were asked to set up a quantitative quality measure based on the percental coverage of two circles around the chip cell centre with radii 0.2 times the length of the chip cell and 0.4 times the chip cell length, so that the chips cells can be classified into the categories *good*, *satisfactory* and *bad*. Instead of introducing three classes, it would have also been possible to reduce the classification problem to the two categories *good* and *bad*. However, the third, intermediate class is convenient for classifying those chip cells which are not

clearly *good* or *bad*, so that the expert is not asked for a hard classification decision that they cannot give with certainty. Table 5.1 then summarises their definition of a quality measure which is to be tested experimentally in Section 5.7.2.

cell type	good	satisfactory	bad
covered	> 90% of inner circle cov. > 50% of outer circle cov.	> 90% of inner circle cov. > 0% of outer circle cov.	other
uncovered	< 10% of inner circle cov. < 20% of outer circle cov.	< 20% of inner circle cov. < 30% of outer circle cov.	other

Table 5.1: Quality measure defined by experts based on two circular spot features

Another quality measure based on experts can be set up by letting the experts label training images taken from actual syntheses and learn a classifier on this data. But, this method does not work out well, because assigning each chip cell (covered and uncovered ones) a point in feature space, the sets of *good* and *bad* chip cells are highly non separable, as shown in Figure 5.16 on the left, where the features for each chip cell together with their expert labelling are plotted for one representative toner deposition image. Adding the largest connected region of pixels as third dimension does not improve separability, as demonstrated on the same example image in Figure 5.16 on the right. Therefore, the theoretical measure defined in Table 5.1 is used instead of a trained classifier.

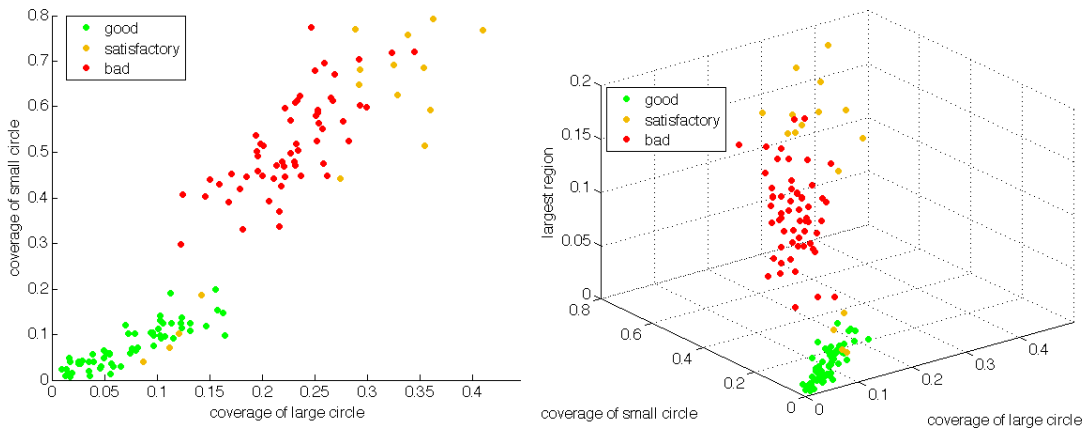


Figure 5.16: Left: Two-dimensional feature space as defined in Section 5.4.3 with user labels for covered and uncovered cells of the deposition shown in Figure 5.1 Right: Three-dimensional feature space with user labels, *good* deposition is encoded in green, *satisfactory* in yellow and *bad* in red

Discriminative feature based quality measure

While the quality measure defined by experts rather arbitrarily combines different features to find a discriminative feature space, image processing methods are capable of determining the most discriminative features systematically. Without further information on the input data, this can be performed in an unsupervised learning step, e.g. by dimensional reduction. First, a higher dimensional feature space consisting of all possible features is built, which is projected onto the k most discriminative feature dimensions. The latter are used to define the quality measure, since the separability of the training data given these features is maximised. Such dimension reduction methods include PCA, which is a linear dimension reduction method, or other non-linear methods as, for instance, a non-linear version of PCA. See [Fod02] for an overview of the most common techniques. If labelled training data is available, a supervised learning of the quality measure must be considered in order to separate the predefined classes, for instance by means of a multi-class Support Vector Machine or a k -nearest-neighbour approach¹. In order to find the most decisive features in theory, feature spaces of all possible feature combinations are tested and the one on which the classifier yields the least number of misclassified samples is chosen. This is a combinatorial problem whose complexity can be easily calculated, assuming that the total number of features is n . Then, the number of all possible combinations consisting of exactly k features is given by the binomial coefficient and the total number of all possible combinations P is

$$P = \sum_{k=1}^n \binom{n}{k} = \sum_{k=0}^n \binom{n}{k} - \binom{n}{0} = 2^n - 1 . \quad (5.12)$$

As the number of possible combinations grows exponentially with the number of available features, solving the combinatorial problem should be restricted to a small number of features.

Furthermore, the number of samples used for training should be calculated in advance. This depends on the number of possible points in feature space², since the samples should be chosen such as to form a representative ensemble of this feature space. Having too few samples compared to the total number of possible points, the training result will strongly depend on the sampled points and hence lead to overfitting. This means, that the trained classifier will not perform well on further test data sets. Determining the extensions of each feature separately, the number of possible points in the feature space of each combination can simply be calculated.

¹These two methods have already been introduced in Section 1.4.2, a two-class SVM will further be used to separate covered from uncovered cells as described in Section 5.6.

²As each feature bases on discrete pixel values, it is easy to see that the number of points in feature space depends on the finite number of values for each feature and is thus finite itself.

feature	formula [number of points]	example
mean grey value	$[0, 255]$	256
small circle coverage	$[0, \pi (k_s \cdot \text{cell length})^2]$, $0 < k_s \leq 1$	302
large circle coverage	$[0, \pi (k_l \cdot \text{cell length})^2]$, $0 < k_l \leq 1$	1207
number of particles	$[0, (1/2 \cdot \text{cell length})^2]$	601
size of largest particle	$[0, (\text{cell length})^2]$	2402

Table 5.2: Possible values of each feature, defined in Section 5.4, depending on the cell length of each chip cell and the fraction parameters k_s and k_l that define the radii for the circular features. The values determined by the formulas should be rounded to the next integer value below the actual value, those in the rightmost column are calculated for $k_s = 0.2$, $k_l = 0.4$ and cell length = 49 pixels

In Table 5.2, the intervals of possible values in dependence of the cell length and the fraction parameters are listed and the number of possible values is calculated for one specific example set of parameters. Due to correlations between the single features, multiplying the dimensions of the features to create the combined feature space is an overestimate of possible values, but shows in which range the number of values is. For instance, the smallest combination of two features in this example already yields a feature space of 77312 possible values when using the mean grey value and the small circle coverage. Assuming only half of these values to be realistic, approximately 4000 chip cells would have to be used in the training step in order to form an ensemble to represent 10% of this reduced feature space. This means, that, apart from the available 600 chip cells already labelled by experts, at least 3400 more would have to be labelled for this systematic feature selection. But, taking into account that this feature selection would solely base on expert assumptions with no bio-chemical indication whether the chips cells classified as *good* really show a high density of correctly coupled amino acids, the focus concerning quality measures is put on finding bio-chemical groundtruth first, which itself then determines the features of relevance.

5.5.2 Bio-chemically implied quality measure

Extraction of the chip cells

The TOF-SIMS surface analysis results gained as described in Section 4.3.1 are investigated concerning their usability as a quality measure based on image processing. First, the image for the tryptophan detection, which is 128×128 pixels in size, is converted to grey scale and divided into the single chip cells by the training routine of Section 5.6, yielding 16 chip cells of size 27×27 pixels.

Enlargement of the coupling signal image

In order to correlate this image with the one taken after the deposition of the particles, it has to be enlarged by factor 7 to a cell length of 189 pixels³. The best method to do so is convolving the image to be enlarged with Lanczos-2-kernel, which is one of the optimal interpolation filters for image enlargement, as investigated in [Tur90]. The kernel function with which the image is to be convolved, analogously to Equation 5.1, is given by

$$l(p_1, p_2) = \begin{cases} \operatorname{sinc}(p_1 - p_2) \cdot \operatorname{sinc}\left(\frac{p_1 - p_2}{a}\right) & -a < p_1 - p_2 < a; p_1 \neq p_2 \\ 1 & p_1 = p_2 \\ 0 & \text{otherwise} \end{cases} \quad (5.13)$$

where a can be chosen arbitrarily to fix the window size of the kernel. For Lanczos-2-kernel, $a = 2$. The result of the enlargement can be seen on an example covered cell in Figure 5.17. Compared to using a discretised version of the sinc-function as kernel, Lanczos windowed version yields better results due to anti-aliasing and visual contrast enhancement.

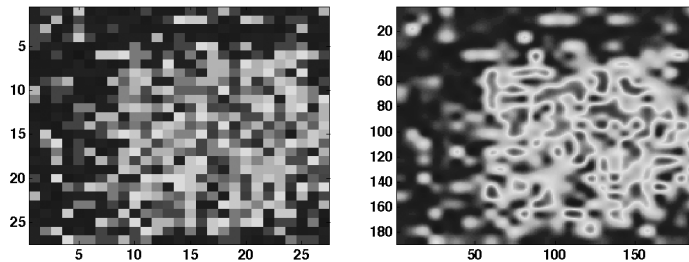


Figure 5.17: Left: Chip cell detail in original size (27×27 pixels) Right: Enlarged by convolution with Lanczos-2-kernel (189×189 pixels)

Alignment and matching

Due to the low contrast in the grid regions of the tryptophan image, translations may still be required for cell alignment before segmenting the single cells. In order to account for that, the image with the potassium signal (see Section 4.3.1) is used for reference. As it is recorded in the same detector as the tryptophan image, they are aligned. Furthermore, dividing the potassium image into the same chip cells as the tryptophan image, the location of the grid becomes clearly visible. So, the translations by which the origins of the chip cells in the tryptophan image must

³Instead, downsampling the deposition image for correlation and then enlarging it again to apply the correlation results to further deposition images leads to two resize steps.

be shifted in order to match the coordinates of the deposition image can easily be determined.

The regions that are finally matched consist of those pixels that start at the translated origin of a chip cell and cover the rest of it in the case of the tryptophan image, while the chip cell in the deposition image is adapted to the number of pixels in the detail of the tryptophan image. Thus, the maximum overlap of each chip cell in both images is used for correlation, as shown for all chip cells in Appendix B.1.

Another option for alignment could have been to match the chip cell contents of the two images directly by applying a registration and matching algorithm, e.g. finding the largest overlap of the segmented foreground in both images. Yet, as the foreground is the object under investigation, it cannot be used for image alignment. Using the grid for alignment thus enables to detect a systematic shift in deposited particles when melting and coupling. This effect may be caused by unbalanced chip fixation on the circuit board or improper positioning in the oven, resulting in a unidirectional melt flow instead of a uniform distribution.

Image segmentation

Then, the image containing the signals of the coupled tryptophan on the chip and the image showing the respective particle deposition are segmented. For the tryptophan image, thresholding and ECEMD as described in Section 5.3 are applied. Since the covered cells are separable from the uncovered ones, the segmentation threshold can be calculated out of the mean grey values (see Section 5.4.2) as the maximum mean grey value for the uncovered cells and the minimum mean grey value for the covered cells. In this case, the result is 0.1970, after normalisation of the image to the highest intensity value. Using ECEMD shows very similar results as thresholding, which can be observed in Figure 5.18.

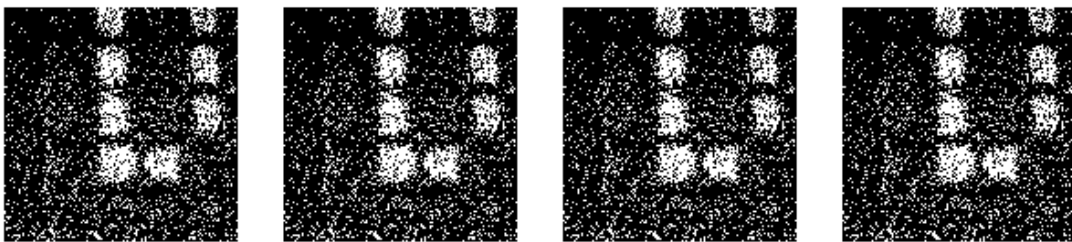


Figure 5.18: Left: Chip cell detail segmented by thresholding Centre Left: Segmented by ECEMD with 4 bins Centre Right: Segmented by ECEMD with 8 bins Right: Segmented by ECEMD with 12 bins, foreground in white, background in black

The low signal to noise ratio of 2:1 is clearly visible, yet, the noise should not be smoothed out as the noise model for the TOF-SIMS data is not known. Hence, smoothing the image with a Gaussian kernel, for instance, would lead to underestimated contamination. Concerning the mean of the noise, Figure 5.18 indicates, that it certainly is above zero. Details of the thresholded segmentation as well as the segmentation for the deposition image can be found in Appendix B.1.

Correlation

Correlating the signal of the coupled tryptophan with the particle deposition before melting, a confusion matrix per chip cell is calculated consisting of the four entries:

- True positive (TP) are those pixels that are segmented as foreground in the deposition and the detection image. They indicate which covered locations actually lead to coupled amino acids.
- False positive (FP) are those pixels that are segmented as background in the deposition image but foreground in the detection image. These pixels show the locations where amino acids from undetected, deposited particles lie.
- True negative (TN) are those pixels that are segmented as background in the deposition and the detection image. These pixels are the negative control group, as they are not covered by particles and show no signal of coupled amino acids.
- False negative (FN) are those pixels that are segmented as foreground in the deposition image and background in the detection image. For them, it is interesting to investigate, why no coupling happens at those locations that are covered with particles.

Figure 5.19 shows the correlations in form of the confusion matrix for one covered cell (chip cell 6 in Appendix B.1).

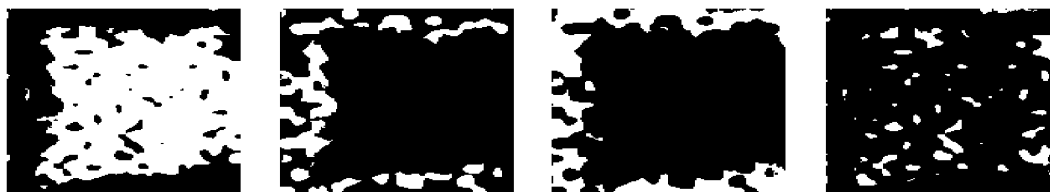


Figure 5.19: Left: Deposited and detected for a covered cell (TP) Centre left: Not deposited but detected (FP) Centre right: Not deposited and not detected (TN) Right: Deposited but not detected (FN), correlation in white, no correlation in black

Despite the low signal to noise ratio, a high number of correlations (white) for the true positive pixels can be observed. The exact number of pixels normalised to the total number of analysed pixels can be found in Table 5.3. The distribution of false positive pixels is subject to noise, however, a systematic concentration of false positive pixels is found at the borders of the particle deposition, which is also the case for the other covered chip cells (see Figure B.6). From this can be deduced that melting the particles enlarges the area of coverage and hence the number of coupled amino acids. On the average, this affects the first 5 to 10 pixels beyond the particle deposition border. As no preferred direction for the extended coupling is visible, the handling of the slides need not include special balancing of the chip on the circuit board or positioning in the oven. The true negative pixels form the largest fraction outside the deposition region. Yet, there are some circular black spots at the borders of the chip cell detail that are not recognised as true negative. They can be enlarged artefacts of one pixel noise or really belong to the class of coupled amino acid signals for whom no particle deposition could be detected. For those pixels belonging to the false negative class, it is possible that some of these signals are artefacts due to the low signal to noise ratio or that no amino acids could couple at these locations.

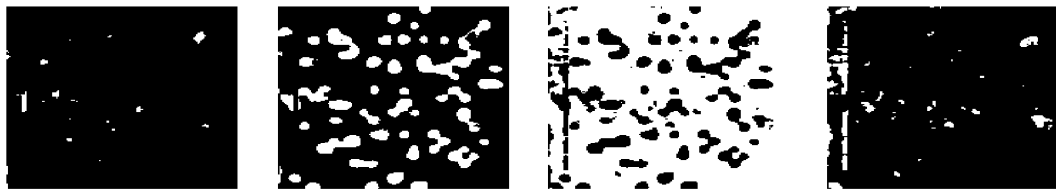


Figure 5.20: Left: Deposited and detected signals for an uncovered cell (TP) Centre left: Not deposited but detected signals (FP) Centre right: Not deposited and not detected signals (TN) Right: Deposited but not detected signals (FN), correlation in white, no correlation in black

Correlating the images for the uncovered chip cells, as shown in Figure 5.20, similar results are obtained. In this case, the number of true positives shows the detected contaminations. Yet, there is a large number of false positives, not only close to the borders of the particle deposition that decrease the true negatives. While conclusions for the covered cells are still possible under these conditions, a higher resolution and better signal to noise ratio are required for the uncovered cells, as the detection limit for contamination is one important parameter for the quality measure to be set up. The number of false negative pixels within the chip cell also has to be investigated further at better experimental conditions in the TOF-SIMS measurement, but, from the uncovered chip cells, a high correlation at the borders of the cells can be detected. This indicates that the number of coupling of deposited amino acids is

highly reduced in the regions of the grid electrode, supporting the assumption that the surface coating is broken or thinned out around the grid electrode, as stated in Section 4.3.1.

chip cell	TP	FP	TN	FN
1	0.0114	0.1777	0.7755	0.0355
2	0.0103	0.1888	0.7791	0.0218
3	0.0059	0.1492	0.8165	0.0284
4	0.0077	0.1724	0.7932	0.0268
5	0.0108	0.1620	0.7953	0.0319
6	0.5535	0.1509	0.2111	0.0844
7	0.4239	0.1595	0.3051	0.1116
8	0.0087	0.1631	0.7989	0.0294
9	0.0089	0.1790	0.7796	0.0324
10	0.3937	0.1579	0.3553	0.0931
11	0.0085	0.1702	0.7808	0.0405
12	0.2965	0.1600	0.4197	0.1238
13	0.0078	0.1143	0.8331	0.0448
14	0.3718	0.1129	0.3972	0.1181
15	0.0028	0.1006	0.8431	0.0534
16	0.2915	0.1068	0.4510	0.1506

Table 5.3: Confusion matrix entries correlating the signals of the deposition image with the coupled tryptophan, TP: deposited & coupled, FP: not deposited & coupled, TN: not deposited & not coupled, FN: deposited & not coupled, numbers all relative to the size of the analysed chip cell detail

Although TOF-SIMS measurements with higher resolution for the chip cells and an increased signal to noise ratio have to be performed, implications for the bio-chemically based quality measure can already be gained from the correlation of the first images. At first, the experts' assumptions that amino acids can be coupled at locations that are uncovered by particles but covered with melt when heating, is experimentally confirmed. Second, one particle deposition per chip cell suffices, a second one, as it used to be performed in the previous experiments, is not required, which reduces deposition time and toner material consumed by one half. Third, even without balancing the chip on the circuit board or the entire board in the oven, no preferred flow direction of the melt is observed.

For covered cells, coupling is possible at locations up to 10 pixels beyond the particle deposition border. Hence, a good feature for the quality measure is the largest connected region of pixels (see Section 5.4.4) with an extended border by this amount. Yet, this is only valid within the chip cell, approximately 5-10 pixels away from

the grid. Extending the border into this region might not lead to coupling as the examples of the false negatives imply. Furthermore, the latter also reveal that not all covered locations show a signal of coupled amino acids. Bio-chemically, this can have several reasons which should also be subject to further experiments. First, it is possible that the coupling sites are difficult to reach for the amino acids due to steric hindrance, this effect is also dependent on the type of amino acid and coupling site. Second, there is a probability that an accumulation of very small particles does not contain any amino acids at all or that some of them are unstable when heating. It is also possible that there are amino acids that cannot propagate through the melted toner matrix and reach the coupling sites.

The correlation of uncovered chip cells shows that the number of possible contaminations in the particle deposition image is overestimated, as many incorrectly deposited particles do not lead to coupled amino acids (comparing the TP signal with the FN signal in Figure 5.20).

Hence, in order to complete the quality measure, for the uncovered cells in particular, further experiments should be performed at the highest lateral resolution possible, which is $0.3\mu\text{m}$ instead of the current $4\mu\text{m}$ and the signal to noise ratio should be increased as well. With this setup, a reliable estimate for the distribution of false positives can be achieved, which will show whether the high number of coupling signals at locations with no particle deposition is realistic or not. Additionally, correlations should be performed for the coupling of different amino acids, in order to investigate the dependence of the coupling probability on the stereometry of the molecules. As a variety of surface modifications with different numbers and structures of coupling sites is available as well, the one with optimum coupling rates can also be determined by the TOF-SIMS supported quality analysis.

5.6 Training phase

5.6.1 Selection of a representative training set

Having introduced the single parts required to build a quality evaluation program, the first part, the training phase, can now be assembled. In the training phase, the user has to provide the algorithm with one image which is representative concerning the deposition quality for the ones to come in the testing phase. A good choice is an image which has an equal number of covered and uncovered cells, so that the number of samples for the two chip cell classes is equal. Since the training is a supervised one, the correct deposition pattern which cell is to be covered and which is to be left uncovered should also be known. Furthermore, the user is expected to specify the image height and image width as well as the microscope objective magnification, the numerical aperture, the maximum wavelength of the light used and the pitch of the camera chip.

Technical requirements, also valid for the test phase, are a uniform illumination of the chip and an alignment of the chip grid with the pixel grid in the camera/ in the image, i.e. a horizontal grid line on the chip should have constant y-coordinate in the image. As a rule of thumb, the maximum rotational distortion should be smaller than one pixel length.

If desired, the image can be enhanced by noise reduction and calibration routines. If the camera software does not offer flat-field correction (see Section 5.2.2), this calibration can be included in the training phase. In this case, the user has to take a white picture first and provide it to the calibration subroutine. For noise reduction, the user can include the convolution with a Gaussian filter as described in Section 5.2.3. Since the test images then have to be preprocessed in the same way as the training image to assure that all parameters are consistently adjusted, the preprocessing parameters are also added to the list of parameters to be saved in the parameter file at the end of the training phase. The parameters for the preprocessing include two booleans that state whether a flat-field correction or a convolution were executed or not. An overview of all the parameters in the parameter file can be found in Appendix B.2.

5.6.2 Training of the parameters

Chip cell length

The first parameter to be determined is the length of one chip cell in units of pixels in the training image. In order to do so, the user is asked to delimit one chip cell in width and height by means of mouse clicks on the training image. Since the cells are quadratic, the mean of the cell width and height is the chip cell length. The goodness of fit can be investigated after the chip cell length calculation because the grid lines that delimit the chip cells are drawn over the entire training image so that the user can decide whether the cell length is appropriate or not. In the former case, the grid lines go through the centre of the metal layer that separates the chip cells from each other as shown in Figure 5.21 on the left. In contrast to that, the right side of Figure 5.21 shows an incorrect cell length that leads to a misplacing of the drawn grid.

For the prototypical setups which are used frequently, the cell lengths are ready to use in the testing algorithm by specifying the so called *series number* that consists of the magnification of the microscope objective and a letter that encodes the microscope and camera used, for instance 10p means the setup with an objective of magnification 10 mounted on the Axiovert 35 microscope with the Progres C5 as image acquisition device. A detailed list of all supported series numbers can be found in Appendix B.4.1.

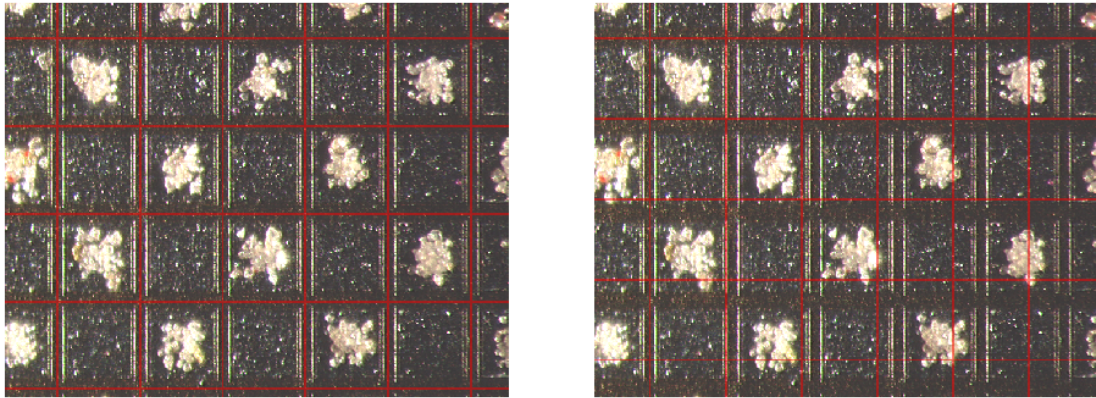


Figure 5.21: Left: Correct cell length Right: Incorrect cell length

Another option could be to use the known deposition pattern and the covered cells in the test image to set the grid. The grid lines could then be placed according to the intensity profile over the image. Having observed a series of at least two covered cells in one row, it is possible to determine the cell length as the interval between the centres of two adjacent covered regions or half of the interval between the centres of two covered regions separated by one uncovered cell, as shown in Figure 5.22. Yet, as Figure 5.22 demonstrates, this method strongly depends on the deposited particles, such that depositions of bad quality result in incorrect cell lengths and grid coordinates. Taking into account that the quality of deposition is to be investigated, this information should not be used to calculate the cell coordinates in order to prevent the analysis from systematic errors.

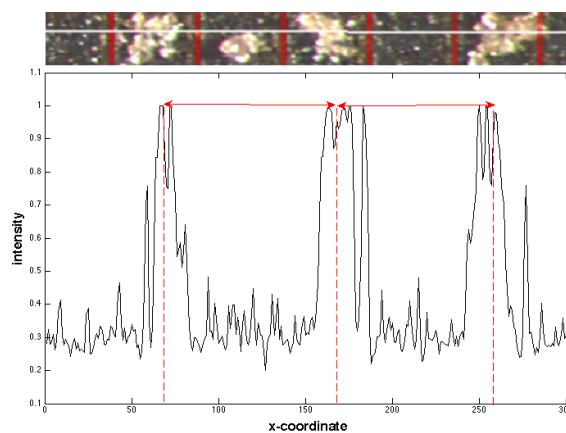


Figure 5.22: Intensity profile for the white line in the image, the red lines in the profile mark intervals twice the cell length as can be read off the image detail on top

Lateral resolution

The next parameter to be calculated is the number of pixels on the camera chip that correspond to the theoretical resolution limit of the optical setup. The latter, already defined as d in Equation 2.7, together with the pixel pitch of the camera p and the total magnification factor M then yields the resolution limit r in units of pixels on the image acquisition device

$$r = \text{ceil} \left(\frac{d \cdot M}{p} \right) = \text{ceil} \left(\frac{\lambda \cdot M}{2N_A \cdot p} \right) \quad (5.14)$$

where $\text{ceil}(\cdot)$ denotes the ceiling function. The user should keep in mind that $r = 1$ can indicate that the resolution of the microscope camera is worse than the resolution limit given by the optical setup. Considering for example the setup with $d = 1.4\mu\text{m}$, $M = 10 \cdot 0.63$ due to the microscope objective and a TV-adaptation in combination with the Zeiss Progres C5 camera, which has a pixel pitch of $p = 3.4\mu\text{m}$, then $r = 3$ pixels.

Knowing r becomes important when counting deposited particles and particle agglomerations (see Section 5.4.4), since then all connected pixels of size smaller than $r \times r$ that indicate deposited particles can be neglected. Downsampling the image according to the calculated resolution is another possibility. This enables fast run times due to the reduced number of pixels, according to Nyquist, however, super-sampling the image prevents aliasing effects and hence implies a higher accuracy of the results.

Classification threshold for covered and uncovered cells

The last parameter to be determined is the classification threshold that separates covered from uncovered chip cells. It is used in the testing phase as a first consistency check, comparing covered and uncovered cells in the image with the correct deposition pattern programmed into the chip.

Assuming a constant illumination over the entire chip, the classification threshold is a constant for all coordinates. It is chosen such that the majority of chip cells in the training image is correctly classified as covered or uncovered. For this classification, the mean grey value feature as described in Section 5.4.2 is used together with the information about the correct deposition pattern, so that the threshold for the mean grey value, i.e. the optimal classification threshold, can be determined. Allowing for misclassified samples on both sides, the problem can be tackled by means of a Support Vector Machine.

Let $\text{mgv}(\cdot)$ denote the function that assigns each chip cell c its mean grey value as described in Section 5.4.2 and let the class labels y_i for the i -th of n chip cells be -1 if it is uncovered, and $+1$ if it is covered. Then, as deduced in Chapter 7.1. of [Bis06], the separation problem of the uncovered and covered classes can be written

in the following form:

$$\inf_{w,b,\xi} \left\{ \frac{1}{2}w^2 + \nu \sum_{i=1}^n \xi_i \right\} \quad \text{subject to} \quad (5.15)$$

$$y_i (w \cdot \text{mgv}(c_i) + b) - 1 + \xi_i \geq 0 \quad \xi_i \geq 0 \quad \forall i = 1, \dots, n \quad (5.16)$$

where w is the vector normal to the separating hyperplane with the offset b , ergo, in this one dimensional case, w is parallel to the axis of mean grey values and b is a scalar number. ν is a free parameter to be set by the user. The higher ν the higher is the penalty ξ_i for misclassified samples.

In the case of separable classes, the distance in grey values by which the covered and uncovered classes are separated can be expressed as the distance between the maximum of the mean grey values for the uncovered cells and the minimum of the mean grey values for the covered cells. The mean of these two values then determines the classification threshold. Half of this distance is called the margin. The margin will also be added to the list of parameters and is used in the testing phase as described in Section 5.7.1. In the non separable case, the classification threshold is calculated⁴ as the fraction of the offset b divided by w and the margin is given as the maximum distance between this threshold and that misclassified sample point which is farthest away from it. As a general rule, the quality of the selected features is determined by the size of the margin and the number of misclassified samples lying within this region. The larger the margin that separates the two classes and the fewer misclassified sampling points lie within its range, the better the discriminative model.

As an example, consider the image as shown in Figure 5.23 on the left. Due to the one uncovered cell which has a large contamination in the upper left part of the image, the sample belongs to the non separable category, which can be confirmed by observing the histogram of mean grey values in Figure 5.23 on the right. This example also shows that the classification threshold is sensible, i.e. covered and uncovered cells can be separated well.

To investigate the effect of the free parameter ν , Table 5.4 shows the positions of the threshold and extend of the margin for varying ν for the example of Figure 5.23 with mean grey values in the interval $[0, 1]$. For $\nu = 0.1$, the classes are not separable, since the penalty term for incorrect assignments is too low, so that the optimal solution is obtained by merging the two classes into one. The maximum mean grey value for the uncovered cells is 0.43, while the minimum mean grey value for the covered cells is 0.42. Knowing this, the thresholds are such that in the optimal case there is only one misclassified sample, namely the uncovered cell with mean grey

⁴The threshold is given as the distance of the hyperplane to the coordinate origin in the Hesse normal form of the plane: $\frac{1}{\|w\|} (w \cdot \text{mgv}(c_i) - b) = 0$.

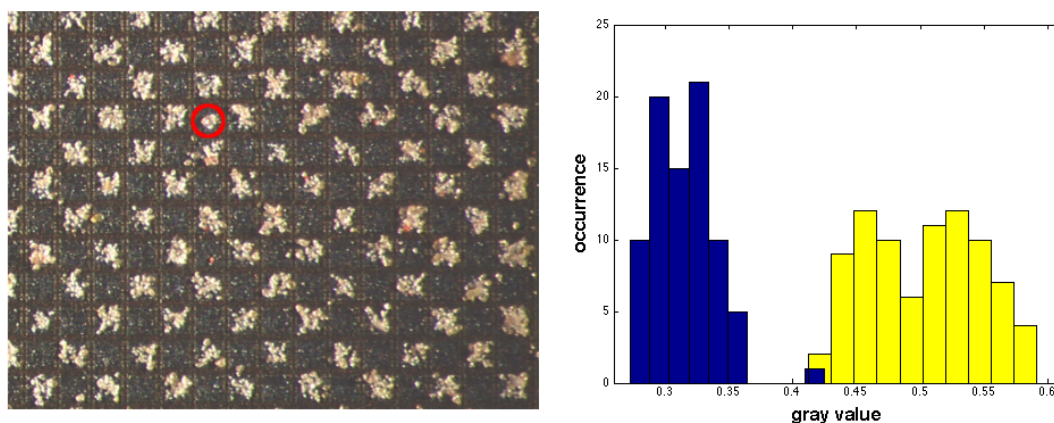


Figure 5.23: Left: Training image Right: Histogram of mean grey values in $[0, 1]$ of covered (yellow bars) and uncovered cells (blue bars), the one misclassified uncovered chip cell with high grey value actually belongs to the covered cells due to the large contamination (marked in red in the image on the left)

value 0.43, which is reached for $\nu = 10$ and does not change when increasing ν . Thus, the penalty weight ν should be of the order of 10 in order to achieve good class separation results. Although being trained on one example this threshold will be proved to work well for other deposition images in Section 5.7.2.

ν	w	b	threshold	margin
0.1	0.0000	1.2500	∞	0.0000
0.5	7.1328	3.0377	0.4259	0.0042
1	7.8624	3.3205	0.4223	0.0077
10	7.8356	3.2912	0.4200	0.0100
100	7.8356	3.2912	0.4200	0.0100

Table 5.4: Dependence of the class separation threshold and margin on ν

At the end of the training phase, all parameters are saved in a text file that is used in the testing phase. This bears the advantage of simple accessibility of the parameters, which will be useful defining interfaces to the automaton in Section 5.8.

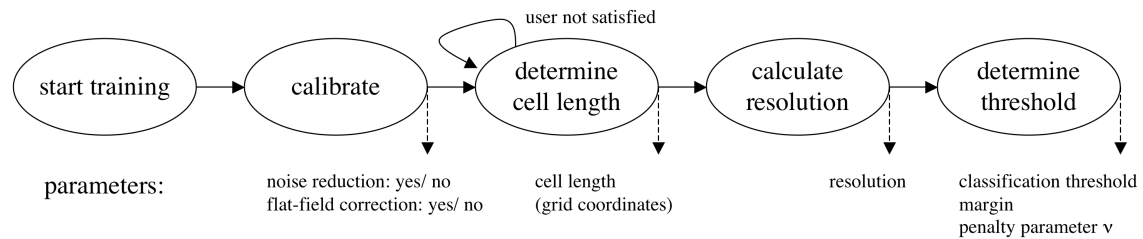


Figure 5.24: Summary of the single training phase steps and their outputs to be saved in the parameter file

5.7 Testing phase

5.7.1 Processing chain

At first, the test image is read in using the specifications in the parameter file. It should be taken under the same conditions as the training image, i.e. the illumination and setup conditions should not have changed. Furthermore, the current deposition pattern which is used to program the chip cell voltages is also loaded. The subsequent preprocessing steps are then performed in the same way as for the training image according to the loaded parameters.

Then, a grid that separates the chip cells from each other is laid over the image and the coordinates of all chip cells are calculated. For this purpose, the cell length of the chip cells from the training phase is needed. The user is asked to mark the origin of the grid (i.e. the upper left crossing point of the horizontal and vertical metal layers that separate the chip cells) by mouse click. Starting from this reference point, the grid lines, i.e. the borders of the chip cells, are drawn at intervals of the trained cell length until the border of the image is reached. The result is printed on the screen. As demonstrated in Figure 5.21 on the left, the grid is correctly positioned if the grid lines are placed in the centre of the metal layer that separates the chip cells. Since the user input will not always be at the same pixel in the image, the subsequent steps must be robust against a horizontal and vertical grid translation of a few pixels. The user input can be omitted if a positioning machine with high spatial accuracy is used to transport the chip to the quality analysis setup. In this case, the coordinates of the grid origin can be added to the list of input parameters. This method will be applied in the automated chip synthesis described in Section 5.8.

Knowing the chip cell coordinates in the image, the preprocessing steps are finished and the actual quality analysis can start. At first, the mean grey value of each chip cell is determined. To make it robust against translations of the grid only the central part of the cell is used for the mean grey value calculation and the pixels at the border up to $0.1 \cdot \text{cell length}$ are not taken into account. This guarantees

that particles on the metal layer that separates the cells have no influence on this first classification. The threshold and the margin determined in the training phase then classify the cells in a first unsupervised classification step as follows, which simultaneously shows the strength of the training parameters:

- uncovered cells are defined by those mean grey values that are smaller than the threshold minus the margin
- covered cells are defined by those mean grey values that are larger than the threshold plus the margin
- those cells whose mean grey values lie in the range of the margin around the threshold could belong to either class. Therefore, they build a third class, indicating how representative the trained parameters really are.

Using the uniquely classified uncovered and covered cells, the next step of the algorithm tries to match the actual deposition pattern to the current particle deposition. If there is a coincidence of more than 50% between the current covered and uncovered cells and the actual deposition pattern, the pattern matching is successful. This excludes coarse malfunctions of the synthesis setup, i.e. it confirms that the voltages on the chip were all correctly set and the conditions of the aerosol (particle density, particle velocity, pressure, etc.) were suitably tuned. If the coincidence is less than 50%, the algorithm terminates with the respective error message. If the training was performed with a representative image, then this error can only be caused by too many incorrectly deposited particles. In this case, the deposition should be repeated, as described in Chapter 5.8. Checking for such coarse errors first allows for fast error detection and correction. After a successful pattern matching, the quality analysis according to the expert labelling quality measure from Section 5.5 is started and the results are saved. The steps of the testing phase are shown in Figure 5.25.

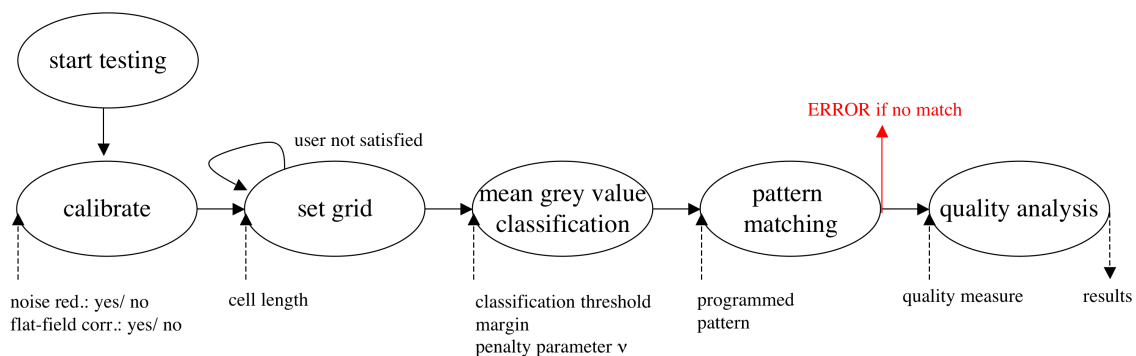


Figure 5.25: Summary of the single testing phase steps and their in- and outputs

5.7.2 Experimental results

Comparing the quality classes defined in Section 5.5.1 that are obtained in the algorithm with those of an expert labelling, four images with altogether 598 chip cells, half of them covered, half of them uncovered, are investigated. Since the vast majority of depositions is of very good quality, as has been shown in [KBN⁺10], there are only few images containing chip cells of all classes on which the strengths of the quality control algorithm can be shown.

In order to compare the classification obtained by the algorithm with that of the expert labelling, the confusion matrix C_{abs} is chosen as the appropriate measure. As the name indicates, its entries measure the misclassification between the different k classes. C_{ij} represents the number of chip cells the expert has labelled as class i but the algorithm has assigned to class j . Normalising the entries in C_{abs} to the total number of manually labelled cells of the same class, the relative confusion matrix C_{rel} is calculated.

$$(C_{rel})_{ij} = \frac{(C_{abs})_{ij}}{\sum_{l=1}^k (C_{abs})_{il}} \quad (5.17)$$

In Table 5.7.2, the results of the comparison between the expert labelling quality measure as defined in Section 5.5.1 and the algorithmic quality assignments are listed.

As can be read off C_{abs} , there are fewer chip cells that are labelled *bad* or *satisfactory* than there are *good* labels, which biases the evaluation in absolute numbers. Therefore, the relative confusion matrix C_{rel} lists the fractions of chip cells for each class.

Taking into account that the quality measure is based only on two simple features, the result is remarkably good, considering the fact that only one chip cell which is labelled *bad* by the expert is found to be *good* in the algorithm. From the row that contains the total number of analysed chip cells, it can be read off that the algorithm coincides with the expert in 98% of all cases of *bad* quality, while it has fewer coincidences for the other classes. The reason for this can be seen in Table 5.6, which shows the pictures of the deposition (left), the expert labels for the single cells (centre) and the class assignments determined by the algorithm (right). Comparing the chip cells that are labelled as *good* and *satisfactory* in the pictures on the left, it is hard to visually distinguish them. Hence, as already mentioned in Section 5.5.1, the confusion in these class assignments does not imply a bad performance of the algorithm but rather the uncertainties in labelling.

The performance of the algorithm itself can be judged by counting the chip cells that have been assigned better classes than those the expert determined. Compar-

image	cells	good	satisf.	bad	C_{abs}	C_{rel}
1	150	122	25	3	$\begin{pmatrix} 62 & 20 & 40 \\ 6 & 4 & 15 \\ 0 & 1 & 2 \end{pmatrix}$	$\begin{pmatrix} 0.51 & 0.16 & 0.33 \\ 0.24 & 0.16 & 0.60 \\ 0.00 & 0.33 & 0.67 \end{pmatrix}$
2	140	66	16	58	$\begin{pmatrix} 45 & 20 & 1 \\ 3 & 1 & 12 \\ 0 & 0 & 58 \end{pmatrix}$	$\begin{pmatrix} 0.68 & 0.30 & 0.02 \\ 0.19 & 0.06 & 0.75 \\ 0.00 & 0.00 & 1.00 \end{pmatrix}$
3	154	142	10	2	$\begin{pmatrix} 135 & 2 & 5 \\ 4 & 6 & 0 \\ 1 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0.95 & 0.01 & 0.04 \\ 0.40 & 0.60 & 0.00 \\ 0.50 & 0.00 & 0.50 \end{pmatrix}$
4	154	128	26	0	$\begin{pmatrix} 87 & 30 & 11 \\ 7 & 12 & 7 \\ 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0.68 & 0.23 & 0.09 \\ 0.27 & 0.46 & 0.27 \\ 0.00 & 0.00 & 0.00 \end{pmatrix}$
total	598	458	77	63	$\begin{pmatrix} 329 & 72 & 57 \\ 20 & 23 & 34 \\ 1 & 1 & 61 \end{pmatrix}$	$\begin{pmatrix} 0.72 & 0.16 & 0.12 \\ 0.26 & 0.30 & 0.44 \\ 0.01 & 0.01 & 0.98 \end{pmatrix}$
mean	150	114	19	17	$\begin{pmatrix} 82 & 18 & 14 \\ 5 & 6 & 8 \\ 1 & 1 & 15 \end{pmatrix}$	$\begin{pmatrix} 0.72 & 0.16 & 0.12 \\ 0.26 & 0.32 & 0.42 \\ 0.06 & 0.06 & 0.88 \end{pmatrix}$

Table 5.5: Comparison of expert labelling and algorithmic quality assignments per test image based on Table 5.1, the number of chip cells classified as *good*, *satisfactory* and *bad* in the left columns are counted with respect to the expert labels

ing them, 22 in total, to the 413 chip cells that have been correctly assigned, it becomes obvious that the algorithm can reliably detect cells of suboptimal quality. On the other hand, there are 163 chip cells that are assigned to lower quality classes by the algorithm than by the expert. This implies that the algorithm still classifies too pessimistically, leading to unnecessary repetitions of depositions which already might have been of sufficient quality.

Improvements on the rate of these unnecessary false alarms could be achieved by changing the requirements for the quality classes defined in Section 5.5.1. In order to investigate the influence of a change in the quality measure on the expert labelled examples of Table 5.6, the relaxation as listed in Table 5.7 is made.

Comparing Table 5.7 with the original quality measure of Table 5.1, the threshold value above which contamination on *good* uncovered cells results in a decrease of quality is relaxed from 20% to 30% coverage. This should cause an increase in the number of correctly assigned *good* chip cells compared with the expert labelling. Yet, the analysis, listed in Table 5.7.2, yields worse results in the number of correctly detected qualities, which can be observed when comparing the increased number of off-diagonal entries in Table 5.7 to Table 5.6.

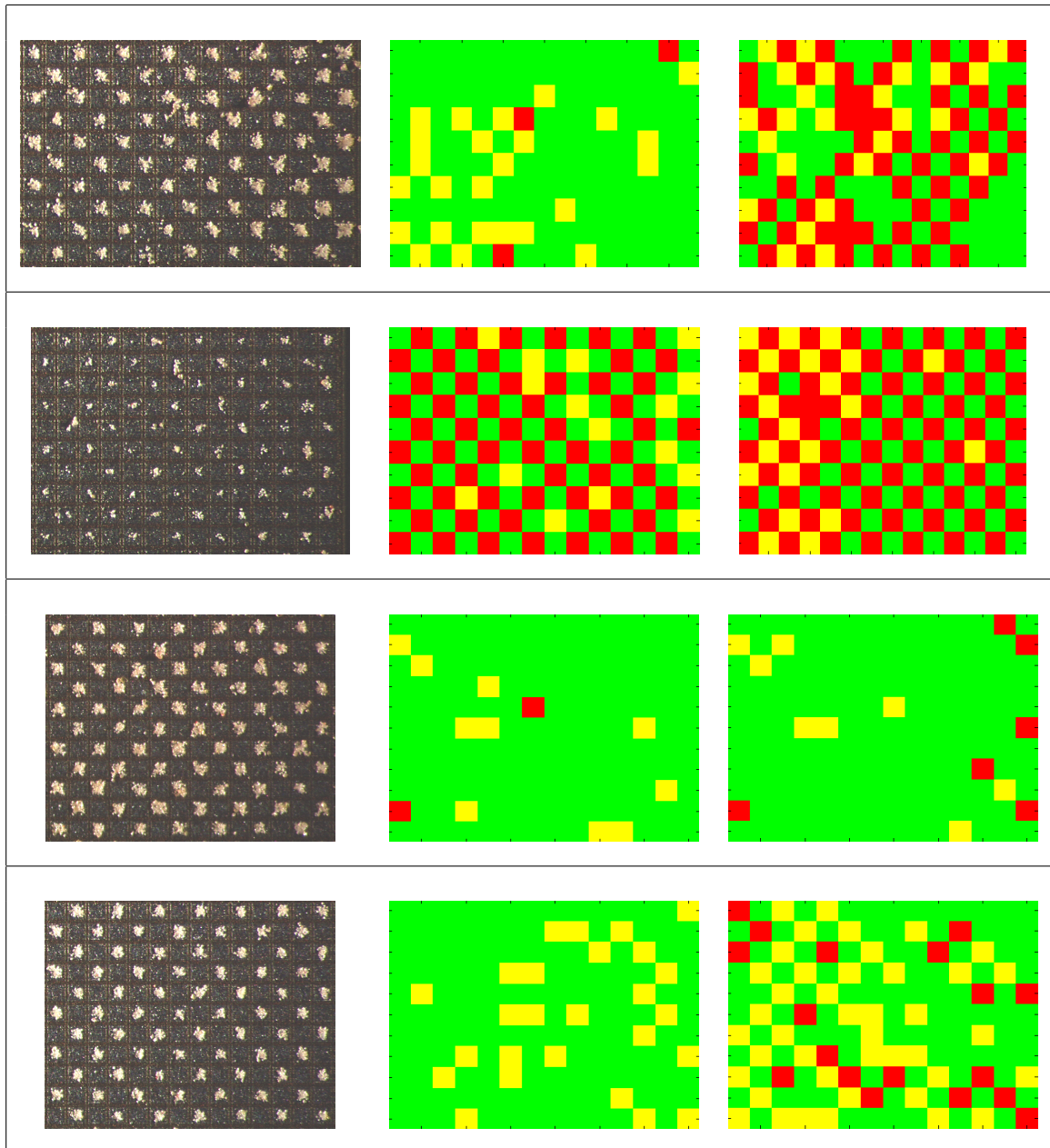


Table 5.6: Left: Analysed part of the test images (1-4 from top to bottom) Centre: Expert labelling of the chip cells Right: Algorithmic labelling of the chip cells (green indicates *good*, yellow *satisfactory* and red *bad*)

cell type	good	satisfactory	bad
covered	> 90% of inner circle cov. > 50% of outer circle cov.	> 90% of inner circle cov. > 0% of outer circle cov.	other
uncovered	< 10% of inner circle cov. < 30% of outer circle cov.	< 20% of inner circle cov. < 30% of outer circle cov.	other

Table 5.7: Quality measure similar to Table 5.6, where threshold value for the coverage of the outer circle on *good* uncovered cells is relaxed from 20% to 30%

image	cells	good	satisf.	bad	C_{abs}	C_{rel}
1	150	122	25	3	$\begin{pmatrix} 56 & 24 & 42 \\ 4 & 4 & 17 \\ 0 & 1 & 2 \end{pmatrix}$	$\begin{pmatrix} 0.46 & 0.20 & 0.34 \\ 0.16 & 0.16 & 0.68 \\ 0.00 & 0.33 & 0.67 \end{pmatrix}$
2	140	66	16	58	$\begin{pmatrix} 42 & 23 & 1 \\ 2 & 2 & 12 \\ 0 & 0 & 58 \end{pmatrix}$	$\begin{pmatrix} 0.64 & 0.35 & 0.01 \\ 0.13 & 0.13 & 0.74 \\ 0.00 & 0.00 & 1.00 \end{pmatrix}$
3	154	142	10	2	$\begin{pmatrix} 131 & 7 & 4 \\ 7 & 3 & 0 \\ 1 & 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 0.92 & 0.05 & 0.03 \\ 0.75 & 0.25 & 0.00 \\ 0.50 & 0.50 & 0.00 \end{pmatrix}$
4	154	128	26	0	$\begin{pmatrix} 87 & 33 & 8 \\ 11 & 13 & 2 \\ 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0.68 & 0.26 & 0.06 \\ 0.42 & 0.50 & 0.08 \\ 0.00 & 0.00 & 0.00 \end{pmatrix}$
total	598	458	77	63	$\begin{pmatrix} 316 & 87 & 55 \\ 24 & 22 & 31 \\ 1 & 2 & 60 \end{pmatrix}$	$\begin{pmatrix} 0.69 & 0.19 & 0.12 \\ 0.31 & 0.29 & 0.40 \\ 0.02 & 0.03 & 0.95 \end{pmatrix}$
mean	150	114	19	17	$\begin{pmatrix} 79 & 22 & 13 \\ 6 & 5 & 8 \\ 0 & 1 & 16 \end{pmatrix}$	$\begin{pmatrix} 0.67 & 0.21 & 0.11 \\ 0.36 & 0.26 & 0.38 \\ 0.12 & 0.21 & 0.67 \end{pmatrix}$

Table 5.8: Comparison of expert labelling and algorithmic quality assignments per test image based on Table 5.7, the number of chip cells classified as *good*, *satisfactory* and *bad* in the left columns are counted according to the expert labels

Other redefinitions of the quality measure do not lead to improvements in matching between expert labelling and algorithmic class assignments, either. Since the measure for the covered cells cannot be relaxed due to the experts' restrictions to maximally cover the chip cell, the quality measure as introduced in Table 5.1 is well-defined in view of a comparison with expert labelling.

The run time required to do the three processing steps (mean grey value classification, chip pattern matching, quality analysis of all chip cells) is also recorded. At

first, the dependence of the run time on the image size, shown in Figure 5.26 on the left, and on the chip cell size, shown in Figure 5.26 on the right, is measured.

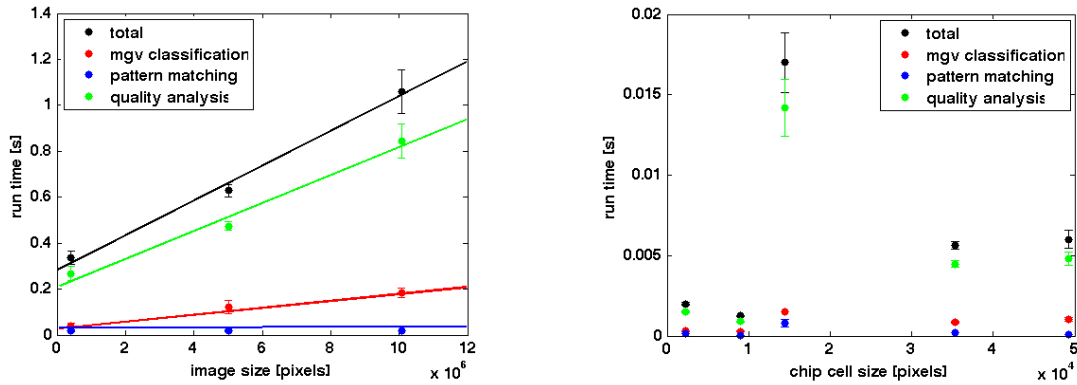


Figure 5.26: Left: Run time measurements performed on 5 images with different cell lengths and image size, mean total run time per image (black), run time for mean grey value (mgv) classification only (red), run time for chip pattern matching only (blue), run time for quality analysis on all chip cells (green), error bars are calculated out of the standard deviations for 10 runs Right: Same measurements per chip cell area for 5 runs

As expected, the run time is linearly dependent on the image size, which is demonstrated by plotting the mean value of 10 runs of the algorithm for each of the three different image sizes, namely 568×760 pixels, 1944×2580 pixels and 2592×3888 pixels. The tests furthermore show that the run times are in the range up to one second, making any tested image size suitable for the application. Comparing the run times of images with different chip cell sizes but the same image size shows that they are very similar. This is indicated by the small error bars in Figure 5.26, since the average is calculated over images with two different chip cell sizes for the first two image sizes (see Table 5.7.2).

While the almost constant time for pattern matching contributes least to the total run time, the mean grey value classification and the quality analysis per chip cell are linearly dependent on the number of pixels in the image. Additionally, it can be read off the graph that the mean grey value classification is about 5 times faster than the quality analysis, which supports its role as quick consistency check before the detailed analysis starts.

Taking the mean value for 5 runs of the algorithm for one image of each cell length summarised in Table 5.7.2 reveals that all run times are in the range of milliseconds, as shown in Figure 5.26 on the right. This result can be further improved by parallelising the calculations, as the chip cells are independent from each other, yielding a gain in the mean grey value classification and the quality analysis.

cell length [pixels]	image size [pixels]	camera	magnification
48	568 × 760	Sony DXC-950P	5x
95	1944 × 2580	Progres C5	5x
120	568 × 760	Sony DXC-950P	10x
188	1944 × 2580	Progres C5	10x
222	2592 × 3888	Canon EOS1000D	5x

Table 5.9: List of microscope and camera setups with their image size specifications tested in run time measurements

On the whole, the experiments prove that the algorithm with the experts' quality measure reliably leads to similar classification results as expert labelling and that the run time requirements are fulfilled, even without special run time tuning or parallelisation. Furthermore, the correlations of the TOF-SIMS analysis with the particle deposition yield a first ground truth apart from expert labellings which refines the quality measure and provides a new labelling against which the algorithm can be tested again.

5.8 Automation

5.8.1 Automated peptide synthesis machine

Peptide syntheses are laborious processes that require high precision and reproducibility in order to guarantee a certain quality standard. Automating all steps including the quality analysis is necessary to match these requirements and additionally to lower the costs. Therefore, a synthesis apparatus as outlined in [CLK⁺10] is currently being built. It will have the options to use the chip as a printing head for synthesis on glass slides or to do on-chip synthesis. To control the deposition on the chip in each layer, a camera and the analysis software for the quality check can be integrated in the framework as described in Figure 5.27.

The automaton will be constructed such that the chip is fixed on a block that can move freely on a rack and is brought to the stations by a motor with spatial precision of $0.1\mu\text{m}$ in each direction and angular precision of $5\mu\text{rad}$. Due to this high precision, the movement of the chip is slow, delimiting the extensions of the entire apparatus. Therefore, the microscope setup described in Section 2.3 is reduced to one stationary objective, a camera and a fixed light source. Such compact setups are commercially available as mono-zoom microscope (MZM) and can be composed individually to fulfil the requirements of Chapter 2. The entire process chain, including the camera, is then to be controlled by a LabView program.

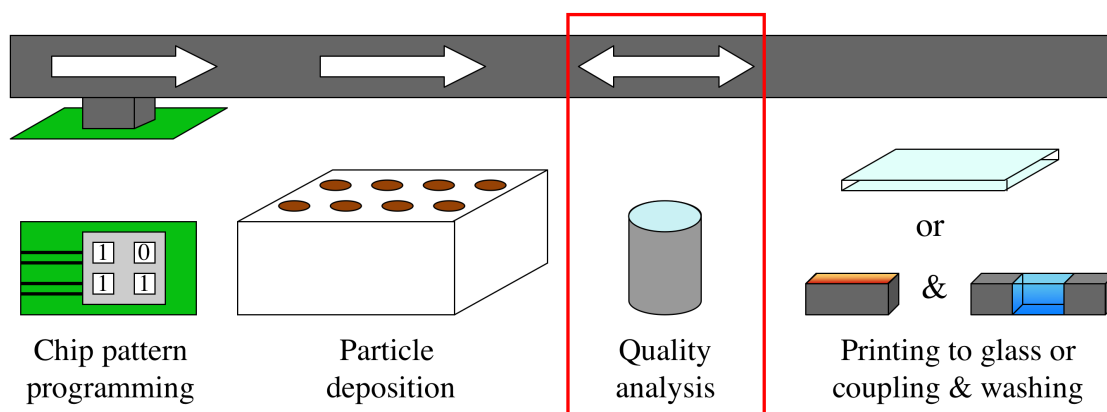


Figure 5.27: Automation process chain (from left to right): programming of the deposition pattern into the chip (shown in green) which is mounted on a block (grey block above the chip) that is guided by a rack (grey bar on top), particle deposition, quality analysis (inserted in the already existing processing chain), then printing the particles onto a glass slide or coupling the amino acids onto the chip surface and subsequent washing of the remnant particles from the chip, the white arrow on top indicate the movement direction: after quality assessment of one amino acid deposition, the next amino acid toner is deposited

5.8.2 Role of the quality analysis program

Given the overall process control of the automaton, the quality analysis should return feedback about the quality of the deposition to the main control so that malfunctions are detected as soon as they appear. Malfunctions include incorrect programming of the voltage pattern of the chip, unstable aerosol conditions that lead to bad deposition qualities or any camera related problems that restrain image acquisition from working properly.

Since each layer is completely assembled before coupling and washing, the automated version of the algorithm has to cope with subsequent depositions of different amino acids. After particle deposition, the first sort of amino acid, the quality analysis algorithm evaluates the contamination of the uncovered chip cells and the quality of coverage for those chip cells that are to be covered, as outlined in Section 5.7.1. In the subsequent deposition step with the second amino acid, the following three cases can be distinguished:

- If the cell was uncovered in the first and second deposition, the quality analysis algorithm evaluates the degree of contamination after the second deposition. So, if the two pictures are aligned, the assigned quality remains the same or decreases.

- If the cell was covered in the first deposition, it stays covered in the second one. Hence, the quality analysis algorithm should assign the same quality label for this cell as in the first step. Yet, it is possible that contaminations from the second amino acid in the previously uncovered area increase the degree of coverage, so that the quality seems to improve in the second step. Comparing the assigned qualities after the first and after the second step thus detects contaminations in uncovered areas of covered cells. Particles of the second amino acid atop the first ones are not detected by the algorithm. Yet, the amino acids contained therein are very unlikely to cause contamination because the amino acids contained in the first deposition layer will reach the surface first.
- If the cell was uncovered in the first deposition and covered in the second, the quality analysis algorithm first detects contamination and then evaluates the degree of coverage. In the case of grave contamination, the algorithm stops the process after the first deposition, so that the second layer will always be deposited on an uncovered cell with tolerable contamination. The latter is mainly caused by small particles in the corner of the chip cells that do not interfere with the synthesis in the spot centre.

Depositing the n -th amino acid in one layer, the same considerations are valid, as the algorithm saves the results from the previous depositions. From the technical point of view, the subsequent analysis of several depositions in one synthesis layer thus requires the combination of the deposition masks for all amino acids in their correct deposition order. The entire procedure is visualised by the example deposition shown in Figure 5.28. The top row shows the images taken after particle deposition, the bottom row lists the respective deposition masks that serve as input for the quality analysis in the respective deposition step. After coupling the first layer of amino acids, the remnants of the toner are removed and the chip surface at the optical resolution of the microscope is the same as at the beginning. Hence, this procedure can be applied to all subsequent layers with the same parameters.

Apart from detecting malfunctions during synthesis, the quality analysis program can also be used to calibrate the automaton at startup, regulating the parameter values of the machine such that the deposition results for a test chip are optimal for combinatorial synthesis. To find optimal values for parameters like particle injection pulse length or pressure under which particles are injected in the deposition chamber, it is possible to set up a calibration routine that subsequently deposits particles with varying parameters onto a chip and let the quality analysis program find out which of those depositions, i.e. which set of parameter values, is best.

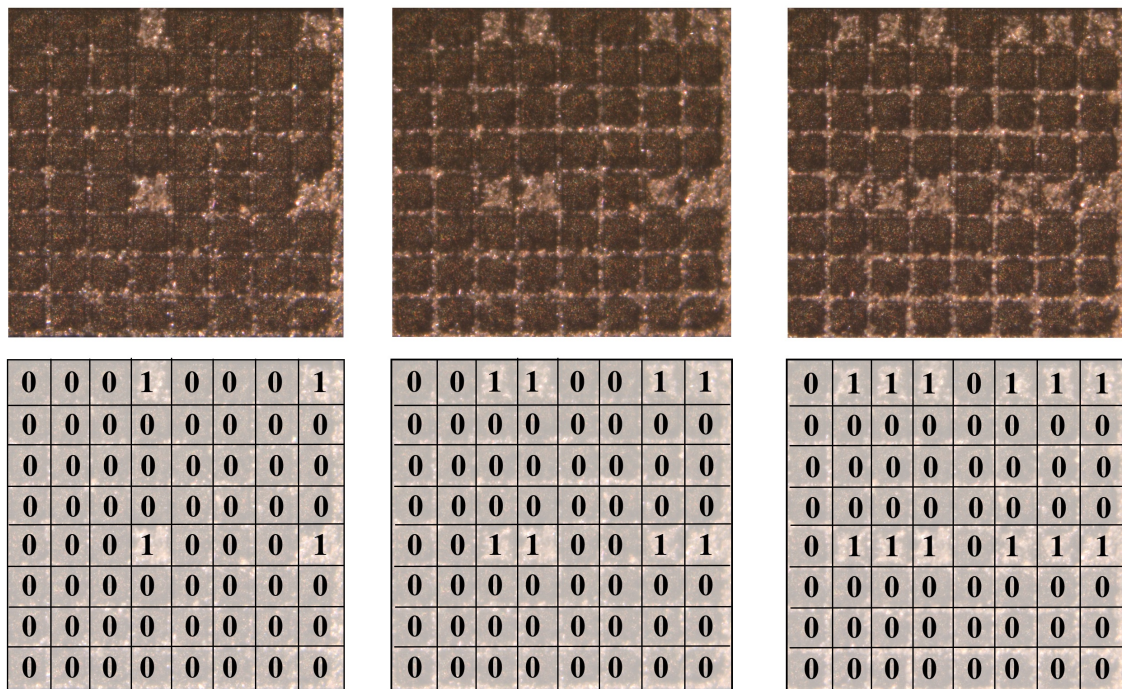


Figure 5.28: Top row: Images of sequential particle depositions without coupling steps in between Bottom row: Particle deposition masks that serve as input for the respective quality analysis

5.8.3 Interface to the quality analysis program

Since the analysis software is written in MATLAB while the machine control is programmed in LabView, an interface has to be set up for the two programs to communicate with each other. Exchanging command and response files over a shared folder is the simplest method that assures compatibility even in the case when the quality analysis routine is translated to another programming language, e.g. to a C++ program.

The LabView process control is supposed to convey the voltage programming files of the chip and the images taken by the camera to the quality analysis program, while the latter sends back responses about the quality to determine the next processing step. Furthermore, the LabView control program is supposed to store the parameter values how to position the chip above the image acquisition setup in order to take optimal pictures. Due to the high positioning accuracy, it is furthermore possible to fix the origin of the grid coordinates in the image and save them to the parameter file of the quality algorithm, so that no user input is required to divide the image into the single chip cells for quality analysis.

Synthesis closed-loop control

The state machine shown in Figure 5.29 depicts the procedure of the synthesis closed-loop control, i.e. started once, no user interaction is required during normal operation.

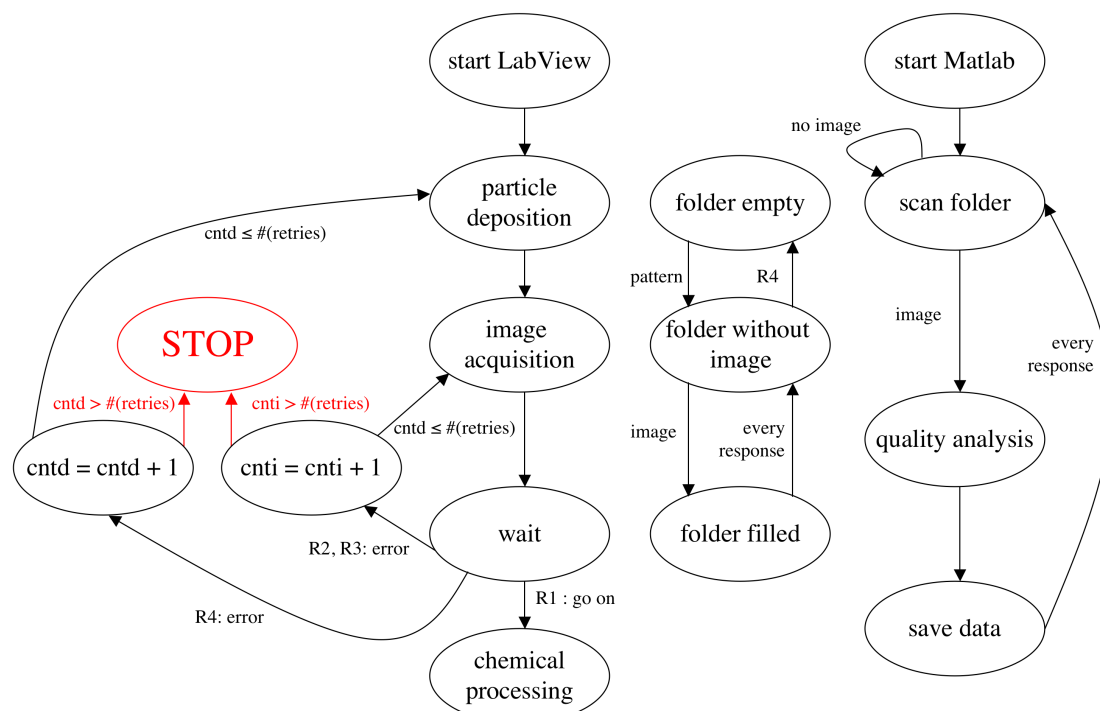


Figure 5.29: Finite state machine diagram of the automated quality analysis process: after starting LabView, particle deposition, image acquisition, quality analysis and chemical processing can be performed and the respective actions in the image folder and in the Matlab routine take place as described in the text.

First, Matlab and LabView are started. Before the automated particle deposition begins, LabView exports the chip programming deposition pattern into a folder shared by Matlab. While LabView controls the particle deposition and takes the image for quality analysis, Matlab scans that shared folder in previously defined time steps until LabView has exported the image to that folder. During the analysis done by Matlab, LabView is in a wait state until it retrieves a response about the quality of the current deposition from Matlab. Alternatively, the chip can already be brought back to the deposition station in the mean time, so that the next step

can be immediately started after receiving the response⁵. The response, together with the analysed data, is saved into another shared folder from where LabView gets one of the four possible responses:

R1 : proceed to next step

R2 : the image data is corrupted so that it cannot be processed

R3 : the image causes a Matlab intrinsic error

R4 : the image is not of expected quality (e.g. due to contamination or sparse particle coverage)

In case of R1, LabView can go on with the next step in the processing queue. For the error messages R2 and R3, the problem could be caused in the image acquisition step. Therefore, this step is repeated once again with a new picture of the same particle deposition on the chip, while the deposition pattern file is kept. Receiving R4, the process has to clean the chip and repeat the steps from particle deposition onwards. In order to assure that no programming error of the chip caused the malfunction, the deposition pattern is also created anew and sent to the shared folder. R4 could also be split into two errors, R4a and R4b, R4a dealing with sparse particle coverage and forcing an additional deposition step and R4b handling the case of contamination requiring the chip to be cleaned prior to a new deposition. If the results do not improve after a previously determined number of retries, LabView stops the processing for human intervention for error inspection. Regardless of the response, Matlab returns to scanning the shared folder waiting for a new image after termination of the analysis of the previous one. Further information about the detailed implementation of the shared folder concept and the MATLAB part of the interface can be found in Appendix B.3.

On the whole, from the quality analysis program side, automation is accomplished and ready to be integrated into the synthesis automaton, as the MATLAB part of the automated quality analysis as described in Figure 5.29 is successfully tested in a stand-alone run, since the LabView program that controls the entire machine has not yet been created and is not required to check the correct processing from image read-in into MATLAB to the output of the data analysis and the response. Only the startup calibration supported by the quality analysis algorithm cannot be implemented until the entire automaton is ready for experiments, in order to determine the parameter space to be tested for optimality.

⁵In the normal case, it is assumed that the quality will be sufficient to continue.

5.9 Test on an entire peptide synthesis

5.9.1 Particle deposition analysis

Despite the lack of automation for the quality analysis algorithm, a manual test on an entire peptide synthesis can be performed to demonstrate the usage of the algorithm. For this test, a recent combinatorial synthesis is used. The test pattern to be monitored consists of 8- and 9-meres, called FLAG and HA epitopes, and one row of permutations of them in the 6th and 7th synthesis layer, as shown in Table 5.10 for further reference.

peptide	support	1	2	3	4	5	6	7	8	9
FLAG	support	Lys	Asp	Asp	Asp	Asp	Lys	Tyr	Asp	
HA	support	Ala	Tyr	Asp	Pro	Val	Asp	Tyr	Pro	Tyr
FLAG perm.	support	Lys	Asp	Asp	Asp	Asp	*	*	Asp	
HA permu.	support	Ala	Tyr	Asp	Pro	Val	*	*	Pro	Tyr

Table 5.10: Sequences of peptides in the order of their assembly in a peptide synthesis, the * in layers 6 and 7 represents the permuted amino acids

The deposition patterns for this chip detail are similar to those shown in Figure 5.15 and can be easily deduced from the actual particle depositions due to the small amount of contaminations⁶. Figure 5.30 and Figure 5.31 summarise the particle depositions for the entire peptide synthesis, each row representing the depositions of one layer. The focus is set on a test of the FLAG and HA epitopes, so that the depositions of the single amino acids in the permutation layers are omitted here because no images were taken in these steps. Inspecting the degree of contamination more closely, it can be noticed that the two columns on the right that are not supposed to be covered acquire strong contamination, starting from layer 1 already. Hence, it is possible that peptides are assembled on these spots. Furthermore, it can be observed that the grid between the synthesis sites becomes also covered with particles, so that the corners of already covered cells become contaminated with particles of another kind, too. In how far these depositions affect the synthesis results is to be investigated in this experiment.

In the first step of the experiment, the quality analysis algorithm is calibrated. As all images in Figure 5.30 and Figure 5.31 are taken under similar conditions, performing the training step described in Section 5.6 on one of the images only is supposed to give reasonable results. Applying the training step to the deposition image of Layer 5-1 Asp, which is chosen arbitrarily, and segmenting all images with this threshold, the results shown in Appendix B.5 are obtained. Comparing them to the actual

⁶There are no uncovered cells that show a larger amount of particles and could be confused with covered cells.

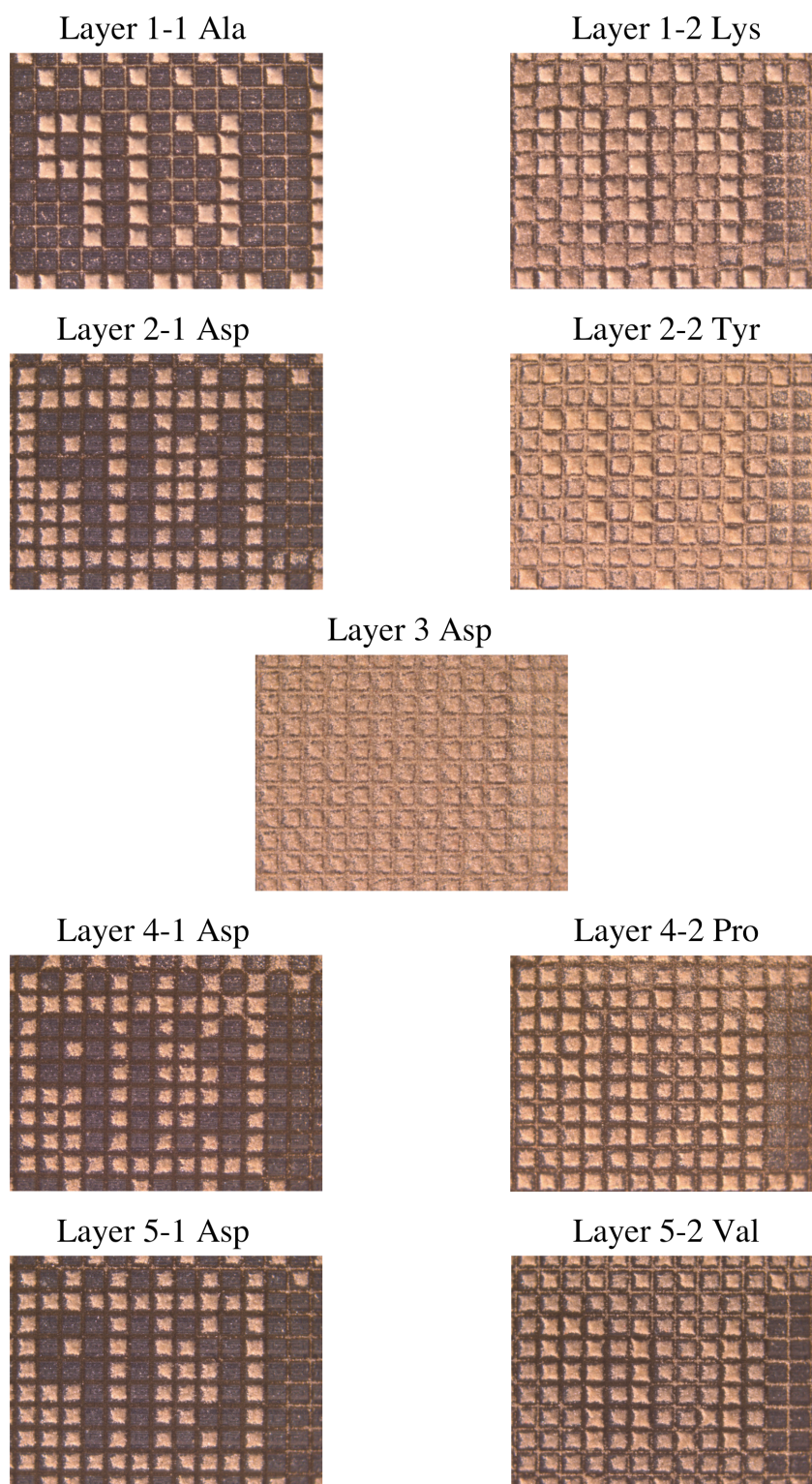


Figure 5.30: Particle depositions for each amino acid of the FLAG and HA epitopes for each layer from layer 1 to layer 5

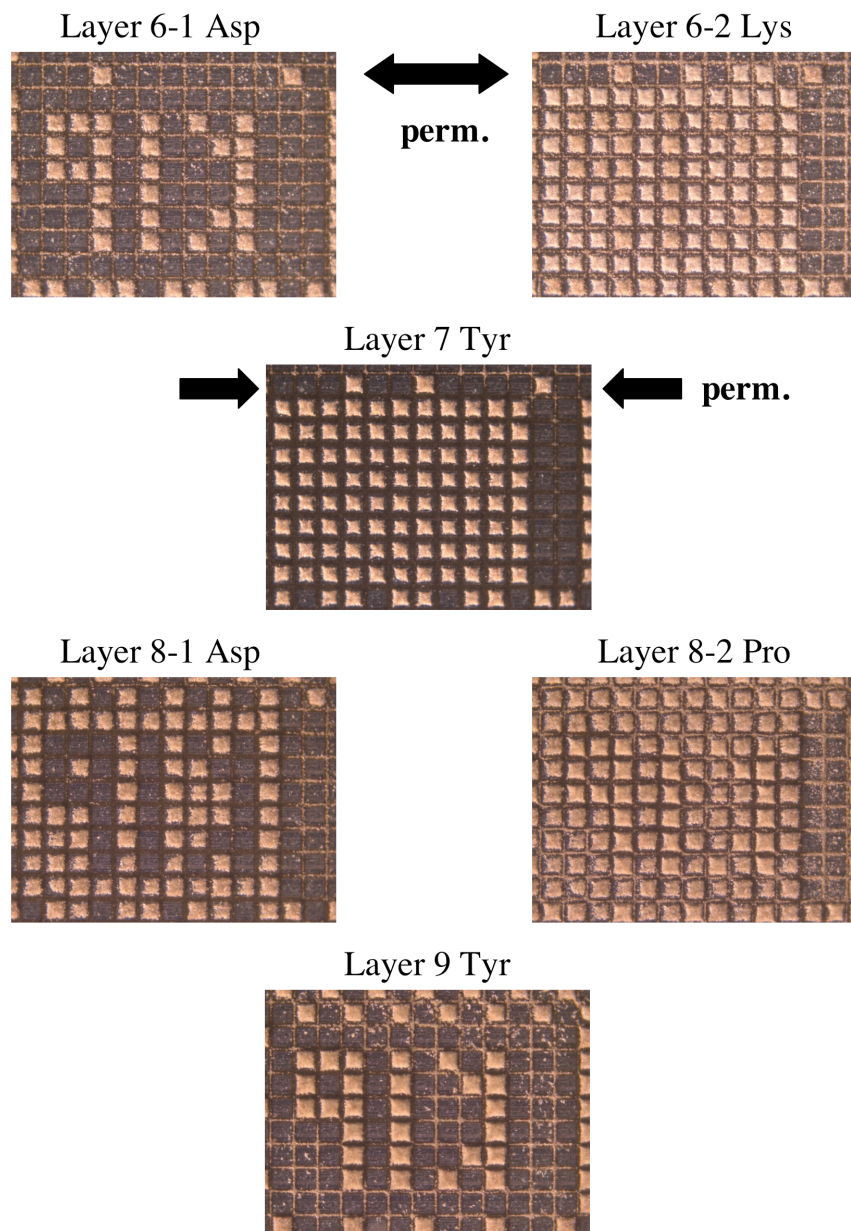


Figure 5.31: Particle depositions for each amino acid of the FLAG and HA epitopes for each layer from layer 6 to layer 8 (for FLAG) and layer 9 (for HA), the black arrows indicate the row with permutations of the FLAG and HA epitopes

deposition images shown in Figure 5.30 and Figure 5.31, the segmentation results coincide with the particle deposition estimated by eye inspection. That the segmentation results highly coincide with the actual particle deposition has been shown in Section 5.3, now, the question about the stability of segmentation over several images is investigated. In order to do so, the segmentation images for the covered cells in the first chip cell row of the first layer (L1-1 Ala and L1-2 Lys) are plotted in Figure 5.32 (top and center) and subtracted from each other, as shown in Figure 5.32 (bottom). From the latter can be deduced that the segmentation results are stable up to a translation. The differences of -1 (blue colour in Figure 5.32, bottom) at the borders of the cells in L1-2 are contaminations that arise as boundary effects when particles are deposited onto the grid of the chip. As the translation can be accounted for by automated image acquisition and contamination at the border of the synthesis sites can be neglected, the successful detection of particle deposition is hereby proven.

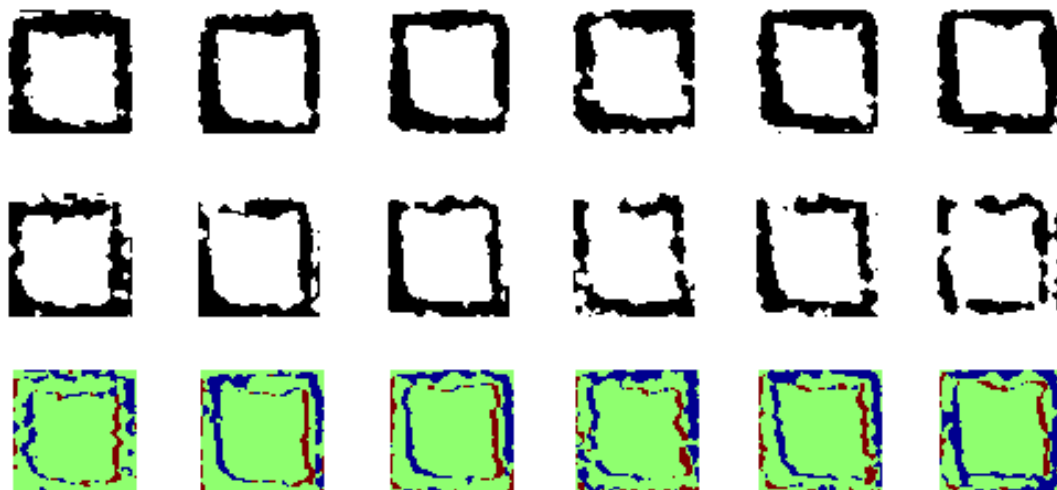


Figure 5.32: Top: Segmentation of the covered cells in the first chip cell row in L1-1 Ala Centre: Segmentation of the same chip cells in L1-2 Lys Bottom: Differences between the former and the latter, indicating that both coincide well up to a translation (blue: difference of -1, green: difference of 0, red: difference of +1)

Secondly, the quality analysis algorithm, i.e. the testing phase according to Section 5.7 is performed. Visualising the results for each chip cell in each deposition image, the 9×13 chip cells are enumerated columnwise from 1 (top left chip cell) to 117 (bottom right chip cell) so that the quality labels for each chip cell in each deposition are plotted columnwise in Figure 5.33. The numbers of assigned *good*, *satisfactory* and *bad* labels are summarised in Table 5.11.

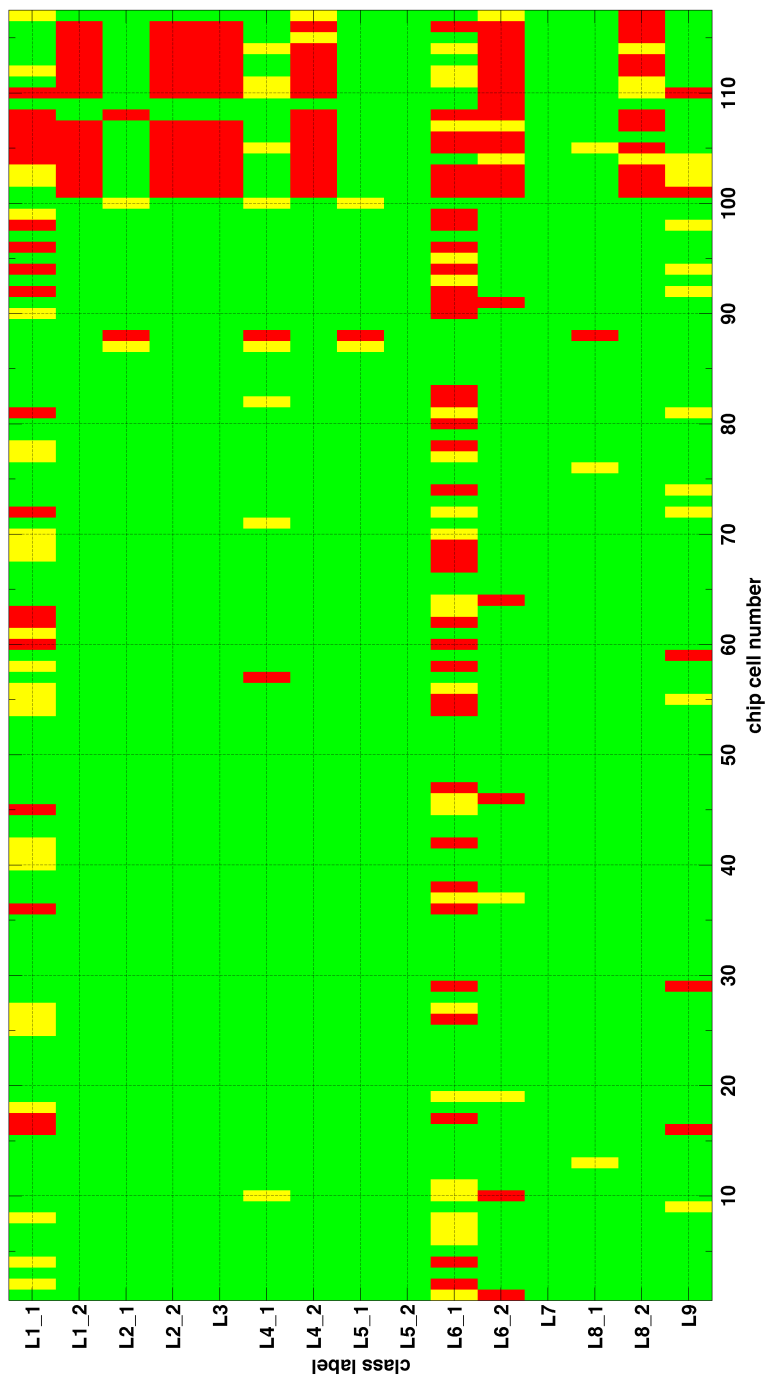


Figure 5.33: Resulting quality classes for each chip cell (enumerated columnwise in a particle deposition image) in each deposition image: green (*good*), yellow (*satisfactory*), red (*bad*)

layer	good	satisfactory	bad
L1-1	72 (61.6%)	26 (22.2%)	19 (16.2%)
L1-2	103 (88.0%)	0 (0.0%)	14 (12.0%)
L2-1	113 (96.6%)	2 (1.7%)	2 (1.7%)
L2-2	103 (88.0%)	0 (0.0%)	14 (12%)
L3	103 (88.0%)	0 (0.0%)	14 (12%)
L4-1	106 (90.6%)	9 (7.7%)	2 (1.7%)
L4-2	101 (86.3%)	2 (1.7%)	14 (12%)
L5-1	114 (97.4%)	2 (1.7%)	1 (0.9%)
L5-2	117 (100%)	0 (0.0%)	0 (0.0%)
L6-1	57 (48.7%)	24 (20.5%)	36 (30.8%)
L6-2	93 (79.5%)	5 (4.3%)	19 (16.2%)
L7	117 (100%)	0 (0%)	0 (0%)
L8-1	113 (96.6%)	3 (2.5%)	1 (0.9%)
L8-2	102 (87.1%)	4 (3.4%)	11 (9.4%)
L9	101 (86.3%)	11 (9.4%)	5 (4.3%)
total	1515 (86.3%)	88 (5.0%)	152 (8.7%)

Table 5.11: Overview of assigned class labels in absolute numbers and in percent relative to the total amount of analysed chip cells per layer

As can be observed from Figure 5.33, the overall quality of the synthesis is *good*, with 86.3% of *good* labels (green) assigned to the analysed chip cells. However, layer L1-1 and layer L6-1 show an increased amount of *satisfactory* (yellow) and *bad* (red) labels, so that a repetition of these steps would have been advisable. Hence, the quality analysis algorithm can be empirically trained to set thresholds for the repetition of an entire step, when the percentage of *good* labels falls below 80% or the percentage of *bad* labels exceeds 20%. Furthermore, assigning more than 80% of all chip cells to the class *satisfactory* should also cause a warning because, in this case, the algorithm is not capable of finding a clear classification.

The analysis of the change of labels from the first deposition to the second deposition in one layer for all layers with two particle depositions (L1, L2, L4, L5, L6, L8) yields the results in Table 5.12. From all possible changes of labels, the one from *good* to *good* is the most frequent, which implies a high quality of the particle deposition in the analysed synthesis. Hence, not only requiring that the number of *good* labels should exceed more than 80% of all chip cells for each deposition, the quality analysis algorithm can also check the deposition quality of an entire layer by determining the ratio of *good* chip cells after the first deposition that remain *good* ones in the second. The practical experience of the analysed synthesis favours a threshold of 80% *good* \rightarrow *good* label assignments, below which the a repetition of the particle deposition for the entire layer seems necessary.

1st → 2nd	L1	L2	L4	L5	L6	L8
<i>good</i> → <i>good</i>	66 (56.4%)	99 (84.6%)	94 (80.3%)	114 (97.4%)	51 (43.6%)	99 (84.6%)
<i>good</i> → <i>satisf.</i>	0 (0.0%)	0 (0.0%)	2 (1.7%)	0 (0.0%)	2 (1.7%)	4 (3.4%)
<i>good</i> → <i>bad</i>	6 (5.1%)	14 (12%)	10 (8.6%)	0 (0.0%)	4 (3.4%)	10 (8.6%)
<i>satisf.</i> → <i>good</i>	23 (19.7%)	2 (1.7%)	5 (4.3%)	2 (1.7%)	14 (12%)	2 (1.7%)
<i>satisf.</i> → <i>satisf.</i>	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	3 (2.5%)	0 (0.0%)
<i>satisf.</i> → <i>bad</i>	3 (2.5%)	0 (0.0%)	4 (3.4%)	0 (0.0%)	7 (6.0%)	1 (0.9%)
<i>bad</i> → <i>good</i>	14 (12%)	2 (1.7%)	2 (1.7%)	1 (0.9%)	28 (23.9%)	1 (0.9%)
<i>bad</i> → <i>satisf.</i>	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
<i>bad</i> → <i>bad</i>	5 (4.3%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	8 (6.8%)	0 (0.0%)

Table 5.12: Overview of the change of class labels from the first to the second deposition in one entire layer for all layers with two depositions, i.e. L1, L2, L4, L5, L6, L8

5.9.2 Fluorescence coupling analysis

From the quality analysis results obtained in Section 5.9.1, a good quality of the entire peptide synthesis is expected. In order to evaluate the quality of the entire synthesis, a fluorescence detection reaction called sandwich assay, as described in [Blo09] is performed: Having coupled the last layer of amino acids to the growing peptide, the protections of the N-termini and the side chains are removed before the peptide array is exposed to a mixture of mouse-anti-HA antibodies and rabbit-anti-FLAG antibodies. Then, a mixture of anti-mouse-AlexaFluor546 and anti-rabbit-AlexaFluor647 is poured over the array to couple the antibodies with fluorescence markers on top of the first antibodies. Under a laser scanner with lasers of the respective wavelengths in the red and green spectrum, the HA synthesis sites shine green, the FLAG sites red, the permutations may be green, red or without a fluorescence signal, depending on their sequence. As shown in Figure 5.34, the synthesis of Figure 5.30 and Figure 5.31 yields clear FLAG and HA signals at the expected synthesis sites, which was correctly predicted by the quality analysis algorithm.

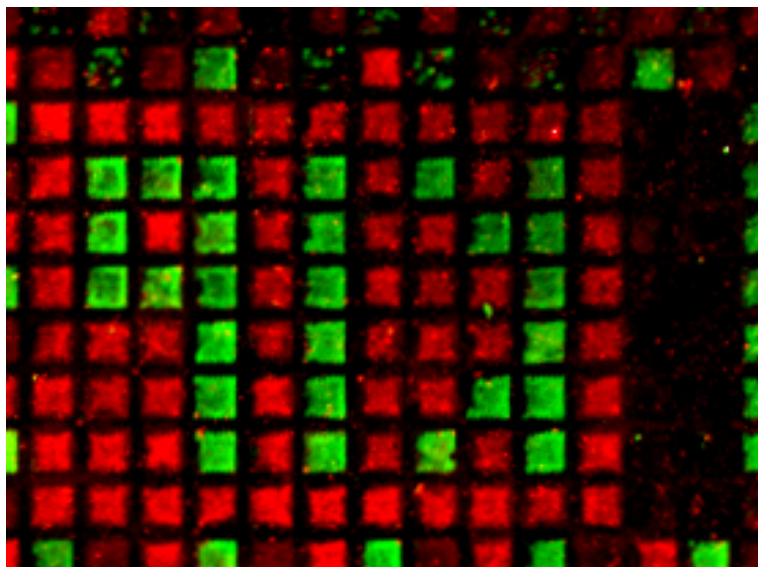


Figure 5.34: Fluorescence detection reaction of FLAG (red) and HA epitopes (green) subsequent to a complete combinatorial particle based peptide synthesis, the weaker signals in the top row of synthesis sites originates from FLAG and HA permutations

Apart from the prediction whether or not fluorescence signals can be expected in the first place, a semi-quantitative predication about the density of correctly assembled peptides is desired. Having successfully accomplished mass spectrometry analyses of assembled peptides by matrix-assisted laser deposition/ ionisation (MALDI), as performed in [Bey05] and [Blo09], a MALDI time-of-flight (spatially resolved) surface imaging, e.g. provided by VOM in Münster, can give more information about the density of correctly assembled peptides. A more cost-effective method that could provide the same information is confocal laser microscopy. However, in order to be able to correlate the density of correctly assembled peptides with the particle depositions in each layer, an automated alignment of the deposition images is required.

What is already possible is an analysis of the mean fluorescence intensities in the red and green intensity channels by means of a modified training step of Section 5.6: The grid is detected as described in Section 5.6. The resulting cell length is 20 pixels, hence, the resolution is $5\mu\text{m}$ per pixel for the scanner used. Subsequently, the mean grey value feature for the detected chip cells is calculated as described in Section 5.4.2, resulting in the mean grey values for the 9×13 chip cells shown in Figure 5.35 left for the green channel and right for the red channel. Hence, gauging the mean intensity values of all chip cells on the array to one cell previously defined as the reference cell (a *standard candle* with known luminous intensity), the strength of antibody couplings to the different peptides can also be determined by means of the quality analysis algorithm in a slightly modified form.

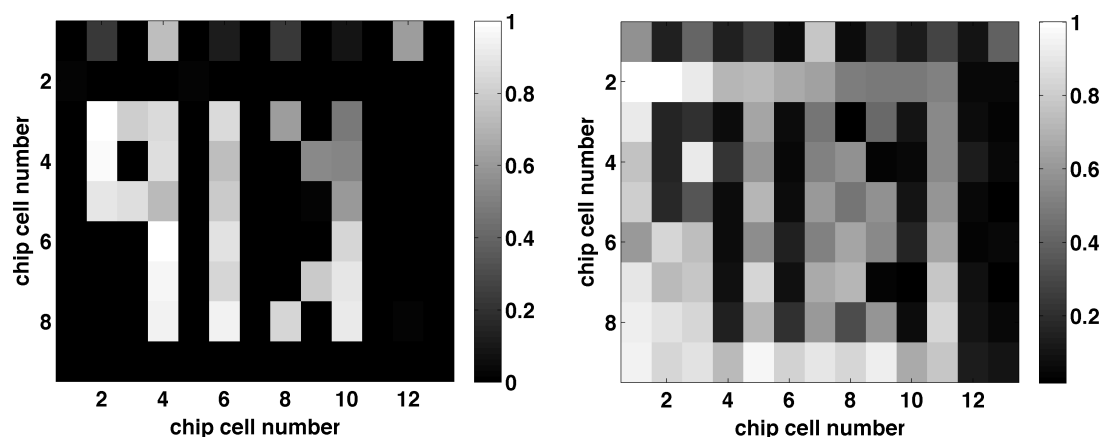


Figure 5.35: Left: Mean fluorescence intensities for all chip cells in the green channel of Figure 5.35 Right: Mean fluorescence intensities in the red channel of Figure 5.35

Although the same peptides should lead to the same intensities in the image, there are gradients in Figure 5.34 that originate from insufficiencies in the chemical surface activation (i.e. the density of coupling sites available), in the deposition and in the coupling of the fluorescent anti-bodies. The most probable cause for the gradients in this array seems to be the quality of the chemical surface activation because the overall intensity of the fluorescence signal is high implying a good deposition and coupling result, but in some regions, the signal shows sharp edges, indicating that the surface activation could have been ripped, an assumption that can also be supported by the TOF-SIMS results obtained in Section 4.3.1.

The effect visible in Figure 5.34 is that the fluorescence intensities per synthesis site are decreasing from left to right. As already discussed in Chapter 4, the effect originates from the density of coupling sites and their accessibility for the amino acids. As a consequence, the strong contamination of the two rightmost columns of synthesis sites that are to be left uncovered in the entire synthesis process does not lead to fluorescence signals of incorrectly assembled peptides, as it would have been expected.

Thus, homogeneity in the surface quality is still an open issue. In [Kön10] compatibility experiments of the chemical surface activation and different kinds of chip surfaces were performed and revealed that adaptations on both sides are necessary. As the chemical surface activation is not planned to be part of the synthesis automation, it has not to be integrated in the quality analysis processing chain developed in this section. Therefore, composition and quality control of the chemical surface activation still under development is investigated by the chemists [Sch10].

Chapter 6

Discussion & outlook

6.1 Discussion

In this work, quality control for the production of peptide chip arrays is established. For each of the peptide synthesis steps, a detailed analysis of the status quo is performed, improvements are tested and, when necessary and possible, the step is optimised. Additionally, an image acquisition setup is assembled that enables automated quality analysis of each layer of amino acids in the synthesis process. An algorithm which evaluates the quality of a representative number of synthesis spots is then created according to the results found when investigating the status quo. Proof-of-principle experiments comparing the assessment of the algorithm with those of experts and bio-chemical measurements reveal that the algorithm is capable of monitoring the synthesis process, detecting contamination and insufficiencies, which leads to a robust and reliable quality control.

Particles

The particles containing the amino acids are suitable for the synthesis process established in this group. This is proved by analysing the form, size and structure of an arbitrarily chosen charge in the electron microscope. These images also show that the optical measurements with a Malvern Mastersizer yield reasonable results. The model for the aerosol based on the numerical solution of the Navier-Stokes-Equation established in this work furthermore leads to systematic investigations of the particle parameters, i.e. particle size and the q/m -value. The optimum parameters that are theoretically determined (mean particle diameter: $2\mu\text{m}$, q/m -value: -3mC/kg) have to be experimentally realised, which means that the width of the particle size distribution has to be narrowed and that tribo-electrically charging the particles has to be improved to narrow the width of the q/m -value distribution as well. Furthermore, the amount of amino acids in the particles is found to be sufficient when analysing the signals from coupled amino acids of the TOF-SIMS measurements.

Particle transfer

The theoretical modelling revealed that the direct transport of negatively charged micro-particles in stationary air requires long deposition times, as the velocity of a particle was estimated to be about $70\mu\text{m/s}$ in an electrical field of $E = 100\text{V}/100\mu\text{m}$. Therefore, the particle deposition out of an aerosol is necessary to keep the electric fields as low as possible (using voltages of 1kV) and shorten the deposition times. So far, this method is the only one leading to precise particle placement (see [Kön10] for a detailed discussion of experimentally investigated alternatives). The theoretical model based on the Navier-Stokes-Equation developed in Chapter 3 and its numerical solution contribute to optimise the aerosol parameters, i.e. voltage values of the electrodes of the synthesis spots and the grid as well as particle velocities. But still, experimental instabilities of the setup, especially related to particle density within the aerosol, particle charge, or influences of the laboratory environment can occur. Hence, an inspection of the chip after particle deposition cannot be replaced by simulation supported parameter optimisation of the particle transfer onto the chip.

Chemical coupling and washing

Since chemical coupling can only be investigated after washing, these two steps have to be treated as one, forming the hardest part of quality control because it is very difficult to find appropriate detection reactions and methods to investigate the chemically activated surface on the chip, namely the number and accessibility of the coupling sites and the coupling rates of the amino acids out of the melted toner matrix. However, in cooperation with the *Verbundzentrum für Oberflächenanalyse Münster*, it is shown that TOF-SIMS enables to spatially resolve the density of coupled amino acids down to $4\mu\text{m}$. These proof-of-principle measurements show that the coupling conditions (accessibility of the coupling sites, number of amino acids per particle, toner matrix, particle coverage, melting temperature and exposure time) are well-tuned to obtain a good coupling rate. It also yields information about the correlation of deposited particles and coupled amino acids. By means of this, a quality measure can be set up that allows for a non-destructive evaluation of the spots during a routine peptide synthesis. Furthermore, the TOF-SIMS measurements reveal impurities in the washing process, indicating that particle remnants should be discarded by washing the chip in such a position that they can flow off the chip immediately in order not to be spread over the entire chip before discarding. This will be realised in the washing automaton currently being built in the group.

Yet, TOF-SIMS measurements at a higher resolution down to $0.5\mu\text{m}$ are possible and should be performed in order to improve the quality evaluation, increase the statistics, as the current measurements are only performed on 16 chip cells. They should be performed for all possible combinations of pairs of amino acids in a first approach, as the coupling rate also depends on the stereometry of the molecules to be coupled together and their environment. Additionally, the washing process

and the coupling efficiency in dependence on the chemical surface modification can be tested. Since the measurements are expensive (about 2000 Euros for an analysis similar to the ones performed for this work), further investigations should be planned in the course of a dedicated project with respective funding.

Image acquisition setup

Assembling the image acquisition setup, the particle characteristics are the most decisive parameter. Adapted to the particles with mean diameter of $10\mu\text{m}$, it is reasonable to have a detection limit of particles with $1.4\mu\text{m}$ diameter. Yet, as the simulations for the particle transfer reveal, finer particles show less contaminations and better coverage. Hence, the detection limit should be adapted according to this finer distribution of particle sizes. However, the simulations also reveal that smaller particles with moderate q/m -value (up to -30mC/kg) do not cause contamination. Since, in normal operation mode, particles with higher charge are very unlikely to be created, the detection limit of $1.4\mu\text{m}$ is still reasonable to keep, as better resolutions come at the cost of disproportionately more expensive equipment. The optimum setup to be found is then given by a microscope with external illumination at a small angle with respect to the chip surface and with an objective with $N_A = 0.25$ and a 10-fold magnification, combined with a TV-adaptation of 0.63 above which a Progres C5 is mounted, so that 108 chip cells can be photographed simultaneously. This results in 188 pixels per chip cell of $100\mu\text{m}$ length, so that the optical resolution is the limiting factor and the supersampling of the picture acquisition prevents imaging artefacts.

Quality analysis algorithm

The core of this work is the quality analysis algorithm that is developed on the basis of the experimental results described above. It is implemented as a module system consisting of an image enhancement part (noise reduction, flat-field correction), an image segmentation part (thresholding or a problem-adapted histogram segmentation called efficient clustering earth mover's distance), a quick consistency check (by extracting the mean grey value per chip cell as the only feature and comparing the covered and uncovered cells with the deposition pattern programmed into the chip) and a detailed quality analysis per chip cell (which is based on the features given by experts' requirements and bio-chemical analyses). The single components are then selected depending on the configuration of the setup. In order to determine all parameters required for this algorithm, a training, i.e. calibration, routine analysing one representative image only is also implemented and can be performed whenever the setup is altered. After each particle deposition, an image of several chip cells is taken and analysed. The algorithm then applies the quality measure defined by experts and the bio-chemical measurements and decides in less than 1s if the quality is sufficient to achieve high coupling rates or whether refinement is necessary. The experimental evidence of the functional capability of the algorithmic processing chain is performed on a complete peptide synthesis.

For automated quality analysis in the processing chain of the peptide synthesis machine to be built, integration into the LabView control framework is achieved by file exchange over designated folders. This allows for flexibility in case the MATLAB quality analysis program is translated into another programming language or the LabView framework will not be implemented as currently planned.

6.2 Outlook

From the physical side, quality analysis is comprehensively investigated, including particle characterisation by electron microscopy and diffraction measurements, performing parameter optimisation of the aerosol dynamics for particle transfer by modelling and comparing the results with experiments, performing surface analysis scattering experiments in order to obtain an estimate for the density of coupled amino acids at sub-spot resolution and implementing an automated image processing quality evaluation which is based on all these results. As soon as the automaton is completed, automated image acquisition will lead to aligned images. The deposition images can then be correlated to estimate the number of correctly assembled peptides.

It is also possible to extend the algorithm to evaluate the fluorescence images from the antigen-antibody detection reactions as outlined in Section 5.9.2. For this task, the single colour channels of the image can be evaluated after establishing a quality measure that bases on the intensity signal of the synthesis spots. Exchanging the classification component in the existing system by this one is simple due to the modular structure. The rest of the algorithm can be reused, since the chip cell structure will also remain the same. Knowing the (integrated) intensities per synthesis spot also enables biological data evaluation like finding the peptide that binds best to a certain test molecule or comparing different peptides with respect to their coupling probability.

Another direction to change the algorithm is changing the array configuration. Currently, the cells to be analysed are squares, but it is also possible to implement an analysis for rectangular or circular spots. The latter could be of high interest to apply the algorithm to peptide syntheses out of liquids or integrate it in the peptide synthesis currently being investigated in [Mär10].

The next step in quality control for the current setup should be bio-chemical quality analysis, setting up quality standards for the surface modifications, the washing procedure and in-situ purification to increase the density of correctly assembled peptides. Figure 6.1 gives an overview of all quality analysis techniques available, the quality control that is currently implemented and the tasks still required to complete the monitoring of particle based peptide synthesis production.

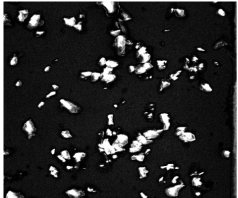
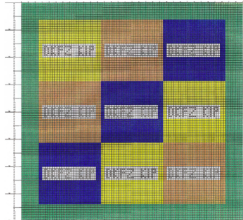
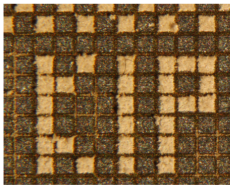

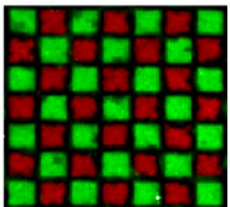
<p>Particle creation</p> 	<p>Quality analysis</p> <p>HPLC (amino acids) Mastersizer (size) <i>EM microscopy (form)</i></p>	<p>Quality control</p> <p>Routine checks guarantee reproducible experimental conditions</p> <p><i>(done)</i></p>
<p>Chip programming</p> 	<p>(I²C read back) <i>Light microscopy</i></p> <p><i>(correct pattern)</i></p>	<p>Correct pattern, particle coverage and contamination can be monitored by using light microscopy and image processing. The algorithm that decides to proceed or improve the deposition is based on the particle deposition simulations and the chemical analysis, correlating particle deposition with the density of coupled amino acids</p> <p><i>(done)</i></p>
<p>Particle deposition</p> 	<p><i>Simulations</i> <i>Light microscopy</i></p> <p><i>(particle coverage, contamination)</i></p>	<p>TOF-SIMS results lead to optimisation in chemical processing</p> <p><i>(partly implementing)</i></p>
<p>Chemical processing</p> 	<p><i>Bromophenol blue</i> <i>Ninhydrin</i> <i>TOF-SIMS</i></p> <p><i>(coupled amino acids)</i></p>	<p>TOF-SIMS results lead to optimisation in chemical processing</p> <p><i>(partly implementing)</i></p>
<p>Peptide array</p> 	<p>Antibody coupling MALDI</p> <p>(correctly assembled peptides)</p>	<p>Reusing the algorithm for particle deposition, the number of correctly assembled peptides could be retrieved from the fluorescence image</p> <p><i>(partly implemented)</i></p>

Figure 6.1: Quality analysis and control for each step of particle based peptide synthesis, italic underlined are the techniques treated in this thesis (I²C is written in brackets as the method is currently not used anymore, see [Kön10] for details)

Acknowledgements

Bernd Jähne for accepting to supervise this thesis, his support, advice and help.

Michael Hausmann for all the things I learnt from him when working on the project, from biology over medicine and microscopes to project funding and university politics and for his scientific and bureaucratic efforts to help me successfully complete this work.

Volker Lindenstruth for all the years of strong support, motivation and encouragement, being the backbone of my thesis and always providing me with the freedom I needed and the challenges I sought.

Kai König and **Simon Fernandez** for their never-ending patience teaching me the important facts and discussing the results and for their good cooperation performing the experiments.

Felix Löffler, **Alexander Nesterov** and **Tobias Förtsch** for their good collaboration in all aerosol and deposition related questions.

Verbundzentrum für Oberflächenanalyse Münster, especially **Heinz Hinkers** for analysing the chip and their support concerning all related issues.

Björn Ommer for supervising the image segmentation part, for his help when writing the ACCV paper and for the knowledge and ideas that he conveyed to me.

Florian Painke and **Nedim Boztepe** for their good collaboration during quality analysis of the chemical washing.

Yun-Chien Cheng for his advice when defining the interfaces to the synthesis automaton.

Eberhard Schmitt for broadening my horizon by giving me the opportunity to evaluate medical data and for his successful efforts to make the best out of every working day in the office.

Chip-based peptide libraries group (DKFZ), its present and former members, for their commitment to provide me with everything I needed as quick as possible.

Fish & Chips group (KIP), for their suggestions and help, especially **Gloria Torralba** for providing me with her work on the topic, which helped me to find an ansatz to start with and **Nick Kepper** for proofreading.

Frankfurt Institute for Advanced Studies and the **HGS MathComp** for funding.

My friends and family, **Sacher Khoudari** for helping me with any Latex related problems and **Rüdiger Vaas** for teaching me how to play Go sometimes meant teaching me how to go on in my research: always look for several options!

Bibliography

- [All97] T. Allen, *Particle size measurement: Powder sampling and particle size measurement*, Springer, 1997.
- [Alt02] H. W. Alt, *Lineare Funktionalanalysis: Eine Anwendungsorientierte Einführung*, Springer, 2002.
- [ARB08] S. Aoyagi, A. Rouleau, and W. Boireau, *TOF-SIMS structural characterization of self-assembly monolayer of cytochrome b5 onto gold substrate*, *Applied Surface Science* **255** (2008), 1071–1074.
- [Bey05] M. Beyer, *Entwicklung und Anwendung neuartiger Trägeroberflächen zur kombinatorischen Peptidsynthese mit Aminosäure-Tonerpartikeln*, Ph.D. thesis, Heidelberg University, 2005.
- [BFB⁺06] M. Beyer, T. Felgenhauer, F. R. Bischoff, F. Breitling, and V. Stadler, *A novel glass-slide based peptide array support with high functionality resisting non-specific protein adsorption*, *Biomaterials* **27** (2006), 3505–3514.
- [Bis06] C. M. Bishop, *Pattern recognition and machine learning (information science and statistics)*, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.
- [Blo09] I. Block, *Herstellung und Anwendung von hochkomplexen Peptidbibliotheken*, Ph.D. thesis, Heidelberg University, 2009.
- [BNB⁺07] M. Beyer, A. Nesterov, I. Block, K. König, T. Felgenhauer, S. Fernandez, K. Leibe, G. Torralba, M. Hausmann, U. Trunk, V. Lindenstruth, F. R. Bischoff, V. Stadler, and F. Breitling, *Combinatorial synthesis of peptide arrays onto a microchip*, *Science* **381** (2007), no. 5858, 1888.
- [BV04] S. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge University Press, New York, NY, USA, 2004.
- [CCN07] V. Caselles, A. Chambolle, and M. Novaga, *The discontinuity set of solutions of the tv denoising problem and some extensions*, *Multiscale Modeling and Simulation* **6** (2007), no. 3, 879–894.

- [CLK⁺10] Y. C. Cheng, F. Löffler, K. König, A. Nesterov, E. Dörsam, and F. Breitling, *Chip printer*, Proceedings of the 2nd WSEAS international conference on Nanotechnology, World Scientific and Engineering Academy and Society (WSEAS), 2010, pp. 19–22.
- [COM] COMSOL AB, *Comsol multiphysics user's guide*, 2008 ed.
- [CW00] W. C. Chan and P. D. White (eds.), *Fmoc solid phase peptide synthesis - a practical approach*, 1 ed., Oxford University Press, 2000.
- [DT05] N. Dalal and B. Triggs, *Histograms of oriented gradients for human detection*, In CVPR, 2005, pp. 886–893.
- [EAG09] Evans Analytical Group, *Analytical resolution versus detection limit*, Website, 2009, http://www.eaglabs.com/techniques/analytical_techniques/.
- [FKS02] H. Fudouzi, M. Kobayashi, and N. Shinya, *Site-controlled deposition of microsized particles using an electrostatic assembly*, *Advanced Materials* **14** (2002), no. 22, 1649–1652.
- [FMG89] P. R. Fisher, R. Merkl, and G. Gerisch, *Quantitative analysis of cell motility and chemotaxis in dictyostelium discoideum by using an image processing system and a novel chemotaxis chamber providing stationary chemical gradients*, *The Journal of Cell Biology (JCB)* **108** (1989), no. 3, 973–984.
- [Fod02] I. Fodor, *A survey of dimension reduction techniques*, Tech. report, Lawrence Livermore National Laboratory, University of California, USA, 2002.
- [FRP⁺91] S. P. Fodor, J. L. Read, M. C. Pirrung, L. Stryer, A. T. Lu, and D. Solas, *Light-directed, spatially addressable parallel chemical synthesis*, *Science* **251** (1991), no. 4995, 767–773.
- [GR94] D. J. Graham and B. D. Ratner, *Multivariate analysis of TOF-SIMS spectra from dodecanethiol SAM assembly on gold: Spectral interpretation and TOF-SIMS fragmentation processes*, *Science* **264** (1994), no. 5157, 399–402.
- [Gre02] B. P. Greimann, *Two-phase flow analysis of sediment velocity*, *Sedimentation and sediment transport: proceedings of the symposium held in Monte Verità*, 2002, pp. 83–86.
- [GWPH93] A. I. Gusev, W. R. Wilkinson, A. Proctor, and D. M. Hercules, *Quantitative analysis of peptides by matrix-assisted laser desorption/ionization*

- time-of-flight mass spectrometry*, Appl. Spectrosc. **47** (1993), no. 8, 1091–1092.
- [Hag00] B. Hagenhoff, *High resolution surface analysis by TOF-SIMS*, Microchimica Acta **132** (2000), no. 2–4, 259–271.
- [HS92] R. M. Haralick and L. G. Shapiro, *Computer and robot vision*, vol. I, Addison-Wesley, 1992.
- [KBN⁺10] K. König, I. Block, A. Nesterov, G. Torralba, S. Fernandez, T. Felgenhauer, K. Leibe, C. Schirwitz, F. Löffler, F. Painke, J. Wagner, U. Trunk, M. Hausmann, F.R. Bischoff, F. Breitling, V. Stadler, and V. Lindenstruth, *Programmable high-voltage cmos chips for particle-based high-density combinatorial peptide synthesis*, Sens. Actuators B **147** (2010), no. 418, 418–427.
- [KCBC70] E. Kaiser, R. L. Colescott, C. D. Bossinger, and P. I. Cook, *Color test for detection of free terminal amino groups in the solid-phase synthesis of peptides*, Analytical Biochemistry **34** (1970), no. 2, 595–598.
- [Kle03] C. Kleinstreuer, *Two-phase flow theory and applications*, Taylor & Francis, 2003.
- [Kön10] K. König, *CMOS-based peptide arrays*, Ph.D. thesis, Heidelberg University, 2010.
- [KR87] A. R. Khan and J. F. Richardson, *The resistance to motion of a solid sphere in a fluid*, Chemical Engineering Communications **62** (1987), no. 1, 135–150.
- [LHW06] W. P. Lehrach, D. Husmeier, and C. Williams, *A regularized discriminative model for the prediction of protein–peptide interactions*, Bioinformatics **22** (2006), no. 5, 532–540.
- [LJLV⁺00] I. Luzinov, D. Julthongpiput, A. Liebmann-Vinson, T. Cregger, M. D. Foster, and V. V. Tsukruk, *Epoxy-terminated self-assembled monolayers: molecular glues for polymer layers*, Langmuir **16** (2000), 504–516.
- [Löf09] F. Löffler, *Physikalische Charakterisierung auf CMOS-Chips von Biopartikeln und deren Ablagerungsverhalten*, Master’s thesis, Heidelberg University, 2009.
- [Low04] D. G. Lowe, *Distinctive image features from scale-invariant keypoints*, International Journal of Computer Vision **60** (2004), 91–110.
- [LW00] D. J. Lockhart and E. A. Winzeler, *Genomics, gene expression and DNA arrays*, Nature **405** (2000), 827–836.

- [LWK⁺10] F. Löffler, J. Wagner, K. König, F. Märkle, S. Fernandez, C. Schirwitz, G. Torralba, M. Hausmann, V. Lindenstruth, F. R. Bischoff, F. Breitling, and A. Nesterov, *High-precision combinatorial deposition of micro particle patterns on a microelectronic chip*, *Aerosol Science and Technology* **45** (2010), no. 1, 65–74.
- [LYW⁺09] H. Lee, S. You, C. G. W. Woo, K. Lim, K. Jun, and M. Choi, *Focused patterning of nanoparticles by controlling electric field induced particle motion*, *Applied Physics Letters* **94** (2009), no. 5, 053104–1–053104–3.
- [Mar70] F. E. Marble, *Dynamics of dusty gases*, *Ann. Rev. Fluid Mech.* (1970), no. 2, 397–446.
- [Mär10] F. Märkle, *Kombinatorische Ablagerung von Biopartikeln mit Hilfe von Laserstrahlung*, Master’s thesis, Heidelberg University, 2010.
- [Mer63] R. B. Merrifield, *Solid phase peptide synthesis. i. the synthesis of a tetrapeptide*, *J. Am. Chem. Soc.* **85** (1963), no. 14, 2149–2154.
- [Mer84] B. Merrifield, *The 1984 nobel prize in chemistry*, Press Release, 1984.
- [MLM⁺01] M. J. McAuliffe, F. M. Lalonde, D. McGarry, W. Gandler, K. Csaky, and B. L. Trus, *Medical image processing, analysis & visualization in clinical research*, CBMS ’01: Proceedings of the Fourteenth IEEE Symposium on Computer-Based Medical Systems (Washington, DC, USA), IEEE Computer Society, 2001, p. 381.
- [Nal04] R. Nallapati, *Discriminative models for information retrieval*, SIGIR ’04: Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval (New York, NY, USA), ACM, 2004, pp. 64–71.
- [Nie09] T. Nierhaus, *Modeling and simulation of dispersed two-phase flow transport phenomena in electrochemical processes*, Ph.D. thesis, Rheinisch-Westfälische Technische Hochschule Aachen, 2009.
- [NLK⁺07a] A. Nesterov, F. Löffler, K. König, U. Trunk, K. Leibe, T. Felgenhauer, F. R. Bischoff, F. Breitling, V. Lindenstruth, V. Stadler, and M. Hausmann, *Measurement of tribo-electric charging of moving micro particles by means of an inductive cylindrical probe*, *Journal of Physics D: Applied Physics* **40** (2007), 6115–6120.
- [NLK⁺07b] A. Nesterov, F. Löffler, K. König, U. Trunk, K. Leibe, T. Felgenhauer, V. Stadler, F. R. Bischoff, F. Breitling, V. Lindenstruth, and M. Hausmann, *Noncontact charge measurement of moving microparticles contacting dielectric surfaces*, *Review of Scientific Instruments* **78** (2007), no. 7, 075111–1–075111–7.

- [NM06] A. Nesterov-Müller, *Ortsgenaue Ablagerung von Aminosäurepartikeln für die kombinatorische Synthese von Peptidarrays auf einen Chip*, Ph.D. thesis, Heidelberg University, 2006.
- [OS88] S. Osher and J. A. Sethian, *Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations*, J. Comput. Phys. **79** (1988), no. 1, 12–49.
- [Pea01] K. Pearson, *On lines and planes of closest fit to systems of points in space*, Philosophical Magazine **2** (1901), no. 6, 559–572.
- [PL98] E. Peirano and B. Leckner, *Fundamentals of turbulent gas-solid flows applied to circulating fluidized bed combustion*, Progress in Energy and Combustion Science **24** (1998), no. 4, 259–296.
- [Ran99] R. Rannacher, *Finite element methods for the incompressible navier-stokes equations*, Tech. report, Institute of Applied Mathematics, University of Heidelberg, 1999.
- [SBK⁺09] C. Schirwitz, I. Block, K. König, A. Nesterov, S. Fernandez, T. Felgenhauer, K. Leibe, G. Torralba, M. Hausmann, V. Lindenstruth, and V. Stadler, *Combinatorial peptide synthesis on a microchip*, Current Protocols in Protein Science **57** (2009), 18.2.1–18.2.13.
- [Sch10] C. Schirwitz, private communication, 2010.
- [Set99] J. A. Sethian, *Level set methods and fast marching methods: Evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science*, Cambridge University Press, June 1999.
- [SFB⁺08] V. Stadler, T. Felgenhauer, M. Beyer, S. Fernandez K. Leibe, S. Güttler, M. Gröning, K. König, G. Torralba, M. Hausmann, V. Lindenstruth, A. Nesterov, I. Block, R. Pipkorn, A. Poustka, F. R. Bischoff, and F. Breitling, *Combinatorial synthesis of peptide arrays with a laser printer*, Angewandte Chemie International Edition **47** (2008), 7132–7135.
- [Ste57] H. Steinhaus, *Sur la division des corps matériels en parties.*, Bull. Acad. Pol. Sci., Cl. III **4** (1957), 801–804 (French).
- [SW72] S. K. Suneja and T. Wasan, *Dispersion of charged particles in a turbulent air stream under transverse flow conditions*, Ind. Eng. Chem. Fundam. **11** (1972), no. 1, 57–66.
- [TM98] C. Tomasi and R. Manduchi, *Bilateral filtering for gray and color images*, ICCV '98: Proceedings of the Sixth International Conference on

- Computer Vision (Washington, DC, USA), IEEE Computer Society, 1998, p. 839.
- [TN97] K. Takahashi and M. Nakazawa, *DNA insight: An image processing system for 2-d gel electrophoresis of genomic DNA*, Genome Informatics (1997), 135–146.
- [Tur90] K. Turkowski, *Filters for common resampling tasks*, Graphics gems (Andrew S. Glassner, ed.), Academic Press Professional, Inc., San Diego, CA, USA, 1990, pp. 147–165.
- [Vap95] V. N. Vapnik, *The nature of statistical learning theory*, Springer-Verlag New York, Inc., New York, NY, USA, 1995.
- [VfOM09] Verbundzentrum für Oberflächenanalyse Münster, *Surface analysis techniques*, Website, 2009, <http://www.vom-muenster.de/analyseverfahren.html>.
- [Vol09] R. Volkmer, *Synthesis and application of peptide arrays: Quo vadis SPOT technology*, ChemBioChem **10** (2009), no. 9, 1431–1442.
- [WGG01] X. Wang, S. Ghosh, and S. W. Guo, *Quantitative quality control in microarray image processing and data acquisition*, Nucleic Acids Research **29** (2001), e75.
- [WJZ09] X. Wang, B. Jin, and W. Zhong, *Three-dimensional simulation of fluidized bed coal gasification*, Chemical Engineering and Processing: Process Intensification **48** (2009), no. 2, 695–705.
- [WKF⁺10] J. Wagner, K. König, T. Förtsch, F. Löffler, S. Fernandez, T. Felgenhauer, F. Painke, G. Torralba, V. Lindenstruth, V. Stadler, F. R. Bischoff, F. Breitling, M. Hausmann, and A. Nesterov, *Microparticle transfer onto pixel electrodes of 45 μ m pitch on HV-CMOS chips – simulation and experiment*, under revision at Sensors and Actuators A, 2010.
- [WO11] J. Wagner and B. Ommer, *Efficiently clustering earth mover’s distance*, ACCV 2010, Part II, LNCS 6493 (R. Kimmel, R. Klette, and A. Sugimoto, eds.), Springer, Heidelberg, 2011, pp. 477–488.

Appendix A

Coupling detection supplements

A.1 Surface coatings

Figure A.1 shows the chemical structure of the surface coatings used for on-chip synthesis. On the left, there is a typical polymer layer as described in [BFB⁺06], on the right, the self-assembled monolayer as described in [LJLV⁺00] is shown.

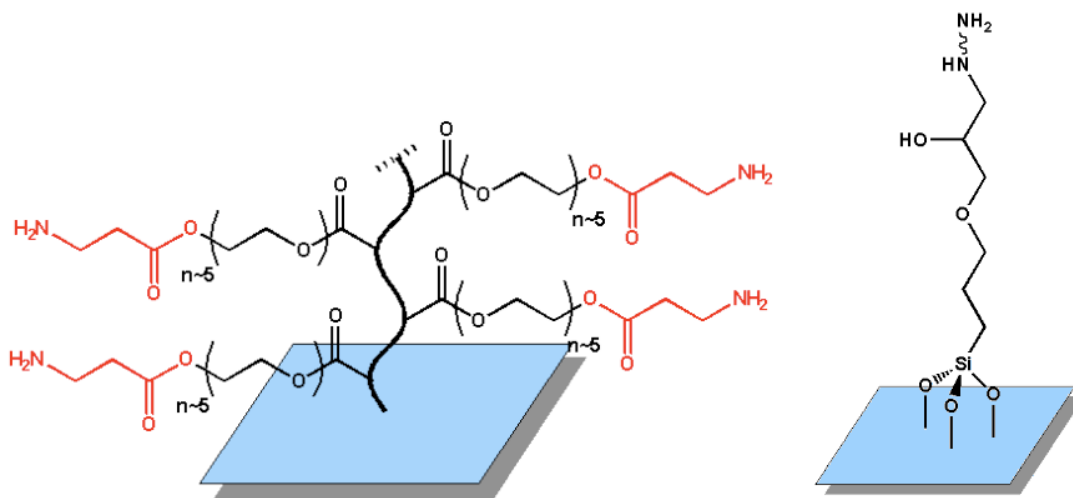


Figure A.1: Left: Polymer surface coating Right: Monolayer surface coating used for chemical activation in on-chip synthesis

Appendix B

Algorithmic supplements

B.1 Correlation related images

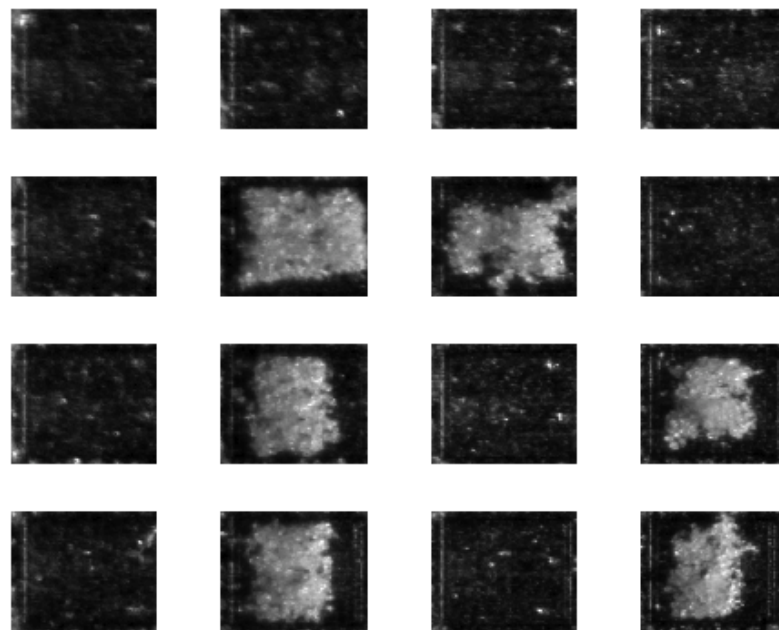


Figure B.1: From top top to bottom: chip cells 1-16 of deposition image after alignment

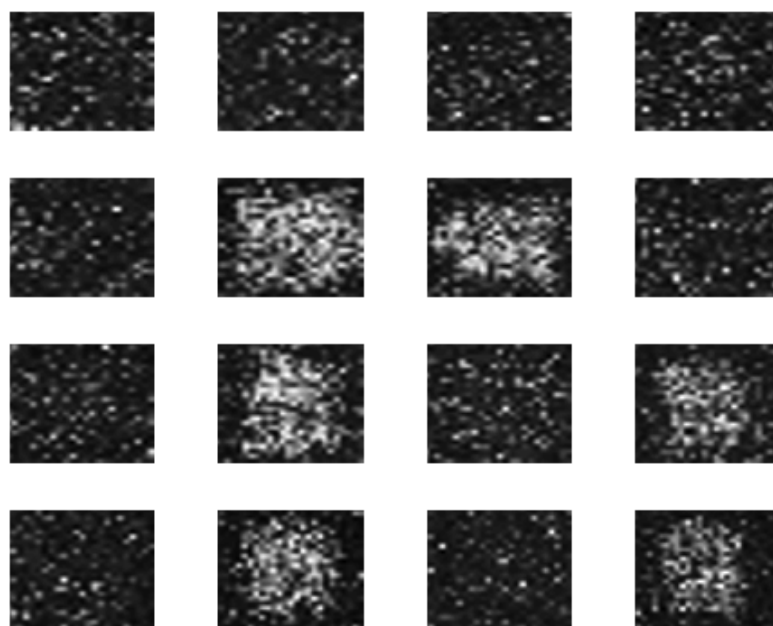


Figure B.2: From top top to bottom: enlarged chip cells 1-16 of tryptophan detection image after alignment

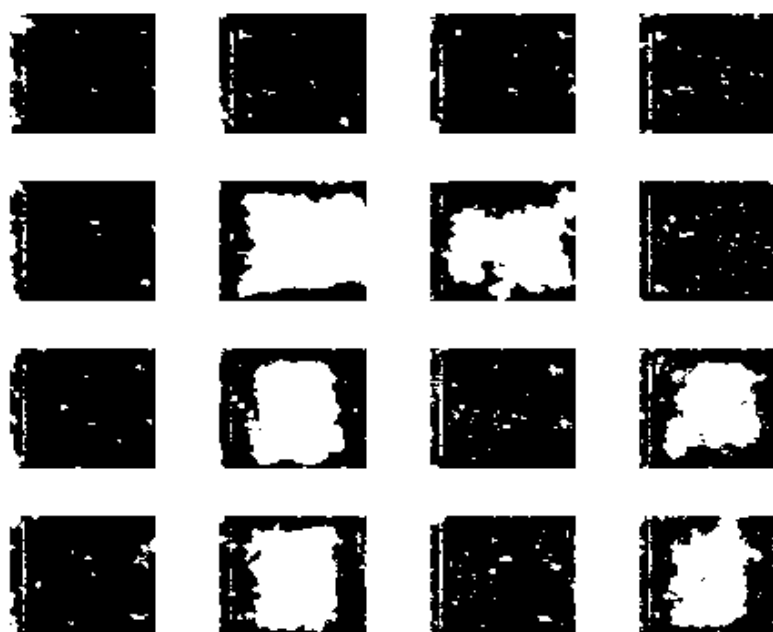


Figure B.3: From top top to bottom: aligned chip cells 1-16 of deposition image after segmentation by thresholding

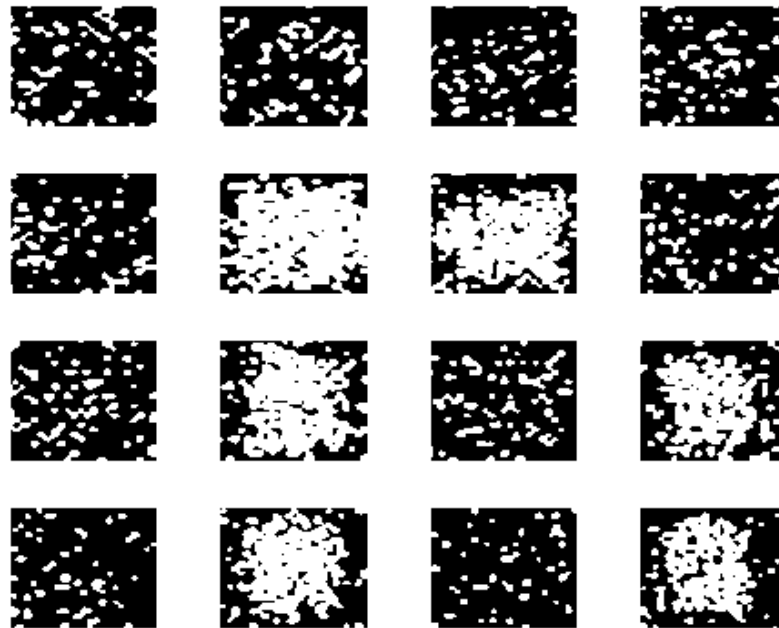


Figure B.4: From top top to bottom: enlarged and aligned chip cells 1-16 of tryptophan detection image after segmentation by thresholding

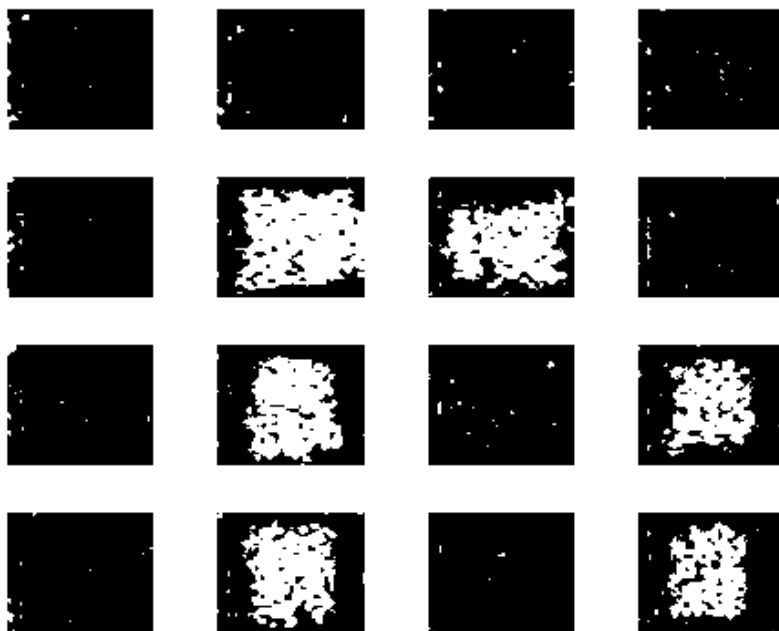


Figure B.5: From top top to bottom: correlation deposited and detected signals (white) from chip cells 1-16 (TP)

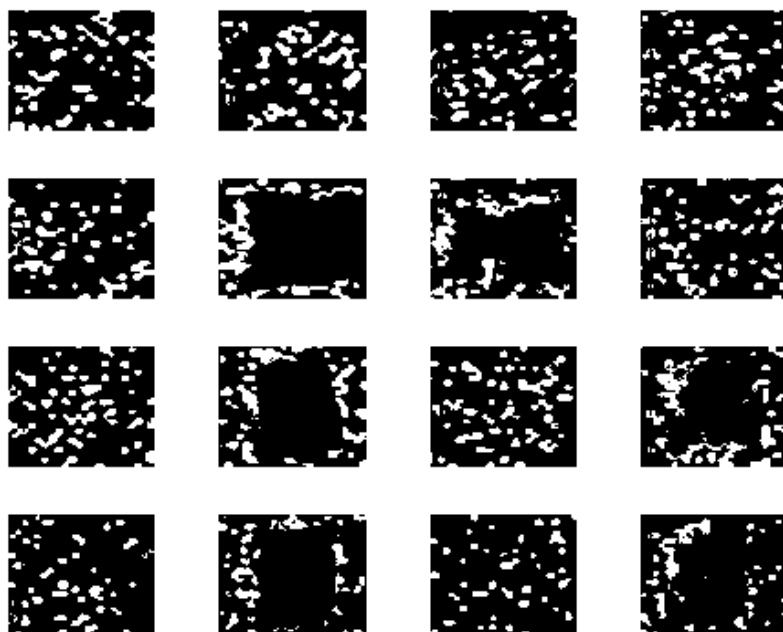


Figure B.6: From top top to bottom: correlation not deposited but detected signals (white) from chip cells 1-16 (FP)

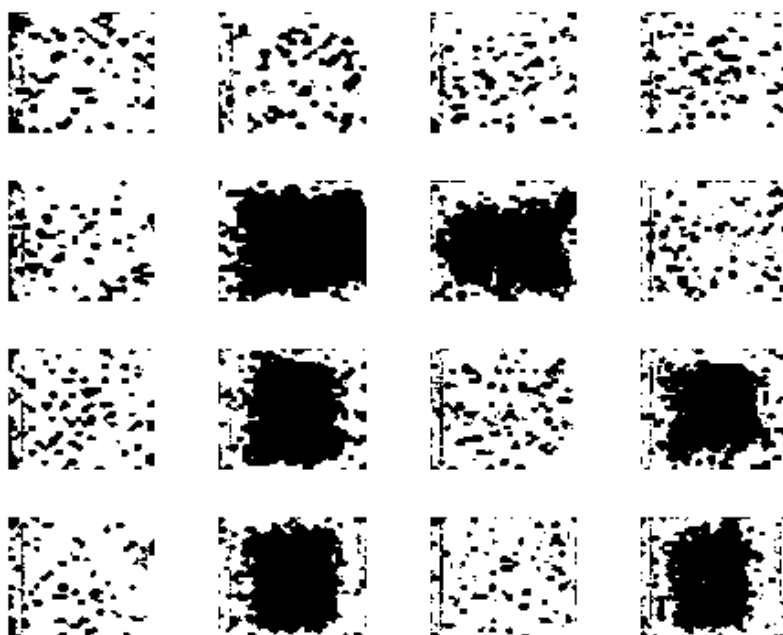


Figure B.7: From top top to bottom: correlation not deposited and not detected signals (white) from chip cells 1-16 (TN)

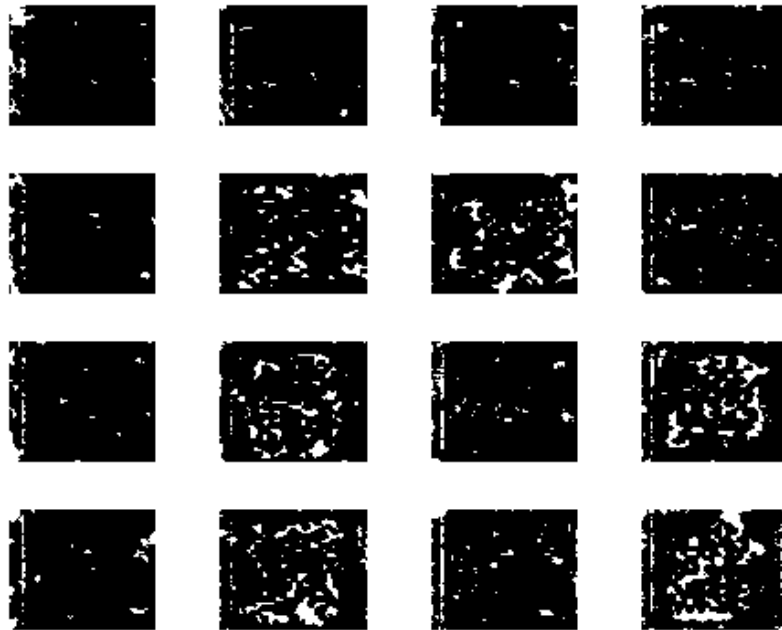


Figure B.8: From top top to bottom: correlation deposited but not detected signals (white) from chip cells 1-16 (FN)

B.2 Usage of the training routine

At first, MATLAB has to be started and training.m has to be loaded in the editor. After having specified the following input values, the routine can be run:

- the image file, which specifies the path to the image to be analysed
- the pattern file, which specifies the path to the file of the deposition pattern; the deposition information is encoded in a text file with the figures 0 for cells not to be covered and 1 for cells to be covered which are separated by one space after each figure
- the width and height of the image to be analysed
- the so called series number, normally set to 'new' for untested setups; this parameter encodes the used setup (microscope, camera, objective) and the cell length of the chip cells (s. B.4.1 for details)
- the calibrate boolean, set to 0 if no calibration with a white image is required or set to 1 if this calibration is desired; in the latter case, the path to the white image also has to be specified, which is added to the parameter file
- the convolve boolean, set to 0 if no convolution of the input image is needed or set to 1 if the image is supposed to be blurred in order to reduce the noise; in

the latter case, the width of the Gaussian kernel should not exceed the optical resolution; the parameters of the kernel are also added to the list of parameters

- the maximum wavelength of the used light in the optical setup in units of μm
- the numerical aperture of the microscope objective
- the total magnification, including the magnification of the objective and possibly mounted TV-adaptations
- the pixel pitch of the microscope camera in units of the μm
- the parameter ν , which should be in the range of 10 in order to separate the uncovered and covered chip cells optimally (s. 5.6 for details)
- the parameter file, which specifies the path to the mat-file in which the parameters are saved

As result, the parameter file is written to the path specified in the last input parameter called parameter file. During the test phase, the variables in this mat-file are loaded into the MATLAB workspace so that the user does not have to specify any further input parameters then.

B.3 Usage of testing routine

Contrary to the training routine, the testing routine does not require many parameters to be set by the user. The following specifications are necessary to be feed into `processing.m` in order to analyse only one image or into `simple_processing.m` in order to start the automated version of the processing routine:

- the parameter file, which contains the path to the parameter file that contains all necessary parameters determined in the training phase
- the image file, which specifies the path to the image to be analysed
- the pattern file, which specifies the path to the file of the deposition pattern; the deposition information is encoded in a text file with the figures 0 for cells not to be covered and 1 for cells to be covered which are separated by one space after each figure

For the automated quality analysis in `simple_process.m` the user additionally has to specify the path to the following folders:

- the image folder, in which the test image is contained before the analysis
- the output folder, in which all analysed images are transferred after a successful analysis and in which all results are written, e.g. the detailed quality analysis results for each cell

- the response folder, in which the response for LabView is written in form of a text file that contains a number which encoded the exit status of MATLAB as described in the chapter 5.8
- the debug folder, in which all images and results that caused an error are saved for later error inspection

B.4 Special parameter specifications

B.4.1 Series numbers

Instead of specifying the series numbers as *'new'* for not yet determined grids, the following series numbers, as shown in Table B.1, are provided for already tested setups.

series number	microscope	camera	cell length [pixels]
04x		Sony DXC-950P	48
10x		Sony DXC-950P	120
05c	Axiovert 35	Canon EOS 1000D	222
10c	Axiovert 35	Canon EOS 1000D	438
05p	Axiovert 35	Zeiss Progres C5	95
10p	Axiovert 35	Zeiss Progres C5	188

Table B.1: Series numbers

If the cell length determined in the training step does not match one of these cell lengths, the program prints a warning message on the screen that the cell length is not yet contained in the data base of series numbers and then continues processing with the new cell length, setting the series number to *'niu'* (not in use).

B.5 Segmentations of synthesis images

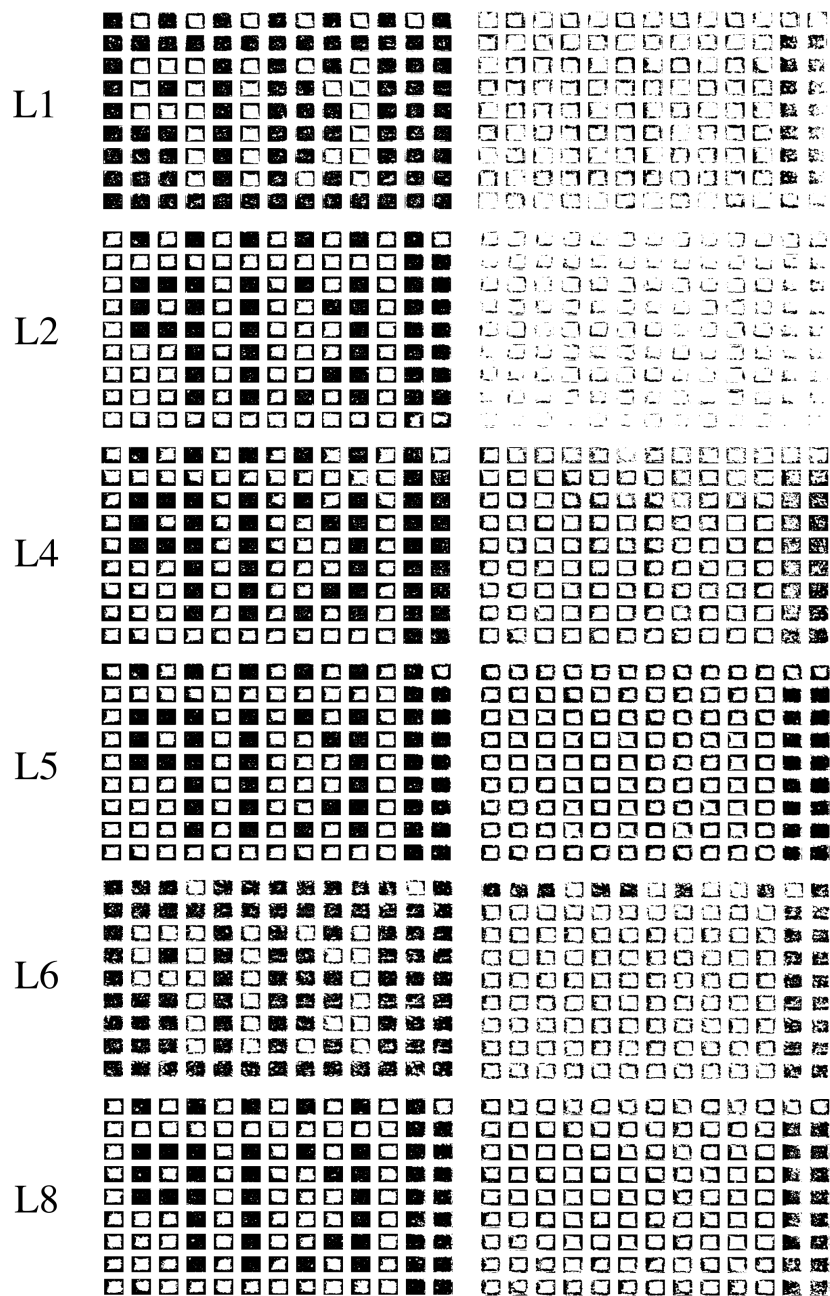


Figure B.9: From top to bottom: segmentations of the particle deposition images shown in Figure 5.30 and Figure 5.31 per layer (layers with one deposition only are omitted here)