

---

**Novel design and data analysis  
concepts for MINFLUX  
nanoscopy**

---

DISSERTATION

for the award of the degree  
“Doctor rerum naturalium”  
of the Georg-August-Universität Göttingen

within the International Max Planck Research School  
“Physics of Biological and Complex Systems”  
of the Georg-August University School of Science

submitted by  
**Hannah Eva Barthel**  
from Kiel

Göttingen, 2021

## **Thesis advisory committee and examination board**

**Prof. Dr. Stefan W. Hell (1st referee)**

Department of NanoBiophotonics,  
Max Planck Institute for Biophysical Chemistry, Göttingen

**Prof. Dr. Silvio Rizzoli (2nd referee)**

Department of Neuro- and Sensory Physiology,  
University Medical Center Göttingen

**Prof. Dr. Sarah Köster**

Research Group for Cellular Biophysics,  
Institute for X-ray Physics,  
Georg-August-Universität Göttingen

## **Further members of the examination board**

**apl. Prof. Dr. Alexander Egner**

Department of Optical Nanoscopy,  
Institut für Nanophotonik Göttingen e.V.

**Prof. Dr. Axel Munk**

Felix-Bernstein-Chair for Mathematical Statistics,  
Institute for Mathematical Stochastics,  
Georg-August-Universität Göttingen

**Prof. Dr. Claus Ropers**

Research Group for Nano-Optics and Ultrafast Dynamics,  
IV. Physical Institute,  
Georg-August-Universität Göttingen

**Date of the oral examination:** November 8, 2021

*“The science of operations, as derived from mathematics more especially, is a science of itself, and has its own abstract truth and value.”*

Ada Lovelace



# Abstract

MINFLUX is a novel concept for super-resolution fluorescence microscopy that achieves unprecedented image resolution despite the limited photon budget of fluorescent molecules. The goal of this thesis is to improve MINFLUX image quality using new optical and analytical approaches. To this end, we aim to increase the information contained in the emitted photons by optimizing the excitation light pattern. Furthermore, by integrating new concepts into the post-processing workflow, we utilize this information in an optimal manner.

In the first part, I examine the suitability of low-order Lissajous figures for 2D MINFLUX illumination schemes that are compatible with alternative, simpler scanning systems based on galvanometric beam deflectors. The width and aspect ratio of these figures are optimized as a function of different imaging parameters in order to maximize the theoretically achievable localization precision. The optimized Lissajous geometries exhibit a substantially more isotropic precision than the previously published 2D excitation pattern.

In the second part, I implement a fast, parameter-free 3D position estimator for MINFLUX nanoscopy based on deep learning. After an optimization of the architecture and the training data, the neural network yields a lower localization error than the currently used maximum-likelihood estimator, especially in the low-photon and high-background regimes. Moreover, the amount of time required for estimating the emitter positions is significantly reduced.

To conclude, this thesis provides new methods for increasing the information content of MINFLUX images by effectively making best use of the available photons. By allowing simpler instrumentation and reducing the number of analysis parameters, the approaches developed herein will reduce the complexity and improve the performance of future MINFLUX nanoscopes.



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Super-resolution microscopy . . . . .	3
1.2	MINFLUX localization concept and statistics . . . . .	5
1.2.1	Statistics of photon detection . . . . .	7
1.2.2	Information theory . . . . .	10
1.2.3	Cramér–Rao bound for 2D localization . . . . .	11
1.3	Scope of this thesis . . . . .	14
<b>2</b>	<b>Lissajous illumination patterns for MINFLUX nanoscopy</b>	<b>15</b>
2.1	Introduction . . . . .	15
2.1.1	Beam positioning in MINFLUX nanoscopy . . . . .	16
2.1.2	Scope of this chapter . . . . .	17
2.2	Parameters of Lissajous illumination patterns . . . . .	18
2.2.1	Number of exposures . . . . .	19
2.2.2	Width of the Lissajous figure . . . . .	23
2.2.3	Aspect ratio of the Lissajous figure . . . . .	24
2.3	Optimization of the geometrical parameters . . . . .	26
2.3.1	Optimal width of the Lissajous figure . . . . .	27
2.3.2	Optimal aspect ratio of the Lissajous figure . . . . .	30
2.4	Localization precision of the optimized Lissajous patterns . . . . .	32
2.4.1	Comparison to the triangular pattern . . . . .	33
2.4.2	Cramér–Rao bound of the six-point Lissajous pattern . . . . .	36
2.5	Summary and discussion . . . . .	39
<b>3</b>	<b>Deep-learning-based data analysis for MINFLUX nanoscopy</b>	<b>41</b>
3.1	Introduction . . . . .	41
3.1.1	Post-processing in MINFLUX nanoscopy . . . . .	42
3.1.2	Deep learning in localization microscopy . . . . .	43
3.1.3	Scope of this chapter . . . . .	46
3.2	Theory and methods . . . . .	46
3.2.1	Maximum-likelihood estimation in MINFLUX . . . . .	46

3.2.2	Fundamentals of deep learning . . . . .	49
3.2.3	Localization performance metrics . . . . .	55
3.3	Data simulation and network design . . . . .	56
3.4	Optimization of the network architecture and training data	62
3.5	Comparison to the maximum-likelihood estimator . . . . .	71
3.6	Summary and discussion . . . . .	82
<b>4</b>	<b>Conclusion and outlook</b>	<b>87</b>
	<b>Bibliography</b>	<b>91</b>
<b>A</b>	<b>Supplementary figures for Lissajous illumination patterns</b>	<b>105</b>
<b>B</b>	<b>Supplementary figures for deep-learning-based data analysis</b>	<b>111</b>
	<b>List of Figures</b>	<b>115</b>
	<b>List of Tables</b>	<b>119</b>

## List of Abbreviations

<b>CRB</b>	Cramér–Rao bound
<b>EBP</b>	excitation beam pattern
<b>EOD</b>	electro-optic deflector
<b>MLE</b>	maximum-likelihood estimator
<b>MSE</b>	mean squared error
<b>numLMSE</b>	numerically unbiased modified least-mean-square estimator
<b>PALM</b>	photoactivated localization microscopy
<b>PSF</b>	point-spread function
<b>ReLU</b>	rectified linear unit
<b>ROI</b>	region of interest
<b>SBR</b>	signal-to-background ratio
<b>SGD</b>	stochastic gradient descent
<b>SMLM</b>	single-molecule localization microscopy
<b>STED</b>	stimulated emission depletion
<b>STORM</b>	stochastic optical reconstruction microscopy



# 1 Introduction

Microscopy facilitates the understanding of biological processes by putting them into spatial and temporal context. These insights can be provided across multiple scales, whether in whole organisms when observing embryonic development [1, 2], on a cellular level when elucidating the structure of the neuronal cytoskeleton [3, 4] or the architecture of mitochondria [5–7], or on the scale of multi-protein assemblies such as the nuclear pore complex [8]. Fluorescence microscopy plays a special role because it allows us to identify specific structures and to study their relationship by labeling them with different markers. However, diffraction fundamentally limits the level of detail at which these structures can be resolved using conventional optical microscopes to a scale of approximately 200 nm. This limitation was postulated by Ernst Abbe in 1873 [9].

More than a hundred years later, in 1994, Hell and Wichmann proposed a method to overcome this barrier by exploiting the photophysical switching properties of fluorophores [10], and super-resolution microscopy, or nanoscopy, was born. This proposal was followed by the first experimental demonstrations of STED<sup>1</sup> microscopy [11, 12], an implementation of so-called *coordinate-targeted* super-resolution microscopy. Later, *coordinate-stochastic* approaches to break the diffraction limit, namely STORM<sup>2</sup> [13],

---

<sup>1</sup> stimulated emission depletion

<sup>2</sup> stochastic optical reconstruction microscopy

PALM<sup>3</sup> [14], and FPALM<sup>4</sup> [15], were independently developed. These novel imaging techniques provided a spatial resolution of around 20 nm [13–16], corresponding to a tenfold increase over diffraction-limited microscopy. For this breakthrough, Eric Betzig, Stefan W. Hell, and William E. Moerner were awarded the 2014 Nobel Prize in Chemistry.

The beginnings of optical microscopy were marked by technical advancements such as aberration-corrected lenses, high-intensity lasers, powerful computers, and efficient electronic detectors to push the resolution towards the diffraction limit [17, 18]. With the advent of nanoscopy, other factors started to determine the achievable resolution, namely the size of the labels and most importantly the photostability and overall switching performance of the fluorophores [19]. Therefore, progress in the recent history of optical microscopy was often driven by advances in labeling concepts, for example protein tags [20–22] and nanobodies [23, 24], photostable and photoswitching fluorescent dyes [25–28], or other strategies to make best use of the limited photon budget.

With the relatively new MINFLUX concept introduced in 2016, a 22-fold reduction of required photon detections compared to STORM-like methods was demonstrated [29]. By combining the particular strengths of coordinate-targeted and coordinate-stochastic super-resolution microscopy, it became possible to push the resolution in fixed and living cells down to a truly molecular scale [30]. In the next section, the main ideas behind STED microscopy and STORM are outlined. Afterwards, the MINFLUX concept and the underlying statistics are introduced, and the scope of the present thesis is defined.

---

<sup>3</sup> photoactivated localization microscopy

<sup>4</sup> fluorescence photoactivated localization microscopy

## 1.1 Super-resolution microscopy

In conventional optical microscopy, the image of a point-like object is a spot with an intensity peak of a certain width, given by

$$\sigma_{\text{PSF}} \approx 0.42 \frac{\lambda}{2\text{NA}}. \quad (1.1)$$

The extent of this so-called diffraction-limited *point-spread function* (PSF) depends on the wavelength  $\lambda$  of the emitted light and on the numerical aperture (NA) of the optical system [9, 31]. Emitters that are close together have overlapping PSFs and cannot be distinguished. As mentioned before, this fundamental resolution limit has been overcome by different methods. They all rely on switching emitters between distinguishable molecular states such that adjacent emitters can be detected sequentially. There is typically a bright *on* state, which permits fluorophores to emit photons, and at least one dark *off* state, where fluorescence is inhibited. However, the approaches differ in how they realize the transition between these states and how they localize the fluorescent emitters in space [32].

Coordinate-targeted super-resolution microscopy techniques such as STED and RESOLFT<sup>5</sup> [33–35] reduce the effective size of the PSF by driving fluorophores to the off state with a beam featuring a local intensity minimum. The off-switching and excitation beams are co-aligned, leaving only a small fraction of molecules at the point of minimum intensity to fluoresce. Images are acquired by scanning both beams over the sample and sequentially switching the fluorophores between the on and off states. Each photon registered by a point detector can be assigned a coordinate defined by the well-known position of the intensity minimum. The achievable resolution is given by the extent of the effective PSF [36], which in turn depends on

---

<sup>5</sup> reversible saturable optical fluorescence transitions

the ratio  $I_{\max}/I_{\text{sat}}$ ,

$$\sigma \approx \frac{\sigma_{\text{PSF}}}{\sqrt{1 + I_{\max}/I_{\text{sat}}}}. \quad (1.2)$$

Here,  $\sigma_{\text{PSF}}$  is the size of the diffraction-limited PSF defined in (1.1),  $I_{\max}$  is the maximum intensity of the off-switching light, and  $I_{\text{sat}}$  denotes the saturation intensity, for which the off-switching probability is 50%. While this relationship yields unlimited resolution in theory, photobleaching of the fluorophores limits the applicable off-switching intensity in practice.

Coordinate-stochastic methods like PALM and STORM acquire a large number of images, each of which contains the diffraction-limited spots of a sparse, random subset of individual fluorescent emitters. Sparsity is achieved by stochastic switching of photoswitchable molecules between the on and off states. In contrast to the coordinate-targeted approach, the origin of each photon is defined as the center of its diffraction-limited fluorescence spot recorded by a camera. How precisely this centroid can be estimated depends on the number  $N$  of detected photons per event [37–40]:

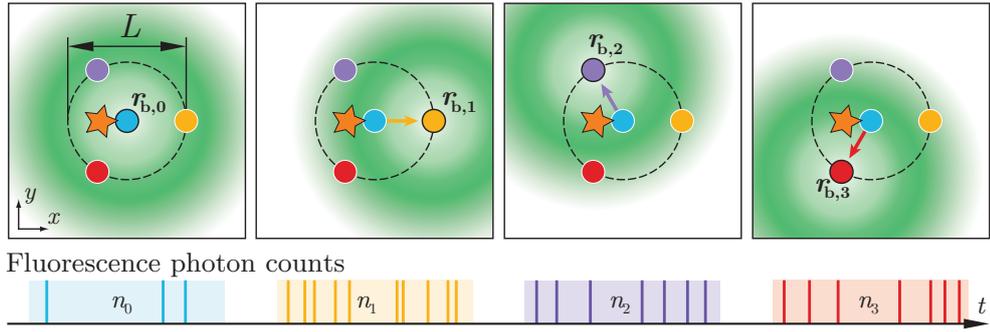
$$\sigma \approx \frac{\sigma_{\text{PSF}}}{\sqrt{N}}, \quad (1.3)$$

where  $\sigma_{\text{PSF}}$  is again the size of the diffraction-limited PSF given in (1.1) and experimental error sources were neglected. Consequently, a high localization precision demands a large number of fluorescence photons, which, however, are only available on a limited budget.

## 1.2 MINFLUX localization concept and statistics

In MINFLUX nanoscopy the position of an isolated molecule is determined by probing its fluorescence with a doughnut-shaped excitation beam and collecting the emitted photons with a point detector. The beam is placed at several positions forming the so-called *excitation beam pattern* (EBP) around the molecule coordinate, whose initial estimate is obtained from a diffraction-limited overview image. The EBP that is typically used for 2D MINFLUX is depicted in Figure 1.1. Sparsity is again achieved by randomly activating only a single emitter at a time within the confocal volume using stochastic switching. The well-known excitation intensity distribution encodes information on the molecule position. The latter is obtained from the relative fluorescence photon counts collected at each *exposure*. The closer the doughnut minimum approaches the emitter, the more photon efficient the measurement becomes, meaning that fewer photons are required to reach a certain localization precision. In theory, localization would be possible without detecting a single photon in the extreme case of a local intensity minimum coinciding with the molecule position [29].

For cellular imaging, the MINFLUX concept is typically implemented in an iterative scheme. The size of the EBP determines the photon efficiency of the measurement but also defines the local field of view. Initially, a large field of view is necessary to find the molecule inside the diffraction-limited activation volume, requiring a large EBP. For the final localization, however, the beam should be scanned around the emitter as tightly as possible in order to maximize the photon efficiency. Therefore, the size of the EBP is iteratively reduced. In each step, a defined number of photons is acquired, from which a preliminary molecule position is derived. This coordinate determines the center of the pattern for the next step. As the measure-

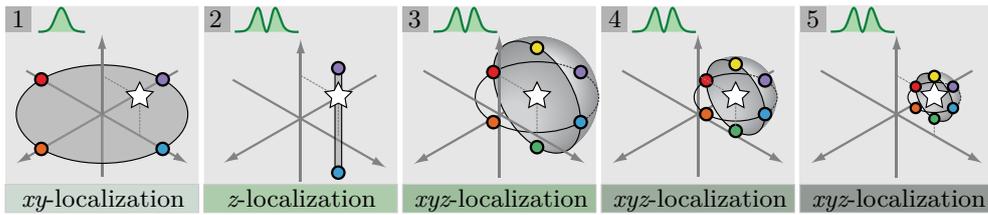


**Figure 1.1: 2D excitation pattern with four doughnut-beam positions.** The fluorescent emitter marked by the orange star is exposed to the doughnut excitation beam shown in green, which is sequentially placed at the four positions  $\mathbf{r}_{\mathbf{b},0-3}$  marked by the colored circles. The coordinates  $\mathbf{r}_{\mathbf{b},1-3}$  are arranged on a circle of diameter  $L$  around the central point  $\mathbf{r}_{\mathbf{b},0}$ . At each exposure a certain number  $n_{0-3}$  of fluorescence photons is detected, resulting in the count trace indicated in the lower panel. Figure adapted from Balzarotti et al. [29]. Reprinted with permission from AAAS.

ment advances, the exposures approach the true emitter position, which makes the detected photons more and more informative. The iteration scheme often employs a combination of coarse alignment with Gaussian beams and precise localization with doughnut beams. In the final iterations a cross-like arrangement of six positions around an additional central exposure point is used, as indicated in Figure 1.2.

MINFLUX has been implemented as a scanning microscope with confocal detection for both imaging and tracking applications. Dual-color measurements of mitochondrial proteins in fixed cells yielded a 3D localization precision of approximately 5 nm [42]. In tracking experiments with single fluorophores, a spatiotemporal resolution of around 20 nm and 100  $\mu\text{s}$  could be achieved [43].

The following sections of this introduction summarize the mathematical



**Figure 1.2: 3D iterative MINFLUX scheme.** The fluorescent emitter marked by the white star is exposed to a sequence of EBPs: 2D localization with a Gaussian beam and four exposures, localization along the optical axis using a 3D-doughnut profile and two exposures, and three iterations with a cross-like seven-point EBP using the 3D-doughnut shape for  $xyz$ -localization. The pattern diameter was iteratively reduced:  $L = \{300 \text{ nm}, 400 \text{ nm}, 150 \text{ nm}, 90 \text{ nm}, 40 \text{ nm}\}$ . Figure adapted from Gwosch et al. [41].

foundations of MINFLUX and derive the theoretical limit for the achievable localization precision. These calculations were published in the original MINFLUX publication [29, Supplementary Materials] and were presented in greater detail by Gwosch [44, Section 2.2]. For simplicity we only consider the case of static, that is, non-iterative, MINFLUX.

### 1.2.1 Statistics of photon detection

In the following, we will consider the case of a single fluorescent emitter located at  $\mathbf{r}_m \in \mathbb{R}^d$ . A sequential exposure to  $K$  different intensity distributions  $\{I_0(\mathbf{r}), I_1(\mathbf{r}), \dots, I_{K-1}(\mathbf{r})\}$  yields a collection of photon counts  $\mathbf{n} = \{n_0, n_1, \dots, n_{K-1}\}$ , which comprise signal photons as well as background and dark count contributions. In the situation of negligible background and detector dark counts, the signal photons obey a Poisson distribution with mean  $\lambda_i^{(0)}(\mathbf{r}_m)$ , which is proportional to the excitation intensity  $I_i(\mathbf{r}_m)$ . We also assume the background and dark counts to follow Poisson distributions, whose sum has mean  $\lambda_i^{(b)}(\mathbf{r}_m)$ . Hence, the detected

counts  $n_i$  also obey a Poisson distribution

$$P(n_i) \sim \text{Poisson}(\lambda_i) = \frac{\lambda_i^{n_i} e^{-\lambda_i}}{n_i!} \quad \forall i \in [0, 1, \dots, K-1] \quad (1.4)$$

with mean

$$\lambda_i(\mathbf{r}_m) = \lambda_i^{(0)}(\mathbf{r}_m) + \lambda_i^{(b)}(\mathbf{r}_m). \quad (1.5)$$

The Poissonian mean of the photon count distribution in absence of background is proportional to the light intensity and given by

$$\lambda_i^{(0)}(\mathbf{r}_m) = c_e q_e \sigma_a I_i(\mathbf{r}_m) \quad \forall i \in [0, 1, \dots, K-1], \quad (1.6)$$

where  $c_e$  is the detection efficiency of the imaging system.  $q_e$  is the quantum yield and  $\sigma_a$  the absorption cross-section of the fluorophore at the excitation wavelength [29, (S1)]. Furthermore, we assume the background contributions to be identical for all exposures:

$$\lambda_i^{(b)}(\mathbf{r}_m) = \lambda_j^{(b)}(\mathbf{r}_m) =: \lambda^{(b)}(\mathbf{r}_m) \quad \forall i, j \in [0, 1, \dots, K-1]. \quad (1.7)$$

Under the condition of a fixed total number  $N = \sum_{i=0}^{K-1} n_i$  of detected photons, the joint probability distribution of the photon counts can be described by a multinomial distribution [29, (S3)–(S4)]

$$P(\mathbf{n}|N) \sim \text{Multinomial}(\mathbf{p}, N) = \frac{N!}{\prod_{i=0}^{K-1} n_i!} \prod_{j=0}^{K-1} p_j^{n_j} \quad (1.8)$$

with a multinomial parameter vector  $\mathbf{p}$  containing the event probabilities

$$p_i(\mathbf{r}_m) = \frac{\lambda_i}{\sum_{j=0}^{K-1} \lambda_j} = \frac{\lambda_i^{(0)} + \lambda_i^{(b)}}{\sum_{j=0}^{K-1} (\lambda_j^{(0)} + \lambda_j^{(b)})} \quad \forall i \in [0, 1, \dots, K-1]. \quad (1.9)$$

Note that for a fixed total photon number  $N$ , there are only  $K - 1$  independent photon counts  $n_i$  and parameters  $p_i$  spanning the so-called *reduced  $\mathbf{p}$ -space*. Then the parameter  $p_{K-1}$  in (1.8) can be replaced by a function of the other  $K - 1$  event probabilities:

$$p_{K-1} = 1 - \prod_{k=0}^{K-2} p_k. \quad (1.10)$$

In order to simplify the expression for the multinomial parameter vector given in (1.9), we define a signal-to-background ratio

$$\text{SBR}(\mathbf{r}_m) = \frac{\sum_{i=0}^{K-1} \lambda_i^{(0)}(\mathbf{r}_m)}{\sum_{j=0}^{K-1} \lambda_j^{(b)}(\mathbf{r}_m)} \propto \frac{\sum_{i=0}^{K-1} I_i(\mathbf{r}_m)}{K \lambda^{(b)}(\mathbf{r}_m)}, \quad (1.11)$$

which is a function of the excitation intensities  $\{I_i(\mathbf{r})\}$  and the emitter position  $\mathbf{r}_m$ . This yields a simple formula for the components of  $\mathbf{p}$  [29, (S30)]:

$$p_i(\mathbf{r}_m) \approx \frac{\text{SBR}(\mathbf{r}_m)}{\text{SBR}(\mathbf{r}_m) + 1} p_i^{(0)}(\mathbf{r}_m) + \frac{1}{\text{SBR}(\mathbf{r}_m) + 1} \frac{1}{K}. \quad (1.12)$$

Here,  $\{p_i^{(0)}\}$  are the components of the background-free multinomial parameter vector

$$p_i^{(0)}(\mathbf{r}_m) = \frac{\lambda_i^{(0)}}{\sum_{j=0}^{K-1} \lambda_j^{(0)}} \approx \frac{I_i(\mathbf{r}_m)}{\sum_{j=0}^{K-1} I_j(\mathbf{r}_m)} \quad \forall i \in [0, 1, \dots, K - 1], \quad (1.13)$$

which are independent of the emitter brightness.

## 1.2.2 Information theory

In order to quantify the information content of the MINFLUX localization scheme, one can calculate the Fisher information that the photon collection  $\mathbf{n}$  holds on the parameter vector  $\mathbf{p}$  [44, 45, (2.32)]:

$$\mathcal{F}_{\mathbf{p},ij} = \mathbb{E}_{\mathbf{p}} \left( \frac{\partial}{\partial p_i} \ln \mathcal{L}(\mathbf{p}|\mathbf{n}) \frac{\partial}{\partial p_j} \ln \mathcal{L}(\mathbf{p}|\mathbf{n}) \right) \quad \forall i, j \in [0, 1, \dots, K-2]. \quad (1.14)$$

The likelihood function  $\mathcal{L}(\mathbf{p}|\mathbf{n})$  for the parameter vector  $\mathbf{p}$  provided a measured photon set  $\mathbf{n}$  equals the multinomial distribution defined in (1.8). Conditioned to a fixed total photon number  $N$  and using (1.10), this yields

$$\mathcal{L}(\mathbf{p}|\mathbf{n}) = \text{P}(\mathbf{n}|\mathbf{p}, N) = \frac{N!}{\prod_{i=0}^{K-1} n_i!} \prod_{j=0}^{K-2} p_j^{n_j} \left( 1 - \prod_{k=0}^{K-2} p_k \right)^{n_{K-1}}. \quad (1.15)$$

Inserting this formula into (1.14) results in the following expression for the Fisher information matrix [29, (S8)]:

$$\mathcal{F}_{\mathbf{p},ij} = N \left( \frac{1}{p_{K-1}} + \delta_{ij} \frac{1}{p_i} \right) \quad \forall i, j \in [0, 1, \dots, K-2], \quad (1.16)$$

where  $\delta_{ij}$  is the Kronecker delta.

However, a MINFLUX measurement aims to estimate the position  $\mathbf{r}_{\mathbf{m}}$  of an emitter, so the quantity we are interested in is the Fisher information of the photon counts on  $\mathbf{r}_{\mathbf{m}}$ . Applying a coordinate transformation from the reduced  $\mathbf{p}$ -space of dimensionality  $K-1$  to the  $d$ -dimensional  $\mathbf{r}$ -space yields the expression [29, (S11)]

$$\mathcal{F}_{\mathbf{r}_{\mathbf{m}},ij} = N \sum_{k=0}^{K-1} \frac{1}{p_k} \frac{\partial p_k}{\partial r_{\mathbf{m},i}} \frac{\partial p_k}{\partial r_{\mathbf{m},j}} \quad \text{with } i, j \in [1, \dots, d]. \quad (1.17)$$

The so-called *Cramér–Rao bound* (CRB)  $\Sigma_{\text{CRB}}$  is the inverse of the Fisher information and sets a lower bound on the covariance of any unbiased estimator. Thus, we can formulate the Cramér–Rao inequality [46, 47] for the covariance matrix  $\Sigma(\mathbf{r}_{\mathbf{m}})$  of the emitter position,

$$\Sigma(\mathbf{r}_{\mathbf{m}}) \geq \Sigma_{\text{CRB}}(\mathbf{r}_{\mathbf{m}}) = \mathcal{F}_{\mathbf{r}_{\mathbf{m}}}^{-1}, \quad (1.18)$$

which fundamentally limits the achievable localization precision.

In this work we will analyze two performance metrics associated with the CRB. These are the square-root arithmetic mean of the eigenvalues  $\{\sigma_i^2\}$  of  $\Sigma_{\text{CRB}}$ ,

$$\sigma_{\text{CRB}}(\mathbf{r}_{\mathbf{m}}) = \sqrt{\frac{1}{d} \text{tr}(\Sigma_{\text{CRB}}(\mathbf{r}_{\mathbf{m}}))}, \quad (1.19)$$

and the CRB isotropy

$$\mathbb{I}(\mathbf{r}_{\mathbf{m}}) = \frac{\min_{i \in [1, d]} \sigma_i(\mathbf{r}_{\mathbf{m}})}{\max_{i \in [1, d]} \sigma_i(\mathbf{r}_{\mathbf{m}})}, \quad (1.20)$$

where  $\text{tr}$  is the trace operator and  $d$  the dimensionality of  $\mathbf{r}_{\mathbf{m}}$ .

### 1.2.3 Cramér–Rao bound for 2D localization

In this section explicit expressions for the CRB will be presented for the 2D localization scheme with four exposures shown in Figure 1.1. The illumination beam has a doughnut-shaped intensity distribution [29, (S17)]

$$I(\mathbf{r}) = I_{\text{max}} \cdot 4e \ln 2 \frac{\|\mathbf{r}\|^2}{\text{FWHM}^2} \exp\left(-4 \ln 2 \frac{\|\mathbf{r}\|^2}{\text{FWHM}^2}\right), \quad (1.21)$$

where  $I_{\max}$  is the peak intensity in the crest of the doughnut and  $\frac{\text{FWHM}}{\sqrt{\ln 2}}$  its peak-to-peak diameter. For simplicity we assume all four intensity distributions to be identical except for a displacement by  $\mathbf{r}_{\mathbf{b},i}$ :

$$I_i(\mathbf{r}) = I(\mathbf{r} - \mathbf{r}_{\mathbf{b},i}). \quad (1.22)$$

Three of the exposures form an equilateral triangle with a circumcircle of diameter  $L$ , whereas the fourth beam position coincides with the center of the triangle, often referred to as the *origin*:

$$\begin{aligned} \mathbf{r}_{\mathbf{b},0} &= (0, 0), \\ \mathbf{r}_{\mathbf{b},i} &= \frac{L}{2}(\cos \alpha_i, \sin \alpha_i) \quad \text{with } \alpha_i = (i - 1) \cdot \frac{2\pi}{3} \text{ and } i \in [1, 2, 3]. \end{aligned} \quad (1.23)$$

The definitions of the doughnut shape and of the beam arrangement can be inserted into the expression for the multinomial parameter vector  $\mathbf{p}^{(0)}$  given in (1.13). Hence, the CRB for localizing a molecule in the absence of background at the center of the EBP with  $K = 4$  exposures in  $d = 2$  dimensions is given by [29, (S27)]

$$\sigma_{\text{CRB}}^{(0)}(\mathbf{0}) = \frac{L}{2\sqrt{2N}} \frac{1}{1 - \frac{L^2 \ln 2}{\text{FWHM}^2}} \stackrel{L \ll \frac{\text{FWHM}}{\sqrt{\ln 2}}}{\approx} \frac{L}{2\sqrt{2N}}. \quad (1.24)$$

Here, we used the definitions of  $\sigma_{\text{CRB}}$  and  $\mathcal{F}_{\mathbf{r}_m}$  provided in (1.19) and (1.17), respectively. For EBP diameters  $L$  that are small compared to the doughnut size, the CRB at the origin scales linearly with  $L$ . Consequently, the photon efficiency of MINFLUX can be tuned by a geometrical parameter, whereas the ultimate limit of camera-based localization given in (1.3) depends on the emission wavelength.

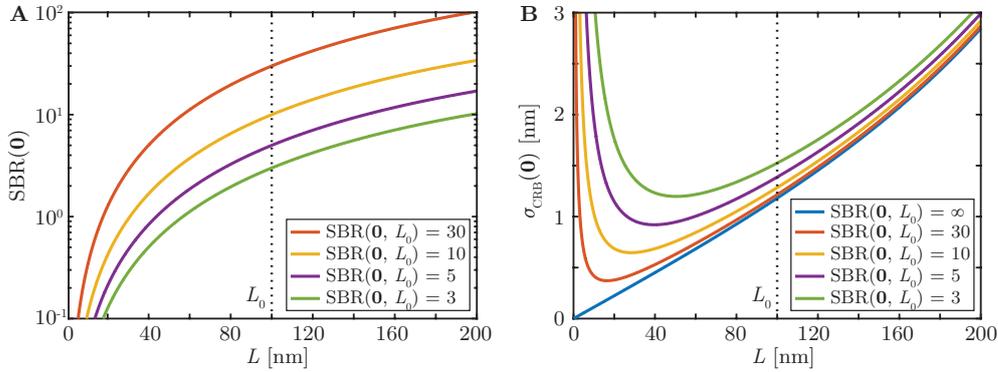
In the presence of background, the expression for the CRB at the center is

extended by a factor that is dependent on the SBR [29, (S31)]:

$$\sigma_{\text{CRB}}(\mathbf{0}) = \frac{L}{2\sqrt{2N}} \frac{1}{1 - \frac{L^2 \ln 2}{\text{FWHM}^2}} \sqrt{\left(1 + \frac{1}{\text{SBR}(\mathbf{0}, L)}\right) \left(1 + \frac{3}{4\text{SBR}(\mathbf{0}, L)}\right)}, \quad (1.25)$$

where the definition of  $\mathbf{p}$  provided in (1.12) was used. Here, the SBR introduced in (1.11) depends on the beam separation  $L$ . This can be explained by the fact that the signal level drops with decreasing EBP diameter. The background contributions  $\lambda^{(b)}$ , however, are typically not affected by a beam displacement. This dependence is shown in the left panel of Figure 1.3. In order to provide a definition of the SBR that is independent of  $L$ , we can express  $\text{SBR}(\mathbf{0}, L)$  as a function of a known SBR for a reference EBP diameter  $L_0$  [29, (S32)]:

$$\text{SBR}(\mathbf{0}, L) = \frac{L^2}{L_0^2} \exp\left(\ln 2 \cdot \frac{L_0^2 - L^2}{\text{FWHM}^2}\right) \text{SBR}(\mathbf{0}, L_0). \quad (1.26)$$



**Figure 1.3: Influence of background on the CRB.** **A**, SBR at the origin as a function of the diameter  $L$  of the four-doughnut EBP. The curves are shown for different SBRs defined at the reference diameter  $L_0 = 100$  nm and a doughnut profile with  $\text{FWHM} = 360$  nm. **B**, Dependence of the CRB at the origin on the beam separation  $L$  for different reference SBRs and a total photon number  $N = 1000$ . Figure adapted from Gwosch [44].

The effect of the beam separation  $L$  on the CRB in the presence of background is visualized in the right panel of Figure 1.3. It becomes clear that depending on the background level, there is a minimum EBP diameter  $L$  that minimizes the CRB at the origin. In contrast to the background-free case, decreasing  $L$  below that value does not increase the photon efficiency any further.

### 1.3 Scope of this thesis

A significant bottleneck in super-resolution microscopy is the effective limitation on image resolution arising due to photobleaching of the fluorescent labels. MINFLUX addresses this problem by increasing the photon efficiency of the imaging process, obtaining a higher resolution than other methods while using fewer photons. In the MINFLUX concept, the excitation light pattern encodes spatial information into the measurement. This information is ultimately retrieved by analyzing the resulting fluorescence photon counts in post-processing. The objective of this thesis is to maximize the information content of MINFLUX images. To this end, we first want to enhance the information contained in the emitted photons by optimizing the excitation pattern, which will be elaborated in Chapter 2. Next, we aim to utilize this information in the optimal way by integrating new concepts into the post-processing workflow, as presented in Chapter 3.

## **2 Lissajous illumination patterns for MINFLUX nanoscopy**

This chapter studies the properties of Lissajous figures used as illumination patterns for 2D MINFLUX. Section 2.1 motivates why Lissajous curves could be an alternative to the currently employed triangular arrangement [29]. Section 2.2 examines different geometric parameters of Lissajous EBPs, which are optimized for various imaging conditions in Section 2.3. Section 2.4 compares these patterns to the existing one as a function of the experimental parameter settings. The results of this chapter are discussed in Section 2.5.

### **2.1 Introduction**

This section gives an overview of different beam-scanning techniques and their implications for the illumination pattern. Based on that, the scope of this chapter is formulated.

### 2.1.1 Beam positioning in MINFLUX nanoscopy

In previously published experimental MINFLUX implementations, the excitation beam was laterally moved within the illumination pattern by electro-optic deflectors (EODs), which enable precise and nearly instantaneous beam positioning [29, 43]. The stabilization time of EODs is dominated by the bandwidth of the driver electronics and was on the order of 5 to 13  $\mu\text{s}$  depending on the operation mode [30]. As EODs have a limited scan range of around 0.4 to 2  $\mu\text{m}$ , a second scanning system for selecting the field of view inside the sample is required. This was achieved with either a piezo-based tip-tilt mirror [29, 30] or a galvanometric scanning system [43]. Overall, using EODs considerably increases the complexity and the cost of optical systems.

Galvanometric scanning units are standard components of many microscopes and have the advantages of a large scan range and a reasonably high scan frequency of a few kHz. Consequently, a single beam-deflection system could be employed for both selecting the field of view and rapidly drawing the EBP around the emitter. This would be a cheaper and notably simpler alternative to the existing MINFLUX implementations. Due to the nature of galvanometric scanners, these devices are typically used for steering the beam in a continuous movement with fixed scan frequencies instead of jumping from position to position. Thus, replacing the EODs by galvanometric deflectors would require an EBP where all exposures lie on a closed path. Excitation could occur along multiple curve segments of defined length, for each of which the collected photons could be assigned a fixed exposure point located within this section.

In order to fulfill the requirement of continuous scanning, the triangular pattern shown in Figure 1.1, which is the current standard for 2D MINFLUX [29, 48], could for example be described by a trifolium, a curve

resembling a three-leaved clover. Besides, Gwosch conducted a brief investigation into Lissajous curves as alternative EBPs and demonstrated that high-order curves may increase the CRB isotropy [44]. However, all these types of patterns have the drawback of taking a relatively long time to iterate through the illumination positions for a given scan frequency.

### 2.1.2 Scope of this chapter

The aim of this subproject is to examine the suitability of low-order Lissajous figures for 2D MINFLUX illumination. To this end, the CRB and its isotropy are employed as metrics for comparing these patterns to the existing triangular EBP as a function of the background level and the size of the region of interest (ROI). In the following sections I examine how certain geometrical parameters of Lissajous-path EBPs affect the theoretically achievable localization precision, and I optimize these parameters to minimize the CRB in the ROI. Particular attention is devoted to the question whether excitation beam positions arranged on a low-order Lissajous curve achieve a higher CRB isotropy than the currently used EBP. For simplicity, stationary conditions are assumed for all calculations, that is, beam displacements during the exposures are not taken into account.

## 2.2 Parameters of Lissajous illumination patterns

Lissajous figures are curves in the  $xy$ -plane generated by the functions

$$\begin{aligned} x &= a \sin(\omega_x t + q_x), \\ y &= b \sin(\omega_y t + q_y), \end{aligned} \tag{2.1}$$

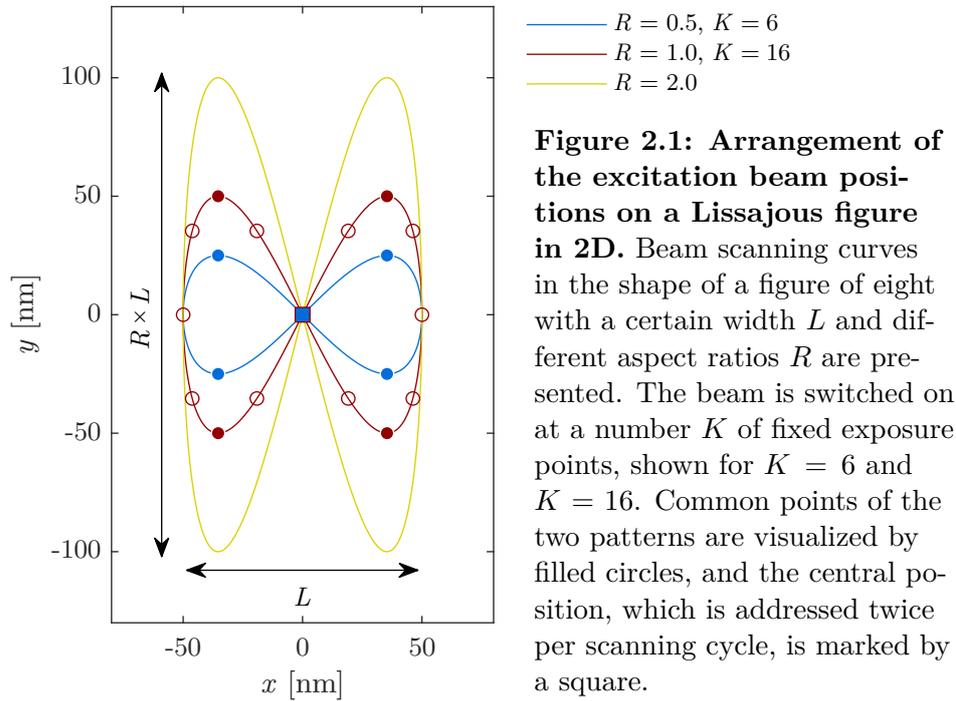
where the frequencies  $w_x$  and  $w_y$  are small integer multiples of a base frequency  $\omega$  [49]. In order to minimize the EBP iteration time at a given scan frequency, we decided to use the lowest-order 2D Lissajous figure that passes through the origin, which is a figure of eight. This curve is generated by setting the frequencies to  $\omega_x = \omega$  and  $\omega_y = 2\omega$  and the phase shifts to  $q_x = q_y = 0$ . Lissajous-path EBPs have a few other geometrical parameters, namely the amplitudes  $a$  and  $b$  of the figure, and the number  $K$  of illumination positions along the curve. In order to make these patterns comparable to the existing triangular MINFLUX EBP, I redefined the amplitudes by the width  $L$  and aspect ratio  $R$  of the Lissajous figure as shown in Figure 2.1:

$$\begin{aligned} a &= L/2, \\ b &= L/2 \cdot R. \end{aligned} \tag{2.2}$$

The  $K$  excitation beam positions can be distributed along the Lissajous curve in many ways. In the special case of equal angular spacing, the positions are calculated according to

$$\mathbf{r}_{\mathbf{b},i} = \frac{L}{2} \cdot \left( \sin\left(\frac{i}{n_x} \cdot 2\pi\right), R \cdot \sin\left(\frac{i}{n_y} \cdot 2\pi\right) \right), \tag{2.3}$$

with  $i \in [0, 1, \dots, K - 1]$  and  $n_{x,y} = \frac{K\omega}{\omega_{x,y}}$ .



This section will elaborate on the parameter choices for the optimization following in Section 2.3. Moreover, it will be demonstrated how the number  $K$  of exposures as well as the width  $L$  and aspect ratio  $R$  of the Lissajous figure affect the CRB on the localization precision that can be obtained with this excitation pattern.

### 2.2.1 Number of exposures

This subsection examines how many exposures are reasonable for MIN-FLUX excitation with a Lissajous EBP and should hence be further analyzed in the following optimization. In principle any number  $K$  of points can be distributed on a figure of eight. However, several criteria should be

taken into account when choosing the exposure points:

1. In order to provide a 2D pattern that is symmetric about both the  $x$ - and  $y$ -axes,  $K$  has to be an even number above 5. Note that for any even number of points equally distributed along a figure of eight, two points lie in the center of the curve.
2. Previous findings suggest that a homogeneous coverage of the ROI with excitation positions increases the isotropy of the CRB [44].
3. The excitation should ideally occur at the points of minimum scanning velocity to minimize the beam displacement during the exposure time. As shown in Figure A.1, these points approximately coincide with the turning points of the scanning pattern along the fast axis, which is the  $y$ -axis in this case.
4. High numbers of photon-count inputs pose a particular challenge to the experiment-control software and electronic hardware. Indeed, existing MINFLUX implementations are limited to eight input values [29, 30]. Though a different technical realization would be possible, our goal was to keep the number  $K$  of excitation positions as small as possible.

In this chapter I compare Lissajous EBPs with different numbers of exposure points. Patterns with  $K = 6$  and  $K = 16$  are shown in Figure 2.1 as blue and red circles, respectively. For  $K = 16$ , on the one hand, the beam positions are distributed relatively evenly throughout the ROI, covering both the periphery and the central region. Additionally, 4 out of 16 positions coincide with the turning points of the scanning pattern in  $y$ -direction when keeping the angular spacing of the positions uniform for simplicity. Those points are visualized by the filled circles.  $K = 6$ , on the other hand, is the minimum number of beam positions that meets the crite-

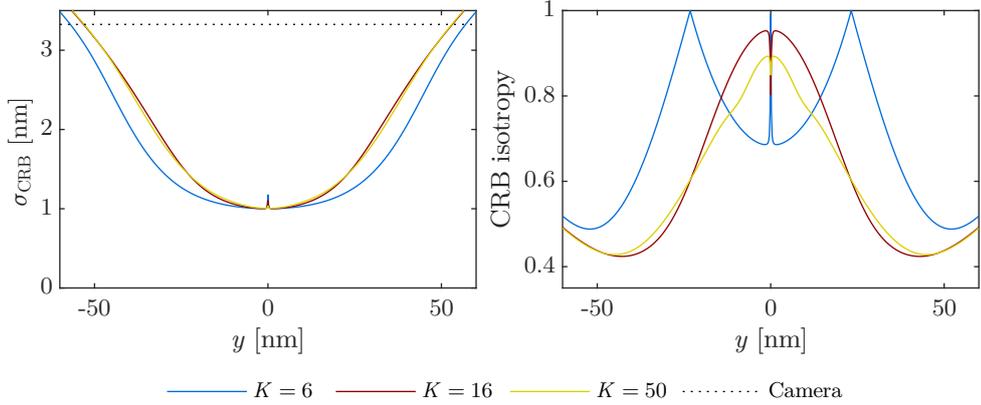
ria named above. For this pattern, I moved the four peripheral exposures to the turning points of the  $y$ -curve instead of distributing them evenly along the path. This makes the pattern a subset of the 16-point geometry and allows us to directly study the effect of the additional ten exposures on the CRB. The  $K = 6$  beam centers can be calculated according to

$$\mathbf{r}_{\mathbf{b},i} = \begin{cases} \mathbf{0} & \text{if } i \in [0, 3], \\ \frac{L}{2} \cdot \left( \sin\left(\frac{2i-1}{8} \cdot 2\pi\right), R \cdot \sin\left(\frac{2i-1}{4} \cdot 2\pi\right) \right) & \text{if } i \in [1, 2], \\ \frac{L}{2} \cdot \left( \sin\left(\frac{2i-3}{8} \cdot 2\pi\right), R \cdot \sin\left(\frac{2i-3}{4} \cdot 2\pi\right) \right) & \text{if } i \in [4, 5]. \end{cases} \quad (2.4)$$

The Lissajous geometry with  $K = 6$  exposure points resembles the quadratic EBP featuring an additional central exposure with increased relative intensity studied by Gwosch [44, Figures 2.7–2.9L]. However, there are two important differences between the two patterns: The Lissajous figure includes two exposures at the origin, which coincide in space but are temporally independent. Moreover, all six exposures have equal intensities. For the quadratic EBP, in contrast, the peripheral intensities were reduced relative to the single central exposure. Thus, the ratio between the total intensity applied in the periphery and at the origin was kept independent of the number of exposures [44, (2.69)].

For all results presented in this chapter, the following parameters were used in the calculations: a doughnut-shaped intensity profile as defined in (1.21) with FWHM = 360 nm and a total number of detected photons  $N = 1000$  if not stated otherwise. The given values for  $\text{SBR}(\mathbf{0}, L_0, R_0)$  were defined for a reference EBP width  $L_0 = 100$  nm and aspect ratio  $R_0 = 1.0$ , and are termed  $\text{SBR}_0$  for simplicity. The CRB of camera-based localization was calculated as previously described [29, (S63)], using the following parameters: a detection PSF of width  $\sigma_{\text{PSF}} = 100$  nm, a camera ROI containing  $K = 81$  pixels, each with a size of  $a = 100$  nm, and  $\text{SBR}_c = 500$ .

Figure 2.2 shows how the number  $K$  of exposures influences the two performance metrics selected for studying the EBPs, namely  $\sigma_{\text{CRB}}$  and the CRB isotropy  $\mathbb{I}$ , in the absence of background.  $\sigma_{\text{CRB}}$  and  $\mathbb{I}$  were defined in (1.19) and (1.20), respectively. In order to assess the potential benefits of further increasing  $K$ , the previously introduced EBPs with  $K = 6$  and  $K = 16$  were compared to a third pattern with  $K = 50$ . Within the area enclosed by the EBP, the CRB obtained with all three geometries lies below the CRB of camera-based localization. The  $\sigma_{\text{CRB}}$  and isotropy profiles along the  $x$ -axis, shown in Figure A.2, look identical to the ones along the  $y$ -axis for  $K = 6$ . This can be explained by the fact that the EBP aspect ratio was set to  $R = \frac{\sqrt{2}}{2} \approx 0.7$ , which is the value producing a square for  $K = 6$  based on the beam positions given in (2.4). For  $K = 16$  and  $K = 50$ , the profiles slightly differ between the  $x$ - and  $y$ -axes, which could indicate a suboptimal aspect ratio. Increasing the number of exposures above 16 seems to have only a minor effect on the CRB, so the following studies were limited to  $K = 6$  and  $K = 16$ .



**Figure 2.2: Influence of the number of exposures in the Lissajous pattern on the CRB and its isotropy.** Central profiles of  $\sigma_{\text{CRB}}$  and the isotropy  $\mathbb{I}$  along the  $y$ -axis for different numbers  $K$  of exposures. The CRB of camera-based localization is visualized by the dotted, black line. Further parameters: Pattern width  $L = 100$  nm, aspect ratio  $R = 0.7$ , and  $\text{SBR}_0 = 1 \times 10^5$ .

It is worth mentioning that the apparent discontinuities at the beam posi-

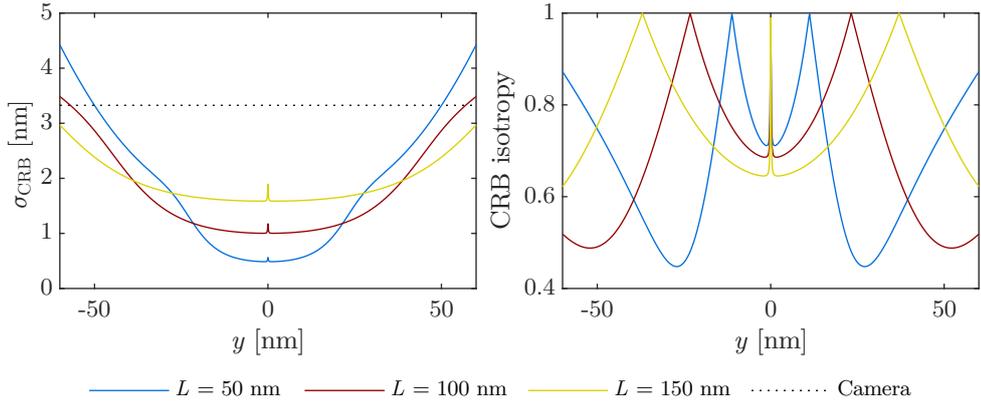
tions are properties of the Fisher information and thus of the CRB. Indeed, the minimum of  $\sigma_{\text{CRB}}$  is not at the position of the central exposure but in its close vicinity. This is due to the fact that each exposure with a doughnut-shaped beam adds information to the Fisher information except at the position of its zero. The effect decreases for increasing background. For infinite SBR the Fisher information is ill-defined at the origin [44]. Therefore, a finite, but sufficiently large  $\text{SBR}_0 = 1 \times 10^5$  was used whenever results were calculated “without background”.

### 2.2.2 Width of the Lissajous figure

The width  $L$  of the figure of eight is the geometrical parameter that probably has the largest effect on the achievable localization precision. As displayed in (1.24) for a four-doughnut EBP with negligible background, the CRB at the origin scales approximately linearly with the EBP width  $L$  for small  $L$ . However, the beam separation does not only tune the localization precision but also defines the field of view. As a matter of fact, the size of the region providing a high localization precision increases with  $L$ , which is illustrated in Figure 2.3. The profiles of  $\sigma_{\text{CRB}}$  and  $\mathbb{I}$  along the  $x$ -axis are identical to the ones shown here.

I would further like to point out that for the shown case of a Lissajous figure with  $K = 6$  exposures, the width  $L$  of the Lissajous path does not equal the actual distance between the peripheral exposures, which can be seen in Figure 2.1. From (2.4) one can derive a relationship between  $L$  and the extent  $L_{\text{rect}}$  of the rectangle formed by the outer illumination positions along the  $x$ -axis, given by

$$L = L_{\text{rect}}\sqrt{2}. \tag{2.5}$$



**Figure 2.3: Influence of the width of the Lissajous pattern on the CRB and its isotropy.** Central profiles of  $\sigma_{\text{CRB}}$  and the isotropy  $\mathbb{I}$  along the  $y$ -axis for different pattern widths  $L$ . The CRB of camera-based localization is visualized by the dotted, black line. Further parameters:  $K = 6$  exposures, pattern aspect ratio  $R = 0.7$ , and  $\text{SBR}_0 = 1 \times 10^5$ .

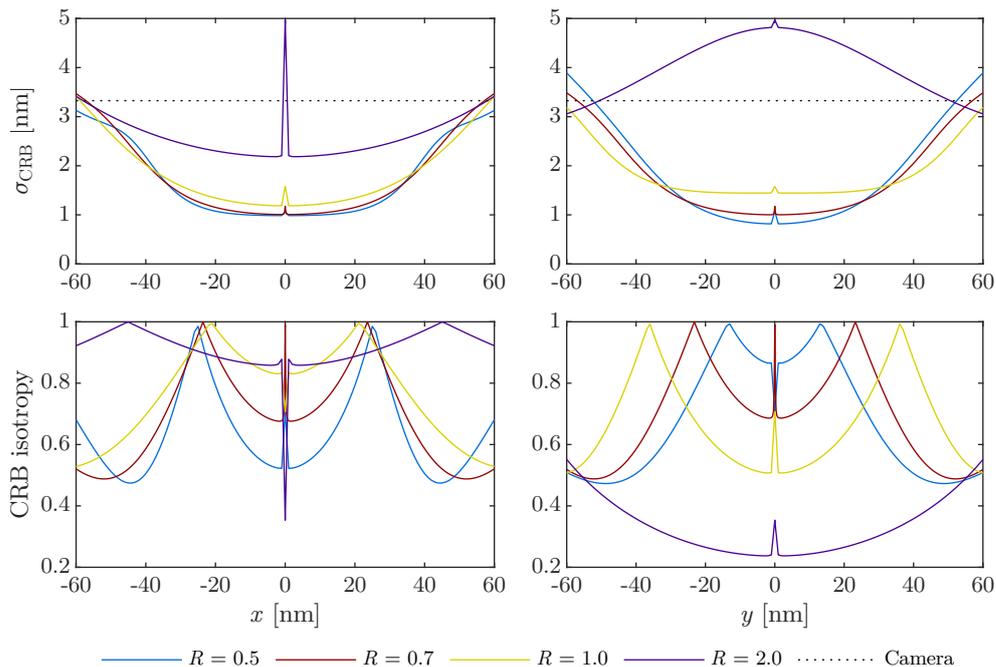
### 2.2.3 Aspect ratio of the Lissajous figure

The third parameter that describes the geometry of the Lissajous-path EBP is the aspect ratio  $R$  of the figure of eight. It determines how evenly the beam positions are distributed around the center. Therefore, it has a considerable effect on the uniformity of  $\sigma_{\text{CRB}}$  and  $\mathbb{I}$  throughout the field of view. This becomes evident in Figure 2.4 for Lissajous figures with  $K = 6$  points and different aspect ratios.

In analogy to the width  $L$ , also the aspect ratio  $R$  of the Lissajous curve differs from the aspect ratio  $R_{\text{rect}}$  of the rectangle defined by the four peripheral beam positions for  $K = 6$ , namely

$$R = \frac{R_{\text{rect}}}{\sqrt{2}}. \quad (2.6)$$

This proves that for a six-point arrangement with a maximum order of rotational symmetry, we expect the aspect ratio of the curve to take a



**Figure 2.4: Influence of the aspect ratio of the Lissajous pattern on the CRB and its isotropy.** Central profiles of  $\sigma_{\text{CRB}}$  and the isotropy  $\mathbb{I}$  along the  $x$ - and  $y$ -axes for different pattern aspect ratios  $R$ . The CRB of camera-based localization is visualized by the dotted, black lines. Further parameters:  $K = 6$  exposures, pattern width  $L = 100$  nm, and  $\text{SBR}_0 = 1 \times 10^5$ .

value of  $R = \frac{\sqrt{2}}{2} \approx 0.7$ .

Furthermore, the diameter of a circle connecting the four peripheral exposures of the six-point Lissajous EBP can be calculated from (2.5) and (2.6):

$$L_{\text{circle}} = L_{\text{rect}} \sqrt{1 + R_{\text{rect}}^2} = L \sqrt{1/2 + R^2}, \quad (2.7)$$

which equals the width  $L$  of the Lissajous figure for  $R = \frac{\sqrt{2}}{2}$  and facilitates the comparison of this pattern to the triangular one defined by a circle as shown in Figure 1.1.

## 2.3 Optimization of the geometrical parameters

As demonstrated in the previous section, the CRB on the achievable localization precision is not uniform throughout the field of view and depends on the size and the shape of the EBP. The optimal geometry of the EBP in turn is determined by the imaging conditions such as the SBR and the ROI diameter  $d_{\text{ROI}}$ . The latter is typically given by the uncertainty about the emitter position and can change in the course of the measurement, especially in iterative MINFLUX settings. In this section the width  $L$  and the aspect ratio  $R$  of Lissajous figures with  $K = 6$  and  $K = 16$  beam positions are optimized as a function of both  $d_{\text{ROI}}$  and the SBR.

The width  $L$  and aspect ratio  $R$  of the figure of eight were optimized by finding the minimum of a CRB-related loss function  $f_{\text{loss}}$ . The choice of the loss function may have a substantial effect on the outcome and the practical relevance of the optimization. A straightforward success metric could be the CRB at the center of the EBP  $\sigma_{\text{CRB}}(\mathbf{0})$ . This metric would have high relevance if the EBP was expected to be always perfectly centered on the emitting molecule, which is, however, generally not the case. Assuming that the emitter is located somewhere in a circular region of diameter  $d_{\text{ROI}}$  in reality, one should rather pick a loss function that takes all CRB values within this ROI into account. This ensures localization with high precision throughout the whole ROI.

There are numerous success metrics that consider the CRB in a certain ROI. Previously the average localization CRB  $\sigma_{\text{CRB}}$  over the edge of a ROI was applied for optimizing the EBP diameter  $L$  in a tracking context [48]. Considering the fact that  $\sigma_{\text{CRB}}$  typically increases with distance from the center, that loss function optimizes the “worst case”, which is localizing a molecule on the edge of the ROI. For MINFLUX imaging applications I

chose the loss function

$$f_{\text{loss}} = \bar{\sigma}_{\text{CRB}}(\text{ROI}), \quad (2.8)$$

which averages the square-root arithmetic mean of the eigenvalues  $\{\sigma_i^2(\mathbf{r})\}$  of  $\Sigma_{\text{CRB}}(\mathbf{r})$  over the entire ROI.

Given our particular interest in maximizing the CRB isotropy, all the optimizations that are presented in this section were performed with a second loss function

$$f_{\text{loss}} = \bar{\sigma}_{\text{max,CRB}}(\text{ROI}). \quad (2.9)$$

Here, only the square root of the largest eigenvalue  $\sigma_{\text{max}}^2(\mathbf{r})$  of  $\Sigma_{\text{CRB}}(\mathbf{r})$ , that is, the spatial dimension with the highest localization uncertainty, is averaged over the ROI. Thus, the goal of using this loss function was to ensure high localization precision in all directions and throughout the whole ROI. However, the optimization results were almost identical to the ones yielded with  $f_{\text{loss}} = \bar{\sigma}_{\text{CRB}}(\text{ROI})$  and are therefore not included in this section. A visual comparison of the two approaches can be found in Figure A.3. The results obtained from the optimization with the second loss function are shown in Figures A.5–A.7, A.10, and A.11.

### 2.3.1 Optimal width of the Lissajous figure

The width  $L$  of the Lissajous figure was optimized together with its aspect ratio  $R$  in a 2D grid search for the minimum of the loss function  $f_{\text{loss}}$ . Following previously presented work [48], the optimization was first conducted as a function of the ROI diameter  $d_{\text{ROI}}$ . For 20  $d_{\text{ROI}}$  values between 1 nm and 100 nm, the grid contained discrete  $L$  values in a slightly bigger range between 1 nm and 300 nm. Two different Lissajous patterns with

$K = 6$  and  $K = 16$  exposures, respectively, were optimized by minimizing  $f_{\text{loss}} = \bar{\sigma}_{\text{CRB}}(\text{ROI})$ , both without background and with a finite  $\text{SBR}_0 = 12$ . This value is realistic for experiments with single fluorescent dye molecules. The values for  $d_{\text{ROI}}$  and  $L$  followed a logarithmic distribution, which resulted in a grid spacing of 0.03 nm to 8.43 nm in the case of  $L$ . All the optimization parameters are listed in Table 2.1.

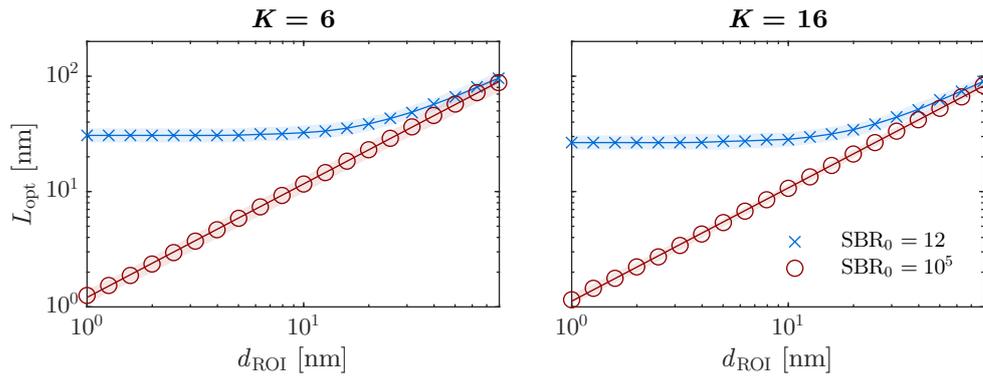
Parameter	Values		
$f_{\text{loss}}$	$\bar{\sigma}_{\text{CRB}}(\text{ROI})$		
$K$	6, 16		
$d_{\text{ROI}}$	Selected values between 1 nm and 100 nm		
$\text{SBR}_0$	Selected values between 1 and 100 000		
	Minimum	Maximum	Number of grid samples
$L$	1 nm	300 nm	201, logarithmic spacing
$R$	0	2	401, linear spacing

**Table 2.1:** Parameters for optimizing width and aspect ratio of the Lissajous figure in a 2D grid search.

The optimal width  $L_{\text{opt}}$  of the Lissajous figure resulting from the 2D grid search is shown in Figure 2.5 as a function of the ROI diameter  $d_{\text{ROI}}$  for various parameter combinations. In the absence of background, the optimal width increases linearly with the size of the ROI according to

$$L_{\text{opt}} \approx \begin{cases} 1.1d_{\text{ROI}} + 0.3 \text{ nm} & \text{if } K = 6, \\ 1.0d_{\text{ROI}} + 0.2 \text{ nm} & \text{if } K = 16. \end{cases} \quad (2.10)$$

For  $\text{SBR}_0 = 12$  and small ROIs, the optimal  $L$  value converges to  $L_{\text{opt}} \approx 31$  nm for  $K = 6$  and to  $L_{\text{opt}} \approx 27$  nm for  $K = 16$ . This is due to the fact that in the presence of background, reducing the EBP width improves the CRB only up to a minimum  $L$  value, as shown in Figure 1.3. As indicated by the shaded areas in Figure 2.5, values 12% above or below the respective optimal width  $L_{\text{opt}}$  ensure  $f_{\text{loss}}(L)$  to stay within 1% above the minimum

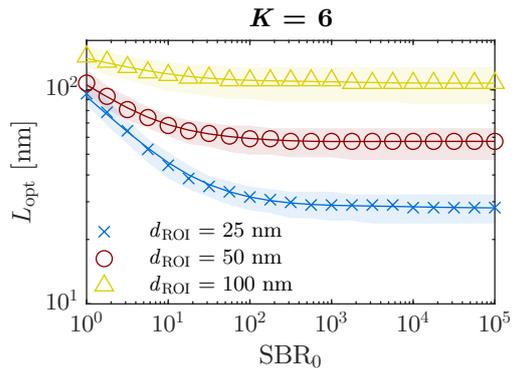


**Figure 2.5: Optimal width of the Lissajous pattern as a function of the ROI size.** The width of the Lissajous figure was optimized with and without background for two different numbers  $K$  of exposures. The optimal width  $L_{\text{opt}}$  minimizes the value of the loss function  $f_{\text{loss}} = \bar{\sigma}_{\text{CRB}}(\text{ROI})$  in a circular ROI of diameter  $d_{\text{ROI}}$ . In this figure and the following ones, a cubic smoothing spline interpolation is shown in addition to the data points for each set. The narrow, shaded areas around the lines show the tolerance ranges for a deviation of 1% from the minimum loss.

on average.

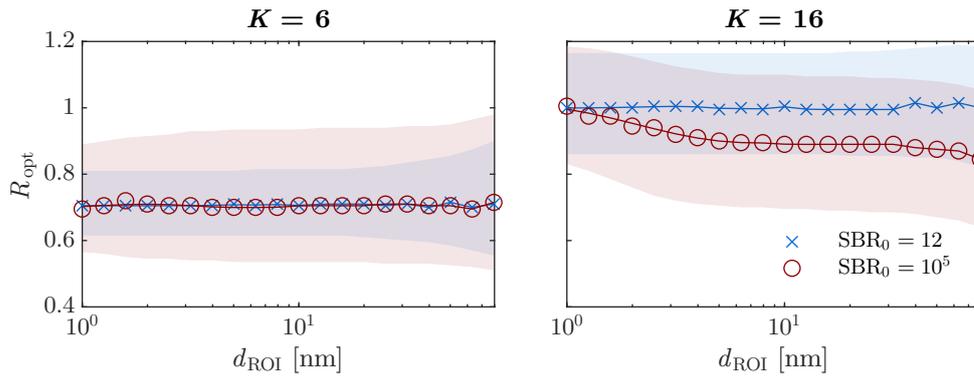
In order to further examine the effect of background on the optimal width of the Lissajous figure,  $L$  was optimized together with  $R$  as a function of the SBR for three different ROI sizes. This time, the 2D grid search was performed for 20 logarithmically spaced SBR values between 1 and 100 000. The results are presented in Figure 2.6, where the findings of Figure 2.5 are confirmed: For large ROIs the optimal width of the Lissajous figure is almost independent of the SBR. The smaller the ROI, the bigger the influence of background on  $L_{\text{opt}}$  gets. With decreasing SBR, the optimal width of the Lissajous figure for a fixed ROI size increases. In the limit of a low SBR,  $L_{\text{opt}}$  becomes nearly independent of  $d_{\text{ROI}}$ . As visualized by the shaded areas,  $L$  has a tolerance range of  $\pm 13\%$  on average.

**Figure 2.6: Optimal width of the Lissajous pattern as a function of the SBR.** Analogous to Figure 2.5. Here, the width of the Lissajous figure was optimized for ROIs of three different diameters  $d_{\text{ROI}}$  and for  $K = 6$  exposures.



### 2.3.2 Optimal aspect ratio of the Lissajous figure

Along with the width of the Lissajous figure, I also optimized its aspect ratio  $R$ . The 2D grid on which the loss function was minimized contained  $R$  values between 0 and 2 as given in Table 2.1. Figure 2.7 presents the optimal aspect ratio  $R_{\text{opt}}$  of the Lissajous figure as a function of the ROI size for different numbers of exposures and SBRs. For patterns with  $K = 6$  illumination positions,  $R_{\text{opt}}$  takes an approximately constant value of 0.7, which yields a quadratic EBP according to (2.6). This confirms our expectation of a geometry with four-fold rotational symmetry as mentioned in Section 2.2.3. For  $K = 16$  and  $\text{SBR}_0 = 12$ , the optimal value is  $R_{\text{opt}} \approx 1.0$  and independent of  $d_{\text{ROI}}$ . In the absence of background, however, the ideal aspect ratio decreases from 1.0 to 0.8 with increasing ROI diameter, which cannot be easily explained. A different loss function  $f_{\text{loss}} = \bar{\sigma}_{\text{max,CRB}}(\text{ROI})$  yielded a more or less fixed optimal aspect ratio  $R_{\text{opt}} \approx 1.0$ , which is visualized in Figure A.7. In summary, the optimization results suggest that except for  $K = 16$  and  $f_{\text{loss}} = \bar{\sigma}_{\text{CRB}}(\text{ROI})$ ,  $R_{\text{opt}}$  depends neither on the size of the ROI nor on the SBR, but only on the number  $K$  of illumination positions. This is confirmed in Figure A.8, where  $R_{\text{opt}}$  is visualized as a function of the SBR. 2D representations of the optimal excitation patterns resulting from the grid search are shown in Figure A.9 for selected parameter settings.

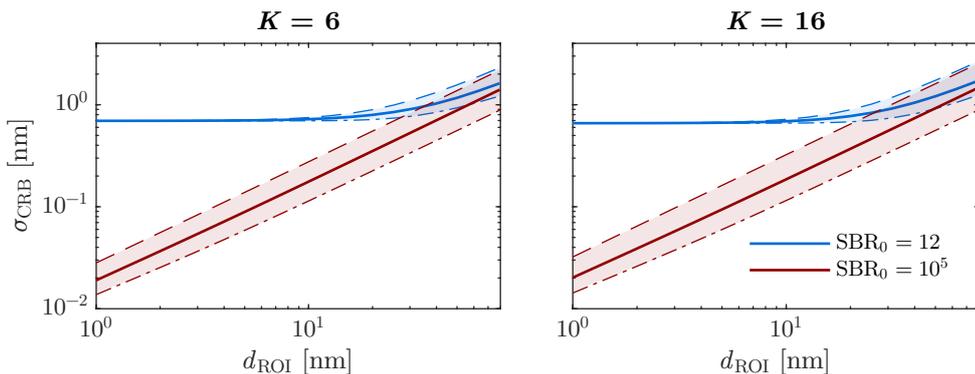


**Figure 2.7: Optimal aspect ratio of the Lissajous pattern as a function of the ROI size.** Analogous to Figure 2.5. The aspect ratio of the Lissajous figure was optimized with and without background for two different numbers  $K$  of exposures.

The shaded tolerance ranges in Figure 2.7 indicate that, compared to the results obtained for the size parameter  $L$ , the loss function is slightly less sensitive to deviations in the aspect ratio: On average,  $R$  values 20% above or below the ideal aspect ratio still yield a loss function value that is within 1% above its minimum. However, this sensitivity varies depending on the number of excitation positions, the SBR, and the ROI diameter. In the absence of background, the CRB is less affected by deviations from  $R_{\text{opt}}$  than for a limited SBR. Moreover, the impact of aspect-ratio changes on the CRB becomes smaller on larger ROIs, especially for  $K = 6$ . Figure A.7 demonstrates that in comparison to  $f_{\text{loss}} = \bar{\sigma}_{\text{CRB}}(\text{ROI})$ , the alternative loss function  $f_{\text{loss}} = \bar{\sigma}_{\text{max,CRB}}(\text{ROI})$  is much more sensitive to deviations from the optimal aspect ratio.

## 2.4 Localization precision of the optimized Lissajous patterns

Having found the optimal values for the width and aspect ratio of the Lissajous figures, the localization precision achievable with these patterns was quantified. Figure 2.8 displays the values of the loss function that was minimized using the 2D grid search described in the previous section,  $f_{\text{loss}}(L_{\text{opt}}(d_{\text{ROI}}), R_{\text{opt}}(d_{\text{ROI}})) = \bar{\sigma}_{\text{CRB}}(\text{ROI})$ , as a function of the ROI diameter  $d_{\text{ROI}}$ .  $\bar{\sigma}_{\text{CRB}}(\text{ROI})$ , the average CRB in the ROI, is visualized by the thick lines and shows a linear dependence on  $d_{\text{ROI}}$  in the absence of background. The shaded bands visualize the full range of values that  $\sigma_{\text{CRB}}$  assumes inside the ROI. For  $\text{SBR}_0 = 12$  and small  $d_{\text{ROI}}$ ,  $\sigma_{\text{CRB}}$  is approximately constant throughout the ROI and converges to 0.70 nm and 0.66 nm for  $K = 6$  and  $K = 16$ , respectively.



**Figure 2.8: CRB obtained with the optimized Lissajous patterns as a function of the ROI size.** The CRB was calculated inside a circular ROI of variable diameter  $d_{\text{ROI}}$  with and without background for 20 different values of  $d_{\text{ROI}}$  and two numbers  $K$  of exposures. The shaded bands indicate the range of values that  $\sigma_{\text{CRB}}$  takes within the ROI. Dash-dotted, dashed, and continuous lines indicate cubic smoothing spline interpolations to the minimum, maximum, and mean values, respectively.

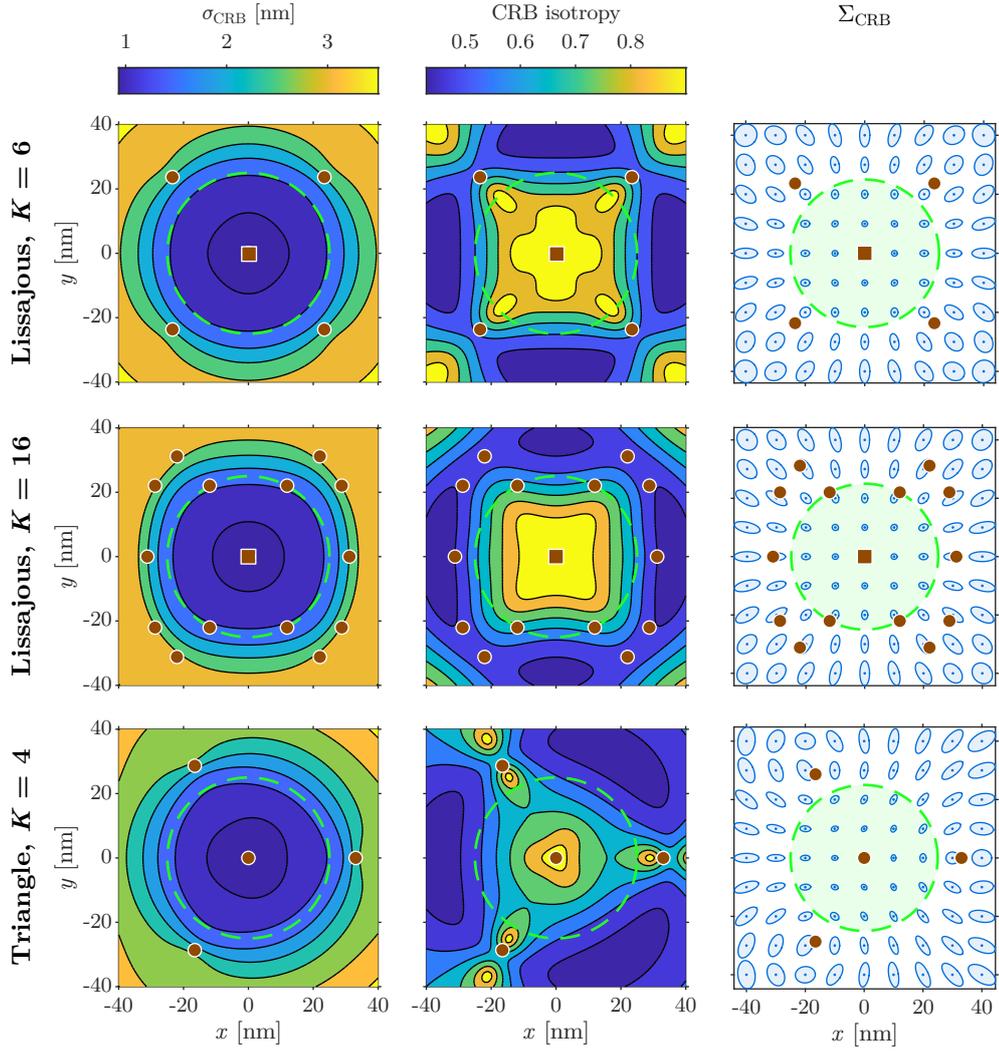
In the following, the Lissajous-path illumination patterns with six and 16 exposures will be compared to the triangular pattern depicted in Figure

1.1 under various imaging conditions. To conclude this section, an explicit expression for the CRB of the six-point Lissajous EBP will be derived.

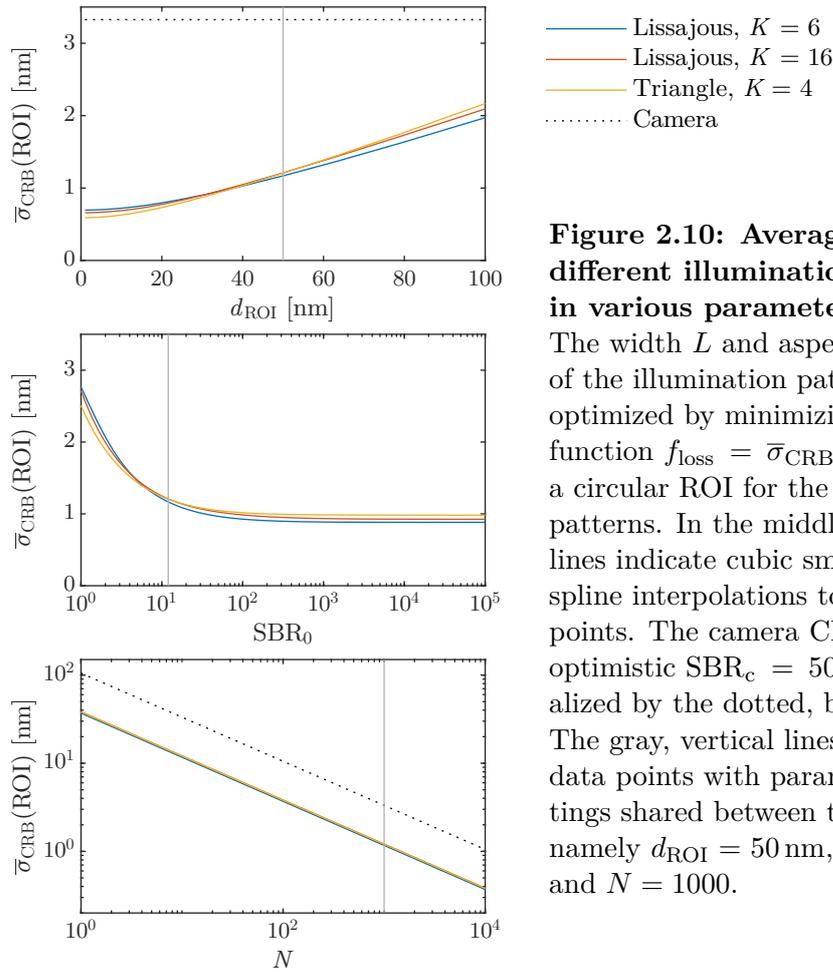
### 2.4.1 Comparison to the triangular pattern

Figure 2.9 compares the 2D distributions of the CRB and its isotropy obtained with the two optimized Lissajous EBPs and the triangular pattern. As demonstrated in Figures 2.5 and A.4, the ideal EBPs are slightly larger than the circular ROIs in which the CRB was minimized. It becomes evident that the CRB distributions of the patterns, visualized in the left column, are very similar. In all three cases  $\sigma_{\text{CRB}}$  is minimal close to the origin and stays below 1.5 nm throughout almost the entire ROI of diameter  $d_{\text{ROI}} = 50$  nm. In contrast, the different degrees of rotational symmetry of the EBPs become clearly noticeable in the 2D maps of the CRB isotropy  $\mathbb{I}$  presented in the central column. For the Lissajous figures,  $\mathbb{I}$  stays above 0.8 and 0.7 throughout almost the entire ROI for  $K = 6$  and  $K = 16$ , respectively. In the case of the triangular pattern, however, the isotropy exhibits a sharp peak at the origin and rapidly decreases down to 0.4 towards the edge of the ROI. The visualizations of  $\Sigma_{\text{CRB}}$  as ellipses of contour level  $e^{-1/2}$  in the right column support this observation.

Figure 2.10 compares the average CRB obtained with the two Lissajous patterns and the triangular geometry as a function of various imaging parameters, namely the ROI size  $d_{\text{ROI}}$ , the SBR, and the total number  $N$  of detected photons. For each data point,  $L$  and  $R$  were set to the corresponding optimal values. In general, the results of the three patterns are very similar under all conditions. Which of the EBPs ultimately achieves the smallest  $\bar{\sigma}_{\text{CRB}}(\text{ROI})$  value depends both on the diameter of the ROI and on the SBR.



**Figure 2.9: 2D maps of the CRB and its isotropy for different illumination patterns with optimal geometry.** The width and aspect ratio of the illumination patterns visualized by the brown circles were optimized by minimizing the loss function  $f_{\text{loss}} = \bar{\sigma}_{\text{CRB}}(\text{ROI})$  on the ROIs marked by the dashed, green circles. This resulted in the following geometries:  $L_{\text{opt}} = 66$  nm and  $R_{\text{opt}} = 0.7$  for  $K = 6$ ,  $L_{\text{opt}} = 62$  nm and  $R_{\text{opt}} = 1.0$  for  $K = 16$ , and  $L_{\text{opt}} = 66$  nm for  $K = 4$ . In the Lissajous patterns the central position marked by a brown square is addressed twice per scanning cycle. The left and central columns each have a fixed color scale for all three plots. In the right column,  $\Sigma_{\text{CRB}}$  is visualized as ellipses with major and minor axes of length  $2\sigma_{1,2}$ , where  $\sigma_{1,2}^2$  are the eigenvalues of  $\Sigma_{\text{CRB}}$ . Further parameters:  $d_{\text{ROI}} = 50$  nm and  $\text{SBR}_0 = 12$ .



**Figure 2.10: Average CRB for different illumination patterns in various parameter settings.**

The width  $L$  and aspect ratio  $R$  of the illumination patterns were optimized by minimizing the loss function  $f_{\text{loss}} = \bar{\sigma}_{\text{CRB}}(\text{ROI})$  on a circular ROI for the indicated patterns. In the middle panel the lines indicate cubic smoothing spline interpolations to the data points. The camera CRB for an optimistic  $\text{SBR}_c = 500$  is visualized by the dotted, black lines. The gray, vertical lines indicate data points with parameter settings shared between the plots, namely  $d_{\text{ROI}} = 50$  nm,  $\text{SBR}_0 = 12$ , and  $N = 1000$ .

The upper panel of Figure 2.10 shows  $\bar{\sigma}_{\text{CRB}}(\text{ROI})$  as a function of  $d_{\text{ROI}}$  for  $\text{SBR}_0 = 12$ . Below  $d_{\text{ROI}} \approx 35$  nm, the triangular pattern achieves a slightly higher precision than the figures of eight, whereas above that value the six-point Lissajous pattern outperforms the other geometries. For small ROIs,  $\bar{\sigma}_{\text{CRB}}(\text{ROI})$  converges to 0.6 nm for the triangular EBP, and to 0.7 nm for the Lissajous patterns. In the limit of a large ROI, a linear relationship emerges between  $d_{\text{ROI}}$  and  $\bar{\sigma}_{\text{CRB}}(\text{ROI})$ , just like in the background-free case. For all EBPs and ROI sizes, the average CRB in the ROI stays

far below the CRB obtained with an ideal camera assuming an optimistic SBR,  $\sigma_{\text{cam}} \approx 3.3$  nm.

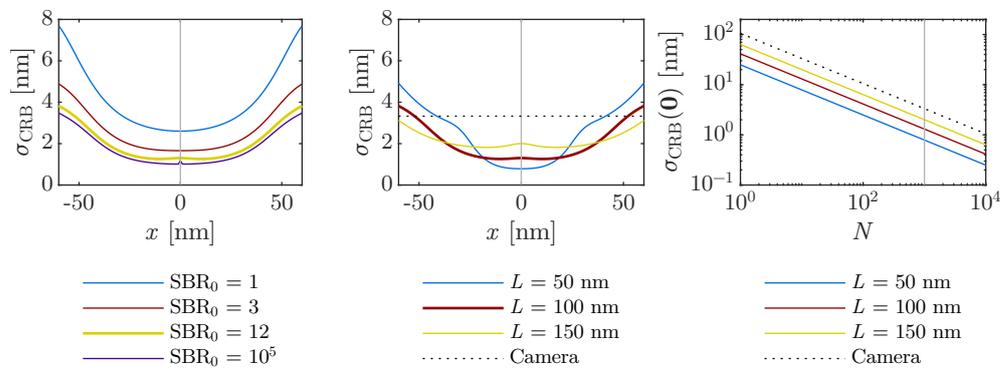
The middle panel exhibits the effect of the SBR on the CRB averaged over a ROI with a diameter of 50 nm. Below  $\text{SBR}_0 \approx 5$ , the triangular EBP yields a slightly smaller  $\bar{\sigma}_{\text{CRB}}(\text{ROI})$  than the Lissajous EBP, whereas above that value the six-point pattern outperforms the other geometries. For  $\text{SBR}_0 = 1$ , the average CRB reaches values between 2.5 nm for the triangular EBP and 2.8 nm for the six-point Lissajous geometry. In the limit of low background, for  $\text{SBR}_0 \gtrsim 150$ ,  $\bar{\sigma}_{\text{CRB}}(\text{ROI})$  converges to 0.9 nm for the Lissajous figures and to 1.0 nm for the triangular pattern.

In the lower panel the CRB dependence on the inverse square root of the total number of detected photons is displayed for  $\text{SBR}_0 = 12$ . The average CRB in the ROI using one of the EBPs is approximately three times as low as the CRB of camera-based localization for an optimistic SBR. For  $K = 6$ ,  $d_{\text{ROI}} = 50$  nm, and  $N = 10\,000$ ,  $\bar{\sigma}_{\text{CRB}}(\text{ROI})$  reaches down to 0.4 nm. To conclude, all panels underpin the previously made observation that the three patterns achieve very similar CRB values in the ROI.

### 2.4.2 Cramér–Rao bound of the six-point Lissajous pattern

The comparison of different EBPs in Figures 2.9 and 2.10 suggests that with both Lissajous patterns a similar localization precision can be achieved. However, as mentioned in Section 2.2, a lower number of exposures would have the advantage of being compatible with the current implementation of the experiment-control software and electronic hardware. Therefore, I further studied the CRB obtained with the optimized six-point Lissajous figure under various imaging conditions.

Figure 2.11 further examines the influence of the imaging parameters, namely the SBR, the EBP width  $L$ , and the total number  $N$  of detected photons, on the CRB of the optimized six-point Lissajous pattern. The left panel shows that for a fixed  $L$ , the CRB drops along with the background level. Moreover, the size of the area providing a low CRB slightly increases with the SBR. The middle panel indicates a proportionality between the EBP width and the CRB at the center for a fixed SBR. At the same time it demonstrates that the region of high localization precision becomes larger with increasing  $L$ . For  $\text{SBR}_0 = 12$  and Lissajous figures with a width between 50 and 150 nm,  $\sigma_{\text{CRB}}(\mathbf{0})$  ranges from 0.8 to 2.0 nm. These values are well below the CRB of camera-based localization assuming an optimistic SBR,  $\sigma_{\text{cam}} \approx 3.3$  nm, and consistent with previous results obtained for the triangular pattern [44, Figure 2.6C–D]. The right panel shows the dependence of  $\sigma_{\text{CRB}}(\mathbf{0})$  on the inverse square root of the total number  $N$  of detected photons, given in (2.12), using the same  $L$  values as in the middle panel. For  $L = 50$  nm and  $\text{SBR}_0 = 12$ , the CRB of the optimized six-point Lissajous figure at the origin is four times lower than the camera limit, reaching down to 0.2 nm for 10 000 photons.



**Figure 2.11: CRB obtained with the optimized six-point Lissajous pattern in various parameter settings.** The left and central panels present profiles of  $\sigma_{\text{CRB}}$  in  $x$ -direction, whereas the right panel shows the CRB at the origin. The camera CRB for an optimistic  $\text{SBR}_c = 500$  is visualized by the dotted, black lines. Parameters if not stated otherwise:  $L = 100$  nm,  $R = 0.7$ , and  $\text{SBR}_0 = 12$ .

In analogy to (1.25), an analytical expression can be derived for the CRB at the center of a Lissajous pattern with six exposures, taking background into account. Using the definitions of  $\sigma_{\text{CRB}}$  and of the multinomial parameter vector  $p$  given in (1.19) and (1.12), respectively, yields

$$\sigma_{\text{CRB}}(\mathbf{0}) = \frac{L}{8R\sqrt{2N}} \frac{(2R^2 + 1)^{3/2}}{\left| \frac{L^2 \ln 2}{\text{FWHM}^2} \frac{2R^2 + 1}{2} - 1 \right|} \sqrt{\left(1 + \frac{2/3}{\text{SBR}}\right) \left(1 + \frac{1}{\text{SBR}}\right)}, \quad (2.11)$$

with  $\text{SBR} := \text{SBR}(\mathbf{0}, L, R)$ . In analogy to (1.26), the SBR for a given EBP width  $L$  and aspect ratio  $R$  can be derived from a known  $\text{SBR}(\mathbf{0}, L_0, R_0)$  for a reference beam width  $L_0$  and aspect ratio  $R_0$ .

The expression for  $\sigma_{\text{CRB}}(\mathbf{0})$  can be simplified by setting the aspect ratio to  $R = \frac{\sqrt{2}}{2}$ , which yields a quadratic EBP according to (2.6) and is approximately equal to the optimum  $R_{\text{opt}} = 0.7$  obtained in Section 2.3.2. Furthermore assuming the EBP to be small compared to the size of the excitation doughnut beam, that is,  $L \ll \frac{\text{FWHM}}{\sqrt{\ln 2}}$ , results in

$$\sigma_{\text{CRB}}(\mathbf{0}) \approx \frac{L}{2\sqrt{2N}} \sqrt{\left(1 + \frac{2/3}{\text{SBR}}\right) \left(1 + \frac{1}{\text{SBR}}\right)}. \quad (2.12)$$

In the absence of background, (2.12) can be simplified to

$$\sigma_{\text{CRB}}^{(0)}(\mathbf{0}) = \frac{L}{2\sqrt{2N}}, \quad (2.13)$$

which equals the expression for the triangular EBP provided in (1.24). Thus, the linear dependence of the CRB at the origin on the width  $L$  of the illumination pattern, an inherent property of the MINFLUX concept, could be confirmed.

## 2.5 Summary and discussion

The aim of this chapter was to examine the suitability of low-order Lissajous figures for 2D MINFLUX illumination schemes that are compatible with galvanometric beam scanning systems. To this end, I optimized the width  $L$  and aspect ratio  $R$  of illumination patterns consisting of six or 16 exposure points distributed along a figure of eight. I compared the optimized geometries to the published 2D excitation pattern [29] regarding the CRB on the theoretically achievable localization precision. Moreover, I addressed the question whether the optimized Lissajous EBPs achieve a higher CRB isotropy than the existing triangular pattern. Finally, I derived an expression for the CRB at the origin of a six-point Lissajous figure.

The CRB values obtained for the Lissajous patterns in various parameter settings are comparable to previously shown results for the triangular pattern [44]. However, the optimization of the EBP width yielded a higher optimal value  $L_{\text{opt}}$  for a given ROI than previous studies [48]. This can be explained by the fact that different loss functions were employed for finding the best  $L$ . While I minimized the average CRB in the entire ROI, Eilers et al. only took the circular edge of the ROI into consideration. Close to the origin smaller EBPs produce a lower CRB, but this behavior changes outside the central region, as demonstrated in Figure 2.3. Therefore, the loss function values obtained in the present thesis cannot be directly compared to the published ones. Using yet another loss function, which also averaged the CRB over the whole ROI but only took the direction of maximum localization uncertainty into account, did not have any significant effect on the results of the optimization.

The newly introduced aspect ratio assumes a single optimal value  $R_{\text{opt}} \approx 0.7$  for  $K = 6$ , which results in a quadratic beam arrangement. In contrast,

$R_{\text{opt}}$  depends on the size of the ROI, the SBR, and the choice of the loss function in the case of  $K = 16$ . This could be due to the fact that for  $K = 16$  an EBP with a lower degree of rotational symmetry is combined with a perfectly symmetric beam profile. However, the loss function  $\bar{\sigma}_{\text{CRB}}(\text{ROI})$  has a broad minimum in the parameter space spanned by  $L$  and  $R$ . Thus, simply using a fixed aspect ratio  $R = 1.0$ , the optimal value obtained for  $\text{SBR}_0 = 12$ , for the 16-point Lissajous figure should still ensure a high localization precision.

A comparison of the four-point triangular EBP and the optimized Lissajous patterns with six and 16 exposures revealed similar 2D CRB distributions for all three geometries. The degree of rotational symmetry of the CRB and its isotropy in the  $xy$ -plane increases with the number of exposures, and, most importantly, both Lissajous EBPs achieve a significantly more isotropic CRB than the triangular pattern. This is consistent with earlier findings obtained for high-order Lissajous curves [44].

Which of the three illumination geometries ultimately minimizes  $\sigma_{\text{CRB}}$  in the ROI depends on the imaging conditions, but the differences between the patterns are marginal. Still, the presented Lissajous EBPs are promising options due to their high CRB isotropy. With its low number of exposures, the six-point Lissajous EBP could be easily implemented using the current design of the hardware control. In combination with galvanometric scanners, this could be a cheaper and simpler alternative to the currently employed beam positioning system.

## 3 Deep-learning-based data analysis for MINFLUX nanoscopy

This chapter presents a novel approach to position estimation in MINFLUX nanoscopy based on deep learning. Section 3.1 elaborates on the motivation for this approach and argues why deep learning could be useful for MINFLUX post-processing. Section 3.2 summarizes the theoretical fundamentals of MINFLUX position estimation and neural networks, as well as relevant methods. Section 3.3 describes strategies for designing, training and evaluating a simple neural network used for predicting 3D positions from simulated photon counts. This preliminary network is further tuned in Section 3.4 in order to minimize both localization error and training time. Finally, the deep-learning approach is compared to the existing position estimator, and its localization performance on experimental data is assessed in Section 3.5. All results presented in this chapter are summarized and discussed in Section 3.6.

### 3.1 Introduction

This section first provides a recapitulation of the established post-processing workflow in MINFLUX nanoscopy and an overview of existing deep-learning applications in single-molecule localization microscopy (SMLM). Based on this introduction, the scope of this chapter is formulated.

### 3.1.1 Post-processing in MINFLUX nanoscopy

Extracting the positions of fluorescent molecules from recorded photon-count traces typically requires three post-processing steps: trace segmentation, position estimation, and localization filtering [30].

In the segmentation step, photon emissions from single molecules can be distinguished from background using a hidden Markov model [29]. Each time point of the trace segment acquired in the final iteration of the MINFLUX cycle is assigned one of the three molecular emission states *on*, *off*, and *blinking*. Successive emissions in the on or blinking states are merged and the resulting events are then split at a predefined total photon number  $N$  in order to obtain several localizations per molecule [30].

Next, the photon counts are accumulated for each event and the position of the fluorescent emitter is estimated based on the resulting count combination. The CRB introduced in (1.19) gives a theoretical limit on the localization precision achievable with any unbiased estimator. The actual localization performance, however, strongly depends on the utilized algorithm. Precise position estimation in post-processing is currently accomplished by an MLE<sup>1</sup>, whose performance converges to the CRB for large photon numbers  $N$ . More precisely, this is the case for  $N \gtrsim 500$  under typical conditions for measuring in cellular environments [50, Figure 3.5]. In the low-photon regime, though, the MLE exhibits a strong bias. Therefore, a numerically unbiased modified least-mean-square estimator (numLMSE) was developed as an alternative estimator for tracking applications, where localization relies on a small number of photons [29]. Further details on the calculation of the MLE and its implementation in a grid-search algorithm will be presented in Section 3.2.1.

---

<sup>1</sup> maximum-likelihood estimator

Finally, the localizations obtained in the previous step need to be filtered in order to remove false-positive events originating from background or thermally activated molecules located outside the EBP. However, it is generally difficult to filter out events with simultaneous emissions from multiple molecules located inside the confocal volume, especially in a cellular context [30]. The employed filtering parameters will be described in Section 3.5.

Overall, numerous parameters are involved in the different steps of the currently employed MINFLUX post-processing workflow, especially in the hidden Markov model, the grid-search algorithm, and the filtering. Setting these parameters requires a high level of user expertise. While MLE-based position estimation given a high number of collected photons is relatively straightforward, localizing emitters under low-photon conditions using the numLMSE becomes extremely complex and entails even more parameters. As a result, MINFLUX would benefit from an alternative data analysis approach that is easy to handle for users with limited experience. Most importantly the workflow should be robust even under challenging conditions such as low photon numbers and simultaneously activated emitters.

### 3.1.2 Deep learning in localization microscopy

Deep learning denotes a set of machine learning approaches that employ artificial neural networks and belong to the feature-learning, or representation-learning, techniques [51]. While conventional machine-learning methods require an experienced human engineer to extract features from the raw data before patterns can be detected in the input, representation learning automatically discovers these features. Deep learning in particular involves multiple levels, called *layers*, of composed transformations, which makes it possible to learn several representations at a time [52]. A neural network

is a universal function estimator, which is generally able to approximate any existing continuous function mapping the input data onto certain parameters of interest to any degree of accuracy [53]. This function may be so complex that it is impossible to find an analytical expression for it, for example in the case of the relationship between pixel intensities of a SMLM image and molecule coordinates. Still, a neural network can deduce an approximation of this function from a sufficient amount of example images.

The beginnings of neural networks and machine learning in general date back to 1957, when Frank Rosenblatt presented a simplified, mathematical model of biological neurons along with an algorithm that enabled these artificial neurons called *perceptrons* to learn from data [54]. However, the initial hype around machine learning came to rest when it emerged that a single layer of perceptrons was not capable of solving certain problems, of which the XOR operator is certainly the most prominent example [55]. Only in the 1980s, neural-network-based machine learning was re-discovered fueled by innovations in algorithms addressing the problem of training multi-layered architectures [56, 57] and by an easier access to computational resources [58]. Famous breakthroughs such as the victory of the neural network AlexNet in the ImageNet Large Scale Visual Recognition Challenge in 2012 [59], where 1.2 million images had to be assigned to 1000 classes, or the development of the program AlphaGo beating a professional Go player [60, 61] finally sparked a revolution in the development of artificial neural networks. In 2018 Yoshua Bengio, Geoffrey Hinton and Yann LeCun received the Turing Award for their breakthroughs in deep learning.

Image analysis was one of the earliest domains where deep learning yielded impressive results, notably in the recognition of handwriting [62, 63]. Medical and biological applications date back to the 1990s [64], and in the past few years, deep learning has spread throughout the entire field of medical image analysis and microscopy [65, 66]. Most recently, also SMLM has

started to take advantage of the power of neural networks to exploit the information encoded by a limited number of photons in the best possible way. The key requirement for training a neural network, a large amount of example data, is generally met in SMLM because signal formation can be accurately simulated using well-established models.

A major challenge in SMLM is the requirement of sparsely activated emitters, which often leads to long acquisition times. Even though localization in dense samples by fitting overlapping PSFs has been demonstrated before [67–70], these algorithms generally involve a high computational effort and sample-dependent parameter tuning. An alternative, deep-learning-based approach to tackle the problem of overlapping PSFs was first demonstrated with the Deep-STORM neural network, which had been trained with simulated 2D STORM images of densely packed molecules [71]. Afterwards, the problem of localizing multiple emitters was extended to 3D and addressed by various groups [72–76]. Numerous other deep-learning investigations focused on extracting additional molecular or imaging parameters from the PSF data, such as the color of the fluorescent emitter [77, 78], molecule orientation [72], stoichiometry of fluorescently labeled protein complexes [79], or structure and intensity of background [80].

Comparing deep learning to existing analysis methods, studies have attested a better performance of neural networks especially under challenging imaging conditions such as low SBR and high emitter density [71, 76]. Regarding computational performance, a direct comparison to the MLE gold standard has proved deep learning superior in estimating multiple output variables [74]. Most importantly, the previously mentioned deep-learning algorithms can be applied to all kinds of samples covered in the training process without making parameter adjustments. Thus, the analysis of SMLM data does not require specific user expertise anymore.

### 3.1.3 Scope of this chapter

Deep learning could be an alternative to the MINFLUX post-processing workflow outlined in Section 3.1.1 and provide parameter-free, fast data analysis. The second part of the current workflow, the position estimation, is a good starting point for integrating a new approach into the existing framework. In addition, replacing the MLE by a neural network can potentially extend the range of precise localization to lower photon numbers and higher background. While there are certain similarities to STORM and related SMLM techniques, the nature of MINFLUX recordings is fairly different from camera-based data. This difference needs to be taken into consideration when designing the architecture of a neural network. Therefore, the aim of this subproject is to verify that deep learning is a valid method for position estimation in MINFLUX nanoscopy. Furthermore, I want to identify conditions where neural networks provide a higher localization performance than the MLE.

## 3.2 Theory and methods

This section summarizes the theory of the MLE and how it was implemented. Next, it gives an overview of the fundamental concepts of deep learning and of the software and hardware used for this work. Finally, a few relevant statistical methods and naming conventions are described.

### 3.2.1 Maximum-likelihood estimation in MINFLUX

The MLE is a widely used estimator due to its advantageous properties:

- Consistency. For a large number of observations, the estimate converges in probability to the true value [81].
- Efficiency. The estimator asymptotically reaches the CRB [81].
- Functional invariance. If  $\hat{\theta}$  is the estimate of the parameter  $\theta$ , then the estimate  $\hat{\beta}$  of a parameter  $\beta = g(\theta)$  is given by  $\hat{\beta} = g(\hat{\theta})$  [82], where  $g$  is a one-to-one function, that is  $g(a) = g(b)$  implies  $a = b$  [83].

In the case of MINFLUX position estimation, the MLE maximizes the likelihood  $\mathcal{L}(\mathbf{r}_m|\mathbf{n})$  of a molecule being located at  $\mathbf{r}_m$  provided a set of photon counts  $\mathbf{n}$ . Using the functional invariance of the MLE, this can be expressed as a maximization of the well-known likelihood function  $\mathcal{L}(\mathbf{p}|\mathbf{n})$  defined in (1.15) for the multinomial parameter vector  $\mathbf{p}$ :

$$\hat{\mathbf{r}}_{m,\text{MLE}} = \arg \max_{\mathbf{r}_m} \mathcal{L}(\mathbf{p}(\mathbf{r}_m)|\mathbf{n}). \quad (3.1)$$

For simplification purposes it is convenient to maximize the logarithm of  $\mathcal{L}(\mathbf{r}_m|\mathbf{n})$  and to omit multiplicative factors that only depend on  $\mathbf{n}$ :

$$\hat{\mathbf{r}}_{m,\text{MLE}} = \arg \max_{\mathbf{r}_m} l(\mathbf{r}_m|\mathbf{n}) \quad (3.2)$$

with the log-likelihood function

$$l(\mathbf{r}_m|\mathbf{n}) = \sum_{i=0}^{K-1} n_i \ln(p_i). \quad (3.3)$$

The multinomial parameter vector  $\mathbf{p}$  is calculated according to (1.12) using an appropriate excitation intensity distribution  $I(\mathbf{r})$ . In 2D the doughnut profile is defined as in (1.21), whereas the 3D doughnut profile is often

approximated by a simple quadratic function [30]

$$I(\mathbf{r}) = x^2 + y^2 + z^2. \quad (3.4)$$

While the MLE for a 1D position can be easily calculated analytically [29], the optimization problem is solved numerically for the 2D and 3D cases.

For the present work the SBR was estimated together with the emitter position as previously described [30]. In iterative MINFLUX the true SBR cannot be simply extracted from the photon-count traces because it depends on the position of the emitter both relative to the EBP and within the confocal volume. The expression for the maximum likelihood estimate given in (3.2) accordingly changes to

$$(\hat{\mathbf{r}}_{\mathbf{m}}, \text{SBR}) = \arg \min_{\mathbf{r}_{\mathbf{m}}, \text{SBR}} \left( - \sum_{i=0}^{K-1} n_i \ln(p_i(\mathbf{r}_{\mathbf{m}}, \text{SBR})) \right). \quad (3.5)$$

The MLE was implemented in an iterative grid-search optimization algorithm following previous studies [30]. The initial grids had a diameter equal to the EBP width  $L$  for the emitter position  $\mathbf{r}_{\mathbf{m}}$ , and a diameter of 10 for the SBR. With  $L = 100$  nm, the following successive pixel sizes were used: 10 nm, 2 nm, 0.4 nm and 0.08 nm for  $\mathbf{r}_{\mathbf{m}}$ , and 1, 0.2, 0.04, 0.04 for the SBR. Only for  $\text{SBR}(\mathbf{0}, L_0) > 10$ , the pixels were scaled such that their initial size equaled  $0.1\text{SBR}(\mathbf{0}, L_0)$ . Events for which the position estimates coincided with any of the grid edges were discarded. Finally, the obtained molecule coordinates were multiplied by a factor of 0.7 along the  $z$ -axis in order to account for an axial scaling caused by a refractive index mismatch between the sample and the cover glass [30].

### 3.2.2 Fundamentals of deep learning

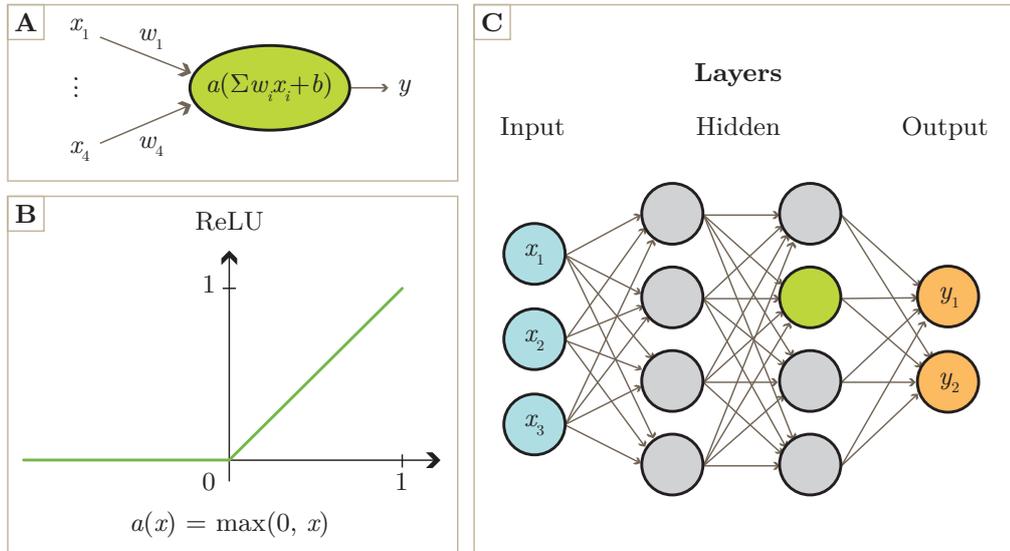
Deep learning is distinct from other, *shallow*, machine-learning techniques in that multiple, increasingly complex features of the data are automatically generated inside the network. By distributing the representation development over many layers, even an extremely complex feature can be broken down into simple transformations. These features are learned jointly, meaning that whenever an internal representation is modified, the other ones adapt accordingly without human intervention [84].

Layers are the fundamental building blocks of neural networks and act like filters on the data. They consist of smaller units termed *neurons* or *nodes*, each of which encodes a mathematical operation. A neuron computes a linear combination of its input values using a set of *weights*, and adds a so-called *bias* as depicted in panel A of Figure 3.1. This intermediate result is passed on to an *activation function*, whose output serves as an input for the following layer. Neural networks where data flows only in one direction, which were employed for our studies, are termed *feedforward*.

Activation functions play an important role in modern deep learning because they introduce nonlinearity to the mathematical operations performed in the different layers. They are an active research area, and the ultimate choice of this function depends on the nature of the learning problem. Among the most common activation functions are the sigmoid function, the hyperbolic tangent, and the ReLU<sup>2</sup>, which is visualized in panel B of Figure 3.1. It is nowadays the default recommendation [58, 85] and was used in this work. Since the proposition of the original ReLU [86–88], several variants have been developed [58]. The activation function in the final layer plays a special role by placing constraints on the output of the network. For a regression to arbitrary values, no last-layer activation

---

<sup>2</sup> rectified linear unit



**Figure 3.1: Building blocks of neural networks.** **A**, Artificial neuron. **B**, ReLU activation function. **C**, Fully connected neural network.

is required, which also applies to the problem modeled in this work.

In general, every neural network has input and output layers and a variable number of *hidden* layers in between. The total number of output and hidden layers denotes the *depth* of the network. There are various network architectures, which consist of different types of layers, depending on the nature of the input data. The most basic kind of layer, in which each neuron is linked to all the nodes of the previous level, is called *fully connected* or *dense* [84]. Networks containing only fully connected layers, as depicted in panel C of Figure 3.1, are effective for computationally simple problems and were used in this work. However, the number of trainable parameters scales quadratically with the number of neurons, the *width*, of each dense layer for networks featuring more than one hidden layer.

In order to limit the number of trainable parameters, image analysis tasks such as the SMLM applications listed in Section 3.1.2 are typically ad-

dressed by convolutional neural networks [62]. These networks additionally contain convolutional and spatial-downsampling layers. Each of their nodes processes only a small part of the input data, which makes it possible to learn local patterns in the images [58].

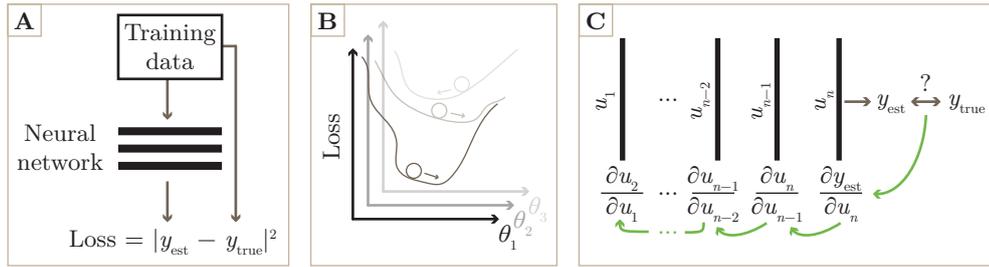
In the case of supervised learning, the deep-learning analysis workflow has two major parts: First, a neural network is trained and validated on data sets with known target outputs, and second, the trained network is used to infer the unknown output parameters from experimental data. The weights and biases are the parameters  $\{\theta_i\}$  to be optimized in the learning process and are initialized with random values. During the evaluation of the *training data*, they are iteratively adjusted in order to minimize the deviation of the *predictions* from the ground truth [58]. In the training process the performance of the network is assessed by calculating a *loss score* as indicated in panel A of Figure 3.2. An example of a common loss function is the mean squared error

$$\text{MSE} = \text{E} \left( \|\hat{\mathbf{y}}(\{\theta_i\}) - \mathbf{y}\|^2 \right) \quad (3.6)$$

of the predictions  $\hat{\mathbf{y}}$  relative to the target outputs  $\mathbf{y}$  [82].

Typically the whole data set is not processed at once, but split into *batches*. Just as the number and size of the layers, the loss function and batch size are *hyperparameters* of the network, which need to be set in the design process. The most important ones are listed in Table 3.1 [89]. For each batch, the loss is determined, and weights are updated before proceeding to the next batch. A single sweep through the training set is termed an *epoch*, and the training of a neural network usually lasts for many epochs. After each of them, the crucial step of rating the performance of the network on *validation data*, which was not part of the training set, takes place [58].

Learning is an optimization problem with the goal of finding the parame-



**Figure 3.2: Key concepts of training neural networks.** **A**, Loss. The MSE is calculated after a training-data batch has been fed forward through the network. **B**, Gradient descent. A minimum of the loss function is approached by taking steps in the opposite direction of the gradient with respect to the trainable parameters  $\{\theta_i\}$  of the network. **C**, Backpropagation. The trainable parameters are adjusted by iterating backwards through the layers  $\{u_i\}$  of the network. Figure adapted from Möckl, Roy, and Moerner [58].

ter combination that minimizes the training loss score. In deep learning, that problem is commonly solved by stochastic gradient descent (SGD) [90, 91], which is sketched in panel B of Figure 3.2. The loss is defined in a high-dimensional space spanned by the trainable parameters  $\{\theta_i\}$  of the neural network, as shown by the example of the MSE in (3.6). In order to minimize the loss, weights and biases are modified by a certain step size following the direction of the negative gradient. The latter is estimated from a randomly selected batch of the input data. In this context, the *learning rate* controlling the step size is another important hyperparameter. Usually its value decreases in the course of the training as the loss function approaches a minimum. There are various optimizers implementing this algorithm, including modern, adaptive SGD algorithms such as RMSprop [92] or Adam [93], which automatically adjust the learning rate for each parameter and over time. Among the hyperparameters of the Adam optimizer, which was used in this work, the initial learning rate is the most important one.

The gradients of the loss function are determined using the backpropaga-

Model	Training
Types of layers	Loss function
Number of layers	Batch size
Widths of the layers	Optimizer
Activation functions	Learning rate
Regularization factors	Number of epochs

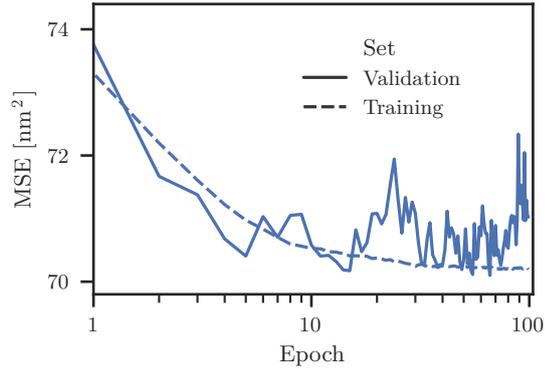
**Table 3.1: Main hyperparameters of feedforward neural networks.**

tion algorithm [57] depicted in panel C of Figure 3.2. It starts by calculating the gradient with respect to the weights and biases of the last layer and applies the chain rule to sweep through the network up to the first layer. This reduces the problem of determining the effect of each parameter on the loss to a series of matrix multiplications [58].

Because of the large number of trainable parameters, mapping the training input onto the ground truth is a highly overdetermined problem. Training a network too well can therefore reduce its ability to generalize and its performance on unknown data, which is called *overfitting*. The opposite phenomenon, a neural net that is not capable of modeling the relevant patterns in the data and is producing poor results on the training set, is termed *underfitting* and characterizes the initial learning phase. Therefore, the validation data is essential for assessing the generalization capacity of the network in the course of the training and finding the sweet spot between underfitting and overfitting. This is illustrated in Figure 3.3, where a neural network initially yields a comparable loss score on the training and validation data, but after a certain learning time, the validation performance degrades. The number of epochs is another important hyperparameter in this context.

There are numerous so-called *regularization* strategies to prevent models from overfitting. The *early stopping* functionality prematurely terminates the learning process as soon as the validation loss score stops improving,

**Figure 3.3: Training and validation loss curves.** MSE of a neural network on the training and validation data sets as a function of the number of training epochs.



which is an easy way to approximate the optimal number of epochs. *Weight decay*, or  $L^2$  regularization, adds a penalty term to the loss function in order to prevent the weights from assuming large values [94, 95]. *Dropout* sets the outputs of random subsets of neurons to zero [96]. All these approaches aim at helping the network to learn a more robust representation of the analysis problem [85].

While the training data is employed for adjusting the weights and biases of a neural network, the validation set is used to tune the model and optimization hyperparameters. In order to make sure that a trained network performs equally well on data that was involved in neither of the two processes, it can make sense to have a third distinct data set. Furthermore, this *test set* is useful for ultimately quantifying the performance of the trained model on new data with a known ground truth [84]. For testing the network on experimental data, this ground truth can be provided by the current gold standard in the field, such as established multi-emitter fitting algorithms for STORM data [67–70] or the MLE in the case of MINFLUX position estimation [29].

To conclude, it is worth noting that a neural network will always produce an output when confronted with experimental data. It can be difficult to detect artifacts in the predictions because these will usually lie within the

space defined by the target outputs of the training data. Therefore, it is crucial for the learning process to cover the sample space of all the data that the network should be able to analyze. In the context of SMLM, for example, the simulated PSFs employed in the training have to be representative of the ones found under relevant experimental conditions [58].

## Software and Hardware

For building, compiling, training, and evaluating neural networks in the present work, the Keras functional API<sup>3</sup> [97] was used, running on top of the TensorFlow library [98] for Python.

Training and evaluation were run on a workstation equipped with 96 GB of RAM<sup>4</sup>, an Intel Core i9-9820X, 3.30-GHz CPU<sup>5</sup> and an NVIDIA Quadro P2000 with 5 GB of memory.

The TensorBoard toolkit was employed for generating the loss curves presented in this chapter. For each epoch the average of all batch losses was calculated using identical batch sizes for the training and validation data sets.

### 3.2.3 Localization performance metrics

In order to quantify the localization performance of an estimator, three metrics were calculated for each 3D emitter position  $\mathbf{r}_m$ , namely the precision  $\sigma(\mathbf{r}_m)$ , the projection  $\tilde{b}(\mathbf{r}_m)$  of the bias averaged over the spatial

---

<sup>3</sup> application programming interface

<sup>4</sup> random-access memory

<sup>5</sup> central processing unit

dimensions – for simplicity called *bias* in this work, and the total error  $\mathcal{E}(\mathbf{r}_m)$ . The latter equals the square root of the MSE introduced in (3.6), which was used as a loss function in the training process. The metrics were calculated according to

$$\sigma(\mathbf{r}_m) = \sqrt{\frac{1}{d} \sum_{i=1}^d \mathbb{E}((\hat{r}_{mi} - \mathbb{E}(\hat{r}_{mi}))^2)}, \quad (3.7)$$

$$\tilde{b}(\mathbf{r}_m) = \frac{1}{\sqrt{d}} \|\mathbb{E}(\hat{\mathbf{r}}_m) - \mathbf{r}_m\|, \quad (3.8)$$

$$\mathcal{E}(\mathbf{r}_m) = \sqrt{\sigma(\mathbf{r}_m)^2 + \tilde{b}(\mathbf{r}_m)^2}, \quad (3.9)$$

where  $d$  denotes the dimensionality of the  $\mathbf{r}$ -space and  $\hat{r}_{mi}$  is the  $i$ th component of the estimated position [44]. The expectation values refer to localizations  $\hat{\mathbf{r}}_m$  assigned to the same molecular emission event or emitter coordinate in the case of experimental or simulated data, respectively. Typically  $\sigma$  was compared to the CRB on the precision of unbiased estimators as defined in (1.19), which is simply called *CRB* in this chapter.

### 3.3 Data simulation and network design

Our goal was to build a neural network for mapping sets of seven photon counts, acquired in the final step of the MINFLUX iteration scheme shown in Figure 1.2, to the 3D coordinates of a fluorophore. This can be modeled as a regression problem with a feature vector  $\mathbf{n}$  and a target vector  $\mathbf{r}_m$  of lengths seven and three, respectively. The simplest architecture to approach this kind of problem is a sequence of fully connected layers. Ini-

tially a single hidden layer with 5000 neurons was incorporated into the network, resulting in around 55 000 trainable parameters. The widths of the input and output layers are defined by the lengths of the feature and target vectors, respectively. I chose the commonly used ReLU activation function shown in Figure 3.1 in the hidden layers and no activation in the output layer because the target vector of our regression problem can assume arbitrary values.

For training the network the MSE loss function defined in (3.6) and the Adam optimizer mentioned in Section 3.2.2 were employed, keeping the default parameter settings except for the initial learning rate  $\alpha$ , which was reduced to  $1 \times 10^{-4}$ . In a direct comparison of different batch sizes, the default value of 32 proved optimal. Higher learning rates and bigger batch sizes both flattened the learning curves, and larger values of  $\alpha$  also increased the final loss score. In order to further automate the training process, the early-stopping functionality was employed for determining the ideal number of epochs. By this criterion, training ended when the validation loss had not improved for 30 epochs, and the model weights from the epoch with the lowest validation loss were restored.

The network was trained with simulated single-molecule, 3D MINFLUX data. For the simulation of photon emission events, the following parameters had to be set: The total number  $N$  of detected photons per localization, the reference  $\text{SBR}(\mathbf{0}, L_0)$  introduced in (1.26), the diameter  $L$  of the EBP in the final iteration step, and the position  $\mathbf{r}_m$  of the emitter relative to the EBP center. As described in Section 1.2.1, both the numbers of acquired signal photons and the background counts obeyed Poisson distributions. In this study, the simulation of photon counts did not depend on the excitation wavelength  $\lambda_{\text{exc}}$  because the simple, quadratic, 3D PSF model introduced in (3.4) was employed. Regarding the emitter position, it is worth noting that the CRB on the theoretically achievable localization precision depends on the position of the molecule relative to the EBP and

strongly increases for distances  $r_m \gtrsim 0.35L$  from the center of the EBP. This was for example shown in Figure 2.11 for 2D MINFLUX localization. Moreover, localizations outside this region are typically filtered out in the current implementation of MINFLUX post-processing to reduce the number of false-positive events [42]. Therefore, the network was trained and validated only within this spherical volume of radius  $r_{\text{train}} = r_{\text{val}} = 0.35L$  around the EBP. The scaling factor along the  $z$ -axis mentioned in Section 3.2.1 was not incorporated into the simulations but was later applied to the coordinates predicted by the trained network.

As mentioned in Section 3.2.2, it is critical for the reliability of a neural network to ensure that the training space entirely covers the sample space of the experimental data that will be presented to the network after training. Thus, a universal position estimator for MINFLUX nanoscopy should eventually be trained with a range of parameters covering the experimental conditions encountered in different types of samples. However, we decided to reduce the complexity of the problem by initially setting each parameter to a fixed value. For further simplification we chose to keep background negligible and to use a total photon number  $N = 1000$  that is sufficiently high for the MLE to produce reliable position estimates according to previous studies [42]. These simplified parameter settings were employed for training the first network and optimizing its architecture in this section and in the following one. The diameter of the EBP in the last MINFLUX iteration was set to  $L = 100$  nm throughout this work, a value that had been previously used for acquiring 3D MINFLUX data [30]. Both the full parameter ranges and the fixed values are provided in Table 3.2.

To create the data sets required for training, validating, and testing the neural network, a set of 3D ground truth positions was specified within the spherical training space. Then, combinations of seven photon counts were generated from the multinomial distribution defined in (1.8) by calculating the parameter vector  $\mathbf{p}$  at the given emitter position with specific SBR and

Parameter	Range of interest	Initial value
Number $N$ of detected photons	10–2000	1000
SBR( $\mathbf{0}, L_0$ )	1–50	1000
EBP diameter $L$	30–100 nm	100 nm
Relative emitter position $\mathbf{r}_m$	$\ \mathbf{r}_m\  \leq 0.35L$	

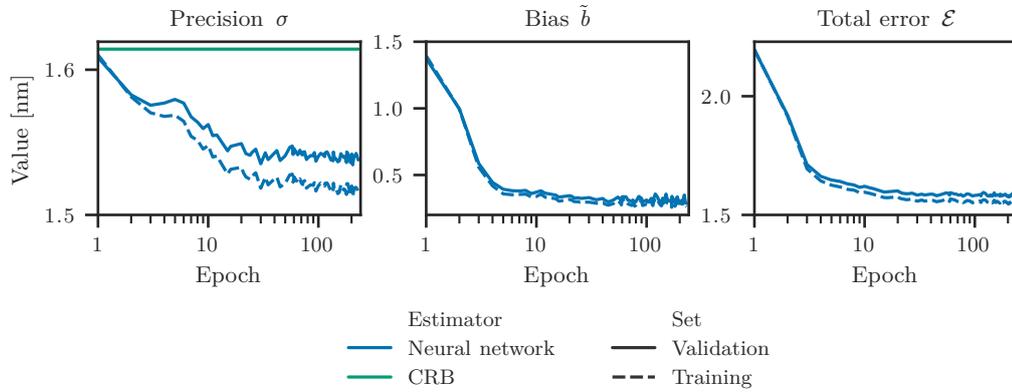
**Table 3.2: Parameters of the simulated MINFLUX data.** In the central column, the full parameter ranges that a universal position estimator should eventually cover are given. On the right side, the fixed parameter values that were initially used in the training data are listed.

total count number  $N$  per localization. The seven features were normalized such that their values ideally stayed between 0 and 1 for the most part in order to avoid losing numeric precision. To this end all photon counts were divided by the total count number  $N$  in the case of data sets with fixed  $N$ . For training with different total count numbers, the mean  $N$  of the data set was used for normalization. This normalization factor must be applied to any data presented to the trained network for evaluation.

The training and validation data set comprised approximately 5000 simulated emitter positions with 100 independently generated photon count combinations per position. Distributed over a spherical volume of radius  $0.35L$ , these positions formed a 3D grid with a spacing of 2.9 nm in each dimension. I manually performed a hold-out validation split by reserving a random subset of the molecule positions for validation. Note that all localizations with the same coordinates were assigned to either of the two subsets. For testing the fully trained network, I created another data set of the same size as the one reserved for validation. Again, there were 100 count combinations per position, but in this case, the coordinates were randomly distributed inside the validation space. Overall, the division of data into training, validation, and test sets followed a 70–15–15 split ratio.

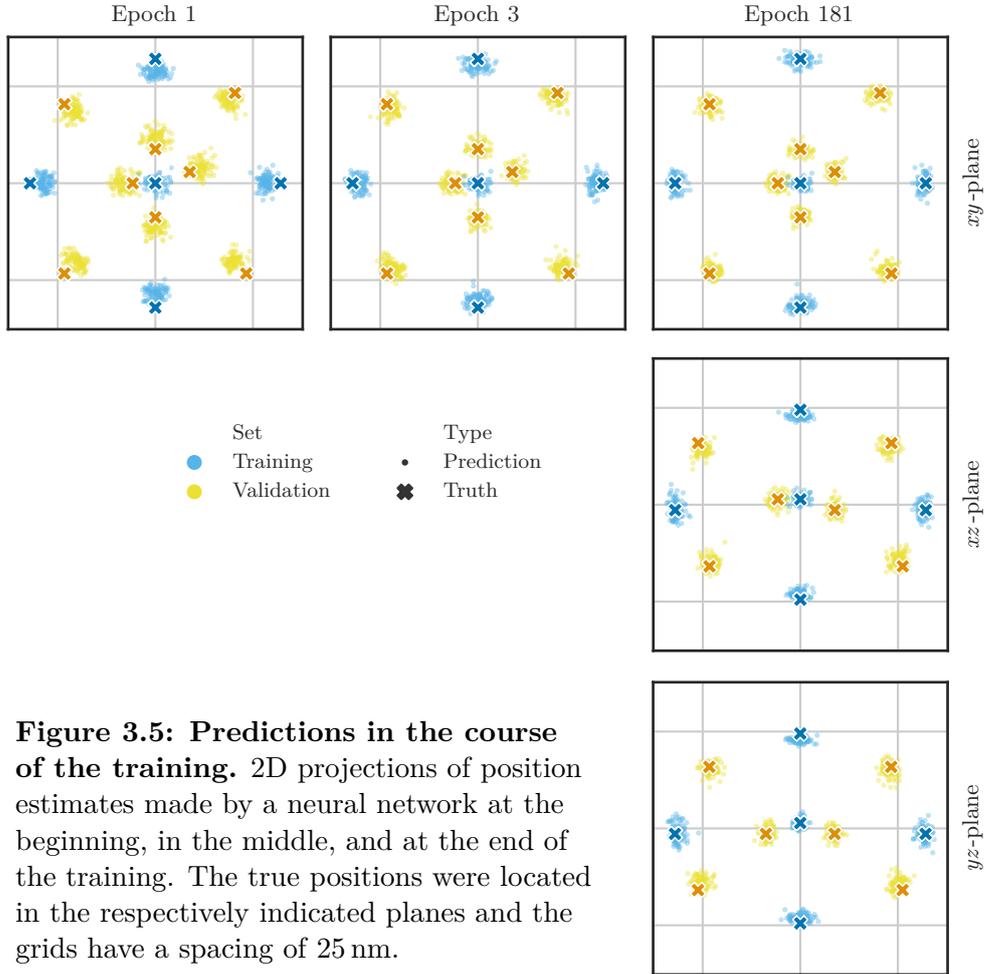
Figure 3.4 demonstrates how the localization performance metrics, which

were introduced in Section 3.2.3, gradually enhance in the course of the training. While the precision  $\sigma$  changes only slightly, the bias  $\tilde{b}$  substantially improves on both training and validation data in the first few epochs. Later, the learning curves, especially for the validation, flatten until the early-stopping criterion is met and training ends in order to prevent overfitting. The composition of the total localization error  $\mathcal{E}$  also changes over time: At the beginning,  $\sigma$  and  $\tilde{b}$  contribute more or less equally to the error, whereas at a later stage of the training,  $\mathcal{E}$  is dominated by the localization precision. Note that due to the presence of bias,  $\sigma$  can fall below the CRB on the uncertainty of unbiased estimators.



**Figure 3.4: Training effect on the localization performance.** Localization precision  $\sigma$ , bias  $\tilde{b}$ , and total error  $\mathcal{E}$  of a neural network on the training and validation data sets as a function of the number of training epochs.

The training effect on the localization performance can also be observed in a visual comparison of selected ground truth positions and the corresponding estimates made by the deep-learning algorithm, which is presented in Figure 3.5. The predictions shown in the left panel were calculated after the first training epoch. Given a training set with approximately 700 000 data points and a batch size of 32, the weights of the network had been updated around 20 000 times at that stage. Both training and validation samples located in the  $xy$ -plane are affected by a localization bias that is



**Figure 3.5: Predictions in the course of the training.** 2D projections of position estimates made by a neural network at the beginning, in the middle, and at the end of the training. The true positions were located in the respectively indicated planes and the grids have a spacing of 25 nm.

pointing outward for events close to the coordinate center and inward at the edge of the training volume. Only two epochs later this bias has almost vanished, especially in the inner part of the sample space, which is shown in the central panel. After this point it took another 178 epochs with a substantially lower learning rate to complete the training. The final results are displayed in the three right panels for positions located in the  $xy$ -,  $xz$ - and  $yz$ -planes. It becomes evident that the bias of the estimator has been further reduced and the point clouds formed by the position estimates are

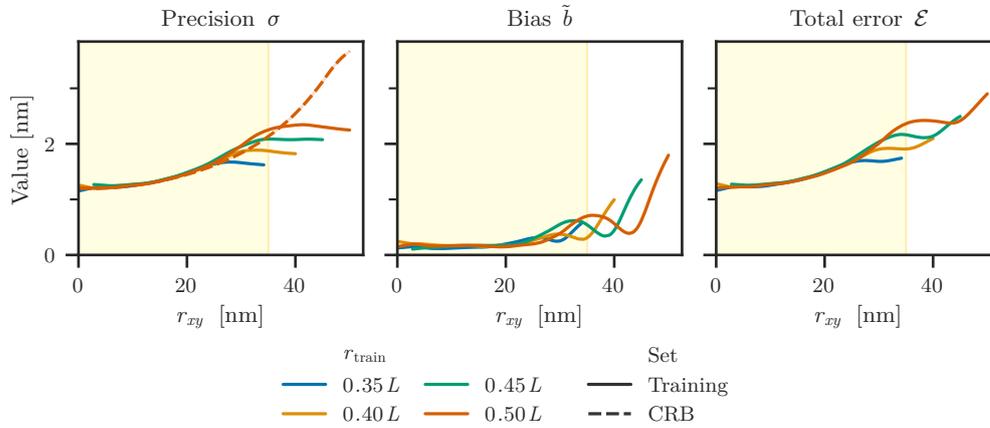
more compact than at the earlier training stages. Still, this improvement is relatively small compared to the one made in the first three epochs. The remaining inward-directed bias is slightly more pronounced along the  $z$ -axis than in the  $xy$ -plane. Otherwise the localization performance seems to be almost isotropic.

### 3.4 Optimization of the network architecture and training data

After I had constructed and trained the first version of a neural network that was able to estimate 3D positions from simulated MINFLUX data, my goal was to improve the localization performance by optimizing the network and training designs. The total localization error  $\mathcal{E}$  introduced in (3.9) is composed of the precision  $\sigma$  and the bias  $\tilde{b}$ . As  $\sigma$  was already surpassing the CRB at the end of the training, which is demonstrated in Figure 3.4, our focus was to first reduce the bias of the deep-learning-based position estimator. To be more specific, we wanted to reduce the inward-pointing bias in the outer regions of the validation space, which can be observed in the right panels of Figure 3.5.

To this end, I studied the effect of extending the training volume beyond the region where the network gets validated. In order to have similar sampling densities, the number of molecule positions in the data sets was scaled along with the extent of the grids. Judging by Figure 3.5, the localization performance is almost isotropic and comparable for training and validation data. Hence, this investigation was limited to training positions in the  $xy$ -plane, and the performance metrics  $\sigma$ ,  $\tilde{b}$ , and  $\mathcal{E}$  were examined as a function of the distance  $r_{xy}$  from the  $z$ -axis. More precisely, for each  $r_{xy}$  the metrics obtained for target positions within 2 nm distance from the  $xy$ -plane were

averaged. From Figure 3.6 it becomes evident that the larger the radius  $r_{\text{train}}$  of the training volume, the larger the area with small bias and a localization precision curve closely following the CRB. The region where  $\sigma$  falls below the CRB moves further towards the edge of the validation space. At the same time, the point where the bias strongly rises moves outwards. However, this bias increase also becomes more pronounced for larger values of  $r_{\text{train}}$ . Therefore,  $r_{\text{train}} = r_{\text{val}} = 0.35L$  minimizes the total error  $\mathcal{E}$ , but  $r_{\text{train}} = 0.4L$  provides the lowest average bias in the validation space and thus best meets our optimization criterion. To conclude, the training volume that minimizes the bias is slightly larger than the validation volume.

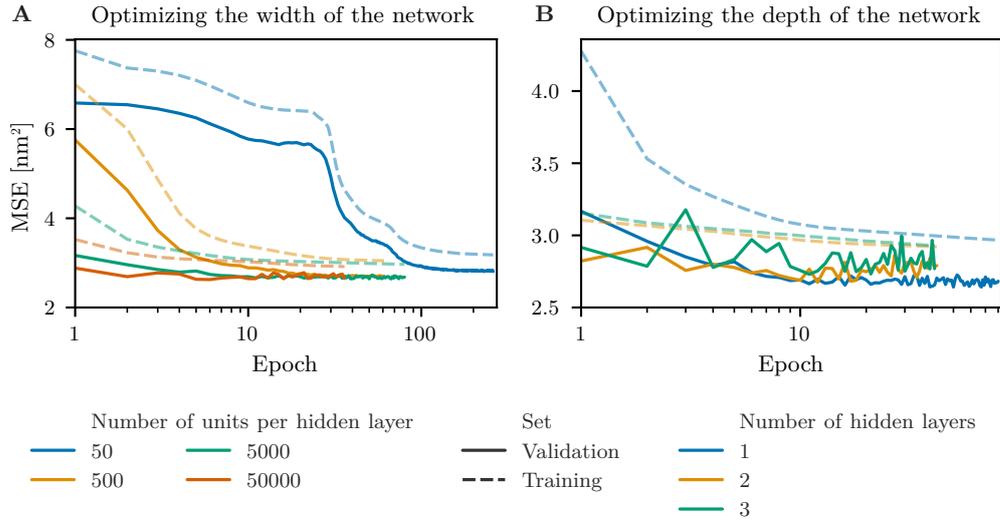


**Figure 3.6: Influence of the extent of the training space on the localization performance.** Localization precision, bias, and total error of a fully connected network with a single hidden layer containing 5000 nodes, trained in regions of different radii  $r_{\text{train}}$ . The localization performance was assessed in the  $xy$ -plane as a function of the distance  $r_{xy}$  from the  $z$ -axis, and the resulting curves were smoothed with a Gaussian kernel of size  $\sigma = 1.5$  nm. The yellow-tinted areas depict the validation space of the neural network with a radius of  $r_{\text{val}} = 0.35L$ , where  $L$  denotes the diameter of the excitation pattern.

Having optimized the extent of the training space, my next goal was to find the ideal size for the network to maximize both localization performance and computational efficiency. The training and validation data sets con-

tained approximately 8000 simulated emitter positions in total, distributed in volumes of radius  $r_{\text{train}} = 0.4L$  and  $r_{\text{val}} = 0.35L$ , respectively. Figure 3.7 compares the loss curves for networks with different numbers of units per hidden layer and with a varying number of hidden layers, shown in the left and right panels, respectively. Our criteria for a suitable architecture were a low validation loss score at the end of the training and a validation curve that is dropping as smoothly as possible. Note that the total number of training epochs varies between the curves due to the early-stopping callback mentioned at the beginning of Section 3.3, which terminates the training as soon as the validation performance stops improving. In general, the validation loss fluctuates in the learning process, whereas the training loss is monotonously decreasing. That is a typical behavior because the gradients of the training and validation loss functions do not necessarily point in the same direction in the space spanned by the trainable parameters. It is also important to indicate that the validation loss is usually lower than the training loss in our example. This results from the fact that molecule positions located at the edge of the training volume, where we expect the localization error to be maximal, were excluded from the validation data set as described in Section 3.3.

The left plot suggests that a single hidden layer with 50 neurons does not provide the required complexity to solve the regression problem. Indeed, the training and validation curves do not steadily decrease but approach a local optimum at first before they further drop to the final loss score. Moreover, the MSE at the end of the training is higher than for the bigger architectures. The networks with 500, 5000 and 50 000 nodes per layer all produce smooth loss curves and reach a similar final validation MSE. I decided to set the width of the hidden layers to 5000 for the following investigations. The right panel of Figure 3.7 presents a comparison of networks with one, two and three hidden layers, each one containing 5000 neurons. The deepest architecture shows an increasing validation loss accompanied by a further decreasing training loss, which is a sign of overfitting to the



**Figure 3.7: Loss curves for networks of different sizes.** Training and validation MSE of fully connected networks with **A**, a single hidden layer containing a varying number of nodes or **B**, different amounts of hidden layers with 5000 nodes each. The loss scores are shown as a function of the number of training epochs.

training data. This cannot be observed in the curves produced by the two shallower designs, which are comparable to each other.

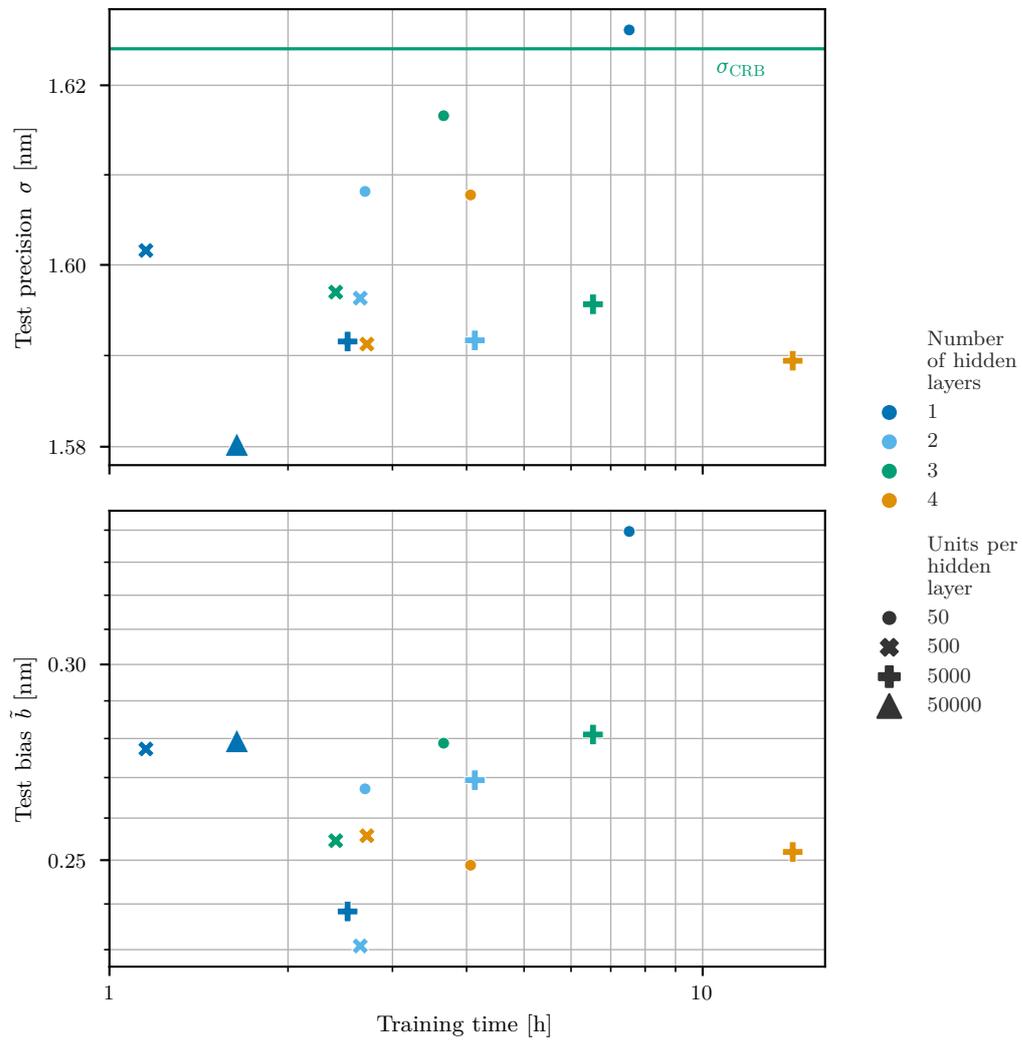
Besides the localization performance, computational efficiency plays another crucial role for the practical applicability of any position estimator. Therefore, I took a closer look at the relationship between the time required for training networks of different sizes and the precision and bias of the resulting estimates. Figure 3.8 displays the precision and bias on simulated test data, which were not used for training or validation, as a function of the training time for 13 different architectures. This detailed analysis reveals that especially the networks with a single hidden layer and 5000 nodes, and with two layers containing 500 neurons each provide precise estimates and minimize the bias. Moreover, their training takes less than three hours. These designs are marked by a blue plus sign and a light-blue cross, respectively. Other medium-sized architectures, such as

the one featuring two hidden layers with 5000 units each and marked by a light-blue plus sign, also show a satisfying localization performance and computational efficiency. The exact values of the obtained bias and the training time can be found in Figure B.1.

It also becomes evident that the training time does not necessarily increase with the number of trainable parameters. As a matter of fact, for the network with a single hidden layer and 50 neurons, indicated by a blue circle, it took almost eight hours to find the optimal values for the 553 trainable parameters. In contrast, the architecture marked by the blue triangle, which had 50 000 nodes in its hidden layer and correspondingly approximately 1000 times more trainable parameters than the smaller model, required less than two hours for training. This can be explained by the fact that the number of epochs needed for this model to converge was significantly smaller than for the narrower design.

Having compared different network architectures, the influence of the training data composition on the localization performance and training time were studied in the next step. More precisely both the number of simulated molecule positions and the number of localizations per position were varied. The results are displayed in Figure 3.9. Like for the previous figure, the average localization precision and bias on simulated test data were calculated. These values could be reproduced with average accuracies of around 0.8% and 14%, respectively, when repeating the training. The exact values of the obtained bias and the training time can be found in Figure B.2.

The two plots suggest that training and validation data with a minimum of approximately 250 distinct positions in total are required for reliably localizing events with a precision  $\sigma$  close to the CRB and a bias  $\tilde{b}$  below 0.5 nm. Placed on a 3D grid, these training coordinates have a spacing of around 11 nm in each dimension. For finer grids a higher number of

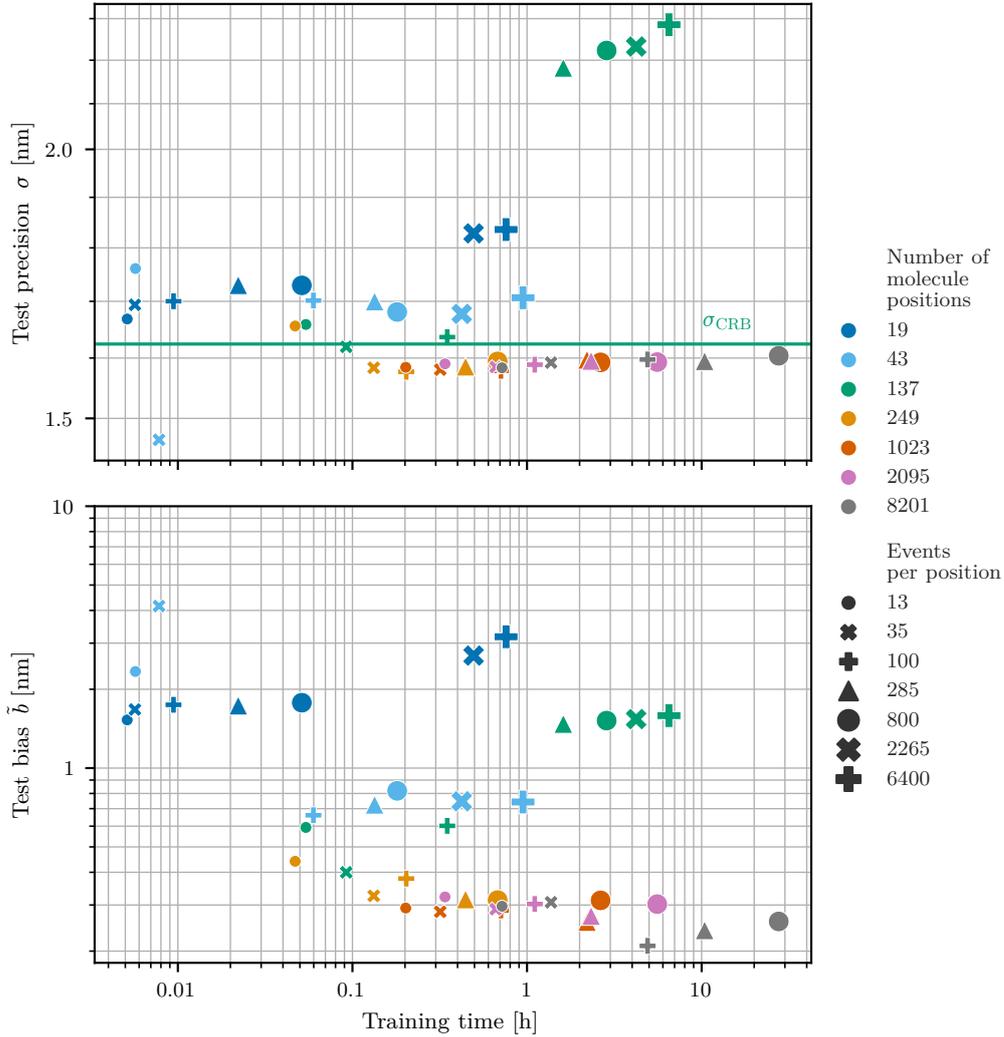


**Figure 3.8: Localization performance and training time for different network sizes.** A training data set with approximately 800 000 data points as described in Section 3.3 was used. The test data consisted of a total of 100 000 simulated events located at 1000 distinct positions, which were randomly distributed in the validation space. The localization precision and bias were calculated at each position, and the mean values over all positions were plotted.

localizations per molecule position is followed by a slight bias reduction. The precision  $\sigma$ , however, does not seem to be affected by that number as long as the grid spacing is sufficiently small. Most prominently the amount of time  $\Delta t_{\text{train}}$  needed for training the network increases with the size of the used data set  $n_{\text{train}}$ . More specifically, this increase can be described by the following relationship:  $\Delta t_{\text{train}} \propto n_{\text{train}}^{0.85}$ .

For data sets containing less than 150 distinct positions, the behavior of the localization precision and bias as a function of the number of events per coordinate observed in Figure 3.9 is less predictable. In some cases, multiplying the number of data points per position seems to lower the localization precision and elevate the bias. This unexpected observation could be reproduced when creating another data set and training a new network with it. Just as described in the previous paragraph, the training time increases with the size of the data set also for lower-density coordinate grids. The relationship between the number of data points provided during training and the time required for it is similar to the one given before:  $\Delta t_{\text{train}} \propto n_{\text{train}}^{0.78}$ .

Regarding the question which training and validation data composition is the best, the choice made earlier was quite conservative in retrospect. As a matter of fact, the data set containing around 8000 simulated positions with 100 localizations each, marked by a small, gray plus sign, produces a network with high precision and the lowest bias on test samples among all the data sets presented in Figure 3.9. However, this comes at the cost of a relatively long training time, namely five hours. Other arrangements such as the one with approximately 1000 grid coordinates and 13 localizations per position, indicated by a small, red circle, provide a similar test precision  $\sigma$ , around 1.6 nm, and a bias  $\tilde{b}$  that is still below 0.3 nm, but the training takes less than 15 minutes. Still, for the results displayed in the next section, the size of the data set was not reduced below 800 000 events in order to ensure optimal learning even under more demanding conditions,



**Figure 3.9: Localization performance and training time for different training data compositions.** A fully connected network with two hidden layers and 5000 nodes per layer was used. The test data consisted of a total of 100 000 simulated events located at 1000 distinct positions, which were randomly distributed in the validation space. The localization precision and bias were calculated at each position, and the mean values over all positions were plotted. The numbers of molecule positions and events per position refer to the size of a common data set for training and validation before splitting it into two parts.

namely lower photon numbers  $N$  and SBR.

## 3.5 Comparison to the maximum-likelihood estimator

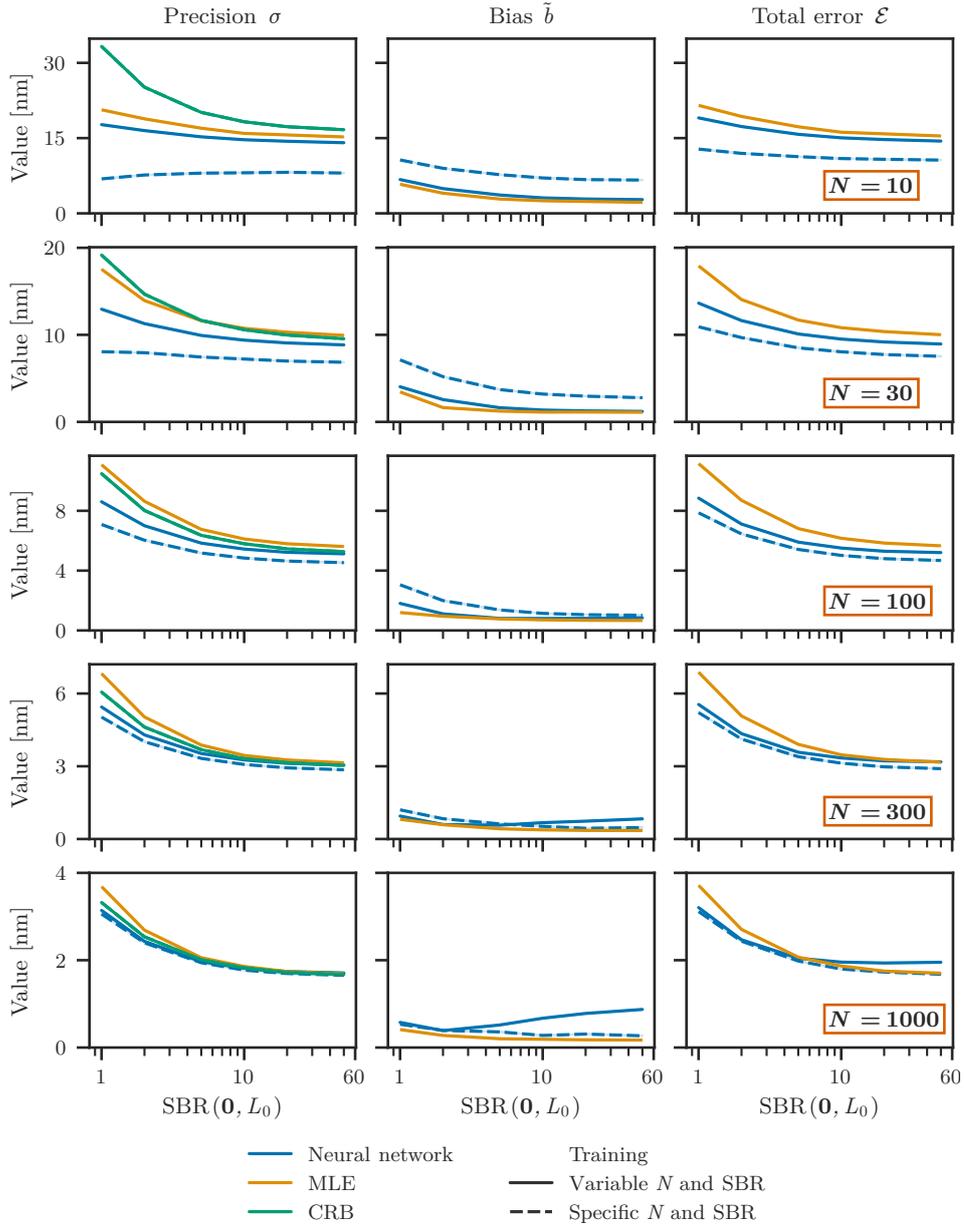
Up to this point the neural networks had been trained and evaluated on data simulated under non-challenging conditions, namely a high number  $N$  of photons per localization and negligible background. Having optimized both the size of the network and the composition of the training data set in order to maximize the localization performance, the next important step was now to compare the deep-learning algorithm to the MLE with respect to their performance in various parameter settings. To this end, 30 separate test data sets were created using a different combination of  $N$  and  $\text{SBR}(\mathbf{0}, L_0)$  for each of them. Every set consisted of a total of 100 000 simulated events located at 1000 distinct positions, which were randomly distributed in the validation space of radius  $r_{\text{val}} = 35$  nm. The total photon number  $N$  per localization and the SBR were within the ranges suggested in Table 3.2.

For the network, I chose a fully connected architecture containing two hidden layers with 500 neurons each, based on the results displayed in Figure 3.8. Two different training approaches were used: First, I trained separate, specialized networks matching the individual combinations of  $N$  and  $\text{SBR}(\mathbf{0}, L_0)$  that could be found in the test data. Each training set consisted of around 800 000 data points distributed over approximately 8000 distinct grid positions. In the second approach, I trained a single, universal network with a range of  $N$  and  $\text{SBR}(\mathbf{0}, L_0)$  values. In order to generate the photon counts, random  $N$  and  $\text{SBR}(\mathbf{0}, L_0)$  parameter values were drawn at about 8 000 000 distinct training and validation grid positions.  $N$  ranged between 10 and 1000 and was drawn from a discrete, uniform distribution, whereas  $\text{SBR}(\mathbf{0}, L_0)$  followed a normal distribution  $\mathcal{N}(1.3, 1.0)$ . The latter was obtained from a fit to previously published experimental MINFLUX

data [30], which will be shown later in this section. Additionally, a lower limit was set to  $\text{SBR}(\mathbf{0}, L_0)$ , equaling the smallest value encountered in the experimental data. A flexible SBR is the most relevant condition in practice because  $\text{SBR}(\mathbf{0}, L_0)$  varies within the sample and the training data of a reliable estimator should cover these variations.

After the learning process had been completed, the test data sets were evaluated using both the differently trained networks and the MLE described in Section 3.2.1. In some cases, especially for small  $N$  and SBR, the MLE predicted the emitter's position at the edge of the optimization grid. These localizations constituted up to 10% of the data points and were filtered out.

Figure 3.10 illustrates how the existing MLE and the different networks compare in terms of localization precision, bias, and total error. As the number of detected photons and the SBR increase, the localization uncertainty and bias of all estimators generally fall along with the CRB on the theoretically achievable localization precision. Above a certain  $N$  and SBR, however, the performance of the universal network starts to deteriorate, which could be a sign of overfitting. In general the neural networks outperform the established MLE judging by the total localization error, especially under challenging conditions, namely small  $N$  and SBR. Only for  $N = 1000$  and  $\text{SBR}(\mathbf{0}, L_0) \gtrsim 5$ , the MLE has a smaller error than the universal network. Moreover, these two estimators have a similar bias, which is small compared to  $\sigma$ , for  $N < 1000$ . In contrast, the specialized nets are dominated by bias especially in the low-photon regime. For the following evaluation of experimental data, we additionally trained a network with variable SBR and fixed  $N = 1000$ , whose localization performance on test data was almost identical to the one of the universal network. Overall, the presented results suggest that compared to the existing MLE, the neural networks designed in this work have a similar or even better localization performance in all experimentally relevant photon and SBR regimes.



**Figure 3.10: Localization performance of neural networks and an MLE in various parameter settings.** The localization precision, bias, and total error of the different estimators were assessed on simulated test data sets with specific SBRs and photon numbers  $N$  per localization. Neural networks were trained in two distinct ways. The performance metrics were calculated at each position, and the mean values over all positions were plotted.

Subsequently, we wanted to gain a deeper understanding of how the number of learning epochs and the amount of training and validation data influence the localization performance of the universal network under different conditions. To this end, we probed the localization performance at different intermediate training stages and, in an independent approach, trained the network with a smaller data set. It becomes evident from Figure B.3 that for large values of  $N$ , the full localization performance is already reached after a single training epoch. For  $N = 1000$  and high SBR, further training even increases the test bias. Moreover, Figure B.4 reveals that reducing the size of the training set lowers the test bias in this regime. For low photon numbers, however,  $\sigma$  and the total localization error increase when less training data is provided. Note that for the network trained with less data, the weights received 23 times less updates than in the case of the larger data set. This difference resulted from the fact that the batch size had not been increased along with the amount of data, and in addition, the larger data set coincided with a higher number of epochs determined by the early-stopping criterion. To conclude, more training data and epochs seem to improve the localization performance under challenging conditions, whereas this effect is reversed under easier conditions.

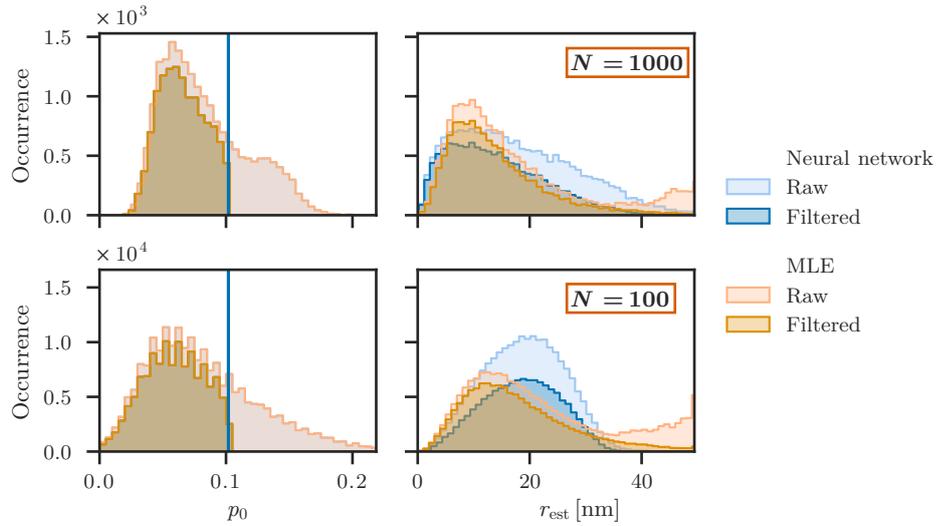
Next, we wanted to extend our findings to experimental data and compare the position estimates of a neural network trained with simulated photon counts to the ones made by the MLE. Again, I chose a fully connected network with two hidden layers consisting of 500 nodes each. It was trained using a data set containing about 800 000 emission events with a fixed total photon number  $N$  and a normally distributed  $\text{SBR}(\mathbf{0}, L_0)$  located at distinct grid positions. The experimental photon counts that were presented to the neural net after training were kindly provided by Dr. Jasmin Pape. The data originated from images of fixed U-2 OS Nup96-SNAP cells stained with Alexa Fluor 647, which had been acquired with a MINFLUX nanoscope [30]. In this data set, 3256 distinct emitters had been identified with around 6500 photons on average detected in the last iteration of the

image acquisition. Photon-count traces were split into localizations with fixed photon numbers, namely either  $N = 1000$  or  $N = 100$ , before making predictions on the emitter positions.

The fluorophore coordinates predicted by the MLE or the neural network were filtered to omit false-positive events, which is displayed in Figure 3.11. In line with previously described approaches [30], we used the emission probability  $p_0$  in the central exposure as a filtering criterion. We set an upper threshold to  $p_0$  in order to discard events originating from either background or molecules emitting outside the EBP that had been accidentally activated. For both kinds of false-positive events, that probability is distributed around  $p_0 = \frac{1}{K}$  with  $K$  being the number of excitation positions. Therefore, we chose an upper limit  $p_0 = \frac{c}{K}$  with  $c < 1$ . By taking the value  $p_0 = 0.10188$ , which had been found in previous studies [30], we were able to discriminate two subpopulations in the distribution of  $p_0$  for  $N = 1000$ . For  $N = 100$  we used the same  $p_0$  threshold even though the two groups were less apparent in this case.

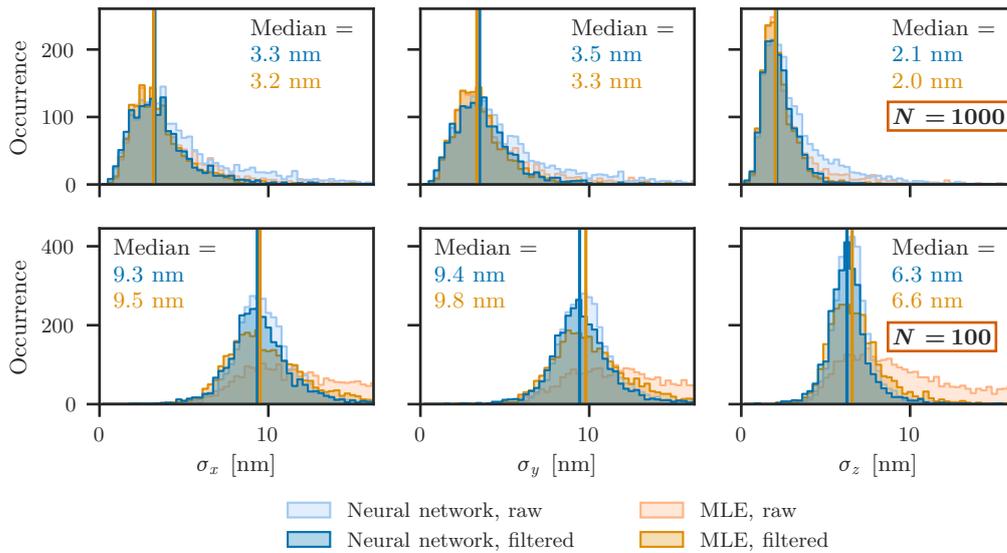
Another potential filtering criterion, which was previously used [30], is the distance  $r_{\text{est}}$  of the estimated position from the EBP center. In the case of the maximum likelihood estimates,  $r_{\text{est}}$  can be exploited to separate two subpopulations of events. However, the right plots in Figure 3.11 indicate that the deep-learning-based estimates do not form these two groups. Therefore and because our goal was to employ identical filters for the molecule positions predicted both by the MLE and by the neural network, we did not apply any threshold to  $r_{\text{est}}$ . Moreover, the distribution of  $r_{\text{est}}$  in the output of the MLE suggests that filtering the data only by  $p_0$  is already sufficient to remove the peripheral subgroup of localizations almost completely.

In order to compare the localization performance of the neural network and of the MLE, our next step was to calculate the 3D standard deviation



**Figure 3.11: Filtering of experimental data.** The figure displays histograms of the relative photon count number  $p_0$  in the central exposure and the distance  $r_{\text{est}}$  of the estimated positions to the EBP center in the last iteration. Localizations were calculated with an MLE or a neural network. The underlying photon-count traces were split at two distinct predefined photon numbers, leading to different amounts of localizations in the top-row and bottom-row data sets. Filtering was performed by applying an upper threshold to  $p_0$ .

$\sigma$  of the position estimates for each molecule. Figure 3.12 displays  $\sigma$  in all spatial dimensions for both filtered and unfiltered predictions made by either of the two estimators. We only evaluated clusters with a minimum of five consecutive localizations at the same position. For  $N = 1000$  detected photons per localization, the neural network and the MLE show almost identical precision, whereas for  $N = 100$ , the network is up to 5% more precise on average. In general, the 3D standard deviation is more or less isotropic in the  $xy$ -plane and slightly smaller along the  $z$ -axis, even before applying the scaling factor of 0.7 along the optical axis as mentioned in Section 3.2.1. Regarding the effect of filtering on the localization precision, it is beneficial for both estimators in the case of  $N = 1000$ , where the median of the precision on raw data is about 20% larger compared to the filtered values given in Figure 3.12. However, for  $N = 100$  a completely different

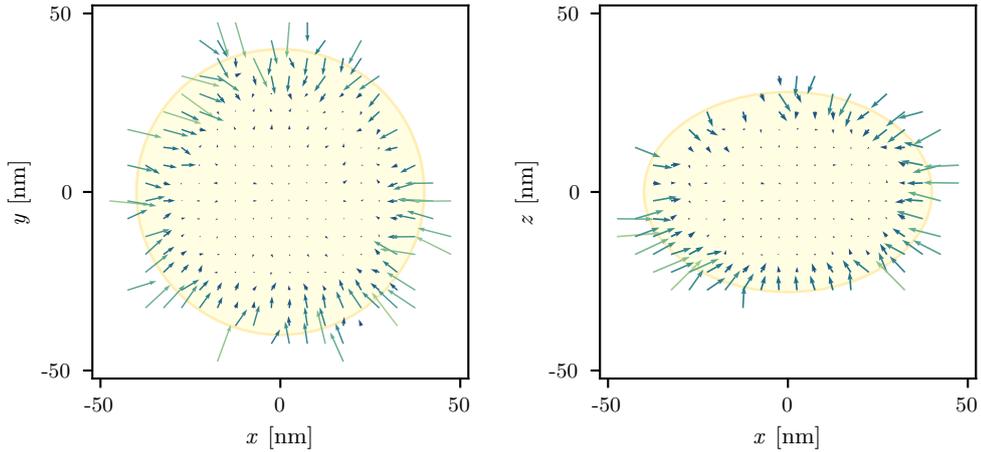


**Figure 3.12: 3D localization precision on experimental data.** The plots show the standard deviations of the position estimates for each fluorophore on data sets with two different photon numbers  $N$  per localization, which are presented in the top and bottom rows, respectively. Histograms are shown for estimates made using either a neural network or an MLE, both before and after filtering as demonstrated in Figure 3.11. The given medians refer to the distributions of the filtered localizations.

picture emerges: For the MLE the localization precision deteriorates by 35 % when omitting the filtering, whereas the neural network stays almost unaffected.

Besides the standard deviation of the position estimates, the bias with respect to the ground truth is another metric for judging the performance of the neural network. However, in the case of experimental data, the ground truth is typically unknown. Therefore, we took the predictions made by the MLE in a photon regime where this estimator is supposed to have a precision close to the CRB and negligible bias [42, 44], namely  $N = 1000$ , as a reference. The offset between the output of the neural network and the reference positions averaged over bins of 5 nm width is visualized in Figure

3.13. The projections into the  $xy$ - and  $xz$ -planes demonstrate that for an EBP diameter of  $L = 100$  nm, the two estimators provide almost identical results within a region of radius  $0.3L$  approximately. Close to the edge of the training volume, the deep-learning-based estimator has an inward-directed offset relative to the MLE of around 6 nm on average. Outside the training space, this value further increases.



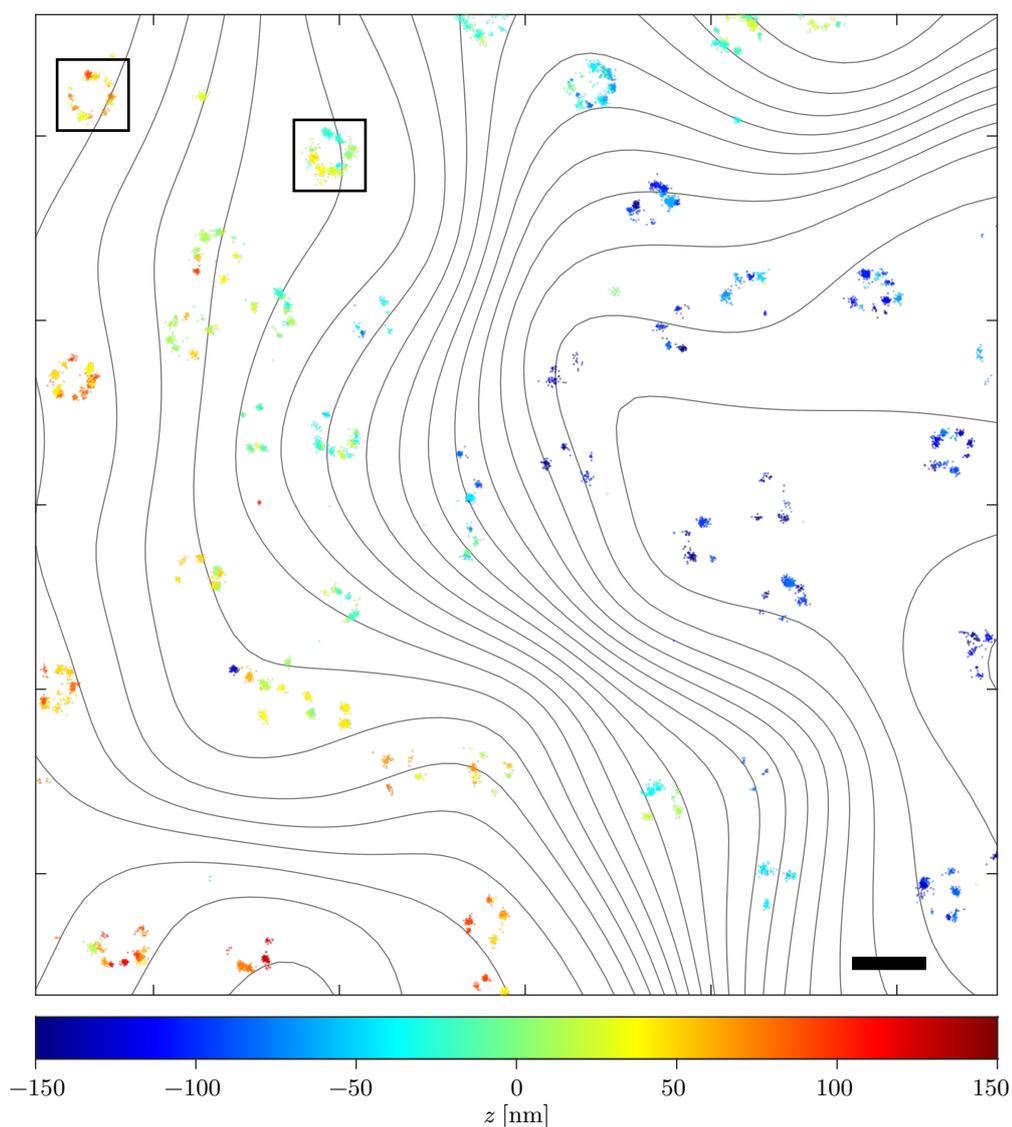
**Figure 3.13: Average offset between position estimates of a neural network and an MLE.** The arrows are pointing from the positions estimated by the MLE to the predictions made by the neural network based on MINFLUX data containing around  $N = 1000$  photons per localization. The arrows' length equals the actual offset in the used coordinate system. For better visualization, the positions relative to the center of the EBP were binned using a bin width of 5 nm. The yellow-tinted areas depict the training space of the neural network with a radius of  $r_{\text{train}} = 0.4L$ , where  $L$  denotes the diameter of the excitation pattern.

The filtered experimental data evaluated by the neural network for  $N = 1000$  photons per localization is presented as a 2D projection with a color-coded  $z$ -coordinate in Figure 3.14. An MLE-based version of this image with  $N = 2000$  was previously published [30]. The used network had a moderate training time of around one hour, and the evaluation of the experimental data took 0.9 seconds. Compared to the MLE, which required five minutes to localize approximately 20 000 emission events contained in the data set, the prediction speed of the deep-learning algorithm was 320

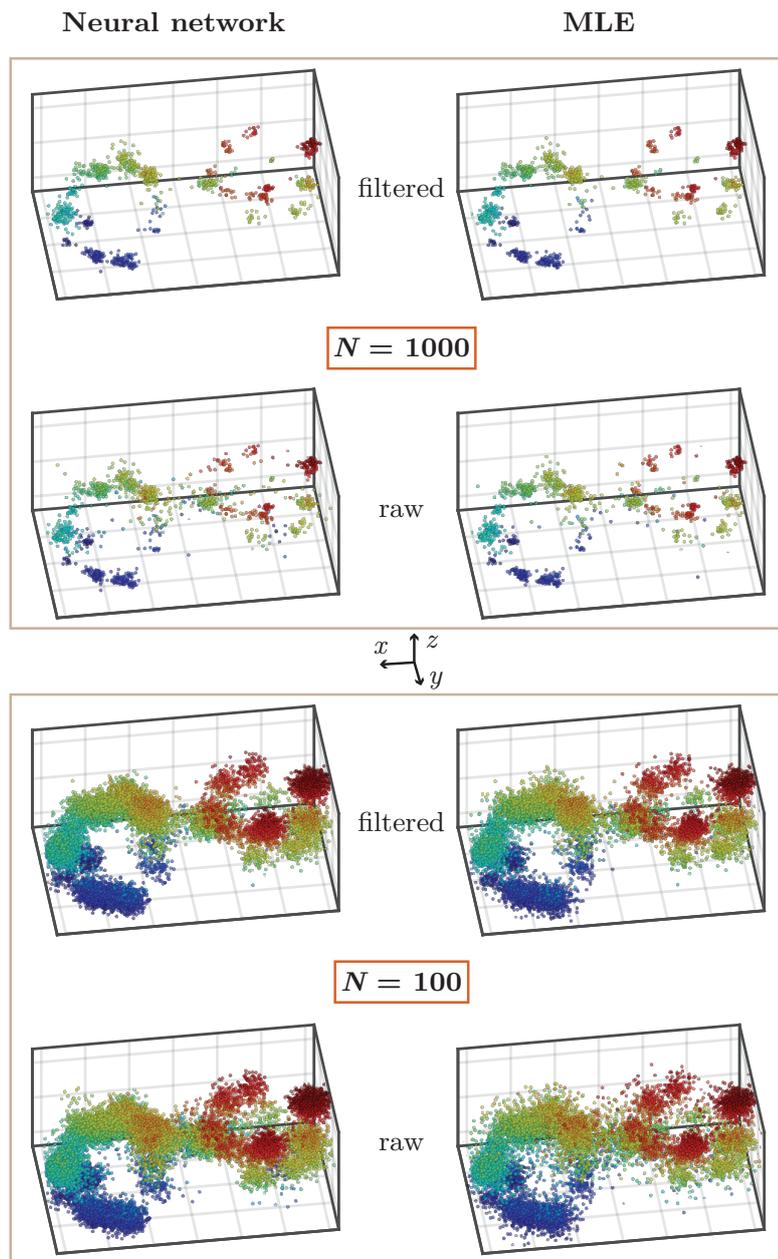
times higher.

Finally, we wanted to compare the image quality produced by the two estimators in 3D and observe how it is affected by filtering and by the number  $N$  of photons per localization. To this end, we examined the two nuclear pores inside the boxes in Figure 3.14 more closely. The 3D renderings in Figure 3.15 visualize the nucleoplasmic and cytoplasmic rings of Nup96 and indicate the eight-fold symmetry of the complex. Each ring consists of 16 copies of the protein, which are divided into eight subunits 42 nm apart [99, 100].

Only small differences can be perceived between the images produced with the neural network and with the MLE for  $N = 1000$ . The events localized by the deep-learning algorithm seem to be slightly more spread out within the clusters, especially when visualizing raw data. In the case of  $N = 100$  however, the representation of filtered MLE localizations shows more data points lying outside the clusters than the rendering of neural-network position estimates. An even more prominent difference between the two estimators is the effect of filtering on the image quality in the low-photon regime: When using the MLE, considerable amounts of outliers appear when exceeding the  $p_0$  threshold. For the neural network though, this effect is rather small: Even in the raw data, the point clouds marking the positions of the fluorescent emitters look relatively compact. These observations are consistent with the results shown in Figure 3.12. In conclusion, the two estimators yield similar results, but the MLE relies more on filtering. In general, it becomes evident that a total number of 100 emitted photons per localization is not sufficient to discern the subunits of Nup96 reliably.



**Figure 3.14: Nup96-SNAP-Alexa-Fluor-647 labels localized by a neural network in a MINFLUX image of fixed nuclear pore complexes.** Position estimates are visualized by circular patches of radius 3 nm and opacity  $\alpha = 0.5$ . Contour lines represent the 3D surface of the nuclear envelope calculated from the weighted average of the  $z$ -coordinates of all position estimates using a 2D Gaussian kernel with  $\sigma_{xy} = 250$  nm. Further parameters:  $N = 1000$  photons per localization and scale bar indicating 200 nm.



**Figure 3.15: 3D rendering of single nuclear pores localized by a neural network and an MLE.** Two images of nuclear pores were taken from Figure 3.14 and rotated  $270^\circ$  against each other in the  $xy$ -plane before being placed side by side. Localizations are displayed as surface renderings of spherical shells with radius 2 nm and opaqueness  $\alpha = 0.4$ . The  $z$ -coordinate is color coded and the grids have a spacing of 50 nm.

## 3.6 Summary and discussion

The aim of this chapter was to validate deep learning as a fast, parameter-free method for 3D position estimation in MINFLUX nanoscopy. Furthermore, we wanted to explore in which circumstances a neural network could be superior to the existing MLE in terms of localization performance. To this end, I constructed a neural network setting its hyperparameters appropriately and trained it with simulated 3D MINFLUX data. Next, I tuned both the network architecture and the training design in order to find a trade-off between learning time and localization error. At first, I simulated emission events under optimal conditions, namely a large number of photons per localization and negligible background. Afterwards, I quantified the localization precision and bias of the neural network and the MLE on simulated data as a function of the number of detected photons and the SBR. Finally, I compared both estimators regarding their localization performance, computation time, and image quality based on the evaluation of experimental MINFLUX data of Nup96.

The comparison of neural networks and the MLE on simulated data presented in Figure 3.10 reveals a similar localization error for large photon numbers or high SBR. Keeping in mind that the MLE is asymptotically efficient, we did not expect the localization performance to improve in this regime. Under more challenging conditions, however, the new position estimators outperform the existing one. This behavior is not surprising because neural networks often develop their full potential when the performance of traditional approaches starts to degrade. We found these observations confirmed in the localization precision and image quality of experimental data shown in Figures 3.12 and 3.15.

Figure 3.13 reveals an offset between position estimates made by the MLE and by the neural network at the edge of the training volume. The values

of this offset are much higher than the ones presented in Figures 3.5 and 3.6, where the neural network shows an inward-directed bias in the outer regions. This could be explained by the significantly lower SBR of the experimental data used in Figure 3.13. Moreover, the true positions of the fluorescent emitters are unknown for this data set and might not be identical to the ground truth estimates provided by the MLE. In any case the bias of the deep-learning model could be reduced by increasing the grid density towards the edge of the training volume.

In general, the size of the training and validation data is a parameter that may be set according to the computational resources and how important slight improvements in localization performance are. With our current choice of around 800 000 samples for trainings with fixed  $N$ , we have already reached a point where we expect only minor enhancements in localization performance from adding more data points in the high-photon regime. Judging by Figure 3.9, these marginal improvements are likely to come at the cost of a much longer training time. In the low-photon regime, however, we suggest increasing the number of training samples based on the results shown in Figure B.4 to reduce the standard deviation of the localizations.

The results shown in Figures 3.10, B.3, and B.4 indicate that the universal network is prone to overfitting in the high-photon, high-SBR regime and to underfitting for small count numbers. A heterogeneous composition of the training and validation data might solve this problem. At the same time, the comparison of different epochs in Figure B.3 implies that due to overfitting effects, the epoch determined by the early-stopping criterion does not always provide the lowest localization bias for all parameter settings. Therefore, it is advisable to make the final choice of the best epoch by examining the localization performance metrics in further detail once the training is complete.

In the comparison of two neural nets in Figure 3.10, the specialized network shows a total error that is dominated by bias in the low-photon and low-SBR regimes, whereas the universal network mostly keeps the bias small. This could be explained by the fact that increasing the bias is generally a good way to reduce the total error under non-asymptotic conditions [101]. The universal network, however, needs to find a single solution to minimize the error under all conditions. This strategy appears to be similar to the asymptotically efficient MLE in the sense that both aim to eliminate bias.

Regarding our final choice of network architecture, further optimization to lower the standard deviation and bias of the localizations under certain conditions might still be possible. On the one hand, different architectures yielded similar outcomes in the high-photon, low-background regime as demonstrated in Figures 3.7 and 3.8. On the other hand, only one of these designs, namely the one featuring two hidden layers with 500 neurons each, was tested using multiple  $N$  and  $\text{SBR}(\mathbf{0}, L_0)$  parameter values. Thus, a different architecture might be more suitable for challenging conditions or for covering a wider range of parameter settings. In general, we noticed that networks of higher complexity do not necessarily require more training time because SGD takes fewer epochs to converge. By employing an early-stopping criterion, I prevented the models from overfitting. However, it is still imaginable that a larger network or extended learning in combination with another regularization approach, such as weight decay or dropout, might further reduce the test error.

In the evaluation of experimental data, the role of filtering is ambiguous: It can improve the general localization precision by discarding false-positive and ambiguous events, but this comes at the cost of losing valid localizations. For high amounts of detected photons, the two populations of true and false-positive events can be easily separated by a  $p_0$  filter. With decreasing photon numbers, an additional threshold on the distance  $r_{\text{est}}$  of the estimated positions to the EBP center becomes important [30]. Figure

3.12 reveals that in the high-photon regime, both estimators benefit from  $p_0$  filtering. However, it is not entirely clear why the localization precision of the neural network does not depend on setting this threshold in the case of lower photon numbers. It seems as if even events localized at the edge of the EBP according to the MLE are assigned a position closer to the center with high precision. As the neural network has only been trained within a limited area of radius  $r_{\text{train}} = 0.4L$ , we expect it to map all events onto positions within this region. This inhibits removing false-positive events with a filter on  $r_{\text{est}}$ . In order to better control the quality of the predictions made by the neural network, the training data set should also include events far outside the EBP in the future.



## 4 Conclusion and outlook

In Chapter 2 of this thesis, I developed an approach to quantify the theoretically achievable localization precision of 2D MINFLUX illumination geometries, and more specifically of Lissajous patterns. The analysis was based on previous investigations into Lissajous illumination geometries and into excitation pattern optimization using a loss function defined on a suitable ROI [44, 48]. I extended these studies by taking into account the aspect ratio of the Lissajous curve as another geometrical parameter, the background level, and different loss functions. The calculations presented in this subproject focused on performance metrics related to the CRB matrix  $\Sigma_{\text{CRB}}$ , namely its eigenvalues and isotropy, that yield the extent and shape of the localization uncertainty. I demonstrated low-order Lissajous figures with the shape of a figure of eight to achieve a substantially more isotropic CRB than the published triangular excitation pattern [29].

In order to achieve an even higher isotropy, it would be possible to rotate the pattern between the illumination cycles. Moreover, it might be worth calculating the effect of omitting the second central exposure on the theoretically achievable localization precision. Modulating the excitation intensity along the beam trajectory could be another way to shift the ratio of radial and angular information as suggested by Gwosch [44]. Finally, an extension of the presented concept to 3D and an integration into the previously published iterative MINFLUX scheme [30] would be of interest. The results obtained in this thesis are of particular importance for setting

the ideal illumination geometries as a function of the ROI in the different iteration steps.

In Chapter 2 all calculations were performed under the assumption of a stationary illumination beam without taking the scanner movement during the exposure into account. Thus, quantifying the effect of the beam displacement on the CRB as a function of the illumination time and the scanning speed would be an important next step. In an entirely different approach, photons could also be acquired continuously along the Lissajous path instead of exciting the emitters only at discrete positions. Such a continuous illumination strategy was employed in the recently published MINSTED concept [102].

In conclusion, I demonstrated that Lissajous figures are suitable patterns for 2D MINFLUX illumination. They could be implemented in an experimental setup using galvanometric mirrors instead of EODs for scanning the excitation beam around the fluorescent emitter as suggested in Section 2.1.1. This would reduce the complexity and the cost of MINFLUX systems and therefore enable a broader range of users to access this technology.

In Chapter 3 I established the foundations for exploiting deep learning in MINFLUX position estimation. I designed and trained a fully connected neural network to predict the 3D positions of fluorescent emitters based on the count sets extracted from the photon traces. Tested on simulated data under various conditions, the new position estimator produced a smaller localization error than the currently used MLE especially in the low-photon and high-background regimes. On experimental MINFLUX data of Nup96, a similar localization uncertainty was obtained with both estimators, but unlike the MLE the precision of the neural network did not rely on prior removal of false-positive events. However, the deep-learning algorithm exhibited an inward-directed bias at the edge of the training volume. This deviation from the target position was on the order of 1 nm for a high SBR

---

and tended to grow with background. Moreover, photon counts detected from molecules located outside the illumination geometry were assigned an origin inside of the training space.

The localization error of the neural network could possibly be reduced even further by providing more training data in the non-asymptotic regime, that is, for low photon counts and high background. It might also be worth considering replacing the MSE by a more sophisticated loss function that puts the error in relation to the CRB. In general, including more information in the loss function can improve the performance of deep-learning algorithms by giving more structure to the data. Most importantly, for an enhanced reliability on experimental data, the neural network should also be trained with data points far outside the excitation pattern. However, the theoretically achievable localization precision, which is fundamentally limited by the CRB, is relatively low in these areas. This can be seen in Figure 2.9 for example. Thus, in order to avoid undue trust in the predictions of the model, a fourth network output may be added to serve as a quality indicator. Of course, simulated training and validation data would have to be labeled accordingly. In the end, this additional output variable could serve as a criterion for filtering the predictions made by the model.

In the future, a neural network could replace further steps of the MIN-FLUX post-processing workflow, notably the computationally demanding segmentation of the photon-count traces currently performed by a hidden Markov model. Predicting the emitter positions directly from the raw data could have several advantages: Photon counts that are currently discarded when splitting events at predefined total photon numbers could be saved; the SBR may be estimated from the trace; count rate steps could be detected in the trace and allow us to discern simultaneously activated emitters; and, last but not least, a much faster prediction is expected from a trained deep-learning model. While the position estimation based on a set of accumulated photon counts is a relatively simple problem, deep learning

could develop its full potential when extracting the complete information encoded in the raw data. Most importantly, an analysis workflow based on a trained neural network would be largely parameter-free and would not require any particular user expertise.

In summary, I was able to increase the information content of MINFLUX images and consequently make better use of the limited photon budget. The studied Lissajous illumination patterns enable a more isotropic localization precision than the existing geometry for 2D MINFLUX. In addition, the neural network developed in this work extends precise position estimation to low-photon and high-background regimes. Both approaches established in this thesis reduce the complexity of MINFLUX nanoscopy. The illumination schemes studied herein are ideal for the implementation with simple scanning systems, and deep learning reduces both the computational effort and the number of parameters in post-processing. Overall, the methods proposed in this work will facilitate access to MINFLUX nanoscopy for the broader research community.

## Bibliography

- [1] J. Huisken, J. Swoger, F. Del Bene, J. Wittbrodt, and E. H. Stelzer. “Optical sectioning deep inside live embryos by selective plane illumination microscopy”. In: *Science* 305.5686 (2004), pp. 1007–9. DOI: [10.1126/science.1100035](https://doi.org/10.1126/science.1100035).
- [2] P. J. Keller, A. D. Schmidt, J. Wittbrodt, and E. H. Stelzer. “Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy”. In: *Science* 322.5904 (2008), pp. 1065–9. DOI: [10.1126/science.1162493](https://doi.org/10.1126/science.1162493).
- [3] K. Xu, G. Zhong, and X. Zhuang. “Actin, spectrin, and associated proteins form a periodic cytoskeletal structure in axons”. In: *Science* 339.6118 (2013), pp. 452–6. DOI: [10.1126/science.1232251](https://doi.org/10.1126/science.1232251).
- [4] E. D’Este, D. Kamin, F. Gottfert, A. El-Hady, and S. W. Hell. “STED nanoscopy reveals the ubiquity of subcortical cytoskeleton periodicity in living neurons”. In: *Cell Rep* 10.8 (2015), pp. 1246–51. DOI: [10.1016/j.celrep.2015.02.007](https://doi.org/10.1016/j.celrep.2015.02.007).
- [5] R. Schmidt, C. A. Wurm, A. Punge, A. Egner, S. Jakobs, and S. W. Hell. “Mitochondrial Cristae Revealed with Focused Light”. In: *Nano Lett* 9.6 (2009), pp. 2508–2510. DOI: [10.1021/nl901398t](https://doi.org/10.1021/nl901398t).
- [6] S. H. Shim, C. Xia, G. Zhong, H. P. Babcock, J. C. Vaughan, B. Huang, X. Wang, C. Xu, G. Q. Bi, and X. Zhuang. “Super-resolution fluorescence imaging of organelles in live cells with photoswitchable membrane probes”. In: *Proc Natl Acad Sci U S A* 109.35 (2012), pp. 13978–83. DOI: [10.1073/pnas.1201882109](https://doi.org/10.1073/pnas.1201882109).

- [7] D. C. Jans, C. A. Wurm, D. Riedel, D. Wenzel, F. Stagge, M. Deckers, P. Rehling, and S. Jakobs. “STED super-resolution microscopy reveals an array of MINOS clusters along human mitochondria”. In: *Proc Natl Acad Sci U S A* 110.22 (2013), pp. 8936–41. DOI: [10.1073/pnas.1301820110](https://doi.org/10.1073/pnas.1301820110).
- [8] A. Szymborska, A. de Marco, N. Daigle, V. C. Cordes, J. A. Briggs, and J. Ellenberg. “Nuclear pore scaffold structure analyzed by super-resolution microscopy and particle averaging”. In: *Science* 341.6146 (2013), pp. 655–8. DOI: [10.1126/science.1240672](https://doi.org/10.1126/science.1240672).
- [9] E. Abbe. “Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung”. In: *Arch Mikrosk Anat* 9 (1873), pp. 413–468. DOI: [10.1007/BF02956173](https://doi.org/10.1007/BF02956173).
- [10] S. W. Hell and J. Wichmann. “Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy”. In: *Opt Lett* 19.11 (1994), pp. 780–2. DOI: [10.1364/ol.19.000780](https://doi.org/10.1364/ol.19.000780).
- [11] T. A. Klar and S. W. Hell. “Subdiffraction resolution in far-field fluorescence microscopy”. In: *Opt Lett* 24.14 (1999), pp. 954–6. DOI: [10.1364/ol.24.000954](https://doi.org/10.1364/ol.24.000954).
- [12] T. A. Klar, S. Jakobs, M. Dyba, A. Egner, and S. W. Hell. “Fluorescence microscopy with diffraction resolution barrier broken by stimulated emission”. In: *Proc Natl Acad Sci U S A* 97.15 (2000), pp. 8206–10. DOI: [10.1073/pnas.97.15.8206](https://doi.org/10.1073/pnas.97.15.8206).
- [13] M. J. Rust, M. Bates, and X. Zhuang. “Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)”. In: *Nat Methods* 3.10 (2006), pp. 793–5. DOI: [10.1038/nmeth929](https://doi.org/10.1038/nmeth929).
- [14] E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess. “Imaging intracellular fluorescent proteins at nanometer res-

- 
- olution”. In: *Science* 313.5793 (2006), pp. 1642–5. DOI: [10.1126/science.1127344](https://doi.org/10.1126/science.1127344).
- [15] S. T. Hess, T. P. Girirajan, and M. D. Mason. “Ultra-high resolution imaging by fluorescence photoactivation localization microscopy”. In: *Biophys J* 91.11 (2006), pp. 4258–72. DOI: [10.1529/biophysj.106.091116](https://doi.org/10.1529/biophysj.106.091116).
- [16] F. Göttfert, C. A. Wurm, V. Mueller, S. Berning, V. C. Cordes, A. Honigmann, and S. W. Hell. “Coaligned Dual-Channel STED Nanoscopy and Molecular Diffusion Analysis at 20 nm Resolution”. In: *Biophys J* 105.1 (2013), pp. L1–L3. DOI: [10.1016/j.bpj.2013.05.029](https://doi.org/10.1016/j.bpj.2013.05.029).
- [17] A. J. Wollman, R. Nudd, E. G. Hedlund, and M. C. Leake. “From Animaculum to single molecules: 300 years of the light microscope”. In: *Open Biol* 5.4 (2015), p. 150019. DOI: [10.1098/rsob.150019](https://doi.org/10.1098/rsob.150019).
- [18] P. O. Bayguinov, D. M. Oakley, C. C. Shih, D. J. Geanon, M. S. Joens, and J. A. J. Fitzpatrick. “Modern Laser Scanning Confocal Microscopy”. In: *Curr Protoc Cytom* 85.1 (2018), e39. DOI: [10.1002/cpcy.39](https://doi.org/10.1002/cpcy.39).
- [19] S. W. Hell, S. J. Sahl, M. Bates, X. Zhuang, R. Heintzmann, M. J. Booth, J. Bewersdorf, G. Shtengel, H. Hess, P. Tinnefeld, A. Honigmann, S. Jakobs, I. Testa, L. Cognet, B. Lounis, H. Ewers, S. J. Davis, C. Eggeling, D. Klenerman, K. I. Willig, G. Vicidomini, M. Castello, A. Diaspro, and T. Cordes. “The 2015 super-resolution microscopy roadmap”. In: *J Phys D Appl Phys* 48.44 (2015), p. 443001. DOI: [10.1088/0022-3727/48/44/443001](https://doi.org/10.1088/0022-3727/48/44/443001).
- [20] A. Keppler, S. Gendreizig, T. Gronemeyer, H. Pick, H. Vogel, and K. Johnsson. “A general method for the covalent labeling of fusion proteins with small molecules in vivo”. In: *Nat Biotechnol* 21.1 (2003), pp. 86–9. DOI: [10.1038/nbt765](https://doi.org/10.1038/nbt765).

- [21] G. V. Los, L. P. Encell, M. G. McDougall, D. D. Hartzell, N. Karassina, C. Zimprich, M. G. Wood, R. Learish, R. F. Ohana, M. Urh, D. Simpson, J. Mendez, K. Zimmerman, P. Otto, G. Vidugiris, J. Zhu, A. Darzins, D. H. Klaubert, R. F. Bulleit, and K. V. Wood. “HaloTag: a novel protein labeling technology for cell imaging and protein analysis”. In: *ACS Chem Biol* 3.6 (2008), pp. 373–82. DOI: [10.1021/cb800025k](https://doi.org/10.1021/cb800025k).
- [22] A. Gautier, A. Juillerat, C. Heinis, Jr. Correa I. R., M. Kindermann, F. Beaufils, and K. Johnsson. “An engineered protein tag for multiprotein labeling in living cells”. In: *Chem Biol* 15.2 (2008), pp. 128–36. DOI: [10.1016/j.chembiol.2008.01.007](https://doi.org/10.1016/j.chembiol.2008.01.007).
- [23] J. Ries, C. Kaplan, E. Platonova, H. Eghlidi, and H. Ewers. “A simple, versatile method for GFP-based super-resolution microscopy via nanobodies”. In: *Nat Methods* 9.6 (2012), pp. 582–584. DOI: [10.1038/nmeth.1991](https://doi.org/10.1038/nmeth.1991).
- [24] T. Pleiner, M. Bates, S. Trakhanov, C. T. Lee, J. E. Schliep, H. Chug, M. Böhning, H. Stark, H. Urlaub, and D. Görlich. “Nanobodies: site-specific labeling for super-resolution imaging, rapid epitope-mapping and native protein complex isolation”. In: *Elife* 4 (2015). DOI: [10.7554/eLife.11349](https://doi.org/10.7554/eLife.11349).
- [25] G. T. Dempsey, M. Bates, W. E. Kowtoniuk, D. R. Liu, R. Y. Tsien, and X. Zhuang. “Photoswitching Mechanism of Cyanine Dyes”. In: *J Am Chem Soc* 131.51 (2009), pp. 18192–18193. DOI: [10.1021/ja904588g](https://doi.org/10.1021/ja904588g).
- [26] G. Lukinavicius, K. Umezawa, N. Olivier, A. Honigmann, G. Y. Yang, T. Plass, V. Mueller, L. Reymond, I. R. Correa, Z. G. Luo, C. Schultz, E. A. Lemke, P. Heppenstall, C. Eggeling, S. Manley, and K. Johnsson. “A near-infrared fluorophore for live-cell super-resolution microscopy of cellular proteins”. In: *Nat Chem* 5.2 (2013), pp. 132–139. DOI: [10.1038/Nchem.1546](https://doi.org/10.1038/Nchem.1546).

- 
- [27] G. Lukinavičius, L. Reymond, E. D’Este, A. Masharina, F. Göttfert, H. Ta, A. Güther, M. Fournier, S. Rizzo, H. Waldmann, C. Blaukopf, C. Sommer, D. W. Gerlich, H.-D. Arndt, S. W. Hell, and K. Johnsson. “Fluorogenic probes for live-cell imaging of the cytoskeleton”. In: *Nat Methods* 11.7 (2014), pp. 731–733. DOI: [10.1038/nmeth.2972](https://doi.org/10.1038/nmeth.2972).
- [28] K. Kolmakov, C. A. Wurm, D. N. Meineke, F. Göttfert, V. P. Boryarskiy, V. N. Belov, and S. W. Hell. “Polar Red-Emitting Rhodamine Dyes with Reactive Groups: Synthesis, Photophysical Properties, and Two-Color STED Nanoscopy Applications”. In: *Chem Eur J* 20.1 (2014), pp. 146–157. DOI: [10.1002/chem.201303433](https://doi.org/10.1002/chem.201303433).
- [29] F. Balzarotti, Y. Eilers, K. C. Gwosch, A. H. Gynnå, V. Westphal, Stefani F. D., J. Elf, and S. W. Hell. “Nanometer resolution imaging and tracking of fluorescent molecules with minimal photon fluxes”. In: *Science* 355.6325 (2017), pp. 606–612. DOI: [10.1126/science.aak9913](https://doi.org/10.1126/science.aak9913).
- [30] K. C. Gwosch, J. K. Pape, F. Balzarotti, P. Hoess, J. Ellenberg, J. Ries, and S. W. Hell. “MINFLUX nanoscopy delivers 3D multicolor nanometer resolution in cells”. In: *Nat Methods* 17.2 (2020), pp. 217–224. DOI: [10.1038/s41592-019-0688-0](https://doi.org/10.1038/s41592-019-0688-0).
- [31] Lord Rayleigh Sec. R. S. “XV. On the theory of optical images, with special reference to the microscope”. In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 42.255 (1896), pp. 167–195. DOI: [10.1080/14786449608620902](https://doi.org/10.1080/14786449608620902).
- [32] S. W. Hell. “Far-Field Optical Nanoscopy”. In: *Science* 316.5828 (2007), pp. 1153–1158. DOI: [10.1126/science.1137395](https://doi.org/10.1126/science.1137395).
- [33] S. W. Hell, S. Jakobs, and L. Kastrup. “Imaging and writing at the nanoscale with focused visible light through saturable optical transitions”. In: *Appl Phys A Mater Sci Process* 77.7 (2003), pp. 859–860. DOI: [10.1007/s00339-003-2292-4](https://doi.org/10.1007/s00339-003-2292-4).

- [34] M. Hofmann, C. Eggeling, S. Jakobs, and S. W. Hell. “Breaking the diffraction barrier in fluorescence microscopy at low light intensities by using reversibly photoswitchable proteins”. In: *Proc Natl Acad Sci U S A* 102.49 (2005), pp. 17565–17569. DOI: [10.1073/pnas.0506010102](https://doi.org/10.1073/pnas.0506010102).
- [35] T. Grotjohann, I. Testa, M. Leutenegger, H. Bock, N. T. Urban, F. Lavoie-Cardinal, K. I. Willig, C. Eggeling, S. Jakobs, and S. W. Hell. “Diffraction-unlimited all-optical imaging and writing with a photochromic GFP”. In: *Nature* 478.7368 (2011), pp. 204–208. DOI: [10.1038/nature10497](https://doi.org/10.1038/nature10497).
- [36] V. Westphal and S. W. Hell. “Nanoscale Resolution in the Focal Plane of an Optical Microscope”. In: *Phys Rev Lett* 94 (2005), p. 143903. DOI: [10.1103/PhysRevLett.94.143903](https://doi.org/10.1103/PhysRevLett.94.143903).
- [37] W. Heisenberg. *The Physical Principles of the Quantum Theory*. Chicago, IL: The University of Chicago Press, 1930.
- [38] N. Bobroff. “Position measurement with a resolution and noise-limited instrument”. In: *Rev Sci Instrum* 57.6 (1986), pp. 1152–1157. DOI: [10.1063/1.1138619](https://doi.org/10.1063/1.1138619).
- [39] R. E. Thompson, D. R. Larson, and W. W. Webb. “Precise Nanometer Localization Analysis for Individual Fluorescent Probes”. In: *Biophys J* 82.5 (2002), pp. 2775–2783. DOI: [10.1016/S0006-3495\(02\)75618-X](https://doi.org/10.1016/S0006-3495(02)75618-X).
- [40] K. I. Mortensen, L. S. Churchman, J. A. Spudich, and H. Flyvbjerg. “Optimized localization analysis for single-molecule tracking and super-resolution microscopy”. In: *Nat Methods* 7.5 (2010), pp. 377–381. DOI: [10.1038/nmeth.1447](https://doi.org/10.1038/nmeth.1447).
- [41] K. C. Gwosch, J. K. Pape, F. Balzarotti, P. Hoess, J. Ellenberg, J. Ries, and S. W. Hell. “MINFLUX nanoscopy delivers multi-color nanometer 3D-resolution in (living) cells”. In: *bioRxiv* 734251 (2019). DOI: [10.1101/734251](https://doi.org/10.1101/734251).

- 
- [42] J. K. Pape, T. Stephan, F. Balzarotti, R. Büchner, F. Lange, D. Riedel, S. Jakobs, and S. W. Hell. “Multicolor 3D MINFLUX nanoscopy of mitochondrial MICOS proteins”. In: *Proc Natl Acad Sci U S A* 117.34 (2020), pp. 20607–20614. DOI: [10.1073/pnas.2009364117](https://doi.org/10.1073/pnas.2009364117).
- [43] R. Schmidt, T. Weihs, C. A. Wurm, I. Jansen, J. Rehman, S. J. Sahl, and S. W. Hell. “MINFLUX nanometer-scale 3D imaging and microsecond-range tracking on a common fluorescence microscope”. In: *Nat Commun* 12.1 (2021), p. 1478. DOI: [10.1038/s41467-021-21652-z](https://doi.org/10.1038/s41467-021-21652-z).
- [44] K. C. Gwosch. “Nanometer resolution imaging with minimal fluorescence photon fluxes”. PhD thesis. 2017.
- [45] R. A. Fisher and E. J. Russell. “On the mathematical foundations of theoretical statistics”. In: *Philos Trans R Soc Lond A* 222.594-604 (1922), pp. 309–368. DOI: [10.1098/rsta.1922.0009](https://doi.org/10.1098/rsta.1922.0009).
- [46] C. R. Rao. “Information and the accuracy attainable in the estimation of statistical parameters”. In: *Bull Calcutta Math Soc* 37 (1945), pp. 81–91.
- [47] H. Cramér. *Mathematical methods of statistics*. Princeton mathematical series 9. Princeton, NJ: Princeton University Press, 1946.
- [48] Y. Eilers, H. Ta, K. C. Gwosch, F. Balzarotti, and S. W. Hell. “MINFLUX monitors rapid molecular jumps with superior spatiotemporal resolution”. In: *Proc Natl Acad Sci U S A* 115.24 (2018), pp. 6117–6122. DOI: [10.1073/pnas.1801672115](https://doi.org/10.1073/pnas.1801672115).
- [49] J. A. Lissajous. “Mémoire sur l’Etude optique des mouvements vibratoires”. In: *Ann Chim Phys* 3.51 (1857).
- [50] J. K. Pape. “Multicolor 3D MINFLUX nanoscopy for biological imaging”. PhD thesis. 2020.

- [51] Y. Bengio, A. Courville, and P. Vincent. “Representation Learning: A Review and New Perspectives”. In: *IEEE Trans Pattern Anal Mach Intell* 35.8 (2013), pp. 1798–1828. DOI: [10.1109/TPAMI.2013.50](https://doi.org/10.1109/TPAMI.2013.50).
- [52] Y. LeCun, Y. Bengio, and G. Hinton. “Deep learning”. In: *Nature* 521.7553 (2015), pp. 436–444. DOI: [10.1038/nature14539](https://doi.org/10.1038/nature14539).
- [53] M. Leshno, V. Y. Lin, A. Pinkus, and S. Schocken. “Multilayer feedforward networks with a nonpolynomial activation function can approximate any function”. In: *Neural Netw* 6.6 (1993), pp. 861–867. DOI: [10.1016/S0893-6080\(05\)80131-5](https://doi.org/10.1016/S0893-6080(05)80131-5).
- [54] F. Rosenblatt. *The perceptron: A perceiving and recognizing automaton*. Research rep. Project Para 85-460-1. Buffalo, NY: Cornell Aeronautical Laboratory, 1957.
- [55] M. Minsky and S. Papert. *Perceptrons: an introduction to computational geometry*. Cambridge, MA: MIT Press, 1969.
- [56] Y. Lecun. “A learning scheme for asymmetric threshold networks”. In: *Proceedings of Cognitiva 85* (1985), pp. 599–604.
- [57] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. “Learning representations by back-propagating errors”. In: *Nature* 323.6088 (1986), pp. 533–536. DOI: [10.1038/323533a0](https://doi.org/10.1038/323533a0).
- [58] L. Möckl, A. R. Roy, and W. E. Moerner. “Deep learning in single-molecule microscopy: fundamentals, caveats, and recent developments [Invited]”. In: *Biomed Opt Express* 11.3 (2020), pp. 1633–1661. DOI: [10.1364/BOE.386361](https://doi.org/10.1364/BOE.386361).
- [59] A. Krizhevsky, I. Sutskever, and G. Hinton. “ImageNet Classification with Deep Convolutional Neural Networks”. In: *Adv Neural Inf Process Syst* 25 (2012). DOI: [10.1145/3065386](https://doi.org/10.1145/3065386).

- 
- [60] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis. “Mastering the game of Go with deep neural networks and tree search”. In: *Nature* 529.7587 (2016), pp. 484–489. DOI: [10.1038/nature16961](https://doi.org/10.1038/nature16961).
- [61] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, T. Graepel, and D. Hassabis. “Mastering the game of Go without human knowledge”. In: *Nature* 550.7676 (2017), pp. 354–359. DOI: [10.1038/nature24270](https://doi.org/10.1038/nature24270).
- [62] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. “Backpropagation Applied to Handwritten Zip Code Recognition”. In: *Neural Comput* 1.4 (1989), pp. 541–551. DOI: [10.1162/neco.1989.1.4.541](https://doi.org/10.1162/neco.1989.1.4.541).
- [63] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. “Gradient-based learning applied to document recognition”. In: *Proc IEEE Inst Electr Electron Eng* 86.11 (1998), pp. 2278–2324. DOI: [10.1109/5.726791](https://doi.org/10.1109/5.726791).
- [64] S.-C. B. Lo, S.-L. A. Lou, J.-S. Lin, M. T. Freedman, M. V. Chien, and S. K. Mun. “Artificial convolution neural network techniques and applications for lung nodule detection”. In: *IEEE Trans Med Imaging* 14.4 (1995), pp. 711–718. DOI: [10.1109/42.476112](https://doi.org/10.1109/42.476112).
- [65] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez. “A survey on deep learning in medical image analysis”. In: *Med Image Anal* 42 (2017), pp. 60–88. DOI: [10.1016/j.media.2017.07.005](https://doi.org/10.1016/j.media.2017.07.005).

- [66] E. Moen, D. Bannon, T. Kudo, W. Graf, M. Covert, and D. Van Valen. “Deep learning for cellular image analysis”. In: *Nat Methods* 16.12 (2019), pp. 1233–1246. DOI: [10.1038/s41592-019-0403-1](https://doi.org/10.1038/s41592-019-0403-1).
- [67] F. Huang, S. L. Schwartz, J. M. Byars, and K. A. Lidke. “Simultaneous multiple-emitter fitting for single molecule super-resolution imaging”. In: *Biomed Opt Express* 2.5 (2011), pp. 1377–1393. DOI: [10.1364/BOE.2.001377](https://doi.org/10.1364/BOE.2.001377).
- [68] J. Min, C. Vonesch, H. Kirshner, L. Carlini, N. Olivier, S. Holden, S. Manley, J. C. Ye, and M. Unser. “FALCON: fast and unbiased reconstruction of high-density super-resolution microscopy data”. In: *Sci Rep* 4.1 (2014), p. 4577. DOI: [10.1038/srep04577](https://doi.org/10.1038/srep04577).
- [69] D. Sage, H. Kirshner, T. Pengo, N. Stuurman, J. Min, S. Manley, and M. Unser. “Quantitative evaluation of software packages for single-molecule localization microscopy”. In: *Nat Methods* 12.8 (2015), pp. 717–724. DOI: [10.1038/nmeth.3442](https://doi.org/10.1038/nmeth.3442).
- [70] R. J. Marsh, K. Pfisterer, P. Bennett, L. M. Hirvonen, M. Gautel, G. E. Jones, and S. Cox. “Artifact-free high-density localization microscopy analysis”. In: *Nat Methods* 15.9 (2018), pp. 689–692. DOI: [10.1038/s41592-018-0072-5](https://doi.org/10.1038/s41592-018-0072-5).
- [71] E. Nehme, L. E. Weiss, T. Michaeli, and Y. Shechtman. “Deep-STORM: super-resolution single-molecule microscopy by deep learning”. In: *Optica* 5.4 (2018), pp. 458–464. DOI: [10.1364/Optica.5.000458](https://doi.org/10.1364/Optica.5.000458).
- [72] P. Zhang, S. Liu, A. Chaurasia, D. Ma, M. J. Mlodzianoski, E. Cukurciello, and F. Huang. “Analyzing complex single-molecule emission patterns with deep learning”. In: *Nat Methods* 15.11 (2018), pp. 913–916. DOI: [10.1038/s41592-018-0153-5](https://doi.org/10.1038/s41592-018-0153-5).
- [73] N. Boyd, E. Jonas, H. Babcock, and B. Recht. “DeepLoco: Fast 3D Localization Microscopy Using Neural Networks”. In: *bioRxiv* 267096 (2018). DOI: [10.1101/267096](https://doi.org/10.1101/267096).

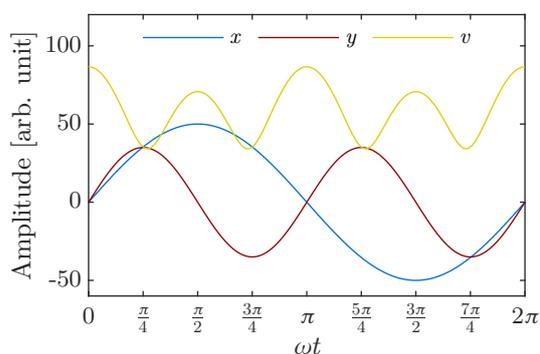
- 
- [74] P. Zelger, K. Kaser, B. Rossboth, L. Velas, G. J. Schutz, and A. Jesacher. “Three-dimensional localization microscopy using deep learning”. In: *Opt Express* 26.25 (2018), pp. 33166–33179. DOI: [10.1364/OE.26.033166](https://doi.org/10.1364/OE.26.033166).
- [75] A. Speiser, L.-R. Müller, U. Matti, C. J. Obara, W. R. Legant, J. Ries, J. H. Macke, and S. C. Turaga. “Teaching deep neural networks to localize single molecules for super-resolution microscopy”. In: *arXiv* 1907.00770 (2020). <https://arxiv.org/abs/1907.00770v2>.
- [76] E. Nehme, D. Freedman, R. Gordon, B. Ferdman, L. E. Weiss, O. Alalouf, T. Naor, R. Orange, T. Michaeli, and Y. Shechtman. “DeepSTORM3D: dense 3D localization microscopy and PSF design by deep learning”. In: *Nat Methods* 17.7 (2020), pp. 734–740. DOI: [10.1038/s41592-020-0853-5](https://doi.org/10.1038/s41592-020-0853-5).
- [77] T. Kim, S. Moon, and K. Xu. “Information-rich localization microscopy through machine learning”. In: *Nat Commun* 10.1 (2019), p. 1996. DOI: [10.1038/s41467-019-10036-z](https://doi.org/10.1038/s41467-019-10036-z).
- [78] E. Hershko, L. E. Weiss, T. Michaeli, and Y. Shechtman. “Multicolor localization microscopy and point-spread-function engineering by deep learning”. In: *Opt Express* 27.5 (2019), pp. 6158–6183. DOI: [10.1364/OE.27.006158](https://doi.org/10.1364/OE.27.006158).
- [79] J. Xu, G. Qin, F. Luo, L. Wang, R. Zhao, N. Li, J. Yuan, and X. Fang. “Automated Stoichiometry Analysis of Single-Molecule Fluorescence Imaging Traces via Deep Learning”. In: *J Am Chem Soc* 141.17 (2019), pp. 6976–6985. DOI: [10.1021/jacs.9b00688](https://doi.org/10.1021/jacs.9b00688).
- [80] L. Möckl, A. R. Roy, P. N. Petrov, and W. E. Moerner. “Accurate and rapid background estimation in single-molecule localization microscopy using the deep neural network BGnet”. In: *Proc Natl Acad Sci U S A* 117.1 (2020), pp. 60–67. DOI: [10.1073/pnas.1916219117](https://doi.org/10.1073/pnas.1916219117).

- [81] W. K. Newey and D. McFadden. “Large sample estimation and hypothesis testing”. In: *Handbook of Econometrics*. Vol. 4. Elsevier, 1994. Chap. 36, pp. 2111–2245. DOI: [10.1016/S1573-4412\(05\)80005-4](https://doi.org/10.1016/S1573-4412(05)80005-4).
- [82] J. G. Kalbfleisch. *Probability and Statistical Inference*. 1st ed. Universitext. New York, NY: Springer, 1979.
- [83] P. R. Halmos. *Naive Set Theory*. Undergraduate Texts in Mathematics. New York, NY: Springer, 1974.
- [84] Francois Chollet. *Deep Learning with Python*. 1st ed. Manning Publications, 2017. DOI: [10.5555/3203489](https://doi.org/10.5555/3203489).
- [85] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. <http://www.deeplearningbook.org>. MIT Press, 2016.
- [86] K. Jarrett, K. Kavukcuoglu, M.A. Ranzato, and Y. LeCun. “What is the best multi-stage architecture for object recognition?” In: *2009 IEEE 12th International Conference on Computer Vision (2009)*, pp. 2146–2153. DOI: [10.1109/ICCV.2009.5459469](https://doi.org/10.1109/ICCV.2009.5459469).
- [87] V. Nair and G. E. Hinton. “Rectified Linear Units Improve Restricted Boltzmann Machines”. In: *Proceedings of the 27th International Conference on Machine Learning*. Madison, WI: Omnipress, 2010, pp. 807–814. DOI: [10.5555/3104322.3104425](https://doi.org/10.5555/3104322.3104425).
- [88] X. Glorot, A. Bordes, and Y. Bengio. “Deep Sparse Rectifier Neural Networks”. In: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. Ed. by G. Gordon, D. Dunson, and M. Dudík. Vol. 15. Fort Lauderdale, FL: Proceedings of Machine Learning Research, 2011, pp. 315–323.
- [89] Y. Bengio. “Practical Recommendations for Gradient-Based Training of Deep Architectures”. In: *Neural Networks: Tricks of the Trade*. Ed. by G. Montavon, G. B. Orr, and K.-R. Müller. 2nd ed. Berlin, Heidelberg: Springer, 2012, pp. 437–478. DOI: [10.1007/978-3-642-35289-8\\_26](https://doi.org/10.1007/978-3-642-35289-8_26).

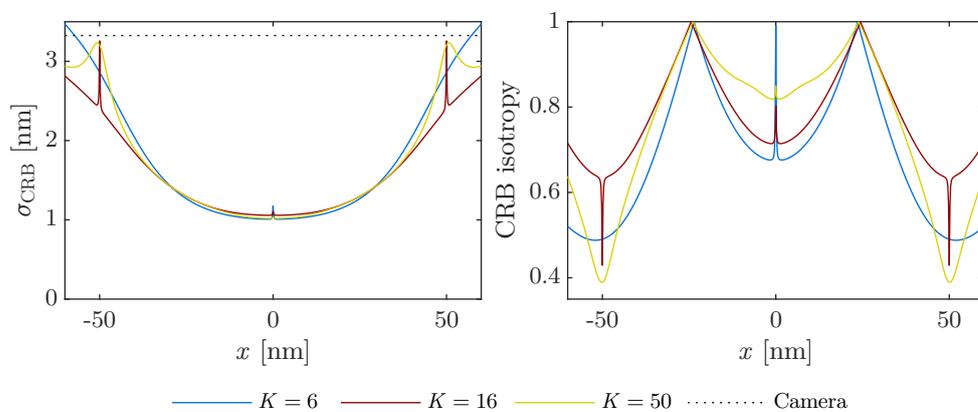
- 
- [90] H. Robbins and S. Monro. “A Stochastic Approximation Method”. In: *Ann Math Stat* 22.3 (1951), pp. 400–407. DOI: [10.1214/aoms/1177729586](https://doi.org/10.1214/aoms/1177729586).
- [91] L. Bottou. “On-Line Learning and Stochastic Approximations”. In: *On-Line Learning in Neural Networks*. Cambridge, MA: Cambridge University Press, 1999, pp. 9–42. DOI: [10.5555/304710.304720](https://doi.org/10.5555/304710.304720).
- [92] T. Tieleman and G. Hinton. *Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude*. Lecture slides. [http://www.cs.toronto.edu/~tijmen/csc321/slides/lecture\\_slides\\_lec6.pdf](http://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf). COURSERA: Neural Networks for Machine Learning, 2012, pp. 26–30.
- [93] D. P. Kingma and J. Ba. “Adam: A Method for Stochastic Optimization”. In: *arXiv* 1412.6980 (2017). <https://arxiv.org/abs/1412.6980v9>.
- [94] A. Tikhonov. “On the stability of inverse problems”. In: *Proc USSR Acad Sci* 39 (1943), pp. 195–198.
- [95] G. E. Hinton. “Learning translation invariant recognition in a massively parallel networks”. In: *PARLE Parallel Architectures and Languages Europe*. Ed. by J. W. de Bakker, A. J. Nijman, and P. C. Treleaven. Berlin, Heidelberg: Springer, 1987, pp. 1–13. DOI: [10.5555/25489.25490](https://doi.org/10.5555/25489.25490).
- [96] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. “Dropout: A Simple Way to Prevent Neural Networks from Overfitting”. In: *J Mach Learn Res* 15.56 (2014), pp. 1929–1958. DOI: [10.5555/2627435.2670313](https://doi.org/10.5555/2627435.2670313).
- [97] F. Chollet. *Keras*. <https://github.com/fchollet/keras>. 2015.
- [98] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mane, R. Monga, S. Moore, D. Murray,

- C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viegas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. “TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems”. In: *arXiv* 1603.04467 (2016). <https://arxiv.org/abs/1603.04467v2>.
- [99] A. von Appen, J. Kosinski, L. Sparks, A. Ori, A. L. DiGiulio, B. Vollmer, M. T. Mackmull, N. Banterle, L. Parca, P. Kastritis, K. Buczak, S. Mosalaganti, W. Hagen, A. Andres-Pons, E. A. Lemke, P. Bork, W. Antonin, J. S. Glavy, K. H. Bui, and M. Beck. “In situ structural analysis of the human nuclear pore complex”. In: *Nature* 526.7571 (2015), pp. 140–143. DOI: [10.1038/nature15381](https://doi.org/10.1038/nature15381).
- [100] J. V. Thevathasan, M. Kahnwald, K. Cieslinski, P. Hoess, S. K. Peneti, M. Reitberger, D. Heid, K. C. Kasuba, S. J. Hoerner, Y. Li, Y. L. Wu, M. Mund, U. Matti, P. M. Pereira, R. Henriques, B. Nijmeijer, M. Kueblbeck, V. J. Sabinina, J. Ellenberg, and J. Ries. “Nuclear pores as versatile reference standards for quantitative superresolution microscopy”. In: *Nat Methods* 16.10 (2019), pp. 1045–1053. DOI: [10.1038/s41592-019-0574-9](https://doi.org/10.1038/s41592-019-0574-9).
- [101] E. L. Lehmann and G. Casella. *Theory of point estimation*. Springer texts in statistics. New York, NY: Springer, 1998. DOI: [10.1007/b98854](https://doi.org/10.1007/b98854).
- [102] M. Weber, M. Leutenegger, S. Stoldt, S. Jakobs, T. S. Mihaila, A. N. Butkevich, and S. W. Hell. “MINSTED fluorescence localization and nanoscopy”. In: *Nat Photonics* 15.5 (2021), pp. 361–366. DOI: [10.1038/s41566-021-00774-2](https://doi.org/10.1038/s41566-021-00774-2).

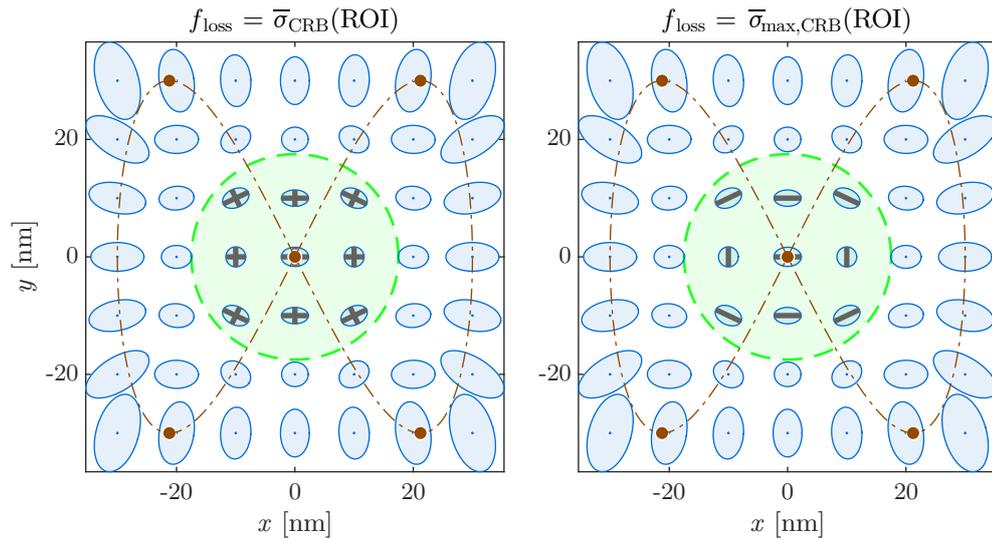
## A Supplementary figures for Lissajous illumination patterns



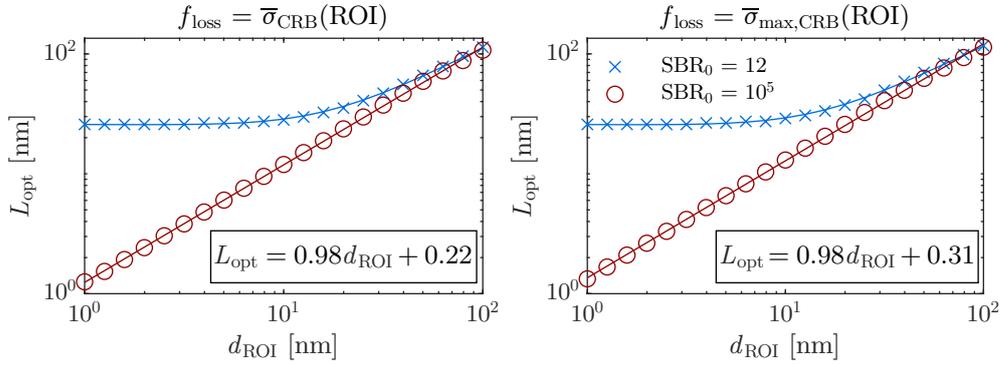
**Figure A.1: Spatial components of a Lissajous figure.**  $x$ - and  $y$ -components of a Lissajous curve given by (2.1) with  $\omega_x = \omega$ ,  $\omega_y = 2\omega$ , and  $q_x = q_y = 0$ , and absolute velocity along the curve.



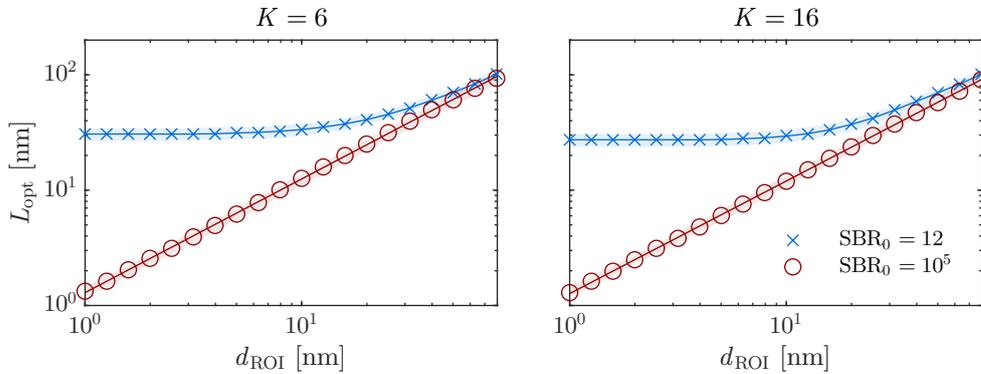
**Figure A.2: Influence of the number of exposures in the Lissajous pattern on the CRB and its isotropy along the  $x$ -axis.** Analogous to Figure 2.2.



**Figure A.3: Criteria for optimizing width and aspect ratio of the Lissajous pattern.** Two different loss functions  $f_{\text{loss}}$  were minimized on a circular ROI marked in green with a dashed periphery.  $\Sigma_{\text{CRB}}$  is visualized as ellipses with major and minor axes of length  $10\sigma_{1,2}$ , where  $\sigma_{1,2}^2$  are the eigenvalues of  $\Sigma_{\text{CRB}}$ . Gray bars indicate the eigenvalues of  $\Sigma_{\text{CRB}}$  that were taken into account in the respective loss functions. The brown, dash-dotted lines represent Lissajous patterns with  $K = 6$  exposures, width  $L = 60$  nm, and aspect ratio  $R = 1.0$ .

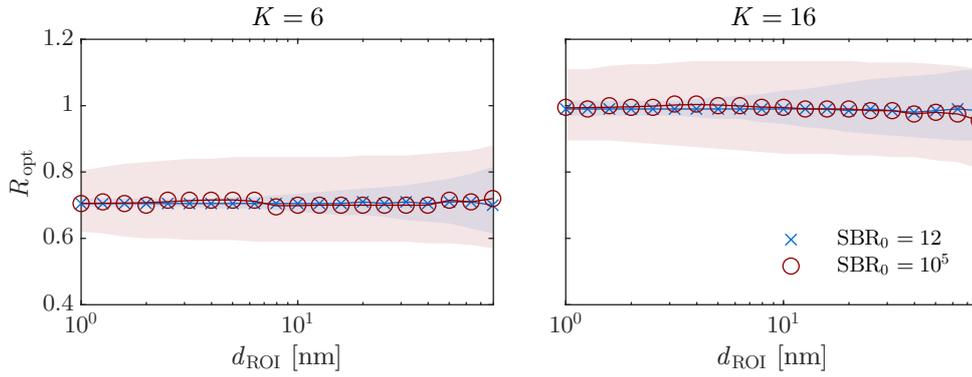
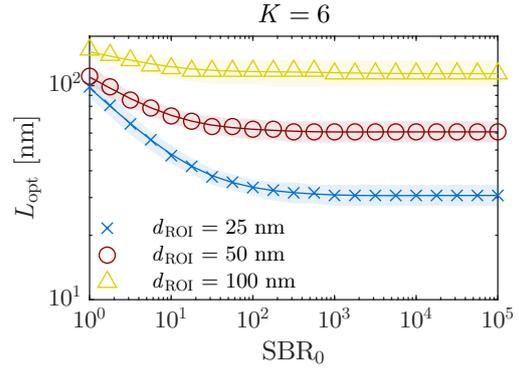


**Figure A.4: Optimal diameter of the triangular illumination pattern as a function of the ROI size.** The width of the four-point illumination pattern was optimized with and without background, using two different loss functions  $f_{\text{loss}}$ . The optimal width  $L_{\text{opt}}$  minimizes the value of the loss function in a circular ROI of diameter  $d_{\text{ROI}}$ . For each set, a cubic smoothing spline interpolation is shown in addition to the data points. The relations of linear regressions performed for the background-free data are given in the boxes.

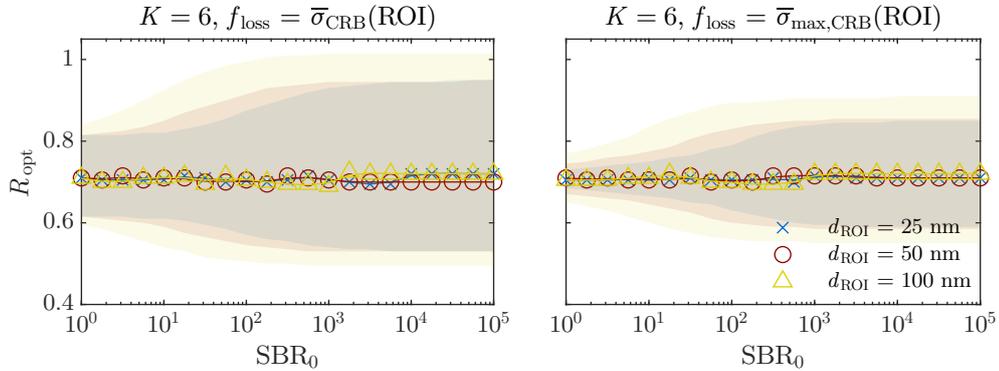


**Figure A.5: Optimal width of the Lissajous pattern as a function of the ROI size for a different loss function.** Analogous to Figure 2.5, but using the loss function  $f_{\text{loss}} = \bar{\sigma}_{\text{max,CRB}}(\text{ROI})$  for optimizing  $L$ .

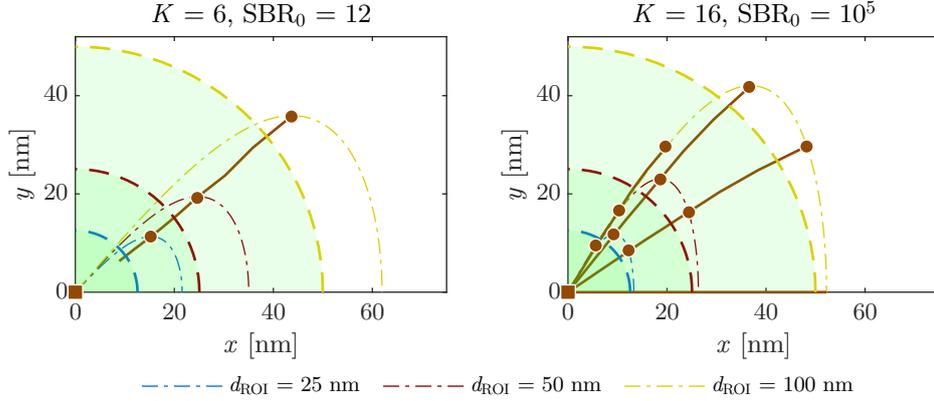
**Figure A.6: Optimal width of the Lissajous pattern as a function of the SBR for a different loss function.** Analogous to Figure 2.6, but using the loss function  $f_{\text{loss}} = \bar{\sigma}_{\text{max,CRB}}(\text{ROI})$  for optimizing  $L$ .



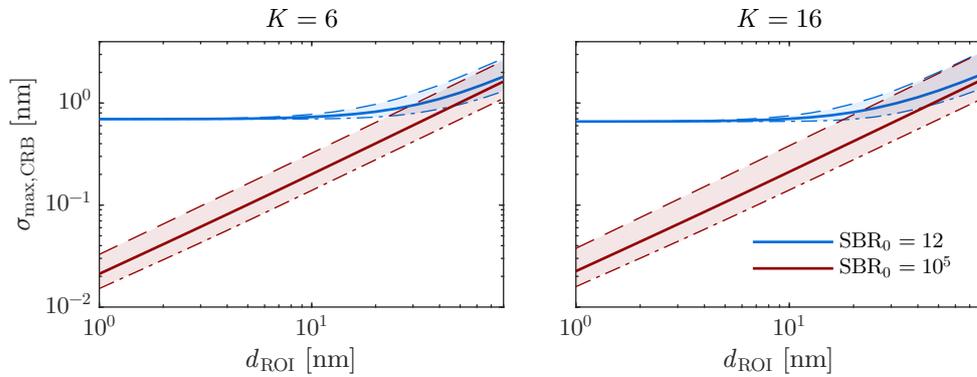
**Figure A.7: Optimal aspect ratio of the Lissajous pattern as a function of the ROI size for a different loss function.** Analogous to Figure 2.7, but using the loss function  $f_{\text{loss}} = \bar{\sigma}_{\text{max,CRB}}(\text{ROI})$  for optimizing  $R$ .



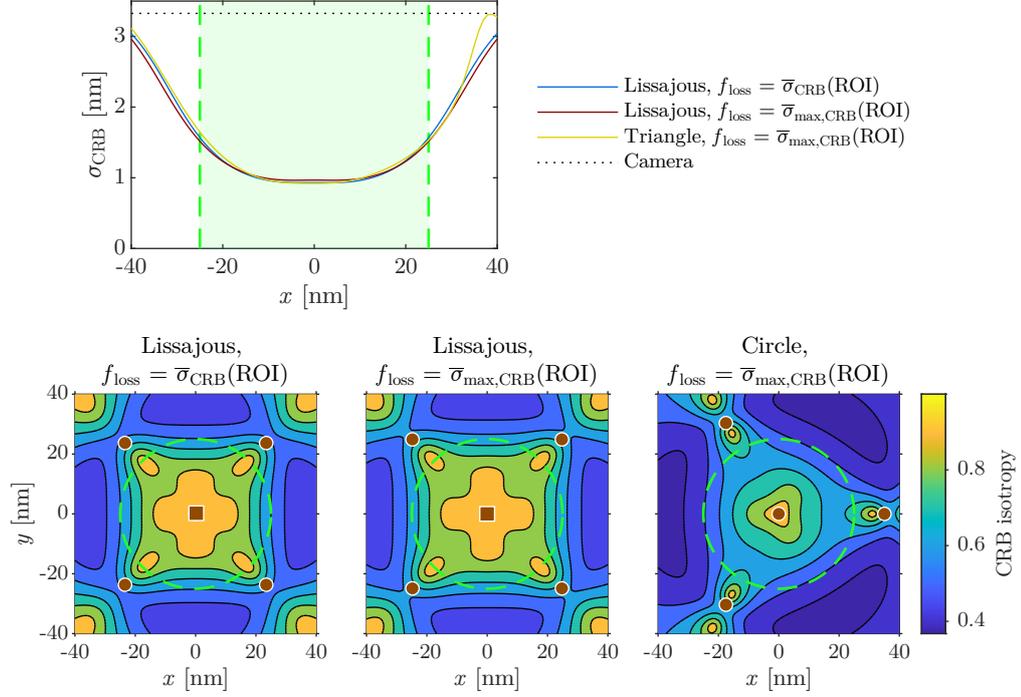
**Figure A.8: Optimal aspect ratio of the Lissajous pattern as a function of the SBR.** Analogous to Figure 2.7. Here the pattern aspect ratio was optimized for ROIs of three different diameters  $d_{\text{ROI}}$ , using  $K = 6$  exposures and two different loss functions  $f_{\text{loss}}$ .



**Figure A.9: Optimal excitation beam positions as a function of the ROI size.** Each  $xy$ -plot shows three exemplary ROIs (dashed lines) and the corresponding optimized Lissajous figures (dash-dotted lines) in the first quadrant, taking the  $x$ - and  $y$ -symmetry of the pattern into account. As the ROI diameter  $d_{\text{ROI}}$  increases, the optimal beam positions move outwards along the brown trajectories. Note that the central exposure points (marked by the brown circles) stay at the origin.  $L$  and  $R$  were optimized by minimizing the loss function  $f_{\text{loss}} = \bar{\sigma}_{\text{CRB}}(\text{ROI})$ .

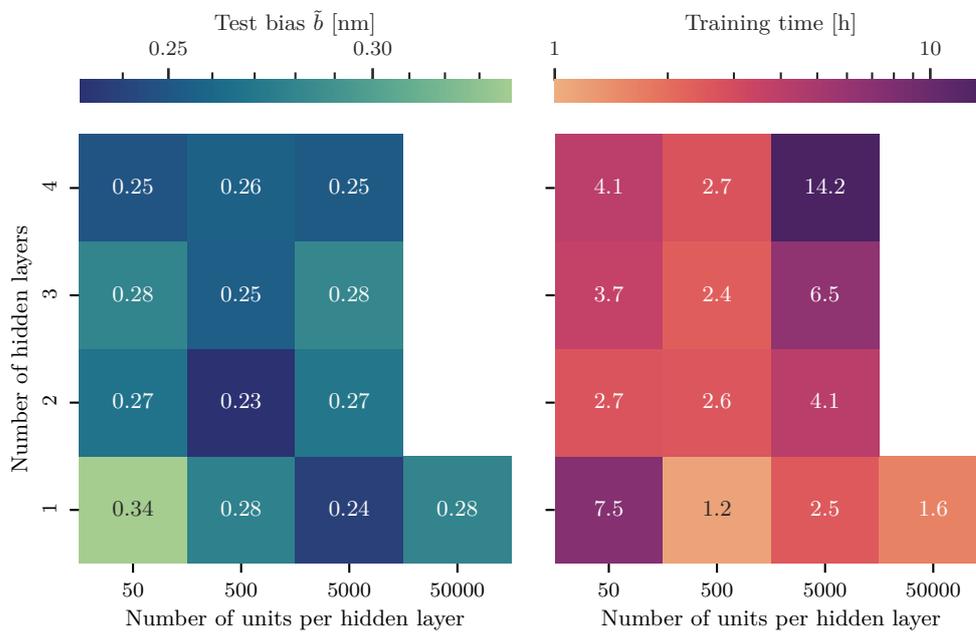


**Figure A.10: Maximum CRB obtained with the optimized Lissajous patterns as a function of the ROI size.** Analogous to Figure 2.8, but using the loss function  $f_{\text{loss}} = \bar{\sigma}_{\text{max,CRB}}(\text{ROI})$  for optimizing  $L$  and  $R$ . The lines indicate cubic smoothing spline interpolations to the minimum, maximum, and mean values of  $\sigma_{\text{max,CRB}}$  inside the ROI.

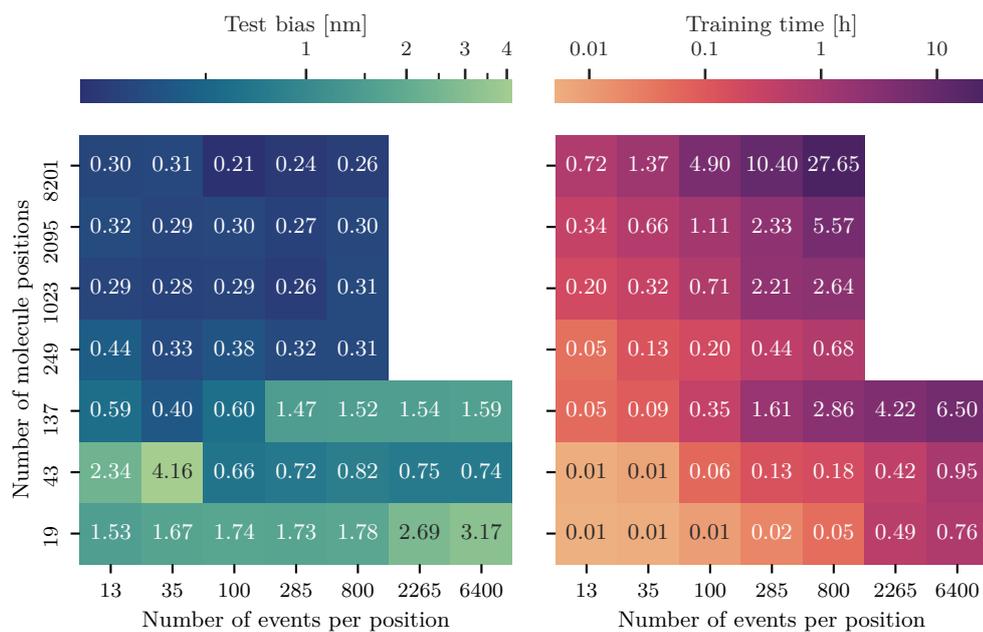


**Figure A.11: CRB and its isotropy for illumination patterns optimized with different loss functions.** The central profiles of  $\sigma_{\text{CRB}}$  in  $x$ -direction and the 2D maps of  $\mathbb{I}$  are shown for two Lissajous patterns with  $K = 6$  and the triangular geometry with  $K = 4$ . The patterns are visualized by brown markers in the lower panels, where the squares at the center of the Lissajous curves indicate double exposure points. All geometries were optimized by minimizing two different loss functions  $f_{\text{loss}}$  on the ROIs marked by the dashed, green circles. This resulted in the following geometrical parameters:  $L = 66/70/70$  nm, and  $R = 0.7$  for the Lissajous curves. In the upper panel the dotted, black line represents the camera CRB for an optimistic  $\text{SBR}_c = 500$ . The lower panels use a common color scale. Further parameters:  $d_{\text{ROI}} = 50$  nm and  $\text{SBR}_0 = 12$ .

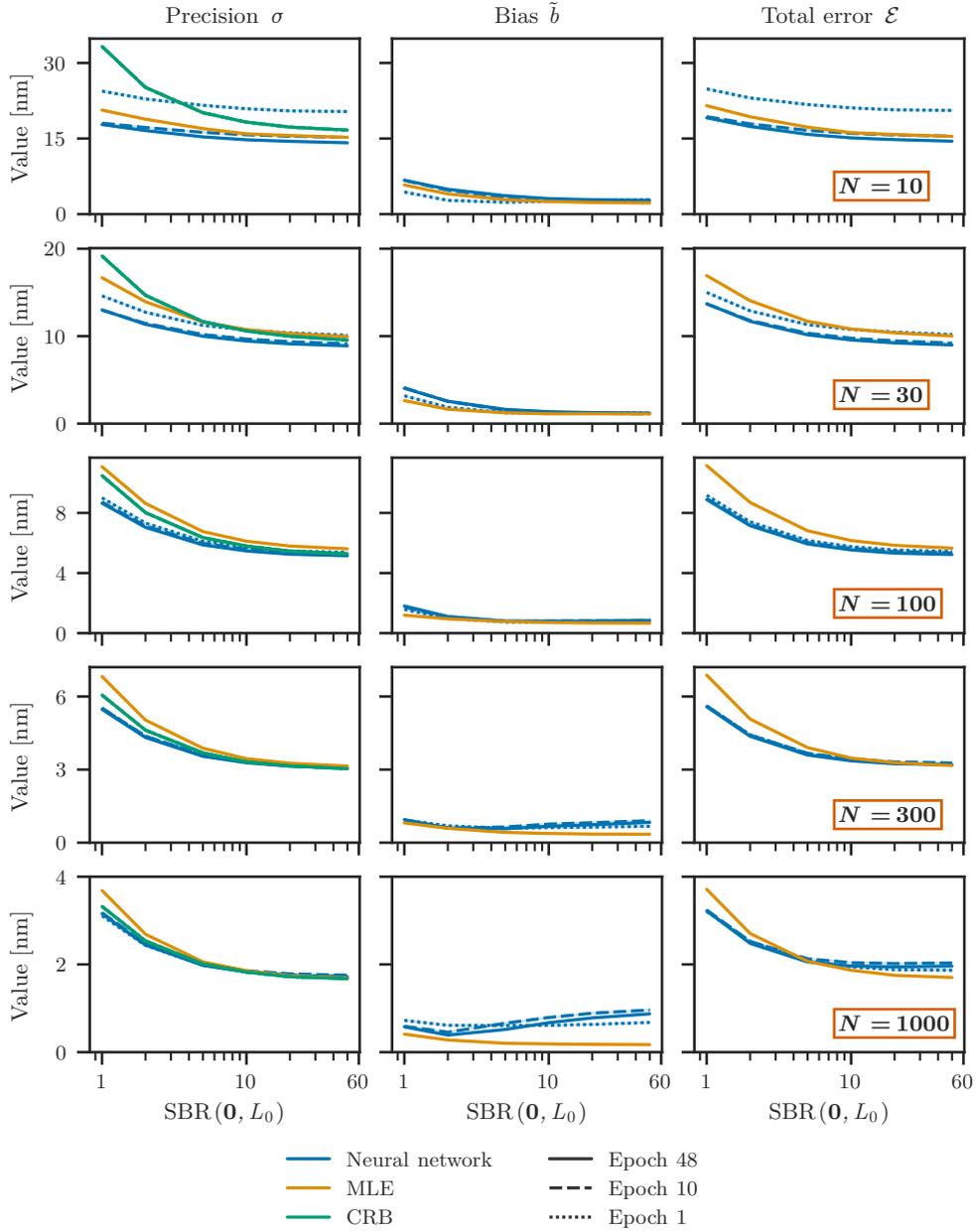
## B Supplementary figures for deep-learning-based data analysis



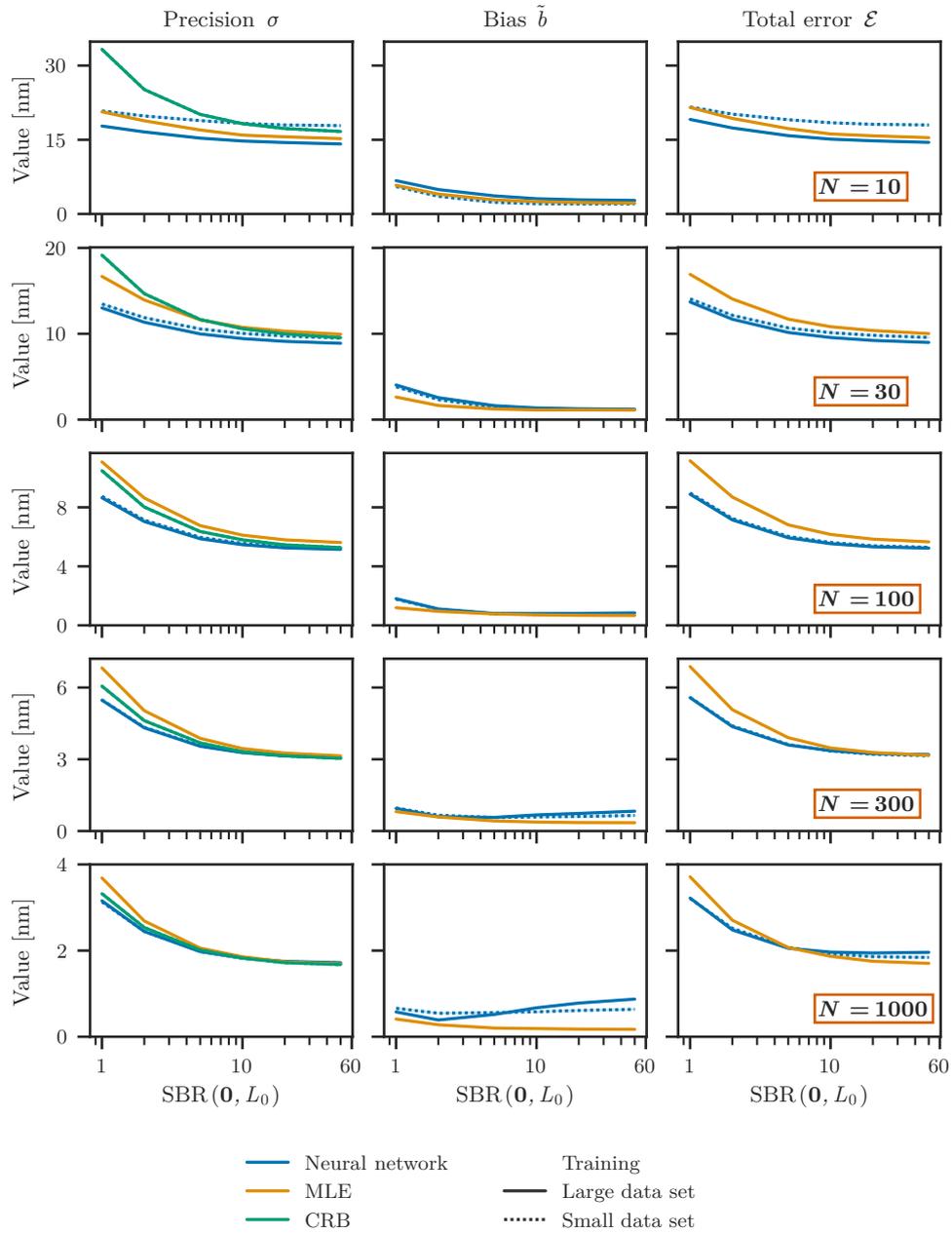
**Figure B.1: Test bias and training time for different network sizes.** Alternative representation of the data shown in Figure 3.8.



**Figure B.2: Test bias and training time for different training data compositions.** Alternative representation of the data shown in Figure 3.9.



**Figure B.3: Influence of the number of epochs on the localization performance of the universal network.** Analogous to Figure 3.10, showing the results of the network trained with variable  $N$  and SBR.



**Figure B.4: Influence of the amount of training data on the localization performance of the universal network.** Analogous to Figure 3.10, showing the results of a network trained with data sets of different sizes and variable  $N$  and SBR. The training and validation sets contained a total of either 8 000 000 or 800 000 data points approximately.

## List of Figures

1.1	2D excitation pattern with four doughnut-beam positions.	6
1.2	3D iterative MINFLUX scheme. . . . .	7
1.3	Influence of background on the CRB. . . . .	13
2.1	Arrangement of the excitation beam positions on a Lissajous figure in 2D. . . . .	19
2.2	Influence of the number of exposures in the Lissajous pattern on the CRB and its isotropy. . . . .	22
2.3	Influence of the width of the Lissajous pattern on the CRB and its isotropy. . . . .	24
2.4	Influence of the aspect ratio of the Lissajous pattern on the CRB and its isotropy. . . . .	25
2.5	Optimal width of the Lissajous pattern as a function of the ROI size. . . . .	29
2.6	Optimal width of the Lissajous pattern as a function of the SBR. . . . .	30
2.7	Optimal aspect ratio of the Lissajous pattern as a function of the ROI size. . . . .	31
2.8	CRB obtained with the optimized Lissajous patterns as a function of the ROI size. . . . .	32
2.9	2D maps of the CRB and its isotropy for different illumination patterns with optimal geometry. . . . .	34
2.10	Average CRB for different illumination patterns in various parameter settings. . . . .	35

2.11	CRB obtained with the optimized six-point Lissajous pattern in various parameter settings. . . . .	37
3.1	Building blocks of neural networks. . . . .	50
3.2	Key concepts of training neural networks. . . . .	52
3.3	Training and validation loss curves. . . . .	54
3.4	Training effect on the localization performance. . . . .	60
3.5	Predictions in the course of the training. . . . .	61
3.6	Influence of the extent of the training space on the localization performance. . . . .	63
3.7	Loss curves for networks of different sizes. . . . .	65
3.8	Localization performance and training time for different network sizes. . . . .	67
3.9	Localization performance and training time for different training data compositions. . . . .	69
3.10	Localization performance of neural networks and an MLE in various parameter settings. . . . .	73
3.11	Filtering of experimental data. . . . .	76
3.12	3D localization precision on experimental data. . . . .	77
3.13	Average offset between position estimates of a neural network and an MLE. . . . .	78
3.14	Nup96–SNAP–Alexa-Fluor-647 labels localized by a neural network in a MINFLUX image of fixed nuclear pore complexes. . . . .	80
3.15	3D rendering of single nuclear pores localized by a neural network and an MLE. . . . .	81
A.1	Spatial components of a Lissajous figure. . . . .	105
A.2	Influence of the number of exposures in the Lissajous pattern on the CRB and its isotropy along the $x$ -axis. . . . .	105
A.3	Criteria for optimizing width and aspect ratio of the Lissajous pattern. . . . .	106

A.4 Optimal diameter of the triangular illumination pattern as a function of the ROI size. . . . .	107
A.5 Optimal width of the Lissajous pattern as a function of the ROI size for a different loss function. . . . .	107
A.6 Optimal width of the Lissajous pattern as a function of the SBR for a different loss function. . . . .	108
A.7 Optimal aspect ratio of the Lissajous pattern as a function of the ROI size for a different loss function. . . . .	108
A.8 Optimal aspect ratio of the Lissajous pattern as a function of the SBR. . . . .	108
A.9 Optimal excitation beam positions as a function of the ROI size. . . . .	109
A.10 Maximum CRB obtained with the optimized Lissajous patterns as a function of the ROI size. . . . .	109
A.11 CRB and its isotropy for illumination patterns optimized with different loss functions. . . . .	110
B.1 Test bias and training time for different network sizes. . . . .	111
B.2 Test bias and training time for different training data compositions. . . . .	112
B.3 Influence of the number of epochs on the localization performance of the universal network. . . . .	113
B.4 Influence of the amount of training data on the localization performance of the universal network. . . . .	114



# List of Tables

2.1	Parameters for optimizing width and aspect ratio of the Lissajous figure in a 2D grid search. . . . .	28
3.1	Main hyperparameters of feedforward neural networks. . .	53
3.2	Parameters of the simulated MINFLUX data. . . . .	59



## Acknowledgments

This is the place where I would like to express my sincere gratitude to supervisors, colleagues, friends, and family members who have accompanied me on the way towards the submission of this thesis.

First of all, I would like to thank Professor Stefan W. Hell for offering me the opportunity to leave picturesque Heidelberg behind in exchange for an exciting project in an environment full of highly skilled people – a decision I have not regretted in all these years. Under his supervision I was able to work on several challenging research questions shining light on super-resolution microscopy from all different angles. He gave me the freedom to try out new methods I was passionate about, and lastly, the chance to achieve my personal goal for this career step: to become a real expert in microscopy. To Professor Silvio Rizzoli, I am grateful for examining my dissertation, and to both him and Professor Sarah Köster as members of my thesis advisory committee for accompanying my doctoral studies and giving me advice on my research. I also want to thank Professor Alexander Egner, Professor Axel Munk, and Professor Claus Ropers for showing interest in my work and being part of my examination board.

I am deeply grateful to Dr. Jan Keller-Findeisen for being a competent, patient, and empathetic supervisor, for invaluable discussions on the results presented in this thesis, for being my sparring partner in all programming-related questions, for continuous proofreading, and for being a great office

mate for many years. Last but not least, he took an active part in my project by implementing the visualizations of Nup96 MINFLUX data.

My special thanks go to Dr. Jasmin Pape, who gave me important insights into MINFLUX post-processing, who juggled ideas with me on multiple occasions, and who ultimately supported me in defining the research question for my deep-learning project. Moreover, she provided me with the experimental MINFLUX data analyzed in this thesis. I would like to thank Dr. Francisco Balzarotti for insightful scientific discussions, for being a patient teacher in Spanish microscopy terminology, and for always supporting me as a colleague and friend. I am grateful to Dr. Klaus Gwosch for project-related advice and especially for providing me with an invaluable reference for the fundamentals of MINFLUX, comprehensively described in his own doctoral thesis. A reliable IT infrastructure formed the basis for my data-driven projects and was kindly provided by Dr. Marco Roose. My sincere thanks also go to Marcel, Volker, Mark, and Steffen for proofreading and helping me to put the finishing touches to my dissertation.

Furthermore, I want to thank current and former colleagues who supported me in earlier projects, namely Dr. Fabian Göttfert, Dr. Sven Sidenstein, Dr. Volker Westphal, Tanja Koenen, and Ellen Rothermel. Another big thank you goes to our secretaries Sarah and Steffi, to the IMPRS-PBCS office, and to Jay and Simon for their reliable help with everyday questions of all kind. The whole department of NanoBiophotonics created an enjoyable and welcoming work atmosphere. Especially Dirk, Lars, Christian, Mark, Christiaan, Daniel, and Daniel will be remembered for Friday runs, Sunday workouts, any-day rooftop coffee breaks, and last-minute pre-submission holiday trips.

Last but not least I want to express my deepest gratitude to my beloved family, who always believe in me and have greatly supported me in finishing this thesis.