UNIVERSITÀ DEGLI STUDI DI TRIESTE

XXXIII CICLO DEL DOTTORATO DI RICERCA IN
Ingegneria Industriale e dell'Informazione

On problems and computational imaging solutions for ptychography

Settore scientifico-disciplinare: ING-INF/01 ELETTRONICA

Dottorando
Francesco Guzzi
Francesco Guzzi

Coordinatore
Prof. Fulvio Babich

Supervisore di Tesi
Prof. Sergio Carrato

Co-Supervisore di Tesi
Dr. George Kourousias

Anno Accademico 2019/2020

Acknowledgements

"No man is an island," wrote Ernest Hemingway.

This activity would have not been possible without the patience, dedication, kindness, gratuity and presence of many important people.

I have to sincerely thank my Supervisor, Prof. Sergio Carrato and my co-Supervisor Dr Georgios Kourousias for their precious and kind advice and their dedicated guidance.

I am thankful to the reviewers Dr Michael Jones (Queensland University of Technology) and Prof. Benedetto Bozzini (Politecnico di Milano) for their feedback. Their constructive suggestions and important advice greatly contributed to creating the final version of the thesis.

I have to thank also Prof. Stefano Marsi and Prof. Gianni Ramponi (Image Processing Laboratory, University of Trieste) for all the gratuitous and insightful advice.

My immense gratitude is also for the members of the IT group (in particular the Scientific Computing and Software For Experiments teams) of Elettra Sincrotrone Trieste as well as for Dr Alessandra Gianoncelli (TwinMic beamline): the insightful discussions and the gratuitous help was essential for this work.

This research has been developed under the Advanced Integrated Imaging Initiative (AI3); project P2017004 of Elettra Sincrotrone Trieste in agreement with University of Trieste.

Contents

List of Acronyms

AI	Artificial Intelligence
ALU	Arithmetic Logic Unit
CAS	Computer Algebra System
CAT	Computer Axial Tomography
CCD	Charge coupled device
CDI	Coherent Diffraction Imaging
CG	Conjugate Gradient
CI	Computational Imaging
CMOS	Complementary Metal Oxide Semiconductor
CNN	Convolutional Neural Network
CPU	Central Processing Unit
CT	Computed Tomography
CUDA	Compute Unified Device Architecture
CV	Computer Vision
DFT	Discrete Fourier Transform
FFT	Fast Fourier Transform
DL	Deep Learning
DM	Differential Map
DNN	Deep Neural Network
DP	Differentiable Programming

DSP	Digital Signal Processing
DSP	Digital Signal Processor
DSSIM	Structural Dissimilarity
ERPY	Embedded Pupil Function Recovery
ER	Error Reduction
eV	electron Volt
FN	Fresnel Number
FOV	Field Of View
FPCNN	Fourier Ptychography Convolutional Neural Network
FP	Floating Point
FP	Fourier Ptychography
FRC	Fourier Ring Correlation
FZP	Fresnel Zone Plate
GPU	Graphics Processing Unit
GT	Ground Truth
GUI	Graphics User Interface
HIO	Hybrid Input-Output
HPC	High Performance Computing
HR	High resolution
IM	Iterated Map
ISA	Instruction Set Architechture
JIT	Just In-Time Compiler
LED	Light Emitting Diode
LEXRF	Low Energy X-ray Fluorescence
LR	Low resolution
M-EPIE	Multi-mode Extended Ptychography Iterative Engine
M-RPIE	Multi-mode Regularised Ptychography Iterative Engine

MF	Magnification Factor
MISR	Multiple Image Super Resolution
ML	Machine Learning
ML	Maximum Likelihood
MPI	Message Parsing Interface
MRI	Magnetic Resonance Imaging
MSE	Mean Square Error
MSSIM	Mean Structural Similarity Index
NA	Numerical Aperture
NMI	Normalised Mutual Information
NN	Neural Network
OSA	Order Sorting Aperture
OZP	Output Zone Plate
PCI-Express	Peripheral Component Interconnect Express
PIE	Ptychography Iterative Engine
POCS	Projections Over Convex Sets
POS	Projections Over Sets
PRELU	Parametrisable Rectified Linear Unit
PTX	Parallel Thread Execution
PWSTD	Pairwise Standard Deviation
RAM	Random Access Memory
RELU	Rectified Linear Unit
RGB	Red Green Blue
ROC	Receiver Operator Curve
ROI	Region Of Interest
SDD	Silicon Drift Detector
SEM	Scanning Electron Microscopy

SIFT	Scale-invariant feature transform
SISR	Single Image Super Resolution
SNR	Signal To Noise Ratio
SR	Super Resolution
SSIM	Structural Similarity Index
STD	Standard Deviation
STN	Spatial Transformer Network
STXM	Scanning Transmission X-ray Microscopy
SXFM	Scanning X-ray Fluorescence Microscopy
SXT	Scanning X-ray Tomography
TEM	Transmission Electron Microscopy
TXM	Transmission X-ray Microscopy
UQI	Universal Quality Index
XFM	X-ray Fluorescence Microscopy
XRF	X-Ray Fluorescence

Abstract

Ptychography is a modern microscopy technique, still in development, within the broader family of Coherent Diffraction Imaging (CDI) based methods. These methods have the potential to revolutionize the study of multiple scientific fields by providing higher resolutions of larger areas at faster speeds, including chemical speciation. The principle of ptychography, in particular, is that a beam of coherent light is shined onto a sample in a scanning fashion and the acquired diffraction patterns are computationally merged to recreate the whole specimen's complex image.

The research activity presented in this thesis is multidisciplinary; with pivotal point in electronic/computer engineering it solves some of the complications of this modern computational imaging technique, which is used in applied physics, with applications in fields such as electrochemistry, biology, nanomaterials and other disciplines that benefit from microscopy. Based on recent advances in Artificial Intelligence, this research makes use of Machine Learning and Optimisation methods for solving certain problems.

The results of this research provided solutions to various problems of ptychography, improving the technique. Moreover, the thesis delivers an implementation of the proposed solutions as a complete software framework [1] based on modern engineering paradigms such as GPU computing.

The development of this work has been carried on in cooperation with an actual laboratory that implements ptychography. This laboratory is the TwinMic X-ray spectromicroscopy beamline at the synchrotron radiation facility Elettra Sincrotrone Trieste, in collaboration with its Scientific Computing group. The thesis also contains the result of two beamtime experiments and discusses the proposals of an additional 2 that have already been granted.

The main contributions of this research range over the topics of spatial coherence, positions refinement, and parameters tuning.

- Partial coherence of the source deteriorates the images that can be reconstructed: we proposed a refined reconstruction algorithm (M-RPIE) where the large computational field of view of the method allows for a sparser scanning [1, 2].
- Position errors impact negatively ptychography, thus a part of the research was devoted to providing a solution to this issue. This led to
    - a metric based approach [3];
    - an analysis of the dynamics of the error signal [1];
    - a method for the automatic control of the position feedback gain [1].
    - a method for including position refinement coefficients within an optimisation process [4]
- Ptychography requires a multitude of parameters which currently are manually tuned. Using advanced Deep Learning techniques, we proposed a reconstruction algorithm which automatically regresses the propagation distance and the position correction coefficients within an optimisation-based process [4]
- Fourier ptychography was also explored and implemented in the software framework. We proposed a CNN model for the generation of a prior which can be effectively used to seed the reconstruction [5].
- Since the research is multidisciplinary, certain results where utilised in other fields such as X-ray Fluorescence [6], Computed Tomography [7], super-resolution for forensics [8], CNNs for face recognition [9, 10, 11, 12, 13] and depth estimation [14].

Introduction

Computer Science and Engineering is the catalyst and enabling factor for almost every modern field of study. Physics and their applications fully rely on cutting edge computing and are often the inspiring factors for new advances. If in the past the research effort in microscopy was fully dictated by the need of rapid development in light sources, optics and sensors technology, in the information age research branched in new areas such as digital signal processing (DSP), computer vision (CV), machine learning (ML), and big data processing, achieving leaps in progress [15]; the combined revolution of these two research areas promoted the birth and development of computational imaging (CI), now a stand-alone discipline that is used in many fields, from photography and medicine to applied physics.

A cutting edge application of it is its use in X-ray microscopy in large accelerators like Elettra in a technique called ptychography. This technique represents a true revolution in the field bringing a sensible gain in the resolution just by processing data that can be acquired with minors, if no, modifications on a typical coherent light beamline setup (Scanning Transmission X-ray Microscopy STXM + conventional Transmission X-ray Microscopy TXM). Nevertheless, this high resolution is available also on a quantitative-valued phase image [16], something that can not be immediately imaged exploiting common imaging methods.

Computational imaging techniques rely on a two-phase process: i) coupled sampling and ii) computational decoupling (i.e., reconstruction), following an image formation model that is known and defined *a priori*. Transmission X-ray ptychography works by shedding a coherent and monochromatic light beam on a sample and observing the scattering exit wave at some distance; the observed diffraction pattern is an intensity image of the field at the detector that encodes a complex intermixing of the real and imaginary part of the specimen refraction index (due to the propagation). If the processing of a single diffraction pattern was the objective of the firsts Coherent Diffraction Imaging (CDI) methods developed in the early 2000s, ptychography (a different approach to CDI) works by intelligently combining (computational decoupling) many diffraction patterns acquired in a scanning fashion (coupled sampling); this "mosaic" not only produces an image of the transmission function with a large Field Of View (FOV) but introduces also a set of constraints that helps in the computational reconstruction procedure.

Until 1999, the race for higher resolution was entirely driven by the use of more performant X-ray optics, and more specifically zone-plates. After the landmark paper [17], the family of Coherent Diffraction Imaging (CDI) methods were formally born: the described experiment amazingly revealed an image from the far-field diffraction pattern formed by a scattered coherent radiation incident on a microscopic synthetic sample. During these 20 years of research in CDI, thanks to both the improved detector/sources characteristics and the exponential increase of computing power, many experimental setups have been brought to reality, leading to transmission X-ray ptychography being one of the most effective and used techniques in synchrotron facilities. From the signal processing point of view, the technique can be studied as a multi-frame super-resolution approach applied to CDI, through which a representation of the complex transmission function of a specimen is reconstructed.

Even from this high-level description of the technique, it becomes evident how huge is the role of the algorithms and their software implementation in reconstructing images "that can not be seen" conventionally.

A computational technique is required because we are trying to solve an inverse problem: from the measurable effects we want to find the cause that originated them. Another huge element in the software complexity (here taken as granted) is the control system of a microscope setup; that makes possible the coupled sampling.
Computed Tomography (CT) is another famous example of computational imaging (and inverse problem), but there is an important distinction: Fourier slice theorem allows for a deterministic single-step reconstruction (inversion). A "Radon transform equivalent" for ptychography is not feasible, instead (the Wigner-distribution deconvolution is a proof of concept and requires lateral movements of the order of the resolution). This means that all the CDI-based methods have to be solved (to date) through an iterative procedure. Algorithms of these type are then computationally expensive and currently require many core parts to be written in a low-level language, such as highly optimised C-like MPI or CUDA kernels.

The results exposed in this thesis provided solutions to various problems of CDI/ptychography to improve the technique. In chapter 1 some fundamental theoretical aspects will be presented, regarding the wave propagation aspects in free-space and within the matter. The electromagnetic aspects will be compared briefly to the particle model.

Chapter 2 will present many fundamental aspects of the imaging performed by using coherent light, namely CDI and Ptychography. Reconstruction methodologies and algorithms will be described with the help of simulations and the ptychography technique will be contextualised by providing a brief summary of the experimental setup. A brief section description of the assigned beamtime experiments will close the chapter.

In chapter 3 we will discuss some solutions that have been proposed to solve some ptychography problems, like positions error, a low degree of beam coherency and parameter tuning.

Chapter 4 will briefly introduce the Fourier Ptychography technique, leading to a Deep Learning-based solution we proposed to increase the reconstruction convergence and quality.

Due to the multidisciplinary nature of this work, in chapter 5 we will describe some solutions based on the same core-ideas of the main part of the thesis, but applied on other topics.

Chapter 1

Framework

X-ray transmission microscopy (TXM) is a well-established and mature family of imaging techniques that aims at investigating the microscopic structure of thin materials. The common approach to this type of imaging consists of the observation of the emerging radiation obtained illuminating a sample, basically looking through it. We will formally describe the setup in the following text. For a more detailed discussion, one can refer to e.g. [18] or other X-ray optics textbooks. However, before going into physics, as this thesis is about computational methods applied to microscopy, a short introduction on inverse problems is mandatory. For the same reason, I also have to briefly introduce how algorithms are currently implemented in a computer software, and that is why Graphics Processing Unit (GPU) are shortly introduced in the framework chapter.

1.1 Synchrotron radiation experiments

"Elettra" is the name of a synchrotron facility where this research activity has been performed. In this laboratory, the synchrotron radiation is generated artificially and exploited to study and analyse the matter, trough various types of experimental setups.

Synchrotron "light" is the electromagnetic radiation generated when an electric charge running at relativistic speed is accelerated radially [19]. In a synchrotron, the electrons are curved from the magnetic fields and a radiation beam is generated tangentially; this is what is called *synchrotron radiation* [20]. The radiation is guided towards the beamlines, which terminate with an experimental laboratory. Synchrotron light possesses many peculiar characteristics [20], such as high spectral radiance, high coherency (with respect to a conventional source), high emission spectrum (which can be later tuned) and the availability to produce radiation with many types of polarisation. These characteristics make this radiation the eligible choice for a large class of experiment.

Elettra is called a "third-generation machine" due to its structure, specialised for the creation of high-quality synchrotron radiation: different elements are in fact exploited to generate (linac), to accelerate (booster), to maintain (storage ring) and to produce the photons. Elettra owns mainly two types of radiation generation components: bending magnets and undulators. This allows accomplishing different types of experiments, from the ones requiring the largest frequency spectrum (bending magnets) to the ones which need spectral purity and coherence (undulators). The storage ring is the structure which keeps constant the characteristics of the electrons beam, mainly providing a fixed temporal electron structure (a bunch) and given kinetic energy, allowing many different beamlines to operate concurrently. This latter fact is in contrast to many Free Electron Laser facilities.

As can be seen in Fig. 1.1, from each photon generation device, a "beamline" departs. A synchrotron radiation beamline usually refers as a whole to the composite system of radiation beam generation, transport, conditioning and use. The optical hutch is the room which contains part of the radiation transport system and the processing of the beam, while the experimental hutch is where the "end-station" is located, the place where "one put the sample" and observes the radiation-matter interaction.

The TwinMic X-ray spectromicroscope is a world-wide unique instrument that combines full-field imaging with scanning X-ray microscope in a single instrument. The microscope is operated currently in the 400 - 2200 eV photon energy range or as equivalent 0.56 - 3 nm wavelengths. According to the energy and X-ray optics, TwinMic can reach sub-100nm spatial resolution.

Beamtime

A "beamtime" is an allocated time for the use of a specific instrument at a synchrotron facility. The cost of 1 hour of time is in the order of magnitude of one thousand euro. Even if some company can purchase the machine time for its own non-disclosed research, the typical way to access to the light is through a proposal submission process, which is repeated twice a year: a principal investigator (PI) together with its experimental team designs an experiment. A beamtime experiment proposal is a technical document where the research team explains the theorised experiment, detailing how

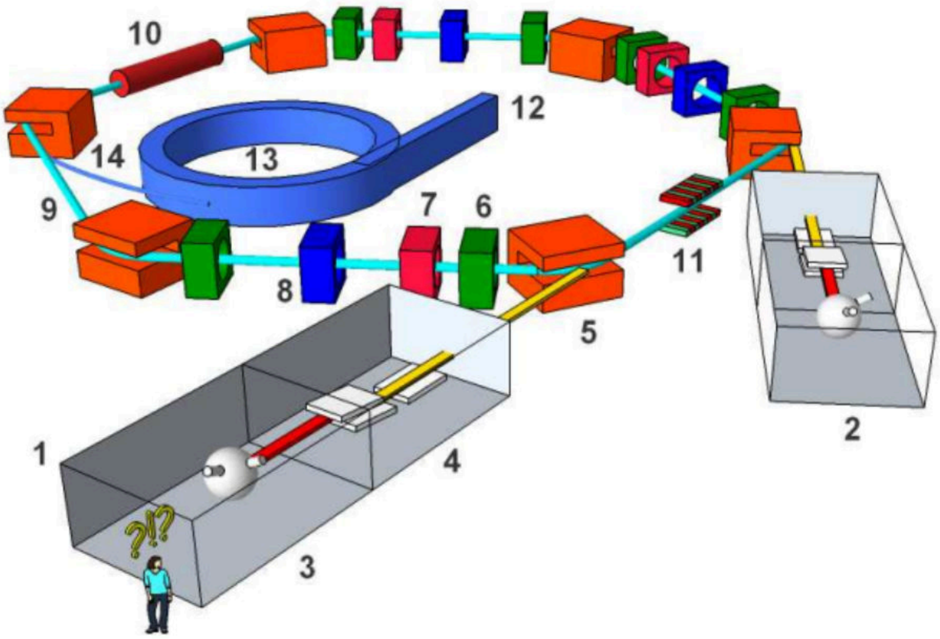


Figure 1.1: Simplified structure of a third-generation light source. 1) undulator based beamline (e.g. Twinmic); 2) bending-magnet based beamline (e.g. Syrmep); 3) experimental hutch; 4) optics hutch; 5) bending magnet; 6) quadrupole magnet; 7) sextupole magnet; 8) steering magnet; 9) storage ring vacuum chamber; 10) RF cavity; 11) insertion device (e.g. undulator) 12) linac; 13) booster; 14) transport line. Image taken from [21]. Thanks to Dr. M. Antonelli

the investigation will take place. Then the experiment proposal is submitted to an external international commission (composed by other researchers in the field) which have to validate the experiment proposal and assign a grade. At this point, if the experiment is accepted, the beamtime can be allocated. At the end of the experiment, a beamtime report is redacted, explaining how the experiments evolved; the user has to demonstrate that the allocated time has been used correctly.

During my research activity, I have been involved in the preparation of 4 beamtime proposal, which received a high grade for the complexity and the exposed methodologies. Two of them have been actually performed and represent the base for the real data used in this work.

1.2 Notes on Inverse problems

The principal aim of collecting data is to gain meaningful information about a particular phenomenon. However, in many cases, the quantities we wish to determine are different from the one we are effectively able to measure; they have been affected by a "processing" that is inherent to the phenomenon we want to study. Thus, starting with the measured data, the problem of trying to reconstruct the interesting quantities is called an *"inverse problem"*. In different terms, we are interested in reconstructing the *cause* of a particular manifested and recorded *effect* [22, 23, 24]. Many are the modern examples:

- **Computer axial tomography (CAT, or CT)**, which aims to reconstruct the 3D absorption matrix that represents the "sample", just by looking at the *total* absorption along a line. In simpler terms, for each acquired pixel, we know the result of a line integral, while aiming at recovering the full line distribution. We help the reconstruction by imposing an image modality which requires the sample to rotate, allowing different views to be acquired (the coupled sampling described in the introduction). In chapter 5.2, more bits of information will be provided on the topic.
- **Model fitting** which aims at recover the parameter of a particular model, by looking at the input/output observations. A physics example is the retrieval of the half-life of a radioactive substance from the measurements of the times at which decay products are detected, but it can be also a simple "curve fitting", as well as a complex echo canceller in our phone (both for audio and multipath propagation). Also, X-ray Fluorescence analysis requires this kind of processing.
- **Inpainting/Interpolation**: given a set of sparse measures (e.g. pixel values of an image) we aim at recover (infer) the missing values *in between*, increasing as much as possible the image fidelity. Solutions for this problems are extensively employed in the field of image coding [25] as well as in scientific imaging [26].
- **Deconvolution**: given a blurred or smeared image, which is the result of a linear, time invariant and spatio invariant process, we want to recover the original image. An even more complex problem is the *blind deconvolution* which have to correctly estimate also the perturbation model that affected the original image. This type of problem is present also in mono-dimensional signals and its solution is at the base of the modern high-speed telecommunications networks.
- **Super-resolution**: in its single-image counterpart, the problem is really similar to a particular mixture of the aforementioned blind con-

volution and inpainting. Indeed, here the reconstructed image domain is forced by definition to be larger than the observable one. The techniques, then changes slightly. In its more rigorous form, instead, the problem is to recover an high resolution image from a bunch of low resolution ones, which are different views of the same scene. Similarly to the case of CT, if dealing with different views practically makes the reconstruction feasible, to reach such result, it is mandatory to use correctly the information, finding a suitable "data fusion" procedure (e.g. [8]). More details will be given in chapter 5.4.

- **Phase retrieval**: it is the core problem of this thesis; it arises from the fact that basically after the infrared portion of the electromagnetic spectrum, we are no more able to detect *fields* as we do with antennae, but only *intensities*. More details were provided throughout the text.

Hadamard well/ill posed problems

Let us take image deconvolution as an example inverse problem. The *forward problem* is a *simulation*, then in our example case is the mapping A from the perfect image f to quantities that can be effectively measured. In all the examples of this thesis, the mappings are based on some physycal theory, and thus can be linear or not. The *data space* is the set of all the possible recorded data. The forward process maps the error-free high frequency image f, to the error-free data d, but due to noise, we measure $d^{'}$

$$d' = A(f) + n = d + n \tag{1.2.1}$$

Howewer, in many cases we can assume than $n = 0$, at least in a first approximation. The inverse problem, thus aims finding the original f from $d^{'}$ and the knowledge of the *a priori* model.

Even if many processes in nature are continuous, a discrete-to-discrete mapping is typically employed, due to how *d* or *d'* are acquired and processed.

The inverse problem of solving $d = A(f)$ for f given d is defined *well posed* if:

1. a solution exist for any data d in the data space;
2. the solution is unique in image space;
3. the inverse mapping from d to f is continuous;

The first two conditions essentially states that the operator A has a well-defined inverse A^{-1} and that the domain of the inverse is all the data space.

The third condition is a necessary but non-sufficient for the stability of the solution, arising from numerical analysis. In a well-posed problem, the propagation of the relative error from the data to the solution is controlled by the *condition number* k:

$$\frac{\|\Delta f\|}{\|f\|} \leq k(A)\frac{\|\Delta d\|}{\|d\|} \tag{1.2.2}$$

where Δd is a variation on the measured data and similarly Δf is the variation on the latent image. For a linear problem such as deconvolution, the condition number is defined as the quantity:

$$k = \|A\|\|A^{-1}\| \tag{1.2.3}$$

which involves the norm of a linear operator $\|A\|$; this kind of norm is defined as:

$$\|A\| = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|} \tag{1.2.4}$$

which means that $\|A\|$ is the *supremum* of a transformation A, found by "testing" all the possible input x.

Then, if k is not too large, we can say that the problem is well-conditioned and that the solution is stable with respect to small variation of data. This arises from the fact that intuitively, for a small change in the inputs there is a small variation in the output. Conversely, for an *Hadamard ill-posed* problem it suffices to not have the inverse, or that the inverse is not unique or that arbitrarily small changes in the data modify abruptly the value in f.

Even if difficult to solve, many ill-posed or ill-conditioned problems such as the ones in the aforementioned list are howewer solved and the CT scans performed routinely are a good example of this fact. Phase retrieval, Coherent Diffraction Imaging and Ptychography are all ill-posed problems, but many algorithm have been proposed as a viable solution.

1.3 GPU computing

The expanding request for *a computational way* in many fields of applied science, drives the demand for higher computing speed and processing throughput. From the origins till 2003, the relentless grow in performance was dominated by the increasing of Central Processing Unit (CPU) clock and cache memory, as well as by using clever methodologies to execute in a rapid way single tasks [27]. Heat generation, energy consumptions and technology limit finally hit, bounding the rise of this kind of advances [28]. Then, a new type of computing approach has to be brought to the masses, the *software parallelism*, in the form of the multi-core CPUs that we use today. The main idea is to provide many computational units and to distribute the tasks among them. CPUs and GPUs both share the same idea, but also present many differences in how the idea is implemented.

Indeed, in a modern CPU each core is a complete *out-of-order* and *multiple instruction issue* processor implementing the full *x86* instruction set [29]. In contrast, the *many-thread* methodology focuses more on execution throughput of parallel applications, inspiring the production of GPU equipped with 10000+ simple, *in-order* computing pipelines. The raw speed gain is potentially about 10 and more, reaching peak performances of 1 teraflop.

CPU - GPU differences

The reason for this huge discrepancy lays in the CPU design: those are heavily optimised for sequential code performance, exploiting sophisticated control logic that can distribute automatically simpler operations in parallel on the few units (Fig. 1.2 panel A), or re-schedule instructions while maintaining the appearance of sequential execution [27]. More importantly, the use of large cache memories reduces data access latency. CPUs are designed to minimise the execution latency trough runtime "tricks" (deeply rooted in digital hardware design, numerical analysis and mathematical logic considerations) and caching mechanisms; then the global time is minimised by reducing the latency of the operation within the same thread. A *thread* of execution is the smallest sequence of programmed instructions that can be managed independently by a scheduler. The scheduler dispatches a job among the available units and as said it can be part of an Operating System as well as an on-die component of a CPU.

GPUs architecture are instead shaped by 3D industries [29]: the prevailing solution is to design the chips in order to maximise the execution throughput of a massive number of threads, each one meant to draw a polygon on the screen; this is achieved by implementing thousands of simple Arithmetic Logic Units (ALUs) onto the chip. The electronic design saves chip area and power by allowing long latency in both the memory access requests and arithmetic operations, compensating with a large memory bandwidth that allows bringing delayed data to all the thousands of units (Fig. 1.2 panel B). GPU Memory bandwidth is currently 5-10 times larger than for CPU, because of this latency-relaxed memory model. An additional latency has to be taken into account for central-memory to GPU-memory data transfer operations which are bottlenecked by the PCI-Express link.

GPUs are then throughput oriented designs, which strives to maximise the total execution throughput of a large number of threads while allowing individual operations to take potentially much longer time to execute. These characteristics provides globally the *"gpu gain"* observed running hardware accelerated algorithms.

GPUs architecture is organized into arrays (called "build blocks") [30, 29] of highly threaded streaming multiprocessors SMs, which shares a per-block control logic and instruction cache. One of the most used programming paradigms on GPUs is the *data-parallelism* and involves to apply the same operation on a large set of input data. If N units are used concurrently, in principle one can obtain a result in $1/N$ of the time. Element-wise matrix operations are typically performed in this way. A piece of parallelisable code is called "a kernel" which represents a list of instruction that has to be computed on some data. Launching a kernel generates a large number of threads, which are the parallelized operations; taken a kernel that has to operate on a set of data, a thread is started for each of the element of the data array and is executed in chunks that span the GPU allocated resources. There are however some caveats: 64 bit floating-point operations have been brought in hardware only recently, and are commonly discouraged in software designs. Unfortunately, Deep Learning (that is a new driving force within the market) does not need higher numerical precision, and 32 or 16 bit are often enough during a model training, dropping to even 1 or 2 bits for microcontroller ready neural networks. That is why this lag exists.

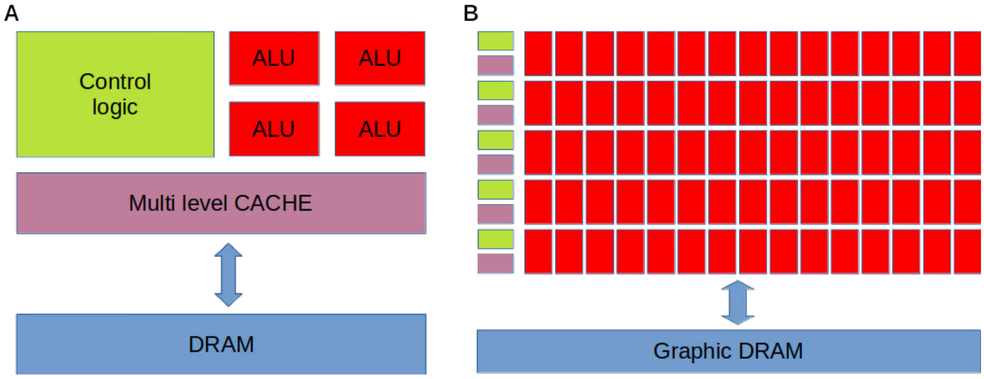


Figure 1.2: Simplified model of a CPU (A) and a GPU (B) architecture. Same color means the same "role". A CPU is equipped with a low number of high complex unit, while a GPU is comprised of an array of simple processing units, which provide the large parallelism.

What emerges from this perspective is that CPUs and GPU are complementary and for a true Scientific Computing acceleration are both necessary. That is why unified application programming interfaces are required, for joint CPU-GPU operations [28].

GPU programming model

CUDA [30] is an extension to the C programming language designed for massively parallel programming in a heterogeneous computing system, such as a modern workstation (or even a smartphone). It introduces new keywords and Application Program Interface (API) that allows the programmer to view the computing system as a host (the CPU) and one or more devices (GPU) with thousands of ALU. A CUDA-extended program is nothing more than a C code with special functions and structures which describes the working of the available device, flagged by special CUDA keywords which are not accepted by a standard C compiler. NVCC is the CUDA compiler which role is to analyse the code and to separate the host and device parts; while the host code is pure straight ANSI C/C++ passed to a tipycal compiler, the CUDA keywords are further processed: the device-specific code is generated in a pseudo-ISA (Instruction Set Architecture) program called PTX which is ultimately compiled, to effectively produce the GPU assembly code [29]. The final executable contains the host machine code, the PTX and the device machine code. During the execution, a JIT compiler can eventually recompile the PTX to accommodate the execution on a different running device. Figure 1.3 shows the scheme of the compilation process for a CUDA code.

Writing GPU code is not an easy task, as it requires profound knowledge on the hardware, programming model, algorithm techniques and numerical methods, but it is required for the modern computing intense scientific research. There are however some caveats which greatly help in the design of GPU software. One alternative has been used to develop the software for this thesis (chapter 2.4): linear algebra libraries and Deep Learning tools nowadays provides a set of reasonable easy to use building blocks which can be employed to *describe* in a high-level language (Python) the GPU operations. The provided acceleration is obviously sub-optimal, but new accelerated and then production-ready algorithms can be easily prototyped.

Multi-GPU programming

The current trend is to move toward a multi-GPU environment; let us take as an example the inverse problem of model fitting discussed previously, imagining that the fitting is done throughout an optimisation process (discussed further in the text in chapter 3.4). To actually take advantages of this new layered parallelism, we need to *reformulate* an algorithm in order

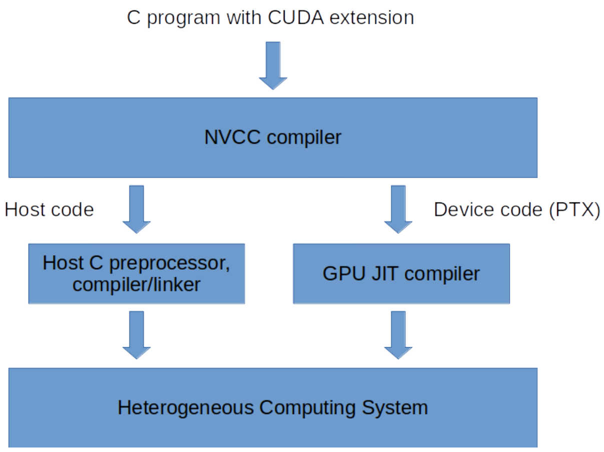


Figure 1.3: Cuda model generation: from a C code with cuda extension, two different paths are generated, one for the host (CPU) and one for the GPU (device).

to implement the following steps:

1. the current model (composed by a computational graph and the parameters found till that iteration) is replicated from a master GPU on all the other GPUs. This should be a one-time only operation.
2. a set of data (a batch) is loaded in the system RAM and is *scattered* among the available GPUs
3. the *forward pass* is executed, then the loss and the gradients are calculated parallelly on each device, exploiting data parallelism on the batches.
4. an *"all reduce"* operation takes all the output from each GPU and actuates modification on the model on the master GPU.

These are the kind of operations that are currently employed for e.g. a ptychography dataset reconstruction on such systems. A second path is however possible if the algorithm is written to exploit better the inherent parallelism of the problem; this is especially true when the degree of independence between the data is high. In a parallel beam CT reconstruction each row is independent from the others, than it can be processed concurrently with the others. In chapter 5.2 a novel multi-GPU system for CT reconstruction will be described. In ptychography, one can increase artificially the independence of each acquisition by intelligently splitting the dataset among many GPU. This solution is described in chapter 2.7.

1.4 Wave propagation basics

The Maxwell equations governs the evolution of classical electromagnetic fields in space and time. Combining the free-space Maxwell equations [31, 32, 18, 33, 34] leads to the following d'Alembert equations for the electric field $\mathbf{E}(x, y, z, t)$ and magnetic field $\mathbf{B}(x, y, z, t)$ in free space:

$$\left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2\right)\mathbf{E}(x, y, z, t) = \mathbf{0}, \tag{1.4.1}$$

$$\left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2\right)\mathbf{B}(x, y, z, t) = \mathbf{0}. \tag{1.4.2}$$

Here, (x, y, z) are three-dimensional spatial coordinates in the Cartesian space, t is time, c is a speed given by

$$c = \frac{1}{\sqrt{\mu_0 \varepsilon_0}}, \tag{1.4.3}$$

the Laplacian in three spatial dimensions is:

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}, \tag{1.4.4}$$

In the case of linearly polarized, monochromatic and fully coherent illumination, the components of the fields in Eq. 1.4.1 and 1.4.2 are completely decoupled, not only mutually between the two equations, but also in their own. The two equations can be split in six scalar equations of the spatial coordinates with the same structure taking into accounts uncorrelated coordinates. Then only one equation in the generic disturbance $\Psi(x, y, z, t)$ can be studied, propagating its solution to the physical quantities of interest (E, B).

$$\left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2\right)\Psi(x, y, z, t) = 0. \tag{1.4.5}$$

From the mathematics point of view, it is convenient [32] to treat $\Psi(x, y, z, t)$ as a complex function, termed "the wave function", which describes the X-ray field, even if only the real part of this wave function has a physical meaning. The information of the field polarisation is lost by transitioning to a scalar wave theory. This assumption is reasonable in many paraxial imaging and diffraction contexts, such as the one encountered here. [35, 33]

Throughout the complex formalism, at each point (x, y, z) in space, and for each instant of time t, $\Psi(x, y, z, t)$ is a complex number, described by magnitude and phase:

$$\Psi(x, y, z, t) = \sqrt{I(x, y, z, t)}e^{j\phi(x,y,z,t)} \tag{1.4.6}$$

Following [34], the magnitude of $\Psi(x, y, z, t)$ has been denoted as

$$|\psi(x, y, z, t)| = \sqrt{I(x, y, z, t)} \tag{1.4.7}$$

where $I(x, y, z, t)$ is the intensity of the field, defined by:

$$I(x, y, z, t) = |\Psi(x, y, z, t)|^2 = \Psi(x, y, z, t)\Psi(x, y, z, t)^* \tag{1.4.8}$$

In the following of the text, the predominance role of this quantity will emerge, as it is the only measurable at X-ray wavelenghts. The phase of $\Psi(x, y, z, t)$ has been denoted $\phi(x, y, z, t)$. Fixing the instant of time t [31], surfaces of constant phase may be identified with wave-fronts of the X-ray field [32]. These fronts move from the source at the speed of light c in vacuum. In the case of full monochromaticity and perfect coherency, the phasor formalism can be exploited like in the case of electrical circuits, therefore the field oscillates at any point in space and time with a fixed angular frequency ω:

$$\Psi(x, y, z, t) = \psi_\omega(x, y, z)e^{-j\omega t} \tag{1.4.9}$$

where

$$\omega = 2\pi f = ck, \tag{1.4.10}$$

f denotes the frequency, and k is the wavenumber corresponding to the wavelength λ in free space. These quantities are related to the following equation:

$$k = 2\pi/\lambda. \tag{1.4.11}$$

The phasor formalism allows decomposing a polychromatic field into its monochromatic components, by using the Fouerier integral [33]:

$$\psi(x, y, z, t) = \frac{1}{\sqrt{2\pi}} \int_0^\infty \psi_\omega(x, y, z)e^{-j\omega t} d\omega \tag{1.4.12}$$

Both in equation 1.4.12 and 1.4.9, each monochromatic component of the field is represented by the product of a spatial part $\psi(x, y, z)_\omega$ and an harmonic time-dependent term.

By direct substitution of Eq. 1.4.9 into Eq. 1.4.5 gives the *Helmholtz equation* in vacuum [31, 18]:

$$\left(\nabla^2 + k^2\right)\psi_\omega(x, y, z) = 0. \tag{1.4.13}$$

Eq. 1.4.13 typically suffices for describing even poly-monochromatic imaging experiments; in fact, the spectral decomposition in 1.4.12 enables the description of imaging processes with a time-independent approach for each monochromatic component, which can be combined to obtain a complete description of the phenomenon [33].

1.5 Angular Spectrum

Assuming that a source of radiation lays on a plane $z = z_0$, a typical electromagnetism problem is to calculate the field produced in the upper half of vacuum-filled three-dimensional space, thus in $z > z_0$. The main approach used to solve *the diffraction problem* is to calculate the field at any distance z, by knowing the initial field at z_0 and the distance Δz. One approach that is helpful in the following of this thesis is the *Angular spectrum method.* This formalism allows to create a *propagation operator* that applied on the boundary field yields its propagated version downstream any other parallel plane. This operator is eventually translated in a GPU-accelerated computer program to compute this computationally intense operation.

Two general solutions of the previous Helmotz equation 1.4.13 are:

- A time independent *plane wave* of the form [18, 36, 35]:

  $$\psi_\omega(\mathbf{r}) = e^{j\mathbf{k}\cdot\mathbf{r}} = e^{j(k_x x + k_y y + k_z z)} \tag{1.5.1}$$

  where **k** is the *wavevector* which describe the propagation direction of the wave $\mathbf{k}/|\mathbf{k}|$, subjecto to:

  $$k = \sqrt{k_x^2 + k_y^2 + k_z^2} = \frac{2\pi}{\lambda} \tag{1.5.2}$$

  and **r** is the position vector

  $$\mathbf{r} = r_x\mathbf{x} + r_y\mathbf{y} + r_z\mathbf{z} \tag{1.5.3}$$

- A *spherical wave* of the form [36, 18]:

  $$\psi_\omega(\mathbf{r}) = \frac{1}{|\mathbf{r} - \mathbf{r_0}|} e^{j\mathbf{k}\cdot[\mathbf{r} - \mathbf{r_0}]} \tag{1.5.4}$$

> in which $\mathbf{r_0}$ is the position vector of the source point around which concentric "shells" of constant phase surfaces emerge. Unlike the plane wave, there is no single well defined direction of propagation, but the field propagates isotropically in the space; each unitary radius $[\mathbf{r} - \mathbf{r_0}]/|\mathbf{r} - \mathbf{r_0}|$ defines in fact a direction of propagation. The denominator is due to the energy conservation principle, that for an expanding wave requires a decaying amplitude of the same factor.

From the Eq. 1.5.3 in the rectangular space, we can define a perpendicular $\mathbf{r}_\perp$ given by components in (x,y):

$$\mathbf{r}_\perp = r_x\mathbf{x} + r_y\mathbf{y}; \tag{1.5.5}$$

the same reasoning can be applied on the $\mathbf{k}$ vector prodcing:

$$\mathbf{k}_\perp = k_x\mathbf{x} + k_y\mathbf{y}; \tag{1.5.6}$$

this formalism allows defining a field by looking at its slices in z. By exploiting the 2D Fourier Theorem it can be observed that the spectrum decomposition (in the spatial frequency meaning) provides a decomposition of the field in its elementary plane wave components, each one directed radially [18, 32]. Each radial direction is defined by: i) an angle α measured between the plane wave component's wave vector and the x axis, and ii) an angle β confined between the wave vector and the y axis. Spatial frequencies, can thus be understood as the Fourier decomposition of a complicated wave into plane waves travelling at different angles with respect to the optical axis.

The Fourier decomposition defined by Eq. 1.5.7

$$\psi_\omega(\mathbf{r}_\perp, z) = \frac{1}{2\pi}\int \hat{\psi_\omega}(\mathbf{k}_\perp, z)e^{j[\mathbf{k}_\perp\cdot\mathbf{r}_\perp]}d\mathbf{k}_\perp \tag{1.5.7}$$

being a wave, has to obey to the Helmotz Eq. 1.4.13. Plugging in Eq.1.5.7 in Eq. 1.4.13 it can be seen that the Fourier Spectrum of a field at a plane z can be calculated (predicted) by knowing its spectrum at another plane $z = z_0$.

$$\hat{\psi}(\mathbf{k}_\perp, z) = \hat{\psi}(\mathbf{k}_\perp, z_0)e^{j(z-z_0)\sqrt{k^2-k_\perp^2}} \tag{1.5.8}$$

Once the initial field at $z = z_0$ is known, the output field depends only on the relative distance $z - z_0$ between the source and observation planes, not on the absolute one.

Discarding the solutions for *evanescent waves* [31] in which the exponential becomes real and describes a decay-to-zero in magnitude, we are interested in propagating solutions given by $k^2 > k_\perp^2$, in which the exponential term oscillates [18]. To obtain the propagated field from its spectrum, we take the inverse 2D Fourier transform

$$\begin{aligned}\psi_\omega(\mathbf{r}_\perp, z) = \psi_\omega(x, y, z) = \\ = \frac{1}{2\pi}\int \hat{\psi}(\mathbf{k}_\perp, z_0) e^{j(z-z_0)\sqrt{k^2-k_\perp^2}} e^{j[\mathbf{k}_\perp\cdot\mathbf{r}_\perp]} d\mathbf{k}_\perp\end{aligned} \tag{1.5.9}$$

The angular spectrum method can be casted in an *operator* following the formalism of [36] with the following expression:

$$\mathcal{D}_{\Delta_z}\{\} = \mathcal{F}^{-1} e^{j\Delta_z\sqrt{k^2-k_\perp^2}} \mathcal{F}\{\}. \tag{1.5.10}$$

The operator notation has to be read from right to left, meaning that for a composite operator such as 1.5.10, the first to apply on the quantity between brackets is the rightmost one.

1.6 Paraxial approximations

From a theoretical perspective, the very nature of the electromagnetic waves precludes the generation of perfectly confined light transport, without having to physically confine the space [33], however it is perfectly possible to generate fields with a *small angular spread* [32, 33, 34], for which the electromagnetic energy is highly confined within a small lateral portion of the space. Such a field is called *paraxial*. In the aforementioned case, the *diffraction problem* presented in the previous section in its accurate from [32, 18], can be solved taking into account many simplifications that become more and more restrictive. By taking the approximation below:

$$\sqrt{k^2 - k_\perp^2} \approx k - \frac{k_\perp^2}{2k} = k - \frac{k_x^2 + k_y^2}{2k} \tag{1.6.1}$$

and exploiting the operator notion defined before, one obtain

$$\mathcal{D}_{\Delta_z}\{\} = e^{jk\Delta_z} \mathcal{F}^{-1} e^{\frac{-j\Delta_z k_\perp^2}{2k}} \mathcal{F}\{\}. \tag{1.6.2}$$

Eq. 1.6.2 is the *Fresnel operator*, which is valid for $k_\perp^2 \ll k^2$, meaning that a the maximum value of $k_\perp^2$ for which the field is not negligible has to be smaller than the wavevector k.

Expressing the relation in its constitutive terms:

$$(\frac{2\pi}{\delta_a})^2 \ll (\frac{2\pi}{\lambda})^2 \tag{1.6.3}$$

it can be seen that the displacement δ_a on the (x,y) plane $\mathbf{r}_\perp$ has to be greater than the wavelength. The equation form of this Fresnel operator 1.6.2 is called *Fresnel-Kirchoff diffraction integral* [18] and its discrete version is used extensively for diffraction patterns computation. In the previous equation, the operator involves using two Fourier Transform. By further manipulation, it is possible to arrive at a convolution based form:

$$\psi_\omega(\mathbf{r}_\perp, \Delta_z) \approx \psi_\omega(\mathbf{r}_\perp, z_0) \circledast P(\mathbf{r}_\perp, \Delta_z) \tag{1.6.4}$$

where the *Fresnel propagator* $P(\mathbf{r}_\perp, \Delta_z)$ is given by:

$$P(\mathbf{r}_\perp, \Delta_z) = -jk\frac{1}{2\pi}e^{jk\Delta_z}e^{jk\frac{\mathbf{r}_\perp{}^2}{2k}}. \tag{1.6.5}$$

The aforementioned operator resembles in its expression the spherical wave in Eq. 1.5.4, if the phase of the latter is approximated applying the condition in Eq. 1.6.1: the formulation of the Fresnel propagation of a monochromatic paraxial field can thus be described by the convolution with a spherical wave, thus reformulating the Huygens's principle.

By exploiting the convolution-Fourier Transform duality, a new operator can be defined which propagates the field in $z = z_0$ producing:

$$\psi_\omega(\mathbf{r}_\perp{}', \Delta_z) \approx \frac{-jke^{jk\Delta_z}}{\Delta_z}\exp[\frac{jkr_\perp^{'2}}{2\Delta_z}]\mathcal{F}\{\psi_\omega(r_\perp, z_0)\exp[\frac{jkr_\perp^2}{2\Delta_z}]\}_{\mathbf{k}_\perp = \frac{\mathbf{kr}'_\perp}{\mathbf{\Delta_z}}} \tag{1.6.6}$$

The Eq.1.6.6 employs just one Fourier Transform of the incident field defined by a mapping of the spatial frequencies given by:

$$\mathbf{k}_\perp = \frac{\mathbf{kr}'_\perp}{\mathbf{\Delta_z}} \tag{1.6.7}$$

The Fresnel approximation is usually implemented numerically following this approach, due to its processing speed.

If we define the *Fresnel Number* with the following expression [18, 32, 37]:

$$F = \frac{D_{max}^2}{\lambda\Delta_z} \tag{1.6.8}$$

where D_{max} is the maximum diameter of the illumination area of a sample, we can set a condition for a further approximation: if $F << 1$, thus if the distance is way larger than the maximum aperture D_{max}, the inner exponential in Eq.1.6.6 can be omitted leading to a negligible error, producing the so-called *Fraunhofer approximation.*

$$\psi_\omega(\mathbf{r}_\perp{}', \Delta_z) \approx \frac{-jke^{jk\Delta_z}}{\Delta_z} \exp[\frac{jkr_\perp^{\prime 2}}{2\Delta_z}]\mathcal{F}\{\psi_\omega(r_\perp, z_0)\}_{\mathbf{k}_\perp=\frac{\mathbf{kr}'_\perp}{\mathbf{\Delta_z}}} \tag{1.6.9}$$

Figure 1.4 shows the variation of the Fresnel number F for a set of distances (x axis) and wavelengths (y axis). As in Fig. 1.4 the colormap is diverging, it is easy to spot (the white part) the transition between the near and the far-field approximation.

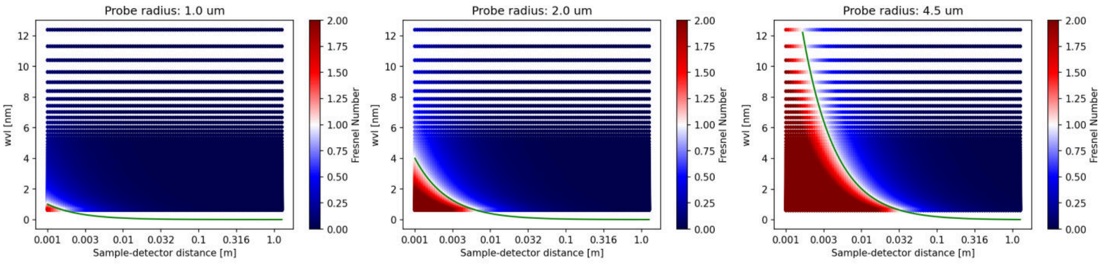


Figure 1.4: Variation of the Fresnel number for different energies and aperture size. The curve for F=1 is highlited, as it acts like a separation line for choosing the correct approximation. For a Fresnel number « 1, the Fraunhofer approximation holds, otherwise the experiment is configured in the near-field regime. The values of F are clipped to 2 to highlight the "separation line".

1.7 X-ray-matter interaction

In order to describe the effect of the medium onto the propagation properties we need to introduce in the previous equation 1.4.13 the complex refraction index $n_\omega(x, y, z)$ that depends upon the wavelength and the position. The presence of the medium, described by the refractive index, alters the wavelength which is observed in vacuum as follows [31]:

$$\lambda = \frac{\lambda_0}{n_\omega(x, y, z)}, \tag{1.7.1}$$

hence

$$k = k_0 \, n_\omega(x, y, z). \tag{1.7.2}$$

The frequency and spatio-variant characteristic of $n_\omega(x, yz)$ describes the sample due to the connection with the electrical and magnetic permittivity ϵ, μ. For a non-magnetic medium, $\mu = \mu_0$, then:

$$n_\omega(x, y, z) = c\sqrt{\epsilon_\omega(x, y, z)\mu_0} = \sqrt{\frac{\epsilon_\omega(x, y, z)}{\epsilon_0}} \tag{1.7.3}$$

With the introduction of the complex refraction index of the medium, the vacuum Helmholtz equation Eq. 1.4.13 then becomes the Helmholtz equation in the presence of i) linear, ii) isotropic, iii) static, iv) non-magnetic, v) spatially slowly varying scattering media, expressed in Eq. 1.7.4 [18, 34]

$$\left[\nabla^2 + k^2 n_\omega^2(x, y, z)\right] \psi_\omega(x, y, z) = 0. \tag{1.7.4}$$

Therefore, the impressed current density $J(x, y, z)$ in the medium is zero everywhere. This means that no sources are present [31].

A monochromatic paraxial field $\psi_\omega(x, y, z)$ can be described by:

$$\psi_\omega(x, y, z) \equiv \tilde{\psi}_\omega(x, y, z)e^{jkz}. \tag{1.7.5}$$

In Eq. 1.7.5, it can be seen the product of a z-directed plane wave e^{jkz} and a perturbing envelope $\tilde{\psi}_\omega(x, y, z)$. The variation of the complex envelope in z is much smaller than its variation in a transverse plane (x,y) [33, 34]. Due to the fact that the magnitude of a complex exponential is unitary, the intensity of the field is:

$$I_\omega(x, y, z) = |\tilde{\psi}_\omega(x, y, z)|^2 = |\psi_\omega(x, y, z)|^2 \tag{1.7.6}$$

where it appears evident that intensity variations in $\psi_\omega(x, y, z)$ (what can be observed) are produced by the variation of $\tilde{\psi}_\omega(x, y, z)$.

By observing that in Eq. 1.7.4, it is possible to define the ∇^2 operator like expressed in Eq. 1.7.7,

$$\nabla^2 = \nabla_\perp^2 + \frac{\partial^2}{\partial z^2} \tag{1.7.7}$$

and by assuming that:

- the variations $\Delta\phi(x, y, z)$ within a longitudinal distance $\Delta z = \lambda$ are much smaller than the field itself [33]:

$$\frac{\Delta\psi(x, y, z)}{\psi(x, y, z)} << 1 \tag{1.7.8}$$

- the term containing the second z derivative is small compared to the others [33]

$$\frac{\partial\phi(x, y, z)}{\partial z} << k\phi(x, y, z) \tag{1.7.9}$$

by substituting Eq. 1.7.5 into Eq. 1.7.4:

$$\left(2ik\frac{\partial}{\partial z} + \nabla_\perp^2 + k^2[n_\omega^2(x, y, z) - 1]\right)\tilde{\psi}_\omega(x, y, z) = 0, \tag{1.7.10}$$

where

$$\nabla_\perp^2 \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \tag{1.7.11}$$

is the transverse Laplacian which considers only the (x, y) plane perpendicular to the optic axis z.

Equation 1.7.10 is extremely important due to the fact that allows to make predictions on the field passing through a medium, by knowing the entering field $\psi_0(x, y, z = 0)$ and the refraction index distribution of the medium $n(x, y, z)$. To do so, we need to define an *entering surface* and an *exit window* which "contain" the medium.

The assumptions made on the slowly varying characteristics of the medium allows to approximate all the light rays by straight lines parallel to the z versor; that is why the assumption is referred to as *projection approximation* [18, 34, 33], which holds if we can neglect the diffraction spread within a sample thickness T when compared to the resolution a [18, 34, 37, 38], following:

$$\frac{\lambda}{a}T \ll a, \tag{1.7.12}$$

that in terms of the Fresnel number 1.7.13 is:

$$N_F = \frac{a^2}{\lambda T} \gg 1. \tag{1.7.13}$$

In this condition, the transverse Laplacian in Eq. 1.7.10 may be neglected producing:

$$\frac{\partial}{\partial z}\tilde{\psi}_\omega(x,y,z) \approx \frac{k}{2i}[1 - n_\omega^2(x,y,z)]\tilde{\psi}_\omega(x,y,z). \tag{1.7.14}$$

which for each fixed point (x, y), is a simple linear first-order ordinary differential equation, that can be integrated with respect to z giving:

$$\tilde{\psi}_\omega(x,y,z=z_0) \approx \exp\left\{\frac{k}{2i}\int_{z=0}^{z=z_0}[1 - n_\omega^2(x,y,z)]dz\right\}\tilde{\psi}_\omega(x,y,z=0).$$

The above equation eq: 1.7.15 shows that the exit wave-field $\tilde{\psi}_\omega(x,y,z=z_0)$ may be obtained from the entrance wave-field $\tilde{\psi}_\omega(x,y,z=0)$ via a simple element-wise multiplication by a *transmission function* $\mathcal{T}(x,y)$, defined by:

$$\mathcal{T}(x,y) = \exp\left\{\frac{k}{2i}\int_{z=0}^{z=z_0}[1 - n_\omega^2(x,y,z)]dz\right\} \tag{1.7.15}$$

which can be simplied, considering that the refraction index can be equivalently expressed by [39]:

$$n_\omega = 1 - \delta_\omega + i\beta_\omega \approx 1, \tag{1.7.16}$$

producing:

$$\mathcal{T}(x,y) \approx \exp\left\{-ik\int_{z=0}^{z=z_0}[\delta_\omega(x,y,z) - i\beta_\omega(x,y,z)]dz\right\} \tag{1.7.17}$$

$$\tilde{\psi}_\omega(x,y,z=z_0) \approx \tilde{\psi}_\omega(x,y,z=0)\mathcal{T}(x,y) \tag{1.7.18}$$

The position-dependent phase shift

$$\arg\mathcal{T}(x,y) \equiv \Delta\phi(x,y), \tag{1.7.19}$$

due to the object is:

$$\Delta\phi(x,y) = -k \int \delta_\omega(x,y,z)dz. \tag{1.7.20}$$

and is then dependent only on $\delta_\omega(x,y,z)$, the real part of $n_\omega(x,y,z)$ minus 1. At X-ray wavelenght, a conseguence of having $\delta_w < 1$ is that the phase velocity within a medium $v = c/n$, is larger than the velocity of light, c. This does not, however, violate the law of relativity, which requires that only signals carrying *information* do not travel faster than c. Such signals move with the group velocity, which is slower than c.

Calculating the intensity $I_\omega(x,y,z=z_0)$ of the output field:

$$I_\omega(x,y,z=z_0) = \tilde{\psi}_\omega(x,y,z=z_0)\tilde{\psi}_\omega(x,y,z=z_0)^* \tag{1.7.21}$$

by exploiting the linear attenuation coefficient μ_ω

$$\mu_\omega = 2k\beta_\omega. \tag{1.7.22}$$

the resulting expression is simply the well-known *Beer-Lambert Law.*

$$I_\omega(x,y,z=z_0) = \exp\left[-\int \mu_\omega(x,y,z)dz\right] I_\omega(x,y,z=0). \tag{1.7.23}$$

From the imaging point of view, Eq. 1.7.23 is typically "linearized" by considering the logarithm of the ratio of the incident and output intensity:

$$\log_e \frac{I_\omega(x,y,z=z_0)}{I_\omega(x,y,z=0)} = -\int \mu_\omega(x,y,z)dz. \tag{1.7.24}$$

which is only a function of the imaginary part of $n_\omega(x,y,z)$. This means that given a sample with a given $n_\omega(x,y,z)$ spatial distribution, measuring the intensity of the field at the exit window of the sample $z = z_0$, only absorption information can be acquired. Measuring at $I(x,y,z=z_0)$ therefore is tipically referred as *contact imaging.* This lack of *phase contrast*, in conventional contact X-ray imaging, is unfortunate, since many important structures of interest (such as soft biological tissues) are transparent and the majority of the information is encoded in the phase. For such samples, we can say that they possess a low absorption-contrast.

1.8 A different view

The description given so far involves only the input/output model of the phenomenons that interlink the radiation/matter interaction. The same phenomenon of absorption and phase shift can be also explored from the particle point of view (due to the particle/wave dual nature of the photons), governed by the interaction of the photons (the incident radiation) with the matter. From the microscopist point of view, an X-ray photon interacting with an atom can be or absorbed or scattered [19]. This can happen in the following way [37]:

- **Absorption**: From an atom in its ground state (the lowest energy configuration), if the photon has sufficient energy, it can remove one electron from inner shells of the atom, moving it to the continuum. At this point, different competing processes can be triggered to restore the energy equilibrium, e.g the emission of other photons (Fluorescence), or electrons (Auger electron emission).
- **Elastic (Rayleigh, coherent) scattering**: the incident photon is scattered by the atom with no transfer of energy; the mechanism is described by electrons that oscillate in the electric field of the incident photon and re-irradiate a wave at the same frequency. The elastically scattered photon is *locked in phase* with the incident photon and the interaction is described by a phase shift
- **Inelastic (incoherent, Compton) scattering**: a photon is scattered inelastically by imparting kinetic energy to an electron. A photon at a different energy is released. Conservation of momentum and energy leads to relationship between the wavelengths of the two photons and the scattering angle θ.

The *cross section* σ [19, 37] is a fundamental quantity in physics (measured in "barns", which is simply an area unit, $1b = 10^{-24} cm^2$), which quantifies the probability that a particular interaction takes place. In the contest of absorption, it describes the ability of a "particle" (an electron, an atom, a molecule, a crystal, etc) to absorb the incident photon. Absorption can also be described in terms of the *mean free path* [37] which represents the thickness over which the unaffected fraction of the beam declines to $1/e = 0.37$ of its original value. Thus, we have just provided a different description of the Lambert-Beer law in Eq. 1.7.23.

For each of the aforementioned interactions (absorption, coherent scattering, incoherent scattering) a different cross section can be calculated (or measured), for different materials at different energies. In the soft X-ray window, at energies below 30 keV, (well within our interest), the total cross section σ experienced by an incident photon [37, 39], is composed by

$$\sigma = \sigma_{Rayleigh} + \sigma_{absorption} \tag{1.8.1}$$

Observing the behaviour of the cross section for each of the aforementioned phenomenons, it can be seen that in X-ray imaging if a photon is scattered, it is far more likely that any subsequent interaction of that photon will be an absorption event [37]. This in turn means that multiple scattering are infrequent. That is why the *single scattering* or the so called *Born approximation, (kinematical scattering)* [18] is valid in most of the cases of interest (of this thesis), where the sample is not so thick and the probability of subsequent scattering interactions is low [40].

If the sample is reasonably thin, the Born approximation states that the field exiting a medium filled region is provided by [18, 37]:

$$\psi(\mathbf{r}) \approx \psi_0(\mathbf{r}) + f[\psi_0(\mathbf{r}), n(\mathbf{r})] = \psi_0(\mathbf{r}) + \psi_s(\mathbf{r}) \tag{1.8.2}$$

where $\psi_0(\mathbf{r})$, is the input field that propagates like in free-space also within the medium and a function $f[\quad]$ which describes the resulting scattered wave in terms of the incident wave $\psi_0(\mathbf{r})$ and the "scattering power" $n(\mathbf{r})$ [37]. The function $f[\quad]$ is the integral of all the spherical waves (free-space outgoing Green functions) which emerge from the scattering centres (where a scattering event occurred) [18].

Observing the scattered radiation in far-field, the spherical wavefronts become plane waves. It is then possible to specify a propagation vector $\mathbf{k}^{'}$ for the scattered radiation. Supposing that $\mathbf{k_i}$ represents the incident radiation wavevector, one can measure the difference in phase between the scattered wave from a volume element around the origin and one around the observation point $\mathbf{r}$ [19]:

$$\Delta_\phi = (\mathbf{k} - \mathbf{k}^{'}) \cdot \mathbf{r} = \mathbf{Q} \cdot \mathbf{r} \tag{1.8.3}$$

where $\mathbf{Q}$ is the so called *scattering vector* or *wavevector transfer*; all the possible scattering vectors $\mathbf{Q}$ have their tips lying on a surface of spherical shape, that is called *Ewald Sphere* [18, 19, 37], which describes the region in reciprocal space (at far field) over which one obtains information during a scattering experiment using a single wavelength λ [37]

1.9 Coherence

The techniques described in this thesis exploits heavily the coherence of the light used. Coherence theory is a complex topic; in this section only some basic but necessary aspects are presented. So far *perfectly known* fields have been presented, completely described by their spatio-temporal relations. A realistic wavefield, instead, is the result of a stochastic process, arising from finite size of the source and limited spectral purity. Formally, coherence refers to the degree of phase correlation existing in the wavefield. From the aforementioned aspects we can define:

- a coherence length l_c which is the distance along the longitudinal axis for which two wavefront are in phase and depends on the degree of spectral/energy purity of the wavefield (longitudinal coherence) (*longitudinal coherence*) [19, 37]

  $$l_c = \frac{\lambda E}{\Delta E} = \frac{\lambda^2}{\Delta \lambda} \tag{1.9.1}$$

  where E is the energy of the radiation and ΔE the "energy bandwidth" (related to frequency and wavelength)

- a coherence width w_c, also expressed as *transverse coherence* which refers to the degree of phase correlation between wavefront positions that are separated transversely to the wave propagation direction

  $$w_c = \frac{\lambda \Delta_z}{2a} \tag{1.9.2}$$

  where a is the lateral size of the source that is distant from the observation point Δ_z.

The degree of coherence can be calculated by the fringe visibility [32, 37] defined as:

$$V = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \tag{1.9.3}$$

by measuring the maximum and minim intensity in a simple interferometry experiment. If the contrast is low, fringes are not visible at all, thus the brightest and darkest intensity coincide, indicating that there is no coherence at all. On the opposite side, when I_{min} is 0, the beam is fully coherent. In a partial coherence situation, the fringes appear "smeared", "blurred". The partial coherence of a diffracting wave can effectively modelled as a Gaussian blur of the intensity that one would have observed if full coherence was on the table [41]. More formally, the coherence of waves is characterised by their spatio-temporal correlation, defined by the *mutual coherence function* [32, 37, 19]:

$$\Gamma(\mathbf{r_1}, \mathbf{r_2}, \tau) = \langle \Psi(\mathbf{r_1}, t+\tau)\Psi(\mathbf{r_2}, t)^* \rangle \tag{1.9.4}$$

which takes into account a wavefield Ψ at the position $\mathbf{r_1}$ and $\mathbf{r_1}$. The time average in the previous equation is done on an interval T that is longer than the fluctuation time of Ψ [42]. By normalising this quantity by the squared product of the field intensity, the *complex degree of coherence* is introduced:

$$\gamma(\mathbf{r_1}, \mathbf{r_2}, \tau) = \frac{\Gamma(\mathbf{r_1}, \mathbf{r_2}, \tau)}{\sqrt{\Gamma(\mathbf{r_1}, \mathbf{r_1}, 0)\Gamma(\mathbf{r_2}, \mathbf{r_2}, 0)}} = \frac{\Gamma(\mathbf{r_1}, \mathbf{r_2}, \tau)}{\sqrt{I_1 I_2}} \tag{1.9.5}$$

the magnitude of γ is a number between 0 and 1 indicating respectively complete incoherence or complete coherence. The magnitude of γ corresponds to the fringe visibility v:

$$V = |\gamma| \tag{1.9.6}$$

In the context of this thesis only *quasi monochromatic* fields are considered, then

$$\frac{\Delta E}{E} < 1 \tag{1.9.7}$$

It is meaningful to denote the *mutual intensity function* which is obtained from the mutual coherence function by setting the lag to 0

$$J(\mathbf{r_1}, \mathbf{r_2}) = \Gamma(\mathbf{r_1}, \mathbf{r_2}, \tau = 0) \tag{1.9.8}$$

It can be proved that the mutual intensity function can be expressed as the decomposition of N mutually incoherent modes ψ_m [43]:

$$J(\mathbf{r_1}, \mathbf{r_2}) = \sum_N \eta_n \psi_n(\mathbf{r_1}) \psi(\mathbf{r_2}_n)^* \tag{1.9.9}$$

where η_n is a real number representing the occupancy of the nth mode [43]. Using this formulation one can define the intensity of a partially coherent field with [43, 44]:

$$I(\mathbf{r}) = \sum_N \eta_n I_n(\mathbf{r}) \tag{1.9.10}$$

This important relation will be used extensively in the thesis.

1.10 Summary

The thesis is built on three pillars which are Inverse problems, Computing and Physics. A software implementation of a reconstruction algorithm for

X-ray microscopy is the final product and represents the thick web that links these topics. I tried to stress out the relations between these elements which require to be all grasped at the same time. Inverse problem theory allows to structure the solution, physics concepts such as wave propagation and coherence are required for the numerical model and computing elements are mandatory for a fully working implementation.

Chapter 2

CDI/Ptychography

In this chapter, conventional microscopy methods will be presented, jointly with the relative TwinMic setup scheme. TwinMic is the spectromicroscopy laboratory at the Elettra synchrotron facility. It provides twofold analisys, trough X-ray fluorescence, and trough conventional (Scanning) transmission X-ray microscopy, therefore the name, "two microscopes". Starting with a rapid description of the microscope and of the conventional image modalities would allow to appreciate the importance of Coherent Diffraction Imaging methods such as plain CDI and ptychography. We will introduce the ptychography forward image model with the help of simulations carried out with the software framework developed during this research. A new reconstruction algorithm will be proposed [1] as a contribution, in order to make the RPIE algorithm capable of dealing with partial coherence and is confronted against other current state-of-the-art reconstruction algorithms. A multi-GPU system will be also described. Many different reconstruction examples will close the chapter.

2.1 Conventional Imaging

X-ray transmission microscopy is a well-established and mature family of imaging techniques that aims at investigating the microscopic structure of thin, non-crystalline materials. Formally, X-ray microscopy allows the imaging of structures on length scales that are intermediate between those probed by optical ($< 1\,\mu\text{m}$) and electron based techniques ($1\,\text{Å}$). This experimental approach is becoming extremely popular for life-science research, due to "a combination of resolution, penetrating power [...] compatibility with wet specimen [...] and ease of image interpretation" [46]. The common approach

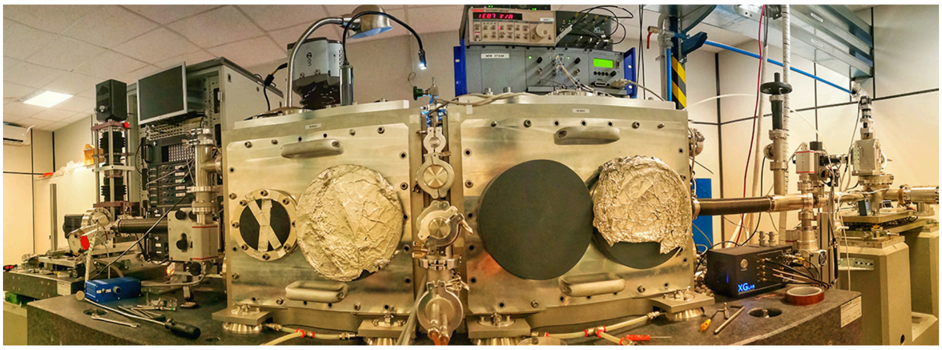


Figure 2.1: The TwinMic end station [45]. The beam arrives from the tube on the right and the detector is the black box at the far left, in line with the aforementioned tube. The experimental chamber is a structure composed by three pieces: i) the optics chamber, on the right; ii) the sample chamber (in between); iii) and the mirror chamber on the left, which contains the optics and the servo-mechanisms related to the conventional STXM experiments. Note that this kind of panorma picture is obtained trough computational imaging.

to this type of imaging consists in the observation of the emerging radiation obtained illuminating a sample, essentially looking through it. This kind of observation is called "transmission".

The X-ray microscope

Common to all microscopies is the need for efficient focusing optics. The increasing demand for high resolution imaging in research areas such as biology, medicine, material science and nanotechnologies has rendered conventional lens-based imaging techniques impractical, due to the fact that the resolution is restricted by the quality and the resolving power of the lens used. At X-ray wavelengths, it is extremely difficult to produce focusing system of high quality: X-ray lenses typically have numerical apertures (N.A.) of 0.005 or less, and efficiencies ranging from 60 percent to as low as single digit percentages [37], while objective lenses for visible-light microscopes, if exploiting oil-immersion provides N.A. up to 1.4, [47, 37] with efficiencies near 100 percent, and excellent achromatic properties.

Diffractive optics

As refractive optics is difficoult to obtain, diffractive optics such as Fresnel zone plate are extensively used as specialised X-ray lenses for x-ray microscopes. These focusing elements are designed to absorb or to phase shift well-defined zones of the wavefront, in order to create a predefined interference pattern at a predefined distance. Figure 2.2 shows a numerical simulation of the field downstream a binary zone plate. The brightest space in panel (B) shows the position of the focus (for the wavelenght and the set of size parameters used in the simulation), which lies correctly on the optical axis (due to the central simmetry of the structure).

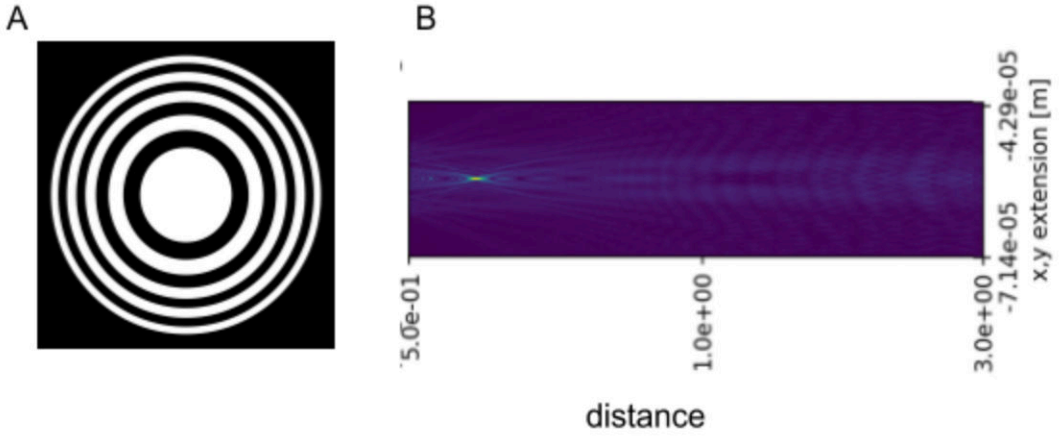


Figure 2.2: A binary zone plate (panel A) and its simulated field (panel B). The zone plate is placed on the xy plane at the origin, while z is the optical axis. Due to the central symmetry of the structure, we can observe a slice in the xz plane that passes on the zone plate diameter. The bright spot in panel B is the focus of the zone plate.

The diameter D and the outermost δ_r zone width are the most critical parameter as they define the highest achievable resolution and the beam width at focus. In a Fresnel lens based microscope, the lithographic process bounds the maximum achievable imaging resolution. Computational techniques like CDI and Ptychography were conceived also to remove the need of such an element, or at least to not rely completely on its performances.

Illumination

Till now we discussed the so-called "end-station". An essential element for a microscope is obviously the light. Figure 2.3 presents the general beamline and optical layout. The X-ray beam generated by an undulator (UND SU 5.6 in the figure) is collimated by a vertically arranged cylindrical mirror (M1) into a plane-grating monochromator (M2 and G) in order to adapt the photon flux and energy resolution to the experimental requirements. A toroidal mirror (M3) refocuses the beam into a secondary source (SS), which consists of an array of apertures (with diameter 15 , 25 , 35 , 50 , 75 , 100 or 150 µm) serving as exit slit and defining the mono chromaticity of the beam delivered into the end-station [45].

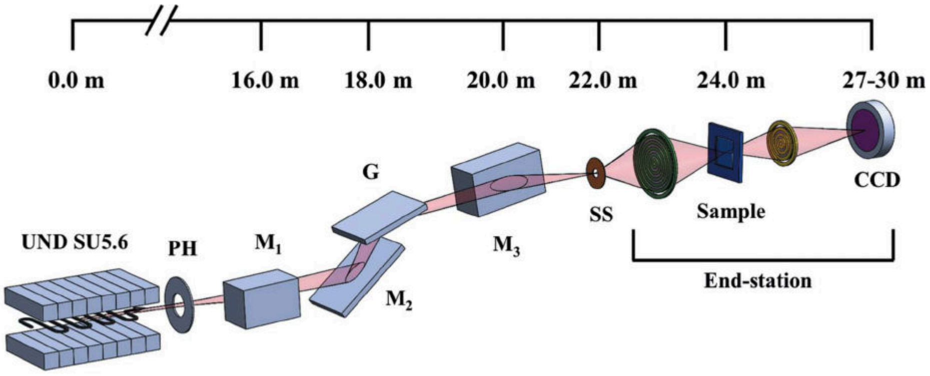


Figure 2.3: Optical scheme of the TwinMic beamline beamline with short undulator X-ray source (SU5.6), pinhole (PH), collimating mirror (M1), plane-grating monochromator (plane mirror (M2) and plane grating (G)), refocusing toroidal-mirror (M3) and secondary X-ray source (SS). Image taken from [45], thanks to Dr. A. Gianoncelli.

When the radiation source has many spatial modes, such as in TwinMic [45], those have to be eliminated to achieve a diffraction-limited focus. The cost is obviously a loss of brilliance, then a trade-off has to be reached. At Twinmic, the solution to increase the coherence consists in using the beamline optics to image the source onto the *secondary source*, which can be adjusted to pass one or several coherent modes [37].

Imaging modalities:

The arrangement of lenses, samples and detector defines mainly two families of microscope techniques:

- in Transmission X-ray Microscope full-field setup (TXM) (figure 2.4), zone plates are used as magnifying lenses which focus the incident light onto the sample (condenser CBS lens in the figure) or focus the transmitted light (objective lens, OZP in the figure) onto the pixelated 2D detector (LECCD in the figure). The Order Sorting Aperture (OSA) is used to remove the high order focuses of the zone plate [37]. The projected image is magnified and is taken in one shot. The ideal magnification of the setup is determined by the ratio of the objective-detector to sample-objective distances, a ratio which can exceed 1000.

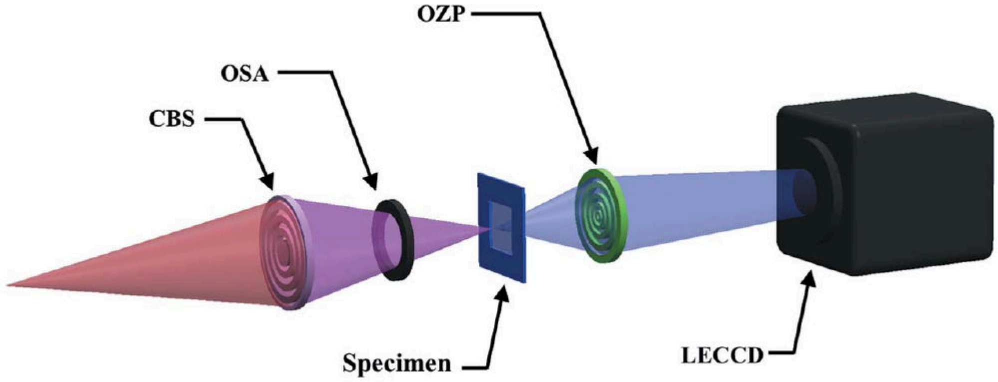


Figure 2.4: The TXM full field setup of the TwinMic beamline. The beam coming from the secondary source is directed towards a condenser zone plane (CBS) which is made "ideal" by the Order Sorting Aperture (OSA) (a small ring). An output zone plate (OZP) is focused on the specimen plane, producing an image directed toward to the detector (LECCD). Image taken from [45], courtesy of Dr. A. Gianoncelli

- In the Scanning Transmission X-ray Microscopy setup (STXM) (figure 2.5), zone plates are used to focus a beam of light onto the specimen, which is mounted on a mechanical stage (sample stage, Fig. 2.6) that allows the specimen to be translated laterally. An STXM acquisition is then completed by raster scanning the sample surface, thus acquiring one pixel at the time. This not only relaxes the sampling requirements of the detector, which can in principle be also "single-pixelated", but also extends the Field Of View (FOV) of each experiment, at the expenses of the acquisition time. However, the use of pixelated detector provides more flexibility and the total intensity can be simply obtained by summing the value at each pixel. A setup of this type can also accomodate X-Ray Fluorescence measurements (XRF), during which the fluorescence spectrum is detected. More details will be presented in chapter 5.1. By raster scanning, a fluorescence map can thus be acquired in conjunction to the absorption image and the two objects can be perfectly superimposed, as have they been collected during the same experiment.

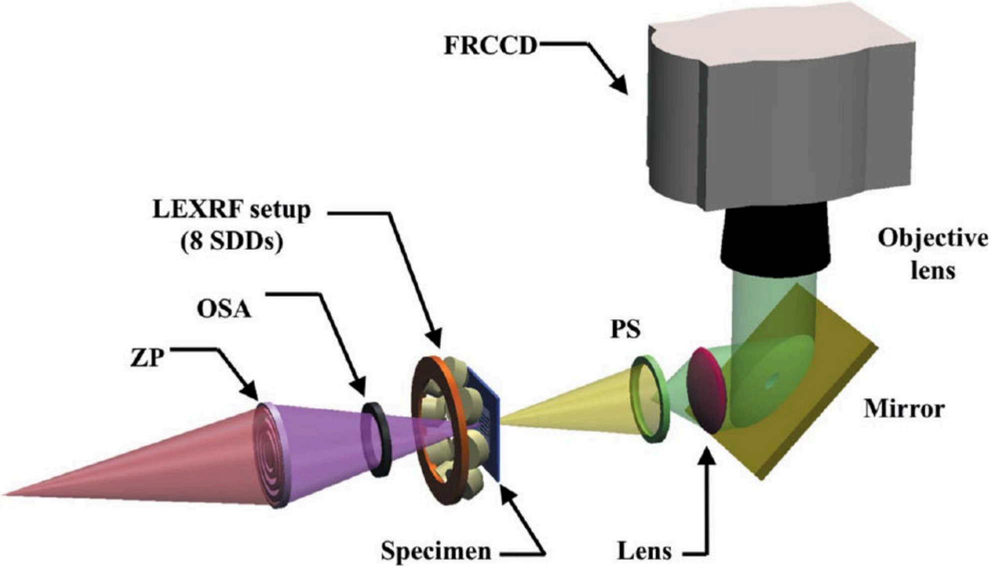


Figure 2.5: The STXM full field setup of the TwinMic beamline. Image taken from [45], courtesy of Dr. A. Gianoncelli

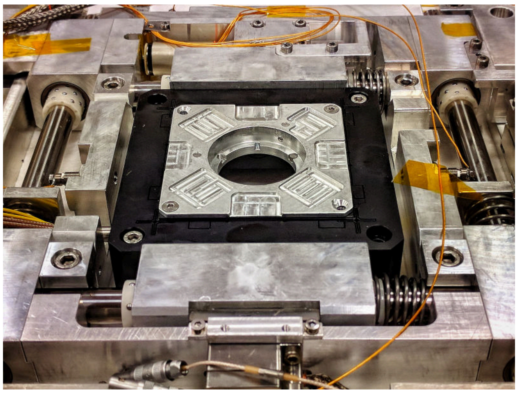


Figure 2.6: TwinMic sample stage, the sample is mounted in the center.

STXM based Phase imaging

Furthermore, an imaging investigation method based on the analysis of the amplitude and phase of the X-rays scattered from materials, provides advantages over conventional absorption-based radiography. Qualitative phase imaging (without an interferometric approach) can be simply accomplished by exploiting an STXM setup: the angular deviation α of a refracted field is in fact directly proportional to the gradient of the phase $\phi(r)$ of the refracted beam [34, 19]

$$\alpha_x = \frac{\lambda}{2\pi}\frac{\partial\phi(x,y)}{\partial x} \tag{2.1.1}$$

$$\alpha_y = \frac{\lambda}{2\pi}\frac{\partial\phi(x,y)}{\partial y}. \tag{2.1.2}$$

The angular deviations can be transformed in a displacement vector [19] by multiplying α_x or α_y by L, which is the sample-to-detector distance.

Indeed, by using a pixelated or a sectorial detector, one can estimate the net displacement of the field and generate a map of the differential phase in x and y. The displacement can be simply estimated by calculating the image centroid (c_x, c_y), that is function of the moment $M_{m,n}$ [23]:

$$M_{m,n} = \sum_{x=0}^{N}\sum_{y=0}^{M} x^n y^m f(x,y) \tag{2.1.3}$$

where x, y are the image coordinates, $f(x,y)$ is the image itself, and (m,n) is the order of the moment. The centroid which describes the total shift in x, y that is correlated to $L\alpha_x$, and $L\alpha_y$ is defined by:

$$(c_x, c_y) = (\frac{M_{1,0}}{M_{0,0}}; \frac{M_{0,1}}{M_{0,0}}) \tag{2.1.4}$$

Theoretically the x, y phase gradients can be used to obtain a quantitative phase information simply by integration. However, this procedure is highly sensitive to noise, causing a well-defined striking artefact that can be seen in figure 2.7: panel a and b shows the horizontal and vertical phase gradients, which integrated by line or by column provides the phase map respectively in panel c and d.

De Jonge method

In the works [48, 49] an interesting Fourier integration method is described, allowing for a way more robust recover of the phase information, in just one complex direct and inverse Fourier transform. The two phase gradients are combined to produce the real image $\Phi(x,y)$ describing the phase:

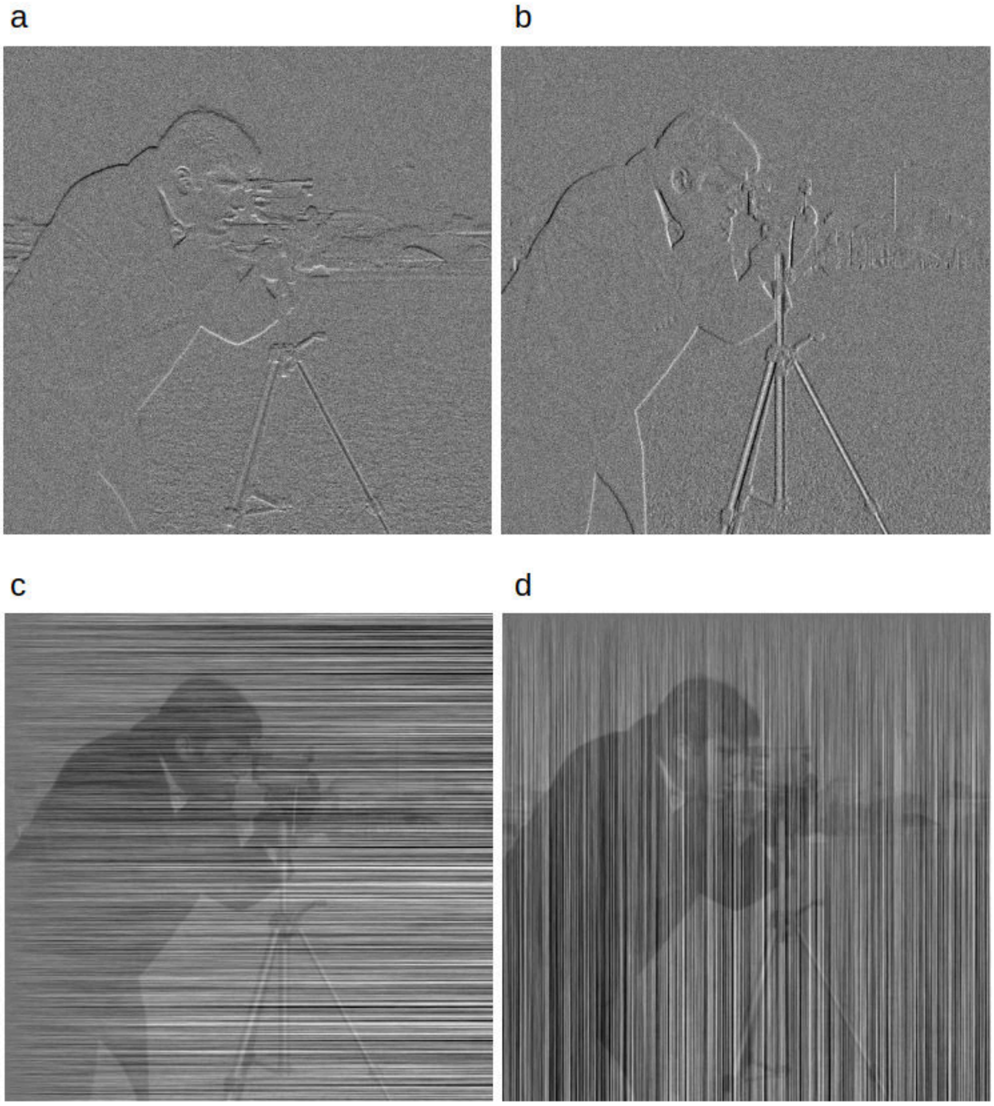


Figure 2.7: Simulation of the retrieval of a phase map using the x, y moments (panel a, b). The integration in the real space lead to poor results, especially in the case of noise. Averaging panel c and d does not help.

$$\Phi(x,y) = \Re\{\mathfrak{F}^{-1}\{\frac{\mathfrak{F}\{L\alpha_x + jL\alpha_y\}(u,v)}{2\pi j(u+jv)}\}(x,y)\} \tag{2.1.5}$$

where L is the sample-to-detector distance, x, y are the coordinates in the real space and u, v are the coordinates in the Fourier space. Eq. 2.1.5 is basically an inverse filtering method applied on the synthetic complex image

$L(\alpha_x + j\alpha_y)$, as the denominator can be seen as the composition of the derivative operator expressed in the Fourier space applied on both the real and imaginary part. Being an inverse filter (e.g. deblur), special care has to be used, in order to avoid the formation of artefacts (e.g. normalisation, division by zero, noise amplification etc.). Figure 2.8 shows the output of the algorithm described in [49] applied on the simulated x, y moments shown in Fig. 2.7, which are the conventional differential phase contrast images that can be obtained; note that to produce those, the operator expressed by the denominator of Eq. 2.1.5 can be used.

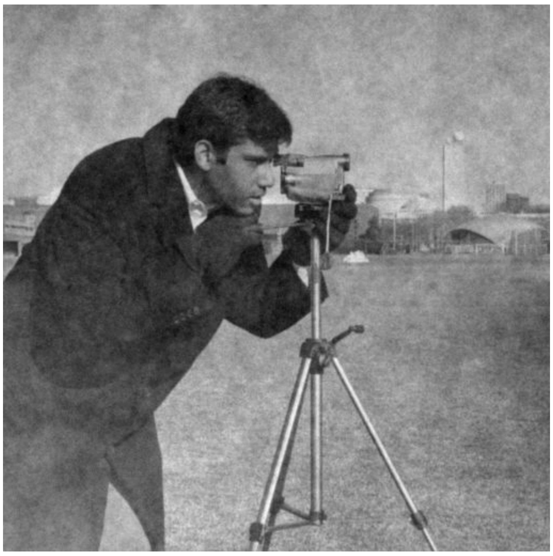


Figure 2.8: For the same data of Fig. 2.7, the method described in [49] provides a better reconstruction.

Figure 2.9 panel a shows the STXM absorbtion image of a diatom obtained during one of our soft-X-ray beamtimes, while panel B shows the output of Eq. 2.1.5 applied on the same raw data. As can be seen, the phase information reveal all the details of the diatom shell. It has to be noted that these images are quite small (100x100 pixels), with a pixel size of 1 µm and also a long measurement is required, as each pixel has to be acquired singularly.

Although the role of X-rays in the field of imaging was quite immediately assigned after their discovery (everyone remembers the famous image of the Röntgen wife finger's), many of these more modern imaging techniques rely tightly on the high-brightness and coherency of modern synchrotron radiation sources and consequently have been developed only in the last decade.

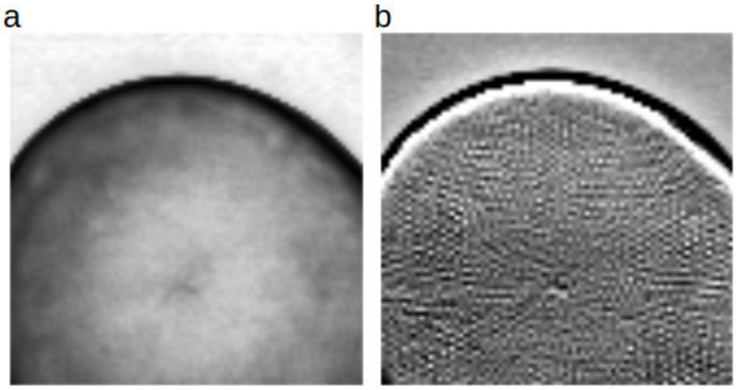


Figure 2.9: STXM scan of a diatom: panel (a) shows the absorption image, while panel (b) shows the phase image obtained by employing the method in [49]. These images have a resolution of 100x100 pixels and shows a FOV of 100 µm × 100 µm.

2.2 Coherent Diffraction Imaging

CDI-based transmission microscopy works by shining coherent light on a specimen and "looking through" it. A sample is illuminated and the diffraction pattern at some distance is recorded by a pixelated detector (Fig. 2.10). Depending on the working regime, on the detector it forms either a direct space (Fresnel diffraction) or a reciprocal space image (Fraunhofer diffraction).

2.2.1 The phase problem

As said in chapter 1.6, we are able to detect only intensities of a field, and this is the image we get on a pixelated detector composed of $N_x \cdot N_y$ pixels. We have clearly an inverse problem, as we want to reconstruct the object (the cause) which produced a given effect (the diffraction pattern); the problem is also ill-posed in the Hadamard sense as the conditions mentioned in chapter 1.2 are not fulfilled.

Given the a priori image formation problem, simply trying to invert the transform matrix is not a feasible solution, due to the fact that additional information is required, since the problem yields $2 \cdot N_x \cdot N_y$ unknowns (real and imaginary part of the wave in each pixel). Indeed, here the amplitude and phase of the wavefront, are only matched by $N_x \cdot N_y$ measured intensities, and an underdetermined system of equations is thus produced. We have then only half of the required information to fully characterise a wavefront. In the literatue this situation is referred as *"the phase problem"*.

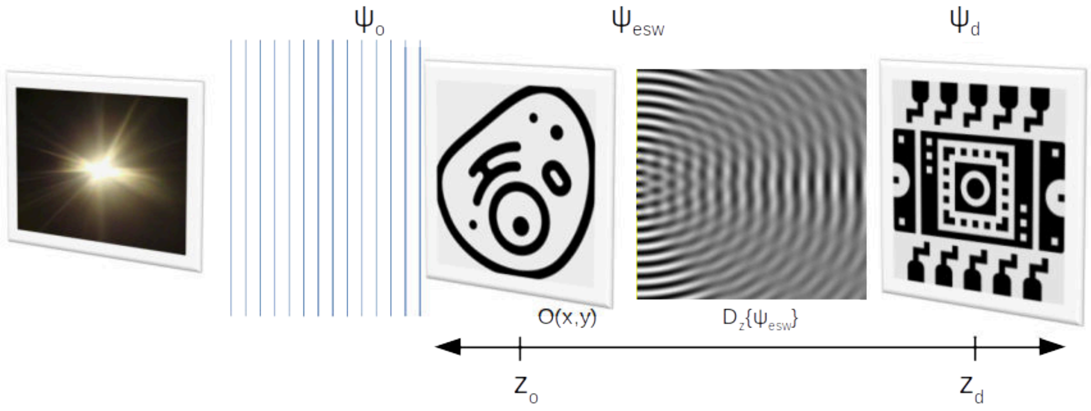


Figure 2.10: A wavefield generated by a coherent source is impressed onto an object characterized by a complex transmission function O(x,y). The illumination interacts with the sample producing an exit-wave that free-space propagates to the detector.

Holography Holography provides one method to solve the phase problem: a coherent wave (reference wave) is introduced and the detected intensity diffraction pattern is related to the *interference* between the scattered field and the "carrier", as described by the following relation:

$$I = |\Psi_s|^2 + |\Psi_r|^2 + |\Psi_s \Psi_r^*| + |\Psi_s^* \Psi_r| \tag{2.2.1}$$

where $|\Psi_s|^2$ and $|\Psi_r|^2$ are the coherent diffraction pattern respectively of the scattered and reference wave, and the other two terms are the holografic interference terms. When the total interference pattern is re illuminated with the reference wave (physically if converted to a transparency film or computationally), the latter is modulated and the effect is to produce two waves: a converging real-image and a diverging virtual-image. The simultaneous presence of this two components generates a distinct artefact denoted as *twin image effect*.

CDI Coherent Diffraction Imaging, provides a different method, entirely computational, to recover the missing information: *phase retrieval*. The phase problem in imaging experiments can therefore be expressed as the search for corresponding phases in a given diffraction pattern which fulfil additional constraints. The entire exit wave at the detector is then reconstructed by retrieving the phase which match the diffraction intensity. This constitutes *a reconstruction algorithm.*

Especially in its keyhole variant [50, 51, 52], transmission CDI is an exceptional technique in the soft-X-ray regime, providing both absorption and phase contrast at a lateral resolution that approaches the Abbe diffraction limit. The image formation model is built on the following assumptions [53, 38]: i) the illuminating light has a coherence width larger than the illuminated area and is single-tone; ii) the interaction X-ray/matter is described by the Born approximation (single scattering approximation); iii) the scattering wave at the exit surface is just the product of the illumination and the object transmission function; iv) paraxial approximation is valid; v) the scattering object has no depth; vi) the propagations distances (source-to-sample and sample-to-detector) are well defined and constant.

2.2.2 CDI forward model

Spatial information of the sample is acquired by exploiting the contrast mechanisms within the sample itself, where the main actors are the electronic densities of each atom. Fig. 2.10 shows the forward image model that describes how an image of the object is produced on a detector. A computational framework for CDI simply implements the models that describe each subsequent phenomenon along the photons' path. At soft x-ray energies, a photon interacts with an atom in one of two ways: it can be elastically scattered, or it can be absorbed [19]. As shown in chapter 1.7, these two complicated processes can be simply modelled and elegantly unified by using the complex transmittance function $O(x, y)$:

$$O(x, y) = \int_{t_0}^{t_1} e^{k[j\delta(x,y,z)-\beta(x,y,z)]} dz \tag{2.2.2}$$

where k is the wavenumber, t is the thickness of the sample, while $\delta(x, y, z)$ and $\beta(x, y, z)$ are respectively the phase-shifting and absorption part of the refraction index distribution $n(x, y, z)$:

$$n(x, y, z) = 1 - \delta(x, y, z) + j\beta(x, y, z) \tag{2.2.3}$$

The complex refractive index 2D distribution $n(x, y)$ and the thickness t describe how an incident field ψ_0 travelling in the z-direction is modulated by an object with such characteristics. The wavefield ψ_{esw} transmitted by the object (exit-wave) becomes:

$$\psi_{esw}(x, y, z_o) = \psi_0(x, y, z_o) \cdot O(x, y) = \psi_o(x, y, z_o) \cdot |O(x, y)| \cdot e^{(j \cdot \phi_{g(x,y)})} \tag{2.2.4}$$

where $O(x, y)$ is the object complex transmission function, expressed also in its magnitude and phase. This pure projection through the specimen's thickness (projection approximation) [18] works for the case where the sample is comprised within the depth of field limit. For thicker objects, one should use the multi-slice method [54, 18].

During the path from the sample to the detector, free-space propagation occurs. In a forward model, this is described by the action of an operator D_z which acts on the 2D field modulating the spatial frequencies of the exit-wave:

$$\psi_d(x, y, z_d) = D_{z_d - z_o}\{\psi_{esw}(x, y, z)\} = D_{z_d - z_o}\{P(x, y) \cdot O(x, y)\}. \tag{2.2.5}$$

As discussed earlier, the propagation operato $D_z\{\quad\}$ has to be chosen as a function of the distance z. For a (virtual/real) setup, given the aperture diameter A of the illumination and given z the observation distance, if the Fresnel Number ($F = \frac{A^2}{\lambda z}$ is greater than 1 then different models of a Fresnel propagation must be used, accordingly to the limits imposed by Nyquist sampling.

In the last equation (Eq. 2.2.5), the exit wave is expressed as the product of the illuminating function (the probe) $P(x, y)$ and the object transmission function $O(x, y)$. We can re-state the importance of modelling the object with a *complex* transmission function: the observed Fraunhofer diffraction pattern of a typical sample does not possess strict central symmetry, then according to Friedel's law [18] a phase-shifting component is present. This fact has implications for the reconstruction algorithm. The Fourier transform of a natural image made by only real components, possesses instead the aforementioned central simmetry.

Finally, as the last element of the forward model, a photon detector is then used to reveal the intensity (magnitude squared) of the incident radiation, producing a real, discrete, and quantised image $I(x, y)$, expressed by:

$$I(x, y) = \psi_d(x, y, z_d) \cdot \psi_d^*(x, y, z_d). \tag{2.2.6}$$

Discarding back-scattering effects, the front-scattered radiation distributes within the Fraunhofer approximation to spatial frequencies up to $u = 1/d$, where d is the minimum periodicity of the sample grating. This means that the setup should accommodate at the same time the desired resolution (limited by the maximum angle at which the detector can sense the scattered radiation) and the FOV. If a plane wave illumination is employed, the CDI "small FOV" problem is exacerbated by the need of an isolated object, because this information is exploited in the reconstruction process [55]. This latter requirement represents a challenging condition for an experiment [55].

2.3 Phase retrieval

Phase retrieval refers to a large family of algorithms which strive to reconstruct the phase of a complex field by matching the inferred data with the measured information. As no direct method currently exists, all the algorithms are iterative, adjusting the estimated solution by multiple iterations. The first example of such type of algorithm was the Error Reduction (ER), published in the famous paper [56]. Even if each algorithm propose its own solution, the general structure is fixed (graphic scheme shown in figure 2.11):

1. The algorithm is initialised with a complex 2D array describing the field at the detector and a real 2D array describing its "mask", the so called *support*. This mask is binary and defines the maximum lateral extent of the object. Beside its simplicity it is an important element in a phase retrieval algorithm. The magnitude of the complex field at the detector is obtained by taking the square root of the acquired diffraction pattern. A random or costant phase is assigned.

2. A backward propagation operator is employed to yield the field at the plane of the object, where the light/matter interaction took part. In the case of far field, an inverse Fourier transform suffices to describe the Fraunhofer diffraction pattern, otherwise the Frensel operator or the angular spectrum (described respectively in chapter 1.5 and 1.6) have to be exploited.

3. The imposed random phase at the detector will produce at the sample plane a wrong field, but it retains part of the solution due to the imposed magnitude. A *real space constrain* is then applied to the inferred complex matrix, by exploiting the support mask we gave as an essential ingredient of the process. The mask is used to selectively substitute the pixel values in the complex field. This is the step were most of the algorithm differs. In the ER algorithm, for pixel values which are

outside of the object extension mask, a 0 is filled (the "finite support"). Inside the region, pixel values are left untouched.

4. The exit wave at the sample plane (actually at its exit window) is propagated forward to the detector plane, following the inverse of the operator used at step (2)

5. Similarly to step 2, *reciprocal space constraints* are applied on this complex field. The most immediate is the substitution of the magnitude values from the measured information (the diffraction pattern). Other types of constrain can be imposed, e.g. due to a masking operation on the detector. At this point one iteration is concluded, and we can proceed propagating the exit wave back again and so forth.

6. The entire process is repeated until convergence is reached. To do so, one can observe the behaviour of an error metric (many more details in chapter 3.4) which provides an estimation of the reconstruction process. Typically, the L2 norm of the difference between the measured $I_m(x,y)$ and the simulated $I_s(x,y)$ diffraction pattern is taken into account, as defined in the following equations:

   $$E = |I_m(x,y) - I_s(x,y)|^2 \tag{2.3.1}$$

   $$E = |I_m(x,y) - \psi_d(x,y) \cdot \psi_d^*(x,y)|^2 \tag{2.3.2}$$

   where $\psi_d(x,y)$ is the field at the detector. Due to noise, this metric tends to converge to a small but non-zero value. One can than observe the behaviour of this metric and realise when this limit has been reached. Optimisation theory and common practises can be used for the stop condition.

The projection method

An iterative method of the type describe earlier is referred to as "projection method", due to the fact that we are projecting the current solution estimate between sets (the two constraints in real and reciprocal space), to make it eventually fall in the intersection. Many types of constraints can be imposed to facilitate the reconstruction (e.g. [57]). More observation on this nomenclature will be given in chapter 3.4.

Figure 2.12 shows the output of a simulated CDI experiment, in which the phase part of the diffraction pattern is retrieved. It is interesting to note in panel C how two solutions are concurrently present, and finally only one

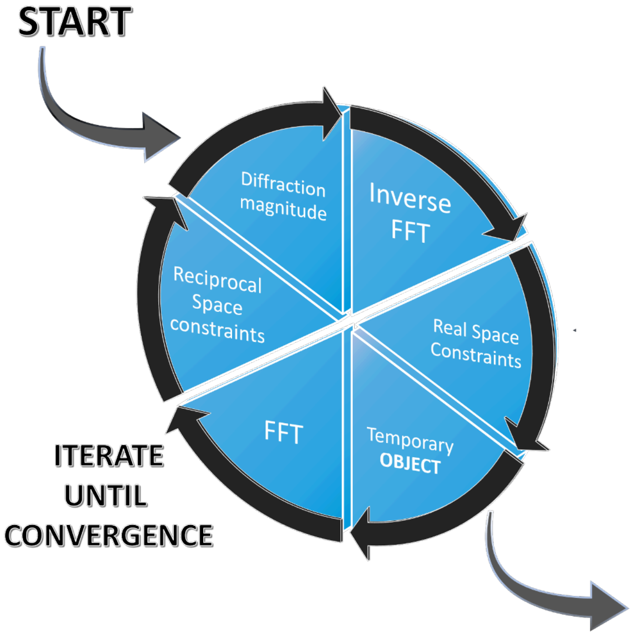


Figure 2.11: Phase retrieval scheme. Details in the text.

prevails. In a phase retrieval process we can observe that we can obtain a solution apart from three trivial ambiguities [58], which derive from the fact that the same far-field diffraction pattern intensity can be obtained in the presence of:

1. a global phase shift $X(\mathbf{r}) \rightarrow X(\mathbf{r})e^{j\phi_0}$
2. conjugate inversion $X(\mathbf{r}) \rightarrow X(-\mathbf{r})$
3. spatial shift $X(\mathbf{r}) \rightarrow X(\mathbf{r} + \mathbf{r_0})$

But in the end, why do we obtain the convergence to a unique solution? The recovery of a signal from its Fourier magnitude alone in general does not yield a unique solution [58]. However, for higher dimension, it has been demonstrated [59] that it is possible to achieve a unique solution, *if* an oversampling of at least a factor of 2 is present.

Many different variants of the ER reduction algorithm have been proposed. The most used is the Hybrid Input Output method (HIO) [61] which can be seen as a particular case of the Iterated Map (DM algorithm) [62]. In

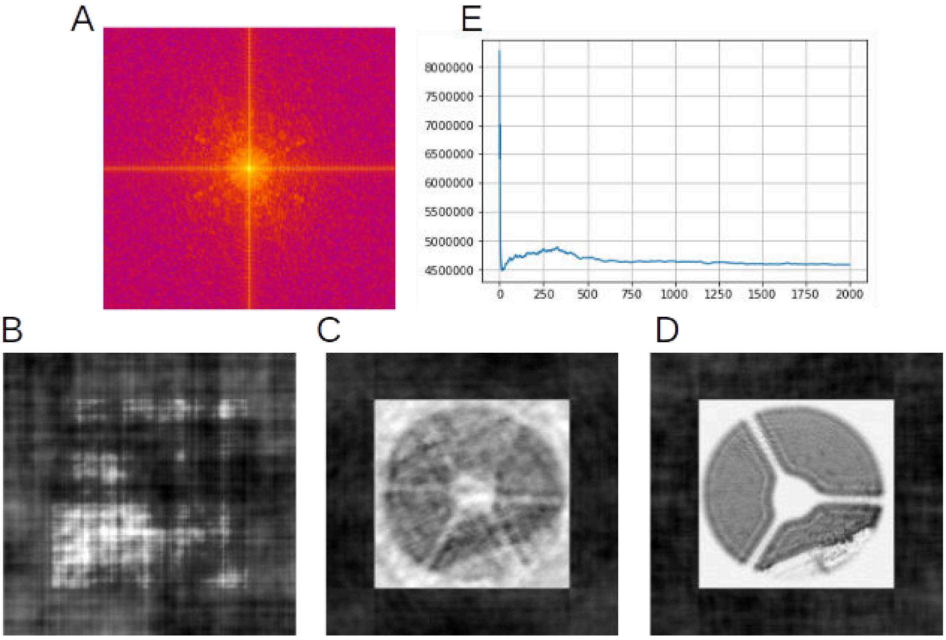


Figure 2.12: Phase retrieval simulation example: the complete Fourier Transform of a real image, whose magnitude is shown in panel A, is reconstructed by the HIO algorithm. Panels B to D shows the convergence of the solution. It is interesting to note how two possible solutions are fighting to prevail on each other: the real image is present as well as its enantiomorph. Images are cropped version of the actual computational box used, which is 4 times larger than the original image. The convergence is shown also by the behaviour of the reconstruction error, which is minimised iteration by iteration. The reconstruction image sample used in this simulation has been taken from data acquired at the Australian Synchrotron and elaborated at Elettra [60, 51].

the simulation of figure 2.12 the HIO method is used; the algorithm flow is similar to the ER algorithm, but the constraints in the object domain are imposed in a different manner given by the following equation:

$$\psi(\mathbf{r}) = \begin{cases} \psi'(\mathbf{r}), & \text{if } \mathbf{r} \in S \\ \psi(\mathbf{r}) - \beta\psi'(\mathbf{r}), & \text{if } \mathbf{r} \notin S \end{cases} \tag{2.3.3}$$

where ψ is the current field estimate), ψ' is the back-propagated field before the application of any real space constraint and S is the support mask.

The role of the support mask

The support mask is essential for a CDI reconstruction and it can be obtained from the measured diffraction pattern $I_m(x, y)$ by calculating the Patterson or autocorrelation map, given by:

$$A(\mathbf{r}) = \mathfrak{F}\{I_m(x, y)\} \tag{2.3.4}$$

which can be binarized by thresholding it at 95 % of its dynamic range to produce the support mask S [40]. The final mask has an extent that is approximately twice as large as the size of the support and this constitutes the first estimate of the support mask which can be iteratively refined (eroded) by employing the "shrink wrap" algorithm [63]. The support estimate emerges iteration by iteration from the updated estimate of the object itself. A final note on the support: a symmetric support mask S is less favourable due to the fact that it does not help in taking the choice between an object and its *enantiomorph*, and this is another explanation for the artefact in figure 2.12 panel C at roughly half of the reconstruction iterations. The support, in fact, was a simple square placed in the middle of the overlapped computational box. On the contrary, spatially separated support regions are very favourable.

The limits of CDI

Although the properties of CDI match ideally the experimental conditions of single-particle coherent diffractive imaging for free-electron lasers, they are not favourable for microscopy application at synchrotron radiation sources, where a large Field of View (FOV) is required. However many successful experiments have been proposed, especially in its keyhole variant,[64, 65, 66, 67, 68, 69, 70, 40, 71, 72, 50] which consists in using a spherical wavefront to produce a magnified image of the sample. A conic beam of X-ray invest the sample and as usual the output diffraction pattern is captured at the detector in far-field. The zone plate is "idealised" by the use of an order sorting aperture which blocks all the other focii and a zero-order stopper which attenuates the non curved radiation.

To accomplish such an experiment it is necessarily to completely characterise the complex illuminating field at the sample plane [64]. To so the so-called "white field" is acquired and the first step consists in recovering such an illuminating function. This process is carried on by an ER phase retrieval in which a priori knowledge on the Fresnel zone plate is used: this allows to disentangle the rapid oscillatory part of the spherical phase curvature, by propagating partly analytically and partly numerically the field from the pupil plane of the zone plate to the detector plane, passing from the focus plane. This steps help the convergence by introducing the a priori information on the first order focal length [65, 40]. As said, white field reconstruction is an additional step which makes longer and more tedious to achieve the final reconstruction.

Instead of exploiting a long projection based algorithm sucha as Error Reduction, both for the whitefield reconstruction and the object reconstruction one can instead use a Conjugated Gradient method, observing that the problem we are trying to solve can be formalised also as an optimisation process (more details in chapter 3.4). This new view of a projection based algorithm (that in the following will be referred to as "non direct" optimisation) allows to calculate the jacobian of the loss function that can be given to a common optimizer such as the "minimise" method in SciPy or Matlab. The result for the white field reconstruction is shown in figure 2.13.

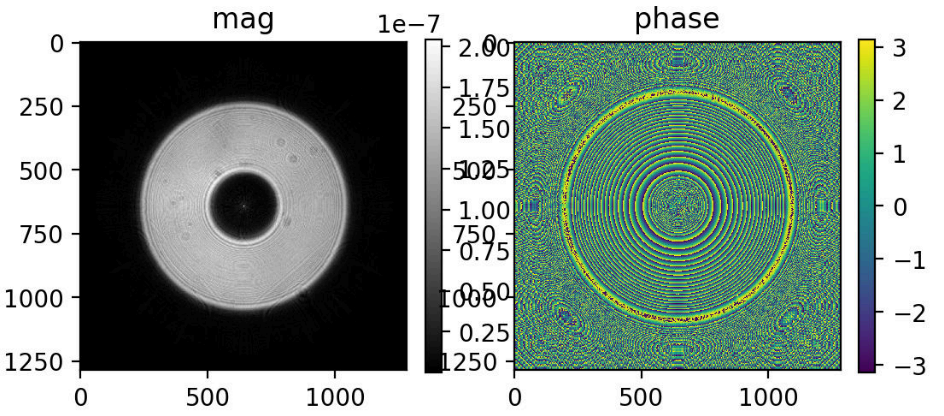


Figure 2.13: Conjugate gradient reconstruction of the white field at Twin Mic

When the pupil function of the lens is reconstructed and then also the complex field at the detector plane, the field at sample plane is readily found by numerical back-propagation and the reconstruction of the sample exit wave can be started. The maximum theoretical real space resolution is determined by the highest angle detected (Abbe theory) [50].

2.3.1 Partial coherence effects

Partial coherence effects can be considered negligible if the blurring that they cause does not extends over more than two pixels at the detector [40]. If the blur is present, solving the problem with a plain image-processing approach (*deconvolution*), to reconstruct such a dataset is extremely complex due to the difficult and time consuming pre-processing involved.

Figure 2.14 shows the effect of partial coherence in the form of an increasing number of modes: fringe visibility in a diffraction pattern (related to the coherence width) decreases as the number of modes increases [37]. The partial coherence model [73] effectively describes the effects of partial coherence as a convolution of the fully coherent diffraction pattern with a Gaussian PSF [37, 74] which is described by an angular semi-width Θ_{ill} given by the expression [37]:

$$\Theta_{ill} = \Theta_{speckle} \frac{\lambda M_s}{4d} \tag{2.3.5}$$

where Θ_{ill} is the angular semi width of the speckle in the diffraction pattern, d is the illumination size and M_s is the number of incoherent modes which describes the illumination [75].

Partial coherence action: The effect of partial coherence on the resolution can be explained by the fact that an analytic function is fully defined by the locations of its complex zeros and the diffraction pattern of a finite object is analytic [76]. The blurring effect of the convolution is to immediately remove any zeros from the measurements and so eliminate the possibility that any reconstruction can be completely consistent with the data [77, 76], due to the fact that the "finite object" assumption is destroyed by the aforementioned convolution [76].

The mutual coherence function [32, 78] can be approximated to be a simple Gaussian function with one adjustable parameter σ. The work [78] is one of the few publications which studies the effect of resolution decrease with partial coherence. Indeed, simulations are used to study the convergence of a ptychography reconstruction as the degree of coherence is reduced. The ratio σ/L is used to define the relative degree of coherence, where L is the width of the object. For each illumination, a real-space error metric is defined to provide absolute quantification [79].

Regarding the spectral purity, to express the resolution ϵ_l as a function of

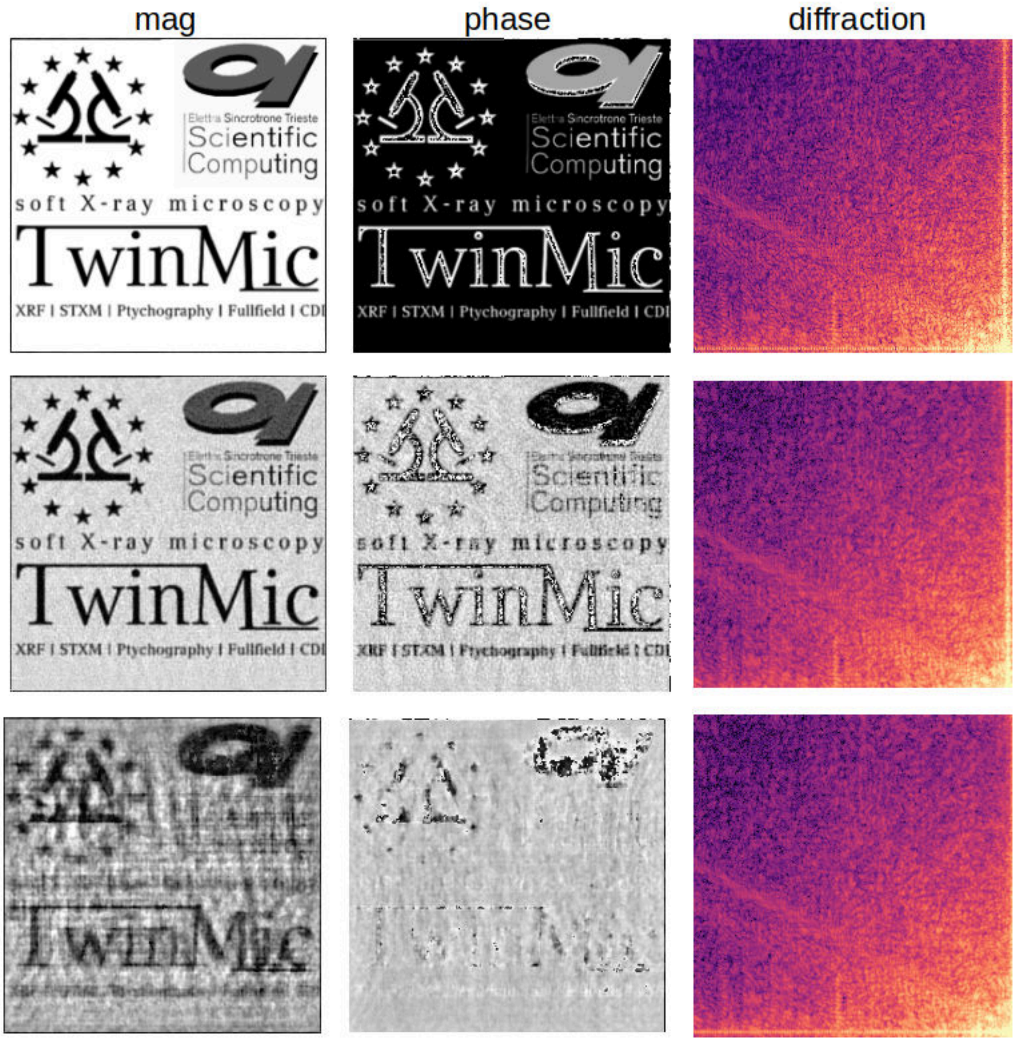


Figure 2.14: The effect of partial coherence on a CDI simulation for 1, 2, 4 modes.

the Coherence length $\Delta\lambda$ is straightforward:

$$\epsilon_l = \frac{\lambda^2}{\Delta\lambda} \tag{2.3.6}$$

where the fraction can be expressed as a function of the fractional bandwidth $\Delta\lambda/\lambda$. This means that to resolve an object of size 5 µm at a resolution of 10 nm, one has to keep the fractional bandwidth within 0.002. To do so a crystal monochromator is essential, as multilayer monochromators do not provide the required spectral purity [80].

2.4 Ptychograpy

Ptychography solves the FOV problem by simply introducing translation diversity into play; the grid scanning procedure is borrowed by the STXM technique, where a large illumination is used in place of a focused beam. The typical scheme of a ptychography setup employed at synchrotron radiation beamline is shown in figure 2.15. As can be seen in Fig. 2.16, the same area is illuminated multiple times, due to the overlapping between different probes (the illumination disk which define the illuminated area).

Ptychography is a hybrid scanning - CDI technique, in which a single extended object is illuminated with a cone of light and the resulting scattered wave is recorded by a detector placed transversely to the optical axis. Following figure 2.15, the sample is shifted while the source is fixed, and the intensity of the field that is scattered by the illuminated portion of the specimen is recorded at the detector. This configuration allows for a large field of view to be acquired, conversely to CDI. The peculiarity of this scanning setup consists in the high overlap factor between adjacent illuminated areas and this is where the magic lies. In a typical image processing super-resolution technique, the low frequency (low resolution) redundancy in the information acts as a cutoff for the multiple competing solutions in the parameter space, "denoising" the update step, while the noninteger shift between adjacent acquisition, associated with some measure of aliasing, permits the reconstruction of details at super-resolution [81]. A similar idea applies to ptychography, in which the noninteger shift appears in the reciprocal space, as we are sampling overlapping portions of the reciprocal space, each one defined by the Ewald sphere of each probe.

During the acquisition, each probe behaves as an independent CDI measure, following the CDI forward model described so far. To form the illuminating beam one can use a pinhole or the first diffraction order of a zone

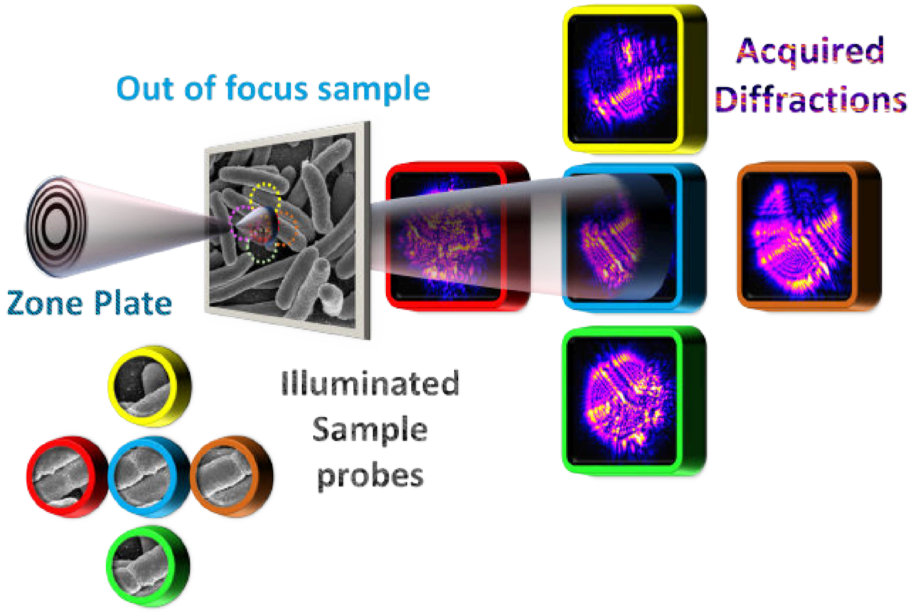


Figure 2.15: Example of a typical ptychography setup used in synchrotron laboratories: a pinhole (or a virtual source constituted by the focus of the zone plate) illuminates a well-defined region on the sample, which is mounted on a motorized stage. The scattered field intensity is recorded on a detector placed at some distance

plate. This will produce on the sample plane a (roughly) spherical wave $P(\mathbf{r})$ of the form:

$$P(\mathbf{r}) = P_{LP}(x,y)\frac{e^{jk\mathbf{r}}}{r} = |P_{LP}(x,y)|e^{j\theta_{P_{LP}}}\frac{e^{jk\mathbf{r}}}{r} \qquad (2.4.1)$$

where:

$$r = \sqrt{x^2+y^2+z^2} \qquad (2.4.2)$$

As usual, the z axis is located on the optical axis and all the setup components (sample plane, optics, detector) are placed perpendicularly to this axis. Fixing z, the spatio variant characteristics of the fields are expressed in (x, y). The spherical wave factor (the high-frequency chirp $e^{jk\mathbf{r}}$) is modulated by a low pass complex component, the pupil function $P_{LP}(x, y)$, incarnating all the low-frequency aberrations from the ideal spherical wave. The illuminated area of the sample can be varied by moving the sample downstream

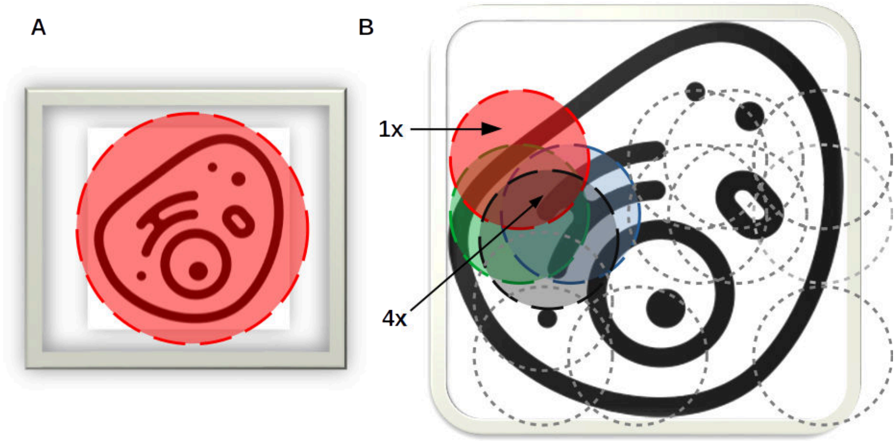


Figure 2.16: Panel A) CDI requires the illumination of an isolated specimen; panel B) ptychography scanning scheme applied to a sample: for each illuminated portion, a CDI experiment is performed. Each zone is not isolated, indeed an area overlap is present between adjacent probes. This allows the reconstruction process.

to the beam. The same type of illumination can be exploited also for the single-frame CDI.

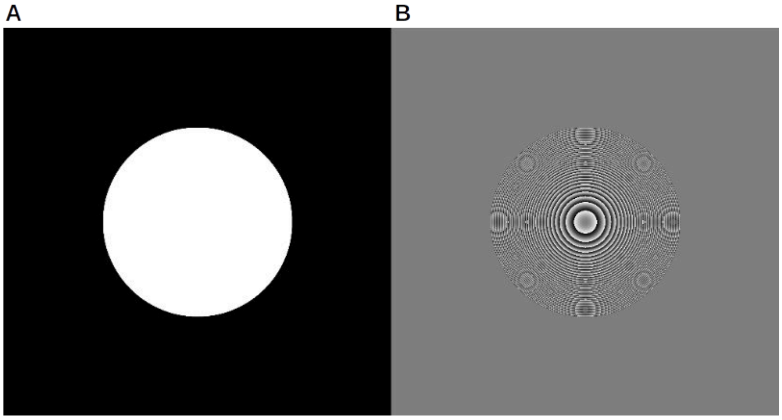


Figure 2.17: A simulated illumination function (probe) P at the sample plane: the magnitude (panel A) is a disk, while the phase shows the typical fringes of the spherical phase factor.

If $P(x, y)$ presents a phase component that is an expanding spherical chirp (spherical phase factor), such as in figure 2.17, the exit wave at far-field has a Fresnel rather than Frauhnofer form [70, 65, 40]; it can be seen as a magnified version of the Fresnel diffraction pattern using an elementary scale transformation [70]. A simulation of this effect is shown in Fig. 2.18, where an object function $O(x, y)$ made as *Lena* (see appendix A) is illuminated by a disk-like illumination of this kind (fig. 2.17. The far-field pattern presents the typical ringing of the Fresnel diffraction.

The scanning component of ptychography is formalised introducing the object ROI $o(x, y)$ (object view, calculation box), cropped from the full object $O(x, y)$. Then the exit-wave of the *jth* diffraction pattern with an ROI centred at the x_j, y_j coordinates is given by:

$$\psi_{esw_j}(x, y, z_o) = P(x, y, z_o) \cdot O(x, y, x_j, y_j) = P(x, y, z_o) \cdot o(x, y) \qquad (2.4.3)$$

where x, y are in the reference system of P while x_j,y_j are coordinates referred to the full object. Note that at the sample plane, which coincides with both the entrance window and exit window of the sample (thin sample approximation) the field ψ_{esw_j} is expressed at $z = z_o$. In the text, when not needed, the z will be dropped as being redundant. For example, an exit wave is defined only at the sample plane, then there is no need to use the z-coordinate. Differently for propagated quantities.

The result of a ptychography scan is a dataset composed of:

- multiple images (the diffraction patterns);
- a set of coordinates for the illuminated Region Of Interest (ROI) on the sample;

These data are contextualized (and made useful) by a set of setup parameters involving:

- the wavelength of the illuminating light;
- the object/detector distance;
- the detector pixel size;
- other parameter that describes the illumination onto the sample (e.g. the fresnel zone-plate diameter, the focal distance, the width of the smallest ring etc.)

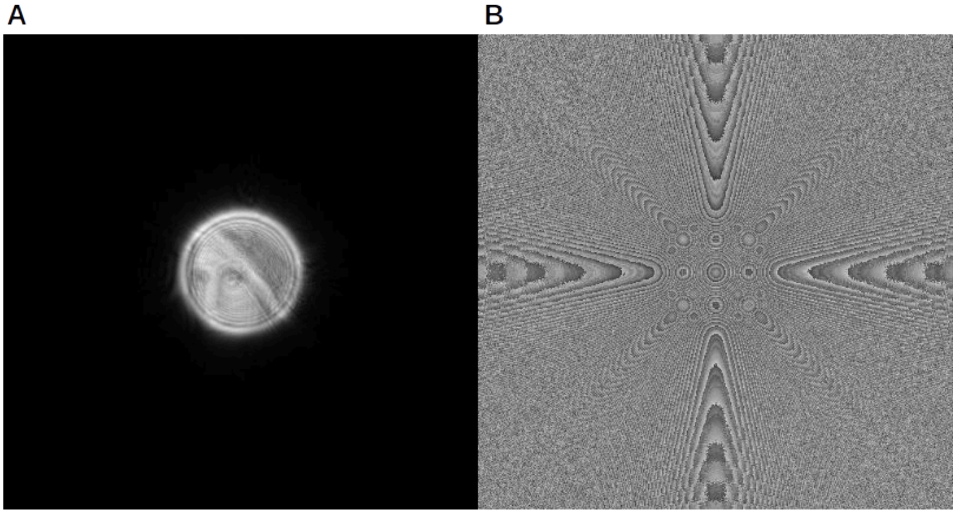


Figure 2.18: Simulated far-field diffraction pattern shown in magnitude (A) and phase (B) of a Lena object illuminated by a cone of light.

The effects of the overlap factor can be appreciated by considering the rows in figure 2.19 and 2.20: for the same degree of coherence (Fig. 2.19 row 1 and 2.20 row 1), the same area is in fact scanned by employing an overlap factor of 1/3 and 2/3. Clearly, considering Fig. 2.20, for both the degree of coherence the reconstruction converged to a better estimate of the object transmission function, shown in magnitude and phase. The last column of each figure, shows the ensemble of the recorded diffraction patterns. Obviously, for a larger overlap more acquisition are required to fill the same area.

Detector noise also affects the ideality of a coherent imaging experiment. Being real objects, detectors are affected by a wide variety of non-deterministic quantities and this reflects on each of the photon capturing, revealing, transforming, and counting mechanisms employed. The main visible effects of these non-idealities is a reduction of the contrast and equivalently in a reduction of the Signal-to-Noise-Ratio SNR in the measured diffraction pattern. The EMVA 1288 standard [82] and [83] provides an insight model for all the detector non-idealities, but for our scopes, it is sufficient considering the photon counting statistics given by Poisson distributions

$$p(n_q|P_r, O_r) = \frac{I_q^{n_q}}{n_q!} e^{-I_q} \tag{2.4.4}$$

where I_q is the average intensity for time interval at the qth pixel and n_q is

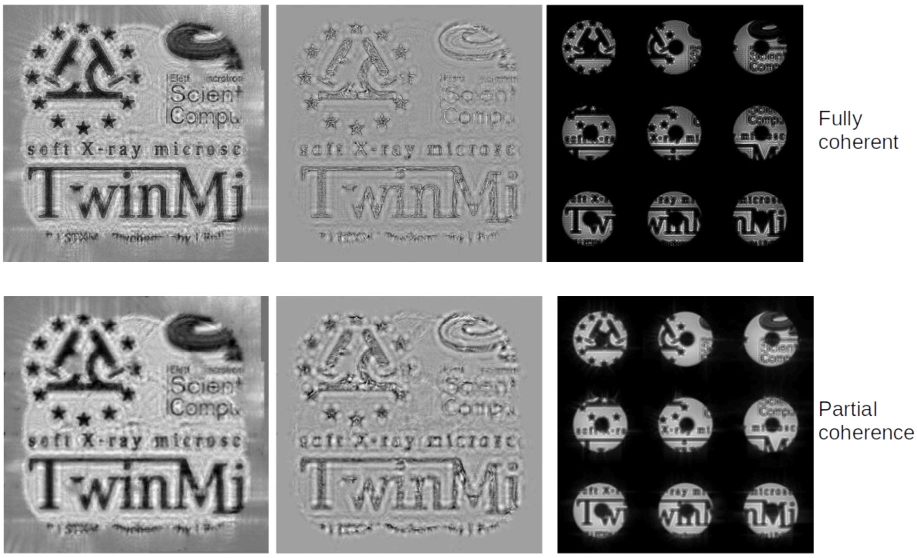


Figure 2.19: Ptychography reconstruction of a large step scan, where the overlap is roughly 1/3.

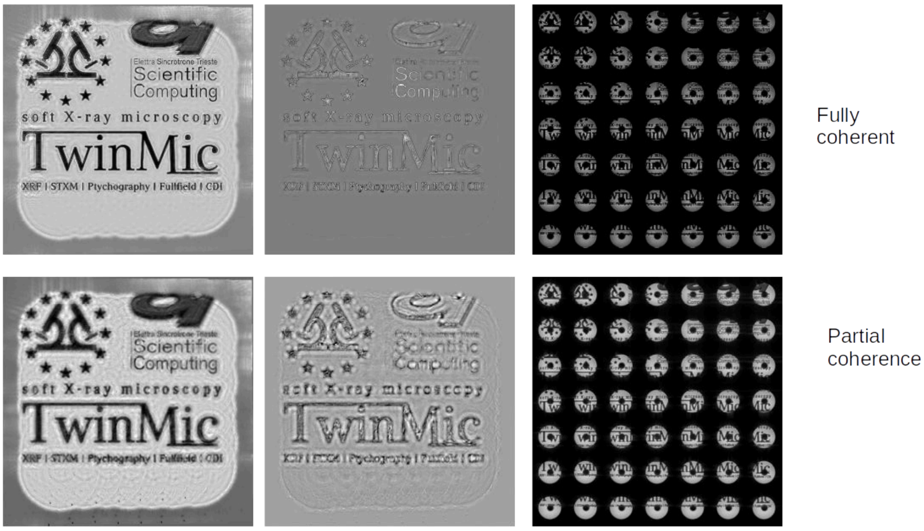


Figure 2.20: The same area of Fig. 2.19 is scanned with an overlap of 2/3, producing a way better result.

the measured value. For moderately large photons counts, the Poisson can be approximated by a Gaussian [84].

By using the relations exposed in this chapter, one can create a ptychography simulator like the one employed in the software framework developed in this research, "SciComPty"; as can be seen in Fig. 2.21 a Lena synthetic object is used in combination with a real whitefield complex image, provided by [85], allowing for a realistic simulation.

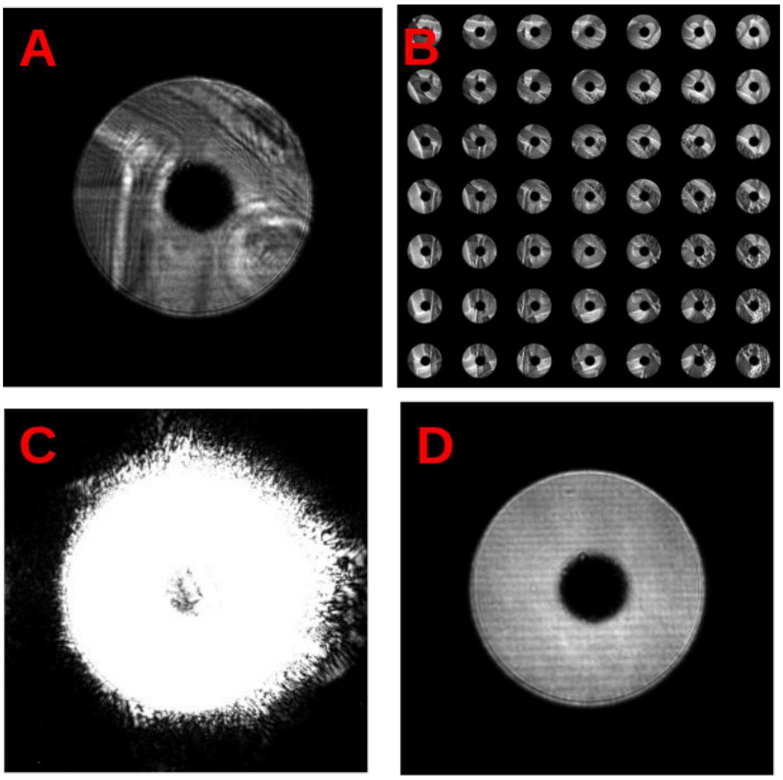


Figure 2.21: Simulated dataset which exploits the complex illumination took from real data [85]. Panel A shows a single diffraction pattern of the central part of the Lena object (A), panel B shows the ensemble of diffraction pattern composing the dataset. Panel C is an higlhy saturated version of a diffraction pattern, showing the high frequency scattering, while panel D shows the average of each acquired diffraction pattern, something that resembles "the white field".

2.5 Reconstruction model - Computational decoupling

During the reconstruction process, the forward model must be inverted to gather back the object transmission function. Since this inversion cannot, in general, be done analytically (even if a Wigner distribution deconvolution approach has been demonstrated as a proof of concept [86, 87, 88, 89]), an iterative procedure is employed, meaning that from initial estimates, a solution is found by altering iteration after iteration the current quantity, following an update rule. As discussed previously, CDI reconstruction is typically performed by using a multiple projections method that corrects the object estimate by sequentially applying a set of constraints that are coherent with the a priori beliefs and the measured information.

Projection method in ptychography The CDI object retrieval involves a) the reconstruction of the illuminating wave $P(x, y)$ at the sample; b) the definition of the object support (a mask) and c) finally the actual object reconstruction. On the other hand, refined versions of the ptychography reconstruction algorithms perform automatically the factorisation of the $P(x, y)$ and $O(x, y)$ functions and thanks to the overlapping condition do not require defining the object support. A computational box calculation step takes the current positions and crops the object accordingly, producing the current object view $o(x, y)$ from $O(x, y)$. In a typical imaging-oriented ptychography experiment like ours, $P(x, y)$ is just a byproduct, though mandatory for the O-P factorisation.

Reconstruction Constraints in ptychography Measured information (diffraction patterns) and overlapping (positions) represent the set of ptychographic constraints in the reciprocal and real spaces; each space is reached by propagating forward or backwards the exit wave. Projecting an estimate on the constraints set means to correct the estimate according to the constraints; this can be done by clearing, substituting, or modifying values. Commonly used ptychography algorithms are again projection-based (with a sparse family of optimisation-based techniques) but they differ in how they deal with each scanned position. The sequential algorithms (PIE family) [90] is shown in figure 2.22: it cycles on the set of positions and diffraction pattern projecting the solution estimate on the constraints set of each acquired image. The parallel ones (DM) [38] find a solution that satisfies concurrently each diffraction pattern. EPIE [91] and DM [38] have been presented extensively in the literature (e.g. [92, 91]), here only an introduction will be presented.

EPIE update equations

The Extended Ptychography Iterative Engine (EPIE) is an iterative algorithm which can retrieve automatically the probe function. It is realised by imposing the following update equations for the current object calculation box o_j (Eq. 2.5.1) and the current estimate of the probe function P (2.5.2). All the quantities are function of the plane coordinates (x, y) which are not shown for clarity. In addition, the z axis is also dropped as all the modifications to $o(\mathbf{r})$ and $P(\mathbf{r})$ are applied on the sample plane. Similar update functions are found in the literature related to Proximal algorithms (e.g. [93]).

$$o'_j = o_j + \alpha_o \frac{(\Psi'_j - P \cdot o_j) \cdot o^*_j}{max\{|P|^2\}} \tag{2.5.1}$$

$$P' = P + \alpha_P \frac{(\Psi'_j - P \cdot o_j) \cdot P^*}{max\{|o_j|^2\}} \tag{2.5.2}$$

Our estimates are updated based on how well we are be able to simulate the diffraction pattern at a given position. Then, the main job of the reconstruction algorithm is to reconstruct the wave at the detector, which provides in a deterministic manner the sample exit wave. We are then reconstructing the exit wave of the object, not the sample. As we will discuss in more detail in chapter 3.4, the main component of each iterative update is given by the error function (relative to the jth diffraction pattern):

$$E_j = \Psi'_j - P \cdot o_j \tag{2.5.3}$$

which intelligently avoids using the propagation operator within the forward model. Indeed, the error metric does not involve a direct confrontation with the measured diffraction pattern, but only between the simulated exit wave and the one corrected with the magnitude information in the measured diffraction pattern, Ψ_j in Eq. 2.5.1, 2.5.2 and 2.5.3. This allows for an easy to calculate gradient, which is:

$$\nabla E_j = -2P^*(\Psi'_j - P \cdot o_j). \tag{2.5.4}$$

The numerator As will be described in more detail in chapter 3.4, the gradient of the error metric is a fundamental component of any iterative algorithm, as one refines the current estimate in a direction which tends to minimise the error itself. The natural way to do this is to go in the opposite direction of the gradient itself. The numerator of Eq. 2.5.1, 2.5.2 is clearly designed on Eq. 2.5.4.

The denominator Anyway, to estimate the object, a reasonable and legit strategy consists in using the information in the exit wave, by dividing it by the current estimate of the illumination. While for high values of P we can effectively trust this update, as higher value of intensity can be produced, this update is poorly conditioned where the probe has a low intensity, due to the fact that we are effectively not able to recover much information; indeed a low amount of light passed trough the sample. As in this second case our estimate will be likely inaccurate, it is better to maintain the current value, or move very near to the current estimate. The combined action of the α_o and α_P coefficients with the denominator, produces a sort of update weighting which guide towards the correct solution faster.

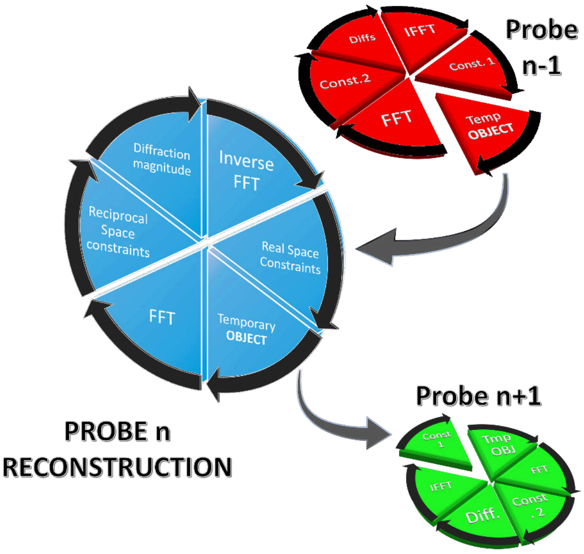


Figure 2.22: A sequential algorithm like EPIE: at each iteration a randomized sequence of probe indexes is generated. Within a for loop, each index is consumed by applying one projection based phase retrieval operation. Here the support mask is not explicit like in CDI but it is more subtle, due to the fact that convergence is guided by the overlap between different (maybe already solved) object parts.

RPIE and EPIE implementation

Figure 2.23 shows the reconstructed illumination obtained the ptychography software framework developed during this research; from the dataset of figure 2.21. The EPIE algorithm has been used, then differently from a ptychography P-FCDI approach [69], no illumination information is provided a priori. It can be seen that the high frequency detail of the illumination is completely resolved.

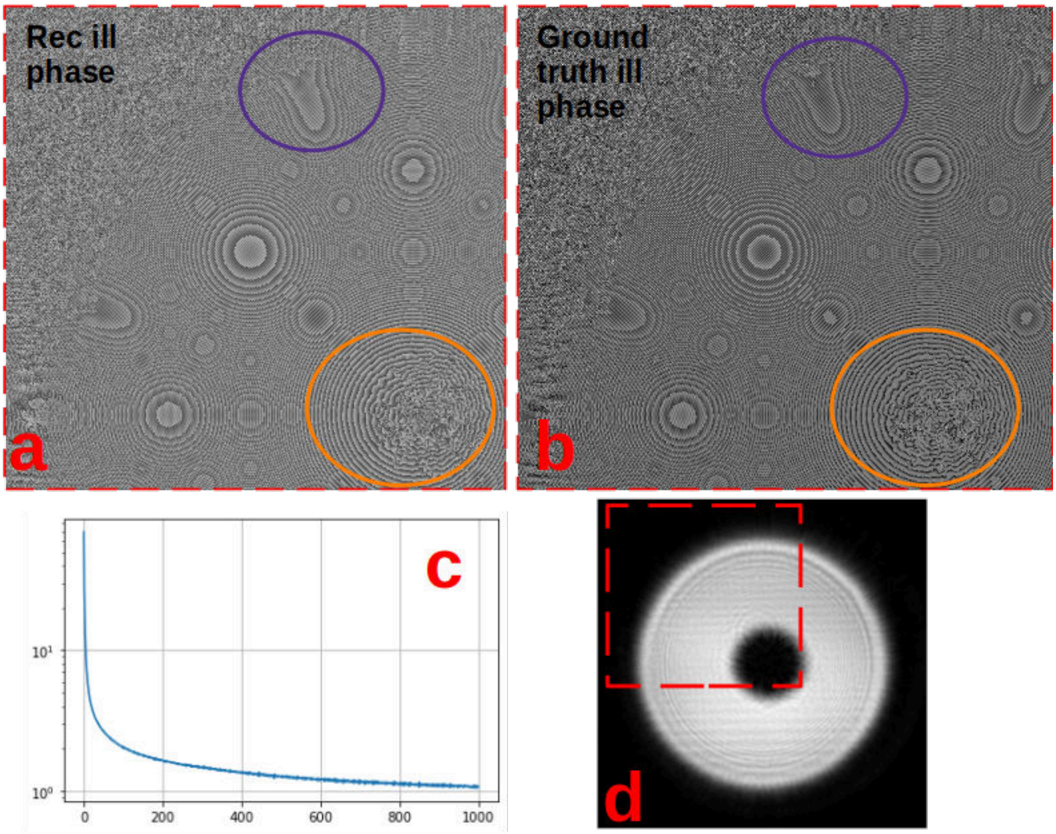


Figure 2.23: From the previous dataset shown in figure 2.21, a ptychography reconstruction using EPIE takes place. No a priori information is given to the algorithm, the illumination is recovered during the process. Panel A and B are magnified version of the phase map depicted in panel D (which shows instead the magnitude). It can be seen that the reconstruction of the illumination phase took place correctly, retrieving the high frequency details highlighted by the two coloured ellipses. Panel C shows the error metric, exposing the convergence of the algorithm.

The software framework developed during this research, SciComPty, provides not only these state-of-the-art algorithms but also a modified version of the newest RPIE variant [94], adapted for a state decomposition of the illumination reconstruction. RPIE update equations (Eq. 2.5.5 and 2.5.6) are remarkably similar to EPIE's but implement a different weight function (the denominator) in which a 2D regularisation term appears. This proved to increase the speed of the convergence [94].

$$o'_j = o_j + \alpha_o \frac{(\Psi'_j - P \cdot o_j) \cdot o^*_j}{(1-\gamma)|P|^2 + \gamma \cdot max\{|P|^2\}} \tag{2.5.5}$$

$$P' = P + \alpha_P \frac{(\Psi'_j - P \cdot o_j) \cdot P^*}{(1-\beta)|o_j|^2 + \beta \cdot max\{|o_j|^2\}} \tag{2.5.6}$$

The additional coefficients γ and β allow to tune between a regularised and a more EPIE-like update step. Those coefficients can be tweaked differently for any dataset, but in our experiments we have seen that a value or roughly 0.04 provides the best results.

The FOV effect

Even if not reported in the literature, from our experiments we noted [1] that the most visible effect is a larger computational FOV, meaning that from areas of a lower degree of overlap, the information extracted is higher. This can be beneficial to reduce the number of scanning probes and consequently, both the acquisition and reconstruction time are shortened.

To test the aforementioned assumption we carried out a simulation. Fig. 2.24 shows the phase of the retrieved object on a real dataset; the use of a test sample helps in the observation of the benefits of the proposed technique. Insets (a) and (b) are the results of respectively an M-EPIE and M-RPIE phase reconstruction of a siemens star, obtained after 150 iterations with $M = 3$ (more details later in 2.6); panel (c) shows the resulting density map [51] which can be obtained employing all the position in the dataset (panel d). The density map shows how a sample is illuminated during the procedure. It can be simply built by correctly stitching and fusing the retrieved probes.

It can be clearly seen observing the first row, that the reconstruction in panel (b) has a larger FOV and is of higher quality than the one in panel (a); in (b) the edge of the siemens star are clearly visible in the left part of the image, while in (a) the figure seems artificially cropped. Observing the second row of Fig. 2.24 the results of an artificially downsampled dataset are

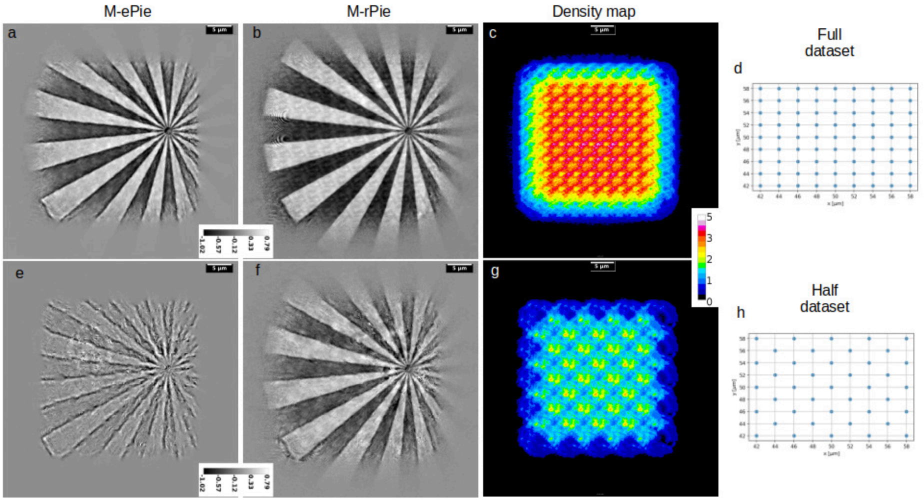


Figure 2.24: Phase reconstruction of siemens star, obtained using the full dataset (first row) or a sparser version of it (second row). Column one shows the results of the M-ePie algorithm, while the second one the reconstruction obtained using M-rPie. The canvas map (panel c and g) related to each dataset (d and h) are also shown.

shown: the dataset has been made more sparse than the first case, taking only the odd indexed data. The coverage of the object plane then become sparser. Panel (e) shows how the M-ePIE algorithm cannot cope with this sparsity, while the M-rPIE provides a reconstruction of better quality (panel f).

Sparser scanning A well-designed regularisation (in conjunction with the required padding on each diffraction pattern) is considered to be the main candidate for a successful data continuation, which is what cause the FOV to enlarge even in areas which are not highly illuminated (such as the borders). This effects automatically translates in a better reconstruction (Fig. 2.24, panel e and f) in the case of sparser scanning (Fig. 2.24, panel g), as the higher resolving power can effectively cope with the missing information between different subsequent measures. Sparse scanning means reduced acquisition and reconstruction times, as well as lower radiation dose, and that is why the implementation of the rPIE algorithm in both its single and multi-probe version is essential for a modern beamline. The fact that these algorithms are implemented on GPU, in a simple way and using open-source tools is for sure a strong plus.

Simulating the sparser scan by dropping half of the dataset describes reasonably well a larger step, even if eventually a different movement-induced statistics can be triggered on the position error.

2.5.1 Multi mode modifications

The redundancy in a ptychography experiment is the key for a super-resolution reconstruction of the entire object, even if for each single CDI projection the sub-wavelength information is washed out by the low-pass filtering action of the free-space propagation (sub-wavelenght information is "non propagated" by evanescent waves which are attenuated exponentially [31]). As we will see in chapter 3.4, the aforementioned redundancy can also be exploited for the correction of some other image formation model parameters; these can include not only setup parameters such as those discussed in this work, but also different co-existing partial coherence states both for the illumination (probe modes) and the objects (object modes) [95]. The ptychography model has to be modified accordingly exposing the parameters to control.

Multi-mode Forward model In the case of many illumination modes, the forward model change is straightforward: given a predefined number of M probe modes, each partially coherent illumination mode is forced to interact under the projection approximation and propagated to the detector. At this point, each complex field is summed in a coherent (field sum) [96] or non-coherent (intensity sum) [43, 73, 97, 98, 44]to estimate the acquired diffraction pattern (respectively eq. 2.5.7 and 2.5.8)

$$\widetilde{I_j} = |\sum_{p=1}^{M} D_z\{P_p(x,y) \cdot O(x,y,x_j,y_j)\}|^2 \tag{2.5.7}$$

$$\widetilde{I_j} = \sum_{p=1}^{M} |D_z\{P_p(x,y) \cdot O(x,y,x_j,y_j)\}|^2 \tag{2.5.8}$$

where x_j, y_j are the shift component of each view, $\widetilde{I_j}$ is the j-th diffraction pattern recorded onto the detector, $P_p(x,y)$ is the illumination probe mode at the sample and finally $O(x,y)$ represents the transmission function of the object shifted by x_j,y_j. D_z as usual represents the propagation operator.

The extension for the N object modes follows the same reasoning:

$$\widetilde{I}_j = |\sum_{p=1}^{M}\sum_{o=1}^{N} D_z\{P_p(x,y)\cdot O_o(x,y,x_j,y_j)\}|^2 \tag{2.5.9}$$

$$\widetilde{I}_j = \sum_{p=1}^{M}\sum_{o=1}^{N} |D_z\{P_p(x,y)\cdot O_o(x,y,x_j,y_j)\}|^2 \tag{2.5.10}$$

and allows taking into account vibrations of the specimen at the sample plane [96].

Multi mode RPIE In general, to cope with partial coherence, the computational problem is relaxed, by solving not only for $O(x,y)$ but also for a set of N mutually incoherent probe functions $P_n(x,y)$. Given the aforementioned feature of RPIE (larger FOV), it can be a great advance to adapt RPIE to such partial coherence situation. The adaptation of the RPIE algorithm to a multi-mode problem (M-RPIE) is summarized by the following new update functions that link the current estimates of the illumination P_n and the object view o to the new estimates $P_n^{'}$ and $o^{'}$:

$$P_n' = P_n + \alpha_P \frac{(\psi_n' - P_n o)\cdot o^*}{\beta\,|o|_{\max}^2 + (1-\beta)\,|o|^2} \tag{2.5.11}$$

$$o' = o + \alpha_o \frac{\sum_{n=1}^{N}(\psi_n' - P_n o)\cdot P_n^*}{\gamma(\sum_{n=1}^{N}|P_n|^2)_{\max} + (1-\gamma)\sum_{n=1}^{N}|P_n|^2} \tag{2.5.12}$$

where:

$$\psi_n' = D_{-z}\{\sqrt{I_j}\frac{\psi_n}{\sum_{n=1}^{N}|\psi_n|^2}\} \tag{2.5.13}$$

is the magnitude corrected scattering wave produced by the nth illumination mode. The x, y dependencies in $P(x,y)$ and $o(x,y)$ have been neglected to simplify the expressions. α_p and α_o represent the update rates of the probe and object estimates and are typically set at 1. The beta and gamma parameters are instead trickier to tweak and define how much regularisation is used in the reconstruction. The role of each denominator is to weight the update where the exit wave is not bright [94], this because for a darker area the error in the *o*-*P* factorisation can be high. The expressions for alpha and beta equal to 1 make the algorithm to simplify in the common M-EPIE, where the denominator is just a positive real number, the maximum in the squared magnitude of $P(x, y)$ and $O(x, y)$. RPIE slows the update increasing the denominator by introducing a loss cost given by the per-pixel magnitude of the o or Pn function. Unfortunately increasing to N the number of modes increases the iteration time, by the same factor, due to the required N forward and backward propagations per view.

2.6 Reconstruction examples

SciComPty provides flexibility in defining the quantities for the simulation parameters, propagation methods and the object/illumination. Simulating a dataset involves defining the energy of the incident radiation, a scanning overlap factor, and finally to define the resolution at the detector and at the sample plane. This would set the coupled sampling operation. To simplify the procedure, one can also define the resolution at the sample or at the detector plane and accept the one provided by optimum sampling, depending on the setup and the propagation method used. In the case of a far-field setup [36], the pixel size δ_s at the sample plane is given by:

$$\delta_s = \frac{\lambda \cdot z_{sd}}{M \cdot \delta_d} = \frac{\lambda \cdot z_{sd}}{S_d} \tag{2.6.1}$$

where M is the detector size in pixels, δ_d the detector pixel size, S_d the detector lateral dimension and z_{sd} the sample-detector distance. In the case of a near field setup with curved wavefront [99, 100] the sample resolution is related to the geometrical magnification factor M_F

$$M_F = \frac{f_d}{f_s} \tag{2.6.2}$$

$$\delta_s = \frac{\delta_d}{M_F} \tag{2.6.3}$$

The sample plane pixel size defines the basis for the scanning movements, that in the simplest case, follows a grid pattern. Commonly, random jitter is added in x and y, to prevent the raster scanning pathology [101, 94, 102]. The overlap factor defines instead the maximum step movement. The object function $O(x, y)$ and the illumination function $P(x, y)$ are assembled in magnitude and phase providing two images each. With four completely different images, this kind of simulation tends to be even harder than a real sample, where the magnitude and phase part of a real object tends to be somehow similar, even if, as said in the introduction, they derive from different physical mechanisms.

During the reconstruction process, the pixel size at the sample plane can be imposed using exactly the aforementioned equations. To reduce the reconstruction time (especially during parameters tuning) it can be beneficial to rebin the acquired diffraction pattern. The detector pixel size of the new images varies linearly with the rebin factor (that practically produces bigger pixels) and this new value has to be used to define the sample pixel size. A zero-padding operation has to be used where the diffraction pattern bright region expands to more than half of the lateral size of the detector; in this case, no pixel size adaptation is required, as the net effect is to have a virtual larger detector S. This new padded dimension will be used during the reconstruction as the computational box size.

In a defocused probe setup, the relative positions of the zone plate, sample and detector are required to establish not only the propagation distance and the pixel size but also the initialisation for the probe function on the sample plane: the high-frequency chirp function of an expanding spherical wave cannot be retrieved during the reconstruction [67, 103], then an explicit form must be included in the initialisation of $P(x, y)$ and this will serve as a seed for the reconstruction. In the far-field case, the Fraunhofer approximation can be simplified even more by using just a 2D DFT to move back and forth.

Figure 2.26, 2.25 and 2.28 shows the results of real dataset reconstruction made with SciComPty: in figure 2.26 (panel A and B) a siemens star is reconstructed with a pixel size of 18 nm. As can be seen we reached the lithography resolution of the specimen, and also imaged a thick layer of dirt near the center. A distinct feature can be observed at a resolution of roughly 50 nm. The observable ptychographic resolution gain [104] G_p can be calculated for this reconstruction by the following expression:

$$G_p = \frac{d}{\delta_s} = \frac{9\mu m}{50nm} = 180 \tag{2.6.4}$$

In Figure 2.26 panel C the phase map of an human lung tissue is shown. The thick diagonal bar is an asbestos fibre included in the sample. In the top left corner of panel C, the reconstructed illumination magnitude is shown.

The use of the modal decomposition (Eq. 2.5.10), also for the object is extremely important in the case of sample vibrations and movements: as an example, the reconstruction obtained in panel A and B of Fig. 2.26 is reasonably good even in presence of a drift of unknown cause during the

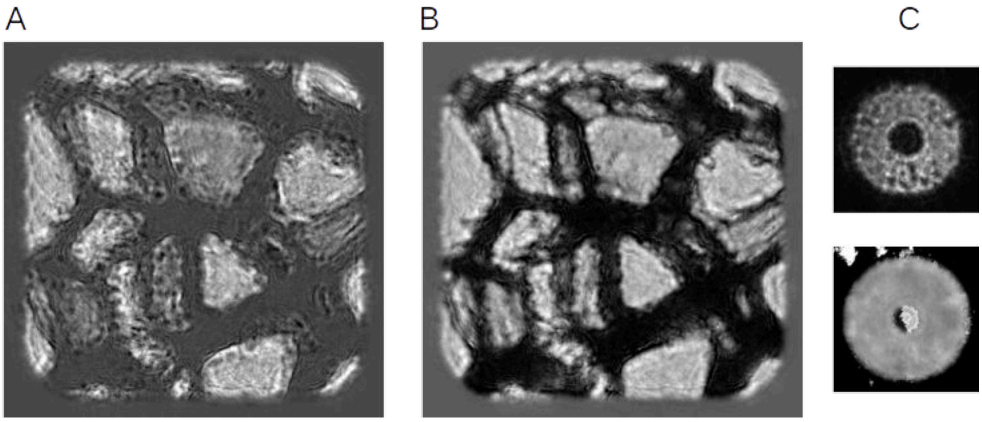


Figure 2.25: EPIE Reconstruction of a stem section of a Solanum lycopersicum (tomato), the sample used in [45, 105], shown in its magnitude (A) and phase (B). Panel C shows the reconstructed probe.

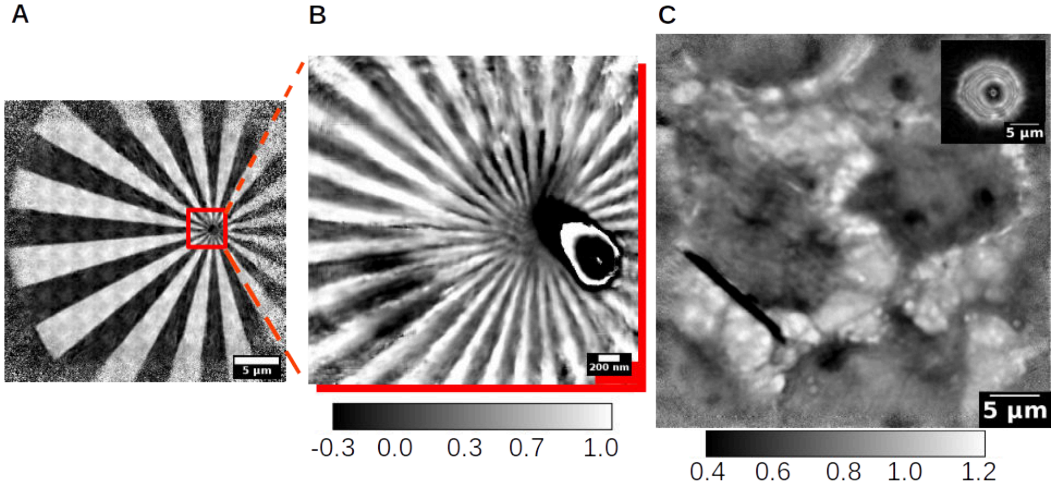


Figure 2.26: Phase reconstruction of a siemens star (panel A and B) and of a human lung tissue contaminated with asbestos (panel C).

experiment. It was not completely clear if it was a thermal drift on the sample or a sample stage error, as a reset of the control system changed the drifting behaviour. Fig. 2.27 shows 4 STXMs took in 20 minutes each one showing a different position of the sample. The scale bar in panel (b) represent 20 um on the image, then the total drift in the shown time window was of about 5 um.

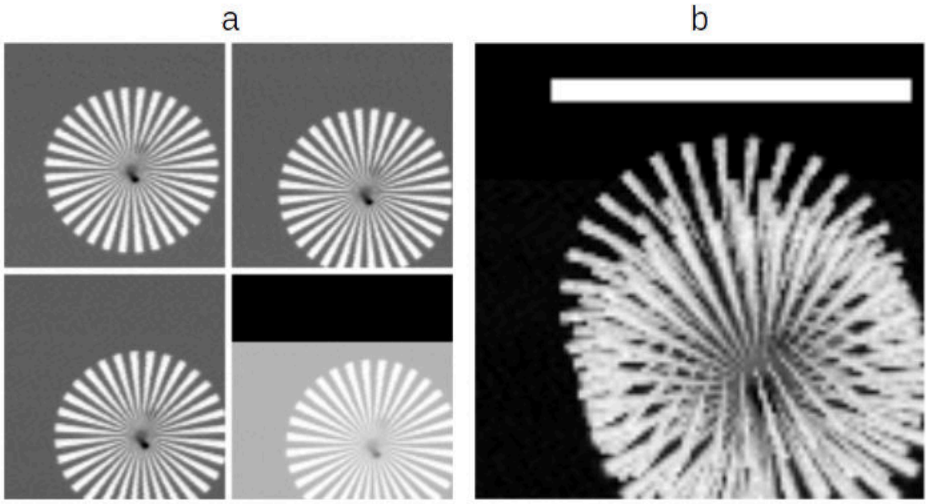


Figure 2.27: The reconstruction in panel A and B of Fig. 2.26 was obtained even in presence of an erratic drift shown by the four STXMs (spot size 250 nm) presented in panel (a), which are overimposed on the same figure in panel (b). The observation time is 20 minutes and the scale bar is of 20 um.

Figure 2.25 shows instead the reconstruction of a stem section of a Solanum lycopersicum (tomato),a specimens representatives of a large sample class used in a wide range of scientific studies.

Figure 2.28 shows different versions of the reconstruction of a different zone of the human lung tissue sample: MET cells were grown in Si3N4 windows and were exposed to asbestos fibres [106, 107]. In this set of reconstructions, a far-field setup is used. Each panel shows the phase of O(x,y) obtained by employing three different algorithms.

The first observable effect is the large FOV provided by the RPIE algorithm. The bottom left fibre in the middle image is way more visible, as well as other cell organelles on the top right side. However the EPIE method provides a higher contrast due to its higher aggressiveness in the iteration update step (as described in the previous section). The leftmost panel shows the output of the proposed M-RPIE method in which 3 illumination modes are employed. This value has been chosen to reduce the computational time, that as has been said before scales linearly with the number of modes. Even if the larger FOV effect is attenuated a bit (presumably due to the necessary tweak in the α_O parameter in eq. 2.5.12, the ringing artefacts is reduced, thanks to the illumination decomposition. The three reconstructed modes are shown in figure 2.29 panel B, against the single mode RPIE $P(x, y)$.

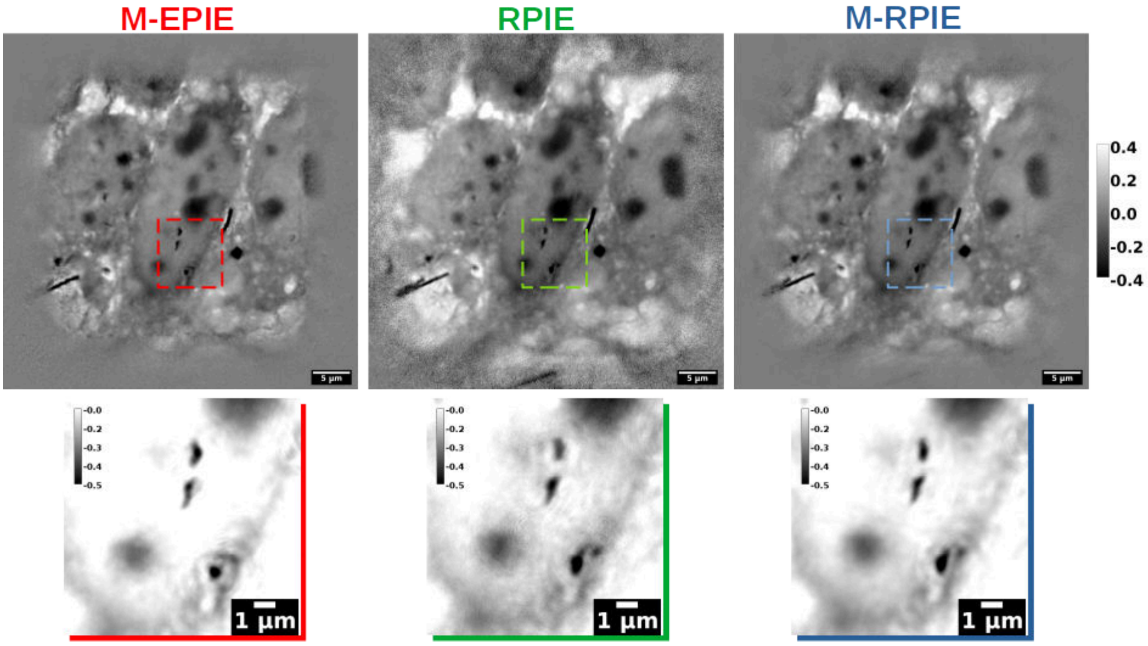


Figure 2.28: Reconstruction of a different area of the same "human lung tissue" sample using multimode-EPIE (left), RPIE (center) and the proposed multimode-RPIE (right): the computational FOV of the single mode RPIE is larger, meaning that more information is extracted in the areas of lower overlap, but speckle noise is more evident. Introducing modal decomposition reduces the noise and provides less ringing artefacts. Part of the larger RPIE FOV is retained also in this reconstruction.

2.7 The software

All the simulations and reconstructions are carried out by using a new software framework, named SciComPty [1], developed during this research and entirely written using PyTorch [108], an open-source deep-learning tool, that here is used as an easy route to fast and efficient GPU programming. The software framework is written as a set of functions and template scripts that can be used to run simulations and reconstruction of synthetic and real datasets. The presented software implements many state-of-the-art reconstruction methods like EPIE [91] and DM [38] and also provides an implementation of the new RPIE [94] algorithm, filling a hole in the current software landscape. SciComPty is not designed as a control-system orches-

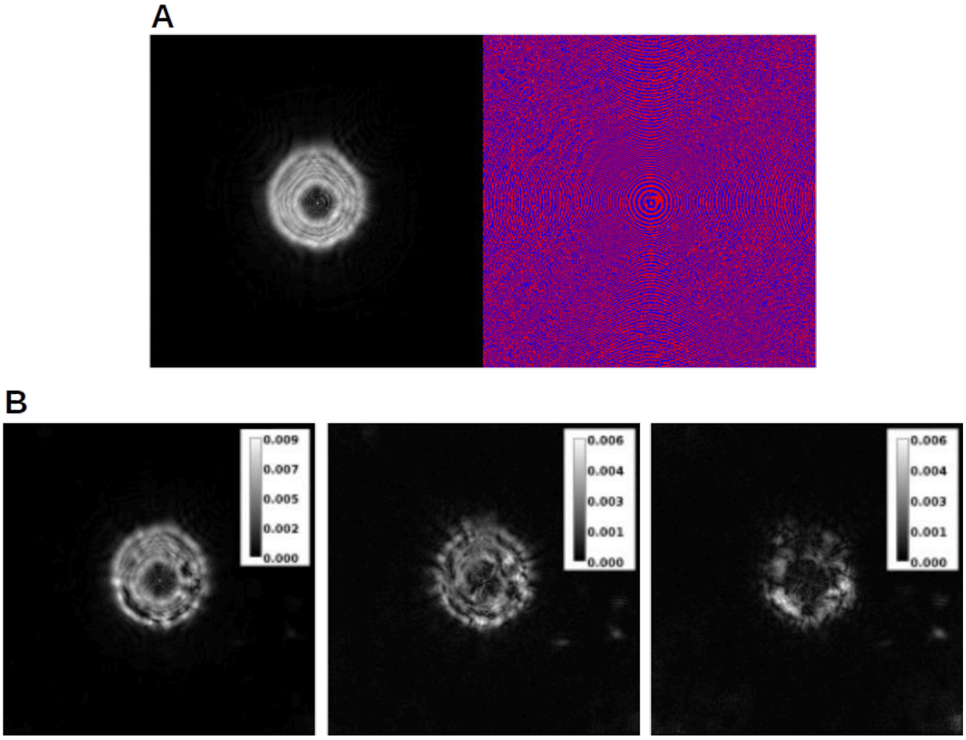


Figure 2.29: The illumination P(x,y) reconstructed at the sample plane in the case of single mode (represented in magnitude and phase) RPIE (panel A) and using M-RPIE with 3 modes (magnitude only).

trator but can only run simulations and offline reconstructions. A screenshot of the reconstruction panel can be seen in figure 2.30.

Fig. 2.31 panel A shows the phase map of another group of chemically fixed mesothelial cells grown in silicon nitride windows. The reconstruction is a remarkable example of the large space-bandwidth product which can be obtained through ptychography, indeed, a large field of view of more than 60 µm is covered with a high resolution (pixel size of 18.32 nm), as can be appreciated by observing panel B, where a scale of 500 nm is shown. To reconstruct this data, four different datasets of adjacent areas were collected in four different scans, for a total of 4x11x11 diffraction patterns; the SciComPty parallel-GPU capable software is used to reconstruct at the same time the four different parts. Then, the partial results are intelligently stitched together by using the illumination density map (Fig. 2.31 panel C) as a weighting factor. Despite its simplicity, it provides no sign of stitching artefacts. The same parallel algorithm can be employed also for a smaller dataset: as shown in Fig. 2.32 (a 9x9 dataset is shown), for a four-GPU

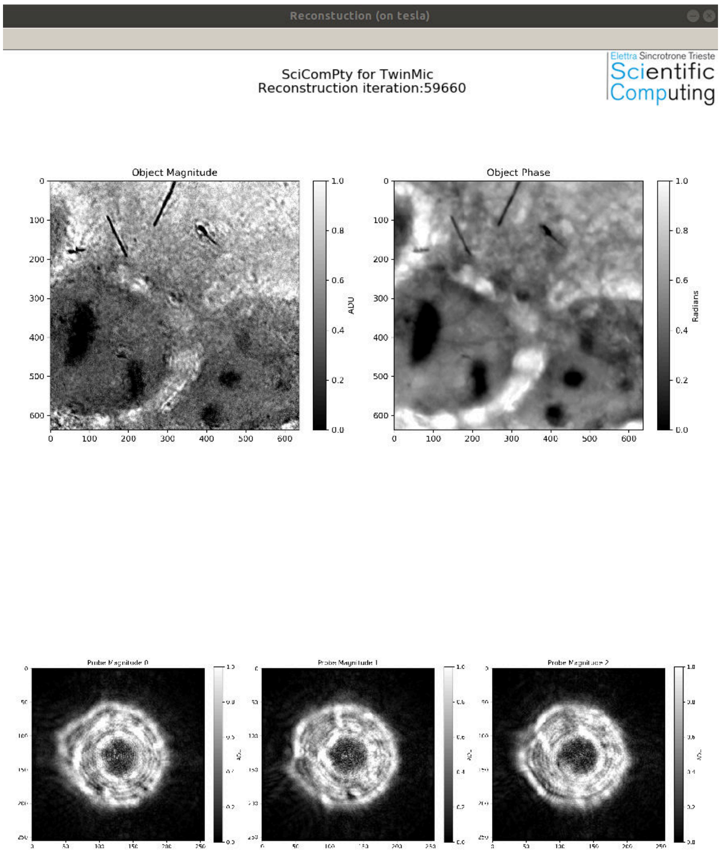


Figure 2.30: SciComPty main window, during a reconstruction.

server, the work is scheduled by dividing the dataset into four sub-regions, with no overlap. Thanks to the revised RPIE algorithm with large FOV, poorly illuminated areas (the borders of each sub-reconstruction) can be effectively merged with no sign of under-sampling.

2.8 Summary

In this chapter, to appreciate the benefits of the coherent diffraction imaging methods, a brief description of the basic microscopy techniques has been given. Phase imaging is time-consuming (STXM), or it requires additional hardware (gratings) or it is inadequate (holography). CDI has been later introduced, exposing the difficulties that the experimenter has to overcome for a successful reconstruction, especially in the form of parameter tuning, support mask creation and white field reconstruction. Ptychography pro-

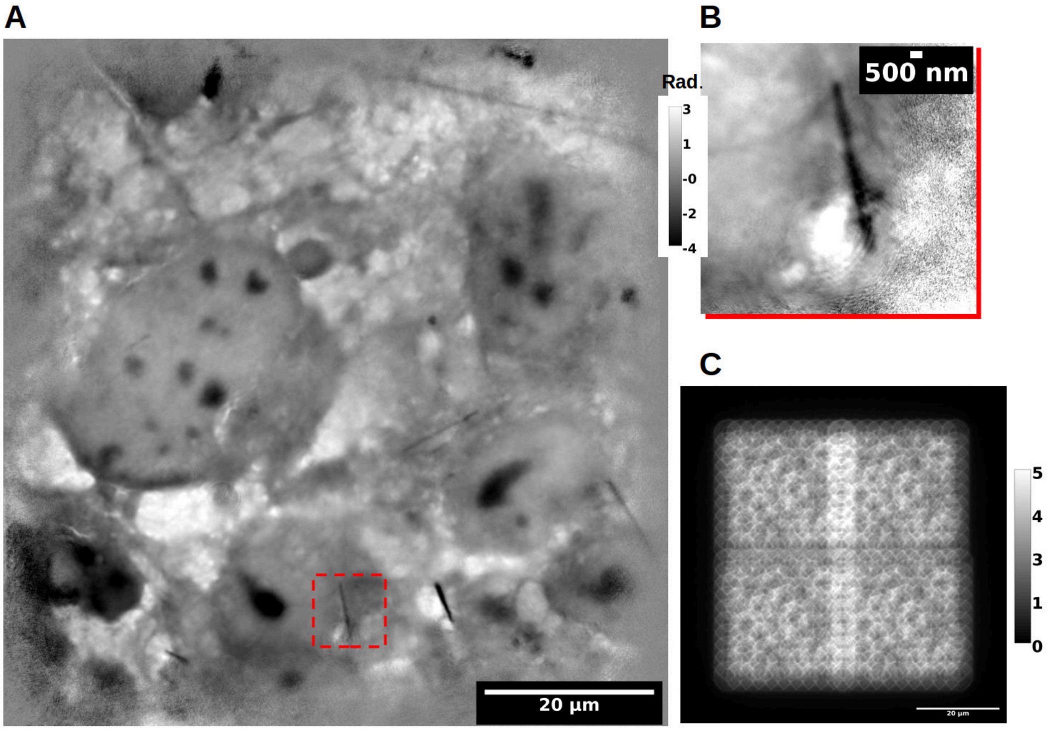


Figure 2.31: Panel A displays the phase reconstruction of a 22 Mpixel reconstruction, obtained by merging 4 different datasets. The red square is shown magnified in panel B. The smart stitching algorithm is based on the use of a density map.

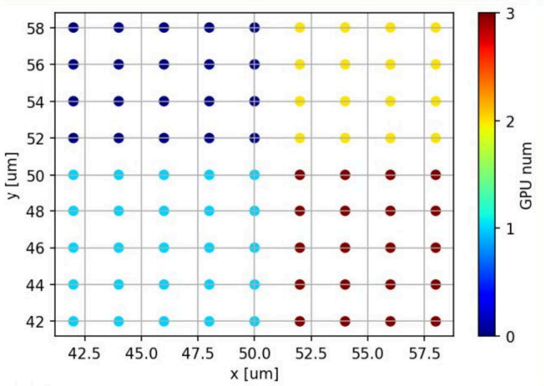


Figure 2.32: Division of the workload among 4 GPUs. No overlap is needed as the M- rPie algorithm provides a good reconstruction also at the borders of each sub-region.

vides an easier access to the coherent diffraction imaging, by employing a streamlined approach. This result is achieved by exploiting the diversity in the dataset. The standard technique has been presented by exposing the forward model of the SciComPty framework. Finally, a novel method based on the RPIE algorithm has been described and tested against other ptychography algorithms. The main advances of the proposed method are related to a larger computational view which allows to implement sparser scanning. This latter result is remarkable in reducing the time for an acquisition and the reconstruction. An additional advantage of the proposed method regards the extension of the RPIE algorithm for the case of multi-mode illumination. These two aspects are crucial for a working ptychography beamline. In the end, many reconstructions of different specimens are provided, showing the capability of the software.

Chapter 3

Solutions for setup errors

In the previous chapter the ptychography technique has been presented through the help of the software framework developed in this research, highlighting the design of a new algorithm to improve a reconstruction in a partial coherence environment. In this chapter other three contributions are presented, as solutions to cope with other setup parameters uncertainties. The first one, proposed in [3], exploits a metric approach to solve for position errors and is suitable for single-probe reconstruction; the second solutions is again designed to cope with position refinement and uses an iterative approach which can be embedded in any sequential type reconstruction algorithm [1]. To effectively use this method, it is necessarily to exploit a GPU compatible registration routine, that we propose [1]; the third one [4] consists of an innovative method to infer the propagation distance, the position refinement coefficients and the mutually incoherent probes, during the reconstruction itself and is formulated as an optimisation problem. The chapter starts with a general view on setup errors in ptychography.

3.1 Setup errors

As in each computational imaging, the knowledge of the a priori model is essential for a good reconstruction. Reconstruction artifacts arise when the model does not describe well the experiment. An iterative reconstruction algorithm will stride to minimize some kind of error function (described in the third subsection of this chapter) but typically stagnates after few iterations due to this kind of setup errors. We can refer to these errors as uncertainties on quantities such as photon energy, scan positions and propagation distance.

Browsing through the ptychography literature, it appears that the majority of the attention in automatic parameters' correction is devoted to the position refinement (next paragraph). As said previously, other efforts have been dedicated to the study of partial light coherence, coloured illumination, or multiple objects mode, while differently from other works, in [109] an algorithm for the determination of experimental parameters is described: apart from the position refinement procedure, the paper describes also the use of an evolutionary algorithm to cope with the uncertainty on the source-to-sample distance, that translates in a spherical aberration of the probe; in electron ptychography, indeed, conventional probe-solving algorithms like EPIE and DM would fail, due to the even higher sensitivity to position errors. Then, the reconstruction is carried out just using the conventional PIE algorithm but coupled with a different probe-solving strategy (an evolutionary algorithm infers the curvature of the illuminating probe).

State of the art on position refinement In many ptychography setups, the positions list of the scan is typically directly taken from the commands that the control system sends to the sample stage; the final movement is composed using a servo for the coarse steps and a piezo for the fine shifts. Even a minor misalignment (wrong positions) within an order of magnitude of the desired resolution (i.e. non-linear thermal-drift induced deformations or motor backlash) can lead to severe degradation of the reconstruction [110]. This is a natural characteristic of any computational imaging technique based on some sort of movement, and in fact, applies e.g. to super-resolution imaging [8]. Frame registration aims at overlaying two or more images onto a coherent coordinate system without user intervention; in the context of ptychography, the registration consists of planar shifts that define the correct area illuminated by the probe for each projection, due to the real sample stage movement.

The literature taxonomy on position refinement techniques presents two main groups:

1. the first one arranges a frame registration task that is defined, either between adjacent estimated object projection on the detector $\Phi_j = D_z\{O_j \cdot P\}$, or directly between the recorded diffraction patterns I_j. Clearly, this latter variant can be used if the ptychography experiment is set in the near field, where a portion of natural images properties (image features) are somehow conserved (or in the case of a curved illumination wavefront, as discussed previously). An in-depth discussion is described in section 2 of this chapter.
2. the second one exploits an optimisation based approach directly using

the main reconstruction metric (e.g. MSE calculated between the estimated and the acquired diffraction pattern (much more details in the third section of this chapter);

In [111], similarly to the third section of this chapter, a non-linear programming framework (NL) is defined for the joint optimisation of the object, the probe and also for the positions. Regularisation strategies have to be taken into account because, as supposed in [112], this approach can easily lead to local minima since several parameters are used. To cope with this problem, a hybrid approach is suggested, as in [112], by combining a conventional reconstruction algorithm such as ePIE [91], Maximum Likelihood (ML) [84] or Differential Map (DM) [38] with an optimisation based positions correction routine. In the case of [92], the probe correction is carried out in a gradientless approach using the Powell method [113], while in [112] the optimisation provided by [111] is used.

The limitation of the Non Linear programming methods is twofold: not only the gradient expressions are available just for the application on far-field setup, but, regarding the implementation, the method is extremely slow because the positions are introduced into the forward model by using the shifting properties of the 2D DFT; this in turn not only increases the computational box by a factor of two (to avoid circular convolution repetition), but also triplicates the total number of the DFTs required to perform the convolution in the frequency domain, where the phase-only term describes the shift.

In a recent work [114] the reconstruction is carried out by EPIE [91], but the position refinement procedure exploits a finite-difference algorithm that results in an approximate but faster update step than the one used in [111]. The same authors recently published a second method exploiting both the combined action of the cross-correlation between re-projected diffractions and a shifted support constraint in an inner Hybrid Input-Output (HIO) reconstruction loop.

In [115] an annealing algorithm is described, essentially by exploring the shifts parameter space and using the main metric as a reference; unfortunately, methods of this type are for their nature computationally expensive. Simple techniques belonging to this first category are applied also in a tomography environment: in [116] a joint reconstruction-alignment is performed reconstructing the object for one iteration and successively registering the acquired tilt-series with the corresponding simulated projection.

The work [47] is a simple and effective method belonging to the aforementioned second category in the methods taxonomy, in which a cross-correlation based method [117] is used as a similarity metric between two adjacent portions of the object; the argmax of the resulting 2D shift map is directly used for the correction. A similar approach is defined in [3], highlighting the use of the Structural Similarity Index [118] as a more suitable similarity metrics. This is the method we proposed and that is described in the following section.

3.2 Metric based approach

In this section, we will discuss about the position refinement of independently reconstructed probes. Considering two images, one as the reference image and one as the movable image, registration (as in alignment) [119, 120] is the process of finding correspondence between all points of them [121], using similarity/dissimilarity measures and metrics. This imaging topic is well described in the literature [122] and is treated historically in the field of motion estimation for video coding, by an ensemble of techniques based on the so called "block matching" [123, 124]. Sub-pixel registration is typically obtained by up-sampling both images [125], leading to an uncertainty in the measure that has been analysed analytically for a phase correlation metric case [126]. Another large part of the literature is devoted to 2D/3D Medical Imaging registration methods for tomographic, ultrasound and magnetic resonance (MRI) scans [127]. In the case of ptychography, accurate registration is crucial for a correct reconstruction [47].

In this work only translation (i.e. horizontal and vertical displacement) will be considered. The alignment is performed in a pairwise manner between adjacent probes. In certain cases there may be other optical issues such astigmatism where only a part of the probe is in focus, but this study assumes typical ptychography beamline setups. The initial strategy is heuristic and exhaustive; it considers an estimated x, y position (i.e. the one provided by the encoders) for a pair of probes. That initial position results in a series of possible positions within a specific range (i.e. x - 10 to x + 10, and y - 10 to y + 10 with steps of one). For each of these positions, an evaluation of the resulting alignment is performed (figure 3.1).

For each possible position, the quality of the alignment is estimated. The main hypothesis is that an imprecision of only few pixels even on a single axis can impact negatively the reconstruction, and that the problems have solutions through error minimisation. Figure 3.2 illustrates the imprecise position impact (panel A,B) and how a basic Standard Deviation based minimisation method can lead to the refined alignment (figure 3.2. Panel C).

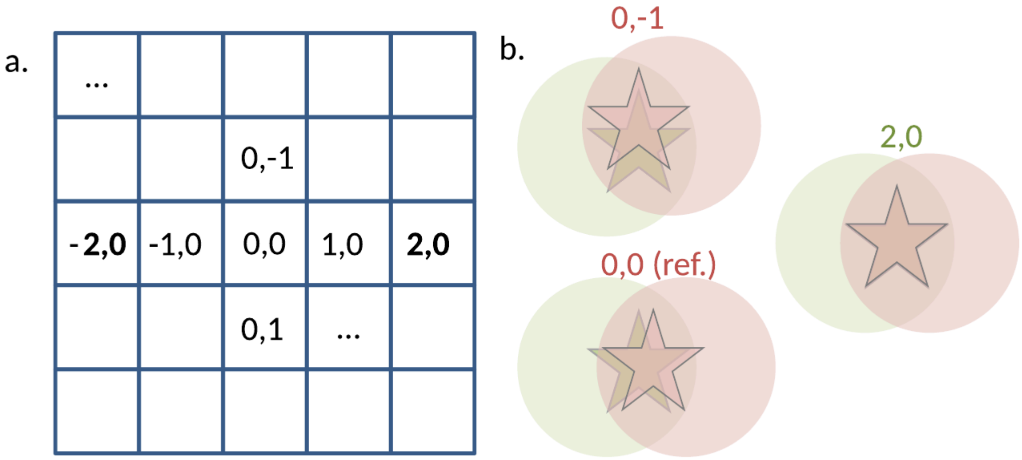


Figure 3.1: Starting from an alignment position as a *reference position* (0,0) a series of possible adjustments is generated (a). For each of these positions the resulting alignment is evaluated (b). The correct alignment (i.e. 2,0) is the *relative correction* that needs to be applied to the reference position of the pair of probes.

3.2.1 Automated Position refinement trough Structural Similarity

After implementing multiple methods for evaluating the probe alignment, we propose the use of the Structural Similarity Index Method (SSIM) as a way to quantitatively measure the alignment and reduce it to an error minimisation problem. Later we will shows the data of other methods used, which proved to be inferior to SSIM. Structural Similarity index (SSIM) [118, 128] is an image measure developed in the context of visual quality ranking as a response for the inadequacy of Mean Square Error (MSE) to mimic the human perception of the image quality. SSIM splits the task of confronting two images in three steps: luminance measurement/comparison, contrast measurement/comparison and structure measure/comparison. Each step produces an index that is combined in an exponentially weighted multiplication manner. The Universal Quality Index (UQI) [129] follows a similar approach but is considered poor in terms of numerical stability; on the contrary SSIM performs better by applying a series of constraints [118].

For a matrix containing a range of positions refinements (figure 3.1, panel a), the quality of the alignment is measured through SSIM on that position resulting to a matrix (panel a) that should yield the relative position correction. This operation can be done pairwise for any adjacent probe. The

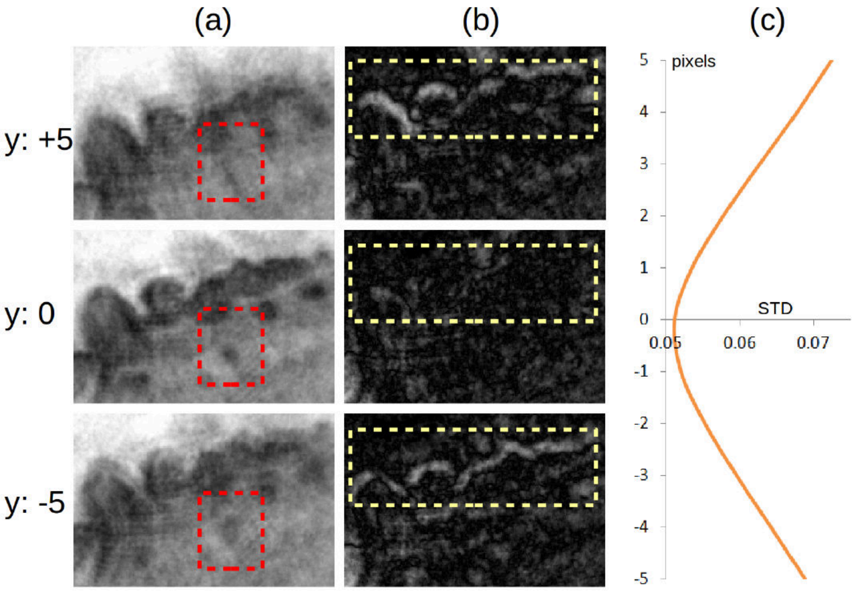


Figure 3.2: Panel (a) shows a ROI of the magnitude resulting from 2 probes at 3 different relative positions. The images are moved vertically +/- 5 pixels for each coordinate of the aligned position. Panel (b) is the Standard Deviation (STD) between the two probes at the different positions highlighting the misalignment. The brighter areas in (b) suggest higher STD thus worse alignment. Panel (c) is the average STD for each position as in (b) demonstrating that the minimum is at the aligned position. The rectangular boxes point out certain areas of interest where the differences are more evident.

calculation takes place only for the common/overlapped part (probe intersection) (figure 3.3 panel b) and results to a value between +1 and -1. Better results are obtained calculating a map of the index using a sliding Gaussian weighted square window along the images, and confronting each block separately; averaging the index for all the blocks is possible to evaluate the overall image quality [118]. In this work, Mean SSIM (MSSIM) has been used for all the images in the dataset.

Additional measures

Even at the early stages of this work, it has been clear that even small inaccuracies in the positions can impact the reconstruction quality. On the other hand, while attempting the automated refinement based on an error minimisation it was rather difficult to choose the right alignment evaluation

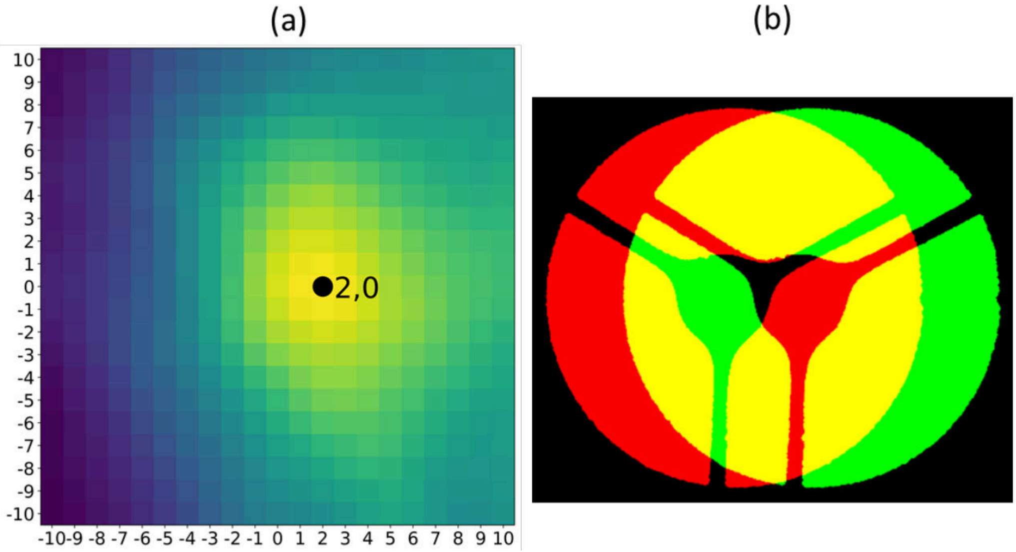


Figure 3.3: Panel (a) is the Mean SSIM for a range of ±10 pixel refinements on the alignment of two probes.The black dot indicates its maximum value which suggests that the alignment should be corrected by+2pixels in one axis while keeping it as it is on the other one. This operation is performed only on the overlappedarea of the masked probes Panel (b). The yellow area indicates the intersection of the two probes

function. The simple Standard Deviation produced encouraging results (see figure 3.2 Panel C) but it was proven not general enough as it works only for few specific datasets. In this section we present a few of the negative results with different methods we tested before selecting the Mean Structural Similarity Index (MSSIM). Note that this difficulty of choosing one method, lead to the idea of applying machine learning to the task. Other than the error minimisation approach, we also tried traditional image registration techniques, but they all produced poor results due to the nature of the reconstructed probes (in particular the presence of beam stop) (figure 3.4).

For the error minimisation approach, this study considered Mean Square Error (MSE), Correlation Ratio, Pairwise Standard Deviation, and Mean Structural Similarity index (SSIM) (figure 3.5). For certain pairs of the ptychographic scan (i.e. for different areas of the sample) one technique could be more performant than the other. Overall, SSIM was the best method for demonstrating the concept of alignment through error minimisation (figure 3.6).

Ongoing work should utilise machine learning techniques in order to com-

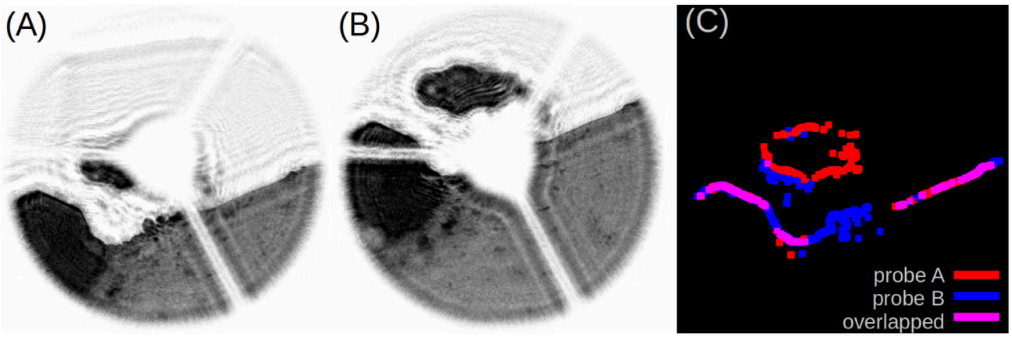


Figure 3.4: (A and B) Two neighbouring reconstructed probes for which it is necessary to know their relative position. Note in the center the presence of the beamstop and its support. The alignment/registration needs to be performed only on their common parts. Panel (C) highlights the edges from the two probes when aligned

bine multiple methods at the same time. Regarding the performance and scalability of the proposed methods, the problem of position refinement in ptychography heavily depends on the number of frames that a scan consists of. The datasets used in this work had less than 100 frames so even a manual alignment can be done in hours. There is a trend for larger scans with hundreds of thousands of frames where the manual option is not applicable. The error minimisation (heuristic) approach can definetely handle larger number of frames due to its automated nature. Still for even larger scans (i.e. $>$ 1000 frames) the problem may need to be split into smaller regions and processed in multi-node computer clusters. Eventually the use of machine learning may be the way for processing very large datasets while using much less resources than what is required by the previous methods.

3.2.2 Software for manual refinement of positions

Ideally, either the positions of the probes are already known or the reconstruction algorithm does not require them to be really precise, however in most cases those positions are imperfect and require refinement. While attempting to develop automated position refinement methods, this research led to the development of a software system for the manual alignment. Even if using such system is a laborious task, it was proven of high value in the field for various reconstructions. The process requires as input the reconstruction of each single probe independently. The alignment is performed through a

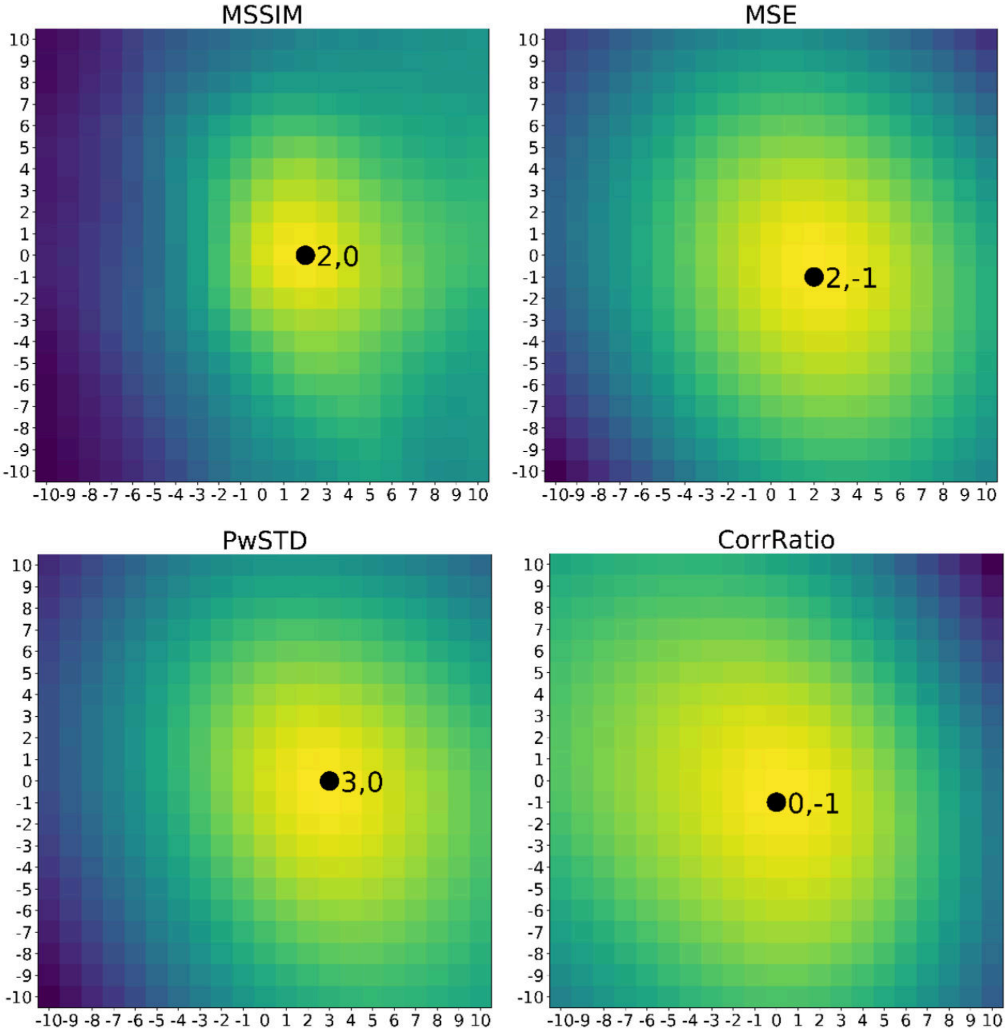


Figure 3.5: Mean Structural Similarity index (MSSIM), Mean Square Error (MSE), Pairwise Standard Deviation (PwSTD), and Correlation Ratio (CorrRatio) for a range of +/-10 pixel refinements on the alignment of two probes (such as the one in figure 3.4). The black dot indicates its maximum value (or its minimum in the case of dissimilarity measures) which suggests that the alignment should be corrected. MSSIM provides the correct answer

graphical tool which eventually outputs the scan positions required for the complete reconstruction (figure 3.7). To our knowledge this tool is the first of its kind; however it presents various technical challenges mostly related to the high pixel count and dynamic range of the images. Each reconstructed probe image can contain several megapixel (e.g. 2130 x 2130) of 32 bit in-

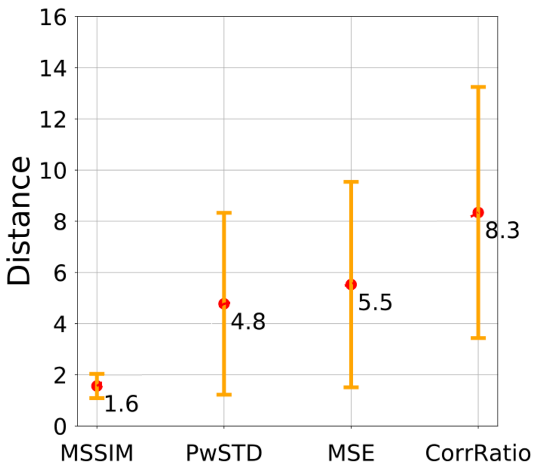
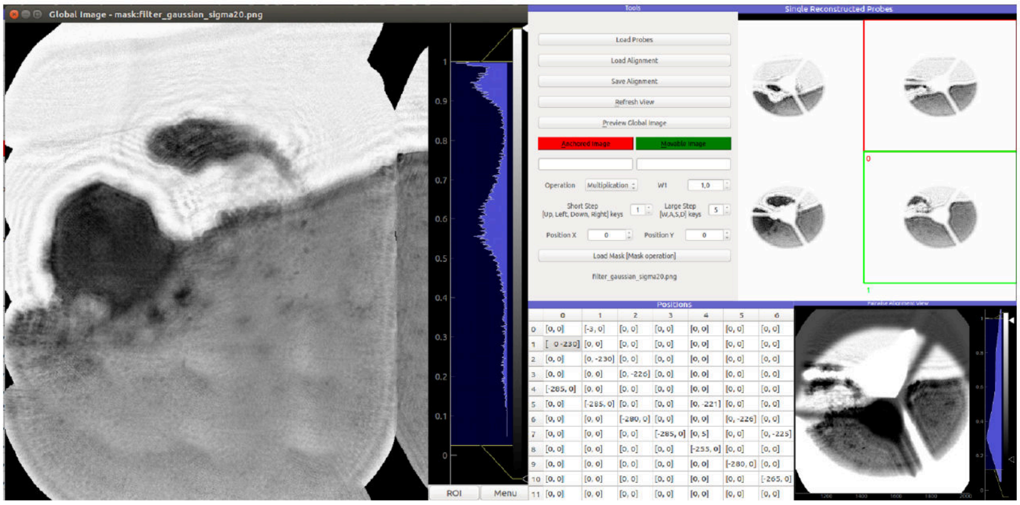


Figure 3.6: Distance between the suggested refinement and the initial position, calculated for all the probe pairs of a ptychographic scan. Other than the mean value for each method, there is a range bar showing how much the distance varied from pair to pair. The lower the value and smaller the variation the better is the method.

formation. Each scan may have dozens of them. Since this is an interactive application, the manual alignment has to be responsive, meaning displaying immediately the result of the new position while the user is moving the images by mouse or keyboard. The system is developed in Python and heavily relies on PyQtGraph whose OpenGL renderer provides a rapid plot update and responsive user interaction. The system is extensible and new functionalities can be easily added-on including the automated approach described in the previous section. Image controls such as zooming, translation, levels, contrast, and changing of look-up tables are implemented. A future version will also implement sub-pixel alignment, which may be useful for really high resolution ptychography (<5 nm).

3.2.3 Machine Learning for the position refining

The issues studied in this research project involve processes that may benefit from Machine Learning techniques (ML), specifically Convolutional Neural Networks (CNNs). CNNs in their general form and after suitable system preparation (referred as training) have been used to provide the correct response for certain application of imaging problems. Indeed CNNs are nowadays used beyond traditional Computer Vision tasks and have multiple applications in physical sciences. CNNs have been used for Tomographic



Operation: Multiplication
W1: 1,0
Short Step [Up, Left, Down, Right] keys: 1
Large Step [W,A,S,D] keys: 5
Position X: 0
Position Y: 0
Load Mask [Mask operation]
Filter_gaussian_sigma20.png

	0	1	2	3	4	5	6
0	[0, 0]	[-3, 0]	[0, 0]	[0, 0]	[0, 0]	[0, 0]	[0, 0]
1	[0 -230]	[0, 0]	[0, 0]	[0, 0]	[0, 0]	[0, 0]	[0, 0]
2	[0, 0]	[0, -230]	[0, 0]	[0, 0]	[0, 0]	[0, 0]	[0, 0]
3	[0, 0]	[0, 0]	[0, -226]	[0, 0]	[0, 0]	[0, 0]	[0, 0]
4	[-285, 0]	[0, 0]	[0, 0]	[0, 0]	[0, 0]	[0, 0]	[0, 0]
5	[0, 0]	[-285, 0]	[0, 0]	[0, 0]	[0, -221]	[0, 0]	[0, 0]
6	[0, 0]	[0, 0]	[-280, 0]	[0, 0]	[0, 0]	[0, -226]	[0, 0]
7	[0, 0]	[0, 0]	[0, 0]	[-285, 0]	[0, 5]	[0, 0]	[0, -225]
8	[0, 0]	[0, 0]	[0, 0]	[0, 0]	[-255, 0]	[0, 0]	[0, 0]
9	[0, 0]	[0, 0]	[0, 0]	[0, 0]	[0, 0]	[-280, 0]	[0, 0]
10	[0, 0]	[0, 0]	[0, 0]	[0, 0]	[0, 0]	[0, 0]	[-265, 0]
11	[0, 0]	[0, 0]	[0, 0]	[0, 0]	[0, 0]	[0, 0]	[0, 0]

Figure 3.7: Example of Graphical User Interface for a set of this software tools for manual refinement of alignment of probe/images.

and Holographic problems [130] but also recent applications in ptychography have produced promising results [131]. In chapter 4, a solution for Fourier Ptychography will be presented.

After examining the position refinement as an error minimisation problem where MSSIM came out as a promising metric, this study also initiated an investigation on possible paths for the use of Machine Learning to the problem of position refinement. This is an on-going research work but the three most promising paths are the following:

1. Use of an extensively trained ML system in order to provide position correction. This had limited success but it is the traditional use of ML methods for such applications.
2. Given a set of metrics like MSSIM, MSE and PwSTD, let the ML system reveal the best one.
3. Given a set of metrics, let the ML system combine them in the most suitable way. This approach and the previous (2) yield the best results so far.

See figure 3.8 for a schematic description. An ML technique is composed of a trainable structure with tweak-able parameters, a training algorithm, a cost function, an optimization algorithm and a relatively high amount of data. One of the most used training paradigms is the supervised approach: training algorithm undergoes many labelled examples to the trainable structure that learn to find out "what matters" for the problem resolution, building its own model.

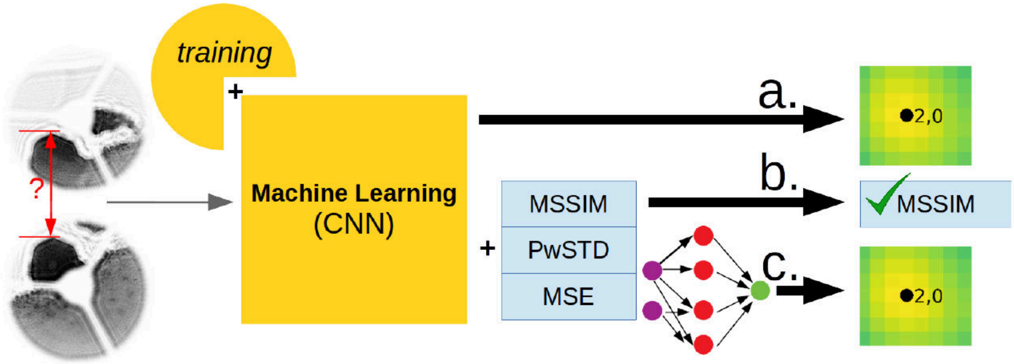


Figure 3.8: Outline of proposed use of Machine Learning for probe position refinement. Give a pair of probes an ML system, after being trained, can be used in three ways: (a) black box where the system should provide the right correction after extensive training, (b) the ML system chooses the right error minimisation method (e.g. MSSIM) from a set of possible ones, and (c) the ML system uses all available measurements and combines them in a suitable way to provide the best position refinement.

Computer Vision (CV) is one of the fields in which ML takes action. Specially constructed NN, Convolutional NN (CNN) smartly use 2D convolutions to reduce the amount of coefficients needed to solve many CV tasks, like object detection/recognition and potentially probe position refinement. Some preliminary results that this research has produced suggest that the following ways on applied ML may assist in solving the problem of position refinement.

- **CNN as a black box** An ImageNet pre-trained residual-CNN [132] is modified in order to take two images as an input and give a 2D map output with the same size of the input images. Every pixel in the output is connected to a SoftMax layer that computes the probability that a particular position is the correct value for the translation. It is important to note that those probabilities are taken from a conjunct model, that is translation probabilities are not independent values. Even if the output is a map, the problem is reduced to a classification problem, as the final response is done by computing the best score class. If a regression approach and MSE loss were used, special care would be needed to avoid the nulling of the output probability variance. The drawback of this method is that understanding what “reasoning” the net is doing is hard. After the training, the network is ready to be used for metric calculation. A similar approach has been used in [133].

- **ML for using multiple error minimisation methods together** In order to understand better what the network does, it is a good idea to use a hybrid approach for a supervised-learned regression in which, giving the two images as an input, the network has to calculate coefficients of a polynomial whose terms are hand defined metrics, calculated on a set of positions. The regressed function's minimum will give the correct translation value. In this case, the training part is the only one connected to the network use, because after the regression, using the calculated coefficients a better measure could be put on work. For this pure regression approach, MSE could be a good candidate for a loss function. Every pair of image input has a label with the ground truth value of translation.
- **ML for identifying the best error minimisation method** ML provides a suitable tool for teaching machines how to choose well. Providing a network with a couple of images, by using properties of the images, the network is trained in order to choose the right metric, i.e. the one whose minimum gives the least MSE with the ground truth value. This type of training could be used even to initiate a CNN based metrics: before a direct training of the embedding in the feature space with siamese network and triplets [134], one of the common methods used for object/face recognition was to use a classification approach, to train on the classes, cut out the classification part of the network and use inner layers outputs as output layer. Those inner layers output the network features that could be used to infer a new metric. Plugging to this network another set of fully connected layers with regression output could lead to a good metric generator. The reason behind this procedure is that a good training of features is difficult to be obtained directly, even using smart techniques.

3.3 Iterative refinement

A fast and reliable way to introduce a simple form of correction exploits a correlation signal calculated at some points in the reconstruction chain. Figure 3.9 shows how an iterative position refinement scheme of this kind can be configured within a projection-based reconstruction algorithm. Where to "sense" for the error? For each *jth* x, y position, current methods [112, 114, 47] observe the cross-correlation at:

- **A)** at the object plane (between a corrected and uncorrected object estimate at the same jth computational box), while other methods, for example in CT [116] acts on the detector plane
- **B)** calculating the correlation between the *jth* acquired diffraction pattern and the *jth* simulated diffraction pattern.

In addition, it seems also unnatural to check between two "artificial" quantities instead of relying at least on one real image (the acquired diffraction pattern). The analysis of the dynamics of the two candidates is the objective of this section; in both cases, the error signal $argmax(xcorr)$ is extremely small (in the order of 0.1, 0.01 pixels) and provides a correction factor that is cumulated at each iteration. A computational box is defined by rounding to the nearest integer and cropping the resulting area. The method of choice for the subpixel shift detection is based on the work [117] and its implementation in SciPy. We adapted the same algorithm in PyTorch, allowing for fast GPU computation and avoiding costly move operation between the two memory systems. The core of the fast procedure relies on the fact that after a single DFT (Discrete Fourier Transform) operation at the first scale, it uses a matrix multiplication implementation of the 2D DFT to refine the initial peak location estimate. The advantage of this approach results from the fact that an upsampled version of the cross-correlation can be computed within just such a neighbourhood without the need to zero-pad.

Update procedure

Each position (the jth diffraction pattern) is updated at each iteration by the following expression:

$$(x,y)'_j = (x,y)_j + \beta_j \cdot \text{argmax}\{XCORR_{i\in[A,B]}\}, \tag{3.3.1}$$

where β_j is a tweakable gain factor (>> 1) for the small error signal (the argmax of the 2D correlation), and i can be set to A or B, meaning that we take the argmax of the crosscorrelation at the detector or at the sample plane.

To investigate the dynamics of the argmax error signal, we used the SciComPty simulator to generate a near field ptychography dataset in which the object function $O(x,y)$ and the illumination function $P(x,y)$ are assembled in magnitude and phase providing two images each. Then the positions are perturbed by adding a gaussian error on the x,y coordinate with standard deviation equal to 5% of each object view. A reconstruction is started

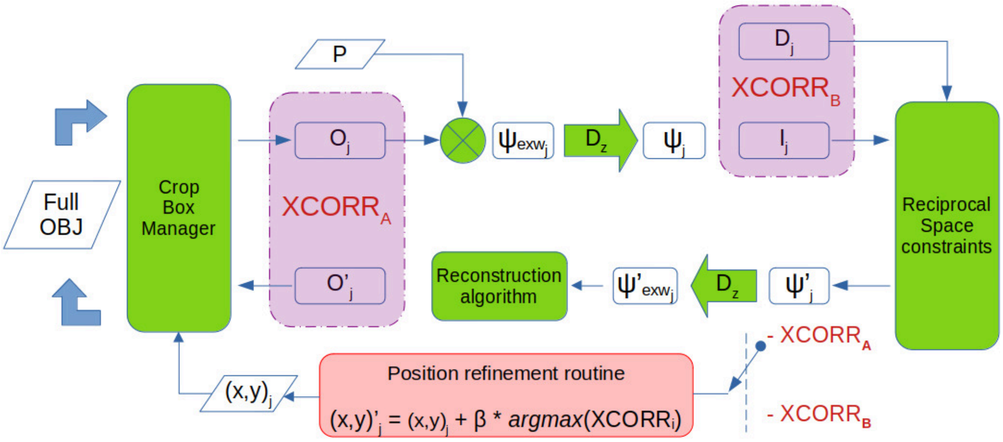


Figure 3.9: Integration of a correlation-based position refinement routine into an iterative algorithm. No alteration on the reconstruction algorithm is requested. The position error signal (argmax of xcorr) can be calculated meaningfully in two points, denoted by XCORRA and XCORRB. The estimated position is then corrected at each iteration by applying a variable gain β on the raw error signal. An automatic method (Adam) is used to accelerate the process by varying beta and then amplifying the error signal.

enabling the position correction routine, based on the error signal XCORRA or XCORRB. The reconstruction is carried out for 500 iterations considering all the $N = 121$ diffraction patterns and the process is monitored iteration after iteration by recording three metrics:

1. the reconstruction error (e.g. defined in Eq. 2.5.3) [38, 91]

$$E = \sum j = 1N|\Psi'_j - P \cdot O_j(\mathbf{r}_{j-est})|^2 \tag{3.3.2}$$

   where as usual P is the probe function, Ψ'_j is the corrected exit wave and $O_j(\mathbf{r}_{j-est})$ the current calculation box that depends on the current estimate of the position $\mathbf{r}_{j-est}$.

2. the average estimated position error average error signal magnitude;

$$E_{estpos} = \frac{1}{N}\sum_{j=1}^{N} \beta_j \cdot \text{argmax}\{XCORR_{i\in[A,B]}\} \tag{3.3.3}$$

3. the ground truth average error, calculated between the estimated po-

sition $\mathbf{r}_{j-est}$ and the ground truth value ($\mathbf{r}_{j-gt}$)

$$E_{gtpos} = \frac{1}{N}\sum_{j=1}^{N} |\mathbf{r}_{j-gt} - \mathbf{r}_{j-est}|. \tag{3.3.4}$$

If beta is a constant equal to 1 and XCORRA is chosen, the output of the reconstruction can be seen in Fig. 3.10: the estimated position error signal (Eq. 3.3.3, green curve) drops rapidly to zero, meaning that no further correction on the positions will be done. As reported in [47], the same effect can be observed for a constant gain. That is why both the reconstruction error (Eq. 3.3.2, blue curve) and the ground truth position error (Eq. 3.3.4, orange curve) converge rapidly but to a high value, flagging a bad reconstruction. Also for the case of XCORRB (Fig. 3.11, only a marginally better reconstruction (see the inset panel of the illumination magnitude) can be gathered. The error signal (green curve in 3.10 and 3.11) is noisier, due to the fact that only half of the information (real images) is present.

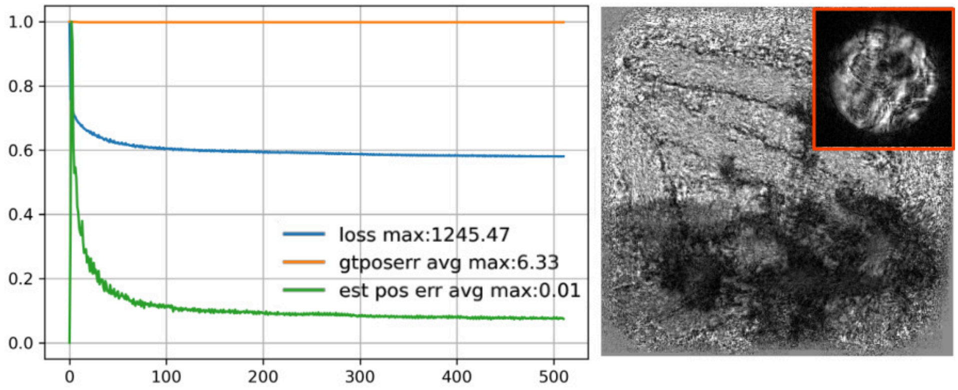


Figure 3.10: XCORRA signal (green curve) with $\beta = 1$ has a low dynamic range, practically not refining the coefficients. The eye of the cat (illumination magnitude) is not correctly retrieved.

By introducing a gain factor (*beta* > 1) the convergence can be greatly improved not only in speed terms but also toward a better solution (hopping local minima). At the same time, a too-large value can cause the reconstruction to stagnate or even to diverge. An adaptive correction is required, ideally per-parameter.

Differently from other solution provided in the literature, here we propose to calculate adaptively a vector of gain factors, one for each position, by using the Adam optimization algorithm [135]: Adam is an extension to

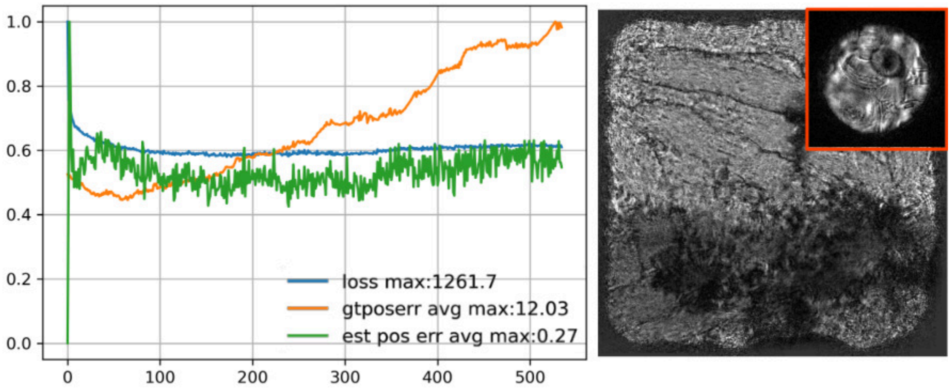


Figure 3.11: XCORRB signal (green) is so noisy that can make the error diverge also for $\beta = 1$. The reconstruction is marginally better than the one in Fig. 3.10.

stochastic gradient descent that is nowadays widely used in deep learning; while in a stochastic gradient descent method the same learning rate is kept constant for all the variables in the optimisation pool and among each iteration, Adam provides a per-parameter factor that is separately adapted as learning unfolds. This is achieved by calculating two parameter-controlled exponential moving averages of the first and second moment estimation. In our method, the argmax of the cross-correlation measure takes the place of the gradient in the Adam algorithm [135], because intuitively its value will maximise the position gradient. The proposed algorithm is shown in Algorithm 1:

```
Data: The cropped object $o_j$ and its refined estimate $o'_j$
Result: Gain parameter $\beta_j$ for the jth crop-box related to the jth
        diffraction pattern $I_j$ and the jth position vector $(x, y)_j$
Initialization as in Alg. 1 of [135];
while reconstruction iteration $> 0$ do
    t = t+1 ;
    $g_t$ = argmax$\{XCORR_i\}$ ;
    proceed as in Alg. 1 of [135];
end
```

Algorithm 1: The modified Adam algorithm uses the argmax of the phase correlation as the gradient of the error, then the exponentially damped moving average are updated as usual.

Figures 3.12 shows how the implementation of Adam is effective in adapting the gain of the estimated position error signal, retrieved at XCORRA, just where in the case of unity gain the signal was ineffective. Conversely, applying Adam at XCORRB (Fig. 3.13) the correction is marginally improved due to its noisier nature.

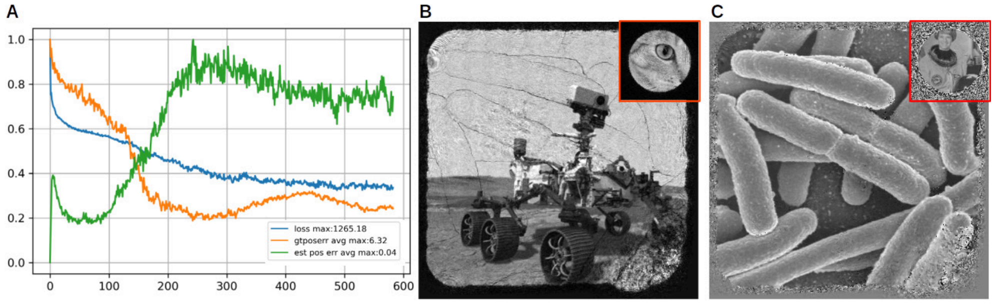


Figure 3.12: Adam is now used as a gain controller, in conjunction with XCORRA (green). The loss (blue) and the ground truth position error (orange) are reduced, providing a good reconstruction for the magnitude and phase of both the object (panel B and C) and the illumination (insets in B and C).

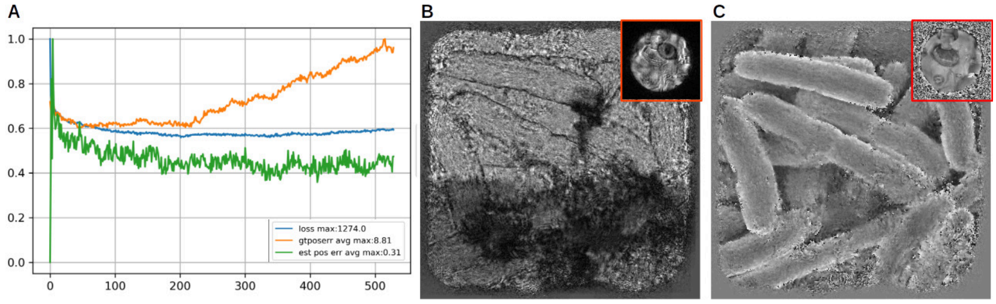


Figure 3.13: Differently from Fig. 3.12, the error signal in XCORRB is noisy allowing only a partial correction.

The choice of an error signal from the object plane finally coincides with what was proposed initially in the literature [112, 114, 47], but only if a gain factor is applied; the statistic-based method proposed in [47], in our experiments and implementation, not only is somehow difficult to tune but does not provide a vector of per-parameter controlled acceleration.

The choice of an error signal from the object plane finally coincides with what was proposed initially in the literature [112], but only if a gain factor is applied; the statistic-based method proposed, in our experience and implementation, not only is difficult to tune, but does not provide a per-parameter controlled acceleration.

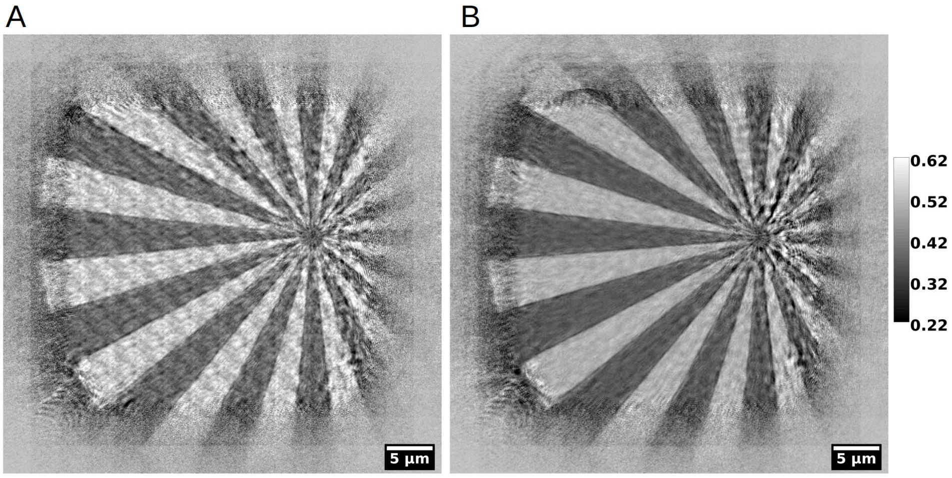


Figure 3.14: Magnitude reconstruction of a siemens star without (panel A) and with (panel B) the proposed position correction.

Figure 3.14 shows the object magnitude reconstruction of the same sample in figure 2.24: considering panel a and b, it can be seen how the proposed position correction procedure is capable of providing way less artefacts on the reconstruction. The raster grid pathology artefacts [38] are also greatly reduced, even if the raw positions commands belong to a regular grid. Indeed, the actual movements performed by the sample stage differs from the regular position commands grid sent by the control software by a small irregular jitter, that is caused by the backlash. That is why a fast implementation of a position refinement algorithm (as deployed by this work) is essential. The analysis of the position error dynamics is also critical to investigate, in order to understand why position refinement techniques such as [116] cannot work in a ptychograpy environment: the use of complex quantities onto the sample plane provides a less noisy estimate then its real counterpart on the detector plane. Also, the images at the object plane tend to be way more intercorrelated than the real and estimated diffraction pattern, providing a more robust shift estimate.

3.3.1 Details on speed performance

In chapter 2.4 we presented a new software framework, named SciComPty entirely written using PyTorch, an open-source Deep Learning tool, that here in this part of the current chapter is used as an easy route to fast and efficient GPU programming.

The performance gain obtained by computing the entire computation on GPU is about 10x (0.1 vs 1.3 s). As we have seen (figure 3.9, the position refinement technique described in this section relies heavily on fast computation of the relative shift between two different images. Currently no implementation of [117] are available for GPU computing. Then during this research we developed an implementation of the algorithm, to avoid invalidate the performance gain obtained so far. Actually if we would have used a CPU algorithm for the registration, the overall speed could have been even worse than a full CPU reconstruction and registration algorithm, due to the overhead given by the data transfer.

As briefly said in chapter 1.3, it is extremely important that all the data transfer between the system and the graphic card RAM are minimised. The PCI-Express link provides a severe throughput bottleneck. Figure 3.15 shows the computation time measured for the registration algorithm implemented on CPU and on GPU: also in this case, the performance gain is of about 10x. This means that the acceleration is huge, due to the fact that in a typical ptychography scan, roughly 100 probes are used, than the total iteration time of 3.15 has to be multiplied by the number of probes.

Due to its large field of application, the subpixel registration routine we developed in this work, will be released as an open source python library.

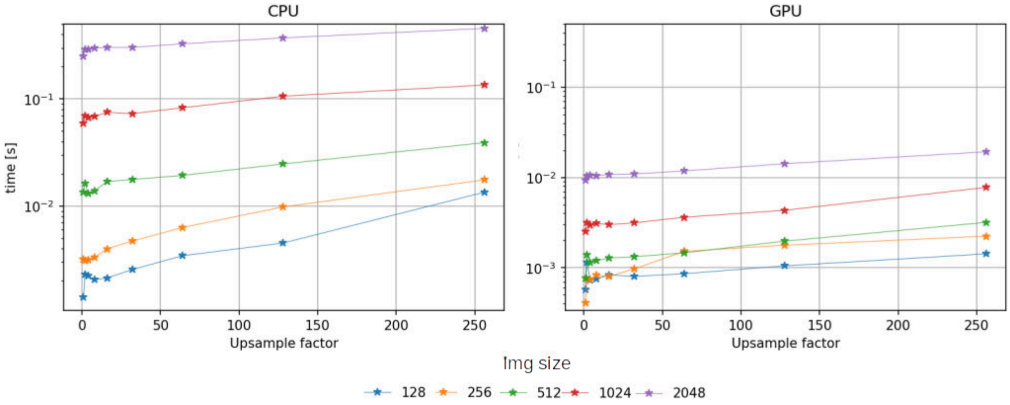


Figure 3.15: A fast routine is essential in order to calculate the relative shift heavily used in this method. Panel A shows the shift computation for a single pair of images (reference, shifted) time done on CPU, while panel B our implementation ported on GPU. The performance gain in terms of speed is about 10x

3.4 Optimisation based algorithm

Many ptychography algorithms are formalised defining a loss function that is minimised varying the latent parameter of the image formation model. In our case, the set of latent parameters is made up the pixels of the object to be reconstructed, represented as a numeric array. From the optimisation theory, calculating the gradient of this loss function with respect to the latent parameters is the central action of the so-called "gradient-based" methods, since each optimisation step is taken in a direction related to the gradient itself (parallel, conjugate). On the contrary, *"the curse of dimensionality"* restricts gradient-less optimisation to a way smaller problem, essentially because the gradient is estimated numerically (e.g. by the symmetric derivative quotient) and not calculated from an expression.

As we said in chapter 2 and in the early part of this chapter, in CDI, the object O is typically reconstructed using a projection based method. We can see this procedure also in the light of a naive implementation of a constrained optimisation procedure, performed by defining constraints in two different domains: the intensity of the acquired diffraction pattern is imposed in the reciprocal domain by operator π_m, while extension mask and physics constraints are applied in the real space by operator π_s:

$$O' = \pi_s\{D_z^{-1}\{\pi_m\{D_z\{O\}\}\}\} \tag{3.4.1}$$

where the D_z propagation operator is used to move back and forth from the real to the reciprocal space representation of the object.

In order to avoid considering these complex operators in the gradient calculation, many algorithms define a loss function as a dissimilarity metric between the object, considering it before (O) and after (*O'*) the application of a set of the aforementioned constraints.

$$\mathcal{L} = \left\|O - O'\right\|^2 \tag{3.4.2}$$

At convergence, these two quantities will ideally coincide because each parameter will already satisfy the constraints (zero dissimilarity). In this formalism, the back and forth propagations cancel out and do not appear in the loss function. If from one hand this choice allows to treat the problem easily within an optimisation framework, from the other it excludes any automatic tuning of the parameters that model the setup.

Algorithms of this type are generally referred as *Projections Over Sets* (POS methods): a set describes a region in the solution space in which each reconstructed object has a particular characteristic, common among each solution in that region. A constraint operator acts on solution estimates projecting that point towards the validity region (another set) of this operator.

If multiple projectors are applied iteratively, it is highly probable that the resulting points will lie in the intersection of the two regions (if it exists). It is necessary to check that the solution point actually possesses all the required properties to be accepted as a valid solution. If the loss function is an L2 norm, the sets are also convex. It is also possible to think about this modus operandi as a "non-direct" optimisation.

This approach is appealing because:

- there is no need to explicitly implement an *inverse operator*;
- the reconstruction process can be monitored as it progresses;
- the number of iterations can be used as a regularisation;
- the noise effect can be monitored;
- it can be applied easily to complex a priori forward models;

CDI techniques (and thus ptychography) are extremely sensitive to a coarse parameter estimation [110, 109], and so lead to a severely degraded reconstruction; a long, tedious and time-wasting trial and error procedure is typically employed to refine the values of the parameters.

To cope with this situation it is then necessary to modify the optimisation procedure, choosing an image formation model in which the desired parameters do not cancel out so that their gradient can be calculated. Algorithms of this type will be referred to as "direct optimisation", where an estimated quantity $\{O\}$ is directly compared with the real data (I):

$$\mathcal{L} = \left\| \widetilde{O} - \sqrt{I} \right\|^2 \tag{3.4.3}$$

The estimated quantity is then a function of the current object estimate and the setup parameters vector θ:

$$\widetilde{O} = f(O, \overline{\theta}) \tag{3.4.4}$$

where θ can consist for example of position vectors and propagation distances. It has to be kept in mind, however, that for this complex loss function the gradient expression calculation may become an extremely tedious task.

The aim In this section, an automatic differentiation based ptychography reconstruction algorithm is presented. Exploiting the automatic differentiation capabilities of the modern Deep Learning (DL) tools, a loss function is derived by explicitly taking into account both the propagation distance and the scan positions; these two sets of parameters are added to the optimisation pool for a joint optimisation-reconstruction.

Differentiable programming attempts in ptychography Investigation on autograd-based methods is extremely important because they represent a fundamental research tool in applied optimisation, and thus in most computational imaging techniques. Differentiable programming methods are getting lots of attention in the last years: for example, in [136] many inverse problem use-cases are described, while in [137] an entirely new and GPU-optimized programming language is defined for image processing tasks. The first application of autograd-based techniques related to ptychography appeared in an inspiring 2018 paper [138], in which a Fourier Ptychography (FP) setup is described using Tensorflow [139] to reconstruct both the object and the Fourier transforming lens pupil. The first autograd (real space) ptychography algorithm has been published very recently [140], being the milestone for this kind of algorithms. This intriguing paper describes an auto-grad based reconstruction algorithm, testing it on both far-field and near field simulated data, in a transmission and Bragg reflection virtual setup. The paper also provides an introduction to the autograd world as well as a review on the basic ptychography concepts.

Differences with the current state-of-the-art In the present work we independently followed the same general idea of [140] but: i) an angular spectrum propagation model is implemented in place of the Kirchhoff Integral; setup parameter invariancy is added by including both ii) a position refinement routine and iii) the propagation distance regression. From the best of our knowledge, no attempts in the literature have been made to cope with the problem of the distance refinement; in [109] the somehow similar problem of illumination function refocusing is tackled but as a form of simple probe retrieval for electron ptychography. Furthermore, no position refinement techniques have been implemented in an autograd environment, defining a limit for this kind of methods. In [140] a good reconstruction is reported to require thousands of iterations, while in our experiments, with the baseline conditions, just hundreds are needed. The number of iterations may, however, increase to reconstruct perturbed data with high uncertainty in the parameters.

Loss Function

In this section, the idea of joint reconstruction-optimisation of the setup parameters is further investigated, adapting the algorithm to the near field and to an autograd environment and expanding it for the first time to the propagation distance optimisation.

Due to having a complex probe interacting with and a complex object, loss functions for ptychography/CDI reconstruction are of the type:

$$f : C^p \mapsto R^1 \tag{3.4.5}$$

In the case where all the hypotheses defined in chapter 1 and 2 hold, a direct loss function for all the j = 1...N diffraction patterns can be written:

$$\mathcal{L} = \sum_{j=1}^{N} \left\| D_z[P(x,y)O(x-x_j, y-y_j))] - \sqrt{I_j} \right\|^2 \tag{3.4.6}$$

where $P(x,y)$ is the probe, $O(x-x_j, y-y_j)$ is the result of the region cropping operator applied on the object to select the probe illuminated area of coordinates (x_j, y_j) and I_j is the diffraction pattern intensity of the jth object section acquired by the detector. The square root of the recorded data can be computed once for all the diffractions to fasten the implementation. D_z is the angular spectrum propagator [18] defined by the expression:

$$\Psi_o(x,y;z-z_i) = \mathcal{F}^{-1}\{\mathcal{F}\{\Psi_i(x,y;z_i)\} \cdot \mathcal{F}\{h(x,y,z)\}\} \tag{3.4.7}$$

where

$$\mathcal{F}\{h(x,y,z)\} = H(f_x, f_y, z) = e^{jkz\cdot\sqrt{1-(\lambda f_x)^2-(\lambda f_x)^2}} \tag{3.4.8}$$

If no distance optimization is desired, the entire free-space propagation filter in the frequency domain can be precomputed once and later used as in a normal signal processing task. On the other side, complete correction of both the distance and the energy (wavelength) would become extremely costly because each time the entire frequency domain filter should be generated. Limiting the optimisation to a distance regression is a good compromise because at least the frequency variant structure of the filter itself can be precomputed; indeed, the impact this trick can have during the experiment is huge. The filter and the data have to be prepared in a way that also minimises the count of *fftshift* operations, precomputing any operation wherever possible.

The use of an autograd environment allows to easily define a loss function that can be composed by the sum of many fundamental expressions. The main loss term is the one presented in eq. 3.4.6 which in this form already allows the direct computation of the gradients with respect to the object O, the probe P and the distance z. In the following section we will also show how the gradient with respect to the positions shifts can be obtained.

Regularisation A critical component in the optimization theory is the regularisation, a component which penalises *ad hoc* solutions in the parameters space: the concept of "overfitting" in the Machine Learning community can be transposed in the field of CDI explaining that a good solution is the one that represents better the real features of the objects and does not put up a pixel value just to accommodate the diffraction pattern dissimilarity measure. Due to the fact that any inverse problem reconstruction uses an "unsupervised" metric, a loss value can decrease up to a certain value, regardless (in some measure) to the real “correctness” of the results; then eventually the optimizer may remain stuck in a local minimum, but being a one-to-many problem, the same loss value (i.e. the same depth value on the multi-dimensional parameter-loss surface) can be obtained for a different and maybe better configuration of the parameters. This wrong solution with a relatively low loss value will be full of artifacts. Then regularisation helps in determining the right solution just by decreasing the flexibility of the model to adapt to the data. The typical approach consists of adding to the loss function the L1/L2 norm of the full parameters tensor. The physical meaning of this choice is a mere energy constraint. In this work, the adopted regularisation strategy is to penalise the object parameters that are greater than 1, calculating the L2 norm of this subset. Again, the physical interpretation has its roots in an energy constraint, making the algorithm to avoid the representation of a light-generating object.

Differentiable programming approach

The optimisation pool defined in equation 3.4.6 is extremely large and is composed by the tensors O, P, z and the list of positions v_N. The gradient of the loss term $\mathfrak{L}$ has to be calculated for each of these parameters. Being O and P complex variables, it is necessary to rely on the complex derivative: the gradient with respect to these two quantities is the composition of two functions, obtained computing the partial derivative of $\mathfrak{L}$ with respect to the real and imaginary part of each differentiation variable.

In the case of handmade symbolic computation, the easiest way to proceed is to exploit the Wirtinger calculus [141] because its formalism allows treating the complex function as if it were a real one, greatly simplifying the hand calculation. In the case of an automatic differentiation environment, no gradient expression has to be provided manually, because the acyclic graph describing the forward model is computed dynamically upon the execution of each expression (the so-called "eager execution"). On the other side, automatic differentiation is different from numeric differentiation [142] even if a numerical output is generated; a handmade symbolic computation provides an expression that can be calculated with the highest numerical accuracy for a defined value of the independent variable; similarly, a Computer Algebra System (CAS) can also generate a symbolic expression through symbolic computation, but often the output results in "expression swell" [143, 144, 142]. Many are the exploited mechanisms to compute the gradients, but in the library of choice (PyTorch [108]) the main idea lies in the fact that when a mathematical expression is evaluated, each temporary result constitutes a node in the computational graph that records the story of the expression, from the input variables to the generated result. The gradient is calculated simply following backwards the graph (backward mode differentiation [142]), from the results to the input variables, just applying the chain rule to the gradient of each nuclear differentiable operator [143] ("operator overloading" mechanism [142, 144]). To do so, each object (tensor) contains not only the actual data but i) the handler to the last operation that produced that variable, just like a linked list and ii) a second data array containing the gradient. This particular data type is called "*doublet*" in the literature [143].

Automatic differentiation has been studied since the 60s, essentially for optimisation applications. With the increase of the computational power, it becomes applicable also for highly dimensional problems like the one encountered in DL, leading to standardized, highly tested and optimised implementations that can run on GPUs. However, DL autograd tools are not conceived to work with complex numbers, then a basic complex library needs to be written; the natural way to use complex numbers is to just add an extra dimension to each 2D tensor and incorporate the real and the imaginary part in the same object, basically duplicating the number of actual variables. Automatic backward operation is completely legit for this kind of custom made datatype and the resulting gradient is simply

$$\operatorname{grad} f(\mathbf{x}) = 2\nabla_x^* f(\mathbf{x}) = \frac{\partial f}{\partial \mathbf{x}^*} = (\frac{\partial f}{\partial \mathbf{a}} + \jmath \frac{\partial f}{\partial \mathbf{b}}) \tag{3.4.9}$$

where $\mathbf{x} = \mathbf{a} + \jmath\mathbf{b}$, and $\mathbf{a}, \mathbf{b}$ are the real and imaginary parts of the differentiation vector variable. The actual gradient is represented in the same complex data type of the variables, made of a real and an imaginary part. As can be seen in eq. 3.4.9, the result of the automatic differentiation can be written in the Wirtinger formalism just as the derivative with respect to the conjugate of the differentiation variable (except for the constant), that is the typical gradient expression exploited for the optimisation of complex quantities.

Spatial transform layer

If writing a mathematical expression in the parameters O, P and z directly permit the gradient generation, the same does not hold for the spatial shift correction defined onto an integer (discrete) sampling grid. In ptychography, for each tuple composed by the diffraction pattern and the corresponding shift vector, the correct area from the object has to be isolated. This cropping operator can be implemented by simply defining a sequence of indexes that crop the input array. Thanks to modern algebra libraries (Numpy, Matlab, etc.), this method is computationally extremely fast (pointers arithmetic) but the main limitation lies in the fact that it is not differentiable, then no positions gradients can be generated by backpropagating a loss function in which integer indexes appear; integer differentiation is nonsense. In the following, an approach borrowed from the Deep Learning research community is then explored. One of the main point of this research is indeed to propose a viable method to position refinement in an autograd environment, as all the previously mentioned techniques cannot be employed within this framework.

Convolutional neural networks are not invariant to geometric transforms applied to their input. To cope with this problem, a Spatial Transform layer [145] is introduced; the learnable model is trained at inferring the spatial transformation that, applied to the input feature map, maximises the task performance metric. For each input, this layer is responsible for applying the transform, instructed by the transform parameters dynamically chosen by a secondary localisation network. This approach greatly improves classification task performances (e.g. in face recognition [9]). This technique is inspired by texture mapping in computer graphics.

In this work, two are the raw components used: a grid generator and a grid sampler. The first component transforms a regular input coordinate grid (x_i, y_i) (what in Numpy or Matlab is the output of "*meshgrid*") in an output coordinate grid (x_o, y_o), by applying an affine transform to the former; this affine transform generally has 6 degrees of freedom (and 6 parameters), but for what concerns this work, only the 2 translation terms (t_x, t_y) are optimised. The scaling factors (s_x, s_y) are constants used for cropping the central region:

$$\begin{pmatrix} x_o \\ y_o \end{pmatrix}_j = \begin{pmatrix} s_x & 0 & t_x \\ 0 & s_y & t_y \end{pmatrix}_j \begin{pmatrix} x_i \\ y_i \end{pmatrix}_j \tag{3.4.10}$$

Then for each *jth* diffraction pattern and each *jth* shift-vector (t_x, t_y) the *jth* output grid (x_o, y_o) is generated. The grid sampler takes the generated affine grid and outputs a sampled version of the object. Following the formalism introduced in [145], the cropped portion of the object V is generated from the entire object U by:

$$V_i = \sum_{h=1}^{H} \sum_{w=1}^{W} U_{w,h} \cdot k(w - x_{i,s}, h - y_{i,s}) \tag{3.4.11}$$

A well-done sampling (with an anti-aliasing filter) adds a regularisation term that can help during the optimisation. The bilinear sampling exploits a triangular (separable) Kernel which does not present dangerous overshooting artifacts.

$$K(w - x, h - y) = K_1(w - x) \cdot K_1(h - y) \tag{3.4.12}$$

where

$$K_1(t) = \max(0, 1 - |t|) \tag{3.4.13}$$

Figure 3.16 shows the structure of the proposed approach, applied to the ptychography framework. The same spatial transform is enforced on the two channels of the tensor that represent real and imaginary parts of the cropped region of the object.

Within this setup, gradient propagation is possible because derivative expressions can be calculated with respect to both the affine grid and the output pixel values [145].

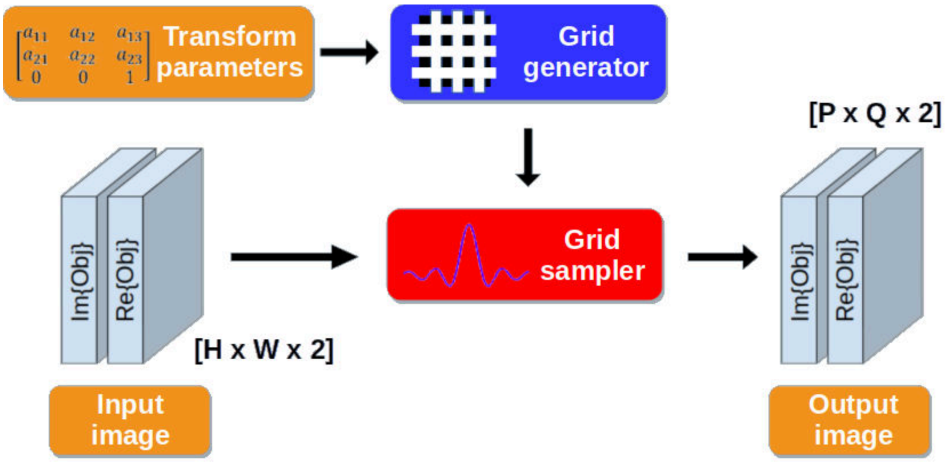


Figure 3.16: Schematics of the differentiable components used in our method. For each positions vector, a grid generator takes as input the corresponding shift transform expressed in an affine transform formalism; the sampling grid is then generated using this information. The object is then sampled at each coordinate defined by the sampling grid, producing the correct cropped region of the object.

3.4.1 Simulation Experiments

In this subsection we will describe the results of many simulated experiments. Firstly we will describe a baseline ptychography reconstruction within the non-linear optimisation approach. Then we will add to the optimisation pool one parameter at the time, in order to discuss how it is retrieved, and finally all the parameters are regressed during the ptychography reconstruction. The experiments are carried out by using synthetic datasets (based on images presented in chapter A) that are discussed in the following section.

Dataset creation

In its simplest usage modality, the reconstruction algorithm is designed to reconstruct just the complex object and the complex illumination, defined on the same plane. A ptychography simulator (which will the topic of a future publication) is used to generate the diffraction patterns, using the forward model presented before. To test the algorithm in many conditions, six datasets have been created using a set of test images that are well-known within the image processing research community (Lena, Cameraman, Barbara, WestConcordphoto, Tiffany from the Kodak CD, Bridge, Chelsea, Astronaut, Bacteria). Each dataset differs from the others for: i) a different set of images representing the object magnitude/phase; ii) a different set of positions; iii) different values of the propagation distance, ranging from z=0.0065 m to z=0.2 m. The two images for the magnitude and phase of the illumination on the sample remain constant among the set, as well as the pixel size of about 1 um and the wavelength of 1 nm. Each image has been contrast stretched to use the full unitary dynamic range, that for the transparency magnitude is [0, 1] and for the transparency phase $[-\pi, \pi]$. This configuration constitutes difficult datasets for the phase retrieval routines, with large black areas, deep borders etc. Due to the abundance of the tests that have to be run, the resolution for each diffraction pattern is limited to 128x128 pixels in order to reduce the computation time. The overlap factor is also kept constant at around 70%, producing a scan pattern of 11x11 positions. A known random jitter is added to the ideal grid-based scan pattern to avoid the well-known "raster scan pathology" [38]. The resulting total computational box (and then the total object size) has a field of view of around 512x512 pixels. Each dataset is saved as a multi-page tiff file containing the diffractions and an array of positions. Other options permit the generation of an HDF5 file which includes the diffractions, positions and parameters or a ptyd file that is compatible with the advanced ptychography reconstruction software PtyPy [146].

A fake reconstruction obtained stitching each object region is shown in Figure 3.18, while Figure 3.19 shows the complex illumination probe in its magnitude/phase and an example of near field diffraction pattern.

Within the context of this section, the term "epoch" is used interchangeably with "iteration", for clarity reasons, because in the context of numerical optimisation an epoch ends when all the available data have been used once.

Baseline reconstructions

In this section, the baseline results are presented, where no correction is needed and just the plain optimisation of the object/probe arrays is performed. In this case, reconstruction quality can be analysed following two approaches: observing the "unsupervised" loss measure that directly guides the optimisation process, or calculating a truth-aware similarity metric; obviously, the latter method can be applied for simulated data only.

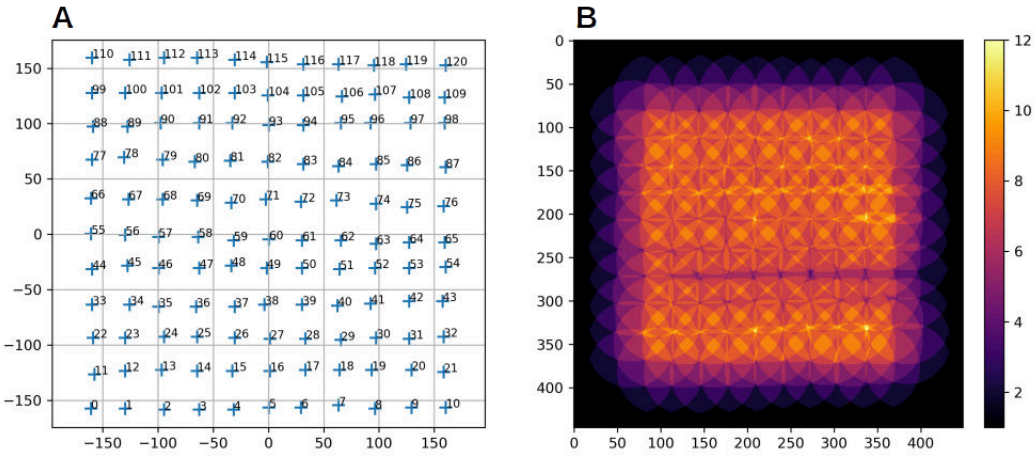


Figure 3.17: Panel A shows an example of the sampling positions at which the simulated object is scanned. The ptychography experiment is carried out overimposing the illumination function onto the sample, at the positions described by the panel A. As in [52] a sampling density mask can be extrapolated.

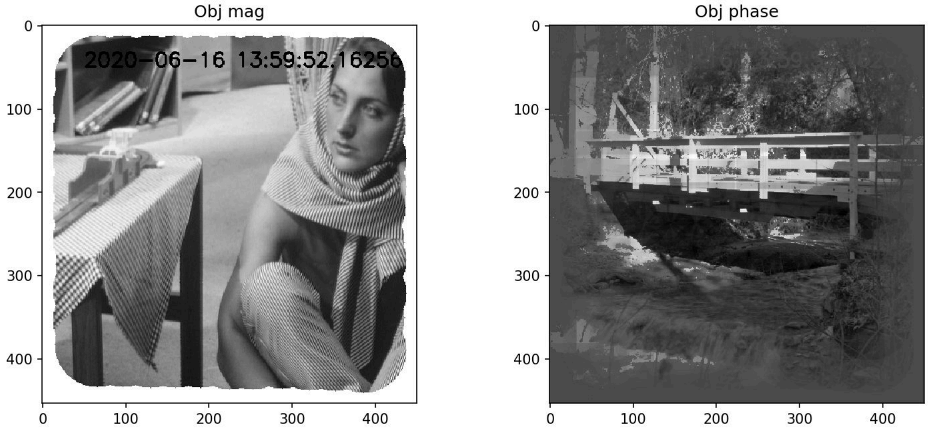


Figure 3.18: Images from dataset 5: "Barbara" is used as the object magnitude, "Bridge" as the object phase, "Chelsea" as the illumination probe magnitude and "Astronaut" as the illumination probe phase (chapter A).

Having many datasets composed of a variety of objects and different parameters allow to observe average performances. Figure 3.20 panel A shows the unsupervised measure that takes into account the dissimilarity between the estimated intensity and the recorded diffraction pattern: in batch-based optimisation, the final loss value for each epoch consists of the sum of the losses accumulated on each batch, that in turn are calculated summing the

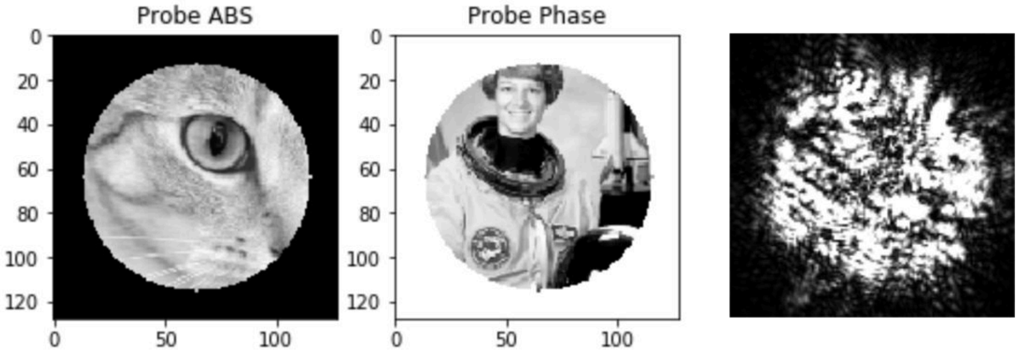


Figure 3.19: The probe magnitude and phase used for the simulations. The third subfigure show one of the simulated diffraction patterns belonging to the object in fig. 3.18

samples' MSE. If for each epoch the complex object is saved, a truth-aware measure can be accommodated. In the literature, typically the MSE between the ground truth object and the current estimate is calculated; a second popular choice is the Fourier Ring Correlation (FRC), a measure taken from the crystallographer's community [147]. In this paper, the structural similarity SSIM [118] is used, due to its better performances in the terms of perceptive quality assessment (PQA). Figure 3.20 panel B shows the average envelope of the Structural Similarity with respect to the running iterations.

Looking at the graphs in figure 3.21, it is apparent how in the first 80 iterations, on average, the results have gained most of the quality, as shown by a massive and rapid decrease of the loss and a corresponding increase in the SSIM index. What is supposed to be the main limiting factor which prevents the loss from approaching zero (and the SSIM approaching 1) is the presence of low transparency areas (defined by black pixels) that limit the extraction of phase information.

Cross-talk between magnitude and phase is inevitable, especially in the case of antipodal values in the magnitudes of the transmission function: if no light passes, no information on the phase can be gathered.

A second cause for a non-perfect SSIM value is the post-processing (contrast stretching) used on the reconstructed images: due to its formulation, SSIM is only moderately sensitive to absolute value differences, but this sensitivity can not be reduced to zero. The high variance in the SSIM is also due to the fact that, for a virtual detector closer to the sample, the convergence of the algorithm is faster because the numerical propagation produces fewer artifacts. Due to the sampling problem, numerical propagation routines have to be ad-hoc chosen for a specific set of physical parameters. Here the angular spectrum is used for a large range of propagation distances.

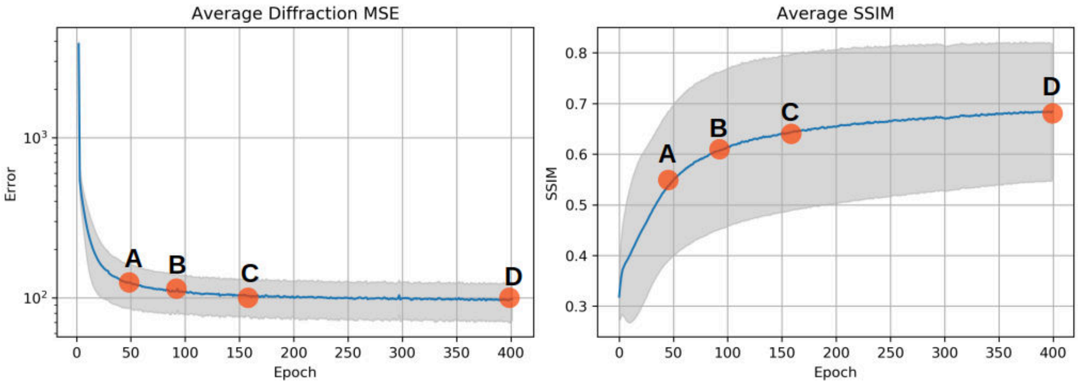


Figure 3.20: (Top left) Evolution of the average MSE between the estimated diffraction pattern of the detector and the measured data. (Top right) Evolution of the SSIM quality index.

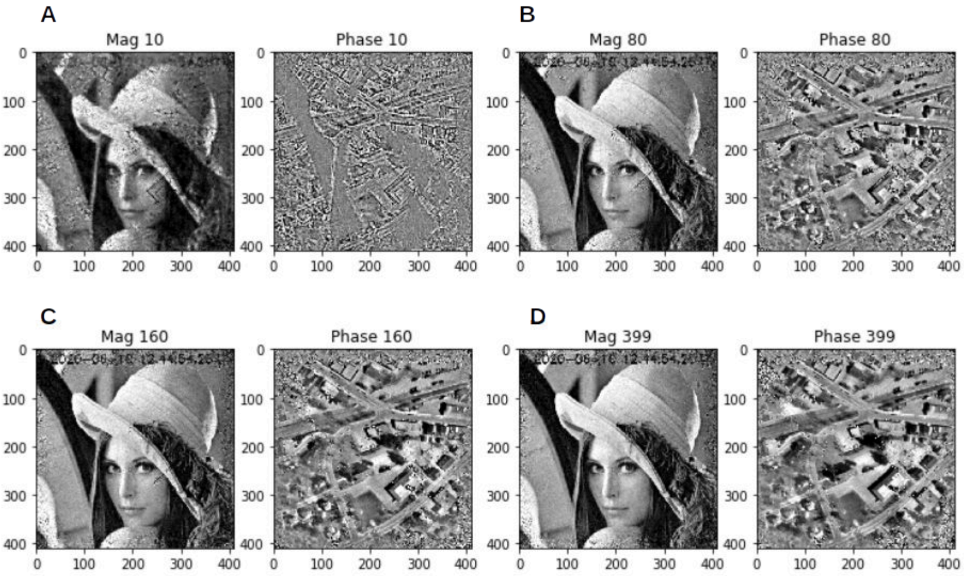


Figure 3.21: Object magnitude and phase reconstruction at 10, 80, 160, 399 iterations, defined respectively by points A,B,C,D on the top graphs. Images chosen for the transparency function are “Lena” and "Westconcordphoto".

Distance regression

In this section, the effect of the propagation distance regression is discussed; as in the previous test, a ptychography reconstruction is carried out in an optimisation framework. No prior information is given, apart from the measured diffractions and the nominal positions that are part of the typical ptychography dataset. The training parameters are as usual the complex object array, the complex illumination array and a floating-point number representing the propagation distance, then just one more variable is added to the optimisation pool. In this test, the process starts from an initial incorrect value that differs from the ground truth by 30%. As it can be seen in the first iterations of figure 3.24, if no correction is applied the optimisation is rich in artifacts, due to the limited invariance of the model to this parameter.

Fig. 3.22 (panel A) shows the effect of a wrong propagation distance estimate on the retrieved probe: speckle-like patterns typically appears.

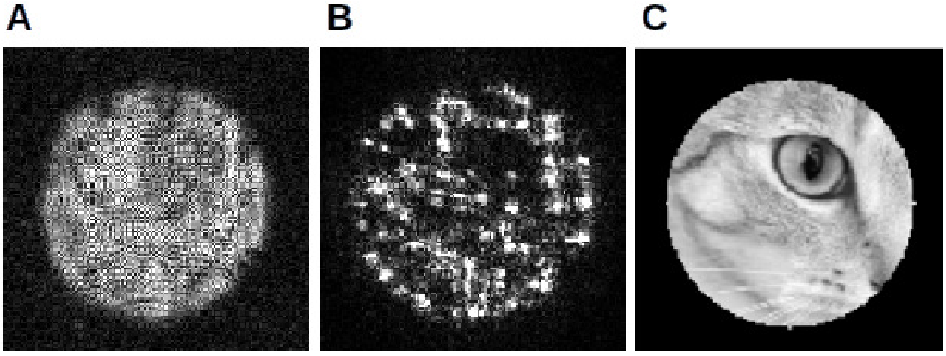


Figure 3.22: Typical artefacts in a probe retrieval procedure (simulated data) in the case of wrong propagation distance (panel A, speckle patterns) and no position refinement (panel B, cluster of dots) Panel C shows the real illumination (magnitude).

Figure 3.23 shows how the object and the probe are retrieved at several different epochs. The reconstruction starts at a $z = 0.07$ m, while the correct propagation distance is $z = 0.1$ m, denoted on figure 3.23 as a red horizontal line. In about 2500 iterations, the process converges to the neighbourhood of the true value.

Figure 3.23 shows the trend of our quality measures (diffraction MSE and SSIM) during the joint arrays/distance optimisation. When the distance is correctly inferred, the improvement both in the loss and in the quality metrics is huge, allowing for almost the complete removal of the crosstalk between magnitude and phase for both the object and the illumination probe.

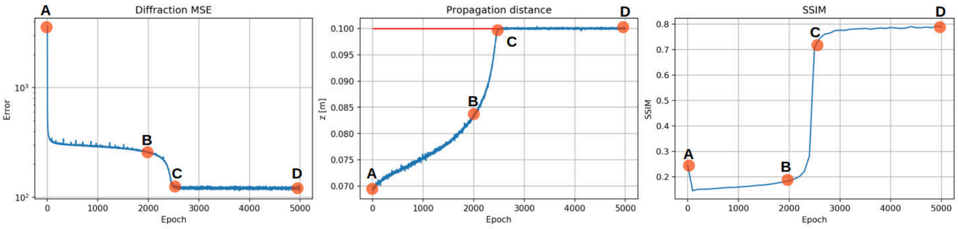


Figure 3.23: Reconstruction converge in terms of MSE (leftomost panel), the regressed propagation distance (center) and the Structural Similarity (SSIM) (rightmost panel). Observing the central panel it can be seen how in roughly 2000 iterations the correct propagation distance is retrieved by the optimisation procedure. For each curve the point A, B, C, D are marked and the relative reconstruction is shown in the next figure (Fig. 3.24)

Figure 3.25 shows the effect of many different initialisations on the overall reconstruction trend: the experiments suggest that a different initialisation affects the convergence speed but not the final accuracy and repeatability. For each initialisation ranging within ±30% of the correct z value, the error is bounded to roughly 10^{-4} m (figure 3.23 panel A), i.e. to a relative error of 0.1%.

Using simulated data in which just one perturbation is added at a time, the effects of the noise on the regressed distance is limited, because the optimisation process can automatically cope with it, exploiting the aforementioned redundancy of the information. Then, even if for smaller z the relative error tends to be higher, the resulting SSIM figures tend to be similar among the different datasets (with different z). The main cause for this behaviour lies in the fact that no significant loss changes are observed, then the algorithm is not able to move significatively to another point. The experiments confirm that the speed of convergence is faster for an absolute lower dynamic range in the needed correction: indeed, the optimisation of the dataset with the lowest z is much faster than the one performed on the $z = 0.2$ m dataset. Figure 3.25 panel B shows how the SSIM index changes with respect to the relative distance error: an SSIM of roughly 0.7 can be obtained from reconstructed objects within a relative error on the distance of about 5%. From this graph, it can be also seen how the convergence speed is lower for large values of z, the points of which span a large area of the graph, defining a hyperbolic-like curve. Convergence speed can be eventually increased with per-dataset criterion tweaking the learning rate of the optimizer; however, in these tests a value of $1e^{-3}$ has been chosen for all the dataset.

Observing the convergence plot in figure 3.23 it is worth noting that while

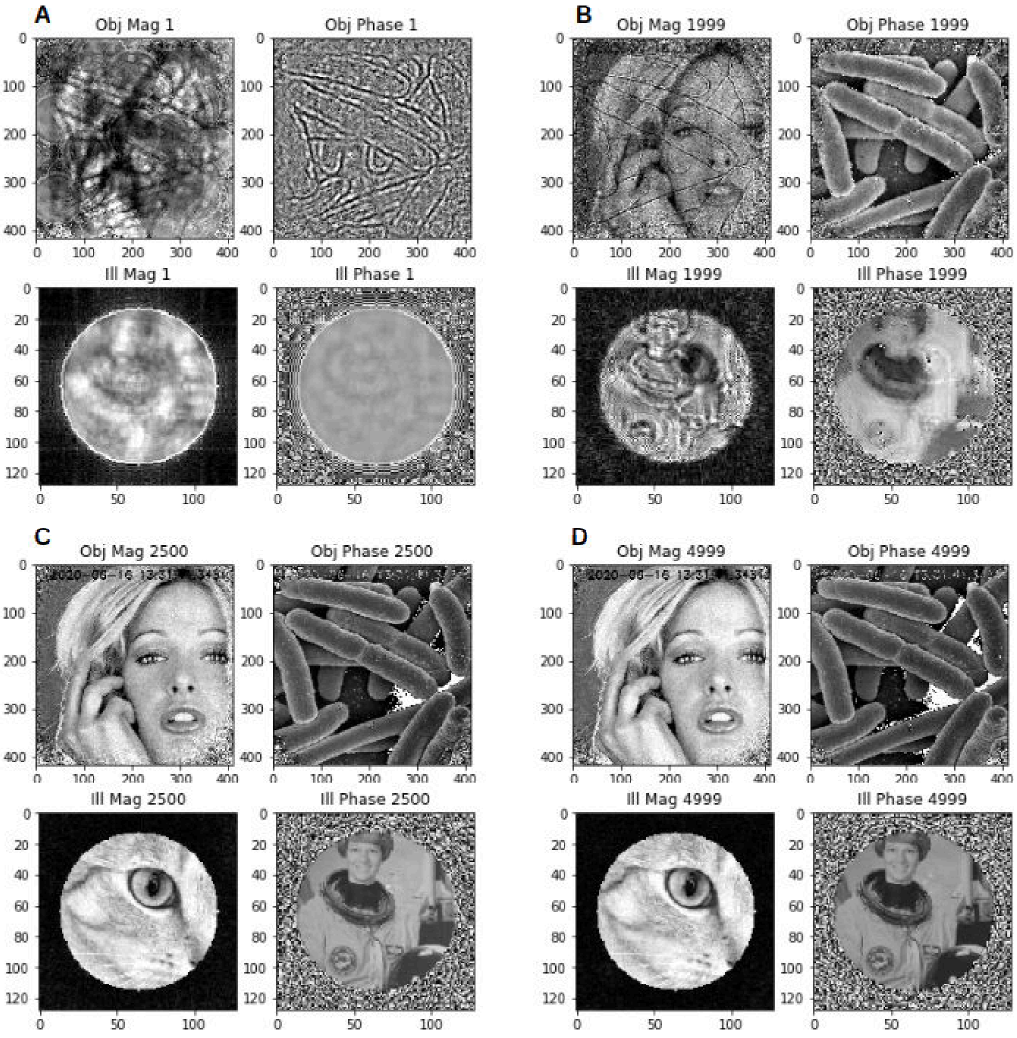


Figure 3.24: Object magnitude and phase reconstruction at 10, 80, 160, 399 iterations, defined respectively by point A,B,C,D on the top graphs. Images chosen for the transmission function are "Tiffany" and "Bacteria".

initially the distance update steps are small, then, as the reconstruction progresses, an acceleration in the convergence is observed. This is a clear effect of the fact that in the first iterations, a fuzzy object can obviously not provide reliable information on the distance gradient, which is noisy from epoch to epoch. Then, the gradient becomes more and more informative, making the average among recent epochs each time more consistent; the update step is then larger. This behaviour can be directly linked to the optimizer of choice, Adam [135, 148], which exploits an exponential moving average of the gradient term, at the cost of an increase in memory consumption.

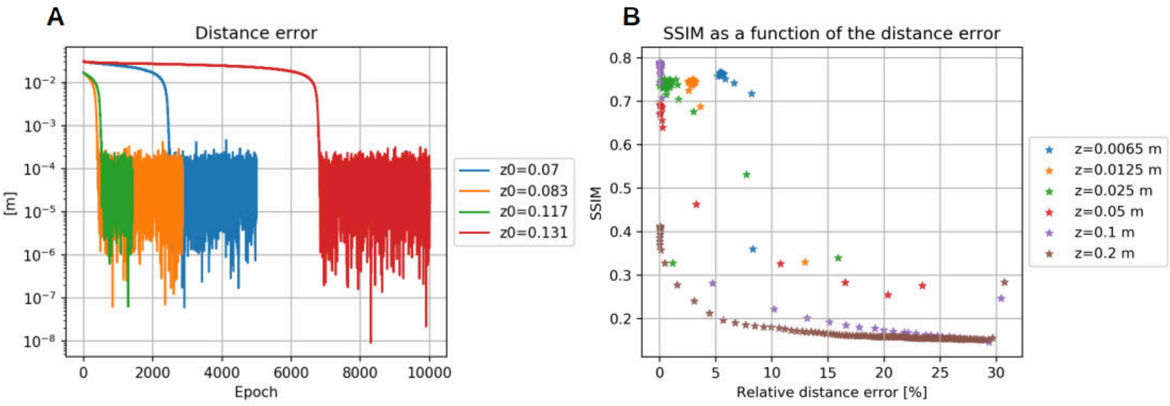


Figure 3.25: Panel A: Regressed distance trend for the reconstruction on the same dataset but with different initialisations z_0. The final error, in this case of about 0.1% of the distance is invariant to the initialisation. The same does not hold for the convergence speed that is significantly dependent on the range of correction. Panel B: performance comparison on all the six datasets; for lower propagation distances the reconstruction process tolerates a relative error on the regressed distance of about 5-10% for the same high SSIM index of 0.7. Instead, for large z, the optimisation requires a fine tweak on this parameter

.

Positions refinement regression

In this section, the position refinement algorithm is tested on six different datasets, whose creation is described in detail at the beginning of this section (3.4.1). The algorithm allows to perform a continuous position refinement while reconstructing the object in an autograd based optimisation process. This task is accomplished without increasing the number of DFTs; $2N$ variables are added to the optimisation pool, where N is the number of the recorded diffraction patterns.

Starting from each created dataset, a random jitter characterized by a standard deviation of 10 pixels is added to the true position coordinates. The resulting distance error is then on average around 12 pixels, with some peaks (within 1 standard deviation) of roughly 20 pixels and few sparse outliers. Incidentally, this jitter has not to be confused with the (known) one purposefully added during the creation of the dataset in order to avoid the "scan pattern pathology" [91, 38]. A new, perturbed dataset is then created and the reconstruction starts with these data, with positions vectors added to the optimizer parameters pool. This means that gradient will be generated also for these variables. Figure 3.26 shows how detrimental can be a position error for the total object box: it can be seen that the limits are completely changed.

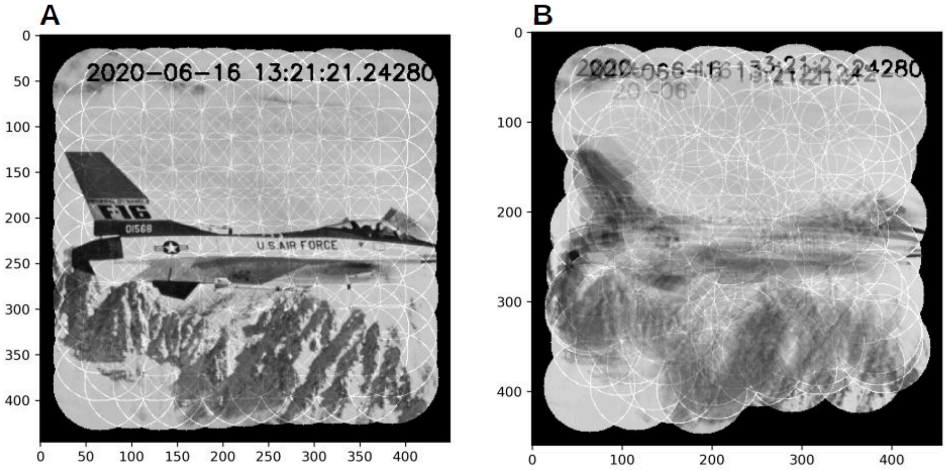


Figure 3.26: Scan positions errors change the supposed maximum size of the computational box. Panel A shows a canvas composed by stitching together different ROI of the transmission function (the area sampled by the illumination probe and that will generate the scattering). In the first case the object can be reconstructed, while with position errors (panel B) transparency information is smeared.

Without regularisation In 2000 iterations (Fig. 3.27, the algorithm is able to refine all the positions, converging to an error of about 2.5 pixels, with a great increase in the quality of the reconstruction. As in this test the exact positions are known, it is possible to calculate a metric for the performance evaluation of the position refining routine, but there is an important difference with respect to the previous cases: a wrong initial position estimate inevitably changes the form and the size of the computational box enclosing the object, because, during the algorithm initialisation, a canvas is created from the min-max shift reported by the positions data, as shown in Fig. 3.26.

Then, being the SSIM index the output of an element-wise operator, the object under test or the reference object have to be resized accordingly.

With regularisation Observing the convergence plot in figure 3.30, the optimisation process is more homogeneous and faster than the one devoted to the propagation distance regression: the update step tends to be bigger at first and then it scales down; this is in contrast with the trend of the distance

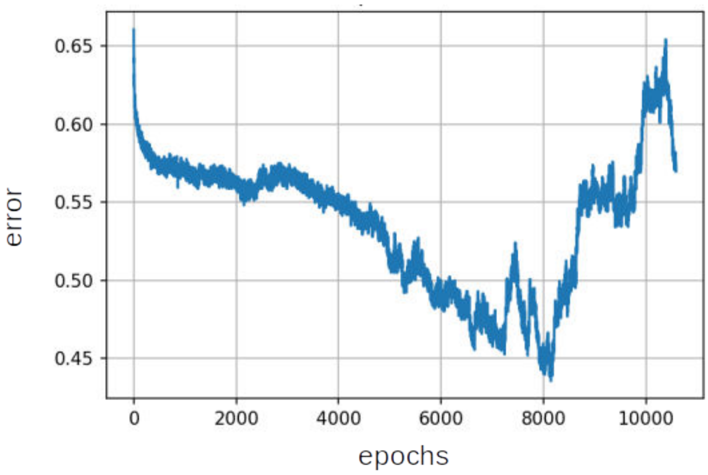


Figure 3.27: Reconstruction error norm in the case of no regularisation. Initially the total error decreases but eventually starts to rise again due to local minima in the loss function. If for most of the positions the error is corrected, for the remaining few, especially at the borders, the correction can assume wrongly large values, and this makes the position error norm to rise. Position error norm alone is not enough to check the performances of the algorithm, that is why the following distribution analysis is also taken into account.

regression, discussed in the previous section. This means that the effects of the cropping operator parameters on the loss are profound, already visible at the start of the optimisation process. Figure 3.29, compared to figure 3.31, shows the importance of the regularisation component, whose inclusion in an autograd environment is as simple as adding the loss component to the main loss. If no control over the regressed position is enforced, there are some difficult cases in which possible large errors cause the algorithm to stall at significantly wrong positions or ease the divergence of some position vectors. This behaviour is spotted by a cumulative position error norm that initially decreases but then raises again, due to some position vectors that assume a large value. This implies that the norm of the position error is not enough to characterize the optimisation process, and a full error distribution has to be considered instead.

To cope with this problem, we add a-priori knowledge about the error to the optimisation algorithm: indeed, we know that the initial positions are rather accurate so that the correction should not be large. As expected, no more bivariate error distributions can be observed during the reconstruction; if the obtained distribution gets narrow this means that the algorithm successfully converges to a value closer to the ground truth.

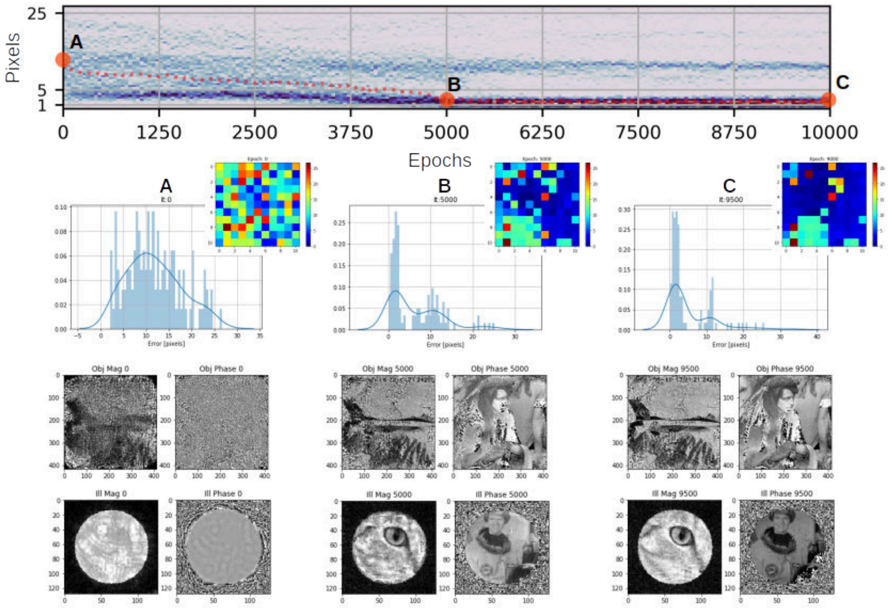


Figure 3.28: Position refinement of the 11x11 acquisition, *without* regularisation: the optimisation reduces the spread of the initial distribution error (A), but having no control on the regressed values some positions are erroneously shifted apart or are not refined in a reasonable amount of time; a bivariate error distribution builds up (B, C). The top bigger graph shows in a waterfall fashion the error distribution. At a given epoch (thus fixing the x coordinate) an histogram can be visualised, such as for panel A, B and C. For each epoch A, B and C the error in all the 11x11 positions is shown by the pixmap (upper-right panel); blue means small error, while red, large error.

Figure 3.30, figure 3.31 and figure 3.32 show the evolution of the position during the reconstruction. In figure 3.32, at iteration 0, current vectors "+" are initialised at the given wrong positions "+". At the end of the reconstruction, the position error norm is greatly reduced, thanks to the iterative position refinement that brings the vectors "+" towards the ground truth positions "+". The effect of the regularisation can be directly observed by comparing figure 3.29 with figure 3.32: the number of green crosses left uncovered by the blue ones is larger without the regularisation. A second effect can also be seen comparing figure 3.28 and 3.31: the convergence speed has improved by a factor 20, as the knee moves from 5000 to 200. For this reason, figure 3.31 shows a simpler graph panel 1 with just the median and the mean of the error distribution, and only the first 1000 epochs (the area highlighted by the green box) are expanded in panel 2 of the same figure, showing the entire error distribution.

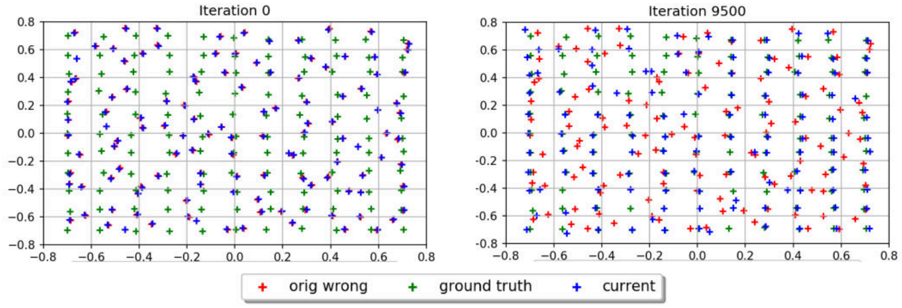


Figure 3.29: Position refinement without position regularisation: start/end (epoch 0/epoch 9500) positions for each projection. Position vectors (blue cross) move from the wrong initial estimates (red crosses) toward the correct values (green). Coordinates are normalized.

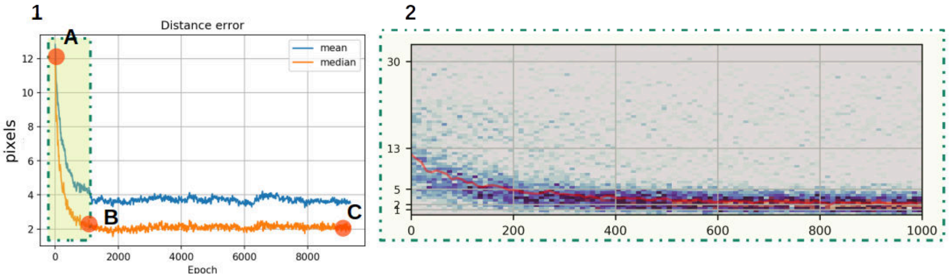


Figure 3.30: Position refinement *with* regularisation: Panel 1 shows the median and the mean estimates of the position error during the reconstruction. Similarly to the previous case, the optimisation reduces the spread of the distribution error, and the regularisation guides the solution towards small values, and no more clustering appears. The optimisation is also faster than as in the previous case: for this reason, the full error distribution (panel 2) is shown only for the first 1000 epochs (the highlighted region of the graph).

Finally, figure 3.33 shows how the SSIM index changes during the reconstruction of each dataset. The actual values of the SSIM are meaningless due to the resize and have been thus normalized. The knee at an error of 4 pixels corresponds to the one that can be observed on figure 3.23. The convergence speed can be inferred by observing how sparse are the points for a region of error larger than 6 pixels.

The quality of the object reconstruction can be increased by restarting the optimization with the position found during a previous optimization attempt. At this point two strategies can be exploited: i) the reconstruction can be started as is, just from a new initialisation with the updated positions assumed to be true; or ii) during a new reconstruction the position can be

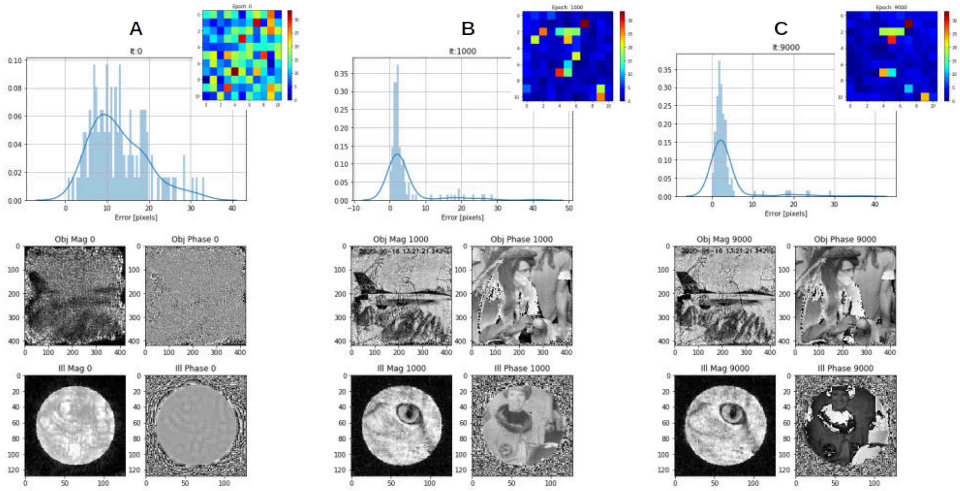


Figure 3.31: Position refinement *with* position regularisation: the optimisation reduces the spread of the distribution error, and the final error distribution shows an error that spans below 5 pixels.

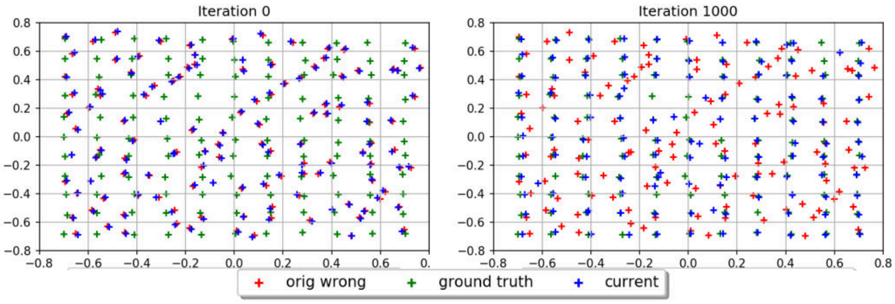


Figure 3.32: Position refinement *with* position regularisation: in the graph at first glance there are just red and blue crosses, indeed the ensemble of current position vectors (blue) lies on top of the ground truth (green), exposing just the wrong position.

refined with a lower learning rate, allowing for an eventually deeper descent into the gradient.

Combined full optimization

In this section, the algorithm is tested against the contemporary presence of perturbations both on the propagation distance and on the positions. The analysis is the same for the previous cases. For each dataset, the propagation distance is initialised to a value corrupted by an 30% error, while each

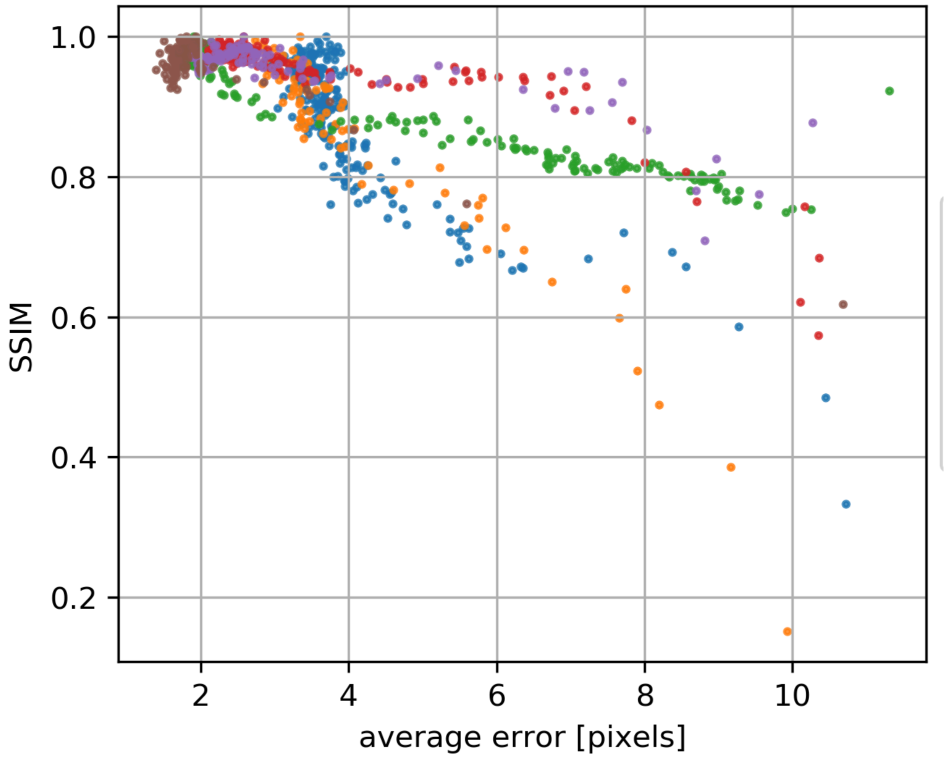


Figure 3.33: Normalised SSIM as a function of the median error: convergence speed in the various phases can be tracked by observing the sparsity of the points for each error value. Each colour is a different dataset.

position is perturbed by random jitter with a variance of 10 pixels. Both the propagation distance and the position vectors are added to the optimisation pool.

As can be seen in figure 3.34, in around 500 epochs both the positions and the distance have recovered more than 90% of initial error. The use of the grid sampler with its inherent regularisation action increases the convergence speed also for the distance. As in the previous graphs, an enlarged version of the position error distribution for the first 1000 epoch is displayed in figure 3.36.

The median of the position error settles again at about 2.5 pixels, denoting the limits of the proposed method.

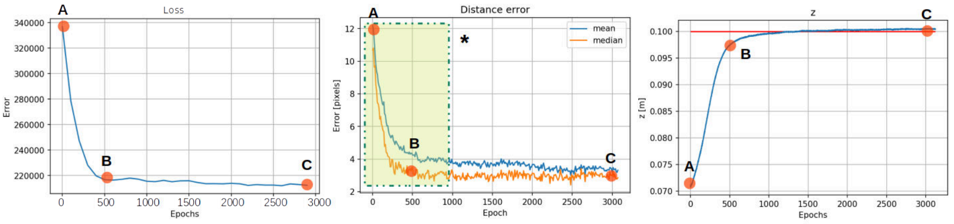


Figure 3.34: Normalised SSIM as a function of the median error: convergence speed in the various phases can be tracked by observing the sparsity of the points for each error value.

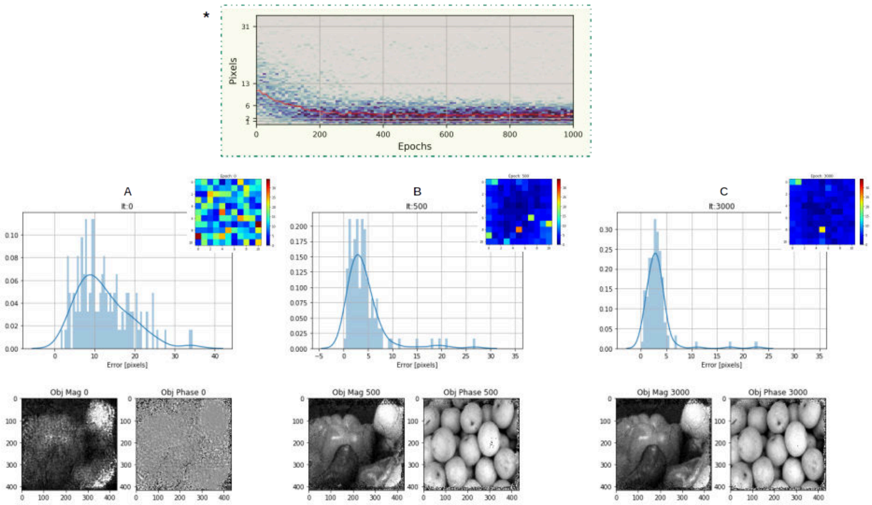


Figure 3.35: Normalised SSIM as a function of the median error: convergence speed in the various phases can be tracked by observing the sparsity of the points for each error value.

Ablation study

The expression “Ablation study” [149] denotes the action of removing components from a system in order to study the individual effects that contribute to the complete arrangement. Having its roots in the field of experimental neuropsychology, where animal body parts are displaced, ablation studies have been recently adopted in the current Deep Learning/Signal Processing literature to describe the act of sequentially removing layers or a functional element from the entire network/algorithm, in order to gain a better understanding of the system as a whole [150]; this is especially important in the Deep Learning field, where researcher are struggling yet to accurately understand how fine variations of the hyper-parameters relate to the final output. The aforementioned analytic approach perfectly fits in the context

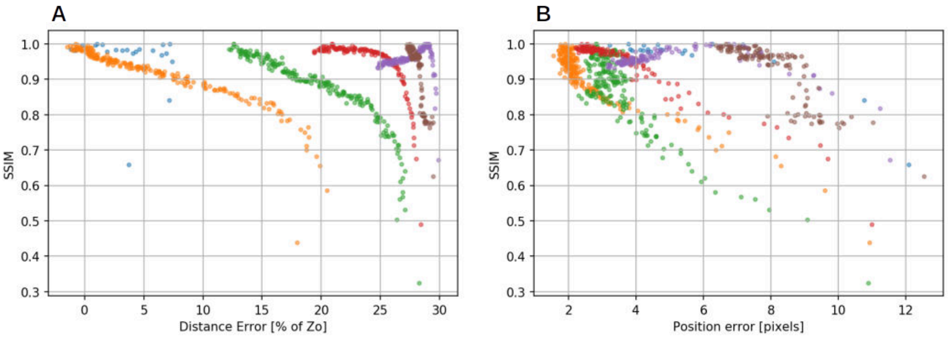


Figure 3.36: Normalised SSIM as a function of the median error: convergence speed in the various phases can be tracked by observing the sparsity of the points for each error value.

of this chapter, where distinct components are devoted to several model corrections.

For each configuration that differs from the baseline approach, the results with or without the correction are compared, using the ratio of their corresponding SSIM indexes. This "gain" provides an insight into how mandatory is the correction in a given setup configuration. During each test, the six datasets described in section 3.4.1 are used; as usual, the object and the probe are saved every 100 iterations. To limit the duration of the tests, the number of iterations is limited to 2000. The best value of the SSIM is considered for both the correction and the no-correction scenarios. In Figure 3.37, the first two bars present the performance gain on datasets perturbed with one error at a time. In the case of the first bar, as already said, an erroneous propagation distance is given to the solver; as already observed previously the amount of gain rises for a dataset with a large propagation distance, because the distance correction assumes more importance when applied. When regularisation is added in an optimisation framework, the artificially added laziness in the data fitting procedure allows for a limited but existent invariance to some reconstruction parameters and that is why the low range of the gain bar is close to 1 (in the case of a small propagation distance). In the second bar, perturbation is applied onto the positions and it is clear how much this correction impacts the reconstruction quality, with a lower bound on the gain around 2. The last two bars report the performance gain if just one correction is applied in a doubly perturbed dataset (positions and distance). It is interesting to note that in a mixed approach the lowest minimum gain is obtained for the regression of the distance only, meaning that the importance of the propagation distance regression (a previously “forgotten” parameter in the literature) is crucial in a real dataset (where reasonably both errors do appear).

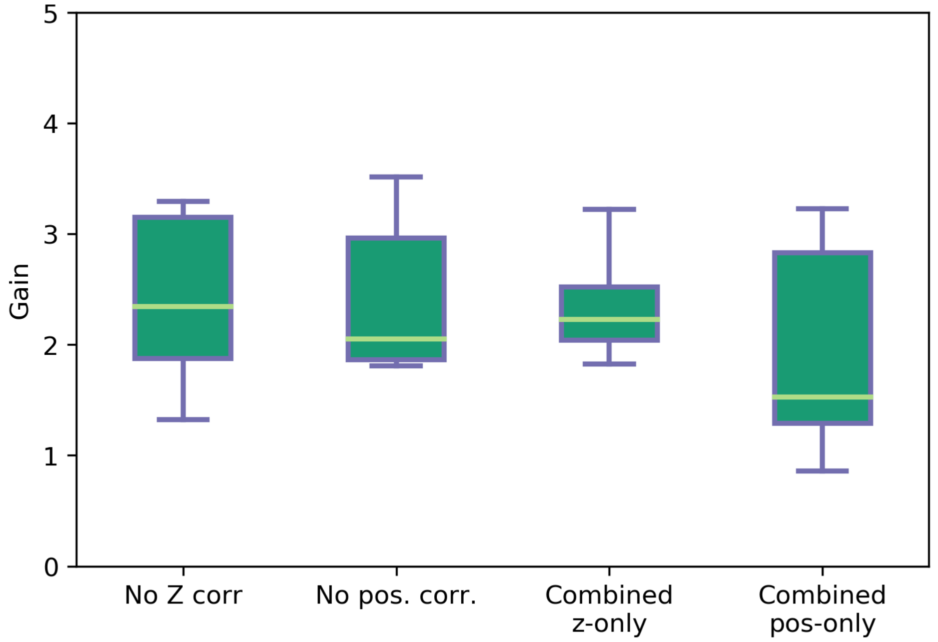


Figure 3.37: SSIM gain for many reconstruction scenarios; the first bar shows the performance gain obtained correcting the propagation distance if an initial wrong estimate is given; here the positions are correct. For the second bar, the opposite situation is depicted. Similarly, the last two bars describe the effect of just one correction in a dataset with wrong assumptions on both the distance and the positions.

3.4.2 Soft X-ray reconstruction

In this section, reconstructions obtained from a soft X-ray experiment are presented. The imaging experiment has been performed at the TwinMic spectromicroscopy beamline at the Elettra synchrotron facility. TwinMic can operate in three imaging modalities: i) STXM; ii) full-field TXM/CDI; iii) scanning CDI (ptychography). Clearly the latter (the one used for this work) is obtained by combining the optic setup of the second modality and the control of the sample stage from the first one.

Similarly to other ptychography experiments performed at the beamline (e.g. [51, 50, 151]), X-ray data were collected using a 1020 eV X-ray synchrotron beam [45, 152] focused with a 600 um diameter Fresnel zone plate (FZP) with an outer zone width of 50 nm placed approximately 2 m downstream a 25 um aperture that defines a secondary source. A Peltier-cooled charge-coupled device (CCD) detector (Princeton MT-MTE) with 1300x1340 20x20 um square pixels was placed roughly 72 cm downstream of the FZP.

According to Abbe theory [50], the limit of resolution, Γ, for coherent illumination is $\Gamma = 0.82\lambda/NA = 50$ nm. As pointed out in chapter 2.4, the resolution in Fresnel CDI is a function of the experimental geometry, namely the distance from the focal point of the FZP to the detector and the physical size of the detector itself, rather than the focusing optics. The scan pattern used for the acquisition is shown in figure 3.39.

From the ptychography configuration point of view, the situation is similar to the one presented in [99], where a point source (obtained by idealising the focus of an FZP through an OSA) illuminates the sample. Following the approach of [99], during the data analysis, the beam has been parallelized by using the Fresnel Scaling Theorem [18]. In the following experiment, as pointed out in section 2, the sample is considered sufficiently thin [38, 54] to be modelled by a multiplicative complex transmission function h defined by the expression

$$h(\bar{r}) = e^{j\frac{2\pi}{\lambda}[n(\bar{r})-1]t(\bar{r}))} = e^{j\frac{2\pi}{\lambda}t(\bar{r})[-\delta(\bar{r})+j\beta]} \tag{3.4.14}$$

where $\bar{r}$ is the planar coordinate on the sample plane, $n(\bar{r})$ is the complex refraction index and $t(\bar{r})$ a real

$$R^2 \mapsto R^1$$

function defining the local thickness. From the reconstruction-inferred $h(r)$ (the objective of the reconstruction), the magnitude map corresponds to:

$$\ln|h(\bar{r})| = \ln\sqrt{I(\bar{r})} = \ln\sqrt{I_o(\bar{r})} - \frac{2\pi}{\lambda}\beta(\bar{r})\cdot t(\bar{r}) \tag{3.4.15}$$

where I_o is the flat field intensity and I the sample intensity.

The phase map instead corresponds to:

$$\arg\{h(\bar{r})\} = \phi(\bar{r}) = -\frac{2\pi}{\lambda}\delta(\bar{r})\cdot t(\bar{r}) \tag{3.4.16}$$

If the properties of the material (beta and delta) are known, it is simple to infer its thickness and vice versa.

Diffraction data are acquired in the form of a 16-bit multi-page tiff file. The positions are directly acquired from the nominal positions provided to the control system of the sample-stage. A series of dark field images is acquired for the dark field correction. No white field correction is applied due to the fact that we are mainly interested not in the in-line hologram, but in the high angle scattering information present in the diffraction pattern; a normalisation (basically a division) would destroy this precious information, even if other pre-processing step involving the white field are under study. For the same reason, the area covered by the inline hologram lies only on less than half of the detector area

Figure 3.40 shows the output of many different algorithms applied on the same dataset: for the DM algorithm [38], the advanced and open-source "Ptypy" [146] software has been exploited; for all the other algorithms (epie [91] and rpie [153]), inhouse implementations of the algorithms have been written [1]. For Ptypy, the raw data are encapsulated in an HDF5-like format called ".ptyd", that contains both the positions and the diffractions. Ptypy can also load the setup parameters from there, but in order to execute a form of parameter hand-tuning, it is preferable to change them programmatically and visually inspect the results.

One of the samples used during the experiment is a human lung tissue, corrupted by asbestos fibres. Due to the fact that for each algorithm lots of iteration (10000) are needed for this dataset, in order to reduce the computation time each diffraction pattern has been scaled to 256x256 pixels, giving a resulting pixel size of roughly 45 nm (~5 x 9 nm = 45 nm) on a 846x847 pixel reconstructed image. Actually, the rescaling is currently necessary, as many GPUs do not have more than 5/8 GB of RAM. As in all the simulated experiments, for each diffraction pattern the correct value of the padding is inferred by the propagation routine taking into account the wavelength and the current propagation distance value. For biological specimens that are almost transparent, the phase reconstruction typically appears to be an easier problem than the correct magnitude reconstruction, as also reported in [92], so that one of the main indicators that the reconstruction is proceeding well is the evolution of the magnitude. We initially searched for best value in the dataset, by employing the DM algorithm, than we used that value to start the reconstruction with our algorithm.

Observing the quality of the results, the proposed method (Fig. 3.40) clearly surpasses all the aforementioned ones: figure 3.40 shows the red insets of the top left fibre, that is correctly reconstructed with the highest resolution only by the proposed algorithm; in a multi-mode DM and EPIE reconstruction lots of ringing artifacts are visible. RPIE [153] provides the best result among the typical reconstruction algorithms, with an object with a large field of view, which extends also in sparse sampled areas. The proposed method (fourth row) reconstructs the fiber and the cell at the highest resolution in both magnitude and phase (see Fig. 3.38). A second inset (green color) shows a texture in the phase reconstruction, where again the proposed method outperforms the others; cell structures are corrupted by fewer artifacts and are visible in their entire length.

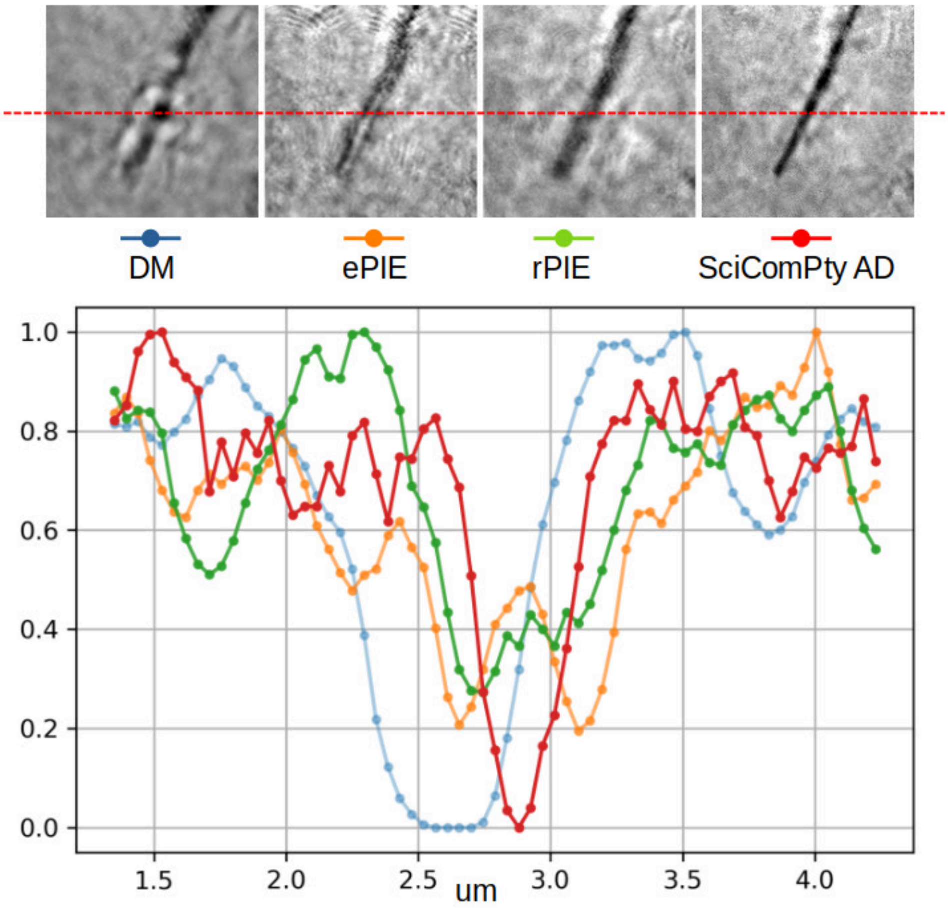


Figure 3.38: Line profile for each of the reconstruction in Fig. 3.40

The higher reconstruction quality can be attributed to the combined action of:

- the correction, e.g the automatic inference of a virtual propagation distance (Fresnel Scaling Theorem) of 0.24 mm instead of the 0.37 mm as obtained by an exhaustive manual search for the other algorithms
- and the use of a stochastic gradient descent based algorithm (Adam) in which the choice of a batch size represents a new hyperparameter that can be tuned in grain steps (then easily).

Then the proposed method provides a reliable reconstruction of both the magnitude and phase of the object transmission function.

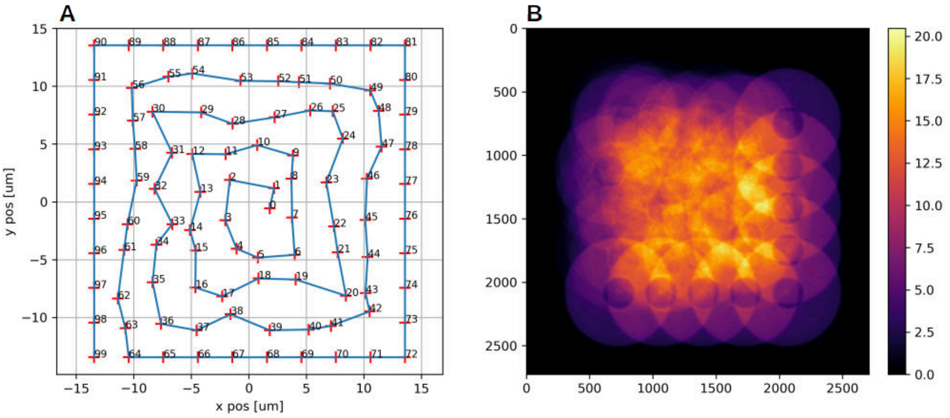
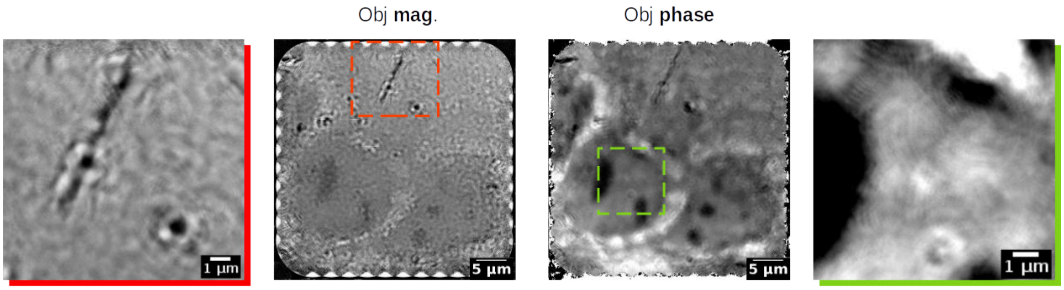


Figure 3.39: Panel A shows the scan positions in micrometres: the base scan pattern is a spiral in which the number of sample stage movements (and thus the backlash) is reduced, moving on a rectangular grid (then no round spiral line appears). A full rounded spiral will result in a larger position error. The corresponding sample density mask [52] is represented in panel B, which denotes a high sampling density especially in the center.

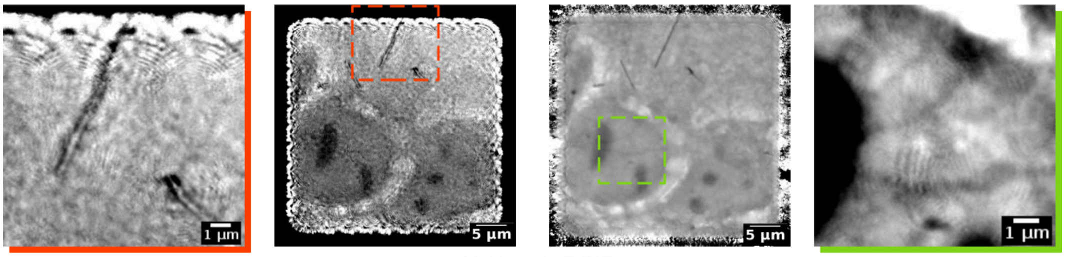
3.5 Summary

In this chapter the major setup problems have been discussed. Currently, the majority of the reconstruction artefacts can be directly attributed to that kind of errors. Three novel different methods are proposed to cope with positon errors, propagation distance uncertainty and multi-mode reconstruction. Each presented method has been implemented in the SciComPty software suite.

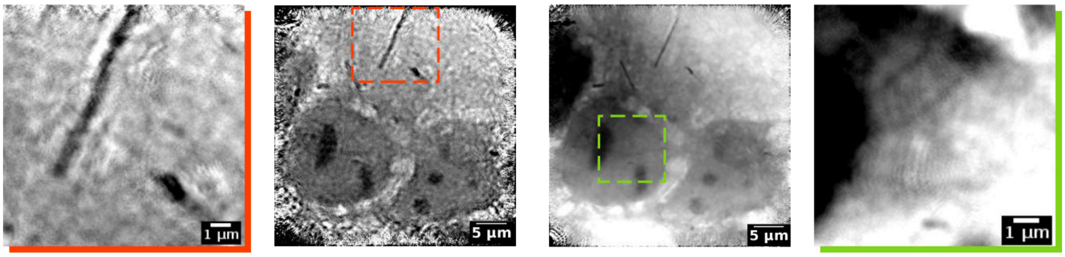
Combining all these three major problems into a unified reconstruction algorithm is quite attractive, especially if limited manual intervention is required. It is especially in this chapter that the framework disccused in chapter 1, the one made by the three pillars of Physics, Computing and Inverse Problem Theory fully takes life, providing again example of the multidisciplinary nature of this research and this thesis.



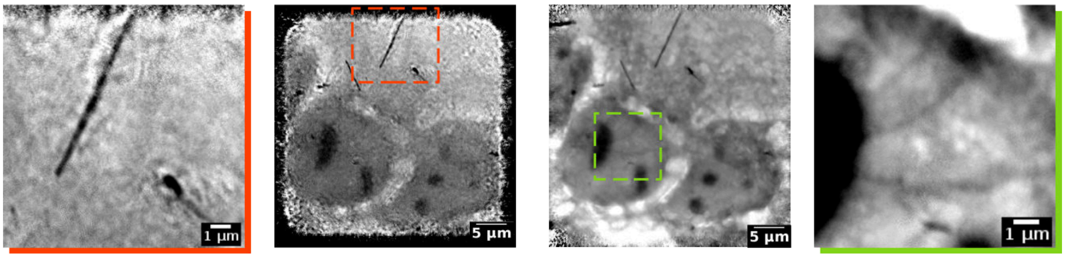
Multi-mode DM



Multi-mode E-PIE



Single mode R-PIE



Proposed "**SciComPty autodiff**"
algorithm

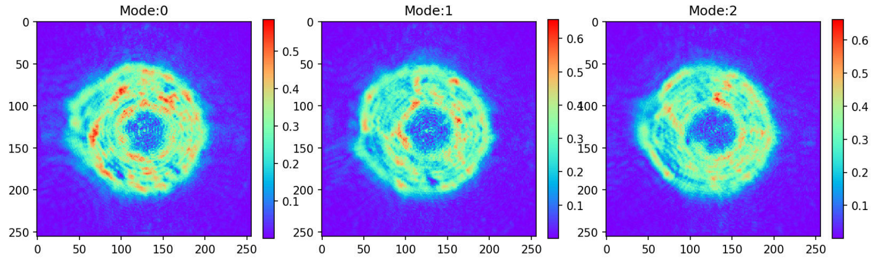


Figure 3.40: Ptychography reconstruction of a human lung cell tissue with asbestos fibres: the proposed algorithm provides the sharpest reconstruction. Two insets provide an enlarged version of the object transmission function magnitude and phase. In the lower panel. the magnitude of the three reconstructed probes is provided.

Chapter 4

Fourier Ptychography

The imaging modality described so far is not the only viable ptychography scheme. A different approach that has been recently transposed to the synchrotron x-ray regime [154] is called “Fourier ptychography”. The technique originated in the field of visible light microscopy where it was proposed not only as a way to produce phase images but also as a path to increase the FOV, while computationally producing large magnification (=high spatial resolution). The net effect is an expansion of the space-bandwidth product.

Visible light lenses are already capable of providing the desired magnification with a high NA, but in that case, FOV is extremely small.

Fourier ptychography embodies a coherent phase-retrieval approach, in which the intensity images acquired under different incidence angles are combined to recover the complex amplitude of the object. The reconstruction offers a resolution beyond the cutoff frequency of the employed optics [155]. This is obtained by combining phase retrieval and aperture synthesis techniques.

In the sample preparation room of a laboratory such as TwinMic, the use of the optical microscope is the routine; not only it can be extremely beneficial to develop an inhouse system providing the aforementioned advantages, but also due to its strict output structure, Fourier Ptychography research can be exploited to design deep learning solution more easily. Indeed, as will be described in the text, by invoking the reciprocity principle, the solutions can eventually be ported to a real-space ptychography scenario.

In this chapter, after a brief description of the technique, we present a contribution to improve the reconstruction speed by using the concept of "Deep

Priors", [5] a single-shot solution to seed a subsequent iterative process. Again, as a computational technique, the forward process will be discussed, followed by the reconstruction procedure.

4.1 Forward model

The Fourier ptychography scheme (Fig. 4.1) is quite straightforward and presents many similarities with the one employed in a camera-equipped optical microscope. A sample is placed in the focal plane of the magnifying system and the detector records pure intensity images [155]. The "computational sampling" element of the technique here is more easily recognizable, due to the fact that the specimen is illuminated sequentially by titled plane waves, which are produced by point sources arranged on a planar array which is placed at a far distance h below the sample. Each plane waves propagates with a different wavevector denoted by $(u_{m,n}, v_{m,n})$. The expression for such a set of waves $\phi_{m,n}$ at the sample plane is given by:

$$\phi_{m,n}(x,y) = \phi_0 e^{jxu_{m,n}+jyv_{m,n}} \tag{4.1.1}$$

where $\phi_{0,(m,n)}$ is the source magnitude (ideally unitary and constant for each source) and

$$(m,n)$$

are the row-column coordinates of the lit source element. In the thin sample regime, the illumination-sample interaction can be again modelled by a pure projection approximation. Thus, the exit wave $\psi_{m,n}(x,y)$ has the usual projection form:

$$\psi_{m,n}(x,y) = o(x,y) \cdot e^{jxu_{m,n}+jyv_{m,n}} \tag{4.1.2}$$

where $o(x,y)$ is the (whole) sample complex transmission function.

With coherent illumination, the behaviour of the magnifying system is completely described by its point spread function $p(x,y)$ or equivalently in the frequency domain by the coherent transfer function $P(u,v)$:

$$P(u,v) = \mathfrak{F}\{p(x,y)\} \tag{4.1.3}$$

which embodies all the spatial frequency response of the system (aberrations).

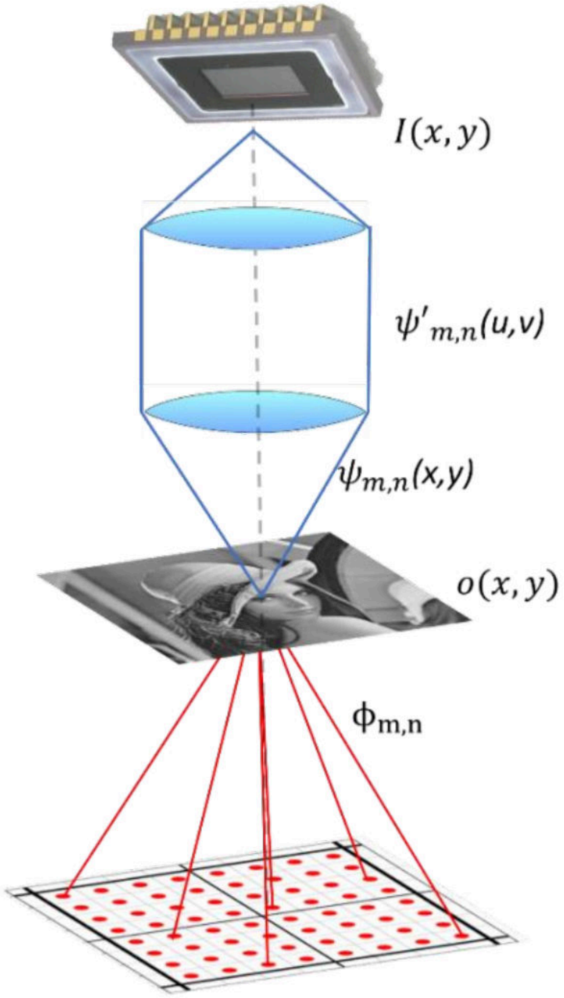


Figure 4.1: If the array-sample distance h is large, the wave-field produced by each point source can be modelled at the sample plane like a tilted plane wave. Many propagation vectors for the 64 LEDs of Fig. 4.2 are shown.

For an ideal magnifying system, $P(u, v)$ is a 2D low-pass filter (a disk) with a cutoff frequency C defined by:

$$C = NA \cdot \frac{2\pi}{\lambda} \tag{4.1.4}$$

where the numerical aperture is NA and the radiation wavelength is Λ.

Other aberrations, like astigmatism, for example, can be described by a more complex form of the coherent transfer function. The recorded image $I(x, y)$ is the intensity of wavefield at the detector plane and is given as usual by:

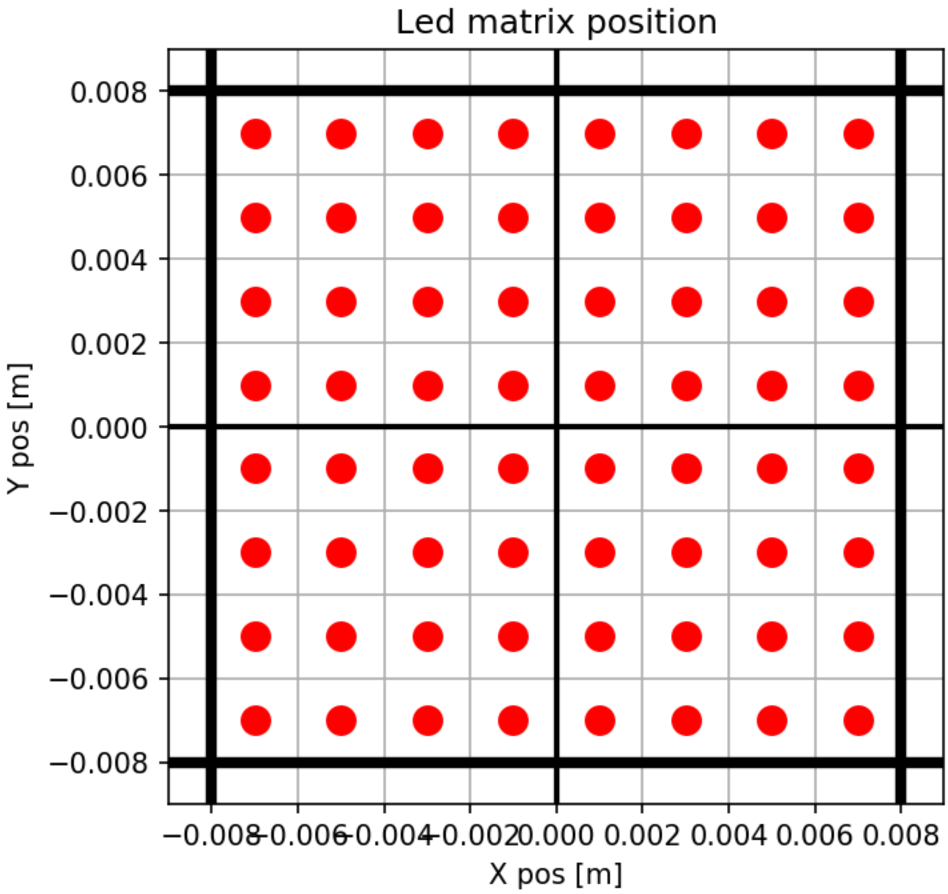


Figure 4.2: A simulated 8x8 LED array scheme is shown.

$$I(x,y) = \psi'^{*}_{m,n}(x,y) \cdot \psi'_{m,n}(x,y) \tag{4.1.5}$$

where the wavefield $\psi'_{m,n}$ is given by the equation:

$$\psi'_{m,n}(x,y) = \mathfrak{F}^{-1}\{\mathfrak{F}\{\psi_{m,n}(x,y)\} \cdot P(u,v)\} \tag{4.1.6}$$

Main principle The processing in the Fourier domain is assured by exploiting the linearity in the complex amplitude of the coherent system [32]. Due to the shifting property of the Fourier Transform, the effect of the tilted plane wave illumination (described by an exponential phase factor) is to translate the exit wave spectrum at the coordinate $(u_{m,n}, v_{m,n})$ in the origin of the Fourier space; the optical system signature is impressed on the input spectrum with its filtering action described by $P(u, v)$, producing the exit wave $\psi^{`}_{m,n}$:

$$\psi'_{m,n}(x, y) = \mathfrak{F}^{-1}\{O(u - u_{m,n}, v - v_{m,n}) \cdot P(u, v)\} \tag{4.1.7}$$

Equivalently, we can follow a model that can be implemented easier in a computational approach: the 2D filter $P(u, v)$ can be moved accordingly on the 2D Fourier plane of $O(u, v)$; a computational box having the size of the intensity image is shifted on the larger $O(u, v)$ Fourier plane. This approach allows using indexes calculation and a simple crop operator as a fast way to act on discrete arrays. The shift coefficients depend on the tilt angle and in turn by the position (m, n) of the lit point source on the array plane. Following simple geometrical reasoning, the wavevector $u_{m,n}$ and $v_{m,n}$ are defined by the following equations:

$$u_{m,n} = \frac{2\pi}{\lambda} \frac{x_{m,n}}{d} \tag{4.1.8}$$

$$v_{m,n} = \frac{2\pi}{\lambda} \frac{y_{m,n}}{d} \tag{4.1.9}$$

where $x_{m,n}$ and $y_{m,n}$ are the coordinates of the points sources on the array plane, while d is the source-sample distance :

$$d = \sqrt{x_{m,n}^2 + y_{m,n}^2 + h^2} \tag{4.1.10}$$

The appearance of each intensity view $I_{m,n}(x, y)$ varies with the illuminating wave vector that produced it: the central LEDs, which produce quasi-perpendicular wavefront, generate images that appear more natural and similar to a typical transparency image, where just the absorbance effects can be spotted. Observing the lateral images, dark-field-like images are instead produced, with an intermixing of the magnitude and phase part marked by strong directional edges due to the phase. In the common optical microscopy application, the point source array is typically made by a 64x64 LED array. While the numerical aperture of the magnifying system determines the spectral content of each view, it is the maximum lateral size of the LED array that defines roughly the highest available resolution. Figure 4.4 shows a Fourier ptychography dataset generated by SciComPty, in which a simulated specimen defined in fig. 4.3 is illuminated by exploiting an 8x8 array (represented in fig.4.2).

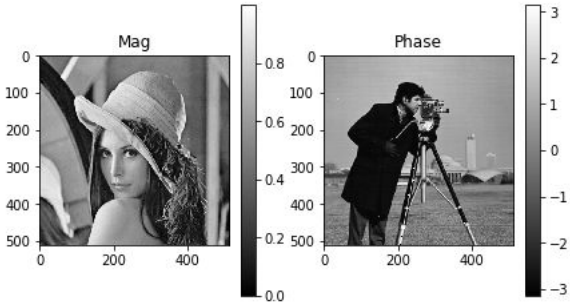


Figure 4.3: The simulated thin object transmission function is composed by enforcing the magnitude and phase array with two images.

4.2 Reconstruction model

Reciprocity principle While at first glance the Fourier ptychography setup appears as completely different from the one described in the previous chapters, the reciprocity principle accounts instead for an equivalence: one optic experiment and its reverse are equivalent and encode the same information; thus swapping the input and output field and the source/detectors roles, nothing change [73]. In a typical ptychography setup (the one of TwinMic), that from now on will be called simply "ptychography" and not "real space ptychography", following the angular spectrum reasoning, a set of infinite, tilted plane waves is produced from the sample plane to the detector plane [32, 36]. The output of such a system exactly coincides with the input of the Fourier ptychography setup. The reversal appears in all the other aspects: instead of scanning in the real space we are scanning in the Fourier domain, and the recording happens in the real space instead of in the Fourier space. This means that the same algorithms can be applied for both the ptychography setups even if on different quantities.

The reconstruction process in Fourier ptychography follows the typical phase retrieval approach, in which a high-resolution sample estimate is sought while holding consistency with the measured pieces of information. An iterative algorithm similar to EPIE [91] is than used in the process: EPRY [156] (Embedded Pupil Function Recovery) applies the illumination correction capability of EPIE [91] to recover its counterpart, the pupil function $P(u, v)$ of the magnification system (referred with the letter P also for its resemblance in the algorithm role with the illumination probe).

The reconstruction is initialised:

- A reasonably good estimate of the object sample is generated by up-

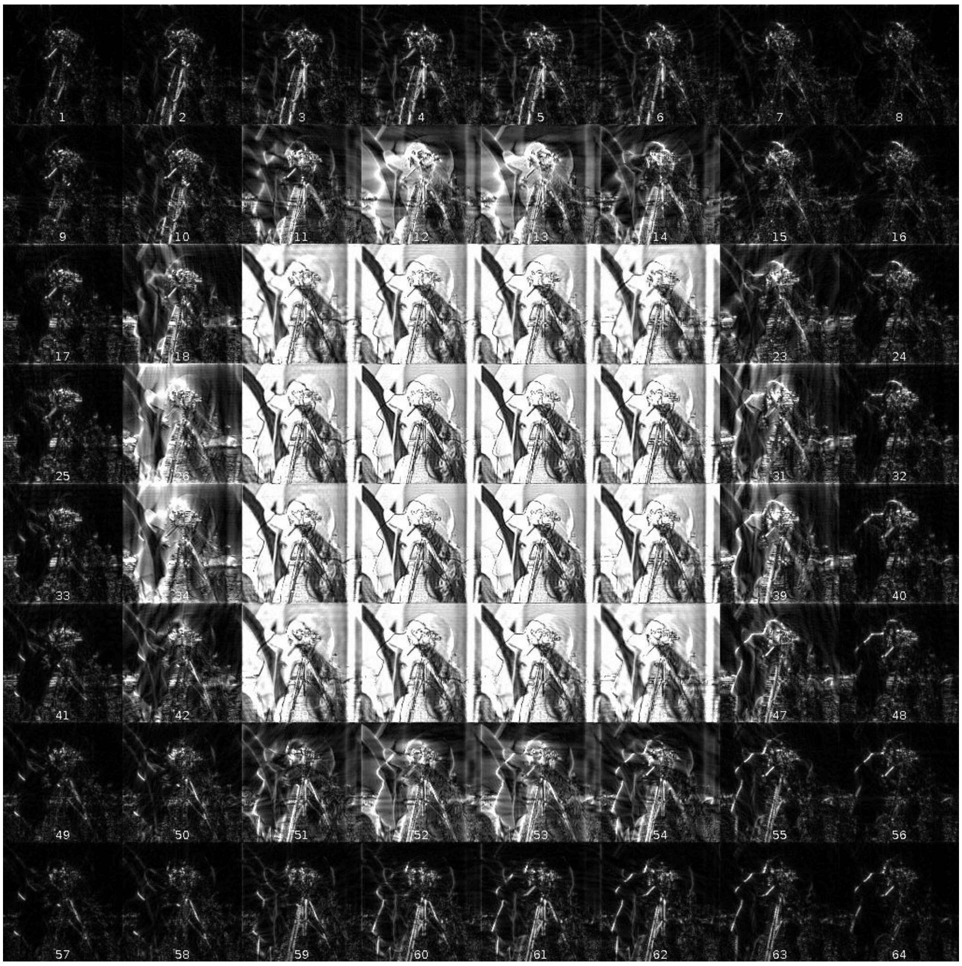


Figure 4.4: A matrix of the 8x8 views at the detector is shown; turning on a specific LED m,n in the array shown in fig. 4.2 produces the image here shown in the box (m,n). Note that the inner images (produced by quasi-perpendicular illumination) are more like a normal transparency, while the one obtained by turning on the lateral leds assume the form of dark-field-like images. The simulated specimen is shown in fig. 4.3. For an 8x8 LED array, 64 intensities are acquired.

sampling the average of all the recorded images. The element-wise average is produced in the 3rd dimension of the image stack. The phase can be set to zero or to a random number. The reconstruction convergence speed is correlated with this initial estimate.

- The pupil function is initialised by generating the disk filter relative to the nominal magnification system NA.

Like all the real space ptychography sequential algorithms, a full iteration of EPRY comprehends the entire set of acquired intensity images:

- For each scanning wavevector $\phi_{m,n}$, the corresponding filter is generated by shifting accordingly the estimate of $P(u,v)$ in the Fourier space
- We walk across the image formation model of equation 4.1.6, by computing the spectrum region of $O(x,y)$ (computational box) and filtering the object estimate in the frequency domain. Then the intensity of the inverse Fourier transform is obtained producing what in the real space ptychography was called “a projection”. A synthetic intensity image is generated.
- The magnitude constraint of the real space wave at the detector plane is enforced in the usual way producing a corrected exit wave of the form

  $$\psi^{"}_{m,n}(x,y) = \frac{\psi^{'}_{m,n}(x,y)}{|\psi^{'}_{m,n}(x,y))|} \cdot \sqrt{I_{m,n}(x,y)} \tag{4.2.1}$$

- The corrected exit wave is propagated back to the Fourier plane where EPIE-like equations are used to update the pupil function $P(u,v)$ and the current object spectrum region $O(u-u_{m,n}, v-v_{m,n})$:

  $$O_{(u-u_{m,n},v-v_{m,n})} = O_{(u-u_{m,n},v-v_{m,n})} + \alpha\Delta_{\psi}\frac{|P(u,v)|P(u,v)^{*}}{|P(u,v)_{max}|+\beta} \tag{4.2.2}$$

  $$P^{'}(u,v) = P(u,v)+\alpha\Delta_{\psi}\frac{|O(u-u_{m,n},v-v_{m,n})|O(u-u_{m,n},v-v_{m,n})^{*}}{|P(u,v)|_{max}+\beta} \tag{4.2.3}$$

  where Δ_{ψ} is the difference between the corrected and uncorrected wavefield at the detector

  $$\Delta_{\psi} = \psi^{"} - \psi^{\text{‘}} \tag{4.2.4}$$

  and β is the usual parameter to avoid dividing by zero.

Differently from (real space) ptychography, this sequential update does not converge well for a random walk across the recorded data: an energy-based sequence is adopted instead, in descending order. This also ensures that the low-frequency regions of $O(u,v)$ are corrected first. In [157] many different update order criteria are studied.

Figure 4.5 shows the scanning density map in the Fourier plane: the reciprocity principle again assures that like in ptychography, an overlap factor between spectrum regions is required to enforce a reconstruction constraint, otherwise for N LEDs, N completely independent phase-retrieval problems have to be solved. No trace of something like a "Fourier-CDI" seems to exist though. Differently from ptychography, changing a scan in Fourier ptychography is difficult because the illumination array has to be redesigned. Many different ad hoc designs have been proposed to increase the overlap distribution [157, 158] We propose like in [51] to use the kind of density map of figure 4.5 to design the Fourier Ptychography experiment.

The reciprocity principle that links each ptychography and Fourier ptychography setups assure also that all the general principles of ptychography apply for both. Then many types of setup unceratanties and modalities can be accounted and corrected in similar ways as in ptychography:

- the raster grid pathology can be corrected by altering the sampling distribution designing ad-hoc LED arrays [157, 158];
- lens aberration are corrected by ERPY[156], DM or other illumination-enhanced ptychography reconstruction algorithms adapted to Fourier ptychography [159];
- state multiplexing has been studied for partial coherence [160], LEDs intensity correction [161] and as a technique to deal with multiple wavelenghts [155]; RGB LED arrays can than be used to illuminate the sample producing stunning high resolution color images with a high resolution grayscale camera or to correct spectral impurity of the source;
- position correction [161] allows retrieving the correct propagation vector angles for each LED;
- near field reconstruction are possible [162];
- confocal 3D like reconstructions have been generated exploiting the multislice approach [163, 164, 165];

The aim Noticing the aforementioned variability in the reconstruction convergence behaviour, we decided to investigate on a method to provide a gain in speed or that allows a fast preview of the reconstruction. Deep Learning methods represent a viable choiche due to the fact that by design operates in "single-shot". In the following section will be provided a brief general description of this powerful family of information processing techniques, followed by the presentation of the designed deep a priori method.

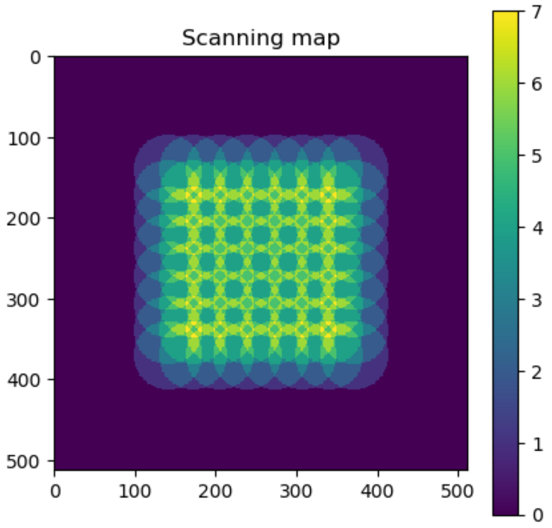


Figure 4.5: In Fourier Ptychography the scanning appears in the Fourier plane rather than in the real space. Conceptually this scanning map is identical to the one presented earlier for real-space ptychography. An analogous overlap constraint has to be held in this domain. Due to the fact that in a typical optical-microscope-based setup the scan array is fixed (only the height can be adjusted), a scheme like this one is critical in designing a Fourier ptychography system.

4.3 Deep Learning methods

In the last decade, regularised iterative algorithms emerged as the standard approach to ill-posed inverse problems [166, 167]. However, in the last year, after the boom of deep learning, many AI-inspired solutions of this kind have been proposed to solve such problems. Artificial Neural Networks (ANNs) are a class of machine learning algorithms that learn from data and specialize in pattern recognition, inspired by the structure and function of the brain.

The main idea is that an ANN can be thought of as generic function approximators [168]. This is achieved through a process called "supervised learning" (the most used method in computer vision and imaging related tasks), during which a set of input data are provided to the system coupled with their corresponding target; iteration by iteration, the trainable system will learn to respond to a given input, imitating as much as it can the provided target. This is not accomplished by "memorising the provided samples", but by generating a computational architecture, an algorithm, that is learned from the I/O relation between data. While self-designing NN are at their infancy, here we discuss about the training of the internal parameters of a model. Supervised learning is thus used also if the function to be learned is not known, as in face recognition, where it is too complex to define a deterministic function between input (a face image) and output (an identity).

Such complex I/O relation is found also in ill-posed inverse problem such as blind deconvolution, inpainting and denoising [169]. Following the previous reasoning, it appears natural proposing NNs methods that can learn to approximate solutions to inverse problems in computational imaging.

To make the previous assumption realistic, starting from the late '90s many crucial advances have been achieved:

- Non-linear activation functions with stronger biological motivations and more complete mathematical justifications have been discovered (Relu), as opposed to the older sigmoidal functions that were differentiable but also prone to stagnation [170].
- Inspired by the visual cortex of the mammals, the concept of compositionality becomes central in information processing (processing is accomplished by composing multiple layered operators, which is the fundamental element within "Deep Learning")
- The introduction of local invariance and local connectivity in processing images (Convolutional [171, 172] and Pooling Layers)
- The definition of efficient backpropagation algorithms designed and tested on Deep Neural Networks (DNN) [148]

From the technology point of view, the availability of a huge amount of data and relatively cheap high-end processing power enabled a speed up in this field. Deep Learning solutions have been applied in a variety of domain, from superhuman performances classification and recognition tasks to autonomous drive. Deep Learning-based computational imaging is one of the hottest topics in the field, even if their actual implementation is philosophically debatable. Indeed, many solutions, especially for deblurring, denoising or super-resolution, employs a component called "generator" which is the main responsible for the stunning results that are achieved; image generation then is not entirely guided by the processing of the actual information, but by a real hallucination (seeded by random generator and only guided by the input data), triggered by the presence of a particular pattern in the input image. That is why this kind of solutions is often rejected in designing forensic algorithms [8].

The drive of DL computational imaging methods is rooted in its "train once - use forever" modus operandi: while in common iterative methods a new time-consuming procedure is launched for each sample, once trained for a specific problem (in an extremely time-consuming procedure) a DNN model should be able to provide a meaningful solution for a vast class of samples, ideally, every possible one. Unfortunately, in many cases, this is far from being true. The generalisation measures the skills of a particular trained model in dealing with different image contents.

In a supervised-learning approach, three datasets are used:

- a training dataset is exploited to make the model learn the complex I/O relationship. To do so a loss function is calculated given the target and the current output of the network. A back-propagation procedure exploits the error gradients to tweak the internal parameter of a model, in order to minimizing the loss.
- A test set is only used to evaluate the performance of a network, by simply calculating the loss function value for the given input-target couple.
- A validation set is used as “fake test data” to tune the hyper-parameter involved in the training. There are, in fact, many parameters for which a gradient is not available (or is too costly to elaborate, or involves profound changes in the model structure) and the tweak has to be made manually.

Evaluating the performance evolution with these three datasets in hand, a trained model can:

- underfit: if the performances on all the three datasets are completely unsatisfactory
- overfit: if the performances on the train dataset suggests a good training, while on the fake (validation) and real test data the model performs poorly. In this case the network "does not generalise well" to unseen data. The operation of the model is then based on peculiar image features that are present only in the training data.

Generalization requires the structure of the network to "resonate" with the structure of the data. However, the nature of this interaction remains unclear, particularly in the context of image processing; this is the most important aspect of DL research.

4.4 Deep Priors

Deep Learning solutions for Computational Imaging are rarely approached successfully by an end-to-end approach, especially at their infancy. Super-resolution, computed tomography and holography are following the same path.

The paper [131] was the first work proposing a DL method for FP. Later a neural network modelling was proposed [138], based on the automatic differentiation capabilities of the Tensorflow framework [139], similarly to what has been described in chapter 3.4. The work [173] explored a single-shot solution, that even if tested on low-resolution data, provides higher resistance to systematic noise.

Having observed how the ERPY method reacts to poorly seeded reconstruction, and a low number of illuminating LEDs, we propose a Deep Prior method to generate a good initialization (a prior) which provides both magnitude and phase values, allowing for faster convergence. This deep prior estimate can be used also as a fast reconstruction preview (obtained in less than 1 s) on a desktop system.

We propose a deep learning model, FPCNN, trained by varying all the simulation parameters. The testing procedure involved the use of a dataset of completely different nature, to check for the generalisation power.

4.4.1 Model structure

A Deep Learning-based solution for FP reconstructions takes as input the data cube composed of the measured low-frequency intensity images and produces a complex image. In recent years CNNs have proved impressive performance on image-to-image map tasks; a particularly successful network for pixel-wise classification (image-to-image mapping) is the so-called UNET architecture [174]. This model was originally designed for biomedical image segmentation and can generate image masks, where each pixel value corresponds to the estimated class number of that pixel. From the original model, two major modification are required to run such a network on Fourier Ptychography datasets:

- The model mapping must be varied to accomplish the required $N \times H \times W$ to $2 \times K \times L$ image mapping, where K and L are larger than H and W
- The training procedure has to be redesigned to accommodate floating

values (magnitude and phase of the predicted reconstruction seed)

UNET [174] (Fig. 4.6) is based on a peculiar encoder-decoder structure: the encoder is composed of a stack of convolutional and (max)pooling layers. Each convolutional layer is configured with zero-padding, thus layer after layer the width and height dimensions are consumed. However, at each layer an increasing number of filters is produced, making the features cube to grow in depth while shrinking in width and height.

Similarly to resnet [132], skip connections are crucial in increasing the throughput of the model: information can move even unprocessed throughout the network skipping a bunch of layers. This means that during the training, the model can implement, if needed, also residual connections, without trading processing power. Thanks to this structure, the expanding (decoder) part of the structure "relocates" on larger but shallower features cubes.

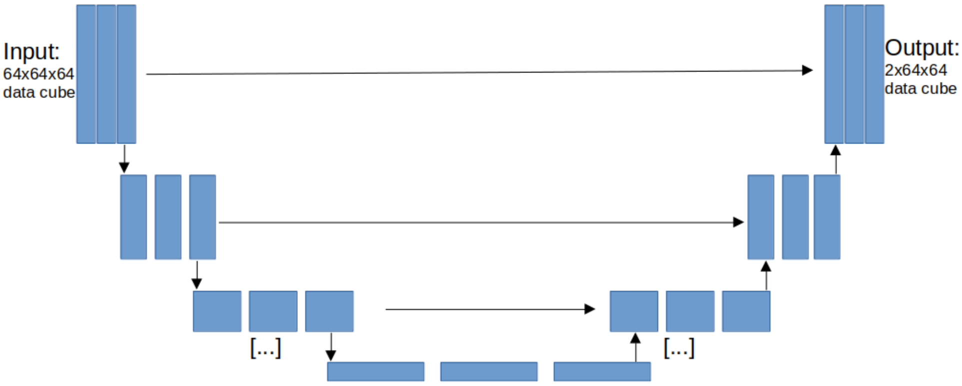


Figure 4.6: Unet structure used; on the left there is the conctracting part which reduces the lateral dimension of each data volume by enlarging the depth. Conversely, on the right side, the expanding part restore the original lateral dimension by concatenation. The horizontal arrows express a concatenation opearation, while the vertical ones denote 1x1 convolution with stride 2, batchnorm and nonlinear function.

Details on the modifications In our implementation, all the convolutional layers have banks of filters composed of 3x3 or 2x2 convolutions followed by a parametric Relu (PRELU) activation [175] (in place of the original RELU), to avoid the dead-neuron effect, a form of vanishing gradient problem. The contracting and expansive paths each contain five convolutional blocks, in place of the original three. A block in the contracting path consists of two convolutional layers followed by a max-pooling layer with a 2x2 pool size, while a block in the expansive path consists of a 2x2 upsampling layer, a concatenation with the corresponding block from the contracting path (i.e. a merge layer), a dropout layer [176] and two convolutional layers. The connecting path consists of two convolutional layers. Finally, the final output layer is made by a 1x1 convolutional layer with no activation, to accommodate also for negative values. This is required to deal with complex number expressed in magnitude and phase. To increase the generalisation power, Dropout layers are added in the decoder structure as well as batch normalisation layers.

4.4.2 Training procedure

During the supervised learning procedure, a Fourier ptychography dataset is synthesised starting from two different images, representing the magnitude and the phase of the object transmission function. Each image is taken from one particular dataset, that varies in the case of the training, validation and testing procedure.

The training dataset is composed of roughly 200 microscope images manually scraped from the internet.

The validation set is instead composed of natural images of the Ukbench dataset [177].

For the testing part, we employed many standard references (Lena, Cameraman, WestConcordPhoto, etc.).

All the simulator parameter (NA, LED plane-sample distance, LED wavelength, LED space) are varied by employing a uniform distribution of parameters. This procedure not only increased the variety in the dataset but forced the network to be not "parameter specific", increasing, in turn, the generalisation power.

Other strategies to increase the "generalisation power" are also used: input images are randomly modified in a procedure called "data-augmentation", which synthetically enlarges a dataset with element-wise operations, rotations, scaling and translations.

Even if the processing cost for each training iteration increases, the training program is written exploiting parallelisation concepts. Each training image is provided to the network through a data-generator that runs in parallel to the training, allowing a fast preparation of the training blocks. Deep Learning application relies heavily on GPU computing (chapter 1.3), due to the amount of operations involeved. To increase the training speed it is fundamental to utilize efficiently the GPU time, without the bottlenecks of a sequential procedure. The aforementioned strategy works in this direction.

The only fixed parameters are then: i) the size of the LR views images (64x64) pixels and ii) the number of the LEDs (64). The input of our CNN is then a cube of 64X64X64 pixels. The output of the model is a complex image of 64x64 pixel that is upscaled to create the reconstruction initialisation. The complex image is represented as two planes of real channels.

The plain Unet is in the literature trained as a classifier since each pixel from the input image is classified between background or "nth-object". Image generation is instead a regression problem requiring a different loss function. Following the work [178], we employed a combination of L1 norm and Structural Dissimilarity (DSSIM) [118] as the objective function.

The structural dissimilarity D between two images I_1 and I_2 is obtained directly from the SSIM index S [118]:

$$D\{I_1, I_2\} = \frac{1 - S\{I_1, I_2\}}{2} \tag{4.4.1}$$

where S is defined as:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1) + (2\sigma_{xy} + C_2)}{\left(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2\right)} \tag{4.4.2}$$

The cumulative loss function $\mathfrak{L}$ employed is then:

$$\mathfrak{L}_\theta = \mathfrak{L}_{1\theta} + \mathfrak{L}_{2\theta} \tag{4.4.3}$$

where $\mathfrak{L}_1$ and $\mathfrak{L}_2$ are the loss functions for the first (magnitude) and the second (phase) channel, that are calculated with the following expressions:

$$\mathfrak{L}_1 = \|M_{gt} - M\| + D\{M_{gt}, M\} \tag{4.4.4}$$

$$\mathfrak{L}_2 = \|P_{gt} - P\| + D\{P_{gt}, P\} \tag{4.4.5}$$

where M_{gt}, P_{gt} are the ground truth images representing the magnitude and phase, while M and P are the inferred images. The optimisation process tunes the parameters θ of the DNN minimizing this loss function, thus making M and P to be similar to M_{gt}, P_{gt}.

Figure 4.7 shows the training history of the model; both the curves exhibits a descending behavior and similar loss values, meaning a correct training procedure.

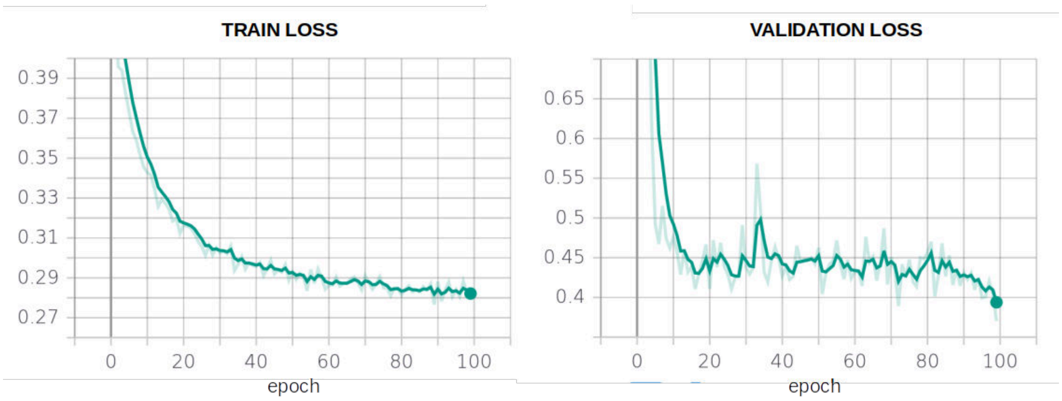


Figure 4.7: Loss evolution for the training (left) and validation (right) phase.

Figure 4.8 also exhibits in both the training and validation curves the same coherent and decaying evolution.

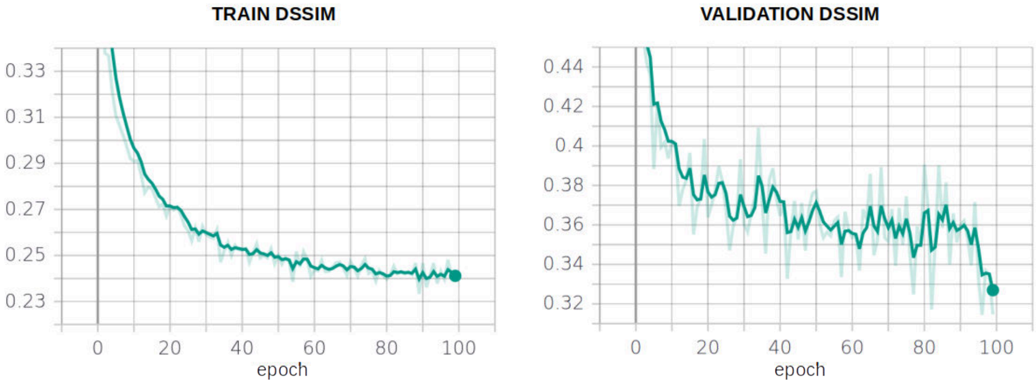


Figure 4.8: DSSIM measure during the training (left) and validation (right) phases.

4.4.3 Results and Discussion

Taking as input a synthetic Fourier Ptychography dataset, the proposed method generates a "single-shot reconstruction" that can be used both as a fast preview and an estimation for an iterative method. The latter point assumes high importance for a limited scanning in the Fourier plane, due to a LED array of limited size. The reconstruction with the simulated array of

8 LED struggles and requires hundreds of iterations.

Figures 4.9, 4.10, 4.11, 4.12, 4.13 shows the output for many validation images; the magnitude and phase are partly decorrelated, but the results will be improved by the iterative procedure. Figure 4.14 shows the output of the iterative procedure applied on a test dataset: after 300 iterations, a conventionally initialised reconstruction provides a reconstruction with poor constrast and artefacts in the phase (marked within the red square). Differently, when the deep prior produced by our method is exploited as a reconstruction seed, the reconstruction quality increases, producing a reconstruction with more constrast and no artefacts on the phase. A quantitative improvement in the reconstruction quality can be measured by computing the difference in the SSIM indexes (whose evolutions are show in Fig. 4.16), calculated for the final reconstruction results: in the presented example it is of about 50% of the max value for a constant initialisation.

The global SSIM index is computed by averaging the result of a local index calculated on small images blocks. Thus, an SSIM map can be also generated: such image is shown in Fig. 4.17. Finally, in the reconstruction metrics presented in Fig. 4.15, a faster convergence can be observed for the deep prior initialisation (the same error value is reached with few iterations) and the reconstruction ends with a smaller value.

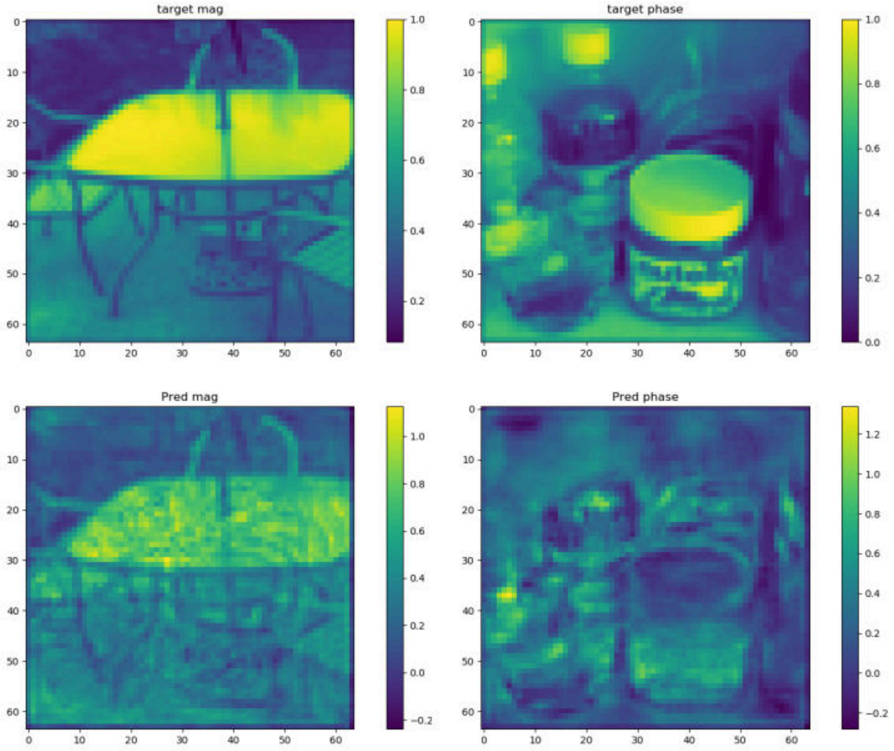


Figure 4.9: Test sample 1: ground truth magnitude and phase (top) and predicted quantities (bottom)

4.5 Summary

Fourier Ptychography is a rapidly improving technique. Having been designed to avoid the limitations of the common optical microscope, there is an extremely high interest in making the technique ready to be used easily. The fact that no movements are implied and that a LED array is the only required additional piece of hardware, is attracting lots of researcher. Despite the recent synchrotron beamline adaptation, the main user basin is still composed by the optical microscopists.

The LED array size limits the technique, providing a limited observation angle and thus a low achievable resolution.

In this chapter we presented a Deep Learning method to provide a fast preview for Fourier Ptychography dataset. Our rapid preview can be combined with a conventional iterative reconstruction to provide a final recon-

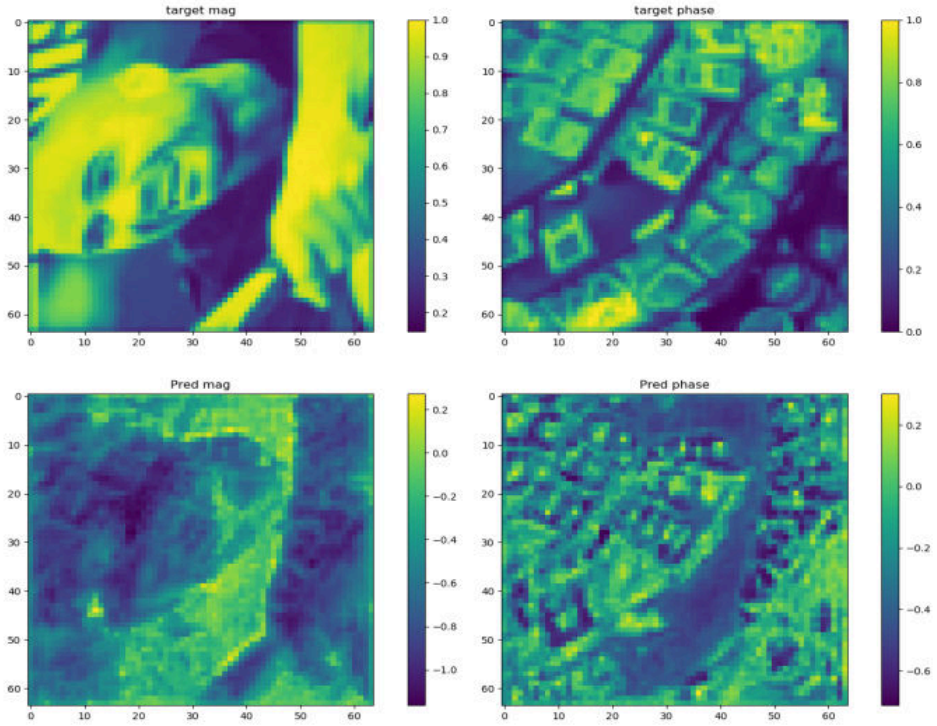


Figure 4.10: Test sample 2: ground truth magnitude and phase (top) and predicted quantities (bottom)

struction with less artefacts and more details.

Fourier Ptychography was first presented, highlighting the similarities with the "real-space ptychography", or simply ptychography. To provide some insight on the method, general principles of Deep Learning for computer vision have been mentioned.

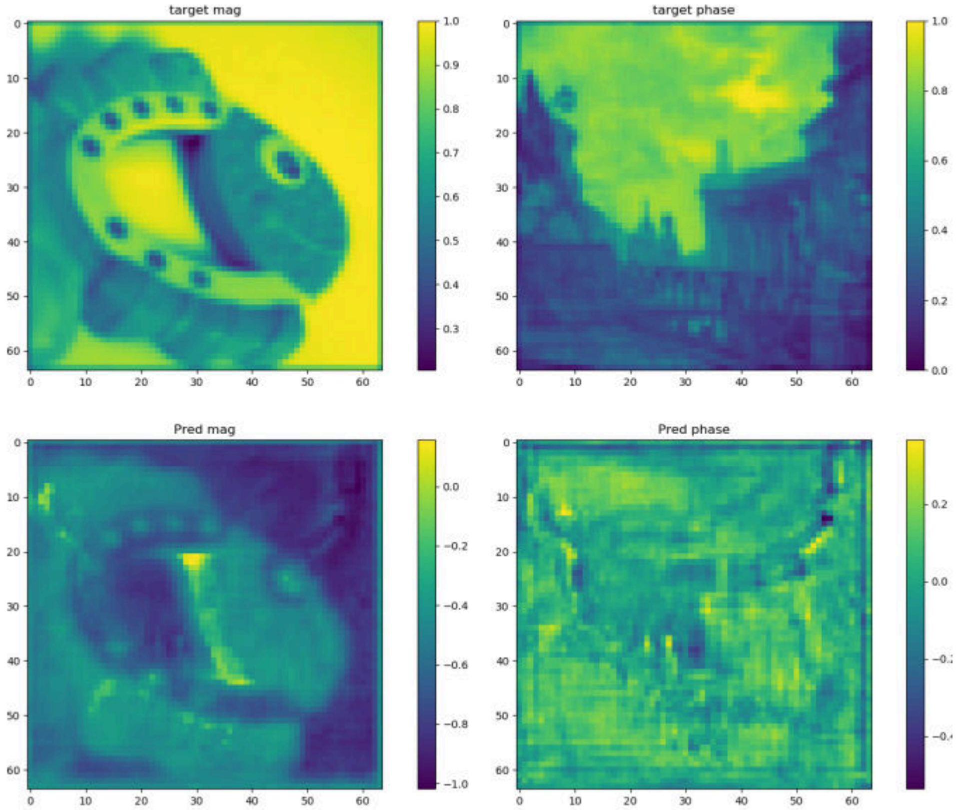


Figure 4.11: Test sample 3: ground truth magnitude and phase (top) and predicted quantities (bottom)

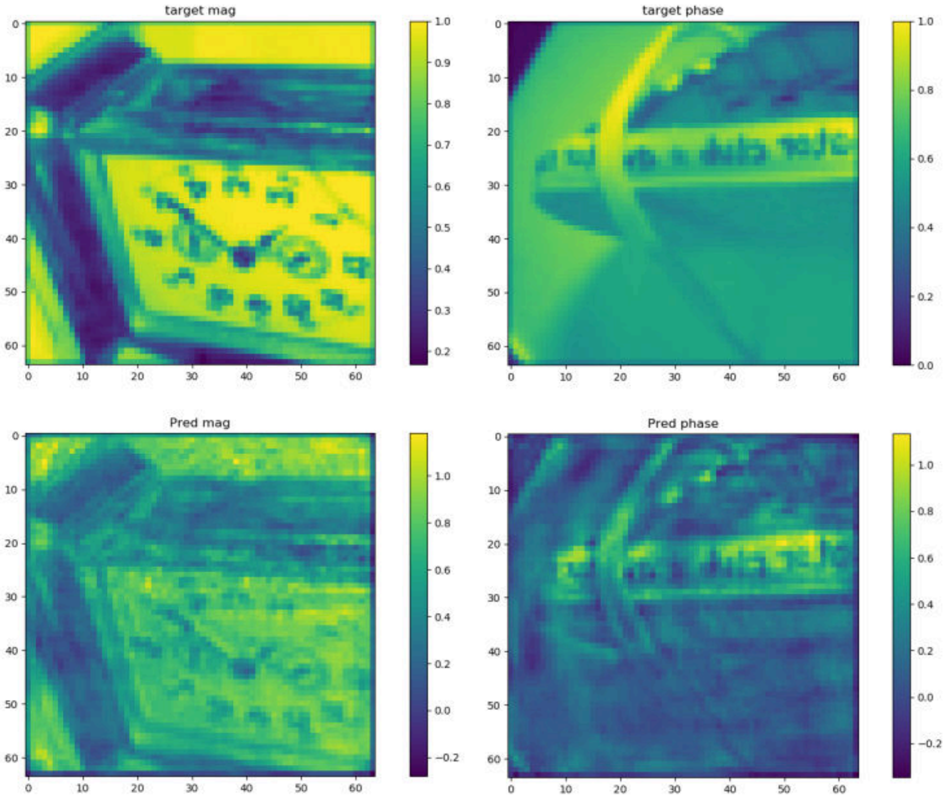


Figure 4.12: Test sample 4: ground truth magnitude and phase (top) and predicted quantities (bottom)

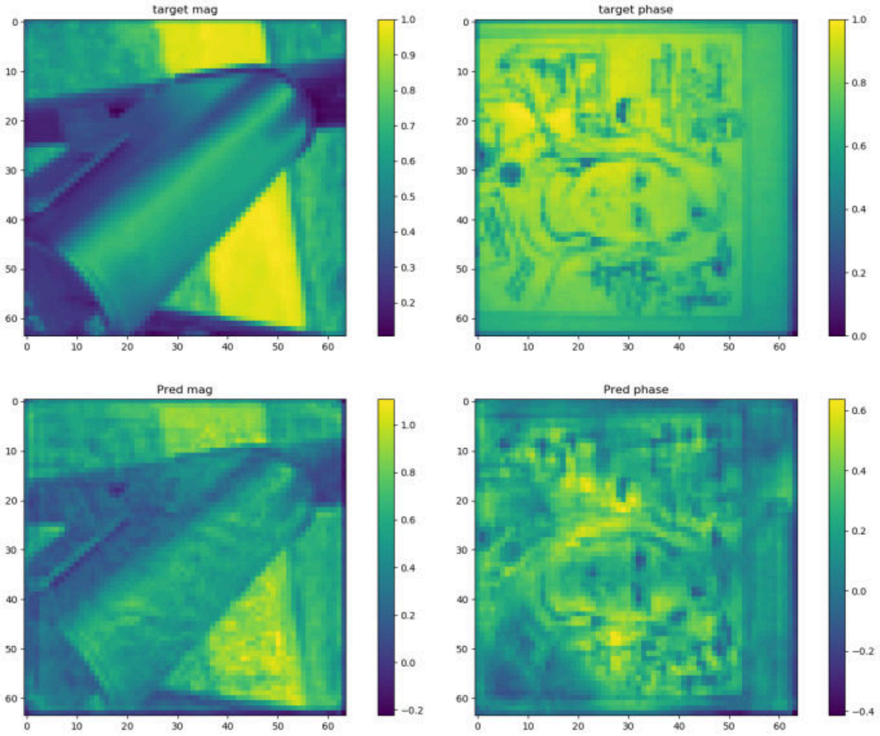


Figure 4.13: Test sample 5: ground truth magnitude and phase (top) and predicted quantities (bottom)

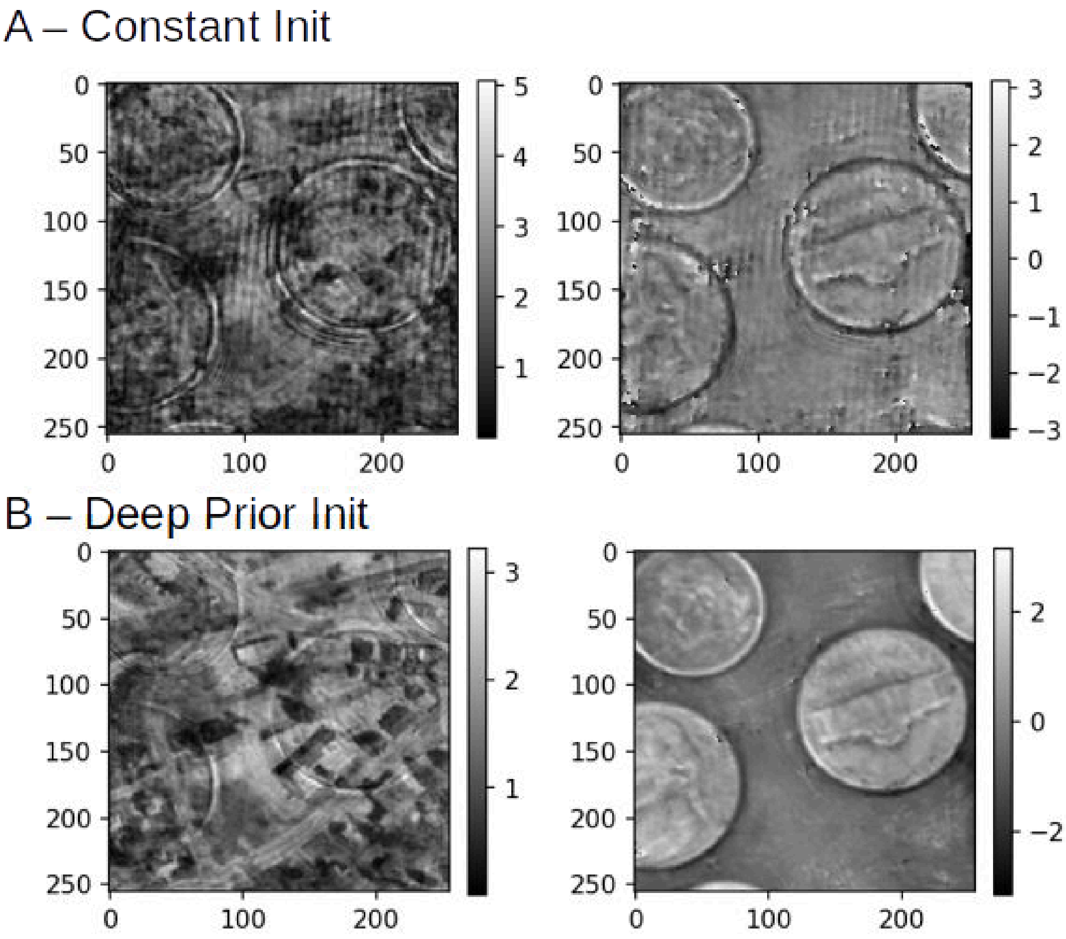


Figure 4.14: Panel(A): The reconstruction after 300 iteration of EPRY using a constant matrix as the initialisation; Panel(B): the reconstruction seeded with our proposed Deep Prior which has a better contrast and less crosstalk in its magnitude and no artefacts in the phase image (red box). These details produce a higher SSIM index (fig. 4.16).

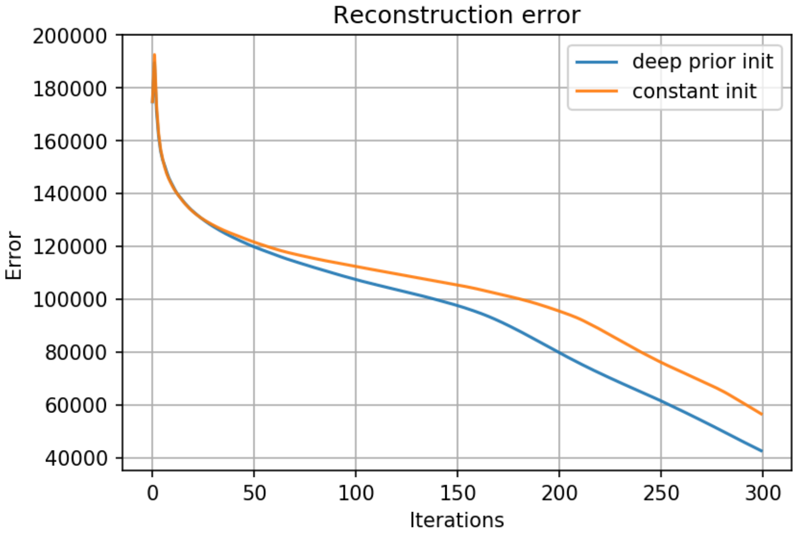


Figure 4.15: Evolution of the reconstruction error in the case of a constant initialization (orange) or a with the use of the Deep Prior (blue).

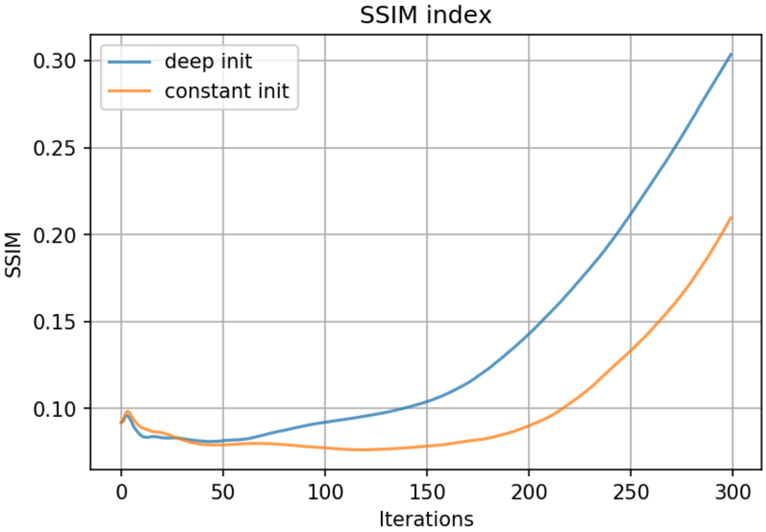
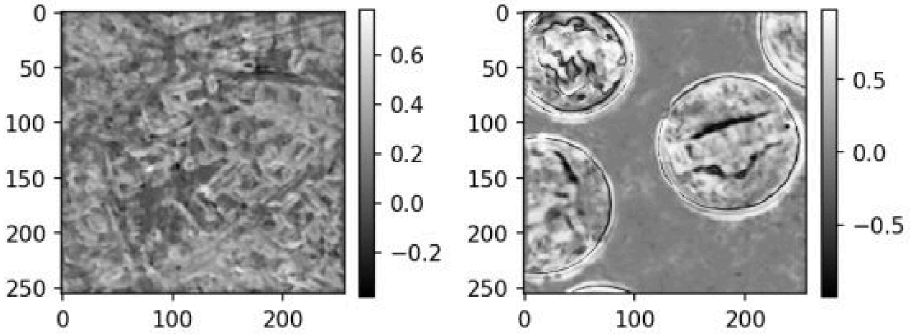


Figure 4.16: Evolution of the SSIM index error in the case of a constant initialization (orange) or a with the use of the Deep Prior (blue). The performance gain can be quantified also in this view.

A – Constant Init



B – Deep Prior Init

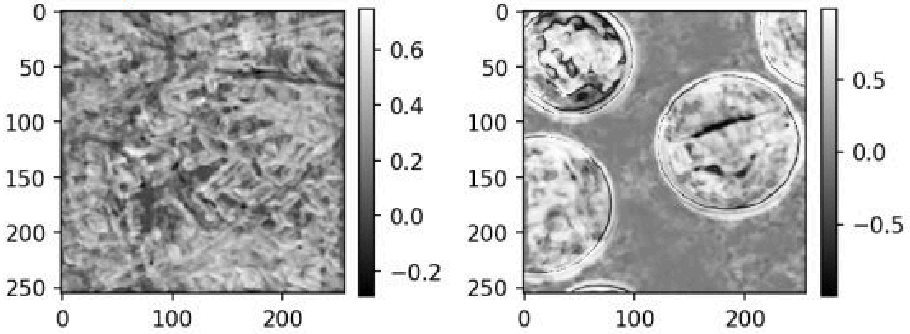


Figure 4.17: SSIM maps for both the constant (panel A) and the Deep Prior initialized (panel B) reconstructions. In panel B the amount of white area (high local SSIM index) is greater than in panel A, indicating thus a higher similarity with the ground truth image.

Chapter 5

Alignment problems and solutions in Computational Imaging

My research is heavily described by the terms "computational methods" and "optimisation", which are recurring themes (and steady carriers, in the telecommunication meaning) in this thesis. It is then obvious that the application of these methods spans many imaging topics in applied physics and engineering. In this chapter, four computational alignment methods based on optimisation are presented, sharing lots of similarities with the previously discussed topics. These methods are novel contributions. Due to the variegated nature of each problem, a brief and concise introduction to the framework will be given, followed by the description of the proposed method.

The first one is a method to align X-ray fluorescence spectra (aka "energy calibration") among different multi-element detectors; the second is about a method to automatically align a CT dataset affected not only by the missing wedge problem, but also by errors due to roll-out and mechanical precision of the rotator stage.

The third method is based on the natural use of the Spatial Transform Network (STN) seen in chapter 3.4 for a face recognition oracle.

Finally a perspective registration algorithm is presented, as a key for fast generation of forensic super-resolution images. It will be shown how the concept of "super-resolution imaging" in ptychography is deeply rooted in the "image processing" super-resolution.

It is important to stress out that even if the last two problems seem too far from what is described in this thesis, the algorithms and methodologies are strictly correlated to the main theme and that is why the sections have not been moved to an appendix.

5.1 XRF energy calibration

X-ray fluorescence (XRF) is a spectroscopic technique that is currently applied in a large variety of fields, spanning from biology, cultural heritage, food science and space exploration [179]. XRF spectra are commonly exploited to gather information on the composition of a specimen, as the fluorescent emissions appear at specific photon energies, making elemental identification a complex but definitely solvable inverse process. Due to its nature, the Low-Energy variant of the technique (LEXRF) poses many more experimental difficulties [180] but is crucial for the identification of light elements.

Silicon Drift Detectors (SDD) [181] are currently employed to measure the photon energy E for each detected photon as it arrives; indeed, the amount of photo-produced electrons Q is a function of the pair creation energy W and the photon energy E, which can then be calculated by measuring the total accumulated charge of the signal pulse [181].

$$E = Q \cdot W. \tag{5.1.1}$$

If a micro-focused X-ray beam (Fig. 5.1 panel a) is raster-scanned onto a sample surface (Fig. 5.1 panel b) [182], the elemental information (provided by the acquired spectra) is also spatially dependent, producing an "elemental map" of the specimen itself (Scanning Fluorescence X-ray Microscopy).

The spectroscopic information is often correlated with the outcome of another imaging technique, such as Scanning Transmission X-ray Microscopy (STXM) that instead provides the sample absorption information (Fig. 5.1 panel e) [179]. Setups such as the ones employed at the TwinMic spectromicroscopy beamline [182] of the Elettra Sincrotrone Trieste synchrotron facility are specially designed to simultaneously perform both the experiments, by employing a transverse XRF detector geometry (Fig. 5.1).

Being the produced XRF radiation isotropic, in order to increase the counting efficiency (but also to acquire topographic information [183]) it is mandatory to increase the observation solid angle, not by enlarging the detector area, but by increasing the number of channels [184] (Fig. 5.1 c1), arranging many small detectors around the specimen (Fig. 5.1, panel c2). Thus, to obtain meaningful information, especially for low counts spectral lines, each spectrum needs to be summed channelwise.

High pixel count planar ring configurations [185] have been superseded by a slanted design with many SDDs suspended around the sample. As a result, both the detector capacity (and thus the Equivalent Noise Charge) and the risk of pile up are reduced. TwinMic currently employs 8 SDDs which cover a solid angle of 0.255 sr, meaning that 4% of the isotropically emitted photons are collected [182]. New detectors and instrumentation design are under active development and will provide a total surface area that covers 22.4% of the entire hemisphere; the proposed solution will employ a total of 32 detectors [184]. Even more complex design with 64 channels are under development [186].

A set of quantised spectra (Fig. 5.1 d) is the starting point for the data analysis (element fitting). Unfortunately, due to the heterogeneity in the signal processing chain, the effective value of each energy bin for each channel is slightly different, meaning that no channel-wise operation (e.g. summation/average) can be directly carried out on the raw spectra. In our case the spectra are not only energy-shifted but a scale transform is also required for the x-axis to correctly align the signals. Many automatic techniques can be fooled [187]: being collected from different viewpoints, each spectrum is only similar to the others and possesses only few common features. Manual alignment is thus currently the preferred method, which however becomes more tedious and error-prone as the number of channels increases.

In this STXM setup, for each detector we thus record maps of $W \cdot H$ spectra. The set of recorded signal is then an array with dimension of $W \cdot H \cdot K$, where K are the frequency bins we are able to discriminate in the photon spectrum.

The detector systems present on Twinmic typically acquire $K = 4000/4096/8000$ bins. To increase the solid angle at which a radiation can be detected, many multi-element detectors can be arranged in an ensemble setup that will collect more fluorescence emission than what one can sense with one detector, as shown in figure 5.2.

The problem What is important in this section is that being collected from different viewpoints, each spectrum originated from the same point on the specimen will have a form that is not equal but only similar to each other. The tacit assumption is that, ideally, each detector should provide the same response for a common stimulus (the same spectrum).

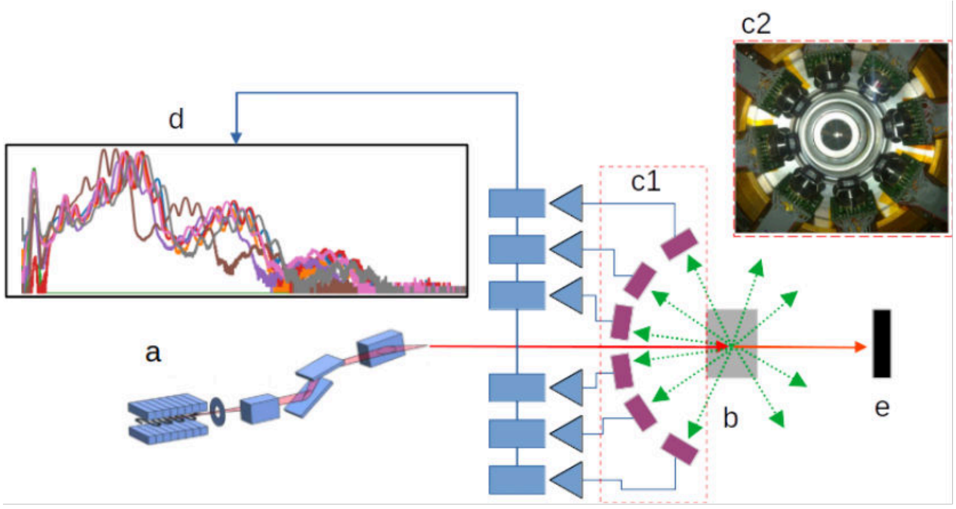


Figure 5.1: Transverse geometry spectro-microscopy beamline: an X-ray beam is produced, focused (a) and projected onto a sample (b). Fluorescence is partially detected by the detector arrangement (c1, c2). The processed signals are raw spectra which require an energy calibration procedure (d) prior to fitting. The absorption image is recorded on a different detector (e).

Energy calibration

An important computational operation on a set of spectra is that of aligning them to a reference spectrum. In X-ray fluorescence, this is referred to as *energy calibration* [188] and may be necessary for fitting low-count acquisitions. For better statistics, especially in trace element studies in the soft X-ray regime, the signal of all detectors is summed before being fitted. In order to do so, it has to be energy calibrated, which is a type of spectra alignment. This is a composite operation that often requires user input for peak identification and setting of parameter values. One of the advantages of using a multi-element detector system is the increased overall photon count, and therefore, an improvement of the system's detection limits.

However, a non-precise energy calibration could cancel out or hide small X-ray fluorescence (XRF) peaks corresponding for instance to trace elements, reducing the detection limits of the overall system. Moreover, when the number of detector cells or elements becomes high, manual energy calibration can be difficult and time-consuming.

The amount of recent work in the literature related to spectra alignment is the measure of how the problem is not completely solved yet [187]. Indeed, even in the case of commercial multi-channel systems, the acquisi-

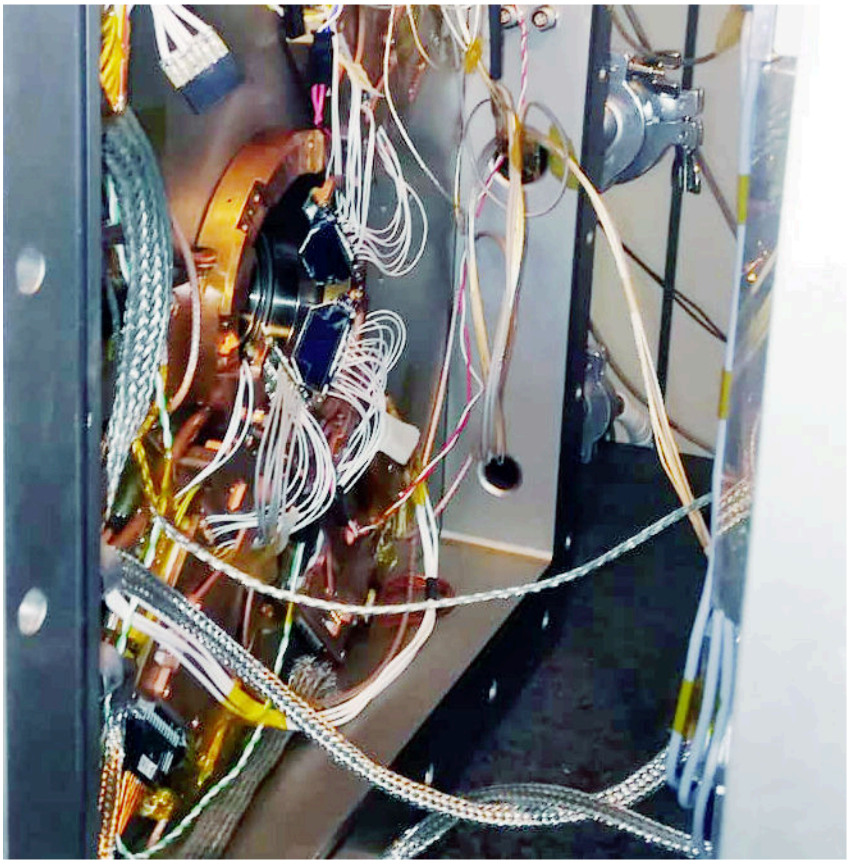


Figure 5.2: The arrangement of a new set of Silicon Drift Detectors (SDDs) within the Twinmic end station; the focused incident beam comes from the left (the photo has been horizontally flipped to have the source at left and the detector at right) and the back-emitted fluorescence radiation (the detector looks to the right) is detected by 4 lateral detector (the blue trapezoids), each one with 8 cells. The signal conditioning electronics is mounted on the same detector PCB.

tion software provides a GUI for semi-automatic methods, which rely on the human-made selection of features in the signals (peaks). Similar reasoning is applied also for open-source solutions, such as PyMCA [189], a widely used software for XRF processing.

From this perspective it is apparent that a common approach to signal alignment is based on three phases:

1. peak detection;
2. peak matching among different signals;

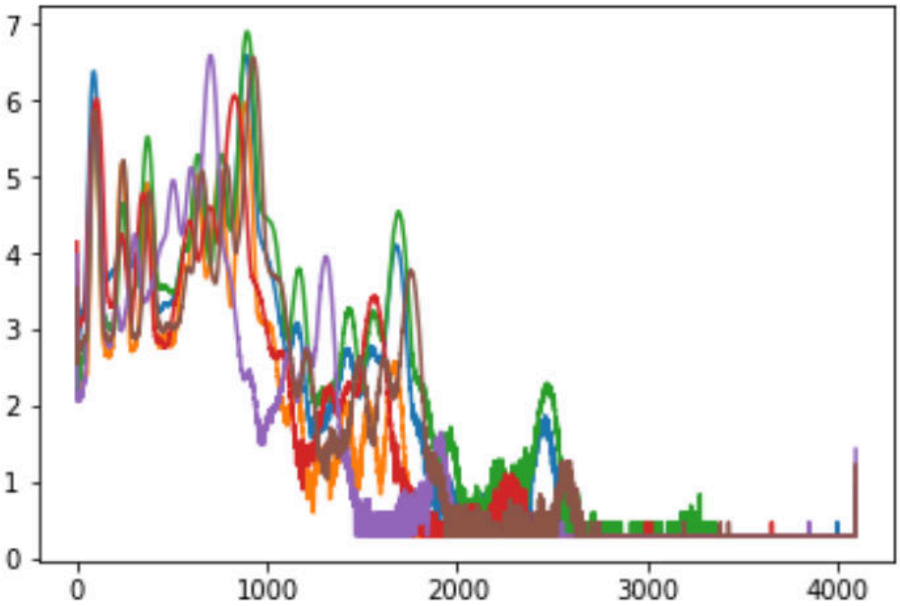


Figure 5.3: Cumulated raw spectra signals (acquired at Twinmic) for each detector (shown in different colours). The x axis represents the energy bins of the discretised fluorescence spectrum, while the y axis shows the logarithm of the cumulated count on each pixel at that energy. The labels are the exponent. It can be seen that a misalignment is heavily present among each detector

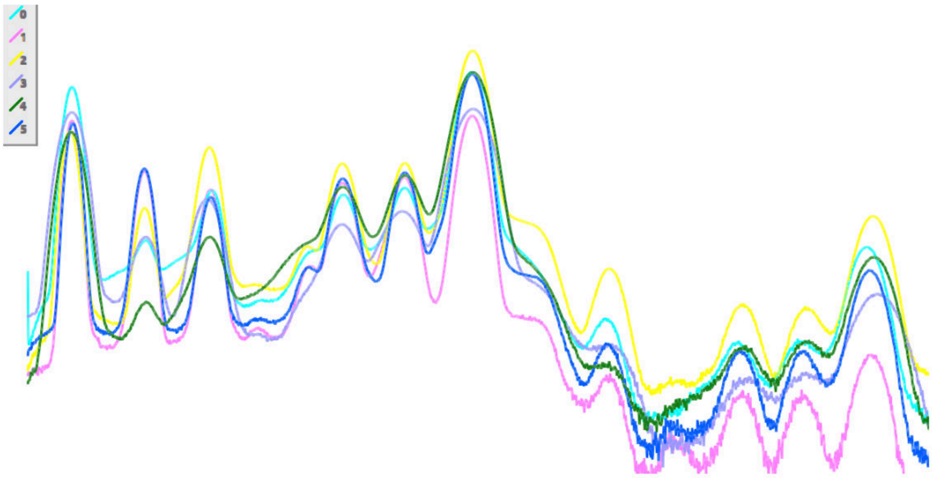


Figure 5.4: For the same data, the manual procedure is carried through the PyMCA software.

3. transform parameters estimation from the peaks list.

Manual intervention is obviously required only in the first two phases where the user is asked to select corresponding features among many spectra. However, automatic methods following the same reasoning are also reported, e.g. in [190]; a second common solution to signal alignment is referred to as “Dynamic Time Warping” [191], a non-linear warp procedure that is typically employed in speech processing. The method is designed to align two signals by creating a warping path in the 2D synthetic space of the x coordinates, which describes the joint coordinates transformation that makes the two signals appear more similar. Unfortunately, this approach is not very suitable for our purpose as it is designed for a pairwise alignment and no fixed reference is maintained. A second major drawback is the severe deformation of the signals (staircase artefact). A third common way to align the signals is to formalise the process in an optimisation framework, where a loss function such as MSE or a form of correlation is iteratively minimised (maximised).

A different type of iterative process is instead used in [192] which is a hybrid technique, where an artificial signal is generated by finding the locations of the peaks on the reference signal and parameters of (3.4.10) are then iteratively searched to maximise the correlation between any other spectra with this artificial reference signal.

The first goal of our work is to align all these spectra on top of each other, using a linear model for the support: the alignment in this modality requires an affine transform, expressed by two parameters: offset and scale. The interpolation can be formulated with the following expression:

$$f_k^{'} = a \cdot f_k + b \tag{5.1.2}$$

where the frequency bin f_k is transposed to $f_k^{'}$ by the affine transform of parameter a (scale) and b (offset).

In this process, one spectrum is elected as "a reference curve"; for n curves, $n - 1$ alignments with the reference signal are necessarily. A reference signal can be choosed at random or one can say that the the best reference would be the one with the maximum energy content. A better choice would be the one with the highest number of peaks, but a peak finding algorithm can be easily fooled.

The dataset that we have to align is a tensor of dimension $N \cdot W \cdot H \cdot K$, where N is the number of detectors, W and H are the size of the map and K is the number of the frequency bins. The characteristic function of each detector remain constant for each pixel of the map, so instead of align $W \cdot H$ signals for each detectors pair, one can sum each $W \cdot H$ spectrum in a single spectrum that characterise what that detector collected. Then, only $N - 1$ alignments have to be done. The procedure typically take 10-15 s for a complete automatic alignment. The final alignment can be judged qualitatively by an xy scatter plot.

Function alignment can be done in many ways that aim at estimating refinement coefficients or by directly accessing to all the observations, or by describing each spectrum with a series of localized features. The tassonomy of the latest approach involves both the analysis in a different domain (some form of signal decomposition and representation (eg. wavelets, Fourier methods, SVD, spline...), or the use of localization features (e.g. peaks locations, Scale Invariant Feature Transform) that will be fed in a direct estimation of the inverse transformation which fullfills the registration. In this work, the 1D functions alignment is casted as an optimization problem in which we find the alignment parameters Θ which minimize a loss function applied directly to the signals s_1 and s_2, by exploiting a similarity metric f:

$$\mathfrak{L}(\Theta) = f(s_1, s_2; \Theta) \tag{5.1.3}$$

Good candidates for the similarity metric are:

- L1 norm
- L2 norm
- Pearson correlation
- Normalized Mutual Information [193]

In the literature a solution has been proposed [188], which exploits the Pearson correlation as a similarity metric. In this work, a similar approach has been developed independently but exploiting 1) a multi scale approach 2) a more effective similarity metric, the Normalised Mutual Information (NMI) [193].

In an information theory mindset, Normalized Mutual Information measures how a given function value in s_1 can be predicted by the corresponding value in s_2. It is defined as [193]:

$$NMI(s_1, s_2) = \frac{H(s_1) + H(s_2)}{H(s_1, s_2)} \tag{5.1.4}$$

where $H(s_1)$ and $H(s_2)$ are respectively the entropies of s_1 and s_2, while $H(s_1, s_2)$ is the joint entropy.

Given the information content I of a signal s of lenght M [194],

$$I(s_i) = \sum_{M} \log_e \frac{1}{p_i} \tag{5.1.5}$$

which is a number expressed in bit, the entropy of the signal is the average information content of s [194]:

$$H(s) = \sum_{M} p_i I(s_i) = \sum_{M} p_i \log_e \frac{1}{p_i} \tag{5.1.6}$$

and is expressed in bit/signal.

The joint entropy $H(s_1, s_2)$

$$H(s_1, s_2) = \sum_{i=1}^{i=M} \sum_{j=1}^{j=N} p_{i,j} \log_e \frac{1}{p_{i,j}} \tag{5.1.7}$$

measures the average information content of a pair of symbols (bit/ signals pair).

Thus, the numerator of the NMI describes the entropy of the two signals, observed independently, while the denominator describes the total entropy of the two signals observed jointly, toghether. NMI is a similarity metric, then in order to minimise it one simple approach consists in negating the similarity to obtain a dissimilarity.

As in many optimisation frameworks, we have found find two problems: the loss function, in general, is not only noisy, but also not convex. Loss function can be made smoother just by reducing the high frequency noise in the signals. The approach works the best if, as in this case, noise and signal can be decoupled with a low-pass filer. The low-pass filter has to be choosen in order to not destroy the detail information in the signal, which helps the solution to converge (the peaks).

The convexity issue problem can be eventually solved only by changing the loss function or the domain observed by it. The best we can do is to help the optimizer not to fall in local minima, giving the algorithm a better view of the loss surface. To do so, we can use a multi-scale approach that will manipulate the domain of the raw signals: at a lower scale, obtained trough a downsampling process, local details in each signals are blurred, producing a loss function with a more convex behaviour. The optimizer will then find easily an approximate estimate of the parameters (especially the offset), that will be refined at each pyramidal level of the multiscale. Each final point of the level n-1 will be the starting point for the subsequent n level (after a 2x scaling). At the finest scale (full resolution), a real pixel wise estimation can be made for the offset. The scale factor of the affine transform does not need to be scaled among different scales.

The described multi-scale approach is highly inspired by the discrete wavelet transform [195], then each scale is half of the signal (4000, 2000, 1000, 500, 250, 125). In order to avoid aliasing, before each downsampling a smoothing process is necessarily: this is carried out through a discrete convolution with a 1, 3, 7, 15, 31, 63 pixel wide Hanning filter.

As in [188], in our experiment, the best performing optimizer is based on the gradient-less "Powell" algorithm [113]. In our experiments, indeed, LBFGS, Adagrad and others are not able to sample correctly the loss surface and are easily trapped in local minima.

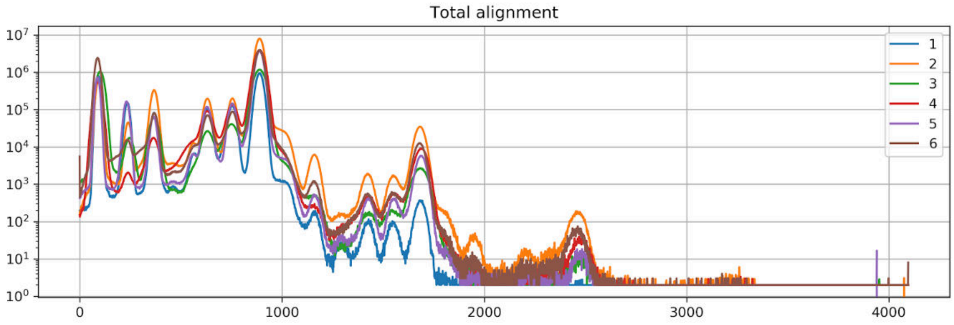


Figure 5.5: The auto-aligned spectra signals

Figure 5.5 shows the final output of the auto-alignment procedure for the set of acquired spectra. As said, the full alignment of n signals is a procedure composed by n-1 alignments with a common reference (figure 5.6, left). In both the mentioned figures, it can be seen that all the peaks have been aligned successfully, as in the manual process. A different view of the same process can be analysed by observing a different representation (figure 5.6 panel B on the right); given the curve s_1 and s_2, for each k frequency bin, the count number is a point in an xy graph:

$$\begin{cases} x = s_1(k) \\ x = s_2(k) \end{cases} \tag{5.1.8}$$

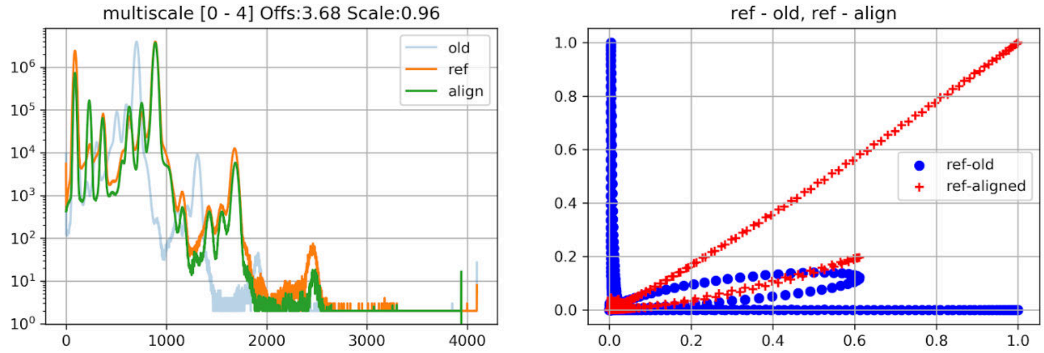


Figure 5.6: Panel A (left) shows the alignment of a single pair of signals, in its raw and aligned version. The alignment process can be appreciated for each curve with an x-y plot (panel B): for each frequency bin, the counts value of the two signals can be represented by a point at a coordinate x(counts at that bin in the reference curve) and y (counts in the aligned signal). Linearity of the produced scatter plot means that at the same frequency bin, the two signals have the same number of counts, that in turn flag a correct alignment (red). The blue scatter plot represents the situation for the reference-raw pair.

Summary

We presented the problem of energy calibration, a crucial pre-processing step in LEXRF analysis. We described the elements that typically fools many automated procedures and finally propose a method, whose preliminary results are really promising. It is the first of its kind based on a fully automatic multi-scale approach. The technique can be used also for many problems involving co-presence of many sources (e.g. brain Electro Encephalographic signals).

5.2 CT alignment

Computer Tomography is another computational imaging technique used extensively in applied science, due to the fact that allows to gather 3D information. In its simplest configuration, this is achieved by recording a series of pure absorption projections through the object [37], with no phase information [34, 33]. Then:

- the first Born approximation applies (chapter 1.7), meaning that the X-ray matter interaction is so weak (single scattering) that the wavefield that reaches the exit window of a slab is essentially the one at the entering window [18, 37];

- a parallel field is used, aligned to the detector [196];
- the detector is placed in the contact region (chapter 1.4), then no phase contrast develops [33, 34];
- the 3D object lies entirely within the depth of field of the microscope [37];

The main point of tomography is to reconstruct the 3D information taking multiple views of the sample, that is rotated at different angles around a *tilt axis.* The set of acquired images is thus called *a tilt series.* In the pure projection approximation, the detected image is a true "shadow" of the sample, which encapsulate absorption information only. With reference to figure 5.7, by looking at a specific view (radiograph, in the $x^{'}, y^{'}$ plane) it is impossible to assign a given absorption coefficient locally (in the 3D space), due to the fact that each pixel value in the detected image is produced by the effect of the absorption interaction along the ray of light ending on that pixel.

Referring to the geometry of figure 5.7, we call a *slice* a specific transverse cut of the reconstruction (or of the object, since the two spaces are somehow similar) which is mapped by a specific row on the detector plane (for a vertical tilt axis). Given a single row of the intensity image on the detector, e.g. $y^{'} = c$ in figure 5.7, the Fourier slice theorem allows to project the 1D Fourier transform of that 1D intensity signal to a specific radial line in the 2D Fourier space of the slice at $z = c$ [37, 196]. By rotating the sample, the Fourier space is filled "diameter after diameter", with a radial density predicted by the Crowther criterion [37].

Cryo nano-tomography is an extremely important imaging tool that allows to analyse biological sample in their native environment, providing ultra-structural three-dimensional information of cells. Laboratories techniques such as Scanning Electron Microscopy (SEM) and Transmission Electron Microscopy (TEM) are extremely helpful tools, but they do not allow for 3D visualization of the whole hydrated and untreated cells [197].

MISTRAL is the soft X-ray transmission microscopy beamline at the ALBA synchrotron light source (Barcelona, Spain) which is primarily dedicated to cryo soft X-ray tomography (cryo-SXT) for three-dimensional visualization of whole unstained cells at spatial resolutions down to 30 nm (half pitch) [198, 197]. Visible-light microscopy allows to visualise biological samples *in*

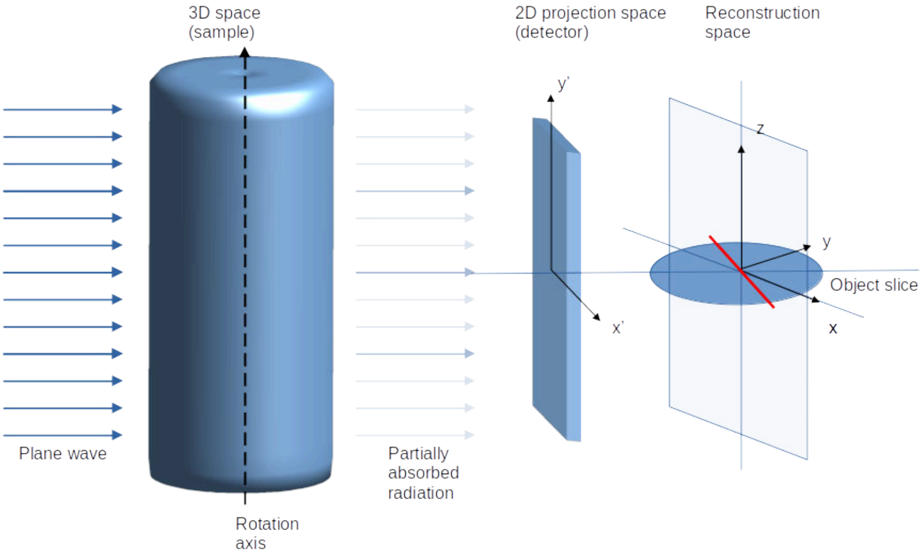


Figure 5.7: General scheme of a parallel field tomography setup.

vivo, but it suffers from the resolution point of view. Transmission electron microscopy offers an extremely high resolution, but the low penetration power of electrons in biological matter makes fast acquisition unfeasible. Soft-X-ray represent then a versatile probe to study biological sample, not only due to their particular behaviour in a specific range of energy, *the water window*, but also because permit imaging in the hydrated state without staining (adding fluorescent components): cryo-SXT needs only the vitrification of water via a cryo-fixation process [198].

The beamline scheme is straightforward: the beam from a bending magnet is processed and directed towards the end-station: a *capillary optic* condenser works as an achromatic lens and is mounted on a xy piezo scanner that allows varying the amplitude and the frequency, thus introducing a new degree of freedom in fixing the exposure of the scene [197]. An objective zone plate is used after the sample to form the image on the detector. The arrangement provides a magnification of roughly 1000x, providing a virtual pixel size of 13 nm. More details on the beamline can be found on [197, 198] and other papers of the same authors.

The sample stage is mounted on top of a rotation table with a run-out below 0.5 µm. This quantity measures how much the rotator does not rotate exactly in line with the main axis. Due to the arrangement of the cryogenic system, the rotation stage provides a limited rotation angle of ±70°. The limited tilting range generates an empty region of information in the Fourier space, producing the so-called *missing wedge problem* [199, 37]. In the reconstruction space it produces an elongation of every point of the reconstructed volume along the beam direction [197]. In addition to this effect, there is an added blurring which is more important than that related to the missing wedge; it occurs again along the beam direction, when the specimen is thicker than the depth of field of the lens [197].

The high resolution of the setup works against us if the errors produced by the limited mechanical precision are not corrected. The effects of many setup misalignments are described in [200]. To correct the misalignment, in a typical cryo-nano-tomography experiment, gold nano-beads are mixed with the specimen solution (prior to vitrification) in order to provide landmarks for the reconstruction procedure, which expects a perfectly aligned tilt-series.

A tool such as "ETOMO" in the IMOD software suite [201] is typically employed to perform a procedure which consists in the manual localization and tracking of the landmarks (the nanobeads). The positions of the nanobeads in the projection space are used to infer the parameters to solve the *tomogram positioning problem*, which involves 3D to 2D mapping [201, 202, 203]. The software then infers the estimated 3D location of the rotation axis as well as its inclination in θ and ϕ and modifies the acquired data to simualate a dataset acquired in an idealised parallel projection setup.

The problem However, in some critical situations, nano beads agglomerate producing large blobs which are of no use during the tracking (limited precision) or are extremely blurred, or are completely non-visible. If no other object in the tilt-serie can be tracked (e.g. lipid-droplets), the dataset has to be trashed.

New algorithms have been proposed in the literature to fix this situation (e.g. [200, 203, 202]) but are extremely computationally demanding, due to the fact that as said in chapter 3.4, the optimisation problems becomes more complex.

The proposed algorithm

I participated in a beamtime, where cryo-SXT was used to image morphological differences on some cells, investigating the effect produced by a particular treatment. The tilt series was acquired at 520 eV. To proceed with the reconstructions of such critical datasets, different techniques have been employed: in some cases the alignment has been performed manually with

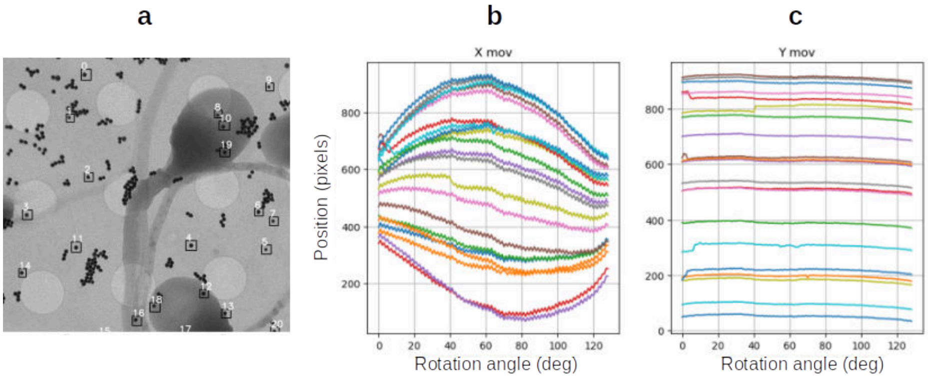


Figure 5.8: If nano-beads are present in the projections (panel a) a tracker can be used for objects with high saliency, producing a scheme of the x (panel b) and y (panel c) movements. Note the high frequency noise over-imposed to the sinogram in panel b and c due to the misalignment.

the IMOD tool, eventually providing the list of beads position produced by a software developed by us. A tracking algorithm described in [204] is used to follow the beads: during the procedure only the first image patch containing a bead has to be provided manually. On the contrary, for the most difficult datasets with no beads at all, or in the case of no objects to track (a dataset that will be otherwise trashed), we designed a simple automatic iterative method to register the frames in the tilt series. The approach is limited to a pure correction of the xy translation factors in the detector plane and aims at registering the frames. Similarly to the case of a cosine stretched cross-correlation between contiguous views, it can only partially compensate the 3D misalignment, assuming that it can be explained by a simpler set of translations on the detector plane.

The algorithm is in principle similar to the one employed in chapter 3.3 for the iterative position refinement of ptychography dataset, and has been inspired by the work [116]; however some modifications are proposed, as explained in the following text. In its Tomopy implementation [205], the algorithm [116] is extremely slow, being CPU based, and is not suitable for tilt-series such as the one we acquired, with a resolution of 1024x1024 pixel. On the other hand, another widely used software, ASTRA Toolbox [206], provides GPU-ready reconstruction algorithms but no alignment option. To apply such class of iterative reprojection algorithm in a reasonable amount of time, a single GPU is not enough. Unfortunately, the MPI version of the ASTRA toolbox [207] is obsolete and difficult to configure.

Similarly to [116], the algorithm we propose works by registering the actual recorded data to the simulated radiograph. To do so, a set of registering coefficients is applied on the real data. As in chapter 3.3, to avoid deteriorating the real data, the correction coefficients are cumulated and a new version of the dataset is created by registering the original dataset using the current value of the coefficients.

We propose the following algorithm:

1. an initial reconstruction is obtained by using the original data, initialising the 3D object;
2. a simulated view is synthesized for each tilt angle in the dataset, by applying the computational forward model (a simulator);
3. a subpixel phase cross-correlation routine (the one described in chapter 3.3) based on [117] is carried out for each pair of (synthetic, real) radiography, producing a set of registration coefficients;
4. the registering coefficients are cumulated for each view and copy of the raw dataset is registered by applying the final coefficients (differently from [116], where each view is built by using the previous data, thus propagating errors and shift-induced artefacts);
5. a new reconstruction is started on this new dataset: for the first 10-20 iterations of the alignment algorithm, the reconstruction starts from a blank estimate, then, differently from [116] the current object is retained and refined at the next iteration, during our experiments;

The final output of the algorithm is an aligned dataset (tilt-series) that can be reconstructed with a method of choice, but the object reconstruction is also provided as a by-product.

Figure 5.9 shows how the proposed algorithm is capable of retrieving the correct position refinement coefficients, producing an aligned sinogram that is generated by stacking each 2D radiography in a 3D matrix, and cutting it at a specific plane.

To minimize the probability of being trapped in local minima, the scheme above is repeated in a multi-scale fashion as in chapter 5.1, by rebinning the original dataset with a decaying factor and repeating the procedure mentioned before. From one scale to the next one, only the coefficients are retained (adapted with a scaling factor), being the object is reinitialised at each change of scale. It is important to note that such important procedure would have been completely unfeasible with the original algorithm implementation, due to the limited execution speed.

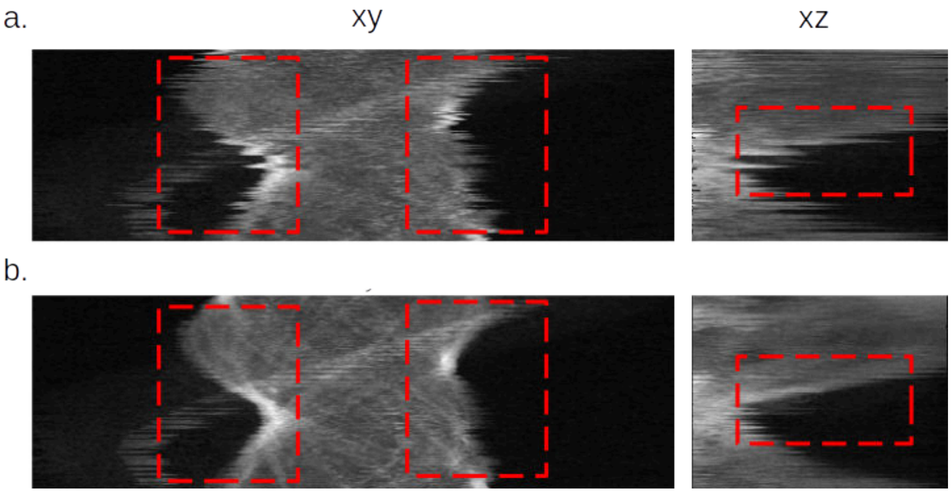


Figure 5.9: The CT alignment algorithm in action on simulated misaligned radiograms. Stacking each 2D radiogram (a radiography) in a 3D matrix, we can slice it on the two orthogonal plane xy, xz, obtaining a sinogram at the current plane. Panel A shows the raw data, while panel B the output of the proposed alignment procedure, that recovered most of the error. Original dataset kindly donated by Prof. F. Brun.

Figure 5.10 shows the effect of a small misalignment, introducing also small in-plane rotation, in order to simulate a model that only partially can describe the real misalignments. The effect of algorithm [116] are catastrophic, with a severe distortion of the sinogram. The reconstruction error diverges after roughly 20 iterations.

Applying the same type of synthetic misalignment, but even with re-binning in the tilt angle (downscaled by a factor of 2), the proposed algorithm produces an aligned dataset (Fig. 5.11).

Figure 5.12 shows the raw and the aligned dataset. Panel B shows the highly fragmented sinogram at the row of the nanobeads on the right (red dotted circle). On the other hand, in panel C, the alignment registers satisfactorily the dataset, producing the expected sine curve. The reconstruction of the registered dataset is provided in figure 5.13 where two reconstructed slices of the object are shown; the tails of the spermatozoa are separated in depth and the morphological details of the collar and the first part of the tail are perfectly visible. In this dataset, the manual tracking was not providing good results, as the only bead that was clearly visible for all the rotation was the one on right. At high tilt angle, the beads on the left become blurry and agglomerated.

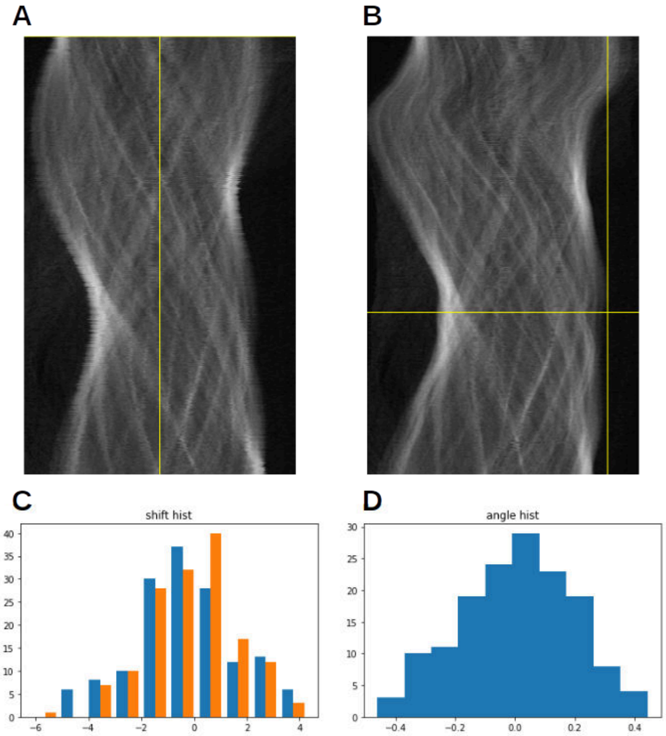


Figure 5.10: A simulated dataset is perturbed with misalignment (panel A). The algorithm in [116] can be easily fooled even by a small amount of noise in the form of translation and small in plane rotation (respectively with statistics in panel C and D), producing a catastrophic effect (panel B).

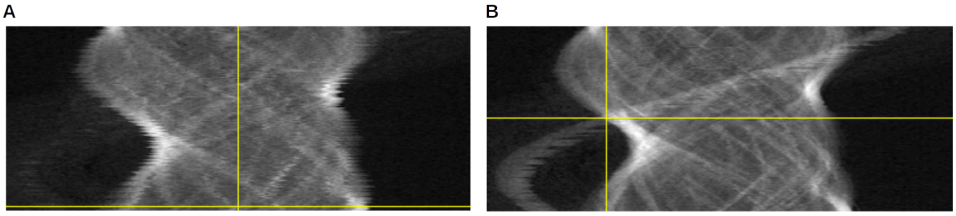


Figure 5.11: For the same amount of absolute synthetic misalignment noise in figure 5.10 panel C and D, but a larger scale (the dataset has been rebinned in x,y and also in the tilt angle step), the proposed algorithm retrieves a good solution.

The missing wedge problem requires a sophisticated iterative algorithm, such as SIRT (e.g. [196]) to be used. Interestingly, during all the alignment experiments (even with [105]), stochastic based algorithms such as Expectation-Maximisation [208] proved to be not robust enough to cope with the misalignment and provided bad results.

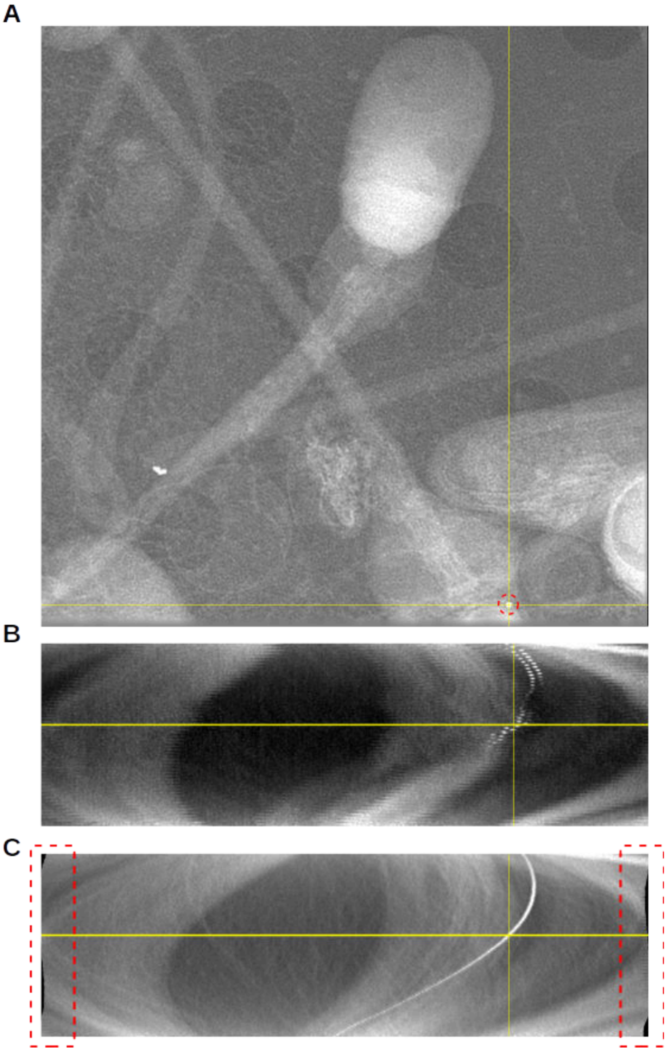


Figure 5.12: From the stacked tilt-series (panel A), one can cut the 3D array in slices which shows the sinogram. If we take a small high absorbing element (nanobeads) as a reference we can clearly see the sinogram curve, in the case of raw data (panel B) and aligned by the proposed method (panel C). Note in panel C the lateral movements row by row.

Implementation details

Looking at figure 5.7, it can be seen that in a parallel field setup, each row produces a completely independent tomography problem, thus making computer tomography reconstructions one of the best examples of problem that can be solved in an embarrassingly parallel framework, just by splitting the data domain. In the implementation of the algorithm, the speed problem

Figure 5.13: Two overlapping slices of the dataset reconstructed with our proposed method. Note the definition of the mitochondria of the cell, a detail that will be difficult to spot in an unaligned reconstruction.

is then solved by employing a multiprocess server which ran the tomography reconstruction by dividing the dataset among the GPUs. When each process has finished the reconstruction, the entire reconstructed stack is recomposed, forming the final object. This method proved to be simple and fast and works best for a single-node multi-GPU system such as the one employed at Elettra.

5.3 CNN distillation and face alignment

The research in deep learning and face recognition methods acted in this research as a driving force to develop the autodiff method for ptychography, and vice versa. In this section an alignment procedure for mobile face recognition applications that we proposed in [9] will be described. The exploited alignment method is exactly the same employed in chapter 3.4, based on an affine spatial transform operator [9]. This section will be short, but it is mandatory.

Face recognition functions are today exploited through biometric sensors in many applications, from extended security systems to inclusion devices; deep neural network methods are reaching in this field stunning performances and represent the best known Narrow Artificial Intelligence use case. The main limitation of these deep learning approaches is an inconvenient relation between the accuracy of the results and the computing power needed. When a personal device is employed, in particular, many algorithms require a cloud computing approach to achieve the expected performances; other algorithms adopt models that are simple by design, meaning that the algorithm can be implemented on mobile devices at the cost of lower accuracy. In this pretty clear trade-off situation, we propose a method to "shrink" an accurate but heavy and complex model into a less demanding structure; this method is called "model (oracle) distillation", and permits to devise of the minimal structure that will enforce (almost) the same input/output relation as the original model. In different terms, distillation is the key to run an accurate but complex neural network in a simple, mobile, hardware.

Face recognition

A face recognition system is typically [9, 10, 11, 12, 13] composed of:

1. a face detector which find faces in the acquired frame (e.g. from a webcam) and crops the ROI of the face
2. an aligning stage which aligns the face on a common set of landmark on a dummy (eyes, mouth, nose, chin on fixed points in the frame)
3. a feature extraction phase in which the face image is processed to devise a feature vector which describes the input image in a particular hyperdimensional space. This *feature space* is the description of a face image operated by the CNN; thus, the "identity" of a particular person is represented by points in this space. All the images of the same identity would then ideally collapse to single point; this never happens in practise, but a good training (and high accuracy) is flagged by the presence of a small cluster.
4. known the structure of the hyperspace, a classifier analyses the generated feature vector and classifies a given image as a particular identity.

The general idea All the details can be found in [9]; here the main point is to show how the same structure and methodology used in chapter 3.4 can be exploited to train a computationally simple model that maintains the same accuracy of a more computationally demanding CNN (Figure 5.14). This results is accomplished through *"model distillation"*.

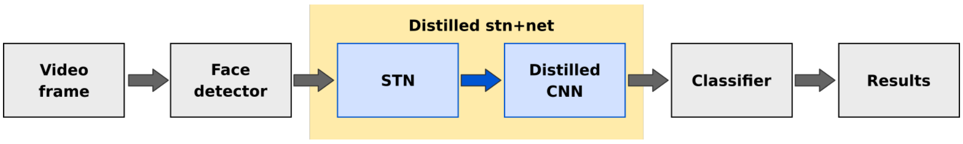


Figure 5.14: Signal chain of the novel hybrid framework composed by an STN component and a custom designed DenseNet feature extractor

5.3.1 Model distillation

The "shrink" technique used in this work (distillation), can be explained in simple terms: what requires model complexity is the extraction of general characteristics from the provided image samples, rather than their actual representation.

Dark knowledge As we have seen in chapter 4.3, a model is trained to generalise well, meaning that is aware of what really cares to accomplish the task. Then, we can say that it keeps a form of knowledge, that in the literature is denoted as "dark knowledge". Mathematically it can be expressed as a latent conditional distribution $p(T|X)$ which relates the input X (an image) to an output T (the response that the model will give). When we as trained humans spot a car, we are quite good at recognising it, then in our mind the distribution $P(I_{car}|T)$ will have a peak for $T = car$ and will be 0 for $T = cake$.

Teacher-student approach In the trained model, it can be shown that when dark knowledge has been inferred by a complex model (*learning is difficult*), it can be eventually represented by a simpler structure that can, in turn, be deployed to mobile hardware. That is why the process is composed of two phases:

1. a complex model (**teacher**) is trained to achieve the best recognition accuracy, following a common procedure;
2. in a second time, a smaller model (**student**) is trained in a teacher-student approach, trying to learn the dark knowledge of the teacher model (distillation);

The aim We proposed to carry out the distillation as a regression process (figure 5.17), forcing the student network to provide the same descriptor generated by the teacher; in the case of an embedding network, this can be directly described in a distance metric framework, where a distance larger than the hypersphere radius of each identity cluster automatically flags bad learning [9]. This motivates us to choose as a loss metric the Euclidean Distance calculated between the target feature vector T and the corresponding predicted descriptor Y.

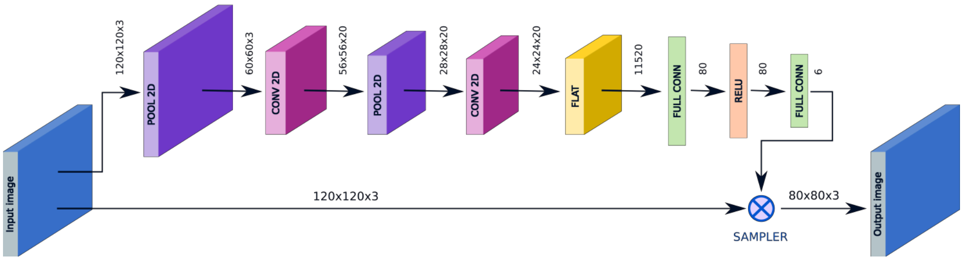


Figure 5.15: Structure of the STN component: an input image is processed by a shallow CNN based localisation network. The convolutional feature maps are then processed trough two fully connected layers that generates the six parameters of the affine transform. A grid generator (not represented) generates the corresponding sampling grid, that will be actually sampled by the sampler, producing an automatically aligned and cropped version of the input.

Alignment

Face alignment is a necessary procedure to improve accuracy. Convolutional Neural Networks, in general, do not have a high degree of spatial invariance. This makes the ROI cropped by a face detector not usable directly without a huge classification accuracy drop. Instead of using a separate procedure which can be prone to error and is slow, we included a face aligner in our CNN model, providing a solution that express clearly an "end-to-end" approach.

STN components As has been described earlier, a Spatial Transform layer [145] is a clever solution that has been introduced to provide spatial invariance to feature maps, by applying a predefined spatial transformation on it; while stride and pool are fixed hyperparameters, STN transformation has parameters that are learned during the training of the entire model.

If in chapter 3.4 the parameters of the spatial transform were directly regressed for the particular sample (as variables in the optimisation pool), here we are training a component, called *localizer* to generate a suitable set of parameters for each different image at its input. Thus, a sampling grid is generated on-the-fly, starting from the on-the-fly inferred transformation parameters.

The efficiency of the STN-equipped structure can be monitored by observing its output: the network focuses its *attention* on the portion of the input frame that it deems relevant for the task at hand. Figure 5.16 shows in fact how, iteration after iteration, a zommed and aligned image is generated. We can observe that the inferred transformation (learned in a completely automatic way) mantains the face at a costant small angle. This increases a bit the "face area" for a given frame size. A second observation is that the angle could have been every other multiple of $\pi/2$, but the regularisation imposed during the training minimises the parameters norm. It is reasonable that this behaviour is forced also by the downsampling of the input image operated by the sampler in the STN. Making this observation on the generated feature maps is called in the literature "co-adaptation study" because involves how the processing of information done at previous layer influence the one "downstream".

Figure 5.16: This figure shows the evolution of the STN output during the training of *'distilled net+stn'*. In less than 10 epochs, the STN correctly localizes the faces, while in 30 epoch the the STN correctly learns to localize and align images for the recognition network.

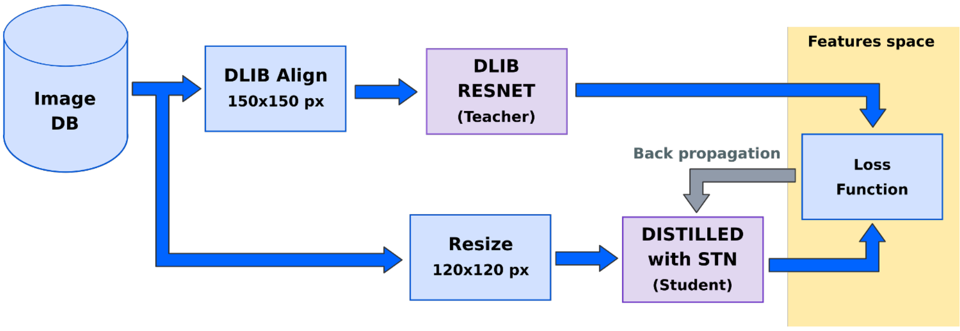


Figure 5.17: Proposed system with Dlib Resnet as teacher and Distilled network with STN as student.

A multidimensional feature extractor network (like the one we trained) cannot be tested "alone". It is necessary to train a classifier that, relying on the generated hyperspace, provides a response for a given image.

The trained models have been tested against the teacher one in both the LFW person verification test and the face recognition test: in the first test, our networks provide a True Positive Rate (at an False Positive Rate of 0.1 %) (Figure 5.18) that is less than 0.01 % less than the teacher value, while in the face recognition test, the performance drop is measured to be less than 2-3 %. These figures are outstanding, considering that the memory consumption of our network is only 1/6 of the reference model while the speed gain is higher than an order of magnitude, thus providing near real-time performance on mobile hardware, not an HPC server.

In conclusion, the reference model relies on a signal processing chain that fails for some hard examples: facial landmarks are wrongly placed in the input image and this results in an incorrect recognition. Our STN equipped model, instead, completely bypasses this face alignment pre-processing step, providing a correct response. The aforementioned STN component is capable to localize faces in an image and to manipulate the frame to "zoom" into it. This behaviour is learned independently during distillation, without human intervention.

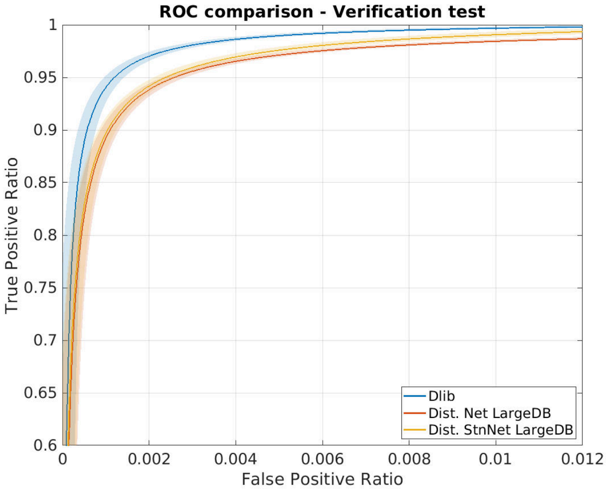


Figure 5.18: Receiver Opearting Curves (ROC) exposing the face identification behaviour of our system against the original one. The reduction in accuracy performances is risible compaired to the shrink in terms of memory and processing power. More details in [9].

5.4 Perspective stabilisation in Forensic Super Resolution

One task often encountered in surveillance videos is the recognition of a target—e.g. the license plate of a vehicle. Often, the quality of a single video frame does not permit a reliable recognition. If multiple frames are available, it is possible to suitably combine them to generate a single image with lower noise (via frame averaging) and/or higher resolution (super-resolution). This last assertion is only valid if the information detail is spread among many pixels (e.g. due to aliasing) [81]; if not, only frame averaging is possible. In both cases the image registration phase is critical because, if done incorrectly, it will lead to an image of lower quality compared to the one obtainable with an interpolation procedure. Several algorithms have been available since the early nineties for the relatively simple case of pure translation and high signal-to-noise ratio (SNR); in turn, the presence of perspective changes in low-quality frames makes the entire procedure extremely challenging.

As said in chapter 1.2, super-resolution is another inverse problem with many similarities with ptychography, which is a super-resolution approach to solve coherent diffraction imaging problems.

In our proposed solution, [8] an object tracking technique suitable for multi-frame super-resolution technique is described. The proposed approach [8] exploits an image registration technique inspired by the Kanade-Lukas-Tomasi algorithm [209, 210] used in a tracking framework. This *modus operandi* is critical in the case of low quality, low resolution, high compression rate and low contrast: this is indeed the typical scenario of surveillance cameras, where e.g. SIFT [211] based methods cannot be used [212]. To solve the problem, we introduce a method capable of accurately computing the perspective transformation [213] that describes the motion of a planar object. The method minimizes the squared distance between the transformed image and a reference, computed over a user-defined region of interest, and uses the partial derivatives in order to significantly speed up the computation.

The essence of the most basic SR algorithms (direct methods) is that each LR image contributes to the reconstruction of the unique HR image with its own information, that is an aliased [214] and differently sampled view of the same HR field [81]. In order to use this information, it is crucial to register each LR view with subpixel accuracy (subpixel with respect to the LR image) and create a higher density (non-uniform) sampling grid, on which information from different frames can be gathered. The high frequency details are restored by applying unblurring operators on this final HR image. As pointed out in [81, 215], the limiting factor in SR techniques performances typically lies in an erroneous registration that can be due not only to geometrical differences in a scene, but also to the presence of different contrast among various frames. All the methods described until now use "model based algorithms", in which a known (but with latent parameters) model is used to generate the estimation of each LR image from an estimation of an HR one; another family of techniques is composed by the learning-based methods.

Among the conventional (non deep-learning based) multiple image super resolution methods [8], L1BTV [214] is one of the most advanced, due to its high robustness with respect to the assumed model of data and noise, a problem that typically affects many computational imaging techniques of these type (ptychography or CDI included). This insensitiveness is obtained by using the combined action of an approach based on the L1 norm minimisation and a robust bilateral prior regularisation that can deal with different data and noise models. The implementation of this method is also computationally inexpensive, making it the ideal candidate for a comparison with our proposed method.

The main points of our method [8] are:

1. The tracking is formulated as an optimization problem: the perspective transformation is sought that minimizes the squared difference between

the warped image and a reference (Eqn. 5.4.3 below).

2. Due to the fact that Eqn. (5.4.3) is a sum of squares of nonlinear functions, we approximate each squared term by a linear function, obtained by computing the partial derivatives.
3. A change of variables is made that improves numerical stability. We exploit a result from matrix calculus for this purpose.
4. A combination of template matching, backtracking, multiresolution and low-pass filtering is used to improve the convergence and to make corrections in case of failure.

As a first step, the ROI has to be defined. The user selects a reference frame in the video sequence—typically, the one in which the object to be tracked is closest to the camera and best visible—and selects four points which outline a quadrilateral region. Figure 5.19 shows an example.

Figure 5.19: Example of user-selected ROI, shown as a red quadrilateral. The quadrilateral should be planar and large (see text).

Let us indicate with $I(x, y, t_0)$ the pixel values of the reference frame of the video and with (x_i, y_i), $i = 1, \ldots, 4$ the coordinates of the four corners of the user-selected quadrilateral. Color videos are converted to grayscale for processing[1]. We also indicate with $I(u, v, t)$ the pixel values of the frame being tracked. The spatial coordinates are Cartesian for all images but it is convenient to indicate their values with different symbols.

---

[1] If necessary, color can be taken into account by extending the summation in (5.4.3) to all the channels of the image in a suitable color space. In most cases, this is not advisable because it triples the computational cost with negligible benefit

We assume that the motion of the tracked planar object is described by a perspective transformation. Therefore, the (u, v) coordinates of a point in the tracked frame are related to the corresponding (x, y) coordinates in reference by the following equation, which has the form of a ratio of first-order polynomials [213]:

$$\begin{cases} u(x,y) = \dfrac{c_1 x + c_2 y + c_3}{c_7 x + c_8 y + 1} \\ v(x,y) = \dfrac{c_4 x + c_5 y + c_6}{c_7 x + c_8 y + 1} \end{cases}, \tag{5.4.1}$$

where $c_1, \ldots, c_8$ are the coefficients to be determined. Finally, we indicate with $\tilde{I}(x, y, t)$ the tracked frame warped to match the reference frame, i.e.

$$\tilde{I}(x,y,t) \triangleq I\big(u(x,y), v(x,y), t\big). \tag{5.4.2}$$

Since the image is sampled, an interpolation algorithm (e.g. bilinear or bicubic interpolation) is used to compute (5.4.2).

For each frame t of the video, the proposed algorithm computes the perspective transformation that minimizes the squared distance between $\tilde{I}(x, y, t)$ and $I(x, y, t_0)$ inside the reference ROI, i.e.

$$\sum_{(x,y)\in Q} \big[\tilde{I}(x,y,t) - I(x,y,t_0)\big]^2 \tag{5.4.3}$$

where Q is the set of pixels belonging to the user-selected ROI. The unknowns depend on t (the perspective transformation is different for each frame) but we shall not indicate it in order to simplify the notation. More details on the optimisation process can be found in [8].

After estimating the displacements and discarding the frames where the estimation has failed, the frames are combined into a single high-quality image. For this purpose, we use a simple geometric approach. The coordinates of each input pixel are warped according to the estimated homography and then mapped onto a high resolution grid. We identify the 4 grid points that surround the target coordinate by taking the floor and ceiling of the x and y components. We add to each of the 4 pixels the value of the source pixel multiplied by a weight dependent on the distance to the target coordinate. We accumulate in this way each pixel of each frame, and finally divide each output pixel by the sum of the contributing weights. In our experiments, we use simple bilinear weights. Figure 5.20 illustrates the concept: the grid represents the high-resolution pixels, the × represents the warped input pixel and (a, b) is the distance between the warped input pixel and the closest upper-left output pixel.

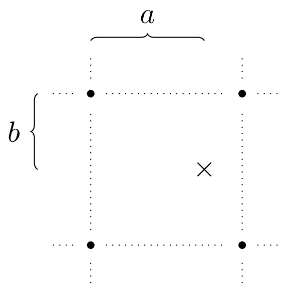


Figure 5.20: Pixel positions in the super-resolution algorithm

The four bilinear weights are therefore:

$$\begin{array}{ccc} (1-a)(1-b) & \cdots & a(1-b) \\ \vdots & & \vdots \\ (1-a)b & \cdots & ab \end{array} \tag{5.4.4}$$

A sharpening effect can be obtained by raising the weights to a positive power. The choice of this method was mainly based on computational efficiency: other more advanced methods have been proposed in the literature; these methods can be used if a higher quality is desired and computational cost is not an issue. Some results of our proposed method are shown if figure 5.21: the proposed method provides higher contrast and lower artefacts than the others.

Figure 5.21: Results comparison for two samples (A) and (B). The super-resolution reconstruction (proposed method and proposed method + blind deconv) are confronted with two upsampling methods as well with other two super-resolution algorithms. The proposed method provides more contrast and allows deciphering correctly the licence plate characters.

Chapter 6

Conclusions

In this thesis, many computational imaging methods were provided as solutions to cope with real effects in imaging experiments, with particular attention to ptychography. The research is based on three pillars: Physics, Inverse Problems and Computing. The intermixing of these three components is essential to produce usable solutions. Modern science requires indeed a high degree of multidisciplinarity, which represents the key for new leaps in progress.

Inverse problems represent the question, the aim. We collect data that do not tell the entire story of what we want to observe and know. At the same time, Inverse Problem Theory also provides the general framework to try to solve a particular demand, but this frame is empty. Here is where Physics comes in help: it contributes by giving the reasoning between the actors, the spirit of the empty box; this role is so precious that the entire machinery will fail in presence of the wrong modelling. Finally, the computing element appears, which acts as the body of this creation and allows the machine to move. We have to remark that this is just an example framework and actually the "motivation" can spark by any of the three pillars. This happened many times during this research.

The thesis started with a brief description of the framework, composed of theoretical elements from physics, applied mathematics and computing paradigms. Starting from chapter 2, many solutions have been proposed to the community, "attacking" the field from many gates.

Firstly a larger computational field of view has been studied and proposed as a way to reduce the number of probes in a ptychography scan. The RPIE reconstruction algorithm is quite unknown in the community and a mature implementation is strategic; if the computational view is enlarged, then larger areas can be reconstructed, or the same area can be observed with a sparser scan. This in turn produces lighter datasets that can be acquired and reconstructed in shorter time, with a series of other positive consequences (rapid parameter tweaking, higher productivity, etc).

Indeed, it has to be mentioned that the amount of data that are currently acquired by an imaging beamline is currently so large that there is an important trend among many synchrotron facilities involving data compression themes. Many proposed approaches employ a lossy compression algorithm for each frame. This is especially true for the online processing of computed tomography and X-ray diffraction data. This is the reason why, the idea of compressed sensing takes the lead, as it proposes to avoid collecting more information than what is required, as in any case part of it will be washed out by a lossy compression method.

Being unknown to the community, the RPIE reconstruction algorithm was not adapted yet in the context of a partial coherence environment, such as the one where this research has been carried out. Then, after having implemented a multi-mode EPIE reconstruction algorithm, it seems perfectly reasonable to extend this important component also to RPIE, producing what in this thesis is called "M-RPIE". The implications of such extension are huge, as we observed a great reduction in ringing artefacts, compared to the M-EPIE counterpart. These ring artefacts essentially determine areas of large uncertainty in the local phase retrieval.

Considering EPIE, M-EPIE, RPIE and M-RPIE we can observe again the intermixing of the three pillars in action: RPIE methods are superior to EPIEs as they keep advanced elements from the optimisation theory: in the text, it was surely described the role of e.g. regularisation in a reconstruction algorithm. On the other hand, in a real environment, it is the better modelling of the multimode M-*PIE variants that provides better reconstructions than the one obtained using single-mode algorithms. Advanced modelling is a key element provided by physics insights. The Computing element is essential to make alive an algorithm that is refined by both mathematical processing and physics insights. As an example, if the algorithm can not run at a reasonable speed is completely useless in a "fast pacing reality" of an imaging beamline. The Computing element has not to be trivialised or misunderstood: it is not only composed by programming principles but is a set of necessary knowledge spanning also e.g. logic and numerical methods.

Another important aspect of this research is related to increasing the setup invariance of the methods, in order to cope with the so-called "incoherence" in the setup conditions. Ptychography, as many computational techniques, requires dozen of parameters to be tweaked and adapted on a dataset basis.

The subtle role of this set of metadata is to describe the reality of an experiment and to contextualise it. When they are correct, we can effectively extract more information from the actual data at hand.

As it has been described in the text, two are the more critical parameters in a ptychography dataset: probe positions and propagation distance. Historically probe positions have been tackled first and a corpus of methods is available in the literature; however, in our experiments the scarce ready-to-use solutions were unsatisfying.

As the research evolved three methods have been proposed: a metric based approach, an iterative approach and a Deep Learning inspired auto-regressive method. The use of such methods is essential in a real case scenario, as the reconstructions are often useless with no correction applied.

Said that, in the presence of an outcome that can be so binary, the fact of really producing a reconstruction will affect in turn all the fields which benefit from this microscopy technique, from electrochemistry and nanomaterials to biology and medicine.

The importance of such methods is strengthened by their vast applicability: an alignment approach similar to the one proposed for ptychography has been also suggested to the computed tomography community, providing a fast method for nanotomography adjustment. As described in the text, also in this application the computing element shined, as a fast implementation is required for real use cases.

The second important parameter in a ptychography dataset is the propagation distance: if the positions are typically or automatically adjusted or not adjusted at all (except for very small datasets), typically propagation distance is manually tweaked by researchers to obtain a good reconstruction that appears "in focus". In this thesis, a method was proposed for the automatic adjustments of such a critical parameter.

All the proposed algorithms and methods have been tested firstly on synthetic test data and then on real datasets acquired at the TwinMic spectromicroscopy beamline. The proposed solutions are concretised in a software framework that is provided as open-source, called SciComPty, which can be downloaded from https://github.com/ElettraSciComp/SciComPty/, currently already available to the TwinMic users.

Multi-disciplinary research is capable of auto-inspire itself. In this thesis, there are many examples of this. Fourier ptychography is a slightly different form of ptychography which is currently vastly employed in the optical microscopist community. As the pool of users is larger than the synchrotron microscopy one, the field is growing faster, even if born later. Due to its particular configuration, this field provides an easier implementation of Deep Learning algorithms, which are nowadays relevant in modern applied science. Even if these two flavours of ptychography are different, the reciprocity principle assures that many of thier actors are symmetric and related. It is then natural to think about new solutions which can be prototyped easily in a Fourier Ptychography environment and then eventually ported to a synchrotron beamline. This is not a new concept, as, in essence, we are enduring the same kind of interlinking that is routinely observed between applications of electronic, optical and x-ray microscopy.

In this thesis a Deep Prior method was proposed to provide a fast preview (less than 1 s) for a Fourier Ptychography dataset. This preview is a suboptimal reconstruction – it is indeed a preview – but it is critical to tweak the illumination parameters and other microscope settings. In this thesis it was also discussed how to seed a conventional iterative Fourier Ptychography reconstruction with this suboptimal preview: the final quality of the reconstruction is largely higher, with fewer artefacts and higher contrast.

Other examples of the auto-inspiration power of the research have been discussed: alignment procedures have also been briefly discussed for X-ray Fluorescence (energy calibration), super-resolution for forensics, and face recognition.

Chapter 7

Future directions

In the previous chapter, the modelling part was described as essential for a good reconstruction algorithm, as it provides the rationale and the real insight in a bag of formulae. However, with the advent of Deep Learning, the tendency is to relax the need for a physical engine. The implications of this approach are profound and at the same time vast: if on one hand the approach is opening many fields to "anyone" who can grab a bunch of data, providing a cluster of solutions that are effectively driving our modern age, it also has many drawbacks. Not only "brainless" solutions are dangerous as we are effectively impoverishing our critical thinking, but we are effectively locking ourselves to the data at hand.

Actually, this is not entirely true: in order to correctly train a system, synthetic data are essential, and in order to produce them a forward model built on physics insights is required. It has to be also noted that many Deep Learning algorithms are currently trained within a mixed procedure, which combines both supervised and unsupervised learning.Physics insight

Then, the future directions of this research will surely apply some Deep Learning concepts; these concepts will be briefly discussed in the following.

During my research activity, indeed, four different beamtime proposals have been submitted, but only two of them have been actually performed. The experiments that have been already carried out were used to test our methodologies, reconstruction and control software and to acquire a large amount of ptychography dataset to continue on the research on the algorithms. In particular the first one was focused on the introduction of automatic methods for retrieving setup parameters and resulted in part of what was described in chapter three of this thesis, while a second one served to gather many ptychography datasets. The algorithm preparation for the two additional remaining beamtimes will form the basis for new developments.

One of these, represents a natural future development of my research in ptychography, with its focus on multi-slice approaches: the first step will be to adapt the forward model (the simulator) and the reconstruction algorithms to a multi-slice formulation [54], allowing to gather the partial 3D information encoded in the diffraction patterns. This approach has already been investigated in the literature, but currently no ready-to-be-used implementations are available, thus limiting individual research.

The second more appealing approach involves the research on less computationally expensive methodologies to accomplish a 3D reconstruction, using a common thin plane approximation: a reconstruction is carried out for any depth plane in the specimen; subsequently an intelligent algorithm is used to remove the artefacts on each plane and segment only the in-focus part of the image. From this information, one can fuse each plane to obtain a single large depth-of-field 2D image or create a 3D object. We started the research on Deep Learning based algorithms as well as traditional image processing techniques to accomplish this; initial promising results are shown in figure 7.1.

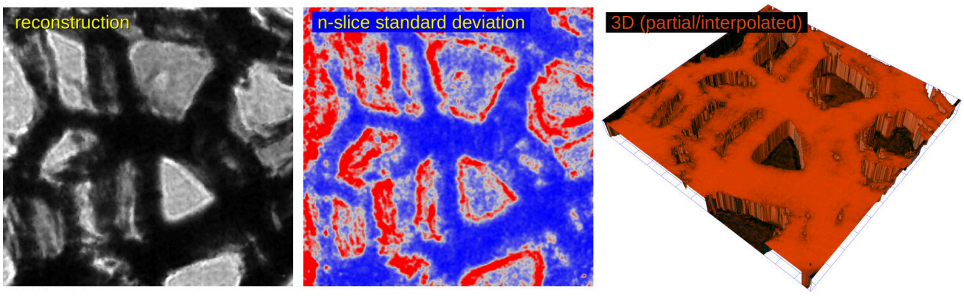


Figure 7.1: Ongoing work of multi-slice reconstructions for revealing depth information and 3D. Tomato-stem sample scanned in TwinMic (proof-of-concept/feasibility).

Another future direction will be related to compressive sensing techniques: a second critical goal for new developments considers the fusion of compressive sensing techniques with the current ptychography methodologies. The possibilities are threefold; the simplest approach consists in locally relaxing the overlap constraints between probes, allowing an extremely large field of view to be acquired (and reconstructed) in a shorter time. A second research path, for which we received a granted beamtime, is in variable dynamic range acquisition: a high dynamic range probe is acquired only "where needed", by automatically selecting how many exposures for probe one will obtain. Again this goes in the direction of reducing the acquisition time. A third way consists in embedding typical L1 norm or greedy algorithms to reconstruct a high-resolution version of each object ROI, similarly with what is currently done in compressive sensing based super-resolution.

The research in Fouerier Ptychography will also continue, involving the creation of true single-shot algorithms, eventually translating the results to real-space ptychography.

Computed tomography will also benefit from the introduction of an autogradient method, to retrieve the tomographic system parameters automatically. A second path is also possible: nanobeads tracking is currently an open question; deep learning will surely provide interesting results

Appendix A

Test images

In this appendix all the test images presented in the text are shown. These test images are extensively used among researchers (and are a de-facto standard). Some notes:

Cameraman, has a distinct Fourier transform that can be quite easily spotted among the literature (the legs of the tripod produces very well defined lines). The coat is not entirely black: applying contrast stretching the buttons appear, indeed the image has been historically used to study dynamic range effects.

Barbara is another interesting test sample, as it is used to check for the high frequency effects of a process: the striped tablecloth allows to easily spot signs of aliasing, or to check if a given process maintains high frequencies. For example, a Total Variation approach can be tweaked on the quantity of "stripes" which are passing through the process. It has been extensively used in many entropy coding works.

Lena is *The test sample.* It has lots of interesting properties, such as the black border of the mirror that represents an hard edge which allows interesting analysis on the reconstruction algorithms and is also easily spotted in a Fourier Transform. It is full of high frequency information due to the feathers and the reticulate part on the hat. There is also an extremely important effect: humans gives extremely attention to faces and *Pareidolia* is one of the most obvious consequences; thus presenting test images with faces allows to easily determine if a results is pleasant or not.

It has to be noted, though, that these test samples are currently technically a bit outdated: we have to think that the digital images that we can easily find on the internet or just by invoking the command camera() or lena() in

Matlab or scikit-image are the digital acquisition of the print, or in the best case of the negative of the aforementioned photos: those images were then digitised with the technology of the late '70s. We can say that these images have outlived their utility on technical ground.

The image processing community need a standardised set of new images. This problem seems even more paradoxical as in the age of Deep Learning lots of datasets are available. However, such datasets are typically made of images of low quality and low resolution, as they are adapted to be consumed by a DNN.

Some of these images are further outdated with regards to the sociological tissue sense: due to its origins (it was a cropped by of a Playboy centrefold), the use of the Lena image is nowadays considered offensive, as it destroys any attempt to encourage diversity and respect in Science and Technology. Despite its historical significance, "the day has come to say farewell to Lena and move on to other more appropriate test images", as reported in a famous SPIE editorial [1].

Currently, many journals do not accept any more manuscripts which use the test image. Maybe this will help in accelerating the community to create a better new dataset, capable of guiding us towards new technological advances.

---

[1] Optical Engineering, 57(12), 120101 (2018). https://doi.org/10.1117/1.OE.57.12.120101

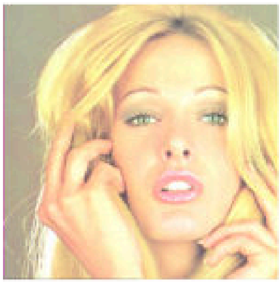

cameraman


f16

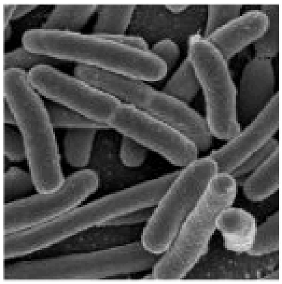

bridge


barbara


man

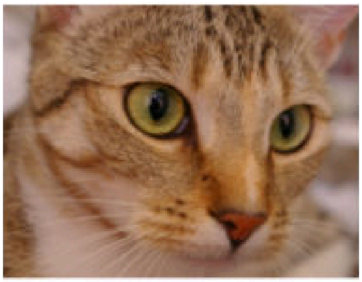

tiffany


astronaut


bacteria


coins


westconcordphoto


chelsea


bridge


onion


lena


pears

Bibliography

[1] F. Guzzi, G. Kourousias, F. Billè, R. Pugliese, A. Gianoncelli, and S. Carrato, “A modular software framework for the design and implementation of ptychography algorithms.” submitted to PeerJ Computer Science.

[2] A. Gianoncelli, V. Bonanni, G. Gariani, F. Guzzi, L. Pascolo, R. Borghes, F. Billè, and G. Kourousias, “Soft x-ray microscopy techniques for medical and biological imaging at twinmic—elettra,” *Applied Sciences*, vol. 11, no. 16, 2021.

[3] F. Guzzi, G. Kourousias, F. Billè, R. Pugliese, C. Reis, A. Gianoncelli, and S. Carrato, “Refining scan positions in ptychography through error minimisation and potential application of machine learning,” *Journal of Instrumentation*, vol. 13, pp. C06002–C06002, June 2018.

[4] F. Guzzi, G. Kourousias, F. Billè, R. Pugliese, A. Gianoncelli, and S. Carrato, “A parameter refinement method for ptychography based on deep learning concepts.” submitted to Optica, preprint https://arxiv.org/abs/2105.08058.

[5] F. Guzzi, G. Kourousias, F. Billè, R. Pugliese, A. Gianoncelli, and S. Carrato, “Start better: Deep prior solution seeding for Fourier Ptychography microscopy.” accepted, Optoelectronics and photonics special session at the 44th international convention on information communication and electronic technology, September 27th - October 1st 2021, Opatija, Croatia.

[6] F. Guzzi, G. D. Credico, G. Kourousias, F. Billè, R. Pugliese, A. Gianoncelli, and S. Carrato, “Signal alignment problems on multi-element x-ray fluorescence detectors.” submitted to International Conference on Applications in Electronics Pervading Industry Environment and Society, September 16-17 2021, Pisa, Italy.

[7] F. Guzzi, G. Kourousias, A. Sorrentino, L. Pasocolo, A. Gianoncelli, F. Billè, R. Pugliese, and S. Carrato, “A fast method for tomography alignment.” in preparation for Applied Science.

[8] G. Guarnieri, M. Fontani, F. Guzzi, S. Carrato, and M. Jerian, “Perspective registration and multi-frame super-resolution of license plates in surveillance videos,” *Forensic Science International: Digital Investigation*, vol. 36, p. 301087, 2021.

[9] F. Guzzi, L. De Bortoli, R. S. Molina, S. Marsi, S. Carrato, and G. Ramponi, “Distillation of an end-to-end oracle for face verification and recognition sensors,” *Sensors*, vol. 20, no. 5, 2020.

[10] F. Guzzi, L. DeBortoli, S. Marsi, S. Carrato, and G. Ramponi, “Distillation of a CNN for a high accuracy mobile face recognition system,” in *42nd International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2019, Opatija, Croatia, May 20-24, 2019*, pp. 989–994, 2019.

[11] S. Marsi, L. De Bortoli, F. Guzzi, J. Bhattacharya, F. Cicala, S. Carrato, A. Canziani, and G. Ramponi, “A face recognition system using off-the-shelf feature extractors and an ad-hoc classifier,” in *Applications in Electronics Pervading Industry, Environment and Society* (S. Saponara and A. De Gloria, eds.), (Cham), pp. 145–151, Springer International Publishing, 2019.

[12] J. Bhattacharya, F. Guzzi, S. Marsi, S. Carrato, and G. Ramponi, “Real-time dnn-based face identification for the blind,” in *Applications in Electronics Pervading Industry, Environment and Society* (A. De Gloria, ed.), (Cham), pp. 119–125, Springer International Publishing, 2019.

[13] L. De Bortoli, F. Guzzi, S. Marsi, S. Carrato, and G. Ramponi, “A fast face recognition cnn obtained by distillation,” in *Applications in Electronics Pervading Industry, Environment and Society* (S. Saponara and A. De Gloria, eds.), (Cham), pp. 341–347, Springer International Publishing, 2020.

[14] S. Marsi, S. Carrato, L. De Bortoli, P. Gallina, F. Guzzi, and G. Ramponi, “An fpga realization for real-time depth estimation in image sequences,” in *Applications in Electronics Pervading Industry, Environment and Society* (S. Saponara and A. De Gloria, eds.), (Cham), pp. 489–495, Springer International Publishing, 2020.

[15] X. Hu, J. Wu, J. Suo, and Q. Dai, “Emerging theories and technologies on computational imaging,” *Frontiers Inf. Technol. Electron. Eng.*, vol. 18, no. 9, pp. 1207–1221, 2017.

[16] M. W. M. Jones, K. Elgass, M. D. Junker, M. B. Luu, M. T. Ryan, A. G. Peele, and G. A. van Riessen, "Mapping biological composition through quantitative phase and absorption x-ray ptychography," *Scientific Reports*, vol. 4, p. 6796, Oct 2014.

[17] J. Miao, P. Charalambous, J. Kirz, and D. Sayre, "Extending the methodology of x-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens," *Nature*, vol. 400, pp. 342–344, Jul 1999.

[18] D. Paganin, *Coherent X-ray Optics*. Oxford: Oxford University Press, 2006.

[19] J. A. D. McMorrow, *Elements of Modern X-ray Physics, Second Edition*. Chicago: John Wiley and Sons, 2011.

[20] S. Mobilio, F. Boscherini, and C. Meneghini, eds., *Synchrotron Radiation, Basics, Methods and Applications*. Berlin: Springer Berlin Heidelberg, 2015.

[21] M. Antonelli, *Photon Beam-Position Monitor basati su diamante e quantum well per sorgenti di luce di terza e quarta generazione*. PhD thesis, XXV Ciclo del dottorato di ricerca in ingegneria dell'informazione, Università degli studi di Trieste, 2012.

[22] C. R. Vogel, *Computational Methods for Inverse Problems*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2002.

[23] R. C. Gonzalez and R. E. Woods, *Digital Image Processing, 4th Editions*. London, England: Pearsons, 2002.

[24] A. Ribes and F. Schmitt, "Linear inverse problems in imaging," *IEEE Signal Processing Magazine*, vol. 25, no. 4, pp. 84–99, 2008.

[25] G. Ramponi and S. Carrato, "An adaptive irregular sampling algorithm and its application to image coding," *Image Vis. Comput.*, vol. 19, no. 7, pp. 451–460, 2001.

[26] G. Kourousias, F. Billè, R. Borghes, A. Alborini, S. Sala, R. Alberti, and A. Gianoncelli, "Compressive sensing for dynamic xrf scanning," *Scientific Reports*, vol. 10, p. 9990, Jun 2020.

[27] G. Pratx and L. Xing, "GPU computing in medical physics: A review," *Medical Physics*, vol. 38, no. 5, pp. 2685–2697, 2011.

[28] J. Nickolls and W. J. Dally, "The GPU computing era," *IEEE Micro*, vol. 30, no. 2, pp. 56–69, 2010.

[29] D. B. Kirk and W. mei W. Hwu, *Programming Massively Parallel Processors, A Hands-on Approach.* Waltham, USA: Elsevier, 2017.

[30] J. Sanders and E. Kandrot, *CUDA by Example.* Upper Saddle River, USA: Addison-Wesley, 2011.

[31] C. A. Balanis, *Advanced Engineering Electromagnetics, 2nd Edition.* Chicago: John Wiley and Sons, 2012.

[32] G. J. W, *Introduction to Fourier optics, 3rd ed.* USA: W.H.Freeman and Co Ltd, 2005.

[33] P. Russo, *Handbook of X-ray Imaging.* CRC Press, Dec. 2017.

[34] D. M. Paganin and D. Pelliccia, "Tutorials on x-ray phase contrast imaging: Some fundamentals and some conjectures on future developments." arXiv: 1902.00364, 2019.

[35] E. Hecht and A. Zajac, *Optics.* Boston, USA: Addison-Wesley, 1974.

[36] J. D. Schmidt, *Numerical simulation of optical wave propagation with examples in MATLAB.* Bellingham, Washington: SPIE, 2010.

[37] C. Jacobsen, *X-ray Microscopy.* Cambridge, UK: Cambridge University Press, 2019.

[38] P. Thibault, M. Dierolf, A. Menzel, O. Bunk, C. David, and F. Pfeiffer, "High-resolution scanning x-ray diffraction microscopy," *Science*, vol. 321, no. 5887, pp. 379–382, 2008.

[39] K. et al., *Center for X-ray Optics and Advanced Light Source X-RAY DATA BOOKLET.* Berkeley, California: Lawrence Berkeley National Laboratory, 2001.

[40] G. J. Williams, H. M. Quiney, A. G. Peele, and K. A. Nugent, "Fresnel coherent diffractive imaging: treatment and analysis of data," *New Journal of Physics*, vol. 12, p. 035020, mar 2010.

[41] I. Vartanyants and I. Robinson, "Origins of decoherence in coherent x-ray diffraction experiments," *Optics Communications*, vol. 222, no. 1, pp. 29 – 50, 2003.

[42] T. Latychevskaia, G. F. Mancini, and F. Carbone, "The role of the coherence in the cross-correlation analysis of diffraction patterns from two-dimensional dense mono-disperse systems," *Scientific Reports*, vol. 5, p. 16573, Nov 2015.

[43] B. Chen, B. Abbey, R. Dilanian, E. Balaur, G. van Riessen, M. Junker, C. Q. Tran, M. W. M. Jones, A. G. Peele, I. McNulty, D. J. Vine, C. T. Putkunz, H. M. Quiney, and K. A. Nugent, "Diffraction imaging: The limits of partial coherence," *Phys. Rev. B*, vol. 86, p. 235401, Dec 2012.

[44] P. Thibault and A. Menzel, "Reconstructing state mixtures from diffraction measurements," *Nature*, vol. 494, pp. 68–71, Feb 2013.

[45] A. Gianoncelli, G. R. Morrison, B. Kaulich, D. Bacescu, and J. Kovac, "Scanning transmission x-ray microscopy with a configurable detector," *Applied Physics Letters*, vol. 89, no. 25, p. 251117, 2006.

[46] J. Kirz, C. Jacobsen, and M. Howells, "Soft x-ray microscopes and their biological applications," *Quarterly Reviews of Biophysics*, vol. 28, no. 1, p. 33–130, 1995.

[47] F. Zhang, I. Peterson, J. Vila-Comamala, A. Diaz, F. Berenguer, R. Bean, B. Chen, A. Menzel, I. K. Robinson, and J. M. Rodenburg, "Translation position determination in ptychographic coherent diffraction imaging," *Opt. Express*, vol. 21, pp. 13592–13606, Jun 2013.

[48] M. D. de Jonge, B. Hornberger, C. Holzner, D. Legnini, D. Paterson, I. McNulty, C. Jacobsen, and S. Vogt, "Quantitative phase imaging with a scanning transmission x-ray microscope," *Phys. Rev. Lett.*, vol. 100, p. 163902, Apr 2008.

[49] C. Kottler, C. David, F. Pfeiffer, and O. Bunk, "A two-directional approach for grating based differential phase contrast imaging using hard x-rays," *Opt. Express*, vol. 15, pp. 1175–1181, Feb 2007.

[50] M. W. Jones, B. Abbey, A. Gianoncelli, E. Balaur, C. Millet, M. B. Luu, H. D. Coughlan, A. J. Carroll, A. G. Peele, L. Tilley, and G. A. van Riessen, "Phase-diverse fresnel coherent diffractive imaging of malaria parasite-infected red blood cells in the water window," *Opt. Express*, vol. 21, pp. 32151–32159, Dec 2013.

[51] G. Kourousias, B. Bozzini, A. Gianoncelli, M. W. Jones, M. Junker, G. van Riessen, and M. Kiskinova, "Shedding light on electrodeposition dynamics tracked in situ via soft X-ray coherent diffraction imaging," *Nano Research*, vol. 9, no. 7, pp. 2046–2056, 2016.

[52] G. Kourousias, B. Bozzini, M. W. M. Jones, G. A. Van Riessen, S. Dal Zilio, F. Billè, M. Kiskinova, and A. Gianoncelli, "Monitoring dynamic electrochemical processes with in situ ptychography," *Applied Nanoscience*, vol. 8, pp. 627–636, Apr 2018.

[53] J. Rodenburg, "Ptychography and related diffractive imaging methods," vol. 150 of *Advances in Imaging and Electron Physics*, pp. 87 – 184, Elsevier, 2008.

[54] A. M. Maiden, M. J. Humphry, and J. M. Rodenburg, "Ptychographic transmission microscopy in three dimensions using a multi-slice approach," *J. Opt. Soc. Am. A*, vol. 29, pp. 1606–1614, Aug 2012.

[55] F. Pfeiffer, "X-ray ptychography," *Nature Photonics*, vol. 12, pp. 9–17, Jan 2018.

[56] R. W. Gerchberg and W. O. Saxton, "Practical Algorithm for determination of phase from image and diffraction plane pictures," *OPTIK*, vol. 35, no. 2, pp. 237–&, 1972.

[57] M. W. M. Jones, A. G. Peele, and G. A. van Riessen, "Application of a complex constraint for biological samples in coherent diffractive imaging," *Opt. Express*, vol. 21, pp. 30275–30281, Dec 2013.

[58] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, "Phase retrieval with application to optical imaging: A contemporary overview," *IEEE Signal Processing Magazine*, vol. 32, no. 3, pp. 87–109, 2015.

[59] Y. Bruck and L. Sodin, "On the ambiguity of the image reconstruction problem," *Optics Communications*, vol. 30, no. 3, pp. 304 – 308, 1979.

[60] B. Bozzini, G. Kourousias, and A. Gianoncelli, "In situ observation of dynamic electrodeposition processes by soft x-ray fluorescence microspectroscopy and keyhole coherent diffractive imaging," *Journal of Physics D: Applied Physics*, vol. 50, p. 124001, feb 2017.

[61] J. R. Fienup, "Phase retrieval algorithms: a comparison," *Appl. Opt.*, vol. 21, pp. 2758–2769, Aug 1982.

[62] V. Elser, "Solution of the crystallographic phase problem by iterated projections," *Acta Crystallographica Section A*, vol. 59, pp. 201–209, May 2003.

[63] S. Marchesini, H. He, H. N. Chapman, S. P. Hau-Riege, A. Noy, M. R. Howells, U. Weierstall, and J. C. H. Spence, "X-ray image reconstruction from a diffraction pattern alone," *Phys. Rev. B*, vol. 68, p. 140101, Oct 2003.

[64] H. M. Quiney, A. G. Peele, Z. Cai, D. Paterson, and K. A. Nugent, "Diffractive imaging of highly focused x-ray fields," *Nature Physics*, vol. 2, pp. 101–104, Feb 2006.

[65] G. J. Williams, H. M. Quiney, B. B. Dhal, C. Q. Tran, K. A. Nugent, A. G. Peele, D. Paterson, and M. D. de Jonge, “Fresnel coherent diffractive imaging,” *Phys. Rev. Lett.*, vol. 97, p. 025506, Jul 2006.

[66] B. Abbey, K. A. Nugent, G. J. Williams, J. N. Clark, A. G. Peele, M. A. Pfeifer, M. de Jonge, and I. McNulty, “Keyhole coherent diffractive imaging,” *Nature Physics*, vol. 4, pp. 394–398, May 2008.

[67] B. Abbey, G. J. Williams, M. A. Pfeifer, J. N. Clark, C. T. Putkunz, A. Torrance, I. McNulty, T. M. Levin, A. G. Peele, and K. A. Nugent, “Quantitative coherent diffractive imaging of an integrated circuit at a spatial resolution of 20 nm,” *Applied Physics Letters*, vol. 93, no. 21, p. 214101, 2008.

[68] J. N. Clark, G. J. Williams, H. M. Quiney, L. Whitehead, M. D. de Jonge, E. Hanssen, M. Altissimo, K. A. Nugent, and A. G. Peele, “Quantitative phase measurement in coherent diffraction imaging,” *Opt. Express*, vol. 16, pp. 3342–3348, Mar 2008.

[69] D. J. Vine, G. J. Williams, B. Abbey, M. A. Pfeifer, J. N. Clark, M. D. de Jonge, I. McNulty, A. G. Peele, and K. A. Nugent, “Ptychographic fresnel coherent diffractive imaging,” *Phys. Rev. A*, vol. 80, p. 063823, Dec 2009.

[70] H. Quiney, “Coherent diffractive imaging using short wavelength light sources,” *Journal of Modern Optics*, vol. 57, no. 13, pp. 1109–1149, 2010.

[71] R. Heine, T. Gorniak, T. Nisius, C. Christophis, M. Pettitt, F. Staier, T. Wilhein, S. Rehbein, M. Grunze, and A. Rosenhahn, “Digital in-line x-ray holography with zone plates,” *Ultramicroscopy*, vol. 111, no. 8, pp. 1131 – 1136, 2011.

[72] C. T. Putkunz, J. N. Clark, D. J. Vine, G. J. Williams, M. A. Pfeifer, E. Balaur, I. McNulty, K. A. Nugent, and A. G. Peele, “Phase-diverse coherent diffractive imaging: High sensitivity with low dose,” *Phys. Rev. Lett.*, vol. 106, p. 013903, Jan 2011.

[73] E. Wolf, “New theory of partial coherence in the space–frequency domain. part i: spectra and cross spectra of steady-state sources,” *J. Opt. Soc. Am.*, vol. 72, pp. 343–351, Mar 1982.

[74] J. Hagemann and T. Salditt, “Coherence-resolution relationship in holographic and coherent diffractive imaging,” *Opt. Express*, vol. 26, pp. 242–253, Jan 2018.

[75] P. Thibault and A. Menzel, “Reconstructing state mixtures from diffraction measurements,” *Nat.*, vol. 494, no. 7435, pp. 68–71, 2013.

[76] G. J. Williams, H. M. Quiney, A. G. Peele, and K. A. Nugent, "Coherent diffractive imaging and partial coherence," *Phys. Rev. B*, vol. 75, p. 104102, Mar 2007.

[77] K. A. Nugent, "Coherent methods in the x-ray sciences," *Advances in Physics*, vol. 59, no. 1, pp. 1–99, 2010.

[78] N. Burdet, X. Shi, D. Parks, J. N. Clark, X. Huang, S. D. Kevan, and I. K. Robinson, "Evaluation of partial coherence correction in x-ray ptychography," *Opt. Express*, vol. 23, pp. 5452–5467, Mar 2015.

[79] D. H. Parks, X. Shi, and S. D. Kevan, "Partially coherent x-ray diffractive imaging of complex objects," *Phys. Rev. A*, vol. 89, p. 063824, Jun 2014.

[80] F. van der Veen and F. Pfeiffer, "Coherent x-ray scattering," *Journal of Physics: Condensed Matter*, vol. 16, pp. 5003–5030, jul 2004.

[81] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: a technical overview," *IEEE Signal Processing Magazine*, vol. 20, pp. 21–36, May 2003.

[82] B. Jähne, "EMVA 1288 standard for machine vision," *Optik & Photonik*, vol. 5, no. 1, pp. 53–54, 2010.

[83] M. Konnik and J. Welsh, "High-level numerical simulations of noise in CCD and CMOS photosensors: review and tutorial." arXiv: 1412.4031, 2014.

[84] P. Thibault and M. Guizar-Sicairos, "Maximum-likelihood refinement for coherent diffractive imaging," *New Journal of Physics*, vol. 14, p. 063004, jun 2012.

[85] C. A. C. of Excellence for coherent X-ray Science, "Nadia Software Project Page." `http://cxscode.ph.unimelb.edu.au/`, 2012. [Online; accessed 25-January-2021].

[86] B. McCallum and J. Rodenburg, "Two-dimensional demonstration of wigner phase-retrieval microscopy in the stem configuration," *Ultramicroscopy*, vol. 45, no. 3, pp. 371–380, 1992.

[87] H. N. Chapman, "Phase-retrieval x-ray microscopy by wigner-distribution deconvolution," *Ultramicroscopy*, vol. 66, no. 3, pp. 153–172, 1996.

[88] J. M. Rodenburg, , and R. H. T. Bates, "The theory of super-resolution electron microscopy via wigner-distribution deconvolution," *Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences*, vol. 339, pp. 521–553, June 1992.

[89] P. Li, T. B. Edo, and J. M. Rodenburg, “Ptychographic inversion via wigner distribution deconvolution: Noise suppression and probe design,” *Ultramicroscopy*, vol. 147, pp. 106–113, 2014.

[90] J. M. Rodenburg and H. M. L. Faulkner, “A phase retrieval algorithm for shifting illumination,” *Applied Physics Letters*, vol. 85, no. 20, pp. 4795–4797, 2004.

[91] A. M. Maiden and J. M. Rodenburg, “An improved ptychographical phase retrieval algorithm for diffractive imaging,” *Ultramicroscopy*, vol. 109, no. 10, pp. 1256–1262, 2009.

[92] O. Mandula, M. Elzo Aizarna, J. Eymery, M. Burghammer, and V. Favre-Nicolin, “*PyNX.Ptycho*: a computing library for X-ray coherent diffraction imaging of nanostructures,” *Journal of Applied Crystallography*, vol. 49, pp. 1842–1848, Oct 2016.

[93] D. P. Bertsekas, “Incremental proximal methods for large scale convex optimization,” *Math. Program.*, vol. 129, p. 163–195, Oct. 2011.

[94] A. Maiden, D. Johnson, and P. Li, “Further improvements to the ptychographical iterative engine,” *Optica*, vol. 4, pp. 736–745, Jul 2017.

[95] P. Li, T. Edo, D. Batey, J. Rodenburg, and A. Maiden, “Breaking ambiguities in mixed state ptychography,” *Opt. Express*, vol. 24, pp. 9038–9052, Apr 2016.

[96] M. Odstrcil, P. Baksh, S. A. Boden, R. Card, J. E. Chad, J. G. Frey, and W. S. Brocklesby, “Ptychographic coherent diffractive imaging with orthogonal probe relaxation,” *Opt. Express*, vol. 24, pp. 8360–8369, Apr 2016.

[97] L. W. Whitehead, G. J. Williams, H. M. Quiney, D. J. Vine, R. A. Dilanian, S. Flewett, K. A. Nugent, A. G. Peele, E. Balaur, and I. McNulty, “Diffractive imaging using partially coherent x rays,” *Phys. Rev. Lett.*, vol. 103, p. 243902, Dec 2009.

[98] D. J. Batey, D. Claus, and J. M. Rodenburg, “Information multiplexing in ptychography,” *Ultramicroscopy*, vol. 138, pp. 13 – 21, 2014.

[99] M. Stockmar, P. Cloetens, I. Zanette, B. Enders, M. Dierolf, F. Pfeiffer, and P. Thibault, “Near-field ptychography: phase retrieval for inline holography using a structured illumination,” *Scientific Reports*, vol. 3, p. 1927, May 2013.

[100] A.-L. Robisch, K. Kröger, A. Rack, and T. Salditt, “Near-field ptychography using lateral and longitudinal shifts,” *New Journal of Physics*, vol. 17, p. 073033, jul 2015.

[101] M. Dierolf, P. Thibault, A. Menzel, C. M. Kewish, K. Jefimovs, I. Schlichting, K. von König, O. Bunk, and F. Pfeiffer, "Ptychographic coherent diffractive imaging of weakly scattering specimens," *New Journal of Physics*, vol. 12, p. 035017, mar 2010.

[102] T. B. Edo, D. J. Batey, A. M. Maiden, C. Rau, U. Wagner, Z. D. Pešić, T. A. Waigh, and J. M. Rodenburg, "Sampling in x-ray ptychography," *Phys. Rev. A*, vol. 87, p. 053850, May 2013.

[103] J. Rodenburg and A. Maiden, *Ptychography*, pp. 819–904. Cham: Springer International Publishing, 2019.

[104] C. Jacobsen, J. Deng, and Y. Nashed, "Strategies for high-throughput focused-beam ptychography," *Journal of Synchrotron Radiation*, vol. 24, pp. 1078–1081, Sep 2017.

[105] A. Gianoncelli, G. Kourousias, M. Altissimo, D. E. Bedolla, L. Merolle, A. Stolfa, and H.-J. Shin, "Combining multiple imaging techniques at the twinmic x-ray microscopy beamline," *AIP Conference Proceedings*, vol. 1764, no. 1, p. 030002, 2016.

[106] A. Gianoncelli, F. Cammisuli, M. Altissimo, M. Salomé, O. Radillo, G. Ricci, S. Giordani, C. Rizzardi, and L. Pascolo, "Iron-related toxicity effects of single-walled carbon nanotubes in human placental cells (bewo) investigated by x-ray fluorescence microscopy," *X-Ray Spectrometry*, vol. 48, no. 5, pp. 413–421, 2019.

[107] F. Cammisuli, S. Giordani, A. Gianoncelli, C. Rizzardi, L. Radillo, M. Zweyer, T. Da Ros, M. Salomé, M. Melato, and L. Pascolo, "Iron-related toxicity of single-walled carbon nanotubes and crocidolite fibres in human mesothelial cells investigated by synchrotron xrf microscopy," *Scientific Reports*, vol. 8, p. 706, Jan 2018.

[108] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32: NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada* (H. M. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. B. Fox, and R. Garnett, eds.), pp. 8024–8035, 2019.

[109] A. Shenfield and J. M. Rodenburg, "Evolutionary determination of experimental parameters for ptychographical imaging," *Journal of Applied Physics*, vol. 109, no. 12, p. 124510, 2011.

[110] F. Hüe, J. Rodenburg, A. Maiden, and P. Midgley, "Extended ptychography in the transmission electron microscope: Possibilities and limitations," *Ultramicroscopy*, vol. 111, no. 8, pp. 1117 – 1123, 2011.

[111] M. Guizar-Sicairos and J. R. Fienup, "Phase retrieval with transverse translation diversity: a nonlinear optimization approach," *Opt. Express*, vol. 16, pp. 7264–7278, May 2008.

[112] A. Tripathi, I. McNulty, and O. G. Shpyrko, "Ptychographic overlap constraint errors and the limits of their numerical recovery using conjugate gradient descent methods," *Opt. Express*, vol. 22, pp. 1452–1466, Jan 2014.

[113] M. J. D. Powell, "An efficient method for finding the minimum of a function of several variables without calculating derivatives," *The Computer Journal*, vol. 7, pp. 155–162, 01 1964.

[114] P. Dwivedi, A. Konijnenberg, S. Pereira, and H. Urbach, "Lateral position correction in ptychography using the gradient of intensity patterns," *Ultramicroscopy*, vol. 192, pp. 29 – 36, 2018.

[115] A. Maiden, M. Humphry, M. Sarahan, B. Kraus, and J. Rodenburg, "An annealing algorithm to correct positioning errors in ptychography," *Ultramicroscopy*, vol. 120, pp. 64 – 72, 2012.

[116] D. Gürsoy, Y. P. Hong, K. He, K. Hujsak, S. Yoo, S. Chen, Y. Li, M. Ge, L. M. Miller, Y. S. Chu, V. De Andrade, K. He, O. Cossairt, A. K. Katsaggelos, and C. Jacobsen, "Rapid alignment of nanotomography data using joint iterative reconstruction and reprojection," *Scientific Reports*, vol. 7, p. 11818, Sep 2017.

[117] M. Guizar-Sicairos, S. T. Thurman, and J. R. Fienup, "Efficient subpixel image registration algorithms," *Opt. Lett.*, vol. 33, pp. 156–158, Jan 2008.

[118] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[119] W. Pratt, *Digital Image Processing, 3rd edition.* Los Altos,USA: John Wiley and Sons, 2001.

[120] A. Bovik, *Handbook of Image and Video Processing.* San Diego, U.S.A: Academic Press, 2000.

[121] A. G. et al., *Image Registration,Principles, Tools and Methods.* London, UK: Springer-Verlag, 2012.

[122] L. G. Brown, "A survey of image registration techniques," *ACM Comput. Surv.*, vol. 24, p. 325–376, Dec. 1992.

[123] G. de Haan, P. W. A. C. Biezen, H. Huijgen, and O. A. Ojo, "True-motion estimation with 3-d recursive search block matching," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 3, no. 5, pp. 368–379, 1993.

[124] G. De Haan, "Progress in motion estimation for consumer video format conversion," *IEEE Transactions on Consumer Electronics*, vol. 46, no. 3, pp. 449–459, 2000.

[125] G. de Haan, "Progress in motion estimation for consumer video format conversion," in *2000 Digest of Technical Papers. International Conference on Consumer Electronics. Nineteenth in the Series (Cat. No.00CH37102)*, pp. 50–51, 2000.

[126] H. Foroosh, J. B. Zerubia, and M. Berthod, "Extension of phase correlation to subpixel registration," *IEEE Transactions on Image Processing*, vol. 11, no. 3, pp. 188–200, 2002.

[127] M. A. Viergever, J. A. Maintz, S. Klein, K. Murphy, M. Staring, and J. P. Pluim, "A survey of medical image registration – under review," *Medical Image Analysis*, vol. 33, pp. 140 – 144, 2016. 20th anniversary of the Medical Image Analysis journal (MedIA).

[128] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? a new look at signal fidelity measures," *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 98–117, 2009.

[129] Z. Wang and A. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81–84, 2002.

[130] L. Waller and L. Tian, "Machine learning for 3D Microscopy," *Nature*, vol. 523, pp. 416–417, Jul 2015.

[131] A. Kappeler, S. Ghosh, J. Holloway, O. Cossairt, and A. Katsaggelos, "Ptychnet: CNN based Fourier Ptychography," in *2017 IEEE International Conference on Image Processing (ICIP)*, pp. 1712–1716, 2017.

[132] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pp. 770–778, IEEE Computer Society, 2016.

[133] S. Miao, Z. J. Wang, and R. Liao, "A cnn regression approach for real-time 2d/3d registration," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1352–1363, 2016.

[134] Y. Deng, C. C. Loy, and X. Tang, “Image aesthetic assessment: An experimental survey,” *IEEE Signal Processing Magazine*, vol. 34, no. 4, pp. 80–106, 2017.

[135] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings* (Y. Bengio and Y. LeCun, eds.), 2015.

[136] B. Nikolic, “Acceleration of non-linear minimisation with PyTorch.” arXiv: 1805.07439, 2018.

[137] T.-M. Li, M. Gharbi, A. Adams, F. Durand, and J. Ragan-Kelley, “Differentiable programming for image processing and deep learning in halide,” *ACM Trans. Graph.*, vol. 37, July 2018.

[138] S. Jiang, K. Guo, J. Liao, and G. Zheng, “Solving Fourier ptychographic imaging problems via neural network modeling and tensorflow,” *Biomed. Opt. Express*, vol. 9, pp. 3306–3319, Jul 2018.

[139] 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. Mané, 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. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, “TensorFlow: Large-scale machine learning on heterogeneous systems,” 2015. Software available from tensorflow.org.

[140] S. Kandel, S. Maddali, M. Allain, S. O. Hruszkewycz, C. Jacobsen, and Y. S. G. Nashed, “Using automatic differentiation as a general framework for ptychographic reconstruction,” *Opt. Express*, vol. 27, pp. 18653–18672, Jun 2019.

[141] R. F. H. Fischer, *Precoding and Signal Shaping for Digital Transmission.* Chicago: John Wiley and Sons, 2002.

[142] B. van Merrienboer, O. Breuleux, A. Bergeron, and P. Lamblin, “Automatic differentiation in ML: where we are and where we should be going,” in *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada* (S. Bengio, H. M. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, eds.), pp. 8771–8781, 2018.

[143] M. Bartholomew-Biggs, S. Brown, B. Christianson, and L. Dixon, "Automatic differentiation of algorithms," *Journal of Computational and Applied Mathematics*, vol. 124, no. 1, pp. 171 – 190, 2000. Numerical Analysis 2000. Vol. IV: Optimization and Nonlinear Equations.

[144] A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind, "Automatic differentiation in machine learning: a survey," *J. Mach. Learn. Res.*, vol. 18, pp. 153:1–153:43, 2017.

[145] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, "Spatial transformer networks," in *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada* (C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, eds.), pp. 2017–2025, 2015.

[146] B. Enders and P. Thibault, "A computational framework for ptychographic reconstructions," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 472, no. 2196, p. 20160640, 2016.

[147] J. You, U. Reiter, M. M. Hannuksela, M. Gabbouj, and A. Perkis, "Perceptual-based quality assessment for audio–visual services: A survey," *Signal Processing: Image Communication*, vol. 25, no. 7, pp. 482–501, 2010. Special Issue on Image and Video Quality Assessment.

[148] S. Ruder, "An overview of gradient descent optimization algorithms." arXiv: 1609.04747, 2017.

[149] R. Meyes, M. Lu, C. W. de Puiseau, and T. Meisen, "Ablation studies in artificial neural networks." arXiv: 1901.08644, 2019.

[150] P. E. Lillian, R. Meyes, and T. Meisen, "Ablation of a robot's brain: Neural networks under a knife." arXiv: 1812.05687, 2019.

[151] B. Bozzini, G. Kourousias, A. Gianoncelli, M. W. Jones, G. V. Riessen, and M. Kiskinova, "Soft x-ray ptychography as a tool for in operando morphochemical studies of electrodeposition processes with nanometric lateral resolution," *Journal of Electron Spectroscopy and Related Phenomena*, vol. 220, pp. 147 – 155, 2017. Proceedings of the "39th International Conference on Vacuum Ultraviolet and X-ray Physics (VUVX-2016).

[152] B. Kaulich, D. Bacescu, J. Susini, C. David, E. di Fabrizio, G. Morrison, P. Charalambous, J. Thieme, T. Wilhein, J. Kovac, D. Cocco,

M. Salome, O. Dhez, T. Weitkamp, S. Cabrini, D. Cojoc, A. Gianoncelli, U. Vogt, M. Podnar, M. Zangrando, M. Zacchigna, and M. Kiskinova, *TwinMic - a European twin x-ray microscopy station commissioned at ELETTRA*, pp. 22 – 25. IPAP Conference Series, Institute of Pure and Applied Physics, 2006. 8th International Conference on X-ray Microscopy;Conference, Himeji, Hyogo, Japan, 26-30 July 2005.

[153] A. Maiden, D. Johnson, and P. Li, "Further improvements to the ptychographical iterative engine," *Optica*, vol. 4, pp. 736–745, Jul 2017.

[154] K. Wakonig, A. Diaz, A. Bonnin, M. Stampanoni, A. Bergamaschi, J. Ihli, M. Guizar-Sicairos, and A. Menzel, "X-ray Fourier ptychography," *Science Advances*, vol. 5, no. 2, 2019.

[155] G. Zheng, R. Horstmeyer, and C. Yang, "Wide-field, high-resolution Fourier ptychographic microscopy," *Nature Photonics*, vol. 7, no. 9, pp. 739–745, 2013.

[156] X. Ou, G. Zheng, and C. Yang, "Embedded pupil function recovery for Fourier ptychographic microscopy," *Optics Express*, vol. 22, no. 5, p. 4960, 2014.

[157] K. Guo, S. Dong, P. Nanda, and G. Zheng, "Optimization of sampling pattern and the design of fourier ptychographic illuminator," *Opt. Express*, vol. 23, pp. 6171–6180, Mar 2015.

[158] S. Pacheco, G. Zheng, and R. Liang, "Reflective Fourier ptychography," *Journal of Biomedical Optics*, vol. 21, no. 2, pp. 1 – 6, 2016.

[159] Z. Bian, S. Dong, and G. Zheng, "Adaptive system correction for robust fourier ptychographic imaging," *Opt. Express*, vol. 21, pp. 32400–32410, Dec 2013.

[160] S. Dong, R. Shiradkar, P. Nanda, and G. Zheng, "Spectral multiplexing and coherent-state decomposition in fourier ptychographic imaging," *Biomed. Opt. Express*, vol. 5, pp. 1757–1767, Jun 2014.

[161] A. Pan, Y. Zhang, T. Zhao, Z. Wang, D. Dan, M. Lei, and B. Yao, "System calibration method for Fourier ptychographic microscopy," *Journal of Biomedical Optics*, vol. 22, no. 9, pp. 1 – 11, 2017.

[162] H. Zhang, S. Jiang, J. Liao, J. Deng, J. Liu, Y. Zhang, and G. Zheng, "Near-field fourier ptychography: super-resolution phase retrieval via speckle illumination," *Opt. Express*, vol. 27, pp. 7498–7512, Mar 2019.

[163] P. Li, D. J. Batey, T. B. Edo, and J. M. Rodenburg, "Separation of three-dimensional scattering effects in tilt-series fourier ptychography," *Ultramicroscopy*, vol. 158, pp. 1 – 7, 2015.

[164] L. Tian and L. Waller, “3d intensity and phase imaging from light field measurements in an led array microscope,” *Optica*, vol. 2, pp. 104–111, Feb 2015.

[165] M. Gu and C. J. R. Sheppard, “Three-dimensional image formation in confocal fluorescence microscopy,” in *Biomedical Image Processing and Three-Dimensional Microscopy* (R. S. Acharya, C. J. Cogswell, and D. B. Goldgof, eds.), vol. 1660, pp. 188 – 198, International Society for Optics and Photonics, SPIE, 1992.

[166] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, “Deep convolutional neural network for inverse problems in imaging,” *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4509–4522, 2017.

[167] A. Chambolle, “An algorithm for total variation minimization and applications,” *J. Math. Imaging Vis.*, vol. 20, no. 1-2, pp. 89–97, 2004.

[168] A. Sinha, J. Lee, S. Li, and G. Barbastathis, “Lensless computational imaging through deep learning,” *Optica*, vol. 4, pp. 1117–1125, Sep 2017.

[169] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, “Noise2noise: Learning image restoration without clean data.” 1803.04189, 2018. [Online; accessed 25-January-2021].

[170] R. H. R. Hahnloser, R. Sarpeshkar, M. A. Mahowald, R. J. Douglas, and H. S. Seung, “Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit,” *Nature*, vol. 405, no. 6789, p. 947–951, 2000.

[171] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[172] V. Dumoulin and F. Visin, “A guide to convolution arithmetic for deep learning.” arXiv: 1603.07285, 2018.

[173] O. Wengrowicz, O. Peleg, T. Zahavy, B. Loevsky, and O. Cohen, “Deep neural networks in single-shot ptychography,” *Opt. Express*, vol. 28, pp. 17511–17520, Jun 2020.

[174] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015 - 18th International Conference Munich, Germany, October 5 - 9, 2015, Proceedings, Part III* (N. Navab, J. Hornegger, W. M. W. III, and A. F. Frangi, eds.), vol. 9351 of *Lecture Notes in Computer Science*, pp. 234–241, Springer, 2015.

[175] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, 2015*, pp. 1026–1034, IEEE Computer Society, 2015.

[176] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[177] D. Nistér and H. Stewénius, "Scalable recognition with a vocabulary tree," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2006), 17-22 June 2006, New York, NY, USA*, pp. 2161–2168, IEEE Computer Society, 2006.

[178] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 47–57, 2017.

[179] A. Gianoncelli *et al.*, "Recent developments at the TwinMic beamline at ELETTRA: An 8 SDD detector setup for low energy X-ray fluorescence," *Journal of Physics: Conference Series*, vol. 425, no. Part 18, pp. 8–11, 2013.

[180] A. Gianoncelli *et al.*, "Simultaneous soft X-ray transmission and emission microscopy," *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 608, no. 1, pp. 195–198, 2009.

[181] L. Strüder *et al.*, "Development of the Silicon Drift Detector for Electron Microscopy Applications," *Microscopy Today*, vol. 28, no. 5, pp. 46–53, 2020.

[182] A. Gianoncelli *et al.*, "Current status of the TwinMic beamline at Elettra: A soft X-ray transmission and emission microscopy station," *Journal of Synchrotron Radiation*, vol. 23, no. 6, pp. 1526–1537, 2016.

[183] G. Kourousias *et al.*, "XRF topography information: Simulations and data from a novel silicon drift detector system," *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 936, no. October 2018, pp. 80–81, 2019.

[184] J. Bufon *et al.*, "Towards a multi-element silicon drift detector system for fluorescence spectroscopy in the soft X-ray regime," *X-Ray Spectrometry*, vol. 46, no. 5, pp. 313–318, 2017.

[185] A. Longoni *et al.*, "A new XRF spectrometer based on a ring-shaped multi-element silicon drift detector and on X-ray capillary optics," *IEEE Transactions on Nuclear Science*, vol. 49 Ii, no. 3, pp. 1001–1005, 2002.

[186] A. Rachevski *et al.*, "The XAFS fluorescence detector system based on 64 silicon drift detectors for the SESAME synchrotron light source," *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 936, no. September 2018, pp. 719–721, 2019.

[187] T. N. Vu *et al.*, "Getting your peaks in line: A review of alignment methods for NMR spectral data," *Metabolites*, vol. 3, no. 2, pp. 259–276, 2013.

[188] G. Kourousias, F. Billè, and A. Gianoncelli, "Automated nonlinear alignment of xrf spectra," *X-Ray Spectrometry*, vol. 46, no. 1, pp. 44–48, 2017.

[189] V. Solé *et al.*, "A multiplatform code for the analysis of energy-dispersive x-ray fluorescence spectra," *Spectrochimica Acta Part B: Atomic Spectroscopy*, vol. 62, no. 1, pp. 63–68, 2007.

[190] C. H. Hsiao *et al.*, "A Dynamic Data Correction Method for Enhancing Resolving Power of Integrated Spectra in Spectroscopic Analysis," *Analytical Chemistry*, vol. 92, no. 19, pp. 12763–12768, 2020.

[191] G. Tomasi *et al.*, "Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data," *Journal of Chemometrics*, vol. 18, no. 5, pp. 231–241, 2004.

[192] P. Monchamp *et al.*, "Signal Processing Methods for Mass Spectrometry," *Systems Bioinformatics*, pp. 101–124, 2006.

[193] C. Studholme, D. Hill, and D. Hawkes, "An overlap invariant entropy measure of 3d medical image alignment," *Pattern Recognition*, vol. 32, no. 1, pp. 71 – 86, 1999.

[194] S. Benedetto and E. Biglieri, *Principles of Digital Transmission.* London (UK): Springer Nature, 2001.

[195] A. Graps, "An introduction to wavelets," *IEEE Computational Science and Engineering*, vol. 2, no. 2, pp. 50–61, 1995.

[196] F. Bleichrodt, T. van Leeuwen, W. J. Palenstijn, W. van Aarle, J. Sijbers, and K. J. Batenburg, "Easy implementation of advanced tomography algorithms using the astra toolbox with spot operators," *Numerical Algorithms*, vol. 71, pp. 673–697, Mar 2016.

[197] A. Sorrentino, J. Nicolás, R. Valcárcel, F. J. Chichón, M. Rosanes, J. Avila, A. Tkachuk, J. Irwin, S. Ferrer, and E. Pereiro, "MISTRAL: a transmission soft X-ray microscopy beamline for cryo nanotomography of biological samples and magnetic domains imaging," *Journal of Synchrotron Radiation*, vol. 22, pp. 1112–1117, Jul 2015.

[198] J. L. Carrascosa, F. J. Chichón, E. Pereiro, M. J. Rodríguez, J. J. Fernández, M. Esteban, S. Heim, P. Guttmann, and G. Schneider, "Cryo-x-ray tomography of vaccinia virus membranes and inner compartments," *Journal of Structural Biology*, vol. 168, no. 2, pp. 234 – 239, 2009.

[199] I. Arslan, J. R. Tong, and P. A. Midgley, "Reducing the missing wedge: High-resolution dual axis tomography of inorganic materials," *Ultramicroscopy*, vol. 106, no. 11, pp. 994 – 1000, 2006. Proceedings of the International Workshop on Enhanced Data Generated by Electrons.

[200] K. Jun and S. Yoon, "Alignment solution for ct image reconstruction using fixed point and virtual rotation axis," *Scientific Reports*, vol. 7, p. 41218, Jan 2017.

[201] J. R. Kremer, D. N. Mastronarde, and J. McIntosh, "Computer visualization of three-dimensional image data using imod," *Journal of Structural Biology*, vol. 116, no. 1, pp. 71 – 76, 1996.

[202] V. Van Nieuwenhove, J. De Beenhouwer, T. De Schryver, L. Van Hoorebeke, and J. Sijbers, "Data-driven affine deformation estimation and correction in cone beam computed tomography," *IEEE Transactions on Image Processing*, vol. 26, no. 3, pp. 1441–1451, 2017.

[203] R. Han, X. Wan, Z. Wang, Y. Hao, J. Zhang, Y. Chen, X. Gao, Z. Liu, F. Ren, F. Sun, and F. Zhang, "Autom: A novel automatic platform for electron tomography reconstruction," *Journal of Structural Biology*, vol. 199, no. 3, pp. 196 – 208, 2017.

[204] A. Lukezic, T. Vojír, L. C. Zajc, J. Matas, and M. Kristan, "Discriminative correlation filter tracker with channel and spatial reliability," *Int. J. Comput. Vis.*, vol. 126, no. 7, pp. 671–688, 2018.

[205] D. Gürsoy, F. De Carlo, X. Xiao, and C. Jacobsen, "TomoPy: a framework for the analysis of synchrotron tomographic data," *Journal of Synchrotron Radiation*, vol. 21, pp. 1188–1193, Sep 2014.

[206] W. van Aarle, W. J. Palenstijn, J. Cant, E. Janssens, F. Bleichrodt, A. Dabravolski, J. D. Beenhouwer, K. J. Batenburg, and J. Sijbers, "Fast and flexible x-ray tomography using the astra toolbox," *Opt. Express*, vol. 24, pp. 25129–25147, Oct 2016.

[207] W. J. Palenstijn, J. Bédorf, J. Sijbers, and K. J. Batenburg, “A distributed astra toolbox,” *Advanced Structural and Chemical Imaging*, vol. 2, p. 19, Dec 2016.

[208] T. K. Moon, “The expectation-maximization algorithm,” *IEEE Signal Processing Magazine*, vol. 13, no. 6, pp. 47–60, 1996.

[209] B. D. Lucas and T. Kanade, “An iterative image registration technique with an application to stereo vision,” in *Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2*, IJCAI’81, (San Francisco, CA, USA), pp. 674–679, Morgan Kaufmann Publishers Inc., 1981.

[210] C. Tomasi and T. Kanade, “Detection and tracking of point features,” April 1991. Carnegie Mellon University Technical Report CMU-CS--91-132.

[211] D. Lowe, “Object recognition from local scale-invariant features,” in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, (Kerkyra, Greece), pp. 1150–1157 vol.2, IEEE, 1999.

[212] I. Rey-Otero, J. Morel, and M. Delbracio, “An analysis of the factors affecting keypoint stability in scale-space,” *Journal of Mathematical Imaging and Vision*, vol. 56, no. 3, pp. 554–572, 2016.

[213] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision.* West Nyack: Cambridge University Press, 2004. OCLC: 1044713766.

[214] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, “Fast and robust multiframe super resolution,” *IEEE Transactions on Image Processing*, vol. 13, pp. 1327–1344, Oct 2004.

[215] M. Amintoosi, M. Fathy, and N. Mozayani, “Precise Image Registration with Structural Similarity Error Measurement Applied to Superresolution,” *EURASIP Journal on Advances in Signal Processing*, vol. 2009, p. 305479, Dec. 2009.

[216] S. L. Smith, P. Kindermans, C. Ying, and Q. V. Le, “Don’t decay the learning rate, increase the batch size,” in *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*, OpenReview.net, 2018.

[217] J. D. Owens, M. Houston, D. Luebke, S. Green, J. E. Stone, and J. C. Phillips, “Gpu computing,” *Proceedings of the IEEE*, vol. 96, no. 5, pp. 879–899, 2008.

[218] J. Bufon, S. Schillani, M. Altissimo, P. Bellutti, G. Bertuccio, F. Billè, R. Borghes, G. Borghi, G. Cautero, D. Cirrincione, S. Fabiani, F. Ficorella, M. Gandola, A. Gianoncelli, D. Giuressi, G. Kourousias, F. Mele, R. Menk, A. Picciotto, A. Rachevski, I. Rashevskaya, M. Sammartini, A. Stolfa, G. Zampa, N. Zampa, N. Zorzi, and A. Vacchi, "A new large solid angle multi-element silicon drift detector system for low energy x-ray fluorescence spectroscopy," *Journal of Instrumentation*, vol. 13, pp. C03032–C03032, mar 2018.

[219] F. Billè, G. Kourousias, E. Luchinat, M. Kiskinova, and A. Gianoncelli, "X-ray fluorescence microscopy artefacts in elemental maps of topologically complex samples: Analytical observations, simulation and a map correction method," *Spectrochimica Acta Part B: Atomic Spectroscopy*, vol. 122, pp. 23 – 30, 2016.

[220] T. Chen, I. J. Goodfellow, and J. Shlens, "Net2net: Accelerating learning via knowledge transfer," in *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*, 2016.

[221] Y. Zhong, J. Chen, and B. Huang, "Toward end-to-end face recognition through alignment learning," *IEEE Signal Process. Lett.*, vol. 24, no. 8, pp. 1213–1217, 2017.

[222] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network." arXiv: 1503.02531, 2015.

[223] H. N. Chapman and K. A. Nugent, "Coherent lensless x-ray imaging," *Nature Photonics*, vol. 4, pp. 833–839, Dec 2010.

[224] A. Menzel, C. M. Kewish, M. Dierolf, P. Thibault, P. Kraft, O. Bunk, K. Jefimovs, C. David, and F. Pfeiffer, "Hard x-ray scanning transmission microscopy with a 2dpixel array detector," *Journal of Physics: Conference Series*, vol. 186, p. 012054, sep 2009.

[225] J. Spence, U. Weierstall, and M. Howells, "Coherence and sampling requirements for diffractive imaging," *Ultramicroscopy*, vol. 101, no. 2, pp. 149–152, 2004.

[226] G. D. Evangelidis *et al.*, "Parametric image alignment using enhanced correlation coefficient maximization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 10, pp. 1858–1865, 2008.

[227] C. Studholme *et al.*, "An overlap invariant entropy measure of 3d medical image alignment," *Pattern Recognition*, vol. 32, no. 1, pp. 71–86, 1999.