University of Zurich[UZH]

Department of Physics

Master Thesis

Sparse sampling for fast quasiparticle interference mapping using a scanning tunneling microscope

and

Deep learning based denoising of low counting statistics X-ray diffraction data

Jens Oppliger

supervised by
Prof. Dr. Fabian D. Natterer

and
Prof. Dr. Johan Chang

June 7, 2022

Abstract

Quasiparticle interference (QPI) mapping using a scanning tunneling microscope (STM) is a notoriously slow measurement technique for gaining insights into the band structure of 2D quantum materials. In conventional QPI mapping the data recording is serial. Thus, complex experiments involving variations of various degrees of freedom such as magnetic field, temperature, doping, or strain become very cumbersome due to the finite hold time of liquid helium cryostats. In this master thesis we build upon the previously developed baseline of sparse sampling and verify its applicability in a low temperature STM by exploiting the natural sparsity of QPI data. We will demonstrate that combining a few sparse measurements with faster point spectroscopy leads to an orders of magnitude faster QPI mapping by utilizing a compressive sensing recovery algorithm. We will furthermore introduce a new adaptive sparse sampling scheme that allows to cumulatively stack individual measurement cycles to gradually improve the QPI quality. We will exemplify these concepts using the noble metal Au(111), the high temperature superconductor $Bi_2Sr_2CaCu_2O_{8+\delta}$ and the topological Weyl semimetal WTe_2. We will also study how experimental nonlinearities influence the developed method and how these can potentially be resolved.

A second part of this master thesis is dedicated to a deep learning-based approach for denoising low counting statistics X-ray diffraction data. Many times the counting time in diffraction experiments is limited due to various circumstances such as beam damage or when performed in pulsed magnetic fields. Thus, the intrinsic signal is often times overshadowed by noise. By training two different deep convolutional neural networks (CNN) we will show its usefulness in denoising X-ray diffraction data. The networks will be trained on both experimental and artificial noise and we will show that training on real experimental data leads to better performance when denoising real low count data. This is in contrast to the predominant methodology in AI-assisted image denoising, that is training on artificially added noise to high counting statistics data.

Acknowledgements

Many people have contributed to this thesis in terms of active measurement support or helpful discussions. I would especially like to thank my main supervisor Prof. Fabian Natterer who motivated me to continue my prior work on sparse sampling. His ambitions and enthusiasm in utilizing advances in computational sciences and mathematics is what paved the road to many results and ideas of this thesis. I also want to thank Danyang Liu, who assisted me in acquiring experimental data, taught me about the operation of the STM and served as a valuable source of feedback. I want to thank Berk Zengin for the collaboration regarding the interconnection of parallel spectroscopy and sparse sampling. Furthermore, I like to thank Kevin Hauser for insightful discussions and his help in identifying problems of the developed techniques and their potential solutions. For the deep learning project, I would like to thank my supervisor Prof. Johan Chang for his motivation about machine learning in experimental condensed matter physics and Michael Denner for giving me insights into neural network training procedures. Lastly, I would like to express my gratitude towards Prof. Nicola Serra for granting me access to his GPU cluster, which was used for training the neural networks.

Contents

1 Sparse sampling for fast quasiparticle interference mapping

1.1 Scanning Tunneling Microscope

1.1.1 Introduction

A scanning tunneling microscope (STM) is a surface sensitive measurement device, which achieves atomic resolution[1] by utilizing a very sharp metallic tip that is preferably terminated by a single atom. Its invention dates back to 1983 when G. Binning and H. Rohrer demonstrated the working principle of the first STM using Au(110), Si(111), and GaAs(111) samples [1]. The basic setup is shown in Figure 1. An STM mainly consists of some piezoelectric actuators (PZA) for moving the tip in both lateral (x,y) and out-of-plane (z) direction. By applying a voltage onto the PZAs one can achieve a very small contraction or expansion of the piezo crystals, resulting in a very precise tip movement close to the sample surface. A metallic tip is attached to the PZA such that by applying a bias voltage between the tip and the sample, a tunneling current can be measured and with that both topographic information as well as electronic properties of the sample.

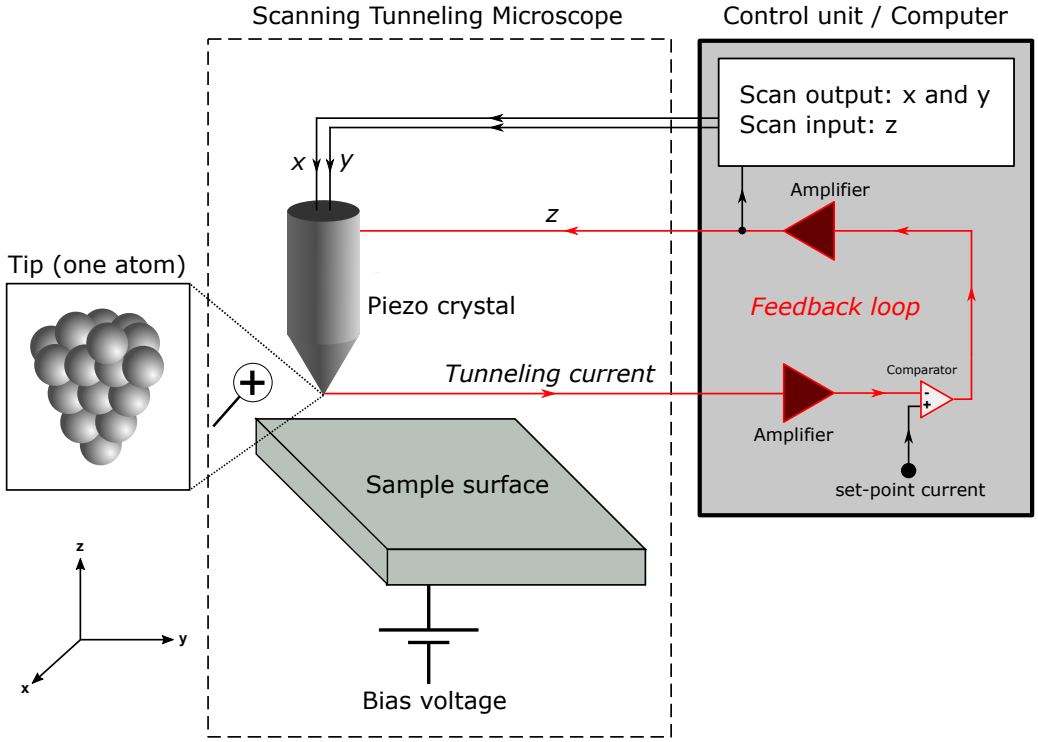

Figure 1: Schematic setup of a scanning tunneling microscope (STM). The tip is usually situated only a few atomic diameters above the sample surface. Applying a bias voltage between the tip and the sample will lead to an exponentially small tunneling current, which can be measured. This allows to directly visualize topographic features by rastering the tip along the surface. Figure adapted from [3].

There are two different operation modes of an STM that are depicted in Figure 2. Instead of keeping the height of the tip constant during the lateral movement (constant height mode) one can also adjust the height (z) in order to maintain a constant tunneling current (constant current model). The latter involves an additional feedback loop and is therefore slower but reduces the risk of the tip crashing into the sample surface. In the constant current mode, the effective tunneling current is amplified[2] and converted into a voltage, which is then compared to a reference value. The resulting difference in voltage is then used to drive the z-piezo, adjusting the height above the surface to ensure a constant current. Due to the fact that the tip hovers only a few atomic diameters above the surface, vibrational isolation of the STM using special dampers is essential. Furthermore, ultra-high vacuum (UHV) and

[1] Range of Å, 1 Å$= 10^{-10}$ m.

[2] Normally, the measured tunneling current is in the range of a few nA such that powerful amplifiers have to be used with an appropriate gain-setting.

cryogenic temperatures are crucial to ensure a stable measurement environment. By rastering the tip along a rectangular grid one can measure the tunneling current respectively the corresponding z-height and create a two-dimensional image of the surface morphology (topography) as illustrated in Figure 2 on the right.

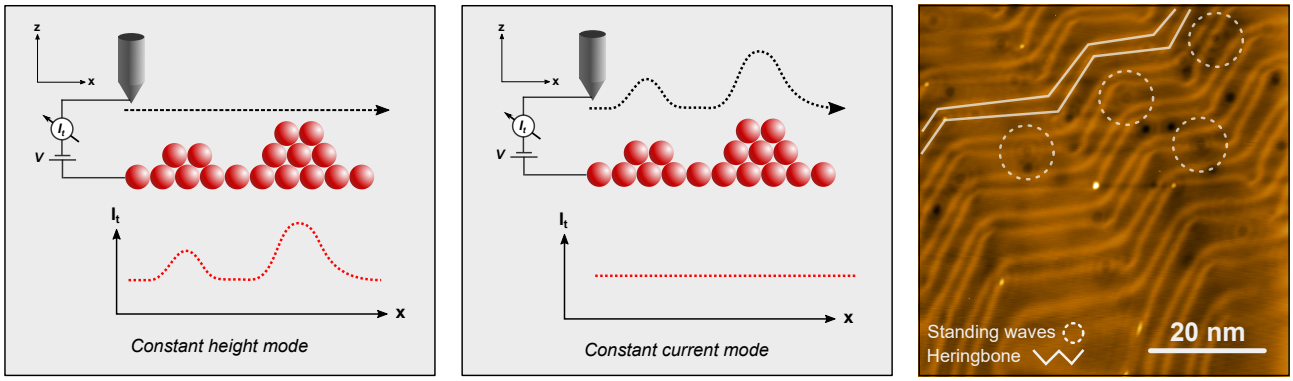

Figure 2: Visualization of constant height and constant current mode. The right panel shows a topographic image of Au(111) with characteristic Herringbone reconstruction and standing wave patterns. The latter is due to electrons that scatter at surface impurities leading to a modulation of the local density of states (LDOS).

In addition to resolving topographic features, an STM is also able to study the electronic properties such as the band structure of 2D quantum materials by utilizing the quantum nature of electrons. It achieves this via a technique denoted as Fourier transform STM (FT-STM) or quasiparticle interference (QPI) imaging [2], which will be described at a later point. In the following I will mostly follow the book of J. Chen [3] unless specified otherwise. The main ability of an STM is to measure a very small tunneling current, which arises from the quantum nature of electrons. When the tip is situated very close above the surface of the sample then the wave functions of the tip electronic states will overlap with the wave functions of the sample electronic states. Due to this overlap, there exists a finite probability for electrons to tunnel through the vacuum-like potential barrier between the tip and the sample. By applying a bias voltage V between the tip and sample one can maintain this tunneling current by shifting the effective Fermi level of the sample with respect to the Fermi level of the tip. Normally, these two Fermi levels would become equal very quickly by the exchange of electrons leading to an energy equilibrium. Depending on the sign of the applied voltage one can either probe the empty states ($V > 0$) or the occupied states of the sample ($V < 0$). Vice versa one probes the occupied tip states when using a positive voltage and the empty tip states when using a negative voltage.

1.1.2 Classical explanation of the tunneling current phenomena

Classically one can describe an electron with energy E that hits onto an external potential barrier of height $U = U(z)$ as

$$\frac{p^2}{2m} + U(z) = E \quad \Leftrightarrow \quad p = \sqrt{2m(E - U(z))} \tag{1}$$

where p is the classical electron momentum and m is the (free) electron mass. The electron should classically not be able to penetrate the potential barrier for $E < U(z)$ (no real solutions exist because of the square root term). However, according to quantum mechanics an electron is in fact not just a particle but also a wave. Thus, it can be described by a wave function $\psi(\vec{r})$, which fulfils the (time-independent) Schrödinger equation

$$H\psi(\vec{r}) = E\psi(\vec{r}) \quad \text{with Hamiltonian} \quad H = E_{\text{kin}} + E_{\text{pot}} = \frac{\hat{p}^2}{2m} + U(\vec{r}) \tag{2}$$

Where now the electron momentum is described by a quantum mechanical operator $\hat{p} = -i\hbar\vec{\nabla}$. In 1D we then obtain

$$\left[-\frac{\hbar^2}{2m}\frac{d^2}{dz^2}+U(z)\right]\psi(z)=E\psi(z) \tag{3}$$

Using the Ansatz of a plane wave, the above differential equation for $\psi(z)$ has solutions in both the classically allowed and the classically forbidden regions

$$\psi(z)=\begin{cases} U>E: & \psi(0)\,e^{\pm ikz} \quad \text{with} \quad k=\frac{\sqrt{2m(E-U)}}{\hbar} \\ U<E: & \psi(0)\,e^{-\kappa z} \quad \text{with} \quad \kappa=\frac{\sqrt{2m(U-E)}}{\hbar} \end{cases} \tag{4}$$

We see that the probability density of finding an electron beyond the tunneling barrier (in $+z$ direction[3]) is finite, even though it decays exponentially fast after passing the potential barrier as

$$A\propto|\psi(z)|^2\propto e^{-2\kappa z}\neq 0 \tag{5}$$

where κ is often denoted as the decay constant. Considering a simple metal-vacuum-metal junction the Fermi energy at absolute zero temperature is given by the work function ϕ of the metal surface, which is the minimum energy required to remove an electron from the bulk of the material into the vacuum region. Let us assume that the work function of the tip and the sample are identical. If we then apply a small negative bias voltage $V\ll\phi/e$ then the Fermi level of the tip with respect to the Fermi level of the sample will effectively be reduced by $-eV$. This means that electronic sample states ψ_n with energies $E_F-eV<E_n<E_F$ are now able to tunnel into the empty tip states as schematically illustrated in Figure 3. In this case only the states that are very close to the Fermi level will play a role in the tunneling current such that the decay constant of a sample state that is situated near the Fermi level is given by the work function

$$I(z)=I(0)\,e^{-2\kappa z} \quad \text{with} \quad \kappa=\frac{\sqrt{2m\phi}}{\hbar}\approx 5\sqrt{\phi\,(\text{eV})}\,\text{nm}^{-1} \tag{6}$$

On average the work function ϕ of metals is in the order of a few eV, which means that the measured tunneling current decreases by one order of magnitude per 0.1 nm (1 Å) height increase. Therefore, an STM is very sensitive to tiny changes of local sample features when the tip is brought very close to the sample surface.

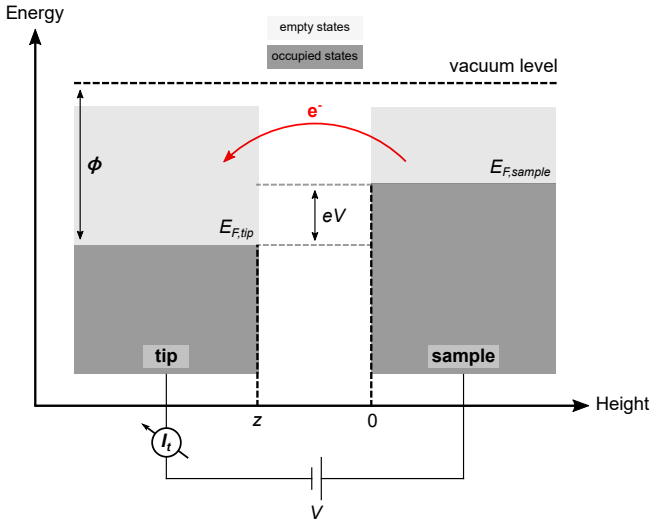

Figure 3: 1D tunneling phenomena for electrons. By applying a negative bias voltage V one can introduce a constant difference in Fermi levels of the sample and the tip, which leads to a tunneling current I_t from the sample to the tip. The opposite scenario is achieved by applying a positive bias voltage. It is assumed that the difference in energy is small compared to the work function ϕ.

[3] Actually, the tunneling behaviour is bi-directional, i.e. quantum mechanics also allows for decaying electron wave functions in $-z$ direction.

1.1.3 Bardeen theory of tunneling

One of the main theories describing the effect of electron tunneling was developed by J. Bardeen [4]. The main idea is to solve the stationary Schrödinger equation of the two systems (tip and sample) separately from which the tunneling current, respectively the electron transmission rate, can be obtained using time-dependent perturbation theory. Near the Fermi level and by applying a positive bias voltage V the electron transmission rate can be seen as a convolution of the density of states ρ of the two electrodes under the influence of thermal broadening of the states[4]

$$I \propto \int_{-\infty}^{\infty} \{f(E-eV)\rho_s(E-eV)\rho_t(E) - f(E)\rho_s(E-eV)\rho_t(E)\}|M|^2\, d\varepsilon \quad \text{with} \quad E = E_F + \varepsilon \tag{7}$$

f is the Fermi-Dirac distribution describing the occupation of electron states as a function of temperature and energy. The amplitude of the electron transfer $|M|^2$ contains the exponentially decaying tunneling probability through the vacuum barrier with increasing separation distance.[5] One can then use some further simplifying assumptions

- The amplitude of the electron transfer $|M|^2$ is constant in the energy interval of interest, which means that the overlap of the wave functions of the two sub-systems is small.
- The thermal broadening k_BT is small compared to the desired energy resolution and therefore the Fermi-Dirac distribution $f(E) = (1 + \exp[(E - E_F)/k_BT])^{-1}$ can be replaced by a step function.
- The density of states of the tip is flat, i.e. its influence on the sample electron wave functions is small and it can thus be seen as constant near the Fermi level.

Using these assumptions, one obtains a simpler expression for the tunneling current

$$I \propto e^{-2\kappa z}\rho_t(E_F) \int_0^{eV} \rho_s(E_F - eV + \varepsilon)\, d\varepsilon \tag{8}$$

A similar expression can be derived for applying a negative bias voltage between tip and sample.

1.1.4 Quasiparticle interference

Next to the ability of mapping topographic features of a sample surface, manipulating single atoms for the creation of the smallest possible nanostructures, an STM is also able gain valuable insights into the electronic properties of quantum materials such as their band structure. Band structures can be derived for electrons situated in a periodic potential (atomic lattice) obeying the Schrödinger equation and provides a relation between the energy of electrons and their momentum $E(\vec{k})$. The main solutions for the time-independent Schrödinger equation are Bloch wave functions $\psi_{\vec{k}}(\vec{r})$, which can be described as linear superposition of plane waves modulated by a periodic function as a consequence of the periodic crystal lattice

$$\psi_{\vec{k}}(\vec{r}) = e^{i\vec{k}\cdot\vec{r}}\, u(\vec{r}) \quad \text{with} \quad u(\vec{r}) = u(\vec{r} + \vec{R}_n) \tag{9}$$

where $\vec{R}_n$ are real space lattice vectors. The dispersion relation can then be calculated using a Bloch wave Ansatz and including the effect of an external potential utilizing second order perturbation theory. Adding an additional perturbing potential such as impurities, point defects or step edges will lead to standing wave patterns in the electron density, i.e. the elastic scattering of electrons at impurities will lead to an interference pattern that modulates the local density of states (LDOS) [2], which is defined as

[4] This is only true under the assumption of elastic tunneling $E_\mu = E_\nu$ (μ: states on the left , ν states on the right).
[5] M is often denoted as the tunneling matrix element.

$$\mathrm{LDOS}(\vec{r},\, E) = \sum_{\vec{k}} |\psi_{\vec{k}}(\vec{r})|^2\, \delta(E - E(\vec{k})) \tag{10}$$

For noble metals such as Cu, Au or Ag, where the dispersion relation has a parabolic shape, such standing waves in the electron density are known as Friedel oscillations and were first observed in 1993 on Cu(111) [5]. Friedel oscillations can be modelled by Bessel functions of the first kind J_0 [6]

$$\mathrm{LDOS}(\vec{r},\, E) \propto 1 - J_0(qx) \tag{11}$$

where q is the magnitude of the momentum transfer during a scattering event when electrons with initial momentum $\vec{k}_i$ scatter and end up with a final momentum $\vec{k}_f$ but the same energy. The momentum transfer for an isotropic parabolic spectrum then becomes

$$\vec{q} = \vec{k}_i - \vec{k}_f = \vec{k}_i - (-\vec{k}_i) = 2\vec{k}_i \tag{12}$$

An STM can probe the mentioned LDOS modulations indirectly in the presence of such scattering events via the measurement of the tunneling current [2]

$$I_t \propto V_t\, \mathrm{LDOS}(\vec{r},\, E_F)\, e^{-2\kappa z} \tag{13}$$

which is true under the assumption of low bias voltages, zero temperature and a tip containing s-wave like states. In practice one measures the differential conductance $G = d\mathrm{I}/d\mathrm{V}$ on every point of a predefined 2D grid, i.e. $G = G(x, y)$ resulting in a 2D $\mathrm{d}I/\mathrm{d}V$ map. Information about the in-plane scattering wave vectors $\vec{q}$ can then be obtained by transforming the obtained $\mathrm{d}I/\mathrm{d}V$ map into Fourier space. This technique is denoted quasiparticle interference (QPI) imaging or Fourier Transform-STM [2] and allows to indirectly gain insights into the band structure of the studied materials via the scattering of electrons at surface impurities. The basic principle of said technique is depicted in Figure 4.

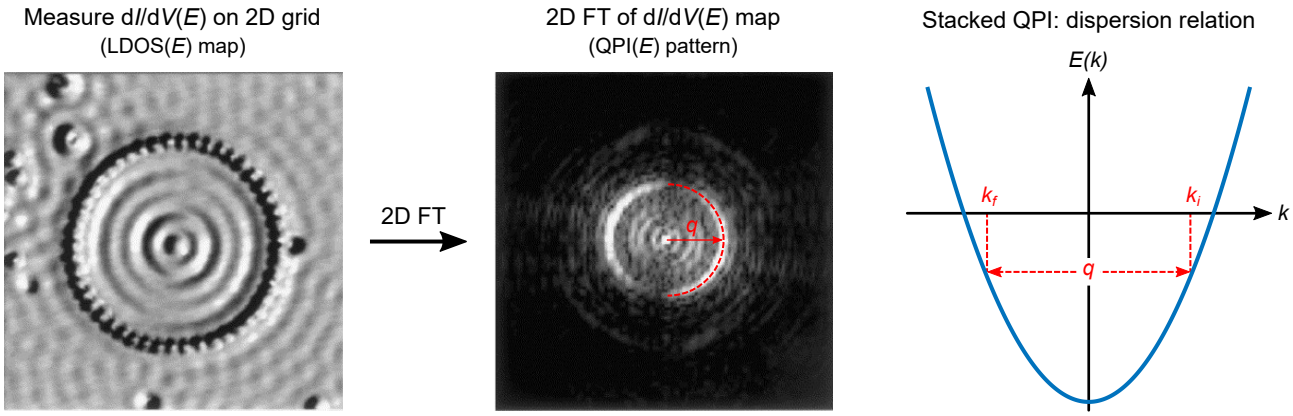

Figure 4: The concept of quasiparticle interference (QPI) imaging and how to achieve in-plane momentum sensitivity with a scanning tunneling microscope (STM). By measuring the differential conductance at various locations (x, y) on a 2D grid (real space) one obtains a $\mathrm{d}I/\mathrm{d}V$ map at a specific energy E, which resembles standing wave patterns respectively a modulation of the local density of states (LDOS). The data presented here has been taken from [7] and shows a ring of 48 Fe atoms, which confine electron states on a Cu(111) surface. A Fourier transform of such LDOS maps from real into scattering space (q-space) where a ring of radius q emerges (Shockley surface state). This is denoted as the QPI pattern at a particular energy. The scattering wave vector q is indicated both in the middle and the right panel. The latter shows a schematic parabolic isotropic dispersion relation (blue) of a typical nearly free electron gas material such as found on the surface of the noble metal Cu(111). It is obtained by stacking QPI patterns at different energies.

1.2 Compressive Sensing

1.2.1 Introduction

The understanding of signal sampling methodologies is of major importance in a ever-growing digitalization of our society and various fields of research. The fundamental concept regarding the sampling of band-limited signals that are continuous in the time domain is dictated by the famous Shannon-Nyquist sampling theorem. It describes that any signal, which satisfies the above conditions, can be exactly recovered by sampling it uniformly with a rate that is twice the maximum contained signal frequency [8][9][10] also known as the Shannon-Nyquist sampling rate $f_s = 2f_{\max}$. This theorem is widely applied in countless data acquisition protocols including, audio and video recording, radio technology, medical imaging and many more [11]. From a technological point of view the Shannon-Nyquist sampling theorem has moved the field of signal processing from the purely analog to the digital territory under the prospect of Moore's law that emerged in 1965 and predicted that the computational power would double every year [12][13].[6] However, in recent years the predicted trend of Moore's law shows to be deviating substantially from the true technological advances in semiconductor transistor manufacturing [14]. The main limitation originates from the very dense packing of transistors in the nanometer regime leading to the emergence of leakage currents. This problem paired with the fact that in many computational tasks the highest signal frequency is very high and thus the required Nyquist sampling rate becomes exceedingly large as well. This results in a vast amount of samples needed for a complete reconstruction of the signal via sinc interpolation [10] whose memory storage requires an ever growing amount of resources. This issue is addressed by the concept of data compression, which aims at reducing the memory requirements for measured high-dimensional data by converting the measured signal into a vector space that allows to describe it in a concise or „compressed" way. Signal compression schemes such as JPEG (for images) or MP3 (for audio) are nowadays omnipresent in various electronic areas. The working principle of natural image compression can be understood in terms of a digital CCD camera that is for example contained in a smartphone. A CCD camera contains millions of pixels each requiring multiple bits of memory. Because the acquired image cannot be stored in its original form[7] the CPU of the device will transform the original image (signal) into the more compact discrete cosine transform (DCT) vector space. The key feature of this transformation is that in the DCT domain the image contains only a few coefficients (pixels) that carry most of the information content, i.e. most of the energy. Removing those coefficients with small energy leads to a dimensionality-reduction of the data. Performing an inverse DCT transform one obtains a compressed version of the original image that differs only a little bit from the original image.

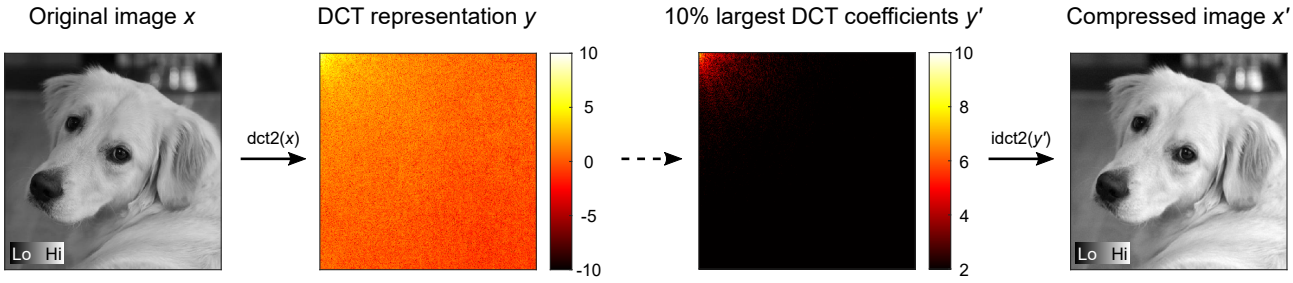

Figure 5: Concept of signal compressibility illustrated at the example of a natural grayscale image. Natural images can be converted into the discrete cosine transform (DCT) basis where only a few coefficients carry most of the energy (the colorbar is in a logarithmic scale). By keeping only a small fraction of said coefficients (the remaining coefficients are set to zero) and performing an inverse transformation one obtains a compressed version of the original image (JPEG algorithm). Note that nowadays the JPEG-2000 algorithm is used as a standard. Instead of the DCT basis it uses specific wavelet transform bases leading to a much better overall compression without compromising much of the quality [15].

[6] The original statement refers to a doubling of the number of transistors per integrated circuit.
[7] In a 10 MP CCD camera where each pixel requires 24 bit of memory, a single image would amount to 30 GB.

1.2.2 Sparsity and the sensing of sparse signals

In Figure 5 we focused on an image that was compressible via the DCT basis. With a compressible representation we refer to some signal that can be well-approximated by another signal with S non-zero coefficients [12]. In mathematical terms we say that a signal $\vec{x}$ of length N is sparse in some basis Ψ if it contains only $S \ll N$ non-zero coefficients in said basis. Another example in one dimension is depicted in Figure 6 and shows that a sinusoidal signal that can be represented with very few coefficients in the Fourier domain. For band-limited signals the Shannon-Nyquist sampling theorem states a minimum number of samples for a perfect reconstruction of such signals by measuring it in its non-sparse form. In this example the non-sparse domain is the time domain as there the signal has non-zero values (nearly) everywhere. However, if the signal has a sparse (or approximately sparse) representation in some vector domain then one could think of sensing the signal directly in its compressed form, which would require much lower sampling rates since most of the coefficients are zero. This is the main idea behind compressive sensing or sparse sampling as we will refer to it synonymously.

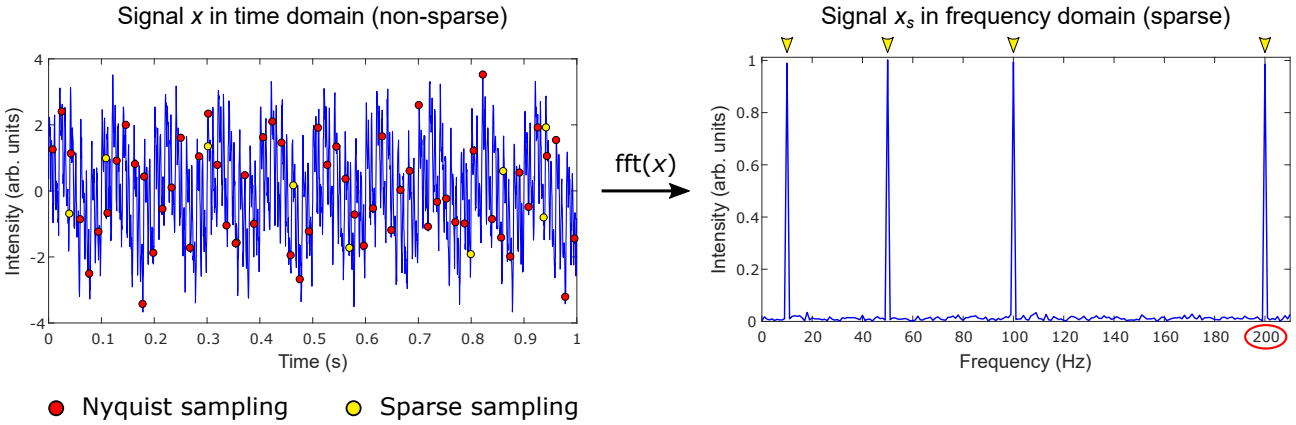

Figure 6: A signal consisting of a superposition of sinusoids is corrupted by noise (left). In its time domain the signal is not sparse because most of the coefficients are not zero. On the other hand the signal is (approximately) sparse in the Fourier domain (right). The traditional Nyquist sampling rate depends on the maximum signal frequency leading to a large amount of samples required for a successful reconstruction of the original signal (schematically depicted in red). The concept of compressive sensing (sparse sampling) allows to sample the signal at a much lower rate because the signal only consists of very few relevant coefficients (four) in the Fourier domain (schematically depicted in yellow).

The „sensing“ of a signal can be expressed via a linear system of equations [16]. When measuring M samples of a signal of length N one has

$$A\vec{x} = \vec{y} \tag{14}$$

$A \in \mathbb{C}^{M \times N}$ hereby models the linear data acquisition process (sensing matrix). $\vec{x} \in \mathbb{C}^{N \times 1}$ is the signal than one tries to reconstruct from a set of measurements $\vec{y} \in \mathbb{R}^{M \times 1}$. According to linear algebra Equation 14 is under-determined if $M < N$, which means that there exists no unique solution for $\vec{x}$. compressive sensing aims at solving such under-determined systems of equations when $M \ll N$ when the signal $\vec{x}$ has a sparse representation $\vec{x}_s$ in some orthonormal basis Ψ where it only contains S non-zero coefficients

$$\Psi\vec{x}_s = \vec{x} \tag{15}$$

One of the major breakthroughs in the field of compressive sensing was made by Candès, Romberg and Tao [17] as well as Donoho [18] in 2006. The authors proved that a sparse signal can be exactly reconstructed using only a few linear, non-adaptive measurements. The condition of non-adaptive or incoherent sampling is of major importance to construct suited sensing matrices and is often stated

as the *restricted isometry property* (RIP) that gives a general measure for the robustness of the signal reconstruction using a specific sensing matrix A [11]

$$(1-\delta_s)||\vec{x}||_2^2 \leq ||A\vec{x}||_2^2 \leq (1+\delta_s)||\vec{x}||_2^2 \tag{16}$$

where $||\vec{x}||_2$ denotes the ℓ_2-norm of $\vec{x}$ (Euclidean norm). The general ℓ_p-norm ($p \geq 0$) of a vector $\vec{x} = (x_1, x_2, ..., x_N)^T$ is defined as

$$||\vec{x}||_p \equiv \left(\sum_{i=1}^{N} |x_i|^p\right)^{1/p} \tag{17}$$

For $p = 0$ one defines the „pseudo-norm“ as a measure for the number of non-zero elements of a vector within the scope of compressive sensing [18]. It can also be shown that $M = 2S$ is the theoretical minimal number of measurements to recover all S-sparse vectors $\vec{x}_s$ via the ℓ_0-norm minimization approach

$$\min ||\vec{x}_s||_0 \quad \text{subject to} \quad \vec{y} = A\vec{x} = A\Psi\vec{x}_s \; (= A\vec{x}_s) \tag{18}$$

However, since the locations of the non-zero coefficients in $\vec{x}_s$ is a-priori not known solving Equation 18 becomes a computational NP-hard problem.[8] Going back to Equation 16 we say that a sensing matrix A obeys the RIP of order S if δ_s is not too close to one, which means that A approximately preserves the ℓ_2-norm of S-sparse signals [11]. The RIP is a necessary and sufficient condition on the sensing matrix A to ensure the exact recovery of all S-sparse vectors $\vec{x}_s$ from a set of linear measurements as stated in Equation 14 [16] via the ℓ_1-norm minimization. The main practical difference between these two norms is that the ℓ_1-norm can be solved very efficiently using linear programming. Similar to the ℓ_0-norm, the ℓ_1-norm also promotes sparse solutions, which makes intuitively sense by looking at the unit spheres using different p-norms in two dimensions as shown in Figure 7. A unit sphere is characterized by unitary length $||\vec{x}||_p = 1$ for all points $\vec{x}$ in the given space. Suppose that we are given a signal $\vec{x} \in \mathbb{R}^2$ that we wish to approximate using another point $\vec{x}'$ situated in a one-dimensional space A. The different ℓ_p-norms will give us the error of this approximation. The main task is again solving a minimization problem where we want to minimize the error $||\vec{x} - \vec{x}'||_p$ for $\vec{x}' \in A$. By increasing the „radius“ of the respective unit spheres (increasing the error) one observes that the approximated solution vector $\vec{x}' \in A$ is most likely sparse when using either the ℓ_0 or the ℓ_1-norm.[9]

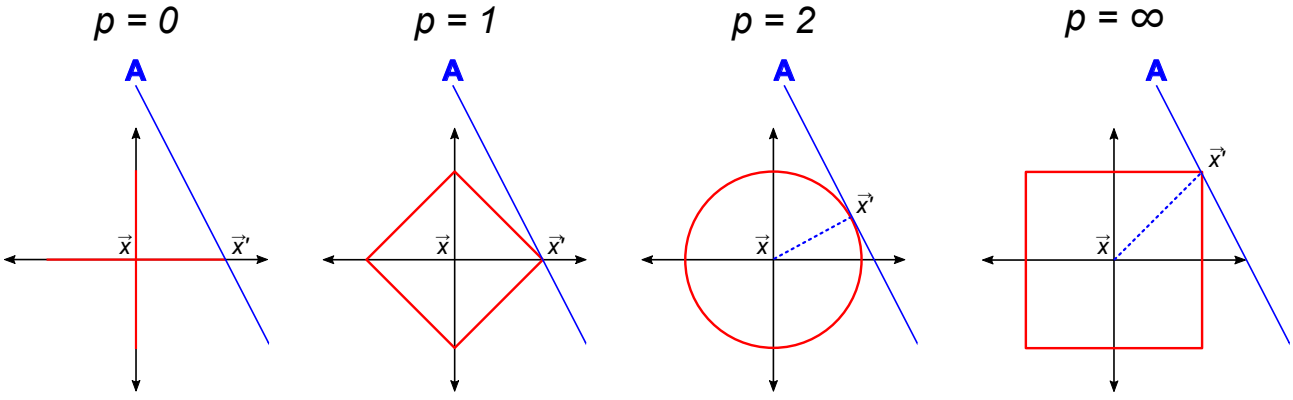

Figure 7: Different unit spheres in two dimensions for different p-norms ($p = 0$ is a „pseudo-norm“). Approximating a point $\vec{x}$ by some other point $\vec{x}'$ situated in a one-dimensional space A involves a minimization of the ℓ_p error. Both the ℓ_0 and the ℓ_1-norm hereby promote sparse solutions (the approximated vector $\vec{x}' \in A$ only has one non-zero component). For $p \geq 2$ the solution $\vec{x}'$ has two non-zero components except for very rare cases.

[8] A NP-hard problem (NP = Non-Polynomial) is characterized by the fact that the needed number of computations grows exponentially (or even faster) when the number of free parameters increases (in the context of this work given by the signal length N). Such problems cannot be solved in polynomial time.

[9] In fact any norm with $p < 1$ will promote sparse solutions as the resulting unit spheres have a convex shape.

If the RIP holds then it was proven that the solution provided by hypothetically solving the ℓ_0-norm is identical to the solution obtained by solving the ℓ_1-norm for the majority of cases [17][18]. This result led to a major increase in attention towards compressive sensing leading to various efficient optimization routines for solving Equation 14 if the number of measurements M is much smaller than the signal length N and also much below the Nyquist sampling regime. One of such algorithms is basis pursuit denoise (BPDN) (or LASSO) [19][38] where one additionally introduces some error threshold $\sigma \in \mathbb{R} > 0$ into the problem

$$\min ||\vec{x}_s||_1 \quad \text{subject to} \quad ||A\vec{x}_s - \vec{y}||_2 \leq \sigma \tag{19}$$

The minimum number of measurements M required for successfully recovering all S-sparse vectors of a signal $\vec{x}_s$ of length N depends on the used sensing matrix A. For example, it is known [11] that symmetric Bernoulli or other sub-gaussian distributions satisfy the RIP as long as

$$M \geq CS \log(N/S) \quad \text{with} \quad C > 0 \tag{20}$$

It turns out that Equation 20 defines a lower limit on the number of measurements required in general. For other bases that are orthonormal such as partial Fourier bases, which will be important for the main part of this thesis, it was shown that more measurements are required for a complete sparse recovery [20].

1.2.3 Applications

Because of the fact that the amount of acquired data will continue to grow while technological advances in manufacturing higher dense-packed transistors stagnates might lead to a future bottleneck. This is where new methods of data handling need to be developed. Going back to Figure 5 we can see such a potential bottleneck already. A CCD camera takes an uncompressed image just to discard most of the information during the compression stage. It is straight-forward to see that such a process is very inefficient. Compressive sensing could hereby serve as a useful tool as it allows to directly acquire the relevant information of the signal in its compressed digital form without the need of measuring the complete image. This creates opportunities for new fields such as super-resolution used for signal enhancements [21]. Today, compressive sensing has been successfully incorporated into medical imaging such as magnetic resonance imaging [22], pattern recognition [23] and radar techniques [24]. Because of the fact that many signals are sparse in some known basis or can be approximated by some sparse signal (compressibility) the applications of compressive sensing are far-reaching and also straight-forward to implement using existing fast algorithms. The key point is that one needs to non-adaptively (randomly) measure the signal in a basis where it is not sparse and then use a linear program after the data acquisition step to directly reconstruct the signal in its sparse domain [11].

1.3 Sparse sampling applied to Scanning Tunneling Microscopy

1.3.1 Introduction

The major work of this thesis is based on the knowledge acquired during my Bachelor thesis, which resulted in a proof of principles of sparse sampling in a simulated quasiparticle interference (QPI) experiment [25]. In the following we will go over the previously developed principles that form the foundation of the novel work presented in this thesis. In Section 1.1.4 we described the experimental procedure in obtaining insights into the band structure of 2D quantum materials via the Fourier transform of $\mathrm{d}I/\mathrm{d}V$ maps for different energies (see also Figure 4). Compared to angle resolved photo-emission spectroscopy (ARPES), which serves as the main tool for band structure analysis, QPI experiments using a Scanning Tunneling Microscope (STM) are an indirect way of measuring the band structure. Some advantages of an STM over ARPES are the compatibility with extreme environments such as magnetic fields, micron-sized samples, cryogenic temperatures and field-effect devices. The main disadvantage of an STM is that the technique is notoriously slow as it contains a single „sensor“ (the tip), which can only be at one place at a time. It therefore requires a lot of time to map out a 2D scan area, which is large enough to achieve the necessary resolution in q-space. This is because the finest momentum space resolution is given by $q_{\mathrm{min}} = 2\pi/L$ where L is the physical size of the scan frame. To complement this condition it also has to be made sure that an adequate amount of pixels is chosen to cover a the desired reciprocal space area [30]. Measuring at different energies introduces another dimension where the acquired point spectrum requires a fine enough energy spacing to distinguish small nuances of dispersive scattering wave vectors. To get a better intuition about the time-consuming nature of QPI mapping experiments consider the case of mapping a 512×512 pixel grid where each point spectrum takes 1 second. The total mapping time would be around 3 days. For a 1024×1024 pixel grid one would need to maintain stable measurement conditions for more than 12 days, which is longer than the normal hold time of liquid helium cryostats (currently around 11 days [31]). Increasing the parameter space to include varying temperatures, magnetic fields, different doping levels etc. will lead to a rapid growth of the measurement time, which makes executing more complex experiments very cumbersome. Nevertheless, an STM can be seen as complementary to ARPES as it excels under conditions where ARPES is limited as mentioned above.

By studying QPI patterns obtained from the Fourier transform of $\mathrm{d}I/\mathrm{d}V$ maps one can identify a common characteristics among all of these patterns: they are very sparse. Figure 8 shows a range of QPI patterns obtained on different materials showing a diverse range of structural complexity. But nevertheless, most of the coefficients in these patterns (i.e. most of the pixel values) are close to zero. This renders the utilization of Compressive Sensing in such experiments overly compelling [32] and forms the motivation of this and prior work [25] where we introduced the concept of sparse sampling for referring to a Compressive Sensing framework that we developed within the scope of STM-QPI mapping experiments.

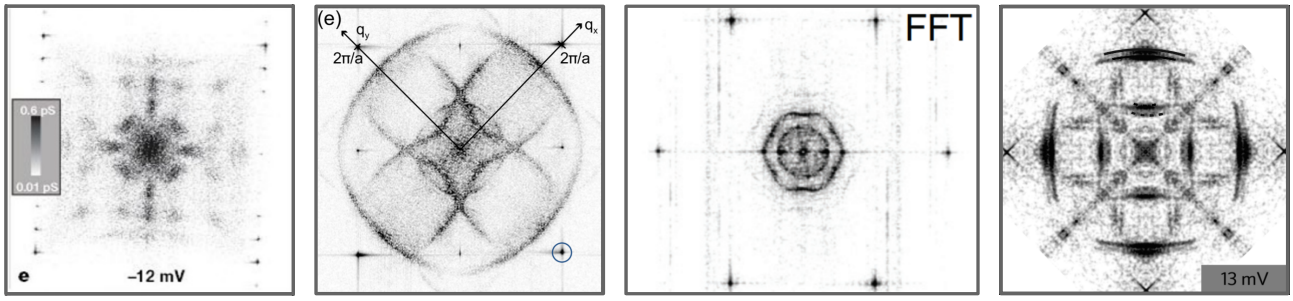

Figure 8: In an STM in-plane momentum sensitivity can be achieved via QPI mapping experiments. Such patterns all show the common characteristic of sparsity as most of the pixels are close to zero (white color). The data presented here has been taken from [26][27][28][29] (from left to right).

In Section 1.2.2 we established an intuitive understanding of compressive sensing. When some signal can be represented in a basis where most of its coefficients are (approximately) zero, i.e. the signal is sparse in that domain, then we can apply a linear algorithm built around an ℓ_1-norm minimization

problem, as described by Equation 19, to fully recover said sparse coefficients from only a few non-adaptive measurements in a domain where the signal is not sparse. In the case of QPI mapping experiments, the non-sparse domain is the real domain (local density of states (LDOS) modulations) and the sparse domain is the Fourier domain (q-space). Figure 9 exemplifies the concept of sparse sampling using a modulation of the LDOS of Cu(111) modelled by a Bessel function of the first kind [2]. By sampling the resulting standing wave pattern only at a few random locations we are able to recover the sparse QPI pattern and from there apply an inverse Fourier transform get back to the real space image. In this thesis we focus on the experimental verification of this sparse sampling concept applied to QPI imaging in a low-temperature STM.

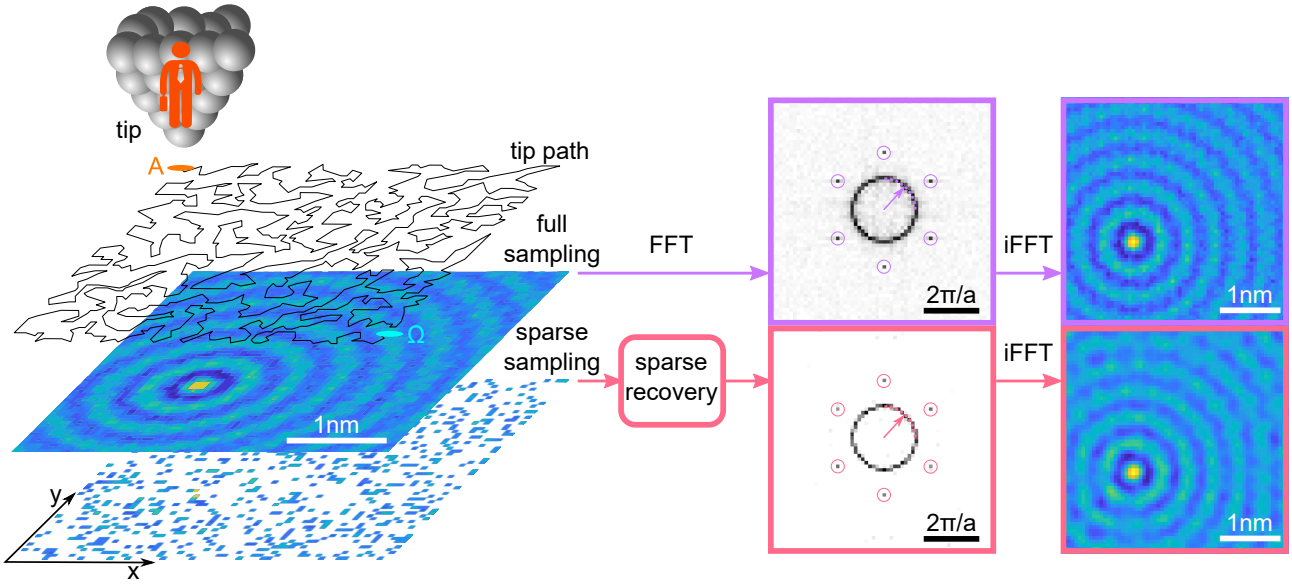

Figure 9: Concept of sparse sampling in an STM-QPI experiment [25]. A simulated standing wave pattern of Cu(111) at 4.3 eV with underlying atomic corrugation and added Gaussian noise (middle panel on the left). We acquire an undersampled set of measurements (bottom left) utilizing a near-optimal traveling salesperson (TSP) tip path (top left). From these measurements (20% of the full 64×64 pixel grid) we can reconstruct the sparse QPI pattern in the Fourier domain via an efficient ℓ_1-norm minimization algorithm [38][39]. Applying an inverse Fourier transform yields again the original LDOS modulations.

1.3.2 Traveling salesperson path implementation

Within the scope of sparse sampling, we are generally interested in measuring a small set of measurements. To satisfy the incoherency requirement (RIP) of compressive sensing, these measurements generally should be randomly distributed on a rectangular (square) grid of size $n_x \times n_y$. Performing a conventional line-by-line motion to visit all the points would be very inefficient as most of the pixels are not measured anyways. To minimize the required sampling time, we want to reduce the overall time required for moving the tip through all of the M selected locations. This task can be formulated as the famous open traveling salesperson (TSP) problem.[10] In our previous work [25] we made use of a genetic algorithm [34] to find the near-optimal path between the selected locations. Furthermore, we enforce a probability density gradient at the borders. This helps to reduce the finite size border artifacts that would otherwise appear after the sparse reconstruction due to the assumed continuity by the Fourier transform. Because of the poor scaling of the TSP calculation (NP-hardness with complexity $\mathcal{O}(M!)$) we added two improvements to enable an efficient path calculation

(1) A cheapest insertion algorithm [35] for less time-consuming path calculations

(2) A binary tree search to sub-divide the original rectangular grid into smaller subtiles

[10] We denote it as „open“ because we do not need to return to the starting position after visiting every point.

The used cheapest insertion and binary tree algorithms is outlined in Appendix A.1. Using a cheapest insertion algorithm slightly increases the overall path length but is limited by a maximum length not larger than twice the path length of the best possible path and has a computational complexity of $\mathcal{O}(M^2 \log_2(M))$ [36]. Because of the computational complexity of the TSP problem, one cannot arbitrary increase the number of points without exponentially increasing the computation time. Therefore, the method of sub-dividing the complete rectangular grid into smaller tiles is necessary if one would like to calculate a path that contains tens of thousands of measurement locations. In Figure 10 we compare the resulting paths on a 1024×1024 pixel grid with 1000 random locations. One clearly observes that the genetic algorithm leads to a more optimized path than the cheapest insertion algorithm. But the latter one is obtained in a much shorter time. We find that on average the paths resulting from the cheapest insertion approach are approximately 10% longer than the ones obtained when using 3000 measurement locations. Under a practical aspect the latter number is close to the maximum amount of locations when using the genetic algorithm due to its poor scalability. Everything larger than that results in potentially day-long calculations when using a conventional PC. Because of the NP-hardness of the TSP problem one can drastically reduce the computation time by calculating the TSP solution for a smaller subset of points and connecting the resulting k tiles with another near-optimal path. In this case the complexity reduces to $O(k(M/k)!)$ where k is the number of tiles and M is the total number of locations to visit.

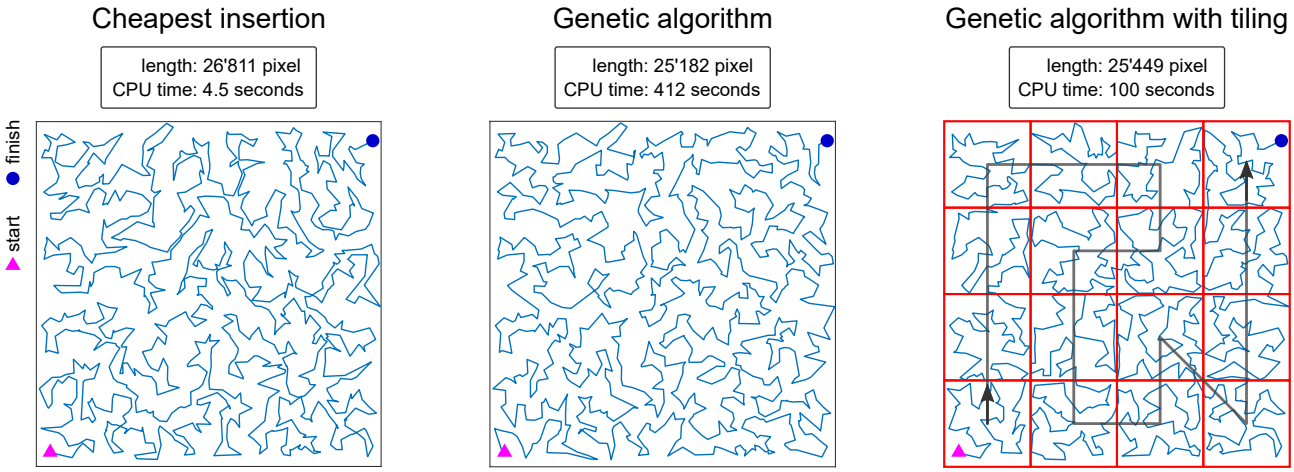

Figure 10: Comparison of traveling salesperson (TSP) paths resulting from different algorithms. A relatively fast path calculation can be achieved using a cheapest insertion algorithm [35] whereas a genetic algorithm [34] leads to a shorter overall path length. In order to measure tens of thousands of points one can subdivide the original grid into smaller tiles to substantially reduce the computation time while making use of a genetic algorithm for a more optimized path. All paths shown are calculated for the same random locations.

To cover a wide range of pixel grids and number of measurement locations we pre-calculate several different TSP paths spanning a range of 10'000 to around 250'000 points. These paths are sub-divided into smaller tiles where every tile contains on average around 500 points for a much more efficient calculation. Therefore, the number of tiles ranges from a few dozens to over 400 for a complete path.

1.3.3 Experimental setup

The experimental data in this thesis has been obtained using a Createc low-temperature STM as shown in Figure 11. The main cooling mechanism is provided by a 4 K liquid helium cryostat that is surrounded by multiple thermal shielding layers containing further coolants such as liquid nitrogen (77 K). The microscope itself it situated in another smaller container below the cryostat where the Joule-Thomson effect is used to cool down the liquid helium to around 1.5 K. The sample can be transferred between the preparation chamber, the load lock and the STM chamber using a manipulator stick. These chambers are all separated by mechanical gate valves that can be opened or closed. Both preparation and STM chamber are in a ultra-high vacuum (UHV) state ($\sim 10^{-9}$ to 10^{-10} mbar).

The preparation chamber is at room temperature and contains various attachments such as an ion gun for sputtering the sample. As a counterpart there exists an annealing station that is mounted on the manipulator stick. The load lock consists of a small volume that can be pumped down to pressures of $\sim 10^{-7}$ mbar relatively fast. It therefore can be used for quick sample transfers from ambient conditions into the UHV of the preparation chamber. The load lock further contains a tip heating station utilizing high voltage applied between the tip and a tungsten filament (not visible in the picture), which was built with the help of Danyang Liu. The motion of the tip within the STM head (inset in Figure 11) is controlled by a Nanonis SPM controller. Next to lateral (x, y) motion the tip can also be moved in the out-of-plane z direction. A coarse walker motor is additionally installed for traversing larger distances on the sample as the piezoelectric tubes only have a limited range in the orders of μm.

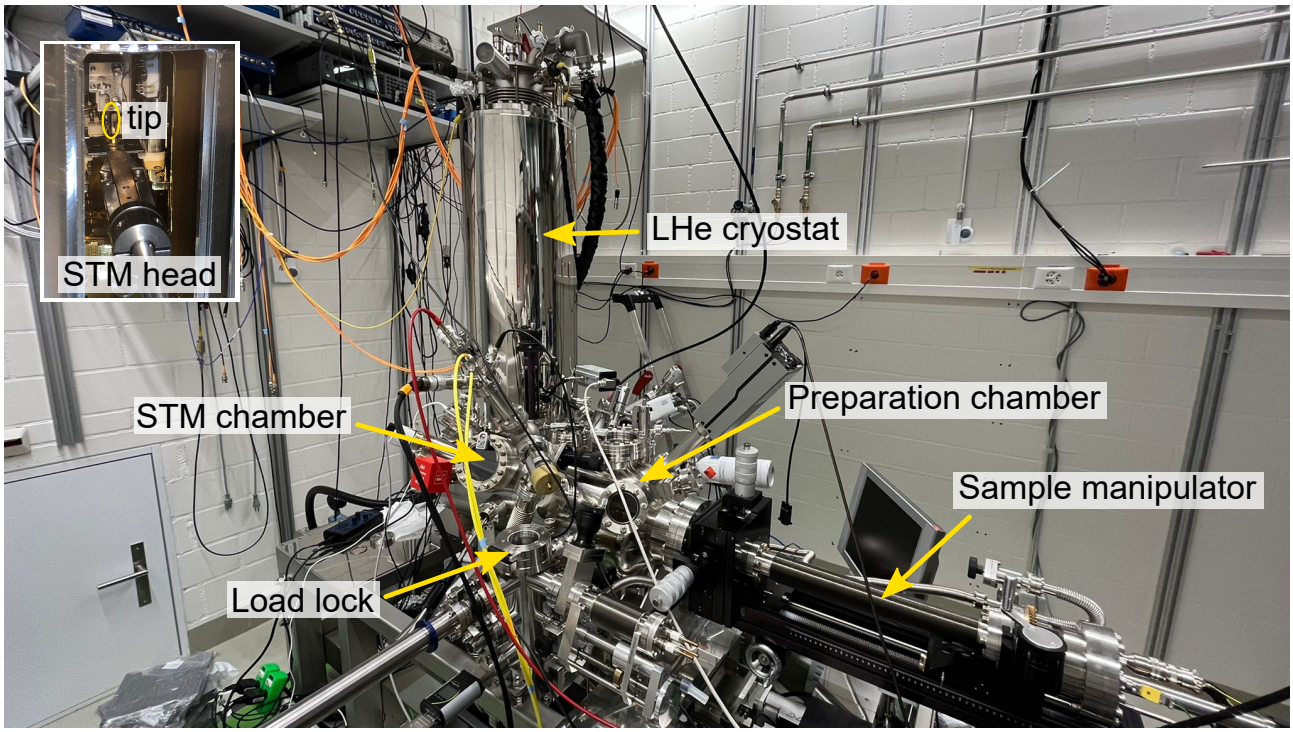

Figure 11: Createc low-temperature STM that has been used to verify the concept of sparse sampling in quasiparticle interference (QPI) imaging experiments. Some key components are highlighted with arrows. The inset shows the STM head containing the actual microscope that is situated below the liquid helium cryostat.

1.3.4 Sparse sampling on Au(111)

In the following we will apply the developed sparse sampling method in a real experiment. Figure 12 illustrates the sparse sampling procedure applied to a real STM environment using a Au(111) single crystal. Prior to its measurement the crystal is prepared using multiple sputtering-annealing cycles. The sputtering has been performed with Ar^+ ions followed by annealing of the sample. Figure 12(a) describes the data acquisition process. A suitable path consisting of M locations on a grid of size $n_x \times n_y$ pixels is chosen depending on the scan frame size and the desired resolution in q-space where the relation has been outlined previously. The STM tip is then controlled to take a spectrum at every of the pre-defined points originating from a near-optimal TSP path, which is loadded into the SPM controller. Such a spectrum is shown in the middle panel of Figure 12(a) and shows both $I(V)$ as well as the $\mathrm{d}I/\mathrm{d}V(V)$ (differential conductance) traces recorded using conventional bias spectroscopy on the Au(111) surface. At one energy one can define differential conductance traces that contain the differential conductance value at that specific energy for all the successively measured locations (right panel in Figure 12(a)). In Figure 12(b) the reconstruction of the sparse QPI patterns using a highly undersampled set of measurements is depicted. In a first step we populate 2D matrices of size $n_x \times n_y$ with the corresponding conductance values at one energy. The locations of these values are defined by the used tip path. We later decompose these matrices into two 1D vectors. One of these vectors contains the measurement values and will be denoted as $y \in \mathbb{R}^{M\times 1}$. The other vector q of same size encodes the designated locations in the original 2D matrix. We then define a normalized 2D Fourier transform operator $\Psi \in \mathbb{C}^{N\times N}$ where $N = n_x \times n_y$. Ψ represents our sparsifying basis of the signal we would like to reconstruct (QPI pattern situated in reciprocal q-space). In a next step we create the measurement matrix A by multiplying a restriction operator $\Omega \in \mathbb{R}^{M\times N}$ with Ψ. Ω is constructed such that row i has a single 1 in column $q(i)$. The named operators are provided within the scope of the SPARCO toolbox [37].

$$A = \Omega\Psi \quad \text{with} \quad \begin{cases} \Psi \in \mathbb{C}^{N\times N} \\ \Omega \in \mathbb{R}^{M\times N} \\ A \in \mathbb{C}^{M\times N} \end{cases} \tag{21}$$

Prior to any reconstruction we perform some data treatment. First, we remove the prominent outliers in the conductance traces that can be observed in the right panel of Figure 12(a). After that we conduct a global data normalization to avoid numeric precision issues in the later reconstruction process. In said data normalization we rescale all the values of the individual measurement vectors y_j to the range of $[0, 1]$ using the global minimum and maximum value of the stacked 2D matrix Y containing the conductance traces at all energies

$$y_j = \frac{y_j - \min(Y)}{\max(Y) - \min(Y)} \quad \text{with} \quad Y = [y_1\ ,\ y_2\ ,\ y_3\ ,\ \ldots\ ,\ y_{n_E}] \tag{22}$$

where n_E is the number of energy increments in a point spectrum. For every individual energy we then pass the corresponding measurement vector y_j and the measurement matrix A_j to the sparse least squares solver SPGL1 [38][39], which solves a basis pursuit denoise (BPDN) problem by minimizing the ℓ_1-norm of the solution x_j with quadratic constraints given as in Equation 19. We choose the error threshold σ_j in terms of the standard deviation of the respective measurement vector y_j. The algorithm will then reconstruct the sparse QPI signal as shown in the middle panel of Figure 12(b) where we show four different QPI patterns of Au(111). Next to the circular nearly free electron Shockley surface state, that is present in many noble metals, we also observe the scattering wave vectors belonging to the $(22 \times \sqrt{3})$ Heringbone surface reconstruction of Au(111) [40]. By applying an inverse Fourier transform we can transform from the sparse q-space back into the non-sparse real space where we originally measured the signal.

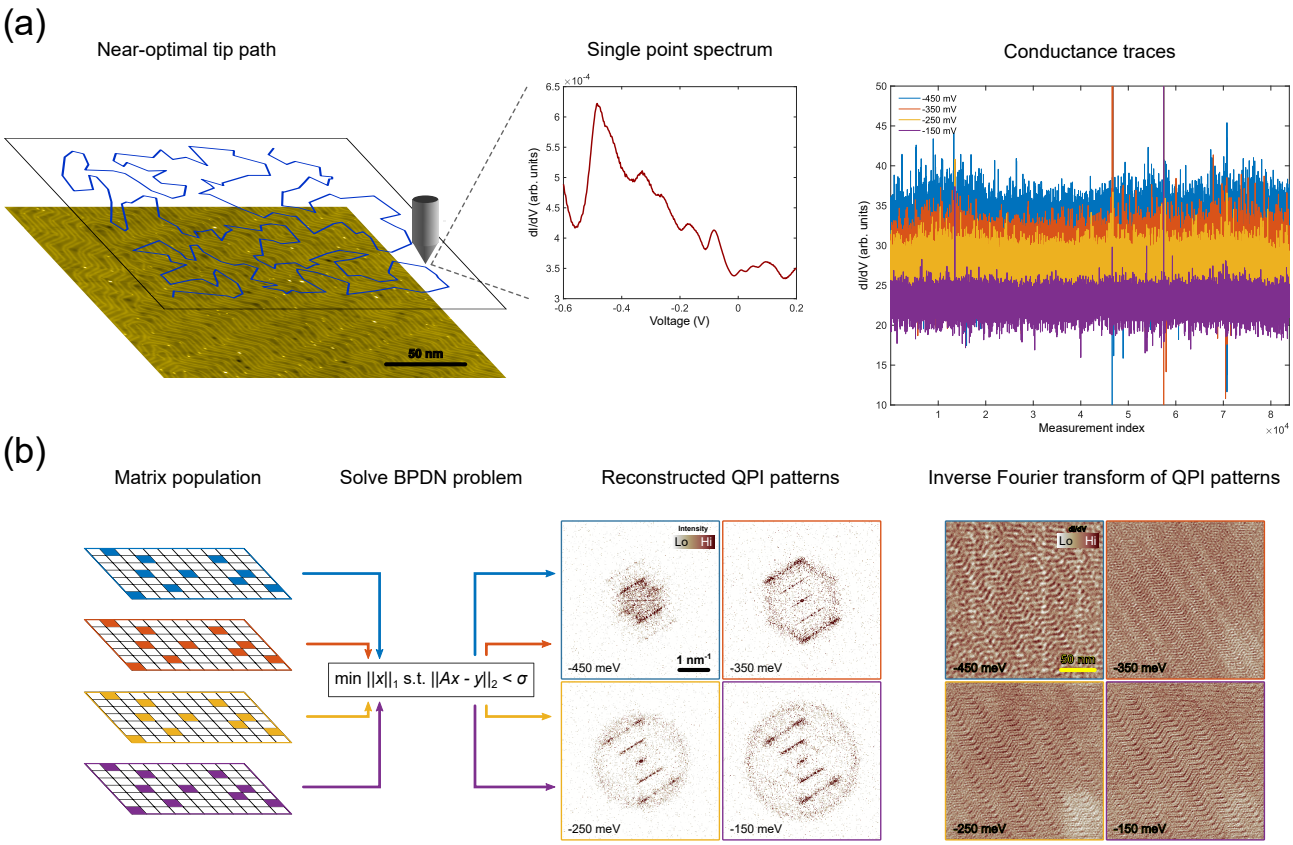

Figure 12: Sparse sampling procedure for speeding up QPI mapping experiments demonstrated on Au(111) at a temperature of 4.2 K. (a) Data acquisition process using conventional bias spectroscopy for measuring the single point spectrum for 1.6 seconds at predefined locations on the sample utilizing a near-optimal tip path. Here we measure on a 205×205 nm region around 84'000 points of a 2048×2048 pixel grid (≈ 2% sampling). The total measurement time is around 38 hours whereas a conventional QPI mapping experiment using the same size and averaging time for a single spectrum would have taken around 78 days. Stacking the conductance values for all the points at the same energy (bias voltage) results in conductance traces. (b) The sparse recovery of the QPI pattern is performed by populating 2D matrices with the measured conductance values from (a). Solving a basis pursuit denoise (BPDN) problem allows to reconstruct the sparse QPI information from a highly undersampled set of measurements. Here we show four different QPI patterns at -450, -350, -250 and -150 meV. Using an inverse Fourier transform we can obtain the equivalent picture in real space that is the modulation of the local density of states (LDOS) showing both the Herringbone surface reconstruction and standing wave patterns corresponding to a Shockley surface state.

The experiment performed in Figure 12 verifies the sparse sampling method by being able to faithfully reconstruct the famous Au(111) QPI signatures while also performing more than 30 times faster than the conventional mapping. We furthermore studied the dispersion relation obtained from the Au(111) sparse sampling measurements. Figure 13(a) shows a single QPI pattern (top) and a zoom-in of the inverse Fourier transform of the QPI pattern corresponding to the LDOS modulations of the mapped region. One observes both the Heringbone supermodulation as well as some circular standing wave patterns. In Figure 13(b) we show the obtained parabolic dispersion relation along two different directions in q-space marked by the two cones of different colors in Figure 13(a). The vertical direction (left side in (b)) corresponds to the direction of the Heringbone where we note the presence of a band gap at around -312 meV when the surface state coincides with the Heringbone periodicity (marked with an arrow). We attribute the creation of this band gap to the additional periodic potential induced by the Heringbone supermodulation [41][42].

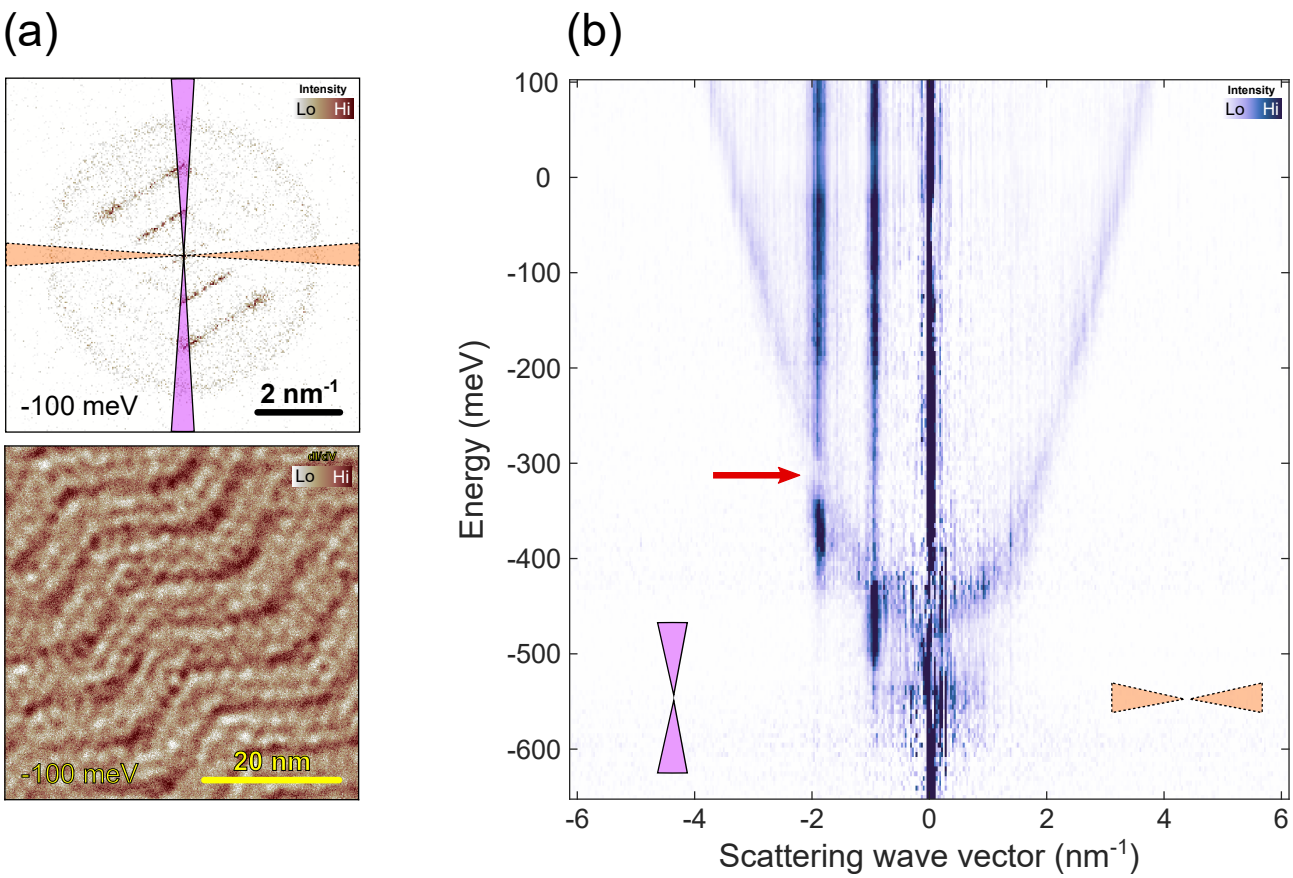

Figure 13: Sparse sampling applied on Au(111). The data corresponds to the one used in Figure 12 (2% subsampling of a 2048×2048 pixel grid on a 205×205 nm region). (a) Reconstructed QPI pattern at -100 meV (top) and corresponding LDOS (zoomed-in) obtained via an inverse Fourier transform of the former. (b) Dispersion relation obtained by averaging the regions described by the cones (opening angle of 20°) depicted in (a) along two different directions. We observe the creation of a band gap at around -312 meV due to the additional Heringbone potential [41][42]. For the nearly free electron dispersion we obtain an effective electron mass of $m^*/m_e = 0.222 \pm 0.002$. The calculation is described in Appendix A.2.

The initial QPI mapping experiments that took advantage of the developed sparse sampling method used conventional bias spectroscopy for obtaining a point spectrum whereas the method was later combined with parallel spectroscopy. The latter allows to decrease the measurement time of an individual spectrum, which in turn means that the overall measurement time is decreased by several orders of magnitude by combining sparse sampling and parallel spectroscopy [33]. This is due to the geometric relation between the two techniques where the total mapping time is given by

$$t_{\text{tot}} = pn^2 t_{\text{spc}} + t_{\text{travel}} + t_{\text{ctrl}} \quad \text{with} \quad p \in (0, 1] \tag{23}$$

where we assumed a square grid with $n \times n$ pixels. The sparse recovery using compressive sensing allows to measure only a small subset $M = pn^2$ of the total measurement region. The time for an individual spectrum t_{spc} can be reduced by a parallel spectroscopy technique, which will be introduced in the next section. Furthermore, moving the tip requires some traveling time represented by t_{travel}. This time can be reduced by using a near-optimal TSP path. Overhead time of the STM controller is contained in t_{ctrl}. It is straight-forward to see that any improvement of either reduction of the amount of measurements M or the reduction of the time required for a single spectrum will lead to strong decrease of the overall total mapping time.

1.3.5 Parallel spectroscopy

In an STM setup one generally makes use of conventional bias spectroscopy where a conventional lock-in amplifier is used to to measure the differential conductance of a sample at a specific location and with a specific bias voltage V_B applied

$$G(V_B, x, y) = (\mathrm{d}I/\mathrm{d}V)_{V_B}(x, y) \tag{24}$$

Instead of first measuring the $I(V)$ curve and then perform numerical differentiation a lock-in amplifier allows to obtain the $\mathrm{d}I/\mathrm{d}V$ curves in real time while increasing the signal-to-noise ratio (SNR) [43]. A normal lock-in amplifier works by applying a small AC-modulation V_B at one frequency while demodulating a narrow frequency range around V_B. The measured signal is then multiplied with a reference sine-wave having the same frequency as the modulation. Integrating the resulting signal and applying a low-pass filter will lead to the removal of all components that are not close or equal to the used modulation frequency. The SNR hereby increases the longer this integration is performed. For multiple applied bias voltages this averaging will lead to a relatively long measurement time needed for a single spectrum in order to maintain a desirable signal quality.

In recent years a novel phase-coherent lock-in technique has been developed, which allows to reconstruct non-linear $I(V)$ characteristics with an AFM [44]. The measurement time resulting from this technique can be orders of magnitudes smaller for a single spectrum than using conventional bias spectroscopy. Instead of applying a small AC-modulation around the applied bias voltage V_B this novel lock-in technique utilizes a large AC-modulation, which covers the whole energy range of interest. When such a large AC-modulation is applied onto an $I(V)$ curve containing non-linearities higher harmonics are being generated that are visible in the Fourier transform of the current. This can be seen by expanding the $I(V)$ curve in a Taylor series

$$I(V) = c_0 + c_1 V + c_2 V^2 + c_3 V^3 + ... \tag{25}$$

One can then apply an AC-modulation $V(t) = A\cos(\omega t)$ with a large amplitude A and $\omega = 2\pi f_{\mathrm{ex}}$ where f_{ex} is a large excitation frequency. Then one can write the AC signal as

$$\begin{aligned} I(V(t)) &= c_0 + c_1 A\cos(\omega t) + c_2 A^2\cos^2(\omega t) + c_3 A^3\cos^3(\omega t) + ... && (26)\\ &\equiv c_0 + a_1\cos(\omega t) + a_2\cos^2(\omega t) + a_3\cos^3(\omega t) + ... && (27)\end{aligned}$$

Using trigonometric identities one can relate the power factors of the cosine terms to cosine terms that contain the higher harmonics of the base frequency ω

$$\begin{aligned} a_2\cos^2(\omega t) &= \frac{1}{2} + \frac{1}{2}\cos(2\omega t) && (28)\\ a_3\cos^3(\omega t) &= \frac{3}{4}\cos(\omega t) + \frac{1}{4}\cos(3\omega t) && (29)\end{aligned}$$

These higher harmonics are therefore generated when a large AC-modulation is applied onto a signal containing non-linearities. The basic concept is depicted in Figure 14 using two different kinds of spectra that both show non-linear characteristics in their current-voltage curves. The higher harmonics are visible in the Fourier transform of the $I(V)$ traces, which were generated with an AC-signal (cosinusoidal) with excitation frequency f_{ex} and an amplitude that covers the complete desired energy range. The simultaneous demodulation of these harmonics allows to reconstruct the original $I(V)$ curves using an inverse Fourier transform and from there obtain the differential conductance $\mathrm{d}I/\mathrm{d}V(V)$ by numerical differentiation. In this work this was achieved by using a multifrequency lock-in amplifier from Intermodulation products. Due to the fact that the tunneling junction is excited with a large voltage amplitude there will be displacement currents generated due to stray capacitances. These need to be compensated appropriately in order to not saturate the pre-amplifier [33].

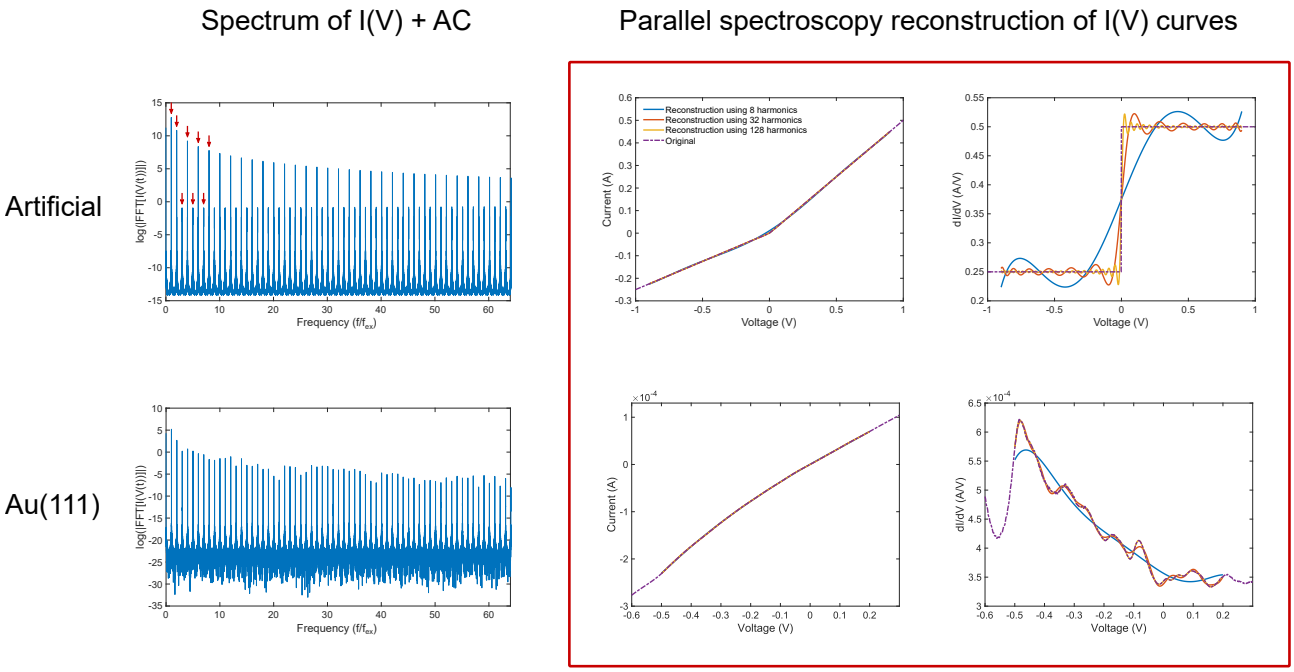

Figure 14: Higher harmonics generation when applying a large AC modulation onto a nonlinear $I(V)$ characteristics visualized using two examples of an artificial and a Au(111) spectrum. Some of the higher harmonics are marked with red arrows. By measuring a certain number of said harmonics in parallel one is able to reconstruct the corresponding $I(V)$ traces and from there the differential conductance $\mathrm{d}I/\mathrm{d}V(V)$. The more harmonics are used the better the reconstructed output becomes.

1.3.6 Sparse sampling on Bi2212

The initial QPI imaging experiments on Au(111) using sparse sampling have proven to be very reliable at extremely low sampling rates. Because of the natural sparsity of QPI patterns one can extend the sparse sampling technique to any other kind of 2D quantum material. One of these materials that has been studied within the scope of this thesis is the high temperature cuprate superconductor $Bi_2Sr_2CaCu_2O_{8+\delta}$ (Bi2212) [33]. Bi2212 has been extensively studied by QPI experiments for many years [26][45][46][47]. While Au(111) is a single crystal, that can be prepared by sputtering and annealing cycles, the used Bi2212 sample was in the form of a flaky single crystal, which had to be glued to the sample plate, cleaved in UHV and directly inserted into the cold STM microscope. Since Bi2212 is a type-II superconductor vortices will emerge upon the application of an external magnetic field. A vortex is a region where the magnetic field lines penetrate the material rendering the region normal conducting. The appearance of vortices will lead to additional vortex-induced quasiparticle states in reciprocal scattering space [46]. In our case the magnetic field was applied using a sacrificial magnet that was fixed on the sample holder plate (Liu et al. in-preparation). In Figure 15 we show the reconstruction of the scattering wave vectors of Bi2212 under the influence of mentioned magnetic field. The individual spectra have been recorded using parallel spectroscopy as described earlier where a single spectrum was recorded in 100 ms. In Figure 15(a) we show two different QPI reconstructions and the corresponding LDOS modulations obtained via an inverse Fourier transform of the former. Specific scattering wave vectors such as the supermodulation Q_{sm} and the Bragg peaks Q_x, Q_y are highlighted. Furthermore, we observe checkerboard structures q_x and q_y at around 0 meV bias voltage [46]. In Figure 15(b) we show the dispersion relation obtained by line profiles along the specified directions in q-space that is along the Bragg peaks Q_x and along the supermodulation Q_{sm}.

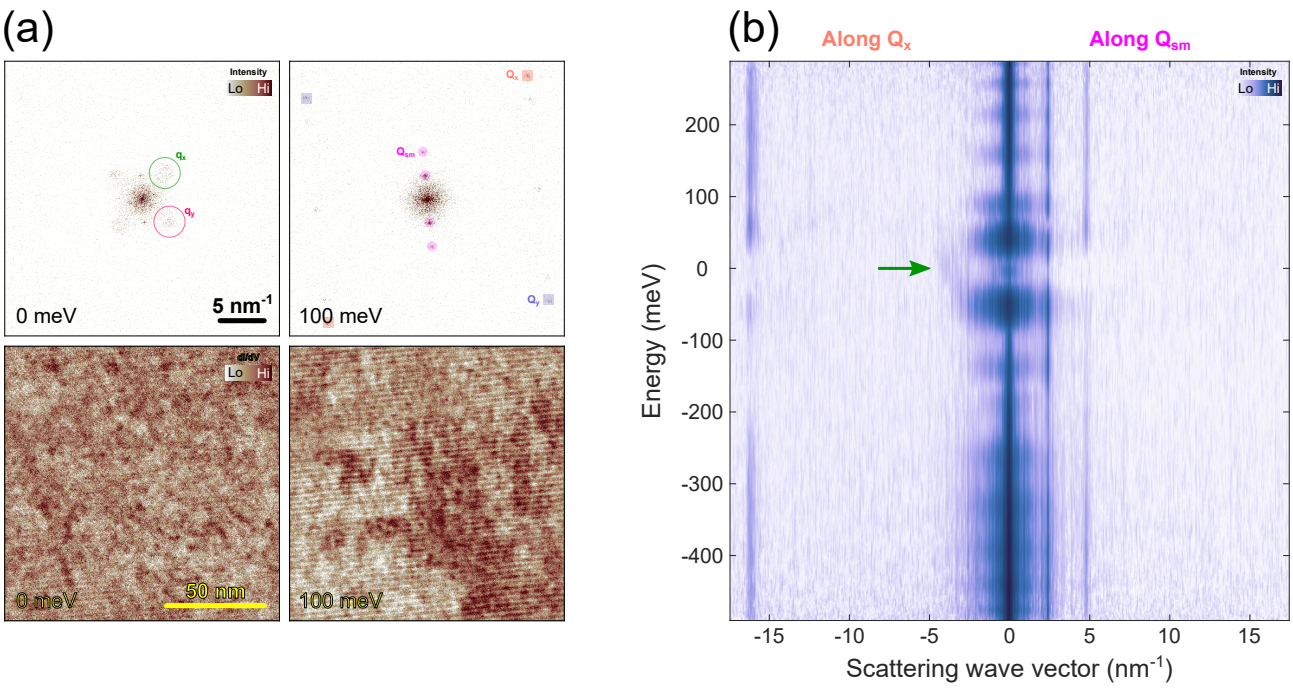

Figure 15: Sparse sampling on the high temperature cuprate superconductor Bi2212 at a temperature of 4.2 K. The grid dimension is 2048×2048 pixel on a 135×135 nm region where a full dI/dV spectrum was obtained on around 3% of the total available locations using parallel spectroscopy. The mapping time was around 7 hours. A regular QPI experiment using conventional bias spectroscopy would have taken over 40 days assuming 1 second per point spectrum. (a) Reconstructed QPI patterns at 0 and 100 meV (top row) and the corresponding LDOS modulations (bottom row). (b) Dispersion relation along the Bragg peak Q_x and supermodulation Q_{sm} direction averaged using a line width of 20 pixels. We note the appearance of checkerboard structures (q_x, q_y) at around 0 meV in the direction of Q_x (indicated with an arrow).

1.3.7 Sparse sampling on WTe2

Another 2D quantum material on which we have tested the sparse sampling methodology is the type-II Weyl semimetal WTe_2 where non-closed Fermi arcs appear on the surface that are related to the non-trivial topological nature of the bulk states [48][49]. Similar to Bi2212, WTe_2 samples are in the form of flaky single crystals obtained by using the exfoliation technique. By cleaving the layered material in UHV one achieves a clean and flat surface to be studied with STM. In Figure 16 we show the result of QPI imaging on WTe_2 where we sparsely sampled the LDOS at a small set of random location using conventional bias spectroscopy. Figure 16(a) shows a single reconstructed QPI pattern (top) and the corresponding LDOS modulations (bottom) at the Fermi energy. We clearly notice the atomic Bragg peaks Q_x and Q_y as well as the appearance of dispersing states close to the zone center in reciprocal space. In the LDOS pattern these states correspond to the additional stripes perpendicular to the orientation of the atomic rows. In Figure 16(b) we plot the obtained energy dispersion relation along the two different indicated directions of the Bragg peaks Q_x and Q_y. We observe a linear dispersion relation around the Fermi energy.

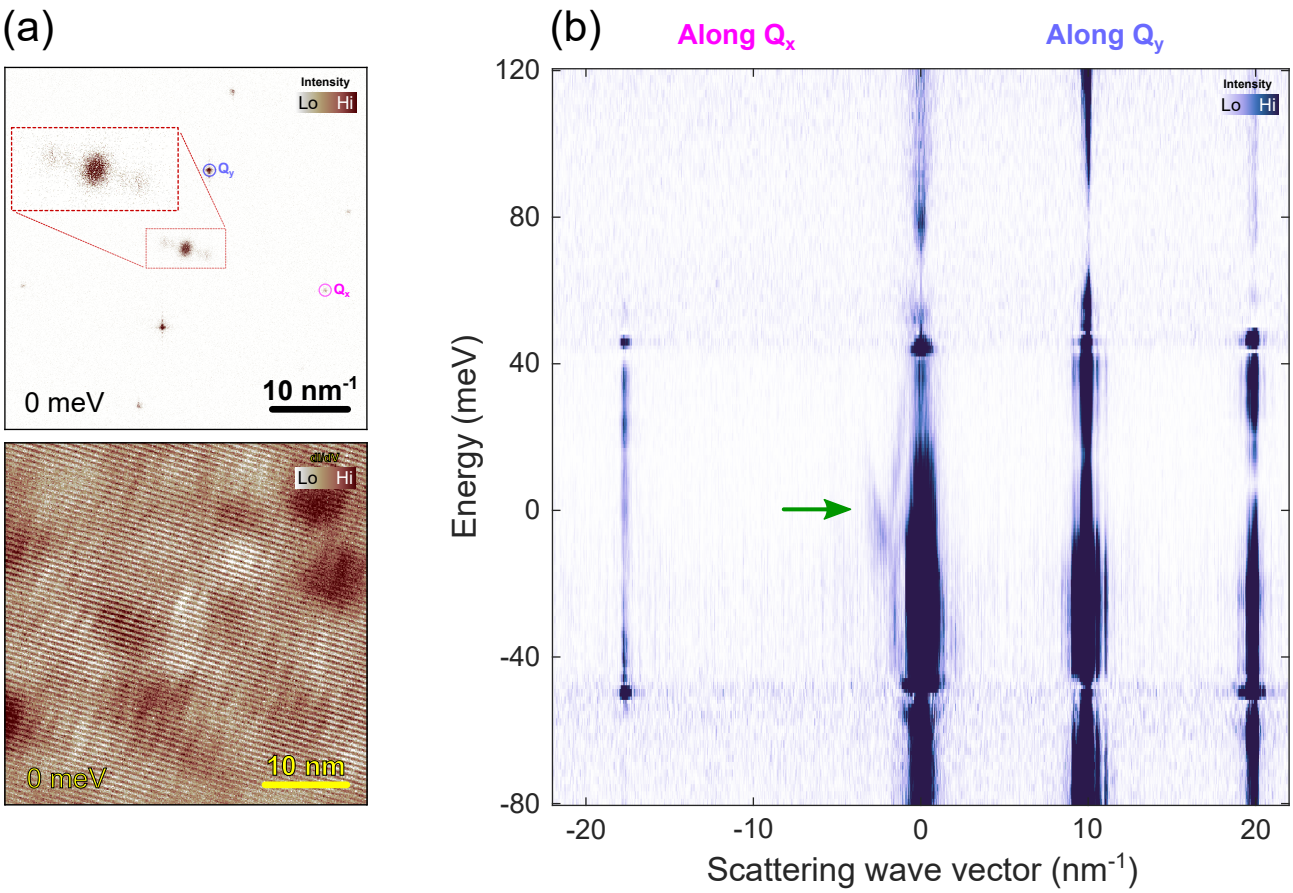

Figure 16: Accelerated QPI mapping on the type-II Wely semimetal WTe_2 using sparse sampling at 4.2 K. The scan frame size was 200×200 nm and a full dI/dV spectrum has been measured at around 3% of the total locations on a 2048×2048 pixel grid using conventional bias spectroscopy. The mapping time was around 40 hours. (a) Reconstructed QPI pattern at 0 meV (top) shows both atomic resolution as well as dispersing states close to the zone center. A zoomed in version of the corresponding LDOS (bottom) shows standing wave patterns corresponding to the dispersive states obtained via an inverse Fourier transform of the top panel. (b) Dispersion relation along the different reciprocal lattice vectors Q_x and Q_y. We find a linear dispersion close to the Fermi energy.

Although we detect some dispersing features close to the zone center (see Figure 16(b)) the full Fermi arc states as for example seen in [48] are not visible. For once this could be attributed to a to low sampling rate since the higher q-vectors of the unseen states are less sparse and thus require more sampling to be reconstructed. Another reason could be a non-ideal tip. However, we also think that experimental nonlinearities have impacted the results. One of such nonlinearities is thermal drift of the sample due to local thermal gradients in the STM chamber. These gradients will lead to a displacement of the sample with respect to the fixed scan frame within we acquire the LDOS measurements over time. Since we utilize a tiled traveling salesperson tip path we effectively measure local smaller sections first before proceeding to the next one. Neighboring tiles might also not be measured consecutively but rather after a longer measurement period. Because of the time-dependent thermal drift there will be a shift of each tile with respect to its initial position. This shift depends on the direction of the thermal drift at the time where the points contained in a tile have been measured. Therefore, there will be a different shift of each tile, which causes a phase-difference at the boundaries of neighboring tiles. This phase-shift generally gets larger for adjacent tiles that show a large difference in their relative acquisition time stamp. The consequence is the appearance of an artifact in the recovered QPI pattern, which we denote as a tiling artifact. In Figure 17 we illustrate the origin of said artifact. Figure 17(a) shows the LDOS corresponding to a sparsely reconstructed QPI pattern of the same WTe_2 dataset as in Figure 16. The LDOS is superimposed with the tiling arrangement and the visiting order of the individual tiles. Because of the potentially large time delay between the measurement of neighboring sections there will be a phase-shift that can be seen along the edges of some of the tiles (best seen by

zooming into Figure 17(a)). Because of the high sparsity of the atomic rows there usually are enough measurements in a single tile to recover the corresponding q-vectors. And due to these phase-shifts one obtains additional Bragg peak satellites in the reconstruction, which are off-setted by some amount with respect to the true Bragg peak as shown in Figure 17(b). This result is furthermore visually verified using simulated data in Figure 17(c)-(d). Note that in Figure 17(b) there are more prominent side peaks in the vertical direction, which can be explained by the fact that the phase-shift of the tiles appears more in horizontal direction in real space. Such artifacts are already visible in the QPI pattern in Figure 16(a) but they are not as prominent at that particular energy.

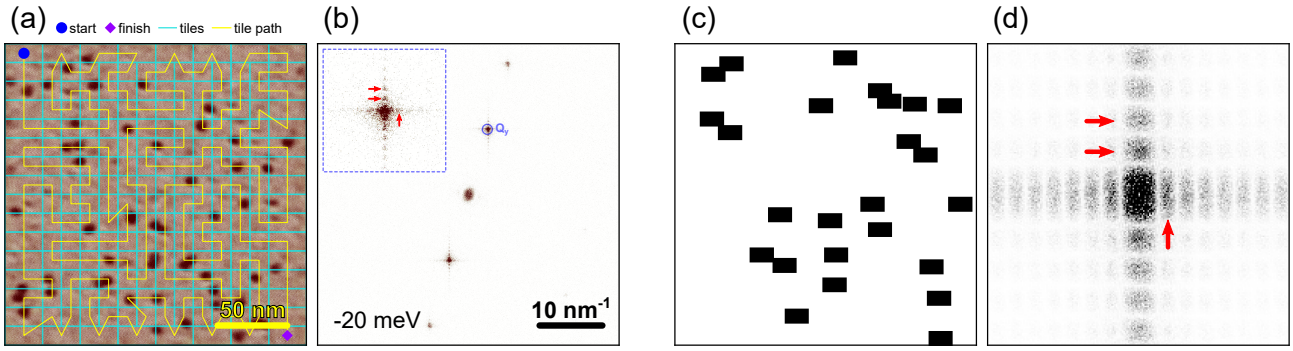

Figure 17: Using a subdivided tip path leads to the appearance of tiling artifact in the reconstruction around the sparse Bragg peaks. (a) 200×200 nm region showing the dI/dV (LDOS) map at -20 meV corresponding to the reconstructed QPI pattern in (b). The individual tiles (light blue) and the order of how the tiles are visited (yellow) is indicated. A phase-shift between some neighbouring tiles emerges because of thermal drift of the sample during the measurement (mapping time of around 40 hours). (b) QPI pattern at -20 meV showing additional Bragg peak satellites around Q_y corresponding to mentioned phase-shift in real space, which we refer to as a tiling artifact (marked with arrows). (c) Simulated tiling artifact using a mask with randomly placed tiles on a square grid. The random placement refers to the individual phase-shift of adjacent tiles in (a). (d) 2D fast Fourier transform of (c) showing a DC component with additional satellites in both horizontal and vertical direction.

We want to point out that the described tiling artifact in Figure 17 is not an artifact of the reconstruction algorithm but rather due to the partial loss of the spatial relationship between the assumed locations (tip path) and the true locations where the LDOS measurements have been obtained. In order to correct this artifact - that is recovering the phase-coherence between the individual tiles - one would need to shift the individual measurement points to their true location relative to all the other points prior to feeding the data into the sparse recovery solver. We will further elaborate on this topic in Section 1.5.

1.4 Adaptive sparse sampling

1.4.1 Introduction

In the previous sections we described the idea behind sparse sampling and how it can be applied onto an experimental setup in the form of quasiparticle interference (QPI) mapping using a Scanning Tunneling Microscope (STM). We have shown that the method allows to fundamentally speed-up the mapping time for QPI measurements. By combining the reduced amount of local density of states (LDOS) measurements with parallel spectroscopy the reconstruction of the very sparse scattering wave vectors becomes orders of magnitudes faster than the conventional grid measurements. This allows to enlarge the potential parameter space, one could probe considering the very limited hold-time of modern liquid helium cryostats. It also implies that the reciprocal space resolution can be enhanced by both measuring a larger sample region and at the same time increasing the pixel density to include very short-range real space modulations. Here we want to note that the benefit of very large pixel grids did barely impact the considered cases as most of the studied wave vectors are situated close to the zone center for which also a smaller grid would have been sufficient. However, for certain types of materials where one has both very long- and short-range modulations it might be of major importance to have both a large scan area but also a large enough pixel grid to enable a large field of view in large reciprocal.

The presented method for fast QPI mapping experiments [25][33] still raises some open questions. On one hand it doesn't provide a quantitative measure for the amount of subsampling required to obtain a desired QPI quality. Even though the compressive sensing theory gives a prescription of the required number of measurements to allow a successful recovery [11][20], in reality obtaining this quantity is oftentimes not possible. The reason is that said prescriptions depend on a a-priori insight about the signal-to-noise ratio and the sparsity level. Both of these measures are generally unknown for the studied systems even though one might get an intuition by studying the statistics of many different systems. This could allow to extract specific patterns that would help specifying said parameters using for example machine learning based approaches. Another issue that was already mentioned in Section 1.3.7 is about the tiling artifact that we encountered in systems with prominent short-range periodicities. These issues are the main incentive for a more general sampling approach that we introduce as adaptive sparse sampling [50]. Similar to the previous procedure we first will illustrate the new sampling method using a simulated environment and then validate it on actual experimental STM data. The underlying concept of adaptive sparse sampling is depicted in Figure 18. Instead of defining the total number of measurements prior to the experiment we iteratively increase the amount of subsampling during a running experiment by cumulatively stacking smaller traveling salesperson (TSP) tip paths that cover the complete scan frame of interest. These are chosen such that every path consists of a few thousand points that can be efficiently calculated without sub-tiling. By pre-calculating a set of dozens of such tip paths we have access to a large pool of measurement locations that can be re-used for different experiments. The adaptive sparse sampling variant is initialized by measuring a regular topography preferably around topographic features that might serve as distinct references for later image registry to account for thermal drift. Such a region is depicted in Figure 18 by the small inset (red square). By then continuously acquiring a subset of LDOS measurements using a small-scale tip path followed by a topographic scan one cumulatively increases the number of measurements as shown in the bottom row of Figure 18. At the same time the reconstruction quality gradually improves. This effect can be observed in the top row of Figure 18 where the surface state ring as well as the Bragg peaks become more and more prominent with an increased number of iterations. This allows to accumulate measurements until a satisfactory signal quality is achieved.

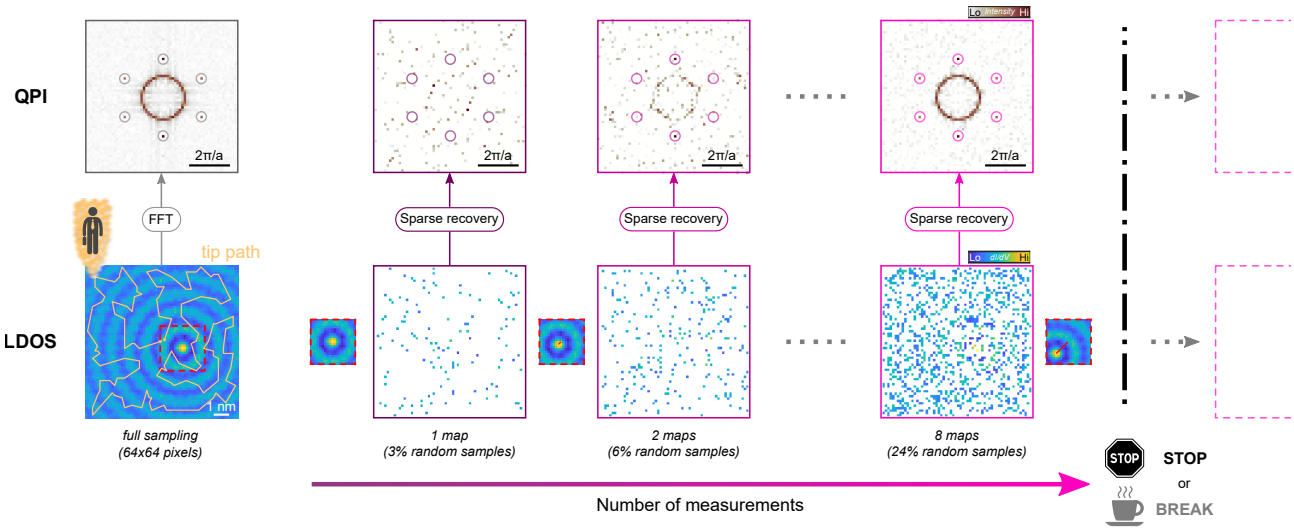

Figure 18: Adaptive sparse sampling concept illustrated using a simulated Shockley surface state with atomic corrugation and added Gaussian noise. The leftmost panel shows the quasiparticle interference (QPI) pattern (top) obtained via a Fourier transform of the total local density of states (LDOS) map (bottom). By adaptively stacking consecutive tip paths one can accumulate LDOS measurements in an iterative fashion. We observe that the quality of the sparsely reconstructed QPI pattern is gradually improved by an increasing number of cumulative measurements. One can further intersect the maps with smaller topographic scans or LDOS maps for aligning the individual measurements prior to the sparse recovery. The red line in said insets indicates the direction of thermal drift over time. This adaptive sparse sampling concept furthermore allows to pause, resume, or interrupt the measurements.

All the small TSP paths have their initial and final location in the center of the frame to minimize the average tip travel distance when switching between the two modes (TSP path to a regular topography scan and vice versa). Since the selected topographic features may be off-center, we generate a random jittery path motion to connect the two modes. This also allows the system to settle piezoelectric creep that might emerge because of larger changes in the scan piezo voltages when repositioning the tip.

Compared to the regular sparse sampling method described in previous sections, this new adaptive sampling scheme has several advantages. One of them is the absence of tiling artifacts because the adaptively stacked TSP paths now cover the complete scan frame. Hence there are less local measurement regions and with that it is more likely that some contribution long range wavelength components are picked up in a single iteration. In scenarios where one has locally defined states one could also make use of an informed sampling approach to sample more dense in such regions [25]. Another advantage compared to regular sparse sampling is the removal of the a-priori knowledge of the exact amount of measurement needed to achieve a certain QPI quality. If in the case of regular sparse sampling the amount of measurements M_{tot} is chosen too small one would have to repeat the experiment with a higher number of measurements. Thus a higher knowledge of the sparsity level is necessary to determine the right amount of subsampling. In the adaptive sparse sampling scheme, a user can use the knowledge of the current reconstruction quality to decide whether to stop or to continue the experiment after every successive iteration. Lastly, the adaptive measurement technique with interleaved topographies potentially enables a post-processing linear drift correction when the time between the individual TSP paths is not too large. This will be further discussed in Section 1.5. In Figure 19 we compare the two sparse sampling methods in terms of their respective flowcharts.

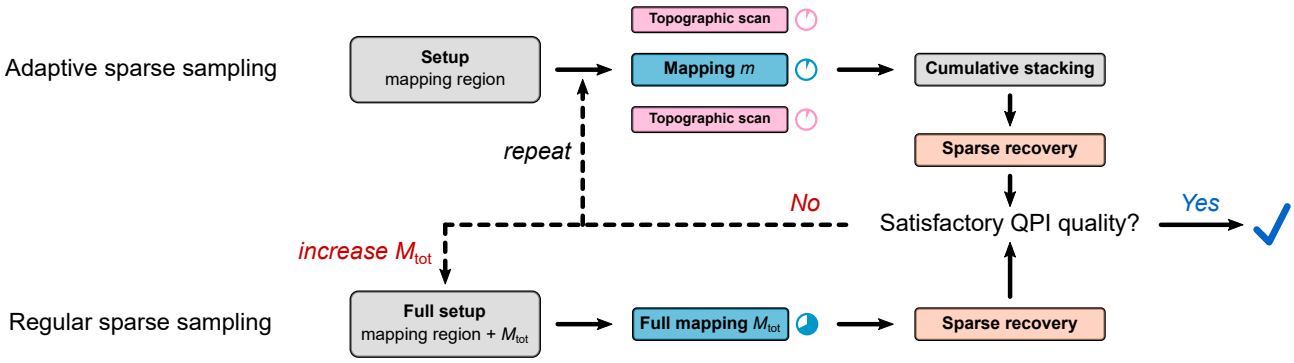

Figure 19: Flowcharts of adaptive and regular sparse sampling. While in a regular sparse sampling experiment one needs to define the exact amount of measurement prior to the experiment, in the adaptive sparse sampling procedure the user can repetitively stack smaller maps until a desired QPI quality is achieved as illustrated in the top panel of Figure 18. The small schematic clocks next to mapping step indicate the massive time difference between the two techniques where the time of choosing an insufficient amount of sampling is in vain in the case of regular sparse sampling.

1.4.2 Adaptive sparse sampling on Au(111)

To validate the new adaptive sampling framework we again have chosen the model system Au(111). In our previous work [33] we measured its dispersion relation using various combinations of conventional bias spectroscopy, parallel spectroscopy and regular sparse sampling. In Figure 20 we show the results from the sparse reconstruction of an iteratively sampled LDOS region with a resolution of 1024×1024 pixel (physical size of 150×150 nm). In the current demonstration we use single TSP paths where each of them contains 3000 points, which have been calculated using the algorithms mentioned in Section 1.3.2. We denote this number of 5×3000 points as one path segment. After each of such path segments we measure a smaller reference region (around 100×100 pixel) where we measure a full spectrum at every point using parallel spectroscopy. Each of these smaller reference maps took around 20 minutes, which can be interpreted as a short interruption of the mapping scheme. Figure 20(a) shows the dispersion relation obtained by azimuthally averaging the reconstructed QPI patterns after 6'000, 15'000 and 60'000 measurements. In Figure 20(b) we show a single QPI pattern consisting of a circular surface state and the Heringbone signature. In Figure 20(c) the respective LDOS maps are shown obtained by an inverse Fourier transform of the respective QPI patterns in Figure 20(b). From the evolution of Figure 20(a) we notice that more and more band structure details start emerging when the number of measurements is increased. From the dispersion plots at lower sampling, we also see that low-momentum states appear earlier than high-momentum states. This is attributed to the sparsity of the surface state that is related to its circumference. At lower energies the surface state is represented by less wave vector values due to its smaller circumference. The signal is therefore more sparse at lower energies and less measurements (equations) are necessary to solve the basis pursuit problem. Similarly, we can say that at higher energies more wave vector values are required to construct a full „circle“ thus decreasing the sparsity level and rendering the reconstruction more difficult. A similar argumentation can be applied to the Heringbone contribution that is visible at low sampling rates as well. Interestingly, we do not see the appearance of a band gap at around -312 meV as we saw it with the regular sparse sampling approach (see Figure 13). The reason for this hasn't been investigated further but it might be related to a non-ideal tip configuration or the usage of parallel spectroscopy, which generally reduces the overall spectral quality.

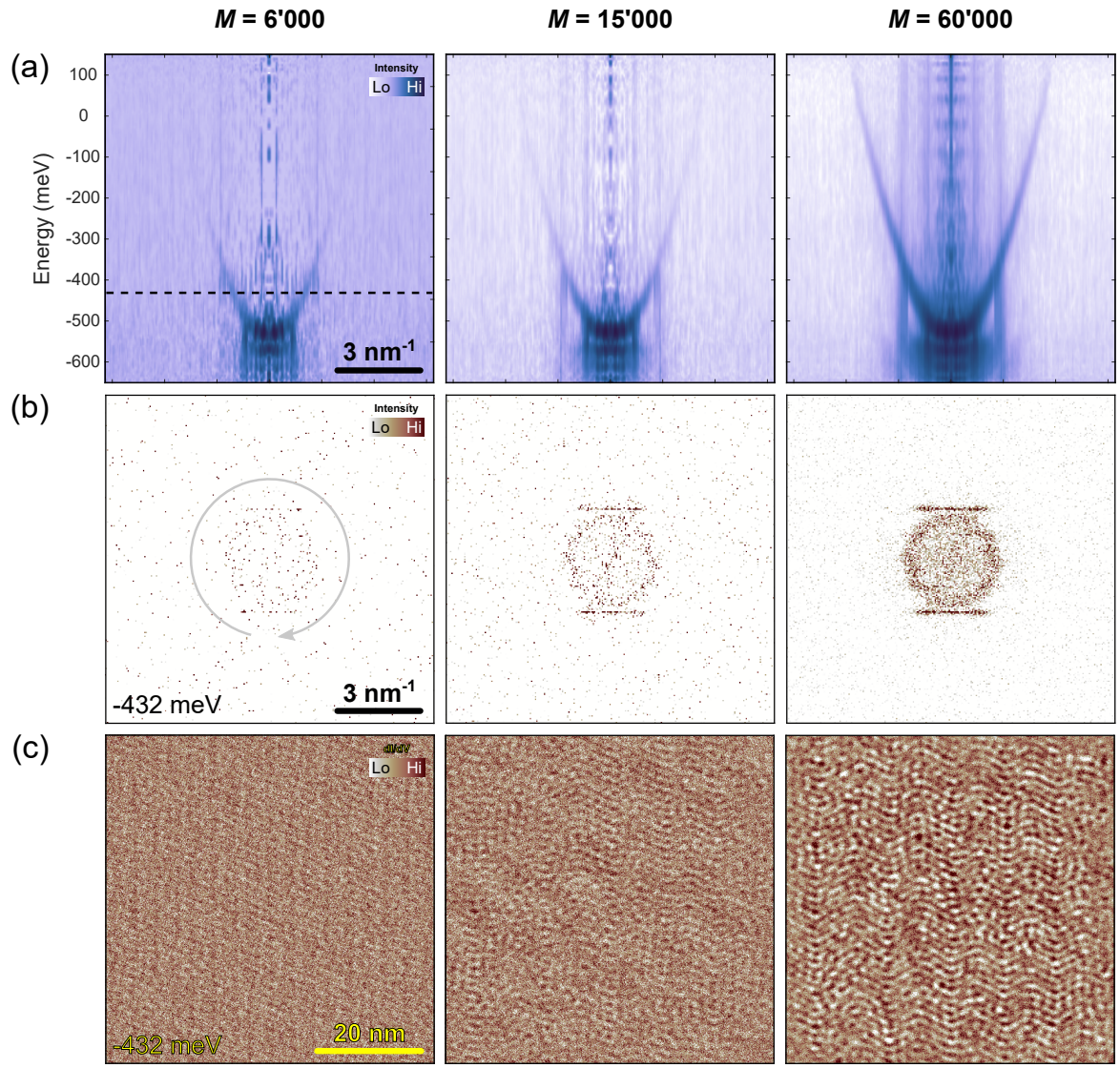

Figure 20: Adaptive sparse sampling on Au(111) at 4.2 K where the individual point spectra have been obtained using parallel spectroscopy. (a) Parabolic dispersion relations of the Shockley surface state obtained after the given number of iterative measurements by azimuthally averaging the reconstructed QPI patterns. The more LDOS measurements (M) are used the better the quality of the reconstruction becomes. From these dispersion plots we obtain an effective electron mass of $m^*/m_e = 0.200 \pm 0.002$, which is around 10% smaller than the effective mass obtained from prior regular sparse sampling measurements. A reason for this might be a wrongly calibrated voltage provided by the multi-frequency lockin amplifier. (b) Evolution of a single QPI pattern at -432 meV with an increasing number of adaptive measurements. (c) The inverse Fourier transform of (b) shows standing wave patterns and the Heringbone reconstruction of the Au(111) surface.

The dispersion plots shown in Figure 20(a) have been obtained after performing a global background subtraction that conserves the scattering wave vectors in q-space. The justification for using said background correction is found in slight and spontaneous tip changes that will lead to abrupt changes in the conductance traces. We assume that the measured conductance spectra $\mathrm{d}I/\mathrm{d}V(\mathrm{E})$ are a convolution of tip and sample density of states

$$\mathrm{d}I/\mathrm{d}V(E) \propto \rho_T \rho_S(E) \tag{30}$$

A change in ρ_T will manifest itself in an energy dependent multiplicative effect on the conductance trace. To combine different LDOS measurements obtained from different path segments and different tip density of states we apply the following energy dependent correction to all segments j

$$[\mathrm{d}I/\mathrm{d}V(E)]_j = \frac{[\mathrm{d}I/\mathrm{d}V(E)]_j - \mu_j}{\sigma_{\mathrm{ref}}}\sigma_j + \mu_{\mathrm{ref}} \tag{31}$$

where μ_j, σ_j are the mean and standard deviation of the conductance trace measured in the j-th path segment. μ_{ref}, σ_{ref} are reference values that we choose to stem from the first segment because in most cases the tip was stable in the beginning of the experiment. This correction shall be applied onto segments that show no tip changes during their data acquisition. If a tip change would occur during a segment one could think about discarding said segment prior to the reconstruction process. We want to note that a different tip configuration could have a detrimental impact on the reconstruction because of the spatial convolution of tip and sample density of states. One example of such a tip configuration would be a nonconducting tip. Furthermore, because this correction treats the individual energies independently the average spectrum might change slightly as shown in Appendix A.4. Depending on the requirements regarding the experimental results one should therefore gauge whether to apply this correction or not. In some sense the just described method of correcting the conductance traces can also be understood as a method of salvaging datasets that, in a normal conventional QPI mapping experiment, probably would be discarded. In Figure 21(a) we show an uncorrected conductance trace of the Au(111) measurement where a sudden tip change can be observed after around three path segments leading to intensity variations in the dispersion plot to the right. The sudden increase of the conductance is most likely due to the tip picking up some mobile molecule from the Au(111) surface. The dashed red lines indicate the location of small interleaved conventional grid regions outlined in Figure 21(b). We also see that the tip changed back to its original state somewhere in the middle of the 5th conventional grid region in Figure 21(b) when it interacted with a defect (marked with purple box). In Figure 21(c) we show the same data after applying described background correction, which leads to the dispersion plot to the right of panel (c) (same as in Figure 20).

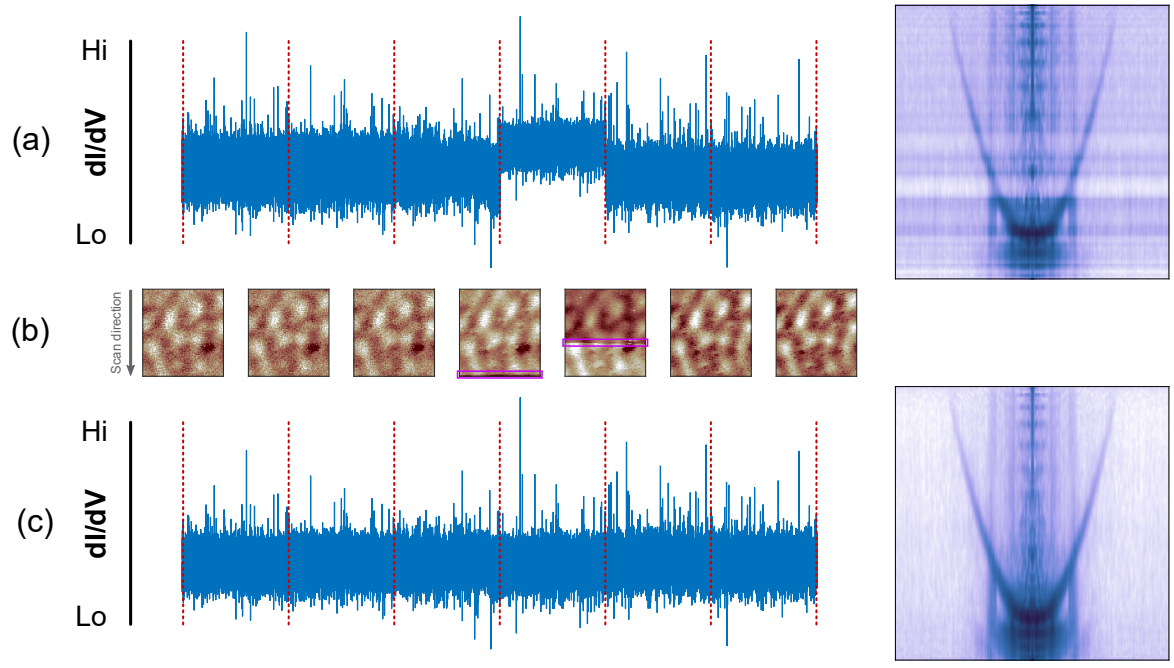

Figure 21: Accounting for tip instabilities using a global background subtraction to remove sudden tip changes visible in the conductance traces. (a) Raw conductance trace hosting a tip change that leads to a sudden jump in the $\mathrm{d}I/\mathrm{d}V$ values. This results in the intensity variations that can be observed in the dispersion plot to the right of (a). (b) Conventionally measured maps with sizes of around 100×100 pixels that have been interleaved after every path segment (marked by the dashed red lines in (a) and (c)). Purple boxes mark the locations where the tip changed. (c) Corrected $\mathrm{d}I/\mathrm{d}V$ trace after applying a background correction that preserves the spatial frequency content. No more intensity variations are visible in the resulting dispersion plot.

The adaptive sparse sampling experiment performed on Au(111) was set up to measure all the path segments with interleaved conventional LDOS scans in one go and later the data was reconstructed using different numbers of segments. The reason for not making the experiment directly interactive as in the flowchart depiction in Figure 19 was related to the lack of software capabilities of the SPM controller. In that sense the actual experiment was not really adaptive but rather served as a proof of concept. To fully make use of the iterative stacking of measurements the user would have to decide whether the QPI quality is sufficient or not at runtime. One option to do this is to visually inspect the result obtained after k segments. However, this in some sense requires an intuition about the sparse wave vectors that one would like to resolve. Another option would be to track a quantitative measure of the reconstruction, which serves as a decision tool for whether to stop or to continue the experiment. In Figure 22 we show that one could for example track the relative mean absolute error (MAE). If we assume two matrices $\hat{X}$ and X with each n rows and m columns, then we define their relative MAE as

$$\text{Relative MAE}(\hat{X}, X) = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{|\hat{X}_{ij} - X_{ij}|}{X_{ij} + \varepsilon} \tag{32}$$

where $\varepsilon > 0$ is a small constant for numerical stability. In Figure 22(a) we use the same simulated surface state as in Figure 18 with known QPI ground truth to conceptually show that the relative MAE between the reconstruction with $k-1$ segments and the one with k segments is decreasing as k increases. We find a similar behaviour when using the measured Au(111) data in Figure 22(b). There we additionally observe that the slope of the decrease reflects the reduced sparsity at higher energies that require more sampling to resolve. We furthermore realize that the data points in both (a) and (b) follow a power law, which is fitted by the solid/dashed lines. We conclude that this or a similar measure can be obtained at runtime and could thus serve as a feedback control about the status of the QPI data quality.

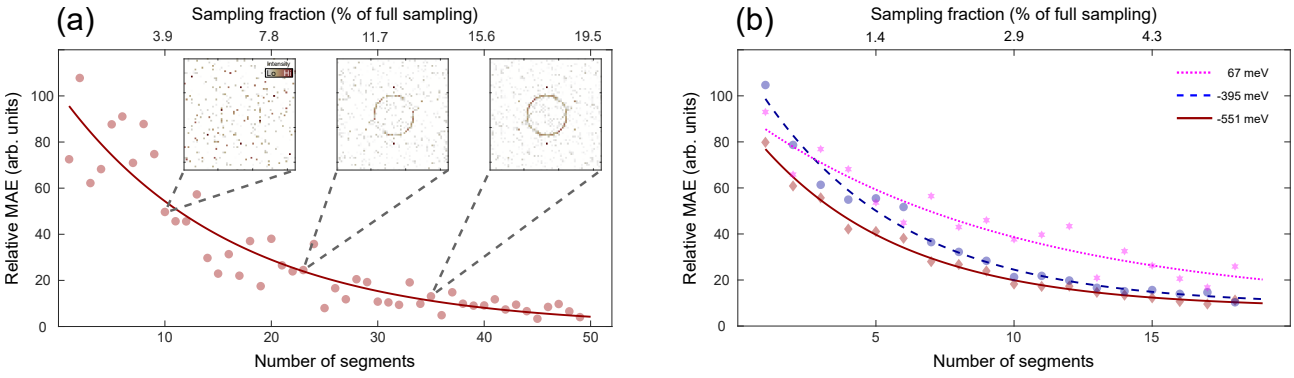

Figure 22: Relative mean absolute error (MAE) as a quantitative measure for steering adaptive sparse sampling experiments using a simulated surface state (a) and experimental data of Au(111) (b). Despite its simplicity this measure could enable an informed decision about how much more sampling would be required (e.g. approaching a plateau).

After performing the above experiment on Au(111) we also continued to study other quantum systems such as WTe_2 and $NbSe_2$ whose surfaces incorporate more short wavelength components such as atomic lattices. There we found that we obtain less reliable results when the effect of nonlinearities during the experiment becomes prominent. These encompass nonlinear thermal drift and the nonlinear response of the piezoelectric actuators.

1.5 Experimental nonlinearities

This part of the thesis is dedicated to discussing systematic issues that were encountered while implementing the sparse sampling techniques for QPI mapping experiments. After a short introduction to the main three nonlinearities that are creep, hysteresis and drift, we will show some correction approaches and ideas that have been developed.

1.5.1 Origin and consequences

The main strength of a scanning tunneling microscope (STM) is that it allows to map a material surface down to the atomic level. The required sensitivity to control the scanning tip along the surface and measure its electronic properties is provided by piezoelectric actuators (PZA). The piezoelectric effect describes the deformation of a material (crystal) whenever an external voltage is applied to that material. Vice versa a voltage is induced whenever the crystal is deformed. In an STM one establishes a very precise „steering device“ by putting different piezoelectric crystals in a specific geometric arrangement („piezo tubes“). By attaching an STM tip to this setup one is able to precisely determine the direction of the tip motion on the sub-atomic level by varying the voltage that is applied to the different piezo tubes. However, there are two main issues when applying said voltages, which reduces the scanning accuracy, namely creep and hysteresis effects [51][52].

When one applies a voltage to a piezoelectric crystal then the individual crystal domains will start to rearrange themselves (flip), which results in the effective deformation of the crystal. However, this rearrangement doesn't happen instantaneously. This process is fast in the beginning but it slows down towards the end, following a logarithmic trend [51]. The displacement of a PZA over time for a fixed applied voltage can be described by

$$L(t) = L_0 \left[1 + \gamma \log_{10}\left(\frac{t}{0.1}\right)\right] \tag{33}$$

L_0 is a constant displacement value that is the displacement of 0.1 s after applying the input voltage. γ is the creep factor determining the rate of the logarithmic slow-down. Practically, creep means that the tip doesn't directly go to its final target location $\vec{r}_f$ but it rather slows down at some point $\vec{r}_m = \vec{r}_f - \vec{\delta}$ close to $\vec{r}_f$ and then slowly moves towards $\vec{r}_f$. The farther one moves the tip, that is the larger the applied voltage, the larger the displacement vector $\vec{r}_m$ will be. This can be observed when scanning a regular topography after moving the tip over a large distance. At the edges of the topography one will observe distorted surface features as shown in Figure 23(a) where a topography of WTe_2 is shown. One clearly observes that the atomic rows are bent at the top edge of the topography where the tip started the scanning motion in a downward direction. This creep behavior is also schematically depicted in the inset of Figure 23(a) where a logarithmic PZA creep displacement over time is shown.

Going back to Equation 33, it is also known that L_0 contains a hysteresis property when applying a voltage loop, which is the second nonlinear effect when moving an STM tip. In Figure 23(b) we show such a hysteresis loop where the displacement at a specific voltage changes after one full voltage loop. Practically this means that if one applies a bias voltage V to a PZA followed by the same negative voltage $-V$ then the tip won't be at the same position as it was before applying the initial voltage. In reality it is difficult to distinguish between creep and hysteresis since most of time these effects are intermixed.

The last important nonlinear effect has already been mentioned in Section 1.3.7. Due to local thermal gradients within the microscope the sample will microscopically drift in a direction depending on these gradients. On a very small scale this motion is most likely nonlinear because of thermal fluctuations leading to varying local thermal gradients. In Figure 23(c) we show two topographic scans of Au(111) that have been acquired at liquid nitrogen temperatures of 77 K. The second topography (labelled „After“) has been obtained 5.5 hours after the first one (labelled „Before“). During this time the sample has drifted along a nonlinear trajectory as depicted in the central panel. In between these

topographies there were many more topographies recorded, which allowed to recreate said trajectory. More details will be given in Section 1.5.3.

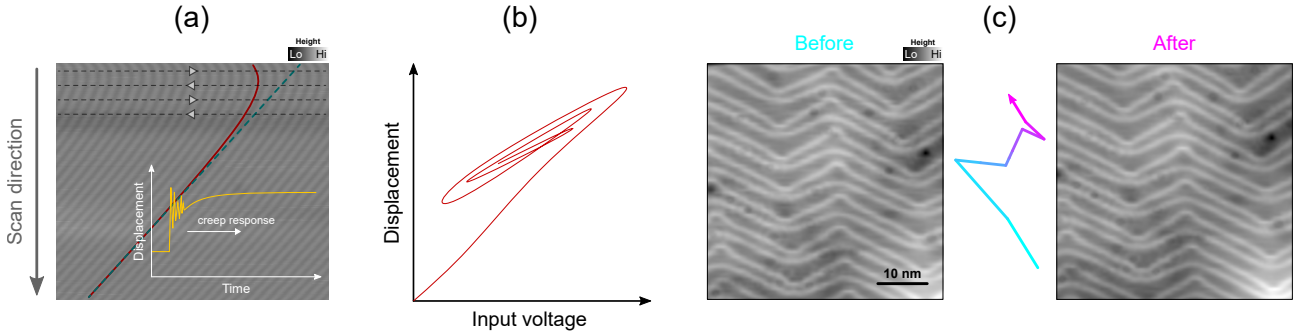

Figure 23: Effects of nonlinearities in a scanning tunneling microscope (STM). (a) Piezoelectric creep can be seen at the edges of topographic scans after moving the tip over a large distance. Here we use the atomic rows of WTe_2 for guidance. The lower right inset shows the logarithmic displacement of a piezoelectric actuator (PZA) with time. Figure adapted from [51]. (b) Piezoelectric hysteresis curve. After applying a voltage loop the tip won't reach its starting position anymore. Figure adapted from [52]. (c) Thermal drift illustrated using two topographies of Au(111) that have been recorded 5.5 hours after each other. Due to local thermal gradients the sample will generally drift in a random trajectory over time (middle panel).

All of the described nonlinear effects might significantly reduce the quality of the reconstructed output using sparse sampling approaches due to the nature of the used near-optimal tip path and the crucial relative correlation between the locations and the measurement values. If one, due to the presence of mentioned nonlinear effects, assumes wrong measurement coordinates then this will directly influence and most likely worsen the output of the sparse recovery. For example, when we have a large creep then the assumed position $\vec{r}_f$, that we would like to visit, will deviate from the actual position $\vec{r}_m$ where the measurement will happen. Furthermore, when we have a long spectral acquisition time then the bias voltage sweep might happen along multiple locations on the surface and thus lead to spectral smearing. For WTe_2 and other quantum materials that show prominent atomic corrugation we have seen that the sub-division of the complete set of measurements locations into smaller tiles will lead to a tiling artifact. The main reason is thermal drift of the sample (see Figure 17) leading to an effective phase-shift of neighboring tiles that haven't been measured successively. In our experiments using the adaptive sparse sampling approach we don't expect to encounter this problem as the individual paths are distributed over the complete scan frame and are not locally constrained. Nevertheless, a strong drift might lead to destructive interference of the very short-range real space modulations such as the Bragg peaks. Piezoelectric effects of creep and hysteresis are also lowering the reconstruction quality in both cases. From experience of successful and failed results we conclude that both creep and hysteresis are less harmful to those scattered electron states that are long-range in real space, i.e. close to the zone center in q-space. This is the reason we think that our model system Au(111) did work so well as most of the scattering wave vectors are close to the zone center. Meanwhile, materials that inhibit very short-range wavelengths will be influenced more by creep and hysteresis but also by thermal drift.

1.5.2 Correction approach for piezoelectric creep

A strong piezoelectric creep as shown in Figure 23(a) preferably occurs after applying a large voltage to the PZAs resulting in a large lateral displacement of the tip. For smaller travel distances the „creeped" distance becomes smaller but it still might have a strong influence on the sampled measurements and their locations as mentioned previously. One of the easiest solutions to avoid creep is simply to wait some time after traversing large distances to settle the creep or not applying too large voltages that is only moving small distances at a time. Since there will be always some residual creep (due to its logarithmic nature) one could correct it by fitting the creep function as described in Equation

33. In Figure 24 we show two different measurements of the tunneling current on WTe_2 that have been obtained with deactivated feed-back loop (constant height) as a function of time. We find that WTe_2 is suited very well for creep studies as it contains periodic and prominent surface features that can be tracked easily. Figure 24(a) shows the creep behavior after moving 50 nm in one direction (45 degrees across the atomic rows) and a tip speed of of 15 nm/s.[11] After a few seconds the tip starts moving and the current follows the shape of the atomic rows as illustrated in the zoomed inset. Shortly after, the initial tip motion is stopped and the current behavior shows a long creep distance traversing another atomic row. The red curve shows the fitted creep function described by Equation 33 that furthermore includes the sinusoidal motion due to the atomic corrugation. The fit is obtained using a least squares approach. The fitted creep parameters are $L_0 = 0.94$ nm and $\gamma = 3.73$. In Figure 24(b) we show another measurement of the current trace performed after moving the tip for 25 nm at the same tip speed. We immediately see that the creep settles much quicker than in Figure 24(a) where it hasn't actually settled even 100 seconds after stopping its initial motion. The fitted creep parameters are $L_0 = 0.83$ nm and $\gamma = 3.96$. By being able to fit the parameters that describe the creep function one would be able to correct the true locations of the spectra in a post-processing manner. By having access to the time stamps, one should furthermore be able to precisely determine the sample location where the bias sweep was initialized. Thus, the spatial relationship between the individual measurements can be restored in this way.

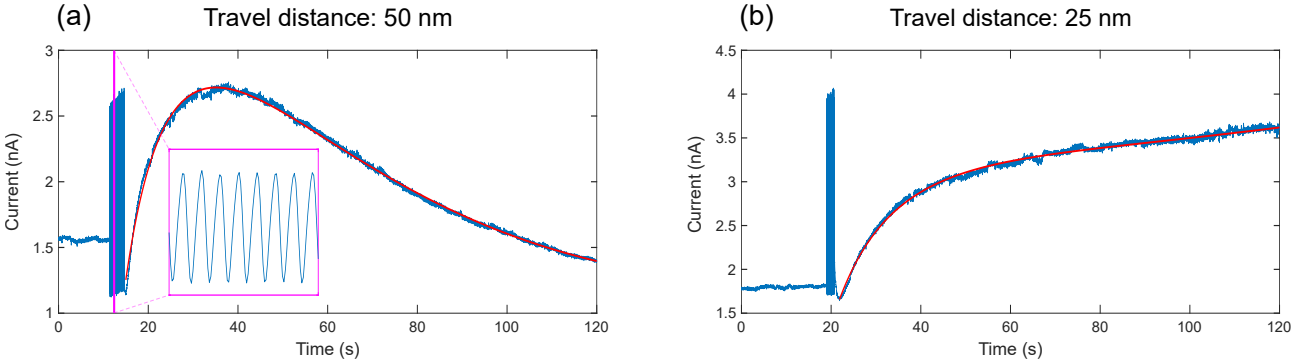

Figure 24: Fitted creep functions illustrating the influence of the tip travel distance on the creeped displacement on WTe_2. A longer travel distance means a larger applied voltage to the PZAs. (a) Travel distance of 50 nm results in a very long creep motion that persists over a long period of time. The inset shows the motion of the tip across the periodic atomic rows. (b) Travel distance of 25 nm shows a much shorter creep motion and comes close to a settling point at the end of the measurement period. In both (a) and (c) we show the fitted creep function as a red line.

1.5.3 Correction approach for nonlinear thermal drift

In our previous work [25] we described a correction approach for linear thermal drift in an STM environment. This assumption might however not be true if the measurement time is on the order of hours as indicated in Figure 23(c). The adaptive sparse sampling approach [50] was partially motivated by this fact. Having many interleaved small-scale topographic scans before and after TSP LDOS maps the deviation from a truly linear drift assumption should become smaller. In this case one should be limited only by the capability of precisely determine the drift vector from consecutive topographies. Let us assume that two topographies I_j and I_{j+1} have been obtained and where some time t has passed in between their respective acquisition. In this case we can determine the relative shift $\vec{v}_{j,j+1}$ (drift vector) between I_j and I_{j+1} using 2D cross-correlation via the 2D fast Fourier transform [53]. Both knowledge of said drift vector and the effective drift trajectory is required to successfully correct the measurement locations. In Figure 25 we show simulations of both linear and nonlinear drift and how deviations from the true parameters will affect the reconstruction quality. Figure 25(a) shows

[11] The tip speed doesn't have an influence on the creep function. A larger tip speed results in a shorter time during which the piezoelectric crystal domains can flip and thus the consequence is a longer creep time but not a loner creep distance.

a simulated Shockley surface state with atomic corrugation as it was used in Figures 9 and 18 on a 256×256 pixel grid. The top panel shows the ground truth QPI pattern and corresponding LDOS modulations. The lower two panels show the reconstruction of the ground truth using a small number of randomly selected LDOS measurements and how a deviation from the true drift parameters will lead to a smearing of the reconstruction. While the circular surface state is more robust to drift due to its overall longer real space modulations, the Bragg peaks vanish already at small deviations. Figures 25(b)-(d) show the result of three simulations for different assumed scenarios such as linear, nonlinear and improperly measured drift vectors. Hereby we compare the ground truth QPI pattern in the top panel of Figure 25(a) with the reconstruction of the named situations using multiscale structural similarity (MS-SSIM) [54]. A higher MS-SSIM value means that the reconstructed output looks more similar to the ground truth (1 means identical). In Figure 25(d) We observe that the obtained phase space (dark-blue region) becomes very small as soon as nonlinear drift is coupled with an imprecise drift vector determination. Figure 25 generally emphasizes the value of operating in the linear drift regime, which means that one should either minimize the overall measurement time or have more interleaved topographies.

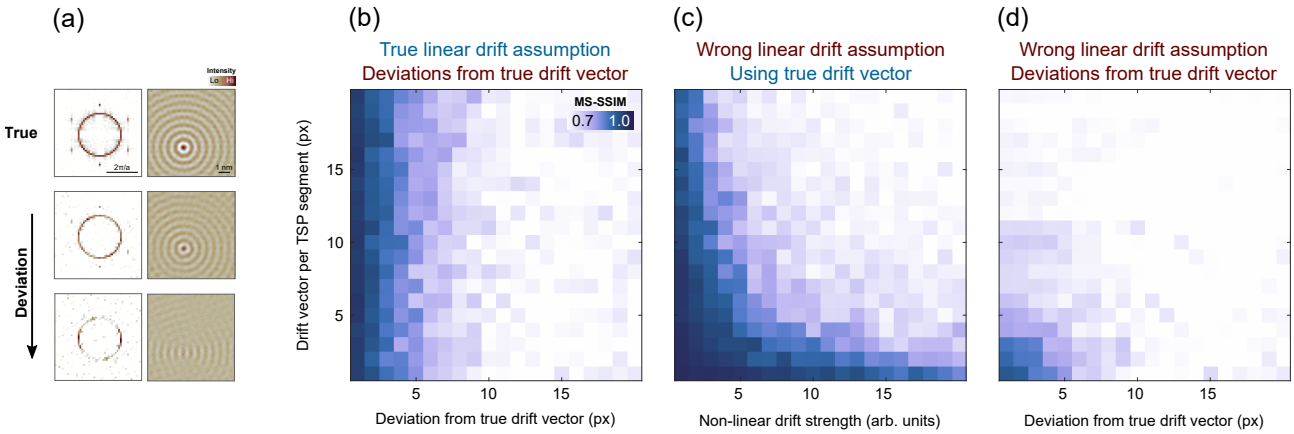

Figure 25: Simulated results illustrating the limits of linear drift correction due to deviations from the true drift vector or a nonlinear drift scenario. (a) Simulated ground truth QPI pattern and LDOS modulations (top). The higher the deviation from the true drift parameters the more smeared the reconstruction becomes (here obtained by randomly sampling 3×1'000 points). (b) Averaged simulation over five runs assuming linear drift but improperly measuring the drift vector between each individual tip path segment. The evaluation measure used is multiscale structural similarity (MS-SSIM) between the ground truth and the reconstructed QPI pattern. (c) Measuring the correct drift vector but encountering nonlinear drift between the individual tip path segments, which is modelled using a quadratic Bézier curve. A stronger nonlinear drift results in a more rapid diminishing reconstruction quality. (d) Combination of both scenarios (b) and (c) that is imprecisely determine the drift vector and encountering nonlinear drift. Even small deviations in the drift vector determination will lead to very fast worsening of the reconstruction quality. This situation most likely resembles real experimental conditions.

The idea of a linear drift correction procedure has further been tested on the 2D transition metal dichalchogenide $NbSe_2$ exhibiting high-momentum states in the form of atomic Bragg peaks. In Figure 26 we show that the deviation from the linear drift assumption will worsen the attribution of said states. The size of the used scan frame is 50×50 nm and the experiment was performed at 4.2 K. On a 512×512 pixel grid we measure a full $\mathrm{d}I/\mathrm{d}V$ spectrum during 484 ms at 2×3000 locations using conventional bias spectroscopy. For every point we measure ten single spectra and take the last one for the reconstruction. During the acquisition of the first nine spectra, we thus hope to settle creep and therefore remove one nonlinear effect. Additionally, we measure a time stamp with pico-second precision for every measurement point using a time-tagger (Swabian instruments). Figure 26(a) shows the sparse reconstruction when we do not post-correct the measurement locations. The only feature visible in the QPI pattern are the atomic Bragg peaks forming a hexagonal pattern. In Figure 26(b)

the reconstruction at the same energy is shown but here we corrected the measurement locations while assuming a linear drift. For this we utilize the above mentioned time stamp information for every measurement point. The drift vectors are obtained by cross-correlating interleaved topographic scans as described previously (here we use a full $\mathrm{d}I/\mathrm{d}V$ map instead of topographies because of the software-related issues as described earlier in this thesis). These scans have a size of 50×56 pixels and are shown in Figure 26(c) where the top row is the raw data and the bottom row has been low-pass filtered to have a better cross-correlation outcome. On the top we state the time that has passed in-between the individual scans. The respective drift vectors are drawn as red arrows in the bottom row clearing showing a nonlinear trend. We observe that the sharpness of the Bragg peaks is not improved after applying the linear drift correction scheme, which can be seen as well using the dispersion plot in Figure 26(d). We attribute the absence of dispersing features more to the very low sampling rate rather than to nonlinear thermal drift whereas the smearing of the Bragg peaks after the correction indicates that the linear drift assumption does not apply in this case due to the hourlong delay between the two individual tip path segments.

Finally, one could think about performing a post-processing correction that would account for all mentioned nonlinear effects [55] by calculating a local displacement field, that is a deviation from perfect atomic structures. This could also be performed in an optimization procedure that would aim at sharpening the atomic Bragg peaks in reciprocal space. The resulting geometrical transformation can then be inverted to obtain new effective coordinates of the measurement points in real space. Performing the sparse reconstruction again with these new positions should yield more sparse coefficients in q-space as now their coherence has been re-established. An ultimate test would be to apply such correction onto materials exhibiting short range periodicities such as WTe_2 to remove the phase-shift in between individual tiles.

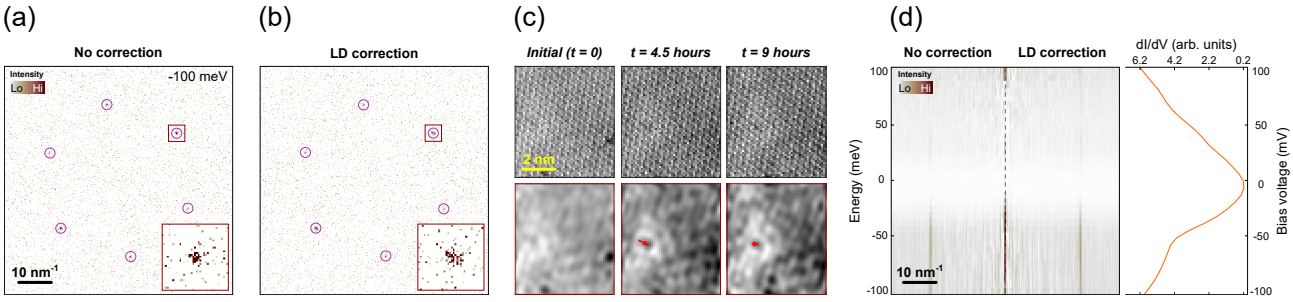

Figure 26: Adaptive sparse applied to $NbSe_2$ illustrating the limit of linear drift correction. (a) QPI pattern at -100 meV using two tip paths containing 3'000 measurements each. No post-processing correction has been applied to the measurement locations. (b) Same as (a) but with linear drift correction applied where we use the drift vectors obtained from the topographic references in (c) as well as time stamp information. (c) Topographic raw scans (top) and their low-pass filtered counterpart (bottom). The drift vectors are drawn as red arrows and the time stamp information is given for every of the three maps. (d) Energy dispersion relation of $NbSe_2$ obtained via an azimuthal average showing just the main Bragg peaks but no actual dispersive features. The latter is attributed to the too low sampling rate that would ensure their successful reconstruction. On the very right the average spectrum of $NbSe_2$ is shown for reference.

1.6 Outlook

1.6.1 Implementation of a different sparsifying basis

In this work we focused on the sparse reconstruction of QPI patterns under the assumption that these have a sparse representation in the Fourier domain. This allowed for a relatively fast recovery process due to the computational efficiency of the fast Fourier transform algorithm. However, it is known that the local density of states (LDOS) is actually proportional to a Bessel function of the first kind [2] (decaying sine wave)

$$\mathrm{LDOS}(\vec{r}, E) \propto 1 - J_0(2kx) \tag{34}$$

where k is the wave vector dictated by the dispersion relation $E(k)$ (depicted in Figure 27(a)). Performing a Fourier transform of a Bessel function actually results in many non-zero coefficients, although most of them are quite small as can be seen in Figure 27(b). A better sparsifying basis to describe a Bessel function of the first kind would be the Hankel domain, which can be seen in Figure 27(c). Therefore it would make sense to implement a new Hankel transform operator (instead of a Fourier operator) that could allow to reconstruct the QPI patterns with even lower samples needed since the signal is more sparse in the Hankel basis. During brief tests we found that calculating the direct Hankel transform for large grids is infeasible as it takes a long computation time compared to the fast Fourier transform. Thus one should aim at implementing a fast Hankel transform if one would like to make full use of efficient sparse recovery algorithms.

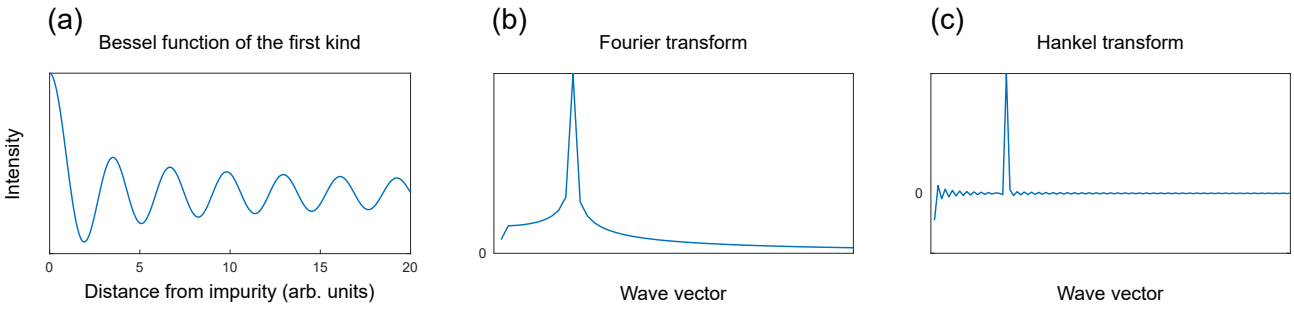

Figure 27: (a) The local density of states (LDOS) can be represented as a Bessel function of the first kind. (b) Fourier transform of (a) showing a large tail of non-zero values. (c) Hankel transform of (b) using [56] offering a much sparser representation of the LDOS than the Fourier transform in (b).

1.6.2 Symmetry operations

Researchers nowadays often apply a posteriori symmetrization to the conventionally acquired QPI data, that is averaging the QPI patterns with their rotated and/or reflected copies. However, this method can introduce artifacts in the form of erroneously emerging symmetries in the QPI patterns [57]. In our previous work [25] we proposed future developments regarding the usage of symmetry operations prior to the actual reconstruction process. Depending on the studied systems different point symmetry operations could be applied to the small set of measured data points. As a consequence the measurement matrix would grow and the overall number of required samples for a successful reconstruction could thus be reduced. In Figures 28(a)-(c) we show the potential of such a symmetrization procedure using a highly symmetric kaleidoscope image, representing a single „unit cell“Ṫhe reconstruction outcome of applying said symmetry operations is hereby strongly dependent on the ability of precisely determining the unit cell locations of the studied sample. This requires topographies of high quality or other kinds of scans where the individual unit cells can be distinguished from each other. In Figure 28(d) we show a simulated hexagonal Si(111) environment (provided by the Nanonis SPM controller software) and a single unit cell that is marked with red lines. Taking random measurements (black dots) one could identify such unit cells and perform mirror and/or rotation operations[12] resulting in more effective measurements (white dots).

[12]Depending on the symmetry of the unit cell.

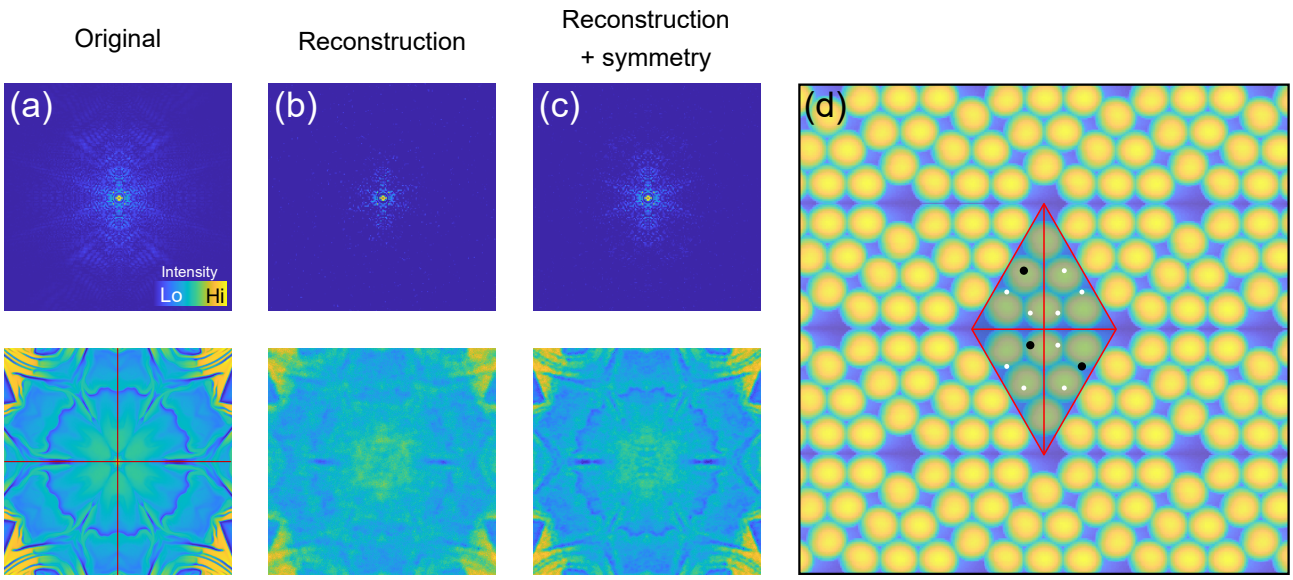

Figure 28: Symmetrization of the acquired data with respect to the individual unit cells prior to the sparse recovery can lead to a big improvement of the reconstruction quality. (a) Fourier transform (top) and corresponding real space image (kaleidoscope) (bottom). The image contains two mirror planes depicted by red lines. (b) Sampling 2% of the complete image in (a) results in a poor reconstruction where many details vanish in the noise. (c) Taking the same measurements as in (b) but performing a data symmetrization prior to the sparse recovery results in much better reconstruction. (d) Simulated topographic surface of Si(111) showing a hexagonal atomic lattice. A single unit cell is marked with red lines containing two mirror planes. The symmetry operations are applied onto the measurements that fall within that unit cell (black dots) resulting in an increase of the total number of measurements (white dots).

1.6.3 Time-tagging for removing coherent noise

In Section 1.5 we talked about how time-tagging the individual measurements can help to more precisely determine the true locations where the measurements have been taken. This would in turn lead to a more precise post-processing correction related to the systematic non-linearities of piezoelectric creep and thermal drift. A further use of such time stamps is related to the removal of coherent noise. If the system is subject to e.g. mechanical (vibrational) noise then one will coherently sample this noise when performing a conventional line-by-line scanning motion. The problem hereby is that every time when the tip changes from one line to the next it requires some amount of time, which effectively leads to a loss of the phase coherence of said noise signal. This will then show itself as rather strong artifacts in the QPI pattern. If one however has access to the time stamps of the measurements then one could restore the lost phase coherence. By sampling only full periods $T_i = n/f_i$ of the noisy frequencies f_i one should be able to sharpen the frequencies f_i. This would allow their removal without the need of filters that also potentially could influence the underlying QPI information.

1.7 Summary and conclusion

In this thesis we have extended upon previous work in the field of sparse sampling [25] focusing on the experimental implementation of the developed method, which allows to significantly reduce the mapping time of quasiparticle interference (QPI) imaging experiments. We realized its verification using various 2D materials such as the noble metal Au(111), the high temperature superconductor Bi2212 and the type-II Weyl semimetal WTe_2. Because of the strongly reduced mapping of up to a factor 30 we were able to utilize large scan frame sizes and pixel grids leading to a high q-space resolution. The randomness of the measurement locations furthermore results in an effective reduction of coherent noise that might be present in the system due to the good denoising properties of the used linear sparse solver. However, we saw that the initial approach of the near-optimal tip motion led to tiling artifacts in the reconstruction because of thermal drift. Since the original sparse sampling method also required the knowledge of the exact sampling rate prior to the experiment we developed a new sparse sampling variant that allows to cumulatively stack measurements obtained from smaller tip paths. This method was verified on Au(111) and allows to track the reconstruction quality at runtime, which should make sparse sampling more accessible in the field of QPI mapping. Additionally, combining sparse sampling with faster parallel point spectroscopy [33] opens the doors for whole new range of complex experiments that would otherwise be very cumbersome to realize.

For future experiments the nonlinear effects and their treatment should be further investigated. This includes robust post-postprocessing corrections for piezoelectric creep and thermal drift. For creep one could fit the underlying creep function for various travel distances and potentially for different angular directions. For thermal drift one should try and automatize the acquisition of regular topographic scans around distinct surface features to situate themselves in the linear drift regime. For the adaptive sparse sampling approach one could think about implementing a decision-making process that would determine after how many path segments the experiment should be stopped. This could be simply achieved by tracking some characteristic measure as shown in Figure 22 or more sophisticated procedures such as machine learning. For example, one could think about training a neural network on a multitude of different QPI patterns such that it would learn to distinguish the sparse scattering wave vectors from experimental noise during the reconstruction process.

2 Deep learning based denoising of X-ray diffraction data

2.1 Introduction

Since the beginning of the digitalization of our society the amount of data acquired has grown exponentially and with that there has been a surge of new computational techniques to enable more efficient data handling. Nowadays, humans are no longer able to control this operation directly by themselves. However, the field of deep learning has proven as a viable tool for studying, interpreting and processing vast amounts of data. One field where deep learning did make major contributions is in the field of digital image processing encompassing tasks such as super-resolution [58], restoration [59] and denoising [60][61][62]. In many of these tasks convolutional neural networks (CNN) are implemented due to their excellent performance in various machine learning problems [63]. A CNN consists of different convolutional layers, which in turn contain a collection of neurons that are defined by unique weights and biases. This allows different layers to extract and learn specific patterns of an image. A convolution is applied by element-wise multiplying a moving kernel of specific size with the underlying values of a given input image. Afterwards a nonlinear activation function is applied onto each neuron resulting in the efficient and general learning behaviour of CNNs. This process depicted in Figure 29.

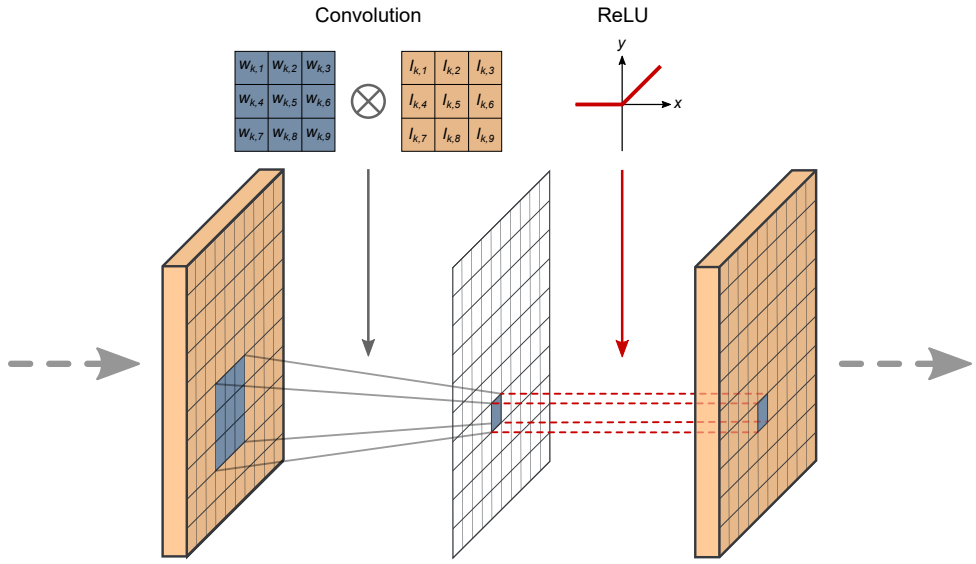

Figure 29: A simplified snapshot of a convolutional neural network (CNN). A kernel of specific size and with weights $w_{k,i}$ (blue) moves along an image I and is convoluted with the underlying image values $I_{k,i}$ (orange). k hereby refers to the convolutional layer index. The result of this convolution is a single number onto which a non-linear activation function such as a rectifying linear unit (ReLU) is applied, which enables the network to learn nonlinear features of the input image. A deep CNN consists of many of such depicted layers.

Here we focus ourselves on the task of image denoising in the field of X-ray diffraction. Generally, the task of image denoising is to remove noise from a noisy image $x_i = s_i + n_i$ where pixel i consists of the intrinsic signal s_i and the noise n_i. Both supervised and unsupervised learning methodologies have been successfully applied on denoising problems. Supervised learning approaches [60][62] rely on image pairs such as the noisy and the clean image („noise-2-truth“) or two different noisy images („noise-2-noise“) [64]. Unsupervised approaches on the other hand do not require pairs of images but rather learn to extract the important features from the noisy image itself such as „noise-2-void“[65] or „noise-2-self“[66]. However, the unsupervised methodologies typically show slightly inferior performance compared to supervised algorithms as less information is available during training. One of the most studied benchmark problems is where Gaussian noise is artificially added to natural photographs. A neural network is then trained to remove said noise. The signal is hereby correlated between neighboring pixels whereas the noise is not. Much less attention has been given to Poisson distributed noise

[67] even though Poisson distributions occur frequently such as in virus-cell infection, radioactivity and particle scattering events. In the noise is signal dependent such that both signal and noise are correlated between neighboring pixels. Normally, the signal-to-noise ratio (SNR) can be increased by using a sufficiently long count time. This solution is however not always feasible. In cryo transmission electron microscopes (TEM) experiments on organic materials it is known that electron beam damage scales with exposure time [68]. Another example where one has strongly limited counting times are diffraction experiments using pulsed magnetic fields [69][70]. The result of low counting times is the presence of a large amount of noise that can easily overshadow the underlying intrinsic signal one is interested in.

One common challenge for many machine learning problems is the scarce availability of training data, which oftentimes results in overfitting of the models [71]. A common approach to many studies is thus to artificially add additive noise (Gaussian, Poisson, Salt, Pepper, etc.) to a known ground-truth signal [72] that is many times a natural image. But also for counting experiments this approach is used [67]. In this part of the thesis, we instead focus on true experimental noise that we use together with a ground truth signal to train two different deep CNNs. The task is to remove the experimental noise from a low counting frame and thus produce a prediction similar to the true high counting frame. Additionally, we will compare the CNN performances on artificial noise.

2.2 Training data

2.2.1 Experimental X-ray diffraction data

The data for this work has been obtained from X-ray diffraction on the high temperature cuprate superconductor $La_{1.88}Sr_{0.12}CuO_4$ [73]. The data has been obtained at the P021 triple axis diffractometer at the PETRA III synchrotron (DESY - Hamburg). The scattering intensities were recorded using a Dectris Pilatus 100K CdTe detector, which offers frames of size 195×487 pixels with a bit-depth of 32 per pixel. The diffractometer was operated with 100 keV photons while the sample was cooled to around 30 K where a charge density wave (CDW) order is fully developed [74][75]. Some of the diffraction data thus contains a weak 2D CDW signal in addition to the fundamental Bragg peaks and powder diffraction lines from the polycrystalline cryostat sample environment. Additionally, spurious and dead pixels are present in the data. In such a counting experiment all the experimentally observed signals are subjected to some sort of Poisson noise. To train a neural network two different kinds of counting statistics data have been acquired, namely low and high counting statistics data, following the „noise-2-truth“ denoising approach. The former has been counted for 2 seconds while the latter has been counted for 21 seconds. We label the experimental pairs of training data $(x_j^{\mathrm{LC}}, x_j^{\mathrm{HC}})_{\mathrm{exp.}}$ for each frame j. These pairs have been measured successively while keeping all other measurement parameters fixed. In Figure 30 we show some examples of both low and high counting statistics X-ray diffraction data on $La_{1.88}Sr_{0.12}CuO_4$. In total 3'567 experimental data pairs have been recorded.

2.2.2 Artificial noise generation

To compliment the measured experimental data, we artificially create low counting statistics frames by adding either Poisson or white Gaussian noise to the corresponding experimental high counting statistics frame. It is known that detecting individual photons can be treated as independent events, which follow a random temporal distribution and can therefore be seen as a Poisson process [76]. Per pixel (sensor), the number of photons N that is measured in some time interval t is hereby given by the discrete Poisson probability distribution

$$\Pr(N = k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!} \tag{35}$$

where λ is the expected number of photons per unit time interval. Therefore, λt is equal to the expected number of photons that hit the detector after some time t. According to the central limit theorem the Poisson distribution will approach a Gaussian distribution in the limit of $N \to \infty$. A Poisson distribution furthermore has the property that its expectation $\mathrm{E}[N]$ is equal to its variance

VAR[N]. This means that Poisson noise is signal dependent and its standard deviation, that is the signal-to-noise ratio, is given by

$$\sigma = \sqrt{\mathrm{VAR[N]}} = \sqrt{\mathrm{E}[N]} = \sqrt{\lambda t} = \sqrt{N} \tag{36}$$

For a fair comparison of the CNN performance on the experimental and the artificial data we would like the datasets to be statistically similar. To achieve this, we define a quantity γ_j as the ratio between the frame integrated low count $N^{\mathrm{LC}}_{j,\mathrm{exp.}}$ and high count $N^{\mathrm{HC}}_{j,\mathrm{exp.}}$. From all the available frames we then define the median value of γ_j over all pairs of experimental training data. This quantity is labelled γ. We then multiply γ with every individual high count frame $x^{\mathrm{HC}}_{j,\mathrm{exp.}}$ and generate an artificial low count frame by adding the associated Poisson noise

$$x^{\mathrm{LC}}_{i,j,\mathrm{art.}} = \Pr(N = \gamma x^{\mathrm{HC}}_{i,j,\mathrm{exp.}}) \quad \text{for every pixel } i \text{ of frame } j \tag{37}$$

For the case of white Gaussian noise, we randomly draw samples from a normal distribution with standard deviation σ and add these samples to the experimental high count data for every frame. σ is hereby defined as the median value of all standard deviations σ_j of all the experimental low count frames $x^{\mathrm{LC}}_{j,\mathrm{exp.}}$. To mimic the experimental noise distribution, we perform a convolution of the artificial noised data using a Gaussian kernel [67] with a radius of two pixels and a standard deviation randomly chosen between 0.3 and 0.5.

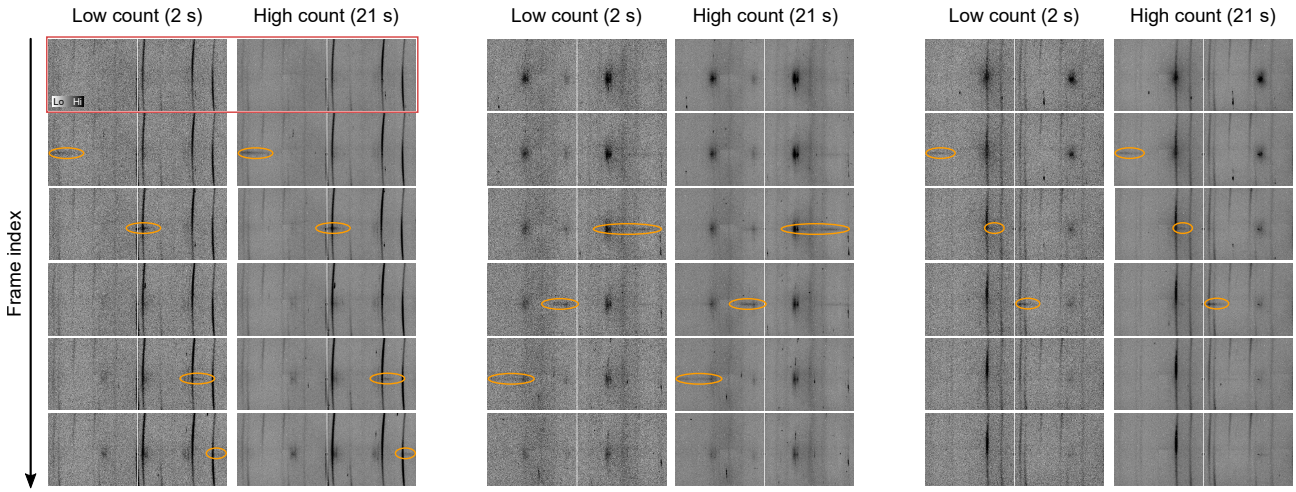

Figure 30: Experimental X-ray diffraction data on $La_{1.88}Sr_{0.12}CuO_4$ at 30 K. Three different examples are shown with different sample orientations (left, middle and right panel). In some frames a charge density wave (CDW) signal can be observed how it translates through the field of view (circled in orange). In the low counting frames these faint signals vanish quickly in the noisy background. Furthermore, vertical powder diffraction lines and circular Bragg peaks are the main features contributing to the overall intensity. A pair of low and high counting frames is marked in with a red rectangle on the top left. The deep CNN is trained on such data pairs.

2.3 Deep neutral network architectures and training process

In this work we implemented two different CNN architectures referred to as VDSR [77] and IRUNet [78] using the tensorflow/keras API implemented in Python. While the VDSR network is a straightforward deep neural network, which only learns the residuals, the IRUNet network is additionally built-upon an autoencoder framework. As described previously we use training pairs containing low and high counting statistics data where the CNNs try to remove the noise from the low count frame to match the ground truth high count frame. Prior to feeding the data into the network pipeline we split all the original frames approximately in half that is from 195×487 to 194×242 pixels to increase the amount of training pairs and allow for a higher batch size to be used during training. The complete dataset is furthermore split into a training (70%), validation (20%) and test set (10%). The training set is used for training the network while the validation set is used as measure for how to adjust the

network weights and biases. The test set is later used to check the model performance. A further pre-processing step includes normalizing each frame to have values between 0 and 1. A schematic illustration of the working principle of the denoising networks is shown in Figure 31. By training the network on many different pairs of low and high counting statistics data it learns the intrinsic features contained within the high count data and is then able produces a denoised output from a noisy low count input. We stress the charge density wave (CDW) signal appearing in the denoised output of Figure 31, even though it is barely visible in the low count input.

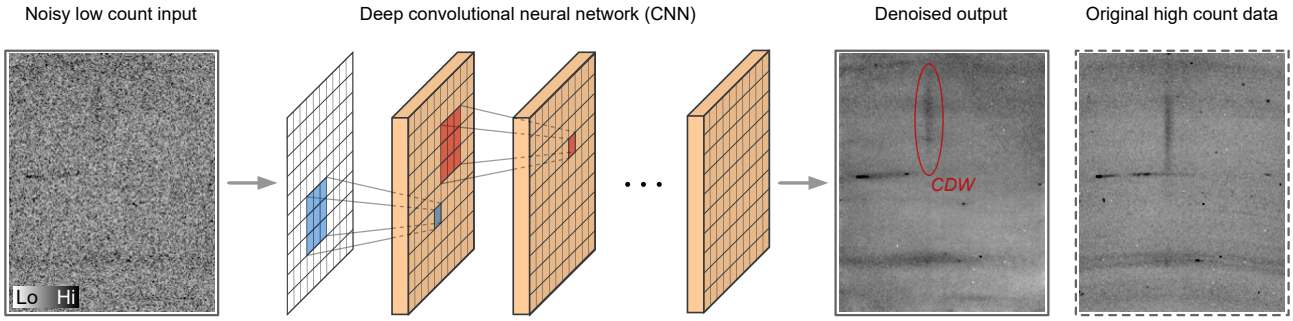

Figure 31: Illustration of the denoising abilities of the trained deep convolutional neural network (CNN) for X-ray diffraction experiments on a $La_{1.88}Sr_{0.12}CuO_4$ single crystal. A real experimental low counting statistic frame (exposure time 2 seconds), belonging to the test dataset, is used as an input to a deep CNN, which is trained to remove the noise. The output shows the denoised low counting input and reveals a charge density wave (marked in red), barely visible in the raw noisy input data. To the very right the real experimental high count frame (exposure time 21 seconds) is shown for comparison.

The named networks both use 20 convolutional layers with 64 filters each and a kernel size of 3×3. The optimization during training is achieved by an Adam optimizer [79] with the AMSGrad variant [80] to stabilize training and convergence. The Adam optimizer is an extension to the stochastic gradient descent technique. Both networks are trained with an initial learning rate of 5×10^{-4} but a different learning rate scheduler to achieve good convergence. For the VDSR model we multiply the initial learning rate by 0.1 after the first 150 epochs and then again after every 50 epochs. On the other hand, we multiply the initial learning rate by 0.5 after every 100 epochs for the IRUNet model. The VDSR model is trained for 250 epochs using a batch size of 8 while the IRUNet model is trained for 300 epochs using a batch size of 16. During each training epoch the performance of the neural networks is evaluated on the validation set by comparing the denoised output x_j^{D} with the actual high counting statistics data x_j^{HC}. The loss function used is a combination of mean absolute error (MAE) and multiscale structural similarity (MS-SSIM) [54]

$$L = (1-\alpha)L_{\mathrm{MAE}} + \alpha L_{\text{MS-SSIM}} \quad \text{with} \quad \alpha = 0.7 \tag{38}$$

as described by [67][81]. During training we apply data augmentation in the form of randomly flipping the frames vertically. The training has been performed on a Nvidia Tesla P100 GPU with 10 GB of VRAM. The training duration was 20 and 12 hours for the VDSR and IRUNet model respectively.

2.4 Denoising results and discussion

As mentioned in Section 2.2.2 we also trained the different deep CNNs on artificial data. This should allow to establish a connection to the usual approach in AI-assisted denoising studies. In the following we will compare the model performances in terms of architectures and training methodology. In Figure 32(a) we compare the statistics of the different noise types in terms of their respective probability density distributions for a single data frame. The true experimental high count shows a strongly asymmetric probability distribution due to the large intensity variation across the detector pixels. We observe that the true experimental low count probability distributions are best described by a skewed Gaussian distribution instead of a pure Poisson distribution. We also see that the convolution

(⊗) of the pure artificial Poisson noise with a Gaussian kernel (G) indeed helps to mimic the true experimental noise profile.

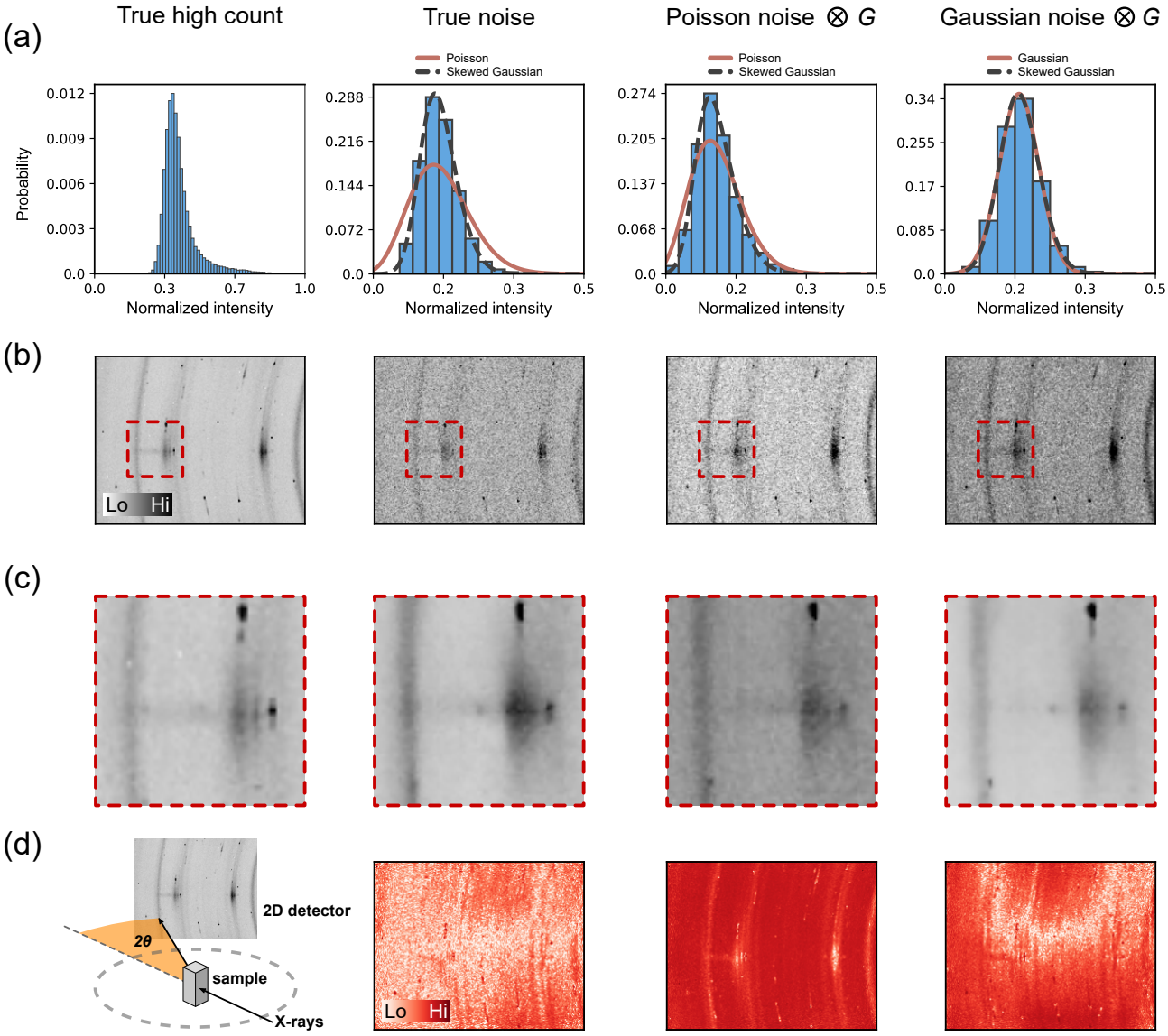

Figure 32: Comparison of different noise statistics. (a) Probability density distributions of the X-ray diffraction frames in (b) with fitted Poisson or (skewed) Gaussian distributions for the noisy situations. We observe that the true experimental noise doesn't follow a pure Poisson distribution but is better described by a skewed Gaussian distribution, imitated by a convolution (⊗) of the pure Poisson noise with a Gaussian kernel (G). (b) Same frame under the influence of different noise statistics. The true experimental high count is shown to the very left for reference. The rectangular region marked in red is situated over a 2D charge density wave (CDW) signal. (c) First panel shows the true experimental high count zoom of the marked region in (b) while the later panels visualize the high count versions of the marked regions in (b) as predicted by the IRUNet neural network trained on the respective noise types. As an input the true noise was used. (d) Experimental setup depicted to the very left while the remaining panels show the heatmaps (absolute relative difference) between the predicted outputs of the individual noise types and the true high counting statistics data frame (Equation 39) on a logarithmic scale. We clearly observe that the model trained on the true noise performs better compared to the models trained on the artificial noise.

The underlying image data to the probability density distributions in Figure 32(a) are shown in Figure 32(b). In Figure 32(c) we show the respective predicted high count statistic version of the rectangular region marked in red in Figure 32(b) using the trained IRUNet model, which has been trained on the individual noise types but the true noise is used as an input. We find that the output in all

cases visually displays the main features of the actual high count data such as the powder diffraction lines whereas the CDW is not really resolved anymore. Note that the first panel in Figure 32(c) simply shows the rectangular marked region of the high count frame in Figure 32(b). To have a better intuition about the structural differences between the predicted high count outputs x^{D} and the true high count x^{HC} we show the corresponding heatmaps y in Figure 32(d) calculated as

$$y = \frac{x^{\mathrm{HC}} - x^{\mathrm{D}}}{x^{\mathrm{HC}} + x^{\mathrm{D}}} \tag{39}$$

A logarithmic colorscale is used to enhance the contrast. We observe that the model trained on the true noise is able to denoise most of the regions within the low count frame while the model trained on the artificial Poisson noise mainly works well for regions with high intensity features such as the powder diffraction lines. The model trained on the artificial Gaussian noise appears to only work in certain regions of the image probably because the noise is not correlated with the signal as it is the case for Poisson and true experimental noise. Therefore the model trained on Gaussian noise has difficulties in distinguishing the noise from the signal when they are correlated. The first panel in Figure 32(d) shows a schematic depiction of the X-ray diffraction experimental setup used to acquire the complete dataset. To quantify the performance of the two denoising CNNs, applied on different noise statistics, we evaluate the trained models on the separate test set. The results are described quantitatively in Table 1 and qualitatively in Figure 33. As a performance measure we use the peak signal-to-noise ratio (PSNR), multiscale structural similarity (MS-SSIM) and the measure that was used during the training process described by Equation 38 (MAE+MS-SSIM). Based on these results we make several observations.

(1) The IRUNet CNN performs much better on Gaussian noise whereas both the IRUNet and VDSR networks yield similar results on Poisson noised data. The same conclusion holds when the models are evaluated on true experimental data. We therefore assume that both training data and training methodology are more important than the actual detailed neural network architecture. At least for the used state-of-the-art models in this work.

(2) Filtering artificial Poisson noise statistics is an easier problem than filtering white Gaussian noise. A better model performance is thus obtained upon removal of Poisson noise. Most likely this stems from the fact that the pixel intensities and the noise are correlated in the case of Poisson noise while they are not for Gaussian noise. Combined with correlation between neighboring pixels, this reduces the amount of information. The overall challenge of the training task is therefore reduced.

(3) Training the deep CNNs on experimental data significantly improves denoising, which means that training on artificially noised data is ultimately insufficient when the models shall later be applied onto real experimental data.

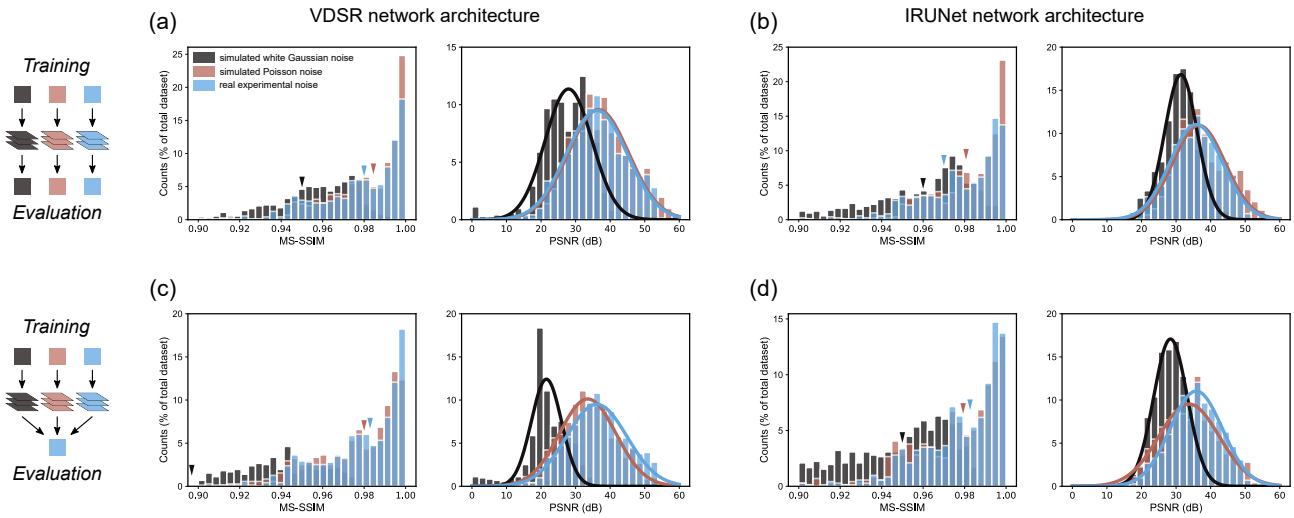

Figure 33: Performance comparison of the different network architectures and noise types using multiscale structural similarity (MS-SSIM) and peak signal-to-noise ratio (PSNR) between the actual measured high count statistic data and the denoised output. The colorcode indicates the different noise types used during the training. The MS-SSIM histograms contain triangular markers that point to the mean value of the respective distribution. The PSNR histograms have additionally been fitted using a Gaussian distribution. The upper panels (a) and (b) show the results of the VDSR and IRUNet architectures introduced in the main text, trained and evaluated on the same noise type. The lower panels (c) and (d) show the evaluation of these networks on the actual true low counting statistics data. Generally, we observe that the models trained on the artificial noise perform worse when evaluated on the true experimental noise compared to the models that have been trained on the latter. This indicates that training on simulated noise, which is the predominant procedure in scattering experiments, is still inferior to training on actual experimental noise.

	VDSR			IRUNet		
Training → Evaluation	PSNR	MS-SSIM	MAE+MS-SSIM	PSNR	MS-SSIM	MAE+MS-SSIM
Gaussian → Gaussian	28.37	0.818 (0.953)	0.816 (0.954)	31.40	0.957 (0.965)	0.960 (0.968)
Poisson → Poisson	36.98	0.987 (0.981)	0.986 (0.982)	36.62	0.982 (0.987)	0.983 (0.987)
Gaussian → Exp.	21.55	0.600 (0.881)	0.511 (0.884)	28.39	0.934 (0.949)	0.942 (0.953)
Poisson → Exp.	33.50	0.971 (0.980)	0.971 (0.980)	33.73	0.971 (0.980)	0.971 (0.980)
Exp. → Exp.	**36.11**	**0.973** (**0.982**)	**0.975** (**0.983**)	**35.84**	**0.974** (**0.982**)	**0.976** (**0.983**)

Table 1: Results of different training and evaluation protocols for the VDSR and IRUNet neural network architectures. The first column indicates the used training and evaluation data. The unit of PSNR is dB while the other measures are unitless. For the measures MS-SSIM and MAE+MS-SSIM both mean and median (in brackets) are provided where the latter also served as the loss function for the training of the networks. Training on experimental data (Exp.) yields a significantly better result as compared to training on data with artificial noise.

2.5 Summary and conclusion

The second part of this thesis was focused on the implementation of a deep convolutional neural network (CNN) for the task of denoising low count X-ray diffraction data, which are subject to some sort of Poisson noise. We compared the performance of two different neural network architectures when trained and evaluated on different noise statistics. We find that training on experimental data leads to significantly better denoising performance when the models are evaluated on experimental data. Artificially adding noise to high counting statistics data, which is the usual approach in AI-assisted image denoising, is thus still inferior to training on true experimental data. This suggests that denoising algorithms should be benchmarked on experimental data sets, which would require further efforts in experimental data acquisition. To generally improve the performance of the trained neural networks one could increase the amount of training data, also including data from other experimental scattering techniques such as small angle neutron scattering (SANS) [82] but also data resulting from angle resolved photo-emission spectroscopy (ARPES) [83] or transmission electron microscopy (TEM) [84][85] could serve as useful additions. With the presented X-ray diffraction data we noticed that the convergence during training was quite slow for both network architectures, which can be seen from the training curves in Figure 34. We assume that the slow convergence is because most of the training data frames have very similar contents respectively very small structural differences. This makes it more difficult for the neural network to filter out new distinct features leading to a slow convergence.

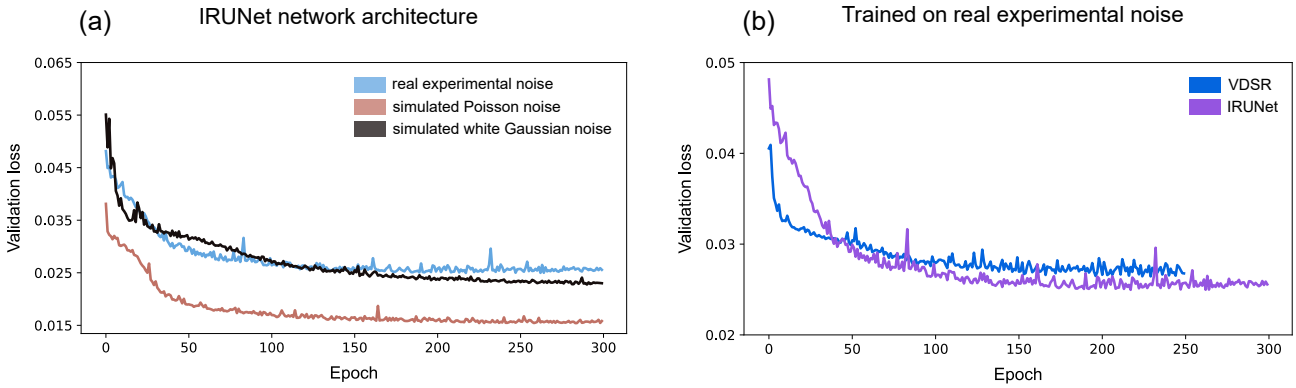

Figure 34: Training history showing the validation loss described by Equation 38 versus training epoch. (a) Comparison of the validation loss obtained by training the IRUNet architecture on different indicated noise types. (b) Validation loss of the different network architectures VDSR and IRUNet trained on the true experimental noise. We observe that the IRUNet neural network converges to a slightly lower loss than the VDSR neural network.

A Appendix

A.1 Algorithms

```
function path = cheapest_insertion(n, M)
  % CALCULATES A TRAVELING SALESPERSON SOLUTION
  %
  % inputs:
  % n (2D int) -> grid size
  % M (int) -> number of measurement locations
  %
  % outputs:
  % path (2D int)

  % Get random measurement locations
  locations = GET_RANDOM_LOCATIONS(grid_size, number_of_locations);

  % Calculate the distance matrix
  a = meshgrid(1:M);
  dmat = reshape(sqrt(sum((locations(a,:)-locations(a',:)).^2,2)),M,M);

  % Define the initial sub-tour
  tour = [START_ID, END_ID];
  open = setdiff(1:M, tour, 'stable');

  % Main algorithm
  while true
      if isempty(open)
          break
      end
      d_shortest = Inf; % minimum distance
      for kk = 1:length(tour)-1
        % Get new subtour distances using node kk
        [d_min, id_min] = GET_MINIMUM_DISTANCE(tour, open);
        if d_min < d_shortest
              id = open(min_id);
              node_to_insert = kk;
              shortest_dist = min_dist; % update distance
          end
      end
      % Remove index from the possible points
      open(open == id) = [];
      % Insert point into the obtained tour node
      tour = [tour(1:node_to_insert), id, tour(node_to_insert+1:end)];
  end

end
```

Listing 1: Cheapest insertion algorithm.

```
function node = binary_tree(node, pts, dim, cut, max_elements)
  % CALCULATES A BINARY TREE FOR A GIVEN NUMBER OF POINTS
  %
  % inputs:
  % node (object) -> tree cell
  % pts (object) -> points
  % dim (int) -> cutting dimension
  % cut (int) -> cutting fraction
  % max_elements (int) -> maximum leaf cell size
  %
  % outputs:
  % node (object) -> tree cell

  % Check if node is a leaf cell
  if node.isLeaf
    return
  end

  % Perform array partitioning
  boundary = cut*(node.rlow(dim) + node.rhigh(dim));
  [index, pts] = partition(pts, node.lower, node.upper, v, dim);
  next_dim = set_other_dimension();

  % Traverse the tree
  if (index > node.lower)
    rhigh = node.rhigh;
    rhigh(dim) = boundary;
    new_node = create_new_left_node();
    node.left = binary_tree(new_node, pts, next_dim, cut, max_elements);
  elseif (index < node.upper)
    rlow = node.rlow;
    rlow(dim) = boundary;
    new_node = create_new_right_node();
    node.right = binary_tree(new_node, pts, next_dim, cut, max_elements);
  end

end
```

Listing 2: Binary tree algorithm used for path tiling implementation.

A.2 Calculation of the parabolic effective electron mass

We start with the assumption that we have a isotropic parabolic dispersion relation in momentum space $E(\vec{k}) = E(k)$

$$E(k) = \frac{\hbar^2 k^2}{2m^*} \tag{40}$$

where m^* is the effective electron mass. In STM we measure the change of momentum $q = k_i - k_f$ with initial state k_i and final state k_f. Due to isotropy of the problem we have that

$$k_f = -k_i \equiv k \quad \Rightarrow \quad q = 2k \tag{41}$$

Equation 40 then becomes

$$E(q) = \frac{\hbar^2 q^2}{8m^*} \equiv aq^2 \quad \text{with} \quad a = \frac{\hbar^2}{8m^*} \tag{42}$$

The coefficient a is obtained by fitting a parabola to the dispersion relation using for example a least squares approach. Having a one can solve Equation 42 for the effective electron mass m^*.

A.3 SPGL1 working principle

In this thesis we utilized the basis pursuit denoise (BPDN) optimization routine provided by SPGL1 written in MATLAB [38][39]. SPGL1 has been developed for large-scale applications and can also solve complex-valued problems. In the presence of noise one can state the BPDN problem as solving an under-determined system of linear equations by minimizing the ℓ_1 norm of the solution with quadratic constraints on its residual

$$\text{(BDPN)} \quad \min ||x||_1 \text{ subject to } ||Ax - b||_2 \leq \sigma \tag{43}$$

where $x \in \mathbb{C}^{N\times 1}$, $b \in \mathbb{R}^{M\times 1}$ and $A \in \mathbb{C}^{M\times N}$ with $M \ll N$. Solutions of x are hereby in the form of a sparse approximation. SPGL1 aims at solving a least absolute shrinkage and selection operator (LASSO) problem [19] via a spectral projected-gradient (SPG) algorithm where one defines the convex one-norm regularized LASSO problem as

$$\text{(LASSO)} \quad \min ||Ax - b||_2 \text{ subject to } ||x||_1 \leq \tau \tag{44}$$

where τ is some constant. If x_τ denotes the optimal solution of Equation 44 then one defines the single-parameter function

$$\phi(\tau) = ||r||_2 \quad \text{with} \quad r = b - Ax_\tau \tag{45}$$

which is called the Pareto curve (see Figure 35). The Pareto curve gives the optimal value of Equation 44 for each $\tau > 0$ and therefore defines the optimal trade-off between the two-norm of the residual r and the one-norm of the solution. For certain parameters of τ and σ the BPDN problem coincides with the LASSO problem. The goal of the SPGL1 algorithm is to find a root of the nonlinear equation $\phi(\tau) = \sigma$. In every iteration an estimate of τ is used to define a new convex optimization problem. The solution to said problem will then yield information about the derivative of the Pareto curve to be used by a Newton root finding algorithm. Since the Pareto curve ϕ is continuously differentiable for all values of τ an efficient root-finding using Newton's method is guaranteed. The SPGL1 algorithm is very efficient with a worst-case complexity of $\mathcal{O}(n \log(n))$.

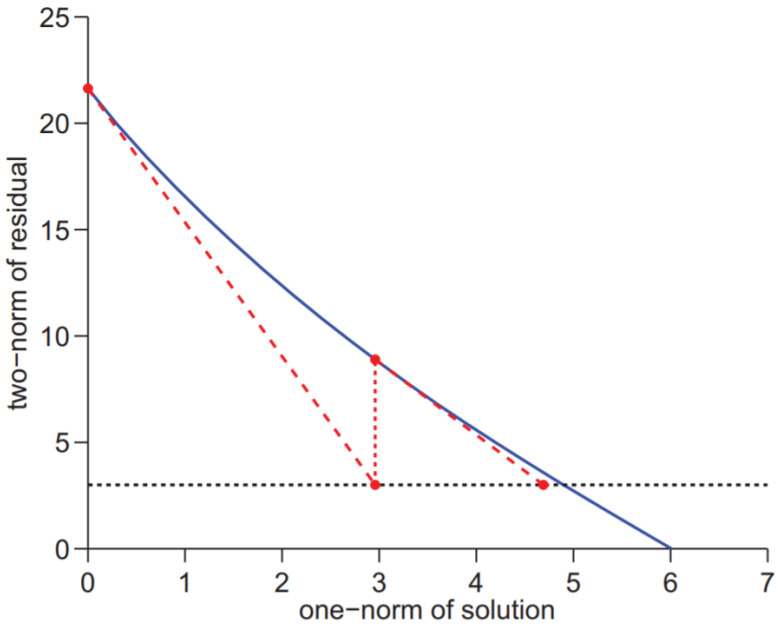

Figure 35: Typical Pareto curve (blue, solid line) showing two iterations of Newton's method (red, dashed line) [38]. Newton's root-finding method is used to find a solution to $\phi(\tau) = \sigma = ||r||_2 = ||x||_1$ that is where the two-norm of the residual („error") is equal to the one-norm of the solution x.

A.4 Effect of background correction on a single spectrum

The applied background described by Equation 31 in Section 1.4.2 was used to level individual path segments measured consecutively according to the adaptive sparse sampling scheme. By equalizing the conductance traces in terms of a reference mean μ_{ref} and standard deviation σ_{ref} one can improve the quality of the reconstruction. However, because this correction is applied onto every conductance trace j at energy E_j this will also lead to a change of the individual spectra as some energies might be shifted upwards while others might get shifted downwards regarding their conductance values. In Figure 36 we show the effect of discussed background correction on a Au(111) and $NbSe_2$ spectrum. Although the differences are not that prominent this approach might also suppress spectral features that are bound to localized states on the surface as these will get overshadowed by the majority of the globally occurring spectra. Furthermore, the deviation from the raw spectrum might increase with more frequent tip changes.

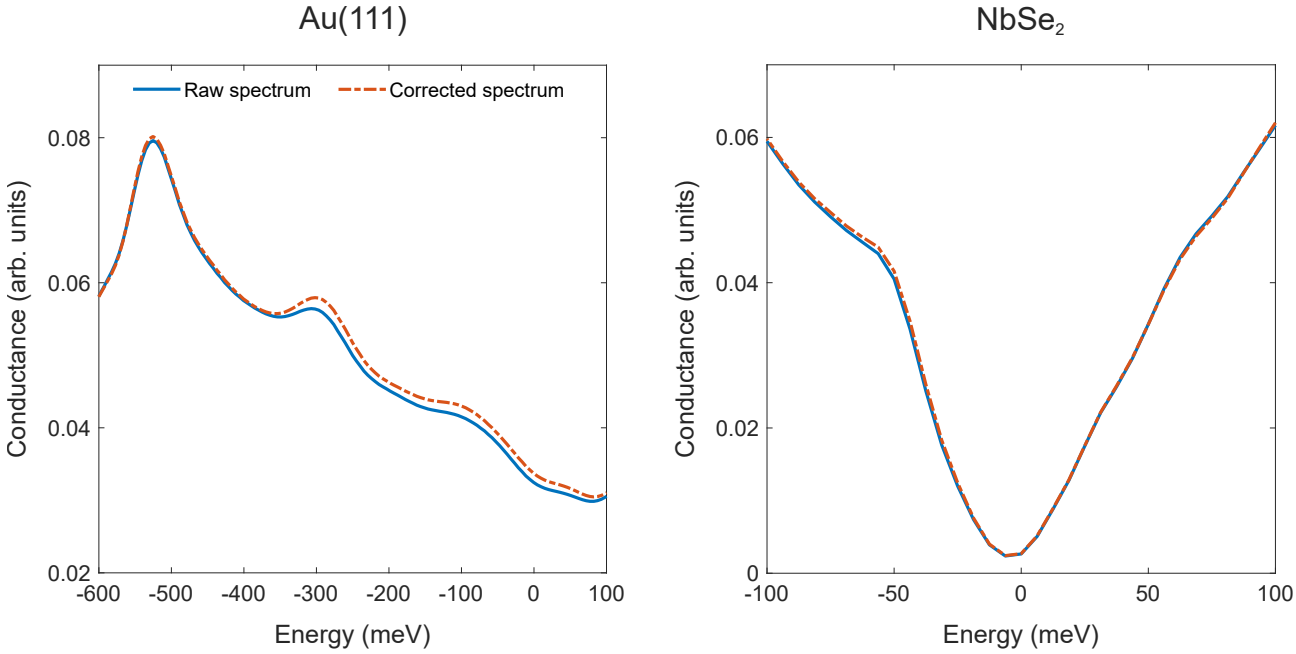

Figure 36: Applying a background correction onto individual path segments independent of energy slightly changes the individual spectra at some energies. Here shown are the average point spectra recorded on adaptively measured Au(111) and $NbSe_2$ samples that were discussed in Section 1.4.2 and 1.5.3. The corrected spectrum of $NbSe_2$ looks very similar to its raw version as there was no obvious tip change during the experiment.

References

[1] G. Binnig, and H. Rohrer. "Scanning Tunneling Microscopy", Surface Science 126, no. 1 (March 2, 1983): 236–44. doi.org/10.1016/0039-6028(83)90716-1

[2] L. Petersen, Ph. Hofmann, E.W. Plummer, and F. Besenbacher. "Fourier Transform–STM: Determining the Surface Fermi Contour", Journal of Electron Spectroscopy and Related Phenomena 109, no. 1–2 (August 2000): 97–115. doi:10.1016/S0368-2048(00)00110-9

[3] C. J. Chen. "Introduction to Scanning Tunneling Microscopy", 3rd ed. Monographs on the Physics and Chemistry of Materials, Oxford: Oxford University Press, 2021. doi:10.1093/oso/9780198856559.001.0001

[4] J. Bardeen, "Tunnelling from a Many-Particle Point of View", Physical Review Letters 6, no. 2 (January 15, 1961): 57–59. doi:10.1103/PhysRevLett.6.57

[5] M. F. Crommie, C. P. Lutz, and D. M. Eigler. "Imaging Standing Waves in a Two-Dimensional Electron Gas", Nature 363, no. 6429 (June 1993): 524–27. doi:10.1038/363524a0

[6] N. Avraham, J. Reiner, A. Kumar-Nayak, N. Morali, R. Batabyal, B. Yan, and H. Beidenkopf. "Quasiparticle Interference Studies of Quantum Materials", Advanced Materials 30, no. 41 (2018): 1707628. doi:10.1002/adma.201707628

[7] M. F. Crommie, C. P. Lutz, and D. M. Eigler. "Confinement of Electrons to Quantum Corrals on a Metal Surface", Science 262, no. 5131 (October 8, 1993): 218–20. doi:10.1126/science.262.5131.218

[8] C. E. Shannon, "Communication in the Presence of Noise", Proceedings of the IRE 37, no. 1 (January 1949): 10–21. doi.org/10.1109/JRPROC.1949.232969

[9] H. Nyquist, "Certain Topics in Telegraph Transmission Theory", Transactions of the American Institute of Electrical Engineers 47, no. 2 (April 1928): 617–44. doi.org/10.1109/T-AIEE.1928.5055024

[10] E. T. Whittaker, "XVIII.—On the Functions Which Are Represented by the Expansions of the Interpolation-Theory", Proceedings of the Royal Society of Edinburgh 35 (ed 1915): 181–94. doi.org/10.1017/S0370164600017806

[11] E. J. Candès, and M. B. Wakin, "An Introduction To Compressive Sampling", IEEE Signal Processing Magazine 25, no. 2 (March 2008): 21–30. doi.org/10.1109/MSP.2007.914731

[12] R. Baraniuk, M. A. Davenport, M. F. Duarte and C. Hedge, "An Introduction to Compressive Sensing", accessed March 30, 2022. legacy.cnx.org/content/col11133/latest/

[13] G. E. Moore, "Cramming More Components onto Integrated Circuits", Reprinted from Electronics, Volume 38, Number 8, April 19, 1965, Pp.114 Ff., IEEE Solid-State Circuits Society Newsletter 11, no. 3 (September 2006): 33–35. doi.org/10.1109/N-SSC.2006.4785860

[14] M. M. Waldrop, "The Chips Are down for Moore's Law", Nature News 530, no. 7589 (February 11, 2016): 144. doi.org/10.1038/530144a

[15] D. Taubman and M. Marcellin, "JPEG2000: Image Compression Fundamentals, Standards and Practice", 2002nd ed. (Boston: Springer, 2001)

[16] S. Foucart and H. Rauhut, "A Mathematical Introduction to Compressive Sensing", 2013th ed. (New York: Birkhäuser, 2013)

[17] E. J. Candès, J. Romberg, and T. Tao, "Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information", IEEE Transactions on Information Theory 52, no. 2 (February 2006): 489–509. doi.org/10.1109/TIT.2005.862083

[18] D. L. Donoho, "Compressed Sensing", IEEE Transactions on Information Theory 52, no. 4 (April 2006): 1289–1306. doi.org/10.1109/TIT.2006.871582

[19] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso", Journal of the Royal Statistical Society. Series B (Methodological) 58, no. 1 (1996): 267–88. jstor.org/stable/2346178

[20] M. Rudelson and R. Vershynin, "On Sparse Reconstruction from Fourier and Gaussian Measurements", Communications on Pure and Applied Mathematics 61, no. 8 (2008): 1025–45. doi.org/10.1002/cpa.20227

[21] P. Sidorenko, O. Kfir, Y. Shechtman, A. Fleischer, Y. C. Eldar, M. Segev, and O. Cohen, "Sparsity-Based Super-Resolved Coherent Diffraction Imaging of One-Dimensional Objects", Nature Communications 6, no. 1 (September 8, 2015): 8209. doi.org/10.1038/ncomms9209

[22] R. Otazo, E. J. Candès, and D. K. Sodickson, "Low-Rank plus Sparse Matrix Decomposition for Accelerated Dynamic MRI with Separation of Background and Dynamic Components", Magnetic Resonance in Medicine 73, no. 3 (2015): 1125–36. doi.org/10.1002/mrm.25240

[23] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan, "Sparse Representation for Computer Vision and Pattern Recognition", Proceedings of the IEEE 98, no. 6 (June 2010): 1031–44. doi.org/10.1109/JPROC.2010.2044470

[24] V. M. Patel, G. R. Easley, D. M. Healy, and R. Chellappa, "Compressed Synthetic Aperture Radar", IEEE Journal of Selected Topics in Signal Processing 4, no. 2 (April 2010): 244–54. doi.org/10.1109/JSTSP.2009.2039181

[25] J. Oppliger and F. D. Natterer, "Sparse Sampling for Fast Quasiparticle-Interference Mapping", Physical Review Research 2, no. 2 (May 1, 2020): 023117. doi.org/10.1103/PhysRevResearch.2.023117

[26] K. McElroy, R. W. Simmonds, J. E. Hoffman, D.-H. Lee, J. Orenstein, H. Eisaki, S. Uchida, and J. C. Davis, "Relating Atomic-Scale Electronic Phenomena to Wave-like Quasiparticle States in Superconducting $Bi_2Sr_2CaCu_2O_{8+\delta}$", Nature 422, no. 6932 (April 2003): 592–96. doi.org/10.1038/nature01496

[27] I. Battisti, G. Verdoes, K. van Oosten, K. M. Bastiaans, and M. P. Allan, "Definition of Design Guidelines, Construction, and Performance of an Ultra-Stable Scanning Tunneling Microscope for Spectroscopic Imaging", Review of Scientific Instruments 89, no. 12 (December 2018): 123705. doi.org/10.1063/1.5064442

[28] M. Steinbrecher, H. Harutyunyan, C. R. Ast, and D. Wegner, "Rashba-Type Spin Splitting from Interband Scattering in Quasiparticle Interference Maps" Physical Review B 87, no. 24 (June 28, 2013): 245436.
doi.org/10.1103/PhysRevB.87.245436

[29] Z. Wang, D. Walkup, P. Derry, T. Scaffidi, M. Rak, S. Vig, A. Kogar, et al., "Quasiparticle Interference and Strong Electron–Mode Coupling in the Quasi-One-Dimensional Bands of Sr_2RuO_4", Nature Physics 13, no. 8 (August 2017): 799–805. doi.org/10.1038/nphys4107

[30] J. E. Hoffman, PhD Thesis, "A Search for Alternative Electronic Order in the High Temperature Superconductor $Bi_2Sr_2CaCu_2O_{8+\delta}$ by Scanning Tunneling Microscopy" (2003)

[31] T. Balashov, M. Meyer, and W. Wulfhekel, "A Compact Ultrahigh Vacuum Scanning Tunneling Microscope with Dilution Refrigeration", Review of Scientific Instruments 89, no. 11 (November 2018): 113707. doi.org/10.1063/1.5043636

[32] Y. Nakanishi-Ohno, M. Haze, Y. Yoshida, K. Hukushima, Y. Hasegawa, and M. Okada, "Compressed Sensing in Scanning Tunneling Microscopy/Spectroscopy for Observation of Quasi-Particle Interference", Journal of the Physical Society of Japan 85, no. 9 (September 15, 2016): 093702. doi.org/10.7566/JPSJ.85.093702

[33] B. Zengin, J. Oppliger, D. Liu, L. Niggli, T. Kurosawa, and F. D. Natterer, "Fast Spectroscopic Mapping of Two-Dimensional Quantum Materials", Physical Review Research 3, no. 4 (November 10, 2021): L042025. doi.org/10.1103/PhysRevResearch.3.L042025

[34] J. Kirk, "Open Traveling Salesman Problem - Genetic Algorithm", MATLAB Central File Exchange (accessed on 2 July 2020). ch.mathworks.com/matlabcentral/fileexchange/21196-open-traveling-salesman-problem-genetic-algorithm

[35] T. A. J. Nicholson, "A Sequential Method for Discrete Optimization Problems and Its Application to the Assignment, Travelling Salesman, and Three Machine Scheduling Problems", IMA Journal of Applied Mathematics 3, no. 4 (December 1, 1967): 362–75. doi.org/10.1093/imamat/3.4.362

[36] D. J. Rosenkrantz, R. E. Stearns, and P. M. Lewis, "Approximate Algorithms for the Traveling Salesperson Problem", in Proceedings of the 15th Annual Symposium on Switching and Automata Theory (Swat 1974), SWAT '74 (USA: IEEE Computer Society, 1974), 33–42. doi.org/10.1109/SWAT.1974.4

[37] E. van den Berg, and M. P. Friedlander, "Sparco/ChangeLog at Master · MPF-Optimization-Laboratory/Sparco", GitHub, accessed September 28, 2021. github.com/MPF-Optimization-Laboratory/Sparco

[38] E. van den Berg, and M. P. Friedlander, "Probing the Pareto Frontier for Basis Pursuit Solutions," SIAM Journal on Scientific Computing 31, no. 2 (January 2009): 890–912. doi.org/10.1137/080714488

[39] E. van den Berg, and M. P. Friedlander, (December 2019) "SPGL1 - Sparse Least Squares", accessed March 30, 2022 friedlander.io/spgl1/

[40] C. Wöll, S. Chiang, R. J. Wilson, and P. H. Lippel, "Determination of Atom Positions at Stacking-Fault Dislocations on Au(111) by Scanning Tunneling Microscopy", Physical Review B 39, no. 11 (April 15, 1989): 7988–91. doi.org/10.1103/PhysRevB.39.7988

[41] F. Reinert and G. Nicolay, "Influence of the Herringbone Reconstruction on the Surface Electronic Structure of Au(111)", Applied Physics A 78, no. 6 (March 1, 2004): 817–21. doi.org/10.1007/s00339-003-2436-6

[42] W. Chen, V. Madhavan, T. Jamneala, and M. F. Crommie, "Scanning Tunneling Microscopy Observation of an Electronic Superlattice at the Surface of Clean Gold", Physical Review Letters 80, no. 7 (February 16, 1998): 1469–72. doi.org/10.1103/PhysRevLett.80.1469

[43] B. Voigtländer, Scanning Probe Microscopy: Atomic Force Microscopy and Scanning Tunneling Microscopy, 2015th ed. (New York, NY: Springer, 2015)

[44] R. Borgani, M. Gilzad Kohan, A. Vomiero, and D. B. Haviland, "Fast Multifrequency Measurement of Nonlinear Conductance", Physical Review Applied 11, no. 4 (April 19, 2019): 044062. doi.org/10.1103/PhysRevApplied.11.044062

[45] J. E. Hoffman, K. McElroy, D.-H. Lee, K. M Lang, H. Eisaki, S. Uchida, and J. C. Davis, "Imaging Quasiparticle Interference in $Bi_2Sr_2CaCu_2O_{8+\delta}$", Science 297, no. 5584 (August 16, 2002): 1148–51. doi.org/10.1126/science.1072640

[46] J. E. Hoffman, E. W. Hudson, K. M. Lang, V. Madhavan, H. Eisaki, S. Uchida, and J. C. Davis, "A Four Unit Cell Periodic Pattern of Quasi-Particle States Surrounding Vortex Cores in $Bi_2Sr_2CaCu_2O_{8+\delta}$", Science 295, no. 5554 (January 18, 2002): 466–69. doi.org/10.1126/science.1066974

[47] Q. Gu, S. Wan, Q. Tang, Z. Du, H. Yang, Q.-H. Wang, R. Zhong, J. Wen, G. D. Gu, and H.-H. Wen, "Directly Visualizing the Sign Change of D-Wave Superconducting Gap in $Bi_2Sr_2CaCu_2O_{8+\delta}$ by Phase-Referenced Quasiparticle Interference", Nature Communications 10, no. 1 (April 8, 2019): 1603. doi.org/10.1038/s41467-019-09340-5

[48] Y. Yuan, X. Yang, L. Peng, Z.-J. Wang, J. Li, C.-J. Yi, J.-J. Xian, Y.-G. Shi, and Y.-S. Fu, "Quasiparticle Interference of Fermi Arc States in the Type-II Weyl Semimetal Candidate WTe_2", Physical Review B 97, no. 16 (April 26, 2018): 165435. doi.org/10.1103/PhysRevB.97.165435

[49] W. Zhang, Q. Wu, L. Zhang, S.-W. Cheong, A. A. Soluyanov, and W. Wu, "Quasiparticle Interference of Surface States in the Type-II Weyl Semimetal WTe_2", Physical Review B 96, no. 16 (October 13, 2017): 165125. doi.org/10.1103/PhysRevB.96.165125

[50] J. Oppliger, B. Zengin, D. Liu, K. Hauser, C. Witteveen, F. von Rohr, and F. D. Natterer, "Adaptive Sparse Sampling for Quasiparticle Interference Imaging" ArXiv:2204.04475 [Cond-Mat], April 9, 2022. arxiv.org/abs/2204.04475

[51] H. Jung and D.-G. Gweon, "Creep Characteristics of Piezoelectric Actuators", Review of Scientific Instruments 71, no. 4 (April 2000): 1896–1900. doi.org/10.1063/1.1150559

[52] J. Gan and X. Zhang, "A Review of Nonlinear Hysteresis Modeling and Control of Piezoelectric Actuators", AIP Advances 9, no. 4 (April 2019): 040702. doi.org/10.1063/1.5093000

[53] D. Padfield, "Generalized Normalized Cross Correlation", MathWorks FileExchange, accessed April 23, 2022. mathworks.com/matlabcentral/fileexchange/29005-generalized-normalized-cross-correlation

[54] Z. Wang, E.P. Simoncelli, and A.C. Bovik, "Multiscale Structural Similarity for Image Quality Assessment", in The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003, vol. 2, 2003, 1398-1402 Vol.2. doi.org/10.1109/ACSSC.2003.1292216

[55] M. J. Lawler, K. Fujita, Jhinhwan Lee, A. R. Schmidt, Y. Kohsaka, Chung Koo Kim, H. Eisaki, et al., "Intra-Unit-Cell Electronic Nematicity of the High-T_c Copper-Oxide Pseudogap States", Nature 466, no. 7304 (July 2010): 347–51. doi.org/10.1038/nature09169

[56] M. Leutenegger (2022). "Hankel transform", MATLAB Central File Exchange, Retrieved May 6, 2022. www.mathworks.com/matlabcentral/fileexchange/13371-hankel-transform

[57] C. Mann, "Commentary on the Interpretation of Fourier-Transform Scanning Tunneling Microscopy Data", ArXiv:1509.07807 [Cond-Mat], September 25, 2015. arxiv.org/abs/1509.07807

[58] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, and M. Norouzi, "Image Super-Resolution via Iterative Refinement", ArXiv:2104.07636 [Cs, Eess], June 30, 2021. arxiv.org/abs/2104.07636

[59] X. Wang, Y. Li, H. Zhang, and Y. Shan, "Towards Real-World Blind Face Restoration with Generative Facial Prior", in 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, 9164–74. doi.org/10.1109/CVPR46437.2021.00905

[60] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising", IEEE Transactions on Image Processing 26, no. 7 (July 2017): 3142–55. doi.org/10.1109/TIP.2017.2662206

[61] V. Jain and S. Seung, "Natural Image Denoising with Convolutional Networks", in Advances in Neural Information Processing Systems, vol. 21 (Curran Associates, Inc., 2008). https://papers.nips.cc/paper/2008/hash/c16a5320fa475530d9583c34fd356ef5-Abstract.html

[62] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a Fast and Flexible Solution for CNN-Based Image Denoising", IEEE Transactions on Image Processing 27, no. 9 (September 2018): 4608–22. doi.org/10.1109/TIP.2018.2839891

[63] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a Convolutional Neural Network", in 2017 International Conference on Engineering and Technology (ICET), 2017, 1–6. doi.org/10.1109/ICEngTechnol.2017.8308186

[64] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2Noise: Learning Image Restoration without Clean Data", ArXiv:1803.04189 [Cs, Stat], October 29, 2018. arxiv.org/abs/1803.04189

[65] A. Krull, T.-O. Buchholz, and F. Jug "Noise2Void - Learning Denoising from Single Noisy Images", ArXiv:1811.10980 [Cs], April 5, 2019. arxiv.org/abs/1811.10980

[66] J. Batson and L. Royer, "Noise2Self: Blind Denoising by Self-Supervision", in Proceedings of the 36th International Conference on Machine Learning (International Conference on Machine Learning, PMLR, 2019), 524–33. proceedings.mlr.press/v97/batson19a.html

[67] Y. Kim, D. Oh, S. Huh, D. Song, S. Jeong, J. Kwon, M. Kim, et al., "Deep Learning-Based Statistical Noise Reduction for Multidimensional Spectral Data", Review of Scientific Instruments 92, no. 7 (July 1, 2021): 073901. doi.org/10.1063/5.0054920

[68] Z. J. W. A. Leijten, A. D. A. Keizer, G. de With, and H. Friedrich, "Quantitative Analysis of Electron Beam Damage in Organic Thin Films", The Journal of Physical Chemistry C 121, no. 19 (May 18, 2017): 10552–61. doi.org/10.1021/acs.jpcc.7b01749

[69] W. Knafo, F. Duc, F. Bourdarot, K. Kuwahara, H. Nojiri, D. Aoki, J. Billette, et al., "Field-Induced Spin-Density Wave beyond Hidden Order in URu2Si2", Nature Communications 7, no. 1 (October 20, 2016): 13075. doi.org/10.1038/ncomms13075

[70] J. P. C. Ruff, J.-H. Chu, H.-H. Kuo, R. K. Das, H. Nojiri, I. R. Fisher, and Z. Islam, "Susceptibility Anisotropy in an Iron Arsenide Superconductor Revealed by X-Ray Diffraction in Pulsed Magnetic Fields", Physical Review Letters 109, no. 2 (July 11, 2012): 027004. doi.org/10.1103/PhysRevLett.109.027004

[71] A. Bansal, R. Sharma, and M. Kathuria, "A Systematic Review on Data Scarcity Problem in Deep Learning: Solution and Applications", ACM Computing Surveys, November 22, 2021. doi.org/10.1145/3502287

[72] C. Tian, L. Fei, W. Zheng, Y. Xu, W. Zuo, and C.-W. Lin., "Deep Learning on Image Denoising: An Overview", Neural Networks 131 (November 1, 2020): 251–75. doi.org/10.1016/j.neunet.2020.07.025

[73] J. Chang, Ch. Niedermayer, R. Gilardi, N. B. Christensen, H. M. Rønnow, D. F. McMorrow, M. Ay, et al., "Tuning Competing Orders in $La_{2-x}Sr_xCuO_4$ Cuprate Superconductors by the Application of an External Magnetic Field" Physical Review B 78, no. 10 (September 25, 2008): 104525. doi.org/10.1103/PhysRevB.78.104525

[74] N. B. Christensen, J. Chang, J. Larsen, M. Fujita, M. Oda, M. Ido, N. Momono, et al., "Bulk Charge Stripe Order Competing with Superconductivity in La2-xSrxCuO4 (x=0.12)", ArXiv:1404.3192 [Cond-Mat], April 11, 2014. arxiv.org/abs/1404.3192

[75] V. Thampy, M. P. M. Dean, N. B. Christensen, L. Steinke, Z. Islam, M. Oda, M. Ido, N. Momono, S. B. Wilkins, and J. P. Hill, "Rotated Stripe Order and Its Competition with Superconductivity in $La_{1.88}Sr_{0.12}CuO_4$", Physical Review B 90, no. 10 (September 26, 2014): 100510. doi.org/10.1103/PhysRevB.90.100510

[76] K. Ikeuchi, ed., Computer Vision: A Reference Guide (Cham: Springer International Publishing, 2021). doi.org/10.1007/978-3-030-63416-2

[77] J. Kim, J. Kwon Lee, and K. M. Lee, "Accurate Image Super-Resolution Using Very Deep Convolutional Networks", in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, 1646–54. doi.org/10.1109/CVPR.2016.182

[78] G. Zuluaga, F. Hernan, F. Bardozzo, J. Ivan Rios Patino, and R. Tagliaferri, "Blind Microscopy Image Denoising with a Deep Residual and Multiscale Encoder/Decoder Network", in 2021 43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC), 2021, 3483–86. doi.org/10.1109/EMBC46164.2021.9630502

[79] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization", ArXiv:1412.6980 [Cs], January 29, 2017. arxiv.org/abs/1412.6980

[80] S. J. Reddi, S. Kale, and S. Kumar, "On the Convergence of Adam and Beyond", ArXiv:1904.09237 [Cs, Math, Stat], April 19, 2019. arxiv.org/abs/1904.09237

[81] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss Functions for Image Restoration With Neural Networks", IEEE Transactions on Computational Imaging 3, no. 1 (March 2017): 47–57. doi.org/10.1109/TCI.2016.2644865

[82] S. Mühlbauer, D. Honecker, É A. Périgo, F. Bergner, S. Disch, A. Heinemann, S. Erokhin, et al., "Magnetic Small-Angle Neutron Scattering", Reviews of Modern Physics 91, no. 1 (March 4, 2019): 015004. doi.org/10.1103/RevModPhys.91.015004

[83] J. A. Sobota, Y. He, and Z.-X. Shen, "Angle-Resolved Photoemission Studies of Quantum Materials", Reviews of Modern Physics 93, no. 2 (May 26, 2021): 025006. doi.org/10.1103/RevModPhys.93.025006

[84] H. S. Kushwaha, S. Tanwar, K. S. Rathore, and S. Srivastava, "De-Noising Filters for TEM (Transmission Electron Microscopy) Image of Nanomaterials", in 2012 Second International Conference on Advanced Computing Communication Technologies, 2012, 276–81. doi.org/10.1109/ACCT.2012.41

[85] N. Mevenkamp, P. Binev, W. Dahmen, P. M. Voyles, A. B. Yankovich, and B. Berkels, "Poisson Noise Removal from High-Resolution STEM Images Based on Periodic Block Matching", Advanced Structural and Chemical Imaging 1, no. 1 (March 25, 2015): 3. doi.org/10.1186/s40679-015-0004-8