A Holistic Approach to Structural Dynamics Using Serial Crystallography

Dissertation with the aim to achieve a doctoral degree

at the Faculty of Mathematics, Informatics and Natural Sciences

Department of Chemistry

University of Hamburg

submitted by Wolfgang Brehm

Hamburg, 2023-01-25

defended

Hamburg, 2023-11-03

Acknowledgments

DECLARATION

I hereby declare to have authored the dissertation with the title:

A Holistic Approach to Structural Dynamics Using Serial Crystallography

independently and without the aide of anything but which was indicated. Passages copied verbatim or with identical meaning from other sources (including from a website and other electronic databases) were indicated in every case by citing the source and quotation marks when the same words were used. Furthermore I declare that this dissertation has not been submitted elsewhere. I archived and will keep available the materials to reproduce the results in this dissertation.

_____   _____

Ort, Datum   Wolfgang Brehm

Abstract

The crystallographic data processing is investigated with the goal of extracting more structural information from serial crysallographic experiments. Systematic errors resulting from uncontrollable aspects in the experimental setup were reduced in several places, but most significantly by modelling the expected intensity of partially observed intensities using Gaussian basis functions to improve merging. These improvements are tied in with several applications, like phasing of continuous diffraction and drug screening for combating SARS-CoV-2.

Zusammenfassung

Der gesamte kristallographische Datenverarbeitungsprozess wird untersucht, um mehr Information über die Struktur von Makromolekülen aus den Experimenten ziehen zu können. Systematische Fehler, die ein unvermeidbarer Aspekt der Experimente sind, können an mehreren Stellen reduziert werden, insbesondere durch die Modellierung nur unvollständig messbarer Intensitäten in Beugungsmustern mittels Gaußscher Basisfunktionen. Projekte aus der Anwendung der strukturellen Kristallographie, wie Methoden zur Phasierung und die gezielte Suche nach Virustatika für SARS-CoV-2, werden vorgestellt, in welchen die Verbesserungen teils zum Zug kommen.

Relevant publications

The following publications in peer-reviewed journals have led to sections of this thesis:

- Section 6.9 is based on 'Crystal Diffraction Prediction and Partiality estimation using Gaussian basis functions' [1]

- Section 9 is based on 'Hash Tables with Pseudorandom Global Order' [2]

Contents

1. Introduction

Chemistry mostly investigates and characterizes ensembles of molecules and their interactions in bulk. The reactivity of atoms is determined by their electronic structure, which is well known and can be derived from basic physical laws in good accordance with experimental results [3]. The reactivity of molecules and especially macromolecules however is determined also by the three dimensional arrangement of their atoms, which is less well known and exponentially harder to derive. Experimental methods that are able to resolve these structures therefore are invaluable, and there are only a few which are able to achieve this with sufficient resolution. The same goes for reactions between molecules. Having insight into the movement of structures during a reaction can lead to a much better understanding. Among the methods that are able to resolve structures of macromolecules at an atomic level, crystallography is the most prolific, but it is being challenged from two sides. With smaller molecules up to small proteins, where NMR is more convenient and provides better insight, and with large macromolecules, where electron microscopy has been successful. All three methods struggle with dynamics and structural diversity differently. This thesis however is primarily concerned with improving the processing of diffraction data, especially from crystallography.

There are fundamental limits of how precisely nanoscopic structures can be measured, because each measurement is an interaction that changes the object, and observations are quantized. In X-Ray crystallography the interaction results in radiation damage and the quantization leads to a statistical distribution of photon counts. Rotational crystallography often gets to within less than a factor 2 of the limiting photon counting statistic, but when rotating the sample is not possible, the number of photons required for a given precision in the determined structure is a magnitude higher. It is not far-fetched to postulate that this is due to not yet sufficiently understood systematic effects, that distort the measurement seemingly randomly. Those effects then have to be marginalized experimentally, thereby making up the lack of understanding with more random observations. This work is the search for the largest of these errors and a proposed solution. Almost every step of the way from diffraction pattern to molecular structure estimate was scrutinized looking for improvements.

The investigation was prompted by the initial success in clustering of crystallographic data to resolve indexing ambiguities [4] showing that it is possible to resolve structural differences between individual diffraction snapshots when they are very large, but not sufficient to resolve structural variation between crystals of the same kind. However, the same algorithm was able to resolve relatively much smaller differences between diffraction datasets involving rotation [5, 6]. The hope is that when enough of these errors can be corrected, finding structural variation between single crystals will become possible.

2. Synopsis

As data processing is essentially statistical inference from experimental results, section 3 will begin more theoretically with an introduction on the statistical framework used later on. Most of this section can be found in a good textbook like Kendall [7] and Gelman et al. [8], it is meant as a summary and refresher, but the derivation of the optimal weights for a weighted median in section 3.7.1 is novel. Fourier transforms are introduced in section 4 because they are one of the main computational tool used in this work besides function optimization. The fundamental and well-known physical nature of diffraction experiments is derived in section 5 from Maxwell's equation, briefly, to enumerate the important limitations of the commonly used approximate results, which serve as the basis for the Gaussian partiality model in section 6.9. Modelling the partiality using Gaussian basis functions is the main contribution of this work to serial crystallography, which is introduced in section 6.8 as a method of structural crystallography (section 6). Minor contributions are an investigation of different merging algorithms in section 6.6, a small adjustment to the Cromer-Mann coefficients to make them easier to work with and more physically sensible in section 6.7, and section 6.10 is an example for the application of structural crystallography, where structural knowledge enabled us to find inhibitors to the SARS-CoV-2 M$^{\text{pro}}$ more rationally. Lastly section 9 introduces a memory efficient hash table that was developed during this work to accelerate detector calibration and the appendix contains the tables for coefficients similar to the Cromer-Mann coefficients and derivations that are too verbose.

3. Statistics

Statistics is the discipline that describes populations and distributions based on data. Especially relevant in the following will be distributions arising from the quantization of light and from other measurement errors. Because quantum effects and measurement errors are hardly avoidable in any scientific experiment, we first need to understand their distributions better, to then be able to infer more abstract properties from noisy data and make the most out of the experimental results we can obtain.

The following is mostly an introduction to statistics as can be found in many introductory books. Section 3.7.1, the optimal weights for the weighted median, is novel and section 3.8 on cuts and projetctions of Gaussian probability densities is unique in its presentation and and generality.

3.1. Probability Distributions

A probability distribution is the mathematical description of the probability of outcomes. A probability can either mean the asymptotic frequency of a given outcome (the frequentist perspective) or encode the belief about the asymptotic frequency of the outcome given limited knowledge (the Bayesian perspective) [7, Chap. 7]. A probability distribution can be expressed by its probability density, a function that, when integrated over a range $[a, b]$ of outcomes gives the probability for any one of these outcomes to occur [7, Chap.1]. Random variables are denoted with capital letters and the corresponding realizations with lower case letters. In this notation the probability P of the random variable X being part of the range $[a, b]$ and the probability density f_X are related by the following integral:

$$P\left(a \leq X \leq b\right) = \int_a^b f_X\left(x\right) dx$$

Or equivalently in terms of the cumulative distribution function F_X, which is the integral of the probability density, when the random variable is continuous, or the partial sum, when the random variable is discrete [7, Chap. 5].

$$F_X\left(a\right) = P\left(X \leq a\right)$$
$$P\left(a < X \leq b\right) = F_X\left(b\right) - F_X\left(a\right)$$

The inverse cumulative distribution function is the quantile function [7, Chap. 2]. It returns the limit a for which samples from the distribution are less than or equal with the given probability P.

The **mean** value μ or expected value of the random variable X is the limit of the average of more and more random samples [7, Chap. 2]. Given a probability density $f_X(x)$ it can be calculated with the following integral:

$$E\left(X\right) = \int x\, f_X(x)\, dx \tag{1}$$

While the mean value μ depends on the location of the density, the central moments are a way to characterize the shape of probability distributions [7, Chap. 3]. Central moments μ_n are defined as the expected

value of powers of deviations of random variables from the mean:

$$\mu_n \equiv \mathrm{E}\left[(X - \mathrm{E}(X))^n\right] \tag{2}$$

The zeroth central moment is 1. The first central moment is 0, if it exists. The second central moment, if it exists, is the variance, its square root is the standard deviation and commonly abbreviated with σ.

$$\mathrm{Var}(X) \equiv \mathrm{E}\left[(X - \mathrm{E}(X))^2\right] \tag{3}$$

The **median** value $\tilde{\mu}$ is the value for which the cumulative distribution reaches $1/2$, therefore the value of the quantile function at $1/2$:

$$\tilde{\mu} = \mathrm{F}_X^{-1}\left(\frac{1}{2}\right) \tag{4}$$

The **mode** is the value of x with the highest probability density [7, Chap. 2].

The density of a probability distribution for values far from the mean is called the **tail** of the distribution [7, Chap. 1]. Distributions that asymptotically have a higher density than the exponential distribution, are said to be heavy-tailed [9]. Super-heavy tailed distributions have a decay asymptotically slower than x^c for any power c smaller than negative one [10]. If c were exactly negative one it would not be a proper probability distribution any more, because the integral of x^{-1} is not finite.

The **covariance** is the variance of two or more random variables that can be inferred from the variance of the other. The covariance between the random variables X and Y is given by the following expectation:

$$\mathrm{Cov}(X, Y) = \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])] \tag{5}$$

3.1.1. Functions of Random Variables

The application of a function to a random variable gives another random variable. Let $Y = g(X)$, then the cumulative distribution function of Y is $\mathrm{F}_Y(y) = \mathrm{P}(g(X) \leq y)$. If the function $g(y)$ is monotonically increasing, it can be inverted and the cumulative distribution function of Y is given by:

$$\mathrm{P}(g(X) \leq y) = \mathrm{F}_X\left(g_n^{-1}(y)\right) \quad \text{if } g(y) \text{ increasing} \tag{6}$$

If the function $g(y)$ is monotonically decreasing, it can be inverted and the cumulative distribution function of Y is given by:

$$\mathrm{P}(g(X) \leq y) = 1 - \mathrm{F}_X\left(g_n^{-1}(y)\right) \quad \text{if } g(y) \text{ decreasing} \tag{7}$$

If the function is not bijective but piecewise monotonic, the inversion is ambiguous and the distribution function needs to be composed of sums of the different branches of the inverse function over the monotonic (and therefore unambiguous) sections [11].

The distribution of functions of several random variables can be derived by rearranging the function to only depend on one variable, which can then be marginalized [12]. In general this can be hard to do, but many functions can be assembled from products and sums. The exact distribution of the sum of two

random variables X and Y can be calculated in the following manner:

$$f(x, y) = x + y$$
$$f^{-1}(y) = f(x, y) - x$$
$$f(x, y) = f(x, f^{-1}(y))$$
$$p_Z(z = x + y) = \int_{\infty}^{\infty} p_{XY}(x, z - x) dx$$

If X and Y are independent, then $p_{XY}(x, y) = p_X(x)p_Y(y)$ and this integral is the convolution of the densities p_X and p_Y:

$$p_Z(z = x + y) = \int_{\infty}^{\infty} p_X(x)p_Y(z - x) dx \tag{8}$$

Similarly, if $Z = XY$ is the product of two independent random variables, then the probability distribution $p_Z(z)$ can be computed with the following integral:

$$p_Z(z = xy) = \int_{\infty}^{\infty} p_X(x)p_Y(\frac{z}{x})\frac{1}{|x|} dx \tag{9}$$

Because this procedure can get complicated quickly, even just for sums and products, one can get an approximation by computing just the mean and variance of functions of random variables instead of the whole function. With help of the principle of maximum entropy (section 3.1.2) and using the same reasoning as in the central limit theorem in section 3.5, knowing just the mean and the variance can give a good approximation of a probability distribution in most cases, so long they are defined and finite.

Just using the following rules for addition and multiplication, the mean and variance of distribution of many functions can be derived, just like the functions themselves can often be composed from addition and multiplication and their inverses alone [13, Chap. 4]. The inverse of addition can be composed by an addition and a multiplication with negative one, but there is no simple relation between the mean and variance of a random variable and its multiplicative inverse. If they exist, these properties can be derived using equations 6 and 7 in most cases.

$$E(X + Y) = E(X) + E(Y) \tag{10}$$
$$Var(X + Y) = Var(X) + Var(Y) + Cov(X, Y) \tag{11}$$
$$E(XY) = E(X)E(Y) + Cov(X, Y) \tag{12}$$
$$Var(XY) = Var(X)Var(Y) + Var(X)(E(Y))^2 + Var(Y)(E(X))^2 \tag{13}$$

Even more cases can be covered by recognizing situations where first-order approximations are sufficient to estimate the error propagation [13, Chap. 4]. This is the case when the errors lie in an approximately linear range of the function. If this is the case, then the expected value a function of a random variable is approximately equal to the function applied to the expected value of the random variable and the standard deviation is approximated by the first derivative evaluated at the expected value of the random variable

times the standard deviation of the random variable:

$$E(f(X)) \approx f(E(X)) \tag{14}$$

$$\text{Var}(f(X)) \approx \left(\frac{df}{dX}(E(X)) \right)^2 \text{Var}(X) \tag{15}$$

Now this still does not cover all functions of random variables and in some cases tedious integrals cannot be avoided, but all techniques that will be used are covered.

3.1.2. Maximum Entropy and Minimum Information

The principle of maximum entropy states that, given some constraints or prior knowledge, the prior probability distribution, representing nothing more than this state of knowledge, is the distribution with the maximum entropy [14]. The probability distribution with maximum entropy is also the one that is the most probable, which aligns with the minimum description length principle, see section 3.4. For this reason many common probability distributions are maximum entropy distributions given some constraints. Entropy is a measure of the average information of all outcomes. The apparent paradox of maximum entropy representing minimum information in this context while usually representing maximum information can be resolved by realizing that this is due to a difference in the point of view [11]. Samples from a high entropy distribution are 'surprising', highly informative and equivalently need a lot of data to encode them, precisely because the knowledge about their distribution is low. The entropy H of a distribution with probability $p(x)$ is the expected value of the negative logarithm of the probability:

$$H(p) = - \sum p(x) \log(p(x)) \tag{16}$$

For continuous distributions, only relative entropy can be calculated, which is defined as:

$$H_c(p) = - \int p(x) \log(p(x)) \tag{17}$$

Which is just the Kullback Leibler divergence for a (possibly improper) uniform prior and with the opposite sign [15]. Therefore maximizing the entropy amounts to minimizing the Kullback Leibler divergence. Table 1 summarizes common probability distributions and the additional constraints under which they can be shown to be of maximum entropy. The least informative prior distribution is the distribution for which the expected information gained per sample is minimized. Sometimes however the least informative prior depends on how the outcomes are measured, which is often not desirable. In this case Jeffreys [16] suggests to use the prior that minimizes the information under the constraint that it does not change under reparametrization. For a probability density $p(x)$ and given a possible smooth and continuous reparametrization $f(x)$, this is constraint is captured by the following equation:

$$p_X(x) = p_{(Y=f(x))}(f(x)) \left| \frac{df(x)}{dx} \right| \tag{18}$$

However, the debate is not settled on the right prior distributions in certain cases, especially when they are discrete. An example for this problem is finding an objective prior for positive integers, see section 3.4.1.

Many commonly encountered probability distributions can be derived from the principle of maximum

entropy. In fact most of the relevant distribution in this work are either maximum entropy distributions themselves or can be derived from maximum entropy distributions. Some of the following distributions are described more readily using the Heaviside step function, which is defined as:

$$\theta(x) \equiv \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{2} & \text{if } x = 0 \\ 1 & \text{if } x > 0 \end{cases} \tag{19}$$

The value at $x = 0$ is taken to be either $1/2$, or 0, 1, or both. The exact choice is inconsequential for everything that follows and so $1/2$ was chosen for the symmetry.

The fundamental constraint, that the probability density must not be negative and must integrate to 1, is not listed, because this is a constraint for all proper probability distributions in general.

The **uniform distribution** is the equal distribution of random variables within a range $[a, b]$.

The **exponential distribution** is the distribution of positive random variables with a finite average, which is parametrized by λ^{-1} [17, equation 19.1].

The **Laplace distribution** or double exponential distribution is the maximum entropy distribution for random variables with a given expected absolute deviation to a mean value μ. The expected absolute deviation is parametrized with b [18, equation 24.1].

The **Poisson distribution** is a discrete distribution with conditions analogous to the exponential distribution. But for large values of λ^{-1} it approaches a Gaussian distribution, which cannot be reached fully because the Gaussian distribution is continuous [7, Chap. 5].

The **Gaussian distribution** is encountered so frequently that it is also known as the *normal distribution*. Its instance with zero mean and unit variance is the *standard normal distribution*. It is the maximum entropy distribution for continuous values with finite mean μ and variance σ^2 [7, Chap. 5]. The reason why it is such a good approximation in so many cases is the central limit theorem, see section 3.5.

distribution	probability density	constraints
uniform	$\dfrac{\theta(x-a)\,\theta(b-x)}{b-a}$	$x \in \mathbb{R} \mid a \leq x \leq b$
exponential	$\lambda \exp(-\lambda x)$	$(x \in \mathbb{R} \mid x \geq 0) \wedge \left(E(X) = \lambda^{-1}\right)$
Laplace	$\dfrac{1}{2b} \exp\left(-\dfrac{|x-\mu|}{b}\right)$	$(x \in \mathbb{R}) \wedge (E(|x-\mu|) = b)$
Poisson	$\dfrac{\lambda^k \exp(-\lambda)}{k!}$	$(k \in \mathbb{N} \mid k \geq 0) \wedge \left(E(K) = \lambda^{-1}\right)$
Gaussian	$\dfrac{\exp\left(-\frac{1}{2}(x-\mu)^2 \sigma^{-2}\right)}{\sqrt{2\pi\sigma^2}}$	$(x \in \mathbb{R}) \wedge (E(x) = \mu) \wedge \left(\text{Var}(x) = \sigma^2\right)$

Table 1: Common probability distributions that can be derived by maximizing the entropy under simple constraints.

3.1.3. Other common probability distributions

In the previous section some common distributions were introduced with the maximum entropy principle. Most probability distributions are maximum entropy distributions in some way, but the constraints under which they are, are often not practical. These distributions arise more naturally from different assumptions, for example by inverting probability distributions via Bayes theorem or by combining random variables of other distributions.

The **Dirac delta distribution** is the most trivial distribution, describing absolute certainty that the random variable X can only be equal to x_0 [19, Chap. III]. It is defined as the function that integrates to 1 over all sets containing x_0 and assumes the value 0 everywhere but at $x = x_0$. Because this definition is difficult to express mathematically without infinite limits, the following definition uses the Heaviside step function instead:

$$\delta\left(x - x_0\right) \equiv \frac{d\theta\left(x - x_0\right)}{dx} \tag{20}$$

While there is a condition under which the **Cauchy distribution** is the maximum entropy distribution[1], it is more typically derived from other distributions or other principles. It arises from the ratio of two normal random variables or as the Fourier transform (see section 4) of exponential or doubly exponential distributions. The probability density of the Cauchy distribution is [17, equation 16.1]:

$$p\left(x, x_0, \gamma\right) = \frac{1}{\pi\gamma\left(1 + \left(\dfrac{x - x_0}{\gamma}\right)^2\right)} \tag{21}$$

The mean is undefined and the variance is infinite. Mode and median are equal to x_0 and the shape parameter γ is equal to the expected median of absolute deviations of samples to x_0 [17, Chap. 16].

Another distribution that arises from composing normal variables is the χ^2-**distribution**. It is the sum of squares of k independent and identically distributed random variables following a standard normal distribution[2] [7, Chap.12]. Its probability density is given by:

$$p\left(x, k\right) = \frac{x^{k/2-1}\exp\left(-\frac{x}{2}\right)}{2^{k/2}\Gamma\left(\frac{k}{2}\right)} \tag{22}$$

Because the mean of each of the squared standard normal variables in the sum is 1, the mean of the sum of k variables is k (equation 10). And similarly, because the variance of each standard normal variable is 1, the variance of the square is 2, because it is the product of two perfectly correlated random variables (equation 13). The variance of χ^2-distributed variables is the variance of the sum of k independent random variables each with variance 2, which is $2k$ (equation 11).

The χ^2-distribution, like the exponential distribution is a special case of the **Gamma-distribution**. It can be derived as the distribution for the sum of α exponential random variables with the rate parameter β.

---

[1] $\mathrm{E}\left[\log\left(1 + (x - x_0)^2\right)\right] = 2\log\left(1 + \sqrt{\gamma}\right)$

[2] But it can also be seen as a maximum entropy distribution with the following two constraints: $\mathrm{E}\left(X\right) = k$ and $\mathrm{E}\left(\log\left(X\right)\right) = \psi\left(k/2\right) + \log\left(2\right)$

The probability density is given by [17, equation 17.1]:

$$p\left(x, \alpha, \beta\right) = \frac{\exp\left(\alpha \log\left(\beta\right) + \left(\alpha - 1\right)\log\left(x\right) - \beta x\right)}{\Gamma\left(\alpha\right)} \tag{23}$$

The mean is α/β and the variance α/β^2. It is the conjugate for the Poisson distribution, see sections 6.6 and 3.3, and the conjugate for the exponential distribution also [20].

3.2. Maximum Likelihood Estimation

Given a probabilistic model, a probability distribution that depends on parameters, we know what outcomes to expect. But if we don't know the parameters, because they are hidden, i.e. not directly available to us, one straightforward way to determine them is maximum likelihood estimation [7, Chap. 17]. The parameters that maximize the likelihood function are the maximum likelihood estimate and represent the most likely explanation for the observed data given a model and that every value of the parameter is equally likely. The Likelihood function is the probability density of the random variable viewed as a function of the parameter. With respect to other parameters than the random variable, a probability density is not a probability density any more, at least in general, and it might not even be integrable. Joint likelihoods of uncorrelated observations are computed by multiplying them like probabilities.

When expressed in terms of parameters other than the random variable, this likelihood function is not necessarily a probability density and might not even be integrable.

Because the logarithm is monotonic for positive real numbers, and likelihoods for observed data, like probability densities, can only be positive, the minimum of the negative logarithm of the likelihood target is equal to the maximum of the likelihood target $\log\left(\mathcal{L}\right)$.

$$\arg\max_{\theta}\left(\prod_{i}^{N}\mathcal{L}\left(X_i|\theta\right)\right) = \arg\min_{\theta}\left(-\sum_{i}^{N}\log\left(\mathcal{L}\left(X_i|\theta\right)\right)\right)$$

The (negative) log-likelihood has two major computational advantages, firstly it is numerically more stable, because additions are computationally more stable than multiplications, and secondly it transforms Gaussian-like likelihood functions into quadratic or approximately quadratic equations, which are much easier to optimize for. Additionally, for Gaussian-like likelihoods, the second order derivative (the Hessian matrix) of the negative log-likelihood conveniently is an approximation for the inverse covariance of the parameters. Here the Gaussian distribution is generalized to higher dimensions, see section 3.8 for details, hence the vectorial parameters $\vec{\mu}$ and the parameter matrix $\mathbf{\Sigma}$.

$$\frac{\partial^2}{\partial\vec{\mu}^2}\left(-\log\left(\exp\left(-\frac{1}{2}\left(\left(\vec{x} - \vec{\mu}\right)^{\top}\mathbf{\Sigma}^{-1}\left(\vec{x} - \vec{\mu}\right) + \log\left(|2\pi\mathbf{\Sigma}|\right)\right)\right)\right)\right) = \mathbf{\Sigma}^{-1}$$

There are however several pitfalls with this method. The maximum likelihood estimate is often biased, that is systematically lower or higher than the true parameters, because the mode and mean don't necessarily align [7, Chap. 17]. Also it has no sound way to select between different models, because often the model with more parameters can give a better fit to the data, and therefore a higher likelihood, but this does not necessarily mean that simply adding parameters leads to a better model. This is known as overfitting and can be tested for by splitting data into two sets, one to fit the data to and another to test how good the

model works on new data [8, Chap. 7]. So long the constraint ratio, the number of data points in relation to the number of parameters, is high, these problems are not that severe. If we know in advance that some values are more likely for the parameters, or if the distribution of parameters is relevant, or non Gaussian, and not narrow, a better way to infer hidden parameters from observations is using Bayes theorem in the following section.

3.3. Bayes Theorem

Bayes theorem relates the probability of an outcome A given another outcome B to the probability of B given A [8, Chap. 1]:

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(B|A)\mathrm{P}(A)}{\mathrm{P}(B)} \tag{24}$$

While this equation is trivially true, it takes on an specific meaning in Bayesian statistics. The fact that it is true can be shown by rearranging the two alternative ways of expressing the probability of A and B happening simultaneously ($\wedge$ is the logical and operator).

$$\mathrm{P}(A \wedge B) = \mathrm{P}(A)P(A|B)$$
$$\mathrm{P}(A \wedge B) = \mathrm{P}(B)\mathrm{P}(B|A)$$
$$\rightarrow \mathrm{P}(A|B) = \frac{P(B)\mathrm{P}(B|A)}{\mathrm{P}(A)}$$

Equation 24 is Bayes theorem and it can be used to update beliefs from observations and therefore can almost serve as the basis of experimental science itself. It gives a clear method of inferring knowledge about parameters or models from knowledge about outcomes and vice versa. Given a probability density $\mathrm{p}_X(x|a)$ that depends on a parameter a, and an outcome X, Bayes theorem allows us to view the parameter itself as a random variable. The probability density of the random variable A given the observation $X = y$ can then be inferred with [8, equation 1.1]:

$$\mathrm{p}_A(a|X = y) = \frac{\mathrm{p}_X(y|a)\mathrm{p}_A(a)}{\mathrm{p}_X(y)} = \frac{\mathrm{p}_X(y|a)\mathrm{p}_A(a)}{\int \mathrm{p}_X(o|a)\mathrm{p}_A(a)da} \tag{25}$$

This is the posterior distribution of A, in a way it turns the likelihood function $\mathrm{p}_X(y|a)$ into a proper probability distribution, thereby forming a *conjugate* distribution to the first distribution as a function of the random variable X. $\mathrm{p}_A(a)$ and $\mathrm{p}_X(x)$ are the prior distributions of A and X respectively, they contain the beliefs prior to the observation. The new assumption about the distribution of X given the observation can be formed by integrating over all possible values of the parameter weighted by its posterior probability, this is the marginalization integral [8, equation 1.4]:

$$\mathrm{p}_X(x) = \int \mathrm{p}_A(a|X = y)\mathrm{p}_X(x|a)da \tag{26}$$

A point of contention often are the probabilities $\mathrm{P}(A)$ and $\mathrm{P}(B)$, and how they are determined in the absence of prior information. This can be resolved by choosing the prior probability distributions that represent the absolute minimum of information as measured by the Kullback-Leibler divergence D_{KL} or

relative entropy:

$$D_{KL}\left(P \parallel Q\right) = \int_{\mathbb{R}} P(x) \log\left(\frac{P(x)}{Q(x)}\right) dx \tag{27}$$

The Kullback Leibler Divergence D_{KL} [15] is an asymmetric measure for the difference between a probability distribution and a reference probability distribution. It quantifies the information gained in a Bayesian update from a prior distribution $q(x)$ to a posterior distribution by the likelihood function $p(x)$. Equivalently it quantifies the additional information required when encoding values of the distribution in $p(x)$ when assuming they are distributed like $q(x)$. Priors that minimize the Kullback Leibler Divergence are known as minimum-information priors and maximum-entropy distributions, see section 3.1.2 [8, Chap. 2.8]. But the major problem of Bayesian induction is the complexity of expressions arising after recursively applying Bayes theorem multiple times or when the underlying relationships are more complicated. In that case full Bayesian inference often becomes infeasible. Approximate methods have been developed and the principle of minimum description length in section 3.4 and maximum a posterior methods just determine the most probable parameters, which is a direct improvement over the principle of maximum likelihood of the previous section.

3.4. Maximum a Posterior and Minimum Description Length Principle

When full Bayesian inference is too tedious, and one still might want to choose the best set of parameters and model, a good compromise and improvement over the principle of maximum likelihood, is the maximum a posterior target and the principle of minimum description length. The principle of minimum description length stipulates that the model, which leads to the shortest description of the observed data including a description of the model itself, in any sufficiently complex language of description, is the best model, because it is the most probable interpretation of the data [21]. This is a generalization of Occam's razor. No further specification of this description language is necessary, because the model with the shortest description in one language is almost identical to the model with the shortest description in some other language, given that both are Turing complete. Any programming language therefore can serve as a stand-in.

Because the length of a sequence is inversely proportional to its probability, the two concepts are equivalent. Bayes principle of multiplying priors and likelihood functions then corresponds to the code length for the model plus the code length for the observed data given the model, respectively. Finding the shortest sequence therefore amounts to finding the model and parameters with the highest posterior probability [21]. Despite the redundancy between these two concepts, the minimum description length principle can be a mental crutch when the other concept seems to fail.

3.4.1. A Prior for Natural Numbers?

Discrete distributions often are uniquely challenging from a Bayesian perspective, because discrete outcomes seem to only justify uniform prior distributions according to the principle of indifference. A uniform prior distribution cannot be proper, because its sum would diverge. Intuitively we might suspect that smaller numbers are more frequent than larger numbers, but by how much? A prior distribution for natural numbers may seem like an outlandish idea, but it is actually useful when modelling outlier distributions.

Outliers, by definition, are those values about which we know the least, and outliers are never the main focus of the analyis. From this it follows that the distribution that must be assumed for outliers is precisely the minimum information prior. Simply not modelling their distribution and ignoring outliers would mean that they will have an outsized effect on the location of the maximum for most likelihood functions. Thus, the following derivation will be used in section 6.9.8 to make a maximum-likelihood approach more robust, by allowing for otherwise unforseen detector readings.

According to the principles laid out by Jeffreys [22, Chap. 4.8, equation 1], the prior distribution for positive integers is proportional to:

$$\mathrm{P}(n|H) \propto n^{-1} + O\left(n^{-2}\right) \tag{28}$$

Without knowing what is behind the term $O(n^{-2})$, this is still not a proper prior distribution, but this does not have to be a problem, because it is still σ-finite[3] and therefore perfectly fine for Bayesian inference[23]. However, if we want a proper prior distribution for natural numbers we might look to different concepts, like the minimum description length principle, because finite encoding guarantees densities with finite integrals. To encode a natural number in any positional numeral system with base larger than one takes a logarithmic amount of digits at a minimum, implying that the universal prior probability of a natural number can be at most $1/n$. In addition to the number itself, the code needs to encode the length of the code, with length proportional to $\lceil \log(n) \rceil$ and so on recursively until the number to encode is less than one. There are so-called universal codes based on this concept, and they imply a probability density approximately and proportional to:

$$u(x) = \begin{cases} \dfrac{u(\log(x))}{x} & \text{if } x > 1 \\ 1 & \text{otherwise} \end{cases}$$

This is the argument laid out by Rissanen [24], but instead based on binary logarithms. The original definition is as follows:

$$\log_2^*(x) = \begin{cases} \log_2(x) + \log_2^*\left(\log_2(x)\right) & \text{if } x > 1 \\ 0 & \text{otherwise} \end{cases}$$

$$u(x) = 2^{-\log_2^*(x)}$$

However, this distribution is still not normalized and it would be preferable if it were smooth. A smoother version can be derived with analytic continuation:

$$u(x) = \begin{cases} \dfrac{u\left(\log_2(x)\right)}{x} & \text{if } x > 1 \\ 0.29 \left[1 - \dfrac{\left(x - \frac{1}{2}\right)^2 - \frac{1}{4}}{1 + 1/\log(2)} \right] & \text{otherwise} \end{cases} \tag{29}$$

But we need to remember that this is still just an approximation to something that might not even exist, because it is not known if all objective priors need to be proper priors.

---

[2] Big-O denotes the asymptotic behavior of a function for large inputs. $O(f(x))$ is the class of functions that are bounded by a constant multiple of $f(x)$ for all possible inputs.

[3] Every integral over any subset is finite.

3.5. Sample Mean and Variance

The mean of a distribution is the first moment of the distribution. The sample mean is an estimator for the mean from samples, essentially the Monte-Carlo integral equivalent to the defining integral of the expected value in equation 1 [25, Chap. 17]. Estimators are not derived from a Bayesian framework, but like the maximum a posterior method (section 3.4) they can be a good approximation for parameter estimation. In analogy to the sample mean, the variance is the second central moment of the distribution and the sample variance is an estimator of the variance of the distribution based on the average squared deviations of the samples to the sample mean μ_S:

$$\mu_s = \frac{\sum_{i=1}^{N} x_i}{N} \tag{30}$$

$$\sigma_s^2 \approx \frac{\sum_{i=1}^{N} (\mu_s - x_i)^2}{N}$$

This estimator for the variance is too low on average however, because the sample mean used in the formula is only an estimate and biased towards the sample. Bessel's correction term $\frac{N}{N-1}$ is used to correct this bias due to the one degree of freedom that is taken up by the estimate of the mean. Using this correction, an unbiased estimate for the sample variance can be formed [25, equation 17.10]:

$$\sigma_s^2 \approx \frac{\sum_{i=1}^{N} (x_i - \mu_s)^2}{N - 1} \tag{31}$$

The central limit theorem states that sums of independent random variables with finite variance approach a Gaussian distribution with the variance given by the sum of individual variances. Accordingly, the average of several random variables with finite variance also approaches a Gaussian distribution [7, Chap. 7.32]. Assuming the sum in the nominator of the fraction in equation 30 contains independent samples, their variances add linearly. The division of the sum by the sample size corresponds to a reduction of the variance by a factor of one over the sample size squared. Therefore the factor N cancels out and the estimate of the variance of the average is:

$$\sigma_{\mu_s}^2 \approx \frac{\sum_{i=1}^{N} (x_i - \mu_s)^2}{N (N - 1)} \tag{32}$$

For non-equal samples the sample mean and sample variance can be generalized to weighted mean and weighted sample variance [26, Chap. 4]. The weighted mean and the weighted sample variance are defined as follows:

$$_w\mu_s = \frac{\sum_{i=1}^{N} w_i x_i}{\sum_{i=1}^{N} w_i} \tag{33}$$

$$_w\sigma_s^2 \approx \frac{\sum_{i=1}^{N} w_i (x_i - {_w\mu_s})^2}{\sum_{i=1}^{N} w_i} \tag{34}$$

The weighted mean is employed when some samples are more important, more reliable, encompass more samples of the same kind, or are more precise, all these aspects can be expressed as different weights for the samples. If the samples are independent of each other, the variance of the weighted mean can be seen

as the sum of variances of the weighted samples normalized by the sum of weights:

$$\sigma^2_{w\mu_s} = \frac{\sum_{i=1}^{N} w_i^2 \sigma^2_{x_i}}{\left(\sum_{i=1}^{N} w_i\right)^2} \tag{35}$$

An estimator with lower variance is more precise, and therefore we seek to minimize the variance in most cases. This is a simple optimization problem with the well-known result that weights inversely proportional to the variance of the samples minimize the variance of the weighted mean. It is no coincidence that the weighted mean and variance calculation is analogous to multiplying Gaussian distributions and the result is a distribution which represents the overlap between the individual distributions [26, Chap. 4]. This result is the main motivation behind most applications of the weighted mean. If the weights are exactly reciprocal to the variance of the samples, the variance of the weighted mean is exactly the reciprocal of the sum of weights:

$$\left(w_i = \frac{1}{\sigma^2_{x_i}}\right) \rightarrow \left(\sigma^2_{w\mu_s} = \frac{1}{\sum_{i=1}^{N} w_i}\right) \tag{36}$$

If the weights are not exactly reciprocal to the variance of the samples or have some other meaning, estimating the variance of the mean gets more complicated. Still there are some estimators worth noting, which will be introduced in the following.

When the weights are observation frequencies, the variance estimate of the weighted mean can be derived from the variance estimate of the unweighted mean in equation 32.

$$\left(w_i = \text{frequencies}\right) \rightarrow \left(\sigma^2_{w\mu_s} \approx \frac{\left(\sum_{i=1}^{N} w_i \left(x_i - {}_w\mu_s\right)^2\right)}{\left(\sum_{i=1}^{N} w_i\right)^2 - \left(\sum_{i=1}^{N} w_i\right)}\right) \tag{37}$$

The weights only need to be inversely proportional to the variance of the samples in order to minimize the variance of the weighted mean. If they are proportional, a maximum likelihood approach can yield an estimate for the proportionality constant. An easier way of estimating it is to compare the observed variance with the value that should be proportional. The sum of weights, which is inversely proportional to the variance times the squared differences between the samples and the mean will be distributed like a Chi-squared distribution. The exact distribution is of no further relevance, but the expected value of this distribution is the proportionality constant. An estimate for the variance of the weighted mean with weights inversely proportional to the variance can be found by cancelling out the proportionality constant from the sum of weights by dividing it by said proportionality constant. The estimate will also be applicable for frequency weights (proportional to the observation frequencies), because observation frequencies are inversely proportional to the variance. With Bessel's correction this leads us to a very similar expression to equation 37.

$$\sigma^2_{w\mu_s} \approx \frac{\left(\sum_{i=1}^{N} w_i \left(x_i - {}_w\mu_s\right)^2\right)}{\left(\sum_{i=1}^{N} w_i\right)(N-1)} \tag{38}$$

However like that, this estimate can be biased and too low, because the uneven weights effectively decrease

the sample size and the proportionality constant constitutes another degree of freedom. The effective sample size of the weighted mean can be estimated using the ratio between the square of the sum of weights and the sum of squared weights according to Kish [27, Chap. 8.2] and it can be used analogously to the sample size also in Bessel's correction term to correct the bias in the weighted sample variance:

$$N_{\text{eff}} = \frac{\left(\sum_{i=1}^N w_i\right)^2}{\left(\sum_{i=1}^N w_i^2\right)} \tag{39}$$

$$_w\sigma_s^2 \approx \frac{N_{\text{eff}}}{N_{\text{eff}} - 1} \frac{\sum_{i=1}^N w_i \left(x_i - {_w}\mu_s\right)^2}{\sum_{i=1}^N w_i} \tag{40}$$

The effective sample size is equal to the sample size when all weights are equal and smaller when they are uneven. This estimate is valid for all of the following cases where the true sample size of the weighted average is unknown, independent of the meaning of the weights. From the weighted sample variance and the effective sample size the variance of the weighted mean can then be estimated as follows [26, equation 4.23]:

$$\begin{matrix} (w_i \propto \text{frequencies}) \\ \left(w_i \propto \sigma_{x_i}^{-2}\right) \end{matrix} \rightarrow \left(\sigma_{_w\mu_s}^2 \approx \frac{\left(\sum_{i=1}^N w_i \left(x_i - {_w}\mu_s\right)^2\right) \left(\sum_{i=1}^N w_i^2\right)}{\left(\sum_{i=1}^N w_i\right) \left(\left(\sum_{i=1}^N w_i\right)^2 - \left(\sum_{i=1}^N w_i^2\right)\right)} \right) \tag{41}$$

When the weights are independent of the variance of the samples, and all samples are identically distributed, the variance of the weighted mean can be estimated from the unweighted sample variance and the effective sample size. This estimator has very limited application, because the unweighted average would be a better estimate for the mean in this case. This also explains why the unweighted mean appears in this estimate:

$$\sigma_{_w\mu_s}^2 \approx \frac{\left(\sum_{i=1}^N w_i^2\right)}{\left(\sum_{i=1}^N w_i\right)^2} \frac{\left(\sum_{i=1}^N \left(x_i - \mu_s\right)^2\right)}{N - 1} \tag{42}$$

Endlich et al. [28] referencing Kish [27] estimate the variance of the weighted mean similarly to equation 41 but instead from the square weighted sample variance. The sum of square weighted deviations is not immediately obvious in Endlich et al. [28] and Gatz and Smith [29], but it can be derived by symbolic simplification. Then replacing the sample size with the effective sample size we get this estimator:

$$\sigma_{_w\mu_s}^2 \approx \frac{\sum_{i=1}^N w_i^2 \left(x_i - {_w}\mu_s\right)^2}{\left(\sum_{i=1}^N w_i\right)^2 - \left(\sum_{i=1}^N w_i^2\right)} \tag{43}$$

This estimator was not derived from inverse variance weights, but from weights that are proportional to the sampling probability. However, as can be seen in figure 1, it performs well in this case too. All estimates are identical when all samples are weighted equally, i.e. all weights are the same. The higher the variance of the weights, the more the different estimates deviate from each other. The expected worst case therefore is that the weights are derived from inverse variance estimates when the variance estimates themselves
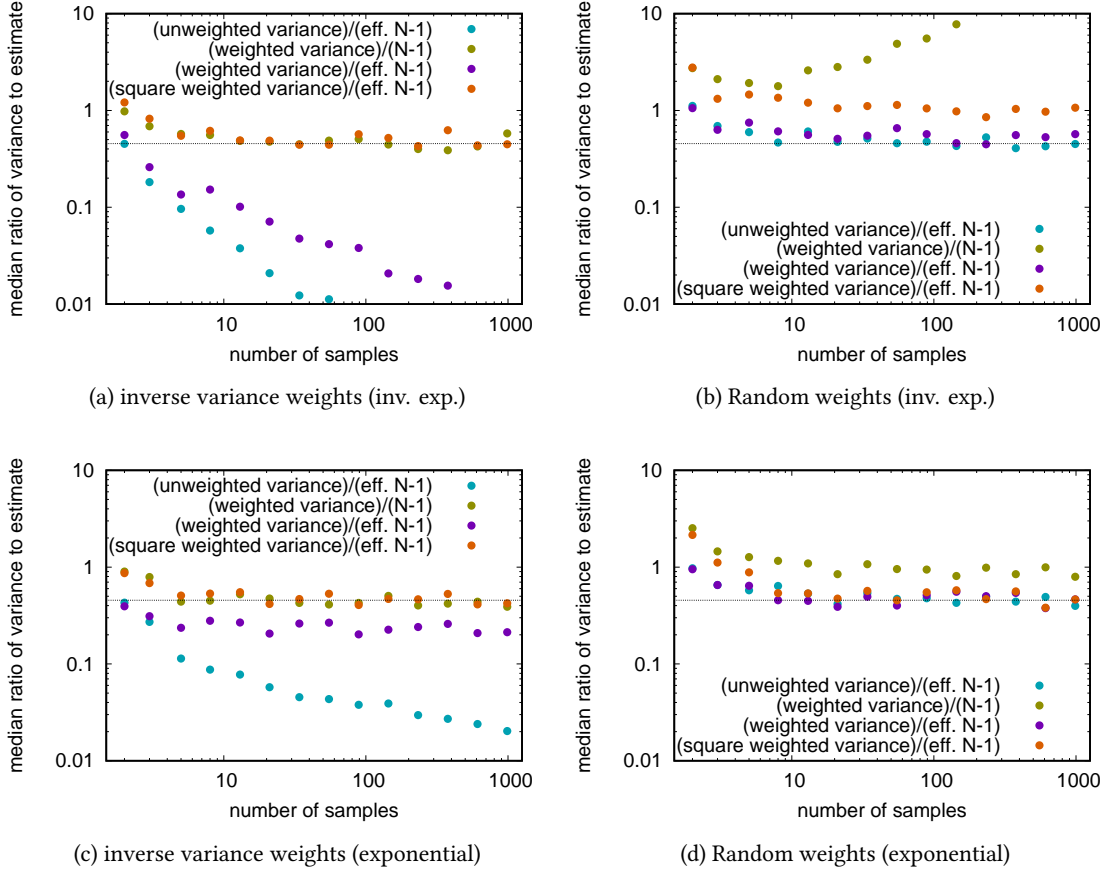
Figure 1: Comparison of the different variance estimators using the median ratio of squared deviation to variance estimate in 255 numerical experiments. In figure (a) the samples have exponentially distributed variance and the weights are set as inversely proportional. In figure (b) the samples have unit variance, but the weights are kept the same as for figure (a), that is they are distributed as inversely exponential. Figure (c) and (d) have exponentially distributed weights, in (c) the variances are set accordingly and in (d) the variance of the samples is kept constant. Ideally the median ratio of estimated variance to actual variance would be close to $\left(\sqrt{2}\operatorname{erf}^{-1}\left(1/2\right)\right)^2 \approx 0.45$ (indicated with the black dotted line), but when the assumptions about the weights are not met, the estimates can deviate significantly. As can be seen, the estimators are sensitive to the nature of the weights.

are exponentially distributed, because the exponential distribution is the minimum entropy distribution for a positive variable with finite mean. The inverse exponential distribution leads to much more disparate weights and worse estimates from all estimators compared to when the weights themselves are distributed exponentially. Four numerical experiments are shown in figure 1. Figures (a) and (b) show the expected worst case. In (a) the weights are inversely proportional to the actual variance of the samples and in (b) the samples have a fixed variance, but the weights are still distributed in the same way as in the first case. Figures (c) and (d) show more well-behaved (and more representative for crystallographic data) examples where the weights follow an exponential distribution, again once in agreement with the actual variance of the samples and once independent. There is no one formula that directly and accurately estimates the variance of the weighted mean for arbitrary weights, but the estimator using the sum of square weighted squared deviations (equation 43) gives the best results in most circumstances. And none of the formulae estimate the variance nearly as well as the sample variance estimates the variance of the equally weighted mean. One potential solution are bootstrapping methods [29], based upon subsampling and computing the variance between the samples, if there are enough samples to subsample different subsets. Preferably the variance of the individual samples can be understood better or possibly modelled. More meaningful weights would not only improve the variance estimate using equation 36, but also the estimate of the mean.

For the computation of the weighted mean, the weighted sample variance and the square weighted sample variance a more numerically stable way is preferred over the literal computation of the previous forumlae. This way also allows incremental computation in a single pass. For each new sample the sum of weights (equation 44), the weighted mean $_w\mu_s$, the weighted variance $_w\sigma_s^2$, the sum of squared weights (equation 47), the square weighted mean $_{w^2}\mu_s$ and the square weighted average of squared deviations from the square weighted mean $_{w^2}\sigma_s^2$ is updated from the previous value. The estimators based on the weighted variance and square weighted variance can be trivially computed on demand from these values. The only non-trivial case, the square weighted variance, is given in equation 50. The following algorithm is adapted from West [30] and Hanson [31]. Equations 44 and 45 are identical to West [30]. Equation 46 is modified to compute the weighted variance instead of the sum of squared deviations directly just like equation 45 does for the mean and then simplified using the identity 51. Equations 47, 48 and 49 apply the same algorithm but for squared weights.

$$\left(\sum_{j=1}^{i+1} w_j \right) = w_{i+1} + \left(\sum_{j=1}^{i} w_j \right) \tag{44}$$

$$_w\mu_{s\,i+1} = {}_w\mu_{s\,i} + (x_i - {}_w\mu_{s\,i}) \frac{w_i}{\left(\sum_{j=1}^{i+1} w_j \right)} \tag{45}$$

$$_w\sigma_{s\,i+1}^2 = {}_w\sigma_{s\,i}^2 + \left((x_i - {}_w\mu_{s\,i+1})(x_i - {}_w\mu_{s\,i}) - {}_w\sigma_i^2 \right) \frac{w}{\sum_{j=1}^{i+1} w_j} \tag{46}$$

17

$$\left(\sum_{j=1}^{i+1} w_j^2 \right) = w_i^2 + \left(\sum_{j=1}^{i} w_j^2 \right) \tag{47}$$

$$_{w^2}\mu_{s\,i+1} = {}_{w^2}\mu_{s\,i} + \left(x_i - {}_{w^2}\mu_{s\,i} \right) \frac{w_i^2}{\left(\sum_{j=1}^{i+1} w_j^2 \right)} \tag{48}$$

$$_{w^2}\sigma_{s\,i+1}^2 = {}_{w^2}\sigma_{s\,i}^2 + \left(\left(x_i - {}_{w^2}\mu_{s\,i+1} \right)\left(x_i - {}_{w^2}\mu_{s\,i} \right) - {}_{w^2}\sigma_{s\,i}^2 \right) \frac{w_i^2}{\sum_{j=1}^{i+1} w_j^2} \tag{49}$$

$$\frac{\sum_{i=1}^{N} w_i^2 \left(x_i - {}_w\mu_s \right)^2}{\left(\sum_{i=1}^{N} w_i^2 \right)} = {}_{w^2}\sigma_s^2 + \left({}_w\mu_s - {}_{w^2}\mu_s \right)^2 \tag{50}$$

There is an alternative to equation 46 (and similarly to equation 49) to compute the weighted variance that does not depend on different values of the mean estimate. This involves two more multiplications, but it is more amenable to symbolic differentiation and it avoids a common mistake in implementations of this algorithm.

$$\left(x_i - {}_w\mu_{s\,i+1} \right)\left(x_i - {}_w\mu_{s\,i} \right) = \frac{\sum_{j=1}^{i} w_j}{\sum_{j=1}^{i+1} w_j} \left(x_i - {}_w\mu_{s\,i} \right)^2 \tag{51}$$

3.6. Empirical Distribution Function and Empirical Quantile Function

Given random samples drawn from a distribution, an approximation to the cumulative distribution function and quantile function can be assembled by sorting the samples. The proportion of samples less than or equal a value x will converge to the value of the cumulative distribution at the value x. The inverse empirical distribution function is the empirical quantile function, which is an estimate for the proportion of values less than a value x, based on the proportion observed so far. The estimate of the value of the q- quantile from a sample has variance [7, equation 9.22]:

$$\frac{q\left(1 - q \right)}{N p \left(P^{-1}\left(q \right) \right)} \tag{52}$$

The variance depends on the proportion q in question, decreases linearly with the number of samples N, and the probability density of the value delimiting the quantile $p\left(P^{-1}\left(q \right) \right)$.

3.7. Sample Median and Median Absolute Deviation

The median value of a distribution is the value where the cumulative distribution equals $1/2$, meaning values less than or equal the median are just as likely as values larger than or equal [7, Chap. 2.8]. Just like the mean of a distribution can be estimated with the sample average, the median of a distribution can be estimated with the sample median. When there is an even number of samples, all values between and including the two middle most values are median values according to the definition. This ambiguity is usually resolved by taking the average of the two middle samples. And like there is the central limit theorem for the median, there is the median theorem for the sample median: The median $\tilde{\mu}_s$ of a sample of size N drawn from a distribution with density $p(x)$ approaches a Gaussian distribution with mean $\tilde{\mu}$ and variance $\left(4N\left(p\left(\tilde{\mu}\right)\right)^2\right)^{-1}$ for large sample sizes, if the density around the median of the distribution is continuously differentiable and non-zero [32, Supp. mat.] [33].

$$\sigma_{\tilde{\mu}_s}^2 = \frac{1}{4N\left(p\left(\tilde{\mu}\right)\right)^2} \tag{53}$$

This is a special case of the variance of quartile estimates in equation 52. The nonzero and continuously differentiable density condition of the median theorem takes the place of the finite variance as a precondition for the convergence of the average in the central limit theorem. Equation 53 also allows a direct comparison between the efficiency of the median and the mean given a distribution that the samples are assumed to follow, for example Gaussian, Laplacian and uniform [25, Chap.17]. For the same number of samples with the same variance, the variance of the median of Gaussian samples is $\frac{\pi}{2} \approx 1.57$ times larger than the respective mean. If the samples come from a Laplacian distribution the median is just as efficient as the mean. The median of uniform samples has three times higher variance than the mean. Despite this apparent inferiority in common scenarios the median is very useful because it is robust with respect to outliers and converges for distributions with infinite variance like the Cauchy distribution, where the sample average does not converge. Outliers are samples with such extreme deviation from the median that it is as if they come from a different distribution entirely. Just one sample with extreme deviation from the mean will lead to an entirely different average, but the median will not be affected nearly as much, up the point where just less than half of the samples are outliers. This is why often, instead of the mean, the median will be used to give a more typical result that is not skewed by a few extreme cases. Just like the sample variance estimates the variance of the samples, the median of absolute deviations from the sample median and the interquartile range are estimates of the typical difference of a sample to the median and it is a robust alternative (although slightly biased depending on the distribution) [7, Chap.2]. To estimate the standard deviation of the sample median of samples from an unknown distribution, Woodruff[34] suggests using the following formula:

$$\sigma_{\tilde{\mu}_s} \approx \frac{x_{\left\lceil \frac{1}{2}\left(N+\sqrt{N}\right)\right\rceil} - x_{\left\lfloor \frac{1}{2}\left(N-\sqrt{N}\right)\right\rfloor}}{2} \tag{54}$$

Because the median, by definition, is the value for which half the values are less than or equal and the other half is greater than or equal, the probability of a value at position i in a ordered list of random samples being closest to the true median follows a binomial distribution with a success probability p of $1/2$. Therefore the values at these positions correspond to $\tilde{\mu} + \sigma_{\tilde{\mu}}$ and $\tilde{\mu} - \sigma_{\tilde{\mu}}$ on average respectively, and their difference, on average, is about two times the standard deviation. Knowing that the probabilities follow a binomial

distribution, one can calculate a weighted squared deviation of the random samples to the sample median, to improve the accuracy of the variance estimate slightly at the cost of robustness:

$$\sigma_{\tilde{\mu}_s}^2 \approx \sum_{i=0}^{N-1} \binom{N-1}{i} \left(\frac{1}{2}\right)^{1-N} (x_i - \tilde{\mu}_s)^2 \tag{55}$$

Bootstrapping methods are available too.

Just like for the sample mean, the sample median can also be generalized to include weights. The weighted sample median $_w\tilde{\mu}_s$ is any value which partitions the weights associated with values less than or equal and the weights of the values larger than or equal so their sums differ the least:

$$_w\tilde{\mu}_s = \arg\min_{w\tilde{\mu}_s} \left| \left(\sum_{\{i|x_i \leq {_w\tilde{\mu}_s}\}} w_i \right) - \left(\sum_{\{i|x_i \geq {_w\tilde{\mu}_s}\}} w_i \right) \right| \tag{56}$$

Or alternatively, if the reader is more familiar with boolean expressions than with set expressions:

$$_w\tilde{\mu}_s = \arg\min_{w\tilde{\mu}_s} \left| \left(\sum_i^N w_i \left(x_i \leq {_w\tilde{\mu}_s}\right) \right) - \left(\sum_i^N w_i \left(x_i \geq {_w\tilde{\mu}_s}\right) \right) \right| \tag{57}$$

The previous two formulations are useful, because they show the partitioning explicitly, but the following expresses the weighted median more compactly:

$$_w\tilde{\mu}_s = \arg\min_{w\tilde{\mu}_s} \sum_i^N w_i \left| x_i - {_w\tilde{\mu}_s} \right| \tag{58}$$

As for the non-weighted median, ambiguities can arise, which have to be resolved consistently, for example using the average of the two values. However, no literature could be found on how to set the weights in order to minimize the variance of the weighted median - unlike for the weighted average, where it is well known that inverse variance weights minimize the variance of the weighted average. That is why in the following section the optimal weights for the best convergence of the weighted median will be derived.

Figure 2: Relative weighting between samples following a Gaussian or uniform distribution with identical variance each. The ratio of the Gaussian density to the uniform density at the median is $\sqrt{6/\pi} \approx 1.38$, this ratio is reached at around 0.58 on the x-axis, coinciding with the minimum of mean squared deviations of the weighted sample median to the actual median.
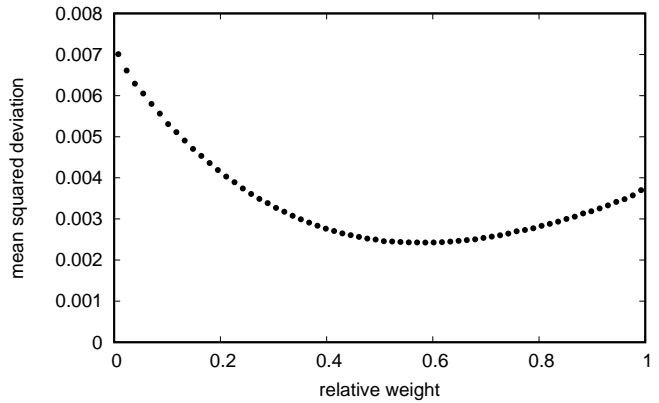
3.7.1. Optimally Weighted Median

We assume all samples x_i can follow a different probability distribution $p_i(x)$, otherwise different weighting would have only very limited application and the non-weighted median would almost always be the preferred choice. Consider a sorted list of weighted samples and the running sum of weights normalized by the sum of all weights. This computes the empirical cumulative distribution of the weighted mixture distribution, let the weighted cumulative distribution be $_wP(x)$.

$$
_wP(x) = \frac{\left(\sum_i^N w_i \int_{-\infty}^x p_i(y)\,dy \right)}{\left(\sum_i^N w_i \right)}
$$

Now consider the sum of weights associated with samples less than or equal to the true weighted median $_w\tilde{\mu}$, normalized by the sum of all weights, let it be c:

$$
c = \frac{\left(\sum_{\{i \mid x_i \le {_w\tilde{\mu}}\}} w_i \right)}{\left(\sum_i^N w_i \right)}
$$

Without knowing the median of each distribution, each weight has probability $1/2$ of being part of this sum or its counterpart. The variance introduced to the nominator by each weight is therefore $1/4$ the weights squared and the variance of c is approximated by the sum of individual variances divided by the normalization factor squared:

$$
\sigma_c^2 = \frac{\left(\sum_i^N w_i^2 \right)}{4 \left(\sum_i^N w_i \right)^2}
$$

The weighted sample median $_w\tilde{\mu}_s$ is the sample where the empirical weighted cumulative distribution reaches $1/2$. The expected value of c is $1/2$ and its variance is the mean squared deviation between the empirical partitioning which determines the weighted sample median and the partitioning which would lead to the value closest to the weighted population median. Therefore the value where the weighted cumulative distribution reaches c has the same variance as the weighted median. Error propagation demands the derivative of $_wP^{-1}$ at $1/2$ which is inverse to the derivative of $_wP$ at the weighted population median.

$$
\sigma_{w\tilde{\mu}_s}^2 = \sigma_c^2 \frac{d_wP^{-1}}{dx}(c)
$$

$$
\sigma_{w\tilde{\mu}_s}^2 = \sigma_c^2 \left(\frac{d_wP}{dx}({_w\tilde{\mu}}) \right)^{-1}
$$

$$
\sigma_{w\tilde{\mu}_s}^2 = \frac{\left(\sum_i^N w_i^2 \right)}{4 \left(\sum_i^N w_i \right)^2} \frac{\left(\sum_i^N w_i \right)^2}{\left(\sum_i^N w_i p_i({_w\tilde{\mu}}) \right)^2}
$$

$$
\sigma_{w\tilde{\mu}_s}^2 = \frac{\left(\sum_i^N w_i^2 \right)}{4 \left(\sum_i^N w_i p_i({_w\tilde{\mu}}) \right)^2}
$$

Figure 3: Numerical experiment with inverse density weighted Gaussian random samples of different size. The mean squared deviation of the weighted median is described well by equation 60 as a function of the sum of squared weights, which is indicated by the red line.
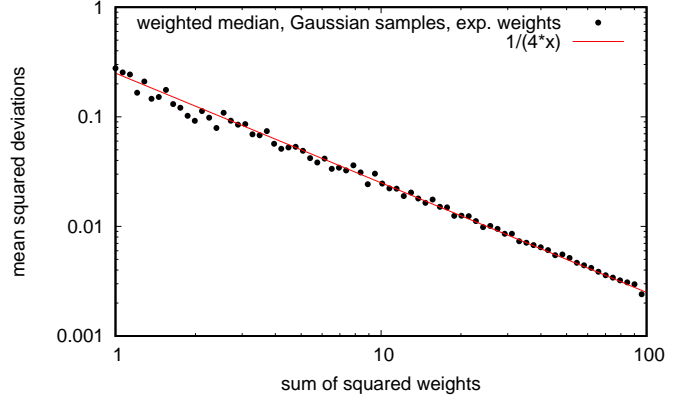
This is the variance we seek to minimize and a very similar optimization problem to the optimal weights for the weighted average (compare equation 35). The weights that minimize the variance of the weighted median are reciprocal to the probability density of the sample distribution at the median, as can be shown by taking the first and second derivatives with respect to the individual weights. The first derivative can only be zero for $w_i = p_i\left({_w}\tilde{\mu}\right)$ and the second derivative is positive, which are the necessary and sufficient conditions for a global minimum:

$$\frac{d\sigma^2_{{_w}\tilde{\mu}_s}}{dw_j} = \frac{w_j\left(\sum_i^N w_i p_i\left({_w}\tilde{\mu}\right)\right) - p_j\left({_w}\tilde{\mu}\right)\left(\sum_i^N w_i^2\right)}{2\left(\sum_i^N w_i p_i\left({_w}\tilde{\mu}\right)\right)^3}$$

$$\frac{d^2\sigma^2_{{_w}\tilde{\mu}_s}}{d^2 w_j} = \frac{1}{2\left(\sum_i^N w_i p_i\left({_w}\tilde{\mu}\right)\right)^2} - \frac{2w_j p_j\left({_w}\tilde{\mu}\right)}{2\left(\sum_i^N w_i p_i\left({_w}\tilde{\mu}\right)\right)^3} + \frac{3\left(p_j\left({_w}\tilde{\mu}\right)\right)^2\left(\sum_i^N w_i^2\right)}{2\left(\sum_i^N w_i p_i\left({_w}\tilde{\mu}\right)\right)^4}$$

$$\arg\min_{w_i}\sigma^2_{{_w}\tilde{\mu}_s} = p_i\left({_w}\tilde{\mu}\right) \tag{59}$$

Numerical experiments as shown in figure 2 also confirm this result. When the weights are set proportional to $p_i\left({_w}\tilde{\mu}\right)$, the variance of the median is equal to $1/4$ the inverse of the sum of probability densities at the median squared (compare figure 3):

$$\sigma^2_{{_w}\tilde{\mu}_s} = \frac{1}{4\left(\sum\left(p_i\left({_w}\tilde{\mu}\right)\right)^2\right)} \tag{60}$$

When the weights are proportional, they are still optimal, but estimating the variance from the weights directly with equation 60 is not possble any more because the proportionality constant is unknown. The variance can still be estimated analogous to equation 55.

$$\sigma^2_{{_w}\tilde{\mu}_s} \approx \sum_i^N \left(x_i - {_w}\tilde{\mu}_s\right)^2 \int_{\sum_j^{i-1} w_i}^{\sum_j^{i} w_i} \exp\left(-\frac{1}{2}\left(\left(y - \frac{1}{2}\right)^2 \sigma_c^{-2} + \log\left(2\pi\sigma_c^2\right)\right)\right) dy \tag{61}$$

22

Figure 4: Variance and kurtosis of the weighted sample median divided by its estimated standard deviation (equation 61). The samples are drawn from a scaled normal distribution, where the weights are drawn from an exponential distribution and the divisors are set accordingly.
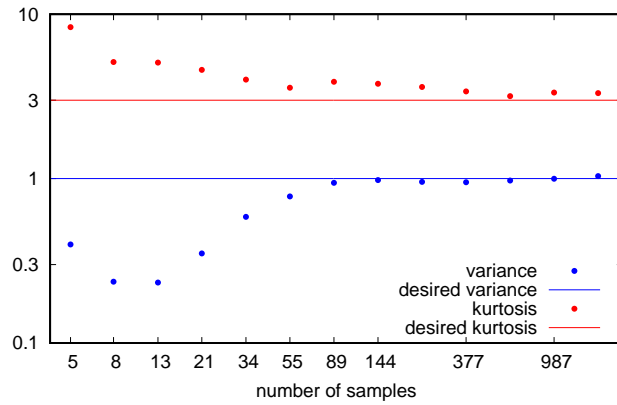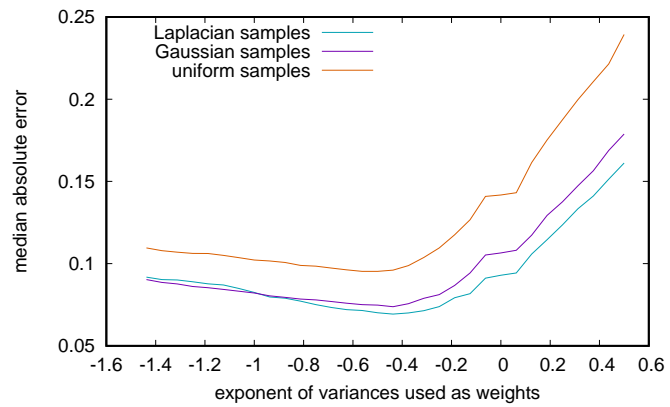

Figure 5: The median of absolute deviations of the sample median of samples following either a Gaussian, a Laplacian or a uniform distribution, with variances following an exponential distribution. The weights are set to a power of the associated sample variances and as can be seen the optimal power is around $-1/2$.

One possibly unexpected consequence of weights proportional to the probability density at the median is that when the distribution of each sample has the same location and shape, and only differs in scaling, the optimal weights are proportional to the inverse standard deviation as demonstrated in figure 5, not inverse variances as for the weighted average. This is because a scaling by a factor amounts to a linear decrease of the density at the median, but an increase of the variance by the same factor squared.

3.8. Gaussian Probability Densities

The Gaussian or normal probability distribution from table 1 can be generalized to any number of dimensions and it is a good approximation in many circumstances, hence the name normal. Its prevalence can be explained by the principle of maximum entropy (see section 3.1.2) and the central limit theorem (see section 3.5). The multivariate probability density $\phi\left(\vec{x}, \vec{\mu}, \boldsymbol{\Sigma}\right)$ is a function of the vector $\vec{x}$, the mean point $\vec{\mu}$ and the symmetric covariance matrix $\boldsymbol{\Sigma}$, which is the multidimensional generalization of the one-dimensional variance. Each diagonal element contains the variance in that dimension, and off-diagonal elements contain covariances. All vectorial quantities are marked with a small arrow on top and matrices have a bold face. The probability density is given by:

$$\phi\left(\vec{x}, \vec{\mu}, \boldsymbol{\Sigma}\right) \equiv \exp\left(-\frac{1}{2}\left(\left(\vec{x} - \vec{\mu}\right)^{\top} \boldsymbol{\Sigma}^{-1} \left(\vec{x} - \vec{\mu}\right) + \log\left(\left|2\pi\boldsymbol{\Sigma}\right|\right)\right)\right) \tag{62}$$

Where

ϕ is a probability density function of a Gaussian distribution

$\vec{x}$ is a point in space

$\vec{\mu}$ is the mean value

$\boldsymbol{\Sigma}$ is the covariance matrix

As a side note, a multidimensional density like this can hardly be written without vectors and matrices in general, specifically because of the matrix inverse. Expressing the matrix inverse $\boldsymbol{\Sigma}^{-1}$ in terms of the individual entries of the matrix $\boldsymbol{\Sigma}$ is extraordinarily cumbersome. Only when the individual values of the vector $\vec{x}$ are uncorrelated, and the off-diagonal values of $\boldsymbol{\Sigma}$ therefore are zero, can the matrix expression in the exponent be rearranged to:

$$\left(\vec{x} - \vec{\mu}\right)^{\top} \boldsymbol{\Sigma}^{-1} \left(\vec{x} - \vec{\mu}\right) = \sum_{i}^{N} \frac{\left(x_i - \mu_i\right)^2}{\boldsymbol{\Sigma}_{ii}} \quad \text{if all } x_i \text{ are uncorrelated}$$

The normalization factor $\left(\left|2\pi\boldsymbol{\Sigma}\right|\right)^{-\frac{1}{2}}$ was pulled into the exponent for a more compact equation. What makes the Gaussian distribution especially useful, aside from its prevalence, is that products, convolutions, cuts and projections can be performed symbolically. The joint probability of several uncorrelated outcomes is given by the product of their probabilities. Equally, the probability density that satisfies all individual probability distributions is computed by a point-wise product of the individual densities. The product of Gaussian distributions is a scaled Gaussian with a mean given by the $\boldsymbol{\Sigma}^{-1}$ weighted arithmetic mean of the individual means and a new covariance given by the inverse of the sum of those weights:

$$\phi\left(\vec{x}, \vec{\mu}_1, \boldsymbol{\Sigma}_1\right) \phi\left(\vec{x}, \vec{\mu}_2, \boldsymbol{\Sigma}_2\right) = \phi\left(\vec{\mu}_1, \vec{\mu}_2, \boldsymbol{\Sigma}_o\right) \phi\left(\vec{x}, \mu, \boldsymbol{\Sigma}_*\right) \tag{63}$$

$$\boldsymbol{\Sigma}_o = \boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2$$

$$\boldsymbol{\Sigma}_* = \left(\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1}\right)^{-1}$$

$$\mu = \boldsymbol{\Sigma}_* \left(\boldsymbol{\Sigma}_1^{-1}\vec{\mu}_1 + \boldsymbol{\Sigma}_2^{-1}\vec{\mu}_2\right)$$

This result can be simplified further, when both densities are identical:

$$(\phi\left(\vec{x}, \vec{\mu}, \mathbf{\Sigma}\right))^2 = \phi\left(\vec{\mu}, \vec{\mu}, 2\mathbf{\Sigma}\right)\phi\left(\vec{x}, \vec{\mu}, \frac{1}{2}\mathbf{\Sigma}\right) \tag{64}$$

The probability distribution for the sum of two independent random variables is given by the convolution of the individual distributions. The rules for combining the means and variances are equivalent to the commonly employed error propagation, the means add, just like the variances.

$$\phi\left(\vec{x}, \vec{\mu}_1, \mathbf{\Sigma}_1\right) * \phi\left(\vec{x}, \vec{\mu}_2, \mathbf{\Sigma}_2\right) = \phi\left(\vec{x}, \vec{\mu}_1 + \vec{\mu}_2, \mathbf{\Sigma}_1 + \mathbf{\Sigma}_2\right) \tag{65}$$

If the individual distributions are correlated, the means still add to form the sum, but there is an additional summand for the variance of the sum, $\mathbf{\Sigma}_{X+Y} = \mathbf{\Sigma}_X + \mathbf{\Sigma}_Y + 2\operatorname{cov}\left(X, Y\right)$. In case of perfect correlation this reduces to $\left(\sqrt{\mathbf{\Sigma}_X} + \sqrt{\mathbf{\Sigma}_Y}\right)^2$.

Marginalization integrals, conditional probabilities and multidimensional Fourier transforms sometimes require projections onto and slices of Gaussian densities on subspaces. Projections on linear subspaces lead to Gaussian densities and slices are Gaussian functions that are no longer proper densities, because they don't necessarily need to integrate to unity. Let the matrix $\boldsymbol{A}$ be the basis of the linear subspace and $\vec{b}$ its origin. The basis is composed of the basis vectors and it maps from the coordinates of the subspace to the full space. Its pseudoinverse $\boldsymbol{A}^+$ maps any point $\vec{x}$ onto the closest point in the subspace, because it is the least squares solution to the following linear equation:

$$\boldsymbol{A}\vec{x}_s + \vec{b} = \vec{x}$$

$$\underset{\vec{x}_s}{\arg\min} \left|\boldsymbol{A}\vec{x}_s + \vec{b} - \vec{x}\right|^2 = \boldsymbol{A}^+\left(\vec{x} - \vec{b}\right) \tag{66}$$

$$\boldsymbol{A}^+ = \left(\boldsymbol{A}^* \boldsymbol{A}\right)^{-1} \boldsymbol{A} \tag{67}$$

Cuts on the subspace can be derived symbolically by substituting $\vec{x}$ in equation 62 with coordinates on the subspace $\vec{x}_s$ mapped back via $\boldsymbol{A}\vec{x}_s + \vec{b}$:

$$\exp\left(-\frac{1}{2}\left(\left(\boldsymbol{A}\vec{x}_s + \vec{b} - \vec{\mu}\right)^\top \mathbf{\Sigma}^{-1}\left(\boldsymbol{A}\vec{x}_s + \vec{b} - \vec{\mu}\right) + \log\left(|2\pi\mathbf{\Sigma}|\right)\right)\right)$$

Expanding the exponent,

$$\exp\left(-\frac{1}{2}\left(\vec{x}_s\boldsymbol{A}^\top\mathbf{\Sigma}^{-1}\boldsymbol{A}\vec{x}_s - 2\vec{x}_s\boldsymbol{A}^\top\mathbf{\Sigma}^{-1}\left(\vec{\mu} - \vec{b}\right) + \left(\vec{\mu} - \vec{b}\right)^\top \mathbf{\Sigma}^{-1}\left(\vec{\mu} - \vec{b}\right) + \log\left(|2\pi\mathbf{\Sigma}|\right)\right)\right)$$

and then completing the square it can be seen that the covariance of the cut is $\left(\boldsymbol{A}^\top\mathbf{\Sigma}^{-1}\boldsymbol{A}\right)^{-1}$. The mean of the cut $\vec{\mu}_c$ needs to fulfill the following equation to complete the square:

$$\boldsymbol{A}^\top\mathbf{\Sigma}^{-1}\boldsymbol{A}\vec{\mu}_c = \boldsymbol{A}^\top\mathbf{\Sigma}^{-1}\left(\vec{\mu} - \vec{b}\right)$$

$$\vec{\mu}_c = \left(\boldsymbol{A}^\top\mathbf{\Sigma}^{-1}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^\top\mathbf{\Sigma}^{-1}\left(\vec{\mu} - \vec{b}\right) \tag{68}$$

Contrary to what one might think $\boldsymbol{A}^{-1}\vec{\mu}$ is not a solution because $\boldsymbol{A}$ is not invertible in general, the correct

solution is $\left(\boldsymbol{A}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\top}\boldsymbol{\Sigma}^{-1}\left(\vec{\mu}-\vec{b}\right)$, because $\left(\boldsymbol{A}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{A}\right)$ is invertible so long $\boldsymbol{A}$ is of full rank. This operation can also be seen as an orthogonal projection, just in a different space that is sheared by the inverse covariance matrix. With these transformations the slice through a Gaussian density on a linear subspace can be restated as:

$$\exp\Bigg(-\frac{1}{2}\bigg($$

$$\left(\vec{x}_s - \left(\boldsymbol{A}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\top}\boldsymbol{\Sigma}^{-1}\left(\vec{\mu}-\vec{b}\right)\right)^{\top}\boldsymbol{A}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{A}\left(\vec{x}_s - \left(\boldsymbol{A}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\top}\boldsymbol{\Sigma}^{-1}\left(\vec{\mu}-\vec{b}\right)\right)$$

$$+ \left(\boldsymbol{A}\left(\boldsymbol{A}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\top}\boldsymbol{\Sigma}^{-1}\left(\vec{\mu}-\vec{b}\right)+\vec{b}-\vec{\mu}\right)^{\top}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{A}\left(\boldsymbol{A}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\top}\boldsymbol{\Sigma}^{-1}\left(\vec{\mu}-\vec{b}\right)+\vec{b}-\vec{\mu}\right)$$

$$+\log\left(|2\pi\boldsymbol{\Sigma}|\right)$$

$$\bigg)\Bigg) \tag{69}$$

Or expressed in terms of ϕ (equation 62):

$$\phi\left(\boldsymbol{A}\vec{x}_s, \vec{\mu}, \boldsymbol{\Sigma}\right) =$$

$$\frac{\phi\left(\vec{x}_s,\, \left(\boldsymbol{A}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\top}\boldsymbol{\Sigma}^{-1}\left(\vec{\mu}-\vec{b}\right),\, \left(\boldsymbol{A}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{A}\right)^{-1}\right)\phi\left(\boldsymbol{A}\left(\boldsymbol{A}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\top}\boldsymbol{\Sigma}^{-1}\left(\vec{\mu}-\vec{b}\right)+\vec{b}, \vec{\mu}, \boldsymbol{\Sigma}\right)}{\phi\left(\vec{0}, \vec{0}, \left(\boldsymbol{A}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{A}\right)^{-1}\right)}$$

$$\tag{70}$$

When projecting a Gaussian density, the projected density is still a proper probability density and the mean of the projected density is indeed just the mean of the distribution projected with $\boldsymbol{A}^{+}$, because a projection corresponds to an integral along all orthogonal directions. The projection of the covariance matrix via $\left(\boldsymbol{A}^{+}\right)^{\top}\boldsymbol{\Sigma}\boldsymbol{A}^{+}$ can be understood with the projection slice theorem (see section 4) and the analytic Fourier transform of the Gaussian probability density (equation 80) or alternatively with the singular value decomposition of the covariance matrix[35]. And so the projection of a Gaussian density onto the linear subspace $\boldsymbol{A}\vec{x}+\vec{b}$ is given by:

$$\phi\left(\vec{x}_s, \boldsymbol{A}^{+}\left(\vec{\mu}-\vec{b}\right), \boldsymbol{A}^{+}\boldsymbol{\Sigma}\boldsymbol{A}^{+\top}\right) =$$

$$\exp\left(-\frac{1}{2}\left(\left(\vec{x}_s - \boldsymbol{A}^{+}\left(\vec{\mu}-\vec{b}\right)\right)^{\top}\left(\boldsymbol{A}^{+}\boldsymbol{\Sigma}\boldsymbol{A}^{+\top}\right)^{-1}\left(\vec{x}_s - \boldsymbol{A}^{+}\left(\vec{\mu}-\vec{b}\right)\right) + \log\left(\left|2\pi\boldsymbol{A}^{+}\boldsymbol{\Sigma}\boldsymbol{A}^{+\top}\right|\right)\right)\right) \tag{71}$$
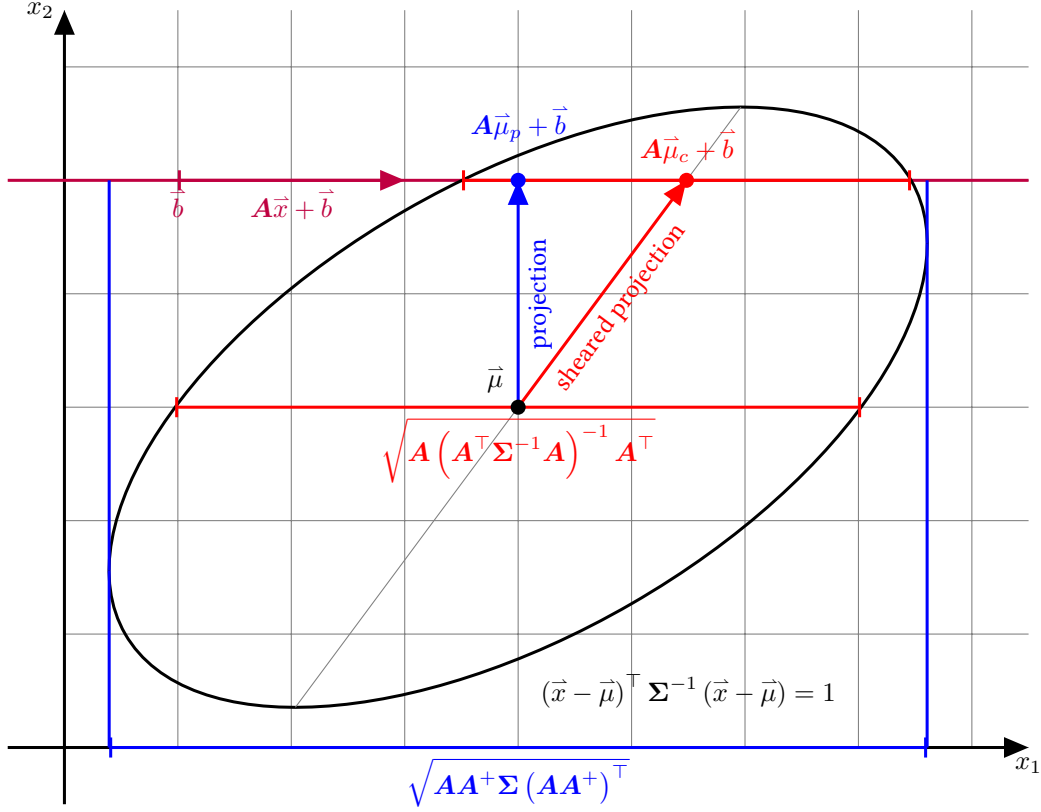
Figure 6: Illustration of cuts and projections of Gaussian densities visualized with an elliptical isoline where $(\vec{x} - \vec{\mu})^\top \Sigma^{-1} (\vec{x} - \vec{\mu}) = 1$. The subspace is given by $\boldsymbol{A}\vec{x}_s + \vec{b}$ and drawn in purple. The slice of the density along this line yields a Gaussian function. Its mean $\vec{\mu}_c$ is given by equation 68 and its location in the superspace is $\boldsymbol{A}\vec{\mu}_c + \vec{b}$. The covariance of the slice is $\left(\boldsymbol{A}^\top \Sigma^{-1} \boldsymbol{A} \right)^{-1}$, which can be mapped back to the original frame with the matrix $\boldsymbol{A}$. The projection of the density onto the subspace gives rise to another density with mean $\vec{\mu}_p$ given by equation 66 and covariance $\left(\boldsymbol{A}^+ \Sigma \boldsymbol{A}^{+\top} \right)$ in the coordinates of the subspace, which mapped back to the original frame is $\boldsymbol{A}\boldsymbol{A}^+ \Sigma \left(\boldsymbol{A}\boldsymbol{A}^+ \right)^\top$.

4. Fourier Transforms

The continuous Fourier transform decomposes a function by projecting it onto a continuous basis of orthogonal complex wave functions. The periodic basis functions are given by:

$$\exp\left(-2\pi i \vec{q}^{\top}\vec{x}\right)$$

If it exists, the Fourier transform of a function is another function, which measures the coefficients necessary to reconstruct the original function with said complex wave functions. The dimensions of the Fourier transformed function are reciprocal to the dimensions of the function. Because of this the Fourier transform can be seen as a bidirectional mapping between 'real space' and the frequency components in 'reciprocal space'. There are multiple definitions that differ slightly in regards to linear scaling factors, but the following definition will be used in this work for maximum symmetry between transform and inverse transform [36, Prologue]:

$$\mathcal{F}_{\vec{x}}\left(f\left(\vec{x}\right)\right)\left(\vec{q}\right) \equiv \int_{-\infty}^{\infty} f\left(\vec{x}\right)\exp\left(-2\pi i \vec{q}^{\top}\vec{x}\right) d\vec{x} \tag{72}$$

$$\mathcal{F}_{\vec{q}}^{-1}\left(f\left(\vec{q}\right)\right)\left(\vec{x}\right) = \int_{-\infty}^{\infty} f\left(\vec{q}\right)\exp\left(2\pi i \vec{x}^{\top}\vec{q}\right) d\vec{q} \tag{73}$$

$$\hat{f}\left(\vec{q}\right) \equiv \mathcal{F}_{\vec{x}}\left(f\left(\vec{x}\right)\right)\left(\vec{q}\right)$$

There are several properties of and identities involving the Fourier transform. In the following only those most relevant to this work will be listed.

Linearity

The Fourier transform is linear, meaning that it is additive and homogeneous:

$$\mathcal{F}_{\vec{x}}\left(f\left(\vec{x}\right)+g\left(\vec{x}\right)\right)\left(\vec{q}\right) = \mathcal{F}_{\vec{x}}\left(f\left(\vec{x}\right)\right)\left(\vec{q}\right)+\mathcal{F}_{\vec{x}}\left(g\left(\vec{x}\right)\right)\left(\vec{q}\right)$$

$$\mathcal{F}_{\vec{x}}\left(af\left(\vec{x}\right)\right)\left(\vec{q}\right) = a\mathcal{F}_{\vec{x}}\left(af\left(\vec{x}\right)\right)\left(\vec{q}\right)$$

This makes it act like a continuous analog to a multiplication with a matrix, while discrete Fourier transforms can be written as a matrix directly.

Differentiation

Differentiation with respect to the dimensions of the transform are given by:

$$\mathcal{F}_{\vec{x}}\left(\frac{d^{n}}{d\vec{x}^{n}}f\left(\vec{x}\right)\right)\left(\vec{q}\right) = \left(2\pi i \vec{q}\right)^{n}\mathcal{F}_{\vec{x}}\left(f\left(\vec{x}\right)\right)\left(\vec{q}\right) \tag{74}$$

This identity will be used to solve a differential equations in section 5.

Convolution theorem

The Fourier transform of a product of two functions is equal to the convolution of the Fourier transform of the functions individually and the other vice versa:

$$\mathcal{F}_{\vec{x}}\left(f\left(\vec{x}\right)g\left(\vec{x}\right)\right)\left(\vec{q}\right) = \mathcal{F}_{\vec{x}}\left(f\left(\vec{x}\right)\left(\vec{q}\right)\right) * \mathcal{F}_{\vec{x}}\left(g\left(\vec{x}\right)\right)\left(\vec{q}\right) \tag{75}$$

$$\mathcal{F}_{\vec{x}}\left(f\left(\vec{x}\right) * g\left(\vec{x}\right)\right)\left(\vec{q}\right) = \mathcal{F}_{\vec{x}}\left(f\left(\vec{x}\right)\left(\vec{q}\right)\right)\mathcal{F}_{\vec{x}}\left(g\left(\vec{x}\right)\right)\left(\vec{q}\right) \tag{76}$$

The shift theorem

The shift of a function f by a vector $\vec{s}$ results in a modulation of its Fourier transform with a linear phase ramp:

$$\mathcal{F}_{\vec{x}}\left(\mathrm{f}\left(\vec{x}+\vec{s}\right)\right)\left(\vec{q}\right) = \exp\left(-2\pi i\vec{s}^{\top}\vec{q}\right)\mathcal{F}_{\vec{x}}\left(\mathrm{f}\left(\vec{x}\right)\right)\left(\vec{q}\right) \tag{77}$$

Conjugation

The Fourier Transform of the complex conjugate of a function $\mathrm{f}(x)$ is equal to the mirrored and complex conjugated Fourier transform of the original function.

$$\mathcal{F}_{x}\left(\overline{\mathrm{f}(x)}\right)(q) = \overline{\mathcal{F}_{x}\left(\mathrm{f}(x)\right)(-q)} \tag{78}$$

The discrete Fourier transform

The discrete Fourier transform (DFT) is the discrete analog to the continuous Fourier transform where both input and output are sampled with the same number of equally spaced samples. The distance between the samples in the output is reciprocal to the size of the input and vice versa. A finite input can be seen to be the product of an input and a top hat function. The Fourier transform of this product is then the convolution of the Fourier transform of the object and the Fourier transform of the top hat function, which is the normalized sinc-function in one dimension. This is how, for finite signals, the continuous Fourier transform can be interpolated from the discrete Fourier transform, but there are computationally more efficient methods that trade some accuracy for more locality [37]. The discrete Fourier transform (DFT) can be accelerated to require $O(n\log(n))$ operations instead of the naive $O(n^2)$, which is why it is useful also for accelerating convolutions and projections computationally. This accelerated discrete Fourier transform is called fast Fourier transform (FFT) [38].

An important link between DFTs and the continuous Fourier transform is the Nyquist–Shannon sampling theorem. The theorem is often stated in terms of band-limited signals, but there is a way to state it without first defining what band-limited signals are. The theorem asserts that a periodic or constrained function can be reconstructed entirely from frequency components at frequencies corresponding to integer multiples of the frequency of repetition in the periodic signal, or corresponding to periodicities at integer fractions of twice the maximum extent of the constrained function. This holds in each dimension of the function independently. Conversely, if the Fourier transform of a function is a constrained function, it is a band limited function, for which the sampling theorem holds in its common formulation. Sufficient

---

[3]Big-O denotes the asymptotic behavior of a function for large inputs. $O(f(x))$ is the class of functions that are bounded by a constant multiple of $f(x)$ for all possible inputs.

sampling for a band-limited function is at intervals with half the periodicity of the frequency component with the highest frequency [39]. There is a generalization of this theorem, that the samples need not be equally spaced. The same number of samples at arbitrary but not identical positions is sufficient too. In practice however, the more uneven the sampling is, the more even small measurement errors and numerical errors are amplified [40].

Projection-Slice theorem

A slice through the origin of the Fourier transform of an object is equivalent to the Fourier transform of a projection along the line orthogonal to the slice. This can be used to efficiently approximate projections using the FFT. There, arbitrary cuts are only approximations, because the slice will not intersect the sample points directly most of the time, and therefore the exact values on the slice need to be interpolated from neighbouring values [41, 42].

The Fourier Transform of Common Functions

The Fourier transform of a Dirac delta distribution centered at 0 is a function, which is constantly 1. If the delta distribution is shifted, the shift introduces a complex modulation according to the shift theorem:

$$\mathcal{F}_{\vec{x}}\left(\delta\left(\vec{x} - \vec{s}\right)\right)\left(\vec{q}\right) = \exp\left(-2\pi i \vec{s}^{\top}\vec{q}\right) \tag{79}$$

The Fourier transform of a shifted Gaussian distribution is a Gaussian function with inverse covariance, centered at zero and instead modulated by a complex exponential:

$$\mathcal{F}_{\vec{x}}\left(\frac{\exp\left(-\left(\vec{x} - \vec{\mu}\right)^{\top}\boldsymbol{\Sigma}^{-1}\left(\vec{x} - \vec{\mu}\right)\right)}{\left|2\pi\boldsymbol{\Sigma}\right|^{\frac{1}{2}}}\right)\left(\vec{q}\right) = \exp\left(-2\pi^2\vec{q}^{\top}\boldsymbol{\Sigma}\vec{q} - 2i\pi\vec{\mu}^{\top}\vec{q}\right) \tag{80}$$

The rectangular function or top hat function is defined as:

$$\mathrm{rect}\left(x\right) = \begin{cases} 1 & \text{if } |x| < \frac{1}{2} \\ \frac{1}{2} & \text{if } |x| = \frac{1}{2} \\ 0 & \text{otherwise} \end{cases} \tag{81}$$

Its Fourier transform is the sinc function:

$$\mathrm{sinc}\left(x\right) \equiv \begin{cases} 1 & \text{if } x = 0 \\ \dfrac{\sin\left(\pi x\right)}{\pi x} & \text{otherwise} \end{cases}$$

The Fourier transform of a Laplace distribution is a Lorentzian function, that is a non-normalized Cauchy distribution. Conversely, the Fourier transform of a Cauchy distribution is similar to a Laplace distribution,

however it is not normalized:

$$\mathcal{F}_{\vec{x}} \left(\frac{\exp\left(-\frac{|x - \mu|}{b}\right)}{2b} \right)(\vec{q}) = \frac{\exp\left(-2\pi i \mu q\right)}{(2\pi b q)^2 + 1} \tag{82}$$

$$\mathcal{F}_{\vec{x}} \left(\frac{1}{\pi \gamma \left(1 + \left(\frac{x - x_0}{\gamma}\right)^2\right)} \right)(\vec{q}) = \exp\left(-2\pi i q - 2\pi \gamma |q|\right) \tag{83}$$

Following Applications

The Fourier transform of a probability distribution is its characteristic function, which can be used as an alternative definition for the Dirac δ-Distribution and in a proof for the central limit theorem. However, the main application of the Fourier transform is the analysis of periodicities. Waves and crystals are inherently periodical and therefore the Fourier transform arises almost naturally when describing diffraction of waves.

5. Diffraction of Waves

Whether measuring neutrons, electrons or photons, there is a wave function whose squared amplitude gives the probability density of measuring a particle at a point in time and space. The classical wave equations, that is neglecting quantization into integer particles, describe most aspects of diffraction. For this work, as for crystallography, the interaction of photons with matter is the most relevant, and because the behaviour of neutron and electron radiation is mostly analogous to the behaviour of electromagnetic radiation, only it will be described in greater detail. In the following, the Fourier diffraction theorem, a simplified description for the interaction between waves and objects, will be derived from Maxwell's equations.

5.1. Maxwell's Equations

From Maxwell's equations all aspects of classical electromagnetism can be derived [43, Chap. I]. There are four fundamental equations. Gauss's law describes the electric field $\vec{E}$ generated by a spatial distribution of charge density ρ. Gauss's law for magnetism stipulates that the magnetic field $\vec{B}$ has a divergence equal to zero, implying that magnetic monopoles do not exist. Faraday's law of induction governs how changes in the magnetic field will lead to electromagnetic induction. Ampère's circuital law with Maxwell's addition relates how electric current $\vec{J}$ and changes in the electric field affect the magnetic field.

The formulations use special operators for a very compact notation. The nabla operator ∇ is denotes the vector of partial derivative operators:

$$\nabla \equiv \begin{pmatrix} \dfrac{\partial}{\partial x_1} \\ ... \\ \dfrac{\partial}{\partial x_n} \end{pmatrix} \tag{84}$$

The divergence operator $\nabla \cdot$, like the dot implies, denotes the scalar product of the nabla operator with a vector-valued function: $\nabla \cdot \vec{f}(\vec{x}) = \sum_i \dfrac{\partial f_i}{\partial x_i}$. The curl operator $\nabla \times$ computes the infinitesimal circulation of a vector-field in three dimensions with the cross product between the Nabla operator and a vector-valued function:

$$\nabla \times \vec{f} = \begin{pmatrix} \dfrac{\partial f_{x_3}}{\partial x_2} - \dfrac{\partial f_{x_2}}{\partial x_3} \\ \dfrac{\partial f_{x_1}}{\partial x_3} - \dfrac{\partial f_{x_3}}{\partial x_1} \\ \dfrac{\partial f_{x_2}}{\partial x_1} - \dfrac{\partial f_{x_1}}{\partial x_2} \end{pmatrix}$$

The Laplace operator ∇^2 is defined as:

$$\nabla^2 f = \nabla \cdot (\nabla f) \tag{85}$$

For orthonormal coordinates this is equal to:

$$\nabla^2 f = \sum_i^N \dfrac{\partial^2 f}{\partial^2 x_i}$$

The Laplace operator applied to a vector valued function, also called the vector Laplace operator often takes the same symbol as the regular Laplace operator, but because the definition is subtly different it will be distinguished by an arrow on top. $\vec{\nabla}^2$ is defined as:

$$\vec{\nabla}^2\vec{f} = \nabla\left(\nabla\cdot\vec{f}\right) - \nabla\times\left(\nabla\times\vec{f}\right) \tag{86}$$

For orthonormal coordinates this reduces to:

$$\vec{\nabla}^2\vec{f} = \begin{pmatrix} \nabla^2 f_{x_1} \\ ... \\ \nabla^2 f_{x_n} \end{pmatrix} = \begin{pmatrix} \sum_i^N \dfrac{\partial^2}{\partial x_i^2}f_{x_1} \\ ... \\ \sum_i^N \dfrac{\partial^2}{\partial x_i^2}f_{x_n} \end{pmatrix}$$

Gauss law	$\nabla\cdot\vec{E} = \dfrac{\rho}{\epsilon_0}$
Gauss law for magnetism	$\nabla\cdot\vec{B} = 0$
Faraday's law of induction	$\nabla\times\vec{E} = -\dfrac{\partial\vec{B}}{\partial t}$
Ampère's circuital law	$\nabla\times\vec{B} = \mu_0\left(\vec{J} + \epsilon_0\dfrac{\partial\vec{E}}{\partial t}\right)$

Table 2: Maxwell's equations in SI units [43, Eqns. I.1a&I.1b].

$\vec{E}$	electric field
$\vec{B}$	magnetic field
ρ	electric charge per unit volume
ρ_f	excess charge per unit volume
ϵ	dielectric permittivity of the medium
ϵ_0	dielectric permittivity of vacuum
μ	magnetic permeability of the medium
μ_0	magnetic permeability of vacuum
$\vec{J}$	current density vector

5.2. The Macroscopic Formulation of Maxwell's Equations

There is an equivalent formulation of Maxwell's equations describing the relations from a macroscopic view. They are just as valid as the previous set and any one of them can be used to describe diffraction. The derivation of the macroscopic formulation will be used mainly to introduces the quantities that are necessary to define what linear isotropic media are.

First we introduce the electric displacement field $\vec{D}$, defined as $\vec{D} \equiv \epsilon_0\vec{E} + \vec{P}$ [43, equation 6.63]. The polarization density $\vec{P}$ is the electric dipole at each point of the medium, whether permanent or induced. The total charge density ρ is the sum of the density of free charge ρ_f and bound charge ρ_b. Only bound charges can be polarized and so $\rho_b = -\nabla\cdot\vec{P}$ which is a different form of Gauss law. With this we can

show that $\nabla \cdot \vec{D} = \rho_f$.

$$\rho = \rho_f + \rho_b \tag{87}$$

$$\rho = \rho_f - \nabla \cdot \vec{P} \tag{88}$$

Substituting equation 88 into Gauss' law we get:

$$\nabla \cdot \vec{E} = \frac{1}{\epsilon_0} \left(\rho_f - \nabla \cdot \vec{P} \right)$$

$$\epsilon_0 \nabla \cdot \vec{E} - \nabla \cdot \vec{P} = \rho_f$$

$$\nabla \cdot \vec{D} = \rho_f \tag{89}$$

Analogously, we introduce the demagnetising field $\vec{H}$ defined as $\vec{H} \equiv \dfrac{\vec{B}}{\mu_0} - \vec{M}$. The magnetization $\vec{M}$ is the density of magnetic dipole moments, whether permanent or induced, and their direction. The total currents $\vec{J}$ in a material can be decomposed into a sum of free currents $\vec{J}_f$, the magnetization current $\vec{J}_m$ and polarization current $\vec{J}_p$. Magnetization $\vec{M}$ contributes to the total current as $\vec{J}_m = \nabla \times \vec{M}$. The separation of paired charges when polarization is induced leads to the polarization current $\vec{J}_p = \dfrac{\partial}{\partial t} \vec{P}$. This allows us to restate Ampère's circuital law as follows:

$$\nabla \times \vec{B} = \mu_0 \left(\vec{J} + \epsilon_0 \frac{\partial \vec{E}}{\partial t} \right)$$

$$\nabla \times \vec{B} = \mu_0 \left(\vec{J}_f + \nabla \times \vec{M} + \frac{\partial}{\partial t} \vec{P} + \epsilon_0 \frac{\partial \vec{E}}{\partial t} \right)$$

$$\nabla \times \frac{\vec{B}}{\mu_0} = \vec{J}_f + \nabla \times \vec{M} + \frac{\partial}{\partial t} \vec{P} + \epsilon_0 \frac{\partial \vec{E}}{\partial t}$$

$$\nabla \times \frac{\vec{B}}{\mu_0} - \nabla \times \vec{M} = \vec{J}_f + \frac{\partial}{\partial t} \left(\epsilon_0 \vec{E} + \vec{P} \right)$$

$$\nabla \times \vec{H} = \vec{J}_f + \frac{\partial}{\partial t} \vec{D}$$

Thereby we arrive at a different set of equations for materials that responds linearly and immediately to electric and magnetic fields. They are listed in table 3.

Gauss law	$\nabla \cdot \vec{D} = \rho_f$
Gauss law for magnetism	$\nabla \cdot \vec{B} = 0$
Faraday's law of induction	$\nabla \times \vec{E} = -\dfrac{\partial \vec{B}}{\partial t}$
Ampère's circuital law	$\nabla \times \vec{H} = \vec{J}_f + \dfrac{\partial}{\partial t} \vec{D}$

Table 3: The macroscopic formulation of Maxwell's equations [43, Eqns. 6.62].

5.3. Linear Media

There is a simplified set of equations that is structurally very similar to Maxwell's equations, but only valid in linear media. Linear media respond instantaneously and linearly to external electromagnetic fields. This is a good first-order approximation in many cases, and makes analytic solutions easier.

In linear media the displacement field and free current and the magnetizing field are directly related to the electric and magnetic field respectively [44, Eqns. 1.2]:

$$\vec{D} = \epsilon \vec{E} \tag{90}$$

$$\vec{B} = \mu \vec{H} \tag{91}$$

$$\vec{J}_f = \sigma \vec{E} \tag{92}$$

In general, the material properties ϵ, μ and σ are tensors. If the medium is locally without directional dependence, they are scalars and the medium is called *linear isotropic*. Applying this simplification to the macroscopic formulation of Maxwell's equations we get an analogous set of equations, which is listed in table 4.

Gauss law	$\nabla \cdot \epsilon \vec{E} = \rho_f$
Gauss law for magnetism	$\nabla \cdot \vec{B} = 0$
Faraday's law of induction	$\nabla \times \vec{E} = -\dfrac{\partial \vec{B}}{\partial t}$
Ampère's circuital law	$\nabla \times \dfrac{\vec{B}}{\mu} = \sigma \vec{E} + \dfrac{\partial}{\partial t} \epsilon \vec{E}$

Table 4: Relations analogous to Maxwell's equations for linear materials, with spatially and temporally varying electric permittivity ϵ, magnetic permeability μ and electric conductivity σ [44, Eqns. 1.1&1.2].

5.4. The Fourier Diffraction Theorem

The two electromagnetic wave equations can be derived from Maxwell's equations without approximations and describe the electromagnetic waves in the most general way [43, Chap. 6.2].

$$\mu_0 \epsilon_0 \frac{\partial^2 \vec{E}}{\partial t^2} - \vec{\nabla}^2 \vec{E} = -\nabla \frac{\rho}{\epsilon_0} - \mu_0 \frac{\partial \vec{J}}{\partial t} \tag{93}$$

$$\mu_0 \epsilon_0 \frac{\partial^2 \vec{B}}{\partial t^2} - \vec{\nabla}^2 \vec{B} = \mu_0 \nabla \times \vec{J} \tag{94}$$

However, they are often too general to give insightful descriptions of diffraction phenomena. There is an approximate solution to the wave equations, known as the Fourier diffraction theorem. The following derivation mostly follows chapter 6 of Kak and Slaney [45]. Gauss law in linear isotropic materials can be

expanded by applying the product rule for derivatives to be more usable in the steps to follow:

$$\nabla \cdot \epsilon \vec{E} = \rho_f$$

$$\epsilon \nabla \cdot \vec{E} + \vec{E} \nabla \cdot \epsilon = \rho_f$$

$$\nabla \cdot \vec{E} + \vec{E} \frac{\nabla \cdot \epsilon}{\epsilon} = \frac{\rho_f}{\epsilon}$$

$$\nabla \cdot \vec{E} + \vec{E} \nabla \cdot \log \epsilon = \frac{\rho_f}{\epsilon} \tag{95}$$

In a source-free medium that is magnetically homogeneous there is no free current ($\vec{J}_f = 0$), no free charge ($\rho_f = 0$) and the magnetic permeability μ is constant. Starting with the curl of Faraday's law of induction in linear isotropic media:

$$\nabla \times \left(\nabla \times \vec{E} \right) = -\nabla \times \frac{\partial \vec{B}}{\partial t}$$

$$\nabla \left(\nabla \cdot \vec{E} \right) - \vec{\nabla}^2 \vec{E} = -\frac{\partial}{\partial t} \nabla \times \vec{B}$$

$$\nabla \left(\nabla \cdot \vec{E} \right) - \vec{\nabla}^2 \vec{E} = -\mu \frac{\partial^2}{\partial t^2} \epsilon \vec{E} \tag{96}$$

$$\nabla \left(\vec{E} \nabla \cdot \log \epsilon \right) + \vec{\nabla}^2 \vec{E} = \mu \frac{\partial^2}{\partial t^2} \epsilon \vec{E} \tag{97}$$

The wave equation for the electric field under these conditions (equation 97) can be derived. It is a specialization of the general wave equation (equation 93). Assuming the electric permittivity is a material property, that does not change over time, the equation can be rearranged slightly:

$$\nabla \left(\vec{E} \nabla \cdot \log \epsilon \right) + \vec{\nabla}^2 \vec{E} = \mu \epsilon \frac{\partial^2}{\partial t^2} \vec{E} \tag{98}$$

This equation however is still too complex for general solutions and we need more approximations. The term $\nabla \left(\vec{E} \nabla \cdot \log \epsilon \right)$ introduces a coupling between the dimensions of the electric field. When the relative spatial changes in the electric field are much larger than the relative spatial changes in the electric permittivity, the term can be neglected, because it is much smaller than the other spatially dependent term $\vec{\nabla}^2 \vec{E}$. This leads us to the following approximation:

$$\vec{\nabla}^2 \vec{E} = \mu \epsilon \frac{\partial^2}{\partial t^2} \vec{E} \tag{99}$$

But in the intended application the spatial changes are on the same order of magnitude as the spatial changes in the electric permittivity. There will be an appreciable error in this approximation, which can largely be corrected for with a polarization correction term. This term will be introduced at a later time and is mostly post hoc derived from geometric considerations.

Since each component of the electric field needs to fulfill equation 99 individually we can narrow our analysis to any one of them and still describe the interaction (neglecting polarization effects) in approximation. This is the approach of scalar theory [43, equation 3-7]. The scalar wave equation, using the scalar

function ψ is:

$$\mu\epsilon\frac{\partial^2}{\partial t^2}\psi = \nabla^2\psi \tag{100}$$

Because we are not interested in the time dependence of the scalar function ψ and only its spatial component $u(\vec{x})$, we can derive the following Helmholtz equation by separation of variables, which all stationary states need to fulfill [46, Chap. 1 D-1, 47, Eqn 2.2]:

$$\psi = u(\vec{x})v(t)$$

$$\mu\epsilon\frac{\partial^2}{\partial t^2}u(\vec{x})v(t) = \vec{\nabla}u(\vec{x})v(t)$$

$$\frac{\mu\epsilon\,\partial^2 v(t)}{u(t)\,\partial t^2} = \frac{\nabla^2 u(\vec{x})}{u(\vec{x})}$$

$$\frac{\partial^2 v(t)}{v(t)\,\partial t^2} = \frac{\mu\epsilon\,\nabla^2 u(\vec{x})}{u(\vec{x})} \tag{101}$$

The left hand side of equation 101 is a function only of time, whereas the right hand side is only a function of space. If both sides are constant, they form a solution to the scalar wave equation (equation 100).

$$\frac{\partial^2 v(t)}{v(t)\,\partial t^2} = -\omega^2 \tag{102}$$

$$\frac{\mu\epsilon\,\nabla^2 u(\vec{x})}{u(\vec{x})} = -\omega^2 \tag{103}$$

The second derivative of the field with respect to the time is the angular frequency ω squared negative. This is the extent to which we are interested in the time dependent part, we want to know the spatial distribution, which can be rearranged:

$$\frac{\nabla^2 u(\vec{x})}{u(\vec{x})} = -\frac{\omega^2}{\mu\epsilon}$$

$(\mu\epsilon)^{-1/2}$ is the speed of light in the medium at that position, therefore $\omega/\sqrt{4\pi^2\mu\epsilon}$ is the inverse wavelength or the local wavenumber $k(x)$, equal to the inverse wavelength.

$$\frac{\nabla^2 u(\vec{x})}{u(\vec{x})} = -4\pi^2 k(\vec{x})^2$$

This can be rearranged to:

$$\left(\nabla^2 + 4\pi^2 k(\vec{x})^2\right)u(\vec{x}) = 0 \tag{104}$$

Note that this equation is linear, therefore a linear combination of solutions is a solution too. To find the

---

[3]This is the crystallographers definition of k. Outside of crystallography k is the angular wavenumber, which differs by a factor of 2π, which would make the following equations slightly more compact, but the result slightly less so.

solutions we manipulate the equation by multiplying out the object dependence on the left-hand side:

$$k\left(\vec{x}\right)^2 = \left(k_0 + \left(k\left(\vec{x}\right) - k_0\right)\right)^2 = k_0^2 + 2k_0\left(k\left(\vec{x}\right) - k_0\right) + \left(k\left(\vec{x}\right) - k_0\right)^2$$

$$o\left(\vec{x}\right) \equiv 4\pi^2\left(2k_0\left(k\left(\vec{x}\right) - k_0\right) + \left(k\left(\vec{x}\right) - k_0\right)^2\right)$$

Defining $o(\vec{x})$ as the variation of k^2 around a median value k_0^2 this can be expressed more compactly as:

$$\left(\nabla^2 + 4\pi^2 k_0^2\right) u\left(\vec{x}\right) = -o\left(\vec{x}\right) u\left(\vec{x}\right) \tag{105}$$

The field $u\left(\vec{x}\right)$ can be seen as composed of the incident field $u_0\left(\vec{x}\right)$ not affected by the inhomogeneities and the scattered field $u_s\left(\vec{x}\right)$. The incident field is a solution to the homogeneous wave equation:

$$\left(\nabla^2 + 4\pi^2 k_0^2\right) u_0\left(\vec{x}\right) = 0 \tag{106}$$

By subtracting this partial solution from equation 105 we get:

$$\left(\nabla^2 + 4\pi^2 k_0^2\right) u_s\left(\vec{x}\right) = -o\left(\vec{x}\right) u\left(\vec{x}\right) \tag{107}$$

Noting that every integrable function can be composed by Dirac δ distributions, the product $o\left(\vec{x}\right) u\left(\vec{x}\right)$ can be expressed like this:

$$o\left(\vec{x}\right) u\left(\vec{x}\right) = \int o\left(\vec{x}'\right) u\left(\vec{x}'\right) \delta\left(\vec{x} - \vec{x}'\right) d\vec{x}'$$

Also, remembering that the equation is linear, a general solution to equation 105 can be found by first finding a solution for a δ distribution. This solution is called Green's function $g\left(\vec{x}, \vec{x}'\right)$ and it is the impulse response of the system:

$$\left(\nabla^2 + 4\pi^2 k_0^2\right) g\left(\vec{x}, \vec{x}'\right) = \delta\left(\vec{x} - \vec{x}'\right)$$

Green's function can be found by Fourier-transforming its definition, rearranging, and inversely transforming its Fourier representation:

$$\mathcal{F}_{\vec{x}}\left(\left(\nabla^2 + 4\pi^2 k_0^2\right) g\left(\vec{x}, \vec{x}'\right)\right)\left(\vec{q}\right) = \mathcal{F}_{\vec{x}}\left(\delta\left(\vec{x} - \vec{x}'\right)\right)\left(\vec{q}\right)$$

$$\left(-4\pi^2 \left|\vec{q}\right|^2 + 4\pi^2 k_0^2\right) \mathcal{F}_{\vec{x}}\left(g\left(\vec{x}, \vec{x}'\right)\right)\left(\vec{q}\right) = \exp\left(-2\pi i \vec{q}^\top \vec{x}'\right)$$

$$\mathcal{F}_{\vec{x}}\left(g\left(\vec{x}, \vec{x}'\right)\right)\left(\vec{q}\right) = \frac{\exp\left(-2\pi i \vec{q}^\top \vec{x}'\right)}{4\pi^2\left(k_0^2 - \left|\vec{q}\right|^2\right)} \tag{108}$$

$$g\left(\vec{x}, \vec{x}'\right) = \frac{-\exp\left(-i k_0 \left|\vec{x} - \vec{x}'\right|\right)}{4\pi \left|\vec{x} - \vec{x}'\right|} \tag{109}$$

The result in equation 108 depends on the number of dimensions and is here stated for three dimensions

only. With this special solution we can synthesize the general solution for the scattered signal as such:

$$\left(\nabla^2 + k_0^2\right) \int o\left(\vec{x}'\right) u\left(\vec{x}'\right) g\left(\vec{x} - \vec{x}'\right) d\vec{x}' = \int o\left(\vec{x}'\right) u\left(\vec{x}'\right) \delta\left(\vec{x} - \vec{x}'\right) d\vec{x}'$$

$$u_s\left(\vec{x}\right) = \int o\left(\vec{x}'\right) u\left(\vec{x}'\right) g\left(\vec{x} - \vec{x}'\right) d\vec{x}'$$

With Born's first order approximation $u\left(\vec{x}'\right) \approx u_0\left(\vec{x}'\right)$ [45, equation 38] we can write the scattered field as:

$$u_s\left(\vec{x}\right) = \int o\left(\vec{x}'\right) u\left(\vec{x}'\right) g\left(\vec{x}, \vec{x}'\right) d\vec{x}' \tag{110}$$

Note the similarity of this result to the Huygens-Fresnel principle, which states that every inhomogeneity acts as a point source of radiation upon stimulation by an incident field. Green's function in equation 109 takes the role of this point source. Because this result contains a convolution it is tempting to apply the convolution theorem (equation 76) because point-wise multiplications are easier to visualize and compute than convolutions.

$$\mathcal{F}_{\vec{x}}\left(u_s\left(\vec{x}\right)\right)\left(\vec{q}\right) = \mathcal{F}_{\vec{x}}\left(\int o\left(\vec{x}'\right) u\left(\vec{x}'\right) g\left(\vec{x} - \vec{x}'\right) d\vec{x}'\right)\left(\vec{q}\right)$$

$$\mathcal{F}_{\vec{x}}\left(u_s\left(\vec{x}\right)\right)\left(\vec{q}\right) = \mathcal{F}_{\vec{x}}\left(g\left(\vec{x}, \vec{x}'\right)\right)\left(\vec{q}\right) \mathcal{F}_{\vec{x}}\left(o\left(\vec{x}'\right) u\left(\vec{x}'\right)\right)\left(\vec{q}\right)$$

$$\mathcal{F}_{\vec{x}}\left(u_s\left(\vec{x}\right)\right)\left(\vec{q}\right) = \mathcal{F}_{\vec{x}}\left(g\left(\vec{x}, \vec{x}'\right)\right)\left(\vec{q}\right) \left[\mathcal{F}_{\vec{x}}\left(o\left(\vec{x}'\right)\right)\left(\vec{q}\right) * \mathcal{F}_{\vec{x}}\left(u\left(\vec{x}'\right)\right)\left(\vec{q}\right)\right]$$

$$u_s\left(\vec{x}\right) = \mathcal{F}_{\vec{q}}^{-1}\left(\mathcal{F}_{\vec{x}}\left(g\left(\vec{x}, \vec{x}'\right)\right)\left(\vec{q}\right) \mathcal{F}_{\vec{x}}\left(o\left(\vec{x}'\right) u\left(\vec{x}'\right)\right)\left(\vec{q}\right)\right)\left(\vec{x}\right) \tag{111}$$

The result in equation 111 is known as the Fourier diffraction theorem [45, equation 108] and it points to a way to compute diffraction in an efficient manner. Discrete Fourier transforms can approximate continuous Fourier transforms and can be accelerated using the FFT. This way diffraction can be simulated by Fourier-transforming the product of source and object, multiplying it with the Fourier transform of Green's function, and inversely transforming the product to retrieve the diffraction. Figure 7 shows this process on the projected electron density of 1,3-Butanediol and a planar incident wave. First the Fourier Transform of Green's function was discretized onto a grid. Mapping the phase to the hue and the amplitudes to the saturation, figure 7b is a visualization of equation 108 in 2 dimensions. The FFT gives Green's function in 2 dimensions, which looks like a spherical wave in figure 7a. This is a dead end in the computation and only depicted for completeness, the process will proceed with the Fourier representation for efficiency. The plane wave in x direction has a constant amplitude but a varying phase. The product of this plane wave and the object has the effect of patterning the object with its phase variation, which can be seen in figure 7c. The convolution in equation 110 of the product of object and source with Green's function can be computed efficiently by way of a multiplication in Fourier space. Figure 7f is the element-wise product of figure 7b and figure 7d. Its Fourier transform gives an approximation of the scattered signal $u_s\left(\vec{x}\right)$ in figure 7e as laid out in the Fourier diffraction theorem (remembering that polarization effects are still neglected).
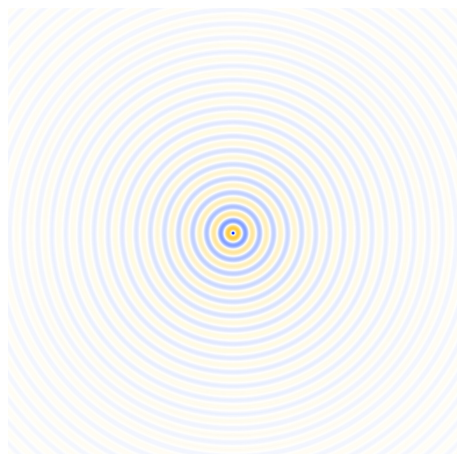
(a) Green's function – a spherical wave
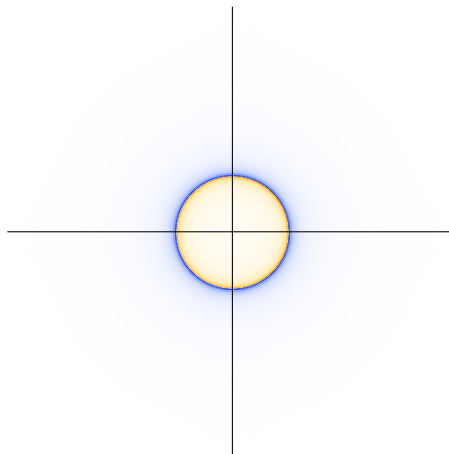
(b) The Fourier transform of Green's function.

(c) Product of source and object.
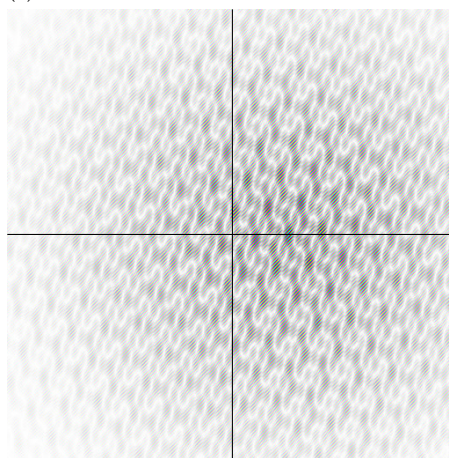
(d) Fourier transform of illuminated object.
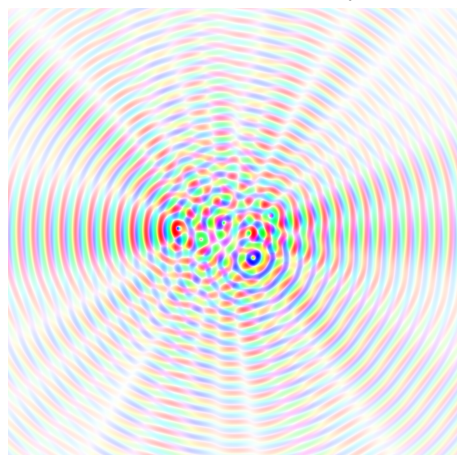
(e) A simulation of the diffracted field.
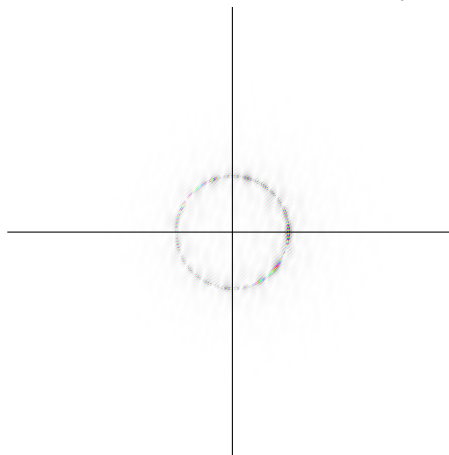
(f) Fourier transform of the diffracted field.

Figure 7: Simulation of diffraction using the Fourier diffraction theorem on a plane wave and the electron density of 1,3-Butanediol. Color saturation encodes amplitudes and the hue encodes the phase. The crosshair marks the origin in Fourier space, but not in realspace, as shifts are not relevant there. The object in this example is the electron density of 1,3-Butanediol projected along one axis. The source was chosen to be a plane wave.

(a) Phases of the simulated diffraction using a discretization of Green's function.

(b) Phases of the simulated diffraction using only the Fourier components of the object that lie on the Ewald sphere.
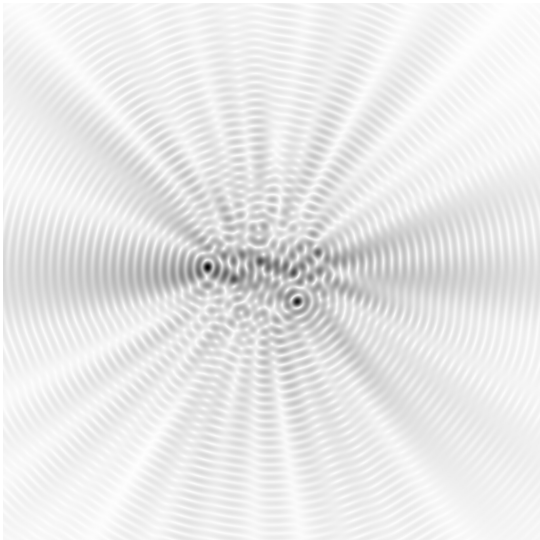
(c) Amplitudes of the simulation using a discretization of Green's function.
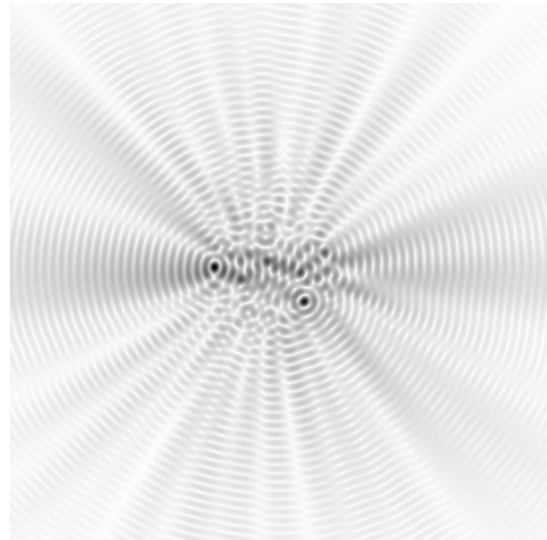
(d) Amplitudes of the simulation using only the Fourier components of the object that lie on the Ewald sphere.

Figure 8: Simulating diffraction from the a projection of 1,3-Butanediol with Green's function and using the Ewald sphere construction. Note that the simulations have qualitatively the same amplitudes, but mostly unrelated phases.

5.5. The Ewald Sphere

The Fourier transform of the Green's function (equation 108) has a pole of order 1 at $k_0^2 = |\vec{q}|^2$. This condition forms a spherical shell and integrals approaching the shell from inside or outside the shell are divergent. That is why the multiplication in equation 111 essentially only depends on the values on this

very shell. In simpler terms and neglecting phase information, the Green's function stamps out a spherical shell from the shifted Fourier transform of the product of object and field. The spherical shell given by the condition is the Ewald sphere [48, Chap. 1.3.4]. Comparing the simulated diffraction using a discretization of Green's function as outlined in figure 7 with the approximation to diffraction that is attained by only leaving the Fourier components of the object that lie on the Ewald sphere, we can see in figure 8 that the amplitudes are very similar, while the phases are more or less uncorrelated.
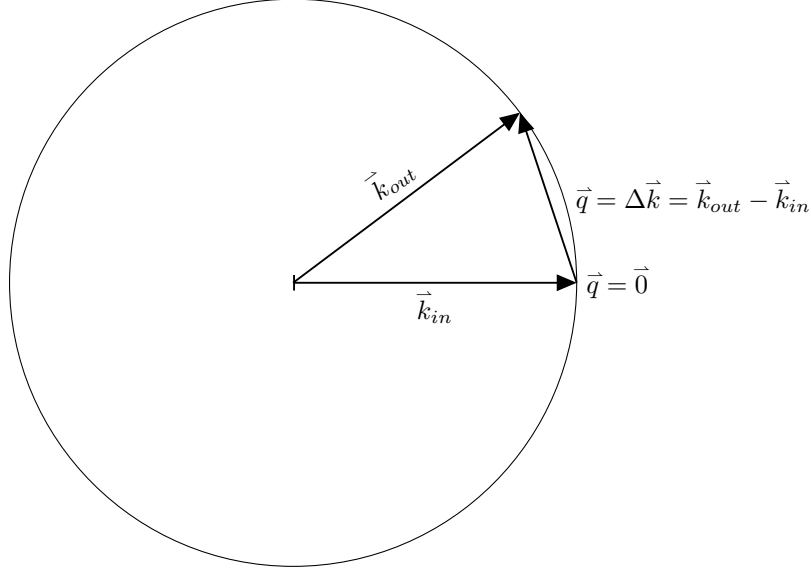
Figure 9: The Ewald sphere construction. The spatial frequencies of the object are selected by the diffraction condition, which is drawn as the Ewald sphere. All points on the Ewald sphere correspond to a difference between a diffracted wave vector $\vec{k}_{out}$ and an incident wave vector $\vec{k}_{out}$, both of which need to have the same length, reciprocal to the wavelength.

Not only are many sources well approximated by a plane wave, but those that are not can be composed by a sum or series of plane waves, and, due to the relation with the Fourier-transform, a plane wave gives a simple solution for the Fourier diffraction theorem. The plane wave needs to have the same frequency as the frequency used in the solution of the wave equation, or it would correspond to a different solution. If $u(\vec{x})$ is a plane wave with wavenumber k_0 pointing in the direction $\vec{w}_{in}$, its crystallographic wave vector is $\vec{k}_{in} = k_0 \vec{w}_{in}$, and its complex field amplitude is given by:

$$u\left(\vec{x}\right) = \exp\left(2\pi i \vec{k}_{in}^{\top} \vec{x}\right) \tag{112}$$

$$\mathcal{F}_{\vec{x}}\left(\exp\left(2\pi i \vec{k}_{in}^{\top}\vec{x}\right)\right)\left(\vec{q}\right) = \delta\left(\vec{q} - \vec{k}_{in}\right)$$

The Fourier transform of this plane wave is a delta function shifted by the wavenumber k_0. The convolution of the object with the source therefore amounts to a shift of the Fourier transform of the object (compare figure 7d) in the case of plane wave incident field, which leads to this particularly simple solution:

$$\mathcal{F}_{\vec{x}}\left(u_s\right)\left(\vec{q}\right) = \frac{\mathcal{F}_{\vec{x}}\left(o\right)\left(\vec{q} - \vec{k}_{in}\right)}{\left(2\pi\right)^2 \left(k_0^2 - |\vec{q}|^2\right)} \tag{113}$$

Because a plane wave just shifts the Fourier transform of the object, the Ewald construction chooses the coordinate system relative to the Fourier transform of the object, so that the $\vec{q}$ vectors directly correspond to spatial frequencies, see figure 9. Instead of the object, the Green's function, now the Ewald sphere, is shifted by the incident wave vector. While this point of view is often convenient because of its direct relation with the object, we need to keep in mind that it is not directly related to the Fourier transform of the diffracted field, precisely because of this shift.

5.6. Bragg's law - a far field approximation

All points on the Ewald sphere correspond to the plane waves that make up the diffracted signal. At the same time the points on the Ewald sphere correspond to Fourier components of the object. Because the Fourier transform of the object is shifted by $k_0 \vec{w}$, the Fourier components of the object on the Ewald sphere are orthogonal to the bisector between the incoming wave vector and the wave vector of the diffracted wave. Therefore we can see by geometric construction, that the corresponding spatial frequencies need to satisfy Bragg's law:

$$\lambda = 2d \sin(\theta) \tag{114}$$

With θ being the Bragg angle equal to half the angle of diffraction, see figure 10b.

Bragg's law can also be shown in real space. Assuming a source and detector far away, the path length difference between two paths of light diffracted at the object and observed at one point is asymptotically equal on all planes orthogonal to the bisector of illumination and diffraction, because the path differences form an isosceles triangle, with the two equal sides being the path difference once before diffraction and once after. The phase difference between two paths can therefore be derived from any two points on different planes, for example the two pine-green planes in figure 10a. If we choose the points to be as close as possible on the two planes, like in the figure, one can easily see that the path difference for parallel incoming and diffracted beams is equal to $2d \sin(\theta)$. For the maximum constructive interference between multiple parallel planes with the same distance, the path difference must be a multiple of the wavelength, and Bragg's law follows.
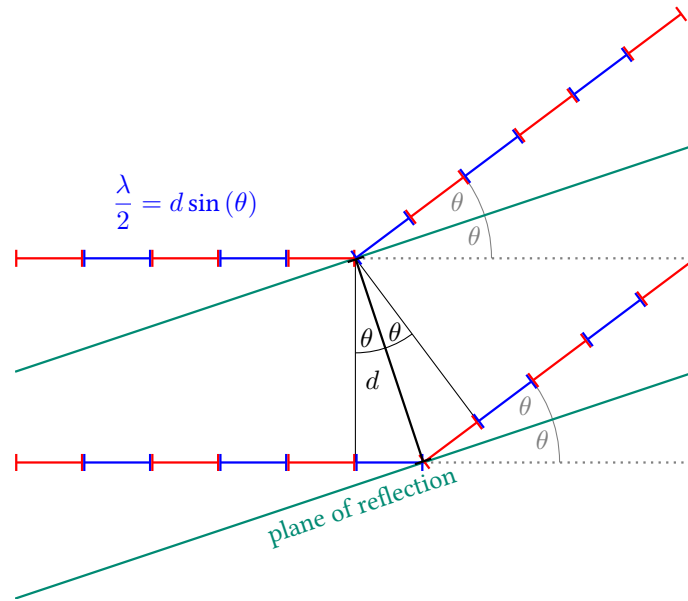
(a) Bragg's law in realspace: Radiation reflects on planes (Fourier components) within the object. When the distance between the planes d, the diffraction angle 2θ and the wavelength λ satisfy Bragg's law, diffraction from all parallel planes are in phase and maximal. The incident and diffracted beams are drawn with red and blue sections corresponding to the positive and negative swings of the wave.
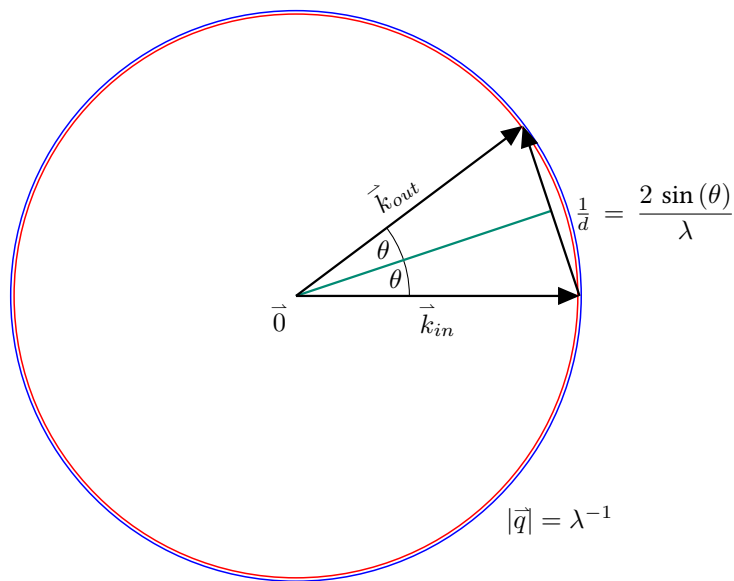
(b) Bragg's law in Fourier space: The difference between the incident wave vector and the diffracted wave vector (both with length reciprocal to the wavelength λ) corresponds to a distance d in the object. The incident and diffracted wave vectors form a equilateral triangle and therefore the length of the base is given by Bragg's law. Green's function is stylized with a red and blue double-struck line corresponding to the pole at $|\vec{q}| = \lambda^{-1}$, similar to figure 9.

Figure 10: Geometry of Bragg's law in real space and reciprocal space.

5.7. Wavefunction Collapse

The only bit of quantum theory we cannot do without is the quantization of electromagnetic radiation into photons. Matter waves are more obviously made up of quanta, like electron radiation and neutron radiation, and the Fourier diffraction theorem is an even more applicable approximation in that case, as matter waves are usually not polarized. But electromagnetic fields are quantized too. Quantum theory starts with the observation, that the difference between energy levels stored in oscillations fundamentally is quantized. And equivalently, waves can transfer energy only in quantized amounts. The unit of energy per oscillation is Planck's constant. This is hardly of any consequence in the previous analysis and especially for the result. But it has obvious consequences in any measurement of the diffracted field. When the field is observed, the observations require energy transfers, which come in the form of photons. The conversion between wave function and observation probability is the squared amplitude of the wave, just like the energy in a harmonic wave is proportional to its amplitude squared. This is called wave function collapse, because the wave function corresponding to this quantum ceases to exist, or rather is converted into a new state corresponding to the observation. Each observation is statistically independent from others, apart from the shared probability of observation. From an experimental point of view this means that the number of photons observed follows a Poisson distribution, which, not incidentally, is the minimum entropy distribution for a non negative integer variable with given expected value. From the observed photon counts and positions, the squared amplitude of the former wave function, that now has collapsed, can be inferred. But crucially, this means that without time resolution higher than the frequency of the waves, the phase is not observed directly.

5.8. Polarization Correction

We now have a scalar approximation for diffraction, but neglecting the term $\nabla \left(\vec{E} \nabla \cdot \log \epsilon \right)$ going from equation 98 to equation 99 did introduce some distortions. Neglecting it allowed us to treat all three components of the vector independently as a scalar each. However, the components are not completely independent, and electromagnetic waves are not a scalar quantity either. Some sound waves that propagate differences in pressure are scalar, but electromagnetic waves propagate directional electromagnetic fields. A change in the direction of the propagation of the electromagnetic field is not equally favorable in all directions. The orientation of the electromagnetic field is the polarization and the effects and distortions introduced by neglecting the vectorial nature of the fields are polarization effects. These effects can usually be accounted for, with a term for polarization correction depending on the diffraction geometry and the polarization only.

Radiation of any mixture of polarizations can be decomposed into orthogonal components. Therefore let us first consider radiation that is perfectly polarized. All charges that can be polarized are polarized to some extent by the electric field. The change in polarization in turn leads to secondary radiation. The exact interplay is given by the electromagnetic wave equations (93, 94), but for the polarization factor it is sufficient to look at the induced radiation from a single charge in isolation. An oscillating charge radiates like a dipole antenna, only the component of the radiation orthogonal to the oscillation can propagate, therefore the amplitude of the radiation depends on the angle between the two like the sinus of the angle between them [49, Chap. 6.2, 50, equation 45]. Another way to express this is to subtract the component

of the inducing field vector $\vec{E}_0$ that is collinear with the direction of secondary radiation $\vec{w}_{out}$ $(= \vec{k} \left| \vec{k} \right|^{-1})$.

$$\vec{E}_s \propto \vec{E}_0 - \left(\vec{w}_{out}^\top \vec{E}_0 \right) \vec{w}_{out} \tag{115}$$

By manipulating the expressions for the intensity of the resulting field $\left| \vec{E}_s \right|^2$, we get:

$$\left| \vec{E}_s \right|^2 \propto \left(\vec{E}_0 - \left(\vec{w}_{out}^\top \vec{E}_0 \right) \vec{w}_{out} \right)^\top \left(\vec{E}_0 - \left(\vec{w}_{out}^\top \vec{E}_0 \right) \vec{w}_{out} \right)$$

$$\left| \vec{E}_s \right|^2 \propto \left| \vec{E}_0 \right|^2 - 2 \left(\vec{w}_{out}^\top \vec{E}_0 \right)^2 + \left(\vec{w}_{out}^\top \vec{E}_0 \right)^2$$

$$\left| \vec{E}_s \right|^2 \propto 1 - \left(\frac{\vec{w}_{out}^\top \vec{E}_0}{\left| \vec{E}_0 \right|} \right)^2$$

Going back to the idea of decomposing any mixture of fields into orthogonal polarized components, we can define a variable p to be the degree of polarization of the incident field and $\vec{n}$ the normal vector to the plane of maximum polarization (containing $\vec{E}_0$ for a single component). Then, by adding the scattered signal of both components, we get:

$$C\left(p, \vec{n}, \vec{w}_{\text{in}}, \vec{w}_{\text{out}}\right) = p \left(1 - \left(\vec{w}_{\text{out}}^\top \left(\vec{w}_{\text{in}} \times \vec{n} \right) \right)^2 \right) + (1-p) \left(1 - \left(\vec{w}_{\text{out}}^\top \vec{n} \right)^2 \right) \tag{116}$$

And with this we have sufficient approximations for diffraction in crystallographic experiments. If anything, the level of detail is too great. Here the source and object were treated as arbitrary and abstract, in section 6.9 they will be approximated.

6. Structural Crystallography

Structural crystallography is one of the most prolific methods of determining macromolecular structures. As alluded to in the introduction, the molecular structure is uniquely informative for chemistry and biology. But all measurements have associated measurement errors and some cannot be avoided. The better the experiment and the associated errors are understood, the less expmerimental data will be needed to draw conclusions from it; up to a certain limit. Data processing nearly reaches the theoretical limit in some cases, but in others it is still far from it. In the following, after introducing structural crystallography in general, all steps of the data processing are revisited to identify possible improvements. Small advancements are described in section 6.6 that were achieved by comparing estimators that are less affected by outliers in detail. An incrementally better way to model the electronic structure of atoms is presented in section 6.7. The most significant contribution is the reduction of required diffraction data achived with an approximated diffraction model, presented in section 6.9.

Using the latest and most powerful radiation sources, diffraction signals from nanoparticles are already sufficient [51], but diffraction from single macromolecules still is too weak by some orders of magnitude, even while free-electron-lasers are getting more powerful. Unordered arrangements of identical macromolecules at best provide a radially averaged signal, but crystals are ordered and provide structural information and diffraction strong enough to be measured. Structural crystallography therefore needs to first produce crystals from the macromolecules, which is an art in itself, and then expose the crystals to radiation. The radiation diffracts at the crystal and the diffraction signal carries the structural information. For details on how the structure relates to the diffracted signal see the Fourier diffraction theorem (equation 111) and its derivation, but in short, the diffracted intensities are directly proportional to the Fourier Amplitudes squared of the object. This is why the Fourier coefficients of the object are called structure factors. For the short wavelengths needed for sufficient resolution of structural details, methods for measuring phase information directly are not available. Only intensities can be observed, which are equal to the amplitudes of the wave function squared. There are ways to reconstruct or substitute the phase information, see section 6.5.6. Together, amplitudes and phases determine the variation of the refractive index of the of the object, by way of the inverse Fourier transform. For electromagnetic radiation and electron beams, the refractive index depends on the electron density, while for neutron radiation, it depends on the nuclei themselves.

6.1. Crystals

Crystals are objects that have spatially ordered structures so that their Fourier transforms are sparse. For almost all crystals this means, that they can be thought of repeating building blocks in three dimensions. The vectors, whose integer linear combinations translate the building blocks to the periodic positions, are the unit cell vectors. The unit cell vectors are often unimaginatively called $\vec{a}$, $\vec{b}$ and $\vec{c}$. Together they form the basis of the unit cell matrix U. In addition, the building blocks can have internal symmetries. The minimum volume, that is needed to reconstruct the crystal using these internal symmetries and the lattice translations, is the asymmetric unit. Structural crystallography aims to determine the structure of crystalline compounds, by determining their symmetry and asymmetric unit using diffraction.

6.2. The Laue Equations

The Laue equations can be seen as an idealized form of the Fourier diffraction theorem (equation 111). Crystals with perfect three-dimensional periodicity have Fourier transforms with non-zero coefficients only for spatial frequencies at integer fractions of combinations of the unit cell vectors. Any other spatial frequencies are not aligned with the periodicity of the crystal. For planar incident waves this means that the difference between the diffracted and the incident wave vector must align with an integer multiple of the unit cell matrix, if diffraction is to be observed in this idealized experiment, see the Ewald construction in figure 9. This set of relations is expressed by the Laue equations (here in matrix form):

$$\boldsymbol{U}\Delta\vec{k} = \begin{pmatrix} h \\ k \\ l \end{pmatrix} \tag{117}$$

The integer vector $(h\,k\,l)$ is called the Miller index. Each structure factor can be addressed with this unique integer triplet. Given the unit cell parameters, initial crystal orientation and experimental geometry, the equation can be rearranged, to give the crystal orientation and the point on the detector where a given reflection can be observed most intensely. Conversely, for a random orientation of the crystal, the probability of any reflection (except the direct beam) to satisfy this diffraction condition is zero because the integer indices on the right side of equation 117 are an infinitesimal subset of the attainable rational vectors on the left side. Experimentally however, there is a neighbourhood close to the ideal diffraction condition where diffraction can be observed at reduced intensity even though the Laue equations are not satisfied. Not knowing which reflections will be observable for a given orientation and how intensely, is known as the partiality problem [52]. Section 6.9 will introduce a way to estimate that neighbourhood and the reduction in intensity, thereby addressing the partiality problem computationally.

6.3. Rotational Crystallography

In conventional rotation measurements each crystal is rotated, exposing it to the beam over a wide range of orientations, while integrating the diffraction over small angular wedges. Under those circumstances the Laue equations have been sufficient approximations for the diffraction condition. The rotation ensures that almost all reflections within the observable resolution range of the diffractometer will reach their optimum at some point during the rotation and can be fully recorded, thereby circumventing the partiality problem. The process of calculating any or all aspects of diffraction patterns (peak position, shape, intensity or full diffraction patterns), given unit cell parameters and experimental geometry is called "prediction" in the context of macromolecular crystallography data processing. For monochromatic rotational crystallography the deviations between measured and predicted peak positions are usually small, except for reflections whose reflection condition is not affected significantly by the rotation. (Those few measurements are typically discarded.) The rotation of the crystal during the exposure about a known axis and with a known angular increment acts as a strong constraint for parameter estimation during the processing of rotational crystallographic data. Using this information, the intensity of a reflection can be integrated and corrected to yield the corresponding squared structure factor amplitude.

6.4. Information Content in a Crystallography Experiment

Aside from the unit cell dimensions with almost negligible information content, the information we are after is in the expected intensity of the structure factors. The structure factors carry the structural information. Viewing the object as a three-dimensional periodic function, the Laue conditions precisely allow sufficient sampling, as introduced in section 4. However, when observing squared amplitudes only, the number of constraints is halved, making structure determination from crystal diffraction inherently dependent on prior or supplementary information. This conclusion will differ in section 7 on continuous diffraction. Typically, the atomic composition, connectivity and bond lengths and angles, is the prior information. Nevertheless it can be interesting to know how much information can be found in an experiment. Knowing that the molecules are made up of atoms, but not knowing where the atoms are and how well ordered the molecules are in the crystal, we know that each structure factor not constrained by symmetry follows a Gaussian distribution, because it is the sum of a relatively large number of values with finite variance. The amplitude squared of a complex Gaussian random variable is the sum of two independent Gaussian variables squared, and therefore distributed as an exponential distribution, which follows from the χ^2-distribution (equation 22). This argument was first presented in Wilson [53] and is since known also as the Wilson distribution. With this assumption about the prior distribution we can calculate the information gained using the Kullback-Leibler-divergence between the prior and the posterior distribution. For a step-by-step calculation, see appendix F, but for now suffice it to say that a typical observation with a signal-to-noise ratio of 10 will yield an information gain of about 3 bit, and an exceptionally well observed intensity with a signal-to-noise ratio of 50 will yield about 5 bit. Thus, the typical information content of a macromolecular crystallography experiment is on the order of 1MB.

6.5. Common Crystallographic Data Processing Procedure

There is no necessity to process crystallographic data in any particular way, but the following procedure has proven to be a successful framework against which any new method is measured. The data processing starts with diffraction images and ends with a set of phased structure factors and a model structure. The detectors need to be calibrated, then, to determine the unit cell and orientation of the crystal, the reflections are indexed. Using the indexing solution, the reflections are integrated. Multiple observations of the same structure factor amplitude squared are merged into a single estimate and the missing phases are determined in the phasing step. The last step is the structure refinement, where a structural model is optimized to fit the observations as well as possible. In the following, the individual steps of this protocol will be explained in more detail.

6.5.1. Detector Calibration

Diffraction detectors measure signal strength as a function of position. The position should be well determined and the way the signal of the detector depends on the diffraction signal should also be known. Detector calibration is the process of determining the relation between detector signal and measured signal and it is conceptually the first step in crystallographic data processing, even though one might return from a later step and start over with a more precise expectation of the diffracted signal. And although some argue that detectors should be well calibrated from the factory, in practice this hardly ever is the case, except for single-photon counting detectors. Several types of detectors are in use for crystallographic experiments:

direct			indirect
electronic		photochemical film	phosphor screen + camera
integrating	photon counting	(naturally integrating)	(naturally integrating)

Table 5: Classification of the detector types commonly employed in crystallography.

Most common are single-photon counting detectors, that detect photons one-by-one sequentially using a pixel array of electrical sensors with threshold circuits. These detectors are popular because they have a high signal-to-noise ratio, high quantum-efficiency, and typically do not require much calibration after their initial calibration. They are, however, not suitable for fast serial crystallography experiments, because the threshold circuit takes longer to reset than each exposure, that way each pixel would saturate at one count [54]. Serial femtosecond crystallography uses electronically integrating detectors. The photons interact with the sensor material and release electrons, the total current is then read out at the end of the exposure and converted to a digital signal.

No matter which detector, there is usually some level of detector calibration involved in the crystallographic data processing pipeline, but the unique combination of precision and variability that can be corrected for, integrating detectors often can profit most from calibration efforts. A common way to calibrate these detectors is using reference signals and using histograms, the latter approach will be addressed in section 8.

6.5.2. Peak finding

The sparse Fourier transform of crystals leads to sharp peaks on the diffraction images in typical crystallographic experiments. This is used in the peak finding step. Simple image recognition algorithms are used to detect peaks, by estimating the noise level of the background and classifying pixels into foreground and background. Contiguous areas of foreground, if not too large, are typically coalesced into a single peak, of which the integrated intensity and center of mass is reported [55, 56, 57].

6.5.3. Indexing

In the indexing step, indexing algorithms find the smallest unit cell that can explain as many of the peaks, found in the peak finding step, as possible. Indexing algorithms typically use autocorrelations, Fourier transforms [58], and in challenging situations with known unit-cell parameters with an exhaustive search for the unit-cell orientation becomes feasible through a rotation space transform, discretization and histogram approach [59]. Several orientations of the unit cell can explain the same peak locations but with different peak intensities in some space groups. When these indexing ambiguities arrive, they can be resolved automatically in most cases, using clustering [4]. If they cannot be resolved, the complete dataset will be twinned, which reduces the contrast and information content of the dataset.

6.5.4. Integration

Using the indexing solution, an expectation of the diffraction pattern is formed, the prediction. The intensity at the expected positions and potentially along images in the rotation is summed up and the background is estimated from the surrounding area. Typically, the integration area is circular or rectangular, and the

background is estimated in a belt around said integration area. The estimated background level is extrapolated to the foreground area and then subtracted from the foreground sum. The standard deviation of the integrated intensity is formed by the sum of the extrapolated background error, estimated from Poisson photon noise and potential excess variability, and the Poisson noise of the foreground sum.

6.5.5. Scaling and Merging

The scaling step determines the scaling factors, that bring the observations from different crystals under different conditions on a common scale. Merging takes the observations and scaling factors and produces an estimate for the structure factors. A closer look at merging can be found in section 6.6. Each crystal can potentially be larger or smaller than the others. Each time the experiment is repeated, the source could fluctuate in intensity and the crystal could be aligned better or worse with the focus of the beam. Whatever it may be, it is very common, that the observations with the same indices between two experiments are related by a linear scaling factor, because the effects just described increase or decrease the intensity of all reflections in each image equally. The diffracted intensities in the peaks is proportional to the modulus square of the Fourier transform of the average electron density in the unit cells. If the unit cells differ from one to the other in a uncorrelated manner, the average electron density appears smoothed.

The Debye-Waller factor describes the apparent smoothing of the electron density $\rho\left(\vec{x}\right)$, caused by disorder in the crystal, has on the measured intensities $I\left(\vec{q}\right)$. The smoothing can be approximated by the convolution with a three dimensional Gaussian distribution, with the expected squared displacement correlation matrix $\boldsymbol{B}$:

$$\rho\left(\vec{x}\right) = \rho_0\left(\vec{x}\right) * \exp\left(-\frac{1}{2}\left(\vec{x}^\top \boldsymbol{B}^{-1}\vec{x} + \log\left(|2\pi\boldsymbol{B}|\right)\right)\right) \tag{118}$$

The measured signal intensity therefore appears dampened [60]:

$$I\left(\vec{q}\right) = \left|F_0\left(\vec{q}\right)\exp\left(-2\pi^2\vec{q}^\top \boldsymbol{B}\vec{q}\right)\right|^2 = \left|F_0\left(\vec{q}\right)\right|^2 \exp\left(-4\pi^2\vec{q}^\top \boldsymbol{B}\vec{q}\right)$$

This dampening factor is called the Debye-Waller factor or B-factor.

6.5.6. Phasing

Because phases are not directly available experimentally, but essential for the Fourier inversion, a multitude of alternatives has been invented. Molecular replacement uses the phases of another structure, and if its structure is sufficiently similar to the one of interest, the resulting electron density is sensible enough to be interpreted as the density of the new molecule. Direct methods use diffraction data with very high resolution and the prior information, that the electron density is essentially sparse at this resolution, because the electron density of atoms forms a sharp peak around the core. Under these conditions the signal strongly overdetermines the atomic structure, which enables direct methods [61, Chap. 16], among which iterative projection algorithms, like it is the case for diffuse diffraction in section 7 and multiple crystal forms [62]. Single anomalous diffraction (SAD) and multiple anomalous diffraction (MAD) exploit the behaviour of certain elements at certain wavelength to significantly shift the phase of the scattered wave when the wavelength is close to resonance. This phase shift breaks the point symmetry of the diffraction signal in

a predictable way that depends only on the position of the anomalous scatterers. When the anomalous scatterers are sparse enough and the signal accurate enough to resolve the small anomalous differences between point-symmetry related structure factors, the anomalous signal again is clearly overdefined and algorithms exist to automatically find the likely positions of the scatterers. The known positions then often restrict the phases of all structure factors enough, to produce an electron density estimate sensible enough to recognize more and more parts of the molecular structure.

6.5.7. Structure Refinement

Given an initial guess of the molecular structure, the atomic coordinates are varied, their average movement might be estimated, and unordered solvent is modeled, to maximize the correspondence of the computed structure factor amplitudes of the model $|F_c|$ to the observed structure factor amplitudes $|F_o|$. The electron density of molecules is approximated e.g. with Cromer-Mann coefficients, because calculating the electron density of molecules from first principles is very compute intensive and more precise than necessary. A closer look on the Cromer-Mann coefficients can be found in section 6.7.

6.6. Merging in detail

In the following, several steps of the crystallographic data processing pipeline are going to be revisited in more detail, starting with merging. As introduced in section 6.5.5, the goal of merging is to estimate the squared amplitude of each structure factor given a set of observations of said quantity. Each observation of a reflection, whether partial or integrated over a range of rotations, has an associated error estimate, first given by the photon counting statistic and the background variance around the peak, and then possibly an improved estimate given by an error model [63, 64]. If the experiments were perfectly reproducible and the only or predominant source of noise was the photon counting statistic, the unweighted average of the observed intensities would be a good estimator. Each observation would stem from the same Poisson distribution and have the same variance, therefore equal weighting is the only sensible choice. The variance of the average would be the sum of all photon counts divided by the number of observations, because the sum of independent Poisson variables is another Poisson variable. For non-scaled observations, these conditions are mostly met, and the unweighted average has worked fairly well. From a Bayesian perspective, the distribution parameter of the Poisson distribution can be inferred by multiplying the prior distribution and the likelihood functions of the independent observations X_i and then normalizing the product if necessary. The result is a gamma distribution with shape parameter $\alpha = 1 + \sum_i^N X_i$ and rate parameter $\beta = N$, if the prior for the distribution parameter λ is the improper uniform prior in the open range $[0, \infty)$. If we instead follow Jeffreys reasoning to an improper prior proportional to $\lambda^{-\frac{1}{2}}$, the shape parameter would be given by $\alpha = \frac{1}{2} + \sum_i^N X_i$.

When the experimental conditions cannot be controlled as precisely, each observation will have a different variance and will need to be multiplied with a scaling factor to bring them all to a common scale. For example in one setting the beam may be half as strong, so all observations will need to be scaled by the factor two and the variance of these observations will be four times as large. Once the observations are scaled, an unweighted average is not appropriate any more. Scaling an observation with a factor scales its associated variance with the same factor squared. The weights that minimize the variance of the mean of samples with different variance are inversely proportional to the variance of each sample. Assuming

observations X_i, corresponding variances σ_i^2 and scaling factors a_i, the weighted mean is:

$$\frac{\sum_i \dfrac{a_i X_i}{a_i^2 \sigma_i^2}}{\sum_i a_i^{-2} \sigma_i^{-2}} = \frac{\sum_i \dfrac{X_i}{a_i \sigma_i^2}}{\sum_i a_i^{-2} \sigma_i^{-2}}$$

So far everything is fairly consistent. A large scaled observation has proportionally more weight, as it should be. Two problems arise from the estimation of the variances. The distribution parameter is unknown, and so the variance has to be inferred from the observation just like the distribution parameter. If, as it is usually done because it is most convenient, the variance is estimated for each observation individually, the variance estimate could be zero, because even for a non-zero flux, observing no photons is possible. But zero variance breaks the weighted mean estimator, the weighted mean would always converge to zero eventually. Thus, the previous Bayesian reasoning serves as the justification for adding one photon count to the variance estimate ($x + 1$ instead of x).

However even then, observations that are larger just by chance have a systematically larger and too high variance estimate and vice versa for observations that are smaller just by chance. This biases the result to smaller values by about one photon count, see figure 11. For the most part, crystallography accepts this bias to keep the parallel and unidirectional nature of data processing, because the bias is usually small compared to the number of photons observed and the other sources of noise. A solution would be to update the variance estimate with the new expected intensity after merging. The other components of the variance estimate of each observation are the photon noise of the background, other background noise and detector readout noise. Because of these other sources of noise, the photon count can sometimes appear to be negative and a common fix is using the absolute value instead ($1 + |x|$).
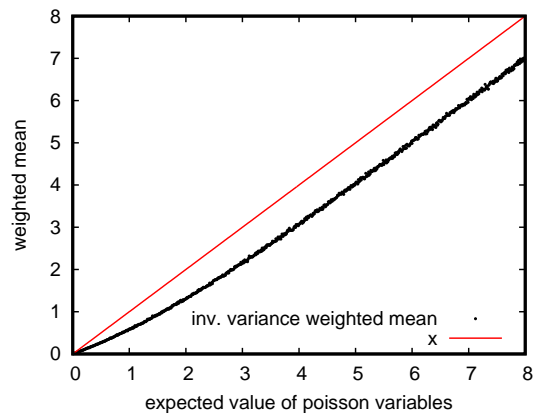
Figure 11: Highlighting the bias introduced when using the Poisson variance estimate $1 + x$ for weighted averaging. The graph shows the weighted mean of random values that were sampled from a Poisson distribution as a function of the distribution parameter as black dots. Without bias the mean values would cluster around the red line, indicating the ground truth of the numerical experiment.

To show the behaviour of different merging algorithms with experimental data, a typical dataset with very abundant observations was chosen, to investigate the rate of convergence using subsampling. A good merging algorithm is one that produces most similar estimates from different subsamples of the observations of the same structure factor, but a dissimilar estimate from a subsample of the observations of another structure factor. However, this is not a definitive measure, just an approximation, the only real measure
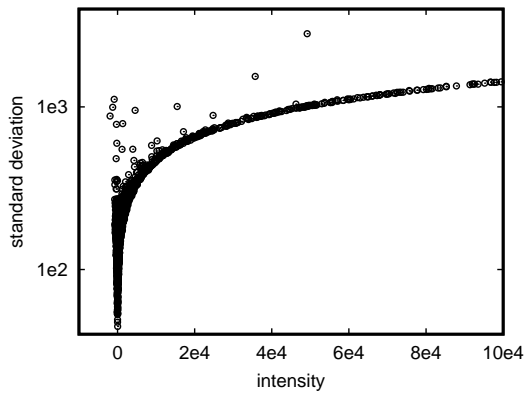
(a) Observed intensities and their standard deviation. As can be seen the standard deviation is at least proportional to the square root of the intensity due to the Poisson photon counting statistic.
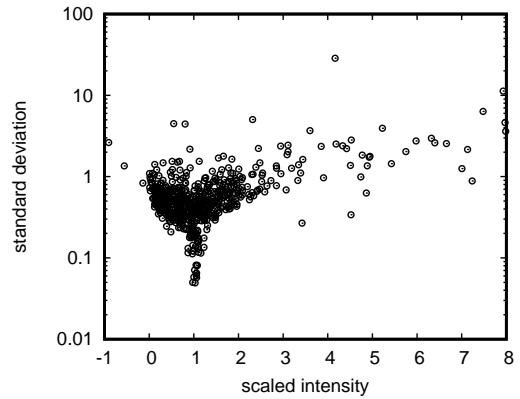
(b) Scaled observations and the corresponding standard deviation that is corrected with an error model.

Figure 12: Data for the reflection $(0\,0\,16)$ in the A_{2a} dataset before and after scaling. This is a typical distribution of observations with estimated variances, except for the extreme abundance of observations, which is the reason why it it was chosen.

crystallographers will accept is how well the merged structure factors can be interpreted as the result of a modelled molecular structure [65]. On the other hand, investigating the convergence rate of different merging algorithms allows a more informed search for a better merging algorithm. Experimental data deviates from the previous assumptions, which is why it is worth to try different approaches, and see how they behave in detail when there are values that are not likely explained by photon counting noise or similar expected phenomena.
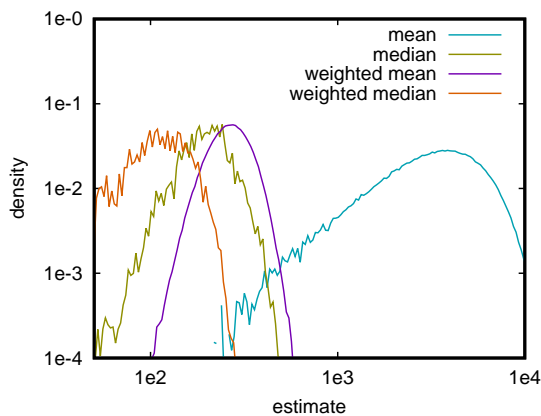
(a) Merging of raw data.
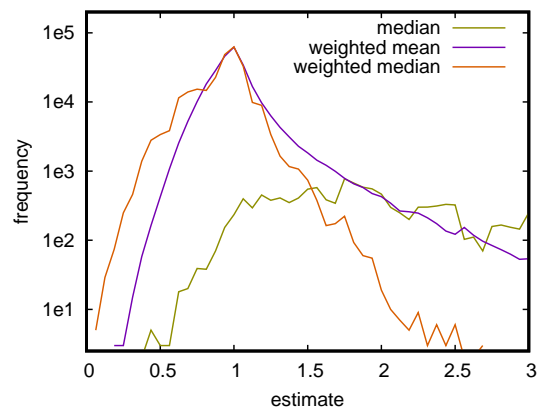
(b) Merging of scaled data.

Figure 13: Comparison of the four fundamental merging strategies in (a) for measured intensities, with errors estimated from background subtraction and photon counting noise, and (b) after scaling, partiality correction, and applying an error model.

The previously discussed bias is much worse when merging raw data using inverse variance weights, because the observations (partially recorded reflections) do not follow a Poisson distribution, not even

approximately. The distribution is much more similar to an exponentially modified Gaussian, and so there are frequent values with low intensity and low variance estimates, which biases the weighted estimates towards zero. Even though the weighted mean is the most consistent estimate in figure 13a, the unweighted mean typically gives better agreement with model structures during structure refinement, because the weighted estimates are compressed towards zero for all structure factors.

After scaling, partiality correction, and error modelling, there are frequent extreme outliers in the dataset, because the partiality estimates are often close to zero. For estimation of the full intensity, that is to correct for the partial nature of each observation, one would have to divide by the partiality estimate, thereby frequently creating extreme values, see figure 12b. Most extreme outliers have associated error estimates that reduce their impact in weighted merging strategies, but more robust strategies still have a lower variance of the result, because not all outliers have a proportionally large error estimate. This is why for scaled data, the histogram of the unweighted means could not be displayed in figure 13b, the estimator is too inconsistent. The median, because it is more robust, is slightly better. The weighted mean suppresses the outliers with a high variance estimate and is a fairly good estimate. The weighted median suppresses the high variance observations and it is very robust. Together, this makes it the most consistent estimator out of the four.

There are countless more advanced ways to estimate structure factors from diffraction measurements, but three are especially relevant for different reasons. Firstly, the median-of-means (a simplified instance of *random subsample consensus*), secondly, a recursive weighted average with a cutoff based on standard deviations, both of which are well-known techniques, and thirdly, a maximum-likelihood estimate based upon a mixture model, included because it fits into a more extensive maximum likelihood modelling approach. As discussed in sections 3.5 and 3.7, the mean is the most efficient estimator for square integrable distributions, and the median has a larger variance, but is robust as long as less than half of the samples are outliers. The median-of-means allows us to interpolate between the two, trading off some robustness for greater efficiency. Median-of-means computes the median of several means of random subsamples.

Figure 14: Comparison of advanced merging algorithms that try to combine robustness and efficiency. The maximum-likelihood estimate is for a mixture model with a Gaussian foreground and a Cauchy outlier distribution. The median of means is the median of size 16 subsamples.
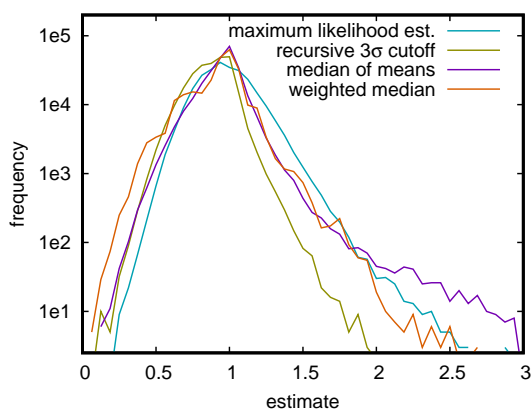
Figure 14 shows the median of subsamples of size 16, which happens to be a good compromise here. The size of the random subsamples is a hyperparameter, a parameter of the algorithm itself. Assuming the samples are either outliers or not, as an abstraction, there is a fraction of outliers ϵ. The plain median would be robust up to a fraction of $1/2$, but a single outlier will contaminate the mean. Therefore the optimal hyperparameter of the median-of-means is the size where the mean is computed from as many

samples as possible while statistically only half the subsamples contain an outlier, which is achieved with a subsampling-size just less than $(\log_2 (1 - \epsilon))^{-1}$. By that logic a subsampling-size of 16 corresponds to an outlier fraction of 4.2%. The recursive σ cutoff merging algorithm computes the mean and variance, in this case using the weighted versions, then excludes all samples that deviate more than a multiple of the standard deviations from the mean and the mean and variance are updated. This process is repeated until the set of samples does not change any further. The multiple is a hyperparameter, it can be be chosen to be any positive number, determining the trade-off between efficiency and robustness. Given an outlier fraction ϵ, a multiple of $\sqrt{2}\,\mathrm{erf}^{-1}(\epsilon - 1)$ includes the $(1 - \epsilon)$ most central samples without including the ϵ most extreme values. The commonly chosen multiple 3 corresponds to an assumed outlier fraction of 0.27%. The maximum likelihood estimator maximizes the following probability $\mathrm{p}(x)$ for all observations x using their variance estimate σ^2 by finding the optimal μ:

$$\mathrm{p}(x) = (1 - \epsilon) \exp\left(\frac{1}{2} \left((x - \mu)^2 \sigma^{-2} + \log\left(2\pi\sigma^2\right) \right) \right) + \frac{\epsilon}{\pi\left(1 + x^2\right)} \qquad (119)$$

It is a mixture distribution of a Gaussian and a Cauchy distribution for the outliers with proportion $1/16$. This estimator was included because a very similar target will be used later. All four algorithms compared in figure 14 performed fairly well. What is essentially just one data point is not sufficient to draw a conclusion when the difference is so small. Nonetheless the weighted median appears especially favourable, because it does not introduce a hyperparameter that needs to be determined first.

6.7. Cromer-Mann Coefficients

The Cromer-Mann coefficients are an approximate description for the electron density of atoms, and therefore play an important role in structure refinement (section 6.5.7). The electron density around atomic nuclei is mostly spherically symmetrical, because the electrostatic potential of the positively charged nucleus itself is spherically symmetrical. Atomic and molecular orbitals are just small nuances on top of this general trend. Taking the hydrogen atom as the most simplistic example, it can be derived analytically that the electron density in the ground state is exponentially decreasing with the distance from the nucleus. Experimental observations are in good agreement, however they show a slightly smoother electron density because of atomic movement and experimental limitations. Atoms with more electrons have a more varying radial distribution. To model electron densities and diffraction on an atomic scale, there is the need for computationally efficient approximations. Due to ever present measurement errors, the requirement for precision of the model is moderate and approximations are acceptable when they are more easy to work with. Cromer and Mann have noted, that the electron density of all atoms in all states can be described well with the sum of four Gaussian functions and a constant offset [66]. The coefficients of this fit are know as the Cromer-Mann coefficients. This is especially convenient, as Gaussian functions have analytical Fourier-transforms and derivatives. This approach is commonly employed when modelling the electron density and diffraction of molecules. The Cromer-Mann coefficients are mostly sufficient and have seen wide adoption, but there are three aspects that can be improved. Firstly, the constant offset is not required for a good fit and it implies a $\delta-$peak of the electron density at the core of the atom, which is unphysical, but more importantly, slightly inconvenient to render onto a discrete grid. Secondly, some of the variances as tabulated in IUCr [49] are negative. This means that the electron densities, that usually can be derived by Fourier transforming the Gaussian model, are not defined in this case. One of the visually most

Figure 15: Tabulated scattering factor of hydrogen atoms [49, table 6.1.1.1] as a function of the reciprocal space distance, in black dots, fitted with the sum of four Gaussian functions. In red, minimizing the logarithmic target in equation 120 and in blue, according to the tabulated parameters[49, table 6.1.1.4] minimizing squared deviations.
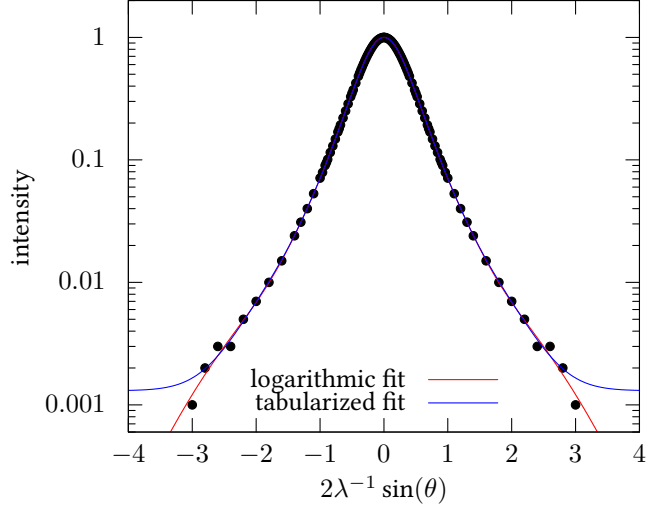
offending cases in this regard is the fit of the singly charged oxide ion, see figure 17. Thirdly, the tabulated fit is the result of the minimization of squared deviations. On the surface, this seems like the sensible thing to do, because minimizing squared deviations in real space also minimizes squared deviations in Fourier space and the other way around. Since the Fourier transform is linear, the transform of a signal with noise is equal to the sum of the transform of the signal and the noise individually. If the errors are uncorrelated and uniform, the Fourier transform will be uncorrelated and uniform. And according to Parseval's theorem the sum of squared errors in the Fourier domain are equal to the sum of squared errors of the signal. Therefore, an approximation minimizing squared errors minimizes (here only approximately because of uneven sampling) the squared errors of the approximation in the Fourier domain. But in practice, measurement errors are often proportional to the square root of the signal or even proportional to the signal instead of being constant. Small errors of a small signal often are more significant than the same error of a large signal. And this is why a new set of coefficients were derived by a constrained optimization procedure using the least squares fit of `gnuplot`, supplying it with relative errors. A small offset was added to avoid infinite weights for intensities approaching zero. Similar to the model of the Cromer-Mann coefficients, the tabulated scattering factors were approximated as the sum of four Gaussian functions, but without a constant offset. Each Gaussian function is multiplied by a factor a_i and scaled along $x = 2\lambda^{-1}\sin(\theta)$ with a parameter b_i.

$$
\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \\ a_3 & b_3 \\ a_4 & b_4 \end{pmatrix} = \underset{a_1,b_1,a_2,b_2,a_3,b_3,a_4,b_4}{\arg\min} \sum_{j=1} \left(\frac{y_j - \left(\sum_{i=1}^{4} a_i \exp\left(-2\pi^2 x^2 b_i^2\right)\right)}{\epsilon + |y_j|} \right)^2 \tag{120}
$$

The scale parameter of the Gaussian is naturally constrained to lead to positive variances because it appears squared in equation 120. Since the sum of the coefficients a_i is physically constrained to the number of electrons of the atom, the sum should be constrained in the optimization procedure also. The method of Lagrange multipliers exists for general constrained optimization problems, but in this specific case the parameters can be transformed as to eliminate one variable and make the constraint implicit. A suitable

57

parametrization that fixes the sum of a_i to s is the following:

$$a_1 = \left(\frac{s}{4} + a'_1 + a'_3\right)$$
$$a_2 = \left(\frac{s}{4} - a'_1 + a'_3\right)$$
$$a_3 = \left(\frac{s}{4} + a'_2 - a'_3\right)$$
$$a_4 = \left(\frac{s}{4} - a'_2 - a'_3\right)$$

The newly derived coefficients can be found in appendix H. They will prove useful in section 7 for modelling and optimizing the electron density of molecules, and the concept of using Gaussian functions as a basis will reoccur in section 6.9.

Figure 16: The electron density of a hydrogen atom projected onto an axis through the core. In black the solution of the non-relativistic Schrödinger equation [67, table 21–3], in red the Fourier transform of the logarithmic fit of the four-Gaussian model to the tabulated scattering factor [49, table 6.1.1.1] and in blue the Fourier transform of the tabulated fit. Note the δ-peak at 0 – a result of the constant offset of the tabulated fit [49, table 6.1.1.4].
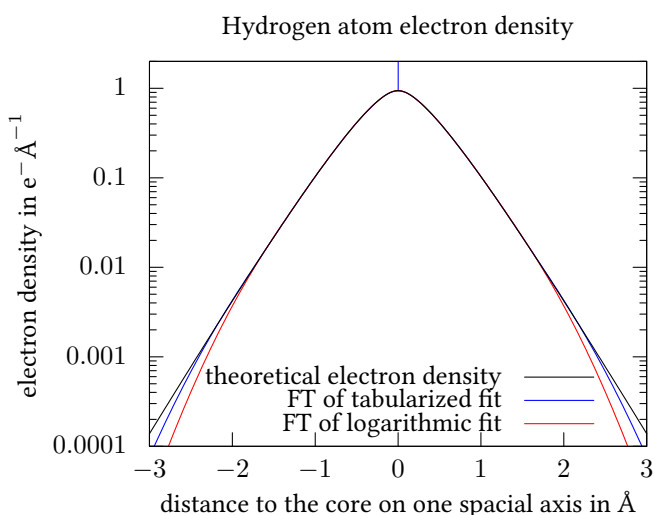
Figure 17: Tabulated scattering factor [49, table 6.1.1.3] of singly charged oxygen ions as a function of the reciprocal space distance in black and the corresponding tabulated fit [49, table 6.1.1.4] in blue and the new fit in red. Mind the logarithmic scale on the y-axis, visualizing relative deviations.
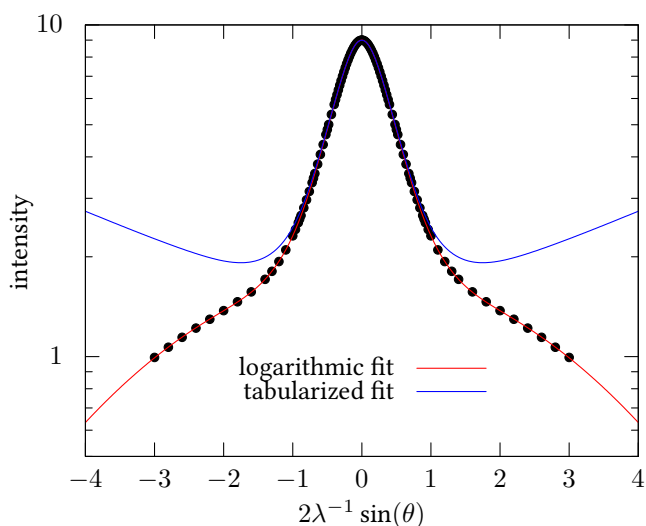
6.8. Serial Crystallography

In the last decades, methods have been employed in the field of macromolecular crystallography, which, for various reasons, deviate from the rotational crystallography setup in significant ways. The most notable among these methods is *serial crystallography*, where crystals are measured only once and consequently many crystals are needed for a complete dataset [68, 69]. An important subclass is *serial snapshot crystallography*, where the crystals are illuminated without rotation. Without the rotation it becomes indispensable to consider not just the ideal diffraction condition, but the partial intensity that can be observed when close enough to the ideal diffraction condition. We know there is a steep fall off of intensity with deviation from the exact condition in a monochromatic experiment with well ordered crystals. This steep fall off makes it easy to define a small range that contains almost all observations of the same structure factor and hardly any observations of anything else, even without knowing the shape of the fall off. Computing the average of these observations with unknown partiality is called Monte Carlo integration in the context of serial crystallography. It has been used to work around the problem of unknown partial intensities with great success [70]. However, for the Monte Carlo integration to converge to an average with a small standard deviation, each reflection needs to be measured multiple times. This approach assumes that the partialities follow the same distribution, with finite first and second moments, for all reflections of a given resolution shell. From this assumption it follows that the average converges to a value proportional to the non-partial intensity, similar to integrating line profiles in powder diffraction.

The development of new methods has not stopped there however. Serial snapshot crystallography has since been carried out with polychromatic, or so-called pink beam sources [71], electron beams [72] and mosaic crystals. More exotic experiments are surely already planned. In these more general cases the Laue equations are not sufficient, because inaccurate predictions of the peak positions and elongated peak shapes cannot necessarily be overcome by just measuring several times more data to make use of Monte Carlo integration. The Laue equations assume point-like peak shapes. In monochromatic experiments the peaks are narrow and compact, so small integration radii or boxes are typically employed, and the Laue equations are sufficient. But when two or more different and equally significant distributions are at play, elongated peak shapes can be observed.

Figure 18 depicts the distributions that are assumed to be relevant and their effect on the diffraction geometry. In polychromatic experiments the distribution of wavelengths, of which the width is called *bandwidth* (pink arrows), together with a distribution in crystal orientation, called *mosaicity* (brown arrows) can lead to elongated peak shapes. The other relevant distributions affecting the diffraction are the size and shape of the crystal (*reciprocal peak size* in green), *convergence* (or *divergence*) (red arrows) of the beam, and different *strain* (cyan arrows) throughout the crystal (that is a variation of unit cell parameters throughout the crystal volume). Once there is more than one relevant distribution, the exact location of the peak on the detector can no longer be determined solely by rearranging the Laue equations.
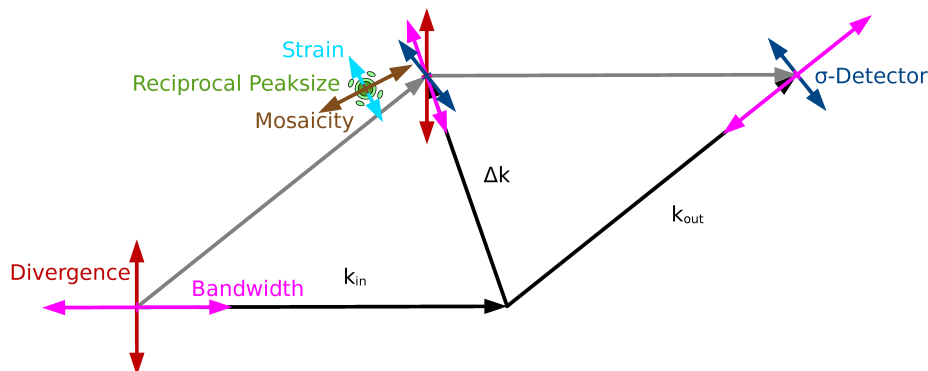
Figure 18: The geometric construction used to generate the covariance matrices of the distributions of diffractive power in reciprocal space and the volume probed by an incident beam. The arrows indicate the components, akin to error bars, that the different distributions contribute to the covariance matrix in a 2D cut. The same contributions have a different effect on k_{in}, Δk and k_{out}, and where they have an effect they are indicated with the same color as where they were introduced. Also compare figure 9

6.9. Gaussian Partiality Modelling and Diffraction Prediction

As alluded to, serial crystallography could profit from a model describing diffraction from crystals in more detail than the Laue equations, but more directly than the Fourier diffraction theorem. If partialities could be estimated, their influence on the merged estimates would not have to be marginalized with so many observations, but it could be corrected for, which would increase the amount of information that can be gained from each diffraction image, which can also be seen in the much faster convergence of the estimates during merging of corrected observations (compare figure 13a and 13b).

6.9.1. Previous approaches

The earliest approaches to dealing with partially recorded reflections relied upon the redundancy afforded by rotation experiments, which makes them unapplicable in serial crystallography. Under those conditions the partiality as a function of the crystal rotation can be reconstructed as a smooth function, because it is overdetermined by the diffraction data. And using the reconstructed profile, the partially observed reflections can be corrected [73, 74, 75].

An early approach in dealing with partial reflections that can be applied to single diffraction patterns [52] assumed reciprocal peaks to be spheres. While the diffraction process is modelled similar to the earlier approaches with the intersection of these small spheres with the Ewald sphere, here the rocking curve is determined entirely by the intersection of the Ewald sphere with the reciprocal lattice spheres. The reduction allows us to use this model even for single diffraction patterns. T. J. Greenhough and J. R. Helliwell continued this approach and have generalized it to ellipsoidal shapes [76, 77, 78]. Andrews et al. [79] showed that this approach can even be applied to Laue diffraction (with very high polychromaticity). The model of Rossmann et al. was generalized by Ginn et al. [80] with a super-Gaussian distribution of Ewald spheres given by the distribution of wavelengths and incidence angles, requiring a numerical integration that is efficiently implemented in CrystFEL [81] as the partiality model `xsphere`. This model has 11 free parameters per crystal in total: 9 for the unavoidable unit cell matrix and one each for the mosaicity radius

and the profile radius.

Holton et al. [82] modelled the most relevant contributions, save the crystal shape transform, based on the principles layd out in Greenhough & Helliwell [83] and Winkler et al. [75] (modelling mosaicity with the intersection of a disk with the Ewald sphere). They also used Gaussian basis functions, but instead of analytical integration of the different distributions, they computed numerical integrals to combine different effects with automatic sampling. No attempt to match measured diffraction data with the proposed model was described, on the contrary the message of the publication was the "untapped potential" that should be realized if a method could be found to fit the simulation to experimental data.

The program package nXDS [50] is another software suite to process serial crystallographic data. The partiality model used assumes an isotropic Gaussian decay of the partiality with the angular offset from the ideal diffraction condition, making for simple symbolic expressions using Gaussians in one dimension and a straightforward optimization of the parameters.

A different approach to computing the integrals that are required for estimating the partiality of reflections in still diffraction patterns uses ray-tracing principles [84]. This approach is much closer to what would be called Monte-Carlo integration outside of crystallography.

An isotropic and simplified partiality model using multidimensional but isotropic Gaussian basis functions has been implemented in CrystFEL and is the default for predicting spot locations and qualitative visibility since version 0.9.0. It uses a simplified version of equation 146 below, but without squaring the exponential term. The scalar projection of the covariance matrix orthogonal to the Ewald sphere is especially simple to calculate in this case. This model can also be used as a partiality model like xsphere and it is selected with the keyword ggpm. This model is most comparable to the one used in nXDS [50]. Notable differences to that model are the formulation using the three-dimensional Gaussian function and the concept of reciprocal peak width, which ascribes an additional constant width to peaks in reciprocal space independent of beam parameters and mosaicity, an effect that is especially significant at low resolution.

The Gaussian-like appearance of peaks on the detector possibly inspired Mendez et al. [85] to impose a Gaussian decay of intensity with distance from the ideal diffraction condition on the detector. The result in equation 4 of Mendez et al. [85] is seen to be proportional to a special case of equation 128 of this work when the covariance matrix Σ_o is uniform in all dimensions and scaled appropriately. Conversely, the result presented in this paper can be seen as a multidimensional generalization of the approach of Mendez et al. [85]. The significance of this difference becomes most obvious when considering elongated peak shapes in pink-beam experiments, which cannot be modelled by the approach of Mendez et al. [85], owing to the isotropic nature of that model.

Dilanin et al. [86] imposed a peak shape on the detector to fit the whole pattern in a similar manner to Mendez et al. [85], but instead of an isotropic Gaussian shape they used an isotropic pseudo-Voigt shape. Pseudo-Voigt functions allow more heavy tailed shapes, and are thereby able to match shape transforms better with their asymptotically inverse-quadratic decay. However, their derivation does not connect these peakshapes with anything but the shape transform of the crystals. Our method generalizes a similar approach to non-isotropic peak-shapes and connects them to mosaicity, non-monochromaticity, the crystal shape transform, the convergence and allows arbitrary compositions thereof. However, it is less general in the sense that only Gaussian shapes are employed. This is a deliberate limitation, because of the analytical difficulties that would be encountered with operations on anisotropic Cauchy distributions.

6.9.2. Inspiration

There are three reasons why one would want to approximate diffraction using Gaussian basis functions. Firstly, they can be combined to describe even distributions that are not Gaussian, as seen in section 6.7 with the Cromer-Mann coefficients, and they are convenient, because the Fourier transform of a Gaussian distribution is another Gaussian. Secondly, a Gaussian distribution is the maximum entropy solution given simplistic assumptions:

1. There is an optimal diffraction condition where the highest intensity is to be expected

2. Measured intensities are non-negative integers and have a finite expected value

3. The fall-off of intensity off the ideal diffraction condition is not immediate (otherwise virtually no reflections could be observed) and the distance of observed intensities to the ideal diffraction condition has a finite expected value and finite variance

From this point of view with limited knowledge, the expected intensity of a given reflection is seen as a probability distribution dependent on $\Delta \vec{k}$. Going by the principle of maximum entropy, the distribution that makes the least assumptions about the intensity values at a given offset from the optimal diffraction condition is the Poisson distribution because this is the maximum entropy distribution of a positive integer variable with finite variance. Experimentally, the same intensity with the same experimental geometry indeed does follow a Poisson distribution, but with different crystals in a very similar orientation and with different beam intensity and alignment, the intensity often approximately follows an exponential distribution convoluted with a Gaussian measurement error, resulting in an exponentially modified Gaussian distribution. This is because small deviations can have seemingly chaotic effects and the experimental geometry cannot be controlled precisely enough.
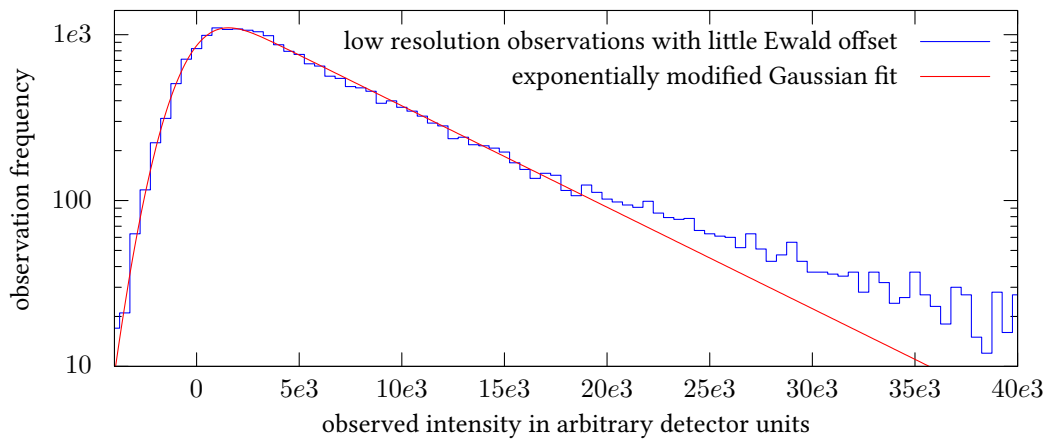
Figure 19: The observation frequency of low resolution intensities with a small offset from the Ewald sphere (less than 1×10^{-5} Å^{-1}) looks like an exponentially modified Gaussian distribution. Such a distribution is expected when an exponentially distributed quantity is measured with some random measurement errors.

From these simplified assumptions it also follows that the dependence on the offset from the optimal diffraction condition should resemble a Gaussian profile. This is because the Gaussian distribution is the maximum entropy distribution for a random real vector with finite mean and variance, or, from an even more restricted point of view, only knowing the absolute distance to the optimal diffraction condition, an isotropic Gaussian profile. Experimentally, we can observe the approximately Gaussian profile as the main trend of the intensity of a single reflection from a large crystal recorded at room-temperature, in very fine rotation increments, see figure 20. A trend like this is typical for room-temperature crystals, cryogenic crystals show a similar trend on average, but typically with even more extreme deviations. The fact that experimental results can be approximated with Gaussian functions is the third reason. Unfortunately, the deviations from the Gaussian profile are significantly greater than what can be explained by measurement errors, thus a Gaussian rocking curve is only a first-order approximation. For small crystals, there cannot be sufficient observations to measure such a profile directly, because they suffer from radiation damage too much for repeatable experiments. But from a large set of small crystals and reflections, an average profile can be determined, by accumulating the mean value of the intensity of the refletctions as a function of the Ewald offset, the distance to the closest point on the Ewald sphere. It is the minimum difference between the left and right-hand side of equation 117 when the direction of $\vec{k}_{out}$ can be chosen freely. Because the crystals are not perfectly homogeneous and the indexing solutions are not equally accurate, we expect an average of many rocking curves with different widths and slightly shifted off centre. As shown in figure 6.9.2 for observations of reflections in several resolutions shells, the average intensity is highest where the Laue equation is fulfilled and is rapidly declining. The observed profile is similar to a Gaussian near the centre, but the tails are more like that of a Laplacian distribution. This could be due to the tails of the individual profiles themselves or it could be an artefact of averaging. Attempts to improve the rate of convergence in merging by using the average profile and with a better understanding of the probability distributions that are involved, only lead to minor improvements. Naturally, the next step is to model the partialities for each crystal individually.

Figure 20: A typical intensity profile of a reflection under rotation, here the reflection (0,1,9) of lysozyme at room temperature, generously measured by Dr. Janina Sprenger, research fellow at CFEL. It can be approximated by a Gaussian profile, but the deviations are larger than the measurement error.
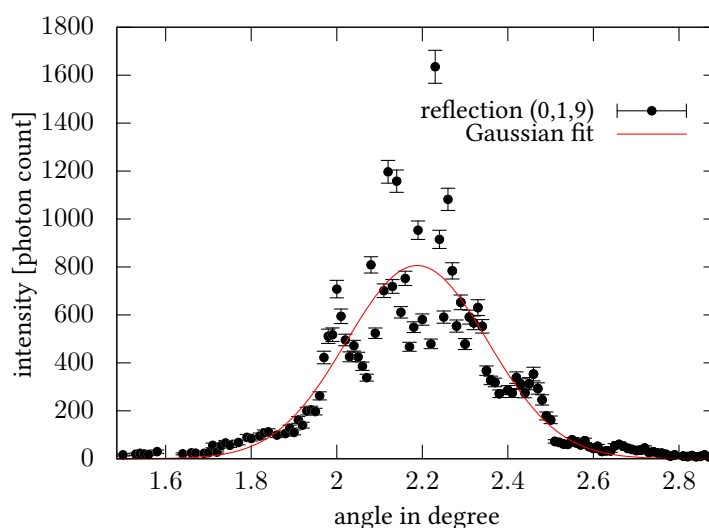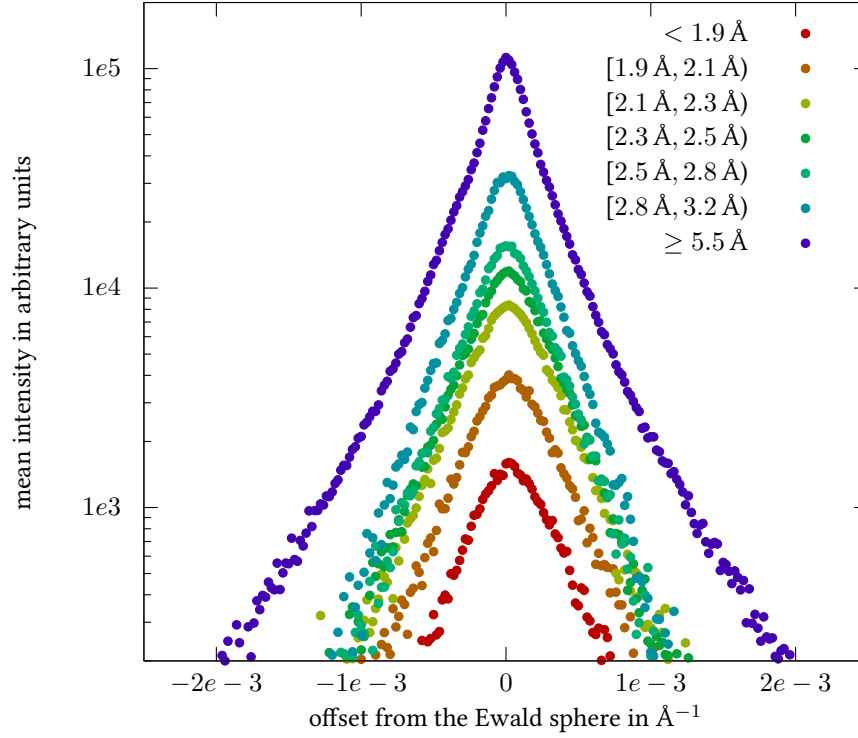
Figure 21: Average observed diffraction intensity of granulovirus hull protein crystals as a function of the distance to the Ewald sphere as given by the indexing solution. Imagine a distribution like in figure 19 for the actual values in the vertical direction for each dot. Different colors indicate the trends for different resolution ranges.

To summarize, the rocking curve is Gaussian according to the principle of maximum entropy and simplistic assumptions. There is some experimental support for this assumption. If the result can be approximated by Gaussian functions, which happen to be a fixpoint of the Fourier transform, can the diffraction process be approximated with these functions too?

6.9.3. Derivation

The incident wave interacts with a three-dimensional object, which is described by its scattering potential, which in turn is mainly determined by its electron density ρ. In the Born approximation and a monochromatic incident wave with flux J_0 (in units of energy per area), the photon flux density j (in units of energy density as a function of solid angle) at each point on the detector can be described as the Fourier transform of the electron density $O(\Delta \vec{k})$, evaluated at points corresponding to the difference $\Delta \vec{k}$ of the incident wave vector $\vec{k}_{\text{in}}$ and scattering wave vector $\vec{k}_{\text{out}}$, a term C correcting for polarization effects [87] and the scattering cross-section as a proportionality constant. The vectors $\Delta \vec{k}$ lie on a sphere with a radius ν reciprocal to the wavelength λ, see figure 9 for the Ewald construction and see section 5 or Slaney and Kak [88] for a derivation of the Fourier diffraction theorem.

In this approximation, diffraction is a linear operation, which means that the superposition principle applies to the complex wave function of the diffraction. The diffraction of several objects is the sum of the diffraction of these objects. The diffraction of an object by multiple sources is the sum of the diffraction

of the object by each source. Depending on whether there is a fixed phase relation between the different contributions to the total diffraction, the contributions add incoherently (assuming an integration over a time interval several times the duration of the oscillation of the wave), that is as modulus squares, or coherently, which is in the complex domain, before the modulus square operation. For a derivation of the resulting average amplitudes of coherently and incoherently interacting waves see section 1.3.2. of Cowley [87].

6.9.4. Decomposition Into Gaussian Basis Functions

Distributions of the sources and the objects are just an even further generalization of the superposition principle; combining these distributions amounts to convolutions of the distributions. Three-dimensional integrals of distributions over potentially curved paths however do not, in general, have a closed solution. Numerical solutions are easy to determine, but compounded, derivative or derived properties (such as those required for least squares minimization) grow in complexity, exponentially. Once one step is numerical, the next steps will most likely have to be numerical too. It is therefore useful and more insightful to have simple closed-form approximations. Gaussian distributions, as well as products and sums thereof, have closed and simple integrals when integrated over the whole domain or along a cut or a projection. Such integrals can likewise be expressed as a sum of Gaussian functions and a constant term. Also their Fourier transforms are well behaved. This way, integrating over multiple distributions still increases the complexity of the result, but starting from a less complex baseline. This means that if one can express all distributions in the model as a sum or series of Gaussian functions, the conditional integration of the resulting distribution can be achieved symbolically. While not every distribution is suitably expressed as a weighted sum of Gaussian distributions, a large family is [89]. Many natural distributions belong to this family. And for most distributions used in the application of the method discussed here, the number of Gaussian basis functions, for sufficient approximation, is very low. See section 3.8 for an overview of relevant identities for products, squares and convolutions of Gaussian distributions.

The identities of equations 63 and 64 can be used to compose the expected flux in a particular diffraction direction from the individual contributions of the source and of the object (see figure 18). As mentioned above, the formulation of this composition depends on whether the distributions are assumed to be in a fixed phase relation (coherent), or to have a randomly varying and uncorrelated phase shift (incoherent). The following two identities, each first expressed using exponential functions and then in terms of ϕ (see equation 62), are at the core of the method for analytical integration used in this work. The first is the

integral of the product of two Gaussian densities, which is then squared (for coherent integration):

$$\left[\int_{\mathbb{R}^n} \exp\left(-\frac{1}{2}\left((\vec{x}-\vec{\mu}_1)^\top \boldsymbol{\Sigma}_1^{-1}(\vec{x}-\vec{\mu}_1) + \log\left(|2\pi\boldsymbol{\Sigma}_1|\right)\right)\right)\right.$$
$$\left.\exp\left(-\frac{1}{2}\left((\vec{x}-\vec{\mu}_2)^\top \boldsymbol{\Sigma}_2^{-1}(\vec{x}-\vec{\mu}_1) + \log\left(|2\pi\boldsymbol{\Sigma}_2|\right)\right)\right) d\vec{x}\right]^2$$
$$= \exp\left(-(\vec{\mu}_1-\vec{\mu}_2)^\top \boldsymbol{\Sigma}_\circ^{-1}(\vec{\mu}_1-\vec{\mu}_2) - \log\left(|2\pi\boldsymbol{\Sigma}_\circ|\right)\right) \tag{121}$$
$$\left[\int_{\mathbb{R}^n} \phi(\vec{x},\vec{\mu}_1,\boldsymbol{\Sigma}_1)\,\phi(\vec{x},\vec{\mu}_2,\boldsymbol{\Sigma}_2)\,d\vec{x}\right]^2$$
$$= \left[\phi(\vec{\mu}_1,\vec{\mu}_2,\boldsymbol{\Sigma}_\circ)\right]^2$$
$$= \phi\left(\vec{0},\vec{0},2\boldsymbol{\Sigma}_\circ\right)\phi\left(\vec{\mu}_1,\vec{\mu}_2,\frac{1}{2}\boldsymbol{\Sigma}_\circ\right)$$

For incoherent integration the integration and squaring operations are reversed:

$$\int_{\mathbb{R}^n} \exp\left(-\frac{1}{2}\left((\vec{x}-\vec{\mu}_1)^\top \boldsymbol{\Sigma}_1^{-1}(\vec{x}-\vec{\mu}_1) + \log\left(|2\pi\boldsymbol{\Sigma}_1|\right)\right)\right)^2$$
$$\exp\left(-\frac{1}{2}\left((\vec{x}-\vec{\mu}_2)^\top \boldsymbol{\Sigma}_2^{-1}(\vec{x}-\vec{\mu}_1) + \log\left(|2\pi\boldsymbol{\Sigma}_2|\right)\right)\right)^2 d\vec{x}$$
$$= \exp\left(-(\vec{\mu}_1-\vec{\mu}_2)^\top \boldsymbol{\Sigma}_\circ^{-1}(\vec{\mu}_1-\vec{\mu}_2) - \log\left(|2\pi\boldsymbol{\Sigma}_\circ|\right) - \frac{1}{2}\log\left(|4\pi\boldsymbol{\Sigma}_*|\right)\right) \tag{122}$$
$$\int_{\mathbb{R}^n} \phi(\vec{x},\vec{\mu}_1,\boldsymbol{\Sigma}_1)^2\,\phi(\vec{x},\vec{\mu}_2,\boldsymbol{\Sigma}_2)^2\,d\vec{x}$$
$$= \int_{\mathbb{R}^n} \phi(\vec{\mu}_1,\vec{\mu}_1,2\boldsymbol{\Sigma}_1)\,\phi\left(\vec{x},\vec{\mu}_1,\frac{1}{2}\boldsymbol{\Sigma}_1\right)\phi(\vec{\mu}_2,\vec{\mu}_2,2\boldsymbol{\Sigma}_2)\,\phi\left(\vec{x},\vec{\mu}_2,\frac{1}{2}\boldsymbol{\Sigma}_2\right)d\vec{x}$$
$$= \phi(\vec{\mu}_1,\vec{\mu}_1,2\boldsymbol{\Sigma}_1)\,\phi(\vec{\mu}_2,\vec{\mu}_2,2\boldsymbol{\Sigma}_2)\int \phi\left(\vec{x},\vec{\mu}_1,\frac{1}{2}\boldsymbol{\Sigma}_1\right)\phi\left(\vec{x},\vec{\mu}_2,\frac{1}{2}\boldsymbol{\Sigma}_2\right)d\vec{x}$$
$$= \phi(\vec{\mu}_1,\vec{\mu}_1,2\boldsymbol{\Sigma}_1)\,\phi(\vec{\mu}_2,\vec{\mu}_2,2\boldsymbol{\Sigma}_2)\,\phi\left(\vec{\mu}_1,\vec{\mu}_2,\frac{1}{2}(\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2)\right)$$

In Eqns. 121 and 122 we have used the definitions:

$$\boldsymbol{\Sigma}_\circ = \boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2$$
$$\boldsymbol{\Sigma}_* = \left(\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1}\right)^{-1}$$
$$\vec{\mu}_* = \boldsymbol{\Sigma}_*\left(\boldsymbol{\Sigma}_1^{-1}\vec{\mu}_1 + \boldsymbol{\Sigma}_2^{-1}\vec{\mu}_2\right)$$

As can be seen from the above expressions, the difference between coherent and incoherent integration amounts to only a difference in scaling when both of the two distributions are single Gaussian distributions (that is, not sums of several Gaussians). As a simplification and because the linear scaling factor is hardly of any consequence, incoherent integration will be the default in the following, but the procedure can be applied with minor modifications for coherent integration as well. Partial coherence can be dealt with by splitting the coherent and the incoherent component into separate Gaussian functions and propagating them appropriately, or by interpolating between the coherent and the incoherent solution based on the degree of coherence, but this will not be considered any further in this work.

6.9.5. Parametrization of the Illumination

The diffraction condition, indicating the spatial frequencies of the object that contribute to the diffraction pattern, forms a spherical shell that passes through the origin, which was referred to above as the Ewald sphere. If the incident beam is convergent or divergent, there is a distribution of incoming directions, leading to a nest of spherical shells of equal radius in reciprocal space, whose centres lie on a spherical cap such that they all intersect at the origin. The normal at the center of this cap is parallel to the mean beam direction. The covariance matrix $\mathbf{\Sigma}_{\text{in}}$ of $\vec{k}_{\text{in}}$ due to convergence or divergence alone cannot really be simplified in general, but if the distribution is isotropic, it can be written as:

$$\mathbf{\Sigma}_{\text{in}} = \sigma_{\text{in}}^2 \nu^2 \left(\mathbf{I} - \vec{w}_{\text{in}} \vec{w}_{\text{in}}^{\top} \right) \tag{123}$$

where σ_{in} is the standard deviation of the incidence angles (i.e. the convergence), $\mathbf{I}$ is the identity matrix and vectors $\vec{w}$ are unit vectors describing beam directions, derived from the wave vectors $\vec{k}$:

$$\vec{w}_{\text{in}} = \frac{\vec{k}_{\text{in}}}{|\vec{k}_{\text{in}}|}$$
$$\vec{w}_{\text{out}} = \frac{\vec{k}_{\text{out}}}{|\vec{k}_{\text{out}}|}$$

Each beam direction, in theory, would need its own polarization correction, and this could be achieved by integrating the polarization correction term for all the beam directions, but as small angles are assumed, the polarization correction of the main beam direction is deemed sufficient for all.

If there are multiple sources with different wavelengths, i.e. if the wavelength distribution has a finite *bandwidth*, the Ewald spheres have different radii and consequently the distribution of sphere centres, previously on a spherical cap, is broadened radially. The 3D distribution of sphere-centres is approximated as a sum of Gaussian kernels. If the angular distribution is assumed to be small and independent of the distribution of wavelengths, it can be calculated by convolving the angle and wavelength distributions to form a cumulative distribution. The convolution of Gaussian kernels amounts to a summation of the respective covariance matrices, see equation 65.

The distribution of $\vec{\Delta k}$ that sample the Fourier transform of the object and contribute to diffraction in a given direction, i.e. a point on the detector, can be derived from the distribution of sphere centres. The distribution of $\vec{\Delta k}$ will be approximated as a Gaussian distribution with mean $\vec{\mu}_A$ and covariance matrix $\mathbf{\Sigma}_A$. Since the diffraction process does not change the wavelength, the outgoing wave distribution is perfectly correlated in wavelength with the corresponding incoming wave distribution. Differences of fully correlated Gaussian distributions require taking the difference of the square root of the respective covariance matrices. Given $\vec{k}_{\text{in}}$ and $\vec{k}_{\text{out}}$ are approximated as Gaussian distributions, $\vec{\Delta k}$ is distributed as a Gaussian around the mean value $\vec{\mu}_A$ corresponding to the difference of the mean of $\vec{k}_{\text{out}}$ and $\vec{k}_{\text{in}}$. The covariance matrix $\mathbf{\Sigma}_A$ of the distribution of $\vec{\Delta k}$ can be computed as the correlated difference between the distribution of $\vec{k}_{\text{in}}$ with covariance matrix $\mathbf{\Sigma}_{\text{in}}$ and the distribution of $\vec{k}_{\text{out}}$ with covariance matrix $\mathbf{\Sigma}_{\text{out}}$ in that particular direction:

$$\sqrt{\mathbf{\Sigma}_A} = \sqrt{\mathbf{\Sigma}_{\text{in}}} - \sqrt{\mathbf{\Sigma}_{\text{out}}} \tag{124}$$

The distribution of $\vec{k}_{\text{out}}$ with the covariance matrix $\boldsymbol{\Sigma}_{\text{out}} = \sigma_\nu^2 \vec{w}_{\text{out}} \vec{w}_{\text{out}}^\top$ is not affected by divergence and only contains the wavelength distribution along $\vec{k}_{\text{out}}$, and where σ_ν is the bandwidth. The distribution of $\vec{k}_{\text{in}}$ is affected by both the wavelength distribution and the angular distribution of incident beams, possibly correlated. In the slightly less general case, where it is assumed that the angular distribution of the incident beam is isotropic and is not correlated to its wavelength, the distribution of $\vec{\Delta k}$ entirely due to polychromaticity is:

$$\boldsymbol{\Sigma}_A = \sigma_\nu^2 \Delta\vec{w}\Delta\vec{w}^\top \tag{125}$$

Combining this equation with equation 123 gives a way to estimate Σ_A under simplified conditions:

$$\boldsymbol{\Sigma}_A = \sigma_\nu^2 \Delta\vec{w}\Delta\vec{w}^\top + \sigma_{\text{in}}^2 \nu^2 \left(\boldsymbol{I} - \vec{w}_{\text{in}}\vec{w}_{\text{in}}^\top\right) \tag{126}$$

If we cannot assume that wavelength and incident angle are uncorrelated, $\boldsymbol{\Sigma}_{\text{in}}$ can be treated as a free parameter instead, and Σ_A can be derived by roatating the component of $\boldsymbol{\Sigma}_{\text{in}}$ that is due to polychromaticity and therefore in line with the incident beam direction to each $\vec{k}_{\text{out}}$. The distribution of $\vec{k}_{\text{out}}$ given $\boldsymbol{\Sigma}_{\text{in}}$ is therefore:

$$\boldsymbol{\Sigma}_{\text{out}} = \text{rotate}\left(\vec{w}_{in}, \vec{w}_{out}\right)\left[\left(\vec{w}_{\text{in}}^\top \boldsymbol{\Sigma}_{\text{in}} \vec{w}_{\text{in}}\right)\left(\vec{w}_{\text{in}}\vec{w}_{\text{in}}^\top\right)\right] \tag{127}$$

where rotate $\left(\vec{w}_{in}, \vec{w}_{out}\right)$ is the rotation matrix of the rotation around the axis orthogonal to $\vec{w}_{in}$ and $\vec{w}_{out}$, that would align $\vec{w}_{in}$ to $\vec{w}_{out}$. Then $\boldsymbol{\Sigma}_A$ is given by equation 124.

6.9.6. Parametrization of the Crystal

Due to its periodicity, the Fourier transform of a crystal is concentrated in peaks. As discussed above, these peaks are broadened by properties of the crystal, such as the finite width of the crystal, mosaicity and strain. Here we define the separate effects that are modelled:

Mosaicity is commonly used to describe a rotational disorder of the crystal and can be seen as a distribution of orientations of the unit cell. Rotational disorder of an object in three dimensions will have six degrees of freedom in general: Rotational disorder around three orthogonal axes and three covariance terms between them.

Strain is the distribution of contractions of unit cells. Generally, for each real-space lattice point in three dimensions there can be a different distribution of displacements in the direction of the origin and traverse to it. In the following it will be assumed that the changes of the structure factors due to strain are negligible.

Mosaicity and strain taken together, considering correlations of the effects in three dimensions, require a higher dimensional tensor that maps each point of reciprocal space to a cross-correlation matrix. In the following, however, mosaicity and strain will be taken as uncorrelated and mosaicity will be approximated to be isotropic. This means that mosaicity is assumed to be equal in all angular directions and mutually independent of crystal strain. The integration in the following subsection (section 6.9.7) will however be applicable with and without this simplification.

Reciprocal peak shape is the parameter that describes the distribution of each lattice point in reciprocal space, possibly due to the transform of the shape of a finite crystal, before being broadened by the effects of mosaicity and strain. In general this is a free parameter, but e.g. if the shape transform

is $\text{sinc}(\pi x)\text{sinc}(\pi y)\text{sinc}(\pi z)$ (the Fourier transform of a cube), it could be approximated by a Gaussian distribution with covariance $\mathbf{\Sigma}_P = \frac{1}{4}\mathbf{I}$. Because of the approximately quadratic decay in the observed diffraction, as opposed to the exponential decay of the Gaussian, shape transforms are not approximated efficiently by sums of Gaussian functions. Therefore, if the reciprocal peak shape is the predominant effect that is broadening the diffraction condition, the approximate nature of the proposed model becomes most obvious. The strength of the proposed method is the ability to combine different effects analytically, where the convolved distributions naturally become smoother.

Integer multiples of the reciprocal unit cell matrix $\mathbf{R}$ span the locations $\vec{\mu}_P$ of the peaks in the Fourier transform of the crystal: $\vec{\mu}_P = \mathbf{R}\begin{pmatrix} h & k & l \end{pmatrix}^\top$ (compare equation 117). The density around $\vec{\mu}_P$ is approximated to be a Gaussian distribution with the covariance matrix $\mathbf{\Sigma}_P$. The cumulative distribution results from the convolution of the individual distributions. Its covariance is therefore the sum of the covariance matrix $\mathbf{\Sigma}_{P_0}$ describing the shape transform, the effect of isotropic mosaicity $\sigma_m^2\left(|\vec{\mu}_P|^2 I - \vec{\mu}_P\vec{\mu}_P^\top\right)$, and the effect of uncorrelated strain $\sigma_s^2\vec{\mu}_P\vec{\mu}_P^\top$. Here σ_m quantifies the mosaicity as the standard deviation of rotational disorder, and σ_s quantifies the strain as the standard deviation of the relative unit cell size variation.

6.9.7. Evaluation of Integrals

Given the distributions defined in Sec. 6.9.5 and 6.9.6, we are now in a position to compute the diffracted flux density in a given direction $\vec{w}_{\text{out}}$. This is done by evaluating particular integrals for each pair of Gaussian basis functions of the distributions, as given below. Polarization and scaling terms were left out at this point for clarity, because they are not affected by the integration. If at least one of the distributions is assumed to have random or chaotic phases, the integration is incoherent. Thus, by using equation 122 and the definition of ϕ in equation 62 we get the following result:

$$\int_{\mathbb{R}^3} \phi\left(\vec{x}, \vec{\mu}_A, \mathbf{\Sigma}_A\right)^2 \phi\left(x; \vec{\mu}_P, \mathbf{\Sigma}_P\right)^2 d\vec{x} \tag{128}$$

$$= \int_{\mathbb{R}^3} \phi\left(\vec{x}, \vec{\mu}_A, \frac{1}{2}\mathbf{\Sigma}_A\right)|4\pi\mathbf{\Sigma}_A|^{-\frac{1}{2}} \phi\left(\vec{x}, \vec{\mu}_P, \frac{1}{2}\mathbf{\Sigma}_P\right)^2 |4\pi\mathbf{\Sigma}_P|^{-\frac{1}{2}} d\vec{x}$$

$$= \phi\left(\vec{\mu}_A, \vec{\mu}_P, \frac{1}{2}\mathbf{\Sigma}_A + \frac{1}{2}\mathbf{\Sigma}_P\right)|4\pi\mathbf{\Sigma}_A|^{-\frac{1}{2}}|4\pi\mathbf{\Sigma}_P|^{-\frac{1}{2}} \tag{129}$$

$$= \exp\left(-\left(\vec{\mu}_A - \vec{\mu}_P\right)^\top \mathbf{\Sigma}_o^{-1}\left(\vec{\mu}_A - \vec{\mu}_P\right)\right)\left|32\pi^3\mathbf{\Sigma}_*^{-1}\right|^{-\frac{1}{2}}$$

If all contributions to the diffraction described by the two distributions have a constant phase relation, the integration is coherent:

$$\left(\int_{\mathbb{R}^3} \phi\left(\vec{x}, \vec{\mu}_A, \mathbf{\Sigma}_A\right) \phi\left(x; \vec{\mu}_P, \mathbf{\Sigma}_P\right) d\vec{x} \right)^2 \tag{130}$$

$$= \phi\left(\vec{\mu}_A, \vec{\mu}_P, \mathbf{\Sigma}_o\right)^2$$

$$= \phi\left(\vec{\mu}_A, \vec{\mu}_P, \frac{1}{2}\mathbf{\Sigma}_o\right) \left|4\pi\mathbf{\Sigma}_o\right|^{\frac{-1}{2}} \tag{131}$$

$$= \exp\left(-\left(\vec{\mu_A} - \vec{\mu_P}\right)^\top \mathbf{\Sigma}_o^{-1} \left(\vec{\mu_A} - \vec{\mu_P}\right)\right) \left|2\pi\mathbf{\Sigma}_o\right|^{-1}$$

where

$$\mathbf{\Sigma}_0 = \mathbf{\Sigma}_A + \mathbf{\Sigma}_P$$

$$\mathbf{\Sigma}_* = \left(\mathbf{\Sigma}_A^{-1} + \mathbf{\Sigma}_P^{-1}\right)^{-1}$$

The result of equation 129 is applied below in section 6.9.8 to compute a diffraction pattern that matches the observed pattern. This requires the appropriate scaling and polarization correction. All in all there are 17 parameters describing each Gaussian function of the crystal (9 for the unit cell, 6 for the shape transform and one each for mosaicity and strain) and 9 describing each Gaussian function in the source (3 parameters for the direction and 6 for a possibly correlated distribution of illumination angles and wavelengths). The source will typically not change for many crystals in a serial crystallography experiment and one Gaussian function will give enough degrees of freedom to describe the diffraction of each crystal.

6.9.8. Pixel-wise diffraction pattern prediction

The first way this model can be used to process data, is to model each pixel of a diffraction pattern, making use of as many constraints as possible in determining the hidden parameters and the structure factor amplitudes. A still diffraction pattern can be calculated by using the result of equation 129 for each point on the detector, by applying a polarization correction C and scaling with the intensity of the incoming beam and with the respective structure factor modulus square $|F|^2$ of each reflection.

$$j = J_0 \left| F \begin{pmatrix} h \\ k \\ l \end{pmatrix} \right|^2 C\left(p, \vec{n}, \vec{w}_{\text{in}}, \vec{w}_{\text{out}}\right) \frac{\exp\left(-\left(\vec{\Delta k} - \mathbf{R}\begin{pmatrix} h \\ k \\ l \end{pmatrix}\right)^\top \mathbf{\Sigma}_o^{-1} \left(\vec{\Delta k} - \mathbf{R}\begin{pmatrix} h \\ k \\ l \end{pmatrix}\right)\right)}{\left|32\pi^3\mathbf{\Sigma}_*^{-1}\right|^{\frac{1}{2}}} \tag{132}$$

$$C\left(p, \vec{n}, \vec{w}_{\text{in}}, \vec{w}_{\text{out}}\right) = p\left(1 - \left(\vec{w}_{\text{out}}^\top\left(\vec{w}_{\text{in}} \times \vec{n}\right)\right)^2\right) + (1-p)\left(1 - \left(\vec{w}_{\text{out}}^\top\vec{n}\right)^2\right) \tag{133}$$

where J_0 is the incident beam flux, p is the degree of polarization, $\vec{n}$ the normal to the polarization plane, and F the structure factor. The flux measured in a pixel is the integral over all directions that fall into the solid angle of that pixel summed up for all Miller indices with significant excitation. If the predicted flux was constant over this area, the integral would be just proportional to the solid angle that the pixel occupies.

The detector is assumed to be composed of rigid panels. Each panel has its own two-dimensional coordin-

ate system consisting of the dimensions fs and ss defined in terms of the memory order, where fs (short for fast scan) is the dimension of values stored consecutively and ss (short for slow scan) is the dimension that is not. Each panel has a local coordinate system given by a 3×2 matrix $\boldsymbol{D}$ for the two dimensions in the plane of the panel and an offset vector $\vec{o}$ for the absolute position in space of the corner corresponding to the origin of the coordinate system of this panel. The solid angle of a pixel can be approximated with the derivative of the normed directionality vector $\vec{w}_{\text{out}}$ with respect to the detector coordinates:

$$\vec{w}_{\text{out}} = \left(\boldsymbol{D} \begin{pmatrix} fs \\ ss \end{pmatrix} + \vec{o} \right) \left| \boldsymbol{D} \begin{pmatrix} fs \\ ss \end{pmatrix} + \vec{o} \right|^{-1} \tag{134}$$

$\vec{w}_{\text{out}} =$ direction in which diffraction is to be predicted

$\boldsymbol{D} =$ matrix translating between panel coordinates and spatial coordinates

$\vec{o} =$ spatial coordinates of the reciprocal space origin in detector coordinates

$\begin{pmatrix} fs \\ ss \end{pmatrix} =$ coordinates of the pixel on the detector

For the following two derivations it will be useful to know the derivative of the direction $\vec{w}_{\text{out}}$ with respect to its two coordinates in the detector panel's coordinate system:

$$\frac{\partial \left(\vec{w}_{\text{out}} \right)}{\partial \begin{pmatrix} fs \\ ss \end{pmatrix}} = \left(\boldsymbol{D} - \vec{w}_{\text{out}} \vec{w}_{\text{out}}^{\top} \boldsymbol{D} \right) \left| \boldsymbol{D} \begin{pmatrix} fs \\ ss \end{pmatrix} + \vec{o} \right|^{-1} \tag{135}$$

The solid angle Ω is approximated by the length of the cross product of the pixel sides projected onto the unit sphere:

$$\Omega \approx \left| \left(\frac{\partial \left(\vec{w}_{\text{out}} \right)}{\partial fs} (\text{fs}) \right) \times \left(\frac{\partial \left(\vec{w}_{\text{out}} \right)}{\partial ss} (\text{ss}) \right) \right| \tag{136}$$

However, the predicted peaks can be very narrow, and therefore the predicted flux can vary substantially within a single pixel. To enable an efficient integration over the area, the predicted flux density can be smoothened analytically without changing the total flux of the whole diffraction pattern. This is achieved by introducing a Gaussian point spread function for the detector (the blue arrows in figure 18) with a covariance matrix corresponding to $1/2$ the extent of a pixel, or for greater accuracy, by oversampling the pixel and applying the same procedure to the subpixels. Simply put, this smoothens the prediction to a level where discretely sampling it only introduces minor artefacts, the main effect being a slightly reduced contrast. The constant $1/2$, of the aforementioned pixel extent, minimizes the maximum Kullback Leibler Divergence D_{KL} [15] between the desired proper integral (b) involving the error function and the estimate (c). Equation 137 shows a proof in one dimension, that can be generalized to higher dimensions for all shapes for which a orthogonalizing coordinate transform can be found. It is natural to assume that the same constant approximately minimizes this difference even when the sides are not strictly parallel. The D_{KL} is an asymmetric measure for the difference of probability distributions, taking into account that

under-estimating a probability is more detrimental than over-estimating it. It was chosen because the predicted flux density is a scaled probability density.

$$D_{KL}\left(P \parallel Q\right) = \int_{\mathbb{R}} P(x) \log\left(\frac{P(x)}{Q(x)}\right) dx$$

$$b(x, \mu, \sigma) = \frac{1}{2}\left(\operatorname{erf}\left(\frac{x - \mu + \frac{1}{2}}{\sqrt{2\sigma^2}}\right) - \operatorname{erf}\left(\frac{x - \mu - \frac{1}{2}}{\sqrt{2\sigma^2}}\right)\right)$$

$$c(x, \mu, \sigma) = \frac{\exp\left(-\frac{1}{2}(x - \mu)^2 \left(\sigma^2 + \sigma_+^2\right)^{-1}\right)}{\sqrt{2\pi\left(\sigma^2 + \sigma_+^2\right)}}$$

$$\underset{(x-\mu),\sigma}{\arg\max}\left(b(x, \mu, \sigma) \log\left(\frac{b(x, \mu, \sigma)}{c(x, \mu, \sigma)}\right)\right) = \left(\frac{1}{2}, 0\right)$$

$$\lim_{\sigma \to 0^+, x - \mu \to \frac{1}{2}^-}\left(b(x, \mu, \sigma)\right) = 1$$

$$\underset{\sigma_+}{\arg\min}\left(-\log\left(q\left(x, x - \frac{1}{2}, 0\right)\right)\right) = \underset{\sigma_+}{\arg\min}\left(\frac{1}{2^2\sigma_+^2} + \log\left(2\pi\sigma_+^2\right)\right) = \frac{1}{2} \tag{137}$$

$$P = \text{precise probability distribution}$$
$$Q = \text{approximation}$$
$$\mu = \text{mean value}$$
$$\sigma = \text{standard deviation from the mean}$$
$$\sigma_+ = \text{the constant to be solved for}$$

Using the results of equations 137 and 135, the resulting covariance matrix of the smoothing function is:

$$\boldsymbol{\Sigma}_D = \frac{\nu^2}{2^2}\left(\frac{\partial\,\vec{w}_{\text{out}}}{\partial\begin{pmatrix} fs \\ ss \end{pmatrix}}\begin{pmatrix} fs \\ ss \end{pmatrix}\right)\left(\frac{\partial\,\vec{w}_{\text{out}}}{\partial\begin{pmatrix} fs \\ ss \end{pmatrix}}\begin{pmatrix} fs \\ ss \end{pmatrix}\right)^{\top} \tag{138}$$

We now have a way of modelling the flux of each pixel. This is good enough for monochromatic experiments, but to model polychromatic experiments we need to take into account that detector response signals of integrating detectors are proportional to the total photon energy impinging on the detector. Integrating detectors are commonly chosen over counting detectors for SX experiments as they are not limited to measuring one photon per pixel at a time. The following derivation uses the wavenumber k, which is proportional to the impinging photon energy.

The average wavenumber of the polychromatic diffracted beam at the particular location of a given pixel can be estimated from the mean point of the distribution resulting from the multiplication of the distribution of the source and the distribution of the peak of the crystal in reciprocal space (compare equation 63) by rescaling the component collinear to the incident beam. We are only interested in the collinear component because the deviation of $\vec{\Delta k}$ in any other direction is not due to the wavelength distribution but due to

other factors like convergence. The rescaling is necessary, because the correlated difference between $\vec{k}_{\text{in}}$ and $\vec{k}_{\text{out}}$, which necessarily have equal wavelengths, leads to a covariance matrix of $\vec{\Delta k}$ that appears sheared with respect to the covariance of $\vec{k}_{\text{in}}$ and compressed along the beam direction. A geometric visualization is offered with figure 22 in lieu of a mathematical proof. The cosine of the angle of diffraction equals the scalar product between the normalized incoming and outgoing wave vectors, leading to the following expression:

$$k = \vec{w}_{\text{in}}^{\top} \left(\mathbf{\Sigma}_A^{-1} + \mathbf{\Sigma}_P^{-1} \right)^{-1} \left(\mathbf{\Sigma}_A^{-1} \vec{\mu}_A + \mathbf{\Sigma}_P^{-1} \vec{\mu}_P \right) \left(1 - \vec{w}_{\text{in}}^{\top} \vec{w}_{\text{out}} \right)^{-1} \tag{139}$$
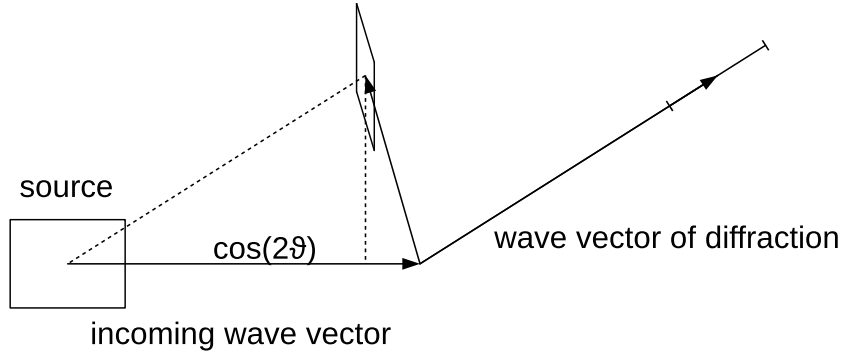
Figure 22: Geometric explanation for formula 139 for the expected wavenumber. Convergence, orthogonal to $\vec{w}_{\text{in}}$, and wavelength dispersion, in line with $\vec{w}_{\text{in}}$, are indicated as a box to highlight the shearing of the covariance when forming the correlated difference between $\vec{w}_{\text{in}}$ and $\vec{w}_{\text{out}}$ and their respective variances. It can be seen that the length of $\vec{\Delta w}$ projected onto $\vec{w}_{\text{in}}$ is $1 - \cos(2\theta)$, where 2θ is the angle of diffraction.

This way, the expected flux and distribution of wavelengths for each pixel in a diffraction pattern can be predicted. As a side-note, this allows us to describe even diffraction patterns where the scattering factors depend on the wavelength, like in MAD phasing, when the beam spectrum contains frequencies close to resonances of atoms within the molecule. A simulation in figure 23 shows that this is possible in principle, but with the introduction of yet another dimension of free parameters for the wavelength dependent structure factors this is an advanced application. For now we will assume that structure factors are not depending on the wavelength.
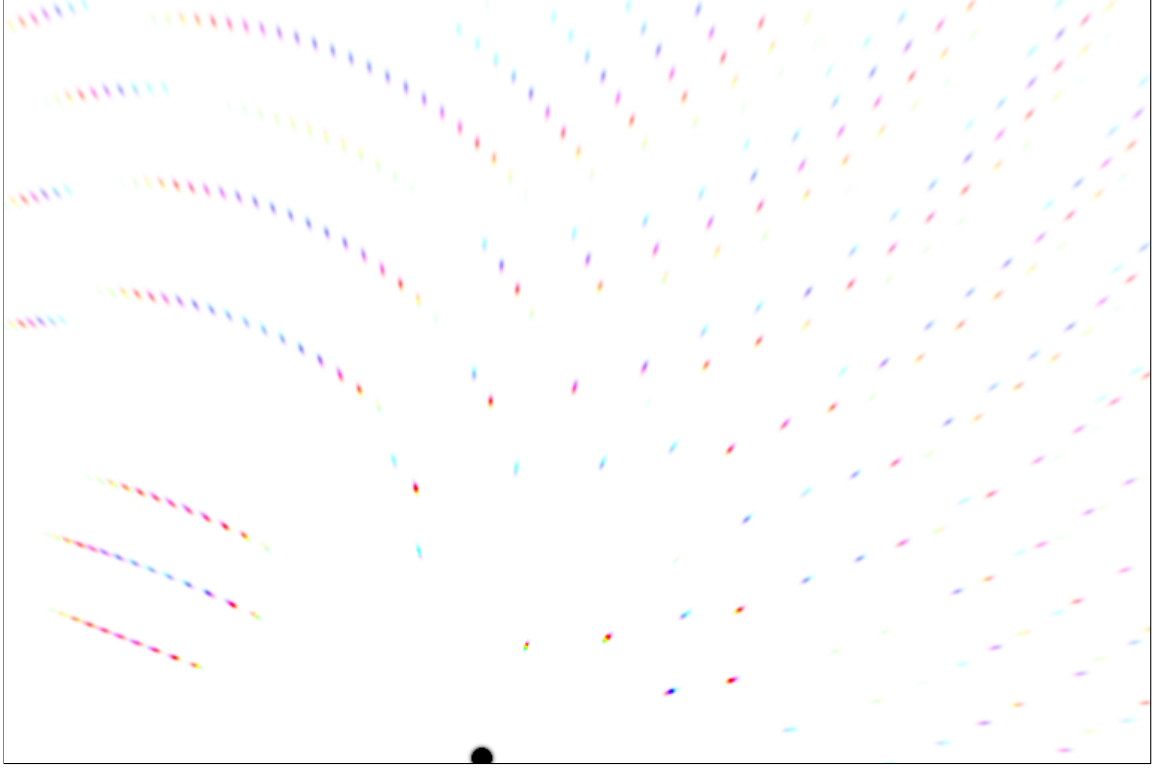
Figure 23: Simulation of a pink-beam diffraction pattern of biotin bound to streptavidin without polarization, shown with false colors. The wavelength spectrum is modelled after the APS pink-beam spectrum. Intensities are mapped to color saturation and the predicted wavelength is mapped to hue. As can be seen, the crystal acts like a monochromator and separates the wavelengths radially. A big black dot was added to show the direct beam position.

Having a distribution of photons of different wavelengths does not change the Poisson photon counting statistic, but it leads to an additional variance in the measured intensity proportional to the width of this distribution, because each photon measured can have a different energy. The width of the wavenumber distribution in each pixel can be estimated from the shape of the product of the two Gaussians in equation 128 by projecting to the incoming beam and rescaling.

$$\sigma_k = \left(1 - \vec{w}_{\text{in}}^\top \vec{w}_{\text{out}}\right)^{-1} \sqrt{\frac{1}{2} \vec{w}_{\text{in}}^\top \boldsymbol{\Sigma}_* \vec{w}_{\text{in}}} \tag{140}$$

This is analogous to the expected wavenumber in equation 139. From the expected photon count, the expected wavelength and the constant g describing the detector response as detector counts per wavenumber, the expected detector reading $\hat{y}$ for a given pixel is given as the product:

$$\hat{y} = jkg \tag{141}$$

To model random and systematic errors, the following two-parameter (α and β) error model is employed

to predict the total variance:

$$\sigma_{\hat{y}}^2 = \sigma_0^2 + g^2 \left(\alpha + \beta \left| j \right| \right) \left| j \right| \left(k^2 + \sigma_k^2 \right) \tag{142}$$

This error model is essentially equivalent to [90, equation 3], but by substituting the measured intensity with the expected intensity, the bias in merging introduced by error estimation based on observed intensities as shown in figure 11 is removed, if the errors before correction σ_0^2 do not contain the photon counting noise of the reflection intensity itself. The factor α models errors proportional to the square root of the expected signal, like those inherent to the photon counting statistic. The factor β models errors proportional to the expected signal, systematic errors like due to a small misalignment of the indexing solution, leading to relative errors. Constant errors are not modelled here, they are assumed to be fully captured by σ_0^2.

To connect the prediction $\hat{y}$ with the measured data y we introduce a probability distribution described by the density function f(y), which enables a maximum likelihood optimization. The probability distribution is a mixture distribution of a smoothened Gaussian that approximates a discrete Gaussian with the additional variance $\frac{1}{2^2}$ using the result in equation 137 and a super-heavy tailed outlier distribution $u(y)$ that models even extreme outliers like defective pixels with an assumed outlier fraction ϵ. Without any further assumptions about the unexpectedly observed values it is natural that they should follow the prior for natural numbers, see section 3.4.1.

$$f(y) = (1 - \epsilon)\, \phi \left(y, \hat{y}, \sigma_{\hat{y}}^2 \right) + \epsilon u(y) \tag{143}$$

Crystal diffraction is sparse and most pixels will not see significant diffraction. The pixels with significant diffraction can be estimated conservatively by finding the potentially excited indices, using a region growing algorithm (see B.1.2) and then projecting the peak shape onto the detector (using equation 158). This accelerates the prediction greatly while not affecting the result in any significant way. Because derivatives can be computed analytically, the predicted diffraction pattern can be optimized by using pseudo-Newton optimization methods like BFGS [91, 92, 93, 94] or gradient descent. In theory this should make the optimization straightforward and efficient, but the target function has many local minima and plateaus. Together with the associated computational cost, this is the reason why pixel-wise refinement of the Gaussian sum model proposed in this paper so far has only been applied to individual diffraction patterns and not full datasets. This method also depends on a pixel-wise background estimate and a detector geometry that is determined well enough, such that predicted pixels coincide mostly with measured pixels. It demands the computation of about 8kpx for a 4Mpx detector. Together this makes it computationally expensive, requiring on the order of 10 single core computing hours per pattern (4GHz AMD A12). Therefore, this method has not yet been connected with structure refinement directly, but has been used to show visually that different diffraction patterns can be predicted accurately.

75

<div align="center">(a)   (b)</div>
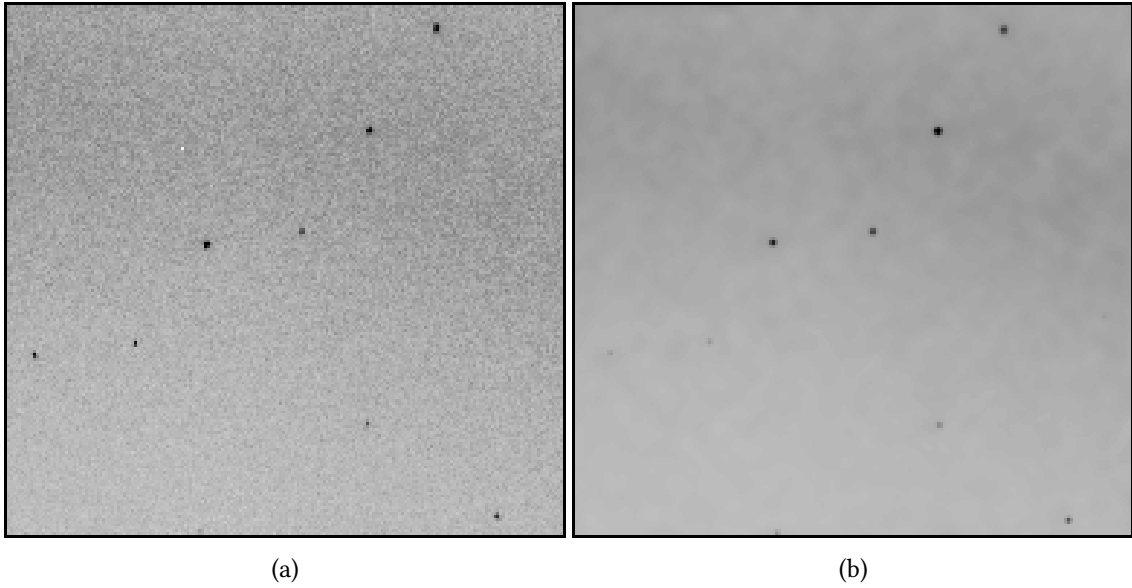
Figure 24: Comparison between (a) previously published diffraction data from a human serotonin receptor [95] (on the left) and (b) predicted diffraction of the same image region after successful optimization, with estimated background added. Diffraction is predicted using equation 116 with the substitution $\Sigma_o \rightarrow \Sigma_o + \Sigma_D$, corrected for the solid angle with equation 136, equation 139 to estimate the expected wavelength and summed up over all significantly excited Miller indices. Intensities are scaled according to the reference intensities deposited in the PDB under 4NC3. The bandwidth of the X-ray beam is estimated to be about 0.1% (LCLS states 0.2% $\Delta E/E$ FWHM for the CXI beamline [96].)

parameter	degrees of freedom	optimization
geometry description	9 for each panel	yes
unit cell	9	yes
reciprocal peak shape	6	yes
mosaicity	1	yes
strain	1	yes
linear scale-factor	1	yes
B-factor	1	yes
error model	2	yes
source description	10	no

Table 6: Parameters that were optimized against pixel values
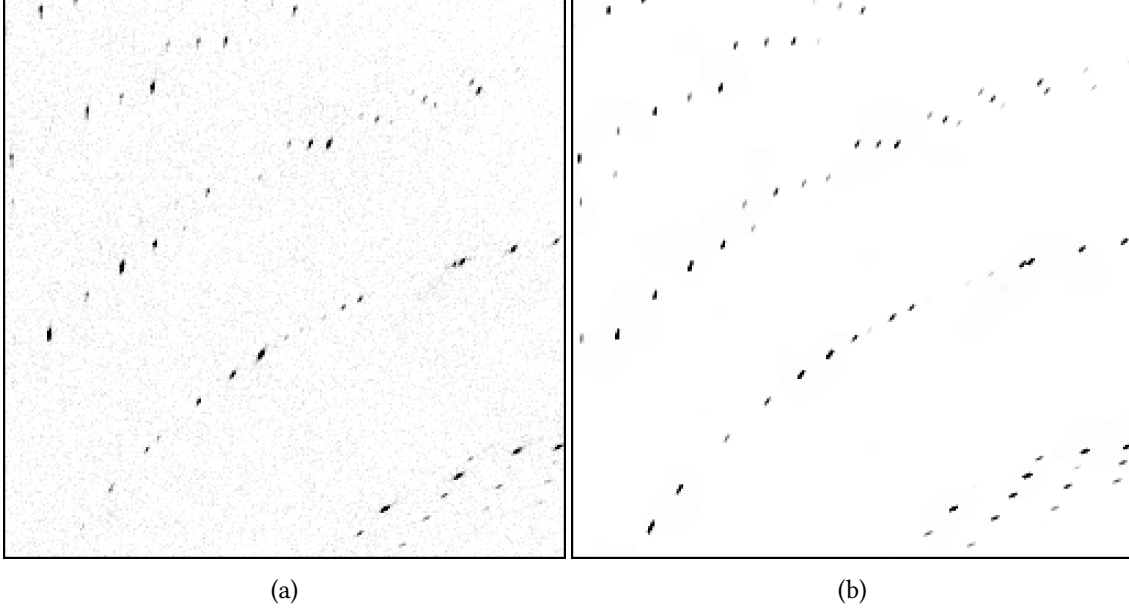
Figure 25: Comparison between (a) diffraction data (unpublished) of selenobiotine bound streptavidin crystals and (b) predicted diffraction of the same image region with estimated background added. Diffraction is predicted using equation 116 with the substitution $\Sigma_o \to \Sigma_o + \Sigma_D$, corrected for the solid angle with equation 136, equation 139 to estimate the expected wavelength and summed up over all significantly excited Miller indices. The diffraction was measured at ESRF with a 1M Jungfrau detector using a pink beam with 5% bandwidth FWHM. The structure factors for the prediction are taken from the Streptavidin-norbiotin complex structure deposited under 1LCV in the PDB [97].

6.9.9. Merging Using Integrated Peak Intensities

This section describes the second application of the model presented in section 6.9.3: merge Gaussian partiality corrected integrated intensities (MGPCII). First we derive an expression for the total intensity of a reflection in a still diffraction pattern and then we describe a method of how to use this expression to reduce the detrimental impact of partially recorded reflections on the estimates of structure factors.

An expression for integrated peak intensities

The total photon energy of one reflection can be computed by integrating the result of equation 129 over all directions. This integral can be approximated when considering that the angular extent of a reflection on the detector is small and the curvature as well as the change in width of the Ewald sphere is negligible for the integral over a single reflection. The density of the Ewald sphere can therefore be approximated as a planar Gaussian, decaying along the direction of diffraction, but constant orthogonal to it. First the double integral is restated by using equation 128:

$$\int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \left(\phi\left(\vec{x}, \vec{k}_{\text{in}} - k\vec{w}_{\text{out}}, \Sigma_A\right) \phi\left(\vec{x}, \vec{\mu}_P, \Sigma_P\right) \right)^2 d\vec{x}\, d\vec{w}_{\text{out}} \tag{144}$$

$$= \int_{\mathbb{R}^3} \phi\left(\vec{k}_{\text{in}} - k\vec{w}_{\text{out}}, \vec{\mu}_P, \frac{1}{2}\Sigma_A + \frac{1}{2}\Sigma_P\right) |4\pi\Sigma_A|^{-\frac{1}{2}} |4\pi\Sigma_P|^{-\frac{1}{2}} d\vec{w}_{\text{out}} \tag{145}$$

The integral along all possible outgoing wave directions is then approximated with a projection onto the outgoing wave direction with the highest intensity $\vec{w}_{\max}$, which can be found by function optimization:

$$\approx \left(\phi \left(\vec{k}_{\text{in}} - k\vec{w}_{\max}, \vec{\mu}_P, \vec{w}_{\max}^\top \mathbf{\Sigma}_\circ \vec{w}_{\max} \right) \right)^2 |4\pi \mathbf{\Sigma}_*|^{-\frac{1}{2}} \tag{146}$$

$$\mathbf{\Sigma}_A^{-1} = d^{-2} \vec{w}_{\text{out}} \vec{w}_{\text{out}}^\top$$
$$d = \text{width of the Ewald sphere at the projection point}$$

The photon flux of each reflection in each pattern is estimated as the product of the result of equation 146 with the incident photon flux J_0, the structure factor amplitude squared, a linear scaling factor a, a B-factor correction term modelling a Gaussian decay of intensities due to random atomic displacements, and a term for the polarization correction (equation 133). This leads to an expression analogous to equation 132, but with an explicit linear and B-factor scaling instead of implicitly assigning those as terms in the structure factors:

$$j = J_0 \left| F \begin{pmatrix} h \\ k \\ l \end{pmatrix} \right|^2 a \exp\left(-B \left| \mathbf{R} \begin{pmatrix} h \\ k \\ l \end{pmatrix} \right|^2 \right) C \frac{\exp\left(-\left(\vec{\Delta k} - \mathbf{R} \begin{pmatrix} h \\ k \\ l \end{pmatrix} \right)^\top \mathbf{\Sigma}_o^{-1} \left(\vec{\Delta k} - \mathbf{R} \begin{pmatrix} h \\ k \\ l \end{pmatrix} \right) \right)}{\left| 32\pi^3 \mathbf{\Sigma}_*^{-1} \right|^{\frac{1}{2}}} \tag{147}$$

The calculation of the mean wavenumber is analogous to equation 139:

$$k = \frac{\vec{w}_{\text{in}}^\top \left(\left(\vec{w}_{\text{out}}^\top \mathbf{\Sigma}_A \vec{w}_{\text{out}} \right)^{-1} \vec{\mu}_A + \left(\vec{w}_{\text{out}}^\top \mathbf{\Sigma}_P \vec{w}_{\text{out}} \right)^{-1} \vec{\mu}_P \right)}{\left(1 - \vec{w}_{\text{in}}^\top \vec{w}_{\text{out}} \right) \left(\left(\vec{w}_{\text{out}}^\top \mathbf{\Sigma}_A \vec{w}_{\text{out}} \right)^{-1} + \left(\vec{w}_{\text{out}}^\top \mathbf{\Sigma}_P \vec{w}_{\text{out}} \right)^{-1} \right)} \tag{148}$$

The width of the predicted wavenumber distribution is analogous to equation 140:

$$\sigma_k = \left(1 - \vec{w}_{\text{in}}^\top \vec{w}_{\text{out}} \right)^{-1} \sqrt{\frac{1}{2} \left(\left(\vec{w}_{\text{out}}^\top \mathbf{\Sigma}_A \vec{w}_{\text{out}} \right)^{-1} + \left(\vec{w}_{\text{out}}^\top \mathbf{\Sigma}_P \vec{w}_{\text{out}} \right)^{-1} \right)^{-1}} \tag{149}$$

The expected detector count for each reflection is the product of wavenumber, flux and a detector constant, as in equation 141. Its variance is estimated with the same two-parameter error model as for the pixel-wise prediction in equation 142.

Parameter Optimization for Merging

The purpose of merging is to produce accurate estimates of the scattering intensities, proportional to the modulus squares of the structure factors, from a set of observed integrated peak intensities, see section 6.6. To make use of equation 146 for that end, its free parameters need to be determined. The scattering intensities are among the parameters to be determined, the other parameters are listed in table 7. To find the optimal parameters a maximum-likelihood approach was chosen because it can be made more robust than least squares, and it is still relatively easy to optimize for, also compare figure 14.

parameter	degrees of freedom	optimization
geometry description	9 for each panel	no
unit cell	9	yes
reciprocal peak shape	6	yes
mosaicity	1	yes
strain	1	yes
linear scale-factor	1	yes
B-factor	1	yes
error model	2	yes
source description	10	no

Table 7: Parameters that were optimized against integrated intensities

$f(y)$ is the probability distribution to be optimized for each observation. Probabilities are assumed to follow a mixture distribution of a Gaussian distribution and an outlier distribution $o(y)$. The outlier distribution should be chosen as to best describe all measured intensities in general, without prediction or scaling. In many cases a Cauchy distribution is a good choice because it fits the shape of the distribution of integrated intensities well for frequently observed values and has an inverse quadratic decay like the positive intensities. The exact shape of the outlier distribution is less relevant; its most important feature is a heavy tail to make the maximum-likelihood approach robust.

$$f(y) = (1 - \epsilon)\, \phi\left(y, \hat{y}, \sigma_{\hat{y}}^2\right) + \epsilon\, o(y) \tag{150}$$

$$o(y) = \frac{1}{\pi\gamma\left(1 + \left(\dfrac{y - y_0}{\gamma}\right)^2\right)} \tag{151}$$

$$o = \text{outlier distribution}$$
$$\epsilon = \text{outlier probability} = \frac{1}{16}$$
$$\gamma = \text{(the scale parameter of the Cauchy distribution)}$$
$$y_0 = 0$$

This approach treats the scattering intensities like any of the other parameters of the model, which is convenient, but it also means that it cannot directly be compared with the other merging algorithms in isolation, because they only determine structure factor amplitudes and scaling is detached. The most comparable method would be the maximum-likelihood that was presented in section 6.6. The critical difference between the two maximum-likelihood approaches is, that here we have a probability distribution for the integrated intensities as observed, whereas in section 6.6 it is a probability for the ratio of observed intensity and estimated partiality times scaling factors. Also, ratios are slightly less well-behaved, and are mathematically not as convenient, as seen in section 3.1.1.

6.9.10. Tests on Experimental Data

To show that equation 146 can be used to correct partially recorded reflections to improve the data quality, two serial femtosecond crystallography datasets were chosen: One calibration dataset of granulin microcrystals[4] (dataset 1) and one dataset [98] that allows SAD phasing (dataset 2). The diffraction patterns of both datasets were indexed and integrated with `indexamajig` of CrystFEL 0.9.1 [99]. To get a baseline for comparison to our method the integrated intensities were merged with `partialator` 0.9.1 and `partialator` 0.8.0, using the partiality models `ggpm`, `xsphere` and `unity` and those merged intensities were chosen that produced the best structure refinement results. The datasets were processed once with and once without over-prediction, that is also integrating peaks further away from the diffraction condition, via the command-line option `--overpredict`. The effect of overprediction is shown exemplary for the first dataset in figure 26, as can be seen, the additional reflections are mostly of low intensity. Overprediction was not helpful when merging with `partialator` in any of the combinations of options that were tested. Therefore, overprediction was not enabled for data points that were used as a comparison to the new method. However, it consistently lead to better structure refinement results when correcting partialities using the generalized Gaussian diffraction model and maximum likelihood parameter optimization during merging. This is why overprediction was enabled for that method.

Figure 26: Histogram of measured intensities of dataset 1 in black (without overprediction) and red (with overprediction) overlaid with the Cauchy outlier distribution ($\gamma = 1967.7$) in blue. The outlier distribution was chosen as to describe the measurements well, but also to reserve some probability especially for the extreme values. Note that the additional intensities due to overprediction are mostly small.
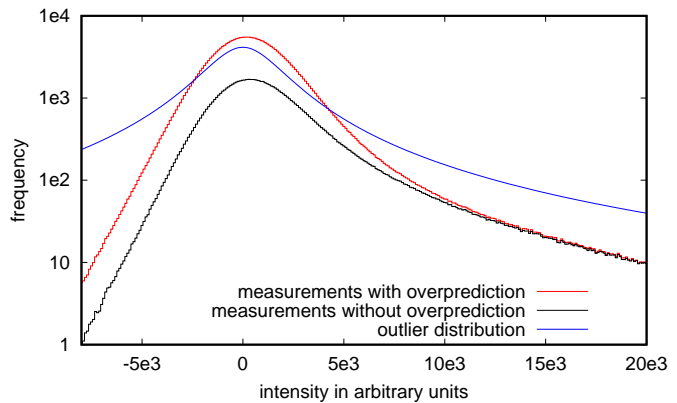
The method described in Sec. 6.9.9 was applied to both datasets and the quality of the intensities was compared to the `partialator` baseline. In addition, dataset 1 was investigated in more detail, in regards to overfitting, to the correlation of prediction and measurement and the distribution of estimated partialities, while the second dataset was used to test how much SAD phasing could be improved.

After optimization of the scaling parameters (in table 7) for dataset 1, the correlation between prediction and measurement is high, see figure 27, but the relative error between prediction and measurement still is about 25% and much larger than the photon counting error. The comparison of predicted and measured partialities in figure 28 shows a strong correlation, which is exploited when correcting the measurements using the partiality estimate. Unknown partialities increase the variance of the intensities before merging and therefore of the merged intensities too. The variance can be reduced by partiality correction.

---

[4]This dataset is previously unpublished and was measured in October 2020 at the SPB beamline of the European XFEL in preparation for bacterial insecticide crystals, by a team lead by Dominik Oberthür and Colin Berry. It has been deposited in the CXIDB with the ID 203.
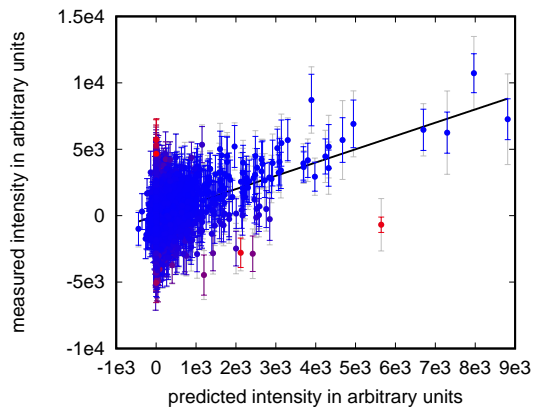
(a) The first 1000 intensities as recorded in the granulin dataset (dataset 1).
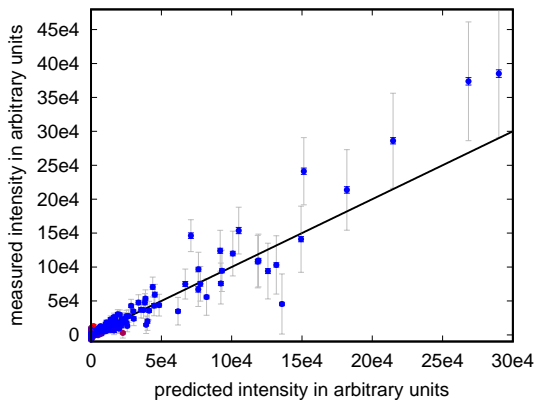
(b) The intensities and predictions for the crystal with the strongest diffraction in the same dataset.

Figure 27: Predicted intensities versus measured intensities with the photon counting error estimates indicated by blue error bars and corrected error estimates by gray error bars. In red are datapoints that were treated as outliers, dots in blue were treated as regular datapoints. The black line shows where the points would lie, if the predictions were in perfect agreement with the measurements.

Figure 28: A scatter plot of the first 10000 predicted versus measured partialities with a measurement error of less than 0.125 in the granulin dataset (dataset 1).
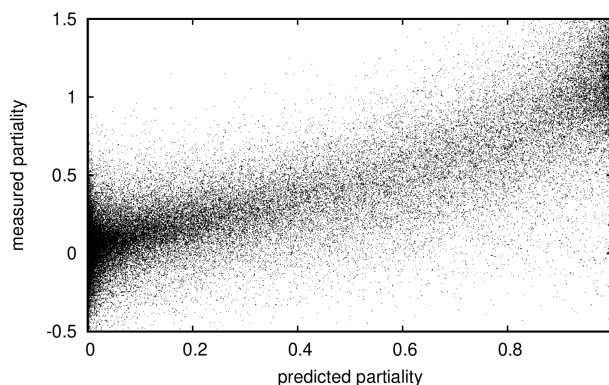
Figure 29: Histogram of partialities measured to a precision better than 1/8 from the granulin dataset (dataset 1). Partiality is the ratio of measured intensity to predicted intensity without partiality estimate, that is only correcting for merged intensity, linear- and B-factor scaling.

81
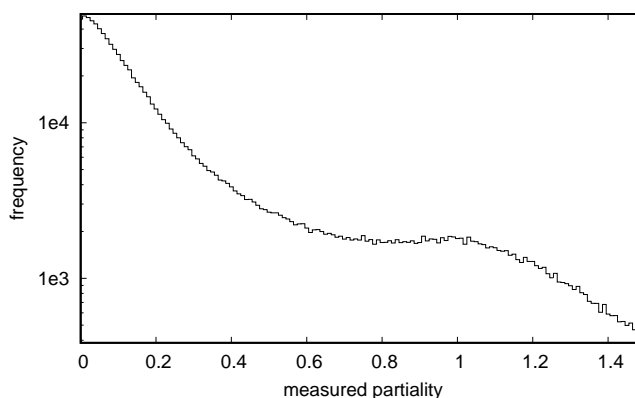
Figure 30: Predicted partialities compared
to measured partialities, with
photon counting error estim-
ates indicated by error bars.
The first 993 values from data-
set 1 in the order they are
recorderd to have an estim-
ated photon counting and back-
ground subtraction error of less
than $1/4$ are displayed. The
black line shows where the
points would lie, if the predic-
tions were in perfect agreement
with the measurements.
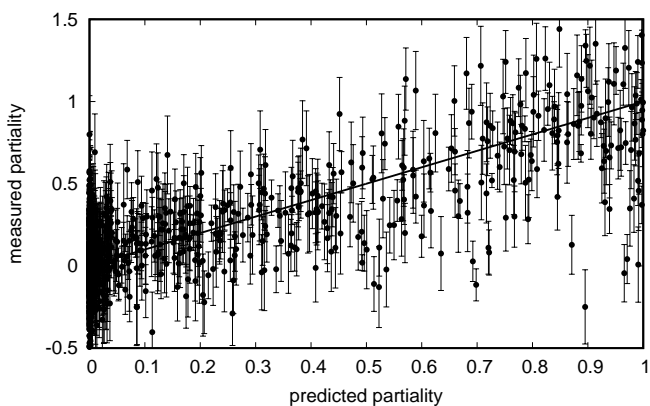
As it would be expected for the smoothened distribution of the function values of a Gaussian function
with uniform input[5], the histogram of the measured partialities (figure 29) has an optimum at 0, corres-
ponding to a reflection that was not observable - most reflections in a given crystal orientation are not
observable, and also there is a very faint optimum at 1. The optimum at 1 corresponds to the flat top of the
intensity curve of an observation near its maximum intensity.

To test the amount of overfitting, dataset 1 was split randomly in two halves. The first half was used
to optimize the parameters of the scaling and partiality model in table 7 and the second half was used to
test the correspondence of prediction and measurement. The median correlation of 256 random prediction-
measurement pairs [6] decreased from 0.59 to 0.56, the reduction in correlation can be observed by comparing
figure 31a to figure 31b. This is evidence of some degree of overfitting, but also means that even half the
number of peaks is sufficient to arrive at roughly the same prediction. Thus, even though this method would
likely profit from additional constraints[7], it still reduced the number of diffraction patterns necessary to
achieve a given data quality by about a factor of 2. R-factors after automatic refinement, see figure 32, were
consistently lower for MGPCII than for `partialator`.

Dataset 2 is of of the Adenosine receptor A_{2A}, measured at LCLS using a wavelength of 2.7 Å [98]. The
protein contains 22 sulphur atoms and the wavelength is close enough to the absorption edge to make SAD
phasing possible. This makes this dataset suitable to see to what extent partiality correction would improve
phasing success. For all merged intensity files a SAD phasing attempt was run using phenix.autosol [100],
the known protein sequence and a resolution cutoff of 2.3 Å.

---

[5]for a derivation of the distribution before smoothing see appendix D

[6]to increase the robustness of the correlation, as there are outliers that skew the correlation to 0, -1 or 1 randomly

[7]Among the constraints that were left unused are the peak positions on the detector and the fact that the different unit cells matrices
are mainly just different rotations of each other.
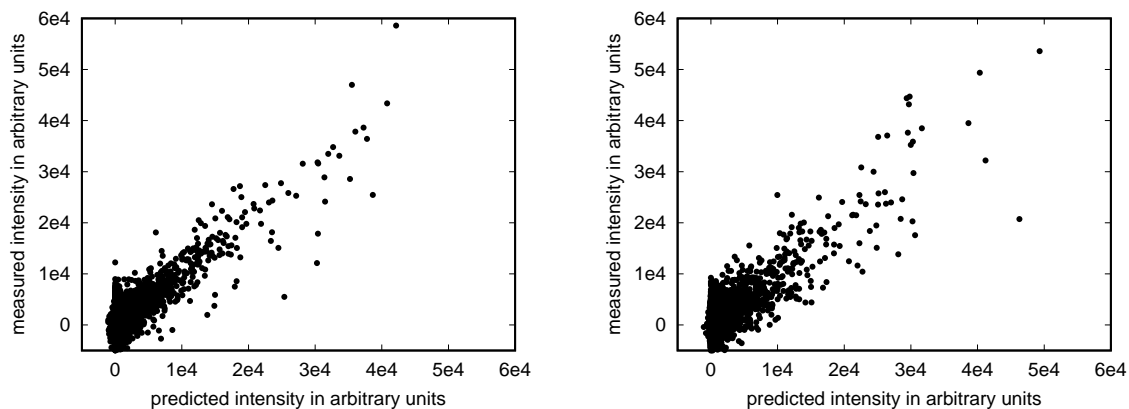
(a) Predicted and measured intensities from the first random half-dataset of dataset 1 that was used to fit all parameters to.

(b) Predicted and measured intensities from the second random half-dataset of dataset 1, using the parameters determined from the random first half-dataset of dataset 1 (also compare figure 31a). Note the slightly reduced correlation compared to figure 31a.

Figure 31: Two subsets of 10000 random pairs of predicted and measured intensities, one for the fit and one to test the fit for overfitting.

Figure 32: Comparison of structure refinement results of the granulin dataset (dataset 1) using `phenix 1.18-3855` to a resolution of $1.8\,\text{Å}$ of MGPCII, in green, and `partialator` 0.9, in violet. The bold dots represent the free R-factor, the small circles represent the R-work. The partiality model `ggpm` gave the best result for `partialator` for all sizes of subsets that were tested.
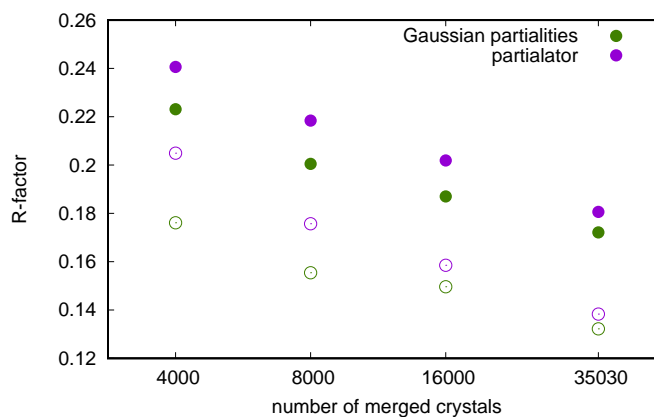
Figure 33: Maximum HySS correlation coefficient found during automatic SAD phasing using `phenix.autosol` from A_{2A} crystals [98] as a function of the number of crystals used during merging. The entries in green are for MGPCII, whereas the violet dots represent the results of `partialator`.
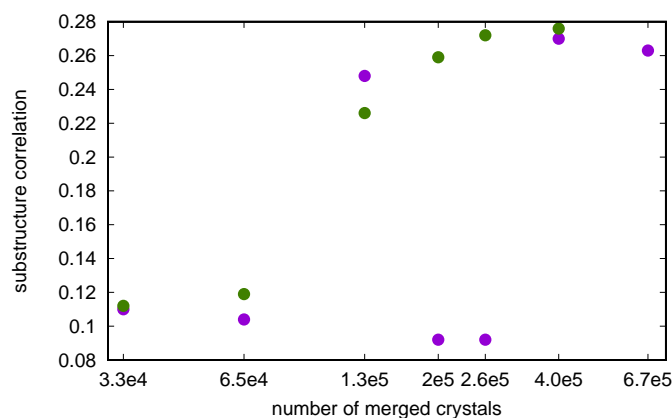
Figure 34: R-factors of the refinement of structures built during automatic SAD phasing using `phenix.autosol` from A_{2A} crystals [98] as a function of the number of crystals used. The entries in green are for MGPCII, whereas the violet dots represent the results of `partialator`. The solid dots are R-free and the small circles are R-work.
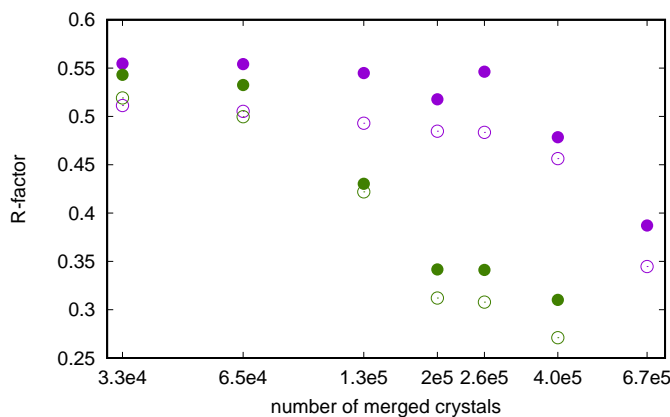
As can be seen from the hybrid substructure search (HySS) correlation coefficient in figure 33 and the R-factors that the automatic structure building and refinement achieved, see figure 34, the improved merging efficiency is reproduced for the anomalous signal too.

6.9.11. Discussion

Using Gaussian basis functions, approximations were developed that have enough degrees of freedom to account for most of the significant effects in macromolecular crystallographic experiments. These approximations were used to simulate diffraction patterns, which were visually very similar to measured diffraction patterns. Partiality estimation and post-refinement using these functions have reduced the number of measurements necessary for a given data quality in merged intensities. In the first example it reduced the number of patterns required to achieve a given R-factor by about a factor of 2 compared to CrystFEL's `partialator`. In the second example S-SAD phasing succeeded with about a quarter of the diffraction patterns. The range of datasets that were tested is not comprehensive, however, and `partialator` is not the only alternative, nor necessarily the best program, just the most commonly used.

There are many differences between MGPCII and `partialator`, partiality estimation being only one of them. Without exhaustive testing we are not able to tell precisely which differences provide the greatest improvement. A significant improvement can however be attributed to the error model used, which has been shown to improve merging on its own using a different approach [64]. Another important difference is that our method profits strongly from overprediction, adding many measurements with mostly insignificant intensities, by integrating reflections even if they are farther away from the ideal diffraction condition. It may seem that overprediction should not improve the precision of the merged result as strongly as it does, especially when the added intensities are mostly small or negative. However, the small intensity values outside of the diffraction condition act as a powerful constraint for determining reciprocal peak shape, mosaicity and strain.

Even though polychromatic diffraction of mosaic crystals can be described qualitatively, automatic refinement has proven to be difficult so far because predicted peak positions can vary by more than half the inter-Bragg distances. There are many more applications of approximating diffraction with Gaussian basis functions that remain to be explored. Pixel-wise refinements, like done in the program diffBragg [85], should lead to even better merging efficiency and a more precise detector geometry refinement at the cost of more computation time. The model could also be used to predict the intensity of peaks per frame

in a rotation series and therefore simplify the visual examination of the effectiveness of data processing, especially for peaks laying along the axis of rotation.

Integrated peak intensities are an easier target to optimize for than pixel-wise intensities, because there are many pixels per reflection. And because peak intensities are integrated over a larger pixel area on the detector, the geometry description only needs to be accurate enough for most of the peak intensity to fall within the integration area. But a consequence of integration is the drastic reduction of the number of constraints. Whereas pixel-wise optimization uses thousands of pixels, albeit with somewhat degenerate information[8], the number of constraints in a traditional cell parameter and orientation refinement during merging of a serial crystallographic datasets is just high enough to be clearly overdefined. This might mean that for datasets of very weakly diffracting crystals and without additional constraints a pixel-wise refinement is the only option.

Lastly, note the generaliy of this model: The same model can be used to simulate diffraction patterns and integrated intensities of serial monochromatic and polychromatic crystallography experiments. The analytical nature of this model makes analytical derivatives available, which is useful for mathematical optimization. It also makes deriving properties like peak locations and shapes and integrals over angular ranges and areas practical. Together this opens up a wide range of experiments where this model can be applied.

6.10. Covid Pharmaceutical Screening

Finally, after revisiting almost all steps in the crystallographic data processing pipeline, we are rewarded with a real life application of macromolecular X-ray crystallography. Structure determination has wide-ranged applications. And due to the global pandemic the need to combat a novel virus arose suddenly. Vaccines were not available yet, but the enzymatic machinery of the virus has one easily identified weak point. The 3C-like protease, or main protease (M^{pro}), is responsible for digesting the primary translation unit into several enzymes and structural proteins. As a peptidase it is a promising target, since there are already several antiviral pharmaceuticals that target peptidases of other viruses. A massive X-ray screening project [101] was initiated to test binding of around 6000 already approved pharmaceutical drugs. If any would bind with high affinity, the likelihood would be great that the enzyme would be inhibited. Since all compounds had already been tested on human subjects for other medical indications, they would likely have acceptable side-effects and could be authorized for more quickly. There are several ways to determine binding affinities experimentally, even inhibition constants can be measured directly, but because our expertise is in crystallography, structure determination was the method of choice for us to test drug binding. Also, crystallization is easily parallelized. While a crystal structure with a bound compound will determine binding only qualitatively, not quantitatively, it does show the mode of binding and thereby enables a more directed search for inhibitors. The following is an account focused only on my personal scientific involvement in the otherwise extensive screening project.

M^{pro} was crystallized in presence of each drug. The crystals were measured with the rotational crystal-lographic method, and processed with XDS [63]. They were phased by molecular replacement, using the known structure without the possibly bound drug. If the electron density allowed it, the structure of the

---

[8]In a single Gaussian approximation each peak on the detector can be described with 6 variables: Height, x and y coordinate of the centre, major and minor axis and orientation of the elliptical shape. Oversampling the shape does not add constraints in this approximation.

drug bound to the protein was modelled into it.

One of the first discoveries was the pharmaceutical drug HEAT, an α-adrenergic receptor antagonist. There extra electron density in the catalytic cleft fitted the tetralone moiety of HEAT, but it seemed as if the phenylethylamine moiety was missing, see figure 35. The methylenetetralone is bound to the catalytic cysteine via the sulphur in two distinct conformations, which are overlayd in the figure, which is why the sulphur appears duplicated. Based upon this observation a reaction mechanism in two steps is proposed, see figure 36. In the first step the drug is cleaved and activated by a base catalyst by deprotonating HEAT in the 2 position of the tetralone moiety. The tertiary hydrogen is assessed to be decently acidic for a C-H bond because of electron-pull and resonance stabilization of the carbocation structure. Phenylethylamine acts as a leaving group, leaving behind the reactive 2-methylenetetralin-1-one. The base $\overset{\ominus}{X}$ was unknown at the time, and is theorized to be M^{pro} itself [102], but it is also known, that the dissociation of HEAT proceeds even in aqueous solutions without any added base, albeit very slowly. Also, 2-methylenetetralin-1-one is a possible intermediate product in the synthesis of HEAT [103] and therefore a potential impurity.
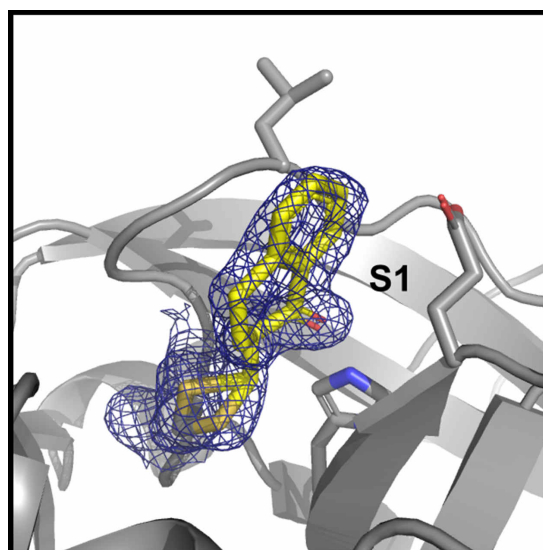
Figure 35: 2-methylenetetralin-1-one (from HEAT) (in yellow) bound to the catalytic pocket S1 of M^{pro} (mostly in gray). The experimentally determined electron density (blue mesh) fits the proposed Michael adducts well. Adapted from Günther et al. [104, figure 3].
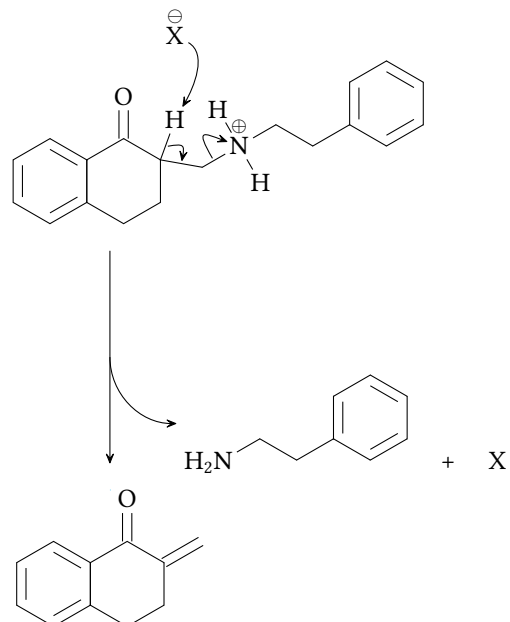
Figure 36: Decay of HEAT to 2-methylenetetralone via deprotonation of the C-H acidic hydrogen in position 2.

The proposed reaction mechanism for the inhibition of M^{pro} by HEAT is as follows: After having been activated by a base, the unsaturated ketone moiety is a good Michael acceptor. Cysteine-145 of M^{pro} can easily be deprotonated and the nucleophilic thiolate reacts with the Michael acceptor at the easily accessible carbon of the unsaturated methylene moiety. A bond is formed and the negative charge shifts through the conjugated system to the oxygen atom to form an enolate. The enolate will isomerize back to the ketone form. The keto-enol tautomery is an equilibrium, but the ketone is favored. As a result, the drug is permanently bound to the catalytically active cysteine-145 of M^{pro}, which will lead to an irreversible inhibition of its peptidolytic activity. Thus we can summarize that the reaction mechanism depends on three properties:

1. A leaving group that is stable enough to deliver the drug, but also cleavable under these conditions.
2. A good Michael acceptor
3. A general affinity to the binding pocket.

The binding pocket is rather hydrophobic and the natural substrates have have the consensus motive [A/V/P/T][X][L/F/V/M][Q]↓[G/A/S/N] [105] using the one-letter codes for the amino acids and '↓' indicating the cleavage site. This motive has a non-polar residue just a few atoms away from where the cleavage site would be, matching with the benzene ring of HEAT.
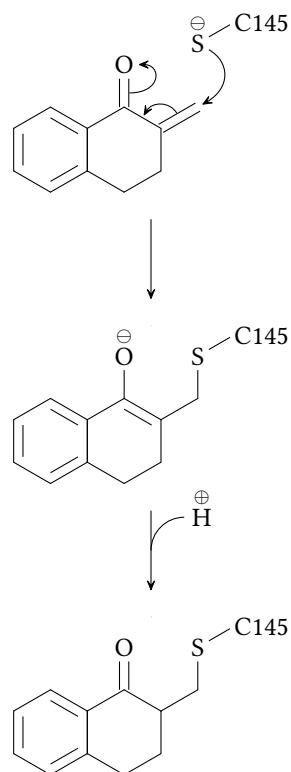
Figure 37: In the proposed mechanism 2-methylenetetralone acts as a Michael acceptor and the sulphur of cysteine-145 of M^{pro} as a nucleophile. This reaction forms a structure that fits the observed electron density very well.
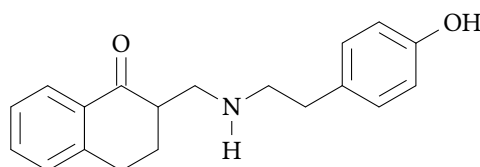
HEAT [IUPAC: 2-(beta-(4-hydroxyphenyl)ethylaminomethyl)tetralone]
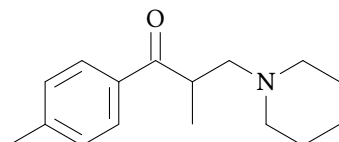
Tolperisone [IUPAC: 2-methyl-1-(4-methylphenyl)-3-piperidin-1-ylpropan-1-one]

Figure 38: Aligned molecular structures of HEAT and tolperisone

Of all the compounds in the cocrystallization assay, only tolperisone has all of the just mentioned three properties, and so the expectation was formed that tolperisone should be binding as well. However, tolperisone did not show binding at first. There are many things that can go wrong and initially the situation was chaotic, so a negative result could not be taken for granted. Also it is not guaranteed that every enzyme-substrate complex can crystallize in the first place. Further investigation then showed that it in fact can cocrystalize, see figure 39, which shows the Michael acceptor 4-MMPPO (2-methyl-1-(4-methylphenyl)-propenone) derived from tolperisone, bound to the same catalytic cysteine as HEAT. Just like with HEAT, it is also a known impurity [106]. The proposed reaction mechanism has since been refined to include putative catalyzing interactions with the catalytic cleft of the enzyme [102]. Later binding assays did confirm these results and both compounds could be shown to be antiviral with a critical concentration EC_{50} of $19.17\,\mu M$ for tolperisone and $24.05\,\mu M$ for HEAT [101, figure 2].
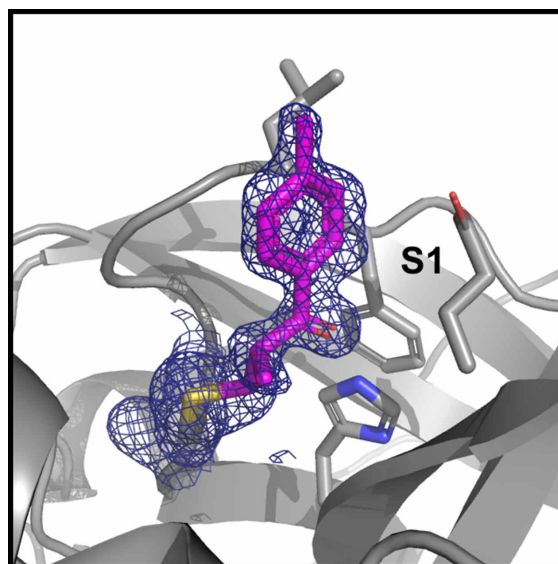
Figure 39: 2-methyl-1-(4-methylphenyl)-propenone (from Tolpersone) (in pink) bound to the catalytic pocket S1 of M^{pro} (mostly in gray). The experimentally determined electron density (blue mesh) fits the proposed Michael adducts very well. Adapted from Günther et al. [104, figure 3].

Even though this project did not lead to a cure for Covid, it shows how powerful structural crystallography and structural knowledge can be. Since then the virustatic drug Paxlovid was developed by Pfizer based on the same principles.

7. Structure Determination from Continuous Diffraction

Continuous diffraction is the non-discretized diffraction from single molecules or multiple molecules of the same kind with the same alignment, but without periodicity. This makes it distinct from crystallography, but many methods can be transferred, especially phasing and structure refinement. On the other hand, continuous diffraction provides more information and therefore other methods were developed that are mostly impossible with crystallographic data. The following is an exploration of how to transfer some of the methods from crystallography to the richer continuos diffraction data.

According to the Nyquist-Shannon sampling theorem, the crystallographic structure factors provide sufficient sampling for arbitrary crystal structures, or stated in another way, the discrete Fourier transform is invertible. Any finer sampling would not provide additional constraints, because the samples would not be independent any more. However, only the squared amplitudes of the structure factors are observed, and so half the constraints, the phases, are missing. Knowing that the object is entirely real does not change this assessment, because the reduction of variables by the factor two is matched by a reduction in the number of constraints because the Fourier transform of a real function has a point-symmetry, this follows from equation 78. Additional constraints are needed. In crystallography this is usually the prior knowledge that molecules are made up of atoms and bonds with very particular angles and distances. If there is more diffraction data, like from different crystal forms, that provide different sampling of the Fourier transform [62], the constraint ratio is such that the solution is unique. Similarly, diffraction from singular objects also provides twice as many constraints in each dimension. This is because the squared amplitudes of the Fourier transform of the object are equal to the Fourier transform of the autocorrelation of the object (equation 76). The autocorrelation of an object is two times as wide as the object in each dimension and each point is a linearly independent combination of the values that describe our object (apart from the point symmetry for entirely real objects). Sufficient sampling according to the sampling theorem is therefore twice of what it was previously in each dimension. A more detailed account of this argument can be found in Morgan et al. [107].

As introduced earlier however, diffraction from singular macromolecules is not strong enough yet to be usable. But until the time comes there sometimes is diffuse diffraction from crystals that looks like the diffraction from single molecules. This is due to a particularity of these crystals: The molecules are randomly shifted as a unit each. The peaks in the diffraction pattern are due to the three-dimensional periodicity, thus they only capture the average position and density, which is smoothed by the random shifts. The correlated movements result in an eletron density average that appears smoother than the electron density of each asymmetric unit. This dampens the peaks intensities, but between the peaks the diffraction from the shifted molecules can be described by the Fourier transform of the difference between each shifted molecule and the average density, like an edge-enhanced version of the electron density of this molecule. In this way we have access to something that is very similar to single molecule diffraction, the diffraction of an oriented arrangement of molecules of the same kind. Methods have been developed that make use of the overconstrained nature of this data to reconstruct the electron density with iterative projection algorithms [108]. Iterative projection algorithms work by iterating between two or more constraints trying to fulfill each at a time until all are fulfilled, provided that they can be fulfilled at the same time. If the problem is highly overconstrained and a solution exists, these algorithms often find solutions quickly even for high dimensional problems. Of active research is the question of how to best adapt these methods when

rotational disorder smears the diffuse diffraction radially, how measurement errors affect the method, and how to help these methods with other prior information, especially about molecular structures.

On the surface iterative projection algorithms seem impossible to combine with structural information, because it is hard to express structural constraints as a projection. In each step we would have to find the structure that describes the diffraction or electron density as closely as possible, while still being structurally sound. This is a very hard problem, especially in contrast to the projections that are usually employed. Error reduction is the most trivial iterative projection algorithm, it iterates between two projections until convergence. The first constraint is the finite support. Because we know the diffraction is from a single molecule, the Fourier transform of the measured squared amplitudes is the autocorrelation of the object, up to measurement errors. The autocorrelation is two times as large as the object itself in each dimension. And with this we know the maximum extent of the object we want to reconstruct, this is the support constraint. The support constraint $S(\vec{x})$ is like a top hat in each direction with the known maximum extent. Because the squared amplitudes are shift invariant (the shift is in the phase), the location is irrelevant and for simplicity the support will be centered at $\vec{0}$. Applying the support constraint amounts to multiplying the putative electron density $\rho(\vec{x})$ with the support:

$$\rho(\vec{x}) \leftarrow \rho(\vec{x}) S(\vec{x}) = \begin{cases} \rho(\vec{x}) & \text{if } S(\vec{x}) = 1 \\ 0 & \text{otherwise} \end{cases} \tag{152}$$

The second constraint is that the Fourier amplitudes $|F(\vec{q})|$ of the current estimate of the electron density need to correspond to the square root of the measured Fourier intensities $\sqrt{I_i}$:

$$F(\vec{q}) \leftarrow \frac{\sqrt{I(\vec{q})} F(\vec{q})}{|F(\vec{q})|} \tag{153}$$

Integrating more constraints into this framework requires expressing them in terms of projections. But there is a connection with function optimization, namely Newton's method. If we take each point of the unknown electron density inside the support as a variable we want to optimize for, and the sum of quadratic distances between measured structure factor amplitudes and the structure factor amplitudes of the reconstructed electron density as a target, the support constraint is always implicit and it can be shown that the second projection corresponds to a Newtons update.

$$x \leftarrow x - f(x) \left(\frac{\partial f}{\partial x}(x) \right)^{-1}$$

Using Newtons method for finding minima and maxima of smooth functions involves the first and second derivatives of the target function f, as a reminder:

$$x \leftarrow x - \frac{\partial f}{\partial x}(x) \left(\frac{\partial^2 f}{\partial x^2}(x) \right)^{-1}$$

Calculating the first and second derivatives in preparation for Newtons method:

$$f(\rho(\vec{x})) = \frac{1}{2} \int_{\text{dom}(\vec{q})} ||\mathcal{F}_{\vec{x}}(\rho(\vec{x}))(\vec{q})| - A(\vec{q})|^2 \, d\vec{q}$$

$$\frac{\partial \mathrm{f}\left(\rho\left(\vec{x}\right)\right)}{\partial \rho\left(\vec{x}\right)}\left(\rho\left(\vec{x}\right)\right) = \int_{\mathrm{dom}\left(\vec{q}\right)} \mathcal{F}_{\vec{q}}\left(\frac{\mathcal{F}_{\vec{x}}\left(\rho\left(\vec{x}\right)\right)\left(\vec{q}\right)}{|\mathcal{F}_{\vec{x}}\left(\rho\left(\vec{x}\right)\right)\left(\vec{q}\right)|}\left(|\mathcal{F}_{\vec{x}}\left(\rho\left(\vec{x}\right)\right)\left(\vec{q}\right)| - \mathrm{A}\left(\vec{q}\right)\right)\right)\left(\vec{x}\right) d\vec{q}$$

$$= \left(\rho\left(\vec{x}\right) - \mathcal{F}_{\vec{q}}\left(\frac{\mathrm{A}\left(\vec{q}\right)\mathcal{F}_{\vec{x}}\left(\rho\left(\vec{x}\right)\right)\left(\vec{q}\right)}{|\mathcal{F}_{\vec{x}}\left(\rho\left(\vec{x}\right)\right)\left(\vec{q}\right)|}\right)\left(\vec{x}\right)\right) \int_{\mathrm{dom}\left(\vec{q}\right)} 1 d\vec{q}$$

$$\frac{\partial^2 \mathrm{f}\left(\rho\left(\vec{x}\right)\right)}{\left(\partial \rho\left(\vec{x}\right)\right)^2}\left(\rho\left(\vec{x}\right)\right) = \int_{\mathrm{dom}\left(\vec{q}\right)} 1 d\vec{q}$$

Newtons update is:

$$\rho\left(\vec{x}\right) \leftarrow \begin{cases} \mathcal{F}_{\vec{q}}\left(\dfrac{A\left(\vec{q}\right)\mathcal{F}_{\vec{x}}\left(\rho\left(\vec{x}\right)\right)\left(\vec{q}\right)}{|\mathcal{F}_{\vec{x}}\left(\rho\left(\vec{x}\right)\right)\left(\vec{q}\right)|}\right)\left(\vec{x}\right) & \text{if } |\mathcal{F}_{\vec{x}}\left(\rho\left(\vec{x}\right)\right)\left(\vec{q}\right)| > 0 \\ \rho\left(\vec{x}\right) & \text{otherwise} \end{cases}$$

Which is virtually identical to the second projection step of the error reduction algorithm, while the first projection is always implicit. We can conclude that the error reduction algorithm is similar to Newtons method minimizing the distance between simulated and measured amplitudes [109]. And more generally, a projection is the solution to a least squares problem in that space. So iterative projection algorithms can be emulated by cycling through complementary targets and optimizing them one by one. Typically in mathematical optimization, multiple targets would be combined into a single target, which is then optimized. This has the benefit of taking the shortest route to the next optimum and finding the best compromise when not all targets can be satisfied completely or when they are not sufficiently constrained on their own. But cycling through the targets has the opposite effect, which can be an advantage too, it explores more of the parameter space, thereby increasing the tendency towards global instead of local optima [110]. A hybrid approach can combine both aspects, by first exploring the space cycling through the targets and then locking in and finding the best solution by optimizing all targets jointly. Intermediate solutions, for example estimating the current uncertainty in the model and equivalently phase errors [111], or hybrid input-output, exist also. What had been unexplored, was a way to include prior structural knowledge in the phasing algorithms. To test the new options a program has been created, that applies these principles to synthetic diffuse diffraction data. The distance between two diffraction patterns has no unit, the energetic state of a structure is in units of energy per mole. But all targets can be expressed in terms of probability, thereby bringing them onto the same scale. The minimum of quadratic distances corresponds to the maximum of Gaussian likelihoods and log-likelihoods, and the conversion between energy and probability in a thermodynamic equilibrium is the Boltzmann-constant k_B and the temperature T:

$$\mathrm{p}(\text{structure}|\text{molecule}) = \exp\left(-E/(k_B T)\right) \tag{154}$$

There are several libraries that can approximate the energetic state of molecular structures including derivatives. The toolbox Open Babel [112] contains all necessary functions and was easy to integrate. Complementary methods can determine the atomic composition and connectivity of the molecule with great accuracy, so we do not need to concern ourselves with the probability of observing a specific molecule p(molecule), only with the structure of a given molecule p(structure|molecule). The conversion factor between quadratic distances (d^2 between prediction and measurement) and log-likelihoods is the standard

deviation σ, if we assume a normal distribution of measurement errors:

$$-\log\left(\mathrm{p}(d^2)\right) = \frac{1}{2}\left(\frac{d^2}{\sigma^2} + \log\left(2\pi\sigma^2\right)\right)$$

Other models, possibly some that are more robust, are possible too. Adding both targets to form a joint target is like applying Bayes theorem:

$$\mathrm{p}(\mathrm{model}|\mathrm{data}) = \frac{\mathrm{p}(\mathrm{data}|\mathrm{model})\mathrm{p}(\mathrm{model})}{\mathrm{p}(\mathrm{data})}$$

The prior probability of the data $\mathrm{p}(\mathrm{data})$ does not need to be determined explicitly. In a full Bayesian inference step, it is simply the missing normalization factor that can be determined by integration. In a maximum a posterior approach it is a constant that does not change the location of the optimum. Using the modified Cromer-Mann coefficients, see section 6.7, the expected diffraction can be expressed analytically and its derivatives can be computed symbolically using computations on the order of $O(nm)$, where n is the number of atoms and m the number of data-points. The diffraction and the derivatives can be approximated with computational cost on the order $O(n + m\log(m))$ using the FFT, by applying the FFT to a sampled electron density and its derivatives, just like it is done in many crystallographic refinement schemes.

Lastly, crystals from biological macromolecules usually contain unstructured solvent reducing the contrast of the electron density. If we want to keep all options we had up to now, we need a way to model the solvent that still allows efficient derivatives with respect to all model parameters, ideally symbolically. A simple solvent model is the following. Everywhere where there are no atoms of the macromolecule, there is solvent, starting at a certain distance, corresponding to the average distance between solvent molecules. The solvent usually is water and the average distance between water molecules is around $3.1\,\text{Å}$ (σ_s). The distance between atoms of the molecule and the solvent should be similar, but it possibly depends on the kind of atoms as well. The average electron density of water (ρ_w) is $0.334\,\text{Å}^{-3}$. Both can be free parameters of the model. The solvent is modeled with a constant density of $0.334\,\text{Å}^{-3}$, and by subtracting a Gaussian with width $3.1\,\text{Å}$ for each atom of the molecule. Mathematically the solvent exclusion can be expressed by a product of complementary probabilities:

$$\rho_s\left(\vec{x}\right) = \rho_w\left(1 - \prod_i^N\left(1 - \exp\left(-\frac{1}{2}\left(\vec{x} - \mu_i\right)^2\sigma_s^{-2}\right)\right)\right) \tag{155}$$

This expression can be expanded to yield an expression that is easy to approximate and Fourier-transform:

$$\begin{aligned}
\rho_s\left(\vec{x}\right) = \rho_w + &\left(\sum_i^N \exp\left(-\frac{1}{2}\frac{\left(\vec{x} - \mu_i\right)^2}{\sigma_s^{-2}}\right)\right) \\
&- \left(\sum_i^N\sum_j^N \exp\left(-\frac{1}{2}\frac{\left(\vec{x} - \mu_i\right)^2}{\sigma_s^{-2}}\right)\exp\left(-\frac{1}{2}\frac{\left(\vec{x} - \mu_j\right)^2}{\sigma_s^{-2}}\right)\right) \\
&+ \left(\sum_i^N\sum_j^N\sum_k^N \exp\left(-\frac{1}{2}\frac{\left(\vec{x} - \mu_i\right)^2}{\sigma_s^{-2}}\right)\exp\left(-\frac{1}{2}\frac{\left(\vec{x} - \mu_j\right)^2}{\sigma_s^{-2}}\right)\exp\left(-\frac{1}{2}\frac{\left(\vec{x} - \mu_k\right)^2}{\sigma_s^{-2}}\right)\right) \\
&- \cdots
\end{aligned}$$

Higher order sums quickly subside and only products between nearby atoms contain significant cross-terms because the Gaussian function is highly localized. Nearby contributions can be filtered out using spatial indexing with only constant overhead. Then, an efficient approximation with a symbolic Fourier transform can be found by neglecting all cross terms further than 3σ apart.
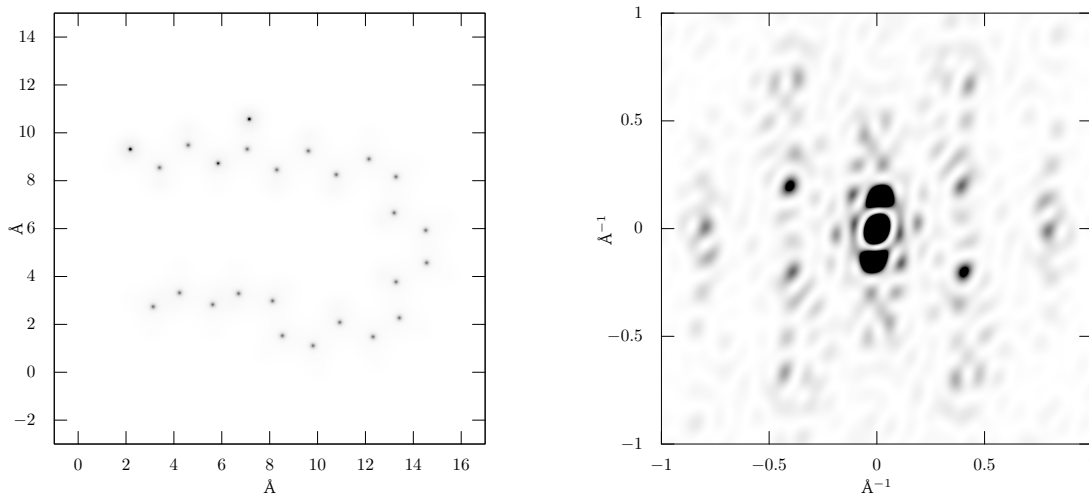
(a) Electron density of anandamide projected onto the xy-plane.

(b) A cut through the Fourier amplitudes squared of the electron density of anandamide. This is the kind of data one could gather if diffraction from single molecules were strong enough.

The methods described were implemented in C++ and tested with synthetic data. The program code has been published under `https://github.com/lykos/diffuse` and preliminary experiments show that structures of small molecules can be reconstructed, even from noisy data and when there is considerable structural variability, like in the case of anandamide. Figure 40a shows the electron density of anandamide, which can be retrieved from continuous diffraction like pictured in figure 40b without additional noise. The additional structural information helps a lot when the data is otherwise too noisy and the initial guess is not too far off. However, even small structures can have several local energetic minima. Optimization of larger structures incur much higher computational cost and have not been investigated, but there is no reason why the methods discussed should not work in principle. In conclusion, the main result of this investigation is to show that iterative projection and function optimization are entirely compatible and structural information can be integrated with iterative projection algorithms just like projection steps can be used in structural refinement.

8. X-ray Detector Calibration with Histograms

In integrating CMOS detectors (a common type of integrating detectors in crystallography) each pixel is potentially slightly different and its associated analog-to-digital-converter (ADC) has slightly varying characteristics, that can even drift over time. Therefore they need to be calibrated [113]. As it is experimentally difficult to provide a consistent calibration signal of known intensity equally at each pixel, what is often done while such a calibration signal is not available, is to record many images of a low signal, so that each pixel sometimes measures no photon, and sometimes one or more. When the precision and resolution of the ADC is high enough - which is usually the case, because they are designed to resolve single photons - the signal levels per photon can be inferred and the offset and gain of each pixel can be calibrated. The location of the first peak in figure 41 is the offset of this pixel, the width of the first peak is the noise floor and the spacing between the peaks is the gain. As an intermediate step it is convenient to accumulate the measurements for each pixel into a histogram. Because a histogram is a compressed representation of all the data for each pixel, many quantities can be calculated more quickly. For gain switching detectors this only works in the highest gain mode, in lower gain modes the precision is lower and the photon counts cannot be inferred any more. Other methods need to be employed there, possibly using histograms as well.
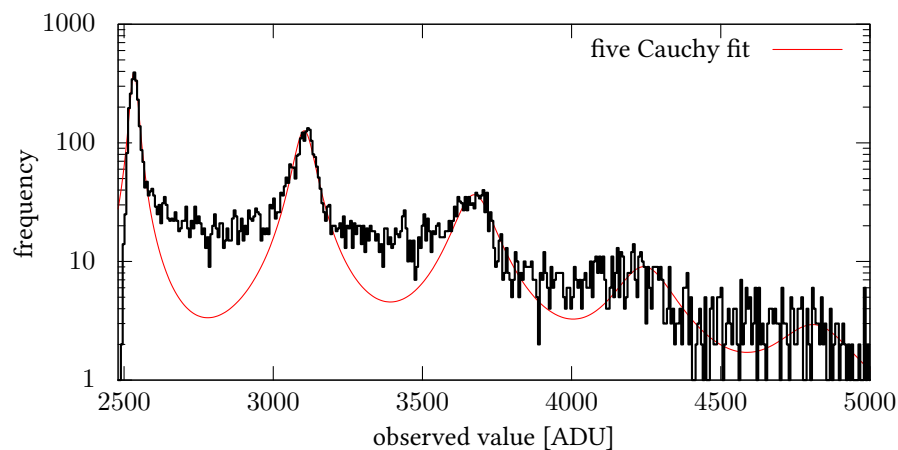
Figure 41: Histogram of observed detector values for one pixel in a series of 10000 images from a integrating CMOS detector (model Jungfrau). The periodic peaks correspond to the typical values measured for 1,2,3... photons. Five regularly spaced Cauchy functions were fitted to the histogram, the fit is drawn in red.

There are several ways to compute a histogram, but when the data is sparse (that is, only a small fraction of attainable signals were measured and it is not known in advance which fraction) the fastest way is to use a hash table. The standard `C++` hash table implementation however is fairly wasteful with computer memory, especially for small data types, when few bytes that are typically recorded for each pixel in each of the tens of thousands of images, and random access memory (RAM) is expensive. There is another way to compute histograms, by first sorting the values, then counting consecutive runs of equal values. Because the read-write patterns of the sorting algorithm quicksort are much more linear than the pseudorandom access patterns of hash-tables, this would be feasible even with slower but cheaper memory, like memory mapped

---

[8]As a consequence of the requirement for stable data pointers, which is a rarely needed feature.

94

files on SSDs. But the sheer volume of transactions will quickly degrade the flash storage cells. However, counting the values while sorting them can drastically reduce the memory usage and even approach the theoretical minimum. This can be achieved by using a sorted array of pairs of values and frequencies. For each value, if it is already in the array, the associated frequency will be incremented. If the value is new, an entry is created at the correct position in the ordered list. This way, the frequent lookups can be done quickly using binary search, but the insertions need to shift half the array on average. Still, this solution is sufficient, very memory efficient and fast enough. It can be much faster, if even just a small fraction of slots scattered across the array were kept empty, so that insertions only need to shift the entries until a free slot is encountered. This only works when the values and empty slots are distributed randomly, otherwise they could cluster pathologically. For this exact purpose hashing is used in hash tables. A hash is a function that distributes input values uniformly onto a range of output values with high probability. The hash table `patchmap` is the result of this synthesis of sorting and hashing to achieve memory efficient hashing.

9. A Memory Efficient Hash Table

A hash table is a data structure commonly used to associate keys with values to implement dictionaries or to test if a key is part of a set of keys to implement set operations. Hash tables are an essential part of many algorithms because they allow associative retrieval of n elements in average constant time and have a space complexity of $O(n)$. The keys are associated to the values by computing the hash value of each key and storing the key-value pair in an array at the location indicated by this value. As long as there are no collisions, that is two different keys that produce the same initial position, different hash map implementations are conceptually identical. This can be the case in the limit for an infinitely large array or in the case of a perfect hashing function that produces different locations for all keys. As memory is constrained and there can be no one predetermined perfect hash function for arbitrary keys, the resolution of hash collisions is the main signature of different hash map implementations. The better the hash collisions are handled, the more of them can be allowed to occur and the smaller the internal array can be. Collision resolution strategies are commonly classified into open addressing and chaining. A probing strategy finds another free position in the same table. The position to store an entry is found by starting from the position indicated by the hash value and then searching for a free bucket with some probing strategy. Chaining means that each position in the array stores a pointer to a secondary data structure that stores all the entries with the same hash value. Testing the initial position incurs a cache miss in most cases. A cache miss is when data, that cannot be found in any of the caches of the CPU, is requested. The time it takes to retrieve random data from the main memory usually determines the performance of a hash table and the first cache miss cannot be avoided due to the pseudorandom nature of hashing. The second step (chaining or probing) occurs more frequently the fuller the table becomes, it is a function of the load factor α. The load factor is defined as the number of elements stored in the table divided by the total number of positions available in the table. Therefore, for a given hash function and collision resolution scheme, the hash table with a lower load factor is faster because it has to resolve the least collisions. The best hash table is the one that enables insertions lookups and possibly deletions at the lowest cost. Cost in computing essentially is two-dimensional - computation time and memory space needed. In different settings memory space and computing time can be constrained differently, but in general the total cost is virtually proportional to the product of time and space. The hash table with the best trade-off therefore is the one that can achieve the

lowest cost given by said product. The `patchmap` is a hash table developed to achieve a good trade-off in this regard.

The central idea of the novel probing scheme [114] is to use an ordering of the keys consistent with the hash value of the key to resolve the collisions that occur in the hash table. Using sorting to resolve hash collisions has been suggested before, most notably by Amble and Knuth [115]. The difference in the proposed algorithm is not using the sorting given by the value of the key but the hash of the key. With linear probing this leads to a global order of the keys stored in the hash table and therefore allows for efficient lookup using interpolation search and binary search. Interpolation search has the best average case and a binary search provides the best worst case complexity for lookups in a sorted array. This means that the lookup, successful or not, of a key will have a worst case time complexity of $O(\log(n))$, and an average complexity of $O\left(\log\left(\log((1-\alpha)^{-1})\right)\right)$. These guarantees are as good as for optimally ordered hash tables [116][9] and stronger than the guarantees that most other probing strategies can give for lookups [117]. And just like for binary tree hashing, which approximates optimal ordering, the faster lookups come at higher cost for insertions of up to $O(n)$ for a full table, but are constant for tables that have a load factor less than 1. Note that the *probable* worst case for advanced probing strategies equally is $O\left(\log(n)\right)$ as it also approaches optimal ordering, and that robin hood hashing [118] with dynamic resizing achieves the same worst case complexity with exceedingly high probability, as the hash table can be resized as soon as the maximum displacement is larger than $O(\log(n))$, but the iterative resizing is not guaranteed to terminate with $O(n)$ space. Robin hood hashing with dynamic resizing is successfully implemented in the `flat_hash_map` and `sherwood_map` [119] of Malte Skarupke for example.

9.1. The Algorithm

The key-value pairs are stored in an array of size m. A binary mask of the same size is used to indicate whether a bucket in the array is free or set. The proposed algorithm requires a hashing function, preferably bijective, an order on the keys to be stored if the hashing function is not bijective and a mapping function that compresses the range of the hash value onto the range of the size of the array while preserving the order of the hash. The hash function can be any function that maps the keys onto positive integers less than some integer, but as the resolution of collisions is essentially a variant of linear probing it is crucial for good performance that the hash values are evenly distributed. The map function used is the product of the hash value h with m and then dividing the result by the maximum hash value $\max_h$ plus one:

$$\text{map}(h) = \lfloor \frac{h\,m}{1 + \max_h} \rfloor$$

Two types of comparisons need to be considered, comparing keys with keys and keys with free bucket positions. When comparing two keys a key is considered to be less if its hash value is less than the other or in case the hash values are equal if the key itself is less than the other key. When comparing a key to an empty position it is considered to compare less if the application of the map function to the result of the hash function of the key is less than the position: position < map(hash(key)) .

---

[9]The authors cautiously claim that the average complexity of successful is bounded by a constant but give no proof.

[9]This can be computed efficiently for fixed width integer types using long multiplication with two fixed width integers of the same length or a simple multiplication with a fixed width integer with double the size, followed by bit shifts, this technique is known as `fastrange` [120].

As long as there have been no collisions the keys stored in the hash table are already in the order given by the hash value. The lookup operation depends on this order. The operations for insertion and deletion therefore need to be implemented in a way so that they uphold this order. This is achieved by finding the closest free bucket starting from the position indicated by map(order(key)). The key-value pair is inserted into the free bucket and then swapped with neighbouring entries until the order of the hash table is restored, see figure 43. This procedure behaves like linear probing in a complexity analysis and therefore needs the same number of probes, $O(1 + (1-\alpha)^{-2})$ [121, 6.4 Hashing]. The term $\frac{1}{4}\left(\frac{\alpha}{(1-\alpha)^2} + \alpha\right)$ has been found to describe the required number of swaps wells. For a numerical simulation see figure 42. Deletion works analogous to insertion, but in reverse. The procedure for inserting is conceptually identical to the insertion step in Algorithm B in Knuth and Amble [115] with linear probing, but instead of comparing the keys directly, the hash value of the keys is being compared.
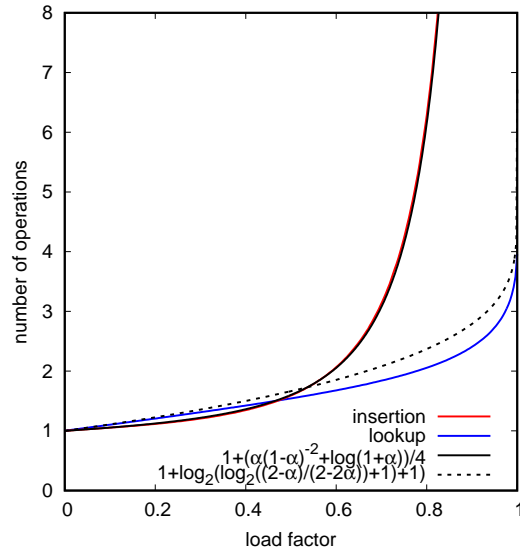
Figure 42: Average cost for insertion and lookup in the hash map with 2^{28} keys. Insertion (in red) shows a strongly quadratic behavior and the function $1 + \frac{1}{4}\left(\frac{\alpha}{(1-\alpha)^2} + \alpha\right)$ is drawn as a solid black line. Lookup (in blue) is almost linear in α. The theoretically derived complexity bound $1 + \log_2(\log_2(\frac{2-\alpha}{2-2\alpha})+1)$ is drawn as a dashed black line.

Given a hash map where the keys are globally ordered, an interpolation search can be employed to retrieve keys. Starting with the first and last element as lower_limits and upper_limit respectively of the array, the range is iteratively subdivided by linear interpolation until the key at midpoint equals the key that needs to be found or the range is empty or the maximum number of iterations has been reached. If the number of iterations exceeds $O(\log(n))$ the search switches to a binary search. Almost always the interpolation search will terminate before that, but switching to a binary search guarantees that the search will terminate in $O(\log(n))$.

For a table that has been filled completely, the average time complexity for key retrieval will be $O(\log(\log(n)))$ and the worst case time complexity will be $O(\log(n))$ as this case reduces to the search in a sorted list of uniformly distributed integers. The algorithm for inserting new keys behaves like linear probing in the way that primary clustering occurs. The complexity for lookup with a linear search would therefore be $\frac{1}{2}(1 + (1-\alpha)^{-1})$ for a successful lookup and at most one probe more for an unsuccessful search. This is similar to algorithm B in Knuth and Amble [115], where the search can be concluded as soon as a key outside of the search range is encountered. In contrast to these ordered hash tables, the proposed hash table has a global order, made up of small patches that are ordered internally *and* with respect to each other. This means probe positions can be interpolated from the keys at the upper and lower limit starting from the beginning and end of the table, reducing the number of probes required for lookups to $1 + \log_2(1 + \log_2(\frac{2-\alpha}{2-2\alpha}))$. The exact term depends on implementation details like the way the position is interpolated. For a numerical simulation see figure 42.

operation	average	worst case
insertion	$O(1 + (1 - \alpha)^{-2})$	$O(n)$
deletion	$O(1 + (1 - \alpha)^{-2})$	$O(n)$
lookup	$O(1 + \log_2(1 + \log_2(\frac{2-\alpha}{2-2\alpha})))$	$O(\log_2(n))$

Table 8: Computational complexities for hash table operations in the `patchmap`. $\alpha = \frac{n}{m}$ is the load factor, m the number of buckets and n the number of elements stored in the table.

operation	average	worst case
insertion	$O(1 + (1 - \alpha)^{-1})$	$O(n)$
deletion	$O(1 + (1 - \alpha)^{-1})$	$O(n)$
lookup	$O(1 + (1 - \alpha)^{-1})$	$O(n)$

Table 9: Computational complexities for hash tables using double hashing. $\alpha = \frac{n}{m}$ is the load factor, m the number of buckets, n the number of elements stored in the hash table.

Step 0, the patch before insertion:

key	x	6	2	5	7	9	3	6	x	x	...
hash	x	1	2	3	3	4	5	6	x	x	...
mask	0	1	1	1	1	1	1	1	0	0	...

Step 1, inserting key 1 with hash 5 into a patch:

key	x	6	2	5	7	9	3	6	1	x	...
hash	x	1	2	3	3	4	5	6	5	x	...
mask	0	1	1	1	1	1	1	1	1	0	...

Step 2, restoring the order in the patch:

key	x	6	2	5	7	9	1	3	6	x	...
hash	x	1	2	3	3	4	5	5	6	x	...
mask	0	1	1	1	1	1	1	1	1	0	...

Figure 43: Three step scheme of inserting the key 1 with the hash value 5 into a patch in the `patchmap`. The hash function is non bijective. The map function in this example is the identity function. The key 1 is inserted at the position closest to map(hash(key))=5, which is position 8, then it is swapped to the left until the order of the patch is restored.

9.1.1. Experimental Results

Using the proposed algorithm a hash table can be implemented that has a good tradeoff between memory requirement and time used for lookup, insertion and deletion. The average time for inserting into, deleting from and searching for a key that has 50% chance of being in the table has been computed for up to 2^{27} keys, for hash tables mapping from 32 bit integers to 32 bit integers. The tests were carried out on a

64 bit linux machine with an Intel Core i5-3320M CPU @ 2.60GHz and a CAS latency of 13.5 ns. The `patchmap` was compared to `khash` from attractivechaos `klib` [122], Malte Skarupke's `bytell` [123] and his `flat_hash_map` [124] using robin hood hashing, `google::sparsehash` from the public github release, `sparsepp` [125], a `sparsehash` fork and `std::unordered_map`. All executables were compiled with the highest optimization level and using `clang` version `7.0.1`.

(a) successful lookup

(b) unsuccessful lookup

(c) insertions
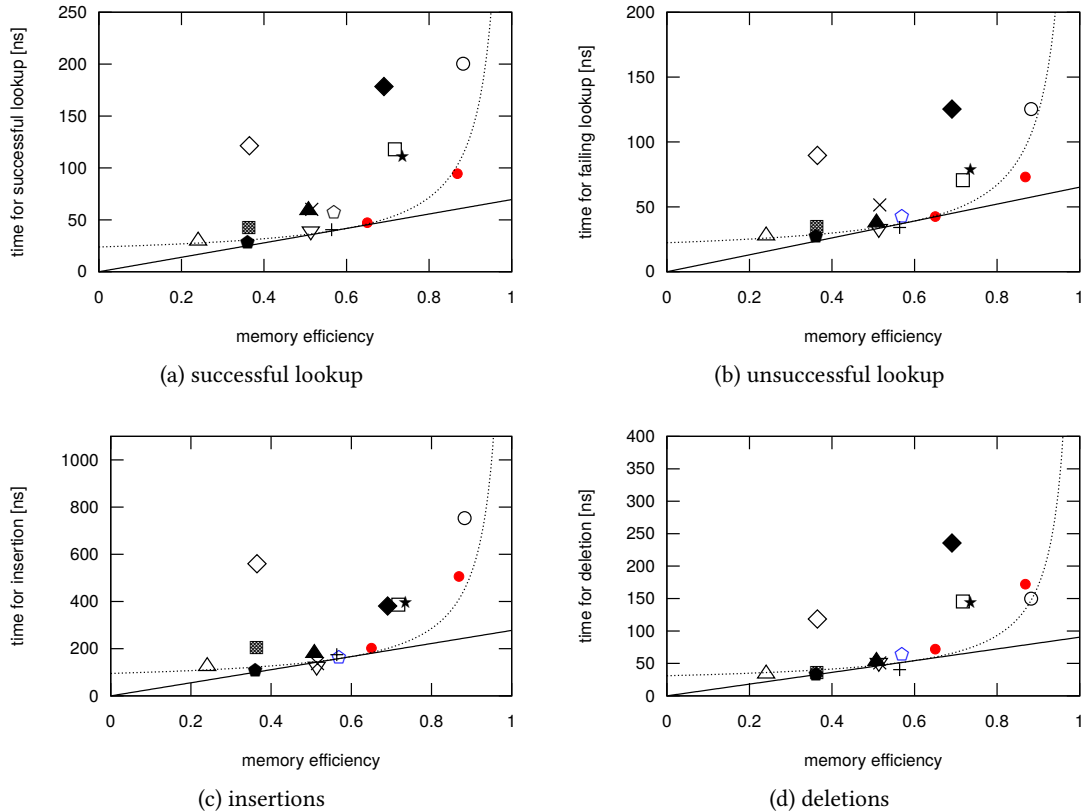
(d) deletions

Figure 44: Selected hash table implementations and their performance given in average memory efficiency on the horizontal axis and average time for lookup, insertion and deletion of a random key in nanoseconds on the vertical axis, computed for up to 2^{27} keys. Insertion contains the time needed for dynamic expansions. There are two points for the `patchmap` with two different tradeoffs between memory efficiency and speed.
patchmap: •, khash: ×, bytell: +, google::sparse_hash_map: ○, google::dense_hash_map ⌂, flat_hash_map: △, std::unordered_map: ◇, sparsepp: □ The dotted line is the empirically determined typically achievable trade-off between memory efficiency and speed as a function of the memory efficiency, given by $O(1 + (1-\alpha)^{-1})$. The black line is an isoline for best trade-off that can be achieved, given this trade-off, at any memory efficieny. Implementations that lie close to or below this line are especially cost effective in the space-time metric.

From figure 44 it can be seen that the `patchmap` sits right at the apparent barrier outlined by other good hash map implementations like `khash`, `bytell` and `google::sparsehash` on a map of memory usage and time efficiency. This broad comparison cannot capture all aspects of hash table performance and it is unfair in several regards. The comparison to the `std::unordered_map` is unfair, as it has to adhere to the `c++` standard requiring that references to elements stay constant for all hash table operations. The

hash tables `sparsepp` and `google::sparsemap` are designed to minimize the peak, instead of the average memory consumption. Deleting elements from the table is done by marking elements as deleted instead of actually removing them from the table in all implementations except the `patchmap`, `bytell` and `std::unordered_map`. This saves time in the short run, but having deleted elements still in the table can lead to a degradation of performance if these elements are allowed to accumulate. Such behavior can be triggered by repeatedly inserting and erasing many keys, see table 10, but it is considered rare in real world applications.

hash table	slowdown	rehashing
`patchmap`	no	none
`bytell`	no	none
`std::unordered_map`	no	none
`flat_hash_map`	no	none
`sparsepp`	minor	frequent
`google::sparse_hash_map`	minor	frequent
`google::dense_hash_map`	severe	rare
`khash`	severe	occasional

Table 10: Performance degradation of hash table implementations when keys from a fixed set are repeatedly chosen at random and then alternatingly inserted or deleted from the hash table to show the effect of many dead keys in the table.

9.1.2. Availability

An implementation of the patchmap as outlined here is available on github under an unmodified MIT license:

`https://github.com/1ykos/ordered_patch_map`

And the accompanying publication can be found here:

`https://infocomp.dcc.ufla.br/index.php/infocomp/article/view/581`

9.1.3. Discussion

Sorting based on the hash is almost identical to robin hood hashing with bidirectional linear probing. When deciding which key to displace and which to keep at a given position, the larger key will also be displaced more than the smaller key. The crucial difference is that when ensuring an absolute order a guarantee can be given for the worst time complexity because a binary search can be employed. The proposed data structure is also very similar to the one used in Gonnet and Ian Munro [126] for sorting. The application is different, and therefore deletions and the opportunity for interpolation search for lookups have not been considered. But regarding the data structure itself, the only relevant difference is that the entries are hashed and sorted based on a hash function.

Using a global ordering of the keys stored in a hash table, a worst case guarantee can be made of a time complexity for associations of $O\left(\log(n)\right)$, and an average complexity of $O(1 + \log_2(1 + \log_2(\frac{2-\alpha}{2-2\alpha})))$ can be achieved. Chaining hash tables can achieve better or at least equal bounds on the complexity of all hash table operations. However, this comes at the cost of a higher degree of indirection due to the secondary data structure containing the elements with the same hash value. This is why in practice chaining usually performs worse in terms of memory efficiency and speed. The Cuckoo probing scheme even guarantees $O(1)$ lookups [127] but also with worse practical performance and only up to a certain load factor. Probing strategies that approximate an optimal arrangement at a higher cost for insertion like Robin hood hashing in conjunction with double hashing [118, 128], binary tree hashing [116] and the algorithm presented here seem to have the same average case complexity for lookups, equally asymptotically approaching optimality for high load factors. In practice the mask structure and the cache-friendly localized access pattern of the linear probing allows to effectively lower the constants associated with insertion and deletion.

The fact that the `patchmap` can perform lookups quickly even when it is very full could make it suitable for use cases where memory is constrained and the hash table itself is virtually static. But the quadratically increasing costs for inserting and deleting elements when the `patchmap` fills up limit its utility when insertion and deletion are frequent at the same time with extremely constrained memory. Other than that the performance characteristics of the `patchmap` make it a good choice in cases when memory is only moderately constrained.

10. Reflection and Outlook

Nearly all steps in the data processing were investigated in order to find the largest systematic errors, to correct them and to find minute structural differences between crystals. Only those steps where significant improvements could be made were presented. Ancillary projects were undertaken to enable others, small improvements could be made in multiple aspects, and more substantial improvements demonstrate the utility of Gaussian functions as a basis for a Fourier-friendly diffraction model. Connecting structure refinement with iterative phasing methods and the sorting-based hash table are useful in their own right. The new coefficients are a straight improvement over the Cromer-Mann coefficients, because they are more easily sampled and Fourier-transformed and reduce the relative errors. Despite great efforts to improve merging using a more detailed description of the probability distributions involved and more robust methods, the new methods were mainly beneficial for few observations, but the rate of convergence with more and more data was not improved by much with these methods alone. However, these methods were instrumental when more of the variance in the measurements could be modelled and then reduced. The more accurate and more detailed model of crystal diffraction derived in section 6.9 enables a more efficient data processing that requires half or less experimental data for the same precision in the result.

As the varying intensity overall had already been corrected for with methods inherited from rotational crystallography (scaling), the partial nature of the measurements in diffraction patterns without rotation was the next largest source of variance. A large part of the variance introduced by unknown partialities could be reduced and it is not known how much more is possible due to the difficulties in parameter optimization and the number of parameters required for even more accurate models, that might not be well-constrained by the data any more. The extent to which partiality estimates can improve clustering remains an open question, even though this was the initial impetus of this project – too much was, and maybe still is, in the way, but with these improvements we are at least closer.

Refinement against the whole diffraction patterns, however, improves the constraint ratio considerably and should improve the data processing efficiency once more. Full Bayesian induction with more informative prior information instead of a maximum likelihood approach with minimal prior information should further improve data processing efficiency too. Alternatively, seeing how well the weighted median fared against the maximum likelihood method for merging observations in section 6.6, minimizing the absolute distances, while not being quite as efficient in theory, could turn out better in practice because of the complications involved in function optimization that can often be circumvented when minimizing absolute distances. With enough improvements in this direction, it is expected that it will eventually be possible to resolve structural differences between the individual crystals. This would mean not just the same information from less data, but a different kind of information in addition.

List of Figures

106

List of Tables

References

[1] Wolfgang Brehm, Thomas A. White, Henry N. Chapman et al. 'Crystal Diffraction Prediction and Partiality estimation using Gaussian basis functions'. Submitted to Acta Crystallographica Section A.

[2] Wolfgang Brehm. 'Hash Tables with Pseudorandom Global Order'. In: *INFOCOMP Journal of Computer Science* 18.1 (June 2019), pp. 20–25. URL: https://infocomp.dcc.ufla.br/index.php/infocomp/article/view/581.

[3] R G Parr. 'Density Functional Theory'. In: *Annual Review of Physical Chemistry* 34.1 (1983), pp. 631–656. DOI: 10.1146/annurev.pc.34.100183.003215. eprint: https://doi.org/10.1146/annurev.pc.34.100183.003215. URL: https://doi.org/10.1146/annurev.pc.34.100183.003215.

[4] Wolfgang Brehm and Kay Diederichs. 'Breaking the indexing ambiguity in serial crystallography'. In: *Acta Crystallographica Section D* 70.1 (Jan. 2014), pp. 101–109. DOI: 10.1107/S1399004713025431. URL: https://doi.org/10.1107/S1399004713025431.

[5] Greta M. Assmann, Meitian Wang and Kay Diederichs. 'Making a difference in multi-data-set crystallography: simple and deterministic data-scaling/selection methods'. In: *Acta Crystallographica Section D* 76.7 (July 2020), pp. 636–652. DOI: 10.1107/S2059798320006348. URL: https://doi.org/10.1107/S2059798320006348.

[6] Kay Diederichs. 'Dissecting random and systematic differences between noisy composite data sets'. In: *Acta Crystallographica Section D* 73.4 (Apr. 2017), pp. 286–293. DOI: 10.1107/S2059798317000699. URL: https://doi.org/10.1107/S2059798317000699.

[7] Maurice G. Kendall. *Kendall's Advanced Theory of Statistics, Volume I*. Kendall's Advanced Theory of Statistics. Charles Griffin & Company Ltd., 1945.

[8] A. Gelman et al. *Bayesian Data Analysis*. Third. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis, 2013. ISBN: 9781439840955. URL: http://www.stat.columbia.edu/~gelman/book/.

[9] S. Asmussen. *Applied Probability and Queues*. Stochastic Modelling and Applied Probability. Springer New York, 2008. ISBN: 9780387215259. DOI: 10.1007/0-387-21525-5_10. URL: https://doi.org/10.1007%2F0-387-21525-5_10.

[10] Michael Falk, Jürg Hüsler and Rolf-Dieter Reiss. *Laws of small numbers: extremes and rare events*. Springer Science & Business Media, 2010.

[11] Dimitri Bertsekas and John N Tsitsiklis. *Introduction to probability*. Vol. 1. Athena Scientific, 2008.

[12] Robert V Hogg and Allen T Craig. 'Introduction to mathematical statistics'. In: *Englewood Hills, New Jersey* (2013).

[13] Eun Sul Lee and Ronald N Forthofer. *Analyzing complex survey data*. Sage Publications, 2006.

[14] J. H. C. Lisman and M. C. A. van Zuylen. 'Note on the generation of most probable frequency distributions'. In: *Statistica Neerlandica* 26.1 (Mar. 1972), pp. 19–23. DOI: 10.1111/j.1467-9574.1972.tb00152.x. URL: https://doi.org/10.1111/j.1467-9574.1972.tb00152.x.

[15] S. Kullback and R. A. Leibler. 'On Information and Sufficiency'. In: *The Annals of Mathematical Statistics* 22.1 (Mar. 1951), pp. 79–86. DOI: `10.1214/aoms/1177729694`. URL: `https://doi.org/10.1214/aoms/1177729694`.

[16] Harold Jeffreys. 'An invariant form for the prior probability in estimation problems'. In: *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* 186.1007 (Sept. 1946), pp. 453–461. DOI: `10.1098/rspa.1946.0056`. URL: `https://doi.org/10.1098/rspa.1946.0056`.

[17] Norman L Johnson, Samuel Kotz and Narayanaswamy Balakrishnan. *Continuous univariate distributions, volume 1*. Second. Vol. 289. John wiley & sons, 1995.

[18] Norman L Johnson, Samuel Kotz and Narayanaswamy Balakrishnan. *Continuous univariate distributions, volume 2*. Second. Vol. 289. John wiley & sons, 1995.

[19] Paul Adrien Maurice Dirac. *The principles of quantum mechanics*. Third. Oxford university press, 1958.

[20] Daniel Fink. 'A compendium of conjugate priors'. In: *progress report: Extension and enhancement of methods for setting data quality objectives*. 46 (1997).

[21] J. Rissanen. 'Modeling by shortest data description'. In: *Automatica* 14.5 (1978), pp. 465–471. ISSN: 0005-1098. DOI: `10.1016/0005-1098(78)90005-5`. URL: `https://www.sciencedirect.com/science/article/pii/0005109878900055`.

[22] H. Jeffreys. *Theory of Probability*. International series of monographs on physics. Clarendon Press, 1983. ISBN: 9780198531937.

[23] Gunnar Taraldsen and Bo Henry Lindqvist. 'Improper priors are not improper'. In: *The American Statistician* 64.2 (2010), pp. 154–158.

[24] Jorma Rissanen. 'A Universal Prior for Integers and Estimation by Minimum Description Length'. In: *The Annals of Statistics* 11.2 (1983), pp. 416–431. ISSN: 00905364. URL: `http://www.jstor.org/stable/2240558`.

[25] Maurice G. Kendall. *Kendall's Advanced Theory of Statistics, Volume II*. Kendall's Advanced Theory of Statistics. Charles Griffin & Company Ltd., 1945.

[26] P. Bevington and D.K. Robinson. *Data Reduction and Error Analysis for the Physical Sciences*. McGraw-Hill Education, 2003. ISBN: 9780072472271. URL: `https://books.google.de/books?id=0poQAQAAIAAJ`.

[27] L. Kish. *Survey Sampling*. Wiley, 1965. ISBN: 9780471489009. URL: `https://books.google.de/books?id=xiZmAAAAIAAJ`.

[28] R M Endlich et al. 'Statistical analysis of precipitation chemistry measurements over the eastern United States. Final report'. In: *OSTI* (July 1986). URL: `https://www.osti.gov/biblio/5419742`.

[29] Donald F. Gatz and Luther Smith. 'The standard error of a weighted mean concentration—I. Bootstrapping vs other methods'. In: *Atmospheric Environment* 29.11 (June 1995), pp. 1185–1193. DOI: `10.1016/1352-2310(94)00210-c`. URL: `https://doi.org/10.1016/1352-2310(94)00210-c`.

[30] D. H. D. West. 'Updating Mean and Variance Estimates: An Improved Method'. In: *Commun. ACM* 22.9 (Sept. 1979), pp. 532–535. ISSN: 0001-0782. DOI: `10.1145/359146.359153`. URL: `https://doi.org/10.1145/359146.359153`.

[31] Richard J. Hanson. 'Stably Updating Mean and Standard Deviation of Data'. In: *Commun. ACM* 18.1 (Jan. 1975), pp. 57–58. ISSN: 0001-0782. DOI: `10.1145/360569.360662`. URL: `https://doi.org/10.1145/360569.360662`.

[32] STEVEN J. MILLER. *The Probability Lifesaver.* Princeton University Press, May 2017. DOI: `10.2307/j.ctvc7767n`. URL: `https://doi.org/10.2307/j.ctvc7767n`.

[33] Paul R. Rider. 'Variance of the Median of Small Samples from Several Special Populations'. In: *Journal of the American Statistical Association* 55.289 (1960), pp. 148–150. DOI: `10.1080/01621459.1960.10482056`. eprint: `https://www.tandfonline.com/doi/pdf/10.1080/01621459.1960.10482056`. URL: `https://www.tandfonline.com/doi/abs/10.1080/01621459.1960.10482056`.

[34] Ralph S. Woodruff. 'Confidence Intervals for Medians and Other Position Measures'. In: *Journal of the American Statistical Association* 47.260 (Dec. 1952), pp. 635–646. DOI: `10.1080/01621459.1952.10483443`. URL: `https://doi.org/10.1080/01621459.1952.10483443`.

[35] Raul Rojas. 'The secret life of the covariance matrix'. In: *Freie Universität Berlin [online], URL: http://www. inf. fu-berlin. de/inst/ag-ki/rojas_home/documents/tutorials/secretcovariance. pdf [cited 1 October 2012]* (2009).

[36] G.B. Folland. *Harmonic Analysis in Phase Space.* Annals of Mathematics Studies. Princeton University Press, 1989. ISBN: 9780691085289.

[37] J.A. Fessler and B.P. Sutton. 'Nonuniform fast Fourier transforms using min-max interpolation'. In: *IEEE Transactions on Signal Processing* 51.2 (2003), pp. 560–574. DOI: `10.1109/TSP.2002.807005`.

[38] Michael Heideman, Don Johnson and Charles Burrus. 'Gauss and the history of the fast Fourier transform'. In: *IEEE ASSP Magazine* 1.4 (1984), pp. 14–21.

[39] Claude E Shannon. 'Communication in the presence of noise'. In: *Proceedings of the IRE* 37.1 (1949), pp. 10–21.

[40] Farokh Marvasti. *Nonuniform sampling: theory and practice.* Springer Science & Business Media, 2012.

[41] Ronald N Bracewell. 'Strip integration in radio astronomy'. In: *Australian Journal of Physics* 9.2 (1956), pp. 198–217.

[42] Ren Ng. 'Fourier slice photography'. In: *ACM Siggraph 2005 Papers.* 2005, pp. 735–744.

[43] John David Jackson. *Classical electrodynamics; 3rd ed.* New York, NY: Wiley, 1999.

[44] Kimball Milton and Julian Schwinger. *Electromagnetic radiation: variational methods, waveguides and accelerators.* Springer Science & Business Media, 2006.

[45] Avinash C Kak and Malcolm Slaney. *Principles of computerized tomographic imaging.* SIAM, 2001.

[46] Claude Cohen-Tannoudji, Bernard Diu and Frank Laloe. 'Quantum Mechanics, Volume 1'. In: *Quantum Mechanics* 1 (1986), p. 898.

[47] Paul Müller, Mirjam Schürmann and Jochen Guck. 'The theory of diffraction tomography'. In: *arXiv preprint arXiv:1507.00466* (2015).

[48] U. Shmueli, ed. *International Tables for Crystallography*. International Union of Crystallography, Jan. 2010. DOI: `10.1107/97809553602060000108`. URL: `https://doi.org/10.1107/97809553602060000108`.

[49] IUCr. *International Tables for Crystallography, Volume C: Mathematical, physical and chemical tables*. 3nd edition. International Tables for Crystallography. Dordrecht, Boston, London: Kluwer Academic Publishers, 2004.

[50] Wolfgang Kabsch. 'Processing of X-ray snapshots from crystals in random orientations'. In: *Acta Crystallographica Section D* 70.8 (Aug. 2014), pp. 2204–2216. DOI: `10.1107/S1399004714013534`. URL: `https://doi.org/10.1107/S1399004714013534`.

[51] Kartik Ayyer et al. '3D diffractive imaging of nanoparticle ensembles using an x-ray laser'. In: *Optica* 8.1 (Dec. 2020), p. 15. DOI: `10.1364/optica.410851`. URL: `https://doi.org/10.1364%2Foptica.410851`.

[52] MG Rossmann et al. 'Processing and post-refinement of oscillation camera data'. In: *Journal of Applied Crystallography* 12.6 (1979), pp. 570–581.

[53] A. J. C. Wilson. 'The probability distribution of X-ray intensities'. In: *Acta Crystallographica* 2.5 (Oct. 1949), pp. 318–321. DOI: `10.1107/S0365110X49000813`. URL: `https://doi.org/10.1107/S0365110X49000813`.

[54] Heinz Graafsma, Julian Becker and Sol M. Gruner. 'Integrating Hybrid Area Detectors for Storage Ring and Free-Electron Laser Applications'. In: *Synchrotron Light Sources and Free-Electron Lasers*. Springer International Publishing, 2016, pp. 1029–1054. DOI: `10.1007/978-3-319-14394-1_37`. URL: `https://doi.org/10.1007/978-3-319-14394-1_37`.

[55] Stefan Zaefferer. 'New developments of computer-aided crystallographic analysis in transmission electron microscopy'. In: *Journal of Applied Crystallography* 33.1 (2000), pp. 10–25.

[56] Anton Barty et al. 'Cheetah: software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data'. In: *Journal of applied crystallography* 47.3 (2014), pp. 1118–1131.

[57] Yaroslav Gevorkov. 'Real-time image analysis and data compression in high throughput X-ray diffraction experiments'. MA thesis. 2016. DOI: `10.3204/PUBDB-2022-03562`.

[58] Yaroslav Gevorkov et al. 'XGANDALF – extended gradient descent algorithm for lattice finding'. In: *Acta Crystallographica Section A* 75.5 (Sept. 2019), pp. 694–704. DOI: `10.1107/S2053273319010593`. URL: `https://doi.org/10.1107/S2053273319010593`.

[59] Yaroslav Gevorkov et al. 'pinkIndexer – a universal indexer for pink-beam X-ray and electron diffraction snapshots'. In: *Acta Crystallographica Section A* 76.2 (Mar. 2020), pp. 121–131. DOI: `10.1107/S2053273319015559`. URL: `https://doi.org/10.1107/S2053273319015559`.

[60] K. N. Trueblood et al. 'Atomic Dispacement Parameter Nomenclature. Report of a Subcommittee on Atomic Displacement Parameter Nomenclature'. In: *Acta Crystallographica Section A Foundations of Crystallography* 52.5 (Sept. 1996), pp. 770–781. DOI: `10.1107/s0108767396005697`. URL: `https://doi.org/10.1107/s0108767396005697`.

[61] Michael G Rossmann and Eddy Arnold. *International Tables for Crystallography Volume F: Crystallography of Biological Macromolecules.* Kluwer Academic Publishers, 2006.

[62] Markus Metz et al. 'Macromolecular phasing using diffraction from multiple crystal forms'. In: *Acta Crystallographica Section A* 77.1 (Jan. 2021), pp. 19–35. DOI: 10.1107/S2053273320013650. URL: https://doi.org/10.1107/S2053273320013650.

[63] W Kabsch. 'XDS'. In: *Acta Crystallogr. D* 66 (2010), pp. 125–132.

[64] Aaron S. Brewster et al. 'SAD phasing of XFEL data depends critically on the error model'. In: *Acta Crystallographica Section D* 75.11 (Nov. 2019), pp. 959–968. DOI: 10.1107/S2059798319012877. URL: https://doi.org/10.1107/S2059798319012877.

[65] P Andrew Karplus and Kay Diederichs. 'Linking crystallographic model and data quality'. In: *Science* 336.6084 (2012), pp. 1030–1033.

[66] D. T. Cromer and J. B. Mann. 'X-ray scattering factors computed from numerical Hartree–Fock wave functions'. In: *Acta Crystallographica Section A* 24.2 (Mar. 1968), pp. 321–324. DOI: 10.1107/S0567739468000550. URL: https://doi.org/10.1107/S0567739468000550.

[67] Linus Pauling and E. Bright Wilson. *Introduction to quantum mechanics.* Reprint. Originally published: New York ; London : McGraw-Hill, 1935. New York, N.Y. : Dover Publications, 1963. URL: http://www.loc.gov/catdir/description/dover032/84025919.html.

[68] Ilme Schlichting. 'Serial femtosecond crystallography: the first five years'. In: *IUCrJ* 2.2 (Mar. 2015), pp. 246–255. DOI: 10.1107/S205225251402702X. URL: https://doi.org/10.1107/S205225251402702X.

[69] J. C. H. Spence. 'XFELs for structure and dynamics in biology'. In: *IUCrJ* 4.4 (July 2017), pp. 322–339. DOI: 10.1107/S2052252517005760. URL: https://doi.org/10.1107/S2052252517005760.

[70] Richard A Kirian et al. 'Structure-factor analysis of femtosecond microdiffraction patterns from protein nanocrystals'. In: *Acta Crystallographica Section A: Foundations of Crystallography* 67.2 (2011), pp. 131–140.

[71] A. Meents et al. 'Pink-beam serial crystallography'. In: *Nature Communications* 8.1 (2017), p. 1281. ISSN: 2041-1723. DOI: 10.1038/s41467-017-01417-3.

[72] Robert Bücker et al. 'Serial protein crystallography in an electron microscope'. In: *Nature Communications* 11.1 (2020), p. 996. ISSN: 2041-1723. DOI: 10.1038/s41467-020-14793-0.

[73] R. Diamond. 'Profile analysis in single-crystal diffractometry'. In: *Acta Crystallographica Section A* 25.1 (Jan. 1969), pp. 43–55. DOI: 10.1107/S0567739469000064. URL: https://doi.org/10.1107/S0567739469000064.

[74] D. F. Grant and E. J. Gabe. 'The analysis of single-crystal Bragg reflections from profile measurements'. In: *Journal of Applied Crystallography* 11.2 (Apr. 1978), pp. 114–120. DOI: 10.1107/S0021889878012856. URL: https://doi.org/10.1107/S0021889878012856.

[75] F. K. Winkler, C. E. Schutt and S. C. Harrison. 'The oscillation method for crystals with very large unit cells'. In: *Acta Crystallographica Section A* 35.6 (Nov. 1979), pp. 901–911. DOI: 10.1107/S0567739479002035. URL: https://doi.org/10.1107/S0567739479002035.

[76] T. J. Greenhough and J. R. Helliwell. 'Oscillation camera data processing: reflecting range and prediction of partiality. I. Conventional X-ray sources'. In: *Journal of Applied Crystallography* 15.3 (June 1982), pp. 338–351. DOI: 10.1107/S0021889882012072. URL: https://doi.org/10.1107/S0021889882012072.

[77] T. J. Greenhough and J. R. Helliwell. 'Oscillation camera data processing: reflecting range and prediction of partiality. 2. Monochromatized synchrotron X-radiation from a singly bent triangular monochromator'. In: *Journal of Applied Crystallography* 15.5 (Oct. 1982), pp. 493–508. DOI: 10.1107/S0021889882012485. URL: https://doi.org/10.1107/S0021889882012485.

[78] T. J. Greenhough, J. R. Helliwell and S. A. Rule. 'Oscillation camera data processing. 3. General diffraction spot size, shape and energy profile: formalism for polychromatic diffraction experiments with monochromatized synchrotron X-radiation from a singly bent triangular monochromator'. In: *Journal of Applied Crystallography* 16.2 (Apr. 1983), pp. 242–250. DOI: 10.1107/S0021889883010316. URL: https://doi.org/10.1107/S0021889883010316.

[79] S. J. Andrews et al. 'The mosaic spread of very small crystals deduced from Laue diffraction patterns'. In: *Acta Crystallographica Section A* 43.1 (Jan. 1987), pp. 70–73. DOI: 10.1107/S0108767387099872. URL: https://doi.org/10.1107/S0108767387099872.

[80] Helen Mary Ginn et al. 'A revised partiality model and post-refinement algorithm for X-ray free-electron laser data'. In: *Acta Crystallographica Section D: Biological Crystallography* 71.6 (2015), pp. 1400–1410.

[81] Thomas A. White et al. 'Recent developments in CrystFEL'. In: *Journal of Applied Crystallography* 49.2 (Apr. 2016), pp. 680–689. DOI: 10.1107/S1600576716004751. URL: https://doi.org/10.1107/S1600576716004751.

[82] James M. Holton et al. 'The R-factor gap in macromolecular crystallography: an untapped potential for insights on accurate structures'. In: *The FEBS Journal* 281.18 (Sept. 2014), pp. 4046–4060. DOI: 10.1111/febs.12922. URL: https://doi.org/10.1111/febs.12922.

[83] Trevor J Greenhough and John R. Helliwell. 'The uses of synchrotron X-radiation in the crystallography of molecular biology.' In: *Progress in biophysics and molecular biology* 41 2 (1983), pp. 67–123.

[84] Loes MJ Kroon-Batenburg et al. 'Accounting for partiality in serial crystallography using ray-tracing principles'. In: *Acta Crystallographica Section D: Biological Crystallography* 71.9 (2015), pp. 1799–1811.

[85] Derek Mendez et al. 'Beyond integration: modeling every pixel to obtain better structure factors from stills'. In: *IUCrJ* 7.6 (Nov. 2020), pp. 1151–1167. DOI: 10.1107/S2052252520013007. URL: https://doi.org/10.1107/S2052252520013007.

[86] Ruben A. Dilanian et al. 'Whole-pattern fitting technique in serial femtosecond nanocrystallography'. In: *IUCrJ* 3.2 (Mar. 2016), pp. 127–138. DOI: 10.1107/S2052252516001238. URL: https://doi.org/10.1107/S2052252516001238.

[87] J. M. Cowley. *Diffraction physics*. Elsevier Science B.V., 1995.

[88] Malcolm Slaney and Avinash C Kak. 'Imaging with diffraction tomography'. In: (1985).

[89] H.W. Sorenson and D.L. Alspach. 'Recursive bayesian estimation using gaussian sums'. In: *Automatica* 7.4 (1971), pp. 465–479. ISSN: 0005-1098. DOI: `10.1016/0005-1098(71)90097-5`. URL: `http://www.sciencedirect.com/science/article/pii/0005109871900975`.

[90] Kay Diederichs. 'Quantifying instrument errors in macromolecular X-ray data sets'. In: *Acta Crystallographica Section D* 66.6 (June 2010), pp. 733–740. DOI: `10.1107/S0907444910014836`. URL: `https://doi.org/10.1107/S0907444910014836`.

[91] C. G. Broyden. 'The Convergence of a Class of Double-rank Minimization Algorithms 1. General Considerations'. In: *IMA Journal of Applied Mathematics* 6.1 (Mar. 1970), pp. 76–90. ISSN: 0272-4960. DOI: `10.1093/imamat/6.1.76`. eprint: `https://academic.oup.com/imamat/article-pdf/6/1/76/2233756/6-1-76.pdf`. URL: `https://doi.org/10.1093/imamat/6.1.76`.

[92] R. Fletcher. 'A new approach to variable metric algorithms'. In: *The Computer Journal* 13.3 (Jan. 1970), pp. 317–322. ISSN: 0010-4620. DOI: `10.1093/comjnl/13.3.317`. eprint: `https://academic.oup.com/comjnl/article-pdf/13/3/317/988678/130317.pdf`. URL: `https://doi.org/10.1093/comjnl/13.3.317`.

[93] D Goldfarb. 'Mathematics of Computation'. In: *American Mathematical Society* 24 (1970), p. 23.

[94] J Schanno. 'Conditions of quasi-Newton methods for function minimization'. In: *Math. Comput* 24.4 (1970), pp. 647–650.

[95] Wei Liu et al. 'Serial Femtosecond Crystallography of G Protein–Coupled Receptors'. In: *Science* 342.6165 (2013), pp. 1521–1524. ISSN: 0036-8075. DOI: `10.1126/science.1244142`. eprint: `https://science.sciencemag.org/content/342/6165/1521.full.pdf`. URL: `https://science.sciencemag.org/content/342/6165/1521`.

[96] LCLS. *LCLS CXI Specifications*. Jan. 2022. URL: `https://lcls.slac.stanford.edu/instruments/cxi/specifications` (visited on 24/01/2022).

[97] Yael Pazy et al. 'Ligand Exchange between Proteins'. In: *Journal of Biological Chemistry* 277.34 (Aug. 2002), pp. 30892–30900. DOI: `10.1074/jbc.m202874200`. URL: `https://doi.org/10.1074/jbc.m202874200`.

[98] Karol Nass. *Advances in long-wavelength native phasing at X-ray free-electron lasers*. en. 2020. DOI: `10.11577/1650020`. URL: `https://www.osti.gov/servlets/purl/1650020/`.

[99] 'CrystFEL: a software suite for snapshot serial crystallography'. In: *Journal of Applied Crystallography* 45.2 (Apr. 2012), pp. 335–341. DOI: `10.1107/S0021889812002312`. URL: `https://doi.org/10.1107/S0021889812002312`.

[100] Dorothee Liebschner et al. 'Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix'. In: *Acta Crystallographica Section D* 75.10 (Oct. 2019), pp. 861–877. DOI: `10.1107/S2059798319011471`. URL: `https://doi.org/10.1107/S2059798319011471`.

[101] Sebastian Günther et al. 'X-ray screening identifies active site and allosteric inhibitors of SARS-CoV-2 main protease'. In: *Science* 372.6542 (2021), pp. 642–646. DOI: `10.1126/science.abf7945`. eprint: `https://www.science.org/doi/pdf/10.1126/science.abf7945`. URL: `https://www.science.org/doi/abs/10.1126/science.abf7945`.

[102] Sebastian Günther et al. 'Catalytic cleavage of HEAT and subsequent covalent binding of the tetralone moiety by the SARS-CoV-2 main protease'. In: *bioRxiv* (2020).

[103] Werner Hansen. 'Mannichbasen aus alpha-Tetralon bzw. dessen Derivaten und Arylalkylaminen und deren Salze'. Jan. 1975.

[104] Sebastian Günther et al. 'Massive X-ray screening reveals two allosteric drug binding sites of SARS-CoV-2 main protease'. In: *bioRxiv* (2020). DOI: `10.1101/2020.11.12.378422`. eprint: `https://www.biorxiv.org/content/early/2020/11/12/2020.11.12.378422.full.pdf`. URL: `https://www.biorxiv.org/content/early/2020/11/12/2020.11.12.378422`.

[105] H. T. Henry Chan et al. 'Discovery of SARS-CoV-2 Mpro peptide inhibitors from modelling substrate and ligand binding'. In: *Chem. Sci.* 12 (41 2021), pp. 13686–13703. DOI: `10.1039/D1SC03628A`. URL: `http://dx.doi.org/10.1039/D1SC03628A`.

[106] Federico Gaeta et al. 'Compositions of tolperisone'. US20090253743A1. Oct. 2009.

[107] Andrew J. Morgan et al. '*Ab initio* phasing of the diffraction of crystals with translational disorder'. In: *Acta Crystallographica Section A* 75.1 (Jan. 2019), pp. 25–40. DOI: `10.1107/S2053273318015395`. URL: `https://doi.org/10.1107/S2053273318015395`.

[108] J. R. Fienup. 'Phase retrieval algorithms: a comparison'. In: *Applied Optics* 21.15 (Aug. 1982), p. 2758. DOI: `10.1364/ao.21.002758`. URL: `https://doi.org/10.1364%2Fao.21.002758`.

[109] Edouard Jean Robert Pauwels et al. 'On Fienup methods for sparse phase retrieval'. In: *IEEE Transactions on Signal Processing* 66.4 (2017), pp. 982–991.

[110] Veit Elser. 'Phase retrieval by iterated projections'. In: *J. Opt. Soc. Am. A* 20.1 (Jan. 2003), pp. 40–55. DOI: `10.1364/JOSAA.20.000040`. URL: `https://opg.optica.org/josaa/abstract.cfm?URI=josaa-20-1-40`.

[111] RANDY J Read. 'Improved Fourier coefficients for maps using phases from partial structures with errors'. In: *Acta Crystallographica Section A: Foundations of Crystallography* 42.3 (1986), pp. 140–149.

[112] Noel M O'Boyle et al. 'Open Babel: An open chemical toolbox'. In: *Journal of cheminformatics* 3.1 (2011), pp. 1–14.

[113] S. Redford et al. 'First full dynamic range calibration of the JUNGFRAU photon detector'. In: *Journal of Instrumentation* 13.01 (Jan. 2018), pp. C01027–C01027. DOI: `10.1088/1748-0221/13/01/c01027`. URL: `https://doi.org/10.1088/1748-0221/13/01/c01027`.

[114] Wolfgang Brehm. 'Hash tables with pseudorandom global order'. In: *INFOCOMP Journal of Computer Science* 18.1 (2019), pp. 20–25.

[115]   D. E. Knuth and O. Amble. 'Ordered hash tables'. In: *The Computer Journal* 17.2 (Jan. 1974), pp. 135–142. ISSN: 0010-4620. DOI: `10.1093/comjnl/17.2.135`. eprint: `http://oup.prod.sis.lan/comjnl/article-pdf/17/2/135/1405739/170135.pdf`. URL: `https://doi.org/10.1093/comjnl/17.2.135`.

[116]   Gaston H Gonnet and J Ian Munro. 'Efficient ordering of hash tables'. In: *SIAM Journal on Computing* 8.3 (1979), pp. 463–478.

[117]   MV Ramakrishna. 'An exact probability model for finite hash tables'. In: *Proceedings. Fourth International Conference on Data Engineering*. IEEE. 1988, pp. 362–368.

[118]   Pedro Celis. 'Robin Hood Hashing'. PhD thesis. Waterloo, Ont., Canada, Canada, 1986.

[119]   Malte Skarupke. *I Wrote The Fastest Hashtable*. 2017. URL: `https://probablydance.com/2017/02/26/i-wrote-the-fastest-hashtable/` (visited on 02/04/2019).

[120]   Daniel Lemire. 'Fast Random Integer Generation in an Interval'. In: *CoRR* abs/1805.10941 (2018). arXiv: `1805.10941`. URL: `http://arxiv.org/abs/1805.10941`.

[121]   Donald E. Knuth. *The Art of Computer Programming, Volume 3: (2Nd Ed.) Sorting and Searching*. Redwood City, CA, USA: Addison Wesley Longman Publishing Co., Inc., 1998. ISBN: 0-201-89685-0.

[122]   attractivechaos. *klib*. 2018. URL: `http://attractivechaos.github.io/klib/#About` (visited on 02/04/2019).

[123]   Malte Skarupke. *A new fast hash table in response to Google's new fast hash table*. 2018. URL: `https://probablydance.com/2018/05/28/a-new-fast-hash-table-in-response-to-googles-new-fast-hash-table/` (visited on 02/04/2019).

[124]   Malte Skarupke. *flat_hash_map*. 2018. URL: `https://github.com/skarupke/flat_hash_map` (visited on 02/04/2019).

[125]   Gregory Popovitch. *sparsepp*. 2019. URL: `https://github.com/greg7mdp/sparsepp` (visited on 02/04/2019).

[126]   Gaston H Gonnet and J Ian Munro. 'The analysis of linear probing sort by the use of a new mathematical transform'. In: *Journal of Algorithms* 5.4 (1984), pp. 451–470. ISSN: 0196-6774. DOI: `10.1016/0196-6774(84)90001-4`. URL: `https://www.sciencedirect.com/science/article/pii/0196677484900014`.

[127]   Rasmus Pagh and Flemming Friche Rodler. 'Cuckoo Hashing'. In: *Algorithms — ESA 2001*. Springer Berlin Heidelberg, 2001, pp. 121–133. DOI: `10.1007/3-540-44676-1_10`. URL: `https://doi.org/10.1007/3-540-44676-1_10`.

[128]   Luc Devroye, Pat Morin and Alfredo Viola. *On Worst Case Robin-Hood Hashing*. 2004.

A. Toxicological Information

compound	symbol	hazard statements	precautionary statements
HEAT		H302	P264 P301+P312
tolperisone		H302, H315, H319, H335	P261, P264 P280
4-MMPPO		H302, H312, H315, H319, H332	P261, P271, P280, P302+P352, P305+P351+P338
2-methylenetetralin-1-one		H302, H312, H315, H319, H332	P261, P271, P280, P302+P352, P305+P351+P338

B. Derivatives and derived properties

B.1. Derived properties

B.1.1. Peak shape on the detector

Looking at the predicted intensity as a function of the position $\begin{pmatrix} fs & ss \end{pmatrix}^{\top}$ on the detector, and assuming that the peak intensity falls into a small angular range ($< 10°$) where the covariance matrices can be approximated as locally constant with good accuracy, an approximation of the peak shape on the detector can be derived by factoring out the (approximately) constant terms from the exponential. For straightforward computation and best approximation the direction with the highest intensity $\vec{w}_{\text{out}}^{max}$ should be determined, this can be achieved with any function optimization algorithm. Newtons method is equivalent to iteratively completing the square for the exponential term and, because the target function can be made very nearly quadratic by taking the logarithm, it converges very fast. The detector coordinate system is commonly given by a 2-by-3 transformation matrix D and an offset vector $\vec{o}$. The outgoing wave direction is therefore given by the normed position vector:

$$\vec{w}_{\text{out}} = \left(D \begin{pmatrix} fs \\ ss \end{pmatrix} + \vec{o} \right) \left| D \begin{pmatrix} fs \\ ss \end{pmatrix} + \vec{o} \right|^{-1} \tag{156}$$

The point $\begin{pmatrix} fs_0 & ss_0 \end{pmatrix}^{\top}$ denotes the peak position on the detector, i.e. the position of maximum flux. Using $\begin{pmatrix} fs_0 & ss_0 \end{pmatrix}^{\top}$ the normed directionality vector can be approximated to first order as:

$$\vec{w}_{\text{out}} \approx \vec{w}_{\text{out}}^{\text{max}} + \frac{\partial (\vec{w}_{\text{out}})}{\partial \begin{pmatrix} fs \\ ss \end{pmatrix}} \begin{pmatrix} fs_0 \\ ss_0 \end{pmatrix} \left(\begin{pmatrix} fs \\ ss \end{pmatrix} - \begin{pmatrix} fs_0 \\ ss_0 \end{pmatrix} \right)$$

Equation 128 for the flux on the detector can be expressed as a scaled Gaussian (or a sum thereof), and using the linearized expression for the directionality vector the intensity on the detector can be expressed as:

$$j \begin{pmatrix} fs \\ ss \end{pmatrix} \approx c \exp \left(-\frac{1}{2} \left(k \left(\vec{w}_{\text{out}}^{max} + \frac{\partial (\vec{w}_{\text{out}})}{\partial \begin{pmatrix} fs \\ ss \end{pmatrix}} \begin{pmatrix} fs_0 \\ ss_0 \end{pmatrix} \left(\begin{pmatrix} fs \\ ss \end{pmatrix} - \begin{pmatrix} fs_0 \\ ss_0 \end{pmatrix} \right) \right) - \vec{\mu} \right)^{\top} \mathbf{\Sigma}^{-1} \left([...] \right) \right)$$

$$\tag{157}$$

$$c = \text{proportionality constant}$$

The scaled Gaussian, that only appears to be three-dimensional, can be rearranged to show the two-dimensional form using suitable substitutions:

$$M = \frac{\partial\left(\vec{w}_{\text{out}}\right)}{\partial\begin{pmatrix} fs \\ ss \end{pmatrix}}\begin{pmatrix} fs_0 \\ ss_0 \end{pmatrix}$$

$$\vec{\Delta x} = \begin{pmatrix} fs \\ ss \end{pmatrix} - \begin{pmatrix} fs_0 \\ ss_0 \end{pmatrix}$$

$$e = \left(k\left(\vec{w}_{\text{out}}^{\max} + M\,\vec{\Delta x}\right) - \vec{\mu}\right)^{\top}\Sigma^{-1}\left(k\left(\vec{w}_{\text{out}}^{\max} + M\,\vec{\Delta x}\right) - \vec{\mu}\right)$$

$$\Sigma' = \left(k^2 M^{\top}\Sigma^{-1}M\right)^{-1}$$

$$\Delta\begin{pmatrix} fs_0 \\ ss_0 \end{pmatrix} = k\Sigma'^{-1}M^{\top}\Sigma^{-1}\left(\vec{\mu} - k\vec{w}_{\text{out}}^{max}\right)$$

$$e = \left(\vec{\Delta x} - \Delta\begin{pmatrix} fs_0 \\ ss_0 \end{pmatrix}\right)^{\top}\Sigma'^{-1}\left(\vec{\Delta x} - \Delta\begin{pmatrix} fs_0 \\ ss_0 \end{pmatrix}\right)$$

$$\qquad + \left(\vec{\mu} - k\vec{w}_{\text{out}}^{\max}\right)^{\top}\Sigma^{-1}\left(\vec{\mu} - k\vec{w}_{\text{out}}^{\max}\right) - \Delta\begin{pmatrix} fs_0 \\ ss_0 \end{pmatrix}^{\top}\Sigma'^{-1}\Delta\begin{pmatrix} fs_0 \\ ss_0 \end{pmatrix}$$

$$\Delta\begin{pmatrix} fs_0 \\ ss_0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{if the outgoing wave vector was optimal}$$

$$f\begin{pmatrix} fs \\ ss \end{pmatrix} \approx c\exp\left(-\frac{1}{2}\left(\vec{\Delta x}^{\top}\Sigma'^{-1}\vec{\Delta x} + \left(\vec{\mu} - k\vec{w}_{\text{out}}^{\max}\right)^{\top}\Sigma^{-1}\left(\vec{\mu} - k\vec{w}_{\text{out}}^{\max}\right)\right)\right) \tag{158}$$

The peak on the detector can therefore be approximated by a scaled two-dimensional Gaussian (or several), potentially broadened by the point spread function of the detector. The shape (without broadening) is given by the two-dimensional covariance matrix Σ'.

B.1.2. Asymptotically Optimal Prediction of a Diffraction Image in Areas With Flux Above Threshold

The naive approach to calculating a diffraction image of a snapshot would be to compute the multiplication of the source with the object and the convolution with the Green's function via the FFT. This holds in general in kinetic far-field approximation even for non-crystals. For a crystal the Fourier transform is sparse and this is usually exploited by iterating over the Miller indices. The computational complexity is $O(N_h N_k N_l)$ where N is the number of indices to be considered in each direction. Because the Ewald sphere essentially is two dimensional, we can come up with a solution to compute this in $O(N^2)$ by using a region growing approach. Every reflection that exceeds a threshold has at least one neighbouring Miller index, which has a virtual reflection at most as far as half the inter-Bragg distance that would exceed the threshold. Picture a curve with some width going through a mesh. The curve intersects only some nodes of the mesh, but for every node that it does there is at least one face (or enclosed volume for higher dimensions) that it intersects. The path of the curve can be traced by testing neighbouring faces (or volumes) for intersection

iteratively.

```
1: initialise list todo
   | by finding the closest Miller index
   | to each midpoint of each detector panel
2: while ( there are elements in list todo )
3:   take one (h,k,l) from the list todo
     add it to set done
4:   if ( no point in volume around (h,k,l)
       | can exceed threshold ) goto 2
5:   add all neighboring indices to list todo
6:   if ( flux of (h,k,l) is below threshold ) goto 2
7:   predict the intensity of (h,k,l) on the detector
```

To implement the set operations efficiently and to actually achieve $O(N^2)$ asymptotic complexity, the hash table `patchmap` [114] was used, but most other data structures with amortized constant lookups and insertions would do as well because the limiting step is checking the overlap in step 4.

B.1.3. Maximum flux of virtual reflection in range $(h \pm \frac{1}{2}, k \pm \frac{1}{2}, l \pm \frac{1}{2})$

The maximum flux of any virtual reflection with fractional coordinates closer to a given Miller Index (h, k, l) than any other Miller Index can be conservatively estimated by taking the reflection at (h, k, l), and convolving its location with a width equal to one unit of (h, k, l) in reciprocal space while not changing the normalization of equation 146. This distance corresponds to a covariance matrix equal to half the reciprocal unit cell times half the reciprocal unit cell transposed, note the similarity to the result in equation 137. The approximation is not sensitive to the assumed direction of maximum diffraction intensity $\vec{w}_{\max}$, a rough estimate is sufficient. For compactness the term $\vec{k}_{\text{in}} - k\,\vec{w}_{\max}$ will be combined as $\vec{\mu}_A$:

$$
\max_{h \pm \frac{1}{2}, k \pm \frac{1}{2}, l \pm \frac{1}{2}} |4\pi\boldsymbol{\Sigma}_*|^{-\frac{1}{2}} \left(\phi \left(\vec{k}_{\text{in}} - k\,\vec{w}_{\max}, \vec{\mu}_P, \vec{w}_{\max}^\top \boldsymbol{\Sigma}_\circ \vec{w}_{\max} \right) \right)^2
$$

$$
= \max_{h \pm \frac{1}{2}, k \pm \frac{1}{2}, l \pm \frac{1}{2}} \frac{\left(\exp\left(-\frac{1}{2}(\vec{\mu}_A - \vec{\mu}_P)^\top \vec{w}_{\max}^\top \boldsymbol{\Sigma}_\circ^{-1} \vec{w}_{\max} (\vec{\mu}_A - \vec{\mu}_P) \right) \right)^2}{|4\pi\boldsymbol{\Sigma}_*|^{\frac{1}{2}} |2\pi \vec{w}_{\max}^\top \boldsymbol{\Sigma}_\circ \vec{w}_{\max}|}
$$

$$
\approx \frac{\left(\exp\left(-\frac{1}{2}(\vec{\mu}_A - \vec{\mu}_P)^\top \vec{w}_{\max}^\top \left(\boldsymbol{\Sigma}_\circ + \frac{1}{2^2} R R^\top \right)^{-1} \vec{w}_{\max} (\vec{\mu}_A - \vec{\mu}_P) \right) \right)^2}{|4\pi\boldsymbol{\Sigma}_*|^{\frac{1}{2}} |2\pi \vec{w}_{\max}^\top \boldsymbol{\Sigma}_\circ \vec{w}_{\max}|}
$$

B.1.4. One Frame in a Rotation Series

The integrated intensity in a given outgoing direction $\vec{w}_{\text{out}}$ can be expressed as proportional to a Gaussian
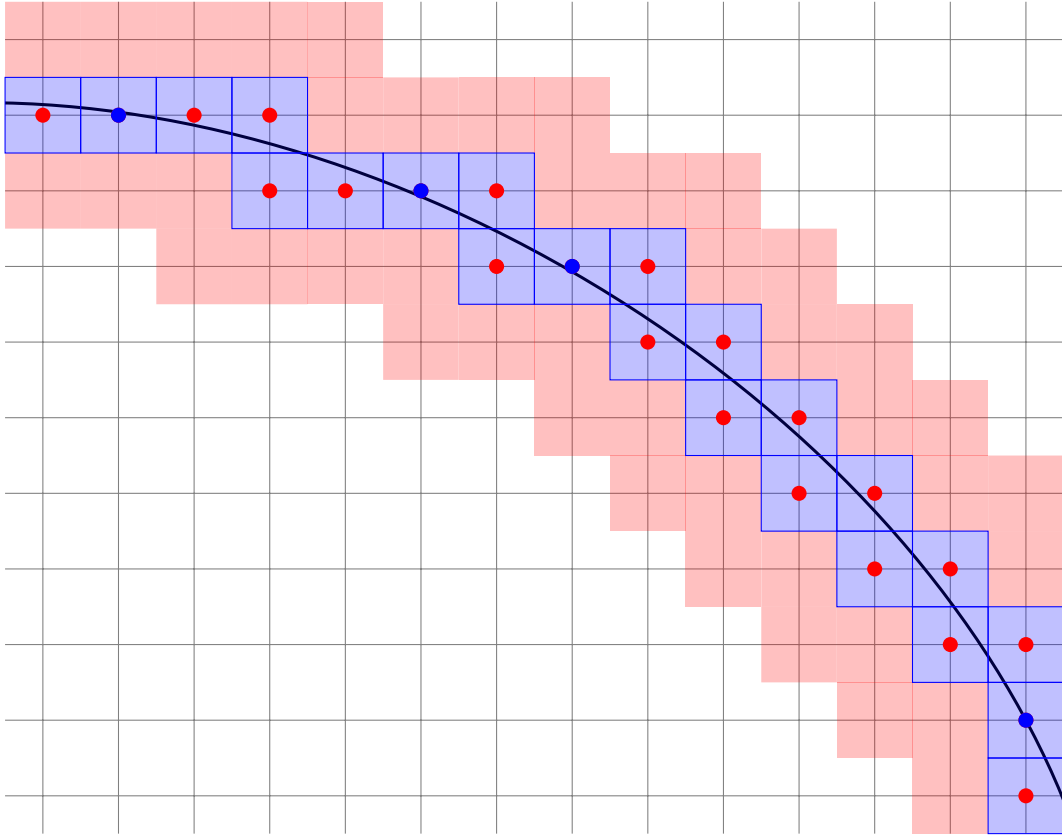
Figure 45: An illustration of region growing for identifying reflections with significant contribution to the diffraction. The gray gridlines intersect at integer combinations that are the Miller indices of the reflections in reciprocal space. The Ewald sphere, or diffraction condition more generally, is assumed to be a smooth function and much thinner in one dimension than the others. It is caricaturized with an ellipse sector in black. The algorithm starts at any of the light red or light blue squares. For each blue square that intersects with the diffraction condition at any point, the diffraction condition at the exact Miller index is evaluated. A significant contribution is indicated with a blue dot, an insignificant contribution with a red dot. For each blue square all new neighbours are inspected for intersections in the same manner. Squares that do not intersect the diffraction condition at any point are colored in light red and don't prompt the inspection of their neighbours.

function, see equation 122. The intensity integrated over an oscillation range $[\beta, \gamma]$ is then proportional to:

$$\int_{\beta}^{\gamma} \phi \left[\vec{\mu}_A, \left((\boldsymbol{G}_\alpha \boldsymbol{U})^{-1} \begin{pmatrix} h \\ k \\ l \end{pmatrix} \right), \boldsymbol{\Sigma} \right] d\alpha \tag{159}$$

$$\boldsymbol{G}_\alpha = \text{rotation matrix with angle } \alpha$$
$$\boldsymbol{U} = \text{unit cell matrix (real-space)}$$

It can be evaluated by first finding the indices values that will be excited to a significant degree in the outgoing arc section described by the position on the detector, the axis of rotation $\vec{g}$ and the oscillation range. Then the target function can be approximated by a one-dimensional Gaussian by developing a small angle approximation around the rotation with the highest predicted intensity and factoring out the constant terms of the three-dimensional Gaussian. This one-dimensional Gaussian integrated for the given range yields a difference of two error-functions.

$$\boldsymbol{G}_{\max} = \text{rotation matrix that yields maximal diffraction}$$

$$\vec{\mu}_P = (\boldsymbol{G}_{\max} \boldsymbol{U})^{-1} \begin{pmatrix} h \\ k \\ l \end{pmatrix}$$

$$\int_{\beta}^{\gamma} \phi \left[\vec{\mu}_A, (\alpha \vec{g} \times \vec{\mu}_P + \vec{\mu}_P), \boldsymbol{\Sigma} \right] d\alpha \tag{160}$$

$$\sigma_g = \left((\vec{g} \times \vec{\mu}_P)^\top \boldsymbol{\Sigma}^{-1} (\vec{g} \times \vec{\mu}_P) \right)^{-\frac{1}{2}} \tag{161}$$

$$\int_{\beta}^{\gamma} \frac{\exp \left(-\frac{1}{2} (\vec{\mu}_S - \vec{\mu}_P)^\top \boldsymbol{\Sigma}^{-1} (\vec{\mu}_S - \vec{\mu}_P) + \frac{\alpha^2}{2\sigma^2} \right)}{|2\pi\boldsymbol{\Sigma}|^{\frac{1}{2}}} d\alpha \tag{162}$$

$$\frac{\exp \left(-\frac{1}{2} (\vec{\mu}_S - \vec{\mu}_P)^\top \boldsymbol{\Sigma} (\vec{\mu}_S - \vec{\mu}_P) \right)}{|2\pi\boldsymbol{\Sigma}|^{\frac{1}{2}}} \int_{\beta}^{\gamma} \exp \left(-\frac{1}{2} \frac{\alpha^2}{\sigma_g^2} \right) d\alpha \tag{163}$$

$$\int_{\beta}^{\gamma} \exp \left(-\frac{1}{2} \frac{\alpha^2}{\sigma_g^2} \right) = \sqrt{\frac{\pi \sigma_g^2}{2}} \left(\text{erf} \left(\frac{\gamma}{\sqrt{2\sigma_g^2}} \right) - \text{erf} \left(\frac{\beta}{\sqrt{2\sigma_g^2}} \right) \right) \tag{164}$$

A similar result is stated with equation 37 in section 3.6 of Kabsch [50].

C. Pixel-wise Backgound Estimation for Pixel-wise Diffraction Pattern Prediction

Background estimation for the pixel-wise diffraction prediction was done by minimizing the following function that acts similar to a boxed median filter or a boxed mean of the middle 75%, which both are much

more easy to compute, but less flexible and slightly less smooth:

$$B(\vec{\mu}, \vec{\sigma}) = \sum^{N} \log((1 - \alpha)\phi(y_i, \mu_i, \sigma_i) + \alpha u(y_i)) + \sum_{\text{adj. } i,j} \log\left(\phi\left(\mu_i, \mu_j, \frac{1}{4}\sqrt{\sigma_i^2 + \sigma_j^2}\right)\right) \qquad (165)$$

α was typically $1/4$, u is the outlier distribution from section 3.4.1 and y_i are the pixel values. The function is minimized by finding the optimal values for μ_i and σ_i. The indices enumerate the pixels and the second sum goes over all pairs of adjacent pixels. This approach is very likely overcomplicated, but it did not turn out to be a bottle-neck and was good enough.

D. Theoretical distribution of partialities

If the intensities of reflections decline like a Gaussian function when leaving the optimal diffraction condition, and because the diffraction condition is essentially random, the distribution of partialities should look like the distribution of function values of a Gaussian distribution with uniform input. The Gaussian function, scaled to a peak height and variance of 1 is $g(X) = \exp\left(-\frac{1}{2}X^2\right)$. Its inverse function, not to be confused with the inverse Gaussian distribution, is $g^{-1}(X) = \pm\sqrt{-2\log(X)}$. There is an ambiguity because $g(X)$ is not strictly increasing or decreasing, but it is symmetric around the axis $X = 0$, and we can therefore restrict the analysis to the increasing branch only. The random variable X is assumed to be uniformly distributed on some region symmetric to 0 on a support $[-a, a]$. The probability density therefore is $f(x) = \frac{1}{2a}$ and the cumulative distribution function $F(x) = \int f(x) = \frac{x}{2a}$. The cumulative distribution function of the random variable $Y = g(X)$ is the distribution function of X applied to the inverse function of g: $P(g(X) < y) = P(X \leq g^{-1}(y)) = F(g^{-1}(y))$, which is $\dfrac{-\sqrt{-2\log(y)}}{2a}$. The density function is its derivative, $\dfrac{\left(-2y^2\log(y)\right)^{-\frac{1}{2}}}{2a}$. In the limit of a large interval $[-a, a]$ this is not a proper density function any more, as the integral $\int_0^1 \left(-2y^2\log(y)\right)^{-\frac{1}{2}} dy$ is divergent.

E. Pseudocode for the Implementation of the `patchmap`

```
insert(key,value) :
1. find the free position p
   closest to map(hash(key))
2. mark p as set
3. insert the key-value pair at p
4. while there is a key to the right
    that is less
  4.1 swap the keys
  4.2 swap the values
  4.3 p ← p-1
5. while there is a key to the left
    to which the key is less
  5.1 swap the keys
  5.2 swap the values
  5.3 p ← p+1
```

```
erase(key) :
1. p ← find(key)
2. while there is a key right of p
    where map(hash(key))>position
  2.1 move the key-value pair left
  2.2 advance p to the right
3. while there is a key left of p
    where map(hash(key))<position
  3.1 move the key-value pair right
  3.2 advance p to the left
4. mark p as free
```

```
find_binary(key,lower_limit,upper_limit) :
1. midpoint ← lower_limit+(upper_limit-lower_limit+1)/2
2. if midpoint is set and (key at midpoint = key)
   then return midpoint
3. if (lower_limit = upper_limit)
   then return not_found
4. if key is less than midpoint
   then upper_limit ← midpoint
5. if midpoint
      is less than
      key to be found
   then lower_limit ← midpoint
6. goto 1



find_interpol(key) :
 1. lower_limit ← 0
 2. upper_limit ← m-1
 3. midpoint ← map(hash(key))
 4. i := 0
 5. if midpoint is set
      if (key at midpoint = key)
      then return midpoint
 6. if (lower_limit = upper_limit)
    then return not_found
 7. if ( key to be found ) is less than ( midpoint )
    then upper_limit := midpoint
 8. if ( midpoint ) is less than ( key to be found )
    then lower_limit := midpoint
 9. i ← i+1;
10. if ( i > 2 log2(n+1) )
    then return find_binary(key,lower_limit,upper_limit)
11. interpolate midpoint linearly
12. goto 5
```

F. Information Content in a Crystallography Experiment

As generally understood and stated in [doi:10.1107/S0567739478001114] the intensities of crystallographic data of an macromolecule follow this distribution:

$$q(I) = \begin{cases} \frac{1}{a} \exp\left(-\frac{I}{a}\right) & \text{for acentric reflections} \\ \frac{1}{\sqrt{2\,\pi\,a\,I}} \exp\left(-\frac{I}{2\,a}\right) & \text{for centric reflections} \end{cases} \tag{166}$$

With a being the expected value of the intensity of a observation of the respective resolution.
This shall be our understanding of the prior knowledge of the intensities and the information we want to compare our measurement with.

F.1. Kullback Leibler divergence of one acentric reflection

The measured probability of the Intensity (m(I)) is assumed to be normally distributed with a σ about some mean μ.

$$m(I) = \frac{1}{\sqrt{2\,\pi\,\sigma^2}} \exp\left(-\frac{(I-\mu)^2}{2\,\sigma^2}\right) \tag{167}$$

Using the prior given in equation 166 and the measurement given in equation 167 the best posterior distribution to assume is in a Bayesian fashion:

$$p(I) \propto p(I)\,p(I) \tag{168}$$

$$p(I) \propto \exp\left(-\frac{\left(I + \frac{\sigma^2}{a} - \mu\right)^2}{2\,\sigma^2}\right) \tag{169}$$

$$\tag{170}$$

The normalization being:

$$N(\mu, \sigma, a) = \sqrt{\frac{\pi}{2}}\,s\,\operatorname{erfc}\left(\frac{\sigma^2 - a\,\mu}{\sqrt{2}\,a\,s}\right) \tag{171}$$

Therefore:

$$p(I) = \frac{\exp\left(-\frac{\left(I + \frac{\sigma^2}{a} - \mu\right)^2}{2\,\sigma^2}\right)}{N(\mu, \sigma, a)} \tag{172}$$

The Kullback Leibler divergence is defined as:

$$x \in \mathcal{X}$$
$$\int_{\mathcal{X}} p(x)\,\ln\left(\frac{p(x)}{q(x)}\right) dx \tag{173}$$

With $p(I)$ given in equation 172 and $q(I)$ given in equation 166 this leads to:

$$\int_0^\infty \frac{\exp\left(-\frac{\left(I + \frac{\sigma^2}{a} - \mu\right)^2}{2\,\sigma^2}\right)}{N(\mu, \sigma, a)} \ln\left(\frac{a\,\exp\left(\frac{I}{a} - \frac{\left(I + \frac{\sigma^2}{a} - \mu\right)^2}{2\,\sigma^2}\right)}{N(\mu, \sigma, a)}\right) dI$$

$$\ln\left(\frac{a}{\mathrm{N}(\mu,\sigma,a)}\right) + \int_0^\infty \frac{\exp\left(-\frac{\left(I+\frac{\sigma^2}{a}-\mu\right)^2}{2\,\sigma^2}\right)}{\mathrm{N}(\mu,\sigma,a)}\left(\frac{I}{a} - \frac{\left(I+\frac{\sigma^2}{a}-\mu\right)^2}{2\,\sigma^2}\right)dI$$

$$\ln\left(\frac{a}{\mathrm{N}(\mu,\sigma,a)}\right) + \int_0^\infty \frac{\exp\left(-\frac{\left(I+\frac{\sigma^2}{a}-\mu\right)^2}{2\,\sigma^2}\right)}{\mathrm{N}(\mu,\sigma,a)}\left(\frac{\mu}{a} - \frac{\sigma^2}{2\,a^2} - \frac{(I-\mu)^2}{2\,\sigma^2}\right)dI$$

$$\ln\left(\frac{a}{\mathrm{N}(\mu,\sigma,a)}\right) + \frac{\mu}{a} - \frac{\sigma^2}{2\,a^2} - \int_0^\infty \frac{\exp\left(-\frac{\left(I+\frac{\sigma^2}{a}-\mu\right)^2}{2\,\sigma^2}\right)}{\mathrm{N}(\mu,\sigma,a)}\left(\frac{(I-\mu)^2}{2\,\sigma^2}\right)dI$$

$$\ln\left(\frac{a}{\mathrm{N}(\mu,\sigma,a)}\right) + \frac{\mu}{a} - \frac{\sigma^2}{2\,a^2} - \frac{1}{\mathrm{N}(\mu,\sigma,a)}\int_0^\infty \exp\left(-\frac{\left(I+\frac{\sigma^2}{a}-\mu\right)^2}{2\,\sigma^2}\right)\left(\frac{(x-\mu)^2}{2\,\sigma^2}\right)dI$$

$$\ln\left(\frac{a}{\mathrm{N}(\mu,\sigma,a)}\right) + \frac{\mu}{a} - \frac{\sigma^2}{2\,a^2} - \frac{1}{\mathrm{N}(\mu,\sigma,a)}\int_{-\mu}^\infty \exp\left(-\frac{\left(I+\frac{\sigma^2}{a}\right)^2}{2\,\sigma^2}\right)\left(\frac{x^2}{2\,\sigma^2}\right)dI$$

$$\int_{-\mu}^\infty \frac{x^2}{2\,\sigma^2}\exp\left(-\frac{\left(x+\frac{\sigma^2}{a}\right)^2}{2\,\sigma^2}\right)dx = \left[\sqrt{2\pi}\left(\frac{\sigma}{4}+\frac{\sigma^3}{4\,a^2}\right)\mathrm{erf}\left(\frac{a\,x+\sigma^2}{a\,\sigma\,\sqrt{2}}\right) + \exp\left(-\frac{\left(x+\frac{\sigma^2}{a}\right)^2}{2\,\sigma^2}\right)\left(\frac{\sigma^2}{2\,a}-\frac{x}{2}\right)\right]_{-\mu}^\infty$$

$$\int_{-\mu}^\infty \frac{x^2}{2\,\sigma^2}\exp\left(-\frac{\left(x+\frac{\sigma^2}{a}\right)^2}{2\,\sigma^2}\right)dx = \sqrt{2\pi}\left(\frac{\sigma}{4}+\frac{\sigma^3}{4\,a^2}\right)\mathrm{erfc}\left(\frac{\sigma^2-a\,\mu}{a\,\sigma\,\sqrt{2}}\right) - \exp\left(-\frac{\left(\mu-\frac{\sigma^2}{a}\right)^2}{2\,\sigma^2}\right)\left(\frac{\sigma^2}{2\,a}+\frac{\mu}{2}\right)$$

$$\ln\left(\frac{a}{\mathrm{N}(\mu,\sigma,a)}\right) + \frac{\mu}{a} - \frac{\sigma^2}{2\,a^2} - \left[\frac{\exp\left(-\frac{\sigma^2}{2\,a^2}+\frac{\mu-I}{a}-\frac{(\mu-I)^2}{2\,\sigma^2}\right)\left(\frac{\sigma^2}{a}+\mu-I\right)}{\sqrt{2\pi}\,\mathrm{erfc}\left(\frac{\sigma^2-a\,\mu}{\sqrt{2}\,a\,\sigma}\right)} - \frac{\left(\frac{\sigma^3}{a^2}+\sigma\right)\mathrm{erf}\left(\frac{a\,(\mu-I)-\sigma^2}{\sqrt{2}\,a\,\sigma}\right)}{2\,\mathrm{erfc}\left(\frac{\sigma^2-a\,\mu}{\sqrt{2}\,a\,\sigma}\right)}\right]_0^\infty \tag{174}$$

$$\ln\left(\frac{a}{\mathrm{N}(\mu,\sigma,a)}\right) + \frac{\mu}{a} - \frac{\sigma^2}{2\,a^2} - \frac{\sqrt{2\pi}\left(\frac{\sigma}{4}+\frac{\sigma^3}{4\,a^2}\right)\mathrm{erfc}\left(\frac{\sigma^2-a\,\mu}{a\,\sigma\,\sqrt{2}}\right) - \exp\left(-\frac{\left(\mu-\frac{\sigma^2}{a}\right)^2}{2\,\sigma^2}\right)\left(\frac{\sigma^2}{2\,a}+\frac{m}{2}\right)}{\mathrm{N}(\mu,\sigma,a)} \tag{175}$$

$$\ln\left(\frac{a}{\mathrm{N}(\mu,\sigma,a)}\right) + \frac{\mu}{a} - \frac{\sigma^2}{2\,a^2} - \frac{2\left(\frac{\sigma}{4}+\frac{\sigma^3}{4\,a^2}\right)\mathrm{erfc}\left(\frac{\sigma^2-a\,\mu}{a\,\sigma\,\sqrt{2}}\right) - \exp\left(-\frac{\left(\mu-\frac{\sigma^2}{a}\right)^2}{2\,\sigma^2}\right)\left(\frac{\sigma^2}{a\,\sqrt{2\pi}}+\frac{m}{\sqrt{2\pi}}\right)}{s\,\mathrm{erfc}\left(\frac{\sigma^2-a\,\mu}{\sqrt{2}\,a\,s}\right)} \tag{176}$$

$$\ln\left(\frac{a}{\mathrm{N}(\mu,\sigma,a)}\right) + \frac{\mu}{a} - \frac{\sigma^2}{a^2} - \frac{1}{2} + \frac{\exp\left(-\frac{\left(\mu-\frac{\sigma^2}{a}\right)^2}{2\,\sigma^2}\right)\left(\frac{\sigma^2}{a\,\sqrt{2\,\pi}}+\frac{m}{\sqrt{2\,\pi}}\right)}{\sigma\,\mathrm{erfc}\left(\frac{\sigma^2-a\,\mu}{\sqrt{2}\,a\,\sigma}\right)} \tag{177}$$

$$\ln\left(\frac{a\,\sqrt{2}}{\sigma\,\sqrt{\pi}\,\mathrm{erfc}\left(\frac{\sigma^2-a\,\mu}{\sqrt{2}\,a\,\sigma}\right)}\right) + \frac{\mu}{a} - \frac{\sigma^2}{a^2} - \frac{1}{2} + \frac{\exp\left(-\frac{\left(\mu-\frac{\sigma^2}{a}\right)^2}{2\,\sigma^2}\right)\left(\frac{\sigma^2}{a}+m\right)}{\sigma\,\sqrt{2\,\pi}\,\mathrm{erfc}\left(\frac{\sigma^2-a\,\mu}{\sqrt{2}\,a\,\sigma}\right)} \tag{178}$$

G. Kullback Leibler divergence of one centric reflection

The measured probability of the Intensity ($\mathrm{m}(I)$) is assumed to be normally distributed with a σ about some mean μ.

$$\mathrm{m}(I) = \frac{1}{\sqrt{2\,\pi}\,\sigma}\,\exp\left(-\frac{(I-\mu)^2}{2\,\sigma^2}\right) \tag{179}$$

Using the prior given in equation 166 and the measurement given in equation 179 the best posterior distribution to assume is in a Bayesian fashion:

$$\mathrm{p}(I) \propto \mathrm{m}(I)\,\mathrm{q}(I) \tag{180}$$

$$\mathrm{p}(I) \propto \frac{\exp\left(-\frac{(I-\mu)^2}{2\,\sigma^2}-\frac{I}{2\,a}\right)}{\sqrt{I}} \tag{181}$$

$$\mathrm{p}(I) \propto \frac{\exp\left(-\frac{\left(I+\frac{\sigma^2}{a}-\mu\right)^2-\left(\frac{\sigma^2}{a}-\mu\right)^2+\mu^2}{2\,\sigma^2}\right)}{\sqrt{I}} \tag{182}$$

$$\mathrm{p}(I) \propto I^{-\frac{1}{2}}\,\exp\left(-\frac{\left(I+\frac{\sigma^2}{a}-\mu\right)^2}{2\,\sigma^2}\right) \tag{183}$$

$$\int_0^\infty \frac{\exp\left(-a\,x^2-b\,x\right)}{\sqrt{x}}\,dx = (2\,a)^{-1/4}\,\Gamma\left(\frac{1}{2}\right)\exp\left(\frac{b^2}{8\,a}\right)\mathrm{D}_{-\frac{1}{2}}\left[b\,(2\,a)^{-\frac{1}{2}}\right] \tag{184}$$

$$\int_0^\infty \frac{\exp\left(-\frac{(x-m)^2}{s}\right)}{\sqrt{x}}\,dx = \begin{cases} \sqrt{-\frac{m}{2}}\,\exp\left(-\left(\frac{m}{2\,s}\right)^2\right)\mathrm{K}_{\frac{1}{4}}\left(-\left(\frac{m}{2\,\sigma}\right)^2\right) & \text{for m<0} \\[2mm] \sqrt[4]{\frac{s^4}{8}}\,\Gamma\left(\frac{1}{4}\right) & \text{for m=0} \\[2mm] \frac{\pi}{2}\,\sqrt{m}\,\exp\left(-\left(\frac{m}{2\,s}\right)^2\right)\left(\mathrm{I}_{\frac{1}{4}}\left(\left(\frac{m}{2\,s}\right)^2\right)+\mathrm{I}_{-\frac{1}{4}}\left(\left(\frac{m}{2\,s}\right)^2\right)\right) & \text{for m<0} \end{cases} \tag{185}$$

I being the Modified Bessel function of 1st kind.

Substitute $m = \frac{\sigma^2}{a} - \mu$ and $s = 2\,\sigma^2$. Therefore the normalization is:

$$\mathrm{N}(\mu,\sigma,a) = \frac{\pi}{2}\,\sqrt{\left|\mu-\frac{\sigma^2}{a}\right|}\,\exp\left(-\frac{\left(\frac{\sigma^2}{a}-\mu\right)^2}{4\,\sigma^2}\right)\left(\mathrm{sign}\left(\mu-\frac{\sigma^2}{a}\right)\mathrm{I}_{\frac{1}{4}}\left(\frac{\left(\frac{\sigma^2}{a}-\mu\right)^2}{4\,\sigma^2}\right)+\mathrm{I}_{-\frac{1}{4}}\left(\frac{\left(\frac{\sigma^2}{a}-\mu\right)^2}{4\,\sigma^2}\right)\right) \tag{186}$$

$$\mathrm{p}(I) = \frac{I^{-\frac{1}{2}} \, \exp\left(-\dfrac{\left(I + \frac{\sigma^2}{a} - \mu\right)^2}{2\,\sigma^2}\right)}{\mathrm{N}(\mu,\sigma,a)} \tag{187}$$

Using the Kullback Leibler divergence as defined in equation 173 and the prior probability (equation 166) this leads to:

$$\int_0^\infty \frac{I^{-\frac{1}{2}} \, \exp\left(-\dfrac{\left(I + \frac{\sigma^2}{a} - \mu\right)^2}{2\,\sigma^2}\right)}{\mathrm{N}(\mu,\sigma,a)} \, \ln\left(\frac{I^{-\frac{1}{2}} \, \exp\left(-\dfrac{\left(I + \frac{\sigma^2}{a} - \mu\right)^2}{2\,\sigma^2}\right)}{\mathrm{N}(\mu,\sigma,a) \, \dfrac{1}{\sqrt{2\,\pi\,a\,I}} \, \exp\left(-\dfrac{I}{2\,a}\right)}\right) dI \tag{188}$$

$$\int_0^\infty \frac{I^{-\frac{1}{2}} \, \exp\left(-\dfrac{\left(I + \frac{\sigma^2}{a} - \mu\right)^2}{2\,\sigma^2}\right)}{\mathrm{N}(\mu,\sigma,a)} \, \ln\left(\frac{\sqrt{2\,\pi\,a} \, \exp\left(\dfrac{I}{2\,a} - \dfrac{\left(I + \frac{\sigma^2}{a} - \mu\right)^2}{2\,\sigma^2}\right)}{\mathrm{N}(\mu,\sigma,a)}\right) dI \tag{189}$$

$$\int_0^\infty \frac{I^{-\frac{1}{2}} \, \exp\left(-\dfrac{\left(I + \frac{\sigma^2}{a} - \mu\right)^2}{2\,\sigma^2}\right)}{\mathrm{N}(\mu,\sigma,a)} \, \left(\ln\left(\frac{\sqrt{2\,\pi\,a}}{\mathrm{N}(\mu,\sigma,a)}\right) + \left(\dfrac{I}{2\,a} - \dfrac{\left(I + \frac{\sigma^2}{a} - \mu\right)^2}{2\,\sigma^2}\right)\right) dI \tag{190}$$

$$\ln\left(\frac{\sqrt{2\,\pi\,a}}{\mathrm{N}(\mu,\sigma,a)}\right) + \int_0^\infty \frac{I^{-\frac{1}{2}} \, \exp\left(-\dfrac{\left(I + \frac{\sigma^2}{a} - \mu\right)^2}{2\,\sigma^2}\right)}{\mathrm{N}(\mu,\sigma,a)} \, \left(\dfrac{I}{2\,a} - \dfrac{\left(I + \frac{\sigma^2}{a} - \mu\right)^2}{2\,\sigma^2}\right) dI \tag{191}$$

$$\ln\left(\frac{\sqrt{2\,\pi\,a}}{\mathrm{N}(\mu,\sigma,a)}\right) + \frac{1}{\mathrm{N}(\mu,\sigma,a)} \int_0^\infty I^{-\frac{1}{2}} \, \exp\left(-\dfrac{\left(I + \frac{\sigma^2}{a} - \mu\right)^2}{2\,\sigma^2}\right) \left(\dfrac{I}{2\,a} - \dfrac{\left(I + \frac{\sigma^2}{a} - \mu\right)^2}{2\,\sigma^2}\right) dI \tag{192}$$

$$\ln\left(\frac{\sqrt{2\,\pi\,a}}{\mathrm{N}(\mu,\sigma,a)}\right) + \frac{1}{\mathrm{N}(\mu,\sigma,a)} \int_0^\infty I^{-\frac{1}{2}} \, \exp\left(-\dfrac{\left(I + \frac{\sigma^2}{a} - \mu\right)^2}{2\,\sigma^2}\right) \left(-\dfrac{\left(I + \frac{\sigma^2}{2\,a} - \mu\right)^2}{2\,\sigma^2} + \dfrac{\mu}{2\,a} - \dfrac{3\,\sigma^2}{8\,a^2}\right) dI \tag{193}$$

$$\ln\left(\frac{\sqrt{2\,\pi\,a}}{\mathrm{N}(\mu,\sigma,a)}\right) + \frac{\mu}{2\,a} - \frac{3\,\sigma^2}{8\,a^2} + \frac{1}{\mathrm{N}(\mu,\sigma,a)} \int_0^\infty I^{-\frac{1}{2}} \, \exp\left(-\dfrac{\left(I + \frac{\sigma^2}{a} - \mu\right)^2}{2\,\sigma^2}\right) \left(-\dfrac{\left(I + \frac{\sigma^2}{2\,a} - \mu\right)^2}{2\,\sigma^2}\right) dI \tag{194}$$

$$\ln\left(\frac{\sqrt{2\,\pi\,a}}{\mathrm{N}(\mu,\sigma,a)}\right) + \frac{\mu}{2\,a} - \frac{3\,\sigma^2}{8\,a^2} - \frac{1}{\mathrm{N}(\mu,\sigma,a)} \int_0^\infty I^{-\frac{1}{2}} \, \exp\left(-\dfrac{\left(I + \frac{\sigma^2}{a} - \mu\right)^2}{2\,\sigma^2}\right) \left(\dfrac{\left(I + \frac{\sigma^2}{2\,a} - \mu\right)^2}{2\,\sigma^2}\right) dI \tag{195}$$

130

$$\text{let} \quad A = \frac{m}{2\,s}$$

$$\text{let} \quad P = m\left(2\,(a-m)^2 + 3\,s^2\right)$$

$$\text{let} \quad Q = 2\,m\,(a-m)^2 + (5\,m - 4\,a)\,\sigma^2$$

$$\int_0^\infty \sqrt{x}\,\frac{(x-a)^2}{2\,s^2}\,\exp\left(-\frac{(x-m)^2}{2\,s^2}\right) dx = \begin{cases} \dfrac{\exp\left(-A^2\right)}{8\,s^2}\sqrt{-\dfrac{m}{2}}\left(Q\,\mathrm{K}_{\frac{1}{4}}\left(A^2\right) - P\,\mathrm{K}_{\frac{3}{4}}\left(A^2\right)\right) & \text{for} \quad m < 0 \\[2ex] \dfrac{\left(2\,a^2 + 3\,s^2\right)\Gamma\left(\dfrac{3}{4}\right) - a\,|s|\,\sqrt{2}\,\Gamma\dfrac{1}{4}}{4\,\sqrt[4]{2}\,\sqrt{|s|}} & \text{for} \quad m = 0 \\[2ex] \dfrac{\pi\,\sqrt{m}\,\exp\left(-A^2\right)}{16\,s^2}\left(P\left(\mathrm{I}_{\frac{3}{4}}\left(A^2\right) + \mathrm{I}_{-\frac{3}{4}}\left(A^2\right)\right) + Q\left(\mathrm{I}_{\frac{1}{4}}\left(A^2\right) + \mathrm{I}_{-\frac{1}{4}}\left(A^2\right)\right)\right) & \text{for} \quad m > 0 \end{cases}$$

$$(196)$$

$$\int_0^\infty I^{-\frac{1}{2}}\,\frac{\left(I + \dfrac{\sigma^2}{2\,a} - \mu\right)}{2\,\sigma^2\,\mathrm{N}(\mu,\sigma,a)}\,\exp\left(-\left(\left(\frac{I + \dfrac{\sigma^2}{a} - \mu}{2\,\sigma}\right)\right)^2\right) dx =$$

for $\quad \mu - \dfrac{\sigma^2}{a} < 0 \quad :$

$$\left(\mu - \frac{\sigma^2}{a}\right)\left(\frac{\sigma}{4\,a}\right)^2 + \frac{\mu}{8} - \frac{3\,\sigma^2}{8\,a} - \frac{\left(\mu - \dfrac{\sigma^2}{a}\right)\left(\left(\dfrac{\sigma}{4\,a}\right)^2 + \dfrac{3}{8}\right)\mathrm{K}_{\frac{3}{4}}\left(\left(\dfrac{\mu - \dfrac{\sigma^2}{a}}{2\,\sigma}\right)^2\right)}{\mathrm{K}_{\frac{1}{4}}\left(\left(\dfrac{\mu - \dfrac{\sigma^2}{a}}{2\,\sigma}\right)^2\right)}$$

for $\quad \mu - \dfrac{\sigma^2}{a} = 0 \quad :$

$$\frac{\left(2\left(\mu - \dfrac{\sigma^2}{2\,a}\right)^2 + 3\,\sigma^2\right)\Gamma\left(\dfrac{3}{4}\right)}{2^{\frac{3}{2}}\,\sigma\,\Gamma\left(\dfrac{1}{4}\right)} - \frac{a}{2}$$

for $\quad \mu - \dfrac{\sigma^2}{a} > 0 \quad :$

$$\frac{P\left(\mathrm{I}_{\frac{3}{4}}\left(\dfrac{\left(\dfrac{\sigma^2}{a} - \mu\right)^2}{4\,\sigma^2}\right) + \mathrm{I}_{-\frac{3}{4}}\left(\dfrac{\left(\dfrac{\sigma^2}{a} - \mu\right)^2}{4\,\sigma^2}\right)\right)}{8\,\sigma^2\left(\mathrm{I}_{\frac{1}{4}}\left(\dfrac{\left(\dfrac{\sigma^2}{a} - \mu\right)^2}{4\,\sigma^2}\right) + \mathrm{I}_{-\frac{1}{4}}\left(\dfrac{\left(\dfrac{\sigma^2}{a} - \mu\right)^2}{4\,\sigma^2}\right)\right)} + \frac{Q}{8\,\sigma^2}$$

H. Diffraction coefficients

Element	a_1	b_1	a_2	b_2	a_3	b_3	a_4	b_4
H	0.469 638	2.796 147	0.331 895	4.488 439	0.175 999	1.655 800	0.022 468	0.805 602
He	1.009 644	2.554 832	0.810 983	1.349 816	0.165 539	0.704 624	0.013 835	0.320 445
H$^-$	0.927 835	7.995 040	0.818 077	3.427 473	0.235 237	1.712 827	0.018 851	0.745 421
Li$^+$	0.855 993	1.422 468	0.748 054	0.864 078	0.198 515	2.198 062	0.197 438	0.439 046
Be^{2+}	0.813 819	0.672 912	0.810 919	1.088 263	0.251 140	0.350 013	0.124 121	1.683 482
Li	1.355 713	1.270 624	1.113 206	7.275 000	0.492 409	0.627 513	0.038 672	0.265 534
Be	2.136 119	5.119 039	1.309 668	0.905 829	0.506 497	0.469 440	0.047 716	0.207 513
B	3.102 097	4.006 238	1.181 633	0.760 291	0.637 362	0.410 679	0.078 908	0.192 048
C	4.011 164	3.301 365	0.945 650	0.400 276	0.904 577	0.749 382	0.138 608	0.183 268
C$_{val}$	2.368 571	2.705 016	1.692 973	4.597 300	1.383 146	0.421 487	0.555 310	0.864 189
N	3.249 948	3.336 847	2.037 919	1.855 856	1.433 762	0.432 659	0.278 371	0.189 957
O	3.352 978	3.029 993	2.909 810	1.699 105	1.397 255	0.388 624	0.339 956	0.177 663
F	3.667 382	1.532 356	3.574 023	2.755 160	1.348 116	0.355 950	0.410 479	0.168 144
O$^-$	3.658 425	2.748 550	2.303 338	1.668 087	1.765 876	0.358 162	1.272 361	5.089 974
Ne	4.164 396	2.406 279	4.044 789	1.387 218	1.339 059	0.330 336	0.451 755	0.156 218
F$^-$	4.009 196	2.458 145	2.886 534	1.486 208	1.776 285	0.321 628	1.327 985	4.488 852
Na$^+$	4.463 336	1.596 495	2.285 662	1.053 944	1.715 259	0.252 221	1.535 744	2.485 176
Mg^{2+}	4.378 824	1.419 565	2.564 082	0.960 053	1.712 221	0.231 010	1.344 874	2.120 374
Al^{3+}	4.458 561	0.989 396	3.302 221	1.476 155	1.849 306	0.240 861	0.389 911	2.105 115
Si^{4+}	4.310 907	0.890 421	3.253 868	1.279 333	1.838 968	0.222 654	0.596 256	1.750 282
Na	7.100 272	1.433 276	1.990 219	4.975 171	1.211 960	0.344 388	0.697 549	0.162 500
Mg	7.285 266	1.267 596	2.789 536	5.242 925	1.110 270	0.332 266	0.814 928	0.157 112
Al	7.353 636	1.129 451	3.707 178	5.012 139	1.004 390	0.324 313	0.934 797	0.152 374
Si	7.346 988	1.016 053	4.688 740	4.560 015	1.088 618	0.149 456	0.875 654	0.328 226
Si$_{val}$	7.032 216	1.071 436	2.397 308	0.311 346	2.303 254	3.959 509	2.267 222	6.208 755
P	7.332 640	0.917 959	5.715 775	4.090 467	1.117 384	0.141 418	0.834 201	0.307 000
S	7.273 003	0.839 896	6.739 758	3.696 799	1.232 711	0.137 954	0.754 528	0.314 454
Cl	7.774 626	3.363 307	7.207 648	0.772 271	1.309 674	0.133 298	0.708 052	0.318 168
Ar	8.816 064	3.080 908	7.131 578	0.713 329	1.372 306	0.128 657	0.680 052	0.323 437
Cl$^-$	7.116 470	0.770 379	6.553 076	3.132 626	2.314 658	5.523 547	2.015 795	0.205 104
K$^+$	8.057 524	2.532 703	7.302 060	0.631 059	1.588 830	0.106 785	1.051 586	4.079 753
Ca^{2+}	8.684 367	2.336 932	5.373 238	0.660 542	3.540 729	0.306 711	0.401 666	4.155 088
Sc^{3+}	8.907 891	2.095 819	5.635 720	0.568 927	3.022 038	0.299 456	0.434 351	3.521 159
Ti^{4+}	9.241 518	1.939 589	5.824 851	0.525 641	2.773 439	0.275 938	0.160 193	3.881 697
V^{5+}	9.399 400	1.797 509	6.139 738	0.484 906	2.399 717	0.248 742	0.061 145	4.486 421
K	8.638 781	2.545 528	7.206 442	0.622 594	1.602 215	0.132 156	1.552 562	8.374 726
Ti^{3+}	9.567 543	1.966 727	6.052 650	0.517 453	2.542 557	0.275 203	0.837 250	3.312 364
Ca	8.693 948	2.282 491	7.147 448	0.579 898	2.564 665	7.610 572	1.593 939	0.125 339
Ti^{2+}	9.649 037	1.983 587	6.304 933	0.509 245	2.288 979	0.272 535	1.757 051	3.444 562
V^{3+}	10.334 381	1.861 758	6.804 518	0.469 648	1.758 907	0.235 611	1.102 195	3.197 846
Sc	9.383 296	2.166 084	7.129 111	0.552 759	2.817 701	6.964 568	1.669 892	0.126 316
V^{2+}	10.283 176	1.883 769	4.482 337	0.304 200	4.228 614	0.564 536	2.005 872	3.319 885
Cr^{3+}	11.202 918	1.777 413	7.672 223	0.430 461	1.223 599	3.148 178	0.901 260	0.158 867
Mn^{4+}	11.873 294	1.692 464	7.833 412	0.413 504	0.806 629	0.131 339	0.486 664	3.401 520
Ti	10.179 911	2.037 452	7.132 532	0.517 294	3.070 890	6.386 433	1.616 667	0.115 598
Cr^{2+}	8.187 313	0.385 965	8.133 178	1.954 180	3.870 581	1.445 772	1.808 928	3.306 830
Mn^{3+}	8.181 363	0.367 702	7.729 015	1.501 534	4.902 977	2.002 181	1.186 645	2.946 096
V	10.991 500	1.923 636	7.101 716	0.488 573	3.295 553	5.914 157	1.611 231	0.110 253
Mn^{2+}	10.642 354	1.587 347	6.301 353	0.410 951	4.074 287	2.726 803	1.982 006	0.231 317
Fe^{3+}	11.176 003	1.506 508	6.461 772	0.388 878	3.546 739	2.446 656	1.815 486	0.218 522
Cr	12.053 584	1.849 519	7.095 113	0.465 262	3.231 251	5.050 360	1.620 051	0.106 205
Fe^{2+}	11.639 034	1.529 107	6.567 351	0.387 355	4.067 058	2.682 187	1.726 557	0.219 492
Co^{3+}	14.037 904	1.623 074	7.394 722	0.416 903	1.675 680	0.227 567	0.891 695	3.389 501
Mn	12.669 058	1.736 774	7.058 290	0.442 391	3.645 676	5.235 619	1.626 976	0.102 473
Co^{2+}	12.633 432	1.493 224	5.105 555	0.271 717	3.869 139	2.663 480	3.391 875	0.467 428
Ni^{3+}	14.863 713	1.573 282	7.641 299	0.411 021	1.628 209	0.224 726	0.866 778	3.390 221
Fe	13.469 371	1.647 444	7.021 916	0.421 012	3.887 224	4.893 841	1.621 489	0.098 215
Ni^{2+}	11.959 629	1.822 263	6.895 023	0.945 746	6.429 426	0.235 207	0.715 922	3.945 902
Co	14.283 606	1.569 695	6.995 793	0.402 696	4.094 384	4.619 481	1.626 217	0.094 890

Cu^{2+}	11.050 786	1.267 041	8.086 054	0.309 981	6.465 339	1.912 235	1.397 820	3.062 180
Ni	15.092 870	1.498 328	6.973 239	0.385 895	4.304 544	4.372 673	1.629 347	0.091 674
Cu^+	11.273 317	1.849 420	8.489 049	1.017 072	6.776 762	0.237 011	1.460 872	3.813 865
Zn^{2+}	12.660 799	1.717 107	8.431 594	0.894 020	6.107 701	0.188 665	0.799 906	3.677 052
Ga^{3+}	11.837 074	1.154 305	8.007 750	0.286 182	6.857 022	1.706 549	1.298 153	2.597 809
Ge^{4+}	12.177 636	1.100 728	7.947 780	0.274 472	6.806 052	1.599 707	1.068 532	2.351 547
Cu	15.861 866	1.435 507	6.951 719	0.371 159	4.549 222	3.687 474	1.637 192	0.088 901
Zn	16.688 703	1.371 272	6.927 053	0.356 739	4.742 990	3.942 481	1.641 254	0.086 169
Ga	17.341 500	1.307 643	6.897 321	0.343 297	5.114 851	4.224 038	1.646 328	0.083 619
Ge	17.665 656	1.237 390	6.846 311	0.329 470	5.848 183	4.199 304	1.639 849	0.080 656
As	17.830 064	1.168 550	6.787 463	0.315 905	6.757 292	4.033 822	1.625 180	0.077 392
Se	17.924 155	1.104 717	7.742 381	3.829 332	6.732 412	0.302 794	1.601 052	0.073 635
Br	17.979 226	1.046 169	8.770 267	3.622 557	6.679 008	0.290 248	1.571 498	0.069 595
Kr	18.010 738	0.992 142	9.830 026	3.425 771	6.635 425	0.277 758	1.523 811	0.064 450
Br^-	15.437 969	0.896 804	7.357 858	4.476 822	7.058 257	0.249 381	6.145 915	1.858 806
Rb^+	17.514 252	0.860 119	7.058 770	0.238 655	6.517 958	2.352 584	4.909 020	3.672 616
Sr^{2+}	17.898 144	0.843 153	10.021 256	2.615 292	7.236 353	0.238 344	0.844 248	4.363 407
Y^{3+}	16.617 093	0.940 599	9.914 059	0.287 196	5.734 424	2.685 174	3.734 424	2.685 296
Zr^{4+}	17.717 084	0.792 322	10.381 566	2.292 058	7.347 274	0.193 219	0.554 077	3.615 006
Nb^{5+}	17.783 062	0.754 853	10.831 963	2.167 269	7.147 105	0.180 536	0.237 869	3.807 918
Mo^{6+}	17.865 655	0.719 244	11.102 167	2.045 986	6.913 625	0.166 458	0.118 553	3.984 778
Rb	18.322 891	0.953 468	10.552 937	3.413 058	6.619 091	0.267 879	1.505 081	0.061 194
Mo^{5+}	17.935 356	0.720 594	11.532 924	2.111 139	6.936 389	0.167 478	0.595 331	3.547 452
Sr	18.732 182	0.921 941	11.148 183	3.460 001	6.608 005	0.259 681	1.511 630	0.059 857
Nb^{3+}	17.899 829	0.757 998	11.115 359	2.270 771	7.197 452	0.182 645	1.787 360	3.835 246
Y	18.960 588	0.886 882	11.961 710	3.405 289	6.603 060	0.250 461	1.474 642	0.055 852
Mo^{3+}	18.032 161	0.723 170	11.844 250	2.199 722	6.982 528	0.169 488	2.141 061	3.709 631
Zr	19.089 263	0.851 759	12.874 035	3.300 137	6.564 667	0.242 255	1.472 035	0.054 437
Ru^{4+}	18.364 515	0.655 397	13.347 439	2.033 358	6.355 173	0.132 064	1.932 873	3.302 506
Nb	18.988 271	0.813 337	14.046 963	3.099 587	6.674 891	0.231 476	1.289 876	0.033 718
Ru^{3+}	18.399 977	0.656 565	13.463 411	2.064 198	6.380 769	0.133 452	2.755 843	3.448 026
Rh^{4+}	18.628 745	0.623 388	14.181 082	1.970 531	5.946 142	0.103 961	2.244 031	3.179 959
Mo	19.034 464	0.778 502	15.095 300	2.968 318	6.635 790	0.221 541	1.234 446	0.025 254
Rh^{3+}	18.662 322	0.624 372	14.316 336	1.998 022	5.968 143	0.105 413	3.053 198	3.319 833
Pd^{4+}	15.660 235	2.056 621	13.824 561	0.726 445	11.778 993	0.289 229	0.736 211	3.960 722
Tc	19.246 530	0.756 495	15.851 553	2.920 806	6.665 848	0.216 759	1.236 069	0.022 204
Ru	17.342 297	0.614 445	12.980 296	1.828 082	7.666 003	3.935 837	6.011 404	0.145 908
Pd^{2+}	16.710 363	2.372 802	12.988 195	0.834 712	11.429 253	0.356 331	2.872 189	0.205 497
Rh	17.802 242	0.606 085	14.948 122	1.935 415	6.127 649	4.178 304	6.121 987	0.144 380
Ag^{2+}	17.161 820	2.089 482	13.013 120	0.300 904	12.942 635	0.738 763	1.882 425	3.992 135
Pd	18.285 102	0.623 182	16.702 826	2.097 766	6.766 207	0.152 292	4.245 864	4.189 627
Ag^+	17.173 750	2.106 563	13.144 553	0.302 849	12.874 187	0.743 302	2.807 511	4.137 271
Cd^{2+}	18.070 381	1.999 237	13.332 503	0.299 797	12.326 299	0.711 421	2.270 818	3.723 391
In^{3+}	18.875 121	1.894 960	13.439 171	0.295 829	11.825 275	0.678 146	1.860 433	3.414 459
Sn^{4+}	20.042 622	1.805 613	13.282 242	0.289 940	11.487 961	0.635 091	1.187 175	3.230 926
Sb^{5+}	20.078 495	1.698 941	13.297 377	0.284 704	11.103 907	0.606 575	1.520 222	2.867 505
Ag	18.315 666	0.598 808	18.194 445	2.041 694	6.639 716	0.144 880	3.850 173	4.731 523
Cd	18.916 563	1.954 597	18.329 119	0.574 567	6.462 895	0.136 124	4.291 423	4.887 115
Sn^{2+}	19.107 667	1.781 495	13.720 301	0.295 095	11.107 774	0.646 238	4.064 258	3.862 564
Sb^{3+}	18.968 920	1.667 488	13.780 631	0.290 205	10.638 430	0.616 938	4.612 018	3.378 141
In	19.542 060	1.896 292	18.063 497	0.565 471	6.815 672	0.145 216	4.578 771	5.232 591
Sn	19.930 558	1.779 296	18.008 421	0.533 292	6.401 586	0.133 543	5.659 436	5.008 159
Sb	20.081 657	1.703 361	17.905 035	0.518 542	6.571 297	4.832 884	6.442 011	0.131 391
Te	20.158 711	1.630 270	17.794 235	0.504 264	7.583 333	4.614 280	6.463 721	0.129 035
I	20.158 673	1.558 040	17.670 863	0.489 361	8.686 104	4.373 731	6.484 359	0.126 800
Xe	20.136 559	1.491 637	17.555 968	0.475 372	9.815 774	4.154 465	6.491 699	0.124 411
I^-	20.749 852	1.525 179	18.497 482	0.455 170	9.763 760	4.642 482	4.988 907	0.085 674
Cs^+	20.758 715	1.325 652	15.506 880	0.399 587	11.112 385	3.597 658	6.622 020	0.221 985
Ba^{2+}	20.576 847	1.267 015	16.266 897	0.380 580	11.466 766	3.293 333	5.689 490	0.208 310
La^{3+}	20.477 686	1.218 631	17.187 606	0.362 830	11.685 045	3.061 163	4.649 664	0.191 311
Ce^{4+}	20.410 528	1.176 557	18.072 818	0.347 351	11.835 613	2.871 742	3.681 041	0.171 442
Cs	20.839 862	1.447 556	17.711 049	0.454 629	10.322 119	4.216 757	6.126 971	0.112 422
Ce^{3+}	21.115 756	1.198 427	18.249 276	0.347 834	12.045 528	2.994 450	3.589 440	0.171 626

Pr^{4+}	21.077 250	1.175 435	16.880 867	0.352 880	11.869 593	2.836 177	5.172 290	0.193 921
Ba	21.538 235	1.441 891	17.607 371	0.451 875	10.360 884	4.433 384	6.493 510	0.119 553
Pr^{3+}	21.731 969	1.193 855	17.086 212	0.352 712	12.147 323	2.951 137	5.034 496	0.194 180
La	21.728 329	1.399 166	17.590 312	0.441 204	11.175 357	4.363 658	6.506 002	0.117 490
Nd^{3+}	21.740 175	1.102 562	21.024 755	0.295 545	9.032 166	2.424 803	5.202 904	3.310 391
Ce	22.447 360	1.373 379	17.599 980	0.431 945	11.431 528	4.242 426	6.521 131	0.115 507
Pm^{3+}	22.412 873	1.084 769	21.022 329	0.289 407	10.201 129	2.416 975	4.363 669	3.342 117
Pr	23.540 254	1.352 596	17.572 828	0.422 717	11.360 438	4.037 463	6.526 481	0.113 343
Sm^{3+}	23.060 332	1.066 243	21.006 650	0.283 379	10.755 233	2.373 879	4.177 785	3.318 962
Nd	24.320 196	1.324 704	17.559 915	0.413 275	11.606 715	3.932 243	6.513 174	0.111 016
Eu^{3+}	23.687 699	1.047 294	20.980 493	0.277 459	11.135 096	2.319 387	4.196 712	3.275 739
Pm	25.125 990	1.294 740	17.517 556	0.403 316	11.877 010	3.827 610	6.479 444	0.108 337
Eu^{2+}	23.938 474	1.055 439	21.039 486	0.278 283	11.295 161	2.336 155	4.726 879	3.530 912
Gd^{3+}	24.312 190	1.028 443	20.946 403	0.271 676	11.498 967	2.264 764	4.242 440	3.236 495
Sm	25.945 873	1.265 903	17.466 242	0.393 907	12.132 576	3.730 575	6.455 309	0.105 867
Tb^{3+}	24.946 581	1.009 614	20.900 872	0.265 993	12.142 240	2.221 608	4.010 307	3.242 896
Eu	26.745 159	1.236 326	17.408 393	0.384 306	12.434 219	3.633 633	6.412 230	0.103 133
Dy^{3+}	25.528 000	0.991 188	20.863 226	0.260 569	11.871 816	2.143 205	4.736 959	3.111 650
Gd	27.419 125	1.205 172	17.367 452	0.374 292	12.870 568	3.646 652	6.342 855	0.099 985
Ho^{3+}	26.121 679	0.973 178	20.820 074	0.255 292	11.868 250	2.077 686	5.189 997	3.031 468
Tb	28.398 703	1.178 655	17.272 243	0.364 758	13.041 363	3.452 056	6.287 691	0.097 130
Er^{3+}	26.729 833	0.955 867	20.778 773	0.250 241	12.067 562	2.024 261	5.423 832	2.975 635
Dy	29.226 260	1.153 302	17.252 510	0.356 071	13.304 453	3.373 036	6.216 777	0.094 051
Tm^{3+}	27.313 951	0.938 644	20.732 451	0.245 310	12.176 988	1.966 492	5.776 610	2.914 185
Ho	29.993 384	1.130 609	17.240 822	0.349 179	13.558 005	3.415 979	6.207 788	0.092 058
Yb^{3+}	27.923 121	0.922 396	20.692 866	0.240 637	12.448 254	1.920 728	5.935 759	2.869 117
Er	30.868 430	1.107 234	17.176 004	0.341 512	13.785 931	3.230 257	6.169 635	0.089 688
Yb^{2+}	28.223 692	0.929 007	20.765 054	0.241 562	13.286 691	1.969 267	5.724 564	3.143 612
Lu^{3+}	28.501 571	0.906 105	20.646 584	0.236 040	12.625 766	1.869 721	6.226 080	2.817 512
Hf^{4+}	28.847 980	0.884 589	20.532 734	0.230 744	12.515 804	1.807 028	6.103 483	2.621 308
Ta^{5+}	29.178 073	0.863 079	20.400 697	0.225 337	12.610 851	1.754 123	5.810 379	2.471 185
W^{6+}	29.521 819	0.842 701	20.276 645	0.220 196	12.984 540	1.714 999	5.216 995	2.359 439
Tm	31.690 965	1.085 800	17.143 092	0.334 824	14.014 143	3.164 613	6.151 800	0.087 707
Yb	32.506 495	1.065 554	17.129 734	0.328 632	14.232 739	3.102 425	6.131 033	0.085 716
Lu	33.411 362	1.047 616	17.119 457	0.322 908	14.336 910	3.165 807	6.132 272	0.084 057
Hf	34.027 279	1.024 429	17.071 349	0.316 592	14.785 836	3.171 410	6.115 537	0.082 258
Os^{4+}	31.114 160	0.811 651	20.102 244	0.211 129	15.295 075	1.815 657	5.488 522	2.963 842
Ta	34.470 518	0.997 571	16.980 327	0.309 267	15.480 430	3.133 490	6.068 725	0.080 176
Ir^{4+}	31.535 331	0.792 897	19.955 609	0.206 054	15.577 774	1.801 446	5.931 286	2.927 505
W	34.789 794	0.971 431	16.922 124	0.302 527	16.274 208	3.076 036	6.013 874	0.077 859
Ir^{3+}	31.741 685	0.795 832	20.012 955	0.206 797	16.002 265	1.845 526	6.243 095	3.152 636
Pt^{4+}	31.921 625	0.774 323	19.798 356	0.200 979	15.997 349	1.790 166	6.282 670	2.885 964
Re	35.076 964	0.945 011	17.142 116	3.011 512	16.866 887	0.294 978	5.914 033	0.074 947
Os	35.281 821	0.919 679	18.068 399	2.939 177	16.805 051	0.288 303	5.844 728	0.072 547
Pt^{2+}	32.281 031	0.780 317	19.939 995	0.202 874	16.983 611	1.876 248	6.795 363	3.350 125
Au^{3+}	32.430 934	0.757 919	19.668 483	0.196 368	16.980 512	1.815 625	6.920 071	3.028 015
Ir	35.458 074	0.894 404	19.059 546	2.861 295	16.770 150	0.280 785	5.712 230	0.068 974
Pt	35.246 233	0.863 775	20.524 789	2.717 948	16.724 185	0.272 301	5.504 793	0.063 875
Au^{+}	32.757 289	0.765 407	19.888 085	0.199 397	18.259 010	1.909 765	7.095 615	3.587 018
Hg^{2+}	32.867 542	0.742 095	19.558 435	0.192 199	18.251 936	1.840 971	7.322 087	3.192 526
Tl^{3+}	33.007 167	0.720 192	19.231 417	0.185 124	18.410 484	1.784 504	7.350 932	2.912 932
Pb^{4+}	33.171 175	0.699 237	18.891 333	0.177 908	18.846 266	1.738 603	7.091 227	2.709 966
Bi^{5+}	33.345 464	0.679 285	19.470 612	1.698 227	18.547 315	0.170 660	6.636 609	2.552 798
Au	35.344 182	0.839 375	21.627 205	2.640 996	16.683 576	0.264 743	5.345 037	0.059 467
Hg^{+}	32.944 717	0.748 446	19.811 879	0.195 808	19.181 911	1.887 315	7.061 492	3.697 184
Hg	35.768 450	0.823 354	22.261 856	2.625 175	16.719 907	0.259 044	5.249 787	0.056 207
Tl^{+}	33.051 873	0.731 194	19.737 842	1.846 164	19.719 260	0.192 133	7.491 025	3.725 910
Pb^{2+}	33.101 449	0.708 869	19.891 879	1.774 963	19.374 679	0.185 023	7.631 993	3.394 062
Bi^{3+}	33.169 526	0.687 827	20.086 620	1.711 763	19.034 095	0.178 040	7.709 759	3.139 209
Tl	36.263 196	0.810 079	22.772 320	2.624 566	16.756 068	0.254 420	5.208 416	0.054 106
Pb	36.792 372	0.798 558	23.214 673	2.632 981	16.830 689	0.250 003	5.162 266	0.051 825
Bi	37.334 220	0.789 142	23.580 435	2.651 119	16.875 340	0.246 998	5.210 006	0.051 816
Po	37.874 954	0.780 763	23.922 445	2.673 664	16.938 384	0.244 258	5.264 217	0.052 051

At	38.343 919	0.773 832	24.255 439	2.697 627	16.918 031	0.243 491	5.482 611	0.055 697
Rn	38.733 863	0.767 366	24.622 392	2.718 367	16.844 462	0.243 881	5.799 283	0.060 444
Ra^{2+}	32.750 466	0.612 715	22.345 558	1.481 191	18.668 772	0.165 015	12.235 204	3.483 634
Ac^{3+}	32.829 095	0.589 953	22.363 535	1.410 141	17.951 046	0.153 370	12.856 324	3.226 624
Th^{4+}	32.953 723	0.568 622	22.404 820	1.347 462	17.231 117	0.141 197	13.410 340	3.017 464
U^{6+}	33.333 367	0.528 766	22.545 179	1.239 026	15.717 529	0.113 184	14.403 925	2.689 223
Fr	39.130 969	0.763 746	24.878 004	2.754 028	16.721 948	0.246 202	6.269 079	0.066 975
Np^{6+}	33.003 560	0.526 755	22.514 616	1.232 003	16.419 334	0.126 042	15.062 490	2.633 769
Ra	39.469 306	0.763 457	25.021 576	2.804 292	16.500 037	0.252 050	7.009 081	0.075 564
U^{4+}	32.444 514	0.559 685	22.558 847	1.323 276	18.280 778	0.154 902	14.715 861	2.894 759
Pu^{6+}	32.625 458	0.525 128	22.537 298	1.226 758	17.171 371	0.137 499	15.665 873	2.580 524
Ac	39.945 130	0.760 506	25.233 314	2.852 104	16.759 683	0.250 608	7.061 873	0.075 237
U^{3+}	31.828 914	0.578 357	22.691 811	1.372 114	19.793 764	0.173 971	14.685 511	3.040 443
Np^{4+}	31.916 461	0.559 177	22.654 285	1.317 758	19.193 491	0.165 482	15.235 763	2.838 745
Th	29.545 645	1.174 559	28.694 760	0.483 106	17.228 973	3.781 848	14.530 622	0.125 618
Np^{3+}	31.258 710	0.578 090	22.847 704	1.366 402	20.738 743	0.182 882	15.154 843	2.979 754
Pu^{4+}	31.268 886	0.559 840	22.798 253	1.313 997	20.237 939	0.176 030	15.694 922	2.784 941
Pa	30.169 065	0.491 229	28.714 068	1.214 832	17.232 441	3.667 592	14.884 426	0.126 024
Pu^{3+}	30.559 529	0.579 257	23.056 376	1.362 607	21.830 070	0.192 093	15.554 024	2.922 697
U	31.239 661	0.494 429	28.545 567	1.247 677	17.067 740	3.655 891	15.147 032	0.125 954
Np	32.197 993	0.497 691	28.565 048	1.282 516	16.779 367	3.653 799	15.457 591	0.126 242
Pu	32.905 564	0.498 690	28.499 560	1.304 627	16.854 673	3.511 603	15.740 204	0.126 388
Am	33.563 940	0.499 222	28.948 487	1.333 049	16.468 228	3.511 851	16.019 345	0.126 519
Cm	34.158 675	0.499 296	29.871 802	1.365 086	16.295 699	0.126 636	15.673 824	3.650 558
Bk	34.597 614	0.498 116	30.603 778	1.385 366	16.553 710	0.126 665	15.244 897	3.655 690
Cf	34.979 821	0.496 577	31.458 289	1.403 780	16.808 138	0.126 685	14.753 753	3.668 383