Chapter 8

Computed Tomography

Authors: Oliver Taubmann, Martin Berger, Marco Bögel,
Yan Xia, Michael Balda, and Andreas Maier

8.1 Introduction

CT is doubtlessly one of the most important technologies in medical imaging and offers us views inside the human body that are as valuable to physicians as they are fascinating (cf. Fig. 8.1).

8.1.1 Motivation

In the previous chapter, we have seen how X-rays can be used to acquire 2-D projection images. However, a single projection image does not retain all spatial information, as it merely shows something akin to "shadows" of the imaged objects. An example is given in Fig. 8.3(a), which shows an X-ray projection image of a luggage bag. Two arrows indicate objects that cannot easily be identified. Using multiple projection images from different angles, we
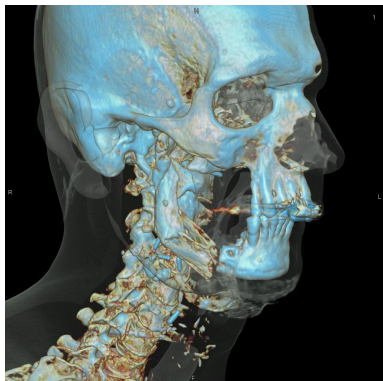
Figure 8.1: Volume rendering of a CT head scan. Image courtesy of Siemens Healthineers AG.
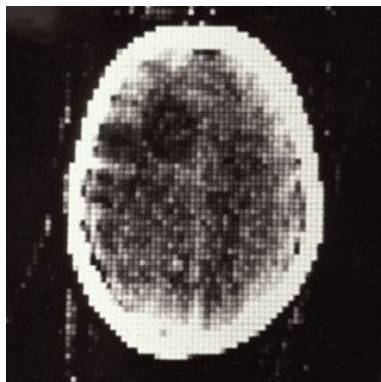

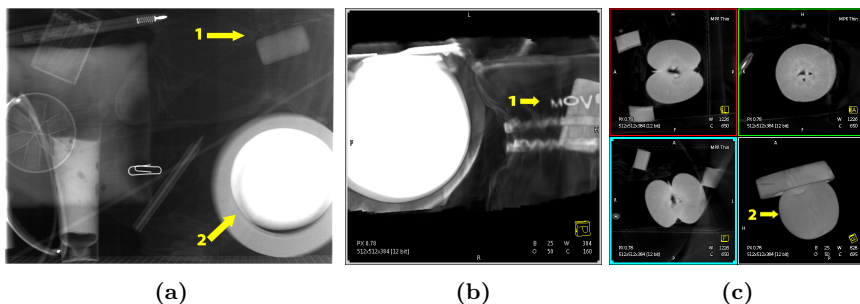**Figure 8.2:** The first clinical CT scan, acquired October 1971 at Atkinson Morley's Hospital in London.

(a)   (b)   (c)

Figure 8.3: 2-D X-ray projection image of a luggage bag **(a)** and a corresponding 3-D reconstruction, visualized with a volume rendering technique and **(b, c)** and as orthogonal cross-sectional slices **(c)**. 1) Indicates a hidden text revealed in **(b)** and 2) an apple that is virtually sliced in **(c)**. Images courtesy of Chris Schwemmer.

are able to perform a 3-D reconstruction and obtain cross-sectional views of the objects. Looking at the reconstructed volume, we can read the letters on the bag (Fig. 8.3(b)) and recognize the bright object as an apple (Fig. 8.3(c)).

8.1.2 Brief History

In 1917, Johann Radon published an article about "the determination of functions by their integrals along certain manifolds," which would not find a

practical application for the following 50 years. The main concepts introduced in his article will be outlined in Sec. 8.2 and used in Sec. 8.3 to explain how CT image reconstruction works.

Only in 1971, the first CT system was built by Sir Godfrey Newbold Hounsfield and Allan McLeod Cormack[1]. Fig. 8.2 shows the result of the first clinical scan of a patient's head performed in the same year. For their seminal invention, they received the Nobel prize in medicine in 1979.

A major advance in the field was the introduction of spiral (or, more accurately, helical) CT by Willi Kalender et al. in 1990. Its name is derived from the novel aspect of its acquisition trajectory describing a helix. Amongst others, this geometry and its advantages will be described in Sec. 8.3.3.

In the early days of CT imaging, data acquisition was fairly slow, taking approximately 4 minutes per rotation. The reconstruction of a single 2-D slice with a low spatial resolution of 80×80 pixels and 3 bit quantization took several hours. By 2002, rotation speed had improved drastically with one rotation performed in only 0.4 seconds. Up to 16 slices in parallel could be reconstructed on-the-fly, at a higher resolution of 512×512 pixels and a quantization depth of 16 bit.

In recent years, this trend continued with temporal and spatial resolutions constantly improving. In this context, the development of dual source CT in 2005 was another significant milestone, featuring two X-ray sources and detectors in a single scanner. In addition to offering additional information when both employed X-ray tubes are operated at different voltages (dual energy scan), it can also be used to speed up the acquisition significantly. The amount of slices acquired in parallel had also increased further, covering a field of view measuring up to 16 cm in axial direction at voxel sizes below one millimeter. This allows imaging of complete organs such as the heart in a single rotation, thus reducing motion artifacts.

In 2014, a modern CT system (cf. Fig. 8.4) was able to acquire up to 128 slices in parallel at a temporal resolution as low as 195 ms with a single X-ray source.

8.2 Mathematical Principles

In this section, we will first introduce the Radon transform as the underlying mathematical principle of the image formation process in CT imaging. Inverting this transform is the fundamental problem solved by image reconstruction methods. Subsequently, we will detail the Fourier slice theorem, a

---

[1] Both researchers were working for EMI at the time, a British music recording and publishing company well known for housing the Beatles' label. This has spawned a widespread belief that the Beatles' success contributed to financing the initial development of CT.

Figure 8.4: At the time of writing, a modern CT scanner can acquire up to 128 image slices in parallel. Image courtesy of Siemens Healthineers AG.

property related to the Radon transform that constitutes the core idea of an important class of reconstruction algorithms.

8.2.1 Radon Transform

Radon's key insight was that any integrable function $f(x, y)$ can be uniquely represented by – and therefore recovered from – all straight line integrals over its domain,

$$p(\ell) = \int_{-\infty}^{+\infty} f(x(l), y(l)) \, \mathrm{d}l, \quad \forall l : (x(l), y(l))^\top \in \text{line } \ell \qquad (8.1)$$

In order to write down all of these integrals without duplicates, a representation of the lines is needed that describes each one uniquely. For this purpose, we can formulate Eq. (8.1) in terms of polar coordinates,

$$p(\theta, s) = \iint_{-\infty}^{+\infty} f(x, y)\delta(x \cos \theta + y \sin \theta - s) \, \mathrm{d}x\mathrm{d}y, \qquad (8.2)$$

with θ the angle between the line's normal vector and the x-axis and s the orthogonal distance between line and origin (cf. Fig. 8.5). Implicitly, this line is described by the equation $x \cos \theta + y \sin \theta = s$. Only those points that satisfy it, i. e., those that fall on the line, are selected by the Dirac function δ in Eq. (8.2), as it vanishes everywhere else (cf. Eq. (2.10)). In that way, the integration of $f(x, y)$ is only performed along the respective line.

The complete set of line integrals $p(\theta, s)$ can now be obtained by going through the angles $\theta \in [0°, 180°]$ and distances $s \in [-\infty, +\infty]$. Apart from orientation which has no influence on the integration, any other line would be
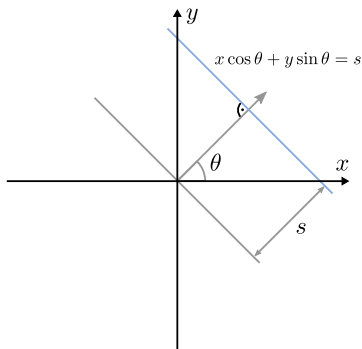
Figure 8.5: The blue line is uniquely described by its distance s to the origin and the angle θ which defines its normal vector $(\cos\theta, \sin\theta)^\top$. This representation immediately gives rise to the implicit line equation.
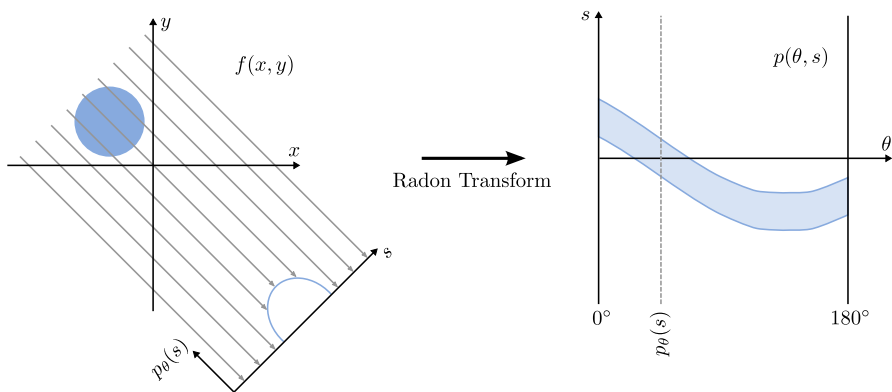

Figure 8.6: $f(x, y)$ has a constant non-zero value inside the blue circle and vanishes everywhere else. We see a single projection on the left and where it fits into the whole sinogram on the right.

equivalent to one of these. For a fixed angle θ, the 1-D function $p_\theta(s) = p(\theta, s)$ is called a projection. It contains all line integrals over $f(x, y)$ with a constant angle θ and variable distance s to the origin. Arranging all projections side-by-side as a 2-D image yields the sinogram. It owes its name to the sinusoidal curves emerging from the underlying geometry. We can see that every point in 2-D except for the origin is found at different distances along the s-axis depending on the angle θ. An example of a sinogram is given in Fig. 8.6.

Turning the function values $f(x, y)$ into line integral values $p(\theta, s)$ is known as the Radon transform in 2-D. The aim of CT reconstruction is the computation of the original function values from measured line integral values, i. e., the inverse Radon transform.
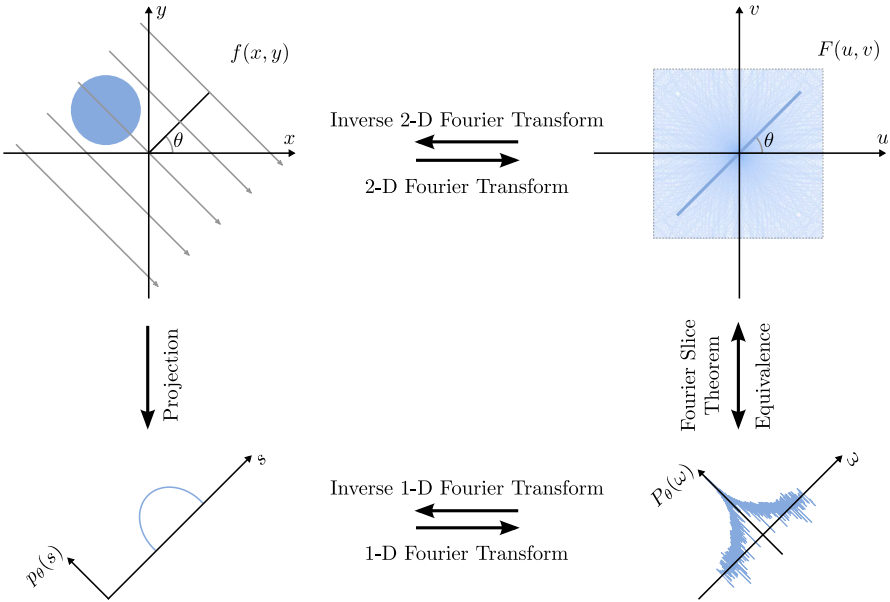
Figure 8.7: The Fourier slice theorem establishes an equivalence between the Fourier transform $P(\xi, \theta)$ of the projection $p_\theta(s)$ and a line in the Fourier transform $F(u, v)$ of $f(x, y)$ which runs through the origin and forms the angle θ with the u-axis. Please note that in the frequency domain images on the right, the magnitudes of the complex numbers were plotted on a logarithmic scale for improved readability.

8.2.2 Fourier Slice Theorem

While it is not immediately clear how to invert the process of projection, we can take a detour through frequency domain. Fig. 8.7 depicts the principle behind the Fourier slice theorem by establishing relationships between the relevant domains. We start by computing the 1-D Fourier transform $P(\xi, \theta)$ of the projection $p_\theta(s)$. The Fourier slice theorem establishes an equivalence that exists between $P(\xi, \theta)$ of the projection $p_\theta(s)$ and a line in the Fourier transform $F(u, v)$ of $f(x, y)$ which runs through the origin and forms the angle θ with the u-axis. A proof of this property can be found in Geek Box 8.1. An intuitive visualization of this relation is displayed in Fig. 8.8. Computation of one 2-D Fourier coefficient is equivalent to projecting the image first, followed by a correlation with the respective 1-D frequency. This is possible as Fourier transform and projection operate in orthogonal direction and are therefore separable as shown in Fig. 8.9. With the complete set of projections, we get many such lines and therefore obtain a good estimate of $F(u, v)$. An inverse 2-D Fourier transform then leads us back to the desired function $f(x, y)$.

Geek Box 8.1: Fourier Slice Theorem

An essential relationship between the projections $p_\theta(s)$ and the function $f(x, y)$ can be established by looking at the frequency domain representations

$$F(u, v) = \mathcal{F}\left\{f(x, y)\right\}, \tag{8.3}$$
$$P(\xi, \theta) = \mathcal{F}\left\{p_\theta(s)\right\}, \tag{8.4}$$

using the Fourier transform $\mathcal{F}$ (cf. Sec. 2.3 (p. 22)). As illustrated in Fig. 8.7, $P(\xi, \theta)$ is equivalent to the part of $F(u, v)$ that falls on a radial line with angle θ. To see why this is the case, we start with the 1-D Fourier transform $P(\xi, \theta)$ of $p_\theta(s)$,

$$P(\xi, \theta) = \int_{-\infty}^{+\infty} p_\theta(s) e^{-2\pi i \xi s} \, \mathrm{d}s. \tag{8.5}$$

Using the definition of the projection $p_\theta(s)$ from the previous section, we obtain

$$P(\xi, \theta) = \int_{-\infty}^{+\infty} \iint_{-\infty}^{+\infty} f(x, y) \delta(x \cos \theta + y \sin \theta - s) \, \mathrm{d}x \mathrm{d}y \, e^{-2\pi i \xi s} \, \mathrm{d}s. \tag{8.6}$$

Rearranging the order of the integrals yields

$$P(\xi, \theta) = \iint_{-\infty}^{+\infty} f(x, y) \int_{-\infty}^{+\infty} \delta(x \cos \theta + y \sin \theta - s) e^{-2\pi i \xi s} \, \mathrm{d}s \mathrm{d}x \mathrm{d}y. \tag{8.7}$$

Eliminating the delta function reads as

$$P(\xi, \theta) = \iint_{-\infty}^{+\infty} f(x, y) e^{-2\pi i (x \cos \theta + y \sin \theta) \xi} \, \mathrm{d}x \mathrm{d}y. \tag{8.8}$$

Variable substitution yields the definition of the 2-D Fourier transform,

$$P(\xi, \theta) = \iint_{-\infty}^{+\infty} f(x, y) e^{-2\pi i (xu + yv)} \big|_{u = \xi \cos \theta, \ v = \xi \sin \theta} \, \mathrm{d}x \mathrm{d}y, \tag{8.9}$$

which finally results in the proposed theorem,

$$P(\xi, \theta) = F(\xi \cos \theta, \xi \sin \theta) = F_{\mathrm{polar}}(\xi, \theta). \tag{8.10}$$

In effect, we can get the complete Fourier transform $F_{\mathrm{polar}}(\xi, \theta)$ of the unknown function $f(x, y)$ in polar coordinates (ξ, θ) by varying θ.

2-D Fourier Transform
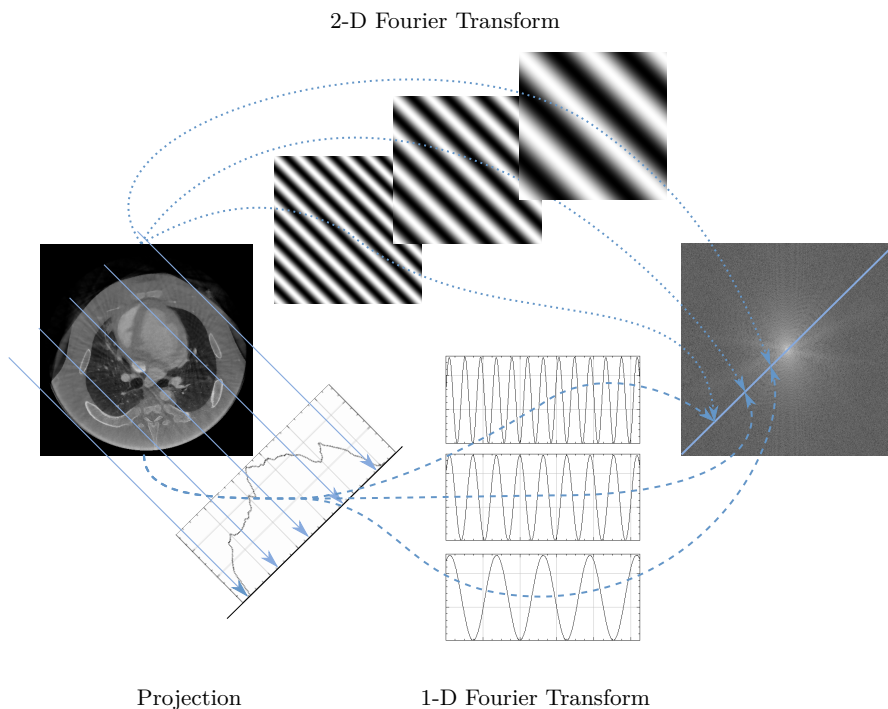

Projection   1-D Fourier Transform

Figure 8.8: Graphical visualization of the Fourier slice theorem. In fact, computation of the projection and correlation with a sinusoidal function (dashed lines) is equivalent to a 2-D correlation with the respective Fourier base function (dotted lines, cf. Fig. 6.9).
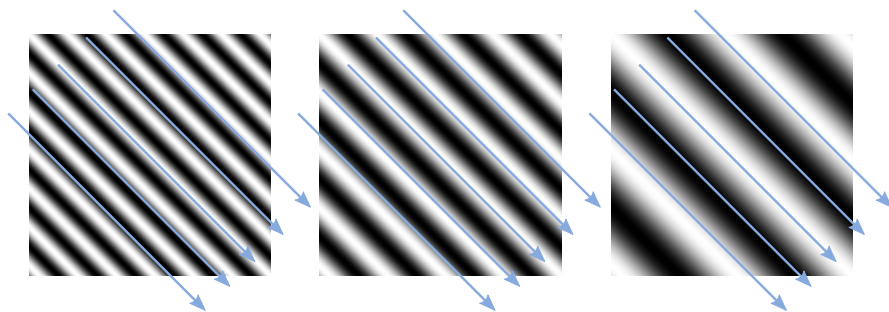

Figure 8.9: A close look at the Fourier base functions reveals that they are actually computing an integration along the wave front. As such projection and convolution operate in orthogonal domains and can therefore be separated into a projection and a 1-D correlation, i. e., a 1-D Fourier transform.

Material / Tissue	HU
Air	-1000
Lung	-600 to -400
Fat	-100 to -60
Water	0
Muscle	10 to 40
Blood	30 to 45
Soft tissue	40 to 80
Bone	400 to 3000

Table 8.1: HUs observed in several materials and tissue classes found in the human body. In general, denser structures exhibit larger HUs.

8.3 Image Reconstruction

As described in Sec. 7.3 (p. 125), X-ray projections can be converted to line integrals using Beer's law, which enables us to apply Radon's ideas to CT reconstruction. A single slice of our imaged object corresponds to the bivariate function $f(x, y)$. More precisely, the function values reconstructed in CT are the linear attenuation coefficients of the imaged material. Typically, they are linearly transformed to the Hounsfield scale, which is normalized such that the absorption of water equals 0 HU,

$$\mu^* = \left(\frac{\mu}{\mu_{\text{Water}}} - 1 \right) \cdot 1000, \tag{8.11}$$

where μ and μ^* denote the coefficients before and after Hounsfield scaling, respectively. Tab. 8.1 lists the HU ranges of several tissue classes found in the human body.

Below, we will discuss the two main methods for 2-D image reconstruction from parallel-beam projections as they have been introduced above. In conventional CT imaging, a 3-D image volume is then obtained simply by acquiring and reconstructing multiple axial slices at slightly offset locations such that they can be stacked on top of each other (Fig. 8.10).

8.3.1 Analytic Reconstruction

Using the Fourier slice theorem, we can derive an analytic reconstruction method known as filtered back-projection.
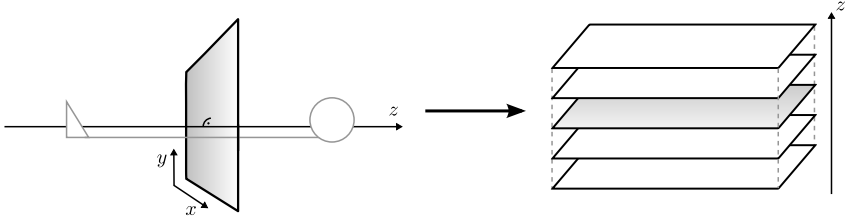
Figure 8.10: In conventional CT, a 3-D image of the body is formed by acquiring, reconstructing, and subsequently stacking 2-D image slices in axial direction. For each slice, all projection rays lie in a plane, which is why we only deal with bivariate functions $f(x, y)$. However, be aware that there are other geometries where this assumption is no longer valid (cf. Sec. 8.3.3).

8.3.1.1 Filtered Back-Projection

It is possible to invert the process of projection directly, without explicitly computing the computations in frequency space suggested by Fig. 8.7. In Geek Box 8.2, it is shown that the required calculations reduce to

$$f(x, y) = \int_0^\pi p_\theta(s) * h(s)|_{s=x \cos \theta + y \sin \theta} \; d\theta, \tag{8.12}$$

where $h(s)$ corresponds to the inverse Fourier transform of $|\xi|$. This amounts to the back-projection of $p_\theta(s)$ convolved with $h(s)$. As a consequence, this method is called filtered back-projection.

Unfiltered back-projection, i. e., just "smearing" line integrals in a projection $p_\theta(s)$ back along their corresponding lines without filtering (cf. Fig. 8.12), is equivalent to adding $P(\xi, \theta)$ to $F(u, v)$, as suggested by the Fourier slice theorem, without considering the factor $|\xi|$. In fact, this is not the inverse, but the dual or adjoint of the Radon transform.

8.3.1.2 Filters

Due to the shape of $|\xi|$, the filter $h(s)$ is typically called ramp filter. Sampling in polar coordinates leads to an oversampling in the center of the Fourier space (cf. Fig. 8.11). Using a ramp filter, this oversampling is corrected by enhancing the high frequency components while dampening the low frequencies in the center of the Fourier space.

According to the sampling theorem, as described in Sec. 2.4.2 (p. 32), with a detector spacing of Δs, the largest frequency that can be detected in $p_\theta(s)$ is $\xi_{max} = \frac{1}{2\Delta s}$. Additionally, noise in the projections is amplified when using an unlimited ramp filter $|\xi|$. Therefore, high frequencies should be limited in the

---

Geek Box 8.2: Filtered Back-Projection

To derive the filtered back-projection algorithm, we start with the inverse Fourier transform of $F(u, v)$,

$$f(x, y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F(u, v) e^{2\pi i(ux+vy)} \, du dv.$$

We can rewrite this equation in polar coordinates $F_{\text{polar}}(\xi, \theta)$ by substituting $u = \xi \cos \theta$ and $v = \xi \sin \theta$. A change in integration variables requires an additional "correction" factor in the integral. This factor is the absolute value of the determinant of the transformation's Jacobian matrix $\boldsymbol{J}$:

$$|\det \boldsymbol{J}| = \left| \det \begin{pmatrix} \frac{du}{d\xi} & \frac{du}{d\theta} \\ \frac{dv}{d\xi} & \frac{dv}{d\theta} \end{pmatrix} \right| = \left| \det \begin{pmatrix} \cos(\theta) & -\xi \sin(\theta) \\ \sin(\theta) & \xi \cos(\theta) \end{pmatrix} \right| =$$

$$= |\xi \cos^2(\theta) + \xi \sin^2(\theta)| = |\xi|.$$

Therefore, performing the change in coordinates, we obtain

$$f(x, y) = \int_0^\pi \int_{-\infty}^{+\infty} F_{\text{polar}}(\xi, \theta) |\xi| e^{2\pi i \xi(x \cos \theta + y \sin \theta)} \, d\xi d\theta.$$

From the Fourier slice theorem, we know $F(\xi, \theta) = P(\xi, \theta)$, thus

$$f(x, y) = \int_0^\pi \int_{-\infty}^{+\infty} P(\xi, \theta) |\xi| e^{2\pi i \xi(x \cos \theta + y \sin \theta)} \, d\xi d\theta.$$

Replacing $x \cos \theta + y \sin \theta$ with s, this reads as

$$f(x, y) = \int_0^\pi \int_{-\infty}^{+\infty} P(\xi, \theta) |\xi| e^{2\pi i \xi s} \, d\xi d\theta,$$

which contains a product of $P(\xi, \theta)$ and $|\xi|$. A product in Fourier space corresponds to a convolution in spatial domain (cf. Sec. 2.3.2 (p. 25)). If we denote the inverse Fourier transform of $|\xi|$ with $h(s)$, we find the following spatial domain representation:

$$f(x, y) = \int_0^\pi p_\theta(s) * h(s)|_{s=x \cos \theta + y \sin \theta} \, d\theta. \qquad (8.13)$$

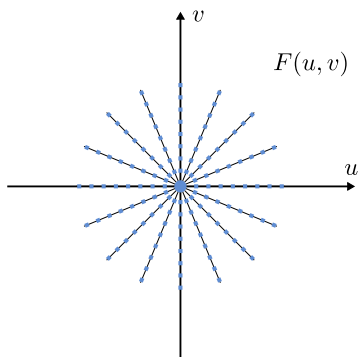This amounts to the back-projection of $p_\theta(s)$ convolved with the filter kernel $h(s)$.

$$b_\theta(x,y) = p_\theta(s)|_{s=x\cos\theta+y\sin\theta}$$

Figure 8.11: Sampling in polar coordinates causes the density of samples to increase with proximity to the origin, whereas the more distant areas are under-represented.
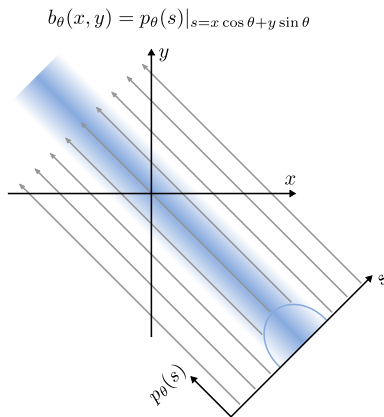
Figure 8.12: The back-projection $b_\theta(x,y)$ of a single projection $p_\theta(s)$ hardly gives us an idea of the original function $f(x,y)$. However, we can reconstruct it by back-projecting a sufficient set of appropriately filtered projections.

filtered projections $\tilde{p}_\theta(s)$. For this purpose, we can generalize the following equation

$$\tilde{p}_\theta(s) = \int_{-\infty}^{+\infty} P(\xi,\theta)|\xi|e^{2\pi i\xi s}\,\mathrm{d}\xi, \tag{8.14}$$

by replacing $|\xi|$ with an arbitrary filter $H(\xi)$:

$$\tilde{p}_\theta(s) = \int_{-\infty}^{+\infty} P(\xi,\theta)H(\xi)e^{2\pi i\xi s}\,\mathrm{d}\xi, \tag{8.15}$$

In practice, various ramp-like filters are used depending on the desired image characteristics, typically involving a trade-off between a smoother image appearance and a higher spatial resolution.

One of the most widely known filters was described by Ramachandran and Lakshminarayanan, in short known as the "Ram-Lak" filter. It corresponds to $|\xi|$ cut off at $\xi_{\max}$ on both sides. In the spatial domain, this results in a filter kernel that reads

$$h(s) = \frac{\mathrm{sinc}\left(\frac{s}{\Delta s}\right)}{2(\Delta s)^2} - \frac{\mathrm{sinc}^2\left(\frac{s}{2\Delta s}\right)}{4(\Delta s)^2}, \tag{8.16}$$

a discretized version of which can be convolved with the discrete projection. A derivation of the Ram-Lak filter is given in Geek Box 8.3.

Geek Box 8.3: Ram-Lak Filter

In order to derive the Ram-Lak filter, we need to start with the inverse
Fourier transform of $|\xi|$

$$h(s) = \int_{-\infty}^{+\infty} |\xi| e^{2\pi i \xi s} \, d\xi.$$

Now, we introduce a band limitation that only allows frequencies $|\xi| \leq B$:

$$h(s) = \int_{-B}^{B} |\xi| e^{2\pi i \xi s} \, d\xi = \int_{-\infty}^{+\infty} |\xi| e^{2\pi i \xi s} \, \mathrm{rect}\left(\frac{\xi}{2B}\right) d\xi$$

Note the we use the rectangular function (cf. Tab. 2.2) to express the
band limitation above. Furthermore, we can also use this function to
express $|\xi|$ as the convolution of two rectangular functions yields a
triangular function (cf. Tab. 2.2):

$$|\xi| = B - \mathrm{rect}\left(\frac{\xi}{B}\right) * \mathrm{rect}\left(\frac{\xi}{B}\right)$$

Now the band-limited inverse Fourier transform of $|\xi|$ takes the fol-
lowing form:

$$h(s) = \mathcal{F}^{-1}\left[\left(B - \mathrm{rect}\left(\frac{\xi}{B}\right) * \mathrm{rect}\left(\frac{\xi}{B}\right)\right) \mathrm{rect}\left(\frac{\xi}{2B}\right)\right]$$

$$= \mathcal{F}^{-1}\left[B \, \mathrm{rect}\left(\frac{\xi}{2B}\right)\right] - \mathcal{F}^{-1}\left[\underbrace{\left(\mathrm{rect}\left(\frac{\xi}{B}\right) * \mathrm{rect}\left(\frac{\xi}{B}\right)\right)}_{\text{support on } [-B,B]} \underbrace{\mathrm{rect}\left(\frac{\xi}{2B}\right)}_{=1 \text{ on } [-B,B]}\right]$$

$$= \mathcal{F}^{-1}\left[B \, \mathrm{rect}\left(\frac{\xi}{2B}\right)\right] - \mathcal{F}^{-1}\left[\left(\mathrm{rect}\left(\frac{\xi}{B}\right)\right) \cdot \mathcal{F}^{-1}\left[\mathrm{rect}\left(\frac{\xi}{B}\right)\right)\right]$$

$$= 2B^2 \, \mathrm{sinc}(2Bs) - B^2 \, \mathrm{sinc}^2(Bs)$$

With $B = \xi_{\max} = \frac{1}{2\Delta s}$, we arrive exactly at Eq. (8.16). In the discrete
case, s needs to be an integer number. Thus, we can simplify above
equation even further to:

$$h_s = \begin{cases} \frac{1}{2\Delta s^2} & s = 0 \\ 0 & s \text{ even} \\ \frac{1}{\pi^2 (s\Delta s)^2} & s \text{ odd} \end{cases}$$
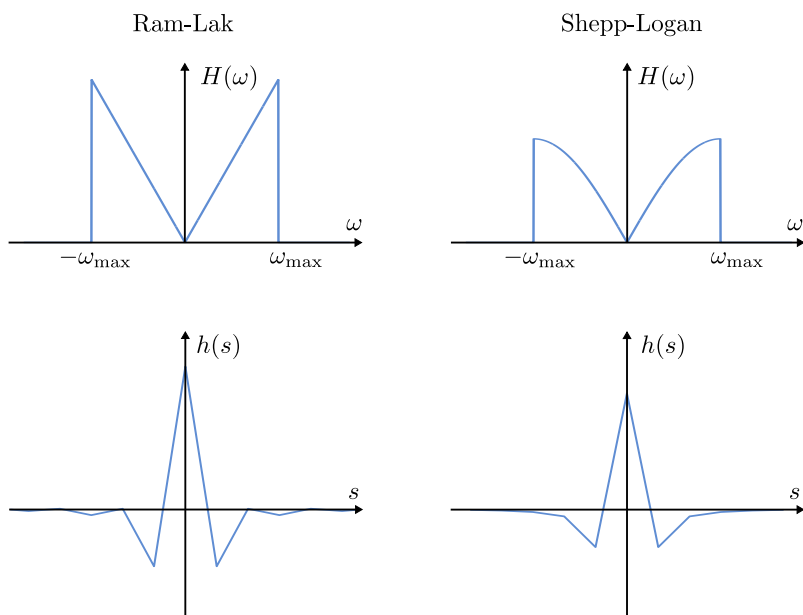
Ram-Lak   Shepp-Logan

Figure 8.13: The responses in frequency domain (top row) as well as the discretized spatial domain kernels (bottom row) of the Ram-Lak and Shepp-Logan filters.

To suppress noise, a windowing function can be multiplied with the filter in frequency domain which lowers its response for frequencies close to ξ_{max}. In the case of the commonly used filter proposed by Shepp and Logan, this windowing function is

$$\left| \mathrm{sinc} \left(\frac{\xi}{2\,\xi_{\mathrm{max}}} \right) \right|. \tag{8.17}$$

This leads to a slightly different function $h(s)$ which can be discretized in the same manner. Fig. 8.13 shows plots of the Ram-Lak and Shepp-Logan filters.

8.3.1.3 Discretization

Through discretization, the convolution

$$\tilde{p}_\theta(s) = \int_{-\infty}^{+\infty} p_\theta(s) \cdot h(s - s')\,\mathrm{d}s' \tag{8.18}$$

becomes

$$\tilde{p}_{\theta,s} = \sum_{s'} p_{\theta,s} \cdot h_{s-s'}\,\Delta s. \tag{8.19}$$
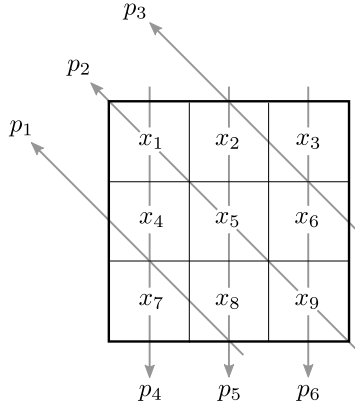
Figure 8.14: Example for an image grid and some projection rays.

An example for a discrete filter h_s is given in Geek Box 8.3.

With an angle increment of $d\theta = \frac{\pi}{N}$, where N is the number of acquired projections, the final back-projection step in Eq. (8.13) can be written as

$$f(x, y) = \frac{\pi}{N} \sum_i \tilde{p}_{\theta_i}(s)|_{s = x \cos \theta_i + y \sin \theta_i}, \tag{8.20}$$

where θ_i denotes the ith angle. Note that the value $\tilde{p}_{\theta_i}(s)$ will generally have to be interpolated from the $\tilde{p}_{\theta_i, s}$ since s is not necessarily an integer number. For each position (x, y), we can find $f(x, y)$ by summing over corresponding (interpolated) values of each filtered projection $\tilde{p}_{\theta_i}$. As a rule of thumb, it is recommended to avoid interpolation in the output space, i.e., in our case, we should sample f directly. This comes naturally with the formulation given in Eq. (8.20). In contrast, back-projecting one $\tilde{p}_{\theta_i}$ at a time to the whole volume would require interpolation on grid points in the domain of f in each step.

8.3.2 Algebraic Reconstruction

A second approach to CT image reconstruction defines the problem as a system of linear equations. Each projection ray corresponds to a linear equation that sums up the image pixels the ray passes through, i.e., computes its discrete line integral, and demands it to equal the measured line integral value. Fig. 8.14 shows an exemplary image grid and a set of projections rays.

Accordingly, we can define the image reconstruction problem as

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{p}, \tag{8.21}$$

where

$$\boldsymbol{x} = (x_1, x_2, ..., x_N)^\top \tag{8.22}$$
$$\boldsymbol{p} = (p_1, p_2, ..., p_M)^\top \tag{8.23}$$

are the sequentially numbered unknown pixels and measured line integrals, respectively. Each element a_{ij} of the system matrix $\boldsymbol{A}$ describes the contribution of a particular pixel to a particular ray. There are many possibilities to model $\boldsymbol{A}$. In the simplest case, a_{ij} is a binary value that is 1 when the ray passed through the pixel and 0 otherwise. The length of intersection can also be used, or even the area of intersection in case we assume the rays to have a non-zero thickness.

Solving this system of linear equations for the solution $\boldsymbol{x}$ directly using matrix inversion (Gaussian elimination, singular value decomposition, etc.) is not feasible in practice as the problems are typically large, ill-conditioned and over-determined. Instead, an iterative solution to this system of linear equations is sought.[2]

The algebraic reconstruction technique (ART) aims to find such an iterative solution using the Kaczmarz method. The basic idea behind this method is that each linear equation defines a line (2-D) or, generally speaking, a hyper plane (higher dimensions) in the solution space, the dimensionality of which equals the number of unknowns. All points on a hyper plane fulfill its corresponding equation. Consequently, the point of intersection of all hyper planes forms the correct solution to the problem. Thus, by repeatedly projecting the current estimate orthogonally onto a different equation's plane, we iteratively improve the solution (Fig. 8.15). A simple mathematical intuition for the ART algorithm is given in Geek Box 8.4.

Using Kaczmarz' method, we can now find an iterative solution for Eq. (8.21). For each line integral measurement p_i and each *row* $\boldsymbol{a}_i$ of the system matrix $\boldsymbol{A}$ we perform the following update step,

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \frac{p_i - \boldsymbol{a}_i \boldsymbol{x}^k}{\boldsymbol{a}_i \boldsymbol{a}_i^\top} \boldsymbol{a}_i^\top, \tag{8.29}$$

and repeat until convergence.

It has been shown by Tanabe in 1971 that if a unique solution exists, this iterative scheme converges to the solution. However, in over-determined systems and in presence of noise, no unique solution might be found and the method might oscillate around the ideal solution. The rate of convergence depends on the angle between the lines. If two lines are orthogonal to each other, the method converges very quickly as the orthogonal projection immediately finds the intersection. Thus, orthogonalization methods can be applied

---

[2] It is worthwhile to note that while $\boldsymbol{A}^{-1}\boldsymbol{p}$ corresponds to the ideal solution, which in the previous section we would obtain by filtered back-projection, $\boldsymbol{A}^\top \boldsymbol{p}$ amounts to an *unfiltered* back-projection.

Geek Box 8.4: 2-D Algebraic Reconstruction Technique Example

In the 2-D case, consider a point $\boldsymbol{x}$ and a line $\boldsymbol{n}^\top \boldsymbol{c} = d$, where $\boldsymbol{n}$ is the normal vector, $\boldsymbol{c}$ a point on the line and d the distance to the origin. Note that $\boldsymbol{n}^\top \boldsymbol{c}$ describes the scalar vector product. The orthogonal projection $\boldsymbol{x}'$ of $\boldsymbol{x}$ on this line must be in the direction of the normal vector $\boldsymbol{n}$:

$$\boldsymbol{x}' = \boldsymbol{x} + \lambda \boldsymbol{n} \tag{8.24}$$

The projected point $\boldsymbol{x}'$ is part of the line and therefore fulfills

$$\boldsymbol{n}^\top \boldsymbol{x}' = d. \tag{8.25}$$

Plugging Eq. (8.24) into this equation we get

$$\boldsymbol{n}^\top (\boldsymbol{x} + \lambda \boldsymbol{n}) = d, \tag{8.26}$$

which can be rewritten as

$$\lambda = \frac{d - \boldsymbol{n}^\top \boldsymbol{x}}{\boldsymbol{n}^\top \boldsymbol{n}}. \tag{8.27}$$

Substituting λ in Eq. (8.24), we arrive at

$$\boldsymbol{x}' = \boldsymbol{x} + \frac{d - \boldsymbol{n}^\top \boldsymbol{x}}{\boldsymbol{n}^\top \boldsymbol{n}} \boldsymbol{n}. \tag{8.28}$$
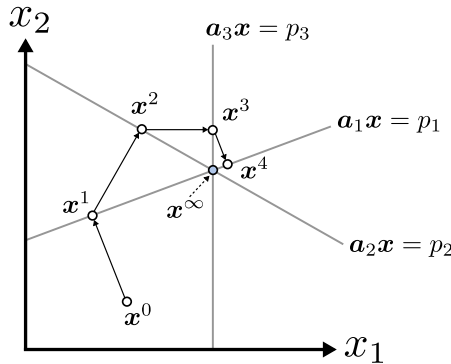

Figure 8.15: Kaczmarz iterations in 2-D space.

in advance in order to improve convergence. However, using such methods is computationally demanding and amplifies noise in the measurements.

Several extensions of this algorithm aim to improve convergence speed. Instead of orthogonalization, for instance, one can also use ordered subsets of the equations to select a better order of the performed projections. Another extension, Simultaneous ART (SART), achieves a speed-up by doing multiple updates at the same time and then combining the results. In each step, the current estimate is orthogonally projected on all lines. The centroid of all projected points is then used for the next iteration. This results in the following update rule:

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \lambda_k \sum_i u_{k,i} \frac{p_i - \boldsymbol{a}_i \boldsymbol{x}^k}{\boldsymbol{a}_i \boldsymbol{a}_i^\top} \boldsymbol{a}_i^\top, \tag{8.30}$$

with

$$\sum_i u_{k,i} = 1, \tag{8.31}$$

where λ_k controls the step size in each iteration.

Other than the presented method by Kaczmarz, there are a multitude of optimization approaches to solve this problem that are not covered here, e.g., Gradient Descent, Maximum-Likelihood Expectation-Maximization, or regularized reconstruction methods. There is also an immediate relation to analytical methods as described in Geek Box 8.5.

8.3.3 Acquisition Geometries

Fig. 8.16 illustrates several important acquisition geometries in CT imaging. Different types of CT scanners have been categorized into generations. CT scanners of the first generation practically realized the parallel beam geometry as introduced above and shown in Fig. 8.16(a) (left). By introducing an array of detectors, the second generation could measure beams from several directions simultaneously. Only by the third generation, however, was this fan of directions (cf. Fig. 8.16(a), right) wide enough to remove the need for a translational motion during acquisition. The projections acquired using a fan beam geometry can be transformed ("rebinned") such that the reconstruction methods described earlier can still be applied. Alternatively, corresponding fan beam versions of the algorithms can be derived in a similar fashion to the ones presented.

Another essential development is the addition of multiple detector rows (Fig. 8.16(b), left), leading to a dramatically increased imaging speed as many slices can be acquired in parallel (multi-slice CT). Another, newer kind of CT systems expands on this notion: by acquiring full 2-D projection images with an image intensifier or – more recently – a flat panel detector, cone beam
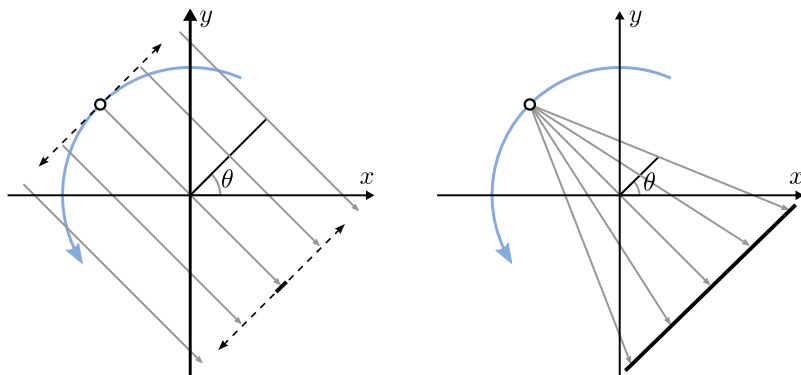
(a) For a parallel beam geometry as introduced in the previous sections (shown on the left), the X-ray source needs to be shifted perpendicularly (dotted line) to the direction of projection, casting pencil beams through the object. If all beams instead emanate from a single position for each angle, we obtain a fan of no longer parallel rays (fan beam geometry; on the right), increasing acquisition efficiency at the cost of a slightly more complicated reconstruction problem. Apart from the flat shape shown here, there also exist curved detectors with an equiangular spacing.
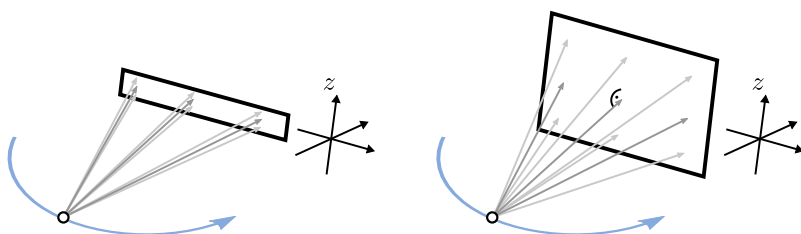
(b) Multiple detector arrays allow for simultaneous acquisition of multiple image slices from one X-ray source position (multi-slice CT; shown on the left). However, in this setup, the beams no longer all lie within the rotation plane. This issue becomes much more important in the case of cone beam CT (shown on the right): Here, the small stack of detector rows gives way to a larger detector matrix, with the beams now forming a cone in 3-D.

Figure 8.16: Basic acquisition geometries in CT imaging. Blue arrows indicate the trajectory of the X-ray source. The detector is depicted by thick black lines.

> ### Geek Box 8.5: ART and its Relation to Filtered Back-Projection
>
> ART formulates the reconstruction problem as a system of linear equations.
>
> $$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{p}$$
>
> For a typical 3-D reconstruction problem with 512 projections with 512^2 pixels each, $\boldsymbol{p} \in \mathbb{R}^{512^3}$ while the corresponding volume $\boldsymbol{x} \in \mathbb{R}^{512^3}$. Consequently, the operator is huge with $\boldsymbol{A} \in \mathbb{R}^{512^3 \times 512^3}$. In order to store such a matrix in floating point precision about 65,000 TB of memory would be required. However, $\boldsymbol{A}$ is very sparse as most entries are equal to 0. In computer implementations, it is typically computed on the fly using *ray-casting*. Thus, general inversion of $\boldsymbol{A}$ is infeasible, even when using the pseudo inverse with
>
> $$\boldsymbol{x} = \boldsymbol{A}^{\dagger}\boldsymbol{p} = \boldsymbol{A}^{\top}(\boldsymbol{A}\boldsymbol{A}^{\top})^{-1}\boldsymbol{p}.$$
>
> However, there are certain geometries for which $(\boldsymbol{A}\boldsymbol{A}^{\top})^{-1}$ can be determined analytically. For the case of parallel-beam geometries, we know that $(\boldsymbol{A}\boldsymbol{A}^{\top})^{-1}$ takes the form of a convolution with the ramp filter. $\boldsymbol{A}^{\top}$ is the adjoint operation of the projection operator. In continuous form, we introduced this operation already as back-projection (cf. Geek Box 8.2). Thus, filtered back-projection is a direct solution for the above system of linear equations.

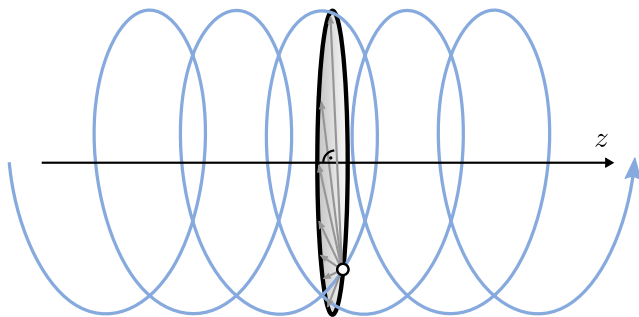**Figure 8.17:** A C-arm system for interventional X-ray imaging. Image courtesy of Siemens Healthineers AG.

Figure 8.18: In spiral CT, although the X-ray source still conveniently rotates in the x-y plane, the trajectory it describes in relation to the imaged object is a helix due to the patient table being slowly moved through the gantry. This enables the acquisition of a large object region while rotating continuously. Projections for an ideal circular trajectory can be interpolated along z from neighboring helical segments.

CT is able to capture a large field of view containing all of the object in a single rotation (Fig. 8.16(b), right). One of the main fields of use for this technology lies in interventional imaging where the X-ray source and detector are mounted on a C-arm device (Fig. 8.17). It has to be noted, though, that for arbitrary objects, an exact reconstruction is only possible in the plane of rotation. The more the beam diverges from this plane, the more artifacts are likely to appear: due to the incomplete data obtained from oblique rays, the reconstruction problem is underdetermined.

For imaging larger parts of the body with few detector rows, it used to be necessary to perform a rotation, then halt and move the table such that the next slice to be acquired is lined up with the detector before starting anew. With the invention of helical CT, a continuous motion of both the rotating gantry and the table became possible. From the point of view of the imaged object, the X-ray source rotates in the x-y plane and moves in the axial direction at the same time, thus following a helix (Fig. 8.18). From the helical rotation, projections for all angles in an axial plane can be interpolated, enabling the use of standard reconstruction methods.

8.4 Practical Considerations

So far, we have described the theoretical background and principles for CT image reconstruction. However, in practice there are several aspects that have to be considered.
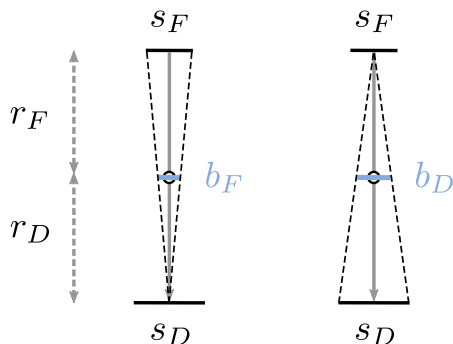
Figure 8.19: The effective sizes b_F of the focus and b_D of the detector in the isocenter can be calculated from the distances to the isocenter r_F and r_D.

8.4.1 Spatial Resolution

In many medical applications, we are not only interested in visualizing large organs, but also smaller structures, such as small blood vessels or calcifications. Visualization of these small structures requires a high spatial resolution. In the following subsection, we will discuss what affects resolution in the x-y scan plane. Resolution in the z-direction typically needs to be considered separately as it depends on different factors.

In the scan plane, resolution depends on several geometrical properties. Focus size, scan geometry, detector element spacing and aperture, and movement of the focus during image acquisition all influence the resolution.

The focus size s_F as well as the detector aperture s_D contribute to image blurring, which can be modeled by

$$b_F = \frac{r_D}{r_F + r_D} \cdot s_F \quad \text{and} \tag{8.32}$$

$$b_D = \frac{r_F}{r_F + r_D} \cdot s_D, \tag{8.33}$$

where r_F represents the distance of the isocenter, i.e., the center of rotation, to the X-ray focus and r_D the distance of the isocenter to the detector center. Effectively, b_F and b_D are the sizes of the focus and detector in the isocenter (cf. Fig. 8.19). Furthermore, the continuous movement of focus and detector during the image acquisition results in additional image blur, which we denote as b_M. The blur that occurs during image acquisition is then described by

$$b_{\text{acq}} = \sqrt{b_F^2 + b_D^2 + b_M^2}. \tag{8.34}$$

However, sampling and image reconstruction also introduce additional blur,
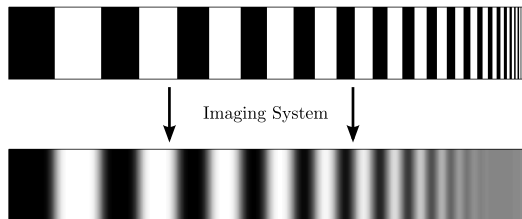
Figure 8.20: A bar phantom can be used to evaluate the spatial resolution of an imaging system. At sufficiently high spatial frequencies, individual lines can no longer be separated after imaging, i.e., we have determined the system's resolution limit.

$$b_A = c_A \cdot \Delta s, \tag{8.35}$$

with Δs the sampling distance and c_A a constant factor which represents the reconstruction algorithm characteristics[3]. Finally, the total blur is modeled by

$$b_{\text{total}} = \sqrt{b_F^2 + b_D^2 + b_M^2 + b_A^2}, \tag{8.36}$$

whereas b_{acq} represents the maximum spatial resolution given by the geometric setup, which could be achieved if we were to use a very fine sampling and a reconstruction algorithm with a sharp kernel. It becomes obvious that as a user, we only have limited influence on spatial resolution. We can decide which convolution kernel we want to use, but the geometrical parameters are defined by the system's scan modes.

Spatial resolution can be measured directly and indirectly. For direct measurement, a bar phantom can be used. Such phantoms consist of alternating black and white bars of varying thickness. The resolution is determined by evaluating whether bars of a certain thickness are still distinguishable after acquisition and reconstruction (cf. Fig. 8.20). A more reliable and objective evaluation is the indirect approach. For this purpose, we scan a thin wire phantom[4], thereby obtaining the system's so called point spread function (PSF). The Fourier transform of the PSF yields the modulation transfer function (MTF) (cf. Fig. 8.21). Frequency is typically measured in line pairs per cm (lp/cm), a unit that can be intuitively understood if we recall the bar phantoms of the direct approach mentioned before. The spatial resolution of a system is often given by the 10% value of the MTF, which represents the

---

[3] E.g., in filtered back-projection, a smooth convolution kernel reduces noise but also spatial resolution, whereas a sharp kernel leads to more noise but yields a high spatial resolution (cf. Sec. 8.3.1).

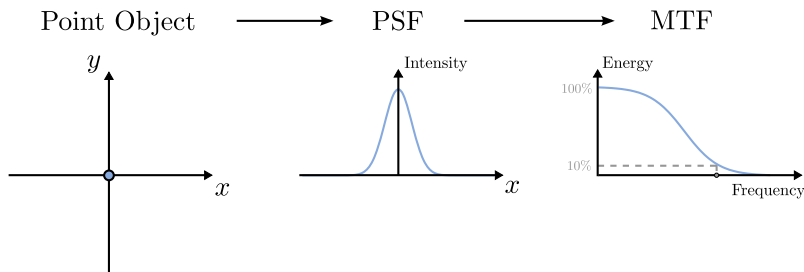[4] Essentially, this mimics a point object for each 2-D slice.

Figure 8.21: If we could scan an ideal point object, the resulting reconstructed image would be the PSF of the system. The Fourier transform of the PSF is the MTF, which shows the relative contrasts achievable for varying object frequencies. In practice, this measurement is typically performed using thin wire phantoms.
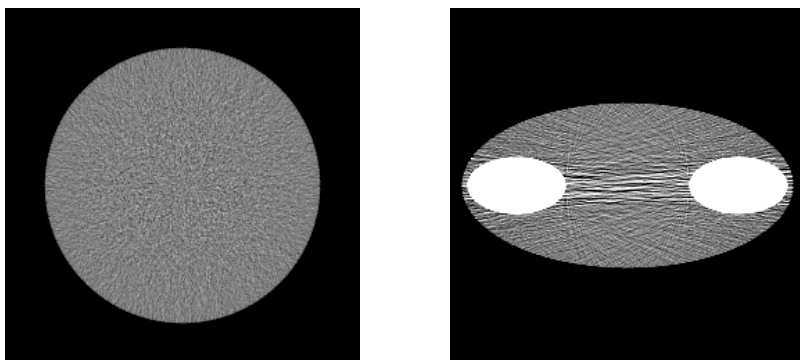

Figure 8.22: The left image shows the reconstruction of a water cylinder phantom. The noise is stronger in the center than in the peripheral regions. On the right side, an elliptic phantom with two high intensity insets is depicted. In its center, streak noise emerges that is caused by the strongly attenuating structures. For both phantom simulations, 90,000 photons per pixel with an energy of 75 keV were used.

frequency at which the contrast has dropped to 10% of the maximum value at $0\,\mathrm{lp/cm}$.

8.4.2 Noise

From the considerations regarding noise in X-ray projections (cf. Sec. 7.4.3 (p. 136)), we know that the number of photons n_s measured by our detector

can be modeled by a Poisson distribution with an expected value of $E[n_s] = N_0 p_a$, where N_0 is the expected value of the number of photons generated by the X-ray tube, and p_a is the probability of a photon passing through the imaged object unaffected. The Poisson process can be approximated by a Gaussian distribution with mean value $\mu = E[n_s]$ and standard deviation $\sigma = \sqrt{E[n_s]}$, if we are dealing with a high number of events, i.e., photons.

For reconstruction, we convert the measured projection images to line integral images by taking the negative logarithm (cf. Sec. 7.3.1 (p. 126)),

$$\int \mu(s)\mathrm{d}s = -\ln \frac{I}{I_0}, \tag{8.37}$$

where $\frac{I}{I_0} = \frac{N_0 p_a}{N_0} = p_a$.

Using the first order Taylor expansion, it can be shown that this transform leads to a new approximate Gaussian distribution with $\mu = -\ln p_a$ and $\sigma = \frac{1}{\sqrt{N_0 p_a}}$. Note that the noise variance increases with object thickness.

During reconstruction, the back-projection step computes a weighted sum of the (filtered) projection values. Hence, the object dependence of the noise statistics is propagated into 3-D. This can be seen in Fig. 8.22, where most of the noise is found in the center of the objects. Additionally, in a non-circular object, streak structures appear in the noise. Therefore, denoising in the reconstructed domain needs to take the directional nature of the noise generation into account.

8.4.3 Image Artifacts

An ideal image reconstruction is only possible in theory. In reality we have to deal with different physical phenomena which are detrimental to image quality and can result in image artifacts. In the following paragraphs, the most common types of image artifacts and ways to reduce their influence will be discussed.

8.4.3.1 Beam Hardening

In practice, CT uses polychromatic X-ray sources, which leads to the attenuation of a homogeneous object being not proportional to the thickness of the object along the ray. A polychromatic X-ray source produces a wide, continuous spectrum of energies and X-ray attenuation coefficients are dependent on the energy. A detailed mathematical description of the spectrum is provided in Sec. 8.5.

When an X-ray passes through an object, lower energy photons are more easily absorbed than higher energy photons. This effect is called "beam hard-
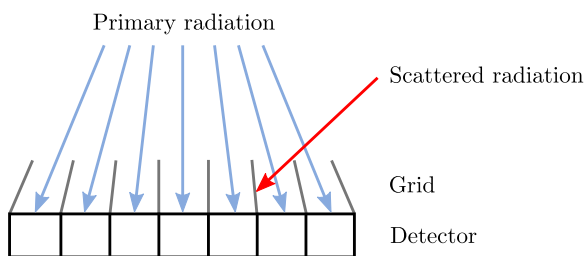
Figure 8.23: Anti-scatter grids can be placed in front of the detector to reject scattered radiation.

ening". Beam hardening results in streak and cupping artifacts. A common approach to deal with beam hardening is to physically pre-filter the X-rays using thin metallic plates, which absorb the low energy photons. An in-detail discussion of beam hardening, including examples, is provided in Sec. 8.5.3.

8.4.3.2 Scatter Artifacts

Scatter, or more specifically Compton scatter, causes X-ray photons to change direction and energy. A scattered photon can therefore be measured in a different detector element than intended. This has an especially large effect when the scattered photon is measured in a detector element that normally would have only few photons, e. g., if a high density object like a metal implant blocks all incoming photons, the corresponding detector element only detects scattered photons. Scatter artifacts are noticeable as cupping and streak artifacts especially between high density structures. Most scanners use anti-scatter grids in front of the detector to reduce scatter. This grid consists of thin lead strips, separated by a X-ray transparent spacer material. It is placed on the detector and aligned towards the focal spot. Thus, a photon that was not scattered can pass through the grid, while most scattered photons will be absorbed by the lead, cf. Fig. 8.23.

8.4.3.3 Partial Volume Effect

Partial volume artifacts appear mostly in low resolution images, especially in thick slice images. With low resolution, it is possible that one pixel consists of two regions with different absorption coefficients μ_1 and μ_2, cf. Fig. 8.24, which leads to streak artifacts in the reconstruction. Geek Box 8.6 describes the problem in more detail. This type of artifact is not often seen with state-of-the-art CT systems as the image resolution and especially slice resolution has improved drastically.
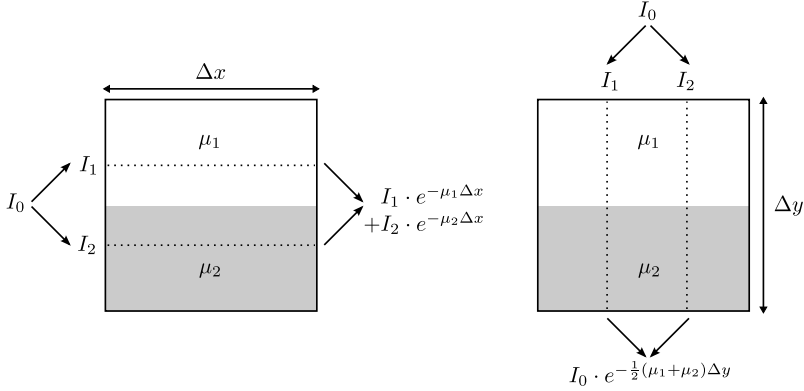
Figure 8.24: Two regions within a pixel with different absorption coefficients result in different measured intensities I for different projection angles.

Geek Box 8.6: Partial Volume Effect

In the first case, we observe two separate regions with the corresponding absorption equations

$$I_1 \cdot e^{-\mu_1 \Delta x}, \tag{8.38}$$
$$I_2 \cdot e^{-\mu_2 \Delta x}, \tag{8.39}$$

where $I_1 + I_2 = I_0$. Thus, the total measured intensity in this pixel is

$$I_x = I_1 \cdot e^{-\mu_1 \Delta x} + I_2 \cdot e^{-\mu_2 \Delta x}, \tag{8.40}$$

which is not equivalent to the average absorption we would expect,

$$I = I_0 \cdot e^{-\frac{1}{2}(\mu_1 + \mu_2)\Delta x} = \tag{8.41}$$
$$= I_1 \cdot e^{-\frac{1}{2}(\mu_1+\mu_2)\Delta x} + I_2 \cdot e^{-\frac{1}{2}(\mu_1+\mu_2)\Delta x} \neq I_x. \tag{8.42}$$

However, in the case of the orthogonal direction, we do arrive at the average absorption,

$$I_y = I_1 \cdot e^{-\mu_1 \Delta \frac{y}{2} - \mu_2 \Delta \frac{y}{2}} + I_2 \cdot e^{-\mu_1 \Delta \frac{y}{2} - \mu_2 \Delta \frac{y}{2}} = \tag{8.43}$$
$$= (I_1 + I_2) \cdot e^{-\mu_1 \Delta \frac{y}{2} - \mu_2 \Delta \frac{y}{2}} = \tag{8.44}$$
$$= I_0 \cdot e^{-\frac{1}{2}(\mu_1 + \mu_2)\Delta y} \neq I_x, \tag{8.45}$$
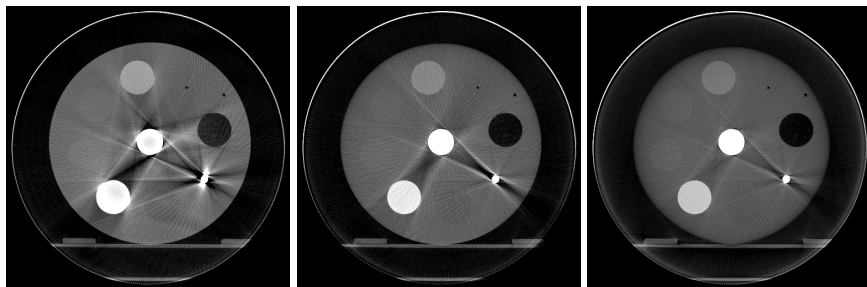
which is not equivalent to Eq. (8.40).

Figure 8.25: Reconstructions of an electron density phantom acquired at tube voltages of 50 kV (left), 80 kV (middle) and 125 kV (right). Photon starvation caused by the titanium rod at around 4 o'clock leads to pronounced streak artifacts. This effect decreases at higher tube voltages as more photons are produced and comparatively fewer of them are absorbed. Images courtesy of Jang-Hwan Choi.

8.4.3.4 Metal Artifacts

Metal artifacts are among the most common image artifacts in CT imaging. This term covers many different types of artifacts that we already discussed. There are various reasons why metal artifacts can occur. Metal causes beam hardening and scatter, which results in dark streaks between the metal objects. Additionally, its very large attenuation coefficient leads to photon starvation behind the metal object; as most photons are absorbed, only an insufficient number of them can be measured, leading to noisy projections. The noise is amplified in the reconstruction and will lead to streak artifacts in these regions, cf. Fig. 8.25.

Metal artifacts can be reduced by increasing the X-ray tube current or with automatic tube current modulation. Alternatively, there are metal artifact reduction algorithms that try to solve this problem without additional dose. Some algorithms aim to remove the metal objects in the reconstructed image and iteratively interpolate the holes in the forward-projected images.

8.4.3.5 Motion Artifacts

If motion, e. g., cardiac, respiratory, or patient motion, is present during an image acquisition, we end up with an inconsistent set of projection images. This can lead to blurring or streak artifacts in the reconstructed images. This type of artifact is especially prevalent with C-arm cone beam CT systems. Due to the slow rotation speed of the C-arm, a typical abdominal or heart scan takes approximately $4-5$ s, during which significant respiratory or cardiac motion can occur. These artifacts can be reduced by estimating
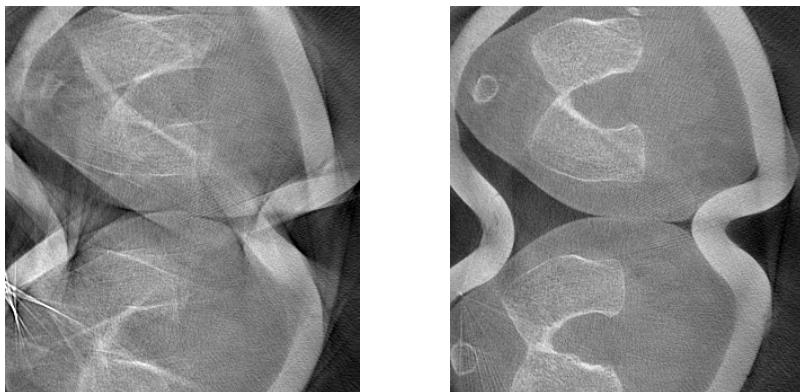
Figure 8.26: An image of human legs reconstructed from motion-corrupted projections using regular filtered back-projection (left) and with an additional marker-based correction (right). By compensating for motion during reconstruction, structures that were originally blurred due to the movement become visible and streak artifacts caused by misalignments are reduced considerably.
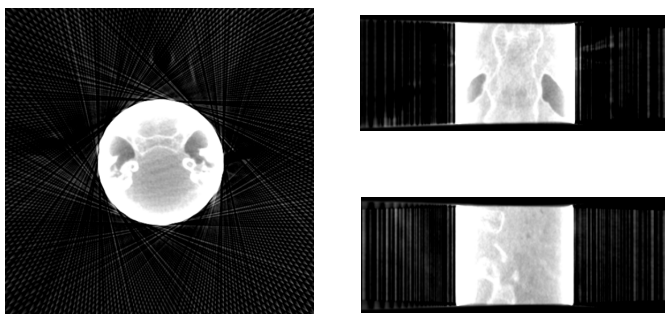

Figure 8.27: Illustration of truncation artifacts in the reconstructed slices.

the motion field and correcting it during image reconstruction. An example is shown in Fig. 8.26.

8.4.3.6 Truncation Artifacts

Truncation occurs when a scanned object is larger than the detector area or X-ray beams are intentionally collimated to a diagnostic region of interest for saving dose. Both cases will result in laterally truncated data. Due to the non-local property of the ramp filter, filtered back-projection reconstruction requires information of the whole projections for each point in the object. This

requirement, however, is not satisfied anymore if projection data are laterally truncated. Thus, a noticeable degradation of image quality manifesting as a cupping-like low-frequency artifact as well as incorrect absorption values will be observed in the reconstruction, as illustrated in Fig. 8.27.

A popular truncation correction is based on estimating the missing data using a heuristic extrapolation procedure. For instance, a symmetric mirroring extrapolation scheme could be used to reduce the truncation artifacts from objects extending outside the measured field of view. Also, the missing measurements can be approximated by integrals along rays through a 2-D water cylinder since it is able to approximately describe a human body.

8.5 X-ray Attenuation with Polychromatic Attenuation

Traditional CT measures the spatial distribution of the X-ray attenuation of an object. The X-ray attenuation of a material is energy dependent, at specific energies it is governed by the composition of the material, more precisely on its mass density and the atomic number and composition of its elements. As described in Sec. 8.3, the common measure for X-ray attenuation in medical CT is the HU. However, for a material other than water and air, the HU value depends on the system design and settings of the CT device as well as the characteristics of the complete scanned object. Fig. 8.28 exemplarily shows the energy dependent X-ray attenuation for bone and soft tissue. This dependency is caused by the non-linear attenuation characteristics of polychromatic radiation.

8.5.1 Mono- vs. Polychromatic Attenuation

When a monochromatic X-ray beam at energy E_0 passes through an object, the measured intensity I_{mono} follows Lambert-Beer's law:

$$I_{\mathrm{mono}}\left(E_0\right) = I_0\left(E_0\right) \cdot e^{-\int \mu(s,E_0)\mathrm{d}s}, \tag{8.46}$$

where $I_0\left(E_0\right)$ refers to the intensity of the incident X-ray at energy E_0, s denotes the path of the X-ray traversing the object, $\mu\left(s,E_0\right)$ denotes the spatial distribution of energy-dependent linear attenuation coefficients.

The attenuation could be obtained by rewriting Equation (8.46) as:

$$q_{\mathrm{mono}}\left(E_0\right) = -\ln\frac{I_{\mathrm{mono}}\left(E_0\right)}{I_0\left(E_0\right)} = \int \mu\left(s,E_0\right)\mathrm{d}s. \tag{8.47}$$
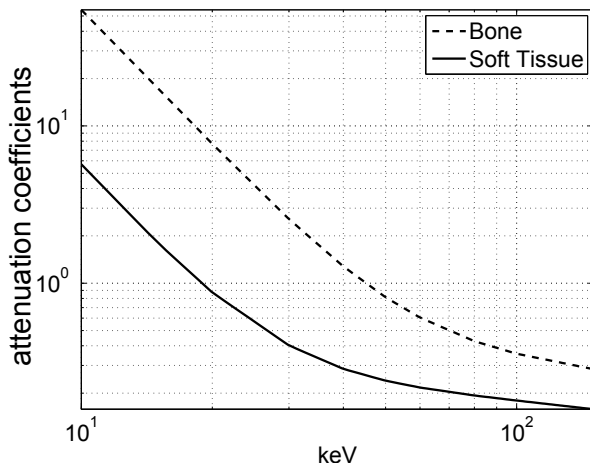
Figure 8.28: Illustration of energy dependent X-ray attenuation for bone and soft tissue.

Thus, for the monochromatic case, there is a linear relationship between the monochromatic attenuation $q_{mono}(E_0)$ and the intersection length of X-ray and the object.

However, the X-ray sources in typical clinical scanners are polychromatic sources. In addition, monochromatic measurements cannot provide real quantitative information as HUs are energy-dependent, i. e., different spectra and filters will result in different HUs. Although there exist physical ways to create monochromatic X-rays at sufficient intensity to perform X-ray CT, e. g., using a monochromator or inserting thick absorption filters to narrow the spectrum, these methods are very expensive and the usage is restricted to research experiments at few institutions. As detailed in Geek Box 8.7, in contrast to the monochromatic X-ray situation, there is no linear relationship between the polychromatic attenuation q_{poly} and the intersection length of X-ray and the object.

Fig. 8.30 depicts the relationship between the intersection length p and the attenuation q when a polychromatic X-ray beam, which is emitted at a tube voltage of 110 kV, penetrates a homogenous aluminium object, and the relationship between the intersection length p and the attenuation q when a monochromatic X-ray beam, which is emitted at the effective energy 45.56 keV of the aforementioned polychromatic X-ray beam, traverses the same object.

> **Geek Box 8.7: Polychromatic Line Integrals**
>
> As aforementioned, in practical setups, rather than a monochromatic X-ray beam, only a polychromatic X-ray beam is available. By summing the monochromatic contributions for each energy bin E in the X-ray spectrum gives ($E \in [0, E_{\max}]$):
>
> $$I_{\text{poly}}(E) = \int_0^{E_{\max}} S(E) D(E) \cdot e^{-\int \mu(s,E)\mathrm{d}s} \mathrm{d}E, \qquad (8.48)$$
>
> where $I_{\text{poly}}(E)$ denotes the measured detector signal of a polychromatic X-ray, $S(E)$ denotes the spectral energy distribution and $D(E)$ denotes the detector energy sensitivity. Fig. 8.29 shows an example of spectrum and the integral under it for the explanation of Equation (8.48).
>
> Consequently, adapting Equation (8.47) to polychromatic situation yields
>
> $$q_{\text{poly}} = -\ln \frac{I_{\text{poly}}(E)}{I_0} = -\ln \int_0^{E_{\max}} N(E) \cdot e^{-\int \mu(s,E)\mathrm{d}s} \mathrm{d}E, \quad (8.49)$$
>
> where
>
> $$N(E) = \frac{S(E) D(E)}{I_0} \qquad (8.50)$$
>
> refers to the normalized energy spectrum with the effective detected intensity (system weighting function) I_0 defined by
>
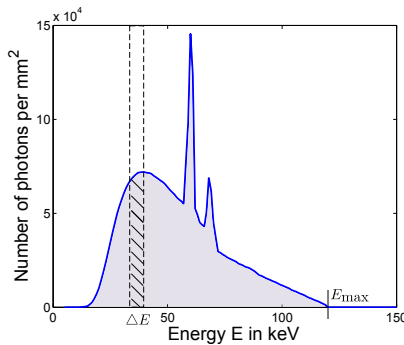> $$I_0 = \int_0^{E_{\max}} S(E') D(E') \mathrm{d}E'. \qquad (8.51)$$

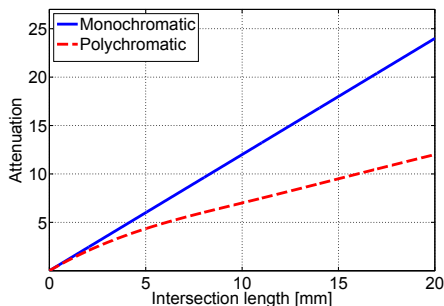**Figure 8.29:** A spectrum and the integral under it.

Figure 8.30: Relationship between intersection length and attenuation.

8.5.2 Single, Dual, and Spectral CT

As mentioned in the previous section, standard single energy CT reconstruction assumes mono-energetic radiation, however, common X-ray sources for medical CT are polychromatic. The fact is that in single energy CT, the energy information of the spectral attenuation coefficient is lost due to the measurement process. Therefore, the polychromatic characteristics of the input spectrum are neglected in single energy CT. Instead of an input spectrum $S(E)$ and a detection sensitivity $D(E)$, an effective detected intensity I_0 is measured in a calibration step, to recover the corresponding effective attenuation. In this manner, single energy CT is unable to provide quantitative information on tissue composition.

On the other hand, if spectral input data is acquired, i. e., multiple measurements with different spectral characteristics are made for each projection ray, real quantitative information on scanned anatomy becomes possible. For instance, dual energy CT measures two image sets at different energy weightings, e. g., by performing two scans with tube voltages set to 80 kVp [5] and 140 kVp, respectively. Fig. 8.31 shows an example of such dual energy CT scanner – Siemens Definition Flash (Siemens Healthineers AG, Forchheim, Germany). It employs two tube-detector pairs and produces two measurements simultaneously at different tube voltages. For spectral CT applications with dual source data, it is desirable to use two spectra with as little overlap as possible in order to ensure the maximal spectral separation between the two acquired datasets. For this task, usually the two tubes are operated at two different kVp settings. Additionally, a special filter can be used to attenuate the lower energy components in the high energy tube spectrum.

Although most spectral CT scanners require two spectral measurements, for some certain scenarios, more measurements are needed. The output quan-

---

[5] The peak acceleration voltage of X-ray tubes is usually given in kVp (kilovolt peak). An acceleration voltage of 120 kVp results in a X-ray spectrum where the individual photon energies are distributed in the range from 0 to 120 keV.

Figure 8.31: A dual energy CT scanner — Siemens Definition Flash. Image courtesy of Siemens Healthineers AG.

tities of these algorithms differ with respect to the diagnostic demands: they range from energy-normalized attenuation values over physically motivated quantities like density and effective atomic number to spatial distributions of whole attenuation spectra. The most popular current diagnostic applications are bone removal, PET/SPECT attenuation correction, lung perfusion diagnosis, or quantification of contrast agent concentrations, for instance in the myocardium.

8.5.3 Beam Hardening

When a polychromatic X-ray beam penetrates an object, photons with lower energies are easier absorbed by the object than photons with higher energies. Consequently, the average energy of an X-ray spectrum shifts toward higher energies while traversing the object, namely, the spectrum of the X-ray beam "hardens". The spectrum becomes harder with increasing intersection length of the X-ray with the object. This effect is called beam hardening.

Now we use an example to illustrate this effect in Fig. 8.32. We assume an X-ray beam is emitted at 120 kVp acceleration voltage and the material of the anode is tungsten. Now, we add additional layers of aluminum filtration to the spectrum. We can see that the corresponding effective energies E_{eff} shifts from 56.57 keV to 74.18 keV with increasing thickness of the aluminum wedge filter.

If beam hardening is not taken into consideration while doing reconstruction, the reconstructed image will be contaminated by beam hardening artifacts, which typically manifest as cupping and streak artifacts. As aforementioned, a spectrum is becoming harder when the intersection length is increasing. Hence, in reconstructed images, the inner part of the object is
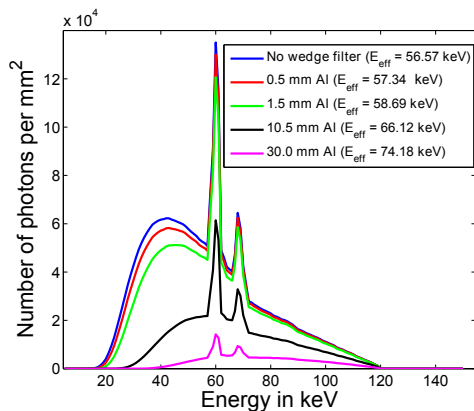
Figure 8.32: Tube spectra with different amount of filtration.
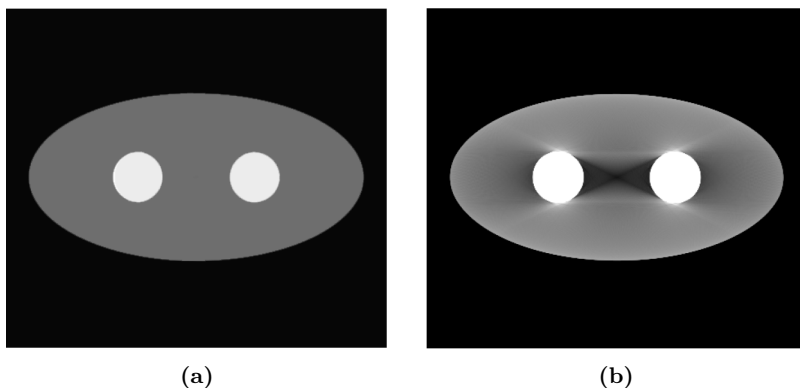

(a)

(b)

Figure 8.33: Beam hardening example: (a) A simple phantom set-up consisting of water (gray), bone (white) and air (black). (b) Reconstruction of the phantom with visible beam hardening artifacts.

darker than the outer part, and a corresponding cupping appears in the reconstruction. Streak artifacts appear as dark bands or streaks in the reconstructed image. The occurrence of such artifacts is due to the fact that X-rays, which pass through only one dense object, are less hardened than those passing through both dense objects. Fig. 8.33 shows a simulated example of an elliptical water phantom with two dense bone insets.

Various beam hardening correction algorithms exist. For a soft-tissue calibration, projection measurements through soft-tissue like materials of variable known thicknesses are performed. For these, the equivalent water thicknesses are known. A simple function is fit through the pairs of measured and expected values. This function is inverted and then used as a look-up table:

for each measured attenuation, the equivalent water thickness is looked up, which then replaces the measured attenuation. If bone-induced beam hardening is also corrected, a separation into bone and water components can be performed. For dual energy CT, special quantitative correction methods exist. These take advantage of the two measurements at different energy weightings and special properties of the attenuation functions of body tissues.

8.6 Spectral CT

8.6.1 Different Spectral CT Measurements

Spectral CT detection refers to producing multiple measurements of the same object with different spectral weightings. The spectral weighting is defined by the tube spectrum and the spectral sensitivity of the detector. In spectral detection techniques, one of these or both are changed between measurements. The spectral weightings should have as little overlap as possible. This enhances the discrimination between the spectral measurements which is beneficial for spectral CT algorithms. Usually, only two spectral measurements are created due to dose limitations and the fact that most spectral CT algorithms do not benefit from additional spectral measurements. This fact can be attributed to the specific attenuation properties of body materials in the CT energy range.

8.6.1.1 Dual KVp

The easiest method for producing spectral measurements is called Dual kVp. For this method, two subsequent CT scans are performed at different tube voltages, e. g., 80 kVp and 140 kVp; see Fig. 8.34(a) for spectra of these two voltages. As mentioned before, it is desirable to use two spectra with as little overlap as possible in order to ensure the maximal spectral separation between the two acquired datasets. To this end, usually a special filter can be used to attenuate the lower energy components in the high energy tube spectrum (140 kVp); see Fig. 8.34(b).

The main advantage is that no special equipment is needed for this method. In medical CT, this method is prone to motion artifacts as the alignment of the two datasets cannot be ensured due to patient motion in between the two scans. However, this is a valid method for evaluating spectral CT algorithms on phantom data.
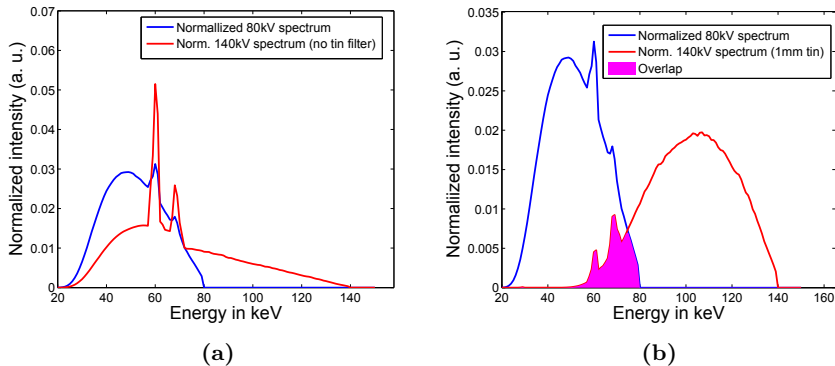
(a)

(b)

Figure 8.34: (a) Spectral measurements with different tube voltages, e. g. 80 kVp and 140 kVp. (b) A special filter was applied to high energy tube spectrum to ensure two spectra with as little overlap as possible.
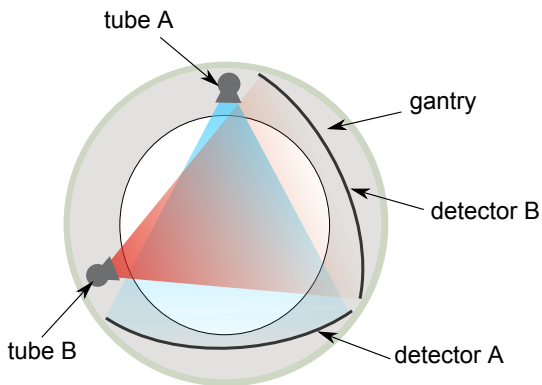

Figure 8.35: Concept of dual source CT.

8.6.1.2 KVp-switching

KVp-switching is another tube-based approach that switches the tube voltage between two readings. As read-out times are typically in the range of hundreds of micro-seconds, a special tube capable of changing the tube voltage very quickly is required. Due to dose efficiency, the tube current should also be adjusted for different kVp-settings as the attenuation properties of human body tissue are very different for different X-ray energies. The projections acquired with this approach are not perfectly aligned as the projections are interleaved. Missing projections may have to be interpolated.
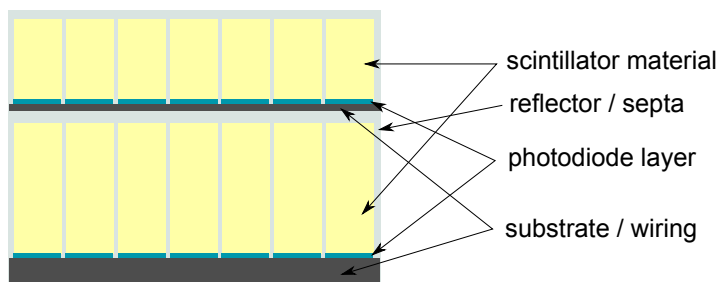
Figure 8.36: Concept of dual layer detector.

8.6.1.3 Dual Source

Dual Source CT is similar to dual kVp with the two CT scans being performed simultaneously by a special CT system. In this system, the gantry houses two tube-detector pairs A and B with a fixed angular offset (see Fig. 8.35). The two X-ray tubes are operated at different tube voltages. More recent systems offer an optional tin filter on one tube to increase spectral separation whereas the two detectors are usually identical in terms of spectral sensitivity. Most available systems, however, use differently sized detectors due to space restrictions within the gantry. So the measurements of the smaller detector offer a limited field of view (FOV). The data from the larger detector can be used to compensate for truncation artifacts in the reconstruction but Dual Energy data is only available for the smaller FOV. Since the two tube-detector pairs are operated simultaneously, scatter radiation from tube A impairs the signal of detector B and vice versa. This is a major drawback of this technology, as this property decreases signal quality and leads to an increased patient dose.

8.6.1.4 Dual Layer Detectors

This technology uses a variation of the detector spectral sensitivity to produce measurements at different energy weightings. Two scintillation detector layers are stacked upon each other and the top detector layer is a pre-filter for the lower one. This technology is also referred to as sandwich detector. Fig. 8.36 shows a possible realization of this concept. The detector efficiency is lowered, as the top layer photodiodes and wiring absorb parts of the X-rays and escape photons may enter the other layer and impair the energy separation of the layers.
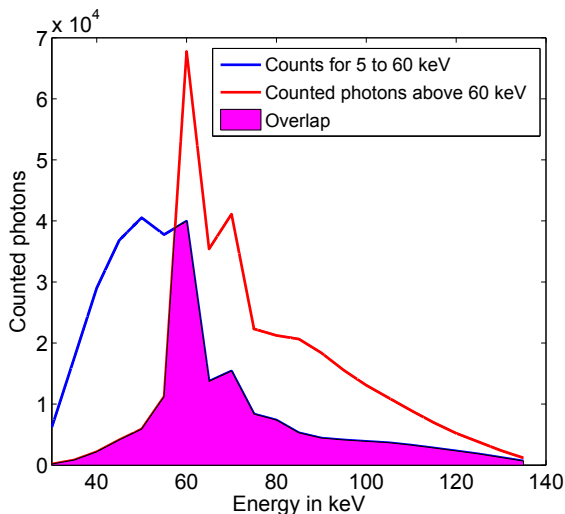
Figure 8.37: Spectral sensitivity example for a counting semiconductor detector with thresholds set to 5 keV and 60 keV. Due to several effects like cross talk, escape photons and signal pile-up, the spectral separation is reduced by a considerable overlap of the sensitivity curves.

8.6.1.5 Counting Detectors

Spectral measurements can be conducted with counting detectors by using multiple energy-thresholded photon counts. Theoretically, X-ray counting for medical CT can be performed with scintillators and semiconductor detectors. As semiconductor detectors have the advantage of being very fast and having very limited cross-talk between channels, a lot of effort has been put in evaluating the suitability of these detectors for medical CT. However, still some issues have to be resolved before this technology becomes commercially available in medical CT scanners. Counting detectors perform especially good at low X-ray flux. At high flux levels, which typically appear in medical CT, several problems arise: Signal saturation prevents distinction of individual detection events and polarization of the semiconductor material affects the signal quality. Due to physical effects, material defects, and technical limitations, the discrimination of X-ray quanta cannot be perfect. This leads to a limited spectral separation between the spectral sensitivities for each threshold signal which is dependent of the incoming X-ray flux. Fig. 8.37 shows spectral sensitivities for thresholds producing photons counts below and above 60 keV at low X-ray flux and their overlap for a 140 kVp tube spectrum.

8.6.2 Basis Material Decomposition

Fully Spectral CT approaches generally yield measures that are directly related to physical properties of the imaged object or tissue. Unlike HUs, these measures should not be system-dependent or be influenced by the surrounding object. The following section introduces basis material decomposition, which yields two or more effective basis material densities to characterize underlying material compositions.

In general, the spectral attenuation coefficient of a material can be expressed as a linear combination of M energy-dependent basis functions $f_j(E)$:

$$\mu(\mathbf{r}, E) = \sum_{j=1}^{M} c_j(\mathbf{r}) f_j(E), \tag{8.52}$$

where $c_j(\mathbf{r})$ denotes the spatially-dependent coefficients, in which $\mathbf{r} = (x, y, z)$ refers to the spatial location information.

The principle of material decomposition is based on the fact that the spectral attenuation coefficients of body materials are dominated by two effects in the energy range of medical CT: photoelectric absorption and Compton scattering, as described in Sec. 7.3 (p. 125). Since two basis materials are sufficient to express $\mu(\mathbf{r}, E)$ for body materials with very small errors, a separation of the the energy-dependent basis functions $f_j(E)$ from the spatially-dependent coefficients $c_j(\mathbf{r})$ is possible. The typical choice for basis functions in medical CT is a set of water and bone mass attenuation functions. We denote the basis functions $f_W(E)$ and $f_B(E)$, with $f_W(E)$-component corresponding to the mass attenuation coefficient of water and $f_B(E)$ to femur bone. The corresponding basis material coefficients are denoted $c_W(\mathbf{r})$ and $c_B(\mathbf{r})$. For this basis material set, Equation (8.52) reads:

$$\mu(\mathbf{r}, E) = c_W(\mathbf{r}) \cdot f_W(E) + c_B(\mathbf{r}) \cdot f_B(E). \tag{8.53}$$

Two methods to recover $c_W(\mathbf{r})$ and $c_B(\mathbf{r})$ are presented in Geek Boxes 8.8 and 8.9.

Further Reading

[1] M. Balda. *Quantitative Computed Tomography*. PhD thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, 2011.

[2] Martin Berger et al. "Marker-free motion correction in weight-bearing cone-beam CT of the knee joint". In: *Medical Physics* 43.3 (2016), pp. 1235–1248. DOI: 10.1118/1.4941012.

[3] T. M. Buzug. *Computed Tomography: From Photon Statistics to Modern Cone-Beam CT*. Springer, 2008. ISBN: 9783540394082.

[4] Frank Dennerlein and Andreas Maier. "Approximate truncation robust computed tomography - ATRACT". In: *Physics in Medicine and Bioloagy* 58.17 (2013), pp. 6133–6148. DOI: 10.1088/0031-9155/58/17/6133.

> **Geek Box 8.8: Projection-Based Basis Material Decomposition**
>
> Inserting Equation (8.53) into the line integral of the spectral attenuation law yields:
>
> $$\int \mu(\mathbf{r}, E) \, \mathrm{d}s = f_W(E) \int c_W(\mathbf{r}) \, \mathrm{d}s + f_B(E) \int c_B(\mathbf{r}) \, \mathrm{d}s. \quad (8.54)$$
>
> We denote the line integral over the water coefficients $A_W = \int c_W(\mathbf{r}) \, \mathrm{d}s$. The integral A_B over the bone coefficients is defined along the same line.
>
> Conducting a dual energy measurement at two energy weightings $S_1(E) D_1(E)$ and $S_2(E) D_2(E)$ gives the following system of nonlinear equations:
>
> $$I_1 = \int_0^\infty S_1(E) D_1(E) e^{-f_W(E) A_W - f_B(E) A_B} \, \mathrm{d}E \quad (8.55)$$
>
> $$I_2 = \int_0^\infty S_2(E) D_2(E) e^{-f_W(E) A_W - f_B(E) A_B} \, \mathrm{d}E \quad (8.56)$$
>
> This system has to be solved for A_W and A_B, which is the scope of current research. Then, the basis material coefficients $c_W(\mathbf{r})$ and $c_B(\mathbf{r})$ can be recovered from A_W and A_B with a plain inverse Radon transform as used for standard CT reconstruction.
>
> A general drawback of projection data-based methods is the requirement of perfectly matched projections and not all dual-energy detection techniques are able to measure line integrals at exactly the same positions for each spectrum.

[5] Joachim Hornegger, Andreas Maier, and Markus Kowarschik. "CT Image Reconstruction Basics". In: *MR and CT Perfusion and Pharmacokinetic Imaging: Clinical Applications and Theoretical Principles.* Ed. by Roland Bammer. 1st ed. Alphen aan den Rijn, Netherlands, 2016, pp. 01–09. ISBN: 9781451147155.

[6] W. A. Kalender. *Computed Tomography: Fundamentals, System Technology, Image Quality, Applications.* Wiley, 2011. ISBN: 9783895786440.

[7] B. Keck. *High Performance Iterative X-Ray CT with Application in 3-D Mammography and Interventional C-arm Imaging Systems.* PhD thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, 2014.

[8] Yanye Lu et al. "Material Decomposition Using Ensemble Learning for Spectral X-ray Imaging". In: *IEEE Transactions on Radiation and Plasma Medical Sciences* (2018). to appear.

[9] Andreas Maier and Rebecca Fahrig. "GPU Denoising for Computed Tomography". In: *Graphics Processing Unit-Based High Performance Computing in Radiation Therapy.* Ed. by Xun Jia and Jiang Steve. 1st ed. Vol. 1. 2015. ISBN: 978-1-4822-4478-6. DOI: 10.1201/b18968-9.

Geek Box 8.9: Image-Based Basis Material Decomposition

Image-based basis material decomposition avoids the matched projection problem by performing the decomposition in the reconstruction domain. For this purpose, the reconstructed attenuation values need to correspond to a constant, known energy weighting throughout the CT volume. For data measured with a polychromatic source, this can be achieved by a quantitative beam hardening correction. It homogenizes the energy weighting throughout the reconstructed CT volume. The homogenized energy weighting is denoted $\tilde{w}_i(E)$. Here, i numbers the N_i spectral measurements. As for projection data-based basis material decomposition, multiple measurements at different energy weightings are required. The relation between spectral attenuation coefficient and measured attenuation coefficient after beam hardening correction $\tilde{\mu}_i(\mathbf{r})$ is defined by the energy weighting:

$$\tilde{\mu}_i(\mathbf{r}) = \int_0^\infty \tilde{w}_i(E)\,\mu(\mathbf{r}, E)\,\mathrm{d}E \tag{8.57}$$

With two basis material decomposition of $\mu(E, \mathbf{r})$ (8.53), we get

$$\tilde{\mu}_i(\mathbf{r}) = \int_0^\infty \tilde{w}_i(E)\,(c_\mathrm{W}(\mathbf{r})\,f_\mathrm{W}(E) + c_\mathrm{B}(\mathbf{r})\,f_\mathrm{B}(E))\,\mathrm{d}E \tag{8.58}$$

Here, we can exchange summation and integration,

$$\tilde{\mu}_i(\mathbf{r}) = c_\mathrm{W}(\mathbf{r})\int_0^\infty \tilde{w}_i(E)\,f_\mathrm{W}(E)\,\mathrm{d}E + c_\mathrm{B}(\mathbf{r})\int_0^\infty \tilde{w}_i(E)\,f_\mathrm{B}(E)\,\mathrm{d}E. \tag{8.59}$$

The complete basis material decomposition with all measurements then leads to the following linear system of equations:

$$\tilde{\mu}(\mathbf{r}) = \tilde{\mathbf{K}} \cdot \mathbf{c}(\mathbf{r}) \tag{8.60}$$

with $\tilde{\mu}(r) = (\tilde{\mu}_1(\mathbf{r}), \tilde{\mu}_2(\mathbf{r}), ..., \tilde{\mu}_{N_i}(\mathbf{r}))^\top$, $\mathbf{c}(\mathbf{r}) = (c_\mathrm{W}(\mathbf{r}), c_\mathrm{B}(\mathbf{r}))^\top$ and

$$\tilde{\mathbf{K}} = \left[\, \int_0^\infty \tilde{w}_i(E)\,f_\mathrm{W}(E)\,\mathrm{d}E \quad \int_0^\infty \tilde{w}_i(E)\,f_\mathrm{B}(E)\,\mathrm{d}E \,\right].$$

The quantitative accuracy of the image-based basis material decomposition approach depends on the accuracy of the beam-hardening correction and the image quality of the resulting basis-material images is reduced since the solution of Equation (8.60) is very sensitive to noise in the input data. So far, more advanced image-based material decomposition methods have been developed to overcome these drawbacks.

[10] Andreas Maier et al. "Fast Simulation of X-ray Projections of Spline-based Surfaces using an Append Buffer". In: *Physics in Medicine and Biology* 57.19 (2012), pp. 6193–6210.

[11] A. Maier et al. "CONRAD - A software framework for cone-beam imaging in radiology". In: *Medical Physics* 40.11 (2013), pp. 111914-1–111914-8.

[12] M. Manhart. *Dynamic Interventional Perfusion Imaging: Reconstruction Algorithms and Clinical Evaluation.* PhD thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, 2014.

[13] K. Müller. *3-D Imaging of the Heart Chambers with C-arm CT*. PhD thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, 2014.

[14] Kerstin Müller et al. "Image artefact propagation in motion estimation and reconstruction in interventional cardiac C-arm CT". In: *Physics in Medicine and Biology* 59.12 (2014), pp. 3121–3138.

[15] Haibo Wu et al. "Spatial-temporal Total Variation Regularization (STTVR) for 4D-CT Reconstruction". In: *Proceedings of SPIE Medical Imaging 2012*. Ed. by Norbert J. Pelc. Town & Country Resort and Convention Center, San Diego, CA, USA, 2012.

[16] Yan Xia et al. "Towards Clinical Application of a Laplace Operator-based Region of Interest Reconstruction Algorithm in C-arm CT". In: *IEEE Trans Med Imaging* 33/2014.3 (2014), pp. 593–606. DOI: 10.1109/TMI.2013.2291622.

[17] Y. Xia et al. "Patient-bounded extrapolation using low-dose priors for volume-of-interest imaging in C-arm CT". In: *Medical Physics* 42.4 (2015), pp. 1787–1796.

[18] Z. Yu et al. "Line plus arc source trajectories and their R-line coverage for long-object cone-beam imaging with a C-arm system". In: *Physics in Medicine and Biology* 56.12 (2011), pp. 3447–3471.

[19] G. Zeng. *Medical Image Reconstruction: A Conceptual Tutorial.* Springer, 2010. ISBN: 9783642053689.