Spectral imaging and tomographic reconstruction methods for industrial applications

Zeegers, M.T.

Citation

Zeegers, M. T. (2023, May 31). *Spectral imaging and tomographic reconstruction methods for industrial applications*. Retrieved from https://hdl.handle.net/1887/3619550

1

Introduction

For well over a century, X-ray imaging has been successfully employed as a non-destructive method to gain insight into the internal structures of various objects. Some well-known applications include baggage scanning at airports and medical diagnosing in hospitals. One particularly vital application area is the food processing chain, where quality preservation and safety checks are the order of the day. Even though most food inspection is surface-based, X-ray imaging has proven to be an essential complement to finding significant anomalies within a food product. Still, a more detailed understanding of the interior of an object can be achieved using Computed Tomography (CT), in which X-ray projections are collected around the imaged object and used to compute a 3D reconstruction of it. These CT acquisitions and algorithms generally take more time than conventional X-ray imaging, which can pose a problem at industrial high-throughput production lines. These conflicting goals of accuracy and processing speed require a trade-off to be made, and new methods and algorithms are needed to optimize this trade-off.

Towards achieving the optimized trade-off, spectral imaging and machine learning are important directions to respectively obtain rich data and achieve faster and better results. Recently, the field of spectral imaging, where the energies within the X-ray beams are taken into account, has been getting more attention. A major driving force is the development of spectral photon-counting detectors, which are able to register the energy properties of each individual photon in the incoming beam independently and can therefore disentangle energetic information into multiple images at different wavelengths. This is a significant information gain compared to conventional X-ray imaging, and can therefore be used to improve X-ray image analysis and to create more accurate CT reconstructions with the same acquisition time. These detectors and reconstruction algorithms are continuously in development. However, the lack of suitable spectral image processing methods and reconstruction algorithms is hampering the application of spectral X-ray imaging in industry in general and food processing in particular. Additionally, recent developments in machine learning open up possibilities for rapid and accurate X-ray image analysis. In particular, deep learning enables rapid recognition of complicated patterns in data and can therefore aid in image processing tasks. The potential of deep learning has yet to be explored for the examination of X-ray images of natural products in industrial settings.

In this dissertation, several research contributions are put forward for spectral X-ray CT and the incorporation of deep learning in (spectral) image processing routines to improve reconstruction accuracy and performance of image processing tasks. The main contribution of this work is the proposal and detailed analysis of a new deep learning workflow for imaging tasks in which the scientific contributions in all chapters are incorporated. The method can be readily implemented in industrial setups and has a particular focus on food processing. This introductory chapter will first provide the background of product inspection in industry, which will be the common thread for all subsequent motivations. After this, X-ray imaging and its application to product inspection are covered. Next, the basic concepts of computed tomography are introduced. Thereafter, an outline of the current state of the art of spectral X-ray imaging and spectral CT is given. Along with this, several machine learning concepts are introduced, which are used in later sections. This chapter is concluded with an overview of all sections in this dissertation, supplemented by the relevant research questions for each non-preliminary chapter.

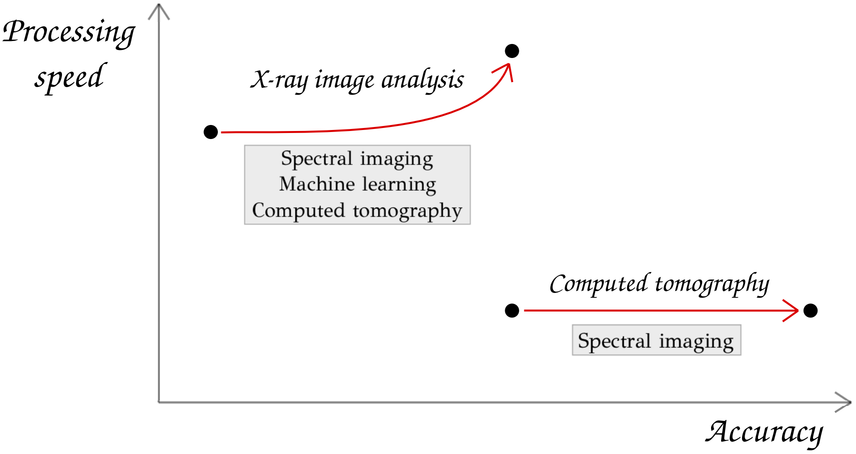


Figure 1.1: Schematic view on the goals in this dissertation. The pursued contributions by the research are indicated by the red arrows. Spectral imaging, machine learning and tomography methods are used to increase the accuracy and speed (i.e. image throughput) of algorithms for tasks on X-ray images. In addition, spectral imaging is used to increase the reconstruction accuracy in computed tomography.

1.1 Product inspection and foreign objects

With the increasing reliance of society on automatic food processing in factories, accurate quality control of food is absolutely essential [305, 327]. Regularly, food products are recalled from supermarkets or other food distributors because of *contamination*: unintended and possibly undetected alterations to the products, which makes them unsuitable for consumption. The contamination and the subsequent recalls waste time, money, effort and energy. On top of that, consumers and product sellers may lose trust in the product manufacturer. It goes without saying that all sorts of contamination should therefore be avoided as much as possible. In general, there are three different distinguishable types of product contamination [89, 287]. Chemical contaminations can appear as pesticides, wrong food additives or inappropriate cleaning substances. Biological contamination encompasses bacterial, fungal, virus and parasite contamination. Both can deteriorate food quality but can also lead to severe food poisoning. In this dissertation, we focus mainly on the third category of *physical contamination*, where food contains items that are not supposed to be there. The existence of this third category is the primary motivator for the realization of the methods presented in this dissertation.

Physical contamination is characterized by the presence of *foreign objects*, also known as foreign bodies. These foreign objects are defined as objects that are not supposed to be present in a product and may harm the end user upon consuming or using it if its presence remains unknown. Foreign objects come in many shapes and conditions. Usually, a distinction is made between intrinsic foreign objects, which are already present in the object but not properly removed (e.g. bones, shells, fruit stalks or pits), and extrinsic foreign objects, which are objects that end up in the product during the food processing stages (e.g. stones, plastic and insects) [76]. Foreign objects of the latter category can be introduced during transport, improperly sanitized factory processing environments, sanitizing processes, automatic or manual deformation processes, and packaging stages. Typical example objects are stones, wood, insects, plastics, hair, pieces of (fallen) gloves, rubber, glass and metal particles. In addition to the harm that can be done after processing, hard foreign objects can also damage the processing machines. Standard control measures such as sieves, filters and magnets can be used to mitigate the problem, but these have technical and operational limitations to minimizing the risks [68]. Furthermore, these methods are automatic and do not necessarily enable the notion of foreign objects being present, as for instance sieves only remove larger foreign objects.

To address the problem of foreign objects, *food inspection methods* are continuously being developed and employed in the food processing industry. Such product inspection methods are in place as well where non-food items are checked for product defects, but also in places such as airports for luggage inspections. Ideally, these methods can eventually be used for assessing products on more advanced quality aspects such as tenderness, ripeness, bacteria presence and other biological properties as well. Nevertheless, all applications have two common goals: the methods should be fast and accurate. These two conflicting goals ask for the continuous development of new inspection methods.

Until the end of the previous millennium, optical sensing technologies were the only well-established methods [106] for foreign object detection. These methods work well when the product is coarse and loose or when the defects are located on the surface, and when foreign objects can be detected based on differences in shapes or colors. For foreign objects that are hidden in a product, the only techniques that are not surface-based and have widespread commercial availability at low cost were metal detection methods. Because these methods can only detect metals, other *noninvasive inspection* methods (also known as nondestructive methods), in which an object can be inspected without damaging it, were highly desired.

Nowadays, many approaches to nondestructive internal product inspection have been developed. The development and availability are due to scientific progress, decreasing machinery costs and the development of sophisticated signal and image processing techniques over the years. Most of the current methods are based on the interaction between the electromagnetic radiation that penetrates the product and the foreign object [106]. As Figure 1.2 indicates, some state-of-the-art approaches include [117, 194, 200, 305] magnetic resonance imaging (in the radiofrequency range), microwave imaging [281], terahertz imaging [324], thermal infrared imaging, optical hyperspectral imaging [85, 89], X-ray imaging and gamma-ray inspection.

The optimal approach and the chosen type of radiation from the electromagnetic spectrum are usually highly problem-specific and are chosen according to the material properties and inspection goal. Important is the *penetration depth* of the active electromagnetic radiation and its interaction that enables distinction of the materials of interest. Radiowaves have high penetration depth but come with low image acquisition speed. Additionally, they need powerful and costly equipment to generate a strong magnetic field, and are prone to motion artefacts. Microwaves only penetrate by a few centimeters. For other electromagnetic radiation types, the penetration depth is typically very limited. However, it increases with the energy of the wave, with X-rays and gamma rays providing sufficient power to fully penetrate through objects. Both gamma rays and X-rays are ionizing and therefore harmful, but - despite common belief - the effect of X-rays on food products is extremely limited, and radiated food is therefore safe to consume. The high penetration power of gamma rays is mostly used to inspect metallic and welded structures. For softer materials it is less suitable, as differences between materials cannot be observed as well as with X-rays. Another method based on acoustic waves is ultrasound and is, together with X-rays, the most widely used inspection method. However, ultrasound is only useful for imaging soft materials, such as soft tissue. It is less suitable when air or other materials with significant differences in acoustic impedance are involved in the object. On the other hand, X-rays can penetrate all common packaging materials [68]. Compared to widespread simple metal detectors,

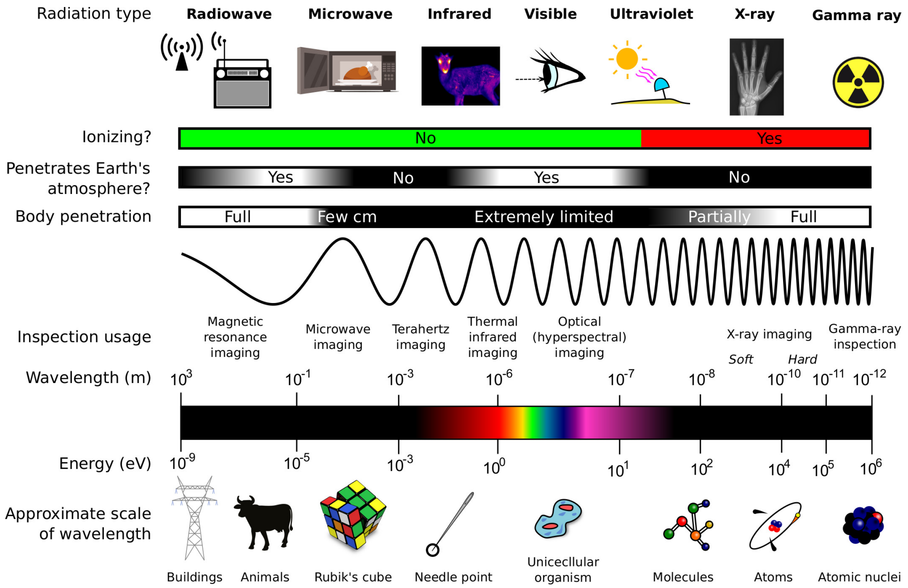


Figure 1.2: Overview of inspection methods based on interaction with the electromagnetic spectrum with radiation types, wavelengths, generic applications, and penetration depth [55, 291].

Method	Speed	Costs	Complexity	Data	Material identification
Standard X-ray imaging	Fast	Low	Low	Small	Low discrimination
Standard X-ray tomography	Slow	Moderately high	Moderate	Moderate	Moderately low discrimination
Spectral X-ray imaging	Fast	High	High	Moderate	Moderately high discrimination
Spectral X-ray tomography	Slow	High	High	Very large	High discrimination

Table 1.1: Different X-ray imaging methods and their properties.

X-ray machinery can expose a wider range of materials including stones, glass, bones, and plastics.

For this reason, this dissertation will focus mostly on X-ray imaging for industrial imaging. State-of-the-art X-ray imaging comes in many modalities. In Section 1.2, *standard X-ray imaging* is explained, which is fast and inexpensive, but does not necessarily yield 2D images with sufficient material differentiation for further analysis. Section 1.3 discusses *X-ray computed tomography*, which by repeatedly capturing radiographs under different angles and reconstructing a 3D volume enables better analysis of an object. Section 1.4 introduces *spectral X-ray imaging*, which generates radiographs at different X-ray energies for better material discrimination than with standard X-ray imaging. Section 1.5 encompasses *spectral X-ray computed tomography*, which combines spectral X-ray imaging and computed tomography for more advanced material discrimination in reconstructed 3D volumes. Table 1.1 gives an overview of all techniques and their relative characteristics.

1.2 Radiography with X-rays

Just as famous as the discovery of X-ray radiation by Wilhelm Conrad Röntgen in 1895 are the subsequently made X-ray images of a hand with a ring, of which the first one Röntgen made revealing the bone structure of his wife's hand. To this day, these images show the enormous potential to visualize the inner structure of objects noninvasively. Therefore, inspection of food is just one of the many applications of X-ray imaging which is, together with gamma-ray imaging, also known as *radiography.* One year after Röntgen's discovery, the first X-ray devices were already in use for clinical observations, and the medical applications of radiography to inspect conditions of various parts of the body (such as possibly broken bones and dental conditions) have been indispensable ever since. With further industrial advances during the 20th century, radiography for baggage screening and cargo inspection, quality inspection of castings and welds and quality control of parts of products such as cars [192] have become crucial as well. Mechanized implementations of these enable *automated X-ray inspection* (AXI). Despite the possibly harmful ionizing radiation of X-rays, radiography has become widely used because it is relatively cheap, fast, easily accessible and has a relatively high penetration depth compared to other inspection methods for such applications [200].

For the electromagnetic spectrum, a single quantum is a *photon.* Depending on the application, photons are described using either properties of particles or waves. Each photon contains a certain amount of *energy* E, measured in electronvolt (eV). The energy of a photon depends on its *frequency* f (in s^{-1}), through

$$E = hf.$$

where h is Planck's constant ($h \approx 4.136 \cdot 10^{-15}$ eV $\cdot\, s$). Photons and their associated energy are usually expressed either in frequency f or in *wavelength* λ, which is

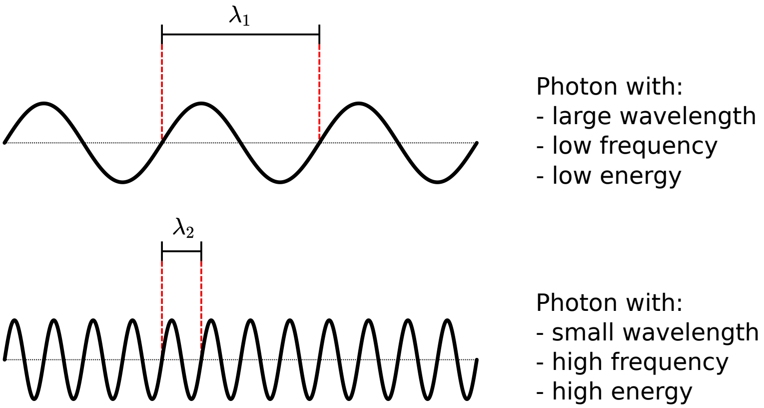


Figure 1.3: Photon model and the comparison of the (equivalent) characteristics between two photons with large wavelength λ_1 and small wavelength λ_2.

inversely proportional to the frequency f through the speed of light c in a vacuum (in $m \cdot s^{-1}$):

$$\lambda = \frac{c}{f}.$$

Electromagnetic radiation is a stream of photons, each carrying its own energy. As indicated in Figure 1.2, the wavelength (or, equivalently, energy or frequency, see Figure 1.3) of the photons determines the type of electromagnetic radiation involved. X-ray beams are made up of photons with wavelengths between 0.01 nm and 10 nm (approximately equivalent to between 100 eV and 100 keV). Photons with high energy - and therefore with small wavelengths - are able to penetrate further into matter than those with low energy. Therefore, a distinction is often made between *soft* and *hard* X-rays. The former consists of photons with maximum energies below 10 keV (above 0.1 nm wavelength) typically used for medical applications, and the latter of photons with maximum energies above 10 keV, which are more prevalent in industrial imaging.

To produce X-rays, an X-ray source consisting of a vacuum tube with a cathode and an anode is typically used (see Figure 1.4). Electrons are emitted from the cathode using the thermionic effect by heating the filament of the cathode to overcome their binding energy, and are accelerated towards an anode due to a voltage applied between the two [51, 55, 192, 242]. The acceleration increases with the *peak voltage* applied (usually indicated in kVp), while the number of released electrons from the filament increases with the applied *current*. The target material of the anode (usually tungsten) is bombarded with electrons, whose kinetic energy is mostly converted to heat, but a small fraction is converted into X-radiation via various processes when the electrons interact with the atoms of the anode material.

In the first of those interaction types, the electron is decelerated due to deflection around the atomic nucleus. The deceleration then converts the kinetic energy of the electron into X-ray radiation, known as braking radiation or *bremsstrahlung*. The energy of the radiation is directly proportional to the amount of deceleration. It is also possible that an incident electron hits an electron in the shell of an atom and kicks it out. When this happens, an electron from an outer shell will take its place, and the atom emits X-rays in the process. The involved shells of the atom characterize the energy of the emitted radiation, and this *characteristic radiation* is therefore highly specific to the material type of the anode. This emission process is known as *X-ray fluorescence* (XRF).

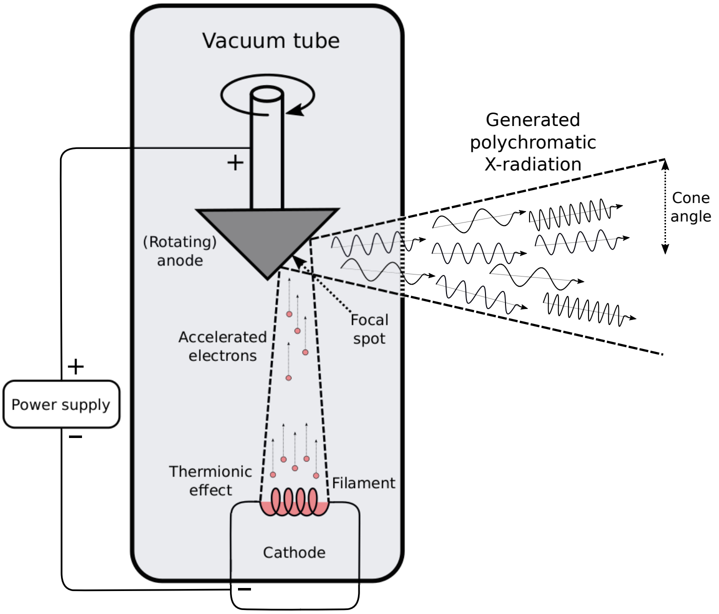
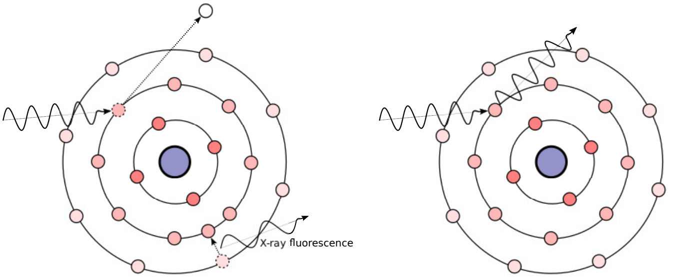
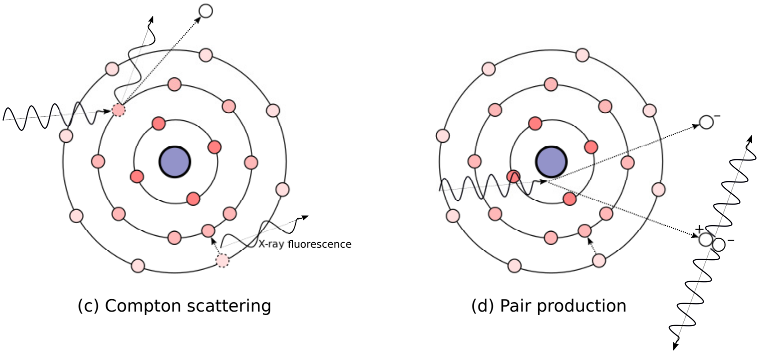


Figure 1.4: Schematic representation of an X-ray source. Electrons are emitted from the filament in the cathode by applying power to heat it up, after which by interaction with the anode a fraction is converted to X-radiation.

The location from which the radiation is emitted in the tube is the *focal spot* of the resulting X-ray beam. Ideally, the focal spot should be a point. In reality, this is an area of a specific size. A small focal spot is better than a large focal spot, as this will reduce the number of blurry edges when forming an X-ray image. On the other hand, a small focal spot will focus the heat in a smaller area and could cause melting of the anode material. For this reason, a rotating anode can be used to distribute the heat more evenly [192]. The generated X-ray beam is *cone*-shaped, and is defined by its *cone angle*. The intensity of the beam is usually referred to as the *flux* of the beam.



(a) Photoelectric absorption (b) Rayleigh scattering



(c) Compton scattering (d) Pair production

Figure 1.5: Schematic view of X-ray interaction with atoms: (**a**) photoelectric absorption with X-ray fluorescence, (**b**) Rayleigh scattering, (**c**) Compton scattering with X-ray fluorescence and (**d**) pair production.

Alternatively, for much higher-quality beams there are *synchrotrons*. In these ring-shaped facilities, (injected) electrons move at high speeds in a closed loop. Powerful magnets are used to bend the directions of the electrons, at each of which intense radiation is emitted. Typically, synchrotrons are mostly used for scientific research. Researchers acquire beam-time at one of the suitable beamlines (with specific beam characteristics), at which obtaining beam-time is a challenging and involved process. The beams that result from this generation process are typically high-intensity *parallel* beams.

Different types of interactions can occur when a photon in an X-ray beam travels through an object (see Figure 1.5). If a photon has low energy, the dominating interactions are *photoelectric absorption* and *Rayleigh scattering*. In photoelectric absorption, a photon collides with an electron in the shell of an atom, and transfers all of its energy to this electron. If this energy is higher than the binding energy of the electron, the photon is completely attenuated and the electron is ejected from the atomic shell. Due to this unstable ionized state, an electron from a higher shell will take its place and the energy surplus will be emitted as characteristic radiation.

This radiation usually has low energy, but is unique to the material's *atomic number* (usually indicated by Z). The second effect of Rayleigh scattering occurs when the photon energy is not sufficient to eject an electron from its shell. In this case, the direction of the photon is changed without changing its energy. At higher photon energies, *Compton scattering* is the most contributing factor. In this process, the photon energies are much higher than the binding energies of the electrons in the shells. Therefore, an electron is emitted from the shell and characteristic radiation will occur (as with photoelectric absorption). In this process, the photon only loses part of its energy, and will continue its path through the material, possibly starting another interaction process. At the highest photon energies, a fourth interaction called *pair production* will take effect (and dominates for photons with gamma ray energies). In this case, the high-energy photon will create an electron-positron pair in the atom, each travelling further in different directions. When the positron meets another electron, two photons with high energy (gamma rays) are formed in opposite directions. This effect may occur more often in industrial imaging, where high-energy radiation is involved for materials with high atomic numbers.

Given a material, the combined effect of its absorption by the aforementioned mechanisms can be collapsed into an *attenuation rate* μ (and here also includes the material mass density). The absorption of X-rays by a homogeneous material with an attenuation rate μ is given by the *Lambert-Beer law*:

$$I = I_0 \exp\left(-\mu \Delta x\right).$$

Here, I_0 is the *intensity* (proportional to the number of photons) of the incident X-ray beam and Δx is the thickness of the object. An illustration of the Lambert-Beer law is given in Figure 1.6a. The Lambert-Beer law states that the intensity of the beam decays exponentially with the thickness of the material (and the attenuation coefficient). Therefore, a higher incident flux cannot easily compensate for an increase in thickness. If the object being imaged is not homogeneous, the attenuation coefficient becomes dependent on the positions over which the ray passes through the object. Therefore, in this case the Lambert-Beer law reads

$$I(s) = I(0) \exp\left(-\int_0^s \mu(x) dx\right),$$

where $\mu(x)$ is the attenuation coefficient at location x, and s is the length of the ray through the object.

The Lambert-Beer law theoretically quantifies the intensity of the beam after it has traversed the object. To measure this quantity, an X-ray detector is needed. Usually, these detectors do not measure the X-rays directly, but rather contain material that converts the X-rays to other detectable radiation. Nowadays, *flat-panel scintillator* detectors are the most commonly used detector type, where the scintillator converts the X-rays to visible light, which in turn can be detected by a charge-coupled device (CCD) that converts the visible light through photodiodes

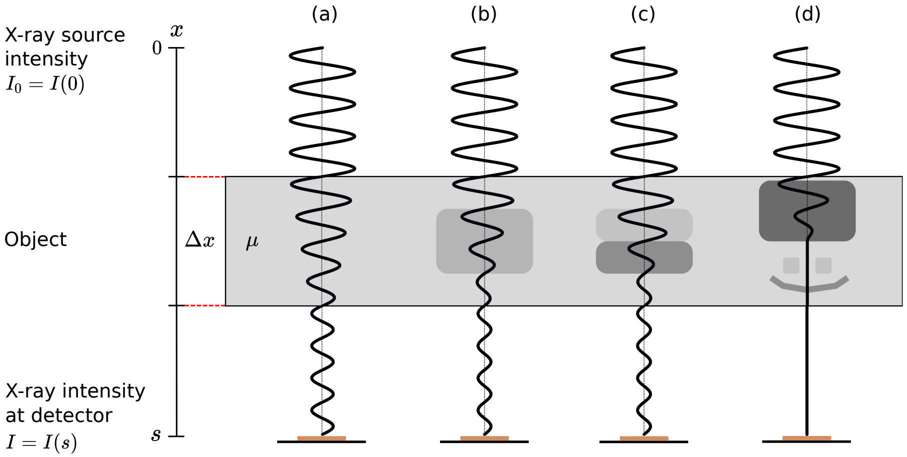


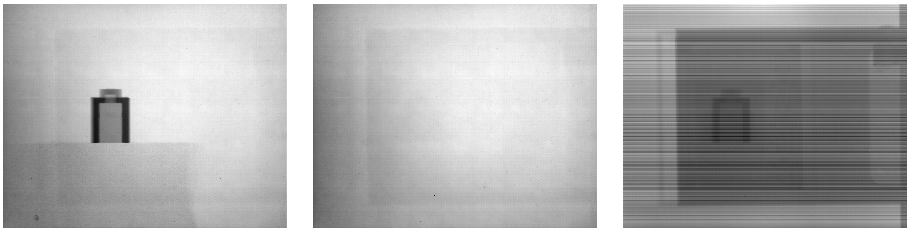
Figure 1.6: Illustration of the Lambert-Beer law: (**a**) ray passing through a homogeneous object with thickness Δx and attenuation μ, (**b**) ray passing through an inhomogeneous object with additional absorption, (**c**) ray passing through a different inhomogeneous object resulting in the same intensity at the detector, (**d**) ray passing through a highly absorbing object, leading to photon starvation and removing any interesting information about the structure of the object.

into electronic signals. The *detector efficiency*, the rate of incoming photons that are detected, depends on the thickness and the material of the scintillator.

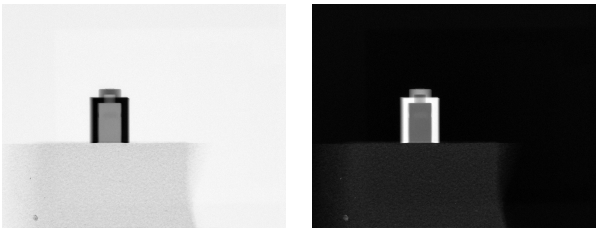
According to the Lambert-Beer law, the detector yields an image that approximately gives the intensity of the incident X-rays. For further analysis, the image is first normalized by applying a *flatfield correction* (with an X-ray image without the object) and a *darkfield correction* (with an image with photon count offsets when the X-ray is not in use) to obtain the transmission image. After this, the image is linearized by applying a negative logarithm to obtain the absorption image. The correction process is visualized in Figure 1.7. The result for an X-ray traversing through the object is approximately:

$$-\ln \frac{I(s)}{I(0)} = \int_0^s \mu(x)dx. \tag{1.1}$$

The images obtained by X-ray radiography are 2D projections, which can be analyzed and processed, such as decision-making on the presence of a foreign object in a 3D food product [191]. However, in the resulting *radiograph* (or *X-ray image*), information about the third dimension (perpendicular to the detector plane) is lost. There are situations where prior information can be exploited well enough to detect desired features, such as the presence of bones. In other situations, absorption of a certain material and thickness cannot be distinguished from absorption from another combination of material and thickness. This is the *superposition problem* (Fig. 1.6b and c). On top of this, when highly absorbing objects are involved in the radiograph acquisition, there is the issue of *photon starvation* (Fig. 1.6d). The



(**a**) X-ray radiograph (**b**) Flatfield radiograph (**c**) Darkfield image



(**d**) Corrected radiograph (**e**) Log-corrected radiograph

Figure 1.7: Demonstration of flatfield correction on radiographs of a small brick. (**a**) The radiograph obtained with the object in the field of view. (**b**) Flatfield radiograph without the object in the field of view. (**c**) Darkfield image without the X-ray tube. Note that the recorded values are in reality much lower compared to the values in radiographs (a) and (b). (**d**) Corrected radiograph using the flatfield and darkfield images. (**e**) Log-corrected radiograph resulting in the typical radiograph representation. Note that the acquired images contain some sort of afterglow resulting from X-ray exposure during a previous acquisition, see the panel afterglow in (a-c). The flatfield correction in (d) removes the panel.

number of emitted and detected photons each follow a *Poisson distribution*. The average number of emitted photons is typically sufficiently high to approximate it by a Gaussian distribution and yields a smooth flatfield image. However, this is not the case when a highly-absorbing object prevents most photons from reaching the detector. In this case, the intensities on the detector pixels may vary heavily, and are likely to lead to high *noise* in the radiograph such that desired features are difficult or impossible to extract. Altering the X-ray source properties, such as the current, the peak voltage, or the *exposure time* may alleviate this problem. However, in many cases, these settings cannot be changed because of time, energy and dosage constraints, or the absorption power of the object is unlikely to be overcome.

1.3 Tomography

To gain better understanding of the inner structure of an object, the procedure of *tomography* noninvasively generates images containing the inner structure of an object. There are many different types of tomography with different material penetration and interaction mechanisms, such as magnetic resonance imaging

(MRI) and ultrasound tomography. In this dissertation, whenever tomography is discussed, we will focus on X-ray transmission *computed tomography* (CT). Since the proof of concept of CT scanning in 1969 by Hounsfield and the subsequent first clinical examination with CT in 1972, the methodology has been hugely important in getting good insight into the three-dimensional condition of (parts of) the human body. Additionally, it finds many applications in scientific research, geological studies [97], cultural heritage [44, 151, 201, 255] and (airport) security [193, 204]. In industry, important applications are materials characterization, nondestructive testing and metrology [55], and tomography also finds more and more applications in food processing (such as quality control of pears [174]).

In X-ray CT, an object is exposed to X-rays from a source, while a detector captures the unabsorbed photons in a resulting radiograph. This process is repeated under different angles, by rotating the object with respect to the source and detector (Fig. 1.8). In some instances, such as with CT scanners in hospitals where patients are imaged, it is more practical to rotate the source and detector rather than the object itself (Fig. 1.9). The precise locations of the source, detector and the object in the direction of the beam are recorded during the scan and are part of the complete *scanning geometry*, which can be circular but also helical or spiral when the object is large.

In a more formal context, the object of interest is modelled by the function $f : \mathbb{R}^n \to \mathbb{R}$ which maps the location to a scalar value that represents the attenuation in that location. In case of a two-dimensional object, we have $n = 2$. The *Radon transform* $\mathcal{R}$ maps the object function f to its projection function $P : [0, \pi] \times \mathbb{R} \to \mathbb{R}$

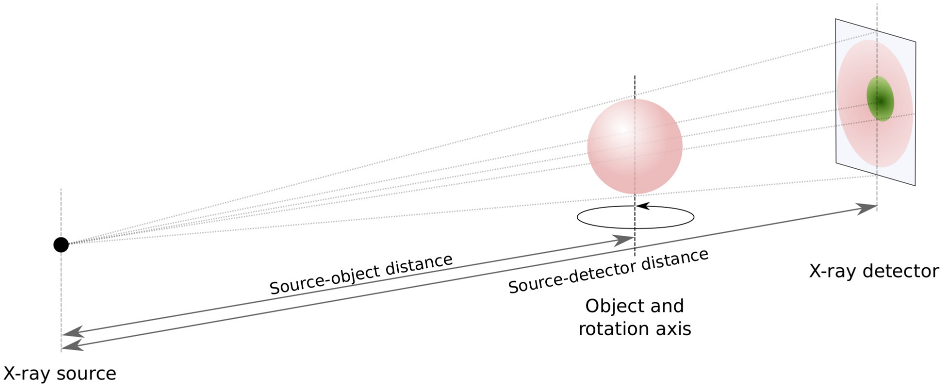


Figure 1.8: The process of obtaining multiple X-ray projections by rotating the source and the detector.

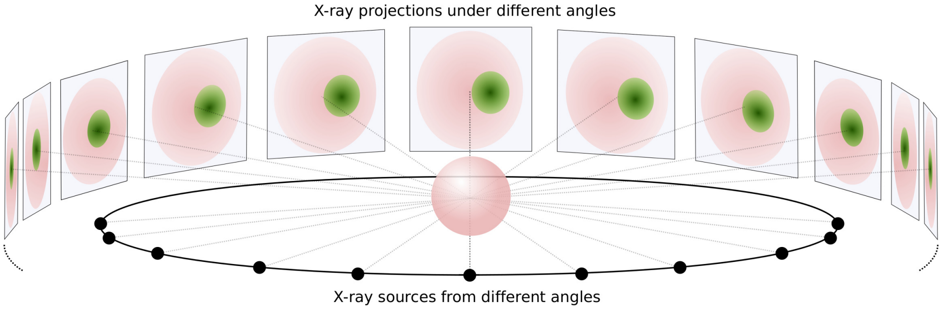


Figure 1.9: The process of obtaining multiple X-ray projections by rotating the object, with the source on the left and the detector on the right.

[84, 209, 232], and is given by the line-integral

$$\begin{aligned}\mathcal{R}[f](\theta,t) &= \int_{L_{\theta,t}} f ds \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1,x_2)\delta(x_1\cos(\theta)+x_2\sin(\theta)-t)\ dx_1\ dx_2.\end{aligned}$$

where $L_{\theta,t}$ is the line possibly passing through the object, given by $L_{\theta,t} = \{x = (x_1,x_2) \in \mathbb{R}^2 : x_1\cos\theta + x_2\sin\theta = t\}$. Note that for a given line, so for fixed θ and t, the Radon transform is equivalent to the Lambert-Beer law, with the left-hand side being the absorption image (see Fig. 1.10a). The Radon transform of an object f is also referred to as the *sinogram* (Fig. 1.10c). For a given object f, the acquired projections can be viewed as a set of line integrals in which detector position t and angle θ are varied:

$$P(\theta,t) = \mathcal{R}[f](\theta,t).$$

The goal of *tomographic reconstruction* is to retrieve the function f from its Radon transform, which is an *inverse problem.* The series of collected radiographs of an object are used as input for a *reconstruction algorithm* [35, 112]. There are various classes of reconstruction algorithms. The first class is the group of *analytical* algorithms in which explicit analytic expressions are derived for the material attenuations in terms of projections. To retrieve the object from its sinogram, intuitively one can smear out each projection back over the object space. The simplest *backprojection* (Fig. 1.10b) is given by the following:

$$f_{\mathrm{BP}}(x_1,x_2) = \int_0^{\pi} P(\theta, x_1\cos\theta + x_2\sin\theta)d\theta.$$

This backprojection operation assumes that all pixels contributed equally to the projections [286]. Therefore, the resulting reconstructed image is typically blurry (Fig. 1.10d). To prevent this, the projection data are usually filtered in the

Fourier domain by a filtering function $h : \mathbb{R} \to \mathbb{R}$, such that the lower frequencies (accounting for smooth edges) have less contribution than the high frequencies (accounting for finer details) to the backprojection. This leads to the *filtered backprojection* (FBP) method (Fig. 1.10e), where the ramp filter in the Fourier space is used, defined by $\overline{h}(\omega) = |\omega|$, and the complete backprojection step including filtering is given by

$$f_{\text{FBP}}(x_1, x_2) = \int_0^{\pi} \int_{-\infty}^{\infty} P(\theta, \omega) e^{2\pi i \omega (x_1 \cos\theta + x_2 \sin\theta)} h(\omega) d\omega d\theta.$$

To remove noise from the reconstruction, the high frequencies are removed by employing the Ram-Lak filter defined by $h(\omega) = |\omega|$ if $|\omega| \leq c$ and $h(\omega) = 0$ if $|\omega| > c$. Similarly, other filters can be employed as well. In the resulting FBP algorithm, the projection $P(\theta, t)$ is repeatedly *backprojected* for each angle θ. Since the detector positions and the scanning angles are discrete in practice, the integrals in the FBP expression above are substituted by the corresponding summations:

$$f_{\text{FBP}}(x_1, x_2) \approx f_{\text{FBP}}^{\text{disc}}(x_1, x_2) = \sum_{\theta \in \Theta} \sum_{t \in T} P(\theta, \omega) e^{2\pi i \omega (x_1 \cos\theta + x_2 \sin\theta)} h(\omega) d\omega d\theta.$$

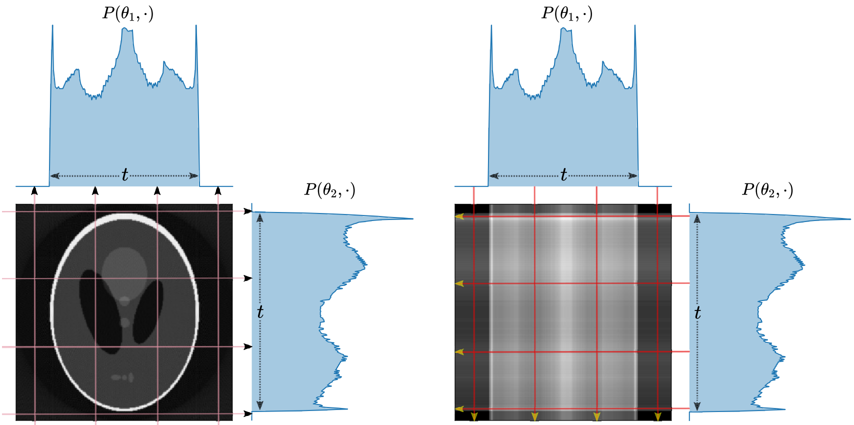
The FBP method is designed for parallel-beam geometries, while most CT setups have a fan-beam or cone-beam geometry. However, for fan-beam geometries the method can be employed by applying coordinate transformations, while generalization to three-dimensional cone-beam geometries results in what is known as the Feldkamp-David-Kress (FDK) algorithm [87].

FBP-derived methods are fast, easy to implement, and perform well on data that contain low noise and are near-complete (no missing angles). Therefore, these have been the method of choice until recently. FDK is the most practical analytic algorithm [286] for circular cone-beam geometries.

Another class of algorithms is that of *algebraic methods*, in which the Radon-transform formulation is changed into a system of equations, and the object function domain is discretized as a collection of *voxels* in a grid. The voxels can be modelled as a vector $\boldsymbol{x} = (x_1, \ldots, x_n)$, where n is the number of voxels. By putting all measurements in a vector $\boldsymbol{p} = (p_1, \ldots, p_l)$, where l is the number of detector pixels times the number of angles, the problem of tomographic reconstruction can be formulated as finding an $\boldsymbol{x}$ such that the following equality holds

$$\boldsymbol{W}\boldsymbol{x} = \boldsymbol{p}.$$

Here, $\boldsymbol{W}$ is the *projection matrix* in which entry w_{ij} indicates how much a ray corresponding to measurement i contributes to the absorption in voxel j (which in the modelling depends on voxel size, ray direction and type of ray-voxel intersection computation). Recovering an accurate image of the scanned object from data p is usually difficult for several reasons. First, the number of measurements l may be too



(a) Shepp-Logan phantom with two forward projections

(b) Two backprojections

(c) Sinogram

(d) Reconstruction with backprojections

(e) Reconstruction with FBP

Figure 1.10: Illustration of tomographic reconstruction algorithms on a 512×512 Shepp-Logan phantom. (**a**) Forward projections under two angles on the Shepp-Logan phantom. White areas indicate regions with high attenuation. These are reflected in the large peaks in the graphs of the projection functions $P(\theta_1, \cdot)$ and $P(\theta_2, \cdot)$. (**b**) Backprojections of the two projections onto the image domain. (**c**) The function values of all 256 projections under equidistant angles combined into a sinogram. (**d**) Reconstruction using backprojections resulting in a blurry figure. (**e**) Filtered backprojection (FBP) reconstruction with significantly less blurry edges.

low compared to the number of voxels. This leads to an underdetermined system, which may have infinitely many solutions. Likewise, the number of measurements may be too high, leading to an inconsistent system with no solution that satisfies all measurement equations. Furthermore, the measurements may be corrupted by noise (for instance, electronic noise in the detector). In this case, $\boldsymbol{p}$ could be written as $\boldsymbol{p} = \overline{\boldsymbol{p}} + \boldsymbol{e}$, where both the uncorrupted data $\overline{\boldsymbol{p}}$ and the noise contribution $\boldsymbol{e}$ are unknown. Working with the known data $\boldsymbol{p}$ instead of $\overline{\boldsymbol{p}}$ is likely to render a potentially solvable system unsolvable.

Instead of solving the above equation, the reconstruction problem is often posed as an optimization problem by finding a minimum-norm least-squares solution $\boldsymbol{x}^*$

that minimizes the *data discrepancy* or *data fidelity* term:

$$\boldsymbol{x}^* \in \underset{\boldsymbol{x}\in\mathbb{R}^n}{\arg\min}\ ||\boldsymbol{W}\boldsymbol{x} - \boldsymbol{p}||.$$

Algebraic iterative methods repeatedly improve the reconstructions with respect to the data discrepancy above, by alternating between forward and backward operations through matrix $\boldsymbol{W}$. One of the most commonly used methods is the Simultaneous Iterative Reconstruction Technique (SIRT). In this algorithm, the current reconstruction is forward projected, after which the difference between this forward projection with the measured data is backprojected using the transpose of the projection matrix. The update step for SIRT is given by the following:

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \boldsymbol{C}\boldsymbol{W}^T\boldsymbol{R}(\boldsymbol{W}\boldsymbol{x}^{(k)} - \boldsymbol{p}).$$

The diagonal matrices $\boldsymbol{C}$ and $\boldsymbol{R}$ contain normalizations with respect to the sums of columns of $\boldsymbol{W}$ and rows respectively. In the ART method, the update step is carried out by taking only one ray each time. For the SART algorithm [15], this is extended to a full update for every projection angle. However, SIRT has a more stable convergence than ART, but an adequate solution takes longer to reach.

Iterative algebraic methods can deal better with inconsistent, incomplete, and noisy projection data than analytical methods. Furthermore, these methods are also flexible with respect to the acquisition geometries, as these can be modelled in the projection matrix $\boldsymbol{W}$. However, the iterative methods require a proper stopping criterion, as running for too long can lead to overfitting to noise and therefore noisy reconstructions may appear. *Variational methods* combine the above optimization problem with a *regularization* term in which prior knowledge about the object can be incorporated:

$$\boldsymbol{x}^* \in \underset{\boldsymbol{x}\in\mathbb{R}^n}{\arg\min}||\boldsymbol{W}\boldsymbol{x} - \boldsymbol{p}||_2 + \lambda\mathcal{F}(\boldsymbol{x}).$$

In this formulation, $\mathcal{F}$ is the regularization operator and $\lambda \in \mathbb{R}_{\geq 0}$ is a parameter that controls the importance of the regularization with respect to the data-fidelity term. Two often-used regularization methods are Tikhonov and Total Variation minimization (TVmin) regularization. The former promotes solutions with small norms and reduces noise by setting the regularization function to $\mathcal{F}(\boldsymbol{x}) = ||\boldsymbol{x}||_2^2$, while the latter promotes sparsity of the gradients and smooth regions (yielding more piecewise smooth solutions), by setting $\mathcal{F}(\boldsymbol{x}) = ||\nabla\boldsymbol{x}||_1$. Regularization can steer towards more robust solutions containing less noise, but needs to be adjusted to each specific problem and requires the regularization parameter to be properly chosen.

Prior knowledge can also be incorporated by restricting the solution space. In *discrete tomography*, the resulting image has discrete values, meaning that all values x_i are contained in a fixed discrete set R. This type of tomography is a vast subfield on its own [121, 122], as the prior knowledge can be exploited for more efficient and robust algorithms (in terms of accuracy, noise handling and required number of projections [30]) when the object of interest is discrete, or even binary [29, 176].

While iterative algebraic methods have advantages over analytic methods, their major disadvantage is that the repeated computation of the forward projections and backprojections makes them *computationally expensive*. However, with the increase in computation power of modern computers, most notably the Graphical Processing Units (GPU), this problem is alleviated. This is reflected in the fact that algebraic methods have been the reconstruction method of choice since 2009 commercially [292], while before that time these were FBP-derived methods. On top of that, advances in parallel computing and efficient splitting of the tomographic reconstruction problem make the computation of algebraic methods more tractable [50].

Tomography overcomes the superposition problem found in 2D radiography, and a precise reconstruction of a product can be made and inspected with this technique. A number of parameters influence the quality of the reconstruction of a given object. During the acquisition, the quality of the resulting radiographs is generally improved by a higher current, peak voltage and exposure time. The detector resolution and the rate at which the detector can detect photons also influence the radiograph quality. For the reconstruction, the given computation time, stopping criterion and the type of algorithm (along with its parameter settings) all influence the result. The reconstruction quality is constrained by the allowed acquisition and computation time, the X-ray dosage limit and the energy budget. As opposed to medical CT, there is less concern about the dosage constraints in industrial applications, allowing for more precision and higher resolution in the reconstruction [55]. Nevertheless, in high-throughput applications, the acquisition times and computational times are severe limiting factors. These constraints can be met by for example reducing the exposure time and increasing the photon flux, and by adjusting the reconstruction algorithms, but this may lead to poor reconstructions. Therefore, even though CT gives more object information than 2D radiography, it is considered to be less suitable for online inspection [186].

1.4 Spectral X-ray imaging

X-rays consist of photons of different wavelengths. Since photons with different energies can exhibit different interactions with matter, this can possibly contribute to a deeper analysis of the imaged object. When the notion of X-ray energy is accounted for in the radiography acquisition, this is referred to as *spectral X-ray imaging*, and encompasses a wide range of approaches.

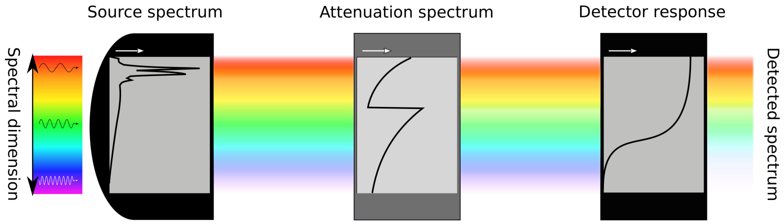


Figure 1.11: Schematic view of the different effects on a polychromatic beam.

To comprehend the methods, it is first important to understand where and how photon energies come into play (see Figure 1.11). First of all, the X-ray source emits a *polychromatic* X-ray beam. The distribution of energies of the photons emitted during the X-ray generation, also called the *source spectrum*, depends on the anode material and the applied peak voltage. An increase in peak voltage increases both the range of photon energies and the overall photon counts per energy. At the same time, the anode material determines the characteristic peaks that emerge from characteristic radiation due to the inner electron shell transitions in the atoms of the anode material when bombarded with fast electrons.

When X-ray photons interact with materials on their path, the resulting absorption is dependent on processes such as photoelectric absorption, pair production and Rayleigh and Compton scattering. The combined contribution of these processes to the beam attenuation depends both on energy and material type. See Figure 1.12 for a number of example spectra. The total attenuation μ_α of the X-ray by a material α for different energies is given by its *attenuation curve*, and is unique to each elementary material. Of particular note in these curves are the *absorption edges* caused by the photoelectric absorption. Due to the ejection of electrons in the shell and the subsequent electron transition from a higher shell, a characteristic radiation with fixed energy is emitted. As an electron will only be ejected when the incoming photon energy exceeds its binding energy, a sharp edge in the attenuation curve will appear. The electron binding energy is proportional to the atomic number, and inversely proportional to the distance to the nucleus of the atom. The electrons reside in shells with a fixed distance to the nucleus, with the innermost shell (K-shell) being the shell with the lowest distance. The binding energy is highest for the K-shell, and therefore requires the highest incident photon energy to be ejected. However, when the energy of the incident photon does not meet the binding threshold, no absorption will occur in this shell. While absorption edges also appear for other electron shells (L_1-, L_2-, M_1- edges, among others), the corresponding *K-edge* is located at the highest energies and provides the highest difference in attenuation, which makes it useful for applications dependent on this contrast (*K-edge imaging*) [94, 249]. Note that the attenuation curve is unique for each element, and so are its K-edge and other absorption edge locations. For elements with a high atomic number, such as metals, the K-edges are located at higher energies. These elements are therefore much more distinguishable than those

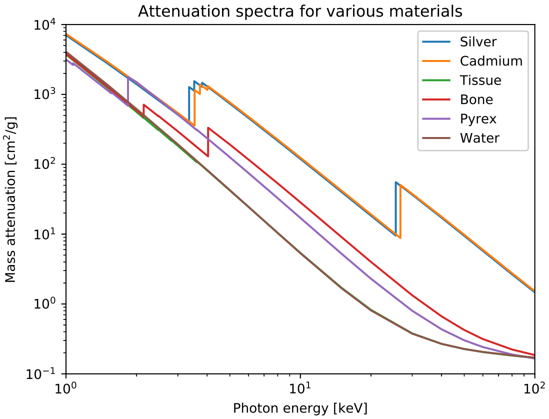


Figure 1.12: Attenuation spectra for various materials. The spectra are unique to each material, but some materials are easier to separate than others: silver and cadmium mostly differ through the slightly different locations of their absorption edges, and the difference in the spectra of water and tissue are almost indiscernible.

of elements with low atomic numbers, whose K-edges may not even be located within the X-ray wavelength range.

As can be observed from the attenuation curves, X-ray photons with low energies tend to be attenuated more strongly than those with high energy. For this reason, when a beam traverses through matter, the mean photon energy shifts to a higher value. This phenomenon is known as *beam hardening*. In standard CT, the X-ray beam is assumed to be *monochromatic*, consisting of photons with the same energy. However, because it is polychromatic in practice, the assumption of monochromaticity leads to *beam hardening artefacts* in the subsequent reconstruction, showing as dark and less attenuating regions in the CT reconstruction. To avoid this, ideally, the beam should be monochromatic, but a sufficiently high flux of such a beam can only be achieved at synchrotron facilities. To mitigate beam hardening artefacts resulting from a wide energy range, the low-energy photons in the incident beam are usually removed by *filtering* the beam with a (metallic) material with known spectral properties. With this approach, beam hardening is actually used to an advantage. Nevertheless, filtering does not completely eliminate beam hardening artefacts but reduces them, and longer scanning times are needed for similar signal-to-noise ratios.

The interactions of different X-ray photon energies with matter can be exploited in various ways. The most elementary way is to scan an object with beams of different energies. In the case of two energies, this is referred to as *dual-energy X-ray absorptiometry* (DEXA). This can be achieved by applying different values for the X-ray tube voltage (and possibly some additional beam filtering). For each detector pixel, this methodology results in two intensity values corresponding to two different voltages. As these define a point in a two-dimensional attenuation space, the projected materials are likely to be more separable from intensity value tuples from other material compositions than in radiographs from single peak voltages.

Instead of modifying the energy distribution in the polychromatic beam from the source, in *detector-based approaches* energetic information of photons reaching the detector can be directly obtained. *Dual-layer* scintillator detectors consist of two layers on top of each other. These spectral detectors have different photoelectric absorptions depending on their sensor materials, and the upper layer typically, on average, detects the low-energy photons while the bottom layer detects the high-energy photons [99, 118]. The advantage of such detectors is that these simultaneously capture images of the same object, avoiding any spatial differences due to motion between subsequent captures.

Similarly, as opposed to regular charge-integrating detectors (also known as energy-integrating detectors), (hybrid) *photon-counting detectors* register single incoming photons and are able to measure the energy of each photon individually [94] instead of integrating the energies of all incident photons [249]. Compared to the two-step process in charge-integrating detectors, a photon-counting detector consists of a semiconductor layer, in which incident X-ray photons generate electron-hole pairs. A bias voltage between a cathode on top of the detector and an anode on the other side separates the charge carriers to the contacts for electronic readout [263]. The process directly converts photons into electric charge, allowing for faster detection speed than with charge-integrating detectors. On top of this, the electronics in these detectors allow for spectral separation of photons into spectral *bins*, in which the number of photons between a preset energy range is counted. Compared to the first dual-energy methods, which were available in the '70s, photon-counting technology is relatively new. Nevertheless, the ability to simultaneously collect radiographs at different energies holds a huge potential for the future [94]. Therefore, several research groups worldwide are developing readout electronics for hybrid photon-counting detectors [91], such as the PILATUS chips at PSI, Medipix [26] and Timepix chips at CERN, PIXIE chips [36] at INFN and the PXD18k at AGH University.

An alternative class of spectral detectors operate by measuring the deposited energy in every pixel during a given exposure time and repeating this for many frames at high rates, yielding a binned energy distribution for each pixel [78]. For this reason, these *hyperspectral* X-ray detectors can attain a high spectral resolution. State-of-the-art detectors, such as the HEXITEC family of detectors [282] and SLcam [213], provide data containing up to hundreds of spectral energy bins and a spectral resolution of less than 1 keV, providing a rich spectral footprint of the measured object. Nevertheless, this relatively new technology requires comparatively longer acquisition times, as well as powerful data processing frameworks, which are currently in development [18].

For photon-counting detectors, different semiconductor materials like silicon (Si), gallium arsenide (GaAs), cadmium telloride (CdTe), or cadmium zinc telluride (CdZnTe) can be used. While Si-based detectors are generally cheaper, it suffers from low absorption, especially at higher energies where the detector layer becomes transparent. In comparison, CdTe and GaAs are more costly but have better absorption at higher photon energies. The extent to which the detector is able to register photons at different energies is given by the *spectral detector efficiency* or *detector response.*

The Lambert-Beer equation (where the line from the source to a the detector pixel is now given by ℓ) can generalized by incorporating the energy dependence :

$$I(E^{\max}) = \int_0^{E^{\max}} I_0(E)D(E)e^{-\int_\ell \mu(x,E)dx} dE.$$

In this formulation, the energy is integrated up to the maximum photon energy $E^{\max}$, which typically is the peak voltage energy or the maximum energy registered by the detector. Now, I_0 models the source spectrum as an incident photon rate as a function of energy, while D is a simple form of modelling the photon detector efficiency. In some formulations, I_0 and D functions are combined into a function known as the effective spectrum. The attenuation μ is now a function, dependent on both location and energy. As opposed to the monochromatic Lambert-Beer law, due to the energy integral, this expression cannot be converted into a linear right-hand side. This means that the data dependency on the energy-dependent attenuation coefficients is nonlinear (as opposed to Equation (1.1)).

While dual-energy absorptiometry has found many applications in food processing [16] and medical imaging, spectral X-ray imaging is especially promising for tasks involving the use of *contrast agents* which enhance the contrast between certain parts of an object. Examples of these include iodine and gadolinium, which have K-edges in the energy range of X-rays used in diagnostic radiography. Photon-counting detectors are valuable when at least two contrast agents need to be detected simultaneously [264]. More generally, spectral X-ray imaging is useful for material identification in an object [96], especially with multiple different materials. Therefore, it opens up new possibilities such as multi-energy X-ray testing [192] and multi-material decomposition [274], for example leading to improvements in tissue discrimination [293].

Even though the photon-counting detector technology is promising for energy-resolved imaging, there are still several challenges to overcome [293]. Firstly, an incident photon may be registered in a location other than where it hits the detector. This can be caused by fluorescence or Compton scattering, or by an effect known as *charge-sharing*, where the incident photon generates charges close to the border of a pixel that may be partly registered by an adjacent detector pixel. Secondly, when multiple photons arrive with too little intervening time, an effect known as *pulse pile-up* may occur [54, 99]. In this effect, the photons will be registered as one photon with an energy equal to the sum of the individual photons. Even though these detectors are able to detect hundreds of millions of photons per second per square millimeter, pulse pile-up is likely to occur. Both charge-sharing and pulse pile-up can lead to artefacts since the spectral distribution of photon counts is perturbed. Furthermore, a trade-off is required for the spatial resolution of the detector. Increased resolution increases the effects of fluorescence, charge-sharing and noise thresholding, and decreases the effects of pile-up [67]. Also, pixels may yield a different output with the same incident photon energy distribution. Therefore, accurate equalization and calibration need to be developed and carried out.

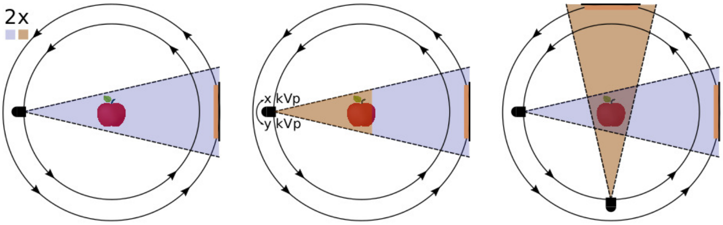
Apart from improvements with respect to energy differentiation and speed, there are more benefits from photon-counting detectors. The thresholding mechanism of photon-counting detectors prevents contribution of electronic noise [264]. This means there is no contribution of dark current, and darkfield correction is therefore not needed. Additionally, depending on the task carried out on the resulting spectral radiographs, the additional spectral information may reduce the total dose compared to regular radiograph acquisition. On top of this, due to the direct photon conversion, photon-counting detectors generally have much higher spatial resolution than energy-integrating detectors [159].

While hyperspectral detectors generate data with high spectral resolution and therefore circumvent the issue of tuning spectral thresholds to obtain a desired result, the registration of photon counts in every small energy bin yields huge data volumes. Additionally, there is the issue of spectral redundancy [60], indicating that spectrally adjacent X-ray images from the data cube are likely to be very similar. Therefore, given a specific task on the hyperspectral data cube, it is not clear in which spectral bins the important features are located.

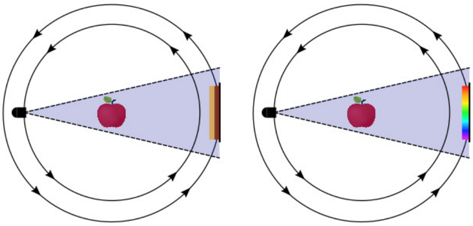
1.5 Spectral tomography

X-ray CT has yet to reach its full potential as spectral information of X-rays is not always recorded [274] or used in reconstruction algorithms. Similar to how radiography paved the way for X-ray transmission tomography, spectral radiography opens up possibilities for *spectral CT*, where the input radiographs are energy-resolved.

Exploitation of spectral characteristics in CT scans can be done in various ways, as demonstrated in Figure 1.13. Historically, the first and most simple spectral CT setup is to perform the same scan twice with different tube voltages. The possibility of artefacts resulting from object motion between the scans and the increased scanning time or X-ray dose are obvious downsides of this approach. Therefore, to reduce the disadvantages of sequential scanning, rapid *tube voltage switching*



(a) Scanning twice (b) Tube voltage switching (c) Two source-detector setups



(d) Dual-layer detector (e) Photon-counter detector

Figure 1.13: Schematic representation of various spectral CT setups.

sources were developed [11, 51, 55, 99]. However, even though the tube voltage switches at high speed at every scanning angle, this still results in considerably more scanning time than with conventional CT scanning, and relatively slow potential switching may still lead to motion artefacts. Another approach is to combine two source-detector combinations into one setup, known as *dual-source CT*, each source with a different peak voltage [11, 51, 55]. This method requires a more expensive setup, requires spatial co-registration of the obtained data, and the scattering resulting from the two different beams may increase the noise in the projections obtained by either detector. However, the setup of each detector-source combination can be configured independently. Both fast-switching potential dual-energy CT and dual-source CT are well-established in clinical practice. More recently, integration of *dual-layer detectors* in clinical settings has also been established [11, 40], with the benefit of simultaneously obtaining perfectly spatially aligned radiographs with different average photon energies [233]. Nevertheless, dual-layer detectors are more costly and designed for specific tasks, which makes these inflexible with respect to various material separation tasks [99]. In medical imaging, dual-energy CT is often used for contrast enhancement, artefact reduction, material decomposition and radiation dose reduction [233]. For industrial CT, dual-energy CT mainly improves reconstructions in terms of beam hardening artefact reduction and feature detection with low contrast [55]. The advent of photon-counting detectors enables *multi-spectral* CT, which generalizes the concept of dual-energy CT to more than two energy bins. As photon-counting detectors record data into narrower energy bins, the polychromatic beam drastically reduces beam-hardening artefacts found in conventional CT. Despite this, industrial CT most often employs energy-integrating detectors [55]. In medical CT, it is expected that dual-energy and spectral CT will be fully integrated into clinical practice and eventually replace conventional CT [99, 274].

In terms of modelling the spectral reconstruction problem, we take the spectral Lambert-Beer law and integrate over the energy interval $[E_c^{\min}, E_c^{\max}]$ of the c-th measurement with $1 \leq c \leq C$:

$$I_c = \int_{E_c^{\min}}^{E_c^{\max}} I_0(E)D(E)e^{-\int_\ell \mu(x,E)dx} dE. \tag{1.2}$$

Here, C is the total number of measurements, each with different energy properties. In *dual CT*, this number is $C = 2$. When prior knowledge of factors contributing to the attenuation is available, a basis decomposition can be incorporated that assumes the dependence of attenuation on location and energy can be written as a (finite) linear combination of functions that depend only on energy or location:

$$\mu(x, E) = \sum_{m=1}^{M} \widetilde{\mu}_m(E)\alpha_m(x).$$

Substituting this assumption into (1.2) yields

$$I_c = \int_{E_c^{\min}}^{E_c^{\max}} I_0(E)D(E)e^{-\int_\ell \sum_{m=1}^M \widetilde{\mu}_m(E)\alpha_m(x)dx} dE. \tag{1.3}$$

The goal of dual CT is to obtain a decomposition of the reconstructed object in terms of the contributions α_1 and α_2. This can be approached in two ways. First, each spectral channel can be reconstructed separately. This can be done by assuming attenuation at an effective energy E_c^{eff} (such as the mean energy in the bin) to remove the energy integral and therefore the nonlinearity of the expression. Then the object can be reconstructed with conventional CT techniques. For every position x, this yields a system of equations

$$\begin{cases} f(x, E_1^{\text{eff}}) & = \sum_{m=1}^M \widetilde{\mu}_m(E_1^{\text{eff}})\alpha_m(x) \\ f(x, E_2^{\text{eff}}) & = \sum_{m=1}^M \widetilde{\mu}_m(E_2^{\text{eff}})\alpha_m(x) \end{cases}.$$

in which $f(x, E_c^{\text{eff}})$ is the solution of the tomographic inverse problem in channel c at position x. With this system of equations, an *image-based decomposition* $\{\alpha_m\}_{1\leq m\leq M}$ can be computed. Note that in dual CT this can only yield a unique solution if $M = 2$. While the image-based decomposition is relatively easy to implement, proper estimations for the attenuations and the photon flux are required. When the energy range in each channel is wide, the reconstructed images will suffer from beam hardening artefacts [274, 283].

Alternatively, by changing the order of the summation and the integral, a *projection-based decomposition* can be attempted:

$$I_c = \int_{E_c^{\min}}^{E_c^{\max}} I_0(E)D(E)e^{-\sum_{m=1}^M \widetilde{\mu}_m(E)\int_\ell \alpha_m(x)dx} dE.$$

In this approach, the integrals $\int_\ell \alpha_m dx$ are computed first by solving the system of nonlinear equations, for instance by polynomial approximations [14, 63, 179, 266]. After the projection-based decompositions are obtained, each of these components are reconstructed separately, usually referred to as Virtual Monochromatic Images. Image-based decomposition is typically found to be faster, but projection-based decomposition yields superior decomposition results [179, 208].

The choice of base decomposition is mostly empirical. From the start of dual-energy CT, the decomposition into photoelectric effect (dominant at low energies) and Compton scattering contributions (dominant at high energies) [14] has been well-investigated. The former can be approximated by a function cubically inversely proportional to the photon energy, and the latter by the Klein-Nishina function. For materials that occur in the body (with an atomic number lower than 25), this decomposition is sufficient for the expression of the attenuation coefficients [99]. The energy-dependent attenuation coefficients can also be decomposed by attenuations of sufficiently different materials. For medical CT, the decomposition of water and bone is helpful, since most materials in the body have attenuation curves similar to either of those [118]. Alternatively, the K-edge in the diagnostic range of iodine makes this material suitable for material decomposition combined with tissue or water, since the attenuation of bone is similar to that of iodine-enhanced blood.

In advanced photon-counting CT, the number of preset spectral bins C is usually larger than two. In this type of spectral CT, the reconstruction techniques can be classified into two categories [143]. The first category is concerned with *multi-channel* reconstruction [64], which opts for the reconstruction of each spectral channel. In general, a channel in multi-channel CT does not necessarily refer to a spectral channel, but to any set of measurements of the same object with different acquisition settings. However, in this dissertation, we will mostly refer to a multi-channel image as a stack of images with each image resulting from different spectral properties.

The second category of spectral CT techniques is *material decomposition* from multi-channel data with more than two spectral bins. Many industrial and medical applications require more than $M = 2$ materials to be separated [172]. However, material decomposition with more than two materials poses challenges for standard binary reconstruction techniques [266], as found in dual CT imaging.

The difficulty of the multi-spectral CT problems is that these consist of many nonlinear equations, resulting in a nonlinear forward operator and an ill-posed problem. Approaches to these problems can be divided into two classes. The first class of methods are *two-step methods*. As a natural extension of a dual-energy CT approach, one subclass of two-step methods consists of methods that first reconstruct each spectral channel separately, after which an *image-based decomposition* (also known as post-reconstruction processing [27]) is carried out by solving for the material distribution with more than two channels (i.e. $C > 2$):

$$f(x, E_c^{\text{eff}}) = \sum_{m=1}^{M} \widetilde{\mu}_m(E_c^{\text{eff}})\alpha_m(x). \tag{1.4}$$

The second subclass performs a projection-based material decomposition (also known as pre-reconstruction processing), followed by the reconstruction of each material separately. The problem here is that the material decomposition in the projection domain is a nonlinear inverse ill-posed problem [6]. A common drawback of all two-step methods is that the separation of the material decomposition and the image reconstruction steps imply loss of information during the first step, for which the other second step cannot compensate [292]. The drawback of image-based decomposition is that it may suffer from beam hardening artefacts in the first step if the energy bins are too wide. In contrast, projection-based material decomposition suffers from sensitivity to noise [251] and the nonlinearity of the problem, of which the introduced errors are computationally costly to alleviate [274].

The class of *one-step methods* is designed to avoid the information loss and related problems inherent to two-step methods. One-step methods attempt to find a solution to the system of equations resulting from Equation 1.3 directly to obtain material maps, but the complexity of the problem (for which finding an analytic solution is challenging, if possible at all) necessitates the use of iterative algorithms. Such algorithms are even more computationally heavy than standard iterative reconstruction methods for non-energy-resolved CT problems. Moreover, many state-of-the-art approaches are limited to only a few materials.

For spectral detectors with a higher energy resolution, the approximation by energy discretization resulting in Equation 1.4 is more accurate than with detectors with lower resolution. Additionally, this modality is useful for K-edge imaging as sharp edges are more clearly visible in the multi-channel reconstruction. Since the bins are narrow, beam hardening problems are also avoided. However, unless extremely high acquisition times are applied, the low signal-to-noise ratios in the channels will have a high impact on the reconstructions by the aforementioned methods.

To mitigate the impact of high noise levels in bins because of low photon counts [244, 320], prior knowledge about the solution can be incorporated into the reconstruction [60]. Regularization is required to obtain satisfactory solutions for the ill-posed spectral CT problems [274]. Similar to the standard iterative techniques, variational methods can be employed with spectral CT, especially in the reconstruction routines of two-step methods. Intensity-gradient sparsity constraints [61], such as Total Variation Minimization (TV) and Non-Local Means (NLM) are among the most commonly used methods in spectral CT [244]. Other variational regularization methods exploit the structural redundancy between the channels [61, 64], with hyperspectral images providing an even stronger structural correlation between channels [288]. Numerous regularization methods have been proposed [274, 302], many of which are multi-channel enhancements of TV or NLM regularization. A few examples include spectral-means (based on NLM) [320], and Total Nuclear Variation (TNV) [64, 126] (which is an extension of TV, leading to low-rank solutions and therefore encouraging common edges in multiple images) [238, 320, 325], as well as parallel level sets methods such as directional Total Variation (dTV) [33, 64, 143]. Another approach is Total Generalised Variation (TGV) (designed for multi-modal and multi-channel imaging) [131, 147], which generalizes TV to higher orders of differentiation, and uses it for spectral CT to combine the channels. Both the sparse gradient and structural redundancy regularization methods can be combined into so-called spatiospectral regularization, for instance by combining spatial TV and spectral TGV [288]. All regularization methods require at least one parameter to be properly set and tuned, while the regularization method of choice is always dependent on the specific application. Regularization is generally easier to incorporate in one-step methods [27]. However, only a few methods incorporate spectral prior information into the one-step method that allows for the reconstruction of more material volumes.

Since photon-counting detectors are a new technology relative to energy-integrating detectors [284], multi-spectral CT has yet to find widespread use in medical and industrial fields. Nevertheless, the possibility to combine multi-spectral imaging with CT has enabled contrast agent imaging by exploiting energy-dependent K-edges of different materials [249]. Additionally, photon-counting detectors are also expected to improve spatial resolution [261]. Therefore, spectral CT and hyperspectral CT are beginning to find applications in medical imaging, such as enhancing tissue contrast [99, 296], and spectral CT is expected to be fully integrated into health care in the coming years [274]. This will further increase the interest in the underlying machinery and methods. However, it also requires further algorithmic developments for spectral image processing and CT reconstruction. At the same time, spectral and hyperspectral CT have great potential for security and industrial imaging [296]. Still, iterative reconstruction algorithms for photon-counting detectors are currently too computationally intensive for fast reconstruction [292], although the continued growth of interest [99] and algorithmic developments in methods with applications in medical CT are expected to improve this. Moreover, the difficulty of large-scale production of affordable high-quality photon-counting detectors still limits their commercial usage [159], and hyperspectral cameras are currently even more expensive. However, the first industrial X-ray cameras able to perform spectral imaging for food processing are scheduled to be launched in the coming years. For this reason, we will mainly focus on spectral X-ray CT in this dissertation. Furthermore, possible downsides of (hyper)spectral imaging, such as long acquisition times and computation times, may be alleviated by machine learning (covered in the next section) which is likewise getting much more attention in recent years in inverse problems and CT reconstruction.

1.6 Machine learning

Many routines in industrial imaging, for example X-ray based foreign object detection and crack detection, can in principle be carried out by humans. As the previous sections show, X-ray imaging and subsequent innovations enable doing these tasks in a noninvasive manner, but the work can be tedious, laborious or even harmful to humans. Moreover, the results are subjective, often irreproducible and possibly severely inaccurate. These aspects are some of the many motivating examples for the development of artificial intelligence, particularly with respect to food processing, with the aim of intelligent agents taking over the decision-making in various stages of the processing chain.

With machine learning, a model has the ability to learn (i.e. increasingly improve performance on a specific task) from data to carry out a task without being explicitly programmed to do so. Machine learning algorithms are concerned with predicting outputs y from inputs x using a function f. In supervised machine learning, this function can be estimated from labelled training data $\{(x_i, y_i)\}_{i=1}^m$ with input features $x_i \in \mathcal{X}$ and corresponding labels $y_i \in \mathcal{Y}$ [205], with $\mathcal{X}$ and $\mathcal{Y}$ being the input and output spaces, respectively. Machine learning aims to minimize the errors on unseen examples, as opposed to minimizing the errors on a training set only such as in optimization. Therefore, the training data should represent the unseen data well and the function f should both predict the training input features correctly as well as generalize to unseen data. As opposed to unsupervised machine learning, the labels y_i are known in *supervised* machine learning, which we will address in this section (for more details we refer to more technical and in-depth work [38]).

More formally, the labelled data space can be written as a product of the input and the output space $\mathcal{Z} := \mathcal{X} \times \mathcal{Y}$. The training data are assumed to be realizations of independent and identically distributed random variables $\boldsymbol{Z} = (Z_1, \ldots, Z_m)$, with $Z_i = (X_i, Y_i)$ from an unknown (joint) probability distribution denoted by $\mathbb{P}_Z$. Let $\mathcal{M}(\mathcal{X}, \mathcal{Y})$ be the set of all (measurable) functions from $\mathcal{X}$ to $\mathcal{Y}$. The *loss* function $\mathcal{L} : \mathcal{M}(\mathcal{X}, \mathcal{Y}) \times (\mathcal{X} \times \mathcal{Y}) \to \mathbb{R}_{\geq 0}$ measures the performance, i.e. how much the prediction $f(x)$ by a given function $f \in \mathcal{M}(\mathcal{X}, \mathcal{Y})$ differs from the target y with respect to a chosen metric. Given the data space $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$, the goal is to find a function $f : \mathcal{X} \to \mathcal{Y}$ that minimizes the *risk* $\mathcal{R}(f)$, which is the expected loss of the function f:

$$\mathcal{R}(f) = \mathbb{E}_{\mathbb{P}_Z}[\mathcal{L}(f, X, Y)].$$

Since the distribution according to $\mathbb{P}_Z$ is unknown, minimizing the risk over all measurable functions (resulting in the Bayes risk $\mathcal{R}^*$) is not possible. Therefore, the idea is to choose a *hypothesis set* or *hypothesis class* (a set of candidate functions) $\mathcal{F} \subset \mathcal{M}(\mathcal{X}, \mathcal{Y})$ and construct a learning algorithm to find a function $\hat{f} \in \mathcal{F}$ for training data $\boldsymbol{Z}$ that minimizes the *empirical risk*, which is given by:

$$\hat{\mathcal{R}}(\hat{f}, \boldsymbol{Z}) = \frac{1}{m} \sum_{i=1}^{m} \mathcal{L}(\hat{f}, Z_i).$$

The empirical risk $\hat{\mathcal{R}}(\hat{f}, \boldsymbol{Z})$ measures the average loss on the training data $\boldsymbol{Z}$ and converges by the law of large numbers almost surely to the true risk $\mathcal{R}(\hat{f})$ as the number of training samples m goes to infinity. Given training data $\boldsymbol{Z}$, an *empirical risk minimizer* is a function $\hat{f}_{\boldsymbol{Z}}$ that minimizes the empirical risk over all functions in the hypothesis set $\mathcal{F}$:

$$\hat{f}_{\boldsymbol{Z}} \in \underset{\hat{f} \in \mathcal{F}}{\arg\min}\, \hat{\mathcal{R}}(\hat{f}, \boldsymbol{Z}).$$

To assess the performance of an empirical risk minimizer $\hat{f}_{\boldsymbol{Z}}$, the difference between the expected risk $\mathbb{E}[\mathcal{R}(\hat{f}_{\boldsymbol{Z}})]$ (where expectation is with respect to $\boldsymbol{Z}$ and the randomness in the learning algorithm to choose $\hat{f}_{\boldsymbol{Z}}$ from $\mathcal{F}$) and the Bayes risk can be examined. This excess risk can be rewritten in the following way:

$$\mathbb{E}[\mathcal{R}(\hat{f}_{\boldsymbol{Z}})] - \mathcal{R}^* = \underbrace{\mathbb{E}[\mathcal{R}(\hat{f}_{\boldsymbol{Z}})] - \inf_{f \in \mathcal{F}} \mathcal{R}(f)}_{\text{estimation error}} + \underbrace{\inf_{f \in \mathcal{F}} \mathcal{R}(f) - \mathcal{R}^*}_{\text{approximation error}} .$$

The first two terms form the estimation error and give the difference between the expected risk of the empirical risk minimizer and the risk minimizer over the hypothesis set $\mathcal{F}$. The last two terms together form the approximation error and give the difference between the risk minimizer over the hypothesis set $\mathcal{F}$ and the Bayes risk (over all measurable functions). In other words, the estimation error reflects how good the data are for finding the best prediction function in $\mathcal{F}$ for the entire data space (with respect to the data distribution). In contrast, the approximation error reflects how restrictive the hypothesis set $\mathcal{F}$ is for finding the optimal prediction function. As illustrated in Figure 1.14, the hypothesis set $\mathcal{F}$ is an important parameter. In case $\mathcal{F}$ is very small, the estimation error can become small, but the approximation error may become large, leading to *underfitting* of the dataset and, therefore, poor results on the training dataset. Conversely, if the hypothesis set is large, the approximation error is low, but the empirical error is high, leading to *overfitting* and bad generalization. Hence, given training data, a well-chosen hypothesis set is necessary for proper learning, of which the complexity provides a good trade-off for function estimation from the data and approximation of the other elements in the distribution. The analysis of this problem is complex and we will refer to literature for more details [38].

While there is no universal learning algorithm for every data distribution, there are many approaches to tackle this problem that provide suitable learning methods in practice. First, given a dataset and a function f, it is not clear whether f overfits

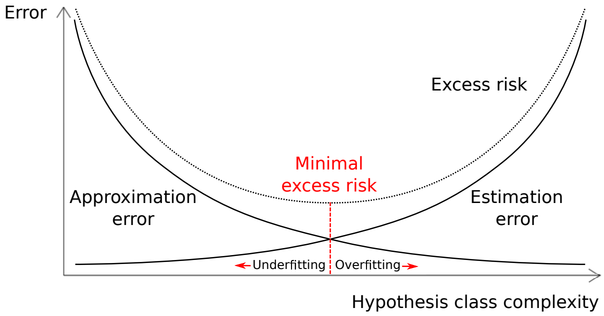


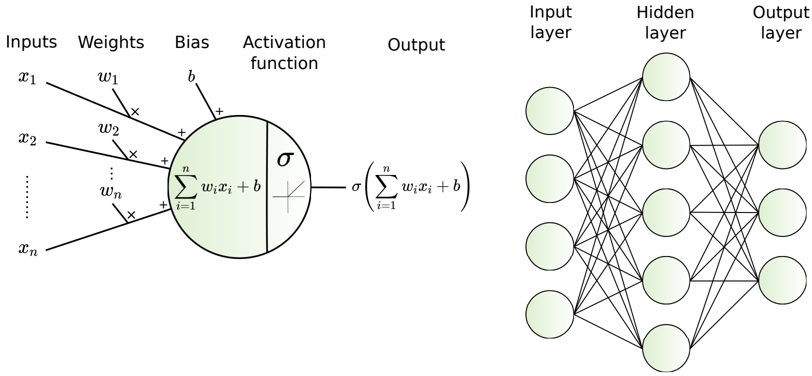
Figure 1.14: Schematic view of the estimation error and approximation error as a function of the hypothesis set complexity. The minimal excess risk is attained at a hypothesis class complexity for which the class contains a function that does not underfit the data, but no function that overfits the data.

or underfits the data. A common approach is to split the dataset into a training set and a validation set. Using a learning algorithm, candidate functions can be computed by minimizing the empirical risk on the training dataset. After that, the performance of a candidate function can be evaluated on the validation set. The idea is that a function overfitting on the training set can be observed by a weak performance on the validation set. Instead, the functions that perform best on the validation set can be used as candidates for the trade-off between estimation and approximation.

In traditional machine learning, there are many approaches for the hypothesis set. These include classes of functions described by decision trees, k-nearest neighbors and support vector machines, to name a few. Due to the increasing data size and complexity of their analysis, traditional machine learning methods have mostly become less suitable for many problems, including those related to food inspection [327]. However, one class of hypothesis sets that deal particularly well with these challenges is that of *neural networks*. Neural networks are inspired by the functionality of biological neurons in the brain. The basic blocks are the artificial *neurons* (Fig. 1.15a), that consist of a number of *input weights* $\boldsymbol{w} = (w_1, \ldots, w_n)$, a *bias* value b and an *activation function* σ, producing an *output value* $\phi(\boldsymbol{x})$ from its input values $\boldsymbol{x} = (x_1, \ldots, x_n)$ in the following way:

$$\phi(\boldsymbol{x}) = \sigma\left(\sum_{i=1}^{n} w_i x_i + b\right).$$

For the activation function σ, the Heaviside (threshold) function is an intuitive choice: only when the sum of the bias and the value of the inputs multiplied by the weights exceeds a certain threshold, the neuron will output a nonzero value. Alternative activation functions include the logistic functions and - more recently used - the Rectified Linear Unit (ReLU) activation function [102, 178], which is the identity for positive numbers and zero otherwise. By connecting the neurons,



(**a**) Artificial neuron structure

(**b**) Example network with fully connected layers

Figure 1.15: Building blocks of a neural network: (**a**) artificial neurons that weigh each input value and passes on the input value modified by a given activation function σ (as an example, the ReLU function is shown) and (**b**) the composition of multiple neurons into (fully connected) layers, and the composition of layers into a network of depth $d = 2$.

a *neural network* is formed (Fig. 1.15b). Different neurons can act on the same input, and can be organized in a *layer* with each neuron producing its own output $\phi_j(\boldsymbol{x})$ on the same input $\boldsymbol{x}$. By placing the layers ϕ_j^i in sequential order, with $i = 0, \ldots, d$ being the index of the layer and d being the depth of the network, a *multilayer perceptron* is created which takes input and propagates it through all the layers. The architecture of the network is given as $A = (C, \sigma)$, where σ is the activation function, and $C = (c_0, c_1, \ldots, c_{d-1})$ specifies the number of neurons per layer. The architecture A gives rise to the set Θ_A, which contains all value combinations of parameters of the network, i.e. the weights and biases. A realization of the network architecture A is a network $F_{A_\theta} : \mathbb{R}^{c_0} \to \mathbb{R}^{c_{d-1}}$ with values $\theta \in \Theta_A$ assigned to the parameters. The hypothesis set of a network architecture is then $\mathcal{F} = \{F_{A_\theta} : \theta \in \Theta_A\}$. The aim is to find a parameter configuration θ^* that minimizes the loss of the predictions of the realization of the network on the training set [188]:

$$\theta^* = \underset{\theta \in \Theta_A}{\arg\min} \sum_{i=1}^{m} L(F_{A_\theta}(x_i), y_i).$$

By feeding the neural network examples from a training set, the loss of the model's predictions and the *ground truth* can be computed. Since the functions ϕ_j^i in the neurons in the neural network are differentiable (almost everywhere), the gradient of the loss can be computed with respect to the network parameters using the *backpropagation* algorithm. Using a stochastic gradient algorithm, the error can be repeatedly used to re-adjust the parameters and optimize the network performance on the training dataset. Often used derivations of stochastic gradient descent for this purpose are AdaGrad [74] and ADAM [146]. The training procedure is carried out and evaluated repeatedly on the validation set. To assess the result of the training procedure, the obtained network is eventually evaluated on a separate test set.

Deep neural networks have recently gained popularity, where *deep* refers to the networks having many layers [205]. Such networks allow for more complex functions and trainable parameters, and deep learning methods can in some cases therefore perform better data analysis than traditional machine learning methods [327]. Deep learning applications include speech recognition, natural language processing, classification, partial differential equation solvers, image generation and image processing [12, 327]. Consequently, these approaches are nowadays used for health care, recommender systems, fraud detection and self-driving cars, to name a few applications.

For imaging tasks, deep Convolutional Neural Networks (CNNs) have found substantial usage for tasks such as denoising, object detection, edge detection, classification and *segmentation* (assigning a label to each pixel in an image) [296]. The key feature of a CNN is that the neurons use *convolution* operators (which generalize the weights of standard neural networks), where the output of a neuron is expressed in the following way:

$$\phi(\boldsymbol{x}) = \sigma\left(\sum_{i=1}^{n} C_{\boldsymbol{h}_i}(x_i) + b\right). \tag{1.5}$$

The convolution $C_{\boldsymbol{h}_i}$ of the neuron from input i is characterized by a *filter* $\boldsymbol{h}_i$. For 2D imaging applications, the convolution operator of h_i with the function $g_{\boldsymbol{x}}$, defined by $g_{\boldsymbol{x}}(k,l) = x_{k,l}$, is given by:

$$\begin{aligned} C_{\boldsymbol{h}_i}(\boldsymbol{x}) &= (\boldsymbol{h}_i * g_{\boldsymbol{x}}) \\ &= \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \boldsymbol{h}_i(m,n) x_{k-m,l-n}. \end{aligned} \tag{1.6}$$

In practice, the summations are taken over a range where the chosen filter is mostly nonzero (which is often indicated by the size of the filter). For instance, a filter can consist of a 3 × 3 matrix *kernel* containing weights that indicate the contribution of pixel $x_{k,l}$ and its neighbouring pixels to the corresponding pixel $C_{\boldsymbol{h}_i}(\boldsymbol{x})_{k,l}$ in the resulting convolved image (see Fig. 1.16 for an example). The expression in Equation 1.5 is an image of one *channel*. When images from multiple channels are combined, these are referred to as *feature maps* (for instance, an RGB image is a feature map with three channels). Each *convolutional layer* creates a new feature map from input feature maps. The input maps could be those of the last previous layer or of all the previous layers combined. By repeatedly passing on an input image through multiple convolutional layers, complicated features can be extracted from images. Deep convolutional neural networks apply many of such

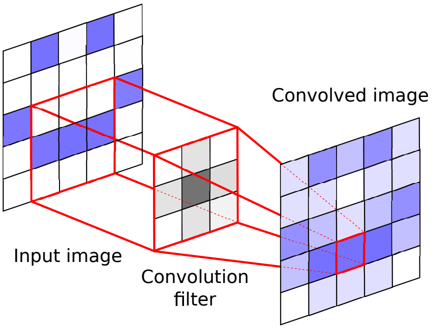


Figure 1.16: Visualization of a convolution. For every target pixel, the input pixel values are multiplied by the corresponding weights in the convolution filter, yielding the convolved image. The colors indicate the values of the pixels. In this case, the resulting image is a blurry version of the input image.

convolutional layers. Along with the biases in the neurons, the weights in the convolution operators are parameters that need to be optimized during the learning process, giving rise to a significantly higher number of trainable parameters than in conventional neural networks with similar architectures.

Nevertheless, despite seemingly overparametrization and the possibility of overfitting of CNNs, these have proven to yield outstanding results for many imaging problems. However, the reason for the success of such deep architectures is not understood well [38]. On top of this, many different strategies can be applied to such networks, giving rise to many classes of CNNs. For instance, the setup of the kernels can be tweaked, such as varying the *stride* (the step size in the summations of Equation 1.6), the *dilation* (the spacing between the kernel points), and *padding* (adding values at the edge of the image), which can all be considered as *hyperparameters* (parameters set beforehand that control the learning process). Additionally, there is a high degree of freedom in the architecture, such as the dependence of convolutional layers on the previous layers. Other examples are the addition of other operations between the convolutional layers, such as up- or downsampling operators or pooling layers (which also reduce the spatial resolution of the feature maps). Chapters 2 and 3 and will go into more detail about relevant CNN architectures and notation, respectively.

The many possibilities in the design of deep CNN architectures results in a large number of CNN classes without a uniform understanding of how each class works. However, this has not prevented its widespread use for imaging problems or limited its suitability for detection and segmentation problems in X-ray imaging and computed tomography. The main assets of CNNs for these fields are the recognition of complicated patterns and the ability to deal well with noisy images, which is for instance useful with (hyper)spectral images that are noisier than normal radiographs. Similar to the feasibility of iterative reconstruction techniques for tomography, learning with convolutional neural networks has become more prevalent due to improved hardware, particularly the GPUs, allowing to train networks with up to millions of parameters. Toolboxes for advanced deep learning, such as PyTorch [222, 223] and TensorFlow [4], provide even easier access to the application of these methods to imaging problems than before.

Deep learning is applied to many problems in 2D X-ray imaging. An important field of application is X-ray security imaging, such as cargo and baggage scanning at airports and other public areas, to detect and classify potentially harmful objects from radiographs [8]. In medical imaging, deep learning has many applications in radiography [195], including the analysis of chest X-ray images [53], as well as bone fracture classification [277] and bone age estimation [155]. For industry, deep learning methods with X-ray imaging have been developed for tasks such as defect segmentation of castings from radiographs [72, 90], weld inspection [90, 185, 192], and food processing [327]. However, deep learning combined with radiography is underused in those application areas, particularly in food processing [192], although it has been getting more attention in recent years [81]. The usage of deep learning is more common in X-ray CT, where it can be applied for 3D reconstruction and segmentation tasks with limited data [120, 187, 226, 326], such as limited angle and limited view measurements [160, 300], noisy data acquisitions [124, 157, 160, 304], as well as reduction of cone-beam artefacts [197] and reduction beam hardening [319, 329] and metal artefacts [32, 157]. On top of that, (parts of) the tomographic inversion operator can also be learned with deep learning [269, 285, 300]. Such approaches can also be combined with spectral imaging for improved image artefact reduction [49]. Another important field of application of deep learning is image compression [177], with network structures that encode data to a lower dimensional space and decode it to either the original data (autoencoder) or data corresponding to a specific task [198, 229]. In hyperspectral imaging, image compression by means of deep learning is carried out to reduce the size of hyperspectral data cubes of satellites and planes that are transmitted to remote locations for further analysis [73]. For the new hyperspectral X-ray imaging methods, such approaches may turn out useful when applied to high-throughput industrial tasks such as food inspection.

Despite the high success of deep learning in recent years, it also comes with many challenges. Apart from the incomplete mathematical understanding [38] of deep learning and their high computational costs [13], another practical problem is the reliance on data to successfully employ deep learning. Deep learning is very data-demanding [183], with the performance on vision tasks increasing logarithmically with the data volume [271]. On top of this, in supervised learning, the data often need to be annotated by experts, which is a tedious and time-consuming process. For food processing, there is a lack of large datasets [8] to employ deep learning for real-time applications. Moreover, very few open datasets are available for X-ray testing [192], most likely due to the limited number of experts for data annotation. Because of this, even though there is much potential, multi-channel acquisitions still need to be fully used in deep learning based X-ray imaging applications [8].

1.7 Research questions and dissertation outline

This dissertation presents a number of tools and techniques as building blocks that can be integrated into automated X-ray inspection systems for industry. As previously pointed out in this chapter, recent advances in X-ray imaging and machine learning enable the use of spectral X-ray imaging and deep learning. This allows for the construction of learning-driven automated spectral X-ray inspection systems. The main driving question throughout this dissertation is therefore the following:

Main question: *How can we use the possibilities of spectral imaging and deep learning in industrial imaging and industrial tomography?*

To provide an answer to this question, we investigate various aspects of spectral imaging, deep learning, or a combination of these. Each of the four forthcoming chapters will address a separate research question.

Chapter 2: *How can tomography assist industrial 2D radiography, in particular foreign object detection?*

In Section 1.1, we have seen that detection of unwanted ('foreign') objects within products is a standard procedure in many branches of industry for maintaining product quality. In Section 1.2, we have seen that X-ray imaging is a fast, non-invasive and widely applicable method for foreign object detection. As noted in Section 1.6, deep learning has recently emerged as a powerful approach for recognizing patterns in radiographs, enabling automated X-ray based foreign object detection at high rates. However, these methods require a large number of training examples and manual annotation of these examples is a subjective and laborious task.

In **Chapter 2**, we will propose a new workflow that solves this manual annotation problem. The workflow combines CT scanning with segmentation to efficiently and objectively produce annotated training data. With this data, a deep convolutional neural network can be trained to learn the foreign object detection task at hand, after which it can be applied to similar tasks. The method relies on high-quality CT scans and subsequent processing. As opposed to medical CT, in industry, there are generally no dosage constraints on the CT scan and therefore higher-quality reconstructions can be achieved by utilizing long scanning times and high penetration power by powerful X-ray sources. The workflow is demonstrated on a real X-ray CT dataset, with results suggesting that the method can be applied to industrial food inspection.

Chapter 3: *How can we use deep learning for task-driven (hyper)spectral data compression?*

In Section 1.4, we have seen that the arrival of (hyper)spectral X-ray imaging may provide additional insights into objects exposed to foreign object detection tasks. However, an important challenge in hyperspectral imaging tasks is to cope with the large number of spectral bins in the generated data. Common spectral data reduction methods do not consider prior knowledge about the task. Consequently, sparsely occurring features that may be essential for the imaging task may not be preserved in the data reduction step. As noted in Section 1.6, convolutional neural network approaches are capable of learning the specific features relevant to the particular imaging task, but applying them directly to the spectral input data is constrained by the computational efficiency.
In **Chapter 3**, we therefore introduce a novel supervised deep learning approach for image analysis. The method combines data reduction and image analysis in an end-to-end network architecture. The proposed data reduction network architecture, which we name DRCNN, consists of a data reduction block paired with a CNN. The network component that performs the reduction is trained jointly with the CNN such that image features most relevant to the task are preserved in the reduction step. The approach is demonstrated on a number of artificial datasets, and suggests the possibility for higher compression and accuracy in various applications, including hyperspectral X-ray imaging.

Chapter 4: *Can we use multi-channel imaging to improve reconstructions in discrete tomography?*

As noted in Section 1.3, discrete tomography is concerned with objects that consist of a small number of materials, which makes it possible to compute accurate reconstructions from severely undersampled projection data. For cases where the allowed intensity values in the reconstruction are known a priori, the discrete algebraic reconstruction technique (DART) has been shown to yield accurate reconstructions from few projections. However, a key limitation is that the benefit of DART diminishes as the number of different materials increases. Since new tomographic imaging techniques can simultaneously record tomographic data at multiple channels and provide multi-channel data, a new reconstruction algorithm can exploit this additional information.
In **Chapter 4**, we present Multi-Channel DART (MC-DART). This method is a generalization of DART to multi-channel data and combines the information for each separate channel-reconstruction in a multi-channel segmentation step. By iteratively applying these steps, a final accurate segmented reconstruction can be produced from multi-channel data. Through a number of simulation experiments, it is shown that MC-DART is capable of producing more accurate reconstructions compared to (single-channel) DART.

Chapter 5: *How can we use prior information on spectral material signatures in tomographic reconstruction and material decomposition algorithms?*

The advances noted in Section 1.5 in multi-spectral detectors are causing a paradigm shift in X-ray CT. Spectral information acquired from these detectors can be used to extract volumetric material composition maps of the object of interest. The image reconstruction step is relatively straightforward if the materials and their spectral responses are known a priori. If these are not known, however, the maps as well as the responses need to be estimated jointly. The most conventional workflows in spectral CT involve performing volume reconstruction followed by material decomposition, or vice versa. However, these methods inherently suffer from the ill-posedness of the joint reconstruction problem.
In **Chapter 5**, we present a new one-step method for tomographic reconstruction in spectral CT. In this method, which we name ADJUST, prior information about spectral attenuation curves of materials that may appear in the object of interest is taken into account by formulating the problem in a specific way. In essence, the to-be reconstructed volume is written as a multiplication of a material map matrix, a material-indicator matrix, and a spectral dictionary matrix. In the proposed optimization method, the first two matrices are jointly iteratively estimated.

In **Chapter 6**, we will return to the main research question and discuss how the methodologies developed in each chapter can support industrial X-ray imaging and tomography. In addition, the relevance of each method is discussed with respect to the fields of spectral and hyperspectral tomography, discrete and multi-channel tomography, hyperspectral imaging, machine learning and dimensionality reduction, along with all their application areas. Furthermore, we outline the contributions, implications and future research directions.