2: Creating three-dimensional images with diffuse light

DOT generates imagery using a diffuse optical technique which is fundamentally different from most other imaging modalities, such as MEG, PET, and fMRI. These "ballistic" imaging modalities provide a more direct spatial mapping between voxel locations within the brain and the signal parameters being measured. Chapter 2 begins by introducing many of the imaging modalities currently in use, along with some of their limitations as neuroimaging tools. Then the steps involved in collecting diffuse optical data and reconstructing three-dimensional DOT imagery are discussed. The three classes of DOT instrumentation: CW, RF, and time-domain, are presented and their relative merits are briefly described.

In order to understand how images can be generated from diffusely scattered light, one must first understand how light propagates through scattering media, of which biological tissue is but one example, so Chapter 2 presents some basic optical concepts. The many forms of optical scattering are then introduced, followed by a discussion of optical absorption and the Beer-Lambert law. In biological tissue, scattering-absorption interactions can complicate DOT image reconstruction, so issues such as the pathlength correction factor and scattering anisotropy are explained as well. The three categories of light propagation, loosely referred to as "ballistic," "snake," and "diffuse," are introduced, and methods for modeling each regime are mentioned. The many ways in which scattered light propagation is modeled are also discussed with reference to DOT. Why these models vary as a function of turbidity is also explored, and the intrinsic sources of optical absorption and scattering in living tissue are presented.

Once the fundamental optical principles have been discussed, the process of generating tomographic imagery from diffuse optical measurements can now be explained. The process of reconstructing DOT imagery involves solving two classes of mathematical problems. The first involves modeling the optical properties of the medium such that the effects of perturbations at known locations can be converted into accurate spatial images. This is referred to as the "forward" problem, and is described, followed by discussions on both direct approaches and numerical methods for solutions. Once the forward problem has been solved and the spatial weight matrices have been generated, the more difficult "inverse" problem – calculating the locations and magnitudes of the absorption changes from the raw optical measurements – is discussed, and the need for regularization is explained. A number of common regularization techniques are also presented to show the variety of ways this problem can be addressed.

2.1 Ballistic imaging modalities

Most neuroimaging modalities in use today operate in a ballistic fashion. The term "ballistic" imaging refers to all modalities in which the excitation energy travels in a straight line from the source through the sample, and from the sample to the detector. With ballistic imaging, there always exists a one-to-one dimensional mapping between the measured parameter(s) and the volume locations within the object being scanned, such that the weight function has a finite value along the source-detector line of sight and nearly zero weight elsewhere. As a result, all ballistic imaging modalities generate well-posed analytical problems (i.e. problems in which the number of measurements is always sufficient to reconstruct an image, so long as the Nyquist criteria are met). Although some of these modalities involve single-scattering, spatial coherence is preserved, and thus direct 1-to-1 spatial images of the object can be formed readily. Figure 2.1 compares the weight distribution of a ballistic modality (C-

T), to that of DOT. Examples of other ballistic imaging modalities include endoscopy, microscopy, optical coherence tomography, PET, ultrasound, and fMRI.

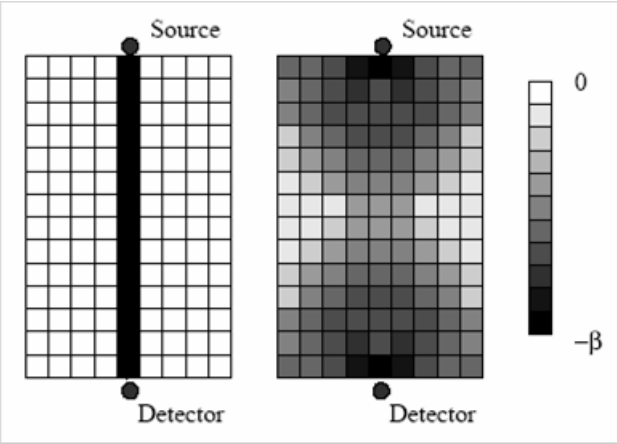

Figure 2.1. Weight distributions for C-T (left) and DOT in transmission-mode (right). The DOT distribution is a broad probability density function, while the C-T distribution gives equal non-zero weights along the source-detector line of sight and zero weights elsewhere [29].

Although image formation with fMRI is far more complex than with these other modalities, there is a direct spatial mapping between every location within the magnet bore and its Larmor frequency in the Fourier domain so the analysis is always well-posed, hence its designation as ballistic [3].

Vision, endoscopy, and confocal microscopy

Optical systems rely on specular reflection to generate 2-D images. The simplest detector is the human eye-brain combination, however electronic imagers are now used to provide a wider spectral range and the ability to store imagery for later retrieval and processing. Confocal microscopy employs spatial filtering to provide sufficient rejection of scattered light so that images of cutaneous tissue structures can be formed 200-300um below the skin surface [30]. The sharp rolloff in spatial sensitivity, achieved by reimaging a small pinhole within the tissue itself (i.e. making it "confocal" with the sample volume) is what allows a specular image to be formed.

Optical Coherence Tomography (OCT)

OCT is a variant of ballistic imaging which improves on confocal microscopy by using confocal geometry in combination with a Michelson interferometer to achieve confocal gating in both lateral (X and Y) dimensions and "coherence" gating in the axial (Z) dimension [31]. Images are reconstructed in a similar manner as with confocal microscopy, and the axial information is achieved through envelope detection of the Doppler-generated carrier frequency.

X-rays/C-T/video fluoroscopy

With X-rays and C-T, two-dimensional images can be formed directly, as shadowgrams on film or on a phosphor-coated surface. In order to form 3-D volume reconstructions, scanning and backprojection are required, however with ballistic imaging modalities like PET and C-T, the image reconstruction is never ill-posed, as it is with diffuse imaging modalities like DOT.

Positron Emission Transaxial Tomography (PET)

PET involves the use of radionuclides whose uptake is metabolically dependent. After enough radioactive material, bound to a metabolic substrate (i.e. deoxyglucose) has preferentially accumulated

within the metabolically active regions of the brain to provide sufficient contrast, the scan is begun. The most common radionuclide, ^{18}Flourine, decays via the emission of a positron (a particle with the mass of an electron, but with 1 unit of positive charge), which subsequently annihilates the first electron it sees. This results in their total conversion to energy, and the release of two 511-keV gamma rays with trajectories exactly 180' apart (i.e. with reciprocal velocities). Coincidence detection using angle-resolving detectors, followed by backprojection similar to C-T scanning, can then reveal the points of origin of these gamma ray pairs [4]. Each detector can be operated in multiple coincidence with many detectors across from it, thereby defining coincidence sampling paths over many angles (fan-beam response). Also, at any given angle many parallel coincidence sampling paths can be defined, resulting in high "linear sampling." The tomographic reconstruction software then takes the coincidence events measured at all angular and linear positions to reconstruct an image that depicts the localization and concentration of the positron-emitting radioisotope within one plane of the cortex This step is repeated to collect 2-D images from enough planes to construct a volumetric, or 3-D image of the entire brain.

Since PET depends on the selective uptake and subsequent decay of a radioactive tracer, the formation of a single volumetric image may take only a minute or two, however the collection of subsequent images requires waiting for the prior radionuclide dose to decay sufficiently, which can take ten minutes or more. This is far too slow for imaging hemodynamic events in real-time [4].

Ultrasound

Acoustic pulses generated by a piezoelectric ceramic transducer travel through the body tissue and fluids until they reach an acoustic boundary, where tissues with slightly different acoustic impedances meet. As a result of this impedance mismatch, some fraction of the acoustic energy will reflect back and the rest will continue to travel onward. Typical values for acoustic velocity and impedance are given in Table 2.1.

Ultrasonic imagery is computed from the echoes, which are detected by the transducer array as analog signals that are then digitized for analysis. Coherent detection allows Doppler velocimetric measurements of axial blood flow rely on weak echoes from erythrocytes in the bloodstream, however blood oxygen saturation is undetectable with current ultrasound technology. Thus, although ultrasound and laser Doppler velocimetry (a related technique using temporally coherent laser light) can measure blood velocity, it cannot be used directly to quantify changes in cerebral metabolism or the cerebral metabolic rate of oxygen consumption ($CMRO_2$).

Table 2.1. Acoustic properties of various tissues within the human body. Acoustic waves refract and reflect from impedance boundaries, creating the backreflected signals used to generate ultrasound images of tissue structures [32, 33].

Material	Acoustic Impedance (Rayls)	Velocity (m/s)
air	0.0004	331
fat	1.38	1450
water	1.54	1540
brain	1.68	1541
blood	1.61	1570
kidney	1.62	1561
liver	1.65	1549

muscle	1.70	1585
lens of eye	1.84	1620
skull-bone	7.8	4080

Functional MRI

MRI employs both static and dynamically scanned magnetic fields to alter the nuclear resonance frequency in such a way that spatial precession maps of tissue – or any other protonated structure – can be formed. Since the proton precession frequency, termed the Larmor frequency (after Larmor, a British physician), is a direct function of the ambient magnetic field strength, a gradient in the B-field will induce a gradient in the Larmor frequencies of the respective protons. This frequency information can then be directly mapped into the spatial dimension, through Fourier transformation, yielding 3-D imagery [3]. Image contrast with MRI can be achieved in many ways. For the sake of brevity, only echo-planar imaging will be discussed here.

Gradient echo-planar imaging (EPI), commonly used for functional MRI measurements, obtains its signal from changes in the T_2 decay time, produced by the precessional dephasing of both intra- and extravascular protons due to inhomogeneities in the local magnetic environment. Most organic substances within the body are weakly diamagnetic. Because of their low magnetic susceptibility they only slightly perturb the local magnetic field, and thus appear magnetically "transparent." A few biological substances, notably deoxyhemoglobin, have greater magnetic susceptibility. These "paramagnetic" materials have a weak magnetic dipole moment. When exposed to a magnetic field, they attempt to align with the field in an effort to reduce their energy state. The effect of this is the creation of local inhomogeneities in the B-field surrounding these particles. Unlike electrostatic fields, magnetic field lines are "closed", so magnetic field strength drops off sharply with distance, approximately as $1/r^3$, which provides excellent localization for imaging purposes. Blood Oxygen Level-Dependent (BOLD) fMRI is based upon detection of local field perturbations produced by Hb within each voxel [17, 34]. Note that it is not the Hb itself that is being detected, rather it is the effect of the magnetic inhomogeneities produced by the paramagnetic Hb molecules on nearby protons. At low (1.5T) static field strengths, much of the precessional dephasing signal is intravascular. As the static field strength is increased, this tendency is reduced as the signal protons become primarily extravascular [12]. So at high field strengths, perturbations from Hb create stronger dephasing signals in the local tissue than in the intravascular water.

The critical factor in creating the BOLD signal is the local Hb concentration within each voxel. It turns out that the intravoxel Hb concentration ([Hb]) is only indirectly related to neural activity, as will be discussed in upcoming chapters, and the interplay between neural, metabolic, and vascular dynamics is the most critical, and yet least understood, aspect of this imaging modality.

If a strongly paramagnetic or ferromagnetic substance, such as a gadolinium compound or a ferric oxide nanocolloid, is introduced into the vascular system, it dominates the weak paramagnetic Hb signature and the resulting MRI signal then becomes sensitive to changes in the total intravascular blood volume [17]. This enables the measurement of cerebral blood volume (CBV), independent of changes in [Hb]. A distinct advantage of these exogenous contrast agents is that they completely eliminate the intravascular blood signal component, and with proper dosing, can provide sensitivity to brain parenchyma that is spatially uniform throughout the brain.

Advantages of ballistic imaging modalities

There are numerous advantages to ballistic imaging. With ballistic imaging, trajectories are easy to establish, so you always know where your particles or waves have been. Image reconstruction is straightforward, since there are enough equations (i.e. measurements) to solve for all of the unknowns,

so standard backprojection techniques serve well. With DOT, photon paths are treated in aggregate, and the result is only a probability distribution function – not a well localized path.

Spatial resolution can be very high, limited only by basic physical principles such as diffraction and coherence, and by the wavelength of the incident radiation (the DeBroglie wavelength, for subatomic particles). With DOT, the photon probability density function (PDF) fundamentally limits spatial resolution. Even with an infinite number of optode pairs, spatial resolution will always be inferior to ballistic techniques.

Image quality can be uniform throughout the entire sample volume, while with DOT the spatial resolution decreases with increasing depth, as the PDF diverges.

Problems with ballistic imaging modalities

Problems can include an increased risk of tissue and genetic damage, limited diagnostic information, poor temporal response, and high cost.

In order to travel ballistically through tissue, photon energies must be very high – typically tens of thousands of electron-volts for X-ray, PET, and C-T. This reduces the frequency of collisions with atoms in the body, but when they do occur, they can deposit tens of eV per interaction, more than enough to rupture the covalent bonds within most biomolecules. High energy photons from X-ray and C-T must be absorbed within the tissue in order to generate contrast. Some of the positrons generated during PET are also absorbed. Random scission of biomolecules can lead to autosomal mutations, cell death, and cancer.

Most ballistic modalities provide excellent structural detail, but only limited metabolic information. X-rays, C-T, and ultrasound have excellent spatial resolution, but provide essentially no information as to the metabolic state of the tissue. This information (blood oxygenation, acidity, metabolic rate, etc.) could be helpful in both detecting and identifying tumors, vascular abnormalities, and other histopathologic states [35].

Those that do provide metabolic information (PET and fMRI) currently offer limited temporal response. fMRI, currently the fastest, can acquire data at rates approaching 10 samples per second, albeit at the cost of a significant reduction in contrast-to-noise ratio, so obtaining rapid frame rates requires the co-addition of multiple trials [3]. PET requires the uptake of radioactively-tagged metabolic substrates (^{18}FDG), so the latency period between consecutive image acquisitions, necessary for decay of the remaining radionuclide, can take tens of minutes. Both fMRI and PET require the purchase and maintenance of very large and heavy instrumentation (superconducting magnets, cyclotrons, magnetically shielded rooms) that can be quite costly. Prices for each generally start at around one million dollars.

2.2 Diffuse imaging and the origin of DOT

In strongly turbid materials where μ_s exceeds μ_a, diffusively propagating light creates a transmission-mode PDF as depicted in Figure 2.1. However, since brain function measurements are often performed in the reflective mode, the three-dimensional PDF resembles a banana. DOT generates imagery based upon local changes in the concentration of chromophores such as Hb and HbO_2 within the tissue volume. In Section 2.2, the methods used to create volume imagery from point optical measurements across the cortex are described, beginning with a brief historical overview of the evolution of diffuse imaging.

A historical timeline of diffuse optics: From oximetry to DOT [36-40]

1860s Felix Hoppe-Seyler demonstrated that the optical absorption changed when blood was mixed with oxygen, and that hemoglobin and oxygen formed a compound that he called oxyhemoglobin.

1864 Georg Gabriel Stokes reported to the Royal Society of London that hemoglobin was in fact the carrier of oxygen in the blood.

1876 Karl von Vierordt used the spectroscope to demonstrate oxygen consumption in his hand. Unfortunately, Von Vierordt's study was ignored at the time and later all but forgotten.

1929 Glen Allan Millikan used a photoelectric blood oxygen saturation meter to measure color changes over time when desaturated hemoglobin solutions were mixed with oxygen solutions in an experimental setting.

1932 Ludwig Nicolai recorded changes in oxyhemoglobin and deoxygenated hemoglobin and noted exponential curves for decreases in the former and increases in the latter. He also demonstrated linear changes over time in the logarithm of the light.

1935 Kurt Kramer, a student of Nicolai's, demonstrated in vivo measurement of blood oxygen saturation using arterial transillumination.

1935 Karl Matthes introduced the first two-wavelength earlobe-mounted blood oxygen saturation meter able to continuously monitor blood oxygen saturation in humans.

1941 Millikan reported on the instrument – and introduced the term "oximeter" to describe it – at the annual meeting of the American Physiological Society. The instrument used an incandescent, battery-operated light and red and green filters.

1948 Earl Wood of the Mayo Clinic improved Millikan's oximeter, including the addition of a pressure capsule, the idea of which had been introduced by Squire during the war.

1950s Brinkman and Zijlstra developed reflectance oximetry. The instruments developed by Wood and by Brinkman and Zijlstra were used briefly in anesthesia, but proved to be too unreliable and difficult to use.

1970s Takuo Aoyagi, at Nihon Kohden Corporation, introduced pulse oximetry, which differs from the previously described oximetry in that it does not rely on absolute measurements, but rather on the pulsations of arterial blood.

1974 Aoyagi and others delivered the first commercial pulse oximeter to Nakajima. The instrument, the OLV-5100, bore some resemblance to Millikan's 1940 oximeter inasmuch as it employed an earpiece with incandescent light, filters and photocells.

1977 The Minolta Camera Company introduced the Oximet MET-1471 pulse oximeter with a fingertip probe and fiberoptic cables.

1**977** F.F. Jobsis published his seminal paper, "Noninvasive, Infrared Monitoring of Cerebral and Myocardial Oxygen Sufficiency and Circulatory Parameters," which is cited universally as the study that introduced near-infrared spectroscopy [41].

1980s Near-infrared spectroscopy benefited from the development of time- and frequency-resolved techniques. These techniques allowed for greater sensitivity and specificity of NIRS measurements, and so spurred a resurgence of interest in the modality.

1986 Fox and Raichle discover a mismatch between CBF and CMRglu during functional imaging with PET [42], creating a huge stir in the neuroimaging community.

1990s Hundreds of studies on near-infrared spectroscopy and DOT were published.

1999 Siegel demonstrates a portable and lightweight continuous-wave DOT imaging system [9].

The "banana" pattern

When a light source and detector are placed at least 2cm apart in contact with the scalp, a small fraction of the incident light will scatter within the scalp, skull, and cerebral cortex, and then repeat this random journey to eventually reach the detector. Although the fraction of light which scatters through the cortex to reach the detector can be as little as 1 part in10^{-10}, this light can provide both spatial and temporal information about the metabolic state of the cortical tissue. For the backscattering geometry just described, a combination of Beer's Law absorption, diffusive scattering, and $1/r^2$ effects leads to the formation of a distinctive "banana"-shaped flux density profile as depicted in Figure 2.2. The shape of this profile is a function of the source-detector separation, the absorption coefficient: u_a, and the reduced scattering coefficient: u_s' within the tissue. Increases in both tissue scattering and absorption act to reduce the amplitude of the detected signal and reduce the penetration depth. In a similar way, static variations in both u_a and u_s' due to tissue structural features within the imaged volume will alter the optical probability density function and will thus affect the quality of the reconstructed image [43].

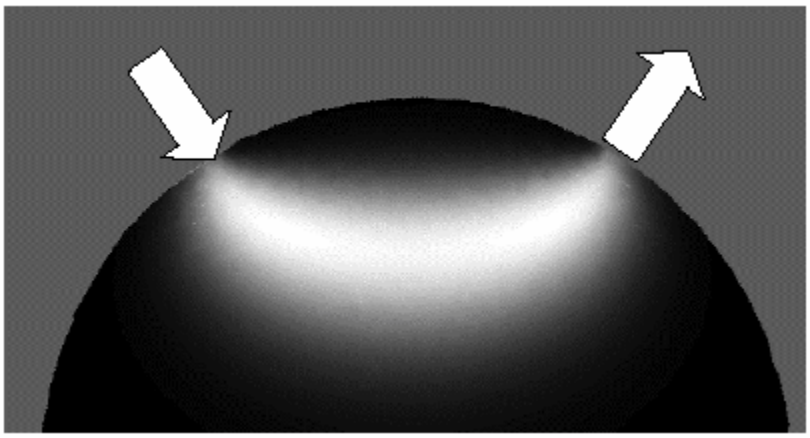

Figure 2.2. The probability density function (PDF) for light entering and exiting an isotropic scattering material at the locations shown by the white arrows. The PDF represents the magnitude of the optical flux density throughout the tissue. Sensitivity to localized changes in tissue absorption is directly proportional to the value of the PDF integrated over the perturbing volume [43]. Not illustrated are the increasingly large PDF values with growing proximity to the source and detector optode locations represented by the white arrows.

Diffuse imaging looks for variations in absorption and scattering

Within the ~650nm to 850nm near-infrared (NIR) spectral band, the absorption of light within tissue is much lower than in any other waveband. The absorption coefficient in tissue at ~800nm can exceed 10cm. The attenuation due to scattering is from 10 to 1000 times greater than due to absorption in this band, so light transport models must address both the scattering and absorbing tissue components.

Since the tissue scattering coefficient (μ_s) is much higher than the absorption coefficient (μ_a) in the NIR band, the transport of scattered light through tissue can be treated statistically, using algorithms based upon standard diffusion equations to process the data. Diffuse imaging measurements can quantify one or both of these parameters.

Of all the possible optical absorbing compounds within tissue, there are only a few chromophores in the near-IR spectral band that vary with metabolic activity. The three chromophores of interest are Hb, HbO_2, and cytochrome oxidase. In normally perfused brain tissue with a blood volume of 3-5%, hemoglobin absorption predominates, and cytochrome oxidase absorption is normally neglected. Since Hb and HbO_2 have different spectral profiles, the relative concentrations of each can be obtained through the measurement of only two discrete optical wavelengths. One fortuitous advantage is the existence of an isosbestic point in the Hb/HbO_2 absorption spectra at ~805nm. Since [Hb]+[HbO2]=[Hbt], an 805nm optical measurement can directly quantify total blood volume without the spatial errors introduced by partial-volume effects. Thus DOT can measure both blood oxygenation and CBV simultaneously, without the need for exogenous contrast agents.

Forming an image from the DOT Measurements

Since μ_s' is much larger than μ_a, light scattering in tissue is best modeled as a diffusive process. As a result, all that can be measured are localized estimates of the bulk tissue properties – not the point measurements possible with ballistic imaging. Thus, DOT can provide μ_s' and μ_a vs. wavelength with limited spatial resolution, and subject to confounding influences from both in-plane and out-of-plane absorbers within the diffusive PDFs. Much effort continues to be spent on developing means to mitigate these confounds and improve the spatial resolution and accuracy of DOT.

Although the tissue optical properties μ_s' and μ_a are the two parameters of interest, these must be determined from a number of discrete optical measurements. In its simplest form, DOT can only measure the amplitude and transit time of the light passing through the tissue at various locations. These point measurements must then be converted to the tissue optical properties μ_s' and μ_a. This conversion involves solving both the "forward problem" (how the light entering at each location scatters within the tissue) and the "inverse" problem (using the measured changes in light intensity and transit time to infer the location of dynamic absorbers and scatterers within the tissue). A depiction of this process is shown in Figure 2.3.

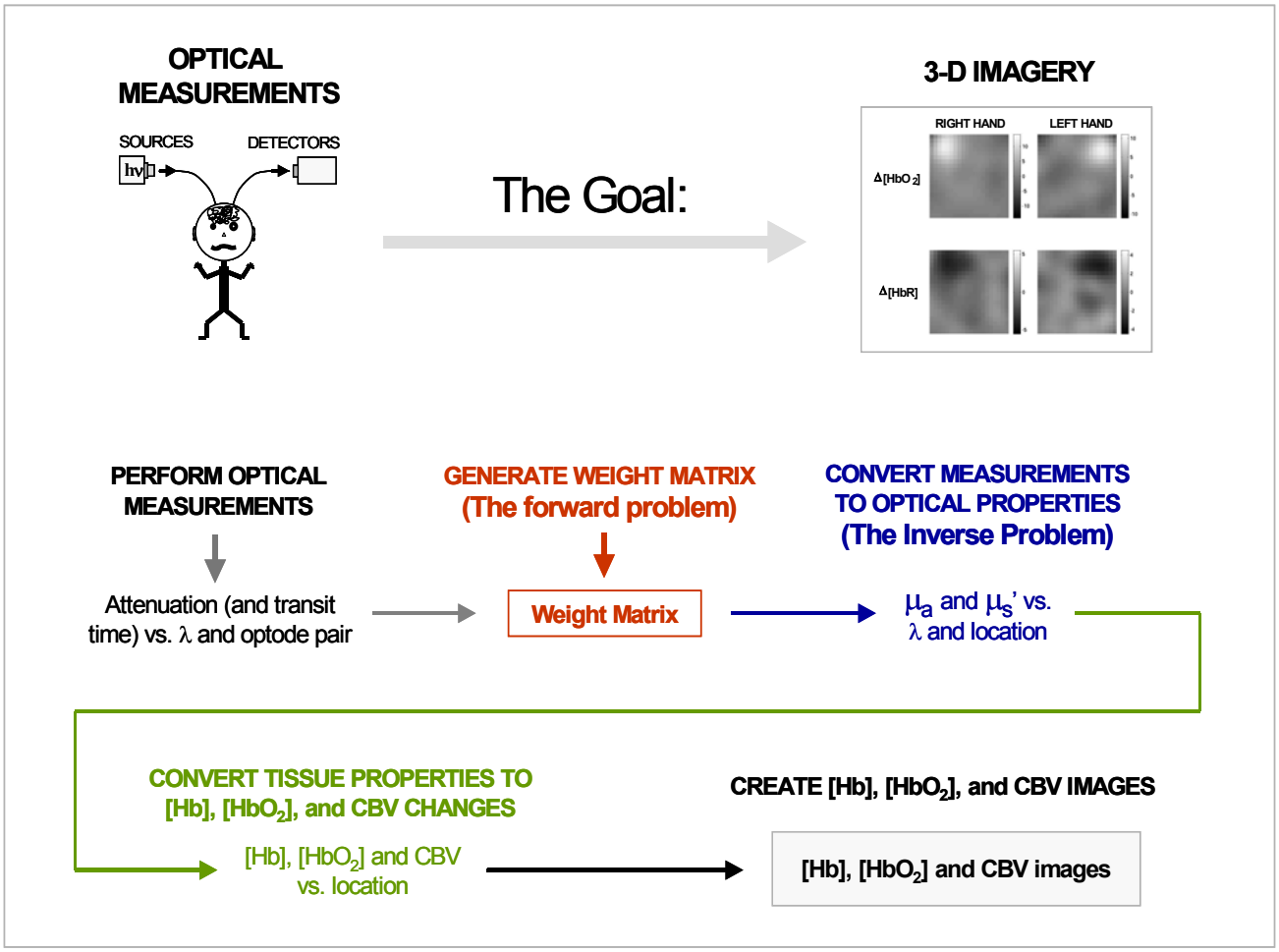

Figure 2.3. A diagram depicting the production of three-dimensional cerebral blood oxygenation and volume imagery from raw optical measurements collected by a DOT instrument. An accurate weight matrix is constructed for Andy's head using the optode locations from the measurement, and assumed optical properties for the scalp, skull, cerebrospinal fluid, and cortex. The weight matrix is then used to convert the raw optical attenuation (CW), or attenuation and transit time (RF and Time-domain) data into the wavelength-dependent tissue optical properties, μ_a and μ_s', as a function of location within the reconstruction volume. The optical properties are then used to determine blood volume and oxygenation changes, which are typically plotted as 2-D spatial images for a given depth. Since DOT measurements can be acquired rapidly, these data can be acquired at frame rates fast enough to capture the temporal evolution of the hemodynamic response.

2.2.1 DOT measurement techniques

A number of DOT measurement techniques have been developed and refined, each with its own set of design trades. The simplest technique involves performing an amplitude-only measurement, and is colloquially referred to as continuous wave, or "CW," to differentiate it from the "RF" technique, which measures both amplitude and transit time in the frequency domain using RF-modulated light sources. The "time-domain" technique uses sub-nanosecond optical pulses to measure both amplitude and transit time within the time-domain, hence its name.

CW instrumentation

Continuous-wave instruments are the easiest to build and operate. CW methods only measure optical attenuation, so there is no way to deconvolve the effects of both absorption and scattering, since both μ_s' and μ_a affect the amplitude of the detected signal. Changes in μ_a clearly affect the signal strength as described by Beer's law, however changes in μ_s' can also indirectly affect the signal strength: if μ_s' increases, then the mean path length of the light will also increase to some extent, thus increasing the pathlength term "L" in the Beer's law equation (eq. 1.3 on p. 14). The result is a decrease in signal strength, as depicted in Figure 2.4.

Thus, CW instruments measure changes in optical attenuation, which results from a combination of changes in μ_s' and/or μ_a. Although μ_a can be estimated if a reasonable value for μ_s' is assumed, this fundamental ambiguity limits the utility of CW instruments somewhat. However this disadvantage is offset by their simplicity and economy.

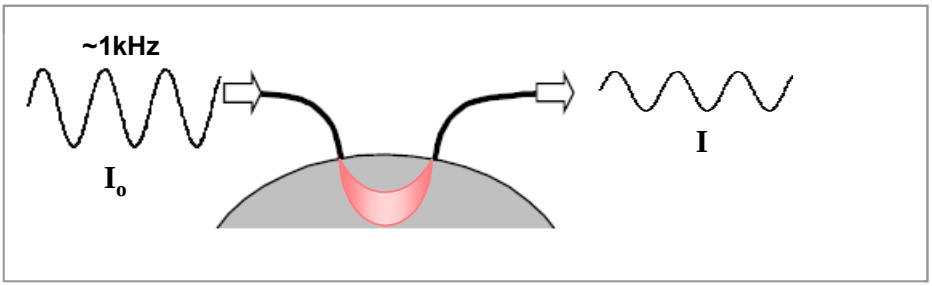

Figure 2.4. With a CW instrument, the light is modulated at an audio frequency to mitigate background flux and to reduce electronic noise through synchronous detection. The optical attenuation is then used to estimate μ_a, given an assumed value for μ_s'.

RF instrumentation

RF-modulated DOT instruments are much more complex to build than CW instruments, but they offer the capability to measure both amplitude changes and transit time changes in the light traveling between each optode pair. This provides two independent unknowns, which can then be used to extract both μ_s' and μ_a independently, as depicted in Figure 2.5. One disadvantage of RF measurements is that the transit time value represents a superposition of the transit times of all the photons arriving at the detector. Thus, only a single "mean transit time" value is actually measured at each optode pair, limiting the available information.

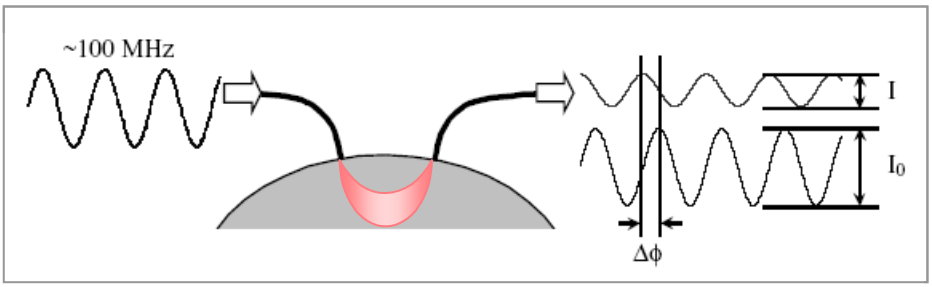

Figure 2.5. With an RF instrument, RF-modulated light enters the tissue and is attenuated and delayed by its diffusive transport to the detector optode. This transit time delay creates a measurable phase shift in the modulation envelope of the detected light relative to the light entering the tissue. The attenuation and mean transit time information is then used to calculate μ_s' and μ_a [43].

Time-domain instrumentation

Optical attenuation can be measured as a direct function of photon transit time using very narrow optical pulses in a manner analogous to LIDAR. Sub-nanosecond optical pulses produced by a fast laser diode or a mode-locked Ti-doped sapphire laser can be launched into the tissue. To the extent that the initial optical pulses resemble a delta function in time, they can be deconvolved from the detected signal to generate the temporal point-spread function (TPSF) of photon flux for each optode pair, as shown in Figure 2.6. These TPSFs provide a significant amount of information – more than is available from an RF measurement – which can be used to provide more accurate spatial estimates of both μ_s' and μ_a vs. λ and time.

Time-domain instruments are the most complex and expensive to construct, but they are inherently capable of providing the most detailed information on tissue optical properties.

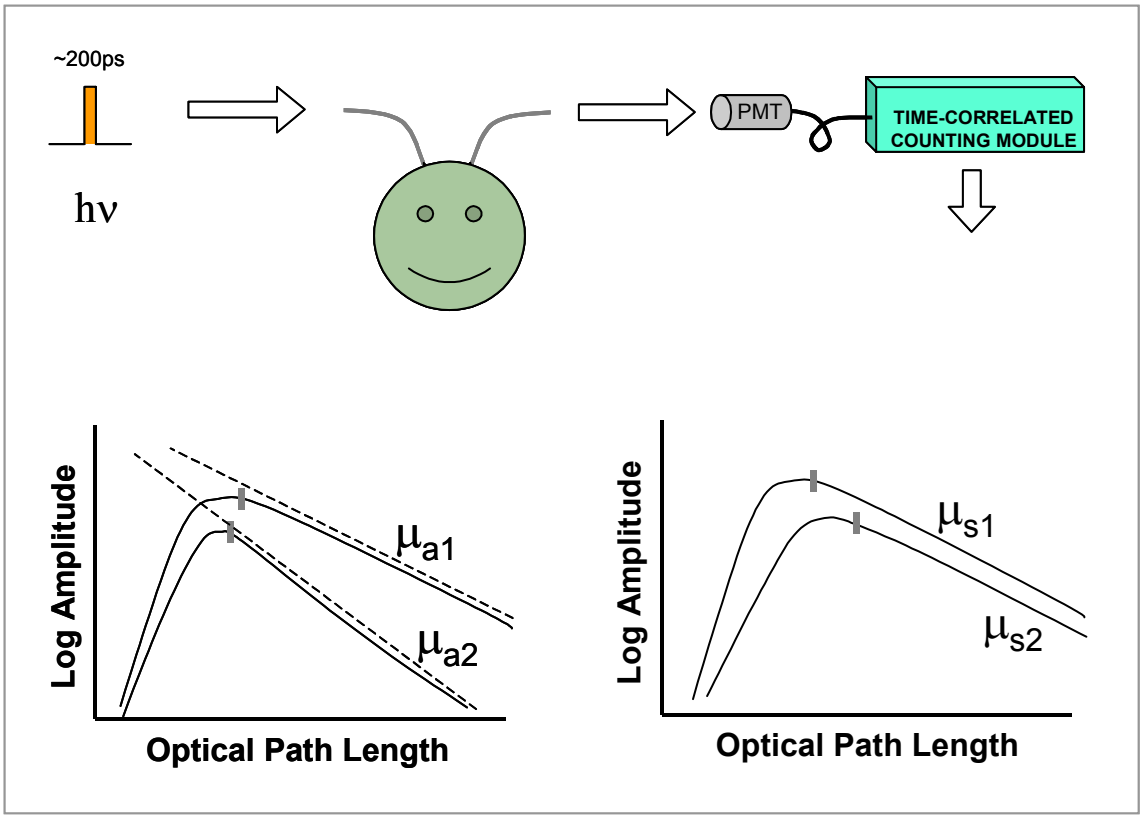

Figure 2.6. With a time-domain instrument, sub-nanosecond laser pulses enter the tissue and are simultaneously attenuated and broadened as a result of absorption and temporal dispersion due to scattering within the tissue. An increase in μ_a generates a steeper decay, but decreases the mean pathlength (=mean transit time) of the temporal point spread function slightly as the banana pattern shrinks in size and moves upward. An increase in μ_s' raises the mean pathlength, and the curve decreases due to an overall increase in attenuation from the additional time the photons spent within the tissue.

2.3 Optical scattering and absorption

In order to understand how images can be generated from diffusely scattered light, one must first understand how light propagates through scattering media. So the optical concepts of ***refraction, reflection, scattering,*** and ***absorption*** are introduced.

Basic optical concepts

Refraction is the deviation of a transmitted light ray upon crossing an abrupt (i.e. << 1λ thick) optical boundary (essentially an optical impedance mismatch) at any angle other than normal to the interface [44]. The angle of refraction can be easily calculated if the angle of incidence and the refractive indices of both optical media are known. When the boundary depth approaches a fraction of a wavelength in thickness, or if there is enough curvature of the boundary surface, then the calculations become more complex. Refraction occurs at every optical interface, and is the dominant source of scattering in materials with a small relative index mismatch or a low scatterer concentration, like tissue, milk, and fog.

Reflection is similar to refraction, but it refers to the portion of light that is directed away from the interface. Reflection is a symmetric process: the angle of incidence is identical to the angle of reflection [44]. The strength of the reflection is a strong function of both the magnitude and geometry of the index mismatch, and can even be eliminated (by phase cancellation). Reflection is the dominant source of scattering in high index-mismatch materials or those with high scatterer concentrations such as white paint (TiO_2 vs. binder), powdered sugar, cumulous clouds, and many refractory materials (magnesia, zirconia, alumina, silica, etc.).

Scattering in this context refers to any refraction or reflection that occurs either at an unpredictable location or deviates in an unpredictable direction. It can be thought of as "angular noise" in the light path of an optical system [44]. Individual photon paths cannot be predicted, but scattering as a whole can be modeled statistically if the initial conditions are known. Forward scattering refers to light scattered more towards the direction of propagation and backscattering refers to light scattered mostly away from the direction of propagation, back towards the light source. Isotropic scattering refers to light that has an equal probability of scattering at any angle.

Absorption is the attenuation of optical power within a material. Macroscopically, this appears as an exponential reduction of the optical intensity with distance traveled through the medium, since each unit length of the medium attenuates the optical power by a constant fraction.

In the time domain, circuitry in which voltage or current exhibits an exponential temporal decay is often characterized by a "time constant," which is defined as the time required for a given signal to decay to 1/e (~37%) of its original value. Although other criteria could also be used, such as the "half-life" used to characterize the decay rate of radioactive materials, the advantage of the time constant is a simplified conversion from the time domain to the frequency domain when calculating the bandwidth of electrical circuitry. In analogous fashion, the optical term for the "length constant" is called the "absorption length," and is defined in exactly the same manner.

2.3.1 Forms of optical scattering

Elastic scattering

Elastic scattering is the dominant scattering mechanism in diffuse imaging. It is caused by spatially abrupt (shorter than a λ) changes in refractive index (optical impedance mismatches). The term "elastic" refers to the fact that the photon energy (i.e. wavelength) is preserved – the scattered photons are at the same wavelength (and hence photon energy) as they were prior to the scattering event [44]. The average angular deviation of elastic scattering is a function of the particle size, shape, and index difference. Mie and Rayleigh scattering are two models for elastic scattering. Elastic scattering is the

mechanism responsible for standard epitaxial and transillumination optical microscopy, as well as DOT.

Inelastic scattering

Inelastic scattering is a result of a nonlinear optical process, and normally occurs at very high flux densities not commonly achieved in DOT. In a nonlinear optical interaction, the photons interact as waves, and can undergo mixing, frequency multiplication, and parametric amplification, and are subject to the same properties and constraints as in the RF domain. Brillouin and Raman scattering are examples of inelastic scattering [44]. Inelastic scattering always involves a change in the photon energy, and hence wavelength.

The Optical Doppler effect involves a form of nonlinear scattering in which the energy difference is provided by the change in photon velocity due to relative axial motion of the scatterers. The frequency shifts are quite small, and are usually detected through self-interference – a variant of homodyne detection – resulting in intensity fluctuations in the scattered optical signal. Doppler scattering is used to measure blood flow, and as a means of generating a carrier frequency for performing optical coherence tomography.

Fluorescence

Fluorescence is a special form of inelastic scattering in which the incident photons induce energy shifts in certain molecules, termed fluorophores, which then re-emit photons of lower energy. Fluorescent scattering, by definition, always involves a shift in the photon energy [44]. In most biomedical applications the re-emitted photon is at a lower energy, a consequence of unavoidable energy losses within the fluorophore. Since the absorption and emission of photon energy destroys the vectorial information of the incident photons, fluorescent emission is isotropic. As a result, isotropically emitting fluorophores have been used as molecular "beacons," flashing out their location even when located deep within living tissue [45].

The excitation wavelengths for fluorophores in biological systems can range from near-infrared to ultraviolet, with the absorption depth decreasing accordingly. Novel techniques, such as multi-photon excitation, can help solve this problem, since the infrared excitation wavelength penetrates tissue far more deeply than light at visible or ultraviolet wavelengths.

2.3.2 Quantifying optical absorption

The Beer-Lambert law

The Beer-Lambert law provides a simple and reasonably accurate model for bulk absorption in isotropic nonscattering media, and is given by:

$$\ln\left(I_0 / I\right) = \sigma\rho d \quad \text{or} \quad \log_{10}\left(I_0 / I\right) = \varepsilon C d = absorbance(OD) \tag{2.1}$$

where: I_0 is the incident intensity,
I is the transmitted light intensity,
σ is the absorption cross section,
ρ is the number density of the absorbing molecules,
C is the concentration of the absorbing molecules (mM),
d is the pathlength (cm),
and ε is the extinction coefficient for a solution ($molar^{-1}\ cm^{-1}$) [44].

However, due to its simplicity, the Beer-Lambert law is accurate only if the following conditions are met:

- The absorbing medium must be homogeneous
- The light source must be monochromatic
- Scattering and reflection must not occur

If the absorption cross-section of the medium varies within the optical pathlength, then σ must be expressed as a function of depth: σ(d). If the light source is polychromatic, then σ must also be expressed as a function of wavelength: σ(d, λ) [46].

The effect of chromophore concentration on optical absorption

With any optical measurement of tissue, a change in the chromophore concentration causes the detected intensity to change. If the hematocrit is assumed to be constant and the subject remains still during the measurement, ε and L can remain fixed. B and G can also be held constant for similar reasons.

Thus, for DOT, the change in optical density can be rewritten as:

$$\Delta OD = -\ln \frac{I_{final}}{I_{initial}} = \sum_i \varepsilon_i \Delta C_i LB \qquad 2.2$$

By considering the contribution of only 2 dynamic chromophores: deoxyhemoglobin (Hb) and oxyhemoglobin (HbO_2), the above equation becomes:

$$\Delta OD^{\lambda} = \left(\varepsilon^{\lambda}_{HbO}\Delta[HbO] + \varepsilon^{\lambda}_{Hb}\Delta[Hb]\right)\cdot L \cdot B^{\lambda} \qquad 2.3$$

[46]

2.3.3 Scattering-absorption interaction

When enough scattering occurs such that the scattering length approaches the absorption length in the medium, then a pathlength correction factor must be applied. Scattering in tissues is commonly expressed in terms of the reduced scattering coefficient:

$$\mu_s' = \mu_s(1-g) \qquad 2.4$$

where μ_s is the scattering coefficient and g is the scattering anisotropy factor, equal to the average cosine of the single scattering phase function:

$$\mathbf{G} = \int_{4\pi} p\cos\theta \, d\Omega \qquad 2.5$$

Thus, if the majority of the light experiences forward-directed scattering, the average optical pathlength is nearly the same as in the unscattered case, and the effect of tissue scattering is "reduced" accordingly. Then the g value is close to unity, and $\mu_s' < \mu_s$. This approximates standard Beer's law attenuation. If, on the other hand, the scattering is nearly isotropic, then g is closer to zero and $\mu_s' \sim \mu_s$.

To address the scattering that occurs with DOT, a modified form of the Beer-Lambert law is used:

$$OD = -\ln \frac{I}{I_0} = \sum_i \varepsilon_i C_i LB + G \qquad 2.6$$

B is a pathlength factor that corrects for increases in the photon pathlength caused by tissue scattering, and G is the measurement geometry factor [46].

The pathlength correction factor

The pathlength correction factor compensates for the additional distance traveled by the diffusing light as it travels between the source and detector optodes. When the scattering length is shorter than the absorption length, as is the case in tissue at near-infrared wavelengths, then light "diffuses" down the flux gradients within the medium, and the radiative transport equation can be approximated by the diffusion equation [46]:

$$-D\nabla^2\Phi(r,t) + c\mu_a\Phi(r,t) + \frac{\partial\Phi(r,t)}{\partial t} = cS(r,t) \qquad 2.7$$

Where: $\Phi(r,t)$ is the photon fluence at position r and time t,

$S(r,t)$ is the source distribution of photons,

$D \approx 1/3\mu_s'$ is the photon diffusion coefficient (μ_s' is the reduced scattering coefficient),

c is the speed of light in the medium,

and μ_a is the absorption coefficient.

Since the refractive indices of the tissue (~1.34) and the air (~1.00) are different, a fraction of the radiation coming from inside the tissue will be reflected back into the turbid medium. The apparent origin of the light can then be approximated by a point source placed at one scattering mean-free path below the surface (i.e. $-1/\mu_S'$) [46]. It has been shown that by setting the fluence due to a point source equal to 0 on an extrapolated boundary, the condition for a mismatched boundary can be satisfied [46]. The fluence due to a point source inside a semi-infinite medium can be forced to 0 by introducing a negative "image source" of light above the boundary, as illustrated in Figure 2.7.

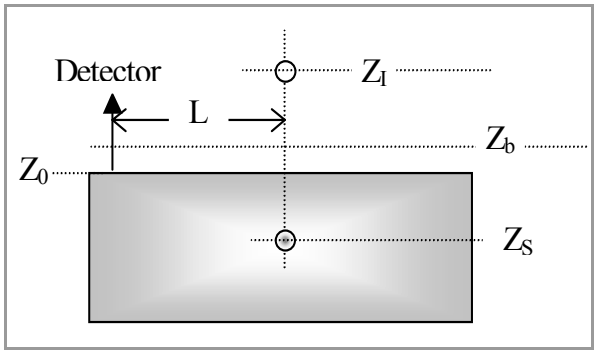

Figure 2.7. The apparent origin of light exiting a diffuse medium can be modeled as appearing one scattering length below the tissue/air interface. z_S is the location where the collimated light source becomes diffuse inside the medium. z_I is the location of the image source used to satisfy the semi-infinite boundary condition. L is the displacement between the collimated light source and the detector, and z_b is the position of the extrapolated boundary.

The fluence for a single source in a semi-infinite medium is given as the superposition of the fluences for the source and the image source calculated in an infinite medium. For a semi-infinite medium, the solution of the diffusion equation becomes [46]:

$$\Phi(r,t) = \frac{c \cdot S}{4\pi D}\left[\frac{\exp(-\sqrt{3\mu_s'\mu_a}\cdot r_1)}{r_1} - \frac{\exp(-\sqrt{3\mu_s'\mu_a}\cdot r_2)}{r_2}\right] \qquad 2.8$$

Where:

$$r_1 = \sqrt{L^2 + (z_0 - z_S)^2} \qquad r_2 = \sqrt{L^2 + (z_0 + 2z_b + z_S)^2}$$

If we set $(z_0 - z_S)$ to be equal to z, and $(z_0 + 2z_b + z_S) = z_i$, then we can rewrite the diffusion equation as:

$$\Phi(r) = \frac{S\exp\left[-(3\mu_s'\mu_a)^{1/2}(L^2+z^2)^{1/2}\right]}{4\pi D(L^2+z^2)^{1/2}} - \frac{S\exp\left[-(3\mu_s'\mu_a)^{1/2}(L^2+z_i^2)^{1/2}\right]}{4\pi D(L^2+z_i^2)^{1/2}} \qquad 2.9$$

If we also assume that L>>z, and L>>z_I (where z and z_I are the positions of the diffuse light source inside the medium and the image source, respectively), then we can employ the Taylor expansion to obtain:

$$\Phi(r) = \frac{S\exp\left[-(3\mu_s'\mu_a)^{1/2}L\right]}{4\pi DL^2} A\left[(3\mu_s'\mu_a)^{1/2} + \frac{1}{L}\right] \qquad 2.10$$

Here A is a constant that depends on the background optical properties and the index mismatch at the air-tissue interface. Assuming that the change in the chromophore concentration is insignificant (i.e. if $\Delta C/C \ll 1$), then we can make another Taylor expansion to obtain the following expression for the change in optical density:

$$\Delta OD = -\ln\frac{\Phi_{final}}{\Phi_{initial}} = \left(\varepsilon_{HbO}\Delta[HbO] + \varepsilon_{Hb}\Delta[Hb]\right)\cdot L \cdot B = \\ = \frac{1}{2}\left(\frac{3\mu_s'}{\mu_a^{initial}}\right)^{1/2}\left[1 - \frac{1}{1 + L\left(3\mu_s'^{(initial)}\mu_a^{initial}\right)^{1/2}}\right]\left(\varepsilon_{HbO}\Delta[HbO] + \varepsilon_{Hb}\Delta[Hb]\right)L \qquad 2.11$$

So for a semi-infinite medium the pathlength factor B, is given by:

$$B = \frac{1}{2}\left(\frac{3\mu_s'}{\mu_\alpha^{initial}}\right)^{1/2}\left[1 - \frac{1}{1 + L\left(3\mu_s'^{(initial)}\mu_\alpha^{initial}\right)^{1/2}}\right] \qquad 2.12$$

This shows that B is a function of the tissue absorption, scattering, and the optode separation. Therefore B is wavelength dependent. Although in practical clinical measurements this dependence is ignored, and B is determined empirically [46].

2.3.4 The three categories of light propagation

The optical properties of biological tissue can vary greatly. In relatively transparent materials such as corneal tissue or cerebrospinal fluid, μ_a greatly exceeds μ_s, and light travels ballistically. In slightly turbid materials, μ_a is comparable to μ_s, and the average photon experiences a few scattering events before it is detected. In strongly turbid materials, μ_s greatly exceeds μ_a, so light travels in a diffuse manner with each photon experiencing many scattering events prior to detection.

Ballistic or "specular" light

The word "ballistic" is used here to imply travel in an inertial manner, unaffected by external forces. All ballistic photons travel along straight lines, as would a projectile exiting a gun barrel or a rocket entering deep space. Ballistic transport implies that both refraction and reflection occur at specific, predictable locations such that an image could be preserved. Mirrors and lenses are designed using ballistic photon transport models. Our binocular vision relies upon ballistic photon transport.

Ballistic models are very simple, since ray paths and pathlengths can be clearly defined. The most common ballistic optical model is Snell's Law, which describes the angle of refraction of light crossing a planar interface created by an optical index mismatch. Beer's Law, which describes the effects of optical absorption, also assumes ballistic transport.

Singly and "fewly" scattered light

This is also referred to as "snake" light. Each photon scatters at least once between the source and detector. The single/few scattering event regime is difficult to characterize. Since the actual ray paths are uncertain, Snell's law cannot be used here, but Beer's Law works reasonably well for predicting attenuation, since the scattering anisotropy ("g") is very high. Yet the scattering is too weak to be accurately modeled with the diffusion equation. The average optical pathlength of snake light is slightly longer than ballistic, and has a statistical distribution with both a narrow temporal and angular profile. Snake light is often encountered in ballistic imaging, where it reduces image contrast.

Multiply scattered or diffuse light

Multiply scattered light is easier to model than snake light, and is best treated statistically, using either the diffusion equation or Monte Carlo simulations – much like heat flow or neutron flux. Pathlengths are completely unpredictable and span a very broad range. Because scattering is so frequent, Beer's Law must be modified to account for the effects of scattering anisotropy, which will affect the mean photon pathlength.

2.3.5 Partial-volume effects

Many DOT systems use two or more source wavelengths to measure the differential spectral absorption in the near-IR band in order to calculate the tissue hemoglobin and deoxyhemoglobin concentrations. One of the problems with differential spectral absorption measurements in scattering media is the "partial-volume" effect. Since each source wavelength is different, each will have a

different μ_a value. As a result, the optical power distribution within the interaction region (i.e. the PDF, or "banana" pattern) for each wavelength will be different. The larger the μ_a, the shallower and narrower the pattern will be, and vice-versa.

This can be understood intuitively. The smaller μ_a, the more the pattern will resemble an unattenuated photon migration profile, with sensitivity extending down to a significant depth. But as tissue absorption increases, μ_a grows, and photons traveling longer than a certain mean free path will be too weak to be detected by the DOT instrument. Thus less and less of the light that travels most deeply will be detected. Therefore, as μ_a increases, this mean free path cutoff decreases, and the banana pattern will shrink to conform to a shorter average pathlength. This means a shift towards a shallower and narrower shape.

The problem, which only arises when trying to perform a multispectral attenuation measurement, is the mismatch in the volume distribution of both banana patterns. Other than at isosbestic points, the μ_a values for Hb and HbO_2 differ to a varying degree across the optical spectrum. In addition, when the blood oxygenation changes, so do both μ_as, and thus so will the geometry of both banana patterns.

Since the spectral absorbance curves are monotonic within most of the near-IR spectral region, the geometric distortion should also be monotonic. So, in principle, it can be modeled and then corrected for if the initial conditions are known, or if enough spatially independent measurements through the same region are taken.

One way to combat this problem is to choose wavelengths that more closely straddle an isosbestic point in the Hb/HbO_2 absorption profile. This way, blood oxygenation changes will cause smaller shifts in both μ_as, thus minimizing the partial-volume effect. What then results is the need to accurately measure a small difference between two or more comparatively large numbers. This places serious performance demands on the measurement system. These include a large dynamic range (to resolve small changes between large flux levels), excellent channel-to-channel gain uniformity and linearity (so that interchannel errors combined with biogenic volume fluctuations do not swamp the measurement), and low interchannel drift. The price that is paid for minimizing the partial-volume error using this method is a reduction in both the accuracy and sensitivity of the oxygenation measurement.

A better solution would be to develop an accurate model of the geometric distortion vs. both wavelength and μ_a for each type of tissue, and then use this model to correct or "back out" the partial-volume effects from the measurements. Although this model would be difficult to develop, it would ultimately permit a larger source wavelength spread about the isosbestic point. This means that μ_a changes would be larger. Thus the oxygenation and volume measurements would be more accurate, and the dynamic range requirement for the hardware could be lowered as well.

The tissue optical properties could be calculated in real-time, using the information gathered by many spatially independent measurements through the same tissue region. This would require a significant amount of computing power, but would avoid the uncertainties of having to predict them ahead of time. This "dynamic" reconstruction technique offers the advantages of flexibility and convenience, however the fidelity of the reconstructed imagery is dependent on the multiplicity and quality of the measured data. Thus numerous semi-redundant measurements must be taken and dense optode fields would be required.

2.4 Modeling light propagation in diffuse media

There are two schools of thought concerning how to model optical scattering in diffuse media. One involves the use of first principles to derive the scattered electromagnetic field, given full knowledge of both the incident field and the composition and shapes of all the scatterers. Another treats light in

bulk form, as optical flux that diffuses from regions of higher flux density to regions of lower flux density, according to standard diffusion theory.

Both approaches have their merits. The "first principles" approach is based upon single particle scattering theory, and is ideal for well-characterized scattering environments, those with few scatterers, and those in which single scattering predominates. The "diffusive transport" approach is based upon diffusion theory, and is well suited for situations where multiple scattering predominates and where the specific nature and shape of the scatterers is unknown. Since the latter is often the case with biological tissues, DOT primarily relies upon diffusion theory for modeling optical transport through tissue.

2.4.1 Intuitive optical transport models

Reflection, refraction, and diffraction are all descriptions for different macroscopic effects of light-matter interactions. In a general sense, all of these involve elastic optical scattering. What differs is the nature and form of the scatterers. If the scatterer is macroscopic (in relation to a photon) and fixed in dimension and composition, then the "boundary" concepts of reflection and refraction and the "wave" concept of diffraction can be applied. If the surface of the scatterer has an uncertain geometry in either time or space (i.e. surface roughness, small size, Brownian motion, turbulence, etc.), then the nature of this scattering cannot be modeled directly. In these cases, macroscopic measures of the system can be used to perform statistical calculations predicting the probable photon distribution – essentially profiling the photonic odds.

When light is initially launched into an uniformly turbid medium, the spatial flux distribution is anisotropic, still resembling the angular emission profile of the fiber (but with a slight refractive collimation due to the tissue index, which is about 34% larger than air). After the light travels a few scattering mean-free-pathlengths within the tissue, the angular flux distribution begins to evolve into a more isotropic pattern. The optical distance required for this to occur is a function of the angular scattering nature of the particles, typically expressed as the "g" value discussed below. Once a near-isotropic flux density is established, the only remaining anisotropy is due to the overarching macroscopic (mm to cm scale) flux gradient within the tissue. Now the photons can be thought of as flowing "en-masse," diffusing through the media down the flux gradient, much as heat flows down a thermal gradient. Diffusive transport has been studied extensively, and the same theoretical models can be applied to photons in a scattering medium.

Fresnel reflection

Fresnel reflection at the optode/tissue interface can affect both the intensity and the angular distribution of light entering and exiting the tissue [44]. The indices of refraction of tissue and silica are ~1.34 and ~1.55 respectively. Any air (n=1.00) present between the fiber and the scalp provides two surfaces across which an index mismatch can occur. Both the fiber/air and air/tissue index mismatches create about a 7% Fresnel loss, however this can be decreased to only about 3% by applying index-matching fluids, such as hair gel, to the optode assembly or to the scalp. Alternately, index-matching may be provided by moisture from natural perspiration.

Index matching fluids can be transparent if the air gap between the optode/tissue interface is thin, (i.e. less than $1/\mu_s$), but if the air gap is larger than ~1mm then the scattering coefficient of the fluid should approach that of the tissue to avoid altering the boundary conditions which could complicate image reconstruction.

2.4.2 Single particle scattering models

The two most commonly used models are the Rayleigh and Mie scattering models. Rayleigh-Gans-Debye (RGD) scattering is a variant of Rayleigh scattering, but it can only be applied if serious constraints on the nature of the scattering media are adhered to.

Rayleigh scattering

Rayleigh derived his theory of light scattering from an intuitive sense. The conclusion he reached was that for particle diameters much smaller than λ, the scattered intensity I_s is:

$$I_s \alpha \frac{V^2}{r^2 \lambda^4} \qquad 2.13$$

Where: V is the particle volume,
r is the distance of the particle from the point of observation,
and λ is the optical wavelength. [47]

The most significant feature of Rayleigh scattering is the λ^4 relationship – shorter wavelengths experience significantly greater scattering in the Rayleigh regime. This is why the sky appears blue, and not white or yellow. Gas molecules are far smaller than .5um, and so act as Rayleigh scatterers. In fact, if it were not for the higher atmospheric absorption in the violet, the sky would appear purple. Clouds, however, are formed from water droplets or ice crystals that are typically far larger than 1um, and so act as Mie scatterers. This is why clouds appear white.

Rayleigh-Gans-Debye (RGD) scattering

The major advantage of RGD theory is that it can be used with nonspherical particles, however it can only be applied in the case of weak scatterers whose refractive index is close to that of the surrounding medium. This results in minimal reflection at the particle/medium interface, which leads to a field pattern within the particle that is nearly identical to that within the surrounding medium [47].

Mie scattering

Mie scattering is based upon the assumption that the scattering particles are spherical. Although this is an ideal model for water droplets in air (i.e. clouds), it is less applicable to tissue, due to the wide variety of scatterers present. Mie scattering characteristics are a function of the ratio of the particle radius to the optical wavelength and the ratio of the refractive index of the particle to the surrounding medium. Mie theory is valid for all sizes of spherical particles, and at very small sizes (<1/4 wavelength), the Mie equations simplify to only a few terms, which are similar to the Rayleigh scattering equations for non-absorbing particles. A typical angular scattering function is shown in Figure 2.8 (c) for a particle with n = 1.40 and radius of 2.76 μm in water (n = 1.33) illuminated at 800 nm. Note the strongly forward biased (i.e. anisotropic) nature of the scattered light at this radius to wavelength ratio [47].

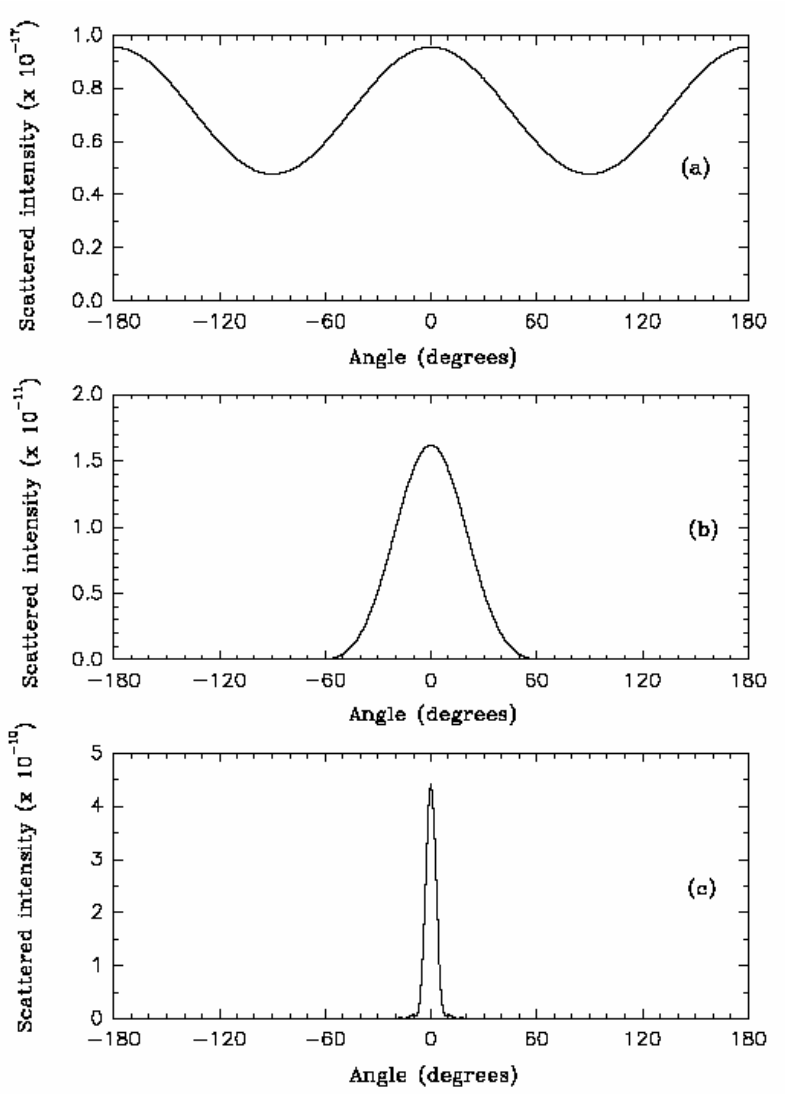

Figure 2.8. Angular scattering models for a single spherical particle. Rayleigh scattering best applies to particles with a diameter (d) far less than the illuminating wavelength λ, shown in (a). Rayleigh-Gans-Debye (RGD) scattering, a variant of Rayleigh scattering is shown in (b). The peak amplitude is lower than the Mie case, but the scattered energy is distributed over a larger solid angle. Mie scattering for particles with $d > \lambda$ is shown in (c). Note the factor of 10^7 difference in scattering crossection and the much narrower angular subtense than both the Rayleigh and RGD cases [47].

2.4.3 Multiple particle scattering models

Basic diffusion theory

Diffusion theory is well-suited for modeling optical transport through tissue, where scattering predominates and the nature and geometry of the scatterers on a subcellular scale is never known. A simple expression for the time independent diffusion equation is:

$$\nabla^2\phi - \frac{\phi}{\Gamma^2} = S$$

where: ϕ is the space irradiance,
Γ is the penetration depth,
and ***S*** represents the source.

The solution for a spherically symmetric geometry is:

$$\phi = \phi_{OS} \left(\frac{\Gamma}{r}\right) e^{-\frac{(r-\Gamma)}{\Gamma}} \qquad 2.15$$

where: ***r*** is the distance from the source
and ϕ_{os} is the irradiance at one penetration depth.

The penetration depth Γ is given by:

$$\Gamma = \frac{1}{\sqrt{3\mu_a\left(\mu_a + (1-g)\,\mu_s\right)}} \qquad 2.16$$

where: μ_a is the transport absorption coefficient in the tissue,
μ_s is the transport scattering coefficient,
and ***g*** is the anisotropy factor. (The term $(1 - g)\,\mu_s$ is the reduced scattering coefficient: μ_s'.) [47]

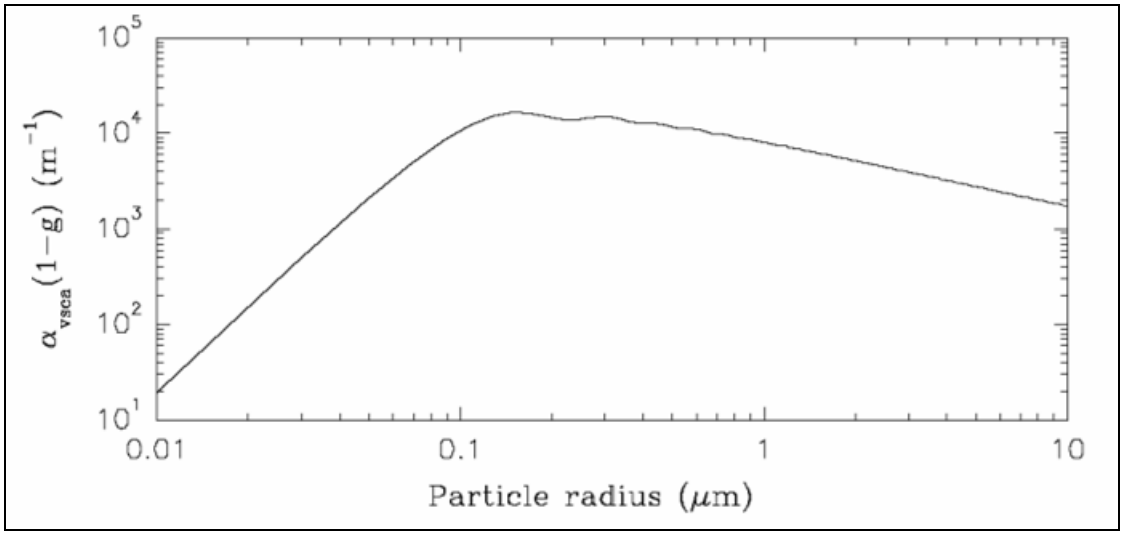

Figure 2.9. A plot of effective multiple scattering efficiency (the product of the volume scattering efficiency α_{vsca} with an anisotropy correction term) vs. particle size, assuming a refractive index ratio of 1.045 and a source wavelength of 800nm. This suggests that particles in the general size range of 0.1um to 1um would be the most efficient scatterers in biological tissue [47].

Empirical measurements of optical extinction

An example of an empirical extinction measurement is shown in Figure 2.10. This measurement reveals a bimodal drop in intensity, which matches diffusion theory predictions. The initial decrease is

very rapid as the highly anisotropic collimated input beam scatters into an isotropically diffusing beam within a depth of about 5mm. Once the light has become diffuse, further attenuation is through absorption by tissue chromophores. The penetration depth is defined as the depth at which the irradiance drops to 37% (i.e. 1/e) of its initial value, and is a function of the tissue optical properties and the wavelength of the incident light.

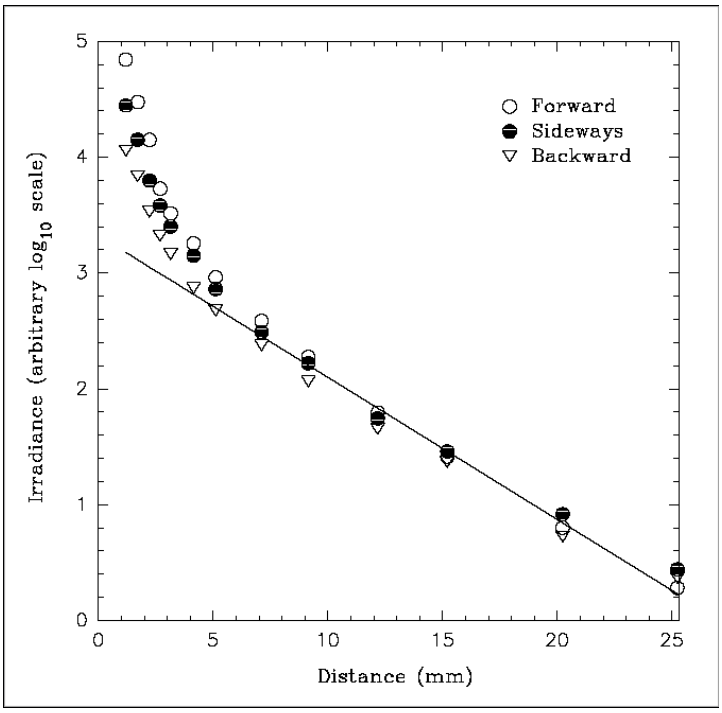

Figure 2.10. Irradiance vs. depth in post-mortem neonatal brain tissue. λ = 660nm [47, 48].

The radiative transport equation

The radiative transport equation models the energy balance inside a volume element of the scattering medium. For conventional DOT, only elastic scattering and absorption are modeled, since Hb and HbO_2 act as simple chromophores in the near-IR spectral band. [Other DOT techniques involving fluorophores such as ICG must be handled differently.] The transport equation can be obtained by considering the total space and time variation of the specific intensity along a direction $\hat{s}$ in an elementary volume and making this equal to the variation of specific intensity due to scattering and absorption inside the medium. The equation for the time-dependent case is given by:

$$\frac{1}{\upsilon}\frac{\partial}{\partial t}I(\quad\hat{}) + \hat{s}\cdot\nabla I(\quad\hat{}) = -\mu_t I(\quad\hat{}) + \frac{\mu_t}{4\pi}\int_{4\pi} p(\hat{s},\hat{s}')I(\quad\hat{}\,)d\hat{s}' + S(\quad\hat{}) \qquad 2.17$$

where: $I(\mathbf{r}, t, \hat{s})$ represents the specific intensity in W/m^2/sr,
υ is the speed of light inside the diffusing medium,
μ_t (= μ_s + μ_a) is the extinction coefficient,
$S(\mathbf{r}, t, \hat{s})$ represents the source spatial and angular distribution in W/m^3/sr,
and $p(\hat{s},\hat{s}')$ is the scattering function that defines the probability of a photon moving in the direction $\hat{s}$ to be scattered into direction $\hat{s}'$ [35].

The diffusion approximation to the radiative transport equation

The radiative transport equation is complex, and analytical solutions with sufficient generality to solve real problems are difficult to obtain, so numerical methods or analytical approximations are generally used. An analytical model commonly used for DOT is the diffusion approximation.

The radiative transport equation can be simplified through a spherical harmonic expansion in s. The result is a set of $(N+1)^2$ coupled partial differential equations, collectively known as the P_N approximation. For odd N, such equations may be reduced to a single (N+1)th-order differential equation. This "P_1" approximation to the radiative transport equation is generally referred to as the diffusion approximation, and is given by:

$$\left(-D\nabla^2 + \upsilon\mu_a + \frac{\partial}{\partial t}\right)\Phi(\quad) = \upsilon S(\quad) \qquad 2.18$$

where: $D = \upsilon/(3\mu'_s)$ is the photon diffusion coefficient in cm^2/s,

$\mu_s' = \mu_s(1-g)$ is the reduced scattering coefficient in cm^{-1},

$\Phi(\quad)$ is the photon fluence rate in W/cm^2,

and $S(\quad)$ is the source term in W/cm^3.

The P_1 approximation is valid when:

- The albedo $c = \mu_s/(\mu_s + \mu_a)$ is close to unity, when $\mu_a << \mu_s'$ (a reasonable assumption for light scattering in tissue)
- The normalized scattering function is not too anisotropic
- The source-detector separation is large compared to the random walk length: $1/\mu_s'$.

The diffusion coefficient has been defined as: $D = \upsilon/3(\mu_s' + \mu_a)$, however better agreement is obtained when the diffusion coefficient independent of μ_a is assumed (i.e. $D = \upsilon/3\mu_s'$), which is consistent with the initial assumption that $\mu_a << \mu_s'$ [35].

2.5 Sources of optical absorption in tissue

Chromophores (molecules which absorb light) are critical to DOT, since they are responsible for modulating the intensity of the incoming light which generates the optical signals used to reconstruct optical tomographic imagery. Some chromophores are present in relatively fixed concentrations, while others vary as a function of physiologic changes.

2.5.1 Static absorbers

Static absorbers are chromophores whose concentrations, and hence their absorbances, remain fixed within the duration of a DOT measurement. Figure 2.11 shows the molar extinction coefficients of the most common chromophores in human tissue. Water, melanin, and lipids are the main tissue absorbers. Other chromophores, such as bilirubin and myoglobin, are present at such low concentrations that their attenuation can often be ignored.

Melanin is the dominant static absorber in the visible and UV region

Melanin is found primarily in skin and hair, and presents a significant source of attenuation for any transcutaneous optical measurements. Although skin color does affect optical throughput, hair color

usually exerts a stronger effect, since the roots penetrate deep into the cutaneous tissue. Thus, light-haired individuals are preferred for cortical DOT measurements.

Water dominates in the near-infrared region, from ~600nm to beyond 1um

The human brain is composed of about 80-90% water [49, 50]. Although melanin absorption dominates between ~300nm to ~1000nm, outside of this range the water absorption dominates. Measurements below 300nm and greater than 900nm become increasingly difficult due to water absorption.

Lipids

In brain tissue, about 8% of gray matter and 17% of white matter is composed of various lipids [50]. The spectral absorption of lipids is relatively uniform and approximates that of water. Since it is present in relatively low concentrations, its total absorbance is less than that of tissue water.

Bilirubin, myoglobin, and other chromophores

The absorption spectrum of blood plasma in the visible band is dominated by bilirubin. Bilirubin is normally present in the blood at around 2.5mg/dl, although it can exceed 20 mg/dl in severe cases of jaundice. It has a relatively low and flat spectral absorption curve in the NIR band at normal plasma concentrations, so it can be safely ignored in most cases [47].

Cerebrospinal fluid

Cerebrospinal fluid surrounds the brain and much of the spinal cord. It is a transparent fluid that contains less protein, bilirubin, and other chromophores than blood plasma, and is optically very similar to water. For most DOT applications, CSF can be modeled as water.

Carboxyhemoglobin

Some carboxyhemoglobin (HbCO) is naturally produced as a result of certain metabolic processes, and constitutes from 2% to as much as 10% (in smokers) of the total circulating hemoglobin. It is nearly transparent in the NIR spectral region, and thus has no effect on DOT measurements (other than altering the "optical" hematocrit when present at high concentrations) [47]. HbCO is incapable of transporting oxygen, and high HbCO concentrations (from exposure to airborne CO) can lead to hypoxic coma and death [51]. The dissociation half-life of HbCO while breathing air (with 21% oxygen, at sea level) is about 4 hours. Higher oxygen tensions can reduce this half-life substantially, which is why victims of CO intoxication are passively "decarboxylated" in hyperbaric chambers equipped with a 100% oxygen breathing mixture.

Methemoglobin

Methemoglobin (Hi) constitutes from 0.5% to 2.5% of the circulating hemoglobin [50]. It is created from regular hemoglobin by increasing the oxidation state of the iron in the heme structure from the normal divalent ferrous (Fe^{2+}) form to the trivalent ferric (Fe^{3+}) form. A small amount of Hi is produced through normal metabolic pathways, and it is rapidly formed through contact with nitrite anions and organic nitrate esters in blood plasma. It is a poor transporter of oxygen, and is usually considered undesirable when present in amounts greater than a few percent. [Except in cases of cyanide exposure, when methemoglobinemia is intentionally induced to reactivate the cytochrome enzymes by drawing the cyanide anions away from the ferric cytochrome iron and on to the ferric methemoglobin iron.] Chronic cigarette smokers often have reduced Hi and elevated HbCO levels as a result of the cyanides and CO present in cigarette smoke.

Hi has a spectral absorption curve that is similar in shape, but more absorbing, than HbO_2, so it should be accounted for if accurate modeling is desired. Its absorption spectrum is pH-sensitive,

however the normal physiological pH range is quite narrow and very stable, so this should not be an issue [47].

Sulfhemoglobin

Sulfhemoglobin (SHb) is not normally present in circulating hemoglobin. It does not transport any oxygen, and is caused by exposure to certain thio- compounds such as H_2S. SHb appears to have a strong absorption in the NIR region, however its presence in the bloodstream is usually a sign of serious pathology or toxicity, and would not be present in normal DOT subjects. The spectral absorption of a number of hemoglobin variants is shown in Figure 2.14 [47].

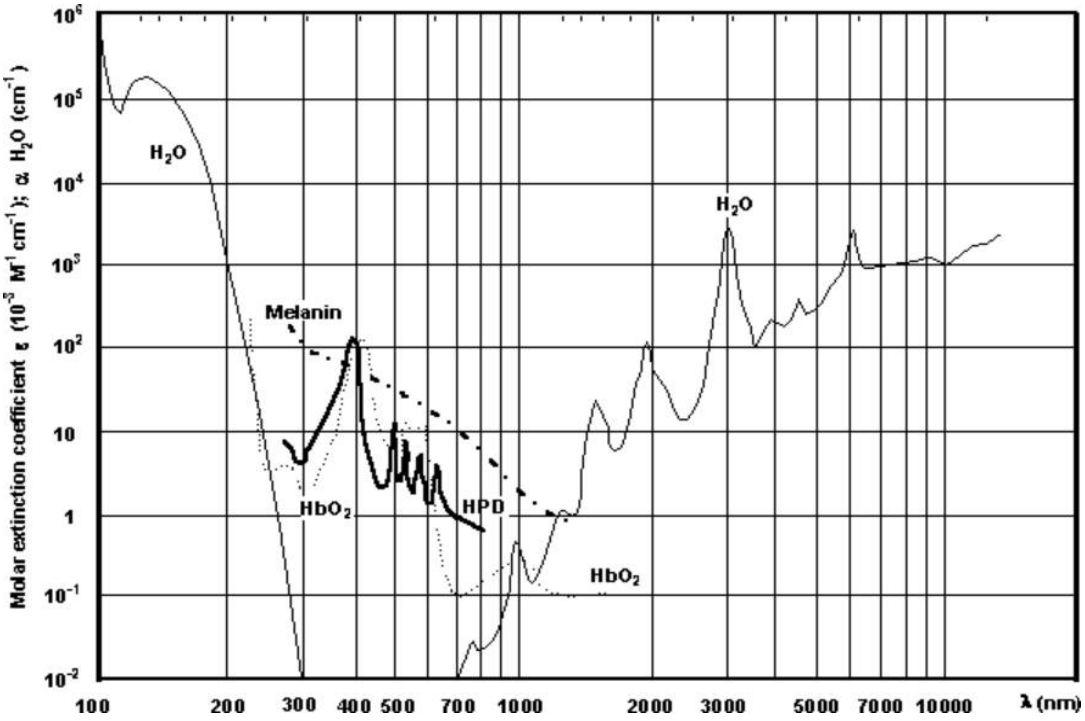

Figure 2.11. The molar extinction coefficient of common chromophores over the full 100nm to 10um spectral range. Water absorption alone limits diffuse optical measurements to a region from ~300nm to ~900nm. Melanin and hemoglobin absorption collectively narrow this region even further – from ~600nm to ~900nm in the near-IR band.

2.5.2 Dynamic absorbers

Dynamic absorbers are chromophores that vary in concentration within the time frame of DOT measurements, and are thus primarily responsible for generating the intensity changes seen with DOT. Some (Hb, HbO_2) are fundamental, while others (Hi, HbCO) provide no functional information, but can be present in unknown amounts, and thus introduce errors into blood volume and oxygenation estimates.

Hemoglobin

Hemoglobin is the strongest dynamic absorber in tissue. It is present in brain tissue at a nominal concentration of around 84μmol/liter [50]. The overall erythrocyte volume fraction within the blood, colloquially referred to as the hematocrit, remains relatively constant in healthy individuals. However the local hematocrit can vary significantly, and is reduced in small vessels due to plasma sheathing. This localized reduction in microvascular hematocrit is referred to as the Fahraeus Effect. Hemoglobin absorption sets the short wavelength limit for practical DOT measurement in humans at around 600nm,

although rodent measurements should be possible at wavelengths below 600nm if small optode separations are used.

The two most common and metabolically significant forms of hemoglobin are deoxyhemoglobin (Hb) and oxyhemoglobin (HbO_2). The absorption coefficients of Hb and HbO_2 are shown in Figure 2.12. Figure 2.13 shows the spectral absorption of whole blood vs. oxygen saturation. Other forms of hemoglobin (HbCO, Hi, SHb) are ineffective at transporting oxygen, and are usually formed as a result of a metabolic disorder. The spectral properties of these chromophores are shown in Figure 2.14.

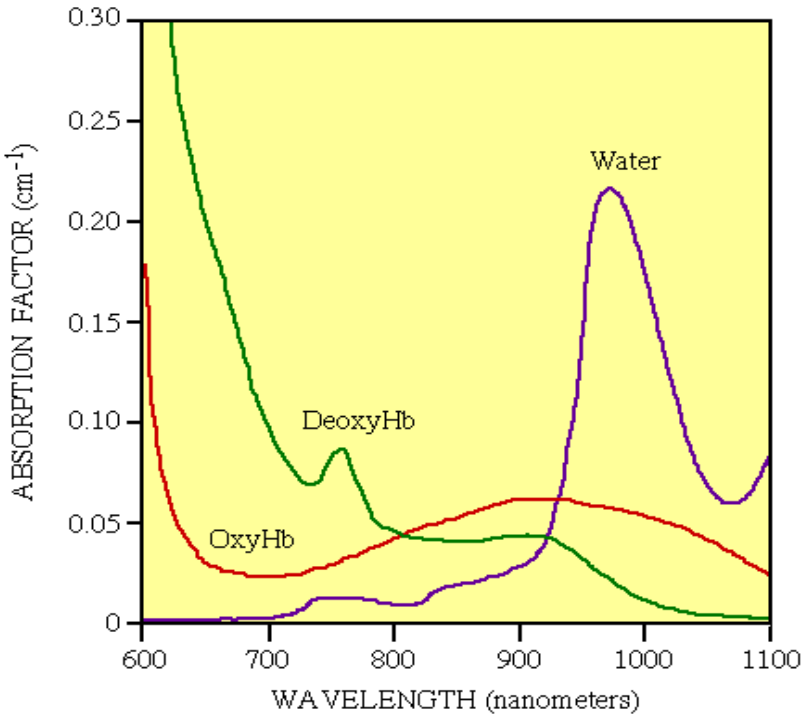

Figure 2.12. The absorption coefficient of hemoglobin, deoxyhemoglobin, and water versus wavelength over the 600nm to 1100nm near-IR spectral band. In this region, the absorption coefficient is very low, so scattering is the predominant loss mechanism.

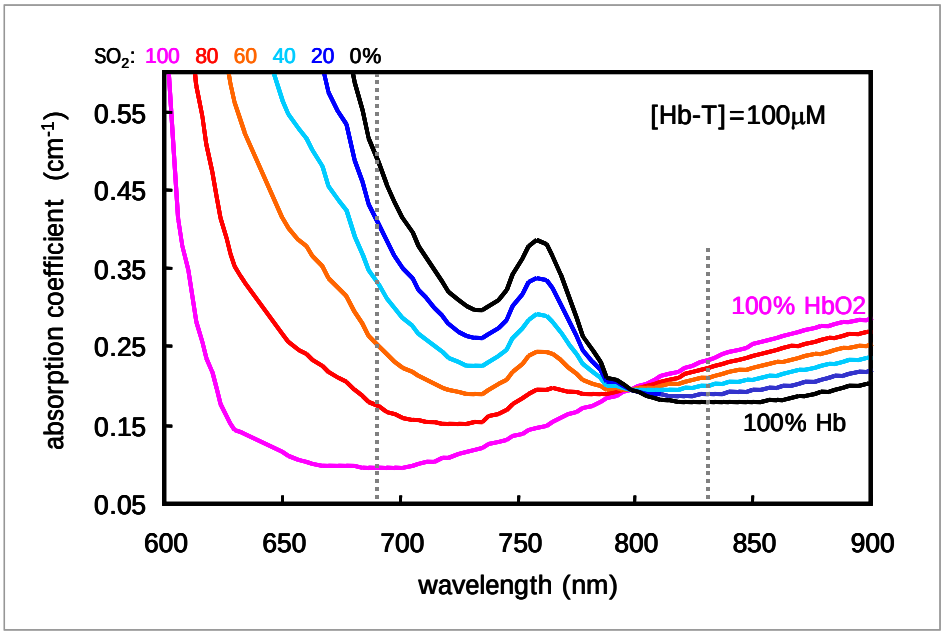

Figure 2.13. A plot showing the spectral absorption properties of blood as a function of oxygen saturation. All the curves converge at the isosbestic point at ~805nm. Optical measurements at this wavelength are sensitive to total blood volume and are unaffected by changes in oxygen saturation.

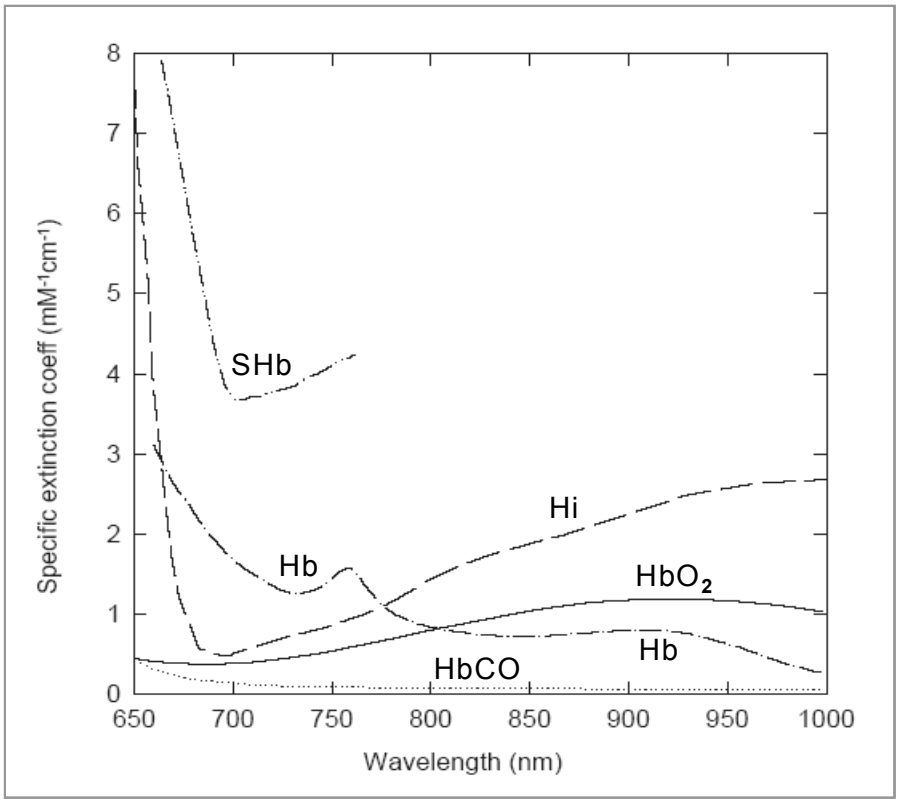

Figure 2.14. Extinction coefficients of various forms of hemoglobin vs. wavelength in the NIR spectral band. SHb = sulfhemoglobin, Hb = deoxyhemoglobin, Hi = methemoglobin, HbCO = carboxyhemoglobin, HbO_2 = oxyhemoglobin [47].

Cytochromes

The cytochrome enzymes consist of a group of iron-containing proteins involved in the phosphorylation process. Many cytochrome enzymes participate in the oxidative chain, however cytochrome c oxidase is the strongest IR absorber of the group. In its "oxidized" form, it has a ~200nm wide absorption band in the NIR, centered at 830nm, however it has negligible absorption in its reduced form. The optical effects of cytochrome are small in comparison to hemoglobin, and polyspectral measurements (i.e. at least three spectrally independent measurements) are required to monitor cytochrome concentration in-vivo [47].

NADH, flavoenzymes, and other chromophores

All of these chromophores combined provide less attenuation in the NIR than that of cytochrome oxidase, so they would be difficult to detect unambiguously. NAD absorbs strongly near 340nm in the near-UV when in its hydrogenated (NADH) form. The flavoenzymes FMN and FAD both absorb in the 400nm to 500nm region of the spectrum. Coenzyme Q absorbs strongly at 275nm in its oxidized form. Succinate dehydrogenase has some absorption in the 600-700nm range, possibly extending further into the NIR, but it is present at very low concentrations [47].

2.6 Sources of optical scattering in tissue

Biological media is highly scattering, and consists of both optically static (connective tissue, collagen, skin, etc.) and optically dynamic (neurons, while firing) scatterers. The static scatterers dominate, and are responsible for the creation of the diffuse optical probability density function within the tissue. Dynamic scattering changes in neurons during activation are thought to be capable of generating minute changes in the optical pathlength that can be detected noninvasively with RF and CW DOT instrumentation.

2.6.1 Static scatterers:

The cell membranes (i.e. both the cytosol/membrane and membrane/extracellular fluid interfaces) are the most important source of scattering in the brain. Mitochondria are also a significant scattering source, and compose up to 20% of the total solid content within the cells. The remaining cell contents (endoplasmic reticulum, Golgi, tubules, filaments) constitute the next largest scattering component. Erythrocytes only constitute about 5% of the tissue volume. Since they are large, they have a low scattering crossection, and are not a significant contributor to the overall scattering coefficient within the tissue. Properties of the major light scatterers in tissue are shown in Table 2.2 [47].

Table 2.2. The major cellular structures and fluids responsible for light scattering within tissue [47].

	Internal lipid content	Internal protein content	Refractive index	Individual size or volume	Scattering properties
Red blood cell	1 g/dl	34 g/dl	1.40	88 fl	Mie
Mitochondria	13 g/dl	39 g/dl	1.428	≈ 0.2 fl	Mie
Lysosomes					Mie
Nucleolus	Lip+pro =	37-47 g/dl	1.40-1.42	≈ 0.2 fl	Mie
Centrioles				≈ 0.03 fl	Mie+Rayleigh
Golgi apparatus			> cytosol		Mie+Rayleigh
Endoplasmic reticulum (Smooth and Rough)			Nissl bodies ≈ 1.5		Mie+Rayleigh
Microtubules & filaments			≈ 1.5		Rayleigh
Cell granules			≈ 1.5		Rayleigh
Lipoprotein membranes			≈ 1.46		See text
Myelin sheath	330-470 mg/g	130-270 mg/g	≈ 1.46	< 1 μm	Mie
Triglyceride			1.491		
Dried protein			1.53-1.54		
Cytosol		10 mg/ml	1.354		
Cerebrospinal fluid		0.28 mg/ml	1.335		
Extracellular fluid		≈ 0.25 mg/g	1.334		
Nuclear sap	Lip+pro =	8.5-12 g/dl	1.350-1.356		
Blood plasma		5.5 g/dl	1.343		

2.6.2 Dynamic scatterers

Isolated cortical brain slices show changes in reflectance when stimulated [52]. Tissue reflectance decreases as the cell volume increases, and membrane depolarization resulted in a reversible increase in cell volume in excitable tissues. There is evidence to suggest that the origin of the observed scattering change was at the intra/extra-cellular barrier [52]. This is possible, however additional effects such as mitochondrial shrinkage may also contribute [47]. Note that these measurements were made on thin (less than 100 μm) tissue slices in reflectance-mode only.

DOT image reconstruction

Reconstructing DOT imagery involves solving mathematical problems of two distinct forms. The forward problem is explained in Section 2.7. Both direct methods and numerical solutions to the forward problem are also discussed. The inverse problem is discussed in Section 2.8, and methods for regularization are also explained.

2.7 Modeling the medium: The Forward Problem

In order to reconstruct optical tomographic imagery, a numerical model is needed for the forward problem. The forward problem relates the measured fluence at each detecting optode to a known perturbation anywhere within the medium, given expressions for both the volume optical properties of the medium and the optode geometry. Generally, the forward algorithm provides a weight (or

importance) to each voxel within the sample volume, and the measured signal is a function of these weights. One can think of the weights as being proportional to the probability of photons from a source reaching a particular voxel, and then continuing on to reach the detector position [29].

Two direct approaches to solving the forward problem include analytical solutions and Monte Carlo simulations. Other standard methods for numerical approximation of partial differential equations include integral or differential techniques, such as finite element and finite difference approximations based upon perturbative approaches such as the Born and Rytov expansions [11].

2.7.1 Direct approaches

As a rule, the more approximate models such as the straight line approximation and the empirical functions are quick to calculate, however the most precise models such as the infinite difference and Monte Carlo can take days to calculate.

Analytical approaches

The analytic solution for a homogeneous system with an embedded spherical or cylindrical object can also be used to generate the weights.

Monte Carlo simulations

Monte Carlo simulation is a robust but time-consuming method of generating the weights. The advantage of these calculations is that they can accurately model irregular boundary conditions and heterogeneous media.

2.7.2 Numerical methods

Linear perturbation techniques such as the Born and Rytov expansions express the imaging problem as a perturbation to a known or estimated background medium. They relate the absorption coefficient to the measured data through a linear system of equations [53].

The Born approximation

The Born approximation expresses the total fluence rate or total field as a combination of the incident field (the field that would have been detected if no optical heterogeneity was present) and the scattered field (the field attributed only to the heterogeneous optical distribution). This scalar Helmholtz equation cannot be solved directly, but a solution can be derived as a convolution of the driving function with the Green's function solution for the homogeneous Helmholtz equation. The Born approximation can be simplified when the scattered field is weak compared to the incident field.

The Born approximation is a straightforward way to obtain a solution to the heterogeneous diffusion approximation, however it imposes a number of theoretical and experimental limitations [54].

The Rytov approximation

The Rytov expansion expresses the total field as the sum of a homogeneous and heterogeneous or scattered exponential complex phases respectively. Subtraction of the homogeneous Helmholtz equation for the homogeneous field yields the heterogeneous Helmholtz equation for the Rytov expansion. The Rytov approximation simplifies when the scattered complex phase is slowly varying.

For a total field that deviates more than 10% from the homogeneous field detected, an inverse solution based on the Rytov approximation is expected to produce a better reconstructed value than the Born approximation solution, since it divides a larger number than the Born solution (normalized or not) [54].

Born vs. Rytov for tissue imaging

Although the structure of the Born and Rytov solutions look very similar, there are some fundamental differences. The Born approximation makes the assumption that the scattered wave is small, and that it scales approximately linearly with the absorption. In biological tissue we are interested in imaging absorption values that vary from about 0.02 cm^{-1} to 0.30 cm^{-1}. In fact, this linear assumption will break down for absorption differences greater than about 0.10 cm^{-1}, which is well within our region of interest. The Rytov approximation does not place a restriction on the magnitude of the scattered wave change, but rather assumes that the scattered field is slowly varying [29].

Generally it has been shown that the Rytov approximation is equivalent to a normalized Born (and by extension to a standard Born) solution for weak scattered fields. For larger scattered fields it was found that the differences encountered between the two approximations was not a result of the physics of the approximation per se but rather on the more efficient formulation of the Rytov scattered field. Thus, for many tissue imaging situations, it is better to use the Rytov approximation (or the equivalent normalized Born approximation for small scattered fields) rather than the standard Born approximation [47].

The calculation of the Born scattered field requires the determination of a gain factor "A", and this generally has to be performed for every source detector pair independently, since individual gains may vary. Determination of A can be done experimentally on a medium with known optical properties and known geometry (i.e. a measurement where A is the only unknown), however the Rytov expression (or the normalized Born) provides a convenient way to cancel out this gain term through differential measurements – measurements where a baseline is obtained prior to a change in optical properties. This would occur where functional activation is monitored, when a contrast agent is administered, or simply when a measurement is performed in a calibration medium either before or after the tissue measurement. Therefore with the Rytov approximation there is no need to explicitly determine A if the experimental protocol is designed to allow for differential measurements.

Taking the ratio of differential measurements also eliminates many systematic errors since both the baseline and activation measurements "see" the same system: with uncertain (but stable) fiber-medium boundaries, imperfect coupling, etc. [54].

2.8 3-D Image reconstruction: The Inverse Problem

The process of reconstructing volume imagery involves calculating the locations and magnitudes of the unknown absorption changes from the raw optical measurements, and is referred to as the "inverse" problem, since, in principle, a matrix inversion is required. It seems that the most straightforward solution to the inverse problem should be to directly invert the forward equations to solve for the unknowns, however the mathematical problem is ill-posed, in that there are too few simultaneous equations to solve for all of the unknowns. Theoretically, a direct matrix inversion should yield the unknowns, but the inversion of a large matrix is a time-consuming calculation, and due to the ill-posed nature of these matrices, the solutions are not unique. Therefore regularization techniques are required to handle the uniqueness problem [29].

2.8.1 Regularization

The equations for the fluence resulting from an absorption or scattering heterogeneity can be discretized to yield the following general system of linear equations:

$$\begin{bmatrix} \delta\phi_a(\quad) \\ \delta\phi\ (\quad) \end{bmatrix} = \begin{bmatrix} A_a(\ ,\ ,\quad) \\ A_D(\ ,\ ,\quad) \end{bmatrix} \begin{bmatrix} \delta\mu_a(\quad) \\ \delta\ (\quad) \end{bmatrix} \qquad 2.19$$

Unfortunately this linear system is ill-posed and underdetermined (i.e. too few measurements available to calculate too many unknowns). Therefore, some type of regularization or stabilization technique is required.

Algebraic regularization approaches include the Algebraic Reconstruction Technique (ART) and a modification of ART referred to as the Simultaneous Iterative Reconstruction Technique (SIRT). Regularization is accomplished by limiting the number of iterations in both algebraic techniques.

Subspace regularization approaches include Truncated Singular Value Decomposition (TSVD) and the Truncated Conjugate Gradient (TCG) approach. Tikhonov regularization, the classic approach, was used to regularize the data presented and discussed in Chapter 6.

The Algebraic Reconstruction Technique (ART)

ART is an iterative process, which means that it "walks in" toward an optimal solution. It begins by projecting the result of an estimated solution as a line onto the hyperplane defined by one row of the system. Then this projection is used as an estimated solution for the next iteration. Once all rows of the forward matrix have been projected upon, the index cycles back to the first row. This can be expressed mathematically as:

$$\hat{\mathbf{x}}_{j+1} = \hat{\mathbf{x}}_j + w\frac{b_i - \mathbf{a}_i\hat{\mathbf{x}}_j}{\mathbf{a}_i\mathbf{a}_i^{\mathbf{T}}}\mathbf{a}_i^{\mathbf{T}} \qquad j = 0, 1, \ldots \quad i = (j \bmod 2m) + 1 \tag{2.20}$$

where: $\hat{\mathbf{x}}_j$ is the j th estimate of the object function,

$\mathbf{a}_i$ is the ith row of the $2m \times n$ matrix $\mathbf{A}$,

b_i is the ith measurement,

and w is a relaxation parameter that adjusts the step size of each iteration [53].

ART is only valid for linear equations [29].

The Simultaneous Iterative Reconstruction Technique (SIRT)

SIRT is similar to ART, but instead of updating the estimate for each row, the average of the update vectors for all rows is calculated. SIRT sequentially projects estimates onto a hyperplane defined by a particular row of the linear system. It provides a smoother reconstruction than ART, albeit at the cost of slower convergence [53].

$$\hat{\mathbf{x}}_{j+1} = \hat{\mathbf{x}}_j + w\frac{1}{2m}\sum_{i=1}^{2m}\frac{b_i - \mathbf{a}_i\hat{\mathbf{x}}_j}{\mathbf{a}_i\mathbf{a}_i^{\mathbf{T}}}\mathbf{a}_i^{\mathbf{T}}. \tag{2.21}$$

Truncated Singular Value Decomposition (TSVD)

The TSVD algorithm is derived from the singular value decomposition (SVD) of the $2m \times n$ stacked system matrix $\mathbf{A}$. The SVD of the system matrix is given by:

$$\tilde{\mathbf{A}} = \mathbf{U}\Sigma\mathbf{V}^{\mathbf{T}}, \qquad \mathbf{U} \in \mathbb{R}^{2m\times 2m}, \quad \Sigma \in \mathbb{R}^{2m\times n}, \quad \mathbf{V} \in \mathbb{R}^{n\times n}$$

where $\mathbf{U}$ and $\mathbf{V}$ are orthonormal matrices and $\boldsymbol{\Sigma}$ is a diagonal matrix with values $\Sigma_{i,i} = \sigma_i \geq 0$. The σ_i are known as the singular values of $\mathbf{A}$ and the decomposition is written such that:

$$\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r \qquad \sigma_{r+1}, \sigma_{r+2}, \ldots, \sigma_{\min(m,n)} = 0$$

where r is the rank of $\mathbf{A}$. The TSVD algorithm computes the reconstruction by using only the largest t non-zero singular values and singular vectors to approximately solve $\mathbf{Ax} = \mathbf{b}$. Mathematically this can be written as:

$$\hat{\mathbf{x}} = \mathbf{V}_t \Sigma_t^{-1} \mathbf{U}_t^{\mathrm{T}} \mathbf{b} \qquad 2.22$$

where $\mathbf{V}_t$ and $\mathbf{U}_t$ are the first t columns of $\mathbf{V}$ and $\mathbf{U}$ respectively, and $\mathbf{\Sigma}_t^{-1}$ is the inverse of the square diagonal submatrix of the largest t singular values. The truncation parameter t controls the amount of regularization in the inverse.

Truncated Conjugate Gradient (TCG)

The TCG algorithm is derived from the conjugate gradient algorithm in a similar manner to the way TSVD follows from the SVD. The conjugate gradient algorithm is an iterative technique to solve a symmetric positive definite linear system of equations. For ill-conditioned systems TCG regularizes by stopping, or truncating, the iterations prior to complete convergence. So for TCG, the degree of regularization is determined by the number of iterations computed.

Tikhonov regularization

Tikhonov regularization involves minimizing the equation:

$$||\vec{u} - \tilde{A}\vec{x}|| + \beta||\vec{x}|| \qquad 2.23$$

where β is a free parameter which controls the level of regularization. Each time we try a different value of β. we create a new matrix to invert. Since inverting a matrix is computationally intensive, choosing the best value for β can be a time consuming process [29].

Optimizing the regularization: The L-Curve

With any regularization technique, one of the primary issues is the selection of the parameter that controls the trade-off between fidelity to the data and some constraint on the quality of the result. There are two general classes of methods: *a priori* methods, which use prior knowledge about the solution, the noise, or both, and *a posteriori* methods, which use only the measurements and the forward model.

For the algebraic techniques, *a posteriori* methods like the L-curve do not work well because standard measures of error, such as the residual error or solution norm, do not change monotonically as we iterate, so we simply present the best possible result for these methods.

For the subspace techniques a well-known *a posteriori* method is the L-curve technique (Hansen 1998), which for subspace methods graphs the log of the 2-norm of the residual versus the log of the 2-norm of the estimate while varying the regularization parameter. Thus the regularization parameter itself is represented only parametrically. An example of an L-curve generated from a TCG reconstruction at a signal-to-noise ratio of 20 dB is shown in Figure 2.15. This graph was generated by plotting the residual norm and reconstruction norm over 300 iterations of the algorithm. The 'corner' of the resulting curve is taken as a good choice of regularization parameter because it identifies a point at which there is a balance between an increase in the residual norm and an increase in the solution norm. The diamond shows the point manually selected as the L-curve corner, which corresponded to 12 iterations [53].

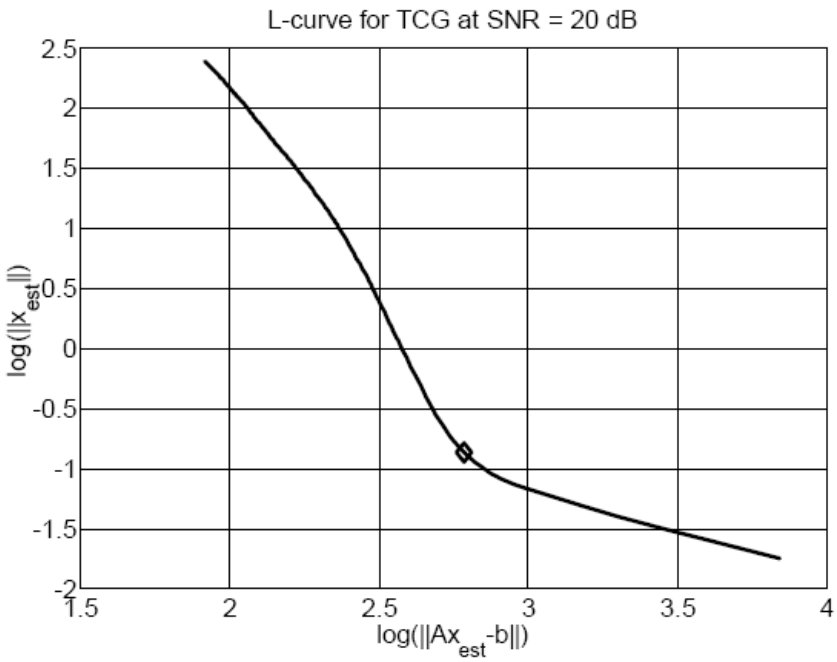

Figure 2.15. The L-curve for a truncated conjugate gradient reconstruction at a signal-to-noise ratio of 20 dB. The diamond at the corner of the graph identifies the point that was selected as the corner of the L-curve [53].