MULTI-MODAL SPECTROSCOPY OF BREAST TISSUE

BY
ZOYA I. VOLYNSKAYA

S.B., Electrical Engineering and Computer Science
University of North Carolina at Charlotte (2003)

Submitted to the Department of Electrical Engineering and Computer Science
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Electrical Engineering and Computer Science
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
May, 2005
[June 2005]

Author_____
Department of Electrical Engineering and Computer Science
May 21, 2005

Certified by_____
Michael S. Feld
Professor of Physics
Director, G.R. Harrison Spectroscopy Laboratory
Thesis Supervisor

Accepted by_____
Arthur C. Smith
Chairman, Department Committee on Graduate Theses

Multi-Modal Spectroscopy of Breast Tissue

by

Zoya I. Volynskaya

Submitted to the

Department of Electrical Engineering and Computer Science

May 21, 2005

In Partial Fulfillment of the Requirements for the Degree of

Master of Science in Electrical Engineering and Computer Science

Abstract

Breast cancer is the most common form of cancer afflicting women in the United States; one out of eight women will be diagnosed with breast cancer during her lifetime. Currently, screening is performed by a combination of annual clinical breast examinations and x-ray mammography. However, only 10 to 25 percent of suspicious lesions detected during mammography are diagnosed as malignant upon biopsy, which implies that a large number of biopsies can be avoided. Although mammography images anatomic changes, it is not sensitive to the underlying morphological and biochemical changes that distinguish benign and malignant breast lesions. Presently employed diagnostic procedures are invasive, time consuming, and expensive. Thus, there is a clinical need to develop new tools for the early diagnosis of malignancy in the breast.

In recent years our laboratory has explored the use of Raman spectroscopy for diagnosing disease; one important area is the detection of breast cancer. Raman spectroscopy provides information about the morphological and biochemical make up of tissue and, with the aid of our diagnostic algorithm, has provided good results in distinguishing between malignant and benign breast lesions, with a sensitivity, specificity, and an overall accuracy of 90, 96, and 86 percent, respectively [Haka,2004]. Although these initial results are promising, we would like to improve the overall accuracy. Another promising spectroscopic technique developed in our laboratory is tri-modal spectroscopy (TMS), the combination of diffuse reflectance (DRS), intrinsic fluorescence (IFS), and light scattering spectroscopy (LSS). This technique has been successfully applied to the diagnosis of epithelial neoplastic tissue, leading to the interest in exploring its application to the diagnosis of lesions in breast tissue. Finally, the Raman and DRS/IFS modalities provide complementary information and the combination of this information into a single diagnostic algorithm may provide superior diagnostic capabilities.

The central theme of this research is to investigate DRS/IFS as a useful technique for the diagnosis of breast cancer and to evaluate the effectiveness of its combination with Raman spectroscopy. Through this research, we hope to aid the medical community in early diagnosis, treatment, and prevention of breast cancer.

Thesis Supervisor: Michael S. Feld

Title: Professor of Physics

Acknowledgements

The work presented in this thesis was performed at the George R. Harrison Spectroscopy Laboratory at the Massachusetts Institute of Technology. I would like to thank my thesis advisor, Professor Michael S. Feld, and Dr. Ramachandra R. Dasari for their guidance and support. I would also like to thank my colleagues: Abi Haka for her collaboration and ability to explain and guide me in the beginning of my graduate life at MIT; Joe Gardecki and Kate Bechtel for their contributions in making me a better scientist and prove reading thesis; Jon Nazemi for showing me the innerworkings of the FastEEM analysis code; Jelena Mirkovic, Sasha McGee and James Tunnell for their participation in the earliest part of the project; Obrad Scepanovic and Wei Shih for making the lab so much fun; and the entire Spectroscopy Laboratory. I would like to thank Madeleine Suprun for creating so many memorable and special moments in my life and for brightening up my life in the USA and everywhere else in the world. Daniel Blinov and Paolo Batoni, I really appreciate your ability to tolerate me daily (thanks for all the 5 am phone calls!). Finally, I would like to thank all my family and friends for their love and support.

"Can miles truly separate us from friends? If we want to be with someone we love, aren't we already there?"
-- *Richard Bach*

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

1.1 Introduction and Motivation

Breast cancer is one of the most common forms of cancer to afflict women in the United States. One out of every eight women will be diagnosed with breast cancer during her lifetime. The mortality rate due to breast cancer is second only to lung cancer. Nearly forty thousand women died in 2003 from this disease [Jemal,2004] and the survival rate critically depends on early diagnosis.

Mammography is currently the leading diagnostic technique used for the detection of non-palpable, highly curable breast cancer [Dahnert,1999]. It relies on finding density changes in the breast tissue. However, because these changes do not uniquely correlate with breast cancer, abnormalities identified with mammography must be further investigated with biopsy. Thus, mammography is used only as a screening technique, not as a diagnostic one. Only 10 to 25 percent of suspicious lesions found during mammography are diagnosed as malignant upon biopsy [NCI,2004]. Currently employed diagnostic procedures are invasive, time consuming and expensive, and this has motivated scientists to investigate new tools for the early diagnosis of malignancy in breast tissue.

1.2 Minimally-Invasive Techniques and Previous Work

Optical techniques offer a means to gain morphological and biochemical information of tissue in a non-destructive manner, since optical radiation is non-ionizing. Furthermore, light delivery and collection is performed with minimal invasion via optical fibers. Optical instrumentation is generally inexpensive, portable, and compact, making it adaptable to a clinical setting. Presently, a variety of optical imaging and spectroscopic techniques are being explored with the aim of improving breast cancer diagnosis, with an emphasis on improving the ability to discern benign from malignant lesions. The techniques, employing visible or near-infrared light, have the potential to provide chemical information and assist pathologists in diagnosing cancer.

Diffuse optical tomography (DOT), an optical analogue of X-ray, studies the propagation of amplitude-modulated pulses of light through the breast, is non-invasive and can detect lesions deep (10 cm) within the tissue. DOT employs diffuse light that propagate through tissue; it provides three-dimensional tomographic images of the internal organs. Some diagnostic ability has been demonstrated based on the detection of oxy/deoxy hemoglobin with a contrast agent [Ntziachristos,2001]. However, DOT can only detect a limited number of chemicals and its low resolution results in small lesions going undetected.

Unlike DOT, optical spectroscopic techniques sample the tissue locally with a volume less than 1 mm^3. Light delivery and collection are accomplished by means of optical fibers that can be incorporated into a biopsy needle. Depth information can be extracted as the needle is inserted or withdrawn. In contrast to a standard biopsy, a spectroscopic transdermal needle measurement has the advantage of providing immediate diagnosis. As a result, spectroscopy has the potential to reduce both the likelihood of a non-diagnostic biopsy, which requires a repeat of needle or surgical biopsies, and patient anxiety by eliminating the currently unavoidable wait for a traditional histopathology diagnosis.

Previously to our own work, several groups have investigated Raman spectroscopy for diagnosing cancer. The first Raman spectroscopy experiment of breast tissue was performed in 1991 [Alfano,1991]. Redd et al. soon followed and examined the application of Raman spectroscopy using visible excitation [Redd,1993]. In our laboratory, Raman spectroscopy has been investigated as a diagnostic modality for breast cancer for nearly 15 years. Our approach has been centered on accurately modeling spectra with a set of morphological basis spectra and achieving diagnostic algorithms that have physical interpretation valuable to pathologists. Results from an *ex vivo* study conducted at the University Hospital in Cleveland confirmed that Raman spectroscopy is able to distinguishing benign and malignant lesions, with a sensitivity and specificity of 90 and 96 percent, respectively, and an overall accuracy of 86 percent [Haka,2002].

Raman spectroscopy has been successful in distinguishing cancerous from non-cancerous lesions. However, the model and algorithm misdiagnose some benign tumors, such as fibroadenoma. Our laboratory has also been successfully exploring the use of trimodal spectroscopy (TMS: diffuse reflectance, light scattering and intrinsic fluorescence spectroscopies) for the diagnosis of epithelial lesions, with excellent success. In a parallel study, I have used fluorescence and reflectance spectroscopies to detect its feasibility and efficacy for the diagnosis of breast cancer.

A few researchers have used diffuse reflectance spectroscopy [Bigio,2000, Yang,1997] fluorescence spectroscopy [Alfano,1987, Nair,2002, Shama,1996] to study breast tissue. They have achieved some success. Gupta and Majumder et al. [Gupta, 1997, Majumder, 1998] analyzed different data sets collected from the same set of breast tissues *ex vivo* and showed that the emission spectra at excitation wavelengths of 340 and 488 nm and excitation spectra at emission wavelengths 390 and 460 nm exhibit significant differences between normal, benign and malignant tissues. Palmer et al. [Palmer,2003] also examined fluorescence and diffuse reflectance spectroscopies. The results from this study are promising: multiexcitation fluorescence spectroscopy is successful in discriminating malignant and nonmalignant tissues, with a sensitivity and specificity of 70 and 92 percent, respectively. The sensitivity (30 percent) and specificity (78 percent) of diffuse reflectance spectroscopy alone is significantly lower. Combining fluorescence and diffuse reflectance spectra did not improve the classification accuracy of an algorithm based on fluorescence spectra alone. The difficulty in the study is the inability to determine the biological basis of the differences observed in the fluorescence spectra of malignant and nonmalignant tissues. Several groups have investigated the Raman spectroscopy [Alfano,1991, Frank,1995, Frank,1994] for diagnosing cancer, but to our knowledge, no researchers have employed the combination of fluorescence, reflectance, and Raman spectroscopies. We expect that the combination of these techniques, which we call multi-modal spectroscopy, will provide a better overall accuracy, defined as the correct prediction of each of the different pathologies.

This thesis will explore the application of DRS/IFS to the diagnosis of breast lesions and will evaluate the combination of DRS/IFS and Raman spectroscopic modalities in terms

of diagnostic success with ex *vivo* data. DRS/IFS spectra of the tissue sample were obtained, immediately followed by the collection of a Raman spectrum. Care was taken in placing the Raman probe at the same site on the tissue as the FastEEM probe. The tissue measurements were taken within half an hour of excision.

1.3 Objectives

The major goal of this thesis is to explore the usage of intrinsic fluorescence and diffuse reflectance spectroscopies for diagnosing breast cancer. The research is performed in several stages, which are examined in following chapters:

The goal of DRS/IFS data analysis is to develop an overall algorithm that will distinguish among five pathologies. The algorithm is developed using leave-one-out cross validation and logistic regression and includes parameters from DRS (diffuse reflectance spectroscopy) and IFS (intrinsic fluorescence spectroscopy).

Another direction of this research is the investigation of different fluorescence excitation wavelengths. DRS/IFS data were collected via the FastEEM instrument developed in the Spectroscopy Laboratory [Motz,2002, Tunnell,2003]. The FastEEM delivers a sequence of ten laser pulses (308- 480 nm) and two white light pulses to the tissue via an optical fiber probe. Different excitation wavelengths excite different fluorophores and probe different tissue depths. We utilize IFS data from three different wavelengths to examine the differences in penetration depth in the breast tissue. The information from the concentration of fluorophores, such as NADH, collagen and others are expected to provide diagnostic information about different pathologies that can be related back to breast anatomy and histopathology. It is anticipated that the application of different wavelengths provides us with additional diagnostic information.

Using an already developed chemical/morphological model of breast cancer with Raman spectroscopy [Haka,2004], Raman spectra acquired

from intact, *ex vivo* samples is examined in a prospective manner. The chemical/morphological model is based on a linear combination of nine basis spectra using an ordinary least-square minimization algorithm with a non-negative constraint, as is explained in more detail in Chapter 5. Additional goals include expanding our diagnostic algorithm to include pathologies not observed in our *ex vivo* laboratory data, such as ductal carcinoma in situ (DCIS).

The main goal of this study is to investigate and develop a multimodal spectroscopy algorithm that combines Raman and DRS/IFS spectroscopic information. In this phase of the research, the sensitivity and specificity of each method are compared, with a focus on the similarities and differences between the outliers obtained with either the Raman or DRS/IFS methods. The results will shed light on the advantages of combining the two techniques, and perhaps on the need for developing a combined Raman/ DRS/IFS probe.

Chapter 2. Breast Cancer

In order to develop a diagnostic algorithm for the early detection of breast cancer, it is important to thoroughly understand the anatomy of breast tissue. Some of the work in this chapter can be found in more detail in [Haka,2005, Rubin,1996].

2.1 Introduction and Epidemiology

Cancer is described as any malignant growth or tumor caused by uncontrolled division of abnormal cells; it also may spread to different parts of body through the blood stream or lymphatic system. In 2003 in United States, there were 211,300 females diagnosed with breast cancer, which accounts for nearly one in every three cancers diagnosed in women. In the same year, there were an estimated 39,800 deaths attributed to breast cancer, including about 400 males [Jemal,2004]. There is a difference in the number of incidences of breast cancer for major ethnic groups (breast cancer is slightly more common in Caucasians). This phenomenon is associated with differences in breast anatomy for different ethnic groups, socioeconomic status and lifestyle, which brings about a difference in treatment and survival rate. The growth of breast cancer tumors is often affected by the presence of estrogen and progesterone. The following risk factors are a result from exposure to these hormones: over the age of 50, first pregnancy after age 30, not having children at all, menstruation before age 12, menopause after age 50. Also a family history of the disease, a personal history of disease, a history of breast biopsy or chest radiation, moderate alcohol use (2 to 5 drinks daily), obesity, and race are considered risk factors.

Over the last several years the mortality rate has decreased slightly likely owing to early diagnoses and improvements in treatment. The earliest diagnosis is accomplished via abnormalities detected by mammography before anything can be felt by a woman or her doctor during a breast examination.

The lack of glandular tissue in the male breast is the primary reason for the rarity of male breast cancer. Because male breast cancer contributes only about one percent to the total numbers of breast cancer, breast cancers in women will be the focus of this research

2.2 Breast Anatomy

The mature breast is made up of secretory glandular tissue and surrounding adipose tissue. Breast tissue extends from below the collarbone to the level of the sixth or seventh rib, and from the breastbone to the axilla. The nipple and areola are covered by stratified squamous epithelium and they are located in the center of the breast. Fibrous connective tissue (fascia) lies between the breast tissue and the skin, and separates breast tissue from the chest muscles. A schematic of the female breast is shown in Figure 2.1.[Rubin,1996] Axilary and intercostals arteries provide blood to the breasts; axillary and intercostals veins carry blood away from the breasts. Breast cancer cells can travel to the lungs through surface veins and metastatic tumors; breast cancer cells can spread to the bone via intercostals veins to the vertebral veins network in and around the spine.

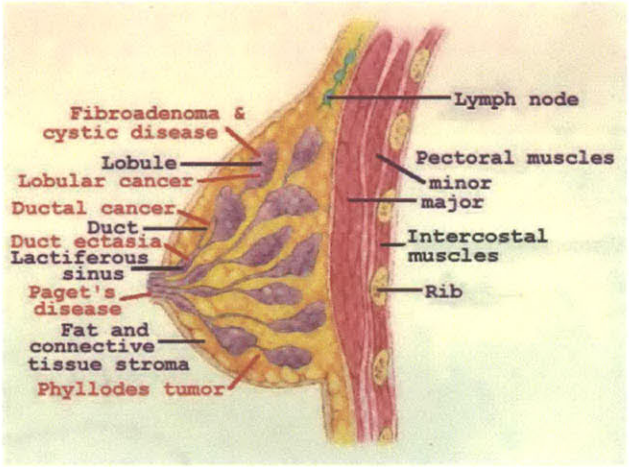

Figure 2.1 Anatomy of the breast

The glandular tissue is composed of 15 to 20 lobes, or milk glands, connected to a conduit of ducts that carry milk to the nipples. Each lobe is autonomous and connects to its own excretory duct, which connects to the nipple. The lobes themselves are divided into smaller units called lobules, which in turn are made up of acini. The lobules are the functional, milk producing, units of the mammary parenchyma. The lobular units and the ducts that connect them are separated from the stroma by a basement membrane, which

composed of collagen. Two layers of fibrous ligaments (Cooper's ligaments) support the mammary glands. These arise from stromal elements in the gland and insert into the skin and pectoral fascia, holding the breast against the chest. The breasts lose their firmness and take on a sagging appearance when these ligaments stretch with the age and multiple pregnancies.

2.3 Normal Breast Tissue

The breast consists of three types of tissue: adipose (fat), fibrous, and glandular. The fatty tissue and connective tissue make up the stroma – the tissue lying between the glands. The fatty tissue encloses cells (adipocytes). The fibrous tissue contains collagen and small amounts of glycosaminogycans. The majority of breast stroma consists of dense fibroconnective tissue admixed with adipose tissue (interlobular stroma) containing elastic fibers supporting large ducts.

2.4 Breast Pathology

Most breast diseases present as palpable masses, inflammatory lesions, nipple secretion or mammographic abnormalities. Of the women with detected abnormalities, approximately 30 percent are found to have no problems, 40 percent are diagnosed with fibrocystic change, 13 percent have miscellaneous benign tumors, 7 percent have fibroadenoma, and only 10 percent of women are diagnosed with cancer. Here we will focus only on FCC (fibrocystic change) and on fibroadenoma.

Fibrocystic Change

Multiple cut sections of breast tissue in Figure 2.2 display classic signs of fibrocystic change, with blue-domed cysts and dense fibrosis (hence, the name fibro-cystic). Fibrocystic change is a benign condition that can be described as fibrosis – formation of fibrous tissue; adenosis – increase in number of ductules; or cyst formation – dilation of ducts and lobules with semi-transparent fluid. These changes can occur alone or in any combination. The cyst size often varies with the phase of menstrual cycle and can

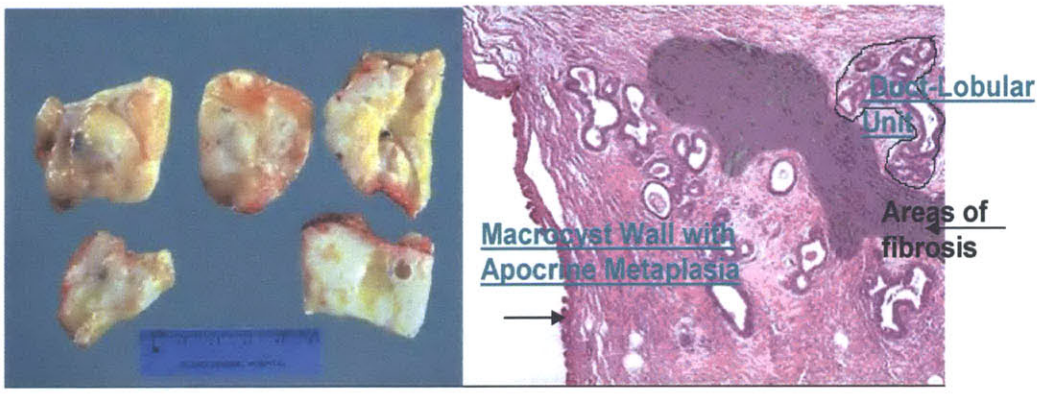

Figure 2.2 Gross and histopathology slide with fibrocystic change

become quite tender when under tense pressure. Fibrocystic changes do not increase the risk of developing cancer. However, if FCC produces palpable lumps, mammographic densities or calcifications, or nipple discharge mimicking carcinoma, then it may come to clinical attention.

Fibroadenoma

Fibroadenoma is the most common benign tumor of the breast. There are several other types of benign tumors such as ductal epithelial hyperplasia (DEH), intraductal papiloma, etc., though they occur at rates far lower than fibroadenoma. Many cases of fibroadenoma occur in younger women. It is described as a new growth composed of both fibrous and glandular tissues. Morphologically, fibroadenoma grows as a spherical

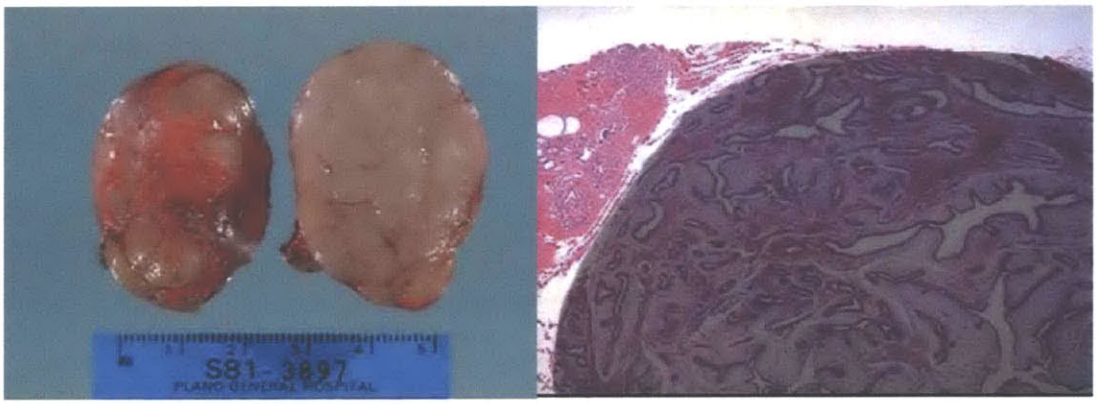

Figure 2.3 Gross and histopathology slide with fibroadenoma

nodule that is usually circumscribed and freely movable from the surrounding tissue. Figure 2.3 displays gross and histopathology slide of fibroadenoma. Lesions diagnosed as fibroadenoma show an accumulation of collagen due to fibroblast proliferation that results in expansion of the stroma. Even though fibroadenoma is a benign lesion, the fact that it often grows rapidly resulting in significant morbidity causes it to be typically excised. These tumors frequently occur in the upper outer quadrant of the breast. They vary in size from less than 1 cm to giant forms of 10 to 15 cm in diameter. Most of them are removed when they are 2 to 4 cm in diameter.

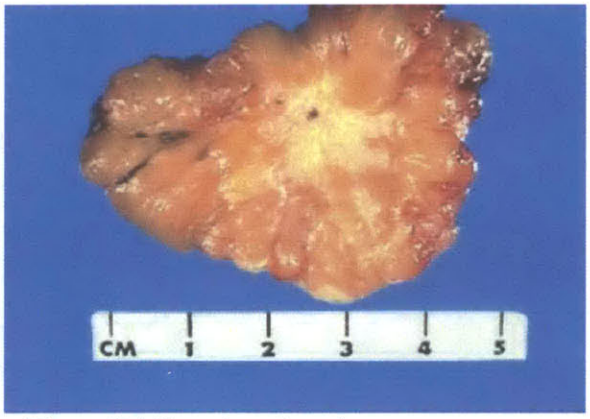

Figure 2.4 Gross breast carcinoma

Cancer

Most breast cancer develops in glandular tissue, epithelium. Cancer is characterized by abnormal cell proliferation as well as abnormal maturation of cells. Malignant cells tend to reproduce even after they have doubled 50 or 60 times, so with time they form a monomorphic tumor that is built up of billions of copies of the original cancerous cell. Breast carcinoma is shown on Figure 2.4. A flowchart of the breast cancer "continuum" from normal breast tissue to invasive cancer is shown in Figure2.5 [Haka,2005].

The earliest form of breast cancer, ductal carcinoma *in situ* (DCIS), develops solely in the milk ducts. Malignant cells are still located within the ductules and have not broken

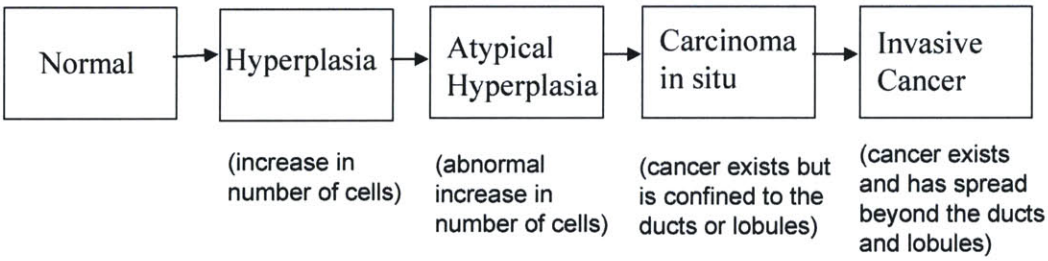

Figure 2.5 Flowchart displaying the breast "continuum"

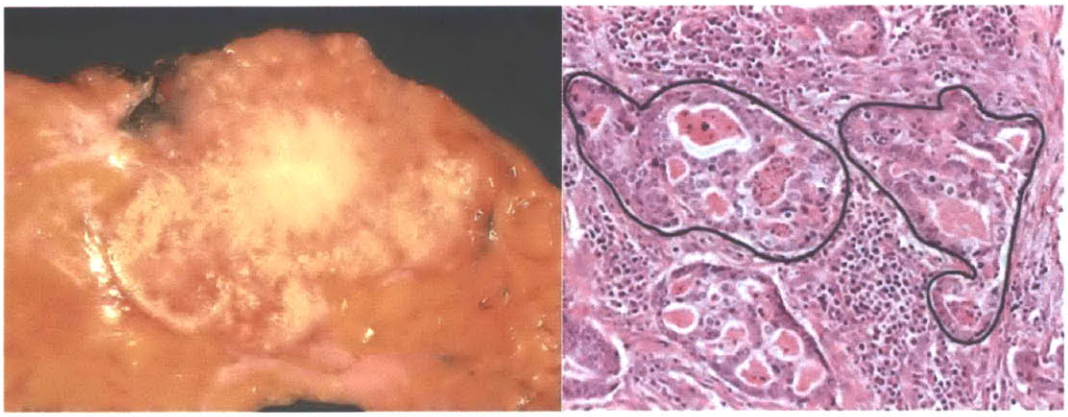

Figure 2.6 Invasive ductal carcinoma. Gross and histopathology slide

through the basement membrane and into the stroma.

Currently women diagnosed with DCIS are treated with mastectomy followed by radiation. It is now believed that many cases of low-grade DCIS and most of the cases of high-grade DCIS progress to invasive carcinoma [Rubin,1996]. This fact once again emphasizes the importance of proper early diagnosis and appropriate treatment.

The most common type of breast cancer, which accounts for 70-80 percent of cases, is invasive ductal carcinoma (IDC). IDC develops from DCIS, spreads through the duct walls, and invades the breast tissue. It can be seen in Figure 2.6.

The central white area on the picture of gross IDC is very hard and gritty because the neoplasm is producing a desmoplastic reaction with lots of collagen. This is often called a "scirrhous" appearance. There is also focal dystrophic calcification leading to the gritty areas. Lesions diagnosed as infiltrating carcinoma exhibit significant hyperplasia and often exhibit increases in collagen content due to fibroblast proliferation in response to stromal invasion by the malignant epithelial cells. Lobular carcinoma *in situ* (LCIS) consists of hyperplasia in the terminal breast ducts and acini. The cells are small and round. Invasive lobular carcinoma (ILC) accounts for 5-15 percent of invasive breast cancers. ILCs often have a diffusely invasive pattern causing tumors to be difficult to detect by either physical examination or mammography. There are several other less common types of breast cancer. The swelling of the breast, inflammatory breast cancer, is an uncommon type of breast cancer. Its symptoms include an increase in the skin

temperature, redness, and swelling of the breast. The skin may show signs of ridges and welts or it may also have a pitted appearance.

2.5 Breast Cancer Screening, Diagnosis, and Treatment

Women between the ages of 20 and 39 should have a breast examination by a physician every 3 years. After 40 years old, women should have a clinical breast exam and mammogram annually. Mammography and ultrasound are techniques that allow many cancers to be discovered at an early stage. They are considered as imaging techniques, which rely on density changes in the breast to detect cancers.

2.5.1 Screening

Mammography

Mammography is a non-invasive imaging procedure that uses x-rays to create images of breast tissue [Dahnert,1999, Haka,2005]. During a mammogram, the breast is compressed between two photographic plates that are used to record the attenuation of x-rays as they traverse the tissue. Different tissues in the breast absorb different amounts of x-rays, producing varying shades of black, gray, and white on the film. Fatty tissue absorbs a relatively small amount of x-rays and thus appears black or dark gray. Normal fibrous and glandular tissues contain water that causes absorption of a moderate amount of x-rays and thus appear light gray. Dense breast tissue, with more collagen, can make mammogram evaluation difficult because the tissue can obscure small cancers. Many mammograms show nontransparent white specks. These are calcium deposits known as calcifications and they are a key component that radiologists look for in a mammogram. Radiologists are looking for asymmetric densities - area in one breast that has a distinctly different appearance than the same area in the other breast. The shape of a mass detected

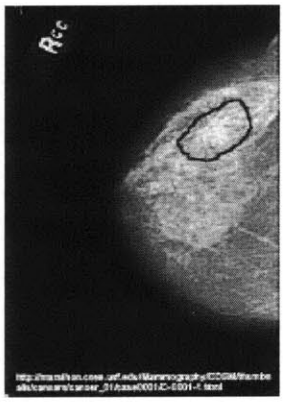

Figure 2.7 Mammogram demonstrating speculated mass

by mammography is important for differentiating benign and malignant lesions. A mammogram is a screening test, not a diagnostic test, which is why mammography cannot be an indicator of the cancer on its own [HST,2004]. The abnormalities found on a mammogram must be examined with further tests. Unfortunately, in most cases further testing shows masses detected via mammography to be benign, and in 10 percent of cases, radiologists misread normal tissue as abnormal.

Ultrasound

Ultrasound, or ultrasonography, is a medical imaging technique that uses high frequency sound waves and their echoes. The technique is similar to the echolocation used by bats, whales and dolphins. It is performed with a transducer. During the ultrasound procedure a radiologist spreads a gel on the skin of the breast and uses the transducer to direct sound waves through and into the breast. Thus, this technique provides freedom in obtaining images of the breast from any orientation. It can be used in determining if a mass is a cyst or a solid lesion that may possibly be cancerous. Even though ultrasound has great contrast, it lacks the spatial resolution of mammography. For this reason it is unable to image small lesions or microcalcifications, which may be a sign of malignancy. Therefore, ultrasound is not approved by the FDA as a screening tool for breast cancer and is only used to aid mammography in detecting abnormalities in the breasts.[Gordon,2002, Haka,2005]

2.5.2 Biopsy

Excisional Biopsy

Biopsy is the procedure of removing tissue from a potentially abnormal site previously found by mammography, ultrasound, or by a physician for further investigation. Traditionally, open excisional biopsies are performed under general anesthesia in the operating room. A surgeon removes the entire lesion, which can vary in size from a few centimeters to the entire breast (mastectomy). After removal the tissue is sectioned, specially stained, and examined by a pathologist. Currently, a diagnosis by pathology is considered the gold standard for diagnostic algorithm development.

Needle Biopsy

In recent years, less invasive techniques such as needle biopsy have been developed. Needle biopsies are less intrusive than excisional biopsies yet they provide the same diagnostic accuracy. The lump is localized with the aid of mammography and ultrasound. During this procedure the woman lies facedown on a special table, as shown in Figure 2.8. Needle biopsies are performed in a radiologist's office on an outpatient basis under local anesthesia. This whole procedure can be done in just few minutes.

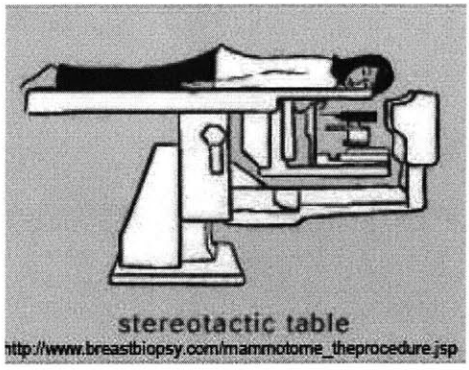

Figure 2.8 Instrumentation used in stereotactic biopsy procedures

Fine Needle Aspiration (FNA)

Different needle biopsies use different sizes of needles. In cases where the doctor wants to extract fluid from a cyst or cells from a solid lesion, they utilize very small needles of 20-25 gauge (0.89-0.51 mm). The procedure is performed under local anesthesia and requires a small incision. The procedure takes a few minutes and typically leaves no scarring and does not require stitches. FNA is the easiest method to provide pathologists with a breast biopsy. Due to the small size of the needle used for FNA, the scarring process will be quick, but as a result the evaluation of the suspicious lesion can be incomplete. A study of fine needle aspiration found that 34 percent of patients were insufficiently sampled and required more biopsies in the future.

Core Needle Biopsy

Core needle biopsies are similar to fine needle aspiration with the exception of the larger size of the needle (14-18 gauge, 2.11-1.25 mm). Increasing the size of the needle enables the doctor to obtain a larger sample that can lead to a better diagnostic accuracy and will require a smaller number of biopsies in the future. During the core biopsy breast tissue is removed while under FNA only cells and fluids are removed.

2.5.3 Staging and Prognosis

Breast cancer has been divided into smaller groups in order to standardize comparisons of results of various therapeutic modalities among clinics and guide treatment. The American Joint Committee on Cancer Staging divides the clinical stages as follows [Rubin,1996]:

> ***Stage 0***. Cancerous lesions diagnosed as DCIS or LCIS have a 5-year survival rate of 92 percent.
>
> ***Stage I***. This stage includes invasive carcinoma with a size of 2 cm or less without nodal involvement and without any distant metastases; the 5-year survival prediction is 87 percent.
>
> ***Stage II***. As the cancer progress in size (invasive carcinoma 5 cm or less) and if there are movable axillary nodes involved, but still there are no metastases, the 5-year survavial rate drops to 75 percent.
>
> ***Stage III***. Different features of breast cancer development correspond to this stage, such as a lesion that is greater than 5 cm in size; breast cancer with fixed azillary nodes or that involves ipsilateral mammary lymph nodes; any breast cancer with skin involvement, pectoral and chest wall fixation, edema, or clinical inflammatory carcinoma, if distant metastases are absent. It relates to a 5-year survival rate of 46 percent.
>
> ***Stage IV***. If the breast cancer already has distant metastases, the 5-year survival rate is about 13 percent.

The older the patient with a single breast lesion, the more likely it is to be cancer. Tumor size is another important prognostic factor also used in staging a lesion, as a larger tumor increases the risk of axillary and systemic metastasis. However, although rare, very small tumors are capable of distant metastasis. The course of treatment depends on both the stage of the disease and the patient's preference. Treatment options include surgery, radiation therapy, chemotherapy, hormone therapy, and immunotherapy. If no axillary metastases are present, some patients may not receive systemic therapy, depending on the characteristics of the tumor. Almost all women with one to three positive axillary nodes receive some form of standard systemic treatment, either hormonal therapy or chemotherapy. Women with four to nine positive axillary nodes are eligible for clinical trials using high-dose chemotherapy. If ten or more nodes are positive, women are

eligible for other experimental treatments, such as autologous bone marrow transplantation.

2.5.4 Surgery

The combination of surgery with radiation and/or chemotherapy is the most common treatment for breast cancer. Lumpectomy and mastectomy are both commonly performed surgical procedures. A partial mastectomy, a lumpectomy, is the surgery of removing a tumor from a breast. Depending on the case, lymph nodes may also be removed during this procedure. To make sure that all of the cancerous cells are excised from the patient, excised tissue is inked with different colors on all faces. If cancer cells are found touching or near the inked regions, it is assumed that malignant cells are likely still present in the patient. The color of ink identifies the region of the surgical cavity which still had malignant cells. Simultaneously with inking excised tissue, the surgeon removes about 6 small samples, called margins, from different spots of the lesion cavity. Margins and the excised tissue are examined by a pathologist to decide if any cells can be presumed present in normal tissue. In cases where those cells are found the doctor recommends the patient undergo a re-excisional biopsy. Generally, lumpectomy procedures take 60 to 90 minutes. Even in the case of clean margins (no cancerous cells have been found in margins) women receive 6 to 7 weeks of radiation therapy in order to eliminate the smallest possibility of remaining cancer cells.

There are three types of mastectomies: simple, modified radical and radical mastectomies. A simple mastectomy includes removal of the breast tissue, skin and nipple but rarely lymph nodes. A most common form of mastectomy is a modified radical mastectomy, under which the entire breast is removed. Also, some amounts of lymph nodes are removed from the underarm region, which is the reason for more difficulties with the range of motion in the arm and shoulder on the operative side after the surgery. Mastectomies require general anesthesia and must be done in the hospital. Patients have a choice of getting breast reconstruction immediately following the mastectomy as a part of the same surgery, or receiving it afterwards.

2.5.5 Therapy

Radiation Therapy

Radiation therapy uses high-energy x-rays to destroy cancer cells. It is performed externally or by radioactive "seeds" that are placed directly into the tumor, called brachytherapy. The patient may be asked to go through radiation therapy in order to shrink the tumor before surgery or she may be asked to undergo radiation after surgery to destroy all the cancerous cells remaining in the breast, underarm, or chest wall, known as adjuvant therapy. Radiation therapy is performed in a hospital or outpatient center. Radiation treatment lasts for 6 to 7 weeks, and each treatment session takes a few minutes.

Chemotherapy

Chemotherapy is a systemic treatment that travels throughout the body via the bloodstream. It often relies on a combination of drugs to slow tumor growth and destroy cancer cells. Drugs may be given to a patient orally or injected into the venous system. Chemotherapy can be used before and after the surgery similar to radiation therapy for the same reasons.

Hormone Therapy

The most effective and least toxic therapy currently available is endocrine manipulations. Estrogen, a hormone produced by women's ovaries, promotes the growth of some breast cancers, particularly those with detectable amounts of estrogen receptor protein. Hormone therapy is based on blocking the effect of estrogen or lowering estrogen levels. It can be prescribed to a patient before and/or after biopsy, as well as to treat cancer that has spread or regenerated after treatment in the past.

2.6 Summary

Breast cancer is the most common female cancer. The survival rate depends on early diagnosis and the right treatment. Pathology is the gold standard for diagnosis, but it is time consuming. Breast cancer is considered to be a heterogeneous disease, i.e. it is a different disease in different women. Owing to the heterogeneity and complexity of the breast, a good understanding of breast pathology is necessary for the correct development of a diagnostic algorithm.

Chapter 3. Theory and Instruments

The research presented in this thesis is focused on the application of diffuse reflectance, intrinsic fluorescence, and Raman spectroscopies to diagnose breast cancer. The theories have been explained in more detail in the theses of Abigail Haka [Haka, 2004] and George Zonios [Zonios, 1998]. In this thesis, we will discuss only the key-points. For more details refer to the works mentioned above.

3.1 Raman Scattering Theory

The Raman Effect was first discovered by C.V. Raman in 1928. For this discovery Raman was awarded the Nobel Prize in Physics in 1930. The Raman Effect involves the exchange of energy between light and matter. When light travels through a medium, the scattering occurs over all directions and both the intensity and polarization of the scattered radiation depend on the direction of observation. Most of the scattered light is of the same wavelength as the incident photons. This type of scattering is elastic and is called Rayleigh scattering. A small part of scattered light is inelastically scattered. The Raman Effect is an inelastic scattering process in which photons incident on the sample take energy from (Stokes) or transfer energy to (anti-Stokes) the sample's rovibrational modes. Raman is a two-photon process that can be viewed as the simultaneous absorption of an incident photon and emission of a Raman photon. The difference between the energies of these two photons corresponds to the transition of the molecule from one energy state to another. Since the energy level diagram is unique for every molecule, Raman spectra are chemically specific. Individual bands in the Raman spectrum are characteristic of specific molecular motions. Also, as Raman scattering involves a net exchange between the radiation and the molecule, it is often classified as an active process. The generation of the Stokes and anti-Stokes components are illustrated in Figure 3.1.

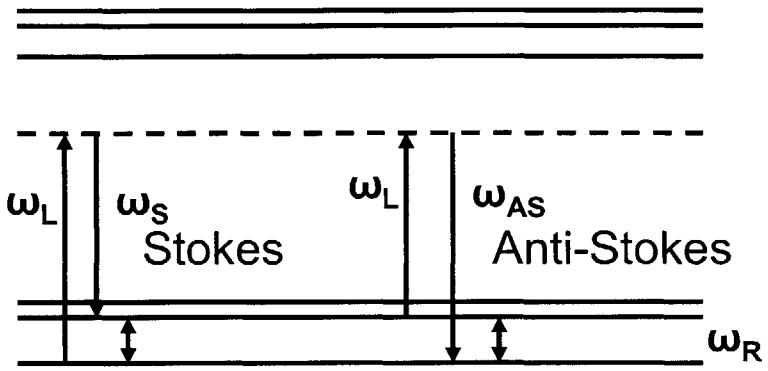

Figure 3.1 Energy level diagram of spontaneous Raman scattering: Stokes and Anti-Stokes generation

The energy difference ($\hbar\omega_R$) between the incoming photon with frequency ω_L and the scattered Stokes photon with frequency ω_S matches a molecular vibration in a gas or liquid, or a phonon in a solid Raman medium. When the scattered light has a lower frequency than the exciting photon it is called the Stokes line ($\omega_S = \omega_L - \omega_R$). When the scattered light has a higher frequency than the exciting photon it is called the anti-Stokes line ($\omega_{AS} = \omega_L + \omega_R$)

The Stokes lines are usually much more intense than the anti-Stokes lines because in thermal equilibrium the population of the excited state is much smaller than the population in the ground state. At thermal equilibrium the fraction of the molecules in one vibrational energy level relative to another is given by the Boltzman distribution:

$$\frac{N_1}{N_2} = \frac{g_1}{g_2}\exp(-\frac{\Delta E}{kT}) \qquad (1)$$

where N_2 and N_1 are the numbers of molecules in a higher and lower vibrational energy level, respectively; g_2 and g_1 are the degeneracies of the higher and lower vibrational levels respectively; ΔE is the energy difference between the higher and lower vibrational energy levels; k is the Boltzman constant; and T is the temperature in degrees Kelvin. The anti-Stokes Raman intensity becomes infinitely small relative to the Stokes Raman intensity for high energy vibrations or low temperatures. Population in the higher energy level is proportional to $\exp(-\hbar\omega_R/kT)$.

3.2 Raman Instrument and Probe

The clinical Raman instrument, developed in the Spectroscopy laboratory, is illustrated in Figure 3.2a. Light from an 830-nm diode is collimated by two cylindrical lenses, directed through a bandpass filter, redirected by a gold-coated mirror and focused onto the Raman probe excitation fiber by a 10×microscope objective.

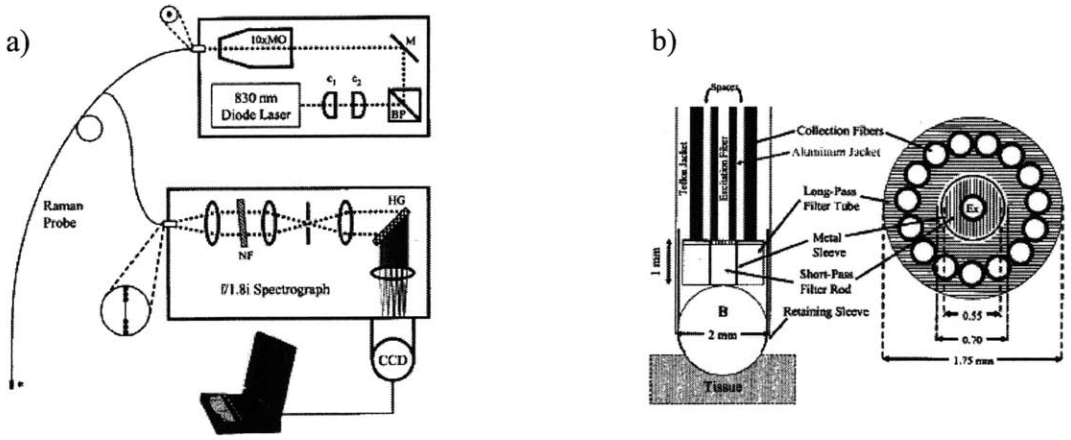

Figure 3.2 a)Schematic of the Raman spectroscopy system. b)Schematic of the Raman probe tip

The proximal linear array of collection fibers from the Raman probe are the input for the *f*/1.8 spectrograph, which collimates the light before it is notch-filtered and focused onto a slit, and then recollimated for dispersion by the holographic grating. Finally, the dispersed light is focused onto a liquid-nitrogen cooled, back-illuminated, deep depletion CCD detector, which is interfaced with a laptop computer. The schematic of the Raman probe used in the study is presented in Figure 3.2b. The probe has a total diameter of 2 mm and a length of 4 m.

3.3 Diffusion theory

Diffuse reflectance spectroscopy (DRS), also sometimes known as elastic scattering spectroscopy, is a non-invasive technique that uses the interaction of light with a sample through absorption and scattering phenomena to produce a characteristic reflectance spectrum that provides information about the structure and composition of the medium.

Diffusion theory is appropriate in a medium dominated by scattering rather than absorption so that each photon undergoes many scattering events before being terminated by an absorption event. The photon has a relatively long residence time that allows it to engage in a random walk within the medium.

Elastic scattering is the most prominent optical property of tissue. The scattering coefficient, μ_s, which represents an average number of scattering events per unit length, is in the range of μ_s=10-100mm^{-1} for soft tissue. This range indicates that light is scattered several tens of times per millimeter of path length while traveling in tissue, i.e. it quickly becomes diffusive. The angle-dependent scattering pattern of the scattering depends on the scatterer size/wavelength ratio. Rayleigh scattering occurs when the scatters are much smaller than the wavelength of light. In such a case, the scattering pattern is isotropic. In the case when the scatterer size is much larger than the wavelength, scattering is highly forward directed, which is the case in tissue.

Biological tissue is a complex and heterogeneous material. The exact origins of scattering in tissue are not well known. Scattering arises from the microscopic variations and inhomogeneities of the refractive index that correspond to various scattering centers. [Zonios, 1998]. The angular distribution of the scattering is described by the scattering phase function, $p(\theta)$, which provides the probability of a photon to be scattered at an angle, θ, with respect to the initial direction. The scattering coefficient is essentially the cross-sectional area per unit volume of tissue. The reduced scattering coefficient is a lumped property incorporating the scattering coefficient μ_s and the anisotropy g:

$$\mu_s' = \mu_s(1 - g) \text{ , where } g = 2\pi \int p(\theta) \cos\theta \sin\theta d\theta \qquad (2)$$

Anisotropy, g, is used to indicate how strongly forward-directed the scattering is. Typical values for tissue are g = 0.8 – 0.95, which corresponds to average scattering angles between 45 and 20 degrees, respectively. The purpose of μ_s' is to describe the diffusion of photons in a random walk of step size of $1/\mu_s'$ [cm] where each step involves isotropic scattering. Such a description is equivalent to the description of photon movement using many small steps $1/\mu_s$ that each involve only a partial deflection angle θ if there are many scattering events before an absorption event, i.e., $\mu_a << \mu_s'$. This is very appropriate and

convenient for modeling multiple scattering processes where the light quickly becomes diffusive.

The problem of scattering of a plane wave by homogeneous spherical objects is known as Mie theory. Mie theory provides an exact solution for scattering of light with wavelength λ by a spherical object with diameter d and refractive index n that is located in a surrounding medium with refractive index n_o. Experimental evidence [Zerull 1976] indicates that by averaging over orientation and/or size, the scatter properties of nonspherical particles can be equivalent to that of spheres, i.e. Mie theory can be applied. In this study μ_s' was modeled using a power law in the following way:

$$\mu_s' = A\lambda^{-B} + C\lambda^{-4} \tag{3}$$

where the first polynomial term represents Mie scattering and the second term represents Rayleigh scattering. In this formula, A can be interpreted as the amount (or concentration) of Mie scatterers, B is related to the sizes of the scatterers, and C is the amount of Rayleigh scatterers. The exact physical meanings of these parameters are still under investigation. However, upon analysis it was discovered that C is equal to zero, which modified formula 3:

$$\mu_s' = A\lambda^{-B} \tag{3a}$$

Absorption

Absorption is measured by means of the absorption coefficient, μ_a, which is defined as the amount of absorption present in units of inverse length. Water, being the major component of soft tissue has negligible absorption in the visible range ($<10^{-4}$ mm^{-1}) [Hale and Querry, 1973]. Tissue absorption, which depends on wavelength, is due to a number of characteristic biochemical molecules. Strong absorption prevents photons from engaging in an extended random walk. Typical values for the absorption coefficient in the visible range are $0.005 < \mu_a < 5$ mm^{-1}.

Nearly all types of tissue contain hemoglobin because its main function is to transport oxygen from the lungs: a necessity for the metabolism and normal functioning of all living tissue. Hemoglobin is a major absorber in the blue region of the visible spectrum. Other biochemical compounds with characteristic light absorption in the visible range

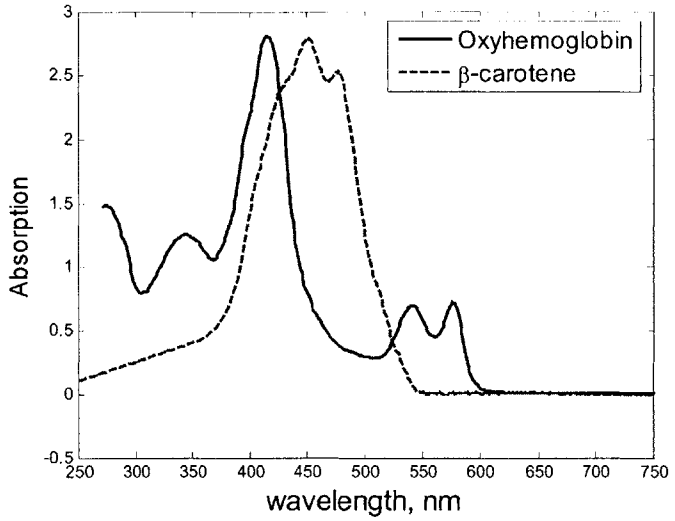

Figure 3.3 Absorption spectra for oxyhemoglobin and β-carotene

include β-carotene, bilirubin, and in the skin, melanin. Many other absorbers are present in the tissue at much smaller concentrations and can be ignored. In this study, oxyhemoglobin and β-carotene will be treated as the major absorbers of light in breast tissue. Their absorption spectra are illustrated in Figure 3.3.

The total absorption coefficient, μ_a, can be expressed as a linear combination of the extinction spectra of the two absorbers as follows:

$$\mu_a = c_{HbO_2} \cdot \varepsilon_{HbO_2}(\lambda) + c_\beta \cdot \varepsilon_\beta(\lambda)) \qquad (4)$$

where c_β is the concentration of β-carotene and c_{HbO_2} is the concentration of oxyhemoglobin.

3.4 Fluorescence

Fluorescence is one of the categories of luminescence and occurs from electronically excited states. The fluorescence lifetime is typically 10^{-9} s. It is a measure of the average time between its excitation and its return to the ground state. Many fluorophores have lifetimes on the order of nanoseconds. Figure 3.4 illustrates the fluorescence phenomenon. A fluorophore is excited to a higher electronic level by absorption of a photon. It relaxes quickly from a higher vibrational level to the lowest level of the excited

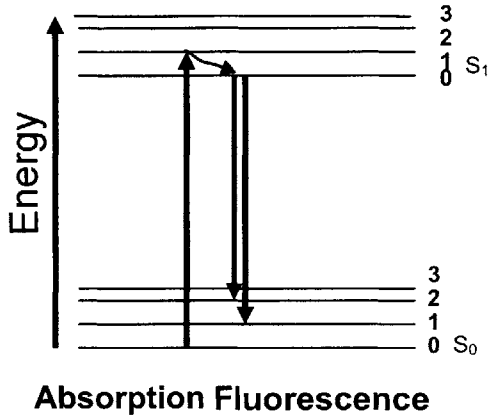

Figure 3.4 Molecular energy diagram depicting fluorescence

state through nonradiative processes and then emits a photon to reach the ground state. From this process it is clear that the energy of emission is typically lower than that of absorption. As a result, fluorescence photons are typically at lower energies or longer wavelengths than the exciting photon. This phenomenon was first observed by Sir G.G. Stokes in 1852.

Fluorophores are divided into two general classes: intrinsic, which occur naturally, and extrinsic, which are added to a sample that does not display the desired spectral properties. Among biological molecules, fluorescence can be observed from reduced nicotinamide adenine dinucleotide (NADH), oxidized flavins (FAD, the adenine dinucleotide, and FMN, the mononucleotide), pyridoxal phosphate, and chlorophyll.

Fluorescence spectroscopy has been widely explored as an important medical diagnostic technique. Promising results have been reported when tissue autofluorescence is used to detect a diversity of diseases, such as atherosclerosis in the aorta and the coronary artery and dysplasia in the colon and other tissues. A majority of the diagnostic methods employed utilize empirical algorithms derived from studying a limited number of specimens. Such empirical algorithms

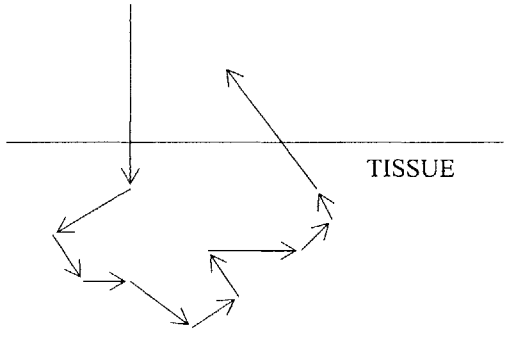

Figure 3.5 Photon migration of fluorescence

ignore the wealth of biochemical and/or morphological information contained in the tissue spectrum. Part of the difficulty has been that fluorescence spectra observed from optically thick tissue are distorted from the intrinsic spectra of individual fluorescence chromophores by the interplay of factors such as scattering, absorption, geometry, and tissue boundary conditions. A model that removes the distorting effects on the fluorescence spectra of optically thick tissues allows connection of the clinically obtained in vivo tissue spectra with their biochemically based fluorophores. Algorithms that incorporate the effects of the intrinsic fluorescence – scattering, absorption, excitation and collection geometries, and the tissue boundary conditions – must be developed in order to move beyond the current empirical approaches. The algorithm determined from the photon migration approach suggests that the distortion in a fluorescence spectrum can be removed by measuring the diffuse reflectance spectrum over the same wavelength range, and in the same manner as the fluorescence spectrum, and by applying the information extracted from the diffuse reflectance spectrum to the fluorescence spectrum in a well-defined manner.

The fluorescent photon is emitted isotropically and undergoes scattering on its way out of the tissue. In the model it is assumed that the secondary fluorescence generated by the fluorescent photons is negligible. Since in the photon migration model the photon path is determined by g and p, the incident excitation photon and the fluorescent photon induced by it will follow the exact same path in the above conditions. Figure 3.5. shows an example of a photon path in which an incident photon is scattered n times along its path and thus induces n fluorescent photons.

Assuming that g is constant over all wavelengths of the fluorescence spectrum, and knowing that in the fluorescence event the photon is always emitted isotropically, the effective anisotropy coefficient g_{eff} can be shown to be the average value of (N-1) forward directed scattering events with anisotropic coefficient g and a single isotropic fluorescence event with $g_{fluorescence}=0$.

For a semi-infinite geometry and index-matched boundary conditions this expression can be further simplified as [Zonios, 1998]:

$$F(\lambda_x, \lambda_m) = \phi(\lambda_x, \lambda_m)\{1 - R[a(\lambda_x), g_{eff}]\}R[a_{eff}(\lambda_m), g_{eff}] \qquad (5)$$

3.5 Optical Spectroscopy of Benign/Malignant Breast Tissues

The IFS spectrum can be decomposed into a linear combination of the spectra of fluorophores associated with morphological structures in the tissue to provide diagnostic information

The endogenous fluorophores present in breast tissue include tryptophan, reduced nicotinamide adenine dinucleotide (phosphate) (NAD(P)H), flavin adenine dinucleotide (FAD), and collagen. NAD(P)H and flavoproteins are indicators of metabolic activity and have excitation/ emission maxima at 351/460 nm and 450/520 nm, respectively. Collagen is the primary structural protein in the extracellular matrix. It has several excitation/emission maxima, one of which occurs at 325/390 nm.

Gupta and Majumder et al. analyzed different data sets collected from the same set of breast tissues *ex vivo* and showed that the emission spectra at excitation wavelengths of 340 and 488 nm and excitation spectra at emission wavelengths 390 and 460 nm exhibit significant differences between normal, benign and malignant tissues. The fluorescence was attributed to reduced nicotinamide adenine dinucleotide (NADH) and collagen. Spectral differences observed in the fluorescence spectra of normal, benign and malignant breast tissues can also be attributed in part to non-fluorescent absorbers and scatters. Diffuse reflectance spectroscopy provides a direct measurement of the tissue absorption and scattering [Gupta,1997, Majumder,1998] .

According to Majumder et al., the fluorophores responsible for the 340, 390, 440, and 520 nm emission bands are amino acids (tryptophan), structural proteins (collagen and elastin), the co-enzyme (NADH), and flavins, respectively [Majumder,1998]. It is known that the excitation spectra recorded from the breast tissue samples for 340, 390, and 460 nm emission consists of spectral bands with peaks around 290, 335 and 340 nm, which are characteristic excitation peaks for tryptophan, collagen and NADH, respectively. The larger intensities of the 340 nm band in the excitation spectra, corresponding to 460 nm

emission, for cancerous tissues would suggest a larger concentration of NADH in cancerous tissues as compared to benign tumors and normal tissues.

Also, it has to be noted that light excited at different wavelengths penetrates to different depths. We define the optical penetration depth as the depth at which the power of light incident on a tissue sample falls to 1/e of its incident value. The sampling depth characterizes the attenuation of both the excitation and the emitted light, which can be at a longer wavelength, as in the case of fluorescence or Raman scattering. Raman signal, achieved using 830 nm excitation wavelength, has larger penetration depth than IFS signal excited at 340 nm or 308 nm wavelengths. Later this information will be used in comparison of Raman and DRS/IFS modalities.

3.6 FastEEM Instrument and Probe

A clinical instrument for DRS/IFS studies, the FastEEM, has been developed in the MIT Spectroscopy Laboratory. This instrument collects white light reflectance and fluorescence excitation-emission matrices (EEM's) within a fraction of a second in order to collect diffuse reflectance, intrinsic fluorescence spectroscopic, and light scattering spectroscopy data.

Diffuse reflectance spectroscopy (DRS) provides information about the morphology and biochemistry of the stromal tissue and epithelium, determining values of the absorption and reduced scattering coefficients, $\mu_a(\lambda)$ and $\mu_s'(\lambda)$, respectively. In the case of breast tissue measurements, we study stromal lesions. Incident white light (300 – 800 nm) undergoes many scattering and absorption events as it propagates through the tissue, and the emerging ("diffusely reflected") light exhibits prominent spectral features caused by the interplay of scattering and absorption. DRS employs a mathematical model based on the diffusion approximation of light propagation in tissue.

The collected fluorescence and reflectance spectra can be used to extract the intrinsic fluorescence spectra (i.e., the fluorescence unaffected by tissue absorption and

scattering). Intrinsic fluorescence spectroscopy (IFS) yields the relative contributions of endogenous tissue fluorophores (e.g. NADH and collagen).

Light scattering spectroscopy (LSS), which is based on Mie theory, extracts epithelial nuclear size distributions from white light reflectance spectra. Light that has been singly scattered in the backward direction can be analyzed to provide the nuclear size distribution. LSS can provide information on number density of nuclei, percentage of enlarged nuclei as well as mean nuclear size. [Tunnell,2003]. The application of LSS technique to diagnose breast cancer will be explored in the future.

A schematic of the FastEEM is presented in Figure 3.6a. It delivers a sequence of ten laser pulses (308- 480 nm) and two white light pulses to the tissue via an optical fiber probe (see below for details of the probe design). The same probe collects the white light reflectance and fluorescence and delivers it to the entrance slit of a diffraction grating spectrometer where it is dispersed onto an intensified CCD detector. All ten laser-induced emission spectra and the two white light reflectance spectra are collected in approximately 0.3 s. Several of these acquisitions can be averaged together to increase the signal-to-noise ratio (SNR). Previously, we found that the acquisition of five measurements provides sufficient SNR in most tissues, making a typical acquisition time

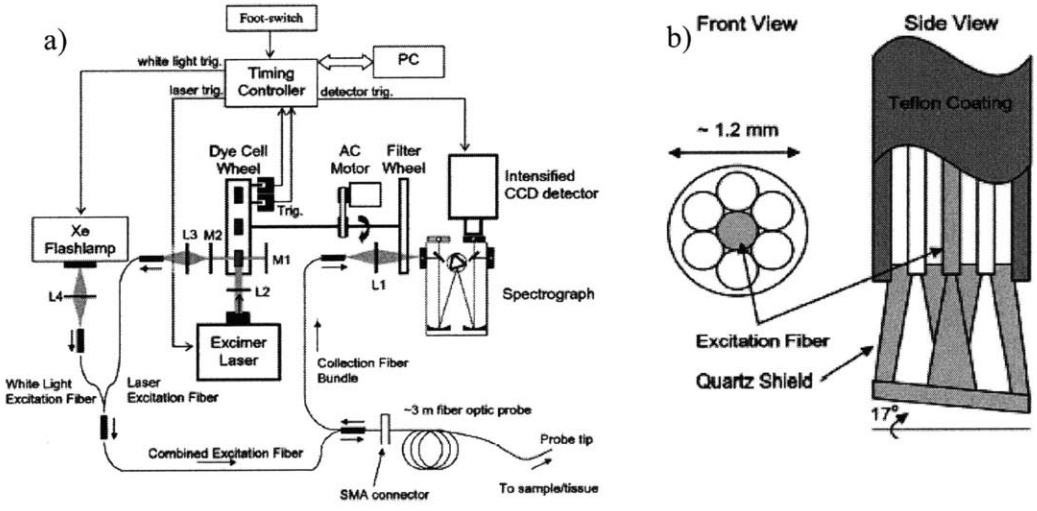

Figure 3.6 a) FastEEM clinical spectrophotometer. b) Schematic diagram of the distal tip of the optical fiber probe

on the order of 1.5 s. The optical fiber probe, which is used for light delivery and collection, is illustrated in Figure 3.6b. It is in the form of a flexible catheter, with an overall length of over 3 m and a diameter of approximately 1.2 mm.

3.7 Summary

In this chapter the theories behind the spectroscopic techniques employed in this work, DRS, IFS and Raman, were presented. The theories are explained in more detailed in [Haka, 2005] and [Zonios, 1998]. A brief overview of the instrumental design and collected data was also presented. The data obtained with Raman spectroscopy will be analyzed in a prospective manner based on the algorithm developed by Abigail Haka. DRS and IFS will provide additional information on the chemical composition of the tissue.

Chapter 4. DRS/IFS

In 2004, my colleague, Abigail Haka, and I collected DRS/IFS and Raman spectroscopy data on 105 samples from 25 patients during an *ex vivo* study at University Hospitals Cleveland in collaboration with Drs. Fitzmaurice, Shenk, and Wang. This study was undertaken to investigate the use of these spectroscopic techniques for the diagnosis of breast cancer. All studies involving human tissue were approved by the University Hospitals of Cleveland and Case Western Reserve University Institutional Review Board and the Massachusetts Institute of Technology Committee On the Use of Humans as Experimental Subjects. Informed consent was obtained from all subjects prior to the surgical procedures.

Measurements were taken in the frozen pathology room within half of an hour of tissue excision. We obtained reflectance and fluorescence spectra from the specimens with the FastEEM instrument first, followed by collection of Raman spectra with the clinical Raman instrument. Care was taken in placing the Raman probe at the same site on the tissue as the FastEEM probe. Overall, we obtained 223 spectra (approximately two spectra per specimen). After the DRS/IFS and Raman data were taken, breast specimens were fixed and given to a pathologist for a diagnosis. The exact spot of probe placement was marked with ink to aid in our comparison of spectroscopy and pathology. The pathologies of the specimens included 33 normals, 59 fibrocystic changes (FCC), 9 fibroadenomas, 13 ductal carcinoma in situs (DCIS), and 9 infiltrating ductal carcinomas (IDC), as determined by Dr. Fitzmaurice, a board-certified pathologist.

The focus of this chapter is on the development of the diagnostic algorithm for DRS/IFS. Subsequent chapters will focus on the use of a previously developed diagnostic algorithm for the Raman data set, and the development of a diagnostic algorithm using combined information from DRS/IFS and Raman spectroscopy.

4.1 Data Processing

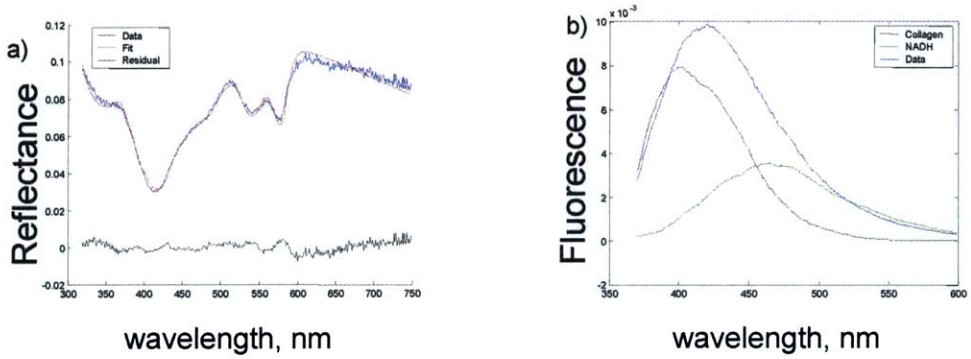

Figure 4.1 a) a representative DRS spectrum and fit; b) representative IFS spectra obtained with 340 nm excitation. The original spectrum acquired from breast tissue is in blue and the contributions of NADH (green) and collagen (black) were found via multivariate curve resolution (MCR)

Reflectance spectra are analyzed using the diffusion approximation to extract tissue morphological properties such as scattering, oxidized hemoglobin concentration, and β-carotene concentration. The intrinsic fluorescence photon-migration model was used to correct the fluorescence spectrum for distortions introduced by tissue absorption and scattering. This corrected spectrum was then used to extract the contributions of the biochemical tissue constituents NADH and collagen. Examples of DRS and IFS spectra are displayed in Figure 4.1. The basis spectra used to model the IFS data shown in Figure 4.1b were extracted using multivariate curve resolution (MCR), an iterative statistical technique that will be discussed further in section 4.2.

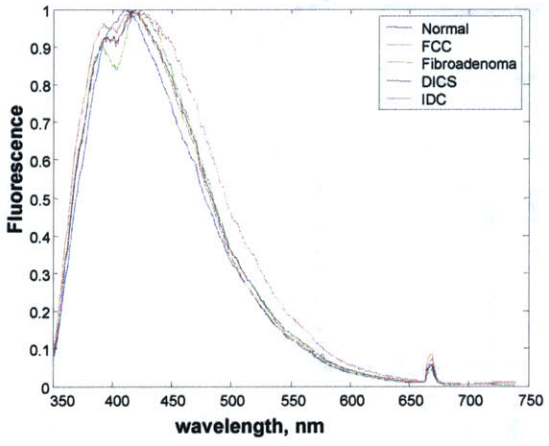

Figure 4.2 normalized intrinsic fluorescence spectra of breast tissue at 340 nm excitation. Peak at 680 nm represents second order laser peak

Figure 4.2. displays average IFS data from each of the pathologies encountered in this study. Differences are observed even without a detailed analysis. The full width at half-maximum (FWHM) is largest for IDC and

then decreases for fibroadenoma, DCIS, FCC and is smallest for normal tissue. A large FWHM possibly indicates a higher concentration of NADH.

Three different excitation wavelengths (308, 340, and 360 nm) are analyzed in order to reveal different fluorophores that could each provide information useful to a diagnostic algorithm. Also, 400 nm excitation wavelength was analyzed, but information from this wavelength did not provide any diagnostic importance. Table 4.1. presents an overview of fluorophores that may be present in the breast tissue and fluoresce at particular excitation wavelengths. However, upon analysis it was discovered that some of these fluorophores are not present at high enough levels in our samples to be detected. These include tryptophan excited at 308 nm, elastin excited at 340 nm, and porphyrin excited at 360 nm.

	Excitation wavelength		
	308 nm	340 nm	360 nm
Fluorophores	NADH	NADH	NADH
	Collagen	Collagen	Collagen
	Tryptophan	Elastin	FAD
			Porphyrin

Table 4.1 Fluorophores at different excitation wavelengths

Representative spectra of DRS and IFS collected with different excitation wavelengths are illustrated in Figure 4.3. Not all of the data collected were subsequently used for analysis. Specifically, DRS data with overall reflectance less than 1 percent were excluded because of the inability to use this information to process the fluorescence data to obtain the intrinsic fluorescence. An example spectrum that was excluded from our data set is shown in Figure 4.4. Furthermore, though most of the fits are observed to be adequate, in cases of high oxyhemoglobin concentration, the region between 660 – 750 nm fits less well. We believe this to be a result of the fitting procedure and is a topic that requires further investigation at a later date.

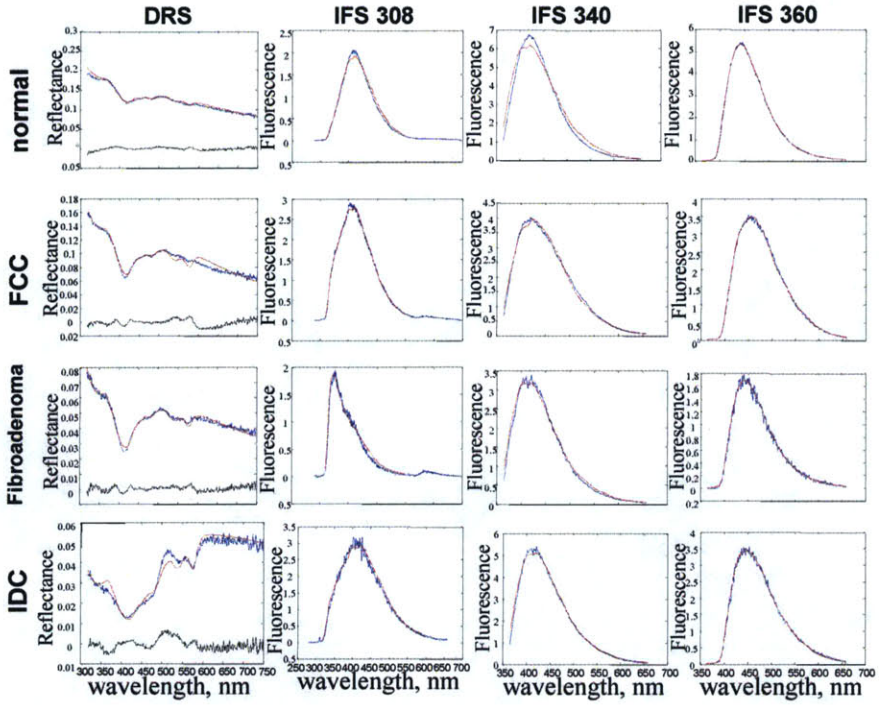

Figure 4.3 Data and fits for DRS and IFS at 308, 340, and 360 nm excitation wavelengths for different pathologies: data (blue), fit (red), residual (black). Spectra at 308 and 360 nm excitation have NADH and collagen, 360 nm excitation spectra fitted with NADH, collagen and FAD.

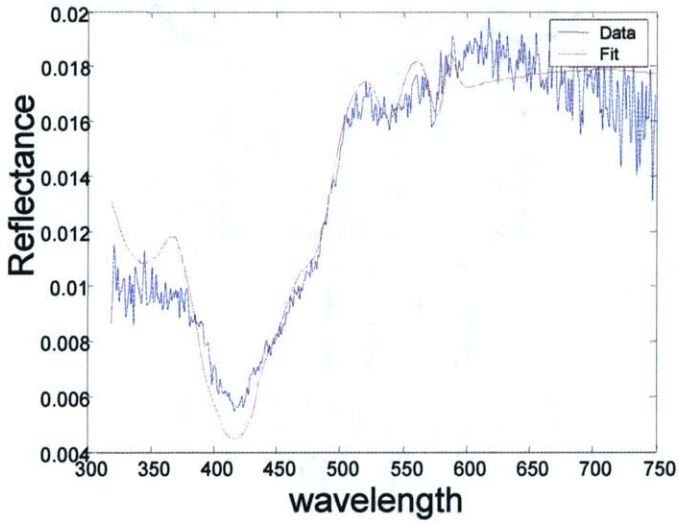

Figure 4.4 Example of a DRS spectrum that was excluded from analyzed data due to reflectance of less than 1%: data shown in blue: fit (red)

4.2 MCR Analysis

In order to investigate the physical meaning of intrinsic fluorescence parameters, a chemical study was performed in the Spectroscopy Laboratory using the FastEEM instrument. The purpose of this study was to ensure that the fit coefficients resulting from an MCR analysis have physical meaning. To test this assumption, fluorescence spectra obtained from pure components (elastin, FAD, tryptophan) were used to fit data obtained from mixtures of these pure components using ordinary least squares (OLS) fitting to confirm linearity. Once the system was shown to be linear, parameters obtained from MCR were used to fit the mixture data and the resulting fit coefficients compared to expected values. Further, we examined whether the initial guess spectra used as inputs in multivariate curve resolution affected these results, as will be discussed in more detail below.

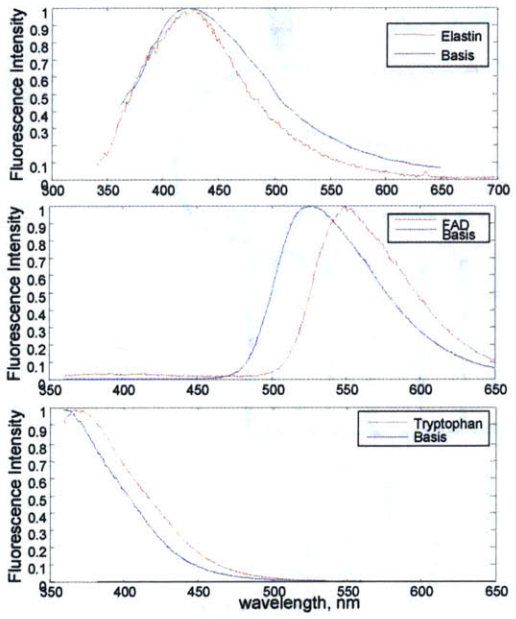

Figure 4.5 Examples of pure component spectra of elastin, and FAD at 340 nm, also tryptophan at 308 nm excitation are shown in red and the previously collected basis spectra from tissue that correspond to these fluorophores are shown in blue.

FAD, tryptophan, and elastin were dissolved in water for two hours. Calibration spectra were collected to ensure the DRS/IFS system was operating correctly. Examples of fluorescence spectra collected from elastin, and FAD at 340 nm excitation wavelength, and tryptophan at 308 nm excitation are compared with the basis spectra previously collected from tissue in Figure 4.5. We observe that the peak intensities of the spectra collected from tissue are shifted to the blue region compared to spectra collected from pure chemicals. This shift in emission wavelength is expected because there are different microclimates, which means that

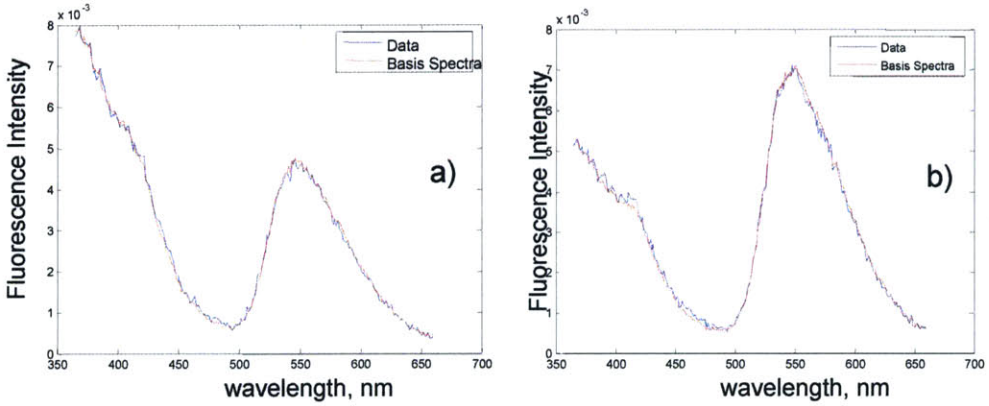

Figure 4.6. Mixture data: FAD, elastin and tryptophan; a) 1:1:1 concentrations ; b) 2:2:1 concentrations, respectively, data is shown in blue, fit is in red

different fluorophores present in the tissue and interact with each other, in the tissue that modify the spectral characteristics of the fluorophores [Richards-Kortrum, 1996].

Mixtures of FAD, elastin and tryptophan were created in concentration ratios of 1:1:1, 1:2:1, and 2:2:1. As expected, a linear combination of the spectra from pure components weighted according to their concentration ratios fit the data well, thus confirming system linearity. Representative spectra and fits are shown in Figure 4.6.

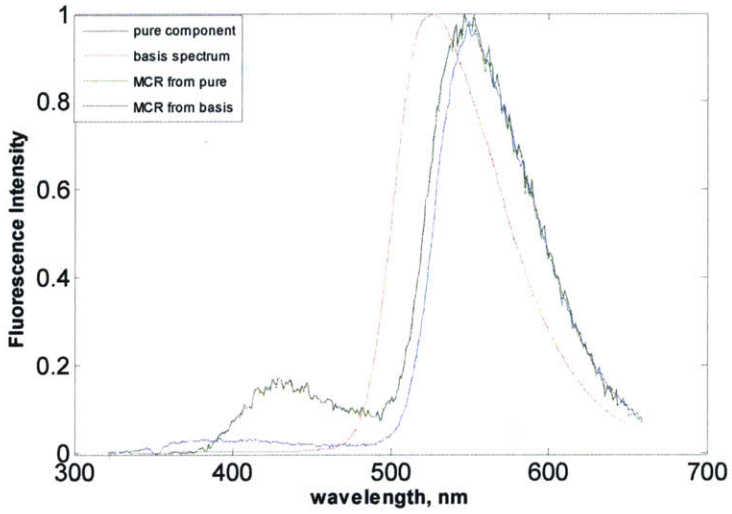

Figure 4.7 FAD: pure (blue), basis spectrum (red), MCR from pure (green), MCR from basis spectrum (black)

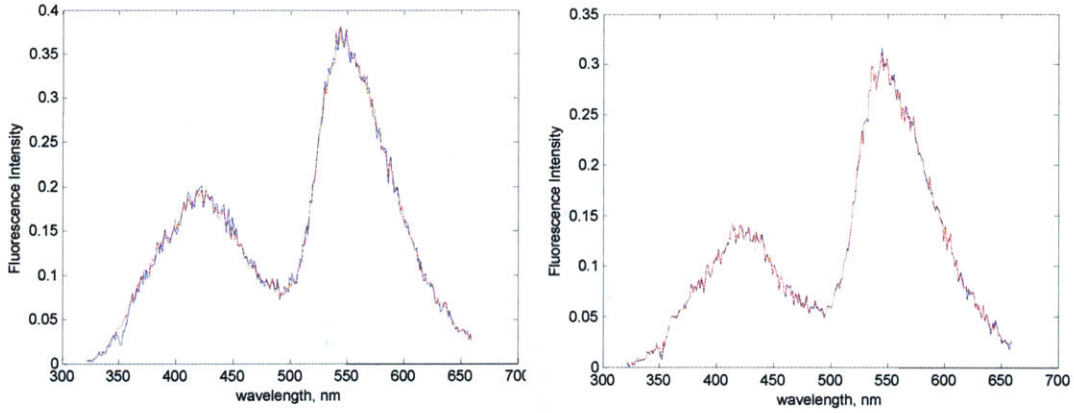

Figure 4.8 MCR from pure components for fitting mixture data: FAD, elastin; a) 1:2 concentrations: b) 1:3 concentrations, respectively. data (blue), fit (red)

MCR calculates basis spectra by minimizing the fitting error of a given spectrum using an initial guess spectrum as the input. For the following analysis we extracted MCR basis spectra twice: once using the pure component spectra as the initial guess and once using basis spectra that were extracted from tissue measurements made by Irene Georgakoudi as an initial guess. The basis spectra that MCR calculates are referred to subsequently as MCR components. Although the original spectra were noisy, we were able to achieve good fits regardless of the initial guess spectra used in MCR. Figure 4.7. shows the FAD spectrum of the pure chemical along with its respective basis spectra obtained from tissue measurements and the MCR components extracted using different initial guess spectra. The resulted MCR spectra are practically identical, which suggests that minimization of the fitting error does not rely heavily on the initial guess spectrum. This is ideal as it is

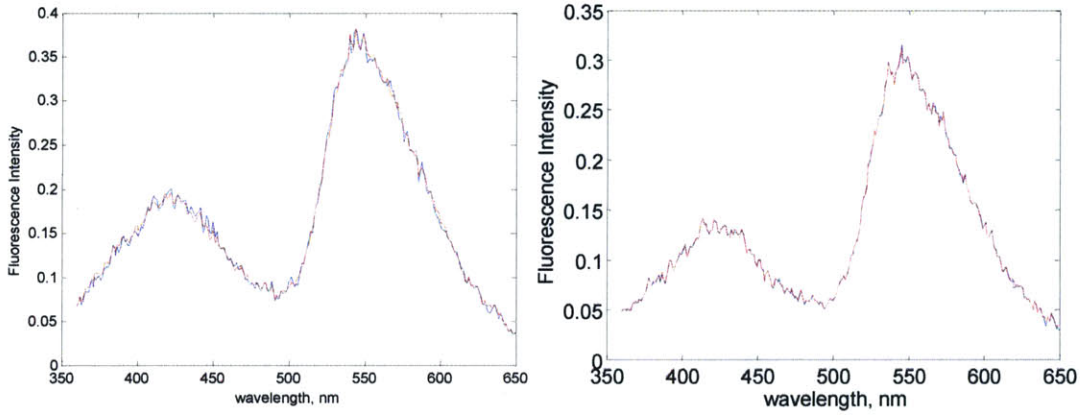

Figure 4.9 MCR from basis spectra for fitting mixture data: FAD, elastin: a) 1:2 concentrations; b) 1:3 concentrations, respectively. data (blue), fit (red)

difficult to know the exact spectral shape of the various fluorophores present in tissue owing to the different tissue microclimates. The data from mixtures of FAD and elastin in ratios of 1:2 and 1:3 are illustrated in Figure 4.8 along with fits obtained from MCR components that used pure component spectra as the initial guess spectra. The fit coefficients for a 1:2 ratio of FAD:elastin were 0.343 and 0.657, respectively. The fit coefficients for a 1:3 ratio of FAD:elastin were 0.255 and 0.745, respectively. The error was no more than 3 percent.

The same data from Figure 4.8 is displayed in Figure 4.9 with fits derived from MCR components that used the tissue basis spectra as the initial guess spectra. The fit coefficients for a 1:2 ratio of FAD:elastin were 0.356 and 0.644, respectively. The fit coefficients for a 1:3 ratio of FAD:elastin were 0.255 and 0.745, respectively. In this case the error was no more than 7 percent.

From this analysis we can conclude that in a simple system MCR is capable of obtaining reasonable fit coefficients with a physical meaning from spectra of mixtures.

Based on the results of this study we analyzed IFS spectra of the *ex vivo* tissue using MCR components. The basis spectra taken from separate tissue measurements were used as the initial guess spectra. The basis spectra subsequently obtained using MCR are thought to represent NADH and collagen because they have similar, though not identical, lineshapes to the spectra of commercially available NADH and collagen. The lineshape and the wavelength maximum of a fluorescence peak obtained from a solution of a pure

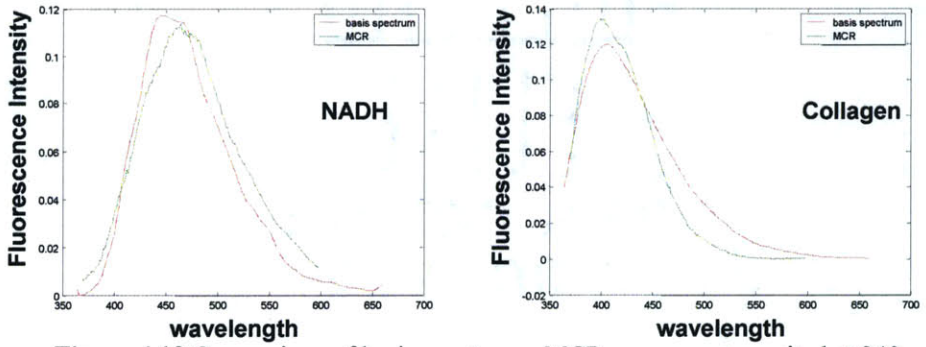

Figure 4.10 Comparison of basis spectra vs. MCR components excited at 340 nm. Basis spectra (red); MCR (green)

component is known to be different than that obtained from the same component in a different chemical environment, such as tissue. [Shafer-Peltier,2001] A comparison of basis spectra vs. MCR components for 340 nm excitation is shown in Figure 4.10.

4.3 Algorithm Development

The DRS/IFS algorithm was developed using leave-one-out cross validation and logistic regression. The desired algorithm must be able to distinguish among the 5 major pathologies. From an examination of breast histopathology, it is known that normal breast tissue consists mostly of adipocytes (fat) while the progression of malignancy includes an increase in the amount of collagen. Therefore, we expect that normal tissue can be separated from the remaining pathologies by the relative presence of collagen and β-carotene, which is fat-soluble. Also from histopathology, fibroadenoma displays an increased cellular density. Because the parameter A is representative of the number of scatterers in the tissue, we expect fibroadenomas to have a relatively high A parameter. Furthermore, we expect the NADH contribution, which is representative of cellular metabolism, to be less than that from cancerous tissue.

By maximizing the sensitivity and specificity of each stage of the algorithm we were able to identify the diagnostic parameters that can distinguish between pathologies in the breast tissue. The diagnostically-relevant parameters from DRS were found to be β-carotene, oxyhemoglobin, and the scattering A parameter. The diagnostically-relevant parameters from IFS were found to be the fit coefficients for NADH at 340 nm excitation and the fit coefficients for collagen at both 340 and 360 nm excitation wavelengths. However, careful examination of the collagen fit coefficients obtained with 340 and 360 nm excitation revealed that only the results from one wavelength were necessary. Because of the slight difference in wavelength between 340 nm and 360 nm, the penetration depths of the light is not sufficient to result in different sampling volumes. However, there is enough variation between the fit coefficients at both wavelengths that averaging them does not provide benefit. This finding is quite important as the number of diagnostic parameters should be minimized in order to prevent overfitting the data set.

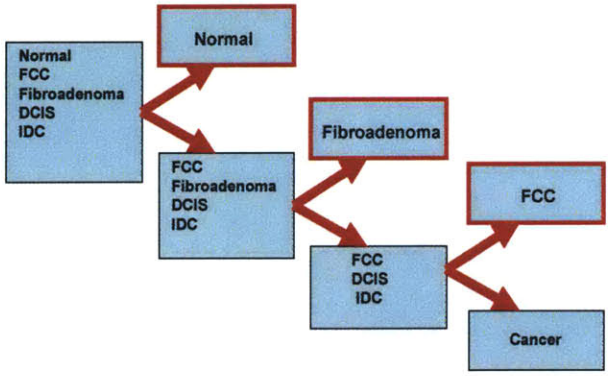

Figure 4.11 Schematic diagram of overall diagnostic algorithm

Therefore, the diagnostic parameters from IFS were reduced to NADH and collagen at 340 nm excitation.

The diagnosis is accomplished in a sequential manner. Normal tissue is first diagnosed and removed from the remaining samples, followed by the diagnosis of fibroadenoma, FCC, and then by the diagnosis of malignant lesions (DCIS and IDC), as depicted in Figure 4.11. This algorithm can be implemented in "real time".

The scatter plots and decision lines for each step of the diagnostic algorithm are shown in Figure 4.12. Figure 4.12a shows that β-carotene versus collagen can separate normal tissue from the rest of the pathologies. After normal tissue was excluded from subsequent steps, fibroadenoma was separated from FCC and IDC using NADH versus the A parameter, as shown in Figure 4.12b. Finally, FCC was distinguished from cancer using collagen versus oxyhemoglobin, as shown in Figure 4.12c. In the latter plot, three

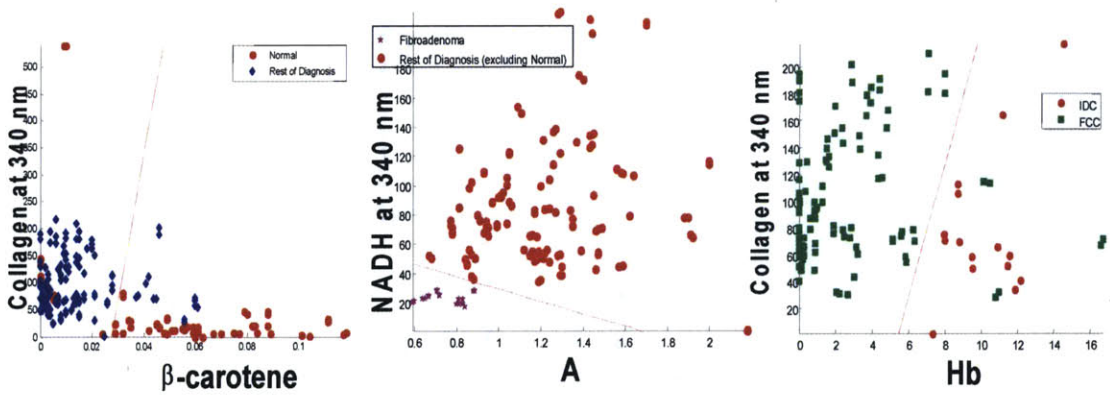

Figure 4.12 Discrimination of pathologies using DRS and IFS parameters

	diagnostic parameters	se, %	sp, %
Normals vs. rest of diagnosis?	-carotene collagen@340	90.5	84.3
Fibroadenoma vs. rest(no normals)	A NADH@340	100	100
FCC vs. IDC	Hb collagen@340	100	97.9

Table 4.2 Sensitivity and specificity using DRS/IFS algorithm

samples (6 spectra) classified as FCC were misdiagnosed as cancer. Two out of those three data samples were previously misclassified as normals. With this knowledge, the sensitivity and specificity of each step are summarized in Table 4.2.

Table 4.3 compares the pathological diagnosis of each sample with the diagnosis from the DRS/IFS algorithm developed for our data set. The sensitivity and specificity for the separation of cancerous and non-cancerous pathologies are 100% and 95.8%, respectively. The overall accuracy (correct prediction of each of the pathologies) is 87.6% (92/105). We note that although we have 100% sensitivity there are only 9 malignant samples.

During the clinical study we also obtained spectra from specimens that were later diagnosed as DCIS. A total of 15 specimens were analyzed in the same sequential manner as explained above. Despite a great deal of effort we were unable to distinguish DCIS from the rest of the pathologies. An example scatter plot of two diagnostic parameters, collagen and β-carotene, for all of the pathologies is shown in Figure 4.13. The DCIS lesions are seen to fall in all the parts of the schematic, making it impossible to separate them from the rest of the diagnoses. One possible explanation for this failure could be the small number of cancerous cells in the tissue that precluded proper measurement. Further studies will be conducted in order to solve this problem.

Pathology / TMS	Normal	Fibrocystic Change	Fibroadenoma	Invasive Carcinoma
	(32 samples)	(55 samples)	(9 samples)	(9 samples)
Normal	27	7	0	0
Fibrocystic Change	2	47	0	0
Fibroadenoma	0	0	9	0
Invasive Carcinoma	3	1	0	9

Table 4.3 Comparison of pathologic diagnosis with that of the DRS/IFS diagnostic algorithm for *ex vivo* specimens. The DRS/IFS diagnostic algorithm resulted in an overall accuracy of 87.6% (92/105)

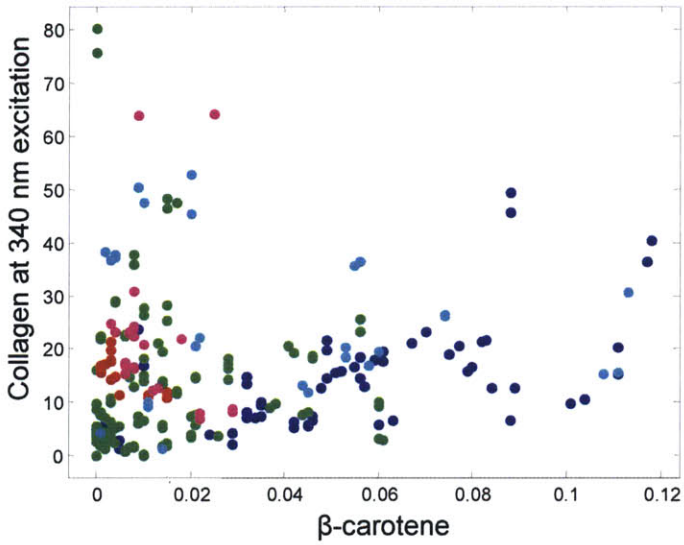

Figure 4.13 Scatter plot for distinguishing amongst various pathologies including DCIS. Normal (deep blue), FCC (green), fibroadenoma (red), DCIS (cyan), IDC (magenta)

4.4 Discussion

This pilot study is the first from our laboratory to use diffuse reflectance and intrinsic fluorescence spectroscopies to examine breast cancer *ex vivo*. This study clearly demonstrates the feasibility of DRS/IFS as a clinical tool for breast cancer diagnosis. Although the results of this study demonstrate the potential benefits of DRS/IFS, it also reveals its weakness, as DRS/IFS was not able to separate DCIS from the rest of the diagnoses. Further investigations will be conducted that will better elucidate the advantages and limitations of DRS/IFS as well as provide a better understanding of the physical meanings of the diagnostic parameters.

Chapter 5. Raman Spectroscopy

We apply Raman spectroscopy to *ex vivo* samples in order to diagnose normal, benign and malignant female breast tissue using a diagnostic algorithm previously developed in our laboratory. Data were collected within half an hour of tissue excision and immediately following DRS/IFS data collection. A total of 105 tissue samples from 25 patients are examined.

5.1 Raman Analysis

As stated in Chapter 3, Raman spectroscopy is a technique capable of providing accurate chemical information of a tissue sample. There are a large number of Raman active molecules in breast tissue and their spectral signatures are sharp and well defined. Previously, we have developed a chemical/morphological model of breast tissue [Shafer-Peltier,2001]. The Raman spectra, sampling a volume of 1 mm^3, are fit to a linear combination of the basis spectra using an ordinary least-squares minimization algorithm with a non-negative constraint. The following basis spectra are used in the algorithm: cell cytoplasm, cell nucleus, fat, beta-carotene, collagen, calcium hydroxyapatite, calcium oxalate dehydrate, and cholesterol-like lipid deposits. The Raman basis spectra are shown in Figure 5.1.

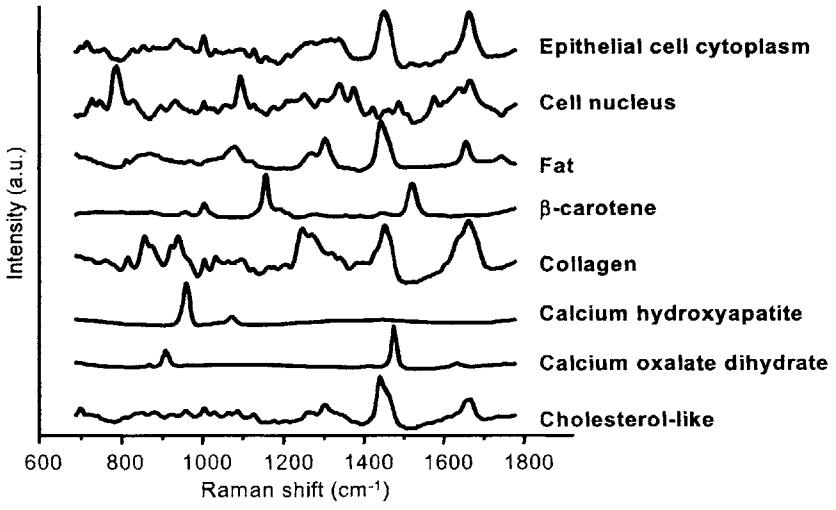

Figure 5.1 Raman morphological model basis spectra

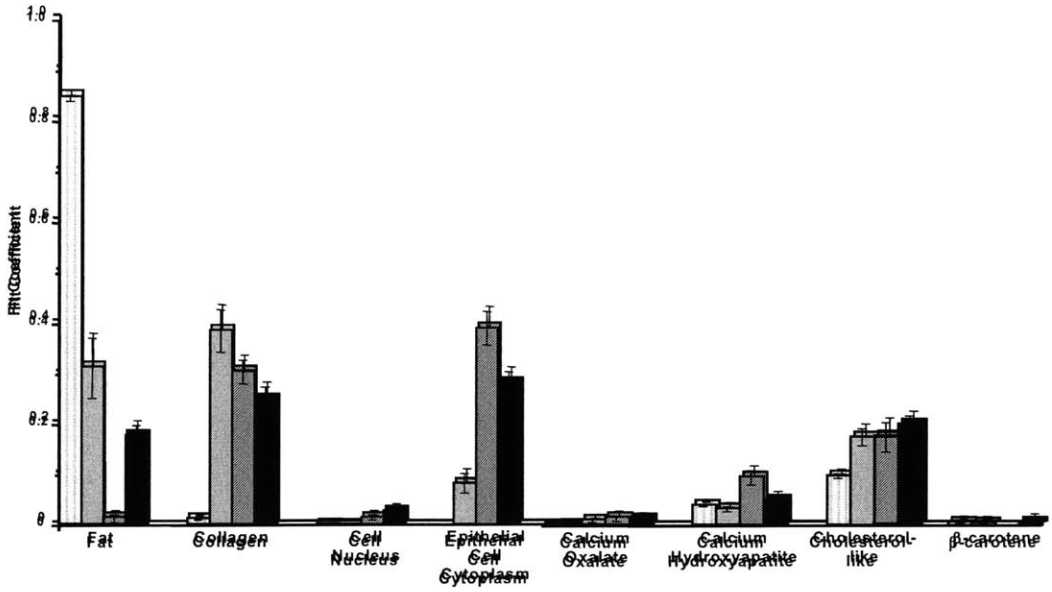

Figure 5.2 Bar graph displaying the average composition of samples diagnosed as normal, fibrocystic change, fibroadenoma, and infiltrating ductal carcinoma.

The fit coefficients, given by the model and normalized to sum to one, represent contributions from chemical components and morphological features to the microscopic tissue spectrum. Figure 5.2 illustrates a histogram of the average fit coefficients associated with normal tissue, fibrocystic change, fibroadenoma, and infiltrating ductal carcinoma. The fit coefficients of the normal breast indicate that it predominantly consists of fat. This is consistent with breast pathology, as normal breast tissue is largely made up of adipocytes cells containing copious amounts of cytoplastic fat, and adipose tissue has a relatively large Raman scattering cross-section. The amount of collagen increases in all abnormal breast pathologies. This happens because lesion formation is often accompanied by scarring process – fibrosis, characterized by an accumulation of collagen.

5.2 Data Measurements and Processing

Data were acquired via a clinical Raman system and Raman optical fiber probe. The average laser excitation power was approximately 150 mW. No tissue damage was observed at this power. The acquisition time varied depending on signal intensity and ranged from 0.5 to 2 seconds. The spectrum of toluene was used to calibrate the

instrument from day to day. The tissue fluorescence background was modeled with a sixth-order polynomial and removed prior to data analysis.

5.3 Diagnostic Algorithm

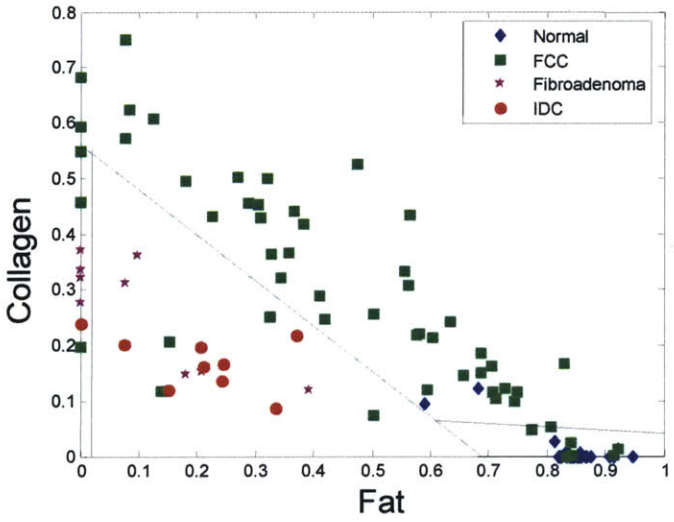

Figure 5.3 Scatter plot displaying the fat and collagen content for all pathologies encountered in this study. Normal (rhombuses), fibrocystic change (squares), fibroadenoma (stars), infiltrating carcinoma (circles).

Data were analyzed using the diagnostic algorithm previously developed on fresh-frozen tissues in our laboratory, as shown in Figure 5.3. Application of this algorithm in a prospective manner resulted in a sensitivity of 88.9 percent, correctly diagnosing 8 out of 9 IDCs, and a specificity of 90.6 percent for cancerous versus non-cancerous tissue. This corresponds to an overall accuracy of 81 percent, (85/105). The limited number of cancerous samples (9 IDCs) greatly influenced the sensitivity of our model. Table 5.11 represents a summary of the prospective use of the Raman diagnostic algorithm

5.4 Discussion

Pathology / Raman	Normal	Fibrocystic Change	Fibroadenoma	Invasive Carcinoma
	(32 samples)	(55 samples)	(9 samples)	(9 samples)
Normal	30	7	0	0
Fibrocystic Change	2	41	0	0
Fibroadenoma	0	3	4	1
Invasive Carcinoma	0	4	5	8

Table 5.1 Comparison of pathologic diagnosis with that of the Raman diagnostic algorithm for ex-vivo specimens. The DRS/IFS diagnostic algorithm resulted in an overall accuracy of 81% (85/105)

The Raman algorithm remained quite robust when applied in a prospective manner. However, this technique failed often for the diagnosis of fibroadenoma with 4 out of 9 specimens misclassified as cancerous. At the same time, it must be noted that due to the nature of fibroadenoma, discussed in Chapter 2, lesions classified as such must be excised. Thus, Raman spectroscopy, while having difficulty distinguishing between fibroadenoma and cancerous lesions, can identify the lesions that must be removed from the body.

This Raman diagnostic algorithm was developed to distinguish between four pathologies at once, and did not include DCIS. An attempt was made to diagnose DCIS specimens without changing the Raman algorithm, but it did not succeed. Studies to incorporate additional fit coefficients and intensity information into the diagnostic scheme are currently underway with the hope that it will aid in diagnosis. However, further work needs to be done in order to properly distinguish among cancerous pathologies. It is clear that additional model parameters need to be investigated to achieve this goal.

Finally, this analysis excludes patients who have undergone pre-operative chemotherapy or who are undergoing re-excision surgery. Such patients harbor pathological tissue changes that we have not previously examined. Pre-operative chemotherapy is used to decrease the size of a large tumor and commonly a tumor is replaced by loose fibrosis. This explains why we see an increase in collagen in all of the specimens from patients who were treated with chemotherapy. If a patient already had an excisional biopsy and is

undergoing a re-excisional surgery it is expected that the amount of collagen will be increased as a result of the scarring process. These two pieces of information are very important because our Raman analysis is based on the amount of collagen and fat present in the breast tissue. The data from pre-operative and re-excisional chemotherapy biopsies need to be analyzed separately from the rest of the data and a new diagnostic algorithm developed. This will be accomplished in the near future.

Chapter 6. Multi-Modal Spectroscopy

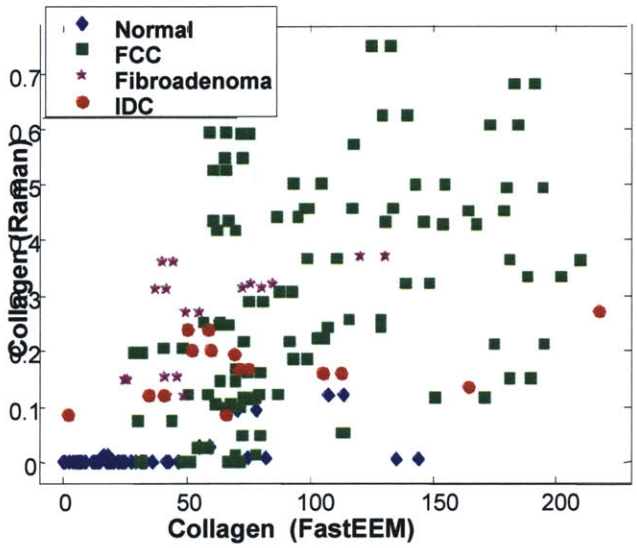

Figure 6.1 Comparison of Raman collagen vs. DRS/IFS collagen

The promising results that were obtained using DRS/IFS and Raman spectroscopic techniques separately raised the possibility of further improving our diagnostic capabilities by combining the respective information into one diagnostic algorithm. Once again, we used the data set from the clinical study performed in Cleveland, OH. In developing the MMS algorithm, only parameters that were diagnostic in each of the modalities were used. The diagnostic parameters from DRS/IFS include A, which represents the amount of scatters, oxyhemoglobin, _-carotene, and NADH and collagen with 340 nm excitation wavelength. The parameters from Raman Spectroscopy are the fit coefficients for fat and collagen. Both of the modalities are sensitive to collagen. However, because each uses a different wavelength of light (Raman at 830 nm and DRS/IFS at 340 nm), the sampling depths are different. As it was mentioned in the Chapter 3, Raman penetrates deeper than IFS excited at 340 or 360 nm wavelength.. This fact explains why collagen fit coefficients extracted via Raman spectroscopy do not strongly correlate with collagen fit coefficients extracted using DRS/IFS, as seen in Figure 6.1.

6.1 Diagnostic Algorithm

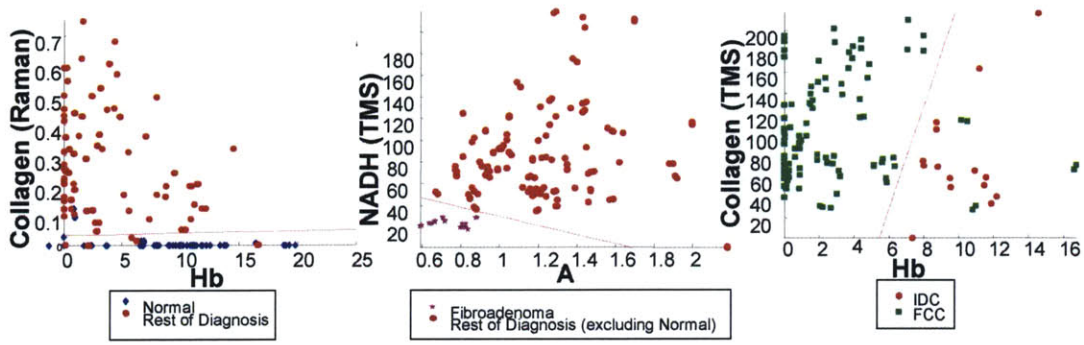

Figure 6.2. Scatter plots for three steps of Multi-Modal Diagnostic algorithm

In order to distinguish among each of the pathologies, leave-one-out cross validation and logistic regression were implemented on the set of data. Because of the fact that more than two parameters must be used to separate all 4 of the pathologies, the analysis was performed in a sequential fashion in the same manner as the DRS/IFS algorithm. Figure 6.2 displays the scatter plots and decision lines for each of the three steps performed in order to achieve separation of the pathologies. In the DRS/IFS diagnostic algorithm developed in Chapter 4, 2 normals were misdiagnosed as FCC and 3 normals misdiagnosed as IDC. By introducing a parameter from the Raman model to the first step of the analysis, which separates out normal pathologies, and incorporating it with the DRS/IFS parameter of oxyhemoglobin, we achieve a greater number of correctly diagnosed normals. Only two remain misdiagnosed as FCC, a benign condition. The two subsequent diagnosis steps remain identical to that used in DRS/IFS because in the

Pathology / Multimodal	Normal	Fibrocystic Change	Fibroadenoma	Invasive Carcinoma
	(**32** samples)	(**55** samples)	(**9** samples)	(**9** samples)
Normal	**30**	**4**	**0**	**0**
Fibrocystic Change	**2**	**49**	**0**	**0**
Fibroadenoma	**0**	**0**	**9**	**0**
Invasive Carcinoma	**0**	**2**	**0**	**9**

Table 6.1 Comparison of pathologic diagnosis with that of the MMS diagnostic algorithm for ex vivo specimens. MMS diagnostic algorithm resulted in an overall accuracy of 92.4% (97/105)

second step all 9 fibroadenoma lesions are correctly diagnosed, while in the third step both DRS/IFS and Raman failed for the same two samples. The fact that both modalities missed the same limited number of samples suggests a possible registration error, which needs to be clarified with the pathologist. Table 6.1 provides an overview of the MMS algorithm performance. This algorithm improves the number of correctly diagnosed normal and fibrocystic changes over that from either DRS/IFS or Raman modalities. Additionally, a fewer number of pathologies are misclassified as cancerous, which is an obvious benefit in that it increases the sensitivity and overall accuracy of this technique.

6.2 Discussion

The results of this study clearly show that a combination of diffuse reflectance, intrinsic fluorescence, and Raman spectroscopies provide results better than that obtained from either technique, as shown in Table 6.2. By improving the sensitivity and specificity for malignant versus non-malignant lesion of our techniques we will provide doctors with the ability to achieve an accurate diagnosis in real time.

	Raman	FastEEM	Multimodal
Sensitivity	88.90%	100%	100%
Specificity	90.60%	95.80%	97.90%
Overall acuracy	81%	87.60%	92.40%

Table 6.2. Comparison of three different techniques. Multi-modal technique provides better sensitivity and specificity, as well as overall accuracy

Chapter 7. Future Directions

This chapter focuses on the future of Raman, DRS/IFS and Multi-Modal spectroscopies in breast cancer diagnosis and treatment. With the successful demonstration of *ex vivo* DRS/IFS and MMS modalities, there are many clinical applications that can be addressed. These clinical studies will be accompanied by a validation of our diagnostic algorithm and the introduction of new pathologies into our data set. Some of the basic experiments will focus on quantitative analysis and spectral refinement. Experiments on determination sampling depth must also be performed in order to clarify the physiological meaning of the diagnostic parameters used in DRS/IFS and MMS algorithms.

7.1 Diagnostic Algorithm Validation and Extension

DRS/IFS, Raman and MMS diagnostic algorithms must be modified to identify DCIS and possibly other pathologies that are present in enough quantity in the data set. As stated previously, our MMS diagnostic algorithm uses the fit coefficients of collagen obtained with the Raman modality to separate normals from the rest of the diagnoses. As illustrated in Figure 6.2.a, the decision line is nearly horizontal, which indicates that only

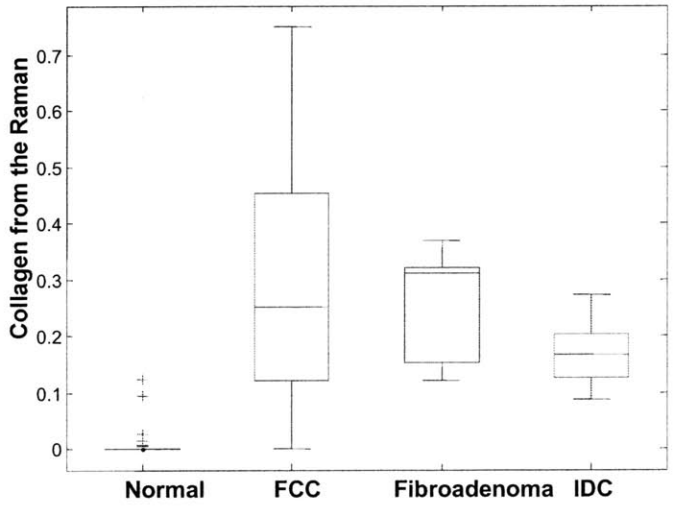

Figure 7.1 Collagen fit coefficient excited at 830 nm for different pathologies.

the collagen content from the Raman technique contributes to the diagnostic algorithm. Figure 7.1 is a boxplot, which illustrates the average values (red line), the interquartile range, which is the difference between the 75th and 25th percentile of the data, (blue box), and outliers (red plusses), of collagen content for each pathology. Given that a clear separation of normals from the other pathologies is possible based on this one value alone, the algorithm may be redeveloped in the future to possibly gain a diagnostic advantage. Previously, the collagen content from DRS/IFS was analyzed in this manner but did not show the same success. This is likely because of the different probing depths of the DRS/IFS and Raman modalities; DRS/IFS could not sample deep enough into the tissue to obtain a meaningful collagen content. Based on this finding, the diagnostic algorithm for Raman spectroscopy may be modified by rationing the fit coefficients for fat and collagen and plotting them versus a third parameter such as cell nucleus content. We hope this modification to the algorithm will allow us to correctly identify DCIS pathologies.

The results of the DRS/IFS and especially the MMS diagnostic models presented in this thesis are promising. However, because they were based on a pilot study and derived in a cross-validated manner, new clinical studies with a greater number of patients must be completed in order to validate the diagnostic algorithms. This clinical study will be undertaken after the construction of an MMS instrument that will have features of both the Raman and DRS/IFS modalities. Currently the FastEEM instrument for DRS/IFS employs 10 different excitation wavelengths, as discussed in Chapter 3. Only one wavelength was ultimately incorporated into our diagnostic algorithm, which means that we can greatly simplify our FastEEM instrument. However, because our laboratory does not focus exclusively on breast tissue (artery, oral and other organs are also being studied), we will develop an instrument with three excitation wavelengths - 308, 340, and 440 nm - to accommodate all organs that are currently under investigation.

7.2 Quantitative Study

Our Raman optical probe provides a reproducible geometry for data collection, allowing us to examine spectral data that has not been normalized. Working with unnormalized data, we can obtain quantitative information by incorporating the Raman scattering cross-sections of the model components into the data analysis routine.

In this study, data will be acquired from thin sections of fresh-frozen breast tissue using our Raman microscope. This data will be acquired using a low magnification objective to ensure that the spectroscopic signal is representative of the entire breast tissue section. Data will be collected for a long period of time to ensure high signal-to-noise ratio Raman data. After spectroscopic data is collected, each thin section of breast tissue will be examined by a pathologist. A combination of staining techniques and morpholometric analysis will be used to obtain an estimate of the volume occupied by each morphological feature represented in our Raman spectral model. These volumes will then be correlated with the fit coefficients of the corresponding spectra. In this manner, the contribution of each morphological component to the overall Raman signal level can be found. Additionally, we expect this study to help elucidate the lower limit of detection of this technique.

7.3 MMS Spectroscopy

Preliminary results examining Multi-modal spectroscopy of breast tissue are very promising. However, a more rigorous understanding of the relationship between a given data parameter and tissue morphology is necessary. Microscopy studies, similar to those used to develop the Raman spectral model, may provide a basis for understanding the fluorescence and reflectance signals obtained from tissue. This study will be performed in the pathology suite as breast tissue cannot be frozen for this experiment. Freezing the tissue will change the fluorescence characteristics of the tissue. Also, in order to successfully combine the optical modalities, a clear understanding of the probe volume and sampling depth is necessary. Monte Carlo simulations will be performed to simulate breast tissue and estimate sampling depth for breast tissue.

7.4 Transdermal Needle

Optical techniques are less invasive than current diagnostic procedures. Our initial results from *in vivo* and *ex vivo* Raman studies illustrate the potential for spectroscopic transdermal needle measurements to provide breast cancer diagnosis and treatment. Due to the fact that the Raman technique was shown to be applicable in a prospective study, we will first build a Raman transdermal needle. A DRS/IFS transdermal needle will be developed in the near future.

Transdermal needle measurements can be performed with either a front-viewing or a side-viewing optical fiber probe. A front-viewing probe can be used in conjunction with a vacuum-assisted biopsy procedure, discussed previously. Studies must be undertaken to investigate whether the vacuum device and overall setup can be used in conjunction with the probe.

A side-viewing probe can also be used for transdermal needle measurements. Our laboratory has developed a side-viewing probe prototype based on a design that replaces the sapphire ball lens with a half-ball lens backed by a mirror. The lens and mirror are configured at a 45° angle, providing collection capabilities similar to the front-viewing probe. The performance of the side-viewing probe is currently under investigation. This type of probe can be used in conjunction with either a spring-deployed or a vacuum-assisted biopsy needle. At the same time, a side-viewing probe will necessitate a method for registering the tissue examined spectroscopically, which will lie somewhere in the middle of the biopsy sample, with that examined by the pathologist. A side-viewing probe is ultimately preferable to a front-viewing probe in terms of spectroscopic advantage as it collects data from a larger tissue volume.

Chapter 8. Conclusions

The overall goal of this thesis was to develop a multi-modal spectroscopic algorithm that can be used to diagnose breast cancer. This process included the development of a diagnostic algorithm based on diffuse reflectance and intrinsic fluorescence, and the validation of a previously-developed algorithm for Raman spectroscopy. Each of these steps has been addressed and results have been presented throughout this thesis.

The DRS/IFS algorithm is developed based on results obtained in a clinical pilot study. This diagnostic algorithm incorporates four different pathologies and results in a high sensitivity and specificity of 100 percent and 95.8 percent, respectively, and an overall accuracy of 87.6 percent. MCR analysis proves that even without perfect initial guess spectra we can achieve good fits and reasonable values for concentration of the intrinsic fluorescence spectra's components. Investigation of different excitation wavelengths reveals that there is limited need for all ten wavelengths currently employed by the FastEEM instrument. In fact, the diagnostic algorithm is based solely on DRS parameters obtained from white light excitation and IFS parameters obtained from 340 nm excitation. In comparison, Palmer et al. [Palmer, 2003] also examined fluorescence and DRS: the sensitivity and specificity in discriminating malignant and nonmalignant tissues were 70 and 92 percent, respectively; at the same time sensitivity (30 percent) and specificity (78 percent) of DRS along was significantly lower; combination of two techniques did not improve the classification accuracy. This finding will help in the development of our next generation FastEEM instrument as well as for a new instrument that combines the Raman and DRS/IFS modalities.

The Raman spectroscopy algorithm is applied in a prospective manner and shown to be an effective diagnostic tool. The sensitivity and specificity of the validated algorithm is 88.9 percent and 90.6 percent, respectively, with an overall accuracy of 81 percent.

A multi-modal diagnostic technique is developed and preliminary results are very promising. It is clear that by combining Raman and DRS/IFS modalities we are able to

increase our sensitivity, specificity, and overall accuracy to 100 percent, 97.9 percent, and 92.4 percent, respectively. A new MMS instrument will be developed and tested in future clinical studies.

References

"Effects of radiotherapy and surgery in early breast cancer. An overview of the randomized trials. Early Breast Cancer Trialists' Collaborative Group", *N Engl J Med*, **333**(22): 1444-55 (1995)

Alfano R, Liu C, Sha W, Zhu H, Akins D, Cleary J, Prudente R and Cellmer E, "Human breast tissues studied by IR Fourier Transform Raman spectroscopy", *Lasers life Sci*, **4** 23-28 (1991)

Alfano RR, Tang G, Pradham A, Lam W, Choy DSJ and Opher E, "Fluorescence spectra from cancerous and normal human breast and lung tissues", *IEEE Journal of Quantum Electronics*, **23** 1806-1811 : (1987)

Angheloiu G, Georgakoudi I, Haka A, JT A, Mueller M, Motz J, Dasari R, Fitzmaurice M, Kramer J and Feld M, "Detection of coronary plaques with superficial foam cells using intrinsic fluorescence at 308 and 480 nm excitation wavelengths." *J. Am. Coll. Cardiol.*, 45-45A (2005)

Bigio IJ, Bown SG, Briggs G, Kelley C, Lakhani S, Pickard D, Ripley PM, Rose IG and Saunders C, "Diagnosis of breast cancer using elastic-scattering spectroscopy: preliminary clinical results", *J Biomed Opt*, **5**(2): 221-8 (2000)

Bloom H and Richardson W, "Histological grading and prognosis in breast cancer", *Br J Cancer*, **11** 359-377 (1957)

Cheng X, Mao J, Bush R, Kopans D, Moore R and Chorlton M, "Breast cancer detection by mapping hemoglobin concentration and oxygen saturation", *Appl Opt*, **42**(31): 6412-6421 (2003)

Cotran R, Kumar V and Collins T. *Robbins Pathologic Basis of Desease*, (1999)

Donegan W, Spratt J. *Cancer of the Breast*, W.B. Saunders Company (1995)

Elmore JG, Barton MB, Moceri VM, Polk S, Arena PJ and Fletcher SW, "Ten-year risk of false positive screening mammograms and clinical breast examinations", *N Engl J Med*, **338**(16): 1089-96 (1998)

Fantini S, Walker S, Franceschini M, Moesta K, Schlag P, Kaschke M and Gratton E, "Assessment of the size, position, and optical properties of breast tumors in vivo by non-invasive optical methods", *Appl. Opt.*, **37** 1982-1989 (1998)

Fornage BD, Sneige N, Ross MI, Mirza AN, Kuerer HM, Edeiken BS, Ames FC, Newman LA, Babiera GV and Singletary SE, "Small (< or = 2-cm) breast cancer treated with US-guided radiofrequency ablation: feasibility study", *Radiology*, **231**(1): 215-24 (2004)

Frank CJ, McCreery RL and Redd DC, "Raman spectroscopy of normal and diseased human breast tissues", *Anal Chem*, **67**(5): 777-83 (1995)

Frank CJ, Redd DC, Gansler TS and McCreery RL, "Characterization of human breast biopsy specimens with near-IR Raman spectroscopy", *Anal Chem*, **66**(3): 319-26 (1994)

Gage I, Schnitt S, Nixon A, Silver B, Recht A, Troyan S, Eberlein T, Love S, Gelman R, Harris J and Connolly JL, "Pathologic margin involvement and the risk of

recurrence in patients treated with breast-conserving therapy", *Cancer*, **78**(9): 1921-8 (1996)

Gupta PK, Majumder SK and Uppal A, "Breast cancer diagnosis using N2 laser excited autofluorescence spectroscopy", *Lasers Surg Med*, **21**(5): 417-22 (1997)

Haka A,"Development of in Vivo Raman Spectroscopy for the Diagnosis of Breast Cancer and Intra-Operative Margin Assessment", 2005, PhD, Cambridge, MIT

Haka A, Shafer-Peltier KE, M. F, Crowe J, R.R. D and Feld M, "Diagnosing Breast Cancer Using Raman Spectroscopy", *Proc. Natl. Acad. Sci USA*, submitted (2004)

Haka AS, Shafer-Peltier KE, Fitzmaurice M, Crowe J, R.R. D and Feld MS, "Identifying Microcalcifications in Benign and Malignant Breast Lesions by Probing Differences in their Chemical Composition Using Raman Spectroscopy", *Cancer Res*, **62** 5375-5380 (2002)

Houssami N, Cheung MN and Dixon JM, "Fibroadenoma of the breast", *Med J Aust*, **174**(4): 185-8 (2001)

Jemal A, Tiwari RC, Murray T, Ghafoor A, Samuels A, Ward E, Feuer EJ and Thun MJ, "Cancer statistics, 2004", *CA Cancer J Clin*, **54**(1): 8-29 (2004)

Johnson AT, Henry-Tillman R and Klimberg VS, "Breast conserving surgery: optimizing local control in the breast with the assessment of margins", *Breast Dis*, **12** 35-41 (2001)

Johnson JM, Dalton RR, Wester SM, Landercasper J and Lambert PJ, "Histological correlation of microcalcifications in breast biopsy specimens", *Arch Surg*, **134**(7): 712-5; discussion 715-6 (1999)

Khan S and Badve S, "Phyllodes Tumors of the Breast", *Curr Treat Options Oncol*, **2**(2): 139-147 (2001)

Long D. *The Raman Effect*, (2002)

Manoharan R, Shafer K, Perelman L, Wu J, Chen K, Deinum G, Fitzmaurice M, Myles J, Crowe J, Dasari RR and Feld MS, "Raman spectroscopy and fluorescence photon migration for breast cancer diagnosis and imaging", *Photochem Photobiol*, **67**(1): 15-22 (1998)

Motz JT, Hunter M, Galindo LH, Gardecki JA, Kramer JR, Dasari RR and Feld MS, "Optical fiber probe for biomedical Raman spectroscopy", *Appl Opt*, **43**(3): 542-54 (2004)

Nair MS, Ghosh N, Raju NS and Pradhan A, "Determination of optical parameters of human breast tissue from spatially resolved fluorescence: a diffusion theory model", *Appl Opt*, **41**(19): 4024-35 (2002)

Nath ME, Robinson TM, Tobon H, Chough DM and Sumkin JH, "Automated large-core needle biopsy of surgically removed breast lesions: comparison of samples obtained with 14-, 16-, and 18-gauge needles", *Radiology*, **197**(3): 739-42 (1995)

Ntziachristos V, Yodh AG, Schnall M and Chance B, "Concurrent MRI and diffuse optical tomography of breast after indocyanine green enhancement", *Proc Natl Acad Sci U S A*, **97**(6): 2767-72 (2000)

Page DL and Dupont WD, "Anatomic markers of human premalignancy and risk of breast cancer", *Cancer*, **66**(6): 1326-35 (1990)

Park CC, Mitsumori M, Nixon A, Recht A, Connolly J, Gelman R, Silver B, Hetelekidis S, Abner A, Harris JR and Schnitt SJ, "Outcome at 8 years after breast-conserving surgery and radiation therapy for invasive breast cancer: influence of margin

status and systemic therapy on local recurrence", *J Clin Oncol*, **18**(8): 1668-75 (2000)

Pask H, "The design and operation of solid-state raman lasers", *Progress in Quantum Electronics*, **27** 3-56 (2003)

Redd D, Feng Z, Yue K and Gansler T, "Raman spectroscopic characterization of human breast tissues: implications for breast cancer diagnosis", *Appl Spectrosc*, **47** 787-791 (1993)

Richards-Kortum R and Sevick-Muraca E, "Quantitative optical spectroscopy for tissue diagnosis", *Annu. Rev. Phys. Chem*, **47** 555–606 (1996)

Rosen P and Oberman H. *Atlas of Tumor Pathology. Tumors of the Mammary Gland*, (1992)

Rubin E and Farber J. *Pathology*, J. B. Lippincott company (1996)

Shafer K, Manoharan R, Dasari RR, Feld MS, Harrison GR, Myles J and Crowe J, "Modeling of NIR Raman Spectra of Human Breast Tissue for Breast Cancer Diagnosis", *Abstracts of Papers of the American Chemical Society*, **216** 344 (1998)

Shafer-Peltier K,"Chemical Basis for Breast Cancer Diagnosis Using Raman Spectroscopy", 2001, PhD, Cambridge, MIT

Shafer-Peltier K, Haka A, Fitzmaurice M, Crowe J, Dasari R and Feld M, "Raman microspectroscopic model of human breast tissue: implications for breast cancer diagnosis in vivo", *J. Raman Spectrosc*, **33** 552-563 (2002)

Shah N, Cerussi A, Eker C, Espinoza J, Butler J, Fishkin J, Hornung R and Tromberg B, "Noninvasive functional optical spectroscopy of human breast tissue", *Proc Natl Acad Sci U S A*, **98**(8): 4420-5 (2001)

Siddiqui MT, Saboorian MH and Ashfaq R, "Combined FNA/core needle biopsy", *Breast J*, **7**(3): 202-3 (2001)

Sternberg S. *Histopathology for Pathologists*, Lippincott-Raven. (1997)

Tunnell JW, Desjardins AE, Galindo L, Georgakoudi I, McGee SA, Mirkovic J, Mueller MG, Nazemi J, Nguyen FT, Wax A, Zhang Q, Dasari RR and Feld MS, "Instrumentation for multi-modal spectroscopic diagnosis of epithelial dysplasia", *Technol Cancer Res Treat*, **2**(6): 505-14 (2003)

Wu Y, Doi K, Giger ML and Nishikawa RM, "Computerized detection of clustered microcalcifications in digital mammograms: applications of artificial neural networks", *Med Phys*, **19**(3): 555-60 (1992)

Yang Y, Katz A, Celmer EJ, Zurawska-Szczepaniak M and Alfano RR, "Fundamental differences of excitation spectrum between malignant and benign breast tissues", *Photochem Photobiol*, **66**(4): 518-22 (1997)

Yang Y, Katz A, Celmer EZ, Zurawska-Szczepaniak M and Alfano RR, "Excitation spectrum of malignant and benign breast tissues: a potential optical biopsy approach", *Lasers life Sci*, **7**(2): 115-127 (1997)

Zonios G, Perelman L, Backman V, Manoharan R, Fitzmaurice M., Crowe J, Dasari RR and Feld M, "Diffuse Reflectance Spectroscopy of Human Adenomatous Colon Polyps In Vivo", *Appl. Opt.*, **38** 6628-6637 (1999)