

# Stochastic decomposition method for modeling the scattered signal reflected of mucosal tissues

Fernand S. Cohen

Ezgi Taslidere

Dilip S. Hari

Drexel University  
Electrical and Computer Engineering Department  
Philadelphia, Pennsylvania 19104

Sreekant Murthy

Drexel University  
College of Medicine  
Philadelphia, Pennsylvania 19104

**Abstract.** The aim of this work is to draw the attention of the biophotonics community to a stochastic decomposition method (SDM) to potentially model 2-D scans of light scattering data from epithelium mucosa tissue. The emphasis in this work is on the proposed model and its theoretical pinning and foundation. Unlike previous works that analyze scattering signal at one spot as a function of wavelength or angle, our method statistically analyzes 2-D scans of light scattering data over an area. This allows for the extraction of texture parameters that correlate with changes in tissue morphology, and physical characteristics such as changes in absorption and scattering characteristics secondary to disease, information that could not be revealed otherwise. The method is tested on simulations, phantom data, and on a limited preliminary *in-vitro* animal experiment to track mucosal tissue inflammation over time, using the area  $A_z$  under receiver operating characteristics (ROC) curve as a performance measure. Combination of all the features results in an  $A_z$  value up to 1 for the simulated data, and  $A_z > 0.927$  for the phantom data. For the tissue data, the best performances for differentiation between pairs of various levels of inflammation are 0.859, 0.983, and 0.999. © 2008 Society of Photo-Optical Instrumentation Engineers. [DOI: 10.1117/1.2982527]

Keywords: biomedical signal analysis; mucosa tissue; medical signal processing; feature extraction; optical signal processing; scattered light.

Paper 07240RR received Jul. 3, 2007; revised manuscript received May 30, 2008; accepted for publication Jun. 15, 2008; published online Oct. 3, 2008.

## 1 Introduction

About 90% of all cancers arise from epithelial cells that cover surfaces of the body and the lining of the internal organs.<sup>1</sup> Malignancies occurring in the epithelium lining the internal surfaces of organs (e.g., aerodigestive tract or colon) typically are discovered late in the course of the disease, usually because they are “hidden” from the physician, and are not usually found in a routine physical exam. Dysplasia is a precursor tissue change prior to epithelial cancer that can be reversible if it is detected at very early stages. Dysplasia is invisible to the eye, and can only be detected using biopsy.<sup>1</sup> Investigators are working on alternatives for biopsy to detect precancerous conditions.<sup>2,3</sup> Noninvasive and minimally invasive optical techniques are becoming staples of modern medical technology.<sup>4</sup> Dysplasia can occur less than 1 mm in size. For such small lesions, optical techniques are a good choice for detecting the structural changes in the tissue that cannot be easily detected visually.

A variety of optical techniques has been developed and include reflectance,<sup>5–9</sup> fluorescence,<sup>10–12</sup> light scattering spectroscopy (LSS),<sup>13–16</sup> time-resolved autofluorescence,<sup>12,17</sup> polarization imaging,<sup>18–20</sup> optical coherence tomography (OCT),<sup>21–23</sup> diffused optical tomography (DOT),<sup>24</sup> near-infrared imaging,<sup>25</sup> Raman scattering spectroscopy,<sup>26–28</sup> and

frequency domain spectroscopy.<sup>29</sup> Of these, reflectance and fluorescence methods are of particular interest for their application to early detection of cancer, particularly those of the aerodigestive tract. For example, in the work reported in Ref. 8, cell size and density are determined through backscattered light using spectroscopy. The scattered light is modeled as Mie scattering by surface cell nuclei. In contrast, in Ref. 30, precancerous detection is performed through fluorescence signatures. Since each of these optical modalities has its own points of advantage, combinations of such methods often produce specificity and sensitivities that are unattainable by individual techniques.<sup>7</sup>

Optical techniques do not necessitate tissue removal and analysis can be made in real time, hence offering the potential of detecting mucosal changes at the microstructural, biochemical, and molecular level. The detailed spatial intensity distribution of light scattered by an individual particle is a complex function of the particle’s size, shape, and orientation with respect to the wavelength as well as the incident illumination.<sup>31</sup> Among other important measures, scattered light provides an objective measure of epithelial nuclear enlargement and crowding, which are the most significant characteristics of dysplasia and early cancer.<sup>1</sup>

Our goal is to establish a textural model that models the reflected scattered signal and shows its sensitivity to differential changes in the morphology and physical characteristics of

Address all correspondence to Ezgi Taslidere, Electrical and Computer Engineering Dept., Drexel Univ., Philadelphia, PA 19104; E-mail: et43@drexel.edu

the tissue being optically scanned. Textural models have been successfully used by us to model radio frequency (rf) ultrasound images of breast tissue,<sup>32-34</sup> where we have proven the model to be quite effective for discrimination between benign, normal, and malignant breast tissues (the  $A_z$  values were in the range of 0.862 to 0.999). Similar to our work on rf modeling of breast tissue, we adopt a stochastic decomposition method (SDM) to model the scattered signal optically reflected from mucosal tissues to track down differential changes in the morphology and physical characteristics of the imaged (scanned) tissue. One of the original contributions of this work is the application of the SDM model to optical data, which is totally new to this application field. A second contribution is the systematic analysis and correlation between the signal parameters in predicting and tracking morphological and physical characteristics of epithelial tissue as it undergoes changes. Simulations and phantom data are used to test the validity of the model. The test scenarios are driven with biological applications (such as dysplasia) in mind. Differentiation based on nuclei sizes and densities (a trademark of nuclear crowding in dysplasia) were embedded in the choice of the simulation parameters of the light scattered from the cells mimicking dysplasia based on the Monte Carlo (stochastic) simulation of Mie scattering. Some of the parameters in our stochastic Mie scattering simulator relate directly to the tissue morphology, such as: number of cells, size of cells, and location of cells; while others relate to the physical staining characteristics of the tissue, which affect and are affected by the absorption and refraction indices. The model is also tested using phantoms with nuclei sizes and density scatterers (spheres) varying in accordance with biological epithelial tissue as it undergoes dysplastic changes. We used phantom data to see the performance of the technique to discriminate between two different morphologies reflecting tissue size of cell nuclei sizes mimicking normal and dysplastic epithelium. The simulations and phantom experiments show that some of the SDM parameters are extremely effective in picking up changes in tissue morphologies.

To further illustrate the feasibility of the model, we have considered a preliminary animal model from mouse colon *in vitro* with different levels of induced inflammation, where we test the ability of the SDM to differentiate between the various stages of inflammation as it develops in the mouse colon epithelium. This is of particular relevance, as there is a functional relationship between inflammation and cancer.<sup>35</sup> Balkwill and Mantovani<sup>36</sup> described that issue as: if genetic damage is the “match that lights the fire” of cancer, some types of inflammation may provide the “fuel that feeds the flames.”<sup>36</sup> For instance, ulcerative colitis (UC) associated colorectal cancer follows a histological sequence starting in the inflamed mucosa as a hyperplastic lesion develops through dysplasia into adenocarcinoma.<sup>37,38</sup> This is sometimes summarized as the “inflammation-dysplasia-carcinoma” sequence.<sup>39</sup>

## 2 Stochastic Decomposition Method and the Extracted Features

The decomposition of the LSS image (2-D scan) is consistent with the general decomposition of regular stochastic fields into predictable (the specular field) components and unpredictable (the diffused field) components, known in the litera-

ture of stochastic processes as the World decomposition, used to decompose the scattered or reflected signal into its two components.<sup>33,40</sup> The specular scattering component is modeled by periodic or quasiperiodic image field of point scatterers corresponding to the cell boundaries, whereas the diffuse component is modeled by an autoregressive field, which corresponds to a linear filter driven by white noise:

$$y(n) = c(n) + d(n), \quad (1)$$

where the diffuse component  $d(n)$  is modeled as a zero-mean autoregressive process of order  $p$  driven by a zero-mean white noise sequence  $w(n)$  (not necessarily Gaussian) with variance  $\sigma^2$ ,

$$d(n) = \sum_{s=1}^p a_s d(n-s) + w(n). \quad (2)$$

The resolvable scattering structure  $c(n)$  can be viewed as a summation of smeared modulated delta functions located at the resolvable scatterers' locations, and of random strength, the coherent component  $c(n)$  can be modeled by a superposition of Gaussian modulated sinusoids:

$$c(n) = \sum_{i=0}^{N_c} \frac{A_i}{\sigma_{ci} \sqrt{2\pi}} \exp\left\{-\frac{(n-m_i)^2}{2\sigma_{ci}^2}\right\} \cos(w_c n), \quad (3)$$

where  $w_c$  is related to the wavelength of the light,  $m_i$  indicates the location of the coherent scatterer,  $\sigma_{ci}$  is a parameter related to the bandwidth,  $A_i$  is related to the strength of the coherent scatterer, and  $N_c$  is the number of coherent scatterers in the window examined. The only known parameter is  $w_c$ .

The SDM first checks for the existence of coherent scatterers by testing the hypothesis of Rayleigh scattering, which can be achieved by using the nonparametric Kolmogorov-Smirnov (KS) test for color field.<sup>41</sup> The rejection of the Rayleigh scattering hypothesis indicates the existence of a coherent component. If a coherent component exists, the signal is decomposed into its coherent  $c(n)$  and diffuse  $d(n)$  components using the wavelet decomposition described in Ref. 32. The SDM consists of taking the wavelet transform of the scattered signal and thresholding its energy. The decomposition uses the continuous wavelet transform and was thoroughly described and tested on simulated ultrasound rf data.<sup>32,33</sup> After the decomposition, we fit the autoregressive (AR) model to the diffuse part and extract the parameters of the diffuse part as well as the coherent part. If the test of presence of a coherent component fails, the signal is totally diffused and is modeled as an AR process.

The parameters for the coherent component are as follows.

*Number of detected coherent scatterers  $N_c$  in a region of interest (ROI). The locations of large fluctuation of the scale-averaged wavelet power around its mean value are the locations of the coherent scatterers in the probed region under study.<sup>33,34</sup> The number of the fluctuations that exceed a threshold is the number of coherent scatterers in the tissue. The number of coherent scatterers is related to the number of resolvable scatterers in the tissue (single scatterers). This parameter is correlated with the size of the window. The larger*

the size of the window is, the more coherent scatterers might be detected.

*Mean energy of the coherent scatterers  $E$ .* This parameter reflects differences in energy due to changes in index of refraction that affect the absorption and reflected energy.

The parameters and features of the diffuse component are estimated from the model parameters and are given as follows.

*Residual error variance of the diffuse component  $\sigma^2$ .* The equivalent of the energy of the coherent scatterers for the diffuse component is the unconditional variance. The unconditional variance is related to the residual error variance (conditional variance of the diffuse component) through the coefficients of the autoregressive model (AR parameters). The values of the unconditional variance are reflected mainly on the residual error variance.<sup>33,34</sup> They are also weakly reflected on the AR parameters. It follows that the same argument made for the mean energy of the coherent scatterers applies for the residual error variance as well, because if the resolution increases drastically, the previously unresolved scatterers will be resolvable. In our previous research for breast cancer characterization using ultrasound, this parameter showed excellent discrimination ability between benign, normal, and malignant tissues.<sup>33</sup>

*Rayleigh scattering degree of the diffuse component  $D$ .* This parameter describes the discrepancy between the empirical distribution of the innovation process of the diffuse component and the Gaussian distribution by calculating the KS distance. The KS distance is the distance between two cumulative probability distribution functions; it is the distance used to test the hypothesis that given data follow a certain nominal distribution. The lower the value of  $D$ , the closer to Rayleigh scattering the diffuse component is.<sup>33,34</sup> In our previous research with ultrasound backscattering, this parameter was found to be tightly connected with the density of scatterers per resolution cell; large scatterers per resolution cell caused Rayleigh scattering and less scatterers caused deviation from that.<sup>33</sup> Based on our past experience, we anticipate this parameter to be able to differentiate between the various stages of dysplasia formation as it directly relates to formation of new dysplastic cells and the increase/decrease in the dysplastic cell density that directly reflects on that KS distance.

*Normalized correlation coefficient of the diffuse component  $\rho_N$*  This parameter reflects the correlations between neighboring diffuse scatterers in the tissue, and depends on the density of the diffuse scatterers in the tissue in relation to the sampling period. The normalized correlation coefficient is the determinant of the covariance matrix normalized to set the diagonal of the matrix equal to 1.<sup>33</sup> The determinant of the normalized covariance matrix is real, nonnegative, and equals 0 if the data are linearly dependent, and equals 1 if the data are independent.

The parameters can be analyzed either individually and/or in combination to construct a detailed map of a tissue region. When using parameters in combination, global decision fusion must be used.<sup>34</sup> The global decision fusion we use is explained in detail in the following paragraphs.

Let  $Y_i$  be the  $i$ 'th feature,  $H_0$  be the null hypothesis (normal tissue), and  $H_1$  be the alternative hypothesis (cancerous tissue). The local Neyman Pearson (NP) decision rule with the

highest detection rate among all other decision rules for any fixed false alarm rate is<sup>34</sup>

$$\frac{p(Y_i|H_1)^{H_1}}{p(Y_i|H_0)^{H_0}} \geq \zeta_i. \quad (4)$$

This reduces to

$$X_i = 1 \text{ iff } T_i > \zeta_i \quad \text{and} \quad X_i = 0 \text{ iff } T_i < \zeta_i, \quad (5)$$

where  $T_i$  is a sufficient statistic associated with the data  $Y_i$ , which could be the data itself  $Y_i$ , as for the case, for example, when  $Y_i$  is Gaussian with different means but the same variance under  $H_0$  and  $H_1$ ; or it could be some function of  $Y_i$ , for example<sup>34</sup>

$$T_i = \frac{(Y_i - \mu_0)^2}{\sigma_0^2} - \frac{(Y_i - \mu_1)^2}{\sigma_1^2}, \quad (6)$$

for the case when  $Y_i$  is Gaussian with different means and variances under  $H_0$  and  $H_1$ .  $\zeta_i$  is the threshold such that the false alarm rate is  $\alpha_i = p(T_i > \zeta_i | H_0)$  and the probability of detection is  $(1 - \beta_i) = p(T_i > \zeta_i | H_1)$ . A global NP decision rule or fusion would be<sup>34</sup>

$$\frac{p(\vec{Y}_H|H_1)^{H_1}}{p(\vec{Y}_H|H_0)^{H_0}} \geq \zeta, \quad (7)$$

where  $\vec{Y}_H$  is the feature vector consisting of the individual parameters ( $H$  set). If  $\vec{Y}_H$  is normally distributed with mean  $\vec{\mu}_0$  and covariance matrix  $\Sigma_0$  under  $H_0$ , and  $\vec{\mu}_1$  and  $\Sigma_1$  under  $H_1$ , then the global NP statistics reduce to<sup>34</sup>:

$$T_g = (\vec{Y} - \vec{\mu}_0)^T \Sigma_0^{-1} (\vec{Y} - \vec{\mu}_0) - (\vec{Y} - \vec{\mu}_1)^T \Sigma_1^{-1} (\vec{Y} - \vec{\mu}_1). \quad (8)$$

Since the NP has the highest probability of detection for a fixed false alarm rate among all other decision rules, it therefore follows that the power (detection rate) of the global likelihood ratio test  $(1 - \beta) \geq (1 - \beta_i)$ ,  $i \in H$ . Hence, the global receiver operating characteristics (ROC) curve, which is a plot of probability of detection versus probability of false alarm, bounds from above all the local ROC curves. Simply stated, this proves that the addition of further information (sum of features) improves the probability of detection over using the individual features alone, and that is true for any given false alarm rate.<sup>34</sup> This means that the fusion of the various SDM features will result in better detection than either alone.

We use the area under the ROC curve  $A_z$  as a metric, since it integrates out the dependence of the performance on which false alarm rate to use. A sample ROC curve and sample class distributions from which the ROC curve is determined are shown in Fig. 1.  $A_z$  values close to 0.5 indicate very weak discrimination power between two classes (total chance decision rule amounting to coin tossing), whereas values close to 1 indicate very strong discrimination power and yield very high sensitivity and specificity values.<sup>34</sup>

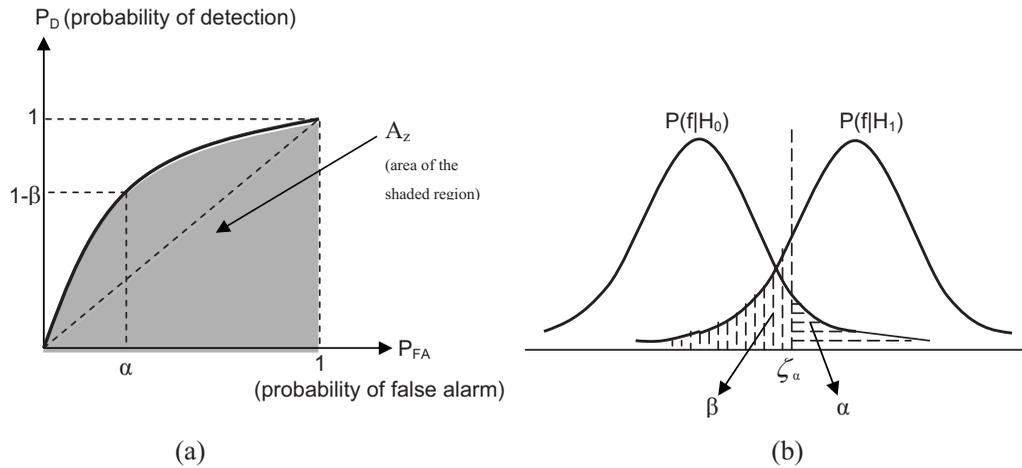


Fig. 1 (a) A sample ROC curve and (b) sample class distributions from which the ROC curve is determined.

### 3 Details of Experiments and Data Collection

In the experiments either based on simulations or on phantom data, we select the parameters and their values to mimic the interrogation of a biological epithelium tissue that is undergoing neoplastic changes, either dysplastic or of a malignant nature. Dysplasia is a term that refers to a precancerous condition. Malignant neoplastic changes mostly follow preexisting dysplastic changes. Removal of adverse environmental stimulus leads to restoration of normal cell growth patterns, which makes dysplasia reversible if detected in its very early stages. Dysplasia is recognized by cells altering their appearance (cytology). As tissue becomes dysplastic, the nuclei enlarge and become crowded.<sup>6</sup> Healthy tissue epithelial nuclei have a characteristic diameter of 4 to 7  $\mu\text{m}$  and are arranged in neat rows.<sup>8</sup> In dysplastic epithelium, the cells proliferate and the nuclei enlarge and appear darker when stained.<sup>1</sup> Nuclei can be as large as 20  $\mu\text{m}$  in diameter.<sup>8</sup> Normal intestinal cells are characterized by uniform nuclear size distribution where malignant cells have larger nuclei and more variation in nuclear size.<sup>18</sup> In Ref. 18, for a normal tissue sample, the average diameter was found to be 4.8  $\mu\text{m}$ , a standard deviation of the sizes was 0.4  $\mu\text{m}$ . For the cancerous tissue sample, the corresponding values were 9.75 and 1.5  $\mu\text{m}$ . The detection of dysplasia is a challenging problem, since it does not form macroscopically observable structures that can be picked during colonoscopy.

Finally, to further show the feasibility of the model, we have also considered a preliminary *in vitro* animal model of a mouse colon with different levels of induced inflammation. Here we test the ability of the SDM to differentiate between the various stages of inflammation as it develops in the mouse colon epithelium.

#### 3.1 Simulation of Scattered Light from Cluster of Spheres Subjected to Changes in Morphology and Physical Characteristics

The simulations are based on linear arrays of light scattering data at a given wavelength. The results on simulated overlapping and nonoverlapping 1-D scans (256 points) of data obeying Mie scattering have been obtained. The theoretical details for the simulations are given in the Appendix (Sec. 6). Each

intensity point [See Eq. (14)] on the 1-D scan results from a cluster of  $N$  multiscattering spheres located at given positions  $(x, y, z)$  within a cube extending from  $-20$  to  $+20$   $\mu\text{m}$  on all three axes. Unless otherwise stated, the term “resolution cell” refers to this volume. Since we do not deal with overlapping spheres, the sphere locations are selected to be sufficiently spaced. The spheres were of a given size and index of refraction (RI) Figure 2 shows an ensemble of sphere scattering clusters, with each cluster comprising  $N$  spheres, each with diameter  $d$ . The strength of the incident light  $I$  at a fixed wavelength  $\lambda$  (580 nm in the experiment) and the direction of the detector  $(\theta, \varphi)$  were control parameters in the simulator. The scattering matrix corresponding to a direction of (45 and 90 deg) is calculated. The output of the simulator is the scattered light received at the detector from the ensemble of the scattering spheres obtained using the Mueller scattering matrix [see Eq. (11)]. The scattered light was calculated at various positions of the resolution cell (centroids of the cube). The contour map is rectangular, with  $\theta$  and  $\varphi$  values representing the vertical and horizontal coordinates. The output of the program is the Mueller scattering matrix map written sequentially for each set of  $\theta$  and  $\varphi$  values. The simulator calculates the scattering matrix for  $nt+1$  values of  $\theta$  between 0 and  $\pi$ , and  $nt+1$  values of  $\varphi$  between 0 and  $2\pi$ . The output

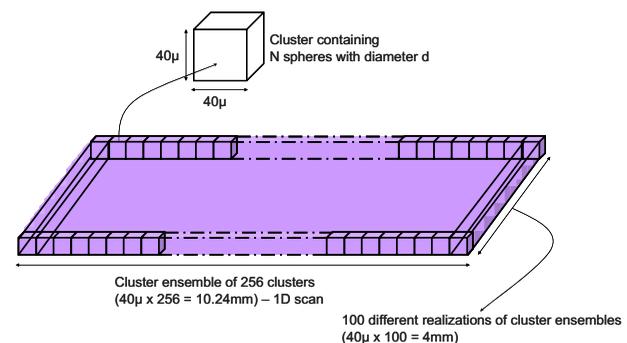


Fig. 2 Distribution of scattering spheres within 256 uniformly distributed cluster ensembles with each cluster having  $N$  spheres each with diameter  $d$ .

considered out of this  $nt+1$  by  $nt+1$  matrix always corresponds to 45 deg ( $\theta$  value) and 90 deg ( $\varphi$  value). The output of the simulation gives the elements of the Mueller scattering matrix of the scattered light, corresponding to  $\theta=45$  deg and  $\varphi=90$  deg [Eqs. (12) and (13)]. The intensity is calculated from the scattering matrix using Eq. (14).

To introduce randomness into Mie scattering,<sup>16</sup> we slightly and randomly perturb the spheres' position coordinates, and number and refractive index (RI) of the scattering spheres around nominal values defining a given tissue structure. The coordinates of the first cluster ensemble are specified and the coordinates for the remaining 256 clusters are calculated, generating from the first cluster by adding a random position component to each of the  $N$  spheres in the first cluster using a Gaussian random number generator. The position of each sphere in the first cluster using a Gaussian random number generator. The position of each sphere in the new cluster is computed as follows:

$$(x, y, z)_i = (x, y, z)_1 + \text{random coordinates} \\ \text{from random number generator.} \quad (9)$$

Random order permutations are obtained from the generated data of length 256. We obtain 100 independent realizations with 256 points of intensity values for each case considered. 100 realizations of 256 points each are deemed sufficient to obtain reliable parameter estimates and good statistical samples for testing and reliable classification. Each realization accounts for one 1-D scan. For each such realization, we compute the parameters of our decomposition. We then proceed to study the discrimination power of the SDM parameters for different scattering conditions corresponding to different tissue structures.

We have considered resolution cells and we picked parameters [cell sizes, number, and index of refraction (RI)] that are close in values to normal and crowding epithelial nuclei that are encountered in dysplasia. Two types of cells are simulated, normal cells and diseased cells. For the simulation of the *normal cells*, the cluster volume is assumed to contain 9 to 11 cells of  $5 \mu\text{m}$  diameter. For the simulation of the *diseased cells*, each cluster ensemble is considered to consist of four cells of  $12 \mu\text{m}$  diameter. The index of refraction (RI) of the tissue is generally in the range of 1.33 to 1.5.<sup>42,43</sup> RI takes a minimal value when the content of water in tissue is maximal and vice versa.<sup>42,44</sup> We used two different indices of refraction,  $RI_1$  and  $RI_2$  ( $RI_1=1.38+0.005i$  and  $RI_2=1.41+0.010i$ ), values that are consistent with many other studies in the literature.<sup>42,45-53</sup>

**Table 1** Parameter values for simulation sets 1 and 2.

		Cell Diameter	Number of Cells	Index of Refraction
Diseased cells	L1	$12 \mu\text{m}$	4	$1.38+0.005i$
	L2	$12 \mu\text{m}$	4	$1.41+0.010i$
Normal cells	S1	$5 \mu\text{m}$	9 to 11	$1.38+0.005i$
	S2	$5 \mu\text{m}$	9 to 11	$1.41+0.010i$

Three different sets of simulations are considered. In simulation sets 1 and 2, four types of data are created, L1, L2, S1, and S2. L1 stands for large spheres with cell diameter of  $12 \mu\text{m}$  with index of refraction  $1.38+0.005i$ , whereas L2 stands for the same configuration except that the index of refraction is increased to  $1.41+0.01i$ . S1 stands for small spheres with cell diameters of  $5 \mu\text{m}$  with index of refraction  $1.38+0.005i$ , whereas S2 stands for the same configuration except that the index of refraction is increased to  $1.41+0.01i$ . L1 and L2 represent diseased cells, while S1 and S2 represent normal cells. Descriptions for the four types of data are summarized in Table 1.

We repeat the experiment (simulation set 2), but now we vary the resolution so that some of the scatterers behave as coherent scatterers as well as diffuse. For simulation set 1, there is no overlap between consecutive samples. However, for this new case, we take samples that are measured every  $10 \mu\text{m}$ , thus producing an overlap between the 256 samples as opposed to the simulation set 1.

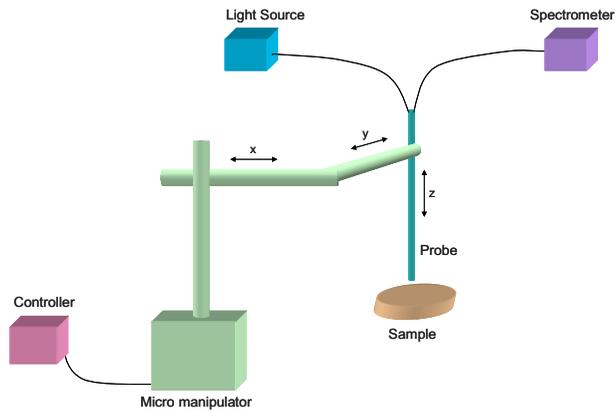
Finally, in simulation set 3 we consider the case of one single scatterer per resolution cell as opposed to many (multiscattering), and there is no overlap between consecutive samples. Three types of data are created, A1, A2, and A3. For A1 and A2, there is a single sphere with diameters 30 and  $10 \mu\text{m}$ , respectively, in the resolution cell. All 256 samples are randomized through small random uniform perturbations on the index of refraction around  $1.38+0.005i$ : positions of the spheres are fixed from cluster to cluster. A3 has ten spheres (multiscattering case) of  $5 \mu\text{m}$  diameter and index of refraction  $1.38+0.005i$ . The descriptions of these three types of data are given in Table 2.

### 3.2 Details of Phantom Experiments

(Probe, spectrometer, and light source are provided by Drexel Photonics Laboratory members Vitol, Kurzweg and Nabet).

**Table 2** Parameter values for simulation set 3.

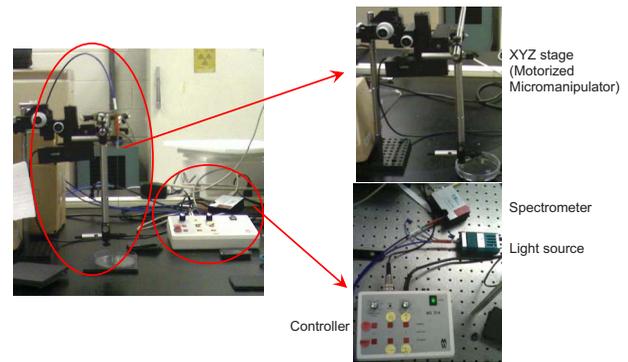
		Cell diameter	Number of cells	Index of refraction
Single sphere scattering	A1	$30 \mu\text{m}$	1	Perturbed around $1.38+0.005i$
	A2	$10 \mu\text{m}$	1	Perturbed around $1.38+0.005i$
MultiSphere scattering	A3	$5 \mu\text{m}$	10	$1.38+0.005i$



**Fig. 3** 3-D schematic figure of the experimental setup used for data collection. The probe is held and moved by the motorized micromanipulator (xyz stage) allowing 2-D scanning of the sample. It is connected to a light source and spectrometer.

Phantom and *in-vitro* animal data are collected using a probe, spectrometer, and light source.<sup>54</sup> The sample is illuminated with white light, and the reflected light spectrum is collected at equal intervals during scanning using a staging station. Intensities collected at one specific wavelength at each data point from an image (2-D scan) of the sample. Various 2-D scans are formed by the intensities at different specific wavelengths. Our method images the tissue area under investigation, and extracts from the data features (SDM parameters) that change with changes in the tissue morphology. A distinctive feature of our technique compared to DOT, OCT, and LSS is the formation of an image (2-D scan), rather than a spot, from which pertinent data are extracted. In this fashion, obtaining more data cannot only lead to a more confident calculation of the relevant features, such as nuclear size distribution, but can also lead to additional information embedded in the spatial texture that our decomposition technique arrives at by modeling the hidden correlations that are obtained only by interrogating a wide sample area.

In our first set of experiments, we use polystyrene latex microspheres (Polysciences Incorporated, Washington, Pennsylvania) suspended in deionized water.<sup>54</sup> We used phantom data to see the performance of the technique to discriminate between two different morphologies reflecting tissue cell nuclei sizes mimicking normal and dysplastic epithelium. The microspheres exist in a 2.5% aqueous suspension with a refractive index of 1.59 to 1.60 at 589 nm. We examined microspheres at diameter sizes of 3 and 10  $\mu\text{m}$ . A schematic diagram for the experimental system setup is shown in Fig. 3, and a picture of the actual system (probe, spectrometer, and xyz stage) is shown in Fig. 4. We use a bifurcated reflectance probe (Ocean Optics, Dunedin, Florida) consisting of one central fiber used for light delivery and six surrounding fibers used for light collection, each with a core of 200  $\mu\text{m}$ .<sup>54</sup> The probe has a special 30 deg window at the end to reduce specular reflectance from the sample surface. The probe is connected to the light source (tungsten halogen lamp; LS-1, Ocean Optics, Dunedin, Florida) and the spectrometer (HR4000, Ocean Optics, Dunedin, Florida) having an output of 3648 points with a wavelength step size of 0.12 nm. To create the image (2-D scan), the fiber is placed on an xyz



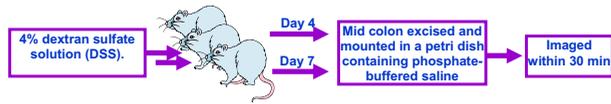
**Fig. 4** Real pictures of the experimental setup used in data collection. This corresponds to the schematic figure in Fig. 3.

stage, which allows the fiber to be moved and data to be collected at different points on the sample. At each data point, the whole spectrum is recorded. After the spectrum is collected at all the data points, the intensity image (2-D scan) is reconstructed by examining one single wavelength from which key parameters are extracted for each performance evaluation.

We use 3- $\mu\text{m}$  spheres to mimic the normal cells, and 10- $\mu\text{m}$  spheres to mimic the dysplastic cells. The surface area over which the data is collected is  $4500 \times 4500 \mu\text{m}$ . Three 1-D scans are taken for each phantom. The scans are taken at 0, 2250, and 3500  $\mu\text{m}$  (the y direction). For each 1-D scan, 51 points are taken 90  $\mu\text{m}$  apart in the x direction. At each point the whole spectrum (at various wavelengths) is computed. Smoothing is used as a preprocessing step for each point. After the spectrum is smoothed, the intensity values for specific wavelengths are taken. As the phantom was limited in size, order permutations are carried out to generate a set of 32 independent 1-D scans from each of the three 1-D scans, resulting into a total of 99 1-D scans used for classification.

### 3.3 Details of Tissue Experiments

The very preliminary animal study is based on the use of mid-colon obtained from three mice fed with 4% dextrin sulfate solution (DSS). The DSS model has been used to study colorectal cancer (CRC). The DSS colitis model can be used in mouse as a model for studying the sequence of chronic inflammation dysplasia in human inflammatory bowel disease.<sup>55</sup> For normal ulcerative colitis, mice are fed with DSS for one cycle; for chronic inflammation one cycle of DSS is followed by two cycles of normal water. For dysplasia, mice are kept on DSS for one cycle, two cycles of water, again DSS, then followed by water. In our model, what we were measuring was uncontrolled acute inflammation. Prior histological studies have shown that during the first seven days, of DSS feeding, these animals show gradual crypt dropout and acute to mixed inflammation starting on day 3 of DSS feeding.<sup>56</sup> Colon was excised on days 4 and 7 of DSS feeding. Clinically, after four days, the disease activity index was 1.3, which increased to 2.7 on day 7. In our experiments, the inflammation certainly increased from day to day while on DSS feed. We named the samples as mouse1, mouse2, and mouse3. Mouse1 represents normal mouse colon *in vitro*, mouse2 represents increased inflammation with the presence of increased



**Fig. 5** Animal study. Colon was excised on days 4 and 7 of 4% dextran sulfate solution (DSS) feeding, causing increased inflammation with feeding. The excised colon was mounted in a petri dish containing phosphate buffered saline and was imaged within thirty minutes after mounting.

neutrophils (polymorph nucleus) on day 4, and mouse3 represents further increase in inflammation with the possible presence of mucosal erosions on day 7. The excised colon was mounted in a petri dish containing phosphate buffered saline. The colon was imaged within thirty minutes after mounting. It takes about 7 to 8 min to complete the scan. A schematic of the animal study is shown in Fig. 5.

Three 1-D scans of the entire piece of the colon ( $10,000 \times 3000 \mu\text{m}$ ) were taken in a zigzag fashion, using a micrometer stage within 30 min after tissue mounting. The data collection process followed a similar protocol as with the phantom data collection. A 3-D figure of the experimental setup for *in-vitro* data collection is shown in Fig. 3. Three scans are taken as in phantom data. The scans are taken at 500, 1500, and 2500  $\mu\text{m}$ . Each scan consists of 50 points. The step size is taken as 200  $\mu\text{m}$ . After the data are collected, the same processing steps as with the phantom data are applied.

## 4 Results and Discussion

The metric that we used for representing the results is the area  $A_z$  under the ROC curve (plot of probability of detection versus probability of false alarm). As explained in Sec. 2, we use the  $A_z$  value under the ROC curve to determine the performance of our classification system.

### 4.1 Performance Evaluation on Simulations

Using the SDM, we calculated five features  $N_c$ ,  $D$ ,  $\sigma^2$ ,  $\rho_N$ , and  $E$  for each realization. The classification performance of each feature is evaluated separately. For simulation set 1, there is no overlap between consecutive samples. At the resolution used, the test for the existence of coherent scatterers was rejected as expected, since for this case only diffuse scatterers were present. The performances ( $A_z$  values) of individual diffuse component features for the four cases for simulation set 1 are given in Table 3. Each value in the table presents the area under the ROC curve for classification between pairs of data. Using a combination of features, the global detector yielded a perfect performance with overall  $A_z$  values greater than 0.991. The combination of performance parameters is in accord with the joint likelihood ratio statistics (NP statistics), which insures the highest detection rate for a fixed alarm rate over any other classifier including the NP based on the individual parameters in isolation (see Sec. 2).

It is interesting to note here that the KS distance parameter  $D$ , which directly relates to the number of scatterers (spheres), was able to differentiate between structures with different numbers and sizes of scattering spheres but failed (as expected) for the case with same density (L1-L2 and S1-S2) but

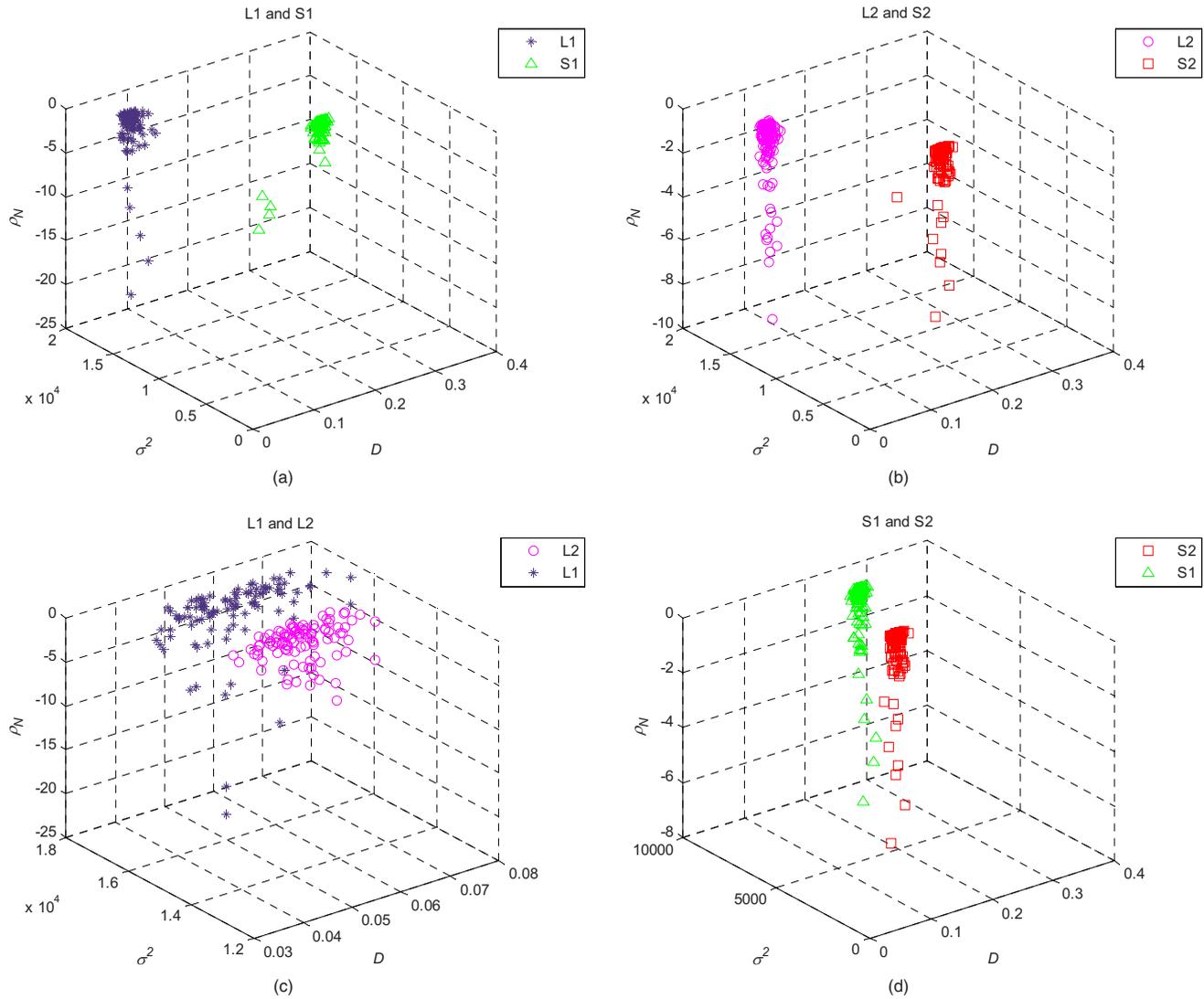
**Table 3** Performance of the system for the extracted features in terms of  $A_z$  (area under the ROC curve) values for differentiation between the pairs in simulation set 1.

	L1-S1	L2-S2	L1-L2	S1-S2
$D$	0.971	0.996	0.672	0.790
$\sigma^2$	1.000	1.000	0.962	1.000
$\rho_N$	0.586	0.583	0.537	0.505
All	1.000	1.000	0.991	1.000

with different indexes of refraction. On the other hand, since the energy of the scattered light (and hence the residual error variance) depends both on the number of scatterers and the index of refraction, as expected, it performed well in all four cases. Finally, as expected, the normalized correlation coefficient was not able to discriminate between the various cases, as the 256 points resulted from nonoverlapping spheres as we moved along the simulated structure. The 3-D plots of the extracted features are given in Fig. 6 to show the good separation between different pairs of classes in simulation set 1.

For simulation set 2, the performances ( $A_z$  values) of individual diffuse component features for the four cases are given in Table 4. Using a combination of features, the global detector yields a perfect performance with overall  $A_z$  values greater than 0.999 for the cases. Unlike the previous case, since there is an overlap between consecutive samples, a coherent component is detected and we have coherent as well as diffuse parameters. In addition, the normalized correlation coefficient becomes a discriminating parameter for cases L1-S1 and L2-S2, where the data between pairs of classes correspond to different numbers of scatterers and size.

To discover the differentiation performance of the method for single and multisphere scattering cases, simulation set 3 is used, and the following results are obtained. The performances ( $A_z$  values) of the individual features for the pairs A1-A2, A1-A3, and A2-A3 are given in Table 5. Using a combination of features, the global detector yields a perfect performance with an overall  $A_z$  value of 1 for the cases. For the case of single scattering, as expected, the number of coherent scatterers as well as the KS distance cannot discriminate between cases A1 and A2, since they both have one scatterer per resolution cell. However, they do successfully differentiate when the number of scattering spheres is different (cases A1-A3 and A2-A3). As expected, the parameters (the energy of the coherent and diffuse scatterers) successfully discriminate between all cases, since they depend on both the index of refraction and the number of scatterers per resolution cell. Finally, as anticipated, the normalized correlation coefficient is not able to discriminate between the various cases due to data permutations that destroyed any discrimination based on correlation. The 3-D plots of the extracted features are given in Fig. 7 to show how good the separation was between different pairs of classes in simulation set 3.



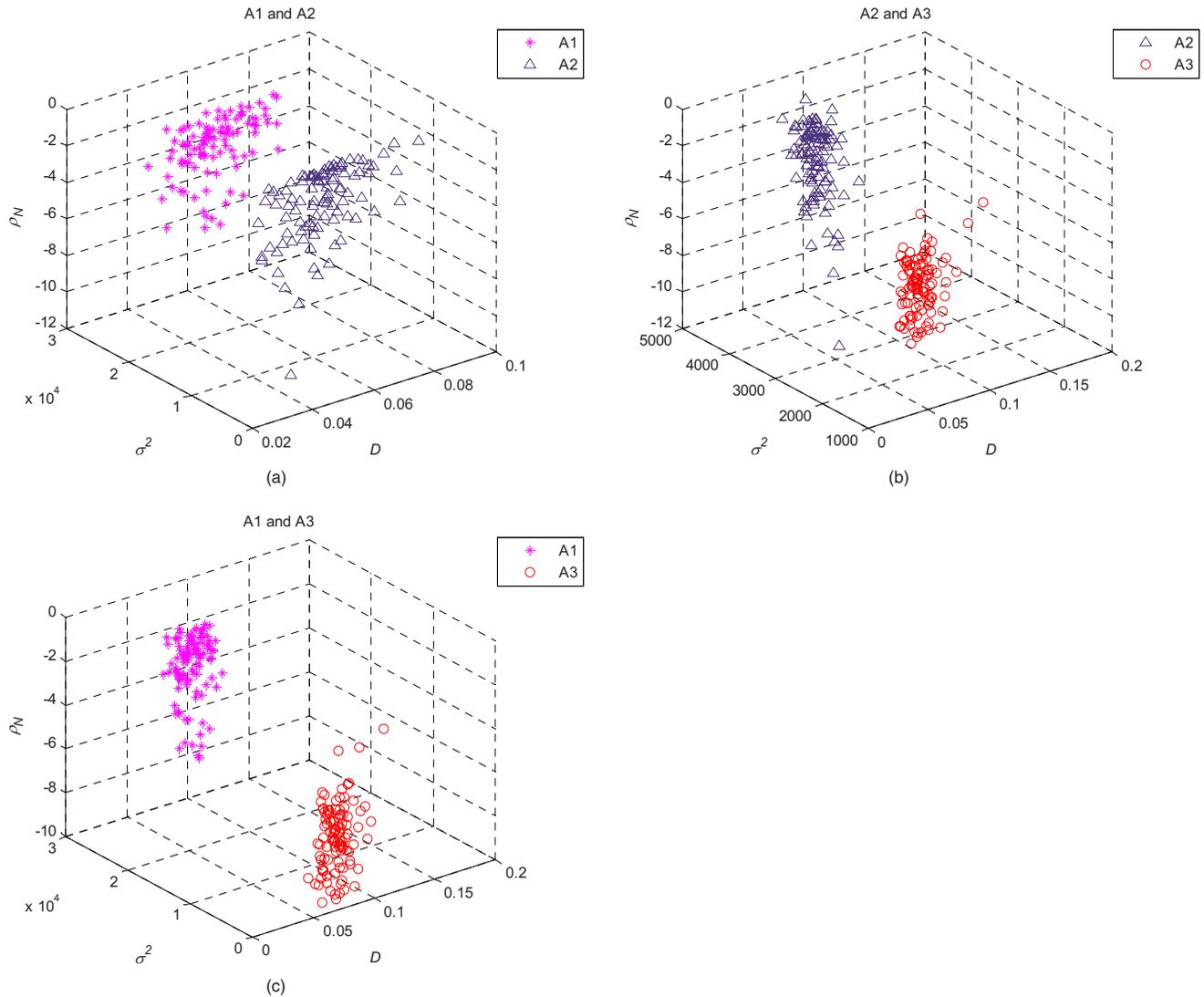
**Fig. 6** 3-D plots of extracted features in 3-D feature space. The features are:  $D$  (Rayleigh scattering degree of the diffuse component),  $\sigma^2$  (residual error variance of the diffuse component) and  $\rho_N$  (normalized correlation coefficient of the diffuse component). These are shown for the simulation pairs: (a) L1-S1, (b) L2-S2, (c) L1-L2, and (d) S1-S2. (refer to Table 1). The good separation between different simulated sets can be observed from the given plots.

**Table 4** Performance of the system for the extracted features in terms of  $A_z$  values for differentiation between the pairs in simulation set 2.

	L1-S1	L2-S2	L1-L2	S1-S2
$N_c$	0.993	1.000	0.639	0.507
$E$	0.999	1.000	0.889	0.965
$D$	0.987	0.968	0.568	0.546
$\sigma^2$	1.000	1.000	0.948	0.983
$\rho_N$	0.890	0.807	0.679	0.607
All	1.000	1.000	0.999	1.000

**Table 5** Performance of the system for the extracted features in terms of  $A_z$  values for differentiation between the pairs in simulation set 3.

	A1-A3	A2-A3	A1-A2
$N_c$	0.998	1.000	0.683
$E$	0.989	1.000	1.000
$D$	0.990	0.975	0.592
$\sigma^2$	1.000	1.000	1.000
$\rho_N$	0.695	0.676	0.578
All	1.000	1.000	1.000



**Fig. 7** 3-D plots of extracted features in 3-D feature space. The features are:  $D$  (Rayleigh scattering degree of the diffuse component)  $\sigma^2$  (residual error variance of the diffuse component), and  $\rho_N$  (normalized correlation coefficient of the diffuse component). These are shown for the simulation pairs: (a) A1-A2, (b) A2-A3, and (c) A1-A3 (refer to Table 2). The good separation between different simulated sets can be observed from the plots.

#### 4.2 Performance Evaluation on Phantom

The performance of the system is calculated for each specific wavelength. For each 1-D scan at a given wavelength, the

**Table 6** Performance of the system for the extracted features in terms of  $A_2$  values for diffuse only model for phantom data at different wavelengths.

	500.04 nm	550.05 nm	571.67 nm	650.00 nm
$D$	0.694	0.551	0.580	0.519
$\sigma^2$	0.981	0.979	0.982	0.924
$\rho_N$	0.500	0.576	0.561	0.504
All	0.990	0.982	0.989	0.927

decomposition parameters were computed, hence a set of 99 imaging parameter datasets for the 3- $\mu\text{m}$  spheres and 99 for the 10- $\mu\text{m}$  spheres. The classification is reported to be the best for wavelengths between 500 to 600 nm. The classification performance between the normal versus the dysplastic mimicking cells is evaluated, as in the simulation case, using the area under the ROC curve as the metric for performance. This is shown in Tables 6 and 7 for a diffuse scatterer only model versus a diffuse plus coherent scatterer model. The last row in both tables shows, the results of classification by fusing all the imaging parameter sets together. The performance varied from 0.927 to 0.994, which is consistent with the simulation results, and is a very strong result in discriminating between normal-mimicking cells from dysplastic-mimicking ones. As with the simulation results, the best feature is still the residual error variance. The overall performance is above 0.9 for both the simulated and phantom data. It is shown that the

**Table 7** Performance of the system for the extracted features in terms of  $A_z$  values for coherent+diffuse model for phantom data at different wavelengths.

	500.04 nm	550.05 nm	571.67 nm	650.00 nm
$N_c$	0.551	0.587	0.517	0.575
$E$	0.986	0.967	0.979	0.865
$D$	0.733	0.530	0.718	0.616
$\sigma^2$	0.984	0.979	0.982	0.924
$\rho_N$	0.572	0.538	0.508	0.510
All	0.994	0.988	0.991	0.931

performances of the individual parameters remain consistent with the simulation results, and the overall performance of the system is still extremely high ( $A_z > 0.927$ ).

### 4.3 Performance Evaluation on Tissue

The classification is reported for wavelengths between 450 to 650 nm in Tables 8–10. As with the phantom results, we observe that the mean energy of the coherent and diffuse scatterer increased as the animals became sicker. The performance is shown in the Tables 8 and 9 in terms of areas under the ROC curves ( $A_z$  values) using again the permutation method used with the phantom data to increase the sample size (i.e., the number of A scans). The best performance using one wavelength (625.24 nm) was at 0.859 combined  $A_z$  value (i.e., using all features) for differentiating between normal tissue and 4-day inflammation, whereas it was at 0.999 combined  $A_z$  value for differentiating between normal tissue versus day-7 inflammation at the same wavelength of 625.24 nm. We have also run the classification algorithm to see if we can differentiate between the 4-day inflammations versus the 7-day ones. The results are shown in Table 10. The best performance using one wavelength was at 0.987 combined  $A_z$  value for differentiating between 4-day inflammation tissue versus 7-day inflammation tissue at a wavelength of 575.02 nm. The best performance of the algorithm was

**Table 8** Performance of the system for the extracted features in terms of  $A_z$  values for tissue data (mouse colon *in vitro*) for classification between case Mouse 1 versus Mouse 2.

	500.33 nm	550.02 nm	575.02 nm	625.24 nm
$N_c$	0.568	0.559	0.590	0.559
$E$	0.811	0.741	0.743	0.843
$D$	0.515	0.505	0.563	0.565
$\sigma^2$	0.809	0.703	0.718	0.821
$\rho_N$	0.510	0.561	0.506	0.513
All	0.812	0.770	0.798	0.859

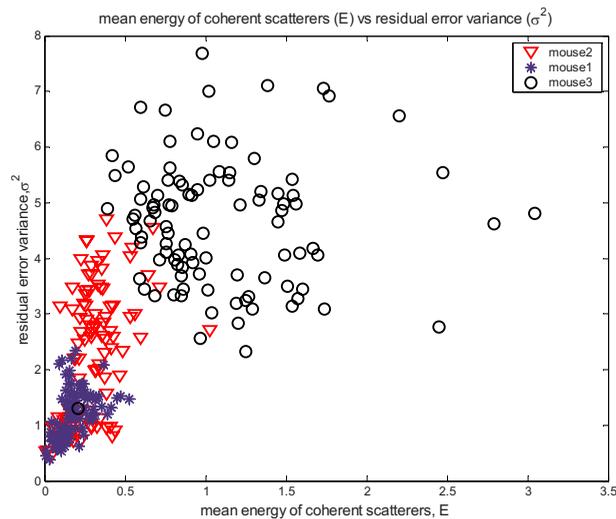
**Table 9** Performance of the system for the extracted features in terms of  $A_z$  values for tissue data (mouse colon *in vitro*) for classification between case Mouse 1 versus Mouse 3.

	500.33 nm	550.02 nm	575.02 nm	625.24 nm
$N_c$	0.541	0.511	0.549	0.508
$E$	0.997	0.997	0.994	0.999
$D$	0.551	0.581	0.538	0.584
$\sigma^2$	0.997	0.998	0.995	0.993
$\rho_N$	0.555	0.560	0.516	0.589
All	0.998	0.998	0.996	0.999

obtained at the wavelengths around the peak values of the spectrum. For the best two performing parameters  $E$  and  $\sigma^2$ , the wavelengths that show the best detection performances show consistency with each other for all the classification pairs. Also, they show consistency with the corresponding wavelengths for the overall performance (performance generated by fusing all the features). Although we have only considered three mice, we have looked at various area or points within each of these samples (three scans per sample at different locations with 50 points each). In Fig. 8, we show the scattered data for the best two features for the three mice at 575.02 nm. The ROC construction and the  $A_z$  values are based on the scattered data points. We can easily see from the scattered data how well separated the data are between the normal and 7-day inflammation, and how it is less so between the normal and the 4-day inflammation, leading to high  $A_z$  value under the ROC curve for the former case and lower for the latter. The separation between the 4-day and the 7-day inflammations is also quite pronounced, leading to high  $A_z$  value under the ROC curve. The scatterer plot shows the direct proportionality of the features to the disease formation. As the disease level increases, the values of the features' "mean energy of coherent scatterers" and "residual error variance" increase as well. This result is promising and consistent with the findings of the simulations and phantom data.

**Table 10** Performance of the system for the extracted features in terms of  $A_z$  values for tissue data (mouse colon *in vitro*) for classification between case Mouse 2 versus Mouse 3.

	500.33 nm	550.02 nm	575.02 nm	625.24 nm
$N_c$	0.541	0.596	0.596	0.571
$E$	0.938	0.961	0.974	0.906
$D$	0.556	0.578	0.509	0.622
$\sigma^2$	0.816	0.896	0.887	0.784
$\rho_N$	0.573	0.591	0.501	0.615
All	0.940	0.968	0.987	0.913



**Fig. 8** Scatter plot of mean energy of coherent scatterers versus residual error variance for the three animals studied (Mouse1, Mouse2, and Mouse3). The scatter plot shows good separation of extracted features between the three levels of inflammation. As the disease level increased, so did the values of these two features.

## 5 Conclusions

A textural model that models the reflected scattered signal is presented, and its sensitivity to simulated changes in the morphology and physical characteristics of mimicking tissue being optically scanned is presented as a proof-of-concept to model scattered signals reflected from mucosal tissues. At the core of that method is the modeling of the scattered light using a SDM texture model that captures how the tissue consisting of diffuse and specular scatterers reflects light shined at different wavelengths to yield textured data and corresponding image model parameters (obtained through the SDM) that correlate closely with these changes in the tissue morphology and physical characteristics (for example in dysplasia formation). In addition, the method statistically analyzes 2-D scans of light scattering data, hence allowing for the assessment of correlation and textural characteristics that otherwise could not be revealed when the analysis of the scattering signal is a function of wavelength or angle, which has been done by other investigators. This proof-of-concept is shown using simulation, phantom data, and on a limited preliminary *in vitro* animal experiment to track mucosal tissue inflammation over time. Although the scope of the animal experiment is limited, it tests the applicability of the model on real tissue. The test scenarios for the simulation and the phantom examples are driven by biological applications such as dysplasia formation and detection. The tissue experiment is geared more toward inflammation tracking, and the ability of the SDM to differentiate between the various stages of inflammation as it develops in the mouse colon epithelium. This is of particular relevance, as there is a functional relationship between inflammation and cancer. The performance is measured by computing the area  $A_z$  under the ROC curve for each parameter separately, and by fusing all the estimated parameter sets together. Combination of all the features results in an  $A_z$  value up to 1 ( $A_z > 0.991$ ) (perfect detection for any false alarm rate) for simulated data, and  $A_z > 0.927$  for the phan-

tom data. The performances for differentiation between pairs of various levels of inflammation are 0.859, 0.983, and 0.999. These encouraging results show that the proposed texture model and technique are excellent candidates to be applied to biological tissues, with the potential for enhanced endoscopy for detection of changes in epithelial tissue.

## Acknowledgments

We would like to thank Drexel Photonics Laboratory members Elina Vitol, Timothy Kurzweg, and Bahram Nabet for providing the optical device (probe, spectrometer, and light source) and the phantoms used in this research. Special thanks are also due to Elina Vitol for help with the phantom data collection. We also would like to thank Amol Karwa for his help during tissue experiments and for sharing useful biological information.

## Appendix: Mie Scattering and Signal Formation for Stochastic Simulator

The Mie scattering theory describes the light scattered by spherical particles.<sup>57</sup> Scattering of light in tissue mainly consists of two components: singly scattered and multiply scattered. Singly scattered light, as predicted by the Mie theory, is not randomized and contains information about individual scatterers. On the other hand, multiply scattered light is thoroughly randomized. Diffusely scattered light from tissue contains information about its basic structures. In the epithelial tissue, the light transport is dominated by multiple scattering.<sup>58–60</sup>

For multisphere scattering simulation, we used the T-matrix approach. The main complication here is how the presence of another scatter center in the neighborhood affects the radiation patterns, scattering matrix, and efficiencies. When two or more identical spheres aggregate into a cluster, the resulting composite particle is nonspherical.<sup>61</sup> The Mie theory, although widely used in the literature as a scattering approximation theory, requires modeling a cell as a homogeneous sphere.<sup>62</sup> Various techniques have been reported for generalizing the Mie theory<sup>62</sup> to deal with the application of it to a volume of spheres with various size distributions,<sup>63</sup> and for modeling a spherical cell containing a homogeneous spherical nucleus.<sup>64</sup> Some used anomalous diffraction approximations,<sup>65</sup> some multiple solutions,<sup>66</sup> some used 3-D finite-difference time-domain (FDTD) models of cellular scattering,<sup>45</sup> and finally some used T-matrix computations.<sup>67</sup> The problem of dependent scattering<sup>68</sup> in the presence of a cluster of spheres has been extensively investigated in the literature.<sup>69–73</sup> All of these reported techniques have important advantages over the Mie theory and helped in alleviating some of its limitations.<sup>62</sup> Recent work has verified the practical applicability of the T-matrix method to clusters of spheres.<sup>74</sup> The T-matrix method surpasses other frequently used techniques in terms of its ability to deal with a range of sphere sizes, as well as to systematically deal with nonspherical scattering based on the calculation of thousands of particles in random orientation.<sup>75</sup> It is for these reasons that we chose the T-matrix approach in our multisphere scattering simulation experiment.

The strategy of the T-matrix solution of the direct light scattering problem can be introduced with the following algo-

rithm: expanding the incident field in vector spherical wave functions regular at the origin; expanding the scattered field outside a circumscribing sphere of scatterers in vector spherical wave functions regular at infinity; and transforming the expansion coefficients of the incident field into those of the scattered field via a T-matrix.<sup>76</sup> The relationship between incident and scattered electric field components perpendicular and parallel to the scattering plane as observed in the farfield is described by the amplitude scattering matrix:

$$\begin{bmatrix} E_{\approx s} \\ E_{\perp s} \end{bmatrix} = \exp \frac{[i\mathbf{k}(r-z)]}{-i\mathbf{k}r} \begin{bmatrix} S_2 & S_3 \\ S_4 & S_1 \end{bmatrix} \begin{bmatrix} E_{\approx i} \\ E_{\perp i} \end{bmatrix}, \quad (10)$$

where  $k=2\pi n/\lambda$  is the wavenumber,  $\lambda$  is the wavelength of the incident light, and  $n$  is the scatterer refractive index. Each element of the scattering matrix  $S_j(j=1,2,3,4)$  depends on the scattering angle  $\theta$ . The elements of the amplitude scattering matrix are found according to the T-matrix theory and equations are explained in detail in Refs. 77–79. The intensity and polarizing properties of the field are described via a Stokes vector<sup>80</sup> that specifies the set of parameters associated with the phase and polarization of radiation. For a cluster of spheres, the relationship between incident and scattered Stokes parameters is described by the Mueller scattering matrix<sup>57,61</sup> as:

$$\begin{bmatrix} I_s \\ Q_s \\ U_s \\ V_s \end{bmatrix} = \frac{1}{k^2 r^2} \begin{bmatrix} S_{11} & S_{12} & S_{13} & S_{14} \\ S_{21} & S_{22} & S_{23} & S_{24} \\ S_{31} & S_{32} & S_{33} & S_{34} \\ S_{41} & S_{42} & S_{43} & S_{44} \end{bmatrix} \begin{bmatrix} I_i \\ Q_i \\ U_i \\ V_i \end{bmatrix}, \quad (11)$$

where  $I$  is the total intensity,  $Q$  is the degree of linear polarization, and  $U$  and  $V$  characterize the phase. The elements of the Mueller scattering matrix are described in terms of the amplitude scattering matrix elements (which are found from T-matrix equations<sup>77–79</sup>) such as:<sup>61</sup>

$$S_{11} = \frac{1}{2}[|S_1|^2 + |S_2|^2 + |S_3|^2 + |S_4|^2], \quad (12)$$

$$S_{12} = \frac{1}{2}[|S_2|^2 - |S_1|^2 + |S_4|^2 - |S_3|^2]. \quad (13)$$

Once the Mueller scattering matrix elements are found, the Stokes parameters can be calculated for any degree of polarization of the incident light. The incident Stokes vector for linear polarized light with angle  $\varphi$  relative to the lab system is  $[I_i \ I_i \cos 2\varphi \ I_i \sin 2\varphi \ 0]^T$ , with  $L_i$  as the intensity of incident beam. For example, if the incident light is polarized perpendicular to the scattering plane, the incident Stokes vector becomes  $[I_i - I_i \ 0 \ 0]^T$  into Eqs. (14) and (15):

$$I_s = [S_{11} - S_{12}]I_i, \quad (14)$$

$$Q_s = [S_{12} - S_{11}]I_i \quad U_s = V_s = 0. \quad (15)$$

## References

1. R. S. Cotran, S. L. Robbins, and V. Kumar, *Robbins Pathological Basis of Disease*, W. B. Saunders, Philadelphia, PA (1994).
2. G. Zonios, R. M. Cothren, J. M. Crawford, M. Fitzmaurice, R. Manoharan, J. Van Dam, and M. S. Feld, *Spectral Pathology*, Annals of the New York Academy of Sciences (1998).
3. J. A. Parrish and B. C. Wilson, "Current and future trends in laser medicine," *Photochem. Photobiol.* **53**, 731–738 (1991).
4. S. A. Boppart, T. F. Deutsch, and D. W. Rattner, "Optical imaging technology in minimally invasive surgery: current and future directions," *Surg. Endosc.* **13**, 718–722 (1999).
5. M. Canpolat and J. R. Mourant, "Particle size analysis of turbid media with a single optical fiber in contact with the medium to deliver and detect white light," *Appl. Opt.* **40**(22), 3792–3799 (Aug. 2001).
6. R. S. Dacosta, B. Wilson, and N. E. Marcon, "New optical technologies for earlier diagnosis of premalignant gastrointestinal lesions," *J. Gastroenterol. Hepatol.* **17**(s1), S85–S104 (Feb. 2002).
7. I. Georgakoudi, B. C. Jacobson, J. Van Dam, V. Backman, M. B. Wallace, M. G. Muller, Q. Zhang, K. Badizadegan, D. Sun, G. Thomas, L. T. Perelman, and M. S. Feld, "Fluorescence, reflectance, and light-scattering spectroscopy for evaluating dysplasia in patients with Barrett's esophagus," *Gastroenterology* **120**, 1620–1629 (2001).
8. L. T. Perelman, V. Backman, M. Wallace, G. Zonios, R. Manoharan, A. Nusrat, S. Shields, M. Seiler, C. Lima, T. Hamano, I. Itzkan, J. Van Dam, J. M. Crawford, and M. S. Feld, "Observation of periodic fine structure in reflectance from biological tissue: a new technique for measuring nuclear size distribution," *Phys. Rev. Lett.* **80**, 627–630 (1998).
9. G. Zonios, L. T. Perelman, V. Backman, R. Manoharan, M. Fitzmaurice, J. Van Dam, and M. S. Feld, "Diffuse reflectance spectroscopy of human adenomatous colon polyps in vivo," *Appl. Opt.* **38**(31), 6628–6637 (1999).
10. L. Coghlan, "Optimal fluorescence excitation wavelengths for detection of squamous intra-epithelial neoplasia: results from an animal model," *Opt. Express* **7**, 436–446 (2000).
11. M. Panjehpour, B. Overholt, T. Vo-Dinh, R. Haggitt, D. Edwards, and F. P. Buckley, "Endoscopic fluorescence detection of high-grade dysplasia in Barrett's esophagus," *Gastroenterology* **111**, 93–101 (1996).
12. B. W. Pogue and G. Burke, "Fiber-optic bundle design for quantitative fluorescence measurement from tissue," *Appl. Opt.* **37**(31), 7429–7436 (1998).
13. A. Amelink, M. P. L. Bard, S. A. Burgers, J. Henricus, and C. M. Sterenborg, "Single-scattering spectroscopy for the endoscopic analysis of particle size in superficial layers of turbid media," *Appl. Opt.* **42**(19), 4095–4101 (2003).
14. J. R. Mourant, J. Boyer, A. H. Hielscher, and I. J. Bigio, "Influence of the scattering phase function on light transport measurements in turbid media performed with small source-detector separations," *Opt. Lett.* **21**, 546–548 (1996).
15. J. R. Mourant, J. P. Freyer, A. H. Hielscher, A. A. Eick, D. Shen, and T. M. Johnson, "Mechanisms of light scattering from biological cells relevant to noninvasive optical-tissue diagnostics," *Appl. Opt.* **37**, 3586–3593 (1998).
16. M. B. Wallace, L. T. Perelman, V. Backman, J. M. Crawford, M. Fitzmaurice, M. Seiler, K. Badizadegan, S. J. Shields, I. Itzkan, R. R. Dasari, J. Van Dam, and M. S. Feld, "Endoscopic detection of dysplasia in patients with Barrett's esophagus using light-scattering spectroscopy," *Gastroenterology* **119**, 677–682 (2000).
17. T. J. Pfefer, K. T. Schomacker, M. N. Ediger, and N. S. Nishioka, "Multiple-fiber probe design for fluorescence spectroscopy in tissue," *Appl. Opt.* **41**(22), 4712–4721 (2002).
18. V. Backman, R. Gurjar, K. Badizadegan, I. Itzkan, R. Dasari, L. Perelman, and M. Feld, "Polarized light scattering spectroscopy for quantitative measurement of epithelial structures in situ," *IEEE J. Sel. Top. Quantum Electron.* **5**(4), 1019–1026 (1999).
19. A. H. Hielscher, J. R. Mourant, and I. J. Bigio, "Influence of particle size and concentration on the diffuse backscattering of polarized light from tissue phantoms and biological cell suspensions," *Appl. Opt.* **36**, 125–135 (1997).
20. J. R. Mourant, T. M. Johnson, S. Carpenter, A. Guerra, T. Aida, and J. P. Freyer, "Polarized angular dependent spectroscopy of epithelial cells and epithelial cell nuclei to determine the size scale of scattering structures," *J. Biomed. Opt.* **7**(3), 378–387 (2002).
21. D. Boas, D. H. Brooks, E. L. Miller, C. A. DiMarzio, M. Kilmer, R. Caudette, and Q. Zhang, "Imaging the body with diffuse optical tomography," *IEEE Signal Process. Mag.* pp. 58–75 (Nov. 2001).
22. J. A. Izatt, M. D. Kulkarni, H. W. Wang, K. Kobayashi, Jr., and M. V. Sivak, "Optical coherence tomography and microscopy in

- gastrointestinal tissues," *IEEE J. Sel. Top. Quantum Electron.* **2**(4), 1017–1028 (Dec. 1996).
23. Y. Pan, H. Xie, and G. K. Fedder, "Endoscopic optical coherence tomography based on a microelectromechanical mirror," *Opt. Lett.* **26**(24), 1966–1968 (Dec. 2001).
  24. V. Ntziachristos, A. H. Hielscher, A. G. Yodh, and B. Chance, "Diffuse optical tomography of highly heterogeneous media," *IEEE Trans. Med. Imaging* **20**(6), 470–478 (June 2001).
  25. M. A. Franceschini, S. Fantini, R. Palumbo, L. Pasqualini, G. Vaudo, E. Franceschini, E. Gratton, B. Palumbo, S. Innocente, and E. Manarino, "Quantitative near-infrared spectroscopy on patients with peripheral vascular disease," *Proc. SPIE* **3194**, 112–115 (1997).
  26. M. G. Shim, L. M. W. K. Song, N. E. Marcon, S. Hassaram, and B. C. Wilson, "Assessment of ex-vivo and in-vivo near-infrared Raman spectroscopy for the classification of dysplasia within Barrett's esophagus," *Proc. SPIE* **3918**, 114 (2000).
  27. U. Utzinger, A. Mahadevan-Jansen, D. Hinzelman, M. Follen, and R. Richards-Kortum, "Near infrared Raman spectroscopy for in vivo detection of cervical precancers," *Appl. Spectrosc.* **55**(8), 955–959 (2001).
  28. T. C. Bakker Schut, M. J. Witjes, H. J. Sterenborg, O. C. Speelman, J. L. Roodenburg, E. T. Marple, H. A. Bruining, and G. J. Puppels, "In vivo detection of dysplastic tissue by Raman spectroscopy," *Anal. Chem.* **72**, 6010–6018 (2000).
  29. M. S. Patterson, J. D. Moulton, B. C. Wilson, K. W. Berndt, and J. R. Lakowicz, "Frequency-domain reflectance for the determination of the scattering and absorption properties of tissue," *Appl. Opt.* **30**, 4474–4476 (1991).
  30. R. Richards-Kortum and M. Follen, "Clinical trial for new cervical cancer detection laser," \$8.9 NIH funding, NIH, Bethesda, MD.
  31. P. H. Kaye, "Spatial light-scattering analysis as a means of characterizing and classifying non-spherical particles," *Meas. Sci. Technol.* **9**, 141–149 (1998).
  32. G. Georgiou and F. S. Cohen, "Tissue characterization using the continuous wavelet transform part I: decomposition method," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **48**(2), 355–363 (Mar. 2001).
  33. G. Georgiou, F. S. Cohen, C. W. Piccoli, F. Forsberg, and B. B. Goldberg, "Tissue characterization using the continuous wavelet transform, part II: application on breast RF data," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **48**(2), 364–373 (Mar. 2001).
  34. E. Taslidere, F. S. Cohen, and G. Georgiou, "Classification of simulated hyperplastic stages in the breast ducts based on ultrasound RF echo," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **55**(1), 50–63 (2008).
  35. R. Virchow, *Die Krankhaften Geschwulste*, August Hirschwald, Ed., Berlin (1863).
  36. F. Balkwill and A. Mantovani, "Inflammation and cancer: back to Virchow?" *Lancet* **357**, 539–545 (2001).
  37. S. H. Itzkowitz and X. Yio, "Inflammation and cancer IV. Colorectal cancer in inflammatory bowel disease: the role of inflammation," *Am. J. Physiol. Gastrointest. Liver Physiol.* **287**(1), G7–17 (Jul. 2004).
  38. D. N. Seril, J. Liao, G. Y. Yang, and C. S. Yang, "Oxidative stress and ulcerative colitis-associated carcinogenesis: studies in humans and animal models," *Carcinogenesis* **24**, 353–362 (2003).
  39. H. Clevers, "At the crossroads of inflammation and cancer," *Cell* **118**, 671–674 (2004).
  40. F. S. Cohen, G. Georgiou, and E. Halpern, "WOLD decomposition of the backscatter echo in ultrasound images of soft tissue organs," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **44**(2), 460–472 (Mar. 1997).
  41. G. Georgiou and F. S. Cohen, "Statistical characterization of diffuse scattering in ultrasound images," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **45**, 57–64 (1998).
  42. S. A. Prahl, "Light transport in tissue," Ph.D. Thesis, Univ. of Texas at Austin (1988).
  43. J. H. Ali, W. B. Wang, M. E. Zavallos, and R. R. Alfano, "Near infrared spectroscopy and imaging to probe differences in water content in normal and cancer human prostate tissues," *Technol. Cancer Res. Treat.* **3**(5), 491–497 (2004).
  44. Bausch and Lomb, Analytic Systems Division, "Refractive index and percent dissolved solids scale."
  45. A. Dunn and R. Richards-Kortum, "Three-dimensional computation of light scattering from cells," *IEEE J. Sel. Top. Quantum Electron.* **2**, 898–905 (1996).
  46. A. Cohen and E. Tirosh, "Absorption by a large sphere with an arbitrary complex refractive index," *J. Opt. Soc. Am. A* **A7**, 323–325 (1996).
  47. R. Aronson, "Diffusion boundary conditions for photon waves," *Proc. SPIE* **2979**, 651–657 (1997).
  48. T. J. Pfefer, L. S. Matchette, and R. Drezek, "Influence of illumination-collection geometry on fluorescence spectroscopy in multilayer tissue," *Med. Biol. Eng. Comput.* **42**, 669–673 (2004).
  49. S. T. Flock, M. S. Patterson, B. C. Wilson, and D. R. Wyman, "Monte Carlo modeling of light propagation in highly scattering tissues—I: model predictions and comparison with diffusion theory," *IEEE Trans. Biomed. Eng.* **36**(12), 1162–1168 (Dec. 1989).
  50. F. Bolin, L. Preuss, R. Taylor, and R. Ference, "Refractive index of some mammalian tissues using a fiber optic cladding method," *Appl. Opt.* **28**(12), 2297–2302 (1989).
  51. S. Patwardhan, A. P. Dhawan, and P. Relue, "Monte Carlo simulation of light-tissue interaction: three-dimensional simulation for transillumination based imaging of skin lesions," *IEEE Trans. Biomed. Eng.* **52**(7), 1227–1236 (2005).
  52. T. J. Pfefer, K. T. Schomacker, M. N. Ediger, and N. S. Nishioka, "Light propagation in tissue during fluorescence spectroscopy with single-fiber probes," *IEEE J. Sel. Top. Quantum Electron.* **7**(6), 1004–1012 (2001).
  53. J. M. Schmitt and G. Kumar, "Turbulent nature of refractive-index variations in biological tissue," *Opt. Lett.* **21**, 1310 (1996).
  54. E. A. Vitol, T. P. Kurzweg, and B. Nabet, "Using white-light spectroscopy for size determination of tissue phantoms," *Proc. SPIE* **5969**, 59691H (2005).
  55. F. Kullmann, H. Messmann, M. Alt, V. Gross, T. Bocker, J. Scholmerich, and J. Rüschoff, "Clinical and histopathological features of dextran sulfate sodium induced acute and chronic colitis associated with dysplasia in rats," *Int. J. Colorectal Dis.* **16**(4), 238–246 (2001).
  56. S. Murthy, A. Flanigan, D. Coppola, and R. Buelow, "RDP58, a locally active TNF inhibitor, is effective in the dextran sulphate mouse model of chronic colitis," *Inflamm. Res.* **51**, 522–531 (2002).
  57. H. C. van de Hulst, *Light Scattering by Small Particles*, Dover Publications, New York (1957).
  58. L. T. Perelman, J. Wu, I. Itzkan, and M. S. Feld, "Photon migration in turbid media using path-integrals," *Phys. Rev. Lett.* **72**, 1341–1344 (1994).
  59. M. A. O'Leary, D. A. Boas, B. Chance, and A. G. Yodh, "Refraction of diffuse photon density waves," *Phys. Rev. Lett.* **69**, 2658–2661 (1992).
  60. A. G. Yodh and B. Chance, "Spectroscopy and imaging with diffusion light," *Phys. Today* **48**(3), 34–40 (1995).
  61. C. F. Bohren and D. R. Huffman, *Absorption and Scattering of Light by Small Particles*, John Wiley and Sons, New York (1983).
  62. R. Drezek, A. Dunn, and R. Richards-Kortum, "Light scattering from cells: finite-difference time-domain simulations and goniometric measurements," *Appl. Opt.* **38**(16), 3651–3661 (1999).
  63. J. Schmitt and G. Kumar, "Optical scattering properties of soft tissue: a discrete particle model," *Appl. Opt.* **37**, 2788–2797 (1998).
  64. A. Brunsting and P. Mullaney, "Light scattering from coated spheres: model for biological cells," *Appl. Opt.* **11**, 675–680 (1972).
  65. G. Streekstra, A. Hoekstra, E. Nijhof, and R. Heethaar, "Light scattering by red blood cells in ektacytometry: Fraunhofer versus anomalous diffraction," *Appl. Opt.* **32**, 2266–2272 (1993).
  66. G. Videen and D. Ngo, "Light scattering multiple solution for a cell," *J. Biomed. Opt.* **3**(3), 212–220 (1998).
  67. A. Nilsson, P. Alsholm, A. Karlsson, and S. Andersson-Engels, "T-matrix computations of light scattering by red blood cells," *Appl. Opt.* **37**, 2735–2748 (1998).
  68. M. Q. Brewster and C. L. Tien, "Radiative transfer in packed fluidized beds: dependent versus independent scattering," *J. Heat Transfer* **104**, 573–579 (1982).
  69. P. Garudadri Venkata, M. M. Aslan, M. Pinar Menguc, and G. Videen, "Surface plasmon scattering by gold nanoparticles and two-dimensional agglomerates," *J. Heat Transfer* **129**, 60–70 (2007).
  70. J. H. Bruning and Y. T. Lo, "Multiple scattering of E. M. waves by spheres. I. multipole expansion and ray-optical solutions," *IEEE Trans. Antennas Propag.* **19**, 378 (1971).
  71. L. Tsang and J. A. Kong, *Scattering of Electromagnetic Waves: Advanced Topics*, John Wiley and Sons, New York (2001).
  72. V. K. Varadan, V. N. Bringi, and V. V. Varadan, "Coherent electromagnetic wave propagation through randomly distributed dielectric

- scatterers," *Physica D* **19**, 2480 (1979).
73. F. Borghese, P. Denti, R. Saija, and G. Toscano, "Optical absorption coefficient of a dispersion of clusters composed of a large number of spheres," *Aerosol Sci. Technol.* **6**, 173 (1987).
  74. D. J. Wilaard, M. I. Mishchenko, A. Macke, and B. E. Carlson, "Improved T-matrix computations for large, nonabsorbing and weakly absorbing nonspherical particles and comparison with geometric optics approximation," *Appl. Opt.* **36**, 4305–4313 (1997).
  75. M. I. Mishchenko, L. D. Travis, and A. A. Lacis, *Scattering, Absorption, and Emission of Light by Small Particles*, Cambridge University Press, Boston, MA (2002).
  76. V. P. Maltsev and K. A. Semyanov, *Characterisation of Bio-Particles from Light Scattering*, Utrecht, Boston (2004).
  77. D. W. Mackowski and M. I. Mishchenko, "Calculation of the T matrix and the scattering matrix for ensembles of spheres," *J. Opt. Soc. Am. A* **13**(11), 2266–2278 (Nov. 1996).
  78. D. W. Mackowski, "Calculation of total cross sections of multiple-sphere clusters," *J. Opt. Soc. Am. A* **11**(11), 2851–2861 (Nov. 1994).
  79. M. I. Mishchenko, "Light scattering by randomly oriented axially symmetric particles," *J. Opt. Soc. Am. A* **8**, 871–882 (1991).
  80. E. Collett, *Polarized Light: Fundamentals and Applications*, Marcel Dekker, New York (1993).