Reconstruction of explicit structural properties at the nanoscale via spectroscopic microscopy

Lusik Cherkezyan
Di Zhang
Hariharan Subramanian
Allen Taflove
Vadim Backman
Reconstruction of explicit structural properties at the nanoscale via spectroscopic microscopy

Lusik Cherkezyan,a Di Zhang,a Hariharan Subramanian,a Allen Tafove,b and Vadim Backman*a,b*

aNorthwestern University, Technological Institute, Department of Biomedical Engineering, E310, 2145 Sheridan Road, Evanston, Illinois 60208, United States
bNorthwestern University, Technological Institute, Department of Electrical Engineering, L359, 2145 Sheridan Road, Evanston, Illinois 60208, United States

Abstract. The spectrum registered by a reflected-light bright-field spectroscopic microscope (SM) can quantify the microscopically indiscernible, deeply subdiffractional length scales within samples such as biological cells and tissues. Nevertheless, quantification of biological specimens via any optical measures most often reveals ambiguous information about the specific structural properties within the studied samples. Thus, optical quantification remains nonintuitive to users from the diverse fields of technique application. In this work, we demonstrate that the SM signal can be analyzed to reconstruct explicit physical measures of internal structure within label-free, weakly scattering samples: characteristic length scale and the amplitude of spatial refractive-index (RI) fluctuations. We present and validate the reconstruction algorithm via finite-difference time-domain solutions of Maxwell’s equations on an example of exponential spatial correlation of RI. We apply the validated algorithm to experimentally measure structural properties within isolated cells from two genetic variants of HT29 colon cancer cell line as well as within a prostate tissue biopsy section. The presented methodology can lead to the development of novel biophotonics techniques that create two-dimensional maps of explicit structural properties within biomaterials: the characteristic size of macromolecular complexes and the variance of local mass density.

Keywords: microscopy; biomedical optics; backscattering; spectroscopy; biophotonics; signal processing.

Paper 150766R received Nov. 11, 2015; accepted for publication Jan. 14, 2016; published online Feb. 17, 2016.

1 Introduction

Visible-light microscopy and spectroscopy techniques remain the principal tools of biological cell and tissue examination in fields from basic science to medical diagnostics. Accordingly, the limitations of these techniques also remain an open problem. Specifically, due to their weakly scattering transparent nature, biomaterials are notoriously difficult to analyze without the use of exogenous labels. In addition, microscopy-based techniques only quantify sample bulk properties, lacking the spatial resolution of obtained information.

It has been recently demonstrated that a tandem application of spectroscopy and microscopy enhances the advantages and mitigates the disadvantages of each technique, showing great promise in a variety of application fields.1 Thus, a spectroscopic microscope (SM), configured to detect interference spectra of backscattered light in the far zone, can quantify the statistics of nanometer scale refractive-index (RI) distribution via the spectral variance (Σ2) of the acquired bright-field image. Further, it has been determined that Σ can sense RI fluctuations at any spatial frequency whatsoever and its lengthscale sensitivity range is limited only by the signal-to-noise ratio (SNR) of the instrument.2

Despite the remarkable ability to sense subtle, microscopically indiscernible structural alterations within weakly scattering label-free media, the quantification of sample’s internal structure via Σ is also associated with a degree of ambiguity. As with most light-scattering markers of structure, it is not always clear which of the two structural properties, the characteristic lengthscale or the amplitude of RI fluctuations, cause a change in Σ during any particular experiment. In addition, the value of Σ is also affected by the sample thickness in a nonlinear manner.2 In this work, we establish that the spectrum registered by an epi-illumination bright-field wavelength-resolved microscope can be analyzed to accurately and explicitly measure sample’s internal structure in terms of physical rather than optical parameters: the standard deviation and characteristic lengthscale of the spatial RI distribution.

2 Theoretical Background

Consider a spatially varying RI object sandwiched between two semi-infinite homogeneous media (Fig. 1). The RIs of the three media are from top to bottom: n0, n1[1 + n2(Δr)] (as a function of location r), and n2. We assume n1 = n2 to approximate the typical case of fixed biological media on a glass slide.3,4 The unit-amplitude plane wave incident normally onto the sample has two distinct sources of reflection: the first is caused by the RI mismatch on one side of the sample (top, air-sample interface in Fig. 1), which is further referred to as reference arm reflectance, and the second is composed of the light scattered from weak RI fluctuations within the sample of interest, which comprises the sample arm. The reference and sample arms are combined to form the wavelength-resolved far-field microscope image. The optical interference of these two components results
in spectral fluctuations of registered intensity, and the variance of these fluctuations is the nanoscale-sensing marker $\Sigma^2$.

We emphasize that no assumptions of one-dimensional light propagation are made in the underlying optics theory, and $\Sigma^2$ is derived from full three-dimensional (3-D) consideration of light scattering and propagation, as well as 3-D specification of RI distribution within the sample. Thus, the instrument specifics of the SM technique are summarized as white-light epi-illumination, bright-field microscope with spectrally resolved image acquisition, small numerical aperture (NA) of illumination (NA < 0.2), moderate NA of collection [NA ∈ (0.3, 0.6)], and with a pixel size of microscope image corresponding to an area in sample space that is smaller than the diffraction limit of light. For simplicity and SNR enhancement, the collection NA used in this work is constant, NA = 0.6. In turn, the sample geometry includes: (i) a weakly scattering sample of interest, (ii) sample thickness not greater than the microscope’s depth of focus (for most setups, 5 to 15 μm), (iii) in the axial dimension, the sample should be RI matched on one side (substrate in Fig. 1) and have a strong RI mismatch on the other (air in Fig. 1) to ensure reference and sample arm light reflections.

Since SM measures interference between a fixed reference-arm reflection and the waves scattered from within the sample, the variance of registered spectral oscillations $\Sigma^2$ can be decomposed into two components:

$$\Sigma^2 = \Sigma_R^2 + \Sigma_L^2,$$

where $\Sigma_L$ is fully described by the RI contrast at the bottom surface, and $\Sigma_R$ is defined by the scattering events occurring within the sample.

In Eq. (1), the optical path difference (OPD) between the interfering waves contributing to $\Sigma_R^2$ is within 0 and $2n_1 L$, and the OPD of interfering waves contributing to $\Sigma_L^2$ is always $2n_1 L$. It follows that it should be possible to independently measure the two components $\Sigma_R^2$ and $\Sigma_L^2$ from the spectral frequency composition of the SM spectrum, i.e., its Fourier transform $\tilde{I}(z)$. Essentially, the Fourier transform of an SM spectrum shows the amount of scattering that has occurred at depth $z = \text{OPD}/2n_1$ inside the sample. For illustration purposes, frequency-space spectrum $|\tilde{I}|$ as a function of depth $z$ corresponding to an infinite spectral bandwidth is shown in Fig. 2.

According to Parseval’s theorem, the spectral variance $\Sigma^2$ is related to the Fourier transform of the spectrum as $\Sigma^2 = \frac{1}{\Delta k} \int |\tilde{F}(k)|^2 dk = \int_{0}^{\infty} |\tilde{I}(z)|^2 dz$. Therefore, $\Sigma_R^2 = E[|\tilde{F}(L)|^2 dz]$ and $\Sigma_L^2 = E[\tilde{I}(L)|^2 dz|$, where $E[.]$ denotes the expected value of a random variable.

However, in practice, the spectral bandwidth is naturally limited to the visible-light wave number range $\Delta k$, and the experimental $\tilde{I}(z)$ is the infinite-bandwidth $\tilde{I}(z)$ convolved with a sinc function $\text{sinc}(z \Delta k/2)$. Hence, a closed-form analytical equation allowing to independently measure $\Sigma_R^2$ and $\Sigma_L^2$ from $\tilde{I}(z)$ does not exist. Nevertheless, this relation can be obtained empirically.

In this work, we develop empirical signal processing algorithm for calculating spectral markers $\Sigma_R$ and $\Sigma_L$ independently. Further, we demonstrate that two optical measures $\Sigma_R$ and $\Sigma_L$ yield two physical measures of sample structure: the standard deviation and the characteristic lengthscale of RI distribution. Using SM data synthesized via finite-difference time-domain (FDTD) solutions of Maxwell’s equations, we validate the developed algorithm on samples with a wide range of structural properties within the example of exponential spatial correlation of RI. We then apply the validated algorithm to experimental data from biological cells and tissues, measuring the explicit physical characteristics of their internal organization.

3 Materials and Methods

3.1 Finite-Difference Time-Domain Simulations

In order to develop the inverse algorithm for $\Sigma_R$ and $\Sigma_L$ determination from the spectral-frequency composition of SM signal, we simulate a physical experiment using rigorous FDTD method, which calculates the light-scattering response of arbitrary inhomogeneous materials via numerical solution of...
Maxwell’s equations of electromagnetics.\textsuperscript{5–8} It is based on the discretization of a 3-D volume into a Cartesian grid consisting of small compared to the wavelength cubic voxels, and the solution of Maxwell’s equations for the evolution of the electric and magnetic field at discrete positions on this Cartesian grid.

The core algorithm of the FDTD method was proposed by Yee\textsuperscript{2} in 1966, and popularized by Taflove in the ‘80s and ‘90s, who also coined the term FDTD. Compared to other electromagnetic approximation methods, such as the finite-element method or the method-of-moments, the FDTD method is more intuitive and simpler to implement. The ease with which inhomogeneous materials are handled in FDTD has made it very attractive for biological applications.\textsuperscript{10–12}

We have an in-house software implementation of the FDTD method, called Angora.\textsuperscript{6,8} It can accurately calculate microscope images of arbitrary inhomogeneous samples under various imaging parameters, incorporating RI fluctuations as fine as 10 nm. We have used Angora to synthesize all reported herein bright-field plane-wave epi-illumination microscope images at 30 different wavelengths between 500 and 700 nm, equally spaced in wave number space.

Sample RI geometry was set to resemble that of fixed biomaterials on glass microscopy slides. RI of dehydrated cells and tissues is reported to be between 1.50 and 1.55,\textsuperscript{3,4} with the exact materials on glass microscopy slides. RI of dehydrated cells and transmission coefficients, \(\Phi_{\text{r}}\), and power spectral density of RI fluctuations \(\Phi_{\text{r}}(k)\) are is isotropic, as they are defined for an unbounded medium \(n^2(r)\), with a physical size much larger than the characteristic length scale of RI fluctuations. The sample, in turn, is defined as a horizontal slice of \(n^2(r)\) with thickness \(L\), where \(L\) can be comparable in magnitude to length scales of sample’s RI fluctuations. We also note that \(\Sigma^2_{\text{R}}\) is a linear function of \(L\). Thus, in this work, we specifically aim to compute the internal-property marker \(\Sigma^2_{\text{R}}/k_cL\), which is independent of sample thickness \(L\).

In an example of exponential functional form of \(B_{n_0}(r)\) with RI fluctuation variance \(\sigma^2_{n_0}\), spatial correlation length \(l_{c_0}\), and \(\Sigma^2_{\text{R}}/k_cL\) is\textsuperscript{1}

\[
\Sigma^2_{l_{c_0}} = \frac{1}{L} \int_{L/2}^{L} E[I(l/2)]^2, \quad (6)
\]

\[
\Sigma^2_{l_{c_0}} = b_1 E[I(l/2)]^2 - b_2 E[I(l/2)]^2, \quad (7)
\]

where \(a = 1.31, b_1 = 2.98, \text{ and } b_2 = 1.39\).

![Fig. 3](https://astronomicaltelescopes.spiedigitallibrary.org/journals/Journal-of-Biomedical-Optics/025007-3/February 2016 Vol. 21(2))

**Fig. 3** \(E[I(l)(z)]\) is obtained by FDTD. Sample thickness 2 \(\mu\)m, exponential functional form of RI spatial distribution with correlation length (a) \(l_{c_0} = 20 \text{ nm}\) (exemplifying a case when \(\Sigma^2_{\text{R}} > \Sigma^2_{l_{c_0}}\)) and (b) \(l_{c_0} = 250 \text{ nm}\) (exemplifying a case when \(\Sigma^2_{\text{R}} < \Sigma^2_{l_{c_0}}\)). On each curve, the depth \(L\) at which \(\Sigma^2_{l_{c_0}}\) is evaluated shown by red circles, and depth \(L/2\) is used for \(\Sigma^2_{\text{R}}\) evaluation by green dots. The shape of \(E[I(l)(z)]\) changes with \(l_{c_0}\) due to the difference in the functional forms of \(\Sigma^2_{\text{R}}(l_{c_0})\) and \(\Sigma^2_{l_{c_0}}(l_{c_0})\).
Since in this simulation, the sample thickness is known a priori, $|\tilde{I}(z)|^2$ was readily evaluated at $z = L/2$ and $z = L$, after which $\Sigma_R/\sqrt{k_c L}$ and $\Sigma_L$ were found according to Eqs. (6) and (7). Figure 3(a) illustrates the match between $\Sigma_R/\sqrt{k_c L}$ and $\Sigma_L$ obtained from the FDTD data according to the derived algorithm with those calculated by analytical Eqs. (4) and (5) for the known sample parameters.

Most importantly, as seen from Eqs. (4) and (5), it is possible to reconstruct the statistics of the internal structure $l_c$ and $\sigma_{n_s}$ once $\Sigma_R/\sqrt{k_c L}$ and $\Sigma_L$ are recovered. As both $\Sigma_R/\sqrt{k_c L}$ and $\Sigma_L$ are linear functions of $\sigma_{n_s}$, the relation between $l_c$ and $\Sigma_L/\sqrt{k_c L}/\Sigma_R$ is only dependent on system parameters controlled by the user ($\Delta k$ and NA):

$$\Sigma_L/\sqrt{k_c L}/\Sigma_R = \sqrt{\frac{\pi}{8}} \left[ 1 \left( 1 + \frac{1}{(xN^A)^2} \right) \right] \frac{1}{x^2(4 + N^A)^2}$$

(8)

While a solution for $x$ can be found numerically, for simplicity and computational speed, we exploit the fact that $\Sigma_L/\sqrt{k_c L}/\Sigma_R(x)$ is well approximated as a linear function for a wide range of correlation lengths above 15 nm and NA: for any NA within 0 to 0.6, the $r^2$ of linear regressions for $\Sigma_L/\sqrt{k_c L}/\Sigma_R$ for $l_c$ between 15 and 600 nm is above 0.98. In the case of NA = 0.6 considered throughout this work, $x$ can be found for any sample thickness simply as

$$x = 1.7 \frac{\Sigma_L}{\Sigma_R} \sqrt{k_c L} + 0.45,$$

(9)

with the fit shown in Fig. 4. The obtained value of $x$ can be substituted into Eq. (4) to find $\sigma_{n_s}$ as

$$\sigma_{n_s} = \frac{2\Sigma_L}{\Gamma \sqrt{1 - 1/\sqrt{1 + (xN^A)^2}}}.$$

(10)

Thus, our computations (results depicted in Fig. 5) postulate that $l_c$ and $\sigma_{n_s}$ can be independently calculated from SM signal. The condition necessary for using the corresponding empirical Eqs. (6) and (7) followed by Eqs. (9) and (10) is the knowledge of sample’s thickness $L$.

3.2.2 Reconstruction of sample thickness from spectroscopic microscope data

We next develop signal processing algorithm for accurate measurement of sample thickness from SM data. Finding sample thickness is complicated in part by the fact that the sample-substrate interface does not always reflect light due to the low RI contrast at the bottom interface typical to fixed biomaterials on glass. Specifically, this is the case when $l_c$ is smaller than the diffraction-limited spot [as in Fig. 3(a)] and the frequency-space spectrum $E(\tilde{I}(z))$ does not necessarily contain an evident peak at $z = L$.

Nevertheless, since no light scattering events occur at $z > L$, $E(\tilde{I}(z))$ always decays at $z > L$. The shape of $E(\tilde{I}(z))$ decay “tail” at $z > L$ has no closed-form analytical expression and depends on the sample’s internal structure as well as the spectral bandwidth of light.

Based on the FDTD-synthesized SM data used to develop Eqs. (6) and (7), a fourth order polynomial was fitted to all calculated frequency-space spectra $E(\tilde{I}(z))$ at $z > L$. Since the shape of $\tilde{I}(z)$ in general, and its decay at $z > L$ in particular, depends on $l_c$, this was done for SM images of samples with all 20 values of $l_c$ (between 20 and 250 nm). As a result, a $4 \times 20$ matrix of corresponding polynomial coefficients was stored. Then, a MATLAB function was created to find a best fit between

![Fig. 5 Accuracy of sample parameters measured from the FDTD-generated SM signal from a sample with a known thickness: (a) $\Sigma_R/\sqrt{k_c L}$ and $\Sigma_L$ as a function of the correlation length $l_c$, obtained for a sample with thickness $L = 2 \mu m$ according to analytical equations (solid lines) and calculated from FDTD simulation (circles with error bars corresponding to the standard deviation between nine ensembles per statistical condition). (b) Corresponding true $\sigma_{n_s}$ and $l_c$ (solid lines) and reconstructed from FDTD-synthesized SM signal (circles with error bars corresponding to the standard deviation between nine ensembles per statistical condition).](image-url)
an arbitrary experimentally measured $E[\tilde{I}(z)]$ decay and the created library of $E[\tilde{I}(z)]$ decays for various values of $l$. Finally, the sample thickness is determined from the location of the best fit of the experimental and FDTD-obtained $[\tilde{I}(z)]$ decay tails.

To summarize, determining sample thickness from an arbitrary experimentally obtained spectrum allows the evaluation of $[\tilde{I}(z)]^2$ at $z = L/2$ and $z = L$, yielding $\hat{\Sigma}_R/\sqrt{k_cL}$ and $\hat{\Sigma}_c$ [Eqs. (6) and (7)], as well as $l_e$ and $\sigma_{sa}$ [Eqs. (8) and (10)], which completes the developed inverse algorithm here.

### 3.2.3 Considerations of sample roughness

It is important to consider that in true experimental conditions the surface roughness properties are most often unknown. Thus, an optimal spectral processing algorithm must universally apply to samples with rough as well as smooth surfaces. In cases when the sample top surface is rough, the registered spectroscopic microscopy signal has an additional, low spectral frequency component. Thus, to remove this spectral feature related to a property of sample surface, a second order polynomial is fitted to the registered spectrum and subtracted from it prior to converting the spectrum into spectral frequency space. We validate this step by applying our inverse algorithm to FDTD-synthesized SM images of weakly scattering samples with a rough surface profile that is characteristic of biological cells and tissues (average thickness of 2 $\mu$m, standard deviation of nanoscale height variations within a diffraction-limited area 22 nm, correlation length of height variations 170 nm, internal RI $l_e = 100$ nm, and $\sigma_{sa} = 0.033$).

### 3.3 Spectroscopic Microscopy Instrument

For SM experimental measurements, we used our in-house instrument previously built for high-throughput partial wave spectroscopy measurements: epi-illumination bright-field microscope with small illumination NA = 0.15, moderate collection NA = 0.6, and 40x magnification (objective lens from LUCPlanFL N, Olympus, Center Valley, Pennsylvania). Koehler illumination scheme was implemented for uniformity of incident light intensity throughout the image. Wavelength-resolved image acquisition was accomplished by using Xenon white light illumination and spectral filtration of the light incident onto the sample via acousto-optical tunable filter (AOTF, HSI-300, Gooch & Housego, Orlando, Florida; filter bandwidth of 3 nm). As a result, each measurement recorded a 3-D ($x, y, \lambda$) data cube consisting of sample bright-field microscope images ($x, y$) obtained at 200 $\mu$m-spaced wavelengths $\lambda$ within the spectral range of 500 to 700 nm.

After the data were collected, spectral noise was removed from spectra corresponding to each pixel of the acquired wavelength-resolved microscope image by a low-pass spectral filter using sixth order butterworth filter with a 0.2 frequency cutoff. Then, data postprocessing was performed in exact accordance with algorithm developed and applied to FDTD-synthesized spectra, which included second-order polynomial fitting and subtraction, Fourier transformation of spectra with the use of Hann window and zero-padding, followed by the analysis of the resultant $E[\tilde{I}(z)]^2$ spectrum.

### 3.4 Colon Cancer Cell Lines

The performance of the proposed analysis was first tested on HT29 human colonic adenocarcinoma cell line models. The experiment included two groups, control vector HT29 (CV) cells and epidermal growth factor receptor (EGFR) knockdown HT29 cells, a less aggressive genetic variant that is histologically indistinguishable from the CV.

HT29 CV and EGFR knockdown cells were collected in centrifuge tubes and centrifuged for 5 min at 1000 rpm. The supernatant was removed, after which the cells were plated on a glass chamber slide: 2 mL of fresh cell culture medium was added to each chamber slide, which was then incubated at 37°C for 6 h. After incubation, the medium was completely removed, the slides were washed and then immediately fixed using 70% ethanol, which completed sample preparation. Slides were stored at 4°C until the spectroscopic microscopy data from 18 CV cells and 17 EGFR knockdown variant cells were acquired. Topography of the same cells was later obtained via an atomic force microscope (AFM) to validate the cell thickness predictions obtained from SM data.

### 3.5 Tissue Section

The second biological model for analysis algorithm application and testing included human prostate tissue biopsy section. Collection of the human sample was approved by the Institutional Review Board at NorthShore University HealthSystem. Sample was obtained from the NorthShore University active surveillance trial initiated in November 2008. Informed written consent was obtained from the participant.

Transrectal biopsy was obtained with 3-D ultrasound guidance, fixed in ethanol and embedded in paraffin. Then, the sample was sectioned, and two sections were applied to a glass slide, after which they were deparaffinized following standard histological procedures. Using one section, both SM and AFM data were collected from the same region of the sample for structural property determination. The other section was stained with hematoxylin and cosin (H&E) to aid in organelle visualization, which was used only for illustration purposes.

### 3.6 Atomic Force Microscope

Height map of the biological samples was determined at room temperature by peak force tapping mode using a Bruker Dimension Icon AFM system with silicon OTESPA-R AFM probes (Bruker AXS).

For cell lines, $30 \times 30 \mu$m$^2$ image was obtained with pixel size of 46.9 nm and for tissue section, $90 \times 90 \mu$m$^2$ image was obtained with pixel size of 176 nm. Image magnification (and, therefore, pixel resolution) for tissue section image was chosen so that the AFM-imaged area of sample surface captures the same area as within the SM field of view. At the same time, larger magnification of cell line images was chosen due to their surface area being much smaller than the microscope’s field of view.

### 4 Results

#### 4.1 Algorithm Validation on Finite-Difference Time-Domain Generated Data

We validate the proposed inverse algorithm by applying it to a validation set of Angora-synthesized SM data for samples with a wide range of internal and surface properties.

First, we confirm that the developed algorithm is accurate for experimentally realistic samples with uneven surface. To accomplish this, we apply the analysis algorithm to FDTD-synthesized...
SM images of 20 rough samples with RI distribution \( l_c = 100 \) nm, \( \sigma_{\Delta} = 0.033 \), and average \( L = 2 \) \( \mu \)m. The roughness of these samples was set to resemble that of biological cells and tissues.\(^{16}\) Our analysis results reconstructed the internal structure characteristics of the imaged samples with excellent accuracy: predicted \( l_c = 93.1 \) nm (6.9% error from true value), \( \sigma_{\Delta} = 0.0315 \) (3.7% error) and predicted \( L = 1.96 \) \( \mu \)m (2.3% error).

This validates that subtraction of low-frequency components from the registered spectra via a second-order polynomial efficiently removes surface roughness contributions to the SM signal and allows accurate calculation of the internal properties. Note that the lowest frequencies removed by the polynomial correspond to small values of \( z \) in \( E[|\tilde{I}(z)|^2] \) profile, which are never sampled when \( \Sigma_R \) and \( \Sigma_I \) are evaluated. As a result, our reconstruction algorithm remains accurate despite this additional spectral processing step.

In order to keep the spectral processing algorithm independent of the sample surface features, we subtract the low-frequency components from spectra obtained from all samples, including the validation set below as well as the biological cells and tissue.

Next, as a validation set for the inverse algorithm, we analyzed SM images of samples with various combinations of all three measured parameters—\( l_c \), \( \sigma_{\Delta} \), and \( L \)—covering the biologically relevant range of structural properties. Figure 6(a) illustrates the excellent accuracy in the prediction of \( \sigma_{\Delta} \) obtained by the inverse algorithm from the FDTD-generated spectrally resolved microscope images. In this set, \( l_c = 100 \) nm and \( L = 2 \) \( \mu \)m were fixed and were predicted within 15-nm accuracy for \( l_c \) and 6% accuracy for \( L \). Figures 6(b)–6(d) present prediction accuracy in a more complex case where RI correlation length was varied within the subdiffractional range, and a new thickness value \( L = 3 \) \( \mu \)m was tested. Note that at \( l_c < 50 \) nm where the percentage of error in \( l_c \), prediction is relatively high, the predicted values of \( l_c \) are still within tens of nanometers from the corresponding true value [Fig. 6(b)].

### 4.1.1 Minimum thicknesses for analysis validity

Importantly, the finite spectral bandwidth of light \( \Delta k \) naturally imposes limitations to the closest spectral frequencies of the SM data that can be resolved. In turn, since OPDs of all interfering waves are within the range of 0 to 2\( n \)L, the sample thickness confines the range of spectral frequencies present in the SM signal. As a consequence, in the limit of very small \( L \), values of \( E[|\tilde{I}(z)|^2] \) at \( z = L/2 \) cannot be resolved from those at \( z = L \), and the key Eqs. (6) and (7) cannot be used. Thus, there must be a lower limit to the sample thickness in order for the developed herein analysis of spectral-frequency profile to apply.

We determine the lower limit of \( L \) for our algorithm accuracy using sets of SM images generated for samples with various thicknesses \( (L \) as low as 500 nm was tested). For every thickness, SM images of samples with 20 subdiffractional \( l_c \) within 20 to 250 nm were obtained in order to ensure that the thickness limitations will be determined for a general case, independent of the internal RI distribution.

We applied the inverse algorithm to calculate all three parameters: \( l_c \), \( \sigma_{\Delta} \), and \( \sigma_{\Delta} \) and subsequently evaluated the error between true and predicted parameters (as in Fig. 6). Results of the error evaluation are summarized in Table 1.

Our analysis shows that accuracy in measuring \( l_c \) is a greater challenge than that in \( \sigma_{\Delta} \) for the inverse algorithm. Moreover, \( \sigma_{\Delta} \) predictions remain extremely robust even in cases when the error in the predicted \( l_c \) is large [see Figs. 5(b), 6(a), and 6(c)]. We believe this to be explained by the fact that while the shape of SM spectral-frequency profile depends strongly on \( l_c \) [due to the difference in \( \Sigma_R(l_c) \) and \( \Sigma_I(l_c) \)], both \( \Sigma_R \) and \( \Sigma_I \) scale linearly with \( \sigma_{\Delta} \). Thus, errors in the quantification of the shape of \( E[|\tilde{I}(z)|^2] \) have a stronger effect on the accuracy of \( l_c \) rather than on that of \( \sigma_{\Delta} \).

To summarize, the inverse algorithm was developed on the testing set of samples with \( L = 2 \) \( \mu \)m and a subdiffractional range of RI correlation lengths 20 to 250 nm. Then, we confirmed that the algorithm is accurate even in the cases when samples have a rough surface. Finally, we applied the algorithm to accurately reconstruct the internal structure from SM data generated for a large set of inhomogeneous weakly-scattering samples, where, in addition to different \( l_c \) values, we varied thickness from 0.5 to 3 \( \mu \)m Fig. 7, and \( \sigma_{\Delta} \) from 0.02 to 0.05. Results of the above testing and validation procedures specify the accuracy as well as the applicability range of the proposed methodology (Table 1), positioning us for data processing from complex experimental samples such as biological cells and tissues.

### 4.2 Cell Lines

After validating our methodology on FDTD solutions of Maxwell’s equations, we proceeded with experiments on
isolated biological cells. As a model, we chose HT29 human colonic adenocarcinoma cell line. We have also used a genetic variant of HT29 cells with EGFR knockdown, which partially suppresses the proliferation aggressiveness of the cell line without changing its microscopically visible morphological qualities.16–20

Our theoretical derivations as well as FDTD simulation have shown that the single necessary condition for obtaining the $l_c$ and $\sigma_{n_a}$ values from SM data is accurate determination of sample thickness. Thus, we specifically ensure that the predictions of $L$ obtained from experimental SM data are accurate by comparing them to the values measured by AFM from the same cells.

Results of this comparison are shown in Fig. 8. White-light epi-illumination microscope image of an isolated HT29 cell [Fig. 8(a)] and the corresponding AFM image of the same cell [Fig. 8(b)] illustrate the spatial height variations and roughness of cell surface that emphasize the complexity of biological cell measurements. Note that to obtain the ensemble average $E[|\tilde{T}(z)|]$, Fourier transforms of independent spectra are averaged. In the case of biological cells, this entails averaging data from areas with different thicknesses. Despite this, our predictions of 35 cell-averaged thickness values with the assumption of average RI to be 1.53 (RI of fixed biological cells and tissues)13,14 show an excellent match with the physical thickness of cells measured with AFM [Fig. 8(d)].

Next, we obtained $l_c$ and $\sigma_{n_a}$ predictions for the same biological cells. Our results showed that the characteristic length scale of RI distribution within the two genetic variants is approximately the same (552 nm inside the CV and 529 nm in EGFR-knockdown variant). However, the standard deviation of RI within those two genetic variants was drastically different: 0.04 inside the CV and 0.02 inside the less aggressive EGFR-knockdown variant (Fig. 9). This difference was statistically significant with a $p$-value of 0.03.

### 4.3 Tissue Section

Last, we demonstrate that the developed methodology can be used for spatially resolved quantification of the internal structure of biological tissue specimens on an example of a sectioned human prostate tissue biopsy. Transmission bright-field microscopy image of the neighboring hematoxylin- and eosin-stained section is shown in Fig. 10(a) and bright-field reflectance microscope image of the unstained sample of interest is shown in Fig. 10(b). Following the above-developed procedure, we first confirm the accuracy of the reconstruction algorithm by comparing AFM-measured sample thickness with that predicted using SM data [Figs. 10(c) and 10(d)]. Here, the large (compared to single isolated biological cells) sample area allowed data analysis in a spatially resolved manner, and the obtained

---

**Table 1**: Range of sample thicknesses for which quantification of the internal ($l_c$, $\sigma_{n_a}$) as well as external ($L$) properties of the sample is accurate.

<table>
<thead>
<tr>
<th>Property</th>
<th>Applicable samples</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L$</td>
<td>$L \geq 1.0 \ \mu m$</td>
<td>$&lt;5%$</td>
</tr>
<tr>
<td>$\sigma_{n_a}$</td>
<td>$L \geq 1.5 \ \mu m, l_c &gt; 50 \ \text{nm}$</td>
<td>$&lt;10%$</td>
</tr>
<tr>
<td>$l_c$</td>
<td>$L \geq 2.0 \ \mu m, l_c &gt; 50 \ \text{nm}$</td>
<td>$&lt;15%$</td>
</tr>
</tbody>
</table>

---

**Fig. 7** Thickness prediction from SM frequency spectra, errors are standard deviations across 20 samples with different RI correlations lengths per thickness.

**Fig. 8** Experimental results from biological cells: (a) representative bright-field reflectance white-light image of a cell; (b) representative height map obtained by AFM from the same cell (EGFR-knockdown variant shown here), scale bar 5 $\mu m$; (c) tail-fitting of experimentally obtained Fourier transform of reflectance intensity averaged throughout the cell area of interest; red dot corresponds to the predicted thickness, where $\tilde{L}_k$ is estimated and the green dot corresponds to the midsample depth, where $\tilde{L}_m$ is estimated for this particular spectrum. (d) Match between the cell thickness evaluated from the frequency content of its reflectance intensity compared to that obtained with AFM; horizontal error bars indicate the range (standard deviation) of thicknesses within the sample as measured by AFM.
thickness map of the sample within microscope’s field of view showed an excellent match with that measured with AFM (pixels with SNR < 1.25 were excluded from the SM data analysis). Accurate evaluation of the observed match between the AFM- and the SM-calculated sample topology is complicated by the differences in pixel sizes, spatial resolutions, as well as sample orientation in the respective images acquired by the two techniques. Thus, after applying a Gaussian blur to the AFM image to approximate a microscope’s diffraction-limited resolution, rotating and extrapolating images (to match the pixel size), we have estimated that 74% of pixels used for internal property reconstruction had an SM-measured thickness within 20% from that measured by AFM.

Spatially resolved values of $l_c(x,y)$ and $\sigma_{nc}(x,y)$ were then obtained, as depicted in Figs. 10(g) and 10(h). As per algorithm limitations determined and reported above based on FDTD data analysis (Table 1), we have calculated $l_c(x,y)$ only for areas with thickness $L(x,y) > 2 \mu m$ (which comprised 48% of sample area) and $\sigma_{nc}(x,y)$ only for areas with thickness $L(x,y) > 1.5 \mu m$ (70% of sample area).

We also note that the relative blurriness of $l_c(x,y)$ and $\sigma_{nc}(x,y)$ distributions is due to the fact that each value is a cumulative statistical characteristic of sample structure within a moving window of $25 \times 25$ pixels, which corresponds to $3.8 \times 3.8 \mu m^2$. For the studied tissue sample, we found the mean and most common $l_c$ values to be 118 and 80 nm correspondingly. In turn, the mean and most common $\sigma_{nc}$ values were found to be 0.020 and 0.012. Based on this very limited dataset, the overall shape of both $l_c$ and $\sigma_{nc}$ value distributions appeared to follow a lognormal functional form.

5 Discussion and Conclusions

In this work, we demonstrate that the spectral-frequency composition of a wavelength-resolved image registered by a reflected-light, bright-field microscope can be analyzed to independently obtain two explicit physical measures of the RI distribution within weakly scattering samples such as biological cells and tissues: the standard deviation and the spatial correlation length. Since the local mass density is a linear function of RI within biomaterials (Gladstone–Dale relation), these measures of RI distribution directly translate into statistics of mass density distribution inside biological cells and tissues: the correlation length of mass density is exactly the same as that of RI, and the standard deviation of mass is that of RI divided by the RI-mass proportionality coefficient $\alpha = 0.18 ml/g$. In biological terms, variance of local mass density $\sigma^2_m(x,y)$ quantifies the compaction degree of macromolecular complexes (folded proteins, chromatin aggregates, etc.) contained within the volume underneath a diffraction-limited area surrounding each pixel $(x,y)$. In turn, $l_c(x,y)$ is the characteristic size of macromolecular complexes within that same volume. Hence, measurement of $\sigma^2_m(x,y)$ and $l_c(x,y)$ is an important tool in studies of structure–functional relationship in crucial biological processes including cancer initiation and progression (epigenetic changes observed in fixed-cell nucleus, cytoplasm, etc.).
extracellular matrix, etc., cell proliferation, as well as genome dysregulation and potential therapy.

Conceptually, the developed algorithm utilizes the physical meaning behind the spectral-frequency profile \( E[|\tilde{I}(z)|^2] \), which allows to obtain several independent parameters of sample’s organization by evaluating \( E[|\tilde{I}(z)|^2] \) at different \( z \). First, no scattering events occur at \( z > L \), which we use to measure \( L \). Second, at \( z = L \), \( E[|\tilde{I}(z)|^2] \) is predominantly defined by the amount of light reflected at the sample–substrate interface, which we use to measure \( \Sigma_z \). Third, at \( z \in (0, L) \), \( E[|\tilde{I}(z)|^2] \) represents the amount of scattering from within the sample, which determines \( \Sigma_R/\sqrt{k_c L} \). Essentially, the reflection at \( z = L \) is defined by the two-dimensional (2-D) statistics of RI distribution, \( n_R(x, y, L) \), and the scattering at \( z \in (0, L) \) is defined by the 3-D statistics of RI, which is why \( \Sigma_R/\sqrt{k_c L} \) and \( \Sigma_z \) probe the sample structure in a truly independent manner. We also note that due to the statistical homogeneity considered here, \( \Sigma_R/\sqrt{k_c L} \) could be computed from the value of \( E[|\tilde{I}(z)|^2] \) for virtually any \( z \) between 0 and \( L \). However, we have chosen to calculate \( \Sigma_R/\sqrt{k_c L} \) at the midpoint of \( z = L/2 \) in order to minimize the inevitable (due the finite spectral bandwidth) contributions from surface roughness at low \( z \) and \( \Sigma_z \) at \( z = L \). Finally, at low \( z \), \( E[|\tilde{I}(z)|^2] \) also contains information about the sample surface roughness profile in addition to its internal inhomogeneity. Since the present work does not aim to measure surface statistics, the surface-related contributions are simply removed from the signal.

Technically, the underlying algorithm is composed of (i) numerical curve fitting to measure \( L \), (ii) an empirical step to obtain the exact values of \( \Sigma_R/\sqrt{k_c L} \) and \( \Sigma_z \) from \( E[|\tilde{I}(z)|^2] \), and (iii) reconstruction of structural properties \( l_e \) and \( \sigma_{na} \) based on analytical closed-form equations.

We have tested and validated the inverse algorithm using FDTD solutions of Maxwell’s equations. The two important advantages of FDTD for algorithm validation are (1) sample exact structure known a priori and thus the technique precision can be readily evaluated and (2) experimental noise and other sources of error are absent. The testing set included SM images synthesized for samples with thickness of 2 \( \mu \)m, and RI correlation lengths of 20 to 250 nm. Then, the algorithm accuracy has been extensively studied and validated on a larger set of samples with thickness of 2 \( \mu \)m, and RI correlation lengths of 20 to 250 nm. Therefore, the algorithm accuracy has been extensively studied and validated on a larger set of samples with thickness of 2 \( \mu \)m, and RI correlation lengths of 20 to 250 nm. Now, the algorithm accuracy has been extensively studied and validated on a larger set of samples with thickness of 2 \( \mu \)m, and RI correlation lengths of 20 to 250 nm. Now, the algorithm accuracy has been extensively studied and validated on a larger set of samples with thickness of 2 \( \mu \)m, and RI correlation lengths of 20 to 250 nm. Now, the algorithm accuracy has been extensively studied and validated on a larger set of samples with thickness of 2 \( \mu \)m, and RI correlation lengths of 20 to 250 nm.
The presented algorithm has been developed with the specific application focus of measuring internal structure of fixed biomaterials. As a result, the sample mean RI was accordingly set to 1.53 throughout the algorithm testing and validation procedures. The observed match between sample thickness determined by AFM (which measures physical thickness) and SM (which measures optical thickness and recovers L using the assumption of mean RI) confirms the accuracy of assumed average RI value for fixed cells and tissue. In addition, the choice of glass slides as sample substrates has also eliminated deterministic light reflection at the bottom interface due to the match between glass and sample average RIs. While these average RI choices do not change the nanoscale structure sensitivities of SM1 and do not affect our algorithm derivations from the physics perspective, they define the scope of the sample/substrate microscope properties tested in the presented work. Following the outlined framework, the algorithms presented herein can be easily extended to other sample/substrate properties as well as other applications of biophotonics. For example, one could introduce an RI mismatch at the sample–substrate interface to accentuate the $E[|I(z)|^2]$ peak at $z = L$ and subsequently remove this deterministic contribution to $E[|I(L)|^2]$ when $\sigma_L$ is calculated.1

From the perspective of structure parametrization, the internal properties $\sigma_L^2$ and $\ell_z$ measured by the inverse algorithm correspond to the height and width of the spatial RI correlation function. Thus, the value of $\sigma_L$ is independent of the sample’s lengthscale composition and, therefore, of the shape of spatial correlation function. At the same time, our definition of $\ell_z$ as the correlation length of RI presented here involves an assumption of exponential RI correlation. We note that previous calculations based on electron microscopy images of biological cell nuclei have shown that the SM signal predicted based on their experimentally measured RI distribution matches that predicted based on an $\ell_z$ value that assumes an exponential RI correlation.2 Thus, even if under certain experimental conditions, the exact value of $\ell_z$ can be prone to error, we still firmly believe that it provides a valuable measure of sample’s lengthscale composition regardless of the functional form of its RI correlation function.

The technique presented herein is unique in its nanoscale sensitivity, versatility, and ease of application: by an automated analysis of a wavelength-resolved far-field microscope image, it can explicitly measure physical properties native to weakly scattering samples. Moreover, it requires no external labels or labor-intensive sample fixation/processing procedures. However, this great advantage also defines the main limitation of the present work, as the exact values of $\ell_z$ and $\sigma_L$ measured from biological samples cannot be corroborated by another independent technique. Thus, corroboration of the $\ell_z$ and $\sigma_L$ values predicted by SM would require imaging of the 3-D native mass distribution within the same biological cells with nanometer resolution, which is prohibitive with current state-of-the-art technology. Nevertheless, the measured spatial standard deviation of RI $\sigma_L = n_0 - n_{\text{true}}$ from both sets of experimental data is within the interval of (0.02, 0.06), which agrees with the estimates based on a discrete-particle model of soft tissue.15 Analogous estimates for correlation length of RI fluctuations inside label-free ethanol-fixed biological cells and tissues are not reported in previous literature, owing by large to the abovementioned technical limitations. Future work will focus on validation of the presented algorithm on experimentally measured data from samples of known internal structure (controlled photons or biological samples quantified via emerging nano-scale-imaging methodologies such as correlative light-electron microscopy).3,2

In summary, we establish that the spectrum registered by a reflected-light microscope can be analyzed to independently reconstruct two physical measures of internal structure within samples such as biological cells and tissues. Applying this approach can lead to the development of novel biophotonics techniques capable of creating 2-D images of intracellular mass-distribution properties such as characteristic size of macromolecular complexes and variance of local mass-density. The ease of utilization as well as the most intuitive physical meaning of measured parameters (as opposed to optical markers of structure) will make this approach widely applicable for users in fields from basic biology, material science to medical diagnostics.

**Acknowledgments**

This study was supported by the National Institutes of Health under Grant Nos. R01CA200064, R01CA155284, R01CA165309, and R01EB016983, and by the Lungevity Foundation. The FDTD simulations in this paper were made possible by a computational allocation from the Quest high-performance computing facility at Northwestern University. H.S. and V.B. are cofounders and/or shareholders in Nanocytollic LLC.

**References**


Lusik Cherkezyan is a postdoctoral fellow in the Department of Biomedical Engineering at Northwestern University, Illinois, where she has also received her PhD. Her work focuses on the use of physics and engineering principles to the study of biological systems. In particular, she is interested in the technology development for the spectroscopic quantification of biomaterials at nanometer scales. Her work has been published in multiple high-profile journals, including *Physical Review Letters, BMC Cancer, Optics Letters, Endoscopy*, etc.

Di Zhang received his BS degree in electrical engineering from Beijing Jiaotong University, China, in 2009. He is currently working toward his PhD degree in biomedical engineering at Northwestern University, Evanston, Illinois, United States. His research interests include computational electromagnetics for modeling light interaction with biological tissue and optical imaging techniques for early stage cancer detection.

Hartharan Subramanian received his PhD in biomedical engineering from Northwestern University and is currently a research professor of biomedical engineering at Northwestern University. He is the cofounder and chief technology officer of NanoCytomics (www.nano-cytomics.com), an in vitro medical diagnostic company developing screening strategies for different types of cancers (e.g., lung, colon, prostate, etc.). He has considerable experience in biomedical optics, cancer biology, and clinical research with numerous publications appearing in leading peer-reviewed journals.

Allen Taflove is a professor in the Department of Electrical Engineering and Computer Science of Northwestern University, Evanston, Illinois. Since 1972, he has pioneered finite-difference time-domain (FDTD) computational solutions of Maxwell’s equations, for which he received the 2014 IEEE Electromagnetics Award. His major publication—*Computational Electrodynamics: The Finite-Difference Time-Domain Method—is the seventh most-cited book in physics.

Vadim Backman is the Walter Dill Scott professor of biomedical engineering at Northwestern University and a program leader at the Robert H. Lurie Comprehensive Cancer Center in Chicago, Illinois. An internationally renowned expert in biomedical optics, he develops revolutionary nanoscale imaging technologies that allow researchers to explore previously intractable questions in biology, disease diagnosis and progression, with a focus on detecting cancer at its earliest and most treatable stages.