Numerical Breast Models for Commercial FDTD Simulators

Guangran Zhu, Boris Oreshkin, Emily Porter, Mark Coates, and Milica Popović

Department of Electrical and Computer Engineering, McGill University
3480 University St, Montreal, QC, Canada
{Guangran.Zhu, Boris.Oreshkin, Emily.Porter}@mail.mcgill.ca
{Mark.Coates, Milica.Popovich}@mcgill.ca

Abstract—This paper presents the development of numerical human breast models suitable for commercial finite-difference time-domain (FDTD) simulators. The geometry of the breast models is derived from images obtained from Magnetic Resonance Imaging (MRI) scans. To avoid assigning tissue properties to every voxel, we apply the regression tree analysis to partition the breast tissue region into cuboid regions (cells) that exhibit similar pixel intensity (and hence have similar tissue structure). The local spatial averaging performed by the analysis addresses the MRI-inherent noise. Secondly, we use dielectric and Debye material to model the heterogeneity and dispersiveness of breast tissue. We find that Debye material offers higher attenuation in the high frequency region than dielectric material. We also confirm that assuming a fixed relaxation time constant in Debye material does not significantly affect the field.
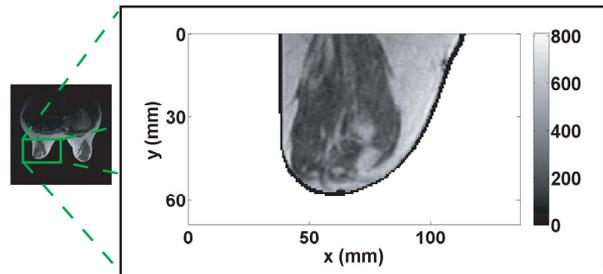
Fig. 1. A slice of the original MRI image and the segmentation stage of the model construction. The skin, outlined in solid black, and the tissue region are segmented from the black background of the original MRI slice.

I. INTRODUCTION

Microwave radar breast imaging exploits the contrast in the dielectric properties between malignant and healthy breast tissues. Numerical simulation, using the finite-difference time-domain (FDTD) technique, remains a critical tool for the further development of imaging methodologies. Early numerical breast models possess over-simplified geometries, rendering the results unreliable. More complicated models with fine-grained structural resolution, e.g. [1], increase the computational complexity to a level that requires custom software. Integrating complex antennas become very difficult, so simulations adopt a plane-wave or point-source excitation.

In this paper, we design a model that retains all important dielectric structure, but intelligently granularizes the resolution of the breast tissue region to reduce the number of dielectric solids that must be simulated. We use regression tree analysis [2] to process MRI images, simultaneously reducing MRI noise by local spatial averaging and identifying the important structural features of the tissue. The regression tree analysis identifies a set of cuboid solids ('cells') to pad the tissue region of breasts.

We also investigate three tissue mapping strategies. The dielectric map and Debye map are from [1]. Zastrow assumes a fixed relaxation time constant when calculating the Debye parameters. We lift this constraint and devise a similar mapping strategy. We apply these three strategies to the breast model and find that 1) Using Debye materials incur larger attenuation than dielectric materials in the high frequency region of signals. 3) The assumption of a fixed relaxation time

constant in the Debye model has little effect on the computed field.

In the next section, we illustrate the process of developing the breast model and explain the tissue mapping strategies. Then we present an example in Section III, and discuss the effect of cell padding and dielectric versus Debye approximation in Section IV. Section V concludes this paper.

II. INTRODUCTION

The breast model development is a three-stage process. The first stage involves segmenting the MRI images; the second stage consists of regression tree analysis to partition the tissue region into non-uniform solids (cells); and the third stage involves tissue property assignment to the cells.

Stage 1: MRI Image Segmentation: The allocation of tissue in our breast model is derived from a set of MRI images. The initial processing involves manual identification of the breast region and its separation from the background. We then segment the skin and tissue regions. The resulting skin and tissue boundaries are triangularly meshed and imported into SEMCAD. Fig. 1 shows a sample of the segmented MRI images.

Stage 2: Regression Tree Analysis: Direct mapping of pixel intensities onto an FDTD grid leads to many tissue solids (approximately 300,000 for our example). Incorporating existing microwave radar antenna designs into an FDTD model with this many solids leads to a vast number of grid points and renders commercial FDTD software unusable. It is highly desirable to reduce the number of solids by developing a much

simpler model that preserves the electrical response to within some small approximation error.

Moreover, it is questionable whether the individual pixel intensity measurements accurately reflect the underlying tissue properties. There is significant noise in the MRI measurements, so the true pixel intensities can differ significantly from the measured values. We can form better estimates of the true intensities by exploiting the local spatial smoothness that should be present in the MRI — the MRI response should not change drastically throughout the breast except where there are structural discontinuities, e.g., transitions from muscular to adipose tissue.

We adopt a regression methodology to estimate the underlying true intensities. Our regression strategy involves partitioning the MRI volume in a tree-structured fashion, and then constructing an estimator $\hat{f}$ to fit the data in each cell of the partition. Such a tree decomposition has been used before in MRI analysis, both for the identification of activity regions in fMRI [3] and in multi-resolution analysis for MRI denoising [4], [5].

We employ a version of the CART (classification and regression trees) algorithm [2]. CART involves a growing phase, in which the input domain is partitioned into successively finer cells, to construct an initial tree that provides a very good, and perhaps perfect, match to the data. Although this excellent match may seem desirable, it is indicative of overfitting; we are building a model that fits the noise inherent in the measurements, rather than revealing the true function lurking beneath the noise. For this reason, the second phase of CART involves pruning, where cells are successively merged. CART then selects the tree that minimizes

$$C(T) = \widehat{L}_n(T) + \alpha|T|,$$

where $\widehat{L}(T)$ is the empirical risk (error measure between the model and the data), $|T|$ is the cardinality of the tree (number of leaf cells), α is a complexity constant that controls the trade-off between fidelity to the data and the complexity of the regression model, and n is the number of data points. The empirical risk we adopt is the mean-squared error,

$$\widehat{L}(T) = \frac{1}{n} \sum_{i=1}^{n} (\widehat{f}_T(x_i) - y_i)^2.$$

In our case, y_i is the original intensity measurement at pixel location x_i, and $\widehat{f}(T)(x_i)$ is the estimator. The estimator we employ is a constant in each cell of the partition, with the constant being set to the mean of all pixel values contained within the cell.

Choosing an appropriate complexity constant α is a challenging task. The intuitive goal is to construct a model that is as simple as possible but still provides a sufficiently accurate description of the data. Usually in regression the goal is to minimize the true risk, which is the mean squared error between the underlying MRI intensity function and our estimate. This true risk can not be directly evaluated because the underlying function is unknown. But if a good model for
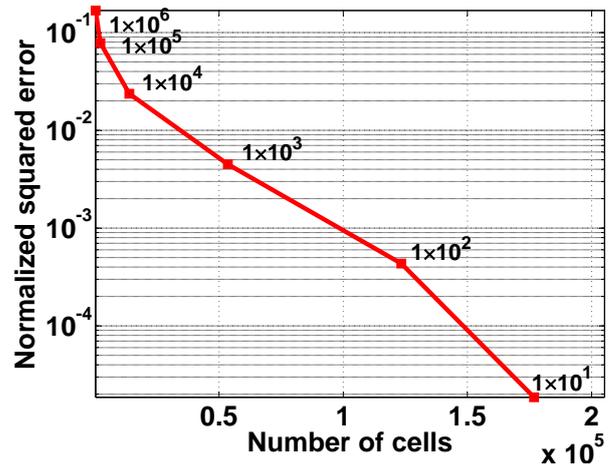

Fig. 2. Pruning the regression tree: the influence of the complexity penalty α on the number of leaf cells in the pruned tree and the corresponding normalized squared error (the sum of the squared differences between the original pixel intensities and the approximations by the cell means, normalized by the variance of the original pixel intensities). Each point in the plot corresponds to a specific pruned regression tree; the labels next to each point are the α-values. When the normalized squared error is 0, the number cells is 205,190, correponding to $\alpha = 0$.

the image noise is available, it is possible to identify a penalty that ensures the true risk is upper-bounded (see [6] for an example).

The primary goal in our work is not really to minimize the mean squared error. Our task is to ensure we build the simplest model that sufficiently captures the structural and dielectric complexity. We have achieved this is the electric fields measured in numerical simulations using the identified model are equivalent to those that arise from more complicated models. In our setting, equivalent means that the discrepancy between the signals (the model noise) is very small relative to the energy of signals we might strive to detect (responses from a tumour, for example). We explore this discrepancy in Section **, which allows us to develop guidelines indicating how many cells should be incorporated in the breast model. This, in turn, identifies a suitable range of α values.

Fig. 2 demonstrates the relationship between the number of cells included in the model and α. Cells are larger where tissue is homogeneous; near transitions from fatty to fibroglandular tissue, there are many small cells to capture accurately the structural complexity.

Stage 3: Assignment of Tissue Properties: Recently, Lazebnik *et al.* reported experimental characterization of the dielectric properties of normal breast tissue [7]. They classify the normal tissue into three groups based on their adipose content: Group 1 (0-30%), Group 2 (31-%85), and Group 3(86-100%) and they publish the data of the 25th, 50th, and 75th percentile of each group. Zastrow *et al.* use a 2-component Gaussian mixture model to approximate the distribution of the pixel intensity as shown in Fig. 3. They calculate the pixel intensities that correspond to the 25th, 50th, and 75th percentile of each Gaussian component.

Breast tissue has been commonly described by the Debye

model, which is

$$\epsilon_r = \epsilon_\infty + \frac{\Delta\epsilon}{1+\omega\tau} + \frac{\sigma_s}{j\omega\epsilon_o} \qquad (1)$$

where $\epsilon_\infty, \Delta\epsilon, \tau, \sigma_s$ are the tissue-dependent Debye parameters. Breast tissue can also be described by the simple dielectric material, which is

$$\epsilon_r = \epsilon' - j\epsilon'' = \epsilon' - j\frac{\sigma}{\omega\epsilon_o}. \qquad (2)$$

Three strategies to map pixel intensities to the tissue property are investigated in this paper using dielectric and Debye material. Zastrow *et al.* has proposed two strategies to map the pixel intensity to the material parameters. In the Debye map, they assume a fixed τ in (1) and interpolate Lazebnik's data of 25th, 50th and 75th percentile of Group 1 and Group 3 using one-pole Debye model. The pixel intensities found at the 25th, 50th and, 75th percentile of Group 1 and Group 3 and the Debye parameters $\epsilon_\infty, \Delta\epsilon, \sigma_s$ as a result of interpolation are one-to-one mapped. This defines 7 intervals, in which pixel intensities in between are linearly mapped to the Debye parameters.

In Zastrow's dielectric map, the ϵ' and σ of dielectric materials for the aforementioned 6 percentiles are obtained by evaluating the Debye model at 6 GHz. They then they are one-to-one mapped to the pixel intensities.

We develop a third tissue map strategy, similar to Zastrow's Debye map. We interpolate Lazebnik's data without any constraint on τ in the Debye model. Therefore, The pixel intensities are mapped to all four of Debye parameters, $\epsilon_\infty, \Delta\epsilon, \sigma_s, \tau$. We refer it as the piecewise-linear (PWL) map in this paper.

Since we have replaced the pixels with cells in the breast tissue region, the mapping is executed from the average pixel intensities of the cells to the material parameters. These cells are imported into SEMCAD and solids are generated by taking the intersection of the cells with the identified tissue boundary. SEMCAD's automatic gridder generates FDTD voxels to model the resultant model.

III. An Example

The MRI image we analyze is that of a 49-year patient. The tissue volume of the breast is approximately 70mm×70mm×60mm. The parameters of the MRI system were as follows: the static magnetic field strength was 1.5 T; a gradient-echo pulse sequence was used, with a pulse flip angle of 25° and a repetition time of 8.816 ms. The 68 MRI slices, each of 512×512 pixels, were acquired in the (x-y) planes perpendicular to the main body (z-) axis (axial run). This results in a spatial resolution of 0.7mm×0.7mm×2mm.

Fig. 4 presents the comparison of the tissue profile of the sagittal cross-section of the breast model shown in Fig. 1 with and without cell approximation.
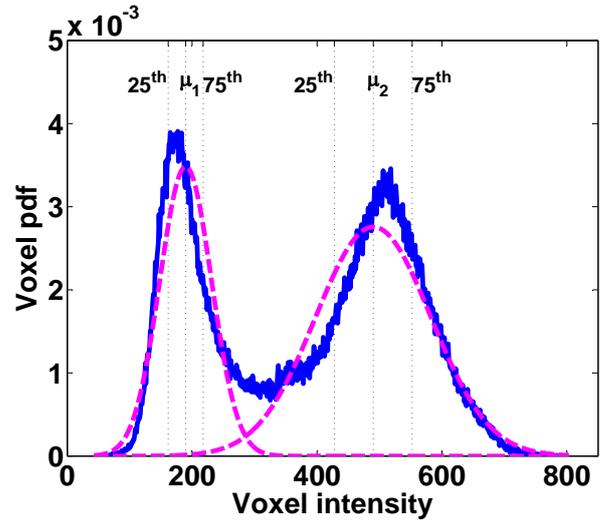

Fig. 3. Two-component Gaussian Mixture Model to fit the distribution of the pixel intensity of a patient's MRI images using Expectation Maximization algorithm. The the 25th, 50th, i.e. mean, and 75th percentile of each component are labeled.
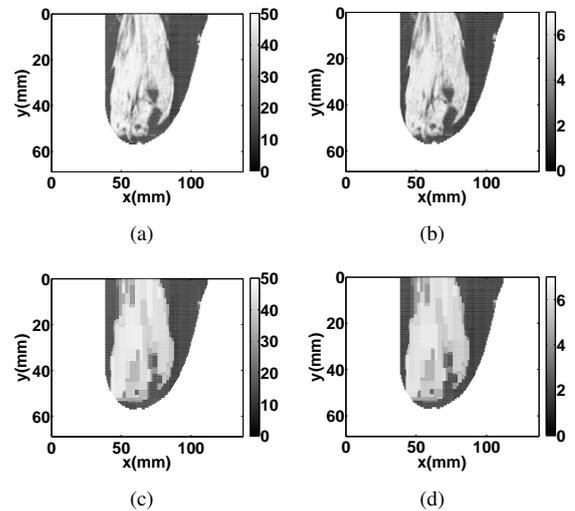

Fig. 4. Sagittal cross-section showing the tissue profile obtained by evaluating the Debye model at 6 GHz using Zastrow's Debye map. (a) and (b) are ϵ' and σ mapped directly from the pixels of MRI images. (c) and (d) show ϵ' and σ when we approximate the tissue with 2194 cells and $\alpha = 1e5$.

IV. Effects of Cell Padding and Dielectric vs. Debye Model

We first investigate the validity of using cells to approximate the heterogenous breast tissue. Fig. 5 and 5(b) illustrate the simulation scenario. We attach a skin layer, a fat layer, and a muscle layer to sandwitch the tissue, which is padded with different number of cells. The Debye parameters of skin, fat, and muscle are obtained from [8]. The breast is surround by four Travelling-Wave Tapered-Line Transmission-Line (TWTLTL) antennas proposed in [Kanj], which offers wideband behavior between 1 to 30 GHz. As labelled in Fig. 5(b), Antenna A transmits a gaussian modulated sinusoid with a bandwidth of 10 GHz centered at 6 GHz. The other three antennas receive and the four field sensors from A to D
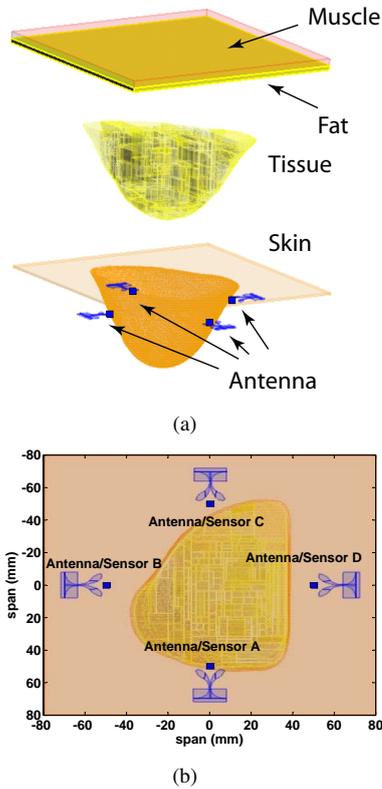
Fig. 5. The simulation scenario constructed inside SEMCAD. a) The top layer is 5-mm thick muscle followed by 3-mm thick fat tissue. The tissue is approximated by the cells illustrated in the previous section. The skin layer around the tissue is obtained from segmenting the MRI images with an average thickness of 1.5 mm. The skin below the fat slab is 1.5 mm thick. (b) The top view of the simulation scenario. The breast is surrounded by four antennas [Kanj] positioned 20 mm below the skin layer and at (0, -70), (-70,0),(0,70),(70,0). Four field sensors indicated by (□) are placed 20 mm in front of the antennas to record the electric field.

record the electric field.

The wave radiate from Antenna A is polarized in the x-direction. We extracted the E_x component of the field and computed the Mean Squared Error (MSE) according to

$$e_i^2 = \frac{\int (E_{x_i} - E_{x_{2194}})^2 dt}{\int E_{x_{2194}}^2 dt} \tag{3}$$

where E_{x_i} is the E_x component of Sensor C when the tissue is padded with i number of cells. $E_{x_{2194}}$, the reference, is when the tissue is padded with 2194 cells, the maximum cells generated for this experiment. Fig. 6 shows the MSE as we increase the number of cells. The seemly convergence behavior in the recorded signals with respect to the increase of the number of cells confirm the effectiveness of our regression-tree-based approximation with all three tissue mapping strategies. We also notice that the heterogeneity of dielectric material can be more effectively approximated than Debye material given the same number of cells.

We have investigated the effect of using dielectric materials and Debye materials in the breast model. In dielectric materials, the attenuation (δ) (dB/m) is

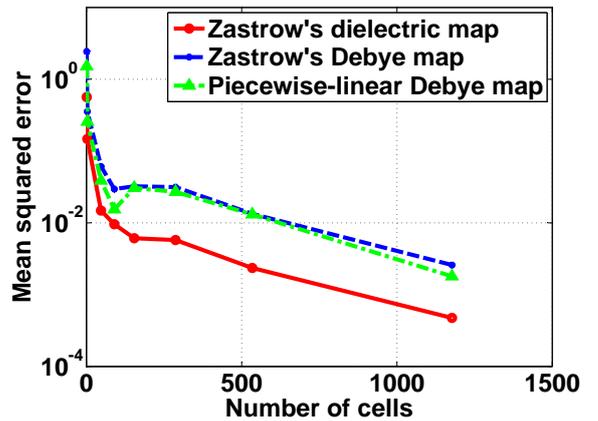$$\delta = 20 \log_{10} e^{\omega/c_o \mathcal{I}m\{\sqrt{\epsilon_r}\}} \tag{4}$$

Fig. 6. Mean-squared error of the E_x-field obtained at Sensor C with different number of padding cells with respect to the E_x-field at Sensor C when the number of padding cells is 2194.

where ϵ_r is given in (2) The lower bound of (4) can be shown as

$$\delta' = 20 \log_{10} e^{\frac{-1}{2c_o} \frac{\sigma}{\sqrt{\epsilon'\epsilon_o}}}. \tag{5}$$

when we assume $\frac{\sigma}{\omega\epsilon_o} < 0$. We notice that the attenuation is independent of frequency in (5). Similarly, the lower bound of antenuation of Debye material is

$$\delta' = 20 \log_{10} e^{\frac{-1}{2c_o} \frac{\omega^2 \tau \Delta\epsilon + \sigma_s/\epsilon_o}{\sqrt{\epsilon_\infty + \Delta\epsilon}}} \tag{6}$$

We notice that the attenntion increases as frequency increases. Fig. 7 shows the unit attenuation inside dielectric and Debye materials for the 25th, 50th, and 75th percentile of Group 1 and Group 3 tissue reported in [7] and their lower bound. It informs us that using Debye material in the breast model, the high-freuqency field components scattered by small tumors will be suppressed. We take the E_x component at Sensor A and Sensor C when the number of cells is 2194 and compute the attenuation along the path between these two points with three tissue mapping strategies. The result is shown in Fig. 8. We notice that, roughly above 6 GHz, using Zastrow's dielectric mapping indeed yields less attenuation than using the two Debye strageties. This implies that using Debye material will create a more difficult problem than using dielectric material since the high frequency components contain information on the geometry of small tumors. Fig. 8 also shows that using either Zastrow's Debye map or our PWL map gives very similar E_x component. Since the only difference is that Zastrow has assumed a constant τ, we conclude that this assumption has minimal effect on the field distortion inside the tissue.

V. Conclusions

This paper has presented the development of numerical breast models suitable for commerical FDTD simulators. The model retains the salient features of the tissue structure but minimizes complexity by dramatically reducing the number of tissue solids. This enables the simultions of complex
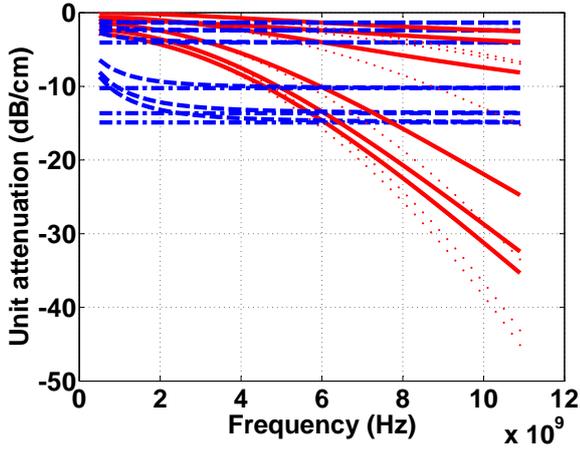
Fig. 7. The dash-dash blue curve shows the attenuation due to the dielectric constant evaluated at 6 GHz of the 6 Debye materials: 25th, 50th, and 75th percentile of Group 1 (0-30% adipose) and 25th, 50th, 75th percentile of Group 3 (85-100% adipose). The solid red curve shows the attenuation of the corresponding Debye materials. The dash-dot and dotted lines are the lower bounds of the attenuation dervied from Taylor's series expansion.
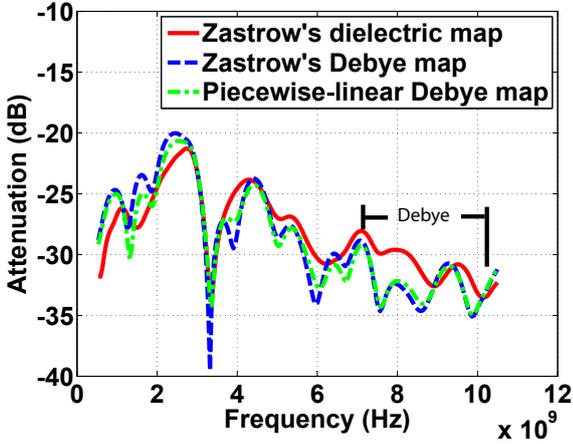

Fig. 8. Attenuation in the E_x-component between Sensor A and Sensor C E-field when the number of padding cells is 2194, under one dielectric material model and two debye material models.

antenna structures together with the anatomically realistic breast model. This paper has also reported the effect of using dielectric map and Debye map to model the heterogeity and dispersiveness of breast model. We find that Debye materials attenuate more in the high frequency component. This poses a more challenging problem for microwave radar imaging to detect geometrically small tumors.

ACKNOWLEDGMENT

APPENDIX A
BUILDING AND PRUNING A REGRESSION TREE

Consider a simple one-dimensional example, a row vector, $\mathbf{t} = [p_1...p_N]$ with mean, $\mu(\mathbf{t})$. The mean-squared error (MSE) of this vector is

$$\xi(\mathbf{t}) = \frac{1}{N} \sum_{i=1}^{N} (p_i - \mu(\mathbf{t}))^2. \qquad (7)$$

We define the left and right subvectors:$\mathbf{t}_L = [p_1...p_S]$ and $\mathbf{t}_R = [p_{S+1}...p_N]$ where S is the split index. We seek the best split S^* to maximize the change in the MSE, that is

$$\overset{\max}{S} \Delta \xi(S, \mathbf{t}) = \xi(\mathbf{t}) - \xi(\mathbf{t}_L) - \xi(\mathbf{t}_R). \qquad (8)$$

We continously apply (8) to $\mathbf{t}_L$ and replace $\mathbf{t}_L$ and $\mathbf{t}_R$ with the subvectors of $\mathbf{t}_L$. This process terminates when the subvector contains only equal-value elements. As a result, we have a binary tree, in which the nodes correspond to the vectors and the branches are denoted by the split index of the parent nodes. The sum of the leaf nodes of this tree represents the root node that corresponds to the input vector.

The resulting regression tree may have many leaf nodes. We want to systematically reduce the number of the leaf nodes. An error complexity measure $\xi_\alpha(\mathbf{t})$ is introduced as

$$\xi_\alpha(\mathbf{t}) = \xi \mathbf{t}_L + \xi \mathbf{t}_R + 2\alpha, \qquad (9)$$

where α is the complexity penalty that controls the trade-off between approximation error and the number of leaf nodes in the pruned tree. We prune the tree according to the rule: if $\xi_\alpha(\mathbf{t}) > \xi(\mathbf{t}) + \alpha$, prune the child nodes at the current node $\mathbf{t}$ and check the parent node of $\mathbf{t}$. Otherwise, stop pruning at the current node and check the next leaf node.

The level of pruning is controlled by α. If α is large, we tolerate the error due to the merge of two child nodes to their parents. There are less leaf nodes and the tree is greatly pruned. If we reduce α, we are less tolerate with the merge error. Few leaf nodes are eliminated and the tree is less pruned. In the two extreme cases, if $\alpha = 0$, no leaf nodes will be merged to their parents. If $\alpha > \xi(\mathbf{t}_{ROOT})$, the entire tree is pruned and left with only the root node.

REFERENCES

[1] E. Zastrow, S. Davis, M. Lazebnik, F. Kelcz, B. V. Veen, and S. Hagness, "Development of anatomically realistic numerical breast phantoms with accurate dielectric properties for modeling microwave interactions with the human breast," to appear, *IEEE Trans. Biomed. Eng.*, 2008.

[2] L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees*. Wadsworth International Group, 1984.

[3] Z. Harmany, R. Willett, A. Singh, and R. Nowak, "Controlling the error in fmri: Hypothesis testing or set estimation?" in *Proc. IEEE Int. Symp. Biomedical Imaging*, Paris, France, May 2008.

[4] D. Healy and J. Weaver, "Two applications of wavelet transforms in magnetic resonance imaging," *IEEE Trans. Info. Theory*, vol. 38, no. 2, pp. 840–860, Feb. 1992.

[5] R. Nowak, "Wavelet-based Rician noise removal for magnetic resonance imaging," *IEEE Trans. Image Processing*, vol. 8, no. 10, pp. 1408–1419, Oct. 1999.

[6] C. Scott, R. Willett, and R. Nowak, "CORT: Classification or regression trees," in *Proc. IEEE Int. Conf. Accoust., Speech, Signal Processing*, Hong Kong, Apr. 2003.

[7] M. Lazebnik, L. McCartney, D. Popovic, C. B. Watkins, M. J. Lindstrom, J. Harter, S. Sewall, A. Magliocco, J. Booske, M. Okoniewski, and S. C. Hagness, "A large-scale study of the ultrawideband microwave dielectric properties of normal breast tissue obtained from reduction surgeries," *Physics in Medicine and Biology*, vol. 52, no. 10, pp. 2637–2656, Oct. 2007.

[8] S. Gabriel, R. Lau, and C. Gabriel, "The dielectric properties of biological tissues. III. parametric models for the dielectric spectrum of tissues," *Physics in Medicine and Biology*, vol. 41, no. 11, pp. 2271–2293, Nov. 1996.