This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

Unsupervised Nosologic Imaging for Glioma Diagnosis

Yuqian Li*, *Student Member, IEEE*, Diana M. Sima, Sofie Van Cauter, Uwe Himmelreich, Anca R. Croitor Sava, Yiming Pi, *Senior Member, IEEE*, Yipeng Liu, and Sabine Van Huffel, *Fellow, IEEE*

***Abstract*—A novel approach to create nosologic images of the brain using magnetic resonance spectroscopic imaging (MRSI) data in an unsupervised way is presented. Different tissue patterns are identified from the MRSI data using non-negative matrix factorization (NMF) and are then coded as different primary colors (i.e. red, green and blue) in an RGB image, so that mixed tissue regions are automatically visualized as mixtures of primary colors. The approach is useful in assisting glioma diagnosis, where several tissue patterns such as normal, tumor and necrotic tissue can be present in the same voxel/spectrum. Error-maps based on linear least squares estimation are computed for each nosologic image to provide additional reliability information, which may help clinicians in decision making. Tests on in-vivo MRSI data show the potential of this new approach.**

***Index Terms*—Nosologic imaging, magnetic resonance spectroscopic imaging (MRSI), blind source separation (BSS), non-negative matrix factorization (NMF), hierarchical non-negative matrix factorization (hNMF).**

Manuscript received August 6, 2012. This work was supported by Research Council KUL: GOA MaNet, CoE EF/05/006 Optimization in Engineering (OPTEC), PFV/10/002 (OPTEC), IDO 08/013 Autism, several PhD/postdoc & fellow grants; Flemish Government: FWO: PhD/postdoc grants, projects: G.0427.10N (Integrated EEG-fMRI), G.0108.11 (Compressed Sensing), G.0869.12N (Tumor imaging), research communities (ICCoS, ANMMM); IWT: TBM070713-Accelero, TBM070706-IOTA3, TBM080658-MRI (EEG-fMRI), PhD Grants; IBBT. Belgian Federal Science Policy Office: IUAP P6/04 (DYSCO, `Dynamical systems, control and optimization', 2007-2011); ESA AO-PGPF-01, PRODEX (CardioControl) C4000103224; EU: RECAP 209G within INTERREG IVB NWE programme, EU HIP Trial FP7-HEALTH/ 2007-2013 (n° 260777); National Natural Science Foundation of China (61271287). Y. Li thanks the China Scholarship Council for the financial support. *Asterisk indicates corresponding author*.

*Y. Li is with the School of Electronic Engineering, University of Science and Technology of China, Chengdu 611731 China; Department of Electrical Engineering and IBBT- Future Health Department, KU Leuven, Leuven 3001, Belgium (e-mail: yuqianli@uestc.edu.cn)

D. M. Sima, A. R. Croitor Sava, Y. Liu and S. Van Huffel are with Department of Electrical Engineering and IBBT- Future Health Department, KU Leuven, Leuven 3001, Belgium (e-mail: Diana.Sima@esat.kuleuven.be; Anca.Croitor@esat.kuleuven.be; Yipeng.Liu@esat.kuleuven.be, Sabine.VanHuffel@esat.kuleuven.be).

S. Van Cauter is with Department of Radiology and Department of Imaging and Pathology, University Hospitals of Leuven, Leuven 3000, Belgium (e-mail: sofie.vancauter@uz.kuleuven.ac.be)

U. Himmelreich is with Department of Imaging and Pathology, KU Leuven, Leuven 3000, Belgium (email: Uwe.Himmelreich@med.kuleuven.be)

Y. Pi is with the School of Electronic Engineering, University of Science and Technology of China, Chengdu 611731 China (e-mail: ympi@uestc.edu.cn).

I. Introduction

ACCURATE diagnosis of brain tumors is of utmost importance in planning therapy and guiding surgery. Magnetic resonance spectroscopy imaging (MRSI) is an advanced non-invasive imaging technique that complements conventional magnetic resonance imaging (MRI) by providing multi-voxel spectra of specific biochemical information relating to the tumor type and grade. Recent studies have utilized MRSI or combined MRI with MRSI to create nosologic images [1] [2] [3] [4] [5]. These methods aim at providing tumor type and grade in a single image, where different tissue types are encoded with different colors. The previous work on nosologic imaging has been based on supervised classification methods. In practice, obtaining large data sets for training classifiers is not always feasible. Moreover, brain tumors such as gliomas, which are the most common primary tumors in adults, can be heterogeneous and infiltrative, especially for the higher grade cases. Thus, MRSI data may contain voxels reflecting contributions from several tissues, mixed in arbitrary percentages. Nosologic images where "mixed tissue" is considered as a separate class [4] ignore the fact that the percentages of each tissue pattern in each "mixed" voxel may vary significantly. In general, visualizing clear contours (e.g., tumor, normal tissue, and, possibly, mixed tissue) is not realistic for heterogeneous brain tumors such as glioma.

With this letter, we present a fully unsupervised method based on blind source separation (specifically, on non-negative matrix factorization (NMF) [6]) to automatically create nosologic images of gliomas. Sajda *et al*. [7], Su *et al*. [8] and Du *et al*. [9] used NMF to differentiate brain tumor tissue from normal tissue without the need of model spectra. Li *et al*. [10] developed a hierarchical tissue pattern differentiation method using NMF (hNMF) and showed that this method is able to differentiate three tissue patterns present in glioblastoma multiforme (GBM, or grade IV glioma). Here we demonstrate how to employ NMF to create nosologic images of the whole excitation volume in an unsupervised way and we address the issue of image reliability by displaying "error-maps" next to each nosologic image. As opposed to the existing nosologic imaging methods [1]–[5], where one color represents one tumor class, our unsupervised nosologic imaging method provides mixtures of primary colors (i.e., red, green and blue) between different tissue patterns; hence the information of mixed tissue,

Copyright (c) 2011 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

i.e., the information about tumor heterogeneity, is preserved.

We illustrate results using MRSI data from low and high grade glioma patients. The accuracy of recovered tissue-specific spectral sources and their corresponding spatial distributions are validated using expert labeling.

II. Materials and Methodology

A. Patients and Data.

MRSI data from 12 glioma patients were acquired in the University Hospital Leuven (UZ Leuven), Belgium. We considered 6 patients with grade II glioma, which typically presents two tissue patterns (normal and tumor, without the presence of necrosis), and 6 patients with grade IV glioma, i.e., GBM, which typically presents three tissue patterns (normal, actively proliferating tumor and necrosis). All the cases included in this paper had histopathological diagnosis according to the WHO 2007 classification [11]. The data were acquired on a 3T MR scanner. The MRSI protocol had the following imaging parameters: point-resolved spectroscopy (PRESS) [12] was used as the volume selection technique, TR/TE: 2000/35 ms, FOV: 16cm*16cm, volume of interest (VOI): 8cm*8cm, slice thickness: 1cm, reconstruction voxel size: 0.5cm*0.5 cm, samples: 2048. Standard anatomical MR images have also been acquired. The institutional review board approved this study. Written informed consent was obtained from every patient before their participation in the study. Data preprocessing was done as in our previous paper [10] using the in-house software SPID [13].

B. Tissue differentiation

For any MRSI data set, we define a matrix X to contain spectra as column vectors, one voxel per column. Each column of this matrix can be approximated as a linear combination of r constituent spectra (i.e., "spectral sources") of specific tissue patterns [8], leading to the factorization:

$$X_{m\times n} \approx W_{m\times r} H_{r\times n} \quad \textit{subject to } W, H \geq 0 \tag{1}$$

Each column of W represents a spectral source. Each row of H contains the linear combination weights. Spatial distribution information of each tissue pattern can be provided by reshaping each row of H back to the original spatial dimensions.

Conventional NMF [6] is typically able to distinguish two tissue patterns (i.e., normal and abnormal) by setting the number of spectral sources r equal to 2. This is sufficient, for instance, in the case of grade II glioma. However, conventional NMF is not always accurate to recover more than two biologically meaningful spectral sources, which is for instance necessary in the case of the three tissue patterns of GBM tumors (i.e., normal, tumor and necrosis). In this situation, a hierarchical NMF method (hNMF) [10] has to be applied. hNMF first separates the brain tissue into normal and abnormal, then by applying an optimized threshold, the abnormal tissue is further separated into actively proliferating tumor and necrosis. In this way, the three most meaningful spectral sources for GBMs are recovered, as well as their spatial distribution information. The decision whether two or three spectral sources are more appropriate for each MRSI data set can be based on the absence or presence of necrosis, which is a hallmark of high-grade glioma (GBM). This can be automatically identified by specific spectral characteristics in MRSI data (i.e., integrating the region encompassing the lipid peaks at 0.9 and 1.3 ppm and comparing these values against the integrated values for the NAA at 2.01 ppm and Cho peaks at 3.22 ppm).

It is worth to note that spectra that are very noisy (i.e. spectra for which the peaks of interest are embedded in noise) or distorted might hinder a successful source separation by NMF. Therefore, all steps involving (h)NMF are typically applied on spectra of sufficient quality from smaller regions of interest (ROI) within the PRESS excitation volume. To ensure sufficient spectral quality, we followed an expert-labeling protocol as in our previous paper [10], although a practical advice for MRSI data acquired under similar conditions to ours is to exclude 2-3 rows/columns of voxels for all borders of the PRESS excitation volume, in order to avoid the most common artifacts: chemical shift displacement effect and lipid contamination.

C. Nosologic Imaging

With (h)NMF, the most relevant tissue-specific spectral sources are recovered as columns of W, as well as the spatial distribution of each tissue pattern, as rows of H. However, the spatial distribution in H only provides localization information for the voxels within the selected ROI. To estimate the contribution of each tissue pattern in the whole PRESS excitation volume (thus even outside of the selected ROI), non-negative least squares (NNLS) [14] is applied on spectra from the whole PRESS excitation volume using the already recovered tissue-specific spectral sources in W. Therefore, an additional non-negative least squares problem is solved for each voxel x_i:

$$\min \| x_i - W h_i \| \quad over\ h_i \geq 0, \quad h_i \in \Re^r \tag{2}$$

By reshaping the collection of all the h_i back into the original matrix size of the PRESS excitation volume， new spatial distributions for all the r tissue patterns are obtained, each dimension representing one tissue pattern. We define the new spatial distribution of each tissue pattern (normalized between 0 and 1) as a color channel in an RGB image by encoding the spatial distribution for necrosis as red channel (if not present, this channel is set to zero), the one for tumor as green channel and the one for normal tissue as blue channel. The regions where the red or green color gets darker are the most aggressive regions for the respective data set. In this way, the spatial distribution maps of different tissue patterns are incorporated into a single nosologic image where tissue-specific regions are interpreted by different colors: blue for normal, green for tumor, red for necrosis, and mixtures of these primary colors for mixed tissues, which could represent tumor

Copyright (c) 2011 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

infiltration. The regions of non-informative signals (i.e., for which the weights corresponding to all tissue patterns are close to zero) appear in black in an RGB color scheme.

The resolution of the spatial distribution is in accordance with the reconstruction voxels size (i.e., 0.5cm*0.5cm in the considered data sets). In order to reach the resolution of the MR images, the RGB image is interpolated using bicubic spline interpolation [15] by a factor of 20, thus reaching the pixel size of 0.025cm*0.025cm.

D. Reliability Investigation Using Standard Errors

For the least squares problem in (2), assuming that W is correctly estimated, we can approximately estimate lower bounds for the standard errors of the linear combination weights h_i at the ith voxel as the diagonal elements of the inverse Fisher information matrix $\sigma_i^2(W^TW)^{-1}$. The residual error $\| x_i - W h_i \|$ divided by the degrees of freedom can be used to estimate σ_i^2, which is an estimate for the noise variance for the spectrum in voxel i. The lower bounds on the standard errors for h_i at voxel i are thus given by:

$$s_i = \sqrt{\sigma_i^{\,2} diag((W^TW)^{-1})} = \sqrt{\frac{\left\| x_i - Wh_i \right\|^2}{m-r} diag((W^TW)^{-1})} \quad (3)$$

where m-r gives the statistical degrees of freedom. By computing s_i (which is an r-dimensional vector) for all voxels i in the PRESS excitation volume, we obtain lower bounds for the standard error estimates for all r tissue patterns at all locations. Then, the standard errors are reshaped and interpolated as "error-maps" of the same size as the nosologic image, and can be interpreted as reliability maps for the nosologic image. As it can be easily derived from formula (3), the spatial error-maps are a function of σ_i^2 and thus are the same for each tissue pattern, but they have different scales, which are given by the diagonal elements of $(W^TW)^{-1}$. This implies that a single error-map can be shown for each nosologic image, but its color range will encode different numerical scales for the r tissue patterns. We report the standard errors as percentages by normalizing the s_i values to the same scale as the corresponding spatial distribution for each tissue pattern. Note that the region of high standard errors does not reflect the failure of the unsupervised nosologic imaging method. They only mark the regions where the nosologic image is more prone to uncertainties due to higher variations in the spectral quality.

III. Experiments and Results

A. Results of Nosologic Imaging

For each patient, the nosologic images (shown in the second column for each patient in Fig. 1) clearly show the relevant tissue patterns and their locations: blue for normal tissue, green for actively proliferating tumor and red for necrosis. Regions of non-informative spectra are present at the borders of the PRESS excitation volume and are rendered in black. Mixtures of primary colors are also present to show mixed tissues, possibly due to tumor infiltration or onset of necrosis in the tumor region, partial volume effect or point spread function. The error-maps are given next to each nosologic image in Fig. 1, where red regions represent lower reliability and blue regions represent higher reliability.

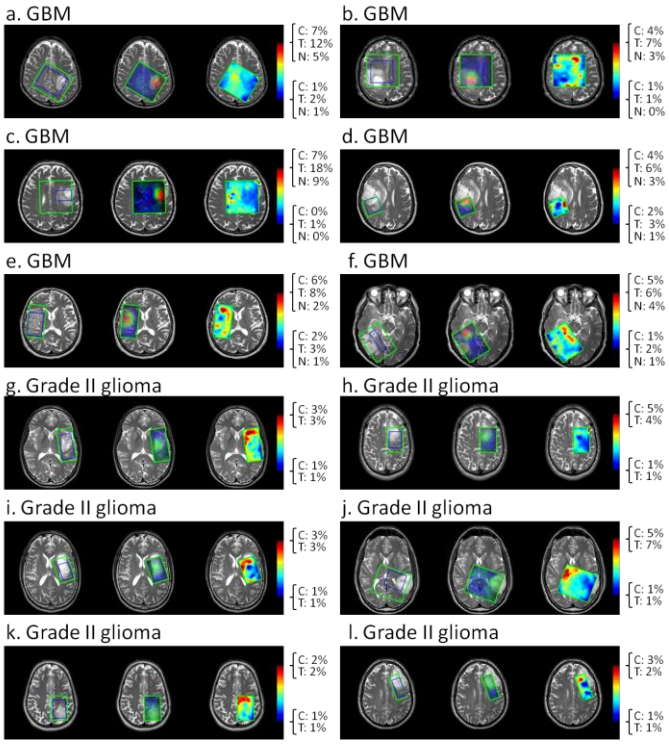


Fig. 1. Nosologic images for 12 data sets (6 from GBM patients and 6 from grade II glioma patients). a-l: First image: the anatomic T2-weighted MR image with the green box showing the PRESS excitation volume and the blue box showing the selected region of interest (ROI). Second image: nosologic images in the PRESS excitation volume, overlaid with the anatomic MR images. The red color indicates the presence of necrosis, green shows active tumor region and blue shows the normal region. Regions that contain non-informative signals are shown in black. Third image: the standard errors for the estimated nosologic images. On the right side of the colorbars, the highest standard errors and lowest standard errors for each tissue ("C" for normal, "T" for tumor and "N" for necrosis) are presented.

B. Results Validation

For result validation, both the nosologic images and recovered spectral sources are evaluated using information from expert labeling. As shown in Fig. 2, the nosologic images are overlapped on the anatomic T2-weighted MR images as a transparent layer. Then the color-coded expert labeling given in the second column also as a transparent layer for each patient is used for visual comparison. We can see that the created nosologic images are in high correspondence with the MR images and the expert labeling. Then the recovered spectral sources are compared with the reference spectra which are calculated as the average of all the labeled tissue-specific spectra for each tissue pattern, both visually (Fig. 2) and numerically using the correlation coefficient between recovered spectral source and reference spectra (Table 1). Patients a.-l. in Fig. 2 and Table 1 correspond to the patients a.-l. in Fig. 1. We

Copyright (c) 2011 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

can observe that all the recovered spectral sources are very similar with the reference spectra and the correlation coefficients are close to the upper bound 1.

IV. Conclusion

Unsupervised nosologic imaging provides a novel way for MRSI data interpretation without the need of large training data sets. The created nosologic image is obtained on the whole PRESS excitation volume and mixed tissues in heterogeneous tumors can be shown as mixtures of primary colors. Furthermore, standard error maps provide extra information about the reliability of the nosologic images. In theory, NMF as a blind source separation tool is supposed to be able to disentangle any mixed signals. Hence, the proposed method could also be feasible for homogeneous tumors. Moreover, homogeneous tumors, which do not contain mixed tissues, should be easier to analyze with this approach than gliomas. Since the proposed method provides a direct and effective first glance at the information contained in MRSI signals, we believe that the proposed method can become a promising non-invasive tool to assist brain tumor diagnosis.

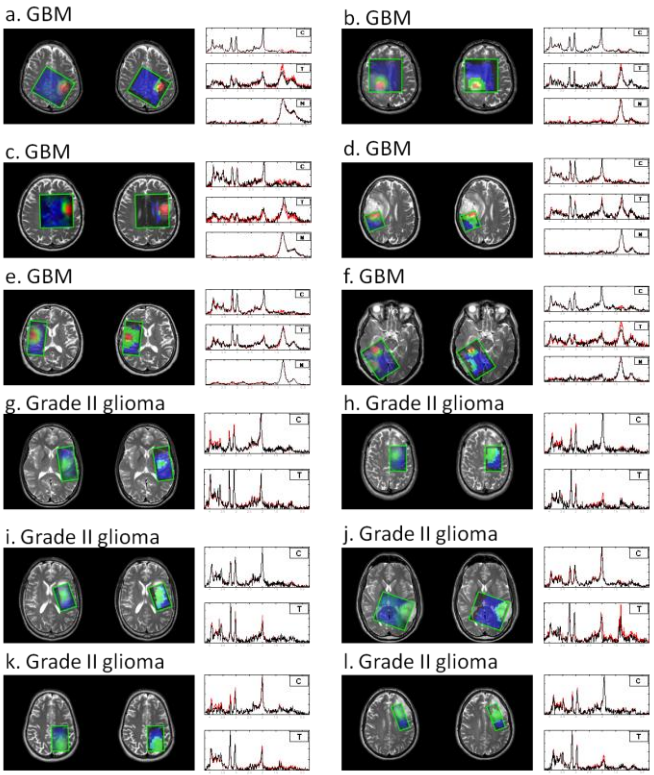


Fig. 2. Validation against expert labeling for 12 data sets (6 from GBM patients and 6 from grade II glioma patients). For each patient, the MR images overlaid with the nosologic images are shown in the first column. The expert labeling translated into colormaps are given in the second column, where red indicates necrosis, yellow indicates necrosis/tumor, green indicates tumor, cyan indicates normal/tumor, blue indicates normal and black indicates spectra of bad quality. In the third column, the recovered spectral sources for each tissue pattern (in black) are overlaid on the reference spectra (red). "C" stands for normal, "T" for actively proliferating tumor and "N" for necrosis.

TABLE I
Correlation Coefficients (bounded between -1 and 1)

Patients	normal	tumor	necrosis
a. GBM	0.99	0.95	1.00
b. GBM	1.00	0.99	0.97
c. GBM	0.96	0.90	1.00
d. GBM	0.98	0.96	1.00
e. GBM	0.96	0.97	1.00
f. GBM	0.99	0.92	0.99
g. grade II glioma	0.92	0.98	
h. grade II glioma	0.98	0.98	
i. grade II glioma	0.99	0.99	
j. grade II glioma	0.96	0.92	
k. grade II glioma	0.94	0.98	
l. grade II glioma	0.99	0.99	

References

[1] F. Szabo de Edelenyi, C. Rubin, F. Esteve, S. Grand, M. Decorps, V Lefournier, J. F. Le Bas, C. Remy, "A new approach for analyzing proton magnetic resonance spectroscopic images of brain tumors: nosologic images," *Nat. Med.(N.Y.)*, vol. 6, pp. 1287-1289, 2000.

[2] A. W. Simonetti, W. J. Melssen, M. van der Graaf, A. Heerschap, L. M. C. Buydens, "A chemometric approach for brain tumor classification using magnetic resonance imaging and spectroscopy," *Anal. Chem.*, vol. 75, no. 20, pp. 5352-5361, 2003.

[3] T. Laudadio, M. C. Martinez-Bisbal, B. Celda, S. Van Huffel, "Fast nosological imaging using canonical correlation analysis of brain data obtained by two-dimensional turbo spectroscopic imaging," *NMR Biomed.*, vol. 21, no. 4, pp. 311-321, 2008.

[4] M. De Vos, T. Laudadio, A. W. Simonetti, A. Heerschap, S. Van Huffel, "Fast nosologic imaging of the brain," *J. Magn. Reson.*, vol. 184, no. 2, pp. 292-301, Jan. 2007.

[5] J. Luts, T. Laudadio, A. J. Idema, A. W. Simonetti, A. Heerschap, D. Vandermeulen, J. A. K. Suykens, S. Van Huffel, "Nosologic imaging of the brain: segmentation and classification using MRI and MRSI," *NMR Biomed.*; vol. 22, no. 4, pp.374-902009.

[6] D. D. Lee, H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788-791, 1999.

[7] P. Sajda, S. Du, T. R. Brown, R. Stoyanova, D. C. Shungu, X. Mao, L. C. Parra, "Nonnegative matrix factorization for rapid recovery of constituent spectra in magnetic resonance chemical shift imaging of the brain," *IEEE Trans. Med. Imag.*, vol.23, pp. 1453-1465, 2004;

[8] Y. Su, S. B. Thakur, S. Karimi, S. Du, P. Sajda, W. Huang, L. C. Parra, "Spectrum separation resolves partial-volume effect of MRSI as demonstrated on brain tumor scans," *NMR Biomed.*, vol. 21, pp. 1030-1042, 2008.

[9] S. Du, X. Mao, P. Sajda, D. C. Shungu, "Automated tissue segmentation and blind recovery of 1H MRS imaging spectral patterns of normal and diseased human brain," *NMR Biomed.*, vol. 21, pp. 33-41, 2008.

[10] Y. Li, D. M. Sima, S. Van Cauter, A. Croitor Sava, U. Himmelreich, Y. Pi, S. Van Huffel, "Hierarchical non-negative matrix factorization (hNMF): a tissue pattern differentiation method for glioblastoma multiforme diagnosis using MRSI". *NMR Biomed*., Epub ahead of print.

[11] D. N. Louis, H. Ohgaki, O. D. Wiestler, W. K. Cavenee, P. C. Burger, A. Jouvet, B. W. Scheithauer, P. Kleihues, "The 2007 WHO classification of tumors of the central nervous system," *Acta. Neuropathol.*, vol. 114, pp. 97-109, 2007.

[12] P. A. Bottomley, "Spatial localization in NMR-spectroscopy in vivo,". *Ann. N. Y. Acad. Sci.*, vol. 508, pp. 333–348, 1987.

[13] J. B. Poullet, "Quantification and classification of magnetic resonance spectroscopic data for brain tumor diagnosis," Ph.D. dissertation, Dept. Elect. Eng., KULeuven., Leuven, Belgium, 2008. Available: http://homes.esat.kuleuven.be/~biomed/software.php#SpidGUI

[14] C. L. Lawson, R. J. Hanson, *Solving Least-Squares Problems*. Prentice-Hall, 1974, ch. 23, pp. 161.

[15] C. De Boor, *A Practical Guide to Splines*. New York: Springer-Verlag, 1978.

Copyright (c) 2011 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.