

Automated Feature Extraction from Power System Transients Using Wavelet Transform

X. Xu, *Student Member, IEEE*, and M. Kezunovic, *Fellow, IEEE*

Abstract— Feature extraction plays an important role for event detection and classification. In this paper, an automated feature extraction method based on wavelet packet transform and best basis algorithm has been proposed. The problem of shift invariance has also been discussed.

Index Terms—Power systems, Wavelet transforms, Feature extraction, Power system transients

I. INTRODUCTION

THE recorded transient waveforms in power system may contain unique signatures revealing the causes of the corresponding transient events. To automatically analyze those recorded waveforms, an important step is to find out those signatures. Due to the non-stationary property of the transient signals, traditional analysis methods such as Fourier Transform are not very suitable for this task. Unlike Fourier Transform, Wavelet Transforms use fast decaying kernel functions, which may better represent and analyze the transient signals.

Wavelet transforms have been applied in solving several problems in power systems, for example, signal analysis [1-7], data compression [8], and numeric solution of differential equations [9]. Among them, the data compression application of wavelet transforms has become very mature and some industrial applications have emerged, e.g. the new JPEG2000 standard for image compression [11]. On the contrary, there are still several open problems for signal analysis applications. In most of the signal analysis applications [1-7], the wavelet transforms are limited to show several fancy pictures and quantitative analysis is seldom, which greatly hinders its application in areas such as automated event analysis.

Compared with Fourier Transform, wavelet coefficients contain time information, which restricts utilizing them as features for event analysis directly. One method to remove the time-related information is using only the energy information at each level, which in essence is equivalent to using wavelet transforms as band pass filters. For example, Gaouda [3] used the energy vector calculated from the decomposition levels as the feature. The similar approach has been utilized in [4], where the frequency spectrum has been cut into six ranges: 0-16.6kHz, 16.7-31.3kHz, 31.4-46.9kHz, 50-65.6kHz, 66.7-

81.3kHz and 81.4-100kHz. The energies at those six frequency ranges are used as the features to train neural networks for phase selection purpose. The main rationale behind the method is that the beating phenomena of the frequency components are larger and higher in the healthy phase than in the faulted phase.

In this paper, an automated feature extraction method that can select the most relevant features from the training data is presented. The method is based on the wavelet packet transform (WPT) and the best-basis algorithm. During the application of the best-basis algorithm, it is important to apply shift-invariant wavelet transforms. Otherwise the results will be too sensitive to time locations.

The paper is organized as follows: the wavelet transform, wavelet packet transform and best-basis algorithm will be briefly reviewed at first, followed by a section discussing the proposed feature extraction method. At last, the problem of shift invariance transform will be discussed.

II. WAVELET PACKET TRANSFORMS AND BEST-BASIS ALGORITHM

The Continuous Wavelet Transform (CWT) of a function $f(t)$ with respect to a wavelet $\psi(t)$ is given by:

$$W_{\psi} f(b, a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \overline{\psi\left(\frac{t-b}{a}\right)} dt \quad (1)$$

Where $\psi(t)$ is the mother wavelet, a is the scaling (or dilation) parameter, and b is the translation parameter.

The mother wavelet $\psi(t)$ may have different forms. And wavelet analysis is a measure of similarity between the basis functions (wavelets) and the original function. The calculated coefficients indicate how close the function is to the daughter wavelet at the particular scale.

CWT involves a fair amount of calculation, and it generates a lot of redundant data. If we choose scales and positions based on powers of two (*dyadic* scales and positions), we obtain the so-called Discrete Wavelet Transform (DWT) as given in equation (2):

$$w(j, k) = \frac{1}{\sqrt{2^j}} \sum_n f(n) \psi(2^{-j} n - k) \quad (2)$$

There is an efficient algorithm to perform the transform using two-channel subband filter banks as shown in Fig. 1.

DWT only partitions the frequency axis finely toward the low frequency, and Wavelet Packet Transform (WPT) is a generalized version, which also decomposes the high

X. Xu is with Texas A&M University, College Station, TX 77843
M. Kezunovic is with Texas A&M University, College Station, TX 77843

frequency bands that are kept intact in DWT.

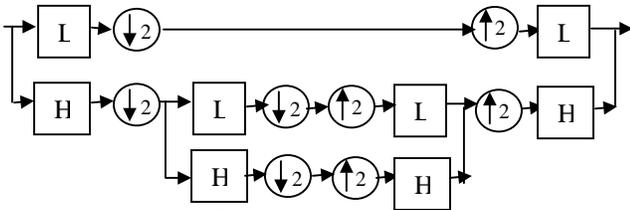


Fig.1. Multistep subband Decomposition and Reconstruction

When applying wavelet transform to analyze signals, the following steps are involved. At first, a proper family of wavelet and the appropriate filters will be selected for a specific type of events. After that, the suitable decomposition levels should be decided. At last, the most relevant wavelet coefficients must be chosen as the features. The *best-basis paradigm* has been proposed in [10] to solve the above problems. This paradigm consists of three main steps:

- Select a "best" basis for the problem at hand from a library of bases
- Sort the coordinates (features) by "importance" for the problem at hand and discard "unimportant" coordinates
- Use the surviving coordinates to solve the problem at hand

III. FEATURE EXTRACTION USING BEST-BASIS PARADIGM

Although several kinds of wavelets have been used in power system event analysis [1-7], how to select an optimal wavelet is still an open question. In this section, we apply the *best-basis paradigm* to search the most suitable wavelet transform and automatically select the most relevant features using a finite-size wavelet library.

A. Entropy-based Best Basis

The primary task of feature extraction is to only select the most significant components and thus reduce the length of the feature vector. A natural thought is to maximize the energy concentration of the coefficients and only choose the ones with big energies. The *Shannon entropy* of a given signal x_i , $i = 1, \dots, n$ is defined by:

$$H(x) = -\sum_{i=1}^n x_i^2 \log_2(x_i^2) \quad (3)$$

As we can see from the definition, if two signals have the same energy but different entropy, the one with smaller entropy will have its energy more concentrated at fewer coefficients. Therefore, it is reasonable to measure the suitability of a basis using entropy.

For signals with different energies, we can normalize them by dividing them with their squared sums respectively. For example, for signal x , its normalized version is

$$x_{Nk} = \frac{x_k}{\sqrt{\sum_{i=1}^n x_i^2}} \quad k = 1, \dots, n \quad (4)$$

Adopting entropy as a performance measure, for a given type of events, we have the following algorithm to select a

best basis:

Best basis algorithm 1:

1. Select a wavelet transform from a given library of wavelets with different mother wavelets and different filter lengths
2. Select a signal from a set of training signals of the given event type
3. Decompose the signal using the selected wavelet until the entropy no longer decreases
4. Calculate the entropy
5. Repeat steps 1 - 4 for all the signals used for training
6. Sum all the entropies obtained in step 4
7. Repeat steps 1 - 6 for all the wavelets in the library
8. Select the wavelet which gives the smallest total entropy

The wavelet library we adopted includes all orthogonal wavelets, i.e. Daubechies, Coiflets, and Symlets with different length of filters. After applying the above procedures on a fault transient, the result is shown in Fig. 2.

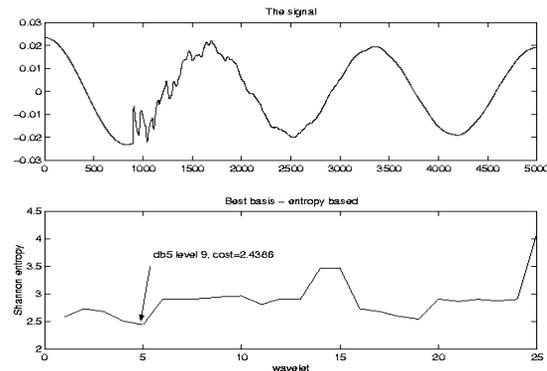


Fig. 2. Best basis - entropy based

We can repeat the above procedures for all the event types of interest and find out the most appropriate wavelet for analyzing each type of events. This is especially useful for compression applications when the waveform is known. After applying the above algorithm on some transient waveforms, the best wavelet for analyzing different types of events can be obtained. The results are summarized in Table I.

TABLE I
BEST WAVELET FOR ANALYZING T TRANSIENTS

Event Type	Number of Cases	Best Wavelet
Phase Fault	120	db4
Ground Fault	50	db4
Capacitor switching	50	db5
Motor starting	60	sym4
Line switching	60	db3
Transformer energizing	40	db1
Overall	440	db6

The above conclusion can be used for compressing and denosing, when the types of events are known. Using the wavelets and decomposition levels listed in Table II makes the resultant wavelet coefficients have the minimum entropy. Of

course, the conclusion is based on a relatively small set of simulation results. More exclusive conclusions can be obtained by applying the algorithm on more simulation data and field recorded waveforms.

When the purpose is classification, the type of an event is unknown at first. In this case, we can modify the above procedures a little by using a set of training waveforms with all types of events of interest present. The wavelet which gives the smallest overall entropy will be selected as the feature extractor, since it can analyze all types of events relatively well.

After calculating the overall entropies for all the wavelets, the db6 wavelet is found as the best one to analyze the transients of fault, capacitor switching, motor starting, line switching, and transformer energizing. The result is shown in Fig. 3.

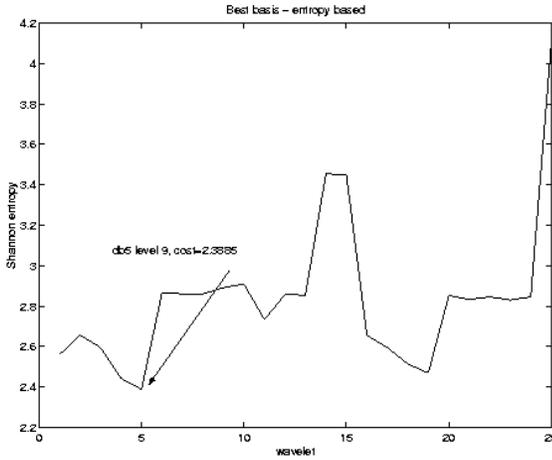


Fig. 3. Best basis for classification - entropy based

B. Best Basis for Classification Purpose

The criteria used to measure the suitability of a wavelet is purpose dependent. Although entropy is a good measure of information cost, thus suitable for measuring compression effect. It may not reflect the classification ability of a wavelet very well. For classification, a basis through which we can maximally separate classes as clouds of points in the n -dimensional space is the best choice. Therefore, the separability "distance" among classes (such as relative entropy) should be used to measure the efficiency of the coordinate system.

The commonly used discriminant measures are: *relative entropy* (or *I-divergence*), *J-divergence*, and *distance*. Given two nonnegative vectors p and q which satisfy $\sum_i p_i = \sum_i q_i = 1$, we can define the relative entropy $I(p, q)$ as:

$$I(p, q) = \sum_{i=1}^n p_i \log\left(\frac{p_i}{q_i}\right) \quad (5)$$

It can be proved that $I(p, q) \geq 0$ and the equality holds if and only if $p = q$. But the relative entropy is not a metric since it is not symmetric and does not satisfy the triangle inequality. This shortcoming can be easily overcome by using *J-divergence*:

$$J(p, q) = \sum_{i=1}^n p_i \log\left(\frac{p_i}{q_i}\right) + \sum_{i=1}^n q_i \log\left(\frac{q_i}{p_i}\right) \quad (6)$$

Also, we can use the distance between the two vectors as the measure:

$$D(p, q) = \sum_{i=1}^n (p_i - q_i)^2 \quad (7)$$

The additive property of those above measures make a fast algorithm to find the best basis possible [10].

To maximally separate the features of several given types of events, the following algorithm can be adopted.

Best basis algorithm 2:

1. Select a wavelet transform from a given library of wavelets with different mother wavelets and different filter lengths
2. Select a signal from a set of training signals
3. Decompose the signal using the selected wavelet transform
4. Repeat steps 2 - 3 for all the signals used for training
5. Compute discriminant measures for each node
6. Select the best basis using the calculated measures
7. Repeat steps 2 - 6 for all the wavelet transforms
8. Select the wavelet which gives the largest measure

Applying algorithm 2 on the cases used in the last section, we find the db5 is the best wavelet as shown in Fig. 4. And corresponding optimal terminal nodes of the wavelet decomposition tree are shown in Fig. 5.

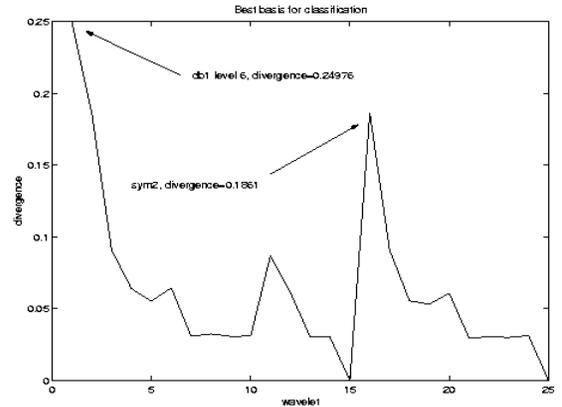
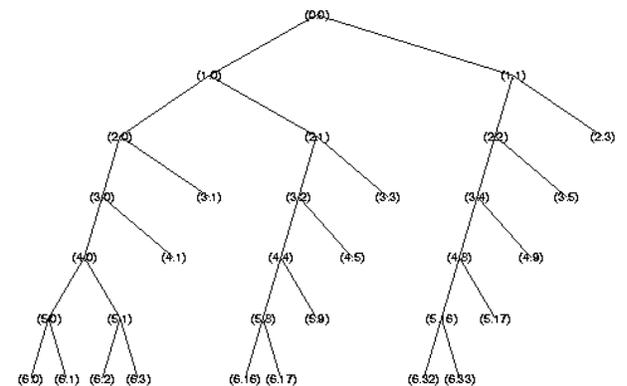


Fig. 4. Best Basis for Classification - discrimination ability based



C. Feature Extraction

The best-basis algorithms give us the optimal decomposition trees. And the wavelet coefficients in each terminal nodes of the trees can be used as the features. But utilizing all the wavelet coefficients is not applicable, since the feature vector will become too long. One important merit of wavelet transform is that only a few of the coefficients are significant and others are almost negligible. The applications of wavelet transforms on compressing and denoising are mainly taking advantage of this merit. Therefore, we need a method to select only those coefficients which contribute most and discard others.

A wavelet coefficient is determined by four parameters: (1) the wavelet, (2) the decomposition level, (3) the node in the decomposition tree, and (4) the position inside the node. The first two parameters determine the full size decomposition tree, and the full size tree can be "trimmed" to an optimal one using the best-basis algorithms mentioned in the last two sections. Since not all the nodes have the same contribution, we can order the terminal nodes by their significance. The coefficients inside a node can also be ordered by their importance. After the ordering processes, we can select several top nodes at first, and then select several top coefficients inside those top nodes as the features.

In summary, the steps for automated feature extraction using wavelet transform are:

1. Perform the Best basis algorithm 2
2. Order the terminal nodes by their discrimination ability
3. Order the coefficients in each node by their discrimination ability
4. Select several N_{nd} top terminal nodes
5. Select several N_{cf} top coefficients from each selected top nodes
6. Combine the selected coefficients to form the feature vector with length of $N_{nd} * N_{cf}$

To decide the appropriate values for the number of selected top nodes N_{nd} and the number of selected coefficients N_{cf} , we need to consider the lengths and types of the signals and the classifier we are using. After some experiments, we choose $N_{nd} = 4$, and $N_{cf} = 5$. The feature extracted using this method for a capacitor switching transient is shown in Fig. 6.

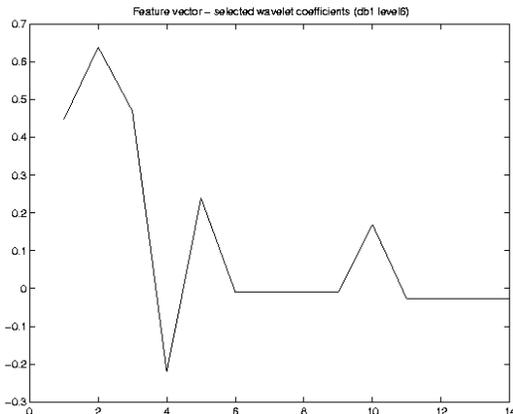


Fig. 6. Feature vector extracted using best basis algorithm

D. Shift Invariance Transform

There are some caveats worth pointing out for the above feature extraction approach. First, the wavelet transforms have to be orthogonal, a requirement of the best basis algorithm. Second, the wavelet transforms should be shift-invariant, which ensures the selected optimal nodes remain unchanged regardless of the shifts of the signals. Otherwise, a small shift in time axis of the transient signal may cause a different set of optimal nodes, which makes selecting the top terminal nodes from the decomposition tree impossible.

It is usually difficult to accurately detect the starting moment of a transient. Shift-invariance ensures the selected terminal nodes in a wavelet decomposition tree are always the best basis regardless of the time origin of the signal. Unfortunately, due to the decimation process and the imperfect frequency response of the wavelet filters, the original DWT is not shift-invariant. As an example, in Fig. 7, a waveform of fault transient has been analyzed using wavelet packet. The signal has been pre-processed, which removed the fundamental frequency component. As can be seen from Fig. 7, the shifted waveform has totally different best-basis representation from its original version. The two pictures in the middle of Fig. 7 give the optimal wavelet decomposition trees while the two pictures at the bottom depict the corresponding wavelet coefficients at all terminal nodes.

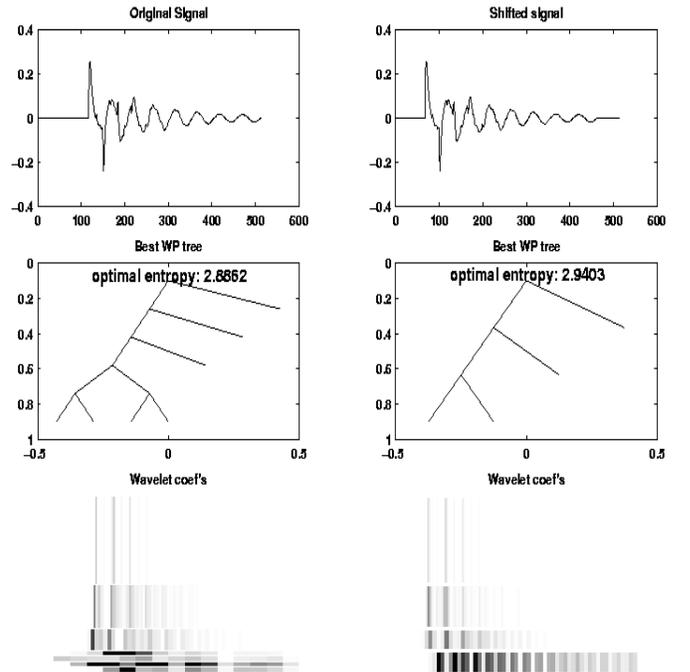


Fig. 7. Shift Variance of Wavelet Packet Transform

Some approaches have been proposed to achieve shiftability, for example, filtering without decimating [12]. Here we applied the method proposed in [13], which simply keeps either the even coefficients or the odd coefficients at each decimation step. Since the fully optimized version is computationally expensive, we use the simplest implementation that looks ahead only one step. The decision rule to choose the even or the odd coefficients is to compare

their entropies. It has been proved that this additional step at each node will make the decomposition results shift-invariant [13].

As shown in Fig. 8, the decomposition patterns remain unchanged regardless of the selection of the time origin, and shift-invariance has been achieved. Also, the optimal entropy has been reduced from 2.8862 to 2.8416.

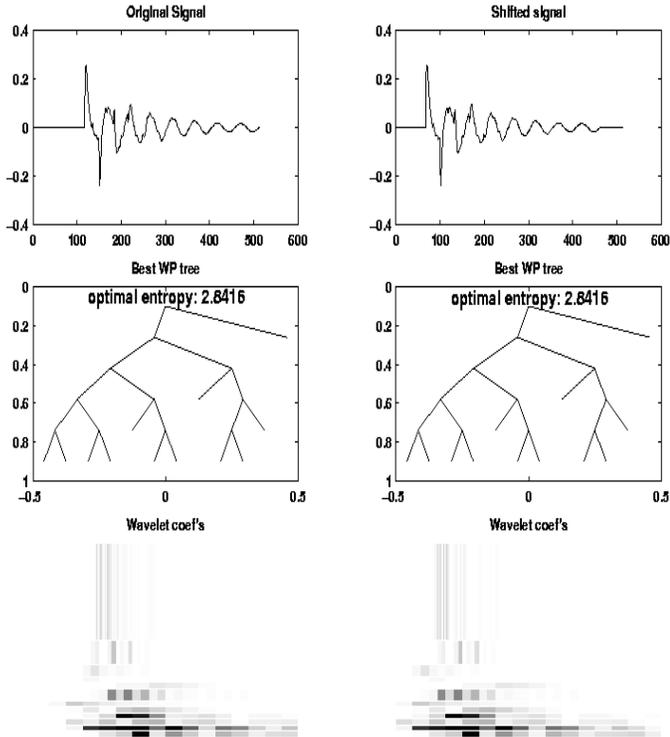


Fig. 8. Shift Invariance of Wavelet Packet Transform

IV. CONCLUSIONS

In this paper, an automatic feature extraction method using wavelet transform has been proposed. The method uses best-basis algorithm to select the most appropriate wavelet filters and decomposition levels. Then the most significant coefficients are chosen according to their discrimination abilities. Those coefficients will compose the feature vectors. To make the best basis algorithm insensitive to signal shift on the time axis, shift invariant wavelet transforms must be used, which in general can also improve the performance.

V. REFERENCES

- [1] A.W. Galli, G. T. Heydt, and P. F. Ribeiro, "Exploring the Power of Wavelet Analysis," *IEEE Computer Applications in Power*, vol. 9, Oct., 1996
- [2] D.C. Robertson, J.S. Mayer, and W.B. Gish, "Wavelets and electromagnetic power system transients," *IEEE Trans. on Power Delivery*, vol. 11, pp 1050-1058, April 1996.
- [3] A.M. Gaouda, "Application of multiresolution signal decomposition for monitoring short-duration variations in distribution systems," *IEEE Trans. on Power Delivery*, vol. 15, April 2000.
- [4] Z.Q.Bo, "A New Approach to Phase Selection Using Fault Generated High Frequency Noise and Neural Networks," *IEEE Trans. on Power Delivery*, vol. 12, Jan. 1997.

- [5] V.L.Pham and K.P.Wong, "Wavelet-transform-based algorithm for harmonic analysis of power system waveforms," *IEE Proc. - Gener. Transm. Distrib.*, vol. 146, May 1999.
- [6] L.Angrisani, et.al., "A measurement method based on the wavelet transform for power quality analysis," *IEEE Trans. on Power Delivery*, vol. 13, October 1998.
- [7] O.Poisson, P.Rioual and M.Meunier, "New signal processing tools applied to power quality analysis," *IEEE Trans. on Power Delivery*, vol. 14, April 1999.
- [8] S. Santoso, E. J. Powers and W. M. Grady, "Power quality data compression using wavelet transform methods," *IEEE Trans. On Power Delivery*, vol. 12, pp 1250-1257, July 1997.
- [9] G.T.Heydt and A.W.Galli, "Transient power quality problems analyzed using wavelets," *IEEE Trans. Power Delivery*, vol. 12, April 1997.
- [10] R. R. Goifman, M. V. Wickerhauser, "Entropy-based algorithm for best basis selection," *IEEE Trans. on Information Theory*, vol. 38, 1992
- [11] JPEG specification, <http://www.jpeg.org/JPEG2000.htm>
- [12] Haitao Guo, "Theory and Applications of the Shift-Invariant Time-Varying and Undecimated Wavelet Transforms", MS thesis, Rich University, May 1995
- [13] I. Cohen, S. Raz and D. Malah, "Orthonormal shift-invariant wavelet packet decomposition and representation", *Signal Processing*, vol. 57, March, 1997

VI. BIOGRAPHIES

Xiangjun Xu (S'99) received his B.E and M.E. degrees from Southeast University and Shanghai Jiaotong University, all in electrical engineering, in 1992 and 1995 respectively. After that, he worked as a teacher in Shanghai Jiaotong University. Since Sep. 1998, he has been with Texas A&M University pursuing his Ph.D. degree. His research interests are computer application on power systems, signal processing, artificial intelligence, and event analysis.

Mladen Kezunovic (S'77, M'80, SM'85, F'99) received his Dipl. Ing. Degree from the University of Sarajevo, the M.S. and Ph.D. degrees from the University of Kansas, all in electrical engineering, in 1974, 1977 and 1980, respectively. Dr. Kezunovic's industrial experience is with Westinghouse Electric Corporation in the USA, and the Energoinvest Company in Sarajevo. He also worked at the University of Sarajevo. He was a Visiting Associate Professor at Washington State University in 1986-1987. He has been with Texas A&M University since 1987 where he is the Eugene E. Webb Professor and Director of Electric Power and Power Electronics Institute. His main research interests are digital simulators and simulation methods for equipment evaluation and testing as well as application of intelligent methods to control, protection and power quality monitoring. Dr. Kezunovic is a registered professional engineer in Texas, and a Fellow of IEEE.