# Wavelet-Based Data Compression of Power System Disturbances Using the Minimum Description Length Criterion

Effrina Yanti Hamid and Zen-Ichiro Kawasaki, *Member, IEEE*

***Abstract*—This paper introduces a compression technique for power disturbance data via discrete wavelet transform (DWT) and wavelet packet transform (WPT). The data compression leads to a potential application for remote power protection and power quality monitoring. The compression technique is performed through signal decomposition up to a certain level, thresholding of wavelet coefficients, and signal reconstruction. The choice of which wavelet to use for the compression is of critical importance, because the wavelet affects reconstructed signal quality and the design of the system as a whole. The minimum description length (MDL) criterion is proposed for the selection of an appropriate wavelet filter. This criterion permits to select not only the suitable wavelet filter but also the best number of wavelet retained coefficients for signal reconstruction. The experimental study has been carried out for a single-phase to ground fault event, and the data compression results of using the suitable wavelet filter show that the compression ratios are less than 11% and are reduced to more than a half of that percentage value by implementing an additional lossless coding.**

***Index Terms*—Data compression, power disturbances, wavelets, wavelet packets.**

## I. Introduction

THE TRANSIENTS due to ground faults, load switchings, and other disturbances may cover a broad frequency spectrum in the order of kilohertz to megahertz. A single captured event recorded for several seconds using monitoring instruments can produce megabytes of data. As a result, the volume of the generated and maintained data increase significantly, which lead to a high cost in storing and transmitting such data. Therefore, it is necessary to develop an effective compression technique which has capability to reduce the volume of data necessary for storing and to speed up the transmitted data for remote monitoring [1]–[3].

Wavelet and wavelet packet transforms have recently emerged as powerful tools for a broad range of applications, in particular, signal compression [2]–[5]. The wavelet transform has good localization in both frequency and time domains, having fine frequency resolution and coarse time resolutions at lower frequency, and coarse frequency resolution and fine time resolution at higher frequency. It makes the wavelet transform suitable for time–frequency analysis. In data compression, the wavelet transform is used to exploit the redundancy in the signal. The performance of a wavelet transform for data compression lies in its ability in concentrating a large percentage of total signal energy in a few coefficients [6]. After the original signal is transformed into the wavelet coefficients, many coefficients are so small so that these coefficients can be omitted without losing significant information after the signal is reconstructed.

During the last three years, power disturbance data compression using wavelet and wavelet packet transforms have been proposed [2], [3]. The choice of which wavelet to use in a compression system plays an important role, because the wavelet affects reconstructed signal quality and the design of the system as a whole. Compared with the actual compression performance of several different wavelets, the previous authors [2], [3] choose only a specific wavelet filter. Improper choice of filter can produce distortions in the reconstructed signal and can cause not optimum compression ratio. An algorithm to optimize the efficiency of compression in the wavelet domain called the minimum description length (MDL) has been proposed by Saito [7]. The MDL criterion aims to gain the compromise between the number of retained wavelet coefficients and the error of signal reconstruction. The algorithm permits one to select the suitable wavelet filter and the best number of wavelet retained coefficients of a signal.

In this paper, we propose a data compression method based on wavelet and wavelet packet for power system disturbances. The method includes the selection of wavelet filter using the MDL criterion to optimize the compression technique. We evaluate several wavelet filters and compare their performances. Although there are many types of wavelet filters, we restrict ourselves to the Daubechies, Coiflets, and Symlets families with a certain level of decomposition. In addition, the results from this wavelet-based compression method are then combined with a lossless coding, e.g., Huffman, Lempel–Ziv–Welch (LZW), or Lempel–Ziv–Haruyasu (LZH), to get more effective compression [3].

## II. Wavelet Transforms

### A. Discrete Wavelet Transform

The wavelet transform of a discrete input data sequence $\boldsymbol{f} = \{f_n\} = \{f_0, f_1, \ldots, f_{N-1}\}$ where $N$ is the length, can be presented in a vector matrix form as

$$\boldsymbol{\alpha} = \mathbf{W}\boldsymbol{f} \tag{1}$$

Manuscript received September 27, 2000.

E. Y. Hamid is with the Department of Electrical Engineering, Osaka University, Osaka, Japan.

Z.-I. Kawasaki is with the Department of Communication Engineering, Osaka University, Osaka, Japan.

Publisher Item Identifier S 0885-8977(02)03277-6.

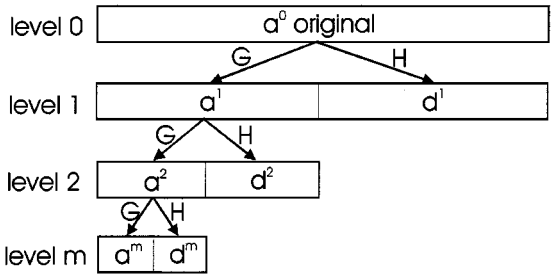


Fig. 1. Decomposition of $\boldsymbol{a}^0$ up to level $m$ using DWT.

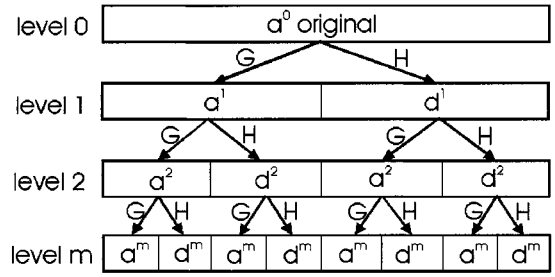


Fig. 2. Wavelet packet decomposition of $\boldsymbol{a}^0$ viewed as a binary tree.

where $\boldsymbol{\alpha}$ contains $N$ wavelet coefficients, and $\mathbf{W}(N \times N)$ is an orthogonal matrix consisting of row basis vectors. The basis vectors are specified by a set of numbers called wavelet and scaling filter coefficients.

Once a specific wavelet has been chosen, we can use its coefficients to define two filters: the low-pass filter and the high-pass filter. Both types of filters use the same set of wavelet filter coefficients, but with alternating signs and in reversed order, meaning this pair of filters is the quadrature mirror filters (QMF). The low-pass and high-pass filters are also called the scaling and the wavelet filters, respectively. These filters are used to construct the filter matrices, denoted as $\mathbf{G}$ and $\mathbf{H}$.

To decompose (or analyze) the signal, Mallat [8] introduced a recursive algorithm which is known as the pyramid algorithm. This algorithm offers the hierarchical, multiresolution of the signal. In this algorithm, the set of $N$ input data is passed through the low-pass and high-pass filters. Each output of the filter consists of $N/2$ wavelet coefficients. The output from low-pass filter is the approximation coefficients ($\boldsymbol{a}^1 = \{a_0^1, a_1^1, \ldots, a_{N/2-1}^1\}$) at the first level of resolution. The output from high-pass filter is the detail coefficients ($\boldsymbol{d}^1 = \{d_0^1, d_1^1, \ldots, d_{N/2-1}^1\}$) at the first level of resolution. The approximation coefficient $\boldsymbol{a}^1$ can now be used as the data input for another pair of wavelet filters (identical with the first pair), generating sets of length $N/4$ of approximation ($\boldsymbol{a}^2 = \{a_0^2, a_1^2, \ldots, a_{N/4-1}^2\}$) and details coefficients ($\boldsymbol{d}^2 = \{d_0^2, d_1^2, \ldots, d_{N/4-1}^2\}$) at the second level of resolution. The process is continued until a desired level of resolution. Since the original input data vector $\boldsymbol{f}$ is the approximation at the lowest resolution (level 0), i.e., $\boldsymbol{a}^0 = \boldsymbol{f} = \{f_0, f_1, \ldots, f_{N-1}\}$, then the DWT algorithm can be presented by the following recursive formula:

$$\boldsymbol{a}^m = \mathbf{G}\boldsymbol{a}^{m-1} \quad \text{and } \boldsymbol{d}^m = \mathbf{H}\boldsymbol{a}^{m-1} \tag{2}$$

where $m$ denotes the resolution level and $m = 1, 2, \ldots, \log_2 N$. Fig. 1 shows this decomposition process.

The different resolution for each level is related to the sampling interval. For level $m$ the sampling interval equals $2^m$. As the sampling interval increases, resolution decreases and each approximation contains gradually less information. The difference in information between the approximations at level $m$ and level $m-1$ is contained in the detail at level $m$.

It is possible to use the approximation and detail coefficients to reconstruct (or synthesize) the original signal. The reconstruction process uses the recursion algorithm in reverse with conjugates of $\mathbf{G}$ and $\mathbf{H}$. For the orthonormal basis the conjugates of $\mathbf{G}$ and $\mathbf{H}$ are equal to the transposed matrices $\mathbf{G}^T$ and $\mathbf{H}^T$, respectively. Thus, the reconstruction formula is as follows:

$$\boldsymbol{a}^{m-1} = \mathbf{G}^T\boldsymbol{a}^m + \mathbf{H}^T\boldsymbol{d}^m. \tag{3}$$

In general, noise suppression is implemented before the signal is reconstructed. This means that the wavelet coefficients $\boldsymbol{d}^m$ and/or $\boldsymbol{a}^m$ whose absolute value is less than a predefined threshold is set, for example, to zero, and then (3) is applied.

### B. Wavelet Packet Transform (WPT)

*1) Theory:* The wavelet packet transform (WPT) is a direct expansion of the structure of the DWT tree algorithm to a full binary tree. In the pyramid algorithm, the detail branches are not used for further calculations, only the approximations at each level of resolution are treated to yield approximation and detail obtained at a higher level. For the wavelet packet, both the detail and approximation coefficients at level $m$ are further decomposed into level $m+1$. The main advantage of the WPT is better signal representation. The search for the best representation of the signal by any subtree of the WPT is called the best-basis selection. Wavelet packet decomposition is shown in Fig. 2, in a tree structure to indicate the decomposition processes. The detail and approximation coefficients in each level for each tree (or subspace) are derived in similar manner to those of DWT using (2).

*2) Best-Basis Selection:* The overcomplete representation of signal by the WPT allows us to choose the appropriate representation of the signal. To find the best-basis or the wavelet coefficients of the best-tree, one first computes its complete detail and approximation (wavelet) coefficients up to a desired level. Then, it is very natural to use the entropy as a measure of efficiency of the basis. Here the entropy of a signal $\boldsymbol{x} = \{x_n\} = \{x_0, x_1, \ldots, x_{N-1}\}$ is defined as

$$H(\boldsymbol{x}) = -\sum_{n}^{N-1} |x_n|^2 \log |x_n|^2 \tag{4}$$

which is known as the nonnormalized Shannon entropy [9]. The best-basis is the basis giving the minimum entropy or maximum information for its distribution of coefficients [6], [9].

The wavelet packet may be efficiently searched for the best-basis. Each tree in the binary tree as shown in Fig. 2 represents a subspace, consisting of the detail or approximation coefficients, of the original signal. Each parent subspace is the orthogonal

sum of its two children's subspaces. The search for the best-basis involves computing entropy using (4) for each subspace, then performing a comparison between the entropy of parent subspace and that of its two children's subspaces. If the parent has a smaller entropy, its two children are omitted from the tree. On the other hand, if the parent has a larger entropy, its two children are kept from the tree. This process is repeated until the original signal at the top level is reached (see also Fig. 4).

## III. Minimum Description Length (MDL) Criterion

The minimum description length (MDL) criterion is an interesting approach to simultaneous noise suppression and signal compression. It is free from any parameter setting such as threshold selection, which can be particularly useful for real data where the noise level is difficult to estimate. The MDL selects the "best" wavelet filter and the "best" number of wavelet coefficients to be retained for the signal reconstruction [7].

The MDL criterion has the following algorithm. Let us consider a discrete model

$$\boldsymbol{f} = \boldsymbol{x} + \boldsymbol{n}$$

where the vector $\boldsymbol{f}$ represents the noisy observed data, vector $\boldsymbol{x}$ is the unknown true signal to be estimated, and vector $\boldsymbol{n}$ is noise. *First*, pick the index $(k, n)$ from the MDL function defined as

$$\mathrm{MDL}(k,n) = \min\left\{\frac{3}{2}k\log N + \frac{N}{2}\log\|\tilde{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_n^{(k)}\|^2\right\} \quad 0 \le k < N; 1 \le n \le M \tag{5}$$

where $\tilde{\boldsymbol{\alpha}}_n = \mathbf{W}_n\boldsymbol{f}$ denotes the vector of the decomposition coefficients of $\boldsymbol{f}$ via the wavelet filter $n$, and $\boldsymbol{\alpha}_n^{(k)} = \boldsymbol{\Theta}^{(k)}\tilde{\boldsymbol{\alpha}}_n = \boldsymbol{\Theta}^{(k)}(\mathbf{W}_n\boldsymbol{f})$ denotes the vector that contains $k$ nonzero elements, and $\boldsymbol{\Theta}^{(k)}$ is a hard-thresholding operation which keeps the $k$ largest elements of $\tilde{\boldsymbol{\alpha}}_n$ in absolute value intact and set all other elements to zero. The $N$ and $M$ denote, respectively, the length of the signal and the total number of wavelet filters used. The $\tilde{\boldsymbol{\alpha}}_n$ and $\boldsymbol{\alpha}_n^{(k)}$ have to be normalized by $\|\tilde{\boldsymbol{\alpha}}_n\|$, so that the magnitude of each coefficient in $\tilde{\boldsymbol{\alpha}}_n$ and $\boldsymbol{\alpha}_n^{(k)}$ is strictly less than one. Note that $\|\boldsymbol{x}\|$ is defined as $(\sum_0^{N-1}|x_n|^2)^{1/2}$. The MDL function is expressed as the sum of two conflicting terms. The first term represents the penalty function, linearly increasing with the number of the retained wavelet coefficients $k$, whereas the second term describes the logarithmic of residual energy between $\tilde{\boldsymbol{\alpha}}_n$ and $\boldsymbol{\alpha}_n^{(k)}$. We see that the log(residual energy) always decreases as $k$ increases (see also Fig. 5). Number of coefficients $k$, for which the MDL function reaches its minimum, is considered as the optimal one. With this criterion, one can optimize the choice of wavelet filter as well. It should be noted that each wavelet filter has different characteristics. A wavelet filter, which is optimal for a given signal, is not necessarily the best for another type of signal.

*Second*, reconstruct the estimated true signal $\boldsymbol{x}$ through the following equation:

$$\boldsymbol{x} = \mathbf{W}_n^T\boldsymbol{\alpha}_n^{(k)} \tag{6}$$

which is exactly the same process as in (3).

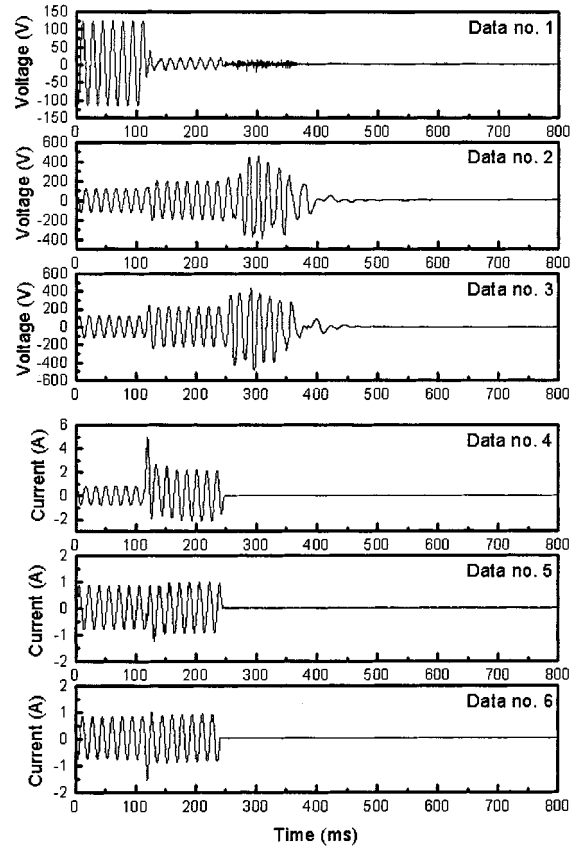


Fig. 3. Fault record from a single-phase to ground of three-phase power system. Data no. 1, 2, and 3 are the voltage of phase a, b, and c, respectively, and the data no. 4, 5, and 6 are for the current of phase a, b, and c, respectively. The fault occurred at 116 ms on phase a.

## IV. Experimental Study

### A. Power Disturbance Data

The experimental study has been carried out for a single-phase to ground fault event, and six power disturbance data have been recorded. The data were obtained from a power system hardware/analog simulator owned by The Kansai Electric Power Company (KEPCO), Japan. The performances of DWT and WPT compression are evaluated using these power disturbance data. Fig. 3 shows these original signals. The length of each signal is $N = 8000$ samples for 800 ms. Each sample requires 12 bytes ASCII and only the magnitudes are stored, so that each signal has a size of 96 000 bytes.

### B. Library of Wavelet Filters

Ten wavelets from the Daubechies family (with two, four, six, eight, ten, 12, 14, 16, 18, and 20 filter coefficients), five wavelets from Coiflets (with two, four, six, eight, and ten filter coefficients), and seven wavelets from Symlets (with four, six, eight, ten, 12, 14, and 16 filter coefficients) are used for the data compression. This corresponds to $M = 22$. The coefficients of each wavelet filter can be found in [9].

### C. Performance Evaluation

To evaluate the compression performance, two performance indexes are employed. The first one is the compression ratio

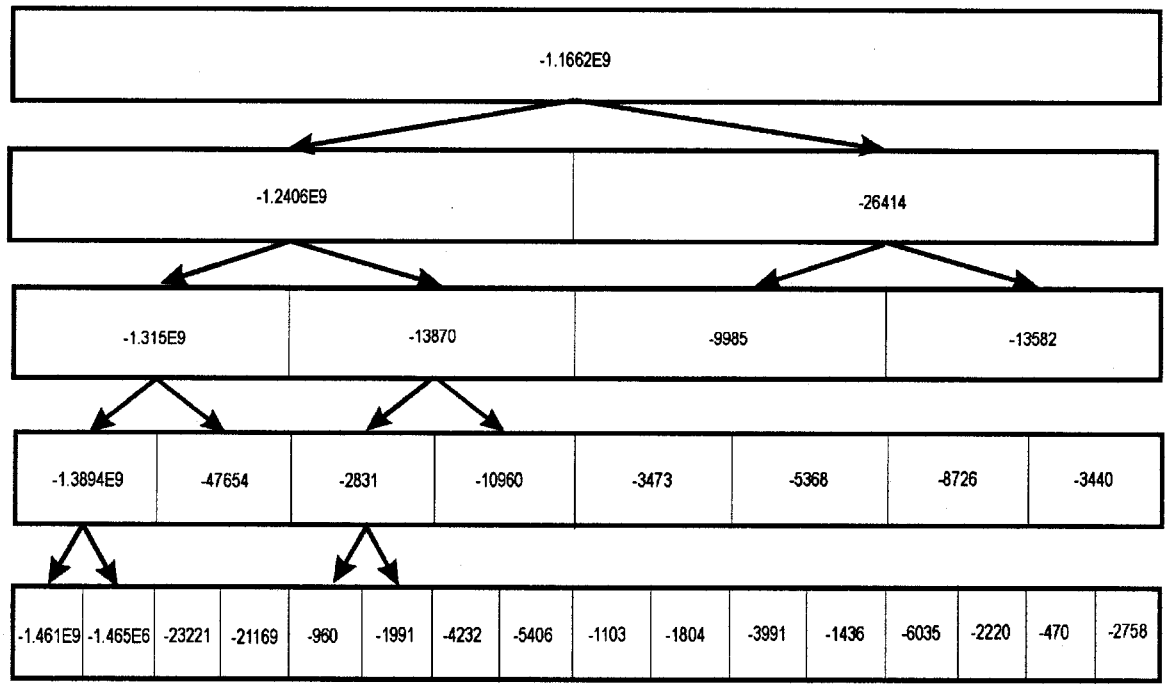


Fig. 4. Entropy values of each subspace where the tree indicates the wavelet packet best-basis of data no. 2.

(CR), i.e., the ratio of the size of the compressed file over the size of the original file, defined as

$$\mathrm{CR}(\%) = \frac{\text{bytes of compressed signal}}{\text{bytes of original signal}} \times 100. \tag{7}$$

The second one is the percentage of mean square error, defined as

$$\mathrm{MSE}(\%) = \frac{\sqrt{\sum_{n=0}^{N-1}(f_n - x_n)^2}}{\sqrt{\sum_{n=0}^{N-1} f_n^2}} \times 100 \tag{8}$$

where $\boldsymbol{f}$ and $\boldsymbol{x} = \{x_n\} = \{x_0, x_1, \ldots, x_{N-1}\}$ are noisy observed (or original) signal and reconstructed signal, respectively.

## V. Results

We compare the performance of 22 wavelet filters for the compression. All signals are decomposed via the DWT and WPT with those filters up to fourth level of resolution $(m = 4)$. For the case of the WPT, the decomposition is performed following the best-basis selection with minimum entropy criterion. The wavelet coefficients from the decomposition is sorted according to their absolute amplitude. The optimal number of retained coefficients $k$ can be calculated based on the MDL criterion.

To simplify the explanation we will give attention to the signal of data no. 2, and we apply the WPT with the Daubechies 5 (Db5) filter. First, the data is decomposed up to a predefined level using (2). The entropy of each subspace is then calculated using (4) to find the best-basis. Fig. 4 shows the result of the best-basis with minimum entropy criterion. Once the best-basis is found, the MDL function is applied to compute the number of wavelet retained coefficients $k$. The result of the MDL function and its components is shown in Fig. 5. The function reaches the minimum at $k = 595$. This means the minimum number of coefficients required for the signal reconstruction with the smallest distortion is 595. The process above is repeated until the last wavelet filter in the library $(n = 22)$, and then, the appropriate filter can be chosen.

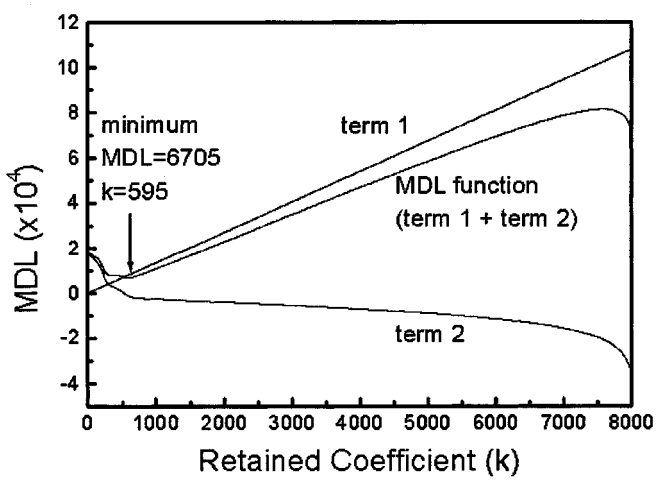


Fig. 5. MDL function and its components for the WPT coefficients of data no. 2 with Db5 filter.

We have applied the MDL criterion to all data to select the suitable filter, and the results are tabulated in Tables I and II for the DWT and WPT, respectively. Both tables show the number of retained coefficients $k$, the MSE, and the minimum value of the MDL function for all wavelet filters. From this point, we can choose the appropriate filter for each corresponding data based

TABLE I
NUMBER OF RETAINED COEFFICIENTS, MSE, AND MDL VALUE FOR 22 WAVELET FILTERS USING DWT

| Filter n | k (1) | MSE (1) | MDL (1) | k (2) | MSE (2) | MDL (2) | k (3) | MSE (3) | MDL (3) | k (4) | MSE (4) | MDL (4) | k (5) | MSE (5) | MDL (5) | k (6) | MSE (6) | MDL (6) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Db1 | 160 | 14.68 | 12904 | 336 | 10.38 | 13891 | 327 | 10.31 | 13740 | 214 | 9.86 | 12041 | 247 | 11.00 | 12922 | 217 | 13.84 | 13435 |
| Db2 | 133 | 10.95 | 11367 | 341 | 3.40 | 9491 | 388 | 2.76 | 9290 | 205 | 3.63 | 7916 | 168 | 6.56 | 9789 | 147 | 9.97 | 11182 |
| Db3 | 135 | 10.13 | 11080* | 598 | 0.89 | 7607 | 629 | 0.73 | 7246 | 163 | 3.26 | 6927 | 591 | 1.14 | 8478 | 556 | 1.72 | 9671 |
| Db4 | 138 | 10.11 | 11114 | 577 | 0.75 | 6644 | 600 | 0.62 | 6205 | 539 | 0.85 | 6638 | 547 | 1.15 | 7931 | 527 | 1.71 | 9246 |
| Db5 | 138 | 10.17 | 11137 | 588 | 0.72 | 6611 | 609 | 0.59 | 6084 | 528 | 0.84 | 6434* | 547 | 1.13 | 7845 | 529 | 1.64 | 9106 |
| Db6 | 140 | 10.02 | 11105 | 577 | 0.73 | 6539 | 601 | 0.59 | 6016 | 537 | 0.83 | 6477 | 541 | 1.11 | 7693 | 533 | 1.61 | 9098 |
| Db7 | 132 | 10.50 | 11184 | 593 | 0.72 | 6696 | 603 | 0.59 | 6041 | 537 | 0.82 | 6445 | 538 | 1.10 | 7629* | 533 | 1.61 | 9091 |
| Db8 | 144 | 10.12 | 11200 | 577 | 0.73 | 6537 | 604 | 0.60 | 6128 | 541 | 0.82 | 6490 | 547 | 1.08 | 7699 | 531 | 1.61 | 9065* |
| Db9 | 143 | 10.16 | 11200 | 583 | 0.74 | 6667 | 601 | 0.59 | 6004* | 539 | 0.82 | 6482 | 548 | 1.08 | 7683 | 539 | 1.59 | 9118 |
| Db10 | 144 | 10.20 | 11229 | 578 | 0.73 | 6550 | 608 | 0.59 | 6112 | 542 | 0.82 | 6516 | 540 | 1.09 | 7631* | 534 | 1.61 | 9115 |
| Coif1 | 136 | 10.70 | 11313 | 331 | 3.49 | 9457 | 377 | 2.75 | 9125 | 201 | 3.67 | 7908 | 165 | 6.47 | 9695 | 633 | 1.92 | 11139 |
| Coif2 | 147 | 9.92 | 11159 | 582 | 0.76 | 6758 | 597 | 0.63 | 6221 | 544 | 0.85 | 6704 | 551 | 1.13 | 7911 | 533 | 1.67 | 9232 |
| Coif3 | 146 | 10.18 | 11248 | 577 | 0.75 | 6611 | 603 | 0.60 | 6068 | 541 | 0.83 | 6534 | 541 | 1.10 | 7683 | 535 | 1.63 | 9162 |
| Coif4 | 151 | 10.11 | 11290 | 582 | 0.73 | 6589 | 610 | 0.59 | 6115 | 545 | 0.82 | 6576 | 552 | 1.09 | 7782 | 540 | 1.60 | 9162 |
| Coif5 | 161 | 9.92 | 11351 | 588 | 0.74 | 6713 | 602 | 0.61 | 6111 | 556 | 0.82 | 6690 | 558 | 1.07 | 7780 | 544 | 1.60 | 9205 |
| Sym2 | 133 | 10.95 | 11367 | 341 | 3.40 | 9491 | 388 | 2.76 | 9290 | 205 | 3.63 | 7916 | 168 | 6.56 | 9789 | 147 | 9.97 | 11182 |
| Sym3 | 135 | 10.13 | 11080 | 598 | 0.89 | 7607 | 629 | 0.73 | 7246 | 163 | 3.26 | 6927 | 591 | 1.14 | 8478 | 556 | 1.72 | 9671 |
| Sym4 | 138 | 9.99 | 11065* | 578 | 0.77 | 6766 | 590 | 0.65 | 6219 | 534 | 0.88 | 6692 | 545 | 1.15 | 7910 | 529 | 1.68 | 9217 |
| Sym5 | 138 | 10.10 | 11111 | 584 | 0.72 | 6584 | 609 | 0.58 | 6054 | 537 | 0.84 | 6518 | 546 | 1.12 | 7828 | 532 | 1.63 | 9120 |
| Sym6 | 141 | 10.02 | 11120 | 587 | 0.71 | 6561 | 601 | 0.59 | 6024 | 536 | 0.83 | 6468 | 536 | 1.12 | 7664 | 528 | 1.62 | 9056* |
| Sym7 | 138 | 10.26 | 11175 | 569 | 0.75 | 6502* | 603 | 0.59 | 6022 | 536 | 0.82 | 6443* | 532 | 1.12 | 7637 | 537 | 1.60 | 9114 |
| Sym8 | 142 | 10.20 | 11202 | 577 | 0.72 | 6486* | 592 | 0.61 | 5971* | 543 | 0.83 | 6554 | 543 | 1.09 | 7671 | 535 | 1.61 | 9114 |

*Note:* The number inside the parenthesis is the data number, and the asterisk (*) indicates the first two minimum MDL.

TABLE II
NUMBER OF RETAINED COEFFICIENTS, MSE, AND MDL VALUE FOR 22 WAVELET FILTERS USING WPT

| Filter n | k (1) | MSE (1) | MDL (1) | k (2) | MSE (2) | MDL (2) | k (3) | MSE (3) | MDL (3) | k (4) | MSE (4) | MDL (4) | k (5) | MSE (5) | MDL (5) | k (6) | MSE (6) | MDL (6) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Db1 | 187 | 13.40 | 12890 | 341 | 9.98 | 13787 | 336 | 9.74 | 13622 | 352 | 5.46 | 11521 | 279 | 9.82 | 12886 | 217 | 13.88 | 13433 |
| Db2 | 135 | 10.90 | 11362 | 354 | 3.26 | 9481 | 387 | 2.76 | 9261 | 206 | 3.63 | 7916 | 169 | 6.56 | 9789 | 148 | 9.97 | 11182 |
| Db3 | 139 | 10.05 | 11090 | 608 | 0.88 | 7670 | 628 | 0.75 | 7287 | 165 | 3.25 | 6926 | 592 | 1.14 | 8478 | 557 | 1.72 | 9671 |
| Db4 | 142 | 10.06 | 11136 | 583 | 0.75 | 6705 | 603 | 0.65 | 6382 | 539 | 0.85 | 6621 | 550 | 1.15 | 7967 | 528 | 1.71 | 9239 |
| Db5 | 135 | 10.44 | 11189 | 595 | 0.72 | 6705 | 620 | 0.58 | 6166 | 530 | 0.84 | 6453* | 547 | 1.12 | 7816 | 529 | 1.64 | 9091 |
| Db6 | 146 | 10.02 | 11173 | 593 | 0.71 | 6626 | 593 | 0.62 | 6060 | 538 | 0.83 | 6480 | 540 | 1.12 | 7704 | 532 | 1.62 | 9082 |
| Db7 | 138 | 10.50 | 11251 | 589 | 0.75 | 6766 | 594 | 0.63 | 6125 | 538 | 0.83 | 6480 | 537 | 1.10 | 7625* | 534 | 1.61 | 9079 |
| Db8 | 150 | 10.18 | 11291 | 595 | 0.71 | 6631 | 614 | 0.59 | 6149 | 541 | 0.82 | 6487 | 549 | 1.08 | 7707 | 531 | 1.61 | 9055* |
| Db9 | 129 | 10.27 | 11044* | 587 | 0.74 | 6712 | 596 | 0.60 | 6007 | 540 | 0.82 | 6482 | 549 | 1.08 | 7683 | 540 | 1.59 | 9117 |
| Db10 | 137 | 10.29 | 11158 | 587 | 0.73 | 6664 | 603 | 0.60 | 6072 | 543 | 0.82 | 6516 | 541 | 1.09 | 7631 | 535 | 1.61 | 9115 |
| Coif1 | 136 | 10.73 | 11313 | 331 | 3.50 | 9463 | 381 | 2.73 | 9137 | 202 | 3.67 | 7908 | 166 | 6.47 | 9695 | 634 | 1.92 | 11147 |
| Coif2 | 148 | 9.92 | 11159 | 586 | 0.74 | 6708 | 597 | 0.63 | 6193 | 541 | 0.86 | 6658 | 549 | 1.14 | 7910 | 534 | 1.66 | 9209 |
| Coif3 | 150 | 10.01 | 11221 | 585 | 0.71 | 6492 | 600 | 0.59 | 5964 | 540 | 0.83 | 6510 | 539 | 1.10 | 7652 | 534 | 1.63 | 9138 |
| Coif4 | 154 | 10.03 | 11286 | 586 | 0.70 | 6487 | 611 | 0.58 | 6028 | 546 | 0.82 | 6576 | 553 | 1.09 | 7782 | 541 | 1.60 | 9162 |
| Coif5 | 162 | 9.94 | 11357 | 597 | 0.71 | 6657 | 601 | 0.59 | 5974 | 557 | 0.82 | 6690 | 559 | 1.07 | 7780 | 545 | 1.60 | 9205 |
| Sym2 | 135 | 10.90 | 11362 | 354 | 3.26 | 9481 | 387 | 2.76 | 9261 | 206 | 3.63 | 7916 | 169 | 6.56 | 9789 | 148 | 9.97 | 11182 |
| Sym3 | 139 | 10.05 | 11090 | 608 | 0.88 | 7670 | 628 | 0.75 | 7287 | 165 | 3.25 | 6926 | 592 | 1.14 | 8478 | 557 | 1.72 | 9671 |
| Sym4 | 140 | 9.93 | 11058* | 589 | 0.75 | 6752 | 591 | 0.65 | 6220 | 532 | 0.89 | 6681 | 544 | 1.15 | 7887 | 531 | 1.67 | 9199 |
| Sym5 | 140 | 10.05 | 11103 | 586 | 0.72 | 6590 | 593 | 0.61 | 6033 | 539 | 0.83 | 6524 | 545 | 1.13 | 7807 | 530 | 1.64 | 9101 |
| Sym6 | 144 | 9.96 | 11121 | 585 | 0.70 | 6459 | 599 | 0.59 | 5946 | 536 | 0.83 | 6474 | 537 | 1.12 | 7664 | 529 | 1.62 | 9056* |
| Sym7 | 146 | 10.01 | 11167 | 569 | 0.73 | 6420* | 598 | 0.58 | 5885* | 538 | 0.82 | 6464* | 532 | 1.12 | 7610* | 536 | 1.60 | 9093 |
| Sym8 | 143 | 10.20 | 11203 | 580 | 0.71 | 6435* | 583 | 0.60 | 5818* | 543 | 0.83 | 6539 | 540 | 1.10 | 7647 | 535 | 1.60 | 9088 |

*Note:* The number inside the parenthesis is the data number, and the asterisk (*) indicates the first two minimum MDL.

on the minimum MDL value, and the results for the first two filters having smallest MDL are tabulated in Table III. We can see that the appropriate filter for a given signal may be different for another type of signal. However, in practice, it is highly preferable to use only one "best" filter for all signals. From Table III, the Symlets 7 and Symlets 8 filters seem to be the candidates for the best filter. We simply select the Symlets 7 filter for the compression of all power disturbance data analyzed here.

TABLE III
APPROPRIATE WAVELET FILTERS BASED ON MDL CRITERION

| Data | DWT | WPT |
|---|---|---|
| 1 | Sym4 - Db3 | Db9 - Sym4 |
| 2 | Sym8 - Sym7 | Sym7 - Sym8 |
| 3 | Sym8 - Db9 | Sym8 - Sym7 |
| 4 | Db5 - Sym7 | Db5 - Sym7 |
| 5 | Db7 - Db10 | Sym7 - Db7 |
| 6 | Sym6 - Db8 | Db8 - Sym6 |

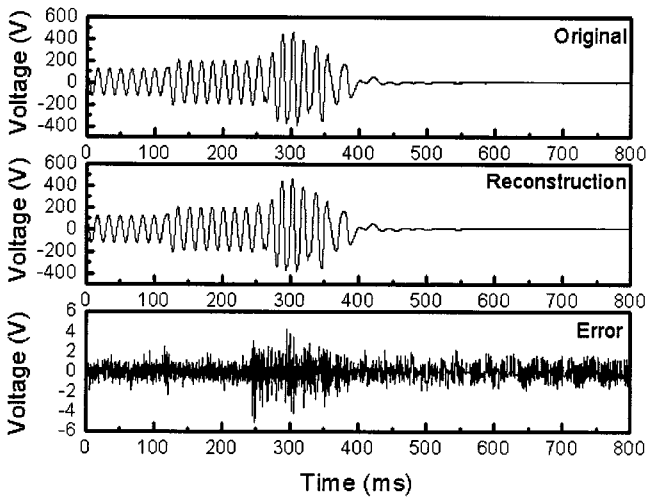


Fig. 6. Original, reconstructed, and residual error signals of data no. 2 using WPT with Sym7 filter.

TABLE IV
CR AND MSE USING DWT WITH SYMLETS 7 FILTER AND LOSSLESS CODINGS

| Data | DWT (%) | DWT+Huff. (%) | DWT+LZW (%) | DWT+LZH (%) | MSE (%) |
|---|---|---|---|---|---|
| 1 | 2.49 | 1.10 | 1.20 | 1.09 | 10.26 |
| 2 | 10.10 | 4.38 | 4.45 | 4.19 | 0.75 |
| 3 | 10.72 | 4.65 | 4.67 | 4.40 | 0.59 |
| 4 | 9.54 | 3.74 | 3.20 | 2.80 | 0.82 |
| 5 | 9.46 | 3.73 | 3.05 | 2.68 | 1.12 |
| 6 | 9.55 | 3.76 | 3.10 | 2.74 | 1.60 |

Using the MDL, we can compute the number of wavelet coefficients to be stored as the compressed data. Here the compressed data contains both magnitude and position of the coefficients. We allocate 12 bytes ASCII for the magnitude and 5 bytes ASCII for its position. For the signal reconstruction, first the wavelet retained coefficients are rearranged according to their positions, and then zero magnitudes are inserted to the rest of the locations. Then, the signal reconstruction from these coefficients is done using (3) or (6). Fig. 6 shows an example of the original signal, reconstructed signal and its residual error of data no. 2 using the Symlets 7 filter.

In addition, more effective compression can be performed by implementing an additional lossless coding (e.g., Huffman, LZW, or LZH) to the results of the DWT and WPT compression. Since the coding has lossless properties, the compression always reproduce the same data when a file is decompressed. Tables IV and V show the comparison of CR and MSE of the analyzed signals using the Symlets 7 filter. The compressed file size (in percentage of original file size) is calculated for the DWT, WPT, and DWT + lossless coding as well as WPT + lossless coding. Both the DWT and WPT compression significantly reduces the original file size of each signal to less than 11%. Further, the tables show that by implementing the lossless coding the CRs are reduced to more than a half of those CRs without the lossless coding.

TABLE V
CR AND MSE USING WPT WITH SYMLETS 7 FILTER AND LOSSLESS CODINGS

| Data | WPT (%) | WPT+Huff. (%) | WPT+LZW (%) | WPT+LZH (%) | MSE (%) |
|---|---|---|---|---|---|
| 1 | 2.75 | 1.21 | 1.32 | 1.19 | 10.01 |
| 2 | 10.23 | 4.43 | 4.50 | 4.22 | 0.73 |
| 3 | 10.75 | 4.66 | 4.67 | 4.39 | 0.58 |
| 4 | 9.72 | 3.80 | 3.25 | 2.85 | 0.82 |
| 5 | 9.55 | 3.76 | 3.07 | 2.70 | 1.12 |
| 6 | 9.64 | 3.80 | 3.13 | 2.78 | 1.60 |

## VI. CONCLUSION

The application of DWT and WPT to compress the data of power system disturbances has been evaluated. Both transforms offer attractive properties for the compression. The experimental results show that better quality reconstruction can be achieved by employing an appropriate wavelet filter to each signal. In practice, it is preferable to use one suitable filter for all signals. Using the MDL criterion, the Symlets 7 filter generally appears superior than other wavelet filters for most power disturbance signals analyzed here. The compression ratios that can be obtained using this filter are varied but less than 11%. Combining wavelet and wavelet packet compression with a lossless coding could result in better compression ratios. Our results show that the compression ratios are reduced to more than a half by implementing an additional lossless coding. Finally, the compression algorithm presented here can be used to compress not only ground fault signals but also a wide variety of one-dimensional (1-D) power system disturbance signals.

## ACKNOWLEDGMENT

The authors would like to thank the Kansai Electric Power Company (KEPCO), Japan, for the permission to use the ground fault data.

## REFERENCES

[1] K. B. Mehta and B. D. Russel, "Data compression for digital data from power system disturbance: Requirements and technique evaluation," *IEEE Trans. Power Delivery*, vol. 4, pp. 1683–1689, July 1989.
[2] S. Santoso, E. J. Power, and W. M. Grady, "Power quality data compression using wavelet transform methods," *IEEE Trans. Power Delivery*, vol. 12, pp. 1250–1257, July 1997.
[3] T. B. Littler and D. J. Morrow, "Wavelet for the analysis and compression of power system disturbances," *IEEE Trans. Power Delivery*, vol. 14, pp. 358–362, Apr. 1999.
[4] M. L. Hilton, "Wavelet and wavelet packet compression of electrocardiograms," *IEEE Trans. Biomed. Eng.*, vol. 44, pp. 394–402, May 1997.

[5] B. Walczak and D. L. Massart, "Noise suppresion and signal compression using the wavelet packet transform," *Chemometr. Intell. Lab. Syst.*, vol. 36, pp. 81–94, Apr. 1997.
[6] R. R. Coifmann and M. V. Wickerhauser, "Entropy-based algorithms for best-basis selection," *IEEE Trans. Inform. Theory*, vol. 38, pp. 713–719, Mar. 1992.
[7] N. Saito, "Simultaneous noise suppression and signal compression using a library of orthonormal bases and the minimum description length criterion," in *Wavelets in Geophysics*, E. Foufoula-Georgiou and P. Kumar, Eds. New York: Academic, 1994, pp. 299–324.
[8] S. A. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 674–693, July 1989.
[9] M. V. Wickerhauser, "Adapted wavelet analysis from theory to software," in *Adapted Wavelet Analysis From Theory to Software*. Piscataway, NJ: IEEE, 1994, pp. 273–298.

**Effrina Yanti Hamid** was born in Medan, Indonesia. She received the B.S. and M.S. degrees in electrical engineering from Institut Teknologi Bandung, Bandung, Indonesia, in 1995 and 1998, respectively. She is currently pursuing the Ph.D. degree in electrical engineering from Osaka University, Osaka, Japan.

Her main research interests include signal processing and its application to power systems.

**Zen-Ichiro Kawasaki** (S'75–M'78) was born in Osaka, Japan. He received the B.S., M.S., and Dr. Eng. degrees in communication engineering from Osaka University (OU), Osaka, Japan, in 1973, 1975, and 1978, respectively.

From 1989 to 2000, he was with the Department of Electrical Engineering, OU. Currently, he is a Professor in the Department of Communication Engineering at OU. His current research interests are in signal processing, diagnosis techniques of power apparatus, and the electromagnetics of lightning discharges.

Dr. Kawasaki is a member of the Institute of Electrical Engineers of Japan (IEEJ), the American Geophysical Union (AGU), and the Society of Atmospheric Electricity of Japan (SAEJ).