To perform experiments for real recorded signals, we used speech, car noise, or music as the noises. Another set of speech was used as the signal s. Korean sentences were recorded for the speech, and the car noise and music were obtained in NOISEX-92 CD-ROMs and a Korean popular song, respectively. Each signal was 10 s long with 16 kHz sampling rate. It is known that speech signal approximately follows a Laplacian distribution. Therefore, sign(·) was used as the score function. Fig. 1 shows the filter h_{12} which was measured in a normal office room, and the number of taps of adaptive filter coefficients was 1024. Table 2 shows the SNRs of the two algorithms for the three different noises after convergence. The SNRs of the ICA-based approach were superior to those of the LMS algorithm. These results show that the ICA-based approach can remove dependent components through higher-order statistics for real recorded signals as well.

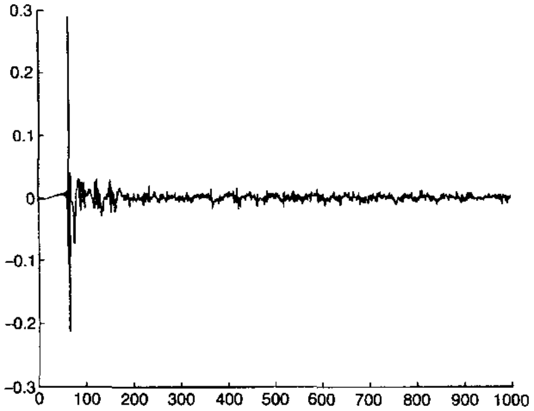

Fig. 1 *Measured filter in normal office room*

Table 2: SNRs of output signals for real recorded signals after convergence (dB)

Signal	Noise	Initial SNRs	LMS algorithm	ICA-based approach
Speech	Car	−3.0	21.0	26.8
Speech	Speech	−3.0	21.5	38.7
Speech	Music	−3.0	21.7	41.8

Conclusion: A method for adaptive noise cancelling based on ICA is proposed and the ICA-based learning rule has been derived. The method is compared with the LMS algorithm through experiments for several noise signals and mixing filters. By including higher-order statistics, the proposed ICA-based approach gives better performances than the conventional LMS algorithm.

Acknowledgment: This work was supported by the Brain Science & Engineering Research Program sponsored by Korean Ministry of Science and Technology.

© IEE 2002 *13 November 2001*
Electronics Letters Online No: 20020593
DOI: 10.1049/el:20020593

Hyung-Min Park, Sang-Hoon Oh and Soo-Young Lee (*Brain Science Research Center and Department of Electrical Engineering and Computer Science, Korea Advanced Institute of Science and Technology, Taejon 305-701, Korea*)

E-mail: mhp@eeinfo.kaist.ac.kr

Sang-Hoon Oh: Now at Department of Information Communication Engineering, Mokwon University, Taejon, 302-729, Korea

References

1 WIDROW, B., GLOVER, J.R., MCCOOL, J.M., KAUNITZ, J., WILLIAMS, C.S., HEARN, R.H., ZEIDLER, J.R., DONG, E., and GOODLIN, R.C.: 'Adaptive noise cancelling: principles and applications', *Proc. IEEE*, 1975, **63**, (12), pp. 1692–1716

2 LEE, T.-W.: 'Independent component analysis' (Kluwer Academic Publishers, Boston, 1998)

3 TORKKOLA, K.: 'Blind separation of convolved sources based on information maximization'. IEEE Workshop on Neural Networks for Signal Processing, 1996

4 DOUGLAS, S.C., and MENG, T.H.-Y.: 'Stochastic gradient adaptation under general error criteria', *IEEE Trans. Signal Process.*, 1994, **42**, (6), pp. 1335–1351

5 NAGUMO, J., and NODA, A.: 'A learning method for system identification', *IEEE Trans. Autom. Control*, 1967, **12**, (3), pp. 282–287

Classification of power disturbances using feature extraction in time-frequency plane

J.Y. Lee, Y.J. Won, J.-M. Jeong and S.W. Nam

An efficient feature extraction in the time-frequency plane is proposed for automatic classification of power disturbances. For that purpose, singular value decomposition and principal component analysis are utilised. Finally, the performance of the proposed approach is tested using a maximum likelihood predictor classifier.

Introduction: Power quality (PQ) has been of great concern recently, due to the increase in the number of loads sensitive to power disturbances, whereby a power disturbance corresponds to any deviation from the nominal value of the input AC power characteristics [1–5]. One of the main issues in PQ problems includes how to localise each disturbance event and recognise its respective type in the disturbance group more efficiently. Since power disturbances are finite energy transient or non-stationary signals, it may not be sufficient to analyse them in the time-domain or in the frequency-domain alone. To solve such problems, several signal processing approaches (e.g. [4, 5]) have been reported for the detection and/or classification of power system disturbances, whereby they may not provide good performance in a noisy environment. In this Letter, a new feature extraction approach, based on the joint time-frequency signal representation [6, 7], is proposed for automatic classification of power disturbances, where the time-frequency structure of each disturbance signal is exploited as its distinguishing feature for the recognition of the respective types of power disturbances. Being the two-dimensional representation of a one-dimensional signal, the time-frequency signal representation encodes in a redundant fashion the information of the one-dimensional signal [7]. Thus, for the effective use of joint time-frequency signal representations, it is of practical importance to apply a data compression procedure to the time-frequency representations. In this Letter, the discrete Wigner distribution (WD) is utilised as a bilinear (or quadratic) time-frequency representation, and effective data compression is accomplished by employing (i) singular value decomposition (SVD) of the discrete WD [7] and (ii) principal component analysis (PCA) [4]. This results in an efficient feature vector extraction for the classification of power system disturbances. In general, the tasks to be performed for the automatic classification of power disturbances include the following: (i) capturing each disturbance event (i.e. detection) and (ii) sorting the captured disturbance into various power disturbance groups and identifying its type in the disturbance group (i.e. recognition) [4, 5]. For the automatic detection of each power disturbance event, the stop-and-go cell-average constant-false-alarm-rate (CA-CFAR) detector [3] is employed in this Letter. Then, along with the power level of each detected disturbance, the WD of each detected disturbance, its SVD, and PCA are utilised for efficient feature vector extraction. Finally, 10-class disturbance data, generated using the power system blockset [4], are tested to demonstrate the performance and applicability of the proposed approach, whereby a maximum likelihood predictor (MLP) neural network [8] is employed as a classifier. Also, simulation results obtained by applying the discrete wavelet transform (DWT)-based approach [4] are also provided for purpose of comparison.

Detection of power disturbances: In this Letter, the stop-and-go CA-CFAR detector [3], which is a modified version of the CA-CFAR detector for the use of power disturbance detection, is utilised to localise each transient power disturbance in additive noise. For example, consider a measured power system signal as in Fig. 1*a*, where an impulse waveshape fault waveform is included, and the power system signal is represented with the maximum magnitude of the pure 60 Hz sinusoidal signal part being set to one (i.e. normalised). Then, application of the stop-and-go CA-CFAR detector yields Fig. 1*b*.

Feature extraction using SVD and PCA: Once each power disturbance waveform is detected, it is necessary to extract, from the detected disturbance waveform, properly chosen discriminant information (called a feature vector) for the efficient disturbance classification. In this Letter, a new feature extraction approach, based on the joint time-frequency signal representation [6, 7], is presented. In particular, among bilinear joint time-frequency representations, the WD, which is optimal in many respects [6, 7], is utilised. When $x(n)$ is an N-point signal, its discrete WD, $\boldsymbol{W}_x$, is an $N \times N$ matrix the (n, m)th element of which is given by

$$\boldsymbol{W}_x(n, m) = 2\,\mathrm{Re}\left(\sum_{k=-(L-1)}^{L-1} x(n+k)x^*(n-k)\exp\left(-\frac{j2\pi mk}{L}\right)\right) - x(n)x^*(n) \qquad (1)$$

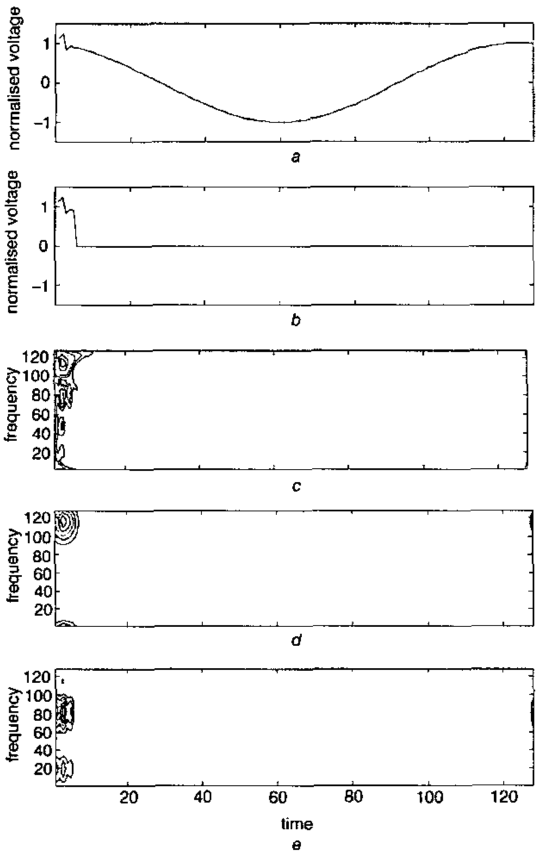

Fig. 1 *Feature extraction using WD and SVD*

a Measured power system signal
b Detected impulse waveshape fault
c Wigner distribution of (*b*)
d $U_1V_1^T$ obtained from SVD of (*c*)
e $U_2V_2^T$ obtained from SVD of (*c*)

In (1), $(n, m) \in \{0, \ldots, N-1\}^2$. However, since the discrete WD includes redundancy due to its two-dimensional representation of a one-dimensional signal, an effective data compression procedure is required, which can be effectively accomplished by means of (i) SVD of $\boldsymbol{W}_x$ and (ii) PCA. More specifically, the SVD of $\boldsymbol{W}_x$ results in the following optimum outer product expansion [7]:

$$\boldsymbol{W}_x = \sum_{i=1}^{N} s_i \boldsymbol{U}_i \boldsymbol{V}_i^T \qquad (2)$$

where $\boldsymbol{U}_i$ and $\boldsymbol{V}_i$ are the $N \times 1$ vectors corresponding to a singular value s_i $(i = 1, 2, \ldots, N)$. For example, in case of the impulse waveshape fault disturbance as in Fig. 1*b*, its WD, $\boldsymbol{U}_1\boldsymbol{V}_1^T$, and $\boldsymbol{U}_2\boldsymbol{V}_2^T$ are presented in Figs. 1*c*–*e*, respectively. Also, in the spectrum of singular values, most energy of each power disturbance is distributed to several major singular values (e.g. s_1, s_2 for power disturbances). From this point of view, $\boldsymbol{U}_1$ and $\boldsymbol{V}_1$ vectors, corresponding to s_1, and $\boldsymbol{U}_2$ and $\boldsymbol{V}_2$ vectors, corresponding to s_2, are chosen, along with the power level l of an N-point disturbance signal, to constitute a potential feature vector $\boldsymbol{F}$ of dimension $4N + 1$:

$$\boldsymbol{F} = [l, \boldsymbol{U}_1^T, \boldsymbol{V}_1^T, \boldsymbol{U}_2^T, \boldsymbol{V}_2^T]^T \qquad (3)$$

In addition, so as to eliminate the redundancy among the elements of (3), Fisher's criterion, which is one of the PCA techniques, is applied to (3) for further data compression, where the degree of distinction of one class from other classes is utilised as the threshold for the appropriate number of feature elements [4].

Classification using MLP classifier: Each extracted feature vector is applied as the input to an MLP, which is one of well-recognised neural network classifiers, to recognise the corresponding class type of it.

Simulations: For the performance test of the proposed approach, 1000 power signal data in additive Gaussian noise (i.e. with 10 classes and 100 data per class; with 40 dB SNR; each data is 128-point long) are generated using the power system blockset of Matlab. Also, an MLP with one hidden layer and 10 output neurons is utilised as a classifier, and 30 data per class (i.e. total: 300 data) are used for training, while 70 data per class (i.e. total: 700 data) are applied for test. Under this condition, the extracted feature, obtained by analysing 300 training data (i.e. 30 data per class), is of a 34 × 1 vector form, where the threshold in the Fisher's criterion is set to 17.0. In this case, the data compression ratio is equal to 15:1 (i.e. reduction from 513 to 34). The simulation results (i.e. classification rates), obtained by applying both the DWT-based approach [4] and the proposed WD-based approach to the 700 test data, are presented in Table 1, where the proposed WD-based approach yields better performance (i.e. 98.0%) than the DWT-based approach (i.e. 93.4%).

Table 1: Simulation results

	Method	DWT-based	WVD-based
Classes		approach	approach
1	Voltage sag	70/70	70/70
2	Voltage swell	62/70	70/70
3	Outage	56/70	70/70
4	Capacitor energising	69/70	70/70
5	Impulse waveshape fault	63/70	64/70
6	Notching	66/70	70/70
7	Harmonic distortion	66/70	66/70
8	Flat top	70/70	70/70
9	UPS	70/70	70/70
10	Phase controlled waveshape	62/70	66/70
Total		93.4%	98.0%

s/t = ratio of s (correctly classified number of data) to t (number of test data)

Conclusion: For the automatic classification of transient or non-stationary power disturbances, WD is utilised as a joint time-frequency representation, and SVD and PCA are employed for effective data compression. The simulation results, obtained using an MLP classifier, demonstrate that the systematic approach results in an efficient feature extraction (e.g. 15:1 compression ratio), also leading to good classification performance.

Acknowledgment: The work was supported by the Korea Research Foundation Grant (KRF-2001-013-E00056)

© IEE 2002 22 February 2002
Electronics Letters Online No: 20020562
DOI: 10.1049/el:20020562

J.Y. Lee, Y.J. Won, J.-M. Jeong and S.W. Nam (*Division of Electrical and Computer Engineering, Hanyang University, Seoul, 133-791, Korea*)

E-mail: swnam@hanyang.ac.kr

References

1 DUGAN, R.C., MCGRANAGHAN, M.F., and BEATY, H.W.: 'Electrical power system quality' (McGraw-Hill, New York, 1996)
2 KEY, T.S., *et al.*: 'IEEE recommended practice for powering and grounding sensitive electronic equipment' (IEEE, New York, 1995)
3 CHUNG, J.H.: 'Applications of digital signal processing to electric power quality and wireless communications'. Ph.D. Dissertation, The University of Texas at Austin, May 2000
4 LEE, C.H., and NAM, S.W.: 'Efficient feature vector extraction for automatic classification of power quality disturbances', *Electron. Lett.*, 1998, **34**, pp. 1059–1061
5 WILKINSON, W.A.: 'Discrete wavelet analysis of power system transients', *IEEE Trans. Power Deliv.*, 1996, **11**, (4), pp. 2038–2044
6 QIAN, S., and CHEN, D.: 'Joint time-frequency analysis' (Prentice-Hall, New Jersey, 1996)
7 MECKLENBRÄUKER, W., and HLAWATSCH, F. (Eds): 'The Wigner distribution – theory and applications in signal processing' (Elsevier, Amsterdam, The Netherlands, 1997)
8 HAYKIN, S.: 'Neural networks' (Prentice-Hall, New Jersey, 1999)

High-speed division architecture for $GF(2^m)$

Chang Hoon Kim and Chun Pyo Hong

A new division architecture for $GF(2^m)$ with standard basis representation is presented. The proposed architecture is based on a modified version of the binary extended greatest common divisor algorithm; it reduces computational delay time and hardware complexity.

Introduction: The implementation of an elliptic curve public key cryptosystem requires division in $GF(2^m)$. Although the division operation can easily be implemented using software, it would be too slow for time-critical applications. Thus many approaches and architectures have been proposed to realise it using hardware [1–3]. Fermat's theorem or Euclid's greatest common divisor (GCD) algorithm or the solution of a set of linear equations can be used to compute division in $GF(2^m)$. Recent research results show Euclid's GCD algorithm is the best choice to compute division using hardware [1–3]. As a result of our research, described in this Letter, we propose a high-speed and low-complexity division architecture for $GF(2^m)$ with standard basis representation (SBR). This architecture is based on a modified version of the binary method [4] that is a different representation of Euclid's GCD algorithm.

New algorithm for division in $GF(2^m)$: Let $A(x)$ and $B(x)$ be two elements in $GF(2^m)$, $G(x)$ be the primitive polynomial used to generate the field and $P(x)$ be the result of the division $[A(x)/B(x)$ MOD $G(x)]$. For each polynomial, the coefficients are binary digits 0 and 1:

$$\begin{aligned}A(x) &= a_{m-1}x^{m-1} + a_{m-2}x^{m-2} + \cdots + a_1x + a_0\\ B(x) &= b_{m-1}x^{m-1} + b_{m-2}x^{m-2} + \cdots + b_1x + b_0\\ G(x) &= x^m + g_{m-1}x^{m-1} + g_{m-2}x^{m-2} + \cdots + g_1x + g_0\\ P(x) &= p_{m-1}x^{m-1} + p_{m-2}x^{m-2} + \cdots + p_1x + p_0\end{aligned} \tag{1}$$

In previous research [4], the division result $P(x)$ is obtained using the binary extended GCD algorithm. The algorithm is described as follows:

```
Input: $G(x)$, $A(x)$, $B(x)$
Output: $V$ has $P(x)=A(x)/B(x)$ MOD $G(x)$
Initialise: $R=B(x)$, $S=G=G(x)$, $U=A(x)$, $V=0$;
 1  while $R \neq 0$ do
 2     while $r_0 = = 0$ then
 3        $R=R/x$, $U=U/x$ MOD $G$;
 4     while $s_0 = = 0$ do
 5        $S=S/x$, $V=V/x$ MOD $G$;
 6     end while
 7     if $S \geq R$ then
 8        $(S, R)=(R+S, R)$, $(V, U)=(U+V, U)$;
 9     else
10        $(S, R)=(S, R+S)$, $(V, U)=(V, U+V)$;
11     end if
12  end while
```

This algorithm is based on five simple facts described in (2–6):

If both S and R are even, then $\mathrm{GCD}(S, R) = x\mathrm{GCD}(S/x, R/x)$ (2)

If S is odd and R is even, then $\mathrm{GCD}(S, R) = \mathrm{GCD}(S, R/x)$ (3)

$$\mathrm{GCD}(S, R) = \mathrm{GCD}(R, S - R) \tag{4}$$

$$\mathrm{GCD}(S, R) = \mathrm{GCD}(S, S - R) \tag{5}$$

If both S and R are odd, then $S - R$ is even (6)

In (6), since subtraction and addition are bit-wise exclusive-OR (XOR) operation in $GF(2^m)$, (7) is satisfied:

$$S - R = R - S = S + R \tag{7}$$

Although the algorithm described above is simple, it is difficult to realise with hardware since the number of iterations is not fixed. In addition, it requires process of comparisons relative to R and S. We solve such problems without affecting basic functions of the binary extended GCD algorithm. The resulting algorithm is described as follows:

```
Input: $G(x)$, $A(x)$, $B(x)$
Output: $V$ has $P(x)=A(x)/B(x)$ MOD $G(x)$
Initialise: $R=B(x)$, $S=G=G(x)$, $U=A(x)$, $V=0$;
            $count=0$, $state=0$;
 1  for $i=1$ to $2m$ do
 2     if $state = = 0$ then
 3        $count = count + 1$;
 4        if $r_0 = = 1$ then
 5           $(R, S)=(R+S, R)$, $(U, V)=(U+V, U)$;
 6           $state = 1$;
 7        end if
 8     else
 9        $count = count - 1$;
10        if $r_0 = = 1$ then
11           $(R, S)=(R+S, S)$, $(U, V)=(U+V, V)$;
12        end if
13        if $count = = 0$ then
14           $state = 0$;
15        end if
16     end if
17     $R=R/x$, $U=U/x$ MOD $G$;
18  end for
```

The differences between the two algorithms are summarised as follows:

(i) In the first algorithm, since the initial value of S equals to $G(x)$, it is always odd at first. In addition, depending on the value of R, we apply two different conditions to get $\mathrm{GCD}(S, R)$. If R is even, we apply (3). If R is odd, we first compute $(S - R)$, and then apply (4) or (5). In this case, the resulting value of S is always odd, and we only need to check the value of R, whether it is even or odd. Based on this result, steps 4, 5 and 6 (see first algorithm) can be removed.

(ii) In the first of the two algorithms described above, since the degree of S is m and the degree of R is less than m, if we reduce the degree of R or S by one for each iteration step, the algorithm will be terminated after $2m$ iterations. In addition, comparison is required as described in step 7 of the first algorithm. In the new algorithm, instead of comparing the value of R and S, we use a different approach. As described in the second algorithm above, we provide new variables *count* and *state*. The variable *count* is used for tracking the difference of degree between R and S, and the variable *state* is used for identifying which one has the larger degree between R and S.

In the second algorithm, depending on the value of *state* and R, we apply four different conditions to get $\mathrm{GCD}(S, R)$. First, when the variable *state* is